Chapitre 2 :
Les concepts de base de Data Warehouse

Les concepts de base de Data Warehouse

Définition

Le concept de data warehousing remonte à la **fin des années 1980** lorsque les chercheurs d'IBM Barry Devlin et Paul Murphy ont développé le « business data warehouse ». Essentiellement, le concept d'entreposage de données visait à fournir un modèle architectural pour le flux de données des systèmes opérationnels aux environnements d'aide à la décision.

Un Data Warehouse (entrepôt de données) est une technologie qui regroupe des données structurées provenant d'une ou de plusieurs sources afin qu'elles puissent être comparées et analysées pour une meilleure business intelligence.

Source : Oracle

Les concepts de base de Data Warehouse
Définition

Le Data warehouse (entrepôt de données) est une collection de données orientées sujet, intégrées, non-volatiles et historisées, organisées pour le support d 'un processus d 'aide à la décision (Inmon, 94).

Source : Copyright J. Akoka - I. Comyn-Wattiau - N.Prat

Les concepts de base de Data Warehouse
Définition

1. **Données orientées sujet :** réorganisation des données par sujets.

- Il n 'y a pas de duplication des informations communes à plusieurs sujets.
- La base de données est construite selon les thèmes qui touchent aux métiers de l'entreprise (clients, produits, risques, rentabilité, ...).
- Les données de base sont toutefois issues des Systèmes d'Information Opérationnels (SIO)

Les concepts de base de Data Warehouse
Définition

1. Données intégrées : IIntegration dans un data warehouse des systèmes sources différents.

- Les données, issues de différentes applications de production, peuvent exister sous toutes formes différentes.
- Il faut les intégrer afin de les homogénéiser et de leur donner un sens unique, compréhensible par tous les utilisateurs.
- Ile doivent posséder un codage et une description unique.

Les concepts de base de Data Warehouse
Définition

1. Données intégrées : IIntegration dans un data warehouse des systèmes sources différents.

- La phase d'intégration est longue et pose souvent des problèmes de qualification sémantique des données à intégrer (synonymie, etc...).

- Ce problème est amplifié lorsque des données externes sont à intégrer avec les données du SIO.

Les concepts de base de Data Warehouse
Définition

1. **Données non-volatiles :** Traçabilité des informations et des décisions prises

- Une information est considérée volatile quand les données sont régulièrement mises à jour comme dans les Systèmes d'Information Opérationnels.

- Dans un SIO, les requêtes portent sur les données actuelles. Il est difficile de retrouver un ancien résultat.

- Dans un DW, il est nécessaire de conserver l'historique de la donnée. Ainsi, une même requête effectuée à deux mois d'intervalle en spécifiant la date de référence de la donnée, donnera le même résultat.

Les concepts de base de Data Warehouse
Définition

1. **Données historisées :** les DW contiennent des données historiques, pas seulement des données actuelles.

- Dans un SIO, les transactions se font en temps réel, et les données sont mises à jour constamment. L'historique des valeurs de ces données n 'est généralement pas conservé car il est inutile.

- Dans un DW, la donnée n'est jamais mise à jour.

- Les données du DW s 'ajoutent aux données déjà stockées => ajout de couches de données successives, à la manière des strates géologiques

Les concepts de base de Data Warehouse
Définition

1. **Données historisées :** les DW contiennent des données historiques, pas seulement des données actuelles.

- Le DW stocke donc l'historique des valeurs que la donnée aura prises au cours du temps.
- Un référentiel de temps est alors associé à la donnée afin d'être capable d'identifier une valeur particulière dans le temps.
- Les utilisateurs possèdent un accès aux données courantes ainsi qu'à des données historisées.

Les raisons de créer un data warehouse

Les concepts de base de Data Warehouse
Les raisons de créer un data warehouse

Deux raisons principales :

- Le premier est de soutenir **la prise de décisions en fonction des données** plutôt que de devoir se fier uniquement à l'expérience et à l'intuition.

