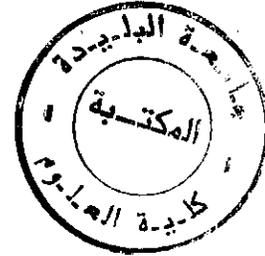


République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahlab, Blida  
USDB.

Faculté des sciences.  
Département informatique .



**Mémoire pour l'obtention  
d'un diplôme d'ingénieur d'état en informatique.**  
Option : Système d'Information

Sujet :

**Conception et réalisation d'un  
tableau de bord basé sur un outil  
d'aide à la décision de la fonction  
Monétique et CCP.**

**Présenté par :** Mazouzi Med Amine  
Atroun Nassima

**Promoteur :** M.Bala  
**Encadreur :** R.Benzouz

**Organisme d'accueil :** Algérie Poste.

**Soutenu le:** date soutenance, devant le jury composé de :

Nom. président du jury, grade, organisme

**Président**

Nom examinateur 1, grade, organisme

**Examineur**

Nom examinateur 2, grade, organisme

**Examineur**

- Numéro /2006-2007

MIG-004-179.1

## REMERCIEMENTS

Tout d'abord, nous souhaiterions remercier Monsieur Ben Azouz notre encadreur pour son professionnalisme et pour nous avoir donné l'opportunité de réaliser notre projet. Et Monsieur Bechar pour son soutien.

Ensuite nous remercions aussi Monsieur Bala pour ses conseils et ses remarques constructives, apportés tout au long du projet.

Aussi, nous remercions toute l'équipe informatique d'Algérie Poste pour leur aide, leur bonne humeur, et la bonne ambiance qui régnait. Rabeh Idir, Hicham, Rafik, Toufik, Telja, Amel, Samia, Said, ....

Enfin nous tenons aussi à exprimer nos gratitudeles plus sincères aux membres du Jury, pour l'honneur qu'ils nous font en acceptant de juger notre travail.

## Résumé

Les entrepôts de données sont utilisés pour l'exploitation et l'analyse de grands volumes de données extraits des systèmes d'informations en exploitation.

Dans un entrepôt de données, les données sont sélectionnées et préparées (pour répondre aux questions vitales de l'entreprise), intégrées (à partir des différentes sources de renseignements) et datées (elles gardent la trace de leur origine). Un modèle multidimensionnel organise l'entrepôt de données autour des tables de faits et dimensions, permettant de mesurer l'activité selon plusieurs axes d'analyses

Les systèmes OLAP (Online Analytic Processing) permettent une exploration interactive des données contenue dans un entrepôt.

Un entrepôt de donnée est une collection de données historiques exploitées par les applications d'aide à la décision. Il centralise des données d'intérêts pour un groupe d'utilisateurs afin de rendre leur accès rapide, peu coûteux et efficace. Dans notre projet, nous nous intéressons à la conception et le déploiement d'un entrepôt de données (Data Warehouse) métier, concernant la Monétique et CCP de l'entreprise d'Algérie Poste.

L'approche retenue pour la réalisation du système est celle du cycle de vie dimensionnel proposée par Ralph Kimball. Notre approche est divisée en quatre phases : Identification des besoins, Conception de la zone de préparation des données, Reporting et enfin Développement et Déploiement du système.

### Mots clé :

Décisionnel, Outil d'Aide à la Décision, Système d'Information Décisionnel, SID, informatique décisionnelle, système d'entrepôtage, Analyse multidimensionnelle, Outil de Pilotage, Tableau de bord, Entrepôt de données, Magasin de données, OLAP, ETL, ODS, Outil de restitution, Cube.

## **Abstract**

The data warehouses are used for working and analysis of big volumes of data extracted from the systems of information in working.

In a data warehouse, data are selected player and prepared (to answer the vital questions of the firm); integrated (from the different sources of information) and dated (they keep the trace of their origin). A multidimensional model organizes the Data Warehouse around tables of facts and dimensions, allowing to measure activity according to several axis of analysis.

OLAP systems (Online Analytic Processing) allow an interactive exploration of data contained in a data warehouse.

A data warehouse is a collection of historical data exploited by the applications of help to decision. It centralizes data of interests for a group of users to return their quick, not very costly and efficient access. In our plan, we are interested in comprehension of the deployment of a Dated Warehouse job, concerning Electronic banking and Count Check Mail of the firm of Algeria Posts.

The approach kept for the realization of the system is that of the dimensional cycle of life offered by Ralph Kimball. Our approach is divided into four stages: Identification of needs, Comprehension of the zone of feeding of data, Reporting and finally Development and Deployment of the system

### **Keywords:**

Business Intelligence, Data Warehouse, Data mart, OLAP, ETL, Reporting  
Operational Data Store, ODS, Dashboard.

# SOMMAIRE

<b>Introduction générale</b>	
I.1 Contexte général .....	1
I.2 Problématique .....	2
I.3 Originalité du travail .....	4
I.4 Organisation du mémoire .....	4
<b>Partie I : Présentation de l'organisme d'accueil</b>	
I.1 Description de l'organisme d'accueil : Algérie Poste..	6
I.2 Présentation de la structure d'accueil : La direction Informatique	8
<b>Partie II : Les systèmes décisionnels : Etat de l'art</b>	
I.1 Informatique décisionnelle .....	12
I.2 Le système d'information décisionnel .....	14
I.3 Composants de base de l'entrepôt de donnée .....	16
I.4 Conclusion .....	29
<b>Partie III : Etude de l'existant</b>	
III.1 Description de l'organisation .....	30
III.2 Les postes de décisions .....	30
III.3 Description du système CCP .....	30
III.4 Description du système Monétique .....	31
III.5 Avantage du système .....	34
III.6 Inconvénients du système .....	35
III.7 Conclusion .....	35
<b>Partie IV : Démarche et choix de conception</b>	
<b>Chapitre I : Conception du niveau de consolidation.</b> .....	44
I.1 Choix du modèle .....	37
I.2 Modélisation dimensionnelle .....	38
I.3 Les différents modèles conceptuels .....	40
I.4 Navigation dans les données .....	42
I.5 Les architectures d'un entrepôt de donnée .....	42
I.6 Conclusion .....	43
<b>Chapitre II: Méthode de conception d'un Entrepôt de données</b> .....	44
II.1 Rappel des caractéristiques du contexte d'application...	44
II.2 Présentation des méthodes de conception .....	44
II.3 Conclusion .....	48
<b>Chapitre III : Application du cycle de vie et choix du formalisme</b> .....	49
III.1 Cycle de vie suivi par l'étude .....	49
III.2 Description des différentes phases du cycle de vie .....	50
III.3 Formalisme à utiliser .....	51
III.4 Résumé .....	58
III.5 Conclusion .....	58
<b>Partie V : Conception de l'Entrepôt de donnée AP</b>	
<b>Chapitre I : Phase définition des besoins</b> .....	59
I.1 Définition des besoins .....	59

I.2 La méthode à suivre .....	59
I.3 Les cas d'utilisation .....	61
I.4 Description des cas d'utilisation .....	61
I.5 Conclusion .....	71
<b>Chapitre II : Modélisation dimensionnelle</b> .....	72
II.1 Identification des Processus .....	72
II.2 Description des modèles .....	72
II.3 Conclusion .....	84
<b>Chapitre III : Conception des éléments de la zone de préparation des données</b> .....	85
III.1 Phase Extraction .....	85
III.2 Phase Transformation et Chargement .....	88
III.3 Diagramme d'états-transitions .....	91
III.4 Diagramme d'activité .....	93
III.5 Conclusion .....	94
<b>Chapitre IV : Reporting</b> .....	95
IV.1 Interface Tableau de bord .....	95
IV.2 Interface administrateur .....	99
IV.3 Conclusion .....	103
<b>Chapitre V : Schéma de mise en œuvre et déploiement</b> .....	104
V.1 Diagramme de composant .....	104
V.2 Diagramme de déploiement .....	108
V.3 Conclusion .....	108
<b>Partie VI : Réalisation</b>	
<b>Chapitre I : Environnement et outils de développement</b> .....	109
I.1 Architecture de déploiement .....	109
I.2 Caractéristiques de l'architecture trois tiers (3/3).....	109
I.3 Environnement Logicielle.....	109
I.4 Introduction aux outils utilisés dans le développement	110
I.5 Implémentation .....	111
I.6 Description de l'utilisation de Talend et AWM.....	112
I.7 Conclusion .....	118
<b>Chapitre II : Description du système</b> .....	119
II.1 Authentification .....	119
II.2 Interface Administrateur .....	119
II.3 Interface Tableau de Bord .....	121
II.4 Conclusion .....	124
<b>Conclusion Général</b> .....	125
<b>Annexe</b>	
<b>Glossaire</b>	
<b>Bibliographie</b>	

## Liste des figures

N°Figure	Désignation	Page
Figure I.1	Organigramme d'Algérie Poste.	7
Figure I.2	Organigramme de la direction Monétique.	10
Figure I.3	Organigramme des postes du CCP concerné par l'étude.	11
Figure II.4	L'évolution de l'utilisation de l'informatique décisionnelle.	12
Figure II.5	Les différents niveaux d'un système d'information.	14
Figure II.6	Différente facette d'un tableau de bord.	15
Figure II.7	Système Source, Base de production	16
Figure II.8	Zone de préparation des données.	17
Figure II.9	Entrepôt de donnée.	18
FigureII.10	Data Mart.	19
FigureII.11	les composant du système d'information décisionnel	21
FigureII.12	OLAP.	22
FigureII.13	Application utilisateurs.	23
FigureII.14	Outil de requêtage.	24
FigureII.15	Les outils de l'entrepôt de donnée : Les Méta-donnée.	24
FigureII.16	Structure de base d'une table de fait.	27
FigureII.17	Structure de base d'une table de dimension.	28
FigureIII.18	Transaction de retrait par Carte.	32
FigureIII.19	Transaction de paiement par Carte.	32
FigureIV.20	Représentation cubique de l'activité gestion des Produits.	39
FigureIV.21	Schéma en Etoile.	40
FigureIV.22	Schéma en Flocon de neige.	41
FigureIV.23	Schéma du cycle de vie dimensionnel.	46
FigureIV.24	Transition des données du modèle relationnel vers le modèle dimensionnel.	48
FigureIV.25	Schéma du cycle de vie suivi par notre étude.	49
FigureIV.26	Exemple de représentation du besoin en indicateur avec le diagramme de cas d'utilisation.	54
FigureIV.27	Représentation de la relation extend dans un cas d'utilisation appliqué a un niveau décisionnel.	55
FigureVI.28	Architecture client/serveur à trois tiers (3/3).	109
FigureVI.29	Environnement logiciel.	110
FigureVI.30	Exemple d'un fichier CSV.	111
FigureVI.31	Exemple d'un fichier Delimited.	112
FigureVI.32	Exemple d'un Fichier Positional.	112
FigureVI.33	Choix du langage de programmation.	113
FigureVI.34	Interface de l'ETL Talend.	113
FigureVI.35	Talend : Création d'un Job.	114
FigureVI.36	Configuration du Job.	114
FigureVI.37	Indication sur le chemin de la table.	115
FigureVI.38	L'utilisation de tMap.	115
FigureVI.39	Exécution du Job.	116
FigureVI.40	Connexion a AWM.	117

## Liste des figures (suite) :

N°Figure	Désignation	Page
FigureVI.41	Création de la correspondance entre les tables et leurs dimensions.	118
FigureVI.42	Interface d'authentification.	119
FigureVI.43	Interface Administrateur.	120
FigureVI.44	Ajouter utilisateur.	120
FigureVI.45	Liste des utilisateurs du système.	121
FigureVI.46	Interface Tableau de Bord.	122
FigureVI.47	Histogramme de l'activité LSB sur tout le territoire algérien.	123
FigureVI.48	Histogramme LSB, Forage vers le Bas.	124

## Liste des Diagrammes

N°Diagramme	Désignation	Page
Diagramme V.1	Diagramme des Cas d'utilisation du Processus CCP.	62
Diagramme V.2	Diagramme des Cas d'utilisation du Processus Monétique.	63
Diagramme V.3	Diagramme des Cas d'utilisation de Processus Compensation.	64
Diagramme V.4	Diagramme des Cas d'utilisation du Processus Acquéreur.	66
Diagramme V.5	Diagramme des Cas d'utilisation du Processus Emetteur.	67
Diagramme V.6	Diagramme des Cas d'utilisation de Processus Front Office	69
Diagramme V.7	Diagramme des Cas d'utilisation de l'Activité Panne.	71
Diagramme V.8	Diagramme Multidimensionnel : Activité CCP.	74
Diagramme V.9	Diagramme Multidimensionnel : Panne CCP.	74
Diagramme V.10	Diagramme Multidimensionnel : Activité Compensation.	75
Diagramme V.11	Diagramme Multidimensionnel : Activité Rejet Compensation.	76
Diagramme V.12	Diagramme Multidimensionnel : Activité Commerçant.	77
Diagramme V.13	Diagramme Multidimensionnel : Activité Gestion Commerçant.	78
Diagramme V.14	Diagramme Multidimensionnel : Activité Carte.	79
Diagramme V.15	Diagramme Multidimensionnel : Activité Gestion Carte.	80
Diagramme V.16	Diagramme Multidimensionnel : Activité LSB.	81
Diagramme V.17	Diagramme Multidimensionnel : Activité Font Office1.	82
Diagramme V.18	Diagramme Multidimensionnel : Activité Front Office 2.	83
Diagramme V.19	Diagramme Multidimensionnel : Activité Panne.	84
Diagramme V.20	Diagramme de cas d'utilisation : Extraction des Fichiers Plats (E).	85
Diagramme V.21	Diagramme de séquence : Extraction des Données (E) (Cas 1).	86
Diagramme V.22	Diagramme de séquence : Extraction des Données (E) (Cas 2).	87
Diagramme V.23	Diagramme de cas d'utilisation : Transformation et chargement des données (TL).	88
Diagramme V.24	Diagramme de séquence : Transformation et chargement des données (Cas 1).	89
Diagramme V.25	Diagramme de séquence : Transformation et chargement des données (Cas 2).	90

## Liste des Diagrammes (suite) :

N°Diagramme	Désignation	Page
Diagramme V.26	Diagramme Etat Transition : Etat du système.	91
Diagramme V.27	Diagramme Etat Transition : Etat de l'entrepôt de données.	92
Diagramme.V.28	Diagramme d'activité du système.	94
Diagramme V.29	Diagramme de cas d'utilisation : Interface tableau de bord.	95
Diagramme V.30	Diagramme de séquence authentification (Cas 1).	96
Diagramme V.31	Diagramme de séquence authentification (Cas 2).	96
Diagramme V.32	Diagramme de Séquence : Visualisation Indicateur (Cas 1).	97
Diagramme V.33	Diagramme de Séquence : Visualisation Indicateur (Cas 2).	97
Diagramme V.34	Diagramme de Séquence : Impression résultat (Cas 1).	98
Diagramme V.35	Diagramme de Séquence : Impression résultat (Cas 2).	98
Diagramme V.36	Diagramme de Séquence : Quitter le système (Cas 1).	98
Diagramme V.37	Diagramme de Séquence : Quitter le système (Cas 2).	98
Diagramme V.38	Diagramme de cas d'utilisation : Interface administrateur.	99
Diagramme V.39	Diagramme de Séquence : Création d'un nouveau compte utilisateur (Cas 1).	100
Diagramme V.40	Diagramme de Séquence : Création d'un nouveau compte utilisateur (Cas 2).	100
Diagramme V.41	Diagramme de Séquence : Modification des droits d'accès aux indicateurs (Cas 1).	101
Diagramme V.42	Diagramme de Séquence : Modification des droits d'accès aux indicateurs (Cas 2).	101
Diagramme V.43	Diagramme de Séquence : Supprimer un compte utilisateur (Cas 1)	102
Diagramme V.44	Diagramme de Séquence : Supprimer un compte utilisateur (Cas 2)	102
Diagramme V.45	Diagramme de Séquence : Chargement des fichiers plats (Cas 1).	103
Diagramme V.46	Diagramme de Séquence : Chargement des fichiers plats (Cas 2).	103
Diagramme V.47	Diagramme de Composent : Authentification.	105
Diagramme V.48	Diagramme de Composent : Tableau de Bord.	106
Diagramme V.49	Diagramme de Composent : Administration.	107
Diagramme V.50	Diagramme de Déploiement du Projet.	108

## - Liste des Tableaux

Tableau	Désignation	Page
Tableau IV.1	Résumé des différences des deux modèles de conception de base de donnée	37
Tableau IV.2	Avantage et Inconvénients d'UML.	57
Tableau IV.3	Diagramme UML par besoin décisionnel.	57
Tableau IV.4	Citation des principaux Diagrammes UML utilisé dans les différentes phases du cycle de vie.	58
Tableau VI. 5	Tableau récapitulatif des outils utilisés pour le développement de l'outil d'aide à la décision.	111

## Lexique Postal

Abréviation	Description
AP	Algérie Poste.
Back Office	Le Back Office désigne l'ensemble des parties du système d'information auxquelles l'utilisateur final n'a pas accès. Il s'agit donc de tous les processus internes à l'entreprise (compensation, oppositions...).
CCP	Compte Chèque Postaux.
Commerçant	Application client (GAB).
Carte (Carte bancaire)	Carte bancaire a piste/puce EMV, pour les paiement et retrait d'argent a partir des commerçant. AP fournit trois types de produit carte : Carte Jaune, Classique et Gold
CIB	Carte Inter Bancaire.
CNEP	Caisse Nationale d'Epargne et de Prévoyance.
DAB	Distributeur Automatique Bancaire.
EMV	Europay Mastercard Visa
EPIC	Etablissement Public à caractère Industriel et Commercial.
Front Office	Le Front-Office (parfois appelé également <i>Front line</i> ) désigne la partie frontale de l'entreprise, visible par la clientèle et en contact direct avec elle, comme les équipes de marketing, de support utilisateur ou de service après-vente. (authentification, autorisation,...).
GAB	Guichet automatique bancaire.
HB	HB technologie, entreprise algérienne spécialisée dans la production et la personnalisation de cartes intelligentes et des documents d'identification.
L'émetteur	Banque du porteur.
L'accepteur	Un commerçant qui accepte le règlement d'un produit ou d'un service par une carte, il est lié à une banque acquéreur, commerçant, GAB.
L'acquéreur	acquéreur du GAB.
Local	Toute opération financière (Transaction) au niveau des GAB d' AP.
LSB	Libre Service Bancaire
Porteur	Client d'une banque, souscrit un contrat carte.
RIB	Relevé d'Identité Bancaire. Il contient les coordonnées bancaires : Nom, Prénom, Adresse, Numéro de Compte.
RSP	Responsable des Services Postaux
SATIM	Société d'Automatisation des Transactions Interbancaires et de Monétique
SPA	Société Par Actions.
SDSSI	Sous Directeur de la Sécurité et des Systèmes Informatiques
TES	Transactions Électroniques Sécurisées.
TPE	Terminal de Paiement Externe.
UPW	Unité Postal de Wilaya.

## Lexique Décisionnel

Abréviation	Description
AWM	Analytic Workspace Manager outil de Création de vues du cube, de définition de la structure dimensionnelle du cube.
Business intelligence	l'informatique décisionnelle
BI	Business intelligence.
CSV	comma-separated values, un des types des fichiers plats
Data Warehouse	Entrepôt de données.
Data mart	Magasin de données.
Dashboard	Tableau de Bord.
ETL	Extract Transforme Load.
ERP	<i>Enterprise Resource Planning</i> , aussi appelés <i>Progiciels de Gestion Intégrés (PGI)</i> .
EIS	Exécutive Information Système.
Fichier plat	Fichier simple contenant une série d'enregistrements organisés en champs structurés ou libres ou encore en colonnes d'informations.
HOLAP	Hybrid OLAP.
IHM	Interface Homme Machine.
Java	Langage de programmation de Sun Microsystems. Ressemble à celui du C++. Parfait pour créer des programmes à intégrer sur le Web.
Job	Tâche.
JSP	Java Server Pages, est une technologie basée sur Java qui permet aux développeurs de générer dynamiquement du code HTML, XML ou tout autre type de page Web.
MOLAP	Multidimensional OLAP.
Méta-données	Des données sur des données.
MERISE	Méthode d'analyse, de conception et de gestion de projet.
ODS	Operational Data Store.
OLAP	Online Analytic Processing.
OLTP	Online Transaction Processing.
OLAP	Online Analytic Processing.
Perl	Practical Extraction and Report Language, Langage de programmation généraliste utilisé dans l'administration système, le développement Web, la programmation réseau, la création d'interfaces graphiques, etc.
ROLAP	Relational OLAP.
Reporting	Restitution.
SID	Système d'Information Décisionnel.
SIAD	Système Informatique d'Aide à la Décision
SGBD	Système de Gestion de Base de Donnée.
SI	Système d'Information.
SQL	Structured Query Language : Langage informatique ayant pour objet le dialogue avec une base de données relationnelle.
SNAPSHOT	vue matérialisée ( <i>materialised view</i> ).
TCP-IP	Transmission Control Protocol-Internet Protocol
UML	Unified Modeling Language.

**Introduction**

**Générale**

## I. INTRODUCTION GENERALE

### I.1 Contexte général :

Dans un environnement économique très pressant, tous les utilisateurs de l'entreprise ayant à prendre des décisions doivent pouvoir accéder en temps réel, aux données de l'entreprise, doivent pouvoir les traiter et en extraire l'information pertinente afin de prendre les bonnes décisions. Ils se posent alors des questions du type : « Quelle est l'évolution des chiffres d'affaires par type de magasin et par période », ou encore : « Quels sont les résultats des ventes par gamme de produit et par région pour l'année dernière ? » ...

Le système d'information opérationnel ne peut satisfaire ces besoins pour au moins deux raisons principales :

- Les bases de données opérationnelles sont trop complexes pour pouvoir être appréhendées facilement par tout utilisateur.
- Le système opérationnel ne peut être interrompu afin de pouvoir répondre à des questions qui nécessitent des calculs importants.

Une nouvelle tendance a commencé à apparaître consistant à développer un système d'information orienté vers la décision et est séparé du système d'information opérationnel. Et pour cela, il faut donc garder un historique et restructurer les données de production.

Les entrepôts de données ou Data Warehouse sont la réalisation de ces nouveaux systèmes d'information. [Gill&Tom, 2000].

Cette apparition a été rendue possible grâce aux progrès technologiques à coûts constants. Parmi ces progrès, on cite :

- L'augmentation importante des capacités de stockage.
- L'introduction des techniques du parallélisme dans l'informatique de gestion, techniques qui permettent des accès rapides à de grandes bases de données.

Le Data Warehouse n'est pas une usine à produire l'information, mais plutôt un moyen de la mettre à la disposition des utilisateurs de manière efficace et organisée.

La mise en oeuvre d'un Data Warehouse est un processus complexe. Il permet la mise en place d'un outil décisionnel s'appuyant sur les informations pertinentes pour l'entreprise, centré sur le métier utilisateur.

L'objectif à atteindre est de recomposer les données disponibles pour en donner :

- Une vision intégrée et transversale aux différentes fonctions de l'entreprise,
- Une vision métier à travers différents axes d'analyse,
- Une vision agrégée ou détaillée suivant le besoin des utilisateurs.

Et de cet objectif découle l'avantage de :

- Rendre l'utilisateur autonome (par rapport au système opérationnel) dans l'exploitation des informations. Il est, dès lors, en mesure d'analyser les procédures de gestion et l'état du système, de prendre des décisions et de mettre en oeuvre des actions stratégiques.
- Consolider et organiser les données historiques d'activités afin de les analyser et d'aider ainsi les gestionnaires à prendre des décisions concernant l'organisation.

Les organisations qui ont mis en oeuvre, avec succès, des projets de ce type ont vu leurs efforts récompensés par un avantage compétitif, significatif et un profit accru.

Le but de notre projet est la conception et le déploiement d'un Data Warehouse au sein d'une organisation, c'est-à-dire sauvegarder les données brutes actuellement existantes (même sur des supports différents), afin de les analyser et de les exploiter dans le futur et ceci pour l'aide à la prise de décisions. Cette étude sera faite avec comme cas « Système d'aide à la décision pour les fonctions Monétique et CCP (Compte Chèque Postaux) » autrement dit :

- Récupérer,
- Organiser,
- Consolider,

Les données historiques des transactions dans le but de prendre des décisions concernant :

- Un commerçant,
- Le lieu de l'installation d'un commerçant,
- L'ajout d'un nouveau commerçant,
- Le suivi de l'activité d'un bureau de poste, ... etc.

## **I.2 Problématique :**

### **I.2.1 Présentation du sujet et problématique :**

Aujourd'hui, Algérie poste a à sa disposition une masse de données importante [RAP, 2006] :

Pour le réseau :

- 3287 bureaux de poste.
- 2636 bureaux informatisés.
- 25550 employés.
- 380 GAB.
- 110 DAB.

Pour le nombre de clients :

- 7.1 millions de comptes CCP.
- 3.6 millions de comptes d'épargne.
- plus de 1 million de visiteurs par jour au niveau des bureaux de postes.
- 175 000 porteurs pour un trafic mensuel de 60 000 transactions

Pour le nombre de carte bancaire :

- 108450 Carte SATIM (Carte CIB : carte de retrait et paiement).
- 569289 Carte HB (carte de retrait) activé.

Pour le nombre de transaction :

- 1887 : Moyenne des transactions journalières des retraits effectués par les cartes de retrait (Carte HB).
- 56625 : Moyenne des transactions mensuelles des retraits effectués par les cartes de retrait (Carte HB).
- 227 : moyenne des transactions mensuelles des cartes CIB.
- 30 : moyenne des transactions journalières des cartes CIB.

Pour les flux financiers, Chiffres-clés :

- 15 000 milliards de dinars de fonds manipulés.
- Chiffre d'affaires : 15 milliards de dinars.
- Courrier : 234 millions d'objets traités.
- Trafic des mandats : 31 millions.

Dans ce contexte, Algérie poste se trouve dans l'incapacité d'analyser par ces moyens disponibles, le volume de données de façon rapide, correcte et efficace.

Confronté à des problèmes de gestion d'une grande masse de données. Elle dispose d'une multitude de systèmes opérationnels au niveau de ces structures (Monétique et CCP).

Les principales difficultés auxquelles elle fait face sont :

- Complexité et diversité de la structure conceptuelle existante (Système opérationnel) rendant l'intégration des données difficiles pour une vue multi-axiale.
- Grande masse de données est stockée dans plusieurs bases de données (dans notre cas deux bases de données hétérogènes), ce qui rend difficile l'interrogation et l'analyse.
- Retard dans l'établissement des rapports d'activité, cela peut prendre quelques jours voire quelques semaines. Ce qui fait que l'information n'est pas disponible au moment voulu.

Pour cela, le développement et l'implémentation d'une architecture Data Warehouse, pouvant regrouper les indicateurs pertinents de pilotage, permettra à l'entreprise de mieux contrôler son activité et maintenir son titre de leader dans un environnement où se développe de plus en plus l'aspect d'incertitude.

### **I.2.2. Objectif :**

Notre projet consiste à développer un outil d'aide à la décision pour les Services Monétique et CCP d'Algérie Poste, partant de deux systèmes d'informations existants conçus pour la gestion de la Monétique et CCP.

L'outil devra exploiter les données provenant des deux bases de données et fournir aux décideurs les moyens de les visualiser et d'en extraire de la connaissance afin d'améliorer la gestion et d'accroître les performances de l'entreprise.

### **1.3 Originalité du travail ( Apport du travail) :**

A l'issu de cette étude, nous avons intégré l'UML a la conception des différentes phase du cycle de vie appliqué, dans le but d'exploité la puissance des diagrammes UML, pour mettre en évidence le système.

UML est un langage formel et normalisé (notation ouverte et riche qui permet un gain de précision et de stabilité), mais c'est aussi un support de communication performant qui cadre l'analyse et facilite la compréhension de représentations abstraites. [Brun&Cam, 2005]

C'est donc un excellent moyen de communication dans une équipe, qui donne à travers sa représentation d'un système une bonne compréhension de son fonctionnement. Enfin UML, de par sa notation ouverte, offre une grande polyvalence dans ses domaines d'applications. [Brun&Cam, 2005]

C'est donc ce qui motive cette étude, donner un sens à la modélisation en UML dans un contexte de conception de Data Warehouse et de modélisation de données.

### **1.4 Organisation du mémoire :**

**La Partie I:** concerne la description de l'organisme d'accueil, ses missions, sa structure et ses objectifs.

**La Partie II,** nous présentons a cette partie les définitions : des système décisionnels, tableau de bord et des composants de base du Data Warehouse. Cette partie s'intitule : "les système décisionnels : l'état de l'art".

**La partie III** concerne l'étude de l'existant, car il est indispensable d'étudier l'organisme d'accueil, de se familiariser avec les termes de l'entreprise, de comprendre le métier concerné par l'étude avant de lancer le projet.

**La partie IV** concerne les Démarches et choix de conception. A cette partie nous choisissons le niveau de consolidation de notre Entrepôt, la méthode choisi pour le développement du projet et le formalisme appliqué au différents phase du processus de conception et développement de ce dernier.

Elle est composée de Trois chapitres :

Chapitre 1 : Méthode de conception du modèle logique et physique.

Chapitre 2 : Méthode de conception d'un Data Warehouse (Entrepôt de données).

Chapitre 3 : Application du cycle de vie et choix du formalisme.

**La Partie V** concerne la conception de notre outil elle s'intitule « Conception du Data Warehouse », a cette partie nous retrouvons la conception de chaque phase du cycle de vie appliqué a notre projet.

Elle est composée de cinq chapitres :

*Chapitre 1* : Phase Définition des besoins.

*Chapitre 2* : Modélisation dimensionnel.

*Chapitre 3* : Conception des éléments de la zone de préparation des données.  
Est composé de deux phases :

- *Phase 1* : Extraction.
- *Phase 2* : Transformation et chargement.

*Chapitre 4* : Reporting.

Est composé de deux Phases :

- *Phase 1* : Interface du tableau de bord.
- *Phase 2* : Interface Administrative.

*Chapitre 5* : Schéma de mise en œuvre et déploiement.

**La partie VI** c'est la phase final : la Réalisation. A cette partie nous introduisons tous les outils qui on participés au développement de notre projet, ainsi qu'une description du future outil d'aide a la décision.

**PARTIE I**

**Présentation de**

**l'organisme d'accueil**

Dans cette première partie nous allons présenter l'organisme d'accueil, où nous avons effectué notre stage de fin d'études.

D'abord nous allons commencer par la description de l'organisme d'accueil, puis nous passons aux principales missions d'Algérie poste, suivi des objectifs d'Algérie poste.

Enfin on termine cette partie par la présentation de l'unité concernée par notre étude : la Direction informatique où sera déployé et installé le futur outil d'aide à la décision, objet de ce mémoire.

## **I. PRESENTATION DE L'ORGANISME D'ACCUEIL**

### **I.1 Description de l'organisme d'accueil : Algérie Poste**

La création d'Algérie poste en 2002 est intervenue dans le prolongement du processus de réforme de la poste et des télécommunications introduit par la loi 2000/03 du 5 août 2000. Ce dispositif légal a consisté principalement à : [RAP, 2006]

- Séparer la poste des télécommunications, en érigeant la poste comme Etablissement Public à caractère Industriel et Commercial (EPIC) et les télécommunications en Société Par Actions (SPA) ;
- Recentrer l'Etat autour de la politique sectorielle ;
- Instituer une autorité de régulation chargée de veiller au respect des règles du marché.

#### **I.1.1 Les principales missions confiées à Algérie Poste : [RAP, 2006]**

- L'émission de timbres-poste et toute autre marque d'affranchissement et valeur fiduciaire postale.
- Les activités relevant du régime de l'exclusivité en matière de poste aux lettres n'excédant pas respectivement : 250g en 2006 et 50g en 2008 (décret 03/397 du 6 décembre 2004) et en matière de services postaux.
- Tout autre service de collecte, de transport et de distribution d'objets et de marchandise.
- La collecte de fonds au profit de la direction générale de la Caisse Nationale d'Épargne et de Prévoyance (CNEP).
- La collecte de fonds à travers les services des chèques postaux.
- L'émission et le paiement des mandats postaux.
- Le paiement des pensions.
- Le paiement des allocations forfaitaires et indemnités d'intérêt général pour le compte du ministère du travail et de la protection sociale.
- La vente de vignettes automobiles.

#### **I.1.2 Présentation organisationnelle :**

Voir la figure 1.

**I.1.2.1 Organigramme d'Algérie Poste :**

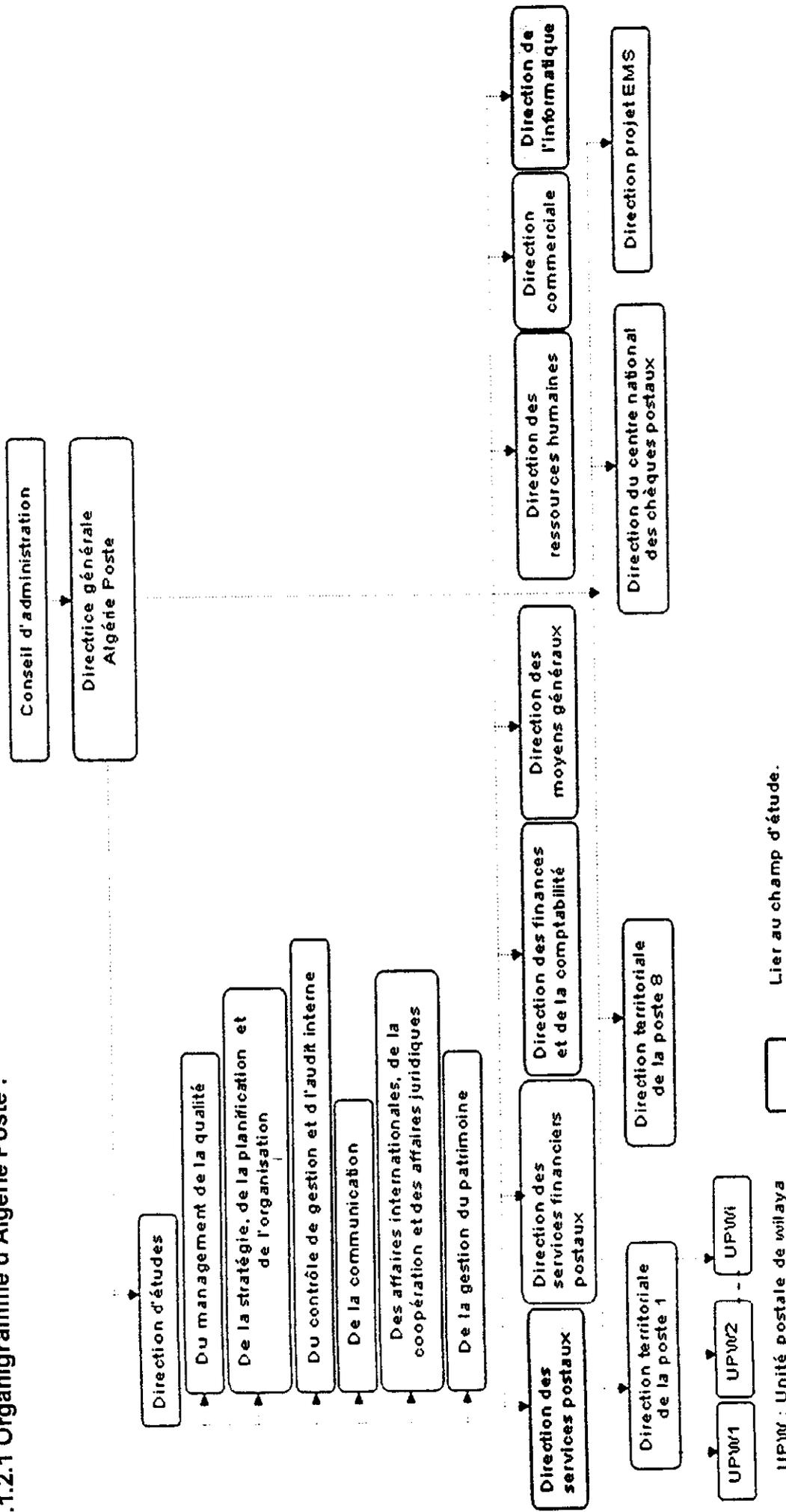


Figure I.1 : Organigramme d'Algérie Poste [RAP, 2006].

### **I.1.3 Objectif d'Algérie Poste : [RAP, 2006]**

Le programme de développement à l'horizon 2009 d'Algérie Poste s'inscrit dans le cadre d'une ambition de développement commercial, déjà marqué par les efforts de modernisation entrepris depuis sa création.

Accroître sa rentabilité, et être toujours performant, améliorer la qualité de ses prestations, assurer sa pérennité et mettre en place un réseau de proximité constituent la vision stratégique d'Algérie Poste pour les deux années à venir.

Cette vision stratégique à l'horizon 2009 vise à améliorer la compétitivité de l'entreprise par le développement de ses parts de marché, de son chiffre d'affaire et la réduction de ses charges.

L'innovation, la diversification des produits et l'amélioration de la qualité de service sont considérés comme les points-clés qui permettront à Algérie Poste de confirmer son orientation client et son engagement à le satisfaire pour bien réussir sa vision stratégique.

Dés à présent, Algérie Poste projette de développer et de moderniser ses métiers de la poste (courrier et services financiers) par l'intégration des nouvelles technologies pour mieux servir la clientèle.

## **I.2 Présentation de la structure d'accueil : La Direction Informatique**

Nous allons décrire la Direction Informatique dans ses deux dimensions organisationnelle et fonctionnelle. La première nous permettra de cerner les différents composants et les liens hiérarchiques qui existent. La deuxième dimension est celle qui nous mènera vers l'identification et la compréhension des différentes missions assignées à la direction Informatique.

### **I.2.1 Présentation fonctionnelle :**

Les missions principales des services Monétique et CCP sont la gestion des comptes des porteurs d'Algérie poste, ainsi que d'assurer les services que chacune d'entre elles offre à leur Porteurs, entre autre :

#### **I.2.1.1 Les objectifs assignés au secteur postal se résument à [RAP, 2006]:**

- Améliorer les conditions de production et de commerce ;
- Réduire les sommes immobilisées ou improductives ;
- Renforcer la sécurité des personnes par la diminution du cash en leur possession ;
- Comblent le retard accumulé dans la scripturalisation des paiements.

Plus de 2700 bureaux sont raccordés au réseau informatique permettant d'offrir les mêmes prestations à toute catégorie de clientèle et en temps réel telles que :

- Retrait, versement sur comptes CCP;
- Transfert électronique de fonds ;
- Imputation des salaires.

## **I.2.1.2 La Monétique [MON, 2004] :**

### **I.2.1.2.1 Définition de la monétique :**

La monétique désigne l'ensemble des traitements électroniques, informatiques et télématiques nécessaires à la gestion des transactions monétaires et des transferts de fonds monétaires.

Toutes les actions et traitement liés de près ou de loin aux Cartes bancaires entrent dans le cadre de la Monétique (retrait, paiement, porte-monnaie électronique, perte de carte, vol, renouvellement, litiges...)

La monétique fait partie du domaine des Transactions Électroniques Sécurisées (TES).

### **I.2.1.2.2 Les services aux Clients :**

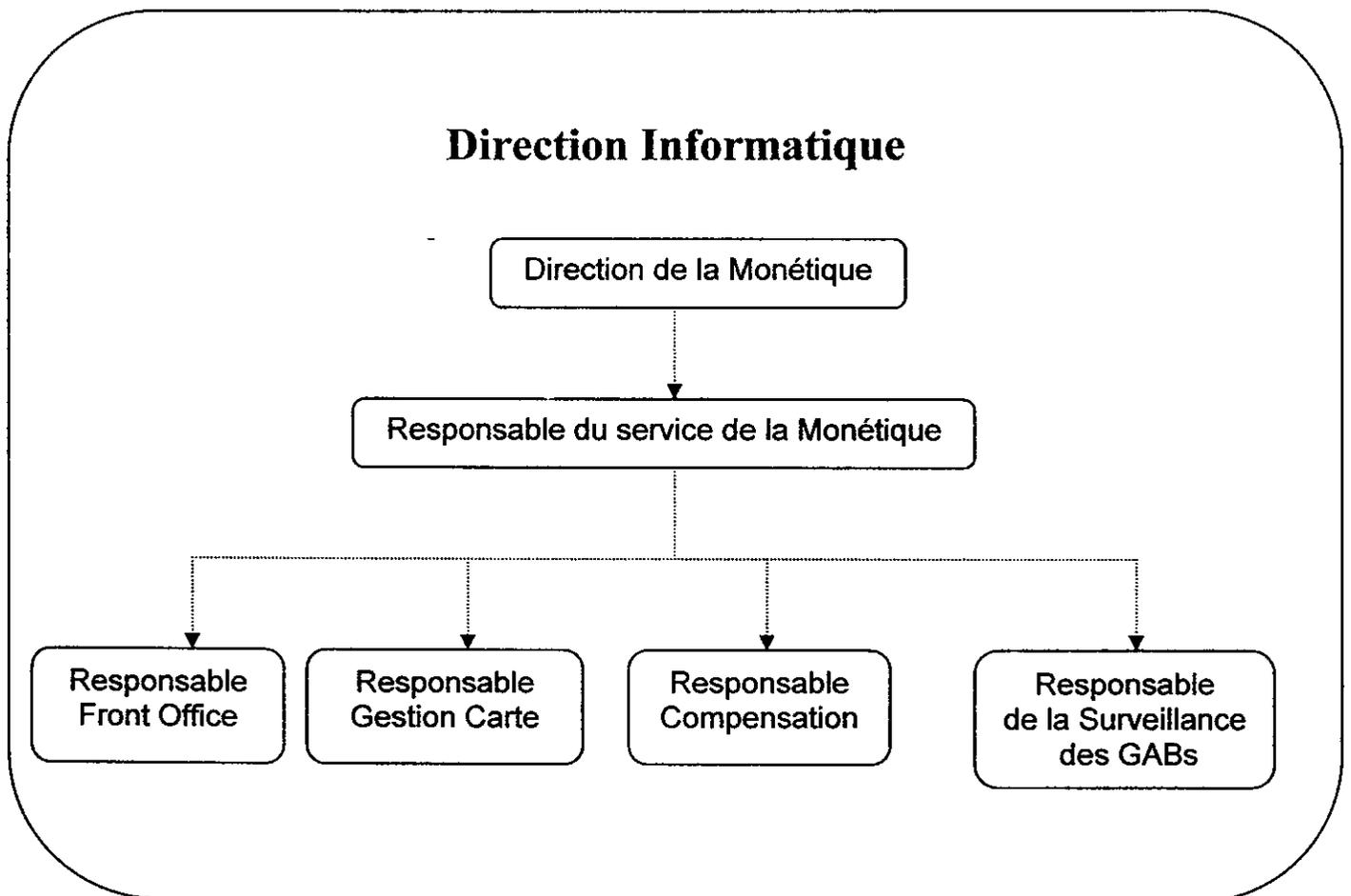
- Services aux Porteurs.
- Services aux Entreprises.
- Services aux Commerçants.
- 1. Services aux Porteurs :**
  - Retrait DAB.
  - Retrait GAB.
  - LSB (Libre Service Bancaire) :
    - Consultations : solde, dernières opérations.
    - Demande de chéquier.
    - Virement compte à compte.
    - Edition de RIB.
    - Demande du mini relevé du compte des 10 dernières opérations.
  - Paiement :
    - Paiement de Proximité
    - Paiement su TPE (Terminal de Paiement Externe)
    - ....
- 2. Services aux Entreprises :**

Consolidation des données et intégration comptable facilités

  - Gestion des frais du personnel
  - Achat non stratégique ou hors production tel que fournitures de bureau, personnel intérimaire,...
  - Gestion des achats de voyages.
- 3. Services aux Commerçants :**
  - Gestion du risque accepteur – sécurisation et garantie des paiements.
  - Télécollecte des paiements.
  - Pré-autorisation.
  - Télécommerce.
  - Fidélité.
  - Centre d'appel pour demandes d'autorisation.

L'organigramme du service Monétique est illustré à la figure I.2, et les postes qui concernent notre étude du service CCP sont illustrés dans la figure I.3.

### I.2.1.2.3 Organigramme de la Monétique :

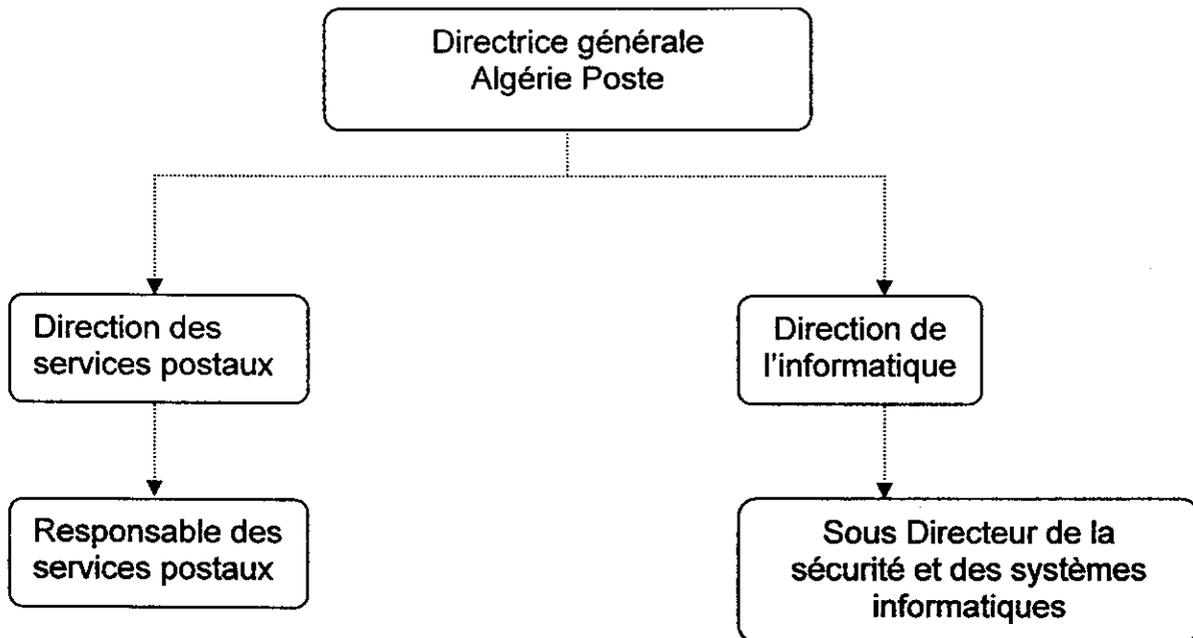


**Figure I.2 :** Organigramme de la direction Monétique. [MON,2004]

La direction de la monétique se trouve au niveau du département de la direction informatique, le Directeur de la monétique dirige à partir de la Direction générale d'Algérie poste.

Au niveau de la direction monétique, la gestion du service revient au Responsable du Service Monétique. Sous sa tutelle les différents postes cités en dessous :

- Responsable du Front Office.
- Responsable Gestion Carte.
- Responsable Compensation.
- Responsable de la Surveillance des GABs.

**I.2.1.3 Postes du service CCP concerné par l'étude :****Figure I.3 :** Organigramme des postes du CCP concerné par l'étude. [RAP, 2006]

 Poste concerné par l'étude.

Le Sous Directeur de la Sécurité et des Systèmes Informatiques (SDSSI) est, comme son nom l'indique le responsable de la sécurité du réseau et des projets informatiques, il est concerné par les indicateurs des modules CCP et de la Monétique.

Le Responsable des Services Postaux (RSP) est le responsable de la gestion du service, de la proposition, la direction et le suivi des projets postaux. Il est concerné par les indicateurs du Module CCP seulement.

**Les systèmes Décisionnels :**

**Etat de l'Art**

Dans cette partie nous donnerons les définitions indispensables à la compréhension et à la lecture du rapport.

Nous commençons par définir l'informatique décisionnelle, sa genèse, ses buts et son contexte d'application. Puis, nous passerons à la définition des tableaux de bords en mettant l'accent sur leur évolution et les défis qu'ils doivent relever actuellement dans le monde du management.

En fin, nous terminerons la partie définitions par une présentation des composants de base de l'Entrepôt de données.

## II. LES SYSTEMES DECISIONNEL : ETAT DE L'ART.

### II.1 Informatique décisionnelle [Gard, 2000]:

Loin d'être un simple phénomène de mode, l'informatique décisionnelle est devenue incontournable pour toute entreprise moderne comme le montre la figure II.4.

Les enjeux économiques et financiers des entreprises sont devenus tels que l'intuition et la réflexion ne suffisent plus à prendre une décision.

Dans un monde où la concurrence fait rage et où chaque choix stratégique peut être vital, les décideurs devraient se munir d'outils informatiques puissants et fiables visant à faciliter la tâche de pilotage de l'entreprise. Et ceci en offrant les moyens d'exploiter les données stockées dans les bases de données de l'entreprise, souvent restées au stade d'archives, et d'en extraire le contenu informationnel qui servira à la compréhension du passé ainsi qu'à une meilleure anticipation du futur.

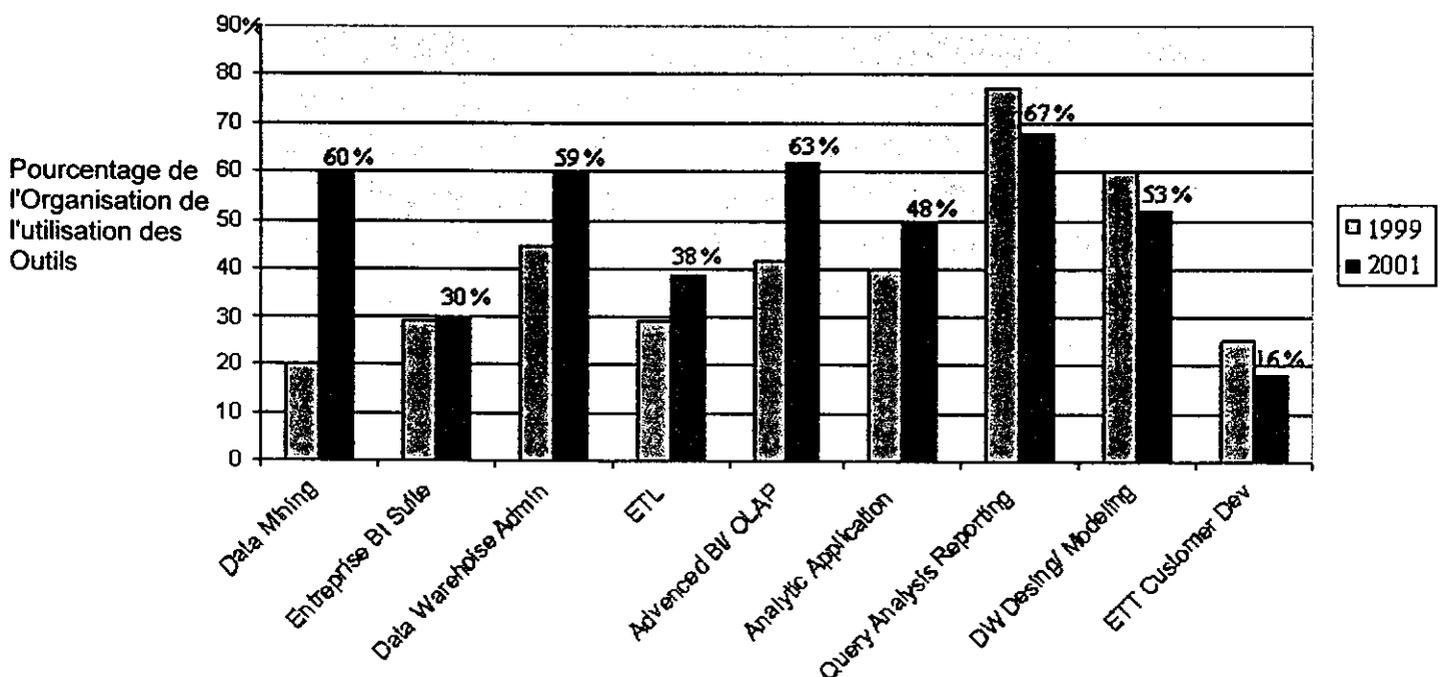


Figure II.4 : L'évolution de l'utilisation de l'informatique décisionnelle. [Card, 2000]

### II.1.1 Définition1 :

Une définition simple consiste à dire que l'informatique décisionnelle, en anglais « *Business intelligence* » est la branche de l'informatique qui permet l'exploitation des données de l'entreprise dans le but de faciliter la prise de décision par les décideurs, c'est-à-dire la compréhension du fonctionnement actuel et l'anticipation des actions pour un pilotage éclairé de l'entreprise. [Liau, 2001]

### II.1.2 Définition2 [Vand&Faul, 2001]:

Une deuxième définition plus technique veut que nous présentions en premier lieu les différents niveaux d'un système d'information et que nous définissions ce qu'est un Système d'Information Décisionnel (SID).

#### II.1.2.1 Les différents niveaux d'un système d'information d'une organisation :

Résumé dans le schéma ci-dessous, un système d'information se décompose en trois niveaux :

**1. Le niveau opérationnel :** Concerne les données relatives aux différentes fonctions de l'entreprise, il s'agit des bases de données résultant des sources d'informations internes.

**2. Le niveau décisionnel :** Constitue une synthèse des données opérationnelles, choisies pour leur pertinence. Ce niveau concerne des données qui, agrégées, intégrées sur les bases de structures particulières de stockage volumineux, résultent en *informations* pertinentes à la décision.

**3. Le niveau stratégique :** le niveau le plus élevé dans la hiérarchie, concerne l'orientation des informations résidant au niveau décisionnel en vue de fournir des indicateurs pertinents.

Ce niveau fournit au décideur d'une part, des systèmes de pilotage qui lui fournissent une série de tableaux de bord et de synthèse, très souvent enrichis de fonctionnalités statistiques et de simulation, et d'autre part sur des outils d'extraction, de gestion de connaissances qui permettent de mettre en évidence des corrélations entre des événements apparemment non liés. La finalité du niveau stratégique est le pilotage de l'entreprise dans une vision stratégique à long terme.

Les différents niveaux sont illustrés par la figure II.5 comme suit :

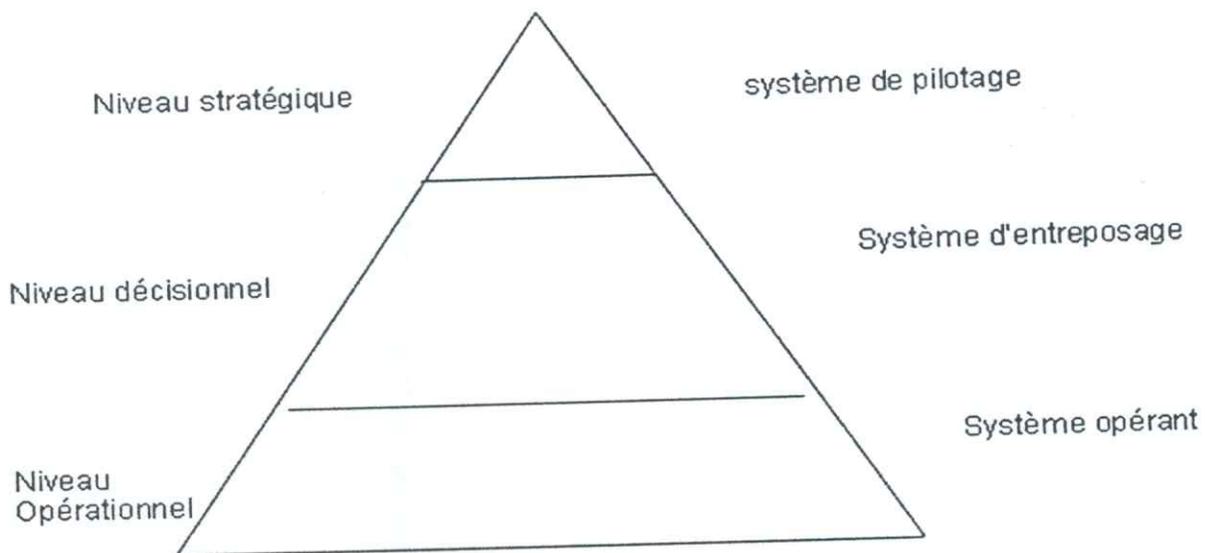


Figure II.5 : Les différents niveaux d'un système d'information. [Vand&Faul, 2001].

