

Chapitre 3

Création de rapports en dimensionnel

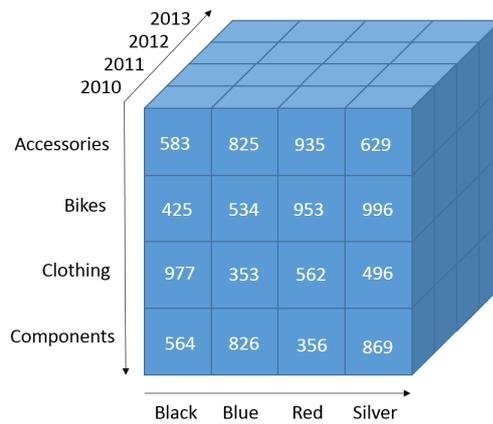
1. Qu'est-ce que le dimensionnel ?

Bien que paraissant sorti tout droit d'un livre ésotérique ou de physique avancée, le développement en dimensionnel, aussi appelé développement multidimensionnel, désigne une fonctionnalité proposée par Cognos (et par d'autres outils) permettant de modéliser les données et de naviguer dans les dimensions grâce à des hiérarchies.

1.1 Le cube

L'ancêtre du modèle multidimensionnel actuel est le cube, aussi appelé cube OLAP. Il était modélisé grâce à l'outil Transformer, dans les anciennes versions de Cognos. Transformer a désormais disparu des versions les plus récentes de Cognos et la modélisation multidimensionnelle a été intégrée à l'outil Framework Manager, qui permet de modéliser les données sur lesquelles s'appuient les rapports.

Un cube peut être assimilé à un tableau multidimensionnel de deux dimensions ou plus, l'exemple de cube le plus basique étant le tableau croisé où les arêtes représentent les axes d'analyse (les dimensions), et le centre du tableau les mesures, c'est-à-dire les faits.



Historiquement, les cubes étaient précalculés, ce qui était à la fois leur point fort et leur point faible. Les cubes étaient généralement calculés pendant la nuit, car en fonction du volume de données, le calcul pouvait durer un certain temps, le logiciel calculant toutes les intersections entre les dimensions et les faits. De cette manière, le temps de réponse des rapports et autres requêtes était optimisé, car les calculs et les appels à la base de données étaient déjà faits. Un cube présentait également l'avantage d'être portable, c'est-à-dire qu'il n'avait pas besoin de connexion à la base de données pour fonctionner, car les calculs avaient déjà été effectués. Avec les logiciels adéquats, il était donc tout à fait possible d'aller chez un client avec le cube copié sur l'ordinateur et de lancer des rapports ou des requêtes et de présenter les données, et ce sans connexion au serveur. Mais au fur et à mesure, ce point fort est devenu un point faible : le volume des bases de données ayant explosé ces dernières années, les cubes sont devenus trop gros et ingérables.

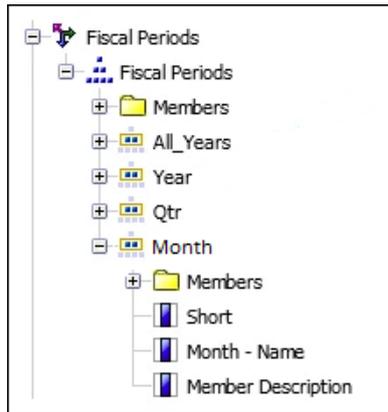
Le cube généré à partir du logiciel Transformer (encore répandu il y a une dizaine d'années) a été abandonné, mais sa logique a été intégrée à Framework Manager. Framework Manager permet désormais de créer des cubes à la volée, c'est-à-dire une représentation multidimensionnelle de la base de données grâce à des dimensions hiérarchisées, les appels à la base de données étant réalisés en direct, comme pour un modèle relationnel classique. Le modèle dimensionnel est exploitable avec les différents outils de Cognos Analytics, notamment Report Studio et Dashboard.

Mais quel est l'intérêt d'un modèle dimensionnel ?

1.2 Structure d'un modèle dimensionnel

Il est important de comprendre qu'à la base, le modèle dimensionnel est né afin de permettre aux utilisateurs de manipuler leurs données. Avant cela, recevoir ou exécuter un rapport préconçu était suffisant pour le bon fonctionnement de l'entreprise. Mais aujourd'hui (et depuis au moins dix ans), dans l'ère du « tout data », les utilisateurs ont besoin de manipuler leurs données, d'aller chercher certaines informations et de tout simplement les analyser. En parallèle de cela, le profil des utilisateurs a évolué puisqu'ils sont également formés à l'utilisation des outils d'informatique décisionnelle alors qu'auparavant ils n'étaient que des récepteurs d'information.