- La seconde est l'idée d'un **guichet unique**. En d'autres termes, les données dont nous avons besoin se trouvent toutes au même endroit plutôt que d'être dispersées entre les applications transactionnelles et opérationnelles d'où nous obtenons ces données lorsqu'il s'agit de prendre des décisions fondées sur les données.

Source : Alan Simon

Les concepts de base de Data Warehouse
Les raisons de créer un data warehouse

Avant L'idée d'un guichet unique, toute tentative de prise de décision basée sur les données nécessite de rechercher des données, soit dans les applications d'origine elles-mêmes, soit dans ce que nous appelons des fichiers d'extraction qui extrait les données d'une ou plusieurs applications. Mais pour la plupart, la prise de décision basée sur les données dans le passé était une grande problématique en raison des données qui sont dispersées.

Si nous considérons toutes ces vues de nos entreprises représentées par nos données ensemble comme une seule, nous avons en fait une discipline connue sous le nom de business intelligence, et par le data warehouse (Les deux sont entrés en scène à peu près au même moment vers 1990).

Source : Alan Simon

Les concepts de base de Data Warehouse

Les raisons de créer un data warehouse

Avec le Data Warehouse, nous intégrons toutes les données en un seul endroit et fournissons un guichet unique pour les données. Donc, nous pouvons nous concentrer sur l'analyse des données plutôt que sur la collecte et l'intégration répétées des données. Et nous avons Business Intelligence et le data warehouse comme une sorte de disciplines qui offrent une valeur énorme aux entreprises.

Source : Alan Simon

Différence entre Data Warehouse et une base de données

Les concepts de base de Data Warehouse

Différence entre Data Warehouse et une base de données

- Une **base de données** est une collection de données organisées. Par exemple, une base de données peut regrouper toutes les informations sur les clients ou sur les transactions.

- Un **data warehouse** est un système de reporting et d'analyse de données. Il fournit des performances élevées pour les requêtes analytiques.

Source : Oracle

Les concepts de base de Data Warehouse
Différence entre Data Warehouse et une base de données

4 différences phare entre database et data warehouse

1. Stockage vs analyse :

Une base de données est conçue principalement pour enregistrer des données. Un entrepôt de données, d'autre part, est conçu principalement pour analyser les données. Une base de données est normalement optimisée pour effectuer des opérations de lecture-écriture de transactions ponctuelles. Il n'est pas conçu pour effectuer de grandes requêtes analytiques de la même manière qu'un entrepôt de données.

Source : Oracle

Les concepts de base de Data Warehouse
Différence entre Data Warehouse et une base de données

4 différences phare entre database et data warehouse

2. Collecte vs catégorie :

Alors qu'une base de données est une collecte de données axée sur les applications, un entrepôt de données est plutôt axé sur une catégorie de données. Une base de données est normalement limitée à une seule application, ce qui signifie qu'une base de données équivaut habituellement à une application ; elle cible habituellement un processus à la fois. Un entrepôt de données, d'autre part, stocke les données d'un nombre quelconque d'applications. Un entrepôt de données comprend un nombre infini d'applications et cible autant de processus que nécessaire.

Source : Oracle

4 différences phare entre database et data warehouse

3. Fournisseur de données vs source d'analyse de données :

L'une des différences pratiques entre une base de données et un entrepôt de données est que le premier est un fournisseur de données en temps réel, tandis que le second est davantage une source d'analyse des données à mesure qu'elles sont enregistrées. Toutes les données peuvent être extraites d'un entrepôt de données pour être analysées chaque fois que cela est nécessaire.

Source : Oracle

Les concepts de base de Data Warehouse
Différence entre Data Warehouse et une base de données

4 différences phare entre database et data warehouse

1. Rapidité de stockage vs temps d'analyse :

Une base de données comporte généralement des tables complexes parce que les données sont organisées de telle sorte qu'aucun élément n'est dupliqué. Cette structure organisationnelle permet un traitement et un stockage très efficaces des données ; une réponse est très rapide. Un entrepôt de données, par contre, n'est pas conçu pour des transactions rapides, mais plutôt pour améliorer les requêtes analytiques, ce qui est obtenu en utilisant moins de tables et une structure plus simple.

