

Projet Business Intelligence

Réalisé par:

- Walid TAOUTAOU
- Brahim OULHAJ



Encadré par:

- Prof. Jamal RIFFI

Plan

- Introduction et problématique
- Analyse des données
- Data Integration: E.T.L
- Création du Cube OLAP
- Reporting
- Data Mining

Introduction et Problématique

Le processus décisionnel est un projet qui se construit, il est né d'un besoin exprimé par les entreprises à cause du volume important des données à manipuler et à analyser. En entreposant les données, le processus décisionnel apporte la solution au problème de la croissance continue des données provenant de différentes sources et de différents formats en les homogénéisant et en les organisant dans un Data-Warehouse où elles sont historiées, résumées et consolidées.

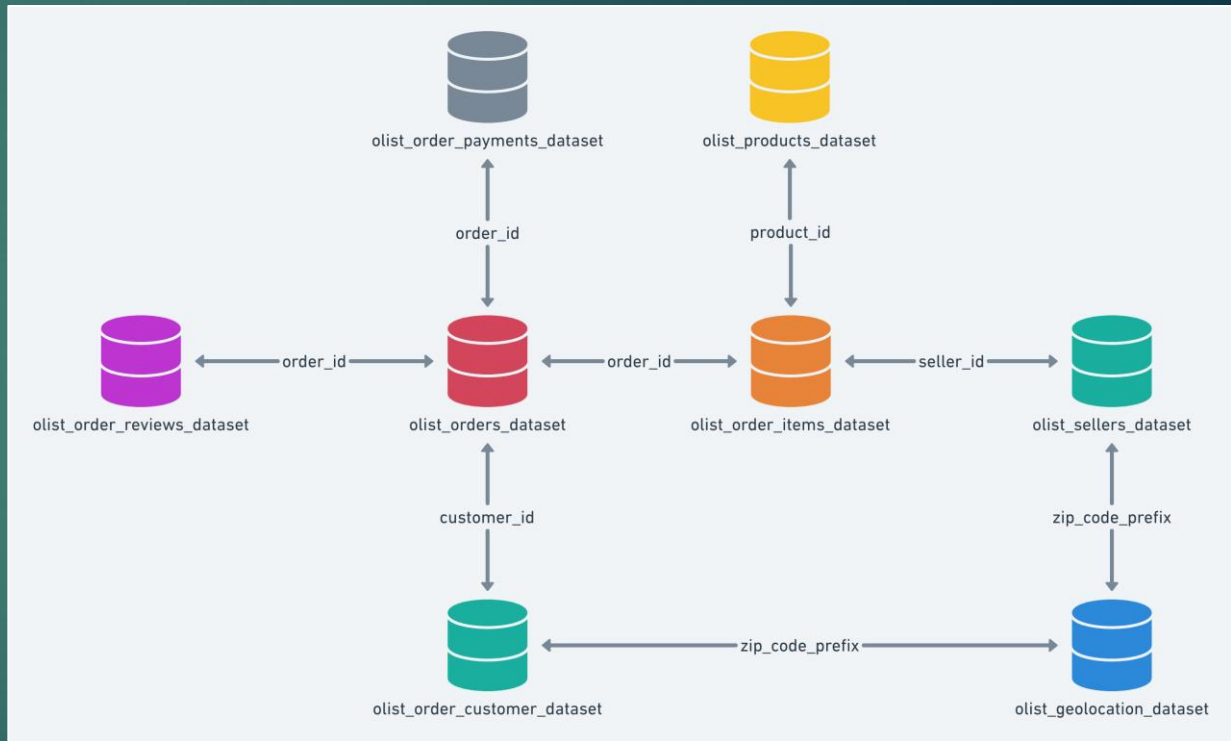
Le volume de données des entrepôts est important et va de centaines de giga-octets à des téraoctets, voir même encore davantage de nos jours.

Notre projet a été réalisé pour faire des statistiques pour un magasin brésilien en ligne, www.olist.com, dont le but d'améliorer le nombre de ventes et d'augmenter le chiffre d'affaire de ce magasin.



