RER - Data Warehouse ETL & OLAP

# Contexte :

Découvrir comment stocker les données dans un data Warehouse, les importer, les trier et les transformer.

Se familiariser avec ETL, OLAP et les Data Warehouse.

# Mots clés :

* ETL :
  + Extraire : identifier et extraire des données depuis des sources hétérogènes
  + Transformer : nettoyer puis standardiser les données extraites (déf format de sauvegarde données, suppression doublons, etc...) afin de les rendre exploitable
  + Charger : transmettre ces données exploitables dans un datawarehouse pour que celle-ci utiliser
* OLTP :classe de logiciels capables de prendre en charge les applications orientées transactions sur Internet.

Généralement, les systèmes OLTP sont utilisés pour la saisie des commandes, les transactions financières, la gestion de la relation client (CRM) et les ventes au détail. Ces systèmes comptent un grand nombre d'utilisateurs qui effectuent des transactions courtes.

* OLAP : OLAP est un acronyme pour Online Analytical Processing. OLAP effectue une analyse multidimensionnelle des données d'entreprise et offre la possibilité de calculs complexes, d'analyses de tendances et de modélisation de données sophistiquées. OLAP permet aux utilisateurs finaux d'effectuer des analyses de données dans de multiples dimensions, leur fournissant ainsi les informations et la compréhension dont ils ont besoin pour prendre de meilleures décisions.
  + Cube: Un cube OLAP est un tableau multidimensionnel de données.
  + Slice : sélectionne une valeur unique pour l'une de ses dimensions et construit un sous-ensemble du cube.
  + Dice : sélectionne des valeurs spécifiques de plusieurs dimensions pour produire un nouveau sous-cube.
* MDX :
* Data Warehouse : une base de données relationnelle hébergée sur un serveur dans un Data Center ou dans le Cloud. Il recueille des données de sources variées et hétérogènes dans le but principal de soutenir l'analyse et faciliter le processus de prise de décision. En matière d’intégration dans le système de données existant, le fonctionnement du Data Warehouse est basé sur le processus ETL (Extract, Tranform, Load) permettant de charger les données issues des différentes applications.
* Data flow :
* Data Mart : sous-ensemble d’un data warehouse
* CRM : base de données c entralisant toutes les informations importantes concernant vos clients, prospects et autres contacts,
* Système d’information dans les entreprises : Le système d’information (SI) est un élément central d’une entreprise ou d’une organisation. Il permet aux différents acteurs de véhiculer des informations et de communiquer grâce à un ensemble de ressources matérielles, humaines et logicielles. Un SI permet de créer, collecter, stocker, traiter, modifier des informations sous divers formats.

L’objectif d’un SI est de restituer une information à la bonne personne et au bon moment sous le format approprié.

* ERP : progiciel qui permet de gérer l’ensemble des processus opérationnels d’une entreprise en intégrant plusieurs fonctions de gestion : solution de gestion des commandes, solution de gestion des stocks, solution de gestion de la paie et de la comptabilité, solution de gestion e-commerce, solution de gestion de commerce BtoB ouBtoC ... dans un système.
* SCM (supply chain management) :
* PowerBI :Power BI est une plateforme évolutive, unifiée pour la business intelligence (BI) en entreprise et en libre-service. Connectez-vous à toutes les données, visualisez-les et insérez les visuels de manière transparente dans les applications que vous utilisez au quotidien.
* Mesure :
* Dimension :
* Granularité des données :
* Dataverse :

# Problématique(s) :

1. Comment utiliser ETL pour stocker les données dans un Data Warehouse
2. En quoi la méthode OLAP permet d’analyser les données dans un Data Warehouse
3. Comment utiliser PowerBI
4. Comment utiliser Python pour le BI

# Hypothèses :

1. Il est difficile d’extraire des données quand elles ne sont pas issues de la même source (Adeline)
2. La granularité des données fait référence aux données éparses (nombreuses et dispersées) (Jean Paul)
3. Il est obligé de passer par la méthode ETL pour constituer un Data Warehouse (Loïc)
4. Un cube de données est un moyen de renvoyer plus rapidement une requête qu’une base de données relationnelle (Briand)
5. On ne peut pas utiliser OLAP avec des données bruts (Tetyana)
6. Avant de créer un flux de données il est important de comprendre comment on va utiliser la donnée. (Adeline 2)
7. PowerBI nous permet de gérer un Data Warehouse et d’utiliser OLAP. (Adrien)
8. La granularité des données est en rapport avec la mesure et la dimension. (Aude)
9. Le Data Mining peut être un outil stratégique. (Seydou)
10. Quarante pourcents du BI correspond à de la visualisation de données. (Étienne)
11. ETL et ELT sont similaires. (Tetyana)
12. OLAP ne sert qu’à visualiser les données. (Nicolas)
13. ETL ne peut s’appliquer qu’a des données structurées (Axel)

FAUX, ETL permet de transformer des données de source hétérogènes en données structurées pour être stocké dans un data warehouse

1. Mesure et dimension sont équivalent. (Solenn)

# Plan d'action :

* Explorer les ressources
* Définir les mots-clefs
* Répondre aux hypothèses
* Faire la comparaison : ETL vs CRUD
* Faire la comparaison : OLTP vs OLAP
* Faire la comparaison : OLAP vs Data mining
* Faire la comparaison : Mesure vs dimension
* Répondre aux problématiques
* Faire les workshops (dans l’ordre suivant)
  1. ETL et TCD (tableau croisé dynamique)
  2. PowerBI
  3. Python ETL
  4. OLAP
  5. MySQL
* Restitution du RER
* Livrables