

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE



Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique
Université AMMAR THELIDJI Laghouat



FACULTE DES SCIENCES
DEPARTEMENT DE MATHEMATIQUE ET INFORMATIQUE

Mémoire de fin d'études
En vue de l'obtention du Diplôme d'ingénieur d'état en
informatique

Option : *systèmes d'informations avancées*

Thème

**Réalisation d'un SOLAP pour l'analyse de
l'historique sylvicole de la Forêt de Montmorency**

Réalisé par :

- ❖ Benguettaf Nouredine.
- ❖ Kachna Attia.

Encadré par :

- ❖ Nardjes Hamini.

Année universitaire : 2011/2012

RÉSUMÉ :

Dans ce sujet, nous abordons les problèmes liés à l'intégration de données spatiotemporelles au sein d'un entrepôt de données. Dans de nombreux cas, les spécifications des jeux de données géospatiales évoluent et les données sont hétérogènes à la fois des points de vue temporel, spatial et sémantique. Afin d'explorer et d'analyser des jeux de données spatiotemporelles dans une application SOLAP (Spatial On-Line Analytical Processing), nous utilisons les notions de dimension spatiale géométrique et de dimensions spécifiques ou génériques. A l'aide de ce type de dimensions, nous présentons deux approches pour modéliser des structures multidimensionnelles. L'objectif final est de permettre l'extraction de connaissances géographiques par l'exploration des données détaillées associées à une époque et des études temporelles sur les données intégrées et comparatives. Le domaine de l'étude est la foresterie et le sujet étudié est l'histoire sylvicole de la forêt de Montmorency.

MOTS-CLÉS : *Modèle Multidimensionnel, Entrepôts de données géospatiales, Spatial On-Line Analytical Processing, Dimension Spatiale Géométrique, Intégration de Données, Foresterie*

ABSTRACT:

This paper presents some problems linked to the integration of data in a spatiotemporal data warehouse. In many cases, the specifications of the data sets have evolved over time and data sources have temporal, spatial and semantic heterogeneity. In order to explore and analyse spatio-temporal data sets in a SOLAP (Spatial on Line Analytical Processing) application, we use geometric spatial dimension and specific or generic thematic dimensions. Using these dimensions, we present two approaches to model multidimensional structures. The final objective is to support geographic knowledge discovery through data exploration of detailed data for an epoch and of integrated comparable data for time-variant studies. The field of study is the subject studied forestry and forestry is the history of the forest of Montmorency.

KEYWORDS: *Multidimensional Model, Geospatial Data Warehouses, Spatial On-Line Analytical Processing, Geometric Spatial Dimension, Data Integration, Forestry*

Liste des figures

Figure 1.1 : Plan de travail.....	05
Figure 2.1 : Cube de données multidimensionnel dans le domaine forestier.....	11
Figure2.2 : Exemple de schéma de dimension (a) et schéma des instances de la dimension (b).....	11
Figure 2.3. Architecture à trois niveaux d'un système d'entrepôt de données.....	14
Figure2.4. : Architecture d'un MOLAP.....	16
Figure 2.5 : a) Schéma en étoile b) Schéma en flocon c) Schéma en constellation.....	18
Figure2.6: Architecture d'un ROLAP.....	19
Figure 2.7: Architecture d'un HOLAP.....	19
Figure 2.8. Clients OLAP : a) Microstrategy b) Cognos c) Oracle.....	20
Figure 3.1. Dimension spatiale géométrique.....	26
Figure 3.2.types de mesure.....	28
Figure 4.1. Localisation de la forêt Montmorency.....	32
Figure 4.2. : Dimension année.....	35
Figure 4.3. : Découpage territorial de la forêt Montmorency.....	36
Figure 4.4: Opération d'overlay.....	38
Figure 4.5: Dimension Age.....	38
Figure 4.6: Dimension Essence.....	39
Figure4.7 : Dimension Hauteur.....	39
Figure 4.8 : Dimension Pente.....	40
Figure 4.9 : Dimension Perturbation.....	41
Figure 4.10: Modèle conceptuel de l'entrepôt de données de la forêt.....	42
Figure 5.1. Architecture de l'application.....	47
Figure 5.2.La base de données.....	49
Figure 5.3 : La carte de la forêt de Montmorency avec Mapinfo 7.0.....	51
Figure 5.4 : La page d'accueil de l'application.....	53
Figure 5.5 : La page principale de l'application.....	53
Figure 5.6 : La page de paramètre des données.....	54
Figure 5.7 : La page de paramètre des cartes.....	55
Figure 5.8 : La page d'interface d'analyse.....	55
Figure 5.9 : La page de résultat de l'analyse.....	56

Liste des Tableaux

<i>Tableau 2.1: Différences entre OLAP et OLTP.....</i>	10
<i>Tableau 2.2. MOLAP Vs ROLAP Vs HOLAP.....</i>	20
<i>Tableau 4.1: Valeurs détaillées de densité.....</i>	38
<i>Tableau 4.2: Valeurs détaillées de pente.....</i>	40
<i>Tableau 5.1 : Les paramètres de la projection.....</i>	51

Table des matières

Chapitre 1 : INTRODUCTION

1.1	Mise en contexte.....	02
1.2	Problématiques.....	03
1.3	Objective.....	04
1.5	Organisation du mémoire.....	04
1.6	Plan de recherche.....	05

Chapitre 2 : ENTREPOTS DE DONNEES ET ANALYSE EN LIGNE

2.1	Introduction.....	08
2.2	Concepts principaux des entrepôts de données et de l'analyse en ligne.....	09
2.2.1	System OLTP VS system OLAP.....	10
2.2.2	Dimensions et hiérarchies.....	10
2.2.3	Faits et mesures.....	11
2.2.4	Hypercube.....	12
2.2.5	Requêtes multidimensionnelles.....	13
2.3	Architecture des systèmes d'entrepôts de données.....	14
2.3.1	Entrepôt de données.....	14
2.3.2	Serveur OLAP.....	15
2.3.2.1	Fonctionnalités d'un serveur OLAP.....	15
2.3.2.2	Implémentation physique d'un serveur OLAP.....	16
2.3.3	Client OLAP.....	20
2.4	Conclusions.....	21

Chapitre 3 : L'OLAP SPATIAL

3.1	Introduction.....	23
3.2	De l'OLAP à l'OLAP Spatial.....	24
3.2.1	Les systèmes d'aide à la décision spatiale.....	24
3.2.2	Les avantages de l'OLAP Spatial.....	25
3.3	Concepts principaux de l'OLAP Spatial.....	25
3.3.1	Dimension spatiale.....	26
3.3.1.1	Les types de dimensions spatiales.....	26
3.3.2	Mesure spatiale.....	27
3.3.3	Opérateurs spatiaux de navigation.....	28
3.4	Les outils OLAP Spatial.....	28
3.5	Intégration des données spatiales dans un entrepôt.....	29
3.6	Conclusions.....	29

CHAPITRE 4 : CONCEPTION

4.1	Introduction.....	31
4.2	Présentation de la forêt de Montmorency.....	32
4.3	Les Buts du notre projet.....	32
4.4	Analyse préliminaire des données de la forêt Montmorency.....	33
4.4.1	Analyse des données descriptives.....	33
4.4.1.1	Par inventaire.....	33
4.4.1.2	Pour l'intégralité des inventaires.....	34

4.4.2	<i>Analyse des données spatiales</i>	34
4.4.2.1	<i>Par inventaire</i>	34
4.4.2.2	<i>Pour l'intégralité des inventaires</i>	34
4.5	<i>Analyse multidimensionnelle</i>	35
4.5.1	<i>Dimension Année</i>	35
4.5.2	<i>Dimension Découpage</i>	36
4.5.3	<i>Dimension Age</i>	37
4.5.4	<i>Dimension Densité</i>	38
4.5.5	<i>Dimension Essence</i>	39
4.5.6	<i>Dimension Hauteur</i>	39
4.5.7	<i>Dimension Pente</i>	40
4.5.8	<i>Dimension Perturbation</i>	41
4.5.9	<i>Le Fait</i>	41
4.5.10	<i>Homogénéité du découpage</i>	41
4.6	<i>Modèles conceptuels de l'entrepôt en étoile</i>	42
4.7	<i>Conception de l'outil d'analyse SOLAP</i>	43
4.8	<i>Conclusion</i>	44
CHAPITRE 5: REALISATION		
5.1	<i>Introduction</i>	46
5.2	<i>Paramètre de PC</i>	47
5.3	<i>Architecture de l'application</i>	47
5.3.1	<i>Entrepôt de données spatiales</i>	48
5.3.2	<i>Serveur OLAP</i>	49
5.3.3	<i>Client OLAP</i>	49
5.4	<i>Réalisation la carte</i>	50
5.4.1	<i>Logiciel utilisé</i>	50
5.4.2	<i>Construction le SIG</i>	51
5.5	<i>L'application finale</i>	52
5.7	<i>Conclusions</i>	57
CHAPITRE 6 : CONCLUSIONS GENERALE		
6.1	<i>conclusion</i>	59
6.2	<i>Perspectives</i>	60

Chapitre 01

Introduction générale

- 1.1. Mise en contexte
- 1.2. Problématique
- 1.3. Objectifs
- 1.4. Organisation de mémoire
- 1.5. Plan de recherche

1.1. Mise en contexte :

Les grands volumes de données sont difficiles à exploiter par les analystes/gestionnaires qui ont besoin d'information de nature «analytique ou décisionnelle».

Le système Géomatique décisionnelle est un mariage du BI (Business intelligence) et des SIG (S...) pour les données spatiales aussi appelée "le géo décisionnel" : "géo" (Terre) + "informatique décisionnelle". "Géo" apporte la référence spatiale avec la cartographie.

L'informatique décisionnelle appelée intelligence d'affaires (BI) apporte des technologies telles que :

- Les Entrepôts de données (data warehousing/datamarts)
- L'OLAP (serveurs MOLAP/ROLAP/HOLAP, clients)
- Le Data mining

Le processus de prise de décision exige de pouvoir faire des prévisions sur le comportement des phénomènes soumis à l'analyse ainsi que de pouvoir analyser les données sur une période de temps significative. Dans le domaine forestier, les outils d'analyse doivent permettre de questionner la base de données sur le comportement, entre différentes époques, sur des mesures telles que le volume de bois, par rapport à certaines essences, à certains peuplements forestiers, à certaines densités de couvert, etc. Les entrepôts de données apparaissent spécialement désignés pour offrir une aide à la décision puisqu'ils contiennent non seulement des données actuelles mais également des données provenant de plusieurs époques, et sont conçus pour permettre le processus d'analyse de ces données. Les outils SOLAP « sont une catégorie de logiciel spécialement conçus pour l'exploration rapide et facile des données multidimensionnelles composées de plusieurs niveaux d'agrégation ». La structure multidimensionnelle qui les caractérise est en opposition avec l'approche transactionnelle car c'est une représentation plus près de la réalité envisagée par l'utilisateur des données (Codd et al. 1993). De plus, cette dernière facilite la réponse aux requêtes (Kimball, 2002a).

Introduction générale

Dans un cube, les dimensions sont les thèmes d'analyse thématiques du modèle (Rivest et al, 2005) ; elles regroupent des données descriptives, spatiales ou temporelles (Bédard et al, 2001), par exemple les *catégories de produits*, la *géographie*, et le *temps*. Les dimensions sont définies par le schéma de la dimension, qui décrit les niveaux de la hiérarchie de la dimension, ainsi que par le schéma des instances contenant les membres de la dimension. Les faits sont les objets de l'analyse (ex : nombre de professeur, surface, total des ventes, etc.).

Dans ce mémoire nous nous intéressant au cas de la Forêt de Montmorency, nous y parleront des défis de gestions qu'elle rencontre et nous y détailleront le système que nous avons conçus pour l'aide à l'analyse de son historique sylvicole.

1.2. Problématique :

La Forêt de Montmorency est une Forêt polyvalente qui doit faire face a plusieurs défis qui des fois sont contradictoires. D'un côté l'aménagement durable et la préservation de la multitude d'espèces animales quelle abrite. De l'autre côté, elle propose une exploitation du bois régulière ainsi que des activités touristiques hivernales et estivales au grand public. A présent l'université Laval dispose d'une importante quantité de données sur l'historique sylvicole de la Forêt. Le besoin est d'un système qui supporte la décision spatio-temporelle qui sera capable de gérer des volumes gigantesques de données et offrant des analyses selon différents critères. Les SOLAP répondent bien à cela en plus d'offrir des affichages adaptés à la GéoDécision (cartes, tableaux croisés et diagrammes statistiques) et des temps de réponse record.

1.3. Objectifs:

L'objectif général de ce mémoire est le Développement d'un Système d'information D'aide à la décision pour La Forêt de Montmorency. Les objectifs spécifiques sont :

- La résolution du problème d'hétérogénéités des données spatiales
- La réalisation et le peuplement de l'entrepôt de données.
- La réalisation de l'interface utilisateur.

1.4. Organisation de mémoire :

Pour présenter notre travail, nous avons retenu pour ce mémoire une organisation en six chapitres :

Le premier chapitre est une introduction générale du sujet, nous y présentons également la problématique et les objectifs.

Le chapitre 2 introduit les principaux concepts des entrepôts de données et de l'analyse en ligne (en temps réel). Nous présentons les principes de la modélisation multidimensionnelle car nous nous appuyerons sur ces concepts pour la formalisation de notre modèle du GéoCube.

Le chapitre 2 présente également l'architecture classique des systèmes décisionnels.

Le chapitre 3 introduit les concepts principaux de l'OLAP Spatial : mesures, faits et dimensions spatiales.

Le quatrième chapitre décrit nos travaux relatifs à la conception de l'entrepôt. Cet espace de stockage de données sources nécessaires aux processus d'aide à la décision doit permettre de stocker les données évoluant dans le temps de manière détaillée ou archivée.

Le chapitre 5 qui représente La réalisation du SIAD sert aussi à finaliser la présentation de nos travaux. Spécialement, il permet d'explicitier les principales fonctionnalités des outils que nous avons développés.

Enfin dans le Chapitre 6 on retrouve une conclusion générale comme fermeture de ce travail.

1.5. Plan de recherche :

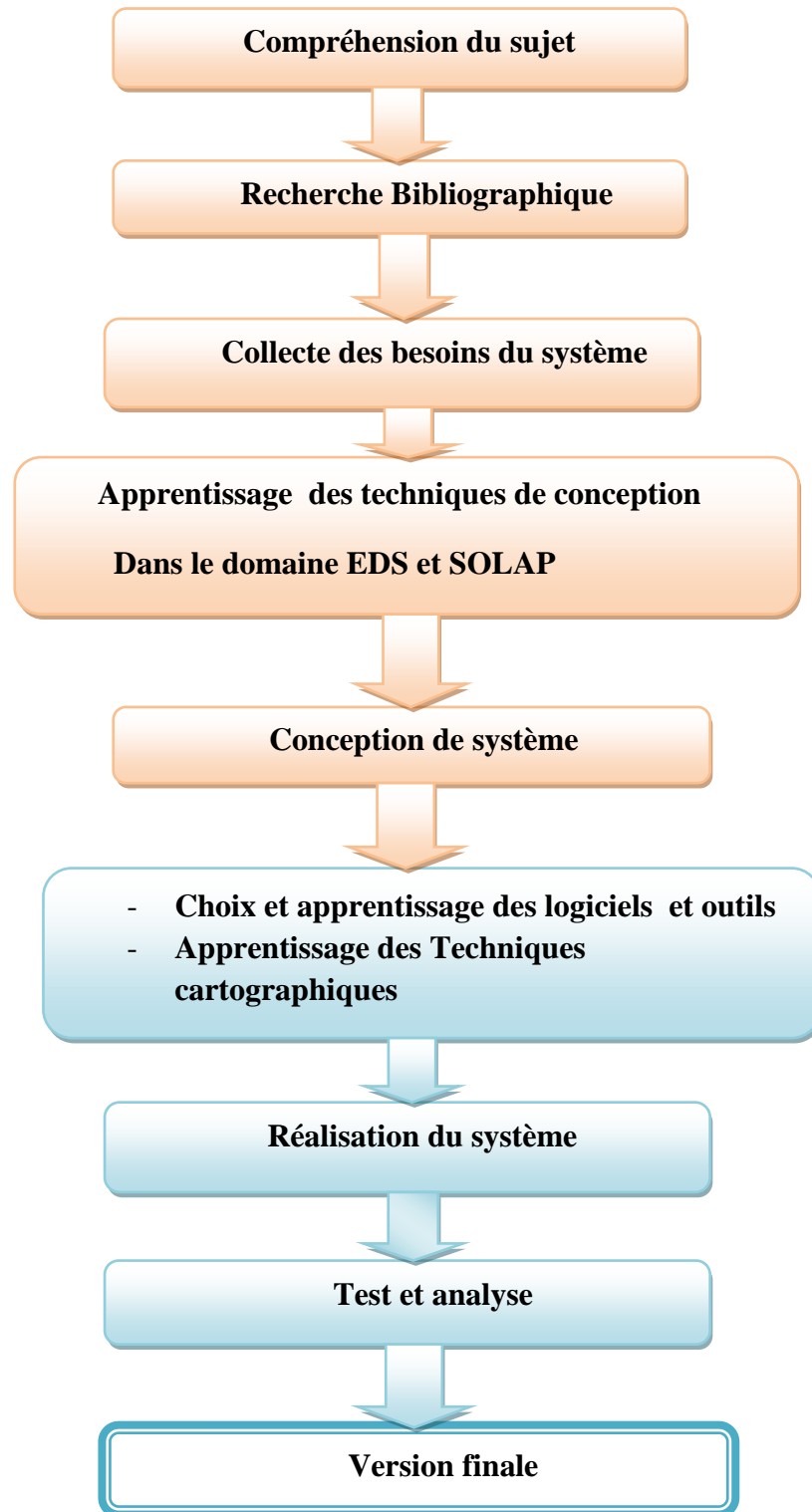


Figure 1.1 : Plan de travail

Introduction générale

La première étape consiste en premier lieu à rassembler les connaissances sur le data warehousing et de comprendre le domaine forestier.

