

b



Gendarmerie Nationale

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب - البليدة 1
Université SAAD DAHLAB - BLIDA 1

كلية العلوم
Faculté des Sciences

قسم الإعلام الآلي
Département d'Informatique



Mémoire de Master

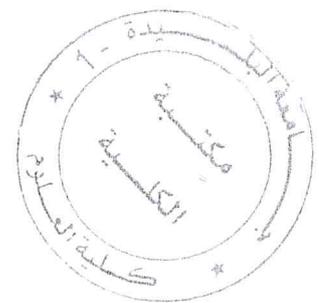
En Informatique

Option

Ingénierie des Logiciels

présenté par

Lamine SLIMANI



Thème :

Conception et Réalisation d'un Data Warehouse
des Accidents de la Route et des Infractions Liées
au Code de la Route Servant de Base pour un
Tableau de Bord Décisionnel

Soutenu le :

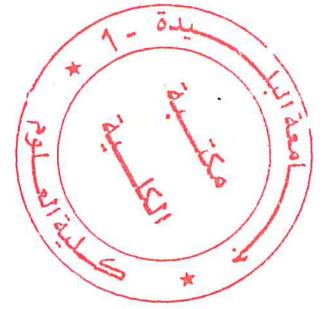
Proposé par : Mr. Mahfoud, BALA.

Jury composé de :

- M(me) Président(e) : M^{me} Abed Hafida
- M(me) Examineur (rice) : M^{me} Rezoug Nachida
- M(me) Examineur (rice) :

MA-004-393-1

Année Universitaire
2016-2017



Remerciements

Je tiens à avancer mes sincères remerciements et gratitudee à Monsieur Bala, maître de conférences à l'USDB, pour avoir accepté d'encadrer ce mémoire et pour son aide précieuse et ininterrompue pour l'accomplissement de mon travail.

Je tiens aussi à remercier du fond du cœur l'ensemble des membres du jury qui me font le grand honneur d'avoir accepté de juger mon travail.

Je remercie particulièrement les enseignants de l'université USDB pour l'effort considérable qu'ils fournissent pour le con déroulement des études.

Je tiens aussi à remercier l'institution de la Gendarmerie Nationale de m'avoir permis de suivre mes études.

*Enfin, je remercie tout particulièrement ma mère, ma femme mes enfants : **Adem Ibrahim Elkhallil et Sarah Nour Errahmène**, qui m'ont toujours soutenu et qui m'ont permis de mener à bien mes études. Je tiens à remercier également ma sœur qui était toujours présente à mes côtés. C'est pour eux que je dédie ce mémoire.*

Lamine

ملخص: إن حوادث المرور، على غرار الحروب، تحصد أرواح الكثير على الطرق. بصفته مسئولاً عن سلامة المواطن و الممتلكات، يجب على الدرك الوطني احتواء هذه الظاهرة. عليه استغلال الكم الكبير من المعطيات، التي تم جمعها من طرف الوحدات الإقليمية، لغرض وضع إستراتيجيات للتدخل و التوزيع الأنجع للقوات عبر شبكة الطرقات. إن مستودعات المعطيات تعتبر التكنولوجيا الأفضل لتحقيق هذه المهمة، حيث أنها توفر القاعدة الرئيسية "لنظم المساعدة لاتخاذ القرارات" و ذلك لتوجيه قرارات المسؤولين على اتخاذ القرارات لتحقيق الأهداف على المستوى القريب، المتوسط و البعيد المدى.

كلمات المفاتيح: حوادث المور، مستودع المعطيات، نظام المساعدة على اتخاذ القرار.

Résumé : Les accidents de la routes, au même titre que les guères, provoquent des génocides sur les routes algérienne. La Gendarmerie Nationale, responsable de la sécurité des biens et citoyens, ce doit de maîtriser ce fléau. Elle doit exploiter les données colossales, recueillies depuis par ces unités territoriales, afin de définir des stratégies d'intervention et de distribution des éléments de l'arme sur le réseau routier. Les data warehouses sont les outils les plus adaptés à cette mission, ils assurent le support pour un système décisionnel aidant les décideurs pour l'accomplissement de leurs courts, moyens et long terme buts.

Mots clés : Accident de la route ; Data warehouse ; Système décisionnel.

Abstract : Road Accidents, as well as wars, are responsible for genocides on algerian roads. The "Gendarmerie Nationale", responsible for citizens and the well-faire security, has to control this scourge. The huge amount of data, gathered by the units on the field, has to be employed in order to put some effecient strategies for intervention and personnel distribution on the road network. The data warehouses are the most suitable tools for this mission accomplishment and fulfilment. They are the backbone for the decisional systemes helping the decision-makers in their short, mid and long-term goals.

Keywords : Road Accident, Data warehouse, Decisional système

Table des matières

Liste des figures :	i
Liste des Tableaux :	vi
Lexique	viii
1. Introduction Générale:	1
1.1. Contexte :	1
1.2. Problématique :	1
1.3. Objectifs du projet :	3
1.4. Présentation du mémoire :	3
2. Les systèmes décisionnels « Decision-making system » :	4
2.1. Introduction :	4
2.2. Les systèmes d'information décisionnels	4
2.2.1. Qu'est-ce que le décisionnel ?	4
2.2.2. Situation du décisionnel au sein du système d'information :	5
2.2.3. Les composantes du décisionnel :	6
2.3. Le Data Warehouse :	7
2.3.1. Définition:	7
2.3.2. Historique des Data Warehouse :	8
2.3.3. Structure des données d'un Data Warehouse :	8
2.3.4. Composantes d'un Data Warehouse :	9
2.3.5. L'architecture globale du data warehouse.....	11
2.3.6. Définitions des concepts [Source : Goglin1998]	12
2.3.7. Le concept OLAP :	14
2.3.8. Fonctionnement du système data warehouse:.....	16
2.3.8.1. L'acquisition des données :	16
2.3.8.2. Le stockage des données dans le data warehouse :	17
2.3.8.3. L'alimentation des datamarts (Les magasins de données) :	17
2.3.8.4. L'exploitation de l'information :	18
2.4. Modélisation des données de l'entrepôt :	18
2.4.1. Caractéristiques d'utilisation: OLTP vs OLAP:.....	18
2.4.2. Les techniques de modélisations et ces concepts :	20
2.4.2.1. Concepts liés à la modélisation dimensionnelle:	21
2.4.2.2. Les techniques de modélisations:.....	22
2.5. Démarche de Construction d'un Data Warehouse :	26
2.5.1. Modélisation et conception du Data Warehouse :	27
2.5.2. Alimentation du Data Warehouse :	28
2.5.2.1. Les phases de l'alimentation « E.T.L. » :	29

2.5.2.2.	Politiques de l'alimentation :.....	30
2.5.3.	Gestion du temps dans le data warehouse :.....	31
2.5.3.1.	Le rôle des données temporelles :.....	32
2.5.3.2.	Les problématiques liées au temps :.....	32
2.6.	Conclusion :.....	33
3.	Etude des besoins :.....	34
3.1.	Introduction :.....	34
3.2.	Description de la démarche d'étude des besoins (Interview) :.....	34
3.2.1.	Localisation et détermination des postes et personnes détenant l'information de l'étendu du projet : ...	35
3.2.2.	Interviews des DBA et développeurs :.....	36
3.2.3.	Planification, préparation et conduite des interviews :.....	36
3.2.3.1.	La planification :.....	36
3.2.3.2.	La préparation :.....	36
3.2.3.3.	La conduite promptement dite :.....	36
3.3.	Autres moyens utilisés pour la détection des besoins :.....	36
3.4.	Étude préliminaire des systèmes sources et des applications.....	37
3.4.1.	Les applications de la Division Sécurité Routière :.....	37
3.4.2.	Le Records Management System (RMS) :.....	37
3.4.2.1.	Module Accident de la route :.....	38
3.4.2.2.	Module contravention :.....	38
3.4.3.	Les états de sorties périodiques de la DSR :.....	39
3.5.	Rédaction et validation du recueil récapitulatif des besoins.....	39
3.6.	Problèmes et obstacles rencontrés :.....	41
3.7.	Conclusion.....	42
4.	Conception du Data Warehouse :.....	43
4.1.	Conception de la zone « entreposage des données » :.....	43
4.1.1.	Introduction :.....	43
4.1.2.	Processus de la modélisation dimensionnelle :.....	43
4.1.2.1.	Volet « Accident de la route » :.....	43
4.1.2.2.	Volet « Chauffeurs » :.....	54
4.1.2.3.	Volet « Infraction routière ».....	62
4.1.2.4.	Volet « Victime ».....	66
4.1.2.5.	Volet « Véhicule ».....	70
4.1.3.	Conclusion :.....	74
4.2.	Conception de la zone « Alimentation » :.....	74
4.2.1.	Introduction :.....	74
4.2.2.	Etude et planification.....	75
4.2.2.1.	Les sources de données :.....	75

4.2.2.2.	Détection des emplacements des données sources :.....	76
4.2.2.3.	Définition de la périodicité de chargement.....	76
4.2.3.	Architecture du processus d'alimentation	77
4.2.4.	Processus de chargement.....	77
4.2.4.1.	Processus de chargement de dimension :.....	78
4.2.4.2.	Processus de chargement des tables de fait :.....	79
4.2.4.3.	Processus de chargement particulier de la dimension « temps » :.....	80
4.2.4.4.	Préparation du data warehouse pour les données cadastrales (SIG) :.....	80
4.2.5.	Conclusion :.....	80
4.3.	Conception des cubes dimensionnels:	81
4.3.1.	Introduction :.....	81
4.3.2.	Définition des niveaux et des hiérarchies :.....	81
4.3.3.	La liste des cubes :.....	83
4.3.4.	Conclusion :.....	86
5.	L'implémentation des DataMarts et du Business Intelligence (BI) :.....	87
5.1.	Introduction :.....	87
5.2.	Implémentation des DataMarts :.....	87
5.2.1.	DataMart Accident:	87
5.2.1.1.	Création des dimensions :.....	87
5.2.1.2.	Mapping et déploiement des dimensions:.....	89
5.2.1.3.	Création, Mapping et déploiement du cube :.....	90
5.2.2.	DataMart « Chauffeurs » :.....	93
5.2.3.	DataMart « Victimes » :.....	95
5.2.4.	Data Mart « Infractions » :.....	96
5.2.5.	Data Mart « Véhicules » :.....	97
5.3.	Validation, création et Déploiement des trois cubes	98
5.4.	Implémentation du BI :.....	99
5.4.1.	Création de référentiel.....	99
5.4.2.	Le chargement de référentiel dans le BI.....	100
5.4.3.	Système d'information géographique :.....	100
5.5.	Les interfaces de la restitutions.....	102
5.5.1.	Les différents sorties du tableau de bord réalisé.....	102
5.5.2.	Le tableau de bord pour androïde :.....	104
5.6.	Conclusion :.....	104
6.	Conclusion Générale et perspectives:.....	105
Annexe 01 :	Diagrammes en étoiles :.....	107
Annexe 02 :	Création, mapping et déploiement d'une dimension	112
Annexe 03 :	Mapping du cube Accident :.....	121

Annexe 04 : Mapping du cube Chauffeurs :.....	123
Annexe 05 : Mapping du cube Victime :	124
.....	122
Annexe 06 : Mapping du cube Infraction	126
.....	122
Annexe 07 : Mapping du cube Véhicule.....	127
Annexe 08 : Création et Déploiement du référentiel dans le OLAP	128
Annexe 09 : Préparation du OBIEE à l'analyse spatiale	132
Annexe 10 : Carte logique entre données sources et données cibles.....	136
Références Bibliographiques :.....	144

Liste des figures :

Figure 2.2.1 :Vue transversale de l'entreprise offerte par le système décisionnel [Source : Goglin 1998]	5
Figure 2.2.2: Le décisionnel au sein du Système d'information	6
Figure 2.2.3 : Les différents composants du décisionnel [Source : Goglin 1998].....	6
Figure 2.3.2 : Structure des données d'un Data Warehouse.....	9
Figure 2.3.3: les Data Marts dans un entrepôt de données selon l'architecture Entreprise Data Warehouse(E.D.W) [Inmon, 2002].....	10
Figure 2.3.4 : les Data Mart dans un entrepôt de données selon l'architecture.....	11
Figure 2.3.5 : L'architecture du data warehouse [Goglin, 1998].....	11
Figure 2.3.6 : L'architecture des systèmes décisionnels [Teste, 2000].....	12
Figure 2.3.7 : Principe de l'architecture ROLAP [Nakache, 1998]	15
Figure 2.3.8 : Principe de l'architecture MOLAP [Nakache, 1998].....	16
Figure 2.3.9 : L'acquisition des données	16
Figure 2.3.10 : Le processus d'acquisition des données.....	17
Figure 2.3.11 : L'alimentation des magasins de données (Datamarts).....	17
Figure 2.3.12 : La restitution des données.....	18
Figure 2.4.2 : Un modèle dimensionnel typique [Kimball, 1996].	21
Figure 2.4.3: Le modèle de données normalisé.....	23
Figure 2.4.5: Le modèle dé normalisé	24
Figure 2.4.6 : Le modèle en étoile.....	25
Figure 2.4.7 : Le modèle en flocon de neige.....	26
Figure 2.5.1: illustration de l'approche « Besoins d'analyse » grâce au cycle de vie dimensionnel de Kimball [Source : Kimball, 2004].....	27
Figure 2.5.2: Illustration de l'approche « Source de données » grâce au cycle de développement du DW de Inmon [Source : Inmon, 2002].....	28
Figure 2.5.3: Illustration de l'approche mixte.....	28
Figure 2.5.4 : Objectif de qualité de données dans un processus ETL [Source : Kimball, 2004]......	31
Figure 3.1.1 : La place de l'étape d'étude des besoins dans un projet Data Warehouse.....	34
Figure 4.1.1 : Analyse des priorités.....	43

Figure 4.1.2 : La dimension Temps du volet « Accident ».....	44
Figure 4.1.3 : La dimension « Type Accident » du volet « Accident ».....	45
Figure 4.1.4 : La dimension « Type Accident » du volet « Accident ».....	46
Figure 4.1.5 : La dimension « Type collision » du volet « Accident».....	46
Figure 4.1.6: La dimension « Construction sur la route » du volet « Accident».....	47
Figure 4.1.7 : La dimension « Commune » du volet « Accident».....	47
Figure 4.1.8: La dimension « Luminosité » (Illumination) du volet « Accident»	48
Figure 4.1.9 : La dimension « caractéristiques de la route » (Road Character) du volet « Accident».....	48
Figure 4.1.10 : La dimension « conditions de la route » (Road Condition) du volet « Accident»	49
Figure 4.1.11 : La dimension «spécificités de la route» (Road_Special_Feature_Dim) du volet « Accident».....	49
Figure 4.1.12 : La dimension « orientations de la circulation » (Road_Flow_Dim) du volet « Accident».....	50
Figure 4.1.13 : La dimension « densité de la circulation » (Road_Flow_Dim) du volet « Accident».....	50
Figure 4.1.14 : La dimension « nature de l'accident » (Nature_Accident_Dim) du volet « Accident»	50
Figure 4.1.15 : La dimension « Météo » (Weather_Dim) du volet « Accident».....	51
Figure 4.1.16 : La dimension « Défauts de la route » (Roadway_Surface_Dim) du volet « Accident»	51
Figure 4.1.17 : La dimension « Cause de l'accident » (Cause_Dim) du volet « Accident»	52
Figure 4.1.19 : La dimension « TYPE CHAUFFEUR » du volet « CHAUFFEUR»	58
Figure 4.1.20 : La dimension « TRANCHE_AGE » du volet « CHAUFFEUR».....	59
Figure 4.1.21 : La dimension « SEX_DIM » du volet « CHAUFFEUR».....	59
Figure 4.1.22 : La dimension «EDUCATION_DIM » du volet « CHAUFFEUR».....	59
Figure 4.1.23 : La dimension «PROPRIETAIRE_DIM » du volet « CHAUFFEUR»	60
Figure 4.1.24 : La dimension «PERMIS_ON_DIM » du volet « CHAUFFEUR»	60
Figure 4.1.25: La dimension «Nationalité » du volet « CHAUFFEUR».....	61
Figure 4.1.27 : La dimension «INFRACTION_DIM » du volet « INFRACTION»	64
Figure 4.1.29 : La dimension «NATURE_VICTIME_DIM » du volet « VICTIME».....	68
Figure 4.1.30 : La dimension «TYPE_VICTIME_DIM » du volet « VICTIME».....	68
Figure 4.1.32 : La dimension «Type du véhicule » du volet « VEHICULE»	72

Figure 4.1.33 : La dimension «Degrés d'endommagement » du volet « VEHICULE»	72
Figure 4.1.34 : La dimension «Marque du véhicule » du volet « VEHICULE».....	72
Figure 4.2.1 : Diagramme d'activité du processus ETL depuis sa mise en service	78
Figure 4.1.2 : Diagramme de processus de chargement d'une table de faits	79
Figure 5.2.1 : Schéma final du mapping.....	89
Figure 5.2.2 : Schéma dans sa vue global du mapping pour le cube « Accident ».	93
Figure 5.2.3 : Le Joiner Final des données de l'accident avec l'introduction de la dimension dégénérée « Accident_Reportum ».....	93
Figure 5.2.4: le mapping global du cube « Chauffeurs ».....	95
Figure 5.2.6 Schéma final du cube « Infraction ».....	97
Figure 5.2.7 Schéma final du cube « Véhicule ».....	98
Figure 5.3.1 Déploiement des cubes	98
Figure 5.4.1 Création des niveaux.....	100
Figure 5.5.1 : Première liste de prises d'écran du tableau de bord décisionnel sous version web	102
Figure 5.5.2 : Deuxième liste de prises d'écran du tableau de bord décisionnel sous version web.....	103
Figure 5.5.3 : Quelques prises d'écran du tableau de bord décisionnel sous version mobile	104
Figure A01.1 : Le modèle en étoile du volet Accident.....	107
Figure A01.2 : Le modèle en étoile du volet Chauffeurs.....	108
Figure A01.3 : Le modèle en étoile du volet Infractions.....	109
Figure A01.4 : Le modèle en étoile du volet « victimes ».....	110
Figure A01.5 : Le modèle en étoile du volet Véhicule.....	111
figure A02.1 : L'ajout d'une nouvelle dimension.....	112
figure A02.3 : Augmentation des tailles des attributs vers 500.....	113
figure A02.4 : Spécification des niveaux.....	113
figure A02.5 : Spécification des attributs des niveaux.....	114
figure A02.6 : Autocréation des dimensions commune_dim, la table Commune_dim_TAB et la séquenceCommune_dim_SEQ	114
figure A02.6 : Déploiement de la Commune_dim, Commune_dim_TAB et Commune_dim_SEQ.....	115

figure A02.7 : L'ajout d'un nouveau mapping	115
figure A02.8 : La sélection des tables sources	116
figure A02.9 : Regroupement des données avec l'opérateur de jointure « JOINER ».).....	116
figure A02.10 : Spécification de la condition de jointure.....	117
figure A02.11 : Liaison entre le JOINER et la dimension« DIM_INFRACTION.....	117
figure A02.12 : Validation et génération des scripts PISQL du « Commune_dim_map ».....	118
figure A02.13 : Déploiement du « commune_dim_map.....	118
figure A02.14 : Démarrage du « commune_dim_map ».....	119
figure A02.15 :Visualisation des données de la dimension « Commune_DIM »	119
figure A02.16 Les données de la dimension «Commune_DIM ».....	120
Figure A03.1 : Section N° 01 du mapping global du cube « Accident »	121
Figure A03.2 : Section N° 02 du mapping global du cube « Accident »	121
figure A03.3 : Section N° 03 du mapping global du cube « Accident ».....	122
Figure A04.1 : Section N° 01 du mapping global du cube « Chauffeurs »	123
Figure A04.2 : Section N° 02 du mapping global du cube « Chauffeurs »	123
Figure A04.3 : Section N° 03 du mapping global du cube « Chauffeurs » Jointure des données du Chauffeur avec celles des accidents, les agréger et les charger dans le cube	123
figure A05.1 Section 01 du Mapping du cube « Victimes ».....	124
figure A05.2 Section 02 du Mapping du cube « Victimes »	124
figure A05.3 Section 03 du Mapping du cube « Victimes ».....	125
Figure A05.4 Section 04 du Mapping du cube « Victimes »	125
Figure A06.1 Section 01 du Mapping du cube « Infractions ».....	126
Figure A06.2 Section 02 du Mapping du cube « Infractions ».....	126
Figure A07.1 Section 01 du Mapping du cube « Véhicule ».....	127
Figure A07.2 : Section 02 du Mapping du cube « Véhicule ».....	127
Figure A08.1 : La couche physique.....	128
Figure A08.2 : La couche logique.....	128
Figure A08.3 : La couche présentation.....	129

Figure A08.4 : Indication du type d'agrégation.....	129
Figure A08.5 : Création d'une dimension logique.....	130
Figure A08.6 : Le montage du référentiel.....	130
Figure A08.7 : Le redémarrage des services du BI.....	131
Figure A09.1 : La couche représentant les communes	132
Figure A09.2 : La couche représentant les wilayas	132
Figure A09.4 : Introduction des métas donnés de la carte de base.....	133
Figure A09.5 : Etablir la connexion du Oracle Mapviewer vers le schéma MVdemo	133
Figure A09.6 : Changer l'état de la carte de base vers « OnLine ».....	134
Figure A09.7 : Mise en correspondance des données de la dimension Commune_DIM avec les données attributaires de la carte correspondant à la commune.....	134
Figure A09.8 : Mise en correspondance des données de la dimension Commune_DIM avec les données attributaires de la carte correspondant à la wilaya	135
Figure A09.9 : Configuration des niveau de zoom pour la vision des couches wilaya et commune	135

Liste des Tableaux :

Tableau 2.4.1: Tableau comparatif entre les tables de faits et les tables de dimensions.....	22
Tableau 3.2.1 : Tableau présentant la population à interviewer.....	35
Tableau 3.5.1 : Synthétisation des besoins recensés.....	41
Tableau 4.1.1 : Tableau descriptif de la dimension « Temps ».....	45
Tableau 4.1.2 : Tableau descriptif de la dimension « Heure_DIM ».....	45
Tableau 4.1.3 : Tableau descriptif de la dimension « Type_Accident ».....	46
Tableau 4.1.4 : Tableau descriptif de la dimension « Type_Accident ».....	46
Tableau 4.1.5 : Tableau descriptif de la dimension « Construction_Dim ».....	47
Tableau 4.4.6 : Tableau descriptif de la dimension « Cmmune_Dim ».....	47
Tableau 4.1.7 : Tableau descriptif de la dimension « Illumination_Dim » (Luminosité).....	48
Tableau 4.1.8 : Tableau descriptif de la dimension «caractéristiques de la route» (Road_Character_Dim).....	49
Tableau 4.1.9 : Tableau descriptif de la dimension «conditions de la route» (Road_Condition_Dim).....	49
Tableau 4.1.10 : Tableau descriptif de la dimension «spécificités de la route» (Road_Special_Feature_Dim).....	49
Tableau 4.1.11 : Tableau descriptif de la dimension «orientations de la circulation» (Traffic_Flow_Dim).....	50
Tableau 4.1.12: Tableau descriptif de la dimension «densité de la circulation» (Traffic_Level_Dim).....	50
Tableau 4.1.13 : Tableau descriptif de la dimension «nature de l'accident» (Nature_Accident_Dim).....	51
Tableau 4.1.14 : Tableau descriptif de la dimension «météo» (Weather_Dim).....	51
Tableau 4.1.15 : Tableau descriptif de la dimension «défauts de la route» (Roadway_Surface_Dim).....	51
Tableau 4.1.16 : Tableau descriptif de la dimension «cause de l'accident» (Cause_Dim).....	52
Tableau 4.1.17 : Tableau descriptif des agrégats utiles du modèle « accident ».....	54
Tableau 4.1.18 : Détection des dimensions communes entre les volets « Accident » et « Chauffeur».....	56
Tableau 4.1.19: Jeux de données des dimensions pour le scénario de démonstration de la dégénération de la dimension « Accident_Reportnum ».....	57
Tableau 4.1.20 : Résultats de l'agrégation sans dégénération : L'accident A1 est calculé trois fois au lieu d'une seule fois	57
Tableau 4.1.20 : Résultat de l'agrégation avec la dégénération de la dimension « Accident_Reportnum ».....	58
Tableau 4.1.21 : Tableau descriptif de la dimension « TYPE CHAUFFEUR » du volet « CHAUFFEUR»	58
Tableau 4.1.21 : Tableau descriptif de la dimension « TRANCHE_AGE » du volet « CHAUFFEUR»	59

Tableau 4.1.22 : Tableau descriptif de la dimension « SEX_DIM » du volet « CHAUFFEUR»	59
Tableau 4.1.23 : Tableau descriptif de la dimension « EDUCATION_DIM » du volet « CHAUFFEUR»	60
Tableau 4.1.24 : Tableau descriptif de la dimension « PROPRIETAIRE_DIM » du volet « CHAUFFEUR»	60
Tableau 4.1.25 : Tableau descriptif de la dimension « PERMIS_ON_DIM » du volet « CHAUFFEUR»	60
Tableau 4.1.26 : Tableau descriptif de la dimension « Nationalité » du volet « CHAUFFEUR»	61
Tableau 4.1.27 : Tableau descriptif des agrégats utiles du modèle « accident ».....	62
Tableau 4.1.28 : Détection des dimensions communes entre les volets « Accident » et « Chauffeur» et « infraction ».....	64
Tableau 4.1.29: Tableau descriptif de la dimension « INFRACTION_DIM » du volet « INFRACTION»	65
Tableau 4.1.30 : Tableau descriptif des agrégats utiles du modèle « infraction ».	66
Tableau 4.1.31 : Détection des dimensions communes entre les volets « Accident » et « Chauffeur», « infraction » et « victimes ».....	68
Tableau 4.1.32 : Tableau descriptif de la dimension « NATURE_VICTIME_DIM » du volet « VICTIMES»	68
Tableau 4.1.33 : Tableau descriptif de la dimension « TYPE_VICTIME_DIM » du volet « VICTIMES»	68
Tableau 4.1.34 : Tableau descriptif des agrégats utiles du modèle « victime»	69
Tableau 4.1.35 : Détection des dimensions communes entre les volets « Accident », « Chauffeur», « infraction », « victimes » et véhicules.....	71
Tableau 4.1.36 : Tableau descriptif de la dimension « type du véhicule » du volet « VEHICULE»	72
Tableau 4.1.37 : Tableau descriptif de la dimension « Degrés d'endommagement » du volet « VEHICULE».....	72
Tableau 4.1.38 : Tableau descriptif de la dimension « Marque du véhicule » du volet « VEHICULE»	73
Tableau 4.1.39 : Tableau descriptif des agrégats utiles du modèle « Véhicule».	74
Tableau 4.3.1 : Liste des hiérarchies des dimensions	83
Tableau 4.3.2 : Liste des cubes avec leurs dimensions et hiérarchies	86
Tableau A10.1 : Tableau des données sources et leurs destination.....	143

Lexique

BI	Business Intelligence.
BSR	Brigade de la Sécurité Routière.
BSR	Bureau de la Sécurité Routière.
BT	Brigade Territoriale.
CGN	Commandement de la Gendarmerie Nationale.
CRGN	Commandement Régional de la Gendarmerie Nationale
CT	Compagnie Territoriale.
DIM	Dimension.
DSR.	Division de la Sécurité Routière.
DW	Data Warehouse (Entrepôt de données).
ED	Etude et développent.
EDW	Entreprise Data Warehouse.
EIS	Executive Information System
ESR	Escadron de la Sécurité Routière.
ETL	Extract-Transform-Load.
FK	Foreign Key.
GT	Groupement Territorial.
HOLAP	Hybrid On Line Analytical Processing.
MOLAP	Multidimensional On Line Analytical Processing.
OLAP	On Line Analytical Processing.
OLTP	On Line Transactional Processing.
PK	Primary Key.
ROLAP	Relational On Line Analytical Processing.
SGBD	Système de Gestion de Base de Données.
SI	Systèmes d'Information.
SIAD	Systèmes d'Information d'Aide à la Décision
SID	Systèmes d'Information Décisionnels.
SIG	Système d'Information Géographique.
SQL	Structured Query Language.

Chapitre Premier

Introduction Générale



1. Contexte
2. Problématique
3. Objectif du projet
4. Présentation du mémoire

1. Introduction Générale:

1.1. Contexte :

Les systèmes décisionnels, basés sur les entrepôts de données (data warehouses), sont apparus depuis le début des années 90 dans le but de mesurer l'efficacité et l'efficience de l'entreprise évoluant dans son environnement. Cette mesure permet de supporter les décideurs dans le processus de prise de décision. Le data warehouse assure la disponibilité de l'information et la met à la portée de tous les décideurs de l'entreprise.

L'information est une ressource très importante dans les entreprises, qui disposent de masses colossales d'informations non-exploitées dans leurs systèmes opérationnels. L'entrepôt de données est une solution efficace pour accéder et exploiter ces informations et en tirer de la plus-value décisionnelle.

Cette information décisionnelle permettra aux décideurs d'exercer leurs missions de prise de décision dans les moments opportuns et dans des délais très courts vu que l'environnement des entreprises est devenu de plus en plus complexe et en perpétuel changement. Prendre une décision rapide et efficace le plus rapidement possible devient primordial.

Ces décisions, qui affecteront la stratégie globale des entreprises, sont critiques et doivent être prises sur la base d'informations fiables et de qualité provenant de ses activités quotidiennes collectées par leurs systèmes opérationnels.

La prise de décision est influencée par l'identification et la présentation d'informations pertinentes ayant des valeurs analytiques et la définition des acteurs ayant besoin de cette information. En effet, le data warehouse est une technologie de stockage dédiée pour cette fonctionnalité. En revanche, les systèmes opérationnels « transactionnels » s'avèrent très limités, voire inaptes à fournir de telles informations et constituer par la même un support appréciable à la prise de décision.

La Gendarmerie Nationale, à l'instar de toutes les institutions de grande envergure, dispose de masses importantes de données affluant quotidiennement depuis ces unités implantées à travers le territoire national. Elle assure une mission critique qui est la sécurisation du territoire et des citoyens. Cette sécurisation passe par la maîtrise de certains fléaux tels les accidents de la route. Pour ce faire, elle doit exploiter les données collectées depuis le terrain afin de prendre des décisions visant à diminuer le nombre des accidents de la route et ainsi réduire leurs impacts sur la vie social, économique et la santé de la communauté.

1.2. Problématique :

Le réseau routier algérien, ne cessant de croître, ainsi que l'ampleur importante du parc automobile ont eu un impact direct sur l'augmentation du nombre des accidents de la route ainsi que les infractions liées au code de la route induisant à l'accident.

Selon les statistiques du Centre national de Prévention et de Sécurité Routières (CNPSR), 10.206 accidents ont eut lieu entre la période s'étalant du moi de Janvier au moi de Mai de l'année 2017, causant 1.697 décès et 17.350 blessés.

En 2016, les accidents de la route ont fait 4.610 morts dont plus de 1.500 âgés de moins de 30 ans, soit près de 34%, et les jeunes ont représenté 41% du nombre global des blessés enregistrés lors de la même

période (55.994), d'après la même source. Cela se traduit par une fréquence moyenne avoisinant les 12 morts et 160 blessés par jour avec tous ce qui s'en suit d'indisponibilité de travail, de nouvelles familles orphelines, des frais des soins pour les assurés, pour ne citer que peut, coutant à l'état 100 milliards DA par an.

Aucune région n'a été épargnée, sur autoroute ou route secondaire, ce n'est pas toujours une question d'infrastructures, mais à 90% à cause du facteur humain. Cela cause des pertes à une échelle nationale dans les domaines sociale, santé et économie que l'Algérie subira en matière de coût directes ou indirectes.

L'Algérie occupe la quatre-vingtième place dans le classement mondial, en termes de nombre de décès dans des accidents de la circulation, où il est enregistré un accident toutes les 20 minutes. Quant aux pertes financières dues aux accidents, celles-ci ont atteints les 100 milliards de dinars durant l'année 2013. Le facteur humain reste la première cause, puis arrive l'état des véhicules et enfin l'état des routes.

Face à ce fléau, la prise des bonnes décisions au moment opportun est devenue cruciale pour les dirigeants de la Gendarmerie Nationale, à tous les niveaux hiérarchiques, afin d'orienter et de mener les efforts de lutte contre l'hécatombe causée par les accidents de la route et les infractions liées au code de la route induisant à ces accidents. L'efficacité de cette prise de décision repose sur la mise à disposition d'informations pertinentes et d'outils adaptés.

Les unités territoriales de la Gendarmerie Nationales collectent les données et les saisissent quotidiennement dans les applications transactionnelles résultant en des grosses masses données fiables et d'un niveau de granularité très fine, offrant une vue cartésienne sur la situation ne permettant pas la prise de décision.

La maîtrise de ce fléau passe par l'analyse de ces grandes masses de données et leur mise en contexte afin de produire de l'information pour supporter la décision. Les bonnes décisions, prises aux moments opportuns permettront d'évoluer des simples solutions curatives vers des solutions préventives assurant : la réduction du nombre d'infractions liées au code de la route induisant vers les accidents routiers, le nombre d'accidents routiers et la gestion rationnelle des efforts des éléments de la Gendarmerie Nationale.

L'informatique décisionnelle apporte des solutions nouvelles pour la modélisation, l'interrogation et la visualisation de données dans un objectif d'aide à la décision. Les modèles multidimensionnels sont des modèles qui permettent de structurer les données pour l'analyse décisionnelle en explicitant la notion de dimension comme axe d'analyse (par exemple la dimension temps, la dimension « type infraction », la dimension période de l'année etc.), et des faits comme mesure ou indicateurs (Par exemple le nombre d'accidents par tranches d'âge dans une période donnée de l'année).

L'objectif de ce projet est de mettre en œuvre un data warehouse pour l'analyse des processus métier des services de sécurité routière de la Gendarmerie Nationale, liés aux accidents de la ruote, afin de consolider les données des systèmes d'information opérationnel (OLTP) vers une vision multidimensionnelle (OLAP). Ce data warehouse servira de base pour un tableau de bord décisionnel ainsi que les requêtes analytiques répondant aux besoins d'analyse et l'affichage sure carte géographique des statistiques.

1.3. Objectifs du projet :

Ce projet s'inscrit dans la finalité d'éliminer les problèmes cités précédemment, ou minimiser leurs impacts sur le processus de prise de décision.

Il a pour but l'instauration d'une démarche informatique décisionnelle, tout en assurant aux décideurs un support fiable pour une meilleure prise de décision. De ce fait, les principaux objectifs attribués au projet peuvent se résumer de la manière suivante:

- Offrir aux décideurs et aux analystes la possibilité de faire des analyses appropriées ;
- Offrir des informations fiables, cohérentes et pertinentes;
- La réduction et la maîtrise de la durée de l'élaboration des rapports statistique dea qualité exploitant toutes les données collectées depuis le terrain;

1.4. Présentation du mémoire :

Le présent mémoire est organisé comme suit : nous présentons, dans le chapitre 2, les concepts de base sur les systèmes décisionnels, particulièrement les entrepôts de données, le processus d'intégration ETL, la modélisation dimensionnelle ainsi que l'analyse en ligne OLAP. Le chapitre 3 est dédié pour l'étude des besoins des décideurs en termes d'analyse des données et d'aide à la décision. La conception du data warehouse est présentée dans le chapitre 4, il y sera question de définir les dimensions et leurs hiérarchies, les tables de faits et les agrégats et aura pour finalité la conception des modèles dimensionnels en étoile. Enfin, dans le chapitre 5, nous avançons les technologies utilisées et l'implémentation de l'ETL, du data warehouse et des data mart. Il y sera question, aussi, de parler du Business Intelligence et du tableau de bord décisionnel.

Chapitre deuxième :



Etude de la littérature

Sur

Les systèmes décisionnels « Decision-making system »

« Un entrepôt de données ne s'achète pas, il se construit. »

Bill Inmon

1. Introduction
2. Les systèmes d'information décisionnels.
3. Le Data Warehouse.
4. Modélisation des données de l'entrepôt.
5. Démarche de Construction d'un Data Warehouse.
6. Conclusion.

2. Les systèmes décisionnels « Decision-making system » :

2.1.Introduction :

Dans notre ère, où tout le monde semble connecté à tout moment ; une quantité colossale de données est générée quotidiennement noyant, ainsi, les entreprises dans une marée infinie de données devant être interprétées et mises en contexte pour en sortir la plus-value et les transformer en informations utiles pour les décideurs. Ces données sont issues soit de sources internes des systèmes transactionnels (OLTP : On-line Transactional Processing) , ou bien de sources externes (web, partenaire, .. etc.).

La quantité titanesque de données brutes et les limites techniques des systèmes OLTP en matière de stockage, gestion et analyse mènent, incontestablement, l'entreprise à converger vers une nouvelle solution informatique dite « décisionnelle » qui met l'accent sur la compréhension de l'entreprise et son environnement et l'exploitation des données et leurs interprétations pour des fins de prise de décisions.

Un besoin nouveau a émergé, poussant les décideurs à exiger une meilleure vision de l'environnement dans lequel leurs entreprises existent et évoluent, et le degré de véracité et de crédibilité des informations mises à leur disposition.

Cela est atteint à travers la définition et la mise en place de ratios ou d'indicateurs « business » suffisamment claires et pertinents permettant la persistance, l'utilisation de la mémoire de l'entreprise et mettant à disposition des décideurs, à tous les niveaux de la hiérarchie, une possibilité de s'y référer pour une bonne mesure des écarts et la correction de la stratégie par la prise des bonnes décisions aux moments opportuns.

2.2.Les systèmes d'information décisionnels

2.2.1. Qu'est-ce que le décisionnel ?

Le système d'information décisionnel 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. Les systèmes de gestion, ou OLTP, sont dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes, et directement opérationnels, car maintenus par les utilisateurs sur le terrain au quotidien. Les systèmes décisionnels, quant à eux, sont dédiés au management de l'entreprise pour l'aider au pilotage de ses activités. Ils constituent une synthèse d'informations opérationnelles, internes ou externes, choisies pour leurs pertinences et leurs transversalités fonctionnelles, et sont basées sur des structures particulières de stockage volumineux (data warehouses, bases OLAP). Le principal intérêt d'un système décisionnel est d'offrir au décideur une vision transversale de l'entreprise intégrant toutes ses dimensions.[Source : Goglin1998]

Cette vue intégrée peut alors être étudiée par fonction ou par métier.

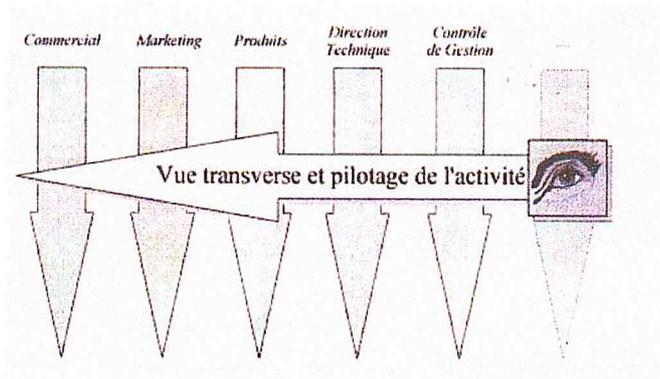


Figure 2.2.1 : Vue transversale de l'entreprise offerte par le système décisionnel [Source : Goglin 1998]

Afin de mieux comprendre la finalité des systèmes décisionnels, nous nous devons de les placer dans leurs contextes et rappeler ce qu'est un système d'information. *«Le système d'information est l'ensemble des méthodes et moyens de recueil, de contrôle et de distribution des informations nécessaires à l'exercice de l'activité en tout point de l'organisation. Il a pour fonction de produire et de mémoriser les informations, de l'activité du système opérant (système opérationnel), puis de les mettre à disposition du système de décision (système de pilotage)»*[Le Moigne, 1977].

Les différences qui existent entre le système de pilotage et le système opérationnel, du point de vue fonctionnel ou des tâches à effectuer, conduit à l'apparition des « systèmes d'information décisionnels » (S.I.D.). Ces différences seront clairement illustrées un peu plus loin dans ce document.

Les origines des SID remontent au début de l'informatique et des systèmes d'information qui ont, tous deux, connu une grande et complexe évolution, cette évolution se poursuit à ce jour.

Parmi les différentes définitions du décisionnel, proposées dans la littérature, on trouve : *«Le Décisionnel est le processus visant à transformer les données en informations et, par l'intermédiaire d'interrogations successives, transformer ces informations en connaissances.»* [Dresner, 2001].

2.2.2. Situation du décisionnel au sein du système d'information :

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. [Source : Goglin1998] Elle s'étend et comprend plusieurs fonctions clés.

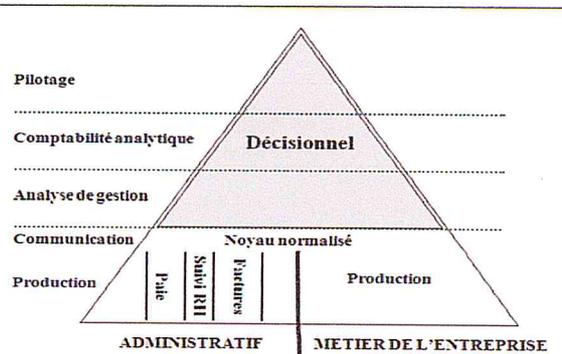


Figure 2.2.2: Le décisionnel au sein du Système d'information

Les frontières du décisionnel commencent une fois les données brutes (opérationnelles) sont valorisées, mises en contexte et transformées en informations. Cette valorisation est effective dès que l'on sort du monde de la production (OLTP).

2.2.3. Les composantes du décisionnel :

Les premiers systèmes de pilotage consistaient en des outils qui, à travers des requêtes, fournissaient des tableaux de bord (Dashboard). Ces outils se sont, ensuite, enrichis de fonctionnalités de simulation et d'interfaces de présentation. Puis, l'arrivée des outils EIS et SIAD. Ces outils, avec toute leur puissance, n'offraient qu'un snapshot à deux dimensions d'une situation particulière. Avec ces outils, on était : « capable d'identifier un dysfonctionnement, mais pas d'en connaître la cause ». [Source : Goglin1998].

Afin de pouvoir déceler les causes, il fallait introduire la dimension de « l'agrégation » permettant d'expliquer la cause de l'information étudiée. Cette nouvelle dimension a été introduite par les systèmes « multidimensionnels » au travers des systèmes OLAP (On-Line Analytical Processing).

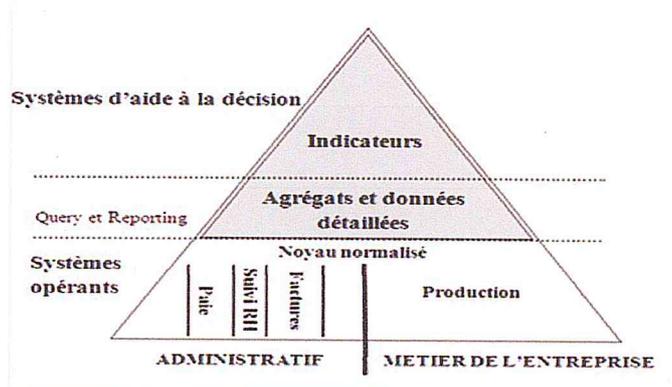


Figure 2.2.3 : Les différents composants du décisionnel [Source : Goglin 1998]

La convivialité et la puissance des outils disponibles ont emmené les décideurs à les adopter et se les approprier. Les décideurs s'ingénierent à trouver de plus en plus d'indicateurs toujours plus astucieux et plus compliqués. La fièvre des indicateurs naquit.

Cependant, cette course vers l'indicateur a eut pour effet néfaste de submerger les décideurs sous une masse colossale de tableaux de bord et de synthèses dont ils n'arrivaient plus à extraire la substantifique moelle. Alors, un moyen visuel a été introduit pour localiser les informations pertinentes plus rapidement, ce fut l'introduction du « color-coding » et des systèmes d'information cartographiques offrant une identification visuelle des informations cruciales selon les critères du décideur, puis de les représenter sous une forme directement exploitable par une équipe de direction.

Un nouveau besoin surgit, alors, qui est la diffusion instantanée, en temps réel, de toutes ces informations et toutes ces analyses vers tous les cadres concernés. C'est ce qui a motivé le développement des réseaux, des activités de « Workflow » et Internet.

Malgré toute sa puissance et son évolution, le décisionnel connaît une manque: Il est en mesure de fournir les informations nécessaires au décideur et de corréler des événements apparemment non liés, mais il est dans l'incapacité d'assister le décideur dans sa prise de décision. Le syndrome de « la non prise de décision » atteint le décideur.

De là jaillit la nécessité de délimiter les seuils à partir desquels les valeurs des indicateurs sont considérées comme anormales, puis définir les règles utilisant ces seuils pour construire un diagnostic. Ce diagnostic servira d'appui pour la prise de la décision finale. Ces nouveaux outils sont appelés les moteurs de règles de gestion ayant pour but d'assister le décideur dans sa prise de décision.

2.3.Le Data Warehouse :

2.3.1. Définition:

« **Bill Inmon** » définit le data warehouse comme suite: « *Un datawarehouse est une collection de données thématiques, intégrées, non volatiles et historiées organisées pour la prise de décision* ».

Un data warehouse est une ensemble de données caractérisées par:

- **Thématiques (orientées sujet)** : la finalité d'un data warehouse est la prise de décisions autour des activités phares de l'entreprise. Un rassemblement des informations par thèmes, contrairement aux modélisations traditionnelles, qui regroupent les informations par fonctions. Cela permet de passer d'une vision verticale de l'entreprise à une vision transversale beaucoup plus riche.
- **Intégrées** : L'efficacité convoitée par la transversalité sera atteinte, réellement, une fois le système d'information sera intégré. Cette intégration nécessitera une forte normalisation, une bonne gestion des référentiels et de la cohérence, une parfaite maîtrise de la sémantique et des règles de gestion s'appliquant aux données manipulées.
- **Non volatiles (pas de suppression)** : les informations stockées dans un data warehouse doivent persister et ne sont pas sujets à des suppressions. Cela étant dans un but de garder la traçabilité informations et des décisions prises.
- **Evolutives dans le temps**: Les différentes valeurs prises par une donnée sont est cruciales dans un système décisionnel, cela offre les comparaisons et le suivi de l'évolution des valeurs d'une

façon chronologique. Dans un système opérationnel, la valeur d'une donnée est simplement mise à jour. Dans un Data Warehouse, chaque valeur est associée à un moment "Every key structure in the data warehouse contains - implicitly or explicitly - an element of time" [Source : Inmon, 2000]. Organisé pour la prise de décision. [Source : Kimball1, 1999]

2.3.2. Historique des Data Warehouse :

Le « Data Warehouse » (entrepôt de données) trouve ses origines dans les années 80, période qui a connu un intérêt croissant au système décisionnel, conséquence de l'émergence des SGBD relationnels, la simplicité du modèle relationnel et la puissance offerte par le langage SQL. Le Data Warehouse se limitait à représenter une copie des données du système opérationnel, périodiquement générée, qui était dédiée à un environnement de support à la prise de décision.

En effet, on puisait les données du système opérationnel, ces données étaient, par la suite, stockées dans une nouvelle base de données, ce qui constituait le concept « *d'infocentre* ». Le but principal de cet infocentre étant de satisfaire aux requêtes des décideurs sans pour autant diminuer les performances des systèmes opérationnels.

Le Data Warehouse, tel que connu de nos jours, a évolué depuis la vue simpliste d'une simple copie périodique des données opérationnels. Il représente une source d'information fiable et crédible, alimenté avec des données recueillies, consolidées, vérifiées et transformées des différentes sources internes et externes.

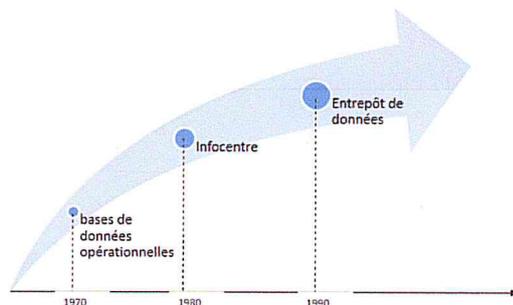


Figure 2.3.1 : évolution des bases de données décisionnelles.

2.3.3. Structure des données d'un Data Warehouse :

Inmon [Source : Inmon, 2000] définit la structure d'un data warehouse selon différents niveaux d'agrégations et de détails des données, comme illustré ci-après :

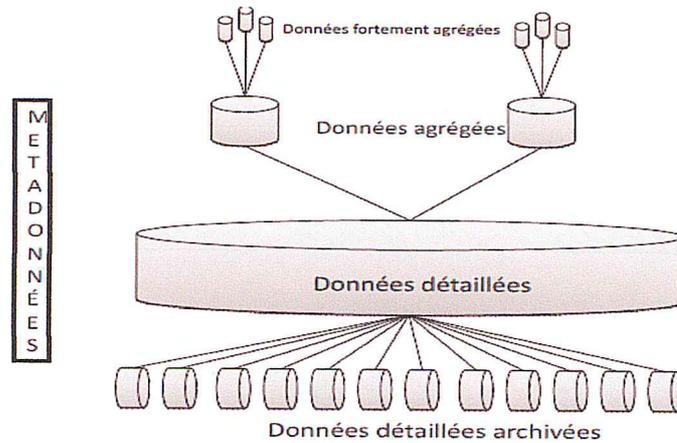


Figure 2.3.2 : Structure des données d'un Data Warehouse.

