

Dans la chronique précédente de cette série, nous avons expliqué que **le Big Data est l'extension de l'informatique décisionnelle**. Cette dernière désigne formellement les méthodes, techniques et outils informatiques utilisés pour piloter une activité et aider à la prise de décision. Les cubes OLAP font partie de l'arsenal des outils informatiques mis à disposition par un Système d'Information Décisionnel pour faciliter l'interrogation des données à large échelle.

Ainsi, pour comprendre comment tirer parti des données d'un point de vue technique, il est indispensable de comprendre ce que c'est qu'un cube OLAP. Dans cette chronique, qui est la 7ème d'une série consacrée à l'interrogation des données à large échelle, nous verrons la définition et l'importance d'un cube OLAP, son fonctionnement, son déploiement en production, ainsi que les solutions technologiques OLAP les plus matures disponibles sur le marché.

Sommaire

- 1. 1 – Analyse décisionnelle : de l'OLTP à l'OLAP
- 2. 2 – Comment fonctionne un cube OLAP ?
 - 2.1. 2.1.1. Définition du concept de cube OLAP
 - 2.2. 2.2.2. Fonctionnement d'un cube OLAP
 - 2.3. 2.3.3. Les opérations multidimensionnelles : les clés de la manipulation d'un cube OLAP
- 3. 3 – Conception d'une base de données multidimensionnelle
- 4. 4 – Implémentation et déploiement d'un cube OLAP
 - 4.1. 4.1.1 – Caractéristiques d'une technologie OLAP
 - 4.2. 4.2.2 – Les différentes technologies OLAP
 - 4.3. 4.3.3 – Les solutions OLAP du Big Data

1 – Analyse décisionnelle : de l'OLTP à l'OLAP

Pour comprendre ce qu'est un cube OLAP et son intérêt dans un système d'information décisionnel, il faut commencer par comprendre qu'un Système d'Information est composé de 2 sous-systèmes : un *Système d'Information Organisationnel* (ou OLTP) et un *Système d'Information Décisionnel* (ou OLAP). Chacun de ces systèmes soutient un processus bien particulier.

Le système organisationnel ou OLTP soutient les processus métiers. A titre de rappel, les processus métiers sont les processus qui dans une entreprise, concourent directement à l'atteinte de ses objectifs. Par exemple la vente, le marketing, la manutention, etc. D'un point de vue informatique, ces processus sont soutenus à l'aide d'un **OLTP** (*OnLine Transactional Processing*). L'OLTP désigne l'ensemble des technologies qui sont utilisées pour automatiser ou informatiser les processus métiers. En premier plan on pense par exemple à l'FF

grand nombre de ses processus métier. Cependant, ils ne sont pas adaptés pour les analyses décisionnelles, spécialement avec les exigences du Big Data. En effet, l'OLTP est lui-même adossé sur un SGBDR, or le SGBDR est construit sur une approche de stockage qui ne lui permet pas d'interroger les données tout en satisfaisant les 4 contraintes ACID de la gestion des transactions. C'est pourquoi un second composant est nécessaire dans le système d'information de l'entreprise, le système décisionnel ou OLAP.

Les processus décisionnels, à la différence des processus métiers, ne concourent ni directement, ni indirectement à l'atteinte des objectifs stratégiques de l'entreprise (par exemple, faire du profit, augmenter le chiffre d'affaire). Leur rôle est de favoriser le suivi de gestion des activités qui forment les processus métiers. *Le suivi de gestion est l'activité informationnelle continue de monitoring de chacune des activités opérationnelles qui forment les processus métiers implémentés dans le système d'information organisationnel ou l'OLTP, menant à des actions ou à des décisions d'ajustement régulier à chaque phase du processus.*

L'élément clé pour faire ce suivi de gestion est ce que l'on appelle **l'indicateur de performance** (en anglais *KPI – Key Performance Indicator*). *Un indicateur de performance se définit comme étant une statistique ciblée et contextualisée selon une préoccupation de mesure, résultant de la collecte de données sur un élément lié au fonctionnement d'une organisation.* Le KPI constitue le maillon INDISPENSABLE du suivi de gestion puisqu'elle permet de mesurer le niveau d'accomplissement des activités qui forment le processus. Armé de la connaissance de cette mesure, les gestionnaires peuvent alors coordonner les activités du processus métier correctement.

Les KPI sont regroupés de façon expressive dans ce que l'on appelle **un tableau de bord** et ensemble ceux-ci permettent de rapprocher le gestionnaire de l'information stratégique nécessaire au pilotage sain de son entreprise. Le tableau de bord et les KPI donnent une vision stratégique de ce qui se passe dans l'entreprise à une période donnée. Cependant, trois conditions sont nécessaires pour rapprocher les décideurs de l'information stratégique nécessaire au pilotage de l'entreprise à l'aide du tableau de bord :

- Le tableau de bord doit être arrimé aux processus métiers de l'entreprise, c'est là qu'il y trouvera sa pertinence et sa crédibilité ;
- Le tableau de bord doit être alimenté par un bon Système d'information, c'est là qu'il y trouvera sa qualité ;
- Le tableau de bord doit être soutenu par une bonne technologie et s'y subordonner, c'est là qu'il y trouvera sa flexibilité et sa facilité d'utilisation.

C'est pour répondre à ces trois conditions indispensables qu'un Système d'Information dit « décisionnel », encore appelé OLAP (Online Analytical Processing) est nécessaire. En fait, la réponse aux trois conditions citées précédemment demande une organisation pas seulement technologique, mais aussi à la fois en termes de ressources humaines et en termes de processus. C'est pourquoi *on définit le système d'information décisionnel comme l'ensemble des processus, des ressources humaines, des procédures et des technologies qui permettent de transformer, restituer et diffuser les informations indispensables au suivi et à la gouvernance stratégique de l'entreprise.* **Le système informatique qui soutient le SID est qualifié d'OLAP (Online Analytical Processing)**, pour dire que les traitements qui y sont effectués, à la différence des OLTP, sont des traitements analytiques.

2 – Comment fonctionne un cube OLAP ?