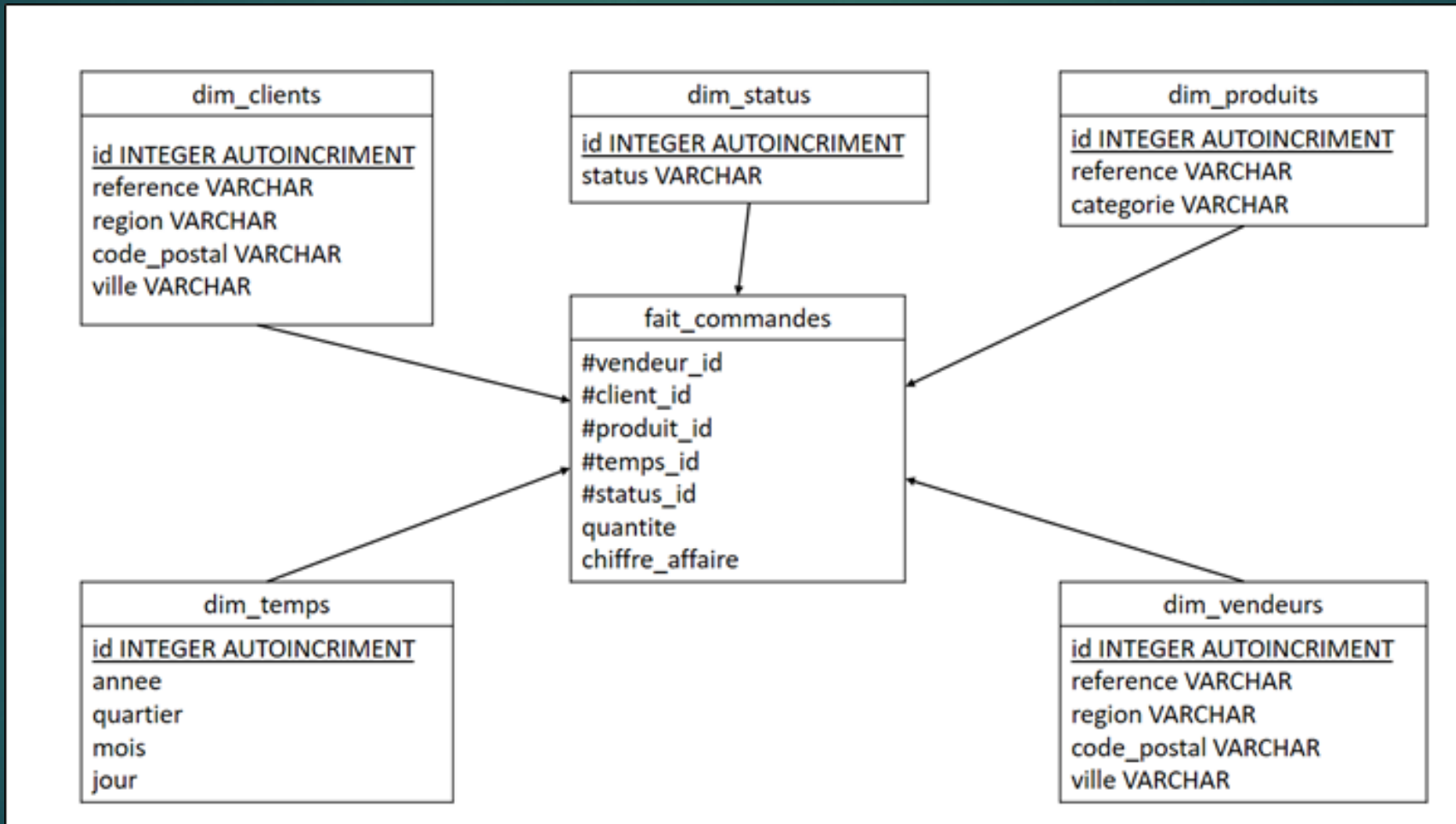
Analyse des données

Il s'agit d'un ensemble de données publiques brésiliennes sur le e-commerce des commandes effectuées à Olist Store. L'ensemble de données contient des informations, en format CSV, sur les 100 k commandes de 2016 à 2018 effectuées sur de multiples marchés au Brésil. Ses caractéristiques permettent de visualiser un ordre de multiples dimensions : de l'état de l'ordre, le prix, les performances de paiement et de fret à l'emplacement du client, les attributs du produit et enfin les commentaires écrits par les clients. Ainsi, un ensemble de données de géolocalisation qui relie les codes postaux brésiliens aux coordonnées lat/lng.



