

Option OLAP d'Oracle Database 10g

*Quand utiliser l'option OLAP pour améliorer
le contenu et les performances d'une
application de Business Intelligence*

*Livre blanc Oracle
Juin 2005*

Option OLAP d'Oracle Database 10g

En résumé.....	3
Oracle Business Intelligence	4
Plate-forme Business Intelligence	4
Outils et applications Business Intelligence.....	5
Oracle Business Intelligence Beans.....	5
OracleBI Discoverer.....	5
OracleBI Spreadsheet Add-In	6
Oracle Enterprise Planning and Budgeting.....	7
Outils de génération d'états.....	7
Option OLAP d'Oracle Database 10g.....	7
Modèle dimensionnel.....	8
Organisation du modèle dimensionnel	8
Expression des requêtes et navigation dans les données.....	9
Définition de calculs	11
Moteur et types de données multidimensionnels.....	12
Interfaces d'interrogation.....	13
API OLAP.....	13
Interface SQL avec les types de données multidimensionnels	13
Quand utiliser les types de données multidimensionnels Oracle	15
Calculs supplémentaires possibles avec les types de données multidimensionnels	15
API de l'espace de travail analytique.....	15
DML OLAP.....	17
Disponibilité des calculs intégrés pour les applications SQL.....	17
Avantages des types de données multidimensionnels en matière de performances.....	18
Types de données multidimensionnels spécialisés	19
Influence des types de requête sur les performances.....	19
Génération d'états analytiques.....	21
Savoir quand utiliser l'option OLAP pour améliorer les performances.....	22
Exemple d'application	25
Outils d'administration des espaces de travail analytiques	28
Analytic Workspace Manager	29
Oracle Warehouse Builder.....	30
Conclusion.....	31

EN RESUME

L'option OLAP d'Oracle Database fournit un modèle dimensionnel adapté aux types de requêtes et d'analyses communs aux applications de business intelligence et aux autres types d'applications analytiques. Oracle prend en charge le modèle dimensionnel aussi bien avec les types de données relationnels qu'avec les types de données multidimensionnels.

Le modèle dimensionnel logique est commun aux deux implémentations. Il décrit les données en termes facilement compréhensibles par les utilisateurs des applications de business intelligence. Il structure les données et comporte des éléments tels que des dimensions, des hiérarchies, des attributs, des cubes et des indicateurs. Grâce à ce modèle, les utilisateurs peuvent facilement interroger les données, y naviguer et définir des calculs complexes incluant des fonctions mathématiques, statistiques, de séries temporelles et prédictives, des prévisions par exemple.

Les fonctionnalités et les avantages qu'apporte le modèle dimensionnel permettent à l'option OLAP de prendre en charge des outils et applications de business intelligence de haut niveau, tels qu'OracleBI Discoverer OLAP, OracleBI Spreadsheet Add-In et Oracle Enterprise Planning and Budgeting, ainsi que des outils et applications de business intelligence de fournisseurs tiers, tels que Business Objects Enterprise et Cognos ReportNet.

Une fois sélectionnée une application prenant en charge le modèle dimensionnel, il faut choisir entre les types de données relationnels et multidimensionnels.

Les types de données relationnels conviennent particulièrement bien à certains usages des applications de business intelligence, notamment lorsque les types de requête sont relativement prévisibles et qu'il n'y a pas de calculs complexes. Dans ces conditions, il est possible d'optimiser l'interrogation des données relationnelles à l'aide de techniques telles que les requêtes en étoile, le précalcul et les tables d'agrégats ou les vues matérialisées.

Lorsque les types de requête des utilisateurs finaux deviennent moins prévisibles et que les besoins en calcul augmentent, il est plus difficile d'optimiser un entrepôt de données relationnelles pour obtenir d'excellentes performances d'interrogation. Dans ce cas, les types de données multidimensionnels offerts par l'option OLAP peuvent être beaucoup plus avantageux en termes de performances, tant pour la préparation que pour l'interrogation des données. L'option OLAP convient particulièrement pour les applications devant prendre en charge des requêtes adhoc.

Ce livre blanc présente les avantages du modèle dimensionnel ainsi que l'architecture, les fonctionnalités et les avantages de l'option OLAP, et indique quels sont les cas les plus favorables pour l'implémentation du modèle dimensionnel à l'aide de l'option OLAP et des types de données multidimensionnels.

ORACLE BUSINESS INTELLIGENCE

La ligne de produits Oracle Business Intelligence se compose d'une plate-forme, d'outils et de composants applicatifs. On peut l'utiliser dans sa totalité, en tant qu'ensemble complet de produits intégrés, ou bien n'utiliser que la plate-forme, en combinaison avec des outils tiers de business intelligence.

Oracle Database 10g comporte des types de données relationnels et multidimensionnels ainsi que des outils d'administration adaptés pour la business intelligence.

Oracle Database 10g constitue la base de la ligne de produits Oracle Business Intelligence et comporte des interfaces ouvertes d'administration et d'interrogation permettant le support complet des outils tiers de business intelligence.

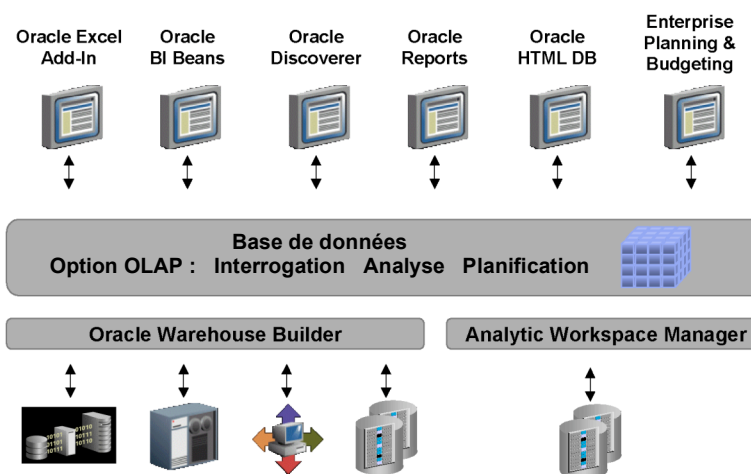


Plate-forme, outils et applications Oracle Business Intelligence

Plate-forme Business Intelligence

La base de données Oracle Database 10g avec l'option OLAP constitue la plate-forme de la solution Oracle Business Intelligence. Il s'agit d'une plate-forme évolutive, complète sur le plan fonctionnel, avec un data warehouse relationnel, un OLAP relationnel et des capacités OLAP multidimensionnelles.

Les capacités du data warehouse relationnel Oracle sont pleinement reconnues. Les data warehouses gérés par Oracle atteignent couramment des dizaines de téra-octets. Oracle a activement développé le langage SQL dans le sens du data warehouse et de la business intelligence. Par exemple, dans Oracle 8, Oracle9i et Oracle Database 10g, des fonctionnalités ont été ajoutées aux types de données relationnels et au SQL pour la gestion des agrégats et la réalisation de calculs propres aux applications de business intelligence. En outre, des progrès considérables ont été réalisés pour rendre excellentes les performances des requêtes dans les applications de business intelligence accédant à des types de données relationnels.

L'option OLAP apporte un modèle dimensionnel et des capacités d'interrogation pouvant être implémentés en tant qu'OLAP relationnel (ROLAP) ou OLAP multidimensionnel (MOLAP). Dans une implémentation ROLAP, l'option OLAP consiste en une couche de modélisation dimensionnelle au dessus des tables relationnelles et en une API d'interrogation dimensionnelle. Dans une implémentation MOLAP, les données sont stockées dans des types de données multidimensionnels dans la base de données Oracle. Ces types peuvent être interrogés via une API d'interrogation dimensionnelle et via le SQL.

La plate-forme Oracle Business Intelligence comporte également deux outils d'administration : Oracle Warehouse Builder et Oracle Analytic Workspace Manager.

Ces deux outils supportent le modèle dimensionnel des implémentations ROLAP et MOLAP.

La plate-forme Oracle Business Intelligence, et plus particulièrement l'option OLAP, sont traitées en détail dans la suite de ce livre blanc.

Outils et applications Business Intelligence

Oracle propose des outils de développement d'applications, des outils pour utilisateur final et des applications complètes de business intelligence. Ce sont notamment Oracle Business Intelligence Beans, OracleBI Discoverer Plus OLAP, OracleBI Spreadsheet Add-In et Oracle Enterprise Planning and Budgeting. Ensemble, ces outils et applications répondent à un large éventail de communautés d'utilisateurs et de besoins en matière de business intelligence.

Oracle Business Intelligence Beans

Oracle Business Intelligence Beans (OracleBI Beans) est un ensemble de composants applicatifs pouvant être utilisés pour créer des applications de business intelligence spécialisées ou pour inclure des fonctionnalités de business intelligence dans des portails ou d'autres applications d'entreprise. OracleBI Beans comporte des composants d'interface utilisateur : générateur de requêtes, générateur de calculs, graphiques et tableaux croisés par exemple, ainsi que des composants d'administration interne : catalogue, gestionnaires de métadonnées et de données par exemple.

Avec OracleBI Beans, un développeur d'applications peut créer une application de business intelligence sophistiquée avec peu de codage, en particulier s'il travaille dans l'environnement de développement Oracle JDeveloper (ce dernier contient des assistants de génération de code pour OracleBI Beans).

OracleBI Beans s'appuie sur le standard J2EE. Les applications peuvent être développées dans un environnement de développement Java quelconque puis déployées sur des serveurs d'applications tels qu'Oracle Application Server, Tomcat ou BEA WebLogic. Les composants BI Beans sont disponibles à la fois en tant que clients lourds Java et en tant que clients légers HTML.

La rapidité du développement d'applications à l'aide d'OracleBI Beans et d'Oracle JDeveloper, ainsi que la possibilité de distribuer des applications où sont intégrés des composants BI Beans sans coût supplémentaire, font d'OracleBI Beans un produit apprécié des intégrateurs de systèmes et des fournisseurs de logiciel indépendants présents sur des marchés verticaux.

Les applications Oracle Business Intelligence telles qu'Oracle Discoverer et OracleBI Spreadsheet Add-In sont elles mêmes construites à partir de composants BI Beans. De ce fait, toutes les applications Oracle Business Intelligence proposent des fonctionnalités de base similaires et ont des interfaces utilisateur cohérentes, et il peut en être de même pour les solutions tierces de business intelligence.