Source : Oracle

Les concepts de base de Data Warehouse

Un système de gestion de base de données (SGBD)

Un système de gestion de base de données **(SGBD)** est le **logiciel qui facilite la gestion des bases de données**. Certains SGBD populaires incluent MySQL, MSSQL, Oracle et PostgreSQL. L'utilisateur peut écrire des requêtes en langage SQL (Structured Query Language) pour manipuler des données dans la base de données. Le processus d'exécution des requêtes dans la base de données s'appelle OLTP ou traitement transactionnel en ligne. Par conséquent, une base de données utilise OLTP. Globalement, une base de données aide à organiser un ensemble de données.

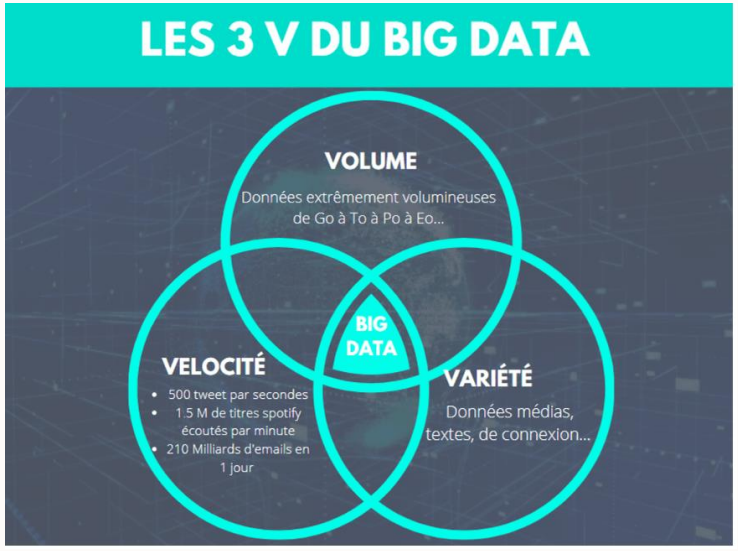
Source : Tibco

Big Data

Les concepts de base de Data Warehouse
Big data

Les Big data (données massives ou méga données) sont des ressources d'information très volumineuses, à vitesse très élevée et/ou de très grande variété qui nécessitent de nouvelles formes de traitement pour permettre une meilleure prise de décision, la découverte d'informations et l'optimisation des processus.

- **Volume:** désigne une très grande masse de données collectées, allant du téraoctet (1 To = 10^(12) octets) au zettaoctet (1 Zo = 10^(21) octets).
- **Vélocité:** désigne une très haute fréquence à laquelle les données sont générées, traitées et mises en réseau.
- **Variété:** désigne une très grande variété de données qui sont soit structurées ou non structurées (textuelles, visuelles ou sonores, scientifiques ou provenant de la vie courante...).



Source : [1, 2, 3]

Les concepts de base de Data Warehouse

Différences entre Data Store, Database, Data Warehouse, Datamart et Data Lake

Une base de donnée (Database) est un type particulier de Data Store. Et un entrepot de données (Data Warehouse) est un type particulier de base de données. Un Datamart est un sous-ensemble d'un entrepôt de données mis en place pour répondre aux besoins précis d'un groupe particulier d'utilisateurs ; par exemple, les ressources humaines, et leur fournir un accès aux informations dont ils ont besoin.
Un Data Lake décrit pour sa part tout réservoir de données de grande envergure dans lequel aucune exigence de schéma et de données n'est définie avant interrogation des données (contrairement aux Data Warehouses).
Un Datamart, un Data Lake et un Data Warehouse sont donc des formes très particulières de Data Store.

Source : Tibco

Qu'est-ce que la Data Virtualization ?

Les concepts de base de Data Warehouse
Data Virtualization

- Le logiciel de Data Virtualization agit comme un pont entre des sources de données multiples et diverses, rassemblant les données essentielles à la prise de décision en un seul endroit virtuel, pour alimenter les analyses.

