



Laboratoire Lorrain de Recherche
En Informatique et ses Applications



UNIVERSITÉ
Université Nancy 2
UFR MI

Eléments méthodologiques pour la construction des bases de données multidimensionnelles

Application dans le domaine du pilotage en milieu bancaire



présenté et soutenu le 18 11 1999

Pour l'obtention du

Diplôme de Recherche Technologique
Option : Systèmes d'Information des Organisations
De l'Université de Nancy 2 - France

Par

Arnaud MALON

Composition du jury

Directrice de recherche : Odile Foucaut, Professeur d'informatique à l'Université Nancy2

Rapporteurs : Odile Thiery, Professeur d'informatique à l'Université Nancy 2
Serge Vendemini, Professeur de Sciences de Gestion à l'ICN

Examineurs : Dominique Lafon, Directeur Général Société AURUS

Didier Bouterin, Ingénieur Banque SNVB
Christine Bourjot, Maître de conférence en informatique à
l'Université Nancy 2

Remerciements

Je tiens tout d'abord à remercier Madame Odile Foucaut pour m'avoir encadré pendant mon travail de DRT, pour m'avoir accordé du temps et s'être investie pour la bonne réalisation de mes travaux à la SNVB et pour la rédaction de ce mémoire.

Je remercie, Odile Thiery, responsable du DRT et rapporteur de ce rapport pour m'avoir convaincu de suivre cette formation.

Je remercie Monsieur Jean Pierre Eckly de la Banque SNVB pour m'avoir accueilli au sein du services des Etudes Informatiques de la SNVB.

Je remercie mes collègues du domaine SIC, Didier Bouterin, Alain Liénard, Nicole Klein, Angelica Aycart, Pascal Baran, Roger Mazetto, Xavier Idoine, Sandrine Toussaint pour ces 18 mois passés en leur agréable compagnie.

Je remercie également Marie Hélène Quil, responsable du projet SIGMA pour m'avoir permis de participer à ce projet.

Je remercie l'ensemble du personnel de la Banque SNVB pour la bonne ambiance.

Je remercie les membres du jury, Messieurs Dominique Lafon, Serge Vendemini et Madame Christine Bourjot pour le temps qu'ils ont consacré à l'étude de ce mémoire.

Je remercie également tout ceux que j'ai oubliés.

Table des matières

PRESENTATION DU CONTEXTE DU DRT	5
CHAPITRE I. ARCHITECTURE ET CONCEPTS DES SYSTEMES DECISIONNELS	6
INTRODUCTION	6
I.1. TYPES D'ARCHITECTURE POUR LES SYSTEMES DECISIONNELS	6
I.1.1. VUE EXTERNE DU SYSTEME DECISIONNEL	6
I.1.2. DE L'INFOCENTRE AU DATAWAREHOUSE	7
I.1.3. PROBLEMES PARTICULIERS	10
I.2. CONCEPTS DES BASES DE DONNEES MULTIDIMENSIONNELLES	11
I.2.1. MESURES, DIMENSIONS, RELATIONS, FORMULES	11
I.2.2. REPRESENTATION GRAPHIQUES DES SCHEMAS DES BDM	16
I.2.3. CONCEPTS RELATIFS A ORACLE EXPRESS	19
CHAPITRE II. METHODES DE CONCEPTION DES BDM	22
INTRODUCTION	22
II.1. APPROCHE DESCENDANTE	22
II.1.1. METHODE D'ORACLE EDUCATION	22
II.1.2. UNIVERSITE DE BOLOGNE	25
II.1.3. CONCLUSION SUR LES DEUX METHODES.	27
II.2. APPROCHE ASCENDANTE	28
II.2.1. DEFINITION DES OBJETS MULTIDIMENSIONNELS	28
II.3. APPROCHE PARALLELE	31
II.4. OPTIMISATION DES BASES MULTIDIMENSIONNELLES	33
II.4.1. VOLUMETRIE	33
II.4.2. PERFORMANCES	35
CONCLUSION	36
CHAPITRE III. REPRESENTATION DES METADONNEES D'UN SYSTEME DECISIONNEL	37
III.1. DEFINITION DES METADONNEES	37
III.2. PROBLEMATIQUE DE LA REPRESENTATION DES METADONNEES	38
III.3. NOTRE CHOIX : UNE BDM POUR LES METADONNEES D'ACCES	38
III.4. METADONNEES MULTIDIMENSIONNELLES	39
III.5. UNE BDM POUR LES METADONNEES	39
III.6. UNE BDM POUR LA GESTION DES PROFILS	40

III.6.1. IMPLANTATION AVEC ORACLE EXPRESS	42
III.6.2. CONCLUSION	45
CHAPITRE IV. APPLICATION EN MARKETING BANCAIRE	46
IV.1. LA BANQUE SNVB	46
IV.1.1. LE DEPARTEMENT DES ETUDES INFORMATIQUES (DEI)	46
IV.1.2. LE DOMAINE SYSTEME D'INFORMATION COMMERCIAL (SIC)	47
IV.1.3. ARCHITECTURE INFORMATIQUE DE LA SNVB	47
IV.2. LE PROJET SIGMA	47
IV.3. ARCHITECTURE DU SYSTEME	49
IV.4. EXEMPLE D'INTERFACES UTILISATEURS	52
IV.5. UN EXEMPLE D'APPLICATION : LA GESTION DES ENCOURS	54
IV.6. EVOLUTIONS DE SIGMA	60
CONCLUSION ET PERSPECTIVES	61
TABLE DES ILLUSTRATIONS	62
BIBLIOGRAPHIE	63

Présentation du contexte du DRT

Le travail présenté dans ce rapport se place dans le cadre d'un projet de refonte du système d'information décisionnel de la banque SNVB. Ce projet avait pour objectif la création d'un nouveau système d'information décisionnel permettant de fournir de l'information de pilotage dans tout le réseau de la banque.

Dans un premier temps, le travail de recherche dans le cadre du Diplôme de Recherche Technologique a porté sur l'aspect veille technologique afin de déterminer l'état de la recherche, universitaire et privée dans le domaine des entrepôts de données et des systèmes de pilotage. Ces travaux ont permis de mettre à jour trois thèmes principaux de recherches : Architecture d'un système décisionnel, modélisation et méthodologie, Datamining et découverte de connaissances.

Dans un deuxième temps, nous nous sommes consacrés principalement au projet de la banque qui s'est déroulé en trois étapes qui correspondent aux trois thèmes de recherches identifiés précédemment : la définition d'une architecture technique et fonctionnelle, la modélisation d'un entrepôt et de ses composants puis la restitution et l'analyse des données.

Une première partie de ce mémoire porte sur les différents types d'architecture possibles pour un système d'informations décisionnel et de là propose et justifie les choix effectués à la SNVB.

Une deuxième partie porte sur la définition d'une méthode de conception d'un système décisionnel : proposition d'un modèle, d'une démarche pour la création du système, d'un langage graphique pour le représenter et d'outils pour l'administrer.

Nous fournissons des éléments de réponse pour trois points particuliers :

- Un modèle général pour les Bases de Données Multidimensionnelle avec un langage graphique
- Des heuristiques pour la démarche de construction
- Des solutions originales pour les métadonnées et la gestion des profils des utilisateurs.

L'aspect Datamining n'a pas été abordé car les travaux de ce type viennent juste de commencer dans la banque, en effet le Datamining n'était envisageable qu'après la mise en place complète de l'architecture globale du système de pilotage. On pourra noter que les nécessités propres à ces activités ont bien été prises en compte lors de la définition de l'architecture du système.

Chapitre I. Architecture et Concepts des Systèmes Décisionnels

Introduction

L'informatique décisionnelle s'est grandement développée ces dernières années. Elle s'est aussi grandement diversifiée.

Aux systèmes décisionnels monolithiques destinés uniquement aux analystes ont succédé les systèmes d'informations stratégiques pour fournir aux cadres et décideurs une information précise, pertinente et rapide. Enfin sont apparus les systèmes de pilotage dont le but est d'irriguer toute l'entreprise avec une information claire et transparente, déjà préparée mais laissant néanmoins aux analystes la possibilité de la retraiter (ces activités de retraitement sont aussi appelées activités OLAP pour On Line Analytical Processing).

On assiste là à la démocratisation de l'information dans les entreprises. L'accès à l'information n'est plus conditionnée par une position hiérarchique mais par des notions transversales de métier et de compétences.

On arrive alors à faire cohabiter dans un même système les utilisateurs finaux qui ont besoin de tableaux de bord pré-construits facilitant leur gestion quotidienne, les décideurs qui ont besoin d'informations plus agrégées leur permettant d'avoir en un coup d'œil une vue globale de l'activité, ainsi que les analystes qui ont accès à toutes les informations et dont les possibilités d'analyse sont décuplées, car elles ne sont plus limitées par leurs outils mais guidées par leur imagination.

Nous verrons ici comment les concepts liés à l'informatique décisionnelle ont évolué en liaison avec les évolutions techniques et structurelles dans les entreprises. De nombreuses études ont été menées mais il s'agit ici d'effectuer une synthèse des concepts généraux communs aux systèmes décisionnels quelle que soit la forme qu'ils prennent dans l'entreprise.

I.1. Types d'architecture pour les Systèmes Décisionnels

Si les caractéristiques et les exigences d'un système décisionnel sont connues depuis longtemps, la manière de les mettre en œuvre a grandement évolué, il s'agit ici de faire un bref historique des différentes formes qu'ont pu prendre les systèmes décisionnels en s'appuyant sur une vue systémique et sur les exigences d'un tel système.

I.1.1. Vue externe du Système Décisionnel

Pour n'importe quelle entreprise, les données représentent un capital stratégique très important, mais ce capital ne devient réel qu'en transformant ces données en informations utilisables. Les entreprises cherchent depuis maintenant plus de 20 ans des solutions pour construire leur *système décisionnel*.

Un système décisionnel doit répondre à cinq exigences :

- ❶ **Assurer l'accès aux informations**, de manière facile et directe
- ❷ **Garantir la qualité des informations**, grâce à la vérification du contenu sémantique des données et à leur cohérence globale.
- ❸ **Découper l'information** en petites unités d'information pour permettre de reconstituer toutes les mesures possibles de l'activité
- ❹ Fournir des **outils de requêtes et d'analyse** pour rendre le plus indépendant possible les analystes

⑤ Diffuser les données de manière fiable et utilisable pour tous les utilisateurs

Quelles que soient la technologie utilisée et l'architecture choisie, on peut donc représenter un Système Décisionnel comme un système répondant à ces cinq exigences réalisant l'interface entre le système de production qui gère et construit les données et les utilisateurs qui les utilisent et les analysent.

La Figure 1 ci-dessous donne la structure classique et théorique d'un système décisionnel. Les 5 points précédents y sont repris. Nous reviendrons par la suite sur chacun des composants de ce système.

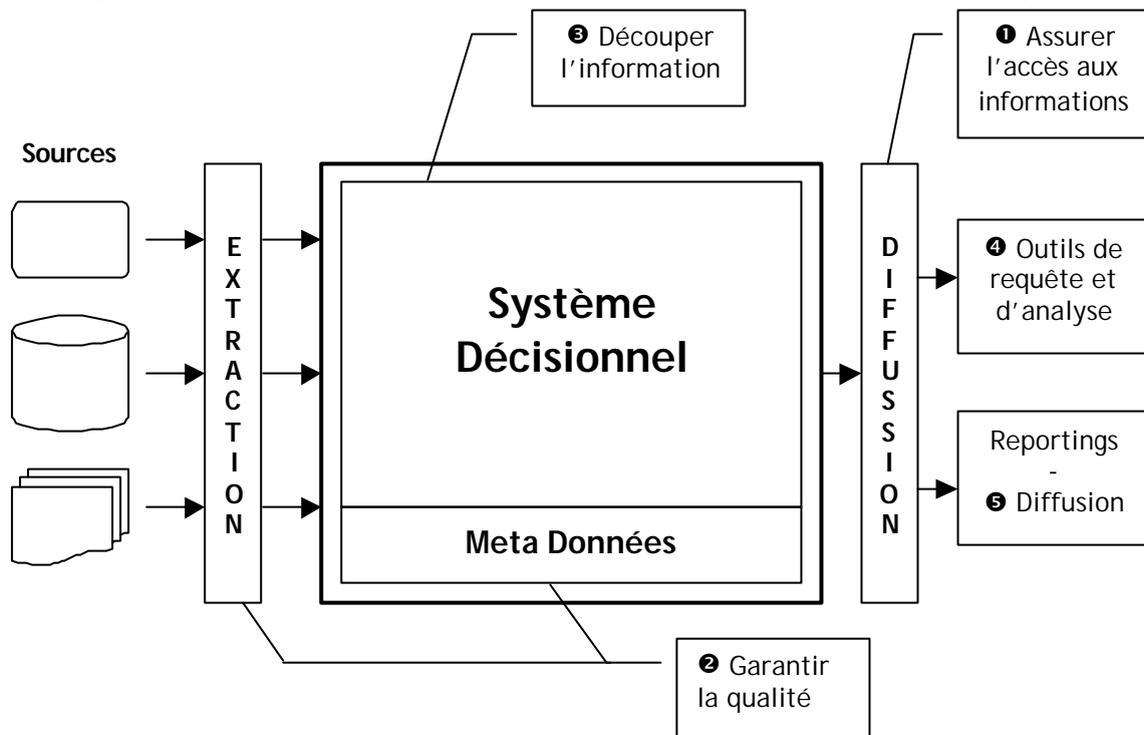


Figure 1. Structure d'un système décisionnel

I.1.2. De l'infocentre au DataWarehouse

1.1.1.1 Petit Historique

Pour répondre aux exigences que nous venons d'énoncer, diverses solutions ont été proposées et implantées dans les entreprises :

Dès les années 80, la première solution, dénommée **Information Center**, consistait à dupliquer les bases de production de manière périodique vers une machine dédiée aux analystes. Mais cette méthode ne s'est réellement développée qu'à partir du développement des SGBD relationnels et du langage SQL.

	Avantages
	Les analystes ont accès à toutes les données disponibles. Ils n'ont pas de contrainte dans la manière dont ils les utilisent car le système est indépendant.
	Inconvénients

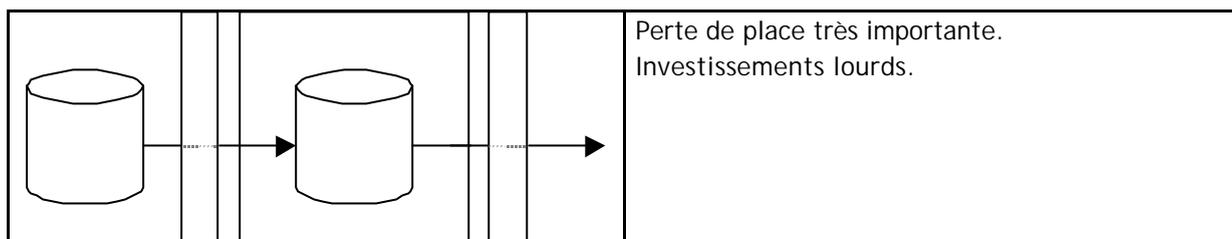


Figure 2. Information Center

Avec l'apparition de ces SGBD Relationnels, de nouveaux outils facilitant la construction des requêtes SQL sont aussi apparus. Ces outils ont permis aux analystes d'accéder directement aux bases de production pour retrouver exactement la donnée qu'ils voulaient, au moment où ils la voulaient. Les utilisateurs n'ont alors plus voulu choisir quelles données dupliquer et avec quelle fréquence.

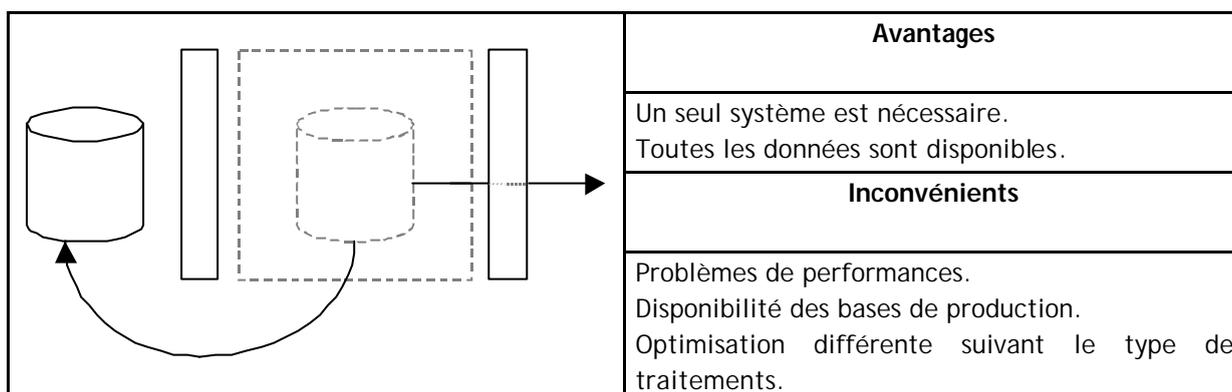


Figure 3. Infocentre

Les analyses ont donc de nouveau été faites sur les bases de production avec en conséquence d'énormes problèmes de performances et de disponibilité. En effet un utilisateur pouvait dès lors, par une seule requête mal construite ou non optimisée, bloquer tout le système pendant des heures, bloquant également les applications de production.

Pour répondre aux demandes en informations, les entreprises ont commencé à investir dans la construction de bases de données séparées des bases de production, pour centraliser et fiabiliser l'accès aux données. Ce concept d'une base de donnée séparée, conçue uniquement pour le besoin des analystes a pris le nom de **Information Warehouse** pour IBM ou de **Data Warehouse** pour Bill Inmon.

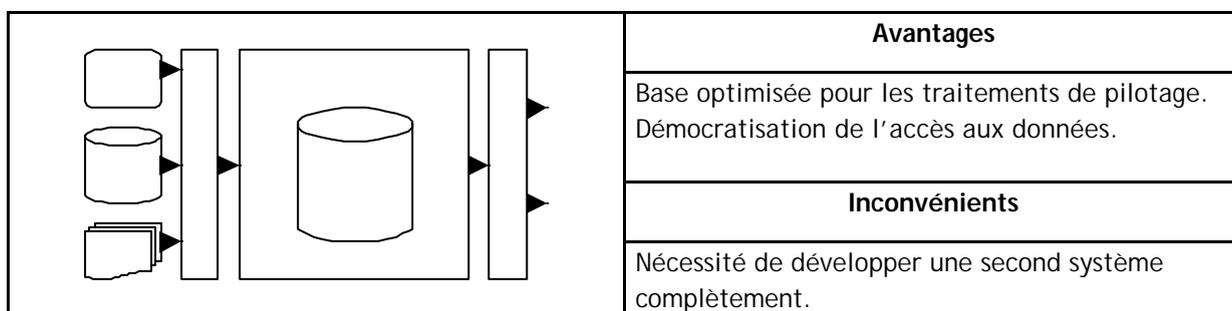


Figure 4. DataWarehouse

La fondation théorique de la conception des entrepôts de données peut être située en 1991 quand Bill Inmon publie son livre *Building a Data Warehouse*¹ où il spécifie les caractéristiques d'un entrepôt de données et la manière d'en concevoir un.

Les Entrepôts de Données constituent la première solution répondant pleinement à la cinquième exigence des Systèmes Décisionnels : Diffuser les données de manière fiable et utilisable pour tous les utilisateurs. C'est ce qui va permettre la démocratisation de l'accès aux données en permettant à toutes les populations d'accéder aux données sans l'aide d'un informaticien. De plus, les systèmes décisionnels vont évoluer pour devenir de véritables *applications décisionnelles* à destination non plus seulement des analystes mais de toutes les populations des entreprises.

En fait, on peut constater que la vision des Systèmes Décisionnels a évolué en même temps que la vision de la structure des entreprises et de leur stratégie. De plus en plus, on s'est rendu compte que la ressource importante qui constituait un avantage stratégique n'étant pas l'Information mais bien la capacité de traitement de l'Information ; Il a donc fallu adapter la structure des entreprises pour tenir compte de cette réalité en offrant une place de plus en plus importante à l'Informatique décisionnelle et en rapprochant l'Informatique de la Direction Générale (un historique de cette évolution peut être trouvée dans l'article de H. Tardieu 'Système d'Information - Stratégique et Système - d'Informations Stratégiques'²).

1.1.1.2 Typologie des entrepôts de données

En 1991, quand Bill Inmon décrit le concept des entrepôts de données, il s'agit d'une base relationnelle conçue pour répondre et optimiser les accès aux données pour l'analyse et le pilotage.

Ce type d'architecture aussi appelée architecture **ROLAP** (pour Relationnal-OLAP) est le choix le plus évident, car le plus souvent, les données proviennent déjà d'un système relationnel et ces technologies sont de loin les mieux maîtrisées dans les entreprises. Dans la plupart des cas, les données sont stockées selon de nouvelles formes de structures que nous présentons dans le paragraphe suivant. Une base normalisée ne serait pas appropriée en raison principalement du besoin de performances. Souvent, on doit même faire appel à des tables stockants les agrégats pour garantir les temps d'accès.

Rapidement, de nouvelles solutions apparaissent utilisant les possibilités des bases de données multidimensionnelles. Les concepts sous-jacents à ces bases de données existent depuis la fin des années 70, mais n'avaient jamais vraiment été utilisés à grande échelle car leur mise en oeuvre nécessite des machines très performantes ainsi qu'énormément de place sur disque. Dans les années 1990, l'évolution technique permet de les utiliser, et leurs caractéristiques en font l'outil idéal pour les systèmes décisionnels.

	Relationnel	Multidimensionnel
Avantages	<ul style="list-style-type: none"> ▪ Maîtrise des SGBD relationnels ▪ Stocke et gère de grandes quantités de données 	<ul style="list-style-type: none"> ▪ Structure de données adaptée pour les analyses ▪ Temps de réponse ▪ Lisibilité du modèle
Inconvénients	<ul style="list-style-type: none"> ▪ Peu adapté aux besoins des analyses ▪ Technicité du modèle 	<ul style="list-style-type: none"> ▪ On ne peut pas gérer dans une seule base de grandes quantités de données

¹ Bill Inmon - Building a Data Warehouse - 1991

² Hubert TARDIEU - Système d'Information - Stratégique et Système - d'Informations Stratégiques

Jusqu'en 1998, les deux approches s'affrontent et tous les concepteurs d'entrepôts de données doivent faire un choix entre l'une ou l'autre. Aujourd'hui, de plus en plus d'outils proposent des architectures hybrides permettant de profiter des avantages des deux mondes pour en tirer le meilleur parti, gérer d'importants volumes de données dans un système relationnel, gérer des informations à haute valeur ajoutée et en garantir un accès rapide et facile dans un système multidimensionnel. Le paragraphe 1.2 est consacré à la présentation des concepts des bases de données multidimensionnelles.

I.1.3. Problèmes particuliers

Les problèmes spécifiques rencontrés lors de la construction des entrepôts de données sont relatifs à leurs particularités que nous avons exposées au paragraphe précédent I.1.1. Ces problèmes concernent soit l'architecture de l'entrepôt, soit la qualité des données ou encore les métadonnées.