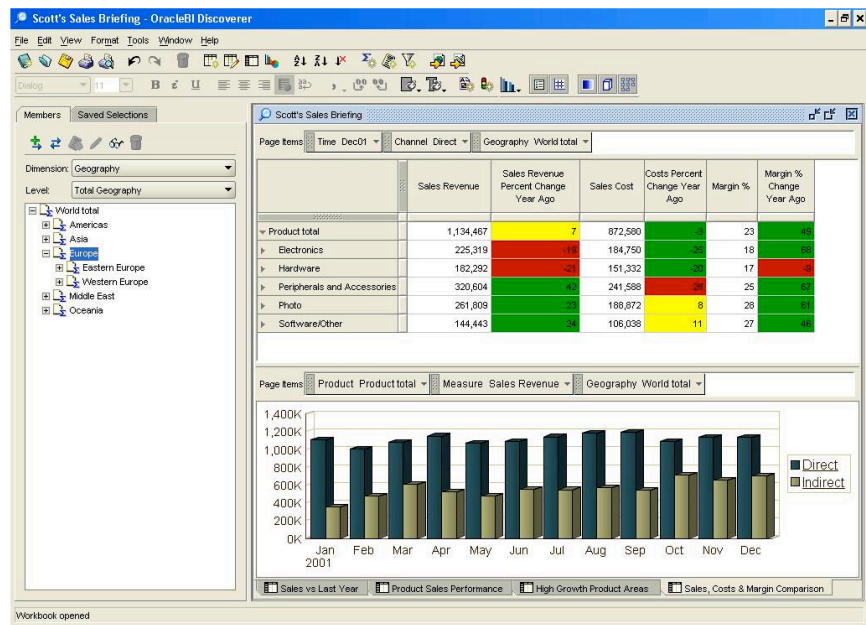
OracleBI Discoverer

OracleBI Discoverer est une famille d'outils de business intelligence permettant de définir et de diffuser des documents créés à des fins de reporting et d'analyse.

OracleBI Discoverer et OracleBI Discoverer OLAP permettent tous deux aux utilisateurs de définir des requêtes et des calculs, de concevoir des classeurs contenant des états et des graphiques, de naviguer dans les données de manière interactive et de partager des documents et des objets d'analyse avec d'autres utilisateurs.

OracleBI Discoverer Plus supporte le modèle relationnel et est mis en oeuvre avec des types de données relationnels. Il ne nécessite pas l'option OLAP. OracleBI Discoverer Plus OLAP supporte le modèle dimensionnel et est mis en oeuvre à l'aide de l'option OLAP. Les documents créés dans Oracle Discoverer Plus et Oracle Discoverer Plus OLAP peuvent être visualisés à l'aide de Discoverer Viewer et d'Oracle Portal.

OracleBI Discoverer Plus OLAP offre toute une gamme d'outils de génération de requêtes, de calcul et de visualisation supportant le modèle dimensionnel.



OracleBI Discoverer Plus OLAP : Analyse des tendances des ventes par produit.

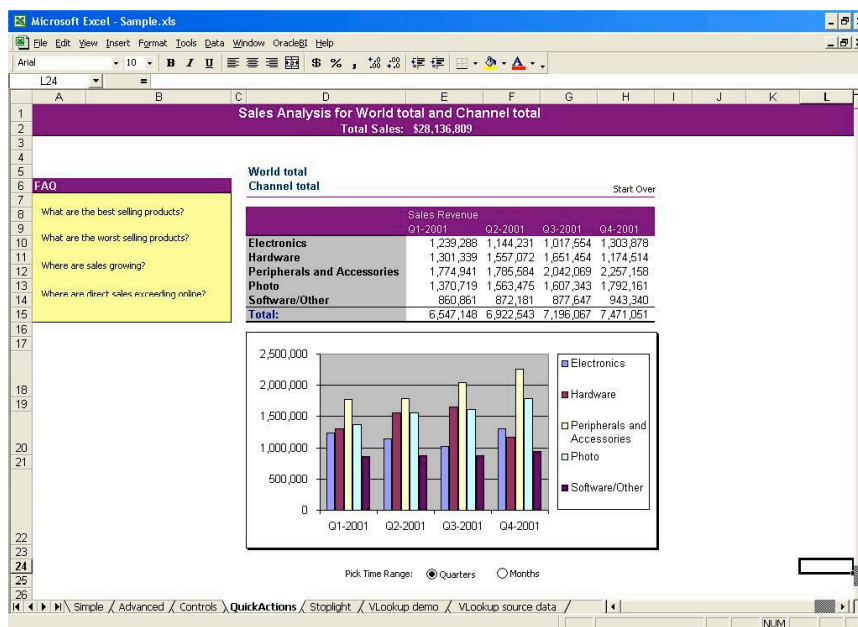
OracleBI Discoverer OLAP convient à tous les types d'utilisateur. L'utilisateur occasionnel apprécie sa facilité d'utilisation et la possibilité d'accéder aux documents préparés par d'autres utilisateurs. Les utilisateurs plus pointus disposent de toute la puissance et l'interactivité dont ils ont besoin pour analyser soigneusement de larges ensembles de données relationnelles et dimensionnelles, et peuvent définir leurs propres calculs.

OracleBI Spreadsheet Add-In

OracleBI Spreadsheet Add-In est conçu pour les utilisateurs qui préfèrent employer Microsoft Excel comme interface avec les données gérées par l'option OLAP. Il leur permet de présenter ces données dans des feuilles de calcul et des graphiques Excel. Ils peuvent alors pousser plus loin l'analyse en appliquant des fonctions Excel aux données et faire référence à ces dernières dans d'autres feuilles de calcul.

OracleBI Spreadsheet Add-In inclut un générateur de requêtes dimensionnelles et un générateur de calculs dimensionnels basés sur les composants OracleBI Beans.

OracleBI Spreadsheet Add-In permet aux utilisateurs d'employer Microsoft Excel comme interface pour les données gérées par l'option OLAP.



Types de données multidimensionnels Oracle interrogés via OracleBI Spreadsheet Add-In.

Oracle Enterprise Planning and Budgeting

Oracle Enterprise Planning and Budgeting (EPB), un composant d'Oracle Corporate Performance Management (CPM), est utilisé par les entreprises pour développer des processus d'analyse et de planification. Comme OracleBI Discoverer, EPB offre des outils évolués de navigation et d'analyse pour données dimensionnelles. Il comporte notamment des éléments structurels de workflow et de planification puisqu'il permet de définir des règles de gestion, des tâches et des ordonnancements, l'accent étant mis sur les processus d'élaboration de budgets et les prévisions.

Outils de génération d'états

Des outils généraux de génération d'états tels qu'Oracle Reports et Oracle HTML Database sont souvent employés pour accéder à diverses sources de données - systèmes transactionnels, data warehouses, business intelligence - dans différents contextes. L'option OLAP comportant une interface SQL avec les types de données multidimensionnels, ces outils ont pleinement accès aux données gérées par l'option OLAP et tirent parti des capacités de calcul de cette dernière et des gains de performances qu'elle apporte.

OPTION OLAP D'ORACLE DATABASE 10g

L'option OLAP d'Oracle Database 10g assure le support de l'interrogation et de l'analyse dimensionnelles. Elle fournit un modèle dimensionnel, un moteur de calcul complet, une API OLAP, un accès SQL aux types de données multidimensionnels et

des gains de performances à toutes sortes d'applications. Si chaque caractéristique est utile par elle-même, c'est leur regroupement qui fait de l'option OLAP une excellente plate-forme pour la plupart des applications de génération d'états et d'analyse adhoc.

Modèle dimensionnel

L'option OLAP offre la possibilité de présenter les données dans le contexte d'un modèle dimensionnel et d'implémenter ce modèle soit avec des data warehouses relationnels, soit avec des types de données multidimensionnels. Nous allons présenter ici les avantages intrinsèques du modèle dimensionnel, indépendamment de l'implémentation qui en est faite. Les avantages supplémentaires résultant de l'implémentation d'un modèle dimensionnel avec des types de données multidimensionnels Oracle sont traités plus loin dans ce livre blanc.

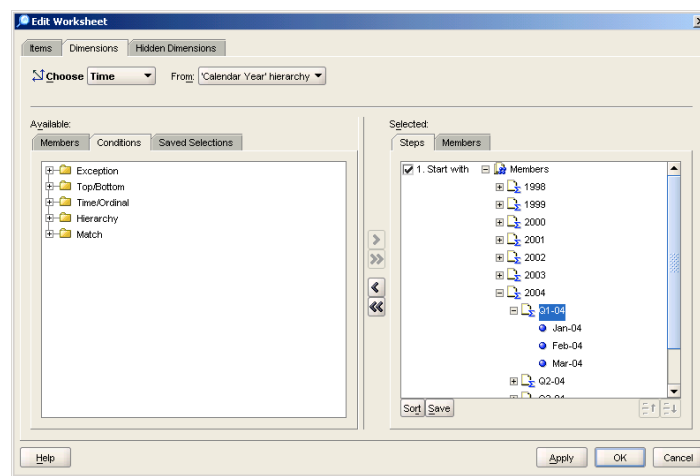
Organisation du modèle dimensionnel

Le modèle dimensionnel organise les données selon une ou plusieurs dimensions. Les groupes d'indicateurs partageant les mêmes dimensions et certaines autres caractéristiques sont appelés *cubes*. Par exemple, les indicateurs Ventes et Nombre d'unités vendues peuvent varier selon les dimensions Temps, Client, Produit et Canal de distribution. Comme ils partagent les mêmes dimensions, ils pourraient appartenir à un même cube.

Le modèle dimensionnel supporté par l'option OLAP est conforme aux standards du marché. Une dimension peut comporter un ou plusieurs *membres* (parfois appelés *éléments*), pouvant être organisés selon une ou plusieurs *hiérarchies*. Les hiérarchies peuvent ou non comporter des *niveaux* (par exemple, mois, trimestre et année). Des *attributs* peuvent être utilisés pour décrire les caractéristiques non hiérarchiques d'un membre de dimension (par exemple, couleur ou taille).

La plupart des modèles dimensionnels comportent une dimension Temps avec des membres tels que Janvier-2004, Février-2004, PremierTrimestre-2004 et Année-2004. Ces membres peuvent être disponibles aux niveaux Mois, Trimestre et Année, et être organisés selon les hiérarchies Année civile et Exercice fiscal. Un attribut non hiérarchique, tel que « Mois de l'année », peut s'appliquer aux membres de la dimension Temps au niveau Mois.

Les dimensions fonctionnent comme des index pointant sur les données factuelles. Elles peuvent comporter des hiérarchies, comportant ou non des niveaux.



Membres d'une dimension Temps affichés dans l'outil de requêtes OracleBI Discoverer

Ce qui s'appellerait un fait dans un data warehouse s'appelle *indicateur* dans le modèle dimensionnel. Un indicateur peut être stocké ou calculé à la volée. Les indicateurs appartiennent à des *cubes*. Un cube peut comporter un ou plusieurs indicateurs. Les cubes comportent généralement 3 à 10 dimensions, mais ils peuvent en avoir beaucoup plus ou n'en avoir qu'une seule. Outre les informations de dimension, la définition d'un cube comporte les règles d'agrégation utilisées par les indicateurs du cube. La règle d'agrégation la plus courante est la somme (SUM) mais l'option OLAP en accepte beaucoup d'autres : moyenne hiérarchique pondérée, premier ou dernier de hiérarchie, sommes mises à l'échelle, etc.

Un modèle dimensionnel peut comporter autant de dimensions, de niveaux, de hiérarchies, d'attributs, de cubes et d'indicateurs que nécessaire.