## II.2 Le système d'information décisionnel :

Un Système d'Information décisionnel est architecturalement la partie qui se superpose au niveau opérationnel en y ajoutant d'une part une couche de données intégrées et organisées en vue d'une prise de décision (système d'entreposage) et d'autre part, une couche d'outils spécialisés permettant de fournir une présentation intelligente de ces données (système de pilotage). Il rassemble donc les niveaux décisionnel et stratégique. [Ako&CW, 2002]

Eclairés par les définitions précédentes, nous pouvons donc donner la définition suivante de l'informatique décisionnelle :

« L'informatique décisionnelle est l'ensemble des processus qui, partant d'un système opérant développeront un système décisionnel en vue de fournir aux entreprises des outils d'aide à la prise de décision performants et agréables ». [Goua, 2001]

### II.2.1 Tableaux de bord [Vand&Faul, 2001]:

Les tableaux de bord au sens classique du terme sont les documents de référence utilisés par les utilisateurs du système d'information de décision. Ils indiquent des valeurs d'indices qui décrivent l'évolution de faits utiles au processus de décision.

De nos jours, nous désignons par tableau de bord l'interface qui vient au bout du développement d'un projet décisionnel comme étant l'unique point d'entrée visible pour le consommateur de l'information.

Les outils de type SIAD (Système Informatique d'Aide à la Décision), ou EIS (Executive Information System) présentent des tableaux de bord plus souples et plus navigationnels construits sur une base de donnée décisionnelle permettant des applications d'extraction de connaissances.

Le tableau de bord doit donner une vision cohérente et représentative de la politique de l'entreprise. Il doit permettre d'anticiper, de prévoir et signaler les dysfonctionnements. Il doit être une référence commune pour l'équipe de direction et les conseils : en ce sens, c'est aussi un moyen de communication. C'est un instrument d'action synthétique, privilégiant les représentations.

C'est aussi un ensemble de clignotants qui doivent permettre de susciter une réaction immédiate.

La construction d'un tableau de bord impose une réflexion approfondie sur les indicateurs nécessaires. C'est à chaque établissement de trouver et de combiner ses indicateurs, ce qui requiert une analyse très fine des données dont l'entreprise dispose dans son système d'informations.

En résumé les buts d'un tableau de bord sont les suivants :

- *Outil de prévision par rapport aux objectifs* : il met en évidence les résultats physiques ou financiers par rapport aux objectifs tracés qui servent de référence.

- *Outils de diagnostic* : il attire l'attention sur tout ce qui est anormal. Ainsi les responsables cherchent les causes de ces dérives et trouvent les actions pour y remédier. Le tableau de bord est en général un excellent révélateur des points faibles d'une organisation d'une société.

- *Outil de dialogue* : il doit servir d'outil de dialogue entre les différents niveaux hiérarchiques dans une entreprise.

En effet chaque gestionnaire doit commenter ses résultats, expliquer les causes des écarts et les mesures prises à son niveau.

- *Outil de performance* : il sert à informer l'encadrement des performances obtenues par d'autres départements ou l'ensemble de l'entreprise.

- *Outil de motivation des responsables* : il est judicieux d'utiliser le tableau de bord comme un miroir reflétant les performances des managers et des défis permanents qu'ils relèvent.

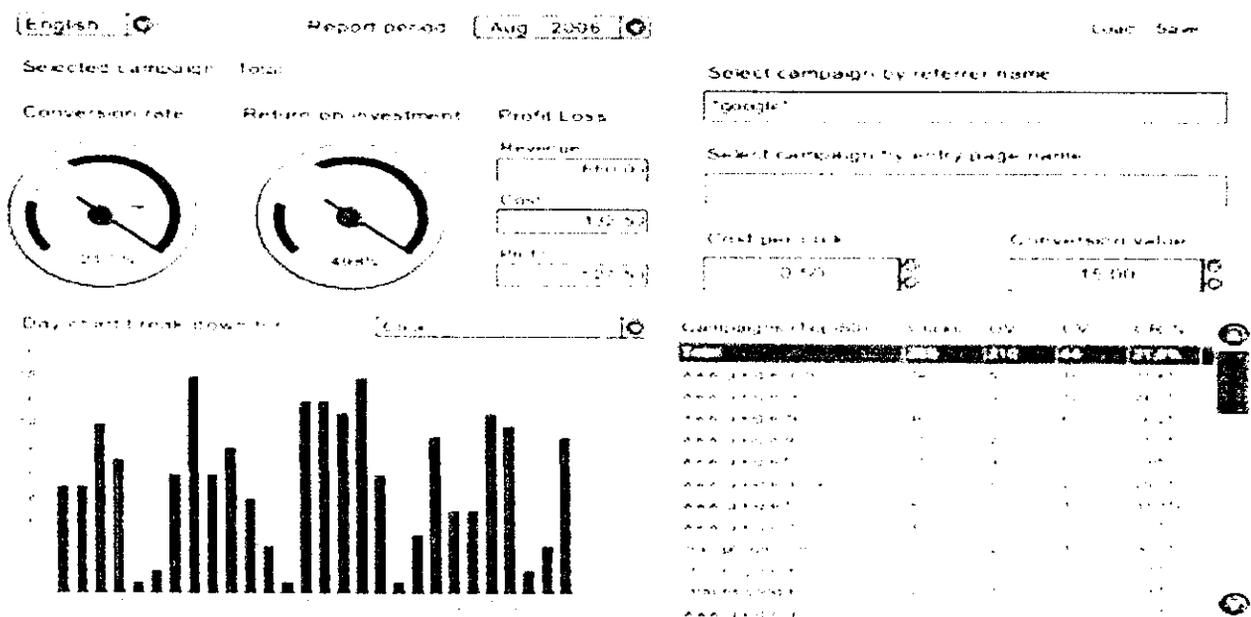


Figure II.6 : Différente facette d'un tableau de bord. [LACO, 2005]

## II.3 Composants de base de l'entrepôt de données :

### III.3.1 Système source

Système opérationnel, dont la fonction consiste à capturer les transactions liées à l'activité. Il s'agit souvent de ce que l'on appelle les applications de gestion dans l'environnement gros systèmes.

La principale priorité du système source est le temps de disponibilité. Les requêtes auprès des systèmes source sont « fondées sur un système de comptes » ; elles font partie du flux normal des transactions et sont fortement limitées dans leur sollicitation du système opérationnel. Kimball part du principe que les systèmes source maintiennent peu de données historisées et que la génération d'états est un fardeau pour ces systèmes.

Il est très important de préciser que ceux-ci ne sont pas interrogés de manières extensives et inattendues, ce qui est souvent le cas des entrepôts de données. Les systèmes source de par leur nature "vivent leur vie" (on parle de « tuyaux de poêle ») et généralement peu d'effort a été consenti pour homogénéiser les dimensions de base, telles que :

- le produit,
- le client,
- le lieu ou
- le calendrier, par rapport aux autres systèmes opérationnels de l'entreprise.

Les systèmes source intègrent des clés qui assurent l'unicité de certains éléments, comme les clés produit ou les clés client. On nomme ces clés de systèmes source des *clés de production*. [Kim&all, 2000].

### Schéma général de construction d'un Data Warehouse

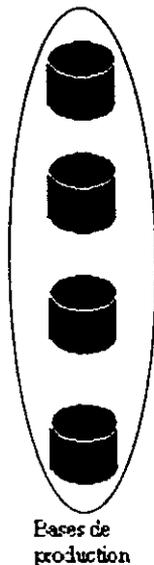


Figure II.7 : Système Source, Base de production. [LACO, 2005]

### II.3.2 Zone de préparation des données

Ensemble des processus qui :

- nettoient,
- transforment,
- combinent,
- archivent,
- suppriment les doublons,

C'est-à-dire préparent les données sources en vue de leur intégration puis de leur exploitation au sein de l'Entrepôt de données. Cette zone comprend tout ce qui se trouve entre les systèmes sources et l'entrepôt de données.

Il serait pratique que cette zone soit centralisée sur un seul environnement, mais elle est le plus souvent répartie sur plusieurs machines.

La zone de préparation des données est le théâtre d'activités simple de tri et de traitement séquentiel et n'a pas nécessairement besoin d'être basée sur les technologies relationnelles.

Toutefois, il arrive souvent que les données arrivent à la porte de la zone de préparation des données sous une forme relationnelle (troisième forme normale).

Dans d'autres cas, les responsables de la zone de préparation se sentent plus à l'aise s'ils organisent les étapes de nettoyage, de transformation et de combinaison dans un contexte de structures normalisées.

La frontière qui détermine le contenu de la zone de préparation est que *la zone de préparation des données ne doit offrir ni services de requête, ni service de présentation*. [Kim&all, 2000].

#### Schéma général de construction d'un Data warehouse

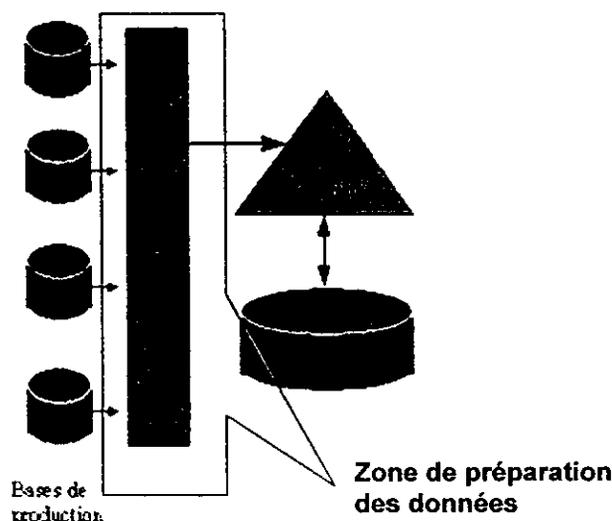


Figure II.8 : Zone de préparation des données. [LACO, 2005].

### II.3.3 Entrepôt de données

Source de données interrogeables de l'entreprise. C'est tout simplement l'union des data marts qui le composent.

L'entrepôt de données est alimenté par la zone de préparation des données.

Précisément l'entrepôt de données est la ressource de présentation *interrogeable* des données d'une entreprise et elle *ne doit pas* être organisée autour d'un modèle entité relation, qui lui ferait perdre sa clarté et ses performances.

En outre l'entrepôt de données est *fréquemment mis à jour* en tenant compte de la charge au fur et à mesure que les données sont corrigées, que les instantanés (*Vue matérialisé*) s'accumulent et que les statuts et les libellés évoluent.

Les données de l'entrepôt de données sont non volatiles ce qui signifie qu'une donnée entrée dans l'entrepôt l'est pour de bon et n'a pas vocation à être supprimée ;

Les données l'entrepôt de données doivent être historisées, donc datées. [Kim&all, 2000].

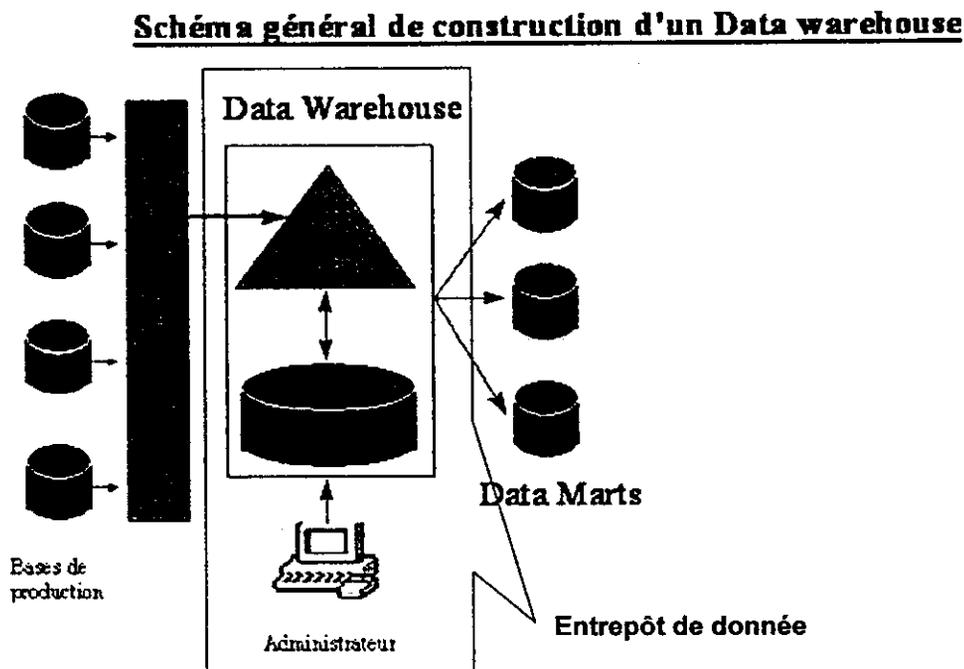


Figure II.9: Entrepôt de donnée. [LACO, 2005].

### II.3.4 Data mart (Magasin de données) :

Sous-ensemble logique d'un Entrepôt de données. Si l'on compare l'entrepôt de données à un camembert, alors un data mart en est une portion.

Il représente un projet réalisable. Au-delà de cette définition relativement simple, on considère le data mart comme la réduction de l'entrepôt de données à un seul processus ou à un groupe de processus, ciblant un groupe métier spécifique.

Un data mart donné sera généralement sponsorisé par un département ou une unité en particulier et, le plus souvent, il sera organisé autour d'un processus précis.

Chaque data mart doit être représenté par un modèle dimensionnel et, au sein d'un entrepôt de données, tous ces data marts doivent être construits à partir de dimensions conformes et de faits conformes. Ceci est à la base de l'architecture en bus décisionnel (Data Warehouse Bus Architecture).

En l'absence de dimensions conformes et de faits conformes, un data mart reste un simple tuyau de poêle, ce qui est le fiéau de l'Entrepôt de données. Si on espère construire un entrepôt de données à la fois robuste et réactif vis-à-vis d'exigences en constante évolution, il est préférable d'adhérer à une architecture en bus (Architecture horizontal).

Lorsque toutes les pièces qui composent un data mart sont décomposées en tables physiques réparties sur plusieurs serveurs de base de données, ce qui est leur configuration finale, la manière physique de combiner les données de ces différentes tables et d'implémenter un entrepôt de données intégré dépend du fait que les dimensions des données représentent la même chose d'une table à l'autre ou pas. C'est ce que l'on appelle des dimensions conformes. Cette architecture en bus est un élément fondamental.

Les data marts sont basés sur des données détaillées et peuvent contenir des résumés (données consolidées) destinés à optimiser les performances, que nous appellerons agrégats. [Kim&all, 2000].

### Schéma général de construction d'un Data warehouse

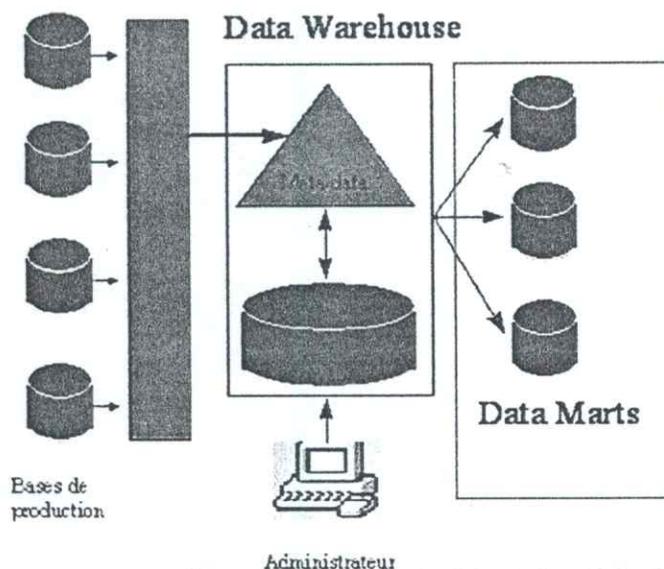


Figure II.10: Data Mart. [LACO, 2005].

### II.3.5 Operational Data Store (ODS):

**II.3.5.1 Définition 1:** Un ODS ("Operational data store") est une base de donnée conçue pour centraliser les données issues de sources hétérogènes afin de faciliter les opérations d'analyse et de reporting. L'intégration de ces données implique souvent une purge des informations redondantes. Un ODS est généralement destiné à contenir des données de niveau fin comme par exemple un prix ou le montant d'une vente, en opposition aux données agrégées tel que le montant total des ventes. Les données agrégées sont stockées dans un entrepôt de données. [Kim&all,2000].

### II.3.5.2 Définition 2 :

Un stockage de données partagées entre applications. Pivot de l'alimentation du système d'information décisionnel en le préservant de toutes les évolutions du système de production (ou système opérationnel). Il centralise la collecte des informations issues des applications de gestion, les transforme et les valorise avant de les injecter dans l'Entrepôt de données. L'ODS gère aussi les rejets ainsi qu'un journal d'opérations pour que l'administrateur puisse prendre en compte et corriger les anomalies rencontrées. [Kim&all,2005].

L'ODS s'appuie généralement sur une solution progiciel de type ETL (*Extraction, Transformation, Loading*) et sur une base de données relationnelle permettant de structurer les tables de référence et de travail nécessaires à son fonctionnement. [LACO, 2005].

L'ODS se définit comme un sas de validation placé en amont de l'Entrepôt de données. Il assure les fonctions suivantes : [LACO, 2005].

- **Préservation du système décisionnel** des évolutions des systèmes de production. De telle sorte qu'une modification dans les structures de données des applications amont n'influe pas directement sur la disponibilité de l'Entrepôt de données.
- **Recensement et synchronisation des flux d'alimentation de l'entrepôt** : Le système décisionnel est un réservoir de données orienté métier/sujet. Il est alimenté à partir de plusieurs flux qui ne sont pas obligatoirement disponibles au même instant. En revanche, l'entrepôt ne peut être alimenté qu'à partir d'un ensemble cohérent d'informations. L'ODS assure le "transit" des données de la production vers l'Entrepôt de données.
- **Historisation des flux prélevés sur les systèmes sources.**
- **Transcodification et contrôle de cohérence** des différents flux entrants.
- **Gestion des rejets et des statistiques sur les données.**
- **Alimentation de la structure de données dimensionnelle** : cela regroupe l'ensemble des procédures de préparations des données : constitution de la table de faits et des dimensions.
- **Constitution des agrégats** : procédures de calculs des données agrégées.
- **Purge et archivage** des données de détails.
- **Préparation des données dans le cas de l'utilisation d'un ou plusieurs outils de Data Mining.**
- **Gestion d'un référentiel complémentaire.** Dans certains cas, il est nécessaire de constituer quelques tables de référence en complément des données de référence extraites des systèmes de production (Ex : l'application amont ne gère que des codes sans correspondance avec des libellés).

Dans ce qui suit, une figure qui résume les différentes phases d'un système d'information décisionnel, et ou l'interaction ODS et Entrepôt de données y est illustré. Le schéma est illustré par la figure II.11 comme suit :

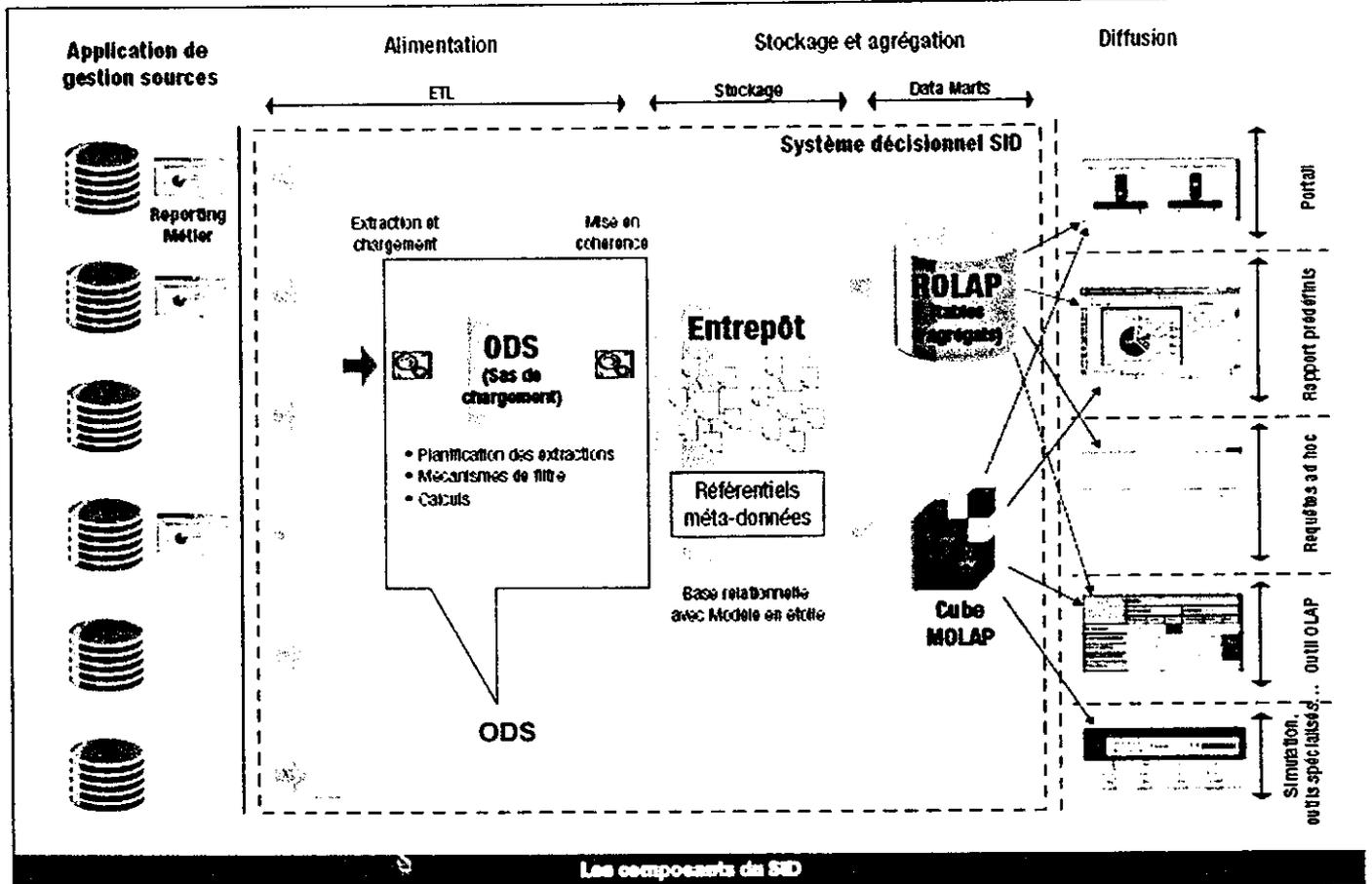


Figure II.11 : les composant du système d'information décisionnel. [LACO, 2005].

### II.3.6 OLAP (Online Analytic Processing):

Activité globale de requêtage et de présentation de données textuelles et numériques contenues dans l'entrepôt de données ; style d'interrogation et de présentation spécifiquement dimensionnel, qui se concrétise par un certain nombre de « vendeurs OLAP ».

La technologie OLAP est non relationnelle et presque toujours basée sur un cube de données multidimensionnel explicite.

Les bases de données OLAP sont également connues sous le terme de bases de données multidimensionnelles. Toutefois, au sein de l'éventail complet des applications de Data Warehouse, on pourrait plutôt les apparenter à de petits data marts indépendants.

Kimball pense que les data marts de type OLAP peuvent être des composants à part entière du bus décisionnel s'ils sont conçus autour de faits conformes et de dimensions conformes. [Kim&all, 2000].

**1. ROLAP (Relational OLAP)**

Ensemble d'interfaces utilisateurs et d'applications qui donnent une vision dimensionnelle des bases de données relationnelles.

**2. MOLAP (Multidimensional OLAP)**

Ensemble d'interfaces utilisateur, d'applications et de technologies de base de données propriétaires dont l'aspect dimensionnel est prépondérant.

**3. HOLAP (Hybrid OLAP) :** est un hybride entre ROLAP et MOLAP. Les parties tables de faits et tables de dimensions sont stockées dans une base relationnelle standard tandis que le reste des données (les calculs) sont stockés dans une base multidimensionnelle.

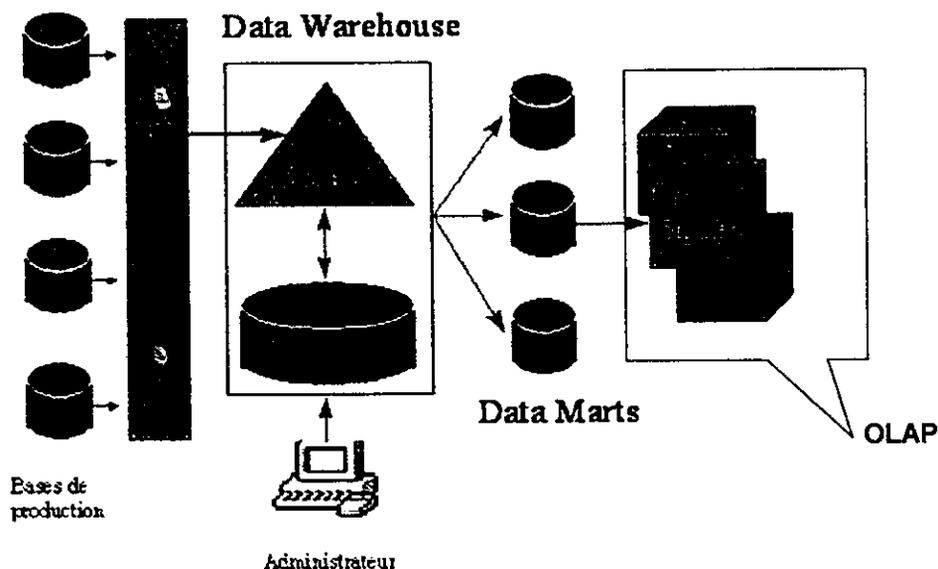
**Schéma général de construction d'un Data warehouse**

Figure II.12: OLAP [LACO, 2005].

**II.3.7 Application utilisateur**

Ensemble d'outils qui interrogent, analysent et présentent des informations répondant à un besoin spécifique. [Kim&all, 2000].

L'ensemble d'outils minimal se compose :

- d'un outil d'accès aux données,
- d'un tableur,
- d'un logiciel graphique et
- d'un service d'interface utilisateur, qui suscite les requêtes et simplifie la présentation de l'écran aux yeux de l'utilisateur.

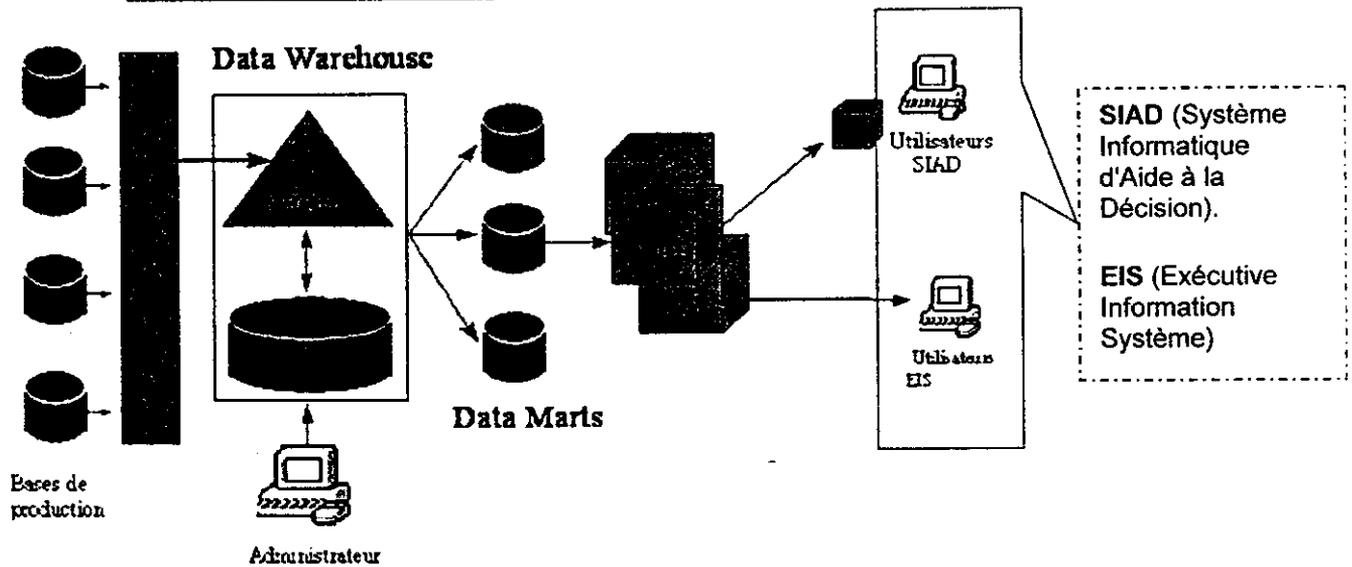
**Schéma général de construction d'un Data Warehouse**

Figure II.13 : Application utilisateurs. [LACO, 2005].

### II.3.8 Outil d'accès aux données

Client de l'entrepôt de données. Dans un entrepôt de données relationnel, ce client maintient une session auprès du serveur de présentation, envoyant un flux de requêtes SQL distinctes au serveur.

Au final, l'outil d'accès entre dans le cadre de la session SQL et sert à présenter un écran affichant des données peut être aussi simple qu'un outil de requête ad hoc ou aussi complexe qu'une application de data mining ou de modélisation élaborée.

Les outils d'accès aux données les plus sophistiqués sont les outils de modélisation ou les outils prévisionnels pouvant charger leurs résultats dans des zones spéciales de l'entrepôt. [Kim&all, 2000]

### II.3.9 Outil de requête :

Type spécifique d'outil d'accès aux données qui invite l'utilisateur à formuler ses propres requêtes en manipulant directement les tables relationnelles et leurs jointures.

Les outils de requête ad hoc, tout puissants qu'ils soient, ne peuvent être exploités efficacement et maîtrisés que par environ 10 % des utilisateurs finaux d'un entrepôt de données.

Les 90 % doivent être servis par des applications pré configurées, qui ressemblent à des « Modèles » finalisés ne demandant pas à l'utilisateur de construire directement une requête relationnelle.

Les outils ad hoc les plus orientés ROLAP portent à 20 % d'utilisateurs potentiels. [Kim&all, 2000]

**Schéma général de construction d'un Data Warehouse**

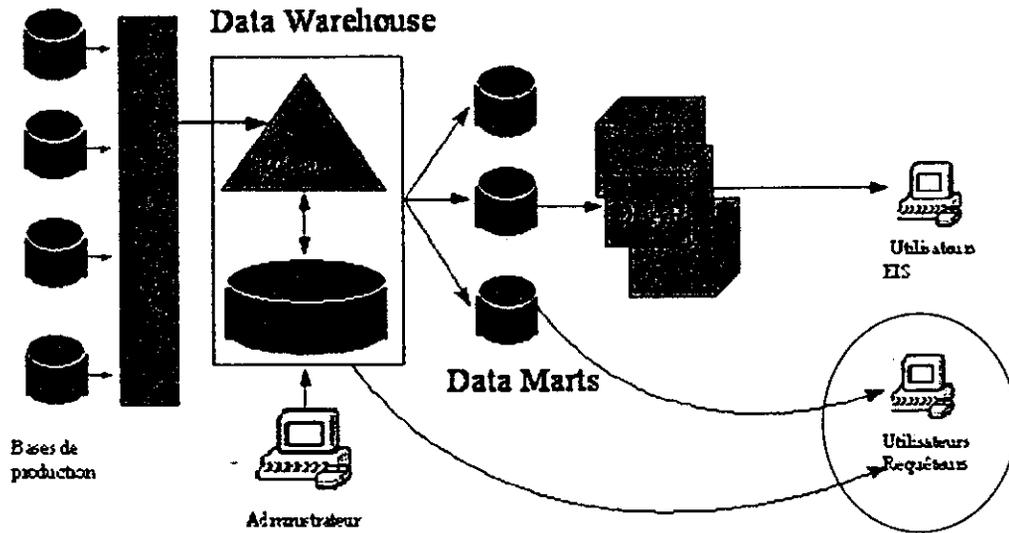


Figure II.14 : Outil de requêtage. [LACO, 2005].

**II.3.10 Méta-données :**

Méta-donnée signifie concrètement « des données sur des données ». Une méta-donnée est une donnée qui a pour but de décrire une autre donnée.

Les méta-données peuvent être des informations complémentaires, nécessaires à la compréhension d'une autre information ou dans le but de permettre une utilisation pertinente.

L'un des grands principes de l'entrepôt de données est de conserver la trace des données produites, après agrégation, consolidation et application des règles de gestion.

Cette fonction est assurée par les méta-données, c'est à dire les « données sur les données ». Dans ces méta-données seront stockées des informations telles que le nom de la base de production dont la donnée est extraite, la date et l'heure de la dernière extraction, la fréquence de mise à jour de cette information, ... [LACO, 2005].

**Schéma général de construction d'un Data Warehouse**

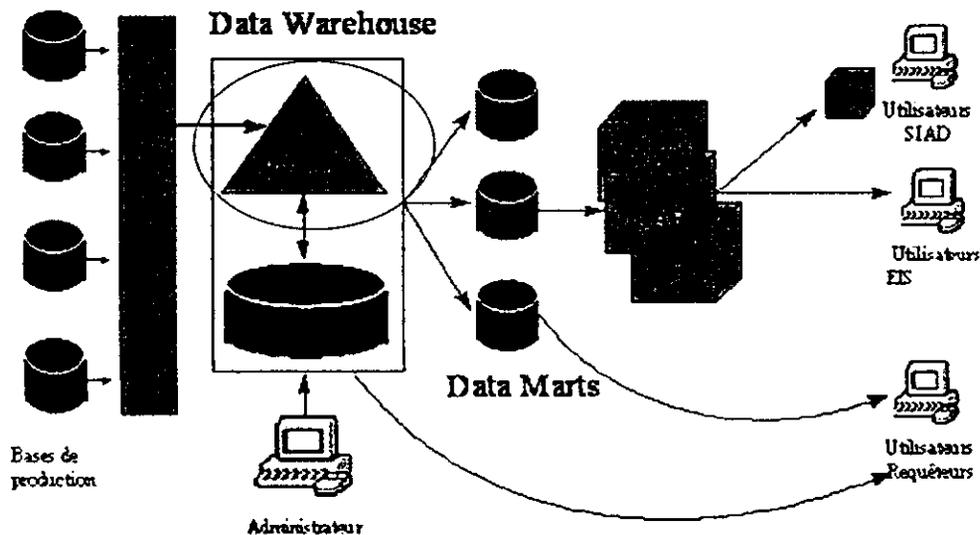


Figure II.15 : Les outils de l'entrepôt de données : Les Méta-donnée. [LACO, 2005].

### II.3.11 Processus de base de l'entrepôt de données:

La préparation des données est un processus essentiel qui comprend, entre autre, les sous processus suivants : [Kim&all, 2000].

- **Extraction** : il s'agit de la première étape de récupération des informations dans l'environnement de l'entrepôt de données. L'extraction comprend la lecture et la compréhension de la source de données, ainsi que la copie des parties nécessaires à une exploitation ultérieure dans la zone de préparation.
- **Transformation** : Une fois les données extraites dans la zone de préparation des données, on rencontre plusieurs étapes de transformation :
  - *Nettoyage* des données, à savoir correction des fautes d'orthographe, résolution des conflits de domaine (noms de villes incompatibles avec le code postal, par exemple), résolution des cas d'informations manquantes et conversion en format standard.
  - *Purge* de certains champs du système source qui sont inutiles à l'entrepôt de données.
  - *Combinaison* des sources de données par mise en correspondance exacte avec des valeurs clés ou par mise en correspondance approximative d'attributs hors clé, y compris la recherche d'équivalents textuels des codes des systèmes sources.
  - *Création des clés de substitution* pour chaque enregistrement dimensionnel afin d'éviter de dépendre des clés définies dans le système source.  
Dans ce cas, c'est le processus de génération des clés qui assure l'intégrité référentielle entre les tables dimensionnelles et les tables des faits.
  - *Construction d'agrégats* pour optimiser les performances des requêtes les plus courantes.
- **Alimentation et indexation** : À la fin du processus de transformation, les données prennent la forme d'image d'enregistrement prêt à être chargées. En règle générale, le processus de chargement des données dans l'entrepôt s'effectue par réplique des tables des faits et des tables dimensionnelles, qui seront ensuite présentées aux services de chargement en masse de chaque data mart destinataire. Le data mart de destination doit alors indexer les données afin d'optimiser les performances de requête.
- **Contrôle qualité** : Lorsque chaque data mart a été chargé, indexé et complété des agrégats appropriés, la dernière étape avant la diffusion des données est celle de l'assurance qualité. Toutes les données doivent avoir été chargées, et l'ensemble des totaux et autre calculs doivent être cohérents par rapport aux données déjà enregistrées. L'état des exceptions est le plus souvent construit à l'aide du générateur d'états de l'utilisateur final du data mart.

- **Diffusion** : Une fois que chaque data mart a été chargé et que sa qualité est approuvée, les utilisateurs doivent être informés de la disponibilité des données. La diffusion fait en outre état des éventuels changements subis par les dimensions sous-jacentes et des nouvelles hypothèses qui ont été introduites dans les faits mesurés ou calculés.
- **Mise à jour** : Les data marts actuels peuvent être mis à jour, même fréquemment. Il est évident que les données erronées doivent être corrigées. Les changements d'intitulés, de hiérarchie, de statut et même d'actionariat de l'entreprise déclenchent souvent des modifications des données initiales stockées dans les data marts qui composent l'entrepôt de données.
- **Requêtes** : L'interrogation des données englobe toutes les activités qui consistent à demander une information à un data mart, y compris les requêtes ad hoc effectuées par les utilisateurs finaux, la génération d'états et les applications complexes d'aide à la décision. Il est évident que les requêtes font tout l'intérêt de l'utilisation d'un entrepôt de données.
- **Réinjection des données** : il existe deux emplacements importants à partir desquels les données « remontent », à l'inverse des flux traditionnels. Tout d'abord, il est possible de recharger une description dimensionnelle nettoyée depuis la zone de préparation, à destination du système opérationnel. Par ailleurs, on peut recharger les résultats d'une requête complexe, d'un modèle ou d'une analyse de data mining en direction du data mart. La méthode semble naturelle pour conserver la valeur d'une requête complexe, qui restitue de nombreuses lignes et colonnes que l'utilisateur souhaite sauvegarder.
- **Surveillance** : il est parfois essentiel de connaître la provenance des données et les calculs qui ont été effectués. Ces enregistrements sont directement liés aux données d'origine, de manière que l'utilisateur puisse interroger à tout moment les enregistrements de surveillance, autrement dit le « pedigree » des données.
- **Sécurisation** : L'équipe du Data Warehouse doit désormais inclure un nouveau membre doté de compétences approfondies : l'architecte de la sécurité de l'entrepôt de données. La sécurité doit être gérée de manière centralisée, depuis une console unique. Les utilisateurs doivent être autorisés à accéder à tous les data marts composant l'entrepôt de données par une identification unique.
- **Sauvegarde et restauration** : Dans la mesure où les données de l'entrepôt représentent un flux provenant des systèmes opérationnels, se dirigeant vers les data marts et terminant leur route sur l'écran de l'utilisateur, une question se pose : où prendre les historiques indispensables à l'archivage et à la récupération en cas de désastre ? En outre, il peut être encore plus compliqué de sauvegarder et de restaurer tous les méta-données qui huilent les rouages de l'Entrepôt de données.

### II.3.12 Modèle dimensionnel :

Discipline de modélisation des données qui se positionne comme solution de rechange à la modélisation entité/relation. [Kim&all, 2000].

Un modèle dimensionnel contient les mêmes informations qu'un modèle entité/relation, mais présente les données dans un format asymétrique dont les objectifs sont :

- la bonne compréhension de l'utilisateur,
- la performance des requêtes et
- l'adaptation au changement.

Il y a trop d'entrepôts de données qui se sont effondrés à cause de conceptions relationnelles trop complexes.

Les techniques de modélisation dimensionnelles ont été appliquées avec succès dans des centaines de situations au cours des quinze dernières années.

Les principaux composants d'un modèle dimensionnel sont les tables des faits et les tables dimensionnelles, dont il s'agit :

#### II.3.12.1 Table de fait :

La table de fait est la table principale de tout modèle dimensionnelle destinée à héberger des données permettant de mesurer l'activité (les mesures).

Une table de faits est une table qui contient les données à analyser. Elle est distinguée grâce à sa grande taille et à son emplacement au centre du schéma.

La table de faits contient les clés des différentes dimensions qui sont reliées à elle ainsi que les mesures de l'activité. Ces derniers sont généralement numériques additifs et valorisées de façons continue.

La clé primaire de la table de faits est la concaténation des clés primaires de toutes les tables dimensions reliées à elle.

Structure de base d'une table de fait

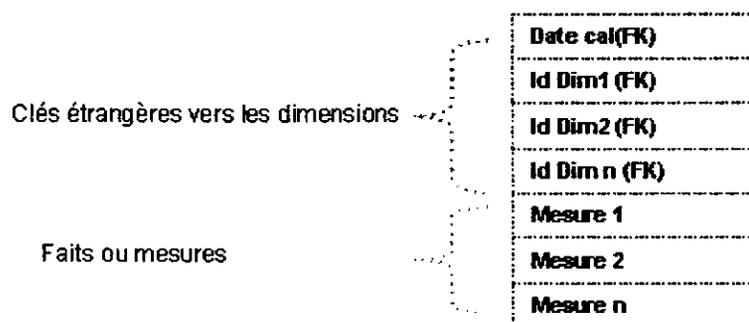


Figure II.16 : Structure de base d'une table de fait. [Ako&CW, 2002].

### II.3.12.2 Les tables de faits sans faits :

Une table de faits sans faits sert à modéliser une activité utile au décideur qui n'a pas de faits mesurables. On trouve deux type de tables de faits sans faits : table de suivi d'événement et table de recouvrement.

1. **Table de suivie d'événement** : Sert à modéliser une activité dont les seules mesures sont obtenues par comptage sur les dimensions. Une table de suivi d'événement relie les dimensions participant à l'événement, Aucune autre tâche que celle de dénombrer les événements n'est demandée. La table des faits ne comporte alors que des clefs étrangères vers les tables de dimension.
2. **Table de couverture** : La table de couverture est une table d'événement qui n'a pas eu lieu, il est alors difficile de dénombrer les événements qui n'ont pas eu lieu. Par exemple quel article en promotion n'a pas été vendu ? La solution est de construire une **table de couverture** qui mémorise un booléen. Toutes les clefs sont représentées, donc tous les produits apparaissent. Reste le problème de la taille potentielle de cette table. Pour cela, on peut limiter la présence de fait à un sous ensemble des produits ou clients concernés...

### II.3.12.3 Les tables de dimension :

Ce sont les tables qui entourent la table de fait dans le schéma en étoile. Petites par rapport à la table de faits, une table de dimension contient une clé et des attributs généralement textuels ou numériques discrets plus ou moins statique décrivant une dimension de l'activité.

Chaque dimension est définie par sa clé primaire, qui assure l'intégrité référentielle avec la ou les tables des faits à laquelle elle est liée.

La plupart des tables dimensionnelles contiennent de nombreux attributs textuels (champs), sur lesquels portent les clauses de conditions et de groupement au sein des requêtes.

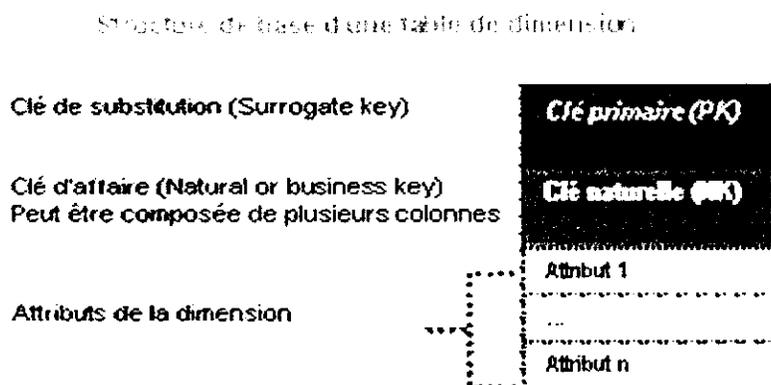


Figure II.17 : Structure de base d'une table de dimension. [Ako&CW, 2002]

Une dimension est un ensemble de valeurs décomposables. Les valeurs d'une dimension sont généralement organisées en hiérarchie.

**Une hiérarchie** : est une décomposition d'une dimension en niveau, afin de permettre à l'utilisateur d'examiner ses indicateurs à différents niveau de détail, allant du niveau global au niveau le plus fin. On aura alors une vision pyramidale des données. La base de la pyramide représentant le niveau le plus détaillé, le plus haut est le niveau le plus global.

**Remarque :**

- Une dimension peut avoir plusieurs hiérarchies.
- L'attribut d'une dimension peut appartenir à plusieurs hiérarchies ou à aucune.
- **Exemple de hiérarchie** : Dans la dimension *Temps*, on peut avoir la hiérarchie suivante : année, mois, semaine, jour, heure.
- **La dimension temps** : la dimension temps est la seul qui figure systématiquement dans tout entrepôt, elle peut enregistrer des dates étalées sur plusieurs années, elle sert a faciliter la gestion du champ date en permettant de décrire les week-ends les jour fériés et autre découpage calendriers.
- **Les dimensions dégénérées** : il s'agit d'une clé de dimension qui n'a pas d'attribut, et donc n'a pas de table de dimension.

## II.4 Conclusion

Nous avons exposé tous les composants de l'environnement de l'Entrepôt de données et expliqué leurs interactions.

Dans l'un des prochains chapitre qui suivent, nous aborderons le cycle de vie dimensionnel, qui constitue le cadre de réflexion du mémoire.

**Etude de l'existant**

Après avoir présenté quelques notions sur le domaine décisionnel, nous passons maintenant à la partie étude de l'existant qui concerne l'identification et la description des postes de décisions, de l'accès ainsi que l'acquisition des données concernées par l'étude.

### **III. ETUDE DE L'EXISTANT :**

#### **III.1 Description de l'organisation :**

Le cadrage du projet par rapport à l'organisation globale de l'entreprise est le suivant :

##### **III.1.1 La structure concernée :**

Direction de la Monétique affiliée à la Direction Informatique (voir figure I.2).  
Direction des Services Postaux affiliée à la Direction générale d'Algérie Poste (voir figure I.1).

#### **III.2 Les postes de décisions :**

Nous énumérons par ordre hiérarchique les postes de prise de décision.

##### **III.2.1 Direction Monétique :**

- 1- Directeur de la direction monétique.
  - 2- Responsable du Service monétique.
- Ensuite vient au même niveau hiérarchique :
- 3- Responsable Front Office.
  - 4- Responsable Carte.
  - 5- Responsable Compensation.
  - 6- Responsable Surveillance des GABs.

##### **III.2.2 Service Postal :**

- 1- Directeur Service Postal.
- 2- Sous Directeur de la Sécurité et des Systèmes Informatiques.

#### **III.3 Description du système CCP :**

Le système opérationnel d'Algérie poste est un système fermé, toutes les transactions effectuées au niveau des bureaux de postes y parviennent à travers un réseau centralisé relié par les liaisons X25, aucune mise à jour ou modification ne peut être effectuée directement sur la base.

Quotidiennement, les données passent du système opérationnel à une base d'archivage, où ils y sont stockés pour une période de 15 ans.

La base d'archivage contient toutes les transactions journalières effectuées par les clients concernant les Comptes CCP.

La base d'archivage est installée sur le SGBD Oracle 10 G.

La transition des données de la base opérationnelle vers la base d'archivage s'effectue par des fichiers plats.

Les fichiers plats se chargent dans la base quotidiennement. Les principaux fichiers générés sont :

- Information administrative CCP.
- Mouvement CCP.

Le logiciel installé à leur niveau permet d'effectuer les opérations suivantes :

- Recherche multicritère.
- Affichage.
- Etablir les relever.

#### **III.4 Description du système monétique :**

Le système adapté est un ERP (Entreprise Resource Planning), sur la demande de la direction informatique pour permettre une meilleure gestion du service monétique.

Nous allons présenter les objectifs et les fonctionnalités principales du système. Cette étape, dans l'étude de l'existant, nous permettra de mieux cerner les capacités du système et nous fournira les informations indispensables à l'évolution du projet.

Le système Monétique gère toutes les opérations qui concernent les porteurs, les commerçants, les pannes et les compensations d'Algérie poste avec les autres banques algériennes.

##### **III.4.1 Les constituant du processus monétique : [MON, 2004].**

- La carte (piste / puce).
- Le front office (authentification, autorisation,...).
- Le Back office (compensation, oppositions...).
- Les terminaux de paiement. (TPE).
- Les GAB (applications clients).

##### **III.4.2 Les intervenants dans une transaction : [MON, 2004].**

- Le porteur (souscrit un contrat carte).
- L'émetteur (banque du porteur).
- L'acquéreur (acquéreur du GAB).
- L'accepteur (est lié à une banque acquéreur, commerçant, DAB, GAB).

##### **III.4.2.1 Transaction de retrait :**

Nous illustrons la transaction de retrait par un schéma, représenté par la figure III.18 comme suit :

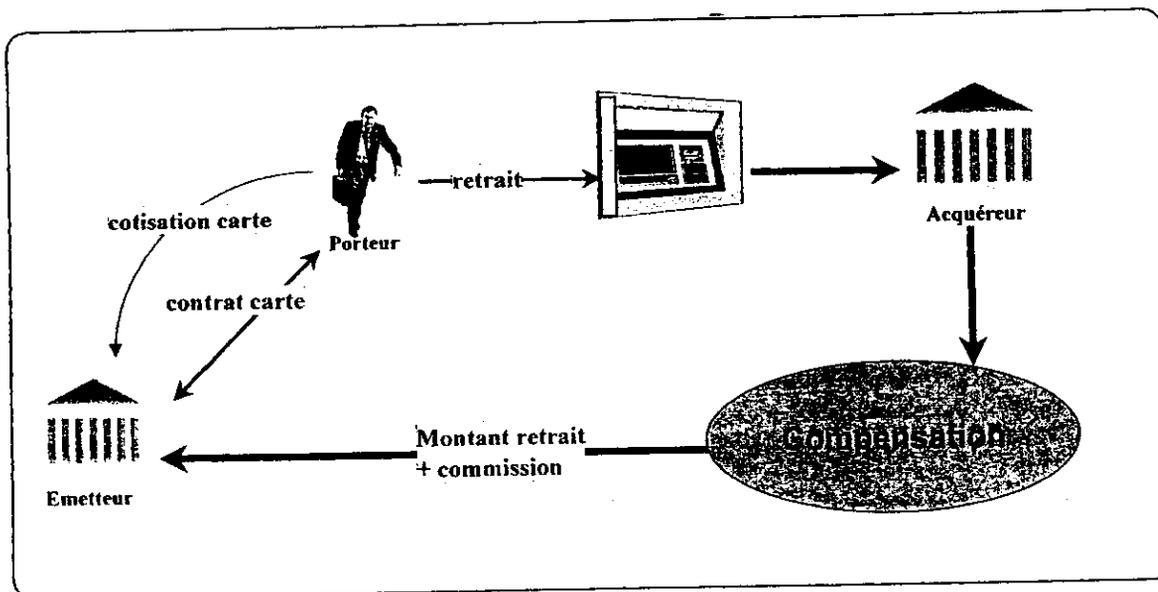


Figure III.18 : Transaction de retrait par Carte. [MON, 2004].

#### III.4.2.2 Transaction de paiement :

Nous illustrons la transaction de paiement par un schéma, représenté par la figure III.19 comme suit :

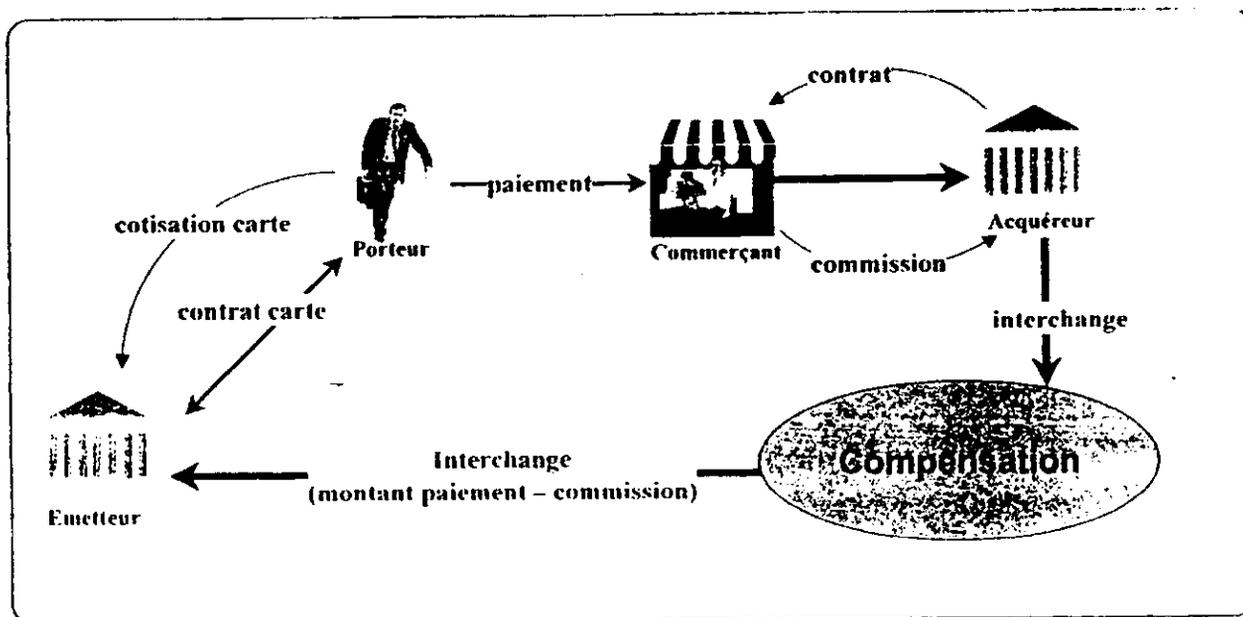


Figure III.19 : Transaction de paiement par Carte. [MON, 2004].

### III.4.3 Les modules du système :

Le système a été conçu d'une façon modulaire où chaque module prenant en compte une activité spécifique de la monétique.

1. *Module de base* : Il permet de renseigner les tables de base nécessaires pour le fonctionnement du système, telle que : Clients, Comptes, Types Comptes ...etc.

2. *Module Activité carte* : concerne les transactions effectuées par les porteurs d'Algérie poste en domestique (transaction de retrait sur les GAB d'AP) et en nationale (transaction de retrait sur les commerçants des banques confrère), activité carte gère tout ce qui est administration et gestion des cartes :

- Comportement de la carte.
- Exception.
- Autorisation, toutes les demandes d'autorisations domestiques.
- Chiffre d'affaire annuel, représenté graphiquement pour suivre l'évolution de l'utilisation de la carte.

3. *Module Commerçant* : concerne les transactions effectuées sur les commerçants de l'acquéreur Algérie poste (GAB), ce module gère :

- Le chiffre d'affaires et commissions sur les commerçants.
- Exception.
- Autorisation.
- Evolution annuelle.

4. *Module Activité Emetteur* : gère et suit le comportement des porteurs détenteurs des cartes Algérie poste. Ce module gère :

- Les transactions reçues et traitées.
- Les transactions en anomalies.
- Les annulations des transactions reçues.
- Le suivi des cartes en circulations.
- Les exceptions traitées et non traitées.

5. *Module acquéreur* : gère et suit l'évolution et le comportement de ces commerçants (GAB). Ce dernier comporte :

- Chiffre d'affaire accompli par ces commerçants.
- Le comportement de ces commerçants (Commerçant en liste noire, en panne ...etc.).
- Les exceptions produites sur ces commerçants.

6. *Module Compensation* : gère et suit les compensation inter bancaire entre Algérie poste et les banques confrères d'Algérie. Ce dernier comporte :

- La gestion et le suivi des compensations par type de transaction (Paiement, Retrait).
- La gestion et le suivi des litiges concernant la compensation.
- La régularisation des litiges par la vérification et le traitement des transactions en anomalies.

7. *Module Panne* : gère et suit le comportement des GAB face aux diverses pannes qui surgissent. Ce module gère :

- Le suivi et la supervision de tous les GAB.
- Afficher et annoncer les différents sites a problèmes ou en souffrances.

#### iii.4.4 Les tableaux de bord du système Monétique :

La partie tableau de bord est en fait la partie IHM du logiciel. Elle donne accès aux données détaillées du système opérationnel sans aucune signification de l'information quelle représente, en effet un directeur n'a pas à savoir que telle carte a retiré tant d'argent a telle heure. Ce qui l'intéresse c'est de savoir s'il a atteint un tel chiffre d'affaire à telle ville sur tel commerçant.

Le tableau de bord du système offre les possibilités de visualiser :

- Situation des effectifs des porteurs de carte.
- Situation du parcours d'une carte d'un porteur.
- Situation du parcours d'un commerçant.
- Situation des effectifs des GABs.
- Situation des effectifs par activité.
- Situation des transactions en litiges.
- Situation du comportement de l'acquéreur.
- Situation du comportement d'un émetteur.
- Situation des litiges en compensation.
- Situation des transactions en compensation.
- etc.