La deuxième étape sert à faire une recherche approfondie sur les documents et les publications existantes afin de rassembler le plus d'idées possibles sur le géo décisionnel pour faciliter la suite du travail.

La troisième étape Collecte des besoins du système, cette étape est la plus difficile car elle comprend le contact avec les personnes (personnel de la direction concernée).

La Conception du système, consiste comme son nom l'indique à mettre en place les idées et les bases sur lesquelles va reposer le système, c'est aussi l'étape la plus délicate du travail car d'elle dépend le bon déroulement du projet et surtout le temps nécessaire pour le mener à bout car si elle est soigneusement effectuée la réalisation du projet sera plus efficace et les résultats de meilleure qualité.

L'étape de la réalisation du système est une implémentation des concepts mis au point dans l'étape précédente et qui va permettre de faire une première série de tests afin de mieux les adapter ou de les changer.

Lorsque les principes sont bien établis et validés, la version finale du logiciel peut alors être mise à la disposition des décideurs des services concernés.

Chapitre 02

Entrepôt de données et OLAP

- 2.1. Introduction:
- 2.2. Concepts principaux des entrepôts de données et de l'OLAP
 - 2.2.1. Système OLTP versus système OLAP
 - 2.2.2. Dimensions et hiérarchies
 - 2.2.3. Faits et mesures
 - 2.2.4. Hypercube
 - 2.2.5. Requêtes multidimensionnelles
- 2.3. Architecture des systèmes à entrepôts de données
 - 2.3.1. Entrepôt de données
 - 2.3.2. Serveur OLAP
 - 2.3.2.1. Fonctionnalités d'un serveur OLAP
 - 2.3.2.2. Implémentation physique d'un serveur OLAP
 - 2.3.3. Client OLAP
- 2.4. Conclusion

2.1. Introduction:

Le but principal des Systèmes de Gestion de Bases de Données (SGBD) est de stocker, d'organiser et de permettre d'interroger au moyen de requêtes de grands volumes d'informations opérationnelles (Elmasri et Navathe, 2003). Ces données cachent souvent des connaissances et des informations stratégiques, toutes deux utiles au processus décisionnel.

Les Systèmes d'Aide à la Décision (SAD) sont des systèmes d'information flexibles et interactifs qui aident les décideurs dans l'extraction d'informations utiles pour identifier et résoudre des problèmes et pour prendre des décisions (Alter, 1980). Cette connaissance est obtenue à partir de données brutes, de connaissances personnelles et de modèles analytiques. Les SAD présentent ces informations, provenant de différentes sources, dans un environnement unique, uniforme et familier à l'utilisateur. Ils combinent, uniformisent et synchronisent les bases de données, les modèles d'analyse et les techniques de visualisation, en permettant de comparer différents résultats et de concevoir et valider des hypothèses. Parmi les systèmes d'aide à la décision, les systèmes d'entrepôts de données sont probablement les plus utilisés dans le monde académique et industriel.

2.2. Concepts principaux des entrepôts de données et de l'OLAP :

Un entrepôt de données est « *une collection de données, intégrées, non volatiles et historisées pour la prise de décisions* » (Inmon, 1996).

Dans un entrepôt de données, les données sont organisées en fonction des exigences analytiques des utilisateurs. Les entrepôts de données sont peuplés en utilisant différentes sources de données hétérogènes pour lesquelles ils fournissent une vision unifiée et homogène. Contrairement aux données opérationnelles qui ont une durée de vie limitée, les entrepôts de données doivent permettre une analyse historique.

Les requêtes des bases de données opérationnelles s'effectuent sous forme de transactions qui lisent et écrivent un nombre réduit de lignes dans différentes tables liées par des liens référentiels. Ce type de requêtes est dit « On line Analytical Transactional Processing » (OLTP). Au contraire, le type de requêtes effectuées sur un entrepôt de données est appelé « On line Analytical Processing » (OLAP).

2.2.1. Système OLTP versus système OLAP :

Le tableau 2.1 résume ces différences entre OLTP et OLAP

	OLTP	OLAP
Utilisation	<i>Gestion des transactions</i>	<i>Aide à la Décision</i>
Conception	<i>Orientée applications</i>	<i>Orientée utilisateurs</i>
Fréquence	<i>Quotidienne</i>	<i>Sporadique</i>
Données	<i>Récentes, détaillées</i>	<i>Historiques, multidimensionnelles, agrégées</i>
Source	<i>BD unique</i>	<i>Plusieurs BD</i>
Utilisation	<i>Répétitif</i>	<i>Ad hoc</i>
Accès	<i>Programmes précompilés</i>	<i>Outils d'Analyse</i>
Nb lignes accédées	<i>Dizaines</i>	<i>Milliers</i>
Type d'utilisateur	<i>Opérateurs</i>	<i>Décideurs</i>
Nb utilisateurs	<i>Milliers</i>	<i>Centaines</i>
Performance	<i>Elevée</i>	<i>Basse</i>
Dimension BD	<i>Giga-octets</i>	<i>Téraoctets</i>

Tableau 2.1: Différences entre OLAP et OLTP

2.2.2. Dimensions et hiérarchies

Afin de permettre à l'utilisateur d'analyser facilement et intuitivement les données, les entrepôts de données se basent sur une structure multidimensionnelle, qui permet de représenter les données selon plusieurs axes d'analyse, appelés dimensions. Les dimensions peuvent représenter plusieurs thèmes d'intérêt pour l'utilisateur, par exemple l'âge, le lieu géographique, le temps, etc....

La structure multidimensionnelle peut être représentée par un cube, que nous illustrons par un exemple issu du domaine forestier à la figure 2.1. Un cube est composé d'éléments appelés cellules. Les cellules contiennent les valeurs d'un fait, habituellement appelées mesures. Les axes du cube correspondent aux dimensions et ils sont gradués par des membres. Dans la structure multidimensionnelle, les dimensions sont des hiérarchies et comportent donc un ensemble de niveaux liés par des relations de classification. La structure d'une dimension est représentée par le schéma de la dimension.

Entrepôt de données et OLAP

Le schéma de la dimension, lorsqu'il est instancié, forme le schéma des instances qui explicite tous les membres de la dimension (figure 2.2).

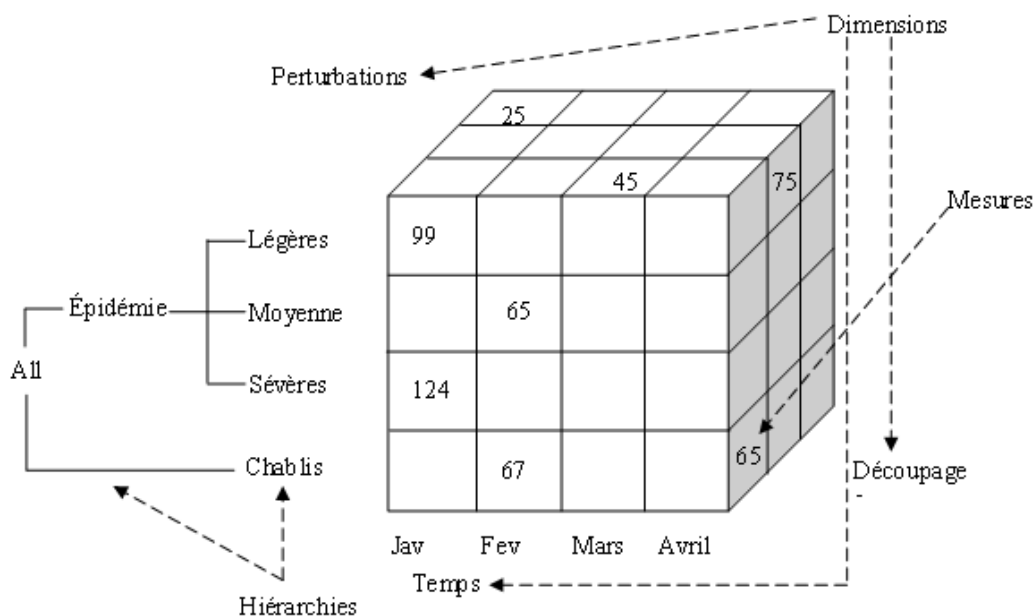


Figure 2.1 : Cube de données multidimensionnel dans le domaine forestier

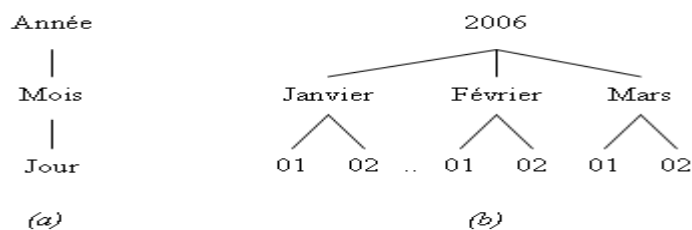


Figure 2.2 : Exemple de schéma de dimension (a) et schéma des instances de la dimension (b)

2.2.3. Faits et mesures

Un fait est un concept relevant du processus décisionnel et, typiquement, modélise un ensemble d'événements d'une organisation. Un fait est décrit par plusieurs mesures. Les mesures représentent usuellement des valeurs numériques qui fournissent une description quantitative du fait. Un fait est associé à une ou plusieurs combinaisons de membres des dimensions. Enfin, certaines mesures peuvent être calculées à partir d'autres mesures ou propriétés de membres. Elles sont appelées mesures dérivées (Blaschka et al. 1998).

Entrepôt de données et OLAP

A chaque combinaison des niveaux des dimensions correspond un niveau différent de détail des mesures. Dans les niveaux moins détaillés des dimensions les mesures sont agrégées en utilisant les fonctions d'agrégations. Les fonctions classiques pour agréger les mesures sont les mêmes que les opérations SQL "COUNT", "SUM", "MIN", "MAX" et "AVG".

Il est fondamental dans un modèle multidimensionnel de contrôler que le processus d'agrégation soit correct, afin de garantir une analyse multidimensionnelle exacte. Dans les systèmes OLAP, l'opérateur d'agrégation le plus utilisé est la somme. La précision de l'agrégation dépend de la sémantique de la mesure et de la structure de l'application multidimensionnelle. Par exemple, sommer des concentrations de pollution n'a pas de sens, en revanche le maximum est significatif. Il est possible de sommer la population des villes d'un pays, mais sommer la population d'une ville dans le temps est erroné, car les mêmes habitants seront comptés plusieurs fois. Cette problématique est connue dans l'OLAP sous le nom de problème d'« additivité » (Kimball, 1996). Une mesure est dite :

- *Additive* si l'on peut lui appliquer la somme sur toutes les dimensions.
- *Semi-additive* si la somme a du sens seulement sur certaines dimensions.
- *Non-additive* si elle n'est sommable sur aucune dimension.

Notons que la mesure peut être prise en compte plusieurs fois dans le processus d'agrégation, cela peut dépendre du type de hiérarchie, par exemple lorsque le modèle présente des hiérarchies non strictes ou multiples (Horner et al. 2004). Dans les bases de données statistiques, cette problématique est appelée « summarizability » (Lenz et Shoshani, 1997). La « summarizability » représente la situation dans laquelle le résultat d'une agrégation pourrait être calculé en utilisant les agrégations précédentes.

Une agrégation correcte implique :

- *Éviter de prendre en compte plusieurs fois la même mesure*
- *Respecter la sémantique de l'agrégation* : il s'agit de fournir un contrôle sur le type d'agrégation en prenant en compte la sémantique de la mesure.

2.2.4. Hypercube :

L'instance d'un modèle conceptuel multidimensionnel est un hypercube. Un hypercube contient dans les cellules les valeurs des mesures détaillées et les axes sont faits par les membres des niveaux les plus détaillés de différentes dimensions. Ensuite, ce cube

de base est décoré avec des cellules qui contiennent l'agrégation des valeurs de mesures pour chaque combinaison de membres des niveaux moins détaillés. Un exemple d'hypercube pour l'application de la Figure 2.1. Sur les axes du cube de base on trouve les membres des niveaux des dimensions et dans les cellules les valeurs des deux mesures.

2.2.5 Requêtes multidimensionnelles

Le processus décisionnel multidimensionnel consiste en l'exploration de l'hypercube, L'utilisateur parcourt les données de l'hypercube selon les différents axes d'analyses à la recherche d'informations utiles, dans un processus fortement interactif, itératif et constructif, qui comprend des étapes de formulation des hypothèses, expérimentation et analyse (Tang et al. 2003). Les utilisateurs interagissent itérativement avec le modèle multidimensionnel pour formuler, modifier et valider leurs hypothèses. Les chemins d'analyse sont imprédictibles, contrairement aux données qui sont définies lors de la conception de l'application. Chaque résultat d'analyse est la conséquence des résultats précédents. Chaque étape du processus d'analyse est représentée par une navigation dans l'hypercube, ou par une requête multidimensionnelle. Ces requêtes utilisent les opérateurs OLAP. Les opérateurs OLAP permettent d'explorer les données multidimensionnelles en utilisant les concepts de dimensions et hiérarchies.

Un panorama des opérateurs OLAP proposés dans la littérature est présenté par Rafanelli en (Rafanelli, 2003). Les plus communs sont :

- **Les opérateurs de forage :**
 - **Roll-up** permet de monter dans les hiérarchies des dimensions, et d'agréger les mesures.
 - **Drill-Down** est l'inverse du Roll-Up et permet de descendre dans une hiérarchie.

- **Les opérateurs de coupe :**
 - **Slice** utilise un prédicat défini sur les membres des dimensions pour couper une partie de l'hypercube limitant le champ d'analyse et permettant à l'utilisateur de se concentrer sur des aspects particuliers du phénomène. En utilisant la terminologie de l'algèbre relationnelle, l'opération de slice est l'équivalent de la sélection.
 - **Dice** réduit la dimensionnalité de l'hypercube en éliminant une dimension. Cette opération est équivalente à la projection de l'algèbre relationnelle.

Entrepôt de données et OLAP

○ *Drill-Accross* met en relation plusieurs hypercubes pour comparer leurs mesures. En effet dans un processus d'exploration et d'analyse, comparer plusieurs phénomènes est fondamental pour aboutir à une connaissance finale. Corréler plusieurs hypercubes pour avoir une vision unique des différentes mesures est donc nécessaire dans le processus d'analyse multidimensionnel.

L'opération de drill-accross fusionne plusieurs hypercubes en utilisant les axes d'analyse en commun.

2.3. Architecture des systèmes d'entrepôts de données

Les architectures des systèmes d'entrepôts de données sont classiquement des architectures à trois niveaux, comme montré en Figure 2.3, constituées par un entrepôt de données, un serveur OLAP et un client OLAP.

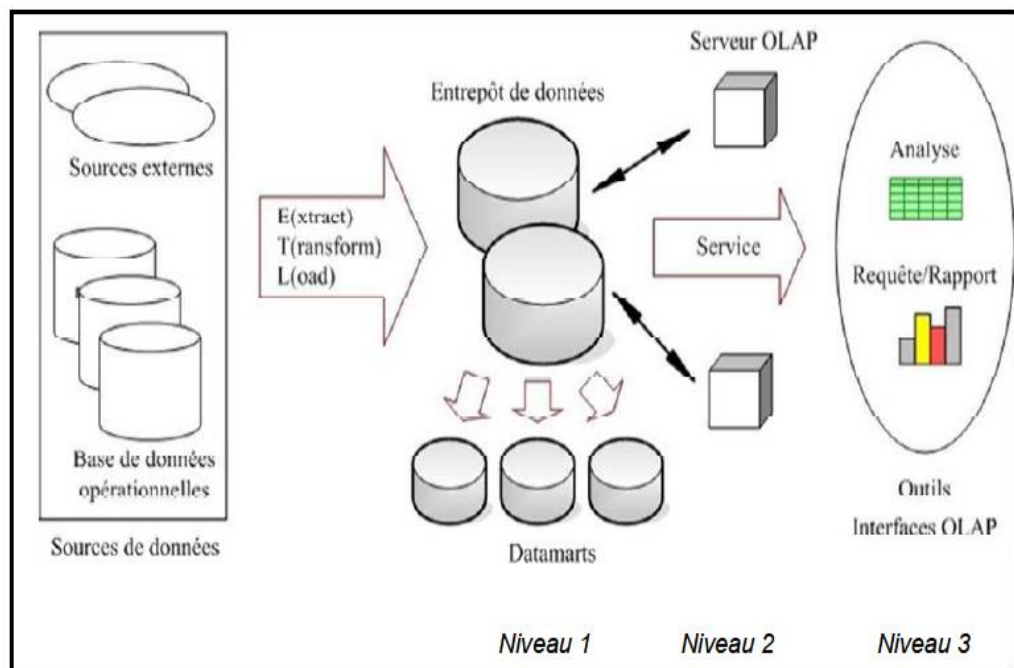


Figure 2.3. Architecture à trois niveaux d'un système d'entrepôt de données.