Expression des requêtes et navigation dans les données

Le modèle dimensionnel peut améliorer considérablement l'ergonomie d'une application en simplifiant la définition des requêtes, la navigation dans les données et la définition des calculs. En effet, les relations existant entre les éléments de données sont nativement intégrées dans le modèle dimensionnel. Les utilisateurs peuvent donc se concentrer sur les données plutôt que d'avoir à définir des relations complexes dans leurs requêtes.

Considérons l'interrogation suivante d'un utilisateur portant sur un jeu de données basé sur les ventes (notez les éléments dimensionnels) :

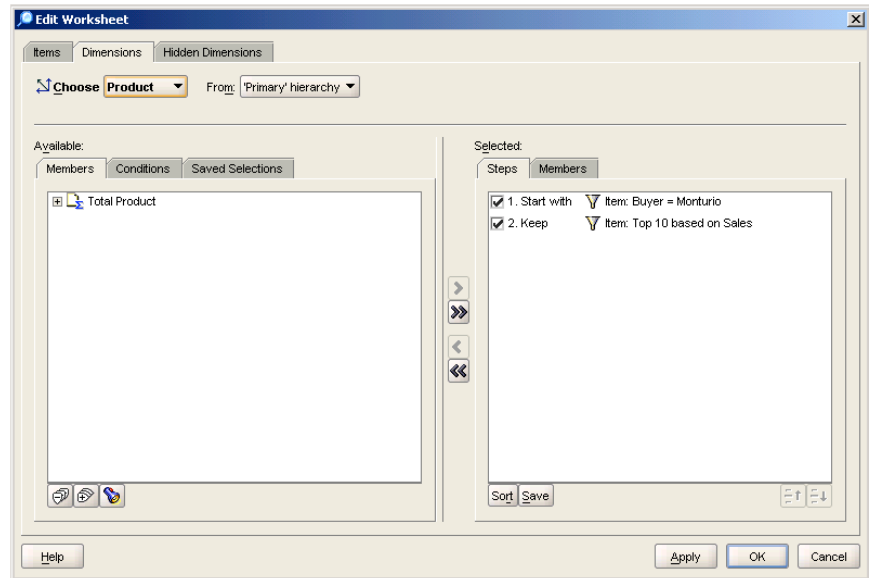
Sélectionner les indicateurs Ventes, Variation des ventes et Variation des ventes en pourcentage pour :

- les 10 meilleurs Produits sur la base de l'indicateur Ventes au niveau Articles, l'attribut Acheteur ayant la valeur « Monturio »,
- les 100 meilleurs Clients au niveau Comptes de la hiérarchie Segment de marché, l'indicateur Variation des ventes en pourcentage ayant une valeur supérieure à 30 %,
- tous les Canaux au niveau Canaux de distribution,
- toutes les périodes de Temps au niveau Trimestre pour les années 2003 et 2004 de la hiérarchie Exercice fiscal.

Les relations à l'intérieur des dimensions (hiérarchies, niveaux et attributs) et entre les dimensions et les indicateurs étant intégrées au modèle, l'utilisateur n'a besoin d'exprimer que les conditions portant sur les données. Il n'a pas à décrire dans sa requête les relations de jointures entre les dimensions et les faits par exemple.

La figure ci-dessous montre l'écran OracleBI Query Builder dans lequel sont effectuées les sélections de la requête pour la dimension Produit. Le modèle dimensionnel permet au générateur de requêtes de réaliser la sélection en deux étapes seulement et de l'exprimer dans un langage facilement compréhensible pour l'utilisateur lambda.

OracleBI Query Builder utilise les relations intégrées au modèle dimensionnel pour simplifier la définition des requêtes.



Utilisation d'OracleBI Query Builder dans Discoverer Plus OLAP pour sélectionner les 10 meilleurs produits de la dimension Produit sur la base des Ventes pour l'acheteur « Monturio »

Dans le modèle relationnel, cette requête nécessiterait une clause WHERE, une jointure, des fonctions d'agrégation et une clause GROUP BY. Dans le modèle dimensionnel, ces relations sont intégrées au modèle. L'utilisateur n'a à indiquer que les caractéristiques des données qui l'intéressent.

La navigation par niveau est réalisée à l'aide des relations hiérarchiques définies dans le modèle dimensionnel. Une fois que des membres ont été sélectionnés et affichés dans un état, l'utilisateur peut naviguer dans la dimension en développant des membres agrégés pour voir les éléments qui les composent ou, au contraire, en réduisant des membres pour revenir à un niveau agrégé ou à un total. Ce concept, illustré ci-après dans le contexte d'Oracle Discoverer, est une caractéristique du modèle dimensionnel en général et il est pris en charge dans OracleBI Beans et dans l'API OLAP.

Un état d'analyse des produits est affiché dans OracleBI Discoverer Plus OLAP. Grâce au modèle dimensionnel, les utilisateurs finaux peuvent facilement définir des états tels que celui-ci.

The screenshot shows the OracleBI Discoverer Plus OLAP interface. The main window displays a table titled "Product Sales, Trend, Rankings and Shares". The table has columns for Sales, Sales Percent Change Prior Period, Product Sales Rank to Parent, Product Sales Rank to Total, Product Sales Share to Parent, and Product Sales Share to Total. The data is organized hierarchically by product type and then by specific product models. The interface also includes a sidebar for navigation and a top menu bar.

	Sales	SALES PERCENT CHANGE PRIOR PERIOD	PRODUCT SALES RANK TO PARENT	PRODUCT SALES RANK TO TOTAL	PRODUCT SALES SHARE TO PARENT	PRODUCT SALES SHARE TO TOTAL
Total Product	130,276,515	41%	1	1	100%	100%
Hardware	117,736,092	39%	1	2	90%	90%
Memory	5,272,678	16%	5	16	4%	4%
CD/DVD	15,083,555	55%	3	6	13%	12%
Portable PCs	19,155,814	-13%	2	5	16%	15%
Desktop PCs	67,900,544	73%	1	3	58%	52%
Sentinel Standard	11,592,953	-4%	3	9	17%	9%
Sentinel Financial	42,630,941	175%	1	4	63%	33%
Sentinel Multimedia	13,676,650	16%	2	7	20%	10%
Monitors	4,346,492	36%	6	18	4%	3%
Modems/Fax	5,977,011	4%	4	14	5%	5%
Software/Other	12,540,422	57%	2	8	10%	10%

Analyse des ventes par produit après navigation dans la dimension Produit. Les indicateurs présentés avec des alerteurs de type feux tricolores sont la variation des ventes en pourcentage par rapport à la période précédente, la part par rapport au parent et la part par rapport au total.

Définition de calculs

Outre les avantages qu'il apporte en matière de définition de requêtes et de navigation dans les données, le modèle dimensionnel améliore également sensiblement la définition des calculs. Outre ses capacités de calcul dans le domaine des agrégations, le modèle dimensionnel embarque aussi des facilités permettant à un utilisateur final de créer des indicateurs personnels en toute simplicité. Le modèle dimensionnel inclut des règles d'agrégation se présentant sous la forme de relations hiérarchiques et d'opérateurs d'agrégation. Par exemple, la valeur des Ventes pour le Premier Trimestre est en général la somme des Ventes pour Janvier, Février et Mars. Les relations hiérarchiques définissent ce qui doit être agrégé, les opérateurs d'agrégation définissent la méthode d'agrégation.

Les indicateurs définis par l'utilisateur sont des extensions courantes dans un modèle dimensionnel. Selon l'application, ils peuvent être définis soit par l'administrateur de base de données, auquel cas l'utilisateur final les voit comme faisant partie du modèle de base, soit directement par l'utilisateur final de l'application de business intelligence. Par exemple, l'administrateur peut définir dans le modèle de base des indicateurs calculés tels que la variation par rapport à la période précédente ou les parts de marché. Ces indicateurs sont alors disponibles pour tous les utilisateurs. Par ailleurs, l'utilisateur final, peut également définir ses propres indicateurs, qui restent alors sa propriété.

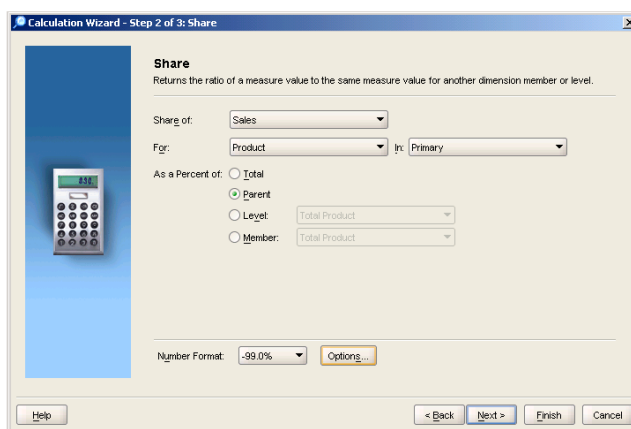
L'implémentation Oracle du modèle dimensionnel présente tous les indicateurs calculés comme des objets à part entière du modèle. Autrement dit, du point de vue de l'utilisateur final, ils se présentent de la même façon que les indicateurs stockés. À ce titre ils peuvent d'ailleurs être imbriqués les uns avec les autres.

Les indicateurs calculés suivants sont couramment utilisés :

- Indicateurs de part de marché : ils représentent le rapport entre les valeurs de deux membres de dimension. Par exemple, l'indicateur « Part de marché du produit en terme de Ventes par rapport au parent » peut être défini comme le rapport entre un produit et son parent immédiat dans une hiérarchie de la dimension Produit.
- Indicateurs de variation en pourcentage : ils représentent la variation en pourcentage d'une valeur entre deux périodes de même niveau de la dimension Temps..
- Indicateurs de classement : ils représentent le classement, de 1 à n, d'un membre de dimension, soit dans la totalité de la dimension, soit au sein d' un niveau soit dans une fratrie (on parle également d'indicateur *Rang par rapport au parent*).

La figure ci-dessous montre l' assistant OracleBI Calculation Builder (que l'on retrouve dans les BI Beans, dans Discoverer et dans Spreadsheet Add-In) dans lequel on définit un indicateur calculant la part des ventes d'un produit en tant que pourcentage des ventes de son produit parent. C'est parce que l'assistant de calculs repose sur un modèle dimensionnel que la définition d'un tel indicateur peut être réalisée directement par un utilisateur final.

OracleBI Beans Calculation Builder, utilisé dans la plupart des applications Oracle Business Intelligence, permet aux utilisateurs finaux de définir de nouveaux indicateurs puissants sans programmation.



Définition d'un indicateur de part de marché dans le modèle dimensionnel

Moteur et types de données multidimensionnels

Le moteur multidimensionnel de l'option OLAP est un moteur de calcul complet basé sur le modèle dimensionnel. D'un point de vue fonctionnel, on peut le comparer aux bases de données multidimensionnelles autonomes classiques. Il en diffère par son intégration complète au processus Oracle Database. Il bénéficie ainsi des capacités d'Oracle Database en matière de sécurité, de haute disponibilité, de montée en charge et de stockage. Par exemple, le moteur et les types de données multidimensionnels sont entièrement compatibles avec Oracle Real Application Clusters et le calcul en réseau (grid computing).