#### III.5 Avantages du système :

- **Centralisé** : le système de production Monétique contient toutes les informations concernant la monétique, susceptible d'être exploitées dans la prise de décision, en regroupant les informations contenues dans les bases de données des directions.
- **Modulaire** : les données sont organisées par activité, ce qui facilite la compréhension.
- **Archivage** : mémorise les données sur 5 années, ce qui offre des informations très utiles dans un projet décisionnel. Les données archivées de la monétique ne concernent que deux ans.
- **Universel** : utilisable dans la gestion de toute la monétique sur le territoire algérien.
- **Client \ Serveur** : permet l'exploitation des données en réseau.
- **Normalisé** : le développement du Système d'Information (SI) est régie par des normes facilitant la maintenance et la compréhension des composants du système.
- **Documenté** : une documentation est disponible concernant le mode d'emploi de l'outil.

### III.6 Inconvénients du système :

L'inconvénient du système Monétique est que les données ne sont pas préparées à un usage décisionnel car il est basé sur un modèle entité-association, de même pour le système CCP.

Inconvénient concernant l'étude, est que l'accès aux données ainsi qu'au schéma relationnel des deux bases de données est strictement interdit, le seul moyen d'accès aux données se fait à travers des fichiers plats dont la structure et le format ont été arrêtés et appliqués par les services concernés.

### III.7 Conclusion :

Nous retrouvons dans le système Monétique toutes les caractéristiques d'un système de production classique, ainsi que quelques caractéristiques d'un ERP, ce qui implique l'existence de beaucoup de fonctionnalités qui ne sont pas encore utilisées et d'autres qui ne seront jamais utilisées. Aussi, certaines caractéristiques qui peuvent être celles d'un entrepôt de données : historique des données, centralisation des données orienté métier, apparaissent sur le système ERP.

Nous pouvons donc considérer que le système Monétique offre une bonne source de données pour la réalisation d'un projet décisionnel, ainsi que pour le système CCP.

Comme nous l'avons évoqué auparavant l'accès direct aux bases de données Monétique ainsi que CCP est interdit, le seul moyen d'extraire les données vers notre entrepôt de données est d'y parvenir à travers les fichiers plats générés par ces deux systèmes.

**PARTIE IV**

**Démarche et Choix de Conception**

Dans cette partie

**Chapitre I : Conception du niveau de consolidation.**

**Chapitre II : Méthode de conception d'un entrepôt de données.**

**Chapitre III : Application du cycle de vie et choix du formalisme.**

**PARTIE IV : Démarche et Choix de Conception**

**Conception du niveau de consolidation**

Dans ce chapitre

**Choix du modèle.**

**Modélisation dimensionnelle.**

**Navigation dans les données.**

**Les architectures d'un Entrepôt de données.**

**Conclusion.**

En se basant sur les définitions apportées dans la partie II et l'étude de l'existant effectuée dans la Partie III, nous allons décrire le processus de conception de l'outil d'aide à la décision.

La première partie concernera la conception de la structure qui permettra l'organisation des données à des fins décisionnelles.

L'étude débutera par la comparaison entre deux modèles existants de modélisation, suivie d'une présentation du modèle adopté.

Nous poursuivrons par proposer une méthode de conception qui sera finalement appliquée sur notre cas, afin de fournir le modèle logique et physique de l'entrepôt.

Chaque partie répondant à des besoins spécifiques, sera traitée indépendamment d'une manière détaillée, en apportant à chaque fois les définitions des concepts utilisés, ainsi que les justificatifs qui nous ont amené à choisir une solution plutôt qu'une autre.

#### IV. CHAPITRE I : CONCEPTION DU NIVEAU DE CONSOLIDATION

Les performances d'un tableau de bord, résident notamment dans la qualité et la pertinence des informations fournies aux décideurs, en plus des critères purement techniques extrêmement importants de vitesse d'exécution des requêtes.

Afin d'arriver à ce niveau de performance, les données exploitées par le tableau de bord doivent répondre à des critères bien précis :

- **Fiabilité** : les données doivent contenir le moins d'erreurs possible.
- **Pertinence** : les données doivent avoir un contenu informationnel important pour le décideur.
- **Accessibilité** : les données doivent être structurées d'une manière à simplifier leur compréhension et la navigation logicielle.

Les systèmes de production en général ne répondent pas à ces critères. Et c'est la raison pour laquelle cette étape de conception est importante.

La première étape vise donc à concevoir un niveau de consolidation des données afin de les préparer à un usage décisionnel, exploitable plus efficacement par les applications de Data Warehouse et de reporting.

Nous appellerons tout au long du document le niveau de consolidation «*Entrepôt*».

À ce niveau une question se pose : Quelle modélisation adopter pour atteindre les buts fixés ?

**I.1 Choix du modèle :**

Deux grandes familles de modèle existent dans la conception des bases de données :

La modélisation transactionnelle : **OLTP** (Online Transaction Processing).

La modélisation dimensionnelle : **OLAP** (Online Analytic Processing).

Afin de choisir la modélisation la plus adaptée à notre travail une description de chacune d'elle s'impose.

**I.1.1 Deux mondes différents :**

Ce titre emprunté à Ralph Kimball [Kim, 2001], dans la comparaison des modèles transactionnels et dimensionnels, nous met directement dans le vif du sujet et nous laisse pressentir que les différences sont de taille.

Nous résumerons ces différences dans un tableau organisé par thèmes de différence qui sont :

- Les utilisateurs.
- Le contenu des données.
- L'organisation des données.
- L'administration du système.

**Tableau IV.1 : Résumé des différences des deux modèles de conception de base de donnée : [Kim, 2001].**

OLTP	OLAP
<b>Utilisateurs</b>	
Agents opérationnels	Managers (Décideurs)
Beaucoup d'utilisateurs concurrents (Avec même une certaine duplication).	Peu d'utilisateurs
Exécutent un grand nombre de fois la même tâche.	Les demandes sont similaires mais différentes (on ne demande pas deux fois la même chose)
Lisent et modifient les données (système de données vivant).	Lisent uniquement les données.
<b>Contenu des données</b>	
Nécessaire au fonctionnement de l'entreprise (c'est la mémoire collective de l'entreprise ; si l'OLTP s'arrête en général l'entreprise s'arrête aussi).	Utilisée pour analyser le fonctionnement de l'entreprise.
<ul style="list-style-type: none"> <li>- Détaillées, actuelles.</li> <li>- Accessibles de façon individuelle.</li> </ul>	<ul style="list-style-type: none"> <li>- Résumées, souvent agrégées (le détail ne nous intéresse pas).</li> <li>- Sont recalculées.</li> <li>- Accessibles de façon ensembliste.</li> </ul>
En règle générale non historisé (mise à jour ; image à un moment donnée).	Le temps est fondamental (historiques).
<b>Organisation des données</b>	
Modèle entité-relation.	Modèle dimensionnel.
Schéma symétrique.	Schéma asymétrique.
Beaucoup de tables.	Peu de tables.
Tables normalisées.	Les tables peuvent être dénormalisées.

Tableau IV.1 (Suite) : [Kim, 2001]

OLTP	OLAP
<b>Administration</b>	
Forte disponibilité.	Disponibilité faible.
Sauvegardes fréquentes (car le système bouge beaucoup).	Sauvegardes peu fréquentes mais très volumineuses.
Beaucoup de petites transactions (paramétrage de la machine, commit et rollback).	Règle générale, une transaction par jour (chargement de données).
Cohérence microscopique des données.	Mesure globale de la cohérence.

### 1.1.2 Résumé :

En effet notre projet de conception d'un tableau de bord s'inscrit dans un environnement décisionnel, ses caractéristiques sont celles d'un modèle dimensionnel, et au vue des principales caractéristiques de chacune des modèles et de la conclusion apportée par Ralph Kimball à propos du modèle entité relation où il dit :

*« Les modèles de données entité-relation sont un désastre pour l'exécution de requêtes parce qu'il ne peuvent pas être compris par les utilisateurs et parce qu'il ne peuvent pas être parcourus utilement par le logiciel de base de données. Les modèles entité-relation ne peut pas servir pour construire des entrepôts de données d'entreprise ».* [Kim, 2001].

Nous opterons pour une modélisation dimensionnelle de notre entrepôt.

### 1.2 Modélisation dimensionnelle :

Dans ce chapitre nous passerons en revue les principales caractéristiques de la modélisations dimensionnelle avant de les mettre en pratique pour la conception de notre entrepôt.

#### 1.2.1 Caractéristiques du modèle dimensionnel :

##### 1.2.1.1 Une approche descendante :

Un modèle dimensionnel fournit une vue globale des données qui correspond aux besoins des décideurs, les données y sont représentées d'une manière simple et compréhensible, de sorte que l'activité est décrite par des indicateurs de performance et des dimensions d'étude.

##### 1.2.2 Représentation graphique des données sous forme d'hyper cubes :

Dans un modèle dimensionnel les données peuvent être vue comme étant un hyper cube à N dimension, où chaque axe représente les dimensions de l'activité modélisée et où chaque cellule (intersection des axes) représente une valeur de l'indicateur de performance retenue par l'activité.

**Exemple :**

Représentation cubique de l'activité de gestion des Produits d'une entreprise.

Les axes représentent les dimensions d'étude suivantes :

La dimension Produit.

La dimension Marché.

La dimension Temps.

Les cellules représentent le nombre de vente et le chiffre d'affaire par produit.

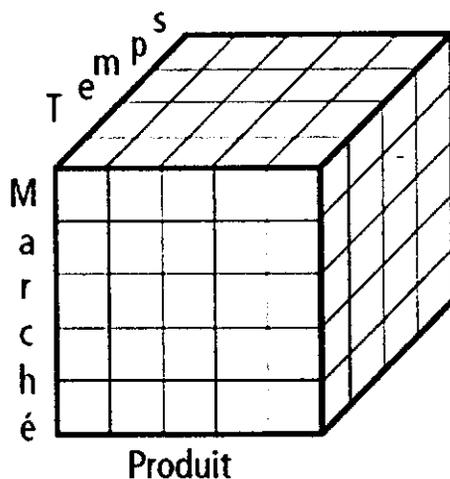


Figure IV.20 : Représentation cubique de l'activité gestion des Produits.

### 1.2.3 Le modèle est asymétrique sous forme d'étoile :

Le modèle dimensionnel est dit asymétrique car contrairement à un modèle relationnel où pratiquement toutes les tables sont de taille égale, il contient deux types de tables.

- De grandes tables appelées « Table de faits », comportant des jointures et les faits calculables de l'activité.
- Des table plus petites appelées « Table de dimension », reliées aux tables de faits contenant des descriptions plus au moins statiques des dimensions de l'activité.

L'avantage de l'asymétrie est qu'il est très simple de distinguer dans un schéma les tables contenant les mesures de celles qui contiennent des attributs statiques.

Un modèle dimensionnel a souvent la forme d'une étoile ou plus généralement d'une constellation d'étoile (table de fait au centre reliée par plusieurs tables de dimensions). La description des tables de faits et dimensions est mentionnée à la partie II.

### I.3 Les différents modèles conceptuels :

#### I.3.1 Schéma en Etoile :

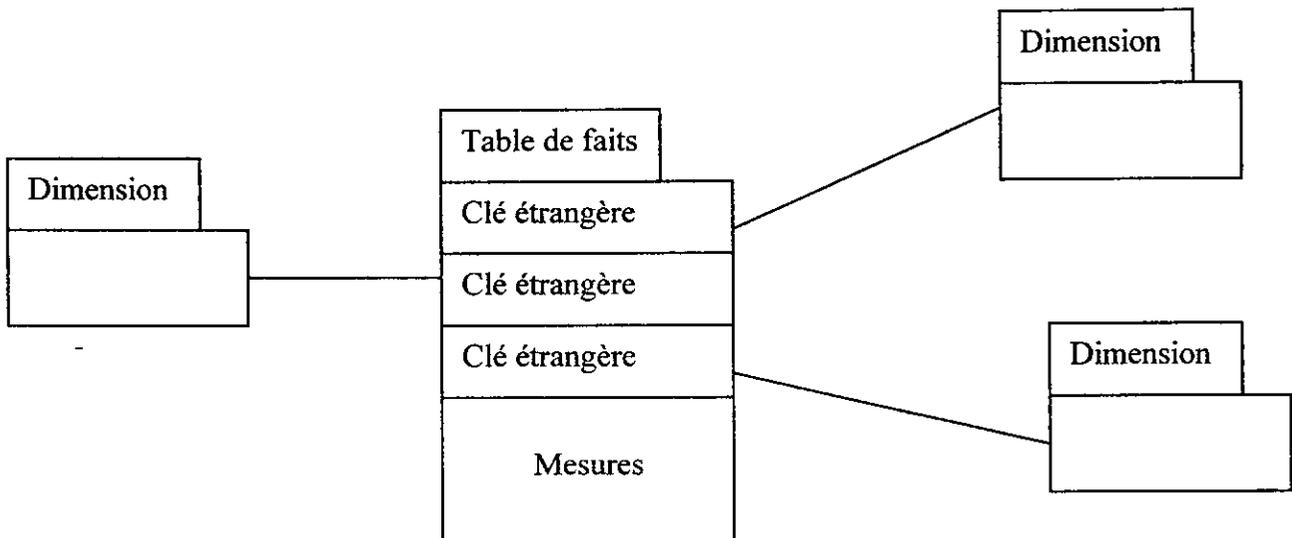


Figure IV.21 : Schéma en Etoile.

#### 1. Description :

- Table de fait au milieu concaténée à un ensemble de tables de dimension.
- Les tables dimensions peuvent ne pas être normalisées.

#### 2. Avantages :

- Représentation simple à la compréhension.
- Réduire le nombre de tables et de jointure.
- Efficacité au niveau de la navigation logiciel.

#### 3. Inconvénients :

- Redondance des données due à la dénormalisation des dimensions, ce qui implique une perte en espace de stockage.

### I.3.2 Flocon de neige :

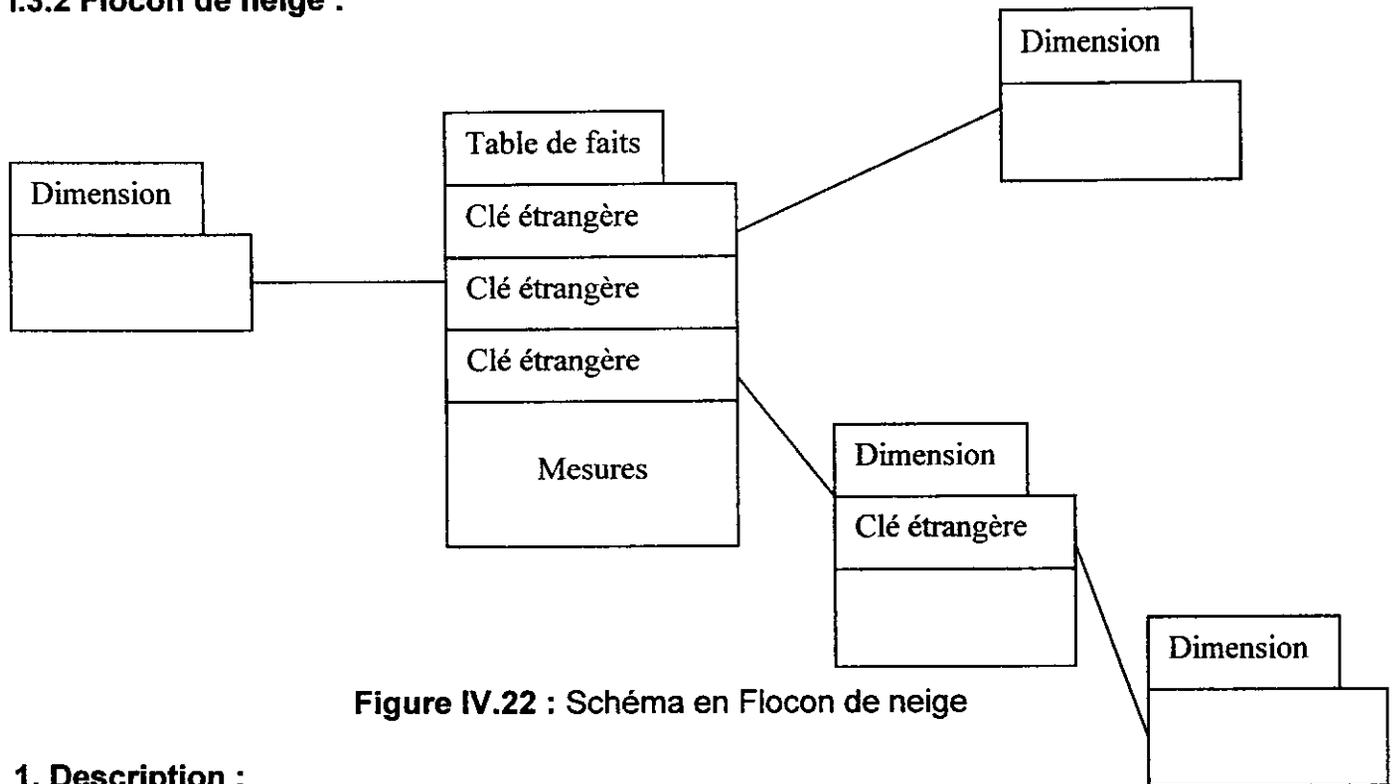


Figure IV.22 : Schéma en Flocon de neige

#### 1. Description :

- Un modèle en étoile dont les dimensions sont normalisées (troisième forme normale).

#### 2. Avantages :

- Aucune redondance des données ce qui implique un gain en espace mémoire par rapport au schéma en étoile.

#### 3. Inconvénients :

- Perte de la simplicité du schéma.
- Navigation logicielle rendue plus complexe.

#### 4. Remarque :

Le gain en espace de stockage qu'implique le floconnage des dimensions s'avère négligeable.

En effet, en général une table dimension a une taille très petite par rapport à la taille de la table de faits ce qui explique aisément l'insignifiance du gain.

Par contre la faiblesse du modèle en flocon en terme de navigation et de simplicité est très importante. C'est ce qui nous amène à la conclusion suivante :

*Le modèle dénormalisé en étoile est le plus adapté à la conception des entrepôts.*

## **5. Note :**

Il est possible que dans certains cas très particulier, le gain induit par la normalisation des dimensions soit significatif.

### **I.4 Navigation dans les données :**

La particularité des schémas dimensionnels implique une navigation particulière dans les données, donc pour clore ce chapitre nous présenterons les principales opérations de navigations dans un modèle dimensionnel.

#### **I.4.1 Le forage vers le bas :**

Désigne la possibilité d'aller d'un niveau globale vers un niveau plus détaillé, en d'autre termes c'est l'opération qui répond à la question des décideurs : « Montrer moi plus de détails ».

Le forage vers le bas se fait en général en passant d'un niveau dans une hiérarchie à un niveau plus bas.

#### **I.4.2 Le forage vers le haut :**

Consiste en l'opération inverse du forage vers le bas, passant d'un niveau dans une hiérarchie à un niveau plus haut. Il s'agit d'une agrégation des données.

#### **I.4.3 Le forage horizontal :**

Permet de donner des informations correspondant à différentes tables de faits de domaines d'activités différents liés par au moins une table de dimension.

##### **I.4.3.1 Glisser\_déplacer :**

Manipulation pour permuter les axes d'analyses.

### **I.5 Les architectures d'un Entrepôt de données:**

Pour implémenter un Entrepôt de données, trois types d'architectures sont possibles [CNAM, 1998] :

- L'architecture réelle.
- L'architecture virtuelle.
- L'architecture remote.

### **I.5.1 L'architecture Réelle :**

Elle est généralement retenue pour les systèmes décisionnels. Le stockage des données est réalisé dans un SGBD séparé du système de production. Le SGBD est alimenté par des extractions périodiques.

Avant le chargement, les données subissent d'importants processus d'intégration, de nettoyage, de transformation.

L'avantage est de disposer des données préparées pour les besoins de la décision et répondant aux objectifs de L'entrepôt de données.

Les inconvénients sont le coût de stockage supplémentaire et le manque d'accès en temps réel.

### **I.5.2 L'architecture Virtuelle :**

Cette architecture n'est pratiquement pas utilisée pour l'entrepôt de données. Les données résident dans le système de production. Elles sont rendues visibles par des produits middleware ou par des passerelles.

Il en résulte deux avantages : pas de coût de stockage supplémentaire et l'accès se fait en temps réel.

L'inconvénient est que les données ne sont pas préparées.

### **I.5.3 L'architecture Remote :**

C'est une combinaison de l'architecture réelle et de l'architecture virtuelle. Elle est rarement utilisée.

L'objectif est d'implémenter physiquement les niveaux agrégés afin d'en faciliter l'accès et de garder le niveau de détail dans le système de production en y donnant l'accès par le biais de middleware ou de passerelle.

*L'architecture réelle est la plus approprier pour la construction de notre entrepôt de données.*

### **I.6 Conclusion :**

Nous avons exposé dans ce chapitre la problématique due au choix du modèle de conception de l'entrepôt ainsi que les caractéristiques du modèle dimensionnel. A présent, il nous reste à choisir une méthodologie de conception adaptée qui sera l'objet du chapitre suivant.

**PARTIE IV : Démarche et Choix de Conception**

**Méthode de conception d'un Entrepôt de données**

Dans ce chapitre

**Rappel des caractéristiques du contexte d'application.**

**Présentation des méthodes de conception :**

- **Présentation de l'approche selon Ralph Kimball.**
- **Présentation de l'approche selon Mark A.R.Kortink et Daniel L.Moody.**

**Conclusion.**

## IV. CHAPITRE II : METHODE DE CONCEPTION D'UN ENTREPOT DE DONNEE

Dans ce chapitre nous aborderons la partie méthodologie ou démarche de la conception dimensionnelles d'un entrepôt.

Nous avons choisi de consacrer un chapitre entier à cet effet par rapport à l'importance du choix qui nous ait donné de faire et de l'importance d'une bonne démarche dans le développement de l'entrepôt.

Nous allons présenter deux méthodes :

- La méthode classique de Ralph Kimball.
- La méthode de conception à partir d'un modèle d'entreprise de Marck A.R.Kortink et Daniel L.Moody.

Chacune des méthodes sera présentées, puis nous opterons ; en fonction des avantages de chacune, pour une démarche adaptée à notre cas.

### II.1 Rappel des caractéristiques du contexte d'application :

La réalisation, de l'entrepôt aura pour but de fournir au tableau de bord des données prêtes à être exploité à un usage décisionnel.

Comme nous l'avons vu dans la partie étude de l'existant, le service Monétique et CCP, disposent, d'un système d'information centralisé et regroupe toutes les informations des bases de données des activités Monétiques et CCP respectivement, mais en raison de sécurité des données et des comptes des clients aucun accès direct au bases opérationnelles n'est autorisé, l'accès au données se fait à travers des fichiers plats bien structurés.

### II.2 Présentation des méthodes de conception :

Une méthode de conception adaptée à l'architecture retenue est primordiale pour atteindre un haut niveau de performance, ceci n'est pas chose aisée car l'inexistence de méthodologie faisant l'unanimité tel que MERISE ; dans la modélisation entité-relation fait qu'au moment du choix de la méthode, une réflexion approfondie s'avère indispensable.

Nous présenterons dans les points suivants la méthode de conception proposée par Kimball ainsi qu'une méthode proposée par Kortink et Moody qui permet de créer un modèle dimensionnel à partir d'un modèle d'entreprise en entité-relation.

#### II.2.1 Présentation de l'approche selon Ralph Kimball :

Considérée comme étant la référence dans le monde de la conception des entrepôts de données, la méthode proposée par Ralph Kimball, décrit les principales étapes de conception ainsi que les neuf points à prendre en considération pour obtenir un bon modèle dimensionnel.

Pour la conception d'une base de données dimensionnelle, il a prévu quatre grandes étapes. En voici une définition initiale : [Kim, 2001].

### 1. Choisir le *processus d'activité* à modéliser :

Un processus d'activité est un processus important pour l'organisation, étayé par une application existante (ou des applications) à partir de laquelle les données peuvent être collectées au profit de l'entrepôt de données.

### 2. Choisir le *grain* du processus d'activité :

Le grain est le niveau de détail fondamental, atomique des données figurant dans les tables de faits pour ce processus. Des grains typiques sont des transactions individuelles, des récapitulatifs individuelles quotidiennes, des récapitulatifs individuelles mensuelles. Il est impossible de passer à l'étape trois sans avoir préalablement défini le grain.

3. Choisir les *dimensions* applicables à chaque enregistrement de la table de faits. Des dimensions typiques sont le temps, l'argent, la région, etc. le choix d'une dimension s'accompagne de la définition de tous les attributs textuels des champs qui garniront la table de dimension.

4. choisir les *faits* mesurés que contiendra chaque enregistrement de la table de faits. Des faits mesurés typiques sont des quantités numériques additives.

#### II.2.1.1 Les neuf décisions majeures :

Les neuf décisions à prendre pour la conception complète d'un entrepôt de données dimensionnel portent sur les points suivants : [Kim, 2001].

1. Les processus métier.
2. Le grain de chaque table de faits.
3. Les dimensions de chaque table de faits.
4. Les faits, y compris les faits pré calculés.
5. Les attributs de dimensions, avec des descriptions complètes et la terminologie adéquate.
6. Comment suivre les dimensions à évolution lente.
7. Les agrégats, les dimensions hétérogènes, les minidimensions, les modes de requêtes et autres décisions sur le stockage physique.
8. L'étendue historique de la base de données.
9. L'urgence avec laquelle les données doivent être extraites et chargées dans l'entrepôt des données.

Les différents choix reposent d'une part sur une bonne compréhension des besoins des utilisateurs finaux, d'autre part sur les caractéristiques des informations dont peut disposer l'équipe chargée de la construction de l'entrepôt de données, à partir des sources représentées par les applications existantes pour cela l'équipe de conception effectue un ensemble d'interviews croisées entre les utilisateurs finaux et les administrateurs des bases de données. [Kim, 2001].

Cet aspect est le talon d'Achille de la méthode, on effet Kimball ne fournit aucune technique permettant d'exploiter les modèles de base entité-relation et est à la merci des erreurs de conception dues à une mauvaise compréhension des besoins utilisateurs et des explications des administrateurs. [Kim, 2001].

Par contre l'auteur de la démarche de conception suivante, Kimball met l'accent sur cet aspect et donne un moyen d'exploiter efficacement un modèle entité-relation afin de créer un modèle dimensionnel fiable. [Kim, 2001].

L'auteur Ralph Kimball expose aussi, que la réussite de l'implémentation d'un entrepôt de données dépend de l'intégration adéquate de nombreux composants et tâches. Il ne suffit pas de posséder le modèle de données parfait ou la meilleure technologie ; il s'agit de coordonner les multiples facettes du projet de Data Warehouse. [Kim, 2001].

Les efforts fournis pour l'implémentation d'un entrepôt de données doivent révéler leur efficacité à travers tous les aspects du projet. Le cycle de vie dimensionnel s'apparente à la partition d'un chef d'orchestre ; il fait en sorte que les éléments du projet soient placés dans le bon ordre au bon moment, il définit la marche à suivre. [Kim&all,2000].

L'approche globale de l'instrumentation d'entrepôts de données par le cycle de vie que propose de suivre Kimball est illustrée par la figure IV.23. Ce schéma représente la succession des tâches de haut niveau (macro tâches) nécessaires à la conception, au développement et au déploiement d'entrepôts de données efficaces. Il décrit le cheminement du projet dans son ensemble ; chaque rectangle sert de poteau indicateur ou de borne. [Kim&all,2000].

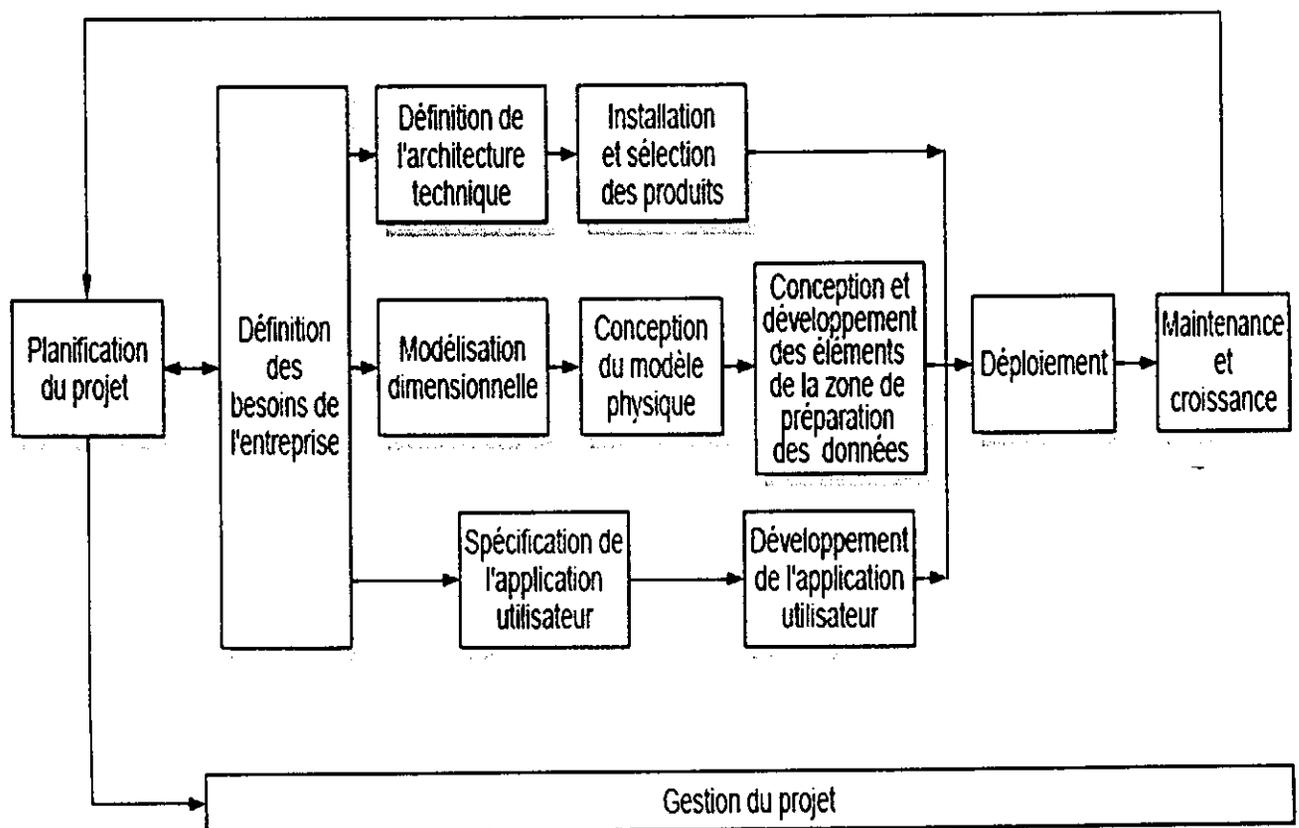


Figure IV.23 : Schéma du cycle de vie dimensionnel. [Kim&all,2000]

## **II.2.2 Présentation de l'approche selon Mark A.R.Kortink et Daniel L.Moody [Kort&Mood, 2000].**

Mark A.R.Kortink et Daniel L.Moody contestent l'idée de Ralph Kimball du fait que les modèles entité-relation ne peuvent pas servir de base pour construire des entrepôts de données d'entreprise.

Pour cela, ils proposent une méthode de conception basée sur un modèle existant.

La première étape de la méthode consiste à faire la classification des hiérarchies qui existent dans le modèle. La dernière étape concerne la réduction des hiérarchies et l'agrégation des entités de transactions pour former des modèles dimensionnels.

### **II.2.2.1 L'approche de conception du schéma en étoile :**

La méthode de conception de Kimball est considérée comme étant la pionnière dans le domaine. Cependant, il existe quelques problèmes avec cette approche : [Kort&Mood, 2000].

1. Les interviews ne répondent pas nécessairement à une réalité. Les besoins et les demandes des utilisateurs sont imprévisibles et sont sujet aux changements. Ce qui fournit une base instable pour la conception.
2. On peut aboutir à une conception incorrecte si le concepteur ne comprend pas les relations qui existent entre les données.
3. Perte d'information due aux agrégations prématurées, qui limitent la manière dont les données peuvent être analysées.
4. l'approche est représentée à travers des exemples plutôt qu'à travers une procédure de conception explicite (la présentation n'est pas formalisée).

L'approche de Kortink et Moody surmonte ces problèmes en utilisant le modèle entité-relation existant (la base de production) comme étant la base de la conception d'un entrepôt de données. En utilisant les relations entre les données, qui ont préalablement déjà été documentées. Ce qui offre une approche structurée pour le développement d'un entrepôt de données. [Kort&Mood, 2000]

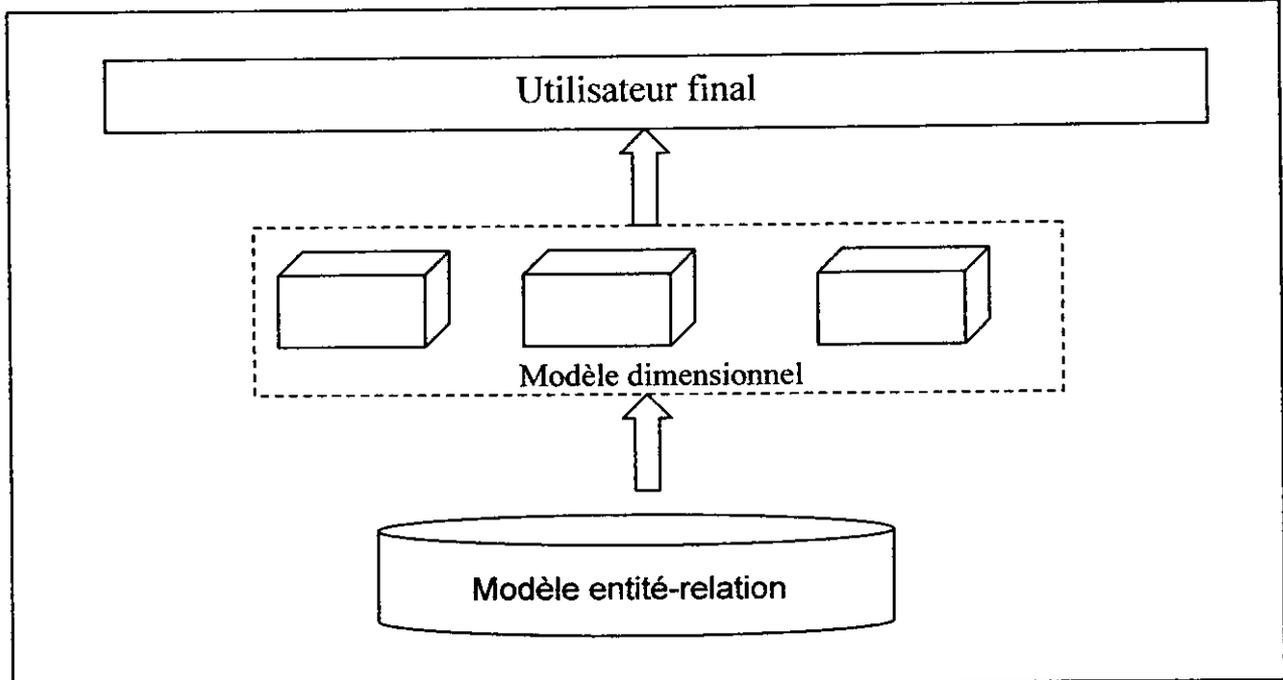
Nous retrouvons ce principe de conception notamment chez Matteo Golfrelli et Stefano Rizzi [Golf&Riz,1998].

### **II.2.2.2 L'approche de conception d'un entrepôt de données :**

Le modèle entité-relation fournit un bon support pour l'alimentation en données car il est détaillé, bien structuré et facilement mis à jour.

Par contre il est incompréhensible par les utilisateurs finaux et non performants aux requêtes. Ce qui nécessite l'ajout d'un niveau supérieur qui consiste en une structure dimensionnelle plus adaptée à un usage décisionnel. [Kort&Mood, 2000]

La figure IV.24 illustre la transition du modèle relationnel au modèle dimensionnel.



**Figure IV.24 :** Transition des données du modèle relationnel vers le modèle dimensionnel.  
[Kort&Mood, 2000]

### II.3 Conclusion :

A la vue des caractéristiques des deux méthodes et du contexte d'application, nous notons les points suivants :

- Kimball ne fournit aucune méthodologie permettant d'exploiter les systèmes existants.
- La deuxième démarche fournit un formalisme qui permet d'exploiter efficacement un système existant et de créer un modèle dimensionnel.
- La critique du deuxième principe de Mark A.R.Kortink et Daniel L.Moody est fondée. Surtout que dans notre cas la réticence et l'inaccessibilité des données directement ainsi qu'au schéma relationnel des deux systèmes, de plus elle ne couvre pas tout le cycle de vie du développement.
- Les deux auteurs sont d'accord sur le fait qu'un modèle dimensionnel est adapté à un usage dimensionnel.
- Les deux méthodes se rejoignent au bout de la première étape.

Nous pouvons conclure en disant que la première méthode (*La méthode classique de Ralph Kimball*) est la plus appropriée à notre cas d'étude, et cela même s'il est très important de pouvoir vérifier les besoins des utilisateurs avec les données opérationnelles.

Nous comblerons ce manque par l'interview des administrateurs de données des deux systèmes.

**PARTIE IV : Démarche et Choix de Conception**

**Application du cycle de vie et choix du formalisme**

Dans ce chapitre

**Cycle de vie suivi par l'étude.**

**Description des différentes phases du cycle de vie :**

- **Définition des besoins.**
- **Modélisation dimensionnelle des données.**
- **Conception de la zone de préparation des données**
- **Spécification de l'application utilisateur.**
- **Développement.**
- **Déploiement.**

**Formalisme a utilisé :**

- **Le point sur l'UML.**
- **Les diagrammes d'UML.**
- **Modélisation multidimensionnelle.**
- **Synthèse.**
- **Résumé.**

**Conclusion.**

## IV. CHAPITRE III : APPLICATION DU CYCLE DE VIE ET CHOIX DU FORMALISME

Dans cette partie du chapitre nous allons exposer le cycle de vie que nous avons suivi et appliqué dans notre cas d'étude où nous définissons chaque phase de la démarche brièvement, et dont le détail sera traité dans la partie Conception de l'entrepôt de données.

Nous poursuivons cette partie par définir le formalisme utilisé pour décrire chaque phase du cycle de vie, nous exposons les problématiques de l'expression des besoins, de la modélisation des données et des architectures des systèmes décisionnels. Nous terminons le chapitre par une synthèse sur l'UML décisionnel, proposé par Pascal Brunot et Hadrien Cambazard.

### III.1 Cycle de vie suivi par l'étude :

Le cycle de vie que nous avons suivi est inspiré du cycle de vie proposé par Ralph Kimball.

Le cycle de vie est illustré par la figure IV.25.

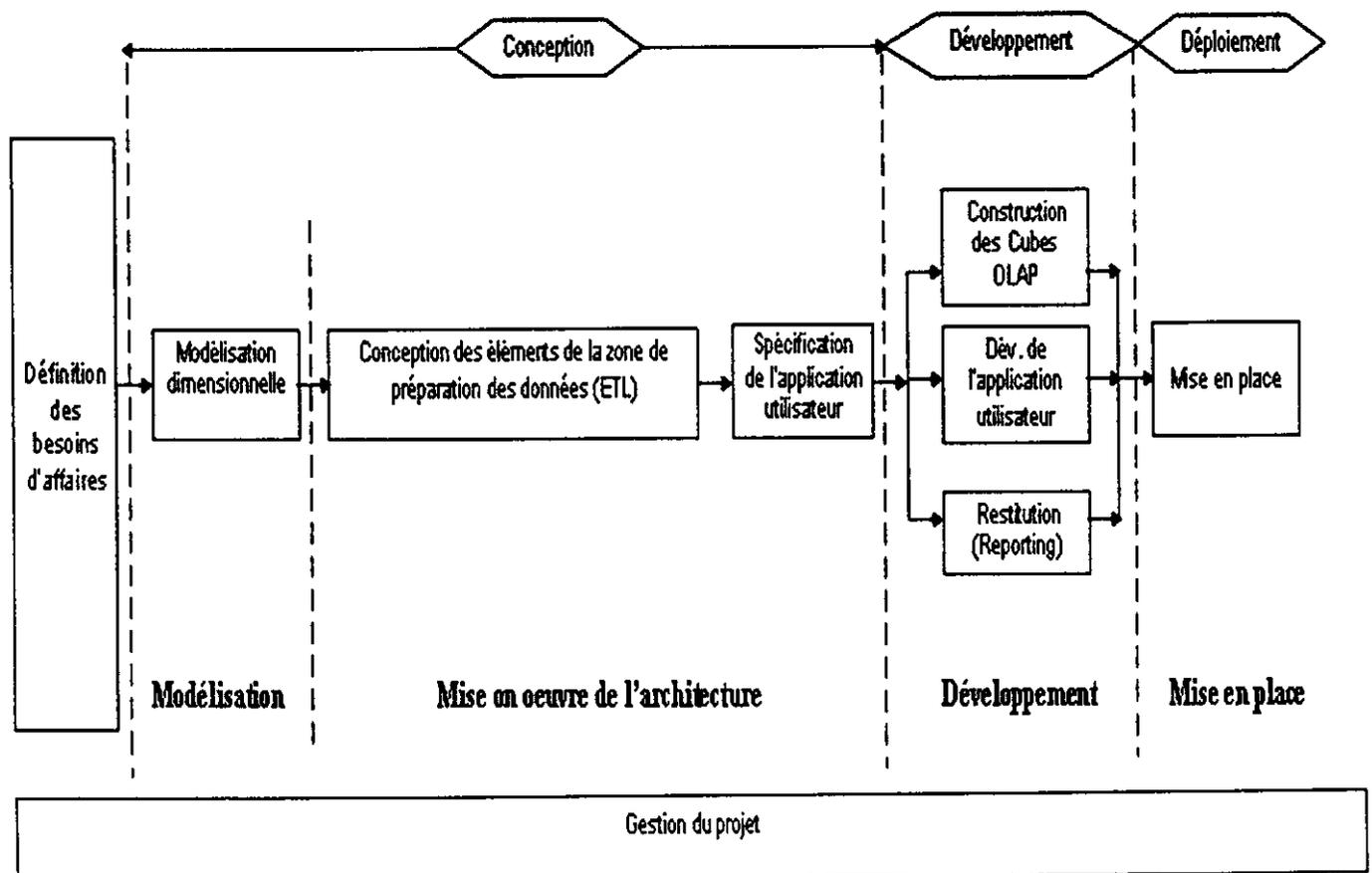


Figure IV.25 : Schéma du cycle de vie suivi par notre étude.

## III.2 Description des différentes phases du cycle de vie: [Kim&all,2000].

### III.2.1 Définition des besoins :

Les chances de succès d'un entrepôt de données se trouvent considérablement accrues par la bonne compréhension des utilisateurs et de leurs besoins. Sans elle, l'entrepôt de données deviendrait rapidement un exercice vain pour les concepteurs.

L'approche utilisée pour identifier les besoins analytiques des employés diffère de manière significative de la traditionnelle analyse des besoins pilotée par les données. Les constructeurs d'entrepôts de données doivent appréhender les facteurs clés qui conduisent l'entreprise à vouloir définir de manière efficace ses besoins et les traduire pour les intégrer lors de la conception. La collecte des besoins est traitée en détail dans la partie V *Conception de l'entrepôt de données* chapitre *Définition des besoins*.

### III.2.2 Modélisation dimensionnelle des données :

La définition des besoins détermine les données requises pour répondre aux besoins d'analyse des utilisateurs. La conception de modèles de données destinés à corroborer ces analyses nécessite une approche différente de celle utilisée lors de la conception de systèmes opérationnels.

A partir des indicateurs et de leurs axes d'analyse qui seront déterminés par la phase étude des besoins, nous effectuerons une analyse plus détaillée des données des systèmes source opérationnels. En couplant cette analyse à la compréhension des besoins précédemment établie, nous développerons ensuite un modèle dimensionnel. Ce modèle identifie la granularité de la table des faits, les dimensions associées, les attributs ainsi que leur hiérarchisation.

La conception logique d'une base de données est complétée par les relations appropriées entre les structures des tables et les clés principales et extérieures. Cet ensemble d'activités s'achèvera sur le développement d'une mise en correspondance des données source et cible.

La modélisation dimensionnelle sera traitée dans la partie V : *Conception de l'entrepôt de données* chapitre *Modélisation dimensionnel*.

### III.2.3 Conception de la zone de préparation des données :

Le processus de conception des éléments de la zone de préparation des données (*data staging*), constituent souvent la tâche la plus sous-estimée du projet de Data Warehouse. Le processus de préparation se déroule en trois étapes majeures : l'extraction, la transformation et le chargement des données. Le processus d'extraction révèle toujours des problèmes de qualité des données, "enfouies" au sein des systèmes source opérationnels. Or, la qualité des données influence de manière significative la crédibilité de l'entrepôt de données ; il convient donc de régler ces problèmes. Pour compliquer encore la chose, il faut concevoir et construire deux processus d'extraction : le premier pour le peuplement initial de l'entrepôt de données, le second pour les chargements réguliers et incrémentiels. La conception de la zone de préparation des données est décrit en détail dans la partie V : *Conception de l'entrepôt de données* chapitre *Conception de la zone de préparation des données*.

### III.2.4 Spécification de l'application utilisateur :

Il est recommandé de définir une série d'applications standard destinée à l'utilisateur final, car tous les utilisateurs n'ont pas besoin d'un accès *ad hoc* à l'entrepôt de données. Les spécifications de l'application décrivent les maquettes d'états, les critères laissés à l'initiative de l'utilisateur et les calculs nécessaires. Elles garantissent que l'équipe de développement et les utilisateurs ont intégré une compréhension commune des applications qui vont être livrées. Les détails concernant les spécifications utilisateur final sont fournis dans la partie V *Conception de l'entrepôt de données* chapitre *Reporting*.

### III.2.5 Développement :

Cette partie du cycle de vie concerne le développement des différentes phases de ce dernier (Construction des cubes OLAP, Développement de l'application utilisateur), cette partie se termine par la phase restitution, qui consiste en la présentation et la représentation de l'outil décisionnel.

### III.2.6 Déploiement

Le déploiement est le point de convergence de la technologie, des données et des applications utilisateur accessibles à partir du poste de travail. Une planification de grande envergure est indispensable afin que toutes les pièces du puzzle se placent correctement. Il est également nécessaire de prévoir une formation des utilisateurs qui intègre tous les aspects de cette convergence. En outre, avant de permettre à l'utilisateur d'accéder à l'entrepôt de données, il faut mettre en place les processus de communication, de support utilisateur, et de prise en compte des demandes d'évolution ou de correction. Nous évoquons le déploiement de notre application au sein de l'entreprise Algérie poste à La partie V Chapitre *Schéma de mise en oeuvre et Déploiement*.

### III.3 Formalisme à utiliser :

Dans ce qui suit nous effectuons une analyse de l'apport d'UML pour la conception de systèmes décisionnels : UML et son approche objet, répondent-ils aux problématiques de la modélisation d'un système décisionnel ?

Après une mise au point sur l'UML, nous nous intéresserons plus particulièrement aux problématiques de l'expression de besoin et de la modélisation de données.

Une conclusion fournie par les deux auteurs Brunot et Hadrien Cambazard, est résumée dans un tableau en fin de synthèse concernant l'apport d'UML dans les systèmes décisionnels.

#### III.3.1 Le point sur l'UML : [Bru&Cam,2004].

UML, Unified Modeling Language, est un langage formel qui permet d'exprimer et d'élaborer des modèles objet, indépendamment de tout langage de programmation.

Né de la fusion des méthodes objet dominantes (OMT, Booch et OOSE), puis normalisé par l'OMG en 1997, UML est rapidement devenu un standard.

Un système est décrit en UML à travers sa *structure statique* et son *comportement dynamique*.

UML propose donc un ensemble de notations graphiques pour capturer les informations relevant des aspects statiques et dynamique du système. C'est donc un langage de modélisation visuel et non une méthode d'analyse et de spécification. UML possède Néanmoins une dimension méthodologique et son utilisation s'inscrit dans une démarche :

- Guidée par l'expression des besoins des utilisateurs du système.
- Centrée sur l'architecture logicielle.
- Itérative et incrémentale (UML supportant très bien l'abstraction)

### III.3.2 Les diagrammes : [Bru&Cam,2004].

#### III.3.2.1. Structure statique

Cette vue du modèle comporte cinq types de diagrammes :

- Diagrammes d'objets
- Diagrammes de classes
- Diagrammes de cas d'utilisation
- Diagrammes de composant
- Diagrammes de déploiement

#### III.3.2.2. Comportement dynamique

Cette vue du modèle comporte quatre types de diagrammes :

- Diagrammes de séquence
- Diagrammes de collaborations
- Diagrammes d'états-transitions
- Diagrammes d'activités

### III.3.3 Données décisionnelles : [Bru&Cam,2004].

L'exercice de ces métiers passe par l'analyse de ces données décisionnelles qui se déclinent en trois types distincts :

- Indicateurs
- Axes d'analyse
- Filtres

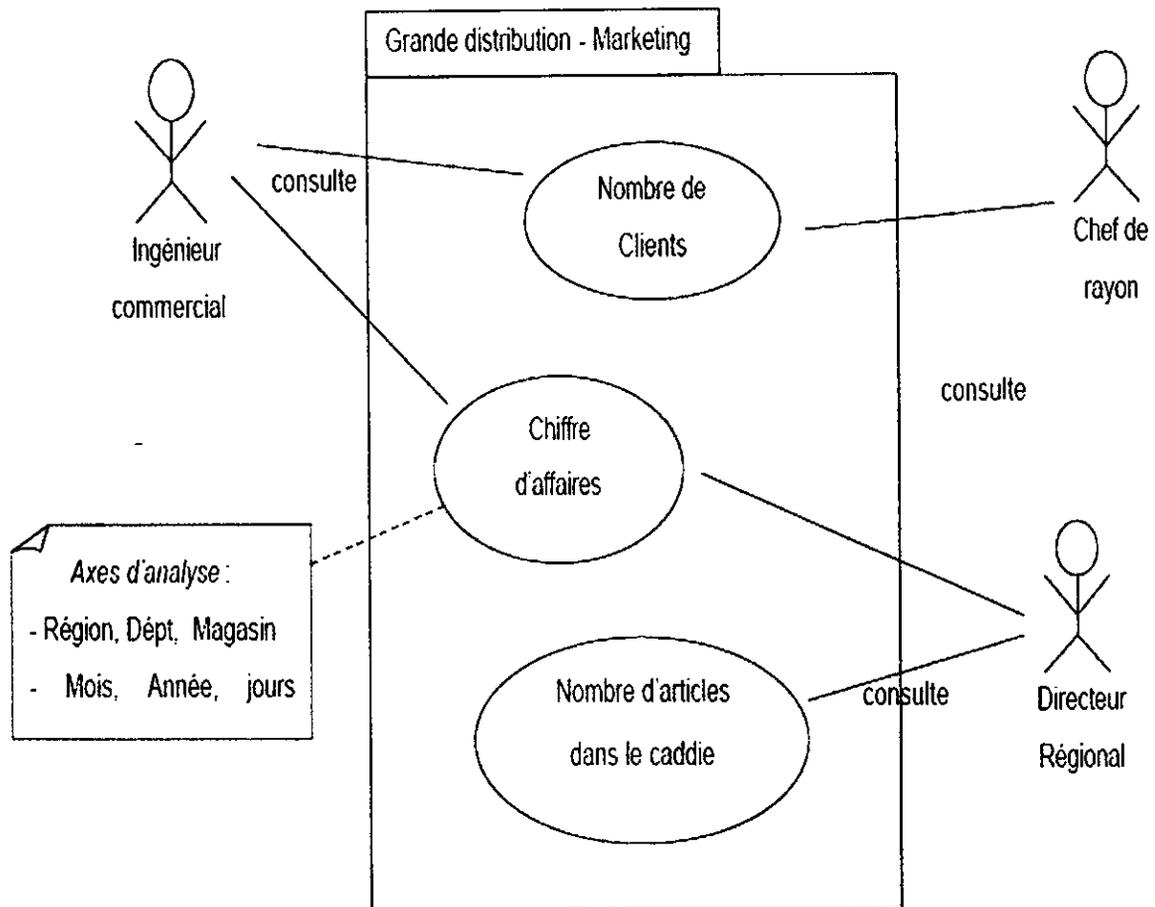
La donnée porteuse de sens qui oriente les choix dans un métier donné est l'indicateur. Or la définition d'un indicateur ou la règle de gestion qui permet de l'obtenir à partir des données opérationnelles peut varier d'un métier à l'autre au sein de l'entreprise.

Elle se rattache à un métier et donc à un acteur ou utilisateur de l'entrepôt de données. On remarque donc à ce niveau de l'expression de besoin que des ambiguïtés sémantiques au niveau des indicateurs existent et que la prise en compte des acteurs est essentielle pour les éliminer. L'acteur est le seul à connaître la règle de gestion qu'il utilise pour calculer tel ou tel indicateur. C'est la modélisation du métier et du savoir faire de chacun qui entre ici en jeu. Chacun à sa propre vision de l'organisation, du sens des termes et indicateurs de son métier.

Les cas d'utilisations qui permettent en UML de formaliser l'expression de besoin s'inscrivent bien dans cette démarche, puisque l'identification des acteurs est une étape cruciale de leur mise en oeuvre. Le système est représenté à l'origine par tous ses cas d'utilisation qui décrivent une interaction avec ses acteurs.

On choisit d'interpréter chaque indicateur comme un cas d'utilisation. Chaque cas d'utilisation est relié à l'acteur qui y fait appel. On regroupe les indicateurs selon des packages (dans l'exemple : Marketing) qui font intervenir les mêmes acteurs et qui touchent un ensemble cohérent de besoins.

L'exemple est illustré à la Figure IV.26.



**Figure IV.26 :** Exemple de représentation du besoin en indicateur avec le diagramme de cas d'utilisation. [Bru&Cam,2004].

La description textuelle des cas d'utilisation est cruciale et correspond à la base de l'analyse des besoins et de la discussion avec les acteurs. Elle concerne ici essentiellement l'association des axes d'analyses et des filtres à chaque indicateur mais aussi de la règle de gestion qui permet de le calculer.

Sur l'exemple précédent, la vue des acteurs met en évidence les questions concernant les indicateurs « Nombre de Clients » ou « Chiffre d'affaires ». Le chef de rayon et l'ingénieur commercial compte-il les clients de la même façon ? On pourrait très bien imaginer que l'ingénieur commercial, prend en compte les personnes morales en plus des personnes physiques que considère le chef de rayon bricolage. Mais tout deux parlent de clients.

Toujours dans le souci d'organiser et clarifier les besoins on peut chercher à structurer les indicateurs et à donner, prévoir le grain des faits déjà à ce niveau.

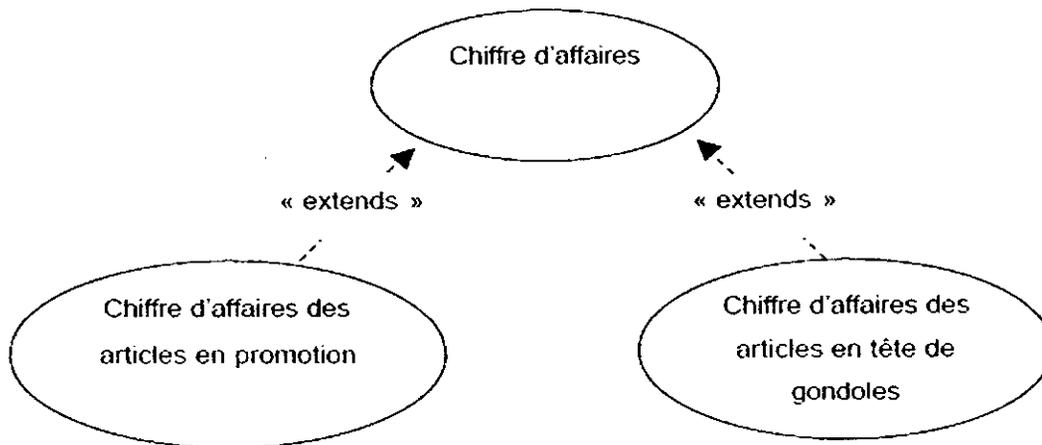


Figure IV.27 : Représentation de la relation extend dans un cas d'utilisation appliqué a un niveau décisionnel. [Bru&Cam,2004].

Le caractère « être en promotion » n'est pas ici un axe d'analyse du chiffre d'affaire, dans la mesure où le chiffre d'affaires des articles en promotion constitue une donnée directement exploitée par un acteur en tant que donnée décisionnelle. Ainsi on pourra choisir par la suite stocker à la fois un chiffre d'affaires et un chiffre d'affaires des articles en promotion dans la table des faits. [Bru&Cam,2004].