Nous espérons qu'à ce stade, vous avez compris l'intérêt de l'OLAP pour le système d'information d'une entreprise. Maintenant, nous allons expliquer le fonctionnement d'un cube OLAP proprement dit. En fait, on a vu l'OLAP, mais pourquoi « cube » OLAP ?

2.1. Définition du concept de cube OLAP

Il faut savoir que tout comme un système OLTP s'appuie sur un SGBDR, le système décisionnel OLAP s'appuie sur le Data Warehouse. Le Data Warehouse est le socle indispensable pour obtenir les réponses aux questions essentielles à la prise de décision et au pilotage de l'entreprise. Si vous prêtez attention, vous vous rendrez compte que conceptuellement, un Data Warehouse est un modèle multidimensionnel. C'est *un modèle relationnel dit dénormalisé.*

Or, quitter d'un modèle de base de données opérationnelle à un modèle de base de données décisionnelle revient à transformer les éléments du modèle de base de données opérationnelle en « dimensions » et en « faits ». Nous y reviendrons dans le point suivant. Cette transformation est nécessaire car comme dit plus haut, le modèle OLTP fournit uniquement une vue aplatie et statique des données d'

```

SELECT p.product, SUM(t.ventes)
FROM t_sales t
INNER JOIN product p ON t.product_id = p.id
INNER JOIN zone_chalandise ch ON t.zone_id= ch.id
WHERE ch.id = 'Z'
GROUP BY p.product

```

Ce type de requête ne peut tout simplement pas être répondu par un SGBD dans des délais raisonnables à cause des contraintes nécessaires pour garantir la cohérence des transactions.

Pour répondre à cette limite, une structure de données qui consiste à stocker les données représentées sous forme de point dans un espace à plusieurs dimensions a été mise sur pied, cette structure particulière de données s'appelle un **hyper cube** ou plus conventionnellement un **cube**.

Plus formellement, **Un cube OLAP est une structure de données multidimensionnelle stockant les faits** (voir définition [Data Warehouse](#)) **comme des mesures** (voir plus bas) **indexées par plusieurs dimensions**. Ainsi, *chaque cellule d'un cube représente la mesure ou valeur quantitative d'un fait sur le croisement de plusieurs dimensions*.

De cette définition, vous pouvez voir que **l'intérêt d'un cube OLAP est d'offrir à l'utilisateur la capacité de faire des analyses multidimensionnelles ou des agrégations par axe de dimension dans l'espace**.

2.2. Fonctionnement d'un cube OLAP

Pour que vous compreniez le fonctionnement d'un cube OLAP, considérons le tableau suivant :

Vendeur	Produit	Date de vente	Prix de vente
Juvéнал	Accessories	01/04/2001	800
<u>Myles</u>	Bikes	09/05/2001	1400
<u>yvan</u>	<u>Clothing</u>	02/02/2002	500
<u>yvan</u>	Components	02/03/2002	1000
Juvéнал	Bikes	15/03/2002	1800
Juvéнал	Bikes	10/03/2003	2000
<u>Myles</u>	Components	12/10/2003	700
<u>Myles</u>	Bikes	25/12/2003	2200
Juvéнал	Components	10/01/2004	500
... * * * *
<u>yvan</u>	Bikes	15/11/2004	2500

Supposons que nous souhaitons calculer la somme des ventes par produit et par année. La représentation de ce tableau sous forme d'une structure multidimensionnelle fournit le cube OLAP suivant.

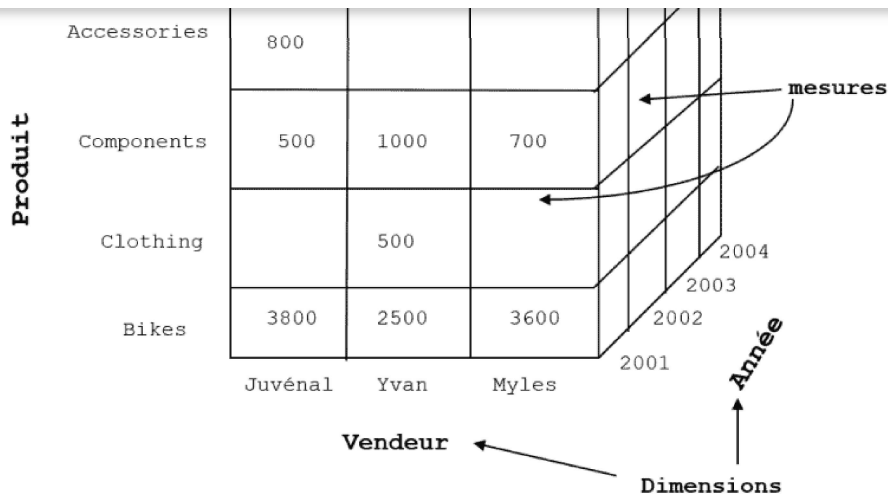


Figure : représentation du tableau en cube OLAP. Les 3 colonnes de catégories deviennent des dimensions dans le cube, tandis que le prix devient la mesure, la valeur correspondant au croisement des 3 dimensions simultanément

Dans le cube, nous avons trois dimensions : Année, Vendeur et Produit ; la mesure c'est le prix de vente. Pour rendre le cube fonctionnel, on applique à ses cellules (donc à la mesure) une fonction d'agrégation, cette fonction peut être soit une somme, une moyenne, un maximum ou un minimum, bref toute fonction qui regroupe ou compare un ensemble de données. Pour faciliter la représentation graphique du cube, nous avons présenté la somme des mesures uniquement sur la première face, mais en réalité, toutes les cellules comportent des valeurs pour chaque recoupement de dimensions et répondent à une requête bien précise. Par exemple la cellule qui est à l'intersection de l'attribut « Bikes » de la dimension Produit et de l'attribut « Juvénal » de la dimension Vendeur a pour somme de ventes 3800; ceci correspond à la requête chiffre d'affaire réalisé par le Vendeur Juvénal sur les Produits « Bikes » depuis 2001. Cependant, certains croisements de dimensions peuvent ne pas avoir de valeur (agrégation nulle dans ce cas), c'est ce qui explique que certaines cellules du cube soient vides.

