



HAL
open science

De la business intelligence interne vers la business intelligence dans le cloud : modèles et apports méthodologiques

Odette Sangupamba Mwilu

► **To cite this version:**

Odette Sangupamba Mwilu. De la business intelligence interne vers la business intelligence dans le cloud : modèles et apports méthodologiques. Web. Conservatoire national des arts et métiers - CNAM, 2018. Français. NNT : 2018CNAM1168 . tel-01793649

HAL Id: tel-01793649

<https://theses.hal.science/tel-01793649v1>

Submitted on 16 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**ÉCOLE DOCTORALE INFORMATIQUE, TELECOMMUNICATIONS ET
ELECTRONIQUE**

**CENTRE D'ETUDE ET DE RECHERCHE EN INFORMATIQUE ET
COMMUNICATIONS**

THÈSE présentée par :

Odette SANGUPAMBA MWILU

soutenue le : **03 AVRIL 2018**

pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**
Discipline/ Spécialité : Informatique

**De la business intelligence interne vers la business
intelligence dans le cloud : Modèles et apports
méthodologiques**

THÈSE dirigée par :

**PR. Isabelle Wattiau
DR. Nicolas Prat**

Professeur, Cnam-Paris
Professeur-Associé HDR, ESSEC Business School

RAPPORTEURS :

**DR Fadila Bentayeb
PR. Christine Legner**

Professeur-Associé HDR, Université Lumière Lyon 2
Professeur, Université de Lausanne

EXAMINATEURS :

**PR. Jacky Akoka
PR. Olivier Teste**

Président de Jury, Professeur émérite, Cnam-Paris
Professeur, IRIT

*A la mémoire de mon père Armand SANGUPAMBA MUKANGU,
A la mémoire de ma mère Constantine MAHENGANU MAMONA,
A la mémoire du Cardinal Joseph Albert MALULA,
A la mémoire de Monseigneur Eugène MOKE,
Je dédie ce travail.*

Remerciements

Il nous est un agréable devoir de rendre hommage à tous ceux qui nous ont apporté une assistance durant notre formation et en particulier dans l'élaboration de ce travail.

Notre gratitude s'adresse avant tout au professeur Isabelle Wattiau et au professeur associé Nicolas Prat qui n'ont ménagé aucun effort pour nous initier au métier de chercheur. C'est grâce à leurs conseils et encouragements que nous avons pu rédiger cette thèse.

Nous remercions le professeur Christine Legner et le professeur associé Fadila Bentayeb, rapporteurs, et leurs collègues Jacky Akoka et Olivier Teste membres du jury, pour leur participation à l'évaluation de ce travail.

Puissent tous les membres de l'équipe ISID trouver ici l'expression de notre reconnaissance pour leur encadrement et leurs encouragements. Nous pensons en particulier à nos amis, les doctorants Quentin, Pierre Henry, Subhi, Anh et Noura.

Que tous nos frères et sœurs SANGUPAMBA, leurs conjoints ainsi que leurs enfants trouvent ici l'expression de notre profonde gratitude pour leurs affection, soutien et encouragements tout au long de notre formation.

A notre famille religieuse, la Congrégation des sœurs de Sainte Thérèse de l'Enfant Jésus de Kinshasa, nous exprimons toute notre reconnaissance pour le soutien moral et spirituel.

Sans l'assistance financière du ministère du budget de la République Démocratique du Congo, nous ne serions pas parvenue au bout de notre formation post-universitaire. C'est ici l'occasion pour nous de rendre un hommage vibrant et d'exprimer toute notre gratitude à tous ses cadres et agents.

Enfin, nos remerciements s'adressent à toute la famille MBUYI MUKADI et à tous nos amis pour l'attention dont nous avons toujours été bénéficiaire.

Résumé

La BI et le cloud computing sont deux grands sujets de recherche en informatique et en système d'information en particulier. Une recherche combinant ces deux concepts est d'un intérêt double : D'une part, dans les entreprises, la BI devient de plus en plus une partie importante du système d'information qui nécessite des investissements en termes de performances de calcul et des volumes de données. D'autre part, le cloud computing offre de nouvelles opportunités pour gérer les données à des fins d'analyse.

Etant donné les possibilités de cloud, la question de la migration de l'ensemble du système d'information y compris la BI est d'un grand intérêt. En particulier, les chercheurs doivent fournir aux professionnels des modèles et des méthodes qui puissent les aider à migrer vers le cloud.

Que faire pour que la BI puisse fournir aux managers un service de mise à disposition de données d'analyse au travers du cloud ? La question de recherche est : Comment aider les organisations à migrer leur BI vers le cloud ?

Dans cette thèse, nous répondons à cette question en utilisant l'approche science de conception (design science). Nous mettons en place une aide à la décision de la migration de la BI vers le cloud qui s'appuie sur les taxonomies. Nous proposons un modèle de guidage opérationnel qui est instancié par une taxonomie de la BI dans le cloud et dont découlent les règles pour la migration de la BI vers le cloud.

Mots clés : business intelligence, cloud computing, science de conception, taxonomies, guidage opérationnel, artefacts.

Résumé en anglais

BI and cloud computing are two major areas of computer science research and in particular in information system. A research combining these two concepts has a double interest : On the one hand, in business, the BI becomes increasingly an important part of the information system which requires investment in terms of computing performance and data volumes. On the other hand, cloud computing offers new opportunities to manage data for analysis.

Given the possibilities of cloud, migration question of the information system including BI is of great interest. In particular, researchers must provide models and methods to help professional in BI migration to the cloud.

The research question is : how to migrate BI to the cloud?

In this thesis, we address this issue using design science research approach. We implement a decision-making help for BI migration to the cloud based on taxonomies. We provide an operational guidance model that is instantiated by a BI taxonomy in the cloud and from that rules for BI migration to the cloud are arised.

Keywords : business intelligence, cloud computing, design science, taxonomies, operational guidance, artifacts.

TABLE DES MATIERES

RESUME	III
RESUME EN ANGLAIS	IV
TABLE DES MATIERES	V
LISTE DES FIGURES	X
LISTE DES TABLEAUX	XI
LISTE DES ANNEXES	XII
CHAPITRE I : INTRODUCTION.....	1
1.1. CONTEXTE	1
1.2. OBJET DE LA RECHERCHE.....	2
1.3. APPROCHE.....	3
1.4. APERÇU DE LA SOLUTION	4
1.5. MODELE DES CONTRIBUTIONS ET PLAN DE LA THESE	8
CHAPITRE 2 : BUSINESS INTELLIGENCE DANS LE CLOUD : ETAT DE L'ART	12
2.1. INTRODUCTION.....	12
2.2. TYPLOGIE DES ARTEFACTS EN SCIENCE DE CONCEPTION	12
2.3. REVUE SYSTEMATIQUE DE LA LITTERATURE SELON KITCHENHAM	15
2.3.1. <i>Objectifs de la revue systématique de la littérature.....</i>	<i>15</i>
2.3.2. <i>Importance de la revue systématique</i>	<i>15</i>
2.3.3. <i>Propriétés des revues systématiques</i>	<i>16</i>
2.3.4. <i>Planification de la revue systématique.....</i>	<i>16</i>
2.3.4.1. Identification de la nécessité de la revue systématique.....	16
2.3.4.2. Protocole de revue	17
2.3.5. <i>Conduite de la revue systématique</i>	<i>17</i>
2.3.5.1. Identification de la recherche.....	18
2.3.5.2. Sélection des études.....	18
2.3.5.3. Evaluation de la qualité des études.....	18
2.3.5.4. Extraction des données	18
2.3.5.5. Synthèse des données	18
2.4. APPLICATION DU PROCESSUS DE REVUE SYSTEMATIQUE SELON KITCHENHAM A NOTRE CAS.....	18
2.4.1. <i>Objectifs de la revue systématique de la littérature de la BI dans le cloud</i>	<i>19</i>
2.4.2. <i>Importance de la revue systématique pour notre travail</i>	<i>19</i>
2.4.3. <i>Propriétés de la revue systématique de la littérature sur la BI dans le cloud.....</i>	<i>19</i>
2.4.4. <i>Planning de la revue systématique.....</i>	<i>20</i>
2.4.4.1. Nécessité d'un examen systématique	20

VI

2.4.4.2.	Protocole de revue de la littérature de la BI dans le cloud.....	20
2.4.5.	<i>Conduite de la revue systématique</i>	22
2.4.5.1.	Identification de la recherche.....	22
2.4.5.2.	Sélection des études.....	24
2.4.5.3.	Evaluation de la qualité des études.....	26
2.4.5.4.	Extraction des données	26
2.5.	RESULTAT DE L'ETAT DE L'ART	27
2.5.1.	<i>Business intelligence</i>	27
2.5.1.1.	Une architecture de la BI	28
2.5.1.2.	Analyse des Big data	33
2.5.2.	<i>Cloud computing</i>	36
2.5.2.1.	L'architecture de cloud	37
2.5.2.2.	Les types de cloud	38
2.5.2.3.	Cloud computing mobile	38
2.5.3.	<i>BI dans le cloud</i>	39
2.5.3.1.	La collecte et la consolidation des données	51
2.5.3.2.	La modélisation et le stockage des données	52
2.5.3.3.	L'analytique	54
2.6.	OPPORTUNITES ET QUESTION DE RECHERCHE	56
2.7.	CONCLUSION.....	60
CHAPITRE 3 : MODELE DE TAXONOMIE, OPERATEURS ET REGLES		61
3.1.	INTRODUCTION.....	61
3.2.	MODELE D'UNE TAXONOMIE A DIMENSIONS HIERARCHIQUES	62
3.3.	DEFINITION FORMELLE D'UNE TAXONOMIE	65
3.4.	LES OPERATEURS ALGEBRIQUES	66
3.4.1.	<i>L'opérateur de création de nouveaux éléments de la taxonomie</i>	68
3.4.2.	<i>Les opérateurs de transformation des éléments de la taxonomie</i>	69
3.4.2.1.	La promotion	70
3.4.2.2.	La rétrogradation.....	70
3.4.2.3.	La scission	71
3.4.2.4.	La fusion	72
3.5.	DU MODELE DE TAXONOMIE AU MODELE DE GUIDAGE OPERATIONNEL	73
3.6.	SYNTAXE EBNF DE MODELE DE GUIDAGE OPERATIONNEL	78
3.7.	COHERENCE, MINIMALITE ET EXHAUSTIVITE DES REGLES DE GUIDAGE OPERATIONNEL	79
3.7.1.	<i>Cohérence</i>	80
3.7.2.	<i>Minimalité</i>	80
3.7.3.	<i>Exhaustivité</i>	81
3.8.	CONCLUSION.....	81

CHAPITRE 4 : METHODOLOGIE DE DEVELOPPEMENT DE TAXONOMIES POUR LES TECHNOLOGIES COMPLEXES EMERGENTES	83
4.1. INTRODUCTION.....	83
4.2. METHODOLOGIE DE DEVELOPPEMENT DE TAXONOMIES DE NICKERSON, VARSHNEY ET MUNTERMANN	83
4.2.1. <i>Présentation de la méthodologie</i>	84
4.2.1.1. Déterminer une méta-caractéristique.....	86
4.2.1.2. Déterminer les conditions de fin	86
4.2.1.3. Choisir une approche et l'appliquer	87
4.2.1.4. Arrêter le processus	87
4.2.2. <i>Difficultés d'application aux technologies complexes émergentes</i>	88
4.3. ADAPTATION DE LA METHODOLOGIE	89
4.4. PRESENTATION DE NOTRE METHODOLOGIE.....	89
4.4.1. <i>Déterminer le sujet de la recherche</i>	92
4.4.2. <i>Déterminer l'objectif de la taxonomie</i>	93
4.4.3. <i>Identifier de la recherche</i>	93
4.4.4. <i>Sélectionner des études</i>	95
4.4.5. <i>Examiner les sources</i>	97
4.4.6. <i>Structurer la taxonomie</i>	97
4.4.7. <i>Vérifier si la source examinée est la dernière de l'échantillon</i>	97
4.4.8. <i>Vérifier la convergence</i>	98
4.4.9. <i>Valider la taxonomie</i>	98
4.5. CONCLUSION.....	99
CHAPITRE 5 : DEVELOPPEMENT D'UNE TAXONOMIE DE LA BUSINESS INTELLIGENCE DANS LE CLOUD	101
5.1. INTRODUCTION.....	101
5.2. APPLICATION DE LA METHODOLOGIE PROPOSEE.....	101
5.2.1. <i>Sujet de la recherche</i>	101
5.2.2. <i>Objectif de la taxonomie</i>	102
5.2.3. <i>Identification de la recherche</i>	102
5.2.4. <i>Sélection des sources</i>	102
5.3.5. <i>Examen des sources</i>	104
5.3.5.1. Taxonomie de contexte.....	105
5.3.5.2. Taxonomie de décision	109
5.3.6. <i>Structuration de la taxonomie</i>	114
5.3.7. <i>Dernière source</i>	115
5.3.8. <i>Convergence</i>	115
5.3.9. <i>Validation de la taxonomie</i>	116
5.3.9.1. Sources et critères de sélection.....	117
5.3.9.2. Examen des sources	118

VIII

5.4.	TAXONOMIES RESULTANTES	122
5.4.1.	<i>Taxonomie de contexte</i>	123
5.4.1.1.	Exigences de sécurité	124
5.4.1.2.	Ressources financières	124
5.4.1.3.	Objectifs	125
5.4.1.4.	Secteur d'activité.....	125
5.4.1.5.	Taille de l'entreprise	125
5.4.1.6.	Ressources humaines	126
5.4.2.	<i>Taxonomie de décision</i>	127
5.4.2.1.	Niveau de migration	128
5.4.2.2.	Couches de cloud	128
5.4.2.3.	Contrat	129
5.4.2.4.	Voie de transmission des services	129
5.4.2.5.	Types de cloud.....	130
5.4.2.6.	Supports des applications BI.....	130
5.5.	CONCLUSION.....	130
CHAPITRE 6: AIDE A LA DECISION DE MIGRATION DE LA BUSINESS INTELLIGENCE VERS LE CLOUD.....		132
6.1.	INTRODUCTION.....	132
6.2.	LES REGLES	132
6.2.1.	<i>Les règles proposées par la littérature</i>	134
6.2.2.	<i>Les règles déduites de la littérature</i>	137
6.2.3.	<i>Les règles d'association</i>	139
6.3.	COHERENCE, MINIMALITE ET EXHAUSTIVITE DE LA BASE DE REGLES	143
6.3.1.	<i>Cohérence</i>	143
6.3.2.	<i>Minimalité</i>	145
6.3.3.	<i>Exhaustivité</i>	147
6.4.	CONCLUSION.....	147
CHAPITRE 7 : IMPLEMENTATION ET VALIDATION		149
7.1.	INTRODUCTION.....	149
7.2.	ARCHITECTURE DE JESS	149
7.3.	ARCHITECTURE DE L'OUTIL DE GUIDAGE OPERATIONNEL	150
7.3.1.	<i>Les taxonomies</i>	151
7.3.2.	<i>Le système expert</i>	152
7.3.2.1.	Mémoire de travail	152
7.3.2.2.	Base de règles.....	155
7.3.2.3.	Moteur d'inférence	157
7.3.3.	<i>L'interface</i>	158
7.4.	DIAGRAMME DE CAS D'UTILISATION	161

7.5.	VALIDATION DE L'APPROCHE.....	162
7.5.1.	<i>Scénario illustratif</i>	162
7.5.1.1.	Enoncé.....	162
7.5.1.2.	Résolution.....	163
7.5.2.	<i>Cohérence et exhaustivité de la base de règles</i>	165
7.6.	CONCLUSION.....	166
	CHAPITRE 8 : CONCLUSION ET PERSPECTIVES.....	168
	PUBLICATIONS	171
	BIBLIOGRAPHIE	172

LISTE DES FIGURES

FIGURE 1: MODELE DE LA SOLUTION	8
FIGURE 2 : ARCHITECTURE DE LA BI (TRADUIT ET ADAPTE DE PHILLIPS-WREN ET AL. (2015))	29
FIGURE 3 : FONCTIONNEMENT DE MAPREDUCE (2017).....	35
FIGURE 4 : CADRE DE LA MIGRATION DE LA BI VERS LE CLOUD (TRADUIT ET ADAPTE DE MURIITHI ET KOTZE (2013))	45
FIGURE 5 : CADRE DES SERVICES DONT LA BI BENEFICIE DANS LE CLOUD (BAARS AND KEMPER 2010).....	46
FIGURE 6 : MODELE DE TAXONOMIE A DIMENSIONS HIERARCHIQUES	64
FIGURE 7 : MODELE DE GUIDAGE OPERATIONNEL	76
FIGURE 8: METHODOLOGIE DE DEVELOPPEMENT DE TAXONOMIES DE NICKERSON, VARSHNEY ET MUNTERMANN (2013)	85
FIGURE 9: NOTRE METHODOLOGIE DE DEVELOPPEMENT DE TAXONOMIES	92
FIGURE 10 : EXAMEN DES SOURCES DE TYPE 1 : NOMBRE D'OPERATIONS	115
FIGURE 11 : TAXONOMIE DE CONTEXTE	123
FIGURE 12: TAXONOMIE DE DECISION	127
FIGURE 13 : INTERFACE RATTLE	140
FIGURE 14 : EXTRAIT DU FICHIER EXCEL	141
FIGURE 15 : ARCHITECTURE DE JESS ADAPTE ET TRADUIT DE FRIEDMAN-HILL (2003).....	150
FIGURE 16 : ARCHITECTURE DE L'OUTIL DE GUIDAGE OPERATIONNEL	151
FIGURE 17 : OUTIL DE GUIDAGE OPERATIONNEL : INTERFACE UTILISATEUR.....	160
FIGURE 18: DIAGRAMME DE CAS D'UTILISATION DE L'OUTIL DE GUIDAGE OPERATIONNEL	161
FIGURE 19: DECISION DE MIGRATION VERS LE CLOUD DE LA CII	165

LISTE DES TABLEAUX

TABLEAU 1: TYPOLOGIE DES ARTEFACTS EN SCIENCE DE CONCEPTION (SANGUPAMBA MWILU, PRAT AND COMYN-WATTIAU 2014)	14
TABLEAU 2 : EXTRAIT DU CADRE DE REFERENCE DE LA BI DANS LE CLOUD	42
TABLEAU 3 : SYNTHESE DES OPPORTUNITES DE RECHERCHE	58
TABLEAU 4 : TAXONOMIE DE CONTEXTE : ITERATIONS DES SOURCES DE TYPE 1	108
TABLEAU 5 : TAXONOMIE DE DECISION : ITERATIONS DES SOURCES DE TYPE 1	114

LISTE DES ANNEXES

I. CLOUD BI FRAMEWORK.....	188
II. EXAMEN DES SOURCES DE TYPE 2 ET TYPE 3	192
III. LES REGLES	201

CHAPITRE I : INTRODUCTION

1.1. Contexte

Le système d'information (SI) est "*un ensemble organisé de ressources : matériel, logiciel, personnel, données, procédures...permettant d'acquérir, de traiter, de stocker des informations (sous forme de données, textes, images, sons...) dans et entre des organisations*" (Reix 2000). Le SI assure donc quatre fonctions au sein d'une organisation : acquérir, traiter, stocker et communiquer les informations. Pour accomplir ces fonctions, le SI a besoin de supports tels que les technologies de l'information.

Il incombe aux chercheurs en SI de promouvoir les connaissances qui permettent l'application des technologies de l'information dans les organisations. Les chercheurs en SI ont pour mission de développer et communiquer des connaissances concernant la gestion des technologies de l'information et leur utilisation à des fins managériales et organisationnelles.

La recherche en SI est à la fois une discipline organisationnelle et une discipline technique qui se consacre à l'analyse, à la construction, au déploiement, à la gestion et à l'évaluation des artefacts du système d'information (Hevner *et al.* 2004). Elle est une discipline organisationnelle car elle analyse et évalue les besoins des entreprises par rapport à leurs processus opérationnels. Elle est une discipline technique car elle utilise les technologies existantes pour combler les besoins des entreprises.

En outre, le domaine des SI suscite plusieurs thèmes de recherche en rapport avec les technologies de l'information. Un des thèmes d'actualité (aussi bien pour les chercheurs que pour les praticiens) est la transformation numérique (des SI). Celle-ci vise notamment à répondre aux besoins d'évolution des SI des organisations. Elle s'inscrit dans le défi de la révolution numérique et de son impact sur les organisations.

Cette transformation numérique (des SI) s'accompagne de quelques termes phares tels que le big data, le cloud computing, la business intelligence (BI), l'analytique, la mobilité, etc.

Elle implique un changement de paradigme au sein des organisations. Ces dernières font de plus en plus le choix de l'externalisation, notamment l'adoption de solutions cloud.

En outre, la transformation numérique (des SI) favorise la combinaison de certains domaines tels que la business intelligence et le cloud, le big data et l'analytique.

Notre thèse dont le sujet est "*de la BI interne à la BI dans le cloud : modèles et apports méthodologiques*" se place dans le cadre de la recherche en système d'information et précisément dans l'optique de ce thème d'actualité qu'est la transformation numérique.

La BI est un terme général. Elle comprend les applications, l'infrastructure, les outils et les meilleures pratiques qui permettent l'accès à l'information en vue de son analyse pour améliorer et optimiser les décisions (Gartner IT Glossary).

Le cloud computing, quant à lui, est un modèle technologique qui permet un accès, via un réseau de télécommunications, à la demande et en libre-service, à des ressources informatiques partagées configurables (exemple : réseaux, serveurs, stockage, applications et services) qui peuvent être rapidement approvisionnées et diffusées avec un effort minimal de gestion ou d'interaction avec le fournisseur de services (Yang and Tate 2012).

Ainsi, la migration de la BI vers le cloud permet à la BI de tirer profit des opportunités qu'offre le cloud computing.

1.2. Objet de la recherche

La BI et le cloud computing sont deux grands sujets de recherche en SI à l'ère de la transformation numérique. Selon les dernières prévisions de Gartner, à la fin de l'année 2020, le marché de la BI et analytiques devrait atteindre 22,8 milliards de dollars (Gartner 2017a), tandis que le marché du cloud public¹ atteindra 383 milliards (Gartner 2017b). Une recherche combinant ces deux concepts est d'un intérêt double :

¹ On parle de cloud public par opposition au cloud privé. En général, dans le cloud public, il y a une mutualisation de ressources externalisées qui sont mises à disposition par un fournisseur tiers. Le cloud privé préfère appliquer les mêmes techniques mais en interne, sur une infrastructure mise en place et gérée directement par l'entreprise.

- D'une part, le cloud computing offre de nouvelles opportunités de gestion des données à des fins d'analyse. Il fournit des infrastructures matérielles et logicielles ainsi que des espaces de calcul qui semblent illimités pour les clients (Baars and Kemper 2010).
- D'autre part, la BI apporte une valeur ajoutée au grand volume de données contenues dans le cloud, on parle de l'analyse de big data (big data analytics). Grâce aux techniques et bonnes pratiques de la BI, on peut identifier de nouvelles connaissances dans les masses de données hétérogènes contenues dans le cloud (Nayem, Fahad and Akhter 2013).

Etant données les possibilités du cloud, les entreprises doivent se poser la question d'y migrer (l'ensemble de) leur système d'information, y compris la BI. Ainsi, les chercheurs doivent fournir aux professionnels des modèles et des méthodes qui puissent les aider à migrer leur BI vers le cloud.

Comment aider les organisations à migrer leur BI vers le cloud ? Notre principale question de recherche est : "Quels sont les nouveaux artefacts à développer pour que la BI tire profit du cloud computing ?".

Cette thèse devra proposer des modèles et des méthodes qui puissent aider les organisations désireuses de passer de la BI interne à la BI dans le cloud comme l'énonce son sujet : "*de la BI interne à la BI dans le cloud : modèles et apports méthodologiques*". Pour ce faire, nous avons adopté une approche de recherche en système d'information que nous présentons à la section suivante.

1.3. Approche

Comme nous l'avons dit plus haut, notre travail se place dans le cadre de la recherche en Système d'Information (SI). Il existe deux paradigmes principaux pour la recherche en SI (Hevner *et al.* 2004) : "science de conception" et "science du comportement". Le premier tient son origine de l'ingénierie et des sciences de l'artificiel. Il produit les artefacts pour résoudre un problème pertinent non encore résolu jusque-là. Les artefacts ainsi produits sont évalués afin de vérifier leur qualité et leur efficacité. Ce premier paradigme est prescriptif. Le second tient son origine des sciences naturelles, il développe et justifie des théories. La recherche en science de conception dans les systèmes d'information vise à étendre les limites

des capacités humaines et organisationnelles en créant des artefacts nouveaux et innovants (Hevner *et al.* 2004). Alors que la recherche en science du comportement vise à comprendre, l'objectif principal de la recherche en science de conception est l'utilité. La recherche en science de conception est maintenant établie comme un paradigme majeur de recherche en SI (Goes 2014). De nombreux chercheurs soutiennent l'idée que la finalité de la recherche en science de conception est la production d'un ou plusieurs artefacts, même si d'autres affirment que son but devrait être la création de théories (Gregor and Jones 2007). Dans notre travail, nous nous plaçons dans le domaine de la recherche en science de conception et adoptons le premier point de vue selon lequel son objectif principal est la production d'artefacts utiles car, nous voulons développer de nouveaux artefacts utiles pour la migration de la BI vers le cloud. Ainsi, ce paradigme nous aidera à répondre concrètement à notre question de recherche qui est "Quels sont les nouveaux artefacts à développer pour que la BI tire profit du cloud computing ?".

Eu égard à ce qui vient d'être dit, cette thèse devra aboutir à la mise en place de nouveaux artefacts pour enrichir la base des connaissances sur la BI dans le cloud. Ainsi dans la section suivante, nous proposons un aperçu général du processus de réponse à notre question de recherche.

1.4. Aperçu de la solution

Comme nous l'avons précisé au point 1.3, nous adoptons l'approche de "design science" (science de conception). Cette approche nous conduit à nous poser la question : Quels artefacts pouvons-nous développer pour répondre concrètement à notre question de recherche ?

Grâce à l'état de l'art sur la BI dans le cloud, nous nous sommes rendus compte que différents artefacts ont été développés pour que la BI tire profit des potentialités du cloud. En vue d'apporter une aide à la décision de migration de la BI vers le cloud, nous nous proposons de développer une taxonomie de la BI dans le cloud.

Etant donné qu'il existe déjà des méthodologies de développement de taxonomies en SI, nous nous sommes inspirés de l'une d'elles assez récente et qui capitalise sur les précédentes.

Il s'agit de la méthodologie de développement de taxonomies de Nickerson, Varshney et Muntermann (2013).

Selon Nickerson, Varshney et Muntermann, une taxonomie comporte deux niveaux : dimension et caractéristique. Ainsi une taxonomie permet de décrire des objets. Chaque objet peut être décrit selon plusieurs dimensions. Pour un objet, une dimension n'a qu'une caractéristique et ses caractéristiques sont mutuellement exclusives.

Exemple 1 : Dans ce premier exemple, nous présentons un extrait de la taxonomie de gestion des risques dans le processus de développement des solutions informatiques au sein des organisations. Cette taxonomie, proposée par Herzfeldt et al (2012), comprend entre autres les trois dimensions suivantes :

- *Environnement des risques*
- *Origine des risques*
- *Risques liés à la solution informatique*

Chaque dimension possède des caractéristiques. Dans la dimension *Risques liées à la solution informatique*, on trouve les caractéristiques suivantes :

- *risque de gestion*
- *risque lié à l'instabilité des besoins*
- *risque lié aux personnes*
- *risque lié à l'appel d'offres*
- *risque lié au contrat*
- *risque lié à l'analyse*
- *risque lié à la conceptualisation*
- *risque lié à la conception et au développement*
- *risque lié à l'intégration*

Dans cette taxonomie, un objet est un risque. Chaque risque ne peut avoir qu'une seule caractéristique dans cette dimension.

Exemple 2 : Ce deuxième exemple est un extrait d'une taxonomie des sites web dynamiques proposée par Gillenson, Sherrell et Chen (2000). Cette taxonomie comprend entre autres les dimensions *types de pages web* et *fonctions de pages web*. La dimension *types de pages web* comprend les caractéristiques ci-après :

- *page de garde*
- *plan de site*
- *page de base*
- *page intermédiaire*
- *page terminale*

Ici, un objet est une page web. Chaque page web ne peut avoir qu'un seul type.

Ce modèle de taxonomie à deux niveaux s'appuie sur l'hypothèse selon laquelle toutes les instances d'une dimension ont des caractéristiques homogènes. Or, en réalité, il existe des instances d'une même dimension qui ne peuvent pas avoir des caractéristiques homogènes. Pour exprimer cela, nous avons regroupé les caractéristiques en catégories. Ainsi, les instances d'une dimension peuvent être réparties en catégories. Pour ce, il importe d'enrichir le modèle de taxonomie de Nickerson, Varshney et Muntermann et de le rendre plus expressif. A cet effet, nous proposons un modèle de taxonomie à dimensions hiérarchiques. Nous introduisons le concept de catégorie. Chaque catégorie regroupe les instances d'une dimension qui ont des caractéristiques homogènes. Ainsi, une catégorie est un niveau intermédiaire entre une dimension et une caractéristique.

Exemple : Dans la dimension *risques liés à la solution informatique*, on trouve trois catégories de risques :

- *risques liés à la gestion de projet*
- *risques antérieurs au projet*
- *risques liés aux phases du projet*

Les caractéristiques de *risques liés à la gestion de projet* sont :

- *risque de gestion*
- *risque lié à l'instabilité des besoins*
- *risque lié aux personnes*

Les caractéristiques de *risques antérieurs au projet* sont :

- *risque lié à l'appel d'offres*
- *risque lié au contrat*

Les caractéristiques de *risques liés aux phases du projet* sont :

- *risque lié à l'analyse*

- *risque lié à la conceptualisation*
- *risque lié à la conception et au développement*
- *risque lié à l'intégration*

Cet exemple montre que les risques sont regroupés en trois catégories dont chacune possède un ensemble de caractéristiques. Chaque catégorie regroupe les risques qui ont des caractéristiques homogènes. Chaque risque appartient à une seule catégorie : il est soit antérieur au projet, soit lié à la gestion du projet, soit lié aux phases de projet.

Par ailleurs, pour développer la taxonomie de la BI dans le cloud, la méthodologie de Nickerson, Varshney et Muntermann montre des limites. En effet la BI dans le cloud est une technologie complexe émergente : Une nouvelle technologie qui est à l'intersection de deux domaines (BI et cloud). Le développement des taxonomies pour ce genre de technologies a ses spécificités. Raison pour laquelle nous avons adapté la méthodologie de développement de taxonomies de Nickerson, Varshney et Muntermann pour la rendre applicable au développement de taxonomies des technologies complexes émergentes. Ainsi, nous avons instancié cette nouvelle méthodologie en l'appliquant au développement d'une taxonomie de la migration de la BI vers le cloud.

En outre, la BI permet d'améliorer et d'optimiser des décisions. Notre ambition est de mettre à la disposition des utilisateurs un outil d'aide à la décision. Ainsi, nous avons proposé un modèle de guidage opérationnel que nous avons instancié par une taxonomie de la BI dans le cloud et des règles pour la migration de la BI vers le cloud.

Afin de valider notre approche, un prototype a été implémenté ainsi qu'un scénario illustratif.

Dans la section suivante, nous présentons schématiquement tous les artefacts précités que nous avons développés tout au long de cette thèse.

1.5. Modèle des contributions et plan de la thèse

A l'aide du formalisme de diagramme de classes UML², nous avons construit un modèle qui regroupe les principaux artefacts que nous avons développés tout au long de notre travail. Ce modèle inclut aussi les deux artefacts proposés par Nickerson, Varshney et Muntermann (2013) car nous nous sommes inspirée de ces deux artefacts et nous les avons enrichis.

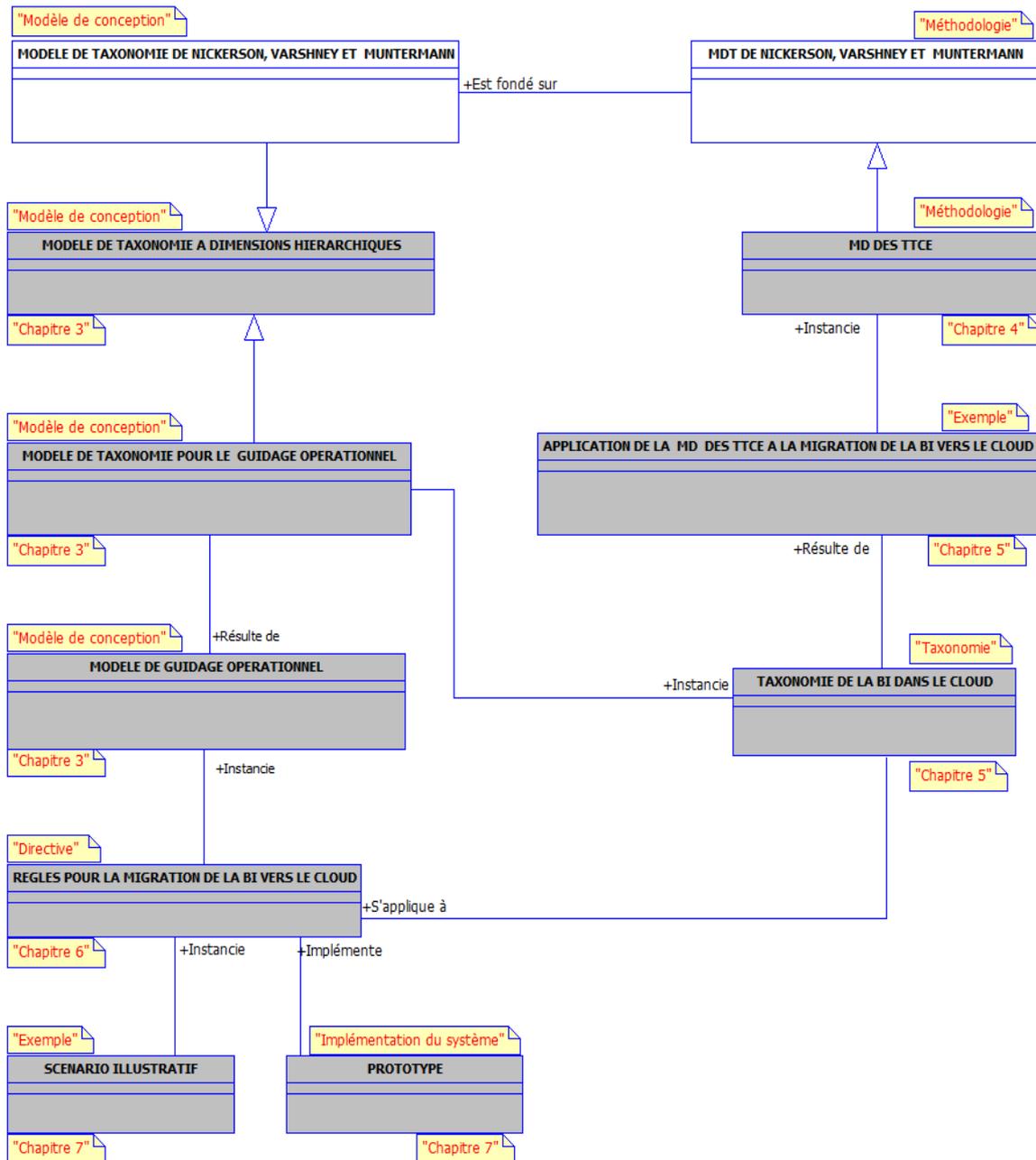


Figure 1: Modèle de la solution

² Unified Modeling Language : <http://www.omg.org/spec/UML/>

Légende

- TTCE : Taxonomie de technologies complexes émergentes
- MDT : Méthodologie de développement de taxonomies
- MD : Méthodologie de développement

Dans la figure 1 :

- Les deux artéfacts représentés par les rectangles blancs sont issus de l'article de Nickerson, Varshney et Muntermann.
- Les artéfacts représentés par les rectangles grisés sont ceux que nous avons conçus et développés dans le cadre de notre thèse.
- Le modèle de taxonomie de Nickerson, Varshney et Muntermann (*modèle de conception*) est fondé sur la *méthodologie* des mêmes auteurs.
- La taxonomie avec dimensions hiérarchiques (*modèle de conception*) généralise le modèle de taxonomie de Nickerson, Varshney et Muntermann (*modèle de conception*).
- La *méthodologie* de développement de taxonomies pour les technologies complexes émergentes spécialise la *méthodologie* de Nickerson, Varshney et Muntermann.
- La *méthodologie* de développement des taxonomies pour les technologies complexes émergentes est instanciée pour être appliquée à la migration de la BI vers le cloud.
- L'application de la *méthodologie* de développement des taxonomies pour les technologies complexes émergentes à la migration de la BI vers le cloud résulte en une *taxonomie* de la BI dans le cloud.
- Le modèle de taxonomie pour le guidage opérationnel (*modèle de conception*) spécialise le modèle de taxonomie à dimensions hiérarchiques.
- Le modèle de taxonomie pour le guidage opérationnel (*modèle de conception*) est instancié en une taxonomie de la BI sur le cloud.
- Le modèle de guidage opérationnel (*modèle de conception*) résulte du modèle de taxonomie pour le guidage opérationnel.
- Le modèle de guidage opérationnel (*modèle de conception*) instancie des règles pour la migration de la BI vers le cloud (*directives*).
- Les règles pour la migration de la BI vers le cloud (*directives*) sont instanciées dans un scénario illustratif (*exemple*).

- Les règles pour la migration de la BI vers le cloud sont implémentées dans un prototype (*Implémentation du système*).
- Les règles pour la migration de la BI vers le cloud s'appliquent à la *taxonomie* de la BI dans le cloud.

La note au-dessus de chaque artefact spécifie son type. Au point 2.2, nous définirons précisément tous les types d'artéfacts. En dessous de chaque artefact, nous avons mentionné le chapitre où est décrit cet artefact.

Ainsi, ce travail est structuré en 8 chapitres :

Le premier chapitre introduit ce travail et présente l'aperçu général de la solution.

Le deuxième chapitre présente un état de l'art de la BI dans le cloud. La mise en place d'un protocole de revue nous permet de constituer une bibliographie et d'effectuer une revue systématique de la littérature correspondante.

Le chapitre trois traite des taxonomies avec dimensions hiérarchiques et des opérateurs associés. Il propose un modèle de taxonomie avec des dimensions hiérarchiques et met en place un modèle de taxonomie pour le guidage opérationnel ainsi qu'un modèle de guidage opérationnel.

Au quatrième chapitre, nous proposons une méthodologie de développement de taxonomies pour les technologies complexes émergentes.

Dans le cinquième chapitre, nous développons une taxonomie pour la migration de la BI vers le cloud. Nous appliquons la méthodologie proposée au quatrième chapitre.

Au sixième chapitre nous proposons une méthode d'aide à la décision de migration de la BI vers le cloud. Nous présentons les règles pour la migration de la BI vers le cloud.

Au septième chapitre, nous décrivons le prototype que nous avons implémenté à partir des lignes directrices pour la migration de la BI vers le cloud. Nous mettons en place un scénario en vue de la validation de notre approche.

Avant de pouvoir entamer le processus de réponse à notre question de recherche, nous allons d'abord examiner la littérature sur le domaine de la BI dans le cloud. Cet examen permettra de justifier la pertinence et l'importance de notre question de recherche. En effet, les artefacts à développer doivent répondre à un problème pertinent et important qui n'a pas encore de solution (Hevner *et al.* 2004).

CHAPITRE 2 : BUSINESS INTELLIGENCE DANS LE CLOUD : ETAT DE L'ART

2.1. Introduction

Dans ce chapitre, nous présentons une revue systématique de la littérature sur le domaine de la BI dans le cloud. Nous appliquons la procédure de revue systématique de la littérature proposée par Kitchenham (2004). En effet, Kitchenham identifie, évalue et synthétise des revues systématiques de la littérature dans le domaine du génie logiciel. Cependant, la procédure qu'elle propose est bien adaptée pour être appliquée au domaine des systèmes d'information en général.

Nous allons limiter la revue aux articles qui font de la recherche en science de conception car nous avons adopté cette approche d'une part. D'autre part, cette revue vise à souligner l'apport des chercheurs en science de conception sur le domaine de la BI dans le cloud afin d'en élucider les opportunités de recherche.

Pour bien identifier et classer les artefacts contenus dans la littérature, nous proposons tout d'abord une typologie détaillée des artefacts.

2.2. Typologie des artefacts en science de conception

Une recherche en science de conception résulte en un ou plusieurs artefacts créés pour répondre à un problème pertinent qui n'a pas encore de solution. Ce paradigme vise à étendre les limites des capacités humaines et organisationnelles en créant des artefacts nouveaux et innovants (Hevner *et al.* 2004).

On trouve différentes typologies d'artefacts dans la littérature. March et Smith (1995) distinguent quatre types d'artefacts produits par la recherche en science de conception :

- Construit : une conceptualisation utilisée pour décrire des problèmes relatifs à un domaine et spécifier les solutions à ces problèmes.

- Modèle : une représentation abstraite d'une situation du monde réel à l'aide de construits.
- Méthode : un ensemble d'étapes (algorithme, directives) suivies pour effectuer une tâche.
- Instanciation : la mise en œuvre d'un construit, d'un modèle ou d'une méthode dans un environnement.

Cette typologie est largement utilisée, y compris dans l'article de Hevner et al.(2004). Cependant, elle est parfois difficile à appliquer, en raison de l'imprécision relative des concepts de construit, de modèle, de méthode et d'instanciation (Sangupamba Mwilu, Prat and Comyn-Wattiau 2014). Par conséquent, il est utile de spécialiser cette typologie en spécifiant et en définissant des sous-catégories pour les quatre catégories d'artefacts précitées.

Offermann et al. (2010) ont tenté de résoudre le problème d'imprécision des concepts dans la typologie des artefacts de March et Smith en proposant des catégories plus spécifiques. Leur travail fournit une base utile pour une classification des différents types d'artefacts. Cependant, il y manque quelques sous-catégories importantes. De plus, certains types comprennent un grand nombre de concepts différents et les définitions proposées peuvent prêter à confusion. Par exemple, Offermann et al. (2010) définissent une méthode comme un ensemble d'activités qui sont effectuées, éventuellement dans un certain ordre, par des personnes afin de soutenir le développement d'un système. Il y a là une confusion car une méthode englobe plusieurs sous-types : une méthodologie, une directive, un algorithme, un fragment de méthode et une métrique. C'est plutôt une méthodologie qui peut être définie ainsi.

Pour remédier à cette situation, nous proposons notre typologie des différents types d'artefacts (sous-catégories des concepts de construit, modèle, méthode et instanciation). Pour chaque sous-catégorie, nous proposons ou reprenons une définition. Les références dans le tableau 1 indiquent les articles dont les définitions ont été tirées et éventuellement adaptées. Notre typologie, avec les sous-catégories précises et une définition pour chaque sous-catégorie, aide à l'identification et à la caractérisation des artefacts en science de conception.

CONSTRUIT	
Langage	<i>Un ensemble de concepts, ou plus généralement de symboles, de règles pour les combiner (syntaxe) et de règles d'interprétation des combinaisons de symboles (sémantique) (Edwards et al. 2001).</i>
Méta-modèle	<i>Un ensemble de concepts représentés graphiquement, avec des règles pour combiner ces concepts.</i>
Concept	<i>Un nouvel élément ajouté à un langage existant ou à un méta-modèle.</i>
MODELE	
Modèle de conception	<i>Une structure ou une description liée au comportement d'un système, utilisant généralement une notation graphique et parfois du texte (Offermann et al. 2010).</i>
Ontologie	<i>Une spécification formelle explicite d'une conceptualisation partagée (Gruber 1993).</i>
Taxonomie	<i>Une classification des objets d'un domaine, sur la base de caractéristiques communes (Nickerson, Varshney and Muntermann 2013).</i>
Cadre de référence	<i>Une structure logique pour l'organisation d'informations complexes (The Chief Information Officers Council 1999).</i>
Architecture	<i>Une structure générale représentant un système, l'organisation de ses différents composants et les relations entre ses composants (ISO/IEC/IEEE 2010) (Jarke et al. 2011).</i>
Exigence	<i>Une condition ou une capacité qui doit être remplie ou possédée par un système (ISO/IEC/IEEE 2010).</i>
METHODE	
Méthodologie	<i>Un ensemble prédéfini d'étapes et de lignes directrices, avec des techniques et des outils associés. Il est destiné à, ou utilisé par, des personnes qui travaillent dans une discipline (ISO/IEC/IEEE 2010) (Nunamaker et al. 2011).</i>
Directive	<i>Une suggestion concernant le comportement à adopter dans une situation particulière (Offermann et al. 2010). Exemples : principes de conception (grandes lignes directrices), heuristiques, règles (directives détaillées) (Hanseth and Lyytinen 2010).</i>
Algorithme	<i>Une séquence exécutable d'opérations pour effectuer une tâche spécifique (Offermann et al. 2010) (ISO/IEC/IEEE 2010).</i>
Fragment de méthode	<i>Un composant de procédé qui peut être traité comme une unité séparée et réutilisé dans des contextes différents (Kornyshova, Deneckère and Salinesi 2007). Exemple : les patrons de conception.</i>
Métrique	<i>Une fonction qui attribue un nombre ou un symbole à une entité afin de caractériser un attribut ou un groupe d'attributs. La valeur de la métrique est appelée une mesure (Purao and Vaishnavi 2003).</i>
INSTANCIATION	
Implémentation d'un système	<i>Un système logiciel ou matériel mis en œuvre. Exemple : un prototype ou un outil finalisé.</i>
Exemple	<i>Toute autre matérialisation concrète d'un artefact abstrait (construit, modèle ou méthode). Exemples : l'application d'un langage de requête à un scénario illustratif, l'illustration d'un cadre de théorie de la conception avec des exemples concrets de théories de la conception, l'application d'une méthodologie de projet à un projet réel.</i>

Tableau 1: Typologie des artefacts en science de conception (Sangupamba Mwilu, Prat and Comyn-Wattiau 2014)

Cette typologie a fait l'objet d'une présentation lors de l'atelier international sur la modélisation et la gestion des big data (MoBID 2014) (Sangupamba Mwilu, Prat and Comyn-Wattiau 2014) et d'une publication dans le journal *Future Generation Computer Systems* (Sangupamba Mwilu, Comyn-Wattiau and Prat 2016). Dans la suite, nous allons illustrer son utilisation pour classer les artefacts que nous avons repérés dans la littérature sur la BI dans le cloud d'une part. D'autre part, la même typologie nous a permis d'identifier les lacunes et les opportunités de recherche dans ce domaine de la BI dans le cloud.

2.3. Revue systématique de la littérature selon Kitchenham

Une revue systématique de la littérature est un moyen d'identifier, d'évaluer et d'interpréter toutes les recherches pertinentes relatives à une question particulière de recherche ou à un phénomène d'intérêt. Elle est une forme d'étude secondaire, les documents individuels qui contribuent à son élaboration étant appelés études primaires (Kitchenham and Brereton 2013).

Dans le domaine du génie logiciel, elle a été largement popularisée par Kitchenham (Kitchenham 2004). Le fait qu'elle soit applicable à tous les domaines lui donne un caractère universel. Nous en rappelons les principes ici.

2.3.1. Objectifs de la revue systématique de la littérature

Selon Kitchenham (Kitchenham 2004), beaucoup de raisons militent pour la revue systématique, les plus communes sont :

- Résumer les documents existants, les avantages et les limites d'une technologie spécifique ;
- Identifier les lacunes dans la recherche actuelle sur un sujet afin de suggérer des voies de solutions pour les recherches futures ;
- Fournir un cadre afin de classer de manière appropriée les activités de recherche.

2.3.2. Importance de la revue systématique

La plupart des recherches commencent par une revue de la littérature. Toutefois, à moins que la revue de la littérature ne soit approfondie et rigoureuse, elle présente peu de

valeur scientifique. Une revue systématique synthétise les travaux existants d'une manière méthodologique et rigoureuse. Par exemple, les revues systématiques doivent être effectuées conformément à une stratégie de recherche prédéfinie dans un protocole (Kitchenham 2004).

2.3.3. Propriétés des revues systématiques

Certaines caractéristiques différencient une revue systématique d'une revue de la littérature classique, parmi lesquelles nous citons (Kitchenham 2004) :

- Une revue systématique commence par définir un *protocole* qui spécifie la question de recherche abordée et les méthodes qui seront utilisées pour effectuer l'examen.
- Une revue systématique est basée sur une *stratégie de recherche* définie qui vise à détecter la plus grande partie possible de la littérature pertinente.
- Une revue systématique *documente sa stratégie* de recherche afin que les lecteurs puissent évaluer sa rigueur et son exhaustivité.
- Une revue systématique exige des *critères explicites d'inclusion et d'exclusion* pour évaluer chaque étude primaire potentielle.
- Une revue systématique précise *les informations à obtenir de chaque étude primaire*, y compris les critères de qualité permettant d'évaluer chaque étude primaire.

2.3.4. Planification de la revue systématique

La planification de la revue comprend l'identification de la nécessité d'une revue systématique et la constitution du protocole de la revue (Kitchenham 2004) (Kitchenham and Brereton 2013) (Okoli and Schabram 2010).

2.3.4.1. Identification de la nécessité de la revue systématique

La nécessité d'un examen systématique découle de l'exigence des chercheurs de résumer toutes les informations existantes sur un phénomène d'une manière approfondie et impartiale. Cela peut être dans le but de tirer, sur un phénomène donné, des conclusions plus générales que celles découlant des études individuelles. Ce résumé peut aussi servir de prélude à d'autres activités de recherche.

2.3.4.2. Protocole de revue

Le protocole de la revue comprend :

- Le contexte : la justification de l'enquête.
- La question de recherche à laquelle l'examen est destiné à répondre.
- La stratégie qui sera utilisée pour rechercher des études primaires, y compris les termes de recherche et les ressources à rechercher. Les ressources comprennent des bases de données, des revues spécifiques et des actes des conférences.
- Les critères et procédures de sélection de l'étude : ce sont les règles requises pour inclure une étude ou l'exclure de la revue systématique. Le protocole doit décrire la façon dont les critères seront appliqués.
- Les procédures d'évaluation de la qualité de l'étude. Les chercheurs doivent déterminer des critères pour évaluer la qualité des études primaires.
- La stratégie d'extraction de données. Cela doit définir comment les informations requises à partir de chaque étude primaire seront obtenues. Si les données nécessitent une manipulation ou des hypothèses et des déductions à faire, le protocole doit spécifier un processus de validation approprié.
- La synthèse des données extraites. Il s'agit de définir une stratégie de synthèse et de préciser comment affiner l'étude si trop d'études primaires sont identifiées.
- Le calendrier du projet : c'est la définition du planning de l'examen.

2.3.5. Conduite de la revue systématique

Selon Kitchenham (Kitchenham 2004), la conduite d'une revue systématique doit comprendre les cinq étapes suivantes :

- L'identification de la recherche,
- La sélection des études,
- L'évaluation de la qualité des études,
- L'extraction des données,
- La synthèse des données.

2.3.5.1. Identification de la recherche

L'objectif d'un examen systématique est de trouver autant d'études primaires relatives à la question de recherche que possible en utilisant une stratégie de recherche impartiale. La rigueur du processus de recherche est un facteur qui distingue les examens systématiques des examens traditionnels. Cette étape détermine la stratégie de la recherche, la documente et en gère la bibliographie.

2.3.5.2. Sélection des études

Les études primaires potentiellement pertinentes obtenues doivent être évaluées en vue de leur pertinence réelle. C'est ici que les critères et la procédure de sélection des études sont déterminés.

2.3.5.3. Evaluation de la qualité des études

Pour être incluses dans un échantillon, les études primaires sélectionnées seront soumises à un processus d'évaluation de leur qualité.

2.3.5.4. Extraction des données

L'objectif de cette étape est de concevoir des formulaires d'extraction de données pour enregistrer avec précision les informations fournies par les études primaires. Afin de réduire la probabilité de partialité, des formulaires d'extraction de données doivent être définis et testés lorsque le protocole d'étude est défini. La procédure d'extraction des données doit être déterminée au préalable.

2.3.5.5. Synthèse des données

Cette étape consiste à rassembler et résumer les résultats des études primaires incluses. Une synthèse peut être descriptive ou quantitative.

2.4. Application du processus de revue systématique selon Kitchenham à notre cas

Nous allons utiliser la procédure proposée par Kitchenham pour effectuer une revue systématique de la littérature sur la BI dans le cloud.

2.4.1. Objectifs de la revue systématique de la littérature de la BI dans le cloud

Notre revue systématique de la littérature sur la BI dans le cloud poursuit le deuxième objectif défini par Kitchenham : Elle se veut le résumé des études existantes sur la BI dans le cloud en vue de déceler les lacunes pour suggérer des voies de solutions.

En nous appuyant sur notre typologie des artefacts en science de conception, nous allons exhiber les artefacts déjà mis en place dans le domaine de la BI dans le cloud d'une part. D'autre part, nous allons identifier des opportunités de recherche en science de conception pour combler le vide en développant de nouveaux artefacts.

2.4.2. Importance de la revue systématique pour notre travail

Cette revue systématique de la littérature nous permet de structurer l'examen des études primaires sur la BI dans le cloud selon les étapes proposées dans la méthodologie de Kitchenham (Kitchenham 2004).

Le protocole de la revue, établi au préalable, facilite sa conduite car toute la stratégie de la revue y est définie. Grâce à la rigueur qu'impose la revue systématique, nous pouvons recueillir un échantillon des études primaires pertinentes pour assurer plus d'exhaustivité à notre échantillon.

2.4.3. Propriétés de la revue systématique de la littérature sur la BI dans le cloud

Nous avons défini notre revue systématique de la façon suivante :

- Stratégie de recherche de la littérature pertinente : La recherche de la documentation sur la BI dans le cloud sera faite dans les principales sources d'articles universitaires. Les résultats seront triés selon leur pertinence.
- Documentation de la stratégie de recherche : Toute la démarche que nous poursuivrons tout au long de la revue sera décrite de manière chronologique.
- Critères explicites d'inclusion et d'exclusion pour évaluer chaque étude primaire potentielle : Pour qu'un article fasse partie de notre échantillon, il doit traiter de notre sujet de recherche et doit relever de la science de conception.

- Informations à obtenir dans chaque étude primaire, y compris les critères permettant d'évaluer la qualité de chaque étude primaire : Pour chaque étude, nous allons retenir les artefacts proposés. Dans un article on peut trouver un ou plusieurs artefacts. Pour évaluer la qualité des articles de notre échantillon, nous avons déterminé quelques critères expliqués plus loin.

2.4.4. Planning de la revue systématique

En appliquant la méthodologie de Kitchenham, nous allons ici identifier la nécessité de faire une revue systématique de la BI dans le cloud et nous mettrons en place un protocole en vue de cette revue.

2.4.4.1. Nécessité d'un examen systématique

Le but de cette revue systématique n'est pas seulement de synthétiser l'état actuel de la littérature sur la BI dans le cloud, mais aussi de souligner de manière particulière l'apport des chercheurs en science de conception pour la BI dans le cloud. L'analyse de la synthèse permettra d'élucider les opportunités de recherche en science de conception. Nous nous servirons de ces opportunités pour apporter notre contribution dans ce domaine de la BI dans le cloud.

