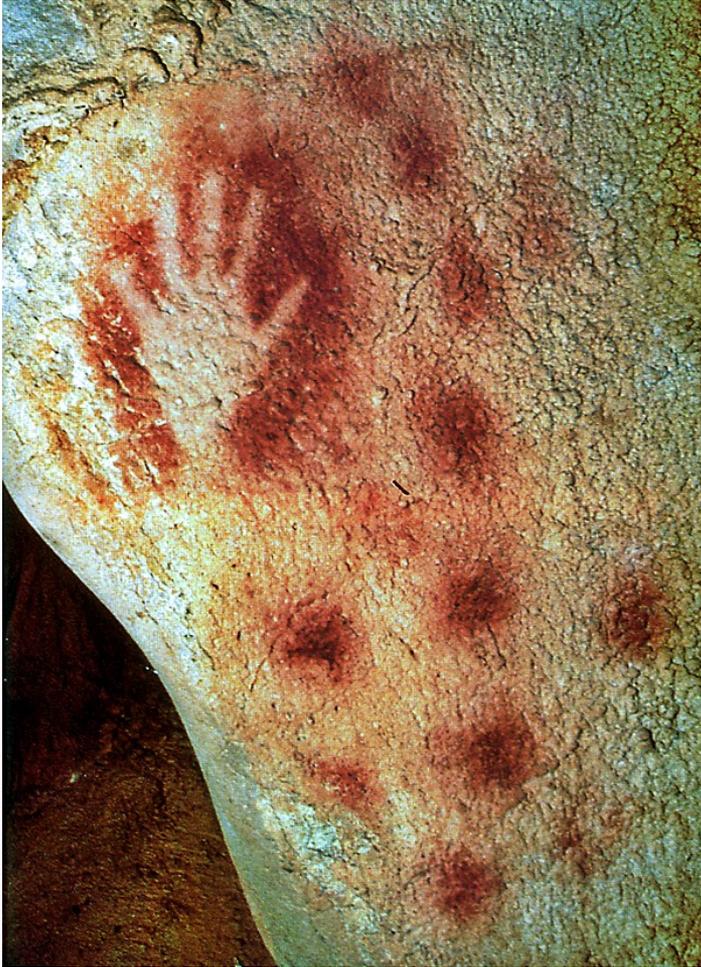


# Cours d'Informatique Décisionnelle

Version de mars 2014

# L'âge de pierre : gestion de l'information



- Dessin et peinture sur la roche

- main négative préhistorique de la Grotte du Pech Merle (Lot, France).

# L'âge de pierre

- **Tout au début de la civilisation, l'homme préhistorique :**
  - Commence à s'adapter à son environnement,
  - Il apprend la façon de tirer profit des ressources qui sont à sa disposition,
  - D'un autre côté, il grave (stocke) l'information sur la pierre (la roche), art, gravures ou peintures rupestres, dans grottes, ou obélisque, totems, etc.

# L'âge du bronze (métal)



image source [http://commons.wikimedia.org/wiki/File:Bois-du-Luc\\_-\\_Fondeurs\\_d%27art.JPG](http://commons.wikimedia.org/wiki/File:Bois-du-Luc_-_Fondeurs_d%27art.JPG) licenced under the Creative Commons Attribution 3.0 Unported license. Attribution : Jean-Pol GRANDMONT

# L'âge du métal

- **A l'âge de bronze, l'homme commence :**
  - à savoir comment transformer la roche en métal,
  - Puis à créer des outils à partir du métal,
  - De nouvelles techniques apparaissent,
  - De nouveaux procédés apparaissent.

# L'ère industrielle

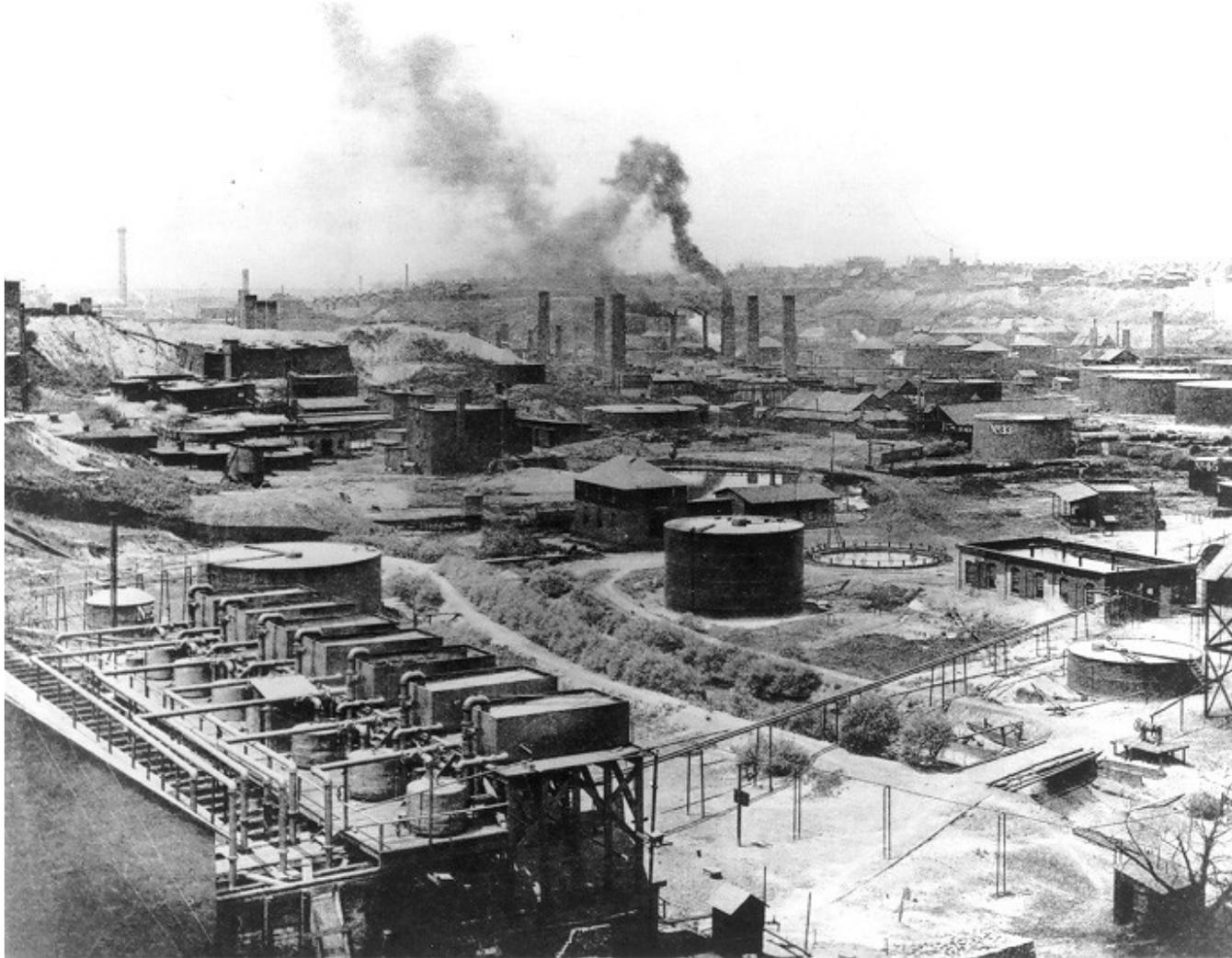


Image source : [http://commons.wikimedia.org/wiki/File:Standard\\_Oil.jpg](http://commons.wikimedia.org/wiki/File:Standard_Oil.jpg)

# L'ère industrielle

- **L'homme :**

- généralise l'approche utilisée pour créer des outils en métal et en bois, des outils de sa propre création :
- Il commence à créer des machines complexes,
  - Début vers 1600, apparition de la machine à vapeur, le train, l'automobile, etc.
  - Les procédés de création de machines sont de plus en plus nombreux, comment les documenter ?

# Besoin de stocker et d'archiver l'information

- **Avec le développement de l'industrie :**
  - Le besoin de stocker et d'archiver les données est créé, afin de :
  - Avoir un moyen de contrôler l'activité des entreprises en général,
  - Garder une trace écrite et,
  - Pouvoir réitérer les mêmes processus organisationnels en question, autant que nécessaire dans le temps,

# Stocker « papier » de l'information



image source [http://en.wikipedia.org/wiki/File:Fondos\\_archivo.jpg](http://en.wikipedia.org/wiki/File:Fondos_archivo.jpg) licenced under the Creative Commons Attribution-Share Alike 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license. Attribution : Archivo-FSP

# Dépôts d'information

- Le premier support de stockage d'information est le papier (papyrus, etc.), il en découle plusieurs dépôts d'information :
- Centre d'archives d'industries, d'entreprises,
  - Registres des entrées/sorties,
  - Historiques des achats/ventes,
- Livres, manuscrits, bibliothèques (privées ou publiques)
- Processus organisationnels et données non documentées (à documenter).

# Informatique et nouvelles possibilités

- Avec l'apparition des premiers ordinateurs en 1930, permettant de stocker et de traiter des données ,
- Il est devenu possible d'envisager le stockage de données sur un support autre que le support PAPIER,
- La généralisation de l'approche mène à la création des centres de données (serveurs).

# Data Center (centre de données)



image source [http://commons.wikimedia.org/wiki/File:BalticServers\\_data\\_center.jpg](http://commons.wikimedia.org/wiki/File:BalticServers_data_center.jpg) Licenced under the Creative Commons Attribution-Share Alike 3.0 Unported licence. Attribution : Fleshas

# Data Center (centre de données)

- **Capacité de stockage des centres de données actuels (2013) :**
  - Par exemple, le centre de données de l'UTAH (centre américain national de cyber sécurité )
  - 1 EB = 1000000000000000000000000B =  $10^{18}$ bytes = 1000 petabytes = 1 billion gigabytes (1 milliard de giga octets = 1 milliard \*  $1000^3$  \* octets).

