

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université SAAD Dahlab de Blida



Faculté des Sciences
Département d'informatique

Projet de Fin d'Etude en vue de l'obtention
du diplôme de MASTER en Informatique
Option : Ingénierie du Logiciels

**Conception et Réalisation
d'un Système Décisionnel
pour le Marketing**
Organisme d'accueil : ATM Mobilis



Réalisé par :

Promotrice :

- ✍ CHAOUA Nabil
- ✍ MEDJBER Hamza

- ✍ Mme Bensettiti

Soutenus le : 06/10/2010 Devant le jury composé de :

Mme Abed	Président	Grade	Maitre de conférence en USDB
Mr Bennouar	Examineur	Grade	Maitre de conférence en USDB
Mr Ould Aissa	Examineur	Grade	Maitre de conférence en USDB
Mme Bensettiti	Rapporteur	Grade	Maitre de conférence en USDB

Année universitaire :
2009 – 2010

Remerciements Chaoua

أبدأ الشكر بحمد الشكور الذي أمد عمري، أحمدته ما وفي حق نعمه التي أنعم بها علينا، فالحمد لله
ملء السماوات وملء الأرض وملء ما بينهما، الحمد لله على نعمة العقل الذي كرمنا به على
سائر خلقه.

ومصدقا لقوله جل وعلا في كتابه "أَنْ اشْكُرْ لِي وَلِوَالِدَيْكَ إِلَيَّ الْمَصِيرُ"، أشكر والديّ على كل
ما قدماه ومازالا يقدمانه إلي من عون مادي ومعنوي، فلهما كل الفضل بعده سبحانه على ما أنا
عليه الآن.

وقيل من لا يشكر الناس لا يشكر الله، فأشكر كل من علمني حرفا خلال حياتي، وأخص بالذكر
الأستاذ مهديم الذي كان له الفضل في تغيير نظرتي إلى العلم، فقد تعلمت منه احترام العقل الذي
وهبني إياه الله.

أشكر كذلك كل من ساهم في هذا العمل من قريب أو بعيد وأخص بالذكر السيدة حابس
(المؤطرة) والسيدة قلبي (رئيسة قسم التسويق الاستراتيجي في مديرية التسويق) وكل العاملين في
القسم التقني وخصوصا سليم أوودي الذي ساعدنا كثيرا في هذا المشروع.
كما أشكر المشرفة السيدة بن ستيي.

أهدي هذا العمل إلى جميع زملائي وخصوصا زملاء الجامعة.
كما أهديه إلى جميع إخوتي، وأختي الوحيدة، وأغتتم الفرصة للترحم على أخي الكبير اعمر الذي
انتقل إلى الرفيق الأعلى فأسأل الله أن يتغمده برحمته ويسكنه فسيح جنانه.
وأخيرا أشكر زميلي في هذا المشروع حمزة مجبر، أشكره على صبره علي خلال فترة إنجازنا لهذا
العمل، أشكره على تحمله إياي، أشكره على صبره علي خلال فترة مرضي.

اللهم لا علم لنا إلا ما علمتنا إنك أنت العليم الحكيم
اللهم علمنا ما ينفعنا وانفعنا بما علمتنا إنك أنت العليم الحكيم
اللهم أرنا الحق حقا وارزقنا اتباعه وأرنا الباطل باطلا وارزقنا اجتنابه

نبيل شعوة

Remerciements Medjber

Louange à Allah le tout-puissant, qui nous aide à tous instances,

Je tiens à remercier tous ceux qui nous ont aidés dans ce projet ou encouragé, je remercie mes parents pour ses soutiens de la Commission permanente avec moi, je remercie la promotrice Mme Bensettiti de l'acceptation de la surveillance sur ce projet, je remercie également les personnels de Mobilis qui nous a ouvert les portes de l'entreprise et nous donne ce que nous avons besoin des informations pour accomplir notre projet et en particulier Mme Habbes (direction des systèmes d'informations), Mme Guelfi (Direction de marketing), Mr Salim Oumedi qui avait nous a beaucoup aidés.

Je remercie également mes collègues pour leurs encouragements et ses conseils

Je remercie aussi tous mes frères et mes amis et tous ceux qui ont aidé à compléter ce travail.

Un grand merci à mon collègue dans ce projet, qui m'a appris la patience

Hamza Medjber

Résumé

Ce projet s'inscrit dans le cadre de conception et mise en œuvre d'un système d'information décisionnel **SID** pour le marketing. Ce travail a été fait durant un stage pratique au niveau de l'ATM Mobilis.

Un SID est défini comme un ensemble d'outils basant sur un stockage des données spécifique (data warehouse) permettant d'intégrer les données éparpillées sur les systèmes de production d'une entreprise sous format multidimensionnelle afin de les exploitées à travers des outils d'analyse. Un SID diffère d'un SI transactionnel par ses objectifs, ainsi par son architecture: Une architecture en couche qui mis en jeu 4 éléments essentiels (sources de données, ETL, Data Warehouse «DW », application de restitution). Chaque composant nécessite le développement d'un outil particulier. Cette particularité requiert une modélisation spécifique c'est la modélisation multidimensionnelle. Pour développer notre système on a utilisé la méthode DWEP (DW Engineering Process) proposé par *Luján Mora*. Cette méthode utilise deux standards : UML comme langage de modélisation et UP (Unified Process) comme méthode de conduit de projet. L'application de restitution est considérée comme la partie visible de l'iceberg d'un système décisionnel, c'est la partie la plus importante de point de vue des utilisateurs. Pour le développement on a utilisé le Kit Visual Studio de Microsoft (Sql Server pour le DW, SS integration service comme ETL, SS Aalysis sercice pour les cubes et ASP.NET pour l'application web).

Abstract

This project is part of design and implementation of a Decision support System (DSS) for marketing. This work was done during an internship at the ATM Mobilis.

A DSS is defined as a set of tools based on specific storage of data (datawarehouse) to integrate scattered data on the operational systems to multidimensional format used throughout the analysis tools. A DSS is different from a transactional system by its goals and its architecture. Layer based architecture contain four essential elements (data sources, ETL, Data Warehouse "DW," Application of restitution). Each component requires the development of a specific tool. This special feature requires a specific modeling: is the multidimensional modeling. To develop our system we used the method DWEP (DW Process Engineering) proposed by Luján Mora. This method uses two standards: UML as a modeling language and UP (Unified Process) as a method of project leads. The restitution application is a part of the iceberg of a decision supports system, is the most important users point of view. For development, we used the Kit Microsoft Visual Studio (SQL Server to DW, SS integration service such as ETL, SS Analysis service for cubes and ASP.NET for web application).

ملخص

هذا المشروع هو تصميم وتنفيذ نظام معلومات مساعد على اتخاذ القرارات موجه لإدارة التسويق. وقد تم هذا العمل خلال فترة تدريب في موبيليس أجهزة الصراف الآلي.

يعرف نظام المعلومات المساعد على اتخاذ القرارات (ن م ا ق) على أنه مجموعة من الأدوات القائمة على تخزين البيانات بصيغة محددة (مستودع البيانات) لدمج بيانات متناثرة على نظم المعلومات الإنتاجية والمسيرة للشركة وتخزينها على شكل متعدد الأبعاد، وذلك من خلال أدوات متعددة الأبعاد التحليل. ويختلف (ن م ا ق) عن أنظمة المعلومات التسييرية في الأهداف والبنية: بنية الطبقات التي تعتمد على 4 عناصر أساسية (مصادر البيانات، محلل البيانات، مستودع البيانات وإطار العرض). كل مكون يتطلب تطوير أداة محددة. تتطلب هذه الميزة نموذج محدد هو النمذجة متعددة الأبعاد. لتطوير نظامنا استخدمنا طريقة DWEP (هندسة العمليات لمستودع البيانات) التي اقترحها لوخان مورا. يستخدم هذا الأسلوب معيارين: UML كلغة للنمذجة، و UP (Unified Process) كوسيلة لإدارة المشاريع. وإطار العرض هو الجزء الظاهر من هذا المشروع، ويعتبر النقطة الأكثر أهمية في نظر المستخدمين. وقد تم استخدام أدوات تطوير مايكروسوفت البصرية ستوديو (Sql Serve) لمستودع البيانات، SS integration service لإنشاء المحمل ETL، SS Aalysis sercice لمكعبات البيانات و ASP.NET لتطبيق (ويب).

Liste des tableaux

TABLEAU 3.1.	POSITIONNEMENT DE LA METHODE [ANNONI ET AL, 2007] DANS LE CADRE DES 4 VUES	47
TABLEAU 3.2.	POSITIONNEMENT DE LA METHODE [LUJAN MORA 2005] DANS LE CADRE DES 4 VUES.....	48
TABLEAU 3.3.	POSITIONNEMENT DE LA METHODE [MAZON ET AL, 2005] DANS LE CADRE DES 4 VUES.....	49
TABLEAU 3.4.	POSITIONNEMENT DE LA METHODE [PRAKASH ET AL, 2003] DANS LE CADRE DES 4 VUES.....	50
TABLEAU 3.5.	POSITIONNEMENT DE LA METHODE [PRAT ET AL, 2002] DANS LE CADRE DES 4 VUES	51
TABLEAU 3.6.	POSITIONNEMENT DE LA METHODE [BONIFATI ET AL, 2001], DANS LE CADRE DES 4 VUES.....	52
TABLEAU 3.7.	POSITIONNEMENT DE LA METHODE [RAVAT ET AL, 1999] DANS LE CADRE DES 4 VUES.	53
TABLEAU 3.8.	POSITIONNEMENT DE LA METHODE [TRYFONA ET AL, 1999] DANS LE CADRE DES 4 VUES.....	54
TABLEAU 3.9.	POSITIONNEMENT DE LA METHODE [KIMBALL ET AL, 1998] DANS LE CADRE DES 4 VUES.....	55
TABLEAU 3.10.	LES DIFFERENTS MODELES UTILISES DANS LA CONCEPTION DES DW. (38)	59
TABLEAU 5.1.	EXTENSION D'UML PAR 15 DIAGRAMMES POUR LES SID (6)	76
TABLEAU 5.2.	DESCRIPTION DES ATTRIBUTS DE SYSTEM SOURCE	91
TABLEAU 5.3.	LES STEREOTYPES DES PACKAGES	95
TABLEAU 5.4.	LES STEREOTYPES DES ATTRIBUTS	95
TABLEAU 5.5.	LES STEREOTYPES DES CLASSES	96
TABLEAU 5.6.	GUIDELINES DE MODIFICATION AVEC LE DWCS	97
TABLEAU 5.7.	LES MECANISMES D'ETL ET LEURS ICONES.....	152

Liste des figures

FIGURE 1.1.	UNE VISION TRANSVERSALE	10
FIGURE 1.2.	LE DECISIONNEL AU SEIN DU SYSTEME D'INFORMATION.....	11
FIGURE 1.3.	DÉFINITION DU DATA WAREHOUSE	12
FIGURE 2.1.	ARCHITECTURE GENERALE D'UN SYSTEME DECISIONNEL.	14
FIGURE 2.2.	LE SYSTÈME D'ALIMENTATION (ETL)	16
FIGURE 2.3.	EXEMPLE DE TRANSFORMATION	17
FIGURE 2.4.	ETL MOTEUR.....	21
FIGURE 2.5.	ETL BASE (GENERATEUR DE CODE).....	22
FIGURE 2.6.	ETL BASE ET MOTEUR.....	23
FIGURE 2.7.	LA TECHNOLOGIE ROLAP	29
FIGURE 2.8.	LA TECHNOLOGIE MOLAP	30
FIGURE 2.9.	LA TECHNOLOGIE HOLAP	30
FIGURE 3.1.	LE CADRE DE REFERENCE DES QUATRE MONDES	41
FIGURE 3.2.	LES QUATRE VUES POUR LES METHODES DE DEVELOPPEMENT DES SID	42
FIGURE 3.3.	FACETTES ET ATTRIBUTS DE LA VUE SUJET	43
FIGURE 3.4.	FACETTES ET ATTRIBUTS DE LA VUE USAGE	44
FIGURE 3.5.	FACETTES ET ATTRIBUTS DE LA VUE SYSTEME	45
FIGURE 3.6.	FACETTES ET ATTRIBUTS DE LA VUE DEVELOPPEMENT	46
FIGURE 3.7.	EXEMPLE D'UN MODELE EN ETOILE	60
FIGURE 3.8.	EXEMPLE D'UN MODELE EN FLOCON	61
FIGURE 3.9.	EXEMPLE D'UN SCHEMA EN CONSTELLATION	61
FIGURE 4.1.	ORGANIGRAMME GENERALDE MIBILIS.....	69
FIGURE 4.2.	ORGANIGRAMME GENERAL DE LA DIRECTION DU MARKETING (DMK)	70
FIGURE 5.1.	LE PROCESSUS UNIFIER (UP) POUR LE DWEP	77
FIGURE 5.2.	LE CYCLE DE VIE MULTIDIMENSIONNEL DE RALPH KIMBALL	77
FIGURE 5.3.	ADAPTATION DE L'UP POUR LA DEMARCHE PROPOSEE	77
FIGURE 5.4.	LA DEMARCHE PROPOSEE	78
FIGURE 5.5.	DIAGRAMME DE CAS D'UTILISATIONS REPARTITION DES ABONNES (PART DU MARCHÉ)	83
FIGURE 5.6.	DIAGRAMME DE CAS D'UTILISATIONS COMPORTEMENT DES ABONNES (CONSOMMATION)	84
FIGURE 5.7.	CAS D'UTILISATION DE L'APPLICATION DE RESTITUTION(SUIVIE DES ABONNES).....	86
FIGURE 5.8.	CAS D'UTILISATION DE L'APPLICATION DE RESTITUTION(CONSOMMATION)	87
FIGURE 5.9.	MODELE (SCHEMA) CONCEPTUEL DES SOURCES DE DONNEES (SCS)	90
FIGURE 5.10.	LES 3 NIVEAUX DU MODELE DWCS UTILISANT DES PACKAGES.....	93
FIGURE 5.11.	EXTENSION D'UML PAR DES STEREOTYPES	93
FIGURE 5.12.	NIVEAU 1 DE DWCS	98
FIGURE 5.13.	NIVEAU 2 : LE SCHEMA TEMPOREL.....	99
FIGURE 5.14.	NIVEAU 2 : LE SCHEMA EFFECTIFS DES ABONNEES.....	99
FIGURE 5.15.	NIVEAU 2 : LE SCHEMA NOUVEAUX ABONNEMENT	99
FIGURE 5.16.	NIVEAU 2 : LE SCHEMA RESILIATION.....	100
FIGURE 5.17.	NIVEAU 2 : LE SCHEMA PENETRATION	100
FIGURE 5.18.	NIVEAU 2 : LE SCHEMA RECHARGEMENT	100

FIGURE 5.19.	NIVEAU 2 : LE SCHEMA RENTABILITE	101
FIGURE 5.20.	NIVEAU 2 : LE SCHEMA APPELS	101
FIGURE 5.21.	NIVEAU 2 : LE SCHEMA ETAT DE MARCHE	101
FIGURE 5.22.	NIVEAU 2 : LE SCHEMA DUREE DES APPELS	102
FIGURE 5.23.	NIVEAU 2 : LE SCHEMA SMS	102
FIGURE 5.24.	NIVEAU 2 : LE SCHEMA MMS	102
FIGURE 5.25.	SCHEMA DE CONSTELATION DES FAITS REPARTITION DES ABONNES	103
FIGURE 5.26.	SCHEMA DE CONSTELATION DES FAITS CONSOMMATION	104
FIGURE 5.27.	NIVEAU 3 : LA DIMENSION PERIODE «HEURE ».....	105
FIGURE 5.28.	NIVEAU 3 : LA DIMENSION OFFRE.....	105
FIGURE 5.29.	NIVEAU 3 : LA DIMENSION DUREE	106
FIGURE 5.30.	NIVEAU 3 : LA DIMENSION SEXE	106
FIGURE 5.31.	NIVEAU 3 : LA DIMENSION OPERATEUR.....	106
FIGURE 5.32.	NIVEAU 3 : LA DIMENSION DEMOGRAPHIQUE	107
FIGURE 5.33.	NIVEAU 3 : LA DIMENSION DATE «MOIS »	107
FIGURE 5.34.	NIVEAU 3 : LA DIMENSION TYPE RECHARGEMENT	108
FIGURE 5.35.	NIVEAU 3 : LA DIMENSION TRANCHE D'AGE	108
FIGURE 5.36.	NIVEAU 3 : LA DIMENSION VALEUR RECHARGEE.....	108
FIGURE 5.37.	NIVEAU 3 : LA DIMENSION TAILLE MMS.....	108
FIGURE 5.38.	NIVEAU 3 : LA DIMENSION PERIODE «JOUR».....	109
FIGURE 5.39.	NIVEAU 3 : LA DIMENSION TYPE TRAFIC.....	109
FIGURE 5.40.	NIVEAU 3 : LA DIMENSION ADRESSE.....	110
FIGURE 5.41.	NIVEAU 3 : LA DIMENSION PERIODE «MOIS »	110
FIGURE 5.42.	NIVEAU 3 : LA DIMENSION POINT DE VENTE	111
FIGURE 5.43.	NIVEAU 3 : LA DIMENSION OPERATEUR NATIONAL	111
FIGURE 5.44.	NIVEAU 3 : LE FAIT EFFECTIF DES ABONNES.....	112
FIGURE 5.45.	NIVEAU 3 : LE FAIT EFFECTIF DES ABONNES.....	112
FIGURE 5.46.	NIVEAU 3 : LE FAIT PENETRATION	112
FIGURE 5.47.	NIVEAU 3 : LE FAIT RESILIATION.....	113
FIGURE 5.48.	NIVEAU 3 : LE FAIT NOMBRE DES APPELS	113
FIGURE 5.49.	NIVEAU 3 : LE FAIT ETAT DE MARCHE	113
FIGURE 5.50.	NIVEAU 3 : LE FAIT RECHARGEMENT	114
FIGURE 5.51.	NIVEAU 3 : LE FAIT DUREE DES APPELS	114
FIGURE 5.52.	NIVEAU 3 : LE FAIT SMS.....	115
FIGURE 5.53.	NIVEAU 3 : LE FAIT MMS	115
FIGURE 5.54.	NIVEAUX DE DATA MAPPING	117
FIGURE 5.55.	EXEMPLE DE DMD (NIVEAU 0)	118
FIGURE 5.56.	EXEMPLE DE DMD (NIVEAU 1)	118
FIGURE 5.57.	EXEMPLE DE DMD (NIVEAU 2)	118
FIGURE 5.58.	EXEMPLE DE DMD (NIVEAU 3)	118
FIGURE 5.59.	DMD (NIVEAU 0).....	119
FIGURE 5.60.	MAPPING DES DIMENSIONS DEMOGRAPHIQUE	119
FIGURE 5.61.	MAPPING DES TABLES INTERMEDIAIRES SUR LES ABONNES	120
FIGURE 5.62.	MAPPING DES TABLES INTERMEDIAIRES SUR LES TRANSACTIONS	121

FIGURE 5.63.	MAPPING DES FAITS.....	122
FIGURE 5.64.	MAPPING DE RECHARGEMENT	123
FIGURE 5.65.	MAPPING DU FAIT ETAT DE MARCHÉ.....	124
FIGURE 5.66.	MAPPING DU FAIT PENETRATION.....	124
FIGURE 5.67.	3 EME NIVEAU DE MAPPING PACKAGE MAPABN1	125
FIGURE 5.68.	3 EME NIVEAU DE MAPPING PACKAGE MAPABN2	126
FIGURE 5.69.	3 EME NIVEAU DE MAPPING PACKAGE MAPABN3	127
FIGURE 5.70.	3 EME NIVEAU DE MAPPING PACKAGE MAPTr1	128
FIGURE 5.71.	3 EME NIVEAU DE MAPPING PACKAGE MAPTr2	129
FIGURE 5.72.	3 EME NIVEAU DE MAPPING PACKAGE MAP_EFFECTA	130
FIGURE 5.73.	3 EME NIVEAU DE MAPPING PACKAGE MAP_NOUVA	130
FIGURE 5.74.	3 EME NIVEAU DE MAPPING PACKAGE MAP_RESIL.....	130
FIGURE 5.75.	3 EME NIVEAU DE MAPPING PACKAGE MAP_DAP.....	131
FIGURE 5.76.	3 EME NIVEAU DE MAPPING PACKAGE MAP_SMS.....	132
FIGURE 5.77.	3 EME NIVEAU DE MAPPING PACKAGE MAP_MMS.....	132
FIGURE 5.78.	3 EME NIVEAU DE MAPPING PACKAGE MAP_NA	133
FIGURE 5.79.	3 EME NIVEAU DE MAPPING PACKAGE MAP_RENT.....	133
FIGURE 5.80.	3 EME NIVEAU DE MAPPING PACKAGE MAP_RECH1.....	134
FIGURE 5.81.	3 EME NIVEAU DE MAPPING PACKAGE MAP_RECH2.....	135
FIGURE 5.82.	3 EME NIVEAU DE MAPPING PACKAGE MAP_AAO.....	135
FIGURE 5.83.	3 EME NIVEAU DE MAPPING PACKAGE MAP_ETATM.....	135
FIGURE 5.84.	3 EME NIVEAU DE MAPPING PACKAGE MAP_PENET	136
FIGURE 5.85.	UNE ETAT D'ACTIVITE	137
FIGURE 5.86.	UN OBJET DU DIAGRAMME D'ACTIVITE	137
FIGURE 5.87.	DIAGRAMME D'ACTIVITE DE CHARGEMENT MENSUEL DES INFORMATIONS DES ABONNES	138
FIGURE 5.88.	DIAGRAMME D'ACTIVITE DE CHARGEMENT QUOTIDIEN DES INFORMATIONS DE TRAFIC.....	139
FIGURE 5.89.	STEREOTYPES POUR LA MODELISATION LOGIQUE SES ICONES.....	141
FIGURE 5.90.	SCHEMA LOGIQUE DU DWS	142
FIGURE 5.91.	SCHEMA LOGIQUE DU MINSAT	143
FIGURE 5.92.	SCHEMA LOGIQUE DU DDATA.....	143
FIGURE 5.93.	SCHEMA LOGIQUE DES FICHIERS PLAT.....	143
FIGURE 5.94.	SCHEMA LOGIQUE DE TOUTES LES SOURCES	144
FIGURE 5.95.	EXEMPLE D'UNE EVOLUTION AVEC REDONDANCE	146
FIGURE 5.96.	SCHEMA LOGIQUE DU DW (DWLS, PARTIE 1).....	147
FIGURE 5.97.	SCHEMA LOGIQUE DU DW (DWLS, PARTIE 2).....	148
FIGURE 5.98.	SCHEMA LOGIQUE DU DW (DWLS, PARTIE 3).....	149
FIGURE 5.99.	SCHEMA LOGIQUE DU DW (DWLS, PARTIE 4).....	150
FIGURE 5.100.	EXEMPLE DU MECANISME «AGREGATION»	152
FIGURE 5.101.	EXEMPLE DU MECANISME «CONVERSION» ET «LOG».....	153
FIGURE 5.102.	EXEMPLE DU MECANISME «FILTER».....	154
FIGURE 5.103.	EXEMPLE DU MECANISME «JOIN»	154
FIGURE 5.104.	EXEMPLE DU MECANISME «LOADER»	155
FIGURE 5.105.	EXEMPLE DU MECANISME «MERGE».....	155
FIGURE 5.106.	EXEMPLE DU MECANISME «CONVERSION»	156

FIGURE 5.107.	ETL PRECESS DES DIMENSIONS D'ABONNE	157
FIGURE 5.108.	ETL PRECESS DES DIMENSIONS OPERATEUR	158
FIGURE 5.109.	ETL PRECESS DE TRANSACTION ET ETAT DE MARCHE	158
FIGURE 5.110.	ETL PRECESS DES FAITS REPARTITION DES ABONNES	159
FIGURE 5.111.	ETL PRECESS DE RECHARGEMENT	159
FIGURE 5.112.	ETL PRECESS DES FAITS CONSOMMATION	160
FIGURE 5.113.	ETL PRECESS DE PENETRATION	161
FIGURE 6.1.	VISUAL STUDIO 2008 AVEC LES ADD IN DUNDAS	163
FIGURE 6.2.	POSITIONNEMENT DE QUELQUE SGBD (FORRESTER WAVE 2009)	166
FIGURE 6.3.	PART DU MARCHE DES EDITEURS D'SGBD (IDC 2009)	166
FIGURE 6.4.	NOMBRE DE FAILLES PAR EDITEUR (ESG 2007)	166
FIGURE 6.5.	LES SERVICES DE SQL SERVER 2008	167
FIGURE 6.6.	L'INTERFACE DE SSIS	168
FIGURE 6.7.	L'INTERFACE DE SSIS	168
FIGURE 6.8.	DIAGRAMME DE DEPLOIEMENT DES SOURCES	170
FIGURE 6.9.	DIAGRAMME DE COMPOSANTS DU DW	171
FIGURE 6.10.	DIAGRAMME DE DEPLOIEMENT DU DW	172
FIGURE 6.11.	DIAGRAMME DE DEPLOIEMENT DE L'ETL	173
FIGURE 6.12.	FLUX DE CONTROLE « ABONNE »	174
FIGURE 6.13.	FLUX DE CONTROLE « TRAFIC »	174
FIGURE 6.14.	INTERFACE DONNE LES INDICATEURS SOUS FORME DE TABLEAU DE BORD	175
FIGURE 6.15.	INTERFACE PERMET L'ANALYSE GEOGRAPHIQUE D'EFFECTIF DES ABONNES	176
FIGURE 6.16.	INTERFACE PERMETTANT L'ANALYSE GEOGRAPHIQUE D'EFFECTIF DES ABONNES PAR CLASSE/OFFRE, SEXE AVEC LA POSSIBILITE DE DRILLE-DOWN	176
FIGURE 6.17.	INTERFACE OFFRANT LA POSSIBILITE D'ANALYSE MULTIDIMENSIONNELLE D'EFFECTIF DES ABONNES	177

Table des matières

INTRODUCTION :	1
1. Contexte générale.....	1
2. Problematique	2
3. Objectifs	3
4. Organisation du mémoire	5
PARTIE I: PARTIE THEORIQUE	6
CHAPITRE I. Présentation du système décisionnel.....	8
I.1. Introduction :.....	8
I.2. Brève Historique sur les systèmes décisionnels:	8
I.3. Définitions :	9
I.4. Positionnement du décisionnel au sein du système d'information :	10
I.5. Le Data Warehouse :.....	11
I.6. Conclusion :	12
CHAPITRE II. Composants du système décisionnel	14
II.1. Introduction :.....	14
II.2. Architecture générale :.....	14
II.3. Systèmes Sources :.....	15
II.4. Zone de préparation des données:.....	15
II.4.2. Extraction des données :.....	16
II.4.3. Transformation :.....	16
II.4.4. Alimentation et indexation :	18
II.4.5. Générations des outils ETL :	18
II.4.6. Types d'ETL :	20
II.5. Serveur de présentation de data warehouse :.....	23
II.5.1. Base de données du Data Warehouse:.....	24
II.5.2. Data mart :	24
II.5.3. OLAP (Online Analytic Processing) :.....	24
II.5.3.1. Les (12 + 6) règles d'OLAP du CODD	25
II.5.3.2. Les technologies OLAP	28

II.6.	Outils de restitution :	31
II.6.1.	Introduction aux outils de restitution :	31
II.6.2.	Restitution (Reporting) :	31
II.6.3.	Analyse multidimensionnelle :	33
II.6.4.	Pilotage :	34
II.6.5.	Outils de statistiques et de Data-Mining :	34
II.7.	Conclusion :	36
CHAPITRE III. Méthodes de conception d'un SID		38
III.1.	Introduction :	38
III.2.	Définition d'une méthode de développement :	38
III.2.1.	Le modèle en cascade :	39
III.2.2.	Le modèle incrémental :	39
III.2.3.	Le processus unifié :	39
III.3.	Etat de l'art des méthodes de développement de SID:	40
III.3.1.	Le cadre de référence des quatre mondes :	41
III.3.2.	Positionnement de quelques méthodes d'ingénierie de SID dans le cadre de référence des quatre mondes :	47
III.3.3.	Modélisation des DW « Modélisation Multidimensionnelle (MD) »:	56
III.3.3.1.	Concepts de bases :	56
III.3.3.2.	Niveau conceptuel :	58
III.3.3.3.	Niveau logique :	59
III.3.3.4.	Niveau physique :	61
III.4.	Conclusion :	64
PARTIE II: Réalisation du système		65
CHAPITRE IV. Présentation de l'organisme d'accueil :		67
IV.1.	Introduction:	67
IV.2.	Organigramme generalde Mibilis:	67
IV.3.	Présentation de la direction de marketing:	69
IV.4.	Missions et objectifs de l'Entreprise :	71
CHAPITRE V. Conception et implémentation du système :		74
V.1.	Introduction :	74
V.2.	Méthode choisie :	74

V.2.1.	Présentation de la méthode DWEP (DW Engineering Process) :	75
V.2.2.	La démarche proposée :	76
V.3.	Conception du système :	79
V.3.1.	Définition des besoins :	79
V.3.2.	Modélisation du système au niveau conceptuel :	88
V.3.2.1.	Schéma conceptuel des sources (SCS) :	88
V.3.2.2.	Schéma conceptuel du DW(DWCS) :	92
V.3.2.3.	Schéma conceptuel de l'ETL (Data Mapping diagram) :	116
V.3.2.4.	Diagramme d'activité d'ETL (ETLAD) :	137
V.3.3.	Modélisation du système au niveau logique:	140
V.3.3.2.	Schéma logique des sources (SLS):	141
V.3.3.3.	Schéma logique du DW (DWLS) :	145
V.3.3.4.	Schéma logique d'ETL (ETL Process Diagram):	151
V.4.	Conclusion :	161
CHAPITRE VI.	Implémentation et Mise en œuvre du système.....	163
VI.1.	Introduction :	163
VI.2.	Choix d'outils de développement :	163
VI.2.2.	Microsoft Visual Studio 2008 :	164
VI.2.3.	Microsoft Sql Server 2008:	166
VI.3.	Modélisation physique du système :	170
VI.3.1.	Schéma physique des sources :	170
VI.3.2.	Modèle physique de DW :	171
VI.3.3.	Modèle physique de l'ETL (Transportation Schema) :	173
VI.4.	Implémentation d'ETL :	174
VI.5.	Quelques interfaces de l'application de restitution :	175
VI.6.	Conclusion :	177
CONCLUSION ET PERSPECTIFS :		179
Conclusion :		179
Perspectives :		180
Bibliographie		185

INTRODUCTION

INTRODUCTION :

1. Contexte générale :

Au début de l'informatisation les entreprises ont commencées par l'automatisation de leurs tâches et activités quotidiennes afin d'augmenter la performance et la rentabilité de ces activités (Vente, stock, production, facturation, comptabilité,...).

Au fil du temps ces applications ont engendrées une masse importante de données. D'autre part les entreprises se sont trouvées face à la mondialisation dont ces dernières doivent envisager leur planification stratégique, leurs productions, leurs stratégies marketing, leurs financements et leurs décisions d'investissement en tenant compte des marchés mondiaux. En outre, l'information pertinente, la capacité d'analyse, de découvert des problèmes, et surtout l'anticipation de future ont devenus des clés de la réussite et la continuité des entreprises.

Pour toutes ces raisons elles ont cherchées à exploiter les données produites par leurs systèmes de productions pour qu'elles servent de base à des analyses, à des prises de décision. Au premiers temps, elles ont commencées à interroger directement ces sources. Cette solution a extrêmement dégradé les temps de réponse de toutes les applications et a perturbée les activités courantes. Cette situation inacceptable a conduit à mettre en œuvre des moyens séparés spécifiques pour répondre aux besoins d'analyse et d'aide à la décision. Ainsi c'est la naissance de la Business Intelligence (BI) ou encore les Systèmes d'information Décisionnels (SID).

La Business Intelligence (BI) se définit comme l'ensemble des technologies permettant de traiter, valoriser et présenter les données à des fins de compréhension et de décision. Elle donne aux dirigeants une visibilité sur la performance de leur entreprise afin d'améliorer la capacité de celle-ci à réagir plus rapidement que ses concurrents face à de nouvelles opportunités ou aux risques du marché. Elle augmente la réactivité et la compétitivité des entreprises. La BI s'appuie sur un système d'information spécifique

Conception et Réalisation d'un Système décisionnel pour le marketing

appelé Système d'Information Décisionnel (SID), par opposition aux systèmes d'informations transactionnels. Les SID comportent plusieurs composants qui se résument autrefois en un entrepôt de données (Data Warehouse « DW »). Un entrepôt de données est une collection de données intégrées et historiées qui sont utilisées pour la prise de décisions stratégiques au moyen de techniques de traitement analytiques.

Toutes les activités de l'entreprise sont concernées par les SID et en sont des utilisateurs potentiels, par exemple :

- + Le contrôle de gestion pour l'analyse des coûts, l'analyse de la rentabilité, l'élaboration budgétaire, les indicateurs de performance, etc.
- + Le marketing pour le ciblage, le pilotage de gammes, les applications de géomarketing, de fidélisation clients, etc.
- + La direction commerciale pour le pilotage des réseaux (directs ou indirects), les prévisions des ventes, l'optimisation des territoires, etc.
- + Les ressources humaines pour la gestion des carrières, la gestion collective, etc.
- + La direction de la production pour l'analyse qualité, la prévision des stocks, la gestion des flux, la fiabilité industrielle, etc.

Ce mémoire présente les travaux d'un projet de fin d'étude pour l'obtention d'un diplôme de master en informatique, ces travaux s'inscrivent dans le domaine de BI et qui ont comme objectif de concevoir et réaliser un système décisionnel pour le marketing, ce projet a été réalisé durant un stage pratique au sein de l'entreprise ATM Mobilis (Opérateur de téléphonie mobile).

2. Problématique :

Durant les cinq dernières années, l'entreprise Mobilis a vu le nombre de ses abonnés augmenter considérablement, passant d'un million d'abonnés en 2005 à environ 8 millions en 2010.

Avec l'ouverture du marché national, l'Algérie a accueilli deux autres opérateurs étrangers de téléphonie mobile, en l'occurrence DJEZZY et Nedjma. Cela a été le prélude d'une longue phase d'érosion de la part de marché de Mobilis.

Conception et Réalisation d'un Système décisionnel pour le marketing

Du cours de l'exercice 2009, le marché de la téléphonie mobile commence à atteindre son niveau maximum. Il devient difficile, pour chaque opérateur, d'atteindre d'autres clients. Avec cette rude concurrence, l'entreprise Mobilis a opté pour une stratégie de maintien de sa part de marché.

Cette stratégie a été soutenue par un plan marketing permettant de maintenir les revenus à leurs niveaux actuels ou d'en limiter au moins l'érosion.

En sachant que le marché de la téléphonie mobile est très flexible (la clientèle est très sensible aux prix, aux services, ...), Mobilis ne ménage aucune effort pour assurer la rétention de sa clientèle en proposant de nouveaux produits et en opérant des promotions.

Mais l'état actuel ne permet pas aux dirigeants de marketing de bien suivre les plans marketing. par exemple, si ils veulent savoir l'effet d'une action marketing (Promotion, Publicité ...), ils s'appuient sur un outil de BI qui n'offre pas, ni l'information pertinente (informations sur les abonnés pré paid seulement sans analyse démographique), ni l'interfaces et les fonctionnalités attendus par un vrai outils d'analyse (interface sous Excel avec simples analyses).

Pour toutes ces raisons, les responsables de l'ATM Mobilis ont optés pour le développement d'un système d'information marketing décisionnel permettant aux responsables de prendre les décisions adéquates aux moments opportuns.

3. Objectifs :

Notre projet s'inscrit dans le cadre de conception, d'implémentation et de déploiement d'un système décisionnel pour le marketing au niveau d'ATM Mobilis, ce système à pour but de permettre aux responsables de marketing à travers des rapports statistiques, des analyses et des géo-analyses de :

- + Gérer et suivre les activités de la structure marketing.
- + Evaluer les plans marketing.
- + Etablir une gestion de la relation clients (CRM) stratégique.
- + Métriser mieux le marché de la téléphonie mobile.

Et cela à travers :

Conception et Réalisation d'un Système décisionnel pour le marketing

- **La spécification des besoins de département marketing**
 - Définition des futurs utilisateurs du système.
 - Spécification des besoins en termes de mesures et axes d'analyses pour chaque utilisateur.
 - Définition des exigences en termes de type de Reporting et interfaces d'analyse.
- **La conception et la réalisation d'un Data Warehouse :**
 - Représentation multidimensionnelle des besoins déjà identifiés.
 - L'implémentation physique de schéma multidimensionnelle dans un SGBD.
- **La conception et la mise en œuvre de l'ETL, qui a pour but d'assurer l'alimentation régulière du Data Warehouse, à partir des systèmes de production qui sont :**
 - **Demographic Data** qui gère les informations personnelles des clients,
 - **DWS** qui gère les informations sur le trafic (appelles, SMS et MMS)
 - Et d'autres systèmes de gestion des abonnés post payé,
- **Le déploiement d'une application décisionnelle (de restitution) :**
 - Permettant aux gestionnaires du système de gérer les utilisateurs et d'attribuer les droits d'accès.
 - Il offre aux utilisateurs finaux du système (Décideurs) multiples interfaces d'analyses
 - ✓ Analyse multidimensionnelle.
 - ✓ Tableau de bord
 - ✓ Graphes (histogramme, Secteur, Courbe ...)
 - ✓ Reporting
 - ✓ Géo-analyse
 - ✓ Un outil de data Mining permettant la segmentation des abonnées et la prédiction des segments sensibles aux churn (résiliation et migration vers les concurrents).

4. Organisation du mémoire :

Ce mémoire est organisé en 2 parties, une partie théorique dont les notions et concepts théoriques du domaine des SID sont présentés et une partie pratique comportant la conception et la mise en œuvre du système pour résoudre la problématique.

Ce mémoire est présenté en 6 chapitres, 3 en première partie et les autres en deuxième, plus une introduction, une conclusion et des perspectives :

✎ Chapitre 1 (Présentation du système décisionnel) : c'est une introduction au domaine des systèmes décisionnels.

✎ Chapitre 2 (Architecture du système décisionnel) : présente l'architecture en couche d'un SID et les différents composants qui le compose.

✎ Chapitre 3 (Méthodes de conception d'un SD) : dresse un état de l'art des méthodes de conception d'un SID et donne les fondements de la modélisation multidimensionnelle.

✎ Chapitre 4 (présentation de l'organisme d'accueil) : définit brièvement l'organisme où notre stage pratique a été fait, en focalisant sur le champ d'étude (département marketing).

✎ Chapitre 5 (conception du système) : présente la conception des divers composants du système (Sources, DW, ETL ...) aux différents niveaux d'abstractions (conceptuel, logique...).

✎ Chapitre 6 (implémentation et mise en œuvre du système) : c'est le fruit de ce projet, décrit les choix techniques et architecturaux plus la présentation du système développé.

PARTIE I: PARTIE THEORIQUE

Introduction :

Dans cette partie, on va présenter les systèmes décisionnels de manière générale par la définition des concepts de ce domaine avec une brève historique dans le chapitre I, on va montrer par la suite dans le chapitre II l'architecture générale d'un système décisionnel avec la présentation de chacun de ses composants et son rôle. On va présenter aussi, dans le chapitre III, les différentes méthodes proposées pour la conception d'un système décisionnel.

CHAPITRE I.

Présentation du système décisionnel

CHAPITRE I. Présentation du système décisionnel

I.1. Introduction :

Ce chapitre est une introduction au domaine de l'informatique décisionnelle (Les systèmes décisionnels). on donnera dans ce chapitre une brève historique au SID, définitions des concepts de domaine et l'emplacement de système décisionnel au sein de système d'information générale d'une entreprise.

I.2. Brève Historique sur les systèmes décisionnels: (1)

Les primitives de l'informatique décisionnelle se sont développées dans les années 70. Historiquement, les grandes firmes ont été les premières à comprendre la valeur ajoutée des outils d'aide à la décision. En effet, disposant de quantités considérables d'informations dans leurs bases de données opérationnelles, elles ont, en premier lieu, commencé par les interroger directement, par des requêtes. Cette solution de remplacement a vite montré ses limites aussi bien en temps qu'en ressources humaines et matérielles. De nouveaux outils apparaissent alors. Ces outils ont connu l'évolution suivante :

- + **L'infocentre** : l'infocentre est une copie à l'identique des bases de données dans un nouvel environnement (années 70 et 80).
- + **L'EIS : Executive Information System** proposant les premiers tableaux de bord dans les années 1990.
- + **Les entrepôts de données (ED)**: qui sont considérés comme étant le lieu de stockage des gros volumes de données devant être analysés.
- + **Les bases de données multidimensionnelles (OLAP)** : une base où chaque indicateur est analysé en fonction de plusieurs critères ou dimensions.
- + et enfin **la Business Intelligence**: qui regroupe les fonctions d'analyse des données et de reporting.

Outre l'évolution des architectures, l'analyse des données dans ces systèmes d'information décisionnels a également connu une évolution. L'informatique décisionnelle est alors essentiellement constituée d'outils d'édition de rapports, de statistiques, de simulation et d'optimisation.

Conception et Réalisation d'un Système décisionnel pour le marketing

Provenant des recherches en Intelligence Artificielle, les systèmes experts voient le jour. Ils sont conçus par extraction de la connaissance d'un ou plusieurs experts et sont des systèmes à base de règles. De bons résultats sont obtenus pour certains domaines d'application tels que la médecine, la géologie, la finance, etc.

Selon (<http://www.grappa.univ-lille3.fr/polys/fouille/sortie002.html>), dans les années 90, deux phénomènes se produisent simultanément. Premièrement, la possibilité de conception d'environnements spécialisés pour l'aide à la décision. Deuxièmement, de nombreux algorithmes permettant d'extraire des informations à partir de données brutes arrivent à maturité. Ces algorithmes ont des origines diverses et souvent multiples. Certains sont issus des statistiques ; d'autres proviennent des recherches en Intelligence Artificielle, recherches qui se sont concentrées sur des projets moins ambitieux, plus ciblés; certains s'inspirent de phénomènes biologiques ou de la théorie de l'évolution. Tous ces algorithmes sont regroupés dans des logiciels de fouille de données ou *Data Mining* qui permettent la recherche d'informations nouvelles ou cachées à partir de données.

I.3. Définitions :

On appelle « Système d'Aide à la Décision (SDD) », ou bien « Système Décisionnel (SD) », ou encore « Business Intelligence (BI) », un ensemble de solutions informatiques permettant l'analyse des données de l'entreprise, afin d'en dégager les informations qualitatives nouvelles qui vont fonder décisions, qu'elles soient tactiques ou stratégiques (2).

D'autres donnent une définition plus précise et plus complète, par exemple l'auteur dans (3) a le définit comme suit :

Le SID est un ensemble de données organisées de façon spécifique, facilement accessibles 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.

Un système décisionnel est dédié au management de l'entreprise pour l'aider au pilotage de l'activité. Il constitue une synthèse d'informations opérationnelles, internes ou

Conception et Réalisation d'un Système décisionnel pour le marketing

externes, choisies pour leur pertinence et leur transversalité fonctionnelles, et sont basés sur des structures particulières de stockage volumineux (Data Warehouses (**DW**), bases **OLAP**). Le principal intérêt d'un système décisionnel est d'offrir au décideur une vision transversale de l'entreprise intégrant toutes ses dimensions.

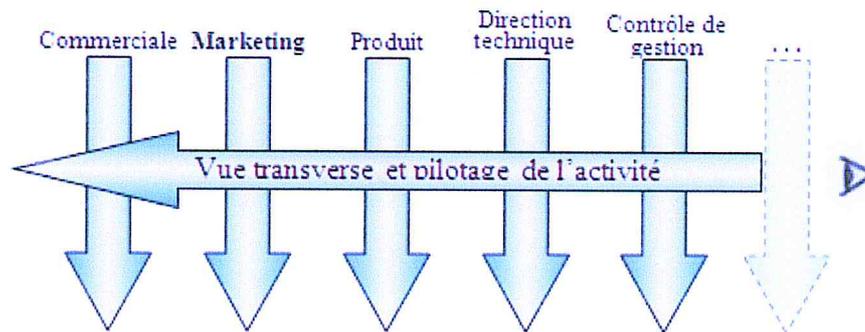


Figure 1.1. Une vision transversale (3)

Cette vue intégrée peut alors être étudiée par fonction ou par métier. Le pilotage est la contribution des événements à un objectif stratégique.

I.4. Positionnement du décisionnel au sein du système d'information :

D'un point de vue architectural, nous considérerons que nous pénétrons dans le monde du décisionnel dès lors que les données de production sont valorisées en informations. Cette valorisation est effective dès que l'on sort du monde de la production. Sur le schéma suivant, décrivant l'architecture fonctionnelle d'une entreprise, on voit la place prise par le décisionnel au sein d'un système d'information.

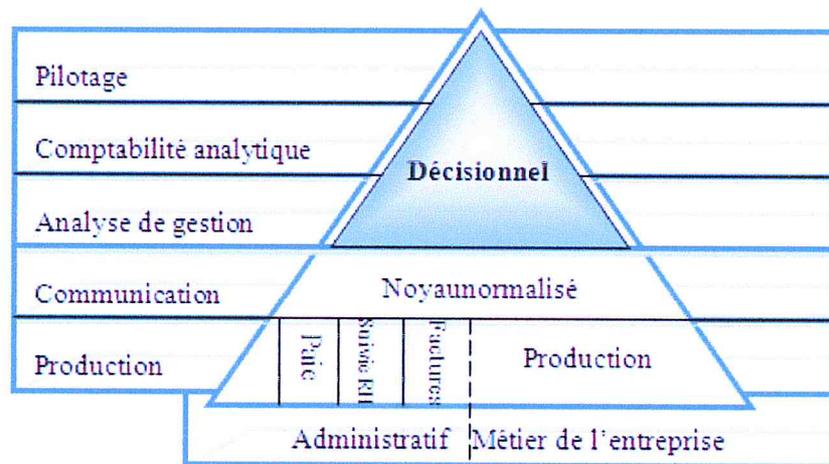


Figure 1.2. Le décisionnel au sein du système d'information

I.5. Le Data Warehouse :

Linguistiquement, Un **Data Warehouse (DW)** est un **entrepôt de données (ED)**. Il s'agit d'un stockage intermédiaire des données issues des applications de production, dans lesquelles les utilisateurs finaux puisent avec des outils de restitution et d'analyse (3).

La définition a été énoncée par Bill Inmon (4) (qui est considéré comme le père du concept) : « *Un data warehouse est une collection de données thématiques, intégrées, non volatiles et historiées organisées pour la prise de décisions* ».

Un data warehouse est une collection de données :

- + **Thématiques (orientées sujet)** : l'objectif d'un DW est la prise de décisions autour des activités majeures de l'entreprise.
- + **Intégrées** : il est clair que l'assurance de la transversalité recherchée nécessite l'intégration des données internes (qui se trouvent dans les systèmes d'information opérationnelles) et d'autres externes (internet, sondages...) qui sont généralement hétérogènes.
- + **Non volatiles (pas de suppression)** : afin de conserver la traçabilité des informations et des décisions prises, les informations stockées au sein du data warehouse ne peuvent être supprimées.

Conception et Réalisation d'un Système décisionnel pour le marketing

- + **Historisées** : outre les problèmes de volumétries, de capacité de stockage et de calcul des machines hébergeant le data warehouse, ce qui nécessite une historisation régulière des informations stockées;
- + **Organisées pour la prise de décisions.**

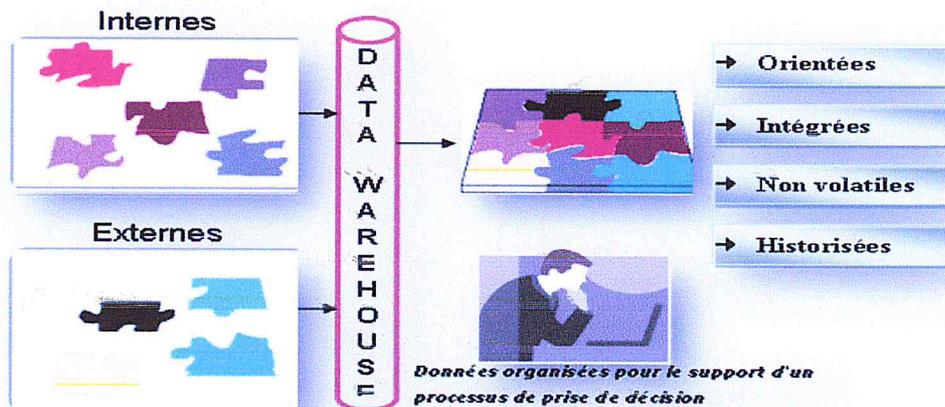


Figure 1.3. Définition du data warehouse (3)

D'autre part, Ralph Kimball (5) a défini un DW comme « une copie des données transactionnelles structurées d'une manière spécifique pour l'interrogation et l'analyse ».

De plus, les experts des systèmes d'aides à la décision (SDD) expriment que « le DW est une base de données (BD) optimisée pour la prise de décision » (6).

I.6. Conclusion :

On a présenté, dans ce chapitre, le domaine de l'informatique décisionnelle et son historique. On a vu un ensemble de définition plus l'emplacement d'un SID sur le plan hiérarchique d'une entreprise.

CHAPITRE III.

Composants du système décisionnel

CHAPITRE II. Composants du système décisionnel

II.1. Introduction :

Ce chapitre montre l'architecture générale du système d'information décisionnel avec la présentation de chaque composant et son rôle dans le système.

II.2. Architecture générale :

L'architecture d'un SID met en jeu 5 éléments essentiels : les sources de données, l'entrepôt de données, les magasins de données et les outils d'analyse et d'interrogation, plus la zone d'alimentation connue sous le nom (Staging Area).

Le SID se compose de plusieurs composants, chaque phase d'implémentation du système nécessite le développement d'un ensemble d'outils (7).

On parle généralement d'architecture n-tiers (en couche) en raison des différentes couches possibles pour gérer les données (source, entrepôt, magasin...) (8).

Le schéma suivant illustre l'architecture générique d'un SID:

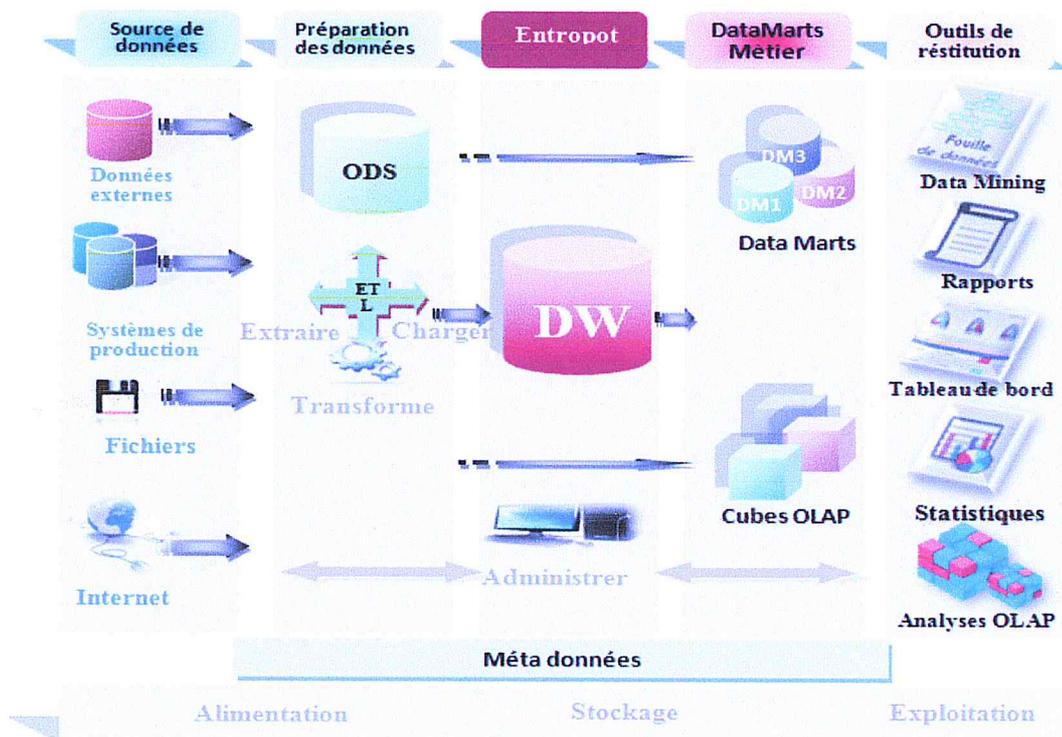


Figure 2.1. Architecture générale d'un système décisionnel.

II.3. Systèmes Sources :

On appelle, d'une façon générale, bases de production toutes les sources (qu'il s'agisse de données de production, d'informations internes ou d'informations externes quel que soit leur mode de stockage « BD, Fichiers plat ou XLM ...») dont il va falloir extraire des données pour alimenter le DW.

On y retrouvera notamment, les données commerciales, les données du système comptable, les données sur le personnel, les informations sur la concurrence, des informations externes produites par divers organismes nationaux ou internationaux de statistiques et autres fournie par des services web (3).

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. On part du principe que les systèmes source maintiennent peu de données historiques 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 manière extensive et inattendue, ce qui est souvent le cas des entrepôts de données(4).

II.4. Zone de préparation des données:

La zone de préparation des données (ou Staging Area) est un ensemble des processus qui nettoient, transforment, combinent, archivent, suppriment les doublons, c'est-à-dire prépare les données sources en vue de leur intégration puis de leur exploitation au sein du data warehouse, Cette zone comprend tout ce qui se trouve entre les systèmes source et le serveur de présentation (5).

A ce niveau, on trouve un outil d'alimentation (Serveur ETL). Est un ensemble de composants logiciels capable d'extraire des données généralement hétérogènes au niveau de des sources, de les transformer en vertu de règles très précises, puis de les injecter au sein du système décisionnel cible (3).

Cette phase est considérée comme le facteur clé de la réussite d'un projet décisionnel ou un projet du DW. De fait, que le développement de L'ETL prendre 80% de temps du développement d'un DW (9).

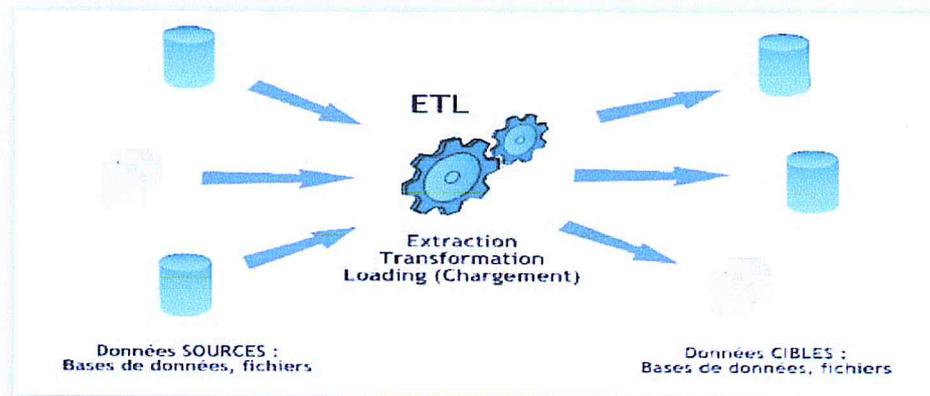


Figure 2.2. Le système d'alimentation (ETL)

II.4.2. Extraction des données :

La première phase de la construction d'un data warehouse consiste à extraire les données utiles des systèmes opérationnels, qui sont dans de nombreux cas hétérogènes (avec plusieurs gestionnaires de bases de données et plusieurs méthodes d'accès), diffuses (avec plusieurs environnements matériels et plusieurs réseaux interconnectés ou non), et complexes (avec plusieurs modèles logiques et physiques bien souvent orientés prioritairement vers les traitements transactionnels). (5)

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.

II.4.3. Transformation :

La deuxième phase du processus ETL consiste à transformer les données extraites de manière à obtenir un ensemble homogène de données, qui deviennent ainsi comparables, additionnables, etc. (3)

Les données peuvent être transformées de nombreuses manières, par exemple nettoyées (correction orthographique, résolution de conflits de domaine, traitement du problème des éléments manquants, conversion à des formats standard), combinées à partir de sources multiples, dédoublées s'il y a lieu et pourvues de clés propres à l'entrepôt de

données. Ces transformations sont le prélude au chargement dans la zone de présentation de l'entrepôt de données. (5)

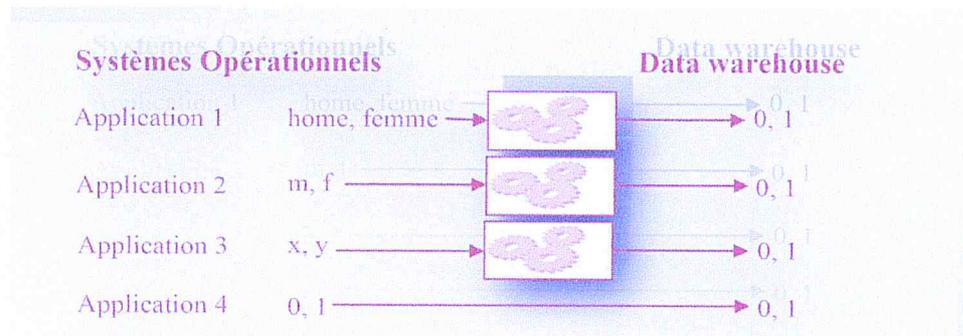


Figure 2.3. Exemple de transformation

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.

Le dé doublonnage : Dès lors que l'on va extraire d'un ensemble de bases des données en vue de les agréger, on va être confronté au problème des redondances. On trouvera par exemple la référence client tant au sein de la base commerciale qu'au sein de la base marketing. On retrouvera le produit tant au niveau des bases de production, qu'au niveau commercial, qu'au niveau comptable qu'au niveau marketing, etc. Ces redondances figureront au sein d'une même source ou au sein des diverses sources. Le dé doublonnage interviendra sur chacun des niveaux requis pour garantir la cohérence et donc la qualité du data warehouse. (3)

Le formatage et la restructuration : Vise à normaliser les informations et à préparer leur injection dans le système cible en les convertissant au format cible. Cette normalisation ou dénormalisation suivant les cas, s'applique fréquemment sur des codes (tel que le code INSEE dont l'ensemble des digits pris séparément ont une signification particulière) ou sur des adresses. Le cas des adresses est typique car on découpera par exemple le champ source « adresse » en n champs cibles : « numéro », « rue », « nom rue », « particularités », « code postal », « ville », « pays ».

Création de 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 référent entre les tables dimensionnelles et les tables des faits.

II.4.4. Alimentation et indexation :

À la fin du processus de transformation, les données prennent la forme d'images d'enregistrement prêtes à être chargées. En règle générale, le processus de chargement des données dans l'entrepôt s'effectue par réplication des *tables des faits* et des *tables dimensionnelles*, qui seront ensuite présentées aux services de chargement en masse de chaque data mart destinataire.

Le chargement en masse est une fonctionnalité très importante à mettre en opposition avec le chargement ligne à ligne, qui est un processus nettement plus lent. Le data mart de destination doit alors indexer les données afin d'optimiser les performances de requête, si ce n'est déjà fait.