Architecture relationnelle, multidimensionnelle ou hybride

En premier lieu, un système décisionnel se doit d'assurer l'accès aux informations de manière facile et directe. Cela pose des problèmes de performances, que ce soit dans les architectures relationnelles ou multidimensionnelles.

Les méthodes de conception de bases de données relationnelles actuelles sont plus tournées vers une optimisation du volume que des performances, les entrepôts de données peinent parfois lorsqu'il s'agit d'effectuer des analyses complexes ou lorsque le nombre d'accès devient trop important.

Les bases de données multidimensionnelles souffrent elles du mal inverse, elles sont très performantes, notamment lorsqu'il s'agit d'analyses complexes, mais elles ne permettent pas de gérer des volumes de données trop importants.

La solution proviendra sûrement des architectures hybrides combinant relationnel et multidimensionnel mais l'interface entre les deux mondes reste encore délicate à mettre en œuvre. On peut d'ailleurs d'ores et déjà trouver dans la littérature des chercheurs préconisant ce type d'architecture³.

De plus, on commence à se rendre compte que les modèles classiques de conception comme le modèle Entités/Association ne conviennent plus pour la conception des entrepôts de données⁴ et qu'il faut donc développer de nouveaux modèles^{5 6}.

Qualité des données

Le deuxième point qui jusqu'à maintenant n'a été que partiellement étudié est la problème du maintien de la cohérence et du contrôle de la qualité des données dans les systèmes d'informations complexes et hétérogènes des entreprises.

Cette garantie passe par une mesure de la qualité des données et par des méthodes de retraitement permettant d'augmenter cette qualité. Si ces méthodes fonctionnent bien au niveau global et sur de grandes quantités de données quantitatives, il reste à généraliser des méthodes permettant d'analyser plus finement les données et de mieux contrôler la qualité des données qualitatives.

³ Judy Rawls, Multitiered Data Warehouses, 1998

⁴ Ralph Kimball, Is ER Modeling Hazardous to DSS, Data Warehouse Architect Octobre 1995

⁵ Neil Raden, Modeling A DataWarehouse, CMP Janvier 1996

⁶ Ralph Kimball, A Dimensional Modeling Manifesto, Août 1997

Métadonnées

Enfin, le dernier problème vient de la nécessité de rendre l'accès aux informations de plus en plus facile pour les analystes et de démocratiser cet accès pour toutes les populations d'une entreprise. Cela passe par la gestion d'un système de métadonnées, fournissant aux utilisateurs une description complète de toutes les données gérées par le système d'information, ainsi que d'outils permettant de s'affranchir des difficultés techniques pour se recentrer sur des notions de métier où l'informatique n'est plus que l'outil qui fournit les données et le système d'information le canal de diffusion, dans un sens comme dans l'autre. La grande difficulté de ce système de méta données est de décrire le système dans un langage assez technique pour permettre aux concepteurs de décrire précisément les données mais assez proche des utilisateurs pour qu'ils puissent le maîtriser.

De nouvelles méthodes pour intégrer efficacement les métadonnées au sein des systèmes décisionnels sont actuellement en cours de développement, la tendance est de diviser les métadonnées en deux catégories : celles servant à contrôler l'origine des données (la traçabilité) et celle décrivant leur sémantique, les premières étant plutôt destinées aux développeurs et les deuxièmes aux analystes⁷. Dans ce cas de figure, le problème du langage qui décrit les métadonnées devient un problème de liaison entre les deux systèmes.

Les travaux présentés dans ce mémoire concernent particulièrement les problèmes de modélisation des bases de données multidimensionnelles et de gestion des métadonnées.

Il s'agit en fait de proposer d'abord un modèle permettant de décrire les bases de données multidimensionnelles, puis à l'aide de ce modèle de proposer une méthodologie de conception des bases de données multidimensionnelles. Cette méthodologie sera d'abord générale pour s'appliquer ensuite aux spécificités du moteur de gestion de bases de données multidimensionnelles Oracle Express.

Enfin, on montrera que la gestion des métadonnées d'un système décisionnel peut être faite dans une base de données multidimensionnelle, garantissant en cela une plus grande cohérence au système.

I.2. Concepts des Bases de Données Multidimensionnelles

I.2.1. Mesures, Dimensions, Relations, Formules

Les nouvelles bases de données multidimensionnelles n'ayant pas encore fait l'objet d'études formelles ni de normalisation présentent des variantes dans les différents logiciels qui les implémentent et chacun d'eux utilise des termes parfois différents pour désigner les mêmes concepts. Dans ce paragraphe, nous proposons un méta-modèle général permettant d'appréhender les concepts qui leur sont communs

Nous décrivons ici les concepts des bases de données multidimensionnelles sous la forme de type d'objets permettant de construire le méta-modèle de ce nouveau type de bases de données. Ce méta-modèle sera dans un premier temps représenté en utilisant la notation UML puis on proposera de nouvelles notations purement multidimensionnelles plus adaptées, enfin on décrira le méta-modèle selon les notations de ce modèle multidimensionnel (I.2.2).

Une base de données multidimensionnelle comprend trois types d'objets : les dimensions, les mesures et les relations que nous définissons successivement.

⁷ Colin While, Managing Distributed Data Warehouse Meta Data, DM Review Magazine Février 1999

Dimension

Une dimension est un axe d'analyse des informations de la BDM.

Exemple :

- Le temps
- Le lieu
- Le type de produit

Une dimension contient toujours des valeurs distinctes. Les dimensions transforment les données brutes stockées en informations.

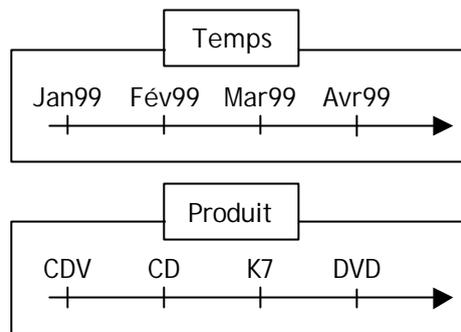


Figure 5. Dimension

Mesure

Une mesure est une grandeur caractérisant le domaine de la BDM défini par rapport aux dimensions.

On associe toujours un type à une mesure (Texte, Entier, Décimal...)

Ex. La quantité vendue est définie par rapport aux dimensions Date, Lieu et Produit

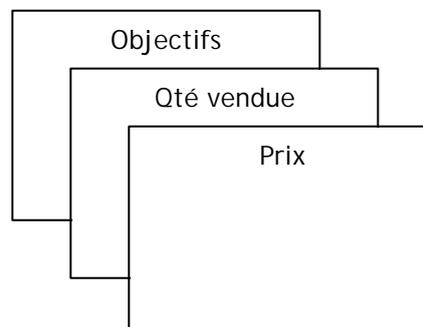


Figure 6. Mesures

Lien entre Dimensions et Mesures : notion d'hypercube

Puis si l'on considère une BDM comme composée d'objets,

- Les mesures sont les objets contenant des données, ces données pouvant être quantitatives ou qualitatives.
- Les dimensions sont des objets contenant des valeurs qui organisent et répertorient les données stockées dans les mesures. Elles permettent de sélectionner et d'utiliser des sous-ensembles de données spécifiques.

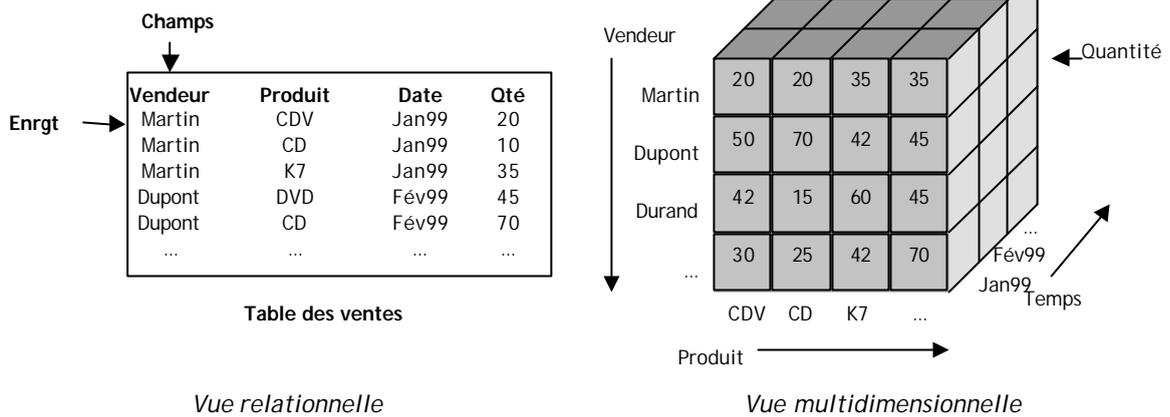
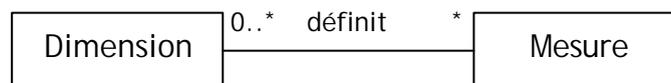


Figure 7. Vue relationnelle / Vue Multidimensionnelle

Un hypercube est donc une structure permettant de croiser des dimensions pour stocker des variables. La forme la plus simple d'hypercube correspond à une liste de mesures, chacune définie sur toutes les dimensions. Rapidement, les bases de données multidimensionnelles ont su gérer des structures plus complexes où les mesures peuvent être définies sur un sous-ensemble des dimensions. Cette extension de la notion d'hypercube est parfois appelée Multicube⁸. Ici, quand nous parlerons d'hypercubes, il s'agira en fait toujours de multicubes.



Une dimension définit plusieurs mesures.
 Une mesure peut être définie par plusieurs dimensions.

Figure 8. Schéma Dimension x Mesure

Extension de la notion de mesure

Cette notion de multicube permet d'étendre la notion de mesure pour inclure la notion de définition par rapport aux dimensions du cube.

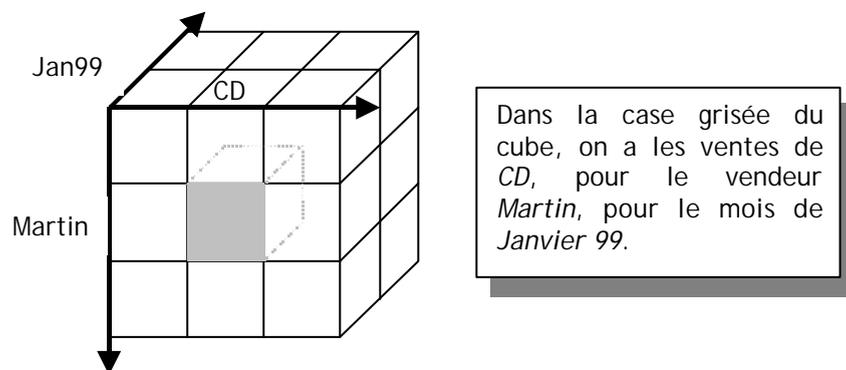


Figure 9. Notion de mesure dans un hypercube

Les prix sont définis uniquement par rapport aux dimensions Temps et Produit

⁸ Nigel Pendse, Multidimensionnal data structures, Janvier 1999

Par extension, on introduit les notions de constante et de propriétés.
 Une mesure qui n'est définie par aucune dimension est en fait une constante.

$\pi = 3.14$

Figure 10. Constante

Une mesure dimensionnée par une seule dimension est une propriété rattachée à chaque position de la dimension.

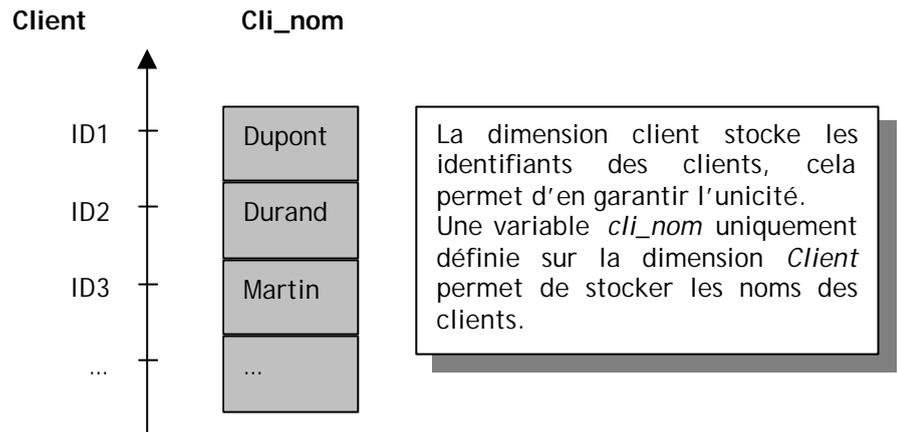


Figure 11. Propriété

Formules & variables

On peut dissocier les mesures en deux grandes catégories : les variables et les formules.

- Les variables stockent des valeurs.
- Les formules ne stockent pas de données, mais une équation. Lors de l'accès à ce type de mesure, l'équation est exécutée et les données calculées sont stockées temporairement et sont disponibles à des fins d'analyse.

Dans le méta-modèle des BDM, Formule et Variable sont deux sous-types de l'objet Mesure.

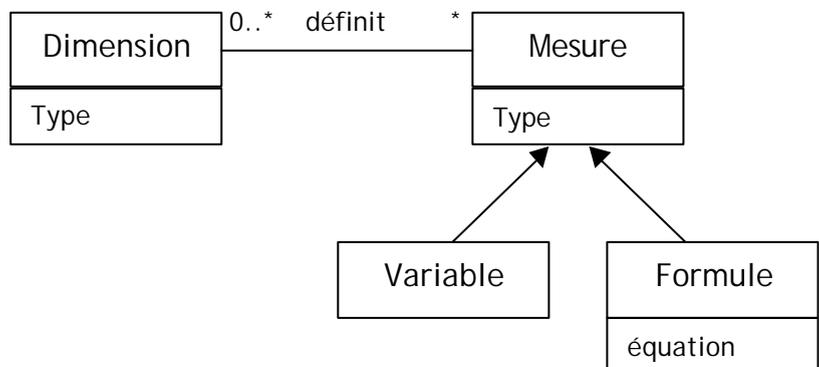


Figure 12. Méta-modèle des BDM

Relations

Les relations constituent une extension de la notion de variable. Une variable renvoie une valeur d'un type classique : booléen, entier, décimal, texte... Une relation est définie sur une ou plusieurs dimensions et renvoie une position d'une dimension, c'est à dire qu'en fait la case du cube concerné ne contient pas une valeur mais un lien (pointeur) vers une position d'une dimension.

Pour décrire le concept de relation, on peut utiliser la représentation suivante :

$$R(d_1, d_2, \dots, d_n) \rightarrow p_i$$

- R étant la relation considérée
- d_1, d_2, d_n étant chacun une position des dimensions D_1, D_2 et D_n
- p_i étant une position d'une dimension P

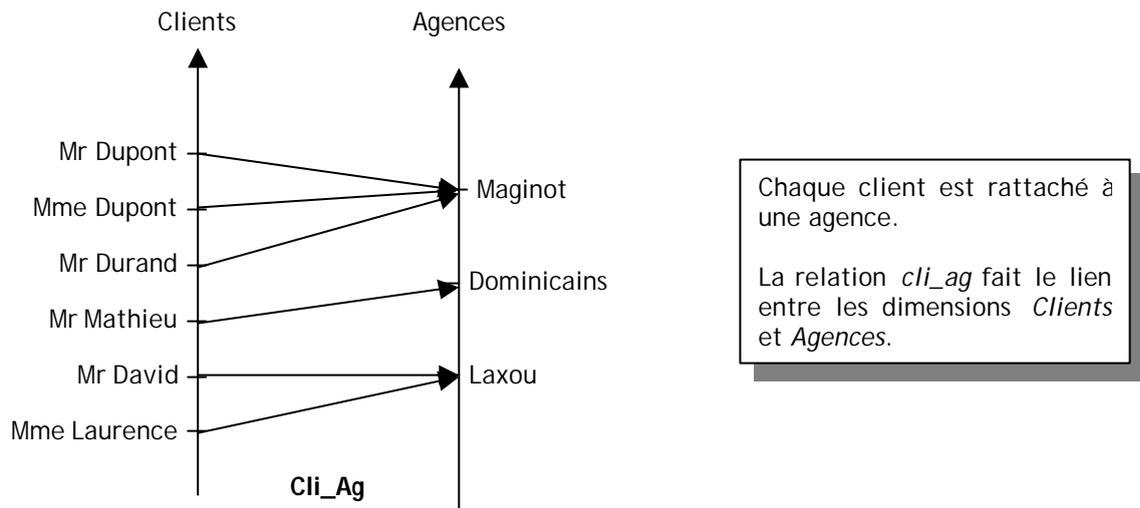


Figure 13. Relation

Les relations permettent notamment de représenter la notion de hiérarchie. Les hiérarchies sont en effet des relations père-fils, c'est à dire des relations entre une dimension et elle-même, qui à chaque position de dimension (fils) associe une autre dimension (son père).

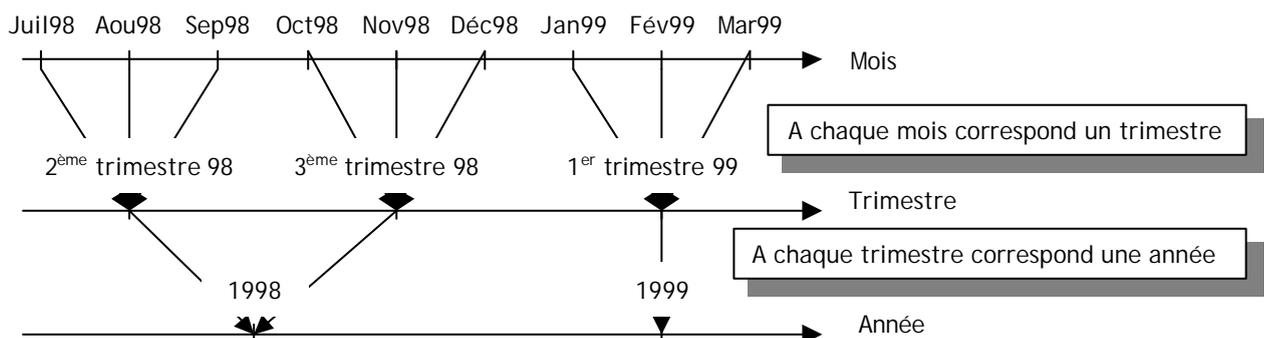


Figure 14. Hiérarchie Père/Fils

Les relations peuvent servir aussi comme outil de sélection. Elles ont alors un double avantage :

- Elles restreignent les valeurs possibles à un nombre fini, facilitent les analyses ultérieures
- Elles offrent une manière de sélectionner plus simple

Méta-modèle des BDM

On aboutit alors au méta modèle suivant, décrit selon la notation UML.

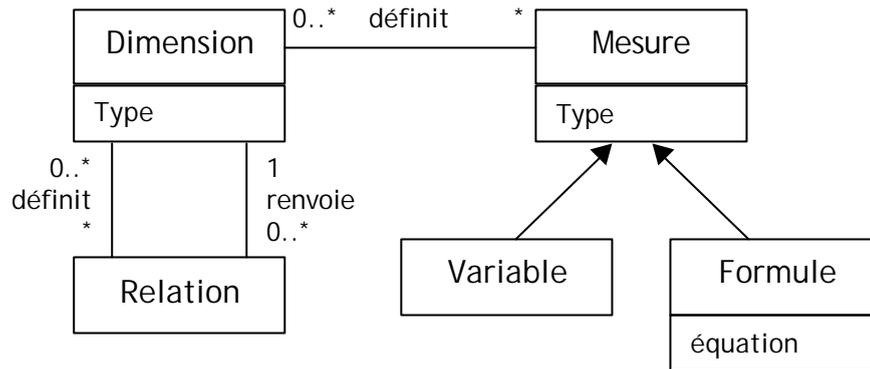


Figure 15. Méta-modèle des BDM

I.2.2. Représentation graphiques des schémas des BDM

A partir du méta-modèle précédent, nous proposons un mode de représentation graphique des schémas des bases de données multidimensionnelles qui facilite leur compréhension par les analystes.

Les dimensions sont les briques de base permettant de décrire le modèle multidimensionnel. Ce sont donc elles qui sont au centre de ce modèle. Une dimension peut contenir des propriétés, chaque propriété étant une variable définie uniquement par cette dimension.



Figure 16. Représentation graphique d'une dimension

La représentation des variables dans le modèle multidimensionnel est la suivante :

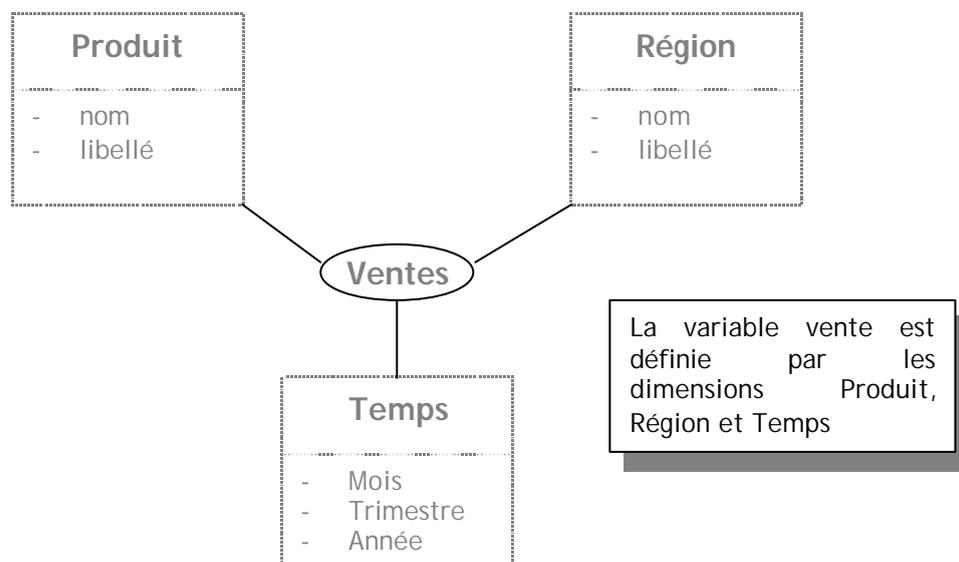


Figure 17. Représentation graphique d'une variable

Les formules ne diffèrent des variables que parce qu'elles sont calculées et non stockées, mais d'un point de vue purement fonctionnel, ce sont des objets très proches. Leurs représentations sont donc très voisines :

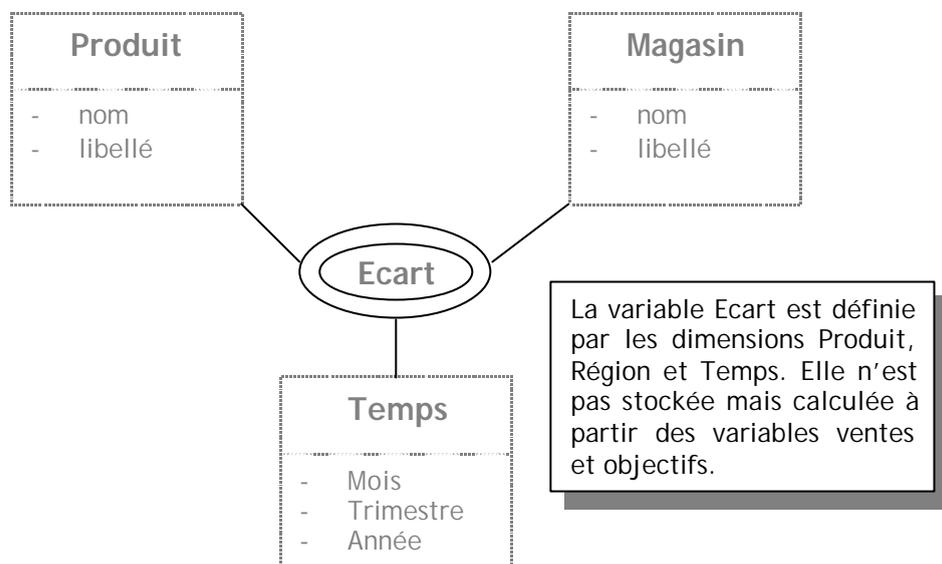
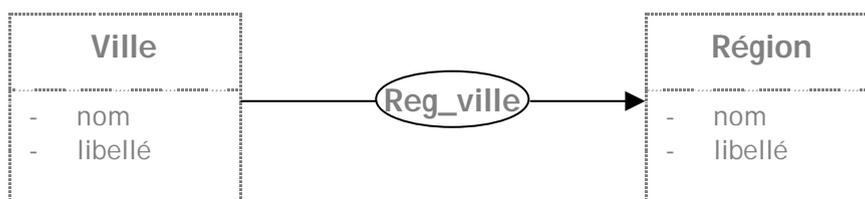
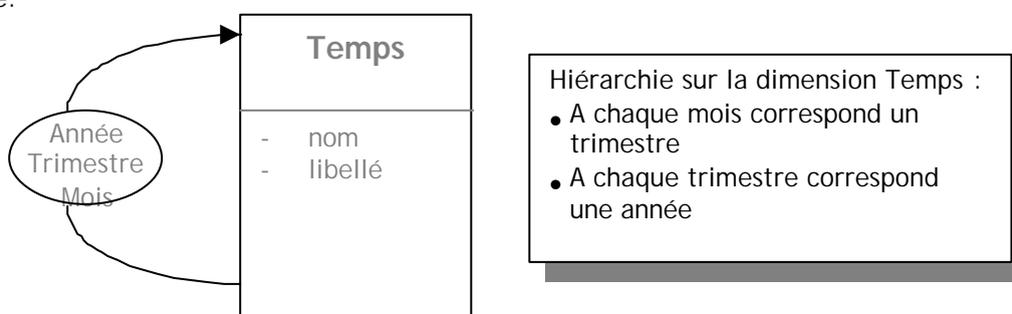


Figure 18. Représentation graphique d'une formule

Le formalisme de représentation d'une relation est également proche de celui d'une variable, à la différence près qu'une flèche indique la dimension la plus agrégée de la relation. La représentation qui suit présente une relation mono-dimensionnelle associant une région à chaque position de la dimension ville.