# Data Center (centre de données)

## Utah Data Center

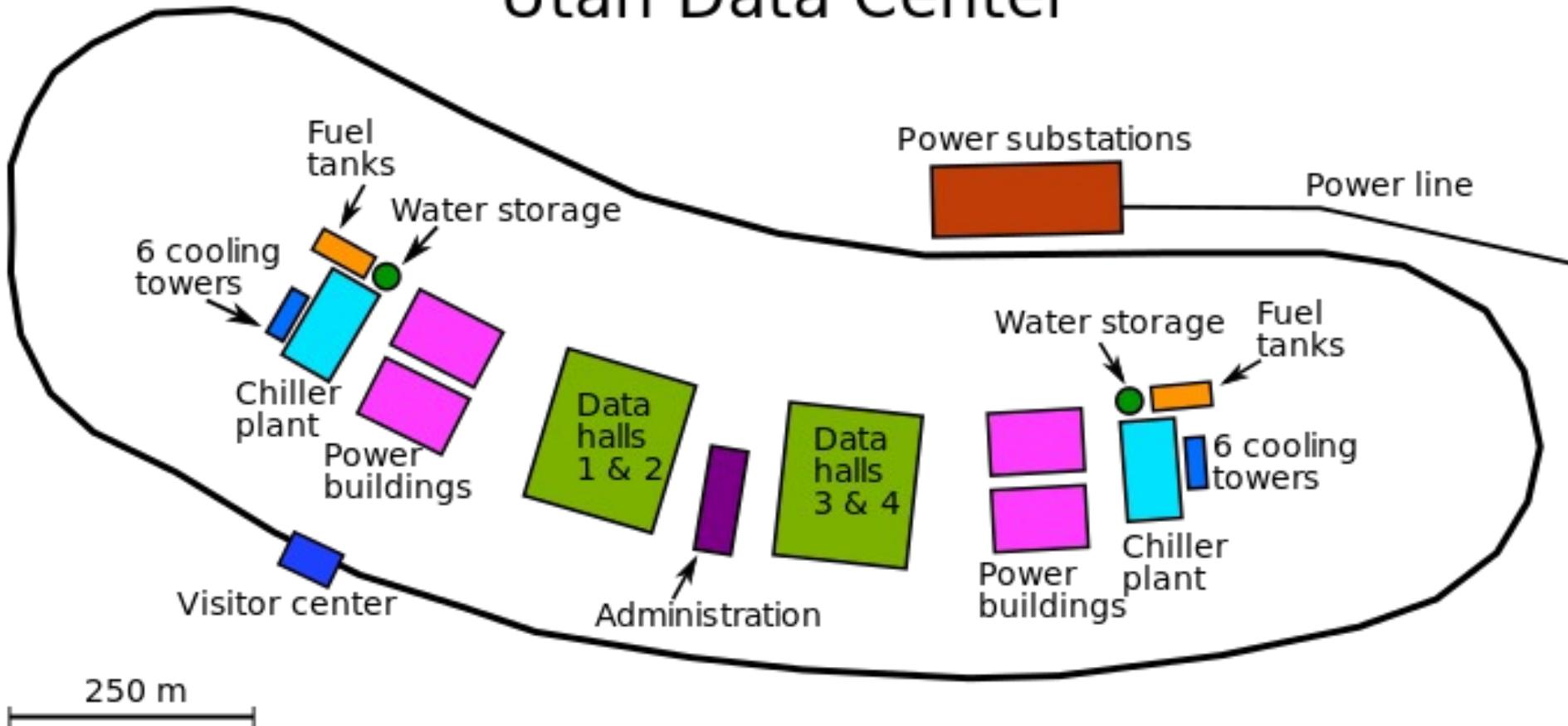


Image source :

[http://en.wikipedia.org/wiki/File:Utah\\_Data\\_Center\\_of\\_the\\_NSA\\_in\\_Bluffdale\\_Utah\\_vector.svg](http://en.wikipedia.org/wiki/File:Utah_Data_Center_of_the_NSA_in_Bluffdale_Utah_vector.svg)

# Nouvelle problématique : Information Overload (surcharge d'information)

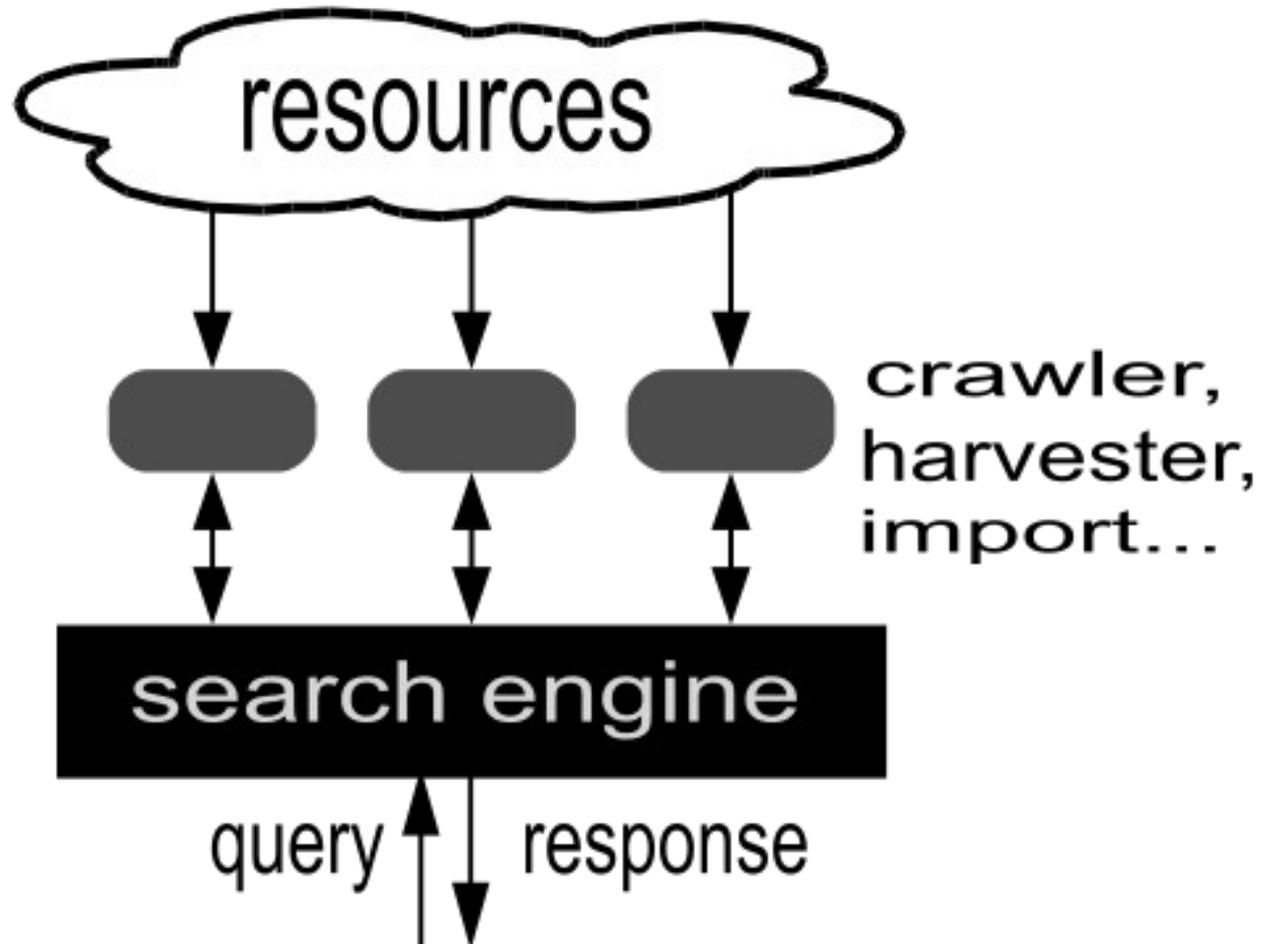


image source <http://commons.wikimedia.org/wiki/File:India-Truck-Overload.jpg> licenced under under the Creative Commons Attribution-Share Alike 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license. Attribution : Peter Krimbacher, Moebius1

# Recherche d'information

- Nouvelle problématique : comment accéder rapidement à l'information ?
- Information à propos d'un sujet donné,
- Solution à un problème donné,
- Décision à prendre,
- Plan d'actions à suivre.

# Principe des moteurs de recherche



# Limites des moteurs de recherche (locaux et Web)

- **Les résultats de recherche sont :**
  - Affichés sous forme de listing,
  - Les résultats sont très nombreux,
  - Et il n'y a pas de conclusion ou de corrélation générale entre ces résultats,
  - La surcharge d'information reste valable,
  - Il y a un nouveau besoin en ce qui concerne l'accès à l'information utile, pertinente.

# Les Ages de l'homme

- Age de pierre,
- Age du bronze,
- Age industriel,
- Age de l'information,
  
- Age de l'intelligence :
  - Intelligence économique

# Intelligence et Intelligence

- Il est nécessaire de différencier entre :
- Intelligence (en anglais), et
- Business intelligence (BI),
  
- La première désigne le degré d'implication + ou - intrusive d'un organisme dans ce qu'on appelle l'espionnage industriel.

# Intelligence d'affaires (économique)

- La deuxième désigne plutôt la capacité d'une entreprise à être en parfait contrôle des données et informations relatives à son propre domaine d'affaires (business).
- ensemble d'actions et de réactions, basées sur l'utilisation « intelligente » de l'information pour des fins à utilité économiques.

# Définition de l'intelligence économique

- ensemble de théories, méthodologies, procédures (fonctionnelles et techniques), processus, architectures et technologies .
- collectent, consolident, modélisent, transforment, et restituent des collections de données brutes
- en informations pertinentes et utiles à destination des acteurs « métier » pour leur permettre d'agir, à des fins économiques et stratégiques.