Le modèle dimensionnel présente l'information de manière intuitive, organisée en mesures et en dimensions. Ces dernières étant organisées en hiérarchies, elles permettent à l'utilisateur de créer facilement son propre rapport, sans connaissances spécifiques en base de données, en syntaxe SQL, etc. en allant chercher l'information dont il a besoin. Le dimensionnel présente directement les valeurs des dimensions hiérarchisées. Au lieu de travailler avec des noms de colonnes comme dans le modèle relationnel, l'utilisateur n'a donc qu'à glisser et à déplacer la ou les valeurs de son choix dans son rapport pour les voir apparaître. Les valeurs d'une dimension sont appelées *membres*. Une *dimension* représente donc un axe d'analyse et est composée d'une ou plusieurs *hiérarchies* contenant différents niveaux (les concepts de membre, de dimension et de hiérarchie sont décrits à la section Quelques concepts en multidimensionnel de ce chapitre). Beaucoup de dimensions n'ont qu'un seul niveau par exemple le type de commande ou la couleur des yeux. Une dimension hiérarchisée classique est le temps, elle est formée de plusieurs niveaux, par exemple l'année, le mois et le jour. Chaque niveau est composé de membres. Pour le niveau année, les membres sont 2020, 2021, 2022, etc. Pour les mois, ce sont janvier, février, etc., et enfin pour les jours, la date elle-même.



Une dimension (un axe d'analyse) peut donc disposer de plusieurs hiérarchies. Il est possible d'ajouter, à la dimension Calendrier, une hiérarchie supplémentaire comprenant plus ou moins de niveaux. Par exemple, il est possible de lui ajouter le semestre et le trimestre, afin de naviguer dans la dimension temps de différentes manières, mais aussi d'afficher des résultats pour d'autres agrégations.

L'utilisateur peut donc facilement manipuler des données (en allant chercher le ou les membres qui l'intéressent) ou encore le niveau entier, puis ajouter des mesures à son graphique ou à son tableau croisé.

Pour comparer le modèle relationnel et le modèle dimensionnel, prenons un exemple. Dans le modèle relationnel, pour récupérer les ventes d'un produit sur une journée donnée, il faut récupérer, dans la requête, la mesure Ventes, la colonne Produit et celle correspondant à la date. Ensuite, il est nécessaire de filtrer la requête par le code ou le nom du produit voulu, et idem pour la date.

En dimensionnel au contraire, il suffit d'aller chercher le membre correspondant au produit voulu (par glissé-déplacé), ainsi que la date, et la mesure ventes. On ne travaille pas ici avec des noms de colonnes à filtrer, mais directement avec des valeurs (les membres). Le fait d'aller chercher un membre crée le filtre en soi.

■ Remarque

Il faut retenir ici la structure des données présentées dans un modèle dimensionnel, composée de dimensions, elles-mêmes composées d'une ou de plusieurs hiérarchies, de niveaux, et enfin de membres. Le fait de sélectionner un membre inclut un filtre (un contexte). Il n'est donc pas nécessaire de filtrer la requête comme dans un modèle relationnel.

1.3 Avantages et inconvénients du modèle dimensionnel

L'une des erreurs de choix de conception souvent rencontrées est l'utilisation non adaptée du modèle dimensionnel. Car ce n'est pas la panacée. Même s'il apporte des avantages et des nouveautés, son mauvais usage peut avoir des effets très négatifs, notamment sur les performances. Il est souvent utilisé « par défaut » parce que c'est le style de développement le plus « moderne » (si c'est nouveau, c'est mieux !) ou parce que l'on croit à tort que le dimensionnel supprime le relationnel, alors que les deux modèles sont complémentaires.

Comme il a été vu précédemment, c'est un outil très puissant pour l'analyse active de données, pour aller chercher l'information et la comparer. Son utilisation est simple et surtout intuitive. Avant cela, pour pouvoir avoir des précisions ou des détails sur une donnée, les utilisateurs devaient faire appel à l'équipe décisionnelle ou à l'administrateur de la base de données afin de réaliser une extraction de données ou encore de générer un nouveau rapport plus détaillé, mais ponctuel. Désormais, avec un minimum de formation, l'utilisateur peut analyser ses données lui-même et faire ses recherches en créant ses propres analyses grâce à de simples gestes.

Bien évidemment, il est possible de faire des analyses très complexes en dimensionnel, mais l'un des avantages importants de ce style de modélisation est le caractère accessible à tous de l'outil.

Un autre avantage lié à la structure du modèle dimensionnel est le fait de pouvoir naviguer dans les données grâce aux hiérarchies. Dans le rapport lui-même, il est possible de cliquer sur une donnée ou d'utiliser le menu de navigation afin d'accéder au parent du membre (le mois étant un niveau supérieur du jour dans la hiérarchie Temps) ou encore à ses enfants (les mois lorsque l'on clique sur une année par exemple). Ces fonctionnalités changent radicalement la manière d'accéder aux données et de les analyser puisqu'il est aisé désormais de prendre du recul dans l'analyse en partant d'une donnée et en accédant au niveau supérieur, ou encore d'entrer dans le détail en un clic. L'équivalent en relationnel nécessiterait un rapport complexe, fait d'accès au détail vers d'autres rapports par exemple puisque, en relationnel, les dimensions sont « plates » et ne sont pas hiérarchisées.