- **Données détaillées** : Représentent les données reflétant les événements les plus récents, omni-consultées, volumineuses de nature due à leur niveau de détail élevé et de leur fine granularité.
- **Données détaillées archivées** : représentent les données anciennes, exceptionnellement sollicitées, persistées généralement dans un disque de stockage de masse, peu onéreux, et qui ont un niveau de granularité équivalent aux données détaillées.
- **Données agrégées** : de granularité moins fine, données agrégées en se basant sur les données détaillées.
- **Données fortement agrégées** : données agrégées à partir des données détaillées, à un niveau d'agrégation plus élevé que les données agrégées.
- **Meta données** : Ce sont des données sur les données. Elles assurent des informations référant à la structure des données, les méthodes d'agrégation et surtout le lien existant entre les données opérationnelles sources et celles du Data Warehouse. Les métadonnées doivent fournir des renseignements sur :
 - Le modèle de données ;
 - La structure des données telle qu'elle est vue par les développeurs ;
 - La structure des données telle qu'elle est vue par les utilisateurs ;
 - Les sources des données ;
 - Les transformations nécessaires ;
 - Suivi des alimentations.

2.3.4. Composantes d'un Data Warehouse :

L'environnement du Data Warehouse comprend quatre composantes : les applications opérationnelles, la zone de préparation des données, la présentation des données et les outils de manipulation et d'accès aux données.

- **Les applications opérationnelles** : ce sont les applications du système opérationnel ou de production de l'entreprise, assurant le fonctionnement, au quotidien, et la performance du système. Ces applications sont extérieures au Data Warehouse.
- **Préparation des données** : C'est le niveau situé entre les applications opérationnelles et la présentation des données. Elle comporte tout ce qu'il y a entre les applications opérationnelles et la présentation des données. Elle est, communément, désignée par l'acronyme ETL, « Extract, transform and Load ». Les données sont extraites et stockées dans un espace tampon pour être assujetties aux transformations nécessaires avant leur chargement. « *Un point très important, dans l'aménagement d'un entrepôt de données, est d'interdire aux utilisateurs l'accès à la zone de préparation des données, qui ne fournit aucun service de requête ou de présentation* » [Kimball, 2002].
- **Présentation des données** : c'est l'entrepôt où les données sont organisées et stockées. A l'encontre des données de la zone de préparation qui sont interdites d'accès aux utilisateurs, la zone de présentation est accessible aux utilisateurs via des outils d'accès. Le data warehouse est constitué d'un ensemble de Data Marts (Magasin de données). Le data mart est une représentation réduite miniature d'un Data Warehouse, construit autour d'un sujet précis d'analyse ou consacré à un niveau départemental¹. Cette différence de construction, autour d'un sujet ou au niveau départemental, définit la façon d'implémentation du Data Mart au niveau de l'entrepôt. On distingue, en effet, deux architectures internes du Data Warehouse :

a. Data Mart indépendant :

Les Data Marts sont des versions miniaturisées du Data Warehouse au niveau départemental, alimentées par le Data Warehouse en informations [Inmon, 2002].

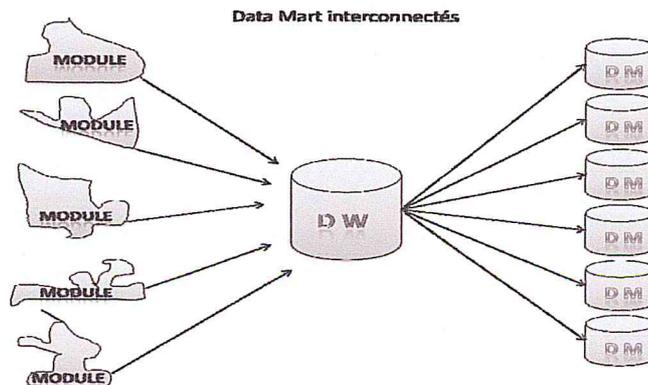


Figure 2.3.3: les Data Marts dans un entrepôt de données selon l'architecture Entreprise Data Warehouse(E.D.W) [Inmon, 2002].

b. Datamart interconnectés :

Ils sont implémentés et construit autour de sujets, interconnectés via les tables des faits (et des dimensions du data warehouse). Dans ce cas, le data warehouse est composé de ces Data Marts et ces tables des faits, appelées bus².

¹ Synthétisation [Chuck, 1998] page 86.

² Appellation proposée par R. Kimball dans son ouvrage [Kimball, 2002].

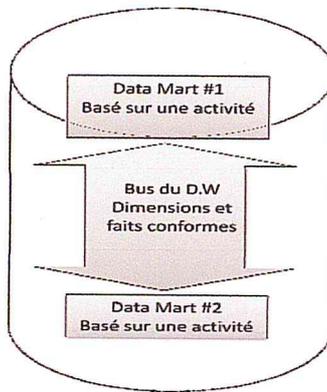


Figure 2.3.4 : les Data Mart dans un entrepôt de données selon l'architecture

- **Zone d'outils d'accès** : représente la myriade des moyens mis à disposition des utilisateurs du Data Warehouse pour l'exploitation et l'interaction avec la zone de présentation des données pour la prise de décision. Ces outils peuvent se manifester sous forme de requêtes « ad-hoc » simples ou sous forme plus sophistiquée à travers des outils permettant l'application de forage de données plus complexes. La majorité des utilisateurs (de 80% à 90%) interagissent avec des applications d'analyses développées au préalable, se manifestant sous forme de requêtes préétablies.

2.3.5. L'architecture globale du data warehouse

Le schéma suivant illustre l'architecture générique d'un système data warehouse :

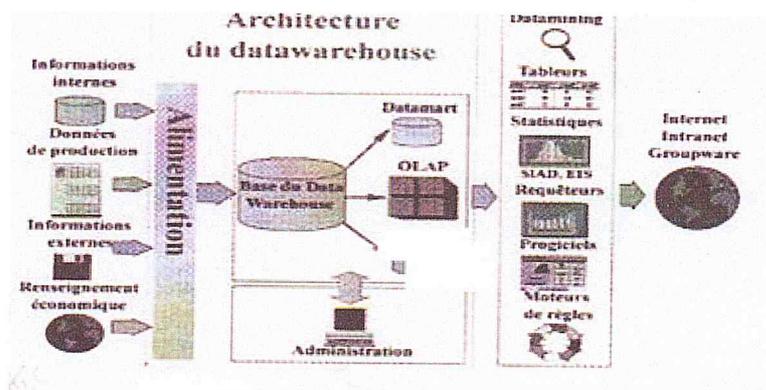


Figure 2.3.5 : L'architecture du data warehouse [Goglin, 1998]

L'architecture des systèmes décisionnels englobe les quatre éléments suivants :

- Les sources de données ;
- L'entrepôt de données ;
- Les magasins de données ;
- Les outils d'analyse et d'interrogation.

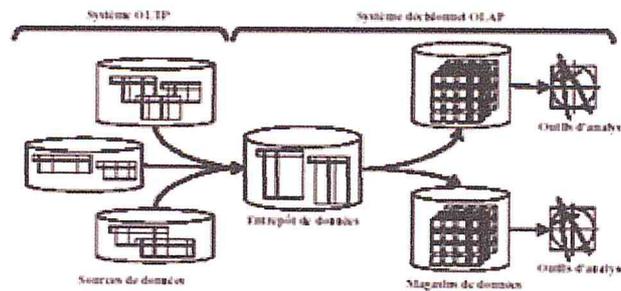


Figure 2.3.6 : L'architecture des systèmes décisionnels [Teste, 2000]

- **Les sources de données** : sont souvent hétérogènes, diversifiées, nombreuses, distribuées et indépendantes. Elles peuvent être internes (bases de production) ou externes (Internet, bases des partenaires) à l'entreprise.
- **L'entrepôt de données (data warehouse)** représente un lieu de stockage centralisé des informations utiles et fiables convoitées par les décideurs. Il met en commun et consolide les données provenant des sources de données hétérogènes en matière de technologies et de modélisations, et persiste leurs évolutions dans le temps (chronologiquement).
- **Les magasins de données (data marts)** sont des vues restreintes de l'entrepôt orientés sujet, dédiés à un département. Les données sont organisées de manière adéquate pour permettre des analyses rapides à des fins de prise de décision, sans pour autant accéder à la totalité du data warehouse.
- **Les outils d'analyse** offrent aux utilisateurs les moyens de manipuler les données suivant des axes d'analyses. L'information est affichée via des interfaces interactives et fonctionnelles vouées à des décideurs souvent non informaticiens (directeurs, chefs de services,...).

2.3.6. Définitions des concepts [Source : Goglin1998]

Dans ce qui suit, une liste des concepts phares utilisés dans le domaine du data warehousing, selon [Goglin 1998] :

a) Base de production :

On appelle, d'une façon générale, bases de production toutes les sources (qu'il s'agit de données de production, d'informations internes ou d'informations externes quel que soit leur mode de stockage) dont il va falloir extraire des données en vue d'alimenter le data warehouse

b) L'alimentation ou transformation :

Les outils d'alimentation sont utilisés pour extraire les données des bases de production et des bases d'informations internes et externes, pour les convertir, les transformer et enfin les stocker dans le data warehouse.

c) La base de données du data warehouse :

Chapitre Deuxième : Etude de la Littératures

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 présentes.

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.

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 sa création.

d) Datamart :

Un datamart est un magasin de données. Il s'agit d'une solution départementale d'entrepôt de données (data warehouse) supportant une partie des données et fonctions de l'entreprise (produit, département, activité, etc.). C'est un sous-ensemble du data warehouse qui ne contient que les données d'un métier de l'entreprise alors que le data warehouse contient toutes les données décisionnelles de l'entreprise pour tous les métiers.

e) OLTP (On-line Analytical 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.

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.

f) Data mining :

Les outils de datamining, également appelés « fouille des données » ou « extraction de la connaissance », s'appuient sur le constat qu'il existe au sein de chaque entreprise des informations dont le sens ou les liens sont cachés dans le gisement des données de l'entreprise. Le datamining permet de faire apparaître des corrélations dans des gisements de données.

Le datamining est l'étape logique suivant le data warehousing puisqu'il consiste en l'analyse des données composant le data warehouse à l'aide d'outils spécialisés en intelligence artificielle, visant à mettre en exergue des corrélations non apparentes par des analyses de premier niveau.

g) EIS :

Un EIS (Executive Information System) est un outil de visualisation des données et de navigation, permettant de constituer des tableaux de bord. Il est constitué d'outils qui permettent aux différents niveaux de management d'accéder aux informations essentielles de leur organisation, de les analyser et de les présenter de façon élaborée. Ces outils sont dotés d'une interface graphique très conviviale et très esthétique.

Chapitre Deuxième : Etude de la Littératures

L'utilisateur final ne peut visualiser ou exploiter que des informations qui ont été prévues par le concepteur des tableaux de bord. A la différence d'un SIAD, l'EIS ne permet pas à l'utilisateur final de poser une question qui n'aurait pas été prévue initialement.

h) SIAD :

Un SIAD (Système Interactif d'Aide à la Décision) est un outil d'analyse et de modélisation des données de l'entreprise qui permet de créer des représentations multidimensionnelles de l'information. Historiquement, il s'agit d'une terminologie et d'outils utilisés avant l'avènement et la maturité du data warehouse.

i) Requêteur :

Un requêteur permet à l'utilisateur final d'accéder aux données de l'entreprise de manière autonome, dans un langage propre à son métier, mais qui nécessite généralement la connaissance de la structure de la base accédée, et ce, en définissant lui-même les informations qu'il veut obtenir ainsi que le format des situations souhaitées.

j) Progiciels :

Ce sont des applications packagées orientées vers un ou plusieurs métiers dédiés (marketing, logistique, finance, ressources humaines, etc.).

k) Moteur de règle de gestion :

Un moteur de règles de gestion est un outil permettant de gérer le patrimoine d'une entreprise qu'est son métier, cristallisé par un ensemble de règles de gestion constituant son expertise et son savoir-faire.

l) Internet/Intranet :

Internet en tant que vecteur de communication normalisé et banalisé répond parfaitement aux problématiques d'accès banalisé et de publication à distance et à faible coût. Internet permet d'envisager, par exemple :

- L'envoi par messagerie électronique et donc la publication des analyses effectuées ;
- La possibilité pour un interlocuteur distant de se connecter pour connaître les dernières évolutions des ventes par exemple ;
- La mise à disposition, au sein d'un serveur Web par exemple, d'informations internes comme les statistiques de production ou externes comme la présentation de la société.

2.3.7. Le concept OLAP :

2.3.7.1. Généralités:

OLAP est un acronyme, en langue anglaise, signifiant « On-Line Analytical Processing ». Le terme correspondant en français est « traitement analytique en ligne ». Il désigne l'ensemble des technologies dédiées à l'accès et l'analyse instantanés (ou presque instantanés « Near-Real-Time ») aux données, dans la finalité d'honorer les besoins des décideurs en terme de reporting et d'analyse.

Dans sa définition, R. Kimball, introduit le concept « OLAP » comme « *Activité globale de requêtage et de présentation de données textuelles et numériques contenues dans l'entrepôt de données; Style d'interrogation spécifiquement dimensionnel* » [Kimball, 2005].

Cependant, le terme OLAP a été introduit, la première fois, par le précurseur *Edgar F. Codd*, en l'année 1993 dans son document technique « Providing OLAP (On-Line Analytical Processing) to User-Analysts : An IT Man-date » [Codd, 1993]³.

2.3.7.2. Architecture des systèmes OLAP

Deux types principaux d'OLAP existent: le MOLAP (Multidimensional Online Analytical Processing) et le ROLAP (Relational Online Analytical Processing), alors que le HOLAP (Hybrid Online Analytical Processing) désigne l'ensemble des technologies qui combinant les deux types MOLAP et ROLAP. Chacun de ces types est caractérisé par ses avantages et ses inconvénients.⁴

a) Les systèmes à architecture ROLAP:

Richard Kimble introduit le ROLAP comme suite: « Ensemble d'interfaces utilisateurs et d'applications qui donnent une vision dimensionnelle à des bases de données relationnelles » [Kimball, 2005].

Ce type manipule les données enregistrées dans des bases de données relationnelles et simule les opérations traditionnelles de «slicing and dicing» d'OLAP. En fait, toute action de «slicing and dicing » est équivalente à ajouter une clause « WHERE » à l'instruction SQL.

Il n'y a pas de pré-agrégation dans ROLAP. Les règles d'agréments sont pré-générées et implémentées dans une table relationnelle. En conséquence, ROLAP mène vers un temps de réponse plus long et une lourdeur d'administration. Cependant, il assure la cohérence lors de l'utilisation. ROLAP est préférable lors de grandes masses de données.

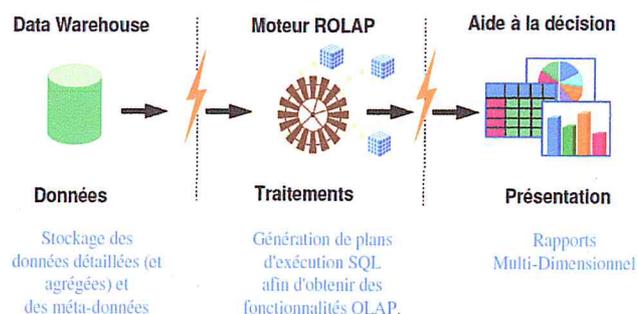


Figure 2.3.7 : Principe de l'architecture ROLAP [Nakache, 1998].

b) Les systèmes à architecture MOLAP:

R. Kimball définit ces systèmes comme étant un « Ensemble d'interfaces utilisateur, d'applications et de technologies de bases de données propriétaire dont l'aspect dimensionnel est prépondérant » [Kimball, 2005].

Ce type représente le mode traditionnel des analyses OLAP. Dans MOLAP, les données sont persistées dans un cube dimensionnel. Le stockage ne se situe pas dans une base de données relationnelle, mais dans des formats propriétaires.

³ Ce document définit les 12 règles de Codd que doit respecter une base de données si elle veut adhérer au concept OLAP (voir Annex A)

⁴ Pour plus d'informations sur MOLAP, ROLAP: avantages et inconvénients, se référer à l'Annex B

Dans MOLAP, l'agrégation des données est faite par défaut, cette agrégation consomme les ressources et induit à une lourdeur proportionnelle à la taille des données à traiter, ce qui le rend un mauvais choix lors des quantités colossales des données à analyser. On parle généralement de volume de l'ordre du giga-octet pas plus.

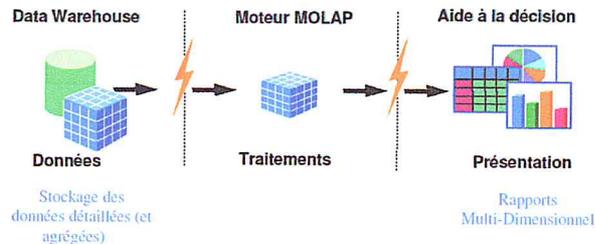


Figure 2.3.8 : Principe de l'architecture MOLAP [Nakache, 1998].

c) Les systèmes à architecture HOLAP:

Le type HOLAP, pour « Hybrid Online Analytical Processing », représente un compromis des avantages des types MOLAP et ROLAP. Pour les données agrégées, HOLAP fait recours aux technologies du cube pour améliorer les performances. Cela dit, HOLAP va au-delà le cube et accède aux données relationnelles sous-jacentes pour puiser des détails non disponibles dans le cube.

d) Autres architectures OLAP:

Les types ROLAP, MOLAP et HOLAP sont les plus répandus dans l'industrie du décisionnel. Cependant, d'autres systèmes à architectures différentes existent en parallèle, tels que OOLAP « Object On-line Analytical Processing », ou alors DOLAP « Desktop On-line Analytical Processing ». Ils ne se connectent, généralement, pas à des serveurs mais vers des cubes conçus, implémentés, stockés et exploités sur machines locales des utilisateurs.

2.3.8. Fonctionnement du système data warehouse:

L'architecture du système data warehouse étant introduite, passons à son fonctionnement qui passe par trois phases:

2.3.8.1. L'acquisition des données :



Figure 2.3.9 : L'acquisition des données

Cette phase représente l'interface entre les systèmes opérationnels et le système décisionnel. Elle détermine la faisabilité de la solution. Elle consiste en l'extraction des données utiles et pertinentes qui sont caractérisées par leurs hétérogénéités, dispersions et complexités.

Les problèmes liés à l'alimentation du data warehouse depuis les sources peuvent être résolus par l'instauration d'un processus en Cinque phases: *découverte*, *extractions*, *transformations*, *transport* et enfin *le chargement*.

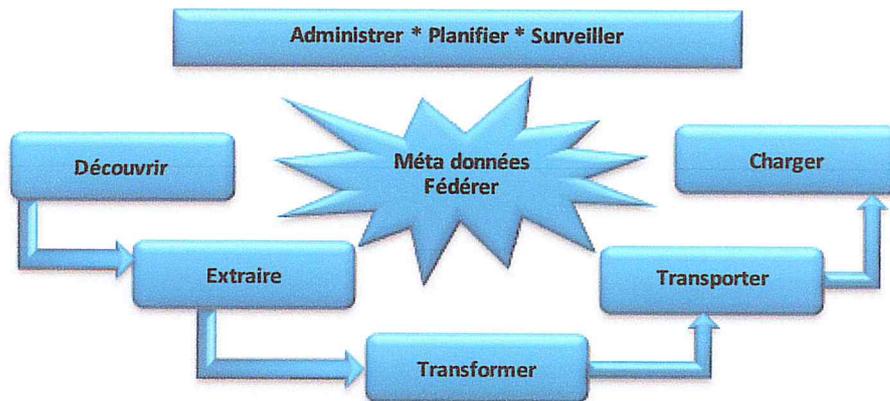


Figure 2.3.10 : Le processus d'acquisition des données

2.3.8.2. Le stockage des données dans le data warehouse :

C'est la phase de persistance des données: orientées sujet, intégrées, non volatiles et historisées afin d'assurer le support du processus décisionnel. Le data warehouse est le point central de stockage de toutes les données de l'entreprise.

2.3.8.3. L'alimentation des datamarts (Les magasins de données) :

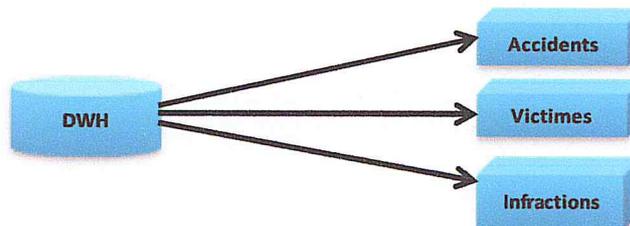


Figure 2.3.11 : L'alimentation des magasins de données (Datamarts)

- Dans l'approche Top-down⁵, le data warehouse représente le point central de contrôle, il garantit la qualité et l'intégrité de l'information. Sa finalité est l'optimisation de l'approvisionnement des data marts.
- Chaque Datamart est conçu et implémenté pour répondre à un sujet bien spécifique de l'entreprise. Les données y sont structurées et filtrées en fonction de la problématique en face. Cela mène à un ensemble réduit de données comparé aux données globales de l'entreprise, ce qui rend le stockage multidimensionnel le plus adéquat.

⁵ Il existe deux approches pour la construction des data warehouse: Top-down et l'approche bottom-up. Ce projet se base sur une approche hybride conçue des deux approches.

2.3.8.4. L'exploitation de l'information :

C'est l'interface avec l'utilisateur final et la finalité du système décisionnel. Le choix des outils de restitution est crucial pour l'adoption du système par les utilisateurs finaux. Il faut opter pour des outils qui répondent au mieux aux attentes des utilisateurs en besoins en informations et aide à la décision sous forme de: Reporting, Cubes analytiques, Requêtage et Data mining.

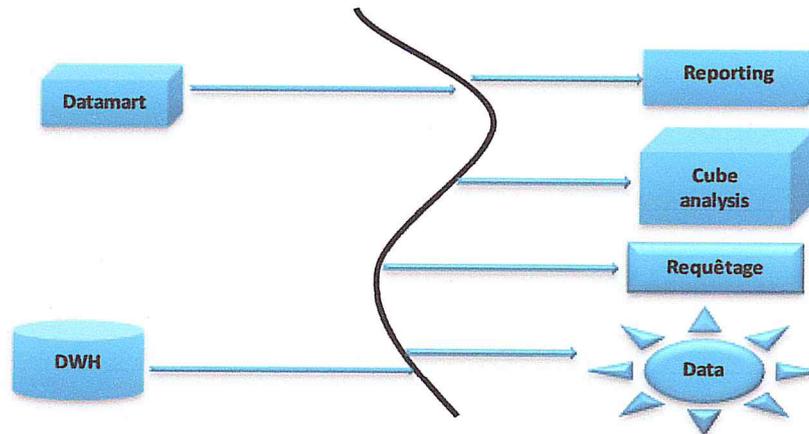


Figure 2.3.12 : La restitution des données

2.4. Modélisation des données de l'entrepôt :

Un data warehouse est avant tout une base de données servant de source pour le système décisionnel. Le modèle de données du data warehouse devient, alors, le cœur du système décisionnel à concevoir. Le système décisionnel diffère radicalement des systèmes d'informations classiques de par son objectif et son utilisation, cela implique l'utilisation d'approches de modélisation spécifiques.

2.4.1. Caractéristiques d'utilisation: OLTP vs OLAP:

Plusieurs techniques de modélisation de données ont initialement été conçues pour satisfaire à des problématiques autres que celles dans le cadre d'un système décisionnel.

Ainsi, pour la modélisation des systèmes d'informations classiques, les approches centrées sur des méthodologies telles que *Merise* sont adoptées, essentiellement, les composantes liées à la modélisation des données. Ces méthodes, qui ont fait leurs preuves, sont caractérisées par leurs précisions, leurs caractères standards ainsi que leurs puissances. Le modèle « entité-association » en est un exemple et fait preuve d'une utilisation très répandue dans entreprises pour la génération des modèles relationnels.

Cependant, ces techniques ont été élaborées dans une finalité d'automatiser des processus à caractère transactionnel, ce qui est désigné par OLTP (On-line Analytical Processing).

L'informatique de décision, ou système OLAP (On-line Analytical Processing), avec ces exigences spécifiques et les nouvelles problématiques qu'elle introduit, requiert une remise en cause de ces méthodes de conception de modèle de données.

a. Caractéristiques d'un modèle dans le contexte OLTP:

- **Persistance des données transactionnelles:** les modèles sont utilisés, principalement, pour garantir le stockage et la persistance des données transactionnelles. La base de données est conçue et implémentée dans la finalité de garder la trace des évènements, apparaissant en temps réel, dans l'entreprise.
- **Intégrité des données:** les modèles de donnée préservent et garantissent la fiabilité, cohérence et intégrité du système, cela passe par la minimisation des redondances. Des théories et des procédés mathématiques découlant de la théorie des ensembles assurent cette intégrité à travers des concepts, tels que les formes normales, les clés uniques, les clés étrangères ou de contraintes d'intégrité référentielle. Du fait que les systèmes transactionnels sont mis à jours en ligne et avec une fréquence très élevée, la minimisation des redondances devient primordiale afin de maintenir l'intégrité du système transactionnel. Il y'a lieu, aussi, d'utiliser des applications jouant le rôle d'interfaces entre la base de données et les utilisateurs finaux.
- **Model orienté processus:** contrairement aux systèmes décisionnels, la conception des systèmes transactionnels est orientée processus, ainsi, le modèle de données doit supporter cette orientation.
- **Requêtes préétablies et pré-optimisées traitant un ensemble réduit de données:** parce que le système transactionnel est orienté processus, il est prévisible et connu en amont. Cela permet de prévoir les requêtes au préalable et même de les intégrer dans des applications. Ces requêtes peuvent être optimisées par la création du bon ensemble d'indexe et l'utilisation des clés uniques améliorant, ainsi, le temps de réponse; cependant, ce temps de réponse reste toujours dépendant du volume de données. La masse de données accédées ou retournées par une transaction est limitée. Les requêtes transactionnelles traitent un ensemble réduit de données et elles ne sont pas le choix idéal pour rassembler ou agréger des informations issues d'un grand nombre de table.

b. Dans le contexte OLAP (modèle décisionnel):

Dans un contexte décisionnel,

- **Optimisation des requêtes par anticipation:** due à la redondance accrue des données, l'absence de la normalisation et la complexité des requêtes, l'optimisation de ces dernières consisterait, dès lors, à l'anticipation sur les chemins d'accès aux données fréquemment employées, au lieu de procéder à l'optimisation requête par requête ;
- **Mise à jours en différé lors de la phase acquisition des données:** Un data warehouse est un lieu de stockage des données décisionnelles dédiées à l'analyse. L'information est mise à la portée des utilisateurs mais sans leur permettre de les mettre à jour en ligne, comme c'est le cas dans les systèmes OLTP. Cependant, le data warehouse est mis à jour depuis les systèmes de production, durant les phases de chargement (processus d'acquisition de données). Les redondances éventuelles seront introduites, à condition de les maîtriser dans le processus d'alimentation ;
- **Manipulation de grandes masses de données:** Dans un contexte décisionnel, les requêtes effectuent des sélections ou des restrictions de population, des regroupements, des calculs, des agrégations, etc. sur de grands ensembles de données, dans le but de répondre aux

besoins des utilisateurs. Il faudra souvent manipuler de gros volumes d'information pour avoir un résultat de quelques lignes. Cela rend plus difficile de prévoir le temps de réponse par rapport aux résultats des requêtes comme est le cas dans le contexte transactionnel. Afin de maîtriser le temps de réponse, les requêtes fréquemment utilisées, peuvent être optimisées en prédéfinissant physiquement des sous-ensembles de données, moins importants en taille que les données plus détaillées, mais suffisants pour résoudre les requêtes les plus courantes ;

- **Découvrir des liens cachés et des modèles (paternes) entre les données apparemment disparates:** le décisionnel permet la corrélation d'éléments dont la relation n'est pas apparente au départ. Pour se faire, des requêtes complexes sont nécessaires, impliquant un nombre important de tables. Le data warehouse doit pouvoir honorer ces requêtes dans des délais raisonnables ;
- Un data warehouse vise à répondre aux besoins des utilisateurs en termes d'informations et non en terme d'applications ;
- **L'imprévisibilité des requêtes:** Dans un contexte décisionnel, les requêtes sont souvent « ad-hoc », générées par l'utilisateur dynamiquement en utilisant des outils, et il est donc impossible d'optimiser chacune de celles-ci au cas par cas ;
- **Modèle intégré et transversal:** le data warehouse permet la mise en place d'un modèle de données intégré et transversale à l'entreprise. Ce modèle se développe de manière incrémentale, au fur et à mesure des réalisations successives des projets décisionnels de l'entreprise.

2.4.2. Les techniques de modélisations et ces concepts :

Les Data Warehouse sont la base et la fondation des systèmes décisionnels. Ces systèmes différents, par leurs vocations, des systèmes OLTP, ont contribué à donner naissance à un modèle de données simplifié et aisément compréhensible, qui est le modèle dimensionnel. Le modèle dimensionnel permet de considérer un sujet d'analyse comme un cube à plusieurs dimensions, offrant des vues en tranches ou des analyses selon différents axes.

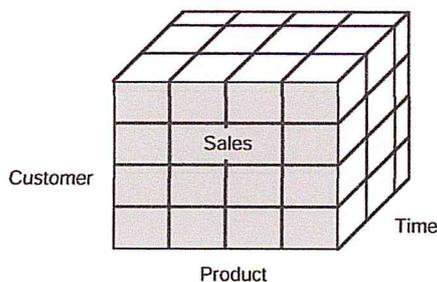


Figure 2.4.1 : Le modèle dimensionnel d'une activité : chaque cellule du cube contient les mesures relatives à une combinaison particulière de produit, client et temps.

La modélisation dimensionnelle, de nature intuitive conjugue simplicité et performance élevée. Le modèle dimensionnel était nommé « schéma des jointures en étoile » à cause de l'aspect rappelant l'étoile du diagramme représentant le modèle dimensionnel. Cette étoile est composée d'une table

centrale « table de fait », et un jeu de petites tables auxiliaires disposées en étoile autour de la table central, qui sont les dimensions. La figure suivante illustre un tel modèle: analyses selon différents axes.

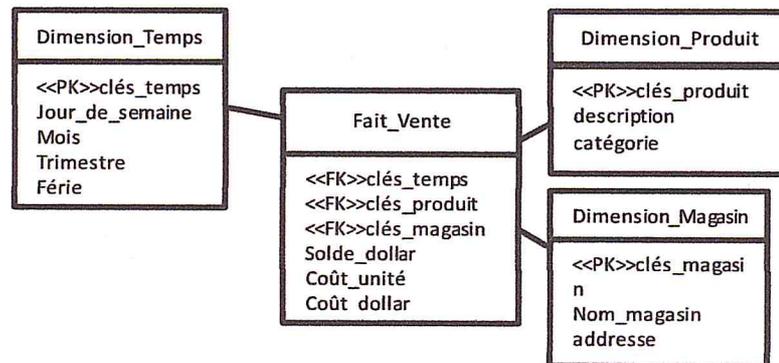


Figure 2.4.2 : Un modèle dimensionnel typique [Kimball, 1996].

2.4.2.1. Concepts liés à la modélisation dimensionnelle:

a. Concept de fait :

Une table de faits est la table contenant les mesures de performances du domaine d'intérêt. Tout le modèle dimensionnel tourne autour de celle-ci. Chaque mesure est représentée par une ligne de la table de faits. Ces mesures se manifestent, généralement, sous forme de valeurs numériques, additives ; elles peuvent être, dans de rares situations, des mesures textuelles. *Une table de faits assure les liens plusieurs à plusieurs entre les dimensions. Elles comportent des clés étrangères, qui ne sont autres que les clés primaires des tables de dimension.*

b. Concept de dimension :

Les tables de dimension sont des tables sous-jacentes aux tables de faits, elles fournissent la description de l'activité sous forme de texte dans une ou plusieurs langues naturelles. A l'instar de toute table relationnelle, une table de dimension est structurée sous forme de plusieurs colonnes décrivant une ligne.

Les tables de dimensions rattachées à une table de fait permettent de fournir la signification aux mesures et de les rendre compréhensibles; elles assurent les analyses en tranches et en dés (Slice & dice). Généralement, une dimension se compose d'une clé artificielle⁶, une clé naturelle⁷ et des attributs. Selon Kimball, « *Une table de dimension établit l'interface homme / entrepôt, elle comporte une clé primaire* » [Kimball, 2002].

c. Tables de fais VS Dimension

Le tableau suivant récapitule les différences au niveau des données de tables faits et dimensions :

⁶ Clé artificielle: dite left de substitution ou surrogate key, elle est une clé générée automatiquement ... voire annex

⁷ Clé naturelle: elle appartient au système OLTP, business, elle permet de ramener des détails depuis l'OLTP à travers des jointures.

	Tables de faits	Tables de dimensions
Structure	Peu de colonnes beaucoup de lignes Elle s'étale en verticale	Peu de lignes beaucoup de colonnes Elle s'étale en horizontale
Données	Mesurable, généralement numérique	Descriptives généralement textuelles
Référentiel	Plusieurs clés étrangères	Une clé primaire
Valeur	Prend de nombreuses valeurs	Plus ou moins constantes
Manipulation	Participe à des calculs	Participe à des contraintes
Signification	Valeurs de mesure	Descriptive
Rôle	Assure les relations entre les dimensions	Assure l'interface homme / data warehouse

Tableau 2.4.1: Tableau comparatif entre les tables de faits et les tables de dimensions.

2.4.2.2. Les techniques de modélisations:

Le modèle décisionnel peut être qualifié grâce aux cinq axes suivants:

1. La clarté et la lisibilité du point de vue de l'utilisateur final;
2. Les performances au chargement;
3. Les performances liées à l'exécution des requêtes;
4. L'administration: la construction du data warehouse n'est pas la difficulté majeure ou la finalité du projet décisionnel, mais c'est de le faire vivre qui pose des problèmes aux entreprises. Il est primordial de tracer les requêtes et d'identifier celles qui sont sollicitées fréquemment, maîtriser et industrialiser tous les processus d'extraction ;
5. L'évolutivité par le développement incrémental du data warehouse, et pas seulement itératif. Un développement itératif pourra, en effet, aboutir à des modules applicatifs indépendants les uns des autres; cependant, le développement incrémental garantit l'intégration inter-modules la mise en œuvre itérative, afin de s'assurer de ce que l'homogénéité globale du système est prise en compte.

Dans ce qui suit, et afin de démontrer les avantages et inconvénients de chacune des approches, normalisation, dé normalisation ou dimensionnel, un modèle sommaire a été avancé qui comporte un cas général dans le monde du commerce.

a. Le modèle de données normalisé:

Considérons, dans un premier temps, le modèle de données normalisé présenté dans la figure suivante. Ce modèle permet, de ventiler les chiffres d'affaires par produit, par client, pour ne citer que peut.

Ce type d'approche est adapté et suivie par les entreprises en matière de modélisation, uniquement si le système à mettre en œuvre permet, à la fois, des sélections et des mises à jour en ligne. Cependant, dans un contexte décisionnel, les mises à jour en ligne ne sont pas d'actualité, ce qui contraint à reconsidérer cette approche.

La sémantique décisionnelle est très faible dans ce modèle. Les informations pertinentes pour l'utilisateur ne sont pas pré calculées et n'existent pas a priori. Les indicateurs devront être recalculés à chaque requête.

Cependant, le modèle est très complet. Il laisse une marge d'autonomie très forte à l'utilisateur.

Un modèle d'entreprise pourrait contenir un nombre très important (centaines ou milliers) d'entités, avec leurs vis-à-vis de tables au niveau physique. Les requêtes seront très complexes et difficiles à formuler par l'utilisateur final, vu qu'elles auront à utiliser des dizaines de tables. Cela affectera au maximum les performances et les temps de réponse.

Pour conclure, ce type d'approche peut avoir ces fruits dans des systèmes décisionnels simples, avec un nombre limités d'utilisateurs effectuant peu de requêtes sur un modèle de données de petite carrure. Autrement, et qui représente la majorité des cas, il faudra recourir à d'autres techniques.

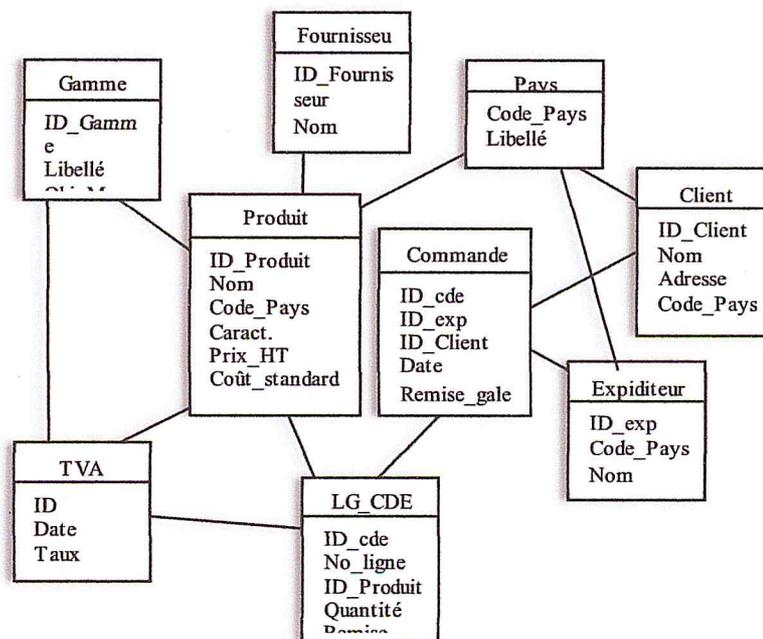


Figure 2.4.3: Le modèle de données normalisé

b. Dé normalisation pour le décisionnel:

Cette approche a pour but l'adaptation du modèle précédent aux besoins liés au décisionnel. Cela passe par le dé normalisation et le pré calcul de certains agrégats, donc à introduire des redondances.

Cette dé normalisation doit être pragmatique, car il n'existe aucune technique formelle pour servir de régulatrices. Cette approche vise à aboutir à un modèle d'une analyse précis des besoins des utilisateurs.

Ainsi, le modèle précédent devient:

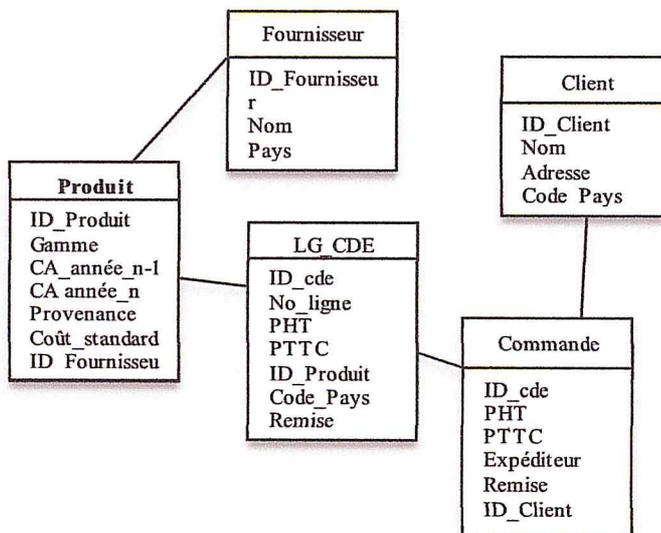


Figure 2.4.5: Le modèle dé normalisé

Le modèle contient un nombre de tables plus restreint associées, chacune, à un sujet d'intérêt. L'orientation « sujet » a pour effet de rapprocher le modèle des besoins des utilisateurs en informations décisionnelles. Le modèle présente un certain nombre d'informations, agrégées, qui sont très fréquemment demandées.

La complexité de ce modèle est réduite d'un facteur de deux (2) comparée au modèle précédent. Cependant, le modèle précédent ne contenait que neuf (9) tables ce qui est peu. Si le modèle normalisé contenait 200 tables, on aurait abouti à une centaine de tables, en appliquant le même facteur, ce qui reste toujours complexe et peu lisible.

La diminution du nombre de tables requiert le nombre de jointures nécessaires pour les requêtes décisionnelles; en revanche, les tables deviennent plus grosses et leurs intégrités difficiles à gérer. Ainsi, les requêtes seront plus simples mais porteront des tables plus volumineuses.

En conclusion, le gain en performance par rapport à un modèle normalisé est très relatif.

c. La modélisation dimensionnelle :

La modélisation dimensionnelle est une approche réservée aux systèmes décisionnels. Elle s'inscrit dans l'objectif phare de ce type de système qui est l'analyse de la ventilation des données quantitatives (les faits) par rapport des données qualitatives (les dimensions).

La modélisation dimensionnelle dérive des concepts des bases de données multidimensionnelles (OLAP), voyant le jour il y a de cela plus de vingt ans (1994). Néanmoins, cette modélisation est indépendante de la technologie, et c'est là la nouveauté, car elle permet l'utilisation de toute base de données : relationnelle, multidimensionnelle, ou objet.

La finalité d'un système décisionnel est l'analyse de la performance, se manifestant à travers des indicateurs dits, justement, « *indicateurs de performance* ». Ces indicateurs n'ont pas de sens propre, mais acquiert les valeurs informationnelles une fois mis en relation avec des dimensions d'analyse.

Les bases de données OLAP permettent l'exploitation rapide et efficace des modèles de type multidimensionnel. Elles fournissent aux utilisateurs des outils assurant la navigation d'une dimension à une autre, de faire des zooms avant (zoom-in) et zoom arrière (zoom-out) sur les informations, etc.

La figure suivante représente un modèle dimensionnel où les indicateurs de base sont groupés dans une table centrale (table de fait). Une table de fait rassemble l'ensemble des indicateurs partageant les mêmes dimensions et qui peuvent être déduits d'autres indicateurs. Ces indicateurs appartiennent au même sujet.

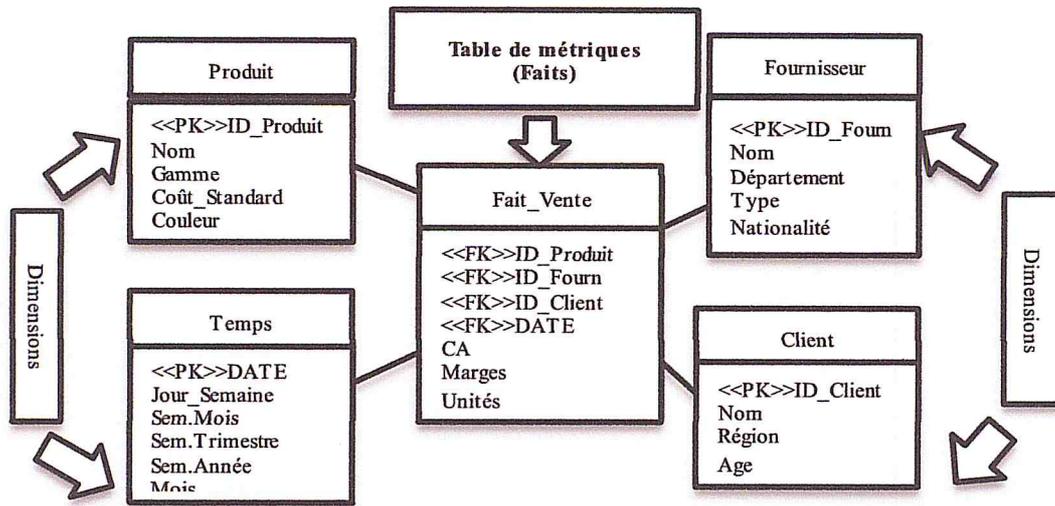


Figure 2.4.6 : Le modèle en étoile

Le schéma illustre un type de modèle nommé, après son aspect, « modèle en étoile ». La table des faits est au centre ayant une clé multiple pour identifiant, conçue par la concaténation des clés de chacune des clés des dimensions d'analyse.

Cette table de faits est entourée par les tables dites « tables de dimension », dont le rôle est de donner du contexte et de la signification aux données quantitatives dans la table des faits.

La force de ce type de modèle réside, essentiellement, dans sa performance et sa lisibilité, même par les non-informaticiens. :

- **La lisibilité** : Ce modèle est facilement interprété par les utilisateurs, même les non-informaticiens. Il est naturellement orienté sujet et définit clairement les indicateurs d'analyse.
- **La performance** : les différents chemins d'accès aux tables pour les requêtes sont prévisibles.

Néanmoins, d'autres techniques de modélisation multidimensionnelles découlant de la modélisation en étoile, existent, notamment la « modélisation en flocon » ou « snowflake ». Le principe du flocon est de garder l'étoile, cependant, ses branches (les dimensions) sont décomposées en sous hiérarchies. Le modèle en flocon revient, donc, à conserver le cœur du modèle en étoile (tables des faits) et d'affiner les branches (dimensions) en les éclatant en sous-tables.

Les avantages du modèle en flocon de neige sont :

- La normalisation des dimensions, réduisant la taille de chacune des tables en éliminant les redondances qui pourraient s'y produire ;
- La formalisation des hiérarchies existant dans une dimension.

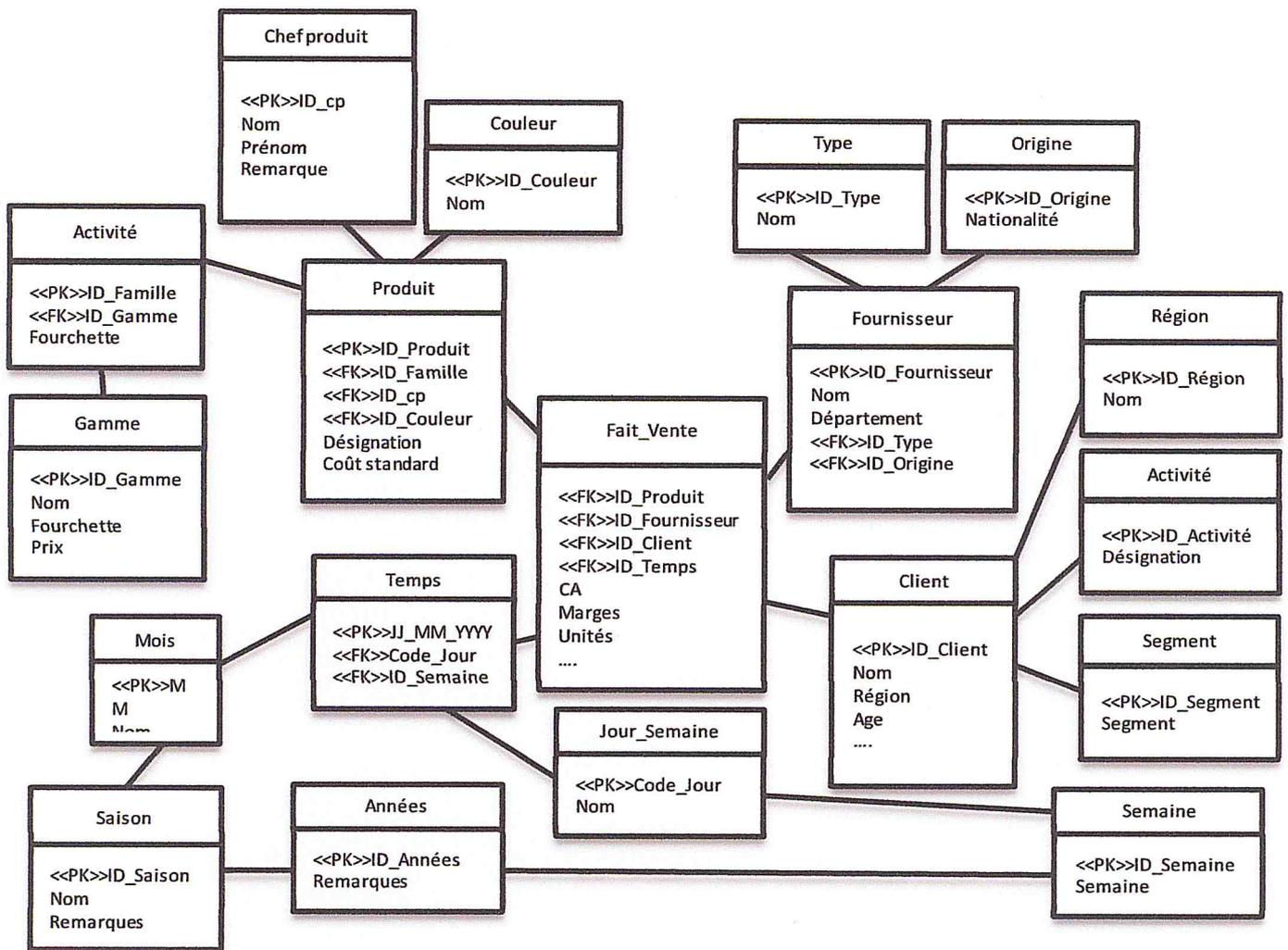


Figure 2.4.7 : Le modèle en flocon de neige

2.5. Démarche de Construction d'un Data Warehouse :

Les démarches, issues des recherches, se réduisent vers les étapes suivantes :

- Modéliser et concevoir le Data Warehouse;
- Alimenter le Data Warehouse;
- Mettre en œuvre le Data Warehouse;
- Administrer et maintenir du Data Warehouse.

Deux approches de conceptions des data warehouses sont les plus renommées :

- L'approche selon les besoins d'analyse (Top-Down);
- L'approche selon les sources de données (Bottom-Up).

C'est Selon le besoin et l'environnement de l'entreprise que l'une des deux approches sera choisie par rapport à l'autre. Une étude doit être faite à priori afin d'opter pour une approche.

Cela dit, le choix de l'approche ne change en rien la définition et la finalité du data warehouse, car il se base sur une architecture multidimensionnelle assurant le support à la décision.