# Définition de l'intelligence économique

- traiter de grandes quantités d'informations (du passé).
- pour aider à identifier et développer de nouvelles opportunités (pour le futur).
- En conséquence, l'usage de ces nouvelles opportunités et mettre en œuvre une stratégie efficace peut fournir un avantage concurrentiel et la stabilité de l'activité ou entreprise à long terme.

# Bases de l'informatique décisionnelle

# Informatique décisionnelle

- L'informatique décisionnelle :
  - en anglais : DSS pour Decision Support System
  - ou encore BI pour Business Intelligence
- désigne les moyens, les outils et les méthodes et les techniques
- qui permettent d'offrir une aide à la décision et
- de permettre aux responsables de la stratégie d'un organisme (économique, administratif, ou politique),
- d'avoir une vue d'ensemble de l'activité traitée, et d'élaborer un processus de prise ou d'aide à la décision.

# Exemple de système de support à l'aide à la décision (géographique)

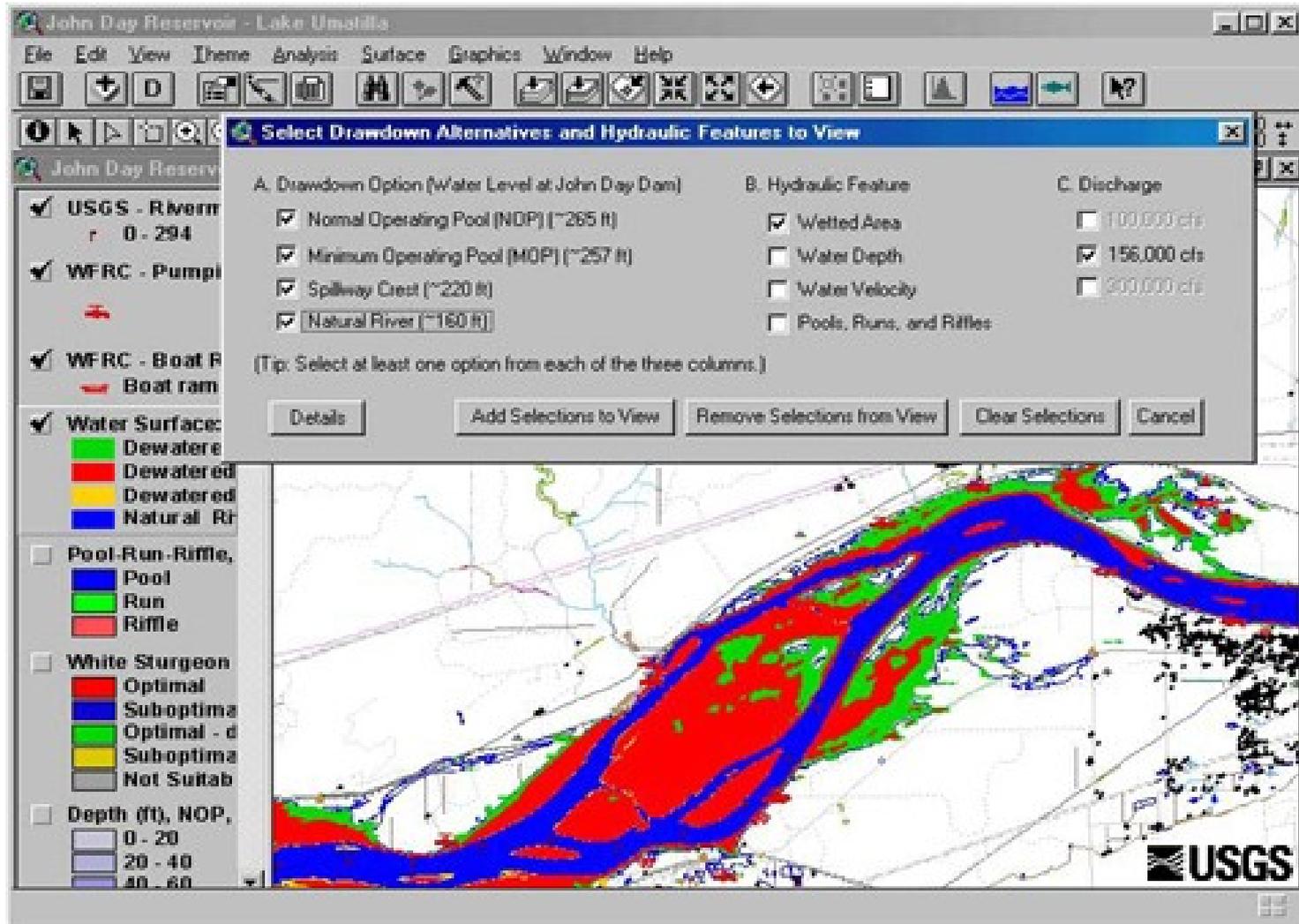


Image source :

[http://en.wikipedia.org/wiki/File:Decision\\_Support\\_System\\_for\\_John\\_Day\\_Reservoir.jpg](http://en.wikipedia.org/wiki/File:Decision_Support_System_for_John_Day_Reservoir.jpg)

# Prise de décision

- La prise de décision peut être considérée comme le processus cognitif ,
- qui entraîne la sélection d'un plan d'action parmi plusieurs scénarios alternatifs.
- Chaque processus décisionnel produit un choix final. La sortie/output peut être une action ou une opinion, ou un choix.