2.3.1 Niveau 1: L'entrepôt de données

Le premier niveau est un SGBD. Les données sont extraites à partir des bases de données transactionnelles, nettoyées et transformées avec des outils ETL (Extract Transform-Load ou en français extraction, transformation et alimentation), et intégrées dans l'entrepôt de données. Le SGBD contient aussi un ensemble de métadonnées

Entrepôt de données et OLAP

concernant les sources de données, les mécanismes d'accès, les procédures de nettoyage et d'alimentation, les utilisateurs, etc.

2.3.2 Niveau2 : Le serveur OLAP

Le deuxième niveau de l'architecture est un serveur OLAP, par exemple Mondrian, DB2 OLAP Server, Oracle OLAP, Microsoft Analysis Services, etc. Le serveur OLAP permet d'effectuer une analyse de données conforme au paradigme multidimensionnel, avec des temps de réponse optimisés.

2.3.2.1 Fonctionnalités d'un serveur OLAP

Un serveur OLAP fournit aux utilisateurs une vue multidimensionnelle des données qui peuvent être analysées grâce aux opérateurs OLAP (Roll-Up, Drill-Down, etc.). De plus, le serveur OLAP permet de gérer les données agrégées et les données détaillées de façon transparente. Les données d'un hypercube peuvent être représentées par un treillis hiérarchique de cuboïdes (Harinarayan et al, 1996). Le cuboïde le plus détaillé contient les valeurs des mesures détaillées. Les autres cuboïdes représentent l'agrégation de mesures détaillées. Si l'application présente n dimensions et chaque dimension L_i niveaux,

Pour optimiser l'accès aux données agrégées, le serveur OLAP pré-calculé un sous ensemble de cuboïdes. Lors d'une requête multidimensionnelle, le serveur OLAP cherche dans un premier temps si ces données ont déjà été calculées et si oui, il renvoie le résultat ; autrement il procède au calcul des cuboïdes nécessaires. En plus des données agrégées et détaillées, un serveur OLAP gère aussi des métadonnées qui, de façon simple et transparente, masquent l'organisation multidimensionnelle complexe de l'entrepôt de données.

Les entrepôts de données contiennent souvent une grande quantité de données détaillée et agrégées, mais grâce à l'utilisation des techniques particulières d'indexation et de politiques de gestion des données pré-agrégées, le serveur permet le passage à l'échelle des données de l'application multidimensionnelle.

Un serveur OLAP ne se limite pas à fournir les opérateurs d'analyse multidimensionnelle, mais il met à disposition de l'utilisateur un puissant moteur de calcul qui présente des fonctions mathématiques et des procédures multidimensionnelles.

Entrepôt de données et OLAP

Ces dernières permettent par exemple de naviguer dans les hiérarchies et de définir des mesures dérivées complexes.

Enfin, puisqu'un entrepôt de données est toujours analysé par plusieurs décideurs, un serveur doit gérer les accès concurrents, et garantir l'intégrité et la sécurité des informations.

2.3.2.2. Implémentation physique d'un serveur OLAP :

Les approches principales pour l'implémentation de serveurs OLAP sont Multidimensionnels OLAP (MOLAP), Relational OLAP (ROLAP) et Hybrid OLAP(HOLAP).

▪ MOLAP (OLAP Multidimensionnel)

Les serveurs MOLAP extraient les données de l'entrepôt de données et les mémorisent en utilisant des structures de données particulières et ils appliquent des techniques d'indexation et de hachage pour localiser les données lors de l'exécution des requêtes multidimensionnelles.

- Les données détaillées de base ainsi que les données agrégées de l'entrepôt sont stockées dans une base de données multidimensionnelle (souvent appelée cube ou hypercube)

- Une base de données multidimensionnelle utilise une structure propriétaire au logiciel utilisé (matrice à N dimensions)

- Le serveur MOLAP extrait les données de l'hypercube et les présente directement au module client

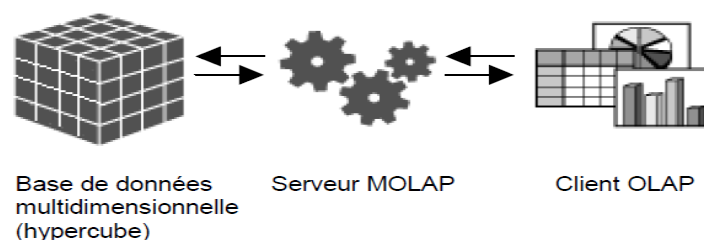


Figure2.4. : Architecture d'un MOLAP

▪ ROLAP (OLAP Relationnel)

Les serveurs ROLAP (Kimball, 1996) utilisent la technologie des bases de données relationnelles pour mémoriser les données. Pour obtenir des performances acceptables,

Entrepôt de données et OLAP

ces systèmes utilisent des structures d'indexation particulières comme l'index bitmap, et les vues matérialisées (Winter, 1998). Une vue matérialisée est la traduction relationnelle d'un cuboïde qui est pré-calculé et stocké dans l'entrepôt de données. La modélisation logique d'une base de données multidimensionnelle selon l'approche ROLAP ne systématise pas l'utilisation de la 3^{ème} forme normale, contrairement aux systèmes OLTP. Le modèle logique le plus utilisé est le schéma en étoile (Kimball, 1996). Le schéma en étoile (Figure 2.5a) se constitue d'une table de faits et de tables de dimensions. Chaque entrée dans la table de faits représente un fait ou une cellule du cuboïde de base, et elle est liée, à travers des clés étrangères, à des dimensions. Les tables de dimensions sont des tables dénormalisées qui contiennent des attributs descriptifs et des attributs qui représentent les hiérarchies des dimensions. Par exemple, un produit est caractérisé par un code, par un nom, qui représente un attribut descriptif, et un type. Le type est utilisé pour regrouper les produits en catégories différentes. La dénormalisation améliore les performances des requêtes car elle réduit le nombre de jointures. Une variante du schéma en étoile est le schéma en flocon (Figure 2.5b). Il présente des dimensions partiellement ou totalement normalisées. La normalisation est utilisée pour réduire la redondance, surtout dans le cas où chaque niveau hiérarchique présente différents attributs et/ou quand la dimension est constituée de nombreux niveaux. Enfin, classiquement, un entrepôt de données est formé par différents hypercubes liés entre eux par des dimensions. La structure logique utilisée dans ce cas est le schéma en constellation. Il s'agit de plusieurs tables des faits qui partagent des tables des dimensions et qui peuvent être vu comme une collection d'étoiles (schéma en galaxie ou constellation de faits) (Figure 2.5c). Cette représentation logique permet d'utiliser l'opération de drill-across.

Entrepôt de données et OLAP

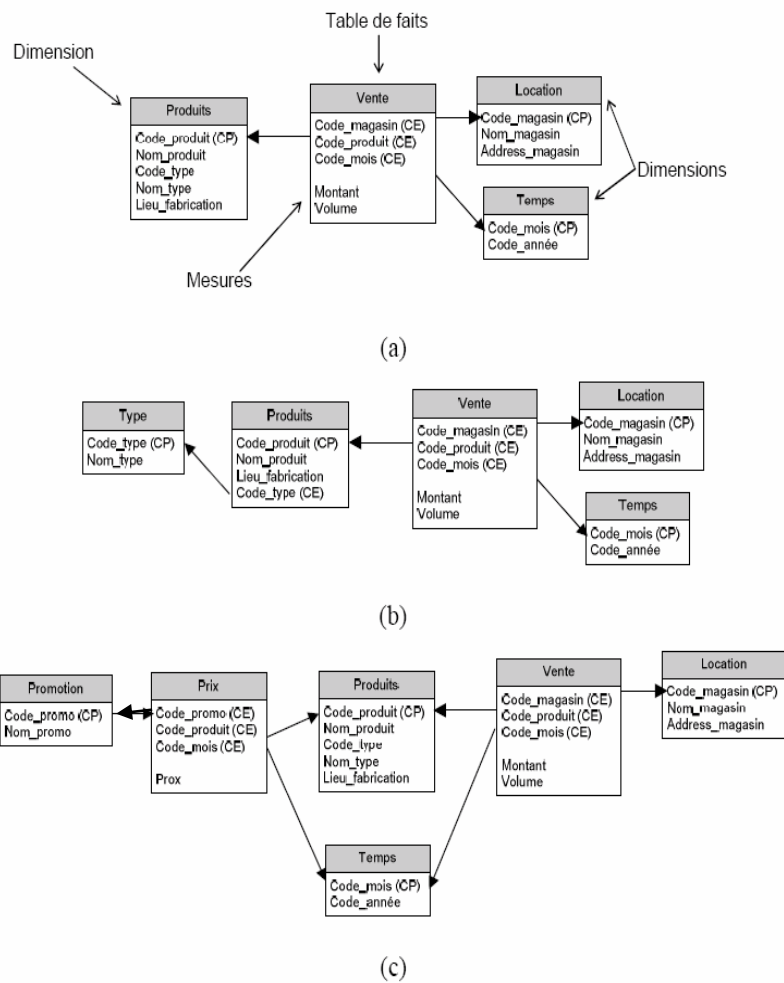


Figure 2.5 : a) Schéma en étoile b) Schéma en flocon c) Schéma en constellation.

- Les données détaillées de base ainsi que les données agrégées de l'entrepôt sont stockées sous forme de tables dans une base de données relationnelle
- La base de données relationnelle doit être structurée selon un modèle particulier (étoile, flocon, ...)

Entrepôt de données et OLAP

- Le serveur extrait les données par des requêtes SQL et interprète les données selon une **vue multidimensionnelle** avant de les présenter au module **client**

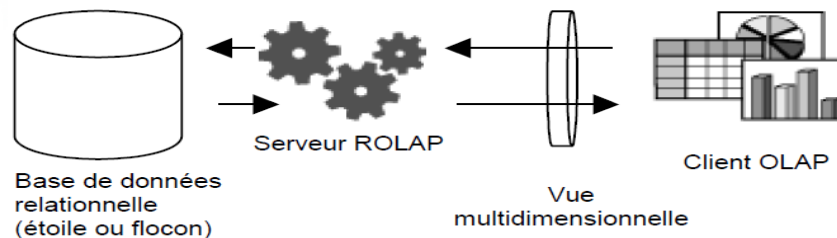


Figure2.6: Architecture d'un ROLAP.

▪ HOLAP (OLAP Hybride)

- Architecture qui consiste en un croisement des architectures MOLAP et ROLAP
- Les données détaillées de base de l'entrepôt sont stockées dans une base de données relationnelle et les données agrégées sont stockées dans une base de données multidimensionnelle
- Le serveur HOLAP accède aux deux bases de données et les présente au module client, selon une vue multidimensionnelle dans le cas des données de la BD relationnelle

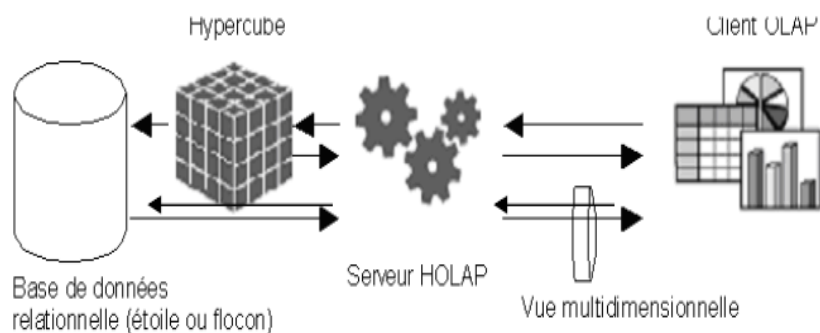


Figure 2.7: Architecture d'un HOLAP.

Différences entre MOLAP, ROLAP et HOLAP

Critère de comparaison	ROLAP	MOLAP	HOLAP
Stockage des données de base (détaillées)	BD relationnelle	BD multidimensionnelle	BD relationnelle
Stockage des agrégations	BD relationnelle	BD multidimensionnelle	BD multidimensionnelle
Performance des requêtes (habituellement)	Performant	Le plus performant (jusqu'à un certain volume)	Très performant

Tableau 2.2. MOLAP Vs ROLAP Vs HOLAP

2.3.3 Niveau 3 : Client OLAP

Le dernier niveau est un client OLAP, par exemple Microstrategy (Figure 2.8a), Cognos (Figure 2.8b), JPivot, Polaris (Stolte, et al, 2002), Advizor (Eick, 2000), etc., qui offrent une interface utilisateur avec des outils de reporting, d'analyse interactive, et parfois de fouille de données. Son rôle est de rendre l'information multidimensionnelle « visible », en d'autres termes, de permettre de découvrir des connaissances grâce à la seule visualisation et interaction avec les données. Parfois, le client et le serveur OLAP sont intégrés. Par exemple, Oracle (OracleBI, 2007) offre des outils d'analyse (Figure 2.8c) et des systèmes spécialisés pour une analyse facile des données issues des systèmes Oracle.

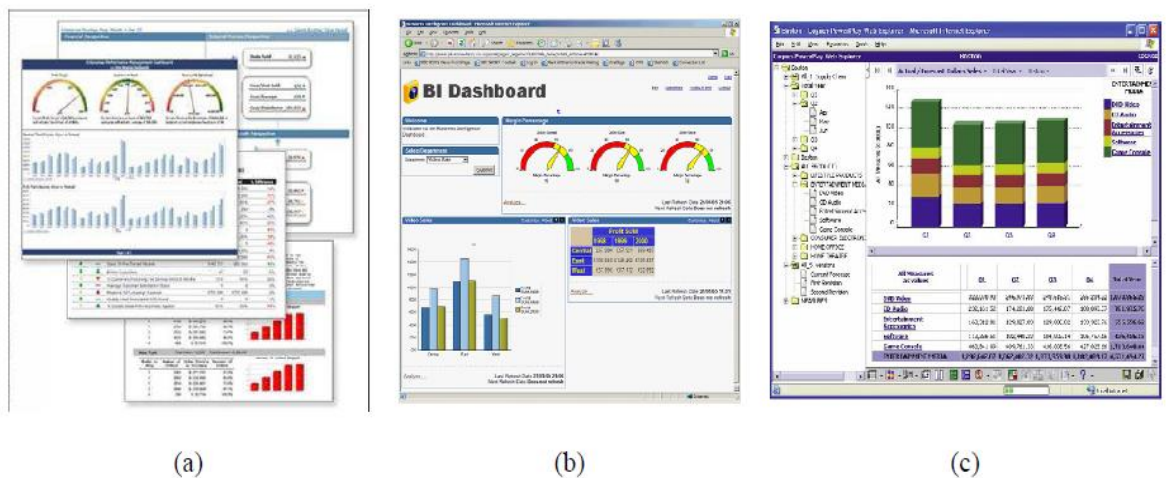


Figure 2.8. Clients OLAP : a) Microstrategy b) Cognos c) Oracle.

2.4 Conclusions

Dans ce chapitre, nous avons présenté un panorama des concepts principaux des systèmes d'entrepôts de données et OLAP. Nous avons décrit les concepts à la base de l'analyse multidimensionnelle, structures de données et opérateurs, et nous avons détaillé les trois niveaux de l'architecture classique des systèmes d'entrepôts de données. Nous avons également présenté les caractéristiques fondamentales des modèles formels pour les bases de données multidimensionnelles, qui sont centrées autour d'un ensemble bien défini des caractéristiques fondamentales.

Chapitre 03

L'OLAP spatial

- 3.1. Introduction
- 3.2. De l'OLAP à l'OLAP Spatial
 - 3.2.1. Les systèmes d'aide à la décision spatiale
 - 3.2.2. Les avantages de l'OLAP Spatial
- 3.3. Concepts principaux de l'OLAP Spatial
 - 3.3.1. Dimension spatiale
 - 3.3.1.1. Les types de dimensions spatiales
 - 3.3.2. Mesure spatiale
 - 3.3.3. Opérateurs spatiaux de navigation
- 3.4. Les outils OLAP Spatial
- 3.5. Intégration des données spatiales dans un entrepôt
- 3.5 Conclusions

3.1. Introduction :

Le terme OLAP Spatial (SOLAP) identifie un type de système d'aide à la décision spatiale qui intègre les techniques d'analyse de l'OLAP et des systèmes d'information géographique. Le SOLAP augmente les capacités d'analyse des systèmes OLAP classiques et il implique une reformulation des concepts des entrepôts de données et de l'analyse en ligne d'un point de vue formel et d'implémentation.