Les hiérarchies sont modélisées par l'objet suivant qui correspond traditionnellement à la relation réflexive.



Exemple de schéma d'une BDM

La Figure 19 présente un exemple de base de données multidimensionnelle représentée selon le formalisme proposé. C'est une base simple correspondant à une structure relationnelle en étoile, c'est

à dire qu'on y trouve au centre du schéma les différentes mesures enregistrées et tout autour les axes d'analyse.

De la même façon, on enregistre ici dans des variables les données, on ajoute des formules pour calculer certains indicateurs supplémentaires, et toutes ces mesures sont définies par les quatre dimensions que sont *Structure*, *Temps*, *Produit* et *Marché*.

Trois relations permettent de gérer les hiérarchies existantes sur les dimensions *Structure*, *Produit* et *Marché*.

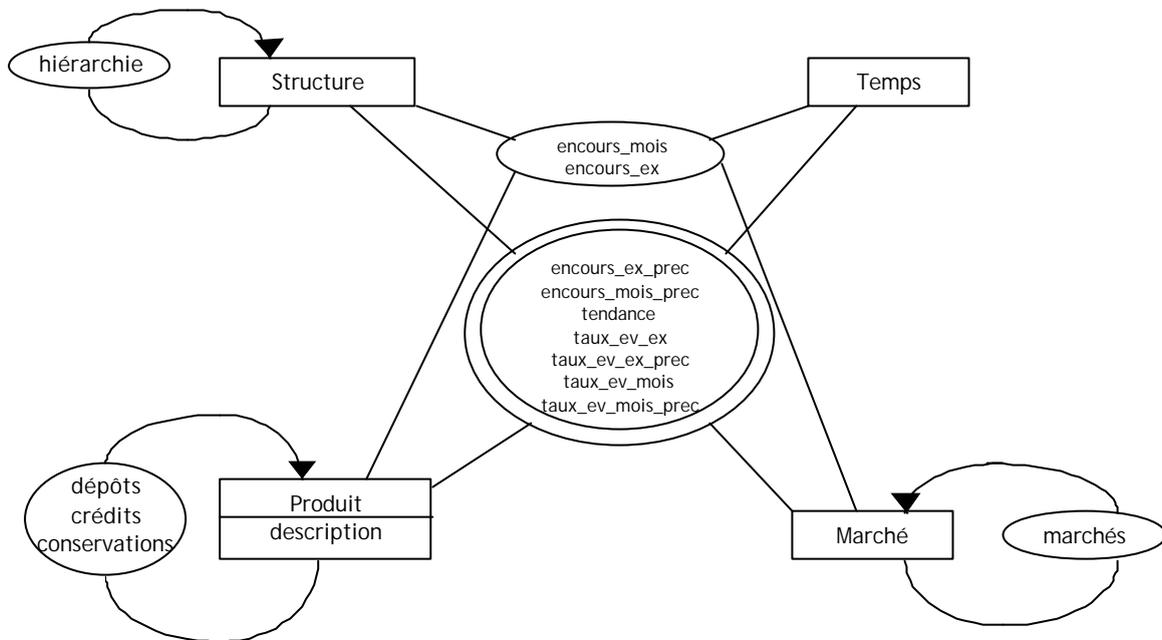


Figure 19. BDM Encours Dépôts Crédits

Modèle multidimensionnel

En appliquant le principe de la métacircularité qui consiste à décrire un modèle à l'aide de ses propres concepts, nous proposons le métamodèle suivant pour exprimer ce que contient toute base de données multidimensionnelle.

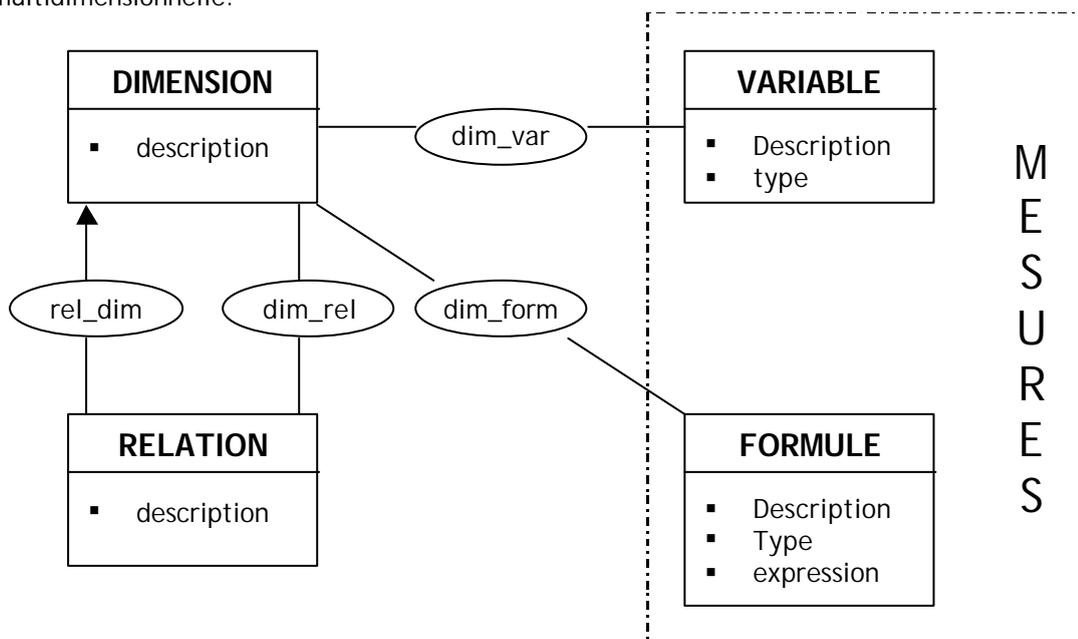


Figure 20. Modèle multidimensionnel

NB :

- Dimension, Variable, Relation et Formule sont des dimensions du méta schéma car ce sont eux qui permettent de structurer les données.
- dim_var, dim_form et dim_rel sont des variables permettant de gérer la dynamique du modèle. Elles relient les dimensions.
- Rel_dim est une relation qui permet de représenter la dynamique existant entre les dimensions et les relations.
- Les hiérarchies forment un sous ensemble des relations qui ne sont définies que par une seule dimension et qui renvoient une position vers cette même dimension (dim_rel et rel_dim pointent alors sur la même dimension et sur la même relation).
- De la même manière que dans les bases relationnelles on peut conserver des procédures stockées qui permettent d'exécuter des manipulations complexes, il existe souvent des objets semblables appelés programmes dans les bases de données multidimensionnelles. On peut considérer cette notion de programme comme un sous ensemble des formules. Un programme serait alors une formule qui ne serait définie par aucune dimension.

I.2.3. Concepts relatifs à Oracle Express

Le modèle de données multidimensionnel décrit peut être adapté selon le moteur de base de données multidimensionnelle que l'on utilise. En effet, chaque moteur a ses contraintes, nous allons ici adapter le modèle aux contraintes du moteur d'Oracle Express.

Ordre des dimensions dans une relation

La première contrainte vient de la définition des variables. Pour Oracle Express, l'ordre dans lequel on énonce les dimensions qui structurent une variable a une grande importance et va influencer grandement les performances. On doit toujours déclarer ces dimensions dans un ordre de complexité d'un point de vue hiérarchique (la hiérarchie possédant le plus grand nombre de niveaux ou le plus d'individus en premier). Pour traduire cela dans la phase de conception, on pourra porter un numéro d'ordre sur les arcs des variables.

Calculs d'agrégations et totalisations

Le fait de stocker toutes les données au niveau le plus fin dans une base multidimensionnelle permet de calculer des cumuls ou agrégats à différents niveaux pour répondre aux interrogations des analystes. Certains cumuls sont stockés dans la base multidimensionnelle, d'autres sont calculés à la demande, cela dépend de la structure de donnée choisie et du type de requête.

Lorsqu'une dimension est fortement structurée à l'aide d'une hiérarchie, Oracle Express se charge de calculer les agrégats dès que les données sont chargées et les enregistre dans la base.

Dans le schéma suivant, Oracle Express utilise la hiérarchie présente sur la dimension temps pour calculer les ventes du premier trimestre.

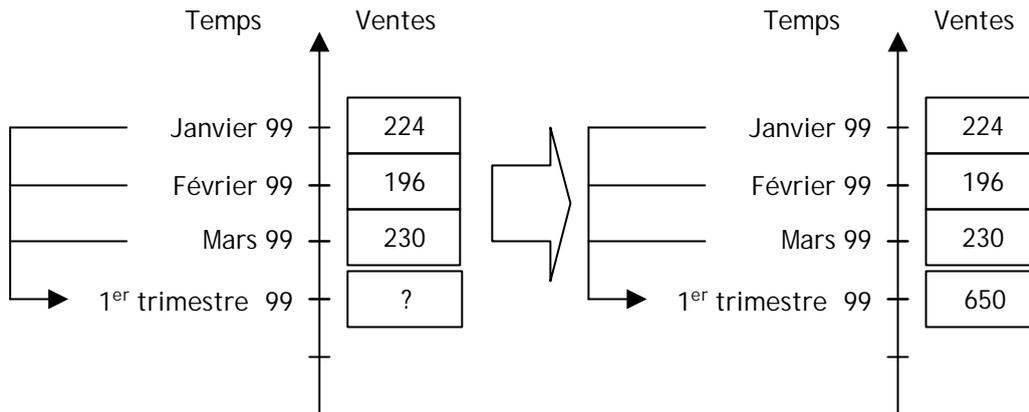
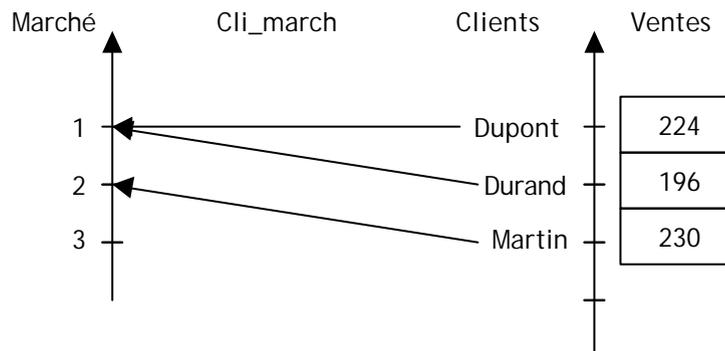


Figure 21. Utilisation d'une hiérarchie

Lorsque l'agrégat demandé est en fait le résultat d'une requête complexe faisant appel aux relations qui lient les dimensions entre elles, la totalisation est faite à la demande.

Dans l'exemple suivant, on a une dimension clients, une dimension marché et une relation cli_march qui à chaque client attribue un marché. La variable vente est définie sur la dimension clients. On peut alors demander un total basé sur la requête : total des ventes pour les clients du marché 1, Oracle Express se charge de ne récupérer que les ventes des clients pour lesquels la relation renvoie au marché 1.



Total des ventes pour les clients du marché 1 : $224 + 196 = 420$

Figure 22. Utilisation d'une relation

Le tableau ci-dessous résume les avantages et inconvénient permettant d'aiguiller le choix d'une hiérarchie ou d'une relation :

Relation

Avantages	Inconvénients
<ul style="list-style-type: none"> ▪ Structure moins complexe ▪ Possibilité de gérer des volumes plus importants (on ne stocke pas les données, donc on n'augmente pas la taille des variables) ▪ Sélection et totalisation possible grâce à la relation. 	<ul style="list-style-type: none"> ▪ Nécessite d'avoir une variable ou une formule pour chaque agrégation ▪ Impossibilité de visualiser simultanément les données détaillées et les données agrégées. ▪ Temps de calculs plus importants ▪ Pas de lecture de niveaux différents sur le même table

Hiérarchies

Avantages	Inconvénients
<ul style="list-style-type: none"> ▪ Possibilité de visualiser simultanément les données détaillées et les données agrégées : drill down / drill up ▪ Consolidation simple ▪ Temps d'accès plus rapide aux données (elles sont déjà calculées et stockées) ▪ Maintenance plus aisée : utilisation des objets et des relations liés aux hiérarchies. 	<ul style="list-style-type: none"> ▪ Lisibilité plus complexe à partir de 3 niveaux ▪ Place occupée plus importante ▪ Les totaux embarqués peuvent introduire des doubles totalisations dans les calculs ▪ La position de plus haut niveau si elle est commune, peut introduire de mauvaises interprétations. ▪ Les positions stockées ne sont pas de même nature

Conclusion

Nous venons de décrire de manière simple le méta-modèle de toute BDM et d'en proposer un mode de représentation graphique.

Ce modèle de représentation des bases de données multidimensionnelles fait pendant aux modèles dit en étoile ou en flocon pour les bases relationnelles⁹. Le concept d'hypercube et de multicubes serait alors à rapprocher du concept de tables de fait particularisées, c'est à dire qu'on définit des sous-ensembles des mesures qui sont définies de manière différentes (plusieurs tables de faits avec des colonnes différentes pour une base relationnelle, des mesures qui ne sont pas toutes définies par toutes les dimensions pour une base multidimensionnelle).

Ce modèle pouvant se décrire lui-même, on peut prouver qu'il est à même de décrire n'importe quelle base multidimensionnelle et peut donc être utilisé dans le paragraphe suivant comme base pour proposer une méthode de conception de bases de données multidimensionnelles.

⁹ Wilburt Juan La Bio, Dallan Quass, Brad Adelberg - Physical Database Design for DataWarehouses - Université de Stanford 1997

Chapitre II. Méthodes de conception des BDM

Introduction

Les bases de données multidimensionnelles reposant sur des concepts différents des bases de données relationnelles, il n'est pas possible d'utiliser pour les construire les méthodes associées aux bases de données relationnelles.

Cependant les BDM sont issues dans un système décisionnel d'une ou plusieurs BDR et les liens qui existent entre les données des unes et des autres sont très forts.

Selon que les BDR préexistent ou non aux BDM, on utilisera une approche descendante ou ascendante pour la construction du système décisionnel.

Dans la réalité de l'entreprise, le raisonnement des analystes et des concepteurs n'est pas linéaire, les deux approches s'autofluencent mutuellement et la conception de l'entrepôt relationnel et des bases multidimensionnelles se développe en parallèle dans un processus incrémental.

Nous présentons successivement ces 3 approches.

II.1. Approche descendante

Le plus souvent, lors du début du processus de conception d'une base multidimensionnelle, on possède déjà une base relationnelle contenant toutes les données utiles, soit que cette base existait au préalable en tant qu'entrepôt de données, soit qu'elle ait déjà été créée pour une précédente base multidimensionnelle.

C'est donc l'approche descendante qui est alors utilisée, consistant à concevoir la base multidimensionnelle à partir de la base relationnelle.

Dans la pratique, la conception d'une base multidimensionnelle se fait plutôt à partir d'une vue partielle orientée-métier de la base relationnelle.

Plusieurs recherches sont en cours pour permettre la transformation automatique ou semi-automatique d'un schéma relationnel, dédié à la production, en un schéma multidimensionnel, dédié à l'analyse et la navigation. De telles recherches permettent de profiter des modèles relationnels déjà en place et très répandus dans les entreprises pour élaborer un système décisionnel.

Nous étudierons ici deux méthodes, la première¹⁰ est tirée des cours de méthodologie de la société Oracle et permet de passer d'un modèle Merise à un modèle Oracle Express, la deuxième vient de l'université de Bologne et prend comme point de départ un schéma Entités/Associations.

II.1.1. Méthode d'Oracle Education

Le principe de cette méthode consiste à partir de la structure relationnelle de la base et à la dénormaliser, c'est à dire à modifier sa structure pour améliorer les performances aux dépens des autres contraintes¹¹.

La première étape consiste à modifier le schéma relationnel de la base pour supprimer toutes les relations d'ordre supérieur à quatre. L'ordre d'une relation correspond au nombre d'occurrences

¹⁰ Oracle Education, Cours méthodologie, 1998

¹¹ Ralph Kimball, Entrepôts de Données : Guide pratique du concepteur de DataWarehouse, Editions Wiley 1997

d'entités participant à une occurrence de l'association. Aussi toutes les relations d'ordre supérieur à 4 devront dans la mesure du possible être éliminées pour les raisons suivants :

- L'ergonomie : une relation d'ordre plus important devient vite illisible
- La maintenance : même si les moteurs multidimensionnelles deviennent de plus en plus puissants et permettent de définir les objets sur un grand nombre de dimension, il n'est pas raisonnable d'utiliser cette possibilité

L'élimination de ces associations d'ordre supérieur à 4 s'effectue de 3 manières :

- En diminuant le nombre d'entités participant à cette association en introduisant une propriété associée à cette association.
- En imbriquant un ou plusieurs individus participant à l'association (ceci est particulièrement valable dans le cas de présence de contraintes d'intégrité fonctionnelle ou de relations d'appartenance.
- En diminuant l'ordre des associations participant à l'association en question.

Il s'agit en fait dans cette transformation préalable de passer d'un modèle 3^{ème} forme normale à un modèle 1^{ère} forme normale. Cette modification est aussi appelée « dénormalisation » au sens de Codd.

Une fois le modèle dénormalisé, l'application de 6 règles simples permet de passer d'un modèle conceptuel Merise à un modèle multidimensionnel. Ces six règles permettent en fait de détecter les dimensions, leurs propriétés, les variables et formules et les relations. Le développeur devra néanmoins faire appel à son expérience pour choisir entre les relations et les hiérarchies.

Règle n°1 : Identifiant

Tout identifiant de chaque entité du modèle conceptuel devient une dimension.

Règle n°2 : Propriétés

Toutes les propriétés de chaque entité du modèle conceptuel de données en dehors de l'identifiant deviennent des variables mono-dimensionnées par cette dimension ou des relations.

Les propriétés non identifiantes d'une entité n'interviennent que dans la définition de cette entité et non dans le reste du modèle, aussi la traduction en variable mono-dimensionnée est la plus adéquate.

Une relation multidimensionnelle peut également être utilisée mais elle nécessite la maintenance d'une dimension supplémentaire (celle qui est associée au type de données).

Règle n°3 : Propriétés des associations

Les propriétés des associations deviennent des variables multidimensionnelles dont les dimensions sont les identifiants des entités constituant l'association.

Règle n°4 : cardinalité (1,1)

Les contraintes d'intégrité fonctionnelle entre 2 entités ou entre 1 entité et une association deviennent des relations ou des hiérarchies.

Règle n°5 : cardinalité (1,n)

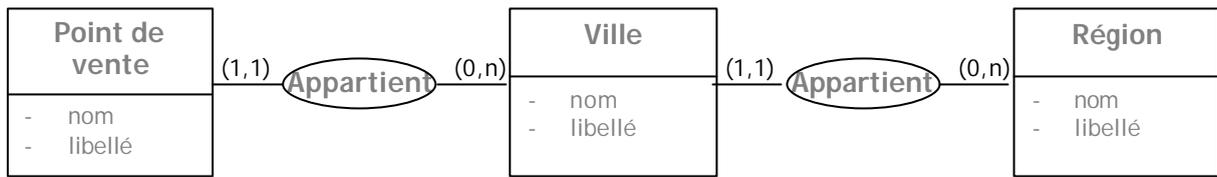
Une relation binaire dont au moins une des cardinalités maximum est égale à 1 devient une relation ou une hiérarchie.

Règle n°6 : cardinalité (n,n)

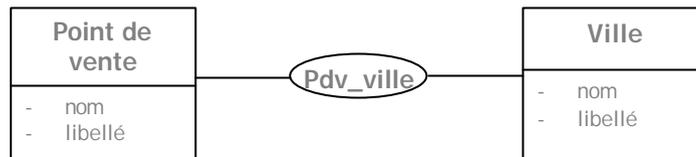
Toutes les autres relations deviennent des variables multidimensionnelles dimensionnées par les identifiants de chaque entité liée à la relation.

Hiérarchie ou relations ?

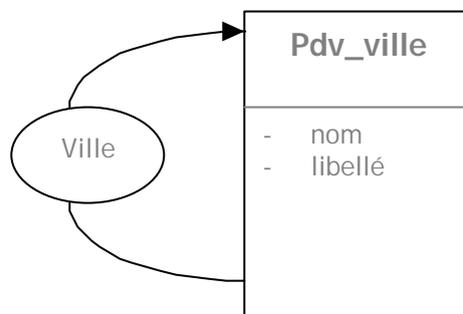
Hiérarchies et relations peuvent parfois être utilisées dans le même contexte. Un exemple type est le cas de l'entité Point de vente relié à l'entité Ville. Comment modéliser cette relation ?