II.4.5. Générations des outils ETL : (10)

+ Première génération :

A l'époque des mainframes en tant que cœur des systèmes d'informations, les outils ETL généraient du code Cobol à partir des données extraites. Coûteuses, ces solutions étaient des plus encombrantes et nécessitaient parfois une forte intervention humaine lors de l'extraction. Leur mise en œuvre était extrêmement consommatrice en temps et en ressources. De nombreux développeurs étaient indispensables pour concevoir l'application mais aussi pour la faire fonctionner et la maintenir. De tels outils sont toujours en production notamment dans le milieu bancaire.

+ Deuxième génération :

Avec l'avènement du Client/serveur dans le cadre d'applications propriétaires et sur mesure, le langage de requêtes SQL a remplacé le Cobol comme format de destination. Mais ces outils ont rencontré leurs limites avec l'avènement d'applications packagées comme les progiciels de gestion intégré (SAP, PeopleSoft, ...). Pour coller du

mieux possible au modèle des données stockées dans la base relationnelle, il fallait aussi extraire le logique métier au niveau de la couche applicative.

+ **Troisième génération :**

Comme chacune des applications citées gère différemment ce logique métier par-dessus des modèles de données spécifiques, les outils ETL ont dû être livrés avec des adaptateurs spécifiques. Leur rôle : fournir une intégration plus étroite avec ces applications. Déjà apparues au cours de la deuxième génération, les interfaces permettant de modéliser graphiquement les échanges et les changements sous-jacents se sont imposées, tout comme la console d'administration centralisée. Il fallait alors tenir compte de la nouvelle complexité engendrée par la transition vers les systèmes d'informations hétérogènes, tout en accueillant les technologies web dans l'entreprise. Ainsi c'est à partir de cette génération que l'on retrouve des interfaces web pour l'administration et le design des processus ETL.

+ **Quatrième génération :**

Avec la nécessité d'être plus réactif dans un environnement concurrentiel très dynamique, les processus de décision se raccourcissent dans le temps. C'est pourquoi ces outils ETL sont désormais capables de gérer simultanément des flux de données en temps réel et en différé. Pour introduire cette dose de temps réel, une série de composants veillent en attendant l'arrivée des requêtes pour les traiter. Dans le même temps, l'ETL se dote d'un distributeur de messages, et empiète déjà un peu sur l'EAI (Enterprise Application Intégration). Dans ce cadre, XML (eXtensible Markup Language) joue aussi un rôle clef grâce notamment au recours à sa technologie sœur XSLT qui permet d'effectuer les transformations à la volée.

+ **Cinquième génération :**

Les précédentes générations d'outils se destinaient essentiellement à une exploitation des données dans un cadre décisionnel avec un entrepôt de données centralisé, localisé (Datamart) ou généraliste (Data Warehouse). La cinquième génération, de son côté, présente une vocation plus large. Les adaptateurs, API (Application Programming Interface) ou autres, sont bi-directionnels : ils sont capables à la fois de lire et d'écrire tout

autant dans la source que dans la cible. Et le "message broker" avec des fonctions de file d'attente est désormais devenu indispensable. Les données exploitées à des fins décisionnelles analytiques restent mises à jour en différé, du fait de leur volume important. Mais celles à caractère opérationnel, dont les applications ont besoin pour rendre compte de la situation présente de l'activité, sont injectées en temps réel ou quasi-temps réel dans la cible.

II.4.6. Types d'ETL : (10)

Nous distinguons généralement trois types ou architectures d'ETL :

+ ETL Moteur :

L'ETL moteur est l'ETL par excellence. Il est monté sur un Serveur du milieu (Serveur ETL). Ainsi toutes les transformations sont traitées par lui. Généralement, la persistance des métadonnées est gérée par un SGBDR (C'est le référentiel). Comme exemple d'ETL de cette famille nous pouvons citer entre autres :

- ✓ *PowerCenter d'Informatica*
- ✓ *DataSage d'IBM Ascential*
- ✓ *Genio de Hummingbird (racheté l'été 2006 par open text)*
- ✓ *Data Integrator de Business Objects*
- ✓ *Ab Initio Software de Ab Initio*
- ✓ *Open source :Kettle, KETL,...*

Avantages :

- Possibilités d'effectuer des opérations multibases. Par exemple, une jointure entre une table Sybase et une autre table Oracle. Notons quand même que dans bon nombre de projets DataWareHouse, on préfère résoudre le problème de « jointure multibase » dans un ODS.
- Gain significatif au niveau de la productivité.

Limites

- Le coût : Ce dernier est souvent fonction du nombre de connecteurs et de machines « moteurs ».
- Le côté « boîte noire » du moteur : En effet les Transformations faites par le moteur ne sont pas accessibles. Donc la seule optimisation possible est celle fournie par le moteur lui-même

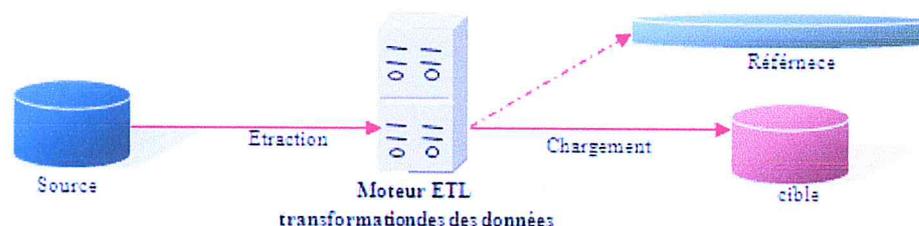


Figure 2.4. ETL Moteur

+ ETL « Base » ou « Générateur de Code » :

Dans ce type d'ETL les transformations sont quasiment toutes déportées dans les SGBDR. D'où l'appellation - DataBase-Embedded - souvent donnée aux ETL de cette famille. L'ETL se charge juste de générer le code SQL idoine. Donc il n'y a pas besoin d'une machine devant héberger l'ETL, mais juste une machine supportant un ordinateur dont le rôle est de vérifier les relations d'ordre entre processus SQL. Généralement la persistance des met a données est gérée par SGBDR (C'est le référentiel). Comme exemple d'ETL de cette famille nous pouvons citer entre autres :

- ✓ WarehouseBuilder d'Oracle Corporation
- ✓ Sunopsis Data Conductor de sunopsis (Racheté fin 2006 par Oracle)
- ✓ DB2 Warehouse Manager d'IBM
- ✓ Open source : Talend

Avantages

- Le côté - boîte blanche- du moteur : En effet, les transformations générées par le moteur sont accessibles (SQL, PL/SQL, T-SQL, ...) et à fortiori optimisable.
- Le prix : Les ETL Base sont les moins chers du marché.

Limite

Ici on suppose que toutes les opérations sont transformables et optimisables par un SGBDR. Nous sommes de ceux qui pensent qu'il existe des opérations qui de par leur nature sont moins performantes si elles sont exécutées sur un SGBDR

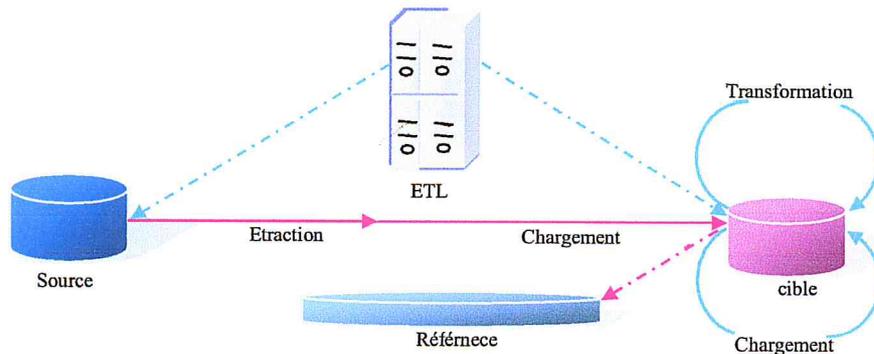


Figure 2.5. ETL Base (Générateur de code)

+ ETL Moteur et Base :

L'ETL moteur et base que nous appelons - ETTL » (ExtractTransform and Transform and Load). Les deux - T » se justifiant par la double transformation. En effet, ici les transformations peuvent être faites dans le moteur ETL ou dans les SGBOR. Généralement, la persistance des métadonnées est gérée par SGBDR (C'est le référentiel). Comme exemple d'ETL de cette famille nous pouvons citer entre autres :

- ✓ *PowerCenter d'Informatica*
- ✓ *DataSatage d'IBM Ascential*
- ✓ *Genio de Hummingbird (racheté l'été 2006 par open text)*

De manière générale, les ETL Moteur convergent aujourd'hui (l'année de rédaction du rapport est 2007) presque tous vers des ETTL

Avantages

Possibilités de répartir des traitements. Tout n'est pas possible avec le SQL. Par exemple ramener la nième ligne d'une requête est une opération dont la difficulté est fonction du SGBDR. Or elle se retrouve faciliter avec un moteur ETL (puisqu'il suffit de boucler sur les lignes résultats pour ramener la ligne recherchée)

Limites

- Le coût : Ce dernier est souvent fonction du nombre de connecteurs et de machines « moteurs ».
- Le côté « boîte noire » du moteur : En effet les Transformations faites par le moteur ne sont pas accessibles. Donc la seule optimisation possible est celle fournie par le moteur lui-même.

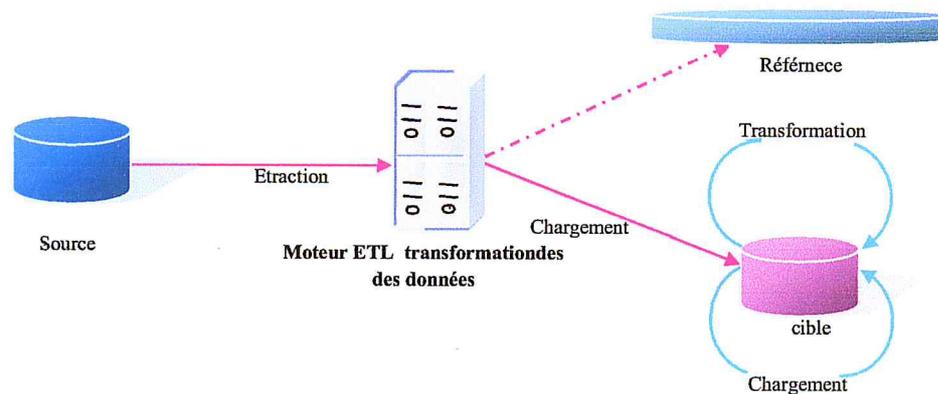


Figure 2.6. ETL Base et Moteur

II.5. Serveur de présentation de data warehouse :

La zone de présentation des données est le lieu où les données sont organisées, stockées et offertes aux requêtes directes des utilisateurs, aux programmes de reporting et autres applications d'analyse. Elle est tout ce que les utilisateurs voient et touchent par le biais des outils d'accès. Sur le serveur de présentation, les données sont présentées et stockées sous une forme *dimensionnelle* :

- ✓ S'il repose sur une base de données relationnelle, les tables y seront organisées sous forme de schémas en étoile ou ses extensions.
- ✓ S'il est basé sur la technologie OLAP (Online Analytic Processing), les données conserveront des dimensions reconnaissables.

La majeure partie des gros data marts sont implémentés sur des bases de données relationnelles. Ce qui explique pourquoi, lorsque nous évoquons les serveurs de présentation, nous partons du principe de la présence de bases de données relationnelles sous-jacentes. (5)

II.5.1. Base de données du Data Warehouse:

Source de données interrogeable de l'entreprise, C'est l'union des data marts qui le composent. L'entrepôt de données est alimenté par la zone de préparation des données (7).

La base de données est le constituant principal du data warehouse puisque c'est dans celle-ci que l'on va stocker les informations extraites des bases de production. C'est au sein du SGBD qu'est stocké le dictionnaire du data warehouse où sont stockées les métadonnées, c'est-à-dire « les données sur les données stockées dans le SGBD » décrivant la manière dont sont constituées les informations stockées.

Le data warehouse est supporté par une base de données relationnelle, multidimensionnelle ou objet, même si celles-ci sont assez rares ou utilisées dans des contextes assez particuliers. La base de données relationnelle présente une organisation de l'information s'appuyant sur des relations existant entre des tables. Ainsi, si la table des clients est en relation avec la table des factures de l'entreprise, alors on pourra établir la liste des factures par client. Une base de données multidimensionnelle est une base dont les données sont stockées de manière à optimiser l'accès aux informations suivant des requêtes non prévues à la création de la base (3).

II.5.2. Data mart :

Un data mart (Magasin de données) est un sous ensemble de l'entrepôt de données orienté sur un sujet particulier (métier précis). Chacun 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. (7)

II.5.3. OLAP (Online Analytic Processing) :

La finalité d'un data warehouse est d'obtenir des vues multidimensionnelles. Ces vues sont représentées sous la forme d'un cube en trois dimensions sachant qu'une base multidimensionnelle peut comporter de nombreuses dimensions (un Hyper Cube). Les systèmes OLAP mettent en œuvre des technologies permettant de rassembler, gérer, traiter et présenter des données multidimensionnelles à des fins d'analyse et de décision.

Un outil OLAP est capable de fournir une information multidimensionnelle partagée pour l'analyse rapide (3).

En 1993, Codd, fondateur des bases de données relationnelles, définit le concept OLAP (On Line Analytical Processing) dans un document technique sous le titre de « **Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate** » comme suit « ... le nom donné à l'analyse dynamique requise pour créer, manipuler, animer et synthétiser l'information, par des modèles d'analyse de données exécutives, contemplatifs et selon des formules » (11).

Dans cet article l'auteur dresse les fameux 12 règles qui gouvernent la technologie OLAP, 6 autres sont ajoutées plus tard en 1995 toujours par le Docteur Edgar F (12).

II.5.3.1. Les (12 + 6) règles d'OLAP du CODD (12):

+ **Caractéristiques Basiques :**

- ✓ *Modèle Conceptuel Multidimensionnel* (Originale règle 1): le modèle OLAP est par nature multidimensionnel, Il se doit de faciliter le forage de données.
- ✓ *Manipulation Intuitive des données* (Originale règle 10): la lecture et la modification des données doivent rester simples. Ici, cette règle n'évoque pas directement les feuilles de calculs mais le rappelle fortement.
- ✓ *Accessibilité* (Originale règle 3): OLAP comme un médiateur : Codd imagine l'OLAP comme un tout capable de se connecter à des sources hétérogènes de données avec un accès utilisateur unique,
- ✓ *Extraction vs Interprétive* (Nouvelle)
- ✓ *Modèle analytique OLAP*(Nouvelle) : pouvoir entre autre paramétrer un rapport, classer, hiérarchiser les données. Codd pensait certainement déjà au datamining.
- ✓ *Architecture client-serveur* (Originale règle 5): avoir la possibilité que plusieurs types de clients puissent se connecter à un serveur de données OLAP.
- ✓ *Transparence* (Originale règle 2): les données stockées sur un serveur OLAP doivent être simples d'accès jusqu'à offrir une interface vers les feuilles de calcul.
- ✓ *Multiutilisateurs* (Originale règle 8): le système doit être capable de répondre à des demandes concurrentes.

+ **Caractéristiques Spéciales :**

✓ *Traitement de données non-normalisées* (Nouvelle) : référence aux sources hétérogènes de données, par exemple les environnements dénormalisés, auxquels peut se connecter un serveur OLAP. Codd imagine l'impossibilité de mettre à jour les données. Cette règle semble de moins en moins respecter. Les éditeurs offrent cette possibilité dans le cadre de la simulation : changer la valeur d'une cellule de son analyse croisée pour en déduire l'impact sur les totaux et la répartition des chiffres.

✓ *Stocker les résultats des analyses OLAP* (nouvelle): les séparer des données sources : ici Codd se penche plus sur la méthode que sur le serveur OLAP lui-même. Il ne conçoit pas la possibilité d'ériger son système OLAP directement au sein de son système opérationnel (transactionnel). De nombreux auteurs ont longtemps défendu ce point de vue. Aujourd'hui les performances machine pourraient transformer cette vision. Sans compter que la nouvelle version Oracle 11 arrive avec « Active Data Guard » qui maintient à jour au sein d'une même instance de base données une réplique qui peut être accédée uniquement en lecture.

✓ *Extraire les valeurs manquantes* (Nouvelle) : toutes les valeurs absentes sont remplacées de façon uniforme. Il s'agit de distinguer les valeurs manquantes de celle qui valent 0.

✓ *Traiter les valeurs manquantes* (Nouvelle) : à rapprocher de la règle précédente. Les valeurs manquantes ne sont généralement pas prises en compte par les analyseurs OLAP : pas d'intersections entre l'ensemble des dimensions !

+ **Caractéristiques de Reporting :**

✓ *Reporting flexible* (Originale règle 11): l'utilisateur peut disposer comme il le souhaite des dimensions. La plupart des produits du marché offrent une manipulation conviviale des dimensions aux utilisateurs.

✓ *Performance uniforme des rapports* (Originale règle 4): le nombre de dimensions, le volume de données manipulées ne doit pas affecter la constance des performances de restitution des résultats. Pour maintenir cette uniformité de la performance, il faut préférer travailler directement avec un niveau agrégé de l'information dans certains cas.

Conception et Réalisation d'un Système décisionnel pour le marketing

✓ *Ajustement Physique Automatique* (Originale règle 7): le serveur OLAP doit être capable d'auto-ajuster l'espace disque qui lui est nécessaire. Peu d'éditeurs ont mis en place cette règle. A défaut d'être automatique de nombreuses bases de données possèdent le « partitionnement » ; il facilite la gestion d'ajout ou de suppression d'unités de stockage.

✚ **Contrôle des dimensions :**

✓ *Dimension générique* (Originale règle 6): une dimension doit avoir la même signification selon les différents contextes auxquels il peut être associé. *Ralph Kimball* l'un des pères du data warehousing parle lui de « dimension conforme ».

✓ *Nombre de dimensions et de niveaux hiérarchiques illimités* (Originale règle 12): techniquement aucun produit ne peut répondre à cette règle. *Ralph Kimball* pense quant à lui que le nombre maximum de dimensions qui décrivent un fait ne doit pas excéder 15. Si le nombre de dimensions et de hiérarchies sont trop important le modèle risque d'être incompréhensible à l'utilisateur.

✓ *Opérations sur les analyses croisées non-restreintes* (Originale règle 9): en OLAP les opérations coïncident d'abord avec des fonctionnalités d'agrégation (somme, comptage, ...). Cependant tous types d'opérations doivent être possibles, même des calculs complexes.

Etant donné que les règles OLAP telles qu'elles sont définies par Codd ont été controversées. D'ailleurs, Nigel Pendse, l'initiateur de l'Olap Report (<http://www.olapreport.com>), évoque que ce terme OLAP n'est pas explicite, il ne fournit pas une définition et ne permet pas de savoir si un outil relève ou non de cette technologie. Par ailleurs, 12 règles ou 18 dispositifs, c'est selon lui, une quantité trop importante pour être facilement retenue. Nigel Pendse suggère alors d'assimiler le terme OLAP à une définition comprenant cinq termes : **Fast Analysis of Shared Multidimensional Information** (le model **FASMI**). Cette définition correspond finalement aux critères retenus pour simplifier les règles de Codd et faciliter l'évaluation des outils OLAP. Elle a été traduite en français par « Analyse Rapide d'Information Multidimensionnelle Partagée ». (8)

II.5.3.2. Les technologies OLAP: (10)

Le marché de l'OLAP se divise en trois grands courants qui se différencient quant à la façon dont sont sauveés physiquement les dimensions.

+ **ROLAP (Relational OLAP) :**

Les dimensions sont stockées au sein d'une base de données relationnelle. Par contre le schéma logique reste multidimensionnel.

Avec cette technologie, les données multidimensionnelles sont stockées dans un SGBDR. Le modèle multidimensionnel est alors traduit de la manière suivante :

- ✓ Chaque fait correspond à une table appelée table de fait.
- ✓ Chaque dimension correspond à une table appelée table de dimension.

Ainsi, la table de fait est constituée d'attributs représentant les mesures d'activité et les attributs clés étrangères de chacune des tables de dimension (ou axe d'analyse). Les tables de dimension contiennent les paramètres et une clé primaire permettant de réaliser des jointures avec la table de fait.

Avantages

- Souplesse (Courbe d'apprentissage faible,...).
- Accès - universel via SQL
- Supporte de très gros volumes.

Limites

- Peu efficace pour les calculs complexes.
- Temps de réponse parfois important.
- Forte consommation en ressources humaines (DBA).

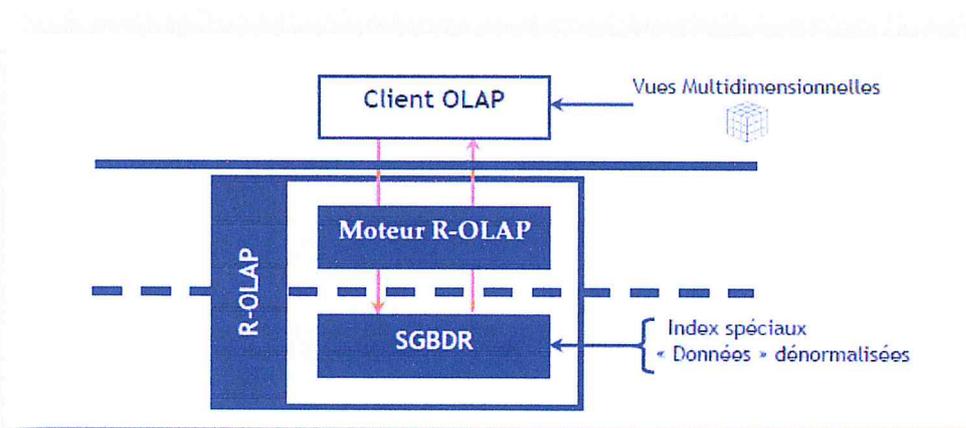


Figure 2.7. La technologie ROLAP (10)

+ MOLAP (Multidimensional OLAP) :

Les données sont physiquement stockées sous forme de dimensions. Avec cette technologie, les données multidimensionnelles sont stockées dans un SGBDM. Les structures multidimensionnelles sont optimisées pour le stockage et le traitement des données. Ces structures natives sont des tableaux à plusieurs dimensions. Dans la littérature, les termes de « cube », « hypercube » et « table multidimensionnelle » sont utilisés de manière interchangeable. Il s'agit ici d'un système multidimensionnel pur.

Cette approche permet de stocker l'ensemble des données de manière multidimensionnelle. L'intérêt est que les temps d'accès sont optimisés.

Avantages

- Rapide (meilleur temps de réponse,...).
- Optimisé pour les calculs complexes.
- Administration simplifiée.

Limites

- Délais parfois longs au niveau de la préparation
- Format généralement propriétaire
- Solution coûteuse et dispendieuse
- Ne supporte pas de très gros volumes de données

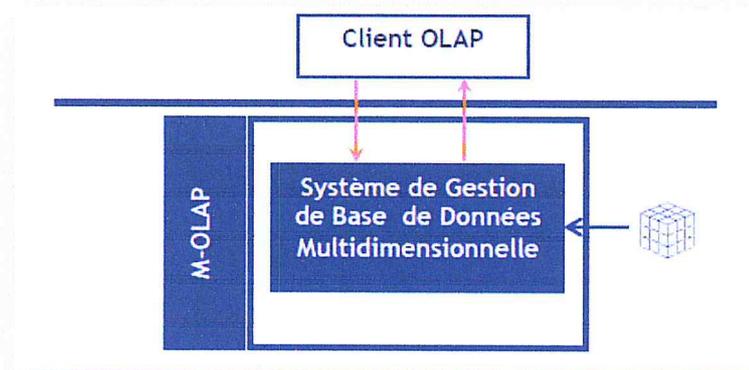


Figure 2.8. La technologie MOLAP (10)

✚ **HOLAP (Hybrid OLAP) :**

Le moteur d'analyse embarque les deux approches. Cette technologie essaye de combiner les avantages des technologies R-OLAP et M-OLAP (l'abréviation - MR-OLAP - est parfois utilisée en lieu et place de - H-OLAP -). Les données agrégées sont stockées en M-OLAP sur un SGBDM alors que les données détaillées sont quant à elles stockées en R-OLAP dans un SGBDR.

Le reproche majeur que nous pouvons faire à cette technologie est qu'elle est complexe et nécessite une plus grande expertise pour son utilisation.

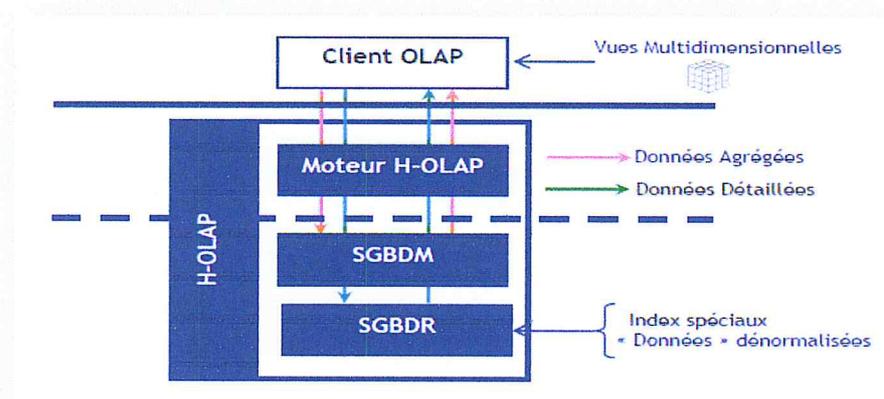


Figure 2.9. La technologie HOLAP (10)

II.6. Outils de restitution :

II.6.1. Introduction aux outils de restitution :

Ce sont les outils et logiciels que les utilisateurs auront pour manipuler les données et profiter des services offerts par l'entrepôt de données et le serveur d'analyse. Ces outils doivent impérativement être aussi ouverts que la philosophie et les principes du BI (Simplicité, Souplesse, esprit analyse, ergonomie ...). Les avancées technologiques permettent maintenant d'avoir des outils très simples et très puissants à la portée de tout le monde.. (13)

Les outils de restitution sont la face visible de l'iceberg de l'informatique décisionnelle. Mais ils ne peuvent éditer des données qui n'ont pas été correctement extraites et analysées auparavant. La valeur ajoutée d'un système décisionnel est bien plus concentrée dans la constitution de la base multidimensionnelle, que dans ces outils de présentation.

Du point de vue de l'utilisateur, ils n'en restent pas moins indispensables, car ils mettent en lumière les informations et connaissances que les étapes précédentes du processus ont contribuées à créer. (14)

Ces outils peuvent être résumés en 3 termes « **Restituer, Analyser et Piloter** ». Ces trois notions correspondent à des usages qui se veulent différents de par la nature des acteurs et de leur responsabilité dans le dispositif d'entreprise (15). Nous pensons que l'ajout des 3 termes « **Classifier, Associer, Anticiper** » est indispensable, ces termes résument les tâches fondamentaux du Data Mining (qui a donné d'autres perspectives à l'informatique décisionnel).

II.6.2. Restitution (Reporting) :

Le premier besoin, le plus demandé par les utilisateurs, celui de produire des rapports. Il est commun de distinguer trois modes de restitution, matérialisés par des rapports aux nuances suivantes :

Conception et Réalisation d'un Système décisionnel pour le marketing

+ *Rapport statique* : il s'agit d'un rapport dont la structure est figée tant en termes de présentation que du périmètre des données présentées. Il présente l'avantage d'être généré en dehors du temps de travail de l'utilisateur et d'être consommable instantanément.

Par exemple, l'acheteur logistique d'une plateforme de commerce voudra disposer de manière hebdomadaire d'un rapport présentant l'évolution de la qualité de service de ses prestataires logistiques (capacité à délivrer la quantité dans les délais et la qualité attendue) afin de peser sur les négociations futures ou établir les pénalités.

+ *Rapport dynamique* : il s'agit d'un rapport ayant une structure de présentation semi statique au sein duquel le périmètre des données peut varier. Cela en offrant à l'utilisateur final la possibilité de choisir les valeurs de paramètres dynamiques intégrés dans le rapport. Cela présente l'avantage indéniable de disposer d'une grande amplitude sur les rapports finaux puisque à partir d'une même structure de rapport on pourra par exemple aussi bien générer un rapport de comparaison annuelle qu'un rapport de comparaison trimestrielle. La notion de paramètre implique cependant d'interroger la BDD ce qui dégrade l'instantanéité.

Par exemple, un responsable commercial régional travaillant pour une société de service en recrutement voudra consulter un rapport d'alerte de consommation (nb d'entreprises clientes n'ayant pas consommé depuis X mois telles offres) en sélectionnant un ou plusieurs codes d'activité INSEE et tout ou partie des offres de service proposées par sa société.

+ *Rapport Ad hoc* : il s'agit d'un rapport librement aménagé par l'utilisateur final. Il aménage les données qui l'intéressent dans une structure type tableau ou graphique en définissant le périmètre de consultation de ses axes et indicateurs. Cette approche nécessite que soit mis en place une « couche métier » (ou semantic layer) d'accès aux données. Cette couche permet de présenter à l'utilisateur une approche fonctionnelle des données et de leur relation, affranchissant ce dernier de la complexité des BDD et du langage SQL. Concrètement, les données sont regroupées au sein de thématiques métiers et reliées entre elles selon ce même logique métier. L'utilisateur n'a plus qu'à sélectionner ses données et les disposer dans son rapport. La encore, la volonté de l'utilisateur se

traduit en une sollicitation de la base de données avec une performance d'affichage variable selon la complexité de la demande.

Par exemple, un chargé d'études au sein d'une société de services immobiliers voudra composer un rapport mettant en scène un nouvel indice illustrant mieux la tendance du marché des bureaux en IDF...

II.6.3. Analyse multidimensionnelle :

Au-delà de l'activité qu'il faut surveiller à tous les échelons opérationnels (pour laquelle les services de restitution remplissent parfaitement leur rôle), il appartient à certains niveaux de management de comprendre plus profondément les mécanismes influant sur cette activité. Il s'agit par exemple de rechercher les causes d'une inflexion afin de mettre en place les mécanismes managériaux, organisationnels ou décider des investissements en vue de contrôler les facteurs déterminants les évolutions de l'activité.

Pour cela, le manager doit disposer d'une structure plus souple d'interrogation des données. Il doit pouvoir de manière libre naviguer dans les données sans à priori initial plus que la simple découverte. Sa navigation peut se faire de manière croisée (interroger la valeur de son indicateur de satisfaction client réparti selon la géographie, le temps et décliné par produit par exemple) et de manière ascendante / descendante en jouant sur le niveau de granularité des axes métiers (vision niveau mois, Europe, famille de produit puis mois, France, famille de produit en remontant ensuite sur une consolidation annuelle etc....)

Un des intérêts reconnus réside dans la capacité à trouver rapidement l'origine, la cause d'un phénomène. Par exemple, l'origine d'une inflexion des ventes constatée à un niveau agrégé (monde, annuel, gamme de produit) peut être identifiée par une navigation dans les niveaux plus fin de la géographie, des périodes ou des produits, aboutissant à situer une partie importante du problème au niveau d'un produit en particulier, d'un mois en particulier et enfin d'une structure commerciale en particulier.

Cette approche est volontairement théorique et non généralisable. En effet, on trouvera certaines catégories d'utilisateurs opérationnels qui peuvent profiter des services offerts pour effectuer leur travail au quotidien. Ca peut être le cas pour un départ-

tement marketing qui cherchera à comprendre les résultats de sa campagne et qui pour cela trouvera un intérêt certain à manipuler des données disponibles dans un format multidimensionnel. En fait, elle peut être considérée comme un prolongement du reporting ad hoc.

Il est commun d'appeler « moteur OLAP » l'outillage permettant ce service. Un moteur OLAP permet à l'utilisateur d'accéder à un environnement fournissant de manière quasi instantanée la valeur des indicateurs au croisement des axes d'analyses. L'utilisateur sélectionne de manière conviviale les axes, le niveau sur les axes et les indicateurs qu'il veut y décliner. Ce système offre des fonctions natives de navigation dans le détail ou dans la consolidation des valeurs.

II.6.4. Pilotage :

Les cadres dirigeants ont des besoins complémentaires. La surveillance de l'activité, non plus au niveau de son délivré opérationnel, mais dans sa capacité à remplir ses objectifs stratégiques, amène à considérer un outillage spécifique.

Les tableaux de bords du pilote (on entendra les termes associés de KPI, Dashboard, Balanced scorecard) sont des formats agrégés de présentation de contenu. Ils mettent en scène les indicateurs clés et les objectifs stratégiques en les confrontant. Très visuel, le tableau de bord permet de savoir rapidement si l'entreprise est en avance ou en retard sur son plan stratégique.

Ce type d'outils permet d'obtenir la vision transversale qu'on a déjà mentionnée dans le chapitre précédent.

II.6.5. Outils de statistiques et de Data-Mining :

« Saviez-vous que les femmes qui achètent des tabliers à carreaux n'achètent jamais de mobilier de décoration par correspondance ? »

C'est ce type de découvertes que permet de faire un outil de data mining. Pour reprendre une expression maintenant tombée dans le domaine public, l'outil de data mining est celui qui vous aidera à «trouver le diamant caché au fond de la mine de charbon, sans vous salir les mains ». Cette allégorie montre bien la fonction principal de ces outils

Conception et Réalisation d'un Système décisionnel pour le marketing

dérivés du monde de la statistique : découvrir la donnée importante, alors qu'elle est dissimulée quelque part au fond de l'entrepôt de données, sans pour autant avoir à la chercher vous-même.

Ces outils vont appliquer à des données, automatiquement, un certain nombre de méthodes statistiques (coefficients de corrélation, réseaux neuronaux, Algorithmes génériques...) pour y détecter des « anomalies ». Ces anomalies sont en réalité des comportements auxquels les dirigeants n'auraient pas pensé initialement.

Les outils de data mining font appel à des algorithmes mathématiques de haut niveau. D'autres outils permettent d'appliquer à vos données des formules statistiques simples. Vous pourrez par exemple distribuer sur une carte à deux dimensions vos clients en fonction de leur chiffre d'affaires sur les deux dernières années. Apparaîtront visuellement les nouveaux clients, les clients perdus et les clients fidèles. (14)

La démarche Data Mining va permettre de fouiller et d'analyser les données brutes afin de répondre à des besoins très opérationnels. Parmi ceux-ci, citons :

+ **La segmentation des clients :**

Qui consiste à créer des segments (groupes) de clients ayant des caractéristiques en communs (comportements d'achat, profils socio-démographiques, ...). L'identification et la caractérisation de ces segments permettent :

- ✓ d'optimiser les actions commerciales et marketing vers les clients les plus rentables.
- ✓ de mieux cibler les actions de fidélisation.
- ✓ de disposer des informations nécessaires à la conception de nouvelles offres et marketing.

De nombreuses méthodes de segmentation existent, l'une des plus simples, la segmentation **RFM** qui permet d'identifier rapidement les principales caractéristiques des clients en les analysant sur trois variables (à adapter selon le métier) :

- **Récence** du dernier achat. (**Récence** du dernier rechargement/appel).
- **Fréquence** des achats. (**Fréquence** des rechargements/appels).
- **Montant** des achats. (**Montant** des rechargements/appels ou **Durée** d'appel).

+ Le Scoring :

Qui attribue un score (une note) aux clients (ou prospects) :

- ✓ Un score mesurant l'appétence à certains produits ou services
- ✓ Un score mesurant le risque que le client vous quitte (on parle de score d'attrition ou churn dans le secteur des télécommunications)
- ✓ Un score d'octroi pour un crédit

II.7. Conclusion :

On a présenté dans ce chapitre l'architecture d'un SID, on a vu qu'un SID inclut plusieurs composants sous forme de couches, chaque couche a un objectif bien précis et nécessite le développement d'un ensemble outils.

Dans le chapitre suivant on présentera un ensemble de méthodes pour la conception d'un SID.

CHAPITRE III.

Méthodes de

conception d'un SID

CHAPITRE III. Méthodes de conception d'un SID

III.1. Introduction :

Parmi Les bons pratiques du développement des logiciels est la conception et la modélisation du logiciel avant son implémentation et mise en œuvre. Ce chapitre a pour objectif de présenter un ensemble de méthodes pour la conception d'un SID. On commencera par l'évocation de quelques notions dans le domaine de méthodologies puis on présentera un ensemble de méthode pour la conception d'SID et en terminant par la présentation des concepts de la modélisation multidimensionnelle.

III.2. Définition d'une méthode de développement :

Il ya une confusion entre les 2 termes « Méthodologie» et « Méthode». Premièrement, les 2 termes sont très utilisés mais un peu définis. Deuxièmement, dans plusieurs cas les 2 termes sont utilisés pour référer à la même chose (processus de développent). D'autres utilisent le terme « Méthode» comme une manière spécifique ou une approche pour résoudre un problème et le terme « Méthodologie» comme la science qui étudier les méthodes, dans ce mémoire on a opté pour la derrière proposition. (9)

Une méthode de génie logiciel décrit les différentes phases et leurs enchaînements dans le processus de développement d'une application informatique. Les phases les plus récurrentes dans les méthodes de génie logiciel sont *l'analyse de l'existant, la spécification, la conception, l'implémentation et l'évaluation*. Une phase d'*intégration* est parfois explicitée lorsqu'il s'agit de combiner deux processus de développement relatifs à deux domaines différents.

Plusieurs modèles existent pour l'enchaînement des différentes phases. Chacun de ces modèles convient plus ou moins à un domaine et répond à des besoins ou à des principes particuliers. Nous citerons, par exemple, le modèle en X (16) qui repose sur le principe de réutilisation et qui est surtout utilisé dans un cadre de développement orienté objet ou le modèle en V (17) qui est adapté pour la prise en compte des facteurs humains dans le développement des interfaces homme-machine.

Les modèles de génie logiciel les plus utilisés sont le modèle en **cascade**, le modèle **incrémental** et le **processus unifié**.

III.2.1. Le modèle en cascade :

Le modèle en cascade a été introduit en 1970 par Royce (18) puis étendu par Boehm (19). Le modèle en cascade définit une réalisation séquentielle des phases dans le processus de développement logiciel avec des retours possibles à la phase précédente pour prendre en compte des lacunes identifiées. Ce modèle est très utilisé par les entreprises et est à la base de plusieurs autres modèles comme le modèle en V (20) ou le modèle en spirales.

III.2.2. Le modèle incrémental :

Le modèle incrémental appelé aussi modèle évolutif consiste à développer un logiciel en procédant par incrément (ou lot). Ce modèle applique itérativement le modèle en cascade mais sans trop détailler les phases pour répondre à des besoins de certains logiciels tellement mouvants qu'il est illusoire de fonder le développement sur une spécification figée. Il fournit pour, Des auteurs comme *Toffolon* (21) différencient le modèle incrémental du modèle évolutif en précisant que dans le modèle incrémental toutes les versions du logiciel doivent être planifiées avant le démarrage des cycles. Chaque lot a une version du logiciel qui sera immédiatement mise en exploitation dans un environnement opérationnel.

III.2.3. Le processus unifié :

Un processus unifié appelé aussi processus UP (Unified Process) est un processus de développement logiciel construit sur UML (Unified Modelling Language). Il est itératif, centré sur l'architecture, conduit par les cas d'utilisation et piloté par les risques (22). UP est un processus générique de développement logiciel c'est-à-dire qu'il est nécessaire de l'adapter au contexte du projet, de l'équipe, du domaine et/ou de l'organisation. Il existe donc un certain nombre de méthodes issues d'UP comme RUP (Rational Unified Process), AUP (Agile Unified Process), EUP (Enterprise Unified Process) ou ICONIX Unified Process.

Plus la dimension enchaînement des tâches une autre dimension est importante dans le domaine de méthodologies de développement des logiciels, c'est les formalismes de représentation d'une méthode.

Ces derniers permettent de fournir une vue abstraite d'un système. Ces formalismes dépendent fortement de l'objectif visé par la méthode et diffèrent d'une méthode à une autre. A titre d'exemple, nous citerons SADT (Structured Analysis and Design Technique) qui utilise des datagrammes et des actigrammes pour la modélisation des *fonctionnalités* des systèmes et Tropos (23) qui propose un formalisme pour une analyse et une modélisation *orientée but* pour le développement des systèmes à base d'agents logiciels.

Depuis le milieu des années 90, le langage UML (Unified Modelling Language), a connu un fort engouement et est devenu une norme OMG (Object Management Group) en 1997. Dans le cadre d'un développement d'un système d'information personnalisé, nous privilégierons les méthodes utilisant le langage UML. Ceci afin de faciliter la prise en main de la méthode par l'utilisation de notations standardisées et connues par un grand nombre d'utilisateurs (développeurs). (24)

III.3. Etat de l'art des méthodes de développement de SID:

La conception d'un SID est une tâche complexe car le SID touche plusieurs unités organisationnelles et peut concerner divers gens selon ses préoccupations. (6)

Un système d'information décisionnel a pour fonction première d'extraire des données des sources opérationnelles afin de les transformer en informations accessibles et pertinentes pour faciliter les prises de décision. Les besoins d'analyse des utilisateurs en vue de ces prises de décision requièrent des modèles qui présentent le sujet d'analyse suivant différents axes. Les modèles de données adaptés et reconnus sont le schéma en étoile et ses extensions, le schéma en flocon de neige, le schéma en constellation. (5)

Les méthodes d'ingénierie des systèmes d'information utilisent d'autres modèles tel que le modèle E-A, elles s'avèrent inadaptées aux développements des SID. Face à ce problème, des méthodes ont été proposées pour l'ingénierie des SID. Cependant, il n'existe pas de méthode qui prend en compte toutes les spécificités des SID. De plus, et en dépit de la notoriété du schéma en étoile, aucune méthode n'est reconnue par la communauté scientifique et la communauté industrielle. (25)

Notant que les méthodes citées dans cet état de l'art seront présentées dans le cadre de référence des quatre mondes qu'a été proposé au départ pour l'ingénierie des systèmes et a prouvé son efficacité pour comprendre diverses disciplines telles que l'ingénierie des systèmes d'information (26), l'ingénierie des exigences (27), l'ingénierie des méthodes et l'ingénierie de processus (27).

III.3.1. Le cadre de référence des quatre mondes :

Le cadre de référence des quatre mondes (Figure 3.1) est utilisé comme base pour étudier et comprendre une méthode d'ingénierie des SID (iSID). Pour chaque monde, une vue particulière de la méthode est adoptée en incluant un ensemble de facettes et d'attributs.

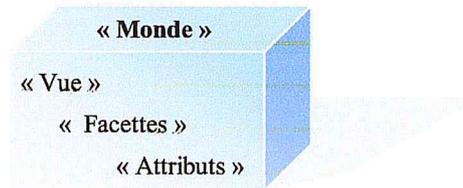


Figure 3.1. Le cadre de référence des quatre mondes (1)

Il est possible d'identifier et d'étudier quatre principaux points de vue relatifs aux méthodes d'ingénierie des SID :

- ✓ Quelle est l'essence de la méthode iSID ?
- ✓ Pourquoi (et quand) utilise-t-on les méthodes iSID ?
- ✓ Avec quoi les méthodes iSID sont mises en œuvre ?
- ✓ Comment les méthodes iSID sont développées ?

Ce cadre de référence permet de :

- Comprendre et étudier l'ensemble des propositions des différentes méthodes de la littérature grâce à la structure proposée en adoptant les notions de multi-vues et multi-facettes.
- Définir précisément et systématiquement le positionnement d'une méthode particulière. En effet, grâce à la richesse de ce cadre de référence en termes de facettes et d'attributs, il est possible de caractériser une méthode particulière avec grande précision et de le faire d'une manière systématique.

Conception et Réalisation d'un Système décisionnel pour le marketing

- Comparer et confronter deux méthodes de manière détaillée pour faire apparaître les similarités et les différences.
- Sélectionner la méthode qui s'adapte le plus à la situation en adoptant un point de vue particulier et en choisissant la méthode qui correspond au mieux aux caractéristiques désirées.
- Positionner un ensemble de méthodes et découvrir par la suite des patrons et des regroupements suivant les vues, les facettes et les attributs. Ceci permet aussi d'identifier des niches de recherches et de localiser les manques par rapport à un aspect en particulier.
- Comme montre la Figure 3.2, chaque vue permet de capturer une connaissance particulière des approches SID. Le monde sujet répond à la question " « Quelle est l'essence de la méthode iSID? ». Deuxièmement, la vue usage explique « Pourquoi (et quand) utilise-t-on les méthodes iSID? ». Puis, la vue système explique « Avec quoi les méthodes iSID sont mises en œuvre? ». Enfin, la vue développement explique « Comment les méthodes iSID sont développées? ».

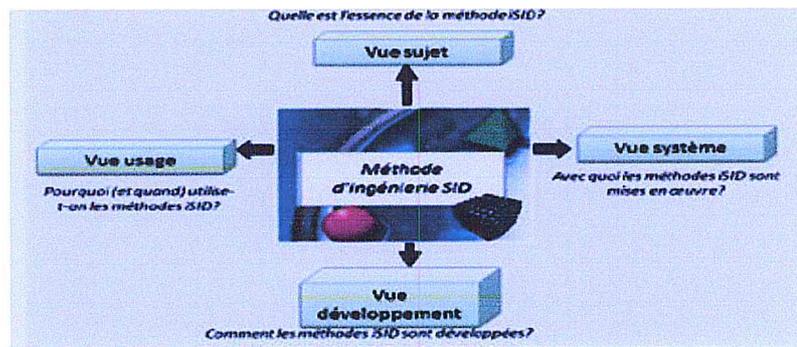


Figure 3.2. Les quatre vues pour les méthodes de développement des SID (1)

Attribuée aux méthodes de développement des SID, une facette représente un moyen de classification. Les facettes ont été utilisées par (28) pour classer les composants réutilisables. Elles sont également employées pour l'ingénierie des exigences pour comprendre et classer les approches basées sur les scénarii. Chaque facette est mesurée par un ensemble approprié d'attributs. Les attributs ont des valeurs qui sont définies dans un domaine. Un domaine peut être de type prédéfini (ENTIER, BOOLÉEN, etc.), un

type énuméré (ENUM {x, y, z}), ou un type structuré (ENSEMBLE ou TUPLE). Par exemple, « approche » est une facette de la « vue sujet » qui aide à classer les approches SID selon l'approche utilisée et qui peut être un des attributs : « Ascendante », « Descendante » ou « Mixte ».

✚ **Vue Sujet :**

La vue sujet comporte la connaissance du domaine au sujet d'une méthode d'iSID (« Quelle est l'essence de la méthode iSID? »). C'est la vue qui fournit des informations sur le sujet. L'intérêt de cette vue est d'étudier

- ❖ l'approche de la méthode
- ❖ la couverture de la méthode définissant l'élément ou les éléments cibles
- ❖ la stratégie de la méthode

Trois facettes sont associées à la vue sujet: l'approche, la couverture et la stratégie. Le reste de la section présente en détail chacune des facettes :

- ✓ La facette « approche » est définie comme suit :
ENUM {Ascendante, Descendante, Mixte},
- ✓ La facette « Couverture » est définie comme suit :
ENSEMBLE DE (ENUM {Exigence, Spécification solution, Système})
- ✓ La facette « stratégie » est défini comme suit :
ENUM {ENSEMBLE DE {Itérative, Incrémentale}, Big bang}

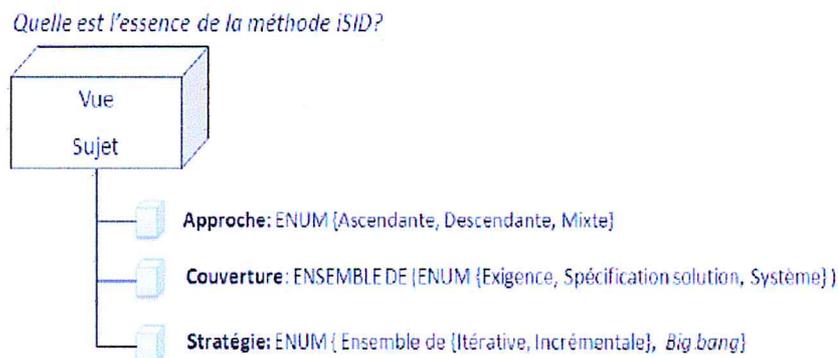


Figure 3.3. Facettes et attributs de la vue sujet (1)

+ **Vue Usage :**

La vue usage décrit la finalité et les buts des utilisateurs du SID. Cette vue répond à la question : « Pourquoi (et quand) utilise-t-on les méthodes iSID? ». L'intérêt de cette vue est d'étudier :

- ❖ les utilisateurs de la méthode
- ❖ la finalité de la méthode
- ❖ le but de la méthode

Trois facettes sont associées à la vue usage: les utilisateurs, la finalité de la méthode et le but de la méthode. Le reste de la section présente en détail chacune des facettes.

- ✓ La facette « utilisateur » est définie comme suit :
ENSEMBLE DE (ENUM {Stratégique, Tactique, Système})
- ✓ La facette « Finalité » est définie comme suit :
ENSEMBLE DE (ENUM {Découverte des exigences, Modélisation du schéma multidimensionnel, Analyse des données}).
- ✓ La facette « But d'usage » est définie comme suit :
ENUM {Prescriptif, Descriptif}

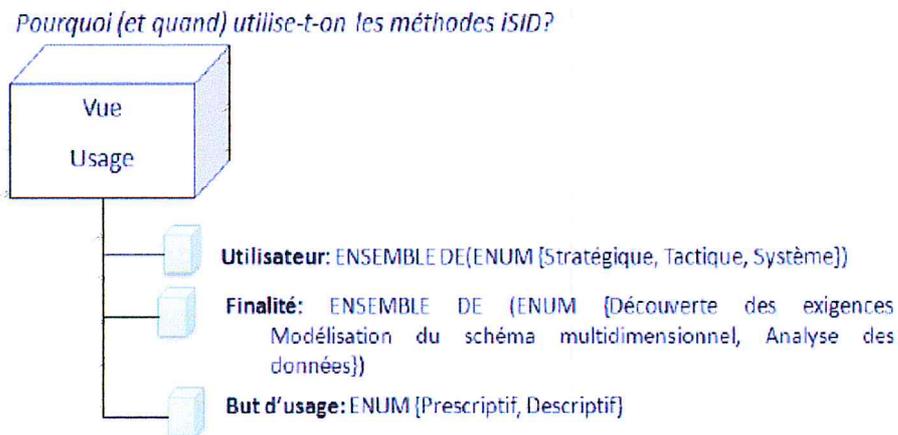


Figure 3.4. Facettes et attributs de la vue usage (1)

+ Vue système :

La vue système répond à la question : « Avec quoi les méthodes iSID sont mises en œuvre? ». L'intérêt de cette vue système est de définir:

- ❖ Niveau de modélisation de la méthode
- ❖ les moyens de mise en œuvre

Deux facettes sont associées à la vue système: le niveau de la modélisation, les moyens de représentations. Le reste de la section présente en détail chacune des facettes.

- ✓ L'attribut « niveau de modélisation » est définie comme suit :
ENUM {Conceptuel, Logique, Physique}
- ✓ L'attribut « paradigme » est définie comme suit :
ENUM {E/A, Objet, spécifique, Non défini}
- ✓ L'attribut « modèle logique» est défini comme suit :
ENUM {Multidimensionnel, Relationnel, Objet, Semi-structuré, Non défini}
- ✓ L'attribut « modèle logique» est défini comme suit :
ENUM {Multidimensionnel, Relationnel, Objet, Semi-structuré, Non défini}
- ✓ L'attribut « modèle des exigences » est défini comme suit :
ENUM {Langage naturel, Modèle classique, Requête, Modèle de but, Non défini}
- ✓ L'attribut « analyse des données » est défini comme suit :
ENUM {Prédictive, Par découverte, Détection de déviation, Non défini}

Avec quoi les méthodes iSID sont mises en œuvre?



Figure 3.5. Facettes et attributs de la vue système (1)

+ **Vue développement :**

La vue développement concerne les différents aspects relatifs au processus du développement des approches SID. La vue s'intéresse au processus qui crée le système SID. Cette vue répond à la question : « Comment les méthodes iSID sont développées? ». L'intérêt de cette vue développement est de définir:

- ❖ comment est l'apprentissage de la méthode iSID
- ❖ comment est l'adaptabilité de la méthode iSID
- ❖ comment est la démarche de méthode iSID

Trois facettes sont associées à la vue développement: apprentissage, adaptabilité et démarche. Le reste de la section présente en détail chacune des facettes.

- ✓ La facette « apprentissage » est définie comme suit :
ENUM {Réutilisation, Capitalisation, À partir de rien}
- ✓ La facette « adaptabilité » est définie comme suit :
ENUM {Rigide, Adaptative}
- ✓ La facette « Démarche » est définie comme suit :
ENSEMBLE de (ENUM {Activité, Produit})
- ✓ La facette « Prototype » est définie comme suit :
BOOLEEN

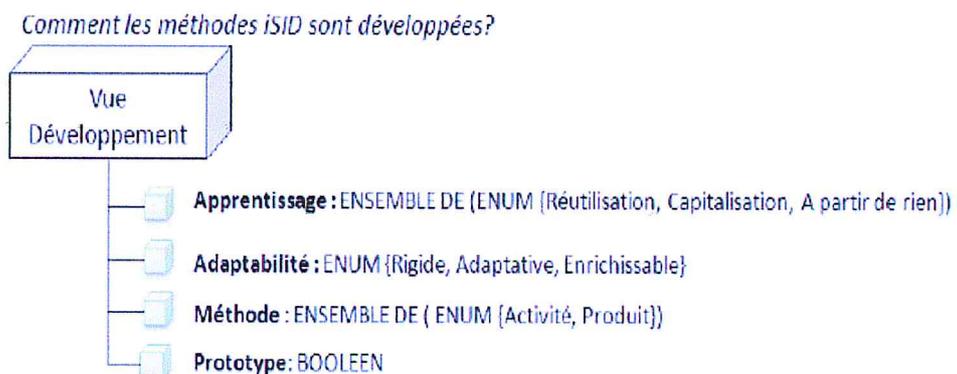


Figure 3.6. Facettes et attributs de la vue développement (1)

III.3.2. Positionnement de quelques méthodes d'ingénierie de SID dans le cadre de référence des quatre mondes :

Dans qui suit on va présenter un ensemble de méthodes de développement de SID dans le cadre de référence des quatre mondes.

+ Méthode de (Annoni et al) (29) :

<u>Vue Sujet</u>
Approche: Mixte
Couverture : exigence, spécification de solution, système
Stratégie: itérative et incrémental
<u>Vue Usage</u>
Utilisateurs : stratégique, tactique, système
Finalités : découverte des exigences (avancée)et modélisation du schéma multi-dimensionnel
But d'usage : prescriptif
<u>Vue Système</u>
Niveau de modélisation : conceptuel, logique
Moyen :
✓ Paradigme conceptuel : objet
✓ Modèle logique: multidimensionnel
✓ Modèle des exigences : matble de requete
<u>Vue Développement</u>
Apprentissage : Réutilisation, capitalisation
Adaptabilité : Adaptative
Démarche : Activité
Prototype : eBIPAD

Tableau 3.1. Positionnement de la méthode [Annoni et al, 2007] dans le cadre des 4 vues

+ **Méthode de (Sergio Luján Mora) (6):**

<p style="text-align: center;"><u>Vue Sujet</u></p> <p>Approche: Mixte</p> <p>Couverture : exigence, spécification de solution, système</p> <p>Stratégie: itérative et incrémental</p> <p style="text-align: center;"><u>Vue Usage</u></p> <p>Utilisateurs : tactique</p> <p>Finalités : découverte des exigences (négligée) et modélisation du schéma multi-dimensionnel, modélisation de l'ETL.</p> <p>But d'usage : prescriptif</p> <p style="text-align: center;"><u>Vue Système</u></p> <p>Niveau de modélisation : conceptuel, logique, physique</p> <p>Moyen :</p> <ul style="list-style-type: none">✓ Paradigme conceptuel : objet✓ Modèle logique: multidimensionnel (extension d'UML)✓ Modèle des exigences : non défini <p style="text-align: center;"><u>Vue Développement</u></p> <p>Apprentissage : À partir de rien</p> <p>Adaptabilité : Adaptative</p> <p>Démarche : Activité</p> <p>Prototype : DWEP</p>

Tableau 3.2. Positionnement de la méthode [Luján Mora 2005] dans le cadre des 4 vues

+ **Méthode de (Mazon et al) (30):**

<p style="text-align: center;"><u>Vue Sujet</u></p> <p>Approche : Descendante Couverture : Exigence, système Stratégie : Incrémental</p> <p style="text-align: center;"><u>Vue Usage</u></p> <p>Utilisateurs : Tactique Finalités : Découverte des exigences, modélisation du schéma multidimensionnel But d'usage : Prescriptif</p> <p style="text-align: center;"><u>Vue Système</u></p> <p>Niveau de modélisation : Conceptuel, logique Moyen : Paradigme conceptuel : Non défini Modèle logique : Multidimensionnel Modèle des exigences : Modèle en but</p> <p style="text-align: center;"><u>Vue Développement</u></p> <p>Apprentissage : À partir de rien Adaptabilité : Adaptable Démarche: Activité Prototype : Non</p>

Tableau 3.3. Positionnement de la méthode [Mazon et al, 2005] dans le cadre des 4 vues

+ **Méthode de (Prakash et al) (31) :**

<p style="text-align: center;"><u>Vue Sujet</u></p> <p>Approche : Descendante Couverture : Exigence, système Stratégie : Big bang</p> <p style="text-align: center;"><u>Vue Usage</u></p> <p>Utilisateurs : Stratégique Finalités: Découverte des exigences, modélisation du schéma multidimensionnel. But d'usage : Prescriptif</p> <p style="text-align: center;"><u>Vue Système</u></p> <p>Niveau de modélisation : Conceptuel, logique Moyen :</p> <ul style="list-style-type: none">✓ Paradigme conceptuel : E/A✓ Modèle logique : relationnel✓ Modèle des exigences : requête <p style="text-align: center;"><u>Vue Développement</u></p> <p>Apprentissage : À partir de rien Adaptabilité : Rigide Démarche: Produit, processus Prototype : Non</p>

Tableau 3.4. Positionnement de la méthode [Prakash et al, 2003] dans le cadre des 4 vues

+ **Méthode de (Prat et al) (32):**

<u>Vue Sujet</u>
Approche: Descendante
Couverture : Spécification solution, système
Stratégie : Incrémental
<u>Vue Usage</u>
Utilisateurs : Tactique
Finalités : Découverte des exigences (réduite), modélisation du schéma multidimensionnel
But d'usage : Descriptif
<u>Vue Système</u>
Niveau de modélisation : Conceptuel, logique
Moyen:
✓ Paradigme conceptuel : spécifique (UMM)
✓ Modèle logique : multidimensionnel (UMM) (idéal : 1)
✓ Modèle des exigences : langage naturel
<u>Vue Développement</u>
Apprentissage : À partir de rien
Adaptabilité: Adaptative
Démarche : Activité, produit
Prototype : Non

Tableau 3.5. Positionnement de la méthode [Prat et al, 2002] dans le cadre des 4 vues

✚ **Méthode de (Bonifati et al) (33):**

<p style="text-align: center;"><u>Vue Sujet</u></p> <p>Approche : Mixte Couverture : Exigence, Spécification solution, Système Stratégie : Itérative</p> <p style="text-align: center;"><u>Vue Usage</u></p> <p>Utilisateurs: Stratégique Finalités : Découverte des exigences (avancée), modélisation du schéma multidimensionnel But d'usage : Descriptif</p> <p style="text-align: center;"><u>Vue Système</u></p> <p>Niveau de modélisation : Conceptuel, logique Moyen :</p> <ul style="list-style-type: none">✓ Paradigme conceptuel : Non-défini✓ Modèle logique: Multidimensionnel (étoile)(idéal : 1, candidat : n)✓ Modèle des exigences: Langage naturel et GQM <p style="text-align: center;"><u>Vue Développement</u></p> <p>Apprentissage : À partir de rien Adaptabilité : Adaptative Démarche : Activité, produit Prototype : Non</p>
--

Tableau 3.6. Positionnement de la méthode [Bonifati et al, 2001], dans le cadre des 4 vues

✦ **Méthode de (Ravat et al) (34):**

<p style="text-align: center;"><u>Vue Sujet</u></p> <p>Approche : Ascendante Couverture : Spécification solution, système Stratégie : Incrémentale</p> <p style="text-align: center;"><u>Vue Usage</u></p> <p>Utilisateurs : Système Finalités : Modélisation du schéma multidimensionnel But d'usage : Descriptif</p> <p style="text-align: center;"><u>Vue Système</u></p> <p>Niveau de modélisation : Conceptuel, logique Moyen :</p> <ul style="list-style-type: none">✓ Paradigme conceptuel : Objet✓ Modèle logique : Objet✓ Modèle des exigences : Non-défini <p style="text-align: center;"><u>Vue Développement</u></p> <p>Apprentissage : A partir de rien Adaptabilité : Rigide Démarche : Activité Prototype : Non</p>

Tableau 3.7. Positionnement de la méthode [Ravat et al, 1999] dans le cadre des 4 vues.

