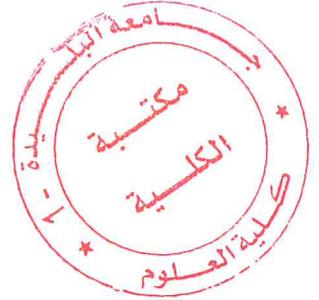


MA.004-439-1

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE Ministère de l'enseignement
supérieur et de la recherche scientifique
Université Saad Dahlab Blida1



Faculté des Sciences
Département d'Informatique



Mémoire de fin d'étude
Pour l'obtention d'un diplôme de Master en Informatique
Spécialité : Système Informatiques et Réseaux

Thème

Plateforme d'ETL pour le traitement de flux de donnée continus en temps réel

Réalisé par :

SECRAOUI Billel & BENREKIA Chawki

Encadré par:

SAKILANI Mohamed

Organisme d'accueil : Optimum Télécom Algérie (Djezzy)

Jury :

Mme. OUAHRA NI Leila (Présidente) Maitre Assistante (A)

Mme. OUKID Lamia Maitre de conférence (B)

M. BALA Mahfoud (Promoteur) Maitre de conférence (B)

Soutenu le : 27/10/2018

MA-004-439-1

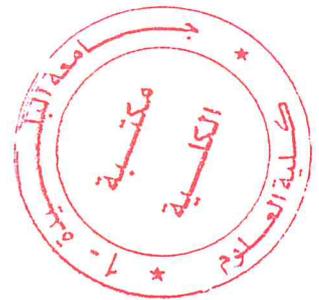
Résumé

Le processus ETL constitue l'usine qui produit l'information pertinente ayant la valeur et la qualité requises exigée par la suite décisionnelle. Toutes les phases qui suivent constituant la chaîne décisionnelle consomment l'information livrée par ce processus d'extraction, de transformation et de chargement.

Pour comprendre et maîtriser le flux de données ainsi que la séquence des tâches de filtrage, nettoyage, conversions, intégration, agrégation et chargement, nous nous intéressons dans le cadre de ce projet de fin d'études à une solution de business intelligence dédiée pour l'analyse des indicateurs clés de performances KPI de l'infrastructure télécom et d'aide à la décision fonctionnant dans un environnement caractérisé essentiellement par des données qui arrivent de façon continue et en temps quasi-réel.

Pour ce faire nous avons choisi de mettre en place une plateforme ELT, cette solution se compose d'une partie d'intégration et de chargement des données bruts avec l'outil PDI et d'une partie transformation faite par SQL server, enfin une application java pour l'analyse des données chargées.

Mot clés : *ETL (Extract Transform load), ELT (Extract Load Transform), PDI (Pentaho Data Integration), KPI (Key Performance Indicator), SQL server (Structured Query Language).*



Abstract

The ETL process is the plant that produces the relevant information with the required value and quality required by the decision-making suite. All the phases that follow constitute the decision-making chain and consume the information delivered by this process of extraction, transformation and loading.

To understand and control the flow of data and the sequence of filtering, cleaning, conversions, integration, aggregation and loading tasks, we are interested in this end of studies project to a dedicated business intelligence solution for analysis of KPI key performance indicators of the telecom and decision support infrastructure operating in an environment characterized mainly by data arriving on a continuous basis and in near real time.

To do this we chose to set up an ELT platform, this solution consists of a part of integration and raw data loading with the PDI tool and a transformation part made by SQL server, finally an application with Java for the analysis of the loaded data.

Key words : *ETL (Extract Transform load), ELT (Extract Load Transform), PDI (Pentaho Data Integration), KPI (Key Performance Indicator), SQL server (Structured Query Language).*

Remerciements

En premier lieu, nous remercions Allah tout puissant de nous avoir donné la force et le courage pour mener à terme ce projet de fin d'étude.

Au terme de ce travail, nous tenons à exprimer notre profonde gratitude à notre cher promoteur M. Mahfoud BALA pour son suivi et pour son énorme soutien, qu'il n'a cessé de nous prodiguer tout au long de la période du projet.

Nous tenons à remercier également notre encadrant M. Mohamed SAKILANI, pour le temps qu'il a consacré et pour les précieuses informations qu'il nous a prodiguées avec intérêt et compréhension.

Nos remerciements vont à tout le personnel que nous avons contacté durant notre stage au sein du service Qualité de Djedda, nous citons : M. Mohamed BETATACHE manager au sein du service Qualité de Djedda, Mme Meriem BELKADI Directrice au sein du service qualité de Djedda, auprès des quelles nous avons trouvé l'accueil chaleureux, l'aide et l'assistance dont nous avons besoin.

Nous adressons aussi nos vifs remerciements aux membres des jurys pour avoir bien voulu examiner et juger ce travail.

Je ne laisserai pas cette occasion passer, sans remercier tous les enseignants de l'USDB pour leurs connaissances et leur expérience qu'ils ont partagées avec nous durant nos années d'étude.

Nous remercions énormément nos familles et nos amis qui nous ont, et continuent à nous soutenir.

Enfin, nous n'oublions pas de remercier toute autre personne qui, de près ou de loin, d'une manière ou d'une autre, a contribué à mener à bien ce travail.

Table Des Matières

1	CHAPITRE : INTRODUCTION GENERALE	13
1.1	CONTEXTE DU TRAVAIL :	13
1.1.1	GSM :	14
1.1.2	UMTS :	15
1.2	PROBLEMATIQUE :	17
1.3	OBJECTIFS :	17
1.4	STRUCTURATION DU MEMOIRE	17
2	CHAPITRE : ETAT DE L'ART SUR LES SYSTEMES D'INFORMATIONS DECISIONNELS	19
2.1	INTRODUCTION.....	19
2.2	LES SYSTEMES D'INFORMATIONS DECISIONNELS :	19
2.2.1	Définition :	19
2.2.2	Architecture :	19
	▪ Phase de collecte ou d'alimentation	20
	▪ Phase d'intégration	20
	▪ Phase de distribution	20
	▪ Phase de restitution	20
2.3	DATA WAREHOUSE :	21
2.3.1	Définition	21
2.3.2	Ancêtre du Data Warehouse :	21
2.3.3	Datamart :	21
2.3.4	Intérêt des Data Warehouse dans l'entreprise :	22
2.3.5	Caractéristiques du Data Warehouse :	22
2.4	ALIMENTATION DU DATA WAREHOUSE PAR LES OUTILS ETL.....	24
2.4.1	Extraction :	25
2.4.2	Transformation.....	25
2.4.3	Chargement.....	26
2.5	MODELE MULTIDIMENSIONNEL.....	26
2.5.1	Table de faits :	26
2.5.2	Table de dimension :	27
2.6	MODELISATION	29
2.6.1	Schéma en étoile.....	29
2.6.2	Schéma en flocon de neige :	29
2.6.3	Schéma en constellation.....	30
2.7	CONCLUSION	30
3	CHAPITRE : PRESENTATION DE L'ORGANISME D'ACCUEIL	31
3.1	PRESENTATION DE DJEZZY :	31
3.1.1	Partie professionnelle :	31
	▪ Introduction	31
	▪ Processus d'évolution de Djazzy.....	31
	– Orascom Télécom Holding :	31
	– Vimpelcom.....	32
	– Orascom Telecom Algérie :	32
	▪ Organisations d'OTA.....	34
	▪ Missions et objectifs d'Optimum Télécom Algérie :	34
	▪ Autorité de Régulation de la Poste et des Télécommunications (ARPT)	Erreur ! Signet non défini.
	– Missions de l'ARPT :	Erreur ! Signet non défini.
3.1.2	Partie technique	35
	▪ La norme Global System for Mobile (GSM).....	35
	– Introduction :	35
	– Historique :	35

—	Le Principe cellulaire:	35
—	Architecture des réseaux GSM :	36
•	Le sous-système radio BSS :	37
•	Le sous-système réseau NSS :	38
•	Le sous-système d'exploitation OSS :	38
—	Les interfaces réseaux:	39
—	Gestion de localisation :	39
—	Sécurité du réseau GSM :	40
—	Interface Radio :	40
•	L'allocation des fréquences :	40
•	Le système d'accès multiples :	41
•	Les canaux dédiés.....	43
—	Conclusion :	43
▪	La norme UMTS.....	44
—	Introduction:	44
—	Historique :	44
—	Définition de l'UMTS:	44
—	Le principe de W-CDMA :	44
—	Les services de l'UMTS:	45
—	Architecture de l'UMTS :	45
•	EU (Equipement Usager) :	46
•	Réseau d'accès :	46
•	Réseau cœur:.....	47
-	La partie Cs (Circuit Switched) :	47
-	La partie Ps (PacketSwitched) :	48
-	Éléments communs :	48
—	Conclusion :	48
3.2	PRESENTATION DE LA STRUCTURE CONCERNEE PAR LE PROJET	48
3.2.1	<i>Présentation fonctionnelle</i> :	48
3.2.2	<i>Présentation organisationnelle</i> :	50
3.3	PROFILS INTERVENANT DANS LE SUIVI DE LA QUALITE (PERSONNES / COMPETENCES)	50
3.4	MOYENS MATERIELS ET OUTILS LOGICIELS UTILISES.....	50
3.4.1	<i>Moyens matériels</i> :	50
▪	Machine :	50
▪	Serveur de données :	51
3.4.2	<i>Outils logiciels</i> :	51
3.5	DESCRIPTION DES FICHIERS SOURCE UTILISES POUR L'ANALYSE DE LA QUALITE DU RESEAU :	51
3.6	DESCRIPTION DU PROCESSUS D'INVESTIGATION POUR L'IDENTIFICATION DES PROBLEMES DE DEGRADATION DE LA QUALITE DU RESEAU : ..	52
3.7	PROBLEMES ET LIMITES DU PROCESSUS ACTUEL :	52
3.8	OBJECTIFS ET BESOIN FUTURS :	53
4	CHAPITRE : SPECIFICATION DES BESOINS & CONCEPTION DU SYSTEME	54
4.1	INTRODUCTION.....	54
4.2	LANGAGE DE MODELISATION	54
4.2.1	<i>Le langage de modélisation UML</i> :	54
4.2.2	<i>Les diagrammes UML</i> :	54
4.3	SPECIFICATION DES BESOINS.....	55
4.3.1	<i>Les diagrammes de cas d'utilisation</i>	55
▪	Définition des diagrammes des cas d'utilisation	55
4.3.2	<i>Les acteurs du système</i>	56
4.3.3	<i>Diagramme de cas d'utilisation du système globale</i>	56
▪	Diagramme.....	56
▪	Description textuelle de chaque cas d'utilisation	57
4.4	MODELISATION DE L'ENTREPOT DE DONNEES.....	57
4.4.1	<i>Introduction</i>	57
4.4.2	<i>Modélisation dimensionnelle</i>	58
4.5	CONCLUSION	69

5	CHAPITRE : REALISATION.....	70
5.1	INTRODUCTION.....	70
5.2	CHOIX DES TECHNOLOGIES.....	70
5.2.1	<i>Outils de développement.....</i>	<i>70</i>
▪	Pentaho Data Integration (PDI).....	70
▪	SQL Server :.....	72
▪	Netbeans :.....	73
5.2.2	<i>Langages de programmation utilisés :</i>	<i>74</i>
▪	Java :	74
5.3	PLATEFORME D'INTEGRATION ET D'ANALYSE DES INDICATEURS DE QUALITE :.....	74
5.3.1	<i>Présentations des interfaces de l'application :</i>	<i>75</i>
▪	3.1.1 Processus d'extraction des données :	75
▪	Transformation des données :	79
▪	Interface d'authentification :	81
5.4	CONCLUSION :	92
6	CONCLUSION GENERALE.....	94

Table des figures

1.1	Modélisation aide à la décision :.....	04
1.2	Architecture 2G et 3G :.....	05
1.3	Architecture 2G (GSM) :	05
1.4	Architecture 3G (UMTS) :.....	06
1.5	Relations qui existent entre la stratégie de l'opérateur, Les indicateurs de performances et le processus d'aide à la décision :.....	07
2.1	Architecture générale du système d'information décisionnelle :.....	11
2.2	Idée de base d'un entrepôt de données :.....	12
2.3	Sous ensemble logique d'un data warehouse :.....	13
2.4	Données orienté sujet dans un entrepôt de données :.....	14
2.5	Données intégrées dans un entrepôt de données :.....	14
2.6	Données non-volatile dans un entrepôt de données :.....	15
2.7	Table de faits :.....	16
2.8	Table de dimension :.....	16
2.9	Exemple de cube de données	17
2.10	Exemple d'un schéma en étoile.....	18
2.11	Exemple d'un schéma en flocon de neige :.....	19
3.1	Processus d'évolution de djezzy :.....	22
3.2	Présentation organisationnelle d'OTA :.....	23
3.3	Autorité de régulation de la poste et des télécommunications :.....	24
3.4	Découpage cellulaire d'un réseau GSM :.....	26
3.5	Architecture d'un réseau GSM :.....	27
3.6	FDMA (Frequency Division Multiple Access) :.....	32
3.7	TDMA (Time Division Multiple Access) :.....	32
3.8	La distribution des fréquences dans les cellules pour les réseaux GSM et UMTS :.....	35
3.9	Architecture du réseau UMTS :.....	36
4.1	Modèle dimensionnel relatif à l'analyse des principaux KPI par heure :.....	48
4.2	Modèle dimensionnel relatif à l'analyse des principaux KPI par jour :.....	49
4.3	Modèle dimensionnel relatif à l'analyse des principaux KPI par semaine :.....	53
4.4	Modèle dimensionnel relatif à l'analyse des principaux KPI par mois :.....	51
4.5	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Lac et Heure :.....	51
4.6	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Lac et jour :.....	52
4.7	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Lac et semaine :.....	52

4.8	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Lac et mois :.....	53
4.9	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Région et heure :.....	53
4.10	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Région et jour :.....	54
4.11	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Région et semaine :.....	54
4.12	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Région et mois :.....	55
4.13	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Réseau et heure :.....	55
4.14	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Réseau et jour :.....	56
4.15	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Réseau et semaine :.....	56
4.16	Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Réseau et mois :.....	57
4.17	Schéma en constellation :.....	58
5.1	Logo Pentaho Data Integration :.....	60
5.2	Logo SQL Server :.....	62
5.3	Logo Netbeans :.....	63
5.5	Logo Java :.....	64
5.6	Architecture du système décisionnel :.....	65
5.7	Processus principal d'extraction des données sources :.....	66
5.8	Transformation (get_date) :.....	66
5.9	Interface de paramétrage de la tâche get_from_ftp_job_SM pour la connexion au serveur FTP.....	67
5.10	Paramétrage des fichiers :.....	68
5.11	Connexion à l'entrepôt de données :.....	68
5.12	Contrôle et chargement des données sources dans l'entrepôt de données :.....	69
5.13	Interface d'authentification d'un utilisateur au moment de l'accès au système :.....	72
5.14	Interface principale du tableau de bord :.....	72
5.15	Menu principale :.....	73
5.16	Interface d'ajout des nouveaux utilisateurs :.....	74
5.17	Paramétrage de l'analyse :.....	74
5.18	Choix de la période :.....	75
5.19	<i>(3G MS init PDP context act fail)</i> :.....	76
5.20	<i>(RANAP RAB Assignment fail)</i> :.....	76
5.21	<i>(3G MS init PDP context act Success) Causes</i> :.....	77
2.22	<i>(RANAP RAB Assignment success) Causes</i> :.....	78
5.23	<i>Other SM KPI fail</i> :.....	78
5.24	KPI 1 et ces causes par région :.....	79
5.25	KPI 2 et ces causes par région :.....	80
5.26	Autres KPI par région :.....	81

Glossaire

SID systèmes d'information décisionnels, se présente en un ensemble de moyens, d'outils et de méthodes permettant de collecter, consolider, modéliser et de restituer les données de l'entreprise dans le but d'apporter une aide à la prise de décision.

ETL (ou en français ETC pour Extraction-Transformation-Chargement), ETL est un outil qui est chargé de l'extraction des données métiers, de leur mise en forme dans le cadre de l'aide à la décision et enfin de leur chargement dans un entrepôt de données.

Data Warehouse (ou entrepôt de données), est une base de données utilisée pour collecter et stocker des informations provenant d'autres bases de données.

Un **Datamart** est un sous-ensemble logique d'un DW.

OLAP (en anglais Online Analytical Processing), désigne les bases de données

Multidimensionnelles destinées à l'analyse.

GSM (en anglais Global System for Mobile), il s'agit d'un standard de téléphone dit : de seconde génération (2G).

UMTS (en anglais Universel Mobile Télécommunication System), est l'une des technologies de téléphonie mobile de troisième génération (3G).

Abréviations

BI Business Intelligence.

SID Système d'information décisionnelle.

DW (Data Warehouse) Entrepôt de données.

OTA Optimum Telecom Algérie.

PDI (Pentaho Data Integration).

figure 1.2, l'architecture 2G et 3G se compose de deux partie radio network et core network dont chacune relie plusieurs éléments.

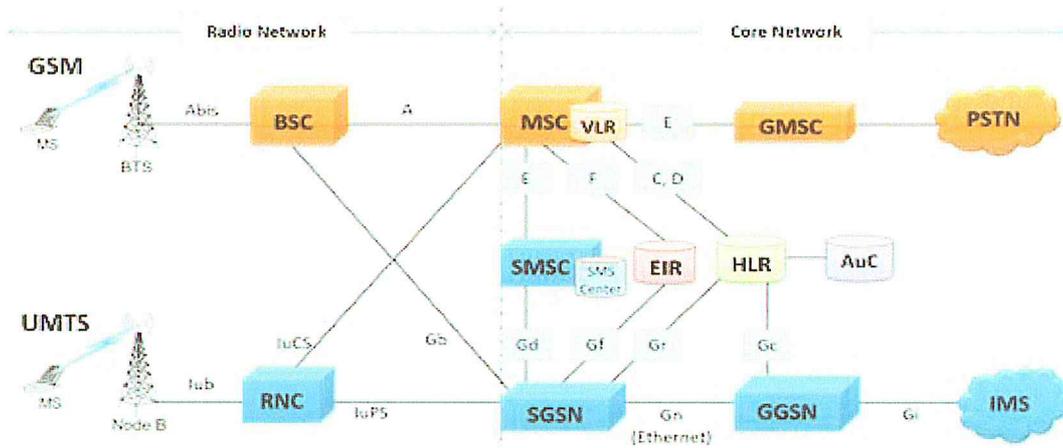


Figure 1.2: Architecture 2G et 3G. [13]

1.1.1 GSM :

La 2G ou GSM est une norme numérique pour les téléphones portables. Il s'agit, pratiquement, d'un réseau permettant une communication de type « voix », la figure 1.3 nous montre une architecture GSM. [1]

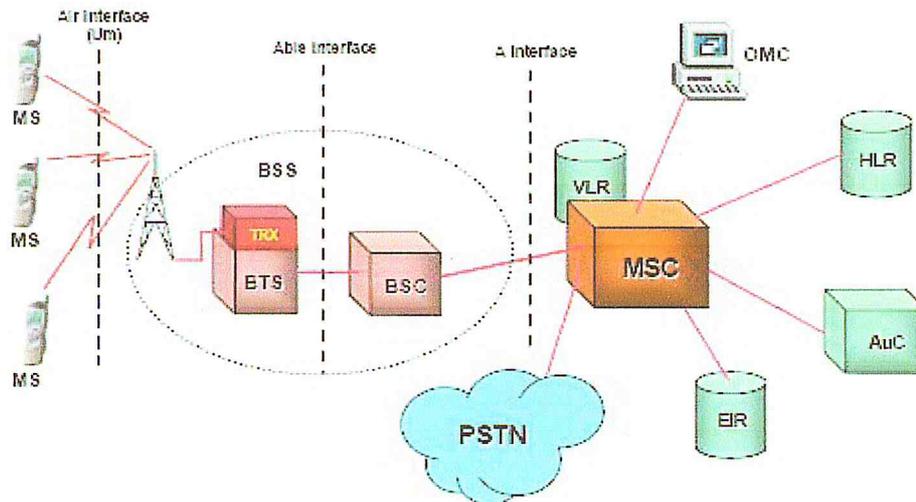


Figure 1.3 : Architecture 2G (GSM). [13]

Il est apparu dans les années 90. Son principe est de passer des appels téléphoniques, s'appuyant sur les transmissions numériques permettant une sécurisation des données (avec cryptage). Ainsi, il contient beaucoup de services tels que l'affichage d'appel, la conférence, ... Il a connu un succès et a permis de susciter le besoin de téléphoner en tout lieu avec la possibilité d'émettre des minimessages (*SMS : Short Message Service*, limités à 80 caractères).

La bande utilisée en GSM est la 900 Mhz et 1800 Mhz. Sa vitesse de transmission maximum vaut 23 kbps (théorique) et 9,6 kbps réel. Cependant, ce débit est insuffisant pour le transfert de fichiers, d'images, de vidéos, accès à Internet, etc.

Les réseaux GSM opèrent avec plusieurs identificateurs :

- L'IMSI ou International Mobile Subscriber Identifier n'est connu qu'à l'intérieur du réseau GSM. Il est d'environ 15 digits (0-9), stocké dans la carte SIM et identifie l'abonné de manière unique au niveau mondial.
- Le TMSI ou Temporary Mobile Subscriber Identity est une identité temporaire utilisée pour identifier le mobile lors des interactions station mobile/réseau. Il est formé par un mot de 04 octets représenté en hexadécimal.
- Le MSISDN ou Mobile Station ISDN est le numéro de l'abonné, c'est le seul identifiant de l'abonné mobile connu à l'extérieur du réseau GSM d'où le numéro de téléphone mobile habituel.
- Le MSRN ou Mobile Station Roaming Number est un numéro attribué lors de l'établissement d'appel. Sa principale fonction est de permettre l'acheminement des appels par les commutateurs (MSC et GMSC).
- L'IMEI ou International Mobile Equipment Identity identifie l'équipement mobile (hardware). Le numéro IMEI permet d'empêcher l'utilisation d'une station volée avec une autre carte SIM. L'IMEI est composé de 60 bits.

1.1.2 UMTS :

La 3G est un système numérique évolué de types UMTS. Il suit la recommandation IMT2000 ou International Mobile Telecommunications-2000. Cette norme européenne est pour la transmission vocale, texte, vidéo ou multimédia numérisée. Elle est basée sur une combinaison de services fixes et radio mobiles comme illustrer dans la figure 1.4. [2]

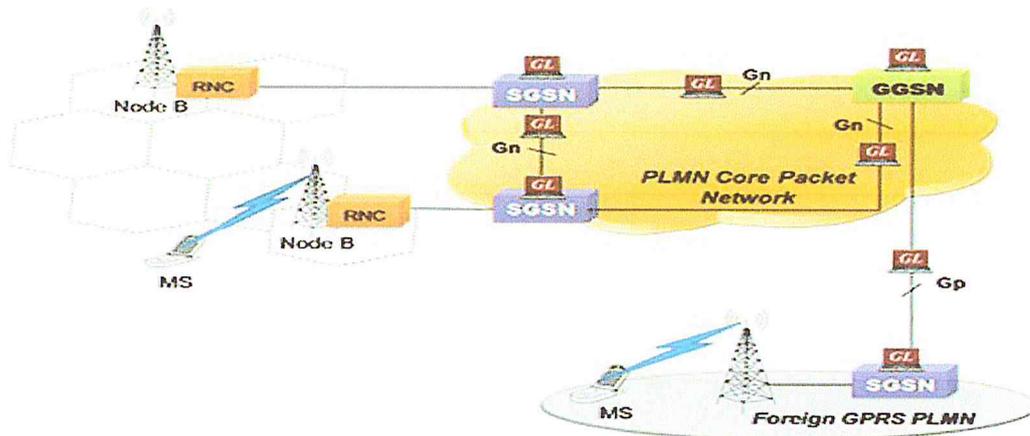


Figure 1.4 : Architecture 3G (UMTS). [13]

La téléphonie standard, l'accès à l'Internet, la téléphonie vidéo et des services spécialement adaptés tels que les actualités et les informations sur la bourse seront mis incessamment à la disposition des utilisateurs, où qu'ils soient et lorsqu'ils sont en déplacement.

Les spécifications techniques de cette norme sont développées au sein de l'organisme 3GPP. Les technologies développées autour de la norme UMTS conduisent à une amélioration significative des vitesses de transmission pouvant arriver jusqu'à 2 Mbit/s. Cette amélioration des débits est obtenue par l'évolution des technologies radio qui admet une meilleure efficacité spectrale.

En général, un réseau UMTS est constitué de trois domaines :

- L'équipement d'abonné UE ou User Equipment.
- Le réseau d'accès UTRAN ou UMTS Terrestrial Radio Access Network.
- Le réseau coeur CN ou Core Network.

L'UTRAN fournit la méthode d'accès sur l'interface constitué par l'espace libre pour les équipements d'abonné. Il détermine l'interface radio (méthode d'accès multiple et de duplexage ainsi que les paramètres correspondants)

utilisée par les UE pour accéder aux services offerts par le réseau UMTS. La station de base est dénommée Node B et l'équipement qui contrôle ce dernier est appelé RNC ou Radio Network Controller.

La principale fonction du cœur du réseau est de fournir la commutation, le routage et le transit des trafics des usagers. Le cœur du réseau contient aussi les bases de données sur les utilisateurs ainsi que les fonctions de gestion du réseau. L'architecture de base du CN a été fortement inspirée du réseau GSM avec l'extension GPRS. Cependant tous les équipements doivent être modifiés et adaptés pour les services et opérations UMTS.

Le réseau cœur de l'UMTS est composé de trois parties dont deux domaines :

- Le domaine CS ou Circuit Switched utilisé pour la téléphonie.
- Le domaine PS ou Packet Switched qui permet la commutation de paquets.
- Les éléments communs aux domaines CS et PS.

Cette infrastructure réseau télécoms ne peut faire face à un flux d'appels et de données devenu de plus en plus volumineux et instantané si le comportement des équipements vis-à-vis des abonnés n'est pas contrôlé, analysé et évalué de façon minutieuse et continue. Pour ce faire, un ensemble de compteurs (KPI) décrivant la qualité de l'infrastructure télécom face sollicitations des abonnés sont définis selon la stratégie des décideurs (top management) au sein de l'opérateur.

La mondialisation et la concurrence qu'engendre l'aide à la décision rendent le pilotage d'une organisation de plus en plus complexe. Cette complexité est liée non seulement à l'augmentation du nombre de paramètres à prendre en compte mais également à la nécessité de prises de décisions rapides afin d'être réactifs à l'évolution de la concurrence et de la demande des clients. L'efficacité de ces prises de décisions repose sur la mise à disposition d'informations fiables, pertinentes et d'outils facilitant cette tâche. Les systèmes traditionnels, dédiés à la gestion quotidienne d'une organisation, s'avèrent inadaptés à une telle activité. Face à ce besoin est né le secteur de l'informatique décisionnelle. La figure 1.5 indique la relation entre stratégies, indicateurs et décision.