# Exemple de prise de décision

## Adding an article to Wikipedia

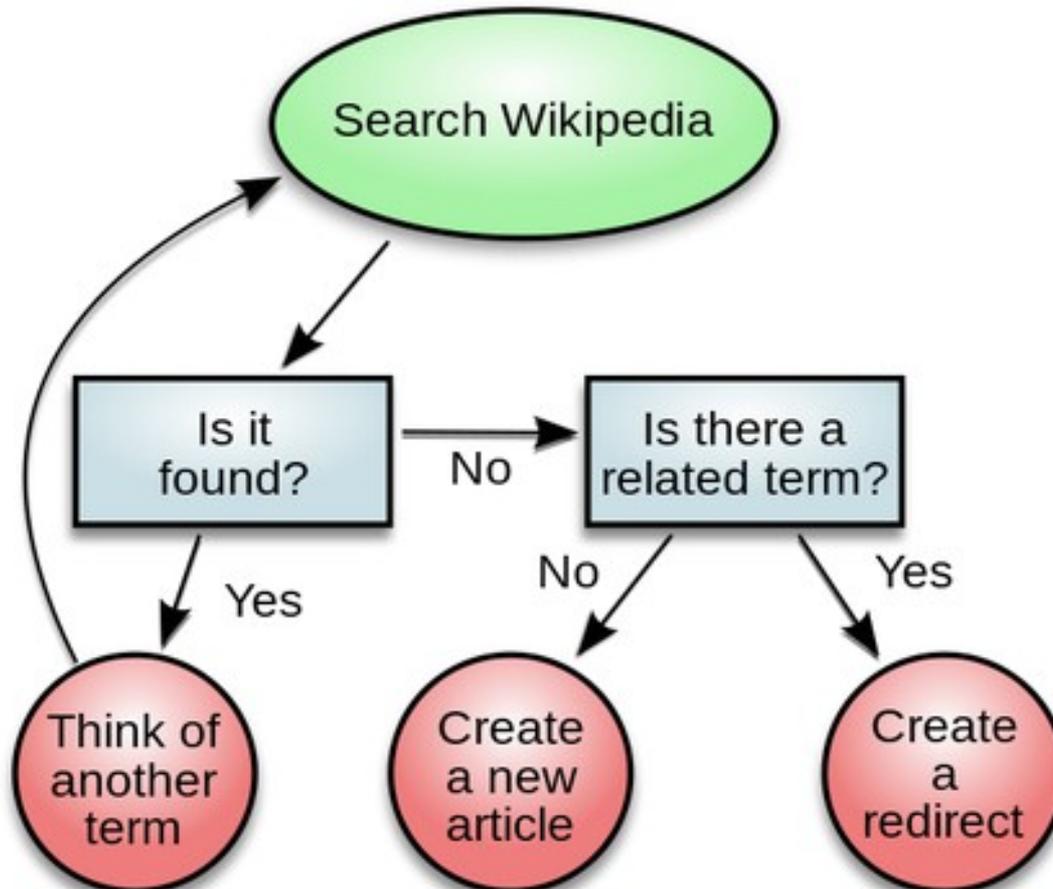


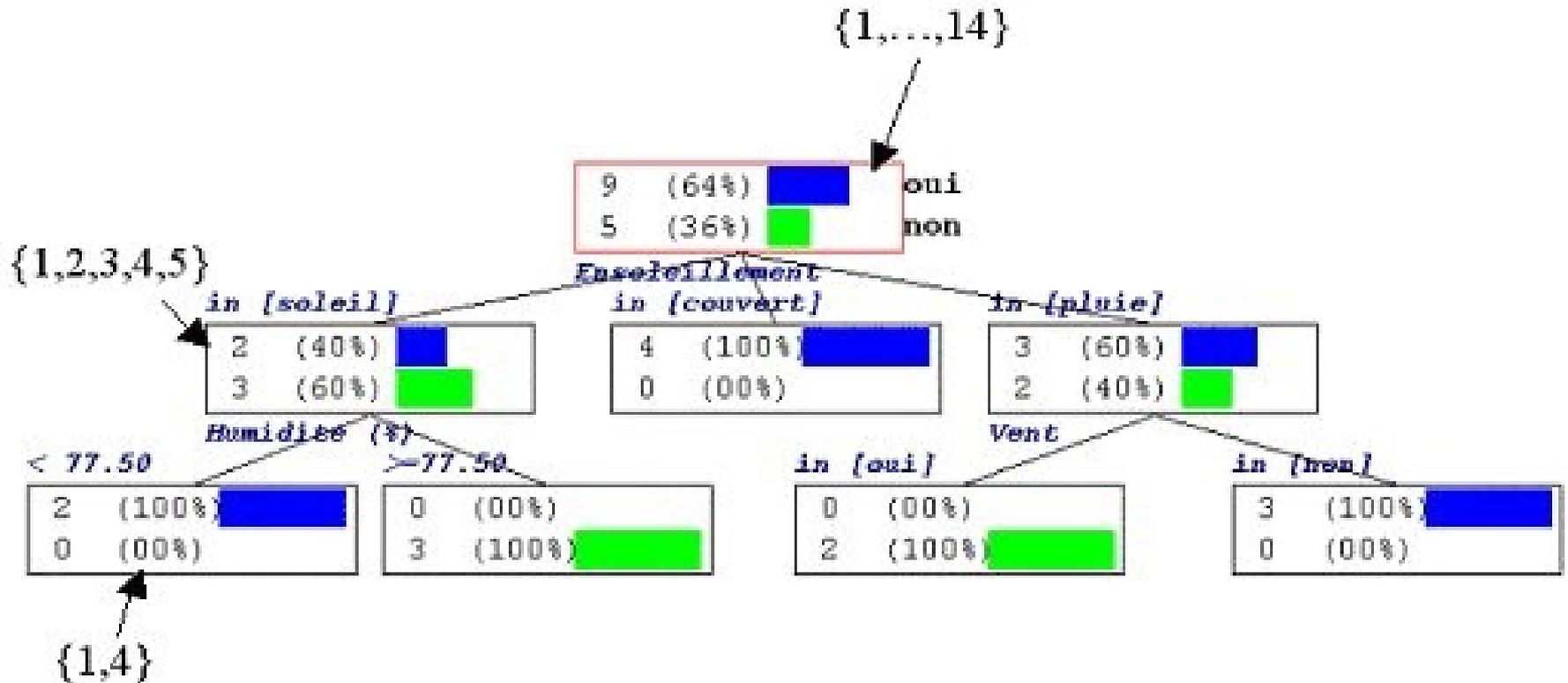
Image source : [http://en.wikipedia.org/wiki/File:Wikipedia\\_article-creation-2.svg](http://en.wikipedia.org/wiki/File:Wikipedia_article-creation-2.svg)

# Arbre de décision :

## +/- calcul de probabilités

Num éro	Ensoleillement	Température (°F)	Humidité (%)	Vent	Jouer
1	soleil	75	70	oui	oui
2	soleil	80	90	oui	non
3	soleil	85	85	non	non
4	soleil	72	95	non	non
5	soleil	69	70	non	oui
6	couvert	72	90	oui	oui
7	couvert	83	78	non	oui
8	couvert	64	65	oui	oui
9	couvert	81	75	non	oui
10	pluie	71	80	oui	non
11	pluie	65	70	oui	non
12	pluie	75	80	non	oui
13	pluie	68	80	non	oui
14	pluie	70	96	non	oui

# Arbre de décision : +/- calcul de probabilités



# Relation entre décisionnel et intelligence économique

- Dans le domaine des finances, et de l'économie, une simple décision ou simple mauvais choix peut coûter à l'entreprise des millions.
- Le décisionnel atteint sa plus grande importance dans le domaine financier.

# Processus de prise de décision dans la business intelligence

- **La prise de décision puis l'action stratégique à suivre sera basée :**
  - Non pas sur les données brutes,
  - Non pas sur les données extraites, et traitées,
  - Mais sur le résultat de transformation, puis de chargement puis représentation de ces données :
    - Sous forme de rapports, diagrammes, etc.

# Systeme d'information

- L'utilisation combinée de moyens informatiques, électroniques et de procédés de télécommunication permet aujourd'hui -
- selon les besoins et les intentions exprimés- d'accompagner, d'automatiser presque toutes les opérations incluses dans les activités en entreprise.

# Systeme d'information

- Un système d'information (SI) est un ensemble organisé de ressources (matériels, logiciels, personnel, données et procédures)
- qui permet de collecter, regrouper, classifier, traiter et diffuser de l'information sur un environnement donné.

# Systeme d'information d'aide à la décision (ou décisionnel)

- Dès que l'on parle de « Informatique décisionnelle », nous sommes obligé d'introduire le terme de SID (Système d'Information Décisionnel).
- Le SID a pour objectif de fournir des indicateurs sur la vie de l'entreprise. Pour cela, il exploite une richesse déjà à sa disposition : l'ensemble des données qu'une entreprise génère dans le cadre de son activité.

# Historique de la BI : Ralph Kimball

- Ralph Kimball (Né en 1944) est un auteur autour du sujet « entrepôts de données » et de la business intelligence. Il est largement considéré comme l'un des architectes du « datawarehouse ».
- Sa méthodologie, aussi connu comme la modélisation dimensionnelle ou la méthodologie de Kimball, est devenu le standard dans le domaine de l'aide à la décision.
- L'apport de la méthodologie Kimball est de montrer la manière de mener à bien un projet décisionnel, c'est à dire comment créer un système d'information décisionnel.

# Historique de la BI : Bill Inmon

- William H. Inmon (né en 1945) est un informaticien américain, reconnu par beaucoup comme le père du datawarehouse.
- Inmon écrit le premier livre, tient la première conférence (avec Arnie Barnett), a écrit la première colonne dans un magazine et a été le premier à offrir des cours sur le sujet.

# Historique de la BI : Bill Inmon

- Inmon donne la définition de ce qu'un datawarehouse : c'est un ensemble de données non volatiles, orienté par un sujet donné, relatif à un temps donné, en support à des décisions pour le management.
- Par rapport à l'approche Ralph Kimball, l'approche de Inmon est souvent caractérisée comme une approche top-down.
- L'apport de l'approche de bill inmon consiste à montrer ce que doit faire un système d'information décisionnel.

# Construction d'un entrepôt de données

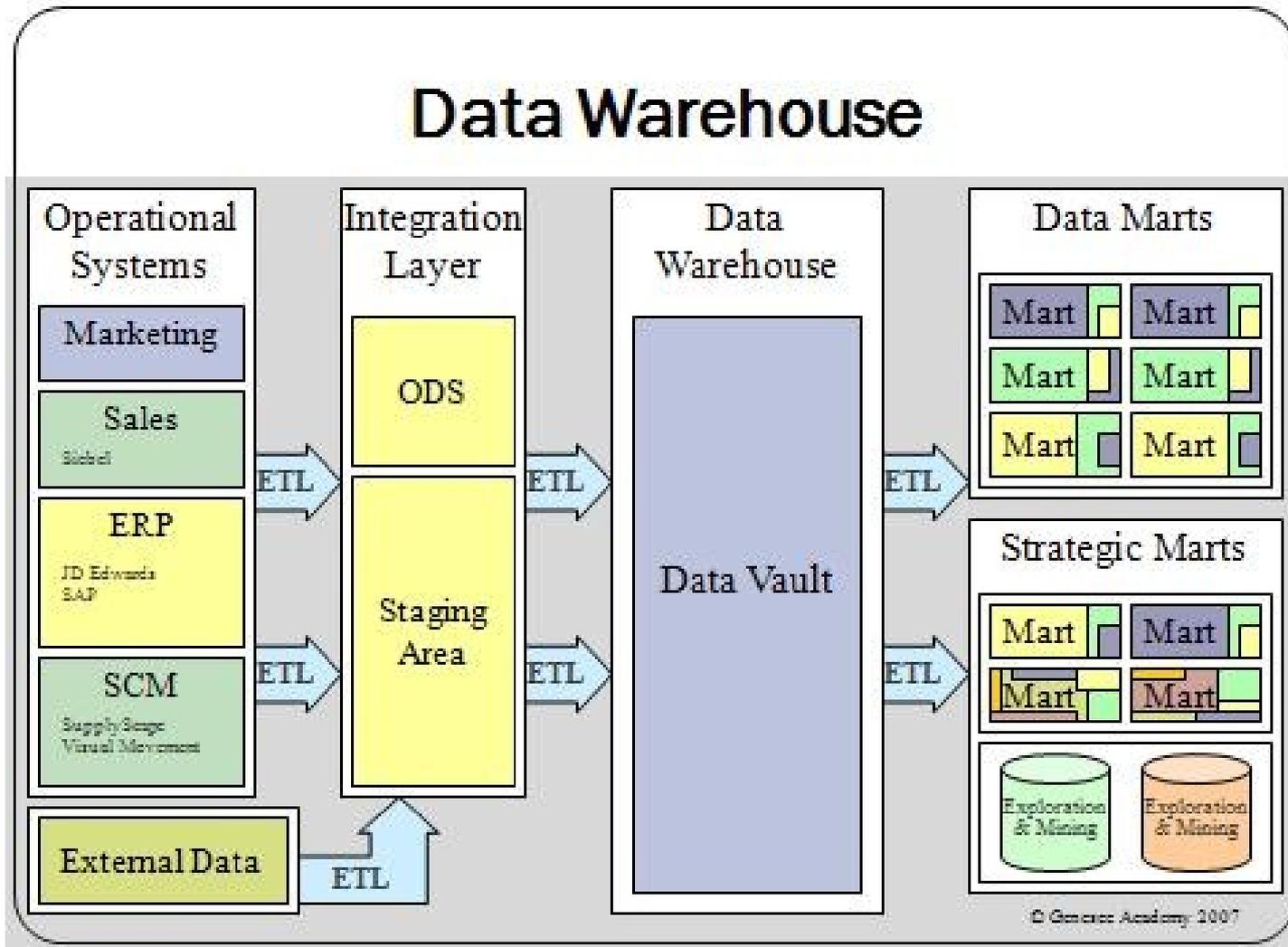


Image source :

[http://en.wikipedia.org/wiki/File:Data\\_warehouse\\_overview.JPG](http://en.wikipedia.org/wiki/File:Data_warehouse_overview.JPG)

# Quelques explications

ODS : operational data storage.

- Serveur de stockage de données opérationnel.