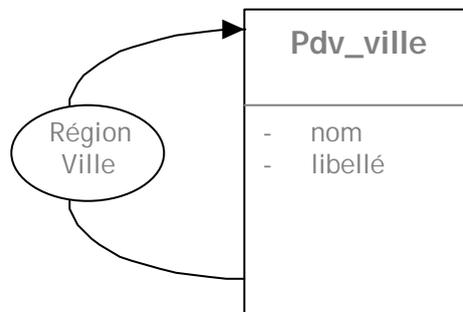
Pour la première relation, si on applique les règles n°4 et n°5, on peut obtenir soit une relation,



Soit une hiérarchie.



Si on inclut la seconde relation (Région), on aura la hiérarchie à 2 niveaux suivante :



On aboutit finalement à un modèle multidimensionnel utilisant la représentation décrite au paragraphe 1.2.2.

II.1.2. Université de Bologne

Les chercheurs Matteo Golforelli, Dario Maio et Stefano Rizzi de l'Université de Bologne proposent eux une méthode algorithmique en deux étapes pour passer d'un modèle entités/associations à une modèle multidimensionnel.

Cette méthode consiste également à détecter les dimensions, variables, formules et relations mais elle commence par définir les faits, c'est à dire les mesures du système.

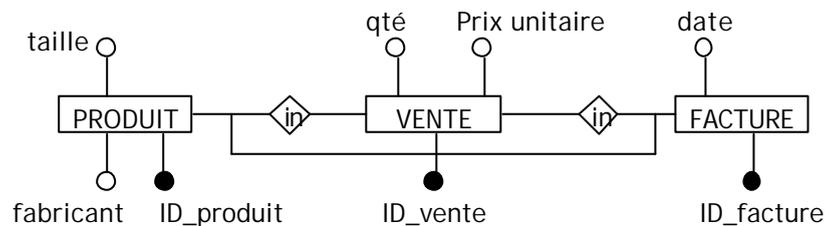
- 1^{ère} étape** Définir les faits
2^{ème} étape Pour chaque fait :
- Construire l'arbre des attributs
 - Elaguer l'arbre des attributs
 - Définir les dimensions
 - Définir les attributs
 - Définir les hiérarchies

La première étape consiste à trouver quels faits déterminent la dynamique du système, c'est à dire quelles mesures sont enregistrées par le système et justifient son existence ; on parle de dynamique car ce sont ces faits qui font évoluer et vivre le système. Un fait matérialise par une donnée un événement survenu dans l'entreprise, par exemple l'entité Vente représente le fait qu'une vente a été conclue pour un magasin, pour un certain produit.

Un fait sera généralement représenté dans un schéma relationnel soit par des entités, soit par des relations n-aires entre entités.

Un algorithme basé sur ces constatations permet de retrouver les faits.

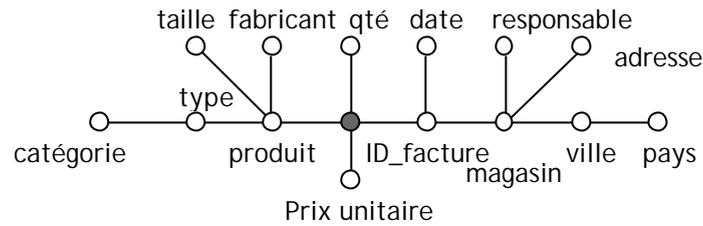
Chaque fait identifié dans le schéma relationnel devient racine d'un « schéma de faits ».



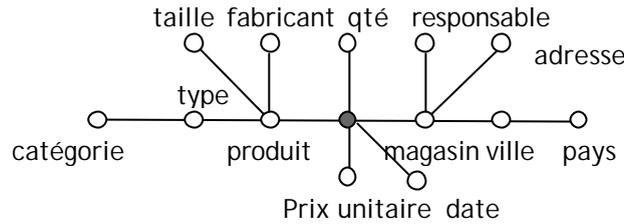
A partir de chacun des faits trouvés, on peut alors construire son arbre des attributs, définir les dimensions qui le structurent, les attributs qui le composent et ses hiérarchies. Cette deuxième étape permet également d'enrichir le modèle par des données externes qui ne sont pas stockées dans la base de données de départ (comme la hiérarchie ville>pays>continent rarement définie explicitement dans les SI).

NB. les entités sont représentées par les rectangles, les associations par les losanges, les propriétés sont représentées par les ronds, noir quand il s'agit d'une clé primaire, blanc sinon.

La même représentation est gardée par la suite pour montrer qu'on n'a plus virtuellement qu'une unique entité autour de laquelle on retrouve les faits enregistrés : la table de faits. Les attributs immédiatement reliés à rond noir central étant les faits à enregistrer dans la table, ceux plus loin les attributs de tables de dimension.



Ensuite on élague l'arbre des attributs, c'est à dire qu'on sélectionne en fait les attributs dont l'analyse peut être intéressante. Ici les problèmes de choix sont les mêmes que dans le cas d'une conception classique.



Les dimensions déterminent comment les faits peuvent être regroupés de manière significative pour l'analyse. Les dimensions doivent être choisies parmi les enfants directs de la racine. Leur choix est crucial pour le futur DataWarehouse car cela déterminera la granularité de chaque instance.

La plupart des schémas relationnels ne gèrent pas le temps mais sont des instantanés. Il faut donc toujours rajouter une dimension temps explicite même si elle n'apparaît jamais dans les arbres de faits.

Les faits doivent ensuite être définis. Ce sont typiquement des agrégats basés sur les attributs numériques des arbres de faits.

La dernière étape consiste à définir les hiérarchies. Les hiérarchies doivent être telles qu'on puisse construire un arbre hiérarchique par niveaux.

C'est donc une méthode semi-automatique qui est proposée, car on a toujours besoin d'un expert qui permet de faire le tri entre les données utiles pour l'analyse et les autres.

De plus, cette méthode ne permet que de partir d'un unique schéma relationnel, or la plupart du temps un DataWarehouse consiste à regrouper plusieurs sources, donc plusieurs schémas. Les auteurs proposent de répéter leur méthodologie pour chaque schéma puis de les regrouper, mais ils ne proposent pas de méthode pour ces regroupements.

II.1.3. Conclusion sur les deux méthodes.

La méthode d'Oracle Education commence par déterminer les dimensions, celle de l'université de Bologne commence par les mesures mais ces deux méthodes arrivent en fait à des résultats similaires. Elles sont toutes les deux basées sur un système de règles qui permet d'automatiser en partie ce passage d'un modèle relationnel à un modèle multidimensionnel mais quelque soit la méthode, on a toujours besoin de l'expertise humaine pour déterminer l'importance de chacune des informations et donc la pertinence de la conserver au sein d'un système décisionnel. De même pour structurer certaines informations comme les notions de hiérarchie, l'expertise humaine permettant de prévoir l'utilisation future de chacune des données reste le seul moyen de faire un choix.

On peut penser que dans les prochaines années, une méthode verra le jour mais il est certain que cette méthode ne pourra faire autrement que de laisser la place en dernier recours à l'expertise humaine dont elle ne pourra en fait que simplifier la tâche en automatisant les tâches simples et répétitives.

II.2.Approche ascendante

L'approche ascendante consiste à définir la structure de la base multidimensionnelle à partir des restitutions et analyses qu'on désire construire. Il s'agira ensuite de réaliser l'interface avec l'entrepôt de données afin de l'alimenter.

II.2.1.Définition des objets multidimensionnels

Pour bien définir les objets multidimensionnels à créer, il est important d'avoir une vue concrète des vues que l'on souhaite obtenir.

Concrètement, une vue peut être décomposée en plusieurs éléments :

Contexte d'affichage

		Colonne	
		données	
Ligne			

Les données sont les différentes informations qui doivent être présentées à l'intérieur du tableau, elles peuvent être qualitatives ou quantitatives.

Les axes d'analyses sont présentes dans trois endroits de la vue :

- En ligne,
- En colonne
- En page, c'est à dire en fait le contexte d'affichage.

Souvent, les utilisateurs visualisent bien les axes présentés en ligne et en colonne mais ils n'ont pas toujours le réflexe de considérer le contexte d'affichage comme un axe d'analyse. Cela tient au fait que jusqu'à maintenant les vues sont en fait des reportings imprimés. En passant d'un mode de présentation « papier » à une consultation « électronique » on permet de modifier bien plus facilement ce contexte d'affichage. Auparavant, pour cela, il fallait imprimer une page pour chaque analyse, maintenant il suffit de proposer à l'utilisateur un moyen de choisir son contexte.

	Marché	Mois	Structure
Tous marchés - Mai 99 - Vandoeuvre	Tous marchés ▼	Mai 99 ▼	VANDOEUVRE (Agence) ▼

Contexte sur papier

Contexte en mode intranet

Ainsi la compréhension de la représentation multidimensionnelle est facilitée pour les utilisateurs.

Dimensions

Chacun des critères d'analyse des données se traduit par une dimension.

Hiérarchies et Relations

Lorsque, pour un critère, il existe une notion de relation hiérarchique (une gamme de produit, des consolidations par mois, trimestre et année, une structure géographique), il faut déterminer si dans les restitutions, cette structure doit figurer sous forme de hiérarchie, c'est à dire représentée dans la même vue par une ligne et ses composants, ou bien sous forme d'un critère permettant de sélectionner un sous-ensemble de données. Dans le premier cas, on crée une hiérarchie pour la dimension, dans le second cas, on crée une relation.

Par exemple, si on considère l'ensemble des clients et un critère comme la catégorie CSP (Catégorie Socio-Professionnelle), on peut soit vouloir afficher dans un tableau un total général puis sa décomposition :

Total	
CSP1	
CSP2	
...	
CSPn	

Soit on veut construire plusieurs reportings, permettre plusieurs analyses avec la catégorie CSP comme critère, dans ce cas, on n'aura pas de position totale fixe mais on pourra plus facilement permettre aux utilisateurs de sélectionner certaines catégories CSP, etc, auquel cas on utilisera une relation.

Variables et Formules

Chaque donnée élémentaire utile aux analystes se traduit par une variable ou une formule dans la base multidimensionnelle.

La donnée peut être uniquement stockée et devient donc une variable ou bien elle peut être calculée à partir d'autres données et devient donc une formule. Parfois également, lorsqu'une donnée peut être calculée mais que le mode de calcul est trop complexe, on peut la précalculer et la représenter.

Il convient également de définir quelles dimensions structurent chacune des variables. Il y a en premier lieu les dimensions facilement identifiables car elles correspondent à des tables de la BD mais il faut aussi réussir à détecter les dimensions comme par exemple la dimension « Temps » qui sont rarement stockées dans des tables.

Passage entre les bases relationnelles et multidimensionnelles

Une fois ce travail visant à décrire les données qu'on souhaite stocker dans la base multidimensionnelle réalisé, il reste à vérifier que ces données sont disponibles dans les bases relationnelles et à réaliser l'interface qui permettra de passer du système relationnel au système multidimensionnel.

La contrôle de l'existence des données est relativement aisé, il suffit pour cela de vérifier :

- Que les dimensions définies correspondent soit à des tables du systèmes relationnel, soit à des données qui représentent des propriétés des objets réels représentés dans les tables.
- Que les variables à fournir ou servant au calcul des formules sont stockées ou calculables à partir des données du système relationnel
- Que les données permettant de construire les hiérarchies et les dimensions existent également. Souvent, ces données ne figurant pas dans les bases de données relationnelles comme par exemple le continent auquel appartient un pays, il est nécessaire de gérer des tables de correspondance spécifiques.

Il y a aussi souvent des données externes à ajouter comme par exemple les cours des devises ou des données boursières. Toutes ces informations sont à incorporer dans la BDM à partir d'autres sources que les bases de données de production.

L'interface entre les deux systèmes se fait en définissant des règles de passages entre les objets de chacun des mondes. Deux étapes sont nécessaires, l'une pour alimenter les dimensions, l'autre pour alimenter les données.

II.3. Approche parallèle

Parfois, il existe déjà dans l'entreprise un entrepôt de données. Pour la création d'un nouvel hypercube correspondant à un nouveau besoin d'analyse, on détermine à partir de la méthode descendante les données dont on aura besoin, on construit une vue de l'entrepôt limitée à ces données puis, à partir de cette vue et après l'avoir complétée par les données externes particulières demandées par les décideurs et de la méthode ascendante, on construit l'architecture de la base de donnée multidimensionnelle.

La Figure 23 représente les liens entre les différents systèmes d'information de l'entreprise.

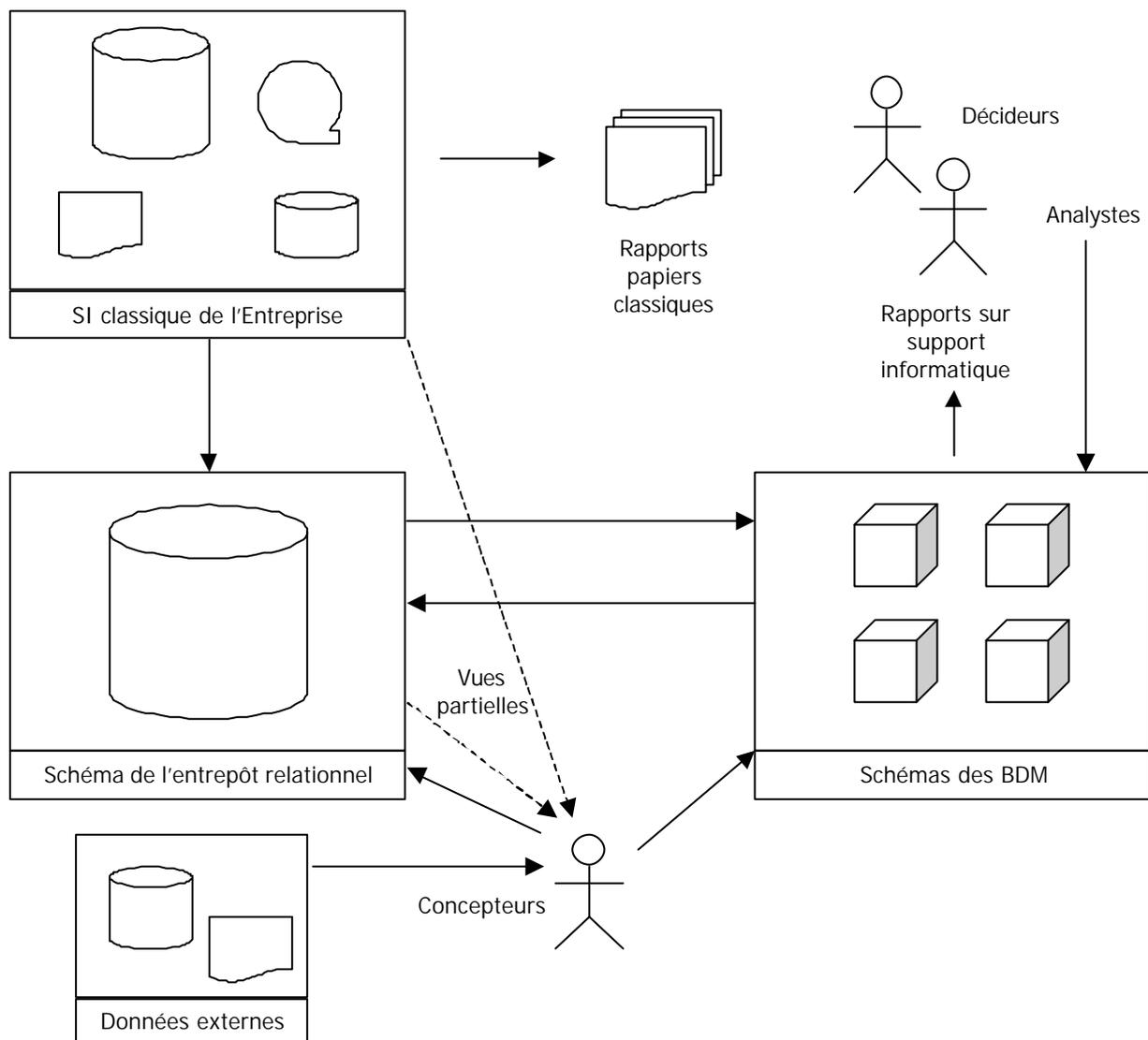


Figure 23. Les systèmes d'informations de l'entreprise

Pour concevoir un système décisionnel efficace, le concepteur doit être capable de transformer cette vision (Figure 23) pour faire apparaître deux systèmes comme le préconise H.Tardieu¹² :

¹² Hubert Tardieu - Système d'Information - Stratégique et Système - d'Informations Stratégiques

- Le système d'informations - stratégique ou SI-S qui désigne un système d'information permettant l'automatisation de l'activité de l'entreprise utilisée à des fins stratégiques
- Le système d'information stratégique (S-IS), c'est à dire un système qui manipule de l'information stratégique pour construire et suivre une stratégie.

D'un coté, on a le SI-S constitué de toutes les sources d'information présentes dans l'entreprise, et plus spécifiquement d'une sous-partie de ces informations utiles pour les activités stratégiques. Ces informations sont structurées pour pouvoir être représentées dans un système unique : l'entrepôt relationnel dont on voit donc qu'il n'est qu'une vue selon un angle stratégique du système d'information opérationnel de l'entreprise.

De l'autre coté, on a le S-IS qui est le système permettant de donner du sens et de rapprocher toutes les informations contenues dans l'entrepôt ; Pour cela, on rajoute des BDM orientées selon l'utilisation qu'on souhaite faire des données, ces BDM matérialisent donc elles aussi des vues de l'entrepôt selon un angle orienté-utilisation.

Lorsqu'on a cette vision des données de l'Entreprise, il devient plus facile d'appréhender la méthode pour concevoir un nouvel hypercube :

- Définir la vue des données de l'entreprise concernées par le problème considéré (méthode descendante) et vérifier si toutes ces données sont stockées dans l'entrepôt, le cas échéant, il convient de les extraire ou de les rechercher à l'extérieur pour les y stocker.
- Définir une vue des données telles qu'elles devront être représentées dans l'hypercube (méthode ascendante).
- Le travail du concepteur revient alors à rapprocher les deux vues, travail grandement facilité par le fait qu'elles sont toutes les deux vues exprimées dans le même langage.

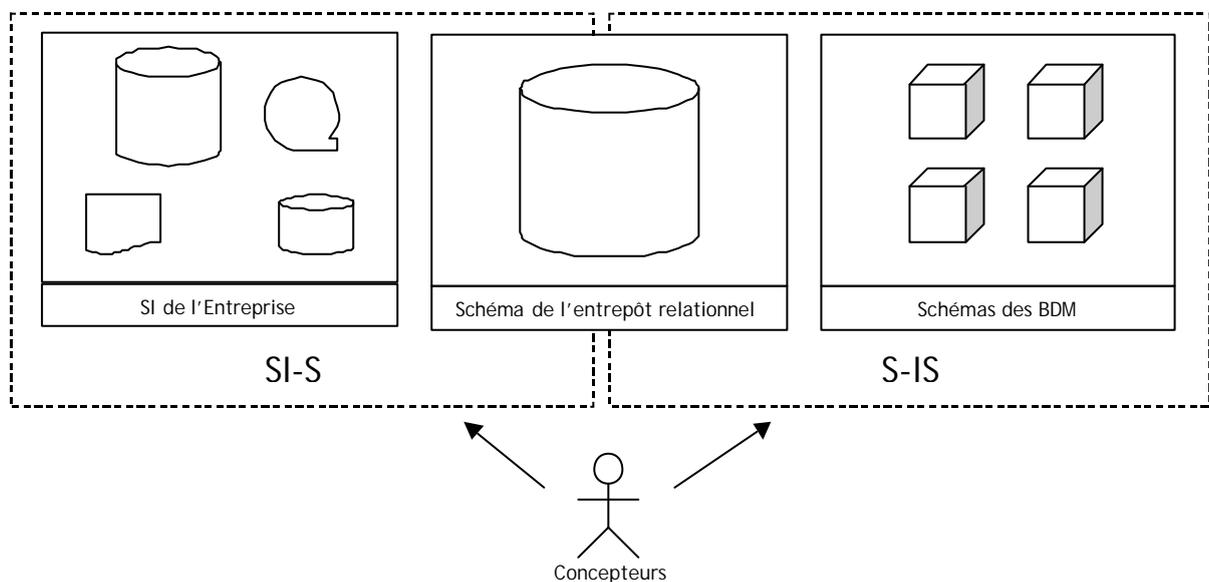


Figure 24. SI-S et S-IS

Nous venons de donner quelques éléments méthodologiques pour la construction d'une BDM, nous sommes tout à fait conscients des limites de ces propositions qui proviennent essentiellement du manque de formalisation associée au modèle multidimensionnel et de la diversité des situations réelles.

Pour effectuer un rapprochement avec les méthodes classiques de conception des BD¹³, nous avons proposé une méthode empirique de conception d'un schéma conceptuel d'une BDM. Ce schéma conceptuel s'exprime avec les concepts du modèle multidimensionnel présenté au paragraphe 1.2.1. Ensuite comme dans le cas des BDR, il faut choisir un SGBDM ou un outil de type EIS pour implanter cette BDM.

Ces logiciels reposent sur un modèle d'implantation physique des hypercubes présentant des points communs. Comme dans le cas des BDR il est nécessaire dans une deuxième étape de la conception de la BDM d'effectuer des choix concernant certains aspects du schéma de la BDM afin d'améliorer les performances. Dans le paragraphe suivant, nous présentons nos résultats concernant cette deuxième étape de la conception d'une BDM.

II.4.Optimisation des bases multidimensionnelles

L'optimisation des bases multidimensionnelles peut s'envisager sous deux aspects souvent incompatibles, la volumétrie ou les performances, l'optimisation de l'un se faisant souvent au dépens de l'autre.

Les principes d'optimisation présentés ici reposent sur le moteur de bases de données multidimensionnel d'Oracle et peuvent ne pas être vérifiés pour d'autres moteurs.

II.4.1.Volumétrie

Les données éparses constituent un élément important à considérer lors de la construction d'une base multidimensionnelle. Les données éparses sont des données qui contiennent une large nombre de cellules non remplies. Dans la plupart des moteurs multidimensionnels, il est important de regrouper ces données pour essayer d'avoir des pages entières de données vides.

Par exemple, une société peut ne pas avoir de ventes pour certaines périodes futures, ou un certain nombre de combinaisons de valeurs de dimensions peuvent ne jamais contenir de ventes (des maillots de bains en Alaska, des anoraks en Floride)

Pour minimiser la place disque utilisée par une base de données multidimensionnelle, il faut réduire la taille des variables contenant des données éparses. Les données éparses sont de deux types :

- La dispersion contrôlée
- La dispersion aléatoire

Données éparses contrôlées

La dispersion contrôlée apparaît quand une variable n'a pas de données sur une plage continue de valeurs pour une ou plusieurs des dimensions. Ce résultat dépend essentiellement de la manière dont la base a été conçue.

Lorsqu'on définit une variable, l'ordre dans lequel on énonce les dimensions qui la définisse est important. On commence toujours par la dimension qui varie le plus.

Prenons un exemple où la variable *ventes* est définie par les dimensions temps, produit et géographie.

Le temps est la dimension qui varie le plus, ensuite viennent les dimensions produit puis géographie qui varie le moins vite.