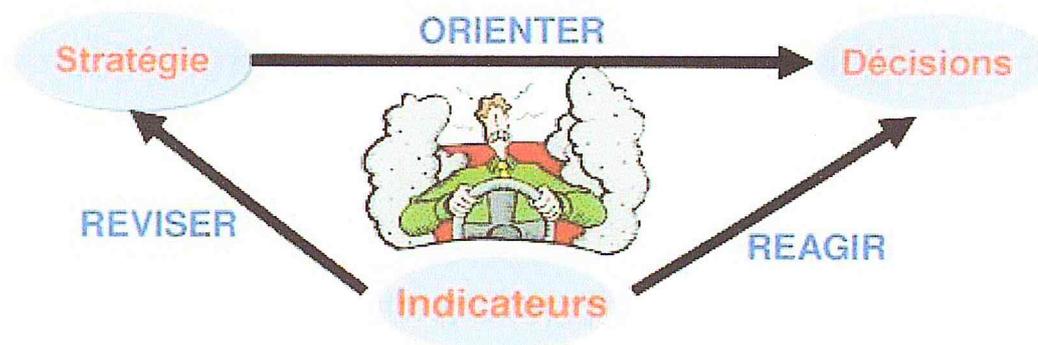


Figure 1.5 : relations qui existent entre la stratégie de l'opérateur, les indicateurs de Performances et le processus d'aide à la décision.

Généralement, une application d'aide à la décision regroupe les informations suivantes:

- Ensemble des informations en cours (y compris celles provenant de sources de données propriétaires ou relationnelles, de cubes, entrepôts de données et data marts).
- Comparaison des chiffres de ventes d'une semaine à l'autre.
- Chiffre d'affaires projeté en prévision de la vente de nouveaux produits.
- Dans un contexte bien précis, conséquences des différentes décisions possibles, compte tenu de l'expérience passée.

1.2 Problématique :

L'opérateur Optimum Telecom Algérie (OTA en abrégé) dispose aujourd'hui d'un système d'information permettant la capture de données brutes et cruciales chaque 24h, celles-ci seront, ensuite, traitées en semi-automatique par une équipe d'ingénieurs du service qualité et ce dans le but de synthétiser, quotidiennement, les événements relatifs au fonctionnement du réseau télécom sous forme de KPI de qualité. Ces indicateurs seront regroupés sous forme de rapports à communiquer au top management. Cette activité est devenue au fil des années un agrégat de tâches fastidieuses, routinières et complexes vu le volume de données traitées dans un processus semi-automatique. En outre, la période de 24H est considérée très longue pour des processus d'analyse d'un niveau critique particulièrement en cas d'événements provoquant une dégradation des performances du réseau télécom. En effet, ceci implique une perte importante en terme de temps et d'argent car la détection de problèmes au niveau de l'infrastructure réseaux exige une réaction immédiate afin d'éviter le mécontentement des abonnés et leur départ vers les concurrents.

1.3 Objectifs :

Notre objectif principal est la création d'une plateforme d'ETL pour le traitement de flux de données continue en temps quasi réel.

Nos objectifs spécifiques sont donc :

- Modélisation et création de la base de données.
- Extraction de données brute depuis un serveur FTP.
- Transformation du format des données.
- Déchiffrement des données.
- Chargement des données.
- Analyse et évaluation des compteurs (KPI) relatifs à la qualité de l'infrastructure télécom

1.4 Structuration du mémoire

Ce stage s'intéresse au développement d'une solution de business intelligence pour un operateur de téléphonie mobile, et ce mémoire synthétise le travail effectué de la manière suivante :

Le chapitre 1 commence par une introduction générale qui situe, tout d'abord, le contexte du travail, définit ensuite la problématique et les objectifs à atteindre.

Le chapitre 2 présente les fondamentaux sur la business intelligence, les processus d'Extraction, Transformation et Chargement (*ETL : Extract-Transform-Load*), traitement en temps réel et flux de données continu.

Le chapitre 3, quant à lui, présente le groupe Optimum Telecom Algérie (OTA) ainsi que le service Qualité du Réseau dans lequel s'est déroulé le stage du PFE,

La conception du système projeté est décrite dans le chapitre 4. Il s'agit de spécifier sous forme de diagrammes UML les besoins fonctionnels et techniques, le schéma des données ainsi que la séquence des traitements.

Le chapitre 5 est consacré pour la description de l'implémentation et la mise en œuvre du système au sein du Service Qualité du Réseau à Djezzy. Il s'agit de présenter l'environnement technique et outils de développement retenus pour l'implémentation et de décrire ensuite de façon sommaire les principaux traitements.

Enfin, nous clôturons ce document par une conclusion générale et un résumé sur la continuité du travail et les perspectives dans le domaine de l'amélioration de la qualité du réseau OTA.

2 Chapitre : Etat de l'art sur les systèmes d'informations décisionnels

2.1 Introduction

Dans cette ère de mondialisation que connaît l'entreprise, l'efficacité est un souci majeur pour les chefs d'établissements. La grande concurrence que connaissent les marchés impose la rapidité dans l'analyse, la planification et la réaction face aux nouvelles conditions sur ces marchés pour assurer la survie de l'entreprise. L'entreprise possède l'information. Cependant, dans la contrainte du temps qui s'impose aux décideurs, le grand volume d'information devra être réduit en des informations stratégiques pour répondre aux besoins d'analyse et de prise de décision.

Une entreprise se doit en permanence de pouvoir se situer par rapport à la concurrence, mais également par rapport à la demande et aux services qu'elle peut offrir. Quelle est l'entreprise qui ne souhaite pas fidéliser ses clients en les identifiant mieux pour leur proposer les produits ou les services appropriés, maîtriser les risques, optimiser ses activités, exploiter intelligemment les données stockées et avoir plus d'informations sur ses concurrents ?

Ces deux dernières décennies ont connu une nouvelle génération dans les systèmes d'information : Les systèmes d'information décisionnels. Ces derniers permettent de donner une vue transversale sur une grande quantité de données issues de différentes sources pour aboutir à une information stratégique.

2.2 Les systèmes d'informations décisionnels :

2.2.1 Définition :

Les *systèmes d'information décisionnels (SID)* se présentent comme un ensemble de moyens, d'outils et de méthodes permettant de collecter, consolider, modéliser et restituer les données de l'entreprise dans le but d'apporter une aide à la prise de décision. [3]

2.2.2 Architecture :

Un système décisionnel ne peut se construire en marge des autres systèmes de l'entreprise. Il nécessite une vision globale du système d'information opérationnel.

On regroupe donc, dans un référentiel commun, des éléments provenant de plusieurs applications orientées métier et représentant la mémoire de l'entreprise.

On parle généralement d'architecture ou modèle n-tiers en raison de différentes couches possibles pour gérer les données et réaliser les analyses. La figure 2.1 représente l'architecture décisionnelle en quatre phases à savoir la phase de collecte des données (*collecter*), la phase d'intégration (*intégrer*), la phase de structuration des données (*distribuer*) et enfin la phase de restitution (*exploiter*).

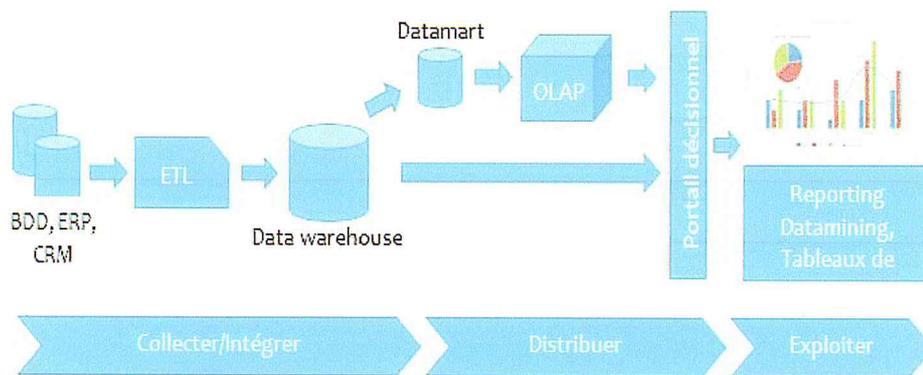


Figure 2.1 : Architecture générale d'un système d'information décisionnel. [3]

Nous décrivons, dans ce qui suit, les quatre phases de l'architecture décisionnelle de la figure 2.1.

- **Phase de collecte ou d'alimentation**

Cette phase fait intervenir des processus d'Extraction, de Transformation et de Chargement *ETL* chargés de récupérer toutes les données nécessaires depuis les différentes sources de stockage. Ces données applicatives sont donc extraites en vue d'être préparées dans la deuxième phase (Transformation).

- **Phase d'intégration**

Une fois les données récupérées par des extracteurs et rendues disponibles dans la zone de préparation (*DSA : Data Staging Area*), cette deuxième étape consiste à nettoyer, filtrer, homogénéiser et agréger les données sources. Il s'agit de l'étape la plus importante et la plus complexe du processus ETL.

- **Phase de distribution**

Après transformation des données et leur adaptation aux objectifs d'analyse, cette étape a pour vocation la structuration des données dans un entrepôt de données (*Data Warehouse*), dans des magasins (*Data Marts*) et des cubes d'analyse. Ces structures de données publient des données nettoyées et précalculées pour des applications d'analyse et d'aide à la décision. Par ailleurs, le système décisionnel prend en charge, dans cette phase la gestion des droits d'accès et respecte donc des schémas correspondant aux profils et aux métiers des différentes classes d'utilisateurs. L'objectif principal de l'étape de diffusion est de segmenter les données collectées en contextes qui soient cohérents, simples à utiliser et qui correspondent à une activité décisionnelle particulière (par exemple aux besoins d'un service particulier). Un entrepôt de données peut héberger des centaines voire des milliers de variables et d'indicateurs alors qu'un contexte de diffusion n'en présente que quelques dizaines pour simplifier son exploitation. Chaque métier peut correspondre à un *Datamart* et chaque contexte d'analyse est modélisé sous la forme d'un hypercube et peut donc être mis à disposition via un outil OLAP.

- **Phase de restitution**

Cette dernière étape, également appelée *Reporting*, se charge de présenter les informations à valeur ajoutée de telle sorte qu'elles apparaissent de la façon la plus lisible possible dans le cadre de l'aide à la décision. Les données sont principalement modélisées par des représentations à base de requêtes afin de constituer des tableaux de bord ou des rapports via des outils d'analyse décisionnelle.

2.3 Data Warehouse :

Les sources de données d'une entreprise proviennent essentiellement des bases de production. Ces données sont éparpillées dans des systèmes multiples, pas nécessairement compatibles entre eux. Ces bases sont conçues pour être efficaces pour les fonctions opérationnelles et de gestion courante de l'activité de l'entreprise. Elles ne sont pas donc appropriées pour l'analyse ; elles sont plutôt efficaces pour conserver l'information. Le Data Warehouse va avoir pour objectif d'agréger et de valoriser ces données provenant de différentes sources. Il va permettre à l'utilisateur d'y accéder de manière simple et ergonomique comme le montre la figure 2.2.

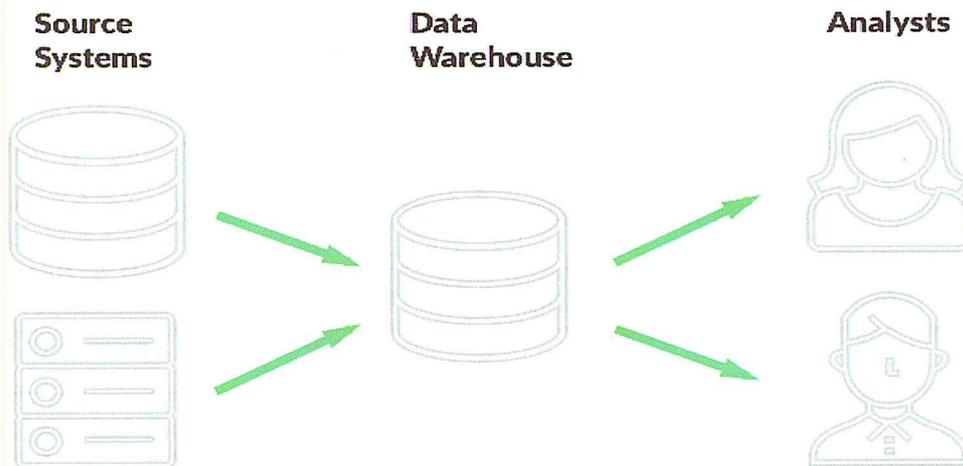


Figure 2.2 : Idée de base d'un entrepôt de données. [14]

2.3.1 Définition

Un entrepôt de données (*Data Warehouse*) est une base de données utilisée pour collecter et stocker des informations provenant d'autres bases de données dans le cadre de la prise de décision. Les bases de données opérationnelles (bases de données de la production) sont collectées, transformées et chargées dans le Data Warehouse via l'outil ETL (*Extract Transform Load*). [5]

2.3.2 Ancêtre du Data Warehouse :

Bien qu'un infocentre puisse être considéré comme un Data Warehouse, il en est plutôt l'ancêtre. L'infocentre était une base de données utilisée, dans un environnement opérationnel, pour des besoins décisionnels. Il contient une copie de travail d'une partie de la base de production, mise à jour périodiquement.

L'infocentre était une première solution pour soulager le système de production des requêtes complexes du décideur. Avec l'accroissement des besoins en matière de décision, que ce soit en termes de quantité de données collectées ou en nombre d'utilisateurs potentiels, l'infocentre se révéla bien insuffisant. Le Data Warehouse l'a rapidement remplacé.

2.3.3 Datamart :

Un Datamart est un sous-ensemble logique d'un Data Warehouse. Il est généralement exploité en entreprise pour restituer des informations ciblées sur un métier spécifique. Il existe, par exemple, des datamarts

compatibilité, RH etc.). Au coeur du Data Warehouse, les données sont organisées par thème. L'entrepôt de données regroupe en son sein des informations de différents métiers (fabrication, achat, qualité) de l'entreprise. Généralement, chaque métier possède des informations de familles communes, mais relatives à un sujet différent (client, produit, contrats). Cet état de fait peut être représenté par le schéma de la Figure 2.4.

Le fait qu'un entrepôt de données soit orienté sujet permet une analyse plus pertinente des données critiques car synthétise toutes les facettes du sujet que l'on étudie.

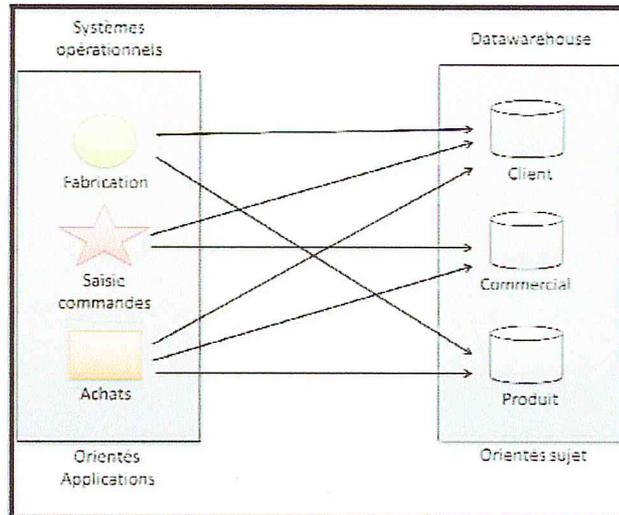


Figure 2.4 : Données orienté sujet dans un entrepôt de données. [15]

- **Intégré :**

Les données proviennent de diverses sources hétérogènes. Avant d'être intégrées au sein du Data Warehouse, elles doivent être mise en forme et unifiées afin d'en assurer la cohérence. Cela nécessite une forte normalisation, de bénéficier d'un référentiel unique et cohérent ainsi que de bonnes règles de gestion. Ces notions sont énoncées et administrées au sein des métadonnées de l'entrepôt de données. Cette phase est très complexe et représente une charge importante dans la mise en place d'un Data Warehouse. Comme illustré par la figure 2.5, les trois sources *Production*, *Employés* et *Facturation* ont été intégrées pour produire un entrepôt sur les *clients*

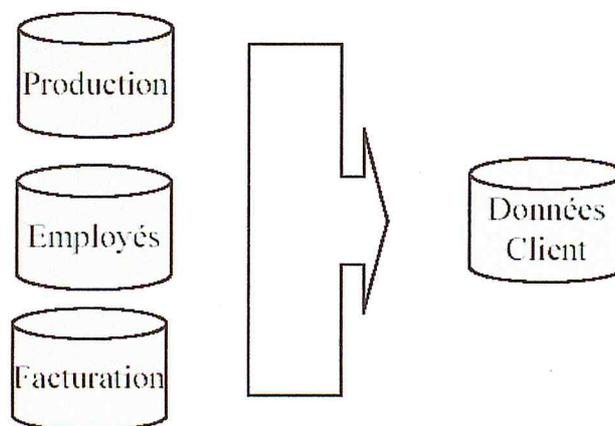


Figure 2.5 : Données intégrées dans un entrepôt de données. [15]

- **Non-volatile :**

Un entrepôt de données (*data warehouse*) permet de conserver la traçabilité des informations et des décisions prises. Comme le montre la figure 2.6, contrairement aux données opérationnelles pouvant être écrasées et supprimées suite aux mises à jour, les données de l'entrepôt, quant à elles, ne sont ni modifiées ni supprimées.

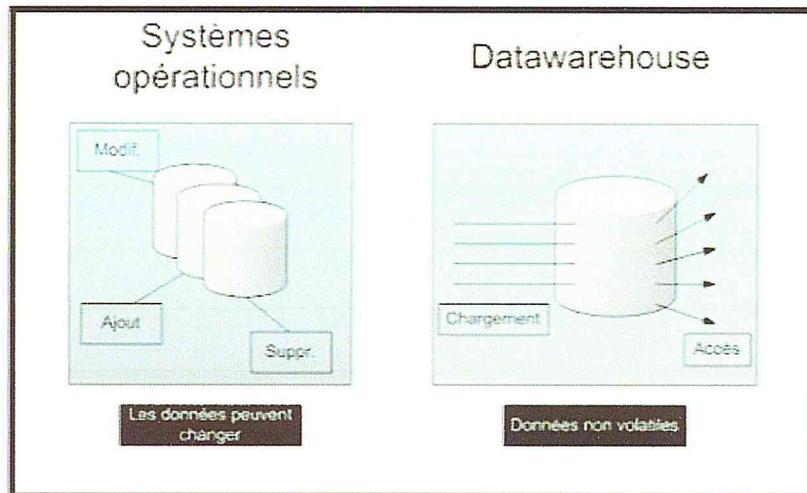


Figure 2.6 : Données non-volatile dans un entrepôt de données. [15]

- **Historisé :**

Contrairement au système de production où les données sont écrasées, les données de l'entrepôt sont toujours préservées et datées afin de tracer leur historique. Chaque nouvelle donnée insérée est accompagnée par une date/heure. Vu l'importance de l'aspect temporel dans un entrepôt de données, un référentiel de temps doit être mis en place afin de pouvoir identifier chaque donnée dans le temps.

2.4 Alimentation du Data Warehouse par les outils ETL

Les données contenues dans un data warehouse sont issues des différentes bases de données de l'entreprise. Ces bases de production, systèmes opérants de l'entreprise, correspondent à l'ensemble des applications informatiques utilisées au quotidien dans l'entreprise pour son activité (gestion de production, gestion bancaire, gestion commerciale,...). Les informations qui y sont stockées, propres à chaque application, peuvent parfois être utilisées par d'autres programmes, par l'intermédiaire de transferts de données, couramment appelés interfaces.

Nous allons donc voir les outils et les méthodes permettant d'alimenter un data warehouse tout en minimisant l'impact sur les systèmes de productions.

La majorité des systèmes d'information d'entreprise sont de nature hétérogène car les systèmes d'informations de l'entreprise s'élaborent le plus souvent sur de longues périodes. Bien que la standardisation des échanges entre les divers outils informatiques avance à grand pas, la disparité des formats des données en circulation est toujours une réalité. C'est le principal obstacle technologique aux échanges d'informations.

Avant d'être utilisables, les données de l'entreprise doivent être mises en forme, nettoyées et consolidées. Les outils ETL (Extract, Transform, Load) permettent d'automatiser ces traitements et de gérer les flux de données qui alimentent l'entrepôt. Ces outils d'alimentation permet aussi de paramétrer des règles de gestion, propres à l'entreprise et à son secteur d'activité. Ces règles visent elles aussi à assurer la cohérence entre les données et à ne stocker dans l'entrepôt de données que des informations préalablement mises en relation les unes avec les autres.

Les outils ETL font référence à plusieurs opérations qui s'articulent autour de trois axes majeurs :

2.4.1 Extraction :

L'extraction des données est la première des étapes des systèmes ETL. Le but de cette étape, est comme son nom l'indique, la lecture et l'extraction des données du système source. On imagine facilement que cette étape s'avère être critique. En effet, dans le cas où le système source doit fonctionner en permanence (24h/24 et 7jours sur 7), il faut que l'extraction, opération coûteuse en ressources, doit être fait le plus rapidement possible et souvent durant un laps de temps précis (souvent nommé « extract window »), décidé en collaboration des équipes chargés de la gestion et/ou de l'exploitation de ce système source.

La complexité de l'extraction n'est pas dans le processus de lecture, mais surtout dans le respect de l'extract window. Cette contrainte est la principale raison de la séparation extraction / transformation. D'autre part, on essaye au maximum d'extraire seulement les données utiles (Mise à jour ou ajoutée après la dernière extraction) et pour ce faire on pourrait s'entendre avec le responsable du système source pour ajouter soit un flag ou encore des dates dans chacune des tables extraites, au moins deux dates : Date de création de l'enregistrement dans la table et la date de mise à jour (En général la plupart des systèmes sources disposent de ces deux dates).

Par ailleurs pour ne pas perdre des données suites à des problèmes d'extraction, il est important de s'assurer que le système source ne purge pas les données avant que l'entrepôt ne les ait extraits. Le processus d'extraction est une des tâches ETL les plus consommatrices de temps car, outre la complexité des différents systèmes source, la détermination exacte des données à extraire est difficile. De plus, ce processus est répétitif, parfois dans des intervalles de temps très rapprochés. La définition du processus d'extraction revient à choisir une méthode d'extraction des données source, un processus de transport et un processus de mise à jour de l'entrepôt (Data Warehouse).

2.4.2 Transformation

C'est l'opération qui réalise le déplacement des données du système source vers le système cible. Par exemple, le transport s'effectue d'un système source vers l'entrepôt ou vers une zone de préparation (zone intermédiaire de stockage), D'une zone de préparation vers l'entrepôt, de l'entrepôt vers un Datamart, etc.

La transformation est la tâche la plus complexe et qui demande beaucoup de réflexion, voici les grandes fonctionnalités de transformation :

- Nettoyage des données.
- Standardisation des données.
- Conformité des données.
- gestion des tables de fait.
- gestion des dimensions.
- affectations des clés de substitution (surrogate key).
- gestion de l'évolution lente (Slowly changing dimension).

- gestion des faits arrivants en retard (Late arriving fact).
- gestion des lookups.

2.4.3 Chargement

Le chargement permet de transférer les données vers leur destination finale. 3 cas de figures se présentent, selon l'architecture mise en place.

De charger les données dans l'entrepôt de données : si la politique retenue a été de construire un entrepôt de données avec une base de données, alors les données seront chargées dans l'entrepôt. Cette approche est proche à celle de Bill Inmon. Il sera dès lors possible d'utiliser des fonctionnalités analytiques comme Oracle le permet.

De les charger dans des cubes de données : la deuxième possibilité est de charger les données directement dans des cubes de données sans les stocker dans un DW. Cette approche est certainement la plus proche à celle de Ralph Kimball. Un bon exemple est l'utilisation directe des cubes de données.

Le mode hybride : la troisième possibilité est celle qui offre le plus d'avantages mais demande par contre plus d'effort. Le chargement des données s'effectue à la fois sur le data warehouse et les data marts.

La transformation de données consiste à :

- La constitution des historiques.
- L'homogénéisation des nomenclatures des différentes sources.
- L'intégration de données externes.
- Filtrage, agrégation, mise à la granularité.
- Nettoyage, suppression d'erreurs.

2.5 Modèle multidimensionnel

L'entrepôt de données héberge toute information utile obtenue à partir des données des systèmes de production et des systèmes externes. Avant d'être intégrées dans l'entrepôt de données, les données doivent être filtrées, nettoyées, adaptées et éventuellement agrégées afin de produire une information de qualité. Celle-ci devra être, ensuite, réorganisée sous forme de dimensions (axes d'analyse) et de mesures (indicateurs) de manière à devenir exploitable facilement pour les besoins d'analyse et d'aide à la décision, on parle alors de modèle multidimensionnel.

Pour arriver à construire un tel modèle qui soit approprié pour un entrepôt de données, nous pouvons choisir entre plusieurs approches basées toutes sur les concepts de faits et de dimensions.

2.5.1 Table de faits :

Une table de fait représente un sujet d'analyse (Chiffre d'Affaire par client, produit et période). C'est une table centrale du modèle dimensionnel constituée de plusieurs mesures appelées aussi indicateurs relatives au sujet traité. Ces mesures sont généralement numériques. [5]

Mais rappelons que certaines tables de faits peuvent ne contenir aucun attribut et représentent que des liaisons entre les tables de dimensions. A titre d'illustration, la figure 3.6 montre une table de fait

contenant le fait (ou la mesure) « quantité vendue » d'un produit (idProduit) durant une date (idDate) et dans une zone donnée (idZone).

Table de faits de Vente
idDate
idZone
idProduit
Montant
Quantité_vendue

Figure 2.7 : Table de faits

- Une table de faits contient les valeurs numériques de ce qu'on désire mesurer.
- Une table de faits contient les clés associées aux dimensions. Il s'agit des clés étrangères dans la table de faits.
- En général, une table de fait contient un petit nombre de colonnes.
- Une table de fait contient plus d'enregistrements qu'une table de dimension.
- Les mesures dans une table de fait sont numériques et sont utilisées dans des fonctions d'agrégation telles que la somme (SUM), la moyenne (AVG), le nombre (COUNT).

2.5.2 Table de dimension :

Une table dimension est une table qui contient les axes d'analyse (les dimensions) selon lesquels on veut étudier des données observables (les faits) qui, soumises à une analyse multidimensionnelle, donnent aux utilisateurs des renseignements nécessaires à la prise de décision. [5]

On appelle donc « dimension » un axe d'analyse. Il peut s'agir des clients ou des produits d'une entreprise, d'une période de temps comme un exercice financier, des activités menées au sein d'une société, etc. La figure 2.8 montre un modèle dimensionnel faisant apparaître les dimensions *Dim_Date*, *Dim_Produit* et *Dim_Zone* nécessaires pour l'analyse des *quantités* vendues de la table de fait *Fait_Ventes*

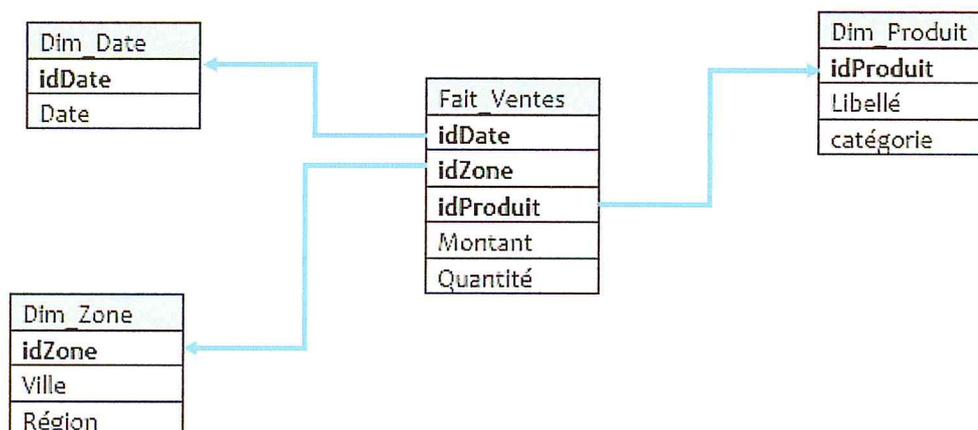


Figure 2.8 : Table de dimension

- Une table de dimension contient le détail sur les faits.