3.2. De l'OLAP à l'OLAP Spatial :

Dans cette section nous introduisons les caractéristiques principales d'un système d'aide à la décision spatiale et les avantages liés à l'introduction de la composante cartographique dans l'OLAP.

3.2.1. Les systèmes d'aide à la décision spatiale :

Dans le contexte de l'exploration et d'analyse interactive de données spatiales, qui représentent 80% des données transactionnelles (Franklin, 1992), les cartes ne sont pas seulement de simples outils de visualisation, mais elles sont des instruments actifs qui stimulent l'utilisateur dans son processus mental d'analyse (MacEachren et Kraak, 2001). Les Systèmes d'Aide à la Décision (SAD), et les systèmes OLAP en particulier, ne présentent aucun instrument pour la gestion des données spatiales. Par conséquent, les outils OLAP, en prenant peu en compte la composante spatiale et son pouvoir d'expression et d'analyse, manquent d'un instrument fondamental d'analyse et d'exploration qui peut aider l'utilisateur dans le processus décisionnel (Caron, 1998). A la base des SIG existe une technologie OLTP, contrairement aux systèmes d'entrepôts de données qui s'appuient sur une technologie OLAP (cf chap 2). Les SIG ne peuvent pas être considérés comme de véritables SAD, car même s'ils incluent des fonctionnalités avancées d'analyse, ils manquent d'une interface simple et intuitive pour visualiser et requêter les données, les temps d'analyse sont longs et ils ne sont pas conçus pour fournir une vision agrégée des données (Keenan, 1996). Un nouveau type de SAD a alors été conçu : les Systèmes d'Aide à la Décision Spatiale (SADS).

Un SADS étend les fonctionnalités des systèmes d'information géographique (gestion de la donnée spatiale, affichage cartographique et graphique, outils d'analyse spatiale), en fournissant des modèles d'analyse, des interfaces simples et flexibles, et gérant des structures complexes de données spatiales (Armstrong et Denshman, 1990), (Goodchild et Denshman, 1990). Ils aident les décideurs à résoudre des problèmes et prendre des décisions dans les applications où la composante spatiale joue un rôle fondamental comme le géomarketing, l'environnement, etc. Différents types de SADS ont été développés, comme par exemple les systèmes « Exploratory Spatial Data Analysis » (Andrienko et al. 2003), qui intègrent des techniques d'interaction et de visualisation cartographique, et des méthodes d'analyse spatiales, ou les systèmes « Spatial Data Mining » (Compieta et al.

2007) qui fournissent des techniques de fouille de données spatiales et visualisation cartographique. Chacun s'adresse à des problématiques décisionnelles spatiales particulières. Parmi ces solutions, on trouve aussi le SOLAP qui vise à intégrer la donnée spatiale dans l'OLAP.

3.2.2. Les avantages de l'OLAP Spatial :

Le Spatial OLAP a été défini par Yvan Bédard comme « *une plateforme visuelle conçue spécialement pour supporter une analyse spatio-temporelle rapide et efficace à travers une approche multidimensionnelle qui comprend des niveaux d'agrégation cartographiques, graphiques et tabulaires* » (Bédard, 1997). La visualisation des mesures sur une carte permet de comprendre la distribution géographique d'un phénomène qui, souvent, peut être différente de l'espace géographique identifié par la structure hiérarchique définie par la dimension géographique. De plus, l'affichage cartographique révèle des informations spatiales (relations spatiales et informations métriques) qu'une simple étiquette textuelle ou un affichage graphique n'aurait jamais montrées. La visualisation cartographique dans un contexte multidimensionnel permet aussi de mettre en relation les différents phénomènes spatiaux par rapport aux axes d'analyse alphanumériques et de comparer ces phénomènes à diverses granularités géographiques.

La composante cartographique dans l'OLAP représente un instrument de visualisation et surtout d'analyse, qui permet à l'utilisateur de voir et comprendre les données spatio_multidimensionnelles, et elle constitue une interface vers l'entrepôt de données spatiale (Bédard et al, 2005). Utiliser une représentation cartographique des membres pour accéder aux opérations de navigation multidimensionnelle permet à l'utilisateur de se concentrer sur les aspects purement géographiques pendant le processus d'exploration et d'analyse sans qu'il ait besoin d'un modèle mental qui trace dans l'espace géographique les membres des dimensions.

3.3. Concepts principaux de l'OLAP Spatial :

Comme c'est le cas pour toute technologie et norme géospatiale moderne, les concepts rapportés ici reposent sur les concepts informatiques standards et apportent les extensions nécessaires pour effectuer de nouvelles fonctions ou pour améliorer l'exécution des fonctions existantes. Ainsi, en comparaison au modèle multidimensionnel conventionnel, le modèle multidimensionnel spatial comprend aussi des faits et des

dimensions non-spatiales. Par contre, comme nous le décrivons ci-après, il existe aussi des dimensions spatiales de différents types, des mesures spatiales et des opérations spatiales. Sans aller dans les détails, nous en présentons les principaux concepts dans les sous-sections qui suivent.

3.3.1. Dimension spatiale :

Le terme de dimension spatiale désigne l'introduction de l'information spatiale dans une application décisionnelle en tant qu'axe d'analyse. Diverses définitions de dimension spatiale existent dans la littérature ou elles sont.

3.3.1.1. Les type de dimensions spatiales :

(Bédard et al. 2001), (Rivest et al. 2003) définissent trois types de dimensions spatiales:

- « *non géométrique* » si les membres sont localisés seulement par un label textuel, comme une adresse ou un code postal.

- « *géométrique* » si tous les membres présentent une composante spatiale (point, ligne, polygone, etc.)

- « *mixte* » s'il y a des niveaux avec la composante spatiale et des niveaux purement textuels.

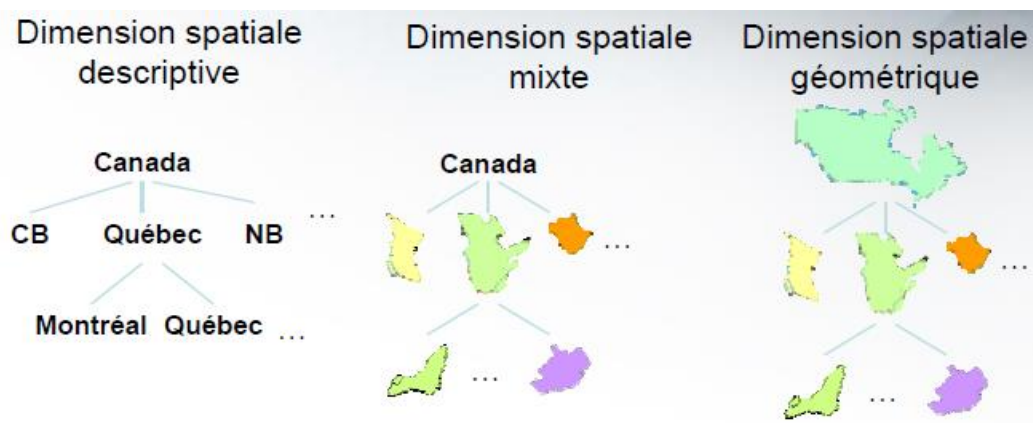


Figure 3.1. Dimension spatiale géométrique

Dans (Malinowski et Zimányi, 2005), les auteurs introduisent le concept de dimension spatiale comme un ensemble de hiérarchies spatiales. Une hiérarchie est spatiale s'il y a au moins un niveau qui contient la composante spatiale. Celui-ci est dit niveau spatial. De plus, entre les membres de deux niveaux spatiaux doit exister une relation topologique

d'inclusion ou d'intersection. Une hiérarchie spatiale peut être totalement spatiale si tous les niveaux sont spatiaux, partiellement spatiale s'il y a au moins un niveau non spatial. De leur côté, Fidalgo et al. (Fidalgo et al. 2004) définissent deux types de dimensions : géographique et hybride. Une dimension géographique contient seulement des niveaux spatiaux. Une dimension est hybride si elle présente à la fois des niveaux spatiaux et alphanumériques. De plus, en se concentrant sur la présence ou non des attributs descriptifs dans un même niveau, ils introduisent une sous-classification de ces dimensions : primitive si seuls les attributs spatiaux sont présents, et composée si des attributs alphanumériques sont aussi inclus.

3.3.2. Mesure spatiale :

Dans un contexte multidimensionnel spatial, il n'y a pas que les dimensions qui possèdent une composante géométrique, mais aussi les mesures. Par conséquent, en plus des mesures conventionnelles supportées dans les systèmes OLAP, il existe les mesures spatiales [Rivest *et al*, 2001].

Le pointeur spatial est le type de mesure spatiale le plus connu [Han et al 1998]. C'est la méthode utilisée par les technologies SIG pour gérer la composante géométrique des objets spatiaux. Il s'agit d'un ensemble de pointeurs (stockés dans le cube de données) vers la géométrie d'un objet spatial stockée dans une autre structure que la structure multidimensionnelle.

Le second type de mesure spatiale est la transposition au monde spatial de la mesure conventionnelle de l'OLAP. Elle permet de dériver des valeurs à l'aide d'un opérateur métrique ou topologique d'analyse spatiale dont le résultat sera ensuite stocké dans le cube de données (ex. surface d'un objet, distance minimale avec l'objet le plus proche, cumul de longueurs sur un réseau).

Finalement, la dernière mesure spatiale consiste à générer des données géométriques sous la forme d'un ou plusieurs objets spatiaux obtenus par la combinaison de dimensions spatiales géométriques (ou mixtes en utilisant les niveaux où les membres qui possèdent une géométrie). Il s'agit d'un ensemble de coordonnées obtenu à partir des opérateurs d'analyses spatiales d'un SIG, par exemple les coordonnées d'un point, ligne ou polygone résultant de l'intersection spatiale des membres de plusieurs dimensions. Ainsi en est-il des

polygones résultant de l'intersection des polygones délimitant les membres des dimensions spatiales frontières politiques et bassins versants.

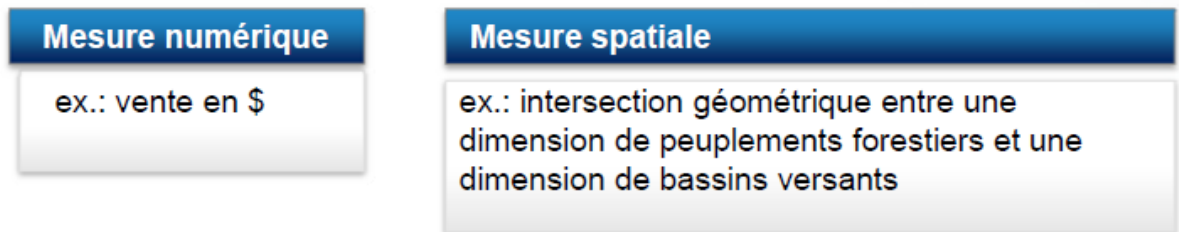


Figure 3.2. types de mesure

3.3.3. Opérateurs spatiaux de navigation :

Les outils SOLAP possèdent des opérateurs de navigation pour explorer via la carte l'ensemble des données spatiales. Les opérateurs spatiaux de navigation proposent différents forages, dont le forage spatial, le remontage spatial et le forage latéral spatial. L'opérateur de forage spatial permet à l'utilisateur de naviguer d'un niveau général à un niveau plus détaillé à l'intérieur d'une dimension spatiale géométrique (ex. cartographier les régions sous-jacentes composant un pays). Une opération de remontage permet la navigation inverse, c'est-à-dire de remonter d'un niveau détaillé des données vers un niveau le plus général (ex. cartographier les données nationales sus-jacentes à une région). Finalement, un opérateur de forage latéral permet de visualiser les différents membres du même niveau de détail d'une dimension spatiale (ex. cartographier pour mieux comparer les mesures de la région sud par rapport à celles de la région nord). Ces opérateurs sont utilisés directement sur la carte.

Les opérateurs spatiaux de navigation peuvent s'appliquer sur un objet individuel (ex. visualiser les régions composant l'objet Canada) ou s'appliquer à l'ensemble des objets d'un niveau de détail (ex. visualiser l'ensemble des régions composant le niveau Pays).

3.4. Les outils OLAP Spatial :

Un outil SOLAP repose sur l'intégration des fonctionnalités SIG et OLAP (Kouba, et al, 2000), (Rivest et al. 2005). La composante cartographique est utilisée pour visualiser les membres de dimensions et/ou les mesures avec une composante spatiale, pour représenter les mesures alphanumériques, grâce à des cartes thématiques, et pour accéder aux opérations de navigation multidimensionnelle.

3.4 .Intégration des données spatiales dans un entrepôt :

Un entrepôt de données spatiales s'approvisionne en données à partir d'autres sources, ces dernières contiennent des données qui peuvent différer dans le format, dans le système de référence, dans l'unité de mesure etc. Pour pouvoir stocker ces données dans une même structure elles doivent avoir les mêmes propriétés, pour cela on procède à une intégration avec des outils spécialisés dans l'intégration de données spatiales comme FME.

3.5. Conclusions :

L'introduction de l'information spatiale dans l'analyse multidimensionnelle implique une reformulation des concepts des entrepôts de données et de l'analyse en ligne.

L'intégration de la composante spatiale en tant qu'axe ou sujet de l'analyse a mené différents auteurs à la définition respectivement de dimension spatiale et mesure spatiale.

Chapitre 04

Conception

- 4.1. Introduction
- 4.2. Présentation de la forêt de Montmorency
- 4.3. Les Buts du notre projet
- 4.4. Analyse préliminaire des données de la forêt Montmorency
 - 4.4.1. Analyse des données descriptives
 - 4.4.1.1. Par inventaire
 - 4.4.1.2. Pour l'intégralité des inventaires
 - 4.4.2. Analyse des données spatiales
 - 4.4.2.1. Par inventaire
 - 4.4.2.2. Pour l'intégralité des inventaires
- 4.5. Analyse multidimensionnelle
 - 4.5.1. Dimension Année
 - 4.5.2. Dimension Découpage
 - 4.5.3. Dimension Age
 - 4.5.4. Dimension Densité
 - 4.5.5. Dimension Essence
 - 4.5.6. Dimension Hauteur
 - 4.5.7. Dimension Pente
 - 4.5.8. Dimension Perturbation
 - 4.5.9. Le Fait
 - 4.5.10. Homogénéité du découpage
- 4.6. Modèles conceptuels de l'entrepôt en étoile
- 4.7. Conception de l'outil d'analyse SOLAP
- 4.8. Conclusion

4.1. Introduction :

La technologie des entrepôts de données, les bases de données multidimensionnelles et les outils d'exploration et d'analyse dédiés font émerger de nouvelles solutions pour le recueil et l'utilisation décisionnelle des données géospatiales. L'utilisation des références spatiales, ou références géographiques, lorsqu'elles sont disponibles, devient même une source d'information supplémentaire pour comprendre et analyser les données. La représentation cartographique, parce qu'elle est synthétique, immédiate et facile à interpréter s'intègre parfaitement bien à un environnement décisionnel. Cependant, la mise en place d'un entrepôt de données et l'exploitation de ces données géospatiales doivent tenir compte des spécificités liées à ce type de données et aux besoins relatifs des utilisateurs afin de tirer le meilleur parti de la richesse des informations contenues dans ces sources de données. En effet, classiquement, l'information géospatiale est organisée selon des attributs descriptifs et géométriques. Un entrepôt de données géospatiales a donc pour vocation de collecter, d'intégrer et d'historiser ce type de données afin de les rendre accessibles aux décideurs. Lorsque les données recueillies portent sur la représentation d'un phénomène sur un territoire à différents temps d'observation, les problèmes d'intégration des données sont à la fois sémantiques, temporels et spatiaux surtout si les différentes sources de données n'ont pas été acquises dans l'objectif d'alimenter un entrepôt. En particulier, les problèmes d'intégration spatiale de données cartographiques hétérogènes nécessitent de nouvelles solutions pour minimiser les impacts négatifs souvent invisibles lors de la seule utilisation de noms géographiques comme référence dans la dimension géographique d'un hypercube (approche habituelle).

Ce chapitre nous allons présenter quelques problèmes et solutions relatifs à l'exploitation de données géospatiales sur plusieurs décennies, particulièrement dans un contexte où les spécifications et les techniques d'acquisition évoluent au cours des années.

4.2. Présentation de la forêt de Montmorency :

La forêt Montmorency est une forêt expérimentale de 6 664 ha située à environ 70 km au nord de la ville de Québec et gérée par la Faculté de foresterie et de géomatique de l'Université Laval. L'Université Laval a reçu ce territoire du Gouvernement du Québec en 1964 par bail emphytéotique de 99 ans. En échange, elle s'est engagée à y développer la recherche et l'enseignement dans divers domaines des sciences naturelles.



Figure 4.1. Localisation de la forêt Montmorency

Localisation :

Cordonnées : (47° 19' 20" Nord, 71° 08' 50" Ouest)

Pays : canada

Province: Québec

Géographie : Superficie=6 664 ha

4.3. Les Buts de notre projet :

La forêt est une ressource naturelle importante au Québec (Canada) de par son étendue (757 900 km²) et le nombre d'emplois générés (plus de 170 000 emplois). Il est donc normal que ce secteur d'activité s'intéresse aux nouvelles technologies pour améliorer la connaissance des évolutions des zones forestières afin de gérer efficacement cette ressource. Nous avons étudié les

apports de la technologie des entrepôts de données géospatiales à ce domaine par l'intégration de plusieurs inventaires forestiers effectués sur la même portion de territoire.