- La Data Virtualization fournit une couche de données moderne qui permet aux utilisateurs d'accéder à des ensembles de données, de les combiner, de les transformer et de les livrer à une vitesse et une rentabilité révolutionnaire.

- La technologie de Data Virtualization permet aux utilisateurs d'accéder rapidement aux données hébergées dans l'ensemble de l'entreprise, y compris dans les bases de données traditionnelles, les sources de big data et les systèmes cloud et IoT, pour une fraction du temps et du coût de l'entreposage physique et de l'extraction/transformation/chargement (ETL)

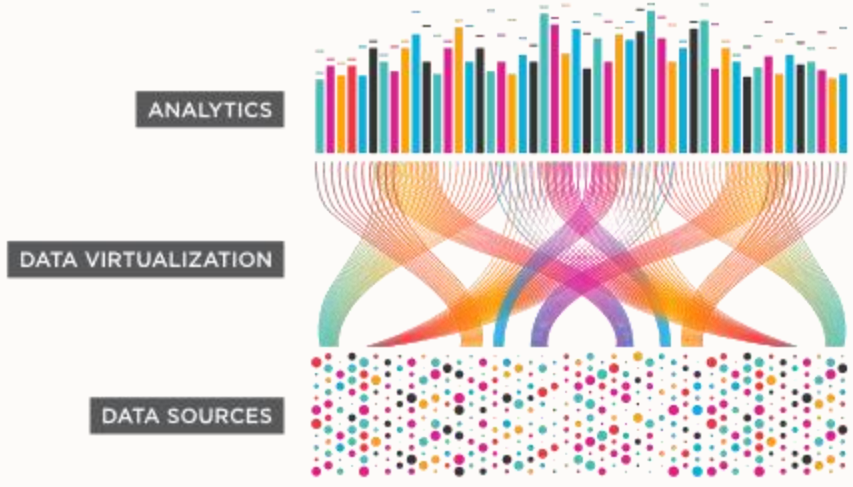
Source : Tibco

Les concepts de base de Data Warehouse
Data Virtualization

- Grace à la Data Virtualization, les utilisateurs peuvent appliquer toute une gamme d'analyses, y compris des analyses visualisées, prédictives et en continu, sur des mises à jour de données fraîches et actualisées à la minute près.

- Grace à une gouvernance et une sécurité intégrées, les utilisateurs de la Data Virtualization sont assurés de la cohérence, de l'extrême qualité et de la protection de leurs données.

- En outre, la Data Virtualization permet d'obtenir des données plus conviviales pour l'entreprise, en transformant les structures et la syntaxe informatiques en services de données faciles à comprendre, élaborés par l'informatique et faciles à trouver et à utiliser via un répertoire professionnel en libre-service.



Source : Tibco

Les concepts de base de Data Warehouse
Data Virtualization / Cloud Computing

- Le Cloud Computing (l'informatique en nuage) est une révolution technologique de cette décennie avec le Big Data. Le Big Data propose des solutions de traitement des données massives alors que le Cloud Computing offre des services de dématérialisation des ressources informatiques.

Source : [3, 17, 19, 20].