2.3. Les opérations multidimensionnelles : les clés de la manipulation d'un cube OLAP

Pour manipuler un cube OLAP, on utilise des *opérateurs multidimensionnels*. Les cubes multidimensionnels disposent de 3 opérateurs multidimensionnels pour leur exploitation à savoir l'opérateur de rotation (ROTATE/ SWITCH), d'extraction (SLICING/DICING) et de sélection (DRILL UP/DRILL DOWN/DRILL THROUGH). Voyons ensemble comment utiliser ces opérateurs.

- **La rotation du cube (ROTATE/SWITCH)** : il est possible d'effectuer une rotation à 90° de 2 dimensions du cube. Cette opération techniquement s'appelle le *ROTATE CUBE* ou *SWITCH CUBE*. Dans l'exemple du cube représentant notre tableau, on a pivoté ou roter ou permuter la dimension Année avec la dimension Produit. Remarquez que les résultats ne sont pas les mêmes, on est maintenant capable d'obtenir le chiffre d'affaire réalisé par le vendeur Juvénal à l'année 2001.

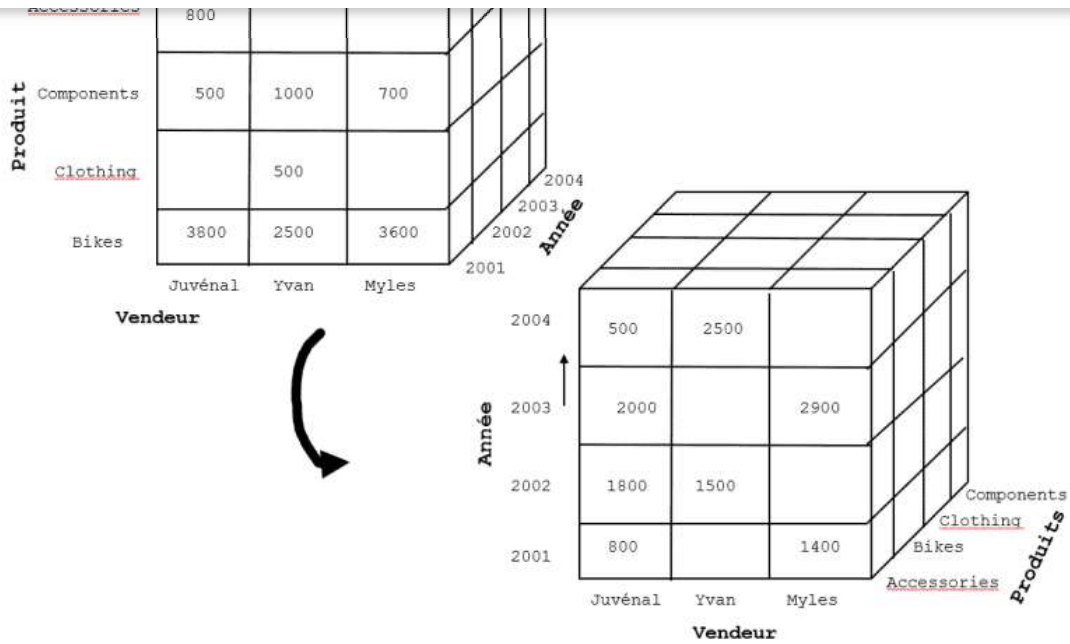


Figure : opération de rotation du cube

- L'opération d'extraction du cube (SLICING/DICING)** : cette opération consiste à extraire du cube un bloc de données correspondant à un croisement entre plusieurs dimensions. Ce bloc permet alors de recalculer plus facilement le cube. On distingue 2 types d'opération d'extraction de données du cube, le **SLICING**, qui consiste à extraire les mesures correspondant à une certaine dimension en s'appuyant sur un critère de valeur, par exemple toutes les valeurs inférieures à 2000 ; et le **DICING**, qui consiste à extraire un bloc de mesures en s'appuyant sur des critères d'attributs de dimensions. Dans la figure ci-après, nous avons effectué une opération de DICING de notre cube précédent, le nouveau cube est formé de l'attribut 2002 de la dimension Année et de tous les attributs de la dimension Produit excepté l'attribut « Accessories » ; cela correspond à la zone grisée de par et d'autre du cube. **Attention !!!** Lorsqu'on fait une extraction de cube, on obtient une nouvelle partition, distincte du cube mère.

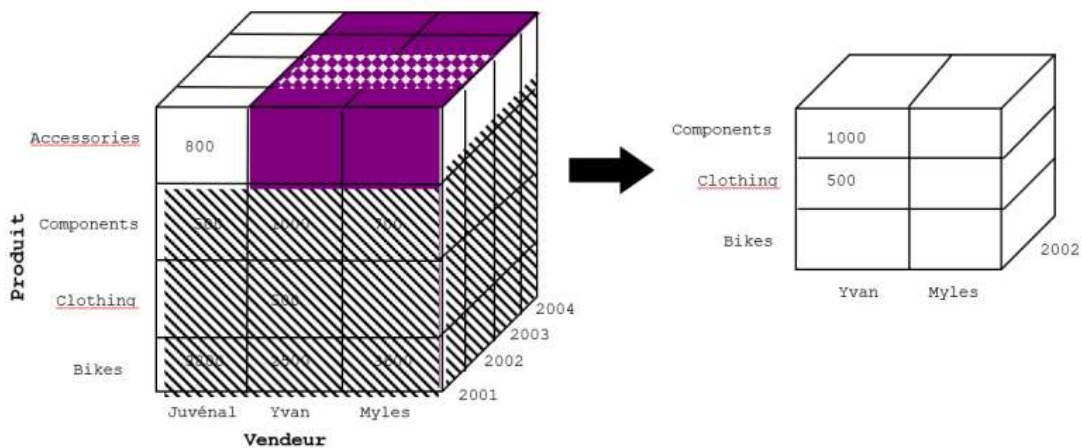


Figure : DICING d'un cube OLAP

- L'opération de sélection** : la sélection est similaire l'extraction, à la seule différence que la sélection navigue simplement à travers les dimensions du cube, sans le partitionner. Les opérations de sélection sont des réponses à des requêtes, alors que l'extraction sort un bloc du cube. Les opérations de SLICING et DICING sont également des opérations de sélection. La sélection permet également de naviguer selon les niveaux de profondeur d'information ou de « zoomer » sur des sur une ou plusieurs dimensions précises. 3 opérations de sélections permettent de naviguer dans le cube : le **DRILL UP**, qui permet de voir la synthèse des

