

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la  
Recherche Scientifique



Université Abderrahmane Mira, Bejaia  
Faculté des sciences exactes  
Département d'Informatique

## Mémoire de Fin de cycle

En vue de L'obtention du diplôme de Master  
professionnel en Informatique.

Option : Administration et sécurité des réseaux

Thème :

Mise en place d'une solution Open  
Source de Business intelligence  
-Cas Entreprise Portuaire de Béjaïa -



Réalisé par :

M. AÏSSA Raouf

M. BOUAICHI Kousseila

Encadrés par :

M. OUZEGGANE Redouane

Devant le jury composé de :

Président : Mme. TAHAKOURT Zineb

Examineur : M. NAFI Mohamed

Promotion : 2014/2015

# *Remerciements*

En préambule à ce mémoire nous souhaitons adresser nos remerciements les plus sincères aux personnes qui nous ont apportées leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette année universitaire.

Nous tenons à remercier sincèrement M. OUZEGGANE, qui, en tant qu'encadrant, s'est toujours montré à l'écoute et très disponible tout au long de la réalisation de ce mémoire.

Nous remercions également M. BOUDJELLABA, qui, en tant qu'encadrant de stage dans l'organisme d'accueil, nous a chaleureusement accueilli et s'est montré très disponible dans la réalisation de notre application.

Nous tenons à exprimer toute notre gratitude aux membres du jury de nous accorder une partie de leur temps afin de juger et d'évaluer notre travail.

Enfin, nous adressons nos plus sincères remerciements à nos parents et ami(e)s, pour leur soutien et encouragements au cours de la réalisation de ce mémoire.

Merci à tous.



# *Dédicaces*

Ce modeste travail est dédié :

A nos chers parents qui nous ont soutenus et encouragés durant toute notre scolarité.

A nos frères et sœurs

A nos enseignants

A nos ami(e)s

A toutes les personnes qui nous ont apportées de l'aide.

# Table des matières

<b>Table des figures</b> .....	vi
<b>Liste des tableaux</b> .....	x
<b>Liste des acronymes</b> .....	xi
<b>Introduction générale</b> .....	1
<b>Chapitre I : Généralités sur la Business intelligence</b> .....	3
I.1. Introduction .....	3
I.2. Présentation de la Business intelligence .....	3
I.3. Historique de la Business intelligence .....	4
I.4. La place d'un système décisionnel dans l'organisation d'une entreprise et ses domaines d'applications .....	6
I.4.1. Finance .....	6
I.4.2. Vente et logistique .....	6
I.4.3. Marketing .....	7
I.4.4. Production .....	7
I.4.5. Ressources humaines .....	8
I.5. Le marché mondial de la BI .....	8
I.5.1. Le big data .....	10
I.5.2. L'analyse prédictive .....	10
I.5.5. La BI est le web sémantique : .....	11
I.5.5. La BI Agile .....	11

I.5.5. La BI mobile et cloud .....	11
I.6. Les outils de la Business intelligence .....	11
I.6.1. Collecte de données .....	12
I.6.2. Stockage .....	13
I.6.1.2 Datawarehouse .....	12
I.6.2.2 Datamart .....	14
I.6.3 Distribution des informations .....	14
I.6.4. Exploitation.....	15
I.6.4.1 Reporting (Générateur de rapports) .....	15
I.6.4.2. Dashboards (les tableaux de bord) .....	17
I.6.4.3. OLAP « On-Line Analytical Processing » .....	17
I.6.4.4 Datamining (Exploration de données) .....	18
I.7. Les démarches à suivre pour la mise en place d’une solution efficace de Business intelligence et ses fonctionnalités .....	18
I.8. Conclusion .....	19
<b>Chapitre II : Comparaison entre divers solutions Open Source de Business intelligence.....</b>	<b>20</b>
II.1. Introduction .....	20
II.2. Les composants décisionnels .....	20
II.2.1. Les ETL .....	21
II.2.1.1. Talend Open Studio .....	21
II.2.1.2. Pentaho Data Integration .....	22
II.2.2. Les outils de reporting .....	22

II.2.2.1. JasperReports .....	22
II.2.2.2. Pentaho Report Designer .....	23
II.2.2.3. QlikView .....	24
II.2.3. Les outils d'analyse multidimensionnelle .....	24
II.2.3.1. Palo .....	24
II.2.3.2. Mondrian .....	20
II.2.4. Les Datamining .....	26
II.2.4.1. WEKA .....	26
II.2.4.2. RapidMiner .....	27
II.3. Les suites décisionnelles .....	27
II.3.1. SpagoBI .....	28
II.3.2. Jaspersoft BI .....	28
II.3.3. Pentaho BI suite .....	29
II.3.4. Comparaisons entre les trois suites décisionnelles.....	30
II.4. Critiques .....	30
II.5. Conclusion .....	31
<b>Chapitre III : Conception.....</b>	<b>32</b>
III.1. Introduction .....	32
III.2. Etude préliminaire .....	32
III.2.1. Présentation du port de Béjaïa .....	32
III.2.2. Historique du port de Béjaïa .....	33
III.2.3. Activités de l'EPB .....	34
III.2.4. Organigramme général de l'EPB .....	34



III.3. Analyse des besoins .....	35
III.3.1. Définition des besoins .....	35
III.3.2. Compte rendu des besoins .....	37
III.4. Processus d'ETL (Extraction-Transformation-Chargement) .....	37
III.5. Conception des éléments de reporting .....	38
III.5.1. Cas d'utilisation « Gestion des sources de données » .....	39
III.5.2. Cas d'utilisation « Ajouter une source de données » .....	41
III.5.3. Cas d'utilisation « Supprimer une source de données » .....	42
III.5.4. Cas d'utilisation « Modifier une source de données » .....	43
III.5.5. Cas d'utilisation « Gestion du tableau de bord » .....	44
III.5.6. Cas d'utilisation « Ajouter une feuille » .....	45
III.5.7. Cas d'utilisation « Supprimer une feuille » .....	46
III.5.8. Cas d'utilisation « Modifier une feuille » .....	47
III.5.9. Cas d'utilisation « Gestion d'un objet » .....	48
III.5.10. Cas d'utilisation « Créer un objet » .....	49
III.5.11. Cas d'utilisation « Supprimer un objet » .....	50
III.5.12. Cas d'utilisation « Modifier un objet » .....	51
III.6. Conclusion .....	51
<b>Chapitre IV : Réalisation.....</b>	<b>52</b>
IV.1. Introduction .....	52
IV.2. Les outils de travail.....	52
IV.2.1. MySQL Community Edition .....	52
IV.2.2. Talend Open Studio for Data Integration 5.6 .....	52

IV.2.3. QlikView .....	53
IV.3. Définition des sources de données .....	53
IV.4. Processus d'ETL au niveau de Talend Open Studio .....	56
IV.5. Création du tableau de bord .....	61
IV.5.1. Graphiques .....	61
IV.5.1. Filtres .....	63
IV.6. Conclusion.....	65
<b>Conclusion générale .....</b>	<b>66</b>
<b>ANNEXE A : Tutoriaux de base de Talend et QlikView .....</b>	<b>67</b>
A.1. Création d'un job sous Talend Open Studio .....	67
A.2. Création d'un tableau de bord sous QlikView .....	70
<b>ANNEXE B : Terminologie des diagrammes UML utilisés .....</b>	<b>75</b>
B.1. Définition d'UML .....	75
B.2. Utilité d'UML .....	76
B.3. Notions générales sur les diagrammes utilisés .....	76
B.3.1. Diagramme de cas d'utilisation .....	76
B.3.2. Diagramme de séquence .....	79
B.3.3. Diagramme d'activité .....	82
<b>Bibliographie .....</b>	<b>83</b>

# Table des figures

## Chapitre I : Généralités sur la business intelligence

I.1 : Pyramide modélisant le processus de BI .....	4
I.2 : Domaines dans lesquels le décisionnel a été déployé .....	9
I.3 : Fréquence moyenne du rafraichissement des applications décisionnelles .....	10
I.4 : Les 4 phases du processus de Business Intelligence, de la donnée à l'information .....	12
I.5 : Extraction, transformation et chargement de données au niveau des outils ETL .....	13
I.6 : Datawarehouse et Datamart .....	14
I.7 : Les étapes à suivre afin de réussir son reporting .....	16
I.8 : Rapport réalisé avec l'outil BIRT .....	16
I.9 : Représentation d'un cube OLAP .....	18

## Chapitre II : Comparaison entre divers solutions Open Source de Business intelligence

II.1 : Transformation et intégration des données sous Talend Open Studio .....	21
II.2 : Interface graphique : JasperReports .....	23
II.3 : Navigation dans un cube Palo sous Excel .....	25
II.4 : Tableau de bord dans MsExcel, constitué avec Palo .....	25
II.5 : Visualisation de données sous WEKA .....	27
II.6 : Les composants décisionnels des trois suites suivantes : SpagoBI, JasperSoft et Pentaho .....	29

Figure II.7 : Analyse des fonctionnalités des suites décisionnelles .....	33
---	----

## Chapitre III : Conception

III.1 : Port de Béjaïa .....	33
III.2 : Organigramme général de l'EPB .....	34
III.3 : Diagramme d'activité du processus d'ETL .....	38
III.4 : Diagramme des cas d'utilisation du reporting .....	39
III.5 : Diagramme de séquence du cas d'utilisation « Gestion d'une source de données » .....	40
III.6 : Diagramme de séquence du cas d'utilisation « Ajouter une source de données » .....	41
III.7 : Diagramme de séquence du cas d'utilisation « Supprimer une source de données » .....	42
III.8 : Diagramme de séquence du cas d'utilisation « Modifier une source de données » .....	43
III.9 : Diagramme de séquence du cas d'utilisation « Gestion du tableau de bord » .....	44
III.10 : Diagramme de séquence du cas d'utilisation « Ajouter une feuille » .....	45
III.11 : Diagramme de séquence du cas d'utilisation « Supprimer une feuille » .....	46
III.12 : Diagramme de séquence du cas d'utilisation « Modifier une feuille » .....	47
III.13 : Diagramme de séquence du cas d'utilisation « Gestion d'un objet » .....	48
III.14 : Diagramme de séquence du cas d'utilisation « Créer un objet » .....	49
III.15 : Diagramme de séquence du cas d'utilisation « Supprimer un objet » .....	50
III.16 : Diagramme de séquence du cas d'utilisation « Modifier un objet » .....	51



## **Chapitre IV : Réalisation**

IV.1 : Schéma de base de données « remorquages » .....	54
IV.2 : Schéma de base de données « escales » .....	55
IV.3 : Processus d'ETL sous Talend Open Studio .....	57
IV.4 : Etape du mapping .....	57
IV.5 : Contenu du fichier de source de données « projetBI » .....	59
IV.6 : Nombre totale de manœuvres des remorqueurs .....	61
IV.7 : Taux d'exploitation de chaque remorqueur .....	62
IV.8 : Durées moyennes de manœuvres par type de navire .....	62
IV.9 : Taux de manœuvres de remorquage par postes .....	63
IV.10 : Les filtres « jour », « mois » et « années » .....	63
IV.11 : Filtre des navires ayant séjourné au port .....	64
IV.12 : Filtre des manœuvres par poste .....	64
IV.13 : Tableau de bord .....	64

## **ANNEXE A : Tutoriaux de base sur Talend et QlikView**

A.1 : Page de démarrage de Talend Open Studio .....	67
A.2 : Création d'un nouveau job .....	68
A.3 : Structure de l'environnement de travail sur Talend Open Studio .....	68
A.4 : Traitement d'un job .....	69
A.5 : Page de démarrage de QlikView .....	70
A.6 : Feuille principale du projet crée sur QlikView .....	71
A.7 : Editeur de sources de données .....	72
A.8 : Création d'un nouvel objet .....	72

A.9 : Les paramètres d'un graphique .....	73
A.10 : Sélection d'une dimension .....	73
A.11 : Saisie de la fonction de calcul .....	74
A.12 : Chargement d'un graphique .....	74

# Liste des tableaux

## **Chapitre III : Conception**

III.1 : Questionnaire pour la direction de remorquage .....	36
III.2 : Compte rendu des besoins .....	37

## **Chapitre IV : Réalisation**

IV.1 : Attributs des bases de données « remorquages » et « escales » .....	56
IV.2 : Les composants utilisés au niveau de L'ETL .....	58
IV.3 : Requêtes SQL pour l'extraction de données .....	59

## **ANNEXE A : Tutoriaux de base sur Talend et QlikView**

A.1 : Type de fichiers tables pouvant être importés sur QlikView .....	71
--	----

## **ANNEXE B : Terminologie des diagrammes UML utilisés**

B.1 : Syntaxe UML : Diagramme de cas d'utilisation .....	79
B.2 : Syntaxe UML : Diagramme de séquence .....	81
B.3 : Syntaxe UML : Diagramme d'activité .....	82

# Liste des acronymes

**API** : **A**pplication **P**rogramming **I**nterface.  
**BDD** : **B**ase **D**e **D**onnées.  
**BI** : **B**usiness **I**ntelligence.  
**CRM** : **C**ustomer **R**elationship **M**anagement.  
**CSV** : **C**omma-**S**eparated **V**alue.  
**DRH** : **D**irection des **R**essources **H**umaines.  
**EPB** : **E**ntreprise **P**ortuaire de **B**éjaïa.  
**ERP** : **E**nterprise **R**esource **P**lanning.  
**ETL** : **E**xtract-**T**ransform-**L**oad.  
**GPL** : **G**eneral **P**ublic **L**icense.  
**HTML** : **H**yper**T**ext **M**arkup **L**anguage.  
**IBM** : **I**nternational **B**usiness **M**achines.  
**MDM** : **M**aster **D**ata **M**anagement.  
**MDX** : **M**ultidimensional **E**xpressions.  
**MIT** : **M**assachusetts **I**nstitute of **T**echnology.  
**OLAP** : **O**nline **A**nalytical **P**rocessing.  
**PDF** : **P**ortable **D**ocument **F**ormat.  
**QVW** : **Q**lik**V**iew **W**orksheet.  
**RTF** : **R**ich **T**ext **F**ormat.  
**SAP** : **S**ystems, **A**pplications, and **P**roducts.  
**SGBD** : **S**ystème de **G**estion de **B**ase de **D**onnées.  
**SQL** : **S**tructured **Q**uery **L**anguage.  
**TXT** : **T**ext.  
**UML** : **U**nified **M**odeling **L**anguage.  
**XLS** : **e**Xcel **S**preadsheet.  
**XML** : **e**Xtensible **M**arkup **L**anguage.



# Introduction générale

L'information est la matière première la plus précieuse pour la compétitivité des entreprises au 21<sup>e</sup> siècle et l'intelligence, humaine soit-elle, ou artificielle, a besoin de cette connaissance pour aider à la prise de décision.

Le partage de données et la diffusion de connaissance sont donc les domaines les plus sollicités de l'informatique. Les systèmes d'information ont permis, grâce aux progrès réalisés en technologies de l'information (logiciels, bases de données,...), le recueillement, traitement, stockage et diffusion de gros volumes d'information, et toutes ces opérations s'effectuent de plus en plus rapidement et à un coût raisonnable.

L'entreprise doit optimiser ses activités de production et ses méthodes et employer une stratégie d'amélioration continue. Elle engendre plus de succès en utilisant un système de gestion de performances afin d'assurer l'efficacité de leur stratégie. Pour cela, La Business intelligence permet d'avoir une vue d'ensemble des différentes activités de l'entreprise et propose une collection de processus et de technologies pour la simplification et l'amélioration de l'utilisation des informations clés au sein des différents domaines fonctionnels de la société. Autrement dit, l'idée c'est alors de les exploiter à des fins décisionnelles et de les organiser pour en extraire de précieux renseignements.

**Problématique :** l'Entreprise Portuaire de Béjaïa comporte différentes structures, dont la direction de remorquages qui gère un ensemble d'opérations qui consistent essentiellement à remorquer les navires entrants et sortants, les données résultantes sont stockées dans différentes bases de données et celles-ci doivent être exploitables pour générer des informations d'aide à la décision, ce qui est un obstacle de taille vu la grandeur des données.

Notre projet consiste donc mettre en place une solution de business intelligence au niveau de l'Entreprise Portuaire de Béjaïa qui comprend une étape de consolidation de données via un outil d'ETL et une étape de reporting qui permettra la visualisation de données

grâce à un tableau de bord contenant différentes informations sur les activités internes et externes de l'entreprise.

Ce mémoire est organisé en quatre chapitres :

Le premier chapitre « Notions sur la Business intelligence » comporte une présentation générale de la Business intelligence, de ses domaines d'utilisation et de ses outils.

Le second chapitre « Comparaison entre divers solutions Open Source de Business intelligence » regroupe différentes solutions Open Source qui s'offrent à un utilisateur pour mettre en place un projet décisionnel et s'achève par des critiques concernant ces dernières.

Le troisième chapitre « Conception et analyse des besoins » comprend une étude préliminaire, une définition et une analyse des besoins, et l'ensemble des diagrammes UML illustrant la partie dynamique de notre projet.

Le quatrième et dernier chapitre « Réalisation » comporte une présentation des outils de travail et des explications sur la démarche à suivre pour réaliser notre projet et quelques illustrations de ce dernier.

Nous concluons par exposer l'ensemble des connaissances acquises au cours de la réalisation de ce projet.

# **CHAPITRE I**

## **Généralités sur la Business intelligence**

### **I.1 Introduction**

La Business intelligence est considéré comme étant « l'informatique à l'usage des décideurs et des dirigeants d'entreprises ». De ce fait, Elle occupe une place importante dans les grandes entreprises.

Ce premier chapitre vise à définir La business intelligence, son évolution et son importance aujourd'hui dans l'entreprise au niveau de la prise de décision, suivi des outils et technologies exploités pour la mise en place d'un système décisionnel.

### **I.2 Présentation de la Business intelligence**

La business intelligence (en français Informatique décisionnelle, BI) couvre un vaste domaine représentant à la fois les moyens , les outils , et les méthodes qui permettent de modéliser ,collecter , consolider et restituer les données d'une entreprise en vue d'offrir une aide à la décision.

La traduction du terme Business intelligence évolue en fonction de l'orientation que souhaite lui donner celui qui y fait référence. Les dictionnaires informatiques nous renvoient à des termes tels qu'intelligence d'affaires, gouvernance d'entreprise ou plus généralement informatique décisionnelle. [1]

La BI est un sujet en pleine évolution, s'adressant à la direction générale tout comme aux métiers. Outil d'aide à la décision, la BI permet d'avoir une vue d'ensemble des différentes activités de l'entreprise et son environnement. Cette vue transversale nécessite de connaître les différents métiers de l'entreprise et implique certaines spécificités



organisationnelles et managériales. En résumé elle offre un ensemble de solutions informatiques permettant l'analyse des données de l'entreprise afin d'en dégager les informations qualitatives nouvelles qui vont fonder des décisions, qu'elles soient tactiques ou stratégiques. [2]

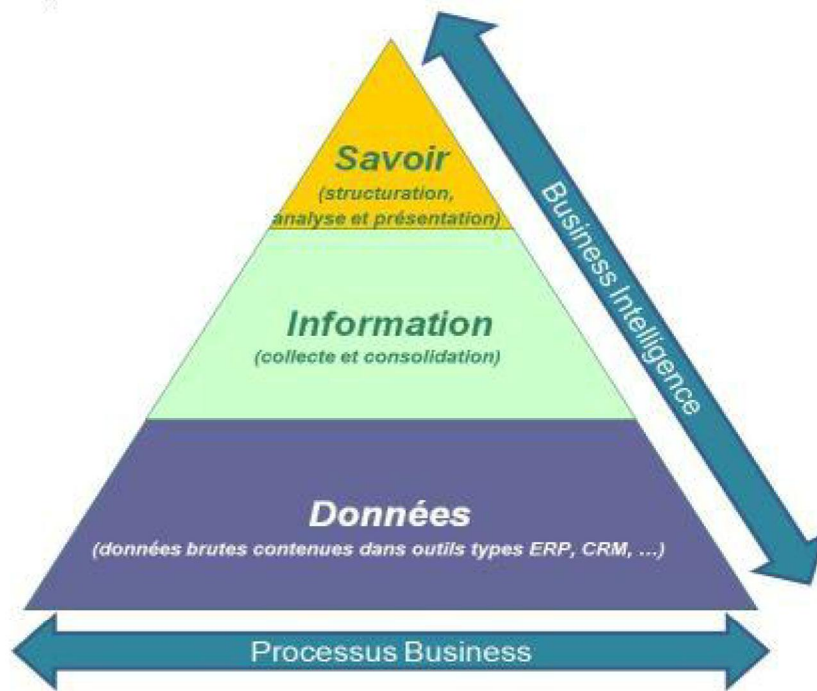


Figure I.1 : Pyramide modélisant le processus de BI. [2]

La figure ci-dessus illustre le processus de Business intelligence qui démarre par la récupération de données brutes (contenues dans des outils de types ERP et CRM provenant des clients, fournisseurs, données de marchés...), puis les consolide et les transforme en informations utiles sous forme de rapports.

### I.3 Historique de la Business intelligence

Le terme Business intelligence est apparu pour la première fois en 1958 grâce à un informaticien allemand nommé Hans Peter LUHN, ce dernier la définit comme « la capacité à appréhender les interrelations des faits présents de manière à orienter l'action vers un but désiré ». Dans les années 70, des chercheurs du MIT qui voulaient développer une nouvelle architecture système en partant de zéro, ont permis une avancée qui



consiste à séparer le traitement opérationnel et analytique , ce qui donne la dispersion des données dans plusieurs Datawarehouses avec des capacités de stockage limités.

Dans les années 80 Digital Equipment Corporation (DEC) permis une autre percée significative, ils ont constitué une équipe multidisciplinaire comportant ingénieurs, financiers et responsable du mangement, partant des recherches du MIT, ils décidèrent d'implémenter plusieurs applications distribuées utilisant des Datawarehouses séparés. C'était la première fois qu'un modèle client/serveur était testé.

Pendant ce temps, IBM poursuit une lutte contre un aspect différent du problème de gestion de l'information. L'un des plus gros problèmes étant d'intégrer des données issues de nombreux systèmes d'informations distincts possédant chacun des codages différents. En 1980, l'informaticien Rakesh Agrawal apporta sa pierre à l'édifice en introduisant le concept du Datamining (exploration de données).

Dans les années 90, Maintenant que l'architecture a été mise en place, le marché a connu développement des premiers outils d'extraction, de transformation, de chargement (regroupé par la suite sous le terme ETL), C'est aussi pendant ces années qu'apparurent des concepts et méthodes tels que OLAP (Online Analytical Processing), ODS (Operational Data store), EIS (Executive Information System) pour désigner des outils de restitution d'information synthétique.

Donc la Business Intelligence a évolué à partir de la recherche opérationnelle qui a commencé dans les années 1960, suivie par la supervision (DSS : Decision Support System et Datawarehouse) dans les années 1970, l'aide à la décision (entrepôts de données, OLAP et BI) et l'exploration de données (Datamining) dans les années 80-90, pour arriver aux applications complexes Agile, collaborative, prédictive, mobile et temps réel tels qu'on les connaît aujourd'hui. [3] [4]

## **I.4 La place d'un système décisionnel dans l'organisation de l'entreprise et ses domaines d'applications**

Pour assurer son bon fonctionnement et sa pérennité, l'entreprise est confrontée à un double défi : Gérer l'immense quantité de données auxquelles elle a accès, et transformer cette même quantité en informations utiles à un pilotage efficace de ses actions et subsister dans l'évolution continue de son environnement.

Dans une entreprise, le système d'information décisionnel a pour objectif de faciliter l'établissement et la mise en œuvre de la stratégie et la réalisation des activités. Il est construit à partir des exigences des métiers, des processus définis par l'entreprise, et il est constitué de l'ensemble des moyens (humains, logiciels, matériels) utilisés pour collecter, stocker, traiter et communiquer les informations. [5]

Les domaines de l'utilisation de la BI au sein de l'entreprise se résument comme suit :

### **I.4.1 Finance**

La compétitivité et la transformation de l'entreprise sont aujourd'hui les enjeux de la fonction Finance. Ce qui fait qu'elle devienne un acteur majeur dans le pilotage global de l'entreprise.

La pression des marchés, des investisseurs et des régulateurs entraîne les directions générales à demander toujours plus à la fonction Finance. Elle doit aujourd'hui aider au pilotage global de l'entreprise en produisant toujours plus rapidement des chiffres précis, opérationnels et adaptés à une réglementation mouvante. [5]

### **I.4.2 Vente et logistique**

L'amélioration de la rentabilité est un des principaux enjeux des entreprises aujourd'hui. Par ses actions sur la réduction des coûts et le contrôle des approvisionnements, la direction achat est en prise directe avec ces objectifs stratégiques. Pour mener à bien ses missions, la mise en place d'un système décisionnel lui permet à la fois de piloter la

performance achat en ayant une vision globale des dépenses mais aussi de communiquer auprès de ses clients internes.

Par conséquent, le rôle de la BI consiste à offrir des solutions d'optimisation de gestion des stocks, de suivi des livraisons, d'analyse des points de ventes, de la rentabilité et de l'impact des promotions. [5]

### **I.4.3 Marketing**

Le marketing est en forte évolution dans le monde de la BI vu la montée en maturité du commerce internet et le succès des réseaux sociaux. Malgré ces progrès, le marketing fait face à des problématiques encore bien ancrées, ce qui confronte les systèmes d'informations à des situations telles que :

- La difficulté de permettre un appariement efficace client / produit.
- La faiblesse des interactions entre les différents canaux de contact ou de distribution encore trop marqués par des architectures en silos et une organisation non coopérative

La direction en collaboration avec l'informatique développe des hypothèses sur les meilleures possibilités de production de valeur ajoutée au sein de l'organisation, dans lesquelles la BI peut jouer un rôle essentiel en clarifiant et développant des hypothèses lors des réunions et séances de travail organisé avec les utilisateurs du domaine. Une fois que les besoins et les capacités effectives sont clairs, que le projet est défini, les informaticiens doivent s'appuyer sur les lignes directrices métier pour construire l'environnement BI pour ce projet particulier. [5]

### **I.4.4 Production**

La planification de la production nécessite une coordination extrêmement fine entre les différents services de l'entreprise afin de répondre à la demande de ses clients, tout en maîtrisant ses coûts, optimiser ses stocks et maintenir en ordre de marche l'outil industriel. Les solutions de Business Intelligence alimentées de données provenant du



système d'information génèrent des gains considérables pour l'entreprise en favorisant sa flexibilité et sa capacité d'anticipation. [5]

### **I.4.5 Ressources humaines**

La fonction Ressources Humaines est un véritable appui aux activités de l'entreprise et donc un contributeur incontournable à la création de valeur, Les ressources humaines doivent donc pouvoir répondre à plusieurs enjeux majeurs :

- Un enjeu stratégique en repensant le travail et en visant des actions à valeur ajoutée pour ses clients, ses salariés, ses managers. Les DRH devront démontrer à chacun l'intérêt qu'il a à orienter ses travaux en ce sens.
- Des enjeux internes : penser et viser la qualité RH en augmentant productivité et qualité des activités administratives comme la paie, la gestion des temps (congés), les questions législatives, la gestion et le suivi des carrières, les axes de mobilisation et de motivation...

Les DRH peuvent désormais proposer des démarches pragmatiques, des méthodes, des solutions, des exemples, des outils. Toutes les actions entreprises peuvent être quantifiables et mesurables. Aussi, la BI propose l'analyse des particularités des effectifs et l'affichage de statistiques concernant les employés, telles que les fonctions et les performances des employés, leurs nombre (hommes/femmes), moyenne d'âge etc... [6]

## **I.5 Le marché mondial de la BI**

Avant de donner quelques chiffres significatifs, il faut savoir que le décisionnel est devenu incontournable dans le monde professionnel, il permet de mieux comprendre le fonctionnement des entreprises, de définir une stratégie, d'aider à avoir une meilleure vision du marché ce qui permet de mesurer le future proche. Aujourd'hui 95% des entreprises sont équipés de logiciels et outils décisionnels.

Il faut aussi savoir que le secteur est en croissance perpétuelle, malgré une légère baisse en 2012 due à la crise économique en Europe, ce taux de croissance atteint les 8% en 2013, soit un chiffre d'affaire de 14.4 milliards de dollars, il surclasse même le taux de

croissance du marché des logiciel qui est de 4%, ce qui affirme que le marché du Business intelligence ne connaît pas la crise. Ce marché est dominé par les grandes firmes mondiales tels que Microsoft, Oracle, IBM, SAP et qui en détiennent les deux tiers, mais il faut aussi parler des nouveaux acteurs qui se développent de plus en plus et qui affichent une croissance moyenne de 12.5% supérieur donc à celle du marché dans son ensemble.

Le taux d'équipement en outils décisionnels est surtout présent dans le domaine de la finance et la gestion 76%, suivi de près par le domaine commercial 71%.