Les types de données multidimensionnels Oracle sont des objets stockés dans Oracle Database et nommés *espaces de travail analytiques*. Les espaces de travail analytiques

sont des ensembles de types de données multidimensionnels tels que les dimensions, les cubes et les indicateurs. Ces types de données sont réellement multidimensionnels : les indicateurs sont stockés dans des structures de type matrice et non dans des structures relationnelles.

Les données sont généralement réparties dans les différents espaces de travail analytiques d'après leur domaine. Par exemple, un espace de travail analytique peut être réservé aux données relatives aux ventes et un autre aux données financières. Une même requête peut faire référence aux données de plusieurs espaces de travail analytiques.

Les espaces de travail analytiques sont stockés dans des tablespaces et fichiers de données standard Oracle. Ils peuvent donc être administrés à l'aide des mêmes fonctions standard Oracle que les autres types de données de la base pour tout ce qui concerne la sécurité, les espaces disques, la haute disponibilité et les récupérations après sinistre.

Interfaces d'interrogation

L'option OLAP propose deux interfaces d'interrogation : l'API OLAP et le SQL. L'API OLAP présente un modèle dimensionnel alors que le langage SQL présente les données dans le contexte d'un modèle relationnel.

API OLAP

L'API OLAP est une API d'interrogation dimensionnelle complète. Elle présente l'application selon un modèle dimensionnel et fournit des méthodes de définition et d'exécution de requêtes et de définition de calculs. L'API OLAP accepte comme sources de données à la fois les types de données relationnels et les types de données multidimensionnels (espaces de travail analytiques). Par conséquent, elle supporte aussi bien l'OLAP relationnel que l'OLAP multidimensionnel.

Dans un contexte ROLAP, l'API OLAP apporte le modèle dimensionnel, des méthodes d'interrogation et un générateur de requêtes optimisé. Ce dernier est spécifiquement adapté à chaque version de la base de données Oracle afin que les performances soient les meilleures possibles. Par exemple, il tire parti de la clause SQL MODEL d'Oracle Database 10g pour traiter efficacement certains calculs d'indicateurs dynamiques.

Dans un contexte MOLAP, l'API OLAP interroge l'espace de travail analytique à l'aide d'un SQL optimisé pour les types de données multidimensionnels. Le SQL étant le langage d'interrogation de base pour les deux types de données, l'API OLAP peut mélanger des types de données relationnels et multidimensionnels dans la même requête. Ceci est l'un des nombreux exemples illustrant les avantages que les applications de business intelligence peuvent tirer d'une base de données relationnelle-multidimensionnelle intégrée.

L'API OLAP est utilisée par les BI Beans, Discoverer Plus OLAP et Spreadsheet Add-In.

Interface SQL avec les types de données multidimensionnels

Toutes les applications de business intelligence ne sont pas basées sur le modèle dimensionnel et sur des API dimensionnelles telles que l'API OLAP. Il existe de nombreuses applications basées sur le SQL qui permettent de présenter des données dans des états, des graphiques ou sous d'autres formats, et ces applications peuvent être très utiles pour présenter les données stockées dans des espaces de travail analytiques.

Les outils de reporting interactif tels que Business Objects et Cognos ReportNet constituent l'une de ces classes d'applications. Grâce à ces outils, qui ne sont pas basés sur le modèle dimensionnel, les utilisateurs peuvent définir des documents, définir des requêtes et naviguer dans les données. Les générateurs d'états constituent une autre classe de ces applications. Ils sont spécialisés dans le formatage de haut niveau et dans la génération d'états par lots ; le langage d'interrogation qu'ils utilisent est le SQL. Ces deux classes d'applications peuvent tirer parti des capacités de calcul et des performances apportées par le moteur multidimensionnel et les espaces de travail analytiques.

Pour que les applications qui ne sont ni des applications dimensionnelles ni des applications de business intelligence puissent plus facilement utiliser des données multidimensionnelles, l'option OLAP propose une interface SQL avec les types de données multidimensionnels. Une application basée sur le SQL peut aborder les types de données multidimensionnels soit en les interrogeant directement, soit en interrogeant des vues relationnelles pointant sur les espaces de travail analytiques.

En définissant une vue relationnelle pour les types de données multidimensionnels, il est possible de masquer complètement la source des données de l'application. Autrement dit, l'outil peut générer du SQL standard par rapport à la vue et la base de données redirige automatiquement la requête vers les types de données multidimensionnels. Tous les calculs définis dans l'espace de travail analytique - par exemple, les agrégations et les calculs d'indicateur - apparaissent comme des données stockées pour l'application effectuant les requêtes. Le moteur multidimensionnel calcule automatiquement les données et les renvoie via le SQL. L'application bénéficie ainsi d'une puissance énorme, bien qu'elle ne soit pas basée sur le modèle dimensionnel et ne sache rien des types de données multidimensionnels.

Lorsqu'une application basée sur le SQL choisit de s'ouvrir aux types de données multidimensionnels, elle dispose de nombreuses opportunités d'interaction avec le moteur et les types de données multidimensionnels :

- Toutes les métadonnées de l'espace de travail analytique sont présentées via des catalogues système OLAP. Elles peuvent donc être accédées à l'aide du SQL.
- Des calculs exprimés en DML OLAP, qui est le langage des types de données multidimensionnels, peuvent être imbriqués dans des instructions SELECT.
- Des applications basées sur le SQL peuvent exécuter des instructions SELECT directement sur les types de données multidimensionnels à l'aide de la fonction de table OLAP_TABLE. Cette dernière redirige les requêtes SQL vers les types de données multidimensionnels et renvoie les données via des curseurs SQL comme si elles provenaient d'un objet relationnel.

Les interfaces offertes par l'option OLAP apportent une grande souplesse et une puissance énorme et font de l'option OLAP d'Oracle Database le serveur OLAP le plus ouvert et le plus accessible du marché.

QUAND UTILISER LES TYPES DE DONNEES MULTIDIMENSIONNELS ORACLE

Après avoir choisi un modèle de données dimensionnel, il faut décider s'il sera implémenté avec des types de données relationnels ou multidimensionnels. Les principaux facteurs déterminant le choix de l'implémentation sont les suivants :

- L'application a-t-elle besoin de calculs autres que ceux pris en charge par l'API OLAP et le SQL ?
- Quels sont les types d'utilisation récurrents dans les applications de business intelligence ?
- Quelles sont les exigences en matière de performances ?

Les types de données multidimensionnels offrent des possibilités de calcul plus étendues que les types de données relationnels et ils peuvent permettre d'améliorer considérablement les performances, tant pour la préparation des données que pour leur interrogation.

Calculs supplémentaires possibles avec les types de données multidimensionnels

L'API OLAP d'interrogation prend en charge les calculs couramment utilisés dans les applications de business intelligence, par exemple les séries temporelles, les parts de marché et les indices, les classements et les agrégations. Ces calculs répondent aux besoins de nombreuses applications de business intelligence et sont disponibles via l'API OLAP, que l'application soit implémentée avec les types de données relationnels ou multidimensionnels.

Parfois, cependant, des calculs plus complexes sont nécessaires. Ces calculs se répartissent généralement dans les catégories suivantes :

- Fonctions définies par l'utilisateur
- Allocations
- Prévisions
- Systèmes d'équations entre membres de dimension
- Flux de données complexes

Les calculs peuvent être intégrés à un espace de travail analytique soit via des formules de calcul OLAP, soit en effectuant des transformations des données en mode batch. Ces deux opérations se réalisent dans l'espace de travail analytique à l'aide de l'API d'espace de travail analytique ou du DML OLAP.

API d'espace de travail analytique

L'API d'espace de travail analytique, nouvelle en Oracle 10g, offre une interface de haut niveau pour l'administration des espaces de travail analytiques. Elle permet de définir un espace de travail analytique en termes de modèle dimensionnel logique, de correspondances avec des sources de données relationnelles, de définition de différents types de calculs. Elle supporte également les activités de gestion du cycle de vie telles que la construction et la régénération des cubes à partir des sources de données et l'exécution de calculs prédéfinis.

L'interaction entre les applications et l'API d'espace de travail analytique passe soit par Java soit par des documents AW/XML et par le PL/SQL. L'API d'espace de travail analytique est une API publique dont peuvent disposer les développeurs d'applications tant Oracle que tierces.

Les calculs définis par l'API d'espace de travail analytique peuvent être simples ou faire intervenir des flux de données entre différents indicateurs du modèle dimensionnel, implémentant chacun différents types de calcul. Un calcul simple est par exemple le calcul d'une part de marché. Quant à un flux de calculs, il peut consister à allouer des données de membres de niveau supérieur à des membres de niveau inférieur, à réaliser une prévision puis à agréger les données vers les membres de niveau supérieur.

L'API d'espace de travail analytique prend en charge les agrégations hiérarchiques, les calculs d'indicateurs, les allocations, les prévisions et les modèles.

Agrégations hiérarchiques

Les agrégations hiérarchiques sont des calculs qui permettent d'agréger au niveau d'un membre de la hiérarchie les données des membres qui lui sont rattachés.

L'option OLAP supporte un large éventail de méthodes d'agrégation dont la somme, les moyennes hiérarchiques pondérées et les sommes mises à l'échelle. Les méthodes d'agrégation peuvent varier selon la dimension. Par exemple, la méthode d'agrégation d'un indicateur Effectif peut être la somme pour une dimension Organisation et la moyenne des jours, des mois, des trimestres et des années pour une dimension Temps.

Calculs d'indicateurs

L'API d'espace de travail analytique peut être utilisée pour définir des fonctions d'analyse courantes, par exemple des calculs de séries temporelles, des parts de marché et des indices, des variations et des classements. Ces calculs sont intégrés au modèle dimensionnel sous forme d'indicateurs et exécutés dans une requête par le moteur multidimensionnel. Ce dernier peut mettre en œuvre ce type d'indicateurs avec une grande efficacité, même s'ils nécessitent des calculs interlignes, des jointures externes et des jointures entre différents cubes.

Allocations

Les allocations permettent de répartir les données depuis un membre agrégé d'une hiérarchie vers les membres qui lui sont rattachés. Par exemple, un système budgétaire peut utiliser les allocations pour distribuer les budgets du prochain exercice de chaque division aux différents groupes de produits puis des groupes de produits aux produits de base.

L'option OLAP propose un large éventail de méthodes de répartition dont les méthodes par copie (copie hiérarchique, minimum, maximum, premier, dernier), les méthodes de répartition à part égale (égalité, égalité hiérarchique) et les méthodes de répartition proportionnelle (y compris les répartitions pondérées).

Prévisions statistiques

L'option OLAP offre un système de prévision statistique complet. Les méthodes de prévision sont notamment la régression non linéaire, le lissage exponentiel simple, le

lissage exponentiel double et Holt-Winters. Des caractéristiques telles que la saisonnalité des données sont également prises en charge.