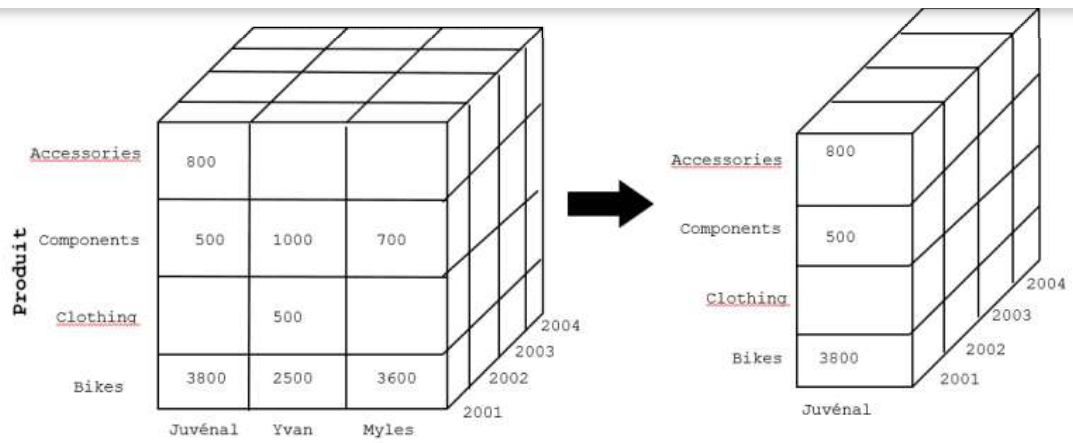


Figure : opération de SLICING/DRILL DOWN d'un cube OLAP

Initiation à l'écosystème Hadoop

Maitrisez l'utilisation des technologies Hadoop

Initiation à l'écosystème Hadoop

Vous travaillez sur un projet Big Data ? Ce livre numérique vous aidera à développer un panorama global des technologies qui permettent de résoudre les challenges soulevés par la valorisation de données à large échelle et vous permettra d'être initié de façon pragmatique à 7 technologies principales de l'écosystème Hadoop, notamment Spark, Hive, Pig, HBase, Oozie, Sqoop et Kafka.

Cliquez sur le lien suivant pour télécharger librement le livre numérique et l'obtenir directement dans votre boîte mail.

Votre email personnelle

3 – Conception d'une base de données multidimensionnelle

Maintenant, vous connaissez l'intérêt d'un cube OLAP ainsi que son mode de fonctionnement. Nous allons vous montrer comment le concevoir.

Comme nous vous l'avons expliqué plus haut, la conception d'un cube OLAP passe par la transformation du modèle de données des bases de données opérationnelles en *modèle multidimensionnel* du Data Warehouse, plus adéquat pour les requêtes multidimensionnelles.

Pour ce faire, on utilise une technique de modélisation particulière qu'on appelle **la modélisation multidimensionnelle**. La modélisation multidimensionnelle est *une technique de modélisation qui consiste à représenter les données sous forme de points dans un espace à plusieurs dimensions*. L'intérêt est d'offrir à l'utilisateur la capacité de faire des **requêtes multidimensionnelles** ou agrégations par axe de l'espace.

Nous avons vu plus haut que les requêtes multidimensionnelles ne peuvent pas être répondues avec un système opérationnel OLTP, car celui-ci privilégie la gestion transactionnelle. C'est pourquoi il faut passer par un modèle multidimensionnel.

Pour modéliser les données qui vont être représentées sous forme de point dans un espace à plusieurs axes, la modélisation multidimensionnelle va définir 2 concepts : la dimension, le fait.

- Les **dimensions** ce sont les regroupements de données caractérisant un sujet métier bien précis. Par exemple le Client, le vendeur, le temps, la Date, la localité sont tous des sujets. La table dans laquelle ces données sont regroupées s'appelle une **table de dimension**, et dans ce cas, les sujets sont des regroupements d'attributs ou colonnes décrivant la même réalité. Par exemple, le sujet Client est composé des attributs nom, prénom, âge, sexe, ...etc. Chaque [table de] dimension constitue un axe d'analyse, qui se traduit par une insertion de l'identifiant unique de la dimension dans la table de fait. Plus les attributs qui composent la dimension seront bien choisies et bien construites (discrètes et finies), plus le niveau de granularité du Data Warehouse sera élevé ;
- Le **fait** : On ne peut pas définir la notion de fait sans définir la notion de mesure. La mesure est la valeur quantitative à laquelle est associé un fait. En réalité, le fait est un événement quantitatif correspondant à une opération métier. Dans la modélisation multidimensionnelle, le fait référence à l'ensemble des données générées par cet événement et la mesure associée. Par exemple, dans un super marché, le fait peut correspondre à un bip de scanner lorsqu'on passe l'article en caisse, cet événement (le bip) génère les données sur la quantité et le prix d'achat de l'article ; en comptabilité, un fait correspond à un enregistrement d'une opération dans le journal, en CRM, le fait correspond à un achat client. L'ensemble des faits sont sérialisés dans une table qu'on appelle en modélisation multidimensionnelle la **table de faits**. Chaque fait est collecté selon un **niveau de grain** ou un niveau de granularité qui dicte le niveau de détail ou de profondeur jusqu'au quel on peut aller dans les analyses de données. Le niveau de granularité est donc la fondation de la performance opérationnelle d'un Data Warehouse, il doit être choisi avec beaucoup de soin. Dans l'exemple précédent, si lors du bip du scanner, vous avez décidé de stocker uniquement l'année de l'achat auquel correspondait le bip, vous ne pourrez pas descendre dans les analyses mensuelles, ni trimestrielles, encore moins hebdomadaires. Faites donc très attention lorsque vous décidez du niveau de granularité !