Le graphique suivant nous présente bien les domaines ou sont déployés ces logiciels :

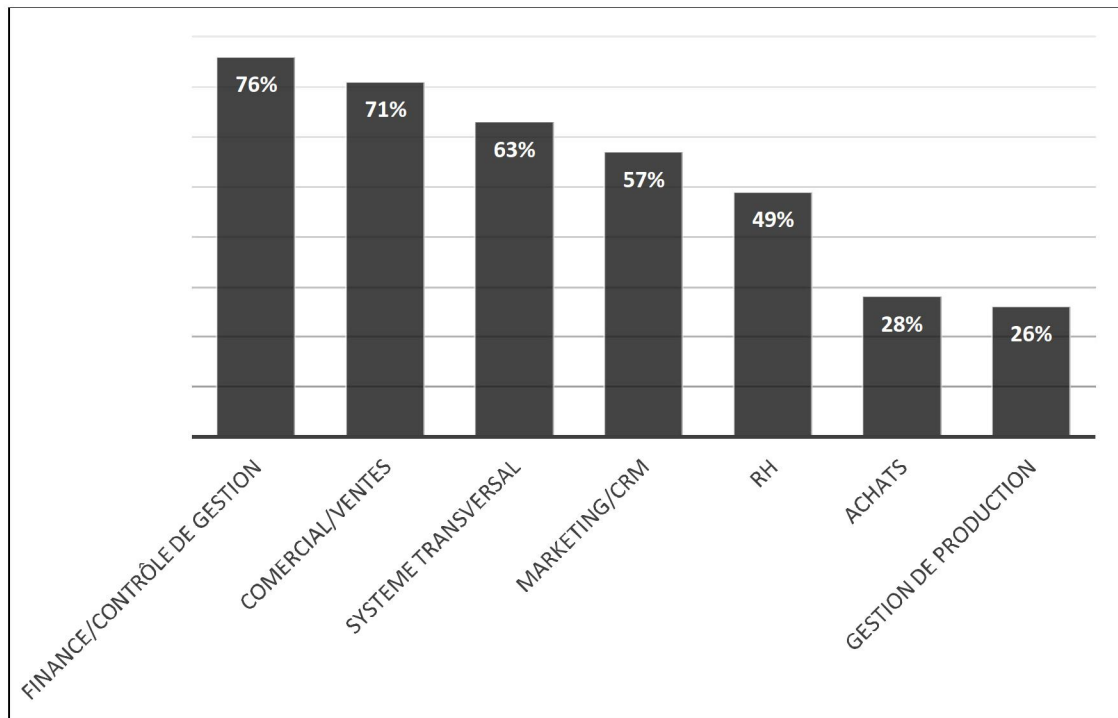


Figure I.2 : Domaines dans lesquels le décisionnel a été déployé. [3]

Quant à la fréquence de mise à jour des données elle est surtout quotidienne, les applications a temps réels sont minoritaires, comme le démontre le graphe suivant :



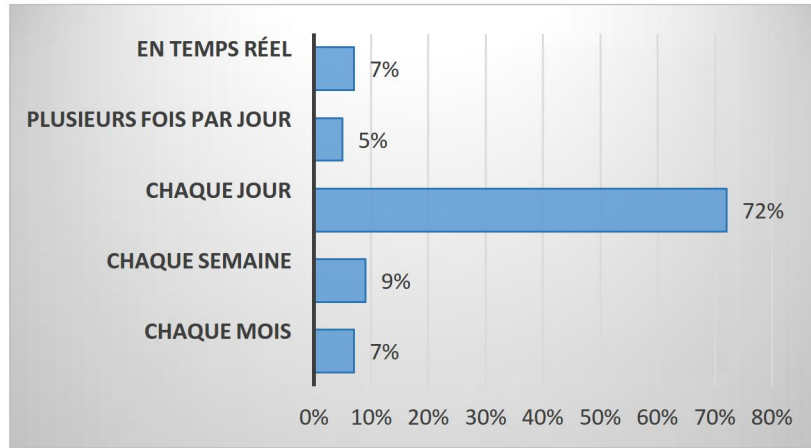


Figure I.3 : Fréquence moyenne du rafraichissement des applications décisionnelles. [3]

Les principaux enjeux du décisionnel est temps réel, sachant la rapidité dont l'information croît avec le Big data, c'est à dire avoir une BI aussi vive que le Web, mais aussi le développement de solutions à bas couts et mobiles. [3]

Dans ce qui suit nous nous concentrerons sur les grandes tendances actuelles du marché de la Business intelligence :

### I.5.1 Le Big data

Le principal défi auquel doit faire face l'informatique décisionnelle est le Big data (ou Megadonnées). Ce problème touche tous les processus du décisionnel (reporting, analyse et prédiction), il faut dire que les entreprises trouvent une énorme difficulté à trier et à analyser les données énormes qu'elles ont sous les mains, ces données qui ont crû d'une façon assez inattendue avec l'avènement du web 2.0 et des medias sociaux, c'est pour cela que les stratégies de collecte, de traitement, d'analyse et de visualisation des données doivent être redéfinies. Des logiciels sont proposés pour faire l'analyse du Big data mais ça reste insuffisant donnant ainsi des perspectives énormes pour ce marché. [7]

### I.5.2 L'analyse prédictive

S'appuyant sur les statistiques et la théorie des jeux pour analyser les données et tirer des hypothèses pour le futur. Il existe des applications pour tous les secteurs, mais elle est surtout utilisée aujourd'hui pour l'évaluation des risques clients. Ce secteur est en pleine expansion et son chiffre d'affaire atteint déjà les 2 milliards de dollars US. [4]

### **I.5.3 La BI et le web sémantique**

Plusieurs éditeurs proposent des outils d'analyse sémantique sur le web. Ces outils sont surtout spécialisé dans la surveillance de l'opinion (ou analyse des sentiments), l'anonymisation, l'extraction sémantique de textes médicaux, l'analyse des CV (curriculum vitae), la classification des documents pour le CRM.

### **I.5.4 La BI Agile**

C'est-à-dire à la réduction des durées de projets pour accéder à une valeur ajoutée plus rapidement. Cela passe par des méthodes, une organisation mais aussi des outils. Nous citerons comme outil Power BI de Microsoft.

### **I.5.5 La BI mobile et Cloud**

Il est maintenant possible de visualiser les données, et avoir les résultats des différentes analyses sur son smartphone, il existe ainsi plusieurs éditeurs qui proposent des versions mobiles de leurs produit qui se présentent sous forme de Dashboard (tableau de bord), ou des outils d'accès, de visualisation et de partage de données via le Cloud, ce qui donne une autonomie aux utilisateurs.[7][8]

## **I.6 les outils de la Business intelligence**

Ayant pour objectif d'offrir une aide à la décision, la Business intelligence couvre un ensemble de technologies et d'outils qui permettent de traiter une somme importante d'informations pour, à la fin du processus, présenter aux décideurs des indicateurs fiables et pertinents. C'est alors que la chaine décisionnelle se présente selon l'architecture suivante :

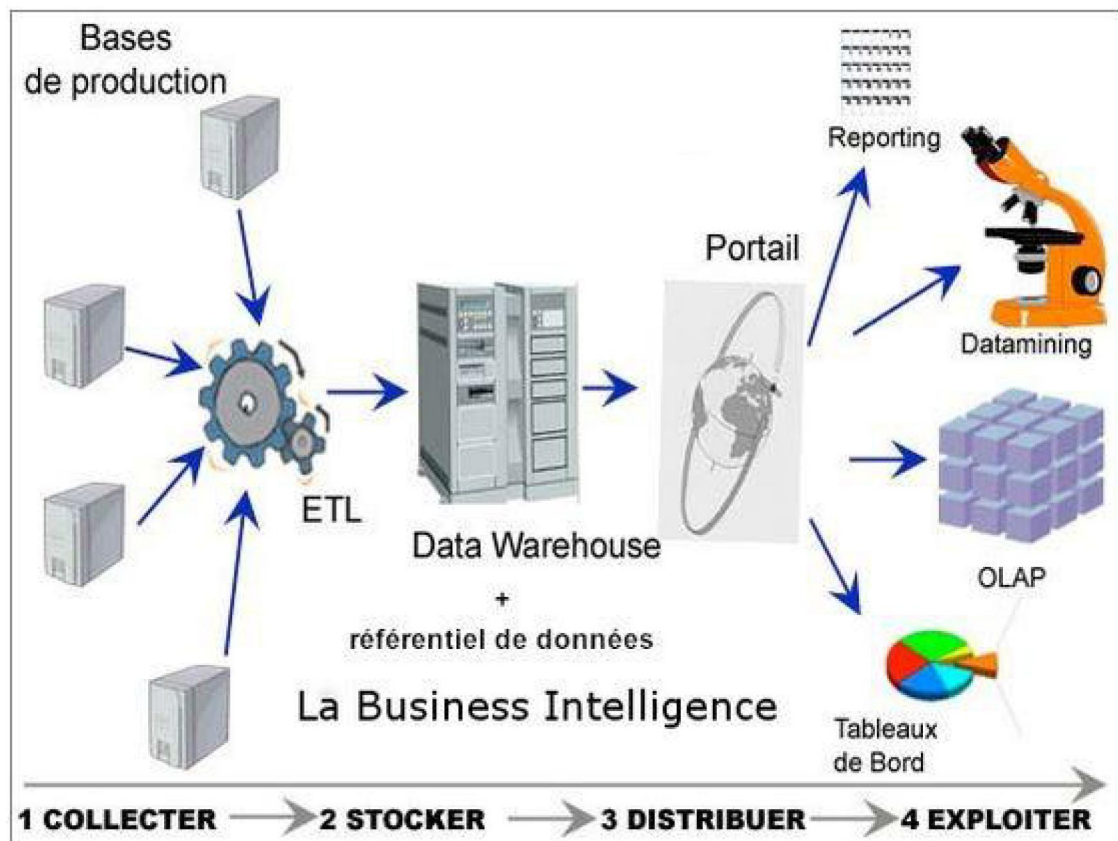


Figure I.4 : Les 4 phases du processus de Business Intelligence, de la donnée à l'information. [10]

Les composants de cette figure forment l'ensemble des outils intervenant dans la mise en place d'une solution de business intelligence, ils s'organisent de la manière suivante :

### I.6.1 Collecte de données

La collecte des données est une fonction remplie par une famille d'outils dénommée ETL. Comme son nom l'indique, les processus ETL (Extraction, Transformation et Chargement) procèdent comme suit :

- ➔ Extract : Extraction des données à partir de différentes sources.
- ➔ Transform : Transformation de ces données afin de les unifier sous un même format.
- ➔ Load : Chargement des données dans le Data Warehouse. [4]



L'ETL peut prendre en charge différentes natures de sources de données, tant en entrée qu'en sortie, les principales étant bien sûr les SGBDR et les flux XML, mais il peut s'agir également de fichiers à formats fixes ou avec séparateurs (CSV). [4]

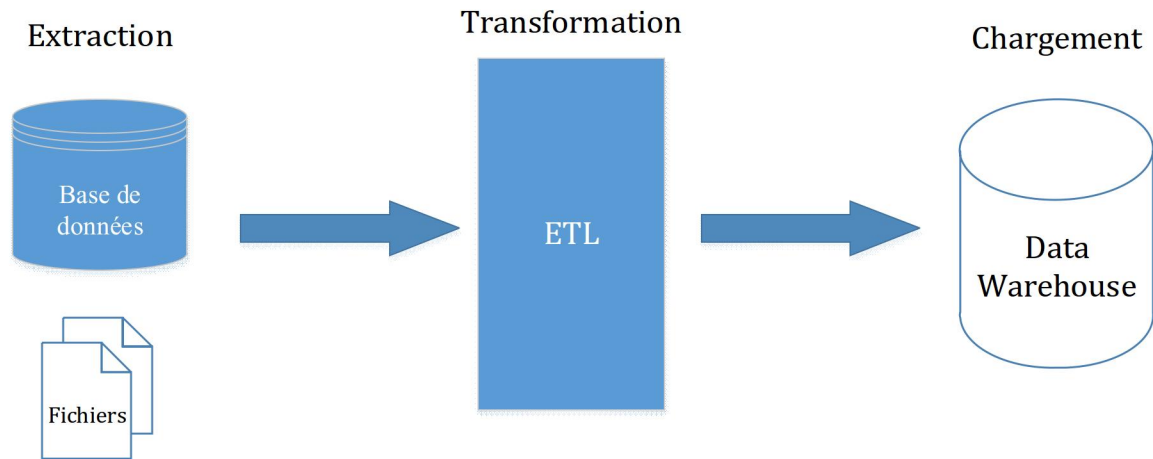


Figure I.5 : Extraction, transformation et chargement de données au niveau des outils ETL. [11]

## I.6.2 Stockage

Les requêtes décisionnelles sont particulièrement gourmandes en ressources machines, et une fois que les données sont nettoyées et consolidées via les outils ETL, Elles seront stockées dans des base spécialisées : le Datawarehouse et le Datamart. [13]

### I.6.2.1 Datawarehouse

Un Datawarehouse (entrepôt de données) centralise les données issues des applications utilisées dans l'entreprise dans le cadre de la prise de décision.

Les principales caractéristiques du Datawarehouse sont les suivantes :

1. **Orienté sujet** : les données sont organisées et triées par thème.
2. **Données Intégrées** : les données provenant de sources hétérogènes, elles utilisent chacune un type de format. Elles doivent donc être intégrées avant d'être proposées à utilisation.

3. **Données non-volatiles** : Les données du Datawarehouse sont non volatiles ce qui signifie qu'une donnée entrée dans l'entrepôt ne disparaît pas et ne change pas au fil du traitement et du temps.
4. **Données historisées** : les données sont datées, afin de visualiser l'évolution dans le temps d'une valeur donnée. [10]

#### **I.6.2.2 Datamart**

Le Datamart est considéré comme une version plus réduite du Datawarehouse contenant des données pour un secteur particulier d'une entreprise.

Plutôt que de viser l'universalité des thèmes comme le fait le Datawarehouse, le Datamart se focalise sur un sujet, un thème ou un métier.

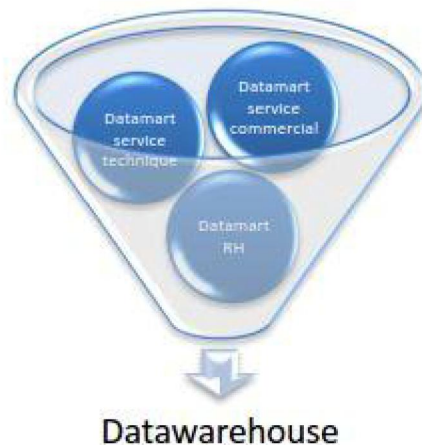


Figure I.6 : Datawarehouse et Datamart. [14]

#### **I.6.3 Distribution des informations**

Avant d'exploiter les données via divers outils, l'information est distribuée auprès de l'ensemble des partenaires à partir du portail d'information d'entreprise (Enterprise Information Portal, EIP).

Si le Datawarehouse est la solution pour stocker et organiser l'information, le portail est incontournable à son tour pour en banaliser l'accès. Les portails d'entreprise (EIP, Enterprise Information Portal) ont initié une nouvelle approche de la conception des systèmes d'information en plaçant au premier plan les besoins des utilisateurs en matière d'accès aux informations essentielles et de partage avec leurs pairs.



L'accès aux informations de l'entreprise se fait avec un simple navigateur Internet. Une fois l'utilisateur connecté et reconnu, le portail affiche sa page d'accueil personnalisée. Chaque poste est personnalisable selon les choix de chacun. [10]

### **I.6.4 Exploitation**

Avant d'entamer cette étape, les données doivent être stockées, nettoyées, consolidées et accessibles. Elles seront ensuite utilisables selon les besoins des utilisateurs. Voici les différents types d'outils d'extraction et d'exploitation existants :

#### **I.6.4.1 Reporting (Générateur de rapports)**

Le terme "Reporting" désigne une technologie de Business intelligence destinés à assurer la réalisation, la publication et la diffusion de rapports d'activité selon un format prédéterminé. Ils sont essentiellement destinés à faciliter la communication de résultats chiffrés ou d'un suivi d'avancement.

Les outils de reporting facilitent la réalisation de rapports automatiques en interrogeant directement les bases de données selon des requêtes SQL préparées au préalable ou ad hoc (requêteurs). Une fois élaboré, le rapport est ensuite diffusé périodiquement ou automatiquement sur le réseau interne auprès des intéressés dûment identifiés. Les outils intègrent en standard un jeu de fonctions afin d'insérer calculs spécifiques et graphiques explicatifs sur le modèle de rapport. Le reporting n'est pas à proprement parler un outil d'aide à la décision. Il est surtout utilisé pour « rendre compte » de l'activité des subalternes auprès de la hiérarchie dans la plus pure tradition de l'entreprise pyramidale. [12]



Figure I.7 : Les étapes à suivre afin de réussir son reporting. [11]

La figure ci-dessus illustre les étapes à suivre pour réussir un rapport : Se fixer des objectifs précis, assurer une bonne collecte de données et les transformer en information, utiliser de bons graphiques, optimiser le rapport en évitant la surcharge d'informations et d'étudier la manière dont le message est perçue par les destinataires.

La figure suivante constitue un exemple des résultats produits lors d'une opération de « reporting ».

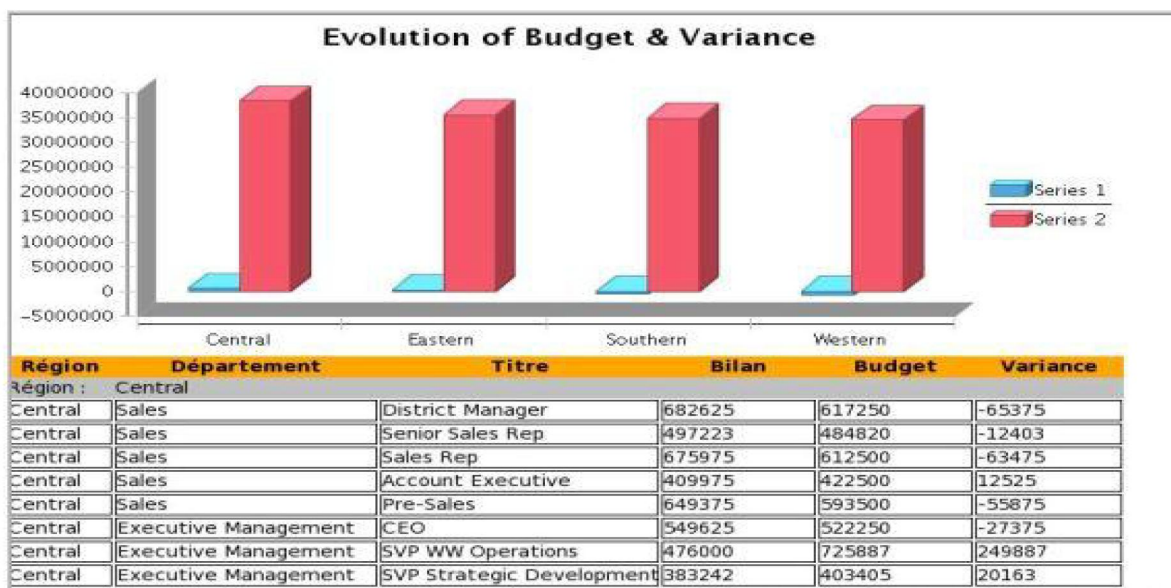


Figure I.8 : Rapport réalisé avec l'outil BIRT [12]

#### **I.6.4.2 Dashboards (Les tableaux de bord)**

Le tableau de bord est une forme particulière de rapport, il présente plusieurs indicateurs de pilotage, construits de façon périodique afin de guider les décisions et offrir une représentation complète de l'activité de l'entreprise. Les tableaux de bord sont considérés comme étant des outils de pilotages et se présentent sous forme de tableaux et de graphiques contrairement aux rapports qui sont des outils de contrôle qui peuvent se contenter de l'un des deux. [12]

#### **I.6.4.3 OLAP « On-Line Analytical Processing » (Analyse multidimensionnelle)**

Les fonctions OLAP permettent, à partir des Datawarehouse, une analyse multidimensionnelle sur des bases de données volumineuses d'une entreprise et de mettre en évidence ses activités grâce à des statistiques : moyennes mobiles, coefficients de corrélation, valeurs cumulées... L'élément principal de l'infrastructure OLAP est le cube. Un cube reprend les mesures de la table de fait que l'on a pu établir lors de la conception du Datawarehouse, et s'en sert pour effectuer des calculs, les mesures étant des données quantitatives.

Le but de l'OLAP est de permettre une analyse multidimensionnelle sur des bases de données volumineuses afin de mettre en évidence une analyse particulière des données.

L'exemple de la figure suivante montre un cube simple à trois dimensions : il présente le chiffre d'affaire que produit un magasin situé sur trois sites (Paris, Toulouse et Bordeaux), proposant trois catégories d'articles (vêtements, sacs et chaussures), ce chiffre étant calculé chaque trimestre. Ainsi, à l'intersection des trois axes se trouve le montant des bénéfices correspondant à la ville, à la période de temps et au type d'article choisis.

[14]



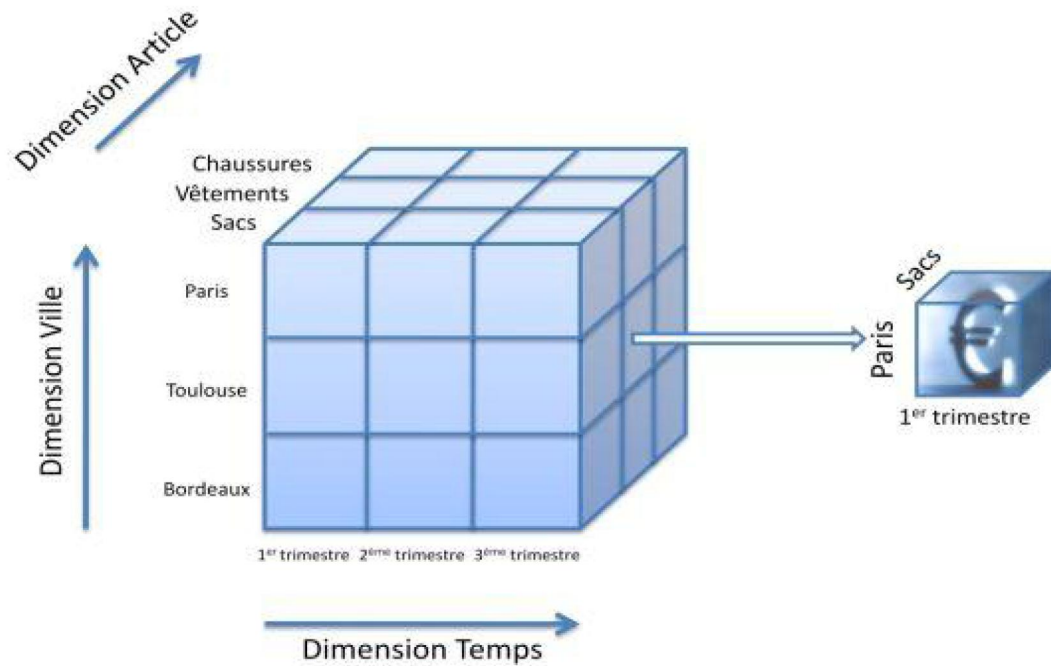


Figure I.9 : Représentation d'un cube OLAP. [14]

#### I.6.4.4 Datamining (Exploration de données)

Le datamining est un processus de fouille de données, il permet d'extraire des connaissances à partir d'un volume de données (moins important que dans le cas d'OLAP). Une fois que le problème en termes de données est identifié, plusieurs étapes sont nécessaires :

- L'accès aux données sélectionnées et préparation en vue de leur future destination.
- Modélisation des données grâce à des analyses et algorithmes de fouille de données.
- On extrait et évalue les connaissances résultant de ces analyses.
- On déploie les connaissances en vue d'une utilisation effective. [14]

### I.7 Les démarches à suivre pour la mise en place d'une solution efficace de Business intelligence et ses fonctionnalités

Comme tout projet informatique, la mise en place d'une solution de BI doit suivre une démarche et une méthodologie bien adaptée. Elle fait appel à une suite de technologies déjà cités et s'organisera comme suit :

1. Planifier et bien étudier la faisabilité du projet.
2. Collecter les besoins en termes de tableaux de bord, de reporting, d'analyse multidimensionnelle...à tous les niveaux de l'entreprise.
3. Concevoir l'architecture technique.
4. Choisir, installer et configurer les technologies à utiliser (SGBD, ETL...) dépendamment des fonctionnalités demandées par les utilisateurs.
5. Concevoir et créer un entrepôt de données (Datawarehouse ou Datamart).
6. Alimenter l'entrepôt de données via des mécanismes d'extraction, transformation et chargement de données à partir des différents systèmes transactionnels.
7. Créer les différentes composantes de restitution à savoir : Les tableaux de bord, les rapports, les analyses...
8. Déployer et maintenir la solution. [15]

## **I.8 Conclusion**

Ce chapitre introductif nous a permis de décrire la Business intelligence et son importance à sélectionner et offrir des informations pertinentes pour les entreprises dans le but de faciliter la prise de décision et la compréhension du fonctionnement et de l'anticipation des actions pour un pilotage éclairé de l'entreprise.

Le chapitre suivant sera consacré à une recherche exhaustive des plus importantes solutions Open Source de Business intelligence qui existent à l'heure actuelle.



## **CHAPITRE II**

# **Comparaison entre divers solutions Open Source de Business intelligence**

### **II.1 Introduction**

D'années en années, l'Open Source n'a cessé de toucher de nouveaux domaines d'applications et le domaine de la Business Intelligence a vu l'apparition des logiciels libres, ceux-ci couvrant tous les aspects du décisionnel : reporting, dashboards, analyse multidimensionnelles, data-minining, et bien sûr ETL.

Le monde du logiciel libre propose un nombre important de solutions et techniques d'aide à la décision, Cela nécessite une recherche exhaustive de meilleures offres décisionnelles Open Source existantes.

Ce second chapitre comporte donc une étude des principales solutions de BI Open Source, une présentation relativement complète de ces outils, de leurs forces, de leurs limites, leur aptitude à satisfaire des besoins opérationnels.

### **II.2 Les composants décisionnelles**

Avant de s'orienter vers la création de solutions décisionnelles complètes, les projets open source se concentraient chacun sur un point bien précis du décisionnel.

Dans cette partie, nous allons présenter les principaux composants décisionnels disponibles en open source, que l'on peut regrouper dans les catégories suivantes :

## II.2.1 Les ETL

À l'origine, les solutions d'ETL sont apparues pour le chargement régulier de données agrégées dans les Datawarehouses, avant de se diversifier vers les autres domaines logiciels. Ces solutions sont largement utilisées dans le monde bancaire et financier, ainsi que dans l'industrie, au vu de la multiplication des nombreuses interfaces. Parmi les solutions les plus réputés nous avons :

### II.2.1.1 Talend Open Studio

Talend Open Studio est un ensemble puissant et flexible de produits Open source pour développer, tester, déployer et administrer des projets de gestion de données et d'intégration d'applications. Talend propose la seule plateforme unifiée simplifiant la gestion de données et l'intégration d'applications en fournissant un environnement unifié pour gérer le cycle de vie complet à travers les frontières de l'entreprise. La productivité des développeurs est considérablement améliorée grâce à un environnement graphique simple à utiliser, basé Eclipse, combinant l'intégration de données, la qualité de données, le MDM, l'intégration d'applications et le Big Data. [16]

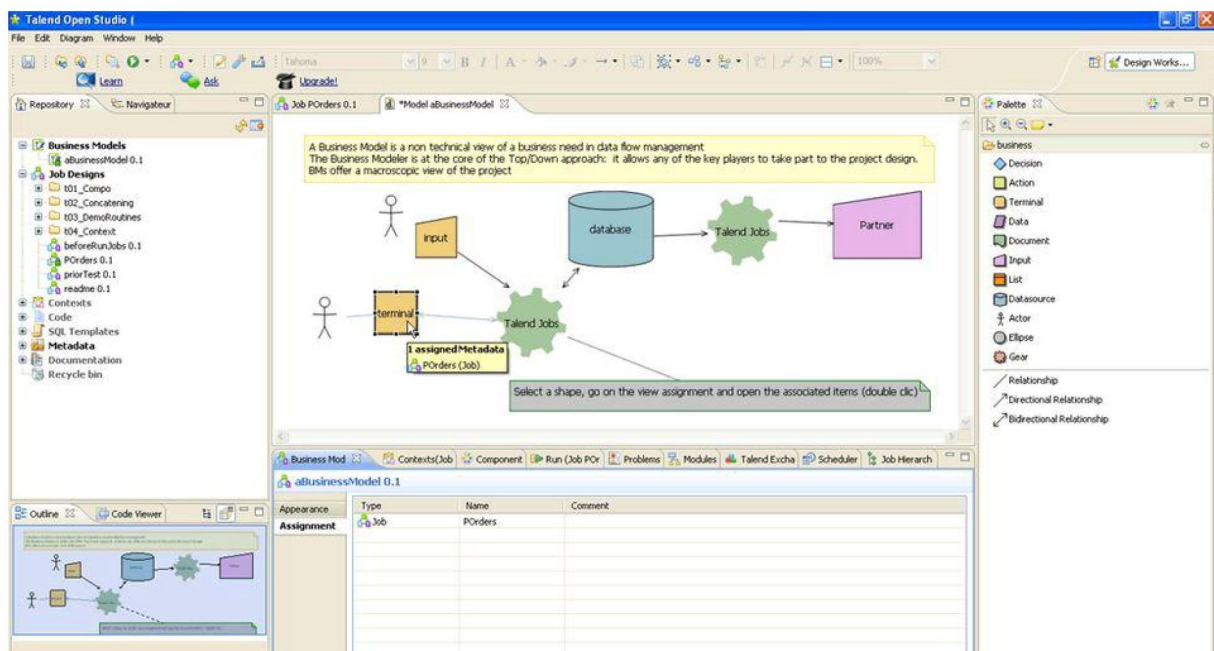


Figure II.1 : Transformation et intégration des données sous Talend Open Studio. [16]

### **II.2.1.2 Pentaho Data Integration**

Pentaho Data Integration (PDI), longtemps connu sous le nom de Kettle, est un ETL open source qui permet de concevoir et d'exécuter des opérations de manipulation et de transformation de données.

