

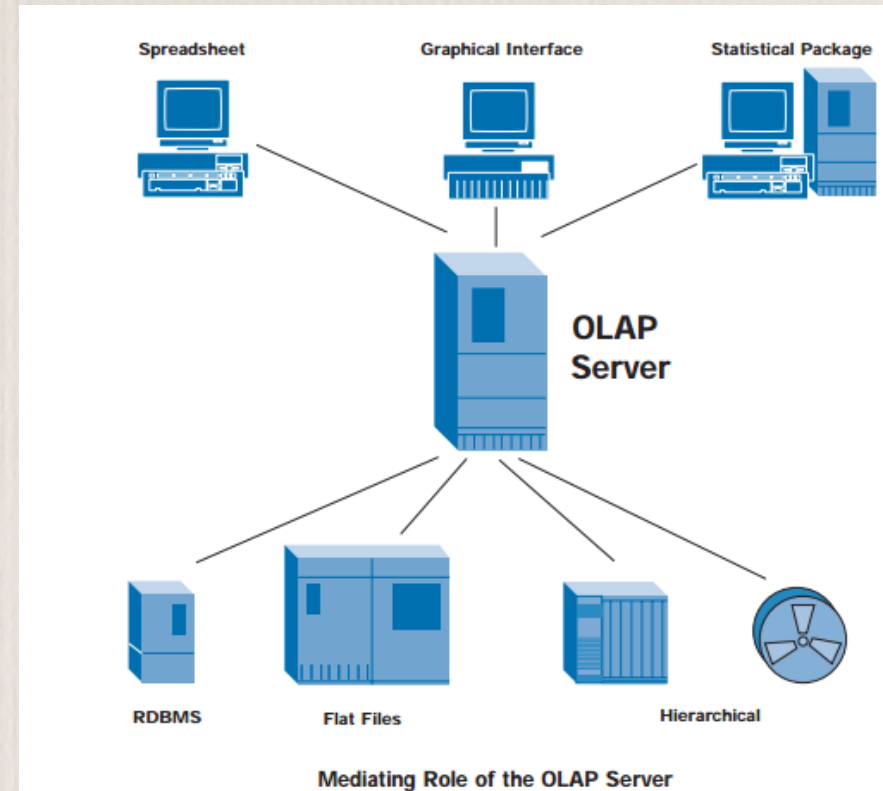
Le mode d'emploi des outils de DataViz

1. Se connecter aux données en utilisant un des nombreux connecteurs prédéfinis (SQL, Excel, Hadoop, Cloud, ...)
2. Décrire les traitements à effectuer: conversion, consolidation et regroupements
3. Choisir parmi les graphiques proposés, le plus adapté

Quelques outils: Excel, Information Builder, Qlik, Tableau, TIBCO Spotfire, SAP, Power BI, SalesForce, Zoomdata, Targit, MyReport, Plotly, ...

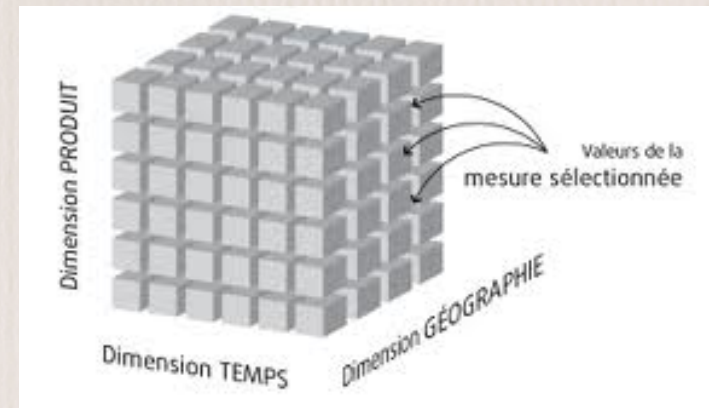
1- La technologie clé: OLAP OnLine Analytical Processing

- Technologie alternative à la BD relationnelle
- Analyse de données organisées en axes structurés
- A l'origine pour le Business Intelligence et la Finance



1- Qu'est ce qu'un hyper-cube de données ?

- dimensions hiérarchisées fixes
 - Géographie: Continent>Pays>Région
 - Temps: année>trimestre>mois
 - Produits: Gamme>Type>Famille
- valeurs élémentaires au plus bas
- règles de consolidation qui « calculent » les niveaux intermédiaires (nombre de, somme, moyenne, min, max, ...)



Bénéfices: choix des axes, filtres, drill-down, ... en temps réel

1-Données <> Format de la donnée

Domestique (à la SQL)

- noms de colonnes représentant le réceptacle des dimensions (et/ou des attributs)
- autant de lignes que de données
- organisé en plusieurs tables qui se référencent mutuellement

Sauvage (quelques exemples)

- horizontal au lieu de vertical: un suivi de projet avec les mois en colonnes, une météo horizontale en jours de la semaine
- une suite d'événement daté
- des données non périodiques (inspection aléatoire, ...)
- ...

2- Les traitements à faire sur les données

Un petit langage intégré pour:

- calculer l'âge à partir de la date de naissance
- format numérique US à convertir en français
- programmer les règles de consolidation
- définir les indicateurs proportionnels (masse salariale/ nb personne)
- csv US -> csv F
- ...