- Une table de dimension contient les informations descriptives des valeurs numériques de la table de faits.
- La table de dimensions contient un plus grand nombre de colonnes.
- Une table de dimension contient en général beaucoup moins d'enregistrement qu'une table de faits.
- Les attributs d'une table de dimension sont souvent utilisés comme « Lignes » et « Colonnes » dans un rapport ou résultat de requête. Par exemple, les attributs textuels d'un rapport proviennent souvent d'une dimension.

- **Cubes OLAP :**

Un cube d'analyse de données en ligne (*OLAP : On Line Analytical Processing*) est une représentation abstraite d'informations multidimensionnelles exclusivement numériques utilisée par l'approche OLAP. Cette structure est prévue pour des fins d'analyses interactives par une ou plusieurs personnes du métier que ces données sont censées représenter. [5]

Les cubes OLAP ont les caractéristiques suivantes :

- Obtenir des informations précalculées (déjà agrégées) selon les besoins de l'utilisateur.
- Simplicité et rapidité d'accès.
- Capacité à manipuler les données agrégées selon différentes dimensions.
- Un cube utilise généralement des fonctions classiques d'agrégation telles que `min()`, `max()`, `count()`, `sum()`, `avg()`, mais peut utiliser également des fonctions d'agrégation spécifiques. L'exemple représenté par la figure 2.9 montre un cube de données relatif aux ventes d'un distributeur par rapport aux dimensions *magasins*, *mois* et *catégories* des produits. Chaque cellule du cube représente une valeur particulière de la mesure « *quantité vendue* » obtenue en prenant un point sur chacun des axes *magasin*, *mois* et *catégorie*

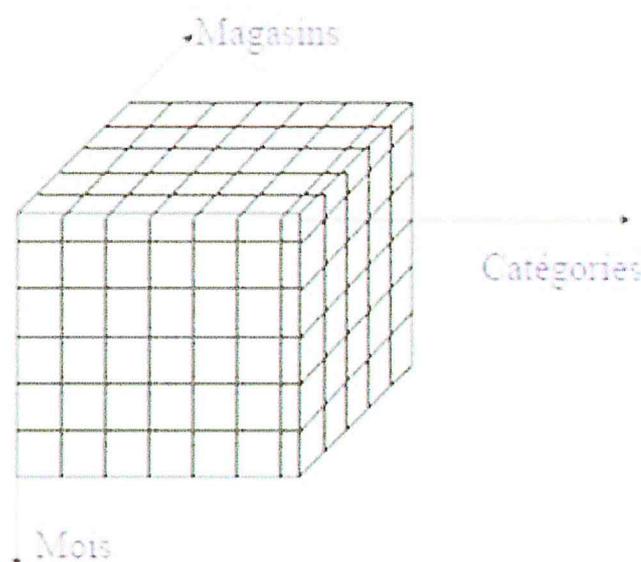


Figure 2.9 : Exemple de cube de données.

2.6 Modélisation

L'un des axes visés pour qualifier un modèle décisionnel est la lisibilité du modèle d'un point de vue de l'utilisateur final. Ce qui qualifie la modélisation multidimensionnelle est qu'elle soit assez proche de la manière de penser des analystes et adaptée à leurs besoins. Cette modélisation consiste à considérer un sujet analyse comme étant un point dans un espace à plusieurs dimensions (axes d'analyse). Les données sont organisées de manière à mettre en évidence le sujet analysé et les différentes perspectives de l'analyse. [4]

On part du principe que les données sont des faits à analyser selon plusieurs dimensions, et il est possible de réaliser une structure de données simple qui correspond à ce besoin de modélisation situé autour.

Au niveau logique, cela peut se traduire par trois schémas différents : en étoile, en flocon de neige ou en constellation.

2.6.1 Schéma en étoile

Ce type de modèle, initié par Ralph Kimball [référence], est traditionnellement représenté par une table de faits centrale autour de laquelle gravitent les dimensions permettant d'analyser les faits qui y sont contenus. Chaque dimension est décrite par une seule table dont les attributs peuvent représenter toutes les granularités possibles. [4]

La figure 2.10 montre un schéma en étoile décrivant les ventes réalisées dans les différents magasins d'une entreprise au cours d'une date. Dans ce cas, nous avons une forme d'étoile avec une table centrale correspondant à la table de faits appelée vente contenant les mesures montant des produits « *Montant* » et la quantité des produits vendue « *Quantité* ». Cette dernière peut être agrégée de diverses façons. La table de faits est entourée de trois dimensions à savoir *temps*, *produit* et *magasin*. Ces dimensions fournissent la base pour l'agrégation de la table de

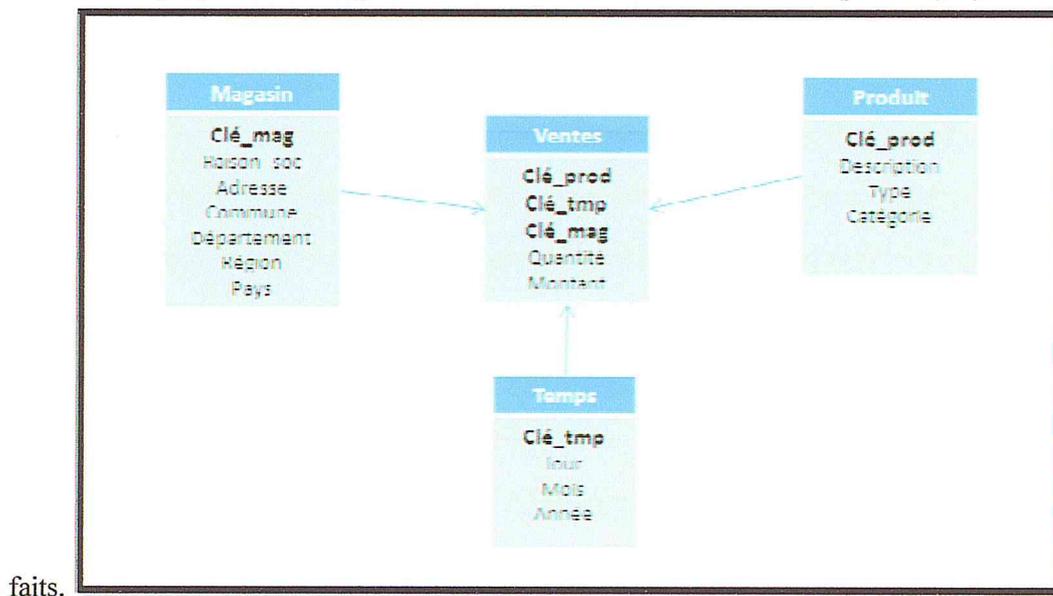


Figure 2.10 : Exemple d'un schéma en étoile.

2.6.2 Schéma en flocon de neige :

Le schéma en flocon de neige dérive du schéma précédent avec une table de faits centrale et autour de laquelle les différentes tables de dimensions sont normalisées donc décomposées pour former des hiérarchies. L'avantage de cette modélisation et de formaliser des hiérarchies au sein d'une dimension, ce qui peut faciliter l'analyse. Un autre avantage est de représenter par la normalisation de ces dimensions, car nous réduisons leurs tailles. [4]

Néanmoins dans [4], l'auteur démontre que c'est une perte de temps de normaliser les relations des dimensions dans le but d'économiser l'espace disque. En effet, cette normalisation rend plus complexe la lisibilité et la gestion dans ce type de schéma qui augmente le nombre de jointures nécessaires dans l'exécution d'une requête.

Les hiérarchies pour le schéma en flocon de neige représenté dans la figure 2.11 sont :

Dimension Temps = Jour ⇒ Mois ⇒ Année

Dimension Magasin = Commune ⇒ Département ⇒ Pays

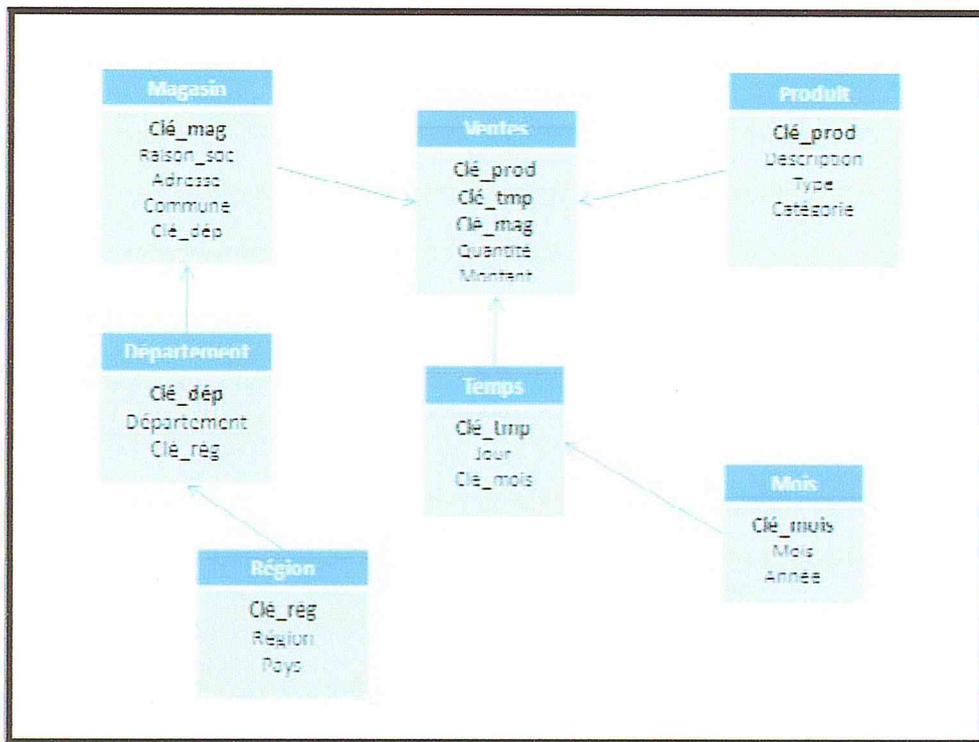


Figure 2.11 : Exemple d'un schéma en flocon de neige.

2.6.3 Schéma en constellation

Le schéma en constellation est basé sur des modèles en étoile. En effet, celui-ci rassemble plusieurs tables de faits avec mise en commun des dimensions similaires. Un modèle en constellation fusionne donc plusieurs modèles en étoile qui utilisent des dimensions communes. [4]

2.7 Conclusion

Nous avons présenté dans ce chapitre le concept de système d'information décisionnel ainsi que les différentes phases qui constituent son architecture. Nous avons ensuite, évoqué les principaux concepts liés à la conception et la construction des entrepôts de données.

Pour alimenter et rafraîchir un entrepôt de données, il faut établir des flux de données entre les bases de production (sources) et l'entrepôt (cible) Ce qui nous a emmené à introduire la notion de processus d'extraction, de transformation et de chargement appelés communément ETL.

3 Chapitre : Présentation de l'organisme d'accueil

3.1 Présentation de Djezzy :

3.1.1 Partie professionnelle :

▪ Introduction

Au début des années quatre-vingt-dix, il fallait attendre des mois et des mois pour acquérir une puce. La rareté du produit le rendait d'ailleurs très convoité. Aujourd'hui, avec trois opérateurs opérationnels sur le marché (Djezzy, Ooredoo et Mobilis), la mobilité et le fait de pouvoir faire des affaires et de communiquer avec une totale liberté de mouvement, la course à l'abonné est lancée.

Djezzy est leader sur le marché algérien des télécommunications, qui connaît une forte croissance due à l'introduction de l'informatique dans les technologies de la télécommunication, offrant une gamme complète de services de voix et de données aux différents clients particuliers et professionnels (GSM et UMTS).

▪ Processus d'évolution de Djezzy

– Orascom Télécom Holding :

OTH (*Orascom Télécom Holding*) est une entreprise égyptienne de téléphonie et des nouvelles technologies présente au Moyen-Orient, en Asie, en Afrique, et en Europe. Elle compte plus de 50 millions d'abonnés dans le monde, dont 18 millions en Algérie, et parmi ses filiales on peut citer :

- Algérie : OTA (*Orascom Telecom Algérie*).
- Bangladesh : Orascom Telecom Bangladesh Ltd, devenu Banglalink.
- Chine : participation minoritaire dans Hutchison Telecom, filiale de Hutchison Wham- poa basée à Hong Kong.
- Corée du Nord : développement d'un réseau 3G à travers la marque commerciale Koryolink depuis 2011.
- Égypte : marque commerciale Mobinil, filiale commune avec Orange.
- Italie : Wind télécom, racheté en même temps qu'Orascom Telecom par Vimpelcom.

– Burundi : U-COM Burundi (leo) est une propriété de (OTA), est une filiale de (OTH)

– Vimpelcom

Vimpelcom est un des plus grands fournisseurs de services de télécommunications mobiles, c'est l'un des plus importants opérateurs de réseau mobile au monde, dont le premier marché est la Russie. Opérant dans 14 pays, il sert plus de 217 millions de clients dont 16.9 millions à OTA comme illustre le tableau 3.1 :

Pays	Opérateurs	Participation	Clients mobiles (millions)	Rang
 Russie	OJSC VimpelCom (Beeline)	100 %	59,0	3/12
Marchés émergents				
 Algérie	Optimum Telecom Algérie (Djezzy)	49 %	16,9	1/3
 Pakistan	PMCL Mobilink	100 %	35,2	1/5
 Bangladesh	Banglalink	100 %	32,3	2/6
Eurasie				
 Arménie	Armenia Telephone Company (Beeline)	100 %	0,8	1/3
 Géorgie	Mobitel	51 %	1,4	3/4
 Kazakhstan	Kar-Tel LLP (Beeline)	98,5 %	9,8	3/9
 Kirghizistan	SkyMobile (Beeline)	%	2,7	2/7
 Ouzbékistan	Unitel (Beeline)	100 %	10,2	2/5
 Tadjikistan	OOO Takom (Beeline)	60 %	1,2	3/4
 Ukraine	Kyivstar	100 %	25,7	1/6
Asie				
 Laos	Vimpelcom Lao (Beeline)	78 %	0,2	3/5
Europe				
 Italie	Wind Telecomunicazioni	92,24 %	21,3	3/4

Tableau 3.1 : Les filiales de Vimpelcom

– Orascom Telecom Algérie :

Orascom Télécom Algérie est une entreprise de droit algérien, Créée en juillet 2001. Elle possède plus de 18 millions d'abonnés. OTA est une filiale d'Orascom Télécom Holding. Elle a bénéficié d'une licence GSM de la part de l'ARPT, acquise au prix de 737 millions de dollars. La mise en place du réseau a coûté 500 millions de dollars.

En 2007, le chiffre d'affaire du holding Orascom a doublé par rapport à ses débuts (2003). OTA a notamment permis au holding de devenir un groupe important du monde des télécommunications, elle est actuellement leader avec plus de 67% du marché de la téléphonie mobile en Algérie. Sa stratégie commerciale est basée sur

quatre marques : Djezzy, Allo ota, OTAxiphone et Imtiyaz.

Djezzy est la dénomination commerciale qui a été retenue pour présenter le réseau GSM d'OTA. Elle met au service de la clientèle algérienne non seulement des produits et services novateurs, mais aussi une qualité de transmission unique grâce à l'utilisation des technologies les plus récentes.

Depuis le 4 octobre 2010, OTH et le russe VimpelCom annoncent leur fusion donc OTH est devenue une partie intégrante de Vimpelcom.

En avril 2014, OTA (*Orascom Télécom Algérie*) est officiellement devenu OTA (*Optimum Télécom Algérie*). Après plus de deux ans de tergiversations et de bras de fer, les autorités algériennes ont réussi à boucler ce dossier délicat. En effet, OTA a changé de statuts et d'actionnaires aussi puisque la nouvelle société mère de Djezzy se met en conformité avec la réglementation algérienne notamment avec la règle des 51%/49%, la figure 3.1 résume le processus d'évolution de Djezzy :



FIGURE 3.1 : Processus d'évolution de Djezzy

▪ **Organisations d'OTA**

Optimum Telecom Algérie se compose de plusieurs services décrits dans la figure ci-dessous.

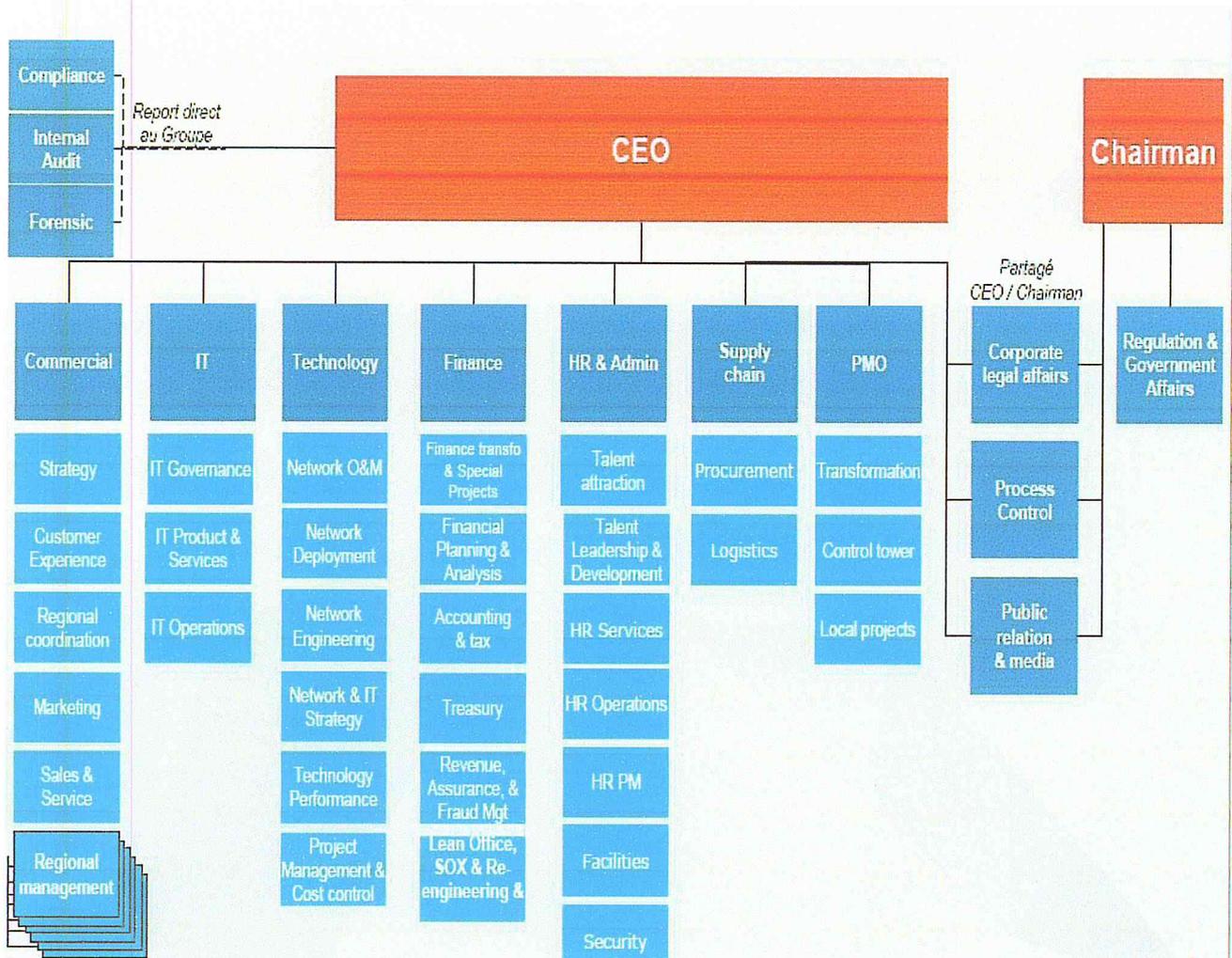


Figure 3.2 : Organisation d'OTA

▪ **Missions et objectifs d'Optimum Télécom Algérie :**

L'ambition d'OTA est d'être l'Opérateur de Télécommunications préféré des Algériens, apportant constamment de la valeur à tous ses partenaires. OTA veut être une référence pour son orientation client, et la qualité de son environnement de travail.

Pour réaliser ses objectifs, OTA s'engage à :

- Offrir les meilleurs produits, de qualité, à des prix compétitifs.
- Déployer des infrastructures à la pointe de la technologie.

- Créer, pour ses employés, le meilleur environnement de travail et d'épanouissement.
- Contribuer activement au bien-être des Algériens.
- Optimiser la création de valeur pour les actionnaires, à travers un contrôle strict des coûts.
- Appliquer rigoureusement sa politique environnementale.
- Améliorer sans cesse son processus interne dans le respect de sa politique qualité.

3.1.2 Partie technique

- **La norme Global System for Mobile (GSM)**

- **Introduction :**

Global System for Mobile (GSM) est une norme internationale constituant un standard pour la téléphonie mobile dit de seconde génération (2G), contrairement à la première génération de téléphones portables, les communications fonctionnent selon un mode entièrement numérique, le GSM permet aussi l'émission de SMS. L'avantage de cette connexion est d'ouvrir un faisceau entre l'appelant et l'appelé qui ne sera fermé qu'en fin de communication. Nous présentons, dans ce chapitre, l'historique, la fonction et l'architecture du réseau GSM. [8]

- **Historique :**

L'histoire de la téléphonie mobile (numérique) débute réellement en 1982. En effet, à cette date le GSM est créé par la conférence européenne des administrations des postes et télécommunications (CEPT) afin d'élaborer les normes de communications mobiles. Le GSM acquiert une influence majeure dans le monde des télécommunications : de nombreux pays l'ont adopté. Aujourd'hui il existe plus de 690 opérateurs GSM répartis dans 213 pays.

- **Le Principe cellulaire:**

Un réseau cellulaire divise la zone à couvrir, généralement un pays entier, en petites zones appelées cellules. Chacune des cellules est desservie par une station de base BS qui reçoit une partie des fréquences disponibles. C'est avec cette station de base que communiquent tous les téléphones mobiles actifs se trouvant dans la cellule concernée. Puisque deux communications radio utilisant la même fréquence interfèrent l'une avec l'autre lorsqu'on se trouve entre les deux émetteurs, les mêmes fréquences ne peuvent être utilisées par deux stations de base voisines, La figure 3.4 montre un exemple des valeurs de fréquence qui diffèrent entre deux cellules voisines.

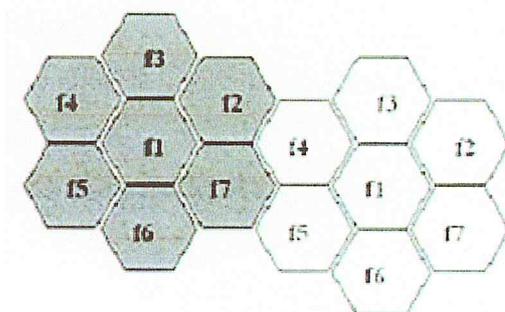


Figure 3.4 : découpage cellulaire d'un réseau GSM

Donc le réseau GSM est un réseau cellulaire dans lequel une zone est divisée en cellules, chacune est constituée de l'infrastructure nécessaire pour relier les abonnés de la cellule au réseau de l'opérateur.

– **Architecture des réseaux GSM :**

L'architecture d'un système GSM se décompose en trois sous-systèmes comme le montre la figure 3.5

- Le sous-système radio (BSS).
- Le sous-système réseau (NSS).
- Le sous-système d'exploitation (OSS).

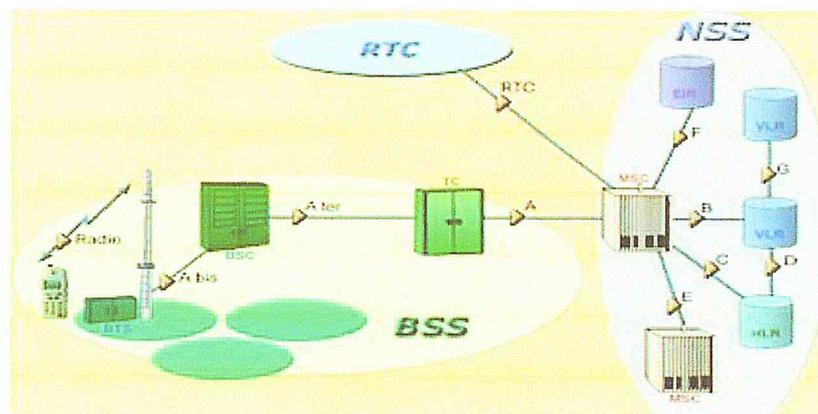


Figure 3.5 : Architecture d'un réseau GSM. [8]

- **Le sous-système radio BSS :**

Sa fonction principale est la gestion de l'attribution des ressources radio et la transmission radioélectrique, indépendamment des abonnés, de leur identité ou de leur communication. On distingue dans le BSS :

- **MS (Mobile Station) :**

C'est le terminal mobile qui émet le premier signal. Il est composé d'un équipement physique et d'une carte SIM authentifiée et autorisée à accéder au réseau mobile.

- **BTS (Base Transceiver Station) :**

Elle est composée d'un ou plusieurs émetteurs et récepteurs (TRX), elle est chargée de la liaison radio avec les stations mobiles. Elle a plusieurs rôles dont l'activation et la désactivation d'un canal radio, la gestion des sauts de fréquence, le codage et la mesure d'un signal radio.

- **BSC (Base Station Controller) :**

Qui est le sous-système intelligent du BSS (il analyse les données et prend les décisions pour assurer la continuité de la communication dans la mobilité), cet équipement assure l'interface avec le segment NSS (Network Sub-System) avec lequel il dialogue au travers de liaisons de type MIC (Modulation par Impulsion et Codage), il assure également le contrôle des BTS qui dépendent de lui, ses fonctions principales sont :

- L'allocation des canaux de communication.
- Le traitement des mesures des niveaux d'émissions BTS et mobiles.
- La concentration de circuits routés vers le MSC.

- La gestion des liaisons de communications.

- **Le sous-système réseau NSS :**

C'est la partie réseau où se déroulent la commutation des appels et les fonctions de gestion de mobilité. Les équipements qui constituent ce segment sont : MSC, HLR, VLR, AUC et EIR.

- ***MSC (Mobile Switching Center) :***

C'est la partie centrale du NSS qui est chargée du routage dans le réseau, de l'interconnexion avec les autres réseaux, elle gère la mobilité et enregistre la location des abonnés (VLR).

- ***HLR (Home Location Register) :***

C'est un enregistreur de localisation géographique des abonnés qui contient une base de données centrale d'un opérateur de réseau mobile. Lorsqu'un abonné achète une carte SIM, toutes ses informations et sa localisation seront enregistrés sur le HLR de l'opérateur, il sert donc de référence aux bases de données locale VLR.

- ***VLR (Visitor Location Register) :***

C'est une base de données qui contient des informations temporaires sur l'utilisateur nécessaires par le MSC afin de servir les abonnés en visite. Le VLR est toujours intégré avec le MSC. Lorsqu'une station mobile traverse vers un autre MSC, Le VLR connecté à cette dernière demande du HLR des données sur le MS.

- ***AUC (Authentication Center) :***

L'AUC est associé à un HLR et sauvegarde une clé d'identification pour chaque abonné mobile enregistré dans ce HLR, cette clé est utilisée pour fabriquer les données nécessaires pour authentifier l'abonné dans le réseau GSM.