Pentaho Data Integration fournit les fonctionnalités suivantes :

- Outil d'ETL (extraction, transformation et chargement) graphique pour charger et traiter les sources de Big Data avec des moyens familiers.
- Vaste bibliothèque de composants préconfigurés permettant d'accéder aux données et de les transformer à partir d'une gamme de sources complète.
- Interface visuelle permettant d'appeler du code personnalisé et d'analyser des images et des fichiers vidéo afin de créer des métadonnées utiles.
- Transformations dynamiques au moyen de variables pour déterminer les règles de mappage, de validation et d'enrichissement des champs.
- Débogueur intégré à des fins de test et de mise au point de l'exécution des tâches.

[17]

### **II.2.2 Les outils de reporting**

Ils existent d'innombrables solutions Open Source de reporting, cependant leur efficacité à générer des rapports de qualités sous divers formats varie d'une solution à l'autre. Voici les meilleures solutions de reporting connues à ce jour :

#### **II.2.2.1 JasperReports**

JasperReports est un moteur de rapport développé par la société JasperSoft et distribué sous une licence open source. Il permet la génération de rapports au format PDF, HTML, XML, CSV, RTF, XLS et TXT. Il utilise JFreeChart afin de générer les graphiques et peut être intégré dans toute application développée avec le langage Java.

Il supporte, en tant que source de données, les bases de données classiques ainsi que les serveurs d'analyse multidimensionnelle, ce qui permet d'exploiter les possibilités du serveur Mondrian directement dans un rapport JasperReports. [12]



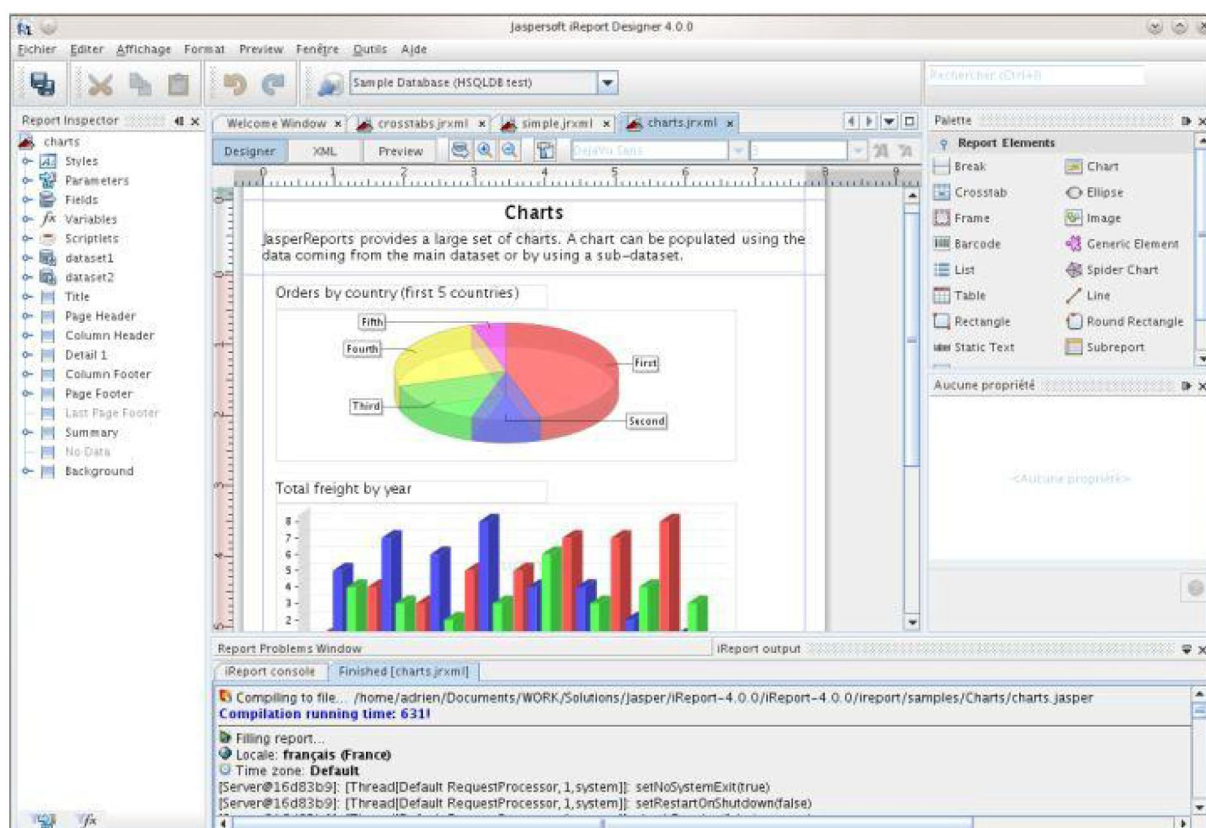


Figure II.2 : Interface graphique : JasperReports. [12]

### II.2.2.2 Pentaho Report Designer

Appartenant à la suite Pentaho, Pentaho Report Designer permet de développer des rapports complexes et, en association avec la plateforme Pentaho, de les publier directement sur le serveur décisionnel. Il supporte les formats PDF, HTML, CSV, Excel, RTF et Texte. Les rapports sont définis en XML, lisibles et pouvant même être manipulés par programmation simple. Ils sont empaquetés, avec les requêtes et les éventuels sous-rapports, dans une archive d'extension .prpt, interprétée par le serveur Web de la suite Pentaho. Il est par ailleurs possible de générer des rapports en masse en utilisant l'ETL Pentaho Data Integration, dans lequel on retrouve une étape de génération de rapports faisant appel au fichier prpt conçu avec Pentaho. [12]



### **II.2.2.3 QlikView**

QlikView est un outil de reporting ayant pour objectif de fournir aux décideurs de l'entreprise des tableaux d'analyse et un ensemble d'indicateurs lui permettant de disposer d'une véritable plateforme d'aide à la décision.

Grace à QlikView, l'utilisateur peut appréhender de façon différente son activité de en :

- Consolidant des données pertinentes issues de différentes sources dans une seule et même application.
- Développant la prise de décisions collaboratives en temps réel et de façon sécurisée.
- Visualisant vos données à l'aide de graphiques soignés et performants.
- Utilisant des applications, des tableaux de bord et des analyses dynamiques. [20]

### **II.2.3 Les outils d'analyse multidimensionnelle**

L'analyse multidimensionnelle est l'un des modes d'analyse les plus courants dans le décisionnel dans le but d'obtenir des rapports de synthèse tels que ceux qui sont utilisés en analyse financière. Parmi les solutions Open Source des analyses OLAP il ya :

#### **II.2.3.1 Palo**

Palo est une base de données multidimensionnelle, en mémoire, développée en C et distribuée sous licence GPL par la société Jedox.

Palo fait partie de la catégorie M-OLAP (Multidimensional OLAP) : toutes les données sont chargées en mémoire et non dans une base de données relationnelle. Les données sont calculées à la volée, ce qui lui permet d'obtenir de très bons temps de réponse.

La base Palo est accessible depuis le tableur Microsoft Excel – ou OpenOffice Calc – avec un plugin, ce qui permet de naviguer dans les données multidimensionnelles directement dans les feuilles de calcul tant appréciées des utilisateurs. [12]

C6		=PALO.ENAME(\$A\$1;"Months";"Year";3;"Year";"English")					
	A	B	C	D	E	F	G
1	localhost/Demo						
2	Sales						
3	2007						
4	Actual						
5							
6			Year			Qtr.1	
7			Gross Profit	Turnover	Cost of Sales	Gross Profit	Turnover
8	Europe	All Products	4 856 435.58	48 416 679.88	43 560 244.31	909 524.88	8 650
9	West	All Products	2 389 264.24	23 889 485.46	21 500 221.22	446 607.44	4 110
10	Germany	All Products	591 481.73	5 684 177.95	5 092 696.22	102 973.82	893
11	France	All Products	457 628.14	4 733 279.97	4 275 651.82	79 394.14	916
12	Switzerland	All Products	233 608.87	2 478 084.49	2 244 475.62	53 376.08	460
13	Netherlands	All Products	190 428.31	1 935 567.46	1 745 139.15	33 205.88	316
14	Belgium	All Products	266 765.64	2 763 440.67	2 496 675.03	50 446.53	438
15	Luxembourg	All Products	49 276.45	516 069.67	466 793.22	9 340.22	88
16	United Kingdom	All Products	513 951.48	4 861 940.52	4 347 989.04	103 414.90	849
17	Ireland	All Products	86 123.62	916 924.74	830 801.12	14 455.86	147
18	East	All Products	937 077.06	8 900 423.24	7 963 346.18	156 939.66	1 579
19	South	All Products	926 395.75	9 651 405.67	8 725 009.92	191 342.58	1 858
20	North	All Products	603 698.53	5 975 365.51	5 371 666.99	114 635.21	1 102
23							

Figure II.3 : Navigation dans un cube Palo sous Excel. [12]

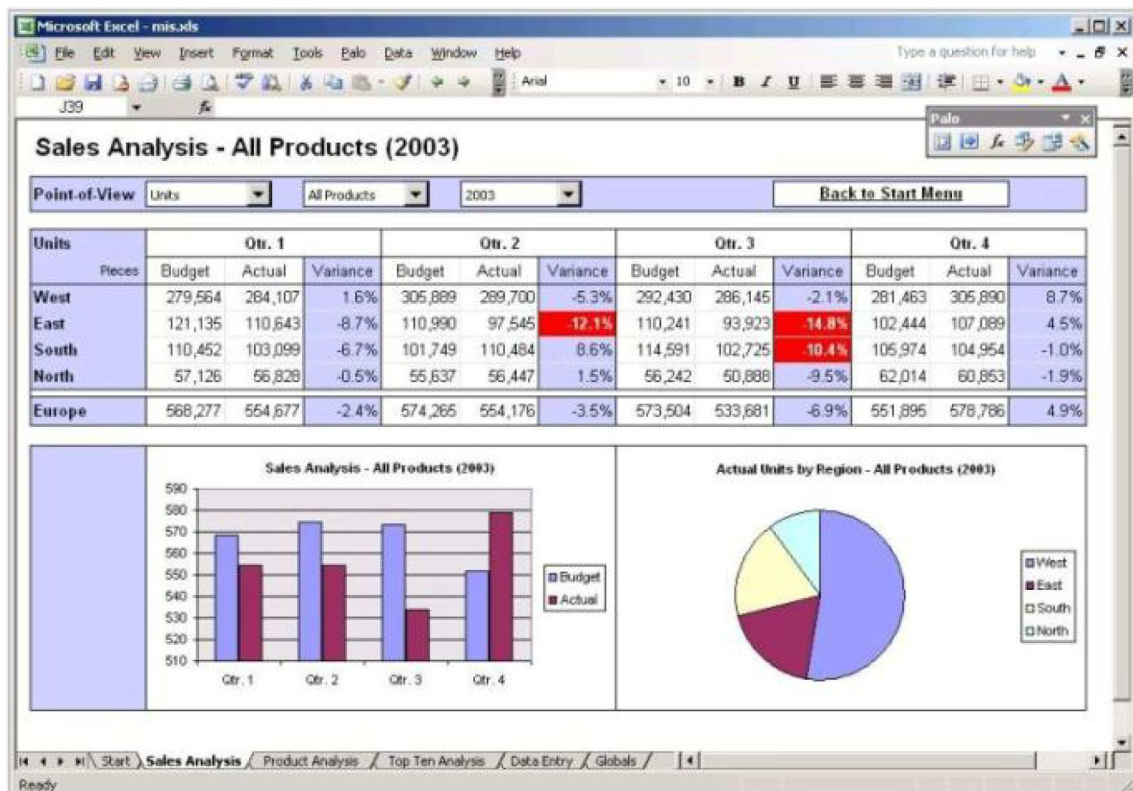


Figure II.4 : Tableau de bord dans MsExcel, constitué avec Palo. [12]



### **II.2.3.2 Mondrian**

Mondrian est un moteur OLAP (Online Analytical Processing) écrit en Java par Julian Hyde qui permet la conception, la publication et le requêtage de cubes multidimensionnels. Mondrian permet l'exécution de requêtes en langage MDX sur des entrepôts de données s'appuyant sur des SGBDR, d'où sa caractérisation de « ROLAP » (Relational OLAP). En matière de ROLAP, Mondrian est la référence open source. Mondrian permet d'accéder aux résultats dans un format multidimensionnel compréhensible par une API de présentation côté client, le plus souvent en mode Web, avec par exemple JPivot, Pentaho Analyzer, Pentaho Analysis Tool, Geo Analysis Tool (G.A.T.). Mondrian s'appuie sur une modélisation OLAP standard et peut donc se connecter à n'importe quel entrepôt de données conçu dans les règles de l'art de la Business Intelligence. [18]

### **II.2.4 Les Datamining**

Les outils de Datamining permettent la construction des modèles à partir des données, c'est-à-dire trouver des structures intéressantes ou des motifs selon des critères fixés au préalable, et d'en extraire un maximum de connaissances utiles à l'entreprise. Nous citons notamment :

#### **II.2.4.1 WEKA**

Weka est un outil permettant d'exécuter des algorithmes de data-mining sur un ensemble de données. Il est ainsi possible d'isoler des populations ou d'extraire des règles à partir des données contenues dans le Datawarehouse. Il se présente sous la forme d'une application indépendante, disposant d'une interface utilisateur graphique ou en ligne de commande. L'utilisateur peut appliquer un à un les différents algorithmes, un workflow de traitements qui pourra par exemple être utilisé dans une plateforme décisionnelle afin d'analyser périodiquement les données. L'utilisation de Weka demande de bonnes connaissances du data-mining et des différents algorithmes statistiques utilisés. [12]

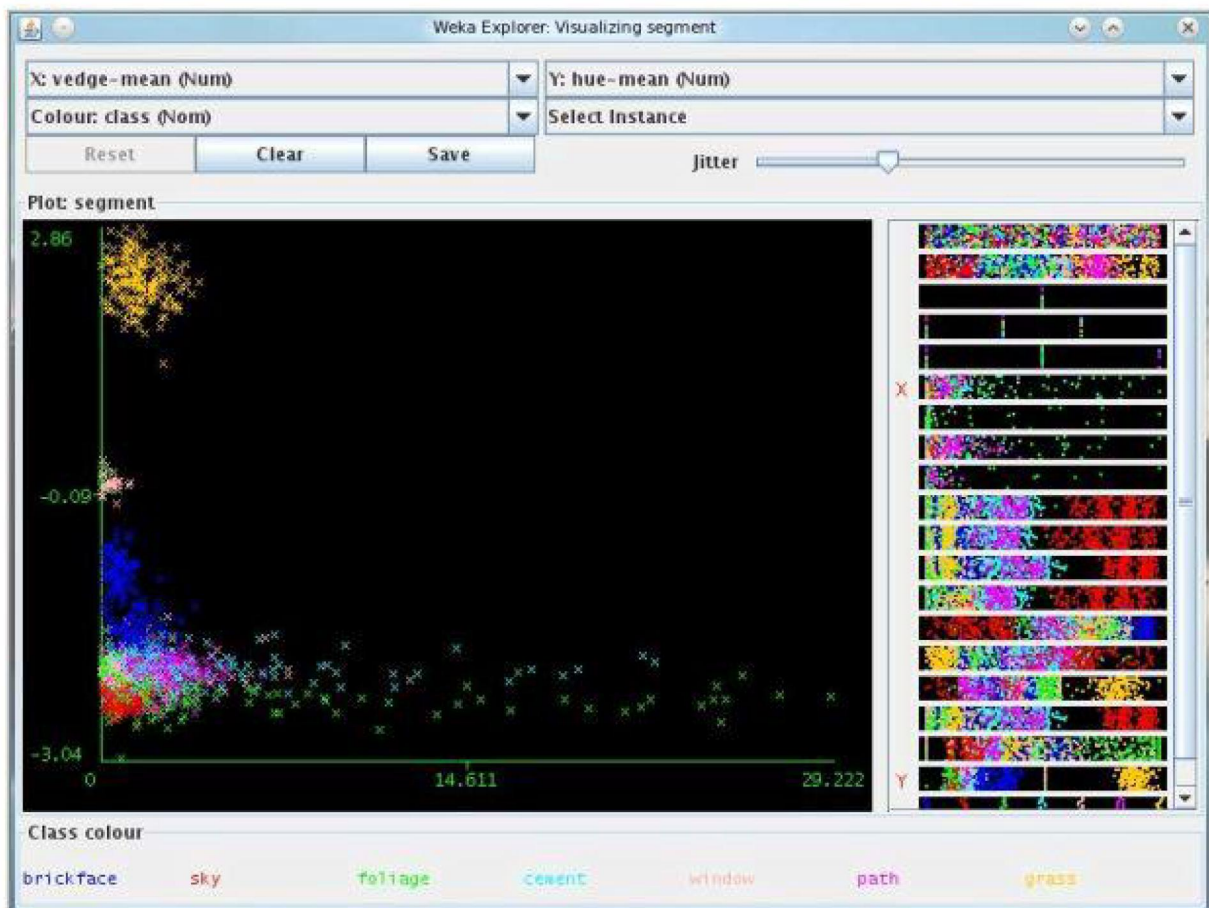


Figure II.5 : Visualisation de données sous WEKA. [12]

#### II.2.4.2 RapidMiner

RapidMiner est un logiciel open source et gratuit dédié au Datamining. Il contient de nombreux outils pour traiter des données : lecture de différents formats d'entrée, préparation et nettoyage des données, statistiques, tous les algorithmes de datamining, évaluation des performances et visualisations diverses. Rédigé avec le langage de programmation Java, cet outil offre des analyses avancées à travers des cadres basés sur un modèle. En plus de l'exploration de données, RapidMiner fournit également des fonctionnalités comme le prétraitement des données et la visualisation, l'analyse prédictive et de modélisation statistique, l'évaluation et le déploiement. [18]

### II.3 Les suites décisionnelles

Une suite décisionnelle est une plateforme logicielle constituée des modules décisionnels présentés ci-avant. Si les suites partageaient autrefois certains des composants open



source que nous venons de présenter, et qu'elles en partagent encore certains, les solutions décisionnelles open source ont maintenant tendance à se différencier afin d'apporter chacune une réponse adaptée aux besoins des projets. [12]

Parmi les solutions décisionnelles open source existantes, nous citons notamment :

### **II.3.1 SpagoBI**

SpagoBI est une suite décisionnelle développée par la société italienne Engineering. Ce projet a été initié en 2005. Elle a comme particularité d'être la seule solution open source 100 % free, une seule version stable avec 100 % des fonctionnalités disponibles.

C'est une suite complète couvrant l'ensemble des besoins de Business Intelligence (ETL, outil reporting, OLAP, dashboard...). Sa force est d'offrir à ses utilisateurs un important panel de solutions analytiques et aux développeurs, testeurs et administrateurs un large éventail d'outils dans leur travail quotidien. Afin de couvrir les différents besoins fonctionnels propre au décisionnel SpagoBI s'appuie sur un ensemble de projet Open Source connexes. Plus qu'une simple solution agrégative, la suite permet de puissantes synergies notamment via son modèle comportemental poussé. [12]

### **II.3.2 Jaspersoft BI**

JasperSoft BI est la plateforme décisionnelle de JasperSoft, société qui développe également le générateur d'états JasperReports disponible depuis 2001. Cette plateforme propose des fonctionnalités de reporting et d'analyse en langage naturel.

En version communautaire, la solution propose la conception et génération de rapports, avec JasperSoft Studio ainsi que leur publication sur un portail web JasperServer.


Contrairement à SpagoBI, Jaspersoft n'est pas entièrement gratuit et comporte une version professionnelle pour la réalisation de rapports avancés, ainsi qu'une version entreprise qui permet à l'utilisateur final de réaliser des analyses sur des cubes OLAP. [12]


### II.3.3 Pentaho BI suite

Pentaho est une suite logicielle décisionnelle complète qui permet la création et la diffusion de documents décisionnels à un grand nombre de destinataires via une interface web. Il comprend en propre tous les éléments nécessaires à un projet décisionnel (ses propres composants décisionnelles déjà cités : Pentaho Data Integration, Pentaho Report Designer...), et est le seul éditeur Open Source capable d'en garantir l'homogénéité dans le temps. [12]

### II.3.4 Comparaisons entre les trois suites décisionnelles

Les deux figures suivantes illustrent les particularités des trois suites décisionnelles « SpagoBi », JasperSoft et Pentaho.

 <b>SpagoBI</b>	
<b>ETL</b>	Talend Open Studio.
<b>Générateur d'état</b>	Jasper Report. BIRT.
<b>Analyse Multidimensionnelle</b>	Jpivot. Mondrian.
<b>Data Mining</b>	Weka.
<b>Portail</b>	eXo platform.
<b>Workflow</b>	Propre à Spago BI.

 <b>Jasper Intelligence</b>	
<b>ETL</b>	Jasper ETL : <ul style="list-style-type: none"><li>• Talend Open Studio.</li></ul>
<b>Générateur d'état</b>	Jasper Report.
<b>Analyse Multidimensionnelle</b>	Jasper Analysis : <ul style="list-style-type: none"><li>• Mondrian.</li><li>• Jpivot.</li></ul>
<b>Planificateur</b>	Quartz.


 <b>Pentaho</b>	
<b>ETL</b>	Kettle.
<b>Générateur d'état</b>	Eclipse BIRT. Jasper Reports. JfreeReport.
<b>Analyse Multidimensionnelle</b>	Jpivot. Mondrian.
<b>Data Mining</b>	Weka.
<b>Portail</b>	Jboss Portal.
<b>Planificateur</b>	Quartz.
<b>Workflow</b>	Enhydra shark.

Figure II.6 : Les composants décisionnels des trois suites suivantes : SpagoBI, JasperSoft et Pentaho. [4]

 <b>SpagoBI</b> ✓ Solution riche et mature, mais à un degrés moindre que Pentaho. ✓ Les meilleures solutions sont intégrées. ✓ A fait le choix de choisir.	 <b>Jasper Intelligence</b> ✓ Accessible, simple de prise en main, fonctionnalités principales présentes. ✗ Absence de BIRT. ✗ JasperSoft met en avant ses solutions.
 <b>Pentaho</b> ✓ Fonctionnalités, communauté, visibilité et taux d'activité très fort. ✓ Logique d'intégration des projets... ✗ ... peut être trop même.	

Figure II.7 : Analyse des fonctionnalités des suites décisionnelles sur [www.osbi.fr](http://www.osbi.fr). [4]

## II.4 Critiques

Le monde de la Business intelligence dispose d'un bon nombre de solutions Open source qui offre à l'utilisateur divers possibilités pour réaliser un projet décisionnel selon ses besoins.

Nous constatons tout de même, en vue de cette recherche, que Talend Open Studio est assez recommandé comme ETL pour l'intégration de données qui comptabilise à l'heure actuelle des millions de téléchargements et distribuant le code source de ses modules socle sous des licences Open Source, offrant ainsi à la communauté de développeurs la possibilité d'améliorer le produit.

JasperReports et QlikView représentent d'excellents outils de reporting et sont souvent recommandés pour leurs rapidités de mise en œuvre et leurs finesses graphiques et simplicité d'utilisation. Par contre QlikView présente l'inconvénient d'être obligé d'acheter la version intégrale pour une intégration au Web contrairement à JasperReports et Pentaho.

Les outils d'analyse multidimensionnelle et de Datamining existent de manière plus amoindris par rapport aux technologies cités ci-dessus mais proposent à leur tour des solutions Open Source assez efficace, à l'image de Jedox Palo et Weka.



Quant aux suites décisionnelles, et malgré la présence de concurrents de taille tels que Pentaho ou Jaspersoft, SpagoBI est considéré comme étant la seule suite décisionnelle intégralement Open Source, c'est-à-dire que l'ensemble de ses fonctionnalités sont librement accessibles et offrent tous les outils nécessaires à l'utilisateur pour mener à bien son projet décisionnel.

## **II.5 Conclusion**

À travers ce chapitre, nous avons fait la découverte d'un ensemble de solutions et de technologies Open Source capable d'assurer la réalisation d'un projet décisionnel de manière complète et fiable, que ce soit en combinant des composants décisionnels de toute sorte, ou en exploitant des suites décisionnelles comportant tous les outils nécessaires.



## **CHAPITRE III**

# **Conception et analyse des besoins**

### **III.1 Introduction**

La modélisation est une étape très importante dans le développement d'un projet informatique, il s'agit d'un processus ayant pour objectif de spécifier les besoins et les exigences des acteurs, le système et l'architecture globale du travail, et ceux, sous forme de schémas et de diagrammes.

Dans ce troisième chapitre intitulé « conception » nous allons tous d'abord présenter l'organisme d'accueil, nous enchaînerons par une analyse des besoins de nos travaux puis nous finaliserons par l'ensemble des diagrammes modélisant notre projet.

### **III.2 Etude préliminaire**

Nous abordons ce chapitre par la présentation de l'organisme d'accueil chargé de notre stage qui est l'Entreprise Portuaire de Béjaïa (EPB).

#### **III.2.1 Présentation du port de Béjaïa**

Le port de Béjaïa est un port algérien se situant dans la région de Kabylie dans le nord du pays. Il est notamment consacré au commerce international et aux hydrocarbures.

Il joue un rôle très important dans les transactions internationales vu sa situation géographique qui se résume comme suit :

- Délimité au nord par la route nationale N°9.
- Au sud par les jetées de fermeture et du large sur une largeur de 2 750m.
- À l'est par la jetée Est.
- À l'ouest par la zone industrielle de Bejaia. [19]



Figure III.1 : Port de Béjaïa. [19]

### **III.2.2 Historique du port de Béjaïa**

Au cœur de l'espace méditerranéen, la ville de Bejaia possède de nombreux sites naturels et vestiges historiques datant de plus de 10 000 ans, ainsi que de nombreux sites archéologiques recelant des objets d'origine remontant à l'époque néolithique.

Bejaia joua un grand rôle dans la transmission du savoir dans le bassin méditerranéen, grâce au dynamisme de son port, la sécurité de la région, la bonne politique et les avantages douaniers. Bougie a su attirer beaucoup de puissants marchands.

La Saldæ romaine devient un port d'embarquement de blé du grenier de Rome, ce n'est qu'au XI<sup>ème</sup> siècle, que Bgayeth, devenue Ennaceria, pris une place très importante dans le monde de l'époque ; le port de Bejaia devient l'un des plus importants de la méditerranée.

La réalisation des ouvrages actuels débuta en 1834, elle fut achevée en 1987. C'est en 1960 qu'a été chargé le premier pétrolier d'Algérie.

Aujourd'hui, le port de Béjaïa est réputé pour sa mixité : hydrocarbures et marchandises générales y sont traités. L'aménagement moderne des superstructures, le développement des infrastructures, l'utilisation de moyens de manutention et de techniques adaptés à l'évolution de la technologie des navires et enfin ses outils de gestion moderne, ont fait évoluer le Port de Béjaïa depuis le milieu des années 1990 pour être classé aujourd'hui second port d'Algérie.

### III.2.3 Activités de l'EPB

L'EPB a pour mission la gestion, l'exploitation et le développement du domaine portuaire du port, et ceux, dans le but de promouvoir les échanges extérieurs du pays. Elle se doit d'assumer la sécurité au sein du pays et est chargée des travaux d'entretien, d'aménagement, de renouvellement et de création d'infrastructures.

L'EPB assure également des prestations à caractère commercial, à savoir le remorquage, la manutention et l'acconage. [19]

### III.2.4 Organigramme général de l'EPB

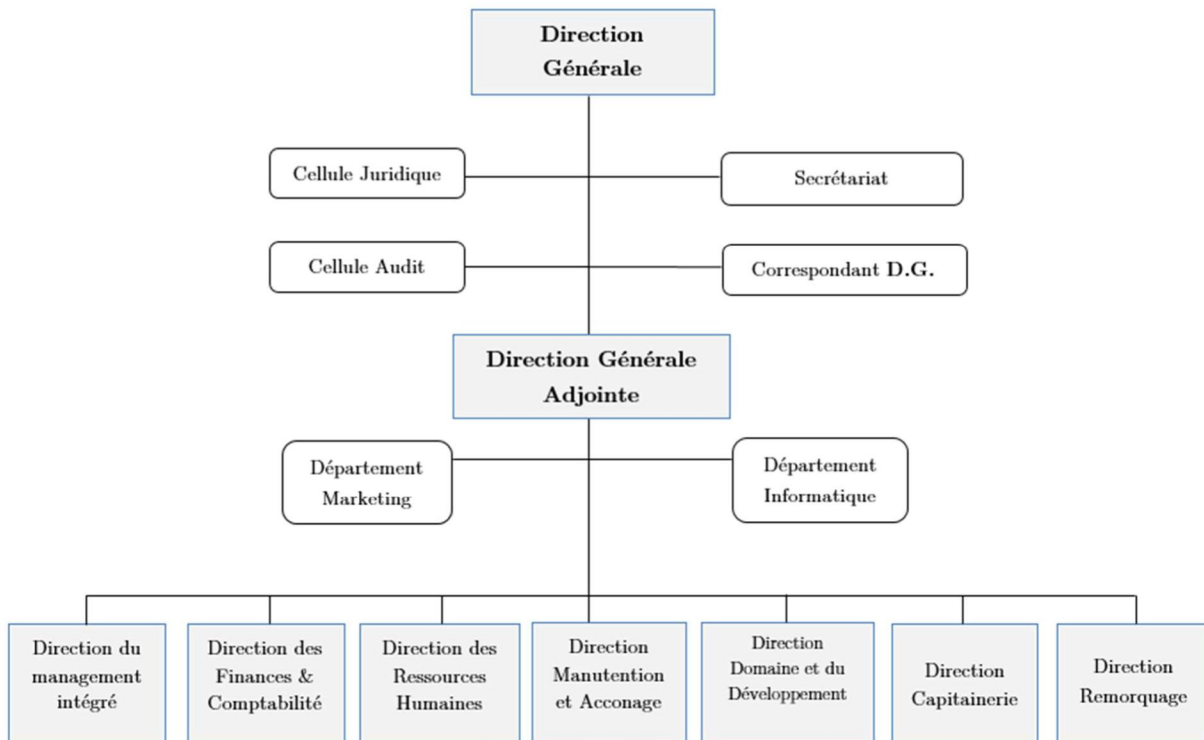


Figure III.2 : Organigramme général de l'EPB. [19]



**La direction de remorquage (DR) :** Elle est chargée d'assister le pilote du navire lors de son entrée et de sa sortie du quai. Son activité consiste essentiellement à remorquer les navires entrants et sortants, ainsi que la maintenance des remorqueurs. Les prestations sont :

- Le Remorquage portuaire.
- Le Remorquage hauturier (haute mer).
- Le Sauvetage en mer.