Une fois que vous avez identifié vos dimensions et vos mesures, il est temps de concevoir votre modèle multidimensionnel. Pour ce faire, vous avez le choix entre 3 approches : vous pouvez le concevoir selon le **modèle en étoile**, selon le **modèle en flocon de neige**, ou selon le **modèle en constellation**. Il n'y a pas vraiment de différence fondamentale entre ces 3 modèles si ce n'est au niveau du nombre de jointures. En général, c'est le modèle en étoile qui est le plus utilisé par les architectes. La figure ci-après illustre un modèle multidimensionnel modélisé en étoile.

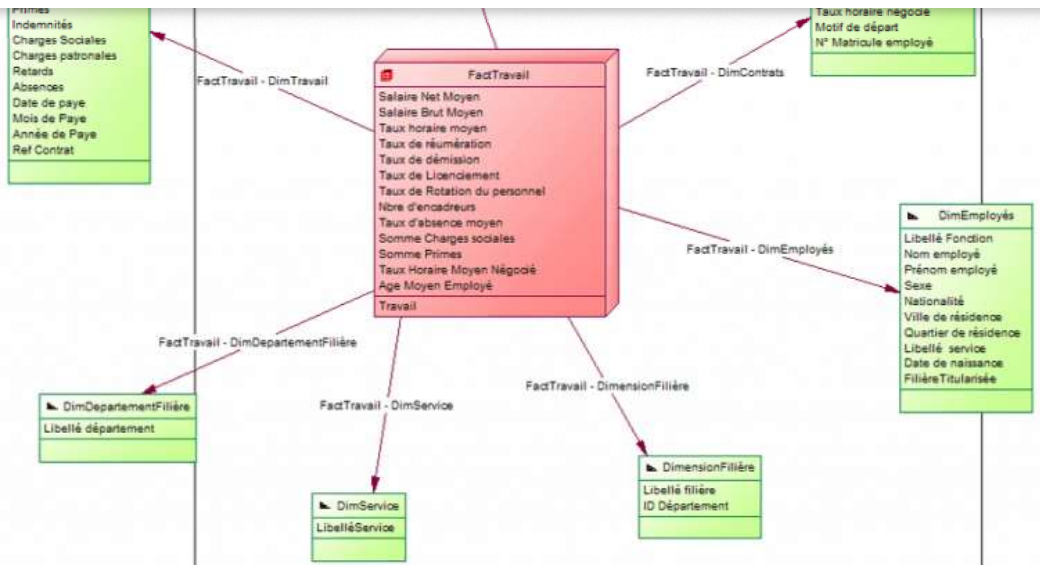


Figure : modélisation multidimensionnelle en étoile

Le modèle multidimensionnel est la première étape dans la construction d'un cube OLAP. Une fois que vous l'avez obtenu, il faut l'implémenter dans une « base de données » multidimensionnelle. C'est l'objet du point suivant.

4 – Implémentation et déploiement d'un cube OLAP

Le point de départ d'un cube OLAP c'est le modèle multidimensionnel. Dans cette partie, nous allons parler des technologies utilisées pour implémenter et déployer les cubes en production, c'est-à-dire les technologies qui sont capables de stocker les données de façon multidimensionnelle et répondre aux requêtes multidimensionnelles. Ces types de technologies sont couramment appelées les **technologies OLAP**.

4.1 – Caractéristiques d'une technologie OLAP

Gérer le stockage des données dans plusieurs dimensions est très compliqué d'un point de vue conceptuel. C'est pourquoi il est difficile de caractériser une technologie OLAP. Mais d'une façon générale, on s'accorde pour dire qu'une technologie peut être qualifiée d'OLAP si elle respecte au moins 5 conditions, qu'on regroupe sur le sigle FASMI – *Fast Analysis of Shared Multidimensionnel Information* :

- **FAST (Rapide)** : cette caractéristique signifie que le système vise à fournir des réponses aux utilisateurs avec une latence inférieure ou égale à 5 secondes. En effet, des études menées aux pays-bas ont montré que les utilisateurs considèrent qu'un traitement de Reporting a échoué si les résultats ne sont pas affichés dans les 30 secondes. Toute technologie qui se dit de type OLAP doit donc fournir des performances suffisamment élevées pour répondre aux requêtes des utilisateurs dans les 5 secondes ;
- **ANALYSIS (Analyse)** : cette caractéristique signifie que la technologie est capable d'effectuer des calculs statistiques et des calculs d'agrégation pour répondre aux besoins métiers des utilisateurs. De plus, l'interface graphique qu'il fournit pour les requêtes est suffisamment intuitive pour les utilisateurs ;
- **SHARED (partagé)** : la technologie OLAP implémente des mécanismes de sécurité jusqu'au niveau de la cellule du cube pour garantir la confidentialité, et garantir le verrouillage en cas d'utilisation concurrente ;
- **MULTIDIMENSIONAL (multidimensionnel)** : cette caractéristique est la fondation d'une technologie OLAP. Toute technologie OLAP offre une vue multidimensionnelle des données, nécessaire pour l'analyse décisionnelle des données ;
- **INFORMATION (information)** : Cette caractéristique fait référence à la quantité de données qui peut être stockée et agrégée par la technologie OLAP ;

3 options s'appliquent au stockage des données en structure multidimensionnelle. Attention, lorsque nous parlons de stockage ici, nous parlons de persistance des données, les données sont persistées dans le moteur OLAP comme elles le seraient dans un SGBD, du moins au moins pour la durée d'une session, et pas simplement le temps d'exécution d'une requête. Il y'a trois options pour le stockage du cube. Soit celui-ci est stocké sous forme d'une base de données relationnelle, soit sous forme de base de données multidimensionnelle ou soit sous forme de fichier client :

- **Stockage sous forme d'une base de données relationnelle** : dans ce cas les données sont dénormalisées et stockées suivant un schéma en étoile dans un SGBD relationnel ;
- **Stockage sous forme d'une base de données multidimensionnelle** : dans ce cas, les données actives sont stockées dans une base multidimensionnelle dans un serveur. Dans la majorité des cas, la base multidimensionnelle est persistée sur le disque dur, dans d'autres cas, la base est montée en mémoire RAM pour une plus grande performance ;
- **Stockage sous forme de fichiers clients** : dans ce cas, les données les plus utilisées sont extraites et enregistrées dans un ensemble de fichiers de petite taille et persistés sur des machines clientes.