2.4.4.2. Protocole de revue de la littérature de la BI dans le cloud

- **Le contexte** : Cette revue se place dans le cadre de la recherche en systèmes d'information en utilisant le paradigme de la science de conception. Elle va synthétiser la littérature sur la BI dans le cloud en vue d'élucider les opportunités de recherche en science de conception dans ce domaine de la BI dans le cloud.
- **La question de recherche** : Quels nouveaux artefacts pouvons-nous développer pour permettre à la BI de tirer profit des avantages du cloud ? Nous allons répondre à cette question tout au long de notre thèse de manière concrète en développant les nouveaux artefacts cités au chapitre deuxième (aperçu général de la solution).
- **La stratégie qui sera utilisée** : Pour constituer notre échantillon d'articles, nous ferons des recherches documentaires sur Google Scholar. En dehors de ce moteur de recherche, les principales sources de documents universitaires relatifs aux technologies complexes émergentes sont ACM, IEEE, DBLP, ScienceDirect, EBSCO, ainsi que la librairie électronique de l'AIS (AISeL). Nous allons sonder toutes ces

sources pour constituer notre bibliographie. Nous lancerons des requêtes constituées des mots clés de notre sujet de recherche. La recherche se fera dans toutes les parties des articles (full text).

Sur la base de notre expérience, nous suggérons de limiter la recherche aux 50 premières pages des résultats pour Google Scholar. Par défaut, ce moteur affiche dix résultats sur une page. Les 50 pages correspondent donc à 500 résultats. Pour les autres sources, nous avons tenu compte du nombre de résultats affichés par défaut sur une page afin de retenir un nombre de pages correspondant aux 500 résultats comme pour Google Scholar. Ces articles seront intégrés (déduplication) : plusieurs articles sont cités dans de multiples sources (en particulier, la plupart des articles sont cités dans Google Scholar).

Les technologies émergentes peuvent rapidement devenir obsolètes. Pour cette raison, nous proposons de limiter la recherche aux onze dernières années : de 2006 à 2016. Les résultats seront affichés selon leur pertinence.

La BI dans le cloud est une technologie émergente et le corpus de connaissances dans ce domaine est aussi émergent. Il y a besoin de recourir à la littérature sur les domaines constitutifs de cette technologie émergente pour compléter les connaissances. Raison pour laquelle nous allons retenir les articles qui traitent de la BI dans le cloud ou qui abordent l'un de ces deux domaines seulement qui sont écrits en anglais car au niveau international, la recherche se fait en anglais. Nous aurons donc deux types différents de sources : Le type 1 sera constitué des articles qui traitent de la BI dans le cloud tandis que le type 2 sera constitué des articles qui traitent de la BI ou bien du cloud computing.

Notre typologie des artefacts présentée à la section 2.2 nous aidera à déterminer le type d'artefacts contenu dans chaque article. Un article peut proposer un ou plusieurs artefacts.

- **Les critères et la procédure de sélection des études :** Pour qu'un article soit retenu, il doit traiter de la BI dans le cloud ou d'un de ces deux domaines constitutifs seulement. Nous procéderons d'abord à la lecture des titres des articles obtenus après l'exécution des requêtes. Tous les articles qui ne se rapportent ni à la BI ni au cloud seront exclus de l'échantillon. Ensuite, les résumés seront examinés pour s'assurer que ces articles relèvent des sciences de conception. L'heuristique de base pour nous aider à décider s'il s'agit de science de conception est d'identifier au moins un artefact que cette étude produit. Au cas où le titre et le résumé ne permettront pas de déterminer si

l'article traite de la BI dans le cloud et qu'il relève des sciences de conception, nous procéderons à la lecture de l'article en intégralité avant de décider de l'inclure ou non dans l'échantillon.

- **La procédure d'évaluation de la qualité des études :** Nous allons éliminer tous les articles qui ne sont pas publiés dans des revues ou des actes de conférences figurant sur les listes de l'ERA (Excellence in Research for Australia) (ERA 2012). En effet, ERA est une entité qui évalue la qualité de la recherche menée dans les universités australiennes par rapport à des indicateurs nationaux et internationaux. Elle publie les listes des revues ainsi que des actes des conférences qui ont été jugés de qualité. Il n'y a pas, à notre connaissance, d'autre liste de référence de ce type dans notre domaine.
- **La stratégie d'extraction des données :** Pour chaque article lu, nous allons retenir l'artefact ou les artefacts qui y sont proposés. Grâce à la typologie des artefacts, nous déterminerons le type de chaque artefact.
- **Synthèse des données extraites :** La synthèse des artefacts trouvés dans la littérature sera représentée dans un cadre référentiel. Ce tableau est présenté dans les annexes de cette thèse (ANNEXE 1).

2.4.5. Conduite de la revue systématique

Nous décrivons ici les cinq étapes de notre revue systématique dérivées de la méthodologie proposée par Kitchenham.

2.4.5.1. Identification de la recherche

Nous avons constitué la bibliographie grâce aux mots clés en faisant la recherche sur les moteurs déterminés dans le protocole de la revue.

Comme le protocole le précise, nous avons deux types différents de sources :

- Les sources de type 1 : les articles qui traitent de la BI dans le cloud,
- Les sources de type 2 : les articles qui traitent de BI ou de cloud computing pris séparément.

Ainsi, les mots clés utilisés sont différents selon les types de sources.

Les sources de type 1

Les termes utilisés pour collecter les articles relatifs à la BI dans le cloud sont les suivants :

("Business Intelligence" OR "Analytics" OR "Data warehouse") AND ("Cloud" OR "SaaS"). La requête combine les mots clés de la BI et ceux du cloud. Le premier ensemble de termes se réfère à la BI et aux termes connexes à la BI alors que le second se réfère au cloud et sa couche la plus typique (SaaS).

Google Scholar ne permet pas des requêtes très complexes : la partie gauche de l'opérateur "AND" ne peut avoir qu'un seul terme. Pour résoudre cette difficulté, nous avons transformé la requête en les deux requêtes suivantes :

- "Cloud" AND ("Business Intelligence" OR "Analytics" OR "Data warehouse")
- "SaaS" AND ("Business Intelligence" OR "Analytics" OR "Data warehouse").

Notons que la recherche était effectuée dans le texte intégral de chaque article. Par ailleurs, nous avons aussi retenu les articles sur les états de l'art.

Les sources de type 2

Pour les sources de type 2, nous nous sommes intéressés uniquement aux articles de référence des domaines constitutifs de la BI dans le cloud, à savoir la BI et le cloud computing. En effet, il s'agit de compléter nos sources contenant des recherches associant la BI et le cloud. Nous n'avons pas jugé utile de considérer tous les articles sur la BI ou sur le cloud computing. On peut prendre en compte les articles qui font des synthèses sur le thème de la BI ou celui du cloud computing. Nous n'avons retenu que les états de l'art pour nous assurer que seuls les documents les plus fondamentaux pour la compréhension des domaines sont sélectionnés.

Les mots-clés sont spécifiques à chacun des domaines concernés. Par conséquent, nous avons lancé deux requêtes sur les mêmes moteurs de recherche comme dans l'étape précédente. Les mots clés utilisés sont les suivants :

- (“Overview” OR “State of the art” OR “Survey” OR “Literature review”) AND (“Business Intelligence” OR “Analytics” OR “Data warehouse”)
- (“Overview” OR “State of the art” OR “Survey” OR “Literature review”) AND (“Cloud” OR “SaaS”).

Tenant compte du fait que Google Scholar n’accepte pas ces requêtes complexes, les deux requêtes conçues pour la recherche des sources de type 2 ont été scindées en cinq :

- “Business Intelligence” AND (“Overview” OR “State of the art” OR “Survey” OR “Literature review”)
- Analytics AND (“Overview” OR “State of the art” OR “Survey” OR “Literature review”)
- "Data warehouse" AND (“Overview” OR “State of the art” OR “Survey” OR “Literature review”)
- Cloud AND (“Overview” OR “State of the art” OR “Survey” OR “Literature review”)
- SaaS AND (“Overview” OR “State of the art” OR “Survey” OR “Literature review”)

Comme pour les sources de type 1, la recherche était effectuée dans le texte intégral de chaque article.

2.4.5.2. Sélection des études

Les requêtes mises en place à la phase précédente nous ont permis de recueillir les articles pour notre échantillon.

Les sources de type 1

Après exécution de chacune des deux requêtes dans Google Scholar, nous avons obtenu respectivement 17 700 et 10 400 résultats. Pour chaque requête, nous avons retenu les cinquante premiers écrans qui contenaient 932 articles.

La deuxième étape de la recherche des articles de type 1 a consisté à lancer la requête telle qu'elle a été conçue initialement ("Business Intelligence "OR "Analytics" OR "Data warehouse") AND ("Cloud" OR "SaaS") à l’aide des six autres moteurs de recherche.

- 1) IEEE Xplore : La recherche dans cette source a retourné 20 991 résultats, nous en avons retenu 500 (50 premiers écrans).

- 2) EBSCOhost : La recherche effectuée dans cette source a retourné 1 700 résultats, nous en avons retenu 500.
- 3) DBLP : Nous avons procédé de la même manière que dans les sources précédentes et nous avons trouvé 129 résultats.
- 4) ScienceDirect : Ici, la requête a retourné 471 résultats.
- 5) ACM Digital Library : Pour cette source, la requête nous a renvoyé 422 résultats.
- 6) AIS Electronic Library (AISeL) : Après l'exécution de la requête, cette source a retourné 279 articles.

A l'issue de la deuxième étape de recherche, notre échantillon comportait 3 233 articles dont nous avons enregistré les références dans un fichier Excel que nous avons classé ensuite par ordre alphabétique des titres. Ce classement nous a permis de supprimer plus facilement tous les doublons. Après ce dédoublonnage, notre fichier comptait finalement 1 030 articles. Ensuite, nous avons parcouru les titres de chacun de ces articles. Cela nous a permis de supprimer tous les articles ne se rapportant pas à la BI dans le cloud. Après cette opération, notre échantillon comportait 890 articles. Enfin, nous avons lu les résumés de ces 890 articles. Cette opération nous a permis de supprimer tous les articles qui ne faisaient pas de science de conception. Nous rappelons que l'heuristique de base pour nous aider à décider s'il s'agit de science de conception est de déceler au moins un artefact que cette étude produit. Après l'application de ce critère, notre échantillon comptait 130 articles. Notons que chaque fois qu'il était difficile de déterminer si l'article traite de la BI dans le cloud par son titre, nous avons procédé à la lecture de son résumé. De plus, chaque fois que la lecture de résumé d'un article ne permettait pas de déterminer si l'article relève des sciences de conception, nous avons parcouru tout l'article avant de décider de l'inclure ou non dans l'échantillon.

Les sources de type 2

Comme pour le premier type, nous avons commencé la recherche par Google Scholar et avons retenu 228 articles.

Dans la deuxième étape consistant à lancer les deux requêtes conçues pour rechercher les sources de type 2 à l'aide des six autres moteurs de recherche, nous avons obtenu les résultats suivants :

- 1) IEEE Xplore : Les requêtes lancées dans cette source ont retourné 187 résultats.
- 2) EBSCOhost : Les requêtes ont retourné 128 résultats.

- 3) DBLP : Nous avons procédé de la même manière que dans les sources précédentes et avons trouvé 109 résultats.
- 4) ScienceDirect : Ici, la requête a retourné 239 résultats.
- 5) ACM Digital Library : Pour cette source, les requêtes ont retourné 91 résultats.
- 6) AIS Electronic Library (AISeL) : Après l'exécution de la requête, cette source a retourné 282 articles.

Toutes les listes d'articles retenus ont été fusionnées et triées pour éliminer les doublons. A l'issue de cette opération, nous avons retenu 108 articles. Après la lecture des titres et des résumés des articles retenus, nous avons supprimé les articles qui ne traitaient pas de la BI ni de cloud computing et ceux qui ne relevaient pas des sciences de conception. Notre échantillon est resté avec 62 articles dont 33 sur la BI et 29 sur le cloud computing.

2.4.5.3. *Evaluation de la qualité des études*

Sources de type 1

Notre échantillon comptait quarante-trois articles à l'issue de l'application du seul critère de qualité : Etre publié dans une revue ou des actes de conférences figurant sur la liste d'ERA.

Sources de type 2

A l'aide du même critère de qualité utilisé ci-haut, nous avons pu élaborer un échantillon du deuxième type de sources comptant dix-neuf articles dont neuf sur la BI et dix sur le cloud computing.

2.4.5.4. *Extraction des données*

Il s'agit ici d'extraire les artefacts contenus dans les soixante-deux articles de notre échantillon : quarante-trois articles de type 1 et dix-neuf articles de type 2. Nous avons procédé à la lecture de chaque article pour ressortir le ou les artefacts proposés. La typologie des artefacts présentée au point 2.2 nous a permis de classer chaque artefact selon son sous-type dans un cadre de référence à deux dimensions. La synthèse de données extraites est présentée au point 2.5.

2.5. Résultat de l'état de l'art

Nous présentons ici l'état de l'art de la BI dans le cloud tel qu'il résulte de notre revue systématique de la littérature. La BI dans le cloud est une technologie émergente qui se situe à l'intersection de deux domaines distincts de l'informatique : la BI et le cloud computing. Il sied donc de présenter d'abord l'état actuel de la littérature sur ces deux domaines constitutifs : la BI et le cloud computing avant de traiter de la BI dans le cloud.

2.5.1. Business intelligence

La BI est une collection de technologies d'aide à la décision visant à soutenir les décideurs pour une rapide prise de décisions pertinentes. Les avantages qu'elle offre poussent les organisations à adopter ses outils pour assurer la prise de décisions adéquates. Il est difficile aujourd'hui de trouver une entreprise prospère qui n'exploite pas la technologie BI car celle-ci fait preuve d'un support incontestable à la prise de décision (Chaudhuri, Dayal and Narasayya 2011).

Par exemple, la technologie BI est utilisée dans les entreprises pour la gestion des commandes et la gestion de la clientèle; dans les services financiers pour l'analyse des sinistres et la détection de fraude, dans le transport pour la gestion de flottes, dans les télécommunications pour identifier les raisons qui poussent les clients à se désabonner et dans le domaine des soins de santé pour l'exploitation des résultats des analyses médicales.

Les deux dernières décennies ont connu une croissance rapide, tant dans le nombre de produits et services offerts par la BI que dans l'adoption de ces technologies par les entreprises. Cette croissance a été motivée par la diminution du coût de l'acquisition et du stockage de très grandes quantités de données. En effet, actuellement, les différentes activités de la société génèrent, rapidement, de gros volumes des données numériques : on parle de big data (Chaudhuri, Dayal and Narasayya 2011).

Outre les données opérationnelles, les organisations font de plus en plus de la BI sur les big data. D'où l'appellation "big data analytique" (big data analytics) qui se réfère à l'application de techniques analytiques sur ces gros volumes de données. Par exemple le big data analytique offre la possibilité de coordonner les informations sur les préférences des

consommateurs grâce aux informations provenant des tweets, des blogs, des données des réseaux sociaux, des transactions de clients dans les secteurs de la banque et du commerce électronique afin de prédire leurs besoins (Chaudhuri 2012). Une des raisons de l'engouement autour des big data est que les utilisateurs pensent identifier des nouvelles connaissances dans ces masses de données hétérogènes afin de prendre des décisions pertinentes pour leur organisation. Par exemple, accéder à des informations qui peuvent être utilisées pour comprendre le comportement du client et prévoir les tendances du marché (Hedgebeth 2007).

Par ailleurs, comme nous l'avons dit ci-haut, le coût de l'acquisition et du stockage de données a considérablement diminué. Cela a augmenté la possibilité pour des entreprises d'acquérir de très gros volumes de données afin d'en extraire autant d'avantages compétitifs que possible en faisant du big data analytique (big data analytics).

Afin d'illustrer comment est structurée la BI et de mettre en évidence chacun de ses composants ainsi que l'interaction entre ces différents composants, nous allons présenter une architecture de la BI.

2.5.1.1. Une architecture de la BI

Une architecture est une structure générale représentant un système, l'organisation de ses différents composants et les relations entre ses composants (Jarke *et al.* 2011). Elle permet de décrire comment un système doit être conçu afin de répondre à sa spécification fonctionnelle.

Il existe plusieurs architectures - types de la BI. Nous présentons ici une architecture de la BI adaptée du cadre de big data analytique proposée par Phillips-Wren et al (2015) (Figure 2). Le choix de cette architecture est motivé par le fait qu'elle comprend tous les éléments relatifs à la BI et, particulièrement, elle tient compte des nouvelles architectures de données et outils analytiques parallèles.

Dans une architecture de la BI, l'élément central est l'entrepôt de données. Celui-ci est une collection de données **thématiques, intégrées, non volatiles et historisées** pour la prise de décision (Inmon 2002).

On dit qu'un entrepôt des données est une collection des données **thématiques** car les données sont organisées par thème tel que client, produit ou vente. C'est une collection de données **intégrées** car celles-ci proviennent de différentes applications de l'entreprise et sont standardisées pour présenter une vue unifiée des données aux utilisateurs. Ces données sont dites **non volatiles** et **historisées** : les nouvelles données s'ajoutent aux anciennes sans les remplacer. On peut suivre dans le temps l'évolution des différentes valeurs des données (Inmon 2002).

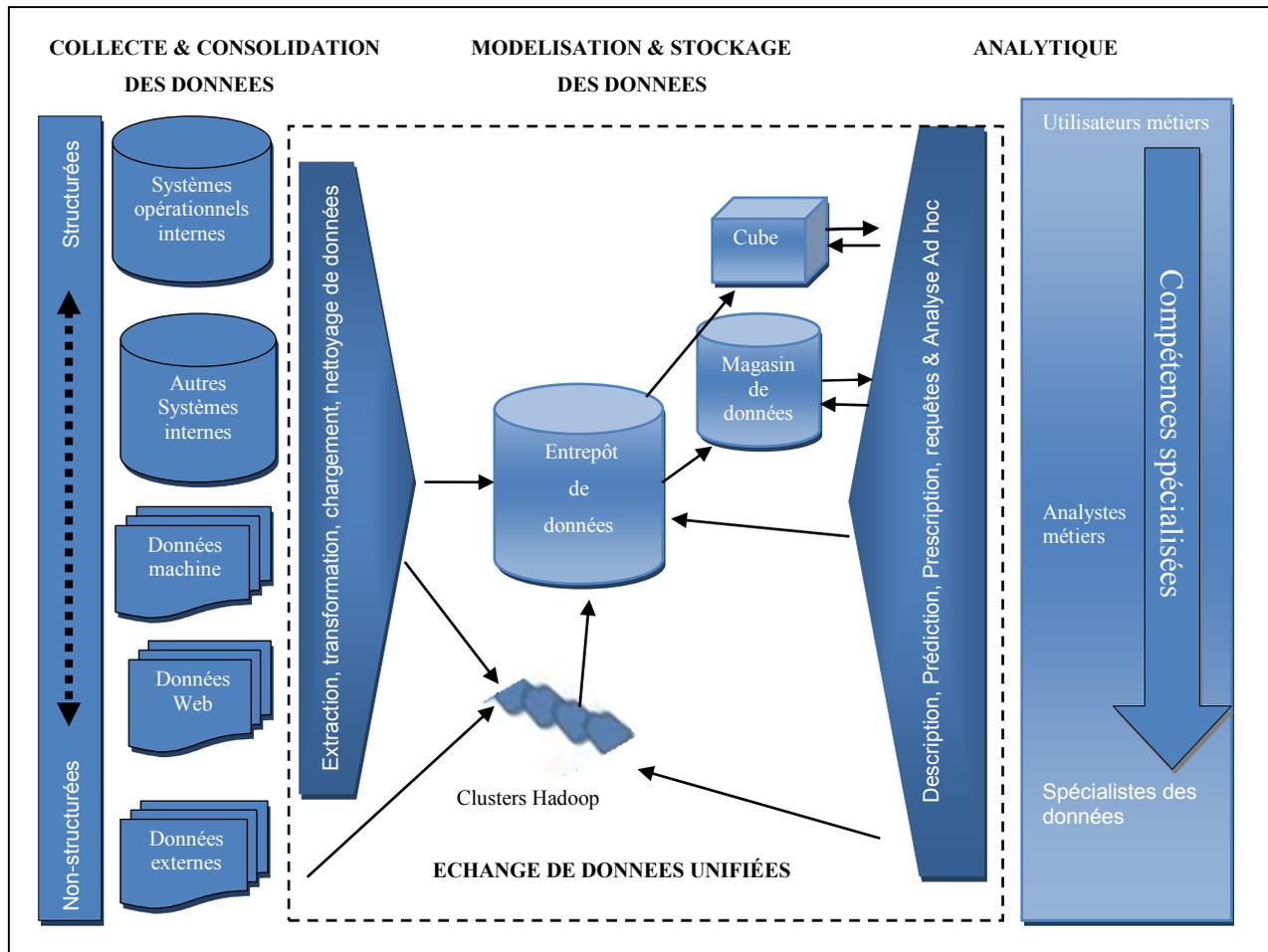


Figure 2 : Architecture de la BI (traduit et adapté de Phillips-Wren et al. (2015))

Comme nous l'avons dit ci-haut, l'architecture que nous présentons a été adaptée du cadre de big data analytique proposée par Phillips-Wren et al (2015). Ce cadre est constitué de composants suivants :

- Sources des données,
- Préparation des données,
- Stockage des données,
- Analyse,
- Accès aux données et leur utilisation,
- Gestion des big data.

Pour nous conformer à la définition d'une architecture et prendre en compte nos besoins de représentation de la connaissance, nous avons apporté les modifications suivantes au cadre du big data analytique tel que proposé par Phillips-Wren et al :

1. Suppression de la partie relative à la gestion des big data : une architecture se limite à présenter les composants d'un système. Il ne décrit pas son fonctionnement.
2. Ajout du composant "cube" : dans un data warehouse, les données sont représentées sous forme de data marts ou plus généralement de cubes.
3. Restructuration en trois grands composants : La BI requiert la capacité d'acquérir l'information, de la stocker et de l'analyser. Ainsi, nous structurons son architecture en trois composants :
 - Collection et consolidation des données,
 - Modélisation et stockage,
 - Analytique.

Ci-dessous, nous présentons chacun de ces trois composants.

1) La collecte et la consolidation des données

Ce composant assure toutes les activités à partir de la collecte des données dans leurs sources différentes jusqu'à la mise à jour de l'entrepôt de données.

Les données proviennent, en premier lieu, des sources internes de l'entreprise. Celles-ci sont généralement des bases de données opérationnelles telles que celles exploitées par les progiciels de gestion intégrée (PGI) ou les progiciels de gestion de la relation client. En second lieu, les données peuvent émaner de sources externes comme Internet. Les données

collectées de sources différentes sont hétérogènes aussi bien par leur format que par leur structure. On retrouve des données structurées (exemple : des bases des données relationnelles), des données semi-structurées (exemple : des documents XML) et des données non structurées (exemple : des textes, des images, du son...).

Les outils ETL (Extract, Transform, Load) permettent de remédier au problème d'intégration et de normalisation des données. Ils sont chargés de mettre les entrepôts de données à jour : ils extraient les données de leurs sources d'origine, les transforment et les chargent dans les entrepôts des données. La transformation consiste principalement à nettoyer les données, à les harmoniser et à les pré-agréger (Al-Aqrabi *et al.* 2015) (Al-Aqrabi *et al.* 2012) en vue de leur stockage.

2) Le stockage des données

Ce composant englobe les modèles, les méthodes, les techniques et les outils de gestion de grands volumes de différents types de données. L'élément important de cette étape est l'entrepôt des données. Celui-ci peut être composé de magasins de données.

Un magasin de données (data mart) est un sous-ensemble d'un entrepôt de données qui contient les données relatives à un domaine fonctionnel d'une organisation.

Dans les entrepôts, les données sont généralement représentées selon une vue multidimensionnelle. Les trois concepts fondamentaux utilisés sont : le fait, la dimension et la mesure (Al-Aqrabi *et al.* 2015). Un fait est représenté comme un hypercube, les dimensions sont les axes du cube et une mesure est un indicateur, une valeur d'intérêt pour le décideur. Une dimension est composée de différents niveaux hiérarchisés. Et chaque niveau possède au moins un attribut.

3) L'analytique

C'est la dernière étape de la BI. Elle comprend tous les types d'analyse ainsi que la transmission des données aux utilisateurs finaux. Les éléments essentiels de cette étape sont des outils de reporting, de data mining et d'OLAP (Online Analytical Processing).

OLAP est basé principalement sur des opérations d'exploration et de cumul (drill-down et roll-up). Les outils OLAP peuvent être classés en trois principaux types, à savoir MOLAP,

ROLAP et HOLAP. On parle de MOLAP ou OLAP Multidimensionnel quand les données sont stockées dans un système de gestion de base de données (SGBD) multidimensionnel. En ROLAP ou OLAP Relationnel, les données sont stockées dans un SGBD relationnel. HOLAP est, quant à lui, un hybride de MOLAP et ROLAP qui combine les avantages de l'un et de l'autre (Brezany *et al.* 2011a).

On distingue quatre types d'analyse : descriptive, prédictive, prescriptive et exploratoire (Chaudhuri, Dayal and Narasayya 2011) (Assunção *et al.* 2014) (Davenport 2006) (Minelli, Chambers and Dhiraj 2012).

Analyse descriptive

L'analyse descriptive utilise des données historiques pour identifier des patrons (patterns) et créer des rapports de gestion. Elle répond à la question "qu'est-ce qui s'est passé et/ou qu'est-ce qui se passe ?". Le résultat de l'analyse descriptive est la modélisation du comportement dans le passé.

Analyse prédictive

L'analyse prédictive tente de prédire l'avenir en analysant les données courantes et historiques. Elle répond à la question : "Qu'est-ce qui va se produire et/ou pourquoi cela se produira-t-il ?". Le data mining est l'outil principal de l'analyse prédictive. Le résultat de l'analyse prédictive est une projection permettant la prévision des événements futurs.

Analyse prescriptive

L'analyse prescriptive aide les analystes à prendre des décisions en déterminant les actions à entreprendre et en évaluant leur impact sur les objectifs, les exigences et les contraintes de l'entreprise. Les outils d'analyse prescriptive incluent la modélisation d'optimisation, la modélisation de simulation, la modélisation de décisions multicritères et les systèmes experts. Le résultat principal de l'analyse prescriptive est un meilleur plan d'action pour une organisation.

Analyse exploratoire

L'analyse exploratoire est souvent utile afin d'identifier une donnée potentielle intéressante pour une étude future ou pour guider la sélection des variables à inclure dans une analyse.

En outre, cette architecture de la BI (figure 2) identifie trois catégories d'utilisateurs impliqués dans l'étape d'analytique : les utilisateurs métiers, les analystes métiers et les spécialistes des données. Les utilisateurs métiers ont des compétences de base sur le domaine. Les analystes métiers sont des spécialistes qui peuvent analyser des données, comprendre comment les données sont organisées, récupérer des données via des requêtes ad hoc, produire des rapports spécialisés, construire des scénarios de simulation et interactivement effectuer une analyse plus approfondie pour appuyer la prise de décision. Un spécialiste des données possède de solides connaissances en mathématiques, en statistique et / ou en informatique. Il peut développer des modèles descriptifs et prédictifs, évaluer ces modèles, les déployer et les tester à l'aide d'études expérimentales (Phillips-Wren *et al.* 2015).

2.5.1.2. Analyse des Big data

Cette tendance du big data est perçue par les entreprises comme un moyen d'obtenir un avantage sur leurs concurrents : si une entreprise est capable de donner un sens aux informations contenues dans les gros volumes de données générées avec une grande vitesse, elle sera en mesure d'obtenir plus de clients, d'augmenter les revenus par client, d'optimiser son fonctionnement et de réduire ses coûts (Assunção *et al.* 2014). On parle alors de big data analytique.

Le big data est souvent caractérisé par sa variété, sa vitesse, son volume, sa véracité et sa valeur. La variété représente les types de données différents qu'il faut concilier. La vitesse se rapporte à la rapidité avec laquelle les données sont produites et traitées et le volume définit la quantité de données. La véracité fait référence à la fiabilité des données en fonction de la fiabilité de leurs sources, alors que la valeur correspond au rendement monétaire qu'une entreprise peut tirer de l'utilisation d'une donnée (Assunção *et al.* 2014) (Abbasi, Sarker and Chiang 2016).

Dans le contexte du big data, les organisations enregistrent plus de données que ne le requièrent leurs besoins immédiats, ce qui peut les exposer davantage aux risques de sécurité et aux problèmes légaux. Des mécanismes de gouvernance appropriés garantissant la conformité réglementaire et juridique sont cruciaux.

Par ailleurs, effectuer des tâches analytiques sur de grands volumes de données nécessite des méthodes spécifiques de stockage, de filtrage, de transformation et de récupération des données. Un aspect important pour la performance des applications de big data analytique est la localisation des données. Le big data étant caractérisé entre autres par son volume élevé, le transfert de ces données vers les différents nœuds (machines) pour leur traitement devient onéreux. MapReduce vient résoudre ce problème. Il présente un modèle plus adapté où la localisation des données est exploitée afin d'améliorer les performances des applications.

MapReduce et Hadoop

MapReduce est un modèle de programmation et une implémentation associée pour la gestion de grands ensembles de données (big data). En effet, les données d'entrée sont généralement de grande taille et les calculs doivent être répartis sur des centaines ou des milliers de machines afin de terminer le travail dans un délai raisonnable (Chaudhuri, Dayal and Narasayya 2011).

Ce modèle est basé sur deux étapes principales Map et Reduce (Wang, Song and Luo 2010) (Abelló, Ferrarons and Romero 2011) et chaque étape exécute une fonction bien déterminée. Dans l'étape de Map, le nœud (machine) sollicité divise le problème en sous-problèmes et délègue ces derniers à d'autres nœuds (ceux-ci peuvent faire la même chose à leur tour). En étape Reduce, les nœuds les plus bas transmettent leurs résultats aux nœuds parents qui les avaient sollicités. A la fin du processus, le nœud d'origine peut reconstituer la solution au problème qui lui avait été posé (Arres, Kabbachi and Boussaid 2013). A noter qu'on appelle cluster un regroupement de nœuds sur un réseau en vue d'augmenter la puissance de calcul.

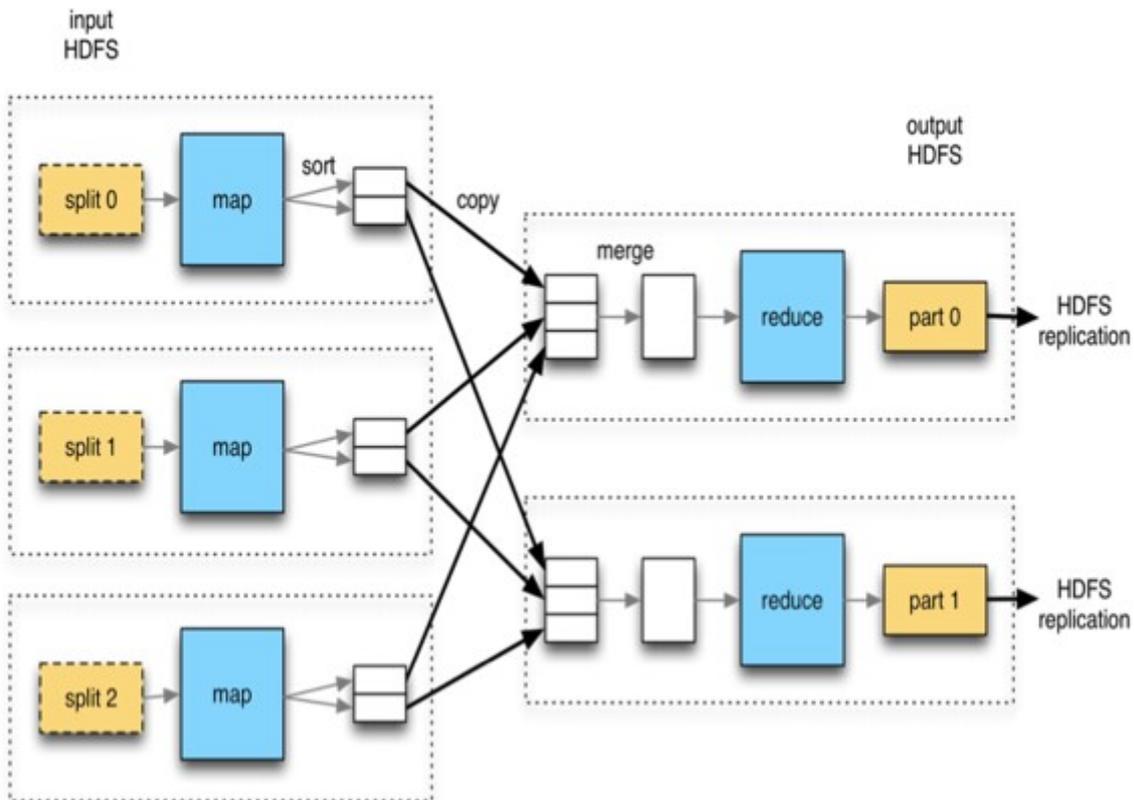


Figure 3 : Fonctionnement de MapReduce (2017)

La mise en œuvre la plus populaire du paradigme MapReduce est le cadre de référence Hadoop (Apache Hadoop 2017). Celui-ci est composé de deux grandes parties qui sont le système de fichier distribué et l'algorithme MapReduce :

- 1) Le système de fichier distribué "Hadoop Distributed File System " (HDFS) est un système de fichiers extensible et portable permettant de stocker de très gros volumes de données. HDFS comprend un nœud maître, un nœud maître secondaire ainsi que plusieurs nœuds esclaves. Le nœud maître gère le système des fichiers et les métadonnées. Le nœud maître secondaire vérifie régulièrement l'état du nœud maître et sauvegarde les métadonnées afin de prendre la relève en cas de panne du nœud maître. Les nœuds esclaves se chargent de stocker et de restituer les données.
- 2) L'algorithme MapReduce qui exécute respectivement les fonctions Map et Reduce comme nous l'avons expliqué plus haut.

Comme nous l'avons énoncé au point 2.5, la BI dans le cloud est la jonction de deux domaines : BI et cloud computing. Après avoir présenté l'état de l'art sur la BI, nous présentons ci-dessous l'état de l'art sur le cloud computing.

2.5.2. Cloud computing

Le cloud computing est un paradigme qui attire beaucoup d'entreprises du fait de ses potentialités. Selon les prévisions de Gartner, le marché du cloud public atteindra 383 milliards en 2020 (Gartner 2017b). Son objectif est de fournir des services innovants à la demande de différents types d'utilisateurs. Ces derniers sont libérés des détails d'infrastructure interne sous-jacente. Bien plus qu'une simple externalisation, dans le cloud computing, il y a deux notions qui sont mises en exergue : la virtualisation et l'agilité. La virtualisation est une abstraction des ressources informatiques (un serveur, un client, un stockage, des réseaux, des applications ou des systèmes d'exploitation) qui masque la nature physique et les limites de ces ressources auprès de leurs utilisateurs (Gartner 2017b). L'agilité consiste en le fait que le cloud est réactif et ses services flexibles par rapport aux besoins des clients.

De la définition de cloud computing découlent essentiellement les cinq caractéristiques suivantes (Yang and Tate 2012) :

- **Self-service à la demande** : On peut accéder aux services informatiques sans la nécessité d'une intervention et d'une interaction humaine supplémentaire du fournisseur.
- **L'accès à un réseau très large** : Il n'existe pas de limites géographiques dans le cloud computing. Partout où l'on se trouve, on peut bénéficier des services de cloud.
- **Mise en commun des ressources** : Des organisations mettent leurs données dans le cloud et celles-ci peuvent être utilisées par tous les autres clients.
- **Elasticité rapide** : Les capacités du cloud computing apparaissent comme illimitées par rapport aux consommateurs.
- **Services mesurés** : Les services sont payés généralement à la consommation.

Par ailleurs, la littérature sur le cloud computing traite essentiellement de l'architecture de cloud, de types de cloud ainsi que de cloud computing mobile.

2.5.2.1. *L'architecture de cloud*

Le cloud computing est généralement décomposé selon les trois couches suivantes (Yang and Tate 2012) (Rimal, Choi and Lumb 2009) :

- **Software as a service (SaaS)** : le cloud fournit des applications complètes au client qui le sollicite via l'Internet. Même des systèmes complexes tels que les progiciels de gestion de la relation client ou "Customer Relationship Management" (CRM) ainsi que les progiciels de gestion intégrés ou "Enterprise Resource Planning" (ERP) peuvent être fournis par le cloud. Les organisations clientes sont de simples consommateurs de ces ressources. Le SaaS fut l'une des premières implémentations du cloud computing.
- **Platform as a service (PaaS)** : les cadres de référence ainsi que les architectures des applications sont fournis et gérés par le cloud ainsi que toutes les infrastructures physiques pour la mise en œuvre des applications complètes ou de certains de leurs composants. C'est une couche logicielle, un environnement de développement intégré. Le PaaS fournit ainsi un langage de programmation avec les compilateurs et les bibliothèques pour les différents systèmes d'exploitation. Cela permet qu'une application développée par un client puisse être déployée dans des systèmes hétérogènes sans qu'il ne soit nécessaire de réécrire une partie ou la totalité du code.
- **Infrastructure as a service (IaaS)** : Les ressources matérielles sont fournies par le cloud. Il s'agit des serveurs de stockage, des serveurs de calcul ainsi que des bandes passantes. Un client IaaS loue les ressources informatiques au lieu de les acheter et de les installer. Il a l'avantage d'utiliser les ressources selon ses besoins et de payer la facture correspondant à sa consommation. L'IaaS peut fournir ses services aux besoins ponctuels d'un client dans le cas d'une tâche en temps réel. Mais le contrat de service comprend une capacité maximale de demandes que le client ne peut pas dépasser. Un type de clients spécifiques de l'IaaS est la communauté des chercheurs. Grâce au large volume d'infrastructure que l'IaaS offre comme service, ces chercheurs peuvent développer des tests et des analyses des données qui ne pourraient pas être possibles sans une infrastructure de calcul à cette grande échelle.

2.5.2.2. *Les types de cloud*

Il existe quatre types principaux de Cloud (Yang and Tate 2012) (Aceto *et al.* 2013) :

- **Cloud public** : Un cloud public fournit des services à de multiples clients en utilisant la même infrastructure partagée. Ses services sont gratuits ou payants à l'usage. Ce type de cloud est conseillé pour les données qui ne sont pas sensibles.
- **Cloud privé** : Un cloud privé délimite son "pool" de ressources informatiques sous-jacentes en créant une plateforme cloud distincte, à laquelle seule une organisation spécifique peut avoir accès. Ce type de cloud est particulièrement conseillé quand il s'agit de données sensibles (Nayem, Fahad, et Akhter 2013).
- **Cloud hybride** : Un cloud hybride est un service cloud intégré utilisant à la fois des clouds privés et des clouds publics pour remplir différentes fonctions au sein d'une même organisation. Ce type de cloud est conseillé quand il s'agit d'une même organisation qui possède à la fois des données sensibles et d'autres moins sensibles.
- **Cloud communautaire** : Un cloud communautaire est un service cloud où l'infrastructure est partagée par plusieurs entreprises qui ont des intérêts communs. On trouvera par exemple un tel cloud pour les universités ou encore pour un domaine d'activité bien précis.

2.5.2.3. *Cloud computing mobile*

Les dispositifs mobiles (par exemple : smartphone et tablette) deviennent de plus en plus une partie essentielle de la vie, des outils de communication les plus efficaces et les plus pratiques, non limités par le temps et le lieu. Les utilisateurs mobiles accumulent une expérience enrichissante de divers services à partir des applications mobiles (par exemple, des applications iPhone et des applications Google), qui fonctionnent sur les périphériques et / ou sur des serveurs distants via des réseaux sans fil. Cette augmentation de l'utilisation des dispositifs mobiles suscite plusieurs questions (Fernando, Loke and Rahayu 2013). Nous citons :

- Les questions liées au niveau opérationnel : ce sont des questions relatives aux technologies sous-jacentes telles que les méthodes pour décharger les périphériques mobiles de leurs tâches de calculs, les modèles coûts-avantages qui aident à décider de décharger ou non les périphériques de leurs tâches de calculs, la façon dont la mobilité des périphériques est gérée et les protocoles de connexion utilisés.
- Les questions liées aux utilisateurs finaux : ce sont celles qui se réfèrent à la collaboration, à la présentation et aux problèmes d'utilisation.

- Les questions liées aux services et applications : ce sont les questions relatives à la tolérance aux pannes pour les exigences de disponibilité, la performance de maintien au niveau de service et des API (Application Programming Interface) dans le cloud.
- Les questions liées à la sécurité, la confidentialité et la confiance : ce sont celles qui se rapportent à la sécurité générale et au respect de la vie privée dans le cloud. En effet, les protocoles de routage sécurisés doivent être utilisés pour protéger le canal de communication entre les dispositifs mobiles et le cloud (Khan *et al.* 2013).
- Les questions liées à la sensibilité au contexte : ce sont celles qui concernent les éléments tels que l'éclairage, le niveau de bruit, la connexion au réseau, les coûts de communication, la bande passante de communication et même la situation sociale. Ainsi, les trois aspects importants du contexte sont l'emplacement de l'utilisateur, d'autres utilisateurs dans le voisinage et les ressources disponibles dans l'environnement de l'utilisateur.
- Les questions liées à la gestion des données : ce sont celles qui concernent l'accès aux données, la portabilité et l'interopérabilité des données vu l'hétérogénéité des dispositifs mobiles.

Pour répondre à ces questions, le cloud computing mobile a été présenté comme une technologie potentielle pour les services mobiles. Il intègre le cloud computing dans l'environnement mobile et surmonte les obstacles liés à la performance (par exemple, la durée de vie de la batterie, le stockage et la bande passante), à l'environnement (par exemple, l'hétérogénéité, l'évolutivité et la disponibilité) et à la sécurité (par exemple, la fiabilité et la confidentialité) discutées dans l'informatique mobile (Dinh *et al.* 2013).

Après avoir présenté l'état de l'art sur la BI au point 2.5.1 et sur le cloud computing au point 2.5.2, ci-dessous, nous présentons l'état de l'art sur la BI dans le cloud.

2.5.3. BI dans le cloud

Comme nous l'avons énoncé au point 2.4.4.1, le but de cette revue systématique de la littérature est, d'une part, de souligner l'apport des chercheurs en science de conception sur la BI dans le cloud. D'autre part, elle permet de synthétiser l'état actuel de la littérature sur la BI dans le cloud.

En ce qui concerne l'apport des chercheurs en science de conception, la recherche dédiée au domaine de la BI dans le cloud est très active. Elle a conduit à de nombreux artefacts pour aider les organisations à adopter la solution cloud. Le déploiement de la BI dans le cloud a généré rapidement de nouveaux besoins de la part des utilisateurs tels qu'avoir des données et des outils analytiques disponibles partout en utilisant des dispositifs mobiles. La mobilité implique aussi la capacité à générer et à prendre en compte des informations contextuelles.

Ainsi, dans les articles de notre échantillon, nous avons répertorié principalement les types d'artefacts suivants :

- Des cadres de référence pour la migration de la BI vers le cloud ainsi que des cadres de référence présentant la structure logique des services que le cloud peut fournir à la BI.
- Des architectures qui représentent la structure générale de la BI dans le cloud ainsi que des architectures pour les outils d'analyse de données dans le cloud.
- Des algorithmes pour organiser les données en vue de leur analyse dans le cloud.
- Des implémentations des systèmes pour l'analyse de données dans le cloud.

Ces artefacts sont regroupés dans un cadre de référence à deux dimensions où chaque ligne du cadre représente un type d'artefact et chaque colonne, une étape de la BI. Nous présentons ici un extrait de ce cadre dont l'intégralité est fournie en annexe (ANNEXE 1).

Artéfacts	Collection et consolidation des données	Modélisation et stockage des données	Analytique
Cadre de référence	<ul style="list-style-type: none"> • Cadre de référence pour l'intégration des données structurées et semi-structurées (Negash 2004) • Cadre de référence pour l'approvisionnement des ressources à partir d'un cloud public (Mian, Martin and Vazquez-Poletti 2013) 	<ul style="list-style-type: none"> • Cadre de référence pour générer des hypercubes (Tapiador <i>et al.</i> 2014) • Cadre de référence pour un système de gestion de base de données dans le cloud (Chi <i>et al.</i> 2011) 	<ul style="list-style-type: none"> • Cadre de référence pour la technologie BI&A (Shanks <i>et al.</i> 2010) • Cadre de référence pour la gestion des big data analytiques dans le cloud (Chaisiri <i>et al.</i> 2012) • Cadre de référence pour d'analyse de données dans le cloud (Yuan and Herbert 2014) • Cadre de référence de la BI et OLAP (Al-Aqrabi <i>et al.</i> 2015) (Al-Aqrabi <i>et al.</i> 2012) • Cadre de référence pour l'analyse des données dans le cloud (Włodarczyk <i>et al.</i> 2010)
<ul style="list-style-type: none"> • Cadre de référence pour la BI dans le cloud (Juan-Verdejo <i>et al.</i> 2014) • Cadre de référence de programmation de recherche dans le domaine de la BI (Dinter, and Lorenz 2012) (Kowalczyk, Buxmann and Besier 2013) • Cadre de référence des services dont la BI bénéficie dans le cloud (Baars and Kemper 2010) • Cadre de référence pour assurer la sécurité dans le cloud (Subashini and Kavitha 2011) 			
Architecture		<ul style="list-style-type: none"> • architecture d'un entrepôt de données dans le cloud (Kaur, Agrawal and Dhiman 2012) • Architecture de MapReduce axée sur les objectifs (AbdelBaky <i>et al.</i> 2012) • architecture d'un système de stockage de données élastique (Cao <i>et al.</i> 2011) • Architecture pour la construction des cubes OLAP dans le cloud (Arres, Kabbachi and Boussaid 2013) 	<ul style="list-style-type: none"> • Architecture pour l'agrégation des données en ligne dans le cloud (Gan, Meng and Shi 2013) • Architecture pour l'analyse des données dans le cloud (Stanoevska-Slabeva, Wozniak and Ristol 2010) • Architecture d'un système qui permet de maintenir l'élasticité d'une plateforme cloud pour OLAP (Brezany <i>et al.</i> 2011b) • Architecture d'un système de mise en échelle de moteur d'exécution optimisé pour l'analyse des données dans le cloud (Gatzioura <i>et al.</i> 2012) • Architecture d'un système pour l'analyse des données dans le cloud (Garey and Johnson 2009)
<ul style="list-style-type: none"> • Architecture d'un système pour assurer la sécurité dans le cloud (Ng <i>et al.</i> 2011) <ul style="list-style-type: none"> • Architecture de la BI dans le cloud (Kasem and Hassanein 2014) • Architecture de la BI (Chaudhuri, Dayal and Narasayya 2011) • Architecture de cloud computing mobile (Khan <i>et al.</i> 2013) (Dinh <i>et al.</i> 2013) • Architecture de cloud computing (Qian <i>et al.</i> 2009) 			

	<ul style="list-style-type: none"> • Architecture d'aide à la décision (Demirkan and Delen 2013) 		
Algorithme		<ul style="list-style-type: none"> • Algorithme pour construire des cubes parallèles basés sur MapReduceMerge (Wang, Song and Luo 2010) • Algorithme pour construire des cubes avec MapReduce (Abelló, Ferrarons and Romero 2011) 	<ul style="list-style-type: none"> • Algorithme pour les requêtes d'agrégation (Hua, Liu and Jiang 2014) • Algorithme pour le traitement de plusieurs requêtes de jointure dans le cloud (Kurunji <i>et al.</i> 2012)
	<ul style="list-style-type: none"> • Algorithme pour la planification des ressources de cloud (Lee <i>et al.</i> 2012) 		
Implémentation d'un system			<ul style="list-style-type: none"> • Outil pour l'analyse des données dans le cloud (Stanoevska-Slabeva, Wozniak and Ristol 2010) • Outil pour l'agrégation en ligne dans le cloud (Gan, Meng and Shi 2013) • Outil d'une mise en échelle d'un moteur d'exécution optimisé pour les données analytiques (Barga, Ekanayake and Lu 2012) • Outil qui permet de maintenir l'élasticité d'une plateforme cloud pour OLAP (Brezany <i>et al.</i> 2011b) • Outil qui permet de combiner la mise en échelle des applications google avec l'analyse des données hors connexion (Chohan <i>et al.</i> 2012)
	<ul style="list-style-type: none"> • Outil pour incorporer BI dans le marché de cloud (Gatzioura <i>et al.</i> 2012) • Outil pour la reconfiguration des applications dans le cloud (Goncalves, Assuncao and Cunha 2012) 		

Tableau 2 : Extrait du cadre de référence de la BI dans le cloud

La BI requiert la capacité d'acquérir l'information, de la stocker et de l'analyser en utilisant différents outils et techniques. Ces trois aspects (acquérir, stocker et analyser l'information) sont souvent traités par différentes équipes de recherche. Par conséquent, un artefact peut être développé pour un aspect spécifique de la BI. Chaque aspect constitue une étape de la BI.

En ce qui concerne la synthèse de l'état actuel de la littérature sur la BI dans le cloud, notons que, comme pour la BI en local, les tâches de la BI dans le cloud se regroupent en trois étapes :

- collection et consolidation des données,
- modélisation et stockage des données,
- analytique.

Avant d'aborder chacune de ces étapes en particulier, nous traiterons d'abord de généralités qui sont valables pour toutes (la BI dans le cloud en général).

Pour la BI dans le cloud en général, la littérature aborde essentiellement les éléments suivants :

- les services dont la BI peut bénéficier dans le cloud,
- l'infrastructure matérielle,
- la sécurité,
- les avantages de la BI dans le cloud,
- les défis à relever.

En effet, les organisations migrent leur BI vers le cloud pour profiter des services que celui-ci fournit. Afin d'accéder à ces services, les organisations ont besoin d'une infrastructure matérielle. Celle-ci doit être connectée à l'Internet qui est la principale voie de transmission des services du cloud. En outre, les ressources dans le cloud sont partagées entre plusieurs clients. Il est donc de bon aloi que des mesures de sécurité soient mises en place. Par ailleurs, la BI dans le cloud ne procure pas seulement des avantages, il y a aussi des défis à relever.

1) Les services dont la BI peut bénéficier dans le cloud

Dans le cadre du cloud computing, un service est défini comme l'application de compétences et de connaissances pour créer de la valeur entre le fournisseur et le client (Baars and Kemper 2010). Un client qui bénéficie des services de cloud est appelé locataire (Muriithi and Kotzé 2013).

Le cloud peut mettre à la disposition de la BI les éléments suivants (Kasem and Hassanein 2014) :

- le matériel : le stockage et les réseaux,
- les logiciels : les systèmes d'exploitation et les pilotes nécessaires pour gérer le matériel, les langages de programmation, les architectures et les cadres de référence,
- les outils : ETL, data warehouses, OLAP, data mining, reporting.

Notons qu'une organisation peut migrer une partie ou la totalité de sa BI vers le cloud (Juan-Verdejo *et al.* 2014). Le processus de migration de la BI comprend trois étapes essentielles (Muriithi and Kotzé 2013)

- L'analyse de la situation existante afin d'identifier les étapes de la BI potentielles pour la migration vers le cloud en cas de migration partielle.
- L'évaluation des impacts de la migration sur l'organisation. Cette évaluation se fait par rapport au bénéfice que cette migration peut apporter à l'entreprise, par rapport à la faisabilité technique de la migration, par rapport au risque que l'entreprise peut courir en envisageant la migration et par rapport à l'impact organisationnel de cette migration.
- La mise en œuvre de la solution cloud pour l'étape ou les étapes identifiées. La performance des services alloués par le cloud est régulièrement évaluée afin de déterminer dans quelle mesure ces services répondent aux besoins de l'organisation.

Nous pouvons représenter schématiquement le processus de migration comme suit (Figure 4) :

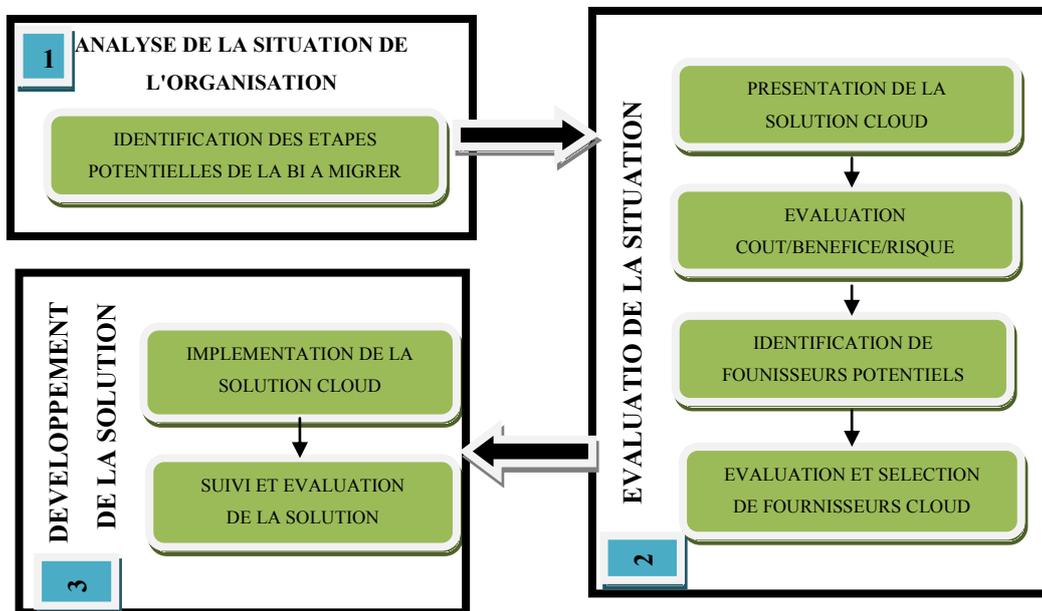


Figure 4 : Cadre de la migration de la BI vers le cloud (traduit et adapté de Muriithi et Kotzé (2013))

Un cadre des services dont la BI bénéficie dans le cloud

Ce cadre (Figure 5) aide à l'identification, à la combinaison et finalement à l'évaluation des services potentiels dont la BI peut bénéficier dans le cloud.

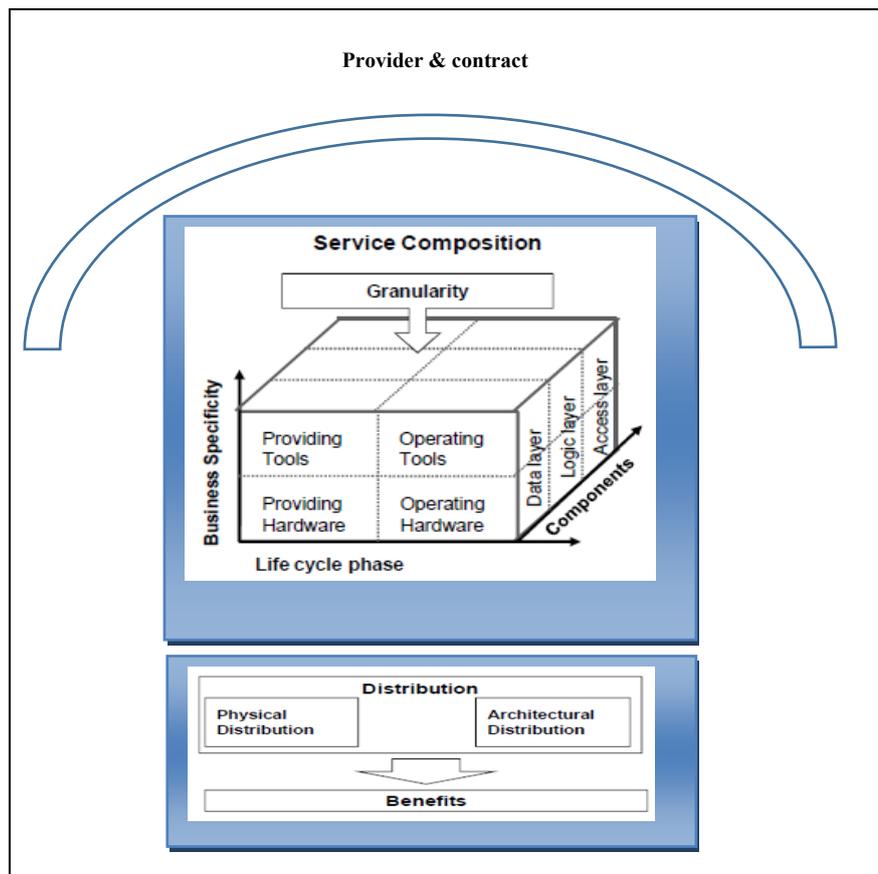


Figure 5 : Cadre des services dont la BI bénéficie dans le cloud (Baars and Kemper 2010)

Ce cadre est constitué de quatre composants : le fournisseur et le contrat, la composition de service, la distribution et les avantages.