Les comparaisons sont un exemple de cas de développement fondamentalement plus simple à mettre en place en dimensionnel qu'en relationnel. Prenons le calcul d'un ratio de résultats entre deux années. Dans ce cas, il existe plusieurs méthodes en relationnel, qui ne sont pas forcément compliquées, mais qui peuvent paraître fastidieuses pour un calcul si simple. Créer une jointure entre deux requêtes, chacune calculant les résultats d'une année, puis calculer le ratio dans une troisième requête est une possibilité. On peut faire également le calcul dans une seule requête en créant une mesure pour chaque année puis en agrégeant les mesures. Mais c'est fastidieux...

Ce genre de calcul en dimensionnel est réalisé de manière beaucoup plus simple, rapide et intuitive. Dans une même requête, il est possible d'aller chercher les deux années voulues en tant qu'objets indépendants, de calculer leurs mesures respectives puis de les diviser pour obtenir le ratio, sans avoir recours à des procédés divers et variés.

Comme il a été vu jusqu'à présent, le développement en dimensionnel offre de nombreuses fonctionnalités, il améliore nettement l'expérience utilisateur (plus de dynamisme et plus d'autonomie) et il est plus intuitif. Tout n'est pas parfait, évidemment, et cette flexibilité a un prix : la performance. Le modèle relationnel est en soi une carcasse vide puisqu'on ne raisonne qu'en termes de colonnes, et sa structure est plus légère. Que les tables soient vides ou alimentées, le « poids » du modèle reste le même. Dans un modèle dimensionnel, l'utilisateur travaille avec des valeurs disponibles lorsqu'il ouvre les dimensions du modèle, et cela implique que Cognos ait déjà effectué les appels à la base de données pour toutes les dimensions du modèle. De plus, l'organisation du modèle en hiérarchies nécessite un surcoût de calcul, assumé par le serveur Cognos, et non par le serveur de la base de données.

■ Remarque

Tous ces facteurs font que les rapports développés en multidimensionnel sont plus gourmands en ressources, alors que les rapports en relationnel utilisent en général davantage le serveur de la base de données, qui est souvent plus puissant ou mieux dimensionné.

Mais « plus gourmand en ressources » ne veut pas forcément dire « moins performants » ! Comme l'exemple du calcul de ratio l'a montré, la méthode dimensionnelle est plus simple dans sa mise en place, elle requiert moins de requêtes et moins d'opérations. Elle n'est donc pas forcément moins performante que le relationnel.

Comme il a été mentionné en début de chapitre, l'erreur souvent commise est de mal identifier le style de développement à utiliser en fonction des besoins. Dans certains cas, les rapports relationnels sont beaucoup plus performants pour un même résultat final. Par exemple, lorsque l'objectif est d'afficher des listes (par exemple, une liste de produits, de type rapport détaillé) comprenant de nombreuses lignes ou encore d'afficher un rapport statique, le modèle relationnel est à privilégier en raison du volume de données à traiter et à afficher.

■ Remarque

Pour faire simple, disons que si les fonctionnalités qu'offre le modèle dimensionnel ne sont pas requises dans un rapport, il vaut mieux opter par défaut pour le modèle relationnel, qui sera plus performant.

En règle générale, le modèle dimensionnel n'apprécie pas les gros volumes de données. Si le rapport à développer a pour vocation d'être dynamique et de permettre la navigation dans les données, alors le choix du modèle dimensionnel est obligatoire. Dans ce cas, il existe des manières d'améliorer les performances, notamment en utilisant des tables agrégées, ce qui réduit le volume des tables de faits.

En résumé, le modèle dimensionnel est à utiliser pour des analyses de haut niveau, pour faire ressortir des tendances et avoir un aperçu global grâce à des graphiques, pour faire des calculs de performances de l'entreprise sur une durée ou encore pour mettre en évidence des chiffres macroéconomiques ou des tendances. Il est plus que recommandé d'utiliser des tables agrégées pour ce type de modèle, le nombre de lignes des tables de faits pouvant fortement impacter les performances.

Une tactique souvent utilisée est de disposer de deux modèles : un dimensionnel se basant sur des tables agrégées et un relationnel basé sur les tables détaillées. De cette manière, les rapports principaux à vocation analytique sont basés sur le modèle dimensionnel dont le volume de données est réduit, et les rapports détaillés (accessibles depuis le rapport principal grâce à des accès au détail) sont basés sur le modèle relationnel permettant d'extraire des données détaillées plus volumineuses.

On peut imaginer un rapport principal affichant les ventes de l'année d'une entreprise sur un graphique, sur lequel il est possible de cliquer afin d'extraire le détail des ventes de la période choisie, grâce aux filtres sélectionnés. L'auteur a expérimenté le rapport détaillé basé sur le modèle dimensionnel : les performances étaient déplorables ! Alors que le même rapport basé sur le modèle relationnel s'affichait quasi instantanément (pour le même résultat final !). De cette manière, les deux solutions sont couplées, chacune étant utilisée là où elle est la plus performante.

■ Remarque

Le modèle dimensionnel est à privilégier pour des rapports basés sur des tables agrégées alors que les modèles relationnels se comportent mieux pour des rapports détaillés de type liste, tableau-croisé, etc. contenant beaucoup de lignes et pour lesquels l'interactivité n'est pas de mise.