SCM : supply chain management.

- Gestion de la chaîne logistique.

ERP : entreprise resource planning.

- Progiciel de gestion intégré.

Data vault : modélisation de BDD relationnelles

# Sources de données

- Il faut savoir qu'un SI habituel traite et génère un ensemble de données « brutes », les sources de données (Finances, ERP, CRM, etc.) utilisées dans l'entrepôt de données peuvent provenir de :
- Données structurées :
  - Base de données relationnelles hétérogènes,
    - MS SQL SERVER, IBM DB2, Oracle, Informix,
    - XML, Excel,
    - Serveurs applicatifs, CRM (gestion de relation client),
- Données non structurées :
  - Fichiers plats, logs de scripts,

# Exemple d'ERP : openERP

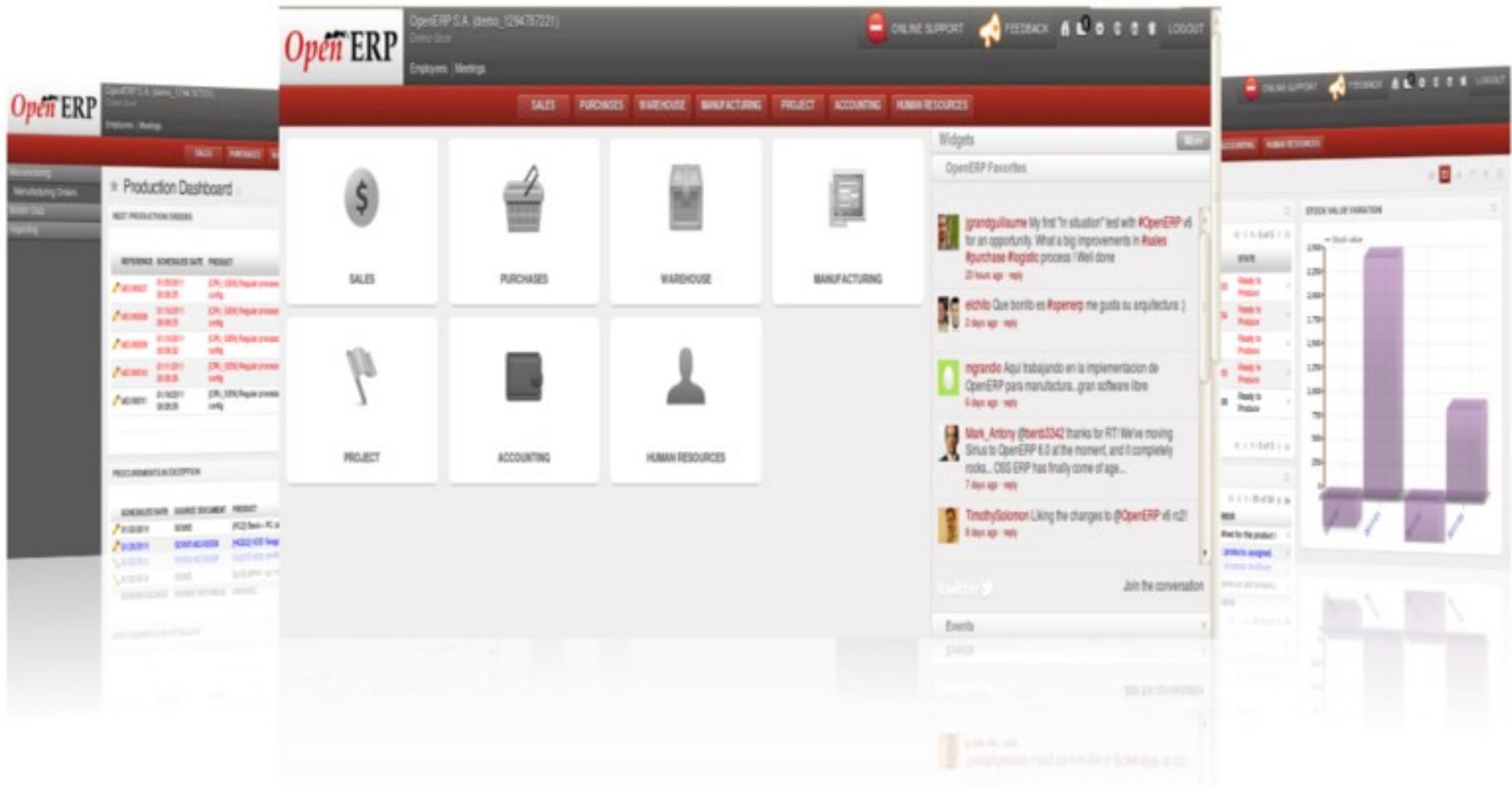


image source [http://commons.wikimedia.org/wiki/File:OpenERP\\_V6.png](http://commons.wikimedia.org/wiki/File:OpenERP_V6.png) licenced under the Creative Commons Attribution-Share Alike 3.0 Unported license. Attribution : Nicos interests

# Fichier plat

- Une base de données orientée texte est un modèle de base de données sous la forme d'un simple fichier.
- Un fichier plat est un fichier texte ou du texte combiné avec un fichier binaire contenant généralement un seul enregistrement par ligne.

# Extract Transform and Load

- l'extraction (récupération/collecte) de données stockées à partir de différentes sources de données, applications de l'entreprise, puis,
- La transformation du résultat pour être adéquat avec le format destination souhaité (voir cube OLAP), puis
- Le chargement/intégration/enregistrement du résultat dans des datamarts indépendantes.

# Exemple d'outil ETL : Talend Open Studio

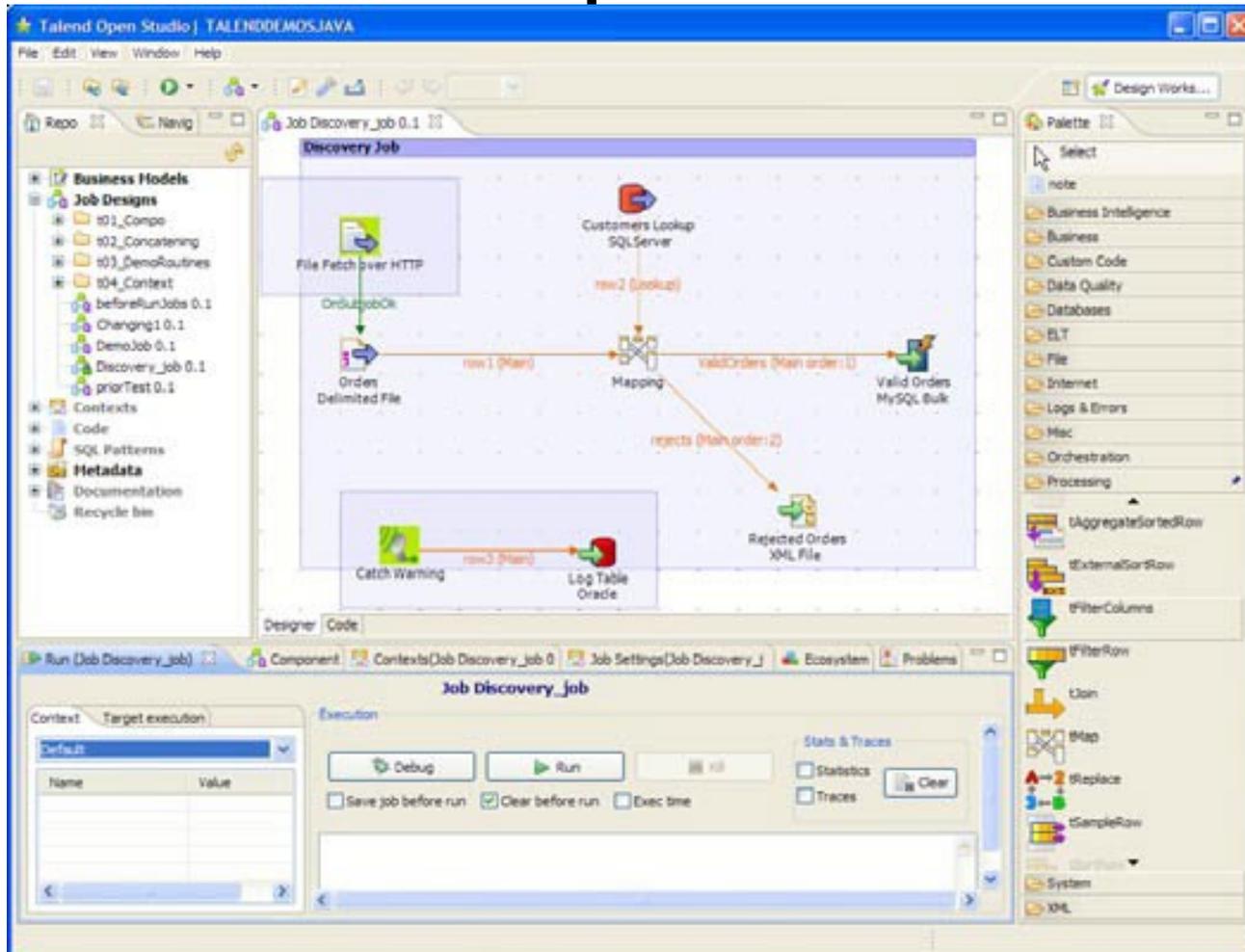


Image source :<http://marketplace.eclipse.org/sites/default/files/Talend-01.jpg>

# Transformation de données

- Le but est de transformer, traduire, convertir les données brutes précédemment extraites au format du datawarehouse au moment de la copie.
- en des informations dotées de sens, et utiles de point de vue économique, stratégique ou organisationnel.
- Trier, résumer, consolider, vérifier l'intégrité, réaliser les indices.

# Intégration de données

- si la donnée unitaire est la matière première, elle n'a que peu d'intérêt en tant que telle.
- Elle ne prend tout son sens que lorsque, regroupée avec d'autres données unitaires, elle forme un indicateur utile.
- D'où l'intérêt de ce que l'on appelle l'intégration de données.