https://www.kaggle.com/olistbr/brazilian-ecommerce#olist_order_items_dataset.csv

Analyse des données



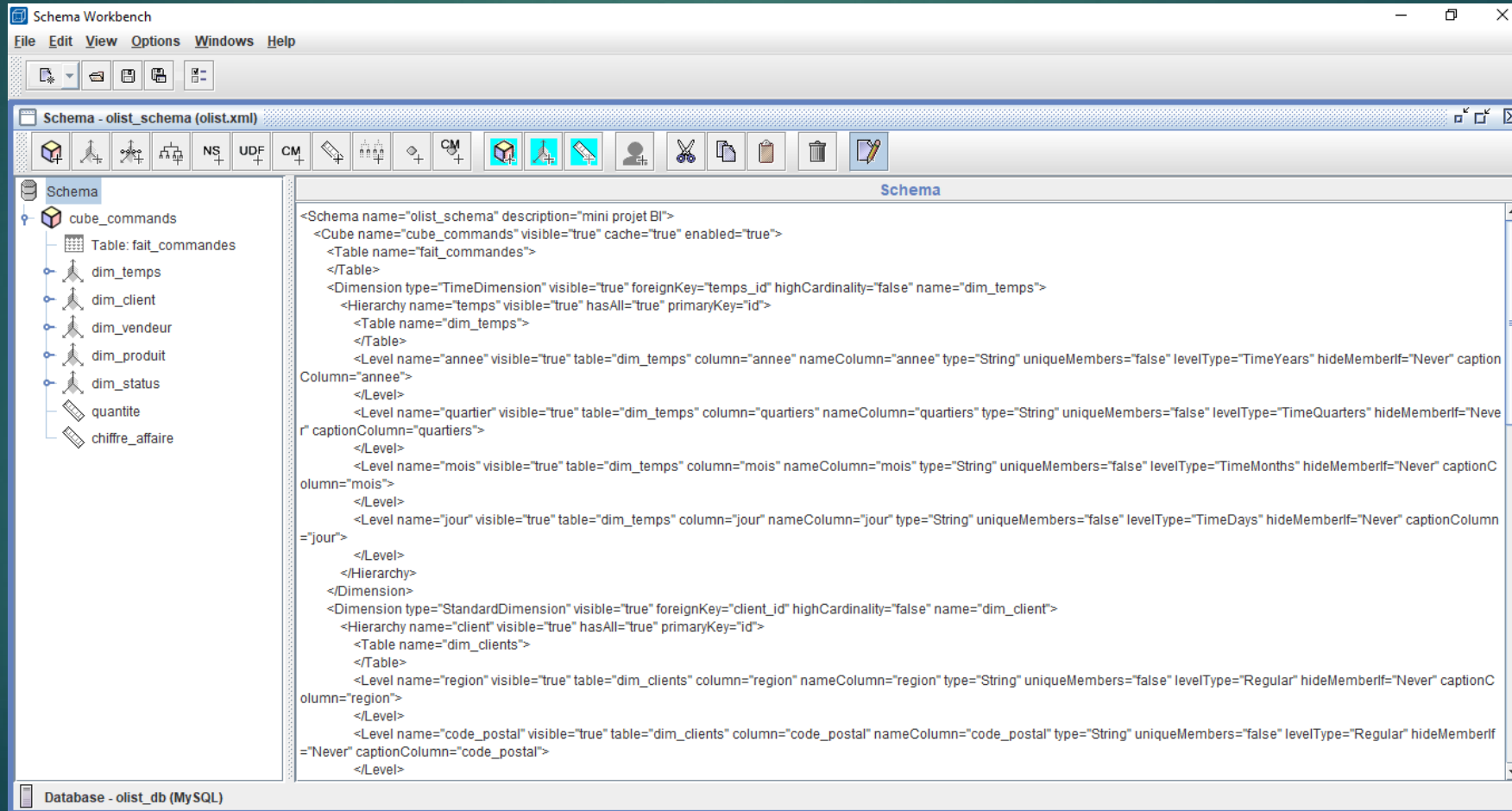
Data Integration: ETL

Talend Open Studio for Data Integration is an **open** source data integration product developed by **Talend** and designed to combine, convert and update data in various locations across a business.