- **Le fournisseur et le contrat** : On doit soigneusement vérifier si le fournisseur est digne de confiance comme dans tous les accords de sous-traitance. Le contrat doit mentionner les promesses du fournisseur de cloud en matière de disponibilité, de sécurité, de flexibilité, d'évolutivité et de fiabilité des services à fournir.
- **La composition du service** : elle peut être obtenue en spécifiant la granularité appliquée sur la couche outils (solution, composant ou service web) et en définissant les services BI à l'aide des dimensions des composants, des spécifications métier et des phases du cycle de vie ciblées.
- **La distribution** : Pour la BI, la distribution a deux faces. On parle de la distribution physique et de distribution architecturale. Dans la distribution physique, le client

sollicite un fournisseur de cloud pour une granularité de service déterminée. En revanche, dans la distribution architecturale, une organisation sollicite plusieurs fournisseurs de cloud. Chacun d'eux fournit un service de bout en bout. Mais ces services combinés forment une seule solution pour le client.

- **Les avantages** : Par exemple, sur le plan financier, on doit constater que la migration de la BI vers le cloud est moins onéreuse que l'externalisation traditionnelle. La flexibilité, l'évolutivité, la performance doivent caractériser les services dont la BI bénéficie de la part du cloud.

En définitive, ce cadre des services dont de la BI bénéficie dans le cloud (Figure 5) illustre tous les éléments qui entrent en ligne de compte quand un client sollicite un service de cloud :

- Le client s'adresse à un fournisseur et un contrat est établi.
- La composition de service que le fournisseur s'engage à fournir au client est un des scénarii que nous présentons ci-après.
- Le client peut solliciter la distribution physique ou la distribution architecturale des services selon ses besoins et ses possibilités.
- Le service dont le client bénéficie doit lui procurer les avantages qu'il n'aurait pas s'il n'avait pas sollicité les services du cloud.

Les scénarii possibles pour la BI dans le cloud

La composition de service est un des éléments du cadre de la BI dans le cloud. Chaque scénario représente une composition de service dont la BI peut bénéficier dans le cadre du cloud. Il existe six scénarii possibles de services pour la BI dans le cloud (Baars and Kemper 2010) :

- **Scénario services complémentaires** : Quelques composants du cloud sont sélectionnés pour l'infrastructure BI, par exemple les composants pour la recherche sur le web. Cette approche présente moins de risque par rapport aux autres scénarii en raison de sa simplicité (SaaS).
- **Scénario remplacement d'outil** : Le cloud prend en charge un outil complet. Il s'agit ici de SaaS. Par exemple un data mart ou un outil OLAP (SaaS).
- **Scénario fourniture de solution** : Le cloud prend en charge une solution bien définie du début à la fin. Il fournit toutes les ressources logicielles et matérielles. La motivation pour ce scénario est similaire au remplacement d'outil, mais ici, le champ

d'application est beaucoup plus large. Il est adapté pour un projet d'exploration des données limité dans le temps ou pour de nouveaux types d'applications (SaaS).

- **Scénario de réseau d'entreprise** : Le fournisseur de solution agit dans les limites d'un réseau d'entreprise. Cela peut être un marché B2B (Business to Business) ou une chaîne d'approvisionnement, etc. L'aspect cloud réside dans l'abstraction physique avec l'infrastructure du fournisseur qui sera virtualisée.
- **Scénario solution mixte** : le remplacement de l'outil est poussé à un plus haut niveau, au point où tous les composants de l'infrastructure BI peuvent être offerts par des fournisseurs externes. Le résultat peut être une infrastructure BI entièrement virtualisée avec tous les avantages d'une meilleure allocation des ressources. En revanche, ce scénario est actuellement entravé par l'incapacité du cloud à fournir certains outils BI. Par conséquent, les outils ou les composants des outils manquants sont mis en place par les organisations elles-mêmes en utilisant la couche PaaS.
- **Scénario application BI composite** : Ici, une solution de BI est librement composée à partir d'Internet. Par rapport au scénario de solution mixte, ce scénario ajoute une granularité plus fine et met l'accent sur la combinaison et le développement rapides. Les avantages supplémentaires résident principalement dans son agilité (PaaS).

2) L'infrastructure matérielle

Pour accéder aux services dont la BI bénéficie de la part du cloud, les organisations ont besoin d'un minimum d'infrastructure matérielle, principalement des ordinateurs de bureau et des dispositifs mobiles connectés à Internet. Ces dernières années, l'utilisation de dispositifs mobiles a considérablement augmenté. Avec l'avènement du téléphone intelligent (smartphone) et des tablettes, le système d'aide à la décision mobile est devenu émergent. Les résultats de la recherche de Gartner indiquent qu'un tiers des fonctionnalités de la BI sera consommé via des dispositifs mobiles (Nayem, Fahad and Akhter 2013). Ce système est bénéfique pour des domaines d'application où les décisions critiques doivent être prises sous la pression du temps et quand les décideurs sont en mobilité.

3) Sécurité

Par sécurité, on entend la confidentialité, l'intégrité et la disponibilité des ressources déployées dans le cloud (Chang *et al.* 2015).

Pour accéder aux services que le cloud lui fournit, le client connecte son matériel à l'Internet. Celui-ci reste la principale voie de transmission des services que le cloud fournit à ses clients.

Les données venant des bases des données relationnelles vers un entrepôt de données implémenté dans le cloud circulent via l'Internet. Les rapports et analyses des données, destinés à des organisations dont l'étape analytique est implémentée dans le cloud, circulent, eux aussi, via l'Internet. Par ailleurs, le partage des ressources dans le cloud pose des problèmes d'intégrité, de confidentialité et de disponibilité (Muriithi and Kotzé 2013) (Gash, Ariyachandra and Frolick 2011). Eu égard à tout ce qui été dit précédemment, il est opportun que des mesures de sécurité adéquates soient mises en place à chaque étape de la BI, c'est-à-dire de la collecte et la consolidation des données à l'analytique.

En outre, la sécurité est souvent considérée comme le principal obstacle à l'adoption de services du cloud. Environ 75% des praticiens dans le domaine de la BI considèrent la sécurité comme le risque majeur pour la migration de la BI vers le cloud (Kasem and Hassanein 2014). Pour certaines entreprises, il est essentiel de conserver les données dans leurs serveurs locaux car leurs données sont confidentielles. Cependant, il existe plusieurs solutions pour sécuriser les données, y compris le cryptage, mais il incombe à l'organisation de crypter correctement les données sur le cloud. Malgré le processus de virtualisation dans toute technologie cloud, on peut constater des brèches de sécurité même lorsque les index des données sont supprimés. Par conséquent, les fournisseurs de cloud sont appelés à relever les défis de sécurité et de confiance pour l'adoption du cloud en général et pour le déploiement de la BI dans le cloud en particulier.

4) Les avantages de la BI dans le cloud

De nos jours, la migration de la BI vers le cloud gagne de plus en plus de popularité parmi les entreprises. Celles-ci valorisent les avantages de l'analyse des données. Les fournisseurs SaaS servent d'interface principale à la communauté des utilisateurs professionnels. Cloud pour la BI (cloud BI) est un concept qui désigne la prestation des capacités de BI en tant que service. Voici les avantages clés du cloud computing pour BI (Al-Aqrabi *et al.* 2015).

- Rapport coût-efficacité : Dans le cloud, les entreprises n'ont pas besoin de budget pour de gros achats de logiciels ou de longues mises à jour sur les serveurs locaux afin de mettre l'infrastructure BI en service. Elles ne paient que les ressources informatiques dont elles ont besoin.
- Flexibilité et évolutivité : Les solutions cloud pour la BI permettent une plus grande flexibilité. Les utilisateurs professionnels peuvent conserver un meilleur contrôle fiscal sur les projets informatiques et ont la flexibilité d'étendre ou de réduire l'utilisation à mesure que les besoins changent. En outre, dans le cloud, les ressources peuvent automatiquement et rapidement s'accroître et elles peuvent supporter un grand nombre d'utilisateurs simultanés. Cela signifie que les clients peuvent facilement augmenter leur utilisation du logiciel sans délai ni coût de déploiement et d'installation de matériel et de logiciels supplémentaires.
- Fiabilité : La fiabilité s'améliore grâce à l'utilisation de plusieurs sites redondants qui peuvent fournir une fiabilité et des emplacements sécurisés pour le stockage de données. En outre, les ressources peuvent être réparties sur un grand nombre d'utilisateurs, ce qui rend le cloud computing adapté à la reprise après sinistre et à la continuité des activités.
- Amélioration des capacités de partage de données : Les applications dans le cloud permettent d'échanger les données à distance. Elles possèdent des fonctionnalités de partage de données déployées via Internet et en dehors du pare-feu d'une entreprise.
- Pas de dépenses importantes : Le TCO ou " total cost of ownership" (coût total de possession) faible est un atout majeur du modèle cloud. En effet, le cloud permet aux entreprises de payer uniquement le service qu'elles utilisent réellement. Avec cette politique, les entreprises contrôlent mieux leurs investissements en consacrant des budgets de fonctionnement aux activités essentielles.

5) Les défis que la BI dans le cloud doit relever

Comme nous l'avons dit ci-haut, la migration de la BI vers le cloud n'apporte pas seulement des avantages, il y a aussi des défis à relever. Nous citons les principaux (Al-Aqrabi *et al.* 2015) :

- La conformité des applications BI aux normes architecturales des services Web et aux normes définies par le fournisseur SaaS ou PaaS (comme les normes Google Apps par

exemple). En effet, un service web englobe un concept et une infrastructure logicielle pour la communication entre les programmes et la livraison de composants d'applications. Le concept de service Web traite le logiciel comme un ensemble de services accessibles sur un réseau universel utilisant des normes et des protocoles basés sur le Web (Gartner IT Glossary).

- Le déploiement d'un système d'entreposage de données massivement parallèles avec une charge de requête uniformément répartie.
- L'architecture du réseau doit être conçue de telle sorte que la charge des requêtes puisse être répartie uniformément parmi les serveurs d'un réseau. Cela assurera de temps de réponse uniforme au traitement des requêtes.

Après avoir parlé de la BI dans le cloud en général, rappelons que les tâches de la BI se regroupent en 3 étapes : la collecte et la consolidation des données, la modélisation et le stockage des données ainsi que l'analytique. Les services que le cloud fournit à la BI peuvent être spécifiques pour chacune de ces étapes. Dans les lignes qui suivent, nous allons ainsi relever les éléments spécifiques à chacune d'elles.

2.5.3.1. La collecte et la consolidation des données

La première étape de la BI consiste à collecter et à consolider les données afin de mettre à jour l'entrepôt des données. Dans cette étape, nous parlerons du processus ETL (Extract, Transform, Load).

Pour aider les décideurs, les données non structurées ou semi-structurées peuvent être tout aussi importantes que les données structurées. Dans les entrepôts de données internes des organisations, les données sont structurées tandis que, dans le cloud, les données sont souvent aussi semi-structurées ou non structurées. Pour manipuler les données de ces deux sources ensemble, il sied de les intégrer. Des processus d'acquisition, de nettoyage et d'intégration sont requis tant pour les données structurées que pour les données semi ou non structurées (Negash 2004). Ainsi, la migration des outils ETL dans le cloud pourra faciliter et automatiser l'intégration des données quels que soient leur format, leur structure et leur source.

Par ailleurs, les outils ETL déployés dans le cloud permettent aussi d'échanger les données entre différents clouds spécialement dans le cas des clouds hybrides (Clay, Shen and Ma 2013).

Notons qu'actuellement, les recherches se poursuivent afin d'ajouter des fonctionnalités supplémentaires aux outils ETL. Ces fonctionnalités permettront aux outils ETL d'effectuer non seulement l'intégration des données et la mise à jour des entrepôts des données, mais également de renvoyer les rapports des analyses à des interfaces utilisateurs (Herwig 2013). Seuls les utilisateurs finaux d'une application auront accès aux rapports relatifs à cette application. C'est au fournisseur de cloud de garantir l'intégrité et la confidentialité des données à ce niveau (Chang *et al.* 2015).

2.5.3.2. La modélisation et le stockage des données

Les données collectées et consolidées lors de la première étape de la BI sont modélisées et stockées dans les entrepôts des données à des fins d'analyse. Nous parlerons ici de la mise en place des entrepôts de données et des systèmes de gestion de bases de données. En effet, l'élément central de cette étape est l'entrepôt de données. Par ailleurs, cette étape englobe les modèles, les méthodes, les techniques et les outils de gestion de grands volumes de différents types de données.

1) La mise en place des entrepôts des données

La BI dans le cloud offre trois possibilités pour mettre en œuvre et gérer un entrepôt de données. La première se produit quand une organisation dispose déjà d'un entrepôt de données et envisage de le transférer vers le cloud. La deuxième consiste à créer un entrepôt de données directement dans le cloud et à le stocker aussi dans le cloud. Enfin, la troisième situation est celle d'une organisation qui gère son entrepôt des données dans ses propres serveurs, mais sollicite de la part du cloud des outils d'administration d'entrepôt de données (Kaur, Agrawal and Dhiman 2012).

2) Les systèmes de gestion de bases des données

Un système de gestion de bases de données (SGBD) est un logiciel permettant de stocker et d'organiser les données des formats et des structures définis (Gartner IT Glossary). Depuis plusieurs décennies, les SGBD les plus usuels sont les SGBD relationnels (SGBDR).

Avec l'avènement des big data, les SGBDR sont devenus inadaptés dans ce cadre. Pour résoudre ce problème, dans le cloud, on utilise généralement des SGBD NoSQL (Not Only SQL) (Arres, Kabbachi and Boussaid 2013) (Cao *et al.* 2011).

En effet, les SGBD NoSQL sont très performants et permettent de stocker de grandes quantités de données hétérogènes dans des environnements distribués. Trois faits essentiels caractérisent la plupart des systèmes NoSQL (Bruchez 2015) :

- Le schéma de données est implicite : Ces SGBD ne contrôlent pas le schéma des données. C'est à l'application cliente de structurer les données qu'elle désire manipuler.
- Le langage de manipulation des données proche de SQL : Les SGBD NoSQL implémentent généralement des langages déclaratifs proches du SQL pour la manipulation des données.
- Le logiciel libre : Beaucoup de SGBD NoSQL sont des logiciels libres.

Les fournisseurs de cloud offrent des infrastructures de stockage fiables et évolutives utilisant ces différents SGBD. Les plus usuels sont (Cao *et al.* 2011) (Bruchez 2015) (Abelló, Ferrarons and Romero 2011) :

- BigTable : fournit un système de stockage distribué orienté colonne pour la gestion des données structurées. Dans un système orienté colonne, les données sont stockées dans des familles de colonnes en tant que lignes. Plusieurs colonnes sont associées à une clé qui les identifie. Une famille peut donc être composée de colonnes différentes pour chaque ligne. BigTable ressemble à une gigantesque table de hachage distribuée qui incorpore des mécanismes permettant de gérer la cohérence et la distribution des données sur GFS (Google File System). Les nœuds (machines) de cluster BigTable sont ajoutés ou supprimés de manière dynamique. Outre son API native basée sur le protocole RPC (Remote Procedure Call), BigTable offre une interface compatible avec Hbase permettant, ainsi, la portabilité des applications entre HBase et BigTable.
- HBase : fournit un système de stockage distribué qui utilise HDFS (Hadoop Data File system) pour stocker les données qu'il gère. Ce système permet un accès aléatoire en temps réel à un très grand ensemble de données. HBase est orienté colonne. Il est évolutif et tolérant aux pannes car la charge de travail, en termes de mémoire et de calcul (CPU) ainsi que de stockage, est distribuée sur toutes les machines du cluster. Il est adapté pour la gestion des données à des fins d'analyse.

- DynamoDB : Ce système met à disposition une API pour accéder aux données à l'aide de langages comme Java ou PHP. DynamoDB est sans schéma. Il gère les données structurées ou semi-structurées, y compris les documents JSON (JavaScript Object Notation). C'est un système de stockage clé-valeur hautement disponible dans un environnement distribué. Dans un système clé-valeur, les données sont représentées par un couple clé/valeur : une clé primaire identifie de façon unique chaque élément de données.
- Cassandra : offre un système orienté colonne. Il s'inspire du modèle BigTable de Google et du modèle Dynamo d'Amazon. Les données sont regroupées en familles de colonnes. Une famille de colonnes peut être comparée à une table dans le modèle relationnel. La différence avec le modèle relationnel réside dans le fait qu'avec Cassandra, le schéma des données n'est pas fixé à l'avance. Cassandra implémente un langage de requête proche de SQL appelé CQL.
- MongoDB : c'est un système de gestion NoSQL codé avec le langage C++ à la différence des autres moteurs NoSQL codés en Java ou en d'autres langages particuliers comme Erlang. Ce système de gestion de base de données est orienté documents. Un document signifie ici une représentation structurée, compréhensible par MongoDB, d'une donnée complexe appelée BSON (Binary Serialized dOcument Notation). Un document BSON représente l'unité stockée par MongoDB. Il équivaut à une ligne dans une table relationnelle. Il est composé de paires clé-valeur. Dans un système MongoDB, une requête est écrite en JSON.

Notons que les données dans le cloud sont distribuées aux différents nœuds (machines) en fonction de la disponibilité des ressources. La localisation des données peut changer de façon dynamique (Kurunji *et al.* 2012). Cela garantit la disponibilité et la performance des services pour les organisations qui migrent leur BI vers le cloud. En outre, les organisations qui choisissent de migrer leur entrepôt de données vers le cloud peuvent également profiter de grandes quantités de données (big data) contenues dans le cloud. Selon la sensibilité de leurs données, elles opteront pour la migration d'une partie ou de la totalité de leur entrepôt de données.

2.5.3.3. L'analytique

Des entreprises comme Google, Facebook, Twitter, LinkedIn, etc., gèrent aujourd'hui de plus grands ensembles de données qui sont analysées en ligne par les utilisateurs dans le but

de trouver les informations nécessaires pour une prise de décisions pertinentes. Les applications analytiques de données sont caractérisées par de grands ensembles de données soumis à une série de phases de traitement. Certaines de ces phases sont exécutées séquentiellement, mais d'autres peuvent être exécutées simultanément ou en parallèle sur des clusters, des grilles ou des clouds. Le modèle de programmation MapReduce est recommandé pour traiter de grands ensembles de données dans les environnements en cluster et dans le cloud.

En outre, l'analyse des données devient un atout pour les décideurs, car elle ouvre de nouvelles possibilités, de nouvelles connaissances pertinentes qui peuvent permettre à une organisation de se positionner dans un marché concurrentiel. Afin de faire face à la grande quantité de données, l'approche à adopter doit être différente de l'approche traditionnelle où les données sont demandées et analysées par la suite en local. Une option est de passer à des environnements de cloud où l'on peut télécharger les flux de travail d'analyse de données afin qu'ils s'exécutent dans un environnement à faible latence (Tapiador *et al.* 2014).

Ainsi, il existe deux possibilités : l'analyse peut être effectuée par le cloud ou les organisations peuvent effectuer l'analyse des données dans le cloud en utilisant la couche PaaS (Patel, Rau-Chaplin and Varghese 2012).

1) Le cloud fournit l'étape analytique

Ici, le cloud analyse les données et retourne les résultats aux utilisateurs via l'Internet. Les organisations ont besoin quand même de personnel qualifié : des analystes pour interpréter les résultats et construire les modèles complexes en cas de besoin (l'analyse prédictive).

2) L'analyse des données effectuée dans le cloud

Dans ce deuxième cas, le cloud fournit une plateforme analytique qui a une puissance de calcul élastique (grâce au parallélisme) avec une réplication de données (même dans différentes régions du monde) et une forte tolérance aux pannes (une nouvelle machine reprend automatiquement le travail d'une autre tombée en panne sans ré-exécuter la requête ou le processus entier) (Abelló, Ferrarons and Romero 2011). Les organisations gèrent elles-

mêmes leurs applications et leurs données. Elles ont besoin de plus de personnel qualifié que dans le premier cas car l'analyse se fait en local. C'est grâce aux applications clientes que les utilisateurs interagissent avec le cloud. Celles-ci fournissent les interfaces permettant aux utilisateurs analystes d'accéder à la couche PaaS pour effectuer l'analyse (Chaudhuri, Dayal and Narasayya 2011). Parmi les applications clientes, on trouve :

- les tableurs,
- les portails de recherche,
- les interfaces pour poster des requêtes ad hoc,
- les tableaux de bord visuels.

Après avoir effectué l'état de l'art, nous avons examiné le cadre de la BI dans le cloud qui se trouve en annexe de ce travail pour déceler les opportunités de recherche en science de conception dans ce domaine de la BI dans le cloud.

2.6. Opportunités et question de recherche

Dans cette section, nous définissons les possibilités de recherche en science de conception pour la BI dans le cloud et nous justifions l'importance de notre question de recherche : "Quels sont les nouveaux artefacts à implémenter pour que la BI tire profit du cloud computing ?" Nous nous basons sur la synthèse des informations résultant de notre revue systématique de la littérature.

Le cadre de la BI dans le cloud qui se trouve en annexe (ANNEXE 1) montre la richesse des apports de la science de conception pour le domaine de la BI dans le cloud d'une part ; d'autre part, elle montre les efforts fournis par les chercheurs pour que la BI tire profit des potentialités du cloud. Dans ce sens, de nombreux artefacts ont été mis en œuvre, principalement des architectures, des cadres de référence, des algorithmes et des implémentations du système.

En revanche, si l'on considère la typologie des artefacts, on peut constater l'inexistence de certains artefacts dans la littérature à notre disposition. Ainsi, aucun auteur n'a proposé un artefact classifiant tous les éléments importants qui peuvent influencer la décision de migration de la BI vers le cloud. La littérature ne propose pas non plus d'artefact qui classifie toutes les options du cloud que les organisations peuvent choisir en tenant compte de

leurs spécificités. Pourtant, la migration vers le cloud mérite d'être étudiée avec beaucoup d'attention pour permettre aux organisations qui migrent vers le cloud d'atteindre les objectifs escomptés. Ainsi, nous pouvons dire que la recherche en matière de BI dans le cloud est en progression. Il est important que d'autres artefacts soient implémentés pour faciliter cette migration.

Pour combler le vide constaté dans la littérature en notre possession, nous avons été conduits à élaborer le tableau 3 qui présente les opportunités de recherche en science de conception dans ce domaine de la BI dans le cloud. Aussi le tableau 3 propose-t-il de nouveaux artefacts pouvant apporter une valeur ajoutée dans le domaine de la BI dans le cloud. C'est ici que notre question de recherche, " Quels sont les nouveaux artefacts à implémenter pour que la BI tire profit du cloud computing ?" trouve son plein sens. En effet, la réponse à cette question donnera lieu à des artefacts inexistant dans la littérature et que nous présentons ci-dessous dans le tableau 3.

Pour élaborer ce tableau 3, nous avons, tout d'abord, recueilli, dans les articles de notre revue systématique de la littérature, les différentes contributions proposées par les chercheurs pour leurs futures recherches. Ce sont des indicateurs précieux, car ils sont généralement les prochaines étapes d'une recherche en cours. En ce sens, il est très probable qu'ils soient réalisables, qu'ils apparaissent bientôt et même qu'ils soient déjà apparus. Ces artefacts sont résumés dans le tableau 3 avec les références des articles qui les mentionnent. Les autres propositions émanent de notre réflexion.

Artéfacts	Collection et consolidation des données	Modélisation et stockage des données	Analytique
Modèle de conception			<ul style="list-style-type: none"> Modèle de processus d'analyse des données dans le cloud
Ontologie	<ul style="list-style-type: none"> Ontologie pour l'intégration des données locales et les données de cloud 		
Taxonomie	<ul style="list-style-type: none"> Taxonomie de la migration de la BI vers le cloud 		
Méthodologie	<ul style="list-style-type: none"> Méthodologie pour effectuer la collection et la consolidation des données dans le cloud 	<ul style="list-style-type: none"> Méthodologie pour modéliser et stocker les données dans le cloud 	<ul style="list-style-type: none"> Méthodologie pour développer des applications orientées clients permettant l'interaction des utilisateurs avec le cloud
Directive	<ul style="list-style-type: none"> Directives pour configurer convenablement les ressources du cloud 		
	<ul style="list-style-type: none"> Directives pour aider à collecter et consolider les données opérationnelles et les données de cloud 	<ul style="list-style-type: none"> Directives pour la modélisation et le stockage des données dans le cloud (Assunção <i>et al.</i> 2014) 	
Implémentation du système	<ul style="list-style-type: none"> Outils permettant d'accomplir les tâches de BI dans le cloud (Gatzioura <i>et al.</i> 2012) (Assunção <i>et al.</i> 2014) 		
Exemple	<ul style="list-style-type: none"> Exemples de migration de la BI vers le cloud 		

Tableau 3 : synthèse des opportunités de recherche

Ainsi, en ce qui concerne toutes les étapes de la BI dans le cloud, en termes de recherche future, la littérature propose la mise en œuvre d'outils pouvant permettre aux utilisateurs de bien accomplir les tâches de BI dans le cloud (Gatzioura *et al.* 2012) (Assunção *et al.* 2014). Quant à nous, nous proposons de définir une taxonomie de migration de la BI

vers le cloud avec des directives pour configurer correctement les ressources du cloud et servir de support dans le choix de migrer vers le cloud ainsi que des exemples concrets. En effet, ces lignes directrices sont d'un grand intérêt pour toutes les entreprises qui veulent migrer leur BI vers le cloud. Structurer ces directives grâce à une taxonomie permettra de fournir aux utilisateurs un processus de décision progressive.

La collecte et la consolidation des données sont des tâches fastidieuses de la BI. Afin de les faciliter dans le cloud, nous proposons aux chercheurs en science de conception de se concentrer sur la mise en œuvre des artefacts suivants :

- Méthodologies : pour permettre aux organisations de suivre pas à pas les étapes requises afin de collecter et consolider les données dans le cloud.
- Ontologies : pour faciliter l'intégration des données opérationnelles et de big data afin d'enrichir l'analyse. Les ontologies permettraient aux fournisseurs et aux consommateurs de données de parler un même langage. Elles faciliteraient aussi l'automatisation de l'intégration des données.
- Directives : Pour aider les organisations à respecter une ligne de conduite dans la collection et la consolidation des données dans le cloud.

Quant à la modélisation et au stockage des données dans le cloud, la littérature propose, comme future recherche, la mise en œuvre des directives pour la modélisation et le stockage des données dans le cloud (Assunção *et al.* 2014). La recherche en science de conception peut aussi conduire à la mise en œuvre des méthodologies pour stocker et modéliser des données dans le cloud.

En ce qui concerne les tâches analytiques dans le cloud, les chercheurs en science de conception peuvent également se concentrer sur la mise en œuvre des artefacts suivants:

- Modèle de processus (modèle de conception) qui pourrait contribuer à une meilleure compréhension des demandes de services initiées par les utilisateurs.
- Méthodologies pour permettre aux entreprises de développer leurs applications personnalisées ou pour ajouter des composants à leurs applications.

2.7. Conclusion

Visant une évaluation plus complète de la littérature liée à notre sujet de recherche, dans ce chapitre, nous avons décrit une revue systématique de la littérature sur la BI dans le cloud. Nous avons mis en place des critères explicites d'inclusion ou d'exclusion d'un article dans notre échantillon. Grâce à cela, nous avons constitué un ensemble de taille raisonnable de documents de haute qualité. Par ailleurs, nous avons effectué la recherche documentaire par des mots clés. Ce processus peut entraîner l'omission de certains documents importants d'une part, les faux négatifs. D'autre part, il peut inclure des documents qui ne traitent pas de notre sujet de recherche, les faux positifs. C'est une des limites de notre revue systématique de la littérature. Pour résoudre le problème des faux négatifs, nous avons inclus dans cet état de l'art certains articles que nous n'avons pas obtenus par notre recherche systématique, mais qui comportent des informations pertinentes relatives à notre sujet de recherche. Ces articles, nous les avons trouvés en parcourant de nouveau les sources de documents universitaires telles que ACM, IEEE, DBLP, ScienceDirect, EBSCO, ainsi que la librairie électronique de l' AIS (AISeL). En ce qui concerne les faux positifs, nous les avons exclus de notre échantillon d'articles lors de la lecture détaillée.

Cet état de l'art nous a permis d'exhiber certains futurs sujets de recherche en science de conception pour le domaine de la BI dans le cloud.

Basés sur le tableau des opportunités de recherche que la BI dans le cloud offre pour la recherche en science de conception, nous voulons définir des directives pour aider à la migration de la BI vers le cloud. Ainsi, nous proposons un modèle de guidage opérationnel que nous allons instancier par une taxonomie de la BI dans le cloud et des règles pour la migration de la BI vers le cloud. Nous aurons besoin de deux taxonomies : une taxonomie de contexte et une taxonomie pour décrire la décision. Les règles vont être implémentées dans un prototype et instanciées par un scénario illustratif.

Pour ce faire, nous présentons le modèle de taxonomie dont résultera le modèle de guidage opérationnel ainsi que les opérateurs qui permettent de développer une telle taxonomie. Ainsi, le chapitre 3 de ce travail porte sur le modèle de taxonomie, opérateurs et règles.

CHAPITRE 3 : MODELE DE TAXONOMIE, OPERATEURS ET REGLES

3.1. Introduction

Notre objectif est de mettre en place un support d'aide à la décision pour la migration de la BI vers le cloud. Ainsi, nous proposons un modèle de guidage opérationnel que nous allons instancier par une taxonomie de la BI dans le cloud et des règles pour la migration de la BI vers le cloud. Pour cela, il nous faut deux taxonomies complémentaires : celle qui regroupe les éléments qui peuvent influencer la décision et celle qui constitue l'objet même de la décision. Pour mettre en place ces taxonomies, nous nous sommes référés au modèle de taxonomie de Nickerson, Varshney, et Muntermann (2013) selon lequel une taxonomie comporte deux niveaux : dimension et caractéristique. Basés sur le fait que toutes les instances d'une dimension ne peuvent pas toujours avoir des caractéristiques homogènes, nous élargissons ce modèle en introduisant un nouveau concept : la catégorie. Cette dernière regroupe les instances d'une dimension qui ont des caractéristiques homogènes. Nous présentons, également, une typologie des opérateurs qui permettent de construire une telle taxonomie.

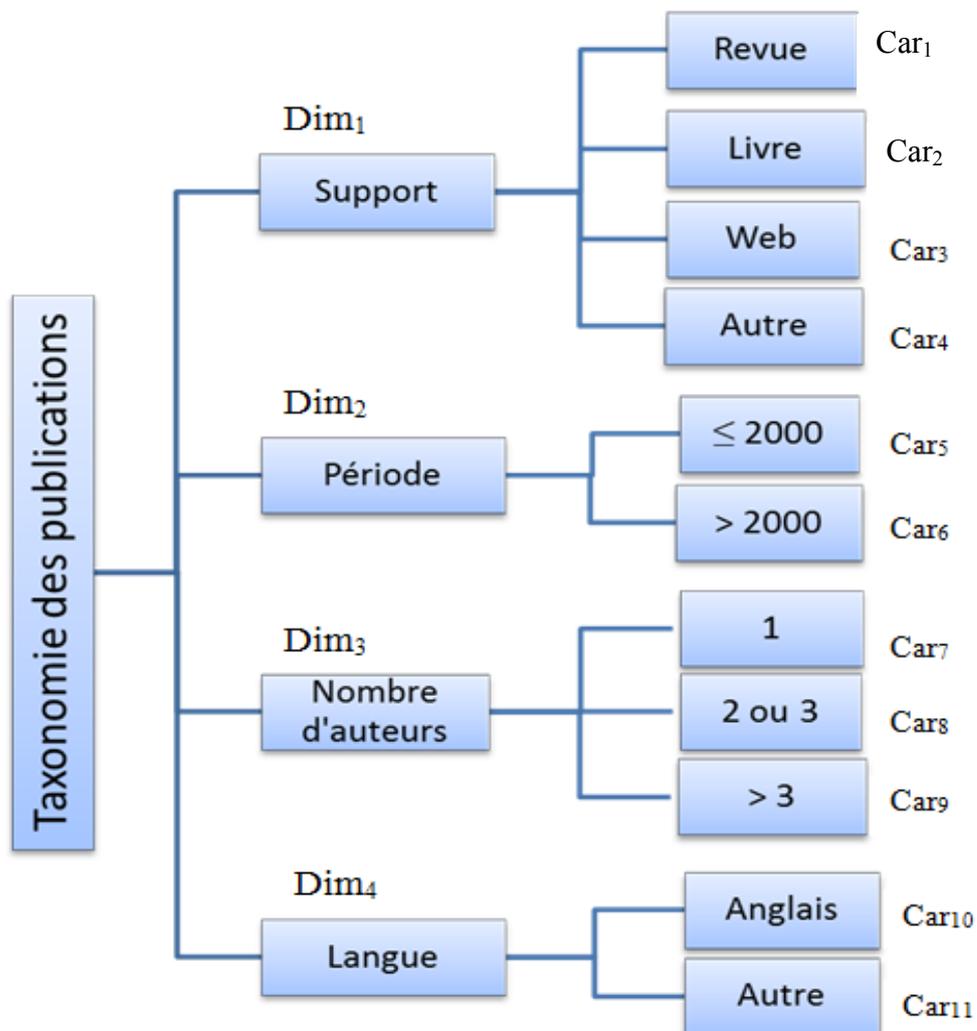
Les taxonomies sont essentielles pour la science en général et le domaine des systèmes d'information en particulier. En classant des objets ou des phénomènes, elles facilitent la compréhension et la prise de décision.

Le modèle de guidage opérationnel aidera à exploiter les taxonomies de contexte et de décision sous forme de règles à appliquer pendant le processus de prise de décision pour la migration de la BI vers le cloud. Nous utilisons le langage EBNF (Extended Backus-Naur Form) (ISO/IEC 14977 1996) pour décrire de manière systématique la syntaxe de ce modèle de guidage opérationnel. Pour vérifier la cohérence, la minimalité et l'exhaustivité des règles, nous mettons en place un ensemble de directives.

3.2. Modèle d'une taxonomie à dimensions hiérarchiques

Comme nous l'avons dit au point 3.1, selon Nickerson, Varshney et Muntermann (2013), une taxonomie est définie par un ensemble de dimensions (Dim). Chaque dimension est un ensemble d'au moins deux caractéristiques (Car), tel que, pour chaque objet, chaque dimension a une et une seule caractéristique.

Exemple : Taxonomie des publications



Pour cette publication : O. Sangupamba Mwilu, I. Comyn-Wattiau et N. Prat, *Future Generation Computer Systems* 63 (2016) 108–122,

- $\text{Dim}_1 = \text{Car}_1 = \text{Revue}$
- $\text{Dim}_2 = \text{Car}_6 = >2000$
- $\text{Dim}_3 = \text{Car}_8 = 2 \text{ ou } 3$
- $\text{Dim}_4 = \text{Car}_{10} = \text{Anglais}$

Dans cet exemple :

- Il n'y a pas de catégorie.
- Deux publications peuvent avoir exactement les mêmes caractéristiques.
- Deux caractéristiques d'une même dimension sont mutuellement exclusives.
- L'union des caractéristiques possibles forme une partition, toute publication possède l'une de ces caractéristiques.

Cette définition de Nickerson, Varshney et Muntermann (2013) est relative à une taxonomie avec dimensions simples.

Nous proposons une taxonomie avec des dimensions hiérarchiques. Nous utilisons le formalisme de diagramme de classe UML³ pour présenter ce modèle qui regroupe les concepts utilisés dans une taxonomie à dimensions hiérarchiques ainsi que les relations qui les lient.

³ Unified Modeling Language : <http://www.omg.org/spec/UML/>

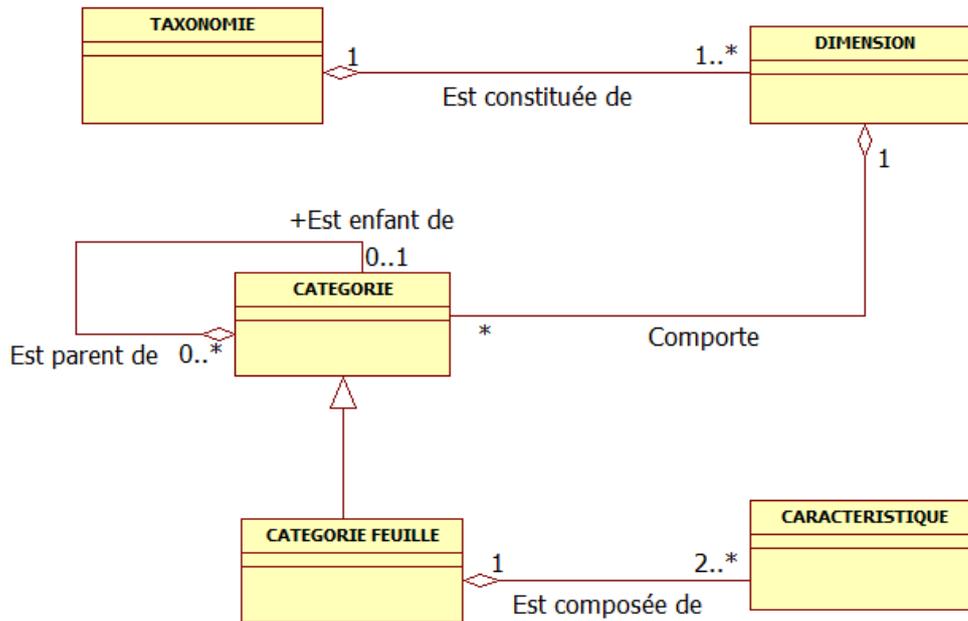


Figure 6 : Modèle de taxonomie à dimensions hiérarchiques

Ce modèle prend en compte nos besoins de représentation de la connaissance : il présente une taxonomie à dimensions hiérarchiques où les caractéristiques sont regroupées en catégories (Cat). Une catégorie peut avoir des sous-catégories : dans ce cas, elle est appelée racine. Les autres catégories sont des sous-ensembles de la racine. La catégorie la plus élevée comprend toutes les caractéristiques et la catégorie de plus bas niveau est appelée catégorie feuille.

Exemple : un extrait de la taxonomie des publications

Cette définition élargie la définition de Nickerson, Varshney et Muntermann (2013), elle illustre ce qui suit :

- Une dimension a au moins deux caractéristiques, éventuellement plus.
- Une dimension peut avoir zéro ou plusieurs catégories.
- Une catégorie a au moins deux caractéristiques.
- Une catégorie peut avoir zéro sous catégories ou plusieurs.
- Si deux caractéristiques relèvent de la même catégorie, elles sont mutuellement exclusives.
- On peut supposer que l'ensemble des caractéristiques d'une dimension ou d'une catégorie n'est pas complet au sens où certains objets ne possèderaient aucune de ces caractéristiques.

Par exemple, si la dimension Langue comprend deux caractéristiques Anglais et Français, toutes les publications en chinois ne possèdent aucune de ces caractéristiques.

3.4. Les opérateurs algébriques

Dans le développement d'une taxonomie, le concepteur a besoin d'opérateurs pouvant l'aider pour la prise en compte des éléments de la taxonomie et sa mise à jour. Dans cette section, nous présentons une typologie des opérateurs pour le développement d'une taxonomie. Ces opérateurs peuvent être combinés pour définir des opérations plus complexes. Chaque nouvel élément qui s'ajoute dans la taxonomie peut nécessiter l'utilisation d'un ou de plusieurs opérateurs. Ces opérateurs répondent au critère d'exhaustivité et de pertinence. Ils sont exhaustifs car ils suffisent pour développer une taxonomie à dimensions hiérarchiques. Ils sont pertinents car ils peuvent être utilisés tous pour construire une taxonomie à dimensions hiérarchiques. Nous l'illustrons au point 5.2.5. Ainsi, pour garantir leur stabilité, nous les formalisons en vue de mettre en place une algèbre y relative. Avant toute chose, nous définissons l'algèbre au sens général.

Une algèbre est un ensemble de n -uplets avec ses opérations (Chen 1984). Elle peut être définie formellement comme suit (Sankappanavar and Burris 1981) :

Pour A , un ensemble non vide et n , un nombre entier positif, on définit $A_0 = \{\emptyset\}$, et pour $n > 0$, A_n est l'ensemble de n -uplets des éléments de A . Une opération n -aire (ou fonction) sur A est une fonction f de A_n vers A , où n est le degré de f .

Un langage (ou type) d'algèbre est un ensemble F de symboles de fonction tel qu'un entier positif n est attribué à chaque membre f de F : Cet entier est appelé l'arité ou le degré (ou rang) de f ; et f est dit : symbole de fonction n -aire.

Si F est un langage d'algèbre alors une algèbre A du type F est une paire ordonnée $\langle T, F \rangle$ où T est un ensemble non vide et F est une famille d'opérations n -aires sur T indexées par le langage F .

Dans tout ce qui suit, Ω est l'ensemble des taxonomies possibles qu'on peut imaginer. Dans le cas d'une taxonomie à dimensions hiérarchiques, une algèbre A sur F est un ensemble ordonné de paires $\langle \Omega, F \rangle$. Tout élément T appartenant à Ω est constitué de trois sous-ensembles. Un entier positif n est attribué à chaque sous ensemble pour indiquer son rang : S_1 , S_2 et S_3 . S_1 est l'ensemble des caractéristiques de la taxonomie, S_2 est l'ensemble des catégories et S_3 est constitué des dimensions de la taxonomie. Enfin, F est l'ensemble des opérateurs qui permettent de créer ou de manipuler les éléments de la taxonomie, f est la variable pour désigner une opération. Formellement, nous pouvons définir ces deux ensembles T (taxonomie) et F (opérateurs) comme suit :

- $T = S_1 \cup S_2 \cup S_3$ et $S_i \cap S_j = \emptyset$ quels que soient i et j
- $F = \{f: \Omega_n \rightarrow \Omega \text{ pour tout } n\}$
(Ω_n est une taxonomie quelconque)

Tout élément x de T est de la forme $x = \langle \text{libellé}, \text{niveau}, \text{parent} \rangle$. x est donc un triplet ordonné où niveau correspond à S_1 , S_2 ou S_3 , et parent est un unique élément appartenant à T . Pour un libellé donné il ne peut y avoir qu'un seul x dans T ayant ce libellé.

Exemple : Taxonomie des publications (voir point 3.2)

- $\langle \text{Support}, S_3, T \rangle$
- $\langle \text{Revue}, S_2, \text{Support} \rangle$
- $\langle \text{Classique}, S_1, \text{Revue} \rangle$
- $\langle \text{En ligne}, S_1, \text{Revue} \rangle$
- $\langle \text{Livre}, S_2, \text{Support} \rangle$

<Monographie, S₁, livre>

<Collectif, S₁, livre>

Nous avons défini la taxonomie de deux manières. Ces deux définitions ne se contredisent pas. La première, énoncée au point 3.3, est une définition théorique où T est un ensemble de dimensions, elles-mêmes étant des ensembles : $T = \{Dim_i, i = 1, \dots, n \mid Dim_i = \{Cat_{ij}, j = 1 \dots, k_i \mid Cat_{i1} = \{Car_{im}, m = 1 \dots, p_i; p_i \geq 2\} \wedge \forall j \geq 2, Cat_{ij} \not\subseteq Cat_{i1}\}\}$ (Prat, Comyn-Wattiau and Akoka 2015). La deuxième définition : $T = S_1 \cup S_2 \cup S_3$, est une définition opérationnelle où S₁, S₂ et S₃ sont composés des triplets comme expliqué plus haut.

Notons que nous avons deux types d'opérateurs : les opérateurs de création de nouveaux éléments de la taxonomie et les opérateurs de transformation des éléments de la taxonomie.

3.4.1. L'opérateur de création de nouveaux éléments de la taxonomie

L'union

L'union est une opération de base en théorie des ensembles. Ici, elle va nous permettre de représenter l'enrichissement des différents ensembles constituant la taxonomie. Alors que les autres opérateurs sont utilisés pour organiser la taxonomie, l'union permet d'ajouter des éléments dans la taxonomie. Considérant qu'on a un ensemble des taxonomies, l'opérateur union opère sur une taxonomie ou plusieurs pour la transformer en une autre. Formellement, nous pouvons le noter comme suit :

$$\forall T, B \in \Omega \mid T \cap B = \emptyset, T \cup B = A \text{ où } A \in \Omega.$$

Exemple 1 : ajout d'une dimension dans une taxonomie

$$T_1 + \text{Dimension} = T_2$$

On peut considérer que c'est l'union de deux ensembles au sens où $T_2 = T_1 \cup \{\text{Dim.}\}$ et Dim est considérée comme une taxonomie ayant une seule dimension.

Etant donné que c'est l'union de deux ensembles, cet opérateur a toutes les propriétés de l'union ensembliste : commutativité, associativité, élément neutre (ensemble vide) et l'idempotence. L'union de deux taxonomies peut avoir un sens ou non, ici nous avons restreint cet opérateur à nos besoins.

Exemple 2 : ajout d'une catégorie dans une dimension est aussi une union puisque une dimension est un ensemble de catégories.

$$\text{Dim}_1 + \text{Catégorie} = \text{Dim}_2$$

$$\text{Dim}_1 \cup \{\text{Catégorie}\} = \text{Dim}_2$$

De la même façon, c'est une union au sens ensembliste, donc nous avons l'assurance de la commutativité, l'associativité, l'élément neutre (ensemble vide) et l'idempotence.

Exemple 3 : ajouter une caractéristique à une dimension ou à une catégorie relève aussi d'une union.

$$\text{Cat}_1 + \text{Caractéristique} = \text{Cat}_2$$

$$\text{Cat}_1 \cup \{\text{Caractéristique}\} = \text{Cat}_2$$

En conclusion, l'opérateur d'ajout qu'il s'agisse de dimension, de catégorie ou caractéristique est une union ensembliste.

Les propriétés de l'opérateur *union*

L'opérateur *union* requiert deux opérands, il est binaire. Il présente les propriétés suivantes (Frécon 2002) :

- **La commutativité** : L'opérateur binaire union est commutatif. L'ordre de deux ensembles à additionner n'est pas important. Soit l'ensemble T des éléments d'une taxonomie et l'ensemble B des éléments tirés d'une source nouvellement examinée. $(T \cup B) = (B \cup T)$.
- **L'associativité** : Le résultat de l'opérateur union de 3 ensembles ne dépend pas de l'ordre dans lequel les opérations sont faites. $((T \cup B) \cup C) = (T \cup (B \cup C))$. T étant les éléments existants dans la taxonomie, B et C, deux ensembles des éléments de deux sources examinées distinctement.
- **L'élément neutre** : L'ensemble vide est l'élément neutre pour l'opérateur union car $\{\emptyset\} \cup T = T$.
- **Idempotence** : L'opérateur union est idempotent car $(T \cup T) = T$.

3.4.2. Les opérateurs de transformation des éléments de la taxonomie

Soit T une taxonomie, rappelons que tout élément x de T est de la forme $x = \langle \text{libellé}, \text{niveau}, \text{parent} \rangle$. x est donc un triplet ordonné où niveau correspond à S_1, S_2 ou S_3 , et parent est un unique élément appartenant à T. Pour un libellé donné il ne peut y avoir qu'un seul x dans T ayant ce libellé. Dans la suite de ce texte, on définit pour tout $x \in T$ et $y \in T'$, $x \equiv y \Rightarrow \text{libellé}(x) = \text{libellé}(y)$.

3.4.2.1. La promotion

En utilisant l'opérateur promotion, un concepteur de la taxonomie peut transférer un concept de S_1 vers S_2 ou S_3 et un concept de S_2 vers S_2 ou S_3 . Une catégorie devient, par exemple, une dimension et une caractéristique devient une catégorie ou une dimension. Les dimensions (éléments de S_3) de la taxonomie ne sont pas concernées par cet opérateur. Nous pouvons représenter l'opération de promotion selon ces différents cas :

- Pour une caractéristique x qui devient une catégorie

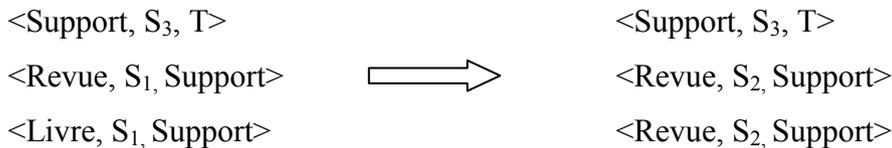
$$T = (T \setminus \{ \langle x, S_1, y \rangle \}) \cup \{ \langle x, S_2, y \rangle \}.$$
- Pour une caractéristique x qui devient une dimension

$$T = (T \setminus \{ \langle x, S_1, y \rangle \}) \cup \{ \langle x, S_3, T \rangle \}.$$
- Pour la catégorie x qui devient une dimension

$$T = (T \setminus \{ \langle x, S_2, y \rangle \}) \cup \{ \langle x, S_3, T \rangle \}.$$
- Pour la catégorie x qui est promue en catégorie de plus haut niveau, elle ne change que de parent

$$T = (T \setminus \{ \langle x, S_2, P_1 \rangle \}) \cup \{ \langle x, S_2, P_2 \rangle \}.$$

Exemple : On promeut Revue et Livre au rang de catégorie et on ajoute les caractéristiques classique, en ligne, monographie et collectif.



La promotion de Revue et Livre au rang de catégorie revient à modifier le deuxième élément du triplet.

3.4.2.2. La rétrogradation

Grâce à l'opérateur de rétrogradation, un concepteur de la taxonomie peut transférer un concept de S_3 à S_2 , de S_3 à S_1 , de S_2 à S_2 ou S_2 à S_1 : une dimension descend au rang de catégorie ou de caractéristique et une catégorie devient une sous-catégorie ou une caractéristique. Les éléments de S_1 ne sont pas concernés par cet opérateur. Formellement, nous pouvons écrire :

- Pour la catégorie x qui est dégradée en caractéristique

$$T = (T \setminus \{ \langle x, S_2, y \rangle \}) \cup \{ \langle x, S_1, y \rangle \}.$$

- Pour la catégorie qui devient une sous-catégorie

$$T = (T \setminus \{ \langle x, S_2, P_1 \rangle \}) \cup \{ \langle x, S_2, P_2 \rangle \}.$$

- Pour la dimension qui est dégradée en caractéristique

$$T = (T \setminus \{ \langle x, S_3, T \rangle \}) \cup \{ \langle x, S_1, y \rangle \}.$$

- Pour la dimension qui est dégradée en catégorie

$$T = (T \setminus \{ \langle x, S_3, T \rangle \}) \cup \{ \langle x, S_2, y \rangle \}.$$

Exemple : Revue et Livre sont dégradés en caractéristique. La rétrogradation est l'inverse de la promotion.

$$\begin{array}{ccc}
 \langle \text{Support}, S_3, T \rangle & & \langle \text{Support}, S_3, T \rangle \\
 \langle \text{Revue}, S_2, \text{Support} \rangle & \longrightarrow & \langle \text{Revue}, S_1, \text{Support} \rangle \\
 \langle \text{Livre}, S_2, \text{Support} \rangle & & \langle \text{Revue}, S_1, \text{Support} \rangle
 \end{array}$$

3.4.2.3. La scission

L'opérateur de scission permet au concepteur de la taxonomie de scinder un concept en deux parties. Les concepts scindés sont de même sous-ensemble (S_1, S_2 ou S_3). Chacun de ces concepts scindés, devenu un élément à part entière de son sous-ensemble, peut être manipulé à l'aide de n'importe quel autre opérateur. Tous les sous-ensembles de la taxonomie sont concernés par cet opérateur. Celui-ci admet un élément en entrée et il retourne deux éléments en sortie. Ainsi, une dimension scindée en deux donne lieu à deux dimensions distinctes. Une catégorie peut aussi être scindée en deux catégories indépendantes l'une de l'autre. Une caractéristique peut être scindée en deux différentes caractéristiques indépendantes. Formellement, nous pouvons écrire :

$$\forall T \in \Omega, T = S_1 \cup S_2 \cup S_3, \forall x \in T, \exists \{y, z\} \in \Omega \setminus \{T\} \mid \text{Scission}(T, \{x\}, \{y, z\}) = \{x' \in ((T \setminus \{x\}) \cup \{y, z\})\}. \text{ Autrement dit, la scission de } x \text{ donne } y \text{ et } z.$$

Exemple : La scission de la caractéristique ≥ 2000 :

$$\begin{array}{ccc}
 \langle \text{Période}, S_3, T \rangle & & \langle \text{Période}, S_3, T \rangle \\
 \langle \langle 2000, S_1, \text{Période} \rangle \rangle & \longrightarrow & \langle \langle 2000, S_1, \text{Période} \rangle \rangle \\
 \langle \langle \geq 2000, S_1, \text{Période} \rangle \rangle & & \langle \text{Comprise entre 2000 et 2010}, S_1, \text{Période} \rangle \\
 & & \langle \langle > 2010, S_1, \text{Période} \rangle \rangle
 \end{array}$$

3.4.2.4. La fusion

L'opérateur de fusion permet au concepteur de la taxonomie de fusionner deux éléments d'un même sous-ensemble de la taxonomie. Tous les sous-ensembles de la taxonomie sont concernés par cet opérateur. Il admet deux éléments en entrée et retourne un seul élément en sortie. On peut fusionner deux dimensions pour en avoir une seule en sortie. On peut aussi fusionner deux catégories ou deux caractéristiques pour obtenir en sortie respectivement une catégorie et une caractéristique. Formellement, nous pouvons écrire la fusion comme suit :

$$\forall T \in \Omega, T = S_1 \cup S_2 \cup S_3, \forall x \in T, \exists \{z\} \in \Omega \setminus \{T\} \mid \text{Fusion}(T, \{x,y\}, \{z\}) = \{x' \in ((T \setminus \{x,y\}) \cup \{z\})\}. \text{ Autrement dit la fusion de } x \text{ et } y \text{ donne } z.$$

Exemple : On fusionne les caractéristiques Français et Arabe.

$\langle \text{Langue}, S_3, T \rangle$		$\langle \text{Langue}, S_3, T \rangle$
$\langle \text{Anglais}, S_1, \text{Langue} \rangle$	\Longrightarrow	$\langle \text{Anglais}, S_1, \text{Langue} \rangle$
$\langle \text{Français}, S_1, \text{Langue} \rangle$		$\langle \text{Autre langue}, S_1, \text{Langue} \rangle$
$\langle \text{Arabe}, S_1, \text{Langue} \rangle$		

Notons que toutes les opérations sur la taxonomie sont initiées par l'opérateur union : c'est quand il y a insertion d'un nouvel élément dans la taxonomie que celle-ci peut nécessiter une restructuration de manière à permettre au nouvel élément de se placer au bon endroit. La restructuration se fait à l'aide des autres opérateurs : promotion, rétrogradation, fusion et scission. Par conséquent, le processus de développement de la taxonomie peut s'arrêter quand on a examiné tous les objets de l'échantillon et qu'on n'a plus de nouveaux éléments à ajouter dans la taxonomie. On peut alors parler de convergence car il n'y a plus d'informations utiles à ajouter dans la taxonomie : l'opérateur union n'est plus utilisé.

Par ailleurs, tous les opérateurs de manipulation de la taxonomie opèrent sur des ensembles et se ramènent à des opérations d'union et de différence. Comme nous l'avons affirmé au point 3.4.1, l'opérateur union renferme toutes les bonnes propriétés de commutativité, associativité, élément neutre et idempotence. En ce qui concerne la différence, elle n'est pas associative, ni commutative, ni idempotente. Elle admet un élément neutre : \emptyset (l'ensemble vide).