✚ **Méthode de (Tryfona et al) (35):**

<u>Vue Sujet</u>
Approche : Ascendante
Couverture : Système
Stratégie : Big bang
<u>Vue Usage</u>
Utilisateurs : Système
Finalités : Modélisation du schéma multidimensionnel
But d'usage : Descriptif
<u>Vue Système</u>
Niveau de modélisation : Conceptuel, logique
Moyen :
✓ Paradigme conceptuel : Objet
✓ Modèle logique : Objet
✓ Modèle des exigences : Non-défini
<u>Vue Développement</u>
Apprentissage : A partir de rien
Adaptabilité : Rigide
Démarche : Produit
Prototype : Non

Tableau 3.8. Positionnement de la méthode [Tryfona et al, 1999] dans le cadre des 4 vues

✚ **Méthode de (Kimball et al) (7) :**

<u>Vue Sujet</u>
Approche : Descendante
Couverture : Spécification solution, système
<u>Vue Usage</u>
Stratégie : Itérative
Utilisateurs : Tactique, système
Finalités : Découverte des exigences (réduite), modélisation du schéma multidimensionnel
But d'usage : Incrémental
<u>Vue Système</u>
Niveau de modélisation : Conceptuel, logique
Moyen :
✓ Paradigme conceptuel : Spécifique
✓ Modèle logique : Multidimensionnel (étoile) (idéal :1)
✓ Modèle des exigences : Non-défini
<u>Vue Développement</u>
Apprentissage : A partir de rien
Adaptabilité : Rigide
Démarche : Activité, produit
Prototype : Non

Tableau 3.9. Positionnement de la méthode [Kimball et al, 1998] dans le cadre des 4 vues

III.3.3. Modélisation des DW « Modélisation Multidimensionnelle (MD) » :

Les DW sont axés sur le traitement analytique en ligne (OLAP), différent des systèmes de traitement transactionnel en ligne (OLTP) par leur architecture et leurs objectifs. Une différence requiert un modèle différent (36), c'est le modèle multidimensionnel MD qu'a été introduit par Ralph Kimball (3). Il consiste en deux nouveaux concepts fondamentaux: *les faits* et *les dimensions*. Le modèle multidimensionnel est la structure de données la plus utilisée et la plus appropriée aux requêtes et analyses des utilisateurs d'entrepôts de données. (1)

Un fait représente un sujet d'analyse, caractérisé par une ou plusieurs *mesures*. Ce fait est analysé selon des axes d'observation qui constituent également ses descripteurs. Ces dimensions peuvent présenter des *hiérarchies* qui offrent la possibilité de réaliser des analyses à différents *niveaux de granularité* (niveaux de détail).

Un entrepôt de données présente alors une modélisation dite « Multidimensionnelle » puisqu'elle répond à l'objectif d'analyser des faits en fonction de dimensions qui constituent les différents axes d'observation des mesures. (8)

III.3.3.1. Concepts de bases :

+ **Fait :**

Le fait modélise le sujet de l'analyse. Un fait est formé de mesures correspondant aux informations qui ne sont autres que des indicateurs décrivant le sujet d'analyse. Les mesures d'un fait sont numériques et généralement valorisées de manière continue (5). Par exemple, dans le fait Ventes, nous pouvons avoir la mesure "Quantité de produits vendus par magasin".

+ **Dimension :**

Le fait est analysé suivant différentes perspectives. Ces perspectives correspondent à une catégorie utilisée pour caractériser les mesures d'activité analysées (37); on parle de dimensions. Une dimension est composée de paramètres correspondant aux informations faisant varier les mesures de sujet analysé. L'emplacement géographique, le temps, les catégories de produit, les employés d'une organisation, les cours d'une université sont tous des exemples de thèmes pouvant correspondre à une dimension.

+ Niveau :

Une dimension peut se diviser en plusieurs niveaux, pour modéliser des phénomènes à différentes échelles ou niveaux de détail. Par exemple, une dimension représentant un découpage administratif peut se diviser en niveaux «Pays», «Province», « Région économique » et « Division de recensement ». (11)

+ Hiérarchie :

Certaines dimensions peuvent comporter plus d'un ensemble de niveaux permettant de décrire un thème d'analyse. Par exemple, on peut décrire le temps selon un découpage jour-semaine ou encore jour-mois-année. On parle alors de hiérarchies multiples ou alternatives. Une dimension peut ainsi comporter plus d'une hiérarchie, offrant des chemins différents pour remonter vers les membres de niveau plus général. En règle générale, le niveau le plus détaillé est commun à toutes les hiérarchies à l'intérieur d'une dimension, puisque ce sont les membres de ce niveau qui définissent le grain de la dimension et par conséquent l'adressage des faits détaillés. (11)

+ Granularité

La granularité est le niveau de détail décrivant un fait. La granularité détermine le volume des données ainsi que le type des requêtes que l'utilisateur peut poser. Prenons l'exemple d'une compagnie de téléphonie cité par (4). Il existe deux systèmes différents ; un premier système avec le détail de tous les appels, un deuxième système avec un résumé mensuel de tous les appels.

Considérons la requête « Quelle est la personne qui a fait le plus d'appels? ». On peut obtenir une réponse dans le premier système mais non dans le deuxième système.

D'un autre côté la requête, « Quelle est la moyenne des appels longue distance ? », a une réponse dans les deux systèmes. Dans le premier système, il faudra calculer la moyenne de plusieurs enregistrements. Dans le deuxième système il suffit d'accéder à un seul enregistrement. Cet exemple montre qu'il y a un compromis à faire entre la performance et le stockage des données. Il est ainsi intéressant de garder les données détaillées dans le cas où il n'y aurait pas de problèmes de stockage et d'avoir des données agrégées pour assurer une meilleure performance. Quel que soit le niveau de granularité

choisie, les données des différentes hiérarchies nécessitent une agrégation pour pouvoir répondre aux questions des utilisateurs. L'agrégation dans les outils OLAP assure la performance. (1)

III.3.3.2. Niveau conceptuel :

La conception et la construction d'un entrepôt de données représentent un enjeu important pour la communauté des bases de données. Cette conception doit concerner les trois niveaux de la modélisation classique, conceptuelle et logique puis physique.

Contrairement à la modélisation logique, il n'existe à ce jour aucun consensus sur la modélisation conceptuelle des entrepôts de données comme cela peut être le cas avec la méthode MERISE pour la conception des bases de données relationnelles. Différents travaux ont été proposés dans la littérature:

Ceux basés sur le modèle E/A (Entité-Association) (35), représentent les concepts décisionnels via des extensions plus ou moins importantes des concepts d'entités et d'associations, par exemple dans (35), les auteurs proposent un modèle multidimensionnel appelé StarER qui a pour concepts de base des faits, des entités, des relations et des attributs.

D'autres approches basées sur les modèles orientées objet, comme UML, (34), (29), , par exemple dans Luján Mora propose un profil (ensemble de stéréotypes, de valeurs typées et de contraintes) pour spécialiser UML à la modélisation multidimensionnelle. Le fait est représenté par une classe avec le stéréotype «Fact». Une dimension est représentée par une classe avec le stéréotype «Dimension». Un autre modèle multidimensionnel appelé YAM2 basé sur la spécialisation du méta-modèle du diagramme de classes UML.

Finalement, des modèles plus spécifiques tels que ceux présentés dans (30), (33), (32), des auteurs proposent un modèle appelé "DFM : DimensionalFact Model". Cette modélisation multidimensionnelle graphique distingue clairement les concepts tels que les faits, les dimensions, les mesures, les attributs faibles et les hiérarchies. Cependant, les modèles spécifiques ne sont pas connus par les concepteurs contrairement à ceux basés sur le modèle E/A et UML, cela peut constituer un obstacle. Le Tableau suivant

résume la classification des travaux traitant le problème de conception et la construction des entrepôts de données des différents modèles :

Les travaux sur les modèles conceptuels	Les différents modèles	Observation
[Sapia et al, 1999] [Tryfona et al, 1999], [Sánchez et al, 1999], [Franconi et Sattler, 1999] [Franconi et Kimble, 2004], [Trujillo et Palomar 1998]	Le modèle Entité - Association (E/A)	Le Modèle le plus utilisé dans la conception des SI, connu par tous les concepteurs (familiarisés avec le E/A) Des caractéristiques de la modélisation multidimensionnelle, comme agrégation, spécialisation, n'existe pas.
[Prat et Akoka, 2002], [Trujillo et al, 2002], [Luján-Mora et al, 2006] [Abello et al, 2002 2006]	Le modèle orienté objets	Profitent de la popularité de la modélisation UML. (agrégation, spécialisation, etc. UML est un standard et extensible Puissance d'expression des aspects statiques et dynamiques des SID.
[Golfarelli et al, 1998], [Cabbibo et Torlone 1998], [Moody et Kortink, 2000], [Husemann et al, 2000], [Tsois et al, 2001], [Vassiliadis et al, 2002, 2005].	Le modèle spécifique	Utilisent exclusivement des concepts liés au décisionnel. Modèle non connu par les concepteurs, nécessite alors la maîtrise de ces modèles.

Tableau 3.10. Les différents modèles utilisés dans la conception des DW. (38)

III.3.3.3. Niveau logique :

Pour construire un modèle approprié pour un entrepôt de données, nous pouvons choisir, soit un schéma relationnel (schéma en étoile, en flocon de neige ou en constellation) où les données sont stockées dans un SGBD relationnel ; soit un schéma multidimensionnel (Cube) où les données sont stockées dans une base de données multidimensionnelle, ces différents schémas relevant d'un niveau logique de conception, en raison du recours à la notion de table (table de faits, table de dimensions) (5), (35).

+ **Modèle en étoile :**

Ce modèle représente visuellement une étoile, on parle de modèle en étoile (Star Schéma (7)), où tous les faits sont définis dans une simple table relationnelle. Cette table va être reliée par sa clé primaire à d'autres tables correspondant aux dimensions. La Figure 3.7 illustre ce modèle. En fait, le modèle en étoile essaie de superposer une structure multidimensionnelle au-dessus d'un modèle relationnel normalisé à deux dimensions. Le modèle en étoile simplifie le modèle logique normalisé en organisant les données de manière optimale pour les traitements d'analyse.

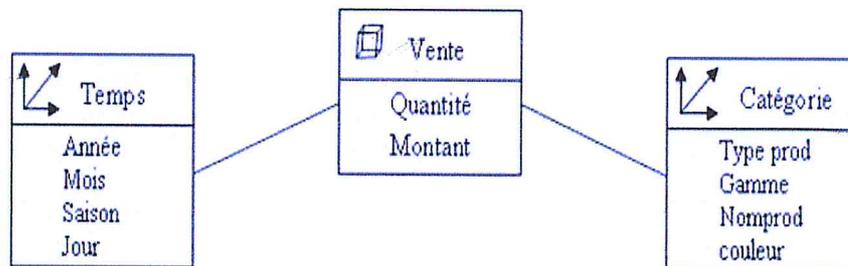


Figure 3.7. Exemple d'un modèle en étoile

+ **Modèle en flocon (snowflake) :**

La modélisation en flocon est une modélisation en étoile pour laquelle nous éclatons les tables de dimension en sous-tables selon la hiérarchie de cette dimension. Il correspond à un schéma en étoile dans lequel les dimensions ont été normalisées, faisant ainsi apparaître des hiérarchies de dimension de façon explicite. La normalisation permet un gain d'espace de stockage en évitant la redondance de données, mais engendre une dégradation des performances, dans la mesure où elle multiplie le nombre de jointures à effectuer pour l'analyse.

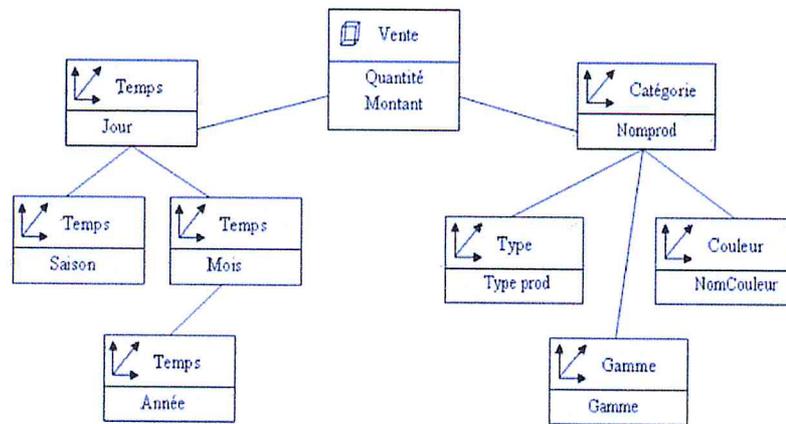


Figure 3.8. Exemple d'un modèle en flocon

✦ **Modélisation en constellation :**

Il s'agit de fusionner plusieurs modèles en étoile qui utilisent des dimensions communes. Un modèle en constellation comprend donc plusieurs faits et des dimensions communes ou non

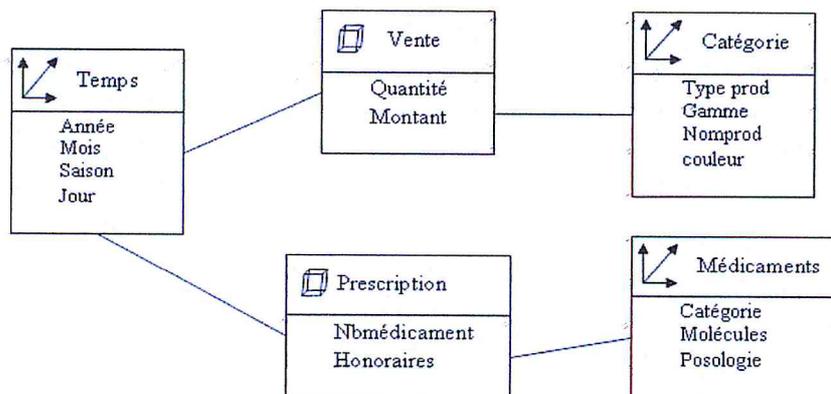


Figure 3.9. Exemple d'un schéma en constellation

III.3.3.4. Niveau physique :

Le stockage de données et les méthodes d'accès associées (les techniques d'indexation) ont été étudiés depuis plusieurs années. Cependant, le gros volume géré par les entrepôts et les besoins d'accès performants lors de l'interrogation ont remis en cause les techniques traditionnelles. Dans cette section nous présentons quelques techniques de stockage et d'accès aux données adaptées aux entrepôts.

+ Le stockage :

Le stockage des modèles multidimensionnels (les cubes) est réduit au problème de stockage d'un vecteur multidimensionnel, où chaque dimension de cube est associée à un axe du vecteur et chaque valeur d'une dimension est associée à une position de l'axe correspondant à la dimension. Les cellules du vecteur, i.e. les intersections entre les positions des axes, contiennent les mesures.

Le problème principal posé par la représentation d'un cube sous la forme d'un vecteur multidimensionnel est que celui-ci est creux. Ceci est dû au fait qu'en général seulement un nombre réduit de cellules d'un cube ont une valeur de mesure associée.

+ L'indexation :

Les techniques d'indexation utilisées dans les bases de données de type (OLTP) ne sont pas bien adaptées aux environnements des entrepôts de données. En effet, la plupart des transactions OLTP accèdent à un petit nombre de n-uplets, et les techniques utilisées (index B+ par exemple) sont adaptées à ce type de situation. Les requêtes décisionnelles adressées à un entrepôt de données accèdent au contraire à un très grand nombre de n-uplets (ce type de requêtes est encore appelé requêtes d'intervalle).

Les différents types d'index utilisés dans les entrepôts de données sont : index binaire, index de jointure et index de jointure en étoile (star join index).

✓ *Index binaire* : un index binaire est une séquence de bits, où chacun d'eux indique si l'enregistrement associé présente une propriété donnée. Actuellement, presque tous les SGBDR commerciaux fournissent des indexes binaires. L'avantage des indexes binaires est qu'il est possible d'exécuter des opérations logiques (par exemple les opérations ET, OU, XOR, NOT) de manière performante. L'autre avantage est que pour des attributs à nombre limités de valeurs, l'espace nécessaire pour leur stockage est réduit. Dans ce cas, l'index binaire peut être géré en mémoire, ce qui améliore les performances du système général. Des techniques de compression de données sont utilisées pour gérer les indexes binaires qui contiennent, presque dans leurs totalité, des bits à zéro.

Index de jointure : Ce type d'index appelé index de jointure (join indices) pour pré-joindre deux tables. L'index matérialise les liens existants entre deux tables par le biais d'une table à

deux attributs contenant les identifiants des n-uplets joints. Cet index peut être vu comme une jointure pré calculée.

✓ *Index de jointure en étoile* : Adapté aux requêtes définies sur les entrepôts de données modélisés en étoile. Un index de jointure en étoile peut contenir toute combinaison des identifiants des n-uplets de la table de faits et des identifiants des n-uplets des tables dimensions pouvant être jointes. Ce type d'index est dit complet s'il est construit en joignant toutes les tables de dimensions avec la table des faits. L'index de jointure en étoile complet est bénéfique à n'importe quelle requête définie sur un entrepôt modélisé en étoile.

✚ **Le partitionnement :**

Les entrepôts de données se prêtent bien à l'utilisation des techniques de partitionnement, car des requêtes complexes sur des grandes quantités de données peuvent être exécutées en parallèle améliorant éventuellement les temps de réponse.

Dans le cas des systèmes relationnels, les techniques de partitionnement de relation sont été étudiées depuis de nombreuses années. Les travaux récents dans le contexte des entrepôts de données visent au partitionnement vertical de la relation de fait d'un schéma en étoile. Dans le système multidimensionnel, il est possible de distinguer : La fragmentation horizontale se décline en deux versions : les fragmentations primaires et dérivées. La fragmentation primaire d'une relation est effectuée grâce à des prédicats de sélection définis sur la relation. La fragmentation horizontale dérivée s'effectue avec des prédicats de sélection définis sur une autre relation. La fragmentation horizontale apparaît plus spécialement appropriée. Une méthodologie de fragmentation horizontale pour décomposer un schéma d'entrepôt en étoile a été proposée, fragmenter quelques/toutes les tables de dimension avec la fragmentation horizontale primaire, puis fragmenter la table des faits en utilisant les schémas de fragmentation des tables de dimension.

PARTIE II: Réalisation du système

Résumé

On a présenté dans la partie précédente qu'est-ce qu'un système décisionnels et quelque méthodes de conception de ce genre de projets. Dans cette partie, on concevoir et réaliser un système décisionnel pour le marketing d'un opérateur de télécommunication mobile. Donc, on va faire une petite présentation de la structure d'accueil, on va présenter la conception du système par la méthode choisie et en fin implémentation mise en œuvre du système

III.4. Conclusion :

Dans ce chapitre on a présenté un ensemble de méthodes de conception d'SID, ces méthodes ont été présentées dans le cadre de référence des 4 mondes, on a vu que chacune de ces méthodes a ces caractéristiques. Des méthodes suivent la démarche ascendante, d'autres la démarche descendante et les nouvelles propositions utilisent une démarche mixte. La plupart des méthodes ne couvrent pas tous les composants d'un SID. On a terminé par présenter les concepts de base de la modélisation aux 3 niveaux d'abstraction.

Dans la partie suivante on va commencer à modéliser notre système, cette partie est divisée en 3 chapitres, (présentation de l'organisme d'accueil, modélisation du système, mise en œuvre du système).

CHAPITRE IV.

Présentation de l'organisme d'accueil

CHAPITRE IV. Présentation de l'organisme d'accueil :

IV.1. Introduction:

Algérie Telecom Mobile Mobilis, filiale d'Algérie Télécom, est le premier opérateur de téléphonie mobile en algérie. Devenu autonome en août 2003, subit une réorganisation en juin 2004, en décembre 2004 Mobilis à lancer le premier réseau de UMTS (Universel Mobile Télécommunication System) en Algérie intégrant dans le club des quarante opérateurs dans le monde qui maîtrisent cette technologie, devenu ainsi un véritable opérateur multimédia en Algérie, il propose à ses clients une large gamme de produits et de services innovants et de haute qualité.

Aujourd'hui plus de 96.70 % de la population Algérienne est couverte par un réseau de haute qualité grâce à 3424 BTS (Station de Base Radio) diffusé sur 3595 sites.

Son parc d'abonnés est compté à plus de 10 millions d'abonnés actif sur ses différentes offres prépayées et postpayées confondue, elle possède une part de marché calculé à 37.6%.

Dans l'esprit de rendre la Téléphonie Mobile accessible à toutes les bourses et à tous les niveaux de vie en Algérie. Ainsi que par souci de toujours satisfaire le moindre besoin et attente de ses clients.

IV.2. Organigramme generalde Mibilis:

Mobilis est une entreprise organisée selon les standards mondiaux de management, Elle est dirigée par un PDG, entouré de 03 grandes divisions :

✚ **Division Affaires Générales :** Elle se compose de 5 direction :

- Direction Affaires Juridique et Contentieux.
- Direction des Achats et de la Logistique.
- Direction de la Qualité.
- Direction de la Formation.
- Direction des Ressources Humaines.

Conception et Réalisation d'un Système décisionnel pour le marketing

+ **Division Réseau et Service** : Elle se compose de 4 direction :

- Direction Ingénierie et Développement Réseau.
- Direction Transmission.
- Direction Déploiement et Opération Réseau.
- Direction Maintenance Réseau.

+ **Division Réseau et Service** : Elle se compose de 4 direction :

- Direction Marketing GP (Grand Publique).
- Direction Distribution et Vente GP.
- Direction Relation Client GP.
- Direction Marché Entreprise

Ainsi que quatre autres directions indispensables (+ une cellule audit) qui sont :

- Direction du Système d'Information.
- Direction de la Marque et de la Communication.
- Direction Stratégie, Programmation et Performance
- Direction des Finances et de la Comptabilité.

De plus, le cabinet du PDG est composé de conseillers chargés chacun de tâches confiées par le PDG. Le PDG est également assisté d'un avocat conseil chargé des résolutions juridiques.

Conception et Réalisation d'un Système décisionnel pour le marketing

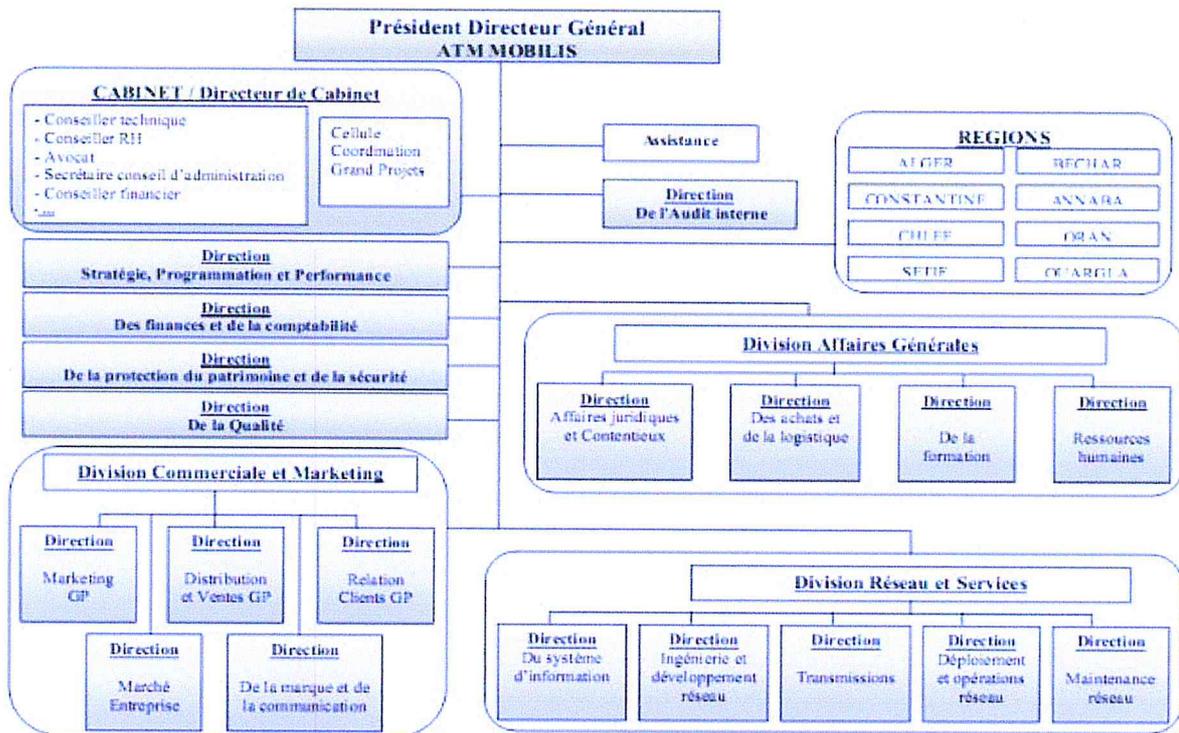


Figure 4.1. Organigramme générale Mibilis

IV.3. Présentation de la direction de marketing:

La direction Marketing grand publique est notre champ d'étude, elle est dévisée en en 4 sous direction (la sous direction qui nous intéresse est la Sous direction Marketing stratégique et relations partenaire) :

✦ **Sous direction Marketing stratégique et relations partenaire** : elle se devise en 3 départements :

- Département Analyse et études de marchés : Analyser les données internes et externes ;
- Département Connaissances clients : étudier et analyser les besoins clientèles.
- Département Relations équipementiers ;

✦ **Sous direction Marketing offres et services** : créer les offres à base des informations tenu depuis la sous direction Marketing stratégique et relations partenaire, la mise en œuvre d'un offre ou un service ce fait en contact avec la direction de la maintenance réseau (DMR).

Conception et Réalisation d'un Système décisionnel pour le marketing

- + **Sous direction Communication Produits** : Après la création d'un offre ou un service, cette sous direction le publie.
- + **Sous direction Interconnexions et roaming** : elle prend en charge la gestion des relations avec les autres opérateurs.

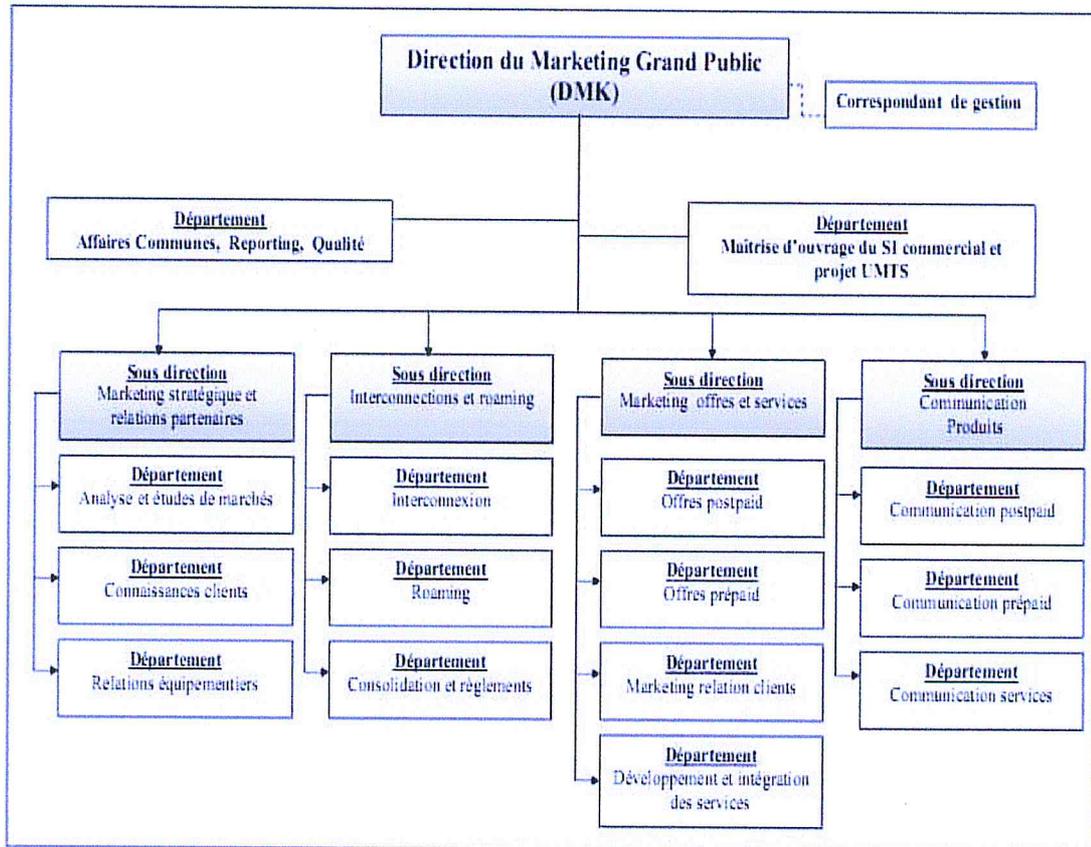


Figure 4.2. Organigramme général de la direction du marketing (DMK)

IV.4. Missions et objectifs de l'Entreprise :

+ Les Missions :

ATM Mobilis a pour mission principale d'offrir un service de téléphonie mobile de grande qualité, simple, efficace et accessible au plus grand nombre de personnes. Elle exerce son domaine d'activité sur l'ensemble du territoire national pour :

- ✓ Installer, développer, exploiter et gérer les réseaux de la téléphonie mobile ;
- ✓ Maintenir et monter des équipements de téléphonie mobile;
- ✓ Fournir des prestations et mettre à la disposition de ses clients des produits de la téléphonie mobile;
- ✓ Etablir, exploiter et gérer les interconnexions permettant à tout opérateur de réseaux de téléphone mobile national ou international d'assurer la communication entre les clients;
- ✓ Assurer en permanence la disponibilité de ses produits pour l'ensemble des clients sur tout le territoire;
- ✓ Exercer ses activités dans le respect de la concurrence;
- ✓ Offrir ses produits et services dans le mobile en provenance et à destination des pays étrangers dans le respect des règles définies par le règlement international dans le domaine d'activité.

+ Les Objectifs :

Chaque entreprise doit définir ses objectifs à tous les niveaux (objectifs commerciaux, financiers, marketing, etc.). Elle doit se fixer des buts pour l'ensemble de l'entreprise, puis pour chaque domaine d'activité stratégique. Les objectifs d'ATM Mobilis sont :

- ✓ Développer le parc d'abonnés et acquérir de nouvelles parts de marché ;
- ✓ Améliorer la couverture radio et la maintenir opérationnelle ;
- ✓ Introduire les nouvelles technologies et devenir leader sur le marché.
- ✓ Devenir l'opérateur multimédia par excellence ;
- ✓ Déploiement du service EDGE
- ✓ Extension du réseau UMTS (3G)

Conception et Réalisation d'un Système décisionnel pour le marketing

- ✓ Devenir opérateur corporatif ;
- ✓ Déploiement des services MVPN, VPN
- ✓ Développer le réseau commercial ;
- ✓ Employer une démarche marketing innovante et une politique de communication efficace ;

Afin de concrétiser ses ambitions, atteindre ses buts et réaliser ses objectifs, ATM Mobilis s'engage envers sa clientèle à :

- ✍ Proposer des offres simples, claires et transparentes qui répondent aux attentes des consommateurs;
- ✍ Améliorer constamment ses produits, ses services et sa technologie;
- ✍ Atteindre tous les segments de la population et rendre le téléphone mobile abordable à ceux au revenu moyen et qui représentent la majorité de la population;
- ✍ Mettre à la disposition du client un réseau de haute qualité;
- ✍ Déployer son réseau uniformément dans les différentes régions du pays;
- ✍ Acheminer dans les meilleures conditions tous les appels des clients quelque soit la destination demandée ;
- ✍ Répondre dans les meilleurs délais à toute réclamation ;
- ✍ Rester en permanence à l'écoute du client dans le but de passer rapidement d'un abonné administré à une véritable relation client ;
- ✍ Essayer de créer un lien solide avec les clients par les centres de services, le centre d'appel et par son site Web ;
- ✍ Entreprendre différentes actions promotionnelles et publicitaires afin de séduire les clients potentiels ;
- ✍ Innover constamment et faire converger son réseau et ses services pour mieux répondre aux besoins de la société algérienne de l'information de demain ;
- ✍ Assurer la diffusion d'une image positive de l'entreprise à travers les médias (slogans, messages) pour établir une relation de fidélité avec les clients.

CHAPITRE V.

Conception et implémentation du système

CHAPITRE V. Conception et implémentation du système :

V.1. Introduction :

Ce chapitre est le cœur de notre travail. Dans ce chapitre on présentera la conception et la modélisation du système. On commencera par la méthode qu'on a suit pour modéliser le système, puis la schématisation des différents composants du système.

La modélisation va toucher le niveau conceptuel et logique. On a laissé la modélisation physique pour le chapitre suivant.

V.2. Méthode choisie :

Dans la partie I, chapitre 3, on a présenté un ensemble de méthodes de conception d'un SID. On a remarqué que presque toutes les méthodes ne couvrent pas toutes les spécificités d'un SID (La plupart des méthodes se focalisent sur la modélisation dimensionnelle et négligent la conception de l'ETL malgré son importance). D'autres méthodes n'utilisent pas un formalisme de modélisation, par exemple « *kimball et al* » (7). Si c'est le cas, l'enchaînement des étapes n'est pas bien détaillé.

Pour la conception de notre système, on a opté pour la méthode **DWEP (Data Warehouse Engineering Process)** de « *Luján Mora* » (6) pour plusieurs raisons parmi elles :

- + La méthode touche presque toutes les spécificités d'un SID (conception des sources, de L'ETL, de DW ...).
- + La méthode utilise 2 standards, **UML** comme Langage de modélisation, **UP (Unified Process)** comme démarche de conduite du projet.
- + La modélisation se fait en 3 niveau d'abstraction (conceptuel, logique, Physique) et selon divers niveaux de détails qui permet d'éviter la difficulté de la modélisation si le système est très complexe (boucaux de faits et de dimensions).

Malgré ces avantages la méthode a quelques lacunes :

- + La définition des besoins n'est pas présentée par l'auteur.
- + L'auteur n'explique pas comment les diagrammes proposés sont intégrés dans la démarche UP.

- + Pas de modélisation pour l'application utilisateur.
Pour résoudre ces problèmes on a :
 - ✓ Pour la définition des besoins, on a utilisé une extension du diagramme de cas d'utilisation pour modéliser les besoins des utilisateurs. On a devisé les besoins en 2 types :
 - Besoins en termes de données : ce type de use case répond à la question : **Qui** besoins **quoi, quand** et **combien** de fois ?
 - Besoins en termes d'interface de restitution (Application utilisateur) : ce type de use case répond à la question : **comment** l'information sera exploiter ?
 - ✓ Pour l'enchaînement des diagrammes (séquencement des étapes) on a proposés une démarche qui combine le cycle de vie dimensionnel proposé par *Ralph Kimball* (7) et l'adaptation d'UP proposée par *Luján Mora* (6). (voir la figure 5.5).
 - ✓ Pour la conception l'interface de restitution on a basé sur l'extension de use case qu'on a proposé pour développer l'application utilisateur (l'application de restitution sera développer comme n'importe quelle application classique (7)).

V.2.1. Présentation de la méthode DWEP (DW Engineering Process) :

Cette méthode utilise UML comme langage de modélisation, *Luján Mora* (6) propose une extension d'UML (**UML profile**) pour répondre à la spécificité d'un SID, cette extension se fait à travers des stéréotypes (icônes), valeurs ciblés et des contraintes.

On a vu en chapitre I que l'architecture d'un SID est présentée en plusieurs couches. Pour modéliser les différents composants d'un SID, L'auteur propose 15 diagrammes, ces diagrammes sont organisés en 5 phases « **stages** » (**Sources, Intégration, DW, Customisation, Client**), chaque phase est analysé en 3 niveaux d'abstraction (**Conceptuel, Logique, Physique**) (voir tableau 5.1).

Conception et Réalisation d'un Système décisionnel pour le marketing

	Source(S)	Integration	DW	customization	Client
Conceptuel	SCS Class diagram Standard UML	DM Class diagram Data Mapping Profile	DWCS Class diagram Standard UML Multidimen- sional Profile	DM Class diagram Data Mapping Profile	CCS Class diagram Standard UML Multidimen- sional Profile
Logic	SLS Class diagram Different data mod- eling profile	ETL Process Class diagram ETL profile	DWLS Class diagram Different data modeling profile	Exporting Pro- cess Class diagram ETL Profile	CLS Class diagram Different data modeling pro- file
Physical	SPS Comp. & deploy. Diagrams Database Deploy- ment profile	Transportation Diagram Deployment diagram Database Deploy baseDeploy- ment profile	DWPS Comp. & de- ploy. Diagrams Database De- ployment profile	Transportation Diagram Deployment dia- gram Database Deploy baseDeploy- ment profile	CPS Comp. & de- ploy. Diagrams Database De- ployment pro- file

Tableau 5.1. Extension d'UML par 15 diagrammes pour les SID (6)

V.2.2. La démarche proposée :

Comme on a cité précédemment, pour modéliser notre système on va utiliser une démarche qui combine entre la démarche et les diagrammes proposés par *Luján Mora* (voir tableau 5.1 et figure 5.1) plus le cycle de vie multidimensionnelle proposé par *Ralph Kimball* (voir la figure 5.2). Cette démarche est illustrée par la figure 5.4.

Sachant qu'on a adapté le processus unifié (UP) pour appliquer les phases (workflows) de la démarche proposée de façon itérative et incrémentale (Figure 5.3).

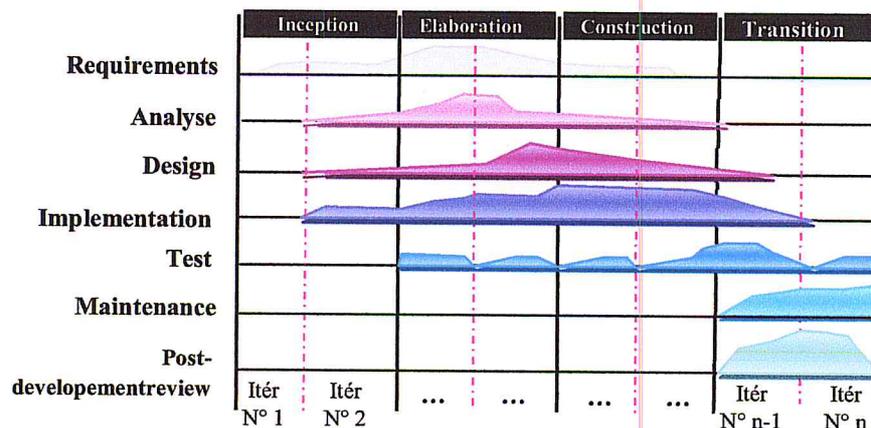


Figure 5.1. Le processus Unifier (UP) pour le DWEP (6)

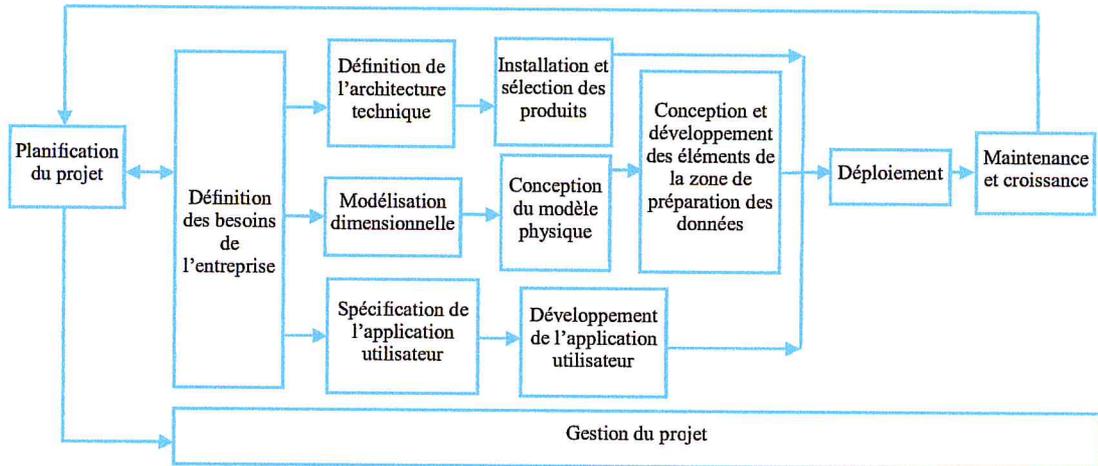


Figure 5.2. Le cycle de vie multidimensionnel de Ralph Kimball (7)

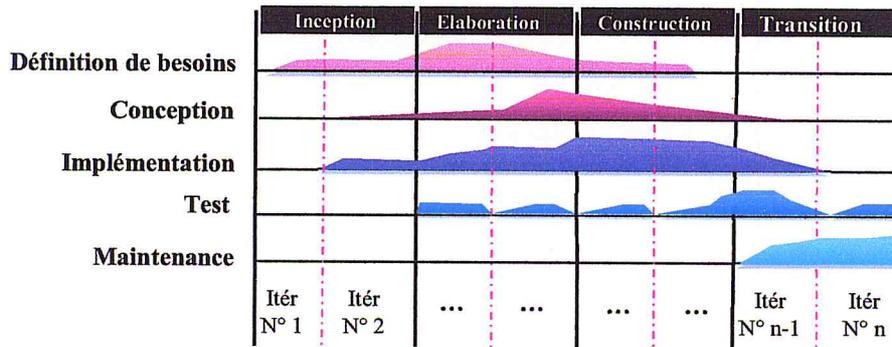


Figure 5.3. Adaptation de l'UP pour la démarche proposée

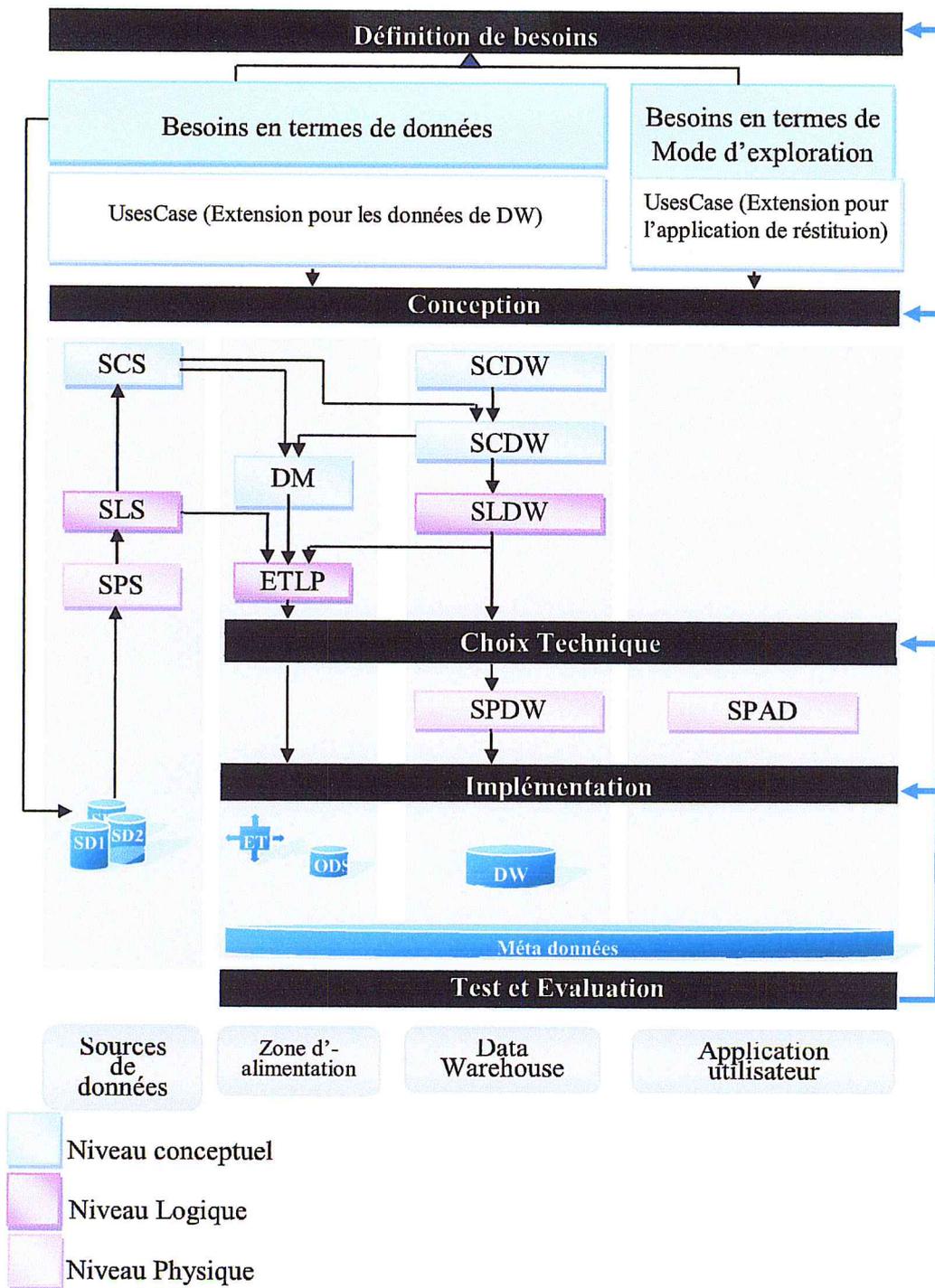


Figure 5.4. La démarche proposée

V.3. Conception du système :

V.3.1. Définition des besoins :

Les chances de succès d'un SID se trouvent considérablement accrues par la bonne compréhension des utilisateurs et de leurs besoins. L'approche utilisée pour identifier les besoins analytiques diffère de manière significative de la traditionnelle analyse des besoins pilotée par les données. Ces besoins constituent le point de départ des trajectoires parallèles pour la construction de système.

Cependant, selon une étude faite par IBM, 80% des projets décisionnels ne parviennent pas à satisfaire les besoins des utilisateurs et 40% d'entre eux échouent à faciliter la prise de décision. Pour cela les récentes propositions de méthodes liées aux SID définissent leur schéma à partir des besoins des utilisateurs et des systèmes sources (Approche mixte).

Pour recenser nos besoins on a utilisé la technique d'interview afin de dégager les exigences analytiques des futurs utilisateurs. Cette technique parvient insuffisante car les utilisateurs finaux ne savent pas bien toutes les fonctionnalités offertes par un SID, pour cela on a prendre l'initiative par la proposition des besoins (fonctionnalités) inspirés des travaux faits dans le domaine de marketing décisionnel et CRM (Customer Relationship Management) analytique. Dans ce qui suit on présentera Ces besoins par des tirées puis ils seront modélisés par une extension (adaptation) de diagramme de cas d'utilisation.

❖ Répartition des abonnés (Part du marché) :

➤ Effectif des abonnés par :

→ Date «Mois» → Trimestre → Semestre → Année

→ Démographie

↳ Sexe (homme, femme)

↳ Âge (année) → Tranche

↳ Adresse (commune → Wilaya → Région)

↳ Secteur d'activité

→ Offre (Gosto, mobicart ...) → Classe (poste paid, prepaid, mixte)

- Nouveaux abonnements (Nombre de nouveaux abonnés) par :
 - Démographie
 - Offre
 - Période «Mois» → Trimestre → Semestre → Année
 - Point de vente
- Pénétration (Taux de pénétration au marché par rapport au nombre d'habitant) par :
 - Sexe
 - Tranche d'âge
 - Adresse
 - Date «Mois»
- Etat de marché (Nombre d'abonnés) par :
 - Opérateur nationale (Mobilis, Djezzy, Nedjma)
 - Date «Mois»
- Résiliation (nombre des abonnés désactivés)
 - Démographie
 - Période «Mois» → Trimestre → Semestre → Année
 - Offre
- ❖ **Consommation (comportement des abonnés) :**
 - Rentabilité (Revenu total, Revenu moyen par abonné)
 - Type_Trafic (Appel, Voix, SMS, MMS ...)
 - Démographie
 - Offre
 - Période «jour» → Mois → Trimestre → Semestre → Année
 - Opérateur

Conception et Réalisation d'un Système décisionnel pour le marketing

- Nombre des appels (Nombre des appels entrants, Nombre des appels sortants, Nombre moyenne des appels entrants, Nombre moyenne des appels sortants)
 - Durée ($d < 1$, $1 < d < 2$, $2 < d < 5$, $5 < d < 10$, $10 < d \dots$)
 - Démographie
 - Offre
 - Opérateur (Opérateur de télécommunication cible)
 - ↳ Opérateur nationale
 - ↳ Opérateur internationale → pays
 - Période «Heure» → jour → Mois → Trimestre → Semestre → Année

- Durée des appels (Durée des appels entrants, Durée des appels sortants, Durée moyenne des appels entrants, Durée moyenne des appels sortants, Durée moyenne des appels entrants par abonné, Durée moyenne des appels sortants par abonné) :
 - Démographie
 - Offre
 - Période «jour»
 - Opérateur

- SMS (Nombre des SMS envoyées, Nombre des SMS reçues, Nombre moyenne des SMS envoyées, Nombre moyenne des SMS reçues)
 - Démographie
 - Offre
 - Opérateur
 - Période «Heure» → jour → Mois → Trimestre → Semestre → Année

- MMS (Nombre des MMS envoyées, Nombre des MMS reçues, Nombre moyenne des MMS envoyées, Nombre moyenne des MMS reçues)
 - Démographie
 - Offre
 - Opérateur

Conception et Réalisation d'un Système décisionnel pour le marketing

- Période «Heure» → jour → Mois → Trimestre → Semestre → Année
- Taille

- Rechargement (Nombres des opérations de rechargement, Valeur totale rechargée)
 - Type de rechargement (Carte, Aresilli, ...)
 - Démographie
 - Offre
 - Valeur rechargée (100DA, 500DA ...)
 - Période «jour» → Mois → Trimestre → Semestre → Année

+ Diagramme de Cas d'utilisations (pour les besoins en termes de données) :

On a utilisé une adaptation du diagramme de cas d'utilisation pour modéliser les besoins des utilisateurs en termes de données, sur cette adaptation les besoins d'analyses sont exprimés par des usages stéréotypés (Fait et Dim), une note est attachée pour le lien entre les acteurs et le use case (Fait) pour spécifier des dimensions utilisées par cet utilisateur. Le lien entre le use case (Fait) et (Dim) est de type « Extend ». On peut associer la priorité et la fréquence d'utilisation au lien entre l'acteur et le use case. Dans notre cas on a plusieurs utilisateurs avec le même profil (les mêmes fonctionnalités),

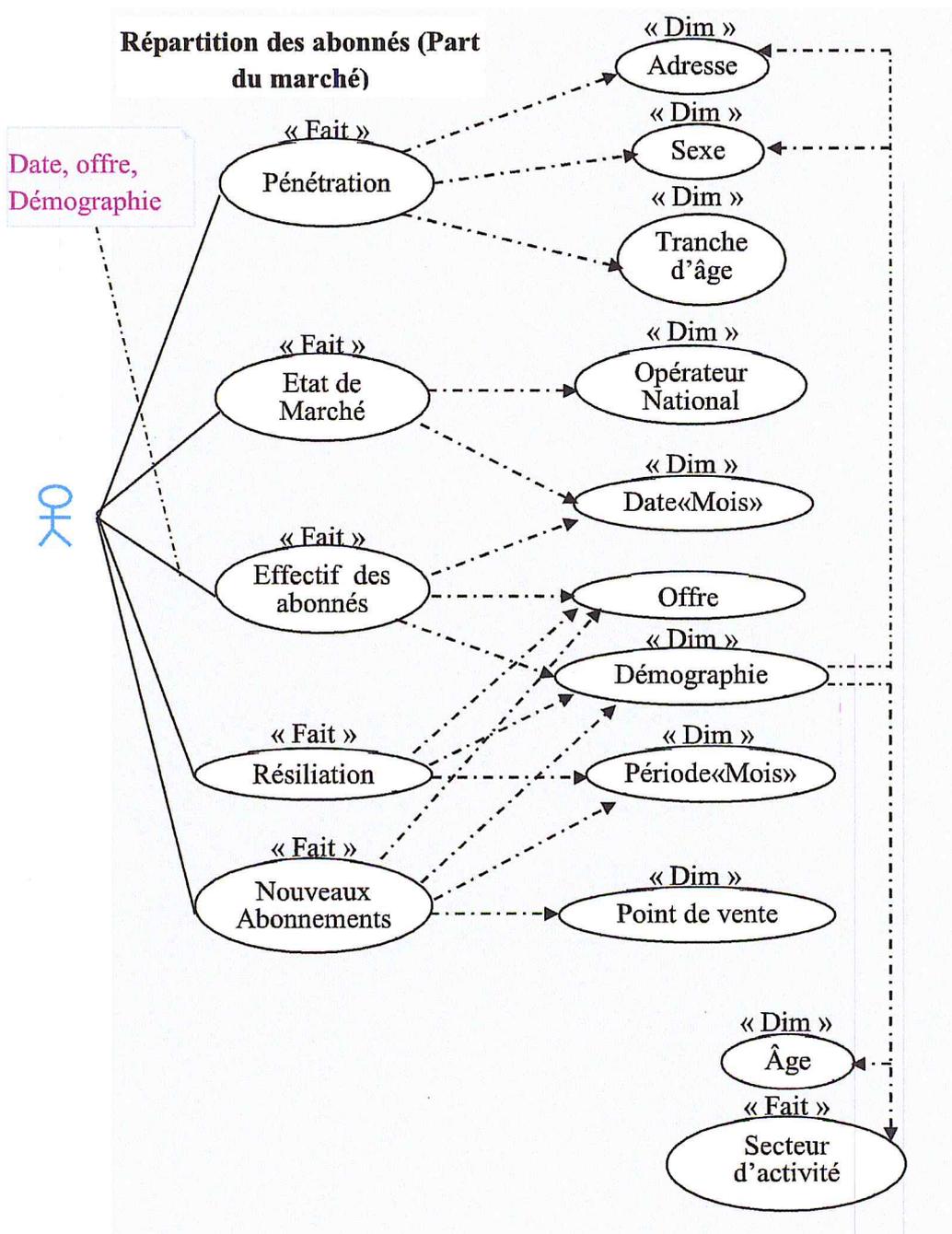


Figure 5.5. Diagramme de Cas d'utilisations Répartition des abonnés (Part du marché)

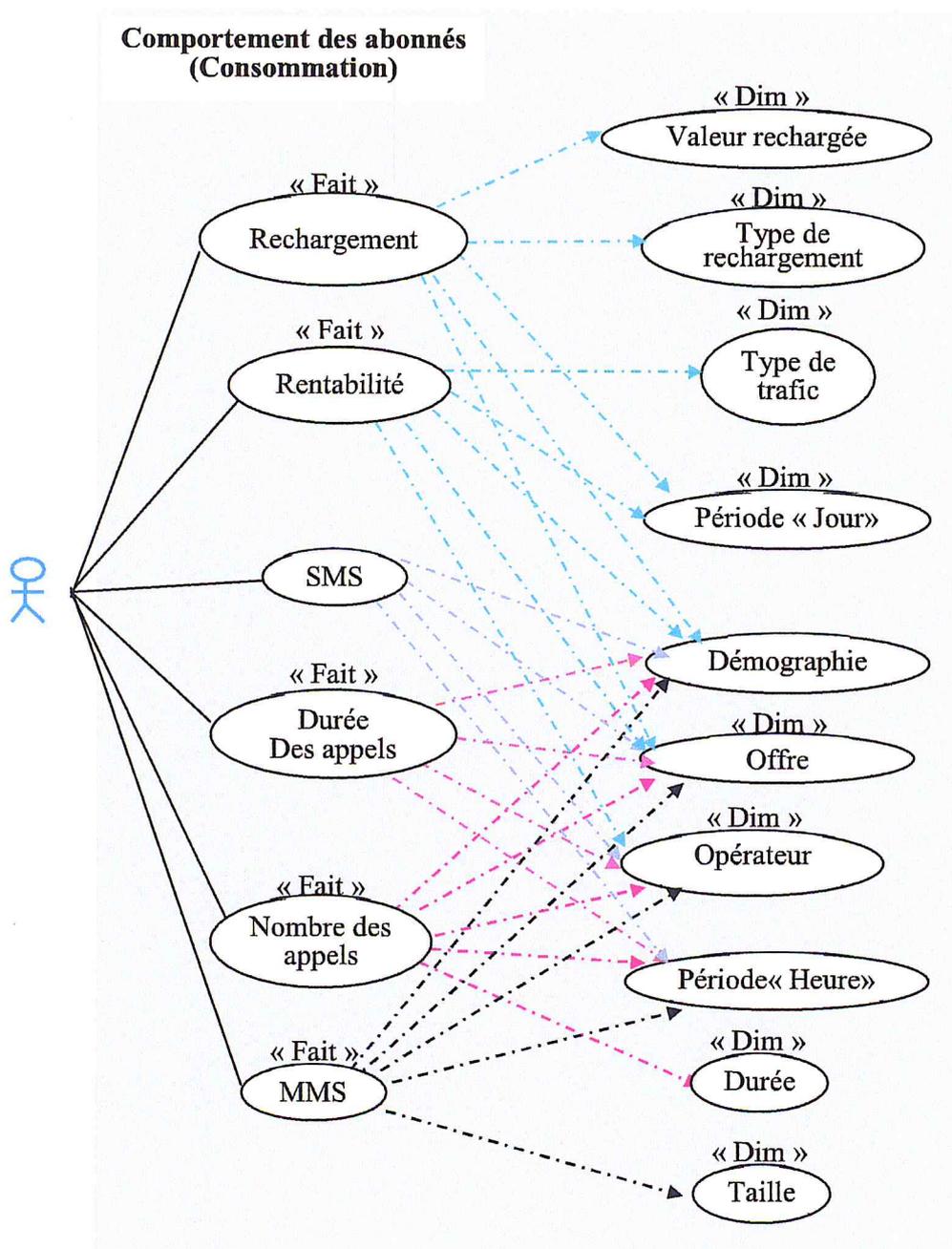


Figure 5.6. Diagramme de Cas d'utilisations Comportement des abonnés (Consommation)

+ **Diagramme de Cas d'utilisations de l'application de restitution:**

Pour décrire les fonctionnalités de l'application de restitution (consommation/ Etude comportementale) on a utilisé le diagramme de cas d'utilisation adapté pour montrer le type d'affichage. Sur ce diagramme les use case sont stéréotypés pour décrire le type de l'interface de restitution. Par exemple «KPI» (Key Performance Indicator) décrit que les données sont présentées sous forme de tableaux de bord qui donne une vue de haut niveau. «Mappe» pour la fonctionnalité de l'analyse géographique. «OLAP» pour l'analyse multidimensionnelle avec la possibilité de Drille-Up, Slice ... les liens entre les use case sont de type « extend » ou l'utilisateur aller d'une vision globale avec des KPI puis il prolonge dans le détail avec des fonctionnalités d'analyse comme l'OLAP.