### **III.3 Analyse des besoins**

Ce projet a été initié en ayant comme objectif de répondre aux attentes des décideurs de l'EPB concernant l'analyse des données .Il est donc impératif de passer par une étape de collection d'informations et de besoins.

Nous focaliserons notre travail sur la direction de remorquage, cette dernière comporte un ensemble de contraintes capables d'être cernés et écartés grâce à la mise en place d'un système décisionnel.

#### **III.3.1 Définition des besoins**

Pour définir les besoins décisionnels de l'entreprise, Nous avons organisé un entretien avec l'ingénieur en recherche opérationnelle chargé des tableaux de bord indiquant les performances de chaque entité au sein de l'EPB. L'entretien a pour but d'élaborer un questionnaire et d'aboutir à des informations sous forme de compte rendu des besoins à produire dans le système décisionnel.

Le tableau suivant comporte l'ensemble des questions posées au décideur et des résumés des réponses qui représentent aussi les besoins de l'entreprise :



Question	Réponse
Quels sont les processus métiers que vous pilotez dans ce secteur ?	<ul style="list-style-type: none"> <li>Elaboration de documents évaluant la prestation de chaque entité au niveau du processus de remorquage.</li> </ul>
Quels sont les objectifs de votre activité ?	<ul style="list-style-type: none"> <li>Veiller à ce que le processus de remorquage se déroule dans les normes grâce au suivi des projets via des indicateurs de performances.</li> </ul>
Comment évaluez-vous la réussite de votre activité ?	<ul style="list-style-type: none"> <li>La réalisation des différentes tâches dans les meilleurs délais.</li> </ul>
Quels sont les problèmes rencontrés dans vos travaux ?	<ul style="list-style-type: none"> <li>La difficulté d'exploiter un grand nombre d'informations issues de différentes bases de données.</li> </ul>
Quels sont les bons indicateurs pour la prise de décision ?	<ul style="list-style-type: none"> <li>Principalement les taux d'activité et de performance de chaque entité.</li> </ul>
Quelles sont vos attentes quant à notre solution décisionnelle finale ?	<ul style="list-style-type: none"> <li>Accélérer la récolte d'informations grâce à une vision consolidée des données à l'intérieur des tableaux de bord.</li> </ul>

Tableau III.1 : Questionnaire pour la direction de remorquage.

### III.3.2 Compte rendu des besoins

Le tableau suivant présente un compte rendu des résultats de la collecte d'informations auprès de la direction de remorquage de l'EPB :

Compte rendu de l'entretien
<b><u>Contexte de recueil :</u></b> La direction de remorquage a fortement approuvé la mise en place d'une solution d'aide à la décision dans le but de faciliter l'exploitation de leurs bases de et aboutir à des résultats efficaces pour la prise de décision.
<b><u>Le choix de la direction de remorquage :</u></b> La direction de remorquage a éprouvé des difficultés durant longtemps à acquérir des informations pertinentes issues de différentes bases de données, et donc une solution informatisée s'impose afin de faire face à ce problème.
<b><u>Les objectifs fixés :</u></b> <ul style="list-style-type: none"><li>• Récolter un maximum d'informations pertinentes et utiles de la base de données concernant le remorquage.</li><li>• Générer des tableaux de bord contenant des rapports, tables ainsi que des graphiques illustrant l'essentiel de la base de données.</li><li>• Faciliter la tâche du responsable des tableaux de bord en exploitant ces informations.</li></ul>

Tableau III.2 : Compte rendu des besoins.

### III.4 Processus d'ETL (Extraction-Transformation-Chargement)

Le processus d'ETL est l'étape qui consiste à collecter des informations à partir de différentes sources de données afin de les unifier et avoir en sortie de nouvelles données utiles et exploitables.

La collecte de données passe par une connexion à la base de données concernée pour acquérir les sources de données et ensuite l'exécution de requêtes SQL pour extraire ces mêmes données. La transformation sera faite au niveau de l'ETL via le mapping qui est un procédé permettant d'accéder aux données quelques soit leurs structures et leurs

formats, et ceux à travers un ensemble de fonctions définies. À la fin, les données seront triées et chargées sous le format choisi.

Le diagramme d'activité suivant illustre le processus d'ETL :

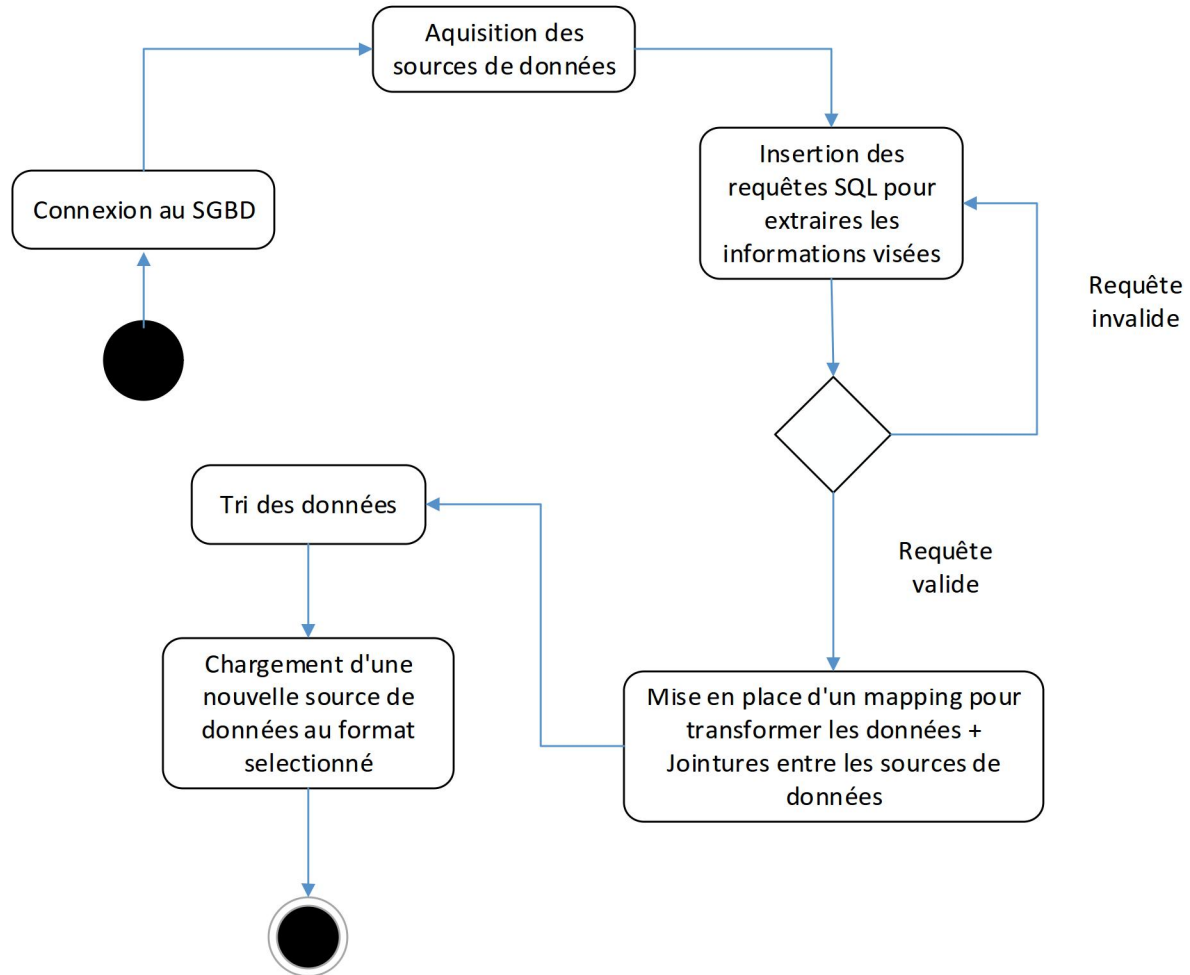


Figure III.3 : Diagramme d'activité du processus d'ETL.

### III.5 Conception des éléments de reporting

Une fois que les nouvelles sources de données sont chargées, Nous utiliserons un outil de reporting pour générer des rapports caractérisés par de multiples objets qui seront visualisés pour l'aide à la décision.

L'utilisateur pourra gérer ses sources de données ainsi que son tableau de bord, Il crée des feuilles de travail et des objets pour diffuser les résultats sous différentes formes.

L'utilisateur peut aussi effectuer des modifications sur les feuilles, les objets et les sources de données, ou les supprimer.

La figure suivante représente le diagramme des cas d'utilisation du reporting :

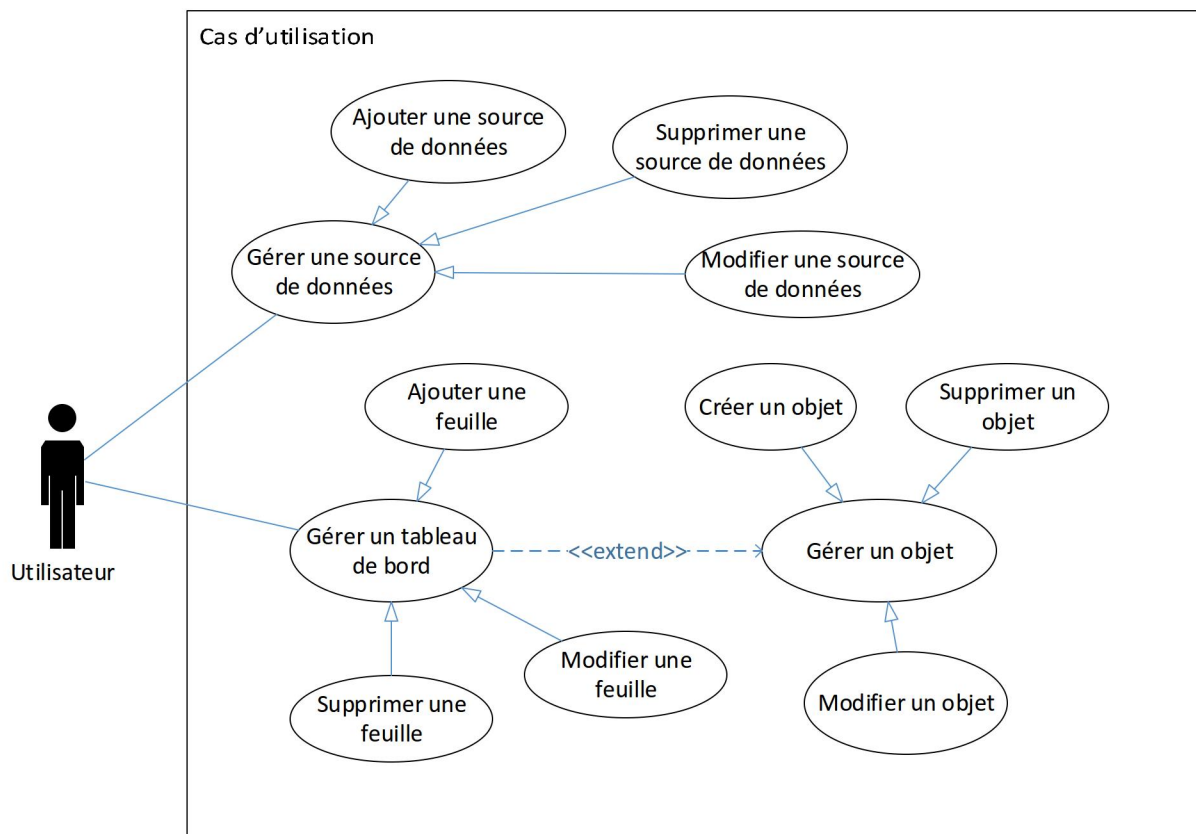


Figure III.4 : Diagramme des cas d'utilisation du reporting.

Nous allons décrire les différents cas d'utilisation comme suit :

### III.5.1 Cas d'utilisation « Gérer une source de données »

Dans les menus de l'outil de reporting, l'utilisateur peut accéder à la liste des sources de données que comporte la base de données du logiciel, il peut ajouter de nouvelles sources, supprimer ou modifier ces dernières.



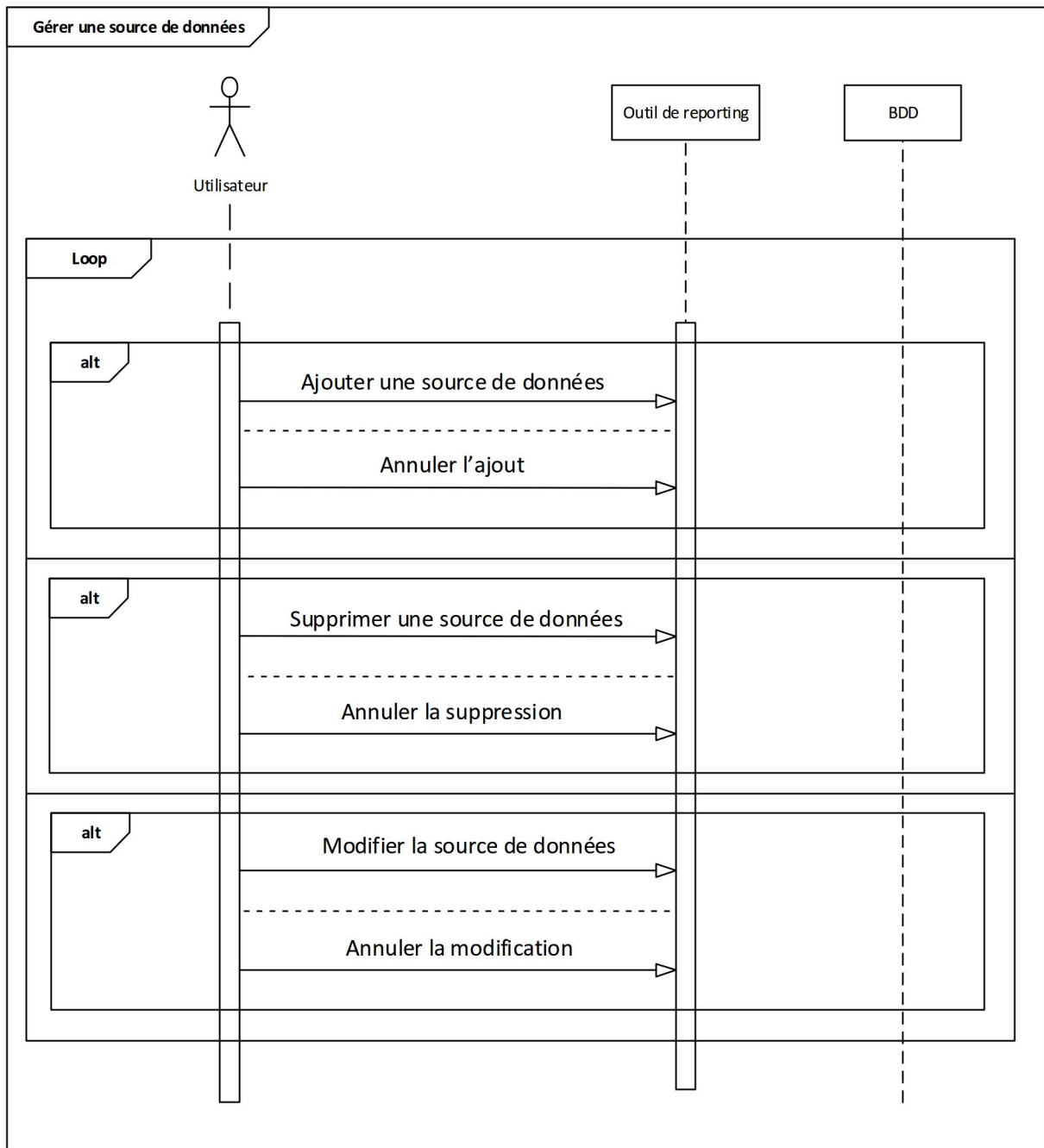


Figure III.5 : Diagramme de séquence du cas d'utilisation « Gestion d'une source de données ».

### III.5.2 Cas d'utilisation « Ajouter une source de données »

L'utilisateur peut ajouter une nouvelle source de données dans la base de données, une fois la requête d'enregistrement faite, la liste des données sera mise à jour.

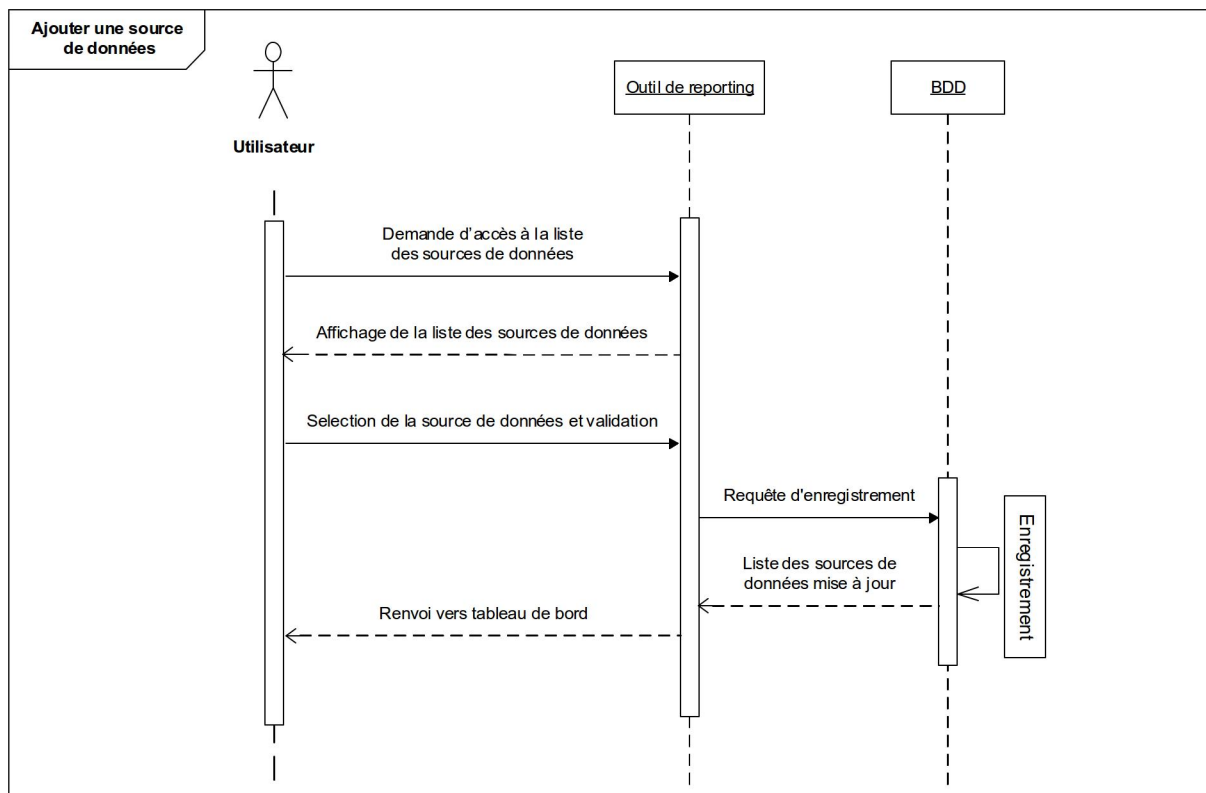


Figure III.6 : Diagramme de séquence du cas d'utilisation « Ajouter une source de données ».

### III.5.3 Cas d'utilisation « Supprimer une source de données »

L'utilisateur peut effectuer une suppression au niveau des sources de données et mettre à jour la liste.

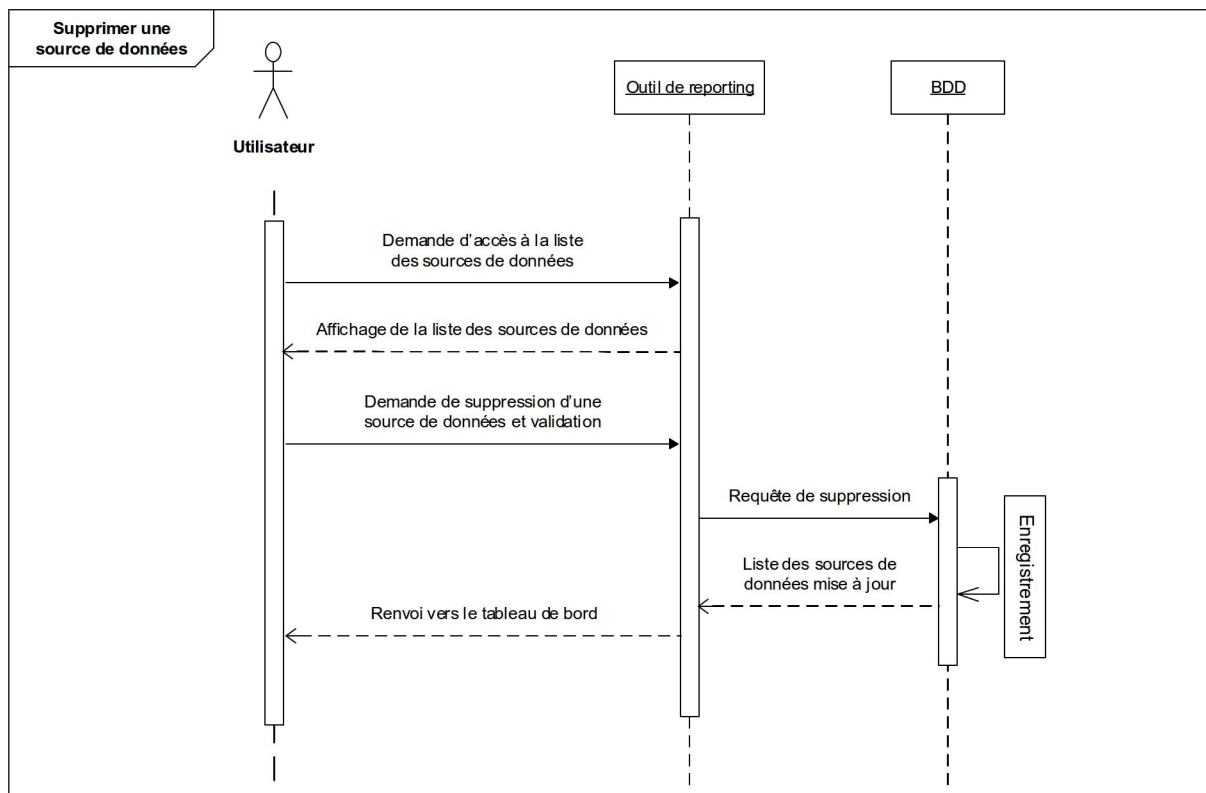


Figure III.7 : Diagramme de séquence du cas d'utilisation « Supprimer une source de données ».



### III.5.4 Cas d'utilisation « Modifier une source de données »

Outre la suppression, l'utilisateur peut aussi effectuer des modifications au niveau de la liste des sources de données et la mettre à jour.

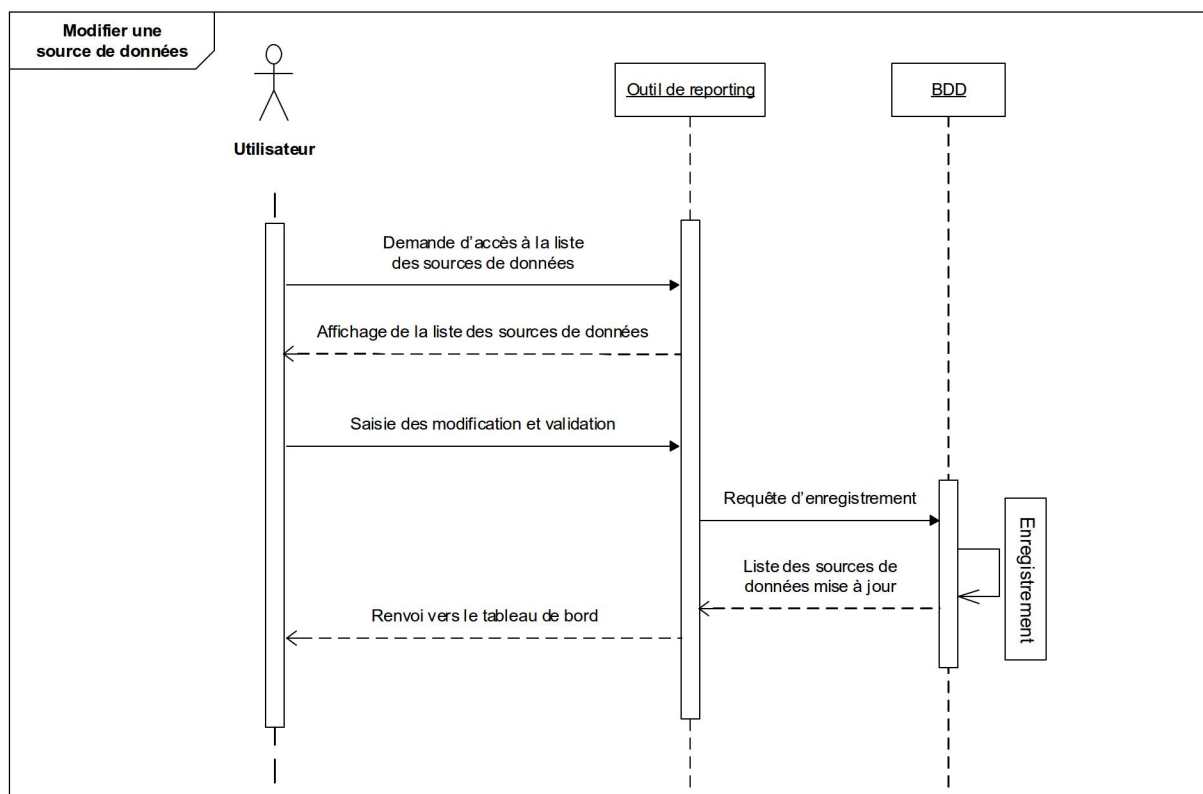


Figure III.8 : Diagramme de séquence du cas d'utilisation « Modifier une source de données ».

### III.5.5 Cas d'utilisation « Gérer un tableau de bord »

Au lancement de l'outil de reporting, l'utilisateur accède au tableau de bord ou il pourra le manipuler en créant des feuilles de travail (et éventuellement la suppression et la modification).

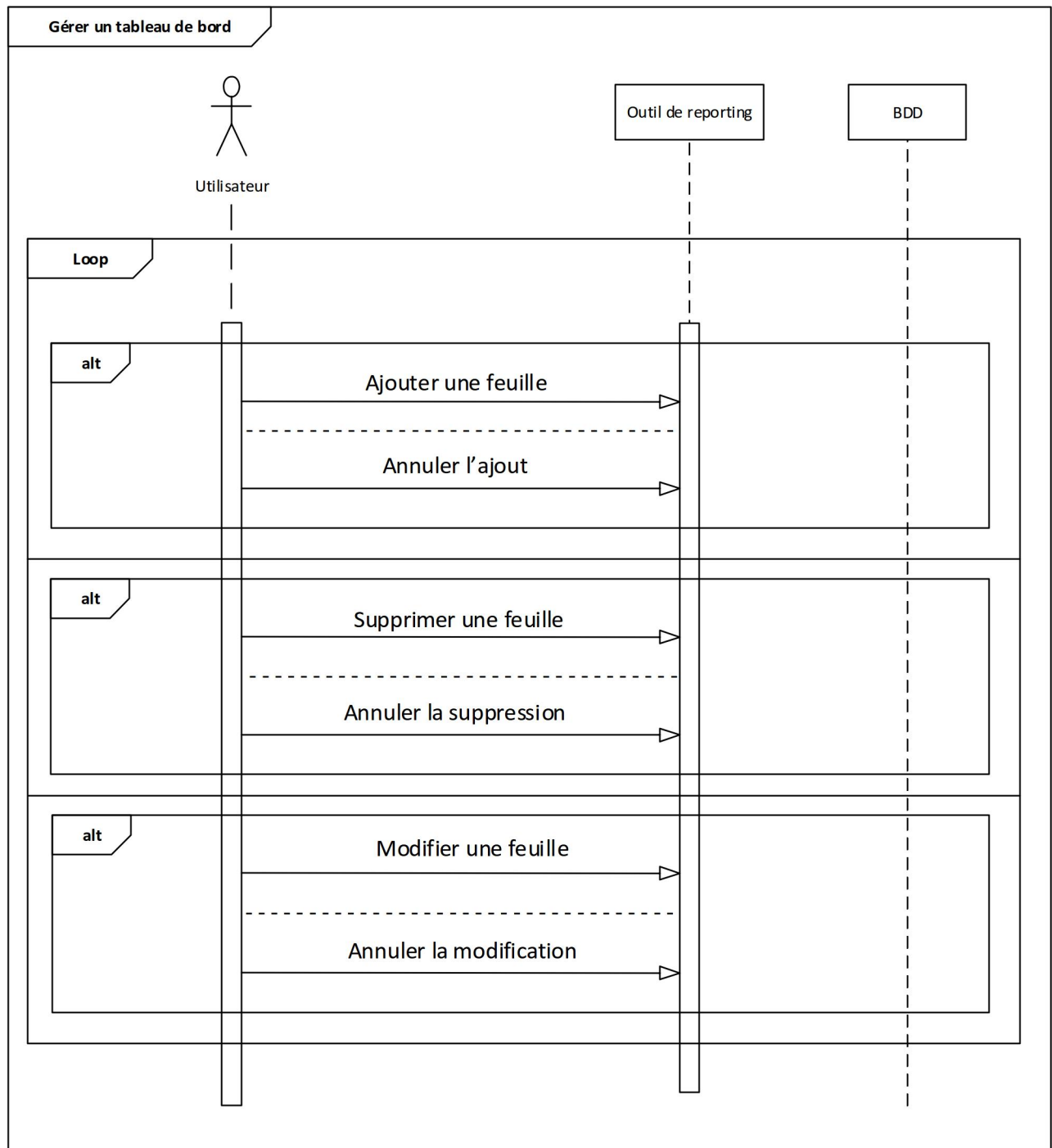


Figure III.9 : Diagramme de séquence du cas d'utilisation « Gestion du tableau de bord ».