Modèles

Les modèles OLAP calculent la valeur de membres de dimension individuels, disposant chacun de leurs propres règles de calcul. Le moteur multidimensionnel ordonne automatiquement les calculs et supporte les équations simultanées. Les modèles sont souvent utilisés dans les applications financières, en particulier pour définir des calculs dans les dimensions non hiérarchiques telles que les plans comptables. Dans les dimensions non hiérarchiques, les valeurs des membres dérivent souvent de la fratrie plutôt que d'agrégations ou d'allocations.

DML OLAP

Le DML OLAP est le langage de programmation procédural du moteur et des types de données multidimensionnels. Il s'agit d'un langage de quatrième génération conçu spécialement pour décrire des calculs dans le contexte du modèle dimensionnel. Le DML OLAP peut être considéré comme l'équivalent dimensionnel du PL/SQL. Il s'adresse plutôt aux administrateurs de base de données et aux développeurs d'applications mais certains utilisateurs pointus l'emploient pour définir des calculs.

Le DML OLAP comporte des opérateurs logiques (IF, THEN, ELSE, CASE, etc.) pouvant être utilisés dans les calculs et les programmes pour diriger le flux de calcul. Il comporte également une riche bibliothèque de fonctions mathématiques, statistiques, financières, prédictives et de séries temporelles.

Le DML OLAP est souvent employé pour enrichir un espace de travail analytique avec des calculs qui vont au-delà de ce que permet de définir l'API d'espace de travail analytique.

Disponibilité des calculs intégrés pour les applications SQL

Un avantage significatif des types de données multidimensionnels est que le moteur multidimensionnel peut présenter l'espace de travail analytique comme étant entièrement calculé, que les indicateurs soient stockés ou calculés au moment de l'exécution. Dans la plupart des espaces de travail analytiques, la majeure partie des informations est calculée dynamiquement.

Le moteur multidimensionnel intègre chaque calcul au modèle dimensionnel sous forme d'objet de haut niveau. Il sait quelles sont les données qui ont été pré calculées et stockées et quelles sont celles qu'il faut calculer à la volée. Par conséquent, les applications n'ont pas à exprimer des règles de calcul dans les requêtes. Il leur suffit de demander des éléments du modèle dimensionnel (membres et indicateurs) et le moteur multidimensionnel exécute automatiquement les calculs nécessaires à la résolution de la requête.

Cette caractéristique de l'implémentation multidimensionnelle est particulièrement utile lorsque des applications basées sur le SQL sont utilisées pour accéder aux données de l'espace de travail analytique car ces applications n'ont pas à exprimer de règles de calcul. Ainsi, même les outils les plus éloignés de l'optique analytique peuvent bénéficier de la puissance analytique des types de données multidimensionnels Oracle.

Considérons l'état suivant produit par un outil de génération d'états. Il comporte une prévision du nombre d'unités vendues pour la prochaine période de vente et des calculs relatifs aux remises sur volume. Les règles de calcul sont intégrées à l'espace de travail analytique. Tout ce que le programme de génération d'états a à faire, c'est de les sélectionner à l'aide d'ordres SQL très simples.

Les applications telles que les générateurs d'états, qui ne savent rien du modèle dimensionnel, peuvent sélectionner des données ayant des types de données multidimensionnels Oracle.

Dans cet exemple, un document créé dans Oracle Reports comporte des indicateurs de tendance, de remise et de prévision calculés par l'option OLAP.

Global Exchange Corporation

Customer: Computer Services Tokyo
10-5, Akasaka 1-chome
Tokyo
Japan
Billing Period: May-03
Billing Date: Aug. 14, 2003

4453-665-78763

Page 1 of 2

Advance Purchase Order Services: *A special service for our Platinum level customers.*

Your total invoice amount for this period was \$2,937,416.17. Based on your purchase history with us we have estimated that your invoice amount for the next period will be \$3,012,650.72, an increase of 2.56%, for total number of 5,577 units.

For your conveniec we have created your purchase order for the next period. Pleas note that this is not an invoice and will not be charged to your account untill you make the purchases. This is just an estimation. Please visit our web site www.globalxchange.com and enter code FCSTINV8782 to make required modifications and purchases.

If you choose this option you will be entitled for a 5% discount on each and every item purchased on or before 03-Sep-2003 for a total saving of upto \$150,632.54. The discounted unit price and sales amounts are included in this estimation.

Product	Forecast Units	List Price	List Amount	Discount Price	Discount Amount	Saving Amount
SIMM- 8MB PCMCIAII card	103	\$275.91	\$28,413.42	\$262.11	\$26,992.75	\$1,420.67
SIMM- 16MB PCMCIAII card	30	\$546.39	\$16,617.59	\$519.07	\$15,786.71	\$830.88
External 6X CD-ROM	21	\$127.80	\$2,707.78	\$121.41	\$2,572.39	\$135.39

Une application basée sur le SQL telle qu'Oracle Reports accède aux 'indicateurs d'un espace de travail analytique

Grâce à cette possibilité d'intégrer des valeurs calculées dans l'espace de travail analytique et de les présenter comme des valeurs pré jointes et précalculées, des outils qui ne sont ni des outils dimensionnels ni des outils de business intelligence, les générateurs d'états par exemple, peuvent être utilisés aisément avec des espaces de travail analytiques.

Avantages des types de données multidimensionnels en matière de performances

Les types de données multidimensionnels Oracle sont conçus pour apporter au modèle dimensionnel d'excellentes performances d'interrogation et une grande efficacité pour le chargement de données, l'agrégation et les calculs. Le gain de performance est d'autant plus sensible que les calculs sont complexes et nombreux et que l'utilisation des données devient une utilisation adhoc.

Types de données multidimensionnels spécialisés

Les types de données multidimensionnels Oracle font appel à des structures de données spécialisées basées sur des tableaux pour stocker les membres de dimension et les données d'indicateur (faits). Ces structures ont des caractéristiques qui assurent d'excellentes performances dans un environnement d'interrogation adhoc. Chaque caractéristique contribue à l'optimisation du modèle dimensionnel.

- Les références aux cellules d'un indicateur utilisent l'adressage relatif et des méthodes d'indexation spécialisées dans lesquelles la dimension fonctionne comme un index des données.
- Les données d'indicateur sont préjointes aux dimensions. De ce fait, les types de données multidimensionnels sont extrêmement efficaces lors de l'application de prédicats de sélection des données, même dans les cas où des jointures externes auraient été nécessaires si le calcul avait été exécuté en SQL (par exemple, lors de certaines analyses de séries temporelles).
- Les cubes d'un espace de travail analytique peuvent partager des dimensions. On obtient ainsi d'excellents temps de réponse lorsque des indicateurs sont calculés à partir de données provenant de plusieurs cubes.
- L'option OLAP est extrêmement efficace pour la résolution de calculs complexes à la volée. Cette efficacité améliore à elle seule les performances des requêtes. En outre, elle réduit la quantité de données à précalculer et à stocker.
- L'option OLAP contient des méthodes brevetées d'optimisation du stockage et d'interrogation d'ensembles de données très creux couramment rencontrés dans les modèles dimensionnels.
- La capacité de navigation dans les agrégats est une caractéristique fondamentale du moteur multidimensionnel de l'option OLAP. On évite ainsi les problèmes de navigation dans les données agrégées que l'on rencontre avec d'autres technologies.
- L'option OLAP supporte le partitionnement des types de données multidimensionnels et la parallélisation des traitements de chargement et d'agrégation des données.

Toutes ces caractéristiques contribuent à l'optimisation du modèle multidimensionnel et en font le meilleur choix pour la plupart des applications de business intelligence.

Influence des types de requête sur les performances

Le fait que les performances d'une application soient améliorées par l'option OLAP dépend des caractéristiques de cette application et du type d'utilisation qu'en font les utilisateurs. Plus les requêtes sont imprévisibles et plus les calculs deviennent intensifs, plus les avantages des types de données multidimensionnels dans le domaine des performances s'accroissent.

Types de requête prévisibles et requêtes adhoc

Lorsque les requêtes sont prévisibles, il est relativement facile d'optimiser les performances d'interrogation d'une application en précalculant les données et en utilisant d'autres techniques telles que la création d'index répondant aux requêtes. Lorsqu'elles le sont moins, la prématisation et les autres techniques

d'optimisation telles que la création d'index pour des requêtes spécifiques deviennent moins pratiques. Il faut alors avoir recours à des optimisations portant sur le modèle dimensionnel et non sur des requêtes spécifiques. Prenons l'exemple de trois catégories d'applications : la génération d'états standard, la génération d'états d'exploration et l'analyse adhoc.

Génération d'états standard

Les systèmes de génération d'états standard tels qu'Oracle Reports et Crystal Reports présentent généralement les caractéristiques suivantes :

- Ils sont employés dans un mode prédéterminé et non dans un mode exploratoire. La plupart des utilisateurs exécutent des états prédéfinis en n'effectuant que quelques sélections de données dans des listes prédéfinies.
- La plupart des états standard ne comportent que peu de calculs. Les utilisateurs ne définissent pas leurs propres calculs.
- Ils interrogent des systèmes transactionnels, des systèmes en troisième forme normale et des data warehouses avec schéma en étoile. La structure dimensionnelle des schémas en étoile n'est généralement pas visible des utilisateurs.

Les performances sont optimisées par indexation et prématisation des données pour des états ou groupes d'états spécifiques. La prématisation peut être réalisée à l'aide d'une fonction d'agrégation (par exemple, somme des ventes journalières pour obtenir un total) ou de tout autre calcul facile à exprimer en SQL.

Génération d'états d'exploration

Avec les outils générant des états d'exploration, l'utilisateur final peut naviguer dans les données en fonction de ses besoins et, éventuellement, définir des calculs. Oracle Discoverer Plus et Business Objects Enterprise sont des exemples de tels outils, qui partagent généralement les caractéristiques suivantes :

- Beaucoup d'utilisateurs préfèrent exécuter des états qui ont été prédéfinis par un administrateur de base de données ou un autre utilisateur. En effet, les modèles de données utilisés par les applications sont difficiles pour les utilisateurs occasionnels et les performances des requêtes sont imprévisibles. Les états prédéfinis permettent d'optimiser la base de données en vue de leur exécution.
- Ces outils peuvent comporter des calculs de business intelligence courants : séries temporelles/tendances, classements et parts de marché par exemple. En général, les calculs se font à l'intérieur d'une même table afin d'obtenir des performances acceptables.
- Ces outils peuvent inclure des concepts du modèle dimensionnel. Ils peuvent par exemple permettre de naviguer dans les hiérarchies et de pivoter le long des axes d'un état d'une façon qui rappelle les dimensions.