### III.3.3.1 Limites :

Un cas d'utilisation est une manière d'utiliser ou d'interagir avec le système. Il est constitué de différents scénarios décrivant les différents cas de figure. C'est la modélisation fondée sur les scénarios qui fait la force d'UML et qui offre une bonne compréhension de « ce qui se passe » dans le système. Or il n'y a pas ici de système à proprement parler, seule l'identification de l'indicateur et des ambiguïtés entre règles de gestion nous intéressent. Les tâches fondamentales que souhaite faire l'acteur se résument ici à consulter des données. [Bru&Cam,2004].

### III.3.3.2 Expression des règles de gestion :

Etant donné que les règles de gestions constituent le cas d'utilisation, on pourrait imaginer, le calcul de la règle de gestion en tant que scénario avec l'expression des cas particuliers dans des scénarios alternatifs ? Ce qui permettrait ainsi de tenir compte de l'existant. [Bru&Cam,2004].

### III.3.3.3 Modélisation du métier :

On s'intéresse à *Pourquoi l'utilisateur veut accéder à tel ou tel indicateur* et modéliser sous forme de scénarios l'analyse possible via des indicateurs et des axes successifs. Les scénarios serviraient ici la modélisation du métier de l'acteur. Imaginons un responsable de fabrication qui gère ses effectifs en effectuant d'une part un calcul de

charge de travail vis à vis des commandes fournies par le service planification et un calcul sur les ressources humaines disponibles. La comparaison de ces deux indicateurs lui permet de prendre la décision d'embauche d'intérimis par exemple. On comprend ainsi pourquoi les indicateurs entrent en jeu, une expression d'un tel besoin pourrait se faire sous forme de diagrammes de séquence. Il s'agit bien ici d'une règle métier qui combinent d'autres règles. [Bru&Cam,2004].

### III.3.4 Modélisation multidimensionnelle

Sur un schéma UML il est possible d'indiquer la nature additive, semi-additive ou non additive des indicateurs. Cela a un intérêt tout particulier en phase de design. La possibilité de définir ces types de relations offre un intérêt dans la représentation de relations historisées entre entités (plus grande légibilité qu'un schéma E/R). [Bru&Cam,2004].

#### III.3.4.1 Etoiles en UML

Il n'est actuellement pas possible de générer d'étoiles à partir de diagrammes UML avec les outils du marché. Des travaux pour définir un modèle UML capable d'assurer la traduction existent ; il est sûr que ces outils se développeront rapidement. [Bru&Cam,2004].

#### III.3.4.2. Modélisation des hiérarchies : [Bru&Cam,2004].

C'est un domaine de recherche actif. Il n'est pas nécessaire de rentrer dans le débat de l'utilité d'un modèle conceptuel de base de données multidimensionnelle.

L'UML permet de modéliser les niveaux de hiérarchies par des classes et de définir des relations spécifiques au multidimensionnel, tel que le *drill-down* sur une classe Cube, qui dérive des concepts objets (agrégation, spécialisation).

Le modèle UML en permettant des types de relations plus fins que les modèles relationnels donnent plus d'informations visuelles à l'utilisateur. Cela peut être pertinent pour des utilisateurs experts, car une hiérarchie explicite et des informations sur la modélisation sont des atouts pour écrire et interpréter des requêtes ad-hoc.

Dans des problématiques de performances, où l'éclatement des tables de dimensions est nécessaire, la modélisation UML permet de construire un modèle plus complexe (dimensions *spécialisant* d'autres dimensions par exemple).

Exemple : un modèle conceptuel en flocon peut faire apparaître une dimension (sous forme d'une classe Personne) qui se ramifie en Employé et Client (sous-classes). Il est possible de générer des requêtes impliquant simultanément les deux entités Employé et Client (l'usage d'un attribut dans Personne est plus limitatif et complique la navigation dans les données). Cela permet entre autre la parallélisation des requêtes (un processeur traitant la partie de requête impliquant Employé, l'autre Client).

### III.3.5 Synthèse

#### III.3.5.1 L'UML Aujourd'hui :

Tableau IV.2 : Avantages et Inconvénients d'UML. [Bru&Cam,2004].

Avantages	Inconvénients
Capture des règles métiers sur les schémas de modélisation de données.	Les notions objets ne sont pas utiles sur une base de donnée relationnelle.
Modélisation des états d'un Entrepôt de données et de la reprise sur incidents.	L'UML n'est que faiblement intégré dans les outils de conception de base de données.
Modélisation des hiérarchies en flocons explicite.	Il n'y a pas de sémantique « universelle ».
Capture des besoins utilisateurs.	Peu d'intérêt du point fort du modèle conceptuel UML : la modélisation par scénario.

#### III.3.5.2 Diagramme UML par besoin décisionnel

Dans le tableau qui suit, les deux auteurs Brunot et Cambazard représentent le degré de nécessité des diagrammes UML dans le système décisionnel, une note de performance a été donnée à chaque diagramme d'UML ainsi que la signification de la note qui vient ci-dessous :

1 : apport significatif ; 2 : apport limité ; 3 : aucun apport ; 4 : perte de temps

Tableau IV.3 : Diagramme UML par besoin décisionnel. [Bru&Cam,2004].

Diagramme UML	Utilisation Possible en décisionnel	Intérêt
Diagramme de classe	Modèle conceptuel de base de données relationnelle	3
	Modèle physique de base de données	4
	Modèle conceptuel de base de données multidimensionnelle	2
Diagramme de séquence	Modélisation des flux de données pour operational data store / Entrepôt de données / Data marts.	2
	Modélisation des règles de gestion, illustration de calcul d'indicateurs (use cases)	2
Diagramme d'objet	Pas d'utilisation	4
Diagramme de collaboration	Modélisation détaillée des flux	3
Diagramme de déploiement	Architecture du système décisionnel d'un point de vue physique.	1
Diagramme de composants	Représentation des éléments de l'architecture physique (les différentes applications qui viennent utiliser le système, les postes utilisateurs)	1
Diagramme d'états-transitions	Représentation de la disponibilité du système	1
Diagramme de cas d'utilisations	Enoncé du besoin utilisateur	2
Diagramme d'activité	Règles de reprise sur incident	1

### III.4 Résumé :

L'apport de l'UML dans le système décisionnel a été résumé dans le tableau de synthèse IV.3. Après mûre réflexion nous avons opté pour certains diagrammes d'UML afin de répondre aux besoins de chaque phase. Les diagrammes choisis et appliqués pour chaque phase sont résumés dans le tableau qui suit :

**Tableau IV.4 :** Citation des principaux Diagrammes UML utilisé dans les différentes phases du cycle de vie. [Bru&Cam,2004].

Phase du cycle de vie	Diagramme Utilisé
Phase Définition des Besoin	Le diagramme des cas d'utilisation.
Phase Modélisation.	
Phase ETL.	Le diagramme des Cas d'utilisation. Le diagramme de séquence. Le diagramme d'état transition. Le diagramme d'activité.
Phase Reporting.	Le diagramme des cas d'utilisation. Le diagramme de séquence.
Phase déploiement.	Le diagramme des composant. Le diagramme de déploiement.

### III.5 Conclusion

L'UML apporte :

- La vue des acteurs
- La structuration des besoins sans chercher l'exhaustivité

Cependant, on notera l'absence du point fort du modèle conceptuel d'UML, les *scénarios* : cela découle que le besoin n'est pas ici réellement la modélisation d'un système ni de processus métiers mais l'identification des données, les indicateurs et leurs axes.

L'UML seul ne présente pas d'intérêt fort au niveau de la modélisation des besoins décisionnels. Il ne serait intéressant que dans la mesure où les cas d'utilisations seraient intégrés automatiquement avec la modélisation, ce qui n'est pas réalisable sans un modèle au-dessus d'UML.

Par rapport au besoin de l'entrepôt de données en terme de modélisation d'architecture, l'approche d'UML sous la forme de diagrammes de composants et de déploiement n'a rien d'exceptionnelle mais répond aux besoins de l'expression des technologies, des flux et des contraintes sur les éléments de l'architecture.

La modélisation multidimensionnelle reste très peu développée en UML, bien qu'il y ait de fortes potentialités dans ce domaine. Nous retenons que l'apport de l'UML dans la conception de base de données multidimensionnelle réside dans la description formelle des besoins avec les cas d'utilisation. La nécessité de définir soi-même sa sémantique UML en multidimensionnel limite néanmoins la portabilité des solutions.

**PARTIE V**



**Conception de l'Entrepôt de données**

Dans cette partie

**Chapitre I : Phase définition des besoins.**

**Chapitre II : Modélisation dimensionnelle.**

**Chapitre III : Conception des éléments de la zone de préparation des données.**

**Chapitre IV : Reporting.**

**Chapitre V : Schéma de mise en œuvre et Déploiement.**

**PARTIE V : Conception de l'Entrepôt de données**

**Phase définition des besoins.**

Dans ce chapitre

**Définition des besoins.**

**La méthode à suivre :**

- Préparation des entretiens.
- Questions aux dirigeants.
- Questions aux responsables ou aux analystes.

**Les cas d'utilisation.**

**Description des cas d'utilisation :**

- Processus CCP.
- Processus Monétique.
- Processus Compensation.
- Processus Acquéreur.
- Processus Emetteur.
- Processus Front Office.
- Activité Panne.

**Conclusion.**

Dans cette partie, nous allons développer et détailler chaque phase du cycle de vie suivi par l'étude, chaque phase indépendamment dans son chapitre respectif.

## V. CHAPITRE I : PHASE DEFINITION DES BESOINS.

### I.1 Définition des besoins

La phase de spécification des besoins a un impact sur tout le processus décisionnel, Ralph Kimball place les besoins au centre de l'univers du processus décisionnel.

En effet la réussite d'un projet décisionnel, repose en majorité sur la bonne compréhension des utilisateurs et de leurs besoins, qui seront définis grâce aux méthodes de collecte d'information.

Dans cette phase, nous devons prendre les facteurs clés qui conduisent l'entreprise à vouloir définir de manière efficace ses besoins, et les traduire pour les intégrer lors de la conception. [Kim&all,2005]

### I.2 La méthode à suivre :

Afin de mieux appréhender les besoins, il faudrait commencer par communiquer avec les utilisateurs, cette recommandation peut paraître évidente, cependant de nombreuses entreprises explorent d'autres possibilités. Il est vrai que nous nous sentons dans notre élément face à des structures de données ou lorsque nous discutons avec les administrateurs de base de données et les experts en systèmes sources. D'autres part, les utilisateurs peuvent se montrer intimidants ou revendicatifs. En dépit des éventuelles réticences, il faudrait commencer par les utilisateurs concernés par l'étude car ils occupent le centre de l'univers de l'entrepôt de données. [Kim&all,2005]

Il existe deux méthodes principales de collecte des besoins et d'étude des données opérationnelles sous-jacentes : les entretiens et les réunions de concertation.

Pour notre cas d'étude nous avons opté pour les entretiens individuels.

#### I.2.1 Préparation des entretiens : [Kim&all,2005].

- Sélection des personnes à interroger.
- Préparation d'un questionnaire.
- Préparer les utilisateurs avant l'entretien : Sensibilisation (Informer et motiver les acteurs de l'entreprise afin qu'ils s'engagent dans le processus de mise en place du projet décisionnel).
- Faire un document de consolidation des entretiens.

Les questions que nous avons posées aux décideurs, proposées par Ralph Kimball, sont de deux types, Les questions posées aux dirigeants nous donnent une vue globale sur ce qui se passe dans l'entreprise, tandis que Les questions posées aux analystes nous donnent des informations détaillées.

Voici un échantillon de questions que nous avons utilisées pendant notre étude des besoins.

**I.2.2 Questions aux dirigeants : [Kim&all,2005]**

- Quels sont les objectifs de votre secteur d'activités ? Qu'essayez vous d'accomplir ?
- Quelles sont vos méthodes pour évaluer la réussite ? Comment savez vous que vous êtes sur la bonne voie ? Quelle est la fréquence des évaluations ?
- Quels problèmes affrontez-vous principalement aujourd'hui ? Qu'est ce qui vous empêche d'atteindre les objectifs ? Quelles en sont les conséquences sur l'entreprise ?
- Comment identifiez vous ou anticipez-vous les problèmes ? Comment identifiez vous l'exception dans votre entreprise ? Quelles sont les perspectives ouvertes pour l'amélioration de l'accès à l'information ? Serai-elle susceptible d'avoir des conséquences significatives sur votre entreprise ? Quel en serait l'impact financier ? Si vous pouviez ..... , de quelle manière votre activité en serait-elle affectée ?
- De quelle manière envisagez-vous une meilleure exploitation de l'information ? Pensez-vous que vos collaborateurs auront un accès à l'information ?

**I.2.3 Questions aux responsables ou aux analystes : [Kim&all,2005]**

- Quels sont les objectifs de votre service ?
- Qu'essayez-vous de réaliser ? Comment vous y prenez-vous pour atteindre cet objectif ?
- Quelles sont vos métriques d'évaluation de la réussite ? Comment savez-vous que vous vous y prenez bien ? Quelle est la fréquence des évaluations ?
- Quels problèmes affrontez-vous en ce moment ? Qu'est-ce qui limite aujourd'hui vos résultats ?
- Quel type d'analyse régulières effectuez-vous actuellement ? Avec quel type de données ? Comment obtenez-vous les données ? Que faites-vous ensuite de l'information obtenue ?
- Quel genre d'analyse souhaiteriez-vous réaliser ? Existe-t-il des améliorations à apporter à vos méthodes/processus actuels ?
- Actuellement, quel type d'analyse réalisez-vous ? Qui fait la demande de ces analyses ? Quel en est l'usage ? Combien de temps cela prend-il ? Avez-vous le temps de donner suite ?
- De quelles possibilités d'analyse souhaiteriez-vous disposer ? -
- De quelles données historisées avez-vous besoin ?
- Quels sont les états actuellement utilisés ? Quelles en sont les informations importantes ? Quel usage faites-vous des informations ?

A la fin des interviews, nous sommes passé à la représentation de ces besoins en utilisant UML.

Le diagramme des cas d'utilisations est le plus adaptée à cet effet, et c'est ce que nous avons choisi pour exprimer les besoins des utilisateurs.

### **I.3 Les cas d'utilisation :**

Le cahier des charges de notre système est de réaliser un outil d'aide à la décision pour la gestion de la Monétique et CCP.

Ses objectifs se décomposent en quatre parties :

- Etude procédurale des processus.
- Visualisation des indicateurs de performance.
- Administration de l'outil de décision.
- Gestion de l'alimentation de l'entrepôt de données.

L'étude des cas d'utilisation nous a permis de formaliser et de détailler les besoins des utilisateurs s'inscrivant dans l'objectifs cités ci-dessus.

Chacun de ces objectifs a été décrit dans la phase qui lui est associée.

L'analyse s'est faite par la création d'un ensemble de cas d'utilisation construit et validé en collaboration avec les utilisateurs et les experts du domaine.

### **I.4 Description des cas d'utilisation :**

La phase définition des besoins se résume à indiquer les indicateurs par axes d'analyse ainsi que les utilisateurs qui les exploiteront, ce qui veut dire qu'à cette phase deux objectifs seront traités :

- Visualisation des indicateurs de performances.
- Etude procédurale des processus.

I.4.1 Processus CCP :

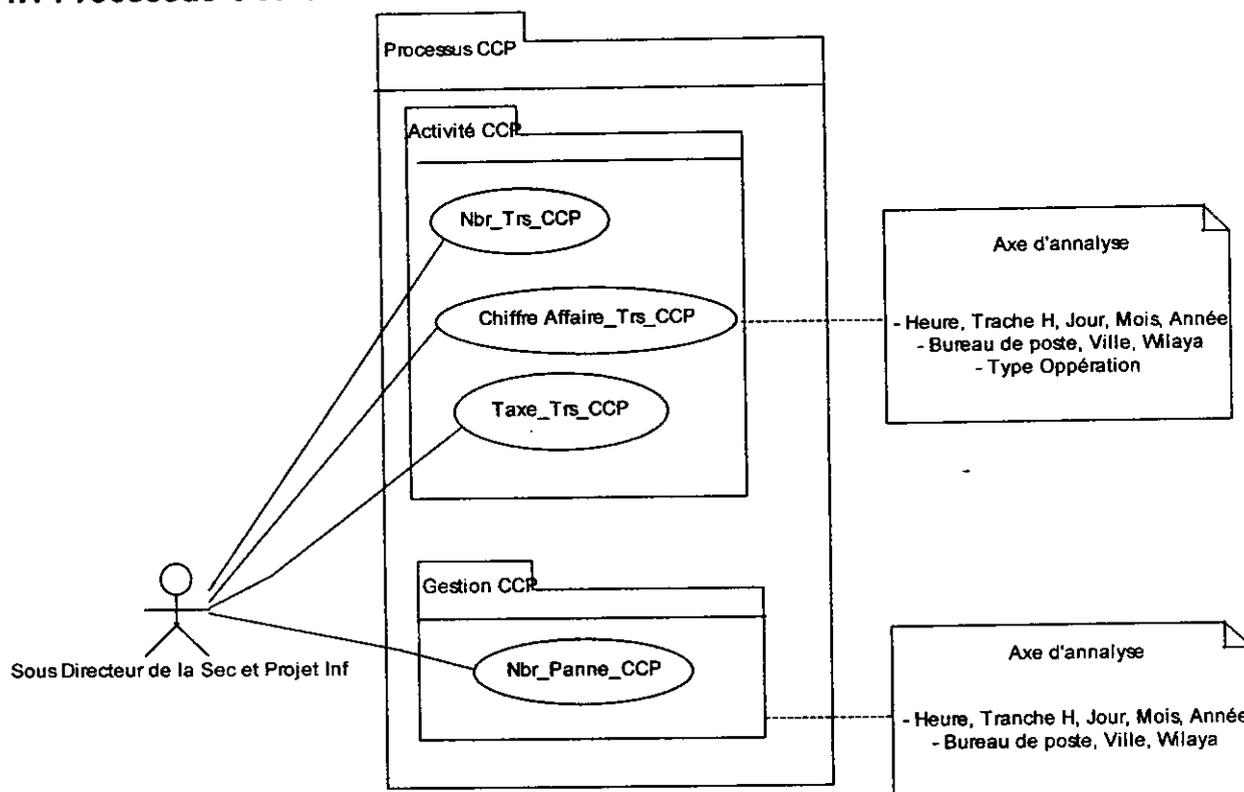


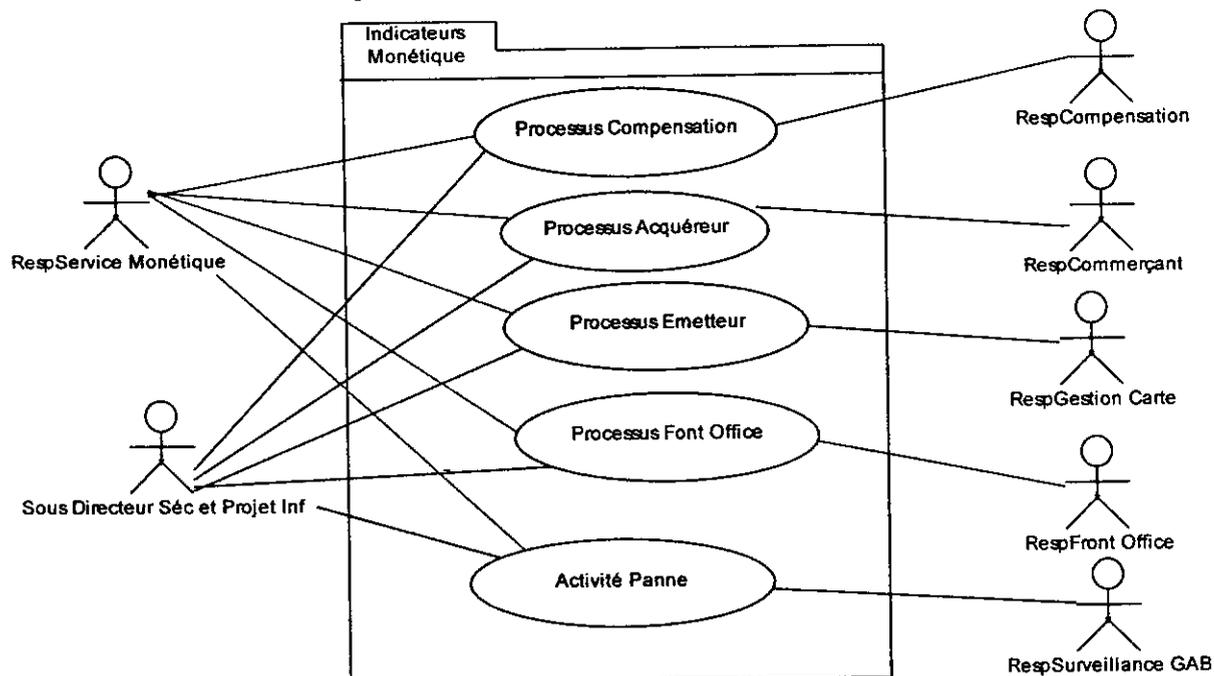
Diagramme V.1 : Diagramme des Cas d'utilisation du Processus CCP

Description :

Acteur	Indicateur	Activité	Action	Axes d'analyse
1. Sous directeur sécurité et projet informatique.	Nombre Transaction CCP	CCP	Visualise	<ul style="list-style-type: none"> <li>- Heure, tranche horaire, jour, mois année.</li> <li>- Bureau de poste, ville, wilaya.</li> <li>- Type Opération.</li> </ul>
	Chiffre d'affaires transaction CCP			
	Taxe transaction CCP			
	Nombre de panne CCP	Gestion CCP	Visualise	<ul style="list-style-type: none"> <li>- Heure, tranche horaire, jour, mois année.</li> <li>- Bureau de poste, ville, wilaya.</li> </ul>

Comme le tableau le résume, l'acteur : «Sous directeur sécurité et projet informatique » visualise les indicateurs du processus CCP, avec choix d'indication des axes d'analyses qui les concernent.

**I.4.2 Processus Monétique :**



**Diagramme V.2 : Diagramme des Cas d'utilisation du Processus Monétique**

**Description :**

Acteur	Processus	Action
1. Sous directeur sécurité et projet informatique. 2. Responsable service monétique. 3. Responsable compensation.	Compensation	Accès
1. Sous directeur sécurité et projet informatique. 2. Responsable service monétique. 3. Responsable commerçant.	Acquéreur	Accès
1. Sous directeur sécurité et projet informatique. 2. Responsable service monétique. 3. Responsable gestion carte.	Emetteur	Accès
1. Sous directeur sécurité et projet informatique. 2. Responsable service monétique. 3. Responsable front office.	Front office	Accès
1. Sous directeur sécurité et projet informatique. 2. Responsable service monétique. 3. Responsable surveillance des GABs.	Panne	Accès

Comme le tableau le résume, les acteurs « Sous directeur de la sécurité et des projets informatique » & « Responsable services monétique » ont les droits d'accès au processus cité dans le tableau, ainsi que chaque responsable du processus lui-même.

- Processus Compensation. A les droits d'accès aussi le « Responsable Compensation ».

- Processus Acquéreur. A les droits d'accès aussi le « Responsable Commerçant ».
- Processus Emetteur. A les droits d'accès aussi le « Responsable Gestion Carte ».
- Processus Front Office. A les droits d'accès aussi le « Responsable Front Office ».
- Activité Panne. A les droits d'accès aussi le « Responsable Surveillance GAB ».

### I.4.3 Processus Compensation :

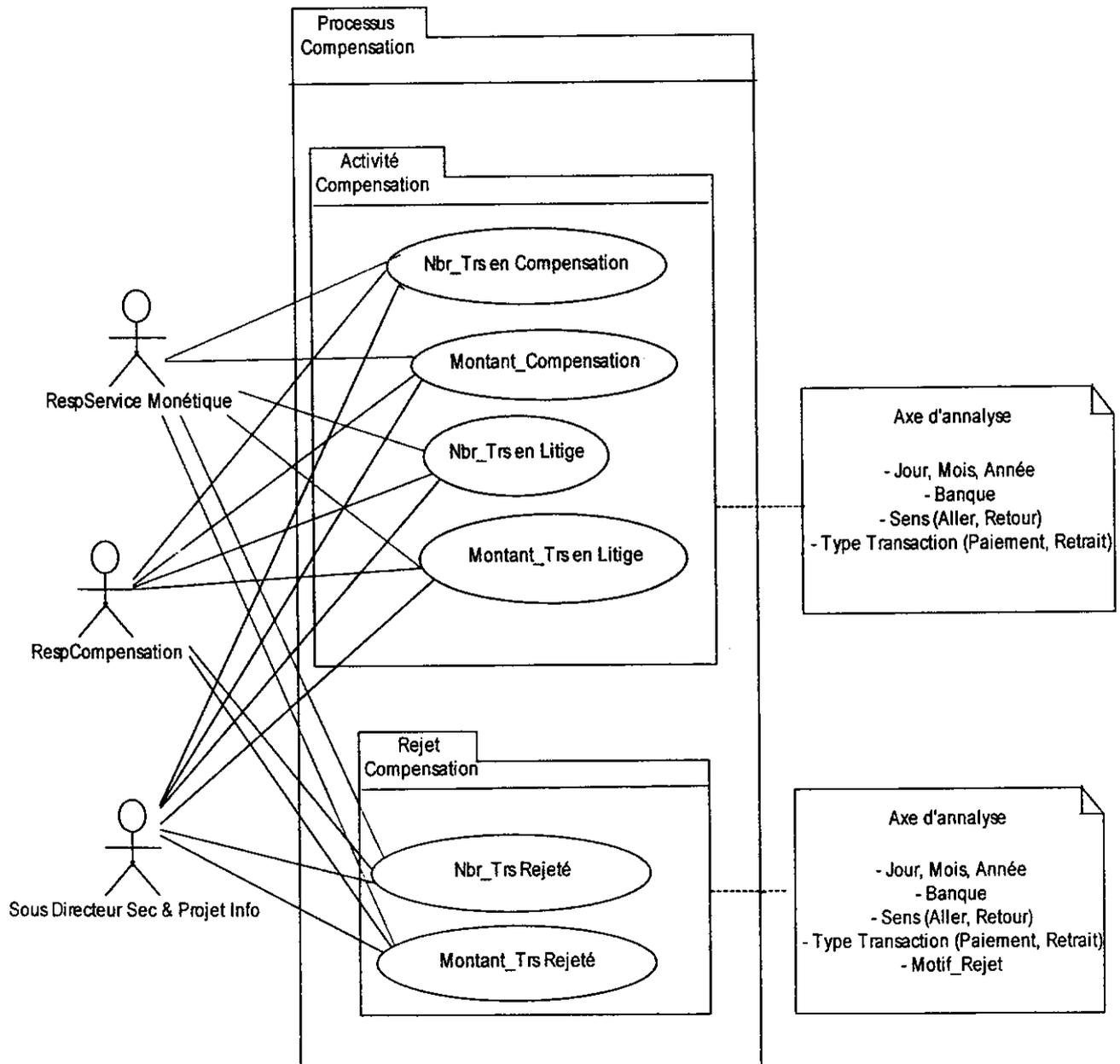


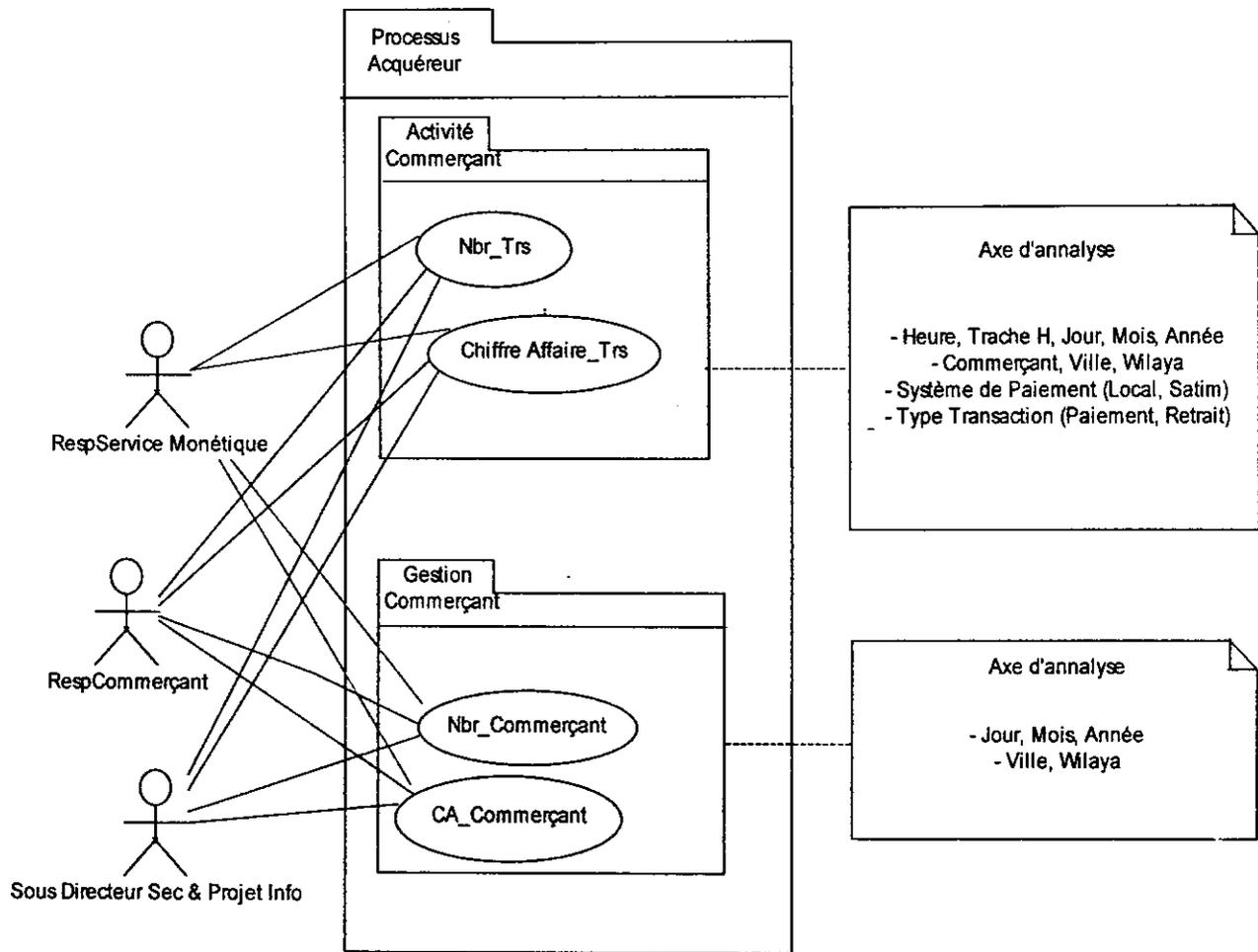
Diagramme V.3: Diagramme des Cas d'utilisation du Processus Compensation.

**Description (Diagramme V.3) :**

Acteur	Indicateur	Activité	Action	Axes d'analyse
1. Sous directeur sécurité et projet informatique.	Nombre Transaction en compensation	Compensation	Visualise	<ul style="list-style-type: none"> <li>- Jour, mois année.</li> <li>- Banque.</li> <li>- Sens (Aller, Retour)</li> <li>- Type Transaction (Retrait seulement)</li> </ul>
	Montant en compensation			
	Nombre transaction en litige			
	Montant transaction en litige			
2. Responsable monétique	Nombre transaction rejetée	Rejet compensation	Visualise	<ul style="list-style-type: none"> <li>- Jour, mois année.</li> <li>- Banque.</li> <li>- Sens (Aller, Retour)</li> <li>- Type Transaction (Retrait seulement)</li> <li>- Motif rejet compensation.</li> </ul>
3. Responsable compensation	Montant transaction rejetée			

Comme le tableau le résume, les acteurs : «Sous directeur sécurité et projet informatique », «Responsable service monétique » et «Responsable Compensation » ont le droit de visualiser les indicateurs du processus Compensation, avec choix d'indication des axes d'analyses qui les concernent.

**I.4.4 Processus Acquéreur :**



**Diagramme V.4 : Diagramme des Cas d'utilisation du Processus Acquéreur**

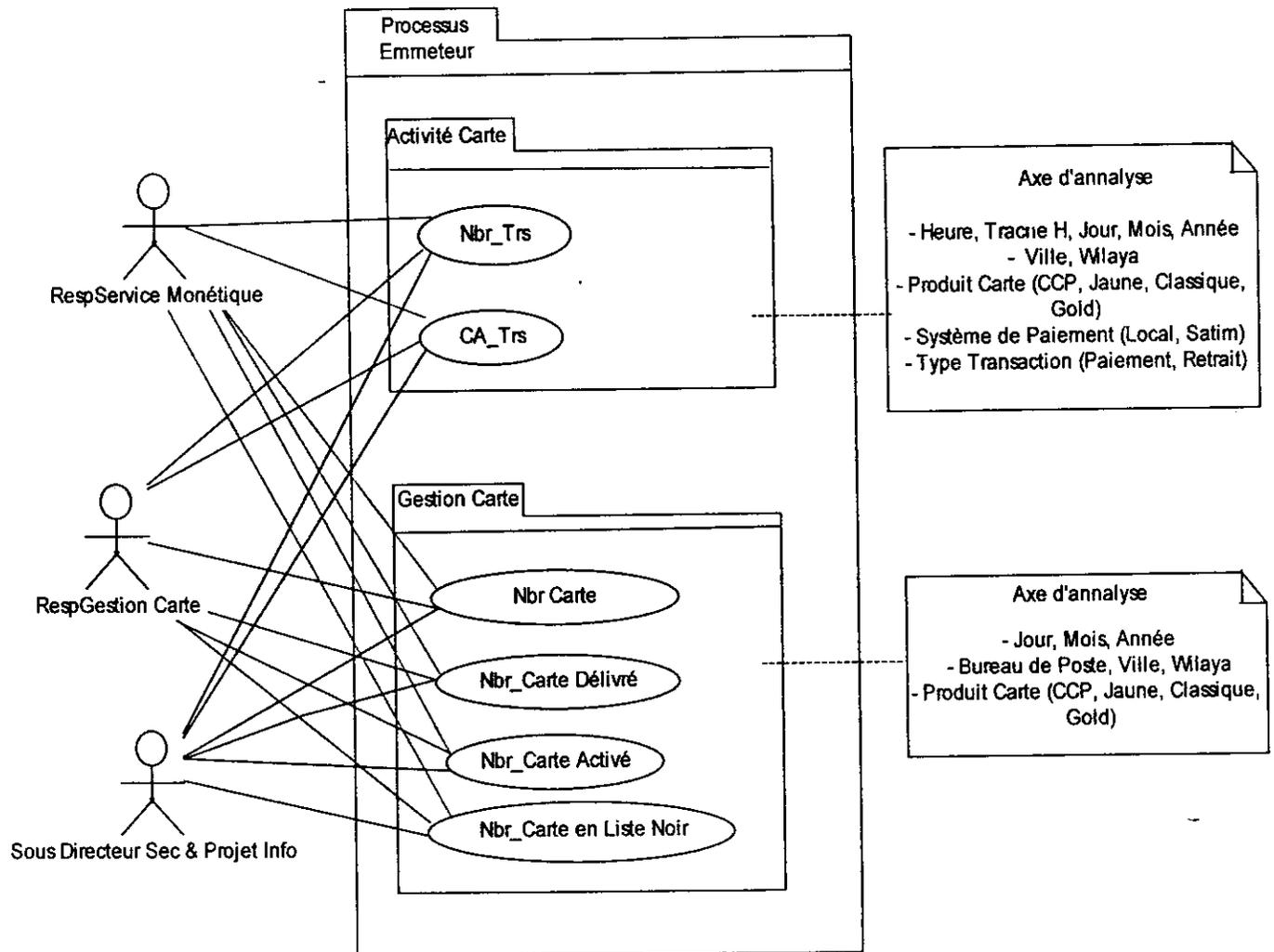
**Description :**

Acteur	Indicateur	Activité	Action	Axes d'analyse
1. Sous directeur sécurité et projet informatique.	Nombre Transaction	Commerçant	Visualise	<ul style="list-style-type: none"> <li>- Heure, tranche horaire, jour, mois année.</li> <li>- Commerçant, ville, wilaya.</li> <li>- Système de paiement (Local seulement)</li> <li>- Type Transaction (Retrait seulement)</li> </ul>
	Chiffre d'affaires transaction			
2. Responsable monétaire				
3. Responsable commerçant	Nombre commerçant	Gestion	Visualise	<ul style="list-style-type: none"> <li>- Jour, mois, année</li> <li>- Ville, wilaya.</li> </ul>
	Chiffre d'affaires commerçant	commerçant		

**Description (suite Diagramme V.4):**

Comme le tableau le résume, les acteurs : « Sous directeur sécurité et projet informatique », « Responsable service monétique » et « Responsable commerçant » ont le droit de visualiser les indicateurs du processus Acquéreur, avec choix d'indication des axes d'analyses qui les concernent.

**1.4.5 Processus Emetteur :**



**Diagramme V.5 : Diagramme des Cas d'utilisation du Processus Emetteur.**

**Description (Diagramme V.5):**

Acteur	Indicateur	Activité	Action	Axes d'analyse
1. Sous directeur sécurité et projet informatique.	Nombre Transaction	Carte	Visualiser	<ul style="list-style-type: none"> <li>- Heure, tranche horaire, jour, mois année.</li> <li>- Ville, wilaya.</li> <li>- Système de paiement (Local seulement)</li> <li>- Type Transaction (Retrait seulement)</li> <li>- Produit carte</li> </ul>
	Chiffre d'affaires transaction			
3. Responsable gestion carte	Nombre Carte	Gestion carte	Visualiser	<ul style="list-style-type: none"> <li>- Jour, Mois, Année.</li> <li>- Bureau de poste, ville, wilaya.</li> <li>- Produit carte.</li> </ul>
	Nombre carte délivré			
	Nombre de carte activé			
	Nombre de carte en liste noir.			

Comme le tableau le résume, les acteurs : «Sous directeur sécurité et projet informatique », «Responsable service monétique » et «Responsable gestion carte » ont le droit de visualiser les indicateurs du processus Emetteur, avec choix d'indication des axes d'analyses qui les concernent.

I.4.6 Processus Front Office :

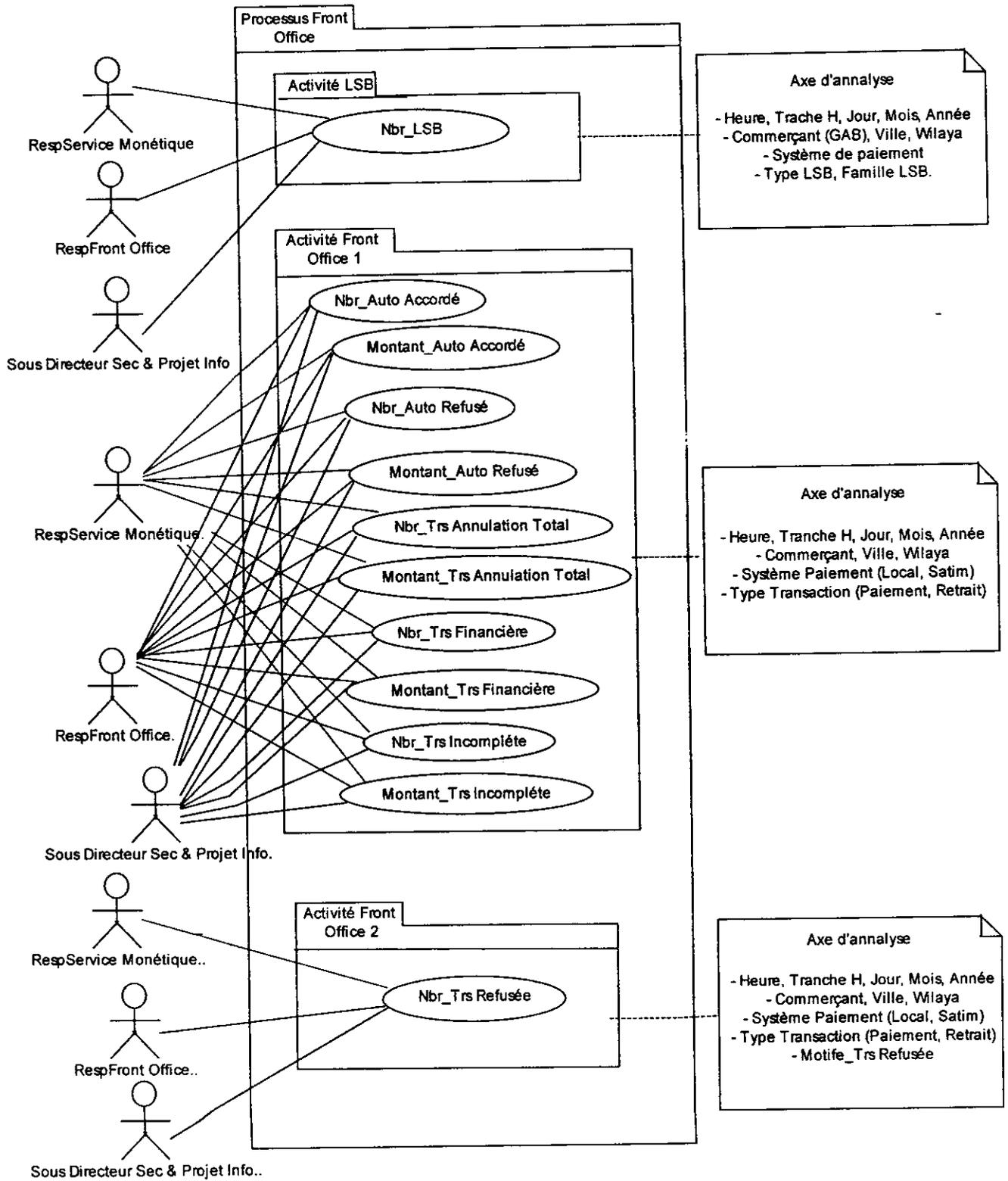


Diagramme V.6: Diagramme des Cas d'utilisation du Processus Front Office

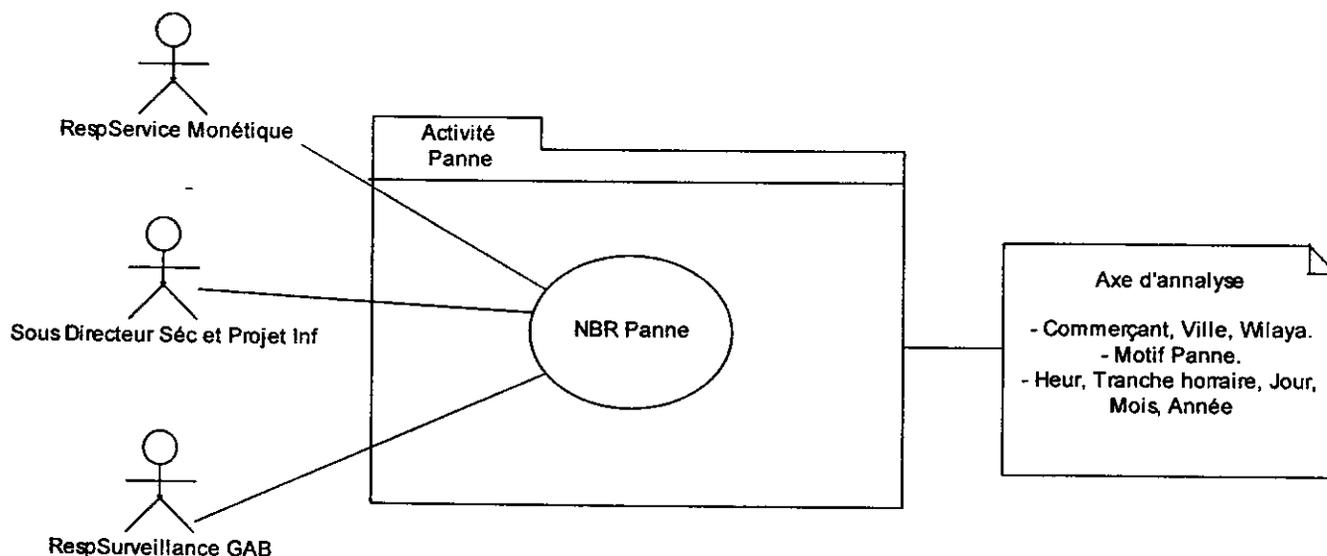
**Description (Diagramme V.6):**

Acteur	Indicateur	Activité	Action	Axes d'analyse
1. Sous directeur sécurité et projet informatique.	Nombre LSB	LSB	Visualise	<ul style="list-style-type: none"> <li>- Heure, tranche horaire, jour, mois année.</li> <li>- Commerçant, ville, wilaya.</li> <li>- Système de paiement (Local seulement)</li> <li>- Type LSB, Famille LSB.</li> </ul>
	Nombre autorisation accordée Montant autorisation accordée Nombre autorisation refusée Montant autorisation refusée Nombre transaction en annulation totale Montant transaction en annulation totale Nombre transaction financière Montant transaction financière Nombre transaction incomplète Montant transaction incomplète	Front Office 1	Visualise	<ul style="list-style-type: none"> <li>- Heure, tranche horaire, jour, mois année.</li> <li>- Commerçant, ville, wilaya.</li> <li>- Système de paiement (Local, Satim)</li> <li>- Type transaction (Retrait seulement)</li> </ul>
3. Responsable front office	Nombre transaction refusée	Front Office 2	Visualise	<ul style="list-style-type: none"> <li>- Heure, tranche horaire, jour, mois année.</li> <li>- Commerçant, ville, wilaya.</li> <li>- Système de paiement (Local seulement)</li> <li>- Type transaction (Retrait seulement)</li> <li>- Motif transaction en refus.</li> </ul>

**Description (suite Diagramme V.6):**

Comme le tableau le résume, les acteurs : «Sous directeur sécurité et projet informatique », «Responsable service monétique » et «Responsable front office » ont le droit de visualiser les indicateurs du processus Front office, avec choix d'indication des axes d'analyses qui les concernent.

**I.4.7 Activité Panne :**



**Diagramme V.7: Diagramme de Cas d'utilisation de l'Activité Panne.**

**Description :**

Acteur	Indicateur	Activité	Action	Axes d'analyse
1. Sous directeur sécurité et projet informatique. 2. Responsable monétique 3. Responsable Surveillance GAB	Nombre Panne	Panne	Visualiser	- Heure, tranche horaire, jour, mois année. - Motif Panne. - Commerçant, Ville, Wilaya.

Comme le tableau le résume, les acteurs : «Sous directeur sécurité et projet informatique », «Responsable service monétique » et «Responsable Surveillance GAB » ont le droit de visualiser l'indicateur de l'activité panne, avec choix d'indication des axes d'analyses qui les concernent.

**I.5 Conclusion :** les diagrammes des cas d'utilisation ont simplifié la compréhension des besoins des décideurs ainsi que leurs lectures.

## Chapitre II

### **PARTIE V : Conception de l'Entrepôt de données**

#### **Modélisation dimensionnelle.**

Dans ce chapitre

##### **Identification des Processus.**

##### **Description des modèles :**

- **Processus CCP.**
- **Processus Compensation.**
- **Processus Acquéreur.**
- **Processus Emetteur.**
- **Processus Front Office.**
- **Activité Panne.**
- **Processus Monétique.**
- **Entrepôt de données.**

##### **Conclusion.**

## V. CHAPITRE II : MODELISATION DIMENSIONNELLE

Une fois les besoins recensés et les données auditées, nous sommes prêts à lancer la conception logique et physique de notre entrepôt de données. La démarche consiste à transformer les données sources en structure de Data Warehouse. Les conceptions logiques et physiques représentent la pierre angulaire de l'entrepôt de données.

Nous avons introduit à la partie II, la modélisation dimensionnelle, les tables de faits, les dimensions et les différents modèles conceptuels (étoile, flocon) au niveau de la Partie IV, Chapitre I. C'est pourquoi nous n'allons pas nous y attarder. Nous avons opté pour la modélisation en étoile, la dé-normalisation des dimension pour l'optimisation du temps de recherche. L'intérêt du schéma en étoile est d'éviter des jointures complexes.

Nous passons directement à l'identification des Processus et à leur représentation dimensionnelle.

### II.1 Identification des Processus :

Les processus et activités qui devront être pris en compte par l'entrepôt ont été recensés lors de l'analyse et définition des besoins que nous rappelons ci-dessous :

- Processus CCP.
- Processus Compensation.
- Processus Acquéreur.
- Processus Emetteur.
- Processus Front Office.
- Activité Panne.
- Processus Monétique.

### II.2 Description des modèles :

Selon les droits attribués aux décideurs, les accès aux indicateurs leurs sont attribués. Ils peuvent les visualiser ainsi que choisir les axes d'analyses les concernant.

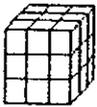
Les décideurs peuvent naviguer dans la granularité des tables de faits et la hiérarchie des dimensions selon leurs grés et leurs besoins, ainsi que les possibilités offerts par l'outil. Par exemple la hiérarchie de la table temps dans notre cas de figure va de Heure jusqu'à année, donc le décideur peut demander à visualiser un indicateur avec un détail qui remonte à heure, ou à voir le général qui peut concerner un mois, une saison ou même l'année.

Et cette option de naviguer entre les hiérarchies des dimensions est appliquée sur plusieurs dimensions qui vont être citées ci-dessous :

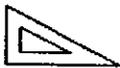
- Dimension Bureau de poste (Bureau de poste, Ville, Wilaya).
- Dimension Commerçant (Commerçant, Ville, Wilaya).
- Dimension Temps (Heure, Tranche horaire, Jour, Mois, Année).
- Dimension Ville (Ville, Wilaya).
- Dimension Type LSB (Type LSB, Famille LSB).

La dimension Temps est décomposée en deux dimensions, La dimension HEURE\_DIM qui contient Heure et Tranche horaire et La dimension JOUR\_MOIS\_ANNEE\_DIM qui contient les jours, mois et année ainsi que les *jours fériés, fiscale* ...etc. nous avons opté pour cette modélisation de la table Temps car c'est ce qui est utilisé sur le terrain de la modélisation dimensionnelle et cela répond au mieux aux besoins des décideurs, ainsi qu'aux besoins de l'étude car pour certains faits leur granularité dans le temps ne descend qu'à : jour.

Dans les diagrammes multidimensionnels qui vont être illustrés dans ce qui suit, nous avons menué les noms des tables de fait du mot : fait et du signe cube,



Et les tables de dimension du mot : DIM et du signe :



Et chaque fait est relié à au moins deux dimensions.

### II.2.1 Processus CCP :

Le processus CCP a à son actif deux activités :

- Activité CCP.
- Activité Panne.

Le décideur peut visualiser les indicateurs des deux activités ainsi que de choisir les axes d'analyse. Par exemple le nombre de transactions CCP effectuées sur tel Bureau de poste, concernant les retraits effectués tel Jour à tel heure.

Le processus CCP est illustré par les deux diagrammes V.8 et V.9 ci-dessous.

II.2.1.1 Activité CCP:

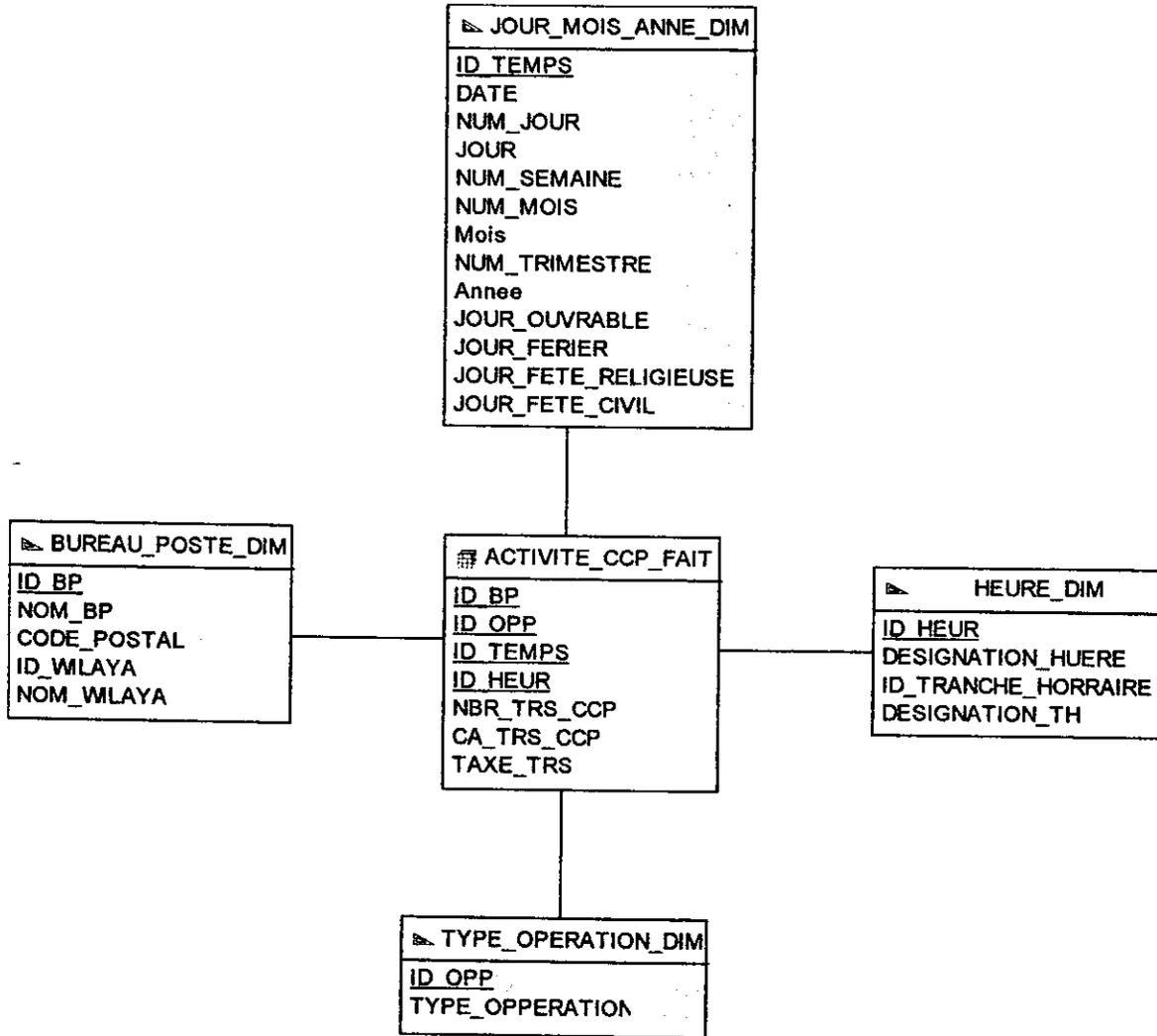


Diagramme V.8 : Diagramme Multidimensionnel : Activité CCP.

II.2.1.2 Panne CCP :

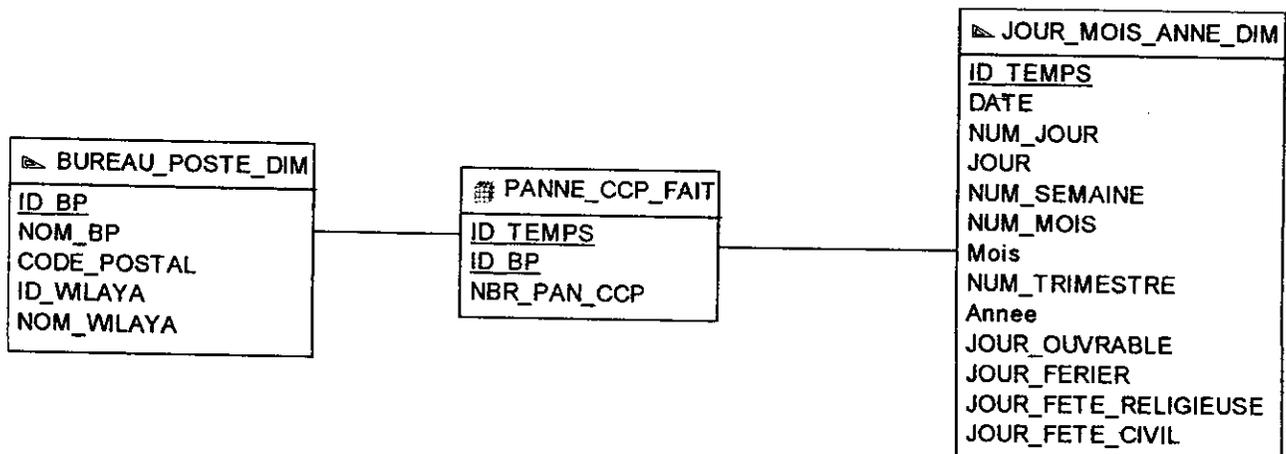


Diagramme V.9 : Diagramme Multidimensionnel : Panne CCP.