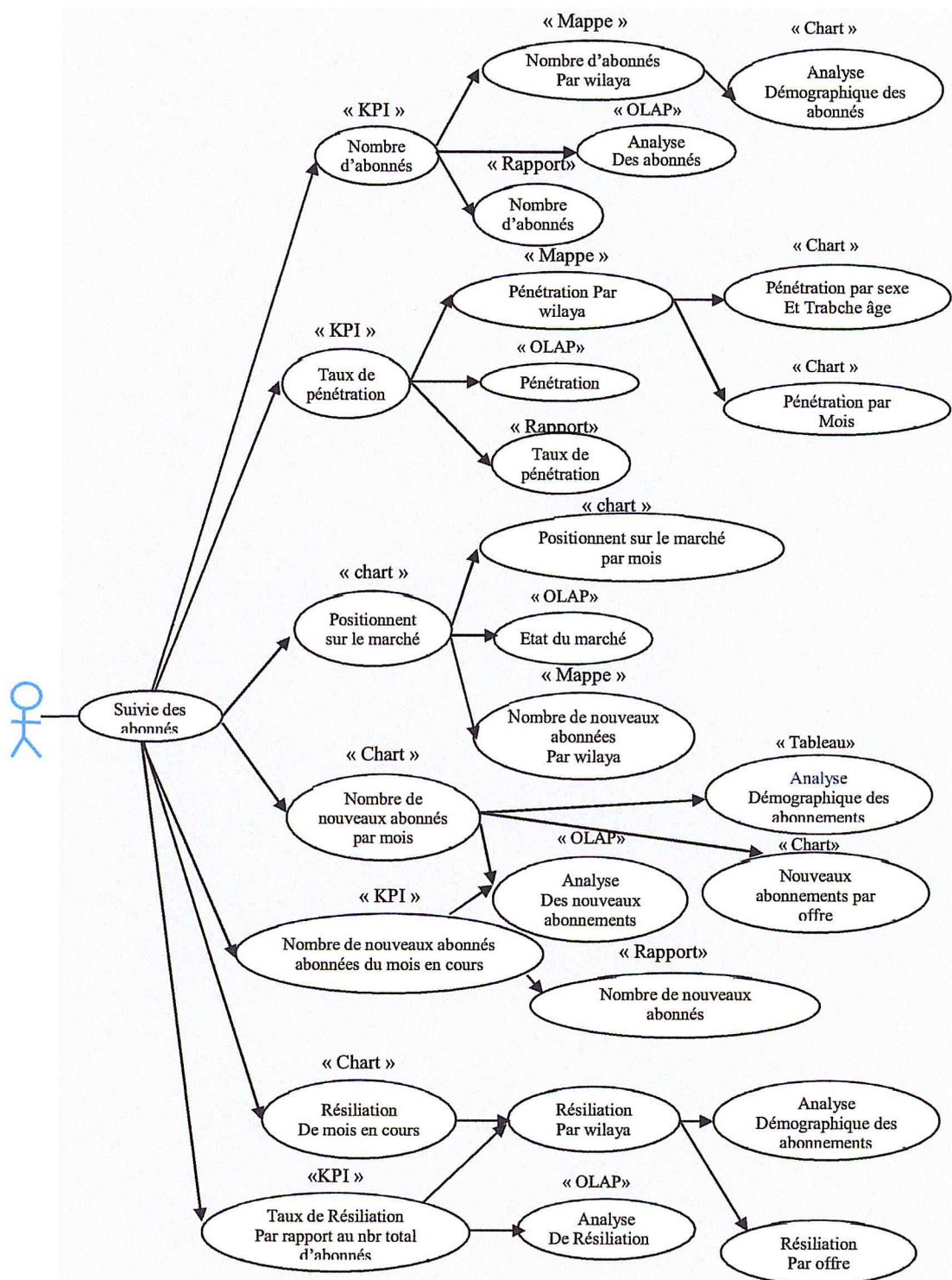


Figure 5.7. Cas d'utilisation de l'application de restitution(Suivre des abonnés)

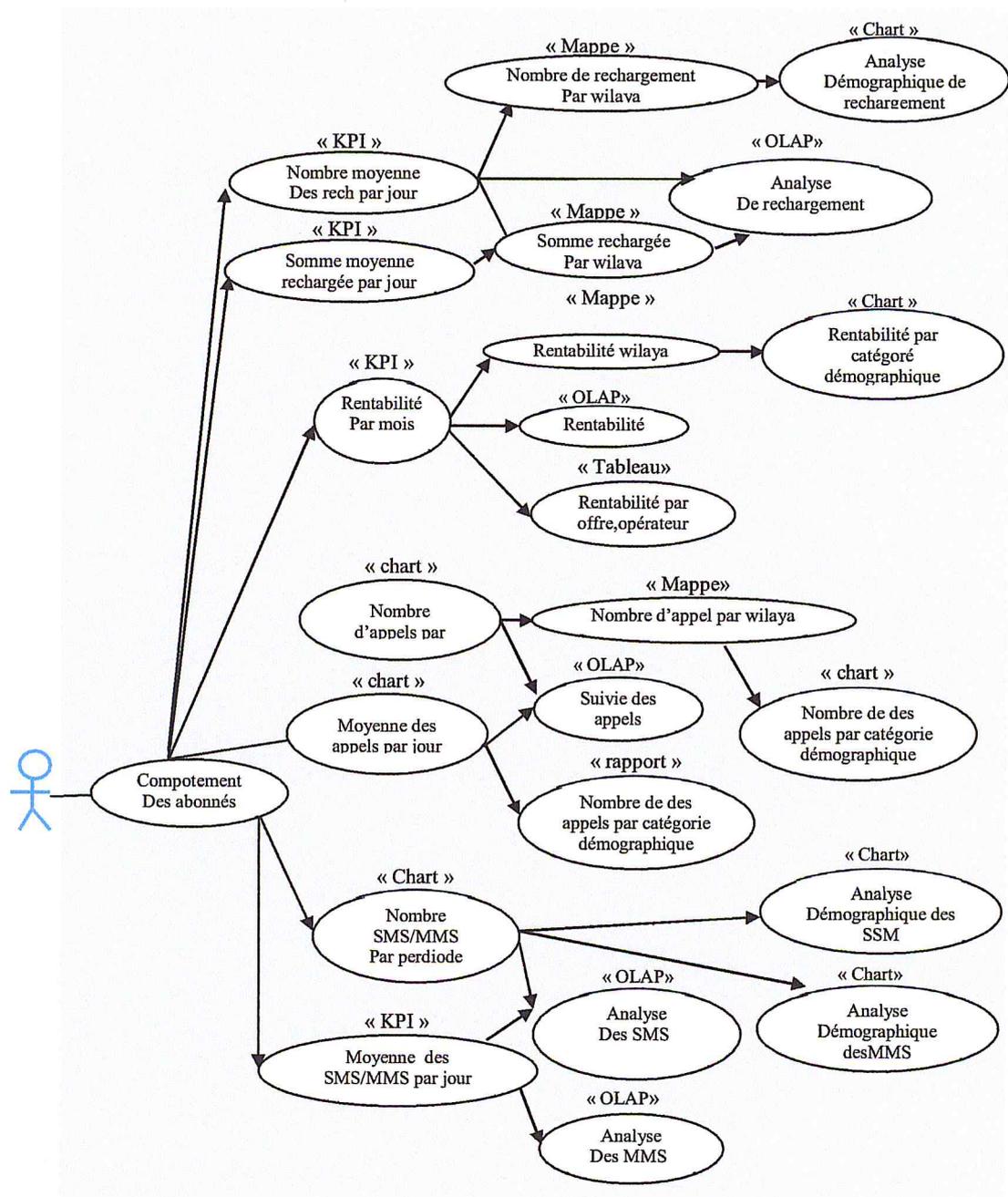


Figure 5.8. Cas d'utilisation de l'application de restitution(Consommation)

V.3.2. Modélisation du système au niveau conceptuel :

Au niveau de la modélisation conceptuelle, les entités importantes et les relations entre elles sont présentées indépendamment d'aucune modélisation ou technologie particulière, la modélisation conceptuelle est fermée sur l'espace du problème (monde réel)

Dans la démarche qu'on a suivie, La modélisation conceptuelle touche tous les composants d'un SID (sources de données, DW, ETL ...).

V.3.2.1. Schéma conceptuel des sources (SCS) :

Après la définition des besoins, 2 tâches parallèles sont déclenchées (figure 5.3). L'identification des sources nécessaires pour la satisfaction de ces besoins et la modélisation conceptuelle du DW.

Pour les sources de données on a suivie une sorte de « reverse engineering » (physique, logique, conceptuel. voir la figure 5.3) afin d'extraire et modéliser les sources.

Après des interviews avec les gens du service technique, on a recensé 4 sources de données principaux (DWS, DData « Demographics Data », MINSAT, Système Poste Paide) plus d'autres sources externes par exemple 2 fichiers « xls », un pour les statistiques démographiques sur la population algérienne et l'autre sur les indicateurs des opérateurs de téléphonie (mobile et fixe) au niveau international.

La figure 5.5 représente le modèle conceptuel des sources (seulement les sources internes). Pour une lisibilité claire et en raison du nombre important des tables (par exemple le DWS contient plus de 50 tables). Sur ce modèle on présente seulement les données utiles pour l'alimentation du DW.

Il faut noter que les tables du système Epix (gère les abonnés poste paid) qu'on a utilisé, ont la même structure que celles de DWS, pour cela dans qui suit on présentera seulement les sources de DWS seulement.

+ Description des systèmes sources :

❖ Demographic DATA (DDATA)

- Serveur : MySQL
- Système : Microsoft windows
- Description : Les informations personnelles des abonnés
- Les tables :
 - Subscriber : Table des abonnés

❖ MINSAT (Mobile IN Service Administration Tool)

- Serveur : SyBase
- Description : Les informations d'états d'abonnés (activation, résiliation ...)
- Les informations concernant notre futur système :
 - Date d'activation
 - Historique de recharge
 - Classe d'abonné
 - Etat (actif, bloqué)
 - Crédit

❖ DWS

- Serveur: Oracle.
- Système: UNIX
- Description : Les informations de trafic effectué sur le réseau de Mobilis des abonnés post paid
- Les tables :
 - air_account_event : Les informations sur les rechargements
 - sdp_account_event : Les informations sur les activations
 - msc_event : Les informations sur le trafic effectué
 - usage_events : historique d'utilisation

Conception et Réalisation d'un Système décisionnel pour le marketing

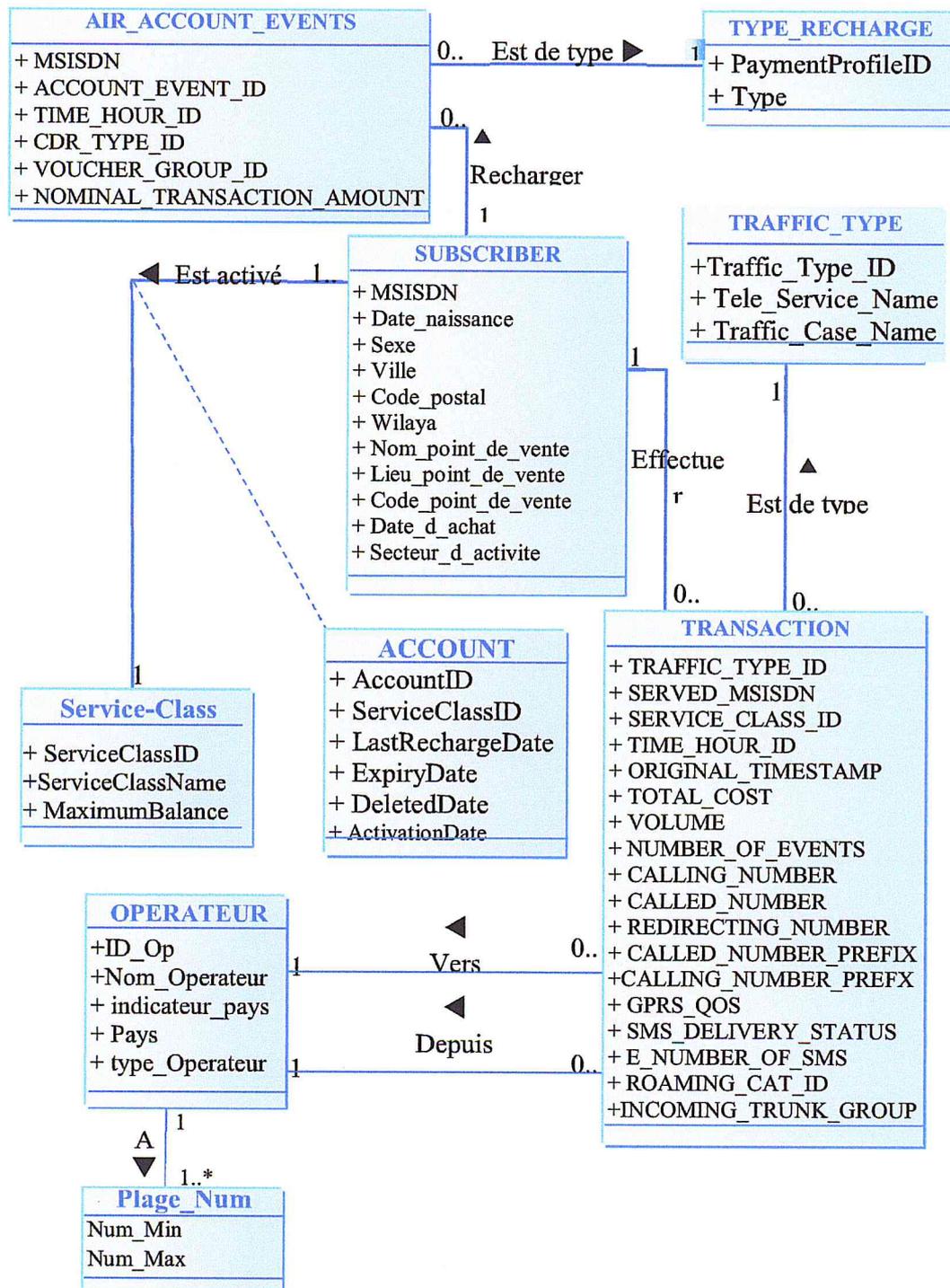


Figure 5.9. Modèle (Schéma) Conceptuel des Sources de données (SCS)

Conception et Réalisation d'un Système décisionnel pour le marketing

Nom de la table	Nom de l'attribut	Description
Subscriber	MSISDN	N° téléphone d'un abonné
	Date_de_Naissance	Date de naissance d'un abonné
	Sexe	Sexe d'un abonné
	Ville	Ville d'un abonné
	Wilaya	Wilaya d'un abonné
	Code postal	Code postal d'un abonné
	Nom_point_de_vente	Nom de point de vente
	Lieu_point_de_vente	Adresse de point vente
	Code_point_de_vente	Code d point de vente
	Date_d_achat	Date d'achat de la puce
	Secteur_d_activité	Secteur d'activité d'un abonné
Transaction	TRAFFIC_TYPE_ID	Type de la transaction (appel, SMS ...)
	SERVED_MSISDN	Le N° Tel d'un abonné
	SERVICE_CLASS_ID	Classe d'abonné (Gosto, Mobicontrol)
	TIME_HOUR_ID	Le temps de la transaction
	ORIGINAL_TIMESTAMP	Le temps GMT de la transaction
	TOTAL_COST	Cout total de la transaction (DA)
	VOLUME	Taille de l'MMS
	CALLING_NUMBER	N° de l'appelé
	CALLED_NUMBER	N° de l'appelant
Trafic type	TRAFFIC_TYPE_ID	L'identifiant de la table
	Tele_Service_Name + Tele_Case_Name	Intitulé de type de trafic
		Description de type de trafic
AIR_ACCOUNT_EVENTS	MSISDN	N° de téléphone d'un abonné
	ACCOUNT_EVENT_ID	Identifiant du compte d'un abonné
	TIME_HOUR_ID	Temps de rechargement
	VOUCHER_GROUP_ID	Identifiant de type d'abonné
	NOMINAL_TRANSACTION	Somme rechargée en DA
TYPE_RECHARGE	PaymentProfileID	Identifiant de type de rechargement
	Type	Intitulé de type de rechargement
Service Class	ServiceClassID	Identifiant d'une classe d'abonnés
	ServiceClassName	Intitulé d'une classe d'abonnés
Opérateur	ID_Op	Indicateur d'un opérateur de téléphonie
	Nom de l'opérateur	Nom d'un opérateur de téléphonie
	Indicateur_pays	Indicateur du pays d'un opérateur
	Pays	Nom du pays
ACCOUNT	Type_Operateur	Type de l'opérateur (Fixe, Mobile)
	AccountID	Identifiant d'un compte utilisateur
	ServiceClassID	Identifiant d'une classe d'abonnés
	LastRechargeDate	La date de dernière rechargement
	DeletedDate	Date de résiliation d'un abonné
	ActivationDate	Date d'activation d'une puce

Tableau 5.2. Description des attributs de system source

V.3.2.2. Schéma conceptuel du DW(DWCS) :

La conception d'un DW est fondée sur La modélisation multidimensionnelle MD, les premiers concepts de cette modélisation sont introduits par *Ralph Kimball* au niveau logique. Dernièrement, il ya des travaux proposent des notations graphiques pour la représentation des propriétés de la modélisation multidimensionnelle au niveau conceptuel. Cependant aucune d'elles n'est acceptée comme standard pour la modélisation des DW au niveau conceptuel.

Luján Mora propose une extension d'UML sous forme de (profil UML) dans le but d'être un standard pour la modélisation des DWs. Ce profil est défini par un ensemble de stéréotypes, valeurs ciblées et contraintes (stereotype, target values and constraints).

Pour modéliser le DW conceptuellement, *Luján Mora* propose un diagramme appelé **DWCS (DW Conceptual Schema)**. Ce diagramme est un extension du diagramme de classe ou les formalismes/éléments (Classe, Relation, Package, Attribut ...) sont stéréotypées pour exprimées les concepts de la MD (Fait, Dimension, niveau, mesure ...) (Figure 5.11 et Tableau 5.3). Le **DWCS** modélise un DW à différents niveaux de détail, ces niveaux de détails est exprimés à travers des packages (Figure 5.10). Cette technique permet de cacher la complexité d'un modèle et de mieux le comprendre. *Luján Mora* propose 3 niveaux de détail et donne un guidelines pour l'utilisation de cette technique (tableau 5.6).

✚ Les Niveaux :

- ✓ **Niveau 1 (Définition du modèle)** : un package (StarPackage) représente le Schéma en étoile pour un seule sujet d'analyse (Star Shcema) d'un modèle multidimensionnel. Une dépendance entre 2 packages à ce niveau indique que les modèles en étoile partagent au moins une dimension.
- ✓ **Niveau 2 (Définition du modèle en étoile)** : un package représente un fait (FactPachage) ou une dimension (DimensionPackage) d'un modèle en étoile. Une dépendance entre 2 packages de dimension à ce niveau indique que les packages partagent au moins un niveau d'une hiérarchie de dimension.

✓ **Niveau 3 (Définition de dimension/fait) :** un package est exploré en un ensemble de classes qui représente les niveaux de la hiérarchie (Base Class) dans un package de dimension (Dimension Class), ou le modèle en étoile en entier dans le cas d'un package de fait (Fact Class).

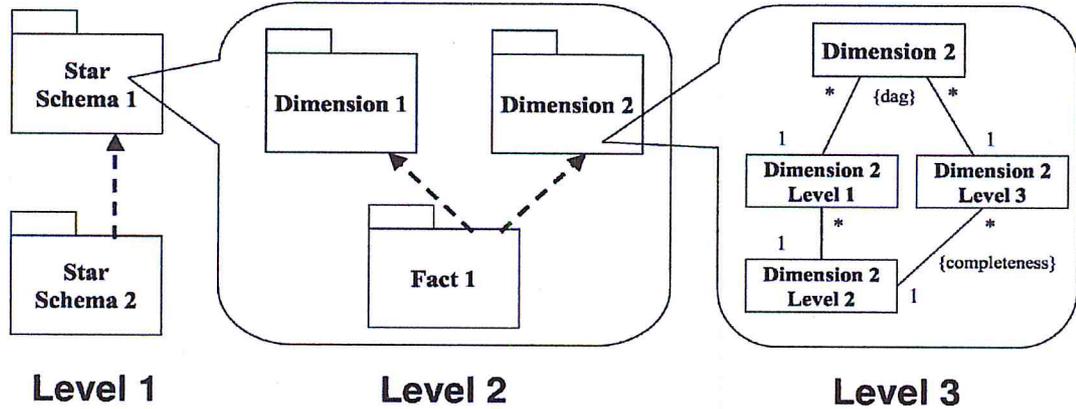


Figure 5.10. Les 3 niveaux du modèle DWCS utilisant des packages

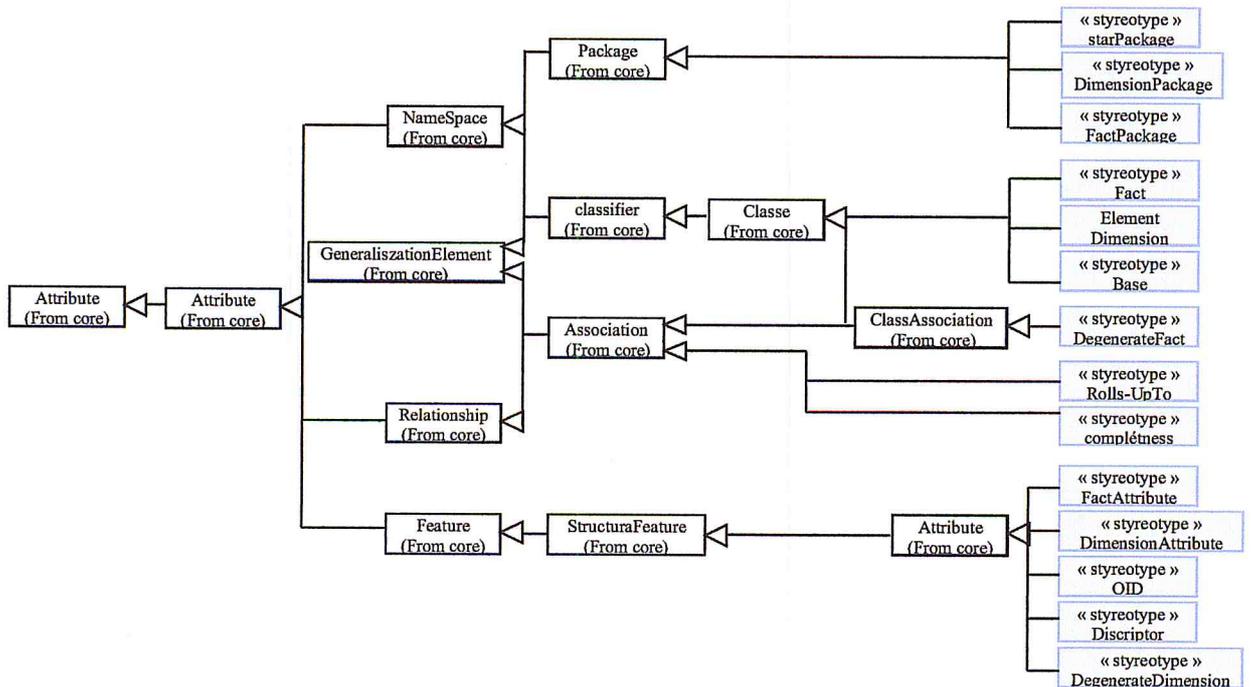


Figure 5.11. Extension d'UML par des stéréotypes

+ Les formalismes (Eléments) :

✓ Fait et Dimension :

Les faits et les dimensions sont représentés par les classes « **Fact** » et « **Dimension** », respectivement. Les classes « **Fact** » sont spécifiés comme classes composés au moyen de relations d'agrégation de n classes « **Dimension** ». **Dimensions et Niveau d'Hiérarchie de Classification.**

Les classes « **Dimension** » sont composées de *niveaux d'hiérarchie de classification* ; chaque niveau d'hiérarchie de classification est spécifier par une classe nommée « **Base** ». Une association (représenté par un stéréotype nommé « **Rolls-upTo** ») entre les classes « **Base** » spécifié la relation entre deux niveaux d'une hiérarchie de classification.

Une « **Dimension** » contient un niveau d'hiérarchie (ou dimension) premier unique nommé *niveau de dimension terminal*. Un chemin roll-up (ou chemin d'agrégation) est une séquence de niveaux des dimensions, qui commence par un niveau de dimension terminal (niveau de détail inférieur) et termine par un niveau de dimension implicite (non représentable graphiquement) qui représente tout les niveaux de dimension.

On représente une association « *Rolls-upTo* » : le rôle **R** représente la direction de rolls-up, et le rôle **D** représente la direction de drills-down.

Attribut :

Les classes «**Fact**» et «**Base**» peuvent avoir des attributs. Les classes «**Dimension**» n'ont pas des attributs, par se qu'elles représentent le concept de dimension, l'information sur une dimension est représenté dans la classe « **Base** » correspondante.

- Une classe « **Fact** » contient des attributs « **FactAttribute** » représentent les mesures
- Une classe « **Base** » contient trois types d'attributs :
 - **OID**: Toute « **Base** » peut avoir un attribut d'identification « **OID** ». Chaque « **Base** » contient un attribut.
 - **«Descriptor»** : Qui sera utilisé comme le libellé par défaut dans l'analyse de données
 - **«DimensionAttribute»** (optionnel): Les informations décrivent la dimension.

+ Stéréotypes :

✓ Stéréotypes de Package :

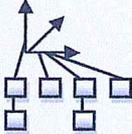
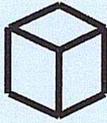
Nom	Description	Icône	Contraintes
Star Package	Représente un model en étoile de la modélisation dimensionnelle		<ul style="list-style-type: none"> ▪ Ne peut contenir que des (DimensionPackage) et/ou un seul (FactPackage). ▪ Ne peut importer depuis un autre StarPackage que des (Dimension Package). ▪ Pas de cycles de dépendance
Dimension-Package	Représente une dimension de la modélisation dimensionnelle		<ul style="list-style-type: none"> ▪ Un (DimensionPackage) ne peut contenir que des classes Base et une seule classe Dimension ▪ Pas de dépendance depuis un (DimensionPackage) vers un (FactPackage) (seulement vers d'autres DimensionPackage) ▪ Pas de cycles de dépendance
Fact Package	Représente une table de faits et ses dimensions de la modélisation dimensionnelle		<ul style="list-style-type: none"> ▪ Un (DimensionPackage) ne peut contenir que des classes Dimension, Base et une seule classe Fact ▪ Pas de dépendance depuis un (DimensionPackage) vers un FactPackage (seulement vers d'autres DimensionPackage) ▪ Pas de cycles de dépendance

Tableau 5.3. Les Stéréotypes des packages

✓ Stéréotypes D'attributs :

Nom	Description	Icône
FactAttribute	Représente un attribut d'une classe Fact	FA
OID	Représente l'attribut OID d'une classe Base	OID
Descriptor	Représente un attribut décrive une classe Base	D
DimensionAttribute	Représente un attribut d'une classe Base	DA

Tableau 5.4. Les stéréotypes des attributs

✓ Stéréotype des classes :

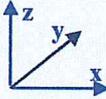
Nom	Description	Icone	Contraintes
Fact	Représente un fait dans un model dimensionnelle		<ul style="list-style-type: none"> ▪ Une classe Fact contienne des attributs FactAttribute ▪ Toute relation d'une Fact doit être une agrégation ▪ Une Fact ne peut en relation qu'avec des classes Dimensions
Dimension	Représente une dimension dans un model dimensionnelle		<ul style="list-style-type: none"> ▪ Ne peut contenir ni des attributs ni des méthodes ▪ Ne peut être relié qu'avec des classes Fact et/ou une seule classe Base ▪ Toute relation d'une Dimension avec une Fact doit être une agrégation vers la classe Fact
Base	Représente un niveau d'hierarchie de classification dans un model dimensionnelle		<ul style="list-style-type: none"> ▪ Les attributs d'une Base doivent être OID, Descriptor ou DimensionAttribute ▪ Une base doit avoir un attribut Descriptor ▪ Une base peut avoir un attribut OID ▪ Une base ne peut être reliée qu'avec des classes Dimension ou Base

Tableau 5.5. Les stéréotypes des classes

+ **Guidelines de modélisation avec le DWCS :**

o	Niveau	Etape
0a		A la fin du processus de conception, le modèle MD sera devisé en trois niveaux : définition du modèle, définition de modèle en étoile, et définition fait/dimension
0b		Définir les faits et dimensions et noter les dimensions partagées et les dimensions qui partagent des niveaux d'hierarchie
1	1	Dessiner un package pour chaque modèle en étoile i.e. chaque fait considéré
2a	1	Décider quel modèle en étoile contiendra la définition des dimensions partagées, et selon cette décision dessiner les dépendances correspondantes
2b	1	Regrouper les dimensions partagées pour minimiser le nombre de dépendances
2c	1	Eliminer les cycles dans la structure de dépendance
3	2	Dessiner un package pour le fait et un package pour chaque dimension du chaque modèle en étoile
4a	2	Dessiner une dépendance du package des faits vers ceux des dimensions
4b	2	Ne jamais dessiner de dépendance d'un package de dimension vers celui des faits
5	2	Ne pas définir une dimension plus qu'une fois (utiliser import)
6	2	Dessiner une dépendance entre les packages de dimension pour indiquer qu'elles font partie d'une même hiérarchie de dimensions
7	3	Dans un package de dimension, dessiner une classe pour la class dimension et une classe pour chaque niveau d'hierarchie de classification (les classes des Bases)
8	3	Dans un package de fait, dessiner une classe pour la class fait et importer les classes de dimension avec les niveaux d'hierarchie correspondants
9	3	Dans un package de dimension, si une dépendance depuis le package courant a été défini au niveau 2, importer le niveau de l'hierarchie partagé correspondant
10	3	Dans un package de dimension, Si un niveau d'hierarchie est importé d'un autre package, il n'est pas nécessaire d'importer tout les niveaux

Tableau 5.6. Guidelines de modélisation avec le DWCS

V.3.2.2.2. Schéma conceptuel du DW (Niveau 1) :

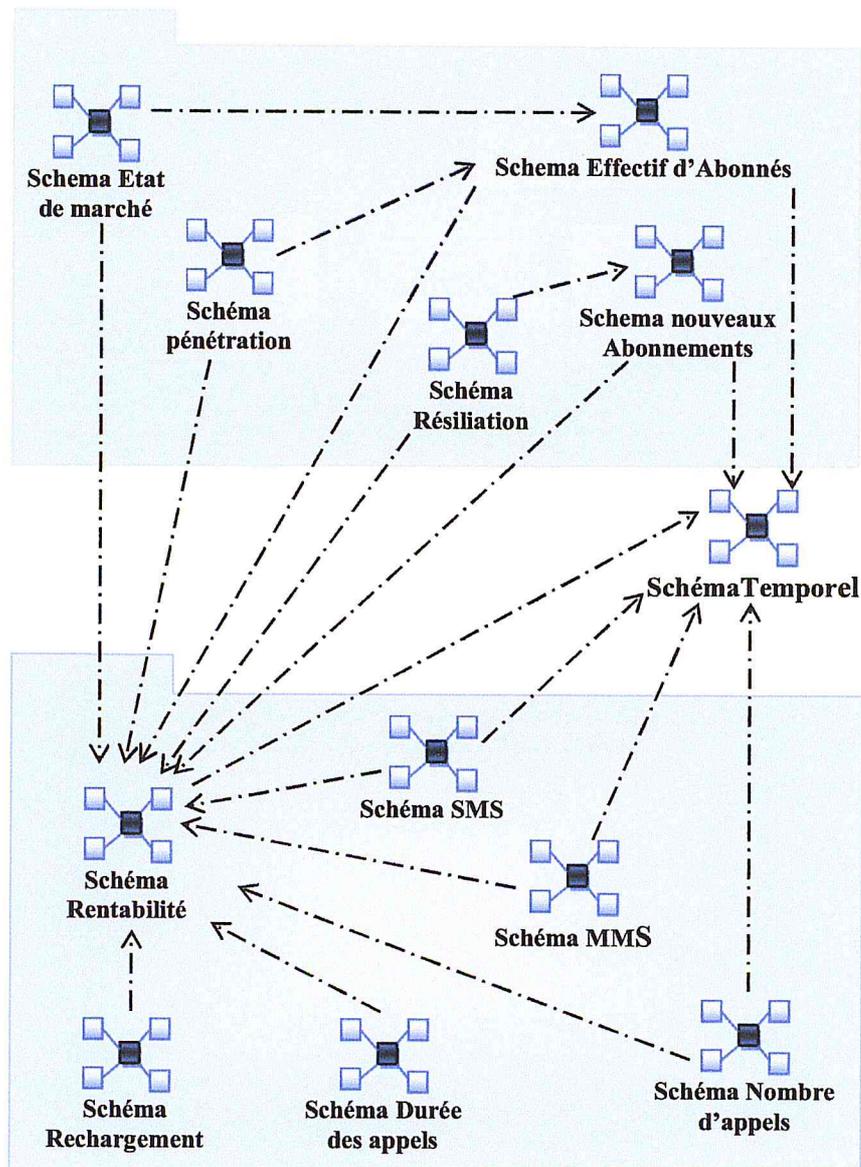


Figure 5.12. Niveau 1 de DWCS

V.3.2.2.3. Schéma conceptuel du DW (Niveau 2)

+ Schéma Temporel :



Figure 5.13. Niveau 2 : le Schéma Temporel

+ Schéma Effectif d'Abonnés :

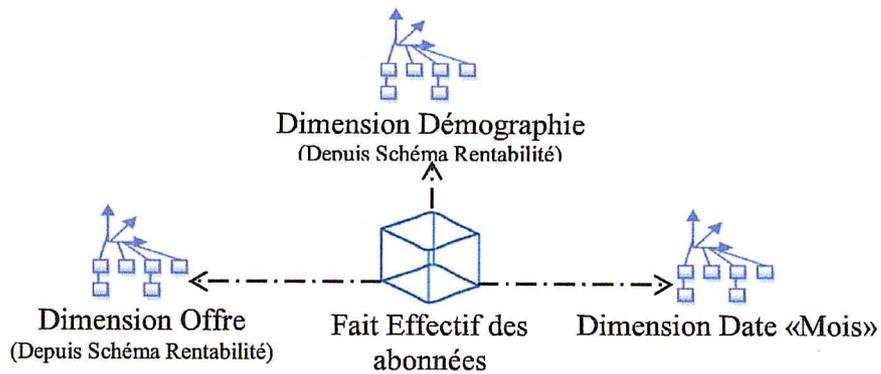


Figure 5.14. Niveau 2 : le Schéma Effectifs des abonnés

+ Schéma Nouveaux Abonnements :

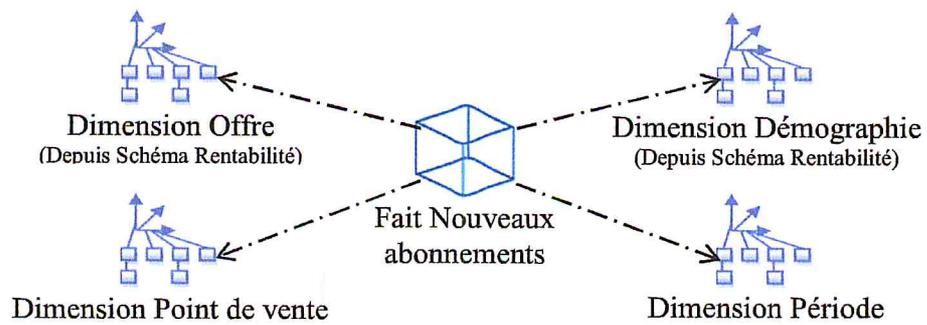


Figure 5.15. Niveau 2 : le Schéma Nouveaux abonnement

✚ Schéma Résiliation :

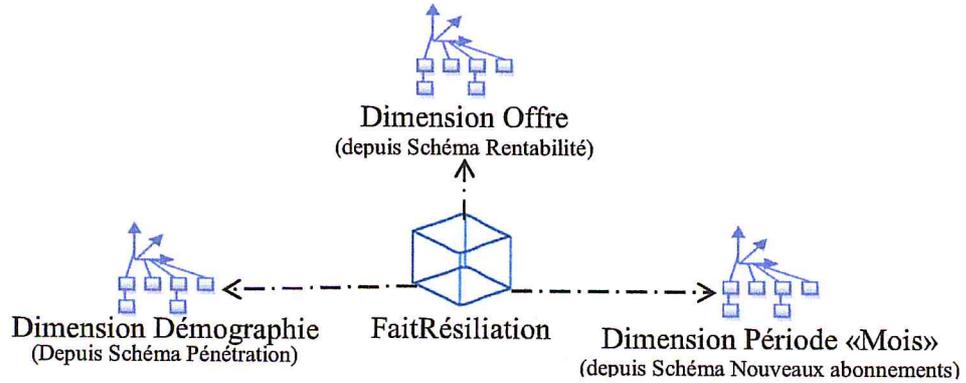


Figure 5.16. Niveau 2 : le Schéma Résiliation

✚ Schéma Pénétration :

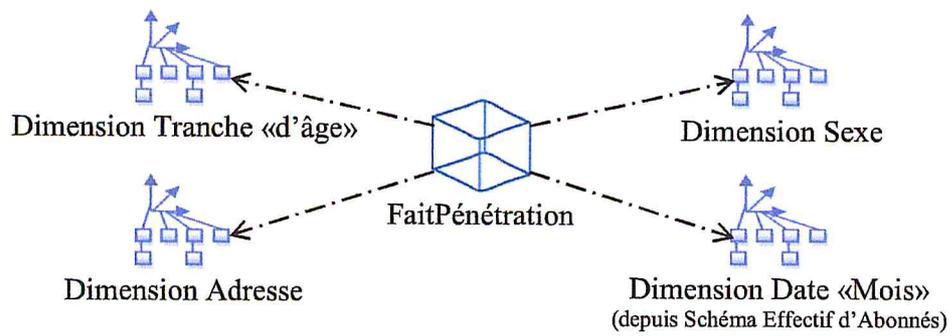


Figure 5.17. Niveau 2 : le Schéma Pénétration

✚ Schéma Rechargement :

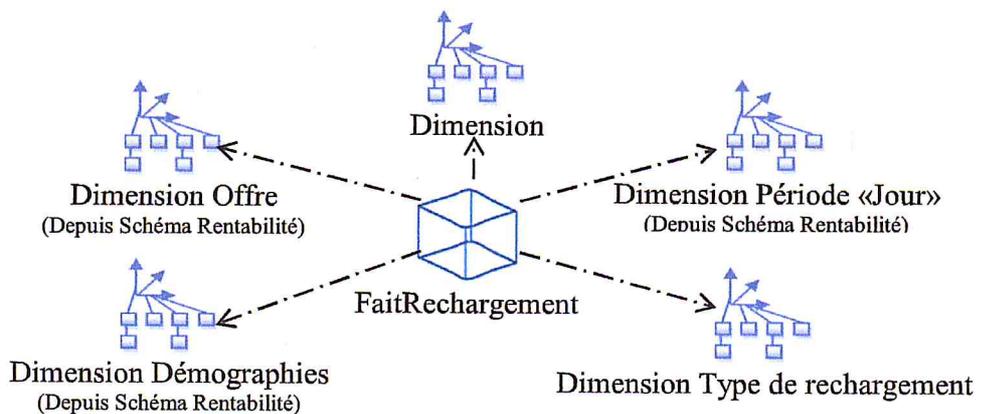


Figure 5.18. Niveau 2 : le Schéma Rechargement

+ Schéma Rentabilité :

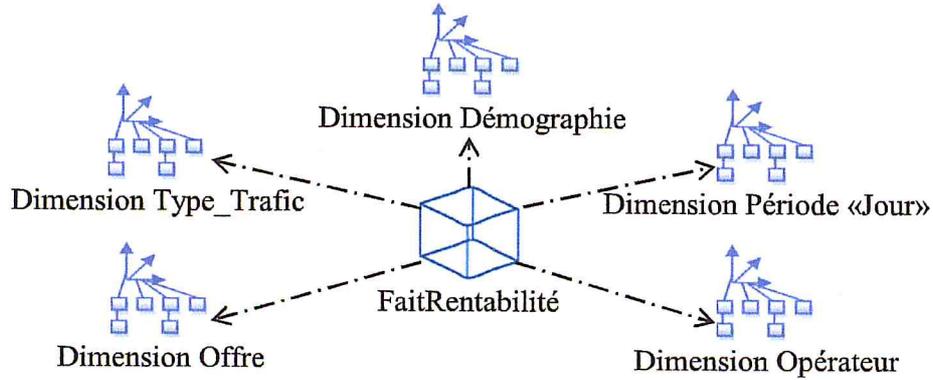


Figure 5.19. Niveau 2 : le Schéma Rentabilité

+ Schéma Nombre des appels :

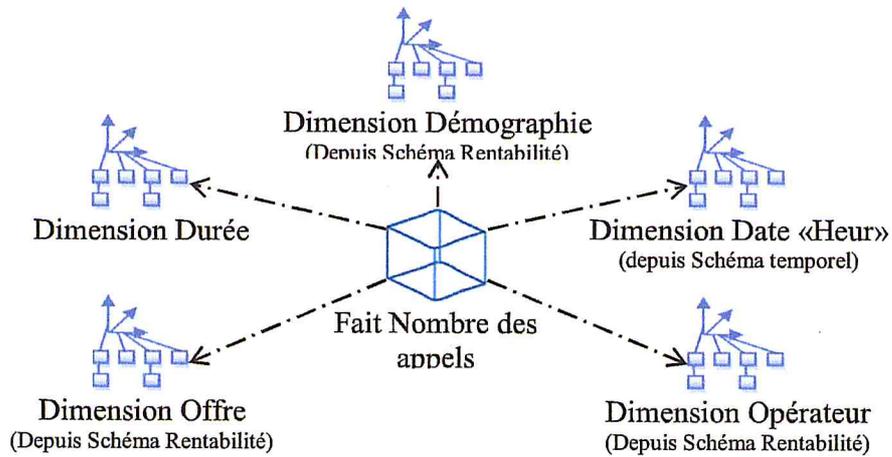


Figure 5.20. Niveau 2 : le Schéma Appels

+ Schéma Etat de marché



Figure 5.21. Niveau 2 : le Schéma Etat de marché

+ Schéma Durée des appels :

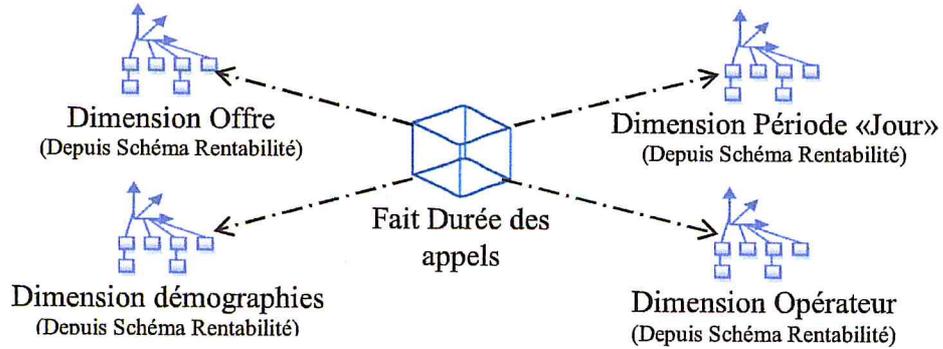


Figure 5.22. Niveau 2 : le Schéma Durée des appels

+ Schéma SMS :

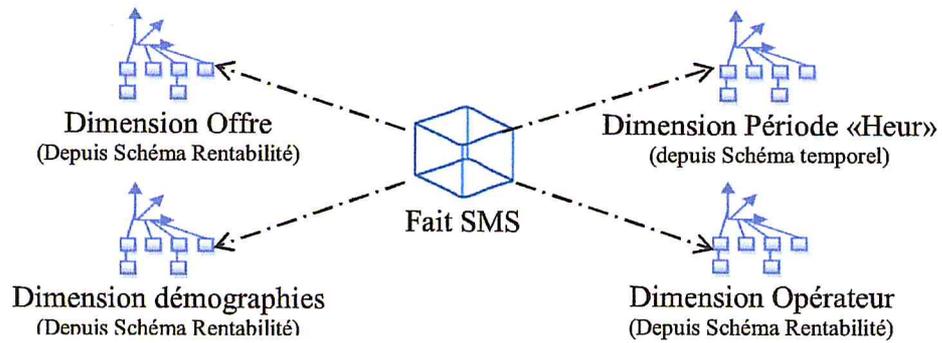


Figure 5.23. Niveau 2 : le Schéma SMS

+ Schéma MMS :

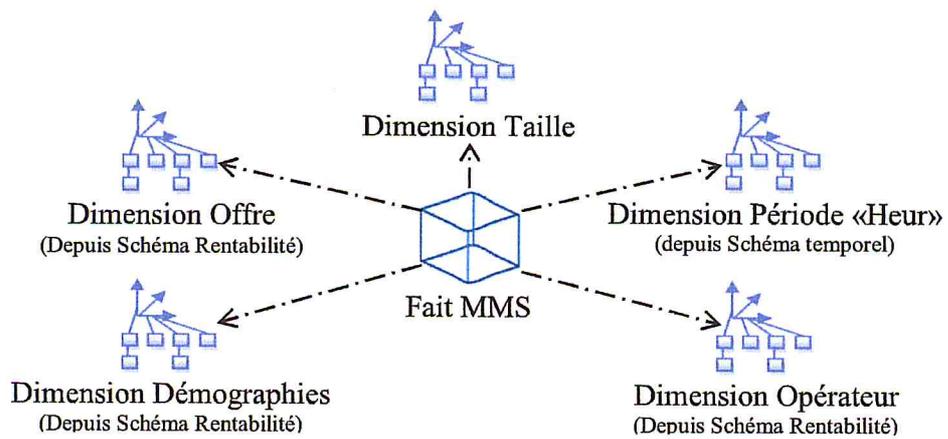


Figure 5.24. Niveau 2 : le Schéma MMS

✦ Schéma Globale des faits Répartition des abonnés

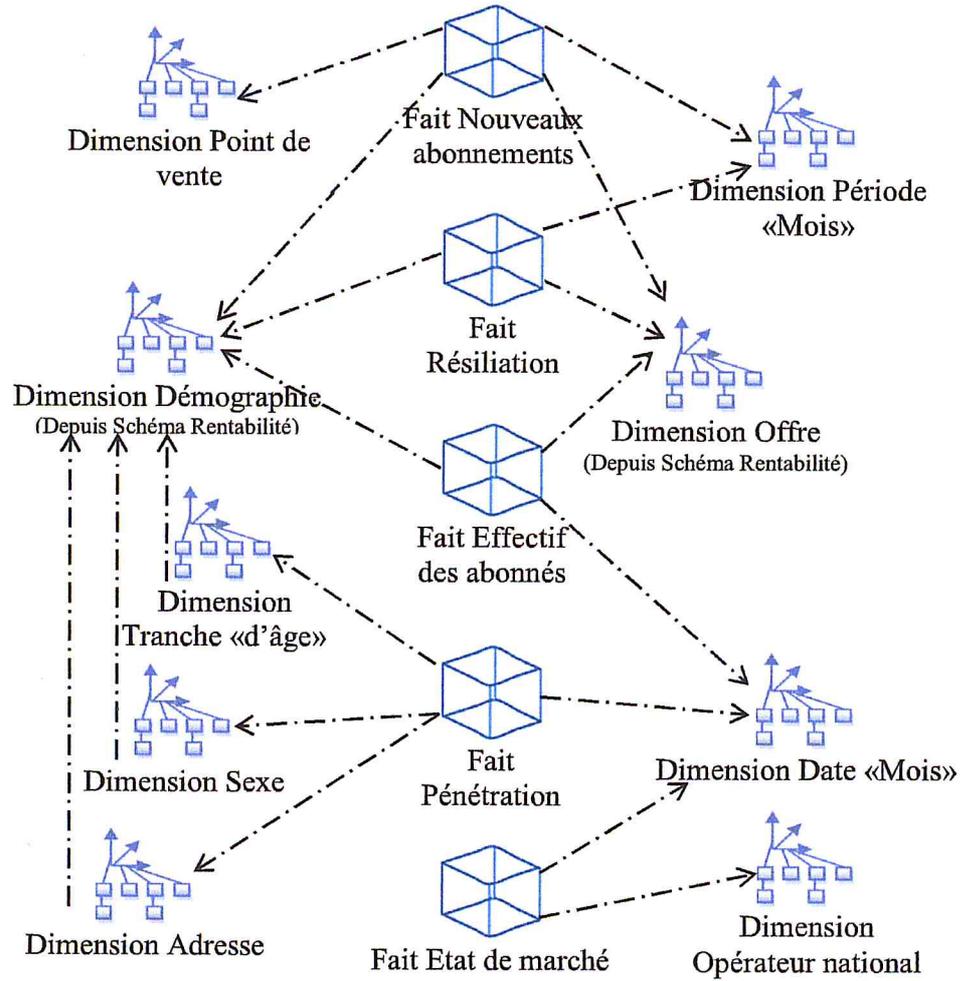


Figure 5.25. Schéma de constellation des faits Répartition des abonnés

✦ Schéma globale de consommation :

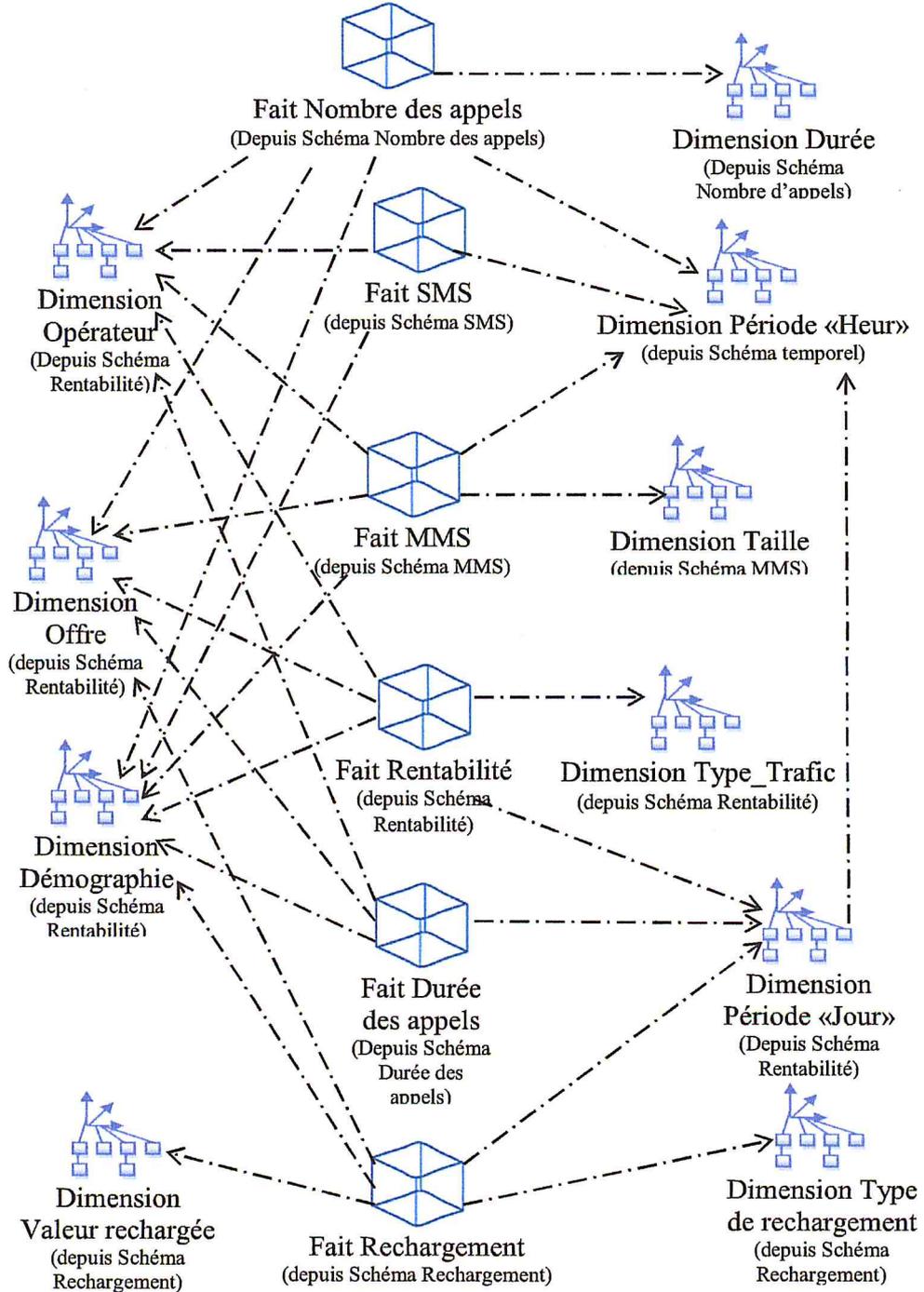


Figure 5.26. Schéma de constellation des faits consommation

V.3.2.2.4. Schéma conceptuel du DW (Niveau 3) :

✚ Dimension Période «Heur» :

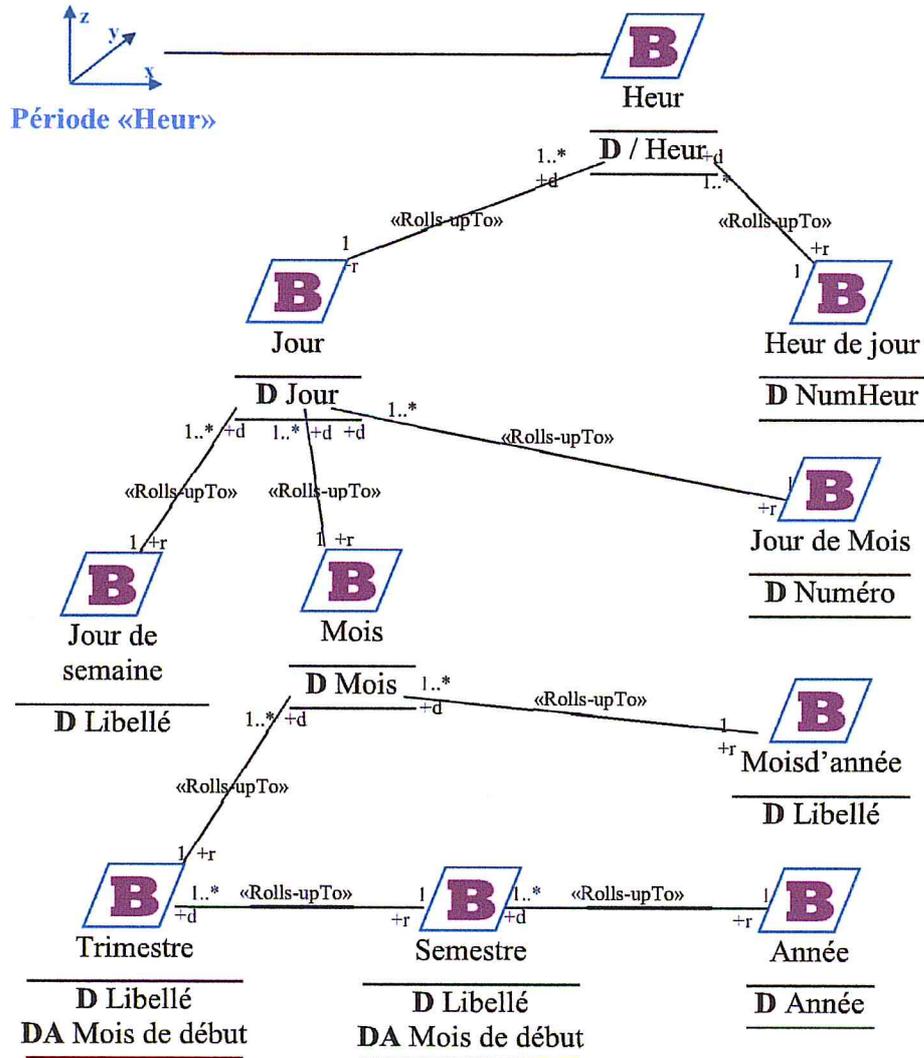


Figure 5.27. Niveau 3 : La dimension période «Heure »

✚ Dimension Offre :

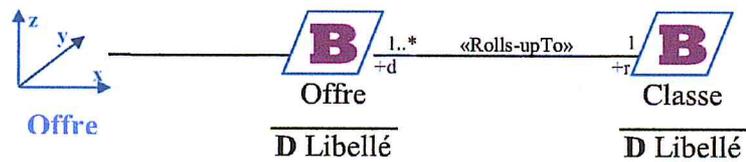


Figure 5.28. Niveau 3 : La dimension Offre

+ Dimension Durée :

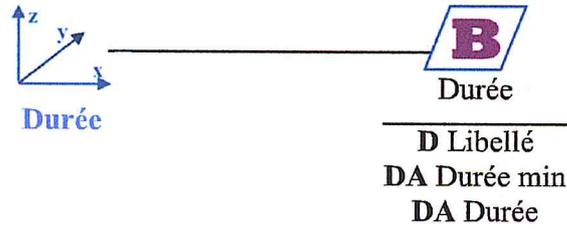


Figure 5.29. Niveau 3 : La dimension Durée

+ Dimension Sexe :

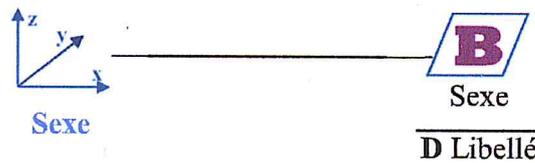


Figure 5.30. Niveau 3 : La dimension Sexe

+ Dimension Opérateur :