2.5.1. Modélisation et conception du Data Warehouse :

a. Approche « Besoins d'analyse »

Ce sont les besoins de l'utilisateur final qui déterminent le contenu du Data warehouse. Connue, aussi, sous l'appellation « approche descendante » ou « Top-Down Approach », elle est illustrée par son cycle de vie dimensionnelle [source : R. Kimball]

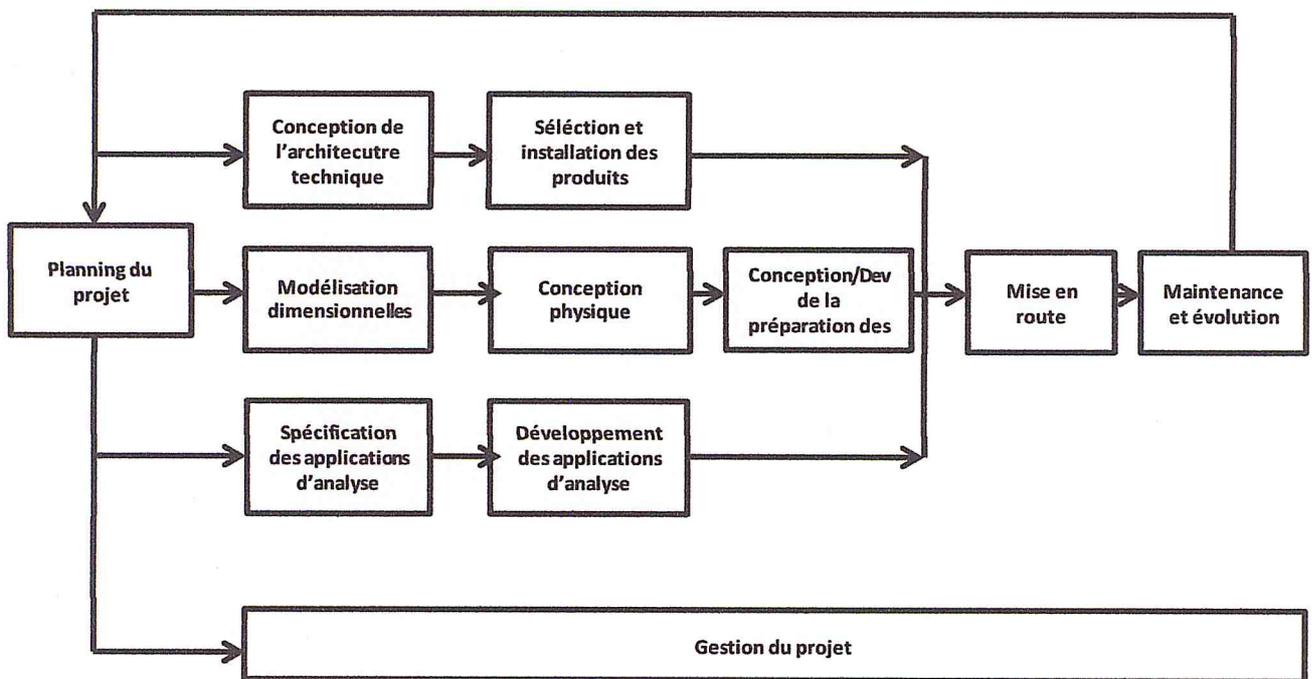


Figure 2.5.1: illustration de l'approche « Besoins d'analyse » grâce au cycle de vie dimensionnel de Kimball [Source : Kimball, 2004].

b. Approche « Source de données » :

Contrairement à l'approche « descendante », cette approche est appelée « approche ascendante » ou « Bottom-up ». Dans cette approche, ce sont les sources de données existantes qui déterminent le contenu du data warehouse.

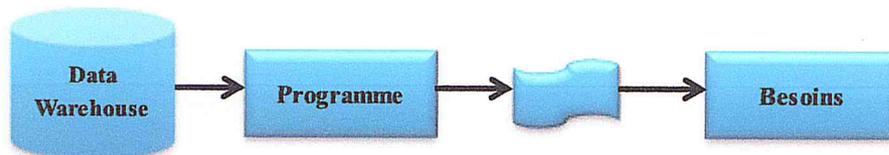


Figure 2.5.2: Illustration de l'approche « Source de données » grâce au cycle de développement du DW de Inmon [Source : Inmon, 2002].

Selon Inmon, les besoins dont en évolution permanente et donc impossible au utilisateur de les déterminer au départ « *Donnez-moi ce que je vous demande, et je vous direz ce dont j'ai vraiment besoin* »⁸, [Source : Inmon, 2002].

c. Approche mixte (Hybride) :

Comme est souvent le cas, un compromis est toujours convoité, d'où la naissance de l'approche hybride ou mixte combinant le « Top-Down » et le « Bottom-up ». L'approche hybride prend en considération et les sources de données et les besoins des utilisateurs.

Elle consiste à construire des schémas dimensionnels en se basant sur les sources de données du système opérationnel, et les valider par rapport aux besoins analytiques des utilisateurs finaux.

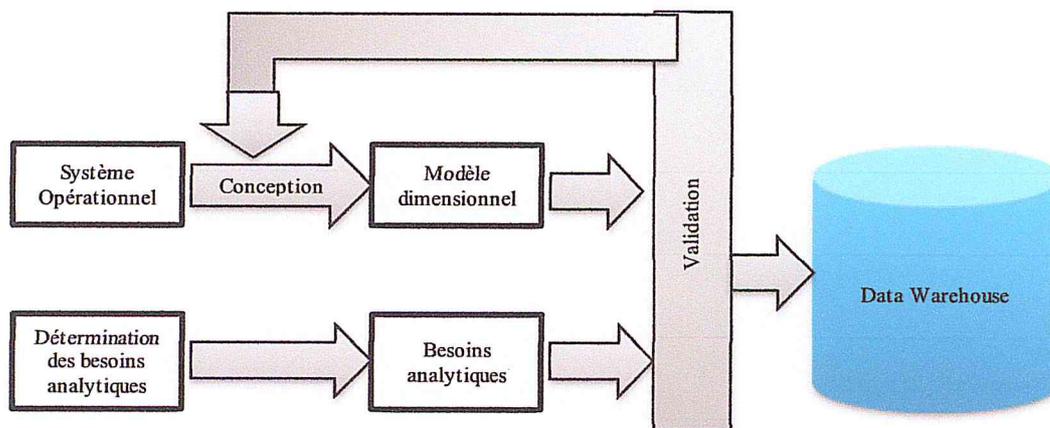


Figure 2.5.3: Illustration de l'approche mixte.

Conclusion :

Le livrable de cette étape consistera en un modèle dimensionnel validé du Data Warehouse. Par la suite, ce modèle dimensionnel sera transformé en modèle physique, qui différera du modèle dimensionnel.

2.5.2. Alimentation du Data Warehouse :

L'étape suivant la conception du data warehouse est son alimentation est chargement en données. Cette alimentation comprend trois étapes, connues sous l'acronyme anglais : ETL pour « Extract-Transform-Load »:

⁸ Give me what I tell you I want, then I can tell you what I really want.”[Inmon, 2002]

- Extraction (Extract) des données sources provenant des systèmes opérationnels de production,
- Transformation (Transform) des données selon le besoin décisionnel et qualité des données sources,
- Le chargement (Load) des données traitées dans le data warehouse.

Les étapes, citées supra, représentent des mécaniques cycliques garantissant l'alimentation du Data Warehouse en données fiables pour assurer la crédibilité du système décisionnel.

2.5.2.1. Les phases de l'alimentation « E.T.L. » :

Les outils, « Extraction-Transformation-Chargement » (E.T.C) ou en anglais E.T.L.), sont des outils facilitant le déroulement des trois phases citées précédemment. Ils représentent une importance accrue dans un projet Data Warehouse.

Comme son nom l'indique, le processus E.T.L. représentent la mécanique d'alimentation du Data Warehouse en trois phases :

a. L'extraction des données

Il paraît évident que l'extraction des données devance toutes les étapes. « *L'extraction est la première étape du processus d'apport de données à l'entrepôt de données. Extraire, cela veut dire lire et interpréter les données sources et les copier dans la zone de préparation en vue de manipulations ultérieures.* » [Source : Kimball, 2005].

Elle se résume en:

- Cibler les données: selon le besoin décisionnel, définir les données qui vont être puisées depuis les systèmes opérationnels ;
- Appliquer les filtres nécessaires: afin de ramener uniquement la vue en données dont le système décisionnel aura besoin ;
- Définir la fréquence de chargement : c'est cette fréquence qui définit quand le système data warehouse devra se connecter aux systèmes opérationnels afin de puiser de nouvelles données ou mettre à jour, en batch, les données modifiées dans les systèmes opérationnels.

Le processus de chargement des données est répétitif, il faudra faire en sorte de ne charger que les données nouvelles ou celles qui ont été mises à jours depuis le dernier chargement. Pour cela, il existe trois stratégies de capture de changement :

- **Stratégie de la colonne d'audit** : c'est une colonne mise à jour soit par des applications soit par des triggers dans la base de données. Elle persiste la date d'insertion ou du dernier changement d'un enregistrement.
- **Stratégie de capture des logs** : il existe des outils ETL exploitant les fichiers logs des systèmes sources pour déceler les changements. En revanche, cette stratégie n'est pas toujours évidente car absente dans certains outils ETL du marché, d'autant plus qu'elle

présente l'inconvénient de la perte des fichiers logs, généralement lors des épurations des systèmes sources.

- **Stratégie de la comparaison avec le dernier chargement : une sauvegarde de copie de l'extraction est effectuée pour servir de référence de comparaison lors de la prochaine extraction.** Il est impossible de rater un nouvel enregistrement avec cette stratégie.

Le processus d'extraction se fait selon différentes manières. C'est selon la disponibilité et l'accessibilité aux sources de données que la politique d'extraction est définie. Il en existe trois⁹ :

b. La transformation des données

La transformation vient en second lieu dans le processus E.T.L., et représente la phase la plus importante. Elle a pour but de garantir la fiabilité des données et leurs qualités. Elle passe par les tâches suivantes :

- **Traiter les données manquantes** : qui sont souvent « nul » et ce par leurs affecter une valeur soit auprès des utilisateurs des systèmes opérationnels, soit en utilisant des données probables selon des statistiques liées à l'environnement de l'entreprises : par exemple la valeur de la donnée météo pourra être mise à défaut à « Ensoleillé », vue que le climat méditerranéen en Algérie ;
- Rectification et correction des données,
- Elimination de toute ambiguïté.
- Elimination des données redondantes.
- Consolidation des données.

La transformation permet d'aboutir à des informations stables, crédibles et de qualité prêtes à être entreposées dans un data warehouse.

c. Le chargement des données

Le chargement représente la dernière phase de l'alimentation d'un data warehouse. Il est indispensable et délicat en même temps, car il demande une certaine connaissance des structures du système de gestion de la base de données (tables et index) pour son optimisation. En effet, le chargement traite un ensemble colossal de données qui doivent être agrégées et cela mène à un traitement intensif nécessitant une optimisation.

2.5.2.2. Politiques de l'alimentation :

- **Politique du Push** : La logique de chargement des données vers la zone de préparation (Staging area) se trouve dans le système de production. Il « pousse » les données concernées vers la zone de préparation quand il en a l'occasion. Cette politique représente un inconvénient majeure : si le système est occupé, les données ne seront jamais envoyées vers le staging area ;

⁹ <http://grim.developpez.com/articles/concepts/etl/>

- **Politique du Pull** : à l'opposé de la politique du « Push », le « Pull » soutire les données des systèmes sources vers la zone de préparation. L'inconvénient majeur de cette politique est que ça peut surcharger le système opérationnel qui est en cours d'utilisation par un nombre important d'utilisateurs effectuant des transactions en ligne ;
- **Politique hybride Push-pull** : C'est le compromis des deux politiques « Push » et « Pull ». cette politique élimine les inconvénients du « push » par le fait que les systèmes opérationnels préparent les données à extraire et notifient la zone de préparation, qui elle, effectuera le « Pull » afin de récupérer ces données.

En conclusion, le processus d'alimentation doit être :

- **Sûr** : garantie l'arrivée des données depuis les sources de données vers le data warehouse dans un format assurant la qualité.
- **Rapide** : il se doit de manipuler de quantités colossales de données pouvant induire à des lenteurs considérables. Le processus d'alimentation devra être traité ces masses de données dans acceptables.
- **Correctif et transparent**: le processus d'alimentation se doit, aussi, de procéder aux corrections nécessaires et combler les manques constatés dans les sources de données dans la finalité d'améliorer la qualité des données. Cela se fait d'une façon transparente à l'utilisateur du système décisionnel.

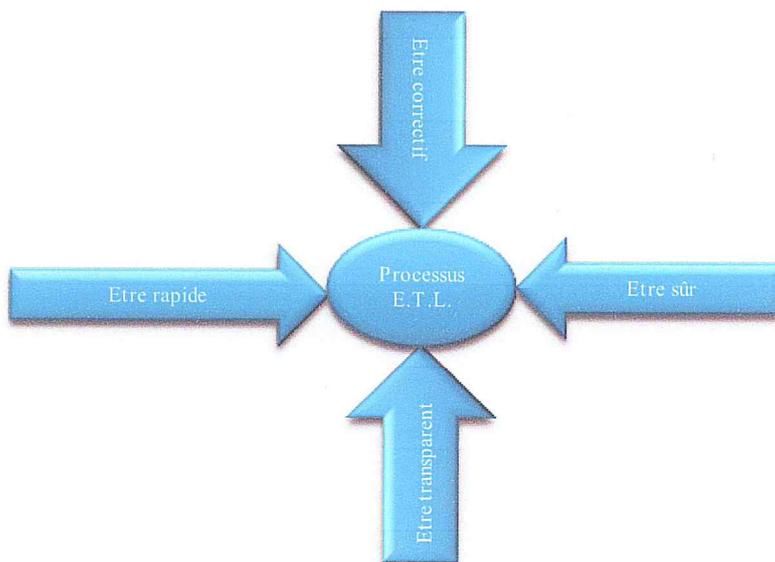


Figure 2.5.4 : Objectif de qualité de données dans un processus ETL [Source : Kimball, 2004].

2.5.3. Gestion du temps dans le data warehouse¹⁰ :

Le système décisionnels consolide les données historique des systèmes opérationnels afin de les analyser et d'en sortir des indicateurs d'aide à la décision. Cela rend les données temporelles

¹⁰ le terme « Temps » sera utilisé, dans ce mémoire, dans le sens de données « Date/Heure ».

primordiales dans le processus de conception des data warehouses. C'est, là, la différence majeure entre les systèmes décisionnels et opérationnels. Dans ce qui suit, une introduction vers les caractéristiques du « Temps » et son utilisation dans les data warehouses sera avancée.

Les applications et solution software des systèmes opérationnels évoluent dans le présent, pour automatiser les tâches quotidiennes. Cela explique le fait que les données temporelles ne nécessitent pas un traitement particulier. Elles seront représentées, dans la majorité des systèmes opérationnels, par de simples attributs « Date/Heures ».

Cependant, dans un data warehouse, le « Temps » influe énormément la structure du système. Les caractéristiques et la finalité des systèmes décisionnels font apparaître des besoins en données temporelles n'existant pas dans les systèmes opérationnels.

Les SGBD présents sur le marché ne supportent pas les données temporelles d'une façon implicite, ce qui incombe aux bases de données data warehouse de prendre en considération les besoins temporels et les implémenter sur des SGBDR classiques. Ce qui mène à la nécessité de définir les données sur le « Temps » d'une manière explicite et de les intégrer dans les structures des tables et des requêtes.

2.5.3.1. Le rôle des données temporelles :

Les données temporelles sont ajoutées, dans un data warehouse, dans un but principal d'historiser (archiver) les données, ce qui permet de découvrir l'aspect d'une entreprise à n'importe quel moment ou période dans le temps., permettant, ainsi, de faire ressortir des habitudes comportementales dans le temps et de faire des comparaisons entre des périodes similaires ou non-similaires.

2.5.3.2. Les problématiques liées au temps :

Lors de la conception du data warehouse, il faudra résoudre les problématiques liées aux données temporelles :

- **Identifier et capturer les besoins temporels :** vu qu'il n'existe pas de méthodes pour identifier et capturer le besoin en données temporelles ;
- **Capturer les changements dimensionnels :** ce sont des changements qui touchent des données déjà présentes et consolidées dans le data warehouse et qui affectent le contenu des données agrégées.
- **La fréquence de captures :** quand et combien de fois les captures des changements se feront elles pour maintenir le data warehouse à jours sans pénaliser les systèmes opérationnels.
- **La synchronisation des changements :** dans un souci de maintenir la fiabilité des données dans le data warehouse et de garantir leur crédibilité, il est primordial de traiter le volet de synchronisation des données liées en cascade.

2.6.Conclusion :

Ce chapitre a été consacré à l'étude de la littérature touchant à tous ce qui est important dans le domaine du décisionnel et du data warehousing.

Il a été question d'introduire les systèmes d'information décisionnels comme moyen d'offrir un vue transversale de l'entreprise et on a situé le décisionnel dans la hiérarchie de l'entreprise et on a introduit ses différent composants.

Ensuite, il était nécessaire de se focaliser sur le système data warhouse, le vif de ce mémoire, depuis sa définition et son historique jusqu'aux techniques de modélisations, puis le document a traité la démarche de construction d'un data warehouse.

Dans ce chapitre, les différentes définitions et mots clés utilisés tout au long de ce mémoire ont été introduit essentiellement les concepts des faits, dimensions, OLAP, ROLAP, et la modélisation en étoile.

Le chapitre suivant, chapitre troisième, sera consacré pour l'étude des besoins de la Gendarmerie Nationale en ce qui concerne les analyses des accidents routiers et les infractions induisant aux accidents.

Chapitre Troisième

Etude des besoins



« Le besoin apprend à prier »

1. Introduction.
2. Description de la démarche d'étude des besoins (Interview).
3. Autres moyens utilisés pour la détection des besoins.
4. Etudes des applications.
5. Rédaction et validation du recueil récapitulatif des besoins.
6. Problèmes et obstacles rencontrés.
7. Conclusion.

3. Etude des besoins :

3.1. Introduction :

L'étude des besoins est une étape primordiale et décisive dans la vie d'un data warehouse. Elle décèle les vrais besoins des utilisateurs en analyse. Ces besoins vont, par la suite, diriger tout le processus de conception et d'implémentation du projet décisionnel. Un produit rejeté par les utilisateurs est le cauchemar de tout développeur et financier. Cette étape est à prendre avec le plus grand intérêt et sérieux.

Cette étude suit deux grands pôles dans le monde du data warehousing : l'approche « **Bottom Up** » et l'approche « **Top Down** ». Cependant, et comme il a été mentionné dans le « **chapitre deuxième** » de ce document, une approche hybride des deux est en mesure de répondre aux mieux aux attentes afin de prendre en considération les besoins des décideurs sans négliger la vue systémique des données sources, et les possibilités analytiques offertes par celles-ci. L'approche « **mixte** » permet de recenser les besoins des décideurs et de déceler, éventuellement, des besoins nouveaux (anticiper les besoins futures) se basant sur les données sources.

L'étude des besoins doit impliquer toutes les parties prenantes et ne doit pas se limiter aux utilisateurs directs du système (les décideurs). Il est impératif d'inclure d'autres parties telles que les utilisateurs producteurs de données dans les systèmes de production ou les DBA¹¹.

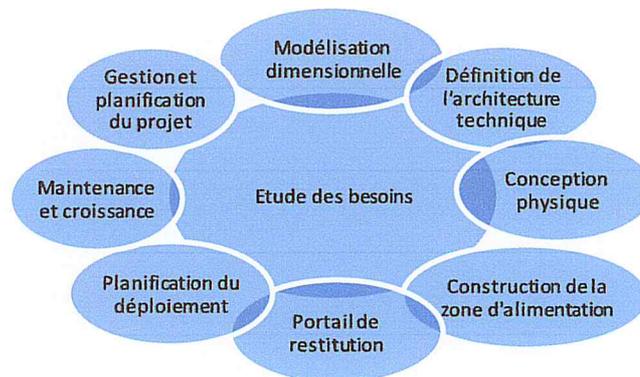


Figure 3.1.1 : La place de l'étape d'étude des besoins dans un projet Data Warehouse.

Dans ce qui suit, il va être question de mettre la lumière sur la démarche adoptée pour recenser et arrêter les besoins des utilisateurs.

3.2. Description de la démarche d'étude des besoins (Interview) :

Dans l'optique de cerner la totalité des besoins, et d'anticiper sur d'autres, tout en suivant une approche mixte entre le top-down et le Bottom-up, il était judicieux d'étudier les systèmes sources et de maintenir des entretiens avec les DBA et développeurs des applications.

Les étapes suivantes ont été tracées afin de cerner les besoins :

- Localisation et détermination des postes et personnes détenant l'information de l'étendu du projet ;

¹¹ « Les DBA sont les principaux experts sur les applications existantes susceptibles d'alimenter l'entrepôt de données. Leurs interviews servent à confronter aux réalités certains des thèmes qui surgissent lors des rencontres avec les utilisateurs finaux. » [Kimball, 96]

- Étude préliminaire des systèmes sources et interviews des DBA et développeurs ;
- Conduite des interviews ;
- Arrêt et validation des besoins avec les utilisateurs.

3.2.1. Localisation et détermination des postes et personnes détenant l'information de l'étendu du projet :

La Gendarmerie Nationale algérienne, est une institution militaire couvrant tous le territoire algérien, s'étalant sur une superficie colossale¹² la mettant au premier rang dans le continent africain. Cette couverture de grande envergure justifie le grand nombre d'éléments de l'arme formant son personnel, chose qui a rendu impossible de s'entretenir avec tous les utilisateurs appartenant aux différents échelons.

Cependant, le travail quotidien est fortement organisé et régit par des procédures suivies rigoureusement. Cette discipline militaire a permis de s'entretenir avec des représentants de chaque type d'unités.

Les services, postes de travail et unités susceptibles de détenteurs des informations portant une plus-value pour le système décisionnel sont listés dans le tableau suivant ¹³:

Niveau Hiérarchique	Structure	Intitulé du poste	Nombre total des postes	Nombre de postes pris comme échantillon
Commandement de la Gendarmerie Nationale	Division Centrale de la Sécurité Routière	Chef Bureau Central Sécurité Routière	01	01
	Service Central de l'Informatique	Chef Bureau Central de Développement	01	01
Commandements Régionaux de la Gendarmerie Nationales	Service Régional de la Sécurité Routière	Chef Bureau Régional Sécurité Routière	06	01
	Bureau de Sécurité Routière (BSR)	Chef Bureau de la Sécurité Routière	48	01
	Compagnie Territoriale	Chef Compagnie	X	03
	Escadron de la Sécurité Routière	Commandant d'Escadron	X	02
	Brigade de la Sécurité Routière	Chef de Brigade de Sécurité Routière	X	02
Les Groupement Territoriaux	Brigade Territoriale	Chef de Brigade Territoriale	X	03

Tableau 3.2.1 : Tableau présentant la population à interviewer.

¹² Une superficie de 2,382 millions km²

¹³ Les éléments listés dans le tableau ne sont pas forcément les utilisateurs du système décisionnel, mais les entités porteuses d'informations judcieuses.

3.2.2. Interviews des DBA et développeurs :

Cette étape est d'une importance inouïe pour un recensement exhaustif des besoins à travers les sources de données existantes. A la fin de cette étude, une idée claire et concise sera établie sur les systèmes opérationnels et leurs environnements dans lesquels ils évoluent.

Les DBAs et les développeurs du service central de l'informatique ont été de grande aide, en parallèle avec le chef de bureau central de la sécurité routière. Mais aussi, le chef de bureau de la sécurité routière du premier commandement régional de Bida et des groupements territoriaux.

Cependant, il a été question de solliciter l'aide des unités territoriales : Escadron Sécurité Routière, Compagnies territoriales, Brigades de sécurité routières ainsi que les brigades territoriales, qui représentent la base de la pyramide mais qui sont les générateurs de données, car c'est eux qui vont saisir les données des accidents et des infractions lors de l'exercice de leurs tâches quotidiennes.

Cette vaste variété d'intervenants ayant différentes responsabilités et appartenant à plusieurs échelons a permis de traiter les besoins selon différents angles de vue.

3.2.3. Planification, préparation et conduite des interviews :

Le contact humain, a bien été, parmi les meilleurs moyens de soutirer de l'information de leurs sources depuis la nuit des temps. C'est dans cette optique que les entretiens avec les personnes porteuses des informations s'est imposée dans ce mémoire. Néanmoins, pour être efficace, une interview doit être préparée en passant par :

3.2.3.1. La planification :

Elle passe par la prise des rendez-vous avec les futures interviewés et la réservation du lieu de de l'entretien.

3.2.3.2. La préparation :

La préparation suit la planification. Elle consiste en la préparation des questions à poser dans un jargon utilisé et compris par les utilisateurs, afin de le rendre plus à l'aise ; mais aussi, les sujets à éviter pour ne pas trop s'égarer du vif du sujet. Les mêmes questions peuvent être formulées différemment selon la personne interviewée.

3.2.3.3. La conduite promptement dite :

La conduite des entretiens est la concrétisation des deux étapes précédentes dans le but de soutirer les informations des interlocuteurs via les questions préparées. Leurs réponses vont décrire les procédures de travaux. De là, découlera la définition des besoins par l'extraction des mesures et métriques utilisés dans la réalisation des procédures, ou produites suite à l'accomplissement de ces procédures.

3.3. Autres moyens utilisés pour la détection des besoins :

Les entretiens sont efficaces pour l'identification de la majorité des besoins ; cependant, les interlocuteurs, peuvent omettre certains besoins soit parce qu'ils pensent qu'ils sont évidents, par oubli ou juste par ignorance de l'importance du besoin.

L'étude des états de sortie ou les rapports représente un trésor riche en données et leurs formats. Il est vivement conseillé de ne pas se tenir uniquement aux interviews, mais d'étudier au préalable les états de sorties surtout dans la phase de préparation.

Par ailleurs, l'étude des données ouvre des perspectives futures, non ressentit dans l'immédiat par les utilisateurs, elle permet de concevoir un data warehouse évolutif (scalable) pouvant répondre aux nouveaux besoins éventuels.

3.4. Étude préliminaire des systèmes sources et des applications

3.4.1. Les applications de la Division Sécurité Routière :

Les services de sécurités routières ont recours à la mise en place de base de données et applications informatiques de sécurité routières, dans un souci d'organisation, et de sauvegarde des informations de plus en plus volumineuses, relatives à l'activité quotidienne des unités territoriales en matière de sécurité routière.

Ces applications constituent une source énorme d'informations, qui peut être exploitée à des fins d'analyse décisionnelles et de reporting présentées sous formes graphique ou même cartographique, afin d'orienter les décisions et les actions de la police de sécurité routière selon des recoupements géographiques, les phénomènes liés à la circulation, le parc roulant mais aussi aux piétons.

Il sera question, dans ce chapitre, de présenter et de décrire les modules de l'application Records Management System (RMS). Les modules traités se rapportent aux activités de la sécurité routière.

La division de Sécurité Routière de la Gendarmerie Nationale est l'entité qui supervise et gère le métier de la sécurité routière et police de la route, elle est en charge notamment de centraliser et de rassembler toutes les données des affaires liées à la sécurité routières sur le territoire national, de procéder aux analyses nécessaires, d'apporter des directives et des orientations aux unités territoriales en charge d'assurer la sécurité sur les routes et de présenter périodiquement des rapports et des statistiques au haut commandement de la Gendarmerie Nationale et également aux autorités publiques quant aux phénomènes routiers, déclinés selon plusieurs axes d'analyses (tranche d'âge, type d'infraction, lieu ...etc.).

Les décideurs exploitent les modules du RMS liés à la sécurité routières pour effectuer leurs missions décisionnelles tant bien que mal.

3.4.2. Le Records Management System (RMS) :

Le RMS est une application Web, développées avec JDeveloper 11g Release, interagissant avec une base de données Oracle 11g Release 2. Il est organisé autour de plusieurs modules, dont les modules accidents et contraventions, pour la saisie des activités opérationnelles des unités de la Gendarmerie Nationale. Cette saisie a pour but de suivre les activités et performances des unités de la Gendarmerie nationale ainsi que l'archivage des données pour une exploitation future pour des raisons statistiques ou la contribution dans les investigations criminelles.

3.4.2.1. Module Accident de la route :

Le module Accident de la route est principalement dédié à la saisie des données des accidents de la route par les unités territoriales de la Gendarmerie Nationale : Escadrons Sécurité Routière (E.S.R.), Compagnies Territoriales (C.T.), Brigades de la Sécurité Routière (B.S.R.) et Brigades Territoriales (B.T.). Ces données seront exploitées par ces unités ainsi que les échelons supérieurs Groupements Territoriaux (C.T.), Commandements Régionaux de la Gendarmerie Nationale (C.R.G.N.) ainsi que le Commandement de la Gendarmerie Nationale (C.G.N.) à travers la Division de la Sécurité Routière (D.S.R.), principalement.

Le module Accident de la route, à l'instar de tous les modules du RMS, se compose d'une entête contenant les données principalement liées à l'accident et des tabulations contenant des liens vers d'autres modules tels que :

- Les véhicules ;
- Les personnes, sous forme de chauffeurs, victimes ou passagers ;
- Les contraventions (infractions) induisant à l'accident.

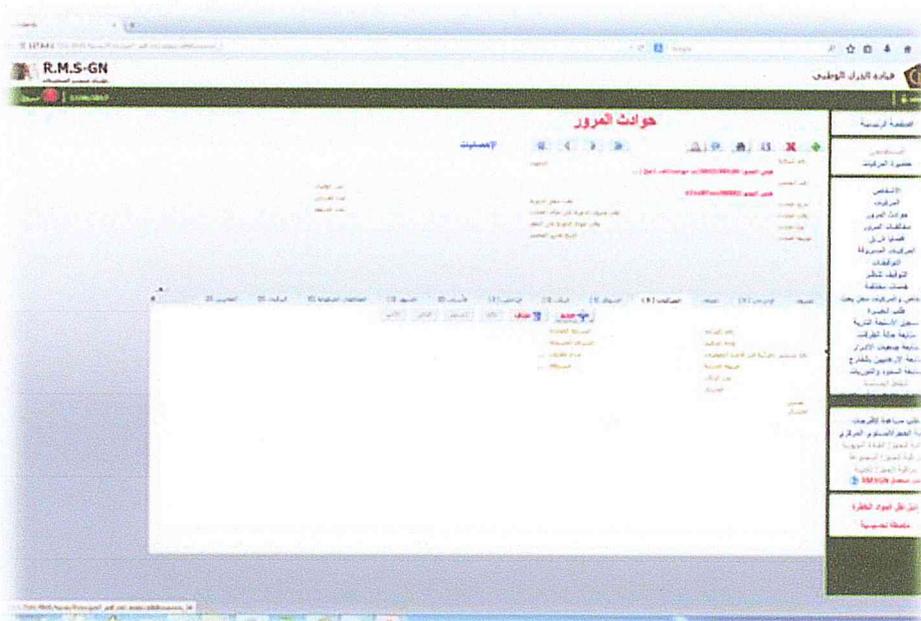


Figure 3.4.1 : Prise d'écran du module Accident de la Route du RMS

3.4.2.2. Module contravention :

Le module Contraventions est exploité par les unités territoriales de la Gendarmerie Nationale : ESR, CT, BT et BSR. Ces données seront exploitées par ces unités ainsi que les échelons supérieurs GT, CRGN ainsi que le CGN à travers le DSR.

Le module Contravention suit la philosophie de la conception du RMS, qui présente les données du module en question dans son entête ainsi que les modules lui étant liés dans des tabulations dans la partie inférieures tels que :

- Les véhicules ;

- Les accidents de la route ;
- Les personnes sous forme de chauffeurs, victimes ou passagers.

Bien que ce module contienne les données de toutes les contraventions liées au code de la route, on s'intéresse ici à celles qui ont induit (causé) à des accidents. Autrement dit, les infractions qui ont fait l'objet d'une contravention et ont induit à un accident de la route.

3.4.3. Les états de sorties périodiques de la DSR :

Plusieurs responsabilités incombent à la DSR, elles ont toutes pour finalité d'assurer la sécurité des citoyens sur le réseau routier. Des citoyens qui peuvent être des chauffeurs, mais surtout des passagers ou des piétons. La sécurité routière touche aussi les biens publics et privés qui peuvent être endommagés suite à des accidents routiers.

Pour cela, la DSR se doit d'exploiter les données saisies par les unités de la GN depuis le terrain via le RMS et de les analyser pour générer des rapports susceptibles de servir de support afin de générer des directives et de corriger la stratégie et la politique à suivre par les unités sur le terrain.

Donc, parmi les missions de la DSR on cite la génération des rapports périodiques sur l'état de la sécurité routière et la prise de décision en matière de distribution rationnelle de l'effectif sur le terrain au moment opportun;

Après constatations, il ressort que la DSR édite périodiquement les états suivants :

- Gravité des accidents par moi, jour, heure et réseau routier ;
- Répartition des accidents par Wilaya ;
- Répartition des accidents par moi, jour et heure ;
- Répartition des accidents matériels, mortels et corporels par moi, jour et heure ;
- Répartition des accidents par causes sur un intervalle de date donnée ;
- Accident routier par type de véhicules répartis sur les mois, jours et heures ;
- Les victimes liées aux accidents de la route par wilaya ;
- Les chauffeurs impliqués dans les accidents de la route par wilaya ;
- Répartition des décès, blessés par mois, jours et heure.

3.5. Rédaction et validation du recueil récapitulatif des besoins

Après les entretiens et la collecte des besoins, et aussi après l'application de l'approche Bottom-up et la consultation du potentiel que peut offrir les sources de données présentes, nous avons arrêté les besoins suivants :

Chapitre Troisième : Etude des Besoins

Besoins	Finalités escomptée par le besoin
Gravité des accidents rapport à l'axe temporel (heure, jour, mois ...)	<ul style="list-style-type: none"> Avoir une vision selon différents critères sur les accidents selon leurs gravités. Cette vision permet de définir le niveau d'urgence et d'alertes qui pourra susciter de déclencher certaines mesures drastiques pour maîtriser le fléau des accidents
Gravité des accidents par réseau routier	<ul style="list-style-type: none"> Cette analyse a pour but d'avoir une appréciation sur la répartition des accidents sur les différents types de routes. La finalité de ce rapport et de contribuer dans la prise de décision sur la répartition de l'effectifs et les ressources des unités territoriales sur le axes routiers pour une gestion rationnelles et de bien répartir les ressources selon les points noirs.
Répartition des accidents par réseau routiers	
Répartition des accidents par Wilaya	<ul style="list-style-type: none"> Un besoin statistique pour l'appréciation et l'évaluation des groupements territoriaux en matière de sécurité routière et maîtrise des accidents de la route La contribution dans le processus de sensibilisation de la communauté civile, qui joue un rôle majeur et primordial dans la maîtrise de cette hécatombe causée par les accidents de la route
Répartition des accidents par rapport à l'axe temporel (mois, jour et heure)	<ul style="list-style-type: none"> Appel à vigilance aux citoyens afin de prendre leurs précautions durant les heures, jours et mois enregistrant le plus grand nombre d'accidents par l'envoi de SMS, émission radio ou télévisée ; La répartition du bon nombre de ressources selon les nombres enregistrés dans les différents intervalles de temps. Doubler d'effectifs durant les périodes enregistrant un nombre important d'accidents, et à l'inverse diminuer le nombre d'effectif sur terrains durant les moments paisibles. Cela entre dans la rationalisation de l'effectif.
Répartition des accidents par causes sur un intervalle de dates données	<ul style="list-style-type: none"> Cette analyse est primordiale. Elle donne une idée claire sur les causes réelles induisant vers l'accident. Souvent, l'être humain est le responsable mais ce n'est pas toujours le conducteur, le piéton peut être la cause ou même l'occupant. Cette analyse a permis de placer des ralentisseurs dans certains axes routiers comme les enlever dans d'autres.
Nombre de chauffeurs par niveau d'éducation sur une période donnée	<ul style="list-style-type: none"> Les conducteurs représentent le maillon faible du processus de sécurisation des routes. La Gendarmerie Nationale doit agir sur cette catégorie de personnes afin d'encapsuler le fléau des accidents. L'étude des conducteurs selon le sexe, tranches d'âge et niveau d'éducation fera ressortir les sous-catégories de chauffeurs impliqués dans le plus grand nombre d'accident dans un but de focaliser les efforts de
Nombre de chauffeurs par tranche d'âge sur une période donnée	
Nombre de chauffeurs par sexe sur une période	

Chapitre Troisième : Etude des Besoins

donnée	sensibilisations, formations ou sanctions.
	<ul style="list-style-type: none"> • Cette analyse permet aussi de faire ressortir les tranches d'âge responsables du plus grand nombre d'accidents, telles les plus jeunes à cause du manque d'expérience ou celle des plus vieux à cause de problèmes de santé et de visions.
Nombre de victimes par tranches d'âge sur une période donnée	<ul style="list-style-type: none"> • Les victimes peuvent être des chauffeurs, passagers ou piétons. Cette analyse fait ressortir des statistiques sur les victimes à utiliser dans les campagnes de sensibilisations ou par les ministères concernées telle que le Ministère de la santé ;
Nombre de victimes par sexe sur une période donnée	<ul style="list-style-type: none"> • Fournir l'information utile, pour certains organismes et services, tels que les hôpitaux et la protection civile. En effet, selon les chiffres communiqués, ces institutions peuvent se mobiliser en temps opportun ;
Répartition des décès, blessés par mois, jours et heure	<ul style="list-style-type: none"> • Cette analyse permet aussi de voir si des tranches d'âge sont touchées plus que d'autres ou pas
Répartition des accidents par type de véhicule	Déterminer la relation entre le type de véhicule et le nombre d'accidents afin d'attirer l'attention du gouvernement avec des alertes
Découpage géographique de la carte de l'Algérie, niveaux Wilayas et Communes, selon différents critères :	Avoir les statistiques d'une façon visuelle afin de mieux les apprécier et de découvrir de nouvelles interprétations et analyses liées à la situation géographique.
	<ul style="list-style-type: none"> • Nombre d'accidents ; • Nombre de décès ; • Nombre de blessés.

Tableau 3.5.1 : Synthétisation des besoins recensés.

Toutes ces analyses sont aussi présentes ventilées sur les trois dernières années. Cette comparaison permettra d'avoir un écho (feed back) pour mesurer l'efficacité des mesures et décisions prises dans le domaine de la sécurité routière.

3.6.Problèmes et obstacles rencontrés :

La démarche adoptée pour la collecte des besoins a permis d'aboutir à un ensemble satisfaisant des attentes des utilisateurs. Cependant, puisque la réalité ne vient pas sans obstacles, il est nécessaire de citer celles qui ont surgies durant cette étape:

- La difficulté d'accès aux données à cause des soucis de confidentialité ;
- Le scepticisme de certaines personnes quant à la réalisation du projet essentiellement à cause de la résistance aux changements, qui est un phénomène connu et qui doit être pris en charge ;

- La difficulté de planification et d'obtention des entretiens ;
- La difficulté de préparation des bonnes questions qui attire les interlocuteurs et les incite à partager leurs connaissances.

3.7. Conclusion

Cette étape a permis d'aboutir à une liste précise des besoins décisionnels de la Gendarmerie Nationale en matière d'accidents de la route et les infractions induisant à ces accidents.

Cette étape a nécessité un temps long et précieux s'étalant sur environ quarante heures et impliquant Trente-trois personnes appartenant à différents niveaux hiérarchiques de la Gendarmerie Nationale. Elle est considérée comme un facteur clé de succès, car elle déterminera l'adoption du projet par les utilisateurs ou son rejet, à travers la détermination de leurs besoins et attentes réels. Elle permettra, aussi, de diriger la construction data warehouse, qui est, justement, le sujet du Chapitre Quatrième.

Chapitre Quatrième

Conception du Data Warehouse

*« Lorsqu'on s'occupe d'informatique il faut faire comme les canards..
Paraître calme en surface et pédaler comme un forcené par en dessous. »*

Richard Lallement

1. Conception de la zone d'entreposage des données.
2. Conception de la zone d'alimentation.
3. Conception des cubes dimensionnels.

4. Conception du Data Warehouse :

4.1. Conception de la zone « entreposage des données » :

4.1.1. Introduction :

Une fois les besoins des utilisateurs connus, la conception des volets du projet data warehouse peut être entamé. Pour cela, nous avons eu recours à la modélisation dimensionnelle qui est souvent associée aux entrepôts de données compte tenu de ses avantages.

Cependant, avant de se lancer dans la modélisation, il est intéressant de classer les sujets recensés selon l'intérêt qu'ils représentent pour l'entreprise et les facilités de réalisation. Ce classement nous aidera à choisir l'activité à modéliser en premier lieu, de manière à garantir des résultats satisfaisants pour l'institution.

4.1.2. Processus de la modélisation dimensionnelle :

La conception d'un modèle dimensionnel passe par cinq étapes essentielles :

Choix de l'activité à modéliser : ce choix se fait après classement des activités dans une matrice dite *d'analyse des priorités* [Source : Kimball, 2004]. Cette matrice permet de connaître quelle activité choisir en premier. Le classement des sujets recensés, qui s'est fait en collaboration avec les décideurs et les techniciens de l'entreprise, est illustré dans la figure suivante :

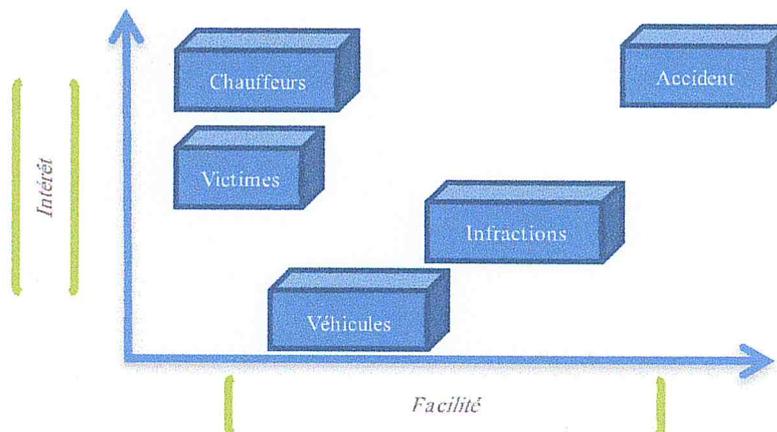


Figure 4.1.1 : Analyse des priorités

- Définition de l'activité et de son grain ;
- Définition des dimensions qui décrivent une ligne de la table de fait ;
- Définition des mesurables du fait ;
- Définition des agrégats.

4.1.2.1. Volet « Accident de la route » :

a. Présentation du volet « Accident de la Route » :

« Un **accident de la route** est un évènement se déroulant sur une voie de circulation et impliquant un véhicule, motorisé ou non, ou un individu, et entraînant des répercussions au niveau matériel ou physique» [DICTIONNAIRE FRANÇAIS].

La Gendarmerie Nationale algérienne, par sa vocation militaire et sécuritaire assume des missions de police administrative, police judiciaire mais aussi de sécurité routière. Elle a pour mission principale d'assurer la sécurité des citoyens, qui passe, entre autres, par la maîtrise en diminution des accidents de la route et la minimisation de l'hécatombe sur les routes.

Cela passe par la prévention qui ne peut se faire sans une aide à la décision sur la base des informations interprétées depuis les systèmes de productions OLTP.

Cette mission est assurée principalement par les unités territoriales : Brigades Territoriales, Compagnies Territoriales, Escadrons de la Sécurité Routières et les Brigades de la Sécurité Routières.

La Gendarmerie Nationale traite et intervient sur les accidents routiers se produisant dans les milieux suburbains, les accidents ayant lieu dans le milieu urbain son traités par la sûreté nationale.

b. Grain du volet accident :

Afin de maintenir une flexibilité maximale, les décideurs ont opté à la granularité permettant d'assurer la génération du nombre d'accidents, nombre de décès, nombre de blessures et nombres des témoins oculaires par : type de collision, construction sur la route, commune, luminosité, caractéristiques de la route, état de la route, type de la route, direction de la circulation, densité de la circulation, type de l'accident, nature de l'accident, météo, défaut de la route et enfin de les ventiler par les données temporelles.

c. Les dimensions participantes du modèle :

Les dimensions ajoutent du sens aux faits et permettent leurs interprétations. Dans ce qui suit, toutes les dimensions participant dans l'étoile seront décrites.

1. Dimension Temps :

La dimension temps est « la seule dimension qui figure systématiquement dans tout entrepôt de données, car en pratique tout entrepôt de données est une série temporelle. Le temps est le plus souvent la première dimension dans le classement sous-jacent de la base de données » [Kimball, 2001].

La dimension temps se présente comme suit :

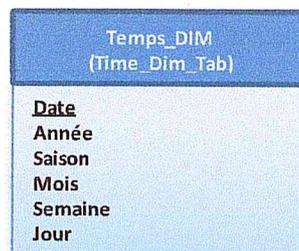


Figure 4.1.2 : La dimension Temps du volet « Accident »

Cette dimension contient plusieurs niveaux de détail depuis l'année jusqu'au jour dans la semaine. Contrairement aux dimensions et aux tables de faits qui vont être listées, cette dimension contient une clé qui n'est autre que la date dans le format court dd/MM/yyyy. Le tableau suivant donne plus de détails concernant la dimension « Temps » :

Désignation	Détail
Date	La date au format complet
Année	Année de la date
Saison	Trimestre de la date
Mois	Nom du mois
Semaine	Numéro de la semaine dans l'année
Jour	Nom du jour de la semaine

Tableau 4.1.1 : Tableau descriptif de la dimension « Temps ».

2. Dimension Heure :

Les décideurs jugent que l'analyse de l'occurrence des accidents ventilées par les heures de la journée et cruciale pour la prise des décisions telle que le renforcement de l'effectif sur les route en les heures enregistrant un pic en nombre d'accident.

La difficulté d'introduction de l'heure dans la dimension « Temps » fais que le recours à la création d'une dimension indépendante, dimension « Heure », soit le plus logique.

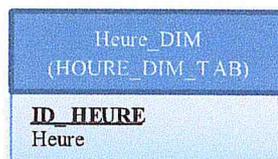


Figure 4.1.3 : La dimension « Type Accident » du volet « Accident »

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_HEURE	Clés artificielle attribuée à un tuple de la dimension heure
Heure	Un numéro sur deux positions de 00 à 23 représentant les heures du jour sur 24h.

Tableau 4.1.2 : Tableau descriptif de la dimension « Heure_DIM ».

3. Dimension Type Accident :

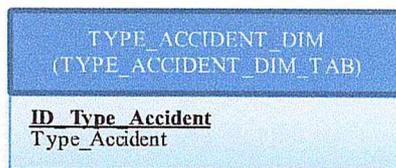


Figure 4.1.4 : La dimension « Type Accident » du volet « Accident »

Cette dimension permet de spécifier le milieu d'occurrence de l'accident : Milieu civil, Milieu militaire ou Zone de la Gendarmerie Nationale.

Différentes mesures et procédures de travail lors de la constatation des accidents et l'élaboration des procès-verbaux sont facteurs des valeurs de cette dimension.

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Type Accident	Clé artificielles ou de substitution gérée par une séquence automatique.
Type_Accident	Milieu d'occurrence de l'accident : militaire, civil ou zone de la Gendarmerie Nationale.

Tableau 4.1.3 : Tableau descriptif de la dimension « Type_Accident ».

4. Type de Collision :

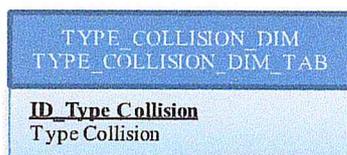


Figure 4.1.5 : La dimension « Type collision » du volet « Accident »

Cette dimension permet aux décideurs de ventiler les mesures selon la nature ou le type de la collision, qui peut être soit: Collision en biais, Collision frontale, Renversement.

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Type Collision	Identifiant ou clé de substitution générée par une séquence automatique.
Type_Collision	Type de la collision, la façon dont l'impact s'est produit

Tableau 4.1.4 : Tableau descriptif de la dimension « Type_Accident ».

5. Construction :

Cette dimension permet d'ajouter le contexte de l'état des constructions sur la route lors de l'occurrence de l'accident. Elle permet de ventiler les faits selon les données suivantes : Constructions signalisées, Constructions non signalisées ou Pas de constructions.

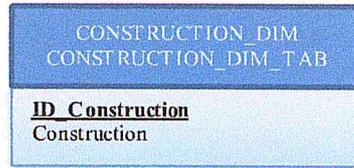


Figure 4.1.6: La dimension « Construction sur la route » du volet « Accident»

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Construction	La clé de substitution pour la dimension. Elle sera gérée par une séquence automatique.
Construction	L'état des constructions sur la route au moment de l'avènement de l'accident.

Tableau 4.1.5 : Tableau descriptif de la dimension « Construction_Dim ».

6. Commune :

La dimension commune, comme son nom l'indique, permet d'apporter des informations sur la commune d'occurrence du fait accident. Elle contient plusieurs niveaux constituant la hiérarchie ascendante suivante : Wilaya, Daïra, et Commune.

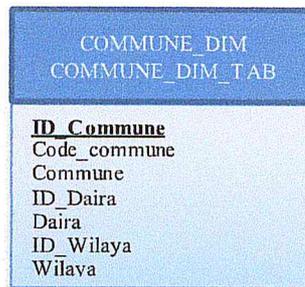


Figure 4.1.7 : La dimension « Commune » du volet « Accident»

Elle dispose d'un identifiant qui est une clé de substitution gérée par une séquence automatique. Cependant, et dans un souci d'éliminer les commune polycèmes qui sont des communes portant le même nom mais appartenant à différentes wilayas ou daïras, un autre niveau a été ajouté qui n'est autre que la clé naturelle de la commune dans le système de production afin de lier la table des fait à la bonne commune.