## II.2.2 Processus Compensation :

Le processus Compensation a à son actif deux activités :

- Activité Compensation.
- Activité Rejet Compensation.

Le décideur peut visualiser les indicateurs des deux activités ainsi que de choisir les axes d'analyse. Par exemple le nombre de transaction en Compensation effectuer dans tel Banque, concernant les retraits, dans le sens Aller, effectuer tel Jour.

Le processus Compensation est illustré par les deux diagrammes V.10 et V.11 ci-dessous.

### II.2.2.1 Activité Compensation :

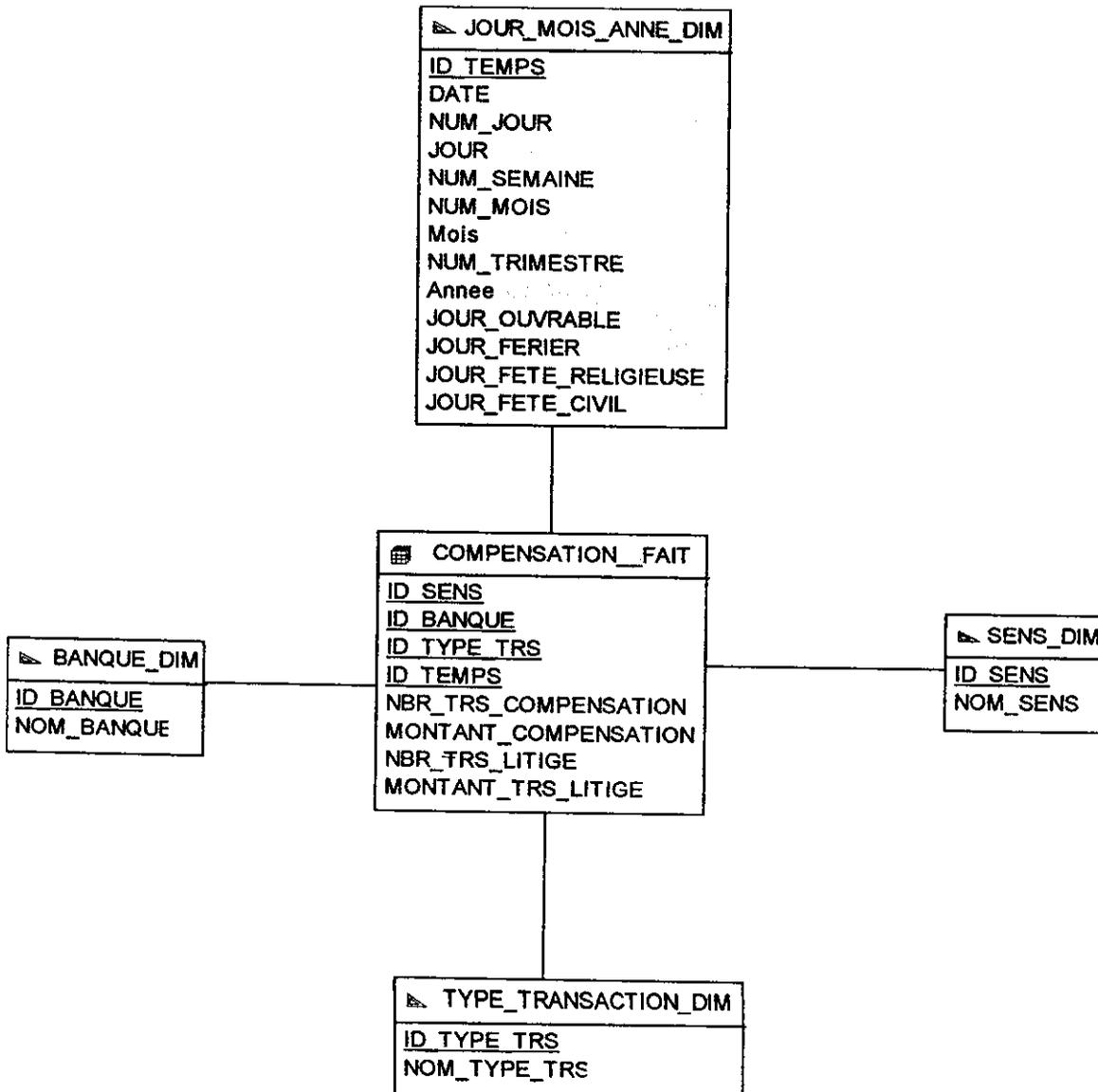
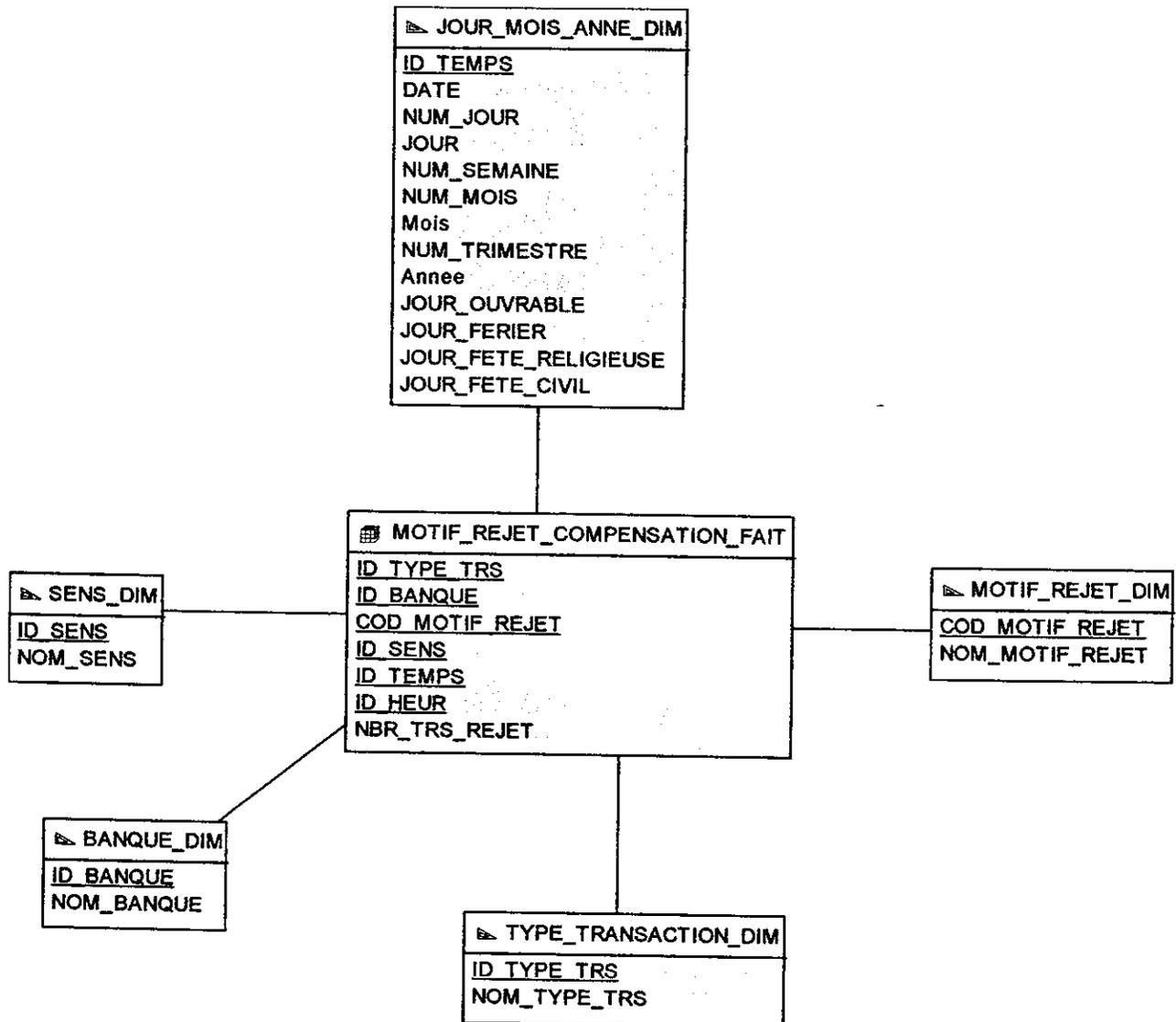


Diagramme V.10 : Diagramme Multidimensionnel : Activité Compensation.

**II.2.2.2 Activité Rejet Compensation :**



**Diagramme V.11 : Diagramme Multidimensionnel : Activité Rejet Compensation.**

**II.2.3 Processus Acquéreur :**

Le processus Acquéreur a à son actif deux activités :

- Activité Commerçant.
- Activité Gestion Commerçant.

Le décideur peut visualiser les indicateurs des deux activités, ainsi que de choisir les axes d'analyse. Par exemple le nombre de transactions effectuées sur tel Commerçant, concernant tel type de transaction (retraits), effectuée tel Jour à tel heure sur tel système de paiement (Local ou Satim).

Le processus Acquéreur est illustré par les deux diagrammes V.12 et V.13 ci-dessous.

II.2.3.1 Activité Commerçant :

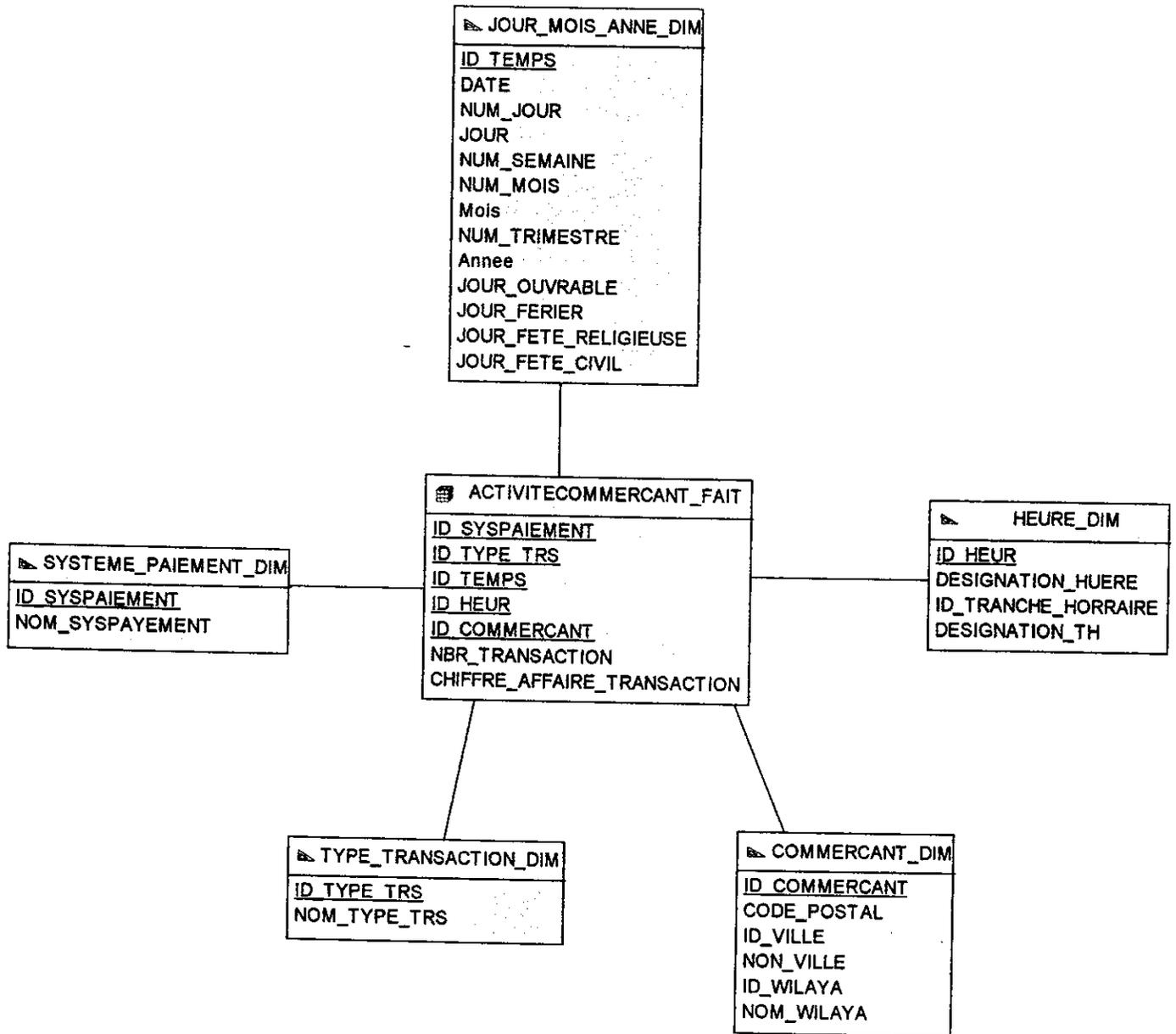


Diagramme V.12 : Diagramme Multidimensionnel : Activité Commerçant.

### II.2.3.2 Activité Gestion Commerçant :

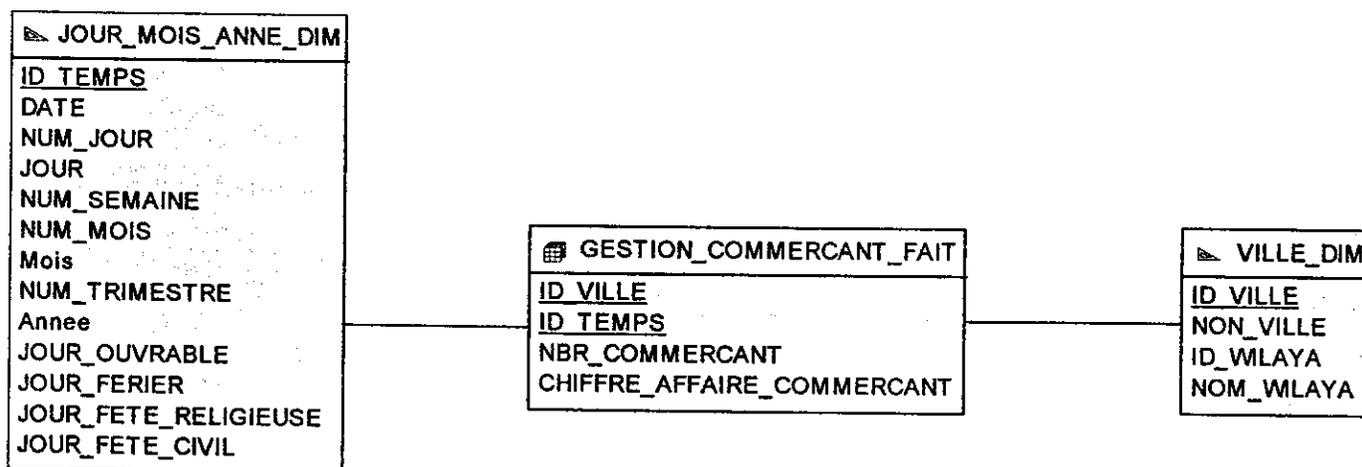


Diagramme V.13 : Diagramme Multidimensionnel : Activité Gestion Commerçant.

### II.2.4 Processus Emetteur :

- Le processus Emetteur a à son actif deux activités :
- Activité Carte.
  - Activité Gestion Carte.

Le décideur peut visualiser les indicateurs des deux activités ainsi que de choisir les axes d'analyse. Par exemple le nombre de transactions effectuées sur tel Type de Carte (Jaune, Classique ou Gold), concernant tel type de transaction (retraits ou paiement), à tel Jour. Où le nombre de cartes activées dans telle ville à tel jour.

Le processus Emetteur est illustré par les deux diagrammes V.14 et V.15 ci-dessous.

II.2.4.1 Activité Carte :

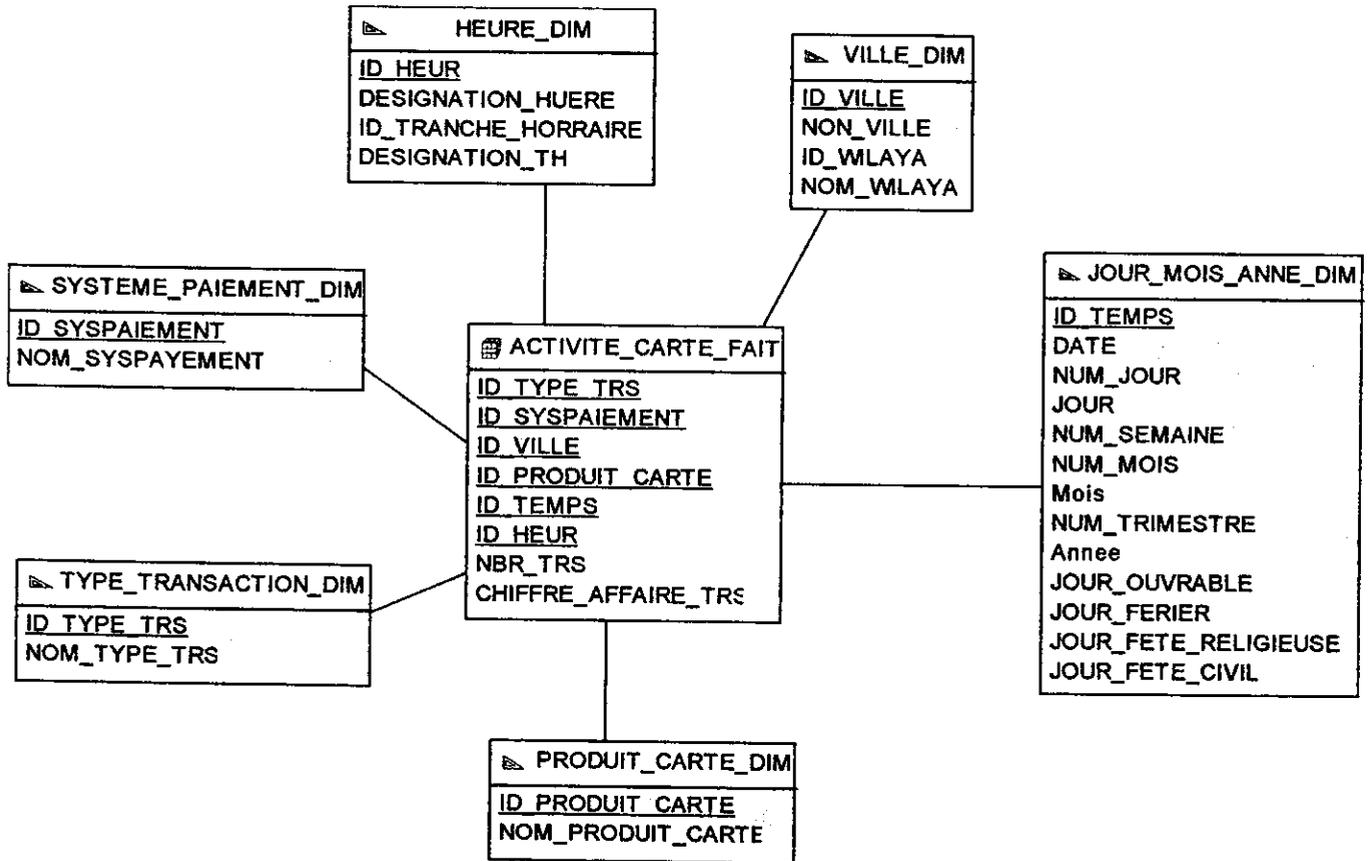


Diagramme V.14 : Diagramme Multidimensionnel : Activité Carte.

### II.2.4.2 Activité Gestion Carte :

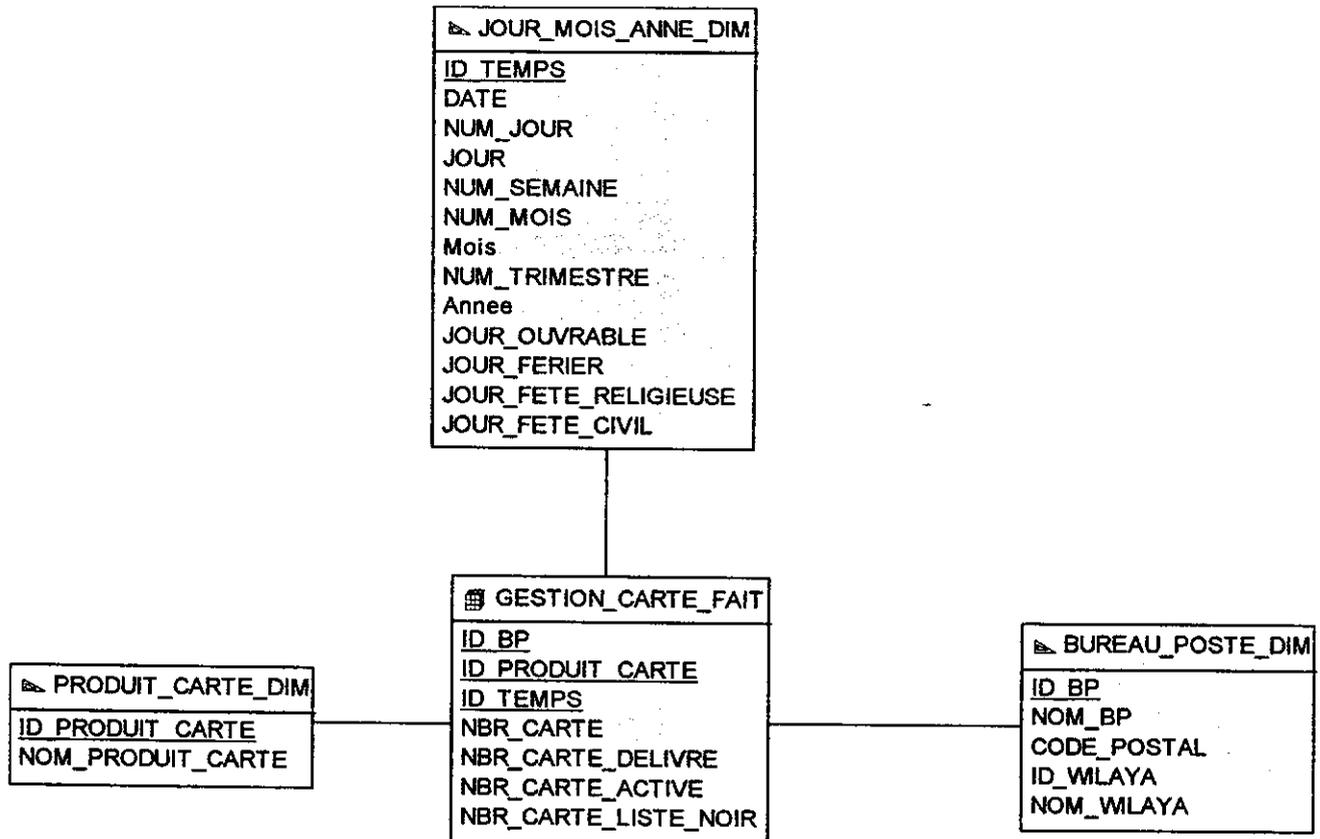


Diagramme V.15 : Diagramme Multidimensionnel : Activité Gestion Carte.

### II.2.5 Processus Front Office :

Le processus Front Office a à son actif Trois activités :

- Activité LSB.
- Activité Front Office1.
- Activité Front Office2.

Le décideur peut visualiser les indicateurs des trois activités ainsi que de choisir les axes d'analyse. Par exemple le nombre d'autorisation effectuées sur tel Commerçant, concernant tel type de transaction (retraits), effectuée tel Jour à telle heure sur tel système de paiement (Local ou Satim). Où le Nombre de Transactions effectuées sur tel type d'LSB (Demande\_Solde, Demande\_Chéquier, Demande\_RIB, Demande\_Mini\_Relevé), concernant telle famille d'LSB (Standard, Etendue) à telle heure de tel jour sur tel commerçant et tel système de paiement (Satim, Local).

Le processus Front Office est illustré par les trois diagrammes V.16, V.17 et V.18 ci-dessous.

II.2.5.1 Activité LSB :

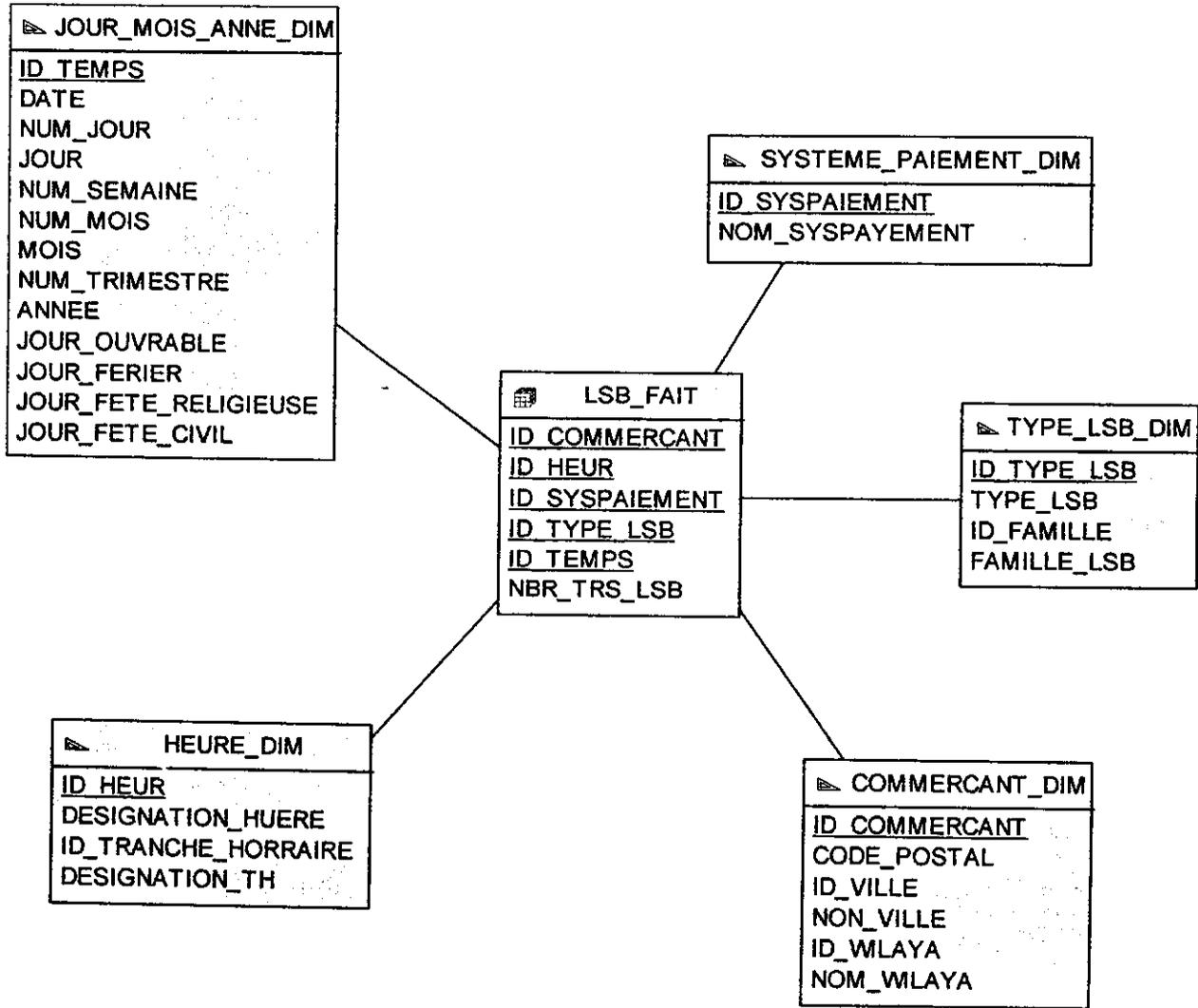


Diagramme V.16 : Diagramme Multidimensionnel : Activité LSB.

II.2.5.2 Activité Front Office 1 :

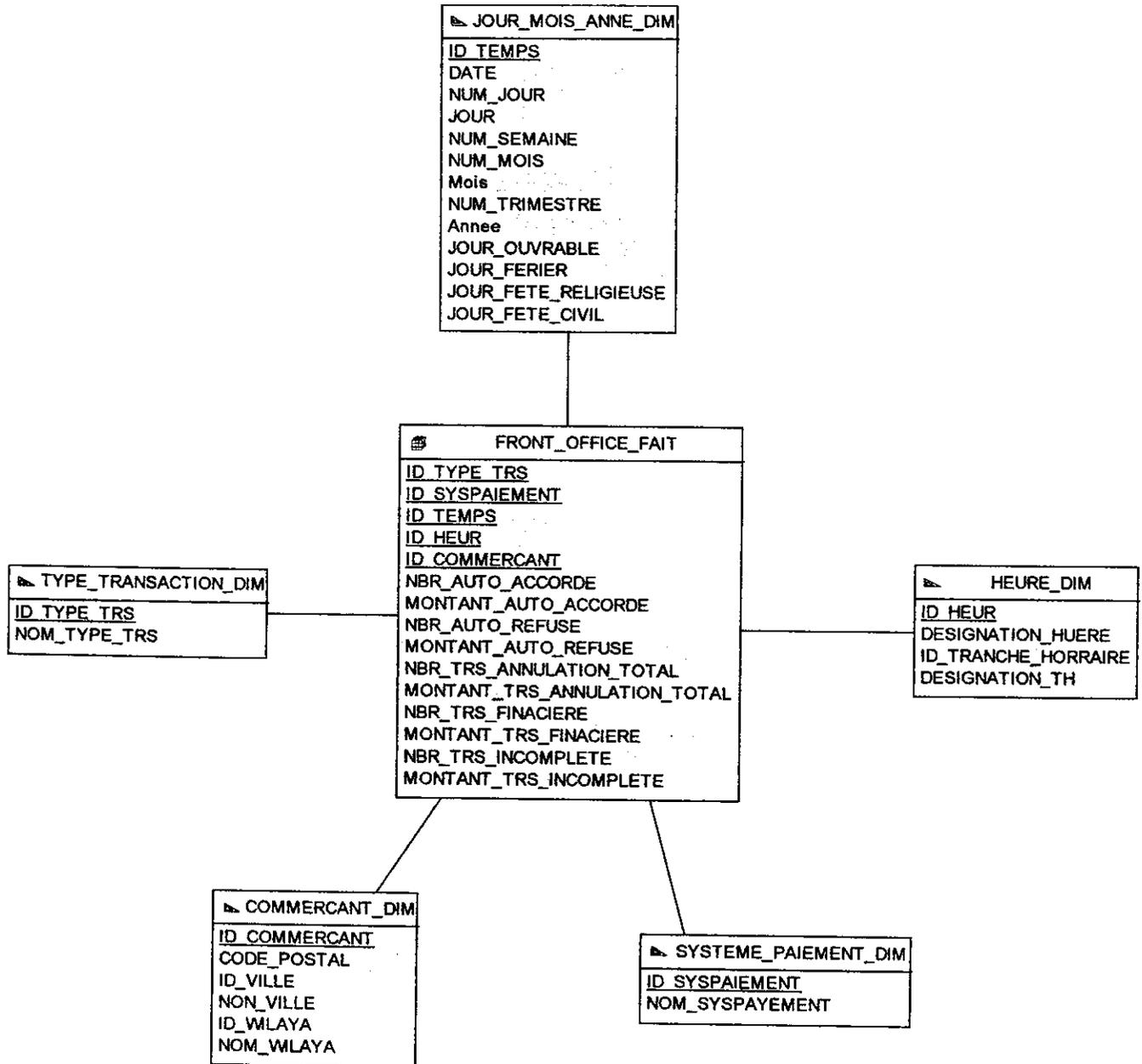


Diagramme V.17 : Diagramme Multidimensionnel : Activité Font Office 1.

II.2.5.3 Activité Front Office 2 :

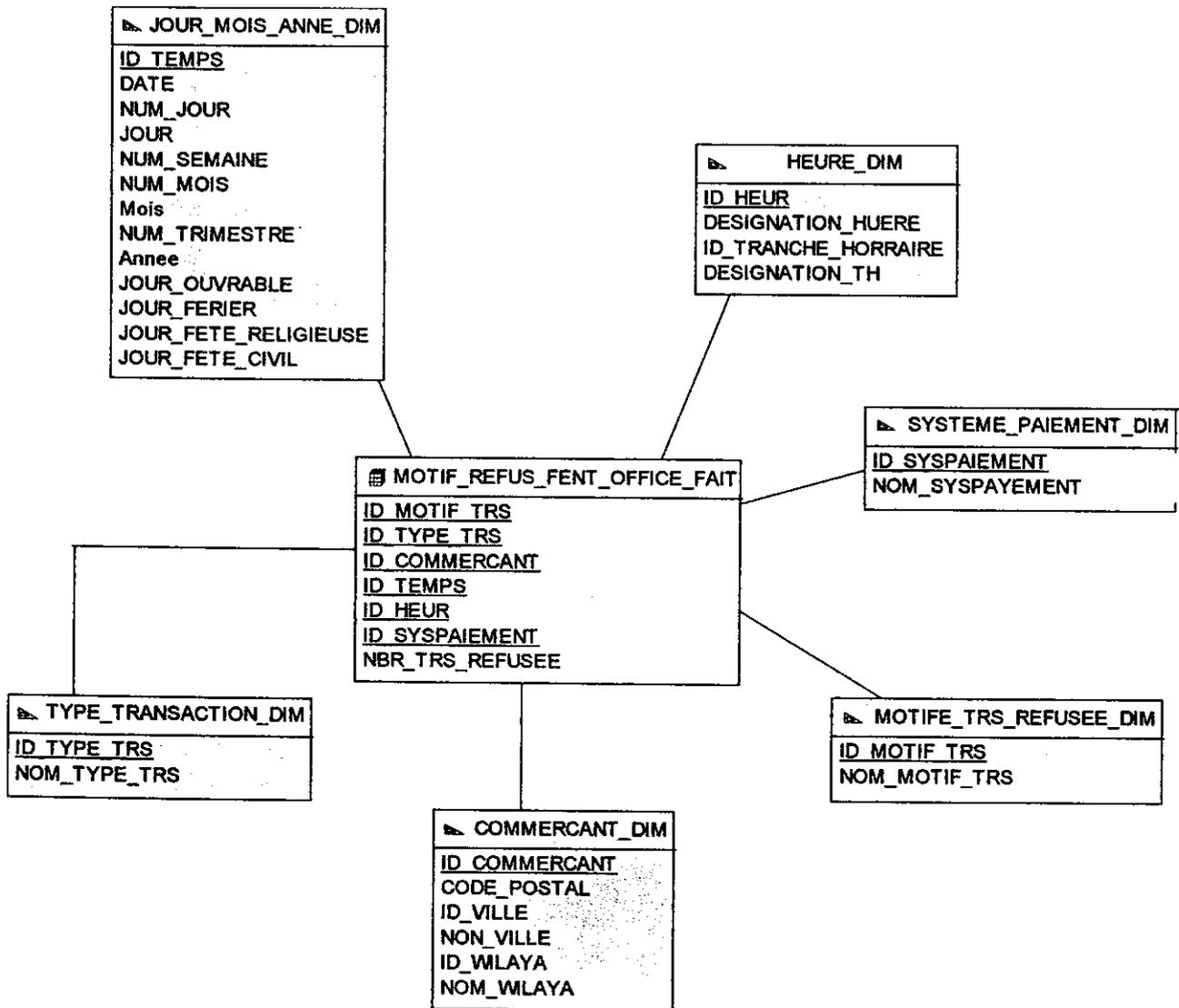


Diagramme V.18 : Diagramme Multidimensionnel : Activité Front Office 2.

II.2.6 Activité Panne :

Le décideur peut visualiser les indicateurs de l'activité ainsi que de choisir les axes d'analyse. Par exemple le nombre de panne arrivée sur tel Commerçant, à telle heure de tel Jour, concernant tel motif de panne.

L'activité panne est illustrée par le diagramme V.19 ci-dessous.

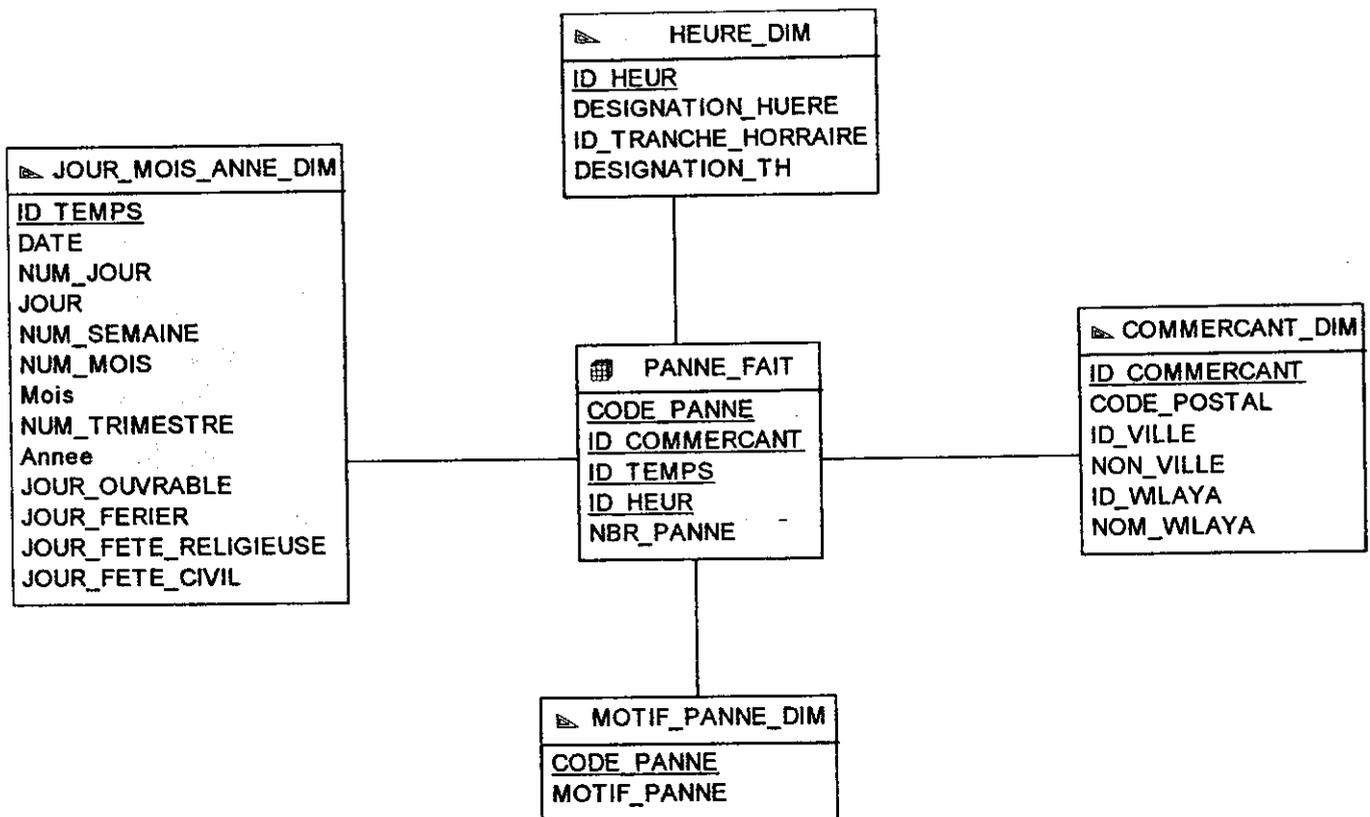


Diagramme V.19 : Diagramme Multidimensionnel : Activité Panne.

## II.2.7 Processus Monétique :

Le processus monétique à a son actif quatre sous processus et une activité :

- Sous Processus Compensation.
- Sous Processus Acquéreur.
- Sous Processus Emetteur.
- Sous Processus Font Office.
- Activité Panne.

Les deux décideurs (Sous directeur de la sécurité et des projet informatique et responsable service monétique) ont les droits d'accès au indicateurs des quatre sous processus ainsi qu'a l'activité panne, avec choix sur les axes d'analyse concernant chaque activité, que d'autres décideurs n'ont pas car ils y sont restreints par leurs droits d'accès.

## II.2.8 Entrepôt de données :

Le diagramme de l'entrepôt de données est la consolidation et l'assemblage des douze derniers diagrammes (V.8 ; V.9 ; V.10 ; V.11 ; V.12 ; V.13 ; V.14 ; V.15 ; V.16 ; V.17 ; V.18 et V.19).

**II.3 Conclusion :** Nous avons opté pour la modélisation dimensionnelle en étoile pour l'optimisation du temps de recherche et pour éviter des jointures complexes.

**PARTIE V : Conception de l'Entrepôt de données**

**Conception des éléments de la zone de préparation des données.**

Dans ce chapitre

**Phase Extraction :**

- **Diagramme des cas d'utilisation de la phase Extraction.**
  - **Scénario 1.**
  - **Scénario 2.**

**Phase Transformation et chargement :**

- **Les diagrammes des cas d'utilisation de la Phase Transformation et Chargement.**
  - **Scénario 1.**
  - **Scénario 2.**

**Diagrammes d'états-transitions :**

- **Diagramme d'état transition général du système.**
- **Diagramme d'état transition détaillé du système.**

**Diagramme d'activité.**

**Conclusion.**

## V. CHAPITRE III : CONCEPTION DES ELEMENTS DE LA ZONE DE PREPARATION DES DONNEES

Le processus de préparation des données, également nommée alimentation des données, constitue la partie immergée de l'iceberg d'un projet de Data Warehouse. Les sources de données et les systèmes qui les hébergent représentent de tels enjeux que la préparation des données prend souvent bien plus de temps que prévu.

Les étapes de ETL se présentent en trois phases :

- Extraction.
- Transformation.
- Chargement.

Pour notre projet nous appliquons ELT (Extraction, chargement, et transformation), on effectue une extraction des données des fichiers plats vers l'ODS, une fois l'opération terminée nous effectuons les différentes transformations qui conviennent aux données de l'ODS. L'ODS ayant subi les transformations nécessaire, le chargement des données s'effectue vers l'entrepôt de données.

Les différentes phases de ELT sont décrites avec les diagrammes des cas d'utilisations et les diagrammes de séquences dans ce qui suit.

### III.1 Phase Extraction :

Chaque cas d'utilisation sera défini textuellement et décrit à l'aide de diagrammes de séquences mettant en évidences les scénarios les plus probables.

#### Remarque :

Il est impossible de dérouler tous les scénarios à cause de leur combinatoire.

#### III.1.1 Diagramme des cas d'utilisation de la phase Extraction :

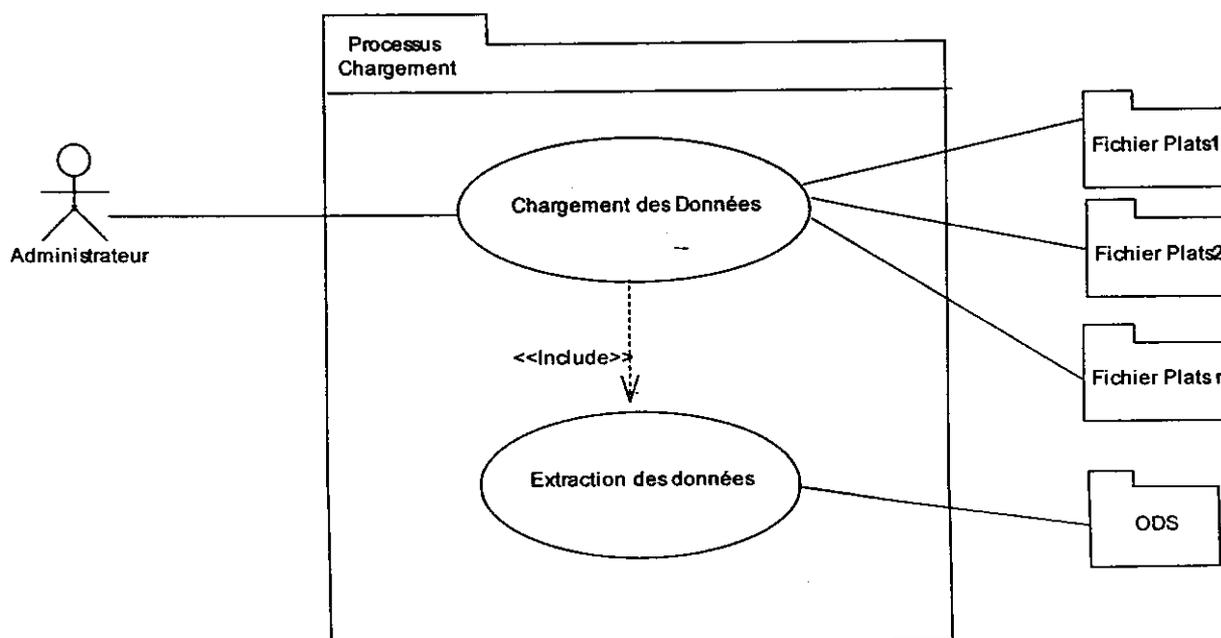


Diagramme V. 20 : Diagramme des cas d'utilisations : Extraction des Fichiers Plats (E)

**Description (Diagramme V. 20):**

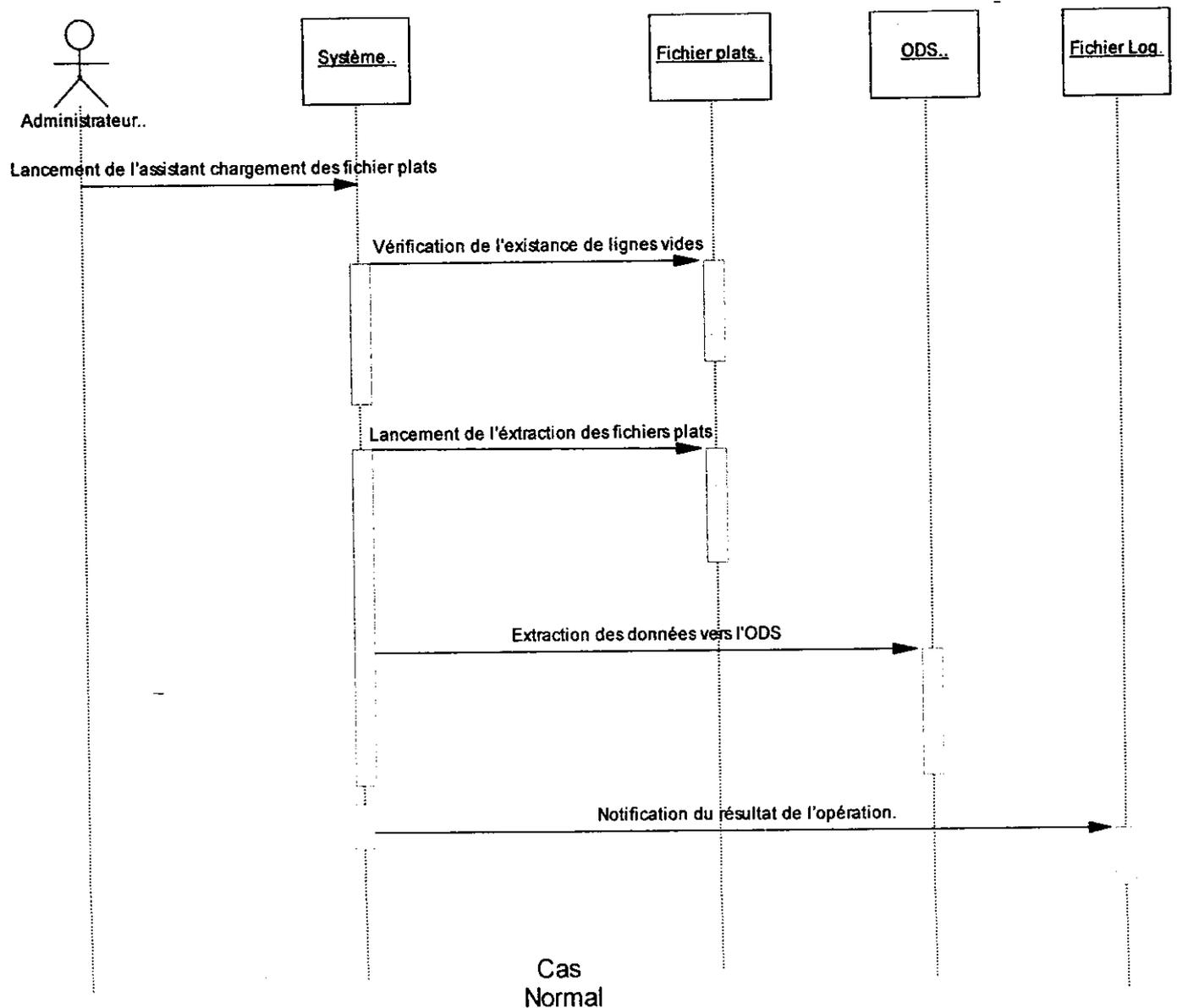
L'administrateur lance le chargement des fichiers plats, une fois l'opération en cours le système commence l'opération de l'extraction des données à partir des fichiers plats vers l'ODS.

**Remarque :** Le système est planifié à lancer le chargement à une heure précise tous les jours, cependant en cas d'erreur l'administrateur peut relancer le chargement.

Scénario 1 : cas normal (Diagramme V.21).

Scénario 2 : cas d'erreur (Diagramme V.22).

**III.1.2. Scénario 1 : Cas normal**



**Diagramme V.21 : Diagramme de séquence : Extraction des Données (E)**

### III.1.3. Scénario 2 : Cas d'erreur

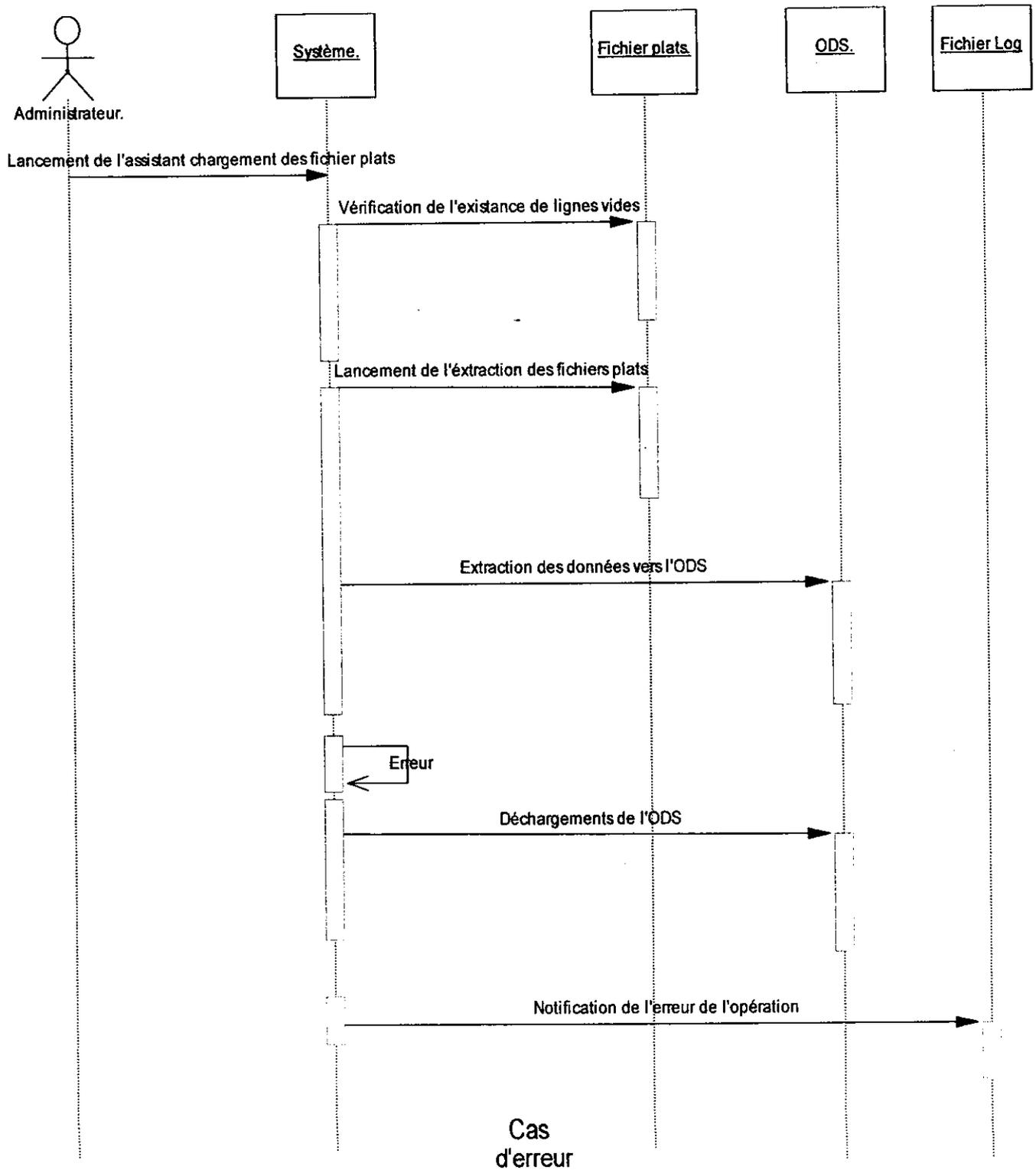
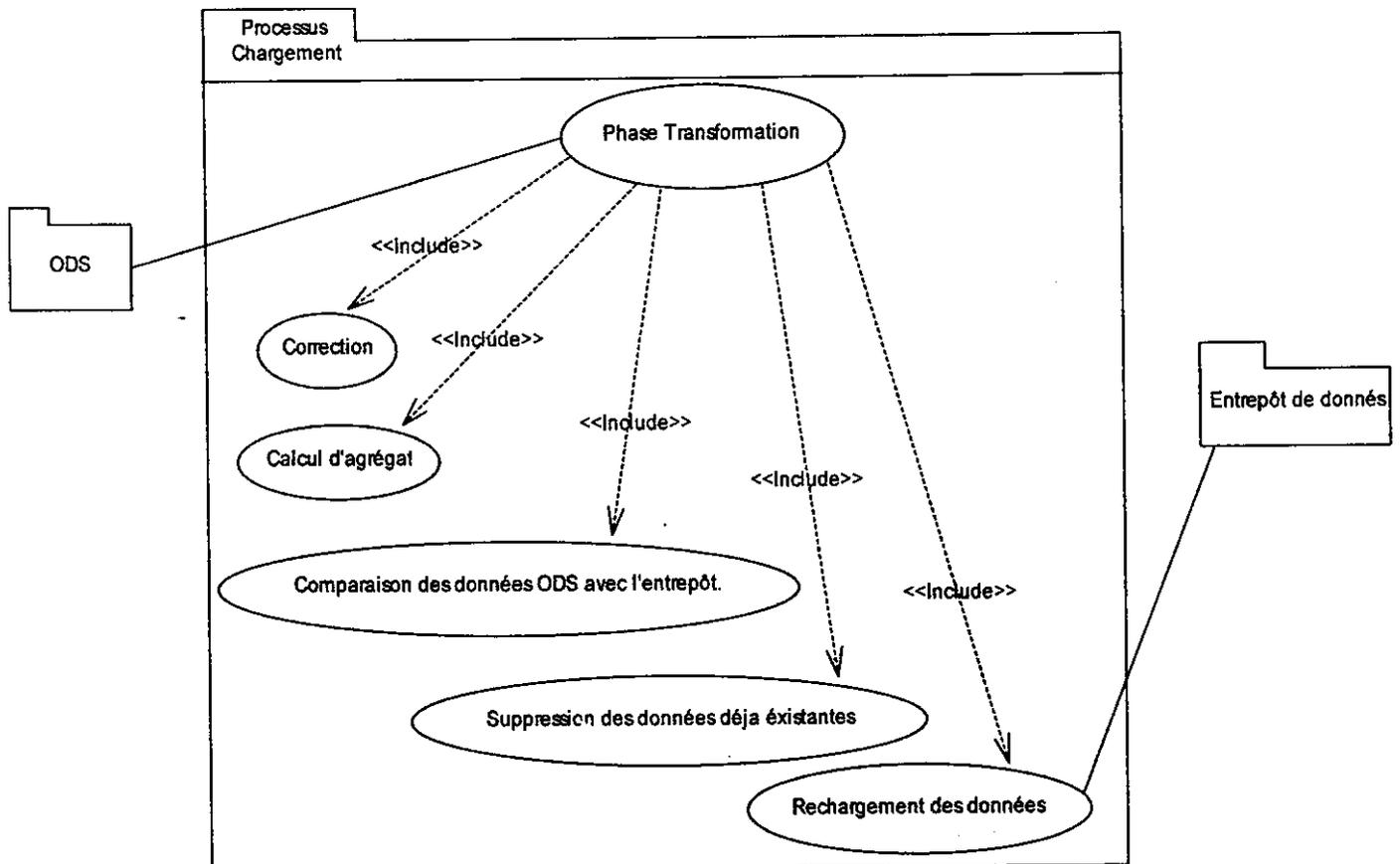


Diagramme V.22 : Diagramme de séquence : Extraction des Données (E)

### III.2 Phase Transformation et chargement :

#### III.2.1 Les diagrammes des cas d'utilisation de la Phase Transformation et Chargement :



**Diagramme V.23 : Diagramme de cas d'utilisation : Transformation et chargement des données (TL)**

#### Description :

Dès que l'opération extraction est finie, le système lance l'opération transformation qui inclut les opérations correction et calcul d'agrégat des données de l'ODS, une fois l'opération transformation est finie, le système vérifie si des données se répètent entre l'ODS et l'entrepôt de données, si oui il les supprime de l'ODS, avant qu'il effectue le chargement vers l'entrepôt de données. Sinon le système lance l'opération chargement des données de l'ODS vers l'entrepôt de données et verrouille le sous système reporting pour les décideurs et l'administrateur.