Si on ne possède pas de données pour les plus récentes périodes, ou les futures, alors la variable *ventes* va conserver les valeurs vides pour ces périodes. La raison est la suivante : les pages* de

¹³ Odile Foucaut, Odile Thiéry, Kamel Smâili - Conception des systèmes d'information et programmation événementielle - InterEditions 1996

données qui contiennent à la fois des données éparses et des données non vides sont stockées comme le montre la figure suivante :

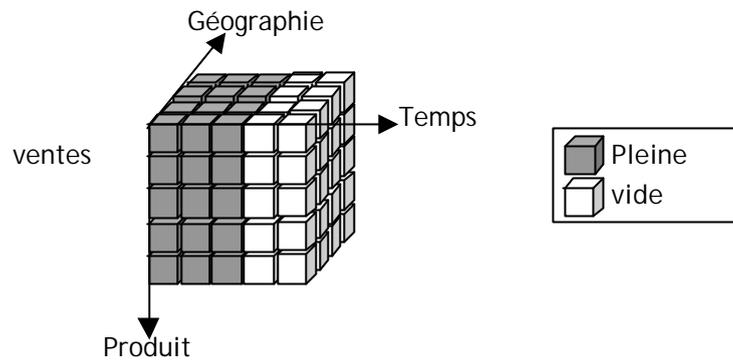


Figure 25. Données éparses non contrôlées

Toutes les cellules vides sont écrites sur le disque car aucune des pages de données n'est intégralement vide (càd aucune des tranches verticales du cube).

Pour que les données éparses ne soient pas stockées sur le disque, on doit définir la variable avec cette dimension en dernier. Dans ce cas, toutes les valeurs sont regroupées sur une (ou plusieurs) page de données communes, et ces pages ne vont pas être sauvegardées.

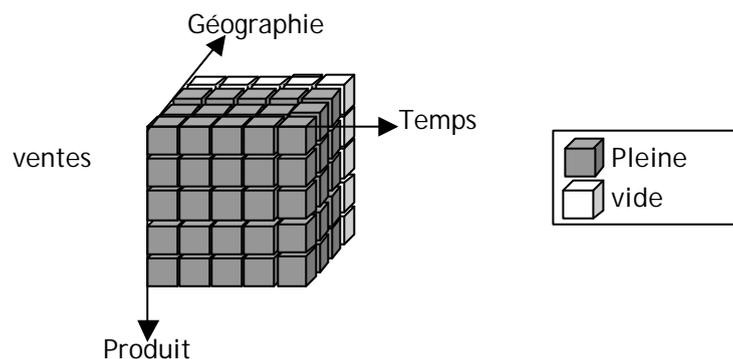


Figure 26. Données éparses contrôlées

Données éparses aléatoires

Lorsque les valeurs vides sont éparpillées aléatoirement dans la variable, on est dans le cas de données éparses aléatoires.

Ceci est dû, la plupart du temps, au fait que certaines combinaisons de positions de dimensions ne comportent jamais de données (on ne vend pas de ski en Polynésie Française par exemple). Ceci dépend donc directement de la nature de l'activité. Dans ce cas, les valeurs vides sont stockées.

Si la part des cellules vides est trop importante par rapport aux cellules non vides, il existe néanmoins une solution pour y remédier, il s'agit des dimensions conjointes. Une dimension conjointe ne stocke que les combinaisons des dimensions où il y a des données.

Les dimensions conjointes permettent de gérer uniquement les croisements significatifs d'une donnée multidimensionnelle. Prenons l'exemple d'une société dont les produits sont distribués dans certaines villes :

* une page correspond à une série de données stockées de manière consécutive.

	Amiens	Nice	Grenoble
Club de golf	785 846	1 908 172	
Bottes de neige			1 292 274
Lunettes de soleil	1 419 442	1 643 619	637 846
Tubas		447 354	
Parkas			1 335 694

Ventes

La création d'une dimension conjointe permettrait de ne retenir que les couples Produit/Ville pour lesquels la variable Ventes contient des données. On créerait une variable ventes2 définie par la dimension conjointe :

<Club de golf	Amiens>	785 846
<Club de golf	Nice>	1 908 172
<Bottes de neige	Grenoble>	1 292 274
<Lunettes de soleil	Amiens>	1 419 442
<Lunettes de soleil	Nice>	1 643 619
<Lunettes de soleil	Grenoble>	637 846
<Tubas	Nice>	447 354
<Parkas	Grenoble>	1 335 694

Ventes2

En fait, on revient aux principes d'un stockage tabulaire relationnel.

Les positions de la dimension sont données entre crochets angulaires. Elles sont composées à partir des positions associées de chaque dimension de base.

La création d'une dimension conjointe est possible quelque soit le moteur multidimensionnelle utilisé. La notion de conjointe entre tout à fait dans le modèle multidimensionnel que nous avons décrit. Par exemple, la dimension conjointe entre Produit et Ville peut se modéliser de la manière suivante :

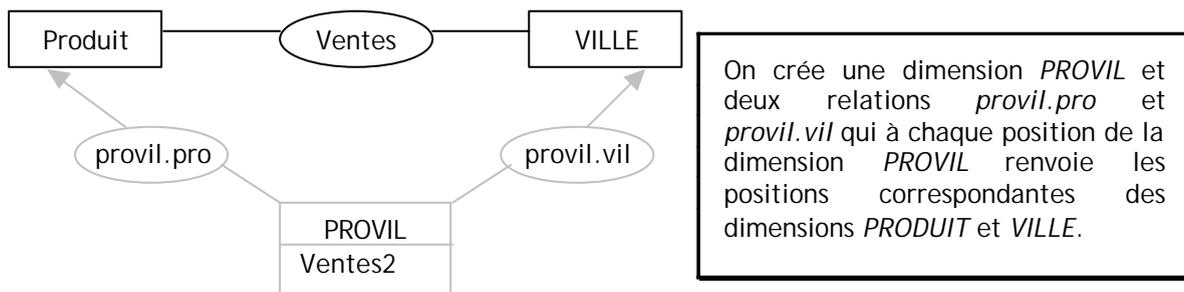


Figure 27. Dimension conjointe

Stockage des variables

L'ordre dans lequel les cellules d'une variable sont stockées sur le disque dépend de l'ordre dans lequel les dimensions de la variable ont été déclarées lors de sa création. Dans le cas de la création de plusieurs variables dimensionnées par les mêmes dimensions, il faut garder le même ordre de déclaration des dimensions des variables afin de conserver les performances.

II.4.2. Performances

La principale opération qui permette d'améliorer les performances consiste à remplacer une formule par une variable. Un programme exécute la formule et va écrire le résultat dans la variable. A chaque modification d'une des variables qui participe à la formule, il faut alors penser à exécuter le programme.

L'avantage peut être considérable si la formule est compliquée, en effet, une formule est calculée à chaque fois qu'un utilisateur y fait appel.

Le peu d'heuristiques pour optimiser les performances d'une base multidimensionnelle s'explique par le fait qu'une telle base est déjà optimisée dans cet objectif.

Alors qu'il existe de nombreuses techniques pour optimiser les performances des bases de données relationnelles et peu de méthodes concernant leur volumétrie (les méthodes de conception sont déjà tournées vers cet objectif), il existe un plus grand nombre de méthodes pour réduire la volumétrie des bases multidimensionnelles et peu pour optimiser leurs performances.

Conclusion

En conclusion, qu'il s'agisse de concevoir ou d'optimiser une base de données multidimensionnelle, plusieurs méthodes partielles existent aujourd'hui pour plusieurs moteurs multidimensionnels.

Dans les prochaines années, les concepts de ces bases devraient se standardiser permettant le développement de méthodes plus générales.

De même, ces différentes méthodes devraient faire place à quelques méthodes intégrant les principes fondamentaux de chacune d'elles.

Mais on peut affirmer que même alors, il sera nécessaire de faire appel à l'expertise humaine pour optimiser les bases multidimensionnelles, la méthode permettant uniquement de faciliter cette tâche et permettant une communication bien plus facile entre les environnements hétérogènes grâce à des représentations et des méthodes plus standards et homogènes. Cela passera aussi par un niveau supplémentaire permettant de décrire la sémantique de ces bases, c'est à dire par les métadonnées comme nous allons le voir dans le prochain chapitre.

Chapitre III. Représentation des métadonnées d'un Système Décisionnel

III.1. Définition des métadonnées

Les métadonnées, ou « *données sur les données* » doivent être au centre du système de pilotage. Elles sont essentielles à tous les niveaux, mais existent et fonctionnent à un niveau différent des données du système lui-même. Les métadonnées utilisées pour gérer et contrôler la création et la maintenance du système de pilotage doivent être gérées à l'extérieur de celui-ci. Les métadonnées utilisées par les utilisateurs du système sont une part du système lui-même et sont disponibles pour contrôler l'accès et les analyses sur le système. Pour l'utilisateur du système de pilotage, les métadonnées sont comme un « *catalogue* » des sujets contenus dans l'entrepôt.

On trouve dans la littérature diverses appellations pour qualifier ces deux types de métadonnées : métadonnées structurelles (structural metadata) et métadonnées d'accès (access metadata)¹⁴, centre des métadonnées (metadata hub) et répertoire des informations de gestion (business information directory)⁷, métadonnées d'arrière-plan (back-room metadata) et métadonnées de premier plan (front-room metadata)¹⁵.

Dans les bases de données relationnelles, on ne traite en général que les métadonnées d'accessibilité appelées dans ce cas *référentiel*.

Métadonnées structurelles

Les métadonnées structurelles sont utilisées pour la création et la maintenance d'un entrepôt de données. Elles décrivent complètement sa structure et son contenu. La base des métadonnées structurelles est le modèle de l'entrepôt qui décrit les entités qui le composent, leurs caractéristiques et comment elles sont liées entre elles. Ce modèle fournit un moyen simple de documenter et d'identifier les métadonnées structurelles. Cela inclut les utilisations stratégiques et opérationnelles des données de l'entreprise, ainsi que les totalisations multidimensionnelles. Les métadonnées structurelles doivent aussi comporter des mesures de performances pour les programmes et les requêtes afin que les développeurs puissent juger de la longueur de leurs programmes et de leurs requêtes. Ces mesures permettront aussi d'optimiser les performances de l'entrepôt.

Métadonnées d'accessibilité

Les métadonnées d'accès sont le lien dynamique entre l'entrepôt et les utilisateurs. Elles contiennent généralement les données gérées par l'entrepôt et un dictionnaire des termes incluant des noms parlant pour les données agrégées. Les métadonnées d'accès incluent aussi l'emplacement et la description des serveurs du système de pilotage et de ses composants (serveurs, hypercubes, objets des hypercubes) et résumant avec une description des données entrantes la source de chacune des données stockées et son mode de calcul. Les métadonnées d'accès fournissent des règles pour l'utilisation des structures hiérarchiques. Elles permettent également de définir des règles pour permettre aux utilisateurs de calculer leurs propres mesures et leurs propres requêtes. Enfin, les métadonnées d'accès contiennent les règles de confidentialité permettant de définir les droits des individus et des groupes à voir, modifier ou diffuser leurs analyses.

¹⁴ Alan Perkins - White Paper - Data Warehouse Architecture : A Blueprint for Success - 1999

¹⁵ Ralph Kimball - Meta Meta Data Data - DBMS, Mars 1998

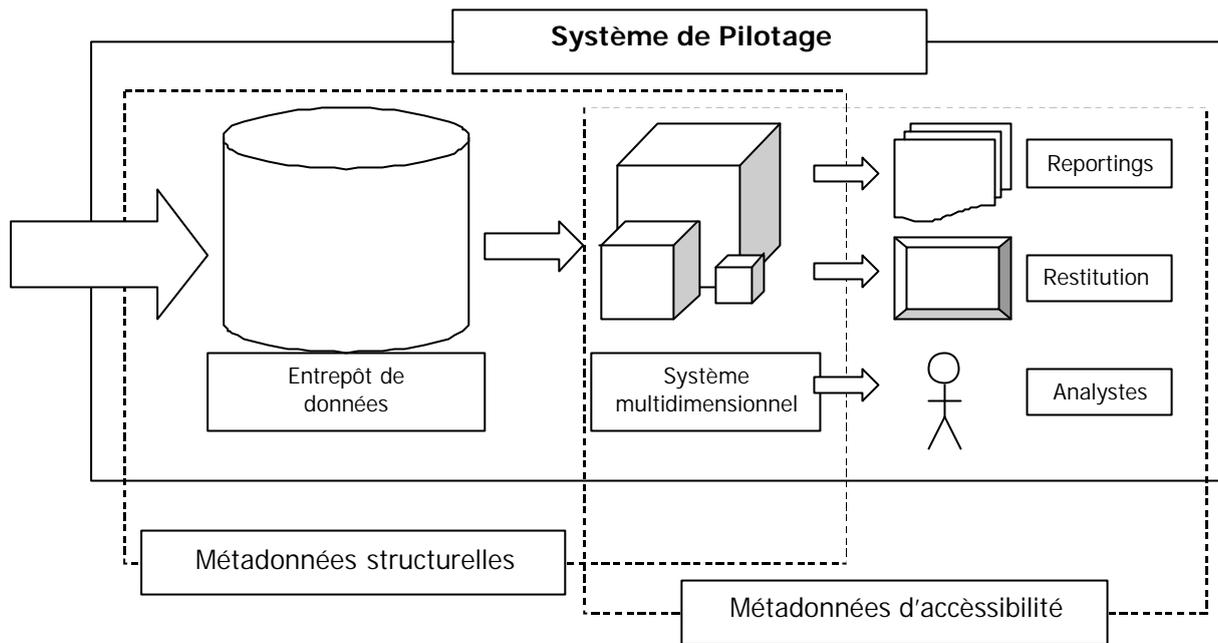


Figure 28. Typologie des métadonnées

III.2. Problématique de la représentation des métadonnées

La question de savoir si les deux types de métadonnées doivent être maintenues dans le même système ou séparément a longtemps été posée. Actuellement, on considère les métadonnées selon la perspective des personnes qui les utilisent, et non plus simplement selon le type de ces données.

En général, les concepteurs et les développeurs utilisent des métadonnées structurales pour le développement du système de pilotage et pour sa maintenance. Ils ont également besoin de certaines informations sur les données d'accessibilité. Les utilisateurs et les analystes, de l'autre côté, utilisent les métadonnées d'accès comme une aide pour trouver l'information dont ils ont besoin, mais aussi comme des métadonnées techniques de haut niveau quand ils effectuent des analyses complexes et cherchent à déterminer la provenance et la signification exacte de chacune des données.

Un autre argument fort plaide en faveur de la séparation des deux types de métadonnées : le mode d'accès à ces données. En effet, l'usage, l'architecture et l'interface d'accès pour les deux types de données sont complètement différents (même s'il existe en général un lien entre permettant de lier les deux systèmes).

III.3. Notre choix : une BDM pour les métadonnées d'accès

Jusqu'à maintenant, tous les systèmes de méta-données étaient structurés d'après le modèle relationnel, mais on a vu que ce modèle ne permettait pas de représenter efficacement les concepts relatifs aux bases de données multidimensionnelles.

Pour contourner cette difficulté, nous avons choisi de structurer ces données grâce au modèle multidimensionnel déjà présenté dans le chapitre 1.2.2. Cela est rendu possible par la méta-circularité de ce modèle que nous avons déjà vu plus haut. De plus, ce choix semble logique puisque le modèle qui structure les données est lui-même un modèle multidimensionnel.

L'autre avantage de ce choix est que cela permet un accès facile à ces données pour toutes les populations car on utilisera les mêmes outils que pour l'accès aux données elles-mêmes.

Le lien entre les métadonnées d'accès et les métadonnées structurelles est géré par un objet commun utilisé dans les deux systèmes : les extractions pour l'entrepôt de données sont les alimentations des hypercubes.

III.4. Métadonnées multidimensionnelles

Les données concernant la structure de la base et les liens entre ses objets sont automatiquement récupérées par des requêtes techniques sur la base, des données sémantiques peuvent être ensuite rajoutées pour permettre aux utilisateurs une meilleure compréhension des données.

Lors de cette phase de récupération des métadonnées, on effectue et on enregistre également quelques comptages simples comme le nombre de positions d'une dimension et par niveau de hiérarchie. Ces comptages permettent d'effectuer un premier contrôle de cohérence après l'alimentation et permet d'améliorer la qualité du système en détectant des dysfonctionnements globaux.

Cette base permet aux analystes de maîtriser le contenu sémantique de chacune des informations qu'ils manipulent (mode de calcul, unité, règles de gestion...).

Cette base de métadonnées pourra également être utile à l'administrateur de la base et aux développeurs. Par exemple, si un objet doit évoluer, par suite d'une évolution du référentiel ou de l'introduction d'une nouvelle notion, d'un nouveau mode de calcul, d'une nouvelle règle de gestion, il pourra facilement faire une analyse d'impact car il saura facilement quels objets sont liés à l'objet qui évolue.

III.5. Une BDM pour les métadonnées

Ayant choisi de représenter les métadonnées dans une BDM, nous donnons Figure 29 le schéma de cette base en utilisant le modèle de représentation des BDM défini au Chapitre III.

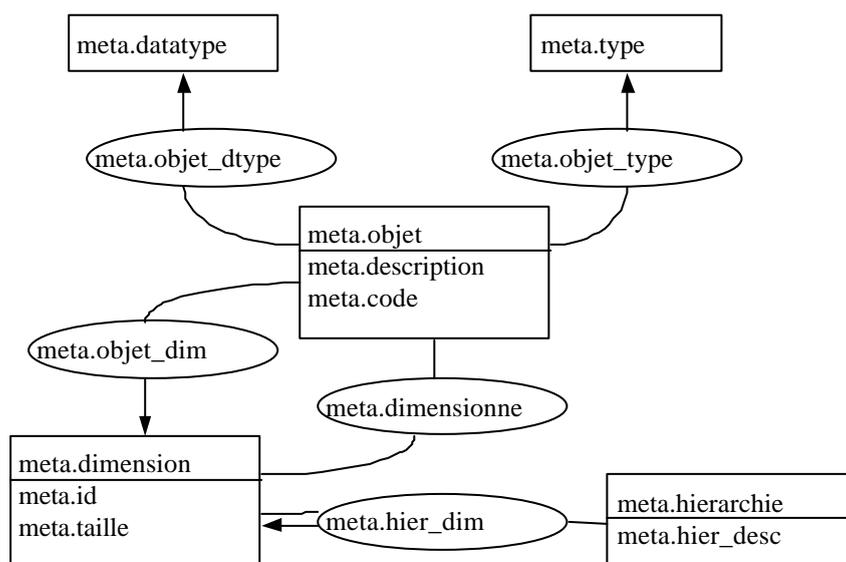


Figure 29. Schéma de la base des métadonnées

Ce schéma a fait apparaître cinq dimensions, une variable et quatre relations. Détaillons les successivement.

Dimensions :

- `Meta.objet`, cette dimension se trouve au centre du modèle et comporte des positions pour chaque hypercube, chaque dimension, variable, formule, relation ou programme. Deux propriétés sont associées à chaque objet : une description et un code.
- `Meta.type`, c'est une dimension technique comportant les types de bases possibles que peuvent prendre les objets : hypercube, dimension, variable, formule, relation ou programme.
- `Meta.datatype`, c'est également une dimension technique qui comporte les types des données gérées par un objet : entier, décimal, booléen, date, texte... On ajoute à ces types de bases chacune des dimensions, en effet le type d'une relation correspond à une dimension.
- `Meta.dimension`, cette dimension contient une copie de tout les objets de type dimension. Elle est rendue nécessaire par l'impossibilité de créer une variable réflexive sur `meta.objet`. Deux propriétés sont associées à chaque dimension : son identifiant Express qui facilite la tâche des développeurs et sa taille qui permet de faire des contrôles de cohérence.
- `Meta.hiérarchie`, c'est une dimension qui contient les hiérarchies existantes sur chacune des dimensions. Une propriété est associée permettant de définir une description pour chaque hiérarchie.

Une seule variable est nécessaire, `meta.dimensionne`, définie entre `meta.objet` et `meta.dimension`. Cette variable de type booléen est en fait une matrice indiquant pour chaque objet s'il est ou non définie par une dimension.

Relations :

- `Meta.objet_dtype` renvoie le type des données gérées par chaque objet.
- `Meta.objet_type` renvoie type de chaque objet.
- `Meta.objet_dim` renvoie, pour les objets de type relation, la dimension renvoyée.
- `Meta.hier_dim` renvoie pour chaque hiérarchie sa dimension associée.

Cette BDM peut être implantée sur un site avec le même logiciel que celui sert aux hypercubes. Les administrateurs et les développeurs peuvent y accéder à travers le même environnement graphique que celui utilisé pour les hypercubes. Enfin, un lien entre la BDM des métadonnées et les BDM d'analyse doit permettre de passer facilement de l'une à l'autre afin d'obtenir pendant une analyse de certaines données des informations sur leur provenance, leurs liens avec d'autres...

III.6. Une BDM pour la gestion des profils

Ayant expérimenté l'intérêt de la modélisation multidimensionnelle pour les métadonnées d'accessibilité concernant leur sémantique, nous avons retenu la même approche pour les métadonnées concernant la confidentialité et les autorisations de manipulation des utilisateurs, ce que nous appelons les « profils ».

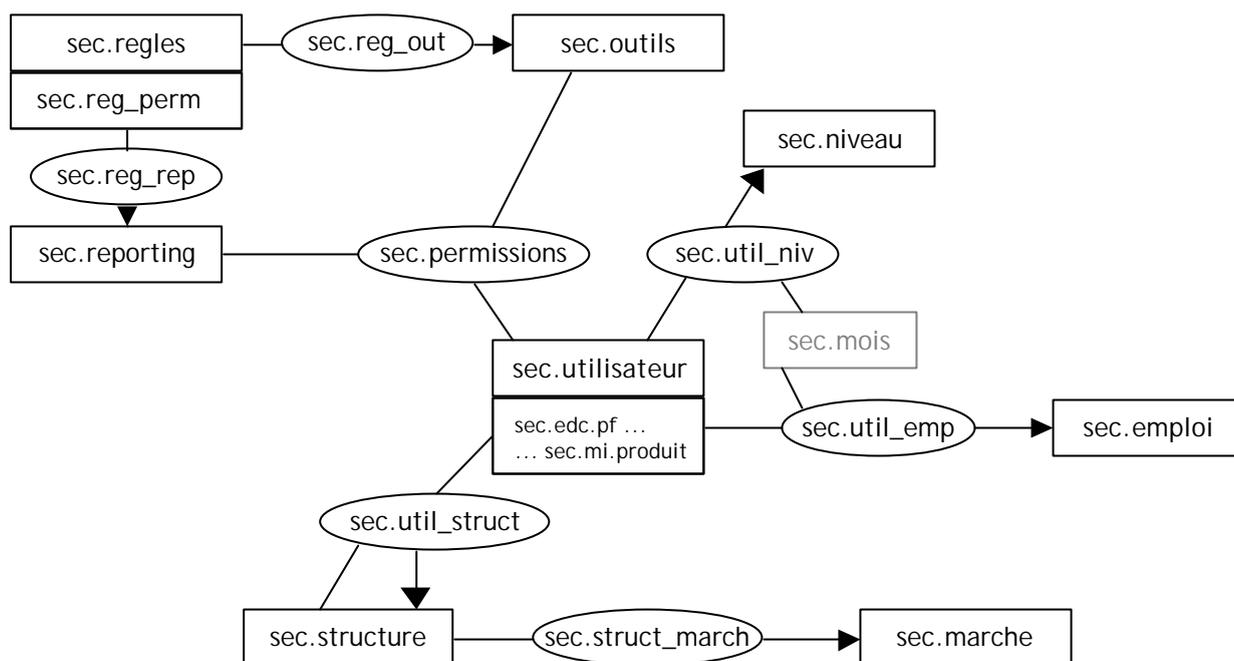


Figure 30. Schéma de la BDM de gestion des profils

La BDM de gestion des profils permet de décrire et d'analyser 7 dimensions :

- Utilisateur
- Structure (CA, Agence, Secteur, Direction Régionale et Banque)
- Marché
- Emploi
- Niveau (au sens d'un niveau de structure : CA, Agence, Secteur, DR ou Banque)
- Reporting
- Outil (des fonctionnalités des reportings qui ne sont pas offertes à tout le monde : exporter vers excel, établir un palmarès... + l'outil Accès pour les droits de consultation)

Une 8^{ème} dimension permet de gérer les profils d'une manière plus globale, il s'agit de la dimension Règle qui permet de stocker les règles à appliquer chaque début de mois pour rétablir les droits (par ex. Toutes les personnes ayant au moins le niveau Secteur ont accès à l'outil Palmarès pour le reporting EDC). Tout les début de mois, on applique donc les règles, ce qui permet de remplir la variable permission.