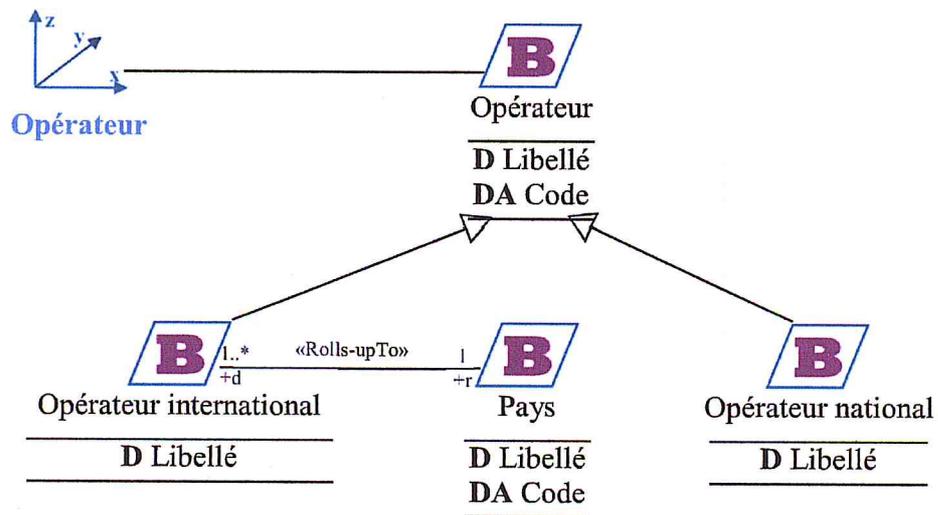


Figure 5.31. Niveau 3 : La dimension Opérateur

+ Dimension Démographique

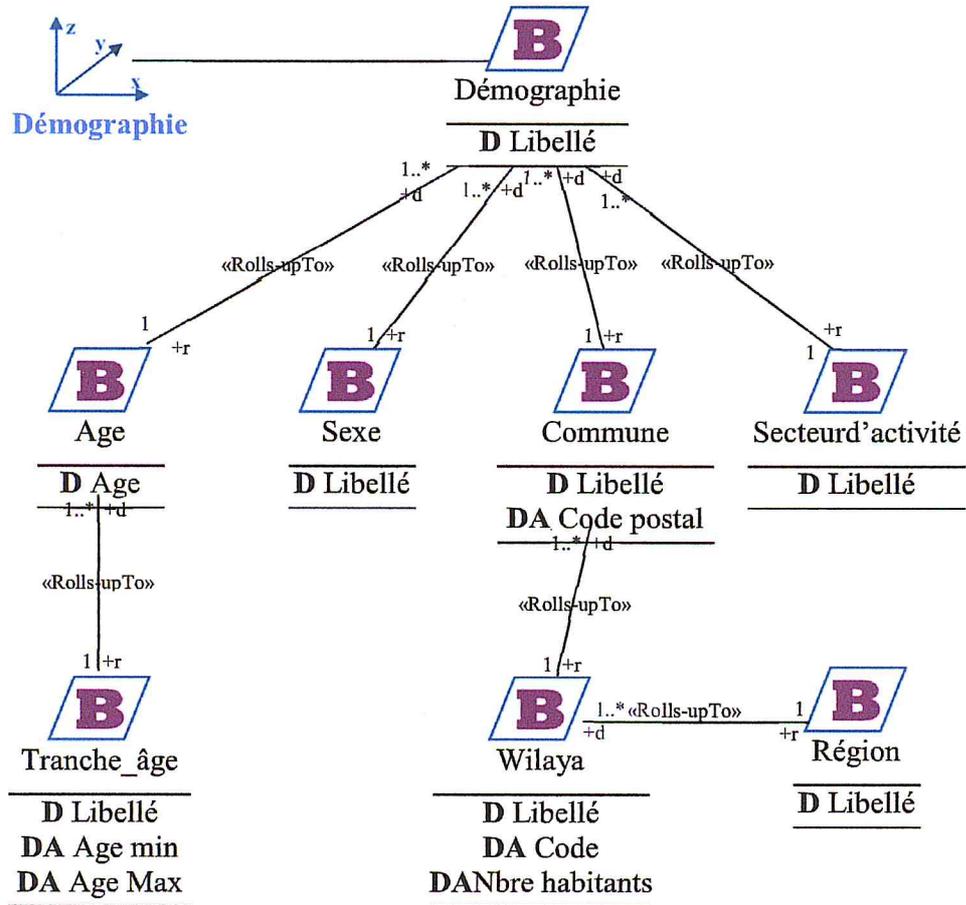


Figure 5.32. Niveau 3 : La dimension Démographique

+ Dimension Date «Mois» :

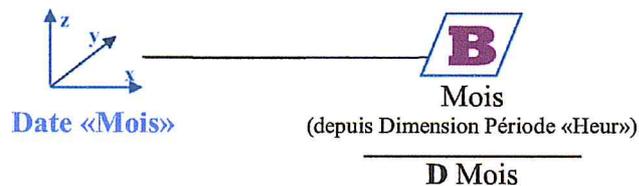


Figure 5.33. Niveau 3 : La dimension Date «Mois »

+ Dimension Type Rechargement :

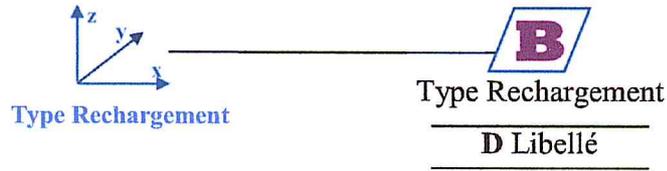


Figure 5.34. Niveau 3 : La dimension Type Rechargement

+ Dimension Tranche «d'âge» :

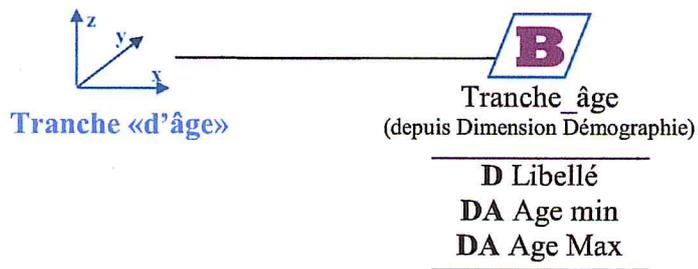


Figure 5.35. Niveau 3 : La dimension Tranche d'âge

+ Dimension Valeur Rechargé :

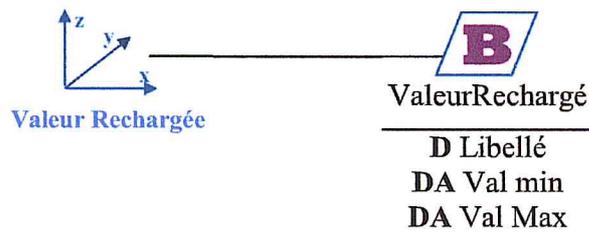


Figure 5.36. Niveau 3 : La dimension Valeur Rechargé

+ Dimension Taille (MMS) :

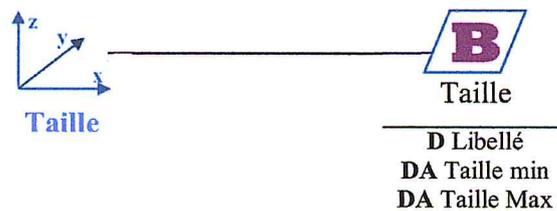


Figure 5.37. Niveau 3 : La dimension Taille MMS

+ Dimension Période «Jour» :

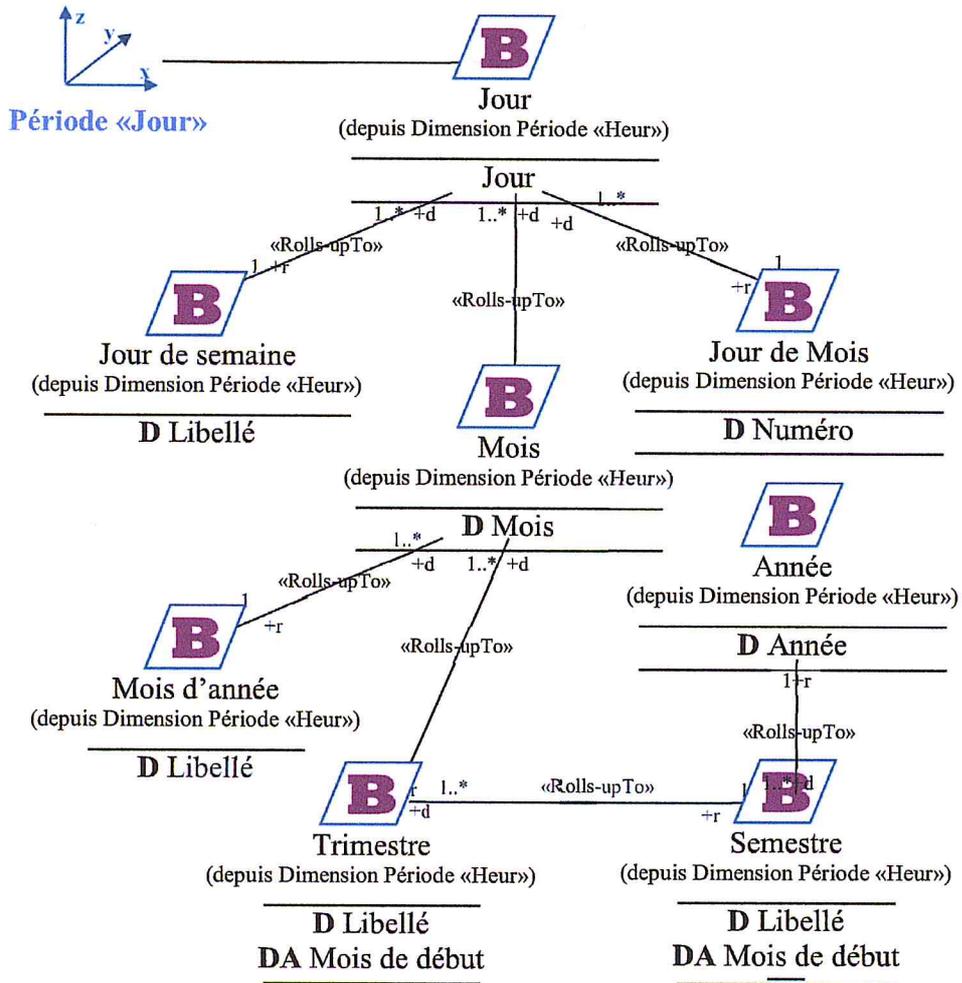


Figure 5.38. Niveau 3 : La dimension période «Jour»

+ Dimension Type Trafic :

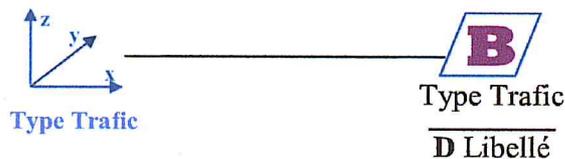


Figure 5.39. Niveau 3 : La dimension Type Trafic

+ Dimension adresse :

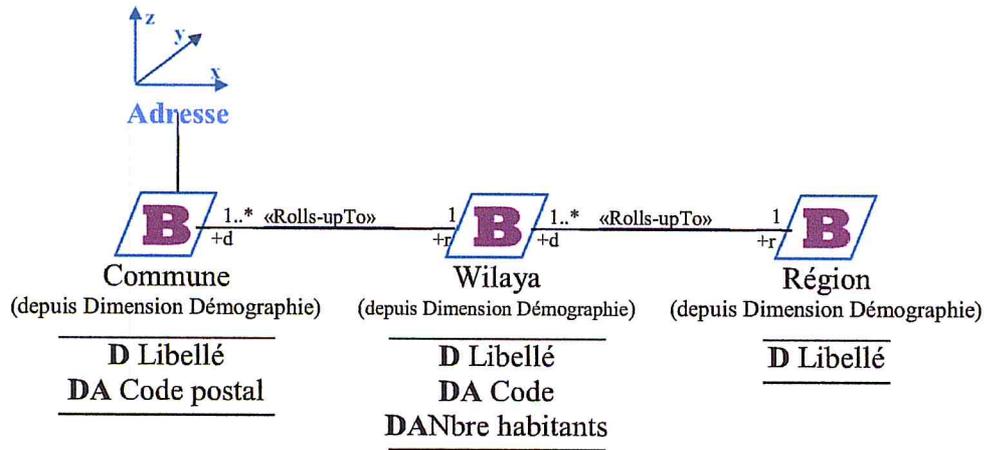


Figure 5.40. Niveau 3 : La dimension Adresse

+ Dimension Période «Mois» :

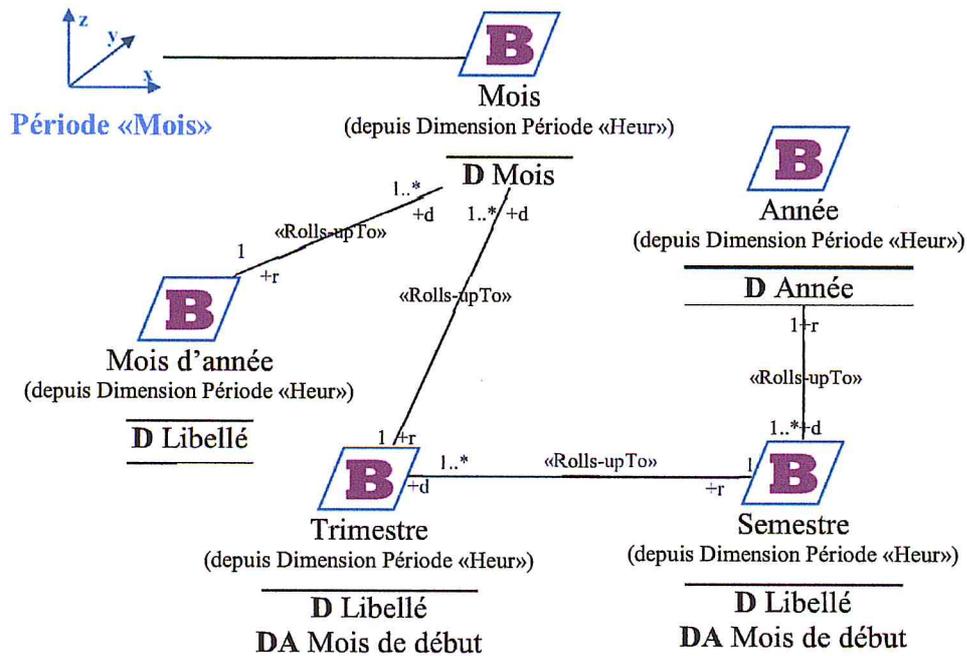


Figure 5.41. Niveau 3 : La dimension période «Mois »

✚ Dimension Point de vente :

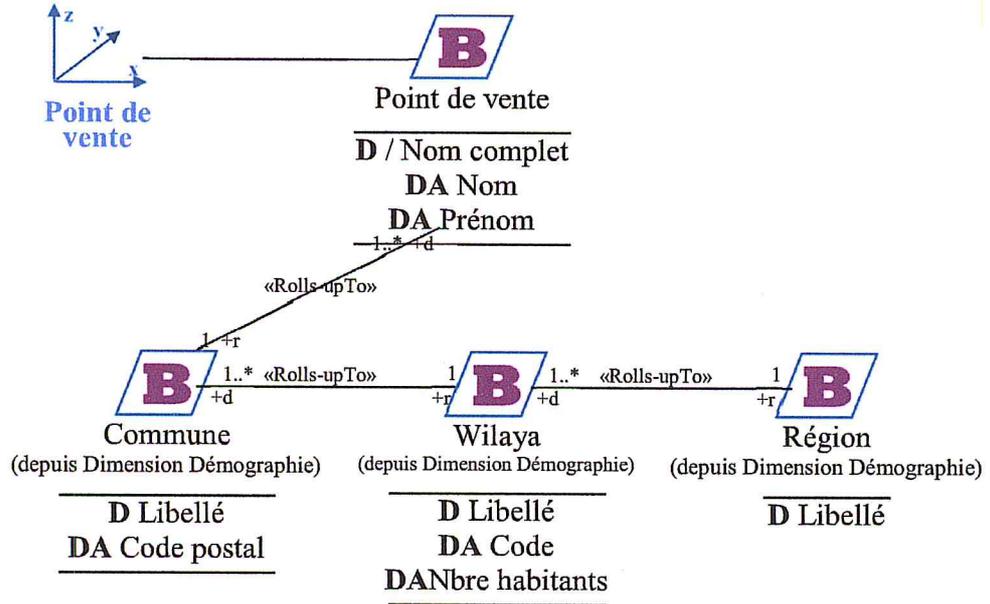


Figure 5.42. Niveau 3 : La dimension point de Vente

✚ Dimension Opérateur national :

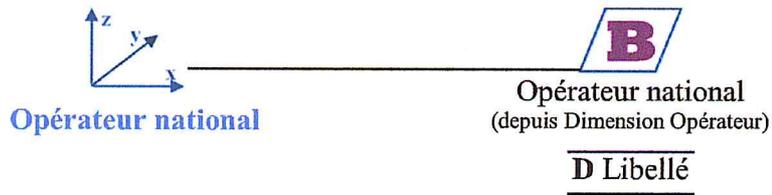


Figure 5.43. Niveau 3 : La dimension Opérateur National

Fait Effectif des abonnés :

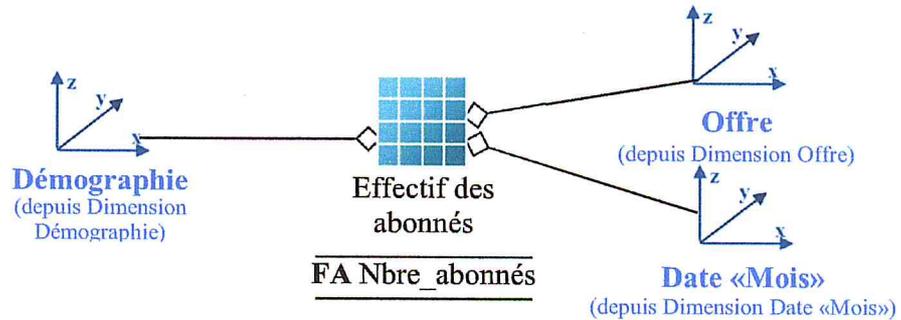


Figure 5.44. Niveau 3 : Le Fait Effectif des abonnés

Fait Nouveaux abonnements :

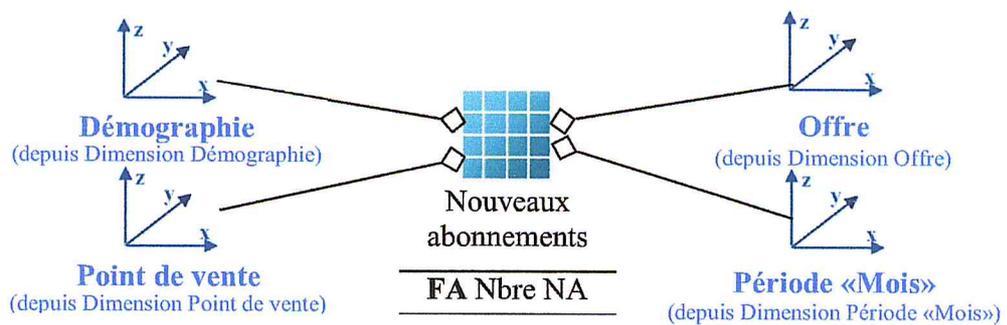


Figure 5.45. Niveau 3 : Le Fait Effectif des abonnés

Fait Pénétration :

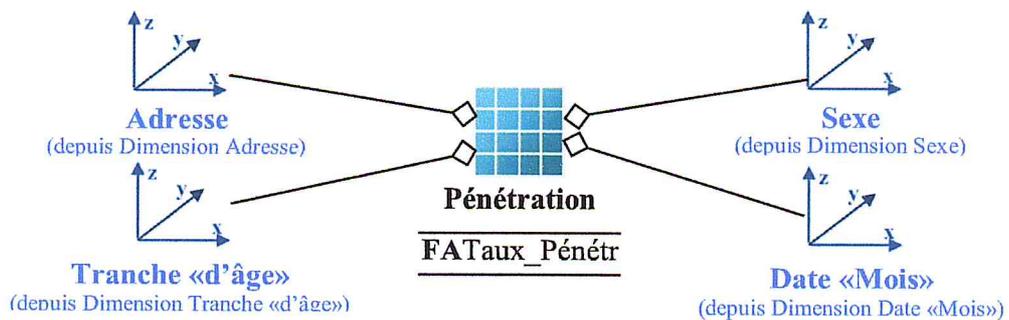


Figure 5.46. Niveau 3 : Le Fait Pénétration

Fait Résiliation :

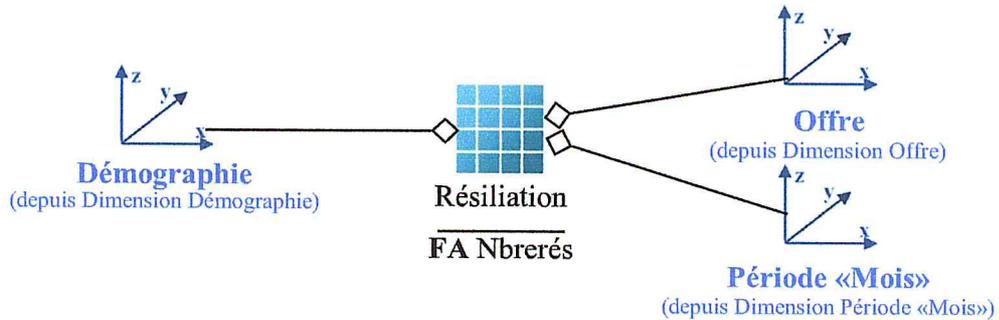


Figure 5.47. Niveau 3 : Le Fait Résiliation

Fait Nombre des appels :

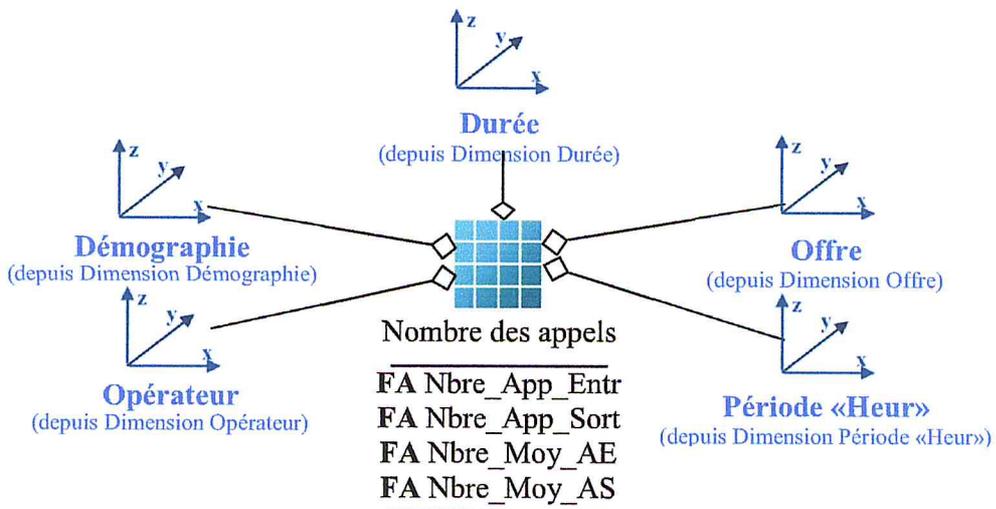


Figure 5.48. Niveau 3 : Le Fait Nombre des appels

Fait Etat de marché :



Figure 5.49. Niveau 3 : Le Fait Etat de marché

✚ **Fait Rechargement :**

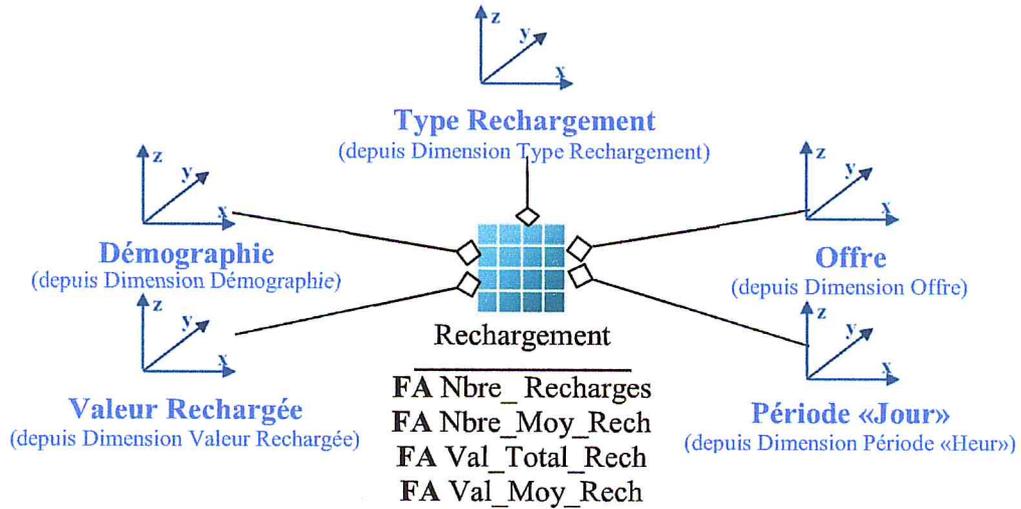


Figure 5.50. Niveau 3 : Le Fait Rechargement

✚ **Fait Durée des appels :**

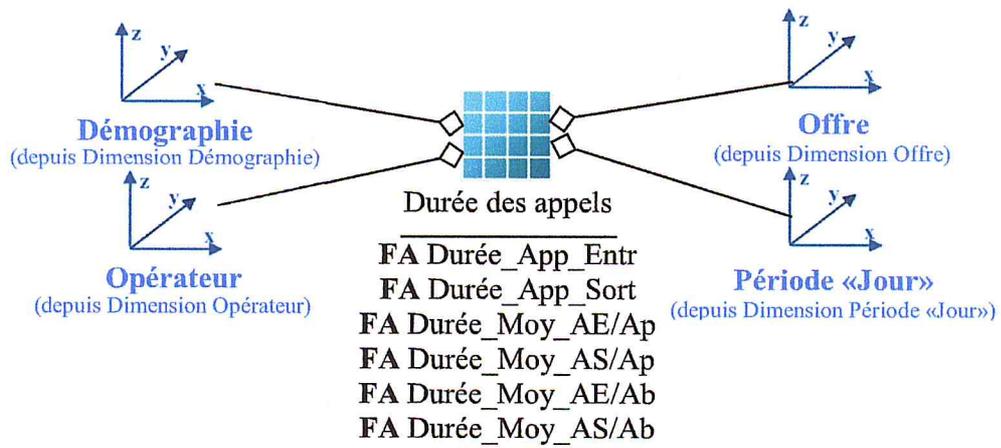


Figure 5.51. Niveau 3 : Le Fait Durée des appels

✚ **Fait SMS :**

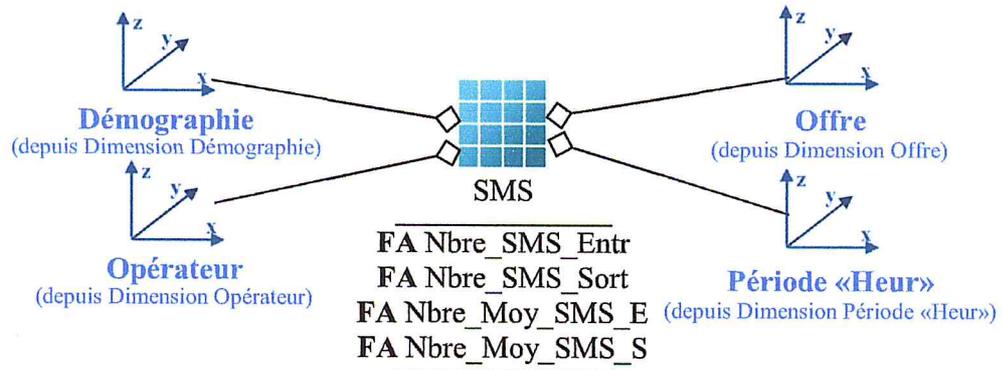


Figure 5.52. Niveau 3 : Le Fait SMS

✚ **Fait MMS :**

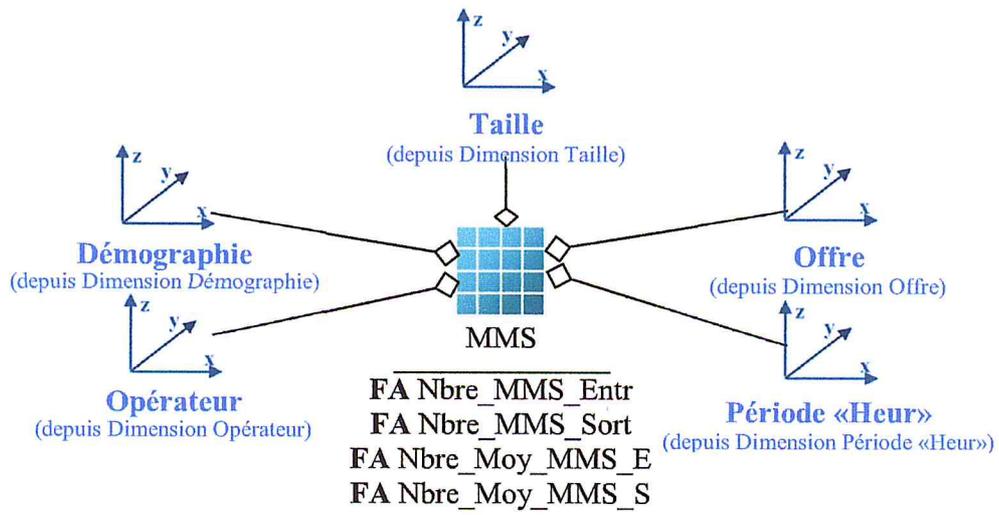


Figure 5.53. Niveau 3 : Le Fait MMS

V.3.2.3. Schéma conceptuel de l'ETL (Data Mapping diagram) :

Un DW est alimenté à partir des sources généralement hétérogènes, ce processus est fait par un outil appelé ETL qui a pour rôle d'extraire des données hétérogènes, de les transformer (nettoyage, transformation, normalisation etc.) et les charger dans le DW. Cette phase est considérée comme le facteur clé de la réussite d'un projet d'SID. Malgré cette importance, elle est négligée par la plupart des méthodes de conception d'SID.

Parmi les valeurs ajoutées de la méthode DWEP, la proposition des diagrammes pour la modélisation d'ETL. *Luján Mora* a adapté UML par un profil pour concevoir l'ETL, toujours *Luján Mora* suivre la philosophie de modélisation en 3 niveaux d'abstraction (conceptuel, logique physique). Au niveau conceptuel la méthode DWEP utilise un diagramme appelé (Data Mapping diagram DMD), ce diagramme permet de tracer le chemin des données allant des sources des données passant par la zone de préparation des données (ODS, ETL) jusqu'à la zone de présentation des données (DW, Data Mart).

Luján Mora a défini un attribut comme l'élément première-classe (first-class) de la modélisation pour permettre de visualiser les liens entre les attributs.

Pour capturer les interconnexions entre les éléments de conception (tables, attributs ...), *Luján Mora* introduit la notion de **Mapping**, le Mapping est défini dans tous les degrés de granularités (niveau de tables, niveau des attributs ...), un Mapping lie 2 éléments de même niveau. Il est exprimé selon le niveau, soit par un package, une relation ou un objet, On a 4 niveaux de détail (Figure 5.53) :

✚ Niveau 0 : Niveau de base de données (DB Level):

À ce niveau chaque schéma de l'environnement du DW (sources de données au niveau conceptuel dans le SCS, schéma conceptuel de DW au niveau conceptuel dans le DWCS) est représenté par un package (« Data Mapping Package »). Le Mapping entre les différents schémas est modélisé par un package, encapsule tous les niveaux inférieurs du Mapping entre les différents schémas (Figure 5.54).

+ Niveau 1 (niveau de Flux de données)

Ce niveau décrit les relations entre les tables de schéma sources (SCS) et de celles de schéma cible (DWCS). Autrement dit, les « Data Mapping diagram (DMD)» est zoomé en un ensemble de DMDs plus détaillés pour capturer comment chaque table est ciblée est reliée avec sa table source. Ce niveau engendre 3 stéréotypes «Mapping», «Input», «Output» (Figure 5.55).

+ Niveau 2 (Niveau des tables) :

Au niveau précédent les liens entre les sources et les cibles sont représentés par un seul Mapping package. A ce niveau en détaillant tous les transformations intermédiaires et cela à travers un ensemble de package permet de segmenter le package de niveau 2 en un séquencèrent d'étapes. Ce niveau permet de visualiser des classes intermédiaires qui représentent physiquement soit une table de l'ODS, soit un objet ou bien une table virtuelle. Ce niveau engendre un ensemble de stéréotypes «Mapping», «Input», «Output», et «Intermediate» (Figure 5.56).

+ Niveau 3 (niveau des attributs)

Ce niveau permet de zoomer sur chaque Mapping package de niveau inférieur pour tracer les liens au niveau des attributs entre chaque tables cible et table source. «Contain», «Map», «MapObj», «Domain», «Range», «Input», «Output», et «Intermediate» (Figure 5.57).

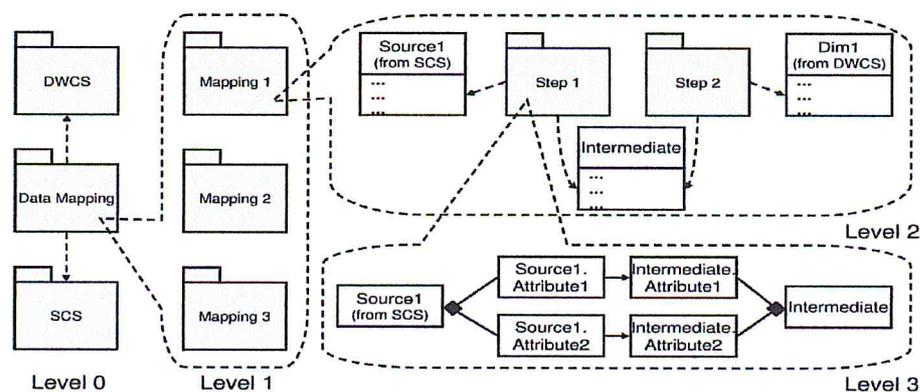


Figure 5.54. Niveaux de Data Mapping

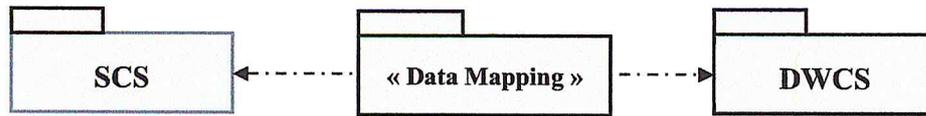


Figure 5.55. Exemple de DMD (niveau 0)

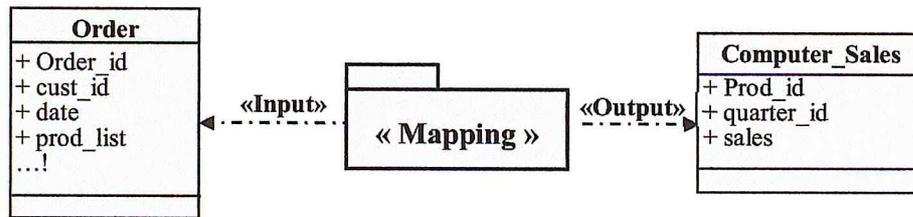


Figure 5.56. Exemple de DMD (niveau 1)

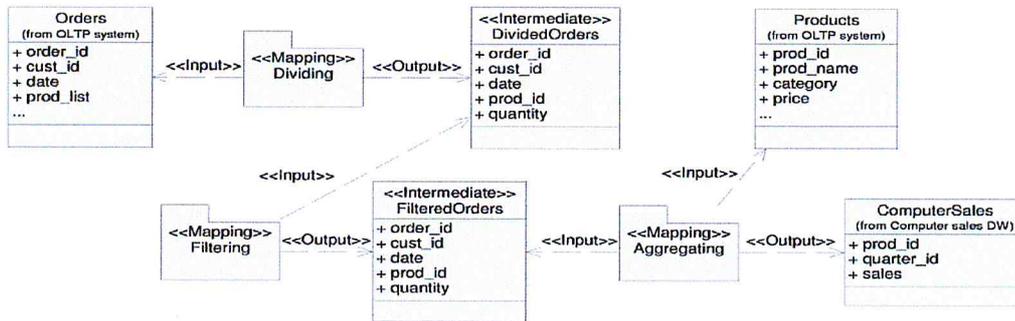


Figure 5.57. Exemple de DMD (niveau 2)

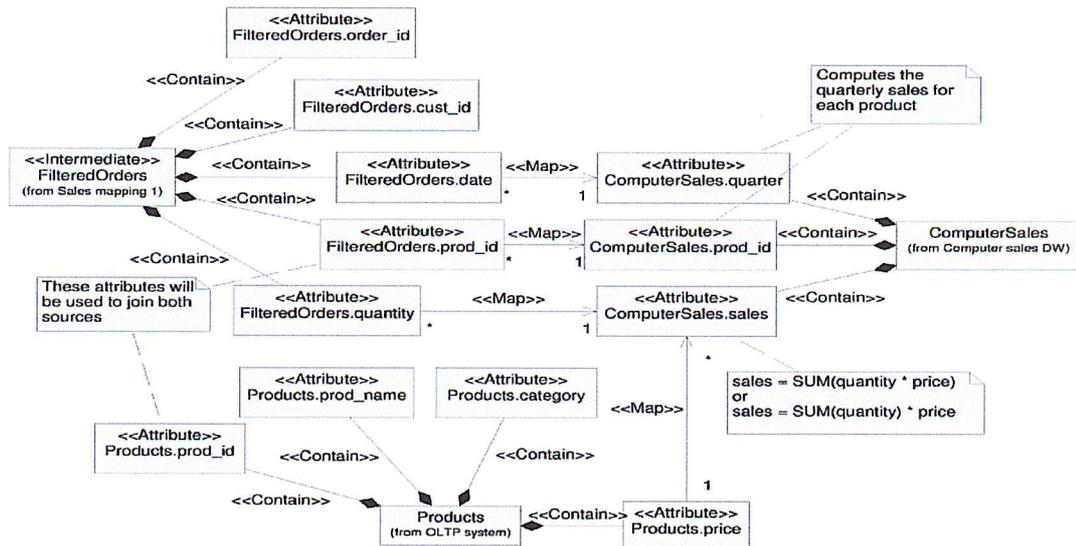


Figure 5.58. Exemple de DMD (niveau 3)

✚ DMD niveau 0 :

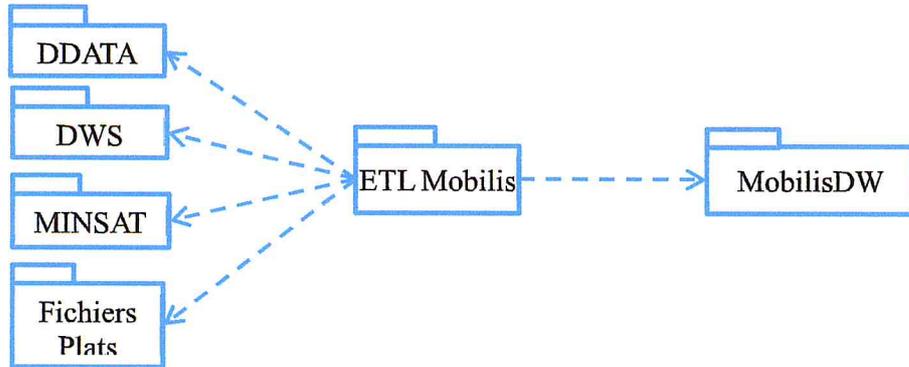


Figure 5.59. DMD (Niveau 0)

✚ DMD niveau 2 :

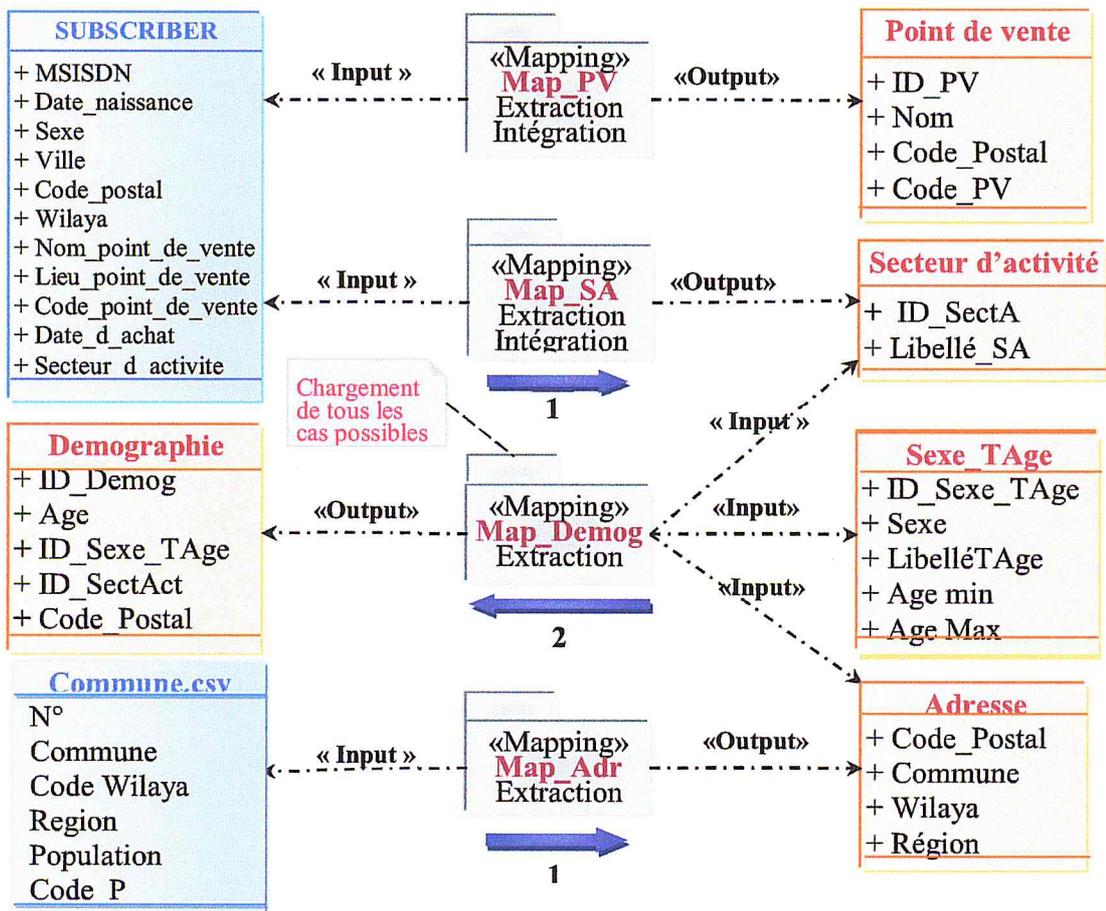


Figure 5.60. Mapping des dimensions démographique

Conception et Réalisation d'un Système décisionnel pour le marketing

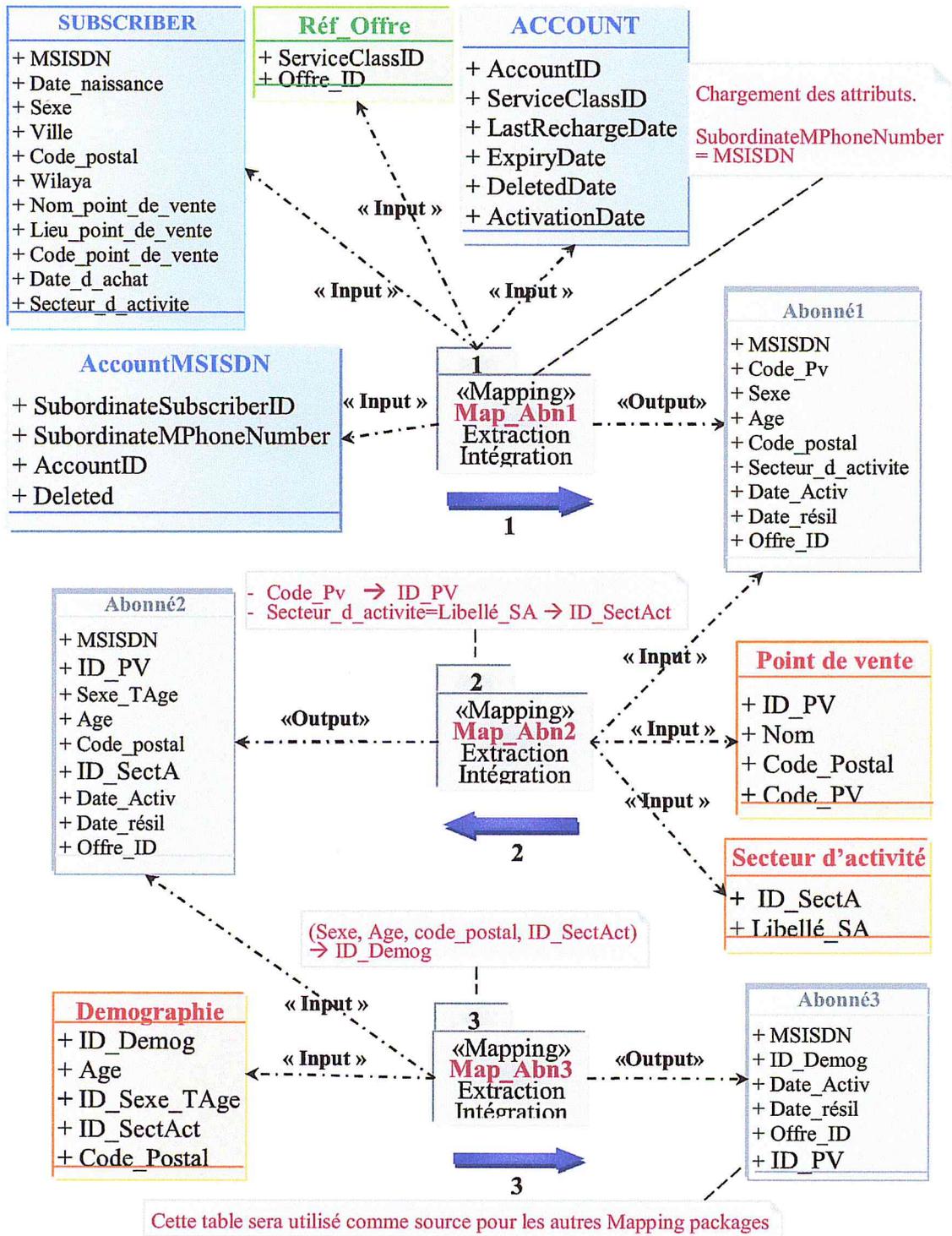


Figure 5.61. Mapping des tables intermédiaires sur les abonnés

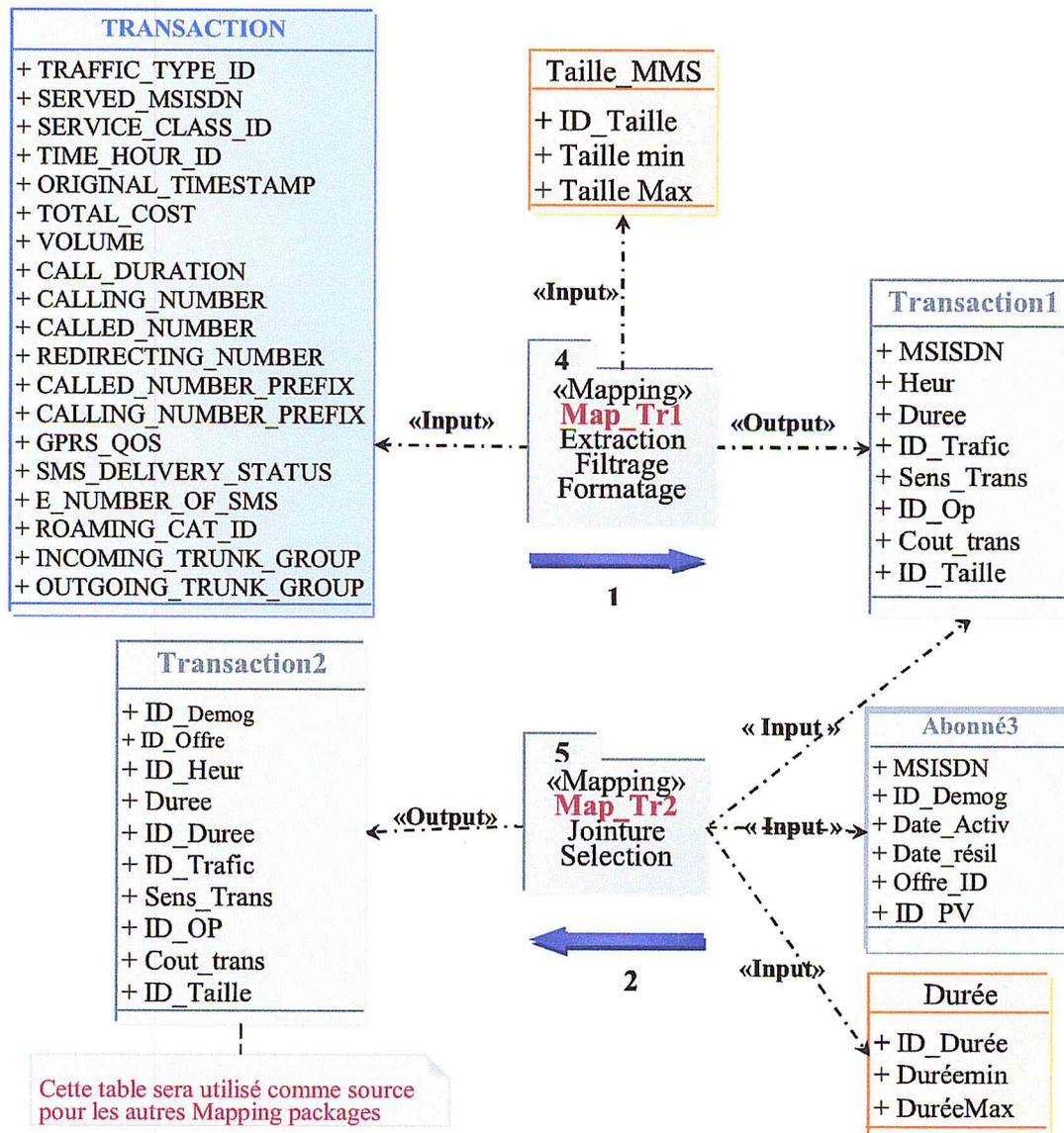


Figure 5.62. Mapping des tables intermédiaires sur les Transactions

Conception et Réalisation d'un Système décisionnel pour le marketing

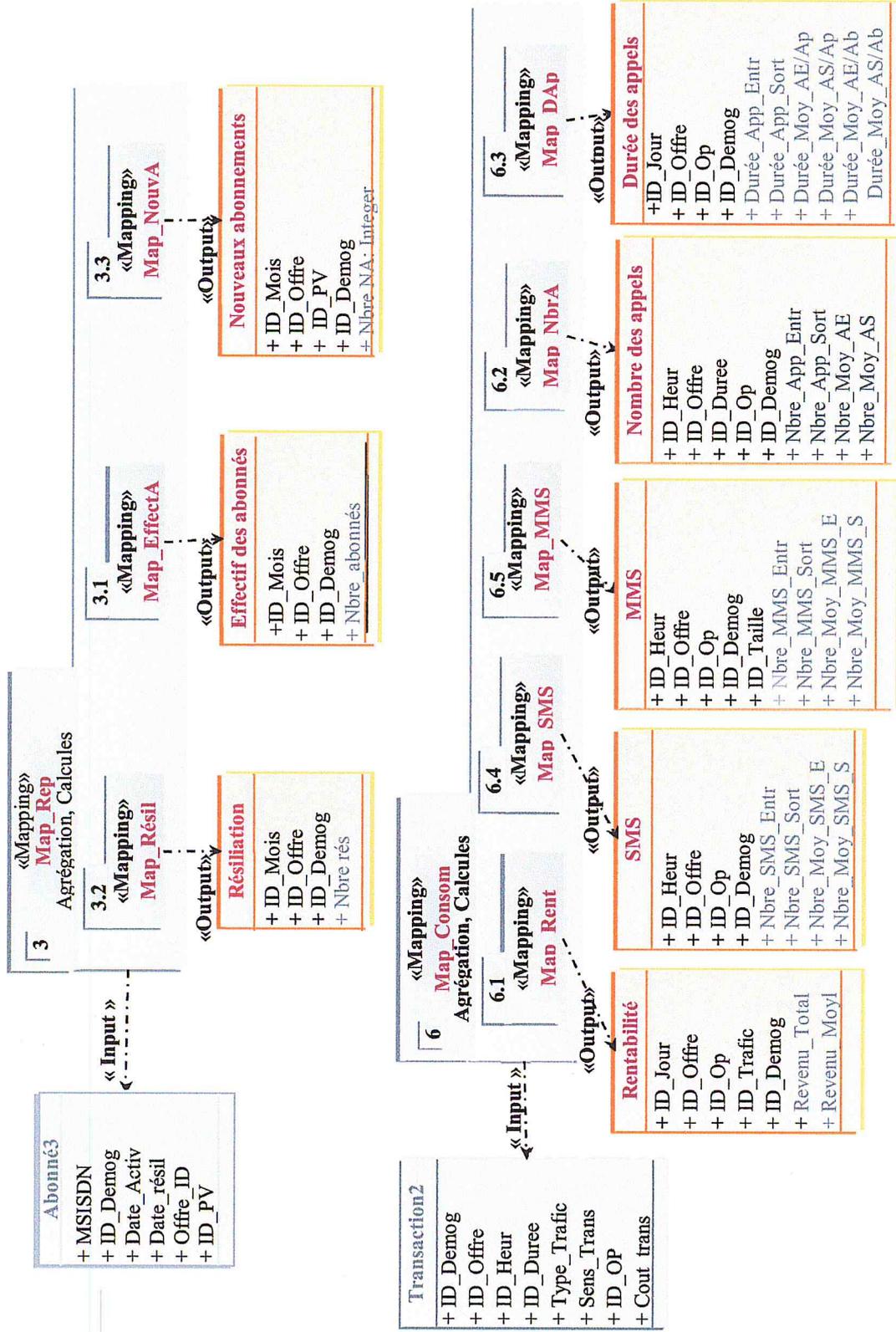


Figure 5.63. Mapping des faits

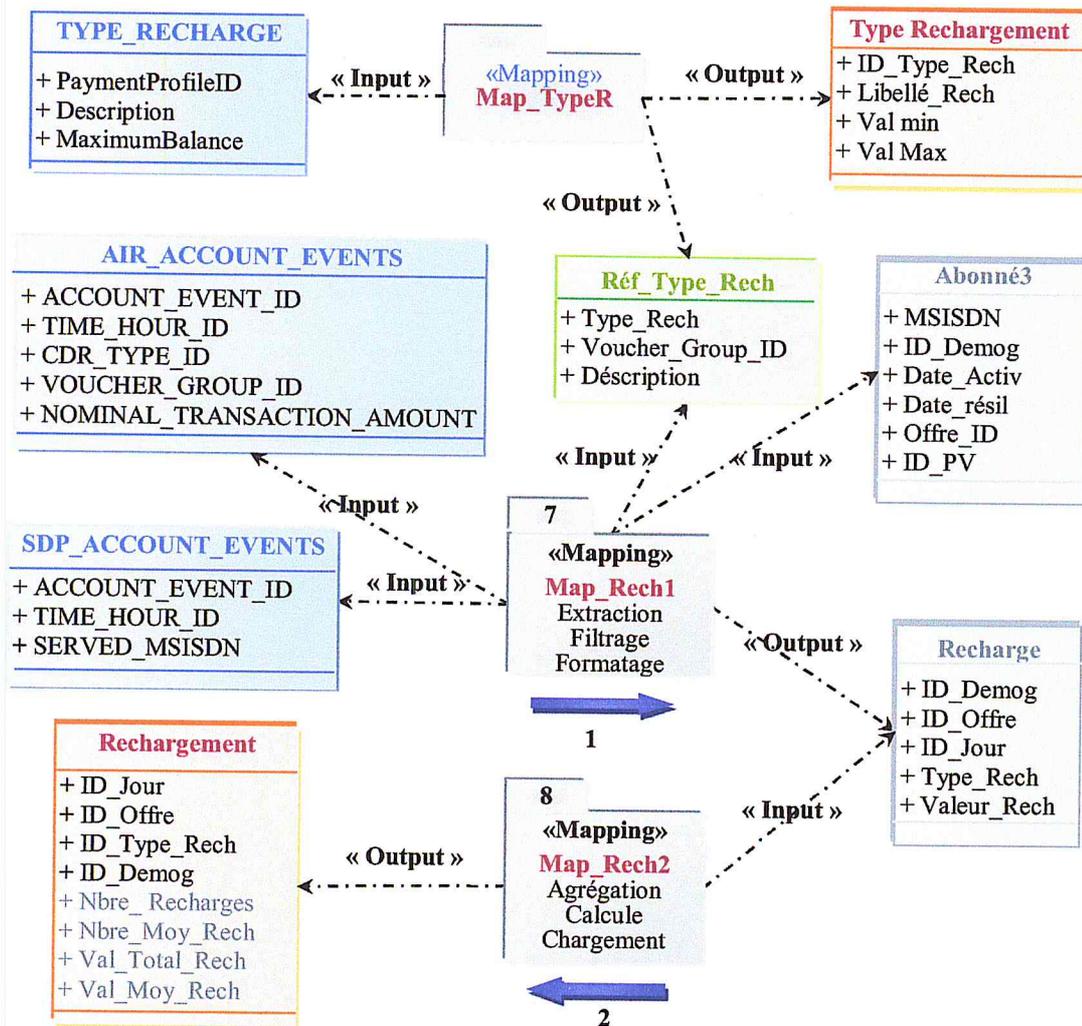


Figure 5.64. Mapping de Rechargement

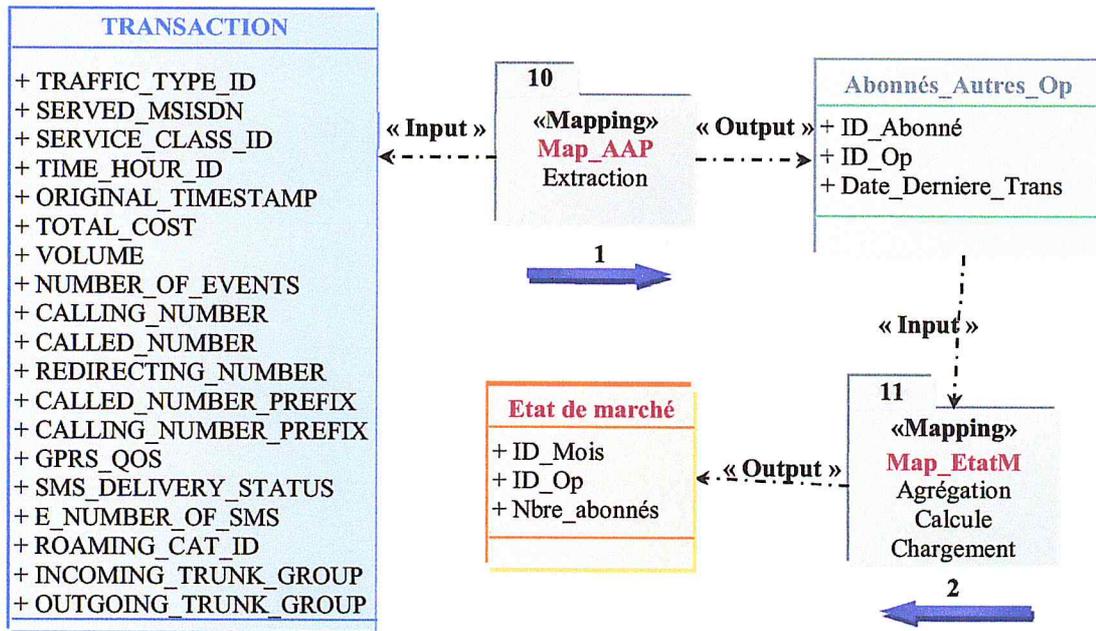


Figure 5.65. Mapping du fait Etat de marché

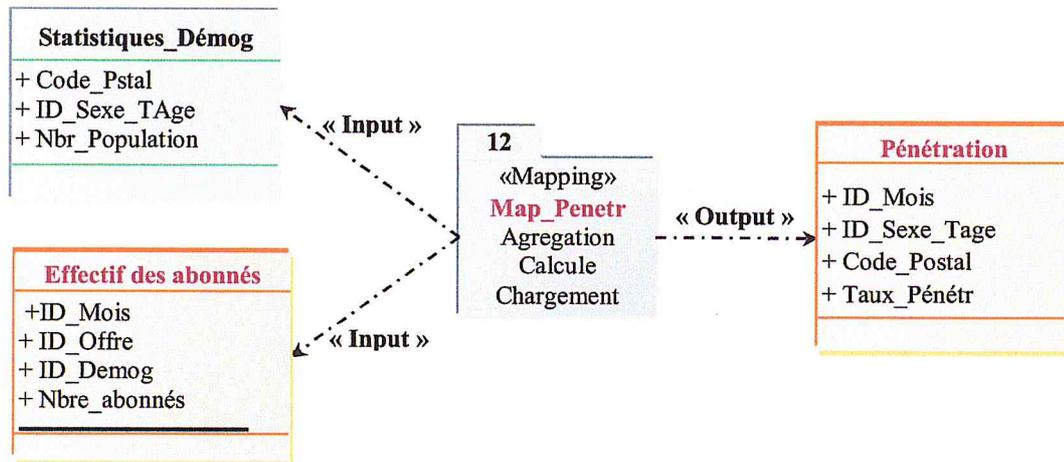


Figure 5.66. Mapping du fait Pénétration

Conception et Réalisation d'un Système décisionnel pour le marketing

➤ DMD niveau 3 :