### III.5.6 Cas d'utilisation « Ajouter une feuille »

L'utilisateur peut créer une ou plusieurs feuilles de travail dans son tableau de bord afin d'y insérer des objets.

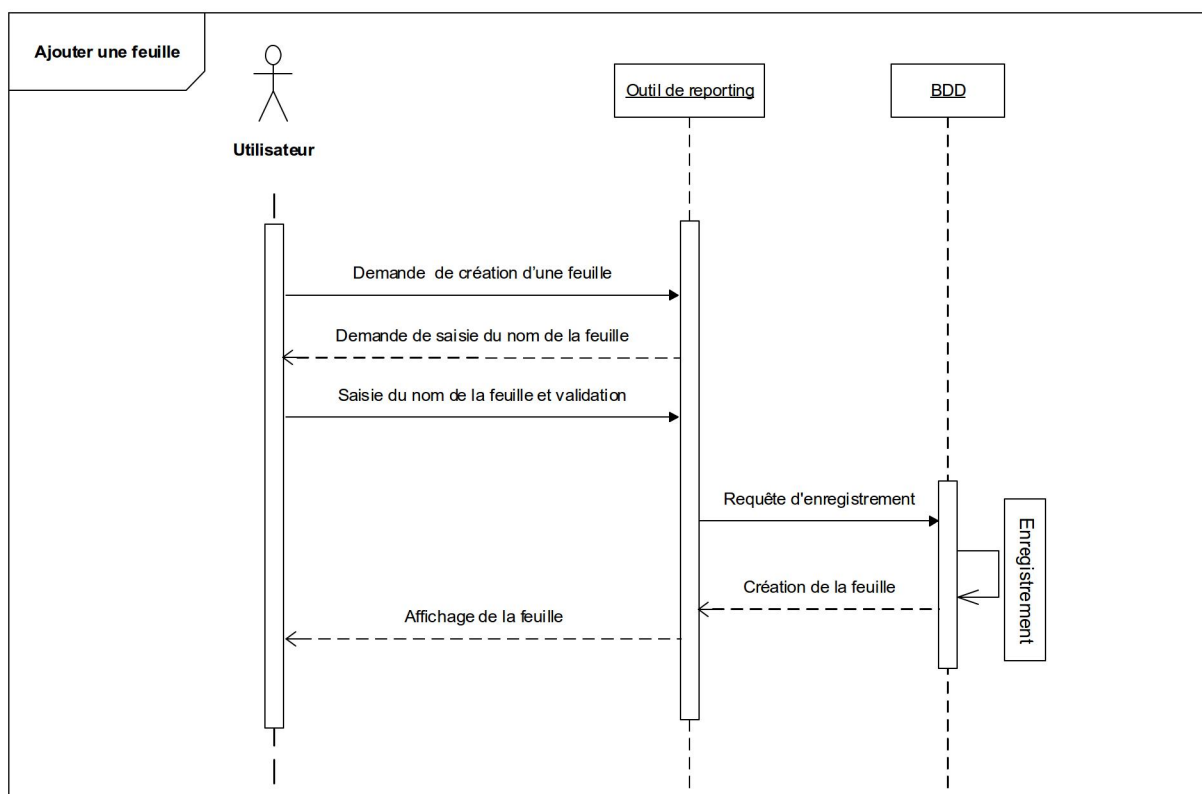


Figure III.10 : Diagramme de séquence du cas d'utilisation « Ajouter une feuille ».



### III.5.7 Cas d'utilisation « Supprimer une feuille »

Si l'utilisateur est désireux de supprimer une feuille, il n'a qu'à la sélectionner et l'enlever.

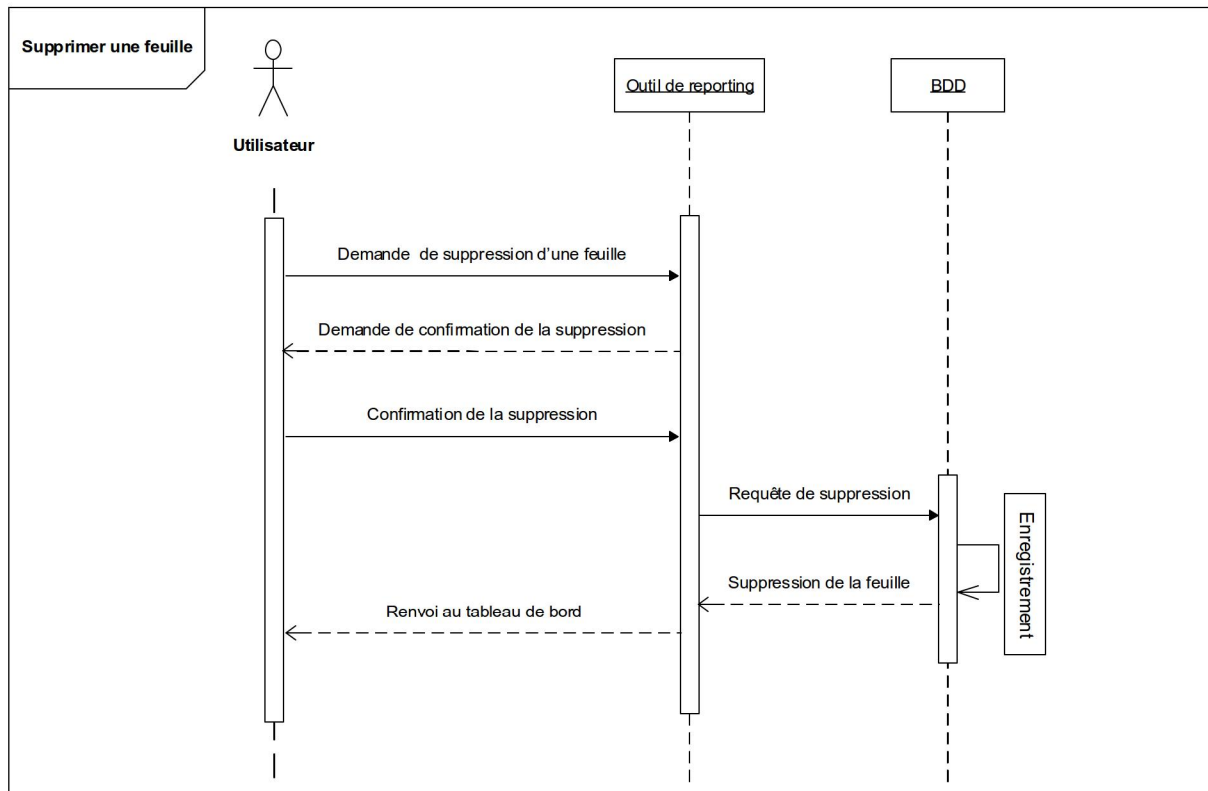


Figure III.11 : Diagramme de séquence du cas d'utilisation « Supprimer une feuille ».

### III.5.8 Cas d'utilisation « Modifier une feuille »

L'utilisateur peut modifier une feuille de travail (nom de la feuille, emplacement, contenu...).

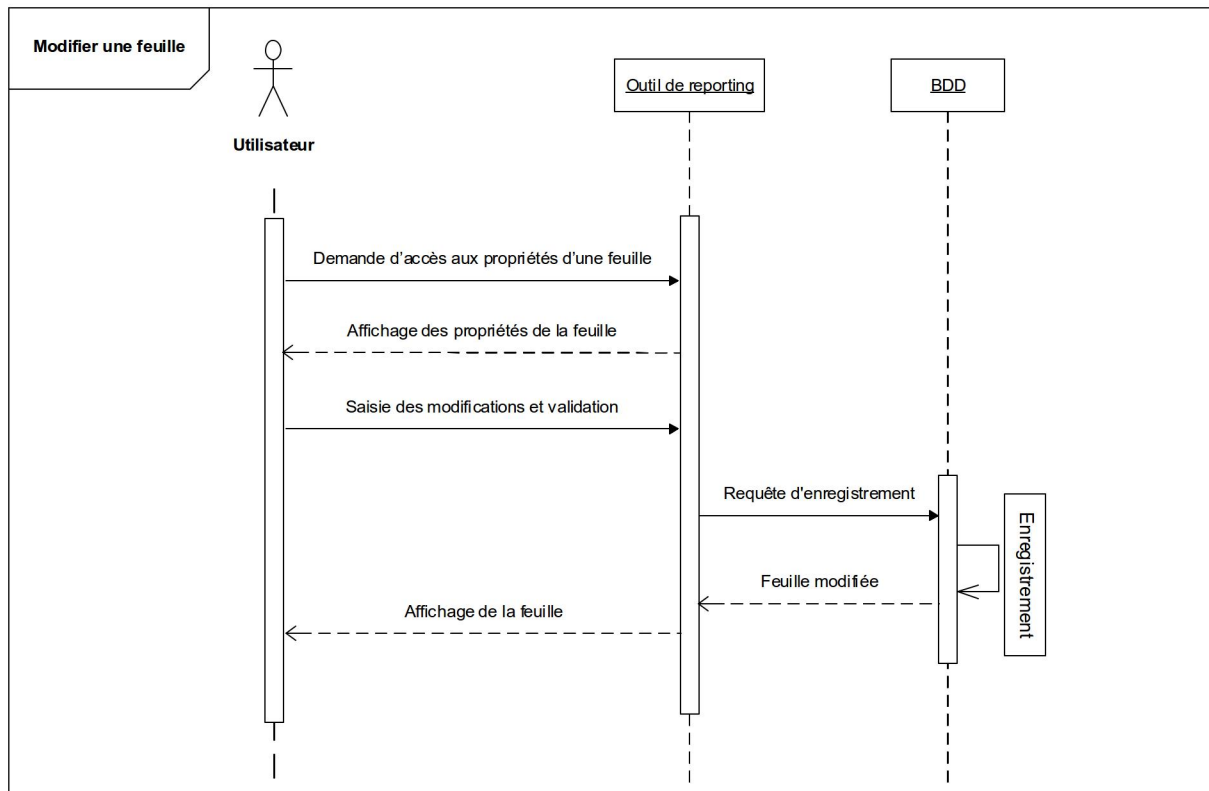


Figure III.12 : Diagramme de séquence du cas d'utilisation « Modifier une feuille ».

### III.5.9 Cas d'utilisation « Gérer un objet »

Une fois que l'utilisateur aura désigné une feuille, celui-ci peut créer de nouveaux objets, et éventuellement supprimer ou modifier ceux qui y sont déjà.

Un objet sur l'outil de reporting peut se présenter sous forme d'un graphique, un tableau, une liste de sélection, un filtre etc...

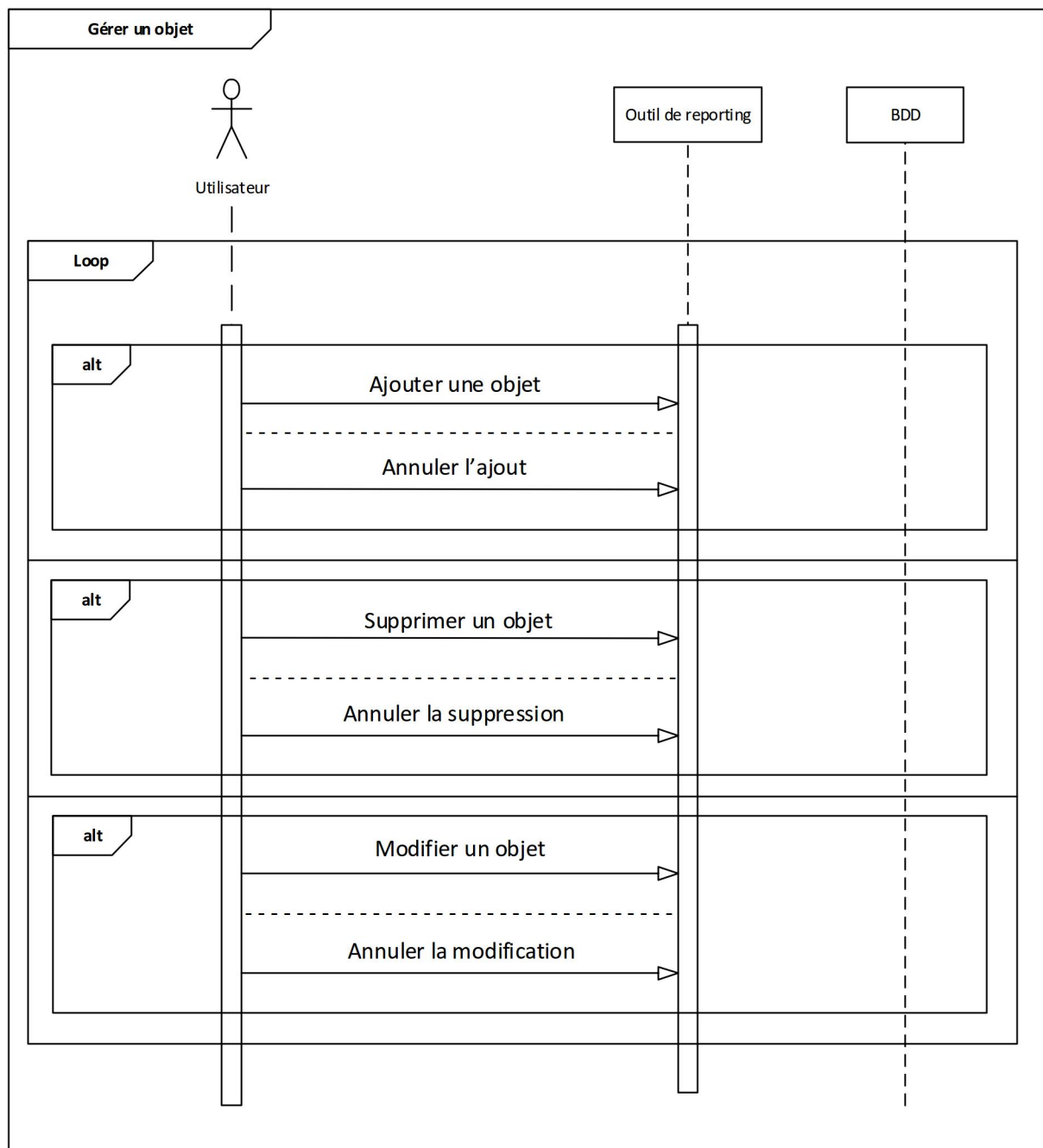


Figure III.13 : Diagramme de séquence du cas d'utilisation « Gestion d'un objet ».



### III.5.10 Cas d'utilisation « Créer un objet »

L'utilisateur exécute la création d'un objet en introduisant les sources de données nécessaire puis en choisissant le type d'objet ainsi qu'une fonction de calcul.

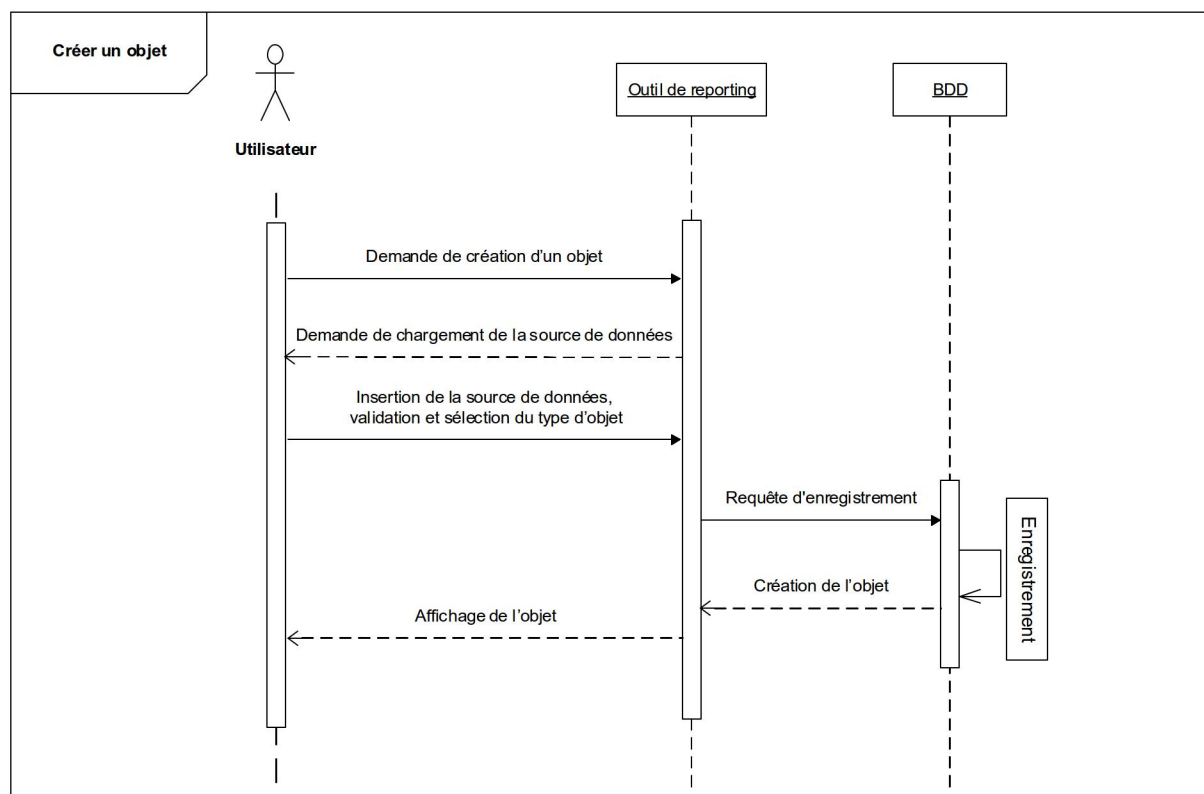


Figure III.14 : Diagramme de séquence du cas d'utilisation « Créer un objet ».

### III.5.11 Cas d'utilisation « Supprimer un objet »

L'utilisateur peut sélectionner un ou plusieurs objets et les supprimer.

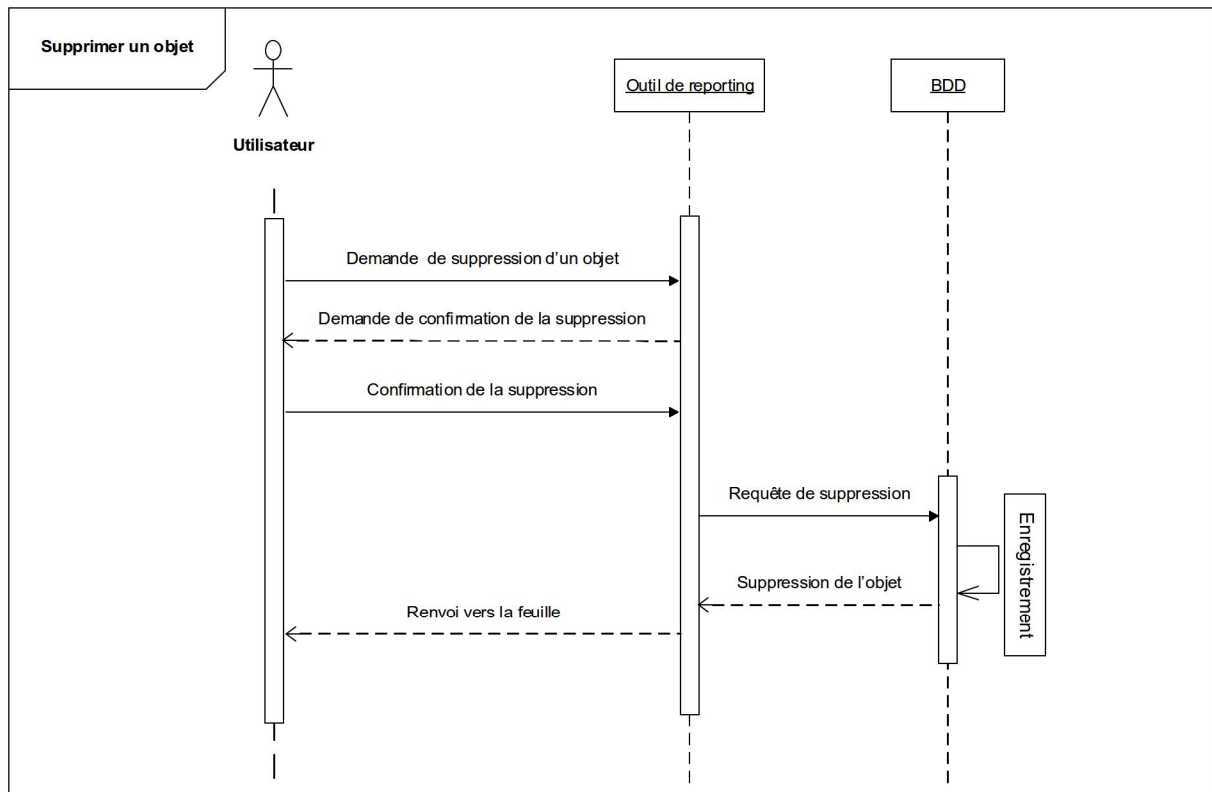


Figure III.15 : Diagramme de séquence du cas d'utilisation « Supprimer un objet ».

### III.5.12 Cas d'utilisation « Modifier un objet »

Outre la suppression, l'utilisateur peut aussi modifier les caractéristiques d'un objet.

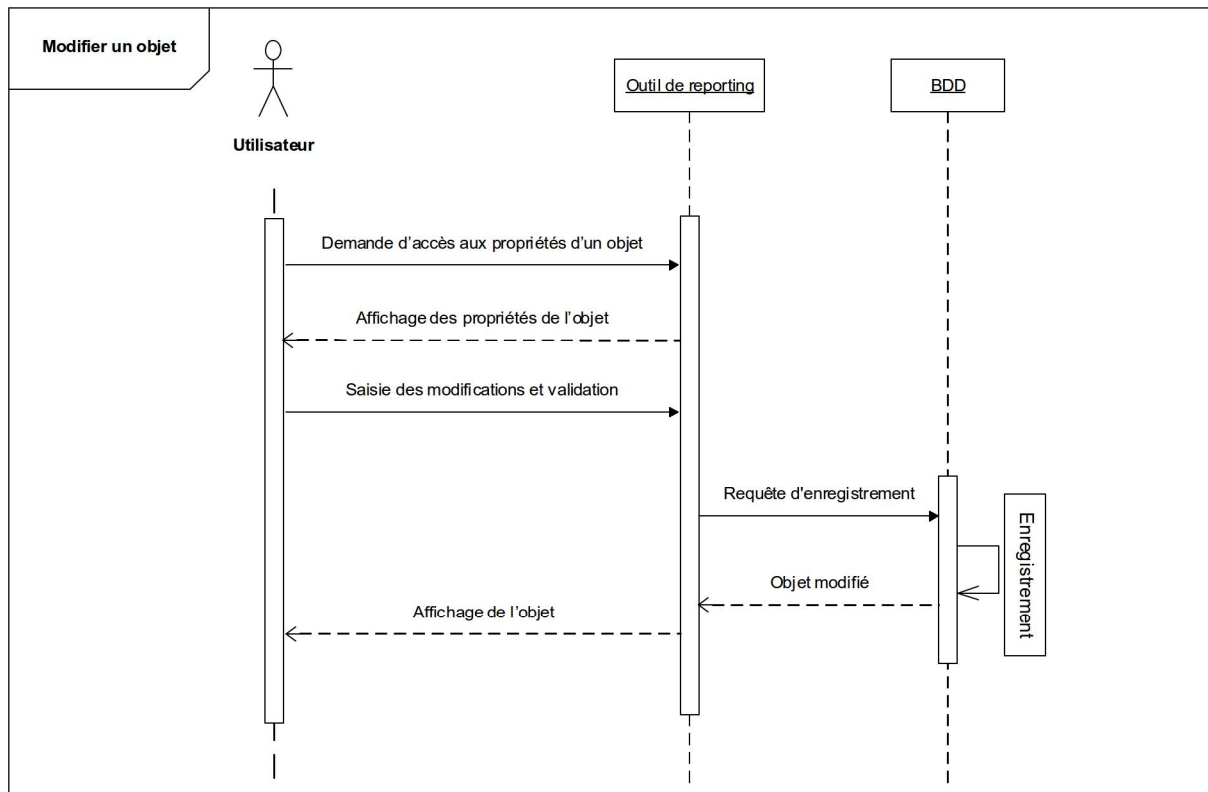


Figure III.16 : Diagramme de séquence du cas d'utilisation « Modifier un objet ».

## III.6 Conclusion

Au terme de ce chapitre, nous avons mis en œuvre une liste des objectifs à atteindre pour les décideurs de l'EPB à travers la définition des besoins, puis nous avons mené une conception détaillée du système à réaliser dans le but de garantir la fiabilité et l'efficacité de la phase de réalisation.

# **CHAPITRE IV**

## **Réalisation**

### **IV.1 Introduction**

Nous consacrons ce chapitre à la partie réalisation en débutant par présenter les composants décisionnels exploités dans le développement de notre de travail, puis expliquer l'élaboration de nos sources de données, et finir par quelques illustrations de notre projet.

### **IV.2 Les outils de travail**

Comme tout projet décisionnel, la mise en place d'une solution de Business intelligence nécessite au préalable la présence d'un ETL pour l'intégration de données. Ensuite nous diffuserons l'ensemble de nos données à travers un outil de reporting. Nos outils se présentent comme suit :

#### **IV.2.1 MySQL Community Edition**

MySQL Community Edition est un SGBD Open Source. Les bases de données de l'EPB sont principalement stockés sous MySQL, d'où l'usage de ce dernier. [22]

#### **IV.2.2 Talend Open Studio for Data Integration 5.6**

Talend Open Studio propose un environnement graphique simple et complet à utiliser, basé sur Eclipse, pour développer, tester, déployer et administrer des projets de gestion de données et d'intégration d'applications. Nos sources de données seront donc intégrés



et consolidé sous Talend afin d'en tirer profit et de les exploiter dans la partie reporting.[16]

### **IV.2.3 QlikView 11**

QlikView nous permettra le développement d'une application personnalisée d'analyse de données sous forme de tableau de bord contenant les indicateurs nécessaires afin de surveiller l'évolution et les performances d'une quelconque activité au sein de l'entreprise.[20]

## **IV.3 Définition des sources de données**

Nous traiterons avec deux différentes bases de données dans notre projet : la base de données « remorquages » qui contient principalement l'ensemble des informations concernant les remorqueurs, les manœuvres, l'équipage..., ainsi que la base de données « escales » qui vient compléter la base de données précédente du fait qu'elle contient des informations sur les navires, les postes etc...Les deux figures suivantes illustrent les schémas de base de données « remorquages » et « escales » :

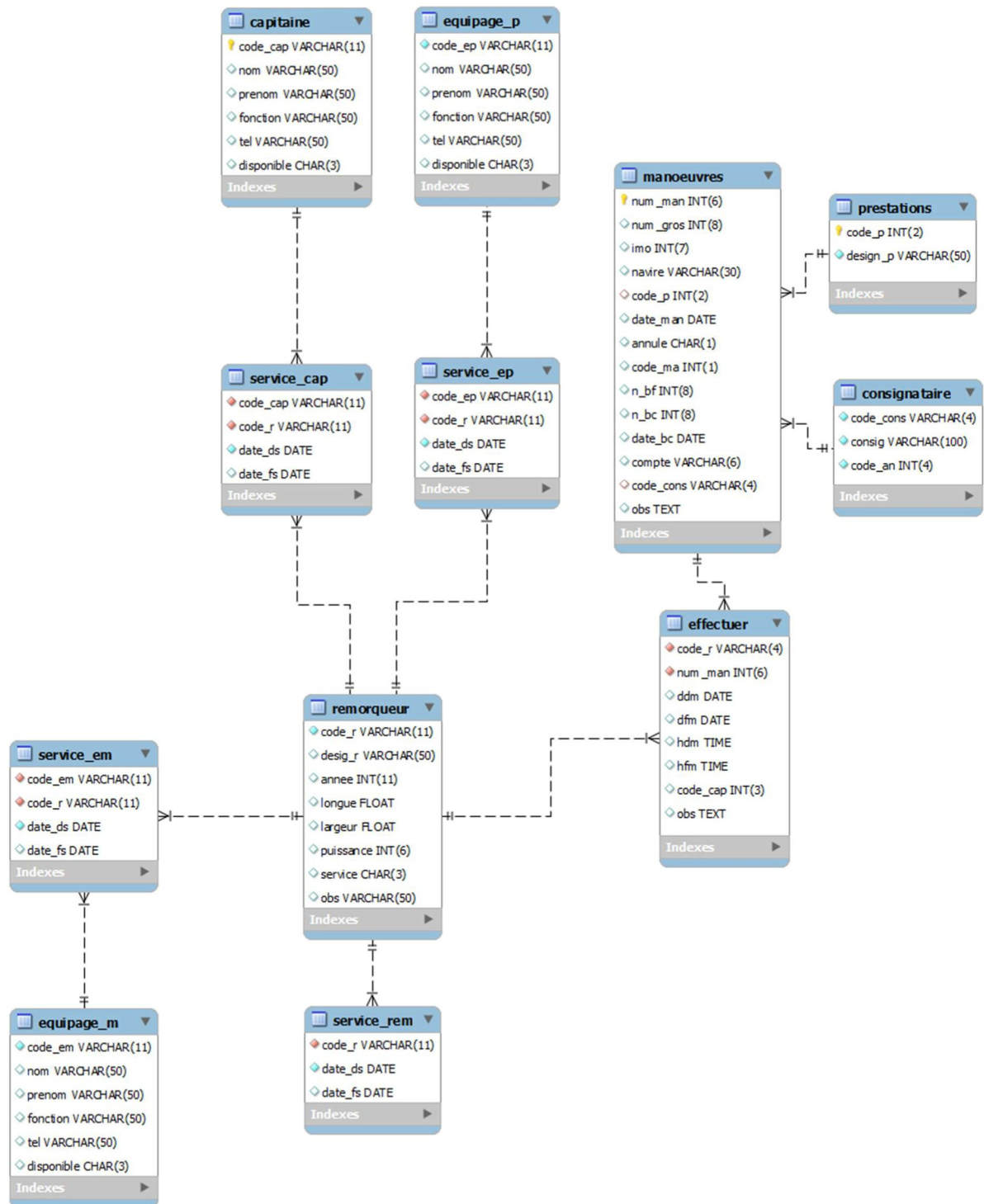


Figure IV.1 : Schéma de base de données « remorquages ».