- ***EIR (Equipment Identity Register) :***

Un EIR sauvegarde toutes les identités des équipements mobiles utilisés dans un réseau GSM. Cette fonctionnalité peut être intégrée dans le HLR.

- **Le sous-système d'exploitation OSS :**

Cette partie du réseau regroupe trois activités principales de gestion : la gestion administrative, la gestion commerciale et la gestion technique. Le réseau de maintenance technique s'intéresse au fonctionnement des éléments du réseau. Il gère notamment les alarmes, les pannes, la sécurité, Ce réseau s'appuie sur un réseau de transfert de données, totalement dissocié du réseau de communication GSM.

– **Les interfaces réseaux:**

Comme le montre le tableau 3.2, les interfaces sont des protocoles permettant de communiquer entre chaque structure du réseau GSM. Elles constituent un élément essentiel défini dans la norme GSM car ce sont elles qui déterminent les interconnexions réseaux au niveau international.

Nom	Localisation	Utilisation
Air	MS-BTS	Interface radio
Abis	BTS-BSC	Divers (transfert des communications...)
A	BSC-MSC	Divers (transfert de données)
B	MSC-VLR	Transfert de données
C	GMSC-HLR	Interrogation HLR pour appel entrant
D (1)	VLR-HLR	Gestion des informations d'abonnés et de localisation
D (2)	VLR-HLR	Services supplémentaires
E	MSC-MSC	Exécution des 'Handover'
F	MSC-VLR	Vérification de l'identité du terminal
G	VLR-VLR	Gestion des informations d'abonnés
H	HLR-AuC	Echange des données d'authentification

Tableau 3.2 : Les interfaces GSM

– **Gestion de localisation :**

- *LU (Location Update):*

Chaque mobile doit effectuer une mise à jour chaque fois que son LAC (*Location Area Code*) est différent en envoyant un au réseau sa nouvelle location et le nouveau TMSI.

- *LAC (Location Area Code) :*

Le réseau GSM est divisé en cellules, Un groupe de cellule forment un LAC, si le mobile passe d'un LAC à un autre donc le mobile doit effectuer une mise à jour (LU) pour informer le réseau de son emplacement exact

– **Sécurité du réseau GSM :**

- *IMSI (International Mobile Subscriber Identity) :*

Lorsqu'un abonné souscrit à un abonnement mobile auprès d'un opérateur, un identifiant unique appelé IMSI (International Mobile Subscriber Identity) lui est affecté. Ce numéro d'IMSI a été préalablement stocké sur la carte SIM (Subscriber Identity Module). Il est attribué à chaque mobile ayant accès au réseau cellulaire et se trouve dans le HLR. Lorsque le mobile change de LAC, il envoie sa nouvelle localisation et son nouveau TMSI (*Temporary Mobile Subscriber Identity*) au VLR correspondant après sa dernière mise à jour (LU).

- *TMSI (Temporary Mobile Subscriber Identity) :*

De manière à conserver la confidentialité de l'identité de l'IMSI, le VLR alloue un numéro temporaire unique à chaque mobile de localisant dans sa zone de couverture ; ce numéro est appelé TMSI et il n'est connu que sur la partie MS \leftrightarrow VLR, le HLR n'en a jamais connaissance, à chaque changement de VLR, un nouveau TMSI est attribué.

– **Interface Radio :**

C'est l'interface la plus importante dans l'architecture du réseau GSM car elle permet de relier un utilisateur mobile au réseau.

• **L'allocation des fréquences :**

La norme GSM a connu une évolution. La première génération utilise la bande de fréquence des 900 MHz, alors que la 2ème génération utilise la bande des 1800 MHz.

Les caractéristiques de chaque génération sont données dans le tableau 3.3:

Caractéristiques	GSM – 900	GSM - 1800
Fréquence d'émission du terminal vers la station de base	890 à 915 MHz	1710 à 1785 MHz
Fréquence d'émission du la station de base vers le terminal	935 à 960 MHz	1805 à 1880 MHz
Espacement du duplex	45 MHz	95 MHz
Nombre de canaux radio	124	375
Largeur des canaux	200 KHZ	200 KHZ
Multiplexage TDMA	8	8
Nombre de canaux logiques	992	92

Tableau 3.3 : Caractéristiques techniques des bandes GSM

- **Le système d'accès multiples :**

Il définit la manière dont différentes communications simultanées, parmi différentes stations mobiles situées dans des cellules différentes, partagent le spectre radio GSM. Un mélange d'accès multiple par répartition en fréquence (FDMA) et d'accès multiple par répartition dans le temps (TDMA), combiné à un saut de fréquence, a été adopté comme système d'accès multiple pour GSM.

- FDMA (*Frequency Division Multiple Access*)

L'accès multiple par répartition en fréquence est la technologie utilisée dans le réseau téléphonique analogique qui divise le spectre en canaux de 30 kHz. En utilisant FDMA, une fréquence est attribuée à un utilisateur. Donc, plus le nombre d'utilisateurs est grand, plus le nombre de fréquences disponibles est important. Le spectre radio disponible limité et le fait qu'un utilisateur ne libère pas sa fréquence assignée jusqu'à ce qu'il n'en ait plus besoin, explique pourquoi le nombre d'utilisateurs dans un système FDMA peut être « rapidement » limité, La Figure 3.6 représente la technologie FDMA.

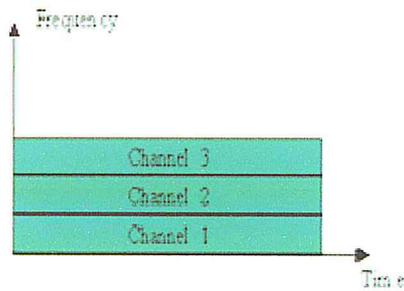


Figure 3.6 : FDMA (Frequency Division Multiple Access).

- TDMA (*Time Division Multiple Access*)

L'accès multiple par répartition dans le temps est une technologie pour la prestation de services sans fil numériques utilisant le multiplexage par répartition dans le temps. Une séquence fixe de créneaux temporels est transmise à plusieurs reprises sur un seul canal de transmission. TDMA fonctionne en divisant une fréquence radio en ces intervalles de temps, puis en allouant les machines à sous à plusieurs appels. De cette façon, une seule fréquence peut supporter plusieurs canaux de données simultanés. En d'autres termes, TDMA permet à plusieurs utilisateurs de partager le même canal. Chacun des utilisateurs, partageant le canal commun, reçoit leur propre éclat dans un groupe de rafales appelé un cadre. Habituellement, TDMA est utilisé Avec une structure FDMA. La figure 3.7 représente la technologie TDMA.

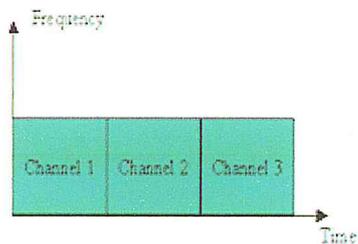


Figure 3.7 : TDMA (Time Divison Multiple Access).

- **Les canaux dédiés**

Un canal logique dédié fournit une ressource réservée à un seul mobile. Ce dernier se verra réserver dans une structure de multi trame, une paire de time slots (un en émission et un en réception) dans laquelle il est le seul à transmettre et à recevoir. Dans la même cellule, aucun autre mobile ne peut transmettre ni recevoir dans un même slot à la même fréquence. Les canaux dédiés sont duplex, et on distingue :

Les canaux TCH et SDCCH :

Ils transportent des informations utilisateurs (voix, données) suivant leurs types, il s'agit des canaux de trafic TCH, Traffic Channel, ou des canaux de signalisation SDCCH, Stand Alone Dedicated Control Channel. Les premiers permettent de transmettre la parole ou les données. Les canaux de signalisation SDCCH ont un débit plus faible que celui des canaux TCH.

Les canaux SDCCH sont requis pour mener à bien les procédures suivantes :

- Mise à jour de localisation : le mobile informe le système dans quelle zone de localisation il se trouve.
- Procédure IMSI Attach, qui permet au mobile de se faire connaître auprès du réseau et d'accéder aux services souscrits.
- Procédure IMSI Detach, qui permet au mobile ou au réseau de s'informer l'un ou l'autre lorsque les services gérés par le MSC ne sont plus accessibles.
- Lorsque le canal dédié est un SDCCH, ce dernier peut écouler tous les types de signalisation, en particulier la signalisation rapide nécessaire au déroulement d'un Handover, dans ce cas il n'y a pas de nécessité d'introduire le FACCH.

- **FACCH :**

Il permet d'écouler différents types de contrôle ou de signalisation. Cependant son débit est très faible. Lorsque le canal alloué est un TCH, on suspend dans ce cas d'urgence, la transmission des informations usagers, et on récupère la capacité libérée afin d'écouler la signalisation. On obtient donc un nouveau canal de signalisation FACCH, Fast Associated Control Channel.

- **Conclusion :**

Le **GSM** a connu un énorme succès et a permis de susciter le besoin de téléphoner en tout lieu avec la possibilité d'émettre des minimessages, mais le besoin d'un débit supérieur et d'autres a impliqué l'évolution de la norme vers d'autres projets de réseau comme le GPRS (2,5G) et le EDGE (2,75 G).

- **La norme UMTS**

- **Introduction:**

Après l'apparition de la 2G, les besoins de déploiement d'un réseau mobile de transport de la voix et de données à haut débit qui sera un siège de nouvelles applications multimédias qui peuvent offrir une couverture mondiale ont mené à l'apparition de la 3G normalisé au plan mondial avec un haut débit , elle est présentée principalement par les normes UMTS (Universal Mobile Télécommunication System).

Dans le reste de ce chapitre nous présentons l'historique, la définition, caractéristiques de l'UMTS son architecture

- **Historique :**

L'UMTS (Universal Mobile Telecommuniation System), fut initié par l'Europe à la fin des années 1980, elle est devenue en 1998 une norme universelle au sein de 3GPP.

Elle fut testé pour la première fois en Europe en 2001, en 2002 elle apparue à Monaco et en 3 mars 2003 l'Angleterre et l'Italie font les premières ouvertures de réseaux 3G.

- **Définition de l'UMTS:**

L'UMTS est une norme de télécommunications de la troisième génération de la famille IMT2000, elle est basée sur la technologie W-CDMA (Wideband Code Division Multiple Access). Elle est conçue pour les communications multimédia de grande qualité et un haut débit pour l'accès aux informations et aux services. [9]

- **Le principe de W-CDMA :**

Chaque paire de Communication se voit attribuer au schéma de codage spécial, complètement différent des autres paires, le W-CDMA permet à plusieurs canaux d'utiliser le même canal physique, les récepteurs réagissent uniquement au code avec lequel ils sont associés avec une même valeur de fréquence comme le montre la figure 3.8 ci-dessous :

- Partie Réseau d'accès.
- Partie cœur.

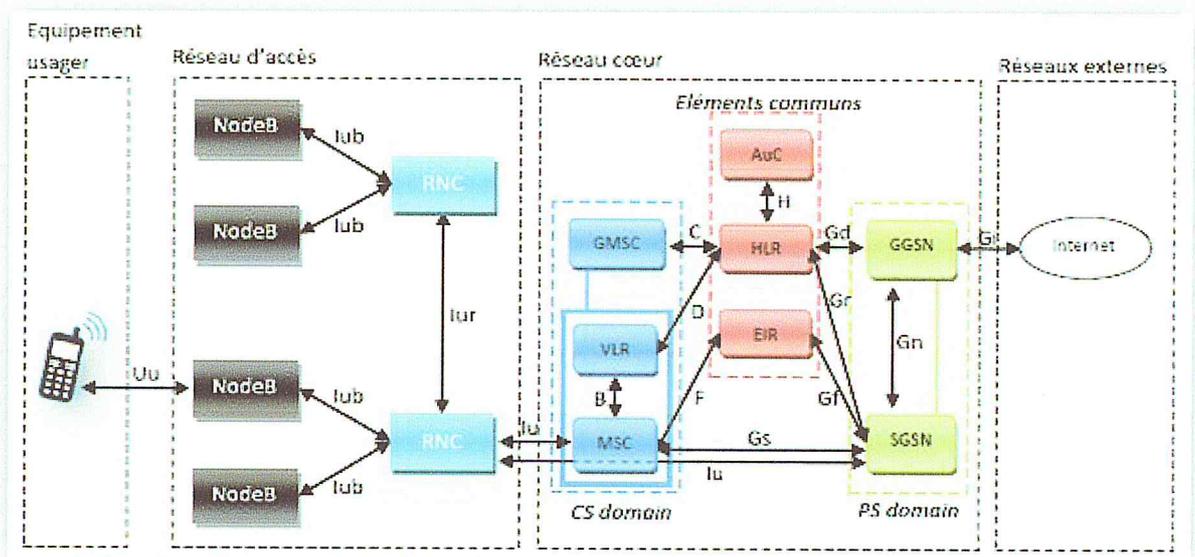


Figure 3.9 : Architecture du réseau UMTS. [9]

- **EU (Équipement Usager) :**

Il transmet de l'information vers le réseau UMTS à travers l'interface radio et applique les fonctions de modulation, de correction d'erreurs, d'étalement de spectre.

- **Réseau d'accès :**

C'est la passerelle entre l'équipement usager et le réseau cœur permettant le transfert des données générées par l'utilisateur via les interfaces Uu et Iu à ce dernier, il assure la sécurité des informations échangées par l'interface radio et chargé d'allouer et maintenir des ressources radio nécessaires à la communication, il est composé de :

- Une ou plusieurs stations de base (Node B).

Node B est équivalent à la BTS du réseau GSM. Il peut gérer une ou plusieurs cellules. Il inclut un récepteur CDMA qui convertit les signaux de l'interface Uu (Interface Air) en flux de données acheminés au RNC sur l'interface Iub.

- Des contrôleurs radio RNC.

RNC (Radio Network Controller) constitue le point d'accès pour l'ensemble des services vis-à-vis du réseau cœur.

- Des interfaces de communication entre les différents éléments au réseau UMTS.
 - Uu: Interface entre un équipement usager et le réseau d'accès UTRAN. Elle permet la communication avec ce dernier via la technologie CDMA.
 - Iu : Interface qui relie entre le réseau d'accès et le réseau cœur de l'UMTS. Elle permet au contrôleur radio RNC de communiquer avec le SGSN.
 - Iur : Interface qui permet à deux contrôleurs radio RNC de se communiquer.
 - Iub : cette interface est présente entre un Node B et le RNC qui le contrôle.

- **Réseau cœur:**

Il est composé de trois parties :

- **La partie Cs (Circuit Switched) :**

Cette partie est utilisée pour la téléphonie, elle est composée de plusieurs modules.

- *Le MSC (Mobile Services Switching Center) :*

Il établit la communication avec l'équipement usager et commute les données.

- *Le GMSC (Gateway MSC) :*

C'est une passerelle entre le réseau UMTS et le réseau téléphonique commuté PSTN.

- *Le VLR (Visitor Location Register) :*

C'est une base de données qui attachée au MSC elle enregistre les usagers dans une zone géographique LA (Location Area).

- **La partie Ps (PacketSwitched) :**

Elle permet la communication de paquet (le transfert de données), et il est composé des modules :

- Le SGSN (Service GPRS Support Node) :

Il est chargé d'enregistrer les usagers dans une zone géographique dans une zone de routage RA (Routing Area).

- Le GGSN (Gateway GPRS Support Node) :

C'est une passerelle vers les réseaux à commutation de paquets extérieurs tels que l'internet.

- **Eléments communs :**

Les éléments communs entre PS et CS sont AUC, HLR et EIR que nous avons déjà vu dans la partie *1.4.2. Le sous système réseau NSS.*

- **Conclusion :**

L'apparition de la 3G a permis l'évolution des systèmes de télécommunication, elle offre aussi un débit bien supérieur au réseau 2G et un téléchargement des données plus performant avec une bonne couverture, mais cela n'a pas empêché l'apparition de la technologie 4G.

3.2 Présentation de la structure concernée par le projet

3.2.1 Présentation fonctionnelle :

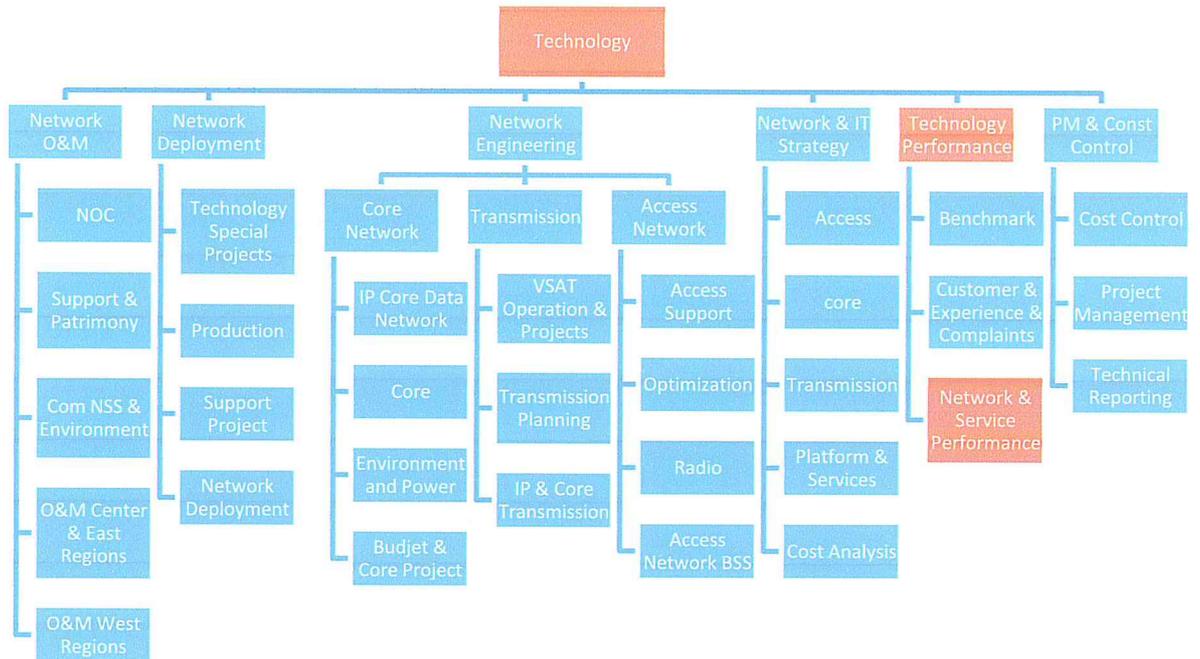
Les principaux devoirs que doit apporter le service technologie de l'information et qualité sont cités dans les points suivants :

- Doit être au courant de tout défaut et / ou comportement anormal dans le réseau pour tout l'équipement de réseau central existant.
- Doit être conscient des conséquences de l'information, de l'escalade et de la résolution des problèmes

- Effectuer des audits et des audits de capacité et de service du réseau central GSM / UMTS, identifier et augmenter la congestion, et signaler les problèmes aux unités respectives de planification et d'optimisation.
- Coordonner les recommandations d'amélioration des KPI pour assurer une intervention rapide en cas de dégradation perçue entre les interfaces de base TDM, IP et de signalisation
- Analyse des données d'information SS7-ISUP et SIGTRAN pour toutes les interfaces et détermination de la violation de seuil pouvant affecter les performances optimales du réseau de signalisation
- Analyser les KPI d'interconnexion (ASR, NER) pour A-Interface, Inter-MSC pour l'optimisation et les interventions
- Assurer la surveillance de la qualité de l'itinérance pour aider à réduire de manière proactive les problèmes d'itinérance
- Surveiller les éléments de réseau de base PS et IP / MPLS dans le but d'identifier les performances médiocres et d'intensifier / suivre avec des parties prenantes responsables pour améliorer la performance et donner aux clients l'expérience souhaitée.
- Audit de la définition et de la formule des indicateurs clés de performance à travers les technologies.

3.2.2 Présentation organisationnelle :

Voici l'organigramme de la structure d'accueil :



3.3 Profils intervenant dans le suivi de la qualité (Personnes / compétences)

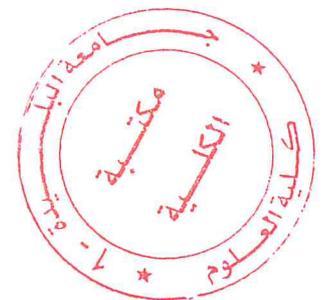
Le service d'accueil se compose de 40 membres dont une présidente et 6 managers et 2 experts et plus de 30 ingénieurs, notre encadreur M. Mohamed SAKILANI est expert dans le management & performance.

3.4 Moyens matériels et outils logiciels utilisés

3.4.1 Moyens matériels :

- Machine :

La structure d'accueil dispose de plusieurs postes de contrôles pour suivre la qualité du réseau commandés par plusieurs administrateurs,



- **Serveur de données :**

La structure dispose d'un serveur FTP hébergeant des données brutes, ces données représentent des valeurs KPI qui servent à l'analyse après avoir passé par une adaptation faite par les administrateurs.

3.4.2 Outils logiciels :

A la base la structure « Network & Performance » travail avec plusieurs outils de traitement, pour commencer les administrateurs doivent extraire les données brutes depuis un serveur FTP pour ce faire ils utilisent un programme conçue par eux même, ce dernier apporte les données d'un jour avant, Ensuite ils utilisent Microsoft Office Excel pour supprimer, ajouter, nettoyer, combiner et calculer plusieurs valeurs afin de les adapter a l'analyse, pour arrivé a ce point important les administrateurs représentent ces données en forme de graphe pour visualiser et suivre la qualité et arrivé a rédiger leurs rapport au supérieur et enfin prendre les bons choix.

3.5 Description des fichiers source utilisés pour l'analyse de la qualité du réseau :

Dans cette partie nous allons décrire les différents fichiers sur les quelles notre analyse se base pour commencer, nous savons que notre système traitera les données provenant des réseaux 2G et 3G, ce qui fait qu'il y aura deux type de fichiers mais une seule et unique extension le « .CSV » (fichier Excel).

Chaque fichier contient plus de 100 colonnes, chaque colonne comporte une valeur de type entiers ou caractère ou chaîne de caractère.

Les fichiers extraits auront un plus grand volume à la fin de la transformation suite a l'ajout de nouvelles colonnes cruciale à l'analyse.

3.6 Description du processus d'investigation pour l'identification des problèmes de dégradation de la qualité du réseau :

L'identification des problèmes est une approche systématique de dépannage. Son objectif est de déterminer les raisons d'un dysfonctionnement et d'expliquer comment résoudre le problème.

La première étape consiste à décrire le problème dans sa totalité. Pour commencer, on doit visualiser les principaux indicateurs (*KPI: Key Performance Indicator*) grâce aux capteurs qui donnent une vue d'ensemble sur le comportement du réseau). Suite à une analyse de ces derniers, nous pourrions remarquer s'il y a une présence d'anomalies ou non.

La présence d'anomalies nous pousse à exploiter les sous-causes qui nous permettent de savoir si le problème est de comportement normal (le système n'est pas défaillant) ou de comportement anormal (le système est défaillant) auquel cas l'équipe concernée est informée par le biais d'un rapport d'analyse.

3.7 Problèmes et limites du processus actuel :

Aujourd'hui, Djezzy dispose de plusieurs administrateurs dédiés pour le suivi de la qualité du réseau où chacun doit extraire, quotidiennement, les données sources dont il est responsable depuis un serveur FTP. Ceci se fait d'une façon manuelle sachant que ces données reflètent l'état du réseau de la journée d'avant (la veille). Dans ces conditions, si la qualité du réseau s'est dégradée, durant le jour J, suite à une panne quelconque, l'administrateur ne pourra le constater qu'un jour après, c.à.d. le jour J+1. Les problèmes engendrés par cette situation peuvent impacter la communication des abonnés durant la journée en question et provoquer éventuellement des réactions chez la clientèle.

En plus, afin de faire face à cette situation, les administrateurs doivent parcourir plusieurs étapes de transformation des données sources générés par l'infrastructure télécom afin de générer des données analytiques capables d'identifier les problèmes de dégradation de la qualité du réseau. . Comme ces transformation de font manuellement, beaucoup de retard est accusé dans l'identification des problèmes de qualité, ce qui engendre beaucoup de pertes pour l'opérateur Djezzy : mécontentement des abonnés, réclamations, immigration chez la concurrence, etc.

3.8 Objectifs et besoin futurs :

Notre projet de fin d'études objective de faciliter l'investigation pour l'identification des problèmes de dégradation de la qualité du réseau. Pour ce faire, nous devons d'abord (1) disposer des données sources qui sont des compteurs générés par les équipements du réseau. (2) extraire ces données sources en temps quasi-réel à partir d'un serveur FTP, (3) opérer sur les données sources un certain nombre de transformations afin de les adapter aux objectifs d'analyse et les charger dans un entrepôt de données, (4) l'analyse de données en ligne en temps quasi-réel pour le contrôle ainsi que l'investigation en cas de dégradation de la qualité du réseau.

Afin d'atteindre ces objectifs et comprendre les attentes des administrateurs Djezzy en termes de suivi de la qualité du réseau, le chapitre suivant est consacré à la spécification des besoins et la conception du système.

4 Chapitre : Spécification Des Besoins & Conception Du Système

4.1 Introduction

La spécification des besoins et la conception sont des étapes indispensables dans la mise en œuvre d'une application logicielle. Au cours de ces étapes, le concepteur doit définir de façon très précise toutes les fonctionnalités que doit fournir le système ainsi que la façon d'y parvenir. Pour ce faire, nous avons opté pour la notation UML devenu un langage universel pour la modélisation objet des systèmes.

Pour la capture des besoins d'un concepteur ETL qui devra utiliser notre système, nous présenterons les diagrammes de cas d'utilisation, nous définissons l'aspect statique de notre système avec les diagrammes de classes.

4.2 Langage de modélisation

4.2.1 Le langage de modélisation UML :

UML (*Unified Modeling Language*) est un langage formel, qui permet d'exprimer et d'élaborer des modèles objets, indépendamment de tout langage de programmation. Né de la fusion des méthodes objet dominantes (OMT, Booche et OOSE), puis normalisé par l'OMG en 1997. UML a plusieurs versions. La dernière version est 2.1.1 sortie en février 2007.

UML est un langage de modélisation visuel et non pas une méthode d'analyse et de spécification. Un système est décrit en UML, à travers sa structure statique et son comportement dynamique. UML propose donc un ensemble de notations graphiques, pour capturer les informations relevant des aspects statiques et dynamiques du système. [10]

4.2.2 Les diagrammes UML :

La version 2.0 d'UML comporte treize diagrammes représentant des vues distinctes pour représenter des concepts particuliers du système d'information. Ils se répartissent en deux grands groupes.

Diagrammes structurels ou diagrammes statiques (*UML Structure*)

- diagramme de classes (*Class diagram*)

- diagramme d'objets (*Object diagram*)
- diagramme de composants (*Component diagram*)
- diagramme de déploiement (*Deployment diagram*)
- diagramme de paquetages (*Package diagram*)
- diagramme de structures composites (*Composite structure diagram*)

Diagrammes comportementaux ou diagrammes dynamiques (*UML Behavior*)

- diagramme de cas d'utilisation (*Use case diagram*)
- diagramme d'activités (*Activity diagram*)
- diagramme d'états-transitions (*State machine diagram*)
- **Diagrammes d'interaction (*Interaction diagram*)**
 - diagramme de séquence (*Sequence diagram*)
 - diagramme de communication (*Communication diagram*)
 - diagramme global d'interaction (*Interaction overview diagram*)
 - diagramme de temps (*Timing diagram*)

Ces diagrammes sont d'une utilité variable selon les cas, et ne sont pas nécessairement tous utilisés dans un même projet. Généralement, les diagrammes d'activités, de cas d'utilisation, de classes, d'objets, de séquence et d'états-transitions sont les plus utilisés.