LA SOLUTION DU POINT DE VUE TRAITEMENT

Au même titre qu'il y'a 3 options possibles pour le stockage des données en OLAP, les 3 options sont disponibles pour le traitement des données. Par contre c'est important de noter que les calculs multidimensionnels ne sont en général pas effectués là où les données sont stockées : le SQL, un serveur multidimensionnel et un client multidimensionnel :

- **Le SQL** : le SQL est utilisé comme moyen de traitement si le cube est stocké dans SGBDR. Le problème avec le SQL c'est qu'il n'est pas adapté à la définition des requêtes multidimensionnelles, du coup la performance en prend un sacré coup !
- **Un serveur multidimensionnel** : dans le cas où le cube est stocké sous forme d'une base multidimensionnelle, un moteur de calcul multidimensionnel séparé de la base multidimensionnel et utilisant son propre serveur est requis pour le traitement, dans ce cas, un langage spécifique est nécessaire pour définir les requêtes multidimensionnelles. Bien qu'il n'y'ait pas un langage standard définir les requêtes multidimensionnelles, on note sur le marche la présence des langages comme le **MDX (Multi Dimensional eXpression)** de Microsoft ;
- **Un client multidimensionnel** : dans le cas où les extraits des données du cube sont stockés dans des fichiers client (c'est-à-dire des fichiers persistés dans l'ordinateur de l'utilisateur métier), un moteur de calcul multidimensionnel dont l'hôte est la machine de l'utilisateur final est requis. Cette option n'est envisageable également que sous l'hypothèse que les utilisateurs finaux possèdent des ordinateurs suffisamment puissants pour effectuer des calculs multidimensionnels directement sur leur poste.

4.2 – Les différentes technologies OLAP

Le croisement de chaque option de stockage et traitement des données multidimensionnelles citées ci-dessus a donné naissance à plusieurs types de technologies OLAP. C'est cette typologie qui permet de classer et différencier les solutions OLAP offertes par les éditeurs. Le tableau ci-dessous en donne le récapitulatif.

Option de traitement	SQL	ROLAP		
	Serveur Multidimensionnel	HOLAP	MOLAP	
	Client Multidimensionnel	DOLAP	DOLAP	DOLAP

Il y'a donc théoriquement 4 approches technologiques OLAP : ROLAP, HOLAP, MOLAP et DOLAP. Les solutions technologiques que vous choisirez pour vos analyses décisionnelles seront toujours dans l'une de ces catégories :

- **ROLAP** : On nomme ROLAP l'approche **Relationnel OLAP**. Les données sont stockées sous la forme de bases de données relationnelles. Elles sont modélisées sous la forme de schémas en étoile ou flocon. Les requêtes multidimensionnelles doivent alors être traduites en requêtes relationnelles (SQL). L'outil **SSAS (SQL Server Analysis Services) de Microsoft** est un exemple de logiciel qui implémente l'approche ROLAP
- **MOLAP**: on nomme MOLAP l'approche **Multidimensionnelle OLAP**. La technologie de stockage est multidimensionnelle. Les données sont stockées sous la forme de tableaux multidimensionnels, des index multidimensionnels sont définis. Elle utilise des techniques de compression face à la faible densité des données. La taille des données pouvant être ainsi stockées est faible par rapport à la solution ROLAP. Cependant, les requêtes sont écrites de manière intuitive et efficace. Cependant, il faut redéfinir un langage de manipulation des données du tableau du tableau. Le **MDX (Multi Dimensional eXpression)** est un exemple de langage multidimensionnel. Dans le domaine du contrôle de gestion et en finance par exemple, ce sont les moteurs MOLAP qui sont utilisés ; **Hyperion ESSBASE**, racheté par Oracle, est un exemple de solution MOLAP ;
- **HOLAP**: On nomme HOLAP l'approche **Hybride OLAP**. Cette technologie combine les deux solutions précédentes. Les données sont stockées dans une base de données relationnelle, et les calculs sont faits dans une base multidimensionnelle.
- **DOLAP** : on nomme DOLAP l'approche **Desktop OLAP**. Dans cette approche, le moteur OLAP est un client lourd installé en local, c'est-à-dire qu'il est installé directement sur le poste du client, les données y sont stockées de façon soit de façon relationnelle, soit de façon multidimensionnelle, ou soit sous la forme de fichiers plats. Ainsi, les utilisateurs peuvent faire leurs analyses directement sous leur poste dans dépendre du service informatique. C'est notamment cette approche technologique qui est à la base du concept de Self Service Business Intelligence. **PowerPivot de Microsoft Power BI**, et **QlikView** de Qlik sont des exemples de DOLAP

Voilà ! Nous avons fait le tour des technologies OLAP. Quelque soit la solution que vous utiliserez, elle se retrouvera toujours dans l'une des 4 catégories. Avant de quitter ce point, il est important que vous ne confondiez pas Data Mart avec cube OLAP. Beaucoup de personnes font souvent cet amalgame.

Le Data Mart et le Cube OLAP sont 2 choses différentes. Tandis que le premier fait référence à une approche de construction de Data Warehouse, l'autre fait référence à une approche de stockage et de requête des données du Data Warehouse. Le Data Mart existe indépendamment d'un cube OLAP. Le cube OLAP est un moyen comme un autre pour stocker et adresser des requêtes au Data Mart. Nous allons conclure cette chronique avec un tour d'horizon des technologies OLAP utilisées en Big Data.

4.3 – Les solutions OLAP du Big Data

Avec l'explosion du volume et de la variété des données qui caractérise le Big Data, l'approche la plus appropriée pour le traitement et le stockage de données consiste à utiliser le **cluster computing**. Avec cette approche, le traitement des données est distribué et parallélisé entre les nœuds d'un cluster. Du coup, la multidimensionnalité propre aux requêtes décisionnelles est gérée facilement grâce aux nœuds de calcul du cluster et il n'y'a pas nécessité de L'OLAP au sens strict du terme.

Les solutions décisionnelles du Big Data tournent toutes autour de 3 stratégies : **Moteurs Hadoop**, **moteur natif Hadoop** ou **SGDB MPP**. Le tableau ci-après les résume très bien.