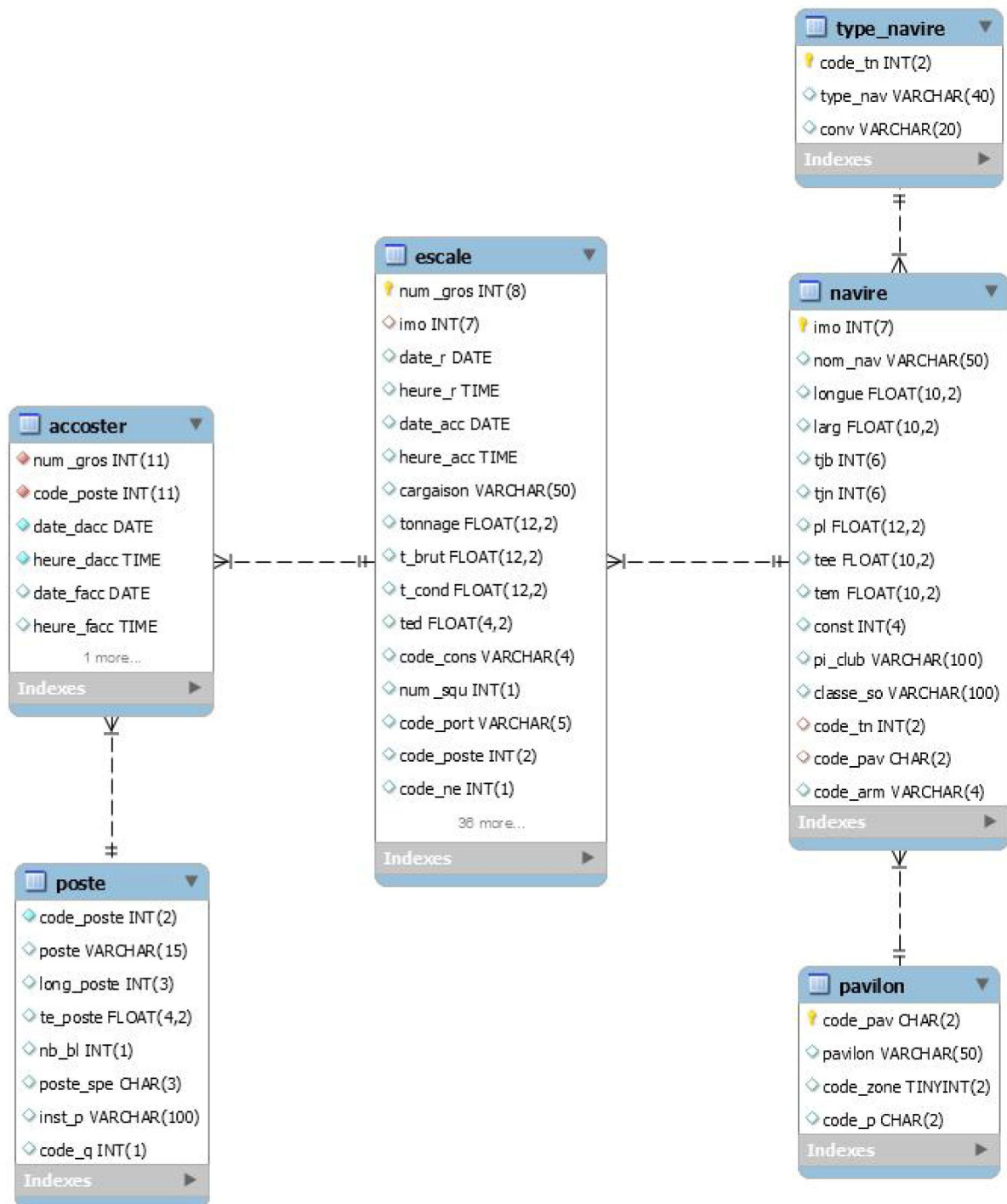


Figure IV.2 : Schéma de base de données « escales ».

Le tableau suivant comporte les informations essentielles extraites dans les deux bases de données pour les besoins de notre projet :

Attribut	Type	Table de l'attribut	Base de données source	Description
<b>code_r</b>	Varchar	remorqueur	remorquages	Identificateur remorqueur
<b>desig_r</b>	Varchar	remorqueur	remorquages	Nom du remorqueur
<b>hdm</b>	Time	effectuer	remorquages	Heure de début de manœuvre
<b>hfm</b>	Time	effectuer	remorquages	Heure de fin de manœuvre
<b>ddm</b>	Date	effectuer	remorquages	Date de début de manœuvre
<b>num_man</b>	INT	manœuvre	remorquages	Numéro de manœuvre
<b>code_p</b>	INT	prestation	remorquages	Identificateur de prestation fournie par le remorqueur
<b>desig_p</b>	Varchar	prestation	remorquages	Nom de la prestation fournie par le remorqueur
<b>num_gros</b>	INT	manœuvre	remorquages	Numéro d'escale
<b>num_gros</b>	INT	escale	escales	Numéro d'escale
<b>poste</b>	Varchar	poste	escales	Nom du poste
<b>type_nav</b>	Varchar	type_navire	escales	Type de navire en question
<b>nom_nav</b>	Varchar	navire	escales	Nom du navire
<b>pavillon</b>	Varchar	pavillon	escales	Pavillon du navire
<b>date_e</b>	Date	escale	escales	Date d'entrée du navire au port
<b>date_s</b>	date	escale	escales	Date de sortie du navire du port

Tableau IV.1 : Attributs des bases de données « remorquages » et « escales ».

## IV.4 Processus d'ETL au niveau de Talend Open Studio

Sachant que les données figurantes dans ces bases de données ne sont pas toutes utilisables dans notre cas (données faisant référence à des factures, ou à des données personnelles qui sont confidentielles), ainsi que des données brutes qui ne sont pas exploitables en l'état, nous avons eu besoin de recourir au processus de l'ETL. Afin d'intégrer les deux bases de données dans une seule source en sortie (sous format CSV



dans notre cas) nous avons créé un Job Talend nommé EPB qui propose une interface graphique simple d'utilisation et qui se présente comme suit :

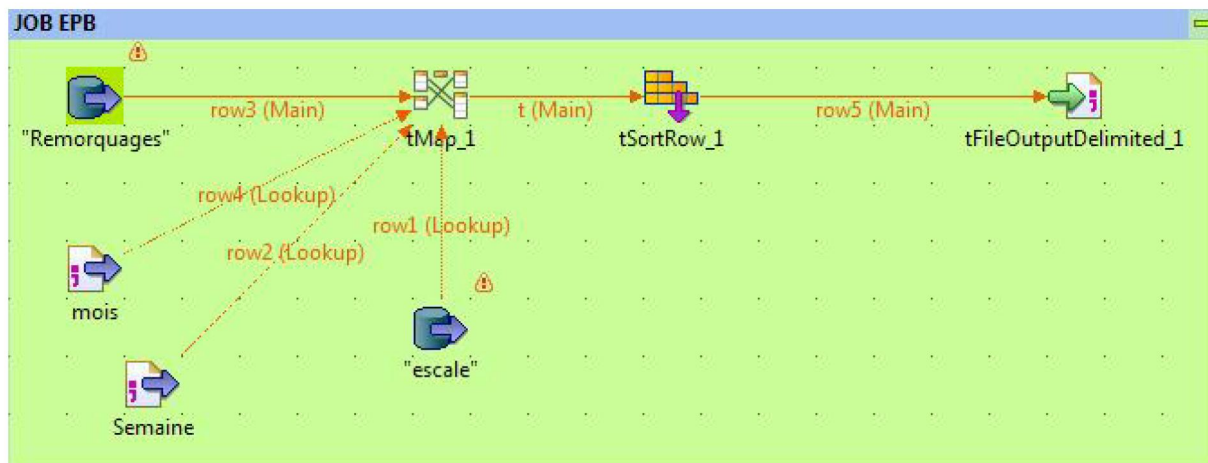


Figure IV.3 : Processus d'ETL sous Talend Open Studio.

La figure suivante est l'étape détaillée du mapping :

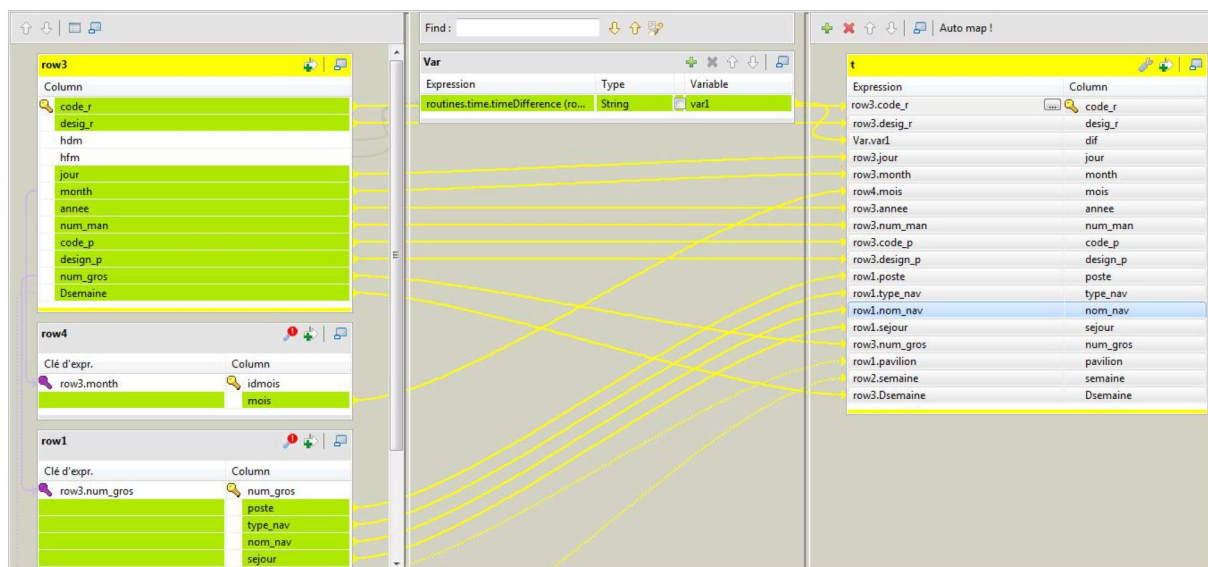


Figure IV.4 : Etape du mapping.

Nous avons eu recours à 5 types de composants Talend dans ce Job :


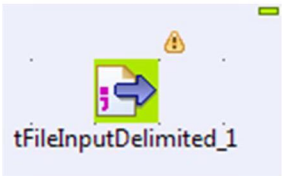

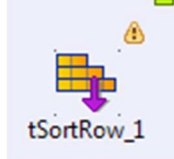
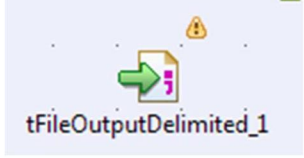
Composant	Nom du composant	Utilité dans notre projet
	tMysqlinput	1.Définir la connexion à la source de données Mysql. 2.Editer des requêtes Sql pour sélectionner et agir sur des attributs précis à extraire.
	tFileInputDelimited	Extraire des données d'un fichier txt, dans notre cas faire correspondre à chaque moi écrit en intégral, son équivalent en Varchar.
	tMap	1. Définir un mapping en extrayant seulement les champs utiles. 2. Définir des jointures entre les 4 sources de données précédentes (dans notre cas Inner Join). 3. Transformer les types et les champs de sortie à l'aide de routines (« time » dans le job).
	tSortRow	Permet les données qui sortent du mapping en choisissant les colonnes à trier et dans quel ordre les trier (ascendant ou descendant).
	tFileOutput	Charger les données en sortie dans un fichier délimité (xlsx, csv...) et choisir son emplacement.

Tableau IV.2 : Les composants utilisés au niveau de L'ETL.

Les deux composants « mois » et « Semaine » nous servent à transformer leurs indices numériques en alphabet (exemple : 1=Janvier).

Le tableau suivant comporte les 2 requêtes essentielles qui interrogent les bases de données et en extraire des données utiles :

Extraction des attributs à partir de la base de données « escales » et calcule des champs séjour à partir des champs date d'entrée et date de sortie	<pre> select e.num_gros,p.poste,t.type_nav,n.nom_nav,DATEDIFF(e. date_s, e.date_e) AS sejour,pa.pavillon from pavillon pa, escale e ,accoster a, poste p, navire n,type_navire t where e.num_gros=a.num_gros and a.code_poste=p.code_poste and e.imo=n.imo and n.code_tn=t.code_tn and n.code_pav=pa.code_pav and e.date_s is not null and e.date_e is not null </pre>
Extraction des attributs à partir de la base de données « remorquages » et extraction du jour du mois et de l'année de la date « ddm » (date début de manœuvre)	<pre> select e.code_r, r.desig_r,e.hdm,e.hfm,DAY(e.ddm) as maDatejour, MONTH(e.ddm) as maDateMois, YEAR(e.ddm) as maDateAnnee,e.num_man,p.code_p,p.design_p,m.num_gros ,DAYOFWEEK(e.ddm) as dsemaine from remorqueur r,effectuer e,manoeuvres m,prestations p where e.num_man=m.num_man and m.code_p=p.code_p and r.code_r=e.code_r </pre>

Tableau IV.3 : Requêtes SQL pour l'extraction de données.

Le processus d'ETL s'achève donc avec la génération du fichier de source de données nommé « projetBI » et qui se présente comme suit :

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
code_r	desig_r	dif	jour	month	mois	annee	num_man	code_p	design_p	poste	type_nav	nom_nav	sejour	num_gros	pavillon	
503 CHELIFF 6	00:30:20	1	11	Novembre	2012	33	1	ENTREE	POSTE 26	RO/RO	TELEGHMA	23	20121218	ALGERIE		
505 CAP CARBON	00:30:20	1	11	Novembre	2012	33	1	ENTREE	POSTE 26	RO/RO	TELEGHMA	23	20121218	ALGERIE		
506 CAP SIGLI	00:30:20	1	11	Novembre	2012	33	1	ENTREE	POSTE 26	RO/RO	TELEGHMA	23	20121218	ALGERIE		
505 CAP CARBON	00:30:08	1	11	Novembre	2012	34	2	SORTIE	POSTE 19	NAVIRE CARI SOPHIA		1	20121209	ILES MARSHALL		
506 CAP SIGLI	00:30:08	1	11	Novembre	2012	34	2	SORTIE	POSTE 19	NAVIRE CARI SOPHIA		1	20121209	ILES MARSHALL		
503 CHELIFF 6	00:30:00	1	11	Novembre	2012	35	1	ENTREE	POSTE 19	CARGO	COSTIS	0	20121213	SAINT-VINCENT ET		
503 CHELIFF 6	00:40:12	1	11	Novembre	2012	36	2	SORTIE	POSTE 01	HUILIER	BARBARICA	8	20121180	ITALIE		
505 CAP CARBON	00:40:12	1	11	Novembre	2012	36	2	SORTIE	POSTE 01	HUILIER	BARBARICA	8	20121180	ITALIE		
506 CAP SIGLI	00:40:12	1	11	Novembre	2012	36	2	SORTIE	POSTE 01	HUILIER	BARBARICA	8	20121180	ITALIE		
503 CHELIFF 6	00:25:00	1	11	Novembre	2012	37	2	SORTIE	POSTE 19	CARGO	COSTIS	0	20121213	SAINT-VINCENT ET		
503 CHELIFF 6	00:35:00	1	11	Novembre	2012	38	2	SORTIE	POSTE 22	CARGO	BF LETICIA	1	20121146	ANTIGUA ET BARBU		
505 CAP CARBON	00:35:00	1	11	Novembre	2012	38	2	SORTIE	POSTE 22	CARGO	BF LETICIA	1	20121146	ANTIGUA ET BARBU		
506 CAP SIGLI	00:35:00	1	11	Novembre	2012	38	2	SORTIE	POSTE 22	CARGO	BF LETICIA	1	20121146	ANTIGUA ET BARBU		
505 CAP CARBON	00:45:00	1	11	Novembre	2012	39	1	ENTREE	POSTE 22	PORTE CONT CONSOUTH		2	20121165	ANTIGUA ET BARBU		
506 CAP SIGLI	00:45:00	1	11	Novembre	2012	39	1	ENTREE	POSTE 22	PORTE CONT CONSOUTH		2	20121165	ANTIGUA ET BARBU		
505 CAP CARBON	00:40:01	1	11	Novembre	2012	40	2	SORTIE	POSTE 17	PORTE CONT SADAN BAYR		2	20121193	TURQUIE		
506 CAP SIGLI	00:40:01	1	11	Novembre	2012	40	2	SORTIE	POSTE 17	PORTE CONT SADAN BAYR		2	20121193	TURQUIE		
505 CAP CARBON	00:55:00	2	11	Novembre	2012	41	2	SORTIE	POSTE 17	PORTE CONT SEA BREEZE		0	20121206	ANTIGUA ET BARBU		
506 CAP SIGLI	00:55:00	2	11	Novembre	2012	41	2	SORTIE	POSTE 17	PORTE CONT SEA BREEZE		0	20121206	ANTIGUA ET BARBU		
505 CAP CARBON	00:25:00	2	11	Novembre	2012	42	2	SORTIE	POSTE 13	CARGO	VELSERDIEP	3	20121167	PAYS-BAS		
506 CAP SIGLI	00:25:00	2	11	Novembre	2012	42	2	SORTIE	POSTE 13	CARGO	VELSERDIEP	3	20121167	PAYS-BAS		
503 CHELIFF 6	00:55:08	2	11	Novembre	2012	43	2	SORTIE	POSTE 24	CARGO	CONTI PYRIT	3	20121166	PANAMA		

Figure IV.5 : Contenu du fichier de source de données « projetBI ».



La liste des attributs la source de données en sortie se résume comme suit :

Attribut	Type	Description
<b>desig_r</b>	Varchar	Nom du remorqueur.
<b>dif</b>	Varchar	Durée en heures, minutes et secondes pris durant la manœuvre (calculé à l'aide de la routine time, et à partir de hdm et hfm) : utilisé pour le calcul de la durée moyenne et du nombre et du temps en somme pris durant les manœuvres.
<b>jour</b>	INT	Jour du déroulement de manœuvre (extrait avec une requête SQL à partir de ddm) : nous sert à trier notre fichier et comme filtre dans nos tableaux de bord.
<b>month</b>	INT	Mois du déroulement de manœuvre (extrait avec une requête SQL à partir de ddm) : sert à trier notre fichier csv.
<b>mois</b>	Varchar	Mois du déroulement de manœuvre (extrait avec une requête SQL à partir de ddm) : nous sert comme filtre dans nos tableaux de bord
<b>annee</b>	INT	Année du déroulement de manœuvre (extrait avec une requête SQL à partir de ddm) : nous sert à trier notre fichier et comme filtre dans nos tableaux de bord.
<b>num_man</b>	INT	Numéro de manœuvre : utilisé pour compter le nombre de manœuvre.
<b>code_p</b>	INT	Code de prestation fournie par les remorqueurs.
<b>design_p</b>	Varchar	Nom de la prestation fournie par le remorqueur : utilisé pour avoir des informations sur les prestations.
<b>poste</b>	Varchar	Poste en question concerné par la manœuvre : pour avoir des informations sur les postes les plus sollicités par exemple.
<b>type_nav</b>	Varchar	Type de navire concerné par la manœuvre.
<b>nom_nav</b>	Varchar	Nom du navire sur lequel se déroule la manœuvre.
<b>sejour</b>	INT	Séjour du navire au port calculé à partir de la différence entre date_e et date_s
<b>num_gros</b>	Varchar	Identificateur d'escale.
<b>pavilon</b>	INT	Pavillon du navire.
<b>Dsemaine</b>	INT	Indice du jour de la semaine
<b>Semaine</b>	Varchar	Jour de la semaine en lettre (ex : lundi)

Tableau IV.3 : Requêtes SQL pour l'extraction de données.



## IV.5 Création du tableau de bord

Notre tableau de bord se divise en quatre feuilles, celles-ci comportent des informations sur les manœuvres de remorquage par rapport aux remorqueurs, aux types de navires, aux postes et aux prestations.

Le tableau de bord se compose principalement de :

### IV.5.1 Graphiques

À partir du fichier « .csv » que nous avons généré grâce à l'ETL, il nous est possible de créer des graphiques, ces derniers comportent principalement des informations sur les manœuvres de remorquages : le nombre de manœuvres et le taux d'exploitation des remorqueurs, la durée moyenne de remorquage.

Voici quelques graphiques générés au niveau du tableau de bord :

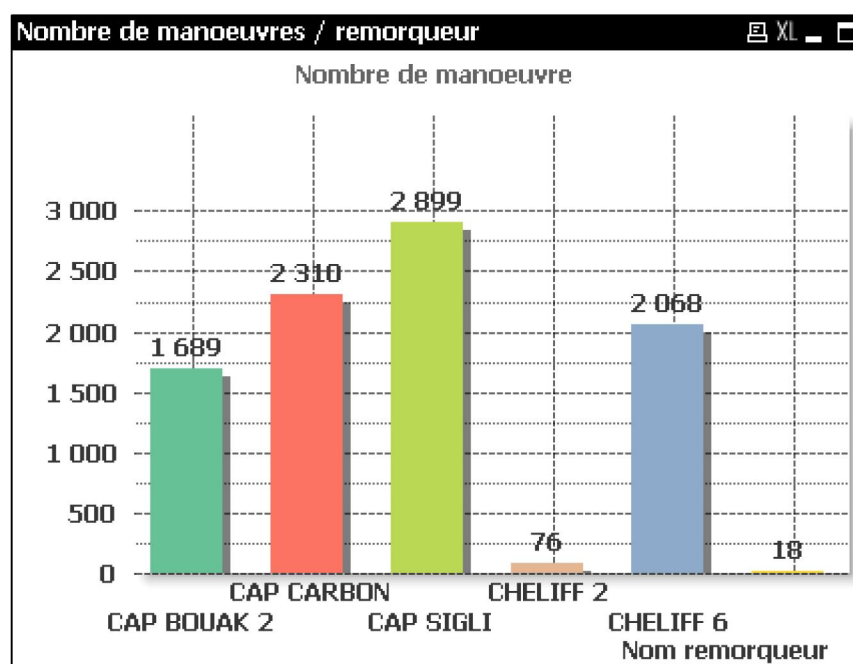


Figure IV.6 : Nombre totale de manœuvres des remorqueurs.

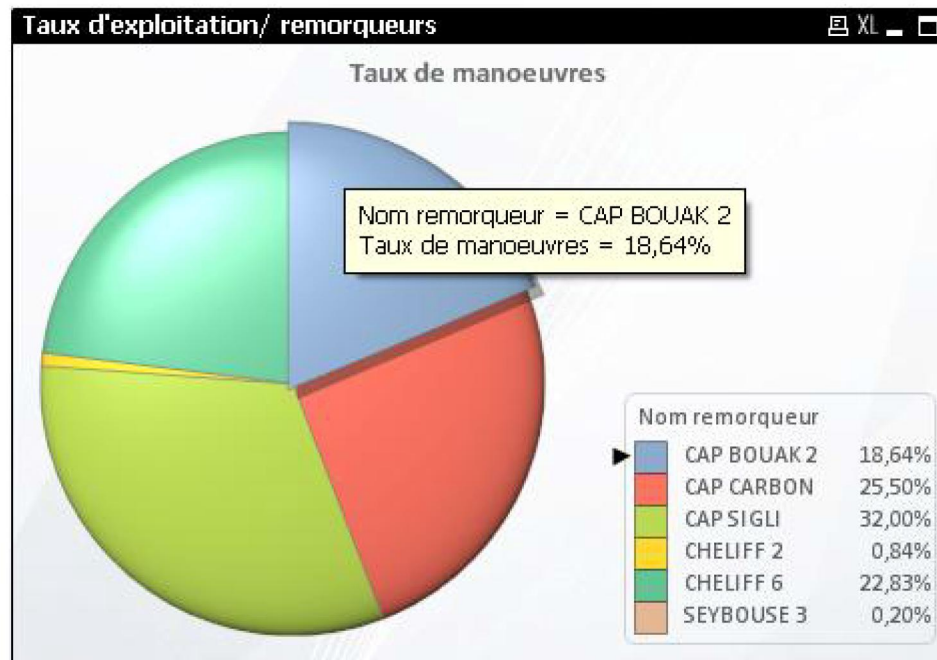


Figure IV.7 : Taux d'exploitation de chaque remorqueur.

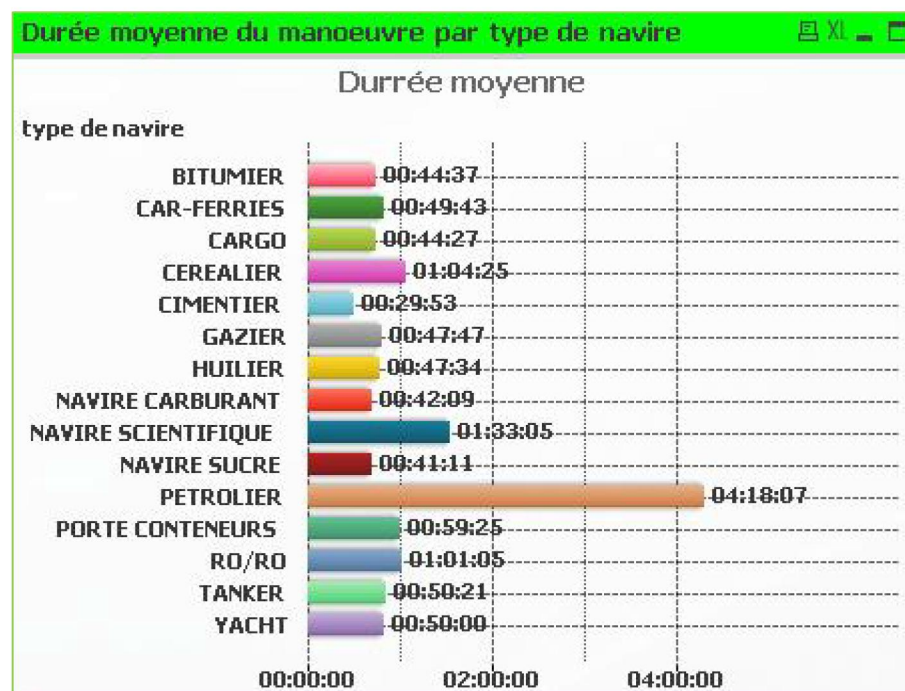


Figure IV.8 : Durées moyennes de manoeuvres par type de navire.