Ces opérateurs nous ont suffi pour procéder à la constitution de nos taxonomies comme nous l'illustrons dans la suite.

3.5. Du modèle de taxonomie au modèle de guidage opérationnel

L'objectif de notre taxonomie à dimensions hiérarchiques est de mettre en place un modèle de guidage opérationnel pour un processus de prise de décision.

Il existe deux formes de guidage, suggestif et informatif. Le guidage informatif fournit des informations pertinentes qui permettent de faciliter le jugement du décideur sans suggérer comment agir. Le guidage suggestif fournit aux décideurs des recommandations judicieuses (quoi faire, quelles valeurs d'entrée utiliser) pour une prise de décision (Silver 1991).

Dans ce travail, nous proposons un modèle de guidage opérationnel suggestif. Ce guidage est opérationnel car il est instancié par des règles qui guident les utilisateurs dans leur processus de prise de décision.

La littérature nous a montré qu'il y a des conseils sur la migration qui dépendent de nombreux éléments de contexte. D'autre part, la migration elle-même se décompose en différentes dimensions. C'est pourquoi nous proposons de construire deux taxonomies complémentaires car une décision est généralement conditionnée par un certain nombre d'éléments. Ceux-ci ne peuvent donc pas être au même niveau que les options de la décision.

Ainsi, les deux taxonomies complémentaires sont :

- La taxonomie des éléments qui peuvent influencer la prise de décision. Elle permet aux organisations de tenir compte de leur situation afin de prendre des décisions qui correspondent à leurs besoins, à leurs moyens et à leurs situations. Cette taxonomie est appelée "taxonomie de contexte".
- La taxonomie des options de la décision. Il s'agit de l'objet même de la décision. Cette taxonomie est nommée "taxonomie de décision".

Eu égard à ce qui vient d'être énoncé, nous pouvons dire que le contexte est la prémisse d'une décision : la décision dépend du contexte.

En vue d'atteindre notre objectif de mettre en place un modèle de guidage opérationnel, du principe ainsi conçu doivent découler des règles de production. Ces dernières instancient le modèle de guidage opérationnel. Elles seront généralement de la forme :

Si "contexte"

Alors "décision"

(A quel service cloud on peut recourir dans un contexte précis).

On peut aussi avoir des règles de la forme :

Si "contexte"

Alors "contexte"

(Un élément de contexte peut avoir de l'incidence sur un autre).

Ou

Si "décision"

Alors "décision"

(Une décision peut influencer la prise d'une autre décision).

Le modèle de guidage opérationnel sert de support pour l'exploitation de ces deux taxonomies complémentaires sous forme de règles. Il est composé de deux parties que nous avons séparées par des pointillés pour des raisons de lisibilité.

La partie supérieure décrit les concepts utilisés dans la taxonomie à dimensions hiérarchiques.

La deuxième partie du modèle de guidage opérationnel est le modèle des règles, il est constitué de tous les concepts relatifs aux règles. Cette partie décrit chaque composant de la règle : prémisse et conclusion.

Pour le contexte, chaque élément qui peut influencer une décision correspond à une expression. De même pour la décision, chacune de ses options correspond à une expression. Ainsi, chacune des deux parties d'une règle (contexte ou décision) peut être constituée de plusieurs expressions. Une expression est constituée des éléments de la taxonomie de contexte ou des éléments de la taxonomie de décision. Elle possède deux parties : gauche et droite.

Les deux parties du modèle de guidage opérationnel sont liées à l'aide de trois associations binaires "se réfère à" et de deux agrégations "fait partie de" :

- *partie gauche "se réfère à" dimension,*
- *partie droite abstraite "se réfère à" catégorie,*

- *partie droite concrète "se réfère à" caractéristique,*
- *expression décision "fait partie de" taxonomie de décision,*
- *expression contexte "fait partie de" taxonomie de contexte.*

La partie gauche d'une expression correspond à une dimension de la taxonomie. La partie droite peut être abstraite ou concrète. Elle est abstraite quand elle correspond à une catégorie. Elle est concrète quand elle correspond à une caractéristique ; la catégorie de cette caractéristique doit aussi être renseignée.

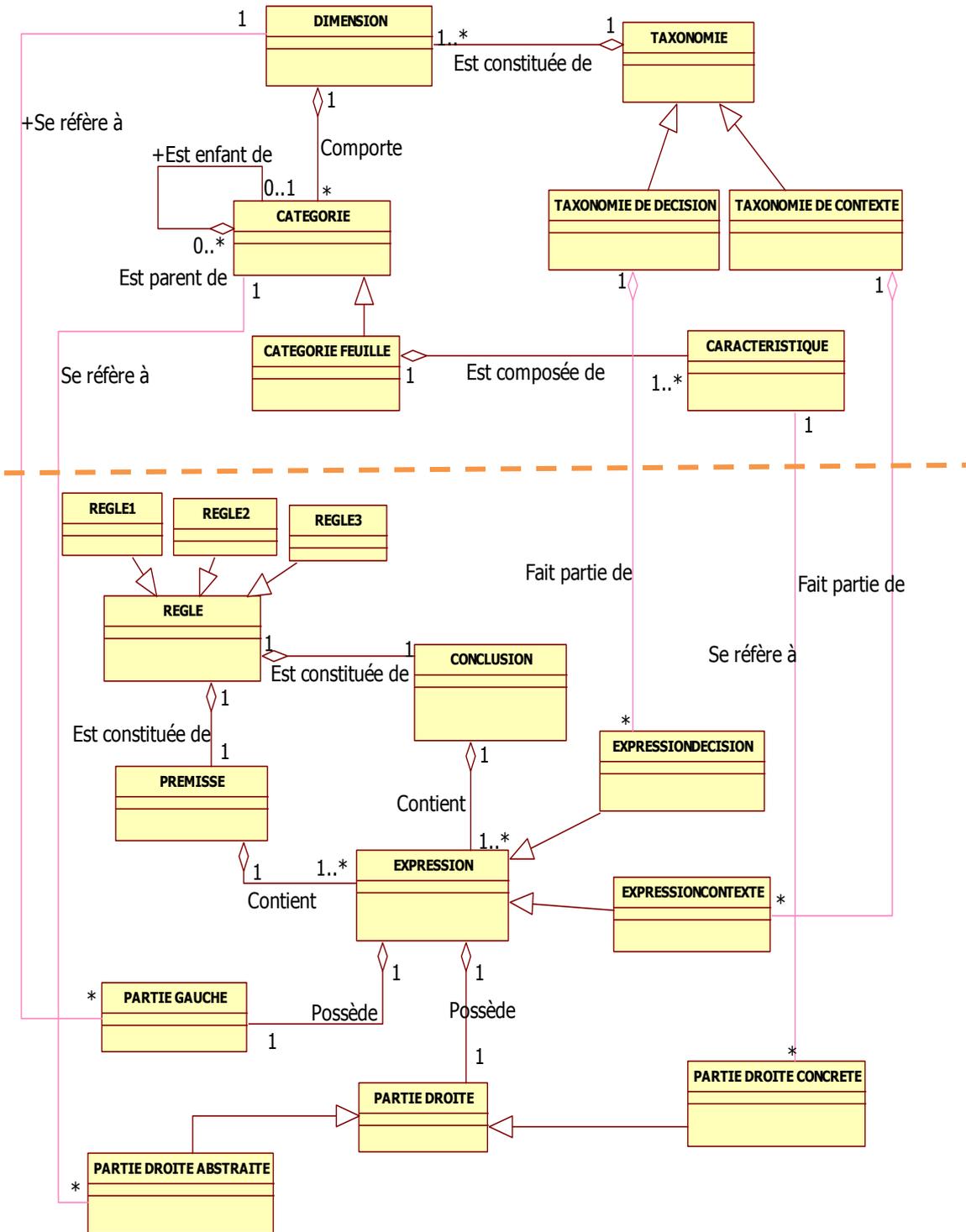


Figure 7 : Modèle de guidage opérationnel

Exemple de règle : Cet exemple est tiré de la base de règles pour la migration de la BI vers le cloud que nous avons mise en place. Les deux taxonomies complémentaires de la BI dans le cloud forment le domaine de ces règles. Dans la taxonomie de contexte, on trouve, entre autres, la dimension *objectifs* (de la migration) et la dimension *taille de l'entreprise*. La dimension *objectifs* possède une catégorie *réduction des coûts* qui comporte les caractéristiques suivantes : *réduction des coûts d'infrastructure*, *réduction des coûts d'implémentation* et *réduction des coûts d'administration*. Cette dimension possède en plus, les caractéristiques : performance des services, mobilité, agilité, manipulation de grandes quantités de données. La dimension *taille de l'entreprise* possède trois caractéristiques : *petite*, *moyenne* et *grande*.

Dans la taxonomie de décision, on trouve entre autres la dimension *niveau de migration*. Cette dimension possède une caractéristique *total* et une catégorie *partiel*. La catégorie *partiel* possède deux sous-catégories : *matériel* et *outils*. Chacune de ces sous catégories possède ses caractéristiques (voir les points 5.3 et 6.2.1).

Si *objectifs* = *réduction des coûts*

et *Taille de l'entreprise* = *petite ou moyenne*

Alors *niveau de migration* = *total*

- Cette règle est composée d'une prémisse et d'une conclusion
 - ❖ Prémisse : **Si** *objectifs* = *réduction des coûts*
et Taille de l'entreprise = *petite ou moyenne*
 - ❖ Conclusion : **Alors** *niveau de migration* = *total*
- La prémisse est constituée de deux expressions :
 - ❖ Première expression : *objectifs* = *réduction des coûts* (expression contexte)
 - ❖ Deuxième expression : *taille de l'entreprise* = *petite ou moyenne* (expression contexte)
- Chacune de deux expressions est constituée de deux parties : gauche et droite
 - ❖ Partie gauche de la première expression : *objectifs*
 - ❖ Partie droite de la première expression : *réduction des coûts*
 - ❖ Partie gauche de la deuxième expression : *taille de l'entreprise*
 - ❖ Partie droite de la deuxième expression : *petite ou moyenne*

- Les deux parties gauches : *objectifs* et *taille de l'entreprise* se réfèrent aux dimensions de la taxonomie de contexte
- La partie droite de la première expression se réfère à une catégorie de la taxonomie de contexte
- La partie droite de la deuxième expression se réfère à une caractéristique de la taxonomie de contexte
- La conclusion est constituée d'une expression : *niveau de migration = total* (expression décision)
- L'expression de la conclusion est constituée de deux parties : gauche et droite
 - ❖ Partie gauche = *niveau de migration*
 - ❖ Partie droite = *total*
- La partie gauche : *niveau de migration* se réfère à une dimension de la taxonomie de décision
- La partie droite : *total* se réfère à une caractéristique de la taxonomie de décision

Afin de compléter et de préciser le modèle de guidage opérationnel, nous allons le décrire par sa syntaxe EBNF.

3.6. Syntaxe EBNF de modèle de guidage opérationnel

EBNF est une notation pour décrire formellement comment écrire des entités dans un langage. Une description EBNF est une liste non ordonnée de règles. Chaque règle comporte trois parties : une partie gauche, une partie droite et les caractères "::<=" séparant les deux parties. La partie gauche est un identificateur de symbole non terminal et la partie droite est une description EBNF d'une construction autorisée faisant intervenir des symboles terminaux et non terminaux ainsi que les opérateurs décrits par le langage EBNF (Dethier 2011). Ces opérateurs sont les suivants :

Supposons que x et y sont des descriptions EBNF de toutes les constructions possibles,

- $()$: délimite une expression EBNF à laquelle un opérateur peut s'appliquer.
- x^* : c'est la répétition d'une construction représentée par x un nombre n de fois avec $n \geq 0$.
- x^+ : c'est la répétition d'une construction représentée par x un nombre n de fois avec $n \geq 1$.

- $x \mid y$: représente soit une construction décrite par x , soit une construction décrite par y .
- $[x]$: la construction décrite par x est optionnelle.

Nous utilisons ici la notation EBNF pour décrire la syntaxe de notre modèle de guidage opérationnel.

Règle ::= Règle1 | Règle2 | Règle3

Règle1 ::= Si prémissescontexte Alors conclusioncontexte (source)

Règle2 ::= Si prémissescontexte Alors conclusiondécision (source)

Règle3 ::= Si prémissedécision Alors conclusiondécision (source)

Prémissecontexte ::= Expression [Et Expression]*

Prémisedécision ::= Expression [Et Expression]*

Expression ::= Expressioncontexte | Expressiondécision

Expressioncontexte ::= Dimensioncontexte \equiv Partiedroitecontexte

Expressiondécision ::= Dimensiondécision \equiv Partiedroitedécision

Partiedroitedécision ::= Catégoriedécision | Caractéristiquedécision

Partiedroitecontexte ::= Catégoriecontexte | Caractéristiquecontexte

Dimensioncontexte ::= Dim₁ | Dim₂ | ... | Dim_n

Dimensiondécision ::= Dim₁ | Dim₂ | ... | Dim_n

Catégoriecontexte ::= Cat₁ | Cat₂ | ... | Cat_n

Catégoriedécision ::= Cat₁ | Cat₂ | ... | Cat_n

Caractéristiquecontexte ::= Car₁ | Car₂ | ... | Car_n

Caractéristiquedécision ::= Car₁ | Car₂ | ... | Car_n

Conclusioncontexte ::= Expressioncontexte [Et Expressioncontexte]*

Conclusiondécision ::= Expressiondécision [Et Expressiondécision]*

Source ::= Référence bibliographique

3.7. Cohérence, minimalité et exhaustivité des règles de guidage opérationnel

Dans cette section, nous mettons en place des directives permettant de vérifier les critères de cohérence, de minimalité et d'exhaustivité des règles qui vont découler du modèle

de guidage opérationnel. Cette vérification se fera au niveau des règles, au niveau des prémisses et des conclusions de chaque règle, au niveau des expressions de chaque prémisses et de chaque conclusion ainsi qu'au niveau des parties de chaque expression.

3.7.1. Cohérence

En termes de cohérence, une règle doit respecter la syntaxe du modèle de guidage opérationnel et elle ne peut pas contredire plusieurs règles existantes. Ainsi,

- Une règle ne peut pas avoir une prémisses issue de la taxonomie de décision et une conclusion issue de la taxonomie de contexte.
- Dans la conclusion d'une règle, les connecteurs "ou" et "non" ne sont pas employés.
- Les deux parties d'une expression (partie droite et partie gauche) doivent être issues de la même taxonomie (de contexte ou de décision).
- Une partie gauche d'une expression ne peut pas se référer à une catégorie.
- La partie droite d'une expression doit se référer à la caractéristique ou la catégorie de la dimension à laquelle se réfère sa partie gauche.
- Deux règles ne peuvent pas avoir une même prémisses et des conclusions contradictoires.

3.7.2. Minimalité

En ce qui concerne la minimalité, une base de règles ne peut pas contenir des règles redondantes. Pour préserver la minimalité, les directives suivantes doivent être appliquées pour la base de règles :

- Si deux règles ont la même conclusion et si leurs prémisses, bien qu'elles soient différentes, sont toutes deux issues de la taxonomie de contexte, ou de la taxonomie de décision, on les fusionne. La prémisses de la nouvelle règle comprend les deux prémisses séparées par la conjonction "ou" et on garde la conclusion car elle est commune aux deux prémisses.
- Si la conclusion d'une règle qui est issue de la taxonomie de décision, se trouve être prémisses d'une autre, par souci de minimalité, on fusionne ces deux règles. La nouvelle règle issue de la fusion aura la prémisses de la première règle et sa conclusion sera constituée de la concaténation des conclusions de deux règles fusionnées.

- Pour deux règles avec une prémisse identique et issue de la taxonomie de contexte, si leurs conclusions sont différentes et sont issues l'une de la taxonomie de contexte et l'autre de la taxonomie de décision, on fusionne ces deux règles. La nouvelle règle aura comme prémisse toutes les expressions de ces deux règles qui sont issues de la taxonomie de contexte et la conclusion toutes les expressions de ces deux règles qui sont issues de la taxonomie de décision.
- Pour deux expressions qui ont des parties gauches identiques et dont les parties droites sont l'une concrète et l'autre abstraite, on supprime l'expression qui a la partie droite concrète et on garde celle qui a la partie droite abstraite.

3.7.3. Exhaustivité

Pour qu'une base de règles soit dite exhaustive, elle doit être capable de fournir une réponse à chaque question.

Ainsi, dans le cas d'une base qui ne comprend pas beaucoup de règles, il est facile de vérifier l'exhaustivité en faisant toutes les combinaisons possibles entre les différents contextes et les différentes décisions. Dans le cas d'une grande base de règles, la vérification de l'exhaustivité est manuellement quasi impossible.

Notons qu'une base de règles peut être exhaustive à un moment donné. Cependant, on peut intégrer de nouvelles règles en tenant compte des éventuelles évolutions du domaine pour lequel une taxonomie est développée.

3.8. Conclusion

Notre objectif étant de mettre en place un support d'aide à la décision, nous avons posé le principe de développer deux taxonomies complémentaires : de contexte et de décision. Pour atteindre cet objectif, nous avons mis en place un modèle de guidage opérationnel qui aidera à exploiter ces taxonomies sous forme de règles. Ce modèle de guidage opérationnel est composé de deux parties :

- Un modèle de taxonomie à dimensions hiérarchiques englobant tous les concepts relatifs à ce genre de taxonomie ainsi que les relations qui les lient.

- Un modèle de règles qui englobe tous les concepts relatifs aux règles ainsi que les relations qui les lient.

En outre, nous avons présenté les opérateurs que le concepteur peut utiliser pour construire et organiser une taxonomie. Nous avons défini les propriétés principales de ces opérateurs, lesquelles garantissent la cohérence de l'approche.

Par ailleurs, nous avons utilisé la syntaxe EBNF pour compléter et préciser le modèle de guidage opérationnel.

Notons que le processus de mise en place d'un modèle de guidage opérationnel pour l'aide à la décision de migration de la BI vers le cloud proposé dans ce chapitre peut s'appliquer à d'autres semblables cas d'aide à la décision.

Rappelons que, pour répondre à notre objectif de mettre en place une aide à la décision de migration de la BI vers le cloud, nous avons besoin de deux taxonomies : de contexte et de décision. Ces deux taxonomies seront développées au chapitre 5 en utilisant la méthodologie de développement de taxonomies des technologies complexes émergentes que nous allons décrire au chapitre 4 et dont la BI dans le cloud est un cas particulier.

CHAPITRE 4 : METHODOLOGIE DE DEVELOPPEMENT DE TAXONOMIES POUR LES TECHNOLOGIES COMPLEXES EMERGENTES

4.1. Introduction

En vue de répondre à notre question de recherche, nous nous sommes proposés de mettre en place une taxonomie de la BI dans le cloud. Etant donné qu'il existe plusieurs méthodologies de développement des taxonomies en système d'information, nous nous sommes inspirés de la méthodologie de Nickerson, Varshney et Muntermann (2013). Celle-ci est une méthodologie récente qui capitalise sur les précédentes.

Par ailleurs, la BI dans le cloud est une technologie complexe émergente. Pour développer une taxonomie relative à ce domaine, il est important de tenir compte de deux défis spécifiques que posent ces genres de technologies. En effet, ces technologies sont souvent à l'intersection de plusieurs domaines et le corpus de connaissances à leur sujet se constitue encore. Face à cela, la méthodologie de développement des taxonomies de Nickerson, Varshney et Muntermann montre des limites.

Ainsi, notre tâche dans le présent chapitre est de proposer une méthodologie appropriée pour le développement de taxonomies pour les technologies complexes émergentes. Pour ce faire, nous allons décrire la méthodologie de Nickerson, Varshney et Muntermann (2013) et en énoncer les limites afin d'en esquisser une version améliorée.

4.2. Méthodologie de développement de taxonomies de Nickerson, Varshney et Muntermann

La méthodologie de Nickerson, Varshney et Muntermann a été publiée en 2013. Elle rassemble, en les enrichissant, les approches définies auparavant. Si elle peut être appliquée à plusieurs domaines, elle montre ses limites lorsqu'il s'agit de développer des taxonomies pour

les technologies complexes émergentes. Nous nous basons sur ces limites pour adapter cette méthodologie.

4.2.1. Présentation de la méthodologie

Au cours des années, les chercheurs en SI ont proposé un certain nombre de taxonomies. Cependant, dans de nombreux cas, ces taxonomies ont été développées avec une approche ad hoc. Par ailleurs, dans le domaine de SI, il y a peu d'articles qui traitent de processus de développement des taxonomies (Nickerson, Varshney and Muntermann 2013).

Pour remédier à cette situation, Nickerson, Varshney et Muntermann proposent une méthodologie de développement de taxonomies qui capitalise sur les précédentes. L'originalité de cette méthodologie réside dans le fait que l'on peut utiliser une approche empirique ou conceptuelle pour développer une taxonomie ou encore combiner les deux approches.

Les étapes de cette méthodologie sont illustrées dans le diagramme ci-dessous.

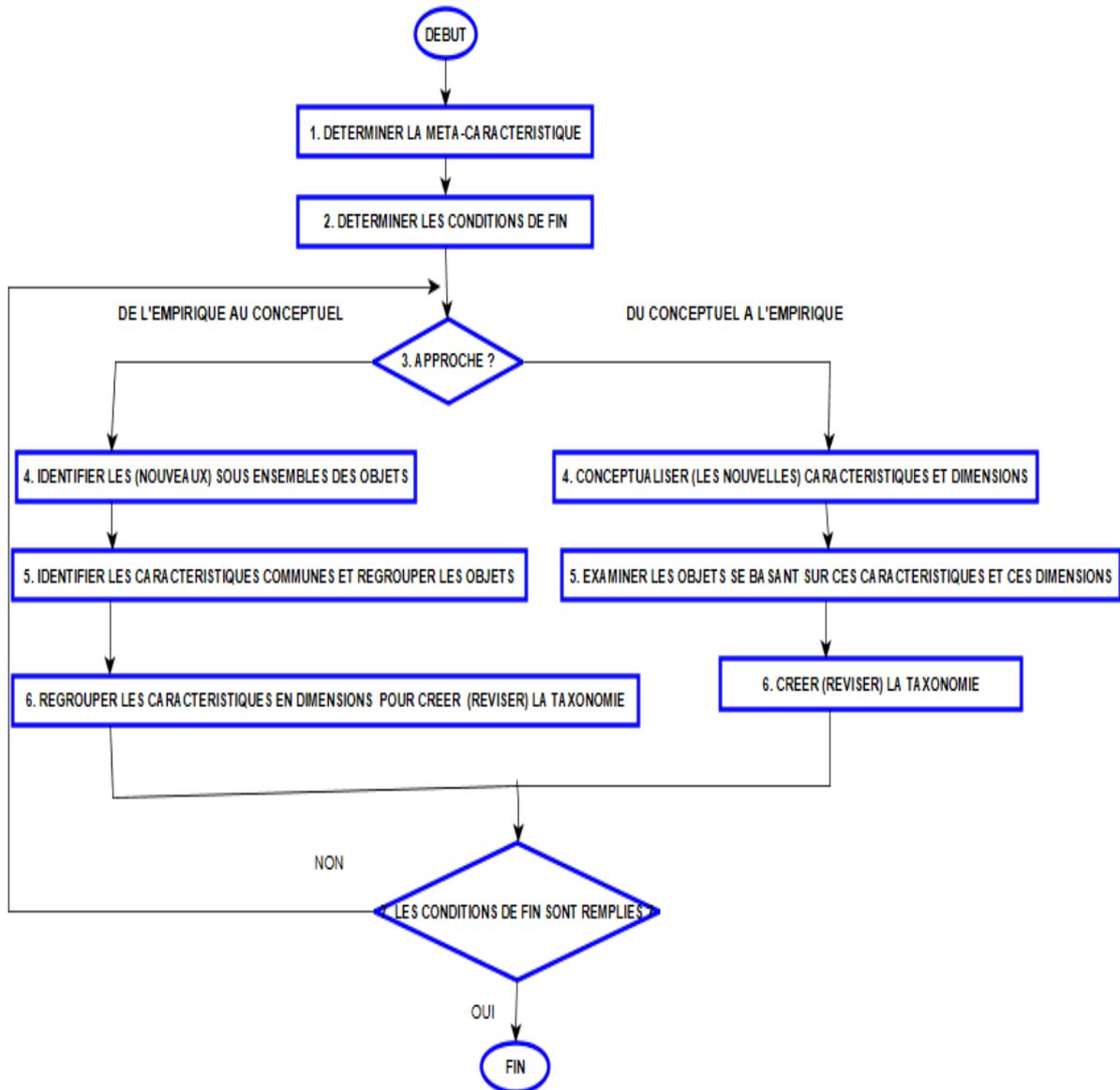


Figure 8: Méthodologie de développement de taxonomies de Nickerson, Varshney et Muntermann (2013)

Nous résumons les étapes de cette méthodologie comme suit :

- Déterminer une méta-caractéristique,
- Déterminer les conditions de fin,
- Choisir une approche et l'appliquer,
- Arrêter le processus (les conditions de fin sont-elles remplies ?).

4.2.1.1. Déterminer une méta-caractéristique

Le développement d'une taxonomie consiste à déterminer les caractéristiques des objets d'intérêt. Or, le choix de ces caractéristiques est un problème central.

Afin de fournir une base à l'identification des caractéristiques d'une taxonomie, une méta-caractéristique doit être spécifiée au début du processus de développement de la taxonomie. En effet, la méta-caractéristique est la caractéristique la plus complète qui servira de base pour le choix des caractéristiques dans la taxonomie. Chaque caractéristique d'une taxonomie devrait être une conséquence logique de cette méta-caractéristique.

Une méta-caractéristique constitue le but pour lequel une taxonomie est mise en place. Ce but doit être basé sur l'utilisation prévue de la taxonomie et pourrait donc être défini par les utilisateurs éventuels de la taxonomie. Le choix de la méta-caractéristique doit être fait avec soin car il influe de manière critique sur la taxonomie qui en résulte. Par exemple, pour une taxonomie des applications mobiles, la méta-caractéristique peut être "interaction de haut niveau entre l'utilisateur de l'application et l'application" (Nickerson, Varshney and Muntermann 2013). Ainsi, toutes les caractéristiques de cette taxonomie vont être relatives à une interaction de haut niveau entre une application mobile et son utilisateur.

4.2.1.2. Déterminer les conditions de fin

Etant donné que cette méthodologie est itérative, des conditions de fin sont nécessaires pour aider le concepteur de la taxonomie à décider de la fin du processus de développement de la taxonomie. Selon Nickerson, Varshney et Muntermann (2013), la première et principale condition de fin demeure que la taxonomie ait répondu à sa définition : qu'elle se compose de dimensions dont les caractéristiques sont mutuellement exclusives et collectivement exhaustives. Les auteurs proposent aussi des conditions de fin subjectives et objectives. En effet, dans ce contexte, les conditions de fin subjectives sont celles qui se rapportent à la taxonomie déjà élaborée, elles se vérifient après toutes les itérations. Une itération consiste à examiner un objet en vue d'identifier les éléments relatifs à la taxonomie, tandis que les conditions de fin objectives se rapportent à l'objet de la taxonomie et se vérifient après chaque itération. Ainsi, les conditions de fin sont subjectives lorsque le développement d'une taxonomie s'arrête parce qu'elle devient concise, robuste, exhaustive, extensible et

explicative. Les conditions de fin objectives sont les suivantes (Nickerson, Varshney and Muntermann 2013) :

1. Tous les objets ou un échantillon représentatif d'objets ont été examinés.
2. Lors de la dernière itération, aucun objet n'a été fusionné avec un autre ni scindé en plusieurs.
3. Au moins un objet a été classé dans chacune des caractéristiques de chaque dimension.
4. Lors de la dernière itération, aucune nouvelle dimension ou caractéristique n'a été ajoutée.
5. Lors de la dernière itération, aucune dimension ou caractéristique n'a été fusionnée ni scindée.
6. Chaque dimension est unique (pas de duplication de dimension).
7. Chaque caractéristique est unique dans sa dimension (pas de duplication de caractéristique dans une dimension).
8. Chaque regroupement de caractéristiques est unique.

4.2.1.3. Choisir une approche et l'appliquer

Cette méthodologie permet de développer une taxonomie tant par une approche conceptuelle que par une approche empirique. Le choix de l'approche à utiliser dépend de la disponibilité des données sur les objets à étudier et de la connaissance que le chercheur possède sur le domaine d'intérêt. L'approche conceptuelle est déductive. Elle commence par identifier, de manière théorique, des dimensions et des caractéristiques à partir desquelles elle détermine les objets. L'application de l'approche empirique est inductive. Elle identifie les dimensions et les caractéristiques à partir des objets concrets.

4.2.1.4. Arrêter le processus

A la fin de chaque itération, le chercheur vérifie si les conditions de fin objectives sont remplies par la version actuelle de la taxonomie. A la première itération, il est probable qu'aucune des conditions objectives ne soit remplie. Par conséquent, le processus doit recommencer. Dans les itérations ultérieures, on évalue les conditions objectives. Si elles ne sont pas remplies, le processus est répété. Dans le cas où elles sont remplies, les conditions subjectives vont être, à leur tour, examinées.

4.2.2. Difficultés d'application aux technologies complexes émergentes

Les conditions de fin objectives précitées fournissent des orientations utiles pour décider de la fin du processus de développement de taxonomie. Cependant, elles sont difficiles à appliquer en général et en particulier dans le cas de développement des taxonomies pour les technologies complexes émergentes. Plus précisément :

- Dans la condition 1, les auteurs proposent que le processus de développement de la taxonomie puisse s'arrêter quand tous les objets ou un échantillon représentatif d'objets ont été examinés. Sur quelle base peut-on décider que tous les objets ou un échantillon représentatif d'objets ont été examinés ? Il est important de commencer le processus de développement d'une taxonomie par la constitution d'un échantillon représentatif d'objets à examiner.
- Les conditions 2, 4 et 5 énoncent qu'au cours de la dernière itération, aucune modification ne doit être apportée dans la taxonomie. Doit-on arrêter le processus de développement de la taxonomie si l'examen d'un objet n'apporte pas de modifications dans la taxonomie ? Non, car le processus de développement de taxonomie est itératif et tous les objets de l'échantillon représentatif doivent être examinés avant d'arrêter le processus. En d'autres termes, si aucun changement n'a été apporté dans la taxonomie au cours d'une itération, cela ne signifie pas nécessairement qu'une nouvelle itération n'apporterait pas non plus de changement dans la taxonomie. On ne peut pas arrêter le processus de développement d'une taxonomie parce qu'une itération n'a pas apporté de modification sur la taxonomie. Il est plutôt important de déterminer le moment où le nombre d'opérations effectuées sur la taxonomie se stabilise et se rapproche de zéro.
- Si aucun objet n'est classé sous certaines caractéristiques (si la condition 3 n'est pas remplie), doit-on obligatoirement trouver un objet à classer sous ces caractéristiques ou les supprimer ? Probablement pas pour tous les cas, en particulier dans le cas des technologies complexes émergentes, où le nombre d'objets à examiner est limité par définition car le corps conceptuel des connaissances à leur sujet se constitue encore : on peut donc imaginer qu'il y aura ultérieurement des objets non encore existants.

Notons que seule la dernière difficulté est liée particulièrement au développement des taxonomies pour les technologies complexes émergentes.

4.3. Adaptation de la méthodologie

Afin de pouvoir être appliquée au développement de taxonomies pour les technologies complexes émergentes, la méthodologie de Nickerson, Varshney et Muntermann nécessite des adaptations dont nous présentons une ébauche dans les lignes suivantes. Notre proposition est de construire une taxonomie des technologies émergentes en nous fondant sur la littérature scientifique adéquate. A cet effet, pour constituer un échantillon représentatif des articles à examiner, nous identifions trois types de sources d'information à partir desquelles on peut mener une revue systématique de littérature pour construire la taxonomie. Nous mettons en place une typologie des opérateurs qui nous aide à compter les opérations effectuées sur la taxonomie pour constater la fin du processus de développement. Tout élément de la taxonomie l'enrichit. Nous suggérons donc de ne pas utiliser l'opérateur de suppression dans le développement des taxonomies en général et en particulier pour le développement de taxonomies pour les technologies complexes émergentes.

En ce qui concerne les conditions de fin subjectives, nous vérifions particulièrement la robustesse en examinant les sources dans un ordre précis qui sera expliqué et justifié plus loin et l'exhaustivité par une diversité des types de sources à examiner.

4.4. Présentation de notre méthodologie

Notre méthodologie est basée principalement sur une revue systématique de la littérature (Kitchenham 2004) car nous avons constaté qu'il y a des similitudes entre les objectifs d'une taxonomie et les objectifs d'une revue systématique de la littérature.

En effet, une taxonomie permet de fournir une structure et une organisation de la connaissance d'un domaine. Elle aide également à comprendre les différents résultats de recherche (Nickerson, Varshney and Muntermann 2013). Une revue systématique de la littérature, quant à elle, permet de résumer les documents existants d'un domaine ou d'une technologie. Elle identifie les lacunes dans la recherche actuelle sur un sujet afin de suggérer des voies de solutions pour les recherches futures. Elle fournit un cadre pour classer de manière appropriée les activités de recherche (Kitchenham 2004).

Ainsi, élaborer une taxonomie ou effectuer une revue systématique de la littérature nécessite ce qui suit :

- Constituer un échantillon des documents existants sur un domaine ou une technologie car, comme pour une revue systématique de la littérature, la taxonomie est constituée à partir des résultats de recherche antérieurs.
- Examiner ces documents pour faire ressortir les éléments relatifs à la taxonomie ou des données relatives à la revue systématique de littérature.
- Résumer les connaissances sous forme de taxonomie ou sous forme d'une synthèse pour la revue systématique de littérature.

La revue systématique de littérature peut enrichir le processus de développement de taxonomie tel que proposé par Nickerson, Varshney et Muntermann. Elle permet de constituer l'échantillon des documents à examiner en vue d'identifier les informations relatives à la taxonomie. Grâce aux recommandations de la revue systématique de littérature, la taxonomie répond aux critères de fin subjectifs tels que l'exhaustivité et la robustesse. En effet, notre méthodologie permet de remédier aux trois difficultés relevées au point 4.2.2 :

- Elle permet de constituer un échantillon représentatif d'objets à examiner avant de commencer le processus de développement de la taxonomie grâce aux principes de la revue systématique de la littérature qui assure la représentativité.
- Elle permet de déterminer le moment où l'on peut arrêter le processus de développement de la taxonomie grâce à la typologie des sources et l'utilisation des opérateurs.
- Elle n'utilise pas l'opérateur de suppression dans le développement de la taxonomie. Les caractéristiques sous lesquelles il n'y a pas d'objets ne sont pas supprimées en vue de garder l'évolutivité.

Ainsi, nous allons rappeler les étapes de la revue systématique de littérature présentées au point 2.3.5 en vue d'illustrer comment une revue systématique de littérature nous permet d'adapter la méthodologie de développement de taxonomie proposée par Nickerson, Varshney et Muntermann.

La conduite d'une revue systématique de la littérature doit comprendre cinq étapes (Kitchenham 2004) :

- l'identification de la recherche,
- la sélection des études,
- l'évaluation de la qualité des études,
- l'extraction des données,
- la synthèse des données.

Pour la construction d'une taxonomie des technologies complexes émergentes, nous allons nous concentrer sur les trois étapes : l'identification de la recherche, la sélection des études et l'extraction des données (examen des sources) car ces trois étapes nous permettent de constituer un échantillon d'objets représentatifs à examiner.

Le diagramme suivant présente les étapes de la méthodologie que nous proposons.

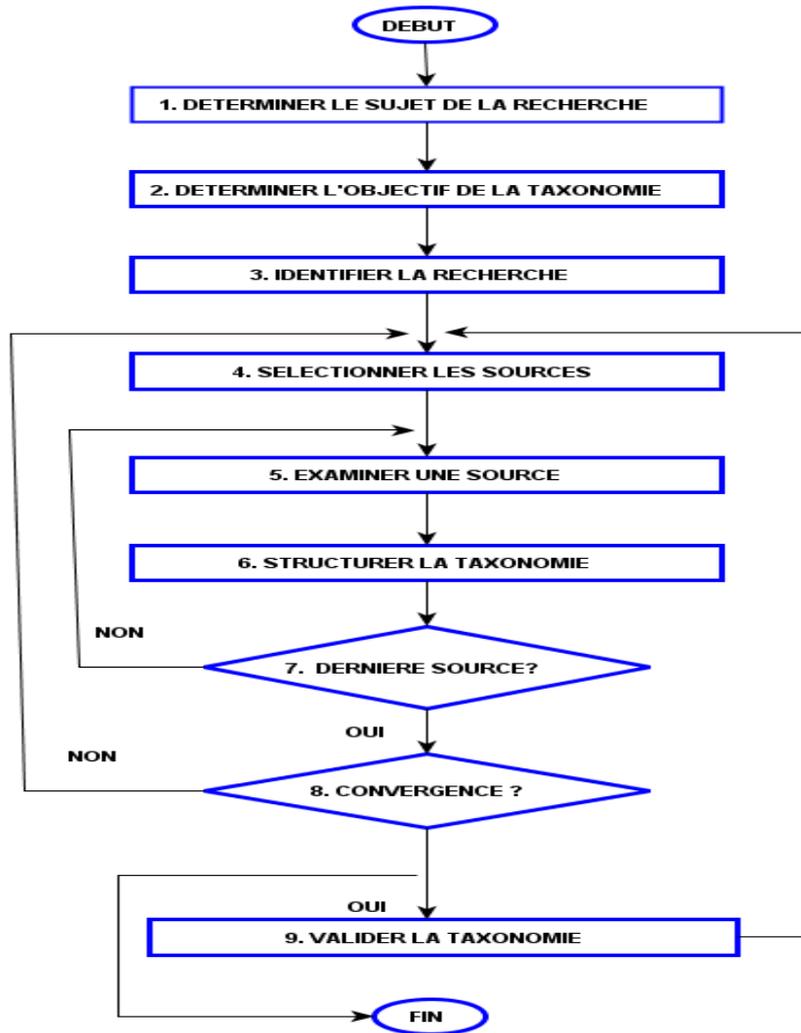


Figure 9: Notre méthodologie de développement de taxonomies

La figure 9 montre les étapes de la méthodologie de développement de taxonomies que nous proposons et qui sont explicitées dans la suite.

4.4.1. Déterminer le sujet de la recherche

Déterminer le sujet de la recherche consiste à situer la taxonomie dans le contexte d'un domaine bien déterminé. Un problème fondamental dans de nombreuses disciplines est la classification des connaissances d'un domaine en une taxonomie. Or, les taxonomies fournissent une structure et une organisation de la connaissance d'un domaine permettant ainsi aux chercheurs d'étudier les relations entre les concepts et, par conséquent, d'émettre des

hypothèses sur ces relations. Grâce aux taxonomies, les chercheurs et les praticiens peuvent comprendre et analyser les connaissances de certains domaines complexes (Nickerson, Varshney and Muntermann 2013).

4.4.2. Déterminer l'objectif de la taxonomie

Avant d'entamer le processus de développement d'une taxonomie, le concepteur détermine l'objectif pour lequel il décide de construire cette taxonomie : Il doit déterminer à quoi servira la taxonomie à développer. Cet objectif guidera le concepteur dans l'identification des éléments nécessaires à la construction de la taxonomie lors de l'examen des sources. Tous les éléments de la taxonomie doivent être relatifs à l'objectif prédéfini.

Ainsi, l'objectif de développement de la taxonomie sert de méta-caractéristique de la taxonomie. Chaque caractéristique des dimensions de la taxonomie doit être une conséquence logique de cette méta-caractéristique (Nickerson, Varshney and Muntermann 2013).

4.4.3. Identifier de la recherche

C'est dans cette étape que toute la stratégie pour constituer un échantillon représentatif des études qui serviront au développement des taxonomies des technologies complexes émergentes est déterminée.

Les technologies complexes émergentes sont souvent situées à l'intersection de plusieurs domaines. Afin de constituer une taxonomie comme cadre d'accueil d'un premier corpus de connaissances de ces technologies complexes émergentes, nous recommandons que la revue systématique de la littérature ne se concentre pas seulement sur les articles relatifs aux technologies complexes émergentes, mais aussi sur chacun des domaines constitutifs de la technologie complexe émergente étudiée.

Par exemple pour la BI dans le cloud, l'échantillon sera constitué par les articles relatifs à la BI dans le cloud mais aussi les articles sur la BI et le cloud computing pris séparément.

En vue de compléter les études menées par les auteurs de publications académiques, nous avons pensé qu'il serait important d'étudier ce qui se passe en pratique grâce à ces nouvelles technologies. Comment les professionnels utilisent-ils ces technologies complexes

émergentes ? Quelles sont leurs recommandations ? Les articles académiques ne peuvent pas traiter intégralement les problèmes pratiques liés au choix et à la mise en œuvre des technologies complexes émergentes. Il existe de nombreux exemples de domaines où, avant toute appropriation par le monde académique, des concepts et des solutions ont été proposés par les entreprises, voire par des individus. L'expérience empirique basée sur ces technologies complexes émergentes peut servir de fondement, à la condition qu'elle soit validée par des experts professionnels. Ainsi, pour le développement d'une taxonomie des technologies complexes émergentes, nous avons inclus également des articles professionnels. Pour chaque type de sources, une requête spécifique sera mise en place.

Nous avons ordonné les sources comme suit :

1. Les articles académiques sur les technologies complexes émergentes (sources de type 1) ;
2. Les articles académiques sur un des domaines des technologies complexes émergentes (sources de type 2) ;
3. Les articles professionnels (sources de type 3).

Cet ordre est important car les principaux concepts se trouvent dans la littérature académique sur les technologies complexes émergentes. Ces concepts aident à structurer les premiers niveaux de la taxonomie. Si ces sources de type 1 ne permettent pas de structurer la taxonomie, les sources de type 2, les articles concernant les domaines connexes aux technologies complexes émergentes peuvent suppléer. Enfin, les articles professionnels publient notamment les expériences basées sur des technologies complexes émergentes.

Par rapport aux conditions de fin subjectives proposées par Nickerson, Varshney et Muntermann (2013), l'examen des sources dans l'ordre croissant des types proposés ci-dessus aide à construire une taxonomie robuste car dans les sources de type 1, certains résultats structurants sont déjà observés et, si ces résultats ne sont pas suffisants, les sources de type 2 aident à les compléter. Le fait d'inclure dans l'échantillon les sources de type 2 et de type 3 permet à la taxonomie de mieux répondre au critère d'exhaustivité.

4.4.4. Sélectionner des études

Elle permet de constituer un échantillon relatif à la question de recherche grâce à la stratégie déterminée à la première étape.

Avec l'aide d'un ensemble de mots clés, nous avons interrogé les bases de données électroniques de ces sources pour recueillir des documents. Tous les résultats ne sont pas toujours pertinents pour la question de recherche. Les principes de la revue systématique de la littérature comportent la définition de critères pour décider d'inclure un article dans l'échantillon (Okoli and Schabram 2010) (Kitchenham 2004).

En vue de constituer un échantillon qui servirait au développement de taxonomies de technologies complexes émergentes, nous avons appliqué les critères suivants :

- **Contenu** : le sujet traité dans la source (article de revue, étude de cas) doit être directement lié à la technologie complexe émergente, ou à l'un de ses domaines constitutifs. Ceci est assuré par le choix des mots-clés utilisés dans la recherche des sources et vérifié ensuite par la lecture du titre, du résumé ou de tout l'article.
- **Langue** de publication : la recherche est limitée aux publications en anglais. Les publications en français ne remplissent pas les conditions de visibilité dans des journaux ou actes de conférences internationales. Ce n'est donc pas un choix direct mais une conséquence du critère de qualité décrit ci-après.
- **Revues** : pour assurer la sélection de sources de haute qualité, nous avons sélectionné les publications académiques de la liste de revues et des actes des conférences de "ERA" (ERA 2012). C'est une garantie de qualité car ERA publie les listes des revues ainsi que des actes des conférences qui ont été jugés de qualité. Cependant, cela peut être insuffisant pour les technologies complexes émergentes. Le manque potentiel peut être compensé par d'autres sources (sources de type 3 comme mentionné ci-dessus).
- **Date** de publication : Etant donné que les technologies complexes émergentes peuvent rapidement devenir obsolètes, nous proposons de limiter la recherche de publications universitaires aux dix dernières années. Pour les publications professionnelles (sources de type 3), un laps de temps encore plus court nous semble judicieux, nous proposons une année. En effet les pratiques professionnelles évoluent très rapidement et c'est tout l'intérêt d'ailleurs de cette presse

professionnelle qui s'attache à partager l'expérience croissante des consultants et experts dans ces nouvelles technologies.

Outre les critères précités, nous proposons une procédure de sélection des sources pour chaque type :

- **Les sources académiques (type 1 et type 2) :** La procédure de sélection des sources des types 1 et 2 est décrite au point 2.4.5.2.
- **Les articles professionnels (type 3) :** Pour les sources de type 3 (les articles professionnels), le processus est spécifique. Nous allons utiliser des sources d'informations spécifiques en adaptant le processus de sélection appliqué pour les sources de type 1 et type 2.

Les articles professionnels sont contenus dans la presse informatique professionnelle (en anglais "magazines"). Pour déterminer les supports les plus fiables qui traitent des technologies complexes émergentes en matière de technologie de l'information, nous avons procédé à une recherche sur Google avec des mots clés spécifiques: " Best IT professional magazines".

Ensuite, pour chaque magazine retenu, nous avons effectué une recherche sur Google avec l'URL du site web du magazine comme nom de domaine pour recueillir les articles relatifs à notre sujet de recherche. Le même ensemble de mots clés que pour les sources de type 1 a été utilisé, mais combiné avec un mot-clé spécifique pour limiter la recherche à des exemples concrets ou des études de cas ("example" ou "case"). Afin de limiter la portée de la recherche, pour chaque combinaison de mots-clés et chaque magazine, seule la première page de résultats a été retenue. Tous les résultats de la recherche ont ensuite été intégrés et vérifiés pour nous assurer que le contenu peut contribuer au développement de la taxonomie. Les articles qui sont très courts et ne contiennent pas d'informations relatives à l'objet de la recherche ont été éliminés. Nous avons de plus limité la recherche aux articles de magazines informatiques ayant moins d'un an.

Etant donné qu'il est particulièrement difficile de déterminer a priori le nombre de sources de type 3 nécessaires pour la construction de la taxonomie, nous avons commencé par un test pilote. Ensuite nous avons étendu la recherche sur la base des résultats de ce test pilote en ne gardant que des magazines informatiques dont les articles contribuaient à la modification de la taxonomie dans le test pilote.

4.4.5. Examiner les sources

Dans notre contexte, une source est un article. Une itération est l'examen d'une source. Examiner une source revient à la parcourir pour en tirer d'éventuelles informations relatives à la taxonomie. Les sources sont examinées selon l'ordre croissant de leurs types respectifs et en fonction de leur ordre décroissant de citations dans Google Scholar pour les sources de type 1 et 2. En ce qui concerne les sources de type 3, elles sont examinées en fonction de leur rang sur la page de résultats tel que retourné par Google. Cela permet de commencer par les sources considérées comme les principales références dans le domaine. Deux sources de même rang sont considérées toutes deux comme pertinentes, et seront examinées dans l'ordre alphabétique du nom de la source.

4.4.6. Structurer la taxonomie

Les informations sont prises en compte et organisées dans la taxonomie grâce aux opérateurs présentés au point 3.4.

Chaque itération conduit à l'un des deux cas suivants :

- Soit l'article ne modifie pas la taxonomie : La source étudiée n'ajoute rien à la taxonomie ou elle confirme un élément qui existe déjà dans la taxonomie. Dans ce cas, on n'a pas besoin d'un quelconque opérateur.
- Soit l'article remet en cause la taxonomie : La source étudiée ajoute un nouvel élément et/ou elle entraîne la modification d'un élément existant par des opérateurs appropriés, à savoir : union, promotion, rétrogradation, fusion et scission.

4.4.7. Vérifier si la source examinée est la dernière de l'échantillon

Toutes les sources de l'échantillon doivent être examinées avant de vérifier s'il y a convergence. Après chaque itération, on restructure la taxonomie en cas de besoin et on vérifie si la source examinée n'est pas la dernière dans l'échantillon. Dans le cas où elle n'est pas la dernière, on passe à l'examen d'une autre source. Si elle est la dernière source, on passe alors à la vérification de la convergence de la taxonomie.

4.4.8. Vérifier la convergence

Les conditions de fin sont nécessaires pour aider le constructeur de la taxonomie à décider de la fin du processus. Contrairement à la méthodologie de développement de taxonomies de Nickerson, Varshney et Muntermann, toutes les sources de l'échantillon doivent être examinées avant d'arrêter le processus du développement de la taxonomie même si certaines itérations intermédiaires n'apportent pas de modifications dans la taxonomie.

Grâce à la typologie des sources et la typologie des opérateurs présentée au point 3.4, le concepteur de la taxonomie peut alors décider de la fin du processus de développement de la taxonomie. Cette fin intervient quand il a examiné toutes les sources de l'échantillon et qu'il n'y a plus de sources qui puissent remettre en cause la taxonomie. On peut alors dire que le degré de convergence est atteint et qu'il n'y a plus de concepts utiles à ajouter dans la taxonomie. En revanche, après avoir examiné toutes les sources d'un échantillon, si la convergence n'est pas toujours atteinte, on peut être amené à constituer un nouvel échantillon de sources de type 1, de type 2 ou de type 3.

4.4.9. Valider la taxonomie

Dans la méthodologie d'échantillonnage, nous avons, dans un premier temps, limité la portée aux articles relatifs à la recherche en science de conception (Hevner *et al.* 2004). En vue de valider la taxonomie issue de cet échantillon des études relatives en science de conception et d'élargir les connaissances dans ce domaine de la BI dans le cloud, nous avons constitué un deuxième échantillon des sources de type 1 liées à la technologie émergente mais qui utilisent les méthodes de recherche quantitative ou qualitative. Notre méthodologie (figure 9) illustre que l'on peut constituer un autre échantillon pour valider une taxonomie en changeant certains critères de sélections des sources. La procédure de validation reprend le processus de développement de la taxonomie à partir de la sélection des sources. Un accent particulier est mis sur l'étape de la sélection des sources car il faut changer certains critères de sélection; et sur l'étape de l'examen des sources pour vérifier si l'examen des sources de ce nouvel échantillon valide effectivement la taxonomie qui a été construite au préalable.

4.5. Conclusion

Pour pouvoir mettre en place une taxonomie de la BI dans le cloud, et plus généralement des technologies complexes émergentes, nous avons adapté la méthodologie de développement des taxonomies proposée par Nickerson, Varshney et Muntermann. En effet, cette méthodologie a montré des limites face au développement de taxonomies pour les technologies complexes émergentes telles que la BI dans le cloud.

Afin de remédier à la première difficulté d'application de la méthodologie de Nickerson, Varshney and Muntermann liée à la constitution d'un échantillon représentatif d'objets à examiner, nous avons proposé une typologie des sources d'information. En nous inspirant des recommandations liées aux revues systématiques de la littérature, nous avons fourni des lignes directrices pour la sélection des sources.

L'ordonnement des sources et l'observation du processus d'exécution des opérateurs de développement des taxonomies ont servi de base pour décider quand le processus de développement de la taxonomie peut être considéré comme convergent. En effet, les sources sont examinées selon leur ordre croissant de type. On peut parler de la fin du processus de développement de la taxonomie quand on finit d'examiner toutes les sources de type 3. Par ailleurs, le processus est considéré comme convergent quand il n'y a plus beaucoup d'informations à ajouter et que le nombre d'opérations à effectuer s'approche de zéro. Ainsi, nous avons résolu la deuxième difficulté d'application de la méthodologie de Nickerson, Varshney et Muntermann relatif à l'arrêt de processus de développement d'une taxonomie.

La troisième difficulté d'application de la méthodologie de Nickerson, Varshney et Muntermann liée aux caractéristiques sous lesquelles aucun n'objet n'est classé, concerne particulièrement les technologies complexes émergentes. Pour la résoudre, nous avons proposé de ne pas utiliser l'opérateur de suppression dans le développement de taxonomies des technologies complexes émergentes. Ainsi, on ne supprime pas les caractéristiques sous lesquelles il n'y a pas d'objet en vue de garder l'évolutivité.

Par ailleurs, nous avons défini notre stratégie de revue systématique de la littérature en ciblant deux objectifs : l'exhaustivité et la robustesse. L'exhaustivité à un moment donné est assurée par la richesse et la variété des sources. La pérennité de cette exhaustivité étant

relative, il y a nécessité d'effectuer régulièrement de mise à jour bien que cela soit un processus fastidieux.

Nous avons examiné les sources en fonction de leur réputation académique. Ainsi, les articles professionnels sont pris en compte lors de la dernière étape. Pour des raisons de robustesse, nous avons proposé de construire d'abord la taxonomie avec les sources de type 1 où certains résultats structurants sont déjà observés et, si ces résultats ne sont pas suffisants, les sources de type 2 aident à les compléter. Cette stratégie permet de rencontrer les conditions de fin subjectives proposées par Nickerson, Varshney et Muntermann.

Dans le chapitre 5, nous utilisons cette méthodologie pour mettre en place notre taxonomie de la BI dans le cloud en calquant le modèle de taxonomie présenté au point 3.2. Ceci nous permettra de plus de démontrer la faisabilité et l'utilité de la méthodologie.

CHAPITRE 5 : DEVELOPPEMENT D'UNE TAXONOMIE DE LA BUSINESS INTELLIGENCE DANS LE CLOUD

5.1. Introduction

Dans ce chapitre, nous utilisons la méthodologie de développement de taxonomies pour les technologies complexes émergentes décrite précédemment, afin d'élaborer une taxonomie de la BI dans le cloud.

L'objectif de cette taxonomie est de soutenir la décision de migration de la BI vers le cloud. Cette décision peut être prise pour une organisation ou une division d'une organisation. Comme nous l'avons spécifié au point 3.5, une taxonomie destinée à soutenir une décision est composée de deux parties complémentaires : une taxonomie de contexte et une taxonomie de décision.

Ainsi, nous avons appliqué la méthodologie pour élaborer la taxonomie des éléments qui peuvent influencer la migration de la BI vers le cloud d'une part. D'autre part, nous avons élaboré la taxonomie des options relatives aux services dont la BI peut bénéficier dans le cloud.

5.2. Application de la méthodologie proposée

Le chapitre précédent fournit un ensemble d'étapes que nous allons suivre pas à pas dans cette section pour mettre en place une taxonomie de la BI dans le cloud.

5.2.1. Sujet de la recherche

Cette taxonomie s'inscrit dans le domaine de la BI dans le cloud. Elle permet de classer et de structurer les connaissances nécessaires en ce qui concerne la migration de la BI vers le cloud. Elle classe les éléments qui peuvent influencer la migration de la BI vers le cloud d'une part. D'autre part, elle classe les options des services dont la BI peut

bénéficier dans le cloud. Ainsi, elle constitue un point de départ pour d'autres chercheurs qui désirent prolonger la recherche dans ce domaine de la BI dans le cloud.

5.2.2. Objectif de la taxonomie

L'objectif de cette taxonomie est de mettre à la disposition des utilisateurs une aide à la décision de la migration de la BI vers le cloud. Par conséquent, la migration de la BI vers le cloud est la méta-caractéristique de la taxonomie que nous allons développer. Ainsi, les éléments constitutifs de la taxonomie doivent être cohérents avec la migration de la BI vers le cloud. La taxonomie de contexte comprendra les éléments qui peuvent influencer cette migration et la taxonomie de décision sera constituée des options possibles en matière de services de cloud dont la BI peut bénéficier dans le cas de la migration.