	Utilisant le SQL ou un langage similaire.	Installe sur le cluster Hadoop	Exécute comme un job Hadoop classique	
Moteur natif SQL sur Hadoop	Catégorie de moteurs SQL sur Hadoop qui exécutent le SQL directement sur le HDFS sans le transformer en job MapReduce ou job TEZ (comme le fait Hive ou Pig par exemple)	Indépendant d'Hadoop, mais installé sur le cluster Hadoop	Le code SQL écrit est transformé en plan d'exécution SQL et directement exécuté sur les données stockées dans le HDFS	Impala
SGBD relationnel distribué sur Hadoop ou Moteur SQL relationnel MPP	SGBDR distribué ou SGBD MPP (<u>Massively Parallel Processing</u>) classique au même titre que Teradata ou <u>GreenPlum</u> qui est capable d'exécuter le SQL sur un cluster Hadoop	Complètement Indépendant du cluster Hadoop	Les tables de la base de données sont partitionnées et distribuées sur les nœuds du cluster Hadoop (et non des nœuds du serveur MPP). Lorsque le SQL de l'utilisateur est soumis, le plan d'exécution du SQL est géré au niveau du master et reparti entre les nœuds du cluster Hadoop qui contiennent les partitions des tables concernées.	Teradata, HAWQ (version Hadoop de <u>GreenPlum</u>), <u>Druid</u>

Figure : solutions OLAP pour le Big Data

Initiation à l'écosystème Hadoop

Maîtrisez l'utilisation des technologies Hadoop

technologies principales de l'écosystème Hadoop, notamment Spark, Hive, Pig, HBase, Oozie, Sqoop et Kafka.

Cliquez sur le lien suivant pour télécharger librement le livre numérique et l'obtenir directement dans votre boîte mail.

Recevoir ma copie

Nous sommes arrivés au terme de cette chronique assez exhaustive sur l'analyse décisionnelle à l'aide d'un cube OLAP. Dites-nous en commentaire laquelle de ces approches vous semble plus appropriée pour votre projet ?

Categories: **PROJET BIG DATA**



Data Transition Numerique

Vous dotez des compétences dont vous avez besoin pour tirer profit des opportunités du numérique

4 comments

Top rated ▼ comments first

Comment as a guest:

or

[Login on website](#)

Gluc0.Amine

Super Article ! est-il possible de le telechargé pour une utilisation ulterieur ?



0

0

votre navigateur web. Cela vous permettra de le consulter ultérieurement.

Cordialement,

Juvénal



0

0



Reply to Juvénal JVC

Lolo

Hello super article ! Cependant ne manque t-il pas l'attribut « Juvénal » de la dimension Vendeur sur votre cube après l'opération de dicing ?



0

0



Reply to Lolo

Juvénal JVC

Bonjour Lolo,

oui tu as raison, il manque « Juvenal » dans l'opération de Dicing. C'est une erreur de notre part. Merci de nous l'avoir signalé.

Bonne fête,

Juvénal



Add your comment...

Rechercher sur le site

Rechercher...



FORMATION SPARK GRATUITE



Vous souhaitez travailler dans le Big Data ? Vous souhaitez vous orienter vers un parcours de Data Engineer ?

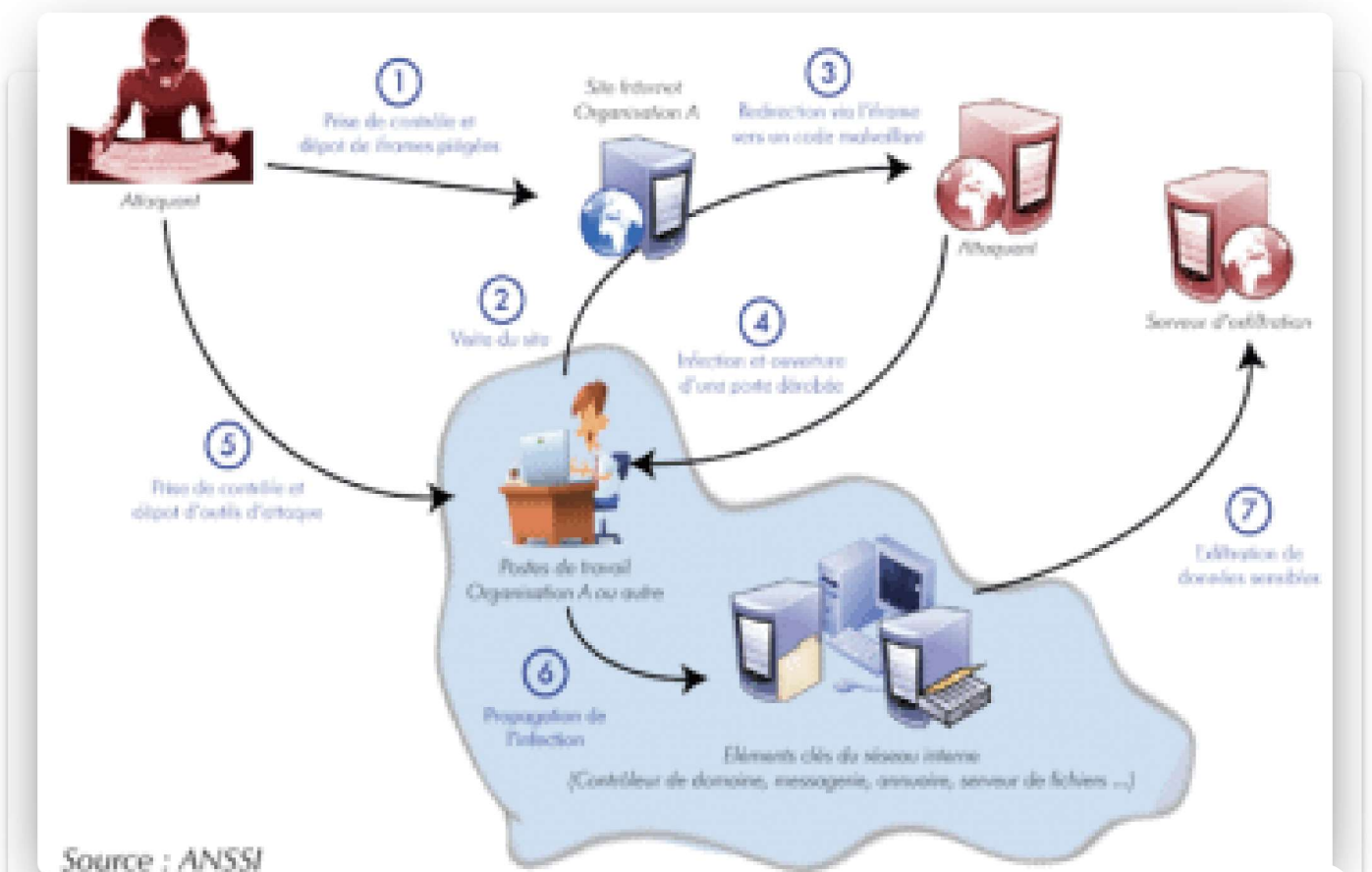
Si c'est le cas, alors vous savez déjà probablement que la maîtrise de Spark est obligatoire dans quasiment toutes les missions Big Data, spécialement pour les postes de Data Engineer !