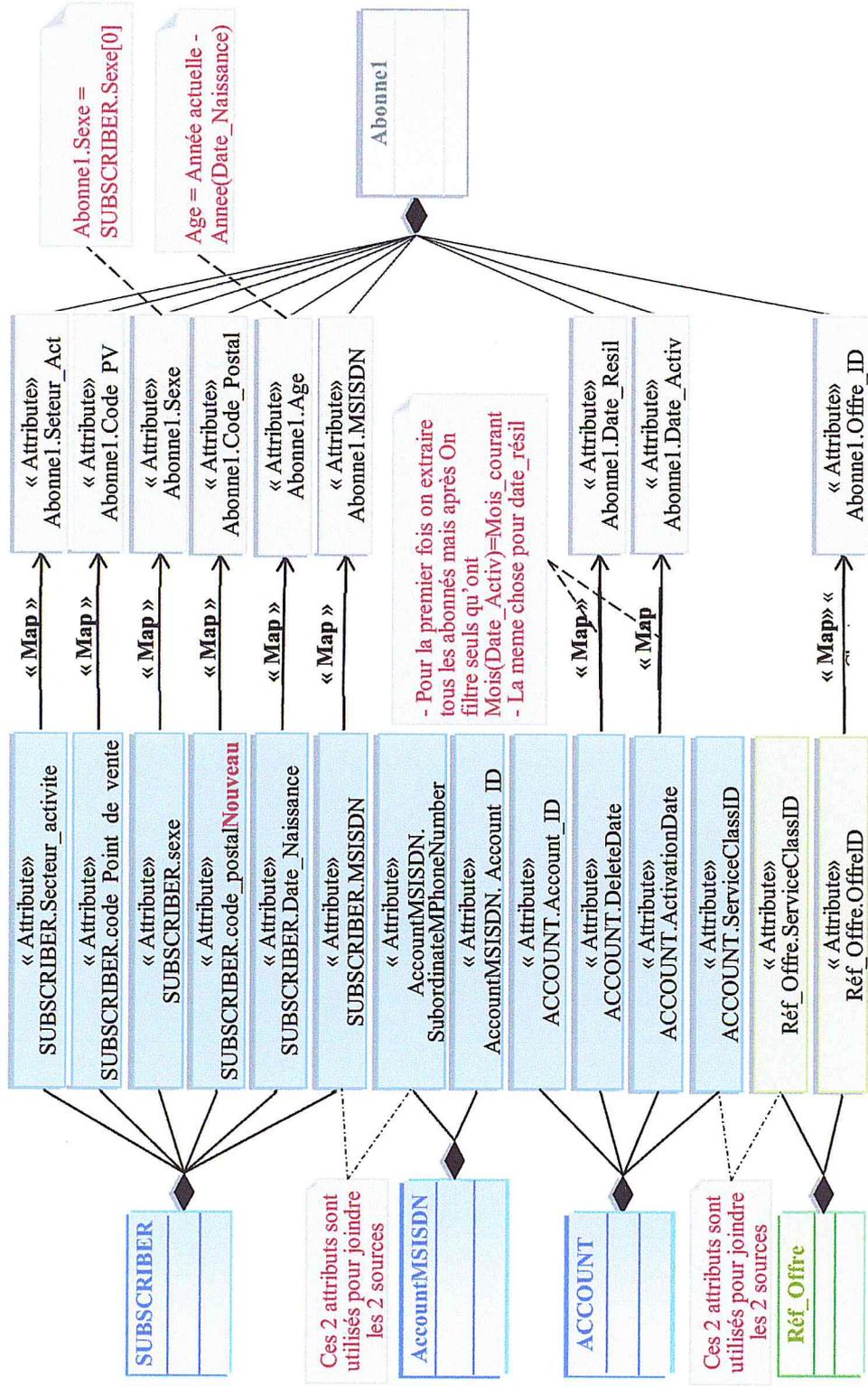


Figure 5.67. 3 ème niveau de Mapping Package MapAbn1

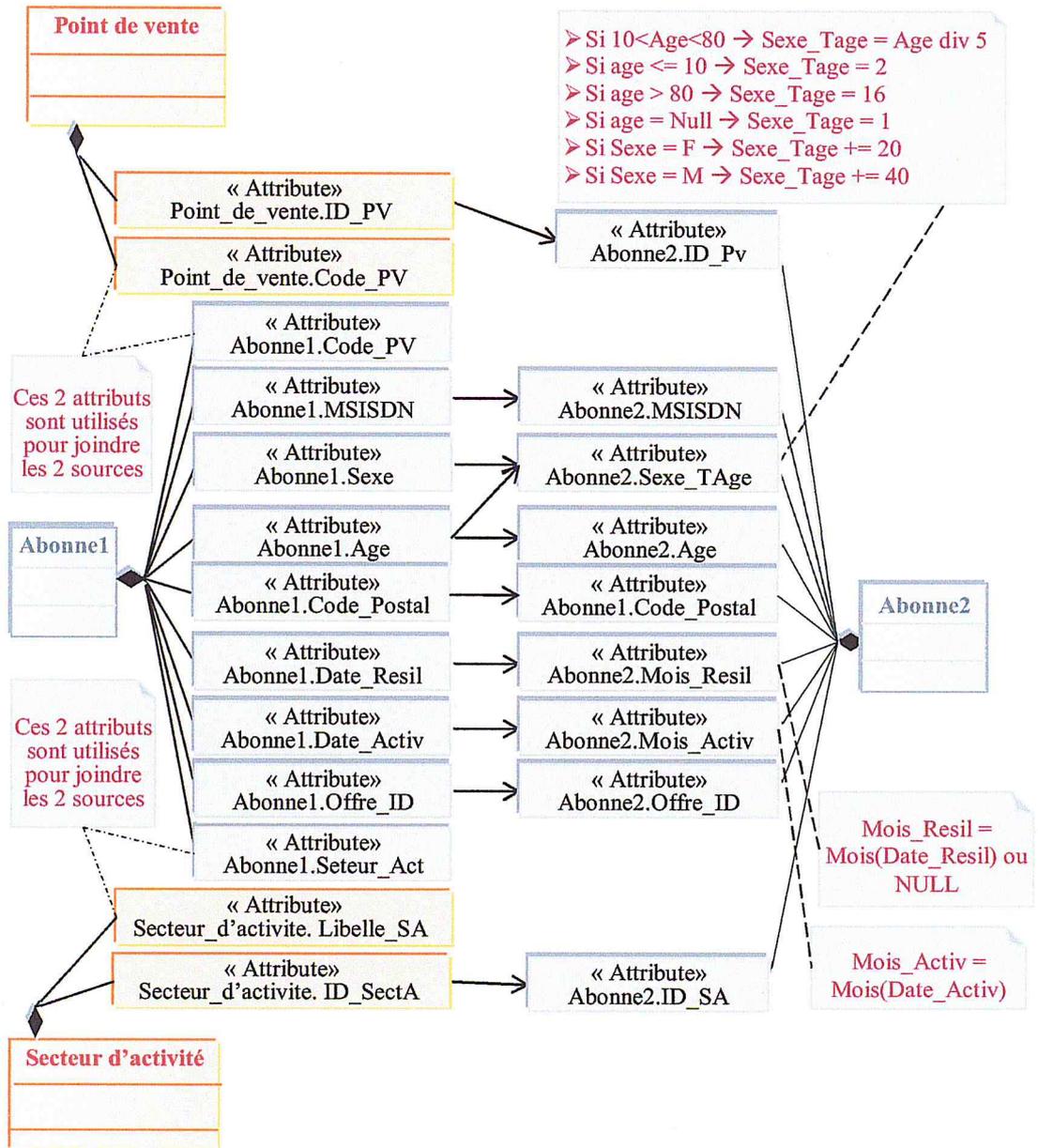


Figure 5.68. 3 ème niveau de Mapping Package MapAbn2

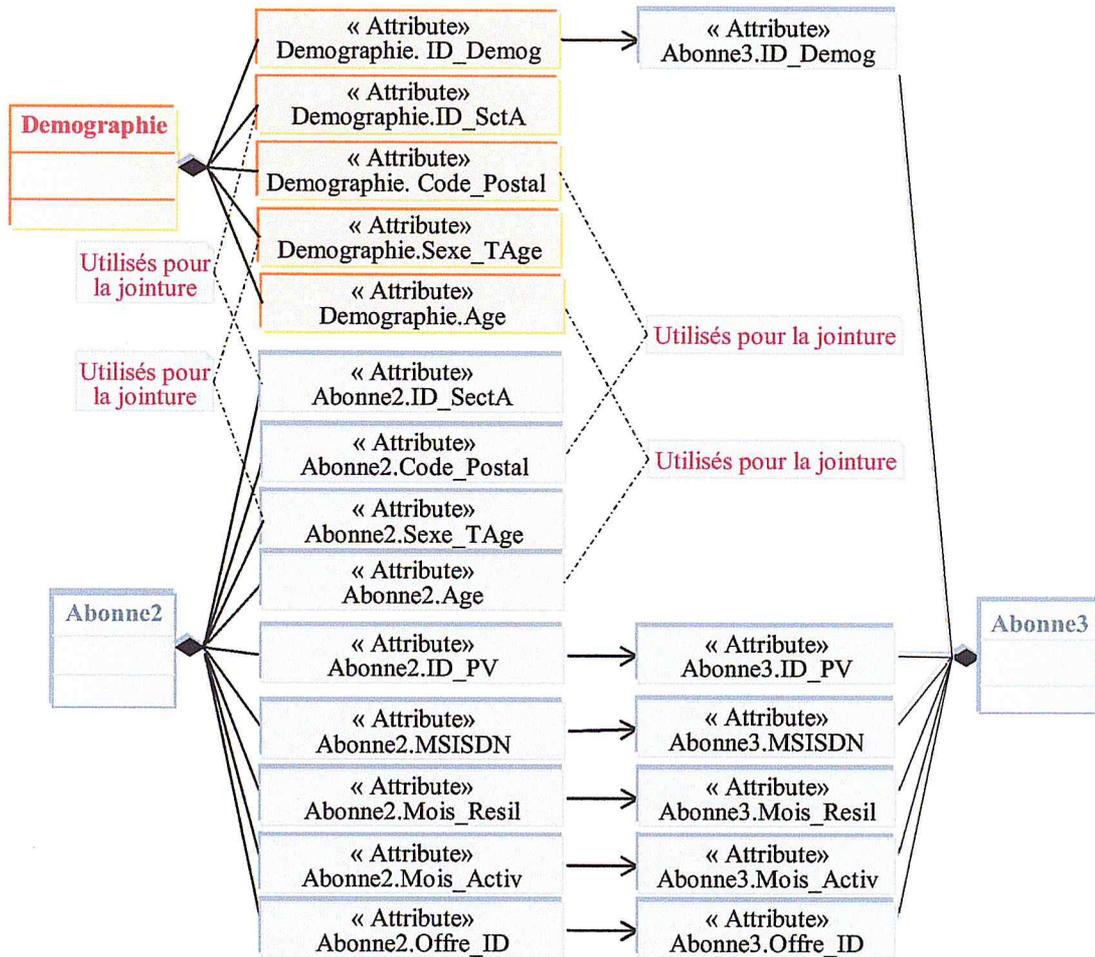


Figure 5.69. 3 ème niveau de Mapping Package MapAbn3

Conception et Réalisation d'un Système décisionnel pour le marketing

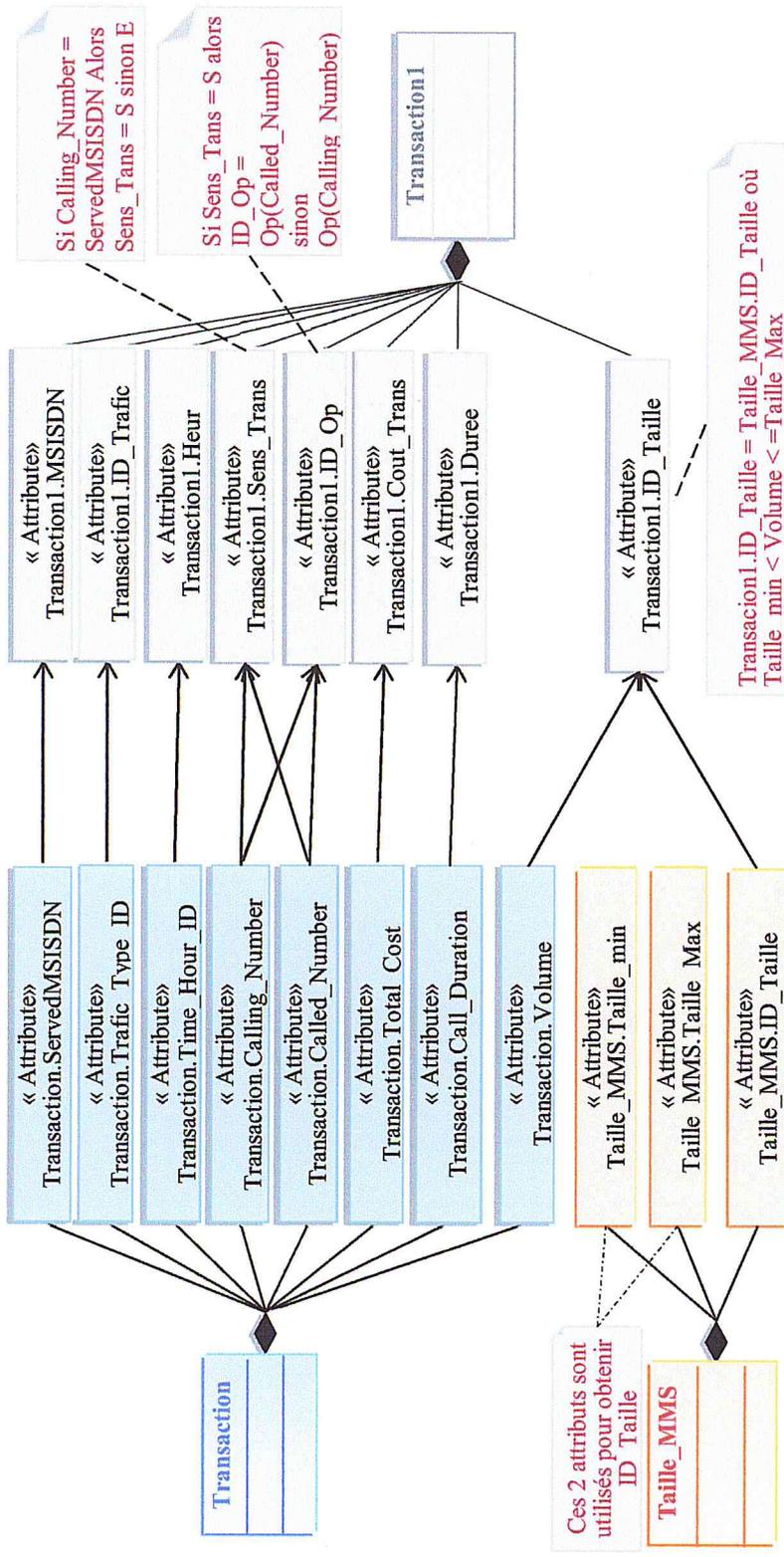


Figure 5.70. 3^{ème} niveau de Mapping Package Map Tr1

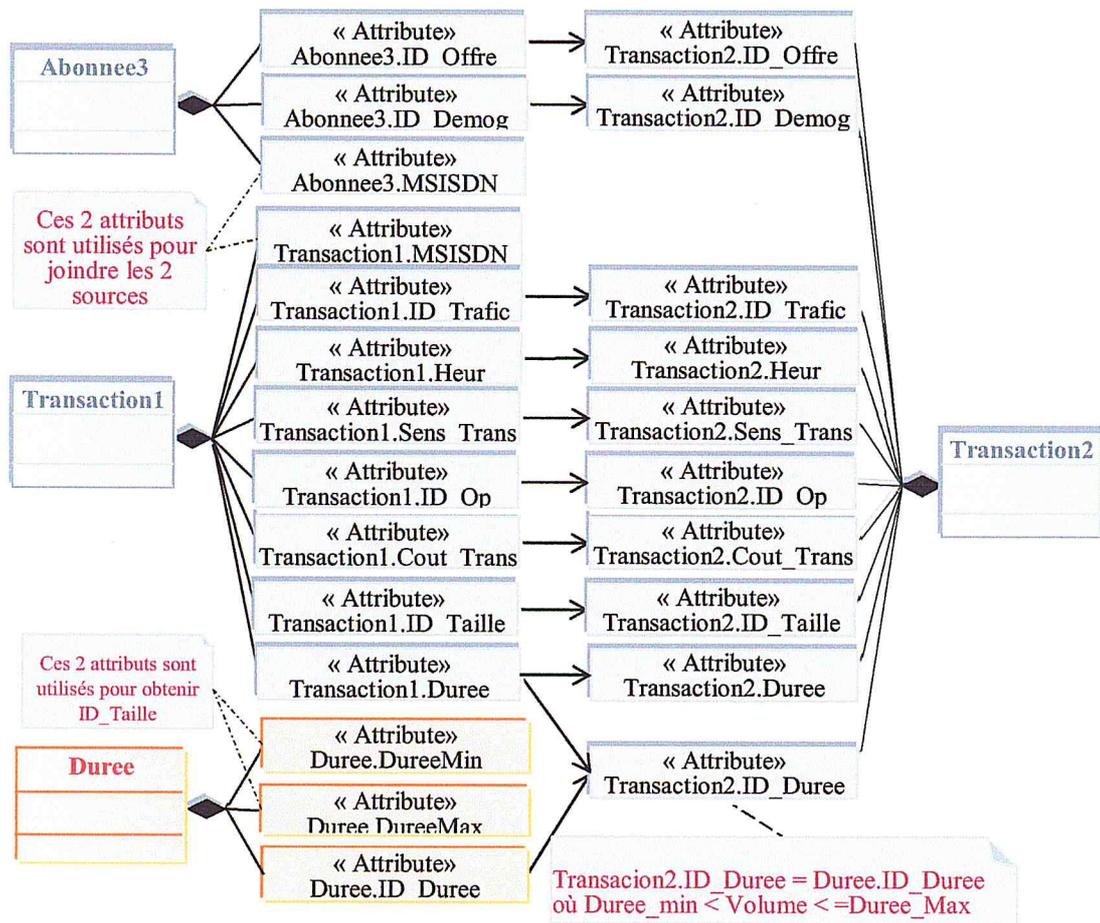


Figure 5.71. 3^{ème} niveau de Mapping Package MapTr2

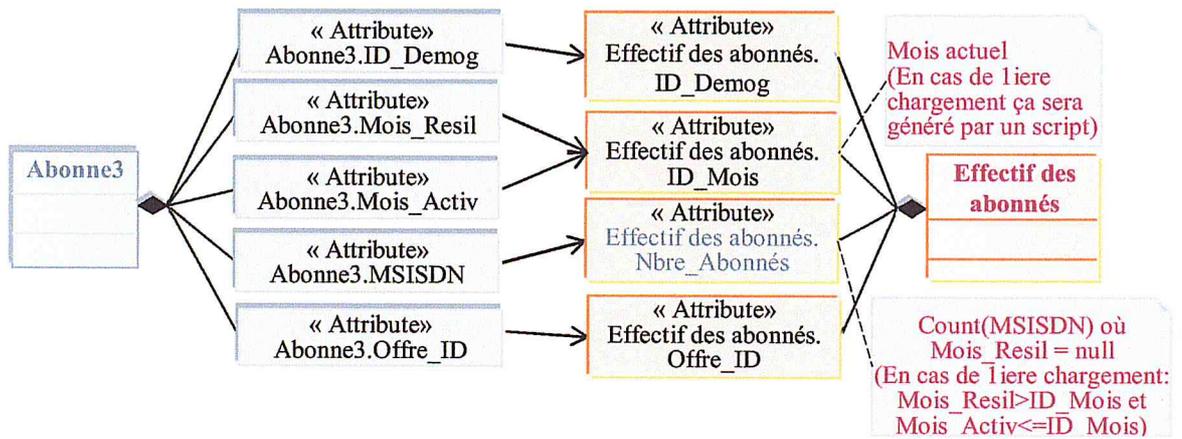


Figure 5.72. 3ème niveau de Mapping Package Map_EffectA

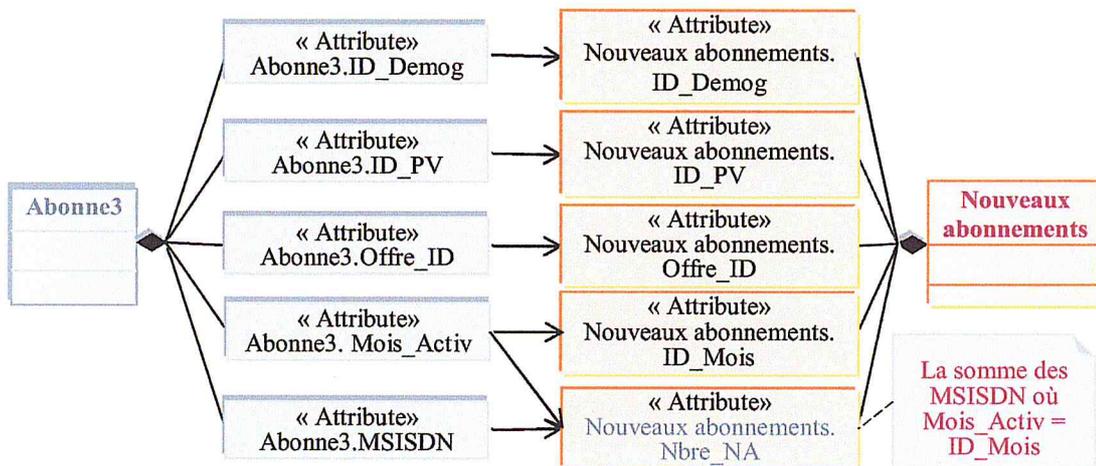


Figure 5.73. 3ème niveau de Mapping Package Map_NouvA

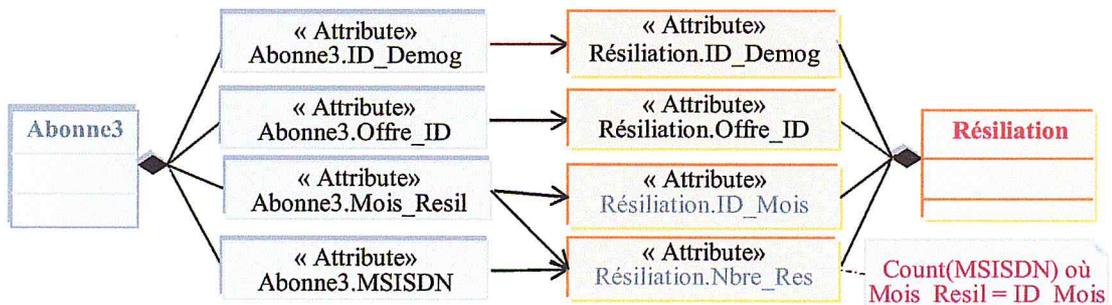


Figure 5.74. 3ème niveau de Mapping Package Map_Resil

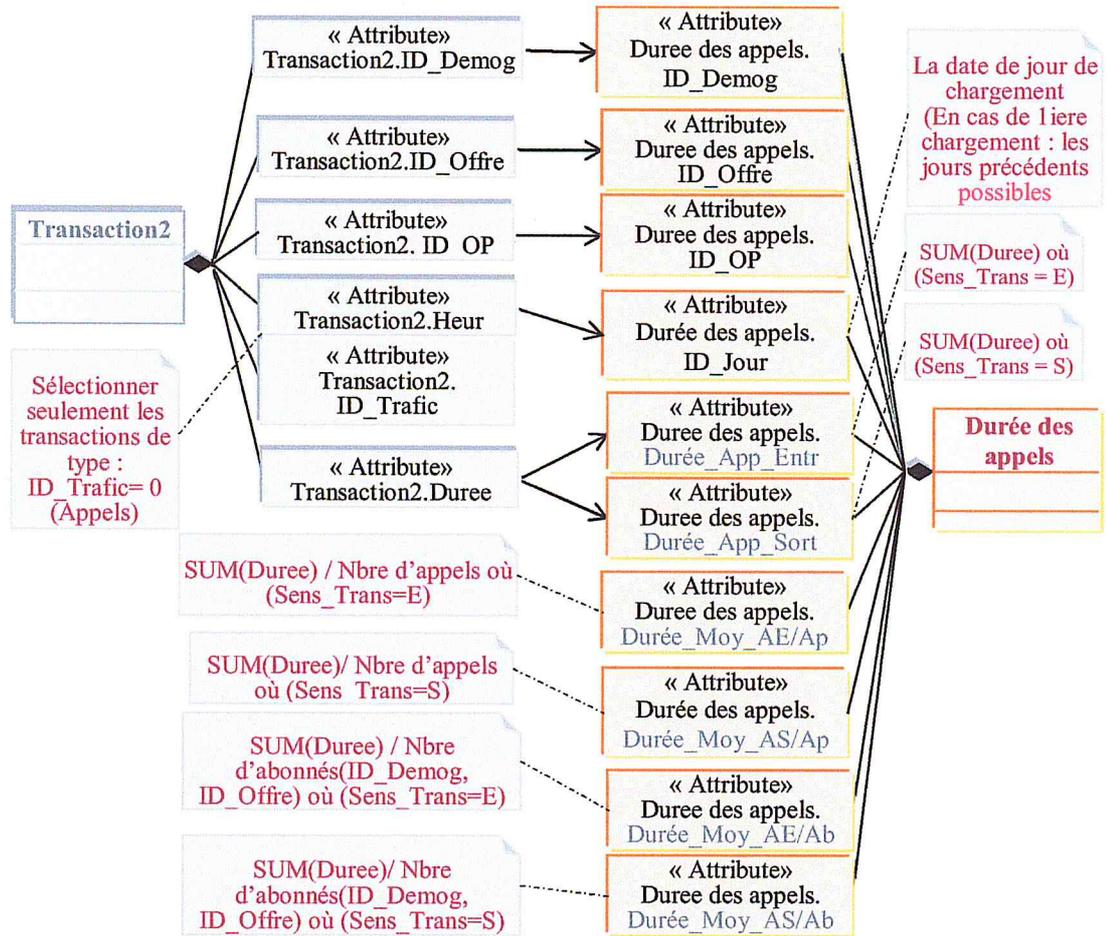


Figure 5.75. 3 ème niveau de Mapping Package Map_DAp

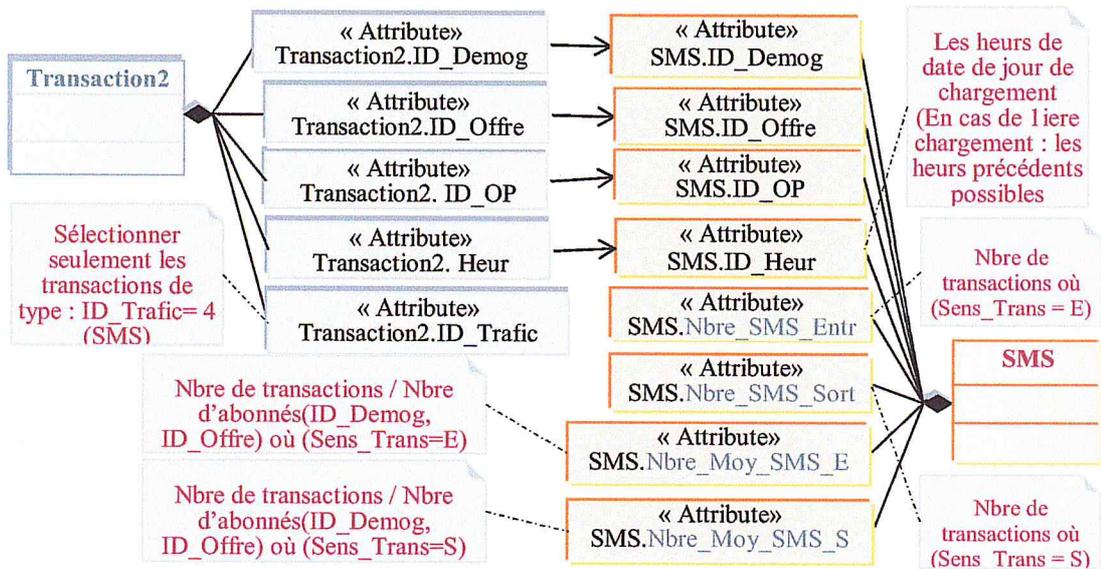


Figure 5.76. 3 ème niveau de Mapping Package Map_SMS

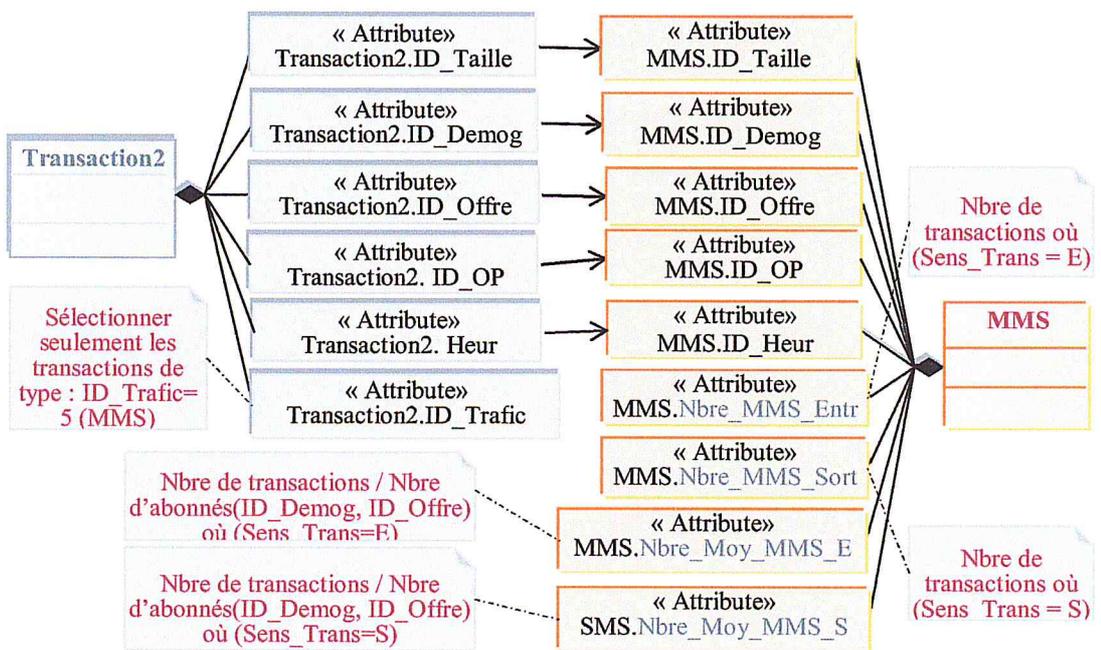


Figure 5.77. 3 ème niveau de Mapping Package Map_MMS

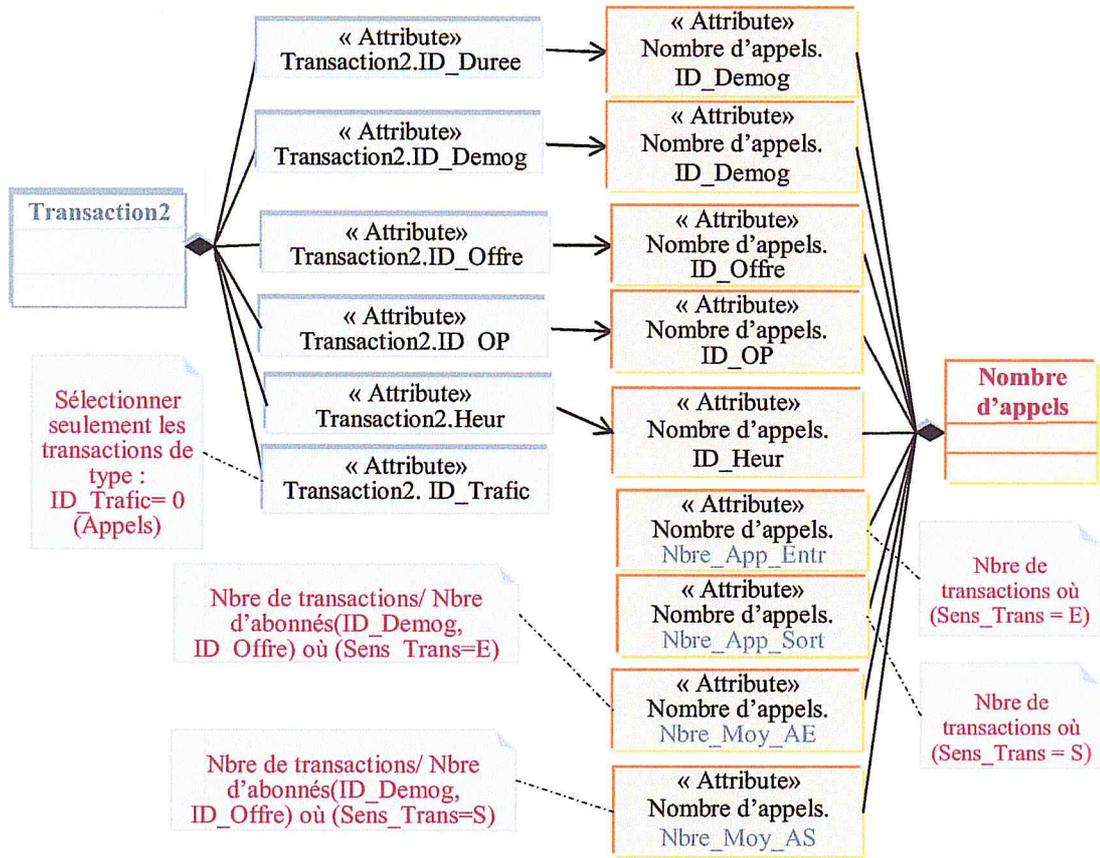


Figure 5.78. 3ème niveau de Mapping Package Map_NA

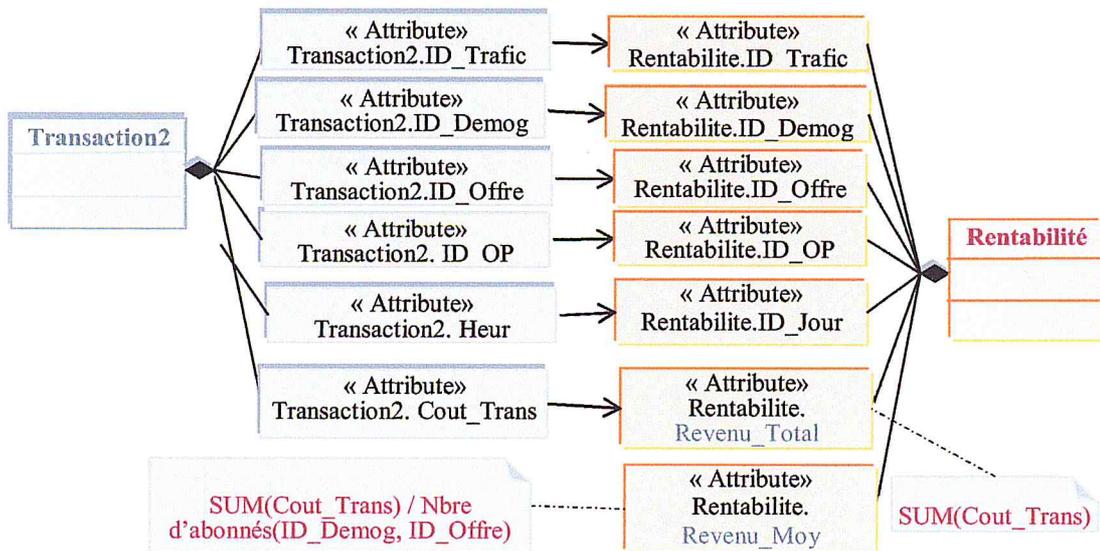


Figure 5.79. 3ème niveau de Mapping Package Map_Rent

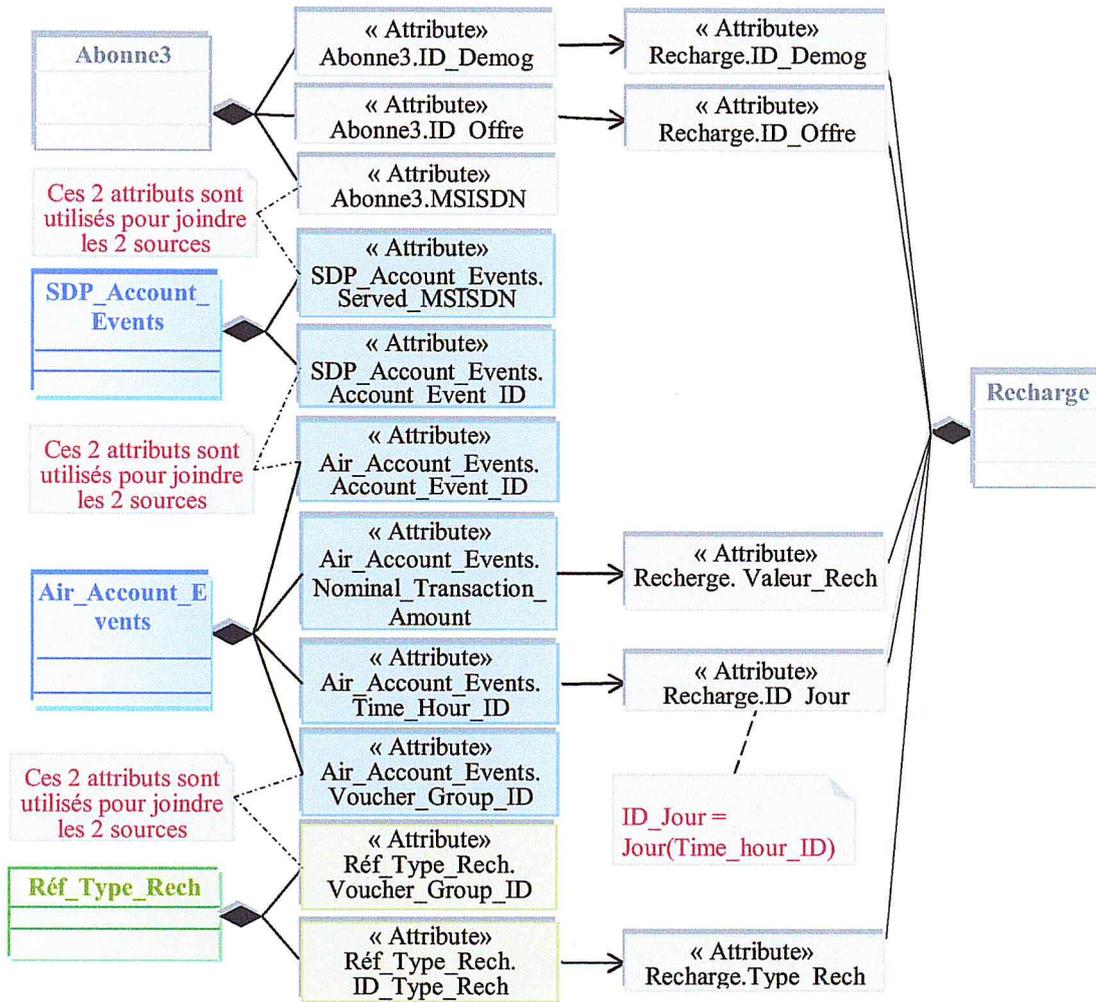


Figure 5.80. 3 ème niveau de Mapping Package Map_Rech1

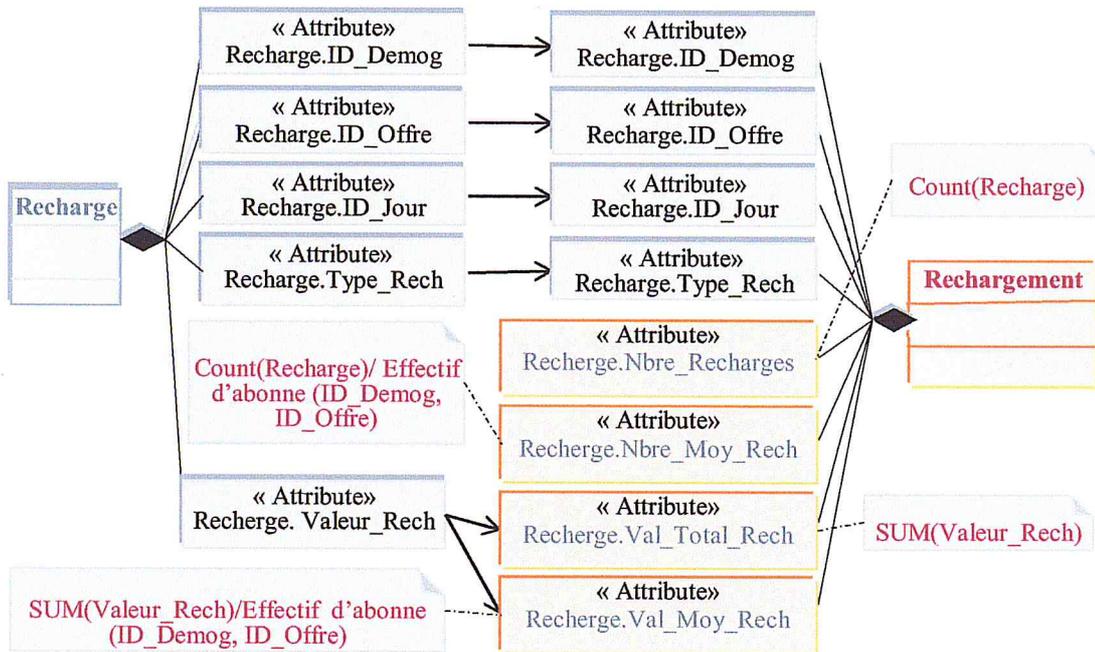


Figure 5.81. 3ème niveau de Mapping Package Map_Rech2

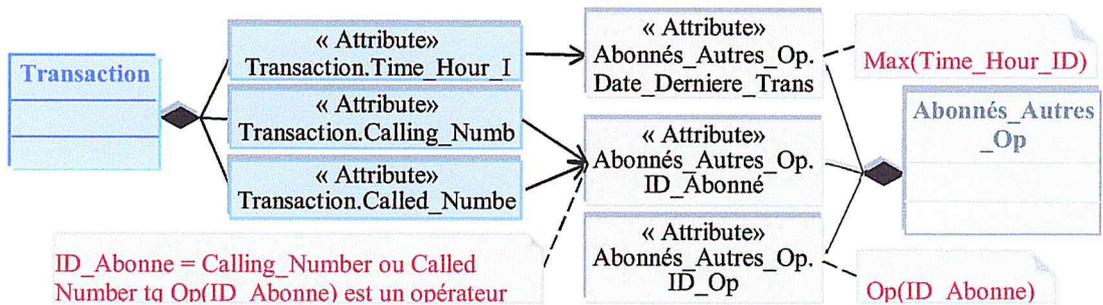


Figure 5.82. 3ème niveau de Mapping Package Map_AAO

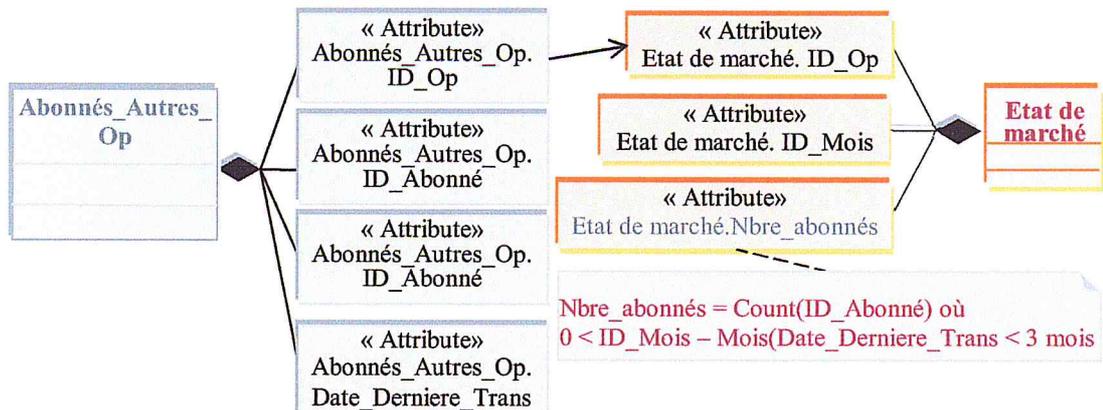


Figure 5.83. 3ème niveau de Mapping Package Map_EtatM

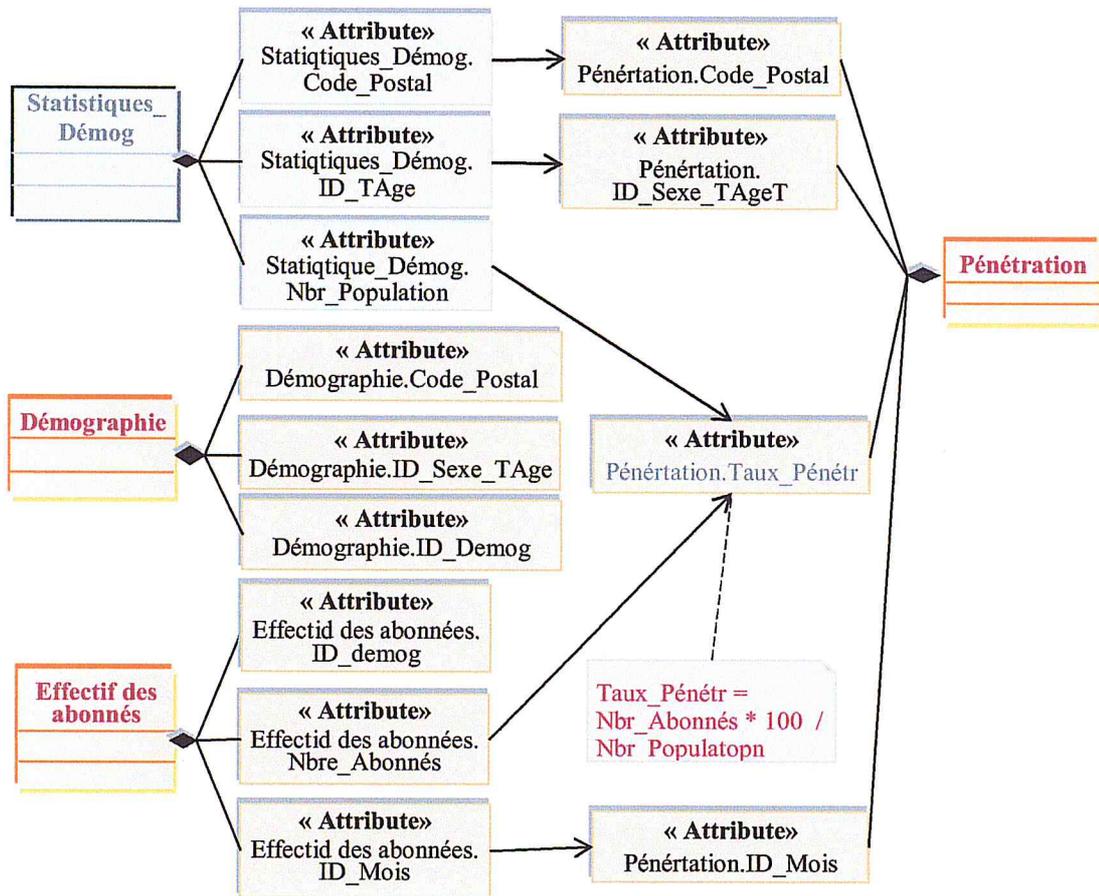


Figure 5.84. 3 ème niveau de Mapping Package Map_Penet

V.3.2.4. Diagramme d'activité d'ETL (ETLAD)

Le diagramme d'activité d'ETL (ETLAD) ne fait pas partie des diagrammes de la méthode DWEP, on a trouvé que c'est un diagramme intéressant pour simplifier et résumer le diagramme DMD d'un côté, et pour visualiser l'aspect dynamique et l'ordre d'exécution des tâches d'ETL d'un autre côté.

Le **diagramme d'activités** est l'un des diagrammes UML qui permet de Visualiser, spécifier, construire et documenter la dynamique d'une société d'objets. Le diagramme d'activités montre le flot de contrôle d'une activité à l'autre.

L'**activité** est une exécution non atomique en cours à l'intérieur d'un automate à état fini. Dans ETLAD, une activité regroupe un ensemble de processus de transformation qui ont la même date d'exécutions (chaque jour, chaque mois, ...).

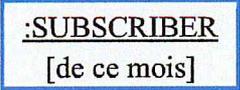
L'**états d'activité** est une tâche non atomique, décomposable. Dans ETLAD, un état d'activité représente un processus de transformation (Mapping Package) du diagramme DMD.



Map_Demog()

Figure 5.85. une état d'activité

Les **objets** peuvent être représentés dans un diagramme d'activité. Ils sont soit entrants (Ressources nécessaires à l'exécution d'une activité) ou sortants (Résultat d'une action) d'une action. Dans ETLAD, un objet représente une table de dimension, de fait, d'un système source ou d'ODS, il est soit une source ou une destination.



:SUBSCRIBER
[de ce mois]

Figure 5.86. Un objet du diagramme d'activité

Un diagramme ETLAD se décompose en trois travées (swimlanes) qui montre le lieu d'un objet ou d'exécution d'une action : Source, ETL et Data warehouse

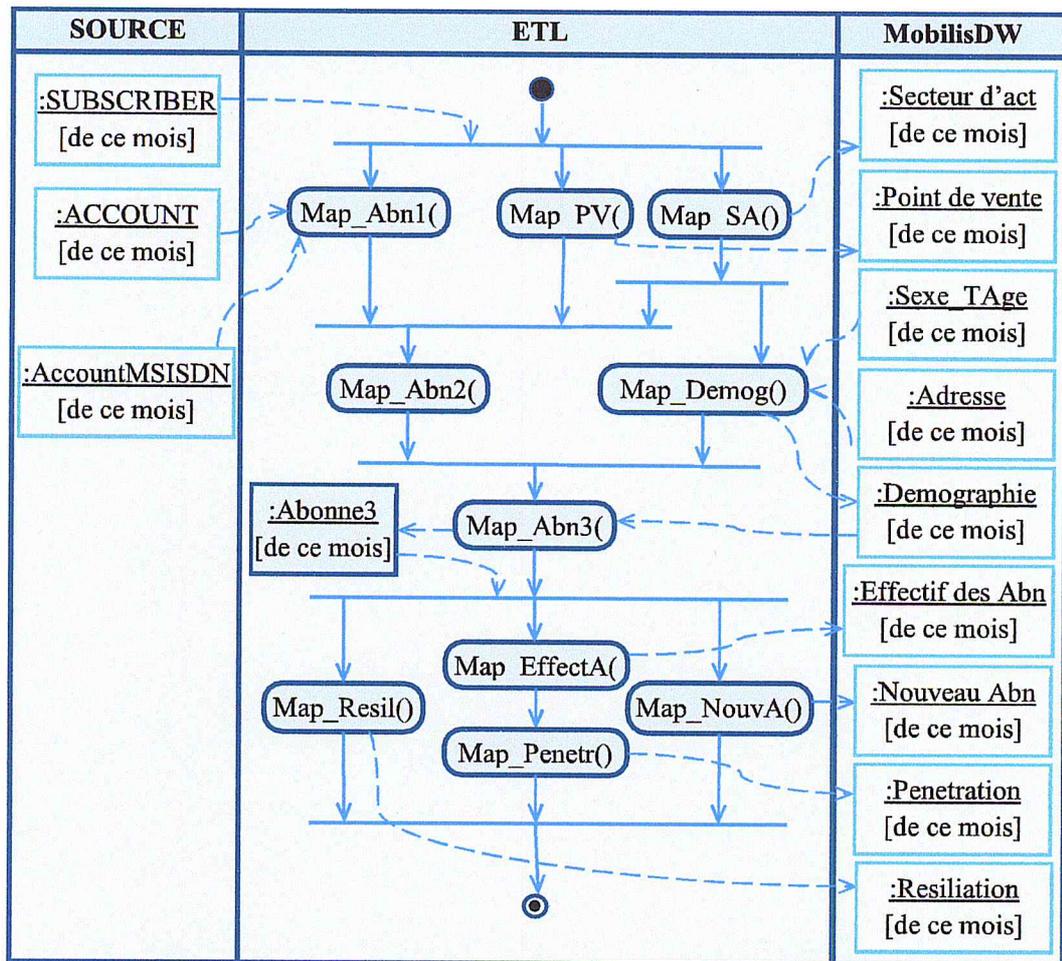


Figure 5.87. Diagramme d'activité de chargement mensuel des informations des abonnés

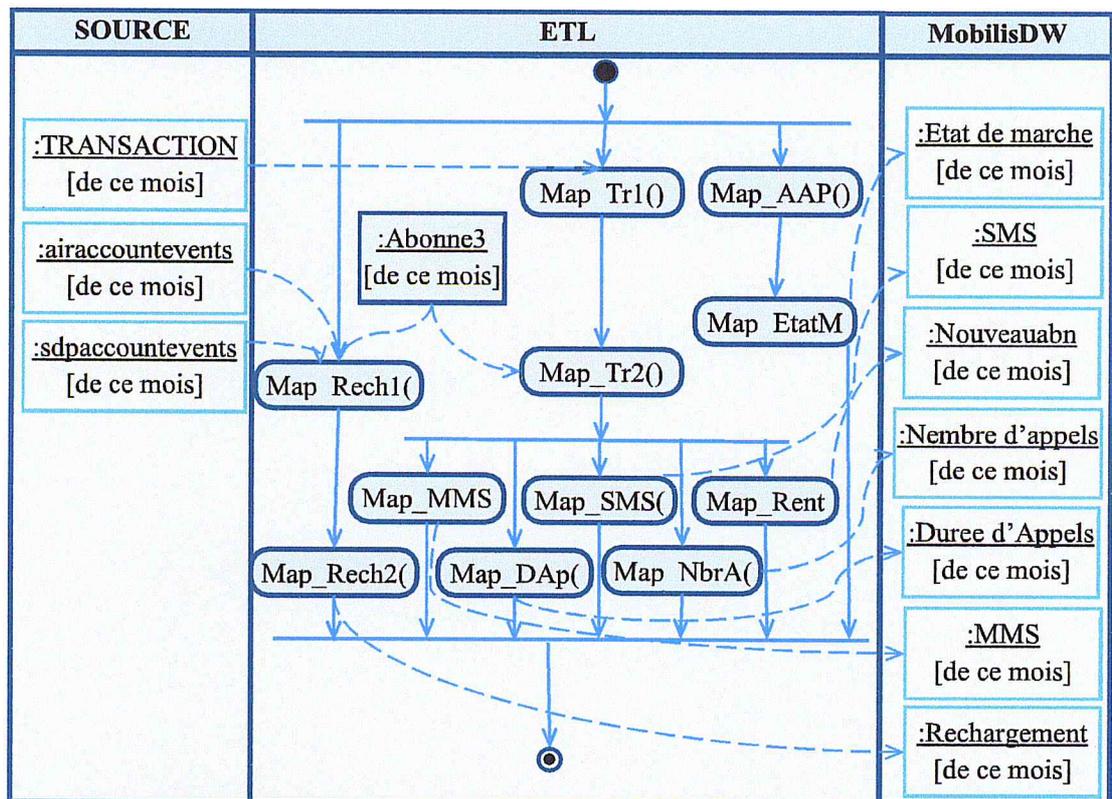


Figure 5.88. Diagramme d'activité de chargement quotidien des informations de trafic

V.3.3. Modélisation du système au niveau logique:

Jusqu'ici, on a modélisé le système conceptuellement, maintenant on entamera une autre vue de modélisation, c'est la vue logique qui est une phase intermédiaire entre le niveau conceptuel (monde réel) et le niveau physique.

Dans les SI classique (SIC) la modélisation logique permet de donner une vision sur les données comment elles seront implémentées et stockées physiquement. À ce niveau on parle seulement sur des tables (relations), des règles bien précises permettant le passage du niveau conceptuel au niveau logique, le type et la taille des attributs sont spécifiés et les règles d'intégrités sont définies.

La modélisation logique des SID repose sur les mêmes principes qu'un SI classique mais elle est plus délicate, de fait qu'un SID se compose d'un ensemble de sous-système (Source, DW, ETL ...), chaque sous système a ses particularités, cette particularité nécessite un modèle logique spécifique. Par exemple pour les systèmes sources la modélisation logique est la même qu'un SIC car ils sont de même nature (système OLTP). Pour le DW, les clés (type et taille), le modèle (en étoile, en flocon ...) sont choisis, pour l'ETL des stéréotypes sont introduits pour représenter les opérations relationnelles (jointure, agrégation ...) de transformation appliquées sur les données durant la phase d'alimentation du DW.