Effectuer un inventaire forestier consiste à partitionner la surface de la forêt étudiée en zones (appelées peuplements) qui présentent des caractéristiques forestières homogènes (essence, âge, densité, hauteur, etc.). Les données géométriques fournissent l'information sur la position et la forme des peuplements.

Les structures de données géométriques sont de type vectoriel. Les données descriptives fournissent des informations qualitatives ou quantitatives sur les caractéristiques des peuplements. Les mesures effectuées sur le peuplement sont par exemple la surface occupée ou le volume de bois. A chaque nouvel inventaire correspond une nouvelle carte forestière. Pour notre expérimentation, nous avons travaillé avec les inventaires de la forêt Montmorency des années 2000 à 2010.

4.4. Analyse préliminaire des données de la forêt Montmorency :

L'analyse préliminaire a permis de prendre connaissance des données sources de la forêt Montmorency, d'analyser la cohérence de ces données et d'évaluer leurs potentiels d'analyse au sein d'un entrepôt de données. Les données ont été étudiées sous deux aspects, spatiaux et sémantiques.

4.4.1. Analyse des données descriptives :

4.4.1.1. Par inventaire :

Chaque inventaire est composé d'une série d'attributs liés à une primitive spatiale sur le territoire. Le nombre d'attributs peut varier selon l'inventaire.

La première étape est de définir clairement chaque attribut de chaque inventaire pour en permettre ultérieurement la comparaison inter-inventaires. La deuxième et dernière étape consiste à déterminer les domaines de valeurs propres à chaque attribut. Il s'agit alors d'établir les contraintes d'intégrité intra-champ, inter-champ et inter-objet. Celles-ci ont permis de détecter des valeurs manquantes, des codages erronés et des combinaisons de valeur invalide. Il peut s'agir, à titre d'exemples, de valeurs d'essence manquantes, de code d'essence inconnu ou des combinaisons de valeur d'âge et de hauteur impossibles. A ce moment, les anomalies furent uniquement détectées et notées. Il faut en effet connaître les résultats de toutes les analyses (analyse des données et analyse multidimensionnelle) avant de déterminer avec précision les types de correction à apporter.

4.4.1.2. Pour l'intégralité des inventaires :

Connaissant la signification de chaque attribut pour chaque inventaire, il est maintenant possible de chercher les correspondances entre inventaires.

La première correspondance vise à établir l'équivalence sémantique entre attributs. Ceci revient à déterminer un groupe d'attributs (sur plusieurs inventaires) ayant sensiblement la même sémantique de façon à ce que ce groupe permette de suivre l'évolution descriptive d'un phénomène.

Ainsi, les attributs suivants furent jugés suffisamment identiques pour être comparés :

- Age
- Hauteur
- Densité
- Perturbation
- Essence

4.4.2. Analyse des données spatiales :

4.4.2.1. Par inventaire :

Chaque inventaire forestier utilise la primitive spatiale peuplement comme unité de base pour la gestion de la forêt. Un peuplement correspond à un ensemble d'attributs où l'homogénéité des valeurs est constatée (i.e. une seule essence par peuplement, une seule densité, etc.) L'ensemble d'attributs servant à former les peuplements a changé au fil des inventaires.

Puisque que le nombre d'attributs servant à délimiter les peuplements a augmenté, il n'est pas surprenant de constater une hausse du nombre de peuplements et une diminution de leur superficie.

4.4.2.2. Pour l'intégralité des inventaires :

Alors que l'analyse des données spatiales par inventaire est relativement simple, l'analyse de ce type de données sur plusieurs époques est, dans le présent contexte, plutôt complexe. La nature même du phénomène cartographié est à la base de cette complexité. Il s'agit d'une cartographie d'un phénomène continu où les critères de délimitation des peuplements varient dans le temps. Il s'ensuit que l'évolution d'un objet (i.e. peuplement) ne peut être suivie puisque celui-ci existe pour

Conception

une seule époque. De plus, par la nature du phénomène observé, la délimitation des peuplements est plutôt floue (précision de quelques dizaines de mètres). Pour ces raisons, un processus d'analyse spatio-temporelle ne peut être réalisé directement sur les peuplements.

4.5. Analyse multidimensionnelle :

Connaissant avec exactitude la nature des données sources, il est maintenant possible d'analyser celles-ci sous l'aspect de dimensions. Les dimensions forment le cœur de l'entrepôt de données impliqué dans le projet d'analyse de la forêt Montmorency:

Voici la liste des dimensions qui furent établie pour cette partie.

- **Dimension temporelle :** Inventaire
- **Dimension spatiale :** Découpage
- **Dimensions thématiques:** Age
Densité
Essence
Hauteur
Pente
Perturbation

4.5.1. Dimension Année :

Des inventaires forestiers de la forêt Montmorency ont eu lieu en 1973, 1984, 1992 et de 2001 à 2010. L'évolution de la forêt est étudiée à l'aide d'analyses temporelles basées sur ces 13 époques. Telle qu'illustrée à la figure ci-contre, la dimension année a un seul niveau : Année. Celui-ci a 13 membres

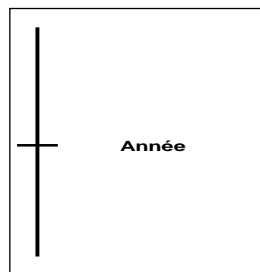


Figure 4.2 : Dimension année

4.5.2. Dimension Découpage :

Depuis 1973, la gestion de la forêt Montmorency a été réalisée à l'aide de trois découpages administratifs. Ces découpages sont constitués d'agréations particulières de peuplements forestiers. La figure suivante illustre les trois découpages utilisant la notion de peuplement comme niveau de détail le plus fin sur le territoire.

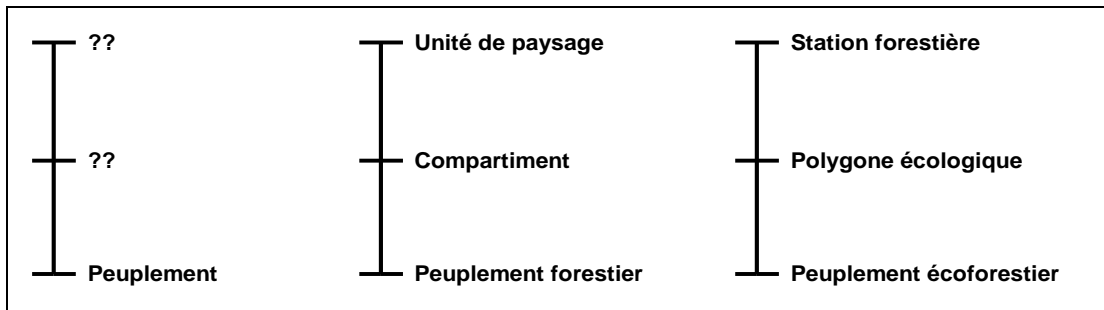


Figure 4.3 : Découpage territorial de la forêt Montmorency.

Mais le découpage qu'on nous a demandé pour cette étude est Peuplement, UP, Forêt. Mais l'utilisation du peuplement comme unité de base pose un problème d'hétérogénéité spatiale.

Les problèmes d'hétérogénéités sont liés à la nature des objets spatiaux considérés lors des inventaires et à l'évolution sémantique des données descriptives. Les objets spatiaux pris en compte par les inventaires forestiers sont les peuplements. Ceux-ci n'étant définis que pour un inventaire donné, ils ne peuvent être utilisés comme objets de référence pour l'étude des évolutions.

Les données descriptives sont déterminées par la législation, le mode d'acquisition et les spécifications propres à chaque inventaire. L'évolution de ces paramètres rend la tâche d'intégration des données particulièrement difficile.

La stratégie à adopter pour solutionner ce problème revient à créer une nouvelle entité, persistante pour tous les inventaires considérés, à laquelle les valeurs descriptives de chaque inventaire sont associées. Ainsi, il est possible de comparer les valeurs descriptives d'une entité territoriale sur plusieurs époques. En ce sens, il existe deux solutions possibles : découper le territoire de la Forêt en une matrice à fine cellules qui resteraient fixes à travers le temps. Ou bien d'utiliser la solution que nous avons choisie et qui consiste en l'utilisation de l'opération d'overlay. La figure suivante résume cette approche.

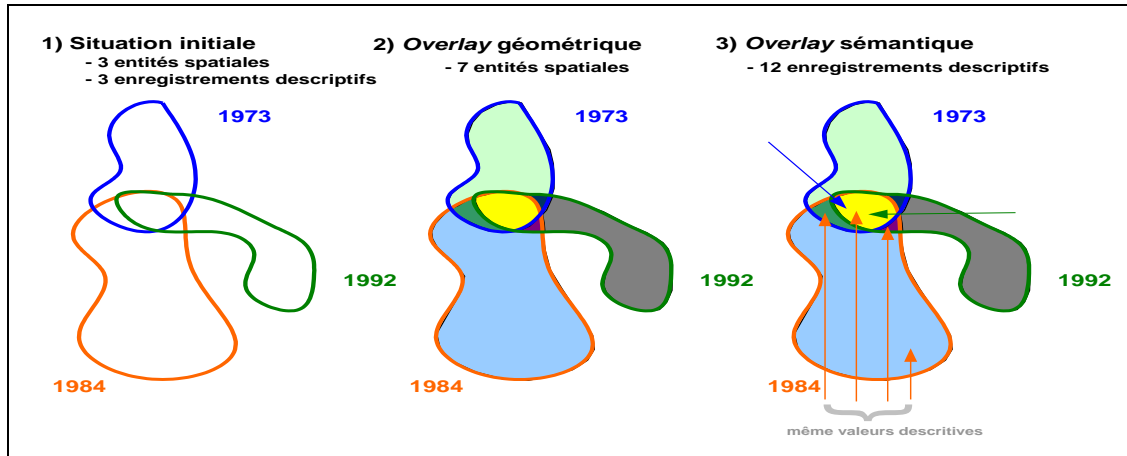


Figure 4.4 : Opération d'overlay pour la génération de sous-peuplement.

Comme on peut le constater au point 3), les peuplements sont éclatés en entités plus petites, nommées sous-peuplements. Ceux-ci sont persistants pour les trois inventaires et l'évolution descriptive peut être constatée.

Un inconvénient majeur doit être mentionné. Puisque la cartographie des sous-peuplements est dérivée des peuplements et que ceux-ci dépendent des inventaires, l'ajout d'un nouvel inventaire entraîne une régénération complète de la cartographie. De plus, le nombre de sous-peuplements croîtra de façon non linéaire pour atteindre des proportions gigantesques. A titre d'exemple, l'opération d'overlay des trois inventaires (environ 8000 peuplements) engendre 50000 sous-peuplements. Tous les liens entre les données descriptives et temporelles des inventaires précédents doivent être refaits. Enfin, la consultation d'une carte de sous-peuplements est difficile car elle est surchargée de polygones irréguliers. Cette approche pourtant fonctionnelle et qui permet de supporter un processus d'analyse spatio-temporelle doit donc être délaissée au profit d'une solution plus adéquate.

4.5.3. Dimension Age :

Cette dimension à 3 niveaux hiérarchiques permet d'analyser la forêt selon l'âge des peuplements. Elle est représentée simplement par la figure ci-contre. Les membres des niveaux « Groupes » et « Classe » sont supposés être homogènes dans le temps. Par contre les membres du niveau « Âge détaillé » changent de valeur en fonction des différents inventaires.

En ce sens, (La figure 4.4) explicite les différents membres possibles pour ce niveau.

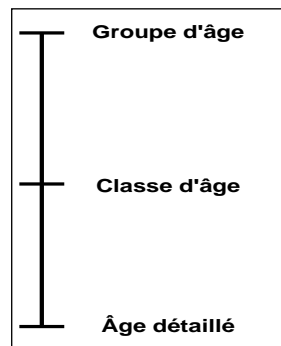


Figure 4.5: Dimension Age.

On peut noter que la correspondance entre les classes d'âge suivant les différents inventaires n'est pas parfaite. Ainsi, en 1984, la classe d'âge «jeune» est composée des 20-40 ans alors qu'en 1992, elle comprend les 15-45 ans. On supposera que l'erreur ou l'approximation faite est acceptable et que cette classe est donc comparable au cours du temps.

4.5.4. Dimension Densité :

Cette dimension est simple puisque qu'elle n'a pas de hiérarchie. La densité a toujours été codifiée de la même façon. Le niveau « Densité détaillée » a donc quatre membres indépendants des inventaires. Le tableau suivant résume cette dimension.

Densité détaillée	Descriptif	densité
1	Peuplement où le couvert est supérieur à 81 %	+ de 80 % de couvert
2	Peuplement où le couvert est supérieur de 61 à 80 %	61- 80 % de couvert
3	Peuplement où le couvert est supérieur de 41 à 60 %	41-60 % de couvert
4	Peuplement où le couvert est supérieur de 25 à 40 %	25-40 % de couvert

Valeurs détaillées de densité.

Tableau 4.1:

Puisque la densité a toujours été codifiée de la même façon, toute densité peut être directement comparée à la densité d'un autre inventaire.

4.5.5. Dimension Essence :

La structure de cette dimension est illustrée par la figure ci-dessous.

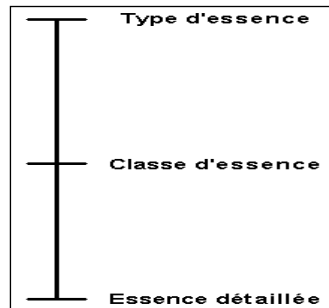


Figure 4.6: Dimension Essence.

Tout comme la dimension « Age », les deux niveaux supérieurs sont indépendants des inventaires. Par contre, les membres du niveau « Essence détaillée » dépendent de l'inventaire.

L'essence détaillée « inclassable » car on peut supposer que les critères de classification s'affinant au cours des inventaires, cette catégorie n'a pas la même signification d'un inventaire à l'autre.

Comme pour la dimension Age, l'homogénéisation du niveau « classes d'essences » résulte d'une approximation admissible.

4.5.6. Dimension Hauteur :

Telle qu'illustrée par la figure suivante, cette dimension est composée de deux niveaux : Groupe hauteur et Hauteur détaillée.

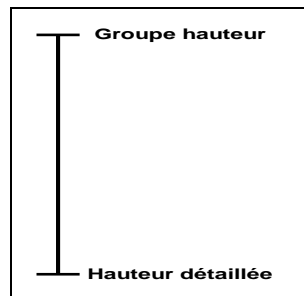


Figure4.7: Dimension Hauteur.

Alors que le niveau « Groupe hauteur » a 3 membres, le niveau « Hauteur détaillée » a plusieurs membres et ceux-ci dépendent de l'année d'inventaire

4.5.7. Dimension Pente :

Tout comme la dimension « Densité », la dimension « Pente » n'a pas de hiérarchie. Un seul niveau compose cette dimension.

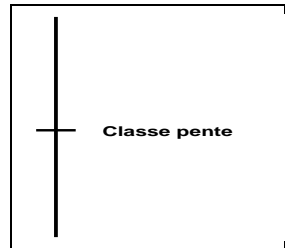


Figure 4.8 : Dimension Pente.

Cette donnée n'est collectée que depuis l'inventaire de 1992 et peut être considérée comme invariante. Il n'est donc pas possible de l'inclure dans les analyses temporelles. Le tableau suivant présente toutefois les différents membres du niveau « Classe pente ».

Classes pentes	Code de pentes	Description
abs		Absence de données
0-3 %	A	Pente nulle
4-8 %	B	Pente faible
9-15 %	C	Pente douce
16-30 %	D	Pente modérée
31-40 %	E	Pente forte

Tableau 4.2: Valeurs détaillées de pente.

4.5.8. Dimension Perturbation :

Cette dimension est composée des 3 niveaux suivants :

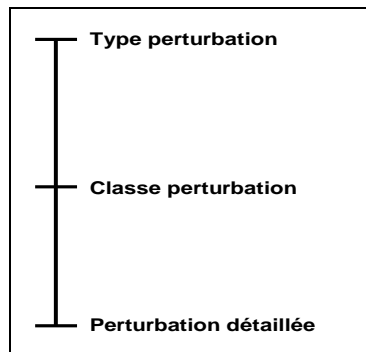


Figure 4.9 : Dimension Perturbation.

4.5.9. Le Fait :

Le fait contient les valeurs à analyser. L'analyse se portera les mesures suivantes : Superficie, Volume de bois résineux, Volume de bois feuillus.

4.5.10. Homogénéité du découpage :

L'une des particularités de ce projet est l'homogénéité des valeurs observées en un lieu donné. En effet, pour un peuplement et un inventaire particulier, il ne peut y avoir deux valeurs de perturbation, de hauteur, d'essence, etc. Une seule valeur est observée. Il en résulte que les dimensions Age, Densité, Essence, Hauteur, Pente et Perturbation sont plutôt des attributs des faits observés (peuplement, temps) que de véritables dimensions. Seules les dimensions Découpage et Inventaire contribuent à générer des faits. Cependant, pour faciliter la navigation dans les données, l'âge, la densité, l'essence, la hauteur, la pente et la perturbation figurent en tant que dimensions.