Cube OLAP

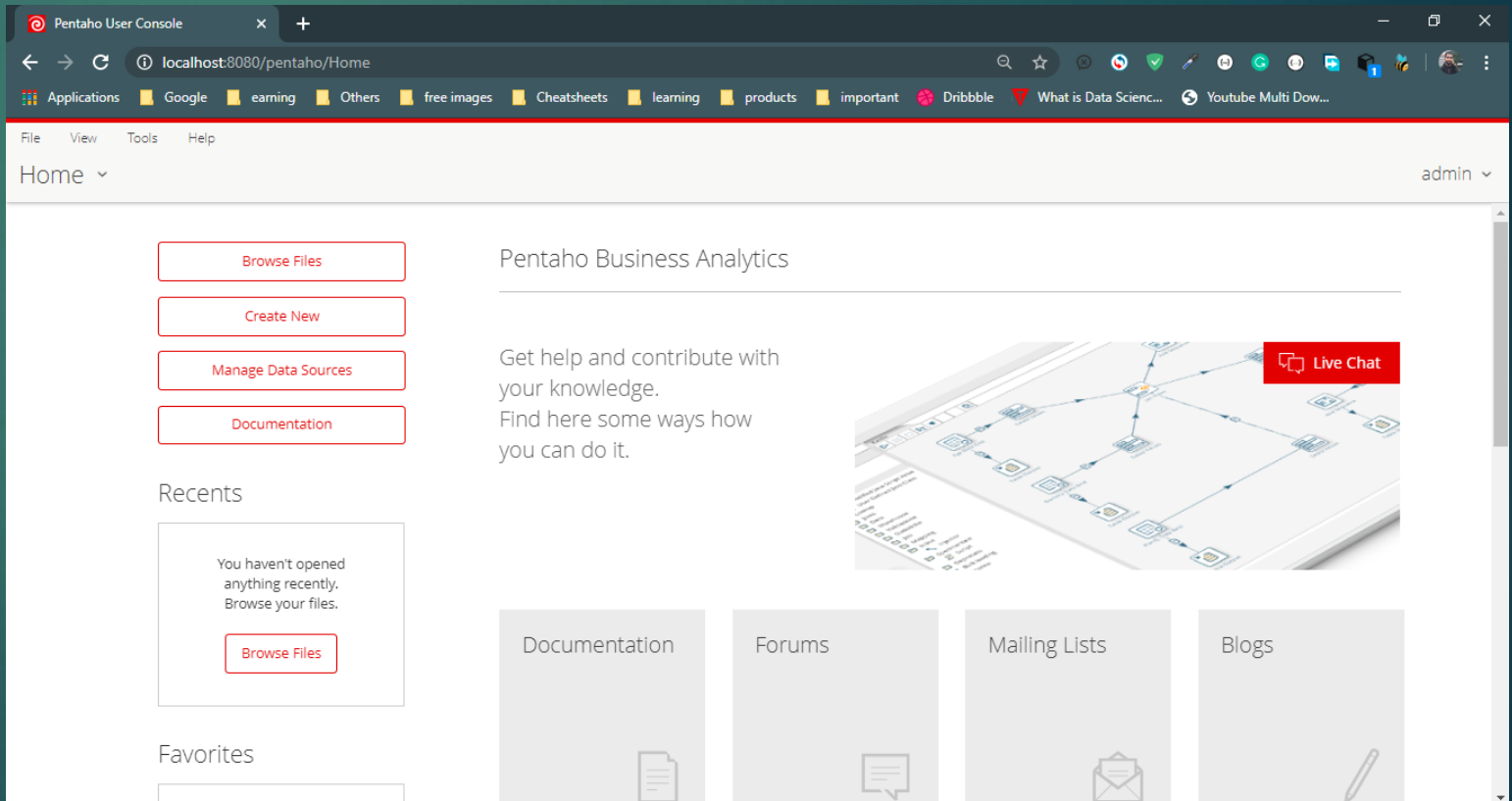
Pentaho Schema Workbench provides a graphical interface for designing OLAP cubes for Pentaho Analysis (Mondrian). The schema created is stored as a regular XML file on disk.



Reporting

Pentaho Server is software that provides reporting.

Pivot4J provides a common API for OLAP servers which can be used to build an analytical service frontend with pivot style GUI. Pivot4J Analytics is a full fledged OLAP client application built with the Pivot4J core API, which can be used as a drop-in replacement for legacy the JPivot web application.



Data Mining

Le *Data Mining* est en fait un terme générique englobant toute une famille d'outils facilitant l'exploration et l'analyse des données contenues au sein d'une base décisionnelle de type Data Warehouse ou DataMart. Les techniques mises en action lors de l'utilisation de cet instrument d'analyse et de prospection sont particulièrement efficaces pour extraire des informations significatives depuis de grandes quantités de données.

La Question qu'on va répondre dans cette partie et la suivante:

Est-ce qu'il est possible, pour un vendeur dans **une région**, de vendre un produit appartient à tel **catégorie** à un client dans une autre **région** dans un **mois** précis?