Les limites:

- les calculs complexes (doivent plutôt être fait en dehors de l'outil)
- les calculs longs en temps handicapent l'UX
- les traitements sur la chronologie temporelle sont quasi impossible à exprimer
- ...

**Le mieux c'est quand tout ce qui est nécessaire est déjà dans les données
.... et qu'il n'y a pas à faire de traitement !!**

3- Choisir le graphique adapté

Choisir le diagramme

Désigner les axes à projeter en 1, 2 (X, Y) ou >3 dimensions (taille, couleur, ...)

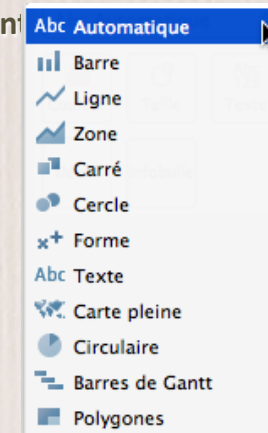
Créer plusieurs diagrammes liés par les données

Filtrer les données sur un diagramme se propage sur les autres

Drill down

Les diagrammes les plus courants

- points et marqueurs
- tableau
- barres (H ou V)
- lignes
- aires
- camembert / donuts
- cartes
- jauges

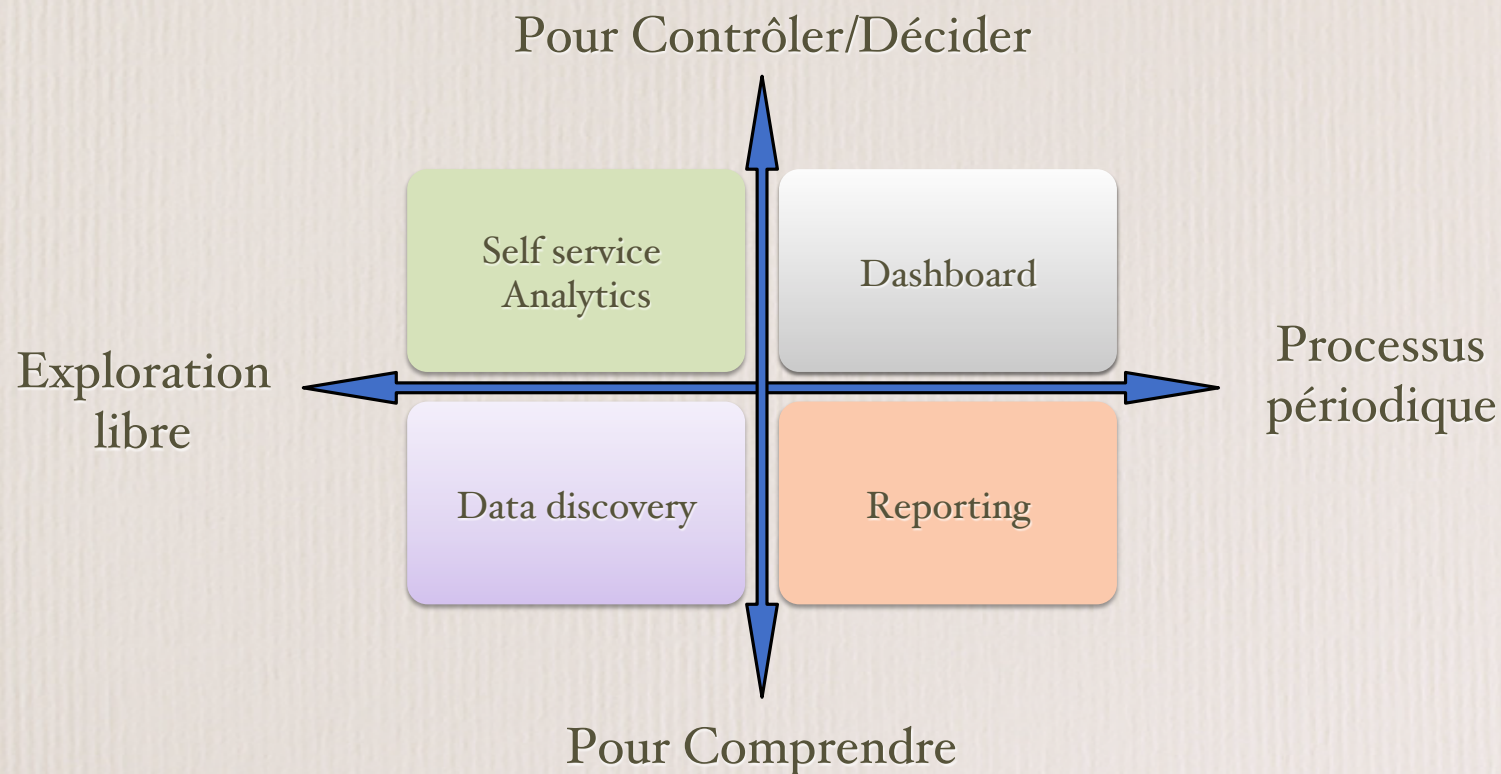


Les attributs les plus courants

- taille, épaisseur
- couleur
- bulle

**L'outil est souvent capable de proposer le graphique qui «convient» aux données
... mais les données sont-elles les «bonnes» ?**

La dataviz a de nombreux objectifs



Et ne faudrait-il pas commencer par énoncer l'objectif ?