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Commune	Clé de substitution de la dimension.
Code_commune	Clé naturelle de la commune des systèmes sources. Elle est utilisée ici afin de différencier entre les communes portant le même nom mais appartenant à différentes Daïras ou Wilayas.
Commune	Nom de la commune.
ID_Daira	Identifiant de la Daïras.
Daïra	Nom de la daïra.
ID_Wilaya	Identifiant de la Wilaya

Tableau 4.4.6 : Tableau descriptif de la dimension « Cmmune_Dim ».

La dimension « Commune » est très importante pour l'intégration des analyses avec les données géographiques. Les faits seront disposés sur la carte géographique sur la base de cette dimension en faisant la liaison entre les shapes files et les valeurs des communes et wilayas.

7. Luminosité (Illumination) :

Cette dimension permet de renseigner les faits sur le degré de luminosité lors de l'occurrence de l'accident.

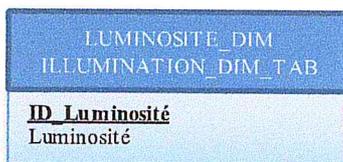


Figure 4.1.8: La dimension « Luminosité » (Illumination) du volet « Accident»

Elle contient des données sur le degré de luminosité au moment de l'accident, telles que : Jour ensoleillé, Temps pluvieux, Vents fort.

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Luminosité	Clé de substitution générée par une séquence automatique.
Luminosité	Le libellé de la luminosité qui peut être : Jour ensoleillé, Temps pluvieux ...

Tableau 4.1.7 : Tableau descriptif de la dimension « Illumination_Dim » (Luminosité).

8. Caractéristiques de la route (Road Character) :

Cette dimension qualifie les faits par les caractéristiques de la route, lieu de l'accident.

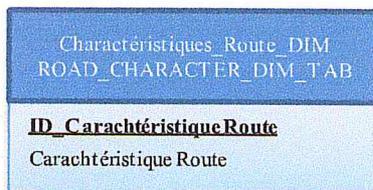


Figure 4.1.9 : La dimension « caractéristiques de la route » (Road Character) du volet « Accident»

Elle contient des données sur les caractéristiques de la route, telles que : Route à un sens unique, Intersection ou Pont.

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Caractéristique Route	Identifiant de substitution
Caractéristique Route	Libellé de la caractéristique de la route, lieu d'occurrence de l'accident : pont, route à double sens, ...etc.

Tableau 4.1.8 : Tableau descriptif de la dimension «caractéristiques de la route» (Road_Character_Dim).

9. Condition de la route (Road Condition) :

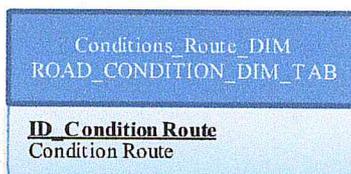


Figure 4.1.10 : La dimension « conditions de la route » (Road Condition) du volet « Accident»

La dimension Condition de la route, permet d'ajouter du contexte aux faits sur la condition de la route en données telles que : Route mouillée, Flaque d'eau, Verglas, Ordures.

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Condition Route	Clé de substitution
Condition Route	Intitulé de la condition de la route : route glissante, verglas, ... Etc

Tableau 4.1.9 : Tableau descriptif de la dimension «conditions de la route» (Road_Condition_Dim).

10. Spécificités de la route (Road Special Feature) :

Elle permet de renseigner sur les spécificités de la route où l'accident a eu lieu.

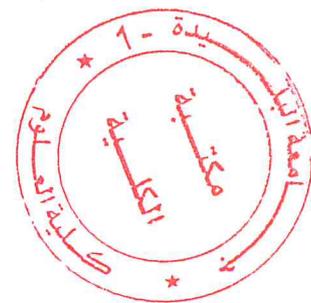
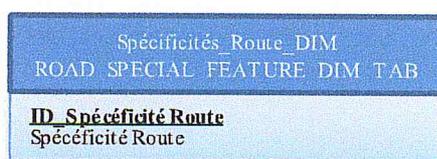


Figure 4.1.11 : La dimension «spécificités de la route» (Road_Special_Feature_Dim) du volet « Accident»

Elle qualifie les faits avec des données telles que : Route nationale, Route Wilayale.

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Spécificité Route	Clé de substitution de la dimension
Spécificité Route	Libellé de la spécificité de la route

Tableau 4.1.10 : Tableau descriptif de la dimension «spécificités de la route» (Road_Special_Feature_Dim).

11. Orientations du trafic routier (Traffic Flow) :

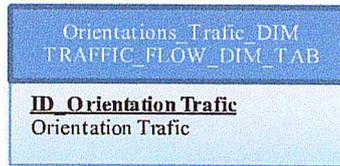


Figure 4.1.12 : La dimension « orientations de la circulation » (Road_Flow_Dim) du volet « Accident»

Cette dimension informe sur la direction ou l'orientation du trafic routier dans le lieu de l'accident, tel que : Nord, Sud, Sud-Ouest. Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Orientation_Trafic	Clé de substitution gérée par une séquence automatique.
Orientations_Trafic	Libellé de l'orientation de la circulation : Nord, Sud, ...Etc

Tableau 4.1.11 : Tableau descriptif de la dimension «orientations de la circulation» (Traffic_Flow_Dim).

12. Densité du Trafic (Traffic level) :

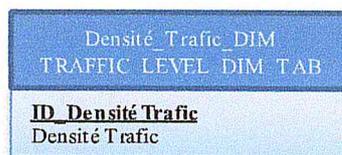


Figure 4.1.13 : La dimension « densité de la circulation » (Road_Flow_Dim) du volet « Accident»

Cette dimension informe sur la densité de la circulation ou du trafic routier dans le lieu de l'accident, tel que : Fluide, Circulation encombrée, Circulation dense, Circulation bloquée. Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Densité_Trafic	Clé de substitution de la dimension.
Densité_Trafic	Intitulé de la densité de la circulation.

Tableau 4.1.12: Tableau descriptif de la dimension «densité de la circulation» (Traffic_Level_Dim).

13. Nature Accident :

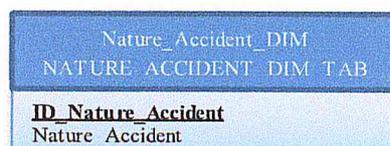
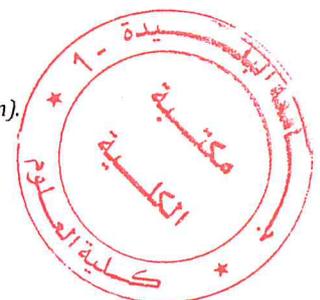


Figure 4.1.14 : La dimension « nature de l'accident » (Nature_Accident_Dim) du volet « Accident»



Elle apporte l'information sur la nature de l'accident, est-il mortel, corporel ou matériel. Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_NATURE_ACCIDENT	Clé artificielle de la dimension, gérée par une séquence automatique.
Nature_Accident	Libellé de la nature de l'accident : Mortel, Corporel ou matériel.

Tableau 4.1.13 : Tableau descriptif de la dimension «nature de l'accident» (Nature_Accident_Dim).

14. Météo (Weather) :



Figure 4.1.15 : La dimension « Météo » (Weather_Dim) du volet « Accident»

La dimension météo décrit les conditions météorologiques lors de l'avènement de l'accident. Elle contient des données telles que : Temps claire, Temps pluvieux, Temps nuageux.

Désignation	Détail
ID_Météo	Clé de substitution de la dimension
Météo	Intitulé de la météo

Tableau 4.1.14 : Tableau descriptif de la dimension «météo» (Weather_Dim).

15. Défauts de la route (Roadway surface) :

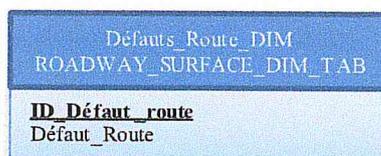


Figure 4.1.16 : La dimension « Défauts de la route » (Roadway_Surface_Dim) du volet « Accident»

Cette dimension permet de renseigner les faits sur les défauts de la route, lieu de l'accident. Elle apporte les données telles que : Bordures de route défectueuses, Nid de poule (Ornière), Affaissement de la route. Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Défauts_Route	Clé artificielle de la dimension.
Défauts_Route	Libellé ou intitulé des défauts de la route.

Tableau 4.1.15 : Tableau descriptif de la dimension «défauts de la route» (Roadway_Surface_Dim).

16. Cause de l'Accident (Cause_dim) :

Cette dimension est très importante, car elle apporte les informations concernant les causes qui ont mené directement ou indirectement à l'accident. Elle contient deux niveaux : la cause et son type, constituant une hiérarchie ascendante « Cause, Type ».

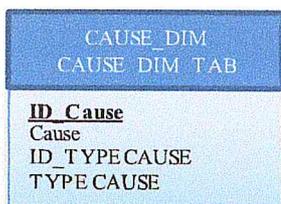


Figure 4.1.17 : La dimension « Cause de l'accident » (Cause_Dim) du volet « Accident»

Elle contient des données comme suite :

- Cause : Circulation en sens interdit avec son type de cause : « Cause liées aux conducteurs » ;
- Cause : Panne mécanique avec son type de cause : « Causes liées aux véhicules ».

Le tableau suivant décrit avec plus de détail les attributs de cette dimension :

Désignation	Détail
ID_Cause	Clé substitution.
Cause	Le libellé de la cause.
ID_TYPE_CAUSE	Clé de substitution du niveau Type de la Cause.
TYPE_CAUSE	Le libellé du type de la cause.

Tableau 4.1.16 : Tableau descriptif de la dimension «cause de l'accident» (Cause_Dim).

d. Les mesurables :

Les mesurables de la table des faits « Accident » peuvent être résumés en quatre :

- Le nombre d'accidents ;
- Le nombre de décès ;
- Le nombre de blessés ;
- Le nombre de témoins.

e. Le modèle en étoile du volet « Accident »¹⁴ :

f. Les agrégats :

Une des techniques d'amélioration des performances des data warehouses est la création des tables d'agrégats. Cela passe par la réduction de quantité des données que le système de gestion de base de données (SGBD) traitera afin de répondre à une requête. Cette technique passe par l'agrégation des données contenues dans les tables de faits détaillées pour être stockées dans de nouvelles tables de faits.

Les nouvelles tables d'agrégats puisent ses données du modèle en étoile détaillé, et cela pourra résulter en :

- **La création de nouvelles dimensions :** Ces dimensions découlent de celles existantes dans le modèle détaillé avec l'omission des niveaux représentant la granularité la plus fine ;
- **L'élimination de quelques dimensions :** selon le nouveau niveau de détail obtenu après l'agrégation, il arrive que certaines dimensions, reflétant un niveau de détail plus fin, soient supprimées.
- **Création de nouveaux faits :** les nouvelles tables agrégées, compte tenue de l'information agrégées qu'elles introduisent, compteront des faits qui n'existaient pas dans le modèle en étoile détaillé.
- **Utilisation des tables pré-jointes :** la pré-jointure signifie qu'une table d'agrégat est construite à partir de la jointure de la table des faits détaillée avec les dimensions impliquées dans l'agrégation, afin de que les résultats soient prêts à être exploités.

La structure de la table d'agrégat détermine sa visibilité à l'utilisateur final :

- Elle est invisible lorsqu'elle reflète exactement le modèle de base
- Elle est visible lorsqu'elle contient des faits supplémentaires.

Cela dit, le recours aux tables agrégées est fait pour améliorer la performance du système, en précalculant les agrégats les plus utilisés. Elles doivent, en revanche, fournir les mêmes résultats que le modèle de base.

C. Adamson, dans son livre « Mastering the Data Warehouse Aggregates, Solution for Star Schema Performance », propose une démarche intéressante qui consiste à :

- 1- Etablir une liste d'agrégats potentiels à partir du modèle en étoile détaillé ;
- 2- Déceler les agrégats utiles d'entre la liste globale ;
- 3- Implémenter le modèle agrégé avec les nouveaux faits, dimensions dérivées et celles à supprimer.

¹⁴ Voir Annexe 01 Figure A01.1

Les agrégats retenus :

Dimension	Agrégats	Agrégats possibles	Agrégats retenus
Time_Dim	Année, trimestre, mois, Semaine, jour	5	4
Heure_Dim	Heure	1	1
Type_Accident_dim	Type accident	1	1
Nature_Accident_dim	Nature accident	1	1
Cause	Cause, type cause	2	1
Roadway_surface	Défaut de la route	1	1
Weater	Conditions météorologiques	1	1
Trafic level	Densité de la circulation	1	1
Traffic flow	Orientation de la circulation	1	1
Road special feature	Spécificité de la route	1	1
Road character	Caractéristiques de la route	1	1
Road condtion	Conditions de la route	1	1
Illumination	Luminosité	1	1
Commune	Wilaya, Daira, commune, code commune	4	3
Construction	Construction sur la route	1	1
Type collision	Type collision	1	1
Type accident	Type accident	1	1

Tableau 4.1.17 : Tableau descriptif des agrégats utiles du modèle « accident ».

Le tableau précédent, les besoins recensés ainsi que les accès fréquents suggèrent les agrégats les plus pertinents et incitent à la création des modèles d'agrégats suivants :

- Gravité des accidents par années, mois, jours, et heures ;
- Répartition des accidents par nature d'accident ventilés par années, mois, jours, et heures ;
- Répartition des décès et blessés dans les accidents de la route sur années, mois, jours, et heures ;

4.1.2.2. Volet « Chauffeurs » :

a. Présentation du volet « Chauffeurs » :

Le volet chauffeur ou conducteur traite des personnes impliquées dans un accident routier et qui conduisaient au moment de l'occurrence de l'accident. Peu importe que la personne soit la propriétaire du véhicule, objet de l'accident, ou pas.

Chapitre Quatrième : Conception du Data Warehouse.

Il est à préciser qu'un conducteur pourra être considéré comme une victime s'il est blessé ou décédé suite à l'impact, et par conséquent sera aussi traité comme victime dans le volet « Victime ».

b. Granularité :

Le grain le plus fin correspond à :

La possibilité de savoir le nombre des chauffeurs impliqués dans les accidents, ainsi que le nombre d'accidents dont ces chauffeurs ont causé, selon les valeurs des dimensions et choisies par les décideurs.

- Pour les accidents : type de collision, construction sur la route, commune, luminosité, caractéristiques de la route, état de la route, type de la route, direction de la circulation, densité de la circulation, type de l'accident, nature de l'accident, météo, défaut de la route et enfin les ventiler par les données temporelles ;
- Pour les chauffeurs : Type du Chauffeurs, Sexe, Tranche d'âge, niveau d'éducation, propriétaire du véhicule ou pas et en enfin portait-il son permis de conduire au moment de l'accident ou pas.

Ce besoin d'avoir le nombre d'accident dont les chauffeurs, répondant à certaines valeurs des dimensions, nécessitera un traitement spécial, que nous verrons plus loin, à cause de ce qui est connu sous l'expression dimensions multivaluées ou relation Many-to-Many entre la table des faits et les dimensions représentant les chauffeurs.

c. Les dimensions participantes du modèle :

Les dimensions communes :

Les utilisateurs auront besoin de ventiler les nombres des chauffeurs selon les spécificités de l'accident. Donc toutes les dimensions listées dans le volet « Accident » seront présentes dans ce volet. Il n'est pas nécessaire de les relister dans cette partie. Il sera question d'introduire, uniquement, les nouvelles dimensions.

Dimension	Volet Accident	Volet Chauffeurs
Temps	✓	✓
Heure	✓	✓
Type Accident	✓	✓
Type de Collision	✓	✓
Construction	✓	✓
Commune	✓	✓
Luminosité	✓	✓
Caractéristiques de la route	✓	✓
Condition de la route	✓	✓

Spécificités de la route	✓	✓
Orientations du trafic routier	✓	✓
Densité du Trafic	✓	✓
Nature Accident	✓	✓
Météo	✓	✓
Défauts de la route	✓	✓
Cause de l'Accident	✓	✓
Accident_Reportnum (Dimension dégénérée)		✓
Nationalité		✓
Tranche d'âge		✓
Sexe		✓
Niveau d'éducation		✓
Propriétaire		✓
Port Permis de Conduire		✓
Type Chauffeur		✓

Tableau 4.1.18 : Détection des dimensions communes entre les volets « Accident » et « Chauffeur ».

1. Dimension dégénérée « Accident_Reportnum » :

La table des faits contient les mesures du sujet, mises en contextes par les dimensions incluses à travers des clés étrangères référençant les clés primaires des dimensions liées au fait dans le modèle en étoile. Cependant, il existe des clés faisant partie de la clé de la table des faits sans, pour autant, référencer une dimension représentée par une table. On parle, alors, de la dégénération de dimension.

La dimension « **Accident_Reportnum** » a été introduite et dégénérée dans le modèle dimensionnel du volet « **Chauffeurs** » (ainsi que tous les volets qui suivent, comme il sera montré plus loin). Elle ne possède aucune description textuelle. Elle existe à travers les clés de la table des faits. En fait, elle a été intégrée afin de résoudre les problèmes des relations dimensionnelles multivaluées ou relations plusieurs-à-plusieurs (many-to-many) entre un accident et ses chauffeurs.

Un accident peut être associé à plusieurs chauffeurs, plusieurs causes, plusieurs infractions et plusieurs victimes, pour ne citer que peu. Lors des analyses, on aura besoin d'avoir le nombre des chauffeurs répondants à certaines valeurs des dimensions ainsi que le nombre d'accidents dont ces chauffeurs ont été impliqués. Elle sera utilisée lors des agrégats pour s'assurer que le même accident soit compté une seule fois avec tous les chauffeurs qui lui sont rattachés.

La formule de l'agrégat sera de type « SUM » sur le nombre des chauffeurs et « COUNT DISTINCT » sur cette clé (Accident_reportnum) pour avoir le nombre d'accident.

Explications à travers un scénario :

Il y'a un besoin de calculer le nombre de chauffeurs répondants à certaines valeurs de dimensions ainsi que le nombre d'accidents en relation avec ce nombre de chauffeurs (Rapport Nombre_CHAUFFEURS/NB_ACCIDENT impliquant ces chauffeurs). On risque de calculer le même accident autant de fois qu'il y'a de chauffeurs

Soit le scénario suivant :

- Dimensions de l'accident : Météo et Type de l'accident ;
- Dimensions du chauffeur : Tranche âge et Niveau éducation ;
- Avec la table des faits ayant le jeu de données suivant :
 - L'accident A1 contient deux chauffeurs ayant la même tranche d'âge et même niveau d'éducation qui sont dans la ligne 1 et 2 ;
 - Un troisième chauffeur appartenant à une tranche d'âge et niveau d'éducation différents.

N° ligne	Accident_reportnum	Météo	Accident_type	Tranche âge	Niveau éducation
1	A1	Ensoleillé	EN BIAIS	20 à 30	UNIVERSITAIRE
2	A1	Ensoleillé	EN BIAIS	20 à 30	UNIVERSITAIRE
3	A1	Ensoleillé	EN BIAIS	31 à 40	PRIMAIRE

Tableau 4.1.19: Jeux de données des dimensions pour le scénario de démonstration de la dégénération de la dimension « Accident_Reportnum »

Résultat des mesures sans dégénération de l'accident reportnum :

- L'agrégat des accidents se fera sur Météo et Accident_Type ;
- L'agrégat des chauffeurs se fera sur Météo, Accident_Type, Niveau_education et Tranche_age.

	Météo	Accident type	Tranche âge	Niveau éducation	NB chauffeur	Nb accident
1	Ensoleillé	EN BIAIS	20*30	UNIVERSITAIRE	2	2
2	Ensoleillé	EN BIAIS	20*30	UNIVERSITAIRE		
3	Ensoleillé	EN BIAIS	31*40	PRIMAIRE	1	1
TOTAL					3	3

Tableau 4.1.20 : Résultats de l'agrégation sans dégénération : L'accident A1 est calculé trois fois au lieu d'une seule fois

La mesure « nb_accident » sera calculée deux fois avec les lignes 1 et 2 du tableau. Elle sera calculée, une autre fois, avec la ligne 3 du tableau. Ce qui résulte en trois accidents, qui ne sont en réalité qu'un seul qui est l'accident A1

On a aucun moyen de regrouper les chauffeurs ayant des valeurs différentes de dimensions mais qui appartiennent au même accident.

Résultat des mesures avec dégénération du Accident_reportnum :

Accident Reportnum	Météo	Accident Type	Tranche Age	Niveau Education	NB CHAUFFEURS	NB_ACCIDENT SUM(COUNT(DISTINCT T (ACCIDENT_REPORT NUM))
1	A1	Ensoleillé	EN BIAIS	20*30	UNIVERSITAIRE	2
2	A1	Ensoleillé	EN BIAIS	20*30	UNIVERSITAIRE	1
3	A1	Ensoleillé	EN BIAIS	31*40	PRIMAIRE	1
Total					3	1

Tableau 4.1.20 : Résultat de l'agrégation avec la dégénération de la dimension « Accident_Reportnum »

Avec cette solution on pourra regrouper les chauffeurs avec leurs accidents de telle sorte à

- Calculer le nombre des chauffeurs selon les valeurs des dimensions Météo, Type Accident, Tranche_age et Niveau_éducation ;
- Faire un COUNT(DISTINCT accident_reportnum) pour s'assurer que le même accident est calculé une seule fois pour la totalité des chauffeurs répondant aux valeurs des dimensions.

Sans la dégénération du « Accident_Reportnum », l'agrégateur aurait inclus tous les accidents répondant aux mêmes valeurs des dimensions Météo et Accident_type, même si les chauffeurs n'y figurent pas.

Remarque : Oracle BI permet de définir les agrégats de type COUNT DISTINCT sur les colonnes séparément. Il est donc pas possible de faire une expression groupée comme suite « COUNT(DISTINCT météo, Accident_Type) ». Cela présente une contrainte technique et qui pose vers la dégénération du « Accident_Reportnum ».

2. Dimension « Type Chauffeur » :

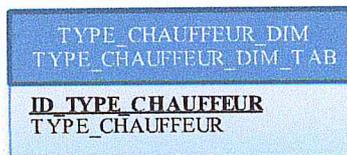


Figure 4.1.19 : La dimension « TYPE CHAUFFEUR » du volet « CHAUFFEUR »

La dimension « Type Chauffeur » liste les différents types que peut prendre un conducteur : Civile, Militaire, Gendarme ou Autre (Inclus les valeurs nulles).

Désignation	Détail
ID_TYPE_CHAUFFEUR	Clé de substitution gérée par une séquence automatique.
TYPE_CHAUFFEUR	Libellé du type du chauffeur.

Tableau 4.1.21 : Tableau descriptif de la dimension « TYPE CHAUFFEUR » du volet « CHAUFFEUR »

3. Dimension Tranche d'âge :

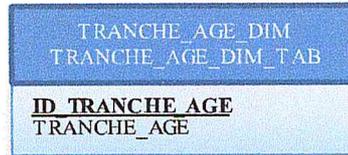


Figure 4.1.20 : La dimension « TRANCHE_AGE » du volet « CHAUFFEUR»

Cette dimension revient avec le volet Victime, car elle décrit un aspect lié aux personnes d'une façon générale. Elle contient les différents intervalles d'âge d'une personne : Entre 0 et 5 ans, Entre 6 et 10 ans, Entre 11 et 15 ans, Entre 16 et 20 ans, Plus de 71 ans ou Autre (Inclus les valeurs nulles).

Désignation	Détail
<u>ID_TRANCHE_AGE</u>	Clé artificielle de la dimension.
TRANCHE_AGE	Libellé de l'intervalle de la tranche d'âge.

Tableau 4.1.21 : Tableau descriptif de la dimension « TRANCHE_AGE » du volet « CHAUFFEUR»

4. Dimension Sexe :

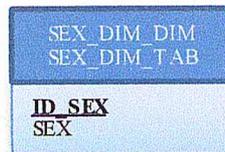


Figure 4.1.21 : La dimension « SEX_DIM » du volet « CHAUFFEUR»

Cette dimension revient, aussi, avec le volet Victime, car elle décrit un aspect lié aux personnes d'une façon générale. Elle contient les différents sexes possibles d'une personne : Homme ou Femme.

Désignation	Détail
<u>ID_SEX</u>	Clé de substitution.
SEX	Intitulé du sexe.

Tableau 4.1.22 : Tableau descriptif de la dimension « SEX_DIM » du volet « CHAUFFEUR»

5. Dimension niveau d'éducation (EDUCATION_DIM):

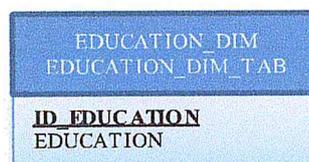


Figure 4.1.22 : La dimension « EDUCATION_DIM » du volet « CHAUFFEUR»

Comme les dimensions « Sexe » et « Tranche d'âge », la dimension « Education » revient avec le volet Victime, car elle décrit un aspect lié aux personnes d'une façon générale. Elle contient les différents niveaux d'éducatons possibles d'une personne : Universitaire, Secondaire, Primaire, Sans niveau.

Désignation	Détail
<u>ID_EDUCATION</u>	Clé artificielle de la dimension.
EDUCATION	Libellé de la dimension « éducation ».

Tableau 4.1.23 : Tableau descriptif de la dimension « EDUCATION_DIM » du volet « CHAUFFEUR »

6. Dimension Propriétaire :

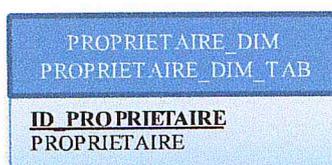


Figure 4.1.23 : La dimension « PROPRIETAIRE_DIM » du volet « CHAUFFEUR »

Cette dimension décrit si le chauffeur, impliqué dans l'accident, est propriétaire du véhicule ou pas : OUI, NON ou AUTRE (Inclus les valeurs non renseignées dans le système source).

Désignation	Détail
<u>ID_PROPRIETAIRE</u>	Clé artificielle de la dimension.
PROPRIETAIRE	Intitulé du propriétaire.

Tableau 4.1.24 : Tableau descriptif de la dimension « PROPRIETAIRE_DIM » du volet « CHAUFFEUR »

7. Dimension Port Permis de Conduire :

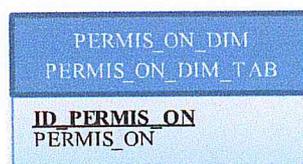


Figure 4.1.24 : La dimension « PERMIS_ON_DIM » du volet « CHAUFFEUR »

Cette dimension décrit si le chauffeur, impliqué dans l'accident, était en possession d'un permis de conduire au moment de l'accident : OUI, NON, Permis de conduire provisoire ou AUTRE (Inclus les valeurs non renseignées dans le système source).

Désignation	Détail
<u>ID_PERMIS_ON</u>	Clés artificielle de la dimension.
PERMIS_ON	Libellé de port du permis de conduire.

Tableau 4.1.25 : Tableau descriptif de la dimension « PERMIS_ON_DIM » du volet « CHAUFFEUR »

8. Dimension Nationalité (Cityzenship_dim) :

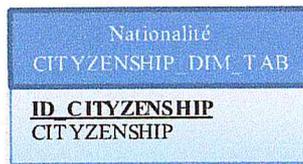


Figure 4.1.25: La dimension «Nationalité » du volet « CHAUFFEUR»

Cette dimension renseigne sur la nationalité du chauffeur. Elle reviendra avec le volet Victime.

Désignation	Détail
ID_CITYZENSHIP	Clé de substitution.
CITYZENSHIP	Libellé de la nationalité.

Tableau 4.1.26 : Tableau descriptif de la dimension « Nationalité » du volet « CHAUFFEUR»

d. Les mesurables :

Les mesurables de la table des faits « CHAUFFEUR» peuvent être résumés en :

- Le nombre des chauffeurs ;
- Le nombre d'accidents impliquant ces chauffeurs. Ce nombre n'est en réalité que la formule SUM(COUNT(DISTINCT Accident_Reportnum)), de la colonne dégénérée.

e. Le modèle en étoile du volet « Chauffeurs »¹⁵ :

f. Les agrégats retenus :

Dimension	Agrégats	Agrégats possibles	
Time_Dim	Année, trimestre, mois, Semaine, jour	5	4
Heure_Dim	Heure	1	1
Type_Accident_dim	Type accident	1	1
Nature_Accident_dim	Nature accident	1	1
Cause	Cause, type cause	2	1
Roadway_surface	Défaut de la route	1	1
Weater	Conditions météorologiques	1	1
Traffic level	Densité de la circulation	1	1
Traffic flow	Orientation de la circulation	1	1

¹⁵ Voir Annexe 01 Figure A01.2

Road special feature	Spécificité de la route	1	1
Road character	Caractéristiques de la route	1	1
Road condtion	Conditions de la route	1	1
Illumination	Luminosité	1	1
Commune	Wilaya, Daira, commune, code commune	4	3
Construction	Construction sur la route	1	1
Type collision	Type collision	1	1
Type accident	Type accident	1	1
Type chauffeur	Type Chauffeur	1	1
Sexe	Sexe	1	1
Tranche d'âge	Tranche d'âge	1	1
Nationalité	Nationalité	1	1

Tableau 4.1.27 : Tableau descriptif des agrégats utiles du modèle « accident ».

Le tableau précédent, les besoins recensés ainsi que les accès fréquents suggèrent les agrégats les plus pertinents et incitent à la création des modèles d'agrégats suivants :

- Nombre des chauffeurs et les accidents correspondant par Gravité des accidents, nationalité, sexe, tranche d'âge des chauffeurs ventilés par années, mois, jours, et heures ;
- Nombre des chauffeurs et les accidents correspondant par nature d'accident, nationalité, sexe, tranche d'âge des chauffeurs ventilés par années, mois, jours, et heures ;

4.1.2.3. Volet « Infraction routière »

a. Présentation du volet « Infractions » :

Ce volet traite des infractions commises par les conducteurs ayant induit à l'accident et qui ont fait l'objet d'une contravention.

Certains accidents ne comptent pas d'infractions, les causes de l'accident étant fortuites, telles que les causes naturelles. Ici, il est questions des infractions commises par les chauffeurs ou piétons et qui ont eu pour effet l'occurrence de l'accident.

b. Granularité :

Le grain le plus fin correspond à :

La possibilité de savoir le nombre de fois une infraction est commise (la fréquence d'occurrence) ainsi que le nombre d'accidents où elles figurent. Par exemple, deux chauffeurs impliqués dans le même accident ayant commis la même infraction « Excès de vitesse » aurait pour résultats :

- Nombre d'occurrence de l'infraction=2 ;

- Nombre d'accidents induit par cette infraction=1 ;

Les utilisateurs auront besoin de ventiler les données liées aux infractions selon deux axes principaux : caractéristiques liées à l'accident et celles à l'infraction :

- Pour les accidents : type de collision, construction sur la route, commune, luminosité, caractéristiques de la route, état de la route, type de la route, direction de la circulation, densité de la circulation, type de l'accident, nature de l'accident, météo, défaut de la route et enfin les ventiler par les données temporelles ;
- Pour les infractions : qualification de l'infraction et l'intitulé de l'infraction.

Ce besoin d'avoir le nombre d'accidents induits par les infractions (répondant aux valeurs des dimensions propres à l'infraction) nécessitera un traitement spécial, comme vu avec le volet « Chauffeurs », à cause de ce qui est connu sous l'expression de dimensions multivaluées ou relation Many-to-Many entre la table des faits et les dimensions représentant les victimes.

c. Les dimensions participantes dans le modèle :

Les dimensions communes :

Les utilisateurs auront besoin de ventiler les nombres d'occurrence des infractions selon les spécificités de l'accident. Donc toutes les dimensions listées au volet « Accident » seront présentes dans ce volet, en y ajoutant le « **Accident_Reportnum** » comme une dimension dégénérée.

Il n'est pas nécessaire de les relister dans cette partie. Il sera question d'introduire, uniquement, les nouvelles dimensions.

Dimension	Volet Accident	Volet Chauffeurs	Volet Infraction
Temps	✓	✓	✓
Heure	✓	✓	✓
Type Accident	✓	✓	✓
Type de Collision	✓	✓	✓
Construction	✓	✓	✓
Commune	✓	✓	✓
Luminosité	✓	✓	✓
Caractéristiques de la route	✓	✓	✓
Condition de la route	✓	✓	✓
Spécificités de la route	✓	✓	✓
Orientations du trafic routier	✓	✓	✓
Densité du Trafic	✓	✓	✓

Chapitre Quatrième : Conception du Data Warehouse.

Nature Accident	✓	✓	✓
Météo	✓	✓	✓
Défauts de la route	✓	✓	✓
Cause de l'Accident	✓	✓	✓
Accident_Reportnum (Dimension dégénérée)		✓	✓
Nationalité		✓	
Tranche d'âge		✓	
Sexe		✓	
Niveau d'éducation		✓	
Propriétaire		✓	
Port Permis de Conduire		✓	
Type Chauffeur		✓	
Infraction			✓

Tableau 4.1.28 : Détection des dimensions communes entre les volets « Accident » et « Chauffeur » et « infraction »

1. La dimension Infraction :

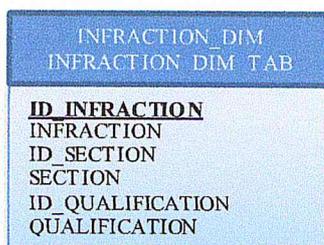


Figure 4.1.27 : La dimension «INFRACTION_DIM » du volet « INFRACTION»

Cette dimension décrit l'infraction commise et induisant à l'accident et ayant fait l'objet d'une contravention, selon trois niveaux :

- Qualification : Crime, délit ou contravention ;
- Section : une organisation selon la législation des infractions.
- Infraction : l'intitulé de l'infraction.

Désignation	Détail
ID_INFRACTION	Clé de substitution de la dimension.
INFRACTION	Libellé de l'infraction commise.
ID_SECTION	Clé de substitution pour le niveau de la section.
SECTION	Intitulé de la section.
ID_QUALIFICATION	Clé artificielle du niveau qualification.
QUALIFICATION	Libellé de la qualification.

Tableau 4.1.29: Tableau descriptif de la dimension « INFRACTION_DIM » du volet « INFRACTION »

d. Les mesurables :

Les mesurables de la table des faits « INFRACTION » peuvent être résumés en :

- La fréquence (nombre de fois) l'infraction est signalée dans les accidents ;
- Le nombre d'accidents dont l'infraction en question figure. Ce nombre n'est en réalité que la formule SUM(COUNT(DISTINCT Accident_Reportnum)), de la colonne dégénérée.

e. Le modèle en étoile:¹⁶

f. Les agrégats retenus :

Dimension	agrégats	Agrégats possibles	Agrégats retenus
Time_Dim	Année, trimestre, mois, Semaine, jour	5	4
Heure_Dim	Heure	1	1
Type_Accident_dim	Type accident	1	1
Nature_Accident_dim	Nature accident	1	1
Cause	Cause, type cause	2	1
Roadway_surface	Défaut de la route	1	1
Weater	Conditions météorologiques	1	1
Traffic level	Densité de la circulation	1	1
Traffic flow	Orientation de la circulation	1	1
Road special feature	Spécificité de la route	1	1
Road character	Caractéristiques de la route	1	1
Road condition	Conditions de la route	1	1
Illumination	Luminosité	1	1

¹⁶ Voir Annexe 01: Figure A01.3

Commune	Wilaya, Daira, commune, code commune	4	3
Construction	Construction sur la route	1	1
Type collision	Type collision	1	1
Type accident	Type accident	1	1
Infraction	Qualification de l'infraction	1	1

Tableau 4.1.30 : Tableau descriptif des agrégats utiles du modèle « infraction ».

Le tableau précédent, les besoins recensés ainsi que les accès fréquents suggèrent les agrégats les plus pertinents et incitent à la création des modèles d'agrégats suivants :

- Nombre de fois l'infraction survient et les accidents correspondant par Gravité des accidents, qualification de l'infraction par années, mois, jours, et heures ;
- Nombre de fois l'infraction survient et les accidents correspondant par nature d'accident, qualification de l'infraction ventilés par années, mois, jours, et heures ;

4.1.2.4. Volet « Victime »

a. Présentation du volet «Victime » :

Les victimes de l'accident de la route sont toutes les personnes ayant été décédées ou blessées suite à l'accident. Elles peuvent être des conducteurs, passagers ou piétons.

b. Granularité :

Le grain le plus fin correspond à :

La possibilité de savoir le nombre de victimes ainsi que le nombre d'accident dont ils y sont impliqués et cela par :

- Pour les accidents : type de collision, construction sur la route, commune, luminosité, caractéristiques de la route, état de la route, type de la route, direction de la circulation, densité de la circulation, type de l'accident, nature de l'accident, météo, défaut de la route et enfin les ventiler par les données temporelles ;
- Pour les victimes : Type de la victime, nature de la victime, Sexe, Tranche d'âge, niveau d'éducation.

Ce besoin d'avoir le nombre d'accident selon les valeurs des dimensions des victimes nécessitera un traitement spécial, comme vu avec le volet « Chauffeurs, infractions et véhicules », à cause de ce qui est connu sous l'expression dimensions multivaluées ou relation Many-to-Many entre la table des faits et les dimensions représentant les victimes.

c. Les dimensions participantes du modèle :

Les dimensions communes :

Les utilisateurs auront besoin de ventiler les nombres des victimes selon les spécificités de l'accident. Donc toutes les dimensions listées dans le volet « Accident » seront présentes dans ce volet, en y ajoutant le « Accident_Reportnum » comme une dimension dégénérée.

Aussi, toutes les dimensions vues dans le volet « Chauffeurs » et ayant un caractère liés aux individus : Sexe, niveau éducation, Nationalité (Cityzenship).

Il n'est pas nécessaire de les reciter dans cette partie. Il sera question d'introduire, uniquement, les nouvelles dimensions.

Dimension	Volet Accident	Volet Chauffeurs	Volet Infraction	Volet Victime
Temps	✓	✓	✓	✓
Heure	✓	✓	✓	✓
Type Accident	✓	✓	✓	✓
Type de Collision	✓	✓	✓	✓
Construction	✓	✓	✓	✓
Commune	✓	✓	✓	✓
Luminosité	✓	✓	✓	✓
Caractéristiques de la route	✓	✓	✓	✓
Condition de la route	✓	✓	✓	✓
Spécificités de la route	✓	✓	✓	✓
Orientations du trafic routier	✓	✓	✓	✓
Densité du Trafic	✓	✓	✓	✓
Nature Accident	✓	✓	✓	✓
Météo	✓	✓	✓	✓
Défauts de la route	✓	✓	✓	✓
Cause de l'Accident	✓	✓	✓	✓
Accident_Reportnum (Dimension dégénérée)		✓	✓	✓
Nationalité		✓		✓
Tranche d'âge		✓		✓
Sexe		✓		✓
Niveau d'éducation		✓		✓
Propriétaire		✓		

Port Permis de Conduire	✓
Type Chauffeur	✓
Infraction	✓
Nature_Victime	✓
Type_Victime	✓

Tableau 4.1.31 : Détection des dimensions communes entre les volets « Accident » et « Chauffeur », « infraction » et « victimes »

1. Dimension « Nature_Victime » :

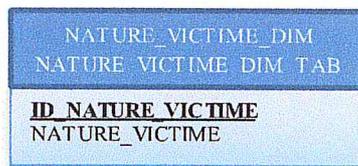


Figure 4.1.29 : La dimension «NATURE_VICTIME_DIM » du volet « VICTIME»

Cette dimension décrit la nature de la victime :Civile, Militaire, Gendarme ou AUTRE (Inclus les valeurs non renseignées dans le système source).

Désignation	Détail
<u>ID_NATURE_VICTIME</u>	Clé de substitution
NATURE_VICTIME	Libellé de la nature de la victime.

Tableau 4.1.32 : Tableau descriptif de la dimension « NATURE_VICTIME_DIM » du volet « VICTIMES»

2. Dimension « Type_Victime » :

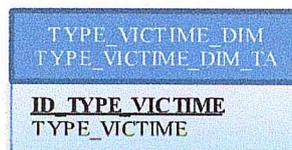


Figure 4.1.30 : La dimension «TYPE_VICTIME_DIM » du volet « VICTIME»

Cette dimension décrit le type de la victime : Passager ou Conducteur (Chauffeur) ;

Désignation	Détail
<u>ID_TYPE_VICTIME</u>	Clé artificielle de la dimension.
TYPE_VICTIME	Intitulé du type de la victime.

Tableau 4.1.33 : Tableau descriptif de la dimension « TYPE_VICTIME_DIM » du volet « VICTIMES»

d. Les mesurables :

Chapitre Quatrième : Conception du Data Warehouse.

Les mesurables de la table des faits « VICTIME» peuvent être résumés en :

- Le nombre des victimes ;
- Le nombre d'accidents où figurent ces victimes. Ce nombre n'est en réalité que la formule SUM(COUNT(DISTINCT Accident_Reportnum)), de la colonne dégrénerée.

e. Le modèle en étoile:¹⁷

f. Les agrégats retenus :

Dimension	Agrégats	Agrégats possibles	Aggrégats retenus
Time_Dim	Année, trimestre, mois, Semaine, jour	5	4
Heure_Dim	Heure	1	1
Type_Accident_dim	Type accident	1	1
Nature_Accident_dim	Nature accident	1	1
Cause	Cause, type cause	2	1
Roadway_surface	Défaut de la route	1	1
Weater	Conditions météorologiques	1	1
Traffic level	Densité de la circulation	1	1
Traffic flow	Orientation de la circulation	1	1
Road special feature	Spécificité de la route	1	1
Road character	Caractéristiques de la route	1	1
Road condtion	Conditions de la route	1	1
Illumination	Luminosité	1	1
Commune	Wilaya, Daira, commune, code commune	4	3
Construction	Construction sur la route	1	1
Type collision	Type collision	1	1
Type accident	Type accident	1	1
Nationalité	Nationalité	1	1
Sexe	Sexe	1	1
Tranche d'âge	Tranche d'âge	1	1
Nature Victime	Nature Victime	1	1
Type Victime	Type Victime	1	1

Tableau 4.1.34 : Tableau descriptif des agrégats utiles du modèle « victime».

¹⁷ Voir Annexe 01 : Figure 01.4

Le tableau précédent, les besoins recensés ainsi que les accès fréquents suggèrent les agrégats les plus pertinents et incitent à la création des modèles d'agrégats suivants :

- Nombre des victimes et les accidents correspondant par Gravité des accidents, nationalité, sexe, tranche d'âge, type et nature des victime ventilés par années, mois, jours, et heures ;
- Nombre des victimes et les accidents correspondant par nature d'accident , nationalité, sexe, tranche d'âge, nature et type des victimes ventilés par années, mois, jours, et heures.

4.1.2.5. Volet « Véhicule »

a. Présentation du volet « Véhicule » :

Le volet « Véhicule » concerne tous véhicule motorisé ou pas et qui a été impliqué dans un accident de la route.

b. Granularité :

Le grain le plus fin correspond à :

La possibilité de savoir le nombre de véhicules impliqués dans combien d'accidents routiers. Ces mesures seront ventilées sur deux axes principaux : caractéristiques liées à l'accident et celles au véhicule :

- Pour les accidents : type de collision, construction sur la route, commune, luminosité, caractéristiques de la route, état de la route, type de la route, direction de la circulation, densité de la circulation, type de l'accident, nature de l'accident, météo, défaut de la route et enfin les ventiler par les données temporelles ;
- Pour les véhicules : Style, Type, Marque du véhicule ainsi que le degré du dommage induit à celui-ci.

Ce besoin d'avoir le nombre d'accidents induits causés par les véhicules (répondant aux valeurs des dimensions propres aux véhicules) nécessitera un traitement spécial, comme vu avec le volet « Chauffeurs et infractions » à cause de ce qui est connu sous l'expression de dimensions multivaluées ou relation Many-to-Many entre la table des faits et les dimensions représentant les victimes.

c. Les dimensions participantes du modèle :

Les dimensions communes :

Les utilisateurs auront besoin de ventiler les nombres des véhicules selon les spécificités de l'accident. Donc toutes les dimensions listées dans le volet « Accident » seront présentes dans ce volet, en y ajoutant le « Accident_Reportnum » comme une dimension dégénérée.

Il n'est pas nécessaire de les reciter dans cette partie. Il sera question d'introduire, uniquement, les nouvelles dimensions.

Chapitre Quatrième : Conception du Data Warehouse.

Dimension	Volet Accident	Volet Chauffeurs	Volet Infraction	Volet Victime	Volet Véhicule
Temps	✓	✓	✓	✓	✓
Heure	✓	✓	✓	✓	✓
Type Accident	✓	✓	✓	✓	✓
Type de Collision	✓	✓	✓	✓	✓
Construction	✓	✓	✓	✓	✓
Commune	✓	✓	✓	✓	✓
Luminosité	✓	✓	✓	✓	✓
Caractéristiques de la route	✓	✓	✓	✓	✓
Condition de la route	✓	✓	✓	✓	✓
Spécificités de la route	✓	✓	✓	✓	✓
Orientations du trafic routier	✓	✓	✓	✓	✓
Densité du Trafic	✓	✓	✓	✓	✓
Nature Accident	✓	✓	✓	✓	✓
Météo	✓	✓	✓	✓	✓
Défauts de la route	✓	✓	✓	✓	✓
Cause de l' Accident	✓	✓	✓	✓	✓
Accident_Reportnum (Dimension dégénérée)		✓	✓	✓	✓
Nationalité		✓		✓	
Tranche d'âge		✓		✓	
Sexe		✓		✓	
Niveau d'éducation		✓		✓	
Propriétaire		✓			
Port Permis de Conduire		✓			
Type Chauffeur		✓			
Infraction			✓		
Nature_Victime				✓	
Type_Victime				✓	
Degré d'endommagement					✓
Type Véhicule					✓
Marque du véhicule					✓

Tableau 4.1.35 : Détection des dimensions communes entre les volets « Accident », « Chauffeur », « infraction », « victimes » et véhicules

1. Type du véhicule (Vehicle_Type_dim) :

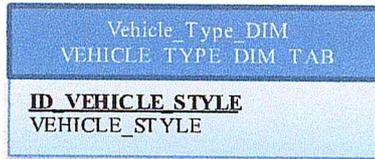


Figure 4.1.32 : La dimension «Type du véhicule » du volet « VEHICULE»

Cette dimension décrit le type du véhicule: Tracteur, Motocyclette, Bus ;

Désignation	Détail
ID_VEHICLE_TYPE	Clé de substitution
VEHICLE_TYPE	Libellé du style du véhicule

Tableau 4.1.36 : Tableau descriptif de la dimension « type du véhicule » du volet « VEHICULE»

2. Degré d'endommagement (Vehicle_Damage_dim) :

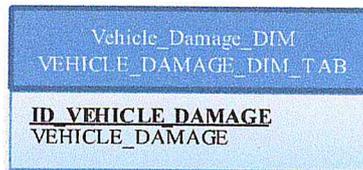


Figure 4.1.33 : La dimension «Degrés d'endommagement » du volet « VEHICULE»

Cette dimension décrit le niveau d'endommagement subit par le véhicule due à l'accident: Sans endommagement, Endommagement sérieux ou Destruction totale ;

Désignation	Détail
ID_VEHICLE_DAMAGE	Clé de substitution.
VEHICLE_DAMAGE	Libellé du niveau d'endommagement du véhicule.

Tableau 4.1.37 : Tableau descriptif de la dimension « Degrés d'endommagement » du volet « VEHICULE»

3. Marque du véhicule (Vehicle_Make_dim) :

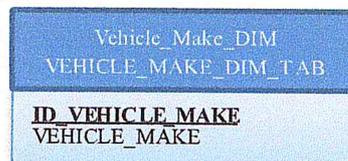


Figure 4.1.34 : La dimension «Marque du véhicule » du volet « VEHICULE»

Cette dimension décrit la marque du véhicule impliqué dans l'accident: Volks-Wagon, Toyota, Mercedes ;

Désignation	Détail
ID_VEHICLE_MAKE	Clé de substitution.
HICLE_MAKE	Intitulé de la marque du véhicule

Tableau 4.1.38 : Tableau descriptif de la dimension « Marque du véhicule » du volet « VEHICULE »

d. Les mesurables :

Les mesurables de la table des faits « VEHICULE » peuvent être résumés en :

- Le nombre de véhicules impliqués dans l'accident ;
- Le nombre d'accidents liant ces véhicules. Ce nombre n'est en réalité que la formule SUM(COUNT(DISTINCT Accident_Reportnum)), de la colonne dégénérée.

e. Le modèle en étoile:¹⁸

f. Les agrégats :

Dimension	agrégats	Agrégats possibles	Agrégats retenus
Time_Dim	Année, trimestre, mois, Semaine, jour	5	4
Heure_Dim	Heure	1	1
Type_Accident_dim	Type accident	1	1
Nature_Accident_dim	Nature accident	1	1
Cause	Cause, type cause	2	1
Roadway_surface	Défaut de la route	1	1
Weater	Conditions météorologiques	1	1
Traffic level	Densité de la circulation	1	1
Traffic flow	Orientation de la circulation	1	1
Road special feature	Spécificité de la route	1	1
Road character	Caractéristiques de la route	1	1
Road condtion	Conditions de la route	1	1
Illumination	Luminosité	1	1
Commune	Wilaya, Daira, commune, code commune	4	3
Construction	Construction sur la route	1	1
Type collision	Type collision	1	1

¹⁸ Voir Annexe 01 : Figure A01.5

Type accident	Type accident	1	1
Type du véhicule	Vehicle_type	1	1
Degrés d'endommagement	Vehicle_damage	1	1
Marque du véhicule	Vehicle_make	1	1

Tableau 4.1.39 : Tableau descriptif des agrégats utiles du modèle « Véhicule ».