Beaucoup d'applications de génération d'états d'exploration sont mises en oeuvre sans OLAP (Discoverer Plus Relational, Business Objects) ou fournissent leur propre serveur OLAP intermédiaire (MicroStrategy). En général, les performances de ces applications sont optimisées à l'aide des méthodes suivantes :

- Restriction de l'accès aux données afin de limiter les problèmes de performances, par exemple en fournissant des états prédéfinis et en contrôlant les données sélectionnables.
- Utilisation de schémas en étoile dans une base de données relationnelle. Les schémas en étoile optimisent les bases relationnelles pour les requêtes présentant des caractéristiques dimensionnelles.
- Indexation soignée, avec index bitmap et index de jointures bitmap.
- Précalcul de données de synthèse dans des tables d'agrégats ou des vues matérialisées.
- Prématisation de certains calculs, c'est-à-dire réalisation de certains calculs dans le processus d'alimentation ETL (Extraction, Transformation, Loading) plutôt que dans le processus d'interrogation.
- Utilisation de stratégies de mise en mémoire cache soit dans l'outil BI soit à un niveau intermédiaire (Business Objects, MicroStrategy).

Les stratégies de mise en mémoire cache et les niveaux intermédiaires OLAP fonctionnent souvent bien, mais soulèvent des problèmes comme la réplication des données, des métadonnées et des règles de gestion (sous forme de calculs), et permettent rarement d'obtenir autant qu'un moteur OLAP intégré.

Génération d'états analytiques

Les applications de génération d'états analytiques sont employées lorsqu'il faut analyser des données de manière interactive à l'aide de méthodes variées d'interrogation, de navigation et de calcul. OracleBI Discoverer Plus OLAP est un exemple de telles applications, qui partagent généralement les caractéristiques suivantes :

- L'utilisateur final définit fréquemment ses propres documents analytiques. Ces documents comportent des états et des graphiques, ainsi que d'autres modes de visualisation des données. Le système peut comporter certains états prédéfinis, mais il est d'usage que l'utilisateur final définisse de nouveaux états à sa convenance.
- L'utilisateur final interroge les données provenant de tout le modèle de données, à tout moment. Les requêtes faisant référence à plusieurs dimensions et à plusieurs tables de faits ou cubes sont courantes.
- Il y a utilisation intensive de calculs personnalisés. Les utilisateurs définissent souvent de nouveaux calculs et ces calculs font fréquemment référence à des données de plusieurs tables de faits ou de plusieurs cubes.

En outre, les utilisateurs de cette classe d'applications ont des exigences élevées en ce qui concerne les performances.

Dans un système non OLAP, les performances des applications utilisant des états analytiques sont optimisées à l'aide des méthodes suivantes :

- Utilisation de schémas en étoile, qui sont optimisés pour certains types de requêtes adhoc.
- Utilisation du partitionnement et du parallélisme.
- Prématisation de grandes quantités de données agrégées et de certains calculs.

- Création de nombreux index.
- Génération de SQL hautement optimisé.
- Utilisation de stratégies de mise en mémoire cache dans l'outil BI ou utilisation d'un niveau intermédiaire.

Même ainsi, les performances risquent de ne pas être à la hauteur des attentes de l'utilisateur final. Elles sont souvent limitées du fait qu'il faut réaliser, lors de l'exécution, des agrégations et d'autres calculs impliquant des jointures coûteuses, en particulier des jointures entre des tables de faits et des jointures externes. L'option OLAP a été spécialement optimisée pour ce type d'applications et ne subit pas ces limitations des performances.

D'une manière générale, on peut constater les faits suivants :

- Les applications de génération d'états standard fonctionnent bien sans l'option OLAP lorsque le nombre d'états est limité car les types de requête sont prévisibles et pris en charge par les techniques d'optimisation disponibles.
- Lorsque le volume et la variété des états augmentent, les optimisations inhérentes aux types de données multidimensionnels peuvent améliorer les performances, à la fois du processus ETL (Extraction, Transformation, Loading) et des requêtes. La valeur ajoutée des applications de reporting standard peut s'accroître du fait des calculs disponibles dans l'espace de travail analytique.
- Les applications générant des états d'exploration ont de bonnes performances lorsque les requêtes sont assez prévisibles et que les besoins en calcul sont limités. L'option OLAP peut apporter à ces applications basées sur le SQL de meilleures performances d'interrogation et un réel contenu analytique.
- Le succès des applications générant des états analytiques est souvent limité sans OLAP car elles ne répondent pas aux exigences en matière de performances et de calcul lorsque les requêtes sont imprévisibles. Celles qui ne sont pas basées sur un modèle dimensionnel sont généralement moins conviviales que celles qui le sont.
- L'option OLAP assure le support du modèle dimensionnel, offre d'excellentes performances d'interrogation, à la fois pour les requêtes prévisibles et non prévisibles, et permet toutes sortes de calculs. Elle augmente considérablement le succès des applications de reporting analytique.

Savoir quand utiliser l'option OLAP pour améliorer les performances

Les performances d'un système de business intelligence sont généralement évaluées en terme de performance des requêtes. Toutefois, cette performance dépend souvent du temps qui a été consacré à la préparation du système pour l'interrogation et de la taille du jeu de données. Plus le temps disponible pour la préparation des données est important, plus la quantité de données agrégées ou précalculées est importante, ce qui aboutit généralement, mais pas toujours, à de meilleures performances des requêtes.

Néanmoins, dans la plupart des cas, le temps disponible pour préparer les données à l'interrogation est limité. En outre, les jeux de données qui contiennent plus de dimensions granulaires ont tendance à être plus volumineux car ils contiennent plus de données de niveau détail et plus de données de niveau agrégé. Par exemple, il est fort probable qu'un jeu de données contenant des données journalières comportera un volume de données de niveau détail plus important qu'un jeu de données contenant seulement des données mensuelles. Par contre, la fenêtre de traitement d'un jeu de données devant être mis à jour quotidiennement sera probablement beaucoup plus réduite que celle d'un jeu de données qui n'est mis à jour que mensuellement.

On voit aisément qu'il est difficile de conjuguer d'excellentes performances des requêtes, une taille réduite du jeu de données et des temps de traitement minimisés. Si les exigences s'orientent en priorité sur les performances des requêtes, les utilisateurs devront se résoudre à accepter un jeu de données plus réduit, moins complet et des mises à jour des données moins fréquentes. L'efficacité des analyses peut en souffrir. Si l'exigence principale est un jeu de données plus volumineux, ce sont les performances des requêtes et par conséquent la satisfaction des utilisateurs qui en pâtiront.

Pour bien choisir l'implémentation d'un système de business intelligence, il faut répondre aux questions suivantes :

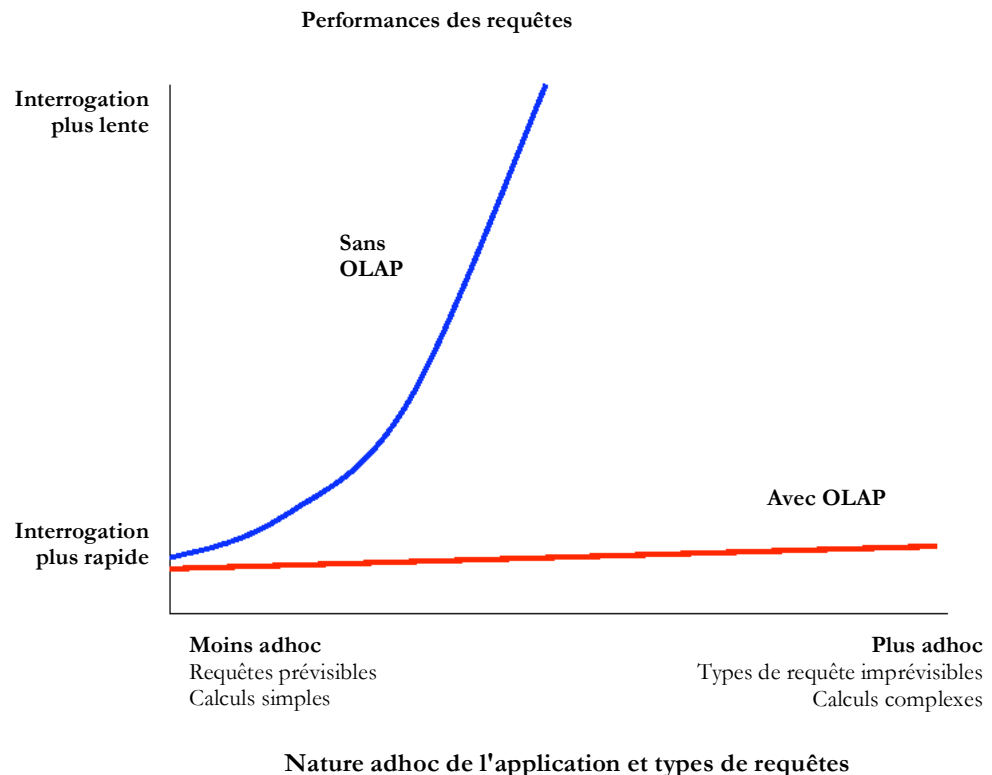
- Quels sont les besoins en matière de calculs ? Les calculs sont-ils simples et prédéfinis ou sont-ils complexes et doivent-ils pouvoir être définis à librement par les utilisateurs ? Faut-il joindre de nombreuses dimensions à des faits ou des faits à d'autres faits ? Faut-il inclure des calculs de séries temporelles nécessitant des jointures externes ?
- Dans quelle mesure l'utilisation des données est-elle prévisible ? Les données sont-elles interrogées par un jeu fixe d'états avec des possibilités de sélection de données réduites, ou bien les utilisateurs définissent-ils leurs propres requêtes sur n'importe quel élément du modèle et emploient-ils des calculs qu'ils ont définis eux-mêmes ?
- Quelle est la taille de la communauté des utilisateurs ? Si les utilisateurs sont nombreux et qu'ils interrogent les données de manière non prévisible, l'utilisation du modèle de données tend plus vers une utilisation adhoc.
- Quelles sont les attentes des utilisateurs en matière de performances ? Les états sont-ils exécutés hors ligne (avec un système de génération d'états par lots par exemple), ou bien les données sont-elles interrogées en mode interactif, les utilisateurs attendant des performances d'interrogation excellentes ?
- Combien de temps et de ressources informatiques sont-ils disponibles pour la préparation des données ? Des fenêtres de mise à jour réduites et un matériel peu puissant peuvent rendre difficiles les mises à jour fréquentes et le préstockage de grands volumes de calculs.

Si l'application de business intelligence est simple, avec des types d'états fixes ou très prévisibles et des calculs simples, l'option OLAP n'est pas forcément nécessaire. Par contre, si le système doit prendre en charge des calculs complexes et offrir d'excellentes performances d'interrogation dans un environnement de requêtes imprévisibles, les avantages apportés par l'option OLAP en matière de performances peuvent être considérables.

Les graphiques qui suivent présentent une modélisation des performances d'implémentation avec OLAP et sans OLAP. Le graphique ci-dessous décrit la relation entre les performances des requêtes et la nature adhoc de l'application.