Pour la conception logique, UML repose sur un profil (profil UML pour les bases de données) proposé par « Rational Rose », ce profil offre un ensemble de stéréotypes pour représenter le schéma relationnel d'une base de données («Table», «Vue», «PrimaryKey» ...). Des icônes peuvent être associés aux stéréotypes (Figure 5.88)

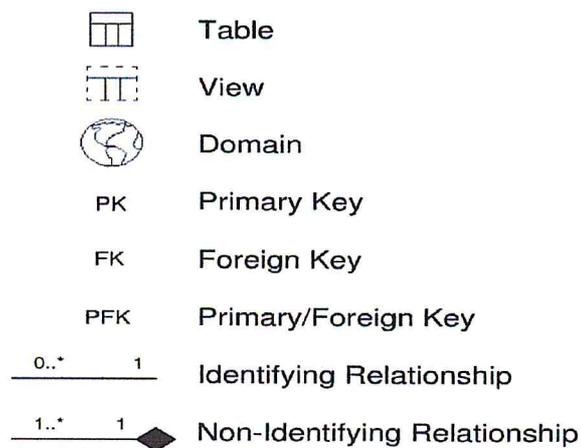


Figure 5.89. Stéréotypes pour la modélisation logique ses icônes

Dans qui suit, on présentera la modélisation logique du système.

V.3.3.2. Schéma logique des sources (SLS):

Les sources de données sont des systèmes de type transactionnel (OLTP), donc la modélisation logique des sources est similaire à n'importe quel SIC.

En réalité ces sources sont des systèmes déjà conçus, donc pour les modélisés on a suivi le sens inverse (physique, logique, conceptuel), l'alimentation du notre DW nécessite 10 tables réparties sur 3 sources (DWS, DData, MINSAT), on présentera pour chaque source ces tables puis on donnera un schéma globale dont les noms des tables sont exprimés sous la forme (Nom de la source. Nom de la table).

Conception et Réalisation d'un Système décisionnel pour le marketing

✚ SLS.DWS :

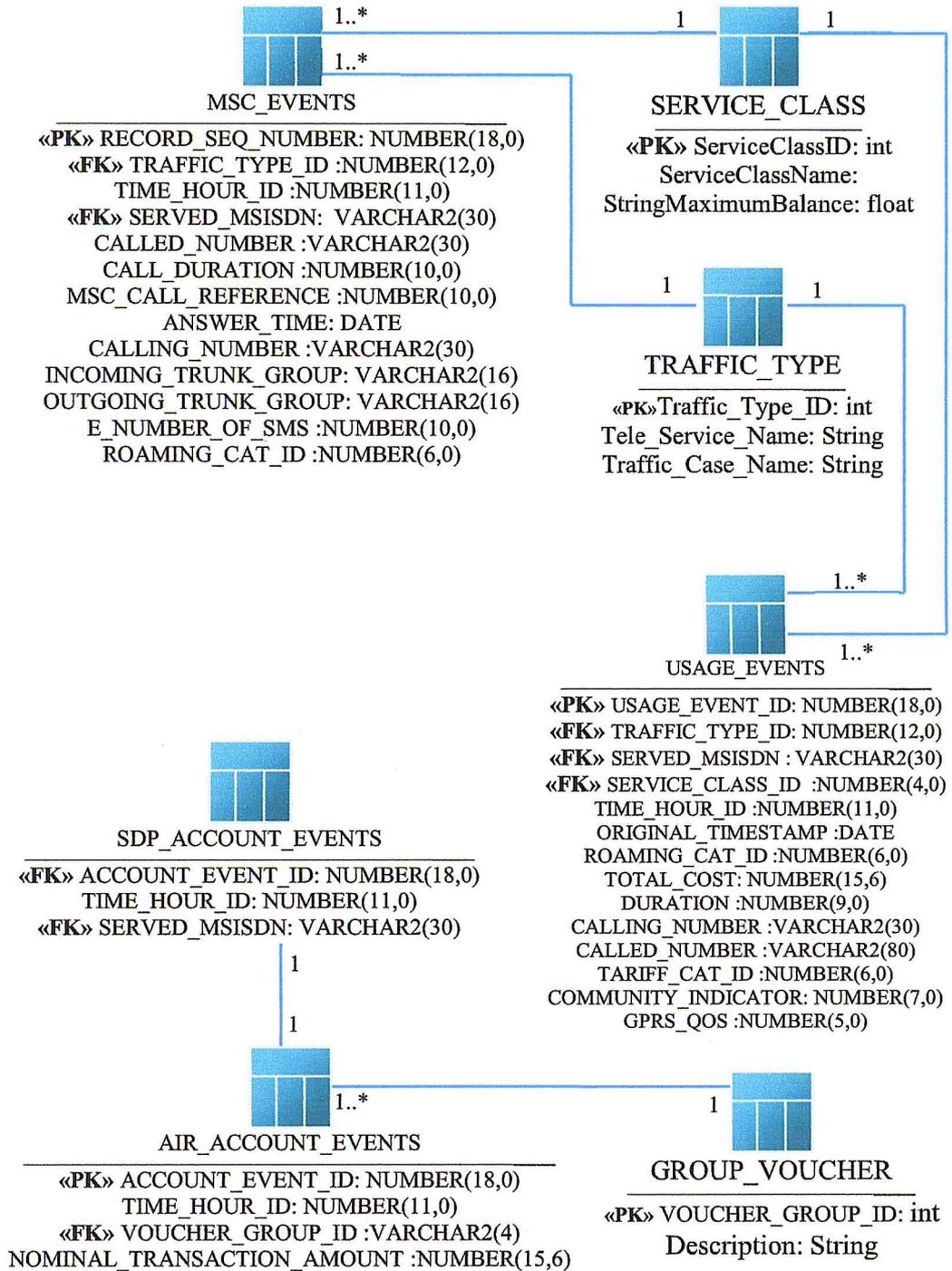


Figure 5.90. Schéma logique du DWS

SLS.MINSAT :

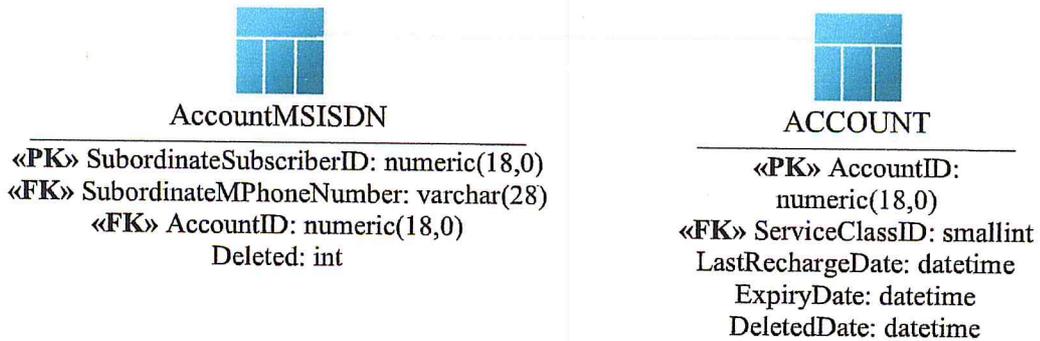


Figure 5.91. Schéma logique duMINSAT

SLS.DDATA

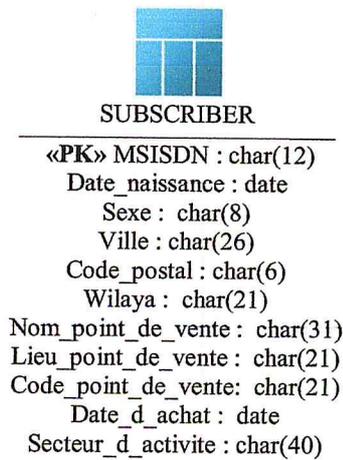


Figure 5.92. Schéma logique duDData

SLS.PLATFILE:

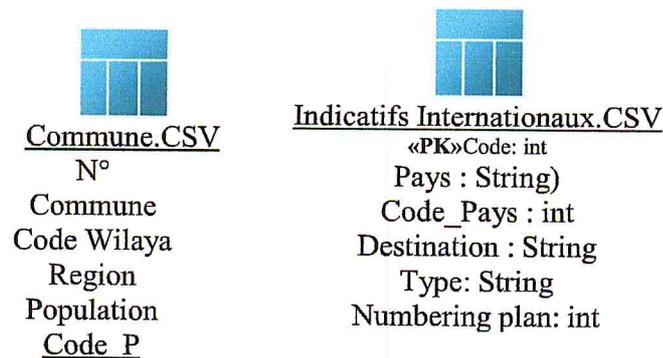


Figure 5.93. Schéma logique des fichiers plat.

✚ Schéma logique des sources (SLS) globale :

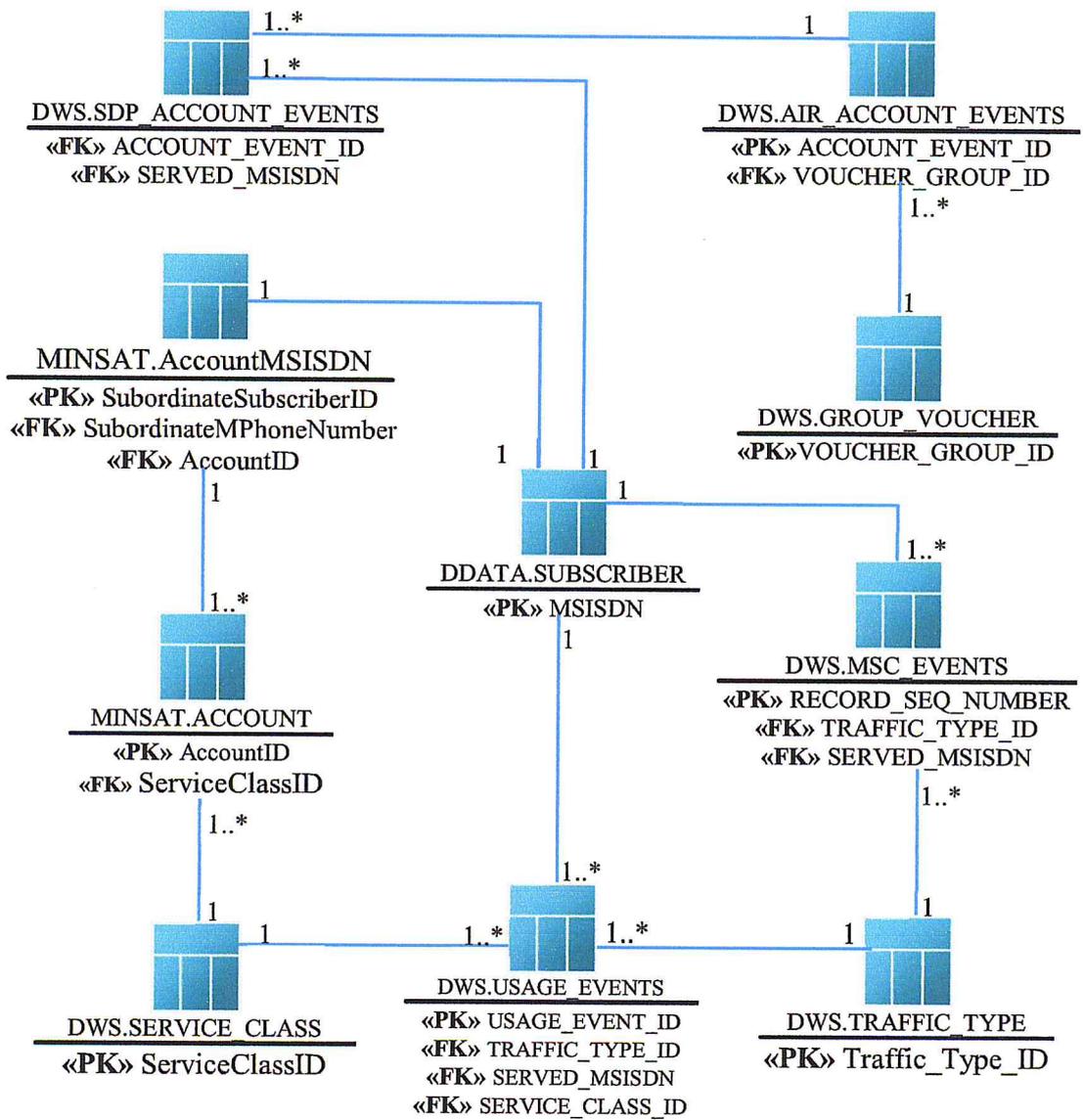


Figure 5.94. Schéma logique de toutes les sources

V.3.3.3. Schéma logique du DW (DWLS) :

Les premières fondations de la modélisation dimensionnelle ont été définies par *Ralph Kimball* au niveau logique. Il a introduit le fameux modèle en étoile et ses variantes (modèle en flocon et modèle en constellation). Contrairement au niveau conceptuel où les notions de la modélisation dimensionnelle ont été présentées indépendamment d'aucune technologie ou modèle spécifique. Au niveau logique, la technologie de stockage est choisie (ROLAP, MOLAP, HOLAP ...). Si la technologie ROLAP ou HOLAP est choisie, alors il faut spécifier le modèle logique (modèle en étoile, modèle en flocon). Plus la technologie de stockage et le modèle, les clés (type et taille), le nombre des niveaux... sont aussi identifiés.

Ces choix reposent sur des critères, comme la taille, la vitesse, le budget, la priorité des faits/dimensions par rapport aux autres ... Par exemple la technologie MOLAP donne plus performances (Vitesse) mais elle n'est pas recommandée si la taille du DW est importante. La même chose pour le modèle logique, le modèle en étoile dénormalise les tables de dimension pour minimiser le nombre de jointures (nombre d'accès au disque) mais cela engendre un gaspillage d'espace de stockage. Donc le compromis temps/taille joue un rôle important pendant les sélections à ce niveau.

Nous pensons que ces choix peuvent être automatisés par un outil qui a comme entrées le modèle conceptuel du DW, les critères de choix, et comme sortie la technologie de stockage et le modèle logique satisfaisants les critères entrées.

Nous pensons que dans le même DW on peut utiliser les 2 modèles (flocon et étoile). C à d : on éclate des dimensions et on laisse des autres, par exemple dans notre cas la dimension temporelle du modèle conceptuel est éclatées en 3 tables (Heure, Jour, Mois) car il y a des faits qui sont analysés selon la granularité (heure), d'autres (jour) ou (Mois). Pour La dimension Adresse on a préféré de la dénormalisée puisque tous les faits utilisent cette dimension avec le même niveau de granularité. Mais une question se pose (Si au future ce DW sera évolué avec des faits analysés par Année ou wilaya, est ce que on modifiera les dimensions existes ?). De notre point de vue, on pense qu'il n'est pas nécessaire de modifier la structure des tables mais il est mieux d'ajouter par exemple

une table « année » ou « Wilaya » et créer une redondance par ce que ces tables ne prennent pas beaucoup d'espace (Figure).

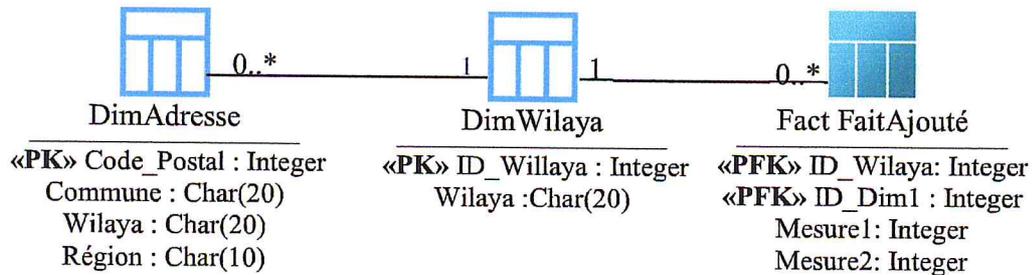


Figure 5.95. Exemple d'une évolution avec redondance

De plus, le choix des clés est un facteur crucial au niveau logique, *Ralph Kimball* conseille de choisir des clés numériques et séquentiels (C à d : il n'est pas essentiel de choisir des clés significatifs) et de grouper des dimensions qui sont souvent utilisées ensembles pour minimiser la taille des tables de faits. Par exemple dans notre cas on a groupé les informations démographiques dans la même table et le Sexe, Tanche d'âge dans la même table.

Dans ce qui suit, on présentera le schéma logique du DW, ce schéma sera présenté en 4 parties.

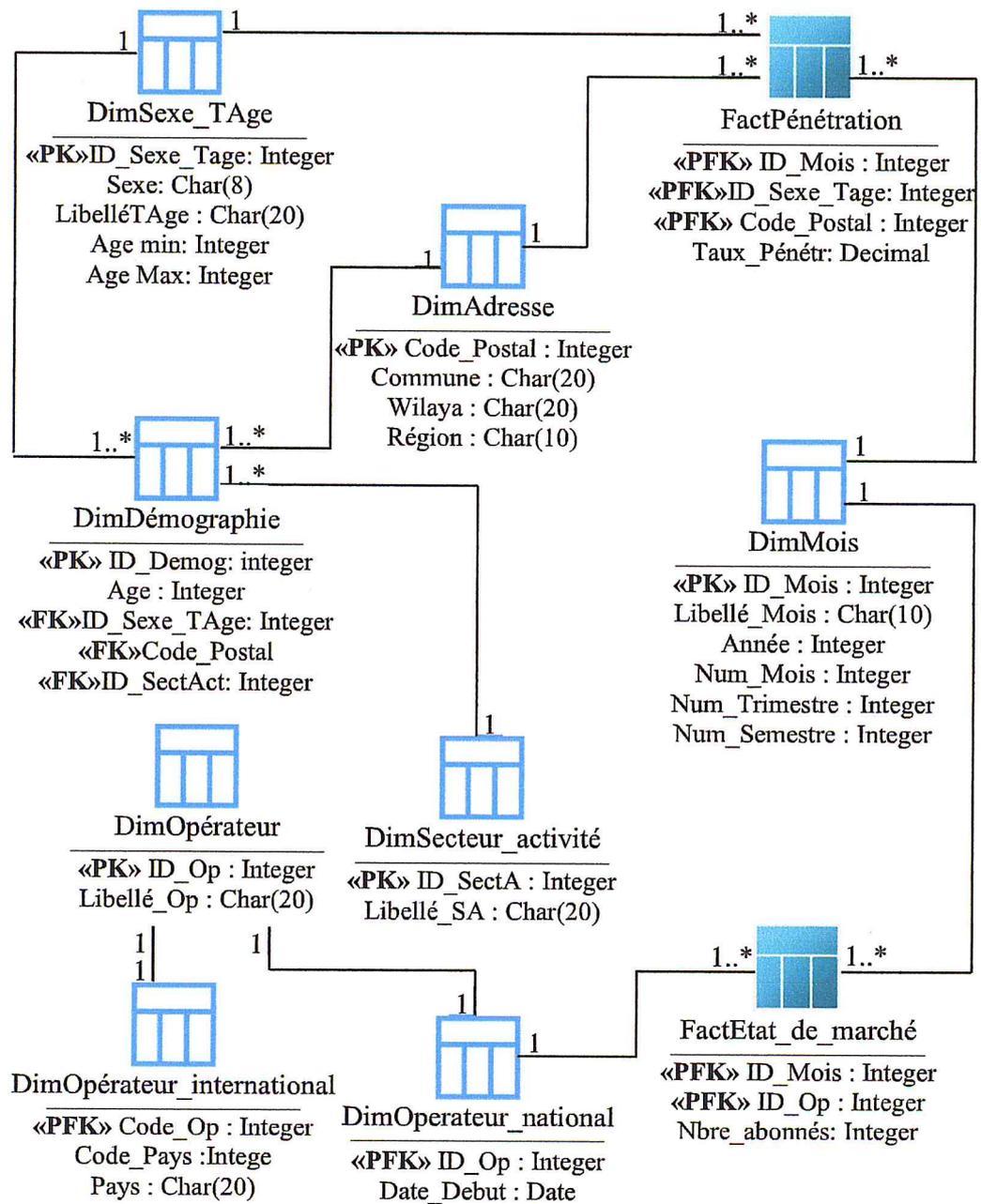


Figure 5.96. Schéma logique du DW (DWLS, partie 1)

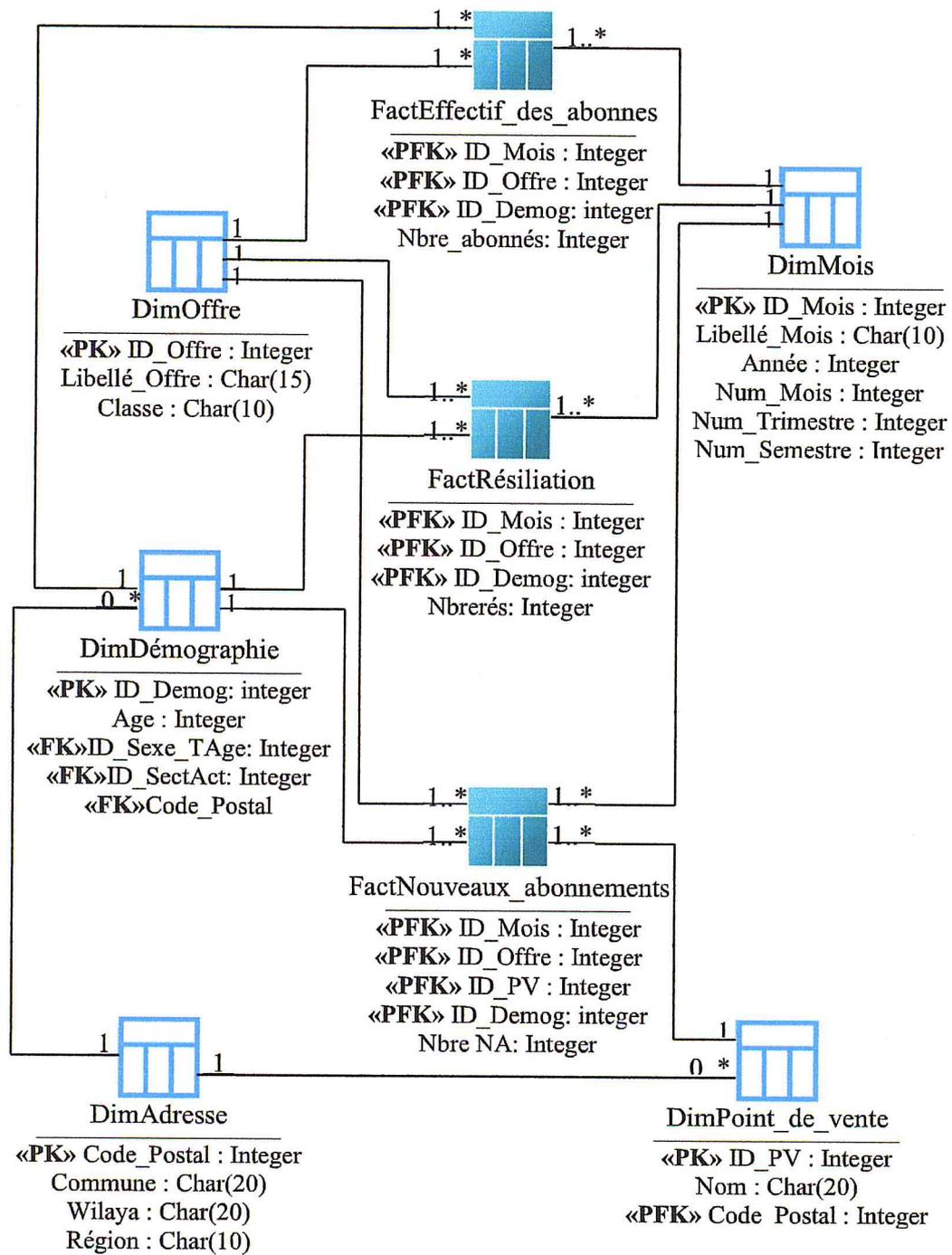


Figure 5.97. Schéma logique du DW (DWLS, partie 2)

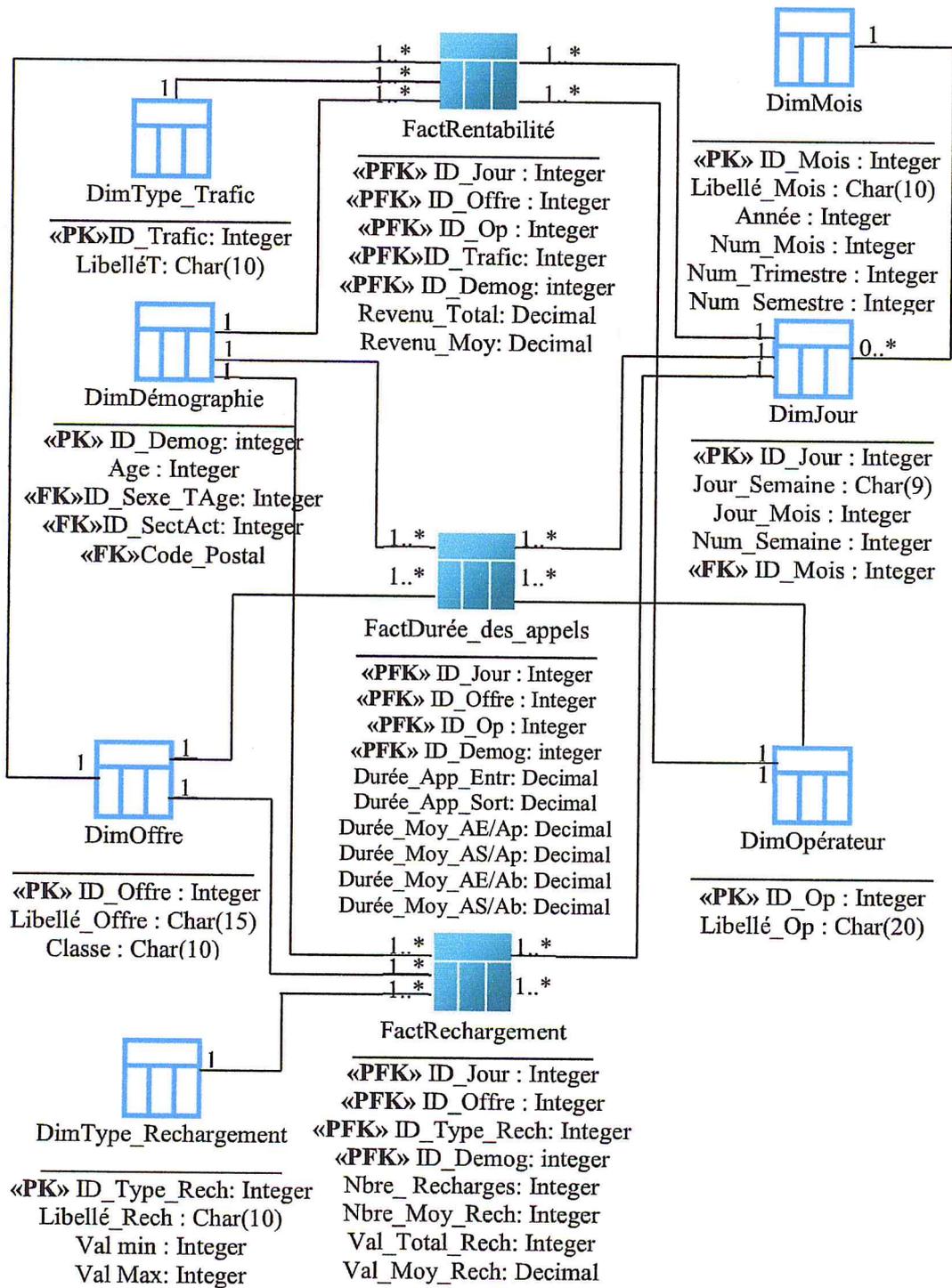


Figure 5.98. Schéma logique du DW (DWLS, partie 3)

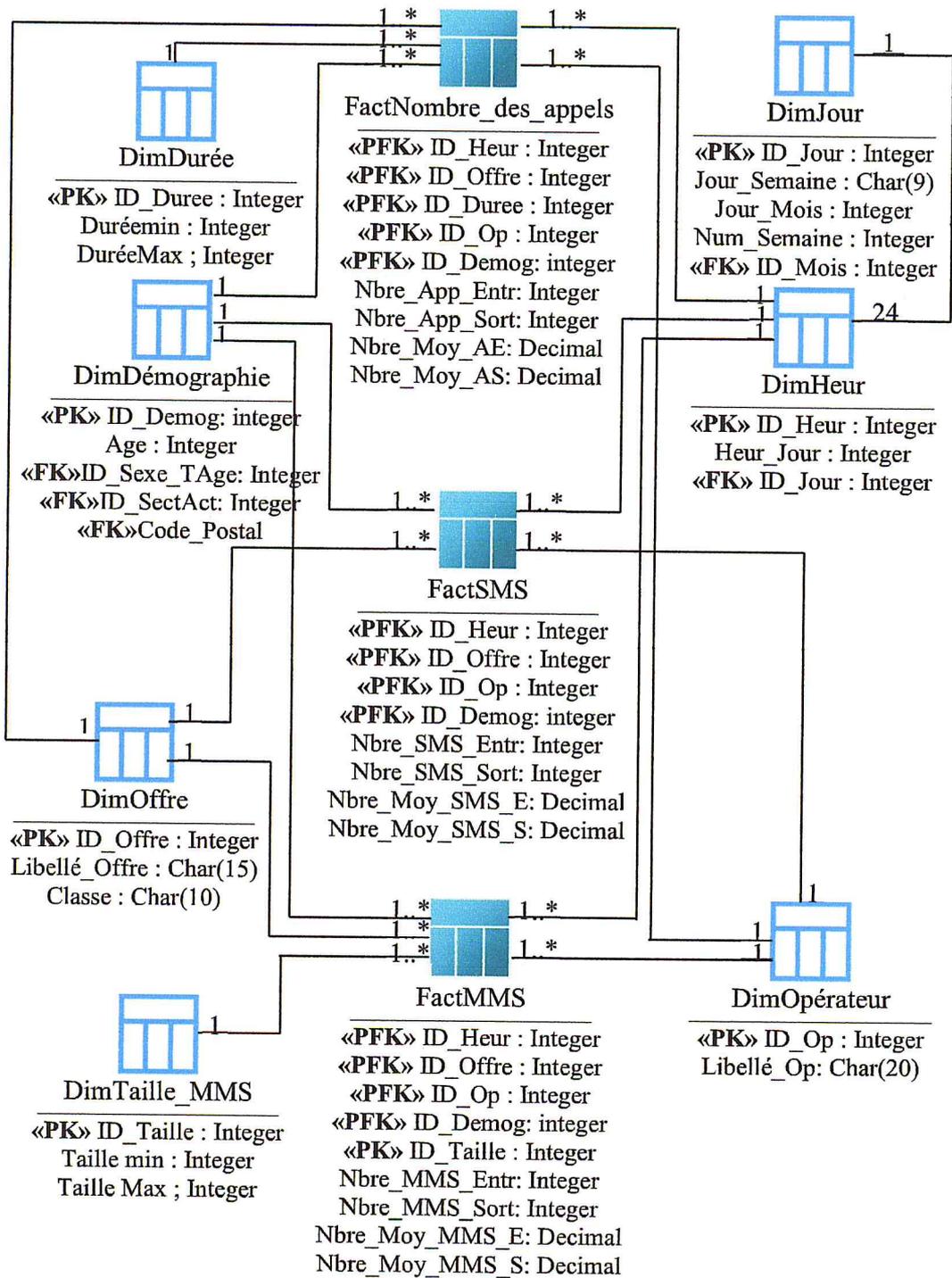


Figure 5.99. Schéma logique du DW (DWLS, partie 4)

V.3.3.4. Schéma logique d'ETL (ETL Process Diagram):

Durant le processus ETL les données sont extraites à partir des sources généralement hétérogènes, en suite elles passent par une séquence de transformations éventuellement très complexes. La conception d'un ETL passe habituellement par 5 phases qui sont :

- ✚ **Extraction des données :** identification des données utiles pour l'alimentation du DW.
- ✚ **Transformation des données :** les données extraites subies d'un ensemble de transformations pour convient à la forme cible, parmi ces transformations (Filtrage des données, nettoyage, conversion de format, calcule des valeurs dérivés, génération automatique des clés (Surrogate keys)).
- ✚ **Jointure des données.**
- ✚ **Sélection des cibles à charger.**
- ✚ **Chargement des données transformées dans les cibles.**

On a dit que le développement et la maintenance de l'ETL prendre des fois 80% des ressources nécessaires pour réaliser un projet d'SID. *Luján Mora* propose un profil UML pour aider les concepteurs à modéliser les ETL au niveau logique appelé (ETL Process «ETLP»). Sa proposition permet de décomposer un processus ETL complexes en un ensemble de processus simples, Il a basé sur son expérience dans le domaine de BI pour définir un ensemble de mécanismes (Transformation) permettant de couvrir les problèmes rencontrés pendant le processus ETL. Ces mécanismes sont présentés par des classes stéréotypées (tableau 5.7) Cette approche permet aux concepteurs de facilement concevoir et maintenir un ETL.

Mécanismes d'ETL:

Mécanisme d'ETL	Description	Icons
«Aggregation»	Agrégation et Calcule des données	
«Conversion»	Conversion de type, format de données ou dérivation des données depuis des données existantes	
«Filter»	Filtrage et vérification des données	
«Incorrect»	Déclare une exception sur des données fausses	
«Join»	Jointure de deux sources de données	
«Loader»	Chargement des données dans la cible	
«Log »	MAJ de journal d'activités des mécanisme ETL	
«Merge»	Intégration de deux sources de données ou plus ayant des attributs compatibles	
«Surrogate»	Génération d'une nouvelle clé	
«Wrapper»	Transforme une source de donnée plat en enregistrement	

Tableau 5.7. Les mécanismes d'ETL et leurs icones

Aggregation

Le mécanisme «Aggregation» grouper les données selon un critère donnée. Le concepteur peut définir les critères d'agrégation et la fonction de calcul (SUM, AVG, MAX/MIN, COUNT ...) dans un commentaire attaché au mécanisme

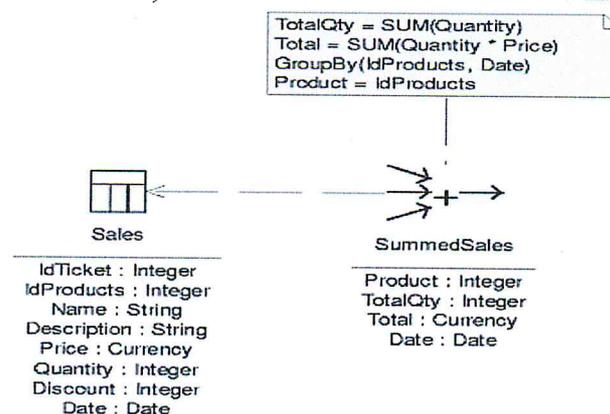


Figure 5.100. Exemple du mécanisme «Agregation»

✚ Conversion

Le mécanisme «Conversion» est utilisé pour changer les types ou les formats des données ou pour calculer et dériver une valeur à partir d'une donnée existe. Les conversions sont définies dans un commentaire comme une fonction appliquée sur l'attribut source. La syntaxe des conversions est: `Attribut_Cible = Fonction (Attributs_Sources)`.

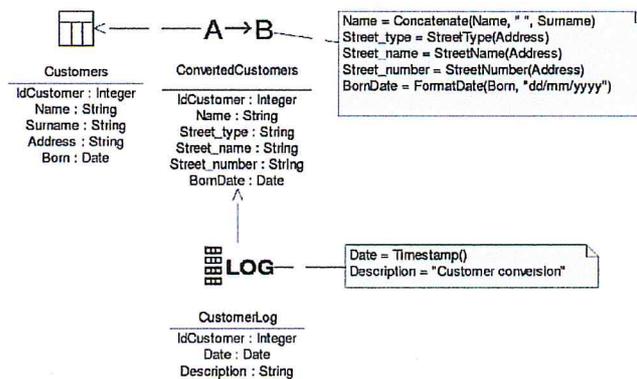


Figure 5.101. Exemple du mécanisme «Conversion» et «Log»

✚ Log

Le mécanisme «log» peut être connecté à n'importe quel mécanisme ETL comme un contrôle d'activité d'un autre mécanisme ETL. Ce mécanisme est utile pour produire des rapports opérationnels. Le concepteur peut ajouter n'importe quelles informations additionnelles dans un commentaire.

✚ Filter

Permet de Filtrer les données non désirées et vérifier si les données sont correctes sur une contrainte donnée. Ce mécanisme permet au concepteur de charger dans le DW seulement les données qui ont un certain niveau de qualité. Les vérifications sont définies par une note qui exprime une expression/fonction booléenne. Parmi ces vérifications (la vérification des valeurs null, valeurs aberrantes, valeurs hors plage ...)

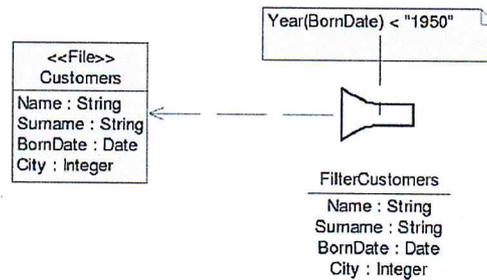


Figure 5.102. exemple du mécanisme «Filter»

✚ Join:

Le mécanisme «Join» Permet de joindre 2 sources reliées entre elles par un nombre d'attributs, les informations suivantes peuvent être ajoutées dans une note :

- ✓ Le type de jointure : Join (Expression de la condition), ou Join peut être (InterJoin, RightJoin, LeftJoin, FullJoin).
- ✓ Si LeftJoin, RightJoin ou FullJoin sont utilisées, alors le concepteur peut définir les valeurs substituent les valeurs inexistantes

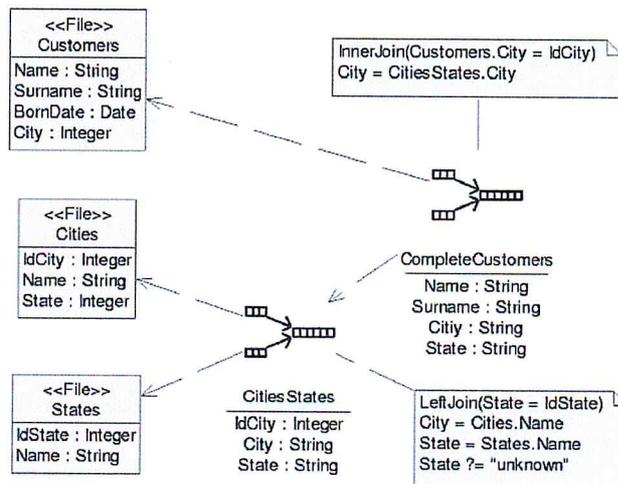


Figure 5.103. exemple du mécanisme «Join»

✚ Loader

Le mécanisme «Loader» charge les données dans les tables cibles (Table de Fait, table de dimension), chaque table cible doit être associée à un mécanisme Loader.

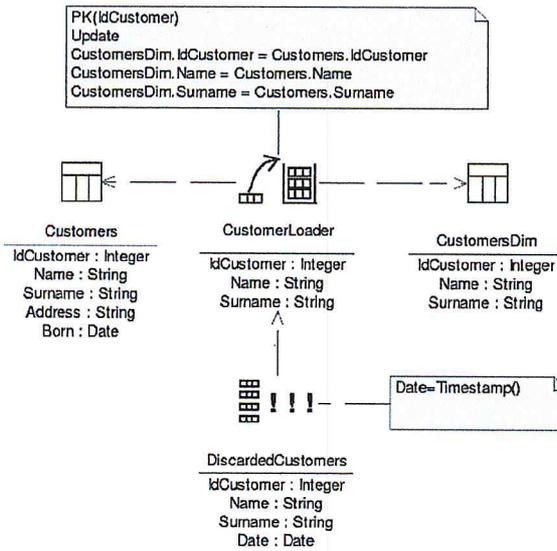


Figure 5.104. Exemple du mécanisme «Loader»

➤ Merge

Le mécanisme «Merge» intègre 2 ou plusieurs sources de données ont des attributs compatibles, les attributs utilisés doivent avoir les mêmes noms. S'ils n'ont pas les mêmes noms, ce mécanisme est précédé par un mécanisme « Convesion».

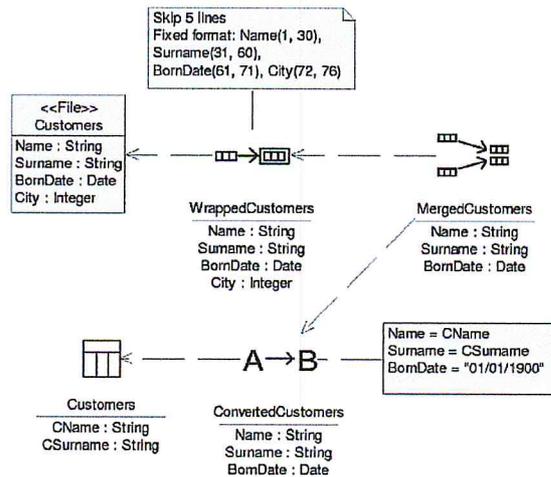


Figure 5.105. Exemple du mécanisme «Merge»

+ Wrapper

Le mécanisme «Wrapper» permet de définir les transformations requises pour extraire des données à partir d'une source de données Native dans une source de données basée sur des enregistrements.

+ Surrogate

Le mécanisme «Surrogate» Génère une clé unique. Les données chargées dans un DW viennent des sources multiples, chaque source à son propre codage, ce mécanisme permet de remplacer les clés des sources par une clé unique.

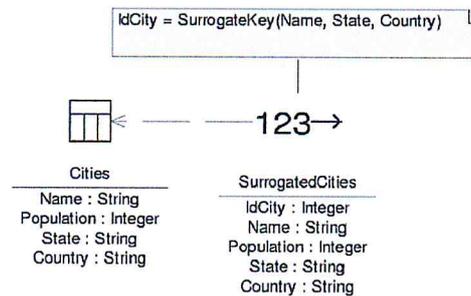


Figure 5.106. exemple du mécanisme «Conversion»

Conception et Réalisation d'un Système décisionnel pour le marketing

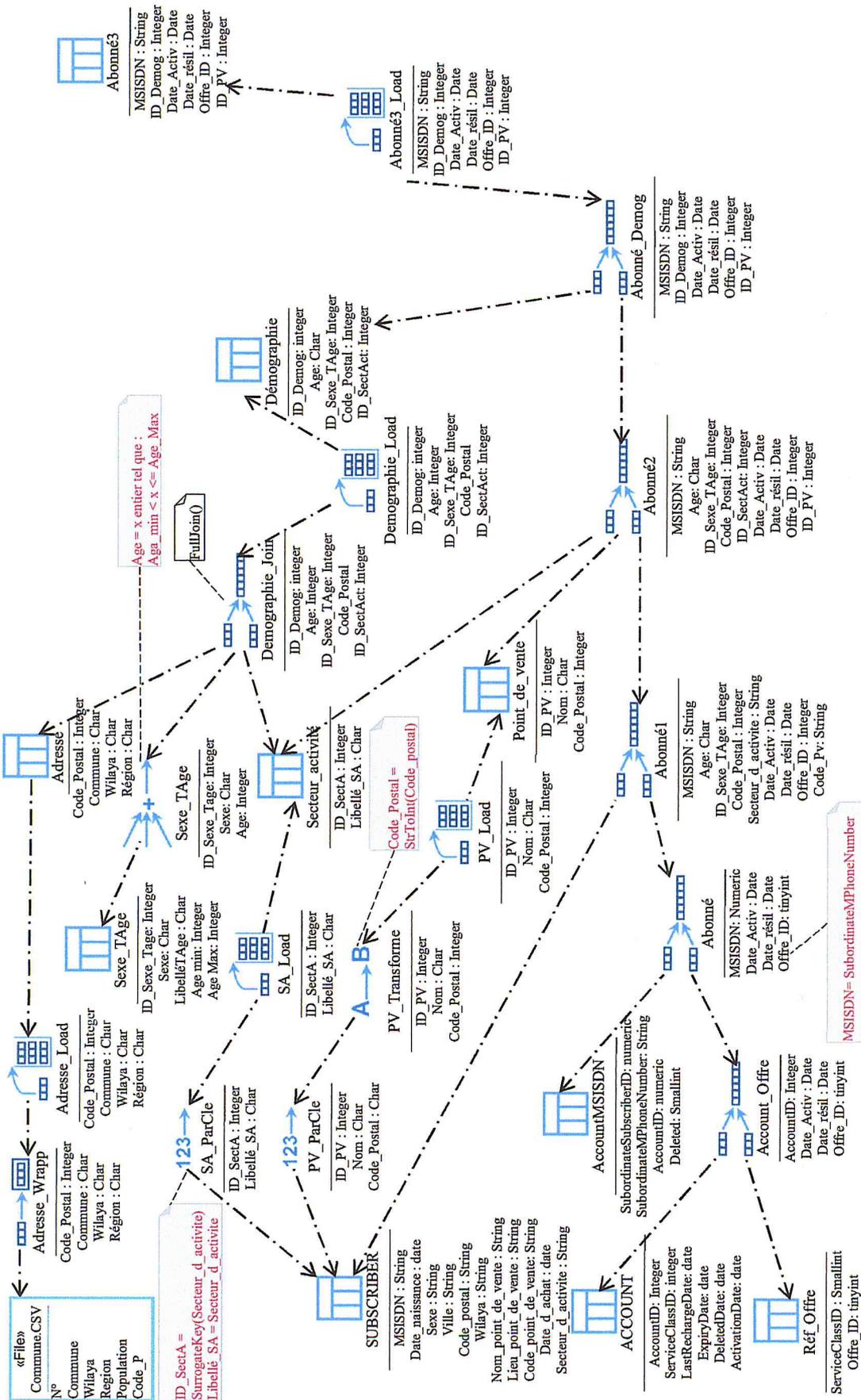


Figure 5.107. ETL Process des dimensionsd'abonné

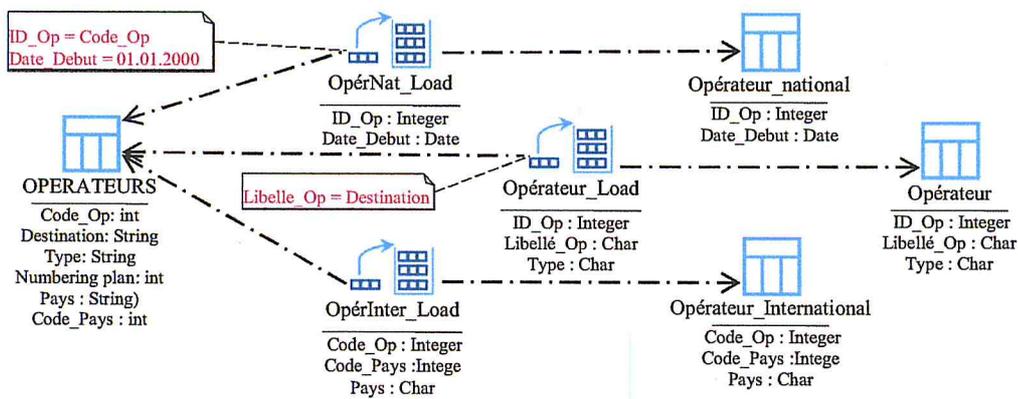


Figure 5.108. ETL Precess des dimensions Operateur

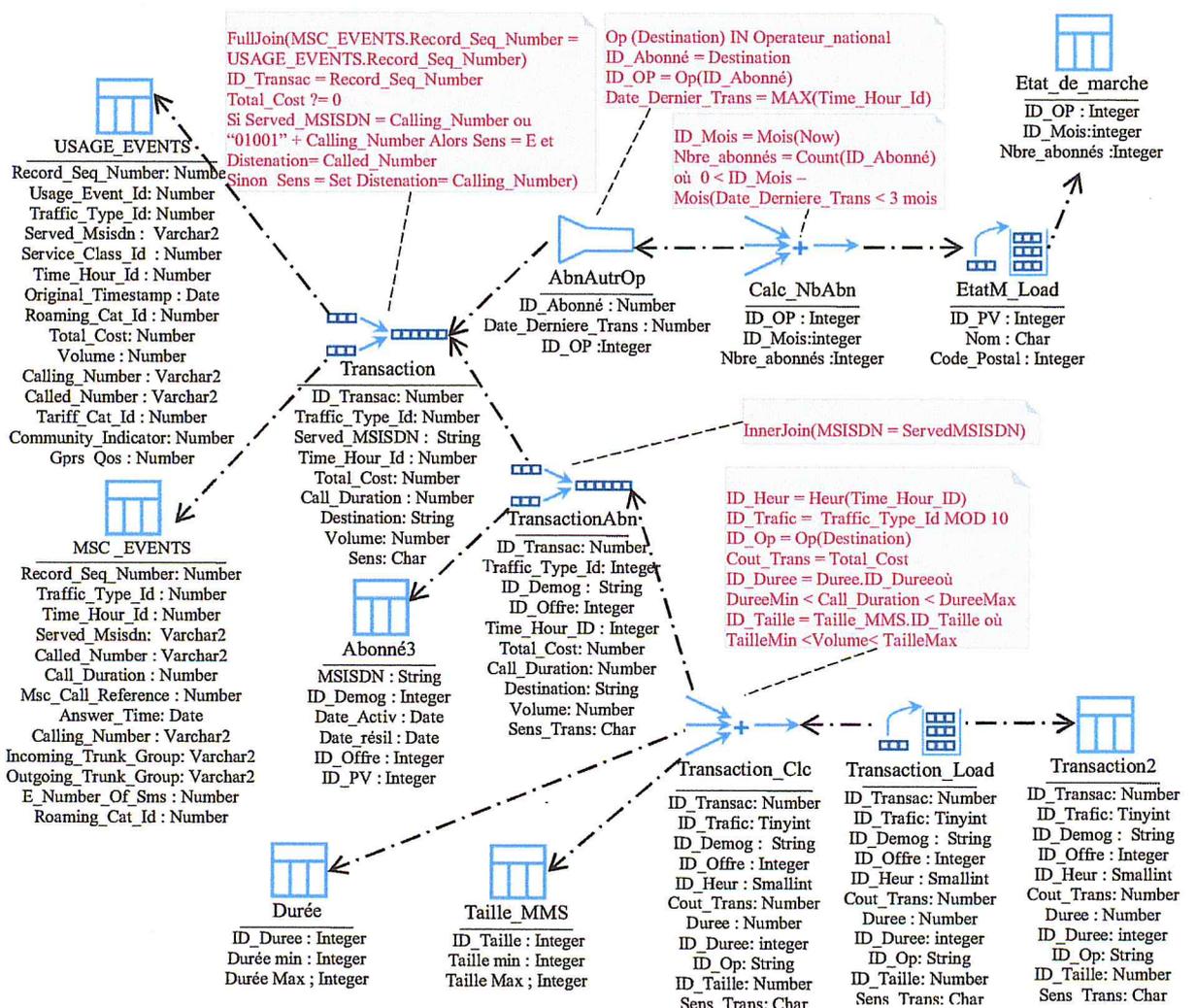


Figure 5.109. ETL Precess de Transaction et Etat de marché

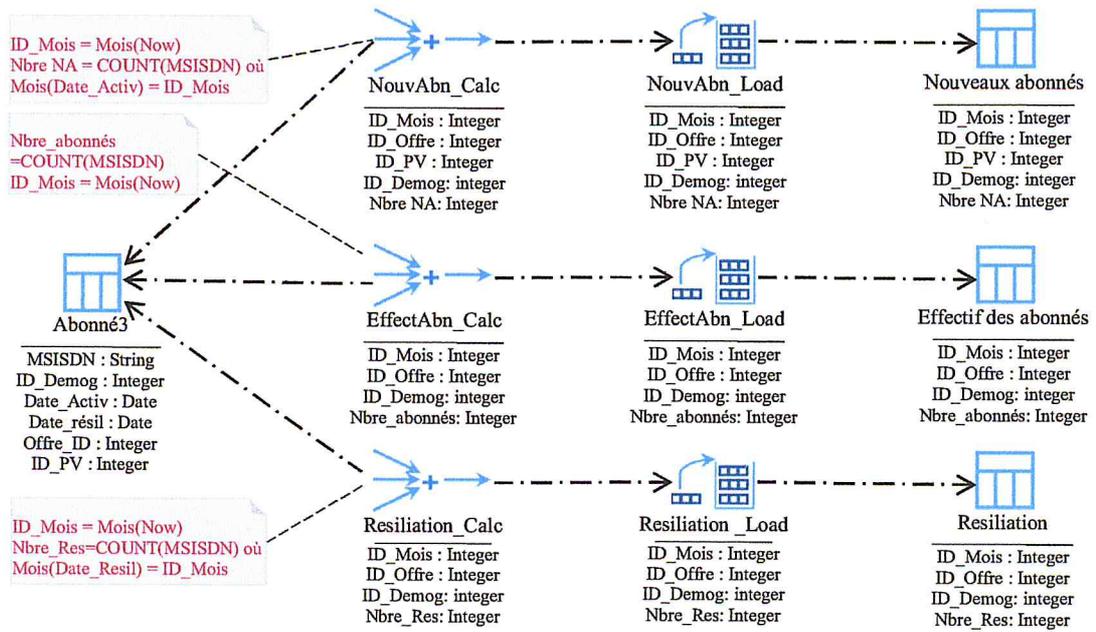


Figure 5.110. ETL Precess desfaits Répartition des abonnés

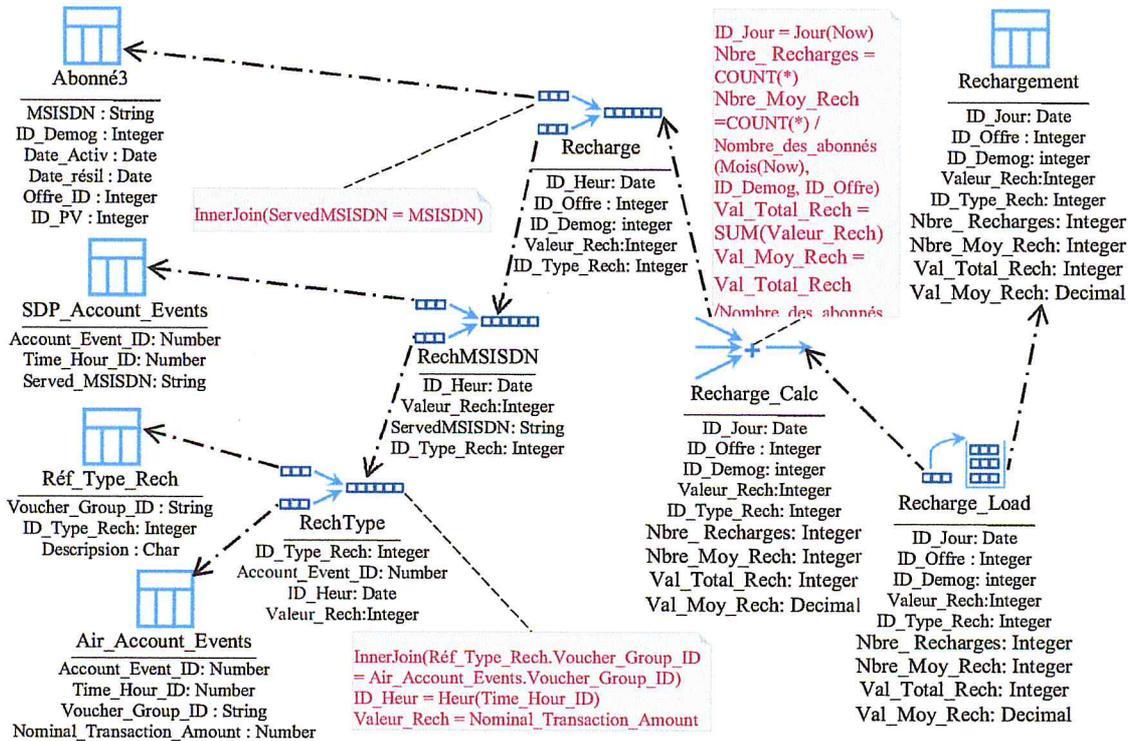


Figure 5.111. ETL Precess de Rechargement

Conception et Réalisation d'un Système décisionnel pour le marketing

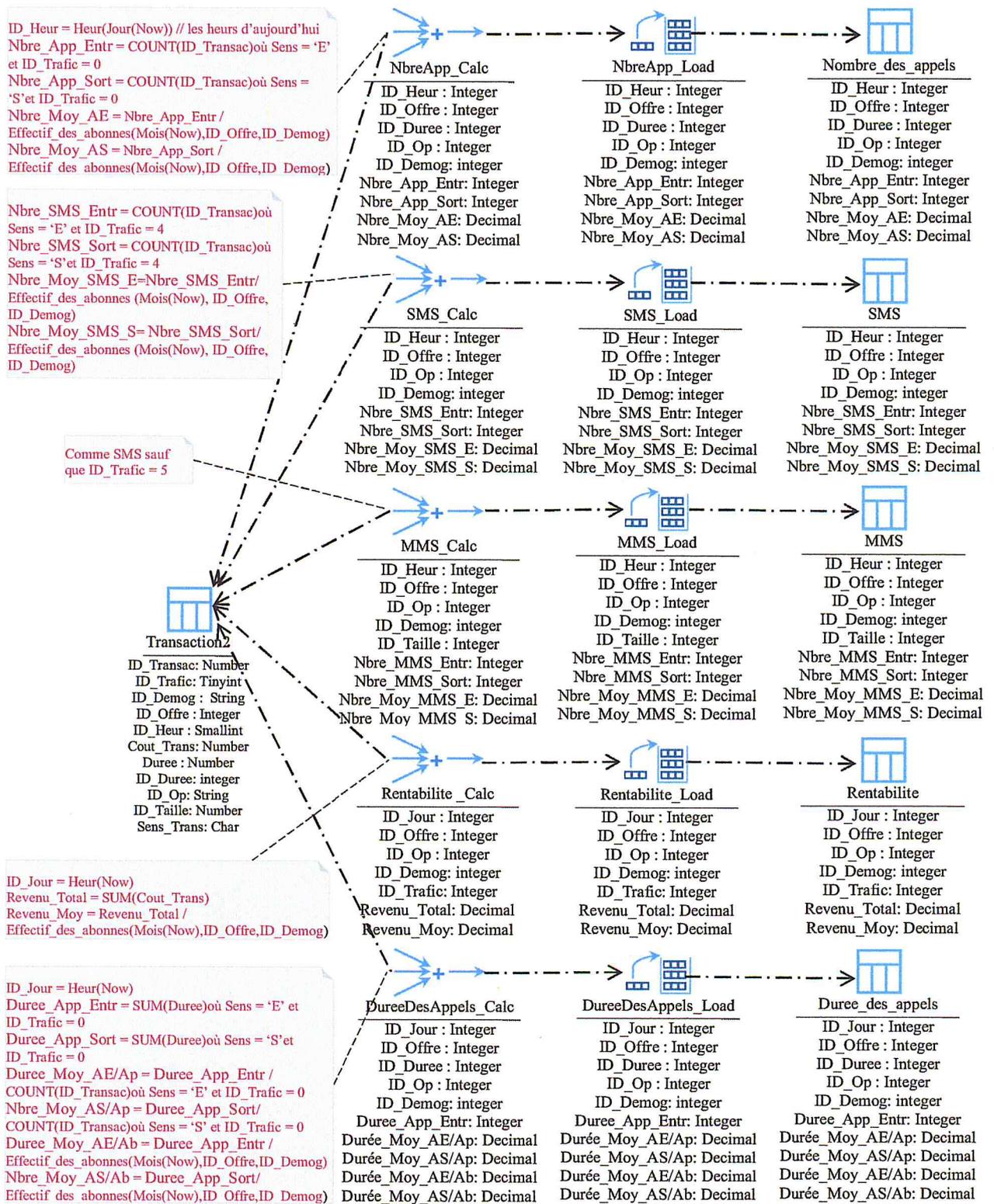


Figure 5.112. ETL Precess desfaits Consommation

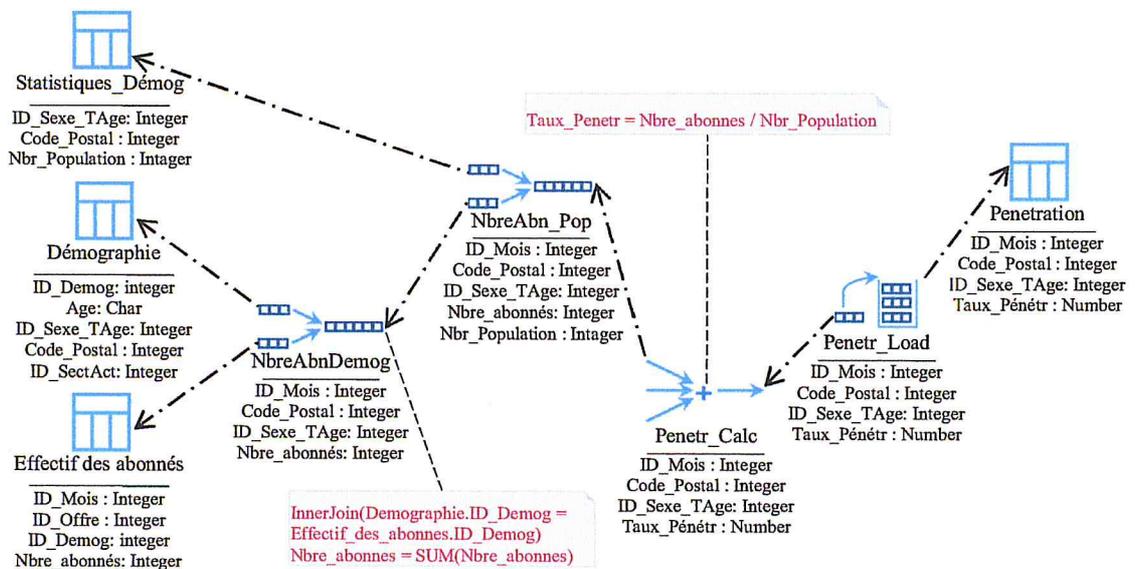


Figure 5.113. ETL Precess de Pénétration

V.4. Conclusion :

Dans ce chapitre, on a conçu et modélisé notre système au niveau conceptuel et logique. On a vu comment les diagrammes de la méthode DWEP permettant de mieux schématiser les différents composants d'un SID et surtout l'ETL qu'été la point faible de toutes les méthodes de conception d'un SID. En plus, on a remarqué que la stratégie de la modélisation en plusieurs niveaux de détail cache la complexité des modèles.