Le tableau précédent, les besoins recensés ainsi que les accès fréquents suggèrent les agrégats les plus pertinents et incitent à la création des modèles d'agrégats suivants :

- Nombre de véhicules et les accidents correspondant par Gravité des accidents, type du véhicule, marque du véhicule et degré d'endommagement ventilés par années, mois, jours, et heures ;
- Nombre de véhicules et les accidents correspondant par nature d'accident type du véhicule, marque du véhicule et degré d'endommagement ventilés par années, mois, jours, et heures.

4.1.3. Conclusion :

La zone d'entreposage constitue la zone exploitable par les utilisateurs. La modélisation de cette zone a été faite grâce à la modélisation dimensionnelle. Cette manière de représenter les données offrira aux utilisateurs des modèles intuitifs et compréhensibles permettant de naviguer et de manipuler les données, détaillées ou agrégées, sans difficulté afin de satisfaire leurs besoins en analyse.

Une fois la modélisation du modèle en étoile de l'entrepôt achevée, il est temps de passer à la conception de la zone d'alimentation. Ça sera l'objet, même, de la section suivante.

4.2. Conception de la zone « Alimentation » :

4.2.1. Introduction :

L'ETL (Extract, Transform and Load), ou le processus Extraction, Transformation et chargement des données dans le data warehouse est une étape très importante et critique dans le projet décisionnel. Elle est la phase représentant la charge de travail maximale. Elle a pour but, comme il a été expliqué dans la littérature, de s'assurer de l'acheminement des données depuis leurs sources opérationnelles jusqu'au data warehouse en passant par les transformations, vérifications et nettoyages nécessaires.

La conception de l'ETL passe par les points suivants :

- Etude et planification ;
- Choix de l'architecture ;
- Conception des processus de chargement :
 - Processus de chargement des tables de dimension ;
 - Processus de chargement des tables de faits ;

- Processus de chargement de la table temps.

4.2.2. Etude et planification :

Durant cette étape, qui est le point d'entrée au processus ETL, il sera question de :

- Etudier les sources de données ;
- Détecter les emplacements des données sources ;
- Définir la périodicité du chargement ;

4.2.2.1. Les sources de données :

Le data warehouse, en cours, puise ses données principalement de l'application Records Management System (RMS) avec ces 38 modules, mais, essentiellement, des modules suivants :

- Module Accidents de la route ;
- Module Contraventions liées aux code de la route ;
- Module Fichier Principal des véhicules ;
- Module Fichier Principal des Personnes ;
- Module Fichier Principal des Adresses ;

Ces modules sont renseignés quotidiennement par les unités territoriales de la Gendarmerie à l'échelle nationale.

Le RMS est une application nationale avec une base de données centralisée, ce qui n'était pas toujours le cas.

Le RMS consistait en onze bases de données indépendantes, mais ayant le même modèle, à raison d'une base de données par groupement territoriale, uniquement, au niveau du 1^{er} Commandements Régional de la Gendarmerie Nationale. Les utilisateurs de chaque groupement avaient accès à une base de données différente. Ces onze bases de données étaient fédérées par un moteur de recherche.

Les groupements territoriaux, des cinq commandements régionaux restants, utilisaient une application centralisée au niveau nationale. Cette application va laisser la place au RMS.

Cependant, le passage vers un RMS national, a contraint les développeurs à importer les données depuis les onze bases de données vers la base de données centrale. L'importation des données depuis l'application des cinq autres régionaux n'a pas été encore entamée.

Cependant, le RMS est en cours d'exploitation depuis 2013, qui répond amplement aux besoins des décideurs, qui font de la comparaison sur les trois dernières années.

L'ETL à mettre en place, utilisera principalement les modules du RMS national, au nombre cinq, avec leurs tables :

- Module Accidents de la route ;
- Module Contraventions liées aux code de la route ;
- Module Fichier Principal des véhicules ;

- Module Fichier Principal des Personnes ;
- Module Fichier Principal des Adresses.

En revanche, et afin de pallier à toute insuffisances dues au processus de migration vers la base central, le processus ETL prendra en considération les onze bases de données dans l'objectif de :

- N'omettre aucun enregistrement, éventuellement, perdu lors de la migration ou leurs absences due à la politique d'archivage des données ;
- Compléter les données en cas d'insuffisances ou de manque dans les champs (uniquement pour les enregistrements importés) ;

4.2.2.2. Détection des emplacements des données sources :

Afin de définir l'emplacement des informations dans les différents systèmes sources et d'en choisir les emplacements les plus fiables, les DBA, développeurs ainsi que les utilisateurs ont été impliqués dans la tâche.

Le nombre important de tables, la redondance des données et l'intervention de différents systèmes, rendent cette tâche très ardue. Afin de mener à bien cette détection, il est judicieux de :

- Lister les données nécessaires à partir des étoiles conçues ;
- Lister les emplacements de chaque donnée ;
- Choisir la source la plus fiable et la valider comme source de chargement ;
- Dresser un tableau¹⁹ qui établit le lien entre données sources et donnée cibles avec les transformations nécessaires.

Cette étape s'achève par l'élaboration d'une carte logique entre données sources et données cibles²⁰.

Il est important de valider les sources de données (donc le tableau cité précédemment), afin de vérifier leurs intégrités et leurs fiabilités.

4.2.2.3. Définition de la périodicité de chargement

La périodicité de chargement est étudiée pour chaque étoile séparément, ce qui n'empêche pas une synchronisation dans les chargements des dimensions communes.

Avant de décider de la périodicité du chargement, les contraintes suivantes doivent être prises considération :

- La quantité de données à charger ;
- Le temps de non activité des systèmes sources.

L'idéal est de procéder à des chargements quotidiens, pendant les heures de non activité des systèmes sources, c'est-à-dire, durant la période qui s'étend entre dix-huit heures et huit heures du matin.

¹⁹ Ce tableau est décrit dans le livre [Kimball 2004].

²⁰ Voir Annexe A10

Pourquoi pas des chargements hebdomadaires ?

Même si pendant les week-ends le nombre d'heures de non activité des systèmes est plus important, la quantité de données à charger sera, elle, plus conséquente, et les erreurs de chargements plus difficiles à gérer, et affectera les performances du processus de chargement. Par ailleurs, la volatilité de certaines données nous contraint à appliquer une telle politique de chargement. Il faut dire que, dans le domaine des accidents de la route, les weekends comptent souvent les plus grands nombre surtout dans la période estivale.

4.2.3. Architecture du processus d'alimentation

Il est essentiel de mettre en place une architecture du processus de l'alimentation « ETL », en amont du projet, car l'architecture affectera toutes ses composantes. La bonne architecture permet de minimiser les répercussions sur l'ensemble du projet lors des éventuels changements et mise à jour dans l'ETL.

L'architecture à mettre en place devra prendre en charge toutes les contraintes auxquelles on doit faire face.

Plusieurs architectures du processus de l'ETL sont envisageables, néanmoins, il a été décidé d'opter pour la méthode «*Push-Pull* ». En plus des avantages qu'elle présente, cette méthode permettra d'exploiter et de tirer profit des avantages offerts par le système d'information de la Gendarmerie Nationale et permet d'exploiter toutes les opportunités offertes:

- L'accès aux systèmes source est garantie grâce à la centralisation des bases de données;
- Réseau disponible et sûr.

Le «*Pull*» sera utilisé avec les enregistrements nouvellement créés, cela se passe par l'extraction des enregistrements créés après le dernier chargement.

Et afin de veiller à garder une performance acceptable du système, la méthode push sera utilisée pour les enregistrements déjà chargés dans le data warehouse mais qui ont fait l'objet de mise à jour. Les données de l'accident ne changent pas, à priori, après leurs saisies dans la base de données, mais dans le cas où ça arrive, cette méthode (Push) est plus appropriée car elle évite au processus ETL de faire la vérification.

4.2.4. Processus de chargement

Le diagramme d'activités suivant décrit le processus général de l'alimentation du data warehouse dès sa mise en service :

Chapitre Quatrième : Conception du Data Warehouse.

TRANCHE_AGE_DIM_TAB	• TRANCHE_AGE		N1→ALL
SEX_DIM_DIM	• ID_SEX	N1	Hierarchie_1 :
SEX_DIM_TAB	• SEX		N1→ALL
EDUCATION_DIM	• ID_EDUCATION	N1	Hierarchie_1 :
EDUCATION_DIM_TAB	• EDUCATION		N1→ALL
PROPRIETAIRE_DIM	• ID_PROPRIETAIRE	N1	Hierarchie_1 :
PROPRIETAIRE_DIM_TAB	• PROPRIETAIRE		N1→ALL
PERMIS_ON_DIM	• ID_PERMIS_ON	N1	Hierarchie_1 :
PERMIS_ON_DIM_TAB	• PERMIS_ON		N1→ALL
CITYZENSHIP_DIM	• ID_CITYZENSHIP	N1	Hierarchie_1 :
CITYZENSHIP_DIM_TAB	• CITYZENSHIP		N1→ALL
INFRACTION_DIM	• ID_QUALIFICATION	N1	Hierarchie_1 :
INFRACTION_DIM_TAB	• QUALIFICATION		N1→N2→N3 →ALL
	• ID_SECTION	N2	Hierarchie_2 :
	• SECTION		N1 →N3 →ALL
	• ID_INFRACTION	N3	
	• INFRACTION		
NATURE_VICTIME_DIM	• ID_NATURE_VICTIME	N1	Hierarchie_1 :
NATURE_VICTIME_DIM_TAB	• NATURE_VICTIME		N1→ALL
TYPE_VICTIME_DIM	• ID_TYPE_VICTIME	N1	Hierarchie_1 :
TYPE_VICTIME_DIM_TAB	• TYPE_VICTIME		N1→ALL
Degré d'endommagement	• ID_Vehicule_DAMAGE	N1	Hierarchie_1 :
	• VEHICULE_DAMAGE		N1→ALL
Type Véhicule	• ID_VEHICULE_TYPE	N1	Hierarchie_1 :
	• VEHICULE_TYPE		N1→ALL
Marque du véhicule	• ID_VEHICULE_MAKE	N1	Hierarchie_1 :
	• VEHICULE_MAKE		N1→ALL:

Tableau 4.3.1 : Liste des hiérarchies des dimensions

- « **All** » : représente le niveau le plus agrégé du fait qui est le niveau le plus haut dans une hiérarchie données. Il n'apparaît pas systématiquement dans toutes les hiérarchies. Exemple : Commune→Daïra→Wilaya→ALL ;
- Les Ni représentent les niveaux, tels que N0 c'est le niveau le plus fin.

4.3.3. La liste des cubes :

A ce stade, il est possible de définir la liste des cubes à mettre en place. Il sera question d'y lister les dimensions incluses et les mesurables présents.

Chapitre Quatrième : Conception du Data Warehouse.

Nom du Cube	Les mesurables	Les dimensions	Les Hiérarchies
Accident	<ul style="list-style-type: none"> • Nombre des accidents ; 	Temps_DIM	Hiérarchie_1
Accient_Header_Cube	<ul style="list-style-type: none"> • Nombre des décès ; • Nombre des blessés ; • Nombre des Témoins oculaires 	TYPE_ACCIDENT_DIM	Hiérarchie_1
		TYPE_COLLISION_DIM	Hiérarchie_1
		CONSTRUCTION_DIM	Hiérarchie_1
		COMMUNE_DIM	Hiérarchie_1
		LUMINOSITE_DIM	Hiérarchie_1
		Caractéristiques_Route_DIM	Hiérarchie_1
		Conditions_Route_DIM	Hiérarchie_1
		Spécificités_Route_DIM	Hiérarchie_1
		Orientations_Trafic_DIM	Hiérarchie_1
		Densité_Trafic_DIM	Hiérarchie_1
		Nature_Accident_DIM	Hiérarchie_1
		Météo_DIM	Hiérarchie_1
		Défauts_Route_DIM	Hiérarchie_1
		CAUSE_DIM	Hiérarchie_1
Chauffeurs	<ul style="list-style-type: none"> • Nombre des chauffeurs ; 	Time_DIM	Hiérarchie_1
Accident_Chauffeur_Cube	<ul style="list-style-type: none"> • Nombre des accidents correspondants aux chauffeurs 	TYPE_ACCIDENT_DIM	Hiérarchie_1
		TYPE_COLLISION_DIM	Hiérarchie_1
		CONSTRUCTION_DIM	Hiérarchie_1
		COMMUNE_DIM	Hiérarchie_1
		LUMINOSITE_DIM	Hiérarchie_1
		Caractéristiques_Route_DIM	Hiérarchie_1
		Conditions_Route_DIM	Hiérarchie_1
		Spécificités_Route_DIM	Hiérarchie_1
		Orientations_Trafic_DIM	Hiérarchie_1
		Densité_Trafic_DIM	Hiérarchie_1
		Nature_Accident_DIM	Hiérarchie_1
		Météo_DIM	Hiérarchie_1
		Défauts_Route_DIM	Hiérarchie_1
		CAUSE_DIM	Hiérarchie_1
		ACCIDENT_REPORTNUM_DIM	NA
		Dégénérée	
		TRANCHE_AGE_DIM	Hiérarchie_1
		SEX_DIM_DIM	Hiérarchie_1
		EDUCATION_DIM	Hiérarchie_1
		CITYZENSHIP_DIM	Hiérarchie_1
		TYPE_CHAUFFEUR_DIM	Hiérarchie_1
		PROPRIETAIRE_DIM	Hiérarchie_1
		PERMIS_ON_DIM	Hiérarchie_1

Chapitre Quatrième : Conception du Data Warehouse.

Victimes	• Nombre des victimes ;	Temps_DIM	Hiérarchie_1
Accident_Victime_Cube	• Nombre des accidents correspondants aux victimes	TYPE_ACCIDENT_DIM	Hiérarchie_1
		TYPE_COLLISION_DIM	Hiérarchie_1
		CONSTRUCTION_DIM	Hiérarchie_1
		COMMUNE_DIM	Hiérarchie_1
		LUMINOSITE_DIM	Hiérarchie_1
		Caractéristiques_Route_DIM	Hiérarchie_1
		Conditions_Route_DIM	Hiérarchie_1
		Spécificités_Route_DIM	Hiérarchie_1
		Orientations_Trafic_DIM	Hiérarchie_1
		Densité_Trafic_DIM	Hiérarchie_1
		Nature_Accident_DIM	Hiérarchie_1
		Météo_DIM	Hiérarchie_1
		Défauts_Route_DIM	Hiérarchie_1
		CAUSE_DIM	Hiérarchie_1
		ACCIDENT_REPORTNUM_DIM	NA
		Dégénérée	
		NATURE_VICTIME_DIM	Hiérarchie_1
		TYPE_VICTIME_DIM	Hiérarchie_1
		Véhicules	• Nombre des véhicules ;
Accident_Vehicule_Cube	• Nombre des accidents correspondants aux véhicules	TYPE_ACCIDENT_DIM	Hiérarchie_1
		TYPE_COLLISION_DIM	Hiérarchie_1
		CONSTRUCTION_DIM	Hiérarchie_1
		COMMUNE_DIM	Hiérarchie_1
		LUMINOSITE_DIM	Hiérarchie_1
		Caractéristiques_Route_DIM	Hiérarchie_1
		Conditions_Route_DIM	Hiérarchie_1
		Spécificités_Route_DIM	Hiérarchie_1
		Orientations_Trafic_DIM	Hiérarchie_1
		Densité_Trafic_DIM	Hiérarchie_1
		Nature_Accident_DIM	Hiérarchie_1
		Météo_DIM	Hiérarchie_1
		Défauts_Route_DIM	Hiérarchie_1
		CAUSE_DIM	Hiérarchie_1
		ACCIDENT_REPORTNUM_DIM	Hiérarchie_1
		Dégénérée	
		Vehicle_Type_dim	Hiérarchie_1
		Vehicle_Make_dim	Hiérarchie_1
		Infraction	• Nombre d'occurrences des infractions ;
Infraction_Cube		TYPE_ACCIDENT_DIM	Hiérarchie_1

<ul style="list-style-type: none"> • Nombre des accidents impliquant ces infractions 	TYPE_COLLISION_DIM	Hiérarchie_1
	CONSTRUCTION_DIM	Hiérarchie_1
	COMMUNE_DIM	Hiérarchie_1
	LUMINOSITE_DIM	Hiérarchie_1
	Caractéristiques_Route_DIM	Hiérarchie_1
	Conditions_Route_DIM	Hiérarchie_1
	Spécificités_Route_DIM	Hiérarchie_1
	Orientations_Trafic_DIM	Hiérarchie_1
	Densité_Trafic_DIM	Hiérarchie_1
	Nature_Accident_DIM	Hiérarchie_1
	Météo_DIM	Hiérarchie_1
	Défauts_Route_DIM	Hiérarchie_1
	CAUSE_DIM	Hiérarchie_1
	ACCIDENT_REPORTNUM_DIM	NA
	Dégénérée	
INFRACTION_DIM	Hiérarchie_1 Hiérarchie_2	

Causes
Causes_Cube

Tableau 4.3.2 : Liste des cubes avec leurs dimensions et hiérarchies

4.3.4. Conclusion :

La bonne conception des cubes dimensionnels est un facteur clé de succès dans le projet Data Warehouse.

Elle permettra, par la suite, une bonne exploitation des données contenues dans l'entrepôt d'une façon fluide et naturelle.

Ce chapitre a été consacré pour la définition des cubes relatifs au projet, les dimensions, les différentes hiérarchies présentes ainsi que les niveaux qui composent ces dernières.

Chapitre Cinquième

L'implémentation des DataMarts et du BI



« La civilisation n'est pas

Un entassement, mais une construction, une architecture. »

Malek Bennabi

1. Introduction ;
2. Implémentation des DataMarts ;
3. Validation, création et Déploiement des trois cubes ;
4. Implémentation du Business Intelligence (BI) ;
5. Les interfaces de la restitution ;
6. Conclusion.

5. L'implémentation des DataMarts et du Business Intelligence (BI) :

5.1. Introduction :

Cette partie du mémoire traite de l'implémentation des magasins de données (DataMarts) ou volets, conçus dans le « chapitre quatrième », cela se fera à travers la création des dimensions de chaque cube et leurs déploiements.

Il sera question de passer à l'implantation du Business Intelligence (B.I.) pour la construction des analyses, graphes, filtres et la représentation de ces analyses sur la carte géographique.

Pour cette solution, les outils oracle suivants ont été choisis :

- Le SGBD Oracle 10g Release2;
- Oracle Warehouse Builder (OWB);
- Oracle BI Administration (OBIA);
- Oracle BI Enterprise Edition (OBIEE).

Ce choix peut être justifié pour les raisons suivantes :

- Puissance des produits Oracle ;
- Possibilités offertes par Oracle ;
- Compatibilité d'Oracle avec les sources de données OLTP des utilisateurs;
- La Gendarmerie Nationale dispose de ces outils d'où la nécessité de rentabiliser l'investissement.

5.2. Implémentation des DataMarts :

5.2.1. DataMart Accident:

La première étape de création du data mart « Accident », est la création des dimensions, leurs mappings avec les sources de données, et leurs déploiements et chargement.

L'étape suivante est la création du cube de données, définition des mesures, son mapping, son déploiement et son chargement.

Ce cube ne contient pas de relations dimensionnelles multivaluées, et donc ne sera pas sujet de dégénération de « Accident_Reportnum ».

5.2.1.1. Création des dimensions :

Dans cette section, nous allons présenter le processus de création d'une seule dimension, à titre illustratif, la création des autres dimensions a été effectuée de la même manière.

Dimension « COMMUNE » :

La dimension « commune » est composée des niveaux Wilaya, Daïra et Commune et code commune comme.

On saisit le nom de la dimension, en respectant une règle unifiée pour nommer les dimensions du data warehouse qui est le nom de la dimension suivie de « _dim » : **Commune_dim**.²¹

On saisit les informations relatives à la dimension (champs de la dimension). Oracle Warehouse Builder (OWB) propose, pour chaque niveau, trois attributs : ID, NAME et DESCRIPTION, qui sont généralement amplement suffisants²².

Il faudra, néanmoins, augmenter la taille des champs afin d'éviter les erreurs, dues à l'insuffisance de la taille des champs de destination, lors de l'exécution du mapping.²³

A noter que pour chaque dimension, nous aurons une clé primaire de la dimension qui sera générée automatiquement de type surrogate, (ID dans notre cas), puis viennent les champs de type business qui sont les clés étrangères des sources de données (applications métier).

Lors de l'utilisation de l'assistant de création des dimensions de OWB, celui-ci crée une seule hiérarchie (La hiérarchie par défaut), dans laquelle on spécifie les différents niveaux. Cependant, si on veut créer plusieurs hiérarchies, il faudra passer par le mode manuel

Il faudra, ensuite, passer à la définition des attributs comme illustré dans *l'annexe 02, figure A02.4 : « Spécification des niveaux »*.

Ensuite, nous aurons à sélectionner les attributs de chaque niveau tel que définit dans *l'annexe 02, figure A02.5 : Spécification des attributs des niveaux*.

A l'issue de la création de la dimension, OWB crée une dimension, la table liée à cette dimension et la séquence utilisée pour la génération des « surrogate key »²⁴. On procède à la validation et le déploiement de la dimension, la table et la séquence lui étant associées, tel qu'illustré dans *l'annexe 02, figure A02.6 : Déploiement de la Commune_dim, Commune_dim_TAB et Commune_dim_SEQ*.

Après le déploiement, la table physique sera créée dans la base de données destination ainsi que la séquence.

La création du reste des dimensions se fera de manière similaire. Nous allons uniquement présenter dans les sections suivantes le mapping et le déploiement de cette dimension, et nous présenterons le résultat du déploiement.

²¹ Voir Annexe 02, figure A02.1

²² Se référer à l'annexe 02 figure A02.2 : Spécification des attributs de la dimension

²³ Voir l'annexe 02 figure A02.3 : Augmentation des tailles des attributs vers 500.

²⁴ Figure A02.6 : Autocréation des dimensions commune_dim, la table Commune_dim_TAB et la séquence Commune_dim_SEQ, Annexe 02

5.2.1.2. Mapping et déploiement des dimensions :

Après la création de la dimension, il faudra procéder à son mapping avec les sources de données correspondantes dans les bases de données sources (exemple : Table TSTATES, TDAIRA et TCOMMUNE)²⁵

On choisit d'abord nos données sources qui alimenteront la dimension.²⁶

Ensuite, on utilise un opérateur de jointure « JOINER » pour regrouper ces données, avec la spécification des conditions de jointure entre les sources de données.

Une fois les sources de données sont réunies dans le JOINER²⁷, on spécifie la clause de jointure des sources de donnée.²⁸

On procède, ensuite, au mapping de la sortie du « Joiner » avec la dimension « Commune_Dim » tel qu'illustré dans la **figure A02.11** : **« Liaison entre le JOINER et la dimension « DIM_INFRACTION », Annexe 02.**

Le schéma final du mapping de la dimension Commune_dim se présente comme dans la figure qui suit :

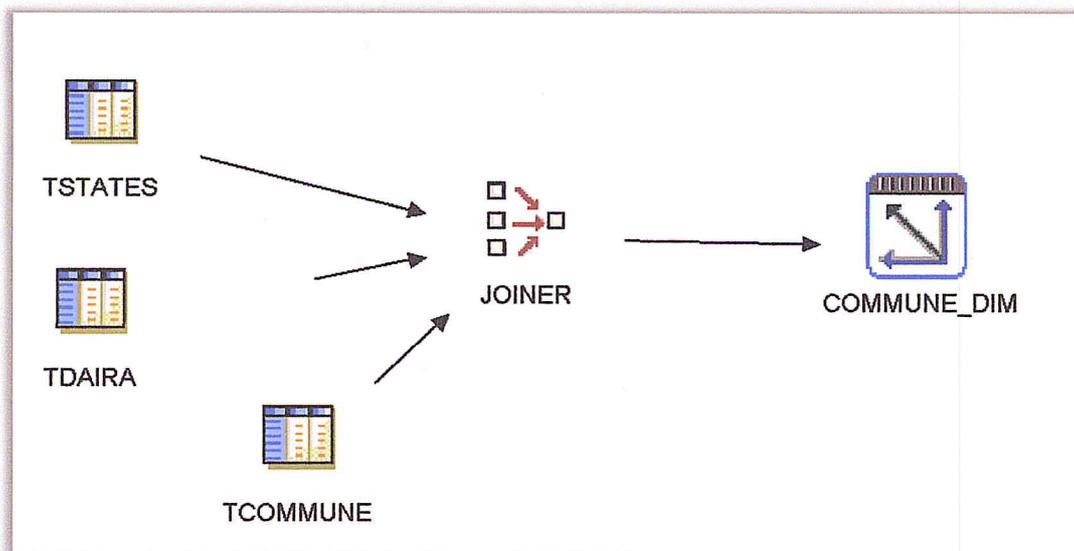


Figure 5.2.1 : Schéma final du mapping.

²⁵ figure A02.7 : L'ajout d'un nouveau mapping, Annexe 02.

²⁶ figure A02.8 : La sélection des tables source, Annexe 02.

²⁷ figure A02.9 : Regroupement des données avec l'opérateur de jointure « JOINER », Annexe 02.

²⁸ figure A02.10 : Spécification de la condition de jointure, Annexe 02.

La **figure A02.12** : « *Validation et génération des scripts PISQL du « Commune_dim_map »* », de l'*annexe 02*, illustre le lancement du processus de validation du mapping de la dimension. Une fois validé avec succès, le mapping doit être déployé vers la base de destination. Cela résultera en la création d'un « package » dans le schéma de destination avec du code PL-SQL correspondant aux composants dans le mapping.²⁹

La **figure A02.14** : « *Démarrage du « commune_dim_map »* » de l'*annexe 02*, montre la phase finale dans la création du mapping : le lancement du mapping. Cela va populer la table oracle correspondant à la dimension, dans l'exemple c'est la table « **COMMUNE_DIM_TAB** » qui sera peuplée selon la définition du mapping.

Afin de vérifier le résultat du mapping, on peut afficher les données correspondantes à la dimension comme le montrent les deux figures de l'*annexe 02* : **figure A02.15** et **figure A02.16**

5.2.1.3. Création, Mapping et déploiement du cube :

Il faudra juste veiller à sélectionner les dimensions du cube de données, puis introduire les mesures voulues, et procéder directement au mapping du cube.

Cependant, un travail de nettoyage et de compléments doit se faire :

- Epuration des données non valides ;
- Jointure avec les tables statiques afin de ramener les libellés correspondant aux codes dans la table accident ;
- Effectuer des expressions de transformations sur les colonnes comme la transformation de la date et heure et le traitement des données nulles.

A noter que lors du mapping du cube de données, on procède globalement comme lors du mapping de la dimension, sauf qu'il faudra utiliser l'opérateur d'agrégation « **AGGREGATOR** », afin d'agréger les données du cube, et de les regrouper par des clauses « **group by** », du fait que notre objectif, et de regrouper les données du cube selon les mesures définies lors de sa création.

Lors du mapping des cubes, le procédé suivant a été entrepris :

- Utilisation d'un composant « **Expression** » pour la vérification de l'état de saisi, de l'existence du code dans la plage des codes de la table statique source et enfin pour forcer la valeur à défaut en cas de données nulles ;
- Utilisation d'un composant « **Joiner** » avec la table principale « **Accident** » et les tables statiques contenant les libellés correspondant aux codes dans les tables principales ;

²⁹ *Figure A02.13* : « *Déploiement du « commune_dim_map »* », *Annexe 02*.

- Relier les sorties (Output) des « Joiners » en cours vers le « Joiner » suivant afin d'effectuer la prochaine jointure dans le but de ramener un autre libellé correspondant à un autre code de table statique.

Cela dit, on aurait pu utiliser toutes les expressions de validation, complètement et forcé à défaut des codes et les relier à un seule « Joiner », et puis faire une jointure avec toutes les tables statiques afin de ramener les libellés correspondants aux codes dans la source de données « accident ». Cela aurait donné un schéma de mapping plus concis.

D'autant plus, que les composants « Joiners » sont gourmands en matière d'espace mémoire lors de leurs exécutions.

Mais d'un autre côté, lors de la définition de la condition de jointure dans le composant « Joiner », si il est relié à « N tables », on aura une condition de jointure comportant « N-Tables » ce qui revient à une jointure gourmande en performances.

Cela reviendrait à (N-1) composants Joiners avec une condition de jointures comportant les deux tables à la fois.³⁰

Cependant, et pour les raisons suivantes, il a été décidé de ne pas économiser le nombre « Joiner » :

- La Gendarmerie Nationale, dispose de data centers très puissants fonctionnant en mode load balancing, ce qui élimine le problème de mémoire et de performance ;
- L'utilisation de « Joiner » différents pour de plus petits groupes de tables et de colonnes permet de déceler les origines des erreurs d'une façon précise et de les corriger ;
- Améliorer la lisibilité du mapping pour les administrateurs qui n'ont pas participé dans le projet ;
- Lors de l'utilisation d'un nombre minimum de « Joiner » pour plusieurs tables différentes, la condition de jointure devient complexe et peut aussi diminuer les performances du mapping ;
- Impossibilité d'effectuer plusieurs jointures externes sur la même table ;

Dans « *figure A03.1 : « Section N° 01 du mapping global du cube « Accident » » de l'annexe 03*, il sera question de préparer les données concernant les causes induisant à l'accident, pour être mappées dans le cube. Les causes représentent l'une des dimensions « *Many-to-Many* », nécessitant la dégénération du numéro de l'accident dans la table des faits. Cependant, lors de l'examen des sources de données, il s'est avéré qu'un accident, certes, peut avoir plusieurs causes, mais uniquement une d'entre elles est marquée comme étant la principale cause. C'est dans un souci d'offrir le maximum de dimensions pour un maximum d'analyse et de flexibilité, qu'il a été décidé, avec les utilisateurs, d'introduire « la cause principale » dans tous les cubes.

Pour ce faire, un filtre et une vue ont été créés au niveau de la base de données source contenant le numéro de l'accident, les causes lui étant associées avec leurs priorités adéquates :

³⁰ (N-1) composants Joiner parce que la table principale est toujours présente dans les jointures résultantes de la décomposition.

- Eliminer les doublons susceptible d'exister afin de garantir la relation one-to-one avec la table des faits ;
- Appliquer un composant « Joiner » pour ramener les libellés des causes depuis les deux tables « TYPE_CAUSE_ACCIDENT » et « CUSES_ACCIDENT » ;
- Appliquer un filtre sur la priorité principale pour éliminer toutes les causes qui ne le sont pas ;
- Utiliser le « Joiner » intitulé « JOINER_CAUSE_TYPE_CAUSE » pour faire une jointure entre les données de l'accident et celles des causes.

Une fois les transformations faites, les vérifications des valeurs nulles achevées et les opérations de jointures effectuées³¹, il faudra agréger les données résultantes du processus avant de les envoyer au cube. Pour cela, un composant « Aggregator » sera utilisé comme illustré dans **figure A03.3 : Section N° 03 du mapping global du cube « Accident » de l'annexe 03.**

La condition d'agrégation est la suivante :

GROUP BY	
NATURE_TYPE_CMG,	ROAD_CHARACTER_LBL,
WEATHER_LBL,	ROAD_SPECIAL_FEATURE_LBL,
HOURE,	DAY,
MONTH,	TYPE_TYPE_ACC,
TRAFFIC_LEVEL_LBL,	TRAFFIC_FLOW_LBL,
ROADWAY_SURFACE_LBL,	ROAD_CONDITION_LBL,
ILLUMINATION_LBL,	CONST_MAINT_ZONE_ON_LBL,
COLLISION_TYPE_LBL,	CAUSE_A_LBL,
WILAYA_LBL,	ACCIDENT_DATE,
CODE_COMMUNE	

Il faudra remarquer que dans ce cube, toutes les dimensions relatives à l'accident sont présentes, sauf la dimension dégénérée « Accident_Reportnum ». Cela étant parce que toutes les relations sont « one-to-one », et donc pas de dimensions multivaluées.

Par contre, dans les mapping des autres cubes, la condition d'agrégation illustrée précédemment sera enrichie par « Accident_Reportnum », représentant une dimension dégénérée, dans le but d'effectuer une jointure entre les données de l'accident et les agrégats des chauffeurs, véhicules, victimes ou infractions, présentant des relations dimensionnelles multivaluées.

³¹ Figure A03.2 : « Section N° 02 du mapping global du cube « Accident » », Annexe 03.

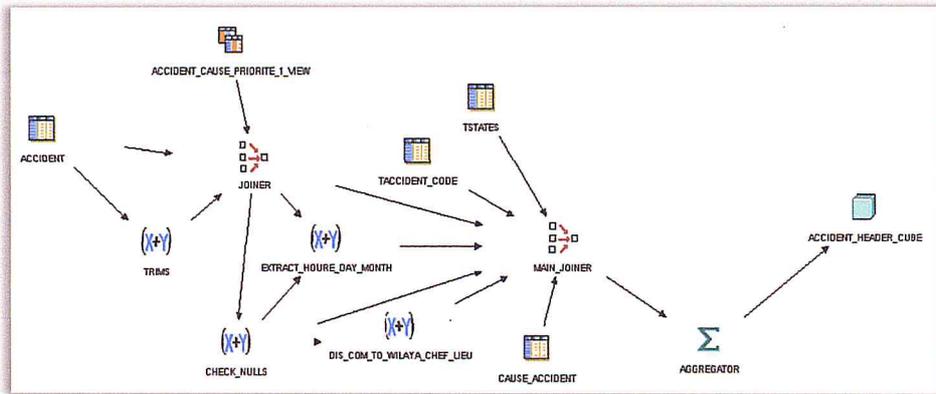


Figure 5.2.2 : Schéma dans sa vue global du mapping pour le cube « Accident ».

5.2.2. DataMart « Chauffeurs » :

On va juste montrer le mapping du cube. Pour ce qui concerne la création des dimensions de ce cube c'est la même procédure que la dimension « COMMUNE_DIM » que nous avons créé précédemment pour le datamart « Accident ».

Pour ce mapping, ainsi que pour tous les autres cubes à venir, représentant des relations dimensionnelles many-to-many avec l'accident, on procèdera selon deux parties :

- Une partie pour ramener les données relatives à l'accident. Cette partie sera identique au mapping du cube « Accident », sauf que cette fois-ci, un nouveau champ qui est « Accident_Reportnum » sera pris en considération. Ce champs représentera la dimension dégénérée et il sera utilisé pour la jointure entre les données de l'accident de la seconde partie qui est le mapping des chauffeurs promptement dit ; et donc, il n'est pas nécessaire d'illustrer cette partie, on ne représentera que l'agrégateur final de l'accident qui sera relié avec un composant « Joiner » avec l'agrégateur des Chauffeur ;

MAIN_JOINER	
#ACCIDENT	
#CAUSE	
#ROAD_CHARACTER	
#ROAD_CONDITION	
#ROADWAY_SURFACE	
#COLLISION_TYPE	
#ILLUMINATION	
#WEATHER	
#TRAFFIC_LEVEL	
#TRAFFIC_FLOW	
#ROADSPECIAL_FEATURE	
#CONSTRUCTION	
#WILAYA	
#OUTGRP1	
RECNUM_ACCIDENT_CAUSE	1, 2
TYPE_CAUSE_PRIORITE_1	1, 2
CAUSE_PRIORITE_1	1, 2
RECNUM_ACCIDENT	7, 8
ACCIDENT_REPORTNUM	1, 2
ACCIDENT_CODE	1, 2
CREATED_BY	1, 2
CREATED_ON	10, 11
HEURE_ACC	31, 32
ROAD_CHARACTER	1, 2
ROAD_CONDITION	1, 2
ROADWAY_SURFACE	1, 2
COLLISION_TYPE	1, 2
ILLUMINATION	1, 2
WEATHER	1, 2
TYPE_ACCIDENT_LBL_AR	1, 2
NATURE_ACCIDENT_LBL_AR	1, 2
TRAFFIC_LEVEL	1, 2
TRAFFIC_FLOW	1, 2

Figure 5.2.3 : Le Joiner Final des données de l'accident avec l'introduction de la dimension dégénérée « Accident_Reportnum »

- La deuxième partie traitera du mapping des données relatives aux chauffeurs. Il sera question de faire une jointure entre l'agrégateur des chauffeurs avec celui de l'agrégateur des accidents sur la base du champ « Accident_Reportnum ».

Les données de l'accident ont été incluses ici, afin de répondre à des questions comme « Quel est le nombre des chauffeurs impliqués dans des accidents de type frontal dans un jour ensoleillé ? » ;

Mais aussi pour ramener le nombre des accidents dont les chauffeurs répondant à certains critères (liés aux chauffeurs) auront causé ou impliqués. Cela nous donne une appréciation exacte sur les nombres. En effet, si on a un nombre de chauffeurs égale à 7, on pourrait facilement penser qu'on est devant une situation de plus d'un accident, mais il peut s'avérer que ces chauffeurs sont impliqués dans un seul accident, en pile, engobant 7 véhicules.

Les figures de l'annexe 04 : figure A04.1, A04.2, A04.3 montrent des prises d'écran sur le processus du mapping de ce cube.

La condition de jointure entre le Joiner des accidents et l'agrégat des chauffeurs se fait sur la dimension dégénérée « Accident_reportnum » :

ACCIDENT.ACCIDENT_REPORTNUM=CHAUFFEUR.ACCIDENT_REPORTNUM

La condition d'agrégation de l'agrégateur « FINAL » est la suivante :

GROUP BY	
INGRP1.ACCIDENT_REPORTNUM ,	INGRP1.TRANCHE_AGE_LBL,
INGRP1.TYPE_CHAUFFEUR_LBL,	INGRP1.PROPRIETAIRE_LBL,
INGRP1.PERMIS_ON_LBL,	INGRP1.EDUCATION_LBL,
INGRP1.CITYZENSHIP_LBL,	INGRP1.SEX_LBL,
INGRP1.ACCIDENT_DATE,	INGRP1.ROAD_CHARACTER_LBL,
INGRP1.ROAD_SPECIAL_FEATURE_LBL,	INGRP1.CODE_COMMUNE,
INGRP1.WEATHER_LBL,	INGRP1.TRAFFIC_LEVEL_LBL,
INGRP1.TRAFFIC_FLOW_LBL,	INGRP1.ROADWAY_SURFACE_LBL,
INGRP1.ROAD_CONDITION_LBL,	INGRP1.ILLUMINATION_LBL,
INGRP1.CONST_MAINT_ZONE_ON_LBL,	INGRP1.COLLISION_TYPE_LBL,
INGRP1.CAUSE_A_LBL,	INGRP1.MONTH,
INGRP1.HOURE,	INGRP1.JOURE,
INGRP1.TYPE_TYPE_ACC,	INGRP1.NATURE_TYPE_CMG,
INGRP1.WILAYA_LBL	

Il faudra remarquer que dans ce cube, toutes les dimensions relatives à l'accident sont présentes, plus la dimension dégénérée « Accident_Reportnum ». Cela permettra par la suite de résoudre le problème des relations dimensionnelles multivaluées, en permettant d'avoir le nombre d'accident dont les chauffeurs ont causé.

En fin, les données de l'accident « MAIN_JOINER » et celles des victimes « AGGREGATOR_VICTIMES » seront jointes avec le composant « JOINER_FINAL » sous la condition de jointure :

« ACCIDENT.ACCIDENT_REPORTNUM = VICTIME.ACCIDENT_REPORTNUM_VICTIME ».

Le tout va être agrégé et mappé avec le cube « Victime_Cube » comme le montre la figure A05.4 de l'annexe 05.

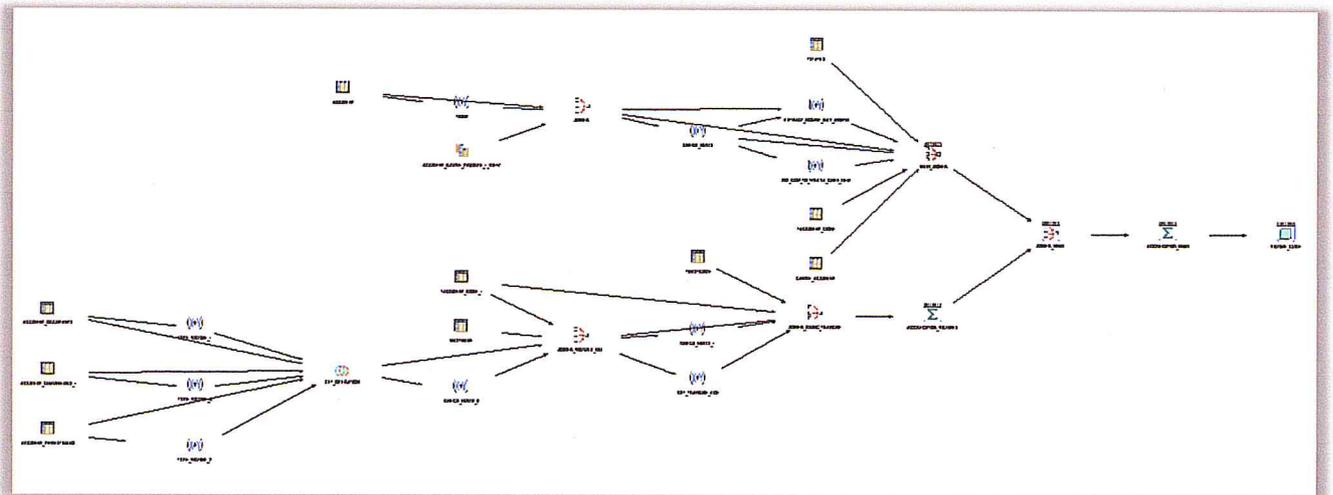


Figure 5.2.5 Schéma final du cube « Victime ».

5.2.4. Data Mart « Infractions » :

Comme pour les chauffeurs et victimes, il est question de montrer juste le mapping du cube, pour ce qui concerne la création des dimensions de ce cube c'est la même procédure que la dimension « Commune_DIM » que nous avons créé précédemment pour le datamart « Accident ».

Comme c'était le cas avec le cube « Chauffeurs » et « Victimes », il sera question de ramener les données des infractions³⁴ et de les joindre avec les données de l'accident qui sont illustrées dans « la figure 5.2.3 ».

Les données de l'accident dans le « MAIN_JOINER » et celles des infractions « AGGREGATOR » seront jointes avec le composant « JOINER_1 » sous la condition de jointure :

« ACCIDENT.ACCIDENT_REPORTNUM = INFRACTION.ACCIDENT_REPORTNUM_CITN ».

Le tout va être agrégé et mappé avec le cube « Infraction_Cube » comme le montre la figure A06.2 de l'annexe 06.

³⁴ Figure A06.1 de l'annexe 06.

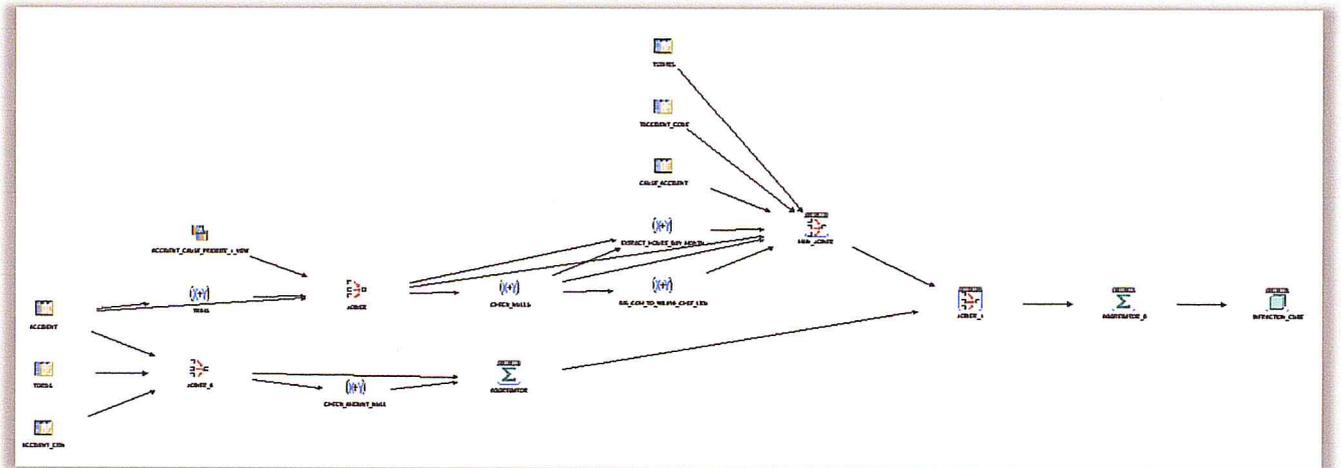


Figure 5.2.6 Schéma final du cube « Infraction ».

5.2.5. Data Mart « Véhicules » :

Il sera de même ici, à l'instar des cubes « chauffeurs », « victimes » et « infractions », il est question de montrer juste le mapping du cube, pour ce qui concerne la création des dimensions de ce cube c'est la même procédure que la dimension « *Commune_DIM* » que nous avons créé précédemment pour le datamart « *Accident* ».

Les données de l'accident vont être ramenées vers le composant « *MAIN_JOINER* » (voir la *figure 5.2.3*) avec l'attribut de la dimension dégénérée « *ACCIDENT_REPORTNUM* ». Il sera utilisé lors de la jointure avec l'agrégat des données des véhicules.

La *figure A07.1 de l'annexe 07* montre la jointure de la table contenant les liens des véhicules impliqués dans des accidents « *ACCIDENT_VEHICULE* » et le fichier principal des véhicules « *MASTAUTO* » afin de ramener toutes les données concernant les véhicules. La table « *Accident_Vehicle* » ne contient que les liens vers le fichier principal des véhicules.

Il sera question de faire des jointures avec les tables statiques afin de ramener les libellés correspondant aux codes ramenés avec la jointure avec le fichier principal des véhicules.

Une fois les libellés ramenés, on passe vers l'agrégation des données des véhicules. Enfin, il sera temps de joindre les données des accidents (*MAIN_JOINER*) avec l'agrégateur des données des véhicules (*AGGRAGATOR*) via le « *Accident_Reportnum* ». On obtient, ainsi, un regroupement des véhicules par accident ; le tout fera objet d'une autre agrégation puis la mapping avec le cube « *VEHICLE_CUBE* »³⁵

³⁵ Figure A07.2 de l'annexe 07.

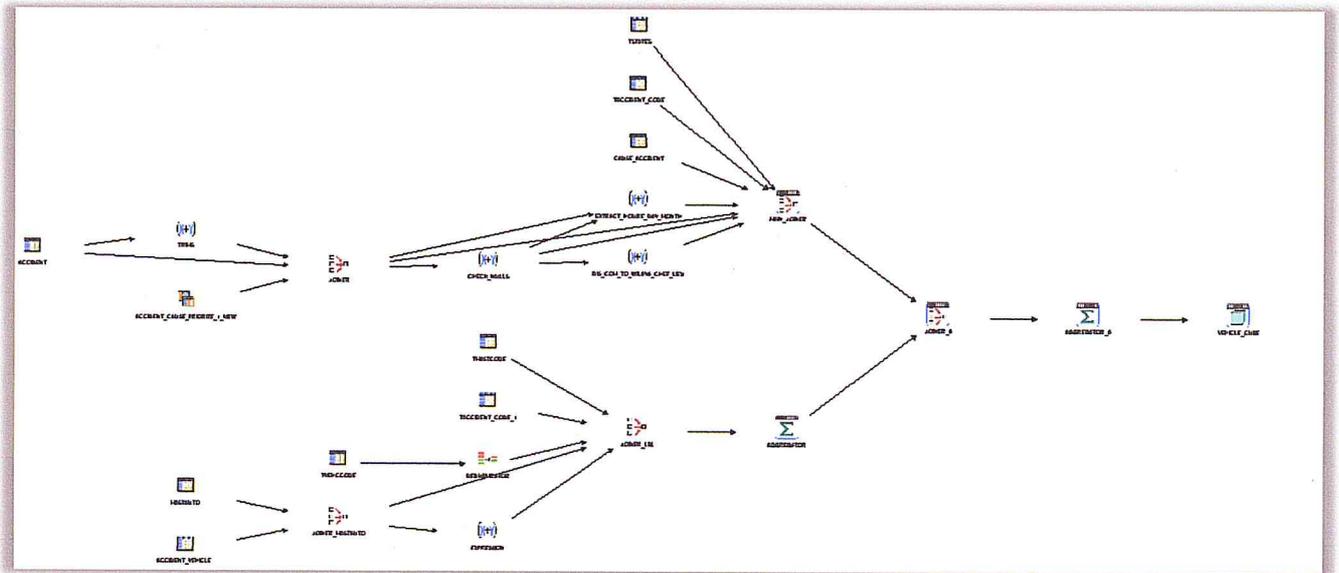


Figure 5.2.7 Schéma final du cube « Véhicule ».

5.3. Validation, création et Déploiement des trois cubes

Cela se fera à travers le « *Centre de contrôle* » du Warehouse Builder, sous l'onglet DWH_LOCATION1 pour sélectionner pour chaque cube trois composants (un mapping, une table et le cube lui-même). On sélectionnera l'action créée dans l'onglet action de déploiement comme la montre la figure ci-dessous enfin on lancera leurs créations, cette action prendra un temps considérable selon l'activité sur les systèmes OLTP mais en moyenne trente (30) minutes.