4.3 Spécification des besoins

4.3.1 Les diagrammes de cas d'utilisation

▪ Définition des diagrammes des cas d'utilisation

Un diagramme de cas d'utilisation capture le comportement d'un système, d'un sous système, d'une classe ou d'un composant tel qu'un utilisateur extérieur le voit. Il scinde la fonctionnalité du système en unités cohérentes, ayant un sens pour les acteurs. Les cas d'utilisation permettent d'exprimer le besoin des utilisateurs d'un système, ils sont donc une vision orientée utilisateur de ce besoin au contraire d'une vision informatique. [11]

4.3.2 Les acteurs du système

Le système interagit avec un seul type d'acteur ou d'utilisateur : « l'administrateur ». L'administrateur prend en charge tous les aspects et les étapes de modélisation du processus ETL.

4.3.3 Diagramme de cas d'utilisation du système globale

▪ Diagramme

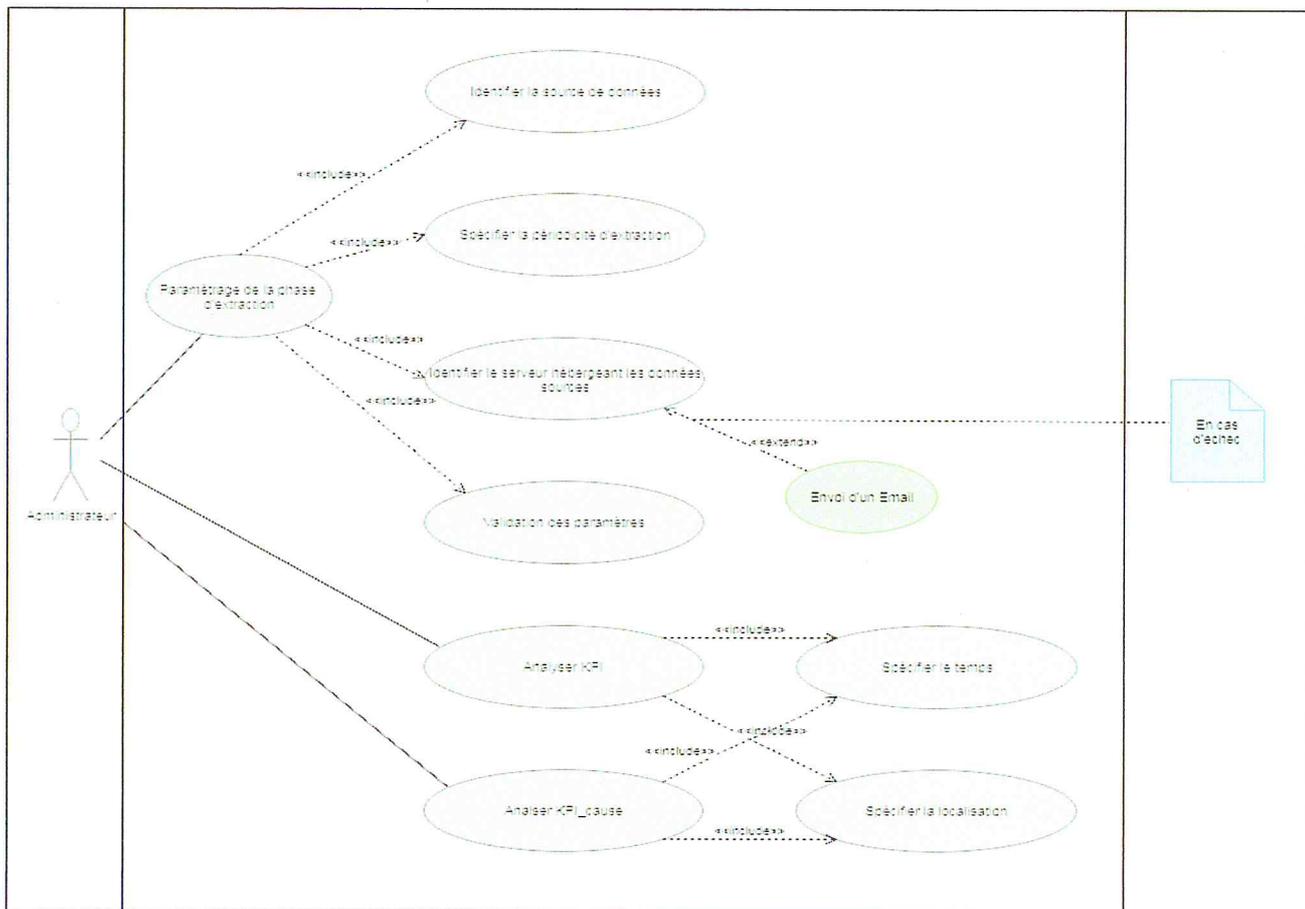


Figure 4.1 : Diagramme de cas d'utilisation du système.

- **Description textuelle de chaque cas d'utilisation**

N	Cas d'utilisation	Désignation
1	Paramétrage de la phase d'extraction	L'administrateur gère l'extraction des données en appliquant les fonctionnalités.
2	Identifier la source de données	L'administrateur doit spécifier le type ainsi que le nom de la source de données.
3	Identifier le serveur hébergeant les données sources	L'administrateur doit spécifier l'adresse IP du serveur et s'identifier par un nom de client.
4	Validations des paramètres	L'administrateur doit valider les paramètres.
5	Envoie d'un Email	L'administrateur reçoit un email automatique en cas d'échec au cours de l'extraction des données depuis le serveur FTP.
6	Analyser KPI	L'administrateur doit analyser les KPI (main KPI) pour savoir quelles causes doit il analyser.
7	Analyser KPI_Causes	L'administrateur doit analyser les KPI_Causes pour détecter la source du problème et rédiger un rapport à l'équipe concernée.

Tableau 4.1 : Description textuelle de chaque cas d'utilisation.

4.4 Modélisation de l'entrepôt de données

4.4.1 Introduction

L'entrepôt des données constitue le noyau dur du système décisionnel que nous mettons en place dans le cadre de la supervision de la qualité du réseau télécom de Djezzy et l'investigation pour comprendre les causes ayant provoqué la dégradation de la qualité. L'entrepôt de données étant une technologie de stockage différente par rapport aux bases de données classiques et nécessite donc un

autre formalisme de modélisation. En effet, nous utilisons une approche de modélisation en étoile où la table centrale représente la table de fait contenant la mesure à analyser et les tables qui l'entourent représentent les dimensions par rapport auxquelles se fera l'analyse.

4.4.2 Modélisation dimensionnelle

Les quatre premiers modèles dimensionnels (modèles 1 à 4) représentent la partie de l'entrepôt relative à l'analyse des principaux indicateurs (Main KPI) par rapport au type des KPI, la localisation (Réseau, Région, Lac) et le temps (heure, jour, semaine, mois) durant lequel ont été générés les compteurs.

un exemple réel de notre base de données qui reflète le modèle présenté dans la figure 4.1 :

Si Time_Heure = Heure et Localisation = Lac

3G MS init PDP context act Fail % (Lac) (Main KPI) = 0.913155437

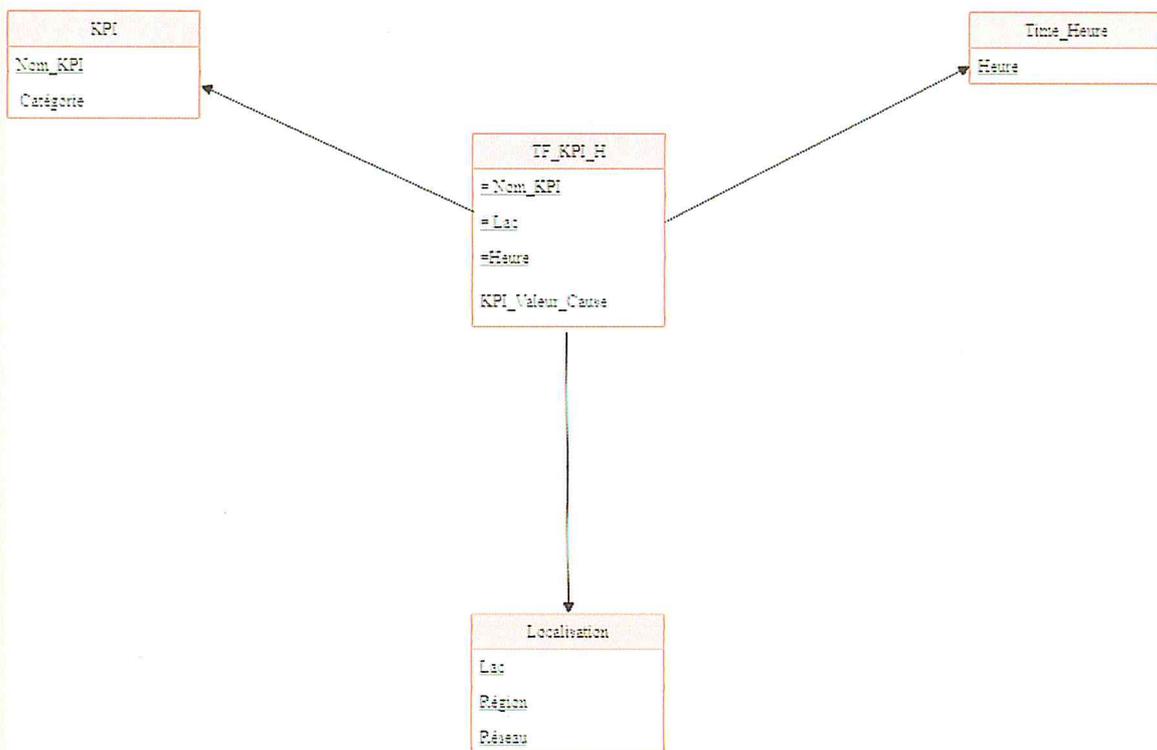


Figure 4.1 : Modèle dimensionnel relatif à l'analyse des principaux KPI par heure

Un exemple réel de notre base de données qui reflète le modèle présenté dans la figure 4.2 :

Si Time_Heure = jour et Localisation = Lac

3G MS init PDP context act Fail % (Lac) (Main KPI) = 0.814166437

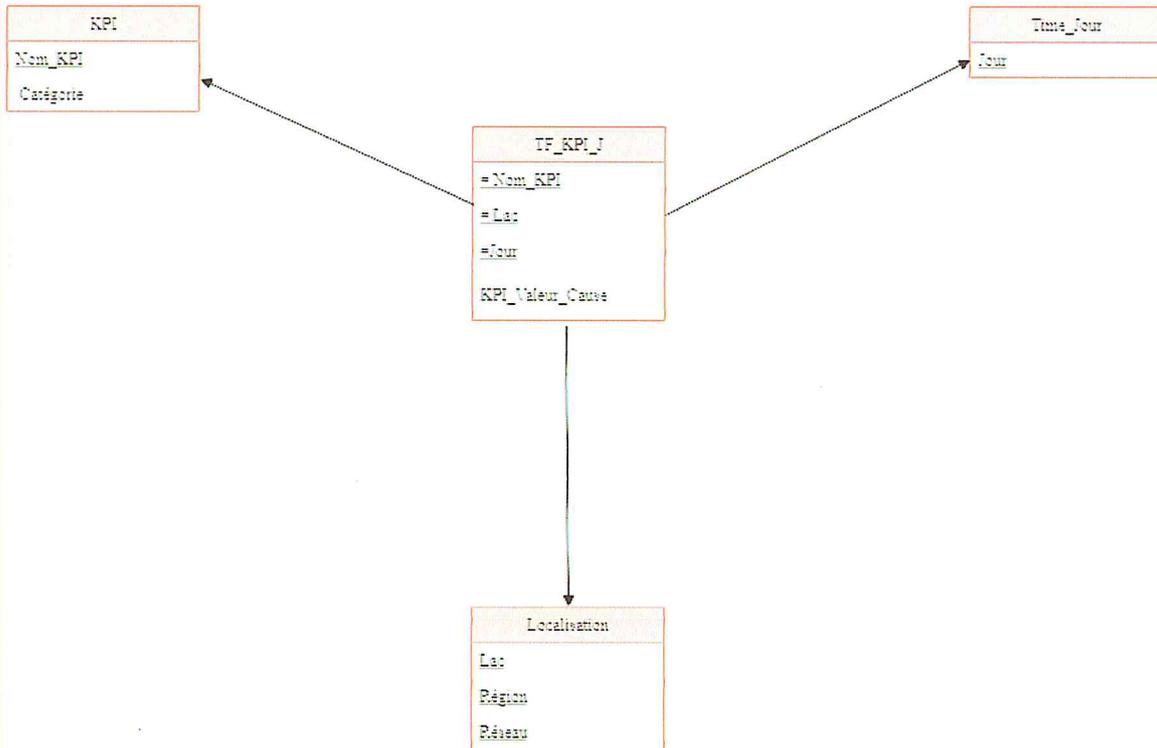


Figure 4.2 : Modèle dimensionnel relatif à l'analyse des principaux KPI par jour

Un exemple réel de notre base de données qui reflète le modèle présenté dans la figure 4.3 :

Si Time_Heure = semaine et Localisation = Lac

3G MS init PDP context act Fail % (Lac) (Main KPI) = 0.914166437

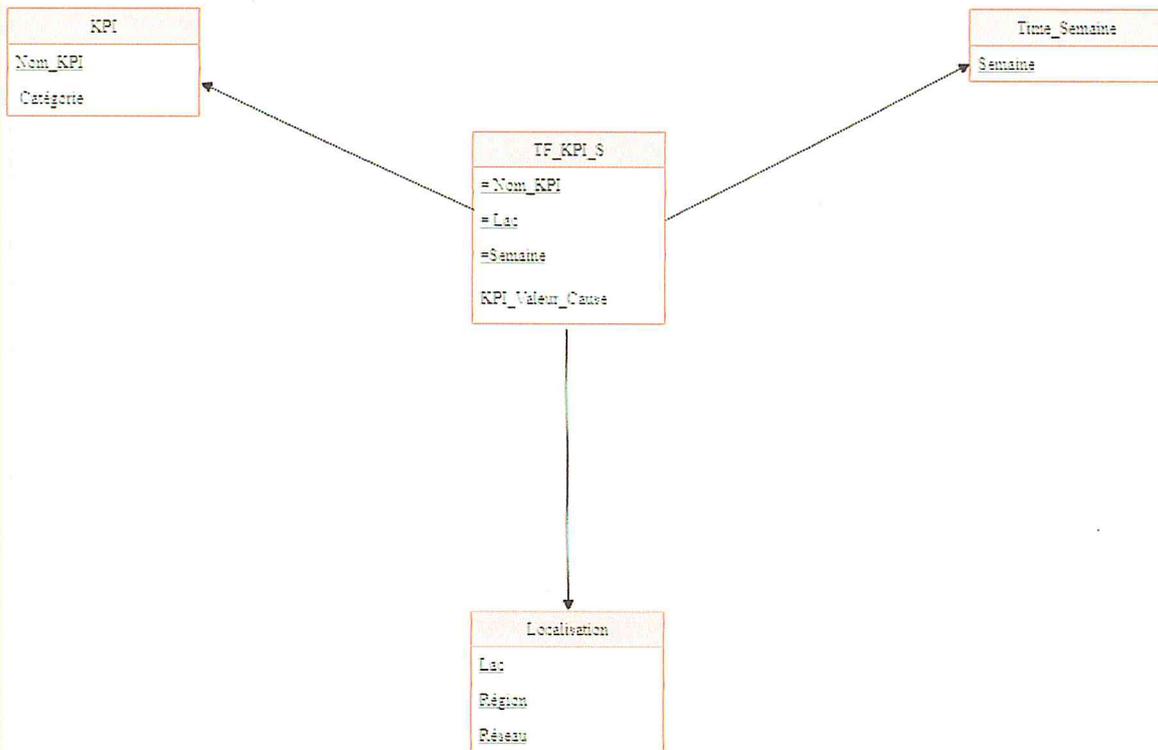


Figure 4.3 : Modèle dimensionnel relatif à l'analyse des principaux KPI par semaine

Un exemple réel de notre base de données qui reflète le modèle présenté dans la figure 4.4 :

Si Time_Heure = mois et Localisation = Lac

3G MS init PDP context act Fail % (Lac) (Main KPI) = 0.619169437

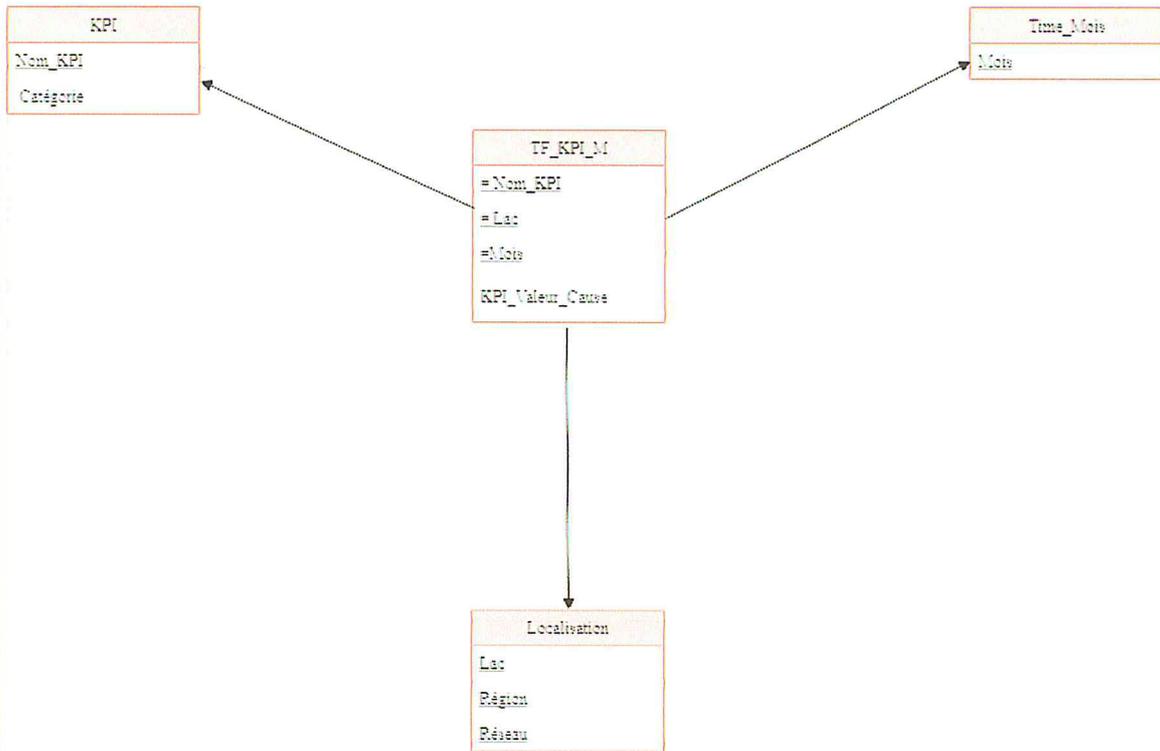


Figure 4.4 : Modèle dimensionnel relatif à l’analyse des principaux KPI par mois

Les seize modèles dimensionnels restants (modèles 5 à 16), représente la relation entre les tables de dimensions et la table de fait afin d’avoir une valeur du KPI (cause KPI) en fonction de la catégorie, la localisation (Réseau, Région, Lac) et le temps (heure, jour, semaine, mois).

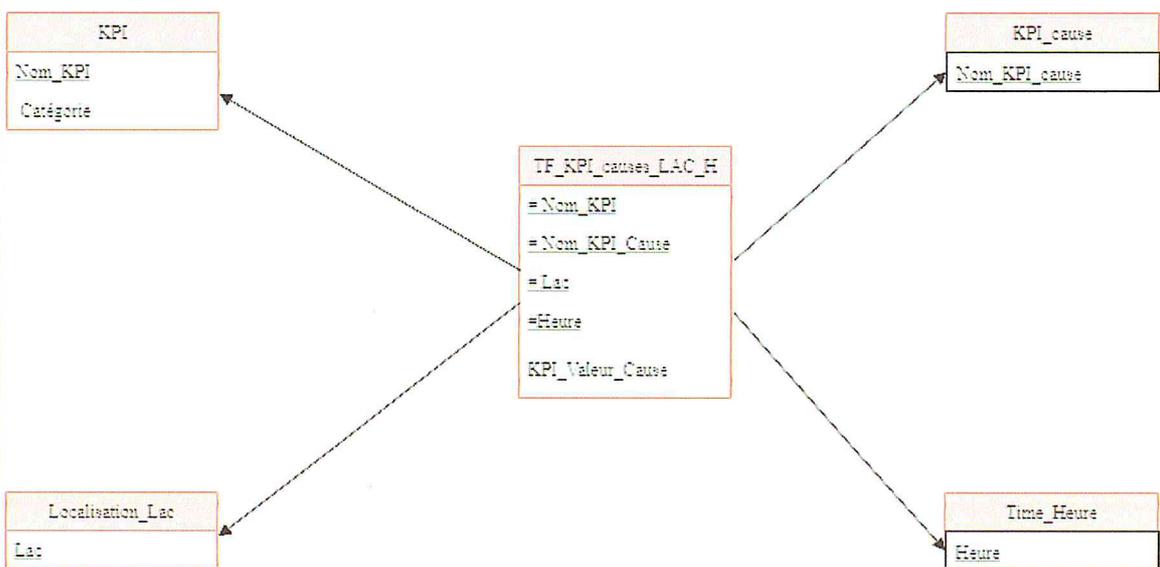


Figure 4.5 : Modèle dimensionnel relatif à l’analyse des causes d’un KPI par Lac et Heure