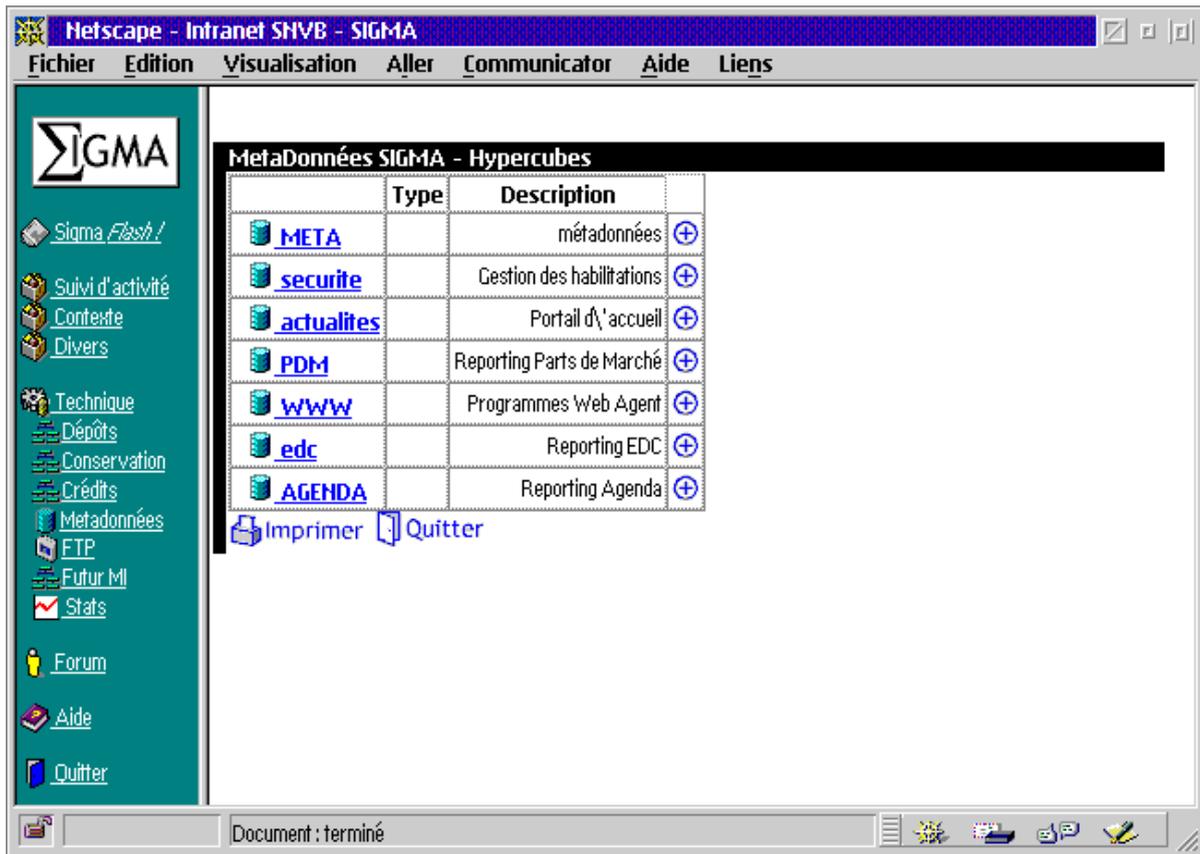
La variable principale est la variable permission qui, pour un utilisateur, un outil et un reporting, indique s'il possède ou non ce droit. Grâce à l'outil « Accès », on peut de plus gérer les droits des utilisateurs à consulter ou non un certain reporting. On pourra alors, par exemple, avoir la règle « Tout les utilisateurs ont l'outil Accès pour le reporting EDC », ainsi, tout les utilisateurs présents dans la base ont l'accès au reporting EDC.

Un utilisateur appartient à une structure. Quand cette structure est la plus fine, c'est à dire le portefeuille, on l'associe de plus à un marché, sinon on l'associe à la position Tous marchés.

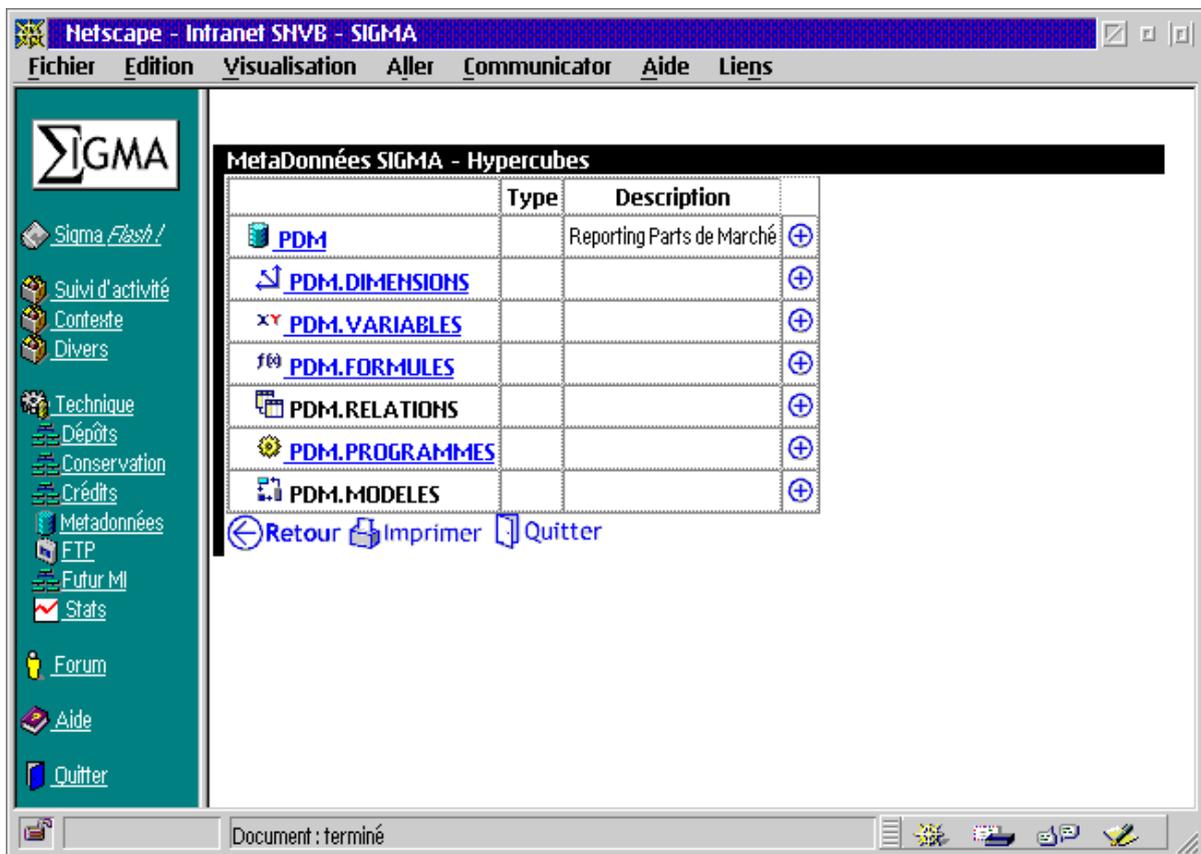
Un utilisateur a un emploi et a un certain niveau. Ce niveau est indépendant de la structure : le rattachement des personnels administratifs à un élément de structure, par exemple, n'illustre pas le niveau auquel ils sont censés avoir accès.

III.6.1. Implantation avec Oracle Express

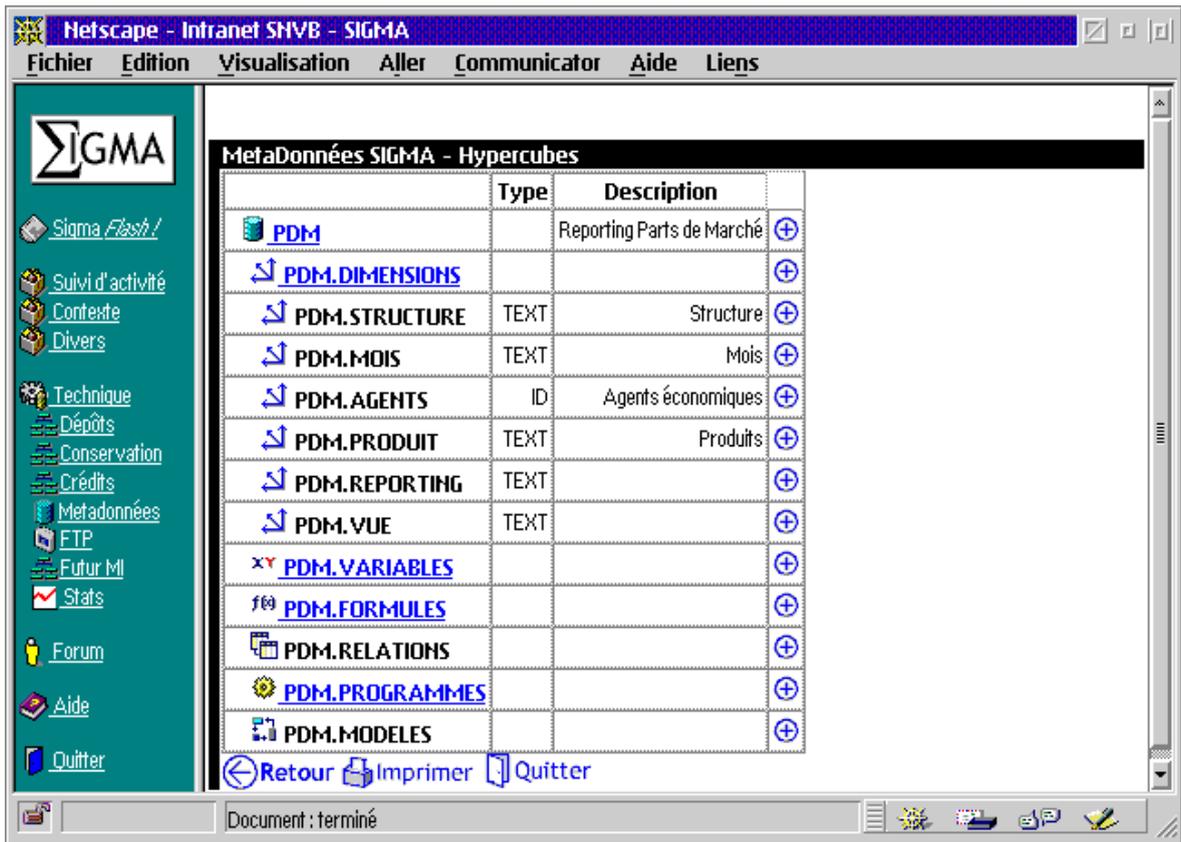
On trouve ci après les réalisations faites à la SNVB pour gérer les métadonnées d'accessibilité. On peut remarquer que le canal utilisé pour diffuser ces métadonnées est l'Intranet, c'est à dire rigoureusement le même que pour l'application de pilotage elle-même.



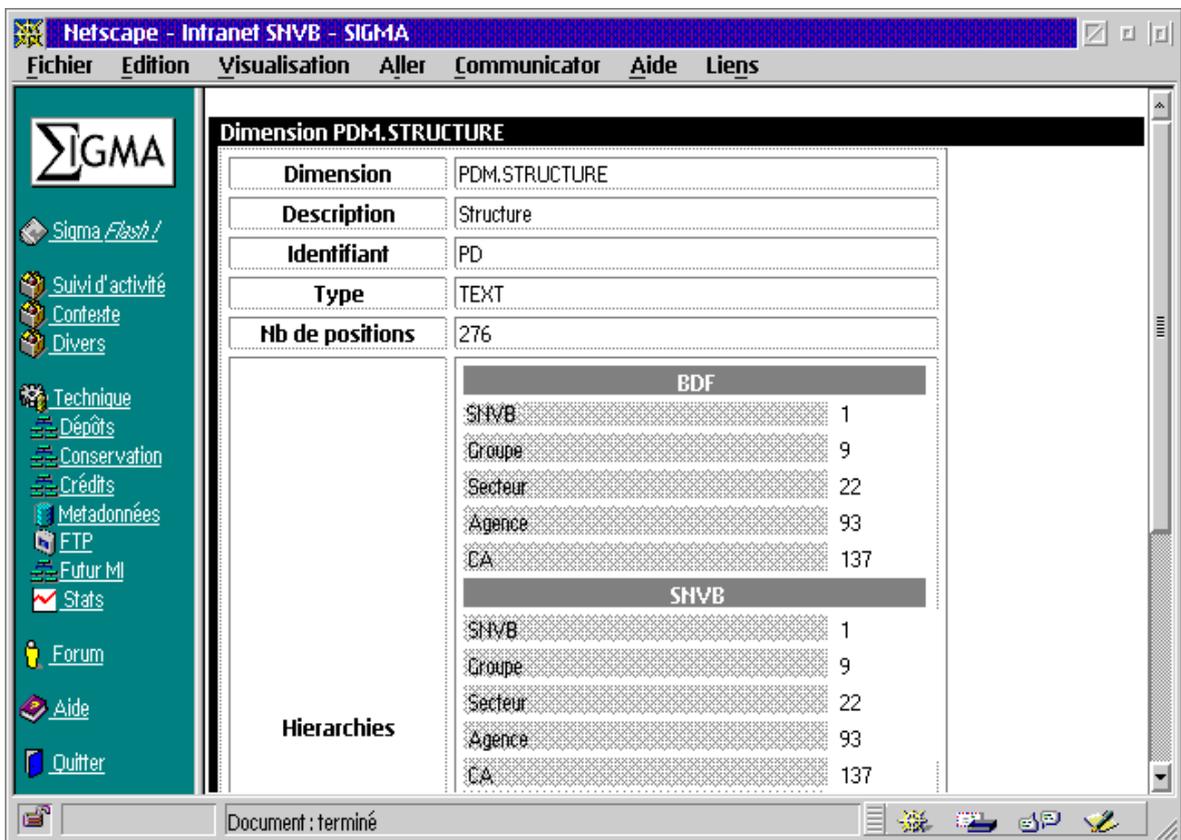
Toutes les bases (et leurs descriptions)



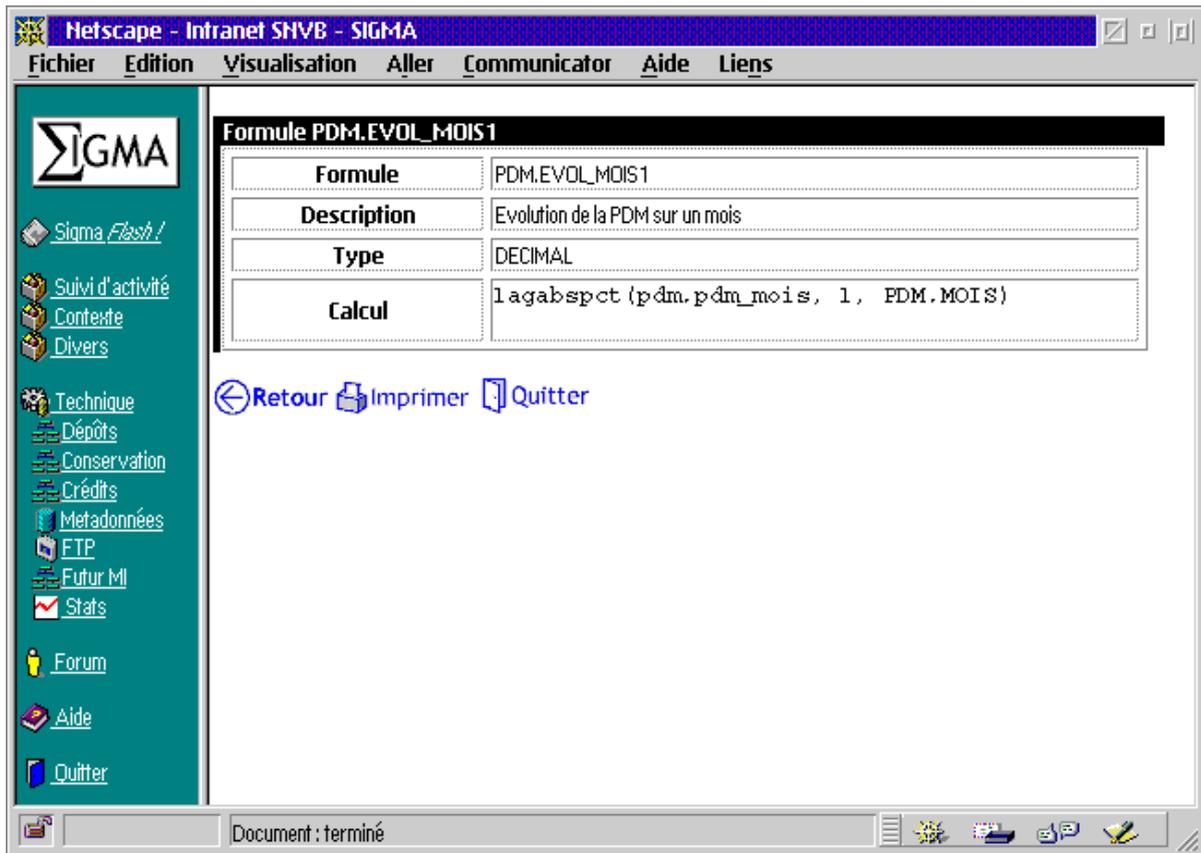
Une base (et ses sous catégories : dimensions, variables, formules, relations, programmes et modèles)



Les dimensions d'une base (leurs noms, leurs types et leurs descriptions)



Une dimension (et ses caractéristiques : nom, description, identifiant, type de données, nombre de positions, ses hiérarchies et les niveaux hiérarchiques)



Une formule (et ses caractéristiques : nom, description, type de données, expression de calcul)

III.6.2. Conclusion

L'exemple que nous venons de présenter montre que les BDM permettent bien plus que la gestion des données et leur analyse comme c'est leur but premier. Gérer également des cubes techniques permet d'avoir un environnement plus cohérent et autorise la création de passerelles entre les différents hypercubes ce qui facilitera grandement les travaux de l'administrateur et des développeurs.

En effet la compréhension de leurs concepts nous a permis de les utiliser pour décrire les bases de données techniques nécessaires à la gestion et à la compréhension des BDM d'un site. Nous offrons ainsi à l'utilisateur des BDM un environnement unique pour consulter ces informations mais aussi pour en comprendre l'origine et le sens au travers de ma BDM des métadonnées et pour connaître ses droits d'accès au travers de la BDM des profils. Un système de diffusion de l'information a même été mis en place en utilisant toujours le même principe et en permettant de cibler son information non seulement sur quelques personnes mais également en s'appuyant sur les attributs déjà utilisés dans la base de gestion des profils.

Remarquons que notre démarche est tout à fait similaire à celle utilisée par les constructeurs des SGBD relationnels qui ont stocké dans les mêmes structures relationnelles à la fois les données et les informations sur les données et les droits d'accès au travers des tables spécifiques dites " tables système ".

Cependant, à notre connaissance, aucun fournisseur de SGBD dédié au multidimensionnel n'utilise encore de façon aussi explicite le principe de métacircularité que dans notre proposition.

Nous consacrons maintenant la dernière partie de ce mémoire à l'application en marketing bancaire dans laquelle nous avons mis en œuvre ces résultats théoriques et grâce à laquelle nous avons pu les découvrir.

Chapitre IV. Application en marketing bancaire

IV.1. La banque SNVB

Créée à Nancy en 1881 à l'initiative d'entrepreneurs et de financiers régionaux, la Société Nancéienne de Crédit Industriel a progressivement étendu ses activités à l'ensemble des départements lorrains. Puis, en absorbant de nombreuses banques locales, elle s'est développée vers l'Ouest jusqu'à Paris. Sa fusion, en 1972, avec la Banque Varin-Bernier a donné naissance à l'actuelle Société Nancéienne Varin-Bernier. Elle est rattachée en 1982 au groupe CIC qui est nationalisé. Le rachat du groupe CIC en mai 1998 par le Crédit Mutuel a permis à la SNVB de réintégrer le secteur privé.

Présente par son réseau de points de ventes dans dix départements, la SNVB répartit ses activités par tiers entre Paris-Ile-de-France, la Champagne-Ardenne et la Lorraine. Banque interrégionale, elle est organisée de façon décentralisée.

Par sa taille, la SNVB est aussi un opérateur financier puissant. Elle met à disposition de sa clientèle des équipes de spécialistes de haut niveau. Grâce à son appartenance au Groupe CIC, elle s'adosse à un large réseau qui lui permet d'accompagner chaque client dans ses besoins à l'échelle nationale, européenne et internationale.

Aujourd'hui, la SNVB compte 189 agences, dans lesquelles sont réparties 2500 salariés. Elle gère les comptes de près de 490.000 clients, principalement des particuliers. Son capital est en 1999 de 400 000 000 Francs soit 60 979 607 Euros.

IV.1.1. Le Département des Etudes Informatiques (DEI)

Le Département des Etudes Informatiques est rattaché à la Direction de l'Organisation des Systèmes d'Information (DOSI). Cette dernière comprend en plus du D.E.I. les départements Organisation des Etudes, Organisation des Activités, Informatique Clientèle, Production Informatique et la Division Architecture Système.