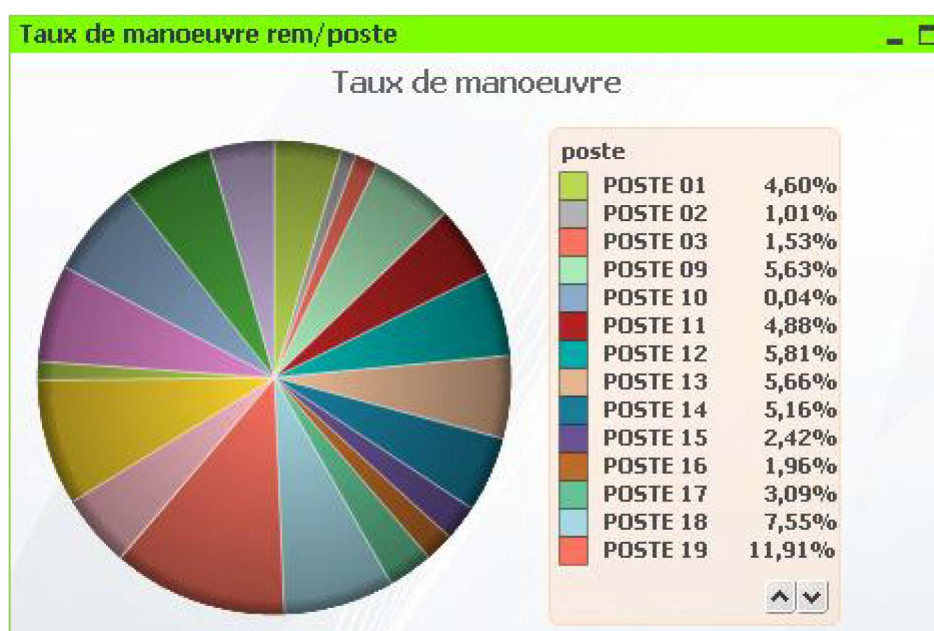


Figure IV.9 : Taux de manoeuvres de remorquage par postes.

## IV.5.2 Filtres

Ils se présentent sous forme de listes de sélection pour réduire les informations du tableau de bord à des conditions précises : sélection d'une date, d'un nom de remorqueur pour spécifier nos analyses sur ce dernier, d'un pays pour afficher ses propres navires etc...

Voici quelques filtres essentiels dans l'analyse de nos données :

Filtre jour													
1	2	3	4	5	6	7	8	9	10	11	12	13	14
15	16	17	18	19	20	21	22	23	24	25	26	27	28
29	30	31											

Filtre mois			
janvier	fevrier	Mars	Avril
Mai	Juin	Juillet	Aout
Septembre	Octobre	Novembre	Decembre

Filtre années		
2012	2013	2014

Figure IV.10 : Les filtres « jour », « mois » et « années ».

Pays   nombre navires	
ALBANIE	1 ▲
ALGERIE	14
ALLEMAGNE	20
ANTIGUA ET...	89
ANTILLES N...	3
BAHAMAS	35
BAHREIN	1
BARBADE	2
BELGIQUE	1
BELIZE	13
BERMUDES	1
CAMBODGE	3 ▼

Figure IV.11 : Filtre des navires ayant séjourné au port.

Poste   nombre manoeuvres	
POSTE 01	213 ▲
POSTE 02	47
POSTE 03	71
POSTE 09	261
POSTE 10	2
POSTE 11	226
POSTE 12	269
POSTE 13	262
POSTE 14	239
POSTE 15	112 ▼

Figure IV.12 : Filtre des manoeuvres par poste.

C'est ainsi que l'ensemble des graphiques et filtres réalisés forme l'ensemble du tableau de bord :

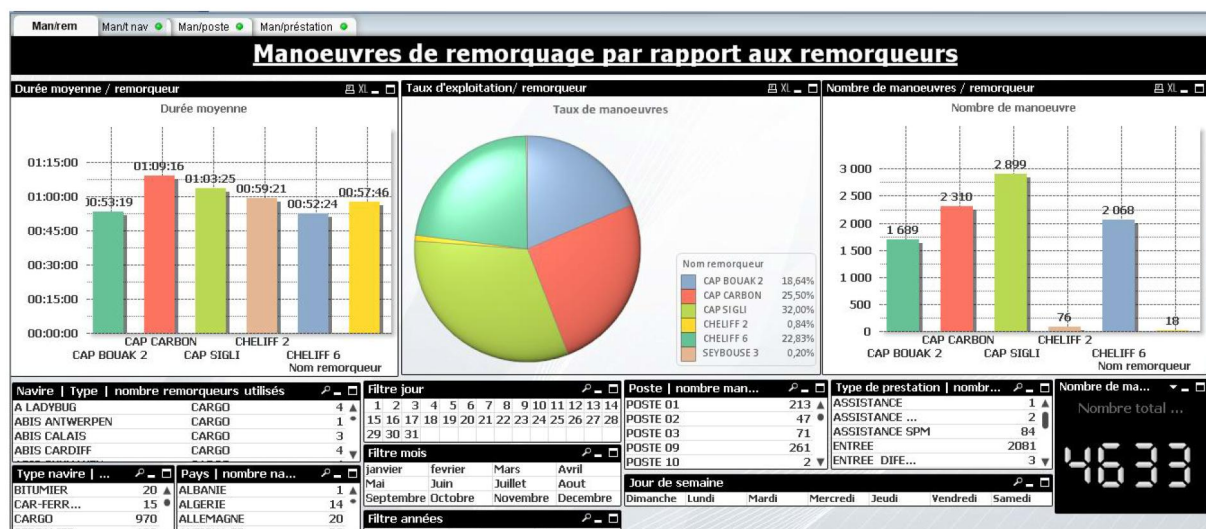


Figure IV.13 : Tableau de bord.



## **IV.6 Conclusion**

Au cours de ce chapitre, nous avons décrit la phase de réalisation à travers la définition de nos outils de travail, du processus d'extraction et de chargement de notre source de données, puis de la création de notre tableau de bord et nous avons illustré notre projet avec différentes captures d'écran de ce dernier.

# Conclusion générale

La Business intelligence est en passe de devenir une arme puissante de contrôle et d'analyse d'information. Beaucoup de spécialistes de la BI prédisent des changements importants dans ce domaine afin de répandre son usage dans toutes les entreprises.

Le travail que nous avons accompli a pour principal objectif de répondre au besoin de la direction de remorquage au niveau de l'Entreprise Portuaire de Béjaïa, qui est de leur offrir une solution à base d'extraction de données, de reporting et d'analyse pour orienter leurs activités quotidiennes.

Ce projet a commencé en premier lieu par une présentation de la Business intelligence afin de nous situer dans ce domaine, nous avons enchainé en effectuant une comparaison globale entre les différentes solutions open source de Business intelligence qui s'offrent à nous, puis par la partie conception comportant une étude préliminaire, l'analyse du besoin, ainsi que la modélisation avec UML. Une fois les fonctionnalités de notre système recensées, le projet se termine par la partie réalisation qui comporte la présentation des outils de travail, les schémas de base de données utilisés ainsi que des explications sur le processus d'ETL produisant la source de données et les différentes illustrations de notre tableau de bord.

Cependant, notre travail reste améliorable, du moment qu'il peut être déployé sur le web et assurer l'accès au tableau de bord à partir de toutes les entités de l'entreprise, et étendre les informations de ce dernier auprès de l'ensemble des métiers portuaires.

La réalisation de ce projet a été bénéfique et enrichissante pour nous dans le sens où il nous a permis d'approfondir et d'acquérir de nouvelles connaissances dans un domaine qui demeure encore inexploré par beaucoup d'informaticiens mais qui offre énormément de services, d'où son utilité à l'avenir.

## ANNEXE A

# Tutoriaux de base de Talend et QlikView

Cette annexe fournit des tutoriaux de base sur la création d'un job au niveau de Talend Open Studio, ainsi que la génération d'objets et de tableaux de bord sur QlikView.

### A.1 Création d'un job sous Talend Open Studio

Un job sous talend consiste à créer un projet d'intégration de données, sa réalisation passe par la connexion à un serveur de base de données (MySQL dans notre cas), sélectionner les bases de données concernées, puis et se servir des composants d'intégration de notre ETL pour manipuler le job.

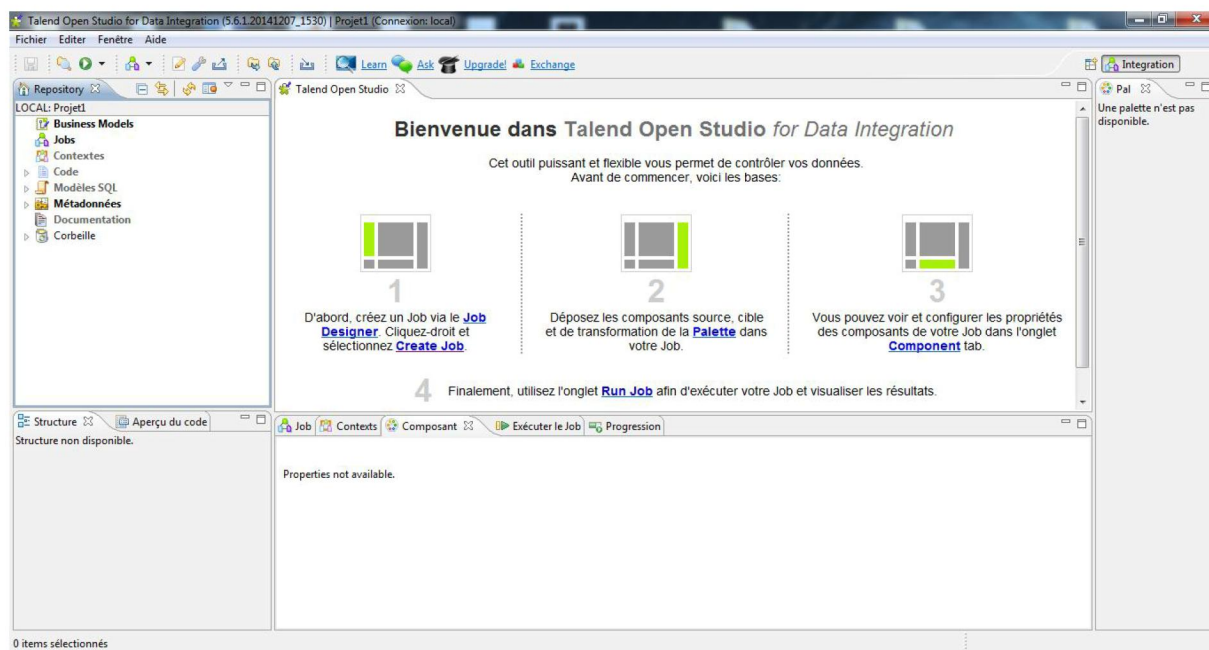


Figure A.1 : Page de démarrage de Talend Open Studio.

Nous allons cliquer sur « Create Job » en gras ci-dessus pour créer un nouveau job, puis nous lui attribuons un nom :

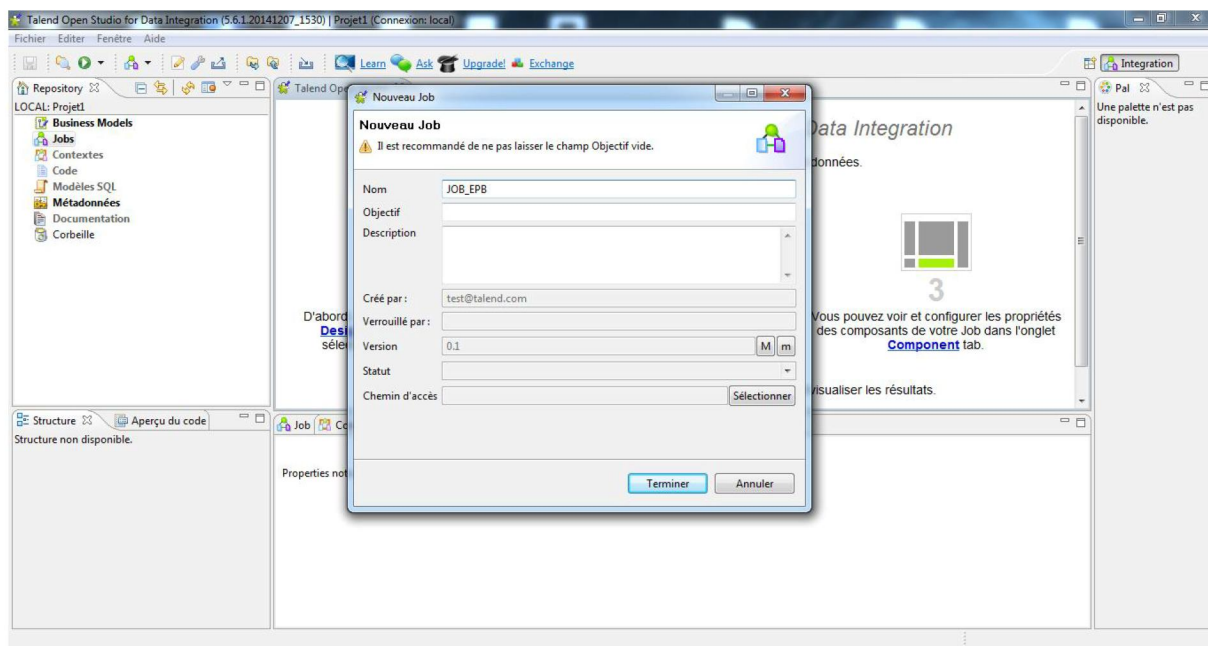


Figure A.2 : Création d'un nouveau job.

Nous voici maintenant à l'intérieur du Job créé (la figure suivante), l'espace centré en blanc est l'environnement de travail pour l'intégration de données, à droite il y a la palette des composants à glisser dans notre job, et à gauche nous avons les informations de notre projet tel que les jobs créés ou les sources de données importées.

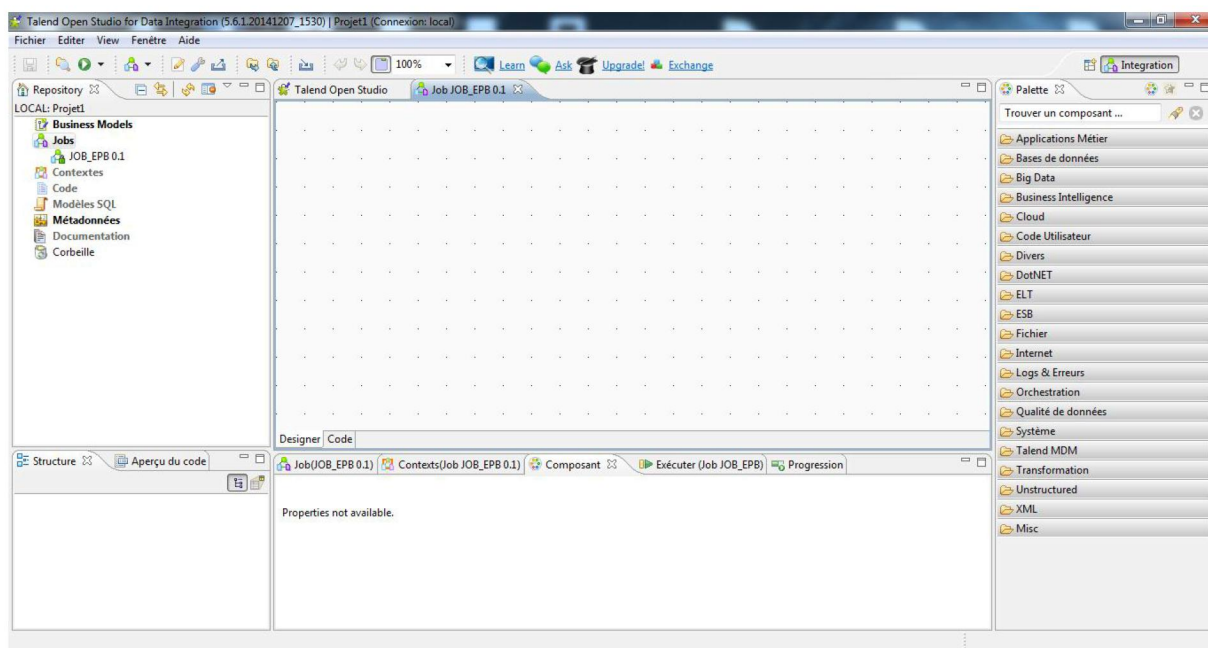


Figure A.3 : Structure de l'environnement de travail sur Talend Open Studio.



Dans l'exemple qui suit, nous allons importer la base de données « remorquage », il suffit de nous connecter au serveur de base de données MySQL puis la glisser dans l'interface du job. Comme l'indique la figure suivante, Il est possible d'insérer une requête SQL pour extraire les informations voulues à partir de la base de données.

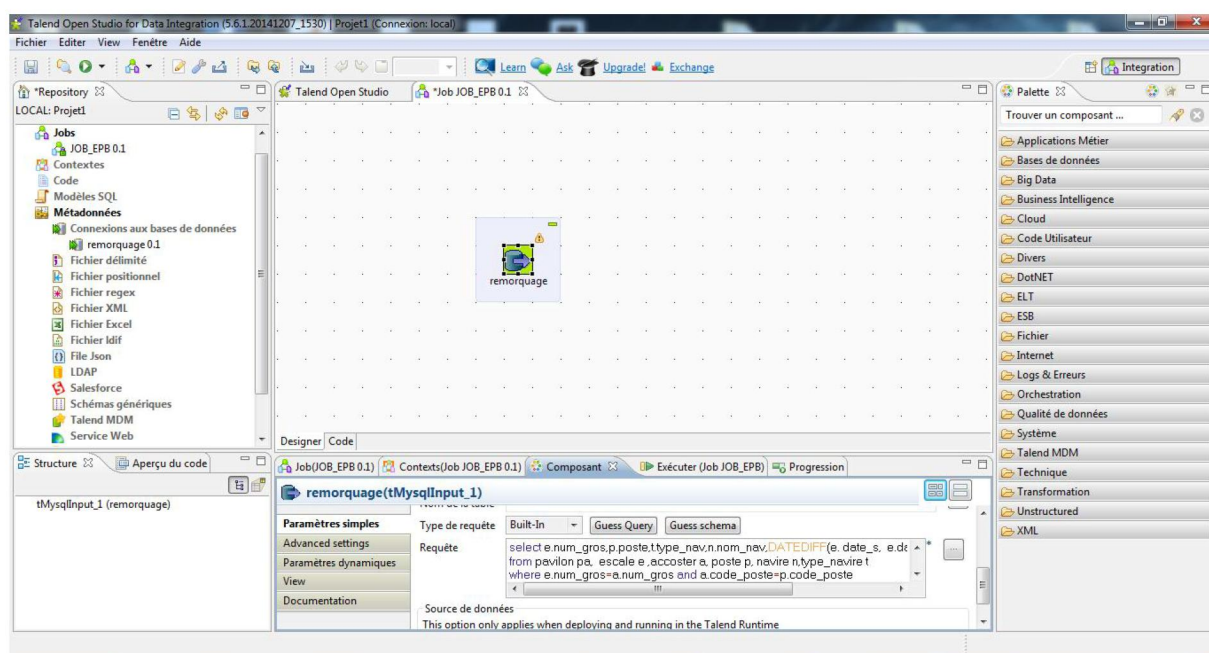


Figure A.4 : Traitement d'un job.

## A.2 Création d'un tableau de bord sous QlikView

Un tableau de bord sous QlikView se compose d'une ou plusieurs feuilles de travail que nous pouvons créer à volonté, sa réalisation passe par l'ouverture d'un nouveau projet, l'importation d'une source de données puis la création des objets qui illustrent nos données en forme de graphiques, liste de sélections, objets textes etc...

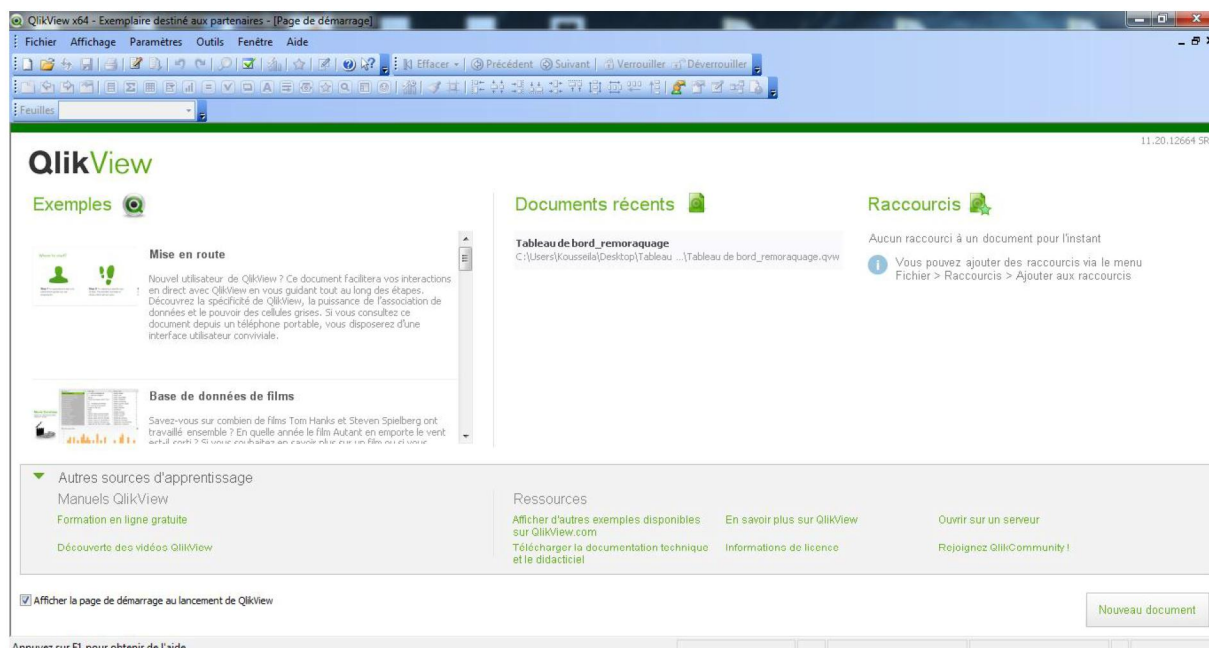


Figure A.5 : Page de démarrage de QlikView.

Nous allons créer un nouveau projet à partir de la page de démarrage, il suffit d'aller vers « fichier » puis « nouveau ». Un nouveau projet doté d'une feuille vierge sera créé (figure A.6)

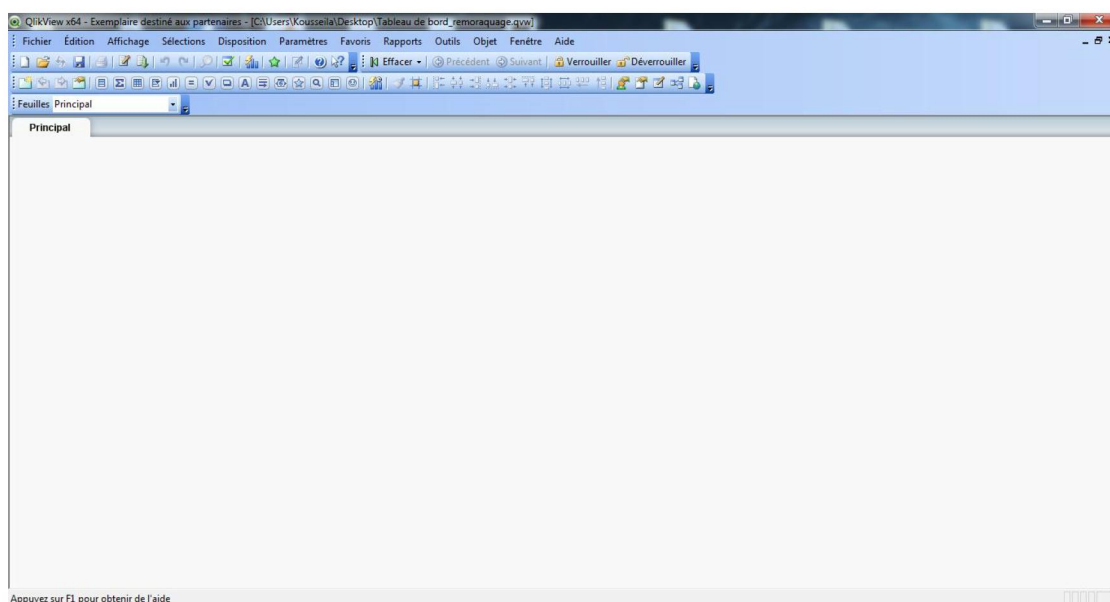


Figure A.6 : Feuille principale du projet créée sur QlikView.

Avant de générer de nouveaux objets, il est primordial d'insérer une ou plusieurs sources de données à notre projet, celles-ci peuvent être sous plusieurs formats et qui se résument dans le tableau suivant :

Type de fichier	Format de fichier
Fichiers délimités	csv, txt, tab, qvo, mem, skv, prn, log
Fichiers Excel	xls, xlw, xlsx, xlsxm
Fichiers HTML	html, htm, php
Fichiers DIF	dif
Fichiers FIX	fix, dat
Fichiers de données QlikView	qvd
Fichiers QlikView Exchange	qvx
Fichiers XML	xml

Tableau A.1 : Type de fichiers tables pouvant être importés sur QlikView.

Nous pouvons insérer une source de données en allant vers fichier/Editer le script, et sélectionner un fichier de table.

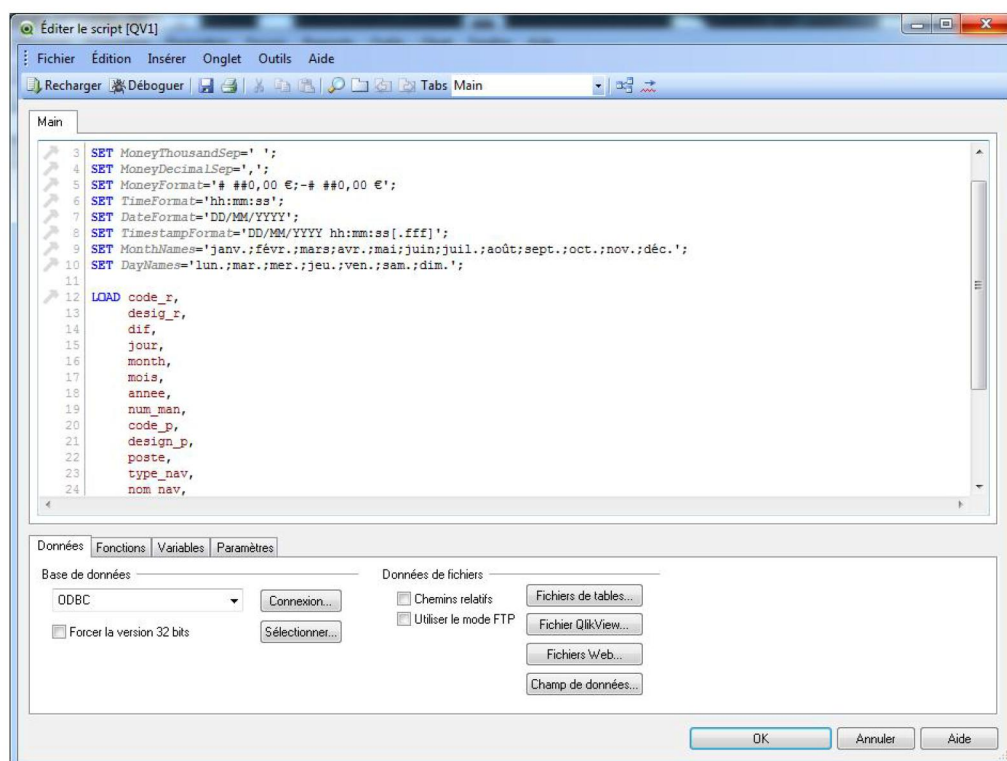


Figure A.7 : Editeur de sources de données.

Une fois que nous avons importé la source de données, il suffit de faire un clic droit de la souris sur la feuille principale et sélectionner n'importe quel objet pour le créer. Nous allons faire une illustration avec la création d'un graphique.

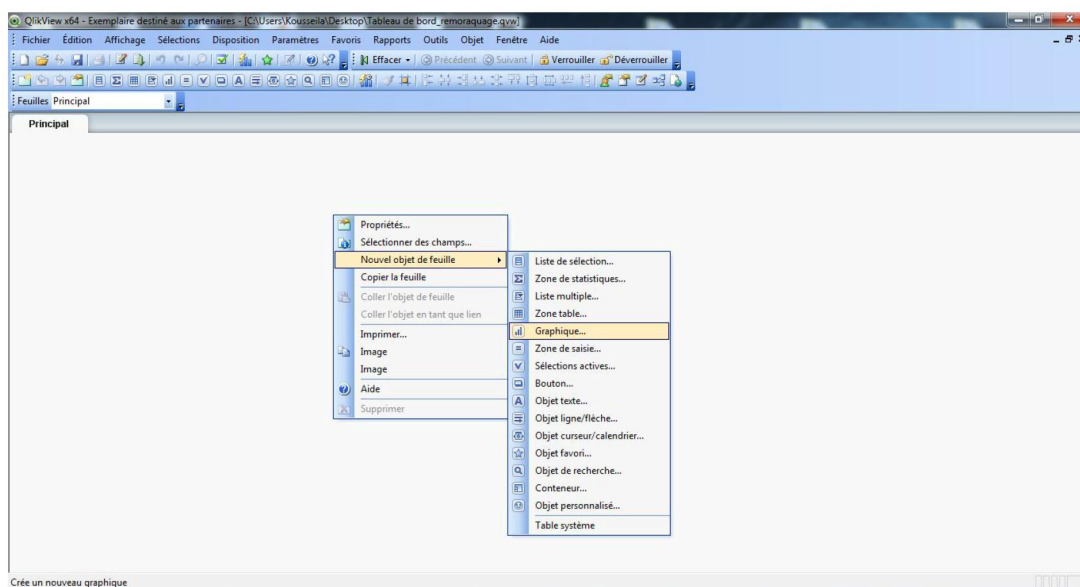


Figure A.8 : Création d'un nouvel objet.