# Intégration de données

L'intégration consiste à concentrer les données collectées dans un espace unifié, dont le socle informatique essentiel est l'entrepôt de données.

Élément central du dispositif, il permet aux applications décisionnelles de bénéficier d'une source d'information commune, homogène, normalisée et fiable, susceptible de masquer la diversité de l'origine des données.

# Intégration de données : fusion de 2 tables contenant les mêmes champs

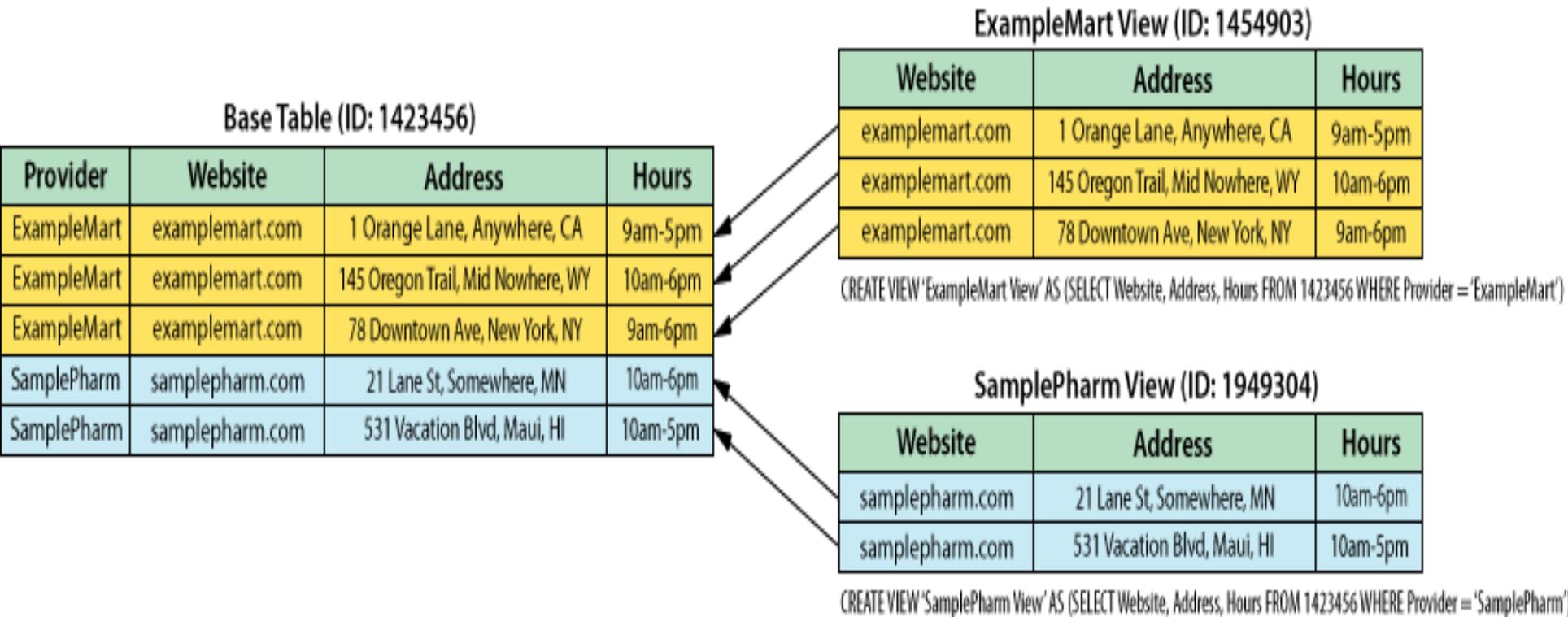


image source [https://developers.google.com/fusiontables/docs/articles/data\\_gathering](https://developers.google.com/fusiontables/docs/articles/data_gathering) licenced under the Creative Commons Attribution 3.0 License. Attribution : Kathryn Hurley

# Nettoyage des données : pourquoi

- Après l'agrégation de l'ensemble des données provenant des différentes sources de données, il se peut que l'on aie comme résultat :
- Des redondances (des n-uplets identiques),
- Des incohérences (des n-uplets contradictoires), erreurs de saisie, etc.
- Des erreurs (une transformation non réussie).

# Nettoyage de données : solution

- Le nettoyage des données permet de:
- vérifier et de corriger les données
- tout en assurant la standardisation des formats des n-uplets des tables de bases de données.
- Les procédés de nettoyage de données permettent d'optimiser la qualité et la cohérence de l'ensemble des données.

# Nettoyage de données : quoi

- **Le nettoyage de données consiste à :**
  - Corriger les erreurs de saisie, de frappe,
  - Vérifier l'intégrité des domaines par ex. les dates, (15/02/2000 vs 02/15/2000) pour les notations francophones et anglophones, les valeurs telles que Centrimètre ou Pouce, KM ou Mile, etc.
  - Proposer une alternative aux données nulles, manquantes, champs vides,

# Nettoyage de données : comment

- **Valeurs manquantes (nulles)**
  - Ignorer le tuple
  - Remplacer par une valeur fixe ou par la moyenne
- **Valeurs erronées ou inconsistantes**
  - Détecter par une analyse de voisinage
    - Écart par rapport à la moyenne
  - Remplacer par une valeur fixe ou par la moyenne
- **Inspection manuelle de certaines données.**

# Datawarehouse et Datamart

- Parmi les termes qui reviennent le plus souvent dans le domaine de l'intelligence économique, on note :
- Datawarehouse : regroupe toutes les données applicatives de l'entreprise, et représente l'élément principal du système d'information décisionnel.

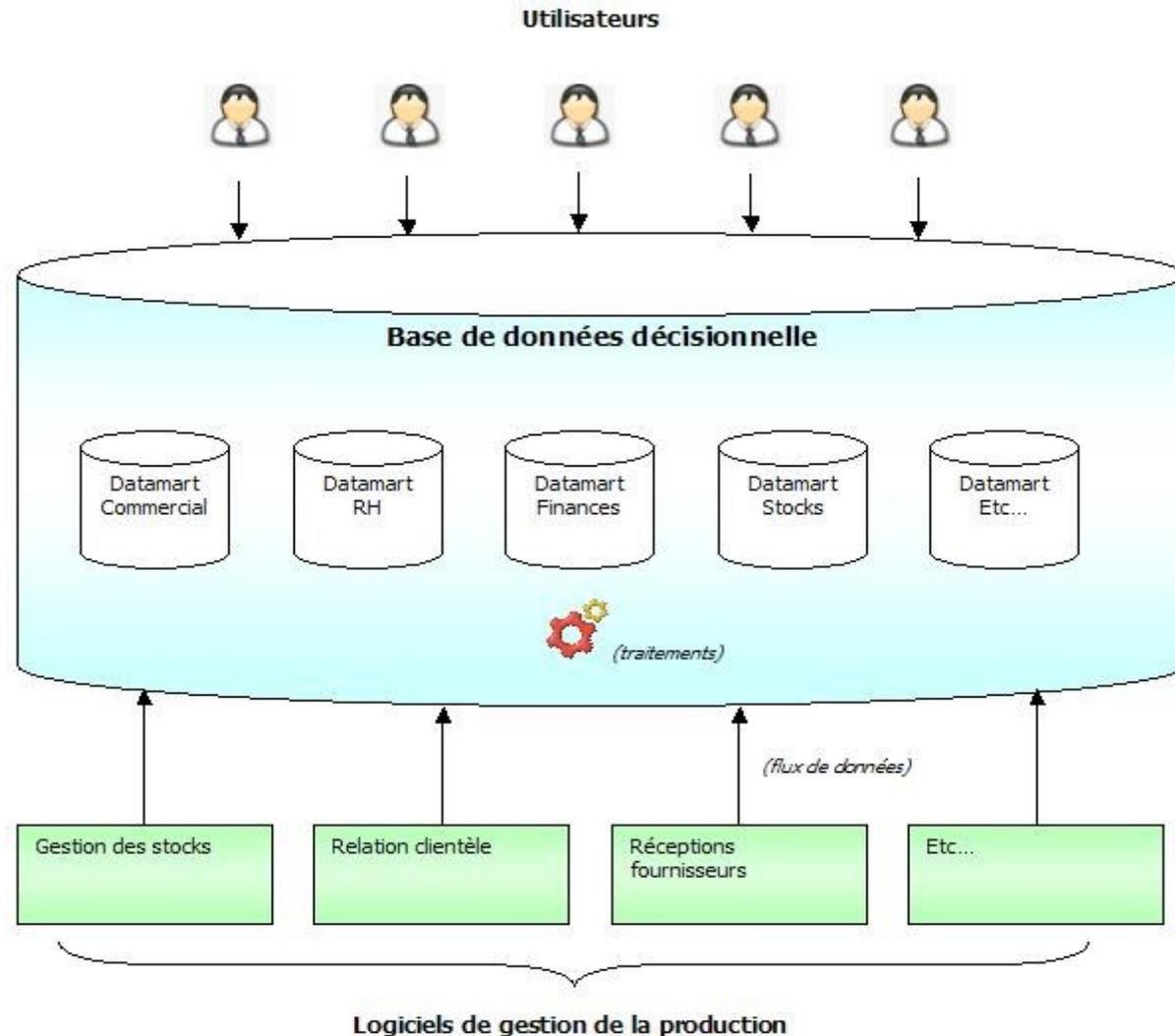
# Datawarehouse (entrepôts de données)

- L'entrepôt de données assure la consolidation des contenus sous forme de structures destinées à en faciliter la manipulation par les outils d'analyses, puis le stockage de ses données.
- base de données centrale utilisée pour le traitement, stockage, et présentation d'un ensemble de données provenant de différentes sources de données indépendantes, et hétérogènes.