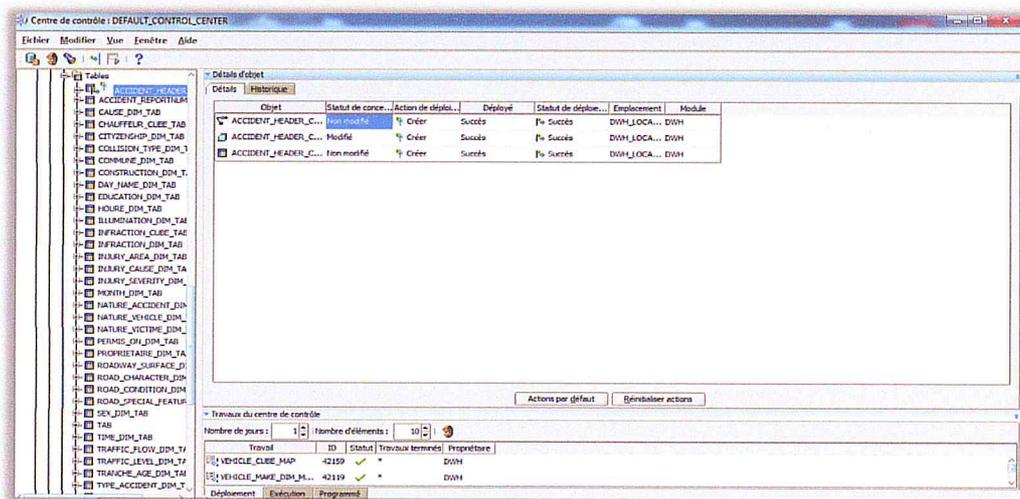


Figure 5.3.1 Déploiement des cubes.

5.4. Implémentation du BI :

Dans cette étape, nous sommes arrivés à l'implémentation des cubes qu'on a conçus précédemment et leurs restitutions aux utilisateurs. Et pour ce faire nous avons utilisé la suite des logiciels *Oracle Business intelligence Enterprise Edition 11gr2* pour la partie OLAP et la construction du tableau de bord qui sera enrichi avec des cartes géographiques pour mieux visualisé les analyses en y intégrant les données cadastrales.

5.4.1. Création de référentiel

L'outil « Administration du BI » permet l'implémentation des cubes dimensionnels à partir des dimensions et tables des faits créés sous le mode ROLAP. Avec cet outil, il est possible de définir les tables d'agrégats pour les analyses fréquemment sollicitées. Il est possible, aussi, avec cet outil de définir les hiérarchies correspondantes aux dimensions.

Pour ce faire, nous créons un nouveau référentiel avec l'outil « Administration du BI », on lui attribue un nom, un mot et la source des données. Après on passe à l'importation des différents cubes et de leurs dimensions.

L'outil administration est constitué de trois couches :

- **La couche physique** : La couche physique contient des informations (Méta Data) sur les sources de données physiques auxquelles « Oracle BI Server » soumet des requêtes. La manière la plus courante de peupler la couche physique est en important Les métadonnées depuis les bases de données et d'autres sources de données. Les sources de données peuvent être de même ou différentes variétés.³⁶
- **La couche modèle de gestion et correspondance (ou Couche Logique)** : spécifie la correspondance entre les modèles métiers et le Schémas de couches physiques. C'est là que les schémas physiques sont simplifiés pour constituer la base pour la vue des utilisateurs sur les données, en d'autre termes c'est ici que le cubes dimensionnels seront implémentés depuis les tables de faits détaillées et les dimensions depuis le ROLAP.³⁷
- **Le modèle présentation** : Ce sont les cubes et les dimensions créés au niveau du modèle logique tels que les utilisateurs vont les percevoir.³⁸

1) Spécification du type d'agrégation pour les différents indicateurs :

Cette étape vient après la création des différents modèles logiques à partir des tables du model physique, la figure suivante montre l'indication du type d'agrégation pour l'indicateur « NB_ACCIDENTS » du cube « Chauffeurs ». En effet, afin de ne pas compter l'accident plus d'une fois avec les chauffeurs correspondant, on utilisera une formule de `Sum(Count Distinct(Accident_Reportnum))`.

C'est dans cette couche que les différents agrégats, correspondant aux accès les plus fréquents, seront implémenté.³⁹

2) Création des dimensions logiques:

³⁶ Figure A08.1, Annexe 08.

³⁷ Figure A08.2, Annexe 08.

³⁸ Figure A08.3, Annexe 08.

³⁹ Figure A08.4, Annexe 08.

Pour chaque dimension contenant une hiérarchie dans le modèle physique, on va créer sa dimension logique, avec une hiérarchie de la dimension qui sera basée sur les niveaux (ici nous montrons juste la dimension logique « Temps », c'est la même chose pour les autres dimensions des différents DataMarts.⁴⁰

3) Création des niveaux de la dimension logique :

Dans cette étape, pour chaque dimension logique, on crée ses niveaux avec la spécification du nombre de lignes pour chaque niveau, voir la figure suivante :

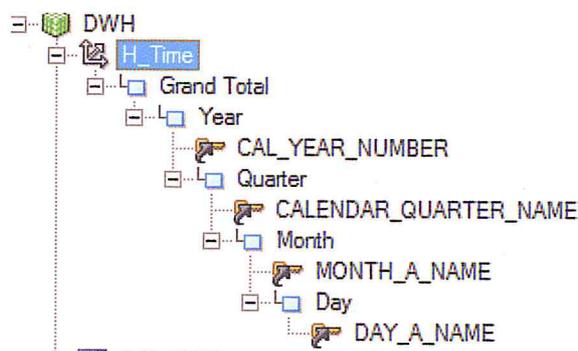


Figure 5.4.1 Création des niveaux.

5.4.2. Le chargement de référentiel dans le BI

Une fois les données des différents DataMarts sont chargées dans le référentiel, nous allons le déployer (charger) au niveau du serveur BI (OLAP)⁴¹. Cela se fait à travers le Enterprise Manager du OBIEE dans la section déploiement.

Après le chargement du référentiel il faut redémarrer tous les services du BI⁴²:

5.4.3. Système d'information géographique :

Le Data Warehouse, en cours de réalisation, devra assurer la représentation des analyses sur carte géographique. Des analyses par rapport aux wilayas et communes à travers la dimension « Commune_DIM ».

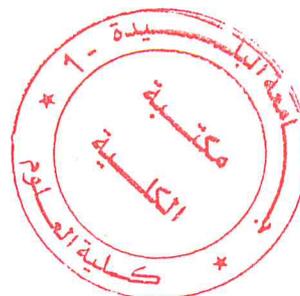
Pour assurer cette fonctionnalité, il a été procédé comme suite :

- Récupération des « Fichier de Formes » ou « Shapefiles »: « Un fichier de formes constitue un format non topologique simple permettant de stocker les informations relatives à l'emplacement géométrique et aux attributs des entités géographiques. Les entités géographiques d'un fichier de formes peuvent être représentées par des points, des lignes et des polygones (ou surfaces).

⁴⁰ Figure A08.5, Annexe 08.

⁴¹ Figure A08.6, Annexe 08.

⁴² Figure A08.7, Annexe 08.



L'espace de travail contenant des fichiers de formes peut également contenir des tables dBASE consignant des attributs supplémentaires associables par jointure aux entités de fichiers de formes. »⁴³. Les fichiers de formes correspondant aux Wilayas et Communes ont été récupérés.

- Importation des shapefiles via « Oracle MapBuilder » : les shapefiles seront importés dans un schéma supportant les données cadastrales (MVDemo). Cela aura pour effet de créer les tables : COMMUNE_ALGERIE, contenant les données cadastrales des communes, et la table LIMITES_WILAYS, contenant les données géographiques des wilayas.

Une fois les tables importées, il sera question de procéder à la création des GEOMETRY_THEMES en se basant sur les dites tables. Deux GEOMETRY_THEMES seront créés, une pour les communes⁴⁴, et l'autre pour les wilayas⁴⁵. Ces deux couches seront intégrées dans la carte nommée BASE MAP servant de base pour les analyses dans OBIEE.

La création des thèmes (couches) et de la carte de base (base map) dans le Oracle MapBuilder, a pour effet de générer de la méta données sous forme de XML qu'il faudra inclure manuellement dans le schéma MVdemo, supportant les données géographiques. Pour ce faire, deux lignes correspondant aux couches Wilaya et Communes seront ajoutées à la table « THEME »⁴⁶ et une autre ligne représentant la méta données de la carte de base (basemap) dans la table MAPS.⁴⁷

- Inclure les données des attributs des entités géographiques dans le processus ETL : Cependant, les attributs des entités géographiques devaient être assujettis au processus ETL afin de les faire correspondre aux données contenues dans la hiérarchie de la dimension commune. En effet, il fallait s'assurer que les communes et les wilayas dans la hiérarchie de la dimension Commune_DIM existent et sont épelées identiquement dans les champs des attributs des entités géographiques.
- Etablir la connexion du « Oracle Mapviewer » vers le schéma MVdemo et mettre la carte de base en ligne : « Oracle Mapviewer » est le composant responsable de la restitution des cartes et la représentation des analyses sur les cartes géographiques. La prochaine étape est donc celle de l'intégration du OBIEE avec le Oracle Mapviewer. Pour ce faire, il faudra se connecter à ce composant à travers sa console et mettre à jour le fichier XML des sources de données cadastrales en y ajoutant les coordonnées du schéma MVDEMO⁴⁸, puis sélectionner la carte de base et changer son état vers « OnLine »⁴⁹.
- Mettre la correspondance entre les attributs de données des entités géographiques avec les données de la dimension « Commune_DIM » : la dernière étape dans l'intégration et la préparation des analyses cartographiques et celles de la mise en correspondance des données des communes et des wilayas avec les données attributaires des entités géographiques.⁵⁰ Dans cette

⁴³ Documentation Online de ArcMap : <http://desktop.arcgis.com/fr/arcmap/10.3/manage-data/shapefiles/what-is-a-shapefile.htm>

⁴⁴ Voir Figure 009.1, Annexe 09

⁴⁵ Voir Figure A09.2, Annexe 09

⁴⁶ Voir Figure A09.3, Annexe 09

⁴⁷ Voir Figure A09.4, Annexe 09.

⁴⁸ Voir Figure A09.5, Annexe 09

⁴⁹ Voir Figure A09.6, Annexe 09

⁵⁰ Voir Figures A09.7 et A09.8, Annexe 09

étape, il sera question aussi de définir les niveaux du zoom de la visibilité des couches wilaya et communes.⁵¹

5.5. Les interfaces de la restitutions

5.5.1. Les différents sorties du tableau de bord réalisé

Les figures, ci-dessous, montrent quelques sorties du tableau de bord réalisé sous formes des graphes, des courbes, des tableaux croisés et une carte géographique :



Figure 5.5.1 : Première liste de prises d'écran du tableau de bord décisionnel sous version web

⁵¹ Voir Figure A09.9, Annexe 09



Figure 5.5.2 : Deuxieme liste de prises d'écran du tableau de bord décisionnel soussa version web

5.5.2. Le tableau de bord pour androïde :

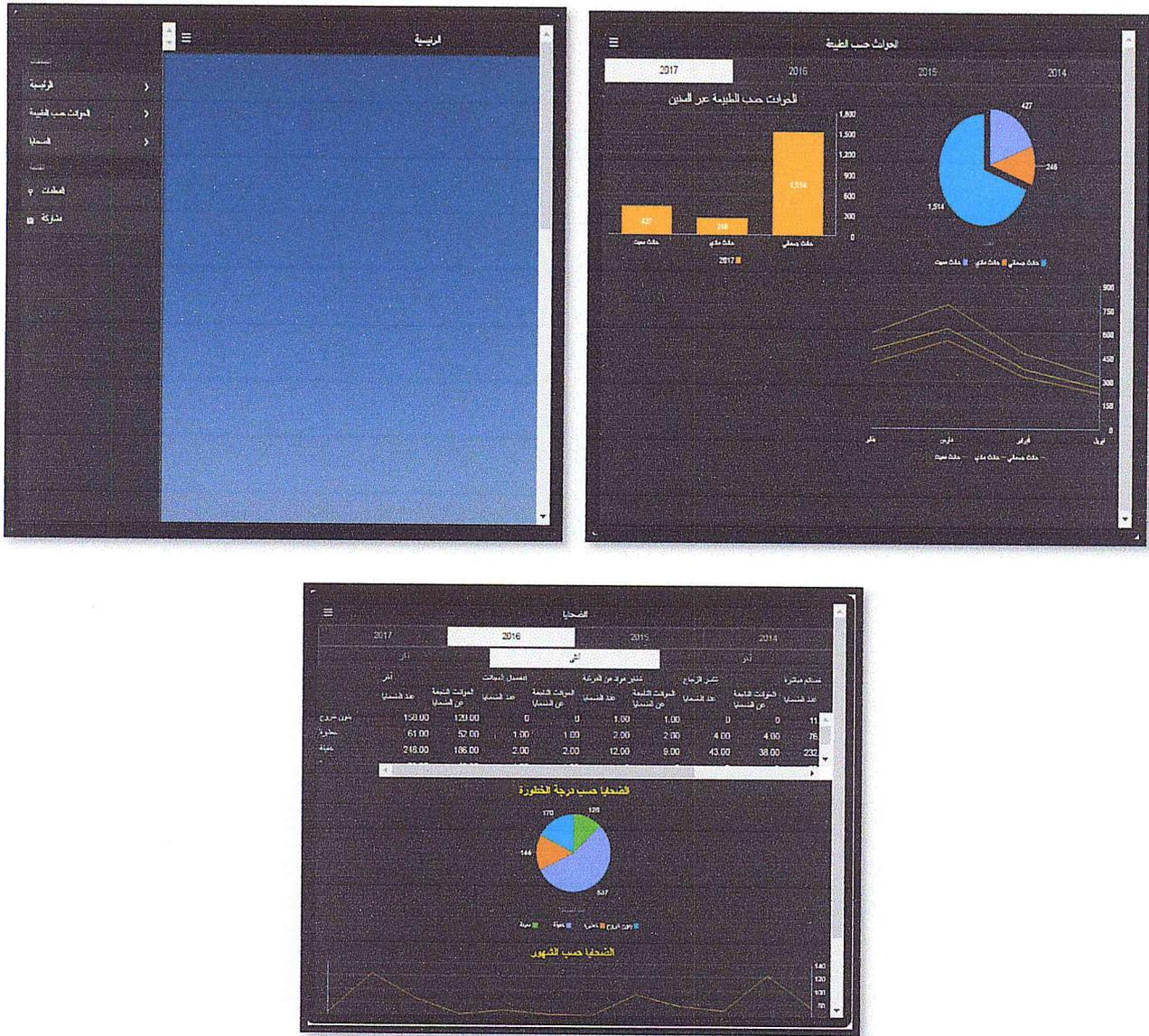


Figure 5.5.3 : Quelques prises d'écran du tableau de bord décisionnel sous version mobile

5.6. Conclusion :

Dans ce chapitre nous avons exploité les outils de travail adoptés on l'occurrence : Oracle 11g, Oracle Warehouse Builder comme ETL pour les données conventionnelles, BI admin pour la création des référentiels et Map Builder pour la création et la préparation des données spatiales, Map Viewer pour le chargement des couches et cartes. Finalement OBIEE pour faire le lien entre les données conventionnelles et spatiales et enrichir notre affichage statistique avec des cartes qui nous informent sur la localisation.

Conclusion Générale et Perspectives



« Le seul conseil en effet qu'une personne puisse donner à une autre à propos de la lecture c'est de ne demander aucun conseil, de suivre son propre instinct, d'user de sa propre raison, d'en arriver à ses propres conclusions. »

L'art du roman - Virginia Woolf

6. Conclusion Générale et perspectives:

L'accès facile et abordables aux technologies de l'information dans notre ère a eu pour effet directe l'immersion des entreprises dans des masses importantes de données brutes, souvent non ou mal interprétées ; ces données affluent à des vitesses accrues.

La Gendarmerie Nationale ne fait pas l'exception à cette situation. Parmi les données produites en grandes quantités et à grande vitesse sont celles des accidents de la route et les infractions liées aux code de la route. Ces données doivent être interprétées afin de créer de l'information utile, et les analyses servant de support pour le processus de prise de décision.

C'est dans cette optique, et dans la finalité de trouver des solutions aux problèmes récurrents dans le processus décisionnel, que la Gendarmerie Nationale a entamé le projet de réalisation d'un Data Warehouse pour permettre la mise en place d'un système décisionnel fiable et efficace intégrant les outils cartographiques dans les analyses présentés sous forme d'un tableau de bord décisionnel.

Afin de cerner au mieux les besoins en analyses décisionnelles des utilisateurs, il fallait tenir des entretiens avec les décideurs, en parallèles avec l'étude des applications et leurs états de sorties, cela a mené au recensement des besoins exprimés d'une façon explicite. Par ailleurs, l'étude des besoins a connu l'implication des développeurs des applications opérationnelles, les administrateurs des bases de données ainsi que les producteurs des données opérationnelles dans l'optique d'anticiper sur les besoins analytiques en se basant sur les sources de données offertes par les systèmes OLTP. Cette étude a fait ressortir quatre sujets d'analyse dignes d'intérêt qui sont : les accidents, les victimes, les chauffeurs, les véhicules et les infractions liées au code de la route.

La conception du data warehouse a été basée sur une approche mixte dérivant de la fusion des deux principales approches, renommées dans le monde des data warehouses, «Besoins d'analyse» (Top-Down) et l'approche «Sources de données» (Bottom-Up). Ce choix hybride a permis d'honorer les exigences des utilisateurs et d'aller au-delà vers l'anticipation des besoins nouveaux ou futures, non exprimés dans l'immédiat, grâce l'étude du potentiel offert par les données générées par les systèmes opérationnels.

La modélisation de la zone de stockage des données suivant les principes de la modélisation dimensionnelle vue la combinaison de simplicité et clareté qu'elle offre, permettant une compréhension

Conclusion Générale et perspectives

facile par les utilisateurs impliqués tout au long de ce processus. Cette étape a permis de définir les modèles en étoile des sujets proposés⁵².

Le data warehouse a été conçu selon le HOLAP, qui est la méthode hybride obtenue en la mixture du MOLAP et du ROLAP, donc il a été question d'implémenter la partie « implémentation physique des modèles dimensionnels sur un SGBD relationnel » représentant l'étape « alimentation de la zone de stockage ». Un temps et un effort considérables ont été consacrés à cette étape, ne faisant pas exception aux projets décisionnels dans le monde des entreprises qui en moyennes prend plus de 80% des projets data warehouses. Les outils Oracles ont été utilisés afin de concevoir et implémenter et automatiser les routines d'extraction, transformation et chargement des données.

Les outils oracle ont été utilisés pour l'implémentation du data warehouse et les solutions de restitution, pour leur puissance, efficacité et fiabilité mais aussi pour le simple fait que la Gendarmerie Nationale en dispose déjà, cela a permis de valoriser l'investissement.

Donc, et afin de réaliser la partie MOLAP⁵³ de l'approche HOLAP, l'outil « Administration BI » a été utilisé afin de créer le référentiel qui sera téléchargé vers le serveur « Business Intelligence » (BI) afin de servir de référentiel « Repository » pour le « Oracle Business Intelligence Enterprise Edition » (OBIEE), qui est l'outil de réalisation du tableau de bord (La partie restitution).

Ce mémoire a eu pour livrable un data warehouse et un tableau de bord. Ces deux produits ne sont pas encore arrivés à leur maturité. Ils devront suivre un processus de la spirale incrémentale. Certaines améliorations devront être apportées concernant les perspectives :

- Apporter les corrections et améliorations selon le « feed back » des utilisateurs ;
- Inclure d'autres systèmes opérationnels dans le processus d'alimentation du data warehouse, notamment les systèmes de la sûreté nationale ;
- Introduire les techniques de « fouille de données » (Data Mining) afin de découvrir des patterns et de prévoir certains comportements et résultats pour une meilleure exploitation des données.
- Continuer le développement du tableau de bord (l'outil de restitution).

⁵² Les accidents, les victimes, les chauffeurs, les véhicules et les infractions liées au code de la route.

⁵³ les modèles agrégés des analyses les plus fréquentes.

Annexe 01 : Diagrammes en étoile

Annexe 01 : Diagrammes en étoiles :

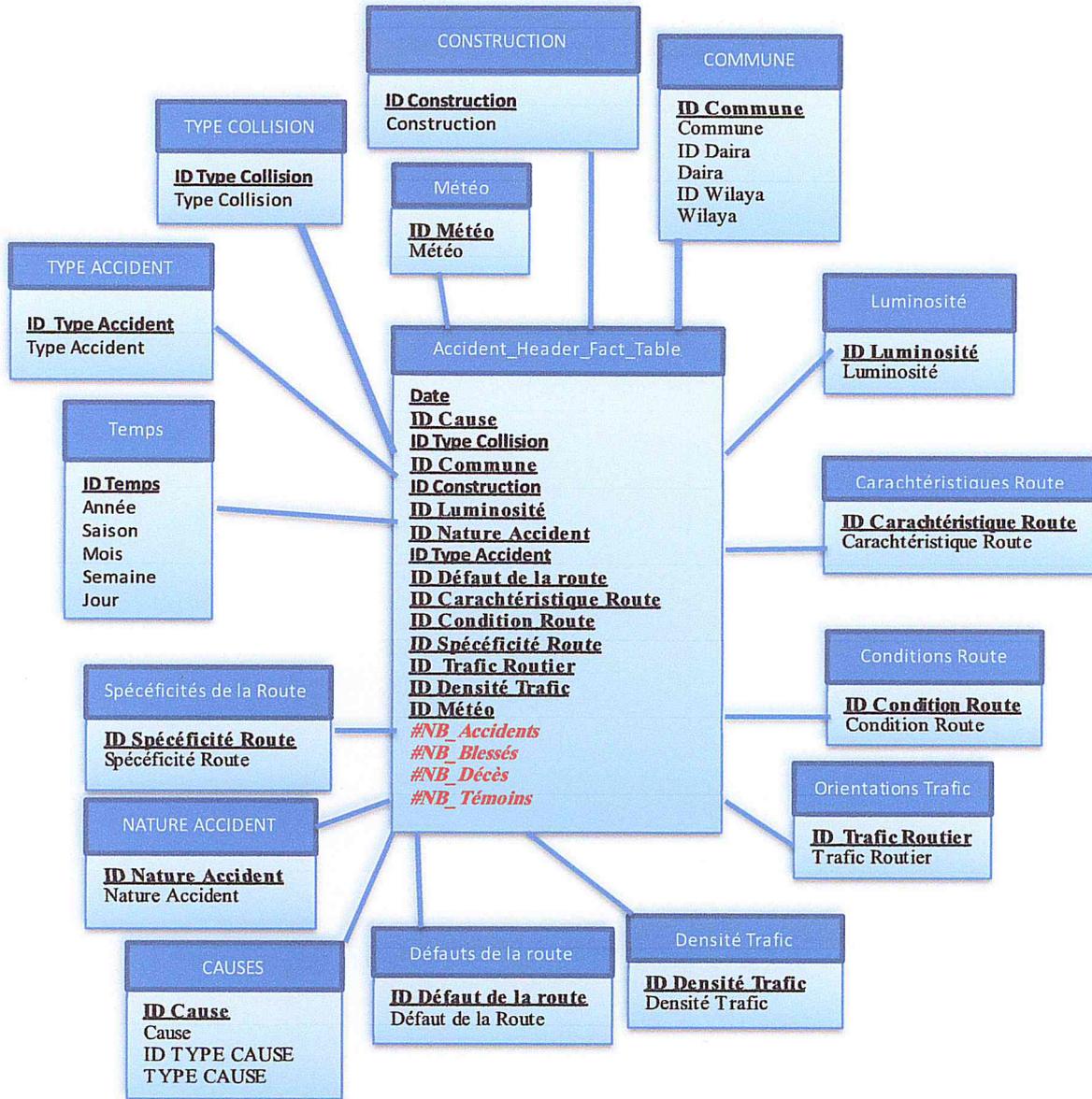


Figure A01.1 : Le modèle en étoile du volet Accident

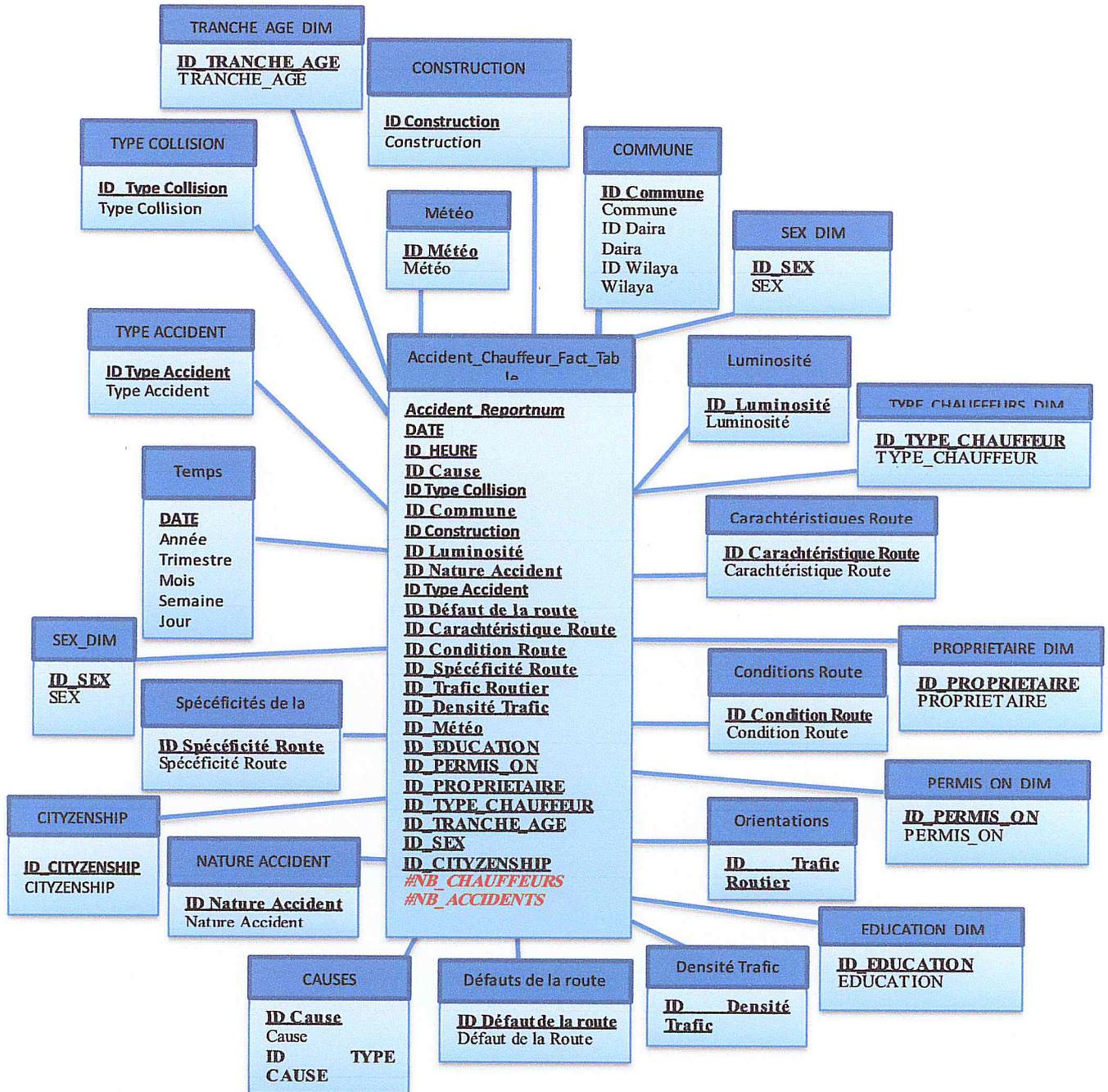


Figure A01.2 : Le modèle en étoile du volet Chauffeurs

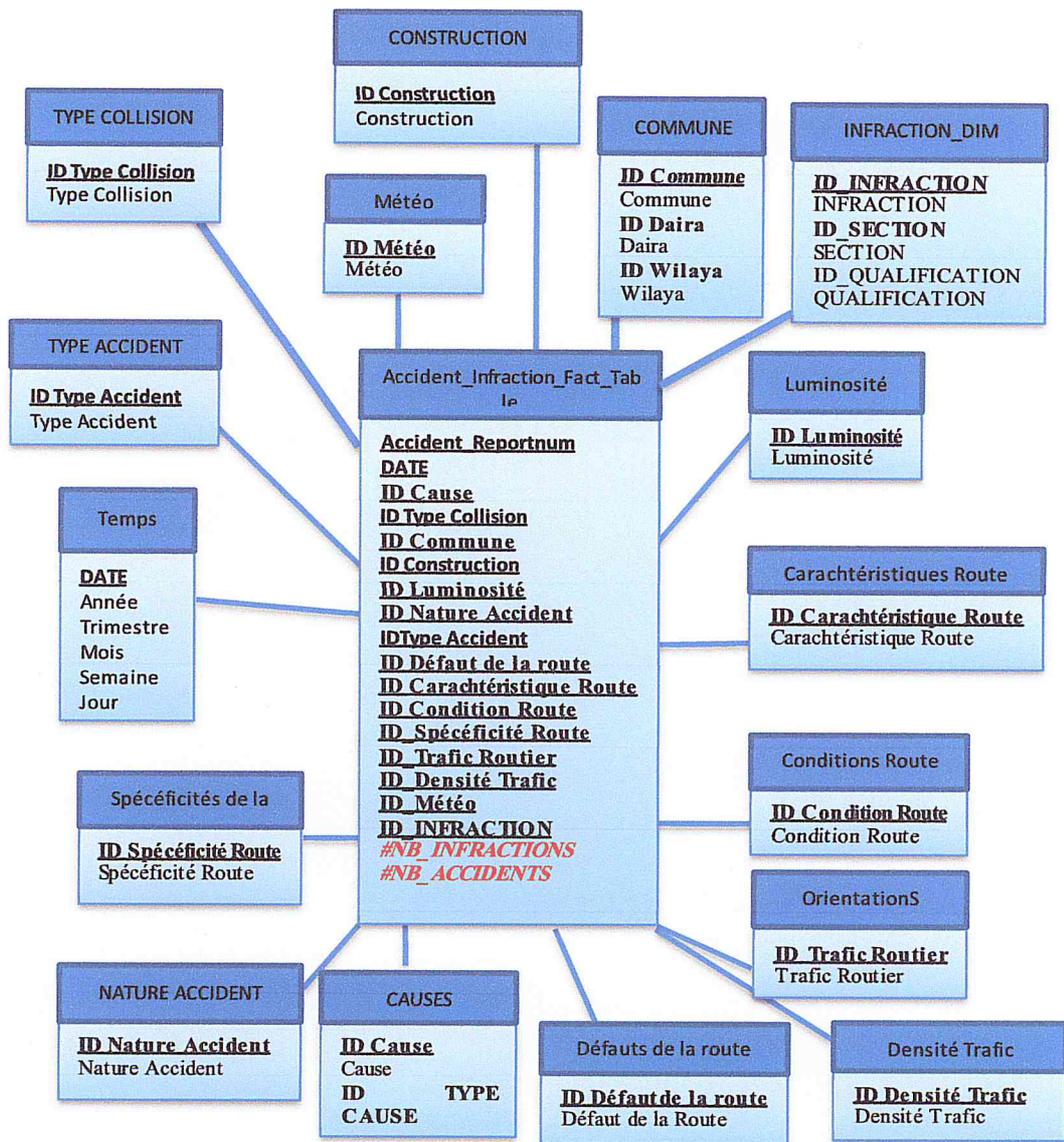


Figure A01.3 : Le modèle en étoile du volet Infractions

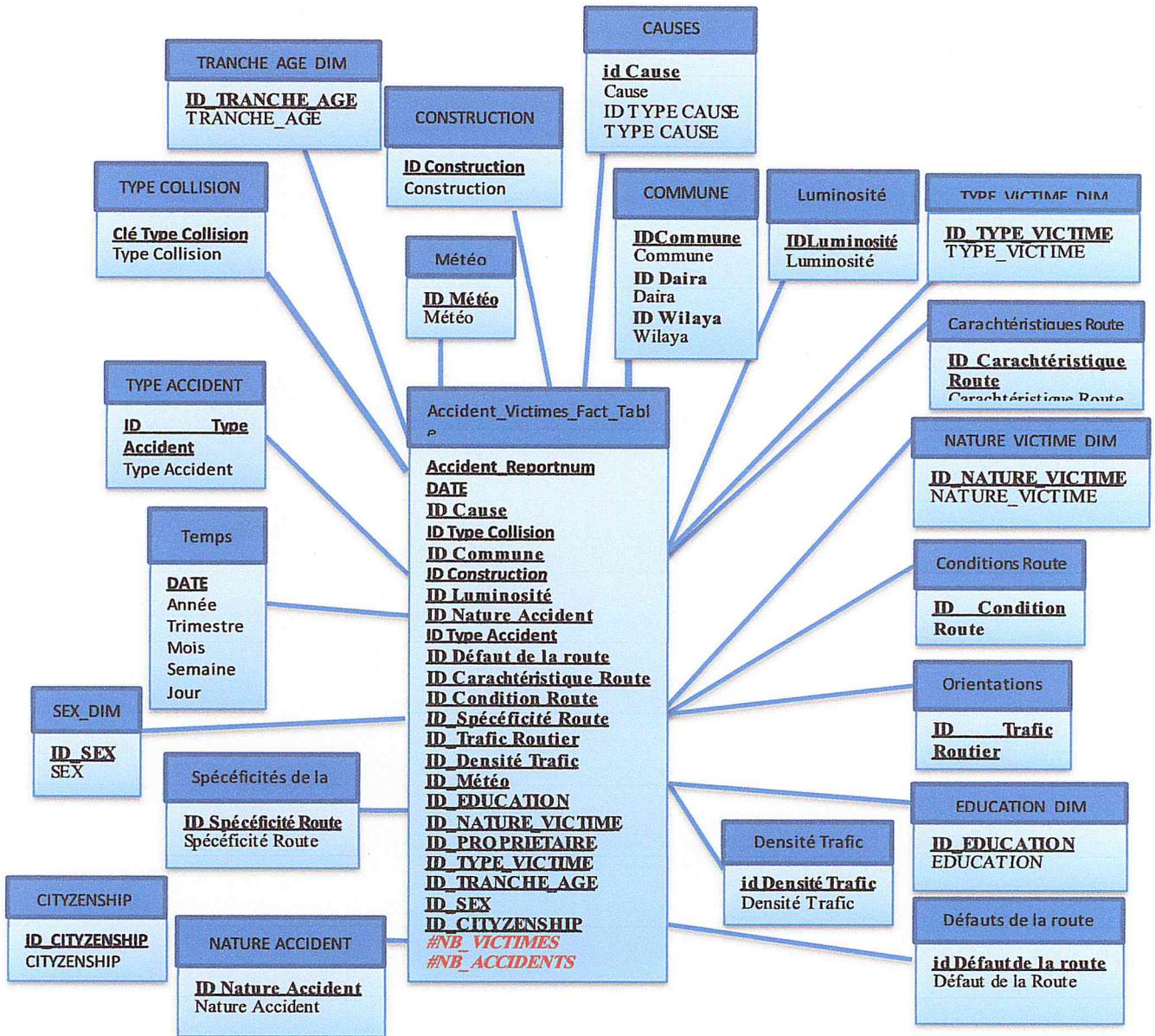


Figure A01.4 : Le modèle en étoile du volet « victimes »

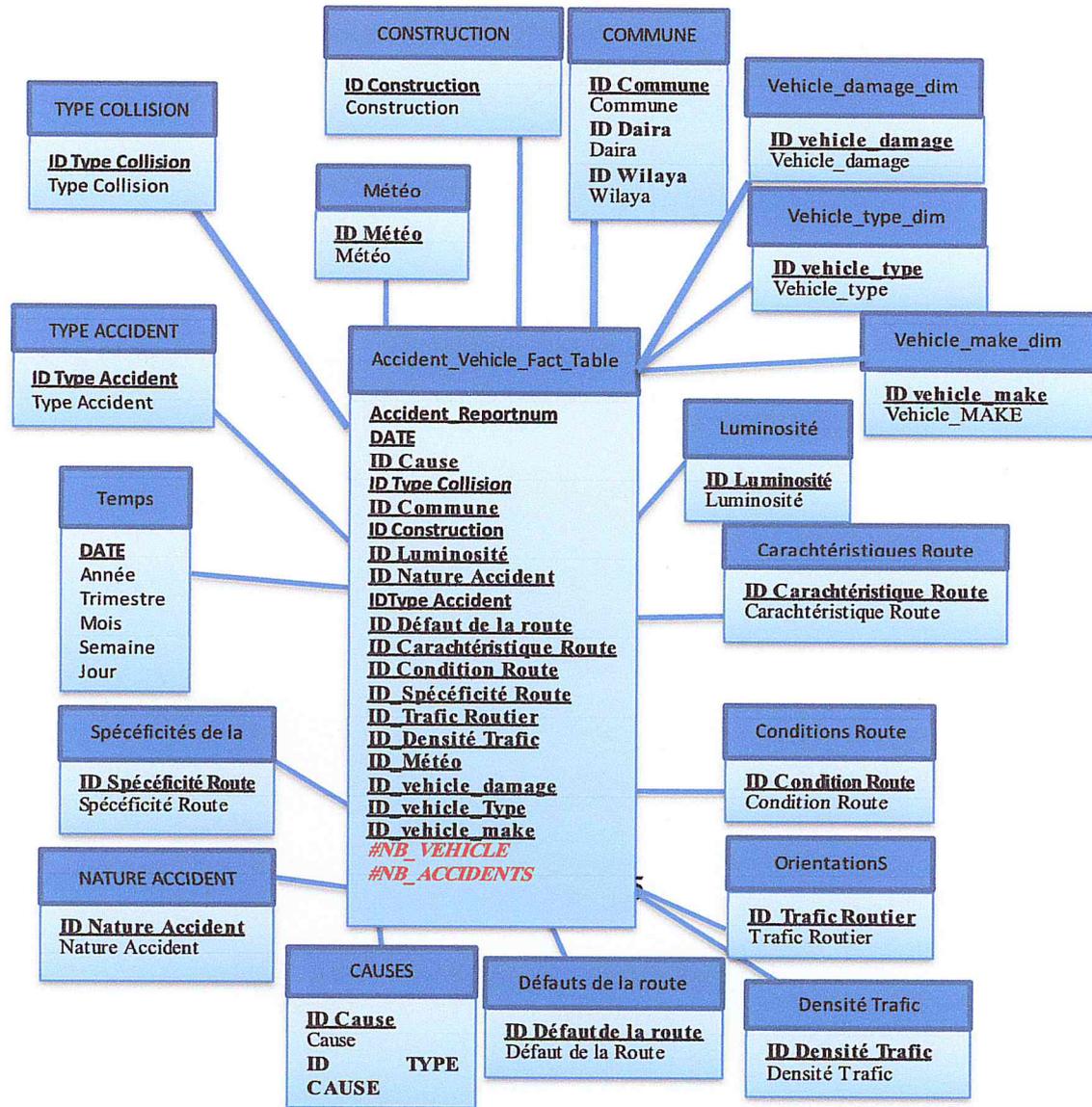


Figure A01.5 : Le modèle en étoile du volet Véhicule

*Annexe 02 : Création, mapping et
déploiement d'une dimension*

Annexe 02 : Création, mapping et déploiement d'une dimension

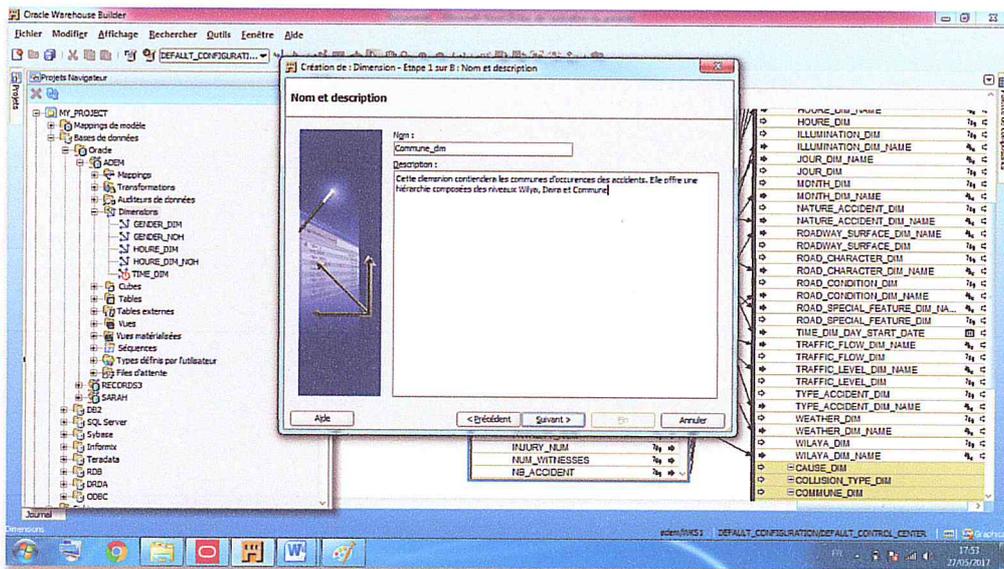


figure A02.1 : L'ajout d'une nouvelle dimension.

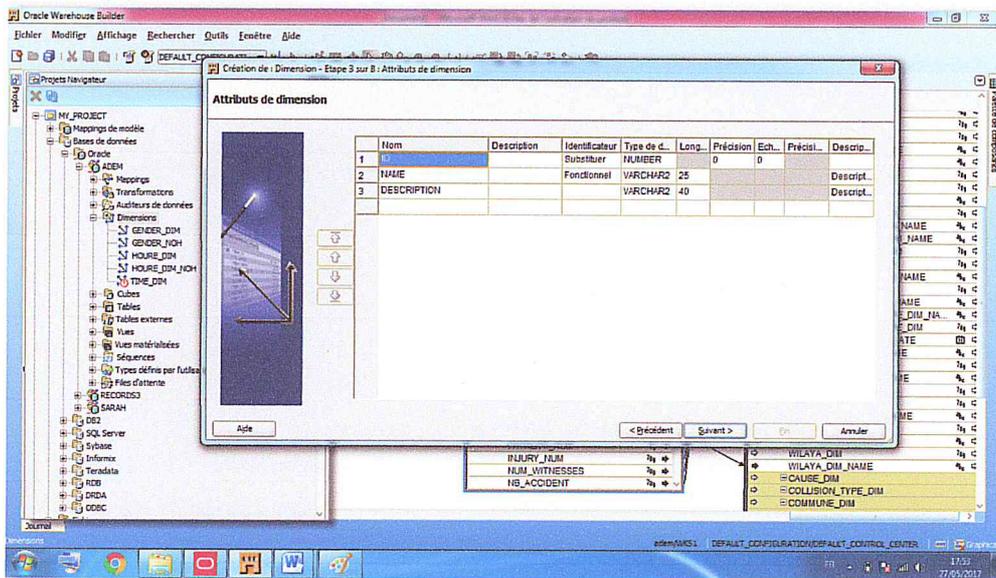


figure A02.2 : Spécification des attributs de la dimension

Annexe 02 : Création, mapping et déploiement d'une dimension

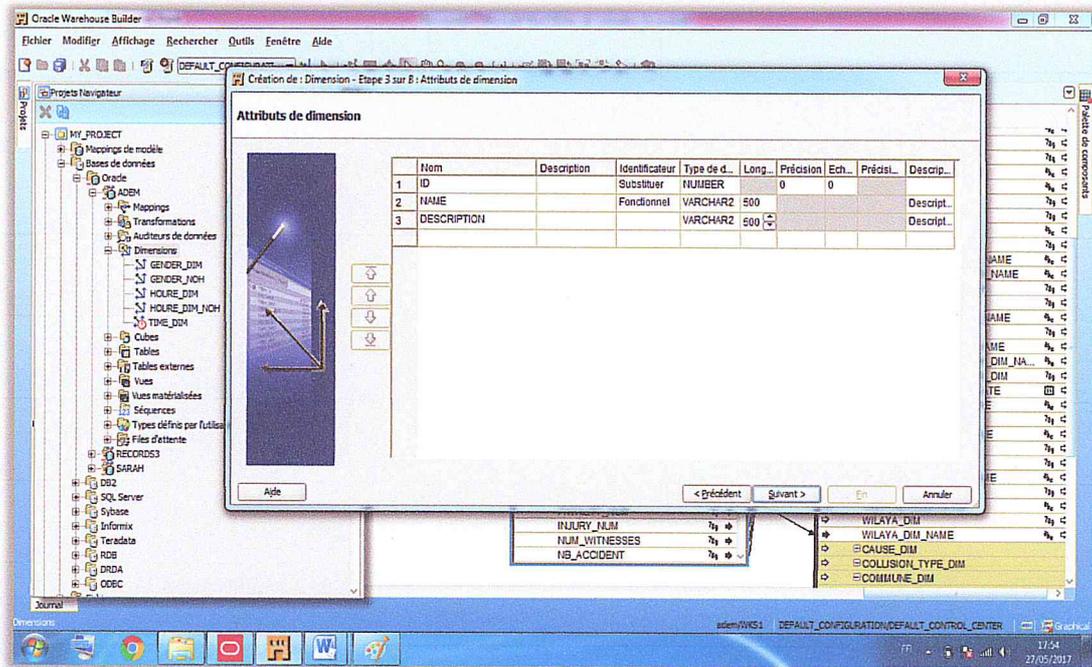


figure A02.3 : Augmentation des tailles des attributs vers 500.