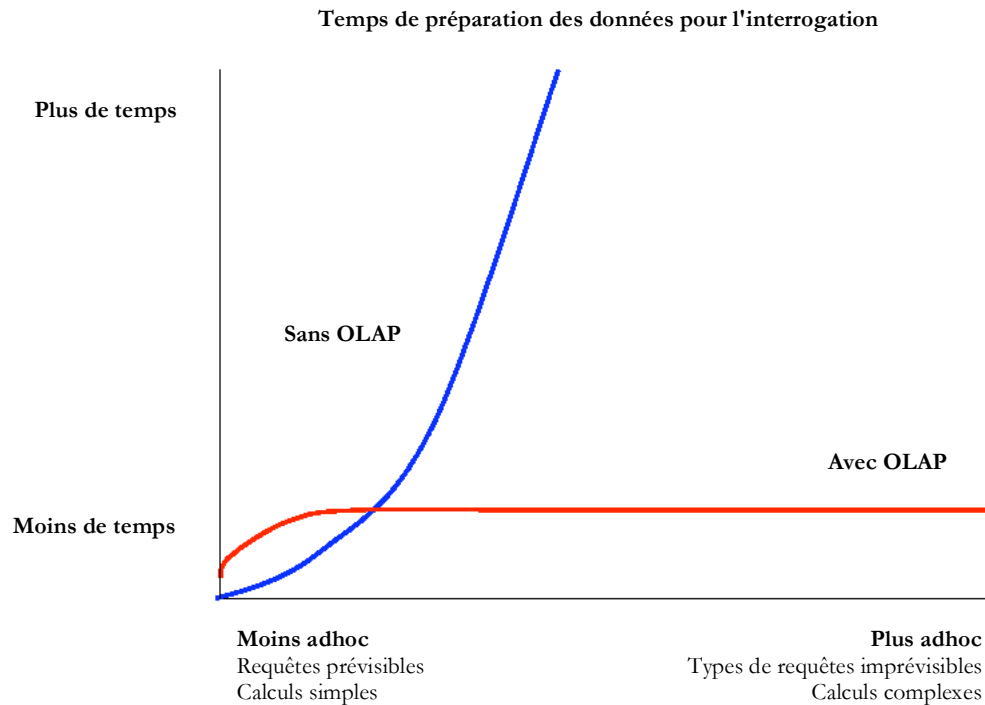
Le premier graphique montre que les applications sans l'option OLAP peuvent obtenir d'excellentes performances d'interrogation lorsque leur nature est telle que les requêtes sont prévisibles et que les besoins de calcul sont relativement simples. Dans ce cas, l'application peut être optimisée spécialement pour ces requêtes, par exemple à l'aide de vues matérialisées et de précalculs d'indicateurs spécifiques.

Lorsque le caractère adhoc de l'application est plus prononcé, il arrive un moment où, sans OLAP, les attentes des utilisateurs en matière de performance ne sont plus satisfaites. A partir de ce point, il faut implémenter l'application avec les types de données multidimensionnels de l'option OLAP pour obtenir des performances d'interrogation correctes.



Ensuite, il est important de tenir compte du temps nécessaire pour préparer les données à l'interrogation car il s'agit de l'un des facteurs déterminant la taille du jeu de données et la fréquence des mises à jour.

Le graphique ci-dessous décrit la relation entre le temps de préparation des données requis pour obtenir des performances d'interrogation acceptables en regard de la nature adhoc de l'application.



Nature adhoc de l'application et types de requêtes

Le graphique montre que le coût de préparation des données est au départ plus élevé avec l'option OLAP, par contre, une fois l'espace de travail analytique préparé, les applications adhoc les plus exigeantes peuvent être satisfaites. Sans OLAP, le coût devient rapidement prohibitif, même pour des applications d'une complexité modérée.

L'option OLAP peut souvent servir toute une gamme d'applications via le SQL, apportant des avantages à la fois en matière de calculs et en matière de performances aux applications dimensionnelles et non dimensionnelles.

Exemple d'application

L'application ci-dessous, correspondant à un cas réel, va nous servir à illustrer le modèle multidimensionnel. Les données contiennent un cube, avec dix dimensions dont les caractéristiques sont les suivantes :

Dimension	Niveaux	Membres
Lieu de livraison	2	52
Fabricant	2	166
Acheteur	2	38
Client	3	4998
Fournisseur	2	3
Produit	3	7099
Temps	4	764
Secteur	2	13
Lieu de livraison	2	53
Lieu d'expédition	2	41

La taille du cube est déterminée par le nombre de combinaisons de niveaux uniques dans le modèle (4608) et la taille logique du cube. La taille logique du cube est l'intersection de toutes les dimensions, ce qui donne le nombre total de cellules, ou points de données, du modèle. Bien que, dans ce cas, aucune des dimensions ne soit particulièrement volumineuse, le nombre de dimensions fait que le cube logique a une taille importante ($7,54 * 10^{20}$ cellules).

Lorsque l'on considère la taille logique du cube, il faut garder à l'esprit le fait que l'on attend du système OLAP d'excellents temps de réponse pour toutes les données du cube, ce qui est considérablement plus ardu que fournir d'excellents temps de réponse pour les seules données stockées dans une table. C'est pourquoi, entre autres, l'option OLAP convient particulièrement pour les types de requête imprévisibles.

Les performances d'interrogation ont été mesurées avec deux jeux de requêtes. Dans le premier, appelé « Faiblement adhoc », les requêtes se limitent à un tout petit ensemble de données du modèle (15 des 4608 combinaisons de niveaux possibles). 25 % des requêtes contiennent au moins un type de calcul allant de l'indice simple (une part de marché par exemple) à des séries temporelles de base (calcul de la période précédente par exemple). Dans le second jeu de requêtes, appelé « Fortement adhoc », les requêtes peuvent sélectionner des données dans environ 75 % du modèle (3456 combinaisons de niveaux). 75 % des requêtes contiennent des calculs.

Chaque jeu de requêtes a été exécuté sur un espace de travail analytique OLAP et sur deux jeux de données non OLAP. Avec le jeu de données OLAP, en l'occurrence un espace de travail analytique, l'optimum des performances a été atteint avec un temps de préparation des données de 64 minutes. Les jeux de données non OLAP ont quant à eux été préparés en 19 minutes et 2800 minutes respectivement. Comme pour le jeu de données OLAP, le temps de préparation a été consacré au précalcul d'une certaine quantité de données de niveau agrégé.

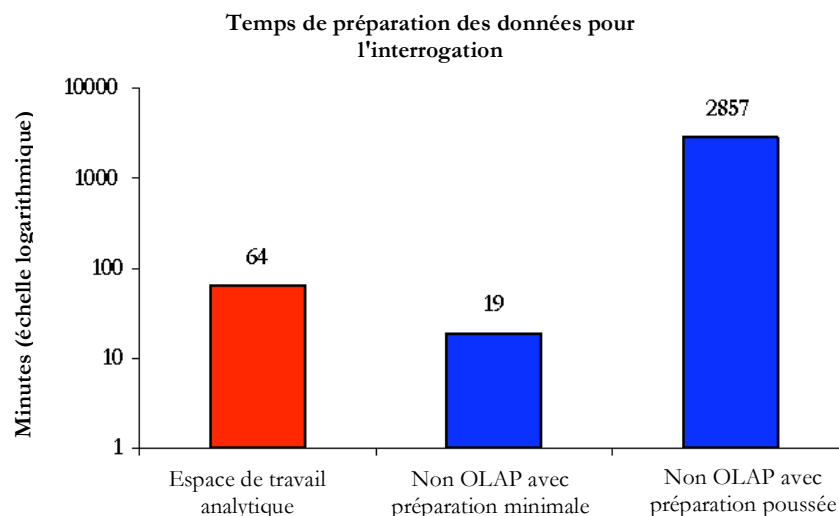
Pour chaque jeu de requêtes, 100 requêtes distinctes ont été exécutées et le temps d'interrogation moyen a été enregistré. Les résultats montrent que le temps de préparation initiale des données est plus élevé avec le jeu de données OLAP mais que les performances sont infiniment supérieures à celles obtenues avec les jeux de données non OLAP. En fait, les performances d'interrogation du jeu de données non OLAP se sont dégradées avec l'augmentation du nombre de vues matérialisées disponibles, en raison de la charge accrue de navigation dans les agrégats. Plus il

existe d'objets avec des données de niveau agrégé, moins la recherche du bon objet à interroger devient efficace.

Implémentation	Temps de préparation en minutes	Espace disque (Go)	Secondes par requête (moyenne sur 100 requêtes)	
			Jeu de requêtes « Faiblement adhoc »	Jeu de requêtes « Fortement adhoc »
Avec OLAP	64	4,3	3,4	3,4
Sans OLAP - Préparation minimale des données	19	0,7	13,3	26,7
Sans OLAP - Préparation poussée des données	2857	42,9	47,0	86,9

Comparaison entre jeux de données OLAP et non OLAP au niveau du temps de préparation des données, de l'espace disque consommé et des performances des requêtes

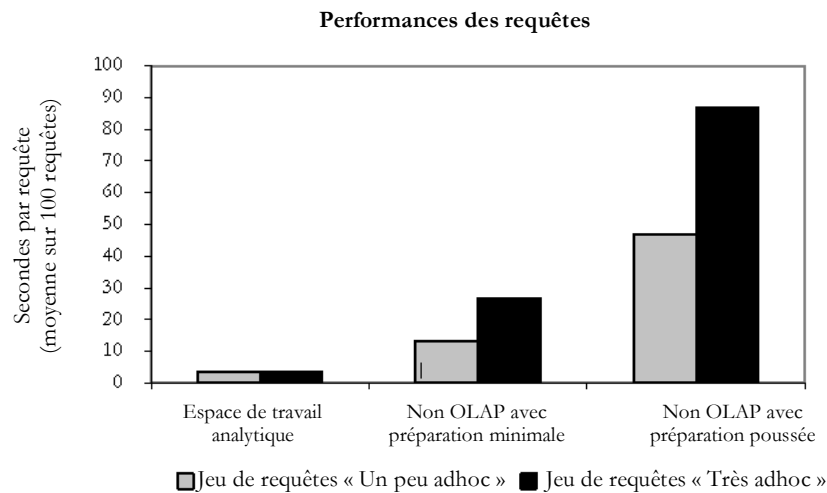
Le graphique ci-dessous compare les temps passés à préparer les données pour l'interrogation. A noter que le temps est présenté en échelle logarithmique.



Temps passé à la préparation de l'espace de travail analytique et des jeux de données non OLAP pour l'interrogation

Les performances avec l'option OLAP sont meilleures que celles sans OLAP, dans le cas le moins optimisé. Néanmoins, l'utilisateur final peut trouver les performances acceptables dans les deux cas si ses requêtes ressemblent à celles du jeu de requêtes « Faiblement adhoc ». (Il s'agit là d'un jugement subjectif des utilisateurs.) Par contre, il est peu probable que l'utilisateur final se satisfasse des performances du jeu de

requêtes « Fortement adhoc » avec les jeux de données non OLAP. En outre, on constate que le précalcul de large volumes de données n'améliore pas toujours les performances des requêtes.