# Datamart (magasin de données)

- Les datamarts (BDDR) généralement alimentés depuis les données du datawarehouse, sont des sous-ensembles d'informations
- destinés à fournir des données (et informations) aux utilisateurs finaux,
- souvent spécialisés vers un groupe, département, ou un type d'affaire (business), ou métier particulier de l'entreprise (marketing, risque, contrôle de gestion.)

# Exemple de datamarts pour le domaine de la gestion de production



# OLAP : représentation des données en SID

- Online Analytical Processing est un terme utilisé dans le domaine du décisionnel, pour désigner le traitement analytique en ligne.
- OLAP est une application/technique informatique orienté vers l'analyse instantané de données économiques selon plusieurs axes (dimensions),
- dans le but d'obtenir des rapports de synthèse tels que ceux utilisés en analyse financière, et aider la direction à avoir une vue transversale de l'activité d'une entreprise.

# Exemple de cube OLAP : tableau multi-dimensionnel

The image displays a BIRT Report Viewer window showing a multi-dimensional OLAP cube table. The table is titled "Look at script tab" and is organized by year (2003, 2004, 2005) and product line (Classic Cars, Motorcycles, Planes, Ships, Trains, Trucks and Buses, Vintage Cars). The "Planes" column is highlighted in red, indicating it is the selected dimension. The table data is as follows:

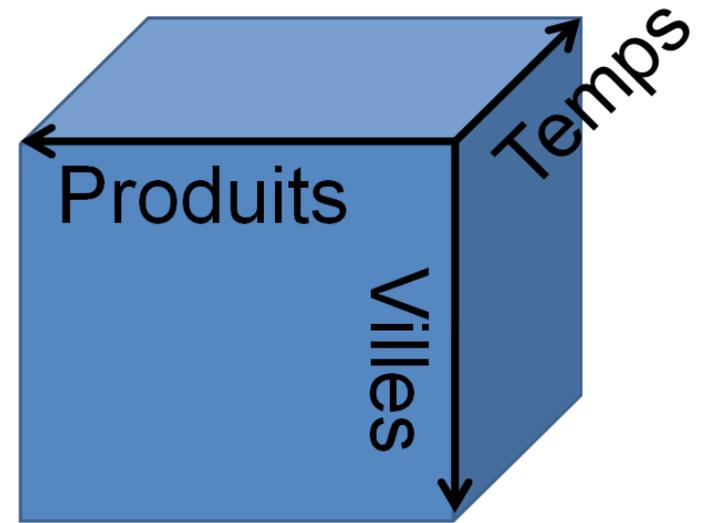
Look at script tab	Classic Cars	Motorcycles	Planes	Ships	Trains	Trucks and Buses	Vintage Cars
2003	152,581.55	33,062.22	37,136.27	24,446.99	7,810.61	43,593.71	106,982.20
	194,291.40	41,629.75	50,387.22	50,171.58	12,144.96	62,804.84	81,959.12
	243,978.40	79,845.21	45,047.15	24,272.54	7,027.48	73,842.46	110,486.95
2004	772,133.31	190,461.56	152,202.57	102,153.37	38,839.00	176,853.58	306,949.80
	317,306.69	85,681.60	65,159.24	66,762.89	21,028.52	67,942.19	136,849.15
	206,721.77	80,101.18	69,780.28	30,719.39	4,862.35	73,696.23	119,997.67
	419,674.86	127,310.80	105,973.57	66,859.21	21,728.25	106,833.60	200,420.71
2005	739,276.89	234,150.26	197,342.41	128,253.85	39,278.34	200,230.67	366,660.42
	374,001.31	112,384.53	76,925.89	44,522.65	18,235.25	101,332.86	159,257.22
	203,634.45	100,300.02	32,775.67	18,466.54	4,076.01	41,874.20	54,609.25

The script editor shows the following code for the "onCreateCell" event function:

```
1 function onCreateCell( cellInst, reportContext )
2 {
3     //Can reference cells by type or id - valid types "header" or "aggregation"
4     if( cellInst.getCellType() == "header" ) {
5         //Get data values see binding tab on crosstab
6         //if( cellInst.getDataValue("PRODUCTLINE") == "Planes" ){
7         if( reportContext.evaluate("dimension['ProductGroup']['PRODUCTLINE']") == "Planes" ){
8             cellInst.getStyle().setBackgroundColor("#FF0000");
9             cellInst.getStyle().setFontSize("12");
10        }else{
11            //Set the rest to yellow
12            cellInst.getStyle().setBackgroundColor("#FFFF00");
13            cellInst.getStyle().setFontSize("12");
14        }
15    }
16    //refer to crosstab header
17    if( cellInst.getCellID() == 24)
18        cellInst.getStyle().setBackg
19 }
```

# OLAP : représentation des tables de données en Cube à 3 dimensions

Produits	Villes	Temps
P1	V1	T1
P1	V2	T2
p2	V1	T1
p3	V3	T3



# Perspective = Dimensions

- Produit, et Région représentent ici des points de vue ou perspectives ou dimensions
- Une dimension est tout ce qui permet de consulter les données (les ventes) selon une catégorie ou repères, et nous donner une meilleure/différente perspective/point de vue.

# Table de dimension

- **La Table de dimension contient :**
  - les axes (dimension) étudiées dans l'analyse
  - selon lesquels on veut étudier des données à mesurer (les faits).
- **Exemple de dimension :**
  - Les **produits** (référence, famille, etc.
  - le temps (jour, mois, période, **année**, etc.
  - La Localisation (**ville**, pays, territoire, etc.,

# Ventes par produit et par période

- Par ex. vous pouvez avoir besoin de consulter les ventes d'un produit selon/durant une année.

Produits/Année	2001	2002
P1	1000	989
p2	12000	1099
p3	1220	1223

# Vente par région et par période

- Puis, vous pouvez souhaiter d'avoir un point de vue différent sur les ventes en consultant les performances d'un produit selon une région.

Produits/Région	Afrique	Europe	Amérique
P1	13000	1989	1245
p2	100	1234	2676
p3	0	1343	342

# Table de faits :

- La Table des faits contient les données observables (les faits, par exemple Quantité, Montant) que l'on possède sur un sujet et que l'on veut étudier,
- selon divers axes d'analyse (les dimensions), par exemple : Produit, Ville, Année.

Les tables situées aux extrémités sont les tables de dimensions ou niveaux de suivi.

# Fait : mesure

- Un fait est l'abréviation de « fait observable », on peut l'appeler également « mesure », qui est l'abréviation de mesure économique.
- Pour exemple, on peut avoir la nécessité de mesurer la quantité, le prix de vente, et le montant (de vente totale) concernant des produits vendus, par type de produit, par date, et par localisation.

# Mesures : intersection de deux dimensions

- Si l'on considère que l'on doit manipuler des tables à deux dimensions, considérons à titre d'exemple les deux dimensions suivantes « produit » et « année »,
- l'information/ou donnée (en vert dans la TDF) qui va résulter de l'intersection de ces deux dimensions peut être :
  - Profit, Vente, Prix, Quantité, etc.

# Modèle en étoile

- le modèle de données « en étoile » est typique des structures multidimensionnelles stockant des données atomiques ou agrégées.
- La table située au centre de l'étoile est la table des faits ou mesures (ou encore métriques) : ce sont les éléments mesurés dans l'analyse comme les montants, les quantités, les taux, etc.