Le Département des Etudes Informatiques compte une cinquantaine de salariés répartis dans 9 domaines :

- **Comptabilité**
Son activité est découpée entre la comptabilité interne à la SNVB, la comptabilité clients (gestion des comptes courants, plans épargne logement, codevi, etc) et la gestion des paiements et contentieux.
- **Monétique**
Sa tâche est de gérer les nouveaux modes de paiement, par exemple les cartes bleues, et les guichets automatiques ainsi que les terminaux de paiement électroniques.
- **Base Personnes et Engagements**
Ce domaine a en charge l'ensemble des travaux d'étude concernant la gestion des données permanentes, entre autres les clients, les contrats, les comptes internes. Il gère également tous les crédits par contrats, les engagements par signatures, des assurances spécifiques, le portefeuille commercial (lettres de changes...) et le portefeuille financier (billets à ordre...)
- **Titres-Etranger-Ressources Humaines**

Il s'occupe de la gestion des produits de la bourse, des transactions avec l'étranger, du personnel (c'est à dire gestion des ressources humaines : édition des bulletins de paie, gestion des congés et de la formation) et de la structure de la SNVB (agences).

- Applications Divers

Il s'occupe des maintenances et de l'adaptation de l'existant.

- Télématique

Il s'occupe de tous les logiciels développés sur Minitel et sur Internet : consultation par les particuliers de la situation de leurs comptes, et pour la bourse des portefeuilles et des opérations effectués. Le serveur vocal Filbanque rentre également dans ses attributions.

- Applicatifs du Poste de Travail

Il s'occupe de la micro-informatique (réalisation d'études et de logiciels, assistance aux utilisateurs), de la maintenance et de l'évolution de l'application Guichet d'agence, sur système local et central (réseau ASTER).

- Logistique

Son activité se compose de la gestion de l'architecture applicative, de l'administration des données, de l'infoservice et des méthodes.

- Système d'Information Commerciale

IV.1.2. Le domaine Système d'Information Commerciale (SIC)

Le domaine Système d'Information Commerciale au sein duquel j'ai effectué mon travail de recherche s'occupe :

- De la mise en place de logiciels orientés vers le commerce (Gestion des Actions Commerciales, gestion des campagnes, des événements) et de leur maintenance.
- De la gestion des risques, de la qualité, des systèmes experts
- Du pilotage : gestion des tableaux de bords, gestion de l'activité, suivi de production

Dernièrement, le SIC a mis en œuvre la bascule progressive de son système de pilotage vers une structure, non plus de type infocentre, mais de type DataWarehouse.

IV.1.3. Architecture informatique de la SNVB

Avant 1990, les agences étaient reliées à 6 IBM 4361, situés au niveau des directions régionales, elles-mêmes connectées au site central. A partir de 1991, les ordinateurs ont été remplacés par des réseaux régionaux appelés « réseaux ASTER » (Architecture de Système de Traitement en Réseau), réseaux connectés en permanence à un IBM 3090. Sur ces réseaux de type Token Ring sont connectés des micro-ordinateurs. Un terminal par agence joue le rôle de serveur.

On trouve principalement deux types de configuration au sein de la SNVB.

Les postes en agences sont des micro-ordinateurs sous OS/2 Warp.

Les postes de travail informatiques sont des PC avec pour système d'exploitation Windows 95.

La mise en place récemment d'un Intranet permet de faire communiquer ces deux mondes avec comme but à terme de faire migrer les postes sous OS/2 vers Windows NT.

IV.2. Le projet SIGMA

Du constat qu'il existe actuellement une multiplicité de tableaux de bord papier dont les contenus, faute d'une base de pilotage unique et partagée, ne sont pas toujours cohérents, un projet a été créé et inscrit au programme stratégique 95-97 «Rénover nos systèmes de pilotage».

L'analyse du problème a entraîné la création d'un système de pilotage complet composé d'un entrepôt de données, d'un système de diffusion des reportings et d'un système d'analyse des données. Ce projet a de nouveau été inscrit au programme stratégique 98-2000 « Mieux guider le développement grâce à un système de pilotage rénové ».

Les objectifs principaux sont la construction et l'exploitation d'un entrepôt de données unique, pour répondre aux besoins de pilotage stratégique et opérationnel des domaines Commercial, Qualité, Risques et Financier.

Les objectifs ont été fixés par le Comité Exécutif :

- Parvenir à la suppression des tableaux de bords papier
- Diffuser une information transparente et identique pour tous
- Accroître les capacités d'analyse en fédérant les différents tableaux de bords
- Gagner en rapidité et en fiabilité en utilisant des outils modernes de transformation et de diffusion des données en informations de pilotage

Notre travail de DRT s'est déroulé en trois étapes où nous avons alterné du travail pratique sur le développement du système de pilotage et des études plus théoriques sur les modèles et méthodes pour construire ce type de système.

Veille technologique

La première étape a été entièrement consacrée à l'acquisition de connaissances sur le domaine des systèmes décisionnels et à une veille technologique approfondie. Elle a donné lieu à un rapport sur « les DataWarehouses » .

Ce premier travail a permis de conforter les objectifs fixés par le comité exécutif et de montrer que le logiciel Oracle Express, auquel nous nous sommes formés, retenu par le groupe CIC était adapté au projet.

Système SIGMA

La seconde étape a comporté deux aspects : le choix et la mise en place de l'architecture technique et applicative du système SIGMA et l'élaboration d'un plan de déploiement progressif.

La troisième étape a conduit à la mise en place opérationnelle d'un premier lot et à une mise en forme des résultats plus méthodologiques dans ce mémoire.

Démarche méthodologique

Le projet a été découpé en quatre lots principaux correspondant aux quatre domaines d'activité Commercial, Qualité, Risques et Financier. Ces lots ont eux-mêmes été découpés en plusieurs sous-lots.

L'entrepôt se construit au fur et à mesure de la réalisation des lots. Une première version du système est construite autour du domaine prioritaire, le domaine Commercial, et la conception se poursuit par itération sur les lots successifs. Dès le premier lot, l'architecture technique et applicative du système a été complètement construite afin d'assurer la couverture fonctionnelle des autres domaines.

Le système de bases de données multidimensionnelle retenu est le système Oracle Express.

Le canal de distribution de l'information retenu est l'Intranet.

Stratégie de lotissement

Le premier lot ne crée pas de nouveaux reportings mais opère la bascule dans le système de reportings existants. Dans un premier temps, on invente rien, mais on organise mieux.

Cette démarche permet de vérifier que l'intégralité de l'architecture retenue fonctionne correctement et de prouver rapidement la plus value apportée par le système.

Ainsi, après la mise en œuvre du premier lot, les autres lots peuvent s'enchaîner facilement.

Dans ce domaine également, l'aspect recherche lié au DRT a été très important car la mise en place du premier lot a été accompagnée d'une recherche de modèle et de méthode dans le cadre même du DRT.

IV.3. Architecture du système

Le système de pilotage SIGMA peut être divisé en trois composantes.

L'entrepôt de données et ses composants

Il s'agit d'une base relationnelle en DB2 sur serveur central MVS. Cet entrepôt a été spécialement conçu pour stocker les informations de base pour le système de pilotage. Il est optimisé pour des traitements lourds OLAP.

Pour stocker des données spécifiques à un seul reporting ou à une seule application, ou pour stocker des agrégats complexes à calculer et afin de ne pas « polluer » l'entrepôt, on lui adjoint des bases annexes appelées Magasins de données. Le référentiel est toujours le même, mais son interprétation est différente, c'est à dire, par exemple, qu'on ne prendra pas en compte les données relatives à certains clients comme ceux étant en contentieux.

Deux modules font également parti de ce système central, un module d'interface avec les sources servant à l'alimentation des données et au contrôle de leur qualité, ainsi qu'un module d'interface avec le système multidimensionnel pour l'extraction des données et parfois une première consolidation. La Figure 31 illustre cette architecture. L'application Mesures Intermédiaires se basant sur une interprétation propre du référentiel, elle est sortie de l'entrepôt pour être stockée dans un Datamart spécifique.

Le système multidimensionnel

Pour chaque besoin de reporting, on a une base multidimensionnelle orientée-métier optimisées pour la diffusion des données et les analyses OLAP.

L'analyse des données et l'Intranet de Pilotage

Enfin, le canal de diffusion des données vers l'utilisateur final est double :

La plupart des utilisateurs accède à des tableaux déjà construits et suivent un chemin préétabli afin de faciliter leurs analyses. Les utilisateurs accèdent à ces tableaux à travers un Intranet de Pilotage.

Les analyses ou experts ont accès directement aux bases multidimensionnelles à travers des outils d'analyse permettant de répondre à la plupart de leurs requêtes. Ils acquièrent ainsi une indépendance relative vis à vis de l'informatique.

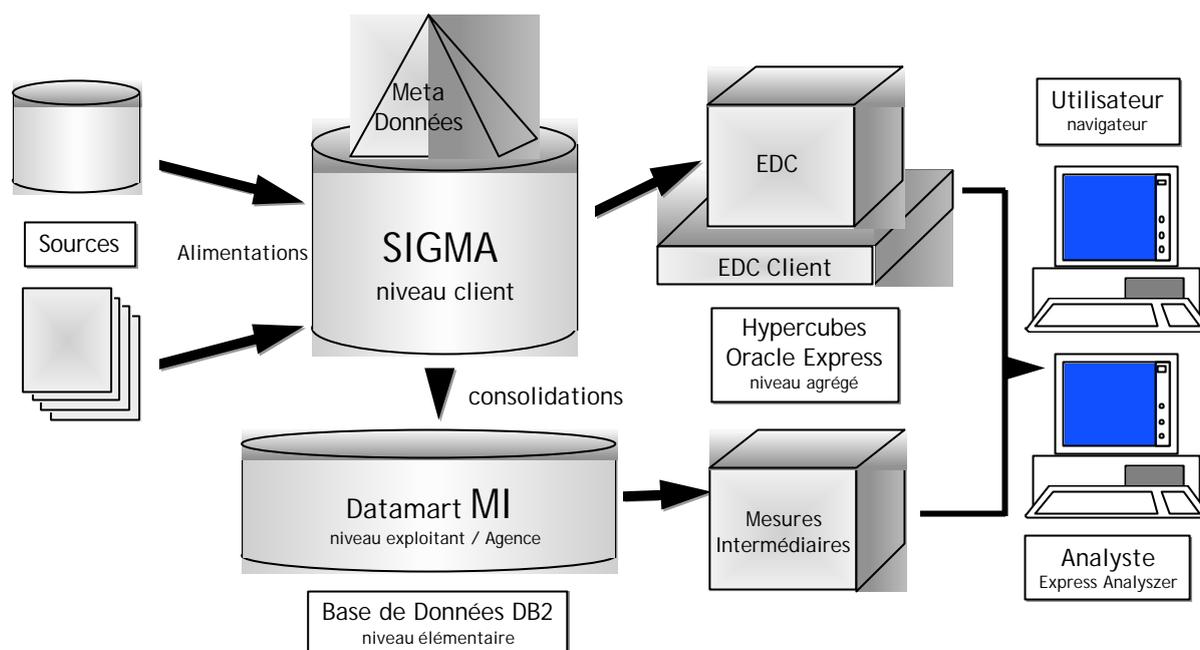


Figure 31. Architecture du système SIGMA

On retrouve la structure théorique d'un système décisionnel comme donnée dans la Figure 1 avec les sources, une interface servant à l'extraction vers le Système décisionnel (les extractions sont en Cobol et le système une base de données DB/2). Conjointement à cette base, on tient à jour un système permettant de gérer les métadonnées à ce niveau. L'interface de diffusion est gérée par des hypercubes Oracle Express qui diffusent leurs informations par l'Intranet et peut également permettre un accès en client-serveur par les analystes.

Tout les composants du système ont été conçus afin d'évoluer facilement.

L'entrepôt central a été conçu pour pouvoir stocker la plupart des données stockées par la banque, ou pour pouvoir rapidement être modifié s'il doit stocker des informations particulières.

Le système multidimensionnel est modulaire et évolue de ce fait également très rapidement.

La diffusion par l'Intranet garantie une évolutivité rapide du système car on s'affranchit ainsi des lourdeurs liées à la télédistribution des applications en mode client-serveur. Ici, chaque modification s'effectue sur le serveur de l'Intranet et est immédiatement utilisable par tous les clients du système.

Développement d'un nouveau lot

Chaque nouveau lot nécessite le développement d'une base multidimensionnelle spécifique et des restitutions qui lui sont associées. Le plus souvent, on cherche à ce que tout les lots se basent sur la même interprétation du référentiel, on pourra alors les intégrer directement dans l'entrepôt et comparer les reportings entre eux. Lorsque cela n'est pas possible, on crée un Datamart.

Actuellement, un premier lot est d'ores et déjà disponible et les lots suivants s'enchaînent à un bon rythme. Il faut savoir qu'une fois l'architecture définie, le premier lot a permis de construire des « modèles » à la fois pour les bases de données multidimensionnelles et pour les restitutions.

Aujourd'hui, un nouveau lot ne nécessite donc qu'une étude fonctionnelle, la plupart des problèmes techniques a déjà été étudiée et la construction d'un nouveau lot est facilitée par l'application des modèles.

Les hypercubes techniques

Nous avons fait le choix de gérer le système par des hypercubes techniques pour stocker les programmes d'accès aux restitutions, pour gérer les habilitations au système ou pour donner une vue globale des métadonnées.

- Gestion des profils

Comme toutes les données du système sont stockées dans des bases multidimensionnelles, les données pour les habilitations sont également gérées dans un hypercube. Le système gagne ainsi en cohérence.

Ce système a pour but de contrôler l'accès au système, de définir un profil d'accès à chaque utilisateur (administrateur, utilisateur, développeur, analyste) et de lui fournir en conséquence les reportings et outils auxquels il a droit. Enfin, le système permet de le positionner sur l'élément de la structure qui l'intéresse et de limiter sa vision selon sa position hiérarchique.

Enfin cette base permet de gérer un «contexte global» pour chaque utilisateur, c'est à dire de conserver chaque position sur chaque axe d'analyse pour permettre de lier toutes les vues et tous les reportings pour une analyse cohérente et facile.

- Métadonnées d'accessibilité

Chaque reporting nécessite un hypercube différent, il existe aussi plusieurs hypercubes techniques. Pour offrir aux développeurs comme aux utilisateurs experts une vue d'ensemble du système, un hypercube a été développé pour stocker toutes les informations relatives aux hypercubes gérés par le système. Automatiquement, cet hypercube stocke la structure et la définition de tous les hypercubes et permet une consultation en mode intranet.

- Accès aux restitutions

L'accès aux restitutions en mode intranet se fait au travers du module Web Agent d'Oracle Express. On stocke alors des programmes qui décrivent les pages à restituer. Ces programmes auraient pu se situer dans chacun des hypercubes de données, mais il nous a paru plus judicieux de séparer les données des traitements et nous avons donc créé un hypercube spécifique pour gérer ces programmes.

IV.4.Exemple d'interfaces utilisateurs

Ecran de login

Le premier écran rencontré par l'utilisateur est l'écran de login où il est invité à se signer pour qu'on puisse vérifier quel niveau d'habilitation il a et quelles données on doit lui fournir.

Pour cela, il rentre son matricule et son agence de rattachement.



SIGMA Flash !

Chaque utilisateur arrive directement à un écran général appelé «SIGMA Flash ! » qui lui présente l'actualité de l'Intranet de pilotage.

Un panneau personnalisé sur la droite lui permet de naviguer facilement dans l'Intranet SIGMA (la page SIGMA Flash !, les reportings, les forums-utilisateurs, l'aide...).

La colonne de gauche donne des informations générales et est personnalisée selon l'utilisateur. L'administrateur du système peut créer des messages pour tous, ou seulement pour un profil d'utilisateur, voire pour un seul utilisateur ; le message peut être un lien vers une vue ou une page quelconque de l'Intranet.

La colonne de droite donne les dates de dernière mise à jour des reportings.

Restitution sous forme de tableau

Outre le tableau présentant les données proprement dites, l'écran possède quatre zones permettant la navigation proprement dite. En effet, le choix a été fait de ne pas donner aux utilisateurs la possibilité de construire leurs propres tableaux et leurs propres analyses. On doit donc leur fournir tous les tableaux qui leur permettront de piloter leur activité et des liens pour naviguer dans les données.

La partie ❶ présente la position actuelle dans la structure SNVB et permet de naviguer entre les niveaux de cette structure.

► **SNVB** ❶ ► **NANCY (Groupe)** ► **VANDOEUVRE (Secteur)** ► **VANDOEUVRE** ►...

La partie ❷ permet de se déplacer sur les axes d'analyse des données.

Marché **Mois** **Structure**
Tous marchés Mai 99 VANDOEUVRE (Agence)

La partie ❸ offre des possibilités de changement de la vue ou de retraitement des données.

On peut modifier le type de représentation des données, du tableau (📊) au graphique (📈), changer l'origine d'un graphique (📍 et 📍), exporter les données vers Excel (📄), imprimer la vue (🖨️) ou demander un palmarès (🏆).

Les possibilités de cette zone dépendent en fait de la vue considérée et du profil de l'utilisateur.

La partie ❹ donne l'accès à d'autres vues, c'est à dire les même données mais présentées différemment, on fait « pivoter » l'hypercube.

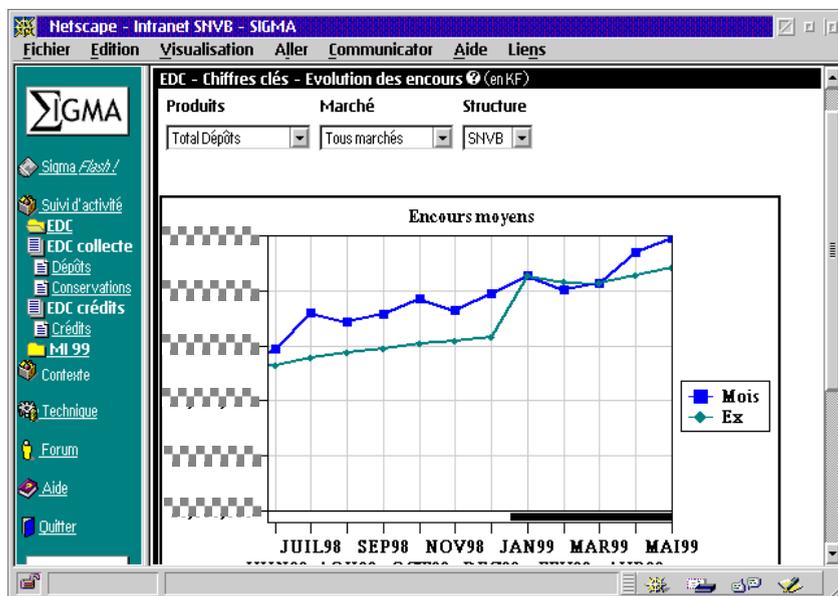
Par exemple, dans la vue « évolution mensuelle », on demande à avoir en colonne les mois plutôt que les indicateurs.

The screenshot shows the Netscape browser window titled "Netscape - Intranet SNVB - SIGMA". The main content area displays "EDC - Chiffres clés Evolution des encours (en KF)". Below this, there are three dropdown menus for "Marché" (Tous marchés), "Mois" (Mai 99), and "Structure" (SNVB). A table follows with columns for "Encours Mois", "Evol. Mois", "Encours Exercice", "Evol. Ex.", and "Tendance". The table rows are "Total Dépôts", "Total Conservations", and "Total Crédits". Below the table, there is a "Courbe d'évolution" link and a section "Autres vues..." containing links for "EDC Evolution mensuelle", "EDC Comparaison structure", "EDC Comparaison marchés", and "EDC Evolution mensuelle de la structure".

	Encours Mois	Evol. Mois	Encours Exercice	Evol. Ex.	Tendance
Total Dépôts					↗
Total Conservations					↗
Total Crédits					↗

Restitution sous forme de graphique

Une vue sous forme de graphique est globalement construite comme une vue sous forme de tableau. Elle n'est là que pour compléter la vue sous forme de tableau.



Aide du système



L'aide du système est une composante complète de l'intranet. On profite ainsi des possibilités offertes par les liens hypertextes dans la gestion des documents.

Une utilisation de pages dynamiques permet en outre une visite guidée du système qui permet de se familiariser très rapidement avec l'application.

IV.5. Un exemple d'application : la gestion des encours

Dans toutes les banques on trouve ce type de reporting qui permet de suivre l'évolution de l'activité. Des encours moyens mensuels et annuels sont calculés afin de gommer les effets journaliers, on a ainsi pour chaque produit des données dont on peut facilement comparer l'évolution.

L'application EDC traite des encours fin de mois de la SNVB. Ce n'est pas un nouveau reporting mais la transposition électronique d'un reporting existant. Les tableaux de bords ont néanmoins été repensés pour profiter des possibilités offertes par la consultation en mode intranet. Cela a permis d'accroître grandement les capacités d'analyse grâce au reporting. De plus, l'historique est reconstruit tous les mois, ce qui permet de neutraliser les effets de mouvement de clients au sein de la banque, ainsi les analyses temporelles sont plus justes et permettent une meilleure analyse.

Ce reporting est publié chaque début de mois.

Les analyses portent sur quatre axes :

- Portefeuille > Centre d'activité > Agence > Secteur > Direction régionale > Banque
- Marchés > tous marchés
- 3 hiérarchies produits : dépôts, crédit, conservations (bientôt assurances)
- Temps

Les données présentées dans ce reporting sont :

- Encours moyens mensuel
- Encours moyen exercice
- Evolution de l'encours du mois n de l'année N par rapport aux mois n de l'année N-1
- Evolution de l'encours de l'exercice N par rapport à l'exercice N-1
- Tendance de l'évolution mensuelle par rapport au mois précédent

Il est à noter que le reporting papier initial ne correspond qu'à la vue principale d'évolution des encours. Les autres vues étaient souvent réalisées par des secrétaires commerciales, par certains commerciaux ou par les analyses. L'intégration de SIGMA doit permettre de standardiser ces vues et de réduire le temps de retraitement des données au profit du temps d'analyse.

Quatre vues principales sont proposées qui correspondent aux quatre axes d'analyse :

Vue	Ligne	Colonne	Profondeur
Evolution des encours	Produit	Mesures	Temps, Structure, Marché
Evolution mensuelle	Produit	Temps	Mesures, Structure, Marché
Comparaison structure	Structure	Mesures	Temps, Produit, Marché
Comparaison marchés	Marché	Mesures	Temps, Produit, Structure

Quelque soit la vue, ce sont les mêmes données qui sont présentées mais sous un angle de vue différent. Les mesures présentées sont les encours du mois et de l'exercice, leurs évolutions et la tendance du mois représentée par une flèche.

Chaque taux d'évolution fait de plus l'objet d'un color-coding permettant de se situer parmi son niveau de structure : en rouge si inférieur au premier quartile, en bleu si supérieur au quatrième quartile.