5.2.3. Identification de la recherche

La méthodologie prévoit de constituer l'échantillon des sources à examiner en vue du développement de la taxonomie grâce aux trois étapes de la conduite d'une revue systématique de la littérature : l'identification de la recherche, la sélection des sources et l'examen des sources.

La typologie des sources se présente comme suit :

- Les sources de type 1 : les articles académiques qui traitent de la BI dans le cloud.
- Les sources de type 2 : les articles académiques qui traitent de BI ou de cloud computing pris séparément.
- Les sources de type 3 : Les articles professionnels qui traitent de la BI dans le cloud.

5.2.4. Sélection des sources

Notre échantillon contient déjà les sources de type 1 et 2 qui nous ont servi pour la revue systématique de la littérature au chapitre 3 (voir le point 2.4.5.2). Il nous reste à constituer la bibliographie des sources de type 3 : les articles professionnels. Pour cela, nous allons déterminer d'abord les magazines (la presse informatique professionnelle) qui publient des articles relatifs à la BI dans le cloud.

La requête : "Top magazines among IT professionals" lancé sur Google, nous a retourné une liste de magazines professionnels classés selon leur ordre décroissant de nombre d'abonnés. De cette liste, nous avons retenu les cinq premiers magazines suivants :

- InformationWeek,
- ComputerWorld,
- CIO,
- ZDNet,
- PCWorld.

Pour constituer notre échantillon d'articles professionnels, nous avons utilisé les mots clés suivants :

("Cloud" OR "SaaS") AND ("Example" OR "Case") AND ("Business Intelligence" OR "Data warehouse" OR "Analytics").

Vu la complexité de la requête, elle ne pouvait pas être exécutée en une seule fois dans Google Scholar. Elle a donc été scindée en les quatre requêtes ci-dessous :

- "Cloud" AND "Example" AND ("Business Intelligence" OR "Data warehouse" OR "Analytics")
- "Cloud" AND "Case" AND ("Business Intelligence" OR "Data warehouse" OR "Analytics")
- "SaaS" AND "Example" AND ("Business Intelligence" OR "Data warehouse" OR "Analytics")
- "SaaS" AND "Case" AND ("Business Intelligence" OR "Data warehouse" OR "Analytics").

Ces quatre requêtes ont retourné quarante articles pour chaque magazine, ce qui a donné un total de 200 articles pour les cinq magazines. Après intégration (déduplication), nous avons retenu 108 articles. Par la lecture, nous avons vérifié que chacun de ces articles traite de la BI dans le cloud. Nous avons supprimé les articles très courts qui n'avaient pas d'informations pertinentes. Après ce processus, nous avons retenu 34 articles relatifs à la BI dans le cloud dont :

- 12 issus d'InformationWeek,
- 8 de ComputerWord,

- 5 de CIO,
- 6 de ZDNet,
- 3 de PcWord.

5.2.5. Examen des sources

Lors de la construction d'une taxonomie, les sources d'information sont examinées l'une après l'autre. Chacune d'elle peut déclencher zéro, une ou plusieurs opérations sur la taxonomie.

Pour chaque article examiné, nous aurons au moins un des cas suivants :

- Cas a : l'article n'apporte rien de pertinent sur le sujet.
- Cas b : l'article confirme une information déjà obtenue.
- Cas c : l'article permet d'ajouter une dimension, une catégorie ou encore une caractéristique.
- Cas d : l'article modifie la taxonomie et va engendrer la fusion, la scission, la promotion ou la rétrogradation d'un élément de la taxonomie.

Les deux premiers cas n'entraînent aucune modification de la taxonomie. Pour les deux derniers cas, nous utilisons les opérateurs appropriés soit pour prendre en compte un nouvel élément, soit pour manipuler les anciens éléments de la taxonomie.

Pour chaque type de sources, nous allons construire deux taxonomies complémentaires : une taxonomie de contexte et une taxonomie de décision conformément à ce qui a été dit au point 3.5.

En revanche, nous présentons ici seulement l'examen des sources de type 1. Le résultat de l'examen des sources de type 2 et 3 est en annexe de cette thèse (ANNEXE 2).

Rappelons que les recherches dans les sources de type 1 ont apporté quarante-trois articles relatifs à la BI dans le cloud. Dans cette section, nous allons examiner chacun de ces articles pour en tirer les informations pertinentes relatives à cette taxonomie.

5.2.5.1. *Taxonomie de contexte*

Chaque élément qui peut influencer la décision de migration de la BI vers le cloud est une dimension et ses constituants, des caractéristiques. Si dans la suite, une autre itération montre qu'une caractéristique précédemment identifiée est, elle aussi, constituée d'autres éléments, cette caractéristique devient catégorie à l'aide de l'opérateur promotion et ses constituants, des caractéristiques. De même pour une dimension déjà établie, si une autre itération montre qu'elle est un composant d'un autre concept, celle-ci sera rétrogradée au rang de catégorie. La typologie des opérations présentée au point 3.4 va servir à structurer la taxonomie en plaçant chaque élément à sa place.

Dans cette section, nous décrivons le déroulement de la méthodologie avec, à chaque itération, l'apport de l'article à la construction de la taxonomie.

- Itération 1 (Cao *et al.* 2011) : l'article n'apporte rien⁴.
- Itération 2 (Demirkan and Delen 2013) : l'article permet d'ajouter la dimension *exigences de sécurité* avec les caractéristiques *intégrité* et *disponibilité*.
- Itération 3 (Chi *et al.* 2011) : l'article n'apporte rien.
- Itération 4 (Baars and Kemper 2010) : l'article permet d'ajouter la dimension *objectif de la migration* avec les caractéristiques *performance des services* et *réduction des coûts*. Il permet d'ajouter la caractéristique *confidentialité* à la dimension *exigences de sécurité*.
- Itération 5 (Mian, Martin and Vazquez-Poletti 2013) : l'article permet d'ajouter la dimension *taille de l'entreprise* avec les caractéristiques *petite* et *moyenne*.
- Itération 6 (Abelló, Ferrarons and Romero 2011) : l'article confirme la dimension *objectif de la migration* avec la caractéristique *réduction des coûts*.
- Itération 7 (Wang, Song and Luo 2010) : l'article n'apporte rien.
- Itération 8 (Muriithi and Kotzé 2013): l'article permet d'ajouter les caractéristiques *manipulation de grandes quantités de données* et *protection des données* à la dimension *objectif de migration*. Il permet d'ajouter la dimension *secteur d'activité* avec les caractéristiques : *éducation et recherche*, *administration publique* et *commerce*.

⁴ Cet article ne traite pas des éléments qui peuvent influencer la migration de la BI vers le cloud.

- Itération 9 (Chang 2014) : l'article confirme la caractéristique *performance des services*. Il permet d'ajouter la caractéristique *institutions financières* à la dimension *secteur d'activité*.
- Itération 10 (Ng *et al.* 2011) : l'article confirme la dimension *exigences de sécurité*.
- Itération 11 (Garey and Johnson 2009) : l'article n'apporte rien.
- Itération 12 (Barga, Ekanayake and Lu 2012) : l'article confirme la caractéristique *performance des services*.
- Itération 13 (Brezany *et al.* 2011a) : l'article confirme la caractéristique *performance des services* dans la dimension *objectif de la migration*.
- Itération 14 (Chaisiri *et al.* 2012) : l'article n'apporte rien.
- Itération 15 (Chohan *et al.* 2012) : l'article confirme les caractéristiques *performance des services* et *réduction des coûts* de la dimension *objectif de la migration*.
- Itération 16 (Al-Aqrabi *et al.* 2012) : l'article confirme la dimension *taille de l'entreprise* ainsi que la caractéristique *performance des services* de la dimension *objectif de la migration*.
- Itération 17 (Assunção *et al.* 2014) : L'article permet d'ajouter la caractéristique *industrie* dans la dimension *secteur d'activité*.
- Itération 18 (AbdelBaky *et al.* 2012) : l'article n'apporte rien.
- Itération 19 (Jun and Jun 2011) : l'article n'apporte rien.
- Itération 20 (Goncalves, Assuncao and Cunha 2012) : l'article n'apporte rien.
- Itération 21 (Hua, Liu and Jiang 2014) : l'article n'apporte rien.
- Itération 22 (Lee *et al.* 2012) : l'article n'apporte rien.
- Itération 23 (Kurunji *et al.* 2012) : l'article n'apporte rien.
- Itération 24 (Tapiador *et al.* 2014) : l'article n'apporte rien.
- Itération 25 (Arres, Kabbachi and Boussaid 2013) : l'article confirme la dimension *secteur d'activité* avec ses caractéristiques *industrie* et *éducation et recherche*.
- Itération 26 (Wlodarczyk *et al.* 2010) : l'article n'apporte rien.
- Itération 27 (Gash, Ariyachandra and Frolick 2011) : l'article permet d'ajouter la dimension *ressources humaines* et la dimension *ressources financières* ainsi que la caractéristique *grande* pour la dimension *taille de l'entreprise*. Il confirme la dimension *exigences de sécurité*.
- Itération 28 (Kaur, Agrawal and Dhiman 2012) : l'article n'ajoute rien.

- Itération 29 (Tancer and Varde 2011) : l'article permet d'ajouter la caractéristique *santé* à la dimension *secteur d'activité*.
- Itération 30 (Han *et al.* 2013) : l'article n'apporte rien.
- Itération 31 (Gan, Meng and Shi 2013) : l'article n'apporte rien.
- Itération 32 (Al-Aqrabi *et al.* 2015) : l'article confirme les caractéristiques *performance des services* et *réduction des coûts*.
- Itération 33 (Nayem, Fahad and Akhter 2013) : l'article confirme les caractéristiques *performance des services* et *réduction des coûts* de la dimension *objectifs* de migration.
- Itération 34 (Bennett *et al.* 2010) : l'article n'apporte rien
- Itération 35 (Juan-Verdejo *et al.* 2014) : l'article confirme les dimensions *exigences de sécurité* et *objectif de la migration* ainsi que la catégorie *réduction des coûts*.
- Itération 36 (Herwig 2013) : l'article confirme la dimension *exigences de sécurité*.
- Itération 37 (Stanoevska-Slabeva, Wozniak and Ristol 2010) : l'article n'apporte rien.
- Itération 38 (Patel, Rau-Chaplin and Varghese 2012) : l'article permet d'ajouter les caractéristiques *analyste et développeur* à la dimension *ressources humaines*.
- Itération 39 (Chang *et al.* 2015) : l'article confirme la dimension *exigences de sécurité*.
- Itération 40 (Chang *et al.* 2013) : l'article confirme la dimension *exigences de sécurité*.
- Itération 41 (Gatzioura *et al.* 2012) : l'article confirme la dimension *exigences de sécurité*.
- Itération 42 (Yuan and Herbert 2014) : l'article confirme la dimension *secteur d'activité* avec sa caractéristique *santé*.
- Itération 43 (Kasem and Hassanein 2014) : l'article confirme les dimensions *exigences de sécurité* et *objectif de migration* avec ses caractéristiques *performance des services* et *réduction des coûts*.

Itération	CAS (a, b, c, d) (voir point 5.2.5)	Si cas c (voir point 5.2.5)	Si cas d (voir point 5.2.5)
1	a		
2	c	Dimension : exigences de sécurité	

		Caractéristiques : intégrité et disponibilité	
3	a		
4	c	Dimension : objectifs Caractéristiques : performance des services et réduction des coûts Caractéristique : confidentialité (dimension exigences de sécurité)	
5	c	Dimension : taille de l'entreprise Caractéristiques : petite et moyenne	
6	b		
7	a		
8	c	Caractéristiques : manipulation de grandes quantités des données et protection des données (dimension objectifs). Dimension : secteur d'activité Caractéristiques : administration publique, éducation et recherche et commerce	
9	c	Caractéristique : institutions financières (dimension secteur d'activité)	
10	b		
11	a		
12 - 13	b		
14	a		
15 - 16	b		
17	c	Caractéristique : industrie (dimension : secteur d'activité)	
18 - 24	a		
25	b		
26	a		
27	c	Dimension : ressources humaines Dimension : ressources financières	
28	a		
29	c	Caractéristique : Santé (dimension secteur d'activité)	
30 - 31	a		
32 - 33	b		
34	a		
35 - 36	b		
37	a		
38	b et c	Caractéristiques : analyste et développeur (dimension ressources humaines)	
39 - 43	b		

Tableau 4 : Taxonomie de contexte : itérations des sources de type 1

Le tableau 4 montre que les articles traitent plus de l'objet même de la migration de la BI vers le cloud que des éléments qui peuvent influencer cette décision. Néanmoins, nous avons pu identifier les éléments nécessaires qui peuvent aider à structurer une taxonomie de contexte dès l'examen des sources de type 1.

Seules neuf itérations sur les quarante-trois ont apporté des informations relatives à la taxonomie de contexte. Dix-sept itérations se retrouvent dans le cas "a" et dix-sept autres dans le cas "b".

5.2.5.2. *Taxonomie de décision*

Tout concept relatif à la BI dans le cloud est une dimension de la taxonomie de décision et ses constituants, les caractéristiques. Grâce aux opérateurs proposés au point 3.4, particulièrement l'opérateur union, nous prendrons en compte les éléments pertinents relatifs à la BI dans le cloud. L'insertion d'un nouvel élément peut nécessiter l'utilisation d'un autre opérateur pour réorganiser la taxonomie dans le but de permettre au nouvel élément de se placer au bon endroit dans la taxonomie. Par exemple, si un ou plusieurs nouveaux éléments sont des constituants d'une caractéristique précédemment identifiée, à l'aide de l'opérateur promotion, cette caractéristique devient catégorie et ses constituants, les caractéristiques. De même pour une dimension déjà établie, si une autre itération montre qu'elle est un composant d'un autre concept, celle-ci sera rétrogradée au rang de catégorie. Deux concepts qui ont les mêmes caractéristiques vont être fusionnés en un seul.

Nous réexaminons les sources de type 1 pour prendre en compte les options de la décision contenues dans ces sources.

- Itération 1 (Cao *et al.* 2011) : L'article permet d'ajouter la dimension *modèle de programmation* avec les caractéristiques : *MapReduce*, *Hadoop*. Il permet de modifier la taxonomie par la promotion de *Hadoop* et permet d'ajouter les caractéristiques *Hive* et *Pig* à la catégorie *Hadoop*.
Il permet d'ajouter la dimension *système de stockage*, les catégories *bases de données relationnelles* et *NoSQL* avec les caractéristiques *Dynamo*, *Cassandra*, *BigTable* et *PNUTS* pour la catégorie *NoSQL*.
- Itération 2 (Demirkan and Delen 2013) : L'article permet d'ajouter la dimension *voie de transmission des services* avec la caractéristique *internet*. Il confirme la dimension *modèle de programmation*.
- Itération 3 (Chi *et al.* 2011) : l'article permet d'ajouter la dimension *contrat* avec la catégorie *SLA (service-level agreement)* et ses caractéristiques *problèmes juridiques* ainsi que *niveau de service à fournir*.

- Itération 4 (Baars and Kemper 2010) : l'article permet d'ajouter la dimension *couche de cloud* ; les catégories : *SaaS, PaaS et IaaS* avec les caractéristiques : *service, outils, solution* (pour *SaaS*); *solution mixte, BI composite* (pour *PaaS*) ; *stockage et réseau* (pour *IaaS*).
Il ajoute la catégorie *durée* à la dimension *contrat* avec les caractéristiques : *long terme* et *court terme*.
Il ajoute la dimension *niveau de migration* ; les catégories *total* et *partiel* ; les sous-catégories *matériel* et *outils* pour la catégorie *partiel* avec les caractéristiques *réseau* et *stockage* pour la sous-catégorie *matériel* et les caractéristiques *ETL, data warehouse, analyse, reporting* pour la sous-catégorie *outils*.
L'article modifie la taxonomie et requiert la promotion des caractéristiques : *service, outils, solution, BI composite, solution mixte* et *stockage*.
L'article permet aussi d'ajouter les caractéristiques : *logiciel de portail, composant ETL, composant OLAP, service web et composant de visualisation des données* (pour la catégorie *service*).
Il permet encore d'ajouter la dimension *couches de BI* avec les caractéristiques *couche données ; couche logique ; couche accès*.
L'article permet de modifier la taxonomie par la rétrogradation de la dimension *système de stockage*, la promotion de la caractéristique *stockage* ainsi que la fusion des catégories *système de stockage* et *stockage*.
Il induit la rétrogradation de la dimension *modèle de programmation*, ainsi que la fusion des catégories *modèle de programmation* et *BI composite*.
L'article permet de modifier la taxonomie par la rétrogradation de la dimension *couche de BI* ainsi que la fusion des catégories *couche de BI* et *solution*.
Il permet d'ajouter les caractéristiques *réseau d'entreprise, centre des données du fournisseur* à la dimension *voie de transmission des services*.
- Itération 5 (Mian, Martin and Vazquez-Poletti 2013): l'article permet d'ajouter la dimension *types de cloud* avec les caractéristiques *privé, public*.
- Itération 6 (Abelló, Ferrarons and Romero 2011) : l'article confirme la catégorie *stockage* et la catégorie *outils*. L'article confirme les sous-catégories *bases de données relationnelles* et *NoSQL*. Il confirme la catégorie *Hadoop* et la caractéristique *MapReduce*.
- Itération 7 (Wang, Song and Luo 2010) : l'article confirme la caractéristique *MapReduce* et la catégorie *Hadoop*. Il confirme la sous-catégorie *outils*.

- Itération 8 (Muriithi and Kotzé 2013) : l'article permet d'ajouter les caractéristiques *hybride* et *communautaire* (à la dimension *types de cloud*). Il confirme la sous-catégorie *outils* avec ses caractéristiques *ETL*, *data warehouse*, *analyse* et *reporting*. Il confirme la dimension *niveau de migration*.
- Itération 9 (Chang 2014) : l'article confirme les dimensions *types de cloud* et *couches de cloud*.
- Itération 10 (Ng *et al.* 2011) : l'article confirme la dimension *les couches de cloud*.
- Itération 11 (Garey and Johnson 2009) : l'article confirme la catégorie *BI composite* avec la catégorie *Hadoop* et la caractéristique *MapReduce*.
- Itération 12 (Barga, Ekanayake and Lu 2012) : l'article confirme la caractéristique *MapReduce*.
- Itération 13 (Brezany *et al.* 2011a) : l'article confirme la catégorie *BI composite* avec sa caractéristique *MapReduce* et sa sous-catégorie *Hadoop*.
- Itération 14 (Chaisiri *et al.* 2012) : l'article confirme la catégorie *outils* et sa caractéristique *analyse*.
- Itération 15 (Chohan *et al.* 2012) : l'article confirme la dimension *types de cloud* ainsi que les catégories *PaaS* et *IaaS* de la dimension *couches de cloud*.
- Itération 16 (Al-Aqrabi *et al.* 2012) : l'article confirme la dimension *couches de cloud*.
- Itération 17 (Assunção *et al.* 2014) : l'article confirme la dimension *types de cloud* et la catégorie *BI composite* avec sa caractéristique *MapReduce* et sa sous-catégorie *Hadoop*.
- Itération 18 (AbdelBaky *et al.* 2012) : l'article confirme la catégorie *BI composite* avec sa caractéristique *MapReduce* et sa sous-catégorie *Hadoop*.
- Itération 19 (Jun and Jun 2011) : l'article confirme la dimension *couches de cloud*.
- Itération 20 (Goncalves, Assuncao and Cunha 2012) : l'article confirme la catégorie *BI composite* avec sa caractéristique *MapReduce* et sa sous-catégorie *Hadoop*.
- Itération 21 (Hua, Liu and Jiang 2014) : l'article confirme la caractéristique *MapReduce* et la sous-catégorie *Hadoop*.
- Itération 22 (Lee *et al.* 2012) : l'article confirme la catégorie *BI composite*.
- Itération 23 (Kurunji *et al.* 2012) : l'article confirme la sous-catégorie *stockage*.

- Itération 24 (Tapiador *et al.* 2014) : l'article confirme la catégorie *BI composite*, sa caractéristique *MapReduce* et sa sous-catégorie *Hadoop* ainsi que les caractéristiques *Hive* et *Pig* pour *Hadoop*.
- Itération 25 (Arres, Kabbachi and Boussaid 2013) : l'article confirme la dimension *types de cloud* et la catégorie *BI composite*, sa caractéristique *MapReduce* et sa sous-catégorie *Hadoop* ainsi que la caractéristique *Hive* pour *Hadoop*.
- Itération 26 (Wlodarczyk *et al.* 2010) : l'article confirme la catégorie *BI composite*, sa caractéristique *MapReduce* et sa sous-catégorie *Hadoop* ainsi que les caractéristiques *Hive* et *Pig* pour *Hadoop*.
- Itération 27 (Gash, Ariyachandra and Frolick 2011) : l'article n'apporte rien.
- Itération 28 (Kaur, Agrawal and Dhiman 2012) : l'article confirme la dimension *types de cloud* et la dimension *couches de cloud*.
- Itération 29 (Tancer and Varde 2011): l'article confirme la dimension *couches de cloud* et la caractéristique *MapReduce* et la sous-catégorie *Hadoop*.
- Itération 30 (Han *et al.* 2013) : l'article confirme la dimension *types de cloud* et la dimension *couches de cloud* ainsi que la catégorie *BI composite* avec sa caractéristique *MapReduce* et sa sous-catégorie *Hadoop*.
- Itération 31 (Gan, Meng and Shi 2013) : l'article confirme la caractéristique *MapReduce* et sous-catégorie *Hadoop*.
- Itération 32 (Al-Aqrabi *et al.* 2015) : l'article confirme la dimension *couches de cloud*.
- Itération 33 (Nayem, Fahad and Akhter 2013) : l'article confirme la dimension *types de cloud*. Il permet d'ajouter la dimension *support des applications BI mobiles* avec ses caractéristiques *tablette* et *téléphone*.
- Itération 34 (Bennett *et al.* 2010) : l'article permet d'ajouter la caractéristique *Hbase* à la sous-catégorie *Hadoop*.
- Itération 35 (Juan-Verdejo *et al.* 2014) : l'article confirme la dimension *couches de cloud* et la dimension *niveau de migration*.
- Itération 36 (Herwig 2013) : l'article confirme la dimension *couches de cloud*.
- Itération 37 (Stanoevska-Slabeva, Wozniak and Ristol 2010) : l'article confirme la dimension *couches de cloud* et la dimension *types de cloud*.
- Itération 38 (Patel, Rau-Chaplin and Varghese 2012) : l'article confirme la catégorie *PaaS* de la dimension *couches de cloud*. Il confirme la sous-catégorie *outils* et sa caractéristique *analyse*.

- Itération 39 (Chang *et al.* 2015) : l'article confirme la dimension *types de cloud*.
- Itération 40 (Chang *et al.* 2013) : l'article confirme la dimension *couches de cloud*.
- Itération 41 (Gatzioura *et al.* 2012) : l'article confirme la dimension *couches de cloud* et la dimension *contrat* avec sa catégorie *SLA*.
- Itération 42 (Yuan and Herbert 2014) : l'article confirme la dimension *supports des applications BI mobile* et la caractéristique *téléphone*.
- Itération 43 (Kasem and Hassanein 2014) : l'article confirme la dimension *couches de cloud*, la dimension *types de cloud* et la sous-catégorie *outils* avec ses caractéristiques *ETL, data warehouse, analyse et reporting*.

Itération	CAS (a, b, c, d) (voir point 5.2.5)	Si cas c (voir point 5.2.5)	Si cas d (voir point 5.2.5)
1	c et d	Dimension : modèle de programmation Caractéristiques : MapReduce et Hadoop Caractéristiques : Hive et Pig (catégorie Hadoop) Dimension : système de stockage Catégories : BDR et NoSQL Caractéristiques : Dynamo, Cassandra, BigTable et PNUTS (catégorie NoSQL)	Promotion de Hadoop
2	c	Dimension : voie de transmission des services Caractéristique : internet	
3	c	Dimension : contrat Catégorie : SLA Caractéristique : problèmes juridiques et normes de conduite	
4	c et d	Dimension : Couches de cloud. Catégories : SaaS, PaaS et IaaS. Caractéristiques : service, outils, solution (SaaS), solution mixte, BI composite (PaaS), stockage et réseau (IaaS). Catégorie : durée Caractéristiques : long et court terme (dimension contrat). Dimension : niveau de migration. Catégories : total et partiel. Sous-catégories : matériel et outils Caractéristiques : réseau et stockage (matériel), ETL, data warehouse, analyse et reporting (outils) Caractéristiques : logiciel portail, composant ETL et composant OLAP, service web, composant de visualisation des données (catégorie service). Dimension : couches de BI. Caractéristiques : couche données; couche	Promotion des caractéristiques : service, outils, solution, BI composite et solution. Rétrogradation : dimension système de stockage. Promotion caractéristique : stockage. Fusion des catégories système de stockage et stockage. Rétrogradation dimension : modèle de programmation. Fusion des dimensions : modèle de programmation et BI composite.

		logique; couche accès; Caractéristiques : réseau d'entreprise et centre de données du fournisseur (dimension voie de transmission des services).	Rétrogradation dimension : couche BI Fusion des catégories : couches BI et solution
5	c	Dimension : Types de cloud Caractéristiques : Privé et Public	
6	b		
7	b		
8	b et c	Caractéristiques : Hybride et Communautaire (pour dimension types de cloud)	
9 - 22	b		
23	a		
24 - 26	b		
27	a		
28 -32	b		
33	c	Dimension : supports des applications BI mobiles Caractéristiques : téléphone, tablette	
34	c	Caractéristique : Hbase (sous-catégorie Hadoop)	
35 - 43	b		

Tableau 5 : Taxonomie de décision : itérations des sources de type 1

Le tableau 5 montre que l'examen des sources de type 1 a fourni un grand nombre d'éléments pour structurer la taxonomie de décision. En général, ces sources proposent les mêmes options pour le déploiement de la BI dans le cloud. Les auteurs se focalisent plus sur les solutions cloud, très peu d'entre eux décrivent les contextes qui peuvent militer pour l'adoption de ces solutions.

Par ailleurs, seules les itérations. 1, 2, 3, 4, 5, 8, 33 et 34 ont apporté des éléments pour la construction de la taxonomie. De manière particulière, l'itération 4 a apporté un nombre considérable de modifications. Plusieurs itérations se trouvent dans le cas "b" (trente-trois articles sur les quarante-trois), les articles confirment des éléments qui existent déjà dans la taxonomie. Seuls deux articles se trouvent dans le cas "a".

5.2.6. Structuration de la taxonomie

A l'aide de la typologie des opérateurs présentés au point 3.4, nous avons structuré les éléments identifiés dans les sources de type 1, de type 2 et de type 3. Le cas "c" (l'article permet d'ajouter une dimension, une catégorie ou encore une caractéristique) a apporté des

modifications dans la taxonomie et a conduit quelques fois au cas "d" (l'article modifie la taxonomie et va engendrer la fusion, la scission, la promotion ou la rétrogradation d'un élément de la taxonomie).

5.2.7. Dernière source

Les sources sont examinées l'une après l'autre selon l'ordre croissant de leurs types respectifs (voir le point 4.4.5). A chaque itération, après la restructuration de la taxonomie au cas où cela s'avérait nécessaire, nous vérifions si la source examinée est la dernière de l'échantillon. Au cas où ce n'était pas le cas, nous examinons une autre source jusqu'à finir toutes les sources de l'échantillon. Etant donné qu'examiner toutes les sources d'un échantillon n'est pas une condition suffisante pour arrêter le processus de développement d'une taxonomie, après cette étape, nous avons vérifié la convergence du processus de construction de la taxonomie.

5.2.8. Convergence

L'examen des quarante-trois sources de type 1 a donné lieu à un total de cinquante-sept opérations sur les deux taxonomies. La figure 10 illustre le nombre cumulé d'opérations après chaque itération. Dans l'itération 4, un nombre important d'opérations a été effectué. Au cours des dernières itérations, la courbe a tendance à converger. Cette convergence montre qu'il n'y a plus beaucoup d'informations nouvelles à ajouter dans la taxonomie jusqu'à la dernière itération.

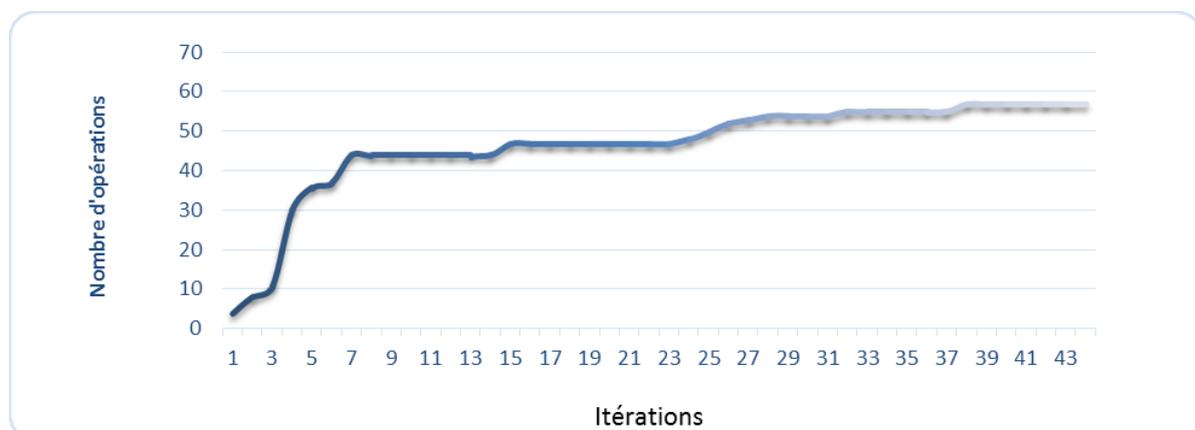


Figure 10 : Examen des sources de type 1 : nombre d'opérations

Nous avons arrêté le processus de développement de la taxonomie après l'examen de toutes les sources de notre échantillon (les trois types des sources). Comme pour les sources de type 1, nous avons construit des graphiques représentant le nombre cumulé d'opérations pour les sources de type 2 et les sources de type 3. Ces trois graphiques ont une caractéristique commune : vers la dernière itération, le nombre d'opérations diminue et s'approche de zéro car il n'y a plus beaucoup d'informations utiles à ajouter dans la taxonomie. Les opérateurs présentés au point 3.4 nous ont permis de compter les opérations effectuées et de constater que lorsqu'on avance vers la fin de processus, il n'y a plus beaucoup d'opérations à effectuer : toutes les sources traitent du même sujet. Par conséquent, les sources examinées vers la fin de processus confirment les éléments qui ont été tirés des sources précédemment examinées. Grâce à la typologie des sources, nous avons constaté que tous les types de sources ont été examinés.

Notons que dans le développement de ces deux taxonomies (de contexte et de décision), en ce qui concerne les sources de type 3, la convergence n'a pas été atteinte avec le premier échantillon de sources constitué avant d'entamer le processus. Nous avons dû constituer un deuxième et un troisième échantillon à des points ultérieurs permettant de découvrir de nouveaux articles afin d'atteindre la convergence dont nous avons parlé ci-haut.

5.2.9. Validation de la taxonomie

Dans cette thèse, nous avons adopté l'approche science de conception (voir point 1.3). Raison pour laquelle nous avons élaboré la taxonomie de la BI dans le cloud à partir des études qui utilisent cette approche. Le résultat de la recherche en science de conception, étant par définition, un artéfact, notre objectif était aussi de repérer dans la littérature, les artéfacts déjà développés pour que la BI tire profit du cloud computing. En vue de valider cette taxonomie, il est utile de compléter nos connaissances par les articles académiques relatifs à la BI dans le cloud qui ne relèvent pas de recherche en science de conception. Ils effectuent des recherches quantitatives ou qualitatives, exhibant des modèles qui peuvent contenir des éléments décrivant la migration de la BI vers le cloud.

Ainsi, nous allons constituer un échantillon d'articles académiques qui utilisent les méthodes quantitatives ou qualitatives. Comme pour le premier échantillon des articles qui présentent une recherche en science de conception, nous allons fouiller dans les moteurs de recherche suivants : Google Scholar, ACM, IEEE, DBLP, ScienceDirect, EBSCO, et la librairie électronique de l'AIS (AISEL). Pour des raisons pratiques, nous allons nous limiter aux dix premiers écrans de résultats.

5.2.9.1. *Sources et critères de sélection*

Cette fois-ci, les critères d'inclusion des sources dans l'échantillon seront les suivants :

- **Contenu** : le sujet traité dans la source doit être lié à la BI dans le cloud (non à la BI et au cloud computing séparément).
- **Langue de publication** : la recherche sera limitée aux publications en anglais.
- **Revues** : pour assurer la sélection des sources de haute qualité, nous proposons de sélectionner les publications académiques de la liste de revues et des actes des conférences de "ERA" (ERA 2012).
- **Conception de la recherche** ou de la méthodologie d'échantillonnage : pour permettre de valider la taxonomie élaborée avec les articles faisant de la recherche en science de conception, cette fois-ci nous avons collecté les documents utilisant une autre approche que la science de conception, par exemple, les articles faisant une étude quantitative ou qualitative. De manière générale, dans ces articles les auteurs développent et vérifient des théories qui expliquent ou prédisent les comportements humains ou organisationnels (Hevner *et al.* 2004).
- **Date de publication**: Etant donné que les technologies complexes émergentes peuvent rapidement devenir obsolètes, nous proposons de limiter la recherche des publications universitaires des dix dernières années: 2006-2016.

Comme pour le premier échantillon constitué grâce au protocole présenté au point 2.3.4.2, les termes utilisés afin de collecter les articles relatifs à la BI dans le cloud sont les suivants :

- ("Business Intelligence "OR "Analytics" OR "Data warehouse") AND ("Cloud" OR "SaaS").

Cette requête sera scindée en deux pour Google Scholar :

- "Cloud" AND ("Business Intelligence" OR "Analytics" OR "Data warehouse")
- "SaaS" AND ("Business Intelligence" OR "Analytics" OR "Data warehouse").

Après l'exécution des requêtes à l'aide de chacun des moteurs de recherche et la déduplication des résultats, nous avons retenu 121 articles qui traitent de la BI dans le cloud avec une autre approche que la science de conception. L'application du critère de qualité "Etre publié dans une revue ou des actes de conférences figurant sur la liste d'ERA" a ramené notre échantillon à vingt-cinq articles. Dans les lignes qui suivent, nous allons examiner ces vingt-cinq articles pour pouvoir valider notre taxonomie telle qu'élaborée avec les articles traitant de la recherche en science de conception.

5.2.9.2. Examen des sources

Rappelons que notre taxonomie a deux parties complémentaires. Nous allons donc relever les informations relatives à ces deux parties : taxonomie de contexte et taxonomie de décision en vue de leur validation.

1) Taxonomie de contexte

Ici, nous allons examiner les vingt-cinq articles pour faire ressortir les éléments qui peuvent influencer la migration de la BI vers le cloud.

- Itération 1 (Böhringer *et al.* 2010) : l'article valide *l'objectif agilité*.
- Itération 2 (Sun *et al.* 2012) : l'article valide *l'objectif réduction des coûts*, la dimension *taille de l'entreprise* et la dimension *ressources humaines*.
- Itération 3 (Cuzzocrea, Song and Davis 2011) : l'article valide *l'objectif performance des services* et *l'objectif manipulation de grandes quantités des données*.
- Itération 4 (Abbasi, Sarker and Chiang 2016) : l'article valide *l'objectif manipulation de grandes quantités de données*.
- Itération 5 (Chen, Chiang and Storey 2012) : l'article valide la dimension *secteur d'activité* avec ses caractéristiques *commerce, administration publique, éducation et recherche* ainsi que *santé*.
- Itération 6 (Thompson and Van der Walt 2010) : l'article valide *l'objectif réduction des coûts* et *l'objectif performance des services*.
- Itération 7 (Dziembek and Ziora 2014) : l'article valide *l'objectif mobilité* et *l'objectif réduction des coûts*.

- Itération 8 (Ereth and Baars 2015) : l'article valide l'*objectif réduction des coûts* et l'*objectif d'agilité*.
- Itération 9 (Simmhan *et al.* 2013) : l'article valide les *exigences de sécurité*.
- Itération 10 (Xu *et al.* 2009) : l'article ajoute le *secteur d'activité télécommunication*.
- Itération 11 (Mahmud, Iqbal and Doctor 2016) : l'article valide le *secteur d'activité santé*, l'*objectif manipulation de grandes quantités des donnée* et l'*objectif performance des services*.
- Itération 12 (Xiufeng, Thomsen and Pedersen 2014) : l'article valide l'*objectif performance des services*.
- Itération 13 (Ari, Olmezogullari and Çelebi 2012) : l'article valide le *secteur d'activité administration publique*, le *secteur d'activité commerce*, le *secteur d'activité institutions financières* et le *secteur d'activité télécommunication* ainsi que l'*objectif performance des services*.
- Itération 14 (Knabke and Olbrich 2015) : l'article valide l'*objectif agilité* et l'*objectif performance des services*.
- Itération 15 (Baur, Bühler and Bick 2015) : l'article valide la dimension *ressources financières*.
- Itération 16 (Rahman, Li and Palit 2011) : l'article valide la dimension *ressources financières* et l'*objectif performance des services*.
- Itération 17 (Fox 2012) : l'article valide l'*objectif réduction des coûts* et l'*objectif performance des services*.
- Itération 18 (Dehne *et al.* 2015) : l'article valide l'*objectif performance des services*.
- Itération 19 (Tan *et al.* 2013) : l'article valide la dimension *exigences de sécurité* ainsi que l'*objectif manipulation de grandes quantités de données*.
- Itération 20 (Ghoshal *et al.* 2014) : l'article valide l'*objectif manipulation de grandes quantités de données*, l'*objectif performance des services* et la dimension *exigences de sécurité*.
- Itération 21 (Ram, Zhang and Koronios 2016) : l'article valide l'*objectif performance des services*, l'*objectif réduction des coûts* et l'*objectif manipulation de grandes quantités de données* ainsi que la dimension *exigences de sécurité*.
- Itération 22 (Kambatla *et al.* 2014) : l'article valide le *secteur d'activité commerce* et le *secteur d'activité santé* ainsi que l'*objectif manipulation de grandes quantités de données*.

- Itération 23 (Baars and Kemper 2011) : l'article valide l'*objectif agilité* et l'*objectif performance des services*.
- Itération 24 (Grossman 2009) : l'article valide l'*objectif performance des services* et les *exigences de sécurité*.
- Itération 25 (Chaudhuri 2012) : l'article valide l'*objectif réduction des coûts*, l'*objectif manipulation de grandes quantités des données* et l'*objectif performance des services*.

Le résultat de l'examen des sources ne relevant pas de la science de conception valide les six dimensions de la taxonomie de contexte. Aucune nouvelle dimension n'a été ajoutée à la taxonomie telle que construite avec les articles relevant de la science de conception. En ce qui concerne les caractéristiques, ce sont majoritairement celles de la dimension *objectifs* qui sont validées. On pourrait dire, de manière générale, que les organisations qui migrent vers le cloud sont motivées par la performance des services et la réduction des coûts. En effet, ce sont ces objectifs qui reviennent le plus souvent.

2) Taxonomie de décision

Ici, nous allons réexaminer les vingt-cinq articles pour faire ressortir les informations relatives à la taxonomie de décision.

- Itération 1 (Böhringer *et al.* 2010) : l'article valide la couche *SaaS*.
- Itération 2 (Sun *et al.* 2012) : l'article valide la couche *SaaS*.
- Itération 3 (Cuzzocrea, Song and Davis 2011) : l'article valide la couche *IaaS*, la sous-catégorie *Hadoop* et sa caractéristique *Hive*.
- Itération 4 (Abbasi, Sarker and Chiang 2016) : l'article valide les *couches de cloud*.
- Itération 5 (Chen, Chiang and Storey 2012) : l'article valide le *système de stockage BigTable* ainsi que les *couches de cloud*. Il valide les *supports des applications BI mobiles*.
- Itération 6 (Thompson and Van der Walt 2010) : l'article valide la *voie de transmission internet* et la couche *SaaS*.
- Itération 7 (Dziembek and Ziora 2014) : l'article valide la couche *SaaS*.
- Itération 8 (Ereth and Baars 2015) : l'article valide les *supports des applications BI mobiles*.
- Itération 9 (Simmhan *et al.* 2013) : l'article valide les *types de cloud public et privé* ainsi que les *couches de cloud PaaS et IaaS*.

- Itération 10 (Xu *et al.* 2009) : l'article valide la couche *PaaS*. Il valide la caractéristique *MapReduce* ainsi que la sous-catégorie *Hadoop* et sa caractéristique *Hbase*.
- Itération 11 (Mahmud, Iqbal and Doctor 2016) : l'article valide la *voie de transmission internet* et les *couches de cloud*.
- Itération 12 (Xiufeng, Thomsen and Pedersen 2014) : l'article valide la caractéristique *MapReduce* et la sous-catégorie *Hadoop* avec ses caractéristique *Pig* et *Hive*. Il valide les *outils ETL* et *data warehouse*.
- Itération 13 (Ari, Olmezogullari and Çelebi 2012) : l'article valide la couche cloud *PaaS/BI composite/Hadoop* et la couche de cloud *IaaS/stockage/Cassandra, MongoDB* ainsi que les *types de cloud privé et public*.
- Itération 14 (Knabke and Olbrich 2015) : l'article valide les *supports des applications de BI mobiles*.
- Itération 15 (Baur, Bühler and Bick 2015) : l'article valide la couche *SaaS*.
- Itération 16 (Rahman, Li and Palit 2011) : l'article valide le *type de cloud hybride* et *voie de transmission internet*.
- Itération 17 (Fox 2012) : l'article valide les couches de cloud *PaaS* et *IaaS*.
- Itération 18 (Dehne *et al.* 2015) : l'article valide les *outils/Data warehouse* et les *outils/analyse*, le *niveau de migration partiel/outils/Analyse* et *DWH*.
- Itération 19 (Tan *et al.* 2013) : l'article valide la catégorie *outils* avec les caractéristiques *ETL* et *data warehouse*. Il valide les sous-catégories *bases de données relationnelles* et *NoSQL*.
- Itération 20 (Ghoshal *et al.* 2014) : l'article valide la dimension *supports des applications BI mobiles*.
- Itération 21 (Ram, Zhang and Koronios 2016) : l'article valide les *outils d'analyse*.
- Itération 22 (Kambatla *et al.* 2014) : l'article valide les *supports des applications BI mobiles* et les *couches de cloud*.
- Itération 23 (Baars and Kemper 2011) : l'article valide les *types de cloud privé et public* ainsi que les *couches de cloud*. Il valide les *supports des applications BI mobiles*.
- Itération 24 (Grossman 2009) : l'article valide la caractéristique *MapReduce* et la sous-catégorie *Hadoop*.
- Itération 25 (Chaudhuri 2012) : l'article valide la dimension *contrat/SLA* et les *couches de cloud*.

Comme pour la taxonomie de contexte, l'examen de ces sources a validé toutes les dimensions de la taxonomie de décision. Aucune nouvelle dimension n'a été ajoutée à la taxonomie telle que construite avec les articles relevant de la science de conception.

Après l'examen de ces sources qui utilisent les méthodes qualitative et quantitative, le constat est : ces sources proposent les mêmes éléments que les sources qui relèvent de la science de conception. Raison pour laquelle l'examen de ces sources n'a pas apporté beaucoup d'éléments nouveaux dans la taxonomie telle qu'élaborée avec les articles qui traitent de la science de conception. Le nombre d'opérations effectuées a été minime : une seule. Nous pouvons donc arrêter le processus de développement de la taxonomie car nous avons examiné toutes les sources de notre échantillon et nous n'avons plus d'informations utiles à ajouter.

5.3. Taxonomies résultantes

Dans cette section, nous présentons les deux taxonomies complémentaires que nous avons développées.

La définition de la taxonomie proposée par Nickerson, Varshney et Muntermann (2013) stipule qu'une taxonomie est un ensemble de dimensions qui possèdent chacune des caractéristiques mutuellement exclusives et collectivement exhaustives telles que chaque objet considéré ne peut avoir qu'une et une seule caractéristique pour chaque dimension. Cette définition ne s'applique pas à ces deux taxonomies complémentaires. En effet, ici, les caractéristiques d'une dimension sont toutes exhaustives à un moment donné et par rapport à la documentation à notre disposition. Cependant, par construction, elles sont exclusives quand elles forment une partition et inclusives quand elles ne forment pas une partition (voir le point 3.3).

5.3.1 Taxonomie de contexte

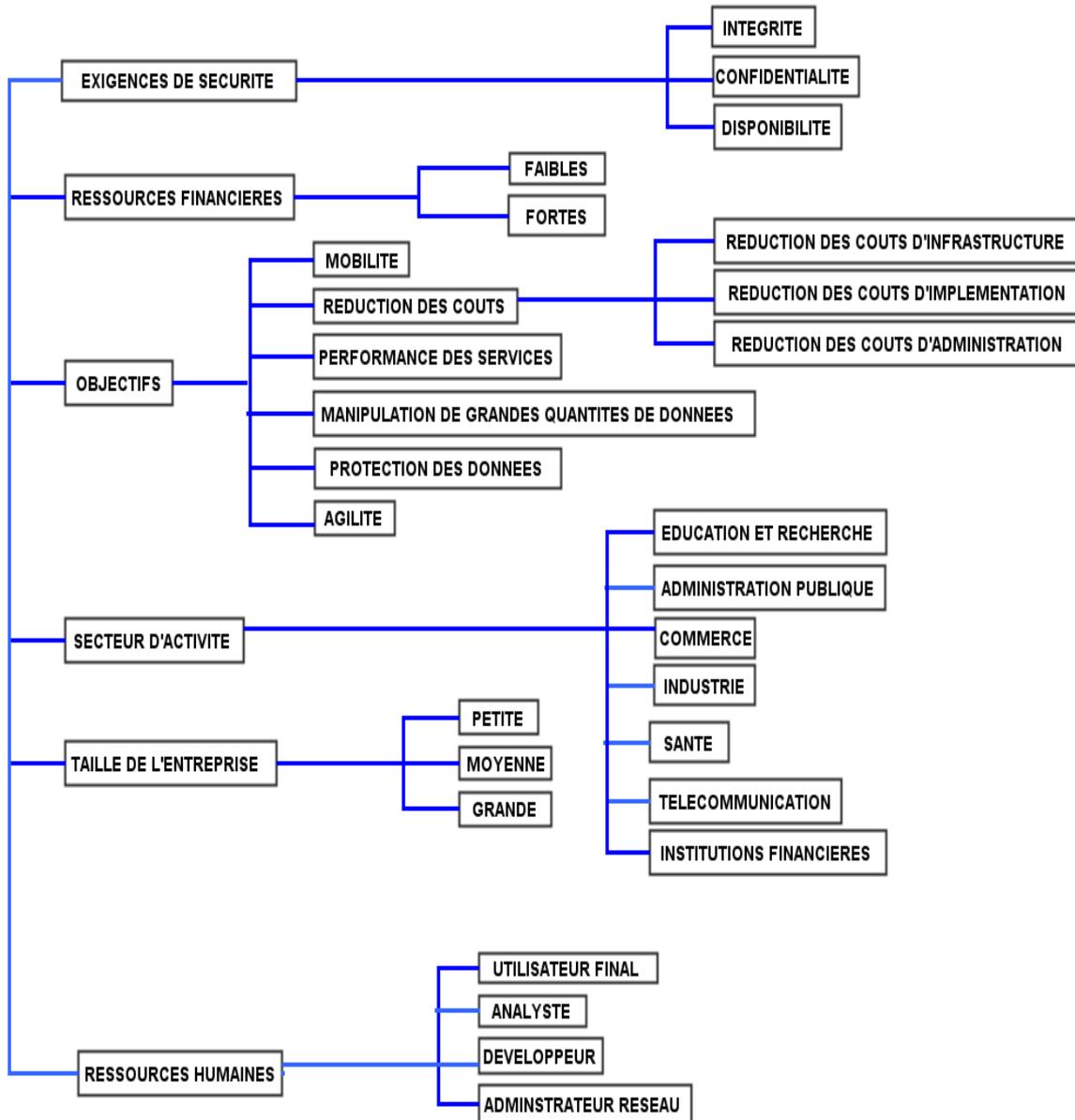


Figure 11 : Taxonomie de contexte

La taxonomie de contexte est constituée de six dimensions suivantes : *exigences de sécurité, objectifs, ressources financières, secteur d'activité, taille de l'entreprise et*

ressources humaines. L'ordre de classement des dimensions dans cette taxonomie n'a pas de signification particulière.

Les dimensions *exigences de sécurité*, *objectifs*, *ressources humaines* ont des caractéristiques inclusives : leurs caractéristiques ne forment pas une partition. Un objet peut cumuler les caractéristiques de l'une de ces dimensions. Tandis que les dimensions *ressources financières*, *secteur d'activité* et *taille de l'entreprise* ont des caractéristiques exclusives les unes des autres : leurs caractéristiques forment une partition, un objet ne peut donc pas cumuler les caractéristiques de l'une de ces dimensions. De manière détaillée, nous présentons ces dimensions ci-dessous.

5.3.1.1. Exigences de sécurité

Comme nous l'avons dit au point 2.5.3, par sécurité on entend la disponibilité, la confidentialité et l'intégrité. Ce sont ces trois critères que nous retrouvons ici. Il s'agit ici de choisir l'exigence (ou les exigences) de la sécurité sur laquelle on aimerait insister au cas où la migration vers le cloud est retenue. Notons que ce choix d'un critère doit tenir compte de la sensibilité des données à gérer et de bien d'autres critères tels que les ressources financières.

5.3.1.2. Ressources financières

Autrefois, la BI était réservée aux plus grandes entreprises qui possédaient des ressources financières importantes et les infrastructures nécessaires pour mettre en œuvre ce qui est souvent une technologie assez complexe. Dorénavant, le cloud place la BI à la portée des moyennes et petites entreprises qui, de manière générale, ont des ressources financières limitées et, par conséquent, ne peuvent pas consentir de gros investissements pour les infrastructures de la BI.

Certes, chaque service du cloud a un coût, mais de manière générale, les services sont payés à la consommation. C'est la raison pour laquelle les organisations peuvent choisir les services dont elles sont capables de supporter le coût. Il est, par exemple, conseillé aux organisations qui ont de faibles ressources financières de choisir la migration totale de leur BI vers le cloud car la migration partielle exigerait d'avoir en plus des infrastructures de la BI en local.

5.3.1.3. Objectifs

Pour déployer sa BI dans le cloud, une organisation doit avoir au moins un objectif précis. La migration peut être motivée par :

- La réduction des coûts (coûts d'infrastructure, d'implémentation ou d'administration),
- La performance des services qu'offre le cloud,
- Le besoin de manipuler les grandes quantités des données du cloud,
- La protection des données,
- L'agilité,
- La mobilité.

5.3.1.4. Secteur d'activité

Le secteur d'activité de l'organisation a un impact sur le choix des services de cloud. Il existe des secteurs dont les informations sont plus sensibles que d'autres. On trouve plusieurs secteurs d'activité. Nous citons entre autres :

- L'éducation et la recherche,
- L'industrie,
- Le commerce,
- La santé,
- Les institutions financières,
- Les télécommunications,
- L'administration publique.

Notons qu'en réalité, une entreprise peut appartenir à plusieurs secteurs d'activité, mais dans le cas précis de ces taxonomies, nous avons proposé que les règles pour la migration s'appliquent par secteur d'activité car chaque secteur a ses spécificités.

5.3.1.5. Taille de l'entreprise

Une entreprise peut être de grande, de moyenne ou de petite taille. Il est plus facile pour les petites et moyennes entreprises de déployer tous leurs composants BI dans le cloud que pour les grandes entreprises. Ces dernières doivent examiner avec soin chacun de leurs composants BI pour déterminer lesquels migrer vers le cloud. En outre, pour les petites et les moyennes entreprises qui n'ont pas de grands moyens financiers afin de mettre en place et maintenir la technologie BI, le cloud permet de réduire les coûts.

5.3.1.6. Ressources humaines

Dans la migration de la BI vers le cloud, le personnel (l'utilisateur) garde sa place primordiale car c'est lui qui manipule les autres ressources (matérielles et logicielles) pour interagir avec le cloud. Ainsi, dans le cadre de cette taxonomie, il s'agit pour une organisation de préciser les compétences dont elle dispose. Cela impacte les choix des services de cloud dont la BI peut bénéficier. On trouve principalement :

- Les utilisateurs finaux qui interagissent avec la couche SaaS. Ce sont de simples consommateurs. Toutes les ressources leur sont fournies et gérées par le cloud;
- Les développeurs et les analystes interagissent avec la couche PaaS. Les cadres et les architectures leur sont fournis et gérés par le cloud ainsi que toute l'infrastructure matérielle pour l'implémentation des applications selon les besoins de leurs organisations;
- Les administrateurs de réseau interagissent avec la couche IaaS. Il n'y a que les ressources matérielles qui leur sont fournies par le cloud.

Comme nous l'avons dit au point 5.3, nous avons développé deux taxonomies complémentaires : de contexte et de décision. Après avoir présenté la taxonomie de contexte, nous présentons ci-après, la taxonomie de décision.

5.3.2. Taxonomie de décision

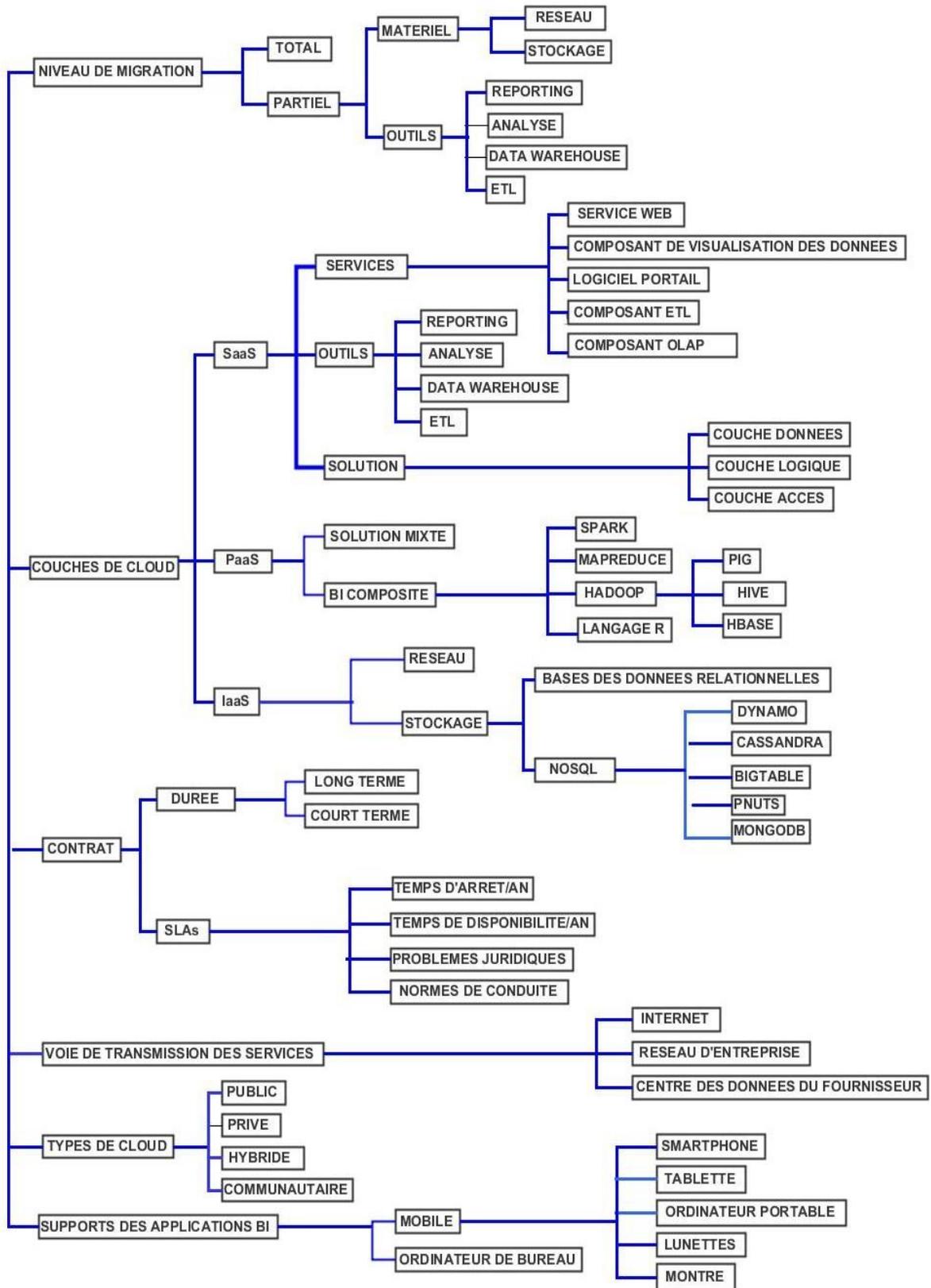


Figure 12: Taxonomie de décision

La taxonomie de décision comprend six dimensions : *niveau de migration, couches de cloud, contrat, voie de transmission des services, types de cloud* et *supports des applications BI*.