Remplissez votre email personnelle ci-bas pour vous inscrire à la formation

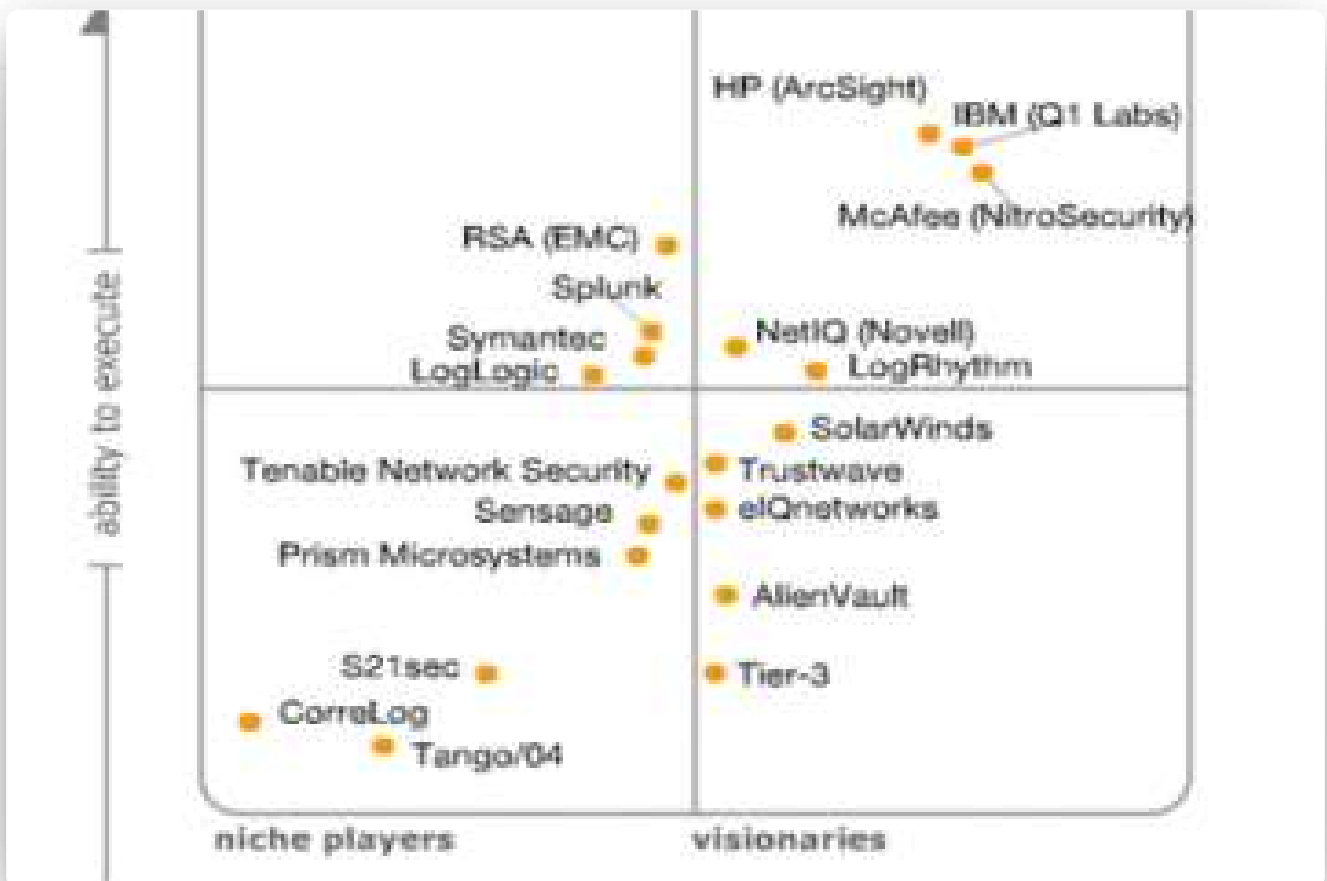
votre adresse mail personnelle

M'inscrire à la formation

Related Posts



Pour continuer à assurer des niveaux de productivité croissants qui leur permettent de faire face aux challenges posés par la nouvelle économie d'aujourd'hui, les entreprises ont compris qu'elles devaient communiquer, communiquer d'une part en interne [Read more...](#)



PROJET BIG DATA

Cybersécurité & Big Data : comprendre l'intérêt et les limites du SIEM

Dans la chronique précédente de cette série sur la cybersécurité et le Big Data, nous avons expliqué que la cybersécurité fournit une vision du niveau d'exposition d'une entreprise à une attaque, en s'appuyant sur les [Read more...](#)



PROJET BIG DATA

Cloudera CDH vs Hortonworks HDP vs MapR CDP : 5 critères pour choisir votre distribution Hadoop

En juillet 2014 dernier, l'ex ministre Arnaud MONTEBOURG présentait les 34 plans de la Nouvelle France Industrielle (NFI), un projet de ré industrialisation dont l'ambition est de positionner la France stratégiquement sur le Big Data. [Read more...](#)