Les concepts de base de Data Warehouse
Data Virtualization / Cloud Computing

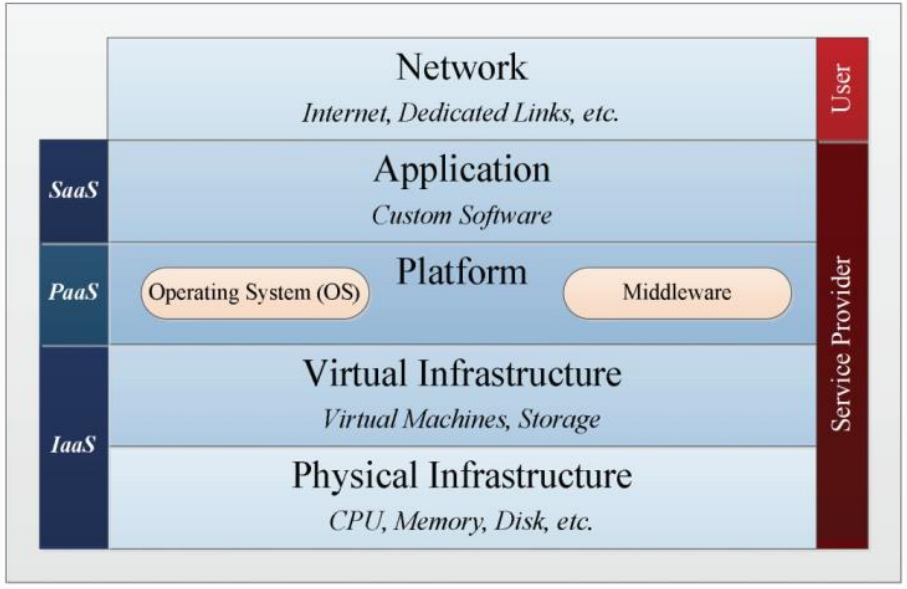
- La définition la plus acceptable du cloud computing a été introduite par l'organisme NIST (National Institute of Standards and Technology): **" Le cloud Computing est un modèle fournissant, à la demande et au travers d'un réseau, un ensemble partagé de ressources informatiques incluant des serveurs, des espaces de stockage, des applications, des traitements et des plates-formes de déploiement qui peuvent être rapidement mises en service avec un effort minimum de gestion et d'interaction avec le fournisseur de ce service ".**

- Une autre définition: **" Le Cloud Computing est un modèle dans lequel les ressources telles que la puissance de traitement, le stockage, la connectivité et le partage, etc. sont proposées sous forme de services par un mécanisme d'accès à distance. Il présente plusieurs caractéristiques souhaitables telles que la distribution et l'élasticité rapide, la sécurité, les self-services à la demande, l'accès omniprésent au réseau et la mise en commun des ressources ".**

Source : [3, 17, 19, 20].

Les concepts de base de Data Warehouse
Data Virtualization / Cloud Computing

- **L'architecture des environnements de Cloud Computing** est composée de cinq grandes couches : Infrastructure physique, infrastructure virtuelle, plateforme, application et réseau. En fonction de ces couches, trois modèles de services de cloud computing ont été définis pour être fournis aux utilisateurs : Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) et Software-as-a-Service (SaaS).



Source : [3, 17, 19, 20].

Environnement simple d'un Data Warehouse

Les concepts de base de Data Warehouse
Environnement simple d'un Data Warehouse

- Pour comprendre comment les différentes pièces d'un data warehouse s'emboîtent, nous allons jeter un coup d'œil à un environnement simple end to end d'un data warehouse.

- Un **data warehouse** est construit en extrayant des **données** d'autres applications et systèmes. Nous identifions nos sources de données ainsi que notre data warehouse. Entre les deux, nous avons un aspect essentiel appelé **ETL.** Ce dérnier signifie extraction, transformation et chargement.