Les dimensions *niveau de migration, couches de cloud* et *supports des applications BI* ont des caractéristiques inclusives. Les dimensions *contrat, voie de transmission des services* et *types de cloud* ont des caractéristiques exclusives les unes des autres. Nous présentons ces dimensions ci-dessous.

5.3.2.1. Niveau de migration

Une organisation peut migrer une partie ou la totalité de ses tâches BI dans le cloud. En ce qui concerne la migration partielle, l'organisation est tenue à préciser le composant matériel ou logiciel qu'elle souhaite migrer dans le cloud. Pour le matériel, il peut s'agir du stockage ou du réseau. Pour les outils, il peut s'agir d'ETL, des outils d'administration de data warehouse et d'analyse ou de reporting. Il est conseillé aux petites et moyennes entreprises, qui n'ont pas de fortes ressources financières, de choisir une migration totale.

5.3.2.2. Couches de cloud

Chaque couche de cloud fournit des services spécifiques. Selon ses besoins et tenant compte de son contexte, une organisation peut solliciter l'une ou l'autre couche de cloud : SaaS, PaaS et IaaS.

La couche SaaS peut être sollicitée par une organisation de petite, de moyenne ou de grande taille. Ses utilisateurs finaux ne sont pas censés avoir une compétence particulière. Le cloud peut fournir à une organisation un service, un outil ou une solution de bout en bout. Du point de vue service, l'architecture de la BI se divise en trois couches (Baars and Kemper 2010):

- Couche données : le cloud développe pour une organisation une solution qui englobe la collecte et la consolidation des données ainsi que la modélisation et le stockage des données.
- Couche logique : le cloud s'occupe simplement de la partie analyse des données.
- Couche accès : le cloud fournit les applications clientes.

La couche PaaS est conseillée pour les organisations qui ont leurs propres analystes et des développeurs. Ceux-ci peuvent développer des applications et effectuer les tâches de BI dans le cloud.

La couche IaaS est conseillée pour des grandes entreprises qui veulent migrer seulement leurs réseaux dans le cloud. Cela nécessite d'avoir le personnel compétent pour l'administration des réseaux. La couche IaaS peut être aussi sollicitée pour toute autre entreprise qui désire conserver ses données dans le cloud.

5.3.2.3. Contrat

Le contrat entre le client et le fournisseur de cloud dépend de la durée estimée du service que le client sollicite. Il existe des contrats à court et à long terme, cette durée est précisée avant le début du service. Le fournisseur est censé déterminer la granularité du service à rendre dans un document souvent appelé SLA (service level agreement) (Baars and Kemper 2010). Ce document précise notamment la durée de disponibilité des services et la durée maximale d'interruption des services de cloud par an. Il décrit les normes de conduite entre client et fournisseur de cloud ainsi que les aspects juridiques.

5.3.2.4. Voie de transmission des services

Il existe trois voies de transmission de services de cloud, à savoir :

- Internet : Généralement, un fournisseur de cloud utilise l'Internet pour la transmission des services à ses différents clients. Dans ce cas, le fournisseur n'est pas connu de ses clients.
- Réseau d'entreprise : Un fournisseur de cloud peut mettre en place un réseau qui le lie avec ses clients et utilise ce réseau pour transmettre à ses clients les services qu'ils sollicitent. Il s'agit de relations entre plusieurs entreprises.
- Centre des données du fournisseur : Dans ce cas, il existe une relation directe entre le fournisseur et le client. Ce dernier est directement connecté au centre des données du fournisseur pour s'approvisionner des services dont il a besoin.

5.3.2.5. *Types de cloud*

Chaque fournisseur cloud est lié à un type de cloud. Choisir un fournisseur cloud revient à choisir un type de cloud bien déterminé (voir point 2.5.2). Le choix du fournisseur de cloud doit être motivé par les activités de l'organisation et la sensibilité de ses données.

5.3.2.6. *Supports des applications BI*

Les organisations ont besoin de dispositifs matériels pour interagir avec le cloud. Les principaux dispositifs sont des ordinateurs (desktop et laptop) et le mobile (les tablettes, les smartphones et montres) (Nayem, Fahad and Akhter 2013).

5.4. Conclusion

Dans ce chapitre, nous avons construit, de manière itérative, deux taxonomies complémentaires pour la migration de la BI vers le cloud. La première est dédiée aux éléments qui influencent la décision de migration et la seconde se rapporte à l'objet même de la migration. Ces taxonomies résultantes ont révélé que ce sont les sources de type 1 qui ont apporté les principales dimensions. Les sources de type 2 n'ont pas apporté beaucoup de modifications dans la taxonomie parce que ce type est constitué d'articles qui sont généralement cités dans les sources de type 1. Les sources de type 3, quant à elles, ont apporté beaucoup plus de caractéristiques que de dimensions ou de catégories. Pour valider ces taxonomies, nous avons constitué un autre échantillon des sources de type 1 en changeant un des critères d'inclusion des articles dans l'échantillon : l'article devrait utiliser une autre approche que la science de conception. L'examen des sources de ce dernier échantillon a apporté très peu de nouvelles informations dans la taxonomie telle qu'élaborée auparavant. Il a néanmoins permis de valider les dimensions, les catégories ainsi que les caractéristiques qui existaient déjà dans la taxonomie.

En outre, pour chaque type de sources, nous avons présenté un graphique du nombre d'opérations effectuées. Ces graphiques se trouvent en annexe (ANNEXE 2). Tous ces graphiques ont illustré comment les courbes convergent quand on arrive vers la fin du processus. Cette convergence illustre le fait qu'en général, les articles proposent les mêmes éléments en ce qui concerne la migration de la BI vers le cloud. Ces éléments sont collectés au début du processus. Plus on avance vers la fin, moins l'examen des sources apporte de modifications dans la taxonomie car les sources confirment les éléments collectés

précédemment et, par conséquent, le nombre d'opérations effectuées à chaque itération tend vers zéro.

Grâce aux principes de la revue systématique de littérature, nous avons constitué un échantillon raisonnable d'articles de qualité. Les éléments tirés de ces articles héritent ainsi de la qualité des sources dont ils proviennent. En définitive, ces deux taxonomies complémentaires contiennent un grand nombre d'informations pertinentes pour la migration de la BI vers le cloud. Cependant, avec l'évolution rapide des technologies complexes émergentes, la pérennité de ces taxonomies est relative. Néanmoins, la méthodologie que nous avons proposée pour développer ces taxonomies reste un atout qui permettra de construire de nouvelles taxonomies adaptées aux éventuelles évolutions.

Le chapitre 6 de cette thèse exploite ces deux taxonomies sous forme de règles grâce au modèle de guidage opérationnel proposé au point 3.5 pour aider les organisations désireuses de migrer leur BI vers le cloud à effectuer le choix des options du cloud adaptées à leur contexte.

CHAPITRE 6: AIDE A LA DECISION DE MIGRATION DE LA BUSINESS INTELLIGENCE VERS LE CLOUD

6.1. Introduction

La décision de migrer la BI vers le cloud mérite d'être bien étudiée car cette migration ne procure pas simplement des avantages, mais elle implique aussi des défis à relever. Il est donc de bon aloi que les organisations, avant de s'engager, aient une information suffisante en ce qui concerne la migration de la BI vers le cloud. Ce chapitre présente un ensemble de règles permettant aux organisations désireuses de migrer leur BI vers le cloud d'effectuer les choix (des services de cloud) relatifs à leur contexte. Cet ensemble de règles constitue une méthode de guidage opérationnel pour la migration de la BI vers le cloud. Il est composé de règles issues de la littérature ou obtenues par data mining (règles d'association). Ces règles s'appuient sur les deux taxonomies complémentaires développées au chapitre précédent (taxonomie de contexte et taxonomie de décision).

Grâce aux directives déterminées au point 3.7, nous vérifions la cohérence, la minimalité et la complétude de notre base de règles.

6.2. Les règles

Une règle est une instruction ou commande qui s'applique dans certaines situations (Friedman-Hill 2003). Dans le cadre de la migration de la BI vers le cloud, les règles sont constituées par des éléments de la taxonomie de contexte et de la taxonomie de décision. Comme nous l'avons dit au point 3.5, une règle se présente généralement sous la forme ci-dessous :

Si "contexte"

Alors "décision".

On peut aussi avoir des règles de la forme :

Si "contexte"

Alors "contexte"

Ou

Si "décision"

Alors "décision"

Afin de constituer notre base de règles, nous avons examiné à nouveau l'échantillon d'articles qui ont servi pour le développement des deux taxonomies complémentaires (de contexte et de décision). Les questions étaient les suivantes :

- Dans un contexte précis, à quel service de cloud peut-on recourir ?
- Un élément de contexte peut-il avoir de l'incidence sur un autre ?
- Une décision peut-elle influencer la prise d'une autre décision ?

Après ce réexamen de la littérature à notre disposition, nous avons constaté que certains auteurs énoncent clairement, sous forme de propositions, les éléments qui peuvent amener à l'adoption d'un service de cloud bien précis. Cependant, leurs propositions ne sont pas absolues mais sont plutôt des indicateurs précieux dans le choix des services cloud.

Exemple : Il est préférable aux petites et moyennes entreprises de migrer totalement leur BI vers le cloud car elles ne possèdent pas de gros moyens d'investissements pour couvrir les coûts relatifs à l'infrastructure, à l'implémentation et à l'administration de la BI en local (Al-Aqrabi *et al.* 2015).

Dans un tel cas, nous pouvons retenir cette règle :

Si *taille de l'entreprise = petite ou moyenne*

et Ressources financières = faibles

Alors *niveau de migration = total.*

Cependant, cette règle n'est pas absolue car on peut retrouver une entreprise de taille moyenne qui a une partie de sa BI en local et une autre partie dans le cloud.

Néanmoins, cette règle permet aux petites et moyennes entreprises, avec des moyens financiers limités, de savoir qu'elles peuvent néanmoins profiter des potentialités du cloud pour exploiter la technologie BI sous cette condition.

Par ailleurs, lorsque, dans un article, on décrit un service de cloud qui est un élément de la taxonomie de décision en mettant en exergue les avantages que ce service procure, nous en déduisons une ou plusieurs règles car ces avantages constituent, de manière générale, ce que les organisations poursuivent pour accéder au service de cloud décrit. Par exemple, on décrit le cloud public comme un type de cloud qui permet d'effectuer des économies car les

organisations ne paient que ce dont elles ont besoin, ce qui réduit le coût total de leur système de BI en réduisant les frais généraux (Kasem and Hassanein 2014).

Nous savons que notre taxonomie de contexte (voir point 5.3.1) comporte une dimension *objectifs* qui regroupe les objectifs de la migration de la BI vers le cloud d'une part. D'autre part, dans la taxonomie de décision (voir point 5.3.2), nous avons une dimension *types de cloud* qui regroupe les différents types de cloud. Nous pouvons donc déduire cette règle :

Si *objectifs* = *réductions des coûts*

Alors *type de cloud* = *public*.

Ces deux règles ne sont pas au même niveau, la première est énoncée clairement et la deuxième est déduite d'un énoncé. Ainsi, sur la base de la littérature, nous distinguons deux types de règles : celles proposées par la littérature et celles déduites de la littérature.

Par souci d'exhaustivité, nous avons mis en place un troisième type de règles : les règles d'association. Ces règles sont obtenues par data mining.

6.2.1. Les règles proposées par la littérature

Les règles proposées par la littérature sont celles énoncées explicitement dans les articles de notre échantillon.

Pour chaque article lu, nous avons relevé les services cloud proposés (décision) et les éléments qui peuvent permettre d'accéder à ces services (contexte) : Cela nous a conduit à mettre en place des règles de la forme :

Si "contexte"

Alors "décision".

Nous avons aussi relevé les décisions qui conduisent à d'autres décisions et les contextes qui impliquent d'autres contextes pour élaborer respectivement des règles de la forme :

Si "décision"

Alors "décision".

Et

Si "contexte"

Alors "contexte".

Rappelons qu'une règle est constituée d'une prémisse et d'une conclusion. Une prémisse et une conclusion sont chacune composée d'expressions. Chaque expression comprend deux parties : gauche et droite séparées par un signe d'égalité. La partie gauche se réfère à une dimension de la taxonomie (de contexte ou de décision) et la partie droite se réfère soit à une caractéristique soit une catégorie.

Nous avons obtenu vingt-huit règles proposées directement dans les articles. Dans les lignes qui suivent, nous présentons, à titre d'illustration, dix de ces règles. Le reste des règles est fourni en annexe (ANNEXE 3).

1. **Si** *objectifs= réduction des coûts/ réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration*
et Taille de l'entreprise = petite ou moyenne

Alors *niveau de migration = total*

Les coûts relatifs à l'infrastructure, à l'implémentation et à l'administration de la BI ne permettent pas de manière générale aux petites et moyennes entreprises d'exploiter cette technologie. Le cloud vient lever cette barrière en donnant la possibilité aux entreprises de payer les services à la consommation. Ainsi, il est conseillé aux petites et moyennes entreprises d'exploiter totalement leur BI dans le cloud car garder une partie de la BI en local nécessiterait des coûts supplémentaires. (Al-Aqrabi *et al.* 2015).

2. **Si** *ressources humaines = analyste*
Alors couches de cloud = PaaS
3. **Si** *ressources humaines = développeur*

Alors couches de cloud = PaaS

Un analyste est considéré comme l'utilisateur typique qui peut interagir avec la couche PaaS. Par ailleurs, les talents de développeurs experts peuvent être mis en évidence afin d'exploiter efficacement l'infrastructure basée sur le cloud (Patel, Rau-Chaplin and Varghese 2012). Pour une organisation, avoir du personnel analyste ou développeur, lui donne la possibilité de solliciter la couche PaaS. Les organisations

peuvent donc migrer leur BI vers le cloud en capitalisant sur les compétences de leur personnel.

4. **Si taille de l'entreprise = moyenne**

Alors types de cloud = public

Les entreprises de taille moyenne sont plus intéressées par le cloud public car, le cloud privé est plus onéreux (Kaur, Agrawal and Dhiman 2012).

5. **Si exigences de sécurité = intégrité et confidentialité**

Alors types de cloud = privé

6. **Si exigences de sécurité = intégrité, confidentialité et disponibilité**

Alors type de cloud = hybride

La plupart des organisations professionnelles ont des données financières et d'autres données sensibles. L'utilisation du cloud public n'est alors pas un bon choix car, dans le cloud public, les mêmes ressources matérielles et logicielles sont partagées entre tous les clients. Cela constitue un risque pour la confidentialité et l'intégrité des données. C'est pour cette raison que les experts suggèrent d'utiliser un cloud privé quand il s'agit de données sensibles ou financières. En revanche, le cloud hybride combine les avantages du cloud privé et ceux du cloud public : les clients peuvent développer et déployer des applications analytiques à l'aide d'un environnement privé pour garantir la confidentialité et l'intégrité, bénéficiant ainsi d'un degré de disponibilité de ressources plus élevé que dans un cloud public (Assunção *et al.* 2014).

7. **Si niveau de migration = partiel/matériel**

Alors couches de cloud = IaaS

8. **Si niveau de migration = partiel/outils**

Alors couches de cloud = SaaS

Une organisation peut choisir de migrer totalement ou partiellement sa BI vers le cloud. Dans le cas de la migration partielle, elle doit préciser les composants BI qu'elle souhaite migrer vers le cloud : outils ou matériel. Si la migration concerne les composants matériels, c'est la couche IaaS qui est sollicitée. Dans ce cas, la virtualisation permet aux entreprises d'accéder à des ressources matérielles haut de gamme qui, par ailleurs, seraient trop coûteuses pour qu'elle puisse se les procurer. Si la migration concerne les outils DWH, ETL et outils d'analyse fournis en tant que

service, c'est la couche SaaS qui est sollicitée (Muriithi and Kotzé 2013) (Herwig 2013) (Baars and Kemper 2010).

9. **Si** *objectifs = performance des services*

Alors *types de cloud = privé*

Le cloud privé offre plus de performance que le cloud public (Kasem and Hassanein 2014). Une organisation qui désire migrer vers le cloud pour profiter de la performance des services que procure le cloud peut solliciter un type de cloud privé. Cependant, il vaut mieux savoir que les prix des services sont généralement fixes pour le cloud privé, contrairement au cloud public où les services sont payés à la consommation.

10. **Si** *objectifs = réduction des coûts*

Alors *couches de cloud = SaaS*

Le modèle SaaS est bien étudié pour permettre aux organisations de s'approvisionner en applications avec un coût réduit et en fonction de leur utilisation (Sun *et al.* 2012). Pour bénéficier du SaaS, les organisations n'ont pas besoin de déboursier les frais en vue de se procurer du matériel spécifique. L'utilisation de SaaS ne nécessite pas de compétences particulières. Le cloud permet donc aux organisations d'exploiter la BI avec des coûts réduits par rapport à la BI en local.

6.2.2. Les règles déduites de la littérature

La compréhension du contenu d'un article peut amener à déduire une ou plusieurs règles. Grâce à la lecture attentive des articles de notre échantillon, nous avons pu déduire seize règles dont les cinq suivantes (le reste des règles se trouve en annexe) : (ANNEXE 3).

1. **Si** *objectifs = réduction des coûts/réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration*

et Exigences de sécurité = disponibilité

Alors *types de cloud = public*

2. **Si** *objectifs = performance des services*

et Exigences de sécurité = intégrité et confidentialité

Taille de l'entreprise = grande

Alors *types de cloud = privé*

En général, le cloud public garantit plus de disponibilité des ressources et utilise un mode de paiement sur mesure : on ne paie que les services consommés, ce qui occasionne une réduction des coûts par rapport à la BI en local. Tandis que le cloud privé garantit beaucoup plus de sécurité et de performance, le prix est généralement fixe (Kasem and Hassanein 2014). Ainsi, le cloud privé est conseillé pour les grandes entreprises ayant des moyens importants pour investir.

3. **Si** *objectifs = performance des services*

et Contrat = court terme

Alors *voie de transmission des services = internet*

et Couches de cloud à solliciter = SaaS/service

La durée du contrat est relative au service sollicité. Par exemple pour un composant, la durée est de court terme à cause de sa simplicité. Par ailleurs, la performance des services est une des principales raisons qui poussent les organisations à migrer leur BI vers le cloud. Ainsi, un composant permet d'augmenter la performance d'une application et il est généralement fourni via l'internet (Baars and Kemper 2010).

4. **Si** *objectifs = (réduction des coûts/réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration) et objectifs = Performance des services*

Alors *contrat = long terme*

et Voie de transmission des services = centre des données du fournisseur

et Couches de cloud = SaaS/outils

Pour un outil, le contrat peut être à long terme vu sa complexité par rapport à un composant. Un client peut se connecter au centre de données du fournisseur pour s'approvisionner en services. Par ailleurs, la réduction des coûts est une des principales raisons qui poussent les organisations à migrer leur BI vers le cloud (Baars and Kemper 2010).

5. **Si** *objectifs = réduction des coûts/réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration et objectifs = Performance des services*

Alors contrat = long terme

et Voie de transmission des services = centre des données du fournisseur

et Couches de cloud = SaaS/solution

Pour une solution, la durée du contrat est sur le long terme car une solution de bout en bout est plus complexe qu'un service ou un composant. Un client peut se connecter au centre de données du fournisseur pour s'approvisionner en services. Comme nous l'avons déjà dit, la réduction des coûts est une des principales raisons qui poussent les organisations à migrer leur BI vers le cloud (Baars and Kemper 2010).

6.2.3. Les règles d'association

A partir des règles issues des articles de notre échantillon, nous avons généré automatiquement un troisième type de règles par une technique de data mining : les règles d'association.

Notre objectif est de rendre la base de règles la plus exhaustive possible : chaque contexte doit correspondre à une décision. Ainsi, les règles d'association sont un complément utile car, grâce à elles, nous pouvons ajouter de nouvelles règles dans la base. Ces règles peuvent contribuer à l'exhaustivité de la base de règles.

Ainsi, nous avons créé un fichier Excel avec deux colonnes dont la première comprend les numéros des articles de notre échantillon. La seconde colonne comprend les différentes règles contenues dans chaque article présentées sous forme des expressions (partie gauche = partie droite). Chaque ligne représente donc un article avec les différentes expressions issues dudit article.

A l'aide du langage R et dans son interface graphique Rattle, nous avons mis en place les règles d'association. Nous avons chargé notre fichier Excel grâce à l'onglet "données" et la commande "exécuter" (Williams 2011).

Deux types d'analyse sont donc pris en charge et le type approprié est choisi à l'aide du bouton de sélection "Paniers". Quand ce bouton est coché, Rattle utilise la méthode d'analyse de panier. Si ce bouton n'est pas coché, Rattle trouve des relations intéressantes entre les éléments des données sans utiliser la notion de panier.

Pour une analyse de panier, on considère deux variables : Identificateur (Ident) et cible. La variable Ident représente le numéro du panier et chaque panier ne peut avoir qu'un seul numéro. La variable cible représente les éléments relatifs à chaque panier. Un panier peut comporter plusieurs éléments. Dans notre cas, un panier correspond à un article. La première colonne du fichier Excel est une variable identificateur et la deuxième colonne est une variable cible.

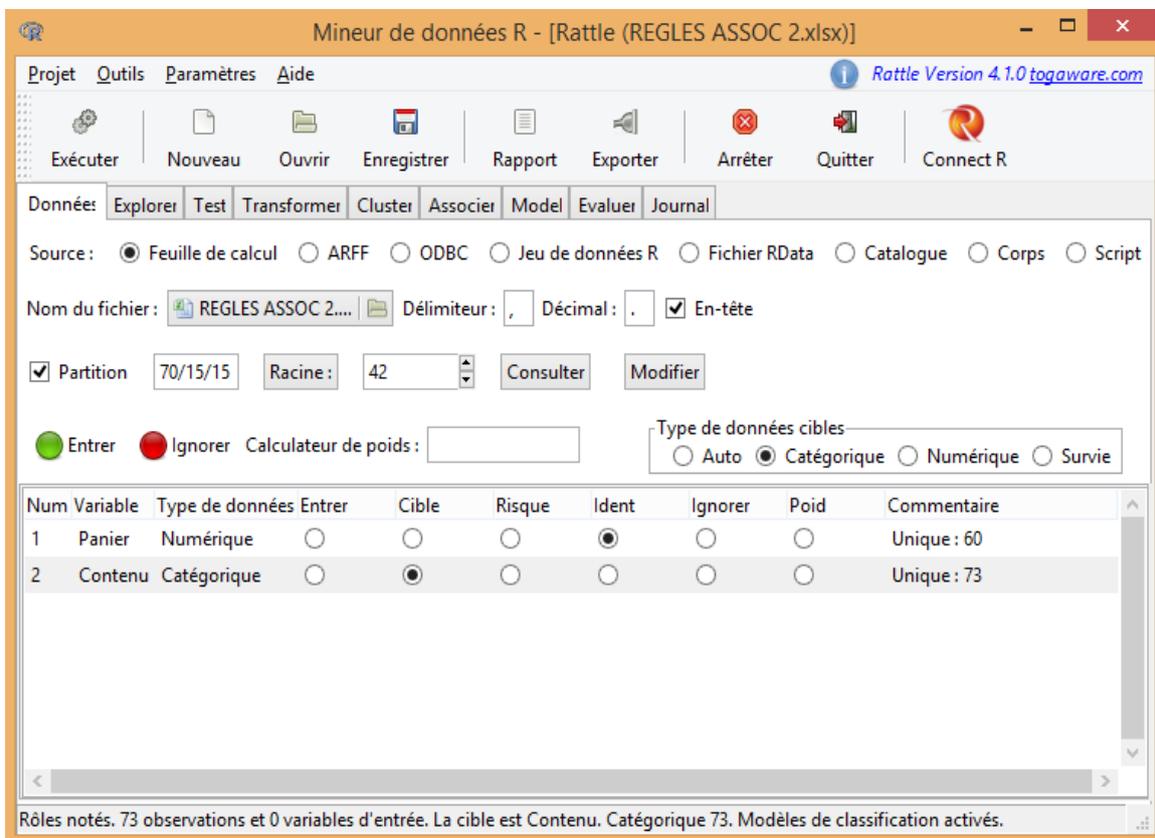


Figure 13 : Interface Rattle

Après avoir chargé notre fichier Excel dans Rattle en utilisant l'onglet "Données", nous avons analysé ces expressions pour trouver des éventuelles règles d'association à l'aide de l'onglet "Associer" et la commande "Exécuter".

Panier	Contenu
1	Objectif = réduction des coûts, Ressources financières = faibles, Taille de l'entreprise = petite ou moyenne, Sécurité = disponibilité, Niveau de migration = total, Couche cloud = SaaS/solution, Moyen de transmission des services = internet.
2	Objectif = réduction des coûts, Sécurité= disponibilité, Type cloud = public. Objectif = performance des services, Sécurité = intégrité et confidentialité, Type cloud = privé.
3	Objectif = performance des services, Contrat = court terme, Moyen de transmission des services = internet, Couche cloud = SaaS/service. Objectif = réduction des coûts et performance des services, Contrat = long terme, Moyen de transmission des services = centre des données du fournisseur, Couche cloud à solliciter = SaaS/outils. Objectif = réduction des coûts et performance des services, Contrat = long terme, Moyen de transmission des services = réseau d'entreprise, Couche cloud = IaaS/réseau. Objectif = réduction des coûts et performance des services, Contrat = long terme, Moyen de transmission des services = internet, Couche cloud = PaaS/meilleure solution.
4	Objectif = réduction des coûts et manipulation de grandes quantités des données, Secteur d'activités = éducation et recherche, Taille de l'entreprise = petite ou moyenne, Niveau de migration = partiel/hardware/stockage et partiel/outils, Moyen de transmission des services = internet.
5	Objectif = réduction des coûts, Sécurité = disponibilité, Taille de l'entreprise = petite ou moyenne, Couche cloud = SaaS/solution.
7	Niveau de migration = partiel/outils/analyse, Couche cloud = PaaS/BI composite.
8	Niveau de migration = partiel/hardware/stockage ou partiel/outils/DWH et Analyse, Couche cloud = IaaS/stockage.
9	Ressources humaines = analyste et/ou développeur, Couche cloud = PaaS/BI composite.
10	Objectif = performance des services, Sécurité = confidentialité, Domaine d'activités = institutions financières, Type cloud = privé.

Figure 14 : Extrait du fichier Excel

Les deux principales métriques utilisées dans l'évaluation des règles d'association sont le support et la confiance (Williams 2011). Un support d'une collection d'items est le nombre de fois que certains items apparaissent toujours ensemble pendant toutes les transactions. La valeur recommandée pour le support varie entre 0,1 et 0,4 (Williams 2011). La confiance minimale est exprimée en proportion du nombre total de transactions dans l'ensemble de données : c'est le nombre de fois que l'item "x" apparaît chaque fois qu'un autre item "y" apparaît dans une transaction. La valeur souhaitable de la confiance peut varier entre 0,83 et 10 (Williams 2011).

Cette technique de data mining a généré onze règles dont nous présentons les six suivantes à titre illustratif (le reste des règles se trouve en annexe) (ANNEXE 3)

1. **Si exigences de sécurité = intégrité et confidentialité**
Alors types de cloud = privé
2. **Si exigences de sécurité = disponibilité**
Alors types de cloud = public
3. **Si exigences de sécurité = intégrité, confidentialité et disponibilité**
Alors types de cloud = hybride

Dans ces trois premières règles, c'est la dimension *exigences de sécurité* qui est mise en exergue : dans le cloud privé, les exigences de la sécurité qui sont exploitées sont l'intégrité et la confidentialité. Dans le cloud public, c'est plutôt l'aspect disponibilité des ressources qui est souligné. Le cloud hybride regroupe les avantages du cloud privé et ceux du cloud public.

4. **Si objectifs = performance des services**
Alors contrat = court terme
et Voie de transmission des services = internet
et Couches de cloud = SaaS/service
5. **Si objectifs = réduction des coûts et performance des services**
Alors contrat = court terme
et Voie de transmission des services = internet
et Couches de cloud = PaaS/BI composite
6. **Si objectif = réduction des coûts et performance des services**
Alors contrat = long terme
et Voie de transmission des services = réseau d'entreprise
et Couches de cloud = IaaS/réseau

Ces trois dernières règles tournent autour de la durée du contrat entre le fournisseur et le client. Par exemple pour une organisation qui sollicite la couche IaaS/réseau, le contrat est généralement à long terme. Pour la règle 4, le contrat est à court terme car il s'agit d'un composant (voir point 6.2.2). Les règles 5 et 6 pourraient paraître contradictoires, mais il n'en est pas question, la différence se trouve au niveau des couches de cloud à solliciter. Dans la règle 5, c'est la couche PaaS qui est sollicitée. Les développeurs ou les analystes construisent

une solution sur mesure pour leur organisation en utilisant les ressources du cloud. Cela prend généralement un temps plus court que lorsqu'une organisation migre son réseau vers le cloud (Baars and Kemper 2010). Ainsi, deux organisations peuvent avoir le même objectif de migration, mais la durée du contrat dépend du service sollicité.

6.3. Cohérence, minimalité et exhaustivité de la base de règles

Une base de règles doit être cohérente, minimale et exhaustive. Dans cette section, nous appliquons les directives proposées au point 3.7 au cas particulier de la base de règles pour la migration de la BI vers le cloud.

6.3.1. Cohérence

En ce qui concerne la cohérence, aucune règle ne comprend une prémisse issue de la taxonomie de décision et une conclusion issue de la taxonomie de contexte.

Exemple :

Si *niveau de migration = total*

Alors *taille de l'entreprise = petite.*

Notre base de règles ne comporte pas de règles de cette forme (si décision, alors contexte).

Dans la conclusion d'une règle, les connecteurs "ou" et "non" ne sont pas employés.

Exemple :

Si *objectifs = réduction des coûts*

Alors *niveau de migration = total ou partiel/outils*

Une conclusion doit être toujours précise : on ne sait pas ici dans quel cas choisir quelle partie droite (total ou partiel/outils). Nous n'avons pas de telles règles dans notre base.

Les deux parties d'une expression (partie droite et partie gauche) doivent être issues de la même taxonomie (de contexte ou de décision).

Exemple :

Si *objectifs = BI/composite*

Alors *couche cloud = PaaS*

Dans la prémisse de cette règle, l'expression est composée de deux parties issues de deux différentes taxonomies (contexte et décision). La partie gauche se réfère à une dimension de

la taxonomie de contexte : *objectifs*. Tandis que la partie droite se réfère à une catégorie de la taxonomie de décision : BI/composite (voir point 5.3). Notre base ne contient pas de règle de cette forme.

La partie droite d'une expression doit se référer à la caractéristique ou à la catégorie de la dimension à laquelle se réfère sa partie gauche.

Exemple :

Si *objectifs* = *confidentialité*

Alors *type cloud* = *privé*

Dans la taxonomie présentée au point 5.3.1, la *confidentialité* est une caractéristique de la dimension *exigences de sécurité*. Elle ne peut donc pas être la partie droite de cette expression dont la partie gauche se réfère à la dimension *objectifs*.

Une partie gauche d'une expression ne peut pas se référer à une catégorie.

Exemple :

Si *réduction des coûts* = *réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration*

Alors *niveau de migration* = *total*

Ici, la prémisse n'est pas correcte car la partie gauche de l'expression de la prémisse se réfère à une catégorie. Par construction, notre base ne comporte pas de règle de cette forme.

Deux règles ne peuvent pas avoir une même prémisse et des conclusions contradictoires.

Exemple

Si *réduction des coûts* = *réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration*

Alors *niveau de migration* = *total*

Si *réduction des coûts* = *réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration*

Alors *niveau de migration* = *partiel*

Par construction, notre base ne comporte pas de règles contradictoires. En effet, au vu du nombre relativement faible de règles, la vérification a été faite manuellement. Dans le cas de

contradiction entre deux règles, nous avons supprimé l'une et gardé l'autre, celle qui est revenue plusieurs dans les articles de notre échantillon.

6.3.2. Minimalité

En ce qui concerne la minimalité, dans notre base de règles, chaque règle est écrite une seule fois.

Si une conclusion d'une règle issue de la taxonomie de décision est en même temps une prémisse d'une autre règle, par souci minimalité, nous fusionnons ces deux règles. La règle issue de la fusion prend la prémisse de la première règle et sa conclusion sera constituée de la concaténation des conclusions de ces deux règles fusionnées.

Exemple :

Si taille de l'entreprise = petite

Alors niveau de migration = total

Si niveau de migration = total

Alors contrat = long terme

La règle issue de la fusion de deux précédentes règles est la suivante :

Si taille de l'entreprise = petite

Alors niveau de migration = total

et Contrat = long terme

Etant donné que dans notre base de règles toutes les prémisses issues de la taxonomie de décision sont des conclusions pour d'autres règles, cette directive supprime de notre base de règles toutes les règles de la forme :

Si décision

Alors décision

L'exemple ci-haut l'illustre.

Si deux règles possèdent une prémisse identique et que cette dernière est issue de la taxonomie de contexte et si leurs conclusions sont différentes et issues l'une de la taxonomie de contexte et l'autre de la taxonomie de décision, nous fusionnons ces deux règles. La prémisse de la nouvelle règle sera constituée de toutes les expressions de ces deux règles qui

sont issues de la taxonomie de contexte et la conclusion, de toutes les expressions de ces deux règles qui sont issues de la taxonomie de décision.

Exemple :

Si exigences de sécurité = intégrité et confidentialité

Alors ressources financières = fortes

Si exigences de sécurité = intégrité et confidentialité

Alors type cloud = privé

La règle fusionnée se présente comme suit :

Si exigences de sécurité = intégrité et confidentialité

et Ressources financières = fortes

Alors type cloud = privé

Cette directive supprime de notre base de règles toutes les règles de la forme :

Si contexte

Alors contexte

L'exemple ci-haut le démontre.

Pour deux expressions qui ont des parties gauches identiques et dont les parties droites sont l'une concrète et l'autre abstraite, on supprime l'expression qui a la partie droite concrète et on garde celle qui a la partie droite abstraite.

Exemple :

Si niveau de migration = partiel/outils (partie droite abstraite)

Alors couches de cloud = SaaS/outils (partie droite abstraite)

Si niveau de migration = partiel/outils/analyse (partie droite concrète)

Alors couches de cloud = SaaS/outils/analyse (partie droite concrète).

La deuxième règle est supprimée.

Nous supprimons toute règle dont les expressions de la partie gauche et les expressions de la partie droite sont toutes incluses dans une autre règle.

Exemple :

Si taille de l'entreprise = petite ou moyenne

Alors niveau de migration = total

et Couches de cloud = SaaS/outils

Si taille de l'entreprise = petite

Alors niveau de migration = total

La deuxième règle est supprimée.

6.3.3. Exhaustivité

Notre base de règles ne possède pas un grand nombre d'éléments dans les deux taxonomies. Pour vérifier son exhaustivité, nous avons effectué différentes combinaisons entre les contextes et les décisions. Nous avons rapproché chaque dimension de la taxonomie de contexte à toutes les dimensions de la taxonomie de décision. Cela nous a conduits à ajouter six nouvelles règles dans la base : des combinaisons qui n'existaient pas avant cette opération.

A titre illustratif, nous avons rapproché les objectifs des types de cloud. Le constat est que chaque objectif influence le choix d'un type de cloud approprié, sauf pour l'objectif *protection de données*. Ce qui nous a conduits à ajouter la règle ci-dessous :

Si objectifs = protection des données

Alors types de cloud = hybride

Le choix de cloud hybride est motivé par le fait que dans une organisation, on trouve généralement des données qui sont sensibles et d'autres qui ne le sont pas. Un cloud hybride répond bien à ce critère car il renferme les avantages du cloud privé et ceux du cloud public. Quand une organisation migre vers le cloud pour protéger ses données, elle aimerait en disposer chaque fois qu'elle en a besoin. L'aspect disponibilité entre en ligne de compte. C'est le cloud public qui répond bien à ce critère de disponibilité. La protection des données implique aussi l'intégrité et la confidentialité. Ces deux aspects sont, de manière générale, garantis par un cloud privé.

6.4. Conclusion

Dans ce chapitre, nous avons présenté un ensemble des règles issues de la littérature et des règles d'association obtenues par data mining. Les règles d'association n'ont pas apporté un grand changement dans la base des règles issues de la littérature : sept sur les onze règles produites se trouvaient déjà dans la base. Pour vérifier la cohérence, la minimalité et la complétude de notre base de règles, nous avons utilisé les directives mise en place au point 3.7 de cette thèse. Ces dernières nous ont permis de vérifier la cohérence, la minimalité et l'exhaustivité de notre base de règles.

Après l'application de ces directives, le constat est que toutes les règles sont cohérentes les unes par rapport aux autres et toutes par rapport au modèle de guidage opérationnel. Par souci de minimalité, nous avons supprimé et fusionné certaines règles. Pour des raisons d'exhaustivité, nous avons ajouté d'autres règles dans la base. Notre base de règles est cohérente, minimale et l'exhaustive à un moment donné. Elle reste évolutive et son exhaustivité est questionnée à chaque mise à jour.

Par ailleurs, grâce aux exigences de minimalité, notre base de règles ne comporte plus que des règles de la forme :

Si contexte

Alors conclusion.

Dans le chapitre suivant nous ferons des vérifications complémentaires de complétude et de cohérence de notre base de règles.

Afin de faciliter l'utilisation de ces règles et de valider notre approche, nous avons développé un outil. Ce dernier permettra de renseigner le contexte d'une organisation désireuse de migrer sa BI vers le cloud afin qu'une décision adaptée lui soit proposée. Le chapitre suivant traitera de l'implémentation de cet outil et de la validation de notre approche.

CHAPITRE 7 : IMPLEMENTATION ET VALIDATION

7.1. Introduction

A partir des règles présentées au chapitre précédent, nous avons développé un système de guidage opérationnel pour la migration de la BI vers le cloud. Un système fondé sur des règles est généralement destiné à remplacer l'expertise humaine. Ce système tire des conclusions pour des prémisses énoncées au préalable (Friedman-Hill 2003).

Nous avons utilisé JESS (Java Expert System Shell) (Friedman-Hill 2003) en interaction avec Java sous l'environnement ECLIPSE pour le développement de cet outil de guidage opérationnel de la migration de la BI vers le cloud. JESS est un générateur système expert écrit complètement en Java. Il est aussi un environnement de script à partir duquel on peut créer des objets et appeler des méthodes Java sans avoir besoin de compilation. Grâce à JESS, on peut développer en Java une application qui a la capacité de raisonner en utilisant les connaissances que l'on possède sous forme de règles déclaratives. JESS est l'un des moteurs de règles les plus simples et les plus rapides.

Dans ce chapitre, nous présenterons l'architecture de JESS ainsi que l'architecture de notre outil. En vue d'illustrer l'utilisation de cet outil, nous mettons en place un diagramme de cas d'utilisation. Pour montrer l'utilité et l'efficacité de l'outil que nous avons développé, nous présentons un scénario illustratif. Enfin, nous procéderons à la validation de notre approche.

7.2. Architecture de JESS

Comme nous l'avons dit ci-haut, nous avons utilisé JESS pour mettre en place un système basé sur des règles. Ci-après, nous présentons l'architecture de JESS.

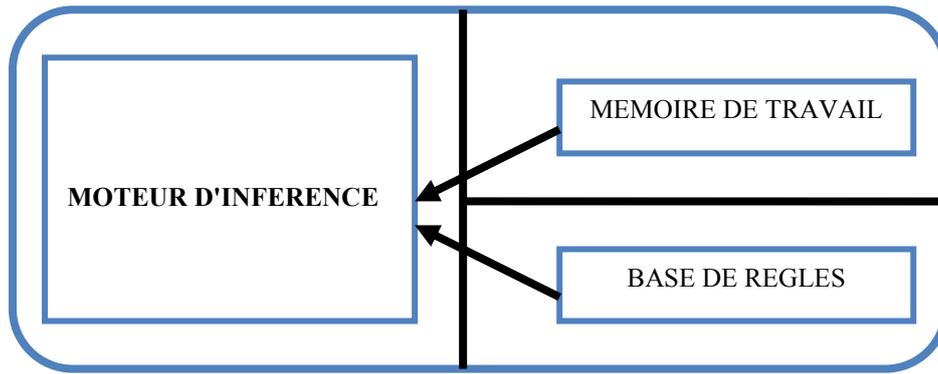


Figure 15 : Architecture de JESS adapté et traduit de Friedman-Hill (2003)

Cette architecture comprend les trois grands composants suivants :

- Une mémoire de travail : C'est une base de faits. Elle comprend tous les faits qui sont utilisés pour former les prémisses et les conclusions des règles.
- Une base de règles : C'est l'ensemble de règles qu'un système peut utiliser.
- Un moteur d'inférence : C'est le composant qui se charge d'exécuter les règles.

Dans la section suivante, nous expliquerons en détail le fonctionnement de chaque composant.

7.3. Architecture de l'outil de guidage opérationnel

Nous présentons ici la structure générale de notre outil de guidage opérationnel, l'organisation de ses composants et les liens entre ses composants.

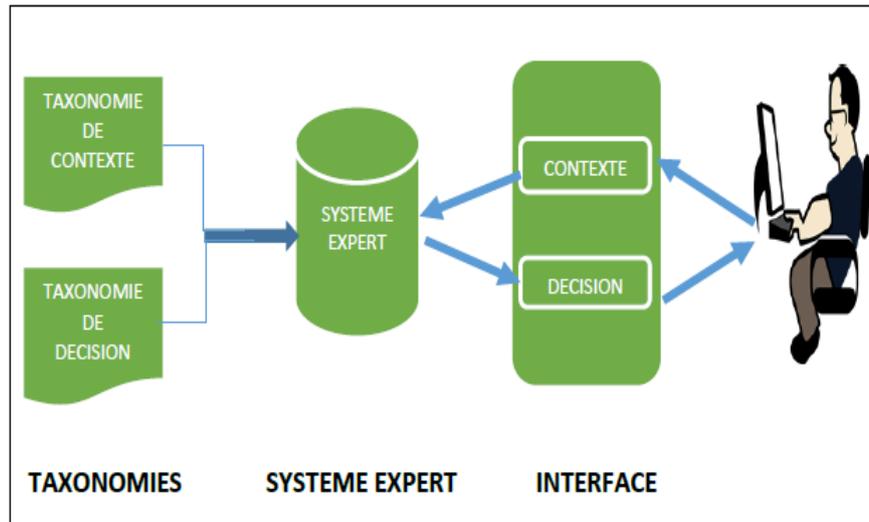


Figure 16 : Architecture de l'outil de guidage opérationnel

L'architecture de l'outil de guidage opérationnel pour la migration de la BI vers le cloud comprend trois grands composants :

- Les taxonomies,
- Le système expert,
- L'interface.

7.3.1. Les taxonomies

L'outil que nous développons est destiné à faciliter l'utilisation des règles décrites au chapitre précédent. Les deux taxonomies complémentaires développées au chapitre 5 forment le domaine de ces règles. En effet, un domaine de règles est l'ensemble de toutes les informations à employer dans l'implémentation des règles (Friedman-Hill 2003).

Ainsi, la taxonomie est le premier composant de l'architecture de l'outil que nous développons. La taxonomie de contexte détermine les différents scénarii de contexte dans lesquels peuvent se trouver les organisations. La taxonomie de décision regroupe les différentes décisions que le système peut proposer à une organisation qui décrit son contexte au préalable.

7.3.2. Le système expert

Le composant système expert est capable de répondre à des questions en effectuant un raisonnement à partir de faits et de règles conçus au préalable, on parle de chaînage avant. Il peut aussi raisonner à partir des faits recherchés et tenter par l'intermédiaire des règles, de remonter à des faits connus, on parle alors de chaînage arrière (Friedman-Hill 2003). Dans le cas de notre outil, nous utilisons le chaînage avant car, après l'application des directives de vérification de la cohérence, de la minimalité et de l'exhaustivité, notre base de règles ne contient que des règles de la forme : Si "contexte", alors "décision". A partir d'un contexte renseigné (prémisse), le système expert fournit au moins une décision (conclusion).

Pour cet outil, c'est JESS qui est utilisé comme système expert. Celui-ci comprend une mémoire de travail ou base de faits, une base de règles et un moteur d'inférence. Nous décrivons ces composants ci-dessous.

7.3.2.1. *Mémoire de travail*

Dans un moteur de règles typique, la mémoire de travail, parfois appelée base de faits, contient toutes les informations que le système basé sur des règles utilise. La mémoire de travail peut contenir les prémisses et les conclusions des règles. Typiquement, le moteur de règle maintient un ou plusieurs index, similaires à ceux utilisés dans les bases de données relationnelles, pour faire, de la recherche dans la mémoire de travail, une opération très rapide.

Il appartient au concepteur du moteur de règles de décider quels types de choses peuvent être stockées dans la mémoire de travail. Certaines mémoires de travail ne peuvent contenir que des objets d'un type spécifique et d'autres peuvent inclure, par exemple, des objets Java.

En ce qui concerne cet outil, la mémoire de travail ou la base de faits contient les éléments de la taxonomie de contexte et de décision représentés sous forme d'expressions en suivant la syntaxe de JESS.

JESS offre la possibilité de définir des templates (emplacements) avant d'enregistrer les faits dans la base. Chaque emplacement est un espace réservé pour une information spécifique. Un emplacement est composé de *slot* ou de *multislot*. Un emplacement est dit *slot* quand il accepte une seule valeur. Il est *multislot* quand il peut accepter plusieurs valeurs. La syntaxe pour définir un template se présente comme suit :

(*def*template nomtemplate
 (slot libellé)
 ⋮
 (multislot libellé))

La syntaxe pour définir un fait se présente comme suit :

(*def*facts nomtemplate
 (libellé "valeur")
 ⋮
 (libellé "valeur" . . . "valeur"))

Pour représenter les éléments de la taxonomie sous forme de faits JESS, nous avons procédé aux transformations suivantes :

1. *def*template nomtemplate devient *def*template contexte ou *def*template décision
 Nous avons défini une template pour chaque taxonomie (contexte et décision) où nous avons réservé les emplacements spécifiques pour chacune de leurs dimensions.
2. *slot libellé* devient *slot nomDimension*
 La réservation de l'emplacement d'une dimension dont un objet ne peut prendre qu'une seule valeur s'effectue en utilisant le mot clé *slot* suivi du nom de la dite dimension. Dans la taxonomie de contexte les emplacements réservés pour les dimensions *ressources financières*, *taille de l'entreprise* et *secteur d'activité* sont de type *slot*. Dans la taxonomie de décision les dimensions *types de cloud*, *contrat*, *voie de transmission des services* et *couches de cloud* sont aussi concernées par ce type d'emplacement.
3. *multislot libellé* devient *multislot nomDimension*
 Pour certaines dimensions, un objet peut cumuler des caractéristiques. Les emplacements de telles dimensions sont de type *multislot*. Dans la taxonomie de contexte, les emplacements des dimensions *objectifs*, *exigences de sécurité* et *ressources humaines* sont de type *multislot*. En ce qui concerne la taxonomie de décision les emplacements de type *multislot* ont été réservés pour les dimensions *niveau de migration* et *supports des applications BI*.

4. *deffacts nomtemplate* devient *deffacts contexte* ou *deffacts décision*

Les faits sont des éléments de la taxonomie de contexte et de décision représentés sous forme d'expressions. La définition des expressions issues de la taxonomie de contexte commence par "deffact contexte" et des expressions issues de la taxonomie de décision commencent par "deffact décision".

5. *libellé "valeur"* devient *nomDimension "nomCaractéristique"*

Il s'agit de renseigner la dimension concernée en précisant l'unique caractéristique concernée. Ce sont les emplacements de type *slot*. Libellé correspond à une dimension et valeur à une caractéristique.

6. *libellé "valeur"..."valeur"* devient *nomDimension "nomCaractéristique"..."nomCaractéristique"*

Un emplacement est réservé pour une dimension dont un objet peut avoir plusieurs caractéristiques. Ce sont des emplacements de type *multislot*.

Le template pour les éléments de la taxonomie de contexte se présente comme suit :

(deftemplate contexte
 (multislot objectifs)
 (slot ressources financières)
 (multislot ressources humaines)
 (multislot exigences de sécurité)
 (slot secteur d'activité)
 (slot taille de l'entreprise))

Le template pour les éléments de la taxonomie de décision se présente comme suit :

(deftemplate décision
 (multislot niveau de migration)
 (slot types de cloud)
 (slot contrat)
 (slot voie de transmission des services)
 (multislot supports des applications BI)
 (slot couches de cloud))

Exemple de faits :

- (defact (contexte (objectifs "réduction des coûts d'implémentation" "réduction des coûts d'infrastructure" "réduction des coûts d'administration")))
- (defact (décision (type cloud "privé")))

7.3.2.2. *Base de règles*

La base de règles contient toutes les règles que le système connaît. Dans ce cas précis, la base de règles comprend toutes les règles qui ont été décrites au chapitre précédent. Les règles peuvent simplement être stockées comme des chaînes de caractère, mais le plus souvent, un compilateur de règles les transpose dans une forme permettant au moteur d'inférence de travailler plus efficacement. Le compilateur de règles de JESS construit une structure de données indexée complexe appelée réseau RETE (Friedman-Hill 2003). Un réseau RETE est une structure de données qui simplifie et rend rapide le traitement des règles. Le compilateur de règles peut ajouter ou réorganiser les prémisses ou les conclusions d'une règle pour la rendre plus efficace.

La syntaxe des règles JESS se présente comme suit:

```
(Defrule nomRègle
  (prémisse)
  =>
  (conclusion))
```

En JESS, pour définir une règle, on utilise le mot clé "**defrule**" qui doit être suivi du nom de la règle à définir. Une règle est composée de deux parties et est entourée de parenthèses. Les deux parties sont séparées par le signe "=>". La partie gauche est la prémisse de la règle. La partie droite en est la conclusion. Le signe "=>" signifie "alors". On suppose que la prémisse est une condition qui doit être remplie pour que la conclusion qui en est le résultat puisse s'exécuter. Ainsi, la prémisse est comparée à tous les faits existants dans la mémoire de travail. Au cas où elle correspond à certains faits, la règle se déclenche et sa conclusion est exécutée.

Dans le cas de la migration de la BI vers le cloud, le contexte décrit avec l'interface de l'outil constitue la prémisse d'une règle. Il est comparé à toutes les expressions ou contextes figurant dans la mémoire de travail. Au cas où il correspond à un ou plusieurs faits, les règles correspondantes se déclenchent et leurs parties conclusions s'affichent sur l'écran.

La partie prémisse d'une règle peut utiliser les opérateurs logiques suivants : Et noté "&", ou noté "|" et non noté "~ ". L'opérateur & est souvent représenté par un vide : " " Dans le cas de deux prémisses qui se suivent sans un connecteur, JESS considère qu'il s'agit du connecteur "&" implicite.

La conclusion commence par le mot clé "assert". Elle peut avoir plusieurs parties. Celles-ci sont toujours connectées par l'opérateur "Et" explicite ou implicite. JESS exécute les conclusions l'une après l'autre en utilisant le connecteur "&".

Pour écrire les règles que nous avons décrites au chapitre précédent avec la syntaxe JESS, nous nous sommes référés aux transformations effectuées lors de la prise en compte des faits dans la mémoire de travail au point 7.3.2.1. Ainsi, chacune des parties de la règle est remplacée par un ou plusieurs faits précédés par le nom de leurs templates respectifs. La syntaxe devient comme suit :

(defrule NomRègle

(contexte ((nomDimension "nomCaractéristique"|"nomCatégorie")... (nomDimension "nomCaractéristique" ... "nomCaractéristique"))

| ((nomDimension "nomCaractéristique"|"nomCatégorie") | (nomDimension

"nomCaractéristique"|"nomCatégorie")|...|(nomDimension

"nomCaractéristique"|"nomCatégorie"))

| ((nomDimension (~"nomCaractéristique"))

=>

(assert (décision (nomDimension "nomCaractéristique"|"nomCatégorie") ... (nomDimension "nomCaractéristique"|"nomCatégorie"))))

Exemple de règle :

(defrule Règle1

(contexte (objectif "réduction des coûts d'implémentation" "réduction des coûts d'infrastructure" "réduction des coûts d'administration") (ressources financières "faibles") (exigences de sécurité "disponibilité") (taille de l'entreprise "petite" | "moyenne"))

=>

(assert (décision (niveau de migration "total") (types de cloud "public") (contrat "long terme") (moyen de transmis des services "internet"))))

7.3.2.3. Moteur d'inférence

Le moteur d'inférence se charge de l'exécution des règles et contrôle leur enchaînement. Il se compose d'un pattern de spécification, d'un agenda et d'un moteur d'exécution.

Le pattern de spécification décide quelles règles déclencher (règles candidates) étant donné le contenu de mémoire de travail (base de faits). Il peut rechercher à travers des millions de combinaisons de faits pour trouver les combinaisons qui satisfont les règles. Il utilise l'algorithme RETE (Friedman-Hill 2003). Cet algorithme assure automatiquement un certain nombre de contrôles de minimalité et de cohérence de la base de règles : Il réduit ou élimine certains types de redondance. Il stocke les correspondances partielles quand l'exécution d'une règle met en relation les différents types de faits. Ceci permet d'éviter la réévaluation complète de tous les faits chaque fois que des changements sont effectués dans la mémoire de travail. L'algorithme RETE permet aussi l'actualisation de la mémoire de travail quand certains faits sont retirés ou supprimés de la base.

Le moteur d'inférence détermine les règles qui doivent être déclenchées, et l'ordre dans lequel cela s'accomplira. La liste des règles pouvant potentiellement être déclenchées (règles candidates) est stockée dans l'**agenda**. Celui-ci est chargé de déterminer laquelle des règles, parmi celles qui sont candidates, a la plus haute priorité et doit être déclenchée en premier.

Enfin, une fois que le moteur de règles décide de la règle à déclencher, il doit exécuter la partie conclusion de cette règle. Le **moteur d'exécution** est le composant de moteur d'inférence qui déclenche les règles.

Un cycle d'un moteur d'inférence se présente comme suit :

- Sélection des règles candidates (potentielles à être déclenchées) : toutes les règles sont comparées dans la mémoire de travail à l'aide de patterns de spécification pour décider lesquelles devraient être activées pendant un cycle. Cette liste non ordonnée de règles activées, ainsi que toutes les autres règles activées dans les cycles précédents, est appelée "ensemble de conflit".
- Résolution de conflit : Il consiste à ordonner la liste des règles activées pour constituer l'agenda. Le processus d'ordonnement de la liste de conflit est appelé "résolution de conflit".
- Déclenchement des règles : pour terminer le cycle, la première règle de l'agenda est déclenchée.