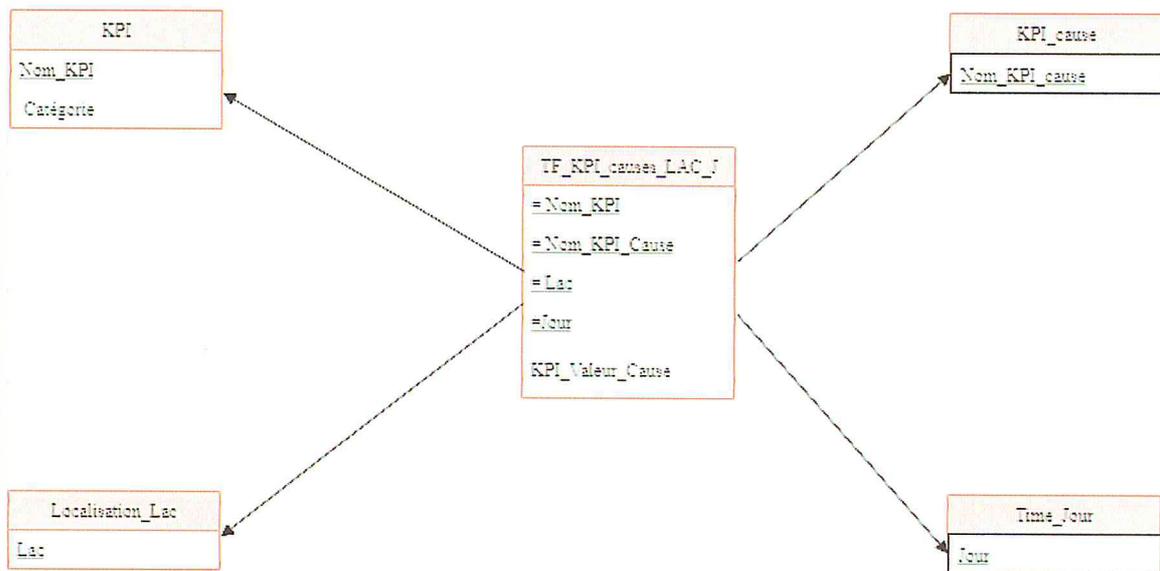


Figure 4.6 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Lac et jour

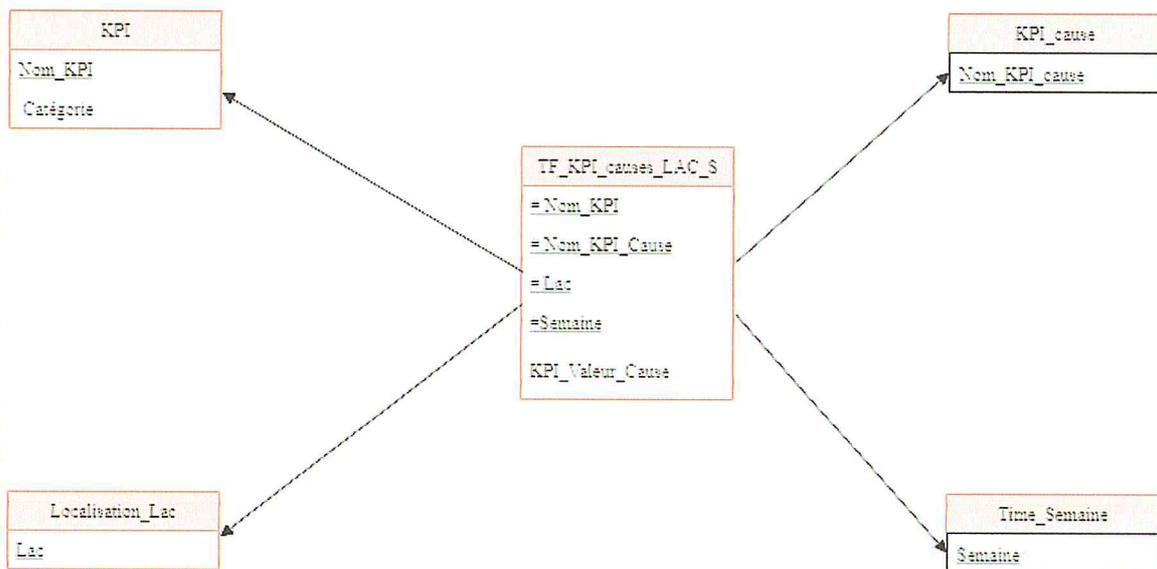


Figure 4.7 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Lac et semaine

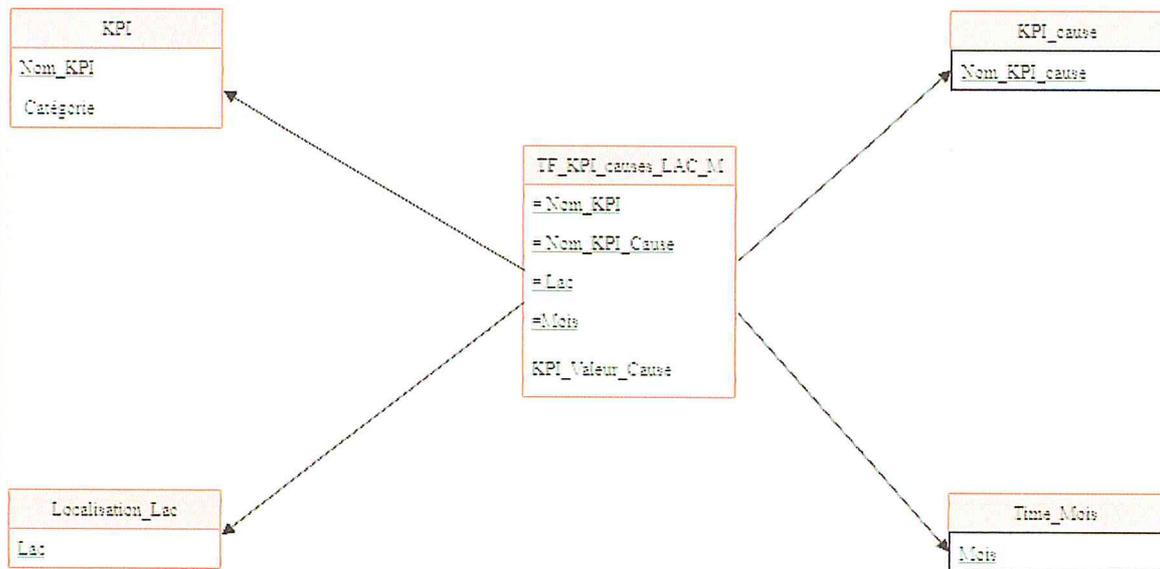


Figure 4.8 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Lac et mois

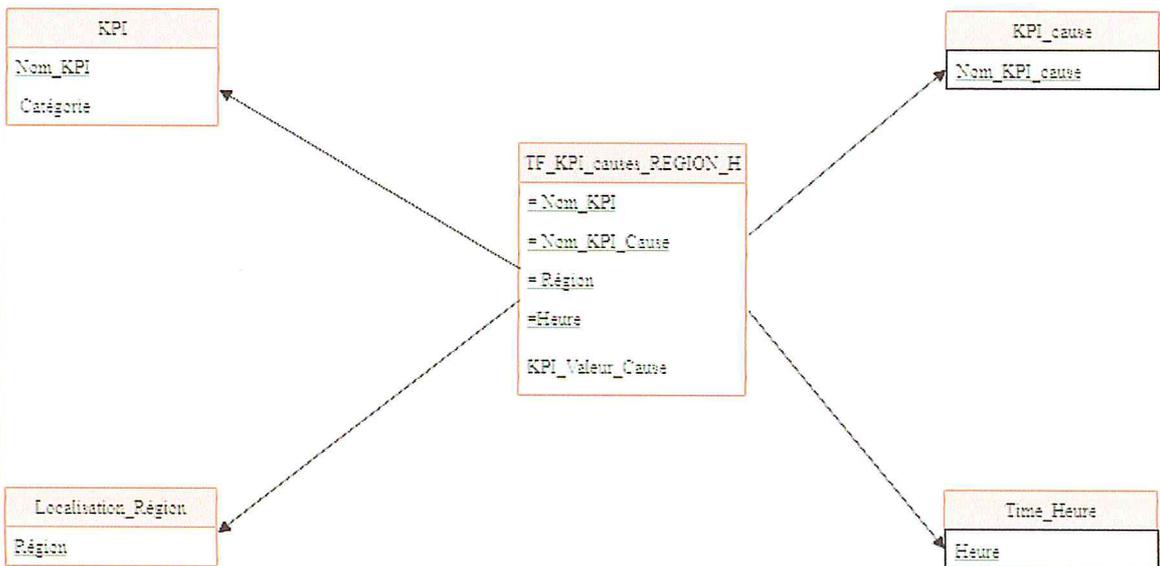


Figure 4.9 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Région et heure

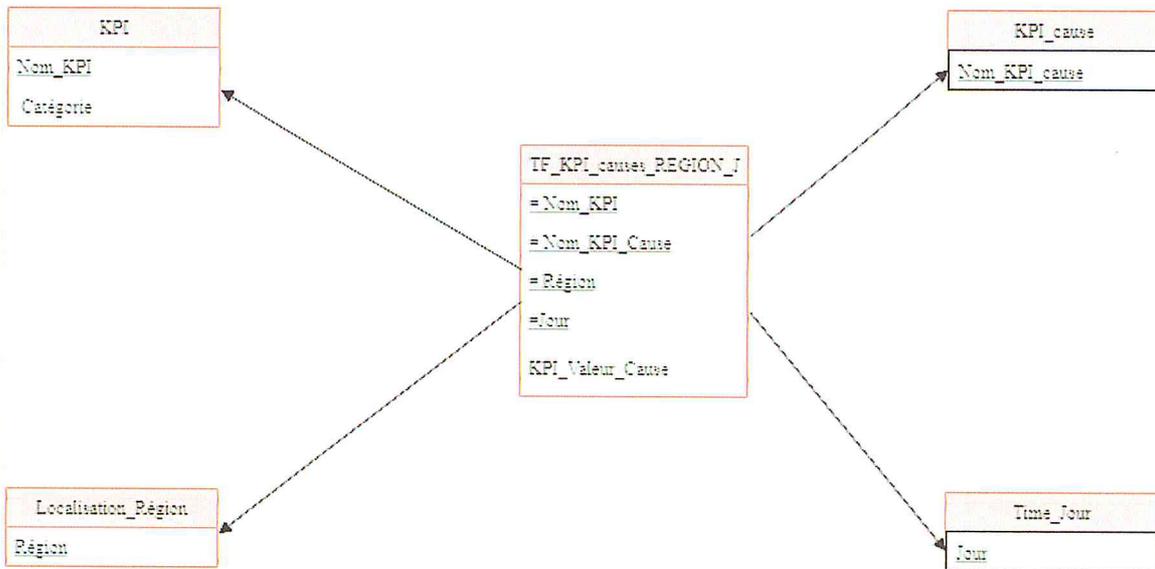


Figure 4.10 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Région et jour

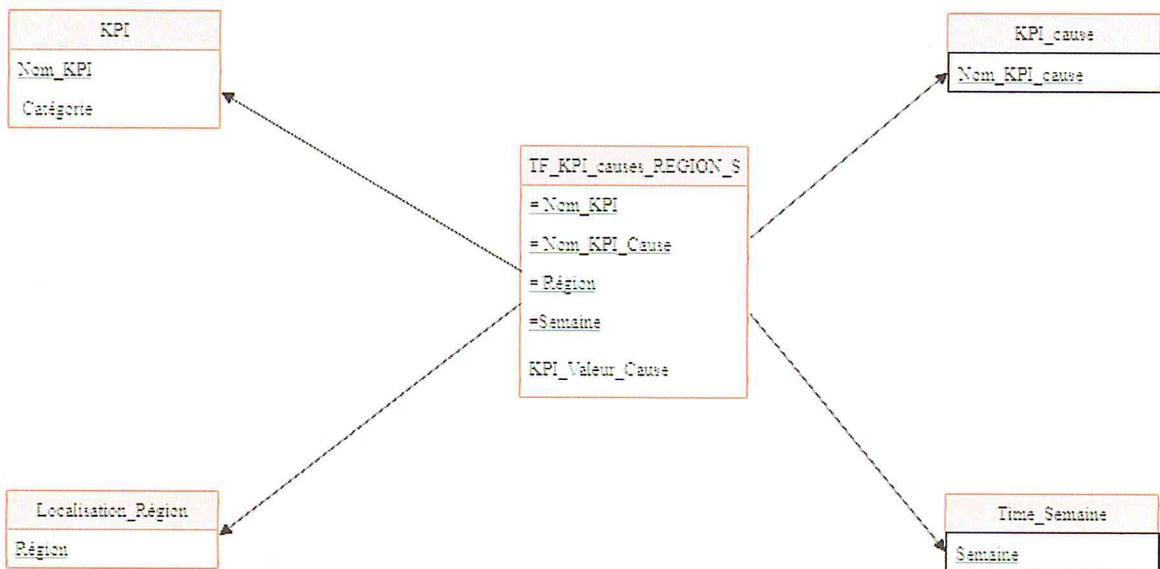


Figure 4.11 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Région et semaine

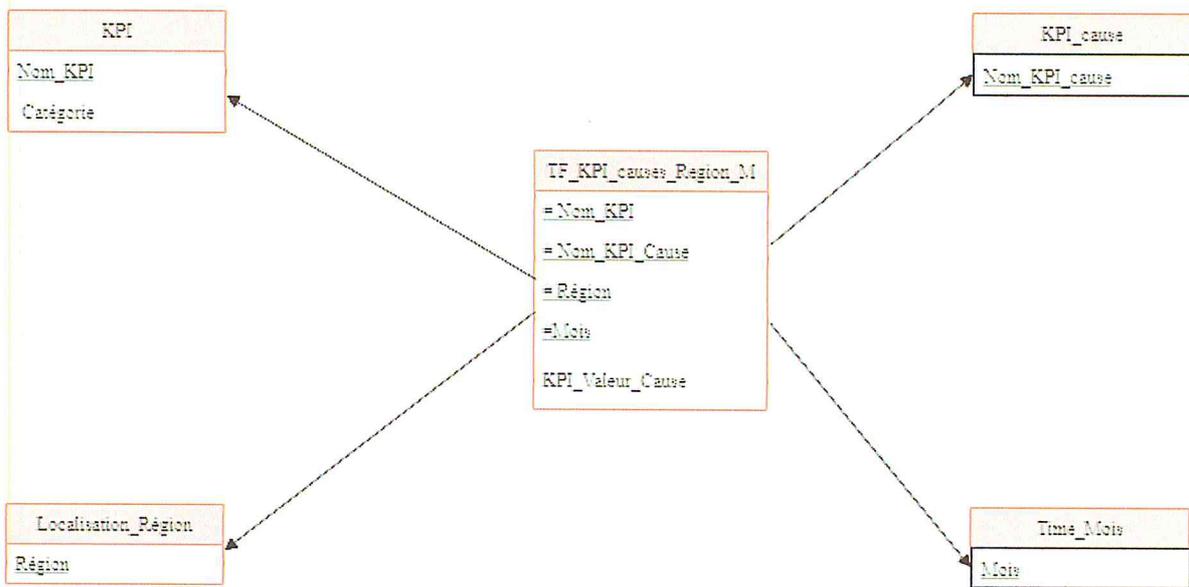


Figure 4.12 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Région et mois

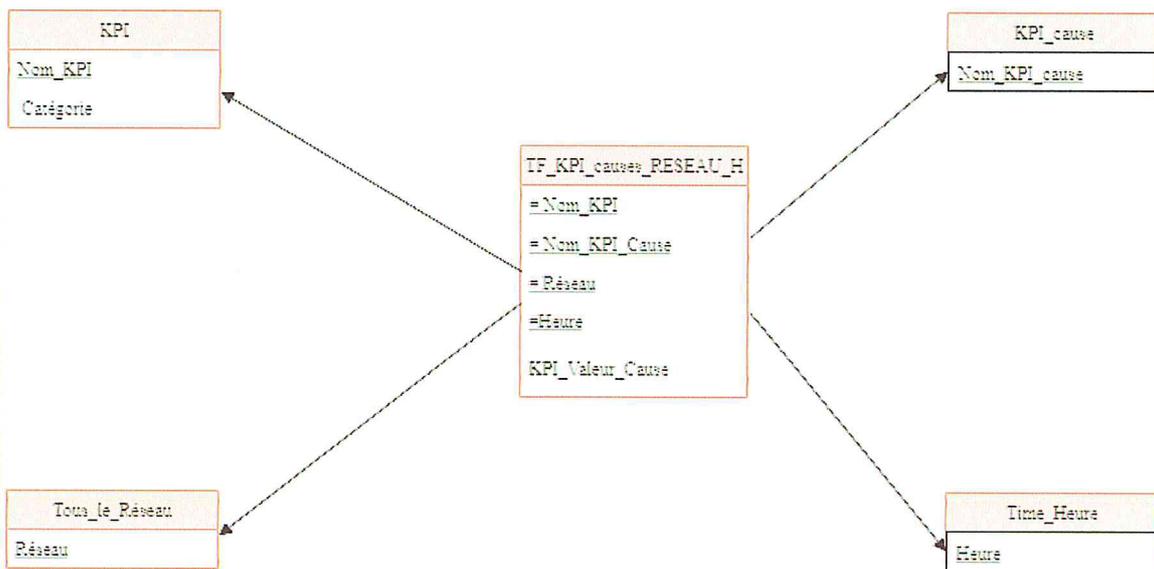


Figure 4.13 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Réseau et heure

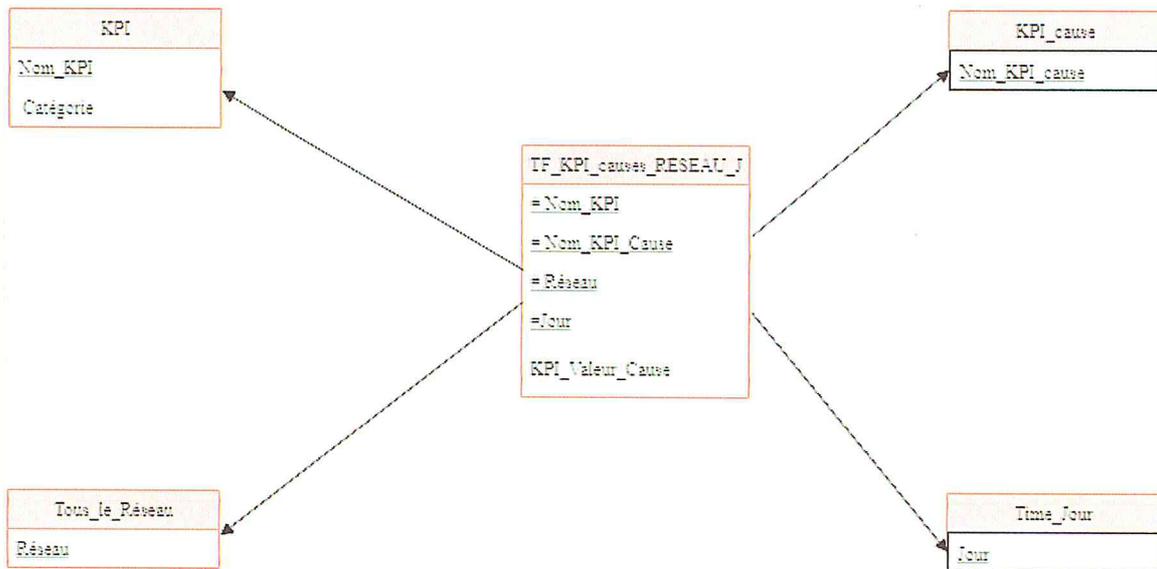


Figure 4.14 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Réseau et jour

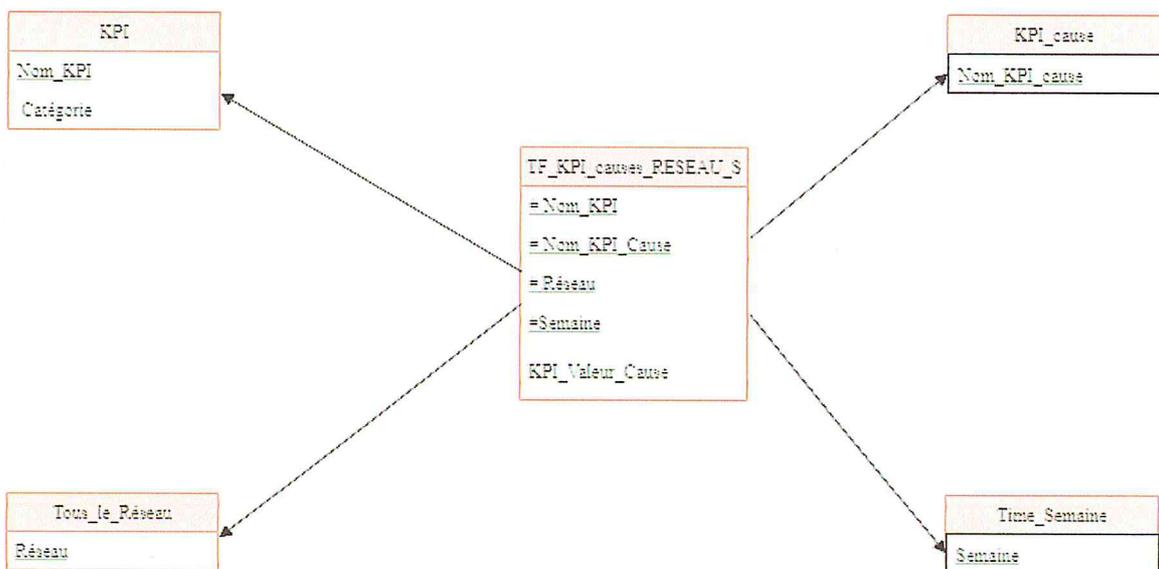


Figure 4.15 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Réseau et semaine

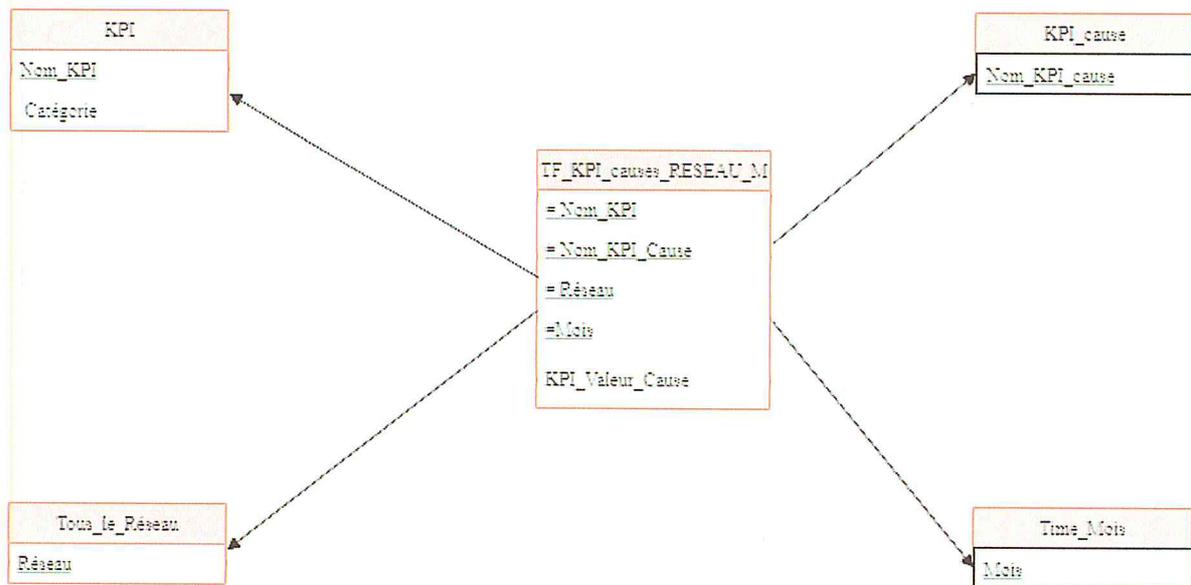


Figure 4.16 : Modèle dimensionnel relatif à l'analyse des causes d'un KPI par Réseau et mois

Schéma en constellation :

Alors que les modèles dimensionnels modélisent les données relatives à un besoin d'analyse particulier, le modèle dimensionnel en constellation modélise de façon exhaustive l'entreposage des données pour des fins d'analyse. Il est obtenu par fusion des différents modèles en étoile en représentant une seule fois, sans duplication, les dimensions communes associées avec plusieurs tables de fait.

4.5 Conclusion

En clair, comme nous venons de le voir, ce chapitre était consacré à la conception des diagrammes des cas d'utilisation qui nous ont permis de dégager l'architecture générale de notre application, nous avons défini aussi les modèles dimensionnels en étoile à base de table de faits et de table de dimensions pour des fins d'analyses, ensuite nous avons décrit le modèle dimensionnel en constellation afin d'avoir une vue générale sur notre entrepôt de données.

Dans la partie suivante nous entamons l'implémentation et la réalisation de notre système décisionnel relatif à la supervision de la qualité ainsi que l'investigation des causes de dégradation de la qualité du réseau.

5 Chapitre : Réalisation

5.1 Introduction

La phase de réalisation constitue l'aboutissement des parties précédentes. En effet, nous allons définir, dans ce chapitre, les choix en termes d'environnement et outils de développement, l'architecture du système ainsi que la description des applications. Nous procédons, ensuite, à des tests de validation du système mis en œuvre.

5.2 Choix des technologies

Notre choix des technologies s'est basé sur des facteurs tels que la consistance des données sources et la périodicité de leur génération, l'infrastructure envisagée par OTA, l'environnement technique du projet et la pérennité de la technologie pour le futur mais surtout la disponibilité des données pour l'analyse en temps quasi-réel.

Nous présentons, dans ce qui suit, l'environnement, les outils ainsi que les langages de programmation retenus pour la mise en place du système décisionnel.

5.2.1 Outils de développement

- **Pentaho Data Integration (PDI)**



Figure 5.1: Logo Pentaho Data Integration.

Pentaho est une entreprise qui propose une suite de logiciels regroupés sous le nom de Pentaho Business Analytics. Elle fut fondée en 2004 et est basée à Orlando, aux Etats-Unis.

Pentaho Business Analytics est une plate-forme décisionnelle qui offre, au travers de divers outils, une couverture globale des fonctionnalités de la Business Intelligence :

- ETL (intégration de données)
- Reporting
- Tableaux de bord
- Analyse ad hoc
- Analyse multidimensionnelle (OLAP)

-Versions communautaire et entreprise

La plate-forme Pentaho existe en version communautaire et en version entreprise. La version entreprise propose des fonctionnalités et des services supplémentaires qui n'existent pas dans la version gratuite, essentiellement des applications sur serveur et un support.

La version entreprise nécessite une souscription annuelle qui prend en compte l'achat de la licence. Les coûts de support et de maintenance sont à ajouter à la licence et s'élèvent à 10 000\$/an pour une licence de base permettant l'accès au logiciel à 25 utilisateurs. Les coûts varient, comme dans la plupart des cas, en fonction du nombre d'utilisateurs et restent très raisonnables par rapport aux tarifs pratiqués par la concurrence.

PDI, que l'on trouve encore communément sous le nom de KETTLE (Kettle ETL Environment), gère la partie intégration de données et constitue ainsi l'outil d'ETL de la plateforme. Cependant, l'application peut être utilisée indépendamment de cette dernière.

-Les modules de PDI

- Spoon

Ce module propose une interface graphique qui permet à l'utilisateur de créer facilement un processus d'ETL sans avoir à saisir de code. C'est le seul module de PDI qui propose une interface graphique. Pour tous les autres, il faut passer par l'invite de commandes.

- Pan

Ce module permet d'exécuter les transformations réalisées dans Spoon en utilisant l'invite de commandes.

- Kitchen

Kitchen permet l'exécution de tâches en mode batch et propose un plan de suivi de ces tâches pour contrôler facilement le processus d'ETL.

- Carte

Ce module est en fait un serveur web qui permet d'exécuter des tâches ET des transformations à distance. Pour cela, il prend en compte un fichier XML (en utilisant une *servlet*) qui contient la transformation ou la tâche à exécuter et la configuration d'exécution. Il permet également de contrôler à distance, démarrer ou arrêter les processus en cours sur le serveur.

Nous avons utilisé Pentaho Data Intégration pour la phase d'extraction des données brutes à partir du serveur ftp ensuite les mettre dans un entrepôt de données d'une façon périodique.

- **SQL Server :**



Figure 5.2 : Logo SQL Server.

SQL Server est le système de gestion de base de données relationnelle (SGBDR) de Microsoft. C'est une base de données complète conçue principalement pour concurrencer les solutions Oracle Database (DB) et MySQL.

Comme tous les principaux SGBDR, SQL Server prend en charge ANSI SQL, le langage SQL standard. Toutefois, SQL Server contient également T-SQL, sa propre implémentation SQL. SQL Server Management Studio (précédemment appelé Enterprise Manager) est l'outil principal de SQL Server et prend en charge les environnements 32 bits et 64 bits.

Nous avons utilisé SQL Server en premier lieu pour la création de notre entrepôt de données, en deuxième lieu pour la transformation des données brutes en données utile à l'analyse

- Netbeans :



Figure 5.3 : Logo Netbeans.

NetBeans est un environnement de développement intégré (IDE), placé en open source par Sun en juin 2000 sous licence CDDL (Common Development and Distribution License) et GPLv2. En plus de Java, NetBeans permet la prise en charge native de divers langages tels le C, le C++, le JavaScript, le XML, le Groovy, le PHP et le HTML, ou d'autres (dont Python et Ruby) par l'ajout de greffons téléchargeable du site : <https://netbeans.org/downloads>.

Nous avons choisi Netbeans en raison des compétences que nous avons acquis au cours de notre cursus universitaire. L'éditeur NetBeans est l'environnement de programmation sous lequel nous avons développé l'application d'analyse des données issues de l'entrepôt de données.

5.2.2 Langages de programmation utilisés :

- Java :



Figure 5.5 : Logo Java

Pour implémenter notre système, nous avons opté pour le langage de programmation java. En effet, Java s'annonce comme une des évolutions majeures de la programmation. JAVA est un langage de programmation orienté objet créé par James Gosling et Patrick Naughton, employés de Sun Microsystems, avec le soutien de Bill Joy (cofondateur de Sun Microsystems en 1982), présenté officiellement le 23 mai 1995 au SunWorld. La société Sun a été ensuite rachetée en 2009 par la société Oracle qui détient et maintient désormais Java.

L'application d'analyse des KPI que nous avons mis en œuvre a été développée avec le langage Java orienté objet

5.3 Plateforme d'intégration et d'analyse des indicateurs de qualité :

Le système à mettre en œuvre consiste en premier lieu à la capture des données sources à partir d'un serveur OTA regroupant tous les compteurs générés par les équipements du réseau de téléphonie mobile. Le processus d'extraction des compteurs est conçu de telle sorte à éviter d'extraire une données plus d'une fois. Pour ce faire, un contrôle sur la date de la dernière extraction est effectué. Ces données sont, ensuite, chargées dans une base cible qui est l'entrepôt de données. Plusieurs transformations sont opérées à ce niveau pour adapter les données aux objectifs d'analyse. Enfin, le système offre une application d'analyse et d'investigation afin d'expliquer les éventuelles dégradations de la qualité du réseau. Pour ce

faire, une dizaine d'indicateurs (KPI) sont définis. Chaque KPI traite sur un aspect particulier du réseau. Ces KPI sont considérés comme principaux et leur analyse met en jeu d'autres KPI secondaires expliquant plus en détail les causes liés aux principaux KPI.

Avant d'aborder le détail de fonctionnement du système, nous présentons tout d'abord l'architecture de celui-ci afin de mettre en relief les principaux composants et leurs relations comme le montre la figure 5.6 :

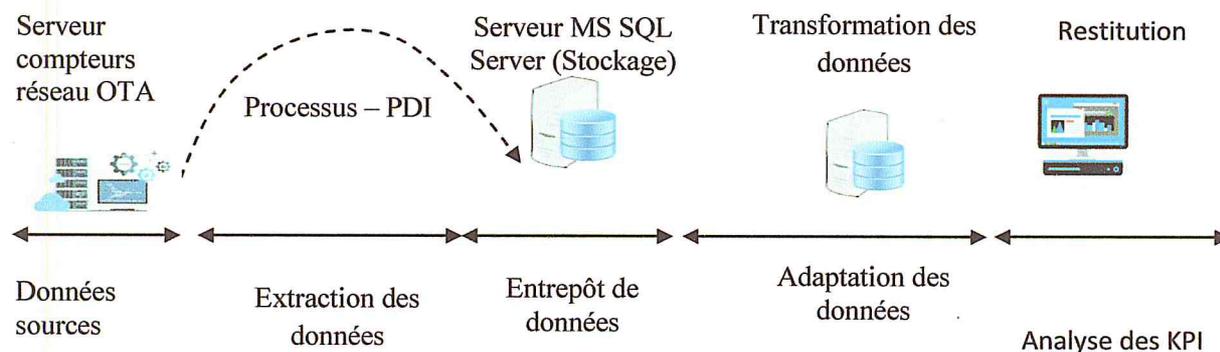


Figure 5.6 : Architecture du système décisionnel

Dans ce qui suit, nous présentons les processus développés pour l'intégration des données sources et l'analyse graphique des KPI. L'extraction des données a été développée sur la plateforme Pentaho Data Integration (PDI). Quant à l'application d'analyse, celle-ci a été mise en oeuvre avec le langage Java sous l'éditeur de développement intégré NetBeans IDE 8.2.