Source : Alan Simon

Les concepts de base de Data Warehouse

Environnement simple d'un Data Warehouse

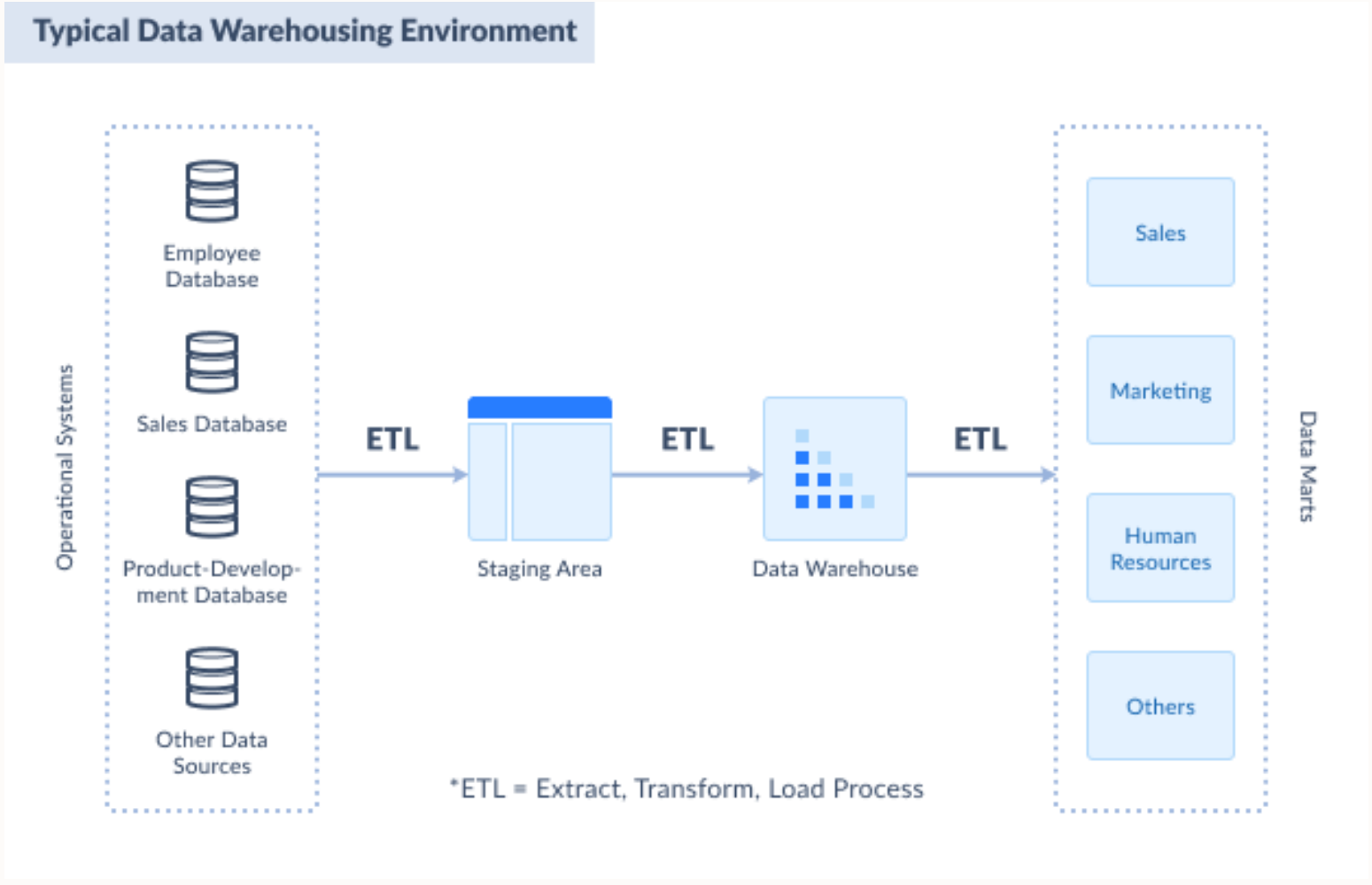
- Parfois, nous ne nous arrêtons pas à extraire et à copier des données de nos sources de données dans un data warehouse via ETL. Parfois, nous continuerons ensuite à regrouper et à copier à nouveau les données, en les envoyant en aval dans des environnements plus petits, généralement appelés Data marts .

- Un data mart est un sous-ensemble d'un entrepôt de données généralement utilisé pour accéder aux informations destinées aux clients. Il s'agit d'une structure spécifique aux paramètres d'entreposage de données. Ainsi, il est généralement axé sur un secteur d'activité ou une équipe et tire des informations d'une seule source particulière.

- Contrairement à la mise en œuvre d'un entrepôt de données d'entreprise qui peut s'étendre sur plusieurs mois, voire plusieurs années, un magasin de données est généralement mis en œuvre en quelques mois, offrant une assistance rapide.

Source : Alan Simon

Les concepts de base de Data Warehouse



Source : skyvia

FIN DE SEANCE