Notons que les résultats obtenus à partir d'un pattern de spécification et du résolveur de conflits de l'agenda peuvent être sauvegardés à travers les cycles, de sorte que seul un travail différentiel puisse être refait au besoin.

7.3.3. L'interface

Comme nous l'avons précisé au point 3.5, nous mettons en place un guidage opérationnel de type suggestif. Celui-ci indique comment les utilisateurs peuvent répondre aux opportunités de jugement qu'ils rencontrent. Le guidage suggestif recommande les paramètres et les autres valeurs d'entrée ainsi que les opérateurs à utiliser (Silver 1991).

Ainsi, dans cet outil de guidage opérationnel, le décideur décrit son contexte à travers l'interface en choisissant les valeurs qui lui sont proposées et demande à la base de règles de lui fournir toutes les décisions y relatives. Un contexte peut être la prémisse d'une ou plusieurs règles dans la base de règles, tandis que les décisions sont des conclusions relatives à la prémisse énoncée.

Un contexte est constitué des six dimensions de la taxonomie de contexte : objectifs, ressources financières, ressources humaines, exigences de sécurité, secteur d'activité et taille de l'entreprise.

En effet, il s'agit pour une organisation de dire l'objectif ou les objectifs qu'elle poursuit en voulant migrer sa BI vers le cloud. En ce qui concerne les ressources financières,

il s'agit de parler de ses capacités financières (faibles ou fortes). Pour les ressources humaines, l'organisation devra préciser les spécialités informatiques que son personnel peut avoir. Les exigences de la sécurité se résument en la confidentialité, l'intégrité et la disponibilité. Pour cela, l'utilisateur doit préciser l'aspect (ou les aspects) sur lequel il voudrait insister. Par exemple dans la migration de son entrepôt des données vers le cloud, une organisation peut insister sur l'intégrité et la confidentialité de ses données. L'organisation doit renseigner le domaine de ses activités et préciser sa taille (petite, moyenne ou grande).

Par ailleurs, proposer les dimensions à renseigner dans ce guidage opérationnel ne signifie pas nécessairement que l'utilisateur doit renseigner toutes les dimensions. Nous avons simplement anticipé les différents contextes qui peuvent se produire. Néanmoins, le contexte est mieux décrit quand l'utilisateur renseigne le plus possible de dimensions de la taxonomie de contexte.

Une fois que le contexte est décrit, l'utilisateur (décideur) peut alors demander au système de lui proposer les différentes décisions relatives à son contexte. Pour un contexte, le système peut proposer une ou plusieurs décisions, c'est au décideur de faire le choix final. Au cas où le contexte est imprécis, le système demande à l'utilisateur de spécifier davantage son contexte.

L'interface utilisateur, construite en Java, qui permet au décideur de manipuler cet outil de guidage opérationnel se présente comme suit :

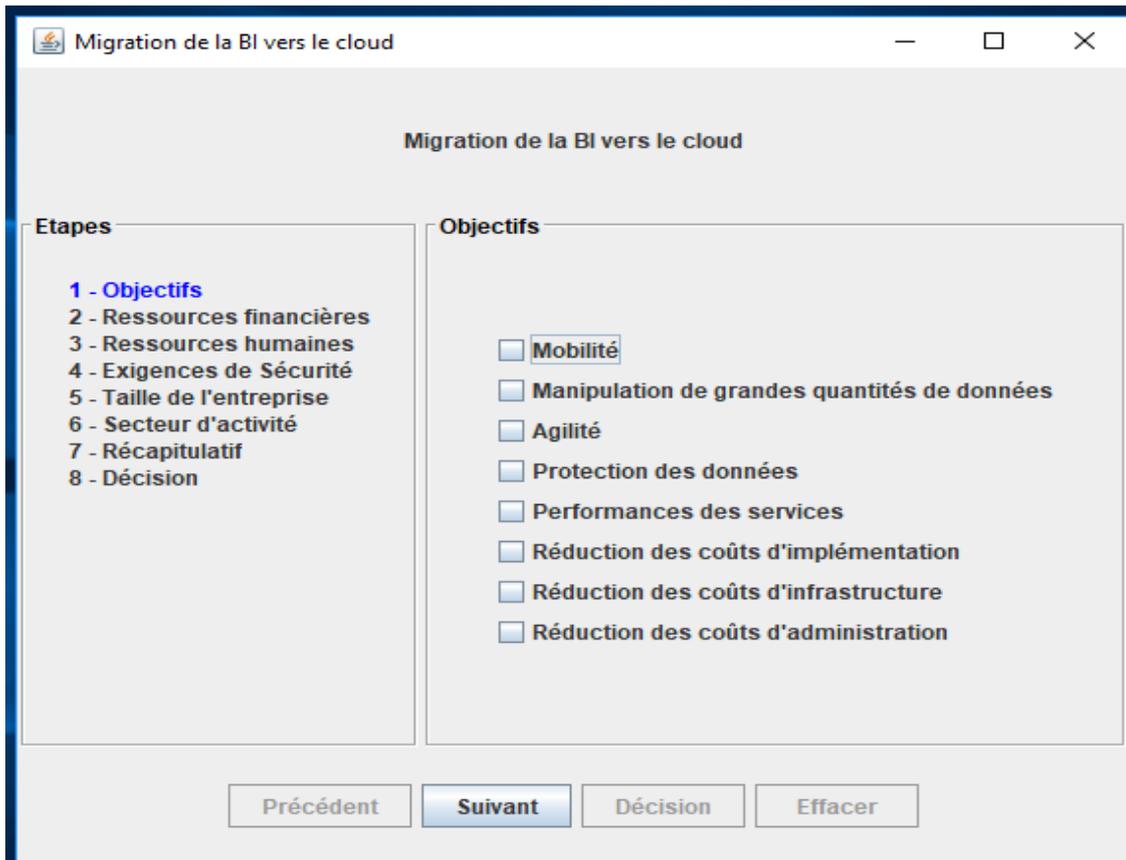


Figure 17 : Outil de guidage opérationnel : interface utilisateur

Cette interface utilisateur est divisée en deux parties dont la partie gauche affiche les étapes à parcourir par l'utilisateur. La partie droite montre l'étape active. C'est dans la partie active que l'utilisateur peut choisir une ou plusieurs caractéristiques (selon les cas) qui décrivent son organisation. Etant donné que pour une dimension, un objet peut avoir ou ne pas avoir de caractéristique (voir point 3.3), l'utilisateur peut passer une étape sans choisir une caractéristique.

A l'aide du bouton "Suivant", on peut progresser vers une autre étape pour renseigner une autre dimension. Le bouton "Précédent" permet de revenir en arrière pour revoir et au besoin apporter une correction sur une étape ou une dimension déjà passée. C'est le bouton "Décision" qui permet au moteur d'inférence d'exécuter la demande en vue d'afficher à l'écran la ou les décisions relatives au contexte décrit. Le bouton "Effacer" permet d'effacer un contexte en vue d'en saisir un autre.

7.4. Diagramme de cas d'utilisation

Pour illustrer comment le décideur pourra utiliser cet outil que nous développons, nous présentons, ci-dessous, le diagramme de cas d'utilisation y relatif.

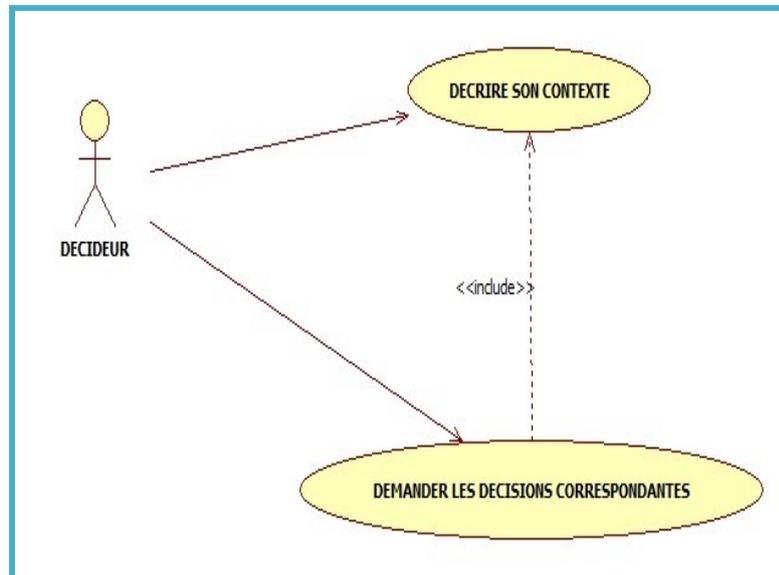


Figure 18: Diagramme de cas d'utilisation de l'outil de guidage opérationnel

Le diagramme de cas d'utilisation ci-haut illustre que l'utilisateur doit d'abord décrire son contexte avant de demander au système de lui proposer les éventuelles décisions. C'est la raison pour laquelle le cas d'utilisation "demander les décisions correspondantes" est lié via «Include» au cas d'utilisation "décrire son contexte".

Après la description du contexte, suivra le cas d'utilisation "demander les décisions correspondantes". Il s'agit d'interroger la base de règles afin d'afficher à l'écran de l'utilisateur d'éventuelles décisions relatives au contexte qui a été décrit.

Dans la section suivante, nous présentons un scénario qui illustre l'application de ce cas d'utilisation afin de valider notre approche.

7.5. Validation de l'approche

Pour valider notre approche, nous présentons un scénario illustratif et nous vérifions la cohérence, la complétude et la minimalité de notre base de règles.

7.5.1. Scénario illustratif

Un scénario illustratif est une méthode d'évaluation qui consiste à décrire une situation artificielle de résolution de problèmes mais de manière réaliste. C'est un scénario détaillé autour de l'artefact pour démontrer son efficacité, son utilité ou sa faisabilité (Prat, Comyn-Wattiau and Akoka 2015).

Dans le cas de l'outil de guidage opérationnel pour la migration de la BI vers le cloud, nous allons utiliser des données réelles en vue de démontrer, au moyen de l'outil, l'utilité et l'efficacité de notre approche opérationnelle pour la migration de la BI vers le cloud.

7.5.1.1. *Énoncé*

La coordination informatique interministérielle (CII) est une entité qui coordonne les dépenses de l'Etat Congolais. Elle centralise toutes les dépenses du gouvernement central et celles des gouvernements provinciaux. Elle est, administrativement, sous tutelle du ministère du budget mais du point de vue de ses attributions, elle est aussi un organe du ministère des finances. Elle s'occupe des quatre phases d'exécution des dépenses publiques : l'engagement, la liquidation, l'ordonnancement et le paiement.

Elle est représentée dans chaque ministère et à la primature de la République par une équipe de fonctionnaires sous la supervision d'un cadre appelé administrateur de budget. Chaque équipe possède de l'équipement informatique qui est connecté au réseau avec le bureau de la CII. Les membres de ces équipes ne sont que des utilisateurs finaux. Seuls les membres de la CII sont informaticiens : analystes-développeurs, ingénieurs concepteurs et administrateurs réseaux.

La CII gère un entrepôt de données qui permet de suivre l'exécution du budget dans chaque institution de l'Etat. Elle analyse toutes les dépenses effectuées à tous les niveaux de l'administration en vue d'établir les statistiques et de proposer le budget pour l'année suivante. La discrétion est une règle d'or de la CII et l'intégrité des données est une recommandation.

La CII éprouve les difficultés suivantes :

- Les coupures d'électricité fréquentes endommagent les serveurs de stockage des données.
- Dans un pays où sévit la guerre, le problème relatif à la protection des données est crucial.
- Beaucoup d'agents dans les provinces ou mêmes dans certains ministères à Kinshasa sont des utilisateurs finaux.
- La CII manque de certains équipements pour collecter, traiter et restituer les informations en temps réel.
- Les coûts relatifs à l'infrastructure et à l'administration de l'entrepôt des données sont très élevés.

Le directeur de la CII, Monsieur Jean Gaston MANYA, pense qu'il est opportun de migrer vers le cloud. Cependant, il n'a pas d'informations suffisantes au sujet de la migration de la BI vers le cloud.

7.5.1.2. Résolution

Le Directeur de la CII a utilisé l'outil de guidage opérationnel de migration de la BI vers le cloud. Il a renseigné les dimensions de la taxonomie de contexte pour décrire la situation de la CII.

- Objectif de migration : Performance des services, protection des données et réduction des coûts d'infrastructure
- Ressources financières : faibles
- Ressources humaines : utilisateurs finaux, administrateurs réseaux, analystes et développeurs
- Exigences de sécurité : Intégrité, confidentialité et disponibilité
- Secteur d'activité : administration publique
- Taille de l'entreprise : grande.

La décision

Au regard du contexte de la CII, particulièrement des compétences qu'elle possède, il serait souhaitable que ses agents développent eux-mêmes un système qui réponde aux besoins de leur organisation. Par ailleurs, il serait souhaitable que la CII choisisse un type de cloud qui garantit la sécurité vu la sensibilité de ses données.

Ayant utilisé l'outil de guidage opérationnel que nous avons mis en place, la décision proposée est la suivante :

- Niveau de migration : "partiel"
- Type de cloud : "hybride"
- Contrat : "court terme"
- Moyen de transmission : "réseau d'entreprise"
- Couche cloud : "PaaS/BI composite".

Etant donné que la CII a du personnel compétent pour l'analyse et le développement, la couche cloud PaaS est conseillée. Le personnel de la CII se chargera de la mise en place d'un système de BI correspondant à la situation particulière de ce service de l'Etat Congolais. Le contrat conclut pour la BI composite est généralement à court terme.

Le domaine des finances est un domaine sensible, le recours au cloud hybride est indiqué car il comprend le cloud privé pour les données sensibles et le cloud public pour les données ordinaires.

Cette décision est efficace car elle capitalise sur les compétences du personnel de la CII et lui permet de résoudre le problème de manque d'infrastructure matérielle qui coûte cher.

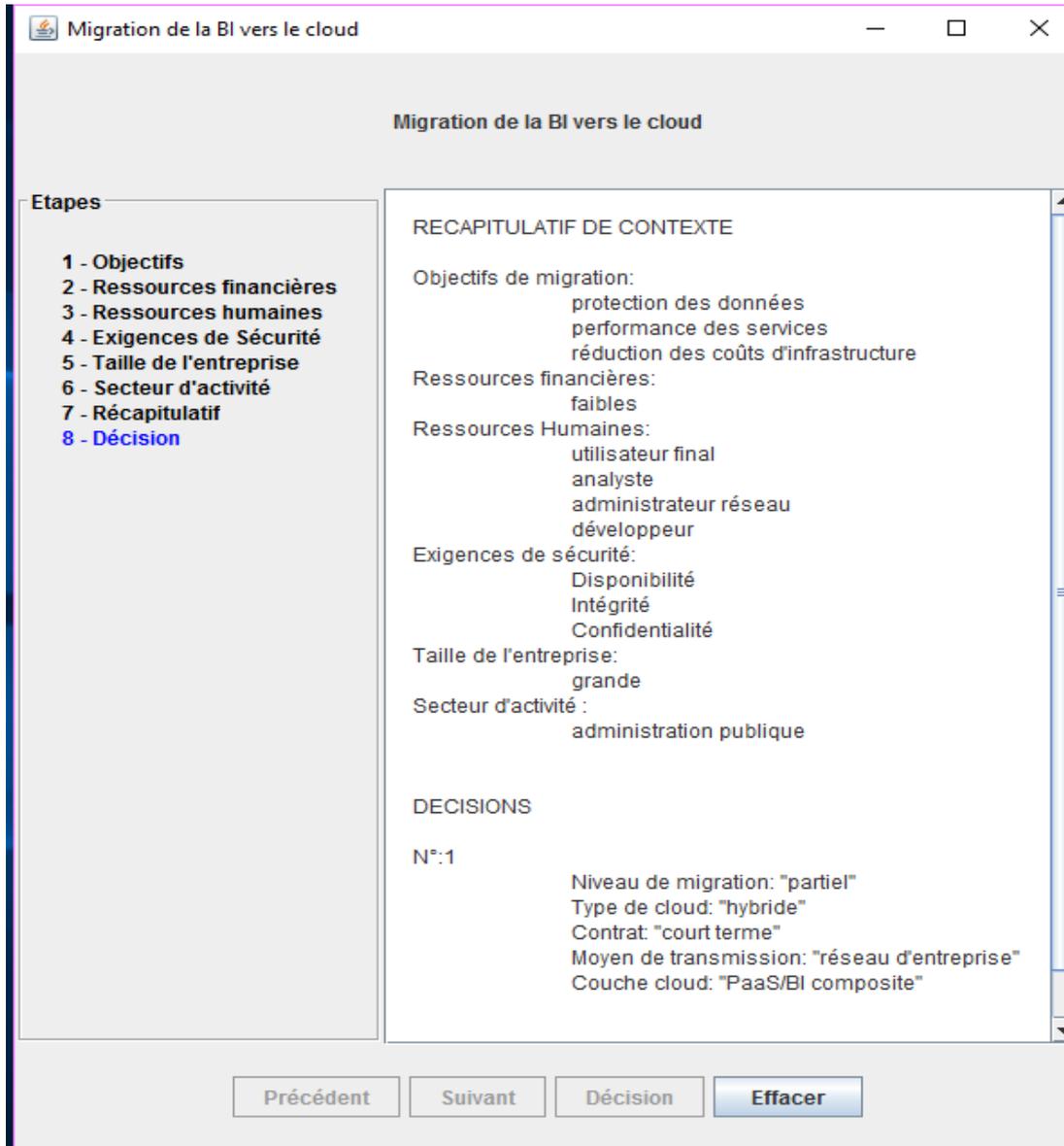


Figure 19: Décision de migration vers le cloud de la CII

7.5.2. Cohérence et exhaustivité de la base de règles

Grâce au pattern de spécification, JESS vérifie la cohérence syntaxique des règles de manière automatique. En revanche, JESS n'est pas en mesure de vérifier une cohérence sémantique des mots. Raison pour laquelle dans le chapitre précédent, nous avons vérifié manuellement la cohérence sémantique de notre base de règles. En effet, la cohérence syntaxique est relative à la syntaxe d'une règle et la cohérence sémantique est liée au sens des

mots utilisés dans une règle. Grâce à cette double vérification (manuelle et automatique), notre base de règles ne produit pas de conclusions contradictoires.

En ce qui concerne l'exhaustivité de notre base de règles, nous avons effectué plusieurs simulations de contextes, notre base de règle reste en mesure de produire au moins une conclusion. Par ailleurs, dans le cas où le contexte n'est pas suffisamment précis, l'outil demande à l'utilisateur de fournir plus d'informations afin de lui permettre de proposer une conclusion. En effet, un contexte est bien décrit quand l'utilisateur renseigne le plus possible de dimensions de la taxonomie de contexte. Par exemple, pour un même objectif de migration, les décisions proposées peuvent être différentes selon les tailles des organisations ou selon le niveau des ressources financières. Par conséquent, il est important qu'un utilisateur décrive son contexte en précisant non seulement l'objectif de migration, mais aussi la taille de l'entreprise et le niveau des ressources financières.

Ainsi, si l'objectif de migration est la réduction des coûts, la décision peut être la migration totale dans le cas d'une petite ou moyenne entreprise. En revanche, si c'est une grande entreprise, la décision serait une migration partielle : matériel ou outils.

7.6. Conclusion

Dans ce chapitre, nous avons présenté l'architecture de l'outil de guidage opérationnel que nous avons développé pour aider à l'utilisation des taxonomies sous forme des règles. Cette architecture comprend trois composants : taxonomies, système expert et interface utilisateur.

En vue de valider notre approche, nous avons mis en place un scénario illustratif : grâce à l'interface de notre outil opérationnel, nous avons renseigné les dimensions de la taxonomie de contexte avec les données de la Direction Informatique Interministérielle (CII) de la République Démocratique du Congo qui désire migrer sa BI vers le cloud. Cet outil a proposé à la CII une conclusion relative à sa situation.

Nous avons proposé cette solution au Directeur de la CII qui a élaboré un projet de migration qu'il a soumis aux autorités compétentes pour l'adoption de cette solution. Ceci montre que notre outil de guidage opérationnel est utile et peut aider les organisations dans la décision de la migration de leur BI vers le cloud.

Cependant, le Directeur de la CII a souhaité savoir l'impact de cette migration au sein de la CII. A partir de ce constat, nous proposons d'affiner notre outil de guidage opérationnel en ajoutant une troisième taxonomie complémentaire : taxonomie des implications. Cela pourra aider les organisations à mieux se préparer à la migration de leur BI vers le cloud.

CHAPITRE 8 : CONCLUSION ET PERSPECTIVES

L'objectif de cette thèse était de mettre en place une aide à la décision de migration de la BI vers le cloud. Pour justifier l'importance et la pertinence de notre question de recherche, nous avons effectué une revue systématique de la littérature dans le but de résumer les études existantes sur la BI dans le cloud. Cette revue systématique avait aussi pour but de déceler les lacunes pour suggérer des voies de solutions. Etant donné que nous avons adopté l'approche de science de conception (design science), nous nous sommes appuyés sur notre typologie des artefacts en science de conception pour exhiber les artefacts déjà mis en place dans le domaine de la BI dans le cloud d'une part. D'autre part, nous avons identifié des opportunités de recherche en science de conception pour combler le vide en développant de nouveaux artefacts en vue de permettre à la BI de profiter des opportunités de cloud.

Basés sur ces opportunités de recherche que la BI dans le cloud offre pour la recherche en science de conception, nous avons choisi de mettre en place un support d'aide à la décision de la migration de la BI vers le cloud qui s'appuie sur les taxonomies. Ainsi, nous avons mis en place un modèle de guidage opérationnel. Ce dernier a permis d'instancier une taxonomie de la BI dans le cloud d'une part. D'autre part, de ce modèle, ont découlé des règles pour la migration de la BI vers le cloud. Pour ce faire, nous avons eu besoin de deux taxonomies : de contexte et de décision. En outre, nous avons utilisé le langage EBNF (Extended Backus-Naur Form) pour décrire de manière systématique la syntaxe de ce modèle de guidage opérationnel. Dans le but de vérifier la cohérence, la minimalité et l'exhaustivité des règles, nous avons mis en place un ensemble de directives.

Pour développer les deux taxonomies sur lesquelles s'appuie le support d'aide à la décision, nous avons eu besoin d'une méthodologie de développement des taxonomies. Pour ce faire, nous nous sommes référés à la méthodologie de Nickerson, Varshney et Muntermann (2013). Celle-ci a montré des limites dans son application en général et en particulier pour les technologies complexes émergentes. En effet, la BI dans le cloud est une technologie complexe émergente. Pour développer une taxonomie relative à ce domaine, il fallait tenir compte des défis spécifiques que posent ces genres de technologies. En tenant compte de ces limites, nous avons proposé une méthodologie appropriée pour le développement des

taxonomies des technologies complexes émergentes. Notre méthodologie s'est basée principalement sur la revue systématique de la littérature (Kitchenham 2004). Celle-ci a enrichi le processus de développement de taxonomie tel que proposé par Nickerson, Varshney et Muntermann.

Grâce à cette méthodologie que nous avons mise en place, nous avons élaboré, de manière itérative, une taxonomie des éléments qui peuvent influencer la migration de la BI vers le cloud d'une part. D'autre part, nous avons élaboré la taxonomie des options relatives aux services dont la BI peut bénéficier dans le cloud.

En outre, nous avons présenté un ensemble de règles qui ont pour domaine, les deux taxonomies complémentaires développées au préalable. Ces règles ont instancié le modèle de guidage opérationnel. Nous avons vérifié la cohérence, minimalité et la complétude de notre base de règles en utilisant les directives que nous avons mises en place lorsque nous avons élaboré le modèle de guidage opérationnel. Toutes les règles sont cohérentes les unes par rapport aux autres et toutes par rapport au modèle de guidage opérationnel. Par souci de minimalité, nous avons supprimé et fusionné certaines règles. Pour des raisons d'exhaustivité, nous avons ajouté d'autres règles dans la base. Notre base de règles est cohérente, minimale, et l'exhaustivité à un moment donné.

Afin de faciliter l'utilisation des règles, nous avons développé un outil de guidage opérationnel. Les trois composants de l'architecture de cet outil sont les taxonomies, le système expert et l'interface. Ce dernier permet de renseigner le contexte d'une organisation désireuse de migrer sa BI vers le cloud afin qu'une décision adaptée lui soit proposée.

Pour valider notre approche, nous avons présenté un scénario illustratif et nous avons vérifié la cohérence, la complétude et la minimalité de notre base de règles.

Par ailleurs, les travaux de cette thèse ont été présentés lors de différentes conférences, dans une revue internationale ainsi qu'à un consortium doctoral. Cependant, comme tout travail de recherche, il y a toujours lieu d'affiner pour apporter des améliorations, de nouvelles perspectives de recherche.

Cette thèse a proposé deux taxonomies complémentaires de la BI pour la migration de la BI vers le cloud qui n'existaient pas auparavant. Cependant, avec l'évolution technologique, ces taxonomies restent évolutives, leur pérennité est relative. Par ailleurs, les deux taxonomies pourraient conduire à la mise en place d'une troisième qui serait la taxonomie des implications. Alors que la taxonomie de contexte regroupe les éléments qui peuvent influencer la migration de la BI vers le cloud, la taxonomie de décision présente les options des services de cloud à choisir pour un contexte donné. La taxonomie des implications, quant à elle, représenterait les implications de la migration de la BI vers le cloud pour un contexte et une décision donnés. Par conséquent, l'outil de guidage opérationnel pourrait aussi être affiné pour prendre en compte les implications de la migration de la BI vers le cloud.

Certes, la démarche pour la mise en place du modèle de guidage opérationnel dont découlent les règles pour la migration de la BI vers le cloud et la méthodologie de développement de taxonomies pour les technologies complexes émergentes possèdent une pérennité certaine, mais les taxonomies et les règles, quant à elles, ont une pérennité relative nécessitant un travail fastidieux de mise à jour.

Dans un futur proche, nous pensons adapter nos deux taxonomies en prenant en compte l'évolution technologique et les compléter par une taxonomie des implications de la migration de la BI vers le cloud car cette dernière ne procure pas simplement des avantages, mais elle implique aussi des défis à relever. Nous pensons qu'il est important que les organisations avant de s'engager, aient une information suffisante en ce qui concerne la migration de la BI vers le cloud. Nous pensons aussi utiliser le chaînage arrière dans notre système expert JESS pour savoir le contexte indiqué pour un service cloud précis.

PUBLICATIONS

1. Sangupamba Mwilu O, Comyn-Wattiau I, Prat N. Design science research contribution to business intelligence in the cloud — A systematic literature review. *Future Gener Comput Syst* 2016; 63 : 108–22.
2. Sangupamba Mwilu O, Prat N, Comyn-Wattiau I. Taxonomy Development for Complex Emerging Technologies – The case of Business Intelligence in the cloud. PACIS 2015, Paper 29, Singapour, Singapore
3. Sangupamba Mwilu O, Prat N, Comyn-Wattiau I. Business Intelligence and Big Data in the Cloud: Opportunities for Design-Science Researchers. *Advances in Conceptual Modeling—Proceedings of ER 2014 Workshops*. GA, USA, 2014, 75–84.

BIBLIOGRAPHIE

- Abbasi A, Sarker S, Chiang RH. Big data research in information systems: Toward an inclusive research agenda. *J Assoc Inf Syst* 2016;**17**:3.
- AbdelBaky M, Kim H, Rodero I *et al.* Accelerating MapReduce Analytics Using CometCloud. *IEEE Fifth International Conference on Cloud Computing*. 2012, 447–54.
- Abelló A, Ferrarons J, Romero O. Building cubes with MapReduce. *ACM 14th International Workshop on Data Warehousing and OLAP*. 2011, 17–24.
- Aceto G, Botta A, de Donato W *et al.* Cloud monitoring: A survey. *Comput Netw* 2013;**57**:2093–115.
- Al-Aqrabi H, Liu L, Hill R *et al.* Taking the Business Intelligence to the Clouds. *IEEE 14th International Conference on High Performance Computing and Communications*. 2012, 953–8.
- Al-Aqrabi H, Liu L, Hill R *et al.* Cloud BI: Future of business intelligence in the Cloud. *J Comput Syst Sci* 2015;**81**:85–96.
- Alison D. Cloud, Analytics Improve Breast Cancer Screening Education
<http://www.informationweek.com/healthcare/analytics/cloud-analytics-improve-breast-cancer-screening-education/d/d-id/1234942>. 2014a.
- Alison D. Better Healthcare Through Analytics
<http://www.informationweek.com/healthcare/analytics/better-healthcare-through-analytics/d/d-id/1252913>. 2014b.
- Apache Hadoop. Hadoop : <http://hadoop.apache.org/docs/>. 2017.
- Ari I, Olmezogullari E, Çelebi ÖF. Data stream analytics and mining in the cloud. *IEEE 4th International Conference on Cloud Computing Technology and Science*. 2012, 857–62.
- Arres B, Kabbachi N, Boussaid O. Building olap cubes on a cloud computing environment with mapreduce. *IEEE International Conference on Computer Systems and Applications*. 2013, 1–5.
- Assunção MD, Calheiros RN, Bianchi S *et al.* Big Data computing and clouds: Trends and future directions. *J Parallel Distrib Comput* 2014;**79–80**:3–15.
- Baars H, Kemper H-G. Business intelligence in the cloud? *PACIS Proceedings*. 2010, paper 145.

- Baars H, Kemper H-G. Ubiquitous Computing-an Application Domain for Business Intelligence in the Cloud? *AMCIS Proceedings*. 2011, paper 93.
- Babcock C. Boxever Implements DevOps, Continuous Delivery On Amazon <http://www.informationweek.com/cloud/software-as-a-service/boxever-implements-devops-continuous-delivery-on-amazon/d/d-id/1318374>. 2014a.
- Babcock C. How Amazon Kinesis Adds Speed, Resilience To Analytics <http://www.informationweek.com/cloud/software-as-a-service/how-amazon-kinesis-adds-speed-resilience-to-analytics/d/d-id/1317590>. 2014b.
- Babcock C. GoGrid Emerges As Cloud's Big Data Specialist <http://www.informationweek.com/cloud/platform-as-a-service/gogrid-emerges-as-clouds-big-data-specialist/d/d-id/1269582>. 2014c.
- Babcock C. IBM, Tencent Offer Enterprise Cloud In China <http://www.informationweek.com/cloud/infrastructure-as-a-service/ibm-tencent-offer-enterprise-cloud-in-china/d/d-id/1317103>. 2014d.
- Babcock C. Cloud SLAs: Improvements Still Needed <http://www.informationweek.com/cloud/infrastructure-as-a-service/cloud-slas-improvements-still-needed/a/d-id/1318730>. 2015.
- Barga RS, Ekanayake J, Lu W. Project Daytona: Data Analytics as a Cloud Service. *IEEE 28th International Conference on Data Engineering*. 2012, 1317–20.
- Baur AW, Bühler J, Bick M. How pricing of business intelligence and analytics SaaS applications can catch up with their technology. *J Syst Inf Technol* 2015;17:229–46.
- Baxter-Reynolds M. Everyone is “going digital,” but just what does that mean? <http://www.zdnet.com/article/everyone-is-going-digital-but-just-what-does-that-mean/>. 2014.
- Bennett C, Grossman RL, Locke D *et al*. Malstone: towards a benchmark for analytics on large data clouds. *ACM Proceedings of the 16th International Conference on Knowledge Discovery and Data Mining*. 2010, 145–152.
- Bertolucci J. Big Data Debate: Do Analytics Trump Intuition? <http://www.informationweek.com/big-data/big-data-analytics/big-data-debate-do-analytics-trump-intuition/d/d-id/1269193>. 2014a.
- Bertolucci J. 5 Big Data Use Cases To Watch <http://www.informationweek.com/big-data/big-data-analytics/5-big-data-use-cases-to-watch/d/d-id/1251031>. 2014b.
- Bertolucci J. Big Data: Matching Personalities In The Call Center <http://www.informationweek.com/big-data/big-data-analytics/big-data-matching-personalities-in-the-call-center/d/d-id/1319108>. 2015.
- Böhringer M, Gluchowski P, Kurze C *et al*. A Business Intelligence Perspective on the Future Internet. *AMCIS Proceedings*. 2010, paper 267.

- Booker E. City Of Austin Broadens Data Analysis Efforts
<http://www.informationweek.com/government/big-data-analytics/city-of-austin-broadens-data-analysis-efforts/d/d-id/1315474>. 2014.
- Brezany P, Zhang Y, Janciak I *et al.* An Elastic OLAP Cloud Platform. *IEEE International Conference on Dependable, Autonomic and Secure Computing*. 2011a, 356–63.
- Brezany P, Zhang Y, Janciak I *et al.* An Elastic OLAP Cloud Platform. *IEEE*, 2011b, 356–63.
- Bruchez R. *Les bases de données NoSQL et le big data : Comprendre et mettre en oeuvre*. Paris: Eyrolles, 2015.
- Cao Y, Chen C, Guo F *et al.* Es 2: A cloud data storage system for supporting both oltp and olap. *IEEE 27th International Conference on Data Engineering*. 2011, 291–302.
- Carlos Perez J. Microsoft buys Capptain for mobile analytics, cloud services
<http://www.computerworld.com/article/2489915/business-intelligence/microsoft-buys-capptain-for-mobile-analytics--cloud-services.html>. 2014.
- Chaisiri S, Bong Z, Lee C *et al.* Workflow framework to support data analytics in cloud computing. *IEEE 4th International Conference on Cloud Computing Technology and Science*. 2012, 610–613.
- Chang L, Ranjan R, Zhang X *et al.* Public Auditing for Big Data Storage in Cloud Computing -- A Survey. *IEEE 16th International Conference on Computational Science and Engineering*. 2013, 1128–35.
- Chang L, Yang C, Zhang X *et al.* External integrity verification for outsourced big data in cloud and IoT: A big picture. *Future Gener Comput Syst* 2015;**49**:58–67.
- Chang V. The business intelligence as a service in the cloud. *Future Gener Comput Syst* 2014;**37**:512–534.
- Chaudhuri S. What next ? : a half-dozen data management research goals for big data and the cloud. *ACM Proceedings of the 31st Symposium on Principles of Database Systems*. 2012, 1–4.
- Chaudhuri S, Dayal U, Narasayya V. An Overview of Business Intelligence Technology. *Commun ACM* 2011;**54**:88–98.
- Chen H, Chiang RH, Storey VC. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Q* 2012;**36**:1165–88.
- Chen PP. An algebra for a directional binary entity-relationship model. *IEEE First International Conference on Data Engineering*. 1984, 37–40.
- Chi Y, Moon HJ, Hacigümüş H *et al.* SLA-tree: a framework for efficiently supporting SLA-based decisions in cloud computing. *ACM Proceedings of the 14th International Conference on Extending Database Technology*. 2011, 129–140.

- Chohan N, Gupta A, Bunch C *et al.* Hybrid cloud support for large scale analytics and web processing. *Proceedings of the 3rd USENIX Conference on Web Application Development*. 2012, 39–50.
- Clay RB, Shen Z, Ma X. Accelerating Batch Analytics with Residual Resources from Interactive Clouds. *IEEE 21st International Symposium on Modelling, Analysis & Simulation of Computer and Telecommunication Systems*. 2013, 414–23.
- Collett S. How analytics empowers business
<http://www.computerworld.com/article/2598507/how-analytics-empowers-business.html>. 2014.
- Cuzzocrea A, Song I-Y, Davis KC. Analytics over large-scale multidimensional data: the big data revolution! *ACM 14th International Workshop on Data Warehousing and OLAP*. 2011, 101–4.
- Davenport TH. Competing on analytics. *Harv Bus Rev* 2006;**84**:98.
- Dehne F, Kong Q, Rau-Chaplin A *et al.* Scalable real-time OLAP on cloud architectures. *J Parallel Distrib Comput* 2015;**79**:31–41.
- Demirkan H, Delen D. Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decis Support Syst* 2013;**55**:412–21.
- Dethier G. Syntaxe et outils de base du langage C
<http://www.montefiore.ulg.ac.be/~dethier/montef/docs/langage-c.pdf>. 2011.
- Dignan L. AWS' Vogels on hybrid cloud, mobile, big data, technical debt
<http://www.informationweek.com/big-data/big-data-analytics/databricks-cloud-next-step-for-spark/d/d-id/1278978>. 2014a.
- Dignan L. AWS: With more than 1 million active customers, we're your stack
<http://www.zdnet.com/article/aws-with-more-than-1-million-active-customers-were-your-stack/>. 2014b.
- Dinh HT, Lee C, Niyato D *et al.* A survey of mobile cloud computing: architecture, applications, and approaches: A survey of mobile cloud computing. *Wirel Commun Mob Comput* 2013;**13**:1587–611.
- Dinter, B, Lorenz A. Social Business Intelligence: a Literature Review and Research Agenda. *Thirty Third International Conference on Information Systems*. 2012.
- Dix J. How UPS Uses Analytics to Drive Down Costs (and No, it Doesn't Call it Big Data)
<http://www.cio.com/article/2852509/data-analytics/how-ups-uses-analytics-to-drive-down-costs-and-no-it-doesnt-call-it-big-data.html>. 2014.
- Duan L, Xu LD. Business Intelligence for Enterprise Systems: A Survey. *IEEE Trans Ind Inform* 2012;**8**:679–87.

- Dziembek D, Ziora L. Business Intelligence Systems in the SaaS Model as a Tool Supporting Knowledge Acquisition in the Virtual Organization,". *Online J Appl Knowl Manag* 2014;**2**:82–96.
- Eastwood B. Healthcare Cloud Use Prevalent, Poised to Spread, HIMSS Says
<http://www.computerworld.com/article/2491061/cloud-computing/healthcare-cloud-use-prevalent--poised-to-spread--himss-says.html>. 2014.
- Edwards S, Lavagno L, Lee EA *et al*. Design of embedded systems: Formal models, validation, and synthesis. *Read Hardwaresoftware Co-Des* 2001;**85**:366–390.
- Enderle R. Are Apple and Gartner Afraid of Analytics?
<http://www.computerworld.com/article/2488909/business-intelligence/are-apple-and-gartner-afraid-of-analytics-.html>. 2014a.
- Enderle R. Why Analytics Makes Tesla Better Than Jaguar
<http://www.computerworld.com/article/2490956/data-center/why-analytics-makes-tesla-better-than-jaguar.html>. 2014b.
- Endler M. Microsoft Privacy Case: What's At Stake?
<http://www.informationweek.com/government/cybersecurity/microsoft-privacy-case-whats-at-stake/d/d-id/1297752>. 2014.
- ERA. Outlet Ranking : <https://research.unsw.edu.au/excellence-research-australia-era-outlet-ranking>. 2012.
- Ereth J, Baars H. Cloud-Based Business Intelligence and Analytics Applications–Business Value and Feasibility. *PACIS Proceedings*. 2015, paper 36.
- Fernando N, Loke SW, Rahayu W. Mobile cloud computing: A survey. *Future Gener Comput Syst* 2013;**29**:84–106.
- Finnegan M. Michelin drives big data strategy with Microstrategy analytics
<http://www.computerworld.com/article/2490321/big-data/michelin-drives-big-data-strategy-with-microstrategy-analytics.html>. 2014.
- Fox GC. Large scale data analytics on clouds. *ACM Proceedings of the Fourth International Workshop on Cloud Data Management*. 2012, 21–24.
- Frécon L. *Éléments de mathématiques discrètes*. Presses Polytechniques et Universitaires Romandes. Lausanne: Presses Polytechniques et Universitaires Romandes, 2002.
- Friedman-Hill E. *Jess in Action: Rule-Based Systems in Java*. Greenwich, CT: Manning, 2003.
- Gan Y, Meng X, Shi Y. COLA: A Cloud-Based System for Online Aggregation. *IEEE 29th International Conference on Data Engineering*. 2013, 1368–71.
- Garey MR, Johnson DS. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. 27. print. New York: Freeman, 2009.

- Gartner. Marché de la Business intelligence : <http://www.gartner.com/newsroom/id/3612617>. 2017a.
- Gartner. Marché de cloud : <http://www.gartner.com/newsroom/id/3616417>. 2017b.
- Gartner IT Glossary. Business Intelligence : <http://www.gartner.com/it-glossary/business-intelligence-bi>.
- Gartner IT Glossary. <http://www.gartner.com/it-glossary/virtualization>.
- Gash D, Ariyachandra T, Frolick M. Looking to the Clouds for Business Intelligence. *J Internet Commer* 2011;**10**:261–9.
- Gatzioura A, Menychtas A, Moulos V *et al.* Incorporating Business Intelligence in Cloud Marketplaces. *10th IEEE International Symposium on Parallel and Distributed Processing with Applications*. 2012, 466–72.
- Ghoshal A, Larson E, Subramanyam R *et al.* The Impact of Business Analytics Strategy on Social, Mobile, and Cloud Computing Adoption. *Thirty Fifth Int Conf Inf Syst* 2014.
- Gillenson M, Sherrell DL, Chen L. A taxonomy of web site traversal patterns and structures. *Commun AIS* 2000;**3**:5.
- Goes PB. Design Science Research in Top Information Systems Journals. *MIS Q* 2014;**38**:iii–viii.
- Gohring N. Cloud BI: Going where the data lives
<http://www.computerworld.com/article/2491281/business-intelligence/cloud-bi-going-where-the-data-lives.html>. 2014.
- Goncalves C, Assuncao L, Cunha JC. Data analytics in the cloud with flexible MapReduce workflows. *IEEE 4th International Conference on Cloud Computing Technology and Science*. 2012, 427–434.
- Gregor S, Jones D. The anatomy of a design theory. *J Assoc Inf Syst* 2007;**8**:312.
- Grossman RL. What is analytic infrastructure and why should you care? *ACM SIGKDD Explor Newsl* 2009;**11**:5–9.
- Gruber TR. A translation approach to portable ontology specifications. *Knowl Acquis* 1993;**5**:199–220.
- Hagen CP. 5 Big Data Trends Government IT Can't Ignore
<http://www.informationweek.com/government/big-data-analytics/5-big-data-trends-government-it-cant-ignore/a/d-id/1316853>. 2014.
- Han H, Lee YC, Choi S *et al.* Cloud-aware processing of mapreduce-based olap applications. *Proceedings of the Eleventh Australasian Symposium on Parallel and Distributed Computing-Volume 140*. 2013, 31–38.

- Hanseth O, Lyytinen K. Design theory for dynamic complexity in information infrastructures: the case of building internet. *J Inf Technol* 2010;**25**:1–19.
- Hedgebeth D. Data-driven decision making for the enterprise: an overview of business intelligence applications. *VINE* 2007;**37**:414–20.
- Heller M. Learn how to crunch big data with R
<http://www.computerworld.com/article/2882819/learn-how-to-crunch-big-data-with-r.html>. 2015.
- Henschen D. Databricks Cloud: Next Step For Spark <http://www.informationweek.com/big-data/big-data-analytics/databricks-cloud-next-step-for-spark/d/d-id/1278978>. 2014a.
- Henschen D. HP Cloud Adds Big Data Options <http://www.informationweek.com/big-data/big-data-analytics/hp-cloud-adds-big-data-options/d/d-id/1317816>. 2014b.
- Henschen D. IBM Applies Analytics To Big Workforce Transitions
<http://www.informationweek.com/software/enterprise-applications/ibm-applies-analytics-to-big-workforce-transitions/d/d-id/1297758>. 2014c.
- Henschen D. IBM Offers 3 Analytic Apps Via Cloud
<http://www.informationweek.com/cloud/software-as-a-service/ibm-offers-3-analytic-apps-via-cloud/d/d-id/1252954>. 2014d.
- Henschen D. IBM Watson Analytics Goes Public <http://www.informationweek.com/big-data/big-data-analytics/ibm-watson-analytics-goes-public/d/d-id/1317887>. 2014e.
- Henschen D. IBM Watson Data Analysis Service Revealed
<http://www.informationweek.com/cloud/software-as-a-service/ibm-watson-data-analysis-service-revealed/d/d-id/1315769>. 2014f.
- Henschen D. Twitter, IBM Partner On Social Insight For Business
<http://www.informationweek.com/cloud/software-as-a-service/twitter-ibm-partner-on-social-insight-for-business/d/d-id/1317059>. 2014g.
- Henschen D. Dell Eyes \$50 Million Savings On BI, Analytics
[http://www.informationweek.com/big-data/big-data-analytics/dell-eyes-\\$50-million-savings-on-bi-analytics/d/d-id/1278581](http://www.informationweek.com/big-data/big-data-analytics/dell-eyes-$50-million-savings-on-bi-analytics/d/d-id/1278581). 2014h.
- Henschen D. SAP To Buy Fieldglass, Cloud-Based Workforce Management Firm
<http://www.informationweek.com/cloud/software-as-a-service/sap-to-buy-fieldglass-cloud-based-workforce-management-firm/d/d-id/1127932>. 2014i.
- Henschen D. SAS Switches From Predictions To Goals
<http://www.informationweek.com/big-data/big-data-analytics/sas-switches-from-predictions-to-goals/d/d-id/1316894>. 2014j.
- Henschen D. Pfizer Connects Dots To Deliver Better Treatments
<http://www.informationweek.com/strategic-cio/executive-insights-and->

innovation/pfizer-connects-dots-to-deliver-better-treatments/d/d-id/1141527?page_number=1. 2014k.

Henschen D. Oracle Unveils Hadoop Data Exploration Tool

<http://www.informationweek.com/big-data/big-data-analytics/oracle-unveils-hadoop-data-exploration-tool/d/d-id/1316198>. 2014l.

Henschen D. IBM Apps For Apple iOS: Not Siri Meets Watson

<http://www.informationweek.com/strategic-cio/digital-business/ibm-apps-for-apple-ios-not-siri-meets-watson/d/d-id/1318041>. 2014m.

Henschen D. Splunk Expands Amazon Cloud, Mobile Options

<http://www.informationweek.com/big-data/big-data-analytics/splunk-expands-amazon-cloud-mobile-options/d/d-id/1316414>. 2014n.

Henschen D. 10 Cloud Analytics & BI Platforms For Business

<http://www.informationweek.com/cloud/software-as-a-service/10-cloud-analytics-and-bi-platforms-for-business/d/d-id/1318724>. 2015a.

Henschen D. 5 Analytics, BI, Data Management Trends For 2015

<http://www.informationweek.com/big-data/big-data-analytics/5-analytics-bi-data-management-trends-for-2015/a/d-id/1318551>. 2015b.

Henschen D. Microsoft Power BI Makeover: 6 Big Improvements

<http://www.informationweek.com/cloud/software-as-a-service/microsoft-power-bi-makeover-6-big-improvements/d/d-id/1318815>. 2015c.

Henschen D. Salesforce Marketing Cloud Gets Predictive

<http://www.informationweek.com/software/enterprise-applications/salesforce-marketing-cloud-gets-predictive/d/d-id/1319353>. 2015d.

Henschen D. IBM Mainframe Makeover: Mobile, Big Data Reality Check

<http://www.informationweek.com/big-data/big-data-analytics/ibm-mainframe-makeover-mobile-big-data-reality-check/a/d-id/1318620>. 2015e.

Henschen D. Apple Watch Meets Salesforce Enterprise Apps

<http://www.informationweek.com/cloud/software-as-a-service/apple-watch-meets-salesforce-enterprise-apps/d/d-id/1319450>. 2015f.

Henschen D. Microsoft Buying Revolution Analytics For Deeper Data Analysis

<http://www.informationweek.com/big-data/big-data-analytics/microsoft-buying-revolution-analytics-for-deeper-data-analysis/d/d-id/1318754>. 2015g.

Henschen D. SAP Cuts Data Visualization Cost, Unites Predictive Suite

<http://www.informationweek.com/big-data/big-data-analytics/sap-cuts-data-visualization-cost-unites-predictive-suite/d/d-id/1319468>. 2015h.

Henschen D. IBM Watson Cloud Gains Eyes, Ears, And A Voice

<http://www.informationweek.com/software/enterprise-applications/ibm-watson-cloud-gains-eyes-ears-and-a-voice/d/d-id/1319008>. 2015i.

- Henschen D. Spark Promoter Databricks Should Let Software Shine
<http://www.informationweek.com/big-data/big-data-analytics/spark-promoter-databricks-should-let-software-shine/a/d-id/1319539>.
- Henschen D. Microsoft's Revolution Analytics Deal Is Finalized
<http://www.informationweek.com/cloud/platform-as-a-service/microsofts-revolution-analytics-deal-is-finalized/d/d-id/1319788>.
- Herwig V. Business intelligence as a service for Cloud-based applications. *IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems*. Vol 1. 2013, 188–192.
- Herzfeldt A, Hausen M, Briggs RO *et al*. Developing a Risk Management Process and Risk Taxonomy for Medium-Sized IT Solution Providers. *ECIS Proceedings*. 2012, paper 165.
- Hevner AR, March ST, Park J *et al*. Design science in information Systems research. *MIS Q* 2004;**28**:75–105.
- Hinchcliffe D. Dreamforce 2014: Converging on cloud, apps, mobile, analytics, and community <http://www.zdnet.com/article/dreamforce-2014-converging-on-cloud-apps-mobile-analytics-and-community/>. 2014.
- Hua Y, Liu X, Jiang H. ANTELOPE: A Semantic-Aware Data Cube Scheme for Cloud Data Center Networks. *IEEE Trans Comput* 2014;**63**:2146–59.
- Inmon WH. *Building the Data Warehouse*. 3rd ed. New York: J. Wiley, 2002.
- ISO/IEC 14977. Information technology - Syntactic metalanguage - Extended BNF. 1996.
- ISO/IEC/IEEE. Systems and Software Engineering – Vocabulary. 2010.
- Jarke M, Loucopoulos P, Lyytinen K *et al*. The brave new world of design requirements. *Inf Syst* 2011;**36**:992–1008.
- Juan-Verdejo A, Surajbali B, Baars H *et al*. Moving business intelligence to cloud environments. *IEEE Conference on Computer Communications Workshops*. 2014, 43–48.
- Jun L, Jun W. Cloud Computing Based Solution to Decision Making. *Procedia Eng* 2011;**15**:1822–6.
- Kambatla K, Kollias G, Kumar V *et al*. Trends in big data analytics. *J Parallel Distrib Comput* 2014;**74**:2561–73.
- Kanaracus C. Tibco scoops up analytics vendor Jaspersoft for \$185 million
<http://www.pcworld.com/article/2149040/tibco-scoops-up-analytics-vendor-jaspersoft-for-185-million.html>. 2014.
- Kasem M, Hassanein EE. Cloud Business intelligence survey. *Int J Comput Appl* 2014;**90**.

- Kaur H, Agrawal P, Dhiman A. Visualizing Clouds on Different Stages of DWH-An Introduction to Data Warehouse as a Service. *IEEE International Conference on Computing Sciences*. 2012, 356–359.
- Khan AN, Mat Kiah ML, Khan SU *et al.* Towards secure mobile cloud computing: A survey. *Future Gener Comput Syst* 2013;**29**:1278–99.
- Kitchenham B. Procedures for performing systematic reviews. *UK Keele Univ* 2004;**33**:1–26.
- Kitchenham B, Brereton P. A systematic review of systematic review process research in software engineering. *Inf Softw Technol* 2013;**55**:2049–75.
- Knabke T, Olbrich S. Exploring the Future Shape of Business Intelligence: Mapping Dynamic Capabilities of Information Systems to Business Intelligence Agility. *Twenty-First Americas Conference on Information Systems*. Puerto Rico, 2015.
- Kornysheva E, Deneckère R, Salinesi C. Method chunks selection by multicriteria techniques: an extension of the assembly-based approach. *Situational Method Engineering: Fundamentals and Experiences*. Springer, 2007, 64–78.
- Kowalczyk M, Buxmann P, Besier J. Investigating business intelligence and analytics from a decision process perspective: A structured literature review. *Proceedings of the 21st European Conference on Information Systems*. 2013.
- Kurunji S, Ge T, Liu B *et al.* Communication cost optimization for cloud Data Warehouse queries. *IEEE 4th International Conference on Cloud Computing Technology and Science*. 2012, 512–519.
- Lee J, Feng T, Shi W *et al.* Towards Quality Aware Collaborative Video Analytic Cloud. *IEEE Fifth International Conference on Cloud Computing*. 2012, 147–54.
- Leimeister S, Riedl C, Böhm M *et al.* The Business perspective of cloud computing: Actors, roles, and value networks. *ECIS Proceedings*. 2010, Paper 56.
- Mahmud S, Iqbal R, Doctor F. Cloud enabled data analytics and visualization framework for health-shocks prediction. *Future Gener Comput Syst* 2016;**65**:169–81.
- MapReduce. <http://blog.cloudera.com/blog/2014/03/the-truth-about-mapreduce-performance-on-ssds/>. 2017.
- March ST, Smith G. Design and Natural Science Research on Information Technology. *Decis Support Syst* 1995;**15**:251–66.
- McTigue J. Cloud Holdouts Face Tipping Point
<http://www.informationweek.com/cloud/platform-as-a-service/cloud-holdouts-face-tipping-point-/d/d-id/1204507>. 2014.
- Mian R, Martin P, Vazquez-Poletti JL. Provisioning data analytic workloads in a cloud. *Future Gener Comput Syst* 2013;**29**:1452–8.

- Mikael Ricknäs. Microsoft leans on data-center strength to get cloud edge
<http://www.computerworld.com/article/2839602/microsoft-leans-on-data-center-strength-to-get-cloud-edge.html>. 2014.
- Minelli M, Chambers M, Dhiraj A. *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*. Hoboken: John Wiley & Sons, 2012.
- Mitchell RL. 8 big trends in big data analytics
<http://www.computerworld.com/article/2690856/8-big-trends-in-big-data-analytics.html>. 2014.
- Moore J. Business Intelligence Takes to Cloud for Small Businesses
<http://www.computerworld.com/article/2490263/enterprise-applications/business-intelligence-takes-to-cloud-for-small-businesses.html>. 2014.
- Moro S, Cortez P, Rita P. Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Syst Appl* 2015;**42**:1314–24.
- Morphy E. Why video-based shopper analytics will bow to mobile
<http://www.computerworld.com/article/2849467/big-data-analytics/why-its-inevitable-video-based-shopper-analytics-will-bow-to-mobile.html>. 2014.
- Muriithi GM, Kotzé JE. A conceptual framework for delivering cost effective business intelligence solutions as a service. *ACM Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*. 2013, 96–100.
- Nayem R, Fahad A, Akhter S. Emerging technologies in business intelligence. *IEEE Proceedings of PICMET'13: Technology Management in the IT-Driven Services*. 2013, 542–547.
- Negash S. Business Intelligence. *Commun Assoc Informati Syst* 2004;**13**:177–95.
- Ng WS, Kirchberg M, Bressan S *et al*. Towards a privacy-aware stream data management system for cloud applications. *Int J Web Grid Serv* 2011;**7**:246–67.
- Nickerson RC, Varshney U, Muntermann J. A method for taxonomy development and its application in information systems. *Eur J Inf Syst* 2013;**22**:336–359.
- Noyes K. IBM to pump \$4B into cloud, mobile and analytics this year
<http://www.computerworld.com/article/2889912/ibm-to-pump-4b-into-cloud-mobile-and-analytics-this-year.html>. 2015a.
- Noyes K. Salesforce brings more analytics power to mobile business users
<http://www.computerworld.com/article/2886674/salesforce-brings-more-analytics-power-to-mobile-business-users.html>. 2015b.
- Noyes K. Salesforce gives Service Cloud a shot of data science
<http://www.computerworld.com/article/2893310/salesforce-gives-service-cloud-a-shot-of-data-science.html>. 2015c.