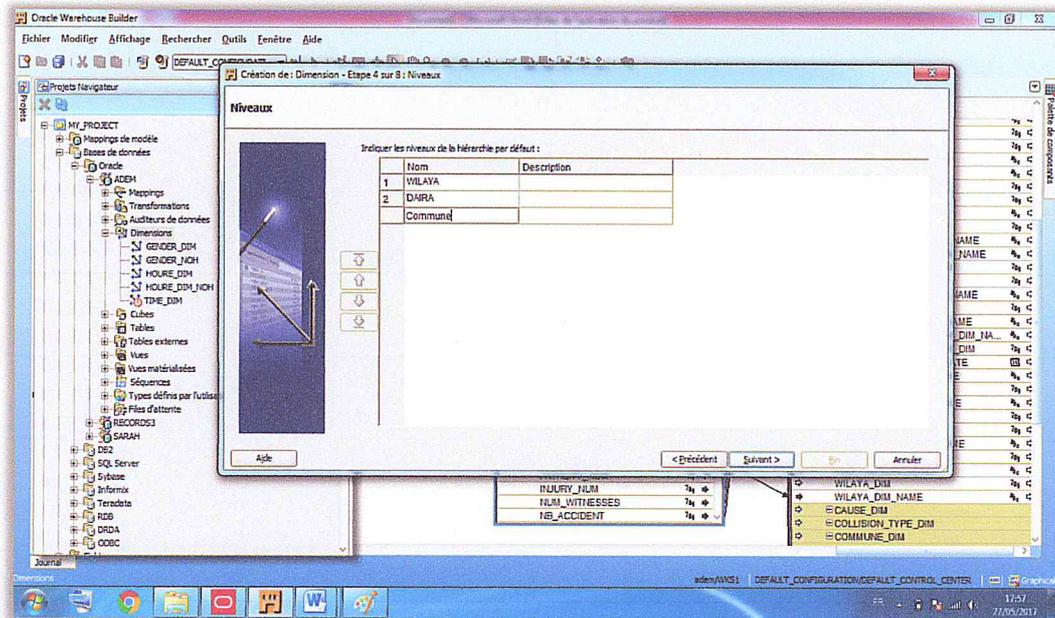


figure A02.4 : Spécification des niveaux

Annexe 02 : Création, mapping et déploiement d'une dimension

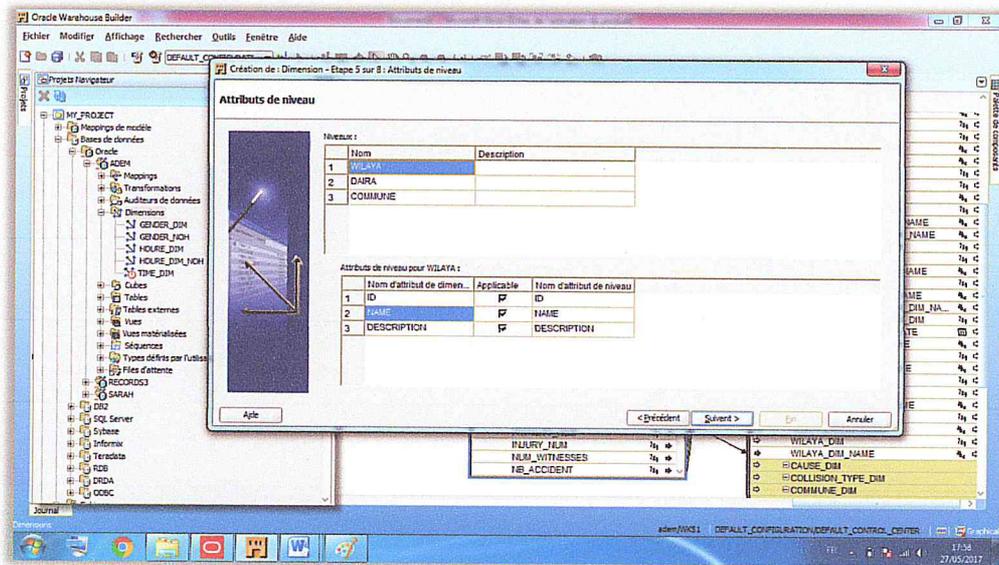


figure A02.5 : Spécification des attributs des niveaux

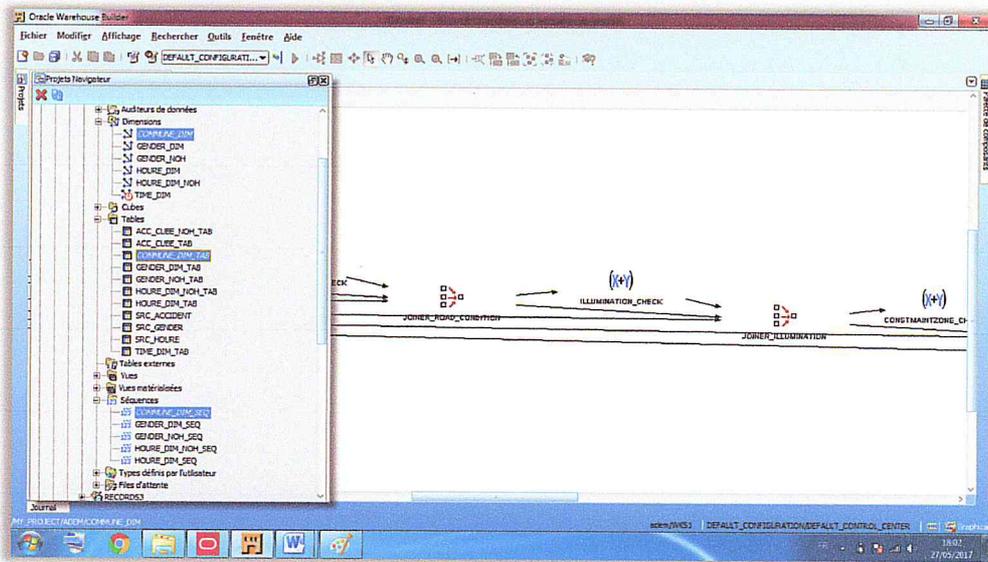


figure A02.6 : Autocréation des dimensions commune_dim, la table Commune_dim_TAB et la séquenceCommune_dim_SEQ

Annexe 02 : Création, mapping et déploiement d'une dimension

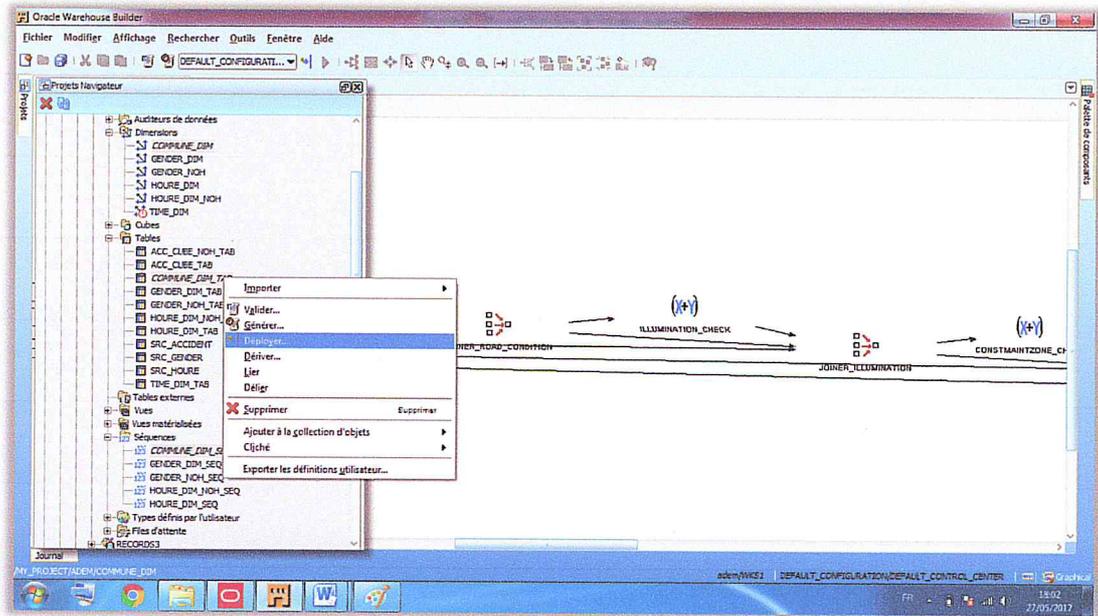


figure A02.6 : Déploiement de la Commune_dim, Commune_dim_TAB et Commune_dim_SEQ

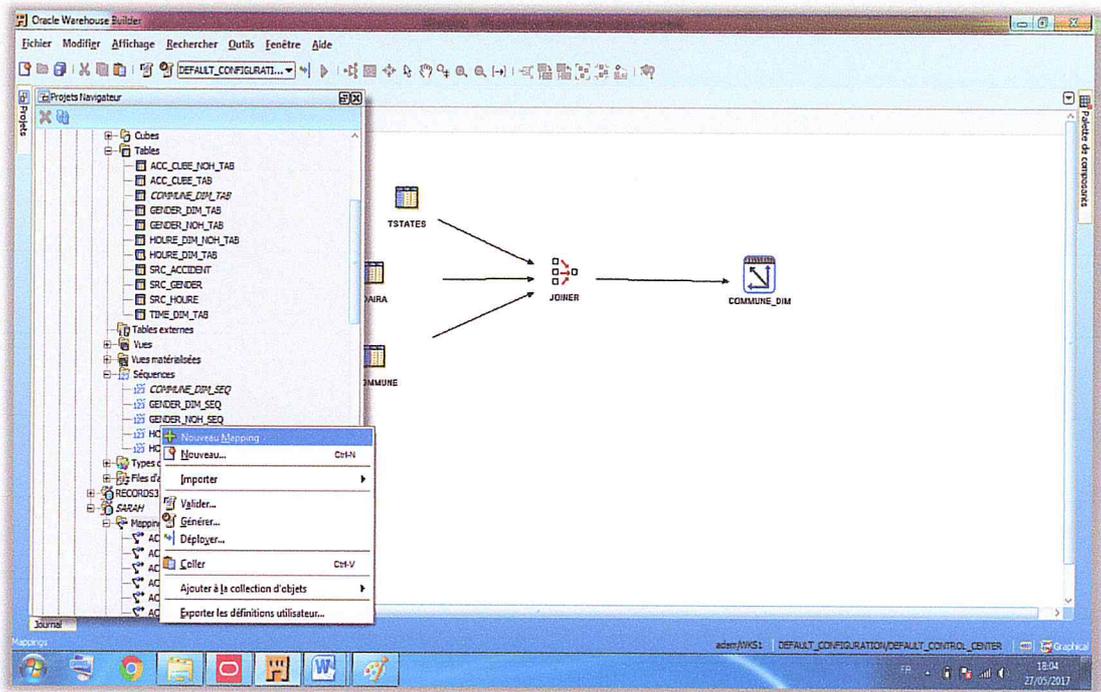


figure A02.7 : L'ajout d'un nouveau mapping

Annexe 02 : Création, mapping et déploiement d'une dimension

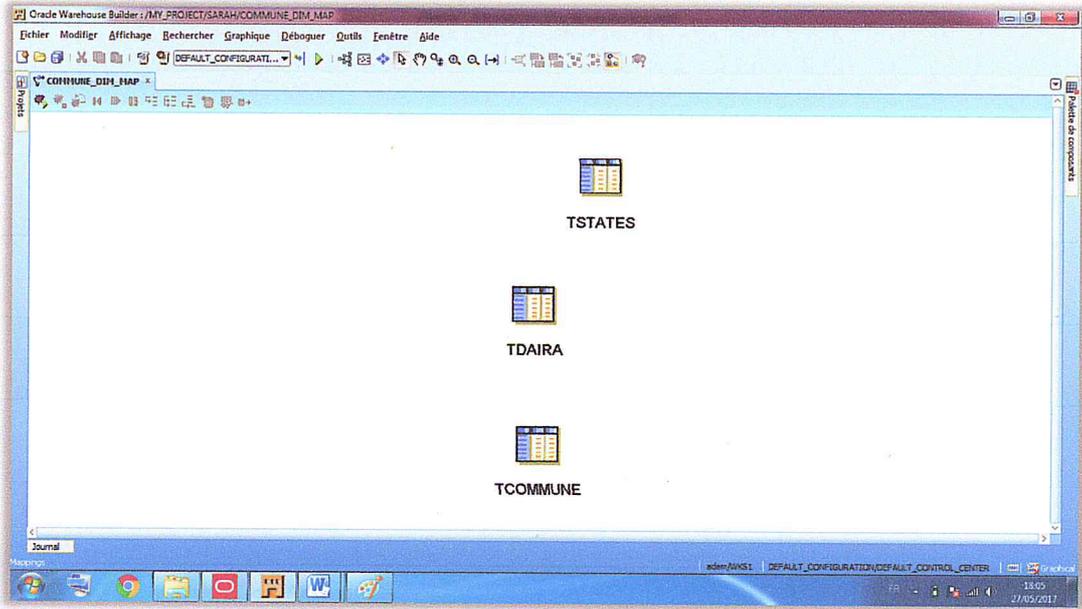


figure A02.8 : La sélection des tables sources

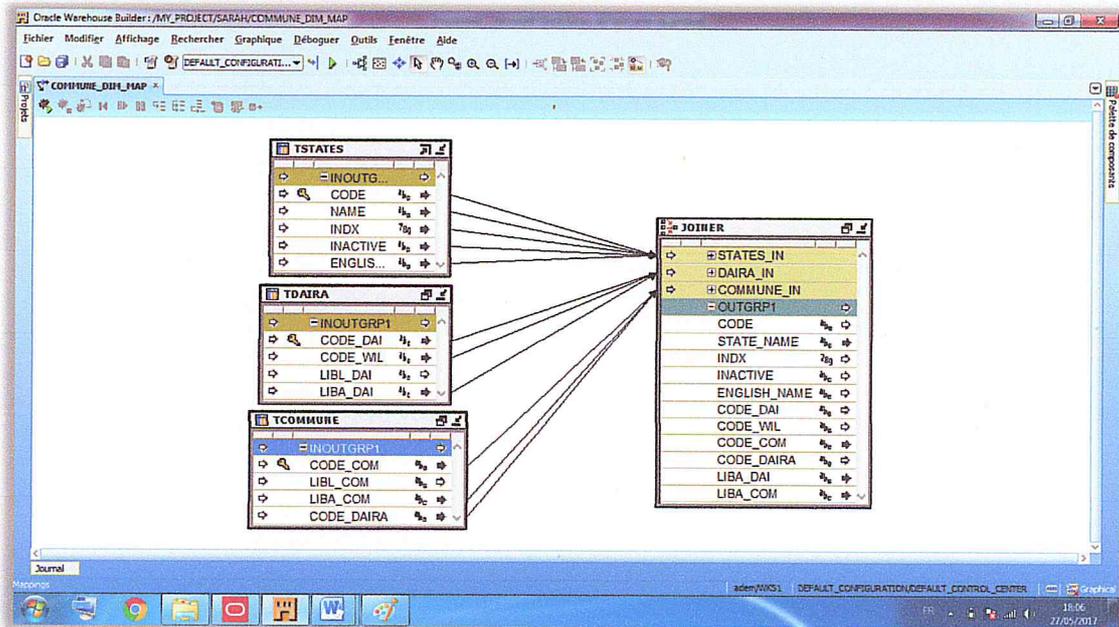


figure A02.9 : Regroupement des données avec l'opérateur de jointure « JOINER ».,,

Annexe 02 : Création, mapping et déploiement d'une dimension

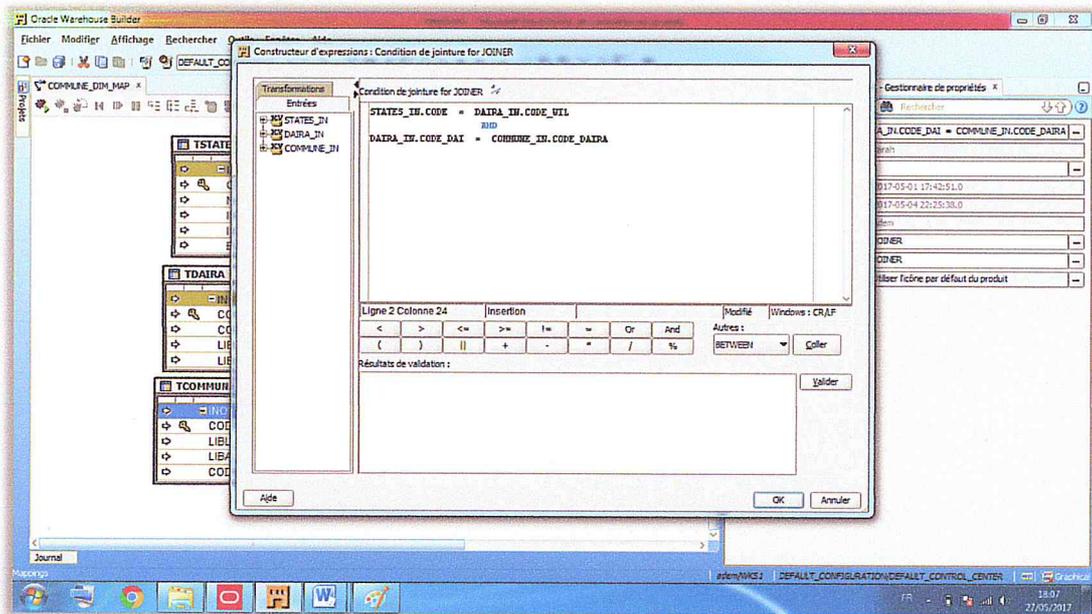


figure A02.10 : Spécification de la condition de jointure

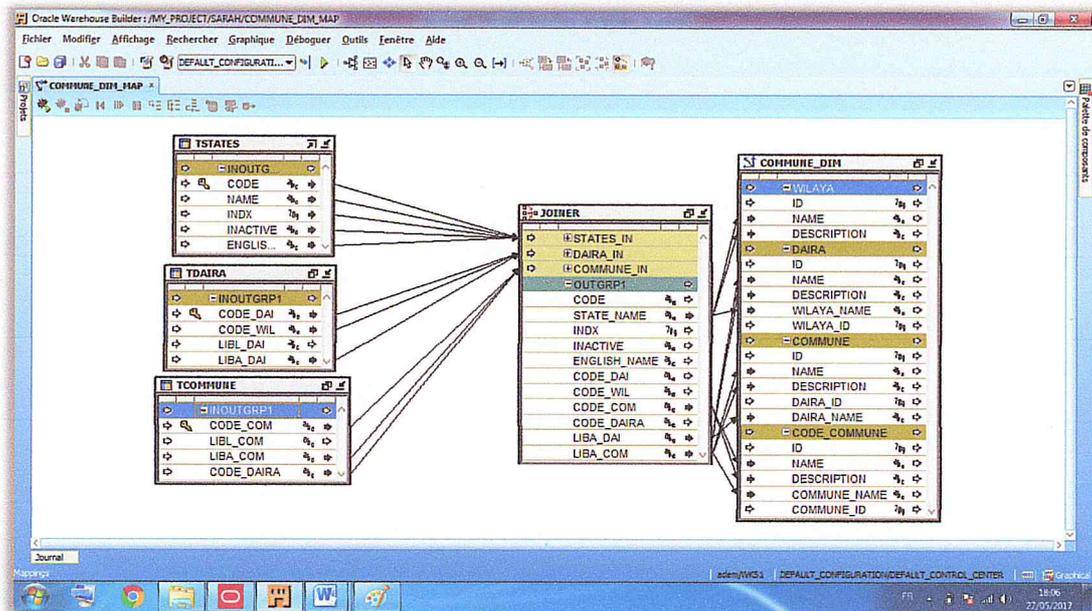


figure A02.11 : Liaison entre le JOINER et la dimension « DIM_INFRACTION ».

Annexe 02 : Création, mapping et déploiement d'une dimension

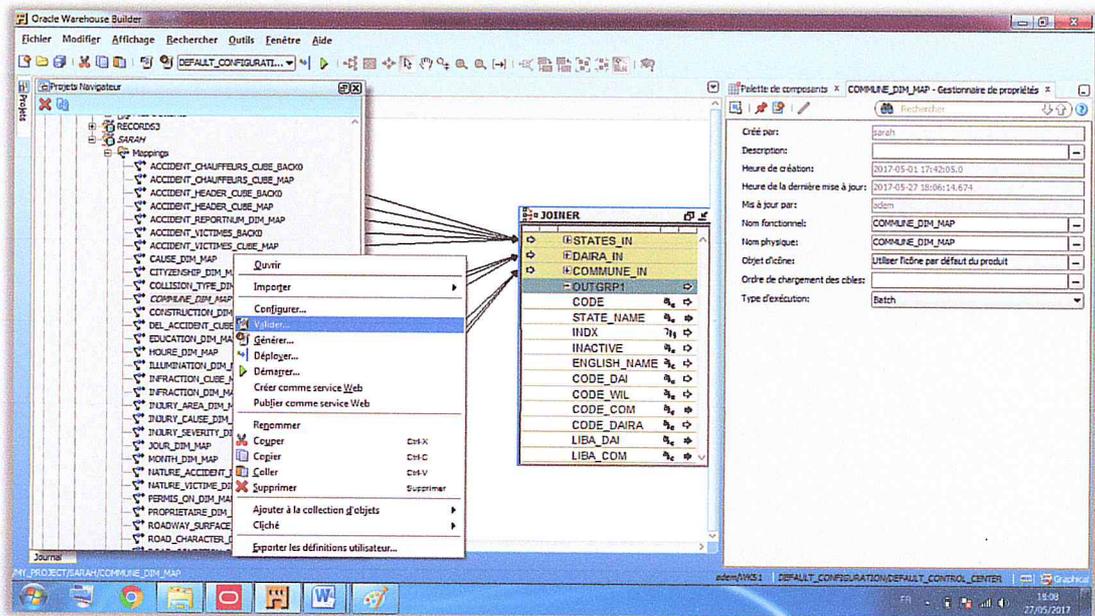


figure A02.12 : Validation et génération des scripts PISQL du « Commune_dim_map »

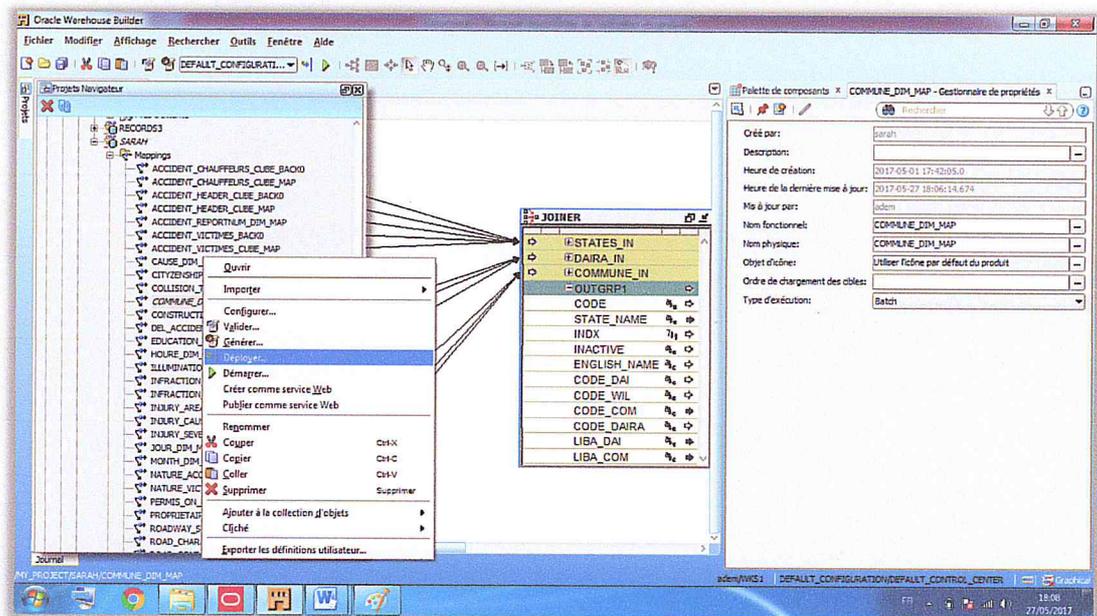


figure A02.13 : Déploiement du « commune_dim_map ».

Annexe 02 : Création, mapping et déploiement d'une dimension

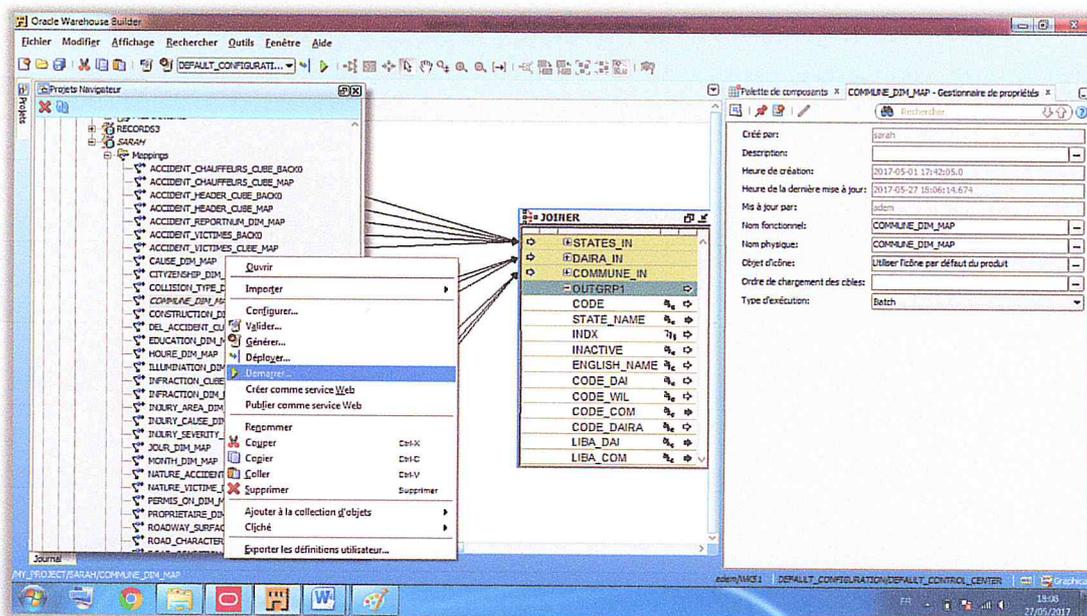


figure A02.14 : Démarrage du « commune_dim_map »

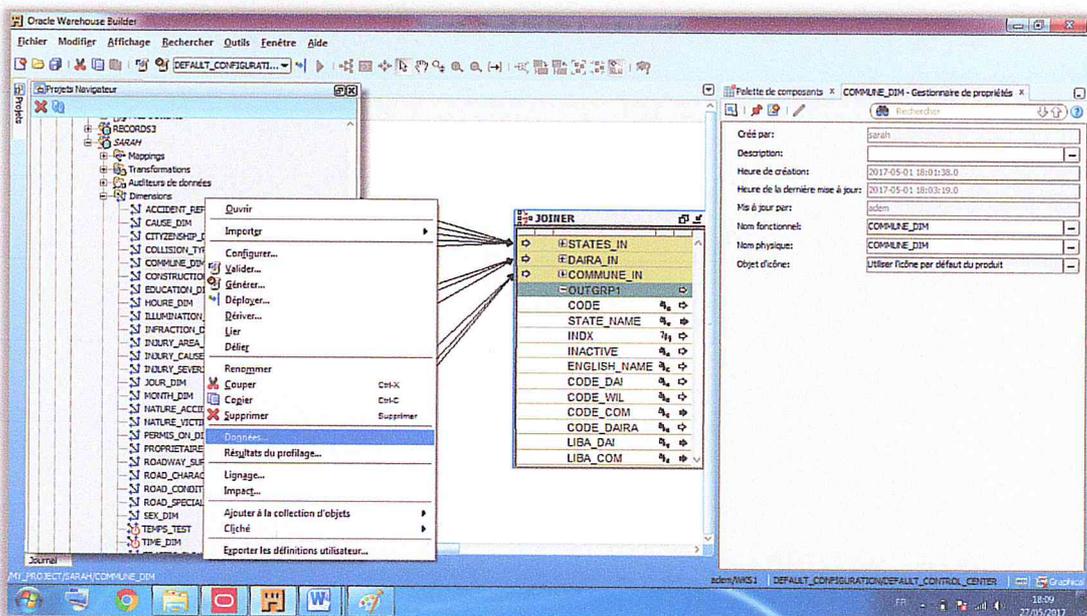


figure A02.15 : Visualisation des données de la dimension « Commune_DIM »

Annexe 03 : Mapping cube Accident

Annexe 03 : Mapping du cube Accident :

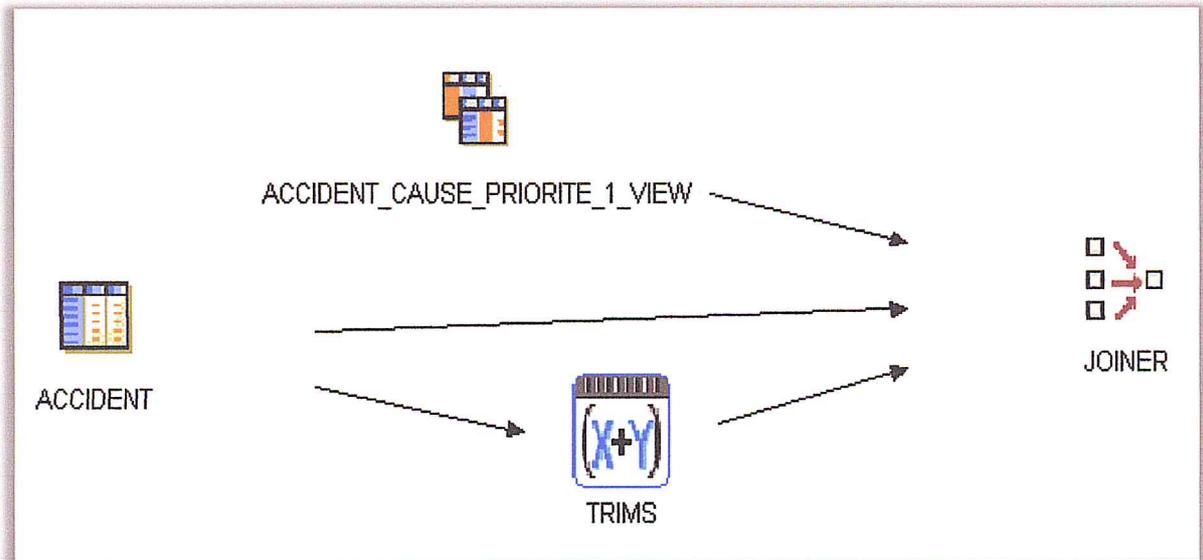


Figure A03.1 : Section N° 01 du mapping global du cube « Accident »

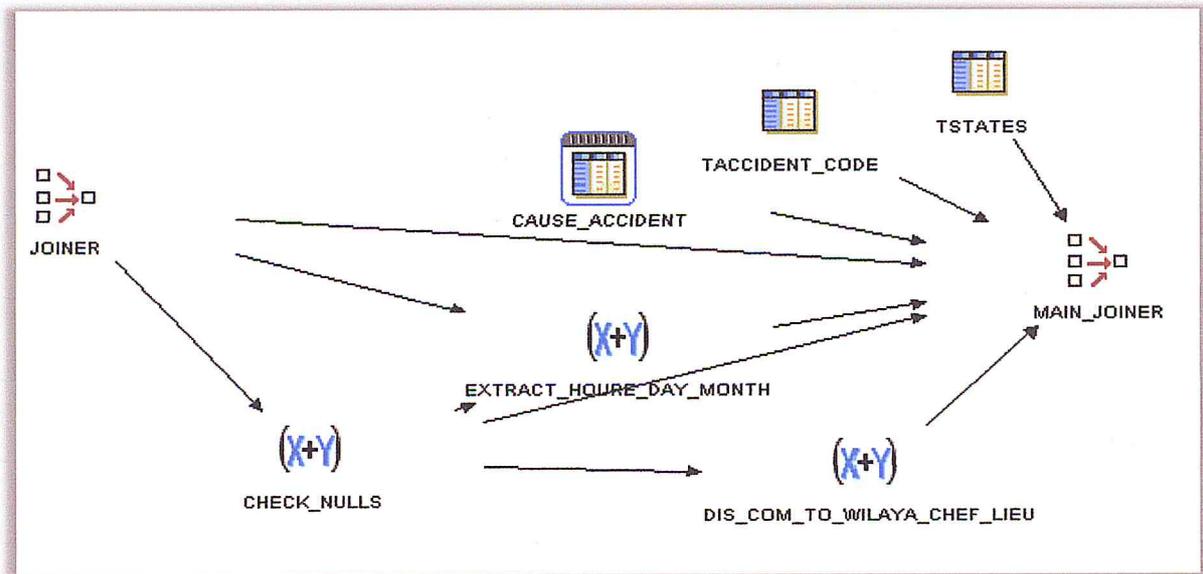


Figure A03.2 : Section N° 02 du mapping global du cube « Accident »

Annexe 03 : Mapping du cube Accident

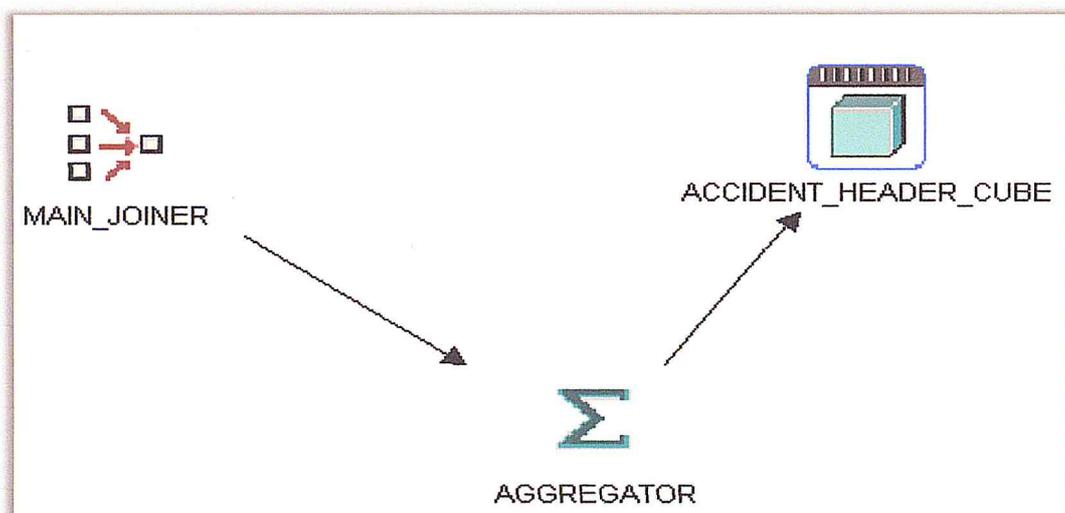


figure A03.3 : Section N° 03 du mapping global du cube « Accident »

Annexe 03 : Mapping du cube Chauffeurs

Annexe 04 : Mapping du cube Chauffeurs

Annexe 04 : Mapping du cube Chauffeurs :

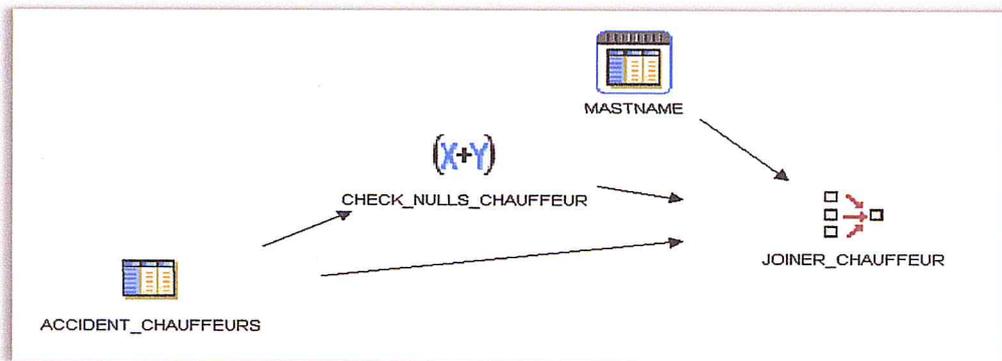


Figure A04.1 : Section N° 01 du mapping global du cube « Chauffeurs »

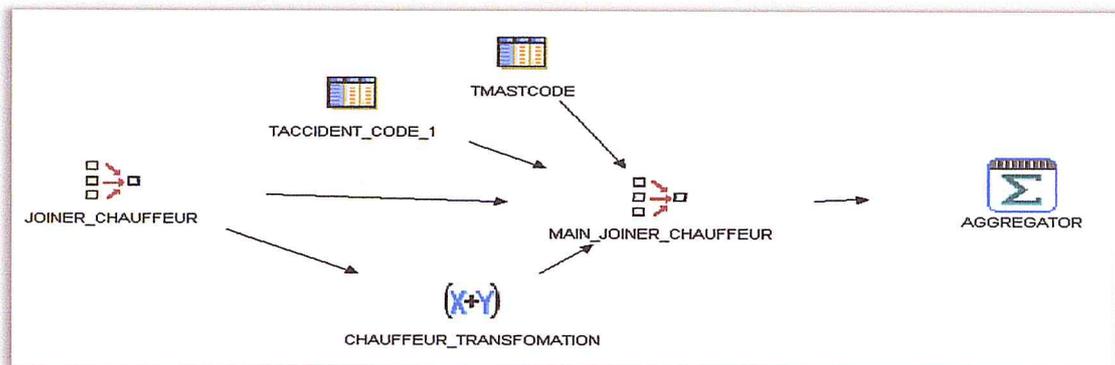


Figure A04.2 : Section N° 02 du mapping global du cube « Chauffeurs »

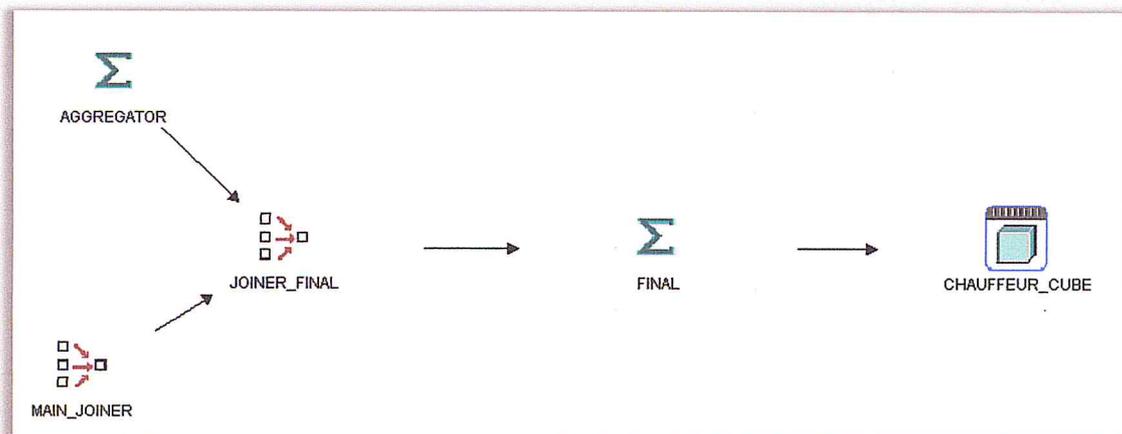


Figure A04.3 : Section N° 03 du mapping global du cube « Chauffeurs » Jointure des données du Chauffeur avec celles des accidents, les agréger et les charger dans le cube

Annexe 05 : Mapping du cube Victime

Annexe 05 : Mapping du cube Victime :

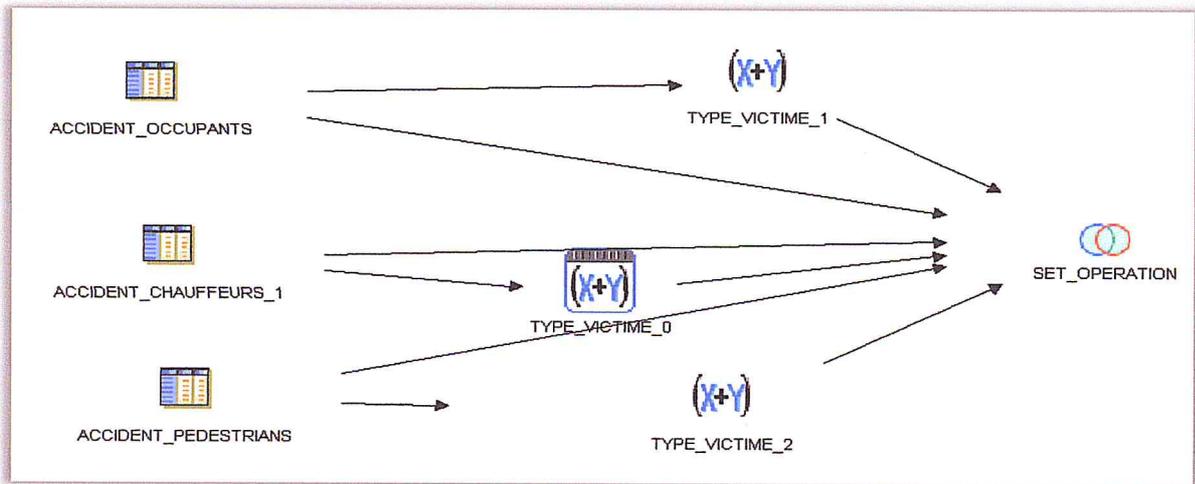


figure A05.1 Section 01 du Mapping du cube « Victimes ».

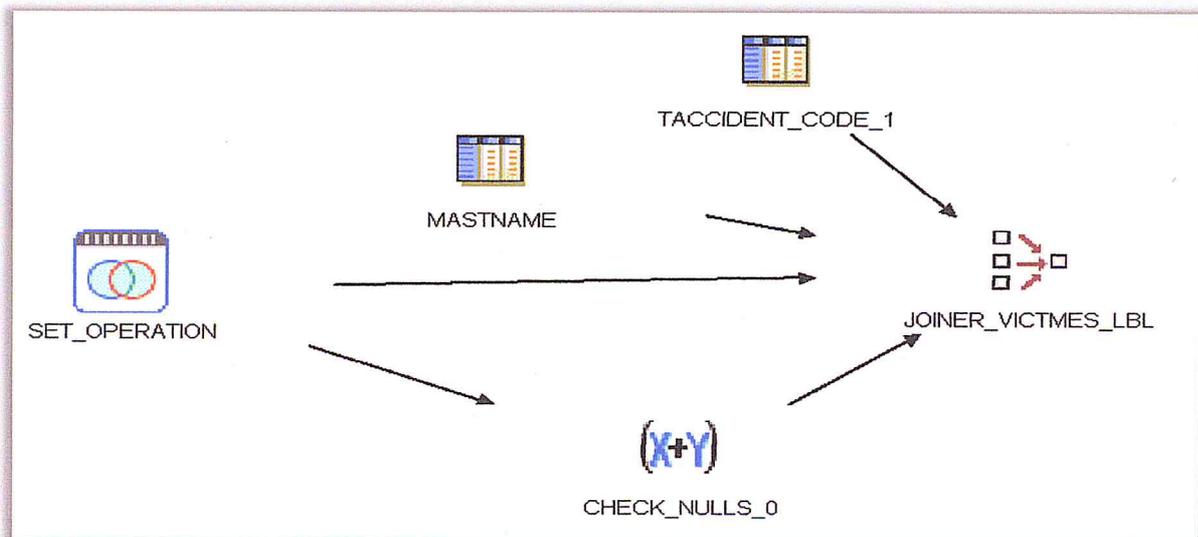


figure A05.2 Section 02 du Mapping du cube « Victimes ».

Annexe 05 : Mapping du cube Victime

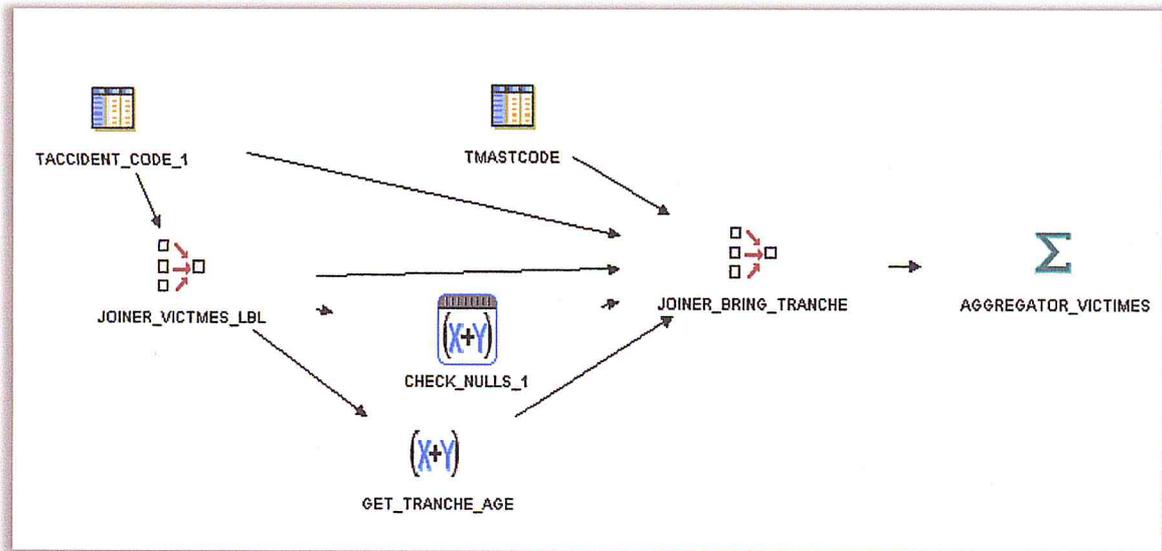


figure A05.3 Section 03 du Mapping du cube « Victimes ».

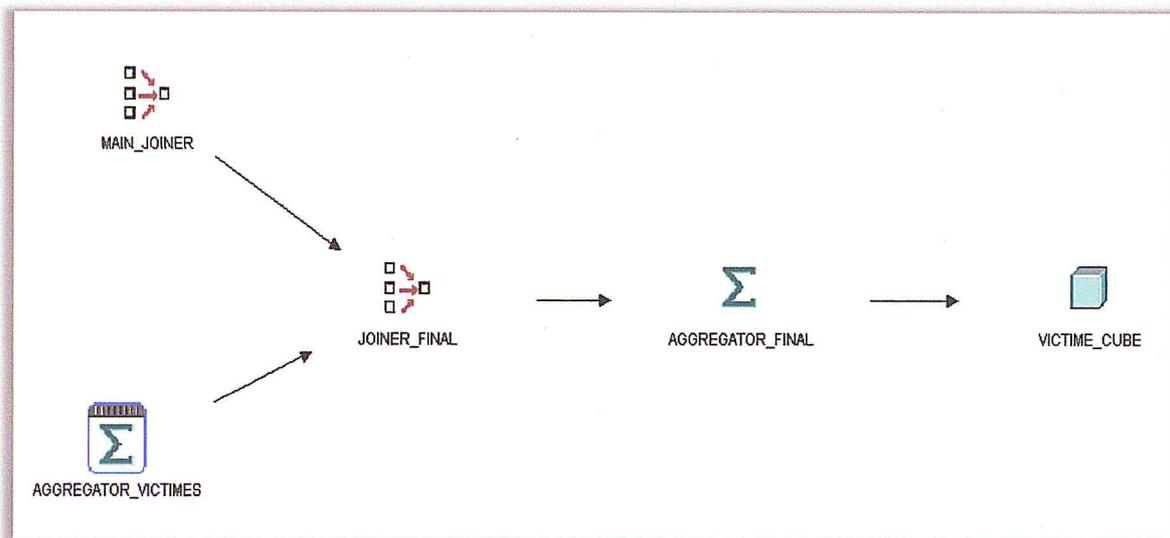


Figure A05.4 Section 04 du Mapping du cube « Victimes ».

Annexe 06 : Mapping du cube infraction

Annexe 06 : Mapping du cube Infraction

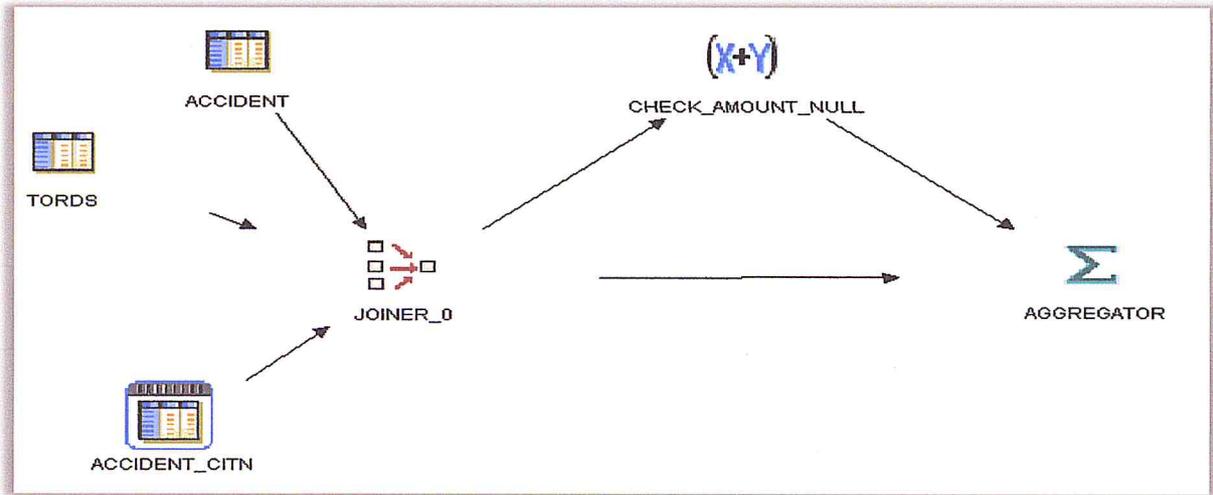


Figure A06.1 Section 01 du Mapping du cube « Infractions ».

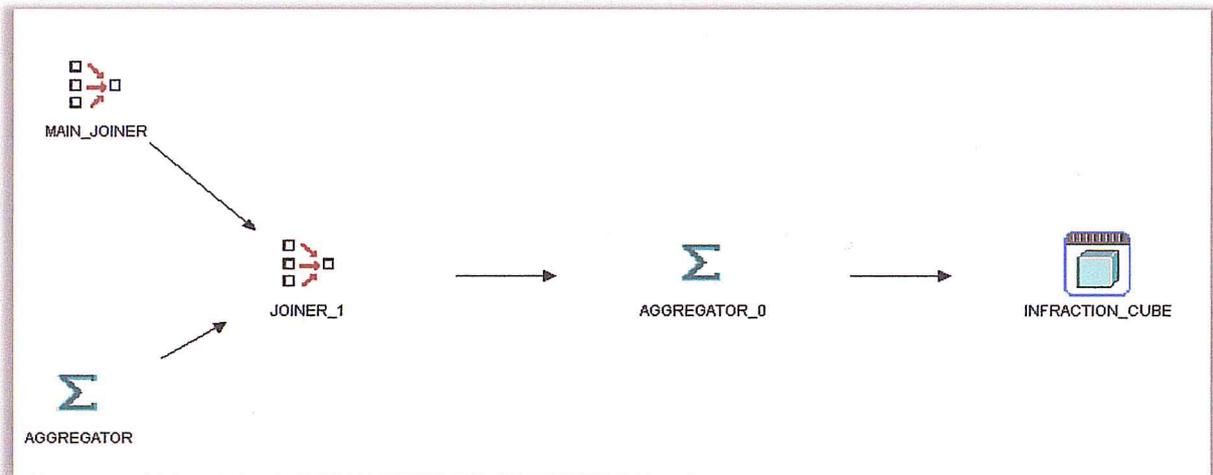


Figure A06.2 Section 02 du Mapping du cube « Infractions ».

Annexe 07 : Mapping du cube Véhicule

Annexe 07 : Mapping du cube Véhicule.

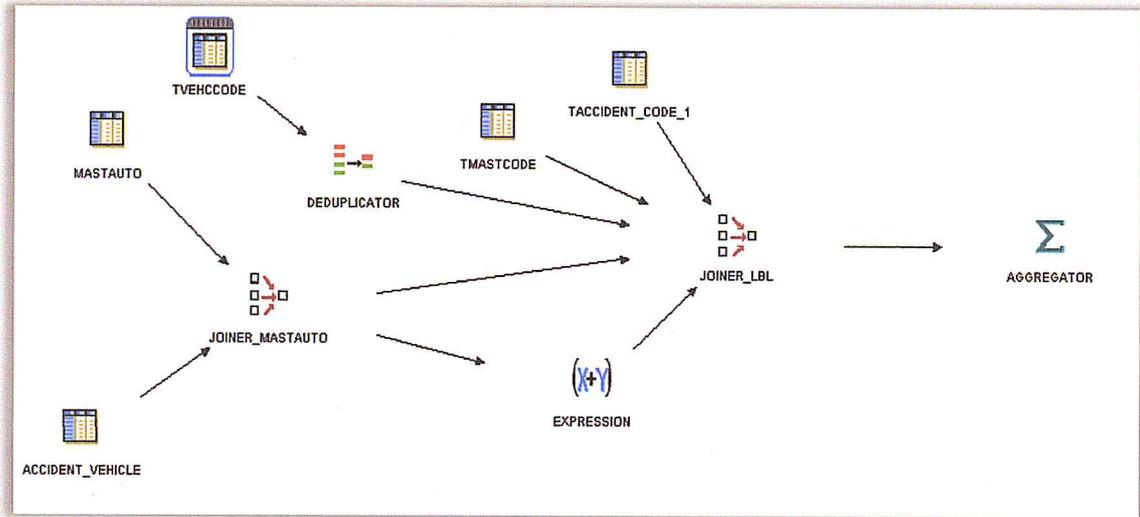


Figure A07.1 Section 01 du Mapping du cube « Véhicule ».