Performances des requêtes en fonction de la quantité de précalculs dans le jeu de données

Chaque ensemble jeu de données et application a des caractéristiques uniques (taille du jeu de données, complexité du modèle, besoins de calcul, dispersion des données, types d'utilisation récurrents, etc.) et donc un profil de performance unique. Le modèle présenté ici donne néanmoins une idée assez précise des caractéristiques des implémentations OLAP et non OLAP en matière de performance, et illustre bien la valeur des optimisations apportées par l'option OLAP.

OUTILS D'ADMINISTRATION POUR ESPACES DE TRAVAIL ANALYTIQUES

Il doit maintenant être clair que le modèle dimensionnel et son implémentation avec des types de données multidimensionnels apportent des avantages significatifs. Il doit également être évident que les types de données multidimensionnels peuvent être utilisés pour améliorer les performances et le contenu d'applications qui ne sont pas basées sur le modèle dimensionnel et qui utilisent le SQL comme langage d'interrogation.

Les types de données multidimensionnels et les espaces de travail analytiques peuvent sembler exotiques à quelqu'un qui est familier des types de données relationnels et du data warehousing, c'est pourquoi il existe des outils qui simplifient la conception et la gestion des jeux de données multidimensionnels.

La plate-forme Oracle Business Intelligence propose deux outils :

- Analytic Workspace Manager
- Oracle Warehouse Builder

Tous deux offrent des fonctions de modélisation dimensionnelle et de gestion des espaces de travail analytiques. Chacun vise un type particulier d'utilisateur et d'utilisation.

Analytic Workspace Manager

L'interface d'administration des espaces de travail analytiques, Analytic Workspace Manager, s'adresse aux spécialistes OLAP devant travailler directement en interactif avec le modèle dimensionnel et des espaces de travail analytiques. C'est typiquement le cas lors du prototypage et du développement d'applications, et dans les implémentations au niveau des services d'une entreprise. Analytic Workspace Manager convient pour un large éventail d'utilisateurs : utilisateur avancé , administrateur de base de données en entreprise, développeur d'applications, etc. Sa conception s'est concentrée sur la facilité d'emploi, la modélisation dimensionnelle et la définition d'agrégations et d'autres calculs.

Analytic Workspace Manager permet de :

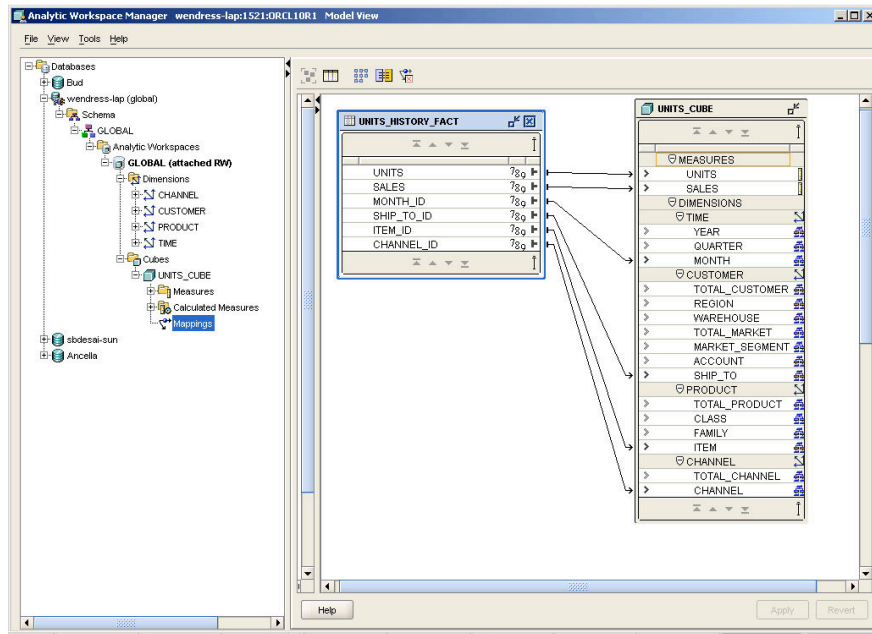
- Définir le modèle dimensionnel logique
- Implémenter le modèle dimensionnel en tant qu'espace de travail analytique
- Mettre le modèle en correspondance avec des sources de données relationnelles (schémas en étoile, en flocon et autres)
- Définir des agrégations et des calculs d'indicateurs
- Effectuer les opérations de maintenance périodique telles que le chargement de données et la ré-agrégation
- Assurer l'administration de manière collaborative, en permettant le partage de modèles , en partie ou en totalité, avec d'autres utilisateurs ou applications telles que Oracle Warehouse Builder.

La création d'un espace de travail analytique s'effectue généralement comme suit :

- Sur la base des besoins de reporting des utilisateurs, conception du modèle dimensionnel logique (dimensions, hiérarchies, cubes, indicateurs de base). Le modèle peut être mis à jour par la suite en fonction des besoins.
- Si nécessaire, mise en correspondance du modèle logique avec des sources dans Oracle Database. Ces sources peuvent être des tables, des vues, des fichiers plats en tant que tables externes ou d'autres objets relationnels.
- Génération initiale de l'espace de travail analytique. Ceci consiste à charger les données et, éventuellement, à agréger certaines données pour optimiser les performances.
- Définition de nouveaux calculs d'indicateurs, si nécessaire.
- Vérification de la conception dans le contexte de l'outil ou de l'application de business intelligence qui sera utilisé pour interroger les données.
- Corrections si nécessaire.

Analytic Workspace Manager permet de réaliser toutes les étapes de ce processus dans un environnement de conception unique, orienté dimensionnel.

Analytic Workspace Manager est conçu pour être accessible aux administrateurs de base de données d'un service ou de toute l'entreprise et aux utilisateurs pointus.



Avec Analytic Workspace Manager, la mise en correspondance d'indicateurs d'un cube avec une table de faits source est une simple opération de glisser-déplacer

Oracle Warehouse Builder

Oracle Warehouse Builder est une application complète d'extraction, transformation et chargement qui peut définir à la fois des modèles relationnels et des modèles dimensionnels et effectuer des déploiements vers les types de données relationnels et multidimensionnels. Oracle Warehouse Builder gère le processus de conception et de déploiement du système source (par exemple, un système de gestion de commandes) vers un data warehouse relationnel ou vers un espace de travail analytique. Pouvoir gérer la totalité du processus - de la source transactionnelle à l'espace de travail analytique - constitue un avantage déterminant pour l'administrateur de base de données de l'entreprise, en particulier dans l'environnement de production.

Il arrive que certains utilisateurs, très au fait des besoins de reporting, conçoivent le modèle dimensionnel à l'aide d'Analytic Workspace Manager, puis transfèrent la maintenance au jour le jour de l'espace de travail analytique à l'administrateur de base de données de l'entreprise, qui utilise Oracle Warehouse Builder. Par exemple, un analyste marketing peut trouver l'outil Analytic Workspace Manager parfait pour un prototypage rapide des dimensions et des cubes. Il peut ensuite confier la maintenance périodique de l'espace de travail analytique à l'administrateur de base de données de l'entreprise, qui, lui, emploie Oracle Warehouse Builder pour définir le processus ETL (Extraction, Transformation, Loading) et gérer l'espace de travail analytique dans son cycle de vie.

CONCLUSION

L'option OLAP d'Oracle Database 10g fournit un modèle dimensionnel, basé sur des dimensions, des cubes et des indicateurs, ainsi que des implémentations relationnelles (ROLAP) et multidimensionnelles (MOLAP). Dans les implémentations ROLAP, les données sont stockées dans des schémas en étoile ou en flocon. Dans les implémentations MOLAP, elles sont stockées dans des types de données multidimensionnels Oracle. Les implémentations ROLAP conviennent pour des applications de reporting exploratoire autorisant les requêtes adhoc et les calculs mais où les types d'utilisation sont relativement prévisibles. Les implémentations MOLAP offrent des possibilités de calcul plus étendues et de meilleures performances, notamment lorsque les types d'utilisation sont imprévisibles, qu'il y a besoin de calculs complexes, que les performances d'interrogation attendues sont élevées et que les jeux de données OLAP doivent être gérés efficacement.

L'avantage principal du modèle dimensionnel est la structure qu'il fournit pour la définition de requêtes, la navigation dans les données et la définition de calculs. Ces avantages profitent tant à l'utilisateur final qu'au développeur d'applications. L'utilisateur final trouve que le modèle dimensionnel améliore le processus d'interrogation et de reporting. Le développeur trouve que le modèle dimensionnel et les composants tels que les OracleBI Beans et l'API OLAP facilitent la création de solutions de business intelligence basées sur ce modèle en les rendant performantes et facilement gérables.

L'interface SQL des types de données multidimensionnels Oracle permet aux applications basées sur le langage d'interrogation SQL, qu'elles aient ou non une approche dimensionnelle, de bénéficier des avantages de l'option OLAP. Les types multidimensionnels peuvent par exemple être utilisés avec des applications SQL telles que Business Objects Enterprise, Cognos ReportNet ou des générateurs d'états. L'interface SQL permet d'employer les types de données multidimensionnels avec la plupart des applications de business intelligence.

Les types de données multidimensionnels peuvent être définis et gérés par toutes sortes d'utilisateurs. Le gestionnaire d'espaces de travail analytiques, Analytic Workspace Manager, rend le modèle dimensionnel et les types de données multidimensionnels facilement accessibles pour les utilisateurs avancés et les administrateurs de base de données d'entreprise. Quant à Oracle Warehouse Builder, il offre un système ETL (Extraction, Transformation, Loading) complet et la possibilité d'effectuer le déploiement directement vers les types de données multidimensionnels.



Option OLAP d'Oracle Database 10g

Juin 2005

Auteurs : Bud Endress, Director of Product Manager - OLAP

Oracle Corporation

Siège social international

500 Oracle Parkway

Redwood Shores, CA 94065

Etats-Unis

Renseignements :

Tél. : +1 650 506 7000

Fax : +1 650 506 7200

www.oracle.com

Copyright © 2005, Oracle. Tous droits réservés.

Le présent document est uniquement fourni à titre informatif, et les informations qu'il contient peuvent faire l'objet de modifications sans préavis. Oracle Corporation ne garantit pas que ce document est exempt d'erreurs, ni qu'il est sujet à des garanties ou conditions, qu'elles soient exprimées oralement ou impliquées par la législation, y compris à des garanties et conditions de commercialisation ou d'adaptation à un but particulier. Oracle Corporation se dégage expressément de toute responsabilité afférente à ce document et ce dernier n'entraîne aucune obligation contractuelle, que ce soit directement ou indirectement. Toute reproduction ou transmission de ce document, sous quelque forme et par quelque procédé que ce soit, quel qu'en soit l'usage, est interdite sans l'autorisation expresse d'Oracle Corporation.

Oracle est une marque déposée d'Oracle Corporation et/ou de ses filiales. Les autres noms cités peuvent être des marques déposées par leurs propriétaires respectifs.