Ainsi, un cube aura le même nombre d'enregistrements dans la table de faits et ce, peut importe s'il à été construit avec une ou plusieurs de ces pseudo dimensions. Cette particularité aura un impact direct lors de la conception de l'architecture de l'entrepôt de données.

4.6. Modèles conceptuels de l'entrepôt en étoile :

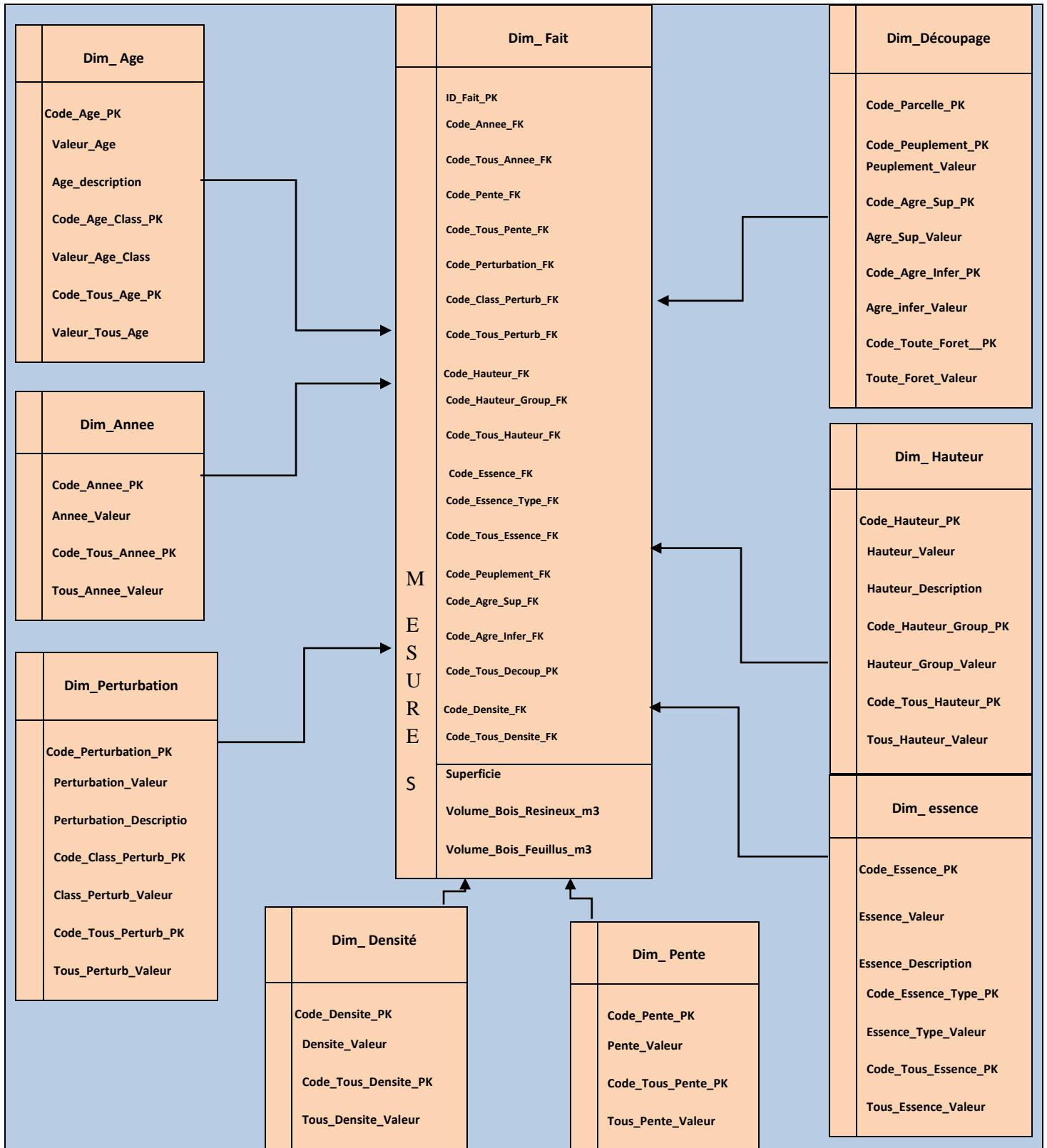


Figure 4.10 : Modèle conceptuel de l'entrepôt de données de la forêt

4.7. Conception de l'outil d'analyse SOLAP :

Pour explorer adéquatement des données à caractère spatial d'un entrepôt de données, un SOLAP s'avère être indispensable.

Il va sans dire que la mise au point d'un SOLAP demande une bonne dose de connaissance de programmation et des API (*Application Programming Interface*) d'un SIG et d'un logiciel OLAP. Un choix de fonctionnalités a été arrêté pour rendre le développement conforme à la charge de travail alloué à ce projet.

Les caractéristiques suivantes furent jugées importantes et nécessaires :

- Opération d'analyse OLAP par les volets cartographique, graphique et tabulaire
- Analyse via le volet descriptif et le volet cartographique de façon indépendante
- Traitement sur une ou plusieurs dimensions spatiales géométriques
- Plusieurs mesures visualisables à la fois

4.8. Conclusion :

Dans ce chapitre, nous présentons des solutions pour concevoir des structures multidimensionnelles lorsque les sources de données sont hétérogènes du point de vue temporel, spatial et sémantique. Pour traiter l'hétérogénéité spatiale, les données en mode vectoriel sont ramenées dans un référentiel fixe. Puis, de nouvelles classifications des données descriptives sont introduites en groupant les attributs descriptifs initiaux en attributs génériques, invariant temporellement. Les dimensions thématiques sont définies à partir de cette nouvelle organisation et des besoins des utilisateurs.

Chapitre 05

Réalisation

- 5.1. Introduction
- 5.2. Paramètre de PC
- 5.3. Architecture de l'application
 - 5.3.1. Entrepôt de données spatiales
 - 5.3.2. Serveur OLAP
 - 5.3.3. Client OLAP
- 5.4. Réalisation la carte
 - 5.4.1. Logiciel utilisé
 - 5.4.2. Construction le SIG
- 5.5. L'application finale
- 5.6. Conclusion

5.1. Introduction :

Les concepts innovateurs de l'OLAP Spatial, comme nous l'avons montré aux chapitres 2 et 3, soulèvent de nombreuses problématiques, en ce qui concerne la gestion et la visualisation des données multidimensionnelles, et la navigation dans l'hypercube. Nous avons réalisé une application, un modèle multidimensionnel, qui, se proposant comme une solution formelle pour OLAP Géographique, intègre dans l'analyse multidimensionnelle la composante sémantique et la composante spatiale de l'information géographique et les opérateurs d'analyse spatiale. Dans ce chapitre, nous présentons notre prototype OLAP Géographique, en utilisant les applications multidimensionnelles.

Notre application est une solution OLAP-SIG intégrées classiques en ajoutant un support pour les mesures géographiques et complexes, les dimensions géographiques.

5.2. Paramètre de PC :

Cette partie décrit les infrastructures déjà en place, en effet c'est une étape à ne pas négliger, car la diversité des sources et leurs plateformes techniques pourront engendrer des problèmes de compatibilité.

Le SIAD sera installé sur des machines INTEL ayant de préférence les caractéristiques :

- Processeur : Core i3
- Mémoire RAM : 4Go
- Espace Disque : de 320 Go
- Système d'Exploitation : Windows Xp 32bits
- Résolution graphique : 1366x768

5.3. Architecture de l'application :

Notre application est basée sur une architecture à trois niveaux (Figure 5.1) composée par un SGBD spatial dans lequel sont stockées les données spatiales et alphanumériques, un serveur ROLAP pour la gestion des requêtes multidimensionnelles et un client. Ce dernier est dédié à la formulation de requêtes multidimensionnelles et à la visualisation des résultats. Il intègre un client SIG et un client OLAP.

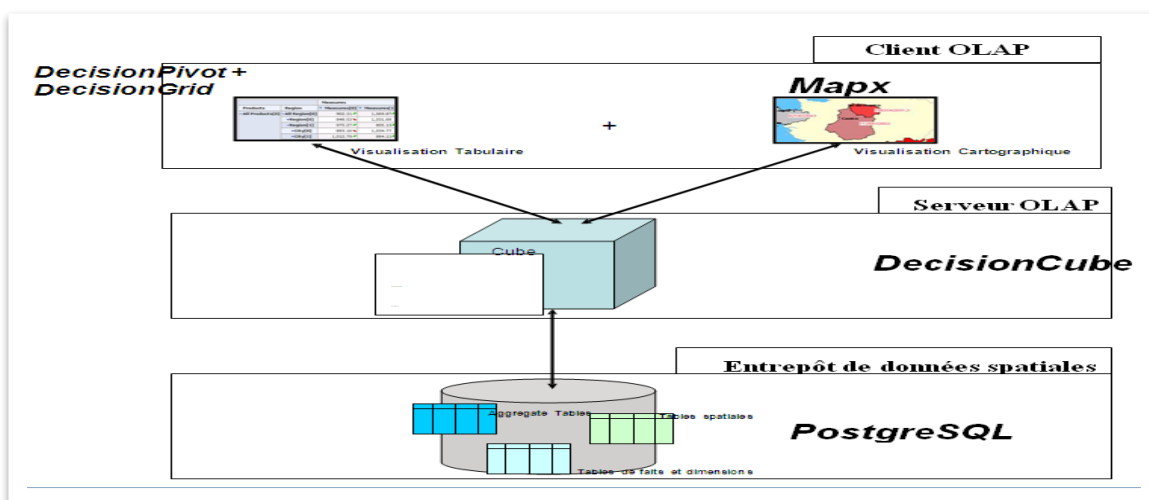


Figure 5.1. Architecture de l'application

Nous distinguons deux différents types d'opérateurs les opérateurs *multidimensionnels non- spatiaux* qui n'utilisent pas la composante spatiale de l'information géographique et les opérateurs *multidimensionnels spatiaux*.

5.3.1. Entrepôt de données spatiales :

L'entrepôt de données spatiales est implémenté en utilisant l'environnement PostgreSQL (PgAdmin3 version 9.0.1). Les données spatiales et alphanumériques de l'application multidimensionnelle sont stockées dans Postgres en utilisant une modélisation en étoile (cf chap 4).

a) Définition l'outil :

PostgreSQL est un système de gestion de bases de données relationnelles objet (ORDBMS) fondé sur *POSTGRES*. Ce dernier a été développé à l'université de Californie au département des sciences informatiques de Berkeley. *POSTGRES* est à l'origine de nombreux concepts qui ne seront rendus disponibles au sein de systèmes de gestion de bases de données commerciales que bien plus tard. PostgreSQL est un descendant OpenSource du code original de Berkeley. Il supporte une grande partie du standard SQL tout en offrant de nombreuses fonctionnalités modernes :

- requêtes complexes ;
- clés étrangères ;
- triggers ;
- vues ;
- intégrité transactionnelle ;
- contrôle des versions concurrentes (MVCC ou multiversion concurrency control).

De plus, PostgreSQL peut être étendu par l'utilisateur de multiples façons. En ajoutant, par exemple :

- de nouveaux types de données ;
- de nouvelles fonctions ;
- de nouveaux opérateurs ;
- de nouvelles fonctions d'agrégat ;
- de nouvelles méthodes d'indexage ;
- de nouveaux langages de procédure.

b) La base de données :

Comme nous l'avons expliqué plus haut, nous avons implémentée l'entrepôt sur des tables de type *postgres* comme il est montré dans la figure 5.2.

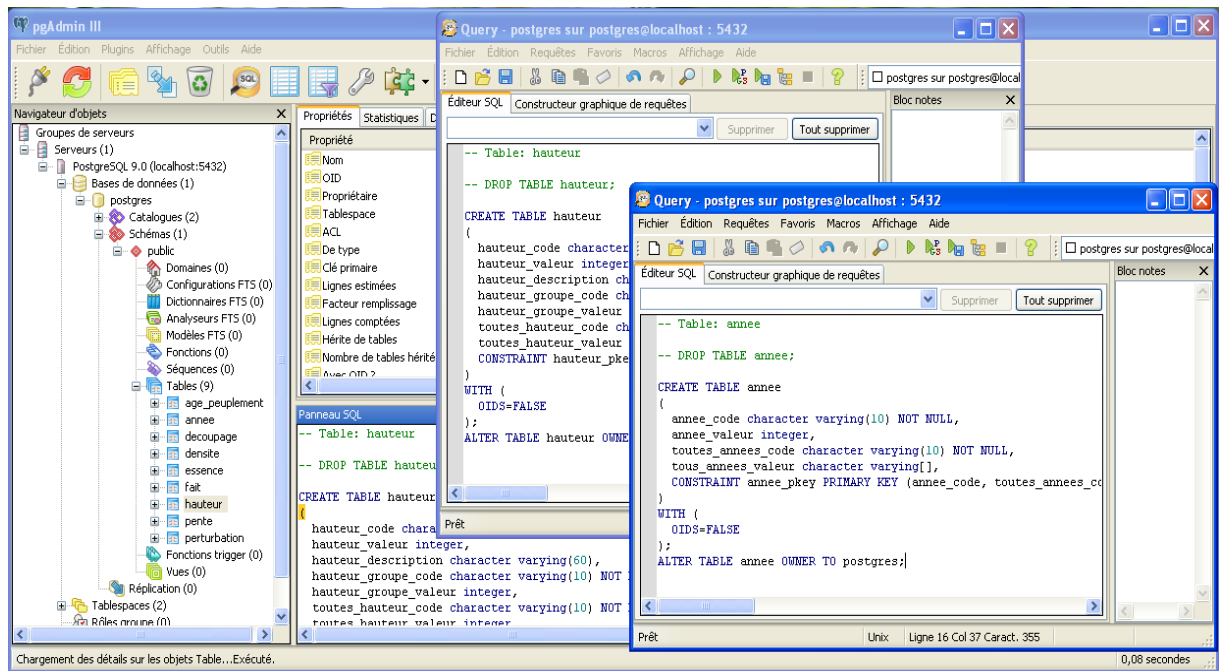


Figure 5.2. La base de données

5.3.2. Serveur OLAP :

Le serveur OLAP a l'importante tâche de gérer les dimensions et les mesures géographiques et complexes, et d'implémenter les opérateurs de Cube. Pour ce niveau, nous avons utilisé la composante intégrée avec l'environnement Delphi (DecisionCube, Delphi 2007). DecisionCube est un serveur OLAP développé pour Delphi, qui s'appuie sur une technologie relationnelle. DecisionCube traduit toutes les requêtes multidimensionnelles en requêtes SQL, et il établit un lien entre le client OLAP et le SGBD.

5.3.3 Client OLAP :

Le client est développé en utilisant la composante Decision Cube. (Decision Cube, Delphi7) Decision Cube c'est une bibliothèque de l'environnement Delphi qui permet d'afficher les résultats des requêtes multidimensionnelles et à des affichages graphiques.

Réalisation

De plus, il permet à l'utilisateur à formuler toutes les opérations de navigation OLAP classiques comme le slice, le dice, le roll-up ou le drill-down.

○ Les outils proposés :

La version Entreprise de Delphi propose dans sa palette d'outils les composants d'aide à la décision

- **DecisionCube** : est une interface permettant de formater l'affichage des données dans un DecisionGrid

- **DecisionGrid** : présente les données structurées et permet la manipulation par l'utilisateur final

- **DecisionQuery** : est un équivalent de TQuery, mais amélioré et proposant un éditeur de requête

- **DecisionSource** : est un équivalent de TDataSource

- **DecisionPivot** : permet la modification de la présentation des données

Nous avons choisi de travailler avec **Mapinfo** pour la réalisation de la carte, en effet il satisfait à tous nos besoins.

De plus pour la réalisation de l'interface utilisateur nous avons choisi l'environnement de développement **Delphi** car c'est l'un des outils qui supporte le composant cartographique **MapX (MapX 5.02)**, **MapX** c'est une bibliothèque de l'environnement Delphi pour la réalisation d'applications cartographiques. MapX permet de créer et visualiser des cartes dans tous les navigateurs web standard. Il implémente toutes les fonctionnalités SIG de base.

5.4. Réalisation la carte :

5.4.1. Logiciel utilisé :

La base de données cartographique est réalisée avec le leader des SIG, (**Mapinfo professionnelle version 7.0**).

Mapinfo est un logiciel permettant d'exploiter un Système d'Information Géographique(SIG).Un logiciel SIG permet l'acquisition, le stockage, la mise à jour, la

Réalisation

manipulation, et le traitement de données géographiques. De plus, il permet de faire de la cartographie et de l'analyse spatiale de façon précise en fonction de l'échelle désirée.

Le principe directeur d'un SIG est le suivant : nous avons d'un coté les données géométriques et de l'autre les données attributaires. Ces données sont stockées sous format numérique et organisées par couches (appelées « **Tables** » dans Mapinfo).

5.4.2. Construction le SIG :

a) Les paramètres de projection :

On utilisant quelques informations cartographiques nécessaires pour réaliser ce travail, le tableau suivant illustre ces informations.