Scénario 1 : cas normal (Diagramme V.24).

Scénario 2 : cas d'erreur (Diagramme V.25).

III.2.2 Scénario 1 : Cas normal

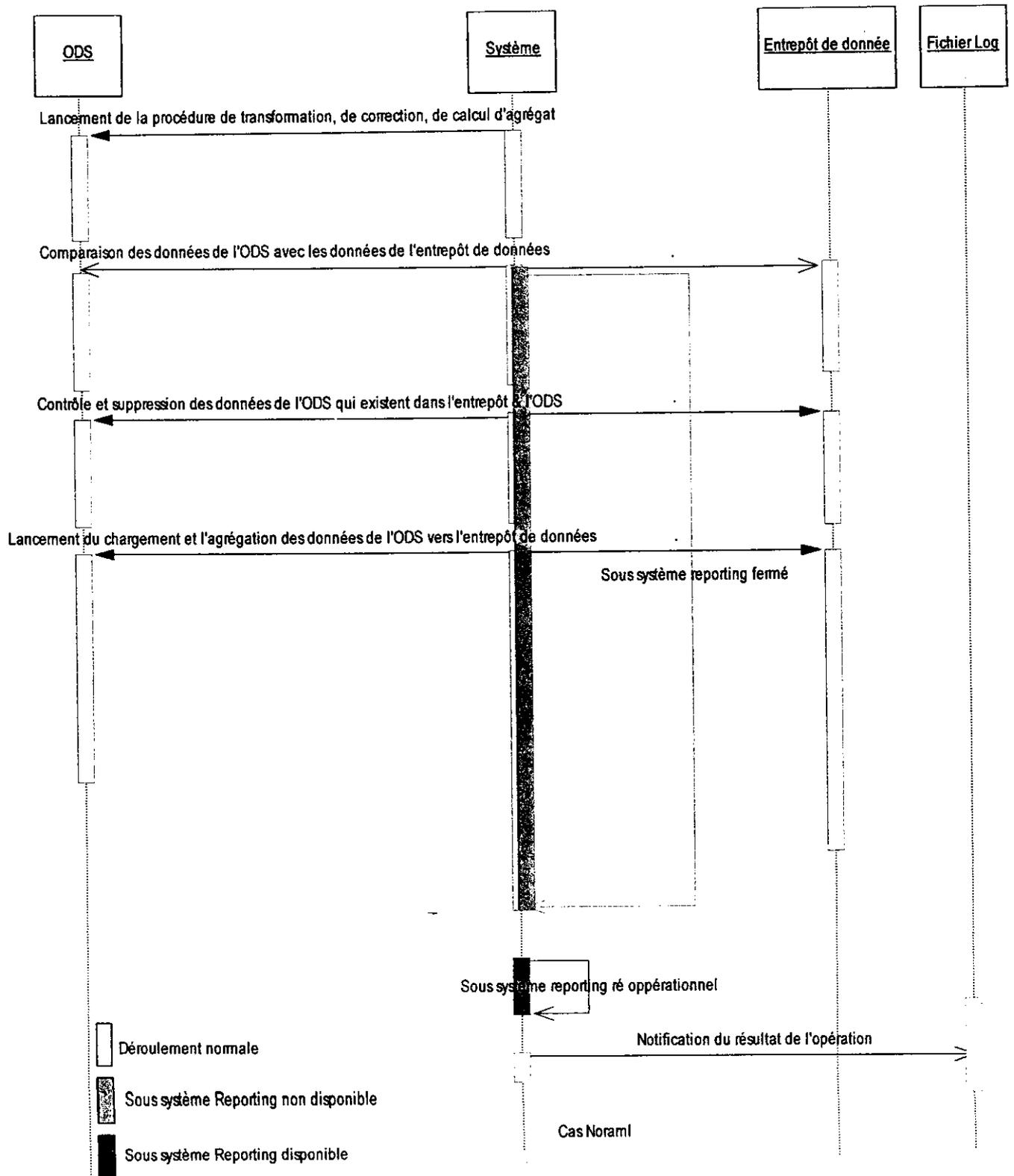
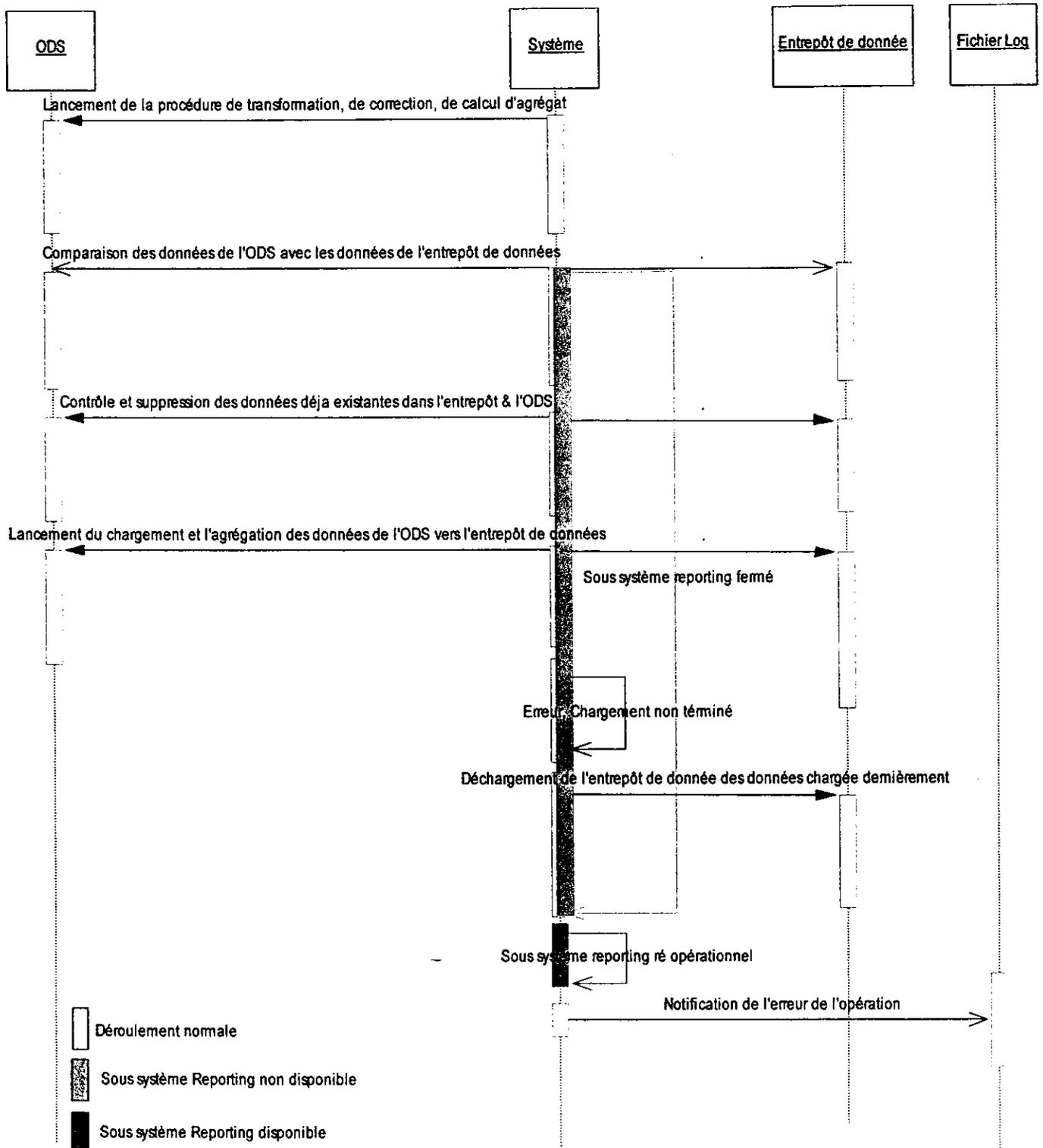


Diagramme V.24 : Diagramme de séquence : Transformation et chargement des données

III.2.3 Scénario 2 : Cas d'erreur



Cas d'erreur

Diagramme V.25 : Diagramme de séquence : Transformation et chargement des données

Après avoir décrit les phase de l'ELT appliquées dans le projet, on passe au diagramme d'états transitions, qui décrit les transactions et les états par lesquels passe l'entrepôt de données. Nous avons choisi la phase ELT pour introduire le diagramme d'état-transition d'UML, car elle est la plus proche et la plus descriptive du changement d'état de l'entrepôt de données.

**III.3 Diagrammes d'états-transitions :** Ils constituent une modélisation du comportement de l'objet et relient des événements à des états en spécifiant la séquence d'états provoquée par une séquence d'évènements. Ils servent donc à représenter des automates à états finis. En UML ces diagrammes sont associés à une classe ! [Bru&Cam,2004].

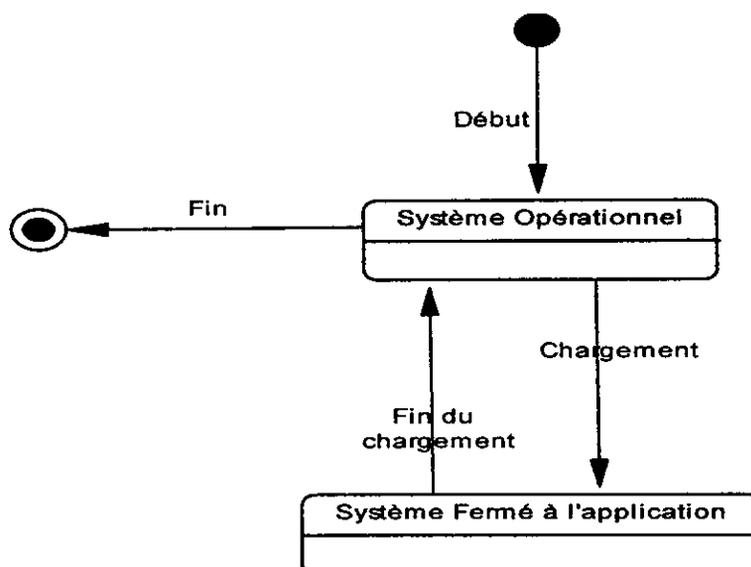
Dans notre cas de figure il s'agit de spécifier les différents états de l'entrepôt de données et la transition entre ces états, ces différents états seront illustrés par les deux diagrammes d'état-transition qui suivent, le premier est plus général par rapport au deuxième qui détaille les différentes étapes du premier diagramme.

### III.3.1 Diagramme d'état transition général du système :

Dans le diagramme qui suit (Diagramme V.26) nous illustrons les deux états majeurs du système (de l'entrepôt de données) :

- Système opérationnel (Toute utilisateur a le droit d'accéder aux indicateurs).
- Système fermé à l'application (aucun utilisateur du système ne peut accéder à l'application).

Traitement : chargement des données.



**Diagramme V.26 :** Diagramme Etat\_Transition : Etat du système

### III.3.2 Diagramme d'état transition détaillé du système :

Dans le diagramme qui suit (Diagramme V.27) nous illustrons les deux états majeurs du système (de l'entrepôt de données) avec des détails sur chaque état :

- Système opérationnel, avec comme transition d'état :
  - Entrepôt de données opérationnel.
  - Fichiers plats en extraction.
  - ODS en chargement.
  - ODS en transformation.
- Système fermé à l'application, avec comme transition d'état :
  - Entrepôt de données en chargement.

Et les traitements sont comme suit :

- Lancement du chargement.
- Extraction des données des fichiers plats.
- Transformation des données ODS.
- Chargement des données de l'ODS vers l'Entrepôt de données.

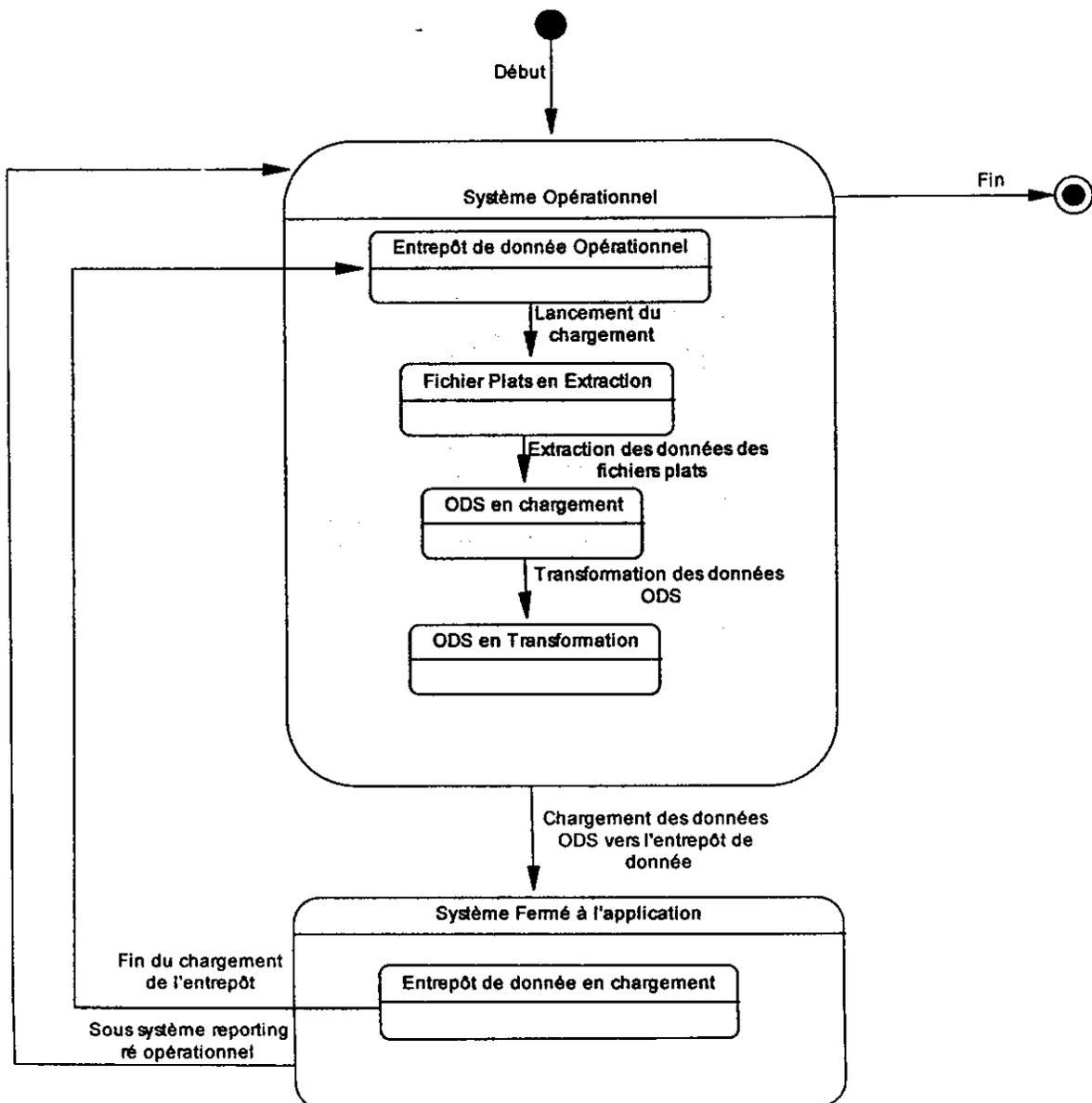


Diagramme V.27 : Diagramme Etat\_Transition : Etat de l'entrepôt de données.

Après avoir décrit les différents états du système à travers les diagrammes d'états-transitions, nous passons à la description de ce dernier par le diagramme d'activité.

Les diagrammes d'activités sont une variante des diagrammes d'états\_transitions organisées par rapport aux actions. Ils permettent de représenter le déclenchement d'événements en fonction des états du système et de modéliser des comportements parallélisables (multi-processus). [Booc&Rumb, 2000].

#### **III.4 Définition des diagrammes d'activités :**

Une activité représente l'exécution d'un mécanisme, un déroulement d'étapes séquentielles. Ces diagrammes permettent de représenter graphiquement le déroulement d'une méthode ou d'un cas d'utilisation. [Bru&Cam,2004].

Un diagramme d'activité UML permet de modéliser un processus interactif, global ou partiel pour un système donné (logiciel, système d'information). Il est recommandable pour exprimer une dimension temporelle sur une partie du modèle, à partir de diagrammes de classes ou de cas d'utilisation, par exemple. Le diagramme d'activité est sémantiquement proche des diagrammes de collaborations ou d'état-transitions, ces derniers offrant une vision microscopique des objets du système. Le diagramme d'activité présente une vision macroscopique et temporelle du système modélisé. [Booc&Rumb, 2000].

Un lot ETL désigne les interactions entre tables sources et cibles, et (où) les procédures stockées qui manipulent le tout.

##### **III.4.1 Diagramme d'activité de notre système :**

Dans le diagramme qui suit (Diagramme V.28) nous allons illustrer les différentes étapes par les quels passe notre système avec indication de l'état de ce dernier à chacune de ses étapes :

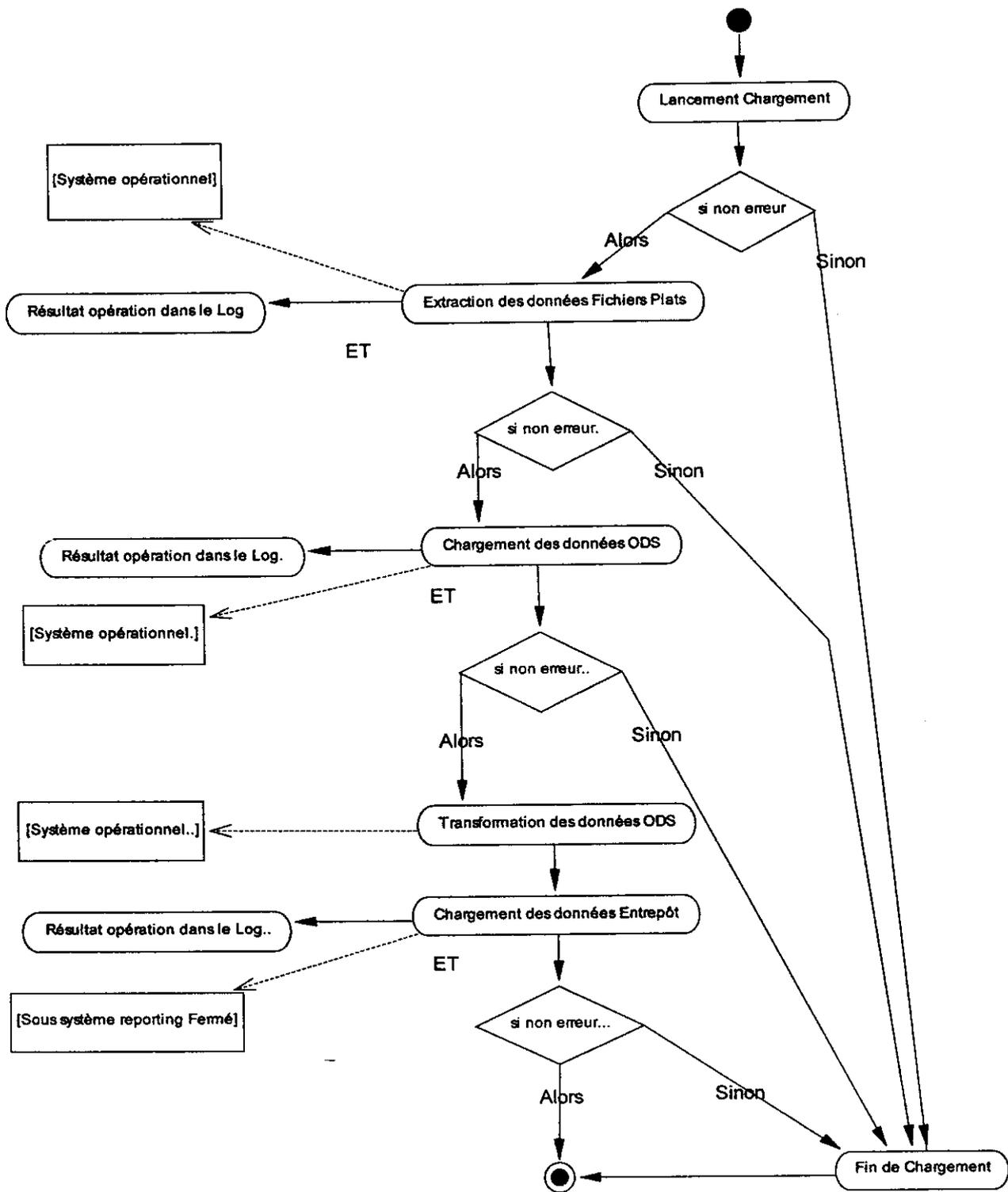


Diagramme V.28 : Diagramme d'activité du système.

III.5 Conclusion : a ce chapitre nous avons utilisé quatre type de diagramme UML (diagramme de cas d'utilisation, diagramme de séquence, diagramme d'état-transition et diagramme d'activité), pour pourvoir décrire formellement tous les aspect de la phase ETL.

**PARTIE V : Conception de l'Entrepôt de données**

**Reporting.**

Dans ce chapitre

**Interface tableau de bord :**

- **Diagramme des cas d'utilisation de l'Interface tableau de bord :**
  - **Connexion au système :**
    - **Scénario 1,2.**
  - **Visualisation d'un indicateur :**
    - **Scénario 1,2.**
  - **Impression du résultat :**
    - **Scénario 1,2.**
  - **Quitter le système :**
    - **Scénario 1,2.**

**Interface administrateur :**

- **Diagramme des cas d'utilisation de l'Interface Administrateur :**
  - **Création d'un nouveau compte utilisateur :**
    - **Scénario 1,2.**
  - **Modification des droits d'accès d'un compte utilisateur :**
    - **Scénario 1,2.**
  - **Suppression d'un compte utilisateur :**
    - **Scénario 1,2.**
  - **Chargement des fichiers plats :**
    - **Scénario 1, 2.**

**Conclusion.**

## V. CHAPITRE IV: REPORTING

Dans cette phase nous allons décrire la partie visible de l'iceberg, il s'agit de l'interface du tableau de bord. Dans cette partie nous allons définir deux sortes d'interface :

- L'interface réservée aux décideurs qui est le tableau de bord.
- L'interface administrateur réservé à l'administrateur de l'outil d'aide à la décision.

### IV.1 Interface tableau de bord :

#### IV.1.1 Diagramme des cas d'utilisation de l'Interface tableau de bord :

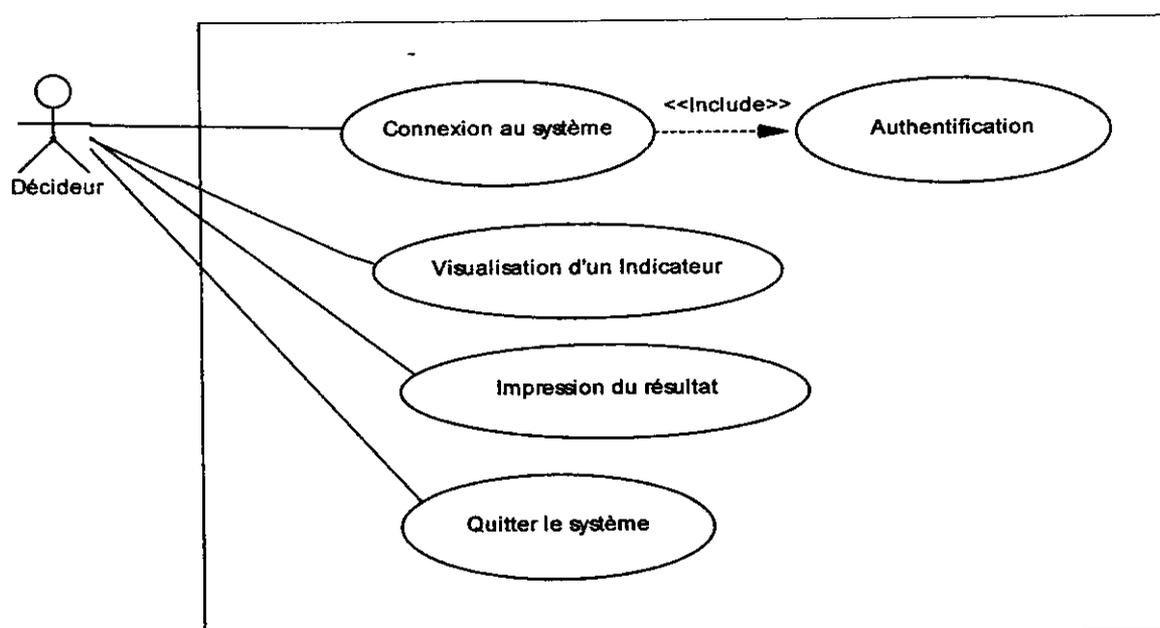


Diagramme V.29 : Diagramme de cas d'utilisation : Interface tableau de bord

#### Description :

Dans ce diagramme des cas d'utilisations, plusieurs cas d'utilisation sont illustrés, ils représentent toutes les fonctionnalités dont le système offre aux décideurs, ils seront décrits dans leurs diagrammes de séquences respectives.

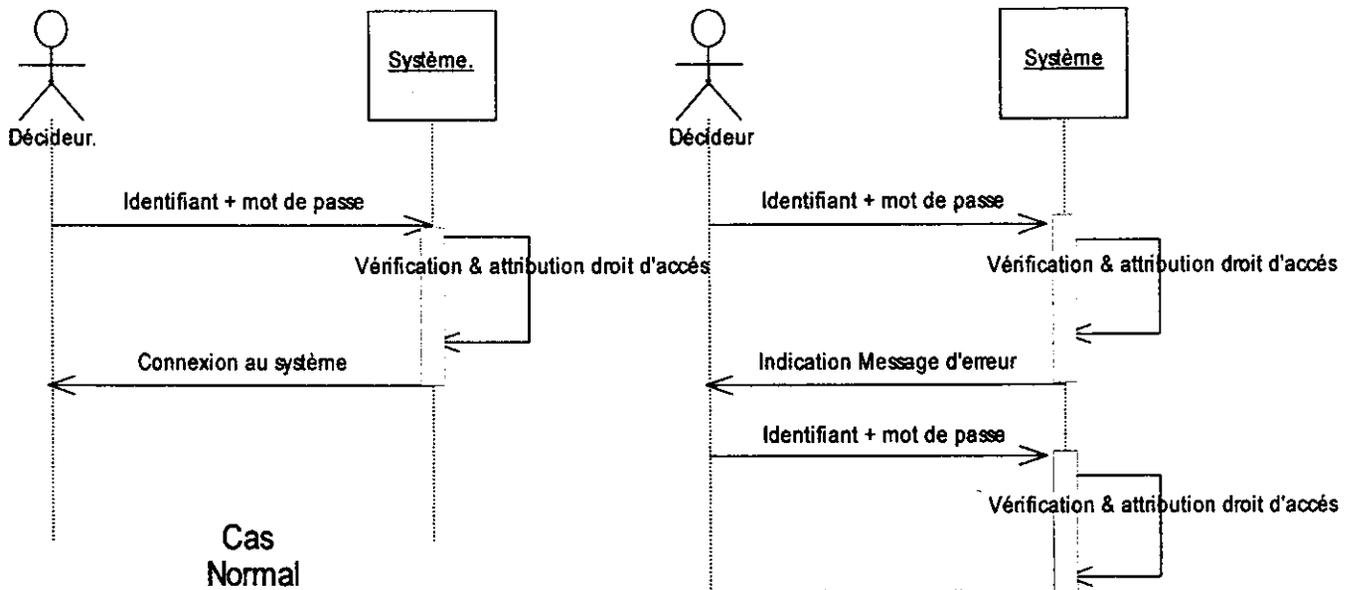
**IV.1.2 Connexion au système :**

Le décideur se connecte au système en s'identifiant par son nom de compte et son mot de passe personnel, s'il n'y a pas d'erreur, la connexion au système s'effectue.

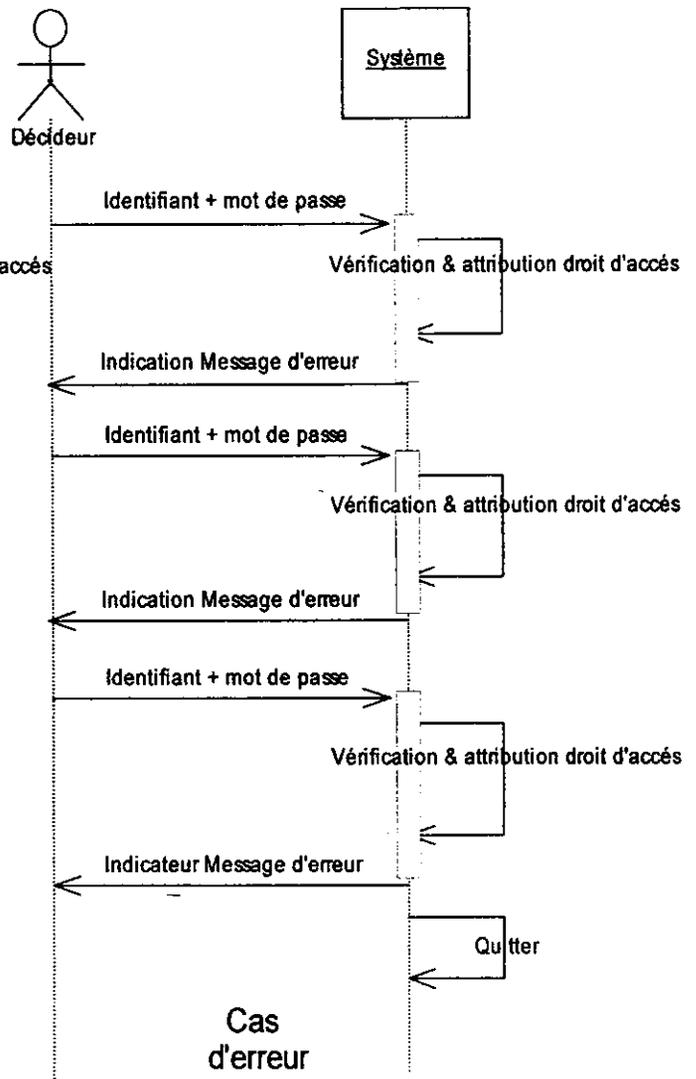
Scénario 1 : cas normal (Diagramme V.30).

Scénario 2 : cas d'erreur, l'utilisateur se trompe dans la saisie de son nom de compte (Diagramme V.31).

**IV.1.2.1 Scénario1, 2 :**



**Diagramme V.30 : Diagramme de séquence authentification**



**Diagramme V.31 : Diagramme de séquence authentification**

### IV.1.3 Visualisation d'un indicateur :

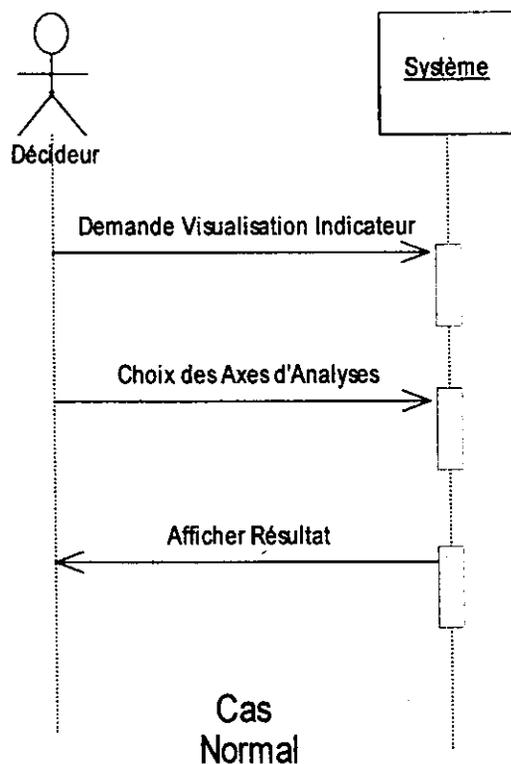
Le décideur une fois identifié, il se fait attribuer des droits d'accès sur les indicateurs qui concernent son compte, il à accès a visualiser les indicateurs le concernant ainsi qu'a choisir les axes d'analyse sur lesquels il voudra travailler.

Deux scénario s'offre a lui :

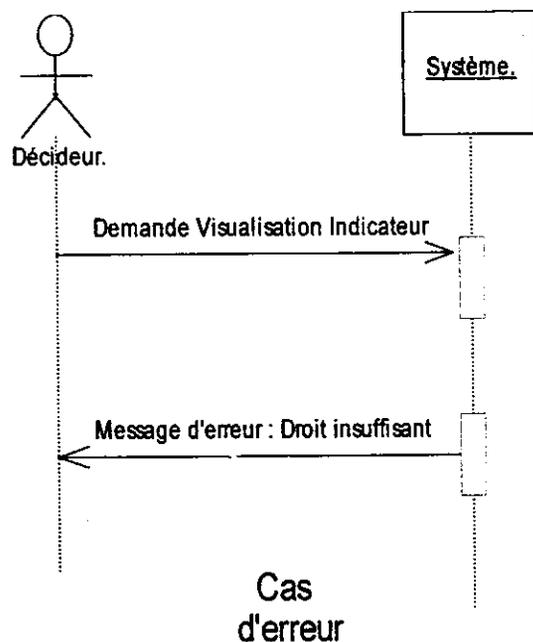
Scénario 1 : cas normal (Diagramme V.32).

Scénario 2 : cas d'erreur, l'utilisateur souhaite accéder à des indicateurs qui se trouvent hors de ces droits d'accès (Diagramme V.33).

#### IV.1.3.1 Scénario1, 2 :



**Diagramme V.32 : Diagramme de Séquence : Visualisation Indicateur**



**Diagramme V.33 : Diagramme de Séquence : Visualisation Indicateur**

### IV.1.4 Impression du résultat :

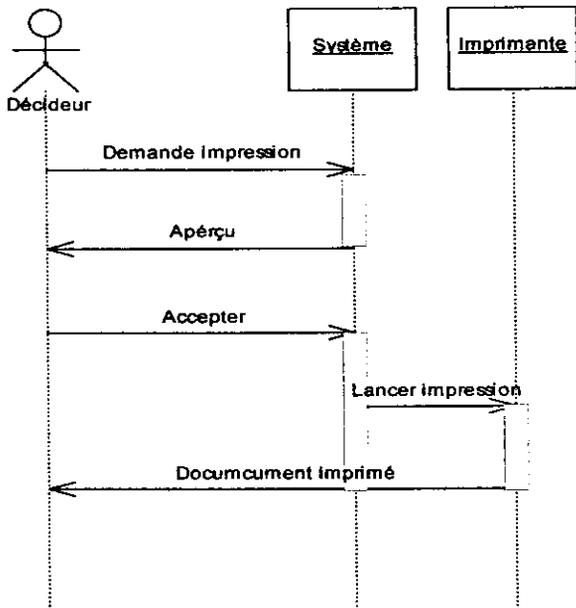
Le décideur à la possibilité d'imprimer le résultat de l'instruction effectuée sur les indicateurs dont il a droit avec spécification des axes d'analyses qui les concernent.

Deux scénario s'offre a lui :

Scénario 1 : cas normal (Diagramme V.34).

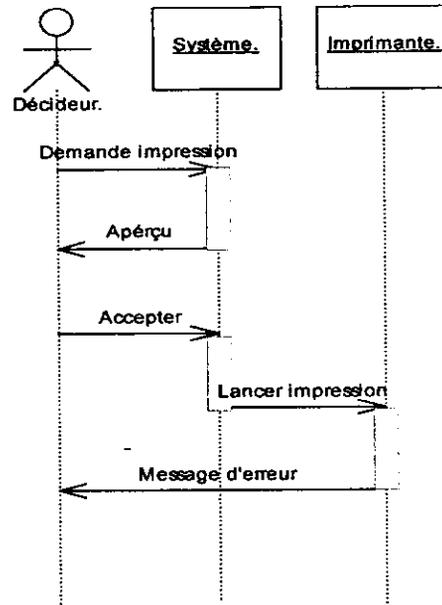
Scénario 2 : cas d'erreur, l'imprimante n'est pas prête (déconnectée, papier ou encre non disponible) (Diagramme V.35).

IV.1.4.1 Scénario1, 2 :



Cas Normal

Diagramme V.34 : Diagramme de Séquence : Impression résultat



Cas d'erreur

Diagramme V.35 : Diagramme de Séquence : Impression résultat

IV.1.5 Quitter le système :

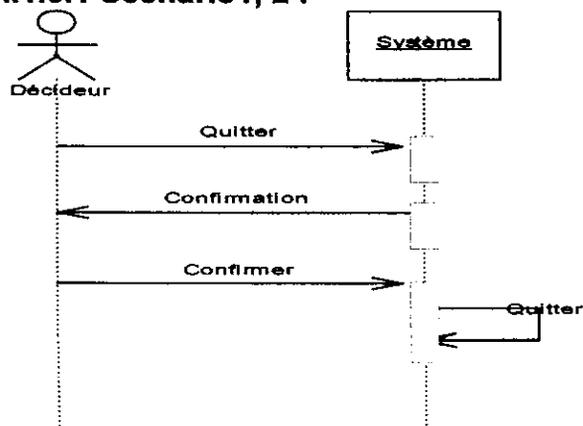
Le décideur souhaite quitter le système, ce dernier demande une confirmation, si elle est validée le système est quitté.

Deux scénario s'offre a lui :

Scénario 1 : cas normal (Diagramme V.36).

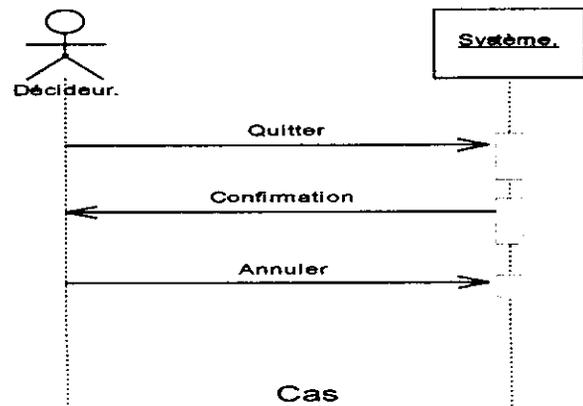
Scénario 2 : cas d'annulation, le décideur ne veut plus quitter le système et annule l'opération (Diagramme V.37).

I.1.5.1 Scénario1, 2 :



Cas Normal

Diagramme V.36 : Diagramme de Séquence : Quitter le système

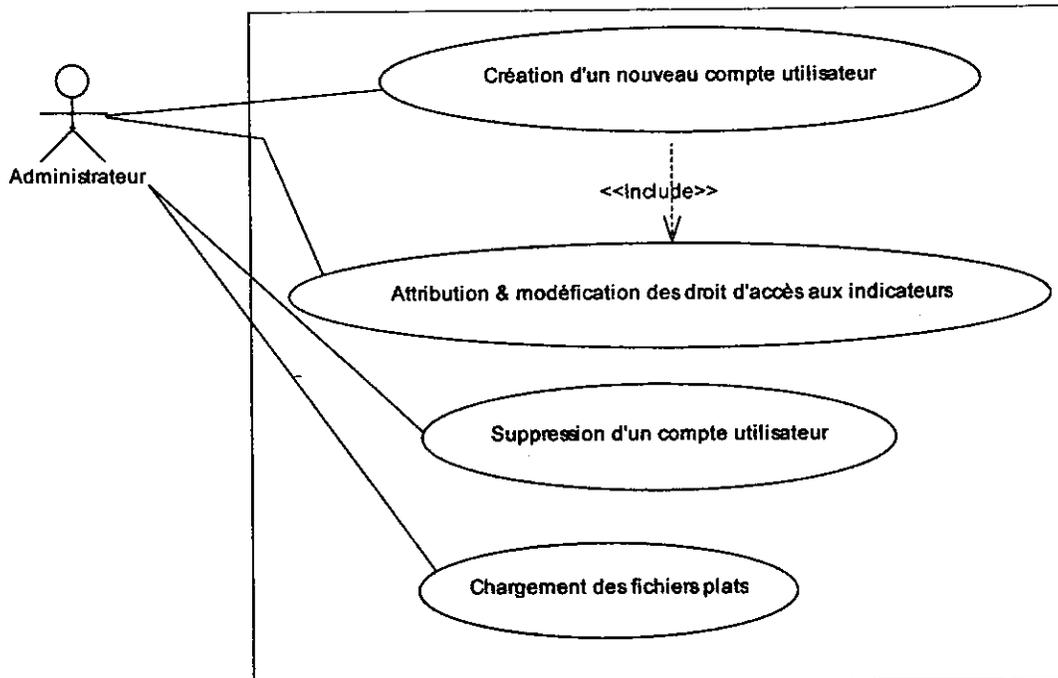


Cas d'annulation de traitement

Diagramme V.37 : Diagramme de Séquence : Quitter le système

## IV.2 Interface administrateur :

### IV.2.1 Diagramme des cas d'utilisation de l'Interface Administrateur :



**Diagramme V.38 :** Diagramme de cas d'utilisation : Interface administrateur

#### Description (Diagramme V.38):

Dans ce diagramme des cas d'utilisations, plusieurs cas d'utilisations sont illustrés, ils représentent toutes les fonctionnalités dont le système offre à L'administrateur afin de piloter et administrer l'outil d'aide à la décision, ils seront décrits dans leurs diagrammes de séquences respectives.

#### IV.2.2 Création d'un nouveau compte utilisateur :

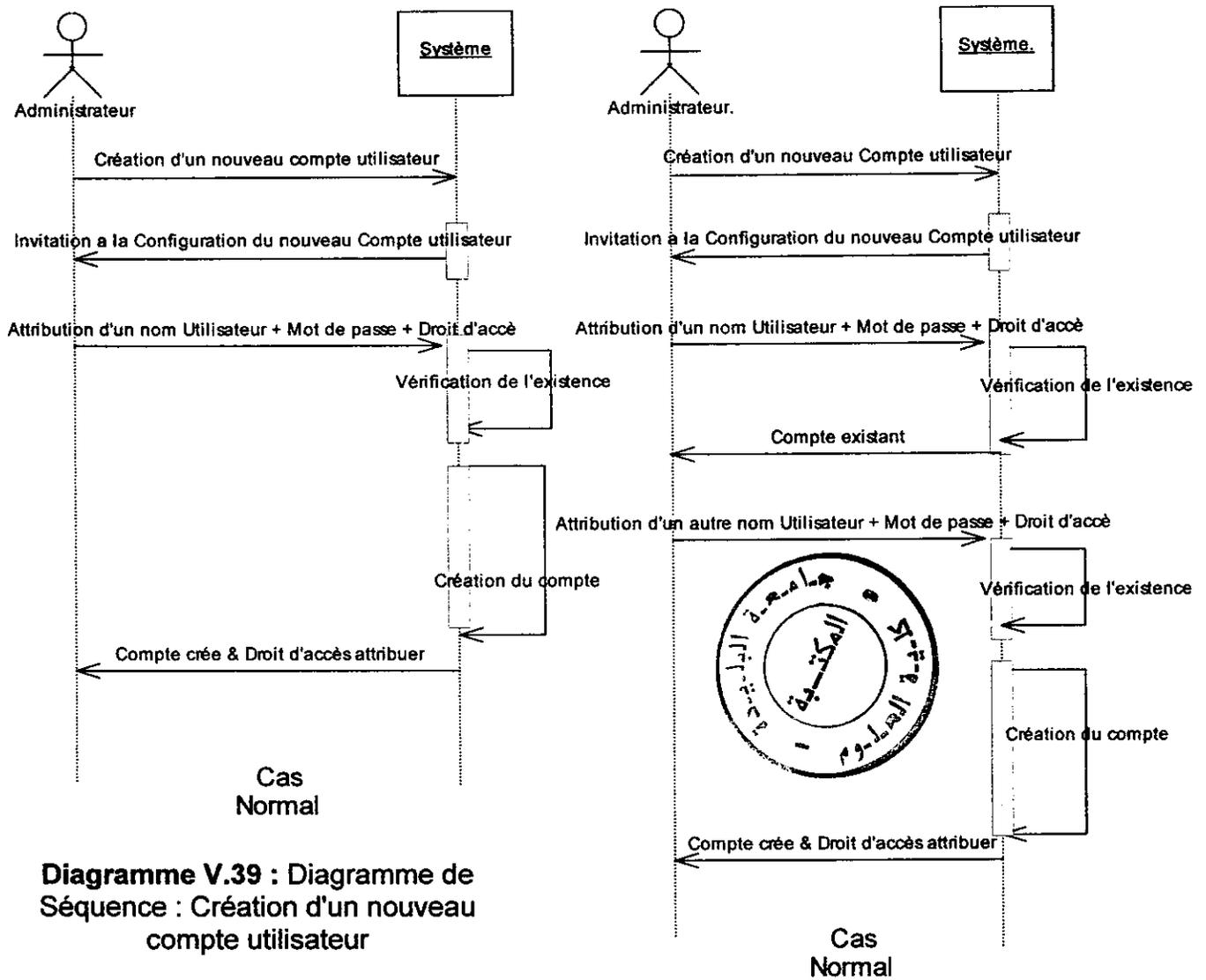
L'administrateur a la responsabilité de créer de nouveaux comptes utilisateurs qui utiliseront le système, ainsi que leur attribuer les droits d'accès aux indicateurs les concernant.

Deux scénarios s'offrent à lui :

Scénario 1 : cas normal (Diagramme V.39).

Scénario 2 : cas d'erreur, dans le cas où il crée un nouveau compte avec un nom de compte déjà attribué au par avant (Diagramme V.40).

IV.2.2.1 Scénario 1, 2 :



IV.2.3 Attribution des droits d'accès :

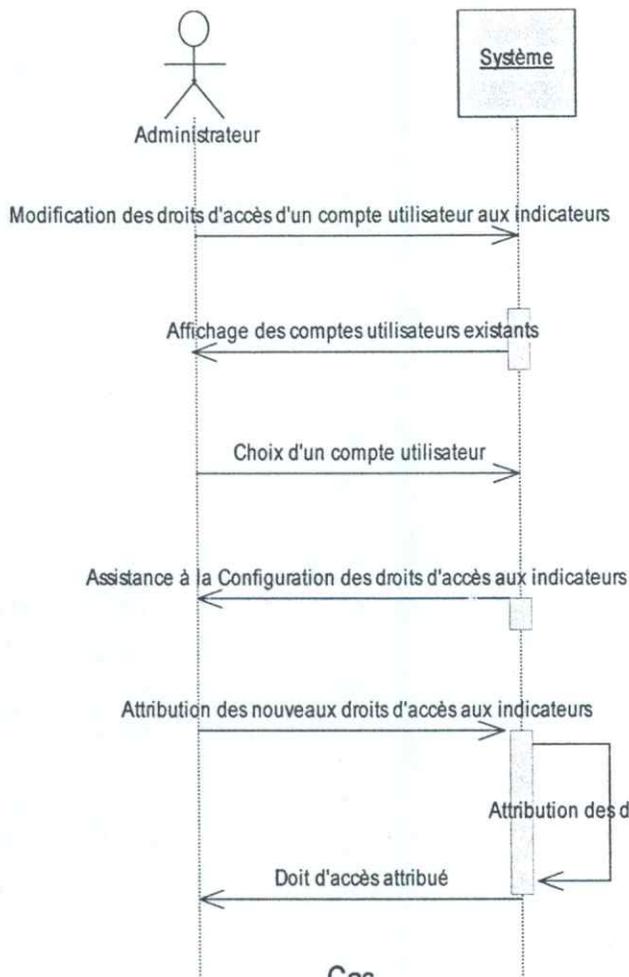
Parmi les responsabilités de l'administrateur, la modification des droits d'accès des décideurs sur les indicateurs, un décideur peut avoir une promotion et changer de grades, ses droits changent aussi.

Deux scénarios s'offrent à lui :

Scénario 1 : cas normal (Diagramme V.41).

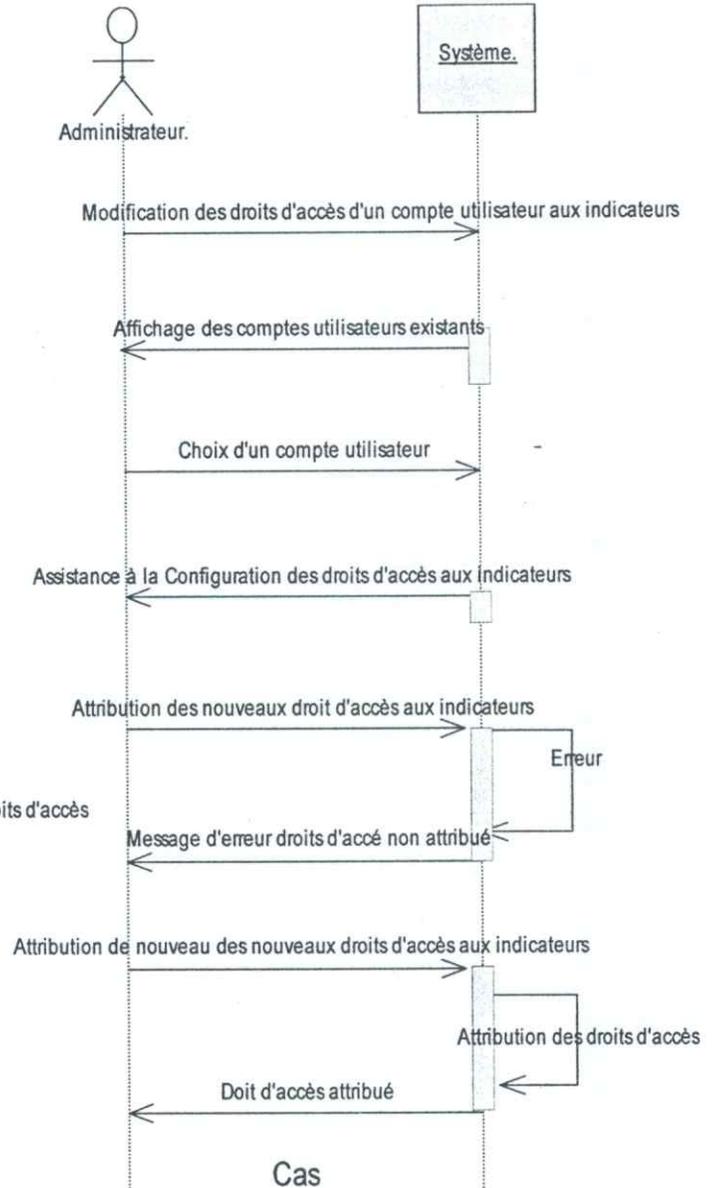
Scénario 2 : cas d'erreur (Diagramme V.42).

IV.2.3.1 Scénario 1, 2 :



Cas Normal

**Diagramme V.41 : Diagramme de Séquence : Modification des droits d'accès aux indicateurs**



Cas d'erreur

**Diagramme V.42 : Diagramme de Séquence : Modification des droits d'accès aux indicateurs**

IV.2.4 Suppression d'un compte utilisateur :

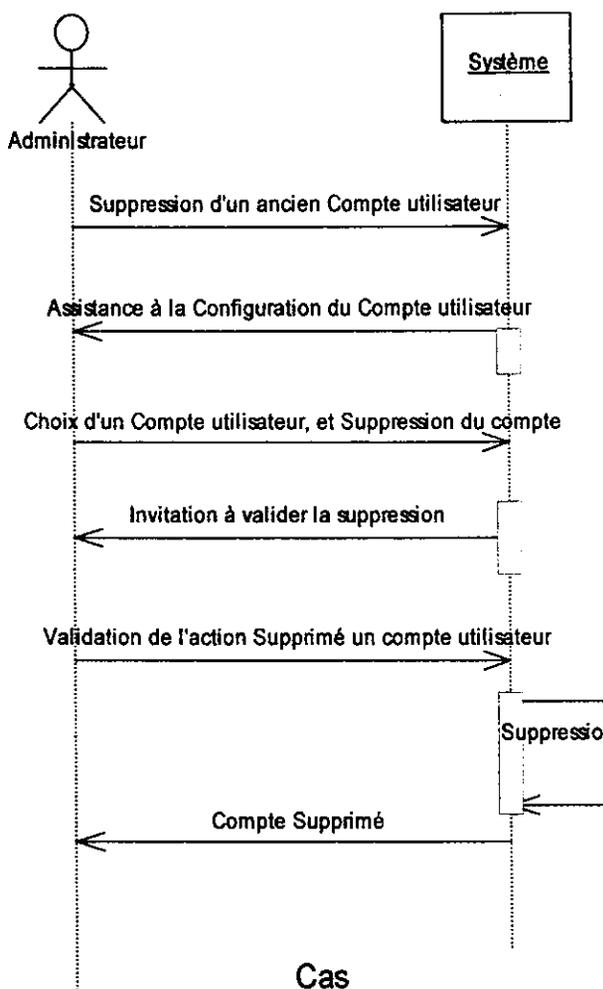
L'administrateur est tenu de supprimer un compte utilisateur en cas de problème avec ce dernier, ou en cas de départ d'un décideur.

Deux scénarios s'offrent à lui :

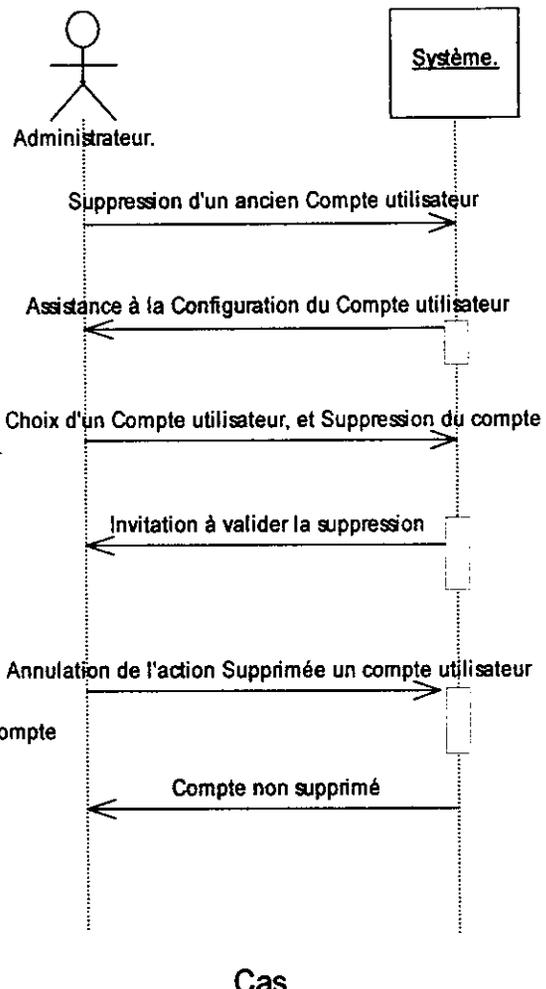
Scénario 1 : cas normal (Diagramme V.43).

Scénario 2 : cas d'annulation de traitement, l'administrateur est revenu sur ses pas et a décidé de ne plus supprimer ce compte la (Diagramme V.44).

**IV.2.4.1 Scénario1, 2 :**



**Diagramme V.43 :** Diagramme de Séquence : Supprimer un compte utilisateur



**Diagramme V.44 :** Diagramme de Séquence : Supprimer un compte utilisateur

**IV.2.5 Chargement des fichiers plats :**

La responsabilité du chargement des fichiers plats est confiée à l'administrateur, c'est lui qui lance l'opération de chargement, a cause de l'instabilité des arrivées des fichiers plats et l'hétérogénéité des sources de ces derniers.

Deux scénarios s'offrent à lui :

Scénario 1 : cas normal (Diagramme V.45).

Scénario 2 : cas d'erreur interruption du chargement (Diagramme V.46).