- Nunamaker JF, Derrick DC, Elkins AC *et al.* Embodied Conversational Agent-Based Kiosk for Automated Interviewing. *J Manag Inf Syst* 2011;**28**:17–48.
- O'Connor F. Regulation on cloud security may spur SaaS use in health care
<http://www.computerworld.com/article/2837836/security0/regulation-on-cloud-security-may-spur-saas-use-in-health-care.html>. 2014a.
- O'Connor F. Tidemark software adds predictive analytics to help CFOs use big data in forecasting, budgeting
<http://www.pcworld.com/article/2846952/tidemark-software-adds-predictive-analytics-to-help-cfos-use-big-data-in-forecasting-budgeting.html>. 2014b.
- O'Connor F. Tidemark adds predictive analytics to help CFOs use big data in forecasting, budgeting
<http://www.computerworld.com/article/2846010/tidemark-adds-predictive-analytics-to-help-cfos-use-big-data-in-forecasting-budgeting.html>. 2014c.
- Offermann P, Blom S, Schönherr M *et al.* Artifact types in information systems design science—a literature review. *International Conference on Design Science Research in Information Systems*. Springer, 2010, 77–92.
- Okoli C, Schabram K. A guide to conducting a systematic literature review of information systems research. *Sprouts Work Pap Inf Syst* 2010;**10**:26.
- Olavsrud T. Enterprises Moving Big Data Workloads to Public Cloud
<http://www.cio.com/article/2846449/big-data/enterprises-moving-big-data-workloads-to-public-cloud.html>. 2014.
- Oltsik J. Big data security analytics mantra
<http://www.computerworld.com/article/2491445/data-center/big-data-security-analytics-mantra--collect-and-analyze-everything.html>. 2014.
- Overby S. How Little Debbie Used Predictive Analytics to Survive Snack Industry Shakeup
<http://www.cio.com/article/2377801/software-as-a-service/how-little-debbie-used-predictive-analytics-to-survive-snack-industry-shakeup.html>. 2014.
- Patel I, Rau-Chaplin A, Varghese B. A Platform for Parallel R-based Analytics on Cloud Infrastructure. *IEEE 41st International Conference on Parallel Processing Workshops*. 2012, 188–193.
- Phillips-Wren G, Iyer LS, Kulkarni U *et al.* Business Analytics in the Context of Big Data: A Roadmap for Research. *Commun Assoc Inf Syst* 2015;**37**:23.
- Power DJ. Mobile decision support and business intelligence: an overview☆. *J Decis Syst* 2013;**22**:4–9.
- Prat N, Comyn-Wattiau I, Akoka J. A Taxonomy of Evaluation Methods for Information Systems Artifacts. *J Manag Inf Syst* 2015;**32**:229–67.

- Pratt MK. University CIO helps boost graduation rates with analytics
<http://www.computerworld.com/article/2490005/it-management/university-cio-helps-boost-graduation-rates-with-analytics.html>. 2014.
- Purao S, Vaishnavi V. Product metrics for object-oriented systems. *ACM Comput Surv CSUR* 2003;**35**:191–221.
- Qian L, Luo Z, Du Y *et al*. Cloud Computing: An Overview. *IEEE International Conference on Cloud Computing*. Springer Berlin Heidelberg, 2009, 626–31.
- Rahman M, Li X, Palit H. Hybrid heuristic for scheduling data analytics workflow applications in hybrid cloud environment. *IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum*. IEEE, 2011, 966–74.
- Ram J, Zhang C, Koronios A. The Implications of Big Data Analytics on Business Intelligence: A Qualitative Study in China. *Procedia Comput Sci* 2016;**87**:221–6.
- Reix R. *Systèmes d'information et management des organisations*. Paris: Vuibert, 2000.
- Ricknäs M. IBM earnings slump as company shifts to cloud, mobile and analytics
<http://www.computerworld.com/article/2835719/ibm-earnings-slump-as-company-shifts-to-cloud-mobile-and-analytics.html>. 2014.
- Rimal BP, Choi E, Lumb I. A Taxonomy and Survey of Cloud Computing Systems. *IEEE Fifth International Joint Conference on INC, IMS and IDC*. 2009, 44–51.
- Ryan MD. Cloud computing security: The scientific challenge, and a survey of solutions. *J Syst Softw* 2013;**86**:2263–8.
- Sangupamba Mwilu O, Comyn-Wattiau I, Prat N. Design science research contribution to business intelligence in the cloud — A systematic literature review. *Future Gener Comput Syst* 2016;**63**:108–22.
- Sangupamba Mwilu O, Prat N, Comyn-Wattiau I. Business Intelligence and Big Data in the Cloud: Opportunities for Design-Science Researchers. *Advances in Conceptual Modeling—Proceedings of ER 2014 Workshops*. GA, USA, 2014, 75–84.
- Sankappanavar HP, Burris S. A course in universal algebra. *Grad Texts Math* 1981;**78**.
- Shanks G, Sharma R, Seddon P *et al*. The impact of strategy and maturity on business analytics and firm performance: A review and research agenda. *Proc 21st Australia Conf. Inf. Syst.* 2010, paper 51.
- Silver MS. Decisional Guidance for Computer-Based Decision Support. *MIS Q* 1991:105--122.
- Simmhan Y, Aman S, Kumbhare A *et al*. Cloud-based software platform for big data analytics in smart grids. *Comput Sci Eng* 2013;**15**:38–47.

- Stanoevska-Slabeva K, Wozniak T, Ristol S eds. *Grid and Cloud Computing: A Business Perspective on Technology and Applications*. Heidelberg: Springer, 2010.
- Subashini S, Kavitha V. A survey on security issues in service delivery models of cloud computing. *J Netw Comput Appl* 2011;**34**:1–11.
- Sun X, Gao B, Fan L *et al*. A cost-effective approach to delivering analytics as a service. *19th International Conference on Web Services*. IEEE, 2012, 512–9.
- Tan W, Blake MB, Saleh I *et al*. Social-Network-Sourced Big Data Analytics. *IEEE Internet Comput* 2013;**17**:62–9.
- Tancer J, Varde AS. The Deployment of MML for Data Analytics over the Cloud. *11th IEEE International Conference on Data Mining Workshops*. 2011, 188–95.
- Tapiador D, O’Mullane W, Brown AGA *et al*. A framework for building hypercubes using MapReduce. *Comput Phys Commun* 2014;**185**:1429–1438.
- The Chief Information Officers Council. Federal Enterprise Architecture Framework. 1999.
- Thompson WJ, Van der Walt JS. Business intelligence in the cloud. *SA J Inf Manag* 2010;**12**:5–pages.
- Vance J. 10 Hot Hadoop Startups to Watch
<http://www.computerworld.com/article/2488369/enterprise-applications/10-hot-hadoop-startups-to-watch.html>. 2014a.
- Vance J. Big Data Wars: How Technology Could Tip the Mid-Term Elections
<http://www.cio.com/article/2838806/big-data/big-data-wars-how-technology-could-tip-the-mid-term-elections.html?page=2>. 2014b.
- Wang Y, Song A, Luo J. A MapReduceMerge-based Data Cube Construction Method. *IEEE Ninth International Conference on Grid and Cloud Computing*. 2010, 1–6.
- Williams G. *Data Mining with R and Rattle*. New York, NY: Springer New York, 2011.
- Williams M. Sony shows off latest SmartEyeglasses prototype
<http://www.computerworld.com/article/2603128/gadgets/sony-shows-off-latest-smarteyeglasses-prototype.html>. 2014.
- Wlodarczyk TW, Rong C, Jia B *et al*. DataStorm An Ontology-Driven Framework for Cloud-Based Data Analytic Systems. *IEEE 6th World Congress on Services*. 2010, 123–7.
- Wolpe T. SAP HANA: Users are struggling to see the business case
<http://www.zdnet.com/article/sap-hana-users-are-struggling-to-see-the-business-case/>. 2014.
- Xiufeng L, Thomsen C, Pedersen TB. CloudETL: scalable dimensional ETL for hive. *Proceedings of the 18th International Database Engineering & Applications Symposium*. ACM, 2014, 195–206.

- Xu M, Gao D, Deng C *et al.* Cloud computing boosts business intelligence of telecommunication industry. *International Conference on Cloud Computing*. Springer, 2009, 224–31.
- Yang H, Tate M. A descriptive literature review and classification of cloud computing research. *Commun Assoc Inf Syst* 2012;**31**:35–60.
- Yuan B, Herbert J. A Cloud-Based Mobile Data Analytics Framework: Case Study of Activity Recognition Using Smartphone. *IEEE Mobile Cloud Computing, Services, and Engineering*. 2014, 220–227.

ANNEXES

I. CLOUD BI FRAMEWORK

Artifacts	Data collection and blending	Data modeling and storage	Analytics	BI as a whole
Language		<ul style="list-style-type: none"> Medical Markup language (Tancer and Varde 2011) 		
Concept			<ul style="list-style-type: none"> BlaaS concept (Chang 2014) 	
System design		<ul style="list-style-type: none"> Data structure in the cloud (Kaur, Agrawal and Dhiman 2012) 	<ul style="list-style-type: none"> Conceptual design of a platform for analytics on the cloud (Patel, Rau-Chaplin and Varghese 2012) 	<ul style="list-style-type: none"> Business process for incorporating BI in cloud market (Gatzioura <i>et al.</i> 2012) Business process for a generic value network (Leimeister <i>et al.</i> 2010)
Taxonomy			<ul style="list-style-type: none"> Taxonomy of data mining algorithms (Duan and Xu 2012) 	<ul style="list-style-type: none"> Taxonomy of mobile cloud computing (Fernando, Loke and Rahayu 2013) Taxonomy of main aspects for cloud monitoring (Leimeister <i>et al.</i> 2010)

Framework	<ul style="list-style-type: none"> • Framework for data integration (Negash 2004) • Framework for provisioning resources from a public cloud (Mian, Martin and Vazquez-Poletti 2013) 	<ul style="list-style-type: none"> • Framework for hypercube generation (Tapiador <i>et al.</i> 2014) • framework to support profit-oriented decisions in database systems in the cloud (Chi <i>et al.</i> 2011) 	<ul style="list-style-type: none"> • The framework for composing, managing, and processing workflows for big data analytics in cloud computing (Chaisiri <i>et al.</i> 2012) • BI and OLAP framework (Al-Aqrabi <i>et al.</i> 2015) (Al-Aqrabi <i>et al.</i> 2012) • Framework for cloud data analytic (Wlodarczyk <i>et al.</i> 2010) 	<ul style="list-style-type: none"> • Framework of cloud BI (Juan-Verdejo <i>et al.</i> 2014) • Framework of social BI research agenda (Dinter, and Lorenz 2012) • Framework of cloud BI services (Baars and Kemper 2010) • Framework for BI&A technology (Shanks <i>et al.</i> 2010) • Framework for healthcare delivery (Yuan and Herbert 2014) • Framework for BI research (Kowalczyk, Buxmann and Besier 2013) • Framework to verify security in the cloud (Subashini and Kavitha 2011)
Architecture		<ul style="list-style-type: none"> • Cloud data warehouse architecture (Kaur, Agrawal and Dhiman 2012) • Elastic data storage architecture (Cao <i>et al.</i> 	<ul style="list-style-type: none"> • Architecture for online aggregation in the cloud (Gan, Meng and Shi 2013) • Architecture of analytics on the cloud (Stanoevska-Slabeva, Wozniak 	<ul style="list-style-type: none"> • Architecture of cloud BI (Kasem and Hassanein 2014) • BI architecture (Chaudhuri, Dayal and Narasayya

		<p>2011)</p> <ul style="list-style-type: none"> • Architecture for building OLAP cubes in the cloud (Arres, Kabbachi and Boussaid 2013) 	<p>and Ristol 2010)</p> <ul style="list-style-type: none"> • Architecture to maintain the elastic OLAP cloud platform (Brezany <i>et al.</i> 2011b) • Architecture for scalable runtime optimized of cloud data analytics (Gatzioura <i>et al.</i> 2012) • Architecture of a system for data analytics in the cloud (Garey and Johnson 2009) 	<p>2011)</p> <ul style="list-style-type: none"> • Cloud computing architecture (Qian <i>et al.</i> 2009) • Architecture of objective-driven MapReduce (AbdelBaky <i>et al.</i> 2012) • DSS Architecture (Demirkan and Delen 2013) • Architecture of a security system (Ng <i>et al.</i> 2011) • Architecture of mobile cloud computing (Khan <i>et al.</i> 2013) (Dinh <i>et al.</i> 2013)
Requirement	<ul style="list-style-type: none"> • ETL requirement (Herwig 2013) 		<ul style="list-style-type: none"> • BaaS requirement (Chang 2014) 	<ul style="list-style-type: none"> • DSS requirement (Jun and Jun 2011) • Requirement for security system in the cloud (Ng <i>et al.</i> 2011)
Methodology	<ul style="list-style-type: none"> • Methodology for data integrity verification (Gash, Ariyachandra and Frolick 2011) 			<ul style="list-style-type: none"> • Methodology for BI migration (Gash, Ariyachandra and Frolick 2011) (Muriithi and Kotzé 2013)

Algorithm		<ul style="list-style-type: none"> • Algorithm for building parallels cubes based on MapReduceMerge (Wang, Song and Luo 2010) • Resources planning algorithm (Lee <i>et al.</i> 2012) • Algorithm for building cube with MapReduce (Abelló, Ferrarons and Romero 2011) 	<ul style="list-style-type: none"> • Algorithm to process multiple join queries in the cloud (Kurunji <i>et al.</i> 2012) • Aggregate query algorithm (Hua, Liu and Jiang 2014) 	
Implemented system			<ul style="list-style-type: none"> • Tool for analysis in the cloud (Stanoevska-Slabeva, Wozniak and Ristol 2010) • Tool for online aggregation in the cloud (Gan, Meng and Shi 2013) • Tool for scalable runtime optimized of cloud data analytics (Barga, Ekanayake and Lu 2012) • Tool to maintain the elastic OLAP cloud platform (Brezany <i>et al.</i> 2011b) • Tool to combine the scale of Google App Engine with offline data analytics (Chohan <i>et al.</i> 2012) 	<ul style="list-style-type: none"> • Tool for incorporating BI in cloud market (Gatzioura <i>et al.</i> 2012) • Tool for application reconfiguration (Goncalves, Assuncao and Cunha 2012)
Example			<ul style="list-style-type: none"> • Example of building OLAP on cloud (Arres, Kabbachi and Boussaid 2013) 	<ul style="list-style-type: none"> • Example of security measures (Ryan 2013)

II. EXAMEN DES SOURCES DE TYPE 2 ET TYPE 3

2.1. Les sources de type 2

Ces sources sont constituées par neuf articles sur la BI et dix articles sur le cloud computing.

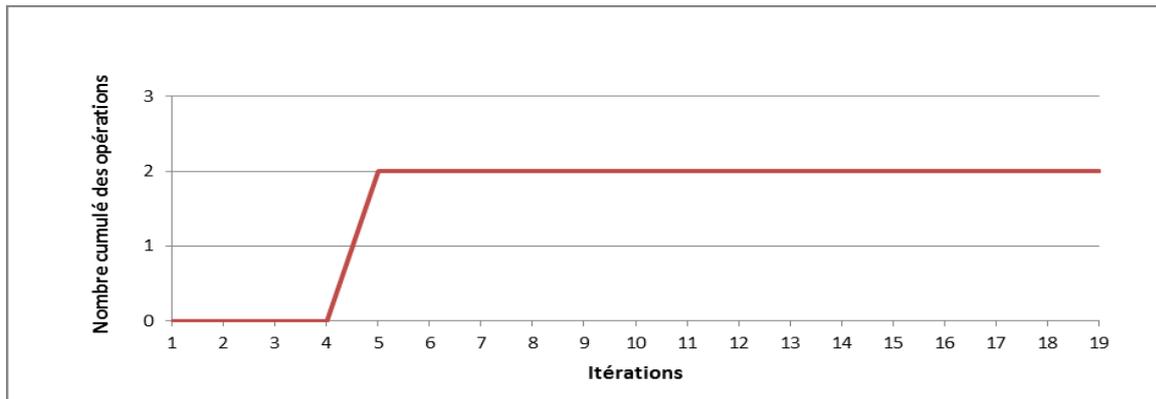
2.1.1. Taxonomie de contexte

- Itération 1 (Subashini and Kavitha 2011) : l'article confirme la dimension *exigences de sécurité*.
- Itération 2 (Rimal, Choi and Lumb 2009) : l'article confirme la dimension *exigences de sécurité*.
- Itération 3 (Dinh *et al.* 2013) : l'article n'ajoute rien.
- Itération 4 (Negash 2004) : l'article n'ajoute rien.
- Itération 5 (Fernando, Loke and Rahayu 2013) : l'article confirme la dimension *exigences de sécurité*.
- Itération 6 (Chaudhuri, Dayal and Narasayya 2011) : l'article n'ajoute rien.
- Itération 7 (Leimeister *et al.* 2010) : l'article confirme la dimension *objectifs*.
- Itération 8 (Qian *et al.* 2009) : l'article n'ajoute rien.
- Itération 9 (Duan and Xu 2012) : l'article n'ajoute rien.
- Itération 10 (Khan *et al.* 2013) : l'article confirme la dimension *exigences de sécurité*.
- Itération 11 (Aceto *et al.* 2013) : l'article n'ajoute rien.
- Itération 12 (Ryan 2013) : l'article confirme la dimension *exigences de sécurité*.
- Itération 13 (Yang and Tate 2012) : l'article confirme la dimension *exigences de sécurité*.
- Itération 14 (Hedgebeth 2007) : l'article confirme la dimension *exigences de sécurité*.
- Itération 15 (Shanks *et al.* 2010) : l'article n'ajoute rien.
- Itération 16 (Moro, Cortez and Rita 2015) : l'article confirme la dimension *exigences de sécurité* et la dimension *secteur d'activité*.
- Itération 17 (Kowalczyk, Buxmann and Besier 2013) : l'article n'ajoute rien.
- Itération 18 (Power 2013) : l'article n'ajoute rien.
- Itération 19 (Dinter, and Lorenz 2012): l'article n'ajoute rien.

2.1.2. Taxonomie de décision

- Itération 1 (Subashini and Kavitha 2011) : l'article confirme la dimension *couches de cloud* et la dimension *types de cloud* ainsi que la sous-catégorie *outils*.
- Itération 2 (Rimal, Choi and Lumb 2009) : l'article confirme la dimension *couches de cloud*.
- Itération 3 (Dinh *et al.* 2013) : l'article confirme la dimension *couches de cloud*.
- Itération 4 (Negash 2004) : l'article confirme la sous-catégorie *solution*.
- Itération 5 (Fernando, Loke and Rahayu 2013) : l'article permet d'ajouter la dimension mobile cloud avec les caractéristiques smartphone, tablette et ordinateur portable.
- Itération 6 (Chaudhuri, Dayal and Narasayya 2011) : l'article confirme la sous-catégorie *solution* et la sous-catégorie *Outils*.
- Itération 7 (Leimeister *et al.* 2010) : l'article confirme la dimension *couches de cloud*.
- Itération 8 (Qian *et al.* 2009) : l'article confirme la dimension *couches de cloud*.
- Itération 9 (Duan and Xu 2012) : l'article confirme la sous-catégorie *outils* et la sous-catégorie *solution*.
- Itération 10 (Khan *et al.* 2013) : l'article confirme la dimension *couches de cloud*.
- Itération 11 (Aceto *et al.* 2013) : l'article confirme la dimension *couches de cloud* et la dimension *types de cloud*.
- Itération 12 (Ryan 2013) : l'article n'ajoute rien.
- Itération 13 (Yang and Tate 2012) : l'article confirme la dimension *couches de cloud* et la dimension *types de cloud*.
- Itération 14 (Hedgebeth 2007) : l'article confirme la dimension *couches de BI*.
- Itération 15 (Shanks *et al.* 2010) : l'article n'ajoute rien.
- Itération 16 (Moro, Cortez and Rita 2015) : l'article n'ajoute rien.
- Itération 17 (Kowalczyk, Buxmann and Besier 2013) : l'article confirme la sous-catégorie *solution*.
- Itération 18 (Power 2013) : l'article confirme la dimension *supports des applications BI mobiles*.
- Itération 19 (Dinter, and Lorenz 2012) : l'article confirme la sous-catégorie *solution*.

2.1.3. Graphe des opérations sur la taxonomie



2.2. Les sources de type 3

2.2.1. Phase 1 : test pilote constitué par les 34 articles de 5 différents magazines informatiques.

2.2.1.1. Taxonomie de contexte

- Itération 1 (Henschen 2015a) : L'article requiert la promotion de la caractéristique *réduction des coûts*. Il ajoute les caractéristiques *coûts d'infrastructure*, *coûts d'implémentation* et *coûts d'administration* à la catégorie *réduction des coûts*. Il permet d'ajouter les caractéristiques *mobilité* et *agilité* à la dimension *objectifs*.
- Itération 2 (Vance 2014a) : l'article n'ajoute rien.
- Itération 3 (Henschen 2015b) : l'article n'ajoute rien.
- Itération 4 (Mitchell 2014) : l'article n'ajoute rien.
- Itération 5 (Enderle 2014a) : l'article n'ajoute rien.
- Itération 6 (Dignan 2014a) : l'article confirme la dimension *exigences de sécurité*.
- Itération 7 (Dignan 2014b) : l'article confirme la dimension *exigences de sécurité*.
- Itération 8 (Vance 2014b) : l'article n'ajoute rien.
- Itération 9 (Bertolucci 2015) : l'article n'ajoute rien.
- Itération 10 (Babcock 2014a) : l'article n'ajoute rien.

- Itération 11 (Moore 2014) : l'article confirme la dimension *objectifs* et la dimension *secteur d'activité*. Il permet d'ajouter la caractéristique *grande* à la dimension *taille de l'entreprise*.
- Itération 12 (Gohring 2014) : l'article confirme la dimension *exigences de sécurité*.
- Itération 13 (Henschen 2014a) : l'article n'ajoute rien.
- Itération 14 (Hinchcliffe 2014) : l'article n'ajoute rien.
- Itération 15 (Olavsrud 2014) : l'article confirme les dimensions *objectifs* et *exigences de sécurité*.
- Itération 16 (Baxter-Reynolds 2014) : l'article n'ajoute rien.
- Itération 17 (Babcock 2014b) : l'article n'ajoute rien.
- Itération 18 (Overby 2014) : l'article n'ajoute rien.
- Itération 19 (Dix 2014) : l'article n'ajoute rien.
- Itération 20 (Henschen 2014b) : l'article n'ajoute rien.
- Itération 21 (Henschen 2014c) : l'article n'ajoute rien.
- Itération 22 (Ricknäs 2014) : l'article confirme la dimension *exigences de sécurité*.
- Itération 23 (Henschen 2014d) : l'article n'ajoute rien.
- Itération 24 (Noyes 2015a) : l'article n'ajoute rien.
- Itération 25 (Henschen 2014e) : l'article n'ajoute rien.
- Itération 26 (Henschen 2014f) : l'article confirme la dimension *exigences de sécurité*.
- Itération 27 (Carlos Perez 2014) : l'article n'ajoute rien.
- Itération 28 (Mikael Ricknäs 2014) : l'article n'ajoute rien.
- Itération 29 (Henschen 2015c) : l'article n'ajoute rien.
- Itération 30 (O'Connor 2014a) : l'article n'ajoute rien.
- Itération 31 (Noyes 2015b) : l'article n'ajoute rien.
- Itération 32 (Wolpe 2014) : l'article confirme la dimension *exigences de sécurité*.
- Itération 33 (Kanaracus 2014) : l'article n'ajoute rien.
- Itération 34 (O'Connor 2014b) : l'article n'ajoute rien.

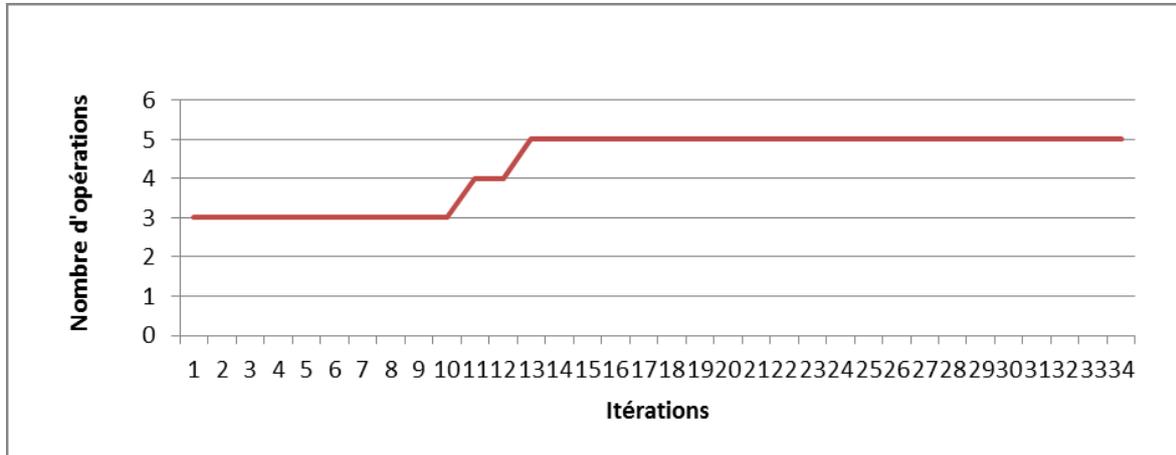
2.2.1.2. Taxonomie de décision

- Itération 1 (Henschen 2015a) : l'article confirme la dimension *niveau de migration*.
- Itération 2 (Vance 2014a) : l'article confirme la sous-catégorie *outils*.
- Itération 3 (Henschen 2015b) : l'article confirme la sous-catégorie *outils*.

- Itération 4 (Mitchell 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 5 (Enderle 2014a) : l'article confirme la sous-catégorie *outils*.
- Itération 6 (Dignan 2014a) : l'article confirme la sous-catégorie *outils* et la dimension *types de cloud*.
- Itération 7 (Dignan 2014b) : l'article confirme la dimension *couches de cloud*.
- Itération 8 (Vance 2014b) : l'article confirme la sous-catégorie *outils*.
- Itération 9 (Bertolucci 2015) : l'article confirme la sous-catégorie *outils*.
- Itération 10 (Babcock 2014a) : l'article confirme la sous-catégorie *outils*.
- Itération 11 (Moore 2014) : l'article confirme la sous-catégorie *solution*.
- Itération 12 (Gohring 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 13 (Henschen 2014a) : l'article confirme la sous-catégorie *service*. Il permet d'ajouter la caractéristique *Spark* à la sous-catégorie *BI composite*. Il confirme la caractéristique stockage *cassandra*.
- Itération 14 (Hinchcliffe 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 15 (Olavsrud 2014) : l'article confirme la dimension *types de cloud*.
- Itération 16 (Baxter-Reynolds 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 17 (Babcock 2014b) : l'article confirme la sous-catégorie *outils*.
- Itération 18 (Overby 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 19 (Dix 2014) : l'article confirme la dimension *types de cloud*.
- Itération 20 (Henschen 2014b) : l'article confirme la catégorie *outils*.
- Itération 21 (Henschen 2014c) : l'article confirme la sous-catégorie *outils*.
- Itération 22 (Ricknäs 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 23 (Henschen 2014d) : l'article confirme la sous-catégorie *outils*.
- Itération 24 (Noyes 2015a) : l'article confirme la sous-catégorie *IaaS*.
- Itération 25 (Henschen 2014e) : l'article confirme la sous-catégorie *outils*.
- Itération 26 (Henschen 2014f) : l'article confirme la sous-catégorie *outils*.
- Itération 27 (Carlos Perez 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 28 (Mikael Ricknäs 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 29 (Henschen 2015c) : l'article confirme la sous-catégorie *outils*.
- Itération 30 (O'Connor 2014a) : l'article confirme la sous-catégorie *outils*.
- Itération 31 (Noyes 2015b) : l'article confirme la sous-catégorie *outils*.
- Itération 32 (Wolpe 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 33 (Kanaracus 2014) : l'article confirme la dimension *couches de cloud*.

- Itération 34 (O'Connor 2014b) : l'article confirme la dimension *types de cloud* et la sous-catégorie *outils*.

2.2.1.3. Graphe des opérations sur la taxonomie



- 2.2.2. Phase 2 : recherche approfondie sur les deux magazines : "InformationWeek" et "ComputerWorld" qui ont fourni des informations conduisant à la modification de la taxonomie pendant le test pilote**

2.2.2.1. Taxonomie de contexte

- Itération 1 (Collett 2014) : l'article confirme la dimension *secteur d'activité* et la catégorie *réduction des coûts*.
- Itération 2 (Henschen 2014g) : l'article n'ajoute rien.
- Itération 3 (Babcock 2014c) : l'article confirme la dimension *secteur d'activité*.
- Itération 4 (Henschen 2014h) : l'article n'ajoute rien.
- Itération 5 (Eastwood 2014) : l'article confirme la dimensions *exigences de sécurité* et *secteur d'activité* ainsi que la catégorie *réduction des coûts*.
- Itération 6 (Babcock 2014d) : l'article n'ajoute rien.
- Itération 7 (Pratt 2014) : l'article n'ajoute rien.
- Itération 8 (Henschen 2014i) : l'article confirme la caractéristique *performance des services*.
- Itération 9 (Morphy 2014) : l'article n'ajoute rien.
- Itération 10 (Oltsik 2014) : l'article confirme la dimension *exigences de sécurité*.
- Itération 11 (Babcock 2015) : l'article n'ajoute rien.

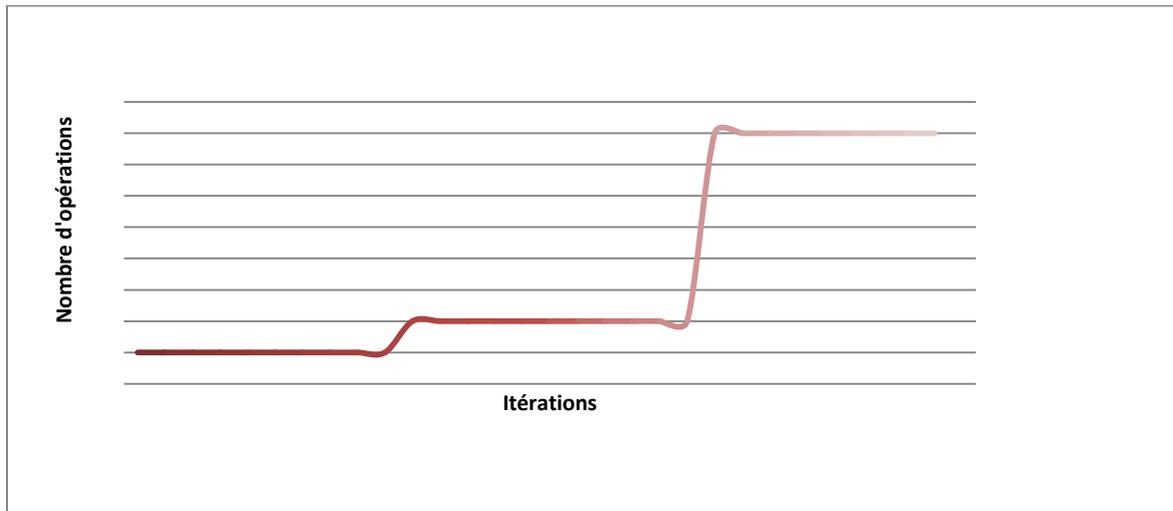
- Itération 12 (Henschen 2015d) : l'article n'ajoute rien.
- Itération 13 (Enderle 2014b) : l'article n'ajoute rien.
- Itération 14 (Bertolucci 2014a) : l'article confirme la dimension *ressources humaines*.
- Itération 15 (Noyes 2015c) : l'article n'ajoute rien.
- Itération 16 (Finnegan 2014) : l'article n'ajoute rien.
- Itération 17 (Henschen 2014j) : l'article n'ajoute rien.
- Itération 18 (Henschen 2014k) : l'article n'ajoute rien.
- Itération 19 (Hagen 2014) : l'article n'ajoute rien.
- Itération 20 (Bertolucci 2014b) : l'article confirme la dimension *exigences de sécurité* et la dimension *secteur d'activité*.
- Itération 21 (Alison 2014a) : l'article confirme la dimension *secteur d'activité*.
- Itération 22 (Henschen 2015e) : l'article n'ajoute rien.
- Itération 23 (McTigue 2014) : l'article n'ajoute rien.
- Itération 24 (Henschen 2015f) : l'article n'ajoute rien.
- Itération 25 (Henschen 2014l) : l'article n'ajoute rien.
- Itération 26 (Henschen 2015g) : l'article n'ajoute rien.
- Itération 27 (Booker 2014) : l'article confirme la dimension *secteur d'activité*.
- Itération 28 (Alison 2014b) : l'article confirme la catégorie *réduction des coûts*.
- Itération 29 (Henschen 2014m) : l'article confirme la dimension *secteur d'activité*.
- Itération 30 (Endler 2014) : l'article confirme la dimension *exigences de sécurité*.
- Itération 31 (Henschen 2014n) : l'article n'ajoute rien.

2.2.2.2. Taxonomie de décision

- Itération 1 (Collett 2014) : l'article n'ajoute rien.
- Itération 2 (Henschen 2014g) : l'article n'ajoute rien.
- Itération 3 (Babcock 2014c) : l'article confirme la dimension *couches de cloud*.
- Itération 4 (Henschen 2014h) : l'article n'ajoute rien.
- Itération 5 (Eastwood 2014) : l'article n'ajoute rien.
- Itération 6 (Babcock 2014d) : l'article confirme la dimension *niveau de migration*.
- Itération 7 (Pratt 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 8 (Henschen 2014i) : l'article confirme la sous-catégorie *outils*.
- Itération 9 (Morphy 2014) : l'article n'ajoute rien.

- Itération 10 (Oltsik 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 11 (Babcock 2015) : l'article permet d'ajouter les caractéristiques *temps de disponibilité/an* et *temps d'arrêt/an* à la catégorie SLA.
- Itération 12 (Henschen 2015d) : l'article confirme la sous-catégorie outils.
- Itération 13 (Enderle 2014b) : l'article n'ajoute rien.
- Itération 14 (Bertolucci 2014a) : l'article confirme la sous-catégorie outils.
- Itération 15 (Noyes 2015c) : l'article confirme la sous-catégorie outils.
- Itération 16 (Finnegan 2014) : l'article confirme la sous-catégorie outils.
- Itération 17 (Henschen 2014j) : l'article confirme la sous-catégorie outils.
- Itération 18 (Henschen 2014k) : l'article confirme la sous-catégorie IaaS.
- Itération 19 (Hagen 2014) : l'article confirme la sous-catégorie Hadoop.
- Itération 20 (Bertolucci 2014b) : l'article confirme la sous-catégorie outils.
- Itération 21 (Alison 2014a) : l'article confirme la sous-catégorie outils.
- Itération 22 (Henschen 2015e) : l'article permet la scission de la dimension *supports des applications BI mobiles*. Les deux dimensions issues de la scission sont *supports des applications BI* et *mobile*. L'article requiert la rétrogradation de de la dimension *mobile* qui devient catégorie de la dimension *supports des applications BI* avec ses caractéristiques *téléphone* et *tablette*. L'article permet d'ajouter de la caractéristique *ordinateur portable* à la catégorie *mobile* et la caractéristique *ordinateur de bureau* à la dimension *supports des applications BI*. L'article requiert la rétrogradation de la dimension *cloud mobile* et sa fusion avec la catégorie *mobile*.
- Itération 23 (McTigue 2014) : l'article permet d'ajouter la caractéristique *montre* à la catégorie *mobile*.
- Itération 24 (Henschen 2015f) : l'article confirme la sous-catégorie outils.
- Itération 25 (Henschen 2014l) : l'article confirme la sous-catégorie outils.
- Itération 26 (Henschen 2015g) : l'article confirme la sous-catégorie outils.
- Itération 27 (Booker 2014) : l'article n'ajoute rien.
- Itération 28 (Alison 2014b) : l'article confirme la sous-catégorie *outils*.
- Itération 29 (Henschen 2014m) : l'article n'ajoute rien.
- Itération 30 (Endler 2014) : l'article confirme la sous-catégorie *outils*.
- Itération 31 (Henschen 2014n) : l'article confirme la caractéristique *montre* à la catégorie *mobile*

2.2.2.3. Graphe des opérations sur la taxonomie



2.2.3. Phase 3 : recherche approfondie un mois après la phase 2

2.2.3.1. Taxonomie de contexte

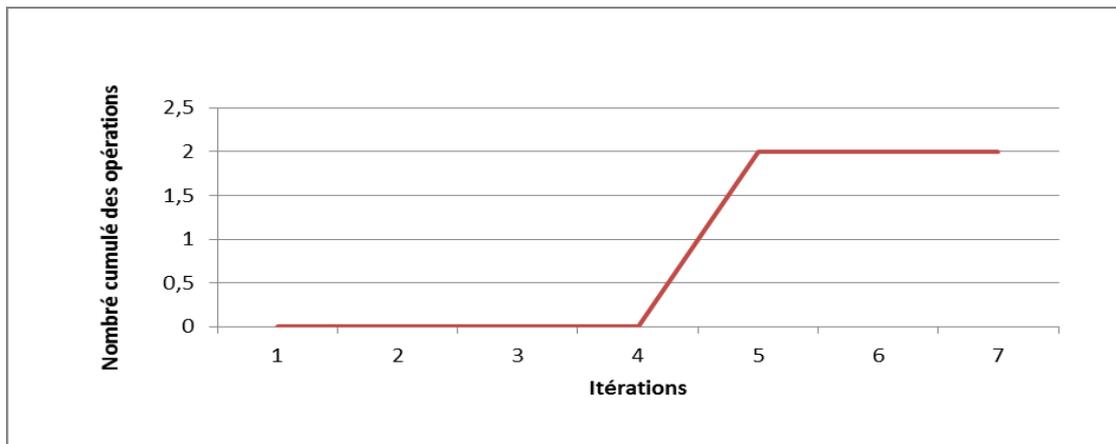
- Itération 1 (Henschen 2015h) : l'article confirme la dimension *taille de l'entreprise*.
- Itération 2 (Henschen) : l'article n'ajoute rien.
- Itération 3 (O'Connor 2014c) : l'article confirme la dimension *secteur d'activité*.
- Itération 4 (Henschen 2015i) : l'article n'ajoute rien.
- Itération 5 (Williams 2014) : l'article n'ajoute rien.
- Itération 6 (Heller 2015) : l'article n'ajoute rien.
- Itération 7 (Henschen) : l'article n'ajoute rien.

2.2.3.2. Taxonomie de décision

- Itération 1 (Henschen 2015h) : l'article confirme la sous-catégorie *outils*.
- Itération 2 (Henschen) : l'article confirme la caractéristique *spark* de la sous-catégorie BIcomposite.
- Itération 3 (O'Connor 2014c) : l'article confirme la sous-catégorie *service*.
- Itération 4 (Henschen 2015i) : l'article confirme la sous-catégorie *service*.
- Itération 5 (Williams 2014) : l'article permet d'ajouter la caractéristique *lunettes* à la catégorie *mobile*.

- Itération 6 (Heller 2015) : l'article permet d'ajouter la caractéristique *langage R* à la sous-catégorie *BI composite*.
- Itération 7 (Henschen) : l'article confirme la caractéristique *langage R*.

2.2.3.3. Graphe des opérations sur la taxonomie



III. LES REGLES

3.1. Les règles énoncées

1. **Si** *objectifs = réduction des coûts/ réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration*
et *Taille de l'entreprise = petite ou moyenne*
Alors *niveau de migration = total* (Al-Aqrabi et al. 2015)
2. **Si** *ressources humaines = analyste*
Alors *couches de cloud = PaaS* (Patel, Rau-Chaplin and Varghese 2012)
3. **Si** *ressources humaines = développeur*
Alors *couches de cloud = PaaS* (Patel, Rau-Chaplin and Varghese 2012)
4. **Si** *taille de l'entreprise = moyenne*
Alors *types de cloud = public* (Kaur, Agrawal and Dhiman 2012)
5. **Si** *exigences de sécurité = intégrité et confidentialité*

- Alors types de cloud = privé** (Assunção *et al.* 2014).
6. **Si exigences de sécurité = intégrité, confidentialité et disponibilité**
Alors types de cloud = hybride (Assunção *et al.* 2014).
7. **Si niveau de migration = partiel/matériel**
Alors couches de cloud = IaaS (Muriithi and Kotzé 2013)
8. **Si niveau de migration = partiel/outils**
Alors couches de cloud = SaaS (Muriithi and Kotzé 2013)
9. **Si objectifs = performance des services**
Alors types de cloud = privé (Kasem and Hassanein 2014)
10. **Si objectifs = réduction des coûts**
Alors couches de cloud = SaaS (Sun *et al.* 2012)
11. **Si objectifs = performance des services**
et Exigences de sécurité = intégrité et confidentialité
Alors types de cloud = privé (Kasem and Hassanein 2014)
12. **Si objectifs = réduction des coûts/ réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration**
et Exigences de sécurité = disponibilité
et Taille de l'entreprise = petite ou moyenne
Alors couches de cloud = SaaS/solution (Al-Aqrabi *et al.* 2015)
13. **Si niveau de migration = partiel/outils/analyse**
Alors couches de cloud = PaaS/BI composite (Brezany *et al.* 2011b)
(Patel, Rau-Chaplin and Varghese 2012)
14. **Si objectifs = Performance des services**
et Secteur d'activité = Institutions financières
Alors types de cloud = privé (Chang 2014)
15. **Si secteur d'activité = institutions financières**
Alors sécurité = intégrité et confidentialité (Ng *et al.* 2011)
16. **Si niveau de migration = partiel/outils/ETL et partiel/outils/DWH**
Alors couches de cloud = SaaS/solution/couche données
(Juan-Verdejo *et al.* 2014)
17. **Si objectifs = mobilité**
Alors supports des applications BI = ordinateur de bureau et mobile (Dinh *et al.* 2013)
(Dignan 2014a)
18. **Si niveau de migration = partiel/outils/analyse**

- Alors couche cloud** = *SaaS/outils/analyse* (Negash 2004)
19. **Si taille de l'entreprise** = *moyenne*
Alors couches de cloud = *SaaS/solution* (Moore 2014)
20. **Si secteur d'activité** = *administration publique*
Alors couches de cloud = *PaaS/BI composite* (Endler 2014)
21. **Si secteur d'activité** = *santé*
Alors couches de cloud = *SaaS/Outils* (Eastwood 2014)
22. **Si couches de cloud** = *SaaS/Solution*
Alors contrat = *long terme* (Ereth and Baars 2015)
23. **Si couches de cloud** = *IaaS/Réseau*
Alors contrat = *long terme* (Ereth and Baars 2015)
24. **Si couches de cloud** = *PaaS/solution mixte*
Alors contrat = *long terme* (Ereth and Baars 2015)
25. **Si couches de cloud** = *SaaS/Service*
Alors contrat = *court terme* (Ereth and Baars 2015)
26. **Si Couches de cloud** = *PaaS/BI composite*
Alors contrat = *court terme* (Ereth and Baars 2015)
27. **Si couches de cloud** = *SaaS/Outils*
Alors contrat = *long terme* (Ereth and Baars 2015)
28. **Si ressources financières** = *faibles*
Alors couches de cloud = *SaaS* (Baur, Bühler and Bick 2015)

3.2. Les règles déduites

1. **Si objectifs** = *réduction des coûts/réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration et Exigences de sécurité= disponibilité*
Alors types de cloud = *public* (Kasem and Hassanein 2014)
2. **Si objectifs** = *performance des services et Exigences de sécurité = intégrité et confidentialité et Taille de l'entreprise = grande*
Alors types de cloud = *privé* (Kasem and Hassanein 2014)
3. **Si objectifs** = *performance des services*
Alors contrat = *court terme*

- et Voie de transmission des services = internet (Baars and Kemper 2010)*
4. **Si** *objectifs = (réduction des coûts/réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration) et objectifs = Performance des services*
Alors *contrat = long terme*
et Voie de transmission des services = centre des données du fournisseur
et Couches de cloud = SaaS/outils (Baars and Kemper 2010)
 5. **Si** *objectifs = (réduction des coûts/réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration) et objectifs = Performance des services*
Alors *contrat = long terme (Baars and Kemper 2010)*
 6. **Si** *objectifs = réduction des coûts / réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration*
et Ressources financières = faibles
et Taille de l'entreprise = petite ou moyenne
et Exigences de sécurité = disponibilité
Alors *niveau de migration = total*
et Couche cloud = SaaS/solution
et Voie de transmission des services = internet (Herwig 2013)
 7. **Si** *objectifs = manipulation de grandes quantités des données*
et Secteur d'activité = Education et recherche ou industrie
et Taille de l'entreprise = grande
Alors *couches de cloud = PaaS/BI composite/Hadoop et hive (Arres, Kabbachi and Boussaid 2013)*
 8. **Si** *secteur d'activité = institutions financières*
Alors *types de cloud = hybride (Moro, Cortez and Rita 2015)*
 9. **Si** *objectifs = manipulation de grandes quantités des données*
et Exigences de sécurité = intégrité et confidentialité et disponibilité
Alors *types de cloud = communautaire (Fernando, Loke and Rahayu 2013)*
 10. **Si** *secteur d'activité = industrie*
et Taille de l'entreprise = grande
Alors *Couche cloud = IaaS/réseau (Babcock 2014c)*
 11. **Si** *objectifs = réduction des coûts et performance des services*
et Secteur d'activité = commerce

Alors couche cloud = *PaaS/solution mixte/hadoop* (Babcock 2014d)

12. **Si secteur d'activité** = *commerce ou institutions financières*

Alors couche cloud = *PaaS/solution mixte/hadoop* (Pratt 2014)

13. **Si secteur d'activité** = *commerce ou institutions financières*

Alors niveau de migration = *partiel/outils/analyse*
et Types de cloud = *hybride* (Henschen 2014i)

14. **Si objectifs** = *réduction des coûts*

et Ressources humaines = *analyste*

et Taille de l'entreprise = *petite ou moyenne*

Alors niveau de migration = *partiel/outils/analyse*

et Couche cloud = *PaaS* (Morphy 2014)

15. **Si objectifs** = *Agilité*

Alors couche cloud = *PaaS/BI composite* (Böhringer et al. 2010)

16. **Si secteur d'activité** = *Télécommunication*

Alors couche cloud = *IaaS/stockage et PaaS* (Xu et al. 2009)

3.3. Les règles d'association

1. **Si exigences de sécurité** = *intégrité et confidentialité*

Alors types de cloud = *privé*

2. **Si exigences de sécurité** = *disponibilité*

Alors types de cloud = *public*

3. **Si exigences de sécurité** = *intégrité, confidentialité et disponibilité*

Alors types de cloud = *hybride*

4. **Si objectifs** = *performance des services*

Alors contrat = *court terme*

et Voie de transmission des services = *internet*

et Couches de cloud = *SaaS/service*

5. **Si objectifs** = *réduction des coûts et performance des services*

Alors contrat = *court terme*

et Voie de transmission des services = *internet*

et Couches de cloud = *PaaS/BI composite*

6. **Si objectifs** = *réduction des coûts et performance des services*

Alors contrat = *long terme*

et Voie de transmission des services = réseau d'entreprise

et Couches de cloud = IaaS/réseau

7. **Si** *objectifs = manipulation de grandes quantités des données*
et Exigences de sécurité = intégrité et confidentialité
Alors *types de cloud = hybride*
8. **Si** *objectifs = performance des services*
Alors *contrat = court terme*
et Voie de transmission des services = internet
et Couche cloud = SaaS/service
9. **Si** *objectifs = réduction des coûts d'infrastructure et performance des services*
et Secteur d'activité = commerce ou santé ou institutions financières
Alors *supports des applications BI = ordinateur de bureau et dispositifs mobiles*
10. **Si** *objectifs = réduction des coûts et performance des services*
Alors *contrat = court terme*
et Voie de transmission des services = internet
et Couche cloud à solliciter = PaaS/BI composite
11. **Si** *objectifs = réduction des coûts et performance des services*
Alors *contrat = long terme*
et Voie de transmission des services = réseau d'entreprise
et Couche cloud à solliciter = IaaS/réseau

3.4. Base de règles après application des directives de cohérence, minimalité et exhaustivité

1. **Si** *objectifs = performance des services*
et Exigences de sécurité = intégrité et confidentialité
Alors *types de cloud = privé*
2. **Si** *objectifs = (réduction des coûts/ réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration) et objectifs = manipulation de grandes quantités de données*
et Secteur d'activité = Education et recherche
et Taille de l'entreprise = petite ou moyenne
Alors *niveau de migration = partiel/matériel/stockage et partiel/outils*

- et Voie de transmission des services = internet*
3. **Si** *objectifs = réduction des coûts*
et Exigences de sécurité = disponibilité
et Taille de l'entreprise = petite ou moyenne
Alors *couches de cloud = SaaS*
 4. **Si** *ressources humaines = analyste ou développeur*
Alors *couches de cloud = PaaS/BI composite*
 5. **Si** *objectifs = Performance des services*
et Exigences de sécurité = confidentialité
et Secteur d'activité = Institutions financières
Alors *types de cloud = privé*
 6. **Si** *taille de l'entreprise = moyenne*
Alors *types de cloud = public*
 7. **Si** *objectifs = Performance des services*
et Taille de l'entreprise = grande
Alors *niveau de migration = partiel/matériel*
et Couches de cloud = IaaS
 8. **Si** *objectifs = (réduction des coûts/ réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration) et objectifs = manipulation de grandes quantité de données*
et Exigences de sécurité = disponibilité
Alors *niveau de migration = partiel/outils*
et Types de cloud = public
et Voie de transmission des services = internet
 9. **Si** *objectifs = manipulation de grandes quantité des données*
et Exigences de sécurité = intégrité, confidentialité et disponibilité
Alors *types de cloud = hybride*
 10. **Si** *objectifs = réduction des coûts/ réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration*
et Taille de l'entreprise = grande
Alors *niveau de migration = partiel/outils*
et Contrat = long terme
 11. **Si** *Exigences de sécurité = intégrité ou confidentialité*

- et Secteur d'activité = institutions financières ou industrie;*
Alors types de cloud = privé
12. **Si** *objectifs = réduction de coût d'infrastructure et Performance des services*
et Secteur d'activité = commerce ou santé ou institution financières
Alors *supports des applications BI = ordinateur de bureau et dispositifs mobiles*
13. **Si** *objectifs = agilité et mobilité*
Alors *supports des applications BI = ordinateur de bureau et dispositifs mobiles*
14. **Si** *objectifs = manipulation de grandes quantités des données*
et Exigences de sécurité = intégrité et confidentialité
Alors *types de cloud = hybride*
15. **Si** *objectifs = performance des services*
et Secteur d'activité = commerce
Alors *couche cloud à solliciter = SaaS/service*
16. **Si** *secteur d'activité = industrie*
et Taille de l'entreprise = grande
Alors *niveau de migration = partiel/matériel*
et Types de cloud = hybride
17. **Si** *secteur d'activité = administration publique*
Alors *couche cloud à solliciter = PaaS/BI composite*
18. **Si** *objectifs = réduction des coûts et performance des services*
et Secteur d'activité = santé
Alors *niveau de migration = partiel/matériel/stockage et partiel/matériel/réseau*
et Couche cloud = IaaS/réseau et IaaS/stockage
19. **Si** *objectifs = réduction des coûts et performance des services*
et Ressources humaines = utilisateur final
et Exigences de sécurité = confidentialité
et Secteur d'activité = Administration publique,
Alors *types de cloud = hybride*
et Niveau de migration = partiel/outils
20. **Si** *objectifs = réduction des coûts/ réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration*
et Ressources financières = faibles
et Taille de l'entreprise = petite ou moyenne

et Exigences de sécurité = disponibilité

Alors niveau de migration = total

et Couche cloud = SaaS/solution

et Voie de transmission des services = internet

et Types de cloud = public

21. **Si** objectifs = (réduction des coûts/ réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration) et objectifs = performance des services

Alors contrat = long terme

et Moyen de distribution = Datacenter du fournisseur

et Couches de cloud à solliciter = SaaS/outils

22. **Si** objectifs = (réduction des coûts/ réduction des coûts d'infrastructure, réduction des coûts d'implémentation et réduction des coûts d'administration) et objectifs = performance des services

Alors contrat = long terme

et Voie de transmission des services = centre des données du fournisseur

et Couches de cloud à solliciter= SaaS/solution

23. **Si** objectifs = manipulation de grandes quantités de données
 et Secteur d'activité = Education et recherche ou industrie
 et Taille de l'entreprise = grande

Alors couches de cloud à solliciter= PaaS/BI composite

24. **Si** Objectifs = performance des services

et Taille de l'entreprise = petite ou moyenne

Alors niveau de migration = partiel/outils

25. **Si** secteur d'activité = industrie

et Taille de l'entreprise = grande

et Couche cloud à solliciter = IaaS/réseau

26. **Si** secteur d'activité = commerce ou institutions financières

Alors niveau de migration = partiel/outils/analyse

et Types de cloud = hybride

et Couches de cloud = PaaS

27. **Si** objectifs = réduction des coûts

et Ressources humaines = analyste

et Taille de l'entreprise = petite ou moyenne

Alors niveau de migration = partiel/outils/analyse

et Couche cloud à solliciter = PaaS

28. *Si objectifs = performance des services,*

Alors couche cloud à solliciter = SaaS/service

29. *Si secteur d'activité = Télécommunication*

Alors couche cloud à solliciter = IaaS/stockage et PaaS

30. *Si objectifs = réduction des coûts d'infrastructure et performance des services*

et Secteur d'activité = commerce ou santé ou institutions financières

Alors supports des applications BI = ordinateur de bureau et dispositifs mobiles

31. *Support Si objectifs = protection des données*

Alors couche cloud à solliciter = IaaS/stockage

Odette SANGUPAMBA MWILU

De la business intelligence interne vers la business intelligence dans le cloud : Modèles et apports méthodologiques

Résumé

La BI et le cloud computing sont deux grands sujets de recherche en informatique et en système d'information en particulier. Une recherche combinant ces deux concepts est d'un intérêt double : D'une part, dans les entreprises, la BI devient de plus en plus une partie importante du système d'information qui nécessite des investissements en termes de performances de calcul et des volumes de données. D'autre part, le cloud computing offre de nouvelles opportunités pour gérer les données à des fins d'analyse.

Etant donné les possibilités de cloud, la question de la migration de l'ensemble du système d'information y compris la BI est d'un grand intérêt. En particulier, les chercheurs doivent fournir aux professionnels des modèles et des méthodes qui puissent les aider à migrer vers le cloud.

Que faire pour que la BI puisse fournir aux managers un service de mise à disposition de données d'analyse au travers du cloud ? La question de recherche est : Comment aider les organisations à migrer leur BI vers le cloud ?

Dans cette thèse, nous répondons à cette question en utilisant l'approche science de conception (design science). Nous mettons en place une aide à la décision de la migration de la BI vers le cloud qui s'appuie sur les taxonomies. Nous proposons un modèle de guidage opérationnel qui est instancié par une taxonomie de la BI dans le cloud et dont découlent les règles pour la migration de la BI vers le cloud.

Mots clés : business intelligence, cloud computing, science de conception, taxonomies, guidage opérationnel, artefacts.

Résumé en anglais

BI and cloud computing are two major areas of computer science research and in particular in information system. A research combining these two concepts has a double interest : On the one hand, in business, the BI becomes increasingly an important part of the information system which requires investment in terms of computing performance and data volumes. On the other hand, cloud computing offers new opportunities to manage data for analysis.

Given the possibilities of cloud, migration question of the information system including BI is of great interest. In particular, researchers must provide models and methods to help professional in BI migration to the cloud. The research question is : how to migrate BI to the cloud?

In this thesis, we address this issue using design science research approach. We implement a decision-making help for BI migration to the cloud based on taxonomies. We provide an operational guidance model that is instantiated by a BI taxonomy in the cloud and from that rules for BI migration to the cloud are arised.

Keywords : business intelligence, cloud computing, design science, taxonomies, operational guidance, artifacts.