Le chapitre suivant c'est le dernier chapitre dont on va engager dans l'implémentation du système.

CHAPITRE V.
Implémentation
et Mise en œuvre
du système

CHAPITRE VI. Implémentation et Mise en œuvre du système

VI.1. Introduction :

Ce chapitre c'est le fruit du notre projet. Dans ce chapitre on entamera le coté technique, architectural et physique du système. On commencera par la sélection technique (sélection d'SGBD, ETL, outils de programmation ...), puis on donnera la modélisation physique du système et on terminera par la présentation de quelques interfaces et fonctionnalités du système.

VI.2. Choix d'outils de développement :

Le choix des outils de développement est l'une des étapes essentielles durant le développement de n'importe quel type de logiciel. Ce choix repose sur des critères comme le cout par rapport au budget dépensé sur le logiciel, l'espace gaspi par l'application et les plates-formes excitantes au niveau de l'entreprise ou le logiciel sera déployé, plus d'autres critères.

Pour notre cas on a choisi le kit Visual Studio et Sql server 2008 de Microsoft:

- + Pour la base de données du DW et les cubes de données on a utilisé SQL Server 2008 et SSAS (Sql Server Analysis Service) de SQL server 2008.
- + Pour l'ETL on a utilisé le SSIS (Sql Server Integration Service) 2008.
- + Pour l'application utilisateur on a utilisé Visual Studio web développer 2008, plus des APIs (le kit **Dundas** pour la BI) pour l'exploitation des données de DW et des cubes à travers des graphes (histogrammes, secteurs, courbe ...), des cartes géographiques, des tableaux et des tableaux croisés pour l'analyse multidimensionnelle.

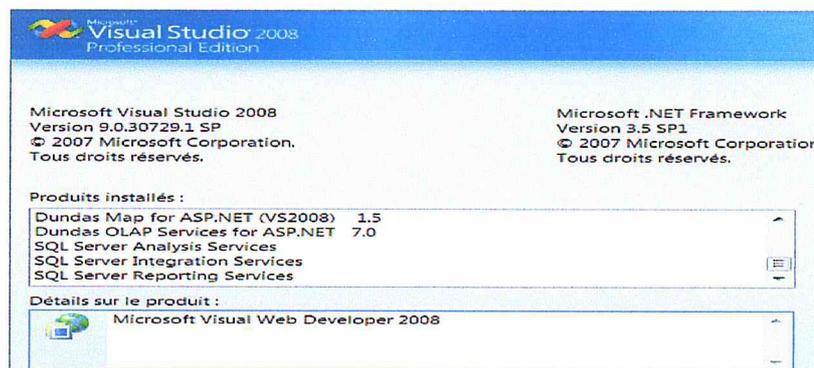


Figure 6.1. Visual Studion2008 avec les Add in Dundas.

VI.2.2. Microsoft Visual Studio 2008 :

Visual studio est un IDE (Integrated Development Environment) conçu par Microsoft. Visual studio, a été pour la première fois publié en 1997. Son but premier était de rassembler plusieurs langages au sein d'un même environnement de développement. Un changement fondamental a été réalisé en 1998 avec l'intégration du Framework .Net. Avec la version 2008, il est possible de programmer avec un très grand nombre de langages tel que : Visual basic, c#, c++, JScript 8.0, base de données (SQL). En plus, il offre la possibilité de programmer avec le Framework .Net 2.0, 3.0, 3.5.

Grâce à un système de plugin il est possible de personnaliser la suite logicielle et de rajouter de nouveaux langages ou de nouvelles fonctionnalités. Par exemple, dans notre cas on a intégré le kit Dundas de plug in pour créer des charts, cartographies et autre interface d'analyse multidimensionnelle. Ces plugins s'intègrent parfaitement dans l'environnement.

Visual Studio 2008 permet aussi de :

- ✚ **Développer des applications pour clients dynamiques :** Visual Studio 2008 offre aux développeurs une nouvelle méthode simple pour créer des clients dynamiques. Les développeurs disposent en effet d'un ensemble d'outils complet et de classes qui simplifient l'intégration de clients dynamiques aux applications Web nouvelles ou existantes. Ils peuvent aussi gérer la mise en mémoire cache locale des données pour des utilisations hors connexion.
- ✚ **Créer des applications Microsoft Office :** Visual Studio Tools for Office (VSTO) est maintenant entièrement intégré à Visual Studio 2008 Professional Edition. Visual Studio permet aux développeurs de personnaliser des applications Office.
- ✚ **Construire des applications Windows Vista :** Les développeurs peuvent facilement exploiter les nouvelles plateformes ; ils peuvent concevoir des applications attrayantes en intégrant sans difficulté les nouvelles fonctionnalités de Windows Presentation Foundation dans les applications Windows Forms existantes et dans les nouvelles applications.

✚ **Gérer les données plus efficacement** : Plus les anciens méthode d'accès aux données, cette version (**ADO et ADO.NET**) introduit le LINQ (Language Integrated Query) et de nouvelles possibilités d'accès aux données permettent aux développeurs de gérer les données de façon plus cohérente, d'accéder aux données via de nouvelles zones de conception et enfin, d'utiliser des classes intégrées dans le cas d'un gabarit de conception non connecté en permanence.

✚ **Web** : En plus de l'infrastructure sécurisée, fiable et extensible d'IIS, les développeurs disposent d'un outil qui leur permet de créer facilement des applications Web interactives. Grâce à l'intégration du modèle familier de programmation AJAX ASP.NET, l'exécution des applications côté client est plus efficace et les utilisateurs bénéficient d'une interface Web plus réactive.

Inconvénients :

- Lent quand beaucoup de projets sont ouverts dans une solution.
- Perte de performance quand beaucoup de tab sont ouverts avec des interfaces.
- La portabilité : c'est l'inconvénient le plus major de Visual Studio et tous le Framework .NET. Visual Studio et toutes les applications développées en l'utilisant ne s'exécutent que sur les plateformes de Microsoft (Windows, Windows server ...).

VI.2.3. Microsoft Sql Server 2008:

Microsoft est l'un des leaders des éditeurs de SGBD avec sa solution SQL Server. Selon des études faites par Garthner, IDC, et Forrester Wave, Sql Server se positionne troisième derrière Oracle et IBM en terme du part de marché (Figure 6.3), première en termes de croissance avec 28 % et aussi première en terme de sécurité selon une étude de ESG en 2006 .(figure 6.4)

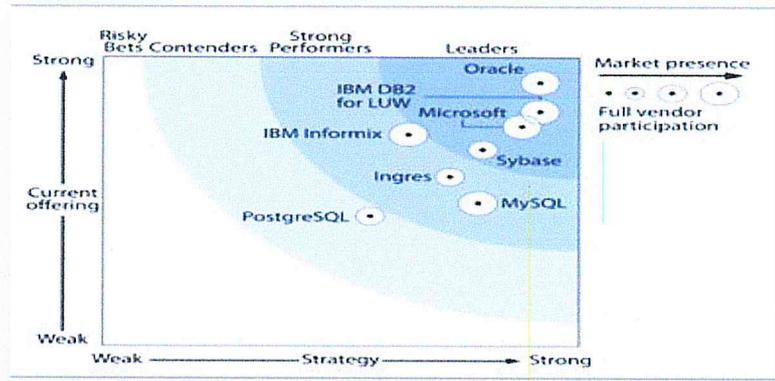


Figure 6.2. Positionnement de quelques SGBD (Forrester Wave 2009)

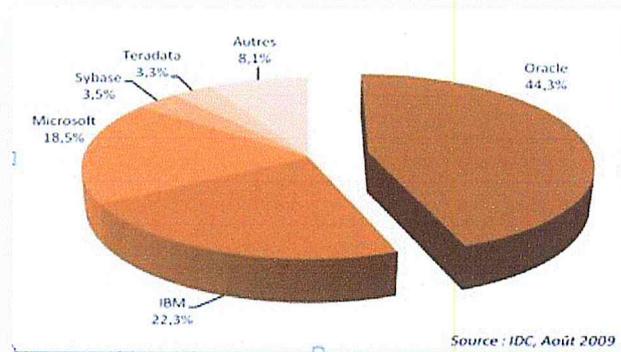


Figure 6.3. Part du marché des éditeurs d'SGBD (IDC 2009)

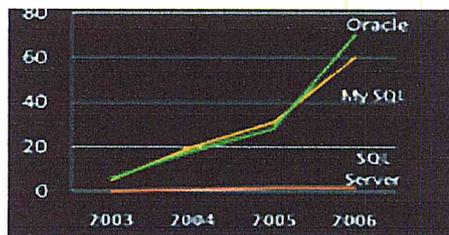


Figure 6.4. Nombre de failles par éditeur (ESG 2007)

Sql server dans sa version 2008 est livré sous forme d'un kit englobe un moteur de base de données, un service d'analyse, un service de reporting et un service de d'intégration (ETL).

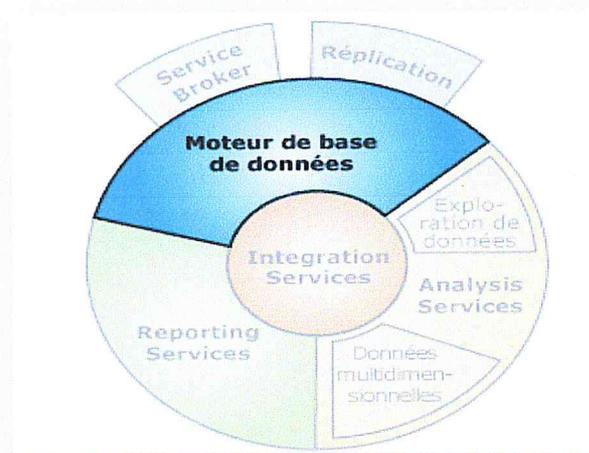


Figure 6.5. Les services de SQL Server 2008

✚ Moteur de base de données :

Le Moteur de base de données est le service central qui permet de stocker, traiter et sécuriser les données. Grâce au moteur de base de données, il est possible de contrôler les accès et de traiter rapidement les transactions pour répondre aux besoins des applications consommatrices de données.

Ce Moteur est utilisé pour créer des bases de données relationnelles, pour le traitement de transactions en ligne ou des données de traitement analytique en ligne (OLAP). Ces opérations comprennent la création de tables pour le stockage des données, ainsi que les objets de base de données tels que les index, les vues et les procédures stockées pour l'affichage, la gestion et la sécurisation des données.

✚ Sql Server Integration Service(SSIS) :

Le SSIS, c'est le service qui permet d'extraire, transformer /intégrer des données de sources hétérogènes et de les charger dans le DW. Dans ce service on appelle « package » l'environnement dans lequel on travaille. On peut construire plusieurs packages Integration Services, et cela sous l'interface « Business Intelligence Development Studio » (Figure 6.6).

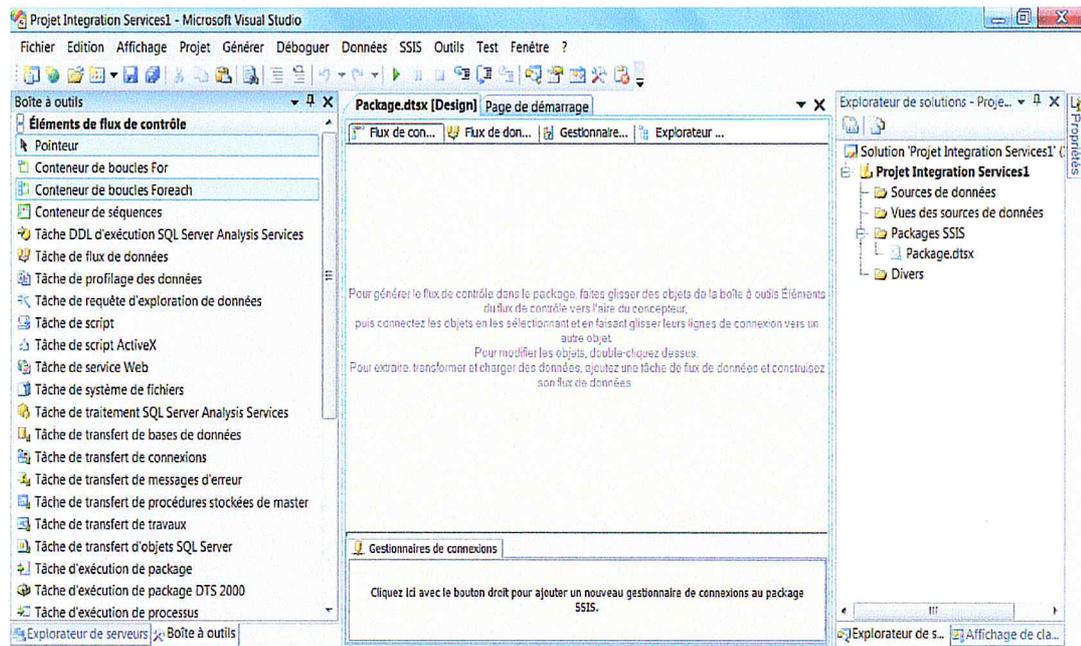


Figure 6.6. L'interface de SSIS

Chaque package contient toutes les tâches d'intégration et fonctionne grâce au principe du « glisser-déposer », ce qui facilite son utilisation. L'enchaînement des tâches d'un package est orchestré par le flux de contrôle (Control Flow). Lorsqu'une tâche a pour objectif d'assurer la transformation des données, elle est nommée « tâche de flux de données ». A l'intérieur de cette tâche se trouve un flux de données (Data Flow) contenant au minimum une source, une transformation et une destination (Figure 6.7).

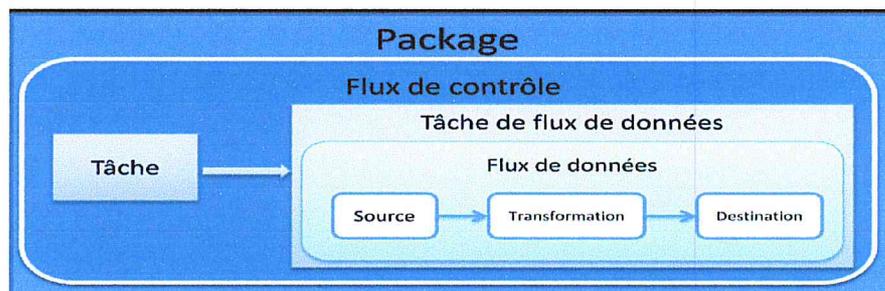


Figure 6.7. L'interface de SSIS

Les tâches SSIS participent chacune à sa manière à l'élaboration de l'entrepôt de données. Comme expliqué précédemment, les tâches de flux de données aident à la transformation des données. Il existe de nombreuses autres tâches, qui concernent souvent soit la maintenance, soit le flux en lui-même. Elles se trouvent toutes dans la boîte à outil du Data Flow. En voici quelques unes :

- **Tâche de service web** : elle permet de lancer une méthode de service web, celui-ci permettant la communication et l'échange de données entre systèmes ou applications, par le protocole http ;
- **Tâches de transfert de bases de données, de connexions, de messages d'erreur, de procédures stockées, et d'objets SQL Server** ;
- **Tâche de script** : elle permet de créer des fonctions non disponibles directement dans les autres tâches de SSIS. (En langage VB par exemple) ;
- **Tâches d'exécution de package, de processus, et de requêtes SQL** ;
- **Tâche XML** : elle permet de travailler avec des données XML ;
- **Tâche FTP** : elle permet à un package de télécharger des fichiers de données depuis un serveur distant ;
- **Tâche « Envoyer un message »** : elle permet l'envoi de messages notifiant le succès ou l'échec d'un événement.

On a d'autres tâches concernant la maintenance :

- **Tâches de nettoyage d'historique ou de maintenance** ;
- **Tâche « Mettre à jour les statistiques » d'une ou plusieurs bases de données** ;
- **Tâche « Sauvegarder la base de données »** ;
- **Tâche « Vérifier l'intégrité de la base de données »**.

Il est possible de regrouper certaines tâches entre elles dans ce qu'on appelle un conteneur. Chaque conteneur contient donc un ensemble de tâches liées entre elles par le même objectif. Il existe ainsi des conteneurs répétitifs qui permettent de créer une boucle qui exécutera leurs tâches un certain nombre de fois.

✚ **Sql Server Analysis Service(SSAS) :**

SQL Server Analysis Services est la plateforme qui permet de créer et gérer des structures multidimensionnelles et des modèles d'exploration de données. Pour cela, Analysis Services fournit des fonctions OLAP (On Line Analytical Processing), et des applications d'exploration de données (data mining). La création des cubes OLAP et des modèles d'exploration se fait avec SQL Server Business Intelligence Development Studio, leur gestion avec SQL Server Management Studio. Ces analyses comprennent un traitement sur des bases de données volumineuses et

permettent de comprendre les métriques et les éléments qui influent sur le fonctionnement de l'entreprise.

Le langage qui permet d'interroger les cubes OLAP est le langage MDX (Multidimensional Expressions). C'est aussi ce langage qui est utilisé par Analysis Services pour construire les cubes. La notion de procédures stockées est également prise en compte dans Analysis Services, elles permettent d'étoffer les fonctions basiques du langage MDX. Il est possible d'en écrire sous différents langages (VB, C++, C#...).

Analysis Services permet une analyse rapide, intuitive et verticale de grandes quantités de données construites sur ce modèle de données unifié, disponible aux utilisateurs dans plusieurs langages et devises. Il fonctionne avec les entrepôts de données, les mini-Data Warehouse, les bases de données de production et les magasins des données opérationnelles, en prenant en charge à la fois l'analyse des données d'historique et en temps réel.

VI.3. Modélisation physique du système :

VI.3.1. Schéma physique des sources :

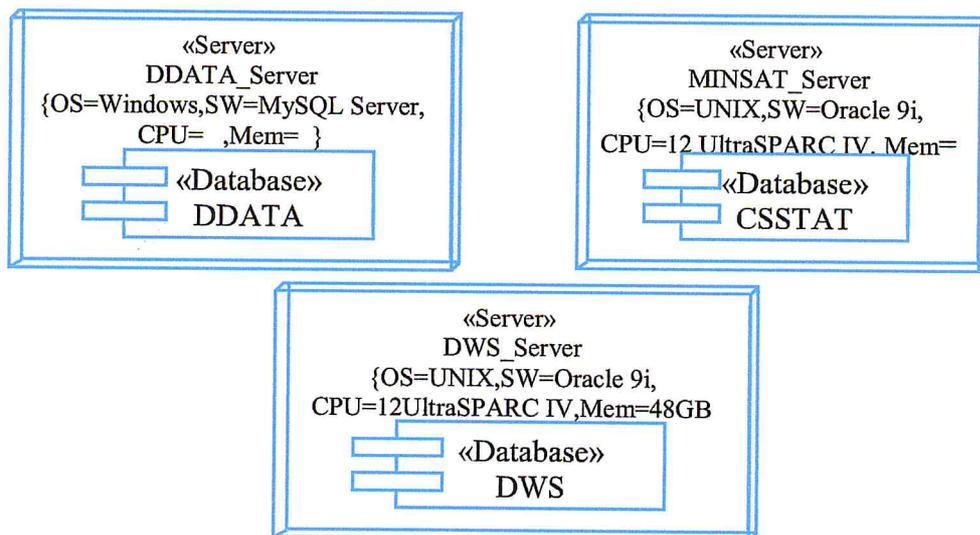


Figure 6.8. Diagramme de déploiement des sources

VI.3.2. Modèle physique de DW :

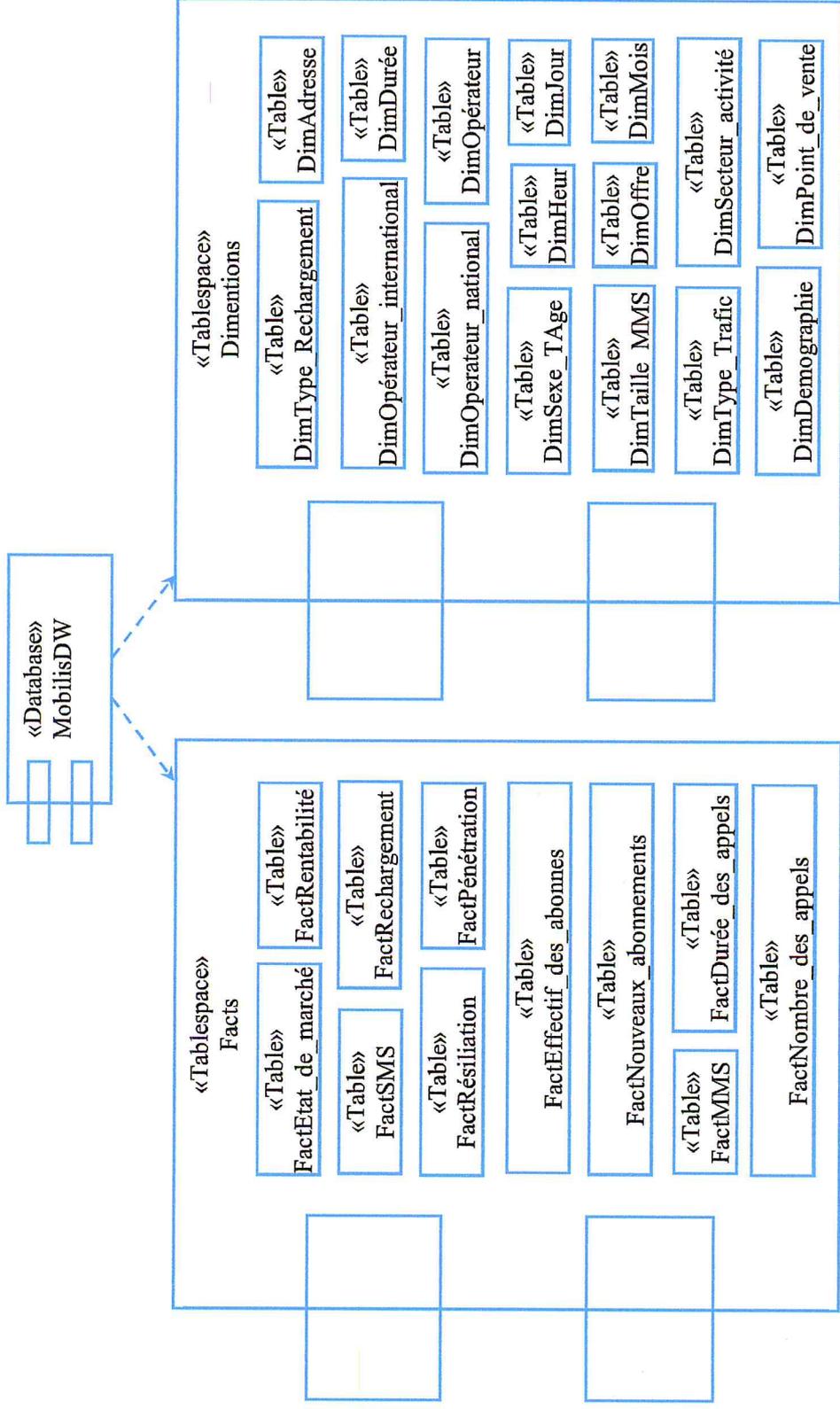


Figure 6.9. Diagramme de composants du DW.

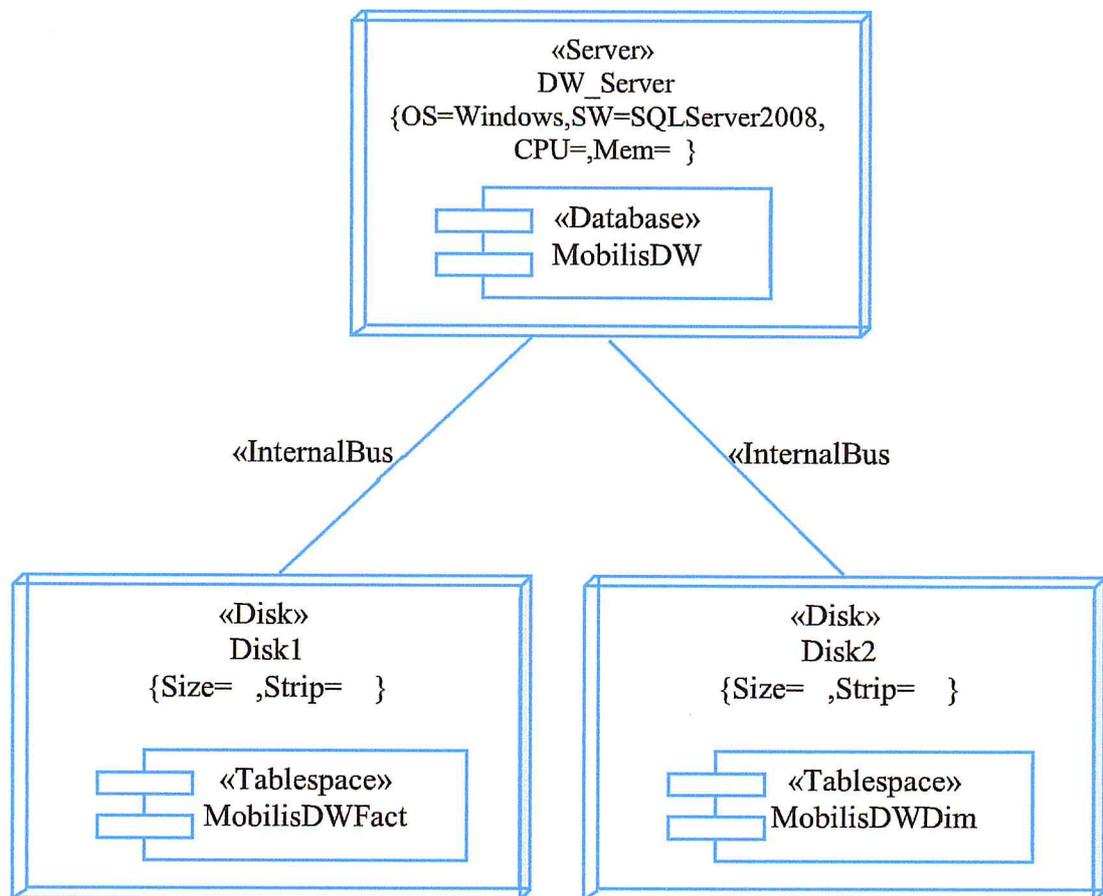


Figure 6.10. Diagramme de déploiement du DW.

VI.3.3. Modèle physique de l'ETL (Transportation Schema) :

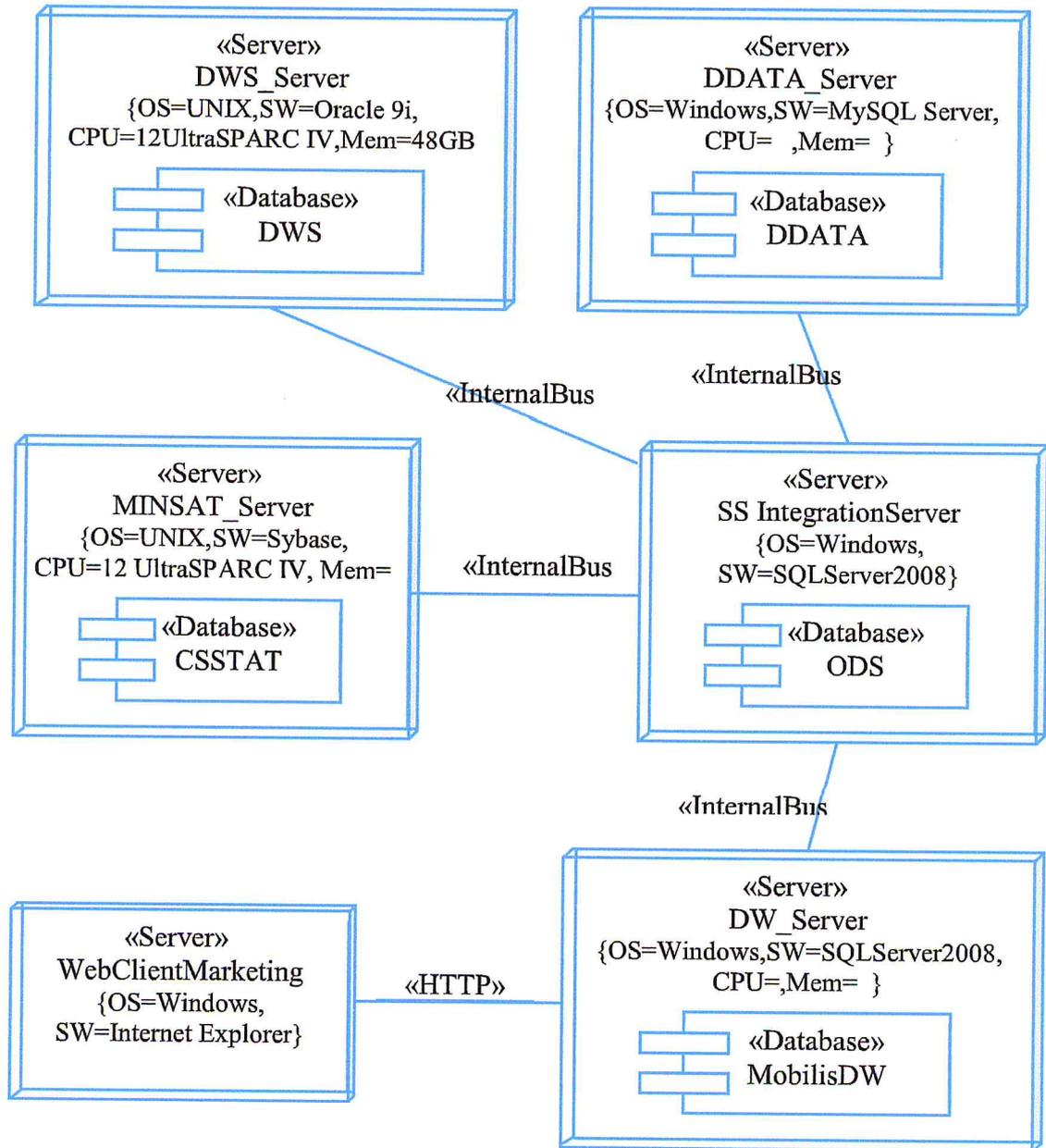


Figure 6.11. Diagramme de déploiement de l'ETL.

VI.4. Implémentation d'ETL :

Utilisant SSIS, nous avons implémenté l'ETL de notre système décisionnel. On a décomposé le processus d'ETL en deux package : abonné et trafic. Un package contient des tâches reliées entre elles et définissent le flux de contrôle, et chaque tâche décrit un flux de données pour le chargement d'une dimension, fait ou une table intermédiaire. Les tâches de package 'abonné' seront exécutées mensuellement, tandis que les tâches de package 'trafic' sont quotidiens.

Grâce à la méthode qu'on a suivie, l'implémentation d'ETL devient facile. On utilise le schéma DMD pour définir le flux de contrôle où les tâches sont les "mapping packages" du schéma DMD (voire Figures 6.12 et Figure 6.13); et les sources, transformations et destinations sont celles de schéma ETL Processe.

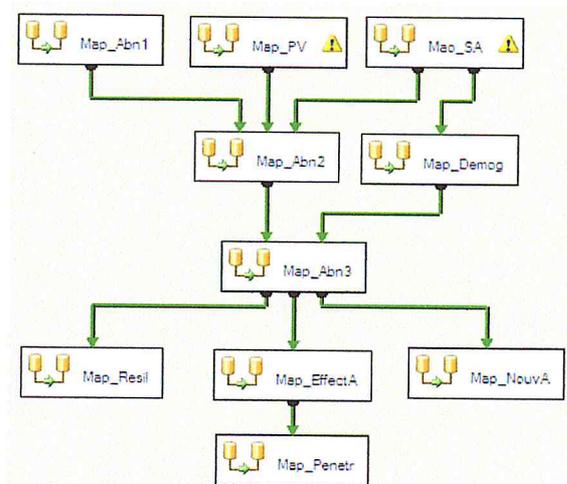


Figure 6.12. Flux de contrôle « abonné ».

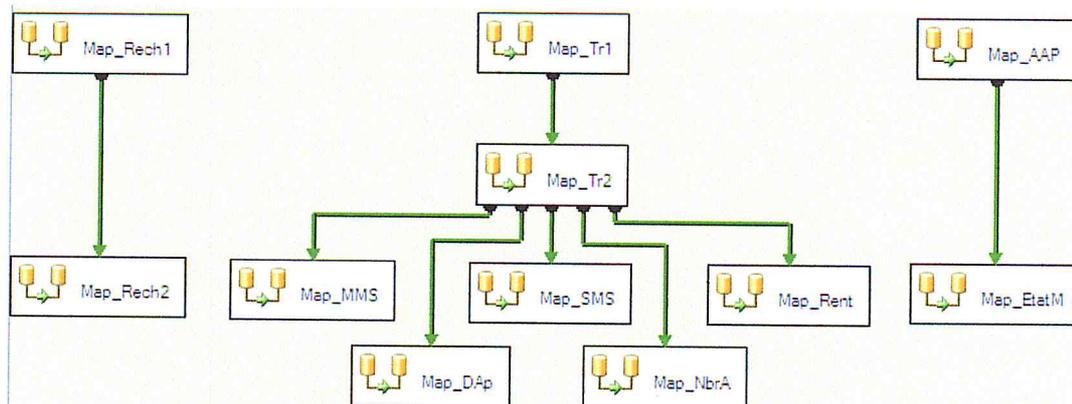


Figure 6.13. Flux de contrôle « trafic ».

VI.5. Quelques interfaces de l'application de restitution :

L'application utilisateur est une application web développée avec Microsoft Visual Web developer, elle est basée sur des pages ASP.NET (ASP.NET est l'environnement de programmation .NET pour la construction d'applications en HTML qui s'exécutent sur le Web). Les graphes, cartes géographique, tableaux croisés ... sont programmés à l'aide d'un ensemble d'API (Dundas Chart, Gauge, Map, Chart for Olap service) développés par la société **Dundas Data Visualisation** (www.dundas.com). Les postes des utilisateurs finaux nécessitent seulement un navigateur Web pour y accéder à l'application. La première interface (page web) de restitution après l'authentification comporte un ensemble de (Gauges et graphes) exprimant des indicateurs comme l'effectif d'abonnés, nombre de du marché, nombre de résiliations de mois en cours.

Cette interface est divisée en 2 sujets d'analyse (répartition des abonnés et Consommation « Etude comportementale ») et permettra aux décideurs (utilisateurs finaux) de prendre une vue de haut niveau sur ces 2 derniers (figure 6.14).

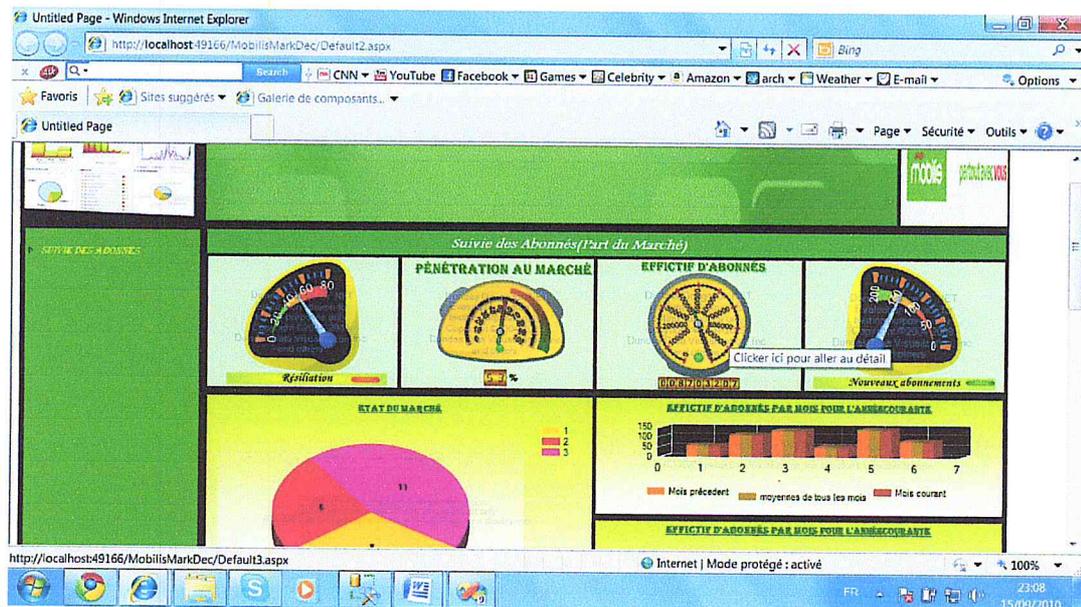


Figure 6.14. Interface donne les indicateurs sous forme de tableau de bord.

Éventuellement l'utilisateur veut aller au détail d'un indicateur par exemple, il veut analyser l'Effectif des abonnés par rapport à la dimension géographique ou démographique. Pour ce raison on a doté l'application par des interfaces qui permettant la géo- analyse (figure 6.15). Ces cartographies peuvent servir comme des cri-

ters d'un Drill-Down sur une Wilaya ou une région, la même chose pour les graphes.

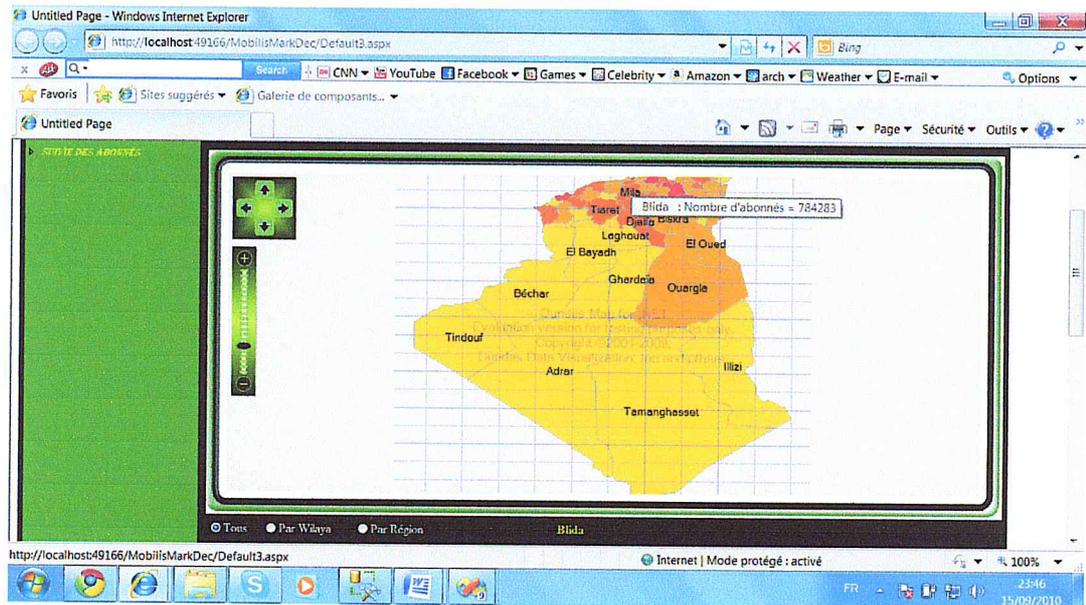


Figure 6.15. Interface permet l'analyse géographique d'effectif des abonnés.

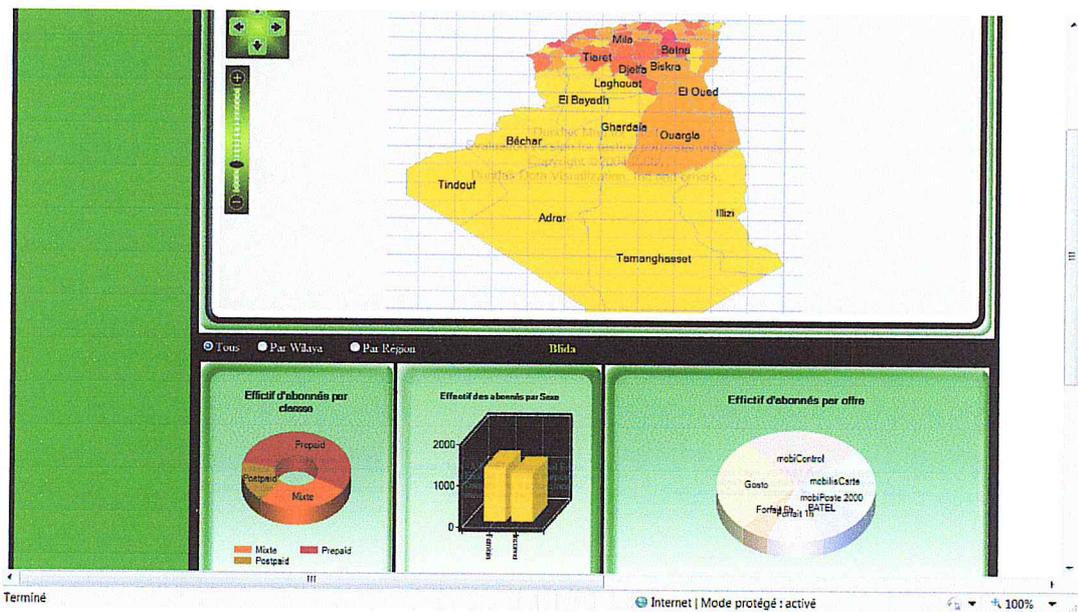


Figure 6.16. Interface permettant l'analyse géographique d'effectif des abonnés par Classe/offre, Sexe avec la possibilité de Drill-Down

Dans plusieurs cas, tous ces genres de restitutions et d'analyses ne suffisent pas, l'utilisateur cherche des fois à fouiller dans les données afin de dégager des réponses à des problèmes ou de trouver des liens de causalités sur des phénomènes. Pour cela on a lui offrir des interfaces d'analyse multidimensionnelle. Ces interfaces

CONCLUSION

ET

PERSPECTIFS

CONCLUSION ET PERSPECTIFS :

Conclusion :

Ce mémoire a abordé la problématique de la conception et la réalisation d'un système décisionnel pour le marketing au niveau de l'entreprise ATM Mobilis afin de répondre à la problématique d'origine qui est la difficulté rencontrée par les dirigeants de l'ATM Mobilis et spécifiquement ceux du marketing pour évaluer et suivre les actions marketing et maîtriser le marché de téléphonie mobile.

Ce mémoire a été organisé en 2 parties, une partie théorique d'écrivant les notions théoriques de l'informatique décisionnelle (définitions, architecture, méthodes de développement ...) et une partie pratique traitant et résolvant la problématique.

Pour la conception du système on a utilisé une méthode introduit par Luján Mora spécifiquement pour les projets des systèmes décisionnels, on a ajouté quelque diagrammes additionnels : diagramme de cas d'utilisation pour montrer les besoins et diagramme d'activité pour visualiser de manière simple le déroulement d'outil d'alimentation de notre nouveau système.

L'implémentation du système est faite sous une plateforme Windows basé sur Microsoft SQL Server comme serveur de base de données, le service SSIS pour l'implémentation d'outil d'alimentation du système et Microsoft Visual Web developer pour l'application. Notre système importer ses données depuis des sources de données multiples : fichiers plat, base de données Oracle, Sybase et MySQL.

Perspectives :

Ce projet ouvre la voie vers différentes perspectives. On peut les catégorisées en 3 types : Des perspectives sur le projet eu même, des perspectives sur les recherches faites dans le domaine de l'informatique décisionnelle et des perspectives dans le domaine de marché BI.

Concernant notre projet, il faut noter que le marketing de nos jours est orienté vers les clients. Où la segmentation des clients jeu un rôle important dans ce contexte. Pour cela le système qu'on a développé peut être enrichir par un outil de segmentation des clients et encore un système de prédiction des churns. En plus, Ce système peut être le premier pas de l'entreprise ATM Mobilis vers la mise en œuvre d'un système décisionnel globale pour tous les autres activités comme (le commerciale, finance, réseaux ...)

Ce projet a nous permet de découvrir un domaine très actif en termes de sujet de recherche : Recherches sur la méthodologie de développement, sur les ETL, sur les DW XML, sur l'évolutivité des DW, sur l'automatisation d'administration et tuning de DW.

Dans ce contexte et pendant le développement de notre système, on a remarqué qu'on a fait plusieurs choix, choix de la technologie de stockage (MOLAP, ROLAP, HOLAP), choix du modèle logique (en étoile ou en flocon) et au niveau physique, choix de type d'indexe et quelles sont les attributs à indexer et encore, le choix de type de partitionnement (horizontal, vertical ...) et quelles sont les tables à partitionner. Tous ça, à nous stimuler à poser cette question (pour quoi pas l'automatisation ou la semi automatisation de processus de développement d'un SID ?). Nous pensons que cette question peut ouvrir les ports sur une nouvelle branche de recherche dans le domaine de l'informatique décisionnel.

Et enfin, ce projet a permet nous aussi de découvrir le marché des solutions BI quelque soit les solutions payante ou open source, pour ces derniers on eu l'occasion de se rapprocher vers les fonctionnalités offertes par les solutions open source comme (Jasper, Pentaho, Spago), on a remarqué que les solutions open

Conception et Réalisation d'un Système décisionnel pour le marketing

sources utilisent presque les mêmes API par exemple les 3 solutions cités utilisent BIRT comme outil de reporting et Talend comme outil ETL avec quelques modifications. Pour ces raisons on a dit (pour quoi pas le développement de notre solution BI ?)

Glossaire

API *Application Program Interface*

Un ensemble de routines, protocoles et outils pour la construction des logiciels. Une bonne API permet de facilement développer un programme.

BI *Business Intelligence*

Un ensemble d'outils dédiés aux responsables afin de les aider à prendre des décisions

CRM *Customer Relationship Management*

Regroupe tous les aspects de l'interaction entre une entreprise et ces clients quelque soit les vente, service ou une nouvelle technologie aide à gérer informatiquement les relations avec les clients

DW *Data Warehouse*

Une collection de données intégrées, historiées, non volatiles, permettant la précise de décisions

DWEP *DW Engineering Process*

C'est un processus de développement des systèmes décisionnels proposé par Luján Mora, elle se base sur 2 standards, UML come langage de modélisation et UP comme méthode de conduit de projet.

E-A *Entité Association*

C'est modèle de données conceptuel qui permet de voir le monde réel comme des entités (avec des attributs) et des relations entre ces entités.

ETL *Extract, Transform, Load*

Un outil permettant d'extraire des donnés d'une ou plusieurs sources de données, de les transformer et les charger dans un cible (généralement un DW)

HOLAP *Hybrid OLAP*

Un des type d'OLAP, HOLAP combine entre la stratégie ROLAP et MOLAP afin d'en profiter des avantages de chacun d'elles.

HTML *HyperText Markup Language*

C'est le langage utilisé pour créer des documents en www. HTML définit la structure en couche d'un document web par l'utilisation d'un ensemble de balises (tag) et attributs, W3C qui a standardisé HTML.

OLAP *On line Analytical Processing*

Une catégorie de logiciels qui permet l'analyser multidimensionnelle d'une très grande quantité de données stocker dans une base de données ou sous forme multidimensionnelle, il ya 3 type principaux d'OLAP (ROLAP, MOLAP, HOLAP).

OMG *Object Modeling Group*

Une organisation interactionnelle de standardisation dans le domaine de l'orienté Objet, elle responsable sur plusieurs spécification comme CORBA, MOF, UML ...

MOLAP *Multidimensional OLAP*

Un des type d'OLAP, dans ce type d'OLAP le stockage est purement multidimensionnel (Cube de données).

SGBD *système de gestion de Base de Données*

SID *Système d'Information Décisionnel*

Un système d'information spécifique, son but est de collecter les informations éparpiller sur les systèmes de production et de les intégrer dans une format appropriée pour l'analyse et la prise des décisions à travers des outils de restitutions.

ROLAP *Relational OLAP*

Un des type d'OLAP produire une adaptation des bases de données relationnelle pour le support d'OLAP. Dans ce type d'OLAP les données sont stockées dans un SGBD relationnel.

UML *Unified Modeling Language*

C'est un langage de modélisation graphique pour visualiser, spécifier, construire et documenter les artefacts. La modélisation se fait à travers des diagrammes permettant de définir un logiciel selon divers points de vue.

UP *Unified Process*

C'est un processus unifié de développement des logiciels, elle est introduite par les fondateurs de l'UML ou UML définit un langage de modélisation visuel et UP spécifie comment développer un logiciel en utilisant UML

XML *Extensible Markup Language*

Méta-langage développé par le W3C permettant de définir des langages de balisage de documents ou de messages, au centre d'un ensemble de standards dédiés à la communication dans les systèmes d'information

Bibliographie

1. **GAM, Ines.** *Ingénierie des Exigences pour les Systèmes décisionnel, Concepts, Modèles et Processus, La méthode CADWE.* THESE DE DOCTORAT, DE L'UNIVERSITE PARIS I – PANTHEON - SORBONNE : s.n., 2008.
2. **NIEUWBOURG, Philippe.** *Quels système décisionnel pour les entreprises agiles* . s.l. : livre blanc, Microsoft, 2002.
3. **Goglin, Jean-François.** *La construction du Data Warehouse.* s.l. : HERMES, 2001.
4. **Immon, Bill.** *Building the Data Warehouse.* s.l. : John Wiley & Sons, 2002.
5. **Kimball, Ralph et Ross, Margy.** *entrepôts de données: Guide pratique de modélisation dimensionnelle.* Paris : Vuibert, 2003.
6. **Luján-Mora, Sergio.** *Data Warehouse Design with UML.* s.l. : These doctorat, Université d'Alicante, 2005.
7. **Kimball, Ralph, et al., et al.** *Concevoir et déployer un data warehouse: Guide de conduite de projet.* Paris : Eyrolles, 2000.
8. **Favre, Cécile.** *Évolution de schémas dans les entrepôts de données :mise à jour de hiérarchies de dimension pour la personnalisation des analyses.* s.l. : these de doctort, Université Lumière Lyon 2, 2007.
9. **Sergio Luján-Mora, Panos Vassiliadis, et Juan Trujillo.** *Data Mapping Diagrams for Data Warehouse.* s.l. : Article Scientifique, 2004.
10. **Benjamine, EPEE NGANDO.** *systeme d'information décisionnel pour le suite de l'activité Médico-Administrative:Modélisation et réalisation.* s.l. : Mémoire d'ingénieure, Conservatoire National des arts et des matiers, Paris, France, 2007.
11. **DUBÉ, ETIENNE.** *Conception et développement d'un service web de constitution de mini cube SOLAP pour client web .* s.l. : mémoire pour l'obtention du grade de Maître ès sciences, Université Laval France, 2008.
12. **L'OLAP, histoire, caractéristiques et état du marché : de Codd à Hyde !!**
<http://www.altic.org>. [En ligne] Avril 2009. [Citation : 27 juin 2010.]
<http://altic.org/altic-blog-buzz/recherche-et-developpement-rad-altic/160-olap1#historique>.
13. **Grim, Yazid.** *Passez en mode BI.* s.l. : www.developer.com, 2007.

14. **NIEUWBOURG, Philippe.** *Quels système décisionnel pour les entreprises agiles*. s.l. : livre blanc, Microsoft, 2002.
15. **UMANIS, SSSL du groupe.** *RESTITUER, ANALYSER ET PILOTER : EVALUER LES OPPORTUNITES OPEN SOURCE*. s.l. : Livre Blanc, OPENBIZZ.
16. **R, Hodgson.** *The X-model: a process model for object-oriented software development*. s.l. : Acte du congrès Le Génie Logiciel et ses applications, Toulouse, 91.
17. **E., Millot P. and Roussillon.** *Man-Machine cooperation in Telerobotics : Problematics and Methodologies*. s.l. : Second Symposium on Robotics, Gif-sur-Yvette, 1991.
18. **W.W., J Royce.** *Managing the Development of Large Software Systems. Proceedings of*. 1970.
19. **B.W., Boehm.** *Software Engineering Economics*. s.l. : Englewood Cliffs : Prentice Hall, 1981.
20. **K, McDermid J. and Ripkin.** *Life cycle support in the ADA environment*. s.l. : Cambridge University Press, 1984.
21. **C, Toffolon.** *Cours de genie logiciel ou comment amener de petits génies à écrire des*. s.l. : Cours de DESS-ICC, Université du Littoral, 1999.
22. **Jacobson I., Booch G. and Rumbaugh J.** *Le processus unifié de développement*. s.l. : Paris : Eyrolles, 2000.
23. **J Giorgini P., Perini A. Mylopoulos J., Giunchiglia F. and Bresciani P.** *Oriented Software Development : A case study*. s.l. : In proceeding of the 13th International, Buenos-Aires,, juin 2001.
24. **Abdouroiahmane, ANLI.** *Méthodologie de développement des systèmes d'information personnalisés*. s.l. : thèse de doctorat, Université de Valenciennes, France, 2007.
25. **Estella Annoni, Franck Ravat , Olivier Teste , Gilles Zurfluh.** *Méthode de Développement des Systèmes décisionnels: Roue de Deming*. s.l. : Rapport scientifique, Laboratoire IRIT(SIG-ED), Université Paul Sabatier, May 2010.
26. **Jarke, M., Pohl, K.** *Requirements Engineering in 2001: Managing a Changing Reality*. s.l. : IEEE Software Engineering Journal, 1994.
27. **Rolland, C., Loucopoulos, P., Grosz, G., and Nurcan, S.** *A framework for generic patterns dedicated to the management of change in the electricity supply*

- industry*. s.l. : 9th International DEXA Conference and Workshop on Database and Expert Systems application, 1998.
28. **Prieto-Diaz R., Freeman, P.** *Classifying software for reusability*. s.l. : IEEE Software, 1987.
29. **Annoni, E., Ravat, F., Teste, O., and Zurfluh, G.** *Automating the choice of decision support*. s.l. : In Database and Expert Systems Applications Conference, 2006.
30. **Mazón, J.-N., Trujillo, J., Serrano, M., and Piattini, M.** *Designing data warehouses : from business requirement analysis to multidimensional modelling*. s.l. : In 13th IEEE International Requirements Engineering Conference Workshop on Requirements Engineering for Business Needs and IT Alignment (REBNITA), 2005.
31. **Prakash, N., and Gosain, A.** *Requirements driven data warehouse development*. s.l. : In Eder and Welzer, 2003.
32. **Prat, N., etand Akoka, J.** *From uml to rolap multidimensional databases using a pivot*. 2002.
33. **Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., and Paraboschi, S.** *Designing data marts for data warehouses*. s.l. : ACM Trans. Softw. Eng. Methodol, 2001.
34. **Ravat, F., Teste, O., Tournier, R., and Zurfluh, G.** *Integrating complex data into a data warehouse*. s.l. : In SEKE International Conference on Software Engineering and Knowledge, 2007.
35. **Tryfona, N., Busborg, F., et Christiansen, J. G. B.** *StarER : A conceptual model for data warehouse design*. s.l. : In DBL, 1999.
36. *Le nouveau look de la Modélisation des Data Warehouses*. **Dimitri Volkmann, Directeur Marketing Produit Outils, Sybase e-Business**. NUMÉRO 2, s.l. : PowerAMC, Blueprint, Bulletin d'information, 2001.
37. **P, Marcel.** *Manipulation de Données Multidimensionnelles et Langages de Règles*. s.l. : Thèse Doctorat de l'Institut des Sciences Appliquées de Lyon, Fance, 1998.
38. **Zoubir, Ouaret.** *Modélisation conceptuelle et logique d'un entrepôt de données XML*. s.l. : thèse de Magister de l'ECOLE NATIONALE SUPERIEURE D'INFORMATIQUE (ESI), Algérie, 2008.
39. **Jean-Michel, Franco et Sandrine, de Lignerolles.** *Piloter l'entreprise grâce au data warehouse*. s.l. : Eyrolles, 2001.

40. **Ponniah, Paulraj.** *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals.* s.l. : John Wiley & Sons, 2001.
41. **Gouarné, Jean-Marie.** *Le Projet Décisionnel Enjeux, Modèles, Architectures du Data Warehouse.* s.l. : Eyrolles, 1997.
42. **Rainardi, Vincent.** *Building a Data Warehouse: With Examples in SQL Server.* s.l. : Apress, 2008.
43. **Imhoff, Claudia, Glemmo, Nicholas et G. Geiger, Jonathan.** *Mastering Data Warehouse Design.* s.l. : Wiley, 2003.
44. **Soutou, Christian.** *UML 2 pour les bases de données.* s.l. : EYROLLES, 2007.
45. **Benjamin Epée Ngando, Matthieu Lefeuvre.** *Apport d'un outil ETL dans la construction d'une solution orientée décisionnelle: Cas du pilotage médico-économique dans un hôpital.* s.l. : Conservatoire National des Arts et Métiers, Centre d'enseignements de Paris, France, 2005.