IV.2.5.1 Scénario1, 2, 3 :

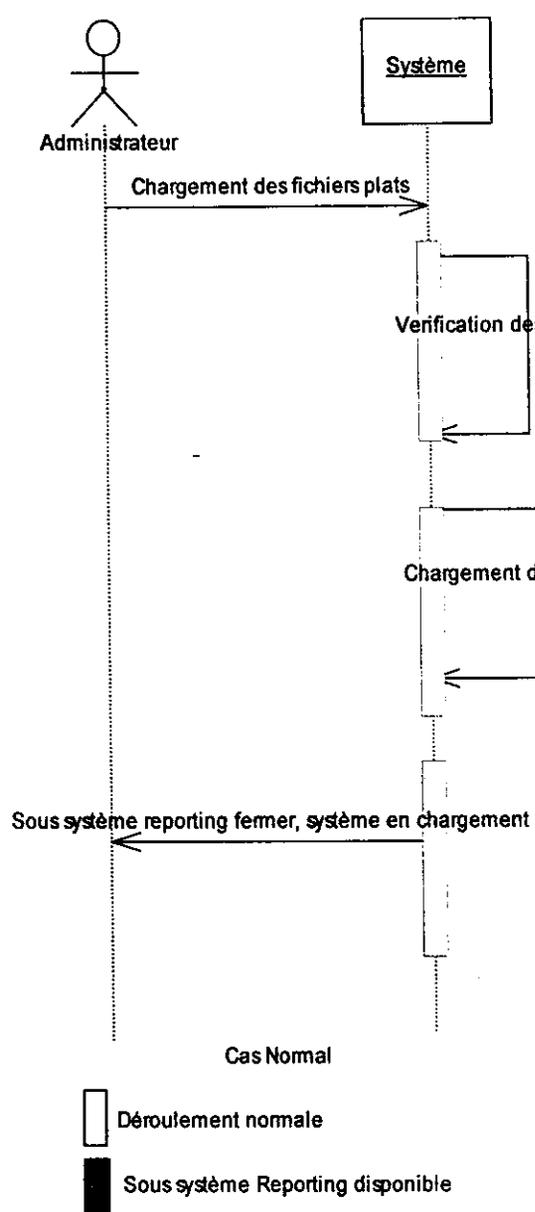


Diagramme V.45 : Diagramme de Séquence : Chargement des fichiers plats

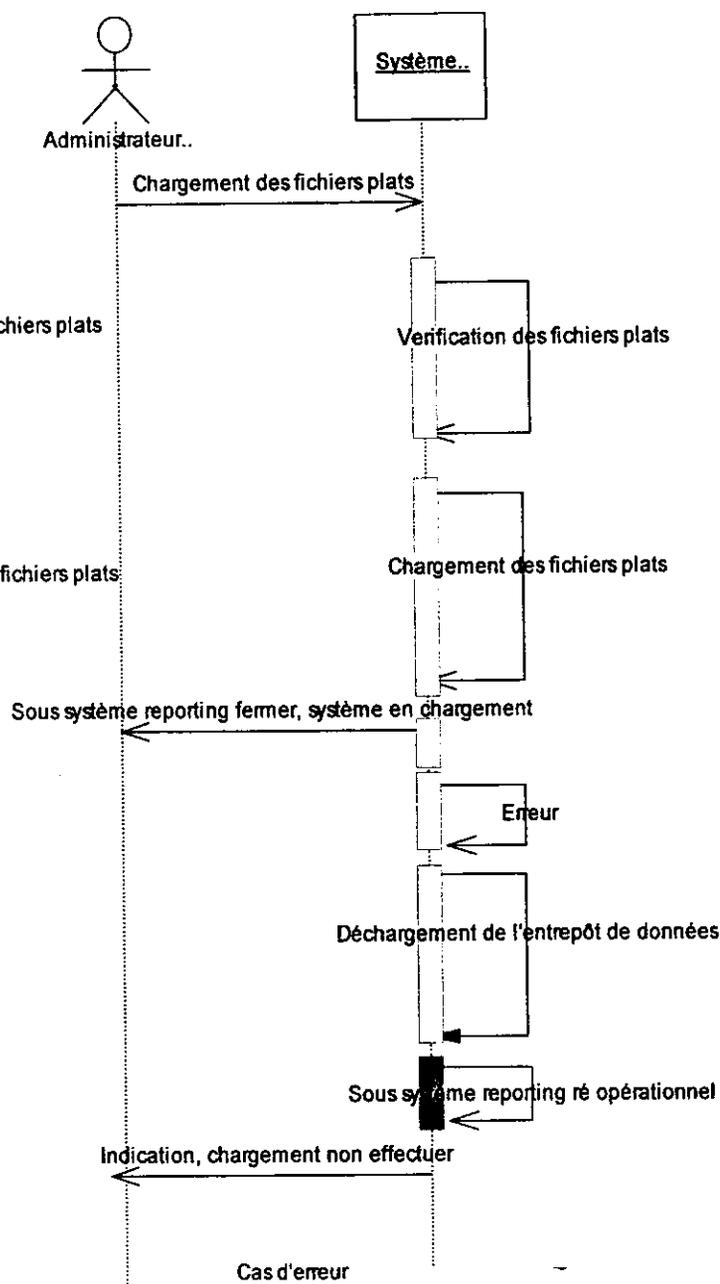


Diagramme V.46 : Diagramme de Séquence : Chargement des fichiers plats

IV.3 Conclusion :

Nous avons décrit à ce chapitre les différents aspects de l'interface de l'outil décisionnel, en utilisant deux des diagrammes UML (Diagramme de cas d'utilisation, Diagramme de séquence).

**PARTIE V : Conception de l'Entrepôt de données**

**Schéma de mise en œuvre et Déploiement**

Dans ce chapitre

**Diagrammes des composants du système :**

- **Diagramme de Composant Authentification.**
- **Diagramme de Composant Tableau de Bord.**
- **Diagramme de Composant Administrateur.**

**Diagramme de déploiement.**

**Conclusion.**

## V. CHAPITRE V : SCHEMA DE MISE EN ŒUVRE ET DEPLOIEMENT

A ce chapitre nous allons entamer la mise en œuvre physique de notre système en utilisant les deux derniers diagrammes d'UML, le diagramme de composant, et le diagramme de déploiement.

Les diagrammes de composants et les diagrammes de déploiement sont les deux derniers types de vues statiques en UML. Les premiers (diagramme de composant) décrivent le système modélisé sous forme de composants réutilisables et mettent en évidence leurs relations de dépendance. Les seconds (diagramme de déploiement) se rapprochent encore plus de la réalité physique, puisqu'ils identifient les éléments matériels (PC, Modem, Station de travail, Serveur, etc.), leur disposition physique (connexions) et la disposition des exécutables (représentés par des composants) sur ces éléments matériels. [Booc&Rumb, 2000].

### V.1 Diagramme de composant :

#### V.1.1 Définition : [Bru&Cam,2004].

Le diagramme de composants permet de décrire l'architecture physique et statique d'une application en termes de modules de différentes nature (fichiers sources, librairie, exécutables...). Ils montrent la mise en œuvre physique du système avec son environnement de développement.

Un composant est un élément physique qui représente une partie implémentée du système. Il peut être du code, un script, un fichier de commande, un fichier de données, une table, etc....

#### V.1.2 Diagrammes des composants du système :

Comme nous l'avons vu au chapitre précédent (Phase Reporting), nous avons présenté deux types d'interface : Interface tableau de bord, Interface Administrateur. Et comme le diagramme des composants représente l'architecture physique et statique de l'application, nous nous retrouvons à représenter trois diagrammes de composants :

- Diagramme des composants Authentification, illustré par le diagramme V.47.
- Diagramme des composants Tableau de Bord, illustré par le diagramme V.48.
- Diagramme des composants Administration, illustré par le diagramme V.49.

V.1.2.1 Diagramme de composant authentification :

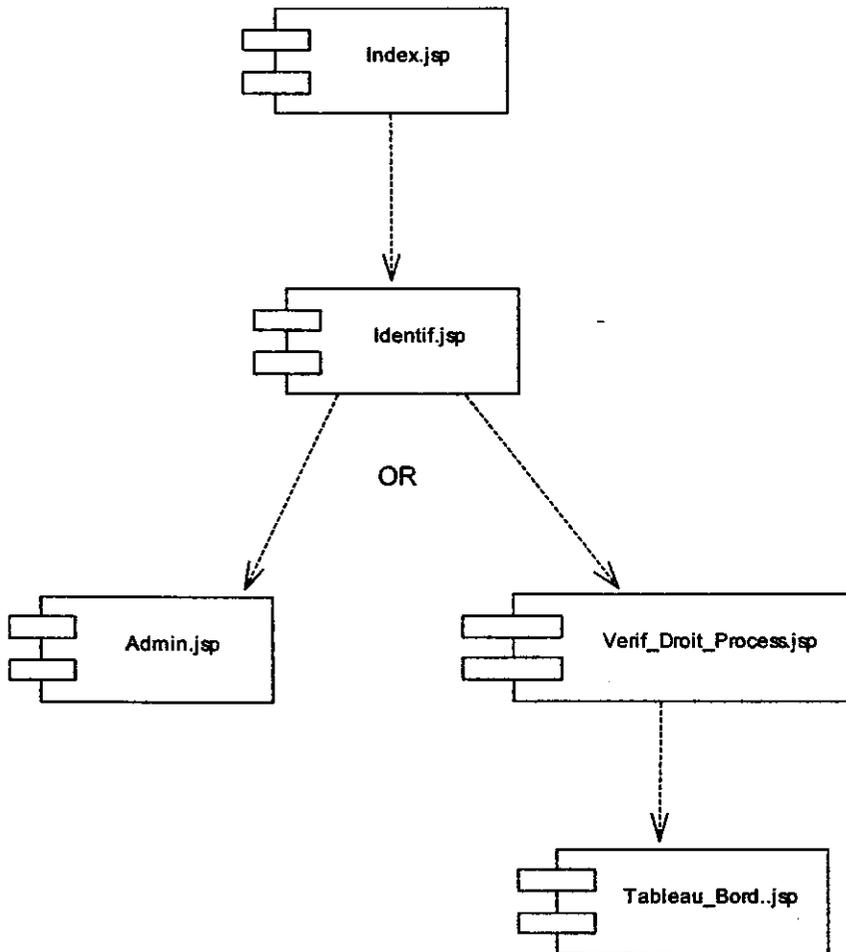


Diagramme V.47 : Diagramme de composant Authentification.

V.1.2.2 Diagramme de composant Tableau de Bord :

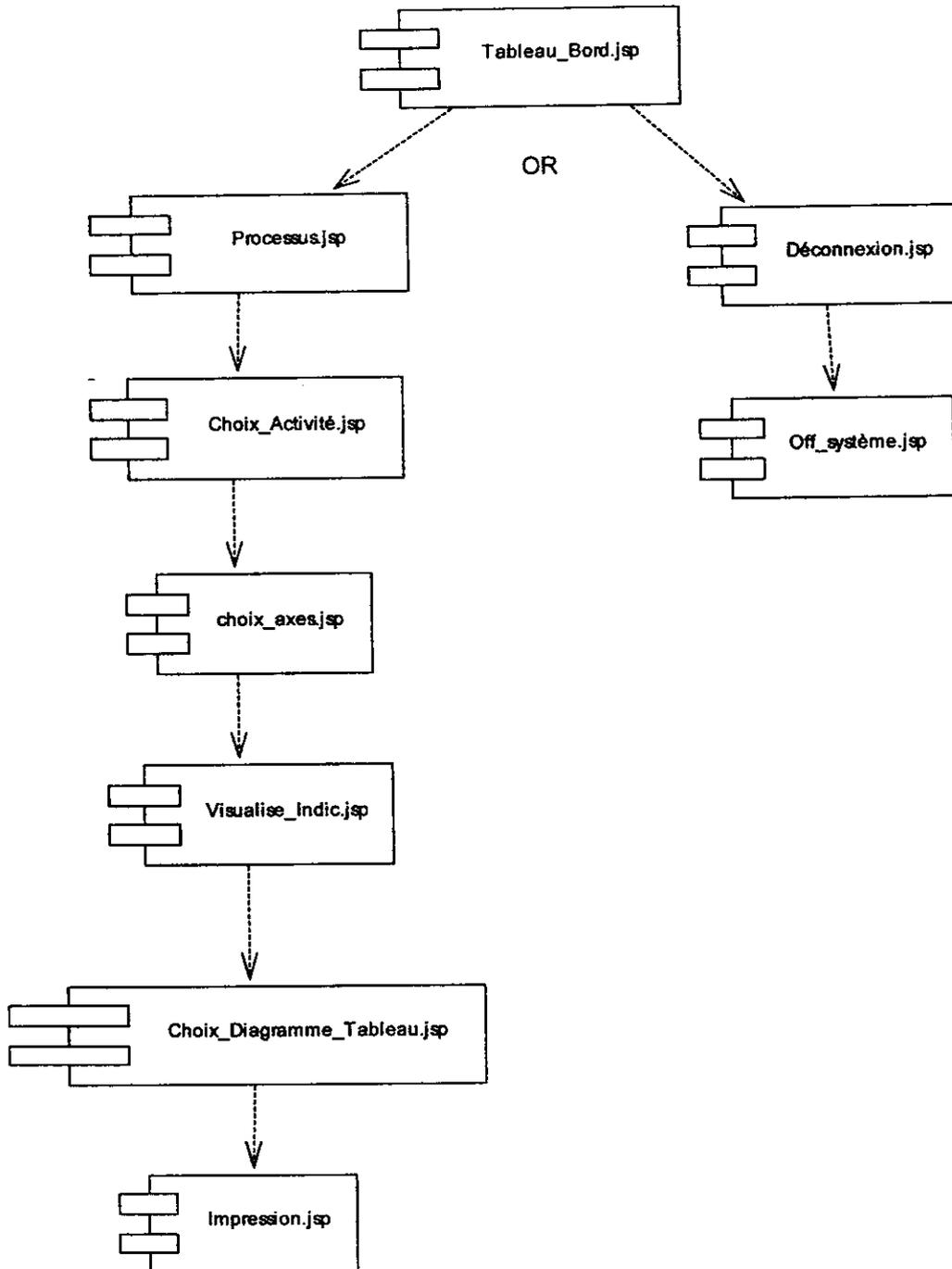


Diagramme V.48 : Diagramme de composant Tableau\_Bord.

V.1.2.3 Diagramme de composant Administration :

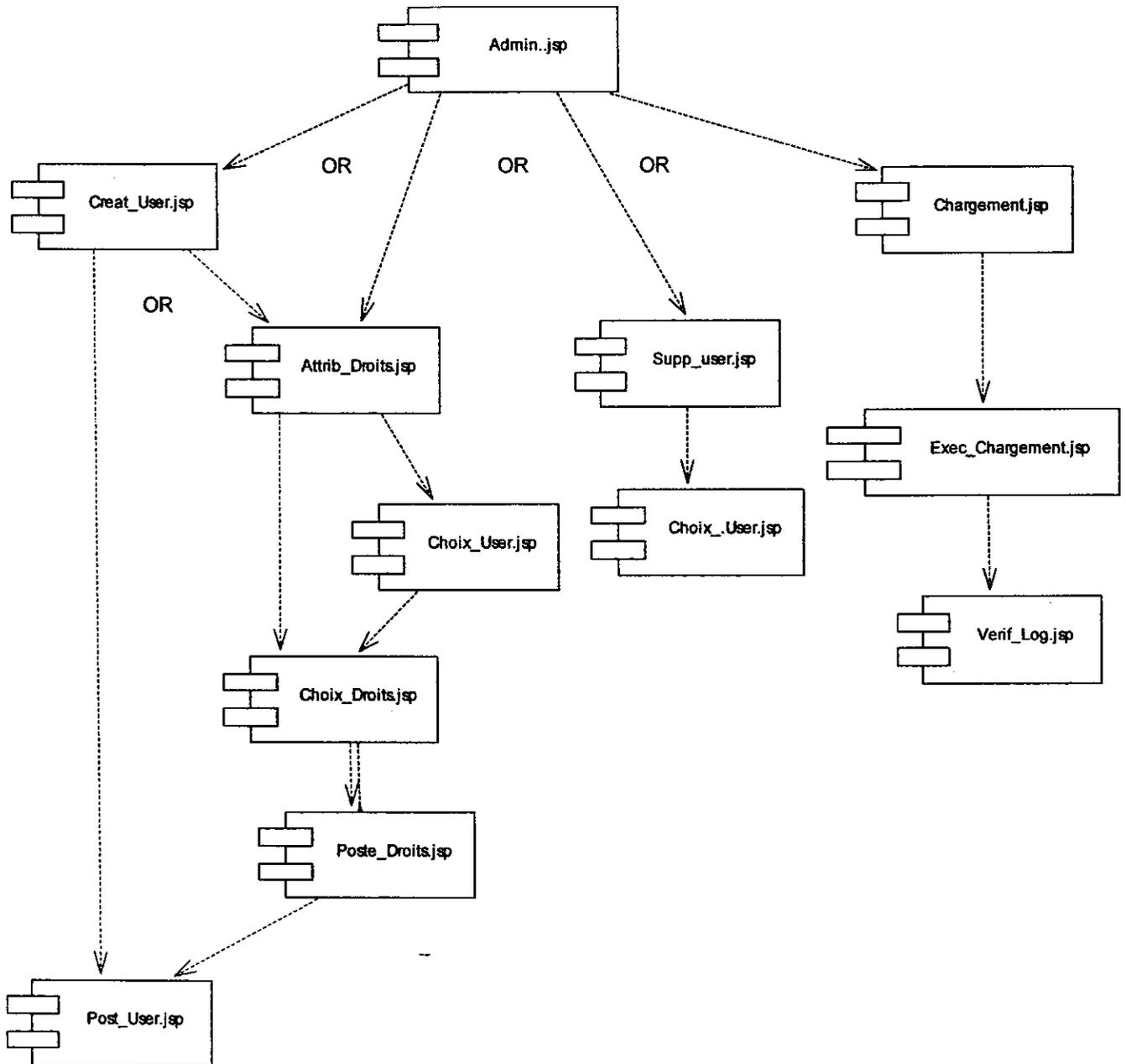


Diagramme V.49 : Diagramme de composant Administration.

## V.2 Diagramme de déploiement :

### V.2.1 Définition : [Bru&Cam,2004].

Les diagrammes de déploiements montrent la disposition physique des matériels qui composent le système et la répartition des composants sur ces matériels. Ils mettent ainsi en évidence l'arrangement physique des ressources d'exécution (les machines et leurs interconnexions)

Il est représenté par un graphe composé de nœuds interconnectés par des liens de communication.

### V.2.2 Diagramme de déploiement du système :

Le diagramme de déploiement de notre système est illustré par le diagramme n° 50 comme suit :

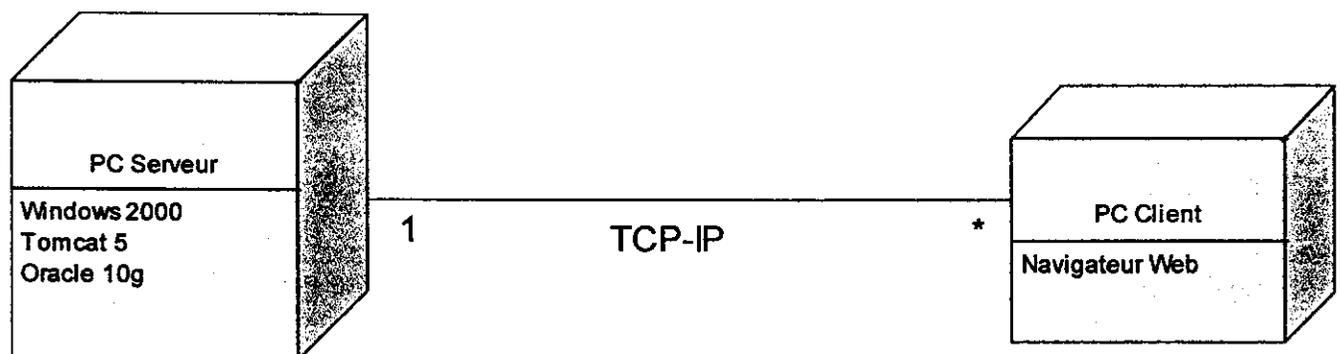


Diagramme V.50 : Diagramme de déploiement.

## V.3 Conclusion :

Nous avons décrit à ce chapitre la mise en oeuvre de notre système en utilisant les deux dernier diagrammes d'UML (Diagramme des composant et le Diagramme de déploiement).

**PARTIE VI**

**Réalisation**

Dans cette partie

**Chapitre I : Environnement et Outil de développement.**

**Chapitre II : Description du Système.**

**PARTIE VI : Réalisation**

**Environnement et Outil de développement**

Dans ce chapitre

**Architecture de déploiement.**

**Caractéristiques de l'architecture trois tiers.**

**Environnement Logicielle.**

**Introduction aux outils utilisés dans le développement.**

**Implémentation**

**Description de l'utilisation des outils :**

- Talend.

- AWM.

**Conclusion.**

Dans cette dernière partie, la phase finale, nous entamons l'architecture de notre système ainsi que son environnement logiciel.

Nous poursuivons par introduire tous les outils de développement utilisés. S'en suit une description de l'utilisation de deux des principaux outils décisionnel (Talend, AWM).

Nous terminons cette partie par une description de notre système.

## VI. CHAPITRE I : ENVIRONNEMENT ET OUTILS DE DEVELOPPEMENT

### I.1 Architecture de déploiement :

Notre application est destinée à être utilisée sous la forme d'«Application Client/Serveur ». Pour notre cas, nous avons utilisé l'architecture client serveur à trois tiers (3/3).

### I.2 Caractéristiques de l'architecture trois tiers (3/3) :

L'architecture trois tiers est composée d'un serveur de base de données exécutant le SGBD, d'un serveur d'application exécutant le corps des applications et enfin de clients responsable des dialogues et de la présentation des données.

Le client dans cette architecture peut être un client Web, s'exécutant dans un browser,

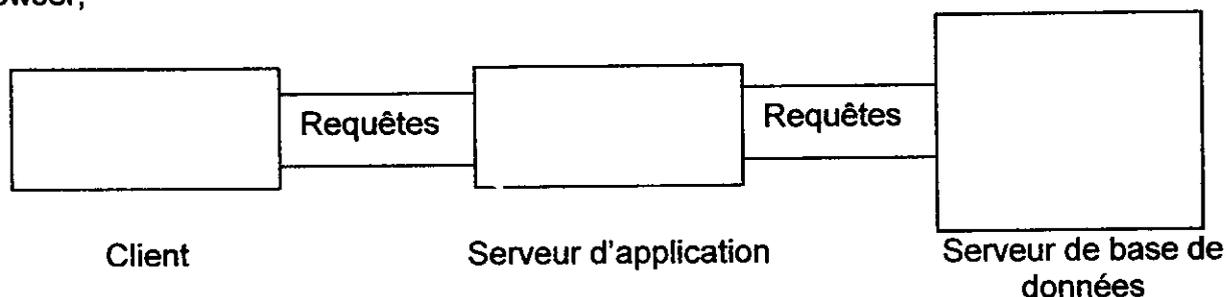


Figure VI.28 : Architecture client/serveur à trois tiers (3/3).

### I.3 Environnement Logicielle :

La réalisation de notre système a mis en jeu une multitude de logiciel, nous présentons à présent la configuration logicielle utilisée :

#### I.3.1 Configuration Logicielle :

##### I.3.1.1 Coté Serveur :

Système d'exploitation : Windows 2000 Serveur.

Gestionnaire de base donnée : Oracle 10g version 10G release2 10.2.0.1.0

Tomcat 5.

##### I.3.1.2 Coté Client :

Système d'exploitation : Windows 2000 PRO et XP.

Tomcat 5.

La configuration Logiciel de notre système est illustré par la figure VI.2 comme suit :

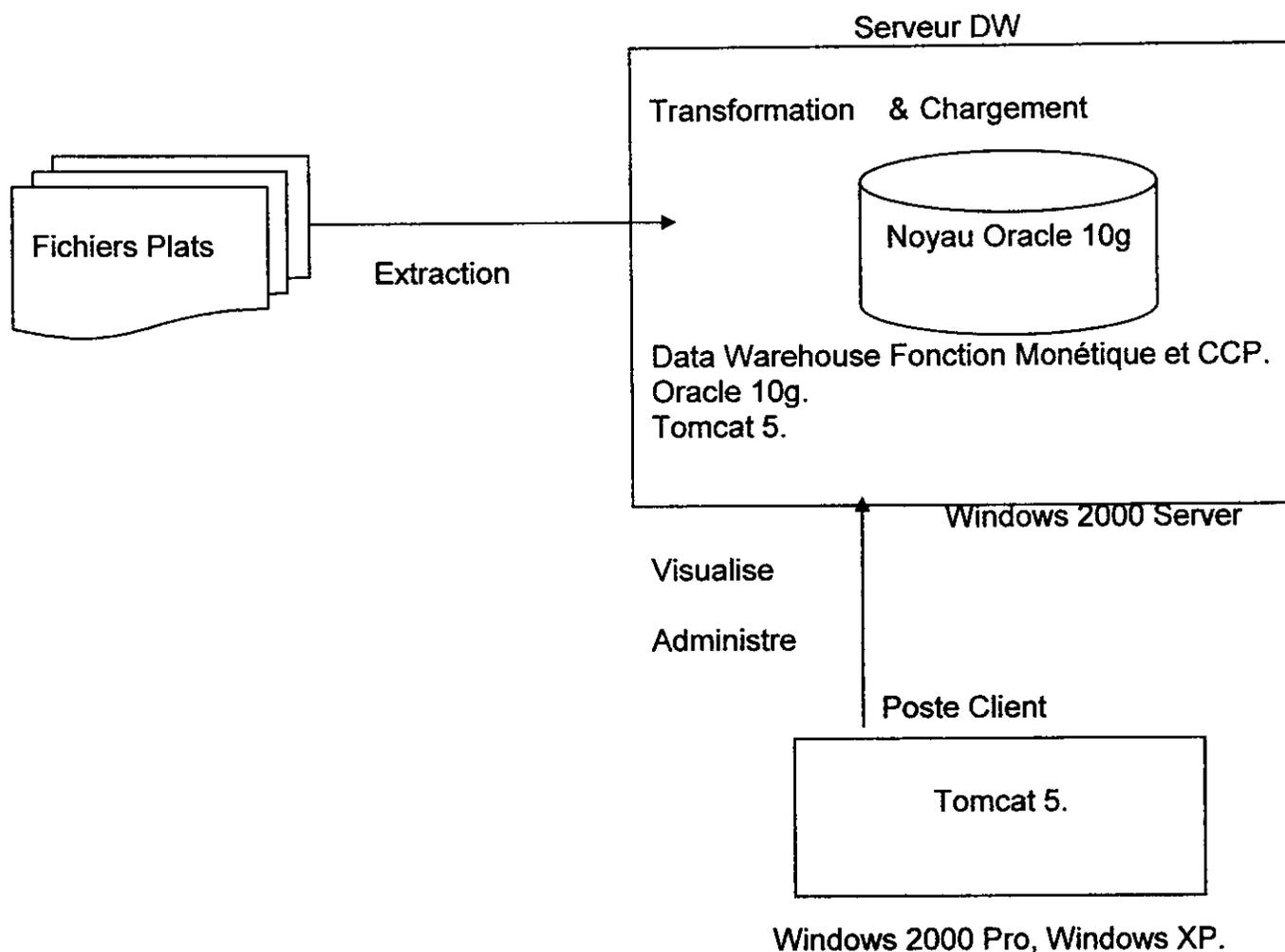


Figure VI.29 : Environnement logiciel.

#### I.4 Introduction aux outils utilisés dans le développement :

Nous commençons par citer tous les outils utilisés à la phase de développement dans un tableau récapitulatif, avant de passer à la description de l'utilisation de deux principaux outils décisionnel : Talend et Analytic Workspace Manager.

Chacun des outils et application utilisés lors du développement on été décrit dans des annexes, qui seront cités ci-dessous, ainsi que deux langage (Annexe -G-).

La tableau récapitulatif est illustré par le Tableau VI.5 comme suit :

**Tableau VI.5** : Tableau récapitulatif des outils utilisés pour le développement de l'outil d'aide à la décision.

Outils / Application	Version Utilisée
Oracle 10g (Data Warehouse) (Annexe –A-).	Oracle 10G release2 10.2.0.1.0.
Analytical Workspace Manager (Cube OLAP) (Annexe –B-)	Analytic Workspace Manager for Oracle OLAP 10.2.0.3.0.
Talend (ETL) (Annexe –C-)	Talend 2.0.4
Jdeveloper (Développement Tableau de bord) (Annexe –D-)	Jdeveloper 10g 10.1.2.2 build 1929
Tomcat (Environnement de développement) (Annexe –E-)	Tomcat 5
SQL Developer (Développement) (Annexe –F-)	SQL Developer 1.1.3.2766.

### 1.5 Implémentation :

Nous allons présentés l'utilisation des deux principaux outils décisionnel que nous avons appliqué dans notre projet : Talend (ETL) et AWM (Construction des Cubes OLAP).

Nous entamons cette partie par la description de l'utilisation d'ETL Talend ainsi qu'une bref description des fichiers plats a partir des quel les données sont extraites.

#### 1.5.1 Les fichiers Plats :

Comme nous l'avons dit au par avant, l'outil Talend extrait les données a partir des fichier plats dont le type est CSV.

CSV est un format informatique ouvert représentant des données tabulaires.

CSV est l'abréviation de l'anglais « comma-separated values », dont la traduction est : valeurs séparées par des virgules.

Nous avons été contraint de travailler avec trois types de fichiers plats, les voici :

Voici un exemple d'un des fichiers plats « CSV » utilisés lors de la population du Data Warehouse, illustré par la figure VI.30 :

```
"01/05/2007","01","16000 ","05:54","3000,0000"
"01/05/2007","01","16008 ","06:38","3500,0000"
"01/05/2007","01","16712 ","06:59","200,0000"
"01/05/2007","01","16308 ","07:03","500,0000"
"01/05/2007","01","16008 ","07:08","500,0000"
```

Cotes

Comma

**Figure VI.30** : Exemple d'un fichier CSV.

Dans l'Exemple qui suit nous illustrons un fichier « Delimited », c'est le même principe que CSV, y'a pas de Cotes et la virgule à été remplacé par un *Point Virgule*, cette exemple est illustré par la figure VI.31 qui suit :

```
0007033;2000.0;10:45;12/06/2007;LOCAL;RETRAIT
0007017;5000.0;10:46;12/06/2007;LOCAL;RETRAIT
0007052;7000.0;10:46;12/06/2007;LOCAL;RETRAIT
0007063;1000.0;10:48;12/06/2007;LOCAL;RETRAIT
0007063;1000.0;10:52;12/06/2007;LOCAL;RETRAIT
```



**Figure VI.31** : Exemple d'un fichier Delimited.

Dans l'exemple qui suit, nous illustrons le dernier type de fichier plat sur lequel nous avons travaillé. Le Fichier Positional ne contient aucun délimiteur. Il est illustré par la figure VI.32 qui suit :

```
114000700220061130405000000000700585
11400070022006113040500000000050585
114000700220061130405000000000200585
114000700220061130405000000000154233
114000700220061130405000000000200585
```

**Figure VI.32** : Exemple d'un Fichier Positional.

## 1.6 Description de l'utilisation des outils Talend et AWM:

### 1.6.1 L'ETL Talend :

Une fois le type de fichier plat déterminé nous passons à la description de Talend. Nous allons présenter les principales étapes de Talend.

- ✓ Choix du langage de programmation : Java/ Perl.
- ✓ Ouverture d'un ancien Job.
- ✓ La configuration d'un Job on utilisant la palette d'outil Talend.
- ✓ L'exécution du Job.

Choix d'un langage de programmation :

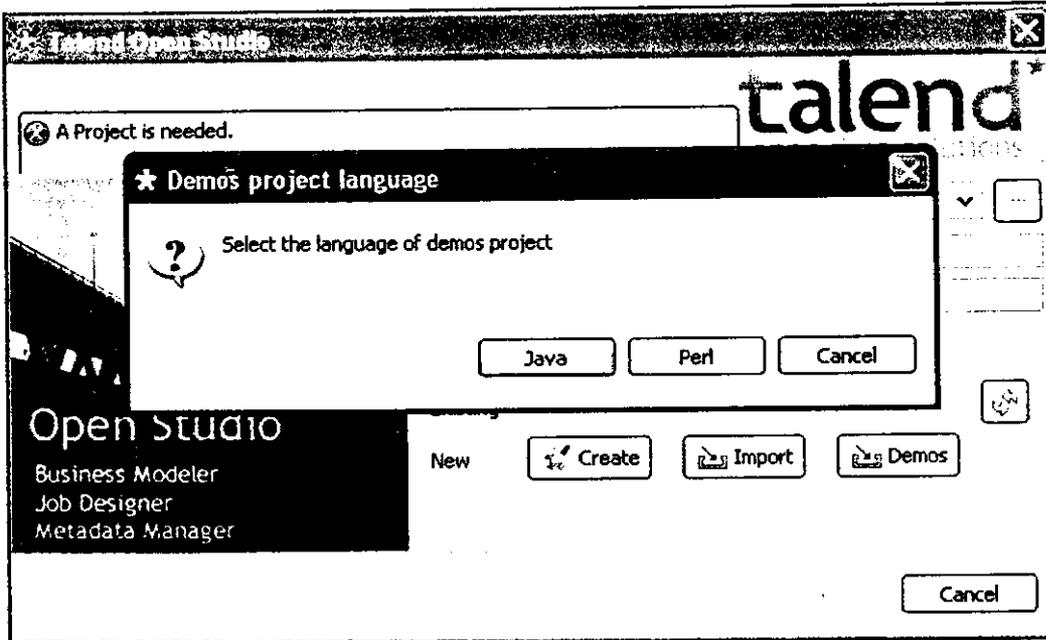


Figure VI.33 : Choix du langage de programmation.

Une fois le langage choisi, nous entamons la programmation.  
Nous voilàmes dans l'interface de Talend :

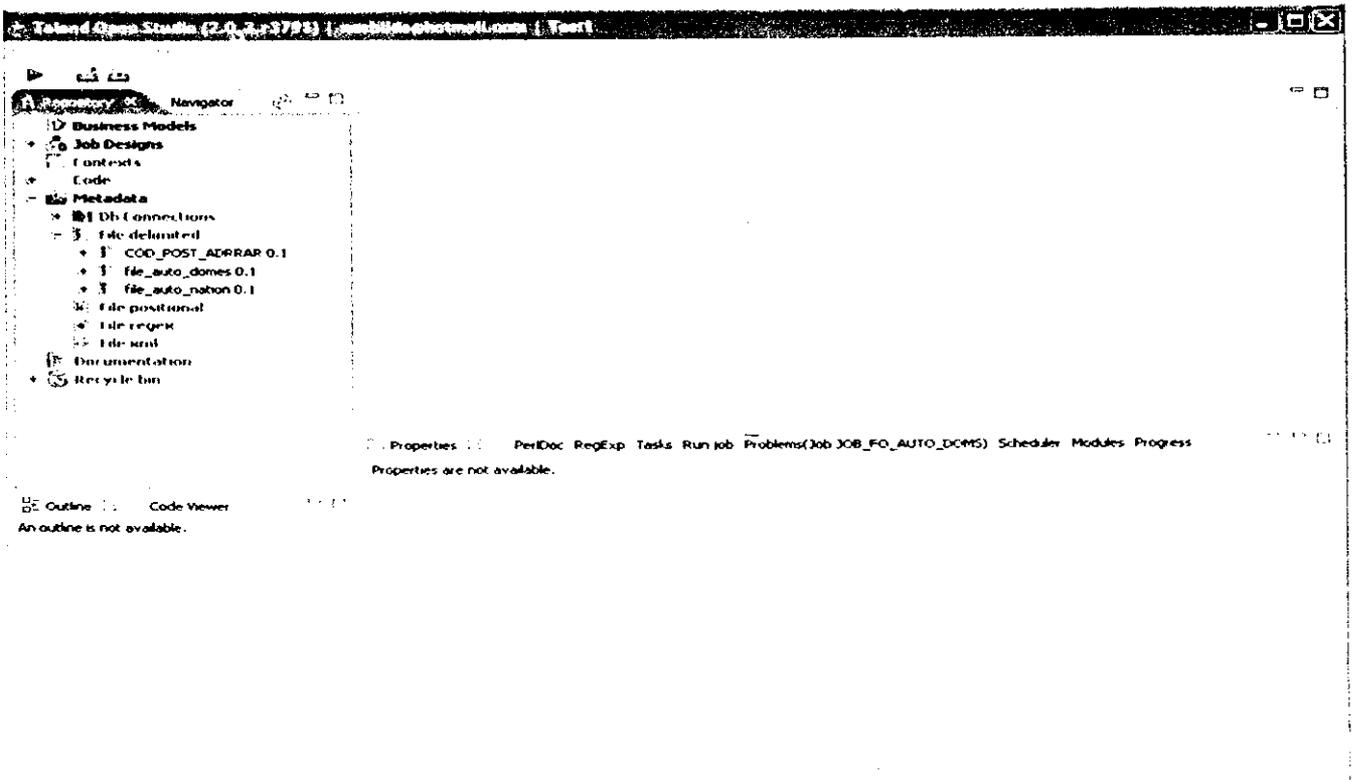


Figure VI.34 : Interface de l'ETL Talend.

## Création d'un Job :

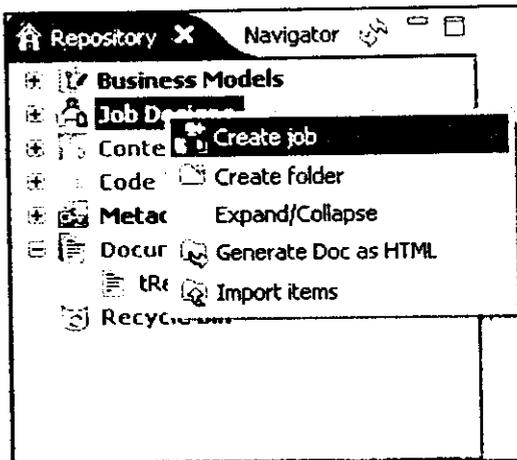


Figure VI.35 : Talend : Création d'un Job.

Une fois le Job Créé nous passons a la principale étape : la configuration de l'extraction des donnée.

Pour pouvoir extraire les données du fichier plats vers une table d'oracle, il faut indiqué le fichier plats par un input file et la table par un Output, et dans notre cas des transformation sur le fichier sont recommandé, cette transformation est intégré par un tMap comme suit :

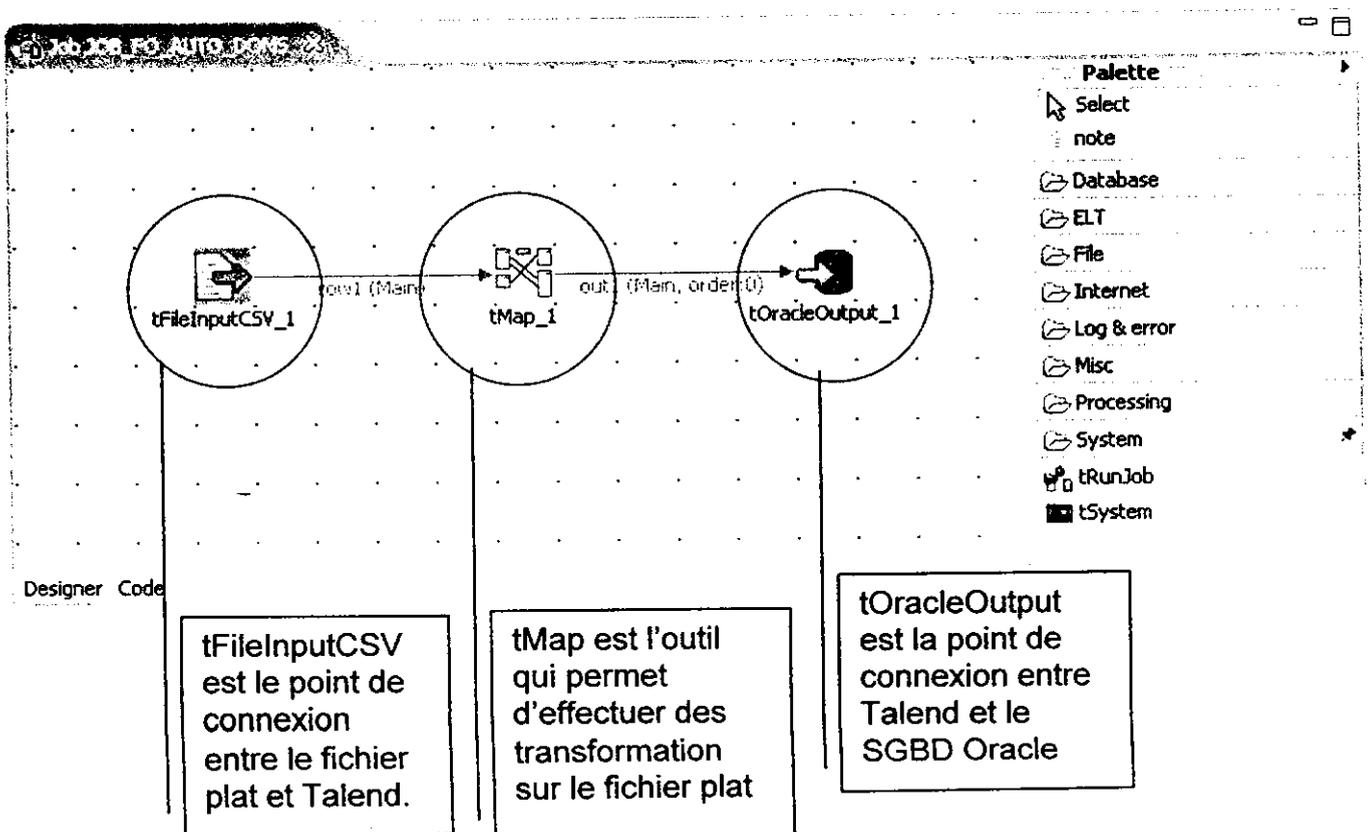


Figure VI.36 : Configuration du Job.

L'emplacement de la table de destination des données est indiqué a l'anglet propriété de l'élément tOracleOutput comme suit :

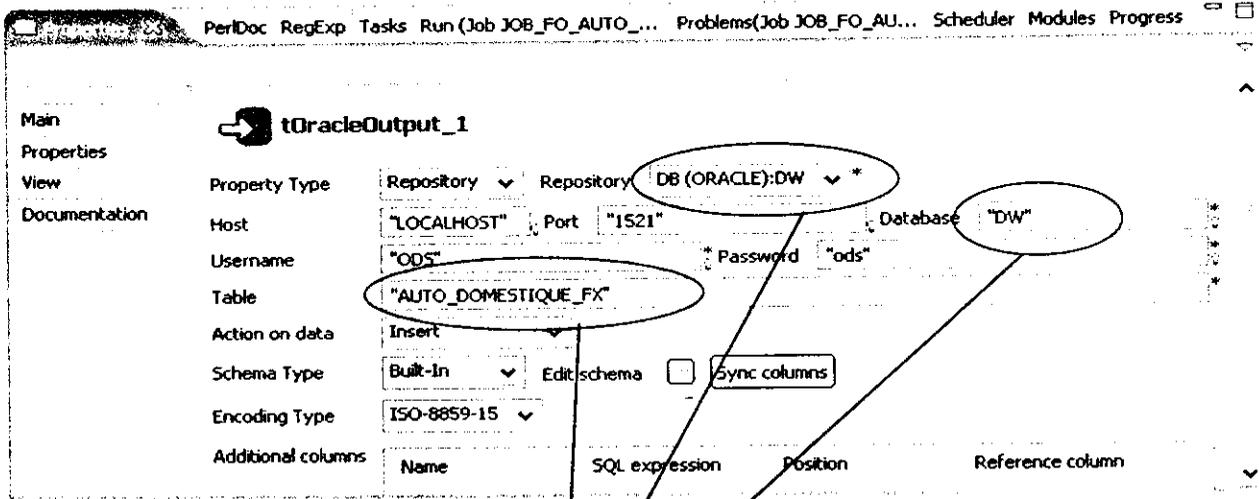


Figure VI.37: Indication sur le chemin de la table.

Pour l'élément tMap nous l'avons utilisé dans cet exemple pour remplacer la virgule dans le champ monnaie par un point, et pour convertir le Champ "DATE\_CMR" qui est de type "String" dans le fichier plat vers le type "Date" dans la table destination d'oracle.

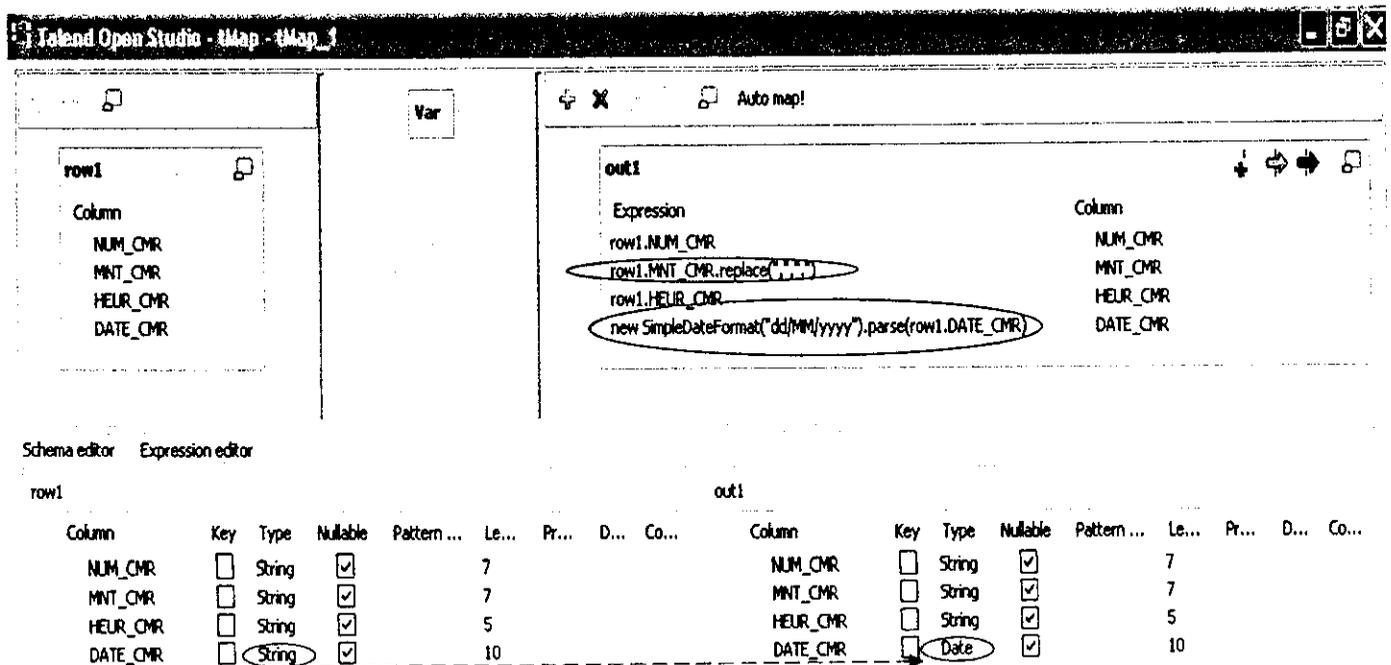


Figure VI.38 : L'utilisation de tMap.

Nous terminons cette phase par l'exécution du job comme suit :

Execution

Stats & Traces

Statistics  Traces

Clear before run  Exec time

Clear

Debug Run

029 | 6977270097  
045 | 0573565414  
091 | 1592849081  
087 | 1085890132  
037 | 8313533186  
009 | 8021041521  
058 | 4628116754  
018 | 9925876363  
087 | 7705131361  
023 | 6891797657  
072 | 7507976951  
061 | 8165006775  
057 | 9159275074

Job inputTable ended at 15:48 02/02/2007. [exit code=0]

Lance l'exécution du Job

Déroulement de l'exécution

Figure VI.39 : Exécution du Job.

**Remarque :**

Talend permet aussi de créer l'exécutable que le système peut exécuter afin de lancer l'extraction de tous les fichiers plats vers l'ODS.

**1.6.2 L'outil Analytic Workspace Manager:**

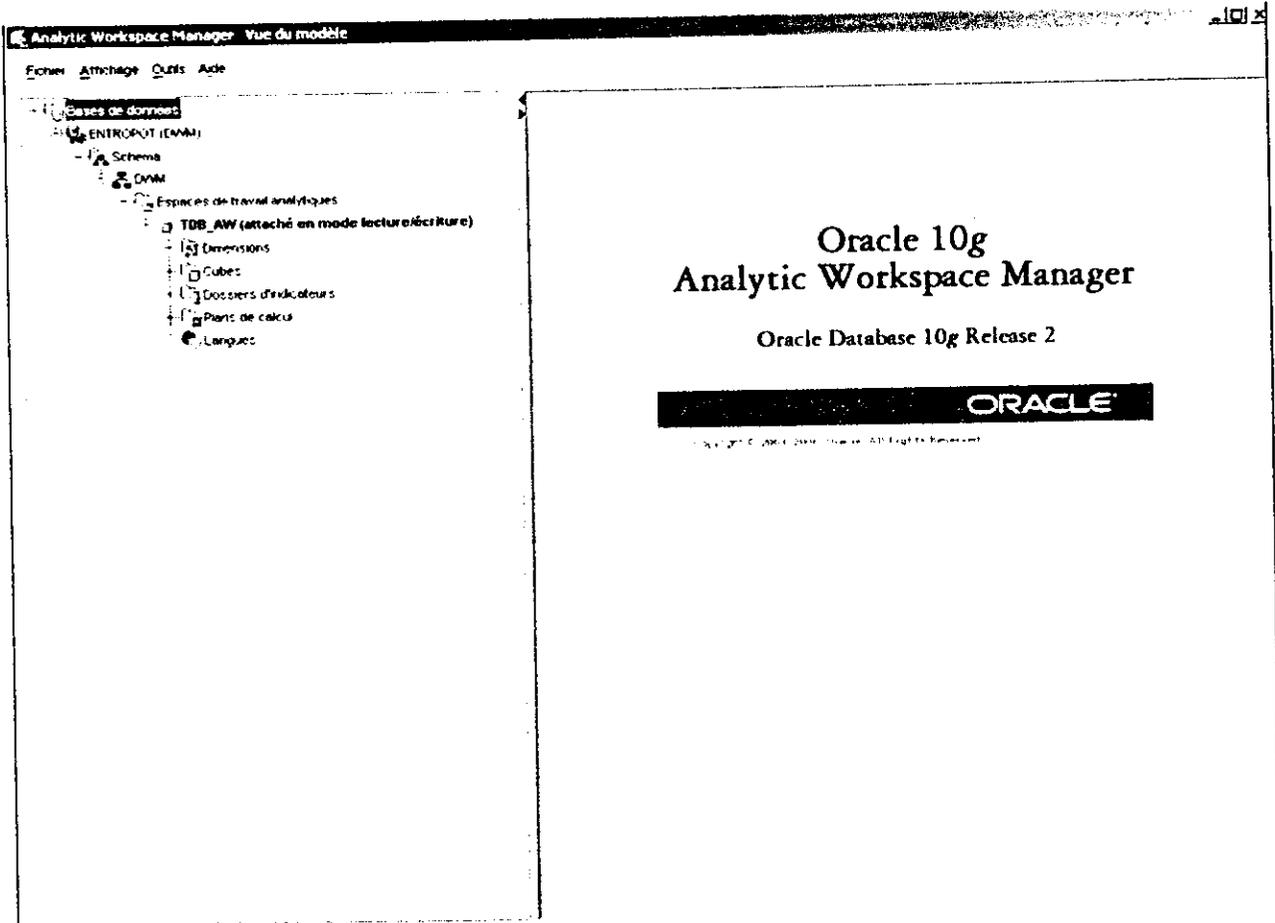
Une fois que la base de données est créée avec l'option OLAP nous passons à la création des cubes OLAP en utilisant AWM. La création se fera en 7 étapes comme suit :

**Etape 1 - Créer un utilisateur**

L'utilisateur "AW" est propriétaire de l'espace d'analyse (Analytic Workspace). Il a les droits en lecture sur le schéma DW, et les rôles qui lui permettent de créer les objets OLAP.

**Etape 2 – Se connecter avec Analytic Workspace Manager (AWM).**

Lancer AWM et se connecter à la base de données sous le compte "AW" comme suit :



**Figure VI.40** : Connexion a AWM.

**Etape 3** - Créer un espace analytique.

**Etape 4** - Créer des dimensions, niveaux, hiérarchies, attributs

Les dimensions sont formées d'un ensemble de valeurs qui permettent de découper les données. Elles forment les axes des cubes. Les dimensions peuvent contenir des niveaux, des hiérarchies et des attributs.

**Etape 5** - Créer le cube et les indicateurs

Un cube contient un ensemble de mesures selon des dimensions communes et qui ont vocation à être gérées de la même manière.

1. créer le cube.
2. Définir les niveaux d'agrégations dans le cube.
3. Créer les indicateurs (ou mesures)

### Etape 6 - Mettre en relation les objets et les données

On vas mettre en relation les objets créés dans les étapes 4 et 5 et les données relationnelles.

- Mettre en correspondance les colonnes de la table et les attributs des dimensions comme dans la copie d'écran ci-dessous :

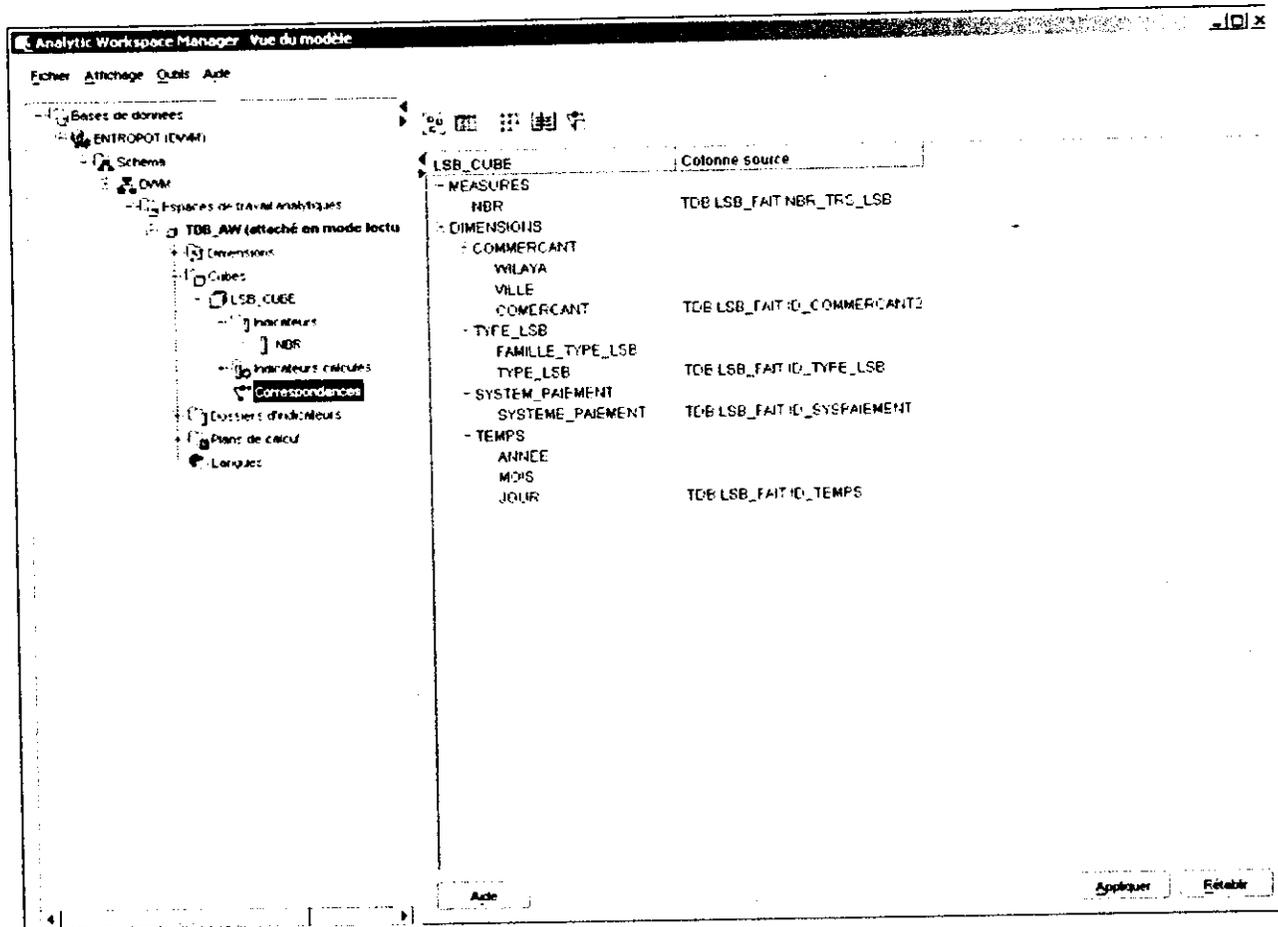


Figure VI.41 : Création de la correspondance entre les tables et leurs dimensions.