La projection	Quebec MTM Zone 7 (NAD 83)
---------------	----------------------------

Tableau 5.1 : Les paramètres de la projection

b) Application de SIG :

La figure suivante (**Figure 5.3**) est représenté l'application de SIG de la forêt de Montmorency qui réalise par l'outil Mapinfo.

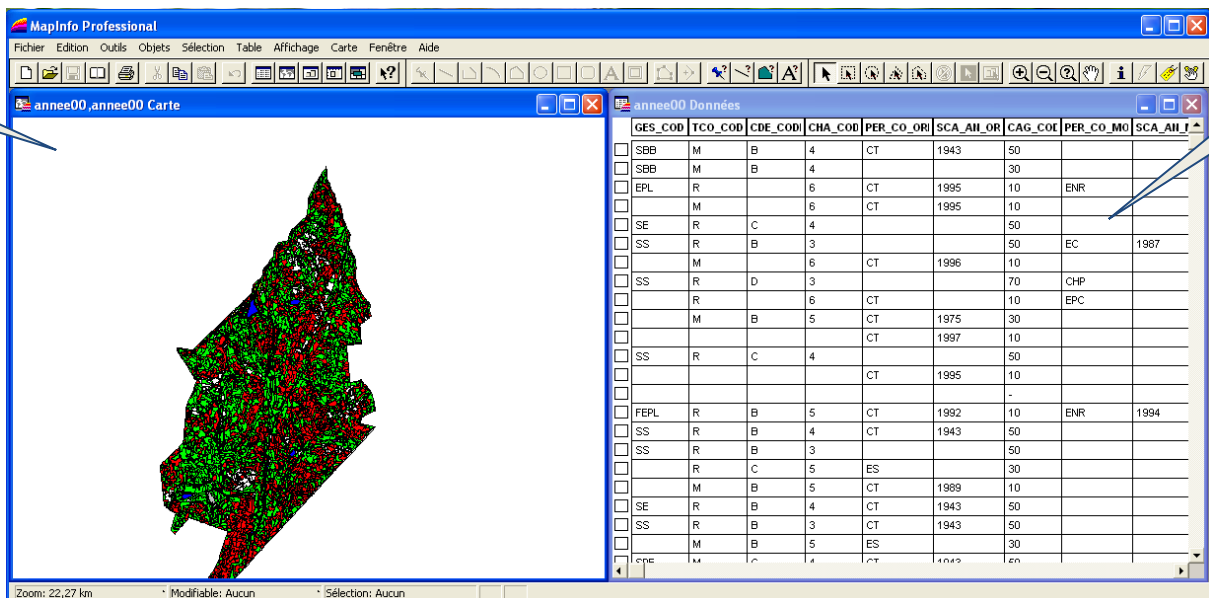


Figure 5.3 : La carte de la forêt de Montmorency avec Mapinfo 7.0

L'application de cartographie qui réalisée est constituée de 11 cartes de l'année 2000 à 2010, chaque cartes avec les données.

Chacune de ces cartes génère un fichier *.TAB* (Voir la Figure Précédente) qui sera exploité par le composant **MapX** pour visualiser la carte dans l'application finale.

- 1) représenté la carte géographique de la forêt de Montmorency.
- 2) représenté les données de la forêt de Montmorency.

5.5. L'application finale :

Nous avons choisi de travailler avec le langage de programmation Delphi car il offre toutes les fonctionnalités dont nous avons besoin en plus d'être le langage qui supporte la bibliothèque spatiale *MapX*, cette dernière fait partie de la même gamme de produit que *Mapinfo* le logiciel SIG que nous avons utilisé pour créer les cartes.

L'interface est construite de manière à être intuitive et facile à utiliser suivant les règles régissant les interfaces SOLAP. Elle se compose de trois types d'affichage de données carte, diagramme et tableau croisé, elle comporte aussi des outils de navigation classiques et des outils d'analyse (navigation SOLAP), une fenêtre de navigation plus grande, une fenêtre d'affichage des données attributaires.

1. la page d'accueil:

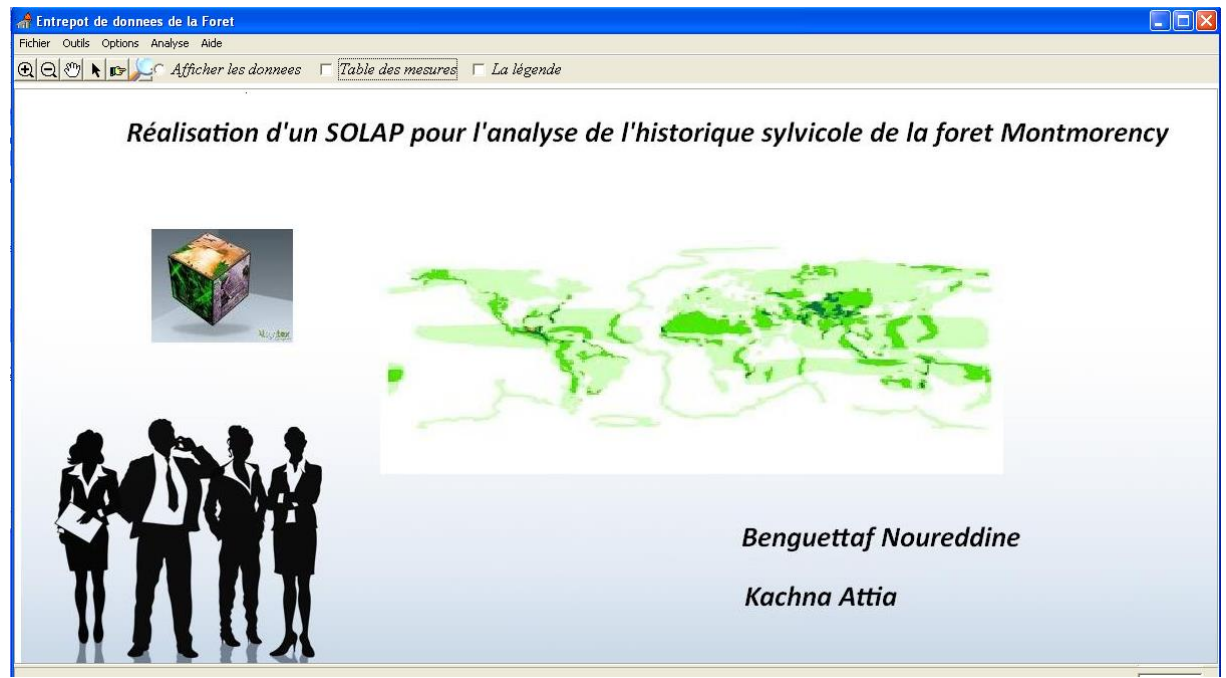


Figure 5.4 : La page d'accueil de l'application

2. La page principale :

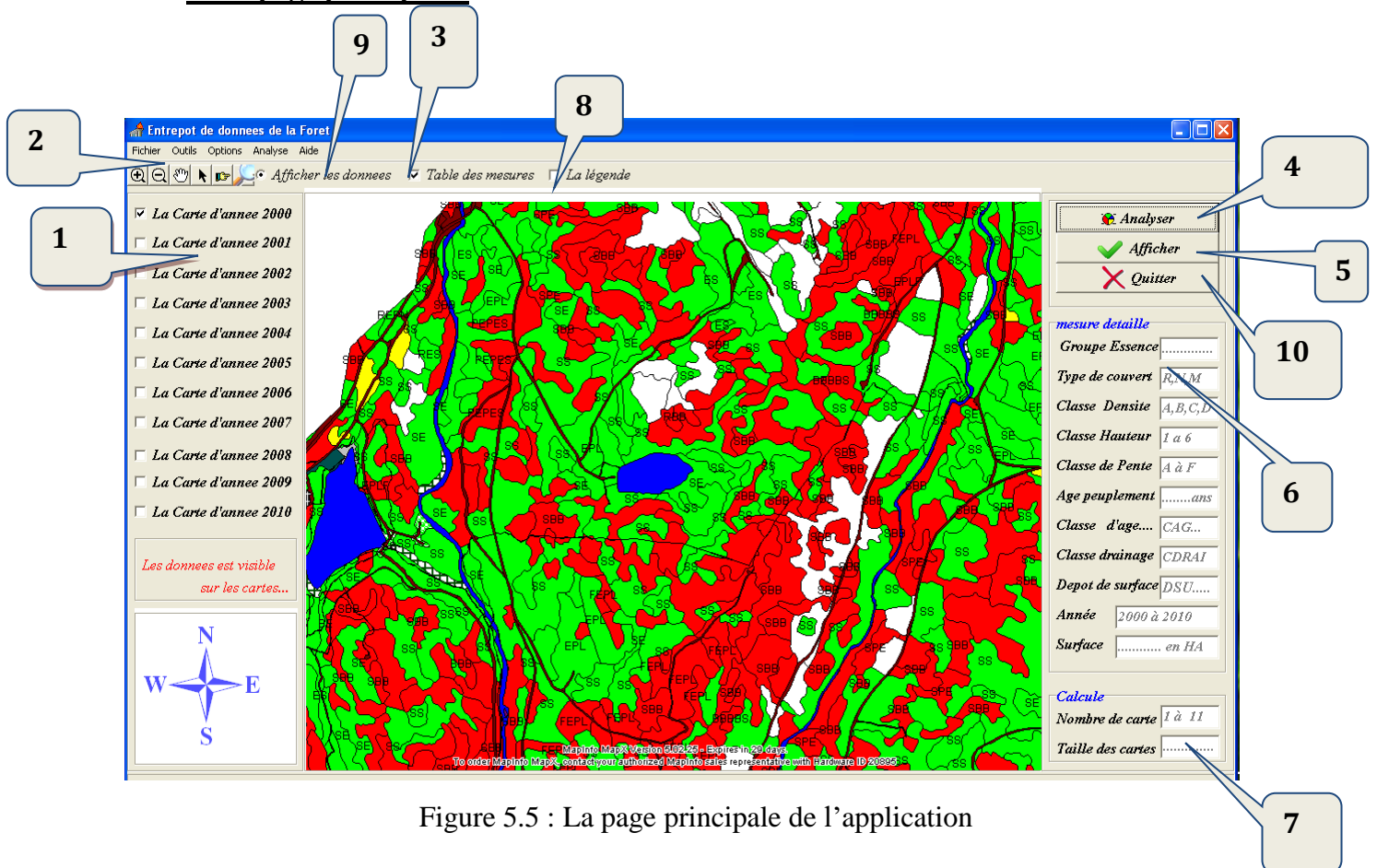


Figure 5.5 : La page principale de l'application

Réalisation

- 1) L'ensemble des cartes affichées avec possibilité d'afficher ou masquer chaque carte.
- 2) L'ensemble d'outils de manipulation de la carte.
- 3) Afficher la table de mesures.
- 4) Afficher l'interface de l'analyse.
- 5) Afficher les données de la carte sur la table des mesures.
- 6) les données qui afficher.
- 7) Le nombre des cartes qui afficher.
- 8) La légende de la carte.
- 9) Afficher les données sur les cartes.
- 10) Quitter l'application.

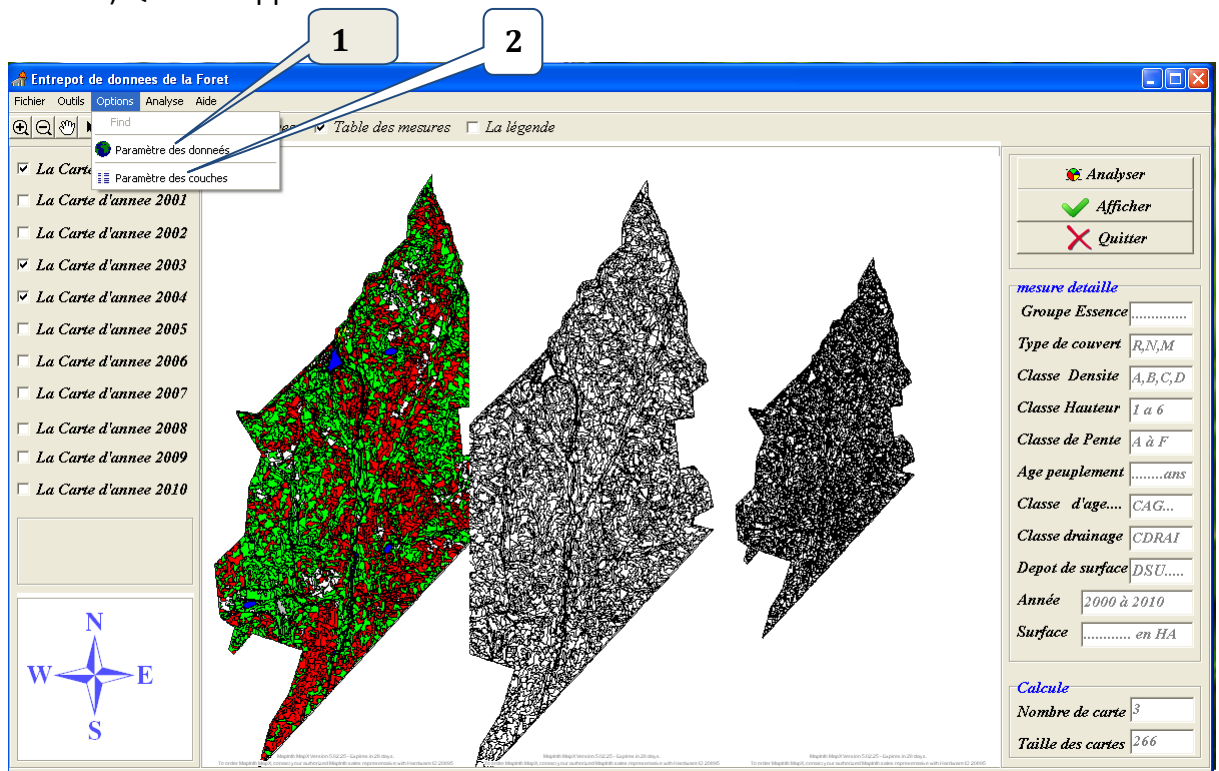


Figure 5.6 : La page de paramètre des données.

- 1) paramètre de créer les données sur les cartes.
- 2) paramètre d'affichage les données et les cartes.

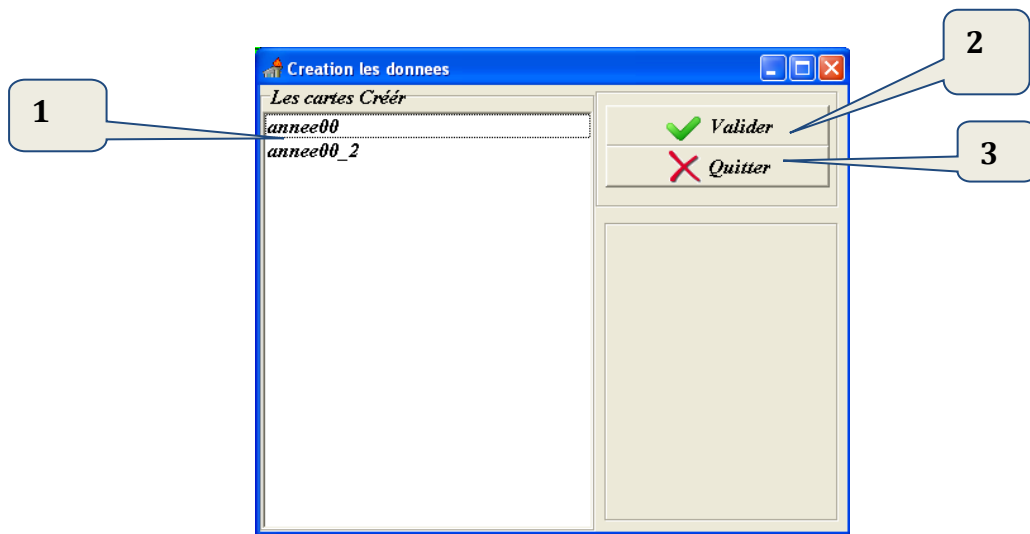


Figure 5.7 : La page de paramètre des cartes.

- 1) les cartes qui visible.
- 2) validation de créé les données.
- 3) Quitter l'application.