5.3.1 Présentations des interfaces de l'application :

▪ 3.1.1 Processus d'extraction des données :

Comme le montre la figure 5.7, le processus principal d'extraction consiste à exécuter les tâches d'extraction des données sources disponibles sur le serveur FTP. Le processus démarre par la tâche *Exécution_Transformation_SIM* qui permet l'exécution d'une transformation *get_date* qu'on expliquera plus tard dans la Figure 5.8, puis la tâche *get_from_ftp_job_SM* qui consiste à se connecter sur le serveur FTP regroupant tous les compteurs générés par les équipements de l'infrastructure télécom d'OTA, Enfin, la tâche *process_download_files_SIM* qui permet l'exécution d'une autre tâche *bulk_load_to_mssql* qui sera détaillé plus tard dans la Figure 5.11.

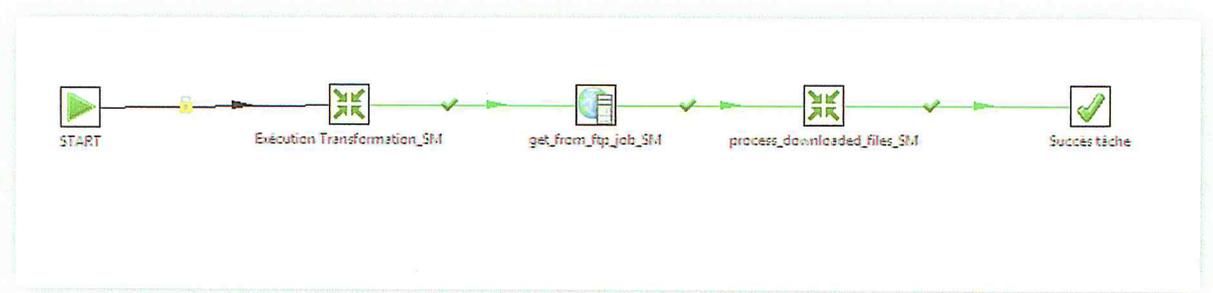


Figure 5.7 : Processus principal d'extraction des données sources.

La figure 5.8 représente une transformation qui permet de récupérer la date et l'heure actuelle de la machine et l'enregistrer dans une variable d'environnement qui sera utilisée dans la tâche suivante (*get_from_ftp_job_SM*).

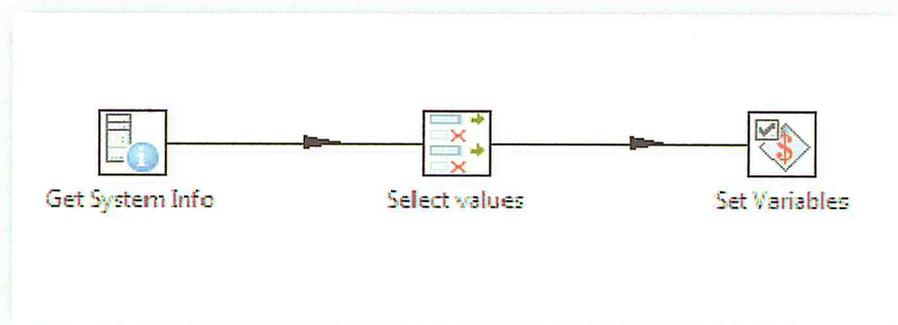


Figure 5.8 : transformation (get_date).

La figure 5.9 représente la tâche *get_from_ftp_job_SM* permettant la connexion au server FTP.

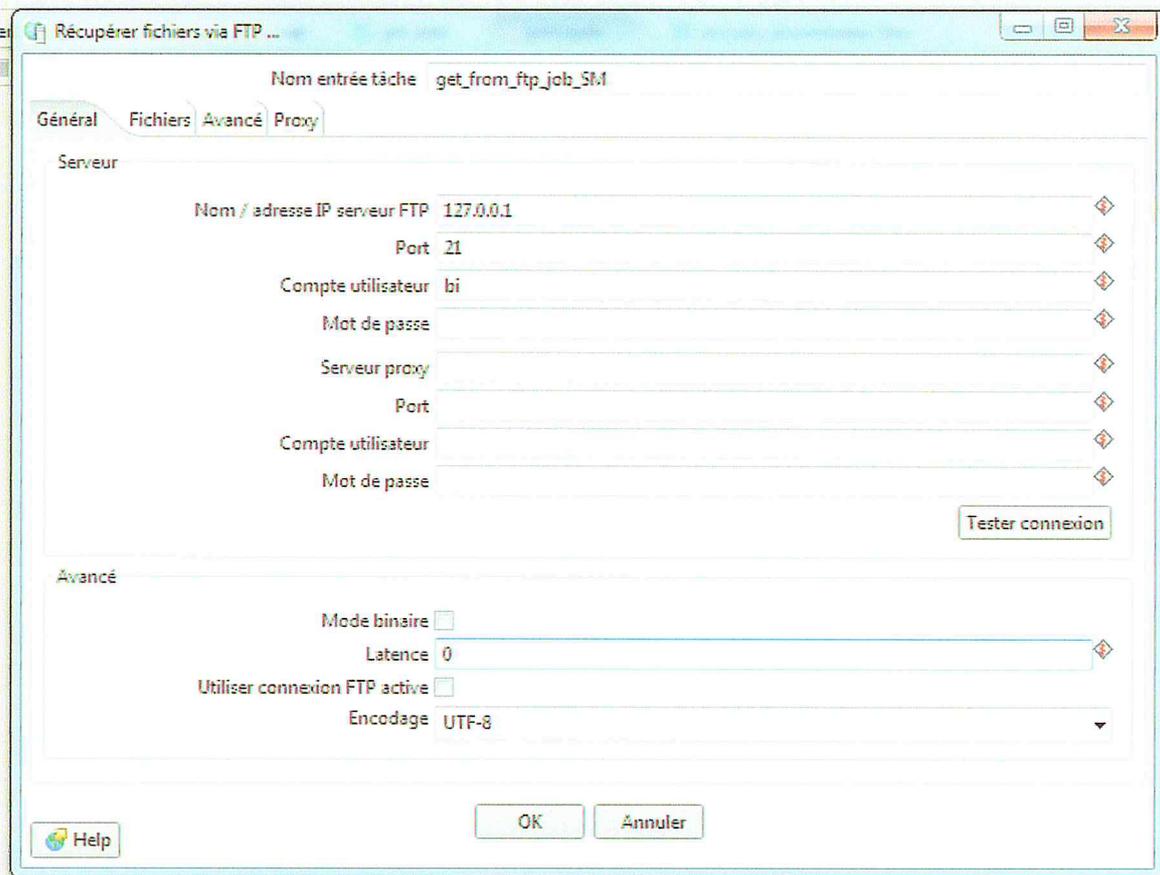


Figure 5.9 : Interface de paramétrage de la tâche `get_from_ftp_job_SM` pour la connexion au serveur FTP

Figure 5.10 représente le paramétrage des fichiers qu'on souhaite récupérer depuis le serveur FTP et cela en donnant d'abord, le chemin du répertoire source en fonction de la variable d'environnement et ensuite le caractère joker à prendre en considération lors du filtrage des fichiers recherchés.

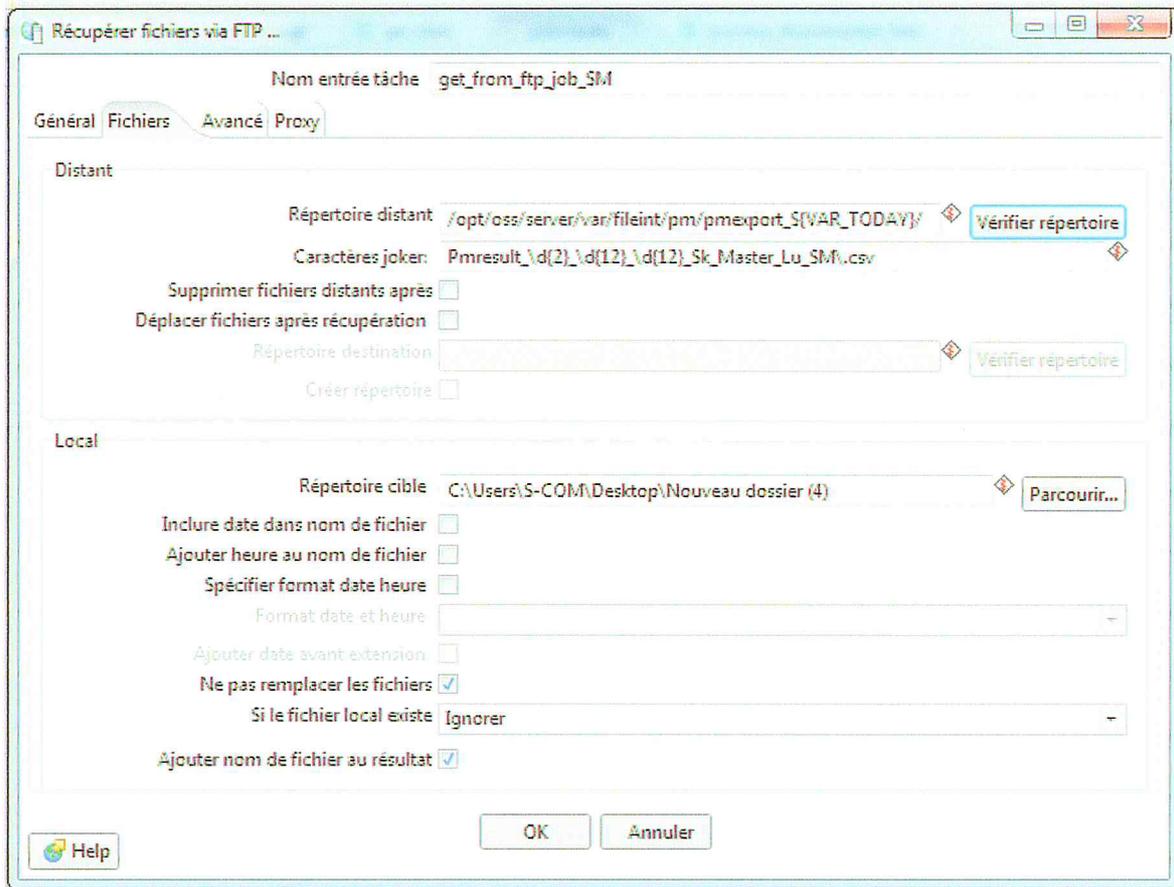


Figure 5.10 : Paramétrage des fichiers.

La tâche *process_upload_file* représentée dans la figure 5.11 permet d'abord la connexion à l'entrepôt de données. Elle permet, ensuite, de filtrer les fichiers pour ne laisser passer que ceux non encore chargés dans l'entrepôt de données.

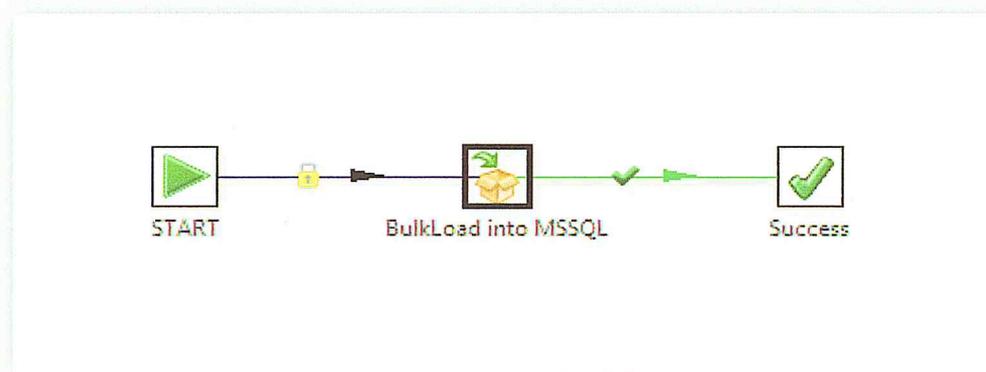


Figure 5.11 : Connexion à l'entrepôt de données.

La tâche représentée dans la figure 5.12 permet de vérifier si les fichiers restitués par la tâche *get_from_ftp_job_SM* ont été déjà chargés dans l'entrepôt de données auquel cas le chargement sera annulé. Dans le cas contraire, les données sont considérées nouvelles et seront alors chargés dans l'entrepôt de données. Cette tâche constitue un filtre qui ne laisse passer que les fichiers contenant de nouvelles données permettant le rafraichissement de l'entrepôt de données

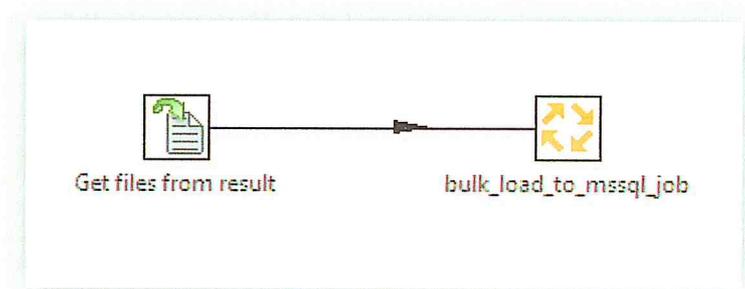


Figure 5.12 : Contrôle et chargement des données sources dans l'entrepôt de données.

▪ **Transformation des données :**

Après l'extraction des données brutes, nous passons maintenant à la transformation de ces dernières ; pour ce faire, nous avons créé tout d'abord une table historique sur laquelle se baseront toutes les autres tables de l'entrepôt de données. Les principales transformations à opérer sur les données peuvent se résumer de la manière suivante :

- a) Suppression des colonnes n'ayant aucune valeur pour l'analyse : Grâce à une opération de projection, les colonnes suivantes seront exclues dans la suite du processus d'analyse des KPI.
 - 3G MS init PDP context sec act fail
 - 3G MS init PDP context sec act success
 - 3G MS init PDP sec context act
 - lu Mode MS Init PDP Active Reject with Dynamic Address per RAI
 - lu Mode MS Init PDP Active Success with Dynamic Address per RAI
 - lu Mode MS Init PDP Active with Dynamic Address per RAI
 - lu Mode PDP Context Secondary Active Fail (Collision with network initiated request) per RAI
 - Intra SGSN GSM to UMTS change request times per RA

- b) Pour une analyse par rapport aux dimensions de localisation, les trois colonnes suivantes seront rajoutées :

- wilaya_code
- wilaya_name
- region_optim

c) Agrégation des données selon les formules suivantes :

- (3G GGSN init PDP context dact fail) = (3G GGSN init PDP context dact) – (3G GGSN init PDP context dact success)
- (3G GGSN init PDP context mod fail) = (3G GGSN init PDP context mod) – (3G GGSN init PDP context mod success)

Toutes les colonnes suivantes seront concernées par des agrégations :

- 3G MS init PDP context act fail
- 3G MS init PDP context dact fail
- 3G MS init PDP context mod fail
- 3G MS totally init PDP context act fail
- 3G NET init PDP context act fail
- 3G NET init PDP context dact fail
- 3G RNC init PDP context mod fail
- 3G SGSN init PDP context dact fail
- 3G SGSN init PDP context mod fail
- 3G totally PDP context dact fail
- 3G totally PDP context mod fail

d) Ajouter de nouvelles colonnes par extraction de sous-chainés à partir de la colonne Object name:

- [Object_name] : la sous-chaine de caractères allant de 1^{er} au 11^{ème} caractère de la colonne Object name
- LAI : la sous-chaine de caractères allant de 28^{er} au 32^{ème} caractère de la colonne Object name
- RAC : la sous-chaine de caractères allant de 34^{er} au 39^{ème} caractère de la colonne Object name.

Une fois les transformations sur la table historique terminées, nous passons ensuite à la création des différentes tables constituant l'entrepôt de données.

- + [table icon] dbo.create_profile
- + [table icon] dbo.Help_SK_3G_Wilaya
- + [table icon] dbo.Help_SK_week_new
- + [table icon] dbo.lu_SM_Etat
- + [table icon] dbo.lu_SM_Hist_D_Lac
- + [table icon] dbo.lu_SM_Hist_D_Lac_KPI
- + [table icon] dbo.lu_SM_Hist_D_Net
- + [table icon] dbo.lu_SM_Hist_D_Net_KPI
- + [table icon] dbo.lu_SM_Hist_D_USN
- + [table icon] dbo.lu_SM_Hist_D_USN_KPI
- + [table icon] dbo.lu_SM_Hist_H_Lac
- + [table icon] dbo.lu_SM_Hist_H_Lac_KPI
- + [table icon] dbo.lu_SM_Hist_H_Lac_ref
- + [table icon] dbo.lu_SM_Hist_H_Lac_ref_PK
- + [table icon] dbo.lu_SM_Hist_H_Net
- + [table icon] dbo.lu_SM_Hist_H_Net_KPI
- + [table icon] dbo.lu_SM_Hist_H_USN
- + [table icon] dbo.lu_SM_Hist_H_USN_KPI
- + [table icon] dbo.lu_SM_Hist_Hr_Lac
- + [table icon] dbo.lu_SM_Hist_Hr_Lac_KPI
- + [table icon] dbo.lu_SM_Hist_Hr_Net
- + [table icon] dbo.lu_SM_Hist_Hr_Net_KPI
- + [table icon] dbo.lu_SM_Hist_Hr_USN
- + [table icon] dbo.lu_SM_Hist_Hr_USN_KPI
- + [table icon] dbo.lu_SM_Hist_W_Lac
- + [table icon] dbo.lu_SM_Hist_W_Lac_KPI
- + [table icon] dbo.lu_SM_Hist_W_Net
- + [table icon] dbo.lu_SM_Hist_W_Net_KPI
- + [table icon] dbo.lu_SM_Hist_W_USN
- + [table icon] dbo.lu_SM_Hist_W_USN_KPI

▪ **Interface d'authentification :**

L'utilisateur doit s'authentifier avec un nom d'utilisateur et un mot de passe pour qu'il puisse ouvrir une session sur le système. Cette phase met en valeur l'aspect de sécurité : la vérification du compte utilisateur et l'attribution des droits et des privilèges relatifs à sa session (voir Figure 5.13).



Figure 5.13 : Interface d'authentification d'un utilisateur au moment de l'accès au système.

La figure 5.14 représente le menu principal de l'application d'analyse qui couvre les trois technologies de la téléphonie mobile à savoir 2G, 3G et 4G. Pour chacune d'elles, nous distinguons les trois sections MM (*Mobility Management*), SM (*Session Management*) et Trafic. Par ailleurs, le menu comporte deux boutons permettant respectivement l'inscription de nouveaux utilisateurs et la déconnexion de la session courante.

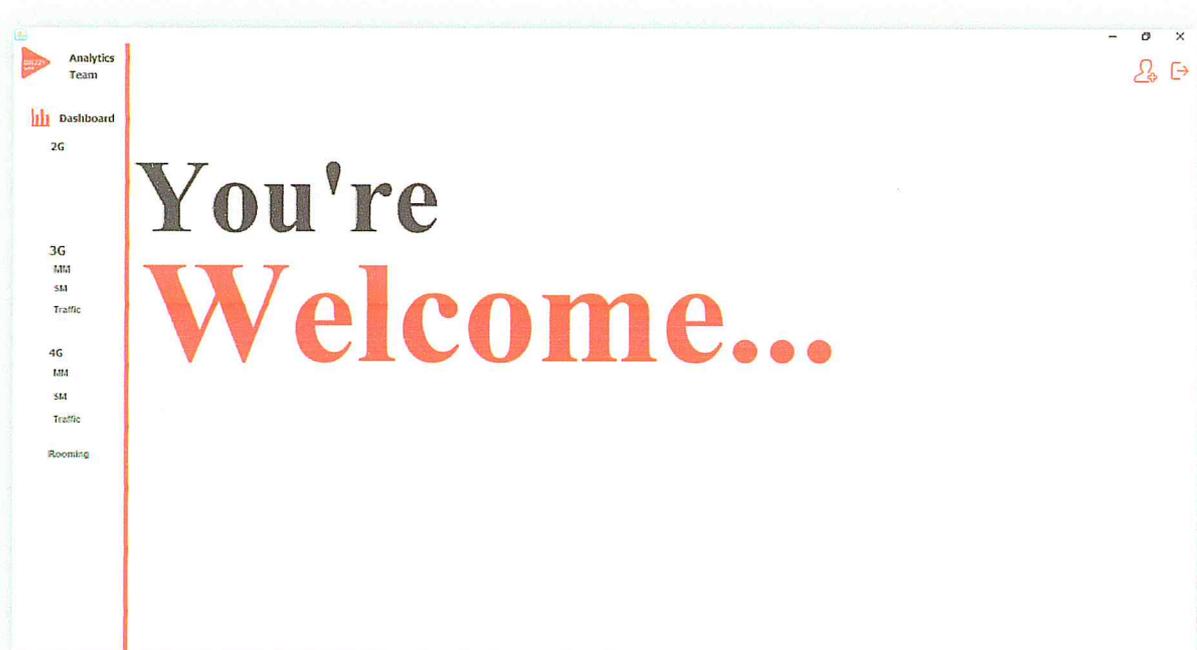


Figure 5.14 : Interface principale du tableau de bord.

Le travail demandé, dans le cadre du projet de fin d'études, est l'analyse des différents KPI ainsi que leurs causes relatives à la sections SM (*Session Management*) de la partie 3G. Cette section consiste à l'authentification, l'autorisation et la compatibilité des services de contrôle d'accès réseaux et permet aux abonnés d'accéder au réseau par l'intermédiaire de la technologie d'accès réseau choisie, qu'il soit mobile, sans fil, fixe ou par câble et garantit une expérience d'accès transparente, même pour les utilisateurs passant d'un type de réseau a l'autre.

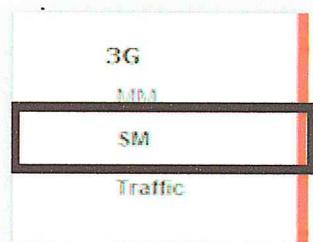


Figure 5.15: Menu principale

La figure 5.16 représente une interface pour ajouter des nouveaux utilisateurs, pour ce faire l'administrateur doit donner un nom d'utilisateur (*User name*), un niveau d'utilisateur (suivant des droit d'accès), un mot de passe (*password*) et le confirmer, et une adresse e-mail. Enfin, valider ou quitter l'interface.

Submit form

User name

Password

Confirm Password

Email

Admin ▼
Admin
User level 1
User level 2
User level 3

Back Create

Figure 5.16 : interface d'inscription de nouveaux utilisateurs.

Dans la figure 5.17 l'administrateur peut choisir le filtre sur la dimension temporelle (*Hourly*, *Daily*, *Weekly*, *Monthly*), ou donner une période (date début et date fin) selon le choix d'usage (jour, mois, année) comme le montre la figure 5.18. Il faut noter que le filtre par jour (*Daily*) est choisi comme filtre par défaut.



Figure 5.17 : Dimension temporelle de l'analyse des KPI.

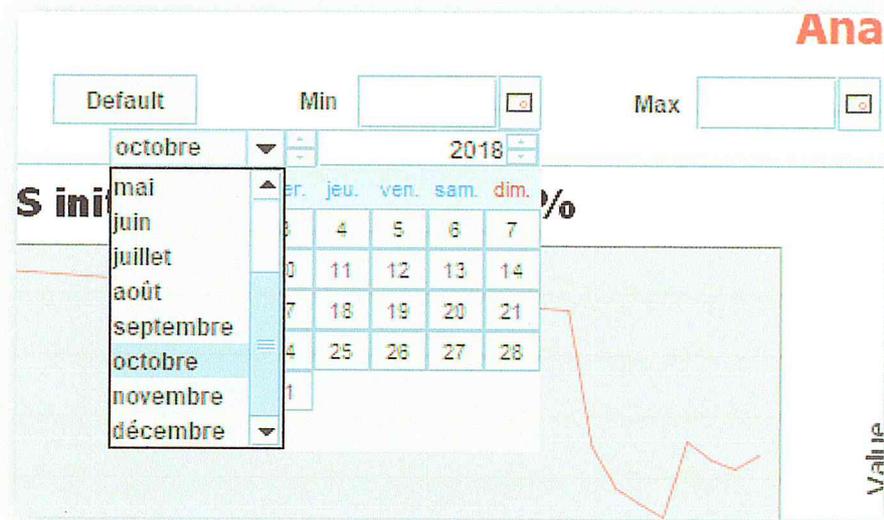


Figure 5.18 : Choix de la période pour l'analyse des KPI.

Scenario 1 : Analyse du KPI 1 :

La Figure 5.19 représente la variation du premier compteur « *Main KPI 1 = 3G MS init PDP context act fail* » par date et par réseau (*All Network*). On peut observer que pendant la période allant du 7 août 2018 au 10 septembre la valeur d'échec est égale à 0.085 ; autrement dit 8.5% du nombre total d'abonnés n'arrivent pas à passer leur appel soit plus de 170 000 personnes vu que le nombre total des abonnés dépasse les deux millions. En revanche, à partir du 11 septembre 2018 jusqu'au 14 septembre 2018, on remarque une baisse de la valeur d'échec qui est égale à 0.049 ; autrement dit 4.9% soit 98 000 personnes n'arrivant pas à passer un appel.

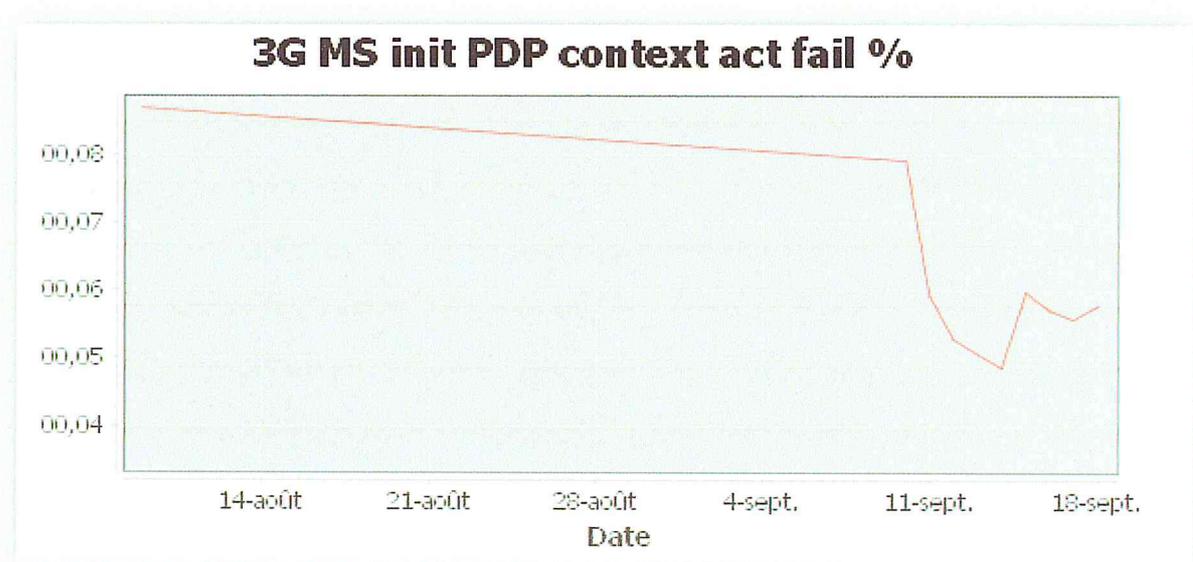


Figure 5.19 : Analyse de KPI 1 « 3G MS init PDP context act fail » par date et par réseau

Scenario 2 : Analyse du KPI 2 :

La Figure 5.20 montre l'évolution du deuxième compteur « *RANAP RAB Assignment fail* » par date et par réseau (*All Network*). Dans ce graphe, la valeur du KPI analysée exprime la défaillance (impossible d'établir une connexion de transport). On remarque que le taux de défaillance augmente du 08 août au 10 septembre 2018 et arrive à presque 0.02 ; autrement dit 2% soit 40 000 échecs pour cause de défaillance.