### Etape 7 - Charger et mettre à jour les données

La dernière étape de la création du cube consiste à lancer l'assistant de maintenance des données qui charge les dimensions et le cube avant de l'utiliser.

#### I.7 Conclusion :

Nous avons cité à ce chapitre tous les outils et application appliqué au développement de l'outil décisionnel, suivi de la description de l'utilisation d'un nouvel outil ETL : Talend, pour l'importance qu'il porte au projet, et de AWM. Ainsi qu'une description des différents fichiers plats exploités.

## Chapitre II

### **PARTIE VI : Réalisation**

#### **Description du Système**

Dans ce chapitre

**Interface d'Authentification.**

**Interface Administrateur.**

**Interface Tableau de Bord.**

**Conclusion.**

## VI. CHAPITRE II : DESCRIPTION DU SYSTEME :

Dans ce chapitre nous allons décrire l'outil de reporting. Nous commençons par décrire l'interface administrateur. Nous terminons par décrire l'interface du Tableau de Bord.

### II.1 Authentification :

La figure VI.42, montre la première page qu'un utilisateur voit s'afficher lorsqu'il accède à l'application. La page d'authentification.

Après la saisie du nom utilisateur et du mot de passe, cette page va faire appel à la page « identif.jsp » qui va vérifier la véracité des informations saisie.

Si c'est le cas, la page « admin.jsp » s'ouvre pour l'administrateur.

Si c'est un décideur la page « verif\_droit\_process.jsp » va vérifier les droits d'accès de ce décideurs et c'est sur ces droits que s'affichera l'interface Tableau de Bord avec accès sur les métiers (processus) le concernant.

Sinon la page « error.jsp » s'ouvre pour signaler une erreur de connexion.

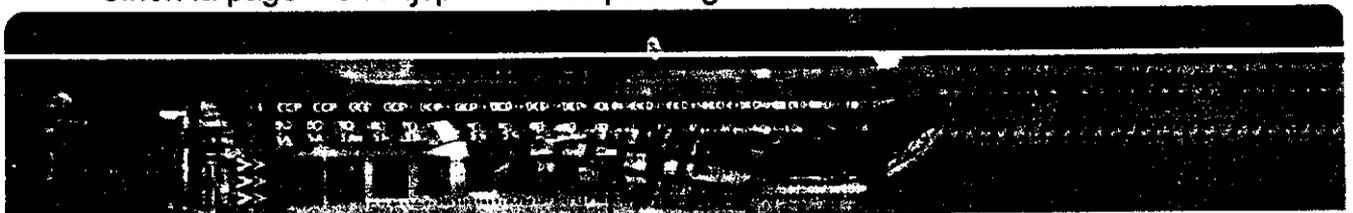


Figure VI.42 : Interface d'authentification.

### II.2 Interface Administrateur :

La Figure VI.43, montre la page du menu de l'administrateur (admin.jsp), qui contient une liste des tâches que ce dernier doit accomplir, entre autre : Ajouter utilisateur (creat\_user.jsp), Modifier droit utilisateur (atrid\_droits.jsp) par clic sur un utilisateur existant, Supprimer utilisateur (supp\_user.jsp).

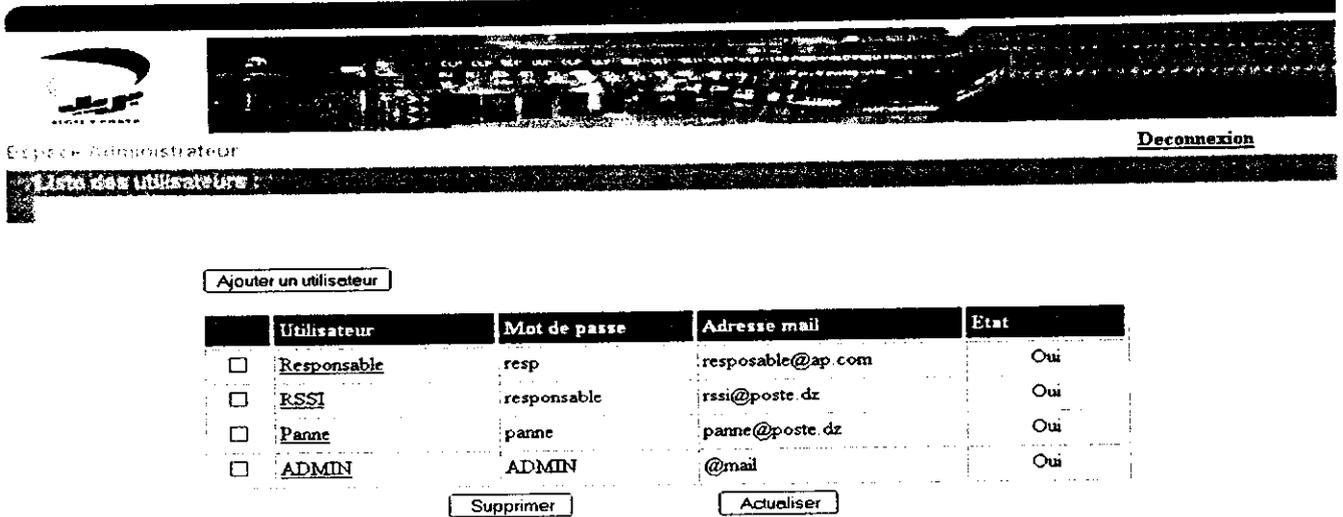


Figure VI.43 : Interface Administrateur

La figure VI.44, montre la page de l'Ajout d'un utilisateur. L'administrateur saisit les information du nouveau décideur, et poursuit pas cocher les droits d'accès de ce dernier, pour enfin valider la création du décideur (poste\_user.jsp).

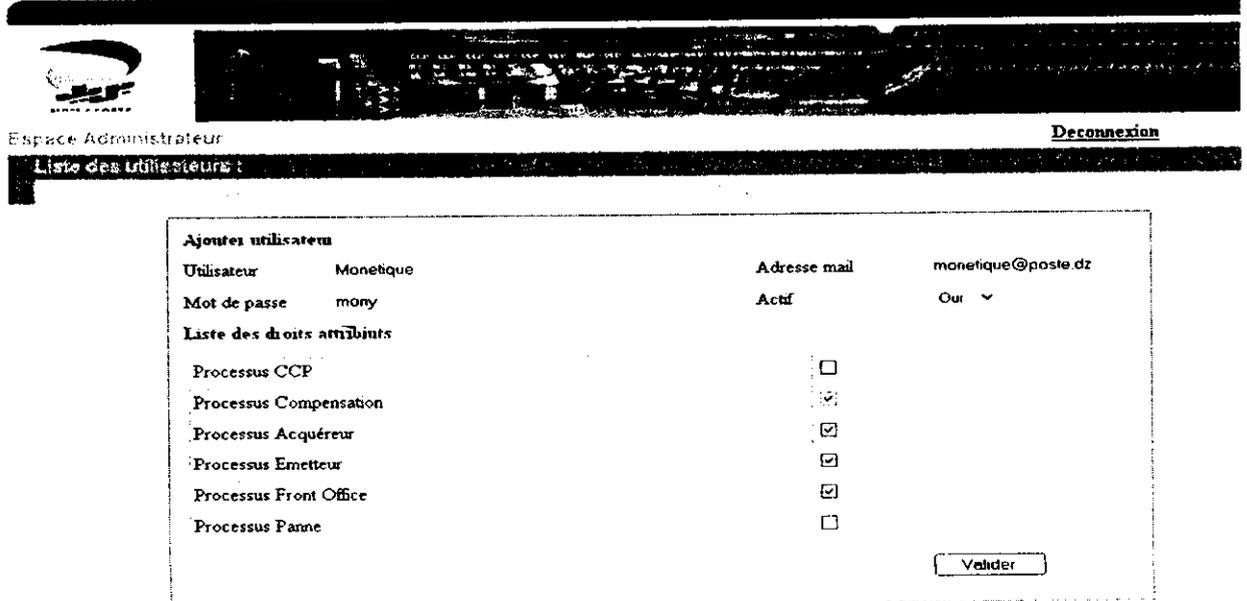


Figure VI.44 : Ajouter utilisateur

La figure VI.45, montre la liste de tous les utilisateurs de l'outil. Une fois la création du nouveau utilisateur validée, son nom s'ajoute a la liste des utilisateurs comme suit :

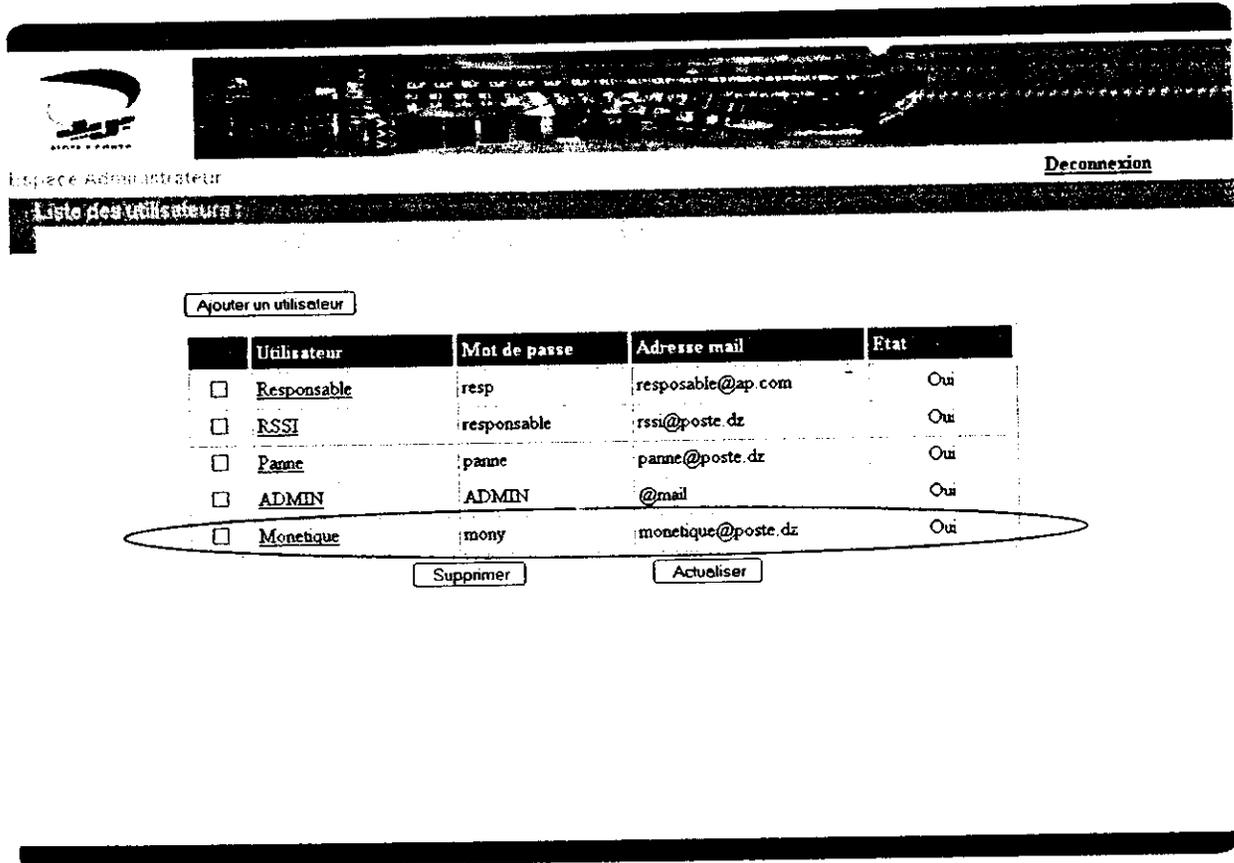


Figure VI.45 : Liste des utilisateurs du système.

### II.3 Interface Tableau :

La figure VI.46 représente l'interface du tableau de bord, d'un décideur qui a droit sur tous les métiers (tableau\_bord.jsp).

Cette interface représente deux métiers : CCP, Monétique.

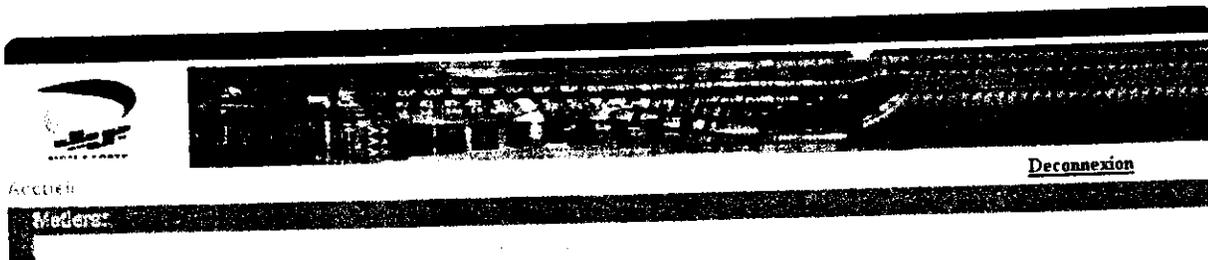


Figure VI.46 : Interface Tableau de Bord.

La figure VI.47, montre l'histogramme de l'activité LSB sur tout le territoire algérien. L'application offre la possibilité de choisir les axes d'analyse, entre autre dans l'activité LSB, le décideur peut :

- Choisir le Type d'LSB : Standard, Etendu ou Tout.
- Année.
- Mois : il peut choisir un mois en particulier, deux mois jusqu'à trois mois.
- L'application lui permet de voir aussi le résultat d'une année entière.
- De choisir une wilaya en particulier, une ville de cette wilaya en particulier.



Accueil > Métier monétaire

Derannexion

Processus:

Acquéreur

Emetteur

Front Office

Compensation

Panne

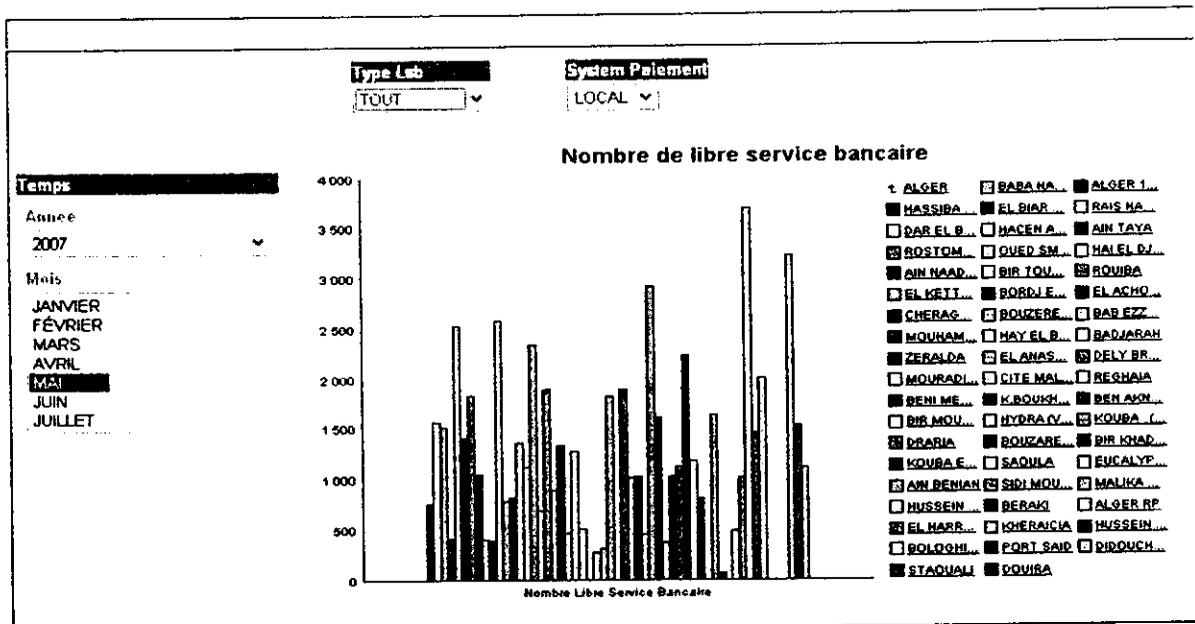


Figure VI.47 : Histogramme de l'activité LSB sur tout le territoire algérien.

La figure VI.48, représente un histogramme de l'activité LSB, après application d'un forage, sur la figure VI.47, d'une des wilayas du pays : Alger, en particulier d'une des villes de cette dernière : Zeralda.

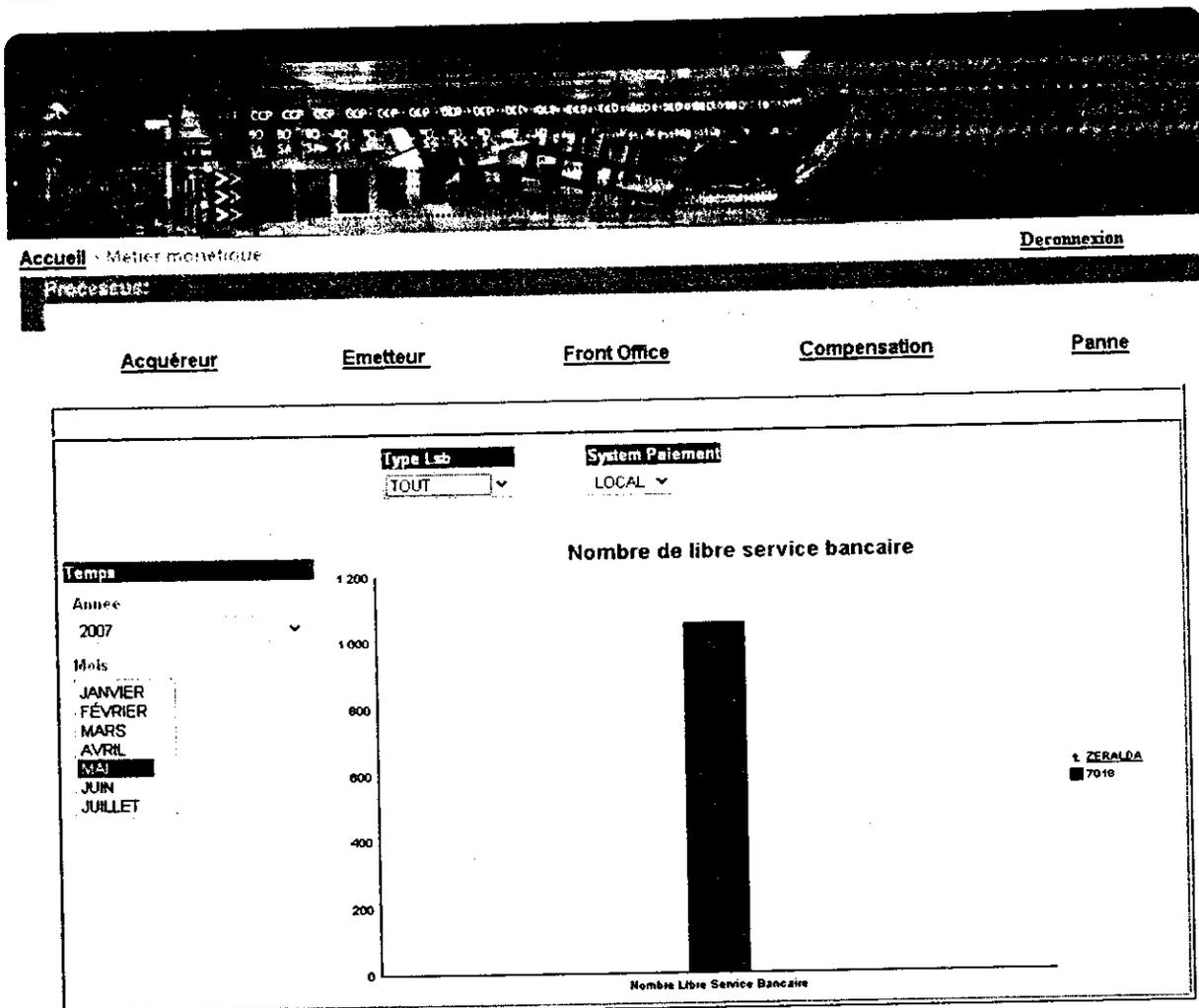


Figure VI.47 : Histogramme LSB, Forage vers le Bas.

#### II.4 Conclusion :

Nous avons présenté quelques aspects des tâches accomplies par l'outil de Reporting.

## CONCLUSION GENERALE

Le décisionnel couvre l'ensemble des activités visant à rassembler les données de l'entreprise et les mettre en forme afin de fournir des éléments d'aide à la décision.

C'est pourquoi l'entreprise Algérie Poste, nous a confié la conception et la réalisation d'un système d'information décisionnel pour les fonctions Monétique et CCP. Qui est une première dans cette entreprise.

Ce projet nous a offert la possibilité d'approfondir nos connaissances d'une manière méthodique dans :

- La conception et la réalisation d'un système d'information décisionnel.
- L'utilisation et l'application d'un outil ETL.
- La création et l'exploitation d'une base de donnée multidimensionnelle, basé sur la technologie MOLAP.
- La programmation d'une interface Web dynamique interagissant avec la base de donnée multidimensionnelle Oracle 10g, en utilisant le langage JSP.
- L'implémentation d'une architecture trois tiers en utilisant un Server Web Tomcat.

Parmi les problèmes rencontrés lors de la réalisation de ce travail, on cite:

- La difficulté dans la récolte des indicateurs. La plus parts des décideurs sont difficiles à interviewer.
- La difficulté dans l'acquisition des données a intégré dans l'Entrepôt de données, sachons que ces données concernent plusieurs responsables.

Afin d'améliorer et d'enrichir ce travail, nous proposons comme perspectives :

- Fixer avec l'entreprise le format définitif des fichiers plats dont le système extraira les données pour le rafraîchissement de l'Entrepôt de données;
- L'extension de l'Entrepôt de données en complétant le système financier par l'ajout des métiers CNEP et Mondat.
- Il y a aussi certain points dont on a pas eu le temps de traité et dont il serait intéressant de rajouté a l'étude : C'est l'étude de l'optimisation et la maintenance d'un Entrepôt de données, l'optimisation concerne le parallélisme des requêtes, le partitionnement et le choix le des indexes, et la maintenance concerne la gestion de l'archivage.

## Annexe -A- : « Oracle 10g »

### A.1 Historique : [Lone&Bryl, 2005].

Oracle 10g est l'avant dernière génération de SGBD de la société Oracle Corporation, un leader dans le domaine des bases de données.

Fondé il y a de cela 28 ans, Oracle a vu son nom changé à plusieurs reprises. En effet, Oracle Corporation a vu le jour sous le nom de SDL (Software Development Laboratories) formée par Larry Ellison, Bob Miner et Ed Oates afin de répondre aux missions proposées par le gouvernement américain. Ce dernier avait un besoin urgent de stocker de gros volumes d'informations et de les interroger rapidement. Ce projet fut considéré top secret, son nom de code fut Oracle.

Les entreprises choisissent la base de données Oracle plus qu'aucune autre pour ses performances, sa fiabilité et sa sécurité. Conçue pour tous les types d'activités, la base de données Oracle offre des avantages tels qu'une installation simple et rapide, et des fonctions complètes d'autogestion. Les grandes entreprises apprécieront quant à elles ses fonctions avancées, comme la mise en cluster.

La société Oracle Corporation est un fournisseur de logiciels de gestion de l'information. Cela comprend la gestion des bases de données, le développement d'application, le *Business intelligence* (Outil décisionnels) et la conception d'application de gestion orientés Internet. Elle se charge de la maintenance des applications pour ses clients. La société offre aussi un vaste éventail de services d'assistance (l'enseignement et la formation...)

L'ensemble des produits Oracle peut aujourd'hui être répartis en cinq catégories :

- La base de données Oracle 10g ;
- Oracle 10g Application Server ;
- Les outils de développement pour Internet ;
- Le datawarehousing et les outils de développement pour l'entreprise ;
- Oracle E-Business suite ;

### A.2 Oracle Data Warehouse: [Bria, 2006].

Oracle 10g a été conçu pour l'entrepôt de données. Pour la création des datawarehouses, quel que soit le volume de données que l'on prévoit de stocker dans l'entrepôt, Oracle 10g offre des fonctionnalités qui vont permettre d'optimiser ses performances.

#### A.2.1 Partitionnement :

Le partitionnement est l'une des fonctionnalités d'Oracle 10g. Il permet de :

- Séparer physiquement des données, optimisant du même coup les requêtes et le stockage.
- D'obtenir une rotation des données.

Le partitionnement se base généralement sur le temps.

Les avantages du partitionnement de données :

- *le volume de données et la facilité de gestion :*

Le partitionnement des tables dans une base de données représente une solution de simplification de gestion des bases de données volumineuses pour les raisons suivantes :

- o Les tables partitionnées sont généralement stockées dans des fichiers de plus petite taille (2 Go par exemple) qui sont faciles à sauvegarder.
- o En cas de défaillance, seul un faible pourcentage des données de la base est affecté si les partitions ont été stockées dans des tablespaces séparés.
- o Possibilité d'accomplir des opérations de maintenance sur une ou plusieurs partitions d'une table pendant que les autres partitions restent accessibles aux utilisateurs.

- *Les performances :*

Lorsqu'une table volumineuse est divisée en partitions, les requêtes qui l'affectent sont traitées plus efficacement car Oracle peut se contenter de lire uniquement la partition qui contient les données recherchées.

Le gain de performance et le principal avantage de partitionnement, quelle que soit la méthode adoptée.

#### **A.2.1.1 Les types de partitionnement :**

Il existe trois types de partitionnement en fonction du but à atteindre :

- Le partitionnement par plages,
- Le partitionnement par listes,
- Le partitionnement par hachages.

#### **A.2.2 Les fonctions d'agrégation :**

Oracle 10g offre de nombreuses fonctionnalités qui peuvent être exploitées à des fins d'analyse de données. L'agrégation ou synthèse de données joue un rôle très important dans la conception d'un entrepôt de données. Elles permettent notamment d'améliorer les performances de la plupart des ces systèmes. Oracle 10g est qualifié de *moteur de synthèse* avec ses vues matérialisées et fonctions d'agrégation qui permettent de créer des agrégats logiques et physiques de données.

##### **A 2.2.1 Objectifs des agrégats :**

- Optimisation des performances du plus grand nombre de catégorie de requêtes utilisateur possible,
- Etre totalement transparent vis à vis du requêteur et du concepteur d'application. (Une requête SQL ne doit jamais référencer un agrégat)

Quelques inconvénients de leur utilisation :

- Nécessiter des besoins de stockage raisonnable dans l'entrepôt.
- Avoir un impact minimum sur leurs reconstructions.
- Doivent être évolutifs et faciles à maintenir.

### A.2.2.2 Principe des agrégats :

Oracle a muni son SGBD d'un navigateur d'agrégat qui à la fonctionnalité d'optimiser et de réécrire la requête en fonction des agrégats définis et des métadonnées d'agrégation.

Les composants d'agrégation d'oracle sont :  
Les vues matérialisées et les dimensions.

#### 1. Les vues matérialisées :

Les vues ne peuvent être stockées dans la base car elles ne contiennent aucune donnée. Pour résoudre ce problème, Oracle Corporation a introduit un nouvel objet de base de données, la vue matérialisée (*materialised view*), contrairement à la vue, la vue matérialisée contient les données provenant de la requête qu'elle définit. Elle peut être dans la même base ou dans une base de données liée. Elle est synonyme de l'objet SQL SNAPSHOT (Table contenant le résultat d'une requête faite sur une table gérée dans une base distante).

Dans un data warehouse, une vue matérialisée stock le plus souvent des données résumées et pré calculées, et permet d'accélérer considérablement le traitement des requêtes.

Oracle peut choisir de réécrire une requête pour qu'elle utilise une vue matérialisée si la récupération des données par ce moyen est plus rapide que d'accéder directement aux tables de la base, ce comportement étant transparent pour l'utilisateur.

#### 2. Journaux de vues matérialisées :

Sont des composants permettant la mise à jour des vues matérialisées. Il contient les modifications apportées à la table à laquelle il est lié. Ses informations seront utilisées par Oracle pour le rafraîchissement des vues matérialisées mettant en jeu la table du journal.

#### 3. Dimension :

Les dimensions sont des méta-données qui définissent les hiérarchies représentées dans les données. Elles sont optionnelles et sont utilisées par le navigateur d'agrégat. Chaque niveau d'une hiérarchie est identifié par un *level key* nommé par un alias. Pour chaque hiérarchie de la dimension, on définit les relations père-fils des différents niveaux. Il est important de valider les dimensions créées afin d'être certain qu'elles seront utilisées par le navigateur d'agrégat.

### A.3 Les fonctions d'agrégats et d'analyse d'Oracle 10g : [Lone&Bryl, 2005].

Oracle 10g offre de nombreuses fonctionnalités puissantes qui facilitent la mise en œuvre et la gestion d'un Data Warehouse, parmi lesquelles on trouve les vues matérialisées que nous venons d'aborder. Il supporte aussi des fonctions SQL d'analyse et d'agrégation. Qui sont essentielles dans un tel environnement.

Avec Oracle 10g, nous pouvons créer des requêtes SQL qui exécutent des opérations semblables à celles que supporte la clause *GROUPE BY* d'une instruction *SELECT*, mais dont la puissance peut être multipliée par dix, grâce notamment aux fonctions *ROLLUP* et *CUBE*.

- La fonction *ROLLUP* :

La fonction *ROLLUP* est une extension simple à la clause *GROUP BY* de l'instruction *SELECT*. Son action est assez simple : elle génère des sous-totaux qui vont du niveau le plus détaillé vers le niveau le plus général.

- La fonction *CUBE* :

Cette fonction permet à une seule instruction *SELECT* de calculer toutes les combinaisons possibles de sous-totaux pour un groupe de dimensions. A noter qu'elle produit aussi un total général.

#### A.4 Les Fonction OLAP :

Entre les outils d'Oracle et le Data Warehouse existe une technique élaborée en 1993 par E.E Codd : OLAP, pour On-line Analytical Processing.

OLAP pour OnLine Analytical Processing, est apparu en 1993 dans le fameux livre blanc écrit par E.F CODD, à la demande de la société Arbor Software, maintenant Hyperion. Le but de ce procédé était de pouvoir sélectionner et de croiser plusieurs données, provenant de sources diverses, afin d'en tirer une information (valeur, statistique par exemple) implicite. Ceci a évolué pour aboutir de nos jours à une méthode d'analyse, offrant aux décideurs un accès rapide et de manière interactive à une information pertinente présentée sous divers angles, dimensions et hiérarchies, en somme, sous forme de cube. [Bria, 2006]

#### A.5 Conclusion :

Oracle est une des entreprises majeures dans le secteur des applications décisionnelles, intégrant ses outils au sein de son système de serveur de base de données Oracle Database 10 g, reconnu par tout le milieu professionnel.

Pour cela et du fait que le SGBD de l'entreprise est Oracle 10g, nous avons opté sur, pour le développement de notre application.

---

## **Annexe –B- : « Analytic Workspace Manager »**

### **B.1 AWM : [Bria, 2006].**

Est utilisé pour définir, peupler et rafraîchir les analytic workspaces. L'Analytic Workspace Manager est une application Java qui permet de créer, gérer, et mettre à jour les analytic workspaces. Cet outil permet aussi de rendre disponible l'analytic workspace pour les OLAP APIs, les BI Beans et les outils de BI. L'Analytic Workspace Manager est disponible sur le cd client de l'Oracle Database 10g. Il ne fait pas partie de l'Entreprise Manager.

L'Analytic Workspace Manager a deux vues des analytics workspaces: une OLAP Catalog View et un Object View. On peut bien entendu permuter entre les deux vues à partir du menu View. La vue OLAP Catalogue nous montre ainsi les metadata OLAP, les noms, les dimensions, les cubes et les mesures créés. La vue Object View, nous permet de naviguer à travers les diverses entités de l'analytic workspace : nom, dimensions, variables, relations notamment.

L'analytic workspace manager dispose d'un assistant pour créer un analytic workspace d'un cube. On peut le lancer depuis le menu Tools. Nous l'utiliserons pour construire l'analytic workspace de nos cubes.

#### **B.1.2 Analytic Workspace :**

Un analytic workspace est une entité qui agrège et stocke les données sous forme multi dimensionnelle au sein de la base de données.

---

## Annexe -C- : « Talend »

### C.1 Talend : [W1]

Talend est la solution d'intégration de données Open Source permettant de répondre avec efficacité à un très large éventail de besoins : alimentation de Data Warehouse, synchronisation de bases de données, transformation de fichiers de divers formats (XML, délimités, positionnels.), etc., c'est un ETL. (Extract, Transform, Load).

Un ELT : Extract, Load and Transform, est un outil où la phase transform est donc assurée par le SGBDR.

Talend est le premier éditeur logiciel Open Source intervenant sur le marché de l'intégration de données (ETL). Il a conclu des partenariats avec des acteurs majeurs du marché comme JasperSoft et MySQL (*MySQL Network Certified Partner*). La société est présente en France, aux Etats-Unis et en Chine, et propose des services de support, de formation et d'expertise autour de ses solutions.

Talend est un outil graphique facile à utiliser, dont le langage de développement est soit Perl, soit Java, il est multiplateforme (Linux et Windows) et permet de tout faire en mode Drag and Drop.

Cet avantage nous a poussé à choisir Talend comme ETL pour notre projet.

---

## Annexe -D- : « Jdeveloper »

### D.1 Jdeveloper: [W2]

Jdeveloper dans sa version 10g 10.1.2.2 build 1929 est un très puissant outil de développement java, qui permet le développement de bout en bout de toute application. C'est-à-dire de la modélisation UML au déploiement de l'application, passant par l'édition du code, le test et le debugage.

Pour simplifier le développement, Jdeveloper fourni aussi un large éventail d'outil couvrant tout le cycle de vie du développement, ainsi que de nombreux wizard très intuitifs utiles au développement rapide des applications.

Les caractéristiques citées existes dans de nombreux environnement, alors pourquoi Jdeveloper ?

La principale raison est la facilité de l'exploitation de la base de données Oracle 10g à travers Jdeveloper. En effet, le fait qu'il soit de la même famille que la base de données 10g lui a donné une longueur d'avance par rapport à ses concurrents, et le titre d'environnement incontournable du développement d'application utilisant le SGBD d'oracle.

Jdeveloper permet de programmer sans aucun problème de conflit et d'exploiter toute la puissance de la BDD grâce aux extensions qu'il offre, notamment dans le domaine du Data Warehouse. Ses interfaces permettent d'interphaser directement avec le serveur Data Warehouse d'oracle 10g grâce à une API commune et de développer des applications Data Warehouse en utilisant les différents wizards disponibles.

Cet avantage que n'offre aucun autre environnement nous a poussé au choix de Jdeveloper comme environnement de développement de notre logiciel.

---

## Annexe –E- : « Tomcat 5 »

### E.1 Tomcat 5 : [Fiel&Kolb, 2001].

Est un Serveur d'application Java permettant d'exécuter des servlets et des pages serveur Java (JSP). Il est développé sous licence open-source par la fondation Apache. Il peut être utilisé ou couplé avec un serveur Web (dont Apache), et porté sur n'importe quel système sur lequel une machine virtuelle Java est installée. Il doit savoir exécuter des applications web pour répondre aux requêtes entrantes.

Cela passe par des procédures de chargement de classes (nous sommes en Java), d'invocation dynamique... Il doit également être capable de convertir une requête en objet Java, pour qu'elle soit exploitable par l'application. Et, en retour, savoir convertir l'objet Java contenant la réponse générée, en réponse compréhensible par le serveur web. Ces objets Java sont régis par une API, qui répond aux spécifications Servlet officielles. On comprend dès lors qu'un serveur d'applications est subordonné à un langage. Hormis le concept, le serveur d'applis PHP n'a rien à voir avec un serveur d'applis Java !

Le coeur d'un serveur d'applications Java est le conteneur de servlets, puisque les servlets sont les éléments essentiels d'une application Web écrite en Java (elles reçoivent les requêtes et renvoient les réponses).

Le conteneur de servlets gère des servlets (sait où se trouvent physiquement les classes Java, pour quelles URL les appeler...), et les exécute lorsqu'elles sont demandées.

#### E.1.1 Servlet : [Fiel&Kolb, 2001]

Programme Java qui s'exécute dynamiquement sur le serveur Web et permet l'extension des fonctions de ce dernier, typiquement : accès à des bases de données, transactions d'e-commerce, etc.

Une servlet peut être chargée automatiquement lors du démarrage du serveur Web ou lors de la première requête du client. Une fois chargés, les servlets restent actifs dans l'attente d'autres requêtes du client.

#### E.1.2 JSP (Java Server Page) : [Fiel&Kolb, 2001]

Extension de la technologie Java Servlet de Sun qui permet de programmer simplement l'affichage de contenus dynamiques sur le Web. JSP consiste en une page HTML incluant du code Java qui s'exécutera soit sur le serveur Web, soit sur le serveur d'application.

Le langage HTML décrit la manière dont s'affiche la page, le code Java servant à effectuer un traitement, par exemple récupérer les informations nécessaires pour effectuer une requête dans une base de données.

---

## Annexe –F- : « SQL Developer »

### F.1 SQL Developer: [W3]

Oracle SQL Developer est un outil graphique gratuit pour le développement de base de données développé par Oracle.

C'est un outil qui améliore la productivité et simplifie les tâches de développement de base de données.

SQL Developer peut communiquer avec n'importe quelle version de Base de données d'Oracle 9.2.0.1 et plus, et fonctionne sur Windows, Linux et Mac OSX.

Cet outil n'est pas vraiment une nouveauté mais la suite de RAPTOR, cette amélioration est proposée par Oracle pour essayer de rivaliser TOAD de Quest Software.

Aujourd'hui disponible dans sa version 1.0, cet outil est complètement gratuit contrairement à son concurrent. Oracle SQL developer permet une exploration facile des différents objets d'une base de donnée, l'exécution de scripts et requêtes SQL.

Oracle SQL developer propose également une plateforme de développement PL/SQL.

Cet utilitaire permet de faciliter les tâches de gestion de base de données, il est multi plateformes, il fonctionne sur Windows, Linux et MacOSX.

Cet avantage nous a poussé au choix de SQL Developer.

---

## Annexe –G- : « Langages de Programmations »

### G.1 Le langage PL/SQL : [Bria, 2006].

Le langage PL/SQL est un langage L4G (un langage de quatrième génération), fournissant une interface procédurale au SGBD Oracle. Le langage PL/SQL intègre parfaitement le langage SQL en lui apportant une dimension procédurale.

Ainsi le langage PL/SQL permet de manipuler de façon complexe les données contenues dans une base Oracle en transmettant un bloc de programmation au SGBD au lieu d'envoyer une requête SQL. De cette façon les traitements sont directement réalisés par le système de gestion de bases de données. Cela a pour effet notamment de réduire le nombre d'échanges à travers le réseau et donc d'optimiser les performances des applications.

D'autre part le langage PL/SQL permet de faire appel à des procédures externes, c'est-à-dire des procédures écrites dans un autre langage (de troisième génération, généralement le langage C).

### G.2 Le langage JSP : [Dele,2005].

Les JSP (*Java Server Pages*) sont un standard permettant de développer des applications Web interactives, c'est-à-dire dont le contenu est dynamique. C'est-à-dire qu'une page Web JSP (repérable par l'extension *.jsp*) aura un contenu pouvant être différent selon certains paramètres (des informations stockées dans une base de données, les préférences de l'utilisateur,...) tandis que la page Web "classique" (dont l'extension est *.htm* ou *.html*) affichera continuellement la même information.

La syntaxe du JSP ajoute des balises XML, appelées *actions JSP*, qui peuvent être utilisées pour appeler des fonctions.

De plus, la technologie permet la création de bibliothèques de balises JSP (*taglib*) qui agissent comme des extensions au HTML ou au XML. Les bibliothèques de balises offrent une méthode indépendante de la plate-forme pour étendre les fonctionnalités d'un serveur HTTP.

Les JSP sont compilées par un compilateur JSP pour devenir des servlets Java. Un compilateur JSP peut générer un servlet Java en code source Java qui peut à son tour être compilé par le compilateur Java, ou peut générer le pseudo-code Java interprétable directement.

## GLOSSAIRE

**Axe d'analyse** : ou dimension. Les ventes d'un produit peuvent par exemple être analysées par région (axe 1), époque (axe 2), magasin (axe 3), etc.

**API** : *Applications Programming Interface*, Interface de programmation d'applications, a pour objet de faciliter le travail d'un programmeur en lui fournissant les outils de base nécessaires à tout travail à l'aide d'un langage donné.

**Analyse multidimensionnelle** : concept qui définit les analyses effectuées par croisement de plus de trois dimensions (ou ensemble de données du même type ou encore axes).

**BI Beans** : Les business intelligence BI beans sont des composants réutilisables conçus pour des développements rapides d'applications OLAP. Les BI Beans peuvent effectuer de multiples opérations : connexion à la base de données, calculs analytiques, affichage de graphiques ou de tableaux de rapports. Ces beans sont bien entendu utilisables au sein d'applications Java ou JSP.

**Cube** : Structure multidimensionnelle permettant l'analyse d'informations observables en les segmentant sur un ensemble d'axes d'analyses.

**Décisionnel** : Processus d'utilisation de connaissances extraites par analyse des informations et des données générées par les processus métiers de l'entreprise pour déterminer la meilleure action à entreprendre, la meilleure décision à prendre.

**Datamart** : base de données, spécifique au monde décisionnel, orientée sujet ou métier. Un Datamart peut contenir des données dupliquées d'un Datawarehouse et/ou des données locales.

**Datamining** : Traitement et analyse statistiques de bases de données permettant d'établir des relations et des comportements types. Avec l'analyse multidimensionnelle classique, on sait ce que l'on cherche tandis qu'avec le datamining, on ne sait pas forcément ce que l'on cherche. On essaye plutôt d'établir des corrélations entre des données afin d'en tirer des renseignements, des indicateurs, des anomalies, des correspondances, etc. qui peuvent mettre en évidence des tendances.

**Data warehouse** : est un entrepôt de données, centralise des informations extraites de différentes applications et bases de données rattachées à l'environnement de production, à l'activité commerciale, etc., pour faciliter la prise de décisions. Synchrones ou asynchrones, les mécanismes d'extraction des données sont régis à travers des méta-données. Ils alimentent une base de données dite multidimensionnelle, dans laquelle le stockage des informations est organisé par axes d'analyse pour accélérer la consultation et les agrégations. Un data warehouse est une Base dans laquelle les données sont centralisées et organisées pour le support d'un processus d'aide à la décision. Le data warehouse peut être segmenté en multiples datamarts, sous-ensembles ou cubes multidimensionnels, dans lesquels le nombre d'axes a été réduit.

**ERP** : Enterprise Resource Planning, aussi appelés Progiciels de Gestion Intégrés (PGI), sont des applications dont le but est de coordonner l'ensemble des activités d'une entreprise (activités dites verticales telles que la production, l'approvisionnement ou bien horizontales comme le marketing, les forces de vente, la gestion des ressources humaines, etc.) autour d'un même système d'information

**ETL** : Extract Transform Loading. Serveur chargé d'extraire, nettoyer et transformer les données émanant de sources diverses pour ensuite les insérer dans une base de données (datawarehouse, datamart, etc.).

**EIS** : Un EIS (Executive Information System) est un outil permettant d'organiser, d'analyser et de mettre en forme des indicateurs afin de constituer des tableaux de bord. Ce type d'outil, facile à utiliser, ne permet de manipuler que des requêtes préalablement modélisées par le concepteur.

**Indicateur** : statistique, suivie au fil du temps, qui présente les tendances d'une condition ou d'un phénomène, au-delà des propriétés de la statistique elle-même. Les indicateurs permettent d'obtenir de l'information supplémentaire. Ils offrent un moyen d'évaluer les progrès en vue d'un objectif. On peut concevoir toute sortes d'indicateurs : mesure de la rentabilité, des ventes, de l'évolution des clients, etc.

**L'informatique décisionnelle** : (en anglais : DSS pour Décision Support System ou encore BI pour Business Intelligence) désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données d'une entreprise en vue d'offrir une aide à la décision et de permettre aux responsables de la stratégie d'une entreprise d'avoir une vue d'ensemble de l'activité traitée, ce qui leur permet de prendre connaissance des données à un niveau global et ainsi prendre des décisions plus pertinentes, d'où le nom d'informatique décisionnelle.

**Le système d'information décisionnel** : est un ensemble de données organisées de façon spécifique, facilement accessible et appropriées à la prise de décision ou encore une représentation intelligente de ces données au travers d'outils spécialisés. La finalité d'un système décisionnel est le pilotage de l'entreprise.

**Modèle de données** : organisation des données dans une base, dans une structure de tables, colonnes, champs etc. pour l'univers relationnel.

**Modèle en étoile** : Technique de modélisation directionnelle, consistant à créer physiquement ce modèle sur une base de données relationnelle. Ce modèle distingue physiquement les tables de fait des tables de dimensions.

**Modèle en flocon** : Technique de modélisation dimensionnelle, dérivée de la modélisation en étoile. Dans ce modèle, les tables de dimensions y sont dénormalisées.

**Modèle relationnel** : Technique de modélisation consistant à modéliser une base de données en la décomposant en entité et en relations corrélant ces entités.

**Meta-données** : Les méta-données sont des informations qui renseignent sur la nature de certaines données. Les méta-données que l'on peut par exemple associer à un document sont : son titre, son auteur, sa date de création, etc. Dans le cadre du décisionnel, elles constituent une sorte de dictionnaire sur lequel le système s'appuie pour comprendre des données utilisées par les différentes applications qui alimentent le datawarehouse. Les intitulés "Client" d'un PGI et "nom" d'une application comptable peuvent contenir les mêmes informations mais le système ne peut le savoir que si un dictionnaire a été conçu pour lui indiquer qu'il s'agit de la même nature d'informations. Les méta-données englobent également l'ensemble des informations relatives à la provenance, à l'historique et aux traitements associés aux données d'un datawarehouse.

**Nettoyage de données** : Processus visant à homogénéiser les données pour les rendre exploitables. Le nettoyage des données assure leur intégrité en éliminant les doublons, en corrigeant l'orthographe et en supprimant ou complétant les champs non renseignés. Les opérations de nettoyage peuvent également couvrir le filtrage, l'agrégation, la vérification de relations, etc.

**OLAP**: On-Line Analytical Processing. Procédé permettant de pré-calculer certains croisements de données afin d'optimiser les performances de l'application décisionnelle. Des variantes de l'OLAP existent. HOLAP (OLAP Hybride) : mixte de SQL et d'OLAP, l'un passant le relais à l'autre en fonction des opérations effectuées.

**Outil de restitution** : ensemble des outils (requêteurs, tableaux de bord, etc.) permettant de restituer le résultat d'une analyse.

**Qualité des données** : Conformité structurelle des données à l'utilisation qu'on souhaite en faire. Améliorer la qualité peut consister par exemple en la correction des occurrences multiples d'un même objet, ou le renseignement de champs vides.

**OLTP** : Online Transaction Processing, Programme ayant une architecture permettant de gérer des transactions en temps réel. Des applications de réservation de billets de train ou de bourse nécessitent ce genre d'architecture.

**Requête** : C'est une demande envoyée au gestionnaire de Base de donnée serveur. Si celui-ci permet la gestion des données, le langage utilisé est le SQL. Dans un contexte d'infocentre, l'exécution des questions sur un serveur est le plus souvent interprétée.

**Requêteur** : Un requêteur (Outil de requêtage) permet à l'utilisateur final d'accéder aux données de l'entreprise de manière autonome, dans un langage proche de celui de son métier, pour effectuer des analyses, croiser des données, etc. Dans la plupart des cas aujourd'hui, ces requêteurs s'appuient sur les méta-données pour proposer à l'utilisateur final des intitulés qui corresponde aux mots qu'il utilise habituellement en lieu et place des intitulés des colonnes des bases de données, le système décisionnel se chargeant de convertir ensuite les requêtes dans un langage compris par la base de données.

**SQL** : langage de Requête Structuré. Le langage SQL est un standard défini par l'ANSI et l'ISO. Il est dérivé de l'algèbre relationnel et de SEQUEL (System R74). Il constitue aujourd'hui le plus petit commun dénominateur des langages du marché.

**SIAD** : (Système informatisé d'aide à la décision), (DSS, en anglais Decision support system) Système informatique intégré, conçu spécialement pour la prise de décision, et qui est destiné plus particulièrement aux dirigeants d'entreprise. Le système d'aide à la décision est un des éléments du système d'information de gestion. Il se distingue du système d'information pour dirigeants, dans la mesure où sa fonction première est de fournir non seulement l'information, mais les outils d'analyse nécessaires à la prise de décision. Ainsi, il est habituellement constitué de programmes, d'une ou de plusieurs bases de données, internes ou externes, et d'une base de connaissances. Il fonctionne avec un langage et un programme de modélisation qui permet aux dirigeants d'étudier différentes hypothèses en matière de planification et d'en évaluer les conséquences.

**Table de faits** : La table centrale du schéma de jointure en étoile.

**Transaction** : Une unité de travail indivisible.

## **Glossaire UML :**

**Abstraction** : L'abstraction est un des piliers de l'approche objet. Il s'agit d'un processus qui consiste à identifier les caractéristiques intéressantes d'une entité en vue d'une utilisation précise. L'abstraction désigne aussi le résultat de ce processus, c'est-à-dire l'ensemble des caractéristiques essentielles d'une entité, retenues par un observateur.

**Agrégation** : l'agrégation permet à une classe de contenir d'autres classes. Par exemple, la classe Segment peut agréger deux objets de classe Point pour son fonctionnement interne. Cela permet de factoriser les caractéristiques communes aux objets (meilleure maintenabilité).

**Interface** : une interface spécifie un ensemble de méthodes que doit implémenter une classe. C'est un contrat que doit respecter la classe.

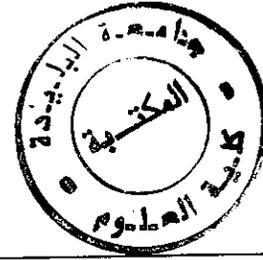
**Modèle** : Représentation schématique d'un processus, d'une démarche raisonnée. Il s'agit donc une abstraction de la réalité (voir abstraction). C'est une vue subjective, mais pertinente de la réalité.

**UML** : Unified Modeling Language, est un langage formel qui permet d'exprimer et d'élaborer des modèles objet

## Bibliographie

[Gard, 2000]	Support de cours : « Explosion de l'informatique décisionnelle » par G.Gardarin. 2000.
[Vand&Faul, 2001]	« Environnement évoluées et évaluation de l'IHM » Chapitre 3 : « présentation des données dans les systèmes d'information opérationnels et décisionnels » par : Jean Vanderdonckt et Stéphane Faulkner. 2001
[Heg, 2003]	Support de cours : informaticien de gestion / HES Option d'école / Base de données informatique décisionnelle. Haute école spécialisée de suisse occidentale.2003
[RAP, 2006]	Rapport annuel d'Algérie poste. 2006
[MON,2004]	Documentation sur la Monétique : La monétique et ses conditions de mise en œuvre. 2004
[Kim, 2001]	« Entrepôt de données, guide pratique du concepteur de Data Warehouse », Ralph Kimball Traduction de Claude Raymond, Edition Vuibert, 2001.
[CNAM, 1998]	Data Warehouse et data minig. Conservatoire National des Arts et Métiers de Lille. Version 1.1 le 15 Juin 1998.
[Bru&Cam,2004]	L'UML Décisionnel. Par Pascal Brunot et Hadrien Cambazard. 2004
[Gill&Tom, 2000]	Découverte de connaissance à partir de Data Warehouse. Gilleron & Tommasi, 2000
[Kim&all,2000]	Concevoir et déployer un data Warehouse, Guide de conduite de projet. Auteur de Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite. Traduction de Véronique Campillo, Emmanuelle Burr, Anne Kenn, Laurent Adam. 2000.
[Ako&CW, 2002]	Ingénierie des systèmes d'information : Entrepôt de données et bases multidimensionnelles. Par Jacky Akoka et Isabelle Comyn-Wattiau. 2002.
[Kim&all,2005]	Le Data Warehouse guide de conduite de projet. Par Ralph Kimball, Laura Reeves, Margy Ross et Warren Thornthwaite. 2005.
[LACO, 2005]	Le décisionnel, l'exception culturelle du système d'information. Par Olivier LACOTE, Directeur de l'activité Business Intelligence ORESYS Solutions. 2005
[Goua, 2001]	Le projet décisionnel, Enjeux, modeles et architecturesq du Data Warehouse. Par Jean-Marie Gouarné. 2001.

## Bibliographie suite :



[Liau, 2001]	E-business intelligence de Bernard Liautaud. 2001
[Kort&Mood, 2000]	« Form Enterprise Models to Dimensional Models : Methodology from Data Warehouse and Data Mart Design ». Daniel L. Moody & Mark A.R.Kortink Proceeding of the International Workshop on Design and Management on Data Warehouses (DMD W2000) Stockhol, Sweden, 2000.
[Golf&Riz, 1998]	"A Methodological Framework for Data Warehouse Design" Matteo Golfarelli et Stefano Rizzi Proceeding ACM First International Workshop on Data Warehousing and OLAP (DOLAP 98), 1998, Washington, D.C., USA.
[Bria, 2006]	Oracle 10g sous Windows. Gilles Briard. 2006.
[Lone&Bryl, 2005]	Oracle 10g Guide du DBA. Kevin Loney et Bob Bryla Edition Campus Press. 2005
[Booc&Rumb, 2000]	Guide de l'utilisateur UML. De Grady Booch, James Rumbaugh. 2000
[Fiel&Kolb, 2001]	JSP Java Server Page, Développement de sites web dynamique De Duane K. FIELDS, Mark A. KOLB Edition Eyrolles 2001
[Dele,2005]	JSP et SERVLETS EFFICACES, Production de sites dynamiques Cas pratiques. De Jean-Luc Deleage Edition DUNOD 2005

## Les thèses:

Auteur	Titre de la these
Ammar TAGUIDA Hamza DJEDDI	Conception et Déploiement d'un Data Warehouse : Cas de la gestion Eleve. De l'école polytechnique militaire. 2005
Sali Said Si Ahmed Hassiba	Conception et réalisation d'un tableau de bord basé sur une modélisation dimensionnelle pour la Gestin des Ressources Humaines à l'entreprise SONATRACH. Années académiques 2000-2001

## Les thèses (Suite) :

Auteur	Titre de la these
Hervé Colasuonno	Conception d'Entrepôts de Données. Examen probatoire en informatique. Années académiques 2000-2001 Sous la direction des profs Michel Simonet et Ana Simonet
Bolognini	« Etude pour la création d'un entrepôt de données dans le cadre de l'assurance vie et transformation des données en information utiles en vue d'une prise de décision », Université de Lausanne école des hautes commerciales Mémoire présenté par Nathalie RYSER BLOGNINI en vue de l'obtention du Diplôme poste grade en informatique et organisation, Années académiques 2000-2002 Sous la direction du prof. Thibault ESTIER.

## Les sites Web:

<a href="http://www.poste.dz/accueil">http://www.poste.dz/accueil</a>	Le site officiel d'Algérie Poste.
<a href="http://www.oracle.com">www.oracle.com</a>	Le site officiel d'oracle
<a href="http://perso.wanadoo.fr/bernard.lupin">http://perso.wanadoo.fr/bernard.lupin</a>	Le site contenant des informations sur OLAP
[W1] : <a href="http://www.talend.com">www.talend.com</a>	Le site Officiel de l'ETL Talend.
<a href="http://www.systemeetl.com/">http://www.systemeetl.com/</a>	Site consacré au système ETL.
[W2] : <a href="http://www.oracle.com/lang/fr/tools/jdev_home.html">http://www.oracle.com/lang/fr/tools/jdev_home.html</a>	Site consacré a Jdeveloper.
<a href="http://www-igm.univ-mlv.fr/~dr/XPOSE2003/tomcat/">http://www-igm.univ-mlv.fr/~dr/XPOSE2003/tomcat/</a>	Site consacré a Tomcat.
[W3] : <a href="http://www.oracle.com/technology/software/products/sql/index.html">http://www.oracle.com/technology/software/products/sql/index.html</a>	Site consacré à SQL Developer.
<a href="http://www.commentcamarche.net/plsql/plsqlintro.php3">http://www.commentcamarche.net/plsql/plsqlintro.php3</a>	Site consacré à PL/SQL.
<a href="http://www.commentcamarche.net/jsp/jspintro.php3">http://www.commentcamarche.net/jsp/jspintro.php3</a>	Site consacré au développement en JSP.
<a href="http://www-adele.imag.fr/users/Didier.Donsez/cours/#dw">http://www-adele.imag.fr/users/Didier.Donsez/cours/#dw</a>	Cours complet sur le Data Warehouse.