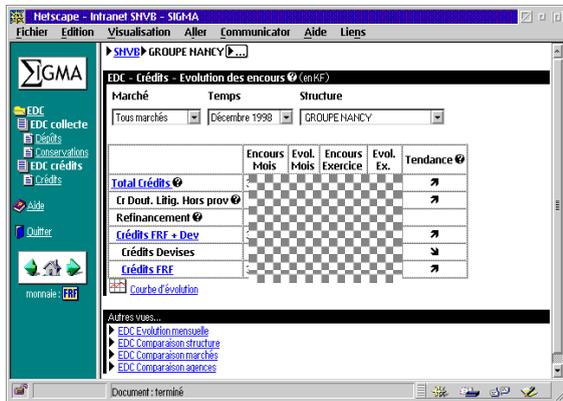
A ces quatre vues principales correspondent cinq vues graphiques pour représenter ces mêmes données :

Vue	Abscisse	Ordonnée	Profondeur
Evolution des encours - Courbe	Temps	Mesures (encours)	Produit, Structure, Marché
Evolution mensuelle - Histogramme	Temps	Mesures (taux)	Produit, Structure, Marché
Comparaison structure - Histogramme	Structure	Mesures (taux)	Temps, Produit, Marché
Comparaison marchés - Histogramme	Marché	Mesures (taux)	Temps, Produit, Structure
Comparaison marchés - Camembert	Marché	Mesures (encours)	Temps, Produit, Structure

Enfin, deux vues ont été rajoutées à la demande des utilisateurs, une première vue correspond à une évolution mensuelle de la structure : La structure en ligne, les mois en colonne et les dimensions Mesures, Marché et Produit en profondeur, ainsi qu'une dimension comparaisons agence qui permet de comparer entre elles toutes les agences d'une Direction Régionale sans avoir à passer par le niveau Secteur.

Chaque vue correspond en fait à quatre vues différentes qui correspondent à quatre sélections de produits : une par hiérarchie, et une « chiffres clés » pour présenter une vue globale.

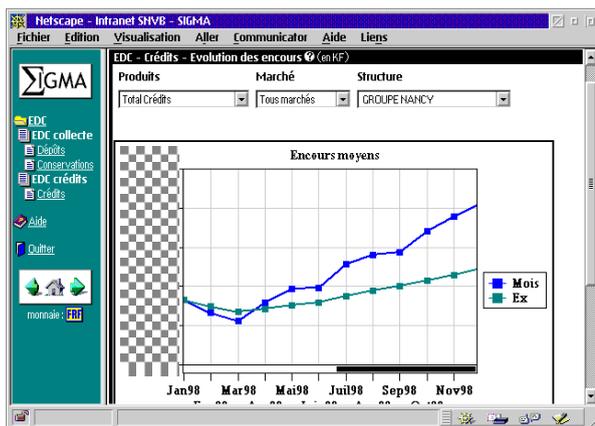
Evolution des encours



Ligne : Production, un clic sur un lien permet de déplier ou de replier la hiérarchie produit
 Colonne : Mesures, encours mensuelle, évolution de l'encours mensuel, encours de l'exercice, évolution de l'encours de l'exercice, flèche de tendance
 Profondeur : Marché, Temps, Structure

Liens vers une courbe d'évolution des encours et vers les autres vues

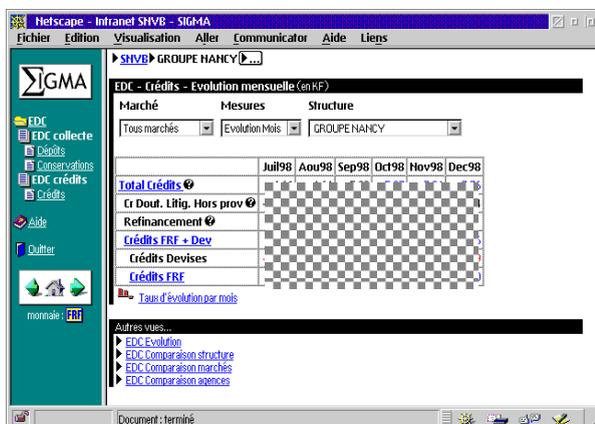
Evolution des encours - Courbe



Abscisse : le temps, on peut zoomer en arrière ou en avant pour avoir une vue générale ou plus détaillée
 Ordonnée : Mesures, les encours du mois et de l'exercice
 Profondeur : Produit, Marché, Structure

Liens vers le tableau d'évolution des encours et vers les autres vues

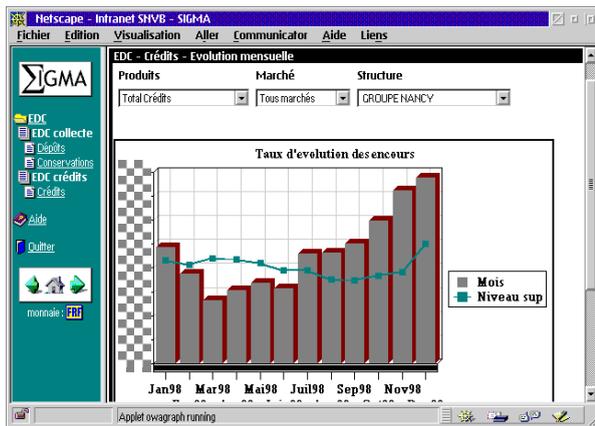
Evolution mensuelle



Ligne : Production, un clic sur un lien permet de déplier ou de replier la hiérarchie produit
 Colonne : Temps, les 6 derniers mois
 Profondeur : Marché, Mesures (taux d'évolution du mois et de l'exercice), Structure

Liens vers un histogramme des taux d'évolutions et vers les autres vues

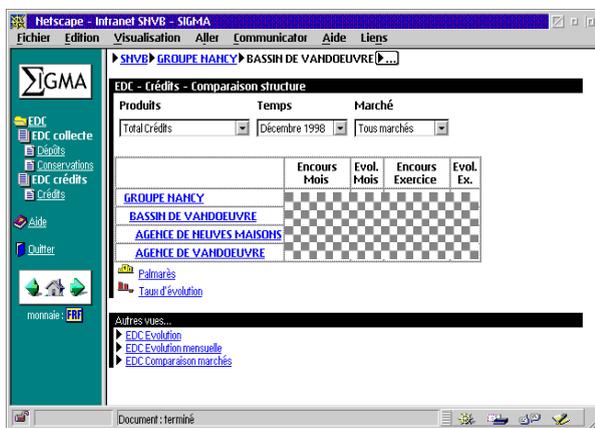
Evolution mensuelle - Histogramme



Abscisse : Temps avec possibilité de zoom
 Ordonnée : Mesures, évolution du mois ou de l'exercice + une courbe représentant l'évolution du niveau de structure supérieur
 Profondeur : Produit, Marché, Structure

Liens vers le tableau d'évolution mensuelle et vers les autres vues

Comparaison structure



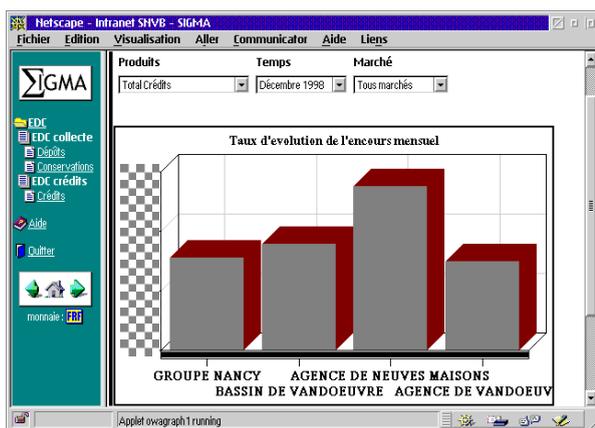
Ligne : Structure, on voit toujours la position et les niveaux supérieurs et inférieurs de la structure.

Colonne : Mesures, encours mensuel, évolution de l'encours mensuel, encours de l'exercice, évolution de l'encours de l'exercice
 Profondeur : Produit, Temps, Marché

Possibilité de demander un palmarès de la structure.

Liens vers une courbe d'évolution des encours et vers les autres vues

Comparaison structure - Histogramme

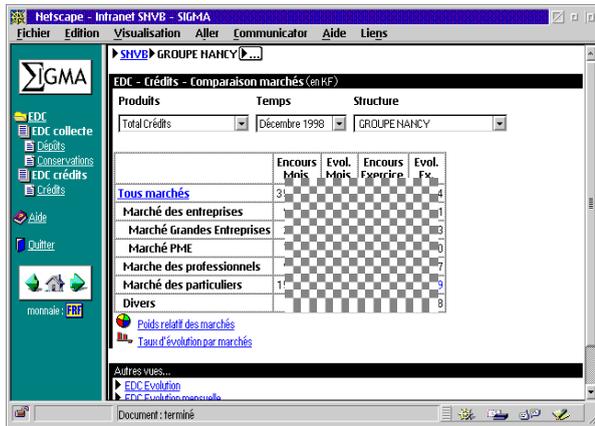


Abscisse : Structure avec possibilité de zoom, on présente la position, le niveau supérieur et le niveau inférieur de la structure

Ordonnée : Mesures, taux d'évolution mensuelle
 Profondeur : Produit, Temps, Marché

Liens vers le tableau de comparaison structure et vers les autres vues

Comparaison Marchés



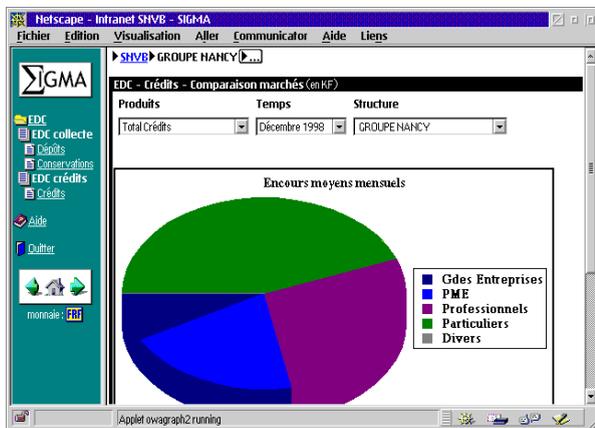
Ligne : Marchés

Colonne : Mesures, encours mensuel, évolution de l'encours mensuel, encours de l'exercice, évolution de l'encours de l'exercice

Profondeur : Produit, Temps, Structure

Liens vers camembert représentant la dispersion des encours, un histogramme représentant l'évolution et vers les autres vues

Comparaison marchés - Camembert



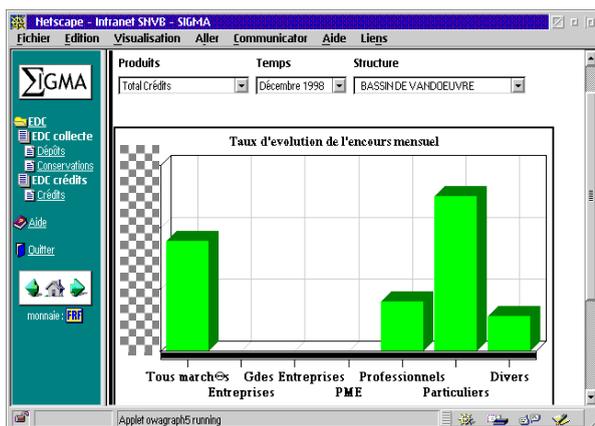
Mesure : encours moyens mensuels

Parts : Marchés

Profondeur : Produit, Temps, Structure

Liens vers le tableau de comparaison des marchés, un histogramme représentant l'évolution des encours et vers les autres vues

Comparaison marchés - Histogramme



Abscisse : Marchés

Colonne : Mesures, Taux d'évolution

Profondeur : Produit, Temps, Structure

Liens vers le tableau de comparaison marchés, un camembert l'évolution des encours et vers les autres vues

Comparaison agences

Produits	Temps	Marché	Encours	Evol.	Encours	Evol.
			Mois	Mois	Exercice	Ex.
GROUPE HANCY						
HANCY ESPACE MAGNOT						
HANCY ESPACE ENTREP.						
AGENCE DE BACCARAT						
AGENCE DE LUNEVILLE						
AGENCE DE HEUFCHATEAU						
AGENCE DE PONT A MOUSSON						
AGENCE DE ST NICOLAS DE PORT						
AGENCE DE TOUL						
AGENCE DE HEUVES MAISONS						
AGENCE DE VANDOEUVRE						
AGENCE DE LAXOU CHLE BOEUF						

Ligne : Structure : un groupe et toutes ses agences

Colonne : Mesures, encours mensuel, évolution de l'encours mensuel, encours de l'exercice, évolution de l'encours de l'exercice

Profondeur : Produit, Temps, Marchés

Possibilité d'obtenir un classement des agences.

Liens vers les autres vues

Lorsqu'on est au niveau du portefeuille, on a de plus accès à une requête permettant d'accéder au niveau client, on doit fournir alors le critère de recherche : plus grosse variation de l'encours mensuel par rapport au mois précédent, par rapport au même mois de l'exercice précédent, par rapport au 31/12 ou bien de l'encours de l'exercice par rapport à l'exercice précédent, on fournit également le produit sur lequel portera la recherche.

On obtient alors une liste des 10 premiers clients répondant au critère, on peut ensuite demander les 10 suivants si besoin est.

Le critère de recherche est rappelé au dessus du tableau avec la tranche de clients que l'on affiche et le nombre total de clients répondant à la requête.

EDC - Total Dépôts - Justification par le niveau client
 Les plus fortes variations pour Variation en volume mois N / mois N-1
 Produit sur lequel la sélection a été effectuée : Total Dépôts
 Nombre de personnes trouvées : 610 - Tranche affichée : 1 - 10

Produits
 Total Dépôts

	Encours Mois	Evol / Mois prec	Encours exercice	Evol / Exer prec	Evol / mois exer prec	Evol / 31-12
GE	291 126	-200 978	442 811	154 311	-11 903	154 311
AR	616 152	-167 345	946 781	471 764	153 494	471 764
A	379 090	-124 013	475 248	271 994	147 168	271 994
	195 948	-110 545	208 048	170 740	133 290	170 740
	1990 239	-96 868	2 353 015	1 062 712	1 176 597	1 062 712
E	140 119	93 859	71 156	-20 347	63 694	-20 347
	399 339	88 475	335 563	34 541	66 381	34 541
HD	769 994	-87 510	774 159	202 629	-7 229	202 629
RA	69 490	-85 670	64 000	37 229	56 665	37 229
IQ	223 342	78 272	114 892	9 428	154 723	9 428

Les 10 suivants

IV.6. Evolutions de Sigma

Le but premier du projet SIGMA, fournir de l'information de pilotage dans tout le réseau, a été atteint. Il reste aujourd'hui à fournir à des utilisateurs plus confirmés tels que des analystes du marketing ou du contrôle de gestion, un système qui leur permette de construire leurs propres états. Ce système doit donc permettre un accès facilité à toutes les informations ainsi qu'aux métadonnées. Enfin, de nouvelles procédures doivent être construites afin de permettre la diffusion facile de ces états.

Une autre voie qui reste à explorer dans le cadre du projet SIGMA est une étude pour faire évoluer le système afin d'en faire profiter également les activités d'infocentre et de datamining. L'architecture très modulaire du système devrait faciliter ces évolutions qui ne se feront alors que par adjonction de modules spécifiques.

Le domaine de l'infocentre nécessite un environnement plus riche que l'environnement proposé actuellement pour les restitutions. De plus, la nature même des requêtes infocentres fait qu'il est très difficile de les prévoir, les responsables de l'infocentre devraient donc avoir accès non seulement aux bases relationnelles du système mais également à l'entrepôt de données.

L'aspect DataMining est également très important mais porte plus sur un problème d'outils. L'offre est de plus en plus importante (SAS, IBM Dataminer...) et se présente sous forme de module qui s'intègre très facilement dans un système comme SIGMA en venant simplement se "brancher" sur l'entrepôt de données.

Conclusion et Perspectives

Le travail présenté dans ce mémoire, réalisé dans le cadre d'un Diplôme de Recherche Technologique de l'Université Nancy 2 en collaboration avec la banque SNVB avait deux objectifs : un objectif de veille technologique et de transfert de connaissances entre l'université et la banque pour tout ce qui concerne les systèmes décisionnels et un objectif d'application de ces connaissances dans le cadre du développement d'un système de pilotage. Ces deux objectifs ont été atteints, il existe aujourd'hui au sein de la banque SNVB une réelle expertise dans ce domaine et le système de pilotage est opérationnel.

Des études plus théoriques sur les modèles et les méthodes pour construire un système décisionnel ont également permis de proposer pour cela un langage de représentation graphique et des heuristiques pour la conception.

On notera cependant que ce travail reste très tourné vers les produits de la gamme d'Oracle Express, il serait intéressant d'approfondir ces concepts et méthodes pour les dégager de cet aspect logiciel. De plus, on assiste ces dernières années à un rapprochement des concepts de chacun des produits présents sur le marché qui devrait conduire à certaines normes et faciliter ainsi le développement de méthodes comme cela a été le cas pour les bases de données relationnelles. Cette perspective ouvre la voie à de nouveaux travaux de recherche qui permettraient de synthétiser ces concepts et de ne garder que le meilleur de chaque modèle et de chaque méthode.

D'autre part, des études sur les autres composants du système décisionnel restent à effectuer pour obtenir un schéma d'architecture plus global et plus cohérent, principalement pour les fonctions d'Infocentre et de DataMining souvent relayées en secondes place ou détachées du système.

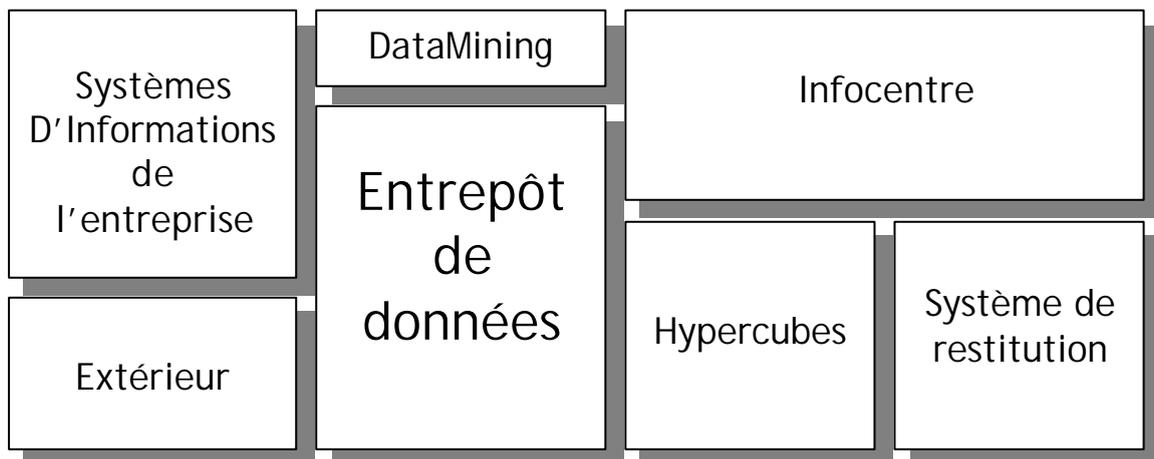


Table des illustrations

Figure 1. Structure d'un système décisionnel.....	7
Figure 2. Information Center.....	8
Figure 3. Infocentre.....	8
Figure 4. DataWarehouse.....	8
Figure 5. Dimension.....	12
Figure 6. Mesures.....	12
Figure 7. Vue relationnelle / Vue Multidimensionnelle.....	13
Figure 8. Schéma Dimension x Mesure.....	13
Figure 9. Notion de mesure dans un hypercube.....	13
Figure 10. Constante.....	14
Figure 11. Propriété.....	14
Figure 12. Méta-modèle des BDM.....	14
Figure 13. Relation.....	15
Figure 14. Hiérarchie Père/Fils.....	15
Figure 15. Méta-modèle des BDM.....	16
Figure 16. Représentation graphique d'une dimension.....	16
Figure 17. Représentation graphique d'une variable.....	16
Figure 18. Représentation graphique d'une formule.....	17
Figure 19. BDM Encours Dépôts Crédits.....	18
Figure 20. Modèle multidimensionnel.....	18
Figure 21. Utilisation d'une hiérarchie.....	20
Figure 22. Utilisation d'une relation.....	20
Figure 23. Les systèmes d'informations de l'entreprise.....	31
Figure 24. SI-S et S-IS.....	32
Figure 25. Données éparses non contrôlées.....	34
Figure 26. Données éparses contrôlées.....	34
Figure 27. Dimension conjointe.....	35
Figure 28. Typologie des métadonnées.....	38
Figure 29. Schéma de la base des métadonnées.....	39
Figure 30. Schéma de la BDM de gestion des profils.....	41
Figure 31. Architecture du système SIGMA.....	50

Bibliographie

- ¹ Bill Inmon - Building a Data Warehouse - 1991
- ² Hubert TARDIEU - Système d'Information - Stratégique et Système - d'Informations Stratégiques
- ³ Judy Rawls, Multitiered Data Warehouses, 1998
- ⁴ Ralph Kimball, Is ER Modeling Hazardous to DSS, Data Warehouse Architect Octobre 1995
- ⁵ Neil Raden, Modeling A DataWarehouse, CMP Janvier 1996
- ⁶ Ralph Kimball, A Dimensional Modeling Manifesto, Août 1997
- ⁷ Colin While, Managing Distibuted Data Warehouse Meta Data, DM Review Magazine Février 1999
- ⁸ Nigel Pendse, Multidimensionnal data structures, Janvier 1999
- ⁹ Wilburt Juan La Bio, Dallan Quass, Brad Adelberg - Physical Database Design for DataWarehouses - Université de Stanford 1997¹⁰ Oracle Education, Cours méthodologie, 1998
- ¹¹ Ralph Kimball, Entrepôts de Données : Guide pratique du concepteur de DataWarehouse, Editions Wiley 1997
- ¹² Hubert Tardieu - Système d'Information - Stratégique et Système - d'Informations Stratégiques
- ¹³ Odile Foucaut, Odile Thiéry, Kamel Smaili - Conception des systèmes d'information et programmation événementielle - InterEditions 1996
- ¹⁴ Alan Perkins - White Paper - Data Warehouse Architecture : A Blueprint for Success - 1999
- ¹⁵ Ralph Kimball - Meta Meta Data Data - DBMS, Mars 1998

Résumé

La gestion des informations de pilotage est cruciale pour l'entreprise. Aujourd'hui, trop d'entreprises continuent à ne confier ces informations qu'à une certaine élite. Les systèmes de pilotage de demain devront permettre de démocratiser l'accès à ces informations. Chaque employé pourra prendre connaissance de manière transparente de chaque information qui le concerne et l'intéresse.

Ainsi apparaît dans les entreprises la superposition de deux systèmes : un système de production, servant à l'activité, et un système de pilotage servant à tous les niveaux pour gérer au mieux cette activité. Si le premier est vital pour la survie de l'entreprise, le deuxième est crucial pour rester compétitif.

Ces systèmes de pilotage ou décisionnels utilisent une nouvelle forme de bases de données, les bases des données multidimensionnelles.

Dans ce travail, nous proposons un modèle général pour la conception des bases de données multidimensionnelles avec un langage graphique adapté. Nous présentons également des heuristiques pour leur démarche de conception. Enfin nous montrons l'intérêt de telles bases dans l'architecture globale d'un système de pilotage pour gérer les informations internes du système de pilotage selon le même modèle que les données elles-mêmes.

L'ensemble des résultats proposés résulte de l'élaboration du système décisionnel de la Banque SNVB basé sur un entrepôt de données en DB2 sur site central et sur les produits de la gamme Oracle Express pour le système de restitutions des tableaux de bord dans tout le réseau à travers un Intranet de pilotage.

Mots clés : Entrepôt de données, système décisionnel, système de pilotage, systèmes d'informations stratégique, tableau de bord, base de données multidimensionnel...

Keywords: DataWarehouse, Executive Information System, Reporting, Briefing Borad, Multidimensional DataBase...