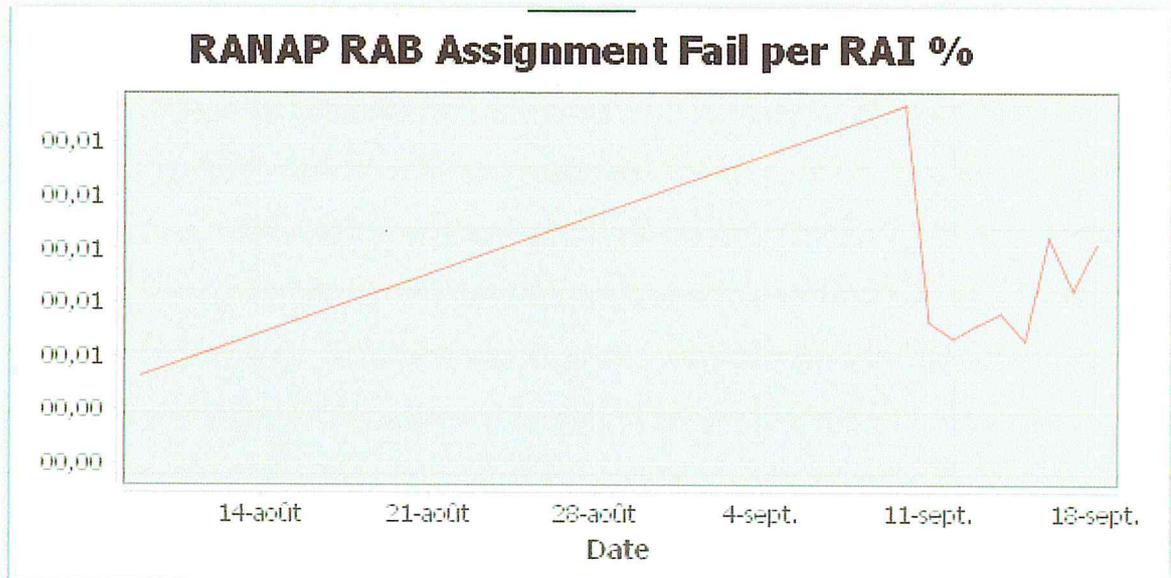


Figure 5.20 : Analyse de KPI 2 (RANAP RAB Assignment fail) par date et par réseau.

Scenario 3 : Analyse des causes du KPI 1 :

La Figure 5.21 montre, sous forme d'histogrammes groupés, la variation des compteurs secondaires (KPI_cause) du KPI 1(3G MS init PDP context act fail) par date et par réseau (All Network). Chaque cause est représentée par une couleur spécifique et par ordre spécifique. Par exemple, le jaune représente le KPI_cause « lu Mode PDP Context Active Fail (APN Restriction) per RAI » On remarque que l'histogramme Jaune a une baisse de 0.99 à 0.98 le 14 août 2018 ce qui fait 0.01 de différence ; autrement dit 1% soit 20 000 tentatives d'activation de « context pdp initier » par l'UE (User Equipement) a échoué en raison d'une restriction APN.

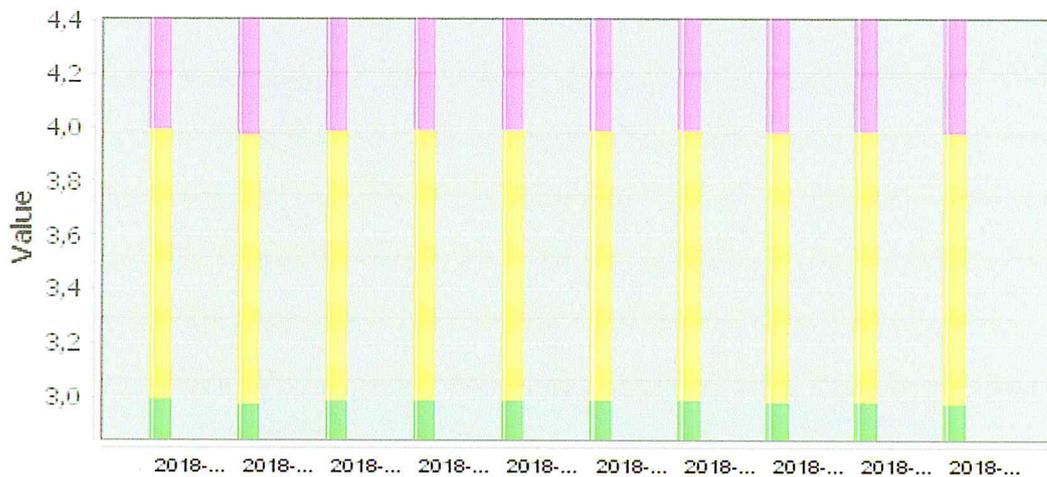


Figure 5.21 : Analyse des causes de KPI 1 (3G MS init PDP context act Success) par date et par réseau.

Scenario 4 : Analyse des causes du KPI 2 :

La Figure 5.22 représente, sous forme d'histogrammes groupés, l'évolution des compteurs secondaires (KPI_cause) du KPI 2 « RANAP RAB Assignment fail» par date et par réseau (All Network). Chaque cause est représentée par une couleur spécifique et par ordre spécifique. Par exemple, le bleu représente le KPI_cause « RANAP RAB Assignment Fail (Lu Transport Connection Failed to Establish) per RAI % » On remarque que l'histogramme bleu est presque a 0.99 tout les jours ce qui fait que 0.01 soit 1% des UE n'arrive pas a établir une connexion a cause d'un problème de transport.

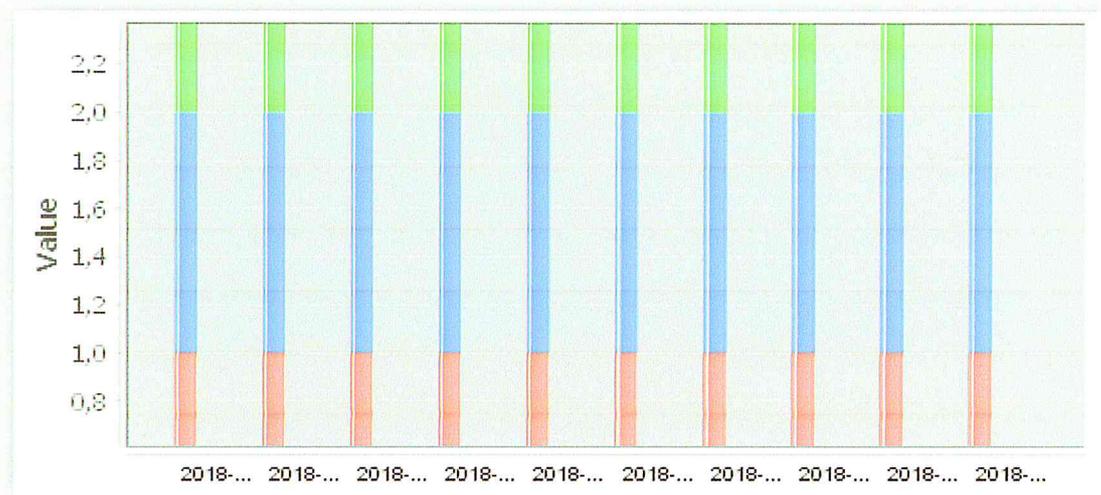


Figure 5.22 : Analyse des causes de KPI 2 (RANAP RAB Assignment success) par date et par réseau.

Scenario 5 : Analyse des KPI 3, KPI 4, KPI 5, KPI 6, KPI 7, KPI 8, KPI 9, KPI 10 par réseau :

La Figure 5.23 représente la variation des autres compteurs (Other KPI) par date et par réseau (All Network). Si on prend par exemple le graphe bleu « 3G NET init PDP context act fail % » retourne le nombre d'activations du contexte PDP initiées par la MS (monitoring system) ayant échoué dans une RAI (Routing Area Identification) spécifiée, on remarque qu'il 'est stable avec une valeur de 0.05 soit 5% des tentatives d'activation on échoué.



Figure 5.23 : Analyse des autres KPI (Other SM KPI fail) par date et par réseau

Scenario 6 : Analyse du KPI 1 et de ces causes par région :

La Figure 5.24 représente la variation du premier compteur (*Main KPI 1*) « 3G MS init PDP context act fail » ainsi que ses causes par date et par région. On remarque que le graphe vert « Iu Mode PDP Context Active Fail(User Authentication Fail) per RAI % » prend une valeur de 0.8 le 15/09/2018 soit 20% d'échec d'authentification.

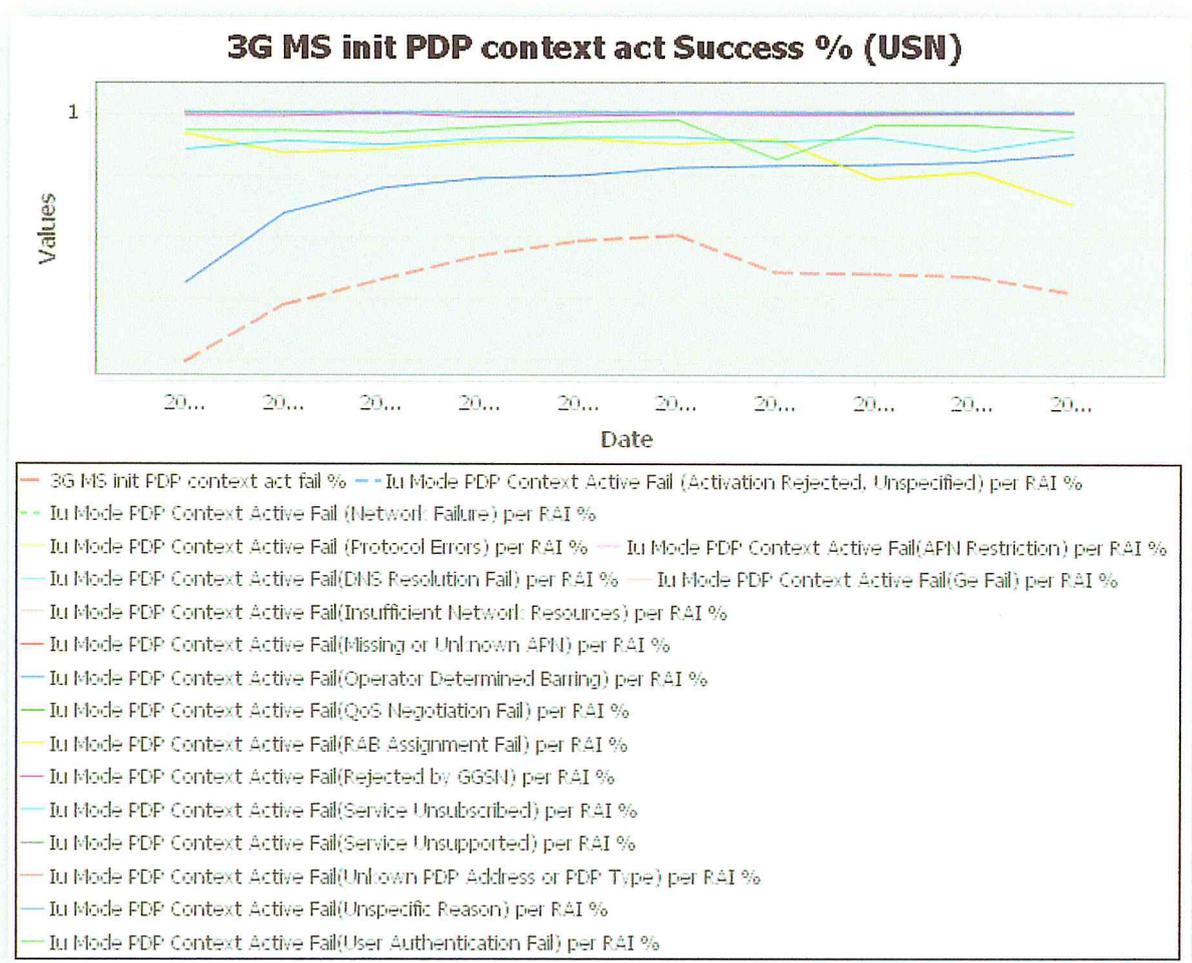


Figure 5.24 : Analyse de KPI 1 et ses causes par date et par région.

Scenario 7 : Analyse du KPI 2 avec ses causes par région :

La Figure 5.25 représente la variation du deuxième compteur (*Main KPI 2*) « *RANAP RAB Assignment Fail per RAI* » ainsi que ses causes par date et par région. On prend par exemple le graphe jaune « *RANAP RAB Assignment Fail (User Plane Versions not Supported) per RAI %* » qui prend presque 1 soit presque 100% de succès de la procédure d'attribution RAB dans le cas contraire (échec) le SGSN reçoit un message RAB ASSIGNMENT RESPONSE contenant la cause de la défaillance « Versions du plan utilisateur non prises en charge ».

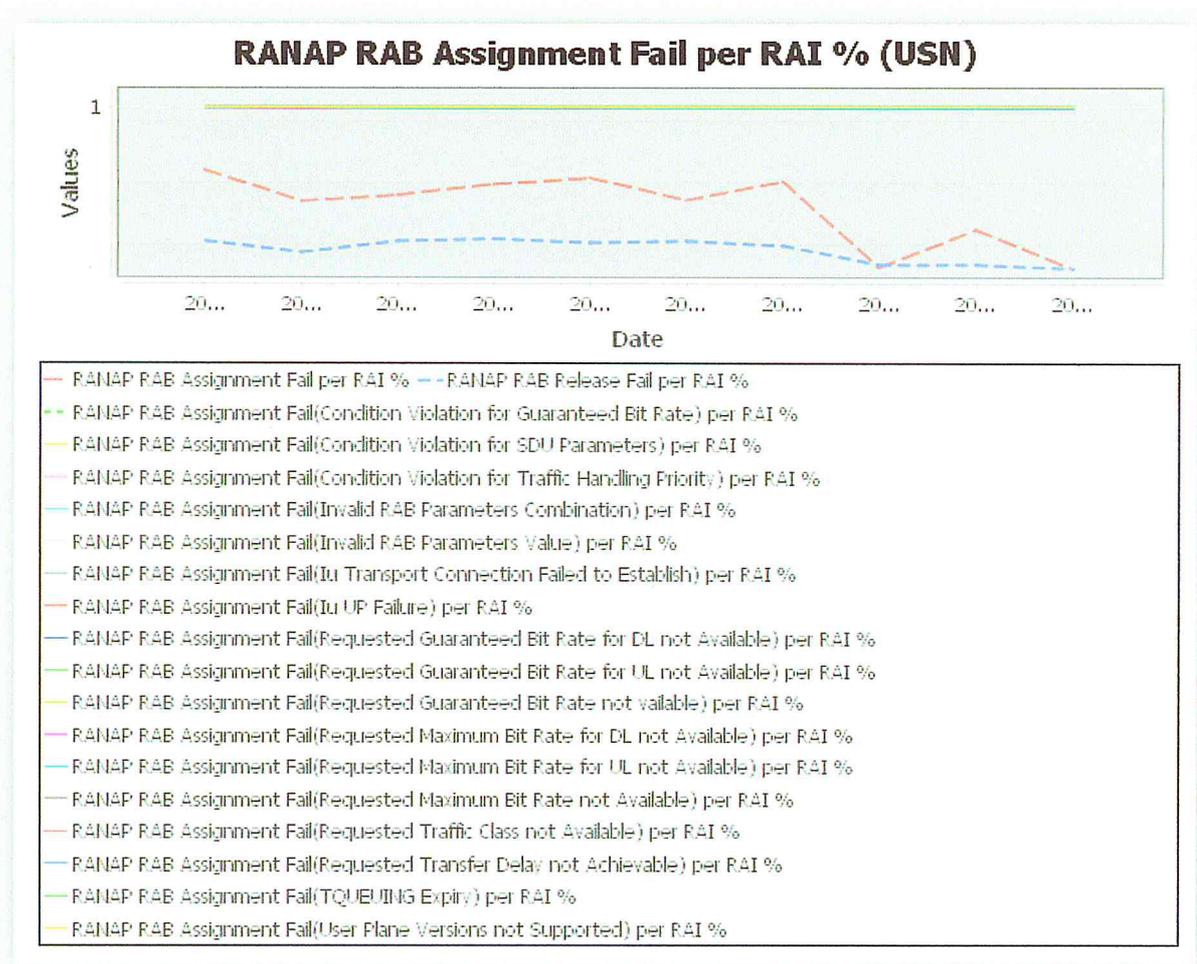


Figure 5.25 : Analyse de KPI 2 et ses causes par date et par région.

Scenario 8 : Analyse des KPI 3, KPI 4, KPI 5, KPI 6, KPI 7, KPI 8, KPI 9, KPI 10 par région :

Figure 5.26 nous montre un histogramme qui représente les valeurs que prennent les autres compteurs (*Other KPI*) par jour et par région.

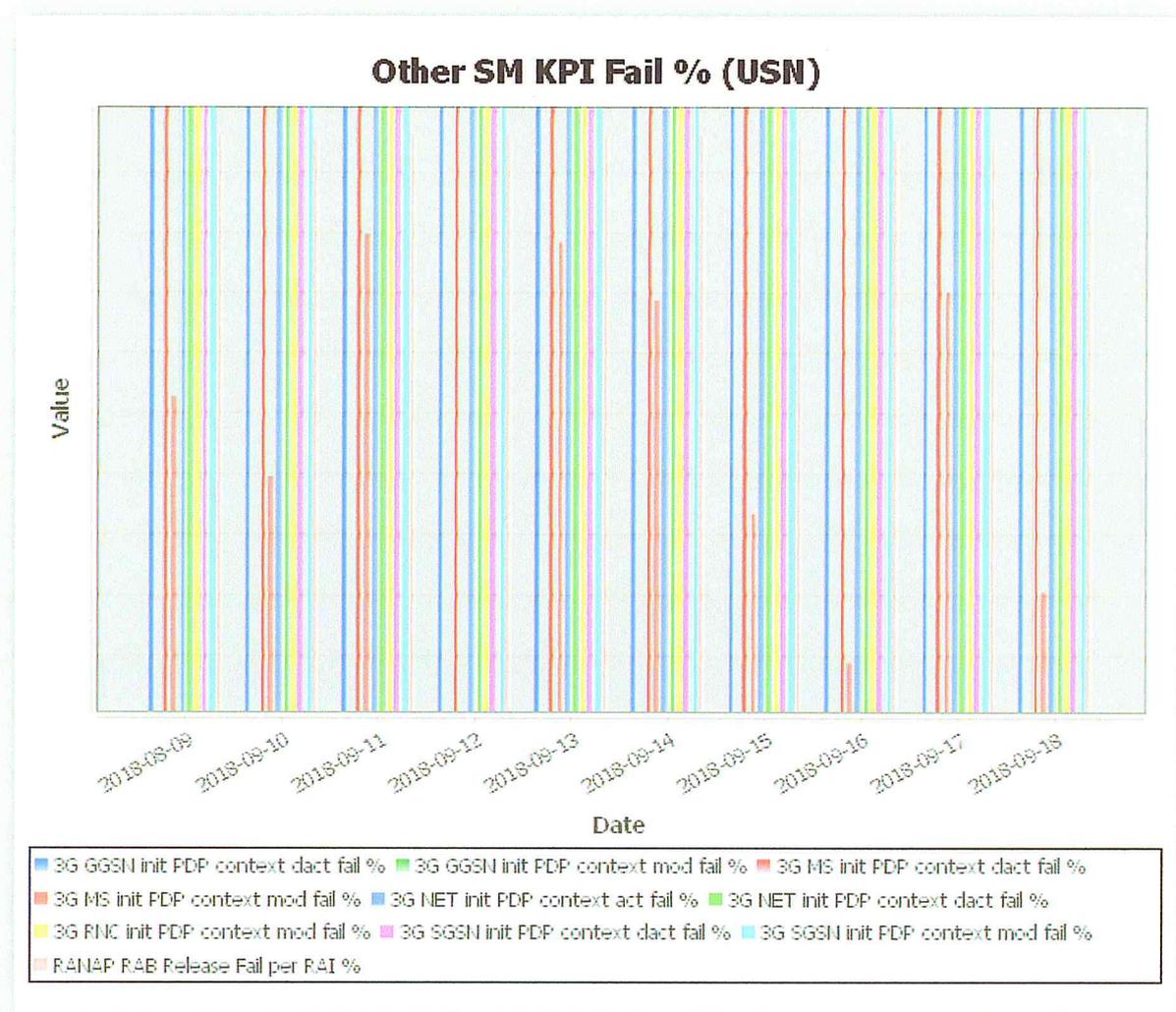


Figure 5.26 : Autres KPI par région.

5.4 Conclusion :

Dans ce chapitre, nous avons introduit la réalisation de notre projet et présenté les choix des technologies ainsi que la justification de nos choix. Nous avons, ensuite, présenté les outils de développement retenus dans la création de notre processus ETL et dans l'implémentation de notre application d'analyse.

Nous avons enchainé par les détails de la plateforme d'intégration et expliquer les différentes tâches de transformations ainsi que toutes les mesures prises en considération dans le processus d'extraction suivi d'une illustration d'interfaces graphiques.

Une fois le processus bien défini, nous avons abordé les transformations à faire subir aux données sources. Ces dernières représentent les valeurs brutes prises par les indicateurs de qualité au fil du temps. Afin de leur donner une valeur analytique et les rendre plus pertinentes, ces données brutes ont

été transformées selon des règles métier. Les données préparées, nettoyées et agrégées seront publiées pour des applications d'analyse et d'aide à la décision. Une application d'analyse a été mise en œuvre afin de tester l'exploitation de l'entrepôt de données. En dehors de celle-ci, d'autres applications peuvent être développées pour analyser d'autres aspects dans le fonctionnement du réseau de téléphonie mobile.

Enfin, nous avons testé notre application d'analyse et illustré différents scénarios qui montrent au quelques résultats sur la qualité du réseau ainsi que les causes permettant d'investiguer sur une éventuelle dégradation de la qualité.

6 Conclusion générale

Dans le cadre du projet de fin d'études, notre travail consiste à réaliser une solution décisionnelle pour le compte de l'opérateur téléphonique Djezzy. Le processus ETL que nous avons mis en œuvre est une solution d'intégration des données en temps quasi-réel.

L'objectif de ce travail est de réaliser un système d'aide à la décision pour l'analyse des indicateurs clés de performances (*KPI: Key Performance Indicator*) de l'infrastructure télécom de Djezzy fonctionnant dans un environnement caractérisé essentiellement par des données qui arrivent de façon continue et en temps quasi-réel.

Pour ce faire, nous avons décomposé notre travail en plusieurs étapes. La première a été consacrée à la modélisation et création de la base de données, la deuxième consistait à l'extraction des données brutes depuis un serveur FTP, la troisième consistait à la transformation des données, (la quatrième consistait au chargement des données dans un entrepôt et enfin la dernière étape consiste en l'analyse et l'évaluation des compteurs (KPI) relatifs à la qualité de l'infrastructure télécom.

Comme tout travail, ce projet de fin d'études devra faire l'objet d'améliorations et d'extensions. Afin de rendre l'analyse de la qualité plus précise et exhaustive, nous recommandons de traiter les points suivants dans la continuité de ce travail :

- a) Traitements des autres fichiers de compteurs générés par l'infrastructure télécom de Djezzy
Nous n'avons traité, dans ce projet, qu'une partie des compteurs générés par le réseau. Il conviendra dans le futur de récupérer les autres compteurs afin de disposer de tous les indicateurs permettant d'analyser de façon exhaustive la qualité du réseau.
- b) Réaliser toutes les transformations dans la plateforme Pentaho Data Integration
Au lieu de traiter une partie des transformations dans PDI et l'autre sur le Serveur MS SQL Server, il sera plus intéressant de centraliser toutes les transformations sur la plateforme PDI.
- c) Génération de rapports
L'application d'analyse pourra être complétée par une fonctionnalité de génération de rapports à destination des responsables. Le rapport devra contenir toutes les informations détaillant la source du problème à résoudre.
- d) Alertes en cas de disponibilité de nouvelles données

Le système pourra être étendu pour lancer dans l'avenir des alertes aux administrateurs dédiés à la supervision du réseau sous forme de SMS ou mails les informant que de nouvelles informations plus récentes sont disponibles sur la qualité du réseau.

Au niveau personnel, nous avons bénéficié pleinement de cette expérience professionnelle. En fait, ce projet a été une opportunité de s'approcher et de connaître de près le monde professionnel et d'acquérir en retour de nouvelles compétences.

Bibliographie

- [1] J.P Muller, «*Le réseau GSM et le mobile*», support de cours, Juillet 2002.
- [2] E. Meurisse, «*L'UMTS et le haut-débit mobile* », Livre, Février 2007.
- [3] Emmanuel Ferragu, «Modélisation des Systèmes d'Information Décisionnels Techniques de modélisation conceptuelle et relationnelle des entrepôts de données», Livre, Septembre 2013.
- [4] Ralph Kimball, «The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data», Septembre 2004.
- [5] SAAD Rachid., BI & Big Data [en ligne] <https://www.saadrachid.net/>, consulté le 25 avril 2018.
- [6] KHOURI. S, « Modélisation conceptuelle à base ontologique d'un entrepôt de données», Avril 2009.
- [7] Bill Inmon, « Modélisation the Architecture for the Next Generation of Data Warehousing», Juillet 2008.
- [8] Joachim Tisal, «*GSM. Réseau et services, 2ème édition*», Livre, Janvier 2005.
- [9] Pierre Lescuyer, «*Réseaux 3G : Principes, architectures et services de l'UMTS*», Livre, Mai 2006.
- [10] Grady Booch, James Rumbaugh, Ivar Jacobson, «*Le guide de l'utilisateur UML*», Livre, Juin 2000.
- [11] Grady Booch, James Rumbaugh, Ivar Jacobson, «*UML - Coffret de 2 livres*», Livre, Juillet 2017.
- [12] Ralph Kimball et Laura Reeves, «Le data warehouse : Guide de conduite de projet», Février 2005.
- [13] André Pérez, « Architecture des réseaux de mobiles : GSM/GPRS, UMTS/HSPA, EPS, NGN, IMS», Octobre 2011.

[14] Ralph Kimball et Laura Reeves, «Le data warehouse : Guide de conduite de projet»,
Février 2005.