2. l'interface d'analyse :

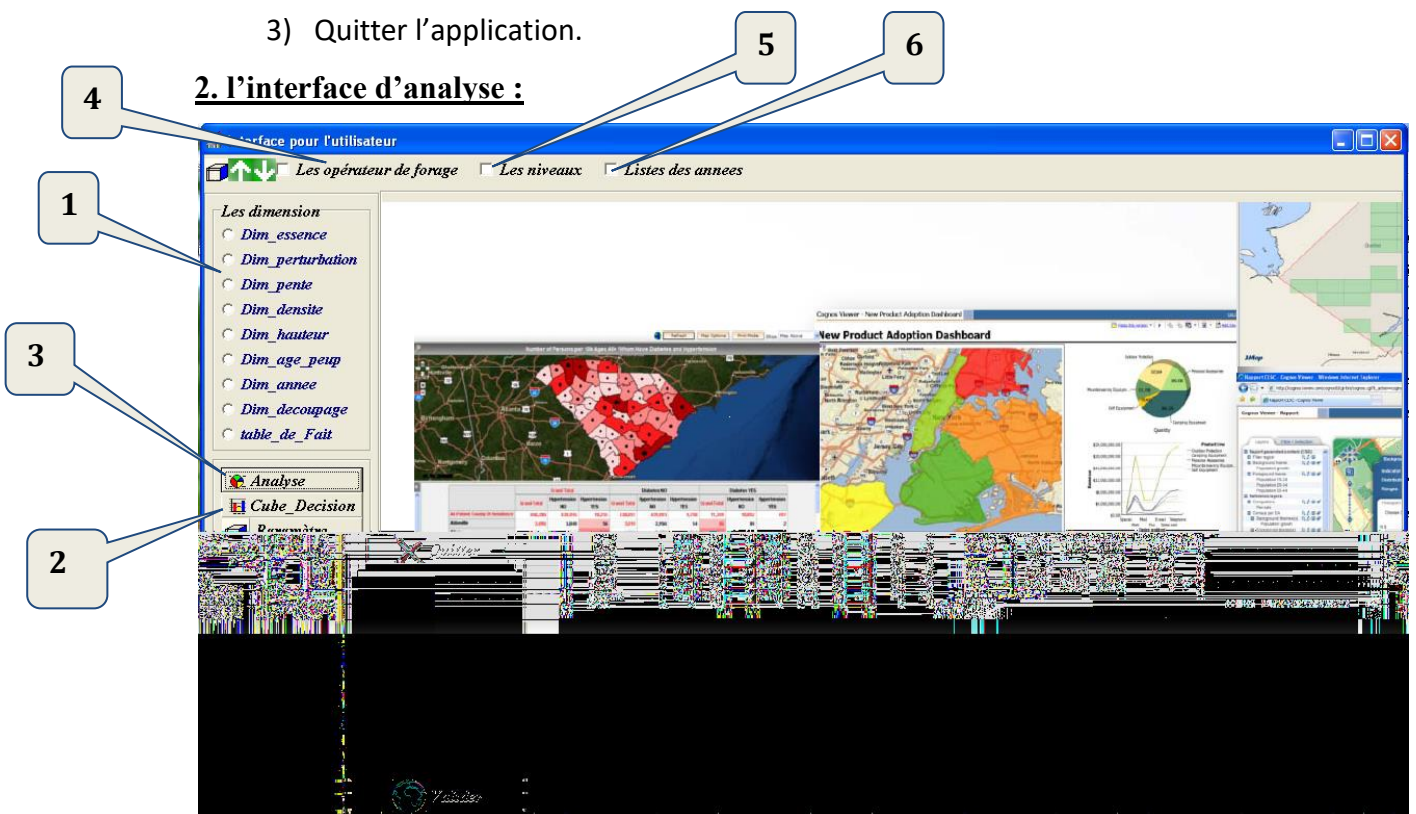


Figure 5.8 : La page d'interface d'analyse.

- 1) choisi les dimensions pour l'analyse.
- 2) Choisi les mesures.

Réalisation

- 3) Commencez l'analyse.
- 4) Afficher les operateurs de forage.
- 5) Afficher les niveaux hiérarchiques du Forêt.
- 6) Afficher la liste des cartes pour chaque année.

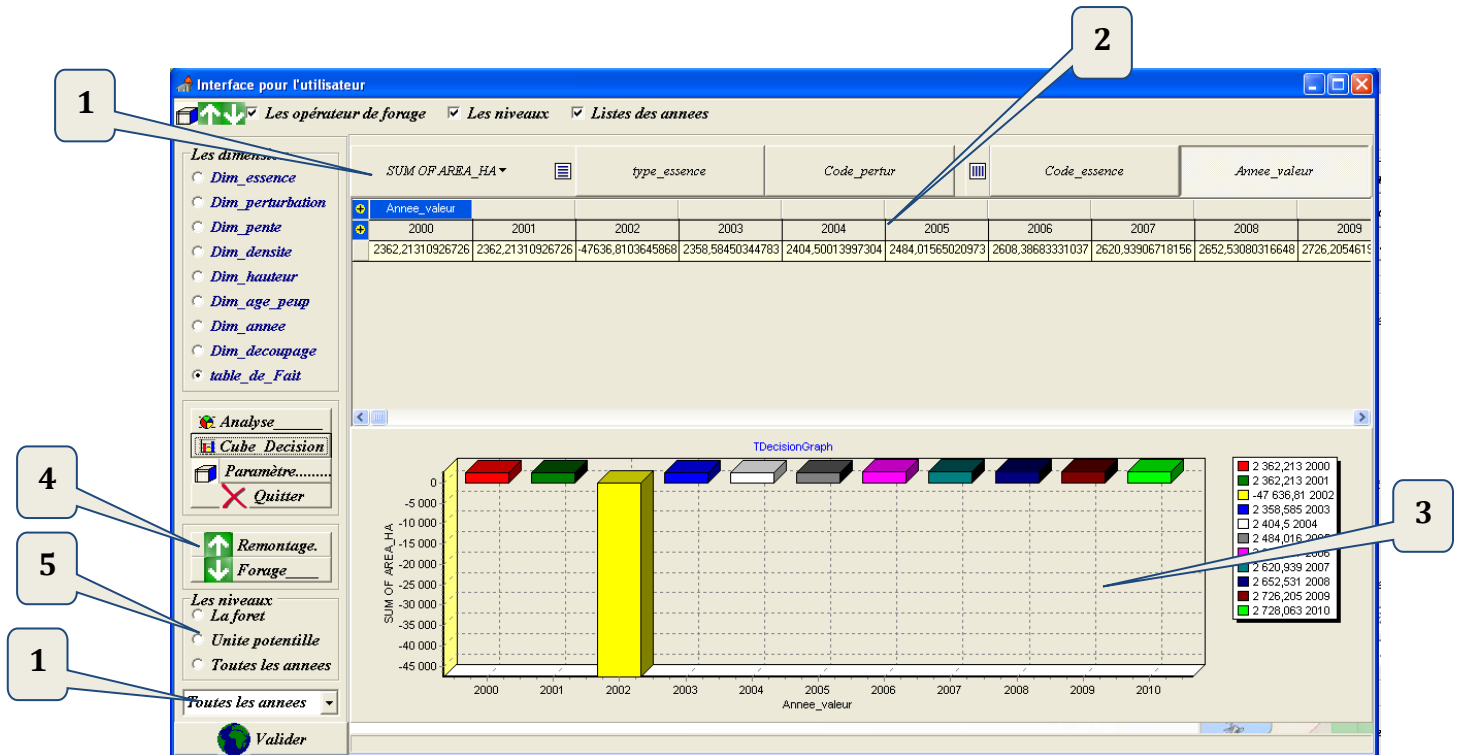


Figure 5.9 : La page de résultat de l'analyse.

- 1) Les dimensions ouvertes lors d'une analyse.
- 2) Le tableau croisé des résultats d'une analyse.
- 3) Les résultats sous forme graphique.
- 4) les buttons des opérateurs de forage.
- 5) les choix de niveau hiérarchique de Forêt.
- 6) les choix des années.

5.6. Conclusions :

Dans ce chapitre, nous avons présenté l'application finale pour une solution OLAP-SIG intégrée pour la Forêt.

Notre travail est basé sur le modèle de Cube. Notre application est gère les mesures géographiques et complexes, et les dimensions géographiques. Il implémente un ensemble d'opérateurs.et présente toutes les caractéristiques fondamentales de visualisation et d'interaction pour une analyse spatio-multidimensionnelle effective, même si la prise en compte des mesures géographiques nécessite une reformulation des techniques de visualisation et d'interaction existantes.

Chapitre 06

Conclusion générale

- 6.1. Conclusion
- 6.2. Perspectives

Conclusion générale

6.1. Conclusion :

Dans le cadre de ce mémoire nous avons présenté La notion des entrepôts de données spatiales associés à des outils d'analyse Spatial On Line Analytical Processing, (SOLAP) qui représentent une solution effective pour l'informatique décisionnelle. Les données qualitatives des inventaires sont organisées dans des hypercubes en axes d'analyses appelés dimensions. Les sujets d'analyse, appelés tables de fait, sont caractérisées par les données quantitatives appelées mesures. Les résultats issus de cette analyse sont calculés à l'aide de fonctions d'agrégations selon les différentes granularités définies par le schéma hiérarchique des dimensions.

Aujourd'hui les outils décisionnels sont beaucoup plus puissants que les systèmes relationnels dotés d'extensions multidimensionnelles (opérateurs CUBE, ROLLUP, ...).

La solution est de collecter ces information dans un lieu centralisé de données (entrepôt de données), structurer ces données de manière a refléter la vision des décideurs (modèle multidimensionnel), mettre cette structure disponible pour l'analyse et l'interrogation.

Cette solution présente une démarche connue comme OLAP (on-line analytical processing) et SIG (system d'information géographique). Notre travail rentre dans le domaine d'organisation de données dans un entrepôt de données tout en se basant sur un modèle multidimensionnel (schéma en étoile). L'objectif est d'offrir à l'utilisateur (décideur) un système qui peut être traduire sur une interface interactive pour l'analyse, l'interrogation et la visualisation de données de manière multidimensionnelle de La forêt Montmorency et La résolution du problème d'hétérogénéités des données spatiales.

Sur un aspect personnel, travailler sur ce projet nous a permis de nous immerger dans un domaine en plein essors, de nous familiariser avec ses techniques et ses technologies et nous constituer une bibliographie riche en ce sujet. Les problématiques rencontrées sur ce projet pourront constituer de nouveau projets à réaliser.

6.2. Perspectives :

Parmi les perspectives que nous envisageons on peut citer:

Extension de la modélisation de l'entrepôt. Dans ce mémoire nous avons utilisé un modèle à un seul fait (modèle en étoile), mais dans l'analyse multidimensionnelle, il est nécessaire d'utiliser plusieurs faits (modèle en constellation).

Opération d'extraction reste difficile à réaliser et nécessite un expert pour élaborer et alimenter le magasin de données. Il est utile de proposer des modèles et des méthodes pour automatiser cette opération.

L'aspect temporel joue un rôle vital dans les systèmes décisionnel, donc la représentation du temps par un simple attribut ne permette pas de mieux exploiter les données de l'entrepôt. L'utilisation des modèles s'appuyant sur cet aspect augmente l'utilité d'analyse multidimensionnelle.

Annexe

Tableau1.Liste des attributs base de données Forêt Montmorency :

Nom du champ	Description	Valeurs	
Fait_ID	Numéro interne	0 à 6763	
ID_peuplement	Numéro commun avec PEP et PET	0 à 6646	
AREA	Superficie en mètres carrés	100.54 à 148327.228	
PERIMETER	Périmètres en mètres carrés	43.444 à 9621.512	
POF_NO	Numéro polygone forestier	01001 à 05601	
APPEL_1	Appellation forestière		
UTR	Unité territoriale de référence	Cascades	
		Eaux-Volées	
		Piché	
		Rivière Noire	
		Rivière Blanche	
AF_RAFINEE	Plan d'affectation raffinée	Bois	Production de bois
		Bois-Pays	Production de bois prioritaire et protection du paysage
		Eau	Eau
		Enc_vis	Encadrement visuel
		MR_0020	Milieu riverain 0 à 20 mètres
		MR_2060	Milieu riverain 20 à 60 mètres
		MR_Cas	Milieu riverain castor
		MR_Fray	Milieu riverain frayère
		MR_Rés	Milieu riverain réservoir
		MR_Ruis	Milieu riverain ruisseau
		Pays-Bois	Protection du paysage et production de bois
		Ref_Bio	Refuge biologique
GES_CODE	Groupement d'essence		

Annexe

PSC_CODE	Code de particularité		
TCO_CODE	Type de couvert	R	Résineux
		M	Mélangé
		F	Feuillus
CDE_CODE	Classe de densité	A-B-C-D	
CHA_CODE	Classe de hauteur	1 à 6	
PER_CO_ORI	Code d'origine	ES	Épidémie sévère
		CT	CPRS
		CHT	Chablis total
		P	Plantation
		PLR	Plantation avec semis en récipients
		PR	Plantation avec semis à racines nues
SCA_AN_ORI	Année d'origine	1932 à 2004	
CAG_CODE	Classe d'âge		
PER_CO_MOY	Code de perturbation	EPC	Éclaircie précommerciale
		ENR	Enrichissement de plantation
		EL	Épidémie légère
		EC	Éclaircie commerciale
		CP	Coupe partielle
		CHP	Chablis partiel
SCA_AN_MOY	Année de la perturbation		
CDR_CODE	Classe de drainage	20 à 60	
CLP_CODE	Classe de pente	A-B-C-D-E-F	
DSU_CODE	Dépôt de surface		
TER_CODE	Code terrain	AL	Aulnaie
		AUT	Autre
		DEF	Défriché
		DH	Dénudé humide
		EAU	Eau
		GR	Gravière
		HAB	Habitation
		ILE	Ile
		INC	Inconnu
		INO	Inondé
		RO	Route
		VIL	Villégiature
SEV_CODE	Série évolutive		

Annexe

CODE_ORI	Code d'origine FM	CT	
ANNEE_ORI	Année d'origine FM	1974-1997	
CODE_TER	Code terrain FM	GR	Gravière
		VIL	Villégiature
CODE_PERT	Code perturbation FM	ENRST	Enrichissement de sentier
ANNEE_PERT	Année perturbation FM	1990 à 2000	
TIGE_RES	Tige résiduelle FM	Contenu avec - de 25% de surface terrière	
MOD_ID	Numéro du modelé écologique	0 à 169	
UN_POT_AM	Unité potentiel d'aménagement	1.1 à 7.3	
REGFIN	Regroupement final	0 à 100	
STD_DEV	Stade de développement	ct	coupe
		Str_>7 m	Strate plus de 7 mètres
		Str_0-4 m	Strate 0 à 4 mètres
		Str_4-7 m	Strate 4 à 7 mètres
		Str_inac	Strate inaccessible

Bibliographie & Webographie

- 1). [**Bimonte**, 2007] : Sandro Bimonte « Intégration de l'information géographique dans les entrepôts de données et l'analyse en ligne : de la modélisation à la visualisation ». Thèse de grade de docteur, L'Institut National des Sciences Appliquées de Lyon, 2007
- 2). [**Y. Bédard**, 2010] : Yvan Bédard «Le géodécisionnel: origine, évolution, état de l'art, enjeux, R&D».Ecole Nationale Supérieure des Mines de Paris, Centre de Recherche sur les Risques et les Crises, *Sophia-Antipolis, France, 29 octobre 2010.*
- 3). [**DJEDDAOUI**, 2006] : DJEDDAOUI Mohamed «Elaboration d'un système multidimensionnel». Thèse de grade magistère UNIVERSITE MOHAMED BOUDIAF DE M'SILA, 2006.
- 4). [**MRNF**, 2008] Ministère des Ressources naturelles et de la Faune. NORME DE STRATIFICATION ÉCOFORESTIÈRE QUATRIÈME INVENTAIRE ÉCOFORESTIER, Forêt Québec Direction des inventaires forestiers, Octobre 2008
- 5). [**Wik**, 2012] : Wikipédia «université Laval» Un article de Wikipédia, l'encyclopédie libre
[http://fr.wikipedia.org/wiki/universite Laval](http://fr.wikipedia.org/wiki/universite_Laval), Dernières visites juillet 2012.
- 6). [**Y. Bédard, PhD**, 2000] Yvan Bédard, PhD «Enrichissement de l'OLAP pour l'analyse géographique: exemples de réalisations et différentes possibilités technologiques». Université Laval, Québec, Canada, G1K 7P4, 2000.
- 7). [**Y. Bédard**, 2010] Yvan Bédard «Entrepôts de données spatiales Notions avancées de bases de données SIG». Université Laval, Québec, Canada, 2010.
- 8). [**Y. Bédard**, 2010] Yvan Bédard «OLAP et SOLAP Notions avancées de bases de données SIG». Université Laval, Québec, Canada, 2010.
- 9). [**Y. Bédard, Marie**, 2009] Yvan Bédard, Marie-Josée Proulx «Approche SOLAP Notions avancées de bases de données SIG». Université Laval, Québec, Canada, 5 novembre 2009.
- 10). [**Jerbi**, 2010] Housseem Jerbi «entrepôt, analyse en ligne et fouille de données». jerbi@irit.fr, 14/12/2010.
- 11). [**dev**, 2012] : développez «informatique décisionnelle»
<http://developpez.com/informatiquedecisionnelle>, Dernières visites 11 avril 2012.
- 12). [**Wik**, 2012] : Wikipédia «forêt Montmorency» Un article de Wikipédia, l'encyclopédie libre
[http://fr.wikipedia.org/wiki/forêt Montmorency](http://fr.wikipedia.org/wiki/forêt_Montmorency), Dernières visites 30 juillet 2012.
- 13). [**Bri**, 2010] Alex Brisebois «Notions avancées d'intégration de données dans un contexte de peuplement d'entrepôt de données spatiales». SCG-65825 Sujet spécial, UNIVERSITE LAVAL, 4 janvier 2002.