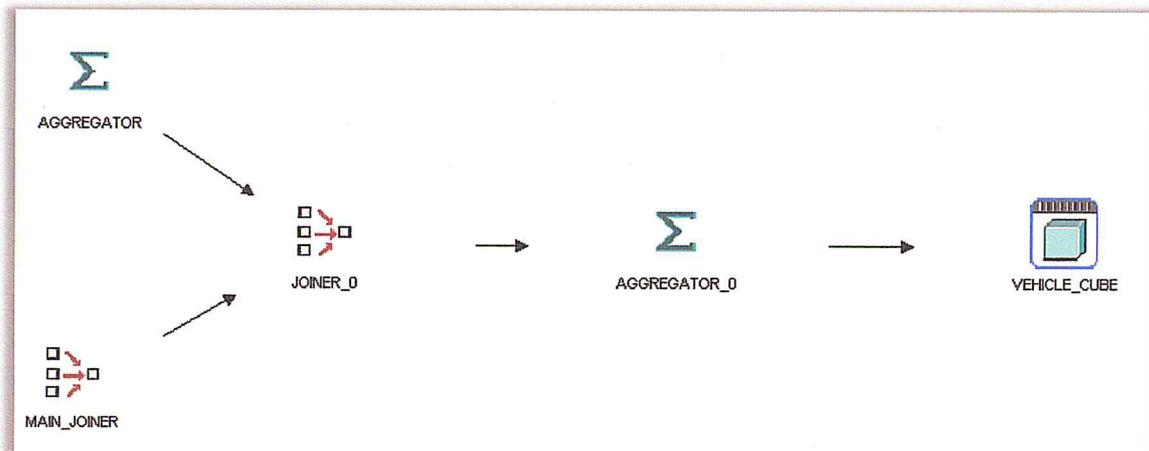


Figure A07.2 : Section 02 du Mapping du cube « Véhicule ».

*Annexe 08 : Création et Déploiement du
référentiel dans le OLAP*

Annexe 08 : Création et Déploiement du référentiel dans le OLAP

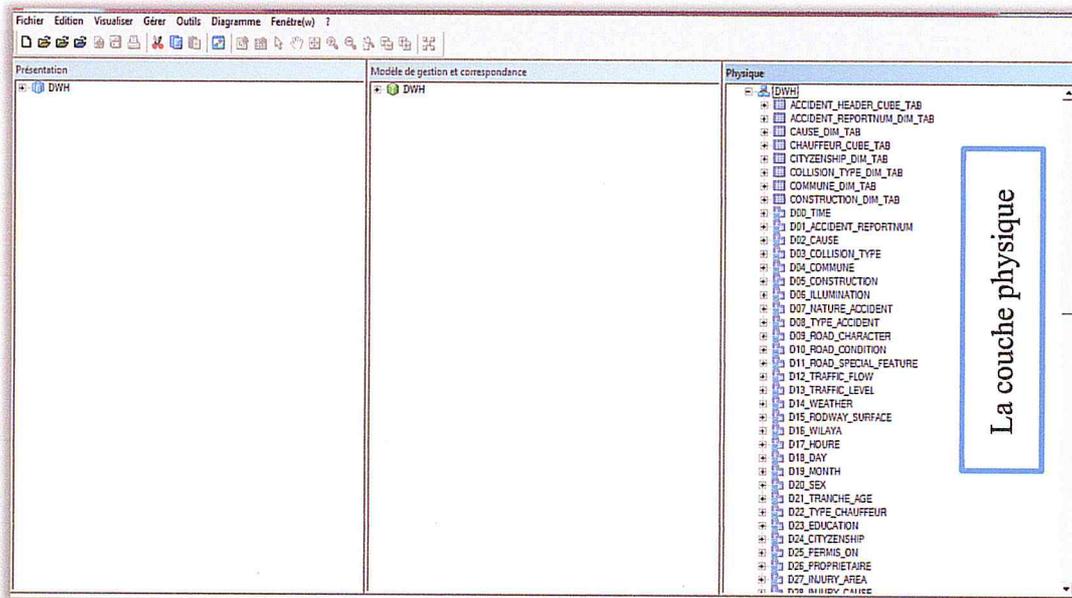


Figure A08.1 : La couche physique.

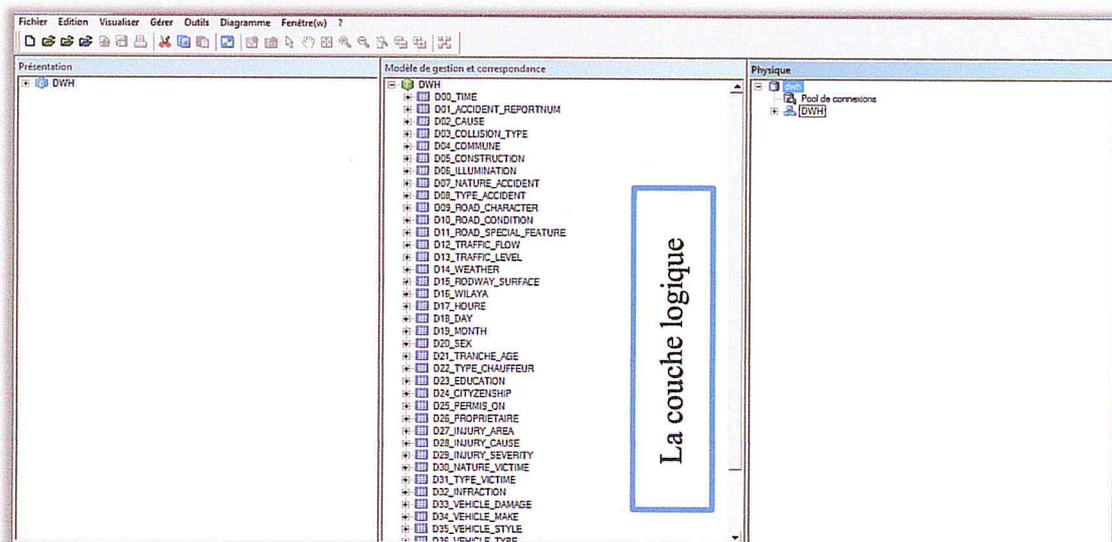


Figure A08.2 : La couche logique.

Annexe 08 : Création et Déploiement du référentiel dans le OLAP

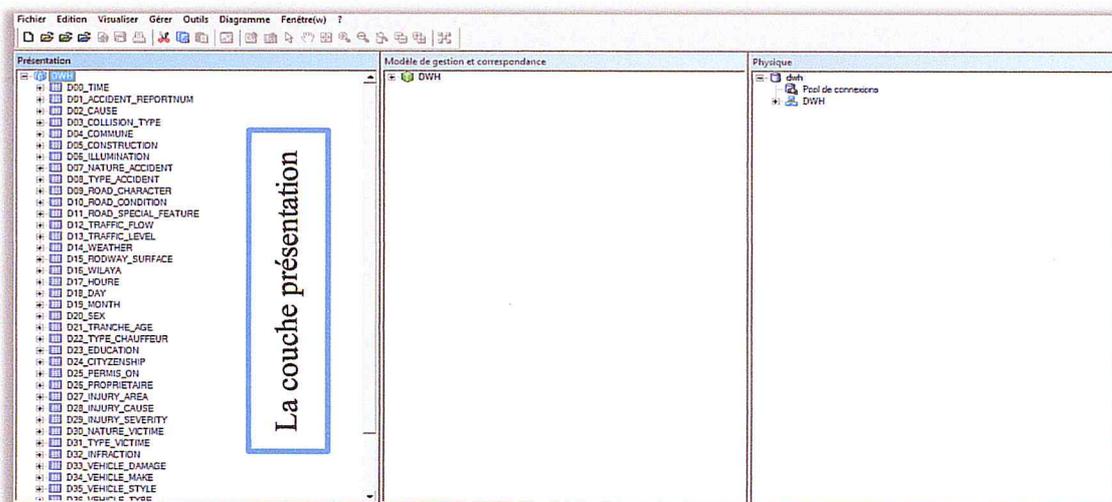


Figure A08.3 : La couche présentation.

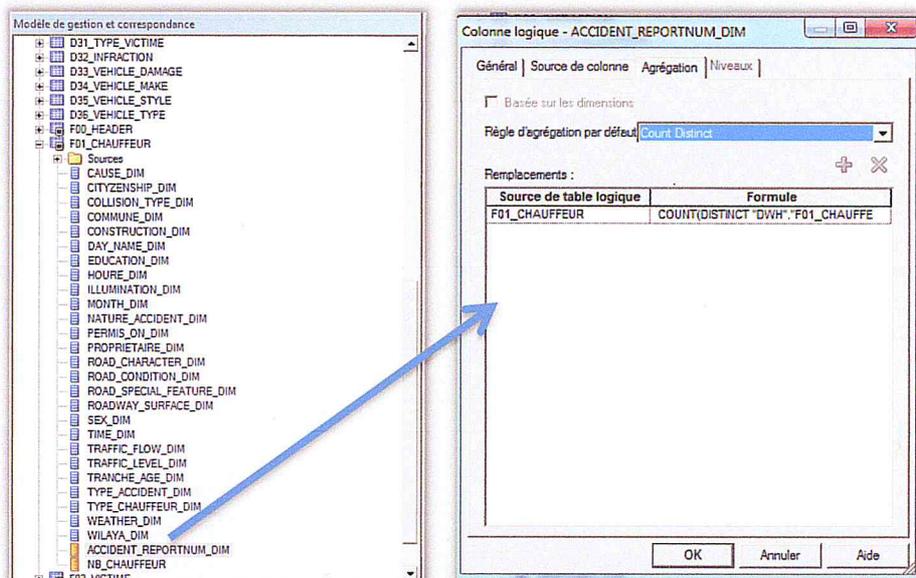


Figure A08.4 : Indication du type d'agrégation.

Annexe 08 : Création et Déploiement du référentiel dans le OLAP

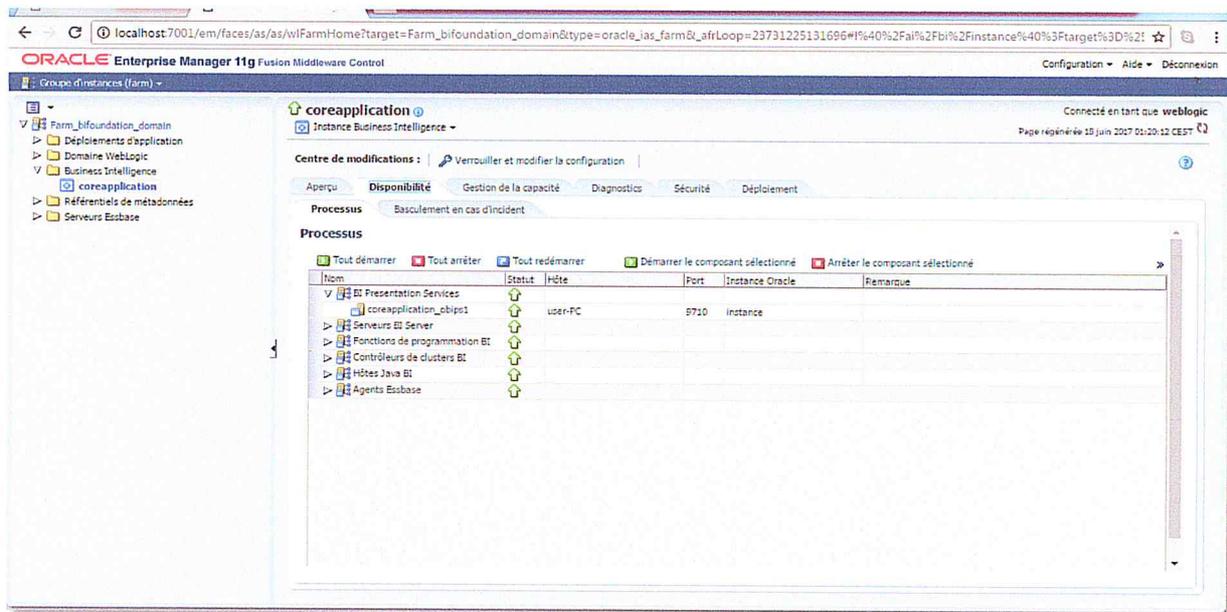


Figure A08.7 : Le redémarrage des services du BI.

*Annexe 09 : Préparation du OBIEE à
l'analyse spatiale*

Annexe 09 : Préparation du OBIEE à l'analyse spatiale

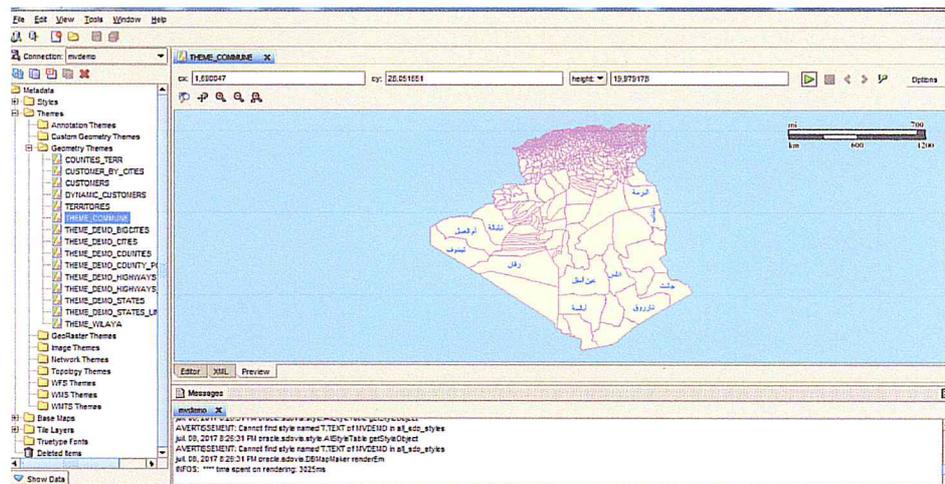


Figure A09.1 : La couche représentant les communes

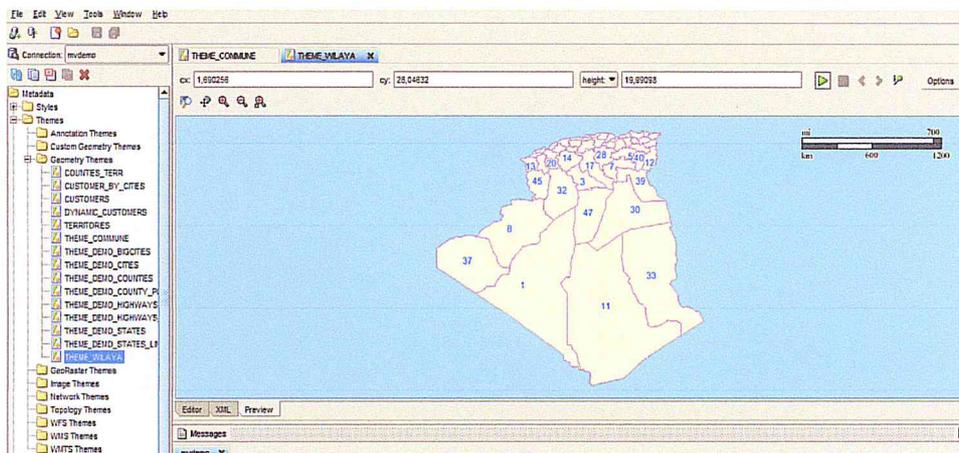


Figure A09.2 : La couche représentant les wilayas

THEMES: Created: 29/05/2017 04:08:33 Last DDL: 29/05/2017 04:08:33

NAME	DESCRIPTION	BASE_TABLE	GEOMETRY_COLUMN	STYLING_RULES
THEME_DEMO_CITIES		CITIES	LOCATION	<?xml version="1.0" stand
THEME_DEMO_COUNTIES		COUNTIES	GEOM	<?xml version="1.0" stand
THEME_DEMO_HIGHWAYS_LINE		INTERSTATES	GEOM	<?xml version="1.0" stand
THEME_DEMO_HIGHWAYS		INTERSTATES	GEOM	<?xml version="1.0" stand
COUNTIES_TERR		COUNTIES_REP	GEOM	<?xml version="1.0" stand
THEME_DEMO_STATES		STATES	GEOM	<?xml version="1.0" stand
THEME_DEMO_COUNTY_POPDENSITY	County population density	COUNTIES	GEOM	<?xml version="1.0" stand
TERRITORIES		TERRITORIES	GEOMETRY	<?xml version="1.0" stand
THEME_DEMO_BIGCITIES		CITIES	LOCATION	<?xml version="1.0" stand
THEME_DEMO_STATES_LINE	Slate boundaries	STATES	GEOM	<?xml version="1.0" stand
THEME_WILAYA		LIMITES_WILAYA	GEOMETRY	<?xml version="1.0" stand
THEME_COMMUNE		COMMUNES_ALGERIE	GEOMETRY	<?xml version="1.0" stand

Figure A09.3 : Introduction de la méta données des couches des wilayas et communes

Annexe 09 : Préparation du OBIEE à l'analyse spatiale

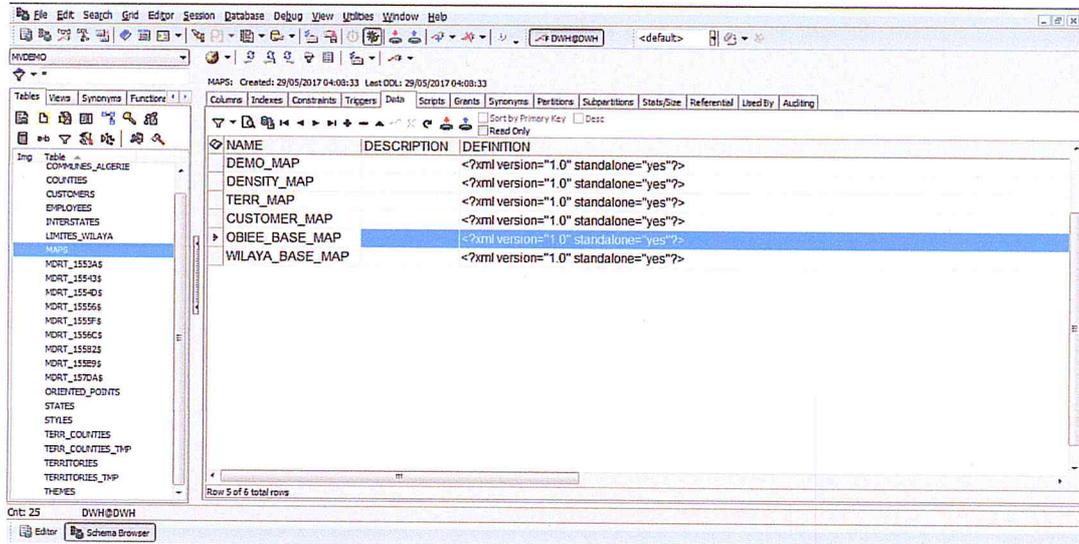


Figure A09.4 : Introduction des métas données de la carte de base

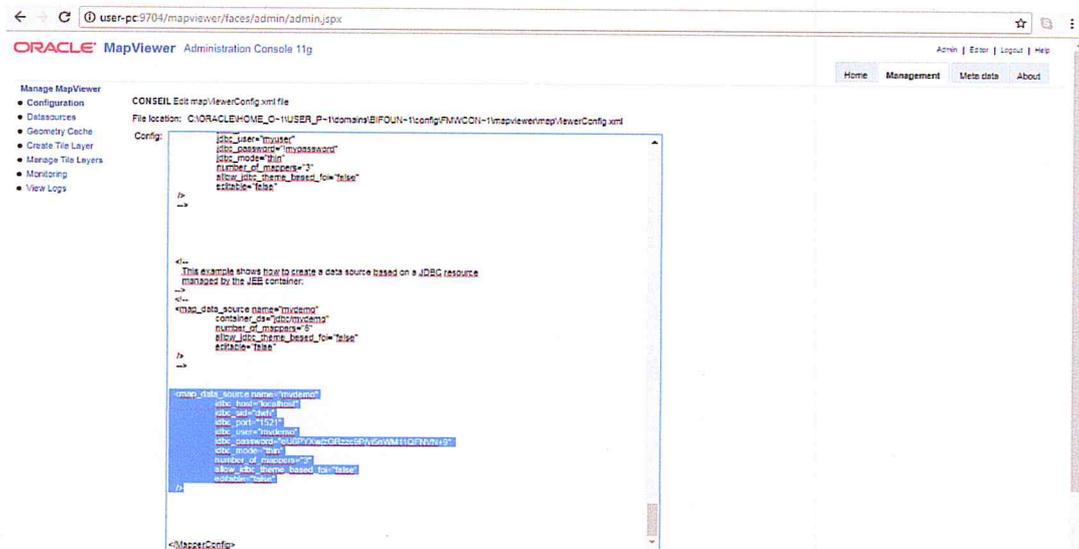


Figure A09.5 : Etablir la connexion du Oracle Mapviewer vers le schéma MVdemo

Annexe 09 : Préparation du OBIEE à l'analyse spatiale

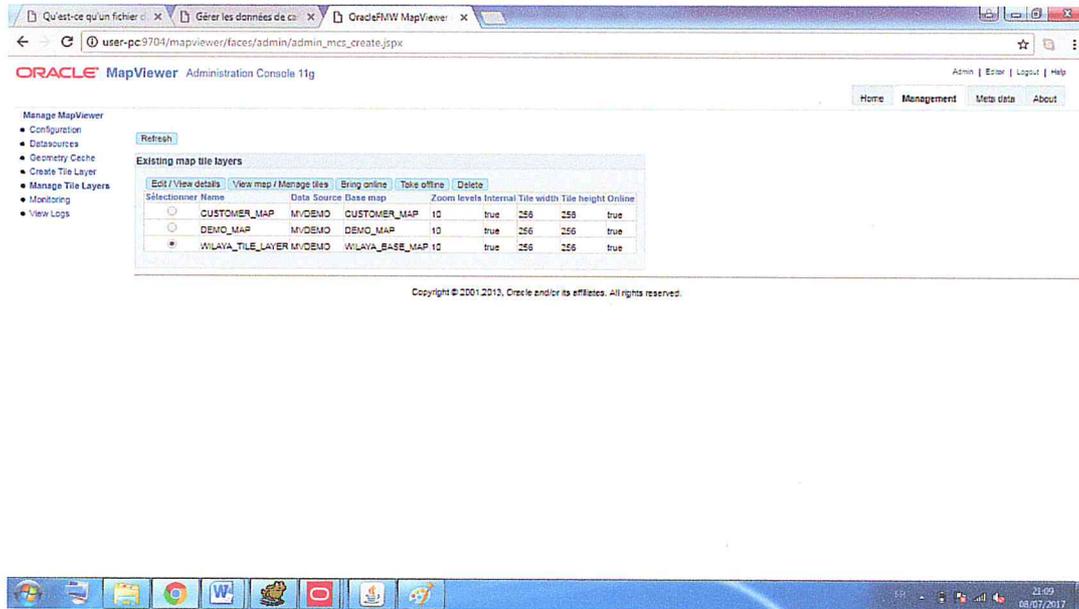


Figure A09.6 : Changer l'état de la carte de base vers « OnLine »

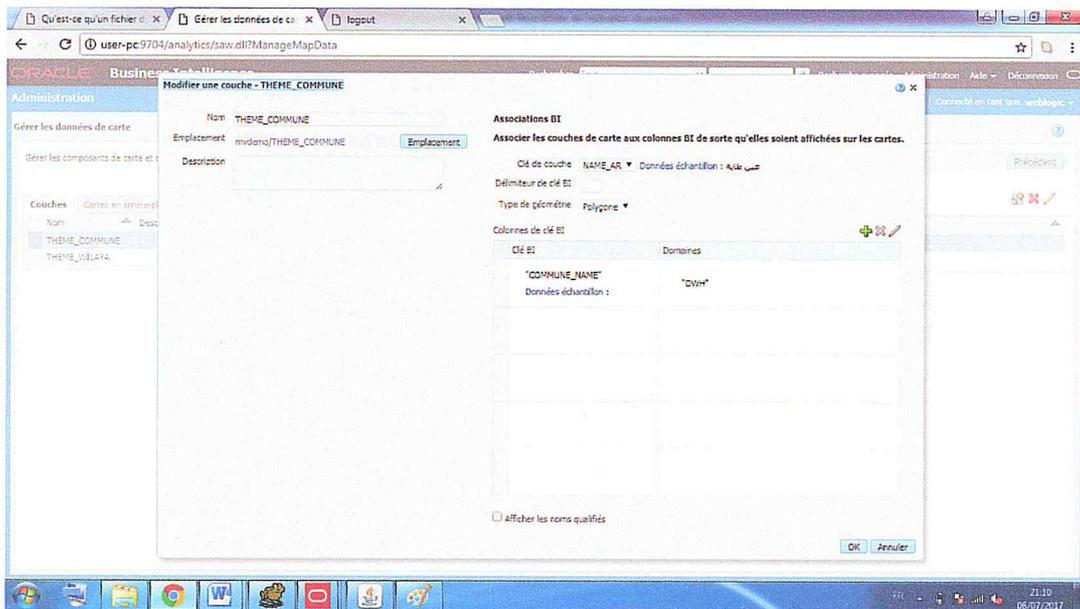


Figure A09.7 : Mise en correspondance des données de la dimension Commune_DIM avec les données attributaires de la carte correspondant à la commune

Annexe 09 : Préparation du OBIEE à l'analyse spatiale

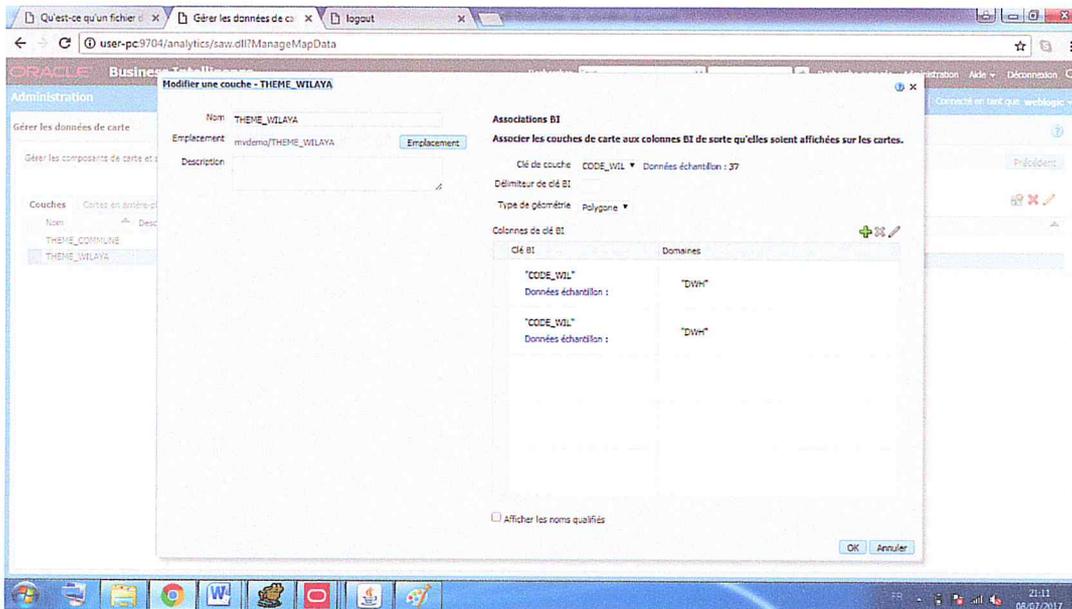


Figure A09.8 : Mise en correspondance des données de la dimension Commune_DIM avec les données attributaires de la carte correspondant à la wilaya

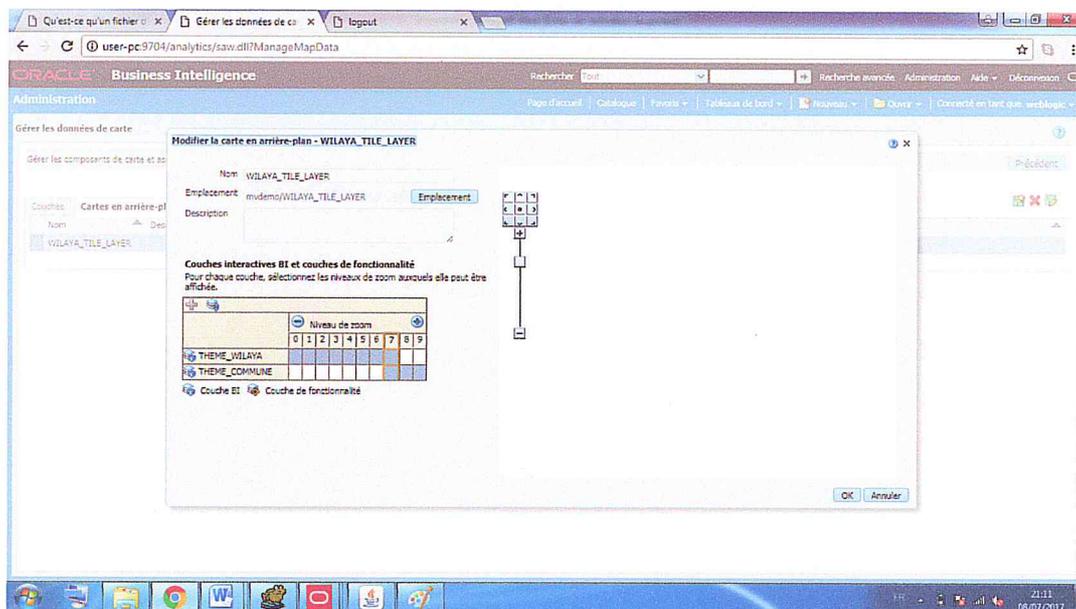


Figure A09.9 : Configuration des niveau de zoom pour la vision des couches wilaya et commune

*Annexe 10 : Carte logique entre données
sources et données cibles*

Annexe 10 : Carte logique entre données sources et données cibles

Les Dimensions					
Dimensions	Tables sources	Colonne filtre	Colonnes sources	Tables Destinations	Colonnes destinations
Type Accident (TYPE_ACCIDENT_DIM)	ACCIDENT	N/A	TYPE_ACC	TYPE_ACCIDENT_DIM_TAB	NAME DESCRIPTION
Type de Collision (COLLISION_TYPE_DIM_DIM)	ACCIDENT	N/A	COLLISION_TYPE	COLLISION_TYPE_DIM_TAB	NAME DESCRIPTION
Construction (CONSTRUCTION_DIM)	TMASTCODE	TYPE=''	NAME	CONSTRUCTION_DIM_TAB	NAME DESCRIPTION
Commune (COMMUNE_DIM)	TCOMMUNE	N/A	LIBA_COM	COMMUNE_DIM_TAB	COMMUNE_NAME COMMUNE_DESCRIPTION
	TDAIRA	N/A	LIBA_DAI	COMMUNE_DIM_TAB	DAIRA_NAME DAIRA_DESCRIPTION
	TSTATES	N/A	CODE_WIL	COMMUNE_DIM_TAB	WILAYA_NAME WILAYA_DESCRIPTION
Luminosité (ILLUMINATION_DIM)	TACCIDENT_CODES	FIELDNAME='ILLUMINATION'	NAME	ILLUMINATION_DIM_TAB	NAME DESCRIPTION
Caractéristiques de la route (ROAD_CHARACTER_DIM)	TACCIDENT_CODES	FIELDNAME='ROAD_CHARACTER'	NAME	ROAD_CHARACTER_DIM_TAB	NAME DESCRIPTION
	TACCIDENT_CODES	FIELDNAME='ROAD_CONDITION'	NAME	ROAD_CONDITION_DIM_TAB	NAME DESCRIPTION
Spécificités de la route (ROAD_SPECIAL_FEATURE_DIM)	TACCIDENT_CODES	FIELDNAME='ROADSPECIAL_FEATURE'	NAME	ROAD_SPECIAL_FEATURE_DIM_TAB	NAME DESCRIPTION
	TACCIDENT_CODES	FIELDNAME='TRAFFIC_FLOW'	NAME	TRAFFIC_FLOW_DIM_TAB	NAME DESCRIPTION
Orientations du trafic routier (TRAFFIC_FLOW_DIM)	TACCIDENT_CODES	FIELDNAME='TRAFFIC_LEVEL'	NAME	TRAFFIC_LEVEL_DIM_TAB	NAME DESCRIPTION
	TACCIDENT_CODES	N/A	TYPE_ACC	NATURE_ACCIDENT_DIM_TAB	NAME DESCRIPTION
Densité du Trafic (TRAFFIC_LEVEL_DIM)	TACCIDENT_CODES	FIELDNAME='ROADWAY_SURFACE'	NAME	ROADWAY_SURFACE_DIM_TAB	NAME DESCRIPTION
	TACCIDENT_CODES	FIELDNAME='WEATHER'	NAME	WEATHER_DIM_TAB	NAME DESCRIPTION

Annexe 10 : Carte logique entre données sources et données cibles

Cause de l' Accident (CAUSE_DIM)	CAUSE_ACCIDENT	N/A	LIBA_CCIRC	CAUSE_DIM_TAB	CAUSE_NAME CAUSE_DESCRIPTION
	TYPE_CAUSE_ACCIDENT	N/A	LIBA_TYPE	CAUSE_DIM_TAB	TYPE_CAUSE_NAME TYPE_CAUSE_DESCRIPTION
Nationalité (CITYZENSHIP_DIM)	TMASTLOCATIONCODE	TYPE='CTRY'	NAME	CITYZENSHIP_DIM_TAB	NAME DESCRIPTION
	TACCIDENT_CODE	FIELDNAME='TRANCHE'	NAME	TRANCHE_AGE_DIM_TAB	NAME DESCRIPTION
Tranche d'âge (TRANCHE_AGE_DIM)	TMASTCODE	TYPE='SEX'	NAME	SEX_DIM_TAB	NAME DESCRIPTION
	TACCIDENT_CODE	TYPE='EDUD'	NAME	EDUCATION_DIM_TAB	NAME DESCRIPTION
Sexe (SEX_DIM)	TACCIDENT_CODE	FIELDNAME='PROPRIETAIRE'	NAME	PROPRIETAIRE_DIM_TAB	NAME DESCRIPTION
	TACCIDENT_CODE	FIELDNAME='PERMIS_ON'	NAME	PERMIS_ON_DIM_TAB	NAME DESCRIPTION
Niveau d'éducation (EDUCATION_DIM)	TACCIDENT_CODE	FIELDNAME='TYPE_CHAUFFEUR'	NAME	TYPE_CHAUFFEUR_DIM_TAB	NAME DESCRIPTION
	TACCIDENT_CODE	N/A	CLASS	INFRACTION_DIM	CLASS NAME CLASS DESCRIPTION
Propriétaire (PROPRIETAIRE_DIM)	TORDS	N/A	ORD_SECT	INFRACTION_DIM	ORD_SECT NAME ORD_SECT DESCRIPTION
	TACCIDENT_CODE	N/A	INFRACTION	INFRACTION_DIM	NAME DESCRIPTION
Port Permis de Conduire (PERMIS_ON_DIM)	TACCIDENT_CODE	FIELDNAME='NATURE_VICTIME'	NAME	NATURE_VICTIME_DIM_TAB	NAME DESCRIPTION
	TACCIDENT_CODE	FIELDNAME='TYPE_VICTIME'	NAME	TYPE_VICTIME_DIM_TAB	NAME DESCRIPTION
Type Chauffeur (TYPE_CHAUFFEUR_DIM)	TACCIDENT_CODE	FIELDNAME='DAMAGE'	NAME	Vehicle_Damage_Dim_Tab	NAME DESCRIPTION
	TACCIDENT_CODE	TYPE='VHTP'	NAME	VEHICLE_TYPE_DIM_TAB	NAME DESCRIPTION
Nature Victime (NATURE_VICTIME_DIM)	TVEH_CODE	N/A	BRAND_NAME	VEHICLE_MAKE_DIM_TAB	NAME
	TACCIDENT_CODE	N/A			

Annexe 10 : Carte logique entre données sources et données cibles
(VEHICLE_MAKE_DIM)

Table de fait	Tables sources	Colonne filtre	Colonnes sources	Tables Destinations	Colonnes destinations	DESCRIPTION
HOURE (HOURE_DIM)	HOURE.TXT	N/A	HOURE	HOURE_DIM_TAB	ACCIDENT_REPORTNUM_DIM	NAME DESCRIPTION
TABLE DES FAITS						
Accident	ACCIDENT	N/A	ACCIDENT_REPORTNUM	ACCIDENT_HEADER_CUBE_TAB	ACCIDENT_REPORTNUM_DIM	
			ACCIDENT_DATE	ACCIDENT_HEADER_CUBE_TAB	DAY_NAME_DIM	
			CREATED_ON	ACCIDENT_HEADER_CUBE_TAB		
			ROAD_CHARACTER	ACCIDENT_HEADER_CUBE_TAB	ROAD_CHARACTER_DIM	
			ROAD_CONDITION	ACCIDENT_HEADER_CUBE_TAB	ROAD_CONDITION_DIM	
			ROADWAY_SURFACE	ACCIDENT_HEADER_CUBE_TAB	ROADWAY_SURFACE_DIM	
			COLLISION_TYPE	ACCIDENT_HEADER_CUBE_TAB		
			ILLUMINATION	ACCIDENT_HEADER_CUBE_TAB	ILLUMINATION_DIM	
			WEATHER	ACCIDENT_HEADER_CUBE_TAB	WEATHER_DIM	
			DIS_COM	ACCIDENT_HEADER_CUBE_TAB	COMMUNE_DIM	
			CREATED_BY	ACCIDENT_HEADER_CUBE_TAB		
			FATALITY_NUM	ACCIDENT_HEADER_CUBE_TAB	NB_FATALITY	
			CONST_MAINT_ZONE_ON_URI	ACCIDENT_HEADER_CUBE_TAB	CONSTRUCTION_DIM	
			ROAD_SPECIAL_FEATURE_URI	ACCIDENT_HEADER_CUBE_TAB	ROAD_SPECIAL_FEATURE_DIM	
			INJURY_NUM	ACCIDENT_HEADER_CUBE_TAB	NB_INJURY	
			NUM_WITNESSES	ACCIDENT_HEADER_CUBE_TAB	NB_WITNESS	
			TRAFFIC_LEVEL	ACCIDENT_HEADER_CUBE_TAB	TRAFFIC_LEVEL_DIM	
			TRAFFIC_FLOW	ACCIDENT_HEADER_CUBE_TAB	TRAFFIC_FLOW_DIM	
			HEURE_ACC	ACCIDENT_HEADER_CUBE_TAB	HOURE_DIM	
			TYPE_ACC	ACCIDENT_HEADER_CUBE_TAB	NATURE_ACCIDENT_DIM	
TYPE_CMG	ACCIDENT_HEADER_CUBE_TAB	TYPE_ACCIDENT_DIM				
CHAUFFEUR	ACCIDENT	N/A	ACCIDENT_REPORTNUM	ACCIDENT_HEADER_CUBE_TAB	ACCIDENT_REPORTNUM_DIM	
			ACCIDENT_DATE	ACCIDENT_HEADER_CUBE_TAB	DAY_NAME_DIM	

Annexe 10 : Carte logique entre données sources et données cibles

INFRACTION	ACCIDENT	N/A	ACCIDENT_REPORTN_UM	INFRACTION_CUBE_TAB	ACCIDENT_REPORTNUM_DIM
			ACCIDENT_DATE	INFRACTION_CUBE_TAB	DAY_NAME_DIM
			CREATED_ON	INFRACTION_CUBE_TAB	
			ROAD_CHARACTER	INFRACTION_CUBE_TAB	ROAD_CHARACTER_DIM
			ROAD_CONDITION	INFRACTION_CUBE_TAB	ROAD_CONDITION_DIM
			ROADWAY_SURFACE	INFRACTION_CUBE_TAB	ROADWAY_SURFACE_DIM
			COLLISION_TYPE	INFRACTION_CUBE_TAB	
			ILLUMINATION	INFRACTION_CUBE_TAB	ILLUMINATION_DIM
			WEATHER	INFRACTION_CUBE_TAB	WEATHER_DIM
			DIS_COM	INFRACTION_CUBE_TAB	
			CREATED_BY	INFRACTION_CUBE_TAB	COMMUNE_DIM
			FATALITY_NUM	INFRACTION_CUBE_TAB	NB_FATALITY
			CONST_MAINT_ZONE_ON	INFRACTION_CUBE_TAB	CONSTRUCTION_DIM
			ROAD_SPECIAL_FEAT_URI	INFRACTION_CUBE_TAB	ROAD_SPECIAL_FEATUR_DIM
			INJURY_NUM	INFRACTION_CUBE_TAB	NB_INJURY
			NUM_WITNESSES	INFRACTION_CUBE_TAB	NB_WITNESS
			TRAFFIC_LEVEL	INFRACTION_CUBE_TAB	TRAFFIC_LEVEL_DIM
			TRAFFIC_FLOW	INFRACTION_CUBE_TAB	TRAFFIC_FLOW_DIM
			HEURE_ACC	INFRACTION_CUBE_TAB	HOURE_DIM
			TYPE_ACC	INFRACTION_CUBE_TAB	NATURE_ACCIDENT_DIM
			TYPE_CMG	INFRACTION_CUBE_TAB	TYPE_ACCIDENT_DIM
			CAUSE	INFRACTION_CUBE_TAB	CAUSE_DIM
			ORD_CODE	INFRACTION_CUBE_TAB	INFACRTION_DIM
ORD_SECT	INFRACTION_CUBE_TAB	INFACRTION_DIM			
DISP_DESC		AMOUNT			
AMOUNT		NB_INFRACTION			
ACCIDENT_CAUSE	N/A				
ACCIDENT_CITN	N/A				
TORDS	N/A				
VEHICLE	ACCIDENT	N/A	ACCIDENT_REPORTN_UM	VEHICLE_CUBE_TAB	ACCIDENT_REPORTNUM_DIM
			ACCIDENT_DATE	VEHICLE_CUBE_TAB	DAY_NAME_DIM
			CREATED_ON		

Annexe 10 : Carte logique entre données sources et données cibles

			ROAD_CHARACTER	VICTIME_CUBE_TAB	ROAD_CHARACTER_DIM
			ROAD_CONDITION	VICTIME_CUBE_TAB	ROAD_CONDITION_DIM
			ROADWAY_SURFACE	VICTIME_CUBE_TAB	ROADWAY_SURFACE_DIM
			COLLISION_TYPE	VICTIME_CUBE_TAB	
			ILLUMINATION	VICTIME_CUBE_TAB	ILLUMINATION_DIM
			WEATHER	VICTIME_CUBE_TAB	WEATHER_DIM
			DIS_COM	VICTIME_CUBE_TAB	COMMUNE_DIM
			CREATED_BY	VICTIME_CUBE_TAB	
			FATALITY_NUM	VICTIME_CUBE_TAB	NB_FATALITY
			CONST_MAINT_ZONE	VICTIME_CUBE_TAB	CONSTRUCTION_DIM
			ROAD_SPECIAL_FEATURE	VICTIME_CUBE_TAB	ROAD_SPECIAL_FEATURE_DIM
			INJURY_NUM	VICTIME_CUBE_TAB	NB_INJURY
			NUM_WITNESSES	VICTIME_CUBE_TAB	NB_WITNESS
			TRAFFIC_LEVEL	VICTIME_CUBE_TAB	TRAFFIC_LEVEL_DIM
			TRAFFIC_FLOW	VICTIME_CUBE_TAB	TRAFFIC_FLOW_DIM
			HEURE_ACC	VICTIME_CUBE_TAB	HEURE_DIM
			TYPE_ACC	VICTIME_CUBE_TAB	NATURE_ACCIDENT_DIM
			TYPE_CMG	VICTIME_CUBE_TAB	TYPE_ACCIDENT_DIM
			CAUSE	VICTIME_CUBE_TAB	CAUSE_DIM
			INJURY_SEVERITY	VICTIME_CUBE_TAB	INJURY_SEVERITY_DIM
			CMG_VIC	VICTIME_CUBE_TAB	NATURE_VICTIME_DIM
			INJURY_AREA	VICTIME_CUBE_TAB	INJURY_AREA_DIM
			INJURY_CAUSE	VICTIME_CUBE_TAB	INJURY_CAUSE_DIM
			LNAME	VICTIME_CUBE_TAB	LNAME et FNAME seront utilisés pour servir de jointure avec la table MASTNAME afin de ramener la date de naissance et de calculer la tranche d'age
			FNAME	VICTIME_CUBE_TAB	
			NAME_REC	VICTIME_CUBE_TAB	TRANCHE_AGE
ACCIDENT_CAUSE		N/A			
ACCIDENT_OCCUPANTS		N/A			

Annexe 10 : Carte logique entre données sources et données cibles

ACCIDENT CHAUFFEUR	N/A	INJURY_SEVERITY	VICTIME_CUBE_TAB	INJURY_SEVERITY_DIM	LNAME et FNAME seront utilisés pour servir de jointure avec la table MASTNAME afin de ramener la date de naissance et de calculer la tranche d'age
		CMG_CHAUFF	VICTIME_CUBE_TAB	NATURE_VICTIME_DIM	
		INJURY_AREA	VICTIME_CUBE_TAB	INJURY_AREA_DIM	
		INJURY_CAUSE	VICTIME_CUBE_TAB	INJURY_CAUSE_DIM	
		LNAME	VICTIME_CUBE_TAB	TRANCHE_AGE	
		FNAME	VICTIME_CUBE_TAB		
		NAME_REC	VICTIME_CUBE_TAB		
		INJURY_SEVERITY	VICTIME_CUBE_TAB	INJURY_SEVERITY_DIM	
		CMG_VIC	VICTIME_CUBE_TAB	NATURE_VICTIME_DIM	
		INJURY_AREA	VICTIME_CUBE_TAB	INJURY_AREA_DIM	
ACCIDENT PEDESTRIANS	N/A	INJURY_SEVERITY	VICTIME_CUBE_TAB	INJURY_SEVERITY_DIM	LNAME et FNAME seront utilisés pour servir de jointure avec la table MASTNAME afin de ramener la date de naissance et de calculer la tranche d'age
		CMG_VIC	VICTIME_CUBE_TAB	NATURE_VICTIME_DIM	
		INJURY_AREA	VICTIME_CUBE_TAB	INJURY_AREA_DIM	
		INJURY_CAUSE	VICTIME_CUBE_TAB	INJURY_CAUSE_DIM	
		LNAME	VICTIME_CUBE_TAB	TRANCHE_AGE	
		FNAME	VICTIME_CUBE_TAB		
		NAME_REC	VICTIME_CUBE_TAB		
		INJURY_SEVERITY	VICTIME_CUBE_TAB	INJURY_SEVERITY_DIM	
		CMG_VIC	VICTIME_CUBE_TAB	NATURE_VICTIME_DIM	
		INJURY_AREA	VICTIME_CUBE_TAB	INJURY_AREA_DIM	

Tableau A10.1 : Tableau des données sources et leurs destinations

Références Bibliographiques :

Les livres

[Adamson/Venerable1998]

« Datawarehouse Design Solutions » ;
C.Adamson, M.Venerable ; John Wiley ans Sons ; 1998.

[English2004]

« Improving Datawarehouse and Business Information Quality : Methods for Reducing Costs and Increasing Profits » ;
L.P. English ; John Wiley and Sons ; 2004.

[Franco/Lignerolles2000]

« Piloter l'Entreprise Grâce au Datawarehouse » ;
J.M Franco, S. De Lignerolles ; Eyrolles ; 2000.

[Goglin1998]

« La Construction du Datawarehouse : du Datamart au Dataweb » ;
J.F Goglin ; Hermes ; 1998.

[IBM1998]

« Data Modeling Techniques for Data Warehousing »
C. Ballard, D. Herreman, D. Schau, R.Bell, A.Valencic ; IBM Redbooks ; 1998.

[Kimble/Ross2002]

« The Datawarehouse Toolkit : The Complete Guide to Dimensional Modeling, 2nd Edition » ;
R. Kimble, M. Ross ; John Wiley and Sons ; 2002.

[Kimble1999]

« Concevoir et Déployer un Datawarehouse : Guide de Conduite de Projet »
R. Kimble, L. Reeves, M. Ross, W. Thornwaite ; Eyrolles ; 2000.

[Rud2001]

« The Datamining Cookbook » ;
O. Parr Rud ; John Wiley and Sons ; 2001.

Articles et thèses :

[Chaudhri/Dayal1997]

« An Overview of Datawarehousing and OLAP Technology » ;
S. Chaudhuri, U. Dayal ; ACM SIGMOD Record ; 1997.

[Codd1993]

« Providing OLAP to User-Analysts : an IT mandate » ;
E.F. Codd ; Technical Reports, E.F. Codd and Associates ; 1993.

Bibliographie

[Dayal/Wuu1992]

« A Uniform Approach to Processing Temporal Queries » ;

Y. Dayal, G.T.J. Wuu ; International Conference on Very Large Databases, Canada, 1992.

[Filali/Kedjnane2010]

« Conception et Réalisation d'un Data Warehouse pour la Mise en Place d'un Système Décisionnel »

A.Filali, S.Kedjnane ; Thèse d'Ingénieur, Institut National de Formation en Informatique, El-Harrache (Algérie) ; 2010.

[Haciane2006]

« Conception d'un Datawarehouse Orienté CRM »

A.Haciane ; Thèse de Magistère, Institut National de Formation en Informatique, El-Harrache (Algérie) ; 2006.

[Teste2000]

« Modélisation et Manipulation d'Entrepôts de Données Complexes et historisées » ;

O. Teste ; Thèse de doctorat, Université P. Sabatier, Toulouse (France) ; 2000.