La première étape est d'attribuer des caractéristiques à notre graphique, tel que le nom du graphique, le type de graphique à afficher, la police de texte etc...

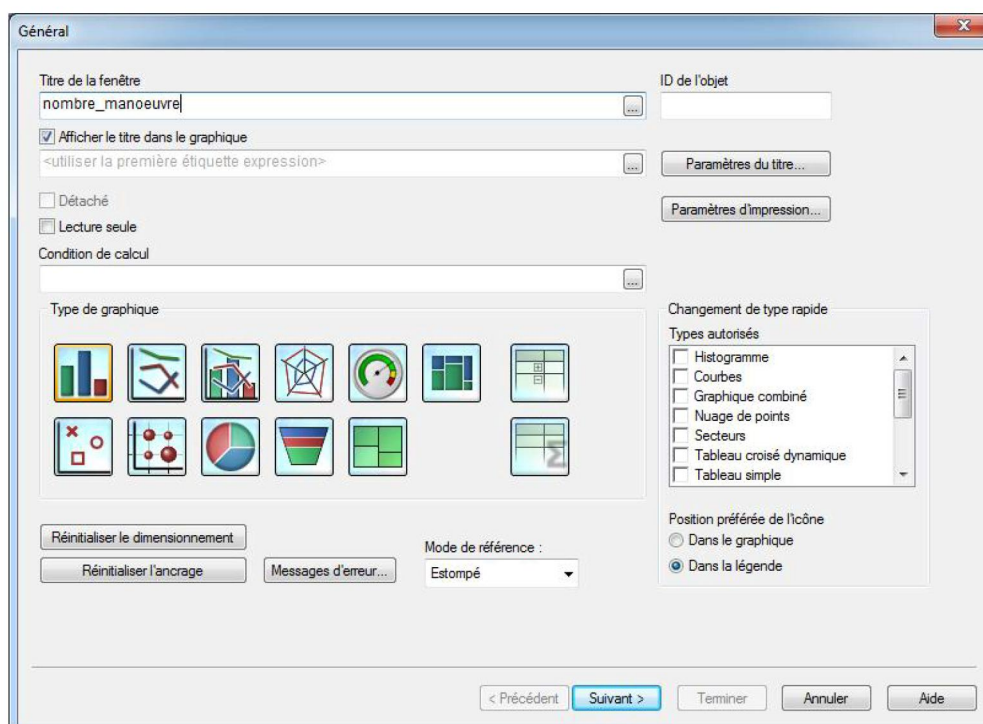


Figure A.9 : Les paramètres d'un graphique.

Nous allons par la suite choisir une dimension parmi celles qui sont disponibles dans la source de données :

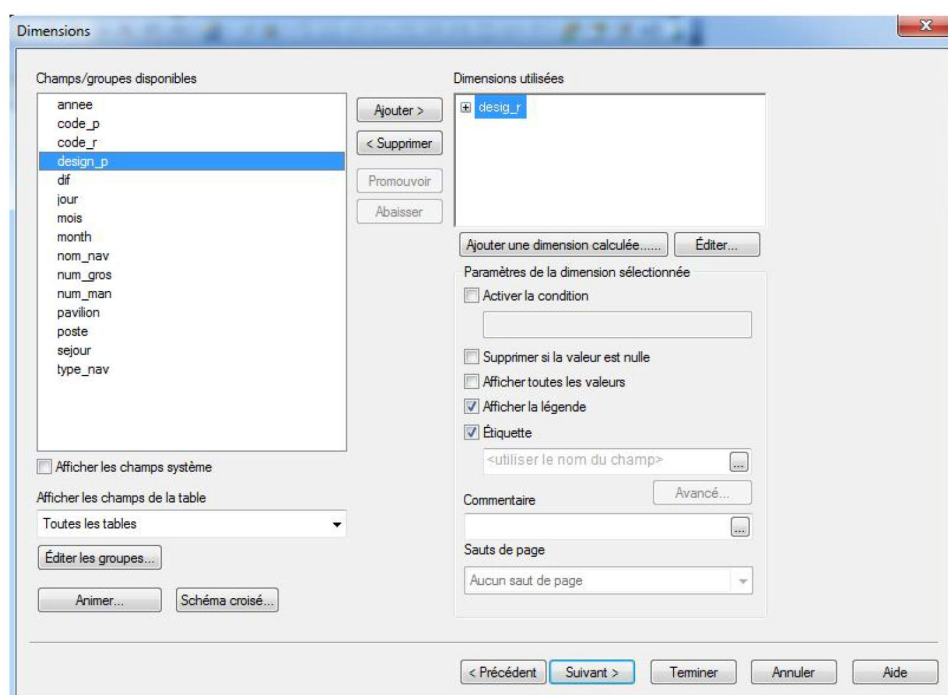


Figure A.10 : Sélection d'une dimension.

Et entrer une fonction de calcul pour l’affichage de notre graphique :

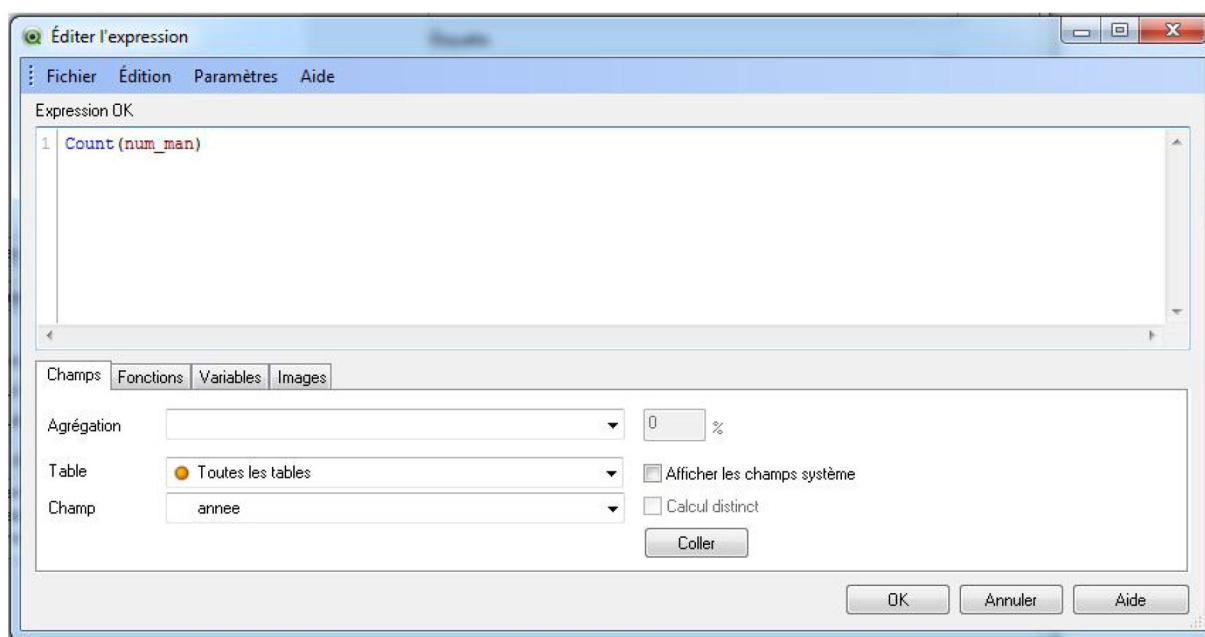


Figure A.11 : Saisie de la fonction de calcul.

Il suffit d’enregistrer les manipulations et nous aurons le graphique affiché dans le tableau de bord.

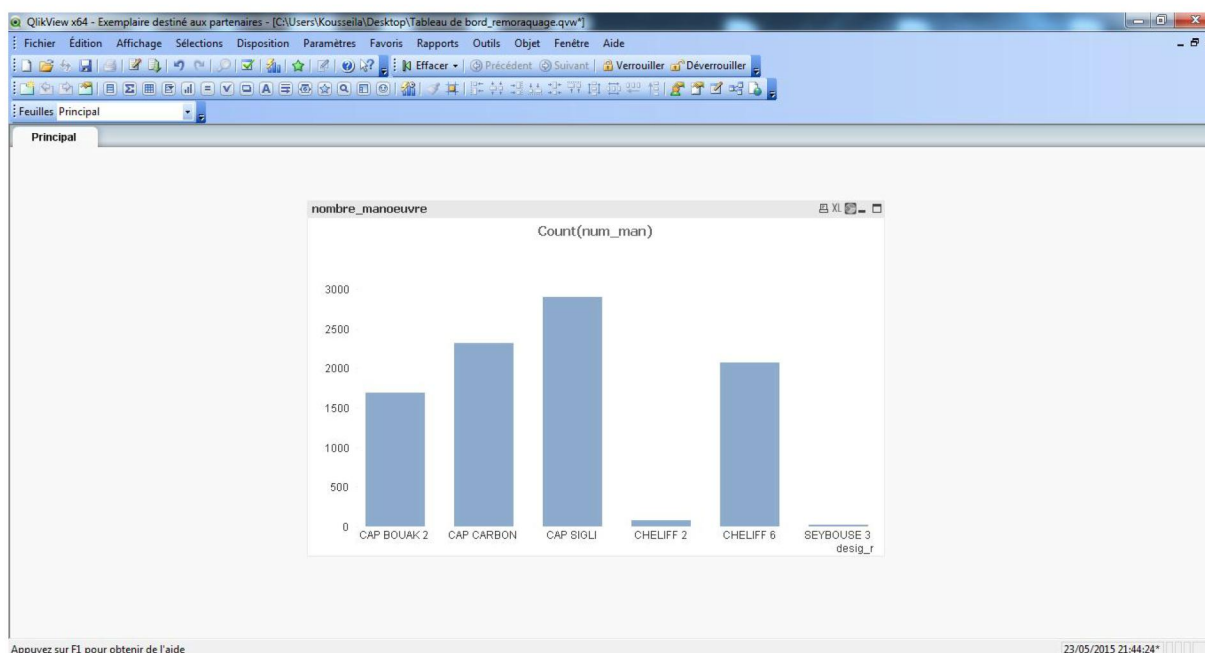


Figure A.12 : Chargement d’un graphique.

Une fois que le projet sera enregistré, un fichier au format « qvw » sera généré qui servira d’exécutable à l’application d’analyse de données sur le tableau de bord.

# **ANNEXE B**

## **Terminologie des diagrammes UML utilisés**

Cette annexe fournit quelques définitions et une liste des différentes figures employées pour la construction des diagrammes UML utilisées lors de la conception.

### **B.1 Définition d'UML**

UML (Unified Modeling Language) se présente comme un langage de modélisation graphique et textuelle orienté objet. Il est destiné à comprendre et décrire des besoins, spécifier et documenter des systèmes, esquisser des architectures logicielles et concevoir des solutions.

En 1994, plus de 50 modèles de représentation orientée objet existaient, tous se ressemblant de plus en plus et utilisant des notations différentes. Par conséquent, les grandes industries avaient besoin de normes et standards avec lesquelles travailler. UML est hérité des trois méthodes qui ont le plus influencées la modélisation objet : OMT (Object Modeling Technic), OOSE (Object Oriented Software Engineering) et Booch, alors que ses principaux auteurs sont GradyBooch, Ivar Jacobson et Jim Rumbaugh.

Depuis 1997, UML est devenu un standard défini par l'Object Management Group (OMG). Depuis Août 2011, cette organisation a diffusé la version 2.4.1.

UML unifie les notations nécessaires aux différentes activités d'un processus de développement et offre, par ce biais le moyen d'établir le suivi des décisions prises, depuis l'expression des besoins jusqu'au codage. [21]

## **B.2 Utilité d'UML**

L'utilité d'UML réside dans :

- Sa normalisation : contrairement à d'autres langages de modélisation, UML est normalisé, c'est à dire pas de profusions de notations qui peuvent être préjudiciable à long terme.
- La possibilité d'utiliser le même atelier de génie logiciel depuis l'expression des besoins jusqu'à la livraison.
- L'utilisation des concepts objets pour enrichir la démarche de conception de systèmes d'information [21].

UML sert entre autre à donner une méthode à suivre pour l'élaboration d'un cahier des charges nécessaires au bon développement d'un logiciel orienté objet. Pour cela, il nous met à disposition différents éléments représentables, par exemple :

- L'activité d'un objet/logiciel.
- Les acteurs.
- Les processus.
- Le schéma de base de données.
- Les composants logiciels [21].

## **B.3 Notions générales sur les diagrammes utilisés**

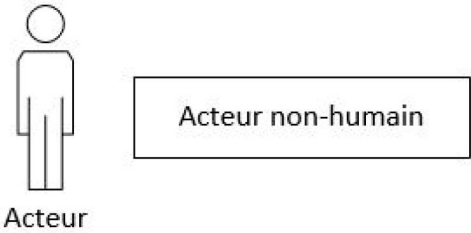

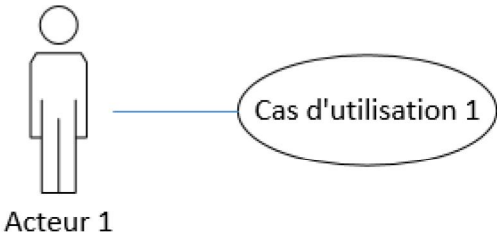
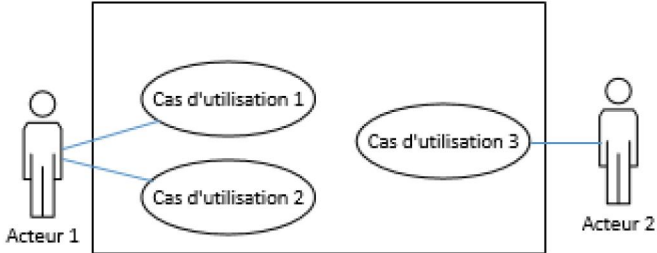
Nous définissons les diagrammes UML utilisés dans la partie « conception » comme suit :

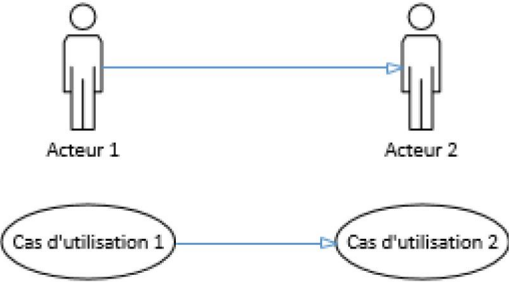
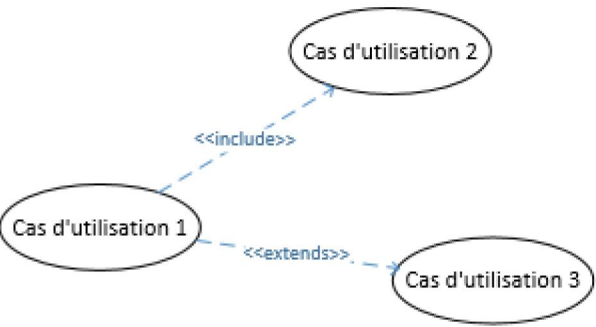
### **B.3.1 Diagramme de cas d'utilisation**

Le diagramme de cas d'utilisation montre les interactions fonctionnelles entre les acteurs (intervenants extérieurs au système) et le système à l'étude. Plus précisément, un cas d'utilisation décrit une séquence d'actions réalisées par le système qui produit un résultat observable pour un acteur. [21]

Le tableau suivant représente la syntaxe graphique présente le diagramme de cas d'utilisation :



Figure	Signification
 <p>Acteur</p>	<p><b>Acteur</b> : rôle joué par une entité externe (utilisateur ou dispositif matériel) qui interagit avec le système étudié.</p>
	<p><b>Cas d'utilisation</b> : représente un ensemble de séquences d'événements qui permettent à un acteur de compléter un processus en utilisant le système.</p>
 <p>Acteur 1</p>	<p><b>Ligne de communication</b> : ligne représentant une interaction entre un acteur et un cas d'utilisation.</p> <p>Un cas d'utilisation doit être relié à au moins un acteur.</p>
 <p>Acteur 1</p> <p>Acteur 2</p>	<p><b>Cadre de diagramme</b> : rectangle qui sert à regrouper tous les cas d'utilisation d'un système et/ou montrer la position des acteurs (internes ou externes) par rapport à ce dernier.</p>

	<p><b>Généralisation</b> : représentée par une flèche orientée d'un élément à un autre.</p> <p><b>Entre acteurs</b> : en plus des actions qui lui sont propres, Acteur 2 fait tout ce que fait Acteur 1, on dit qu'il hérite de toutes ses relations.</p> <p><b>Entre cas d'utilisation</b> : signifie que cas d'utilisation 2 définit un comportement spécifique de cas d'utilisation 1. Dans ce cas il hérite à la fois de sa définition et de ses relations.</p>
	<p><b>Inclusion et extension</b> : flèche en tirets reliant deux cas d'utilisation. Elle est accompagnée d'un stéréotype qui définit la nature de la relation :</p> <p><b>&lt;&lt;include&gt;&gt;</b> : Cas d'utilisation 1 est inclut dans Cas d'utilisation 2 (cas d'utilisation 1 n'est réalisable que si cas d'utilisation 2 est d'abord réalisé).</p> <p><b>&lt;&lt;extend&gt;&gt;</b> : Cas d'utilisation 1 étend Cas d'utilisation 3 (Cas d'utilisation 1 ajoute des séquences</p>


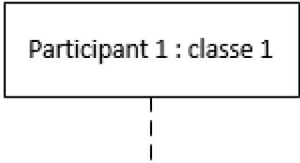
	de traitement à cas d'utilisation 3, qui sont facultatives selon l'acteur relié à cas d'utilisation 3 ou le scenario suivi).
	<b>Fragment de séquence :</b> fragment qui fait référence à un autre diagramme de séquence. Il est modélisé par un bloc avec une étiquette « ref » et un nom identique au nom du diagramme de référence. Les participants dans un fragment de séquence sont les mêmes dans son diagramme.

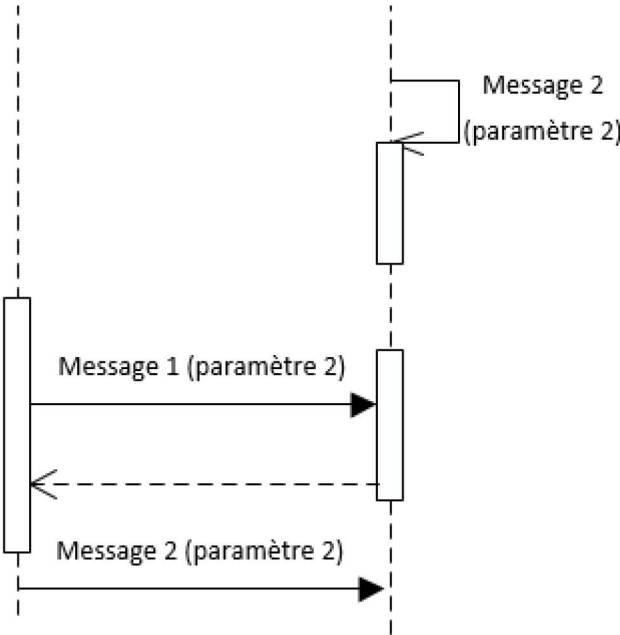
Tableau B.1 : Syntaxe UML : Diagramme de cas d'utilisation. [21]

### B.3.2 Diagramme de séquence

Le diagramme de séquence décrit les échanges de messages passés entre objets, et/ou de ses acteurs, au sein d'une interaction du système selon un ordre chronologique représenté sur un axe vertical. [21]

Le tableau suivant représente la syntaxe graphique présente le diagramme de cas de séquence :

Figure	Signification
	<b>Participants :</b> acteurs (humains ou inhumains) ou objets ayant un rôle précis dans une séquence d'actions.  <b>Ligne de vie :</b> ligne verticale en pointillés sur laquelle sont présentées

	<p>les actions d'un participant dans un ordre chronologique. Tout participant d'un diagramme possède sa ligne de vie.</p>
 <pre> sequenceDiagram     participant L1     participant L2     L1-&gt;&gt;L1: Message 1 (paramètre 2)     L1-&gt;&gt;L2: Message 2 (paramètre 2)     L2-&gt;&gt;L2: Message 2 (paramètre 2)     L2--&gt;&gt;L1: Message 2 (paramètre 2)     L1-&gt;&gt;L2: Message 2 (paramètre 2)     </pre>	<p><b>Barre d'activation</b> : barre positionnée sur une ligne, et qui signifie que le participant est occupé par une action.</p> <p><b>Message</b> : flèche horizontale qui sert à modaliser une communication entre deux participants à un instant donné. Chacune de ses deux extrémités est positionnée sur une ligne de vie. Un message possède un émetteur et un récepteur, un participant peut être à la fois émetteur et récepteur.</p> <p><b>Signature</b> : texte avec le message spécifiant son nom et ses paramètres. Message (param) et une signature.</p> <p><b>Message synchrone/asynchrone</b> : les messages synchrones sont de la forme de message1 et message3 tant dit que les messages asynchrones sont de la forme de message2.</p>



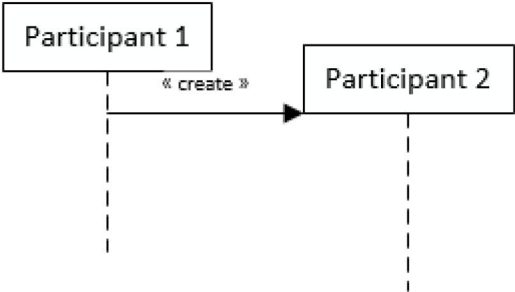
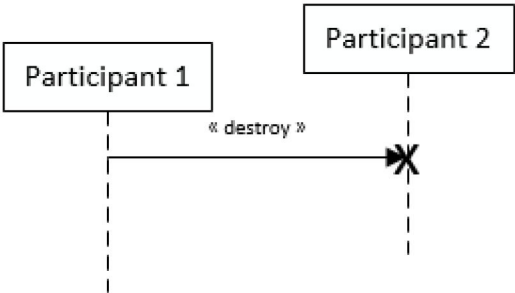
	<p><b>Messages de retour :</b> lorsqu'un participant appelle un autre en lui envoyant un message, l'appelé émet un message de retour à la fin de son action. Selon le schéma, le message de retour est modélisé par une flèche en pointillés.</p>
 <pre> sequenceDiagram     participant P1 as Participant 1     participant P2 as Participant 2     P1-&gt;&gt;P2: « create »     </pre>	<p><b>Messages de création :</b> certains participants n'existent pas depuis le début du diagramme, mais uniquement après un certain temps. Un message de création est un message stéréotypé&lt;&lt;create&gt;&gt;et sert à dire qu'un participant est responsable de la création d'un autre.</p>
 <pre> sequenceDiagram     participant P1 as Participant 1     participant P2 as Participant 2     P1-&gt;&gt;P2: « destroy »     destroy P2     </pre>	<p><b>Message de destruction :</b> un message de destruction est un message Stéréotypé&lt;&lt;destroy&gt;&gt; qui sert à dire qu'un participant est responsable de la destruction d'un autre. on remarque bien que la ligne de vie du participant détruit s'arrête à la croix.</p>

Tableau B.2 : Syntaxe UML : Diagramme de séquence. [21]

### B.3.3 Diagramme d'activité

Le diagramme d'activité représente les règles d'enchaînement des actions et décisions d'un système ou de ses composants au sein d'une activité. Il vient illustrer et consolider la description textuelle d'un cas d'utilisation.

Le tableau suivant représente la syntaxe graphique présente le diagramme de cas d'activité :

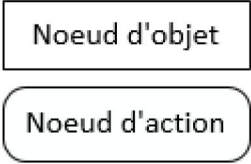
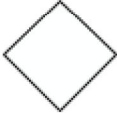
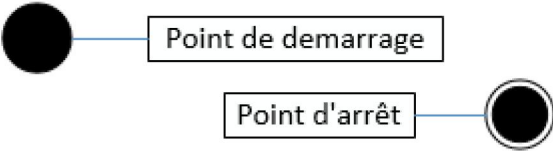
Figure	Signification
	<p><b>Nœud d'action</b> : Étape dans l'activité, dans laquelle les utilisateurs ou logiciels effectuent une certaine tâche.</p> <p><b>Nœud d'objet</b> : Représente des données qui passent le long d'un flux.</p>
	<p><b>Un nœud de fusion de décisions</b> représente un choix dans un flux et la fusion de différents flux suite à ce choix.</p>
	<p><b>Point de démarrage et d'arrêt</b></p> <p>Le diagramme est composé d'un point de démarrage, d'un point arrêt et d'action, qui sont représenté par des cercles noirs.</p>

Tableau B.3 : Syntaxe UML : Diagramme d'activité. [21]

# Bibliographie

- [1] : Noirault C. (2006) : Business Intelligence avec Oracle 10g, ENI Editions, 443 pages.
- [2] : Lafare M. (2006) : Thèse professionnelle "business intelligence" HEC Paris.
- [3] : Bensahla A. (2011) : Veille Technologique, Livre Blanc.
- [4] : Cabrol A. (2008) : Solutions Open Source de Business Intelligence, [www.adullact.org](http://www.adullact.org).
- [5] : Bruney M. (2011) : Propos sur les SI décisionnels, Communauté francophone des utilisateurs de données, <http://www.decideo.fr>.
- [6] : Caporiondo L. (2014) : Mise en place d'un système d'information décisionnel dans une entreprise, Portail national eduscol, <http://www.eduscol.education.fr>.
- [7] : Herr B. (2013) : Année du Big Data ?, Décembre 2013 - Paru dans Solutions & Logiciels N°44.
- [8] : Chaumais R. (2013) : Finyear magazine numéro 29 Novembre, Paris.
- [9] : Chaker H. (2008) : Informatique décisionnelle et management de la performance de l'entreprise, Cahier de recherche - Laboratoire orélanais de gestion - Université d'Orléans.
- [10] : Fernandez A. (2008) : Les nouveaux tableaux de board des managers, Groupe EYROLLES, 476 pages.
- [11] : Fernandez A. (2008) : Le projet décisionnel clés en main "Méthode pour concevoir et réaliser le système décisionnel de l'entreprise, <http://www.decideo.fr>.
- [12] : Collectif SMILE Open source solutions (2012) : Décisionnel : le meilleur des solutions open source, Livre blanc, <http://www.smile.fr>.
- [13] : Bordanave C. (2012) : Business Intelligence – Reporting, Rapport de stage, Université Bordeaux 1.
- [14] : Valentin P. (2009) : Introduction à la B.I Avec SQL Server 2008, Association Dotnet France, <http://www.dotnet-france.com>.
- [15] : El Omari A. (2012) : Business Intelligence : Quels bénéfices pour votre entreprise ?, interview recueillie par Asmâa AYOUB, <http://www.rdvdesdecideurs.com>
- [16] : Talend (2015) : Open Source Integration Software and Data Management Tools, <http://www.talend.com>.
- [17] : Pentaho (2015) : Business analytics and business intelligence leaders, [www.pentaho.fr](http://www.pentaho.fr)
- [18] : osbi.fr (Open Source Business Intelligence) (2010) : Cubes OLAP avec Mondrian, <http://www.osbi.fr>.
- [19] : Document interne de l'EPB (2015), Site officiel du port de Béjaïa : <http://www.portdebejaia.dz/>.
- [20] : Qlik : Business Intelligence and Data Visualization Software (2015), <http://www.qlik.com>
- [21] : Pascal Roques (2007) : UML 2 : Modéliser une application web, Eyrolles, 4ème édition.
- [22] : MySQL (2015) : La base de données la plus populaire au monde, [www.mysql.fr](http://www.mysql.fr).



---

## Résumé

Les premières applications dans les entreprises des outils informatiques d'aide au pilotage et à la décision, désormais appelés «Business-intelligence», datent du début des années quatre-vingt-dix ; les «Executive Information Systems» étaient à l'origine réservés à l'information des dirigeants. La banalisation dans les entreprises, vers le milieu de la décennie quatre-vingt-dix, des technologies de «datawarehouse» et de «datamart» a accru la possibilité de développer des systèmes d'information organisant les données de façon facilement accessible et appropriée à la prise de décisions et les représentant de manière intelligente au travers d'outils spécialisés tournés vers l'analyse multidimensionnelle avec notamment les bases «OLAP». Ce qui a pour effet de faciliter le management de la performance de l'entreprise. Actuellement, l'informatique décisionnelle facilite la modélisation d'une performance multidimensionnelle pour en permettre le pilotage au travers d'un des modèles possibles de sa représentation : le tableau de bord prospectif.

**Mots clés :** Business intelligence, ETL, Reporting, Tableau de bord, UML, Talend, QlikView.

---

## Abstract

The first applications in companies of tools and pilot assistance with the decision, now known as "Business-intelligence", date from the early ninety years; the "Executive Information Systems" were originally reserved to information officers. The trivialization in companies, about the middle of the ninety decade, technology "data warehouse" and "data mart" has increased the possibility of developing information systems easily accessible way of organizing data and appropriate decision-making and representative intelligently through specialized tools oriented multivariate analysis including the basics "OLAP". This has the effect of facilitating the management of the company's performance. Currently, BI facilitates the modeling of a multidimensional performance to allow the steering through a possible models of its representation: the Balanced Scorecard.

**Keywords :** Business intelligence, ETL, Reporting, Dashboard, UML, Talend, QlikView.

---