# Modèle en étoile : schéma

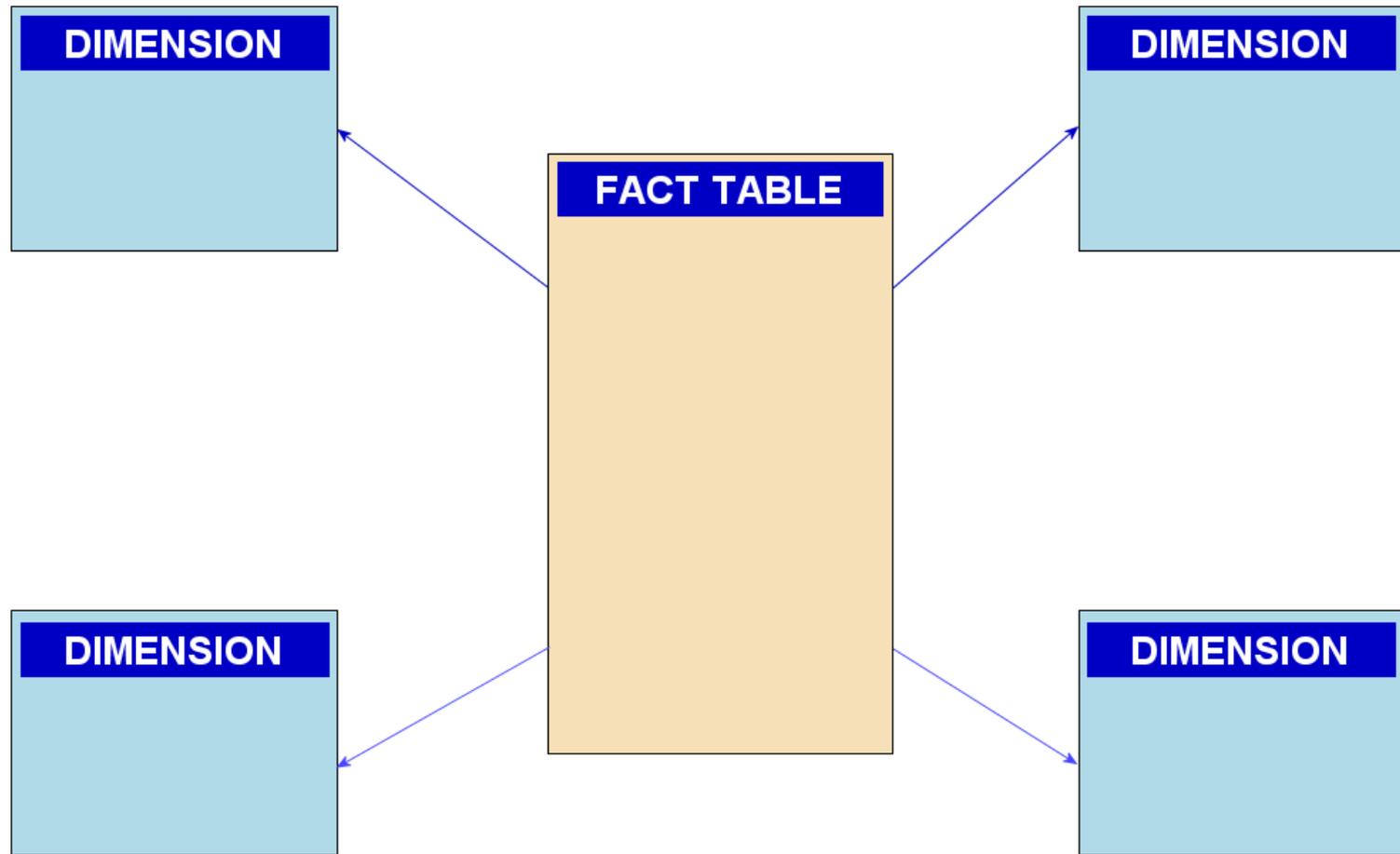
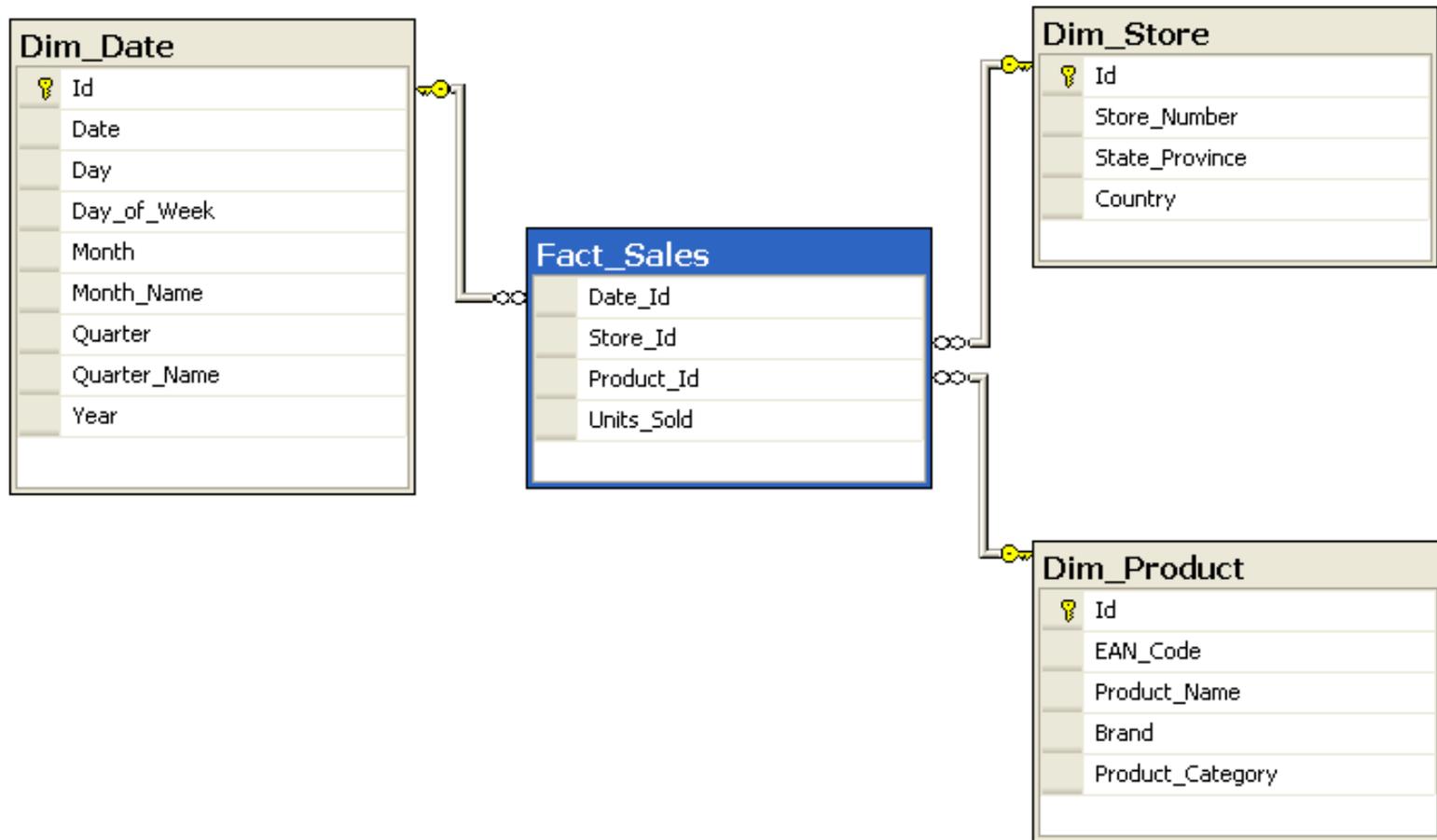


image source <http://commons.wikimedia.org/wiki/File:Star-schema.png> licensed under the Creative Commons Attribution-Share Alike 3.0 Unported licence. Attribution : SqlPac

# Modèle en étoile : exemple



# Hiérarchie : fils de dimension

- Considérons cette fois-ci, que vous voulez consulter les ventes par région, mais que vous voulez approfondir votre recherche par ville, etc.
- C'est ici, que la notion de hiérarchie intervient en décomposant les dimensions, en sous-unités permettant
- d'avoir une meilleure granularité des résultats.
- Flocon : variante du modèle en étoile : chaque table de dimension fait apparaître la hiérarchie sous-jacente (catégorie pour produit).

# Modèle de données en flocon

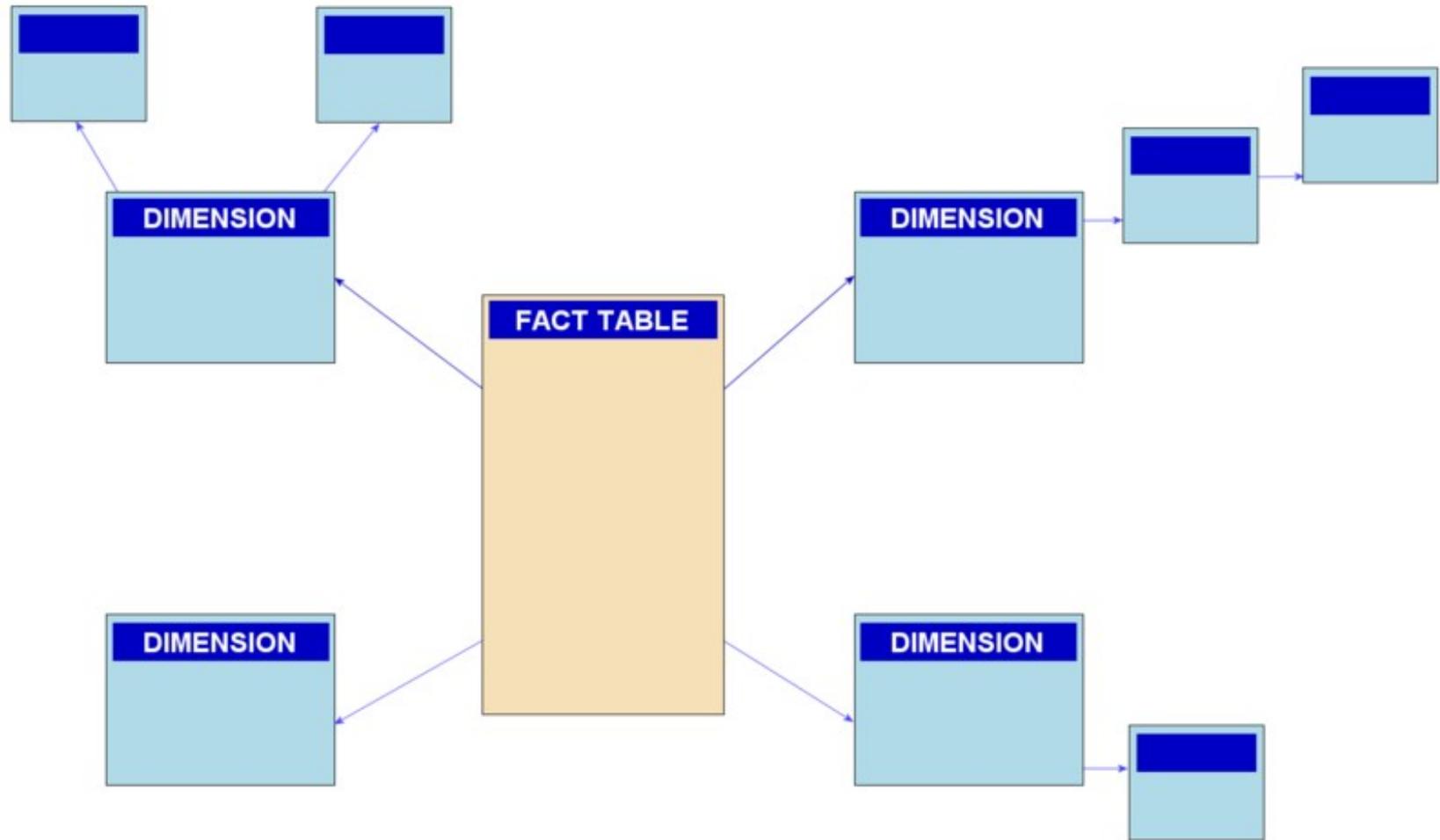


image source <http://en.wikipedia.org/wiki/File:Snowflake-schema.png> licensed under the Creative Commons Attribution-ShareAlike 3.0 License. Attribution : SqlPac

# Modèle de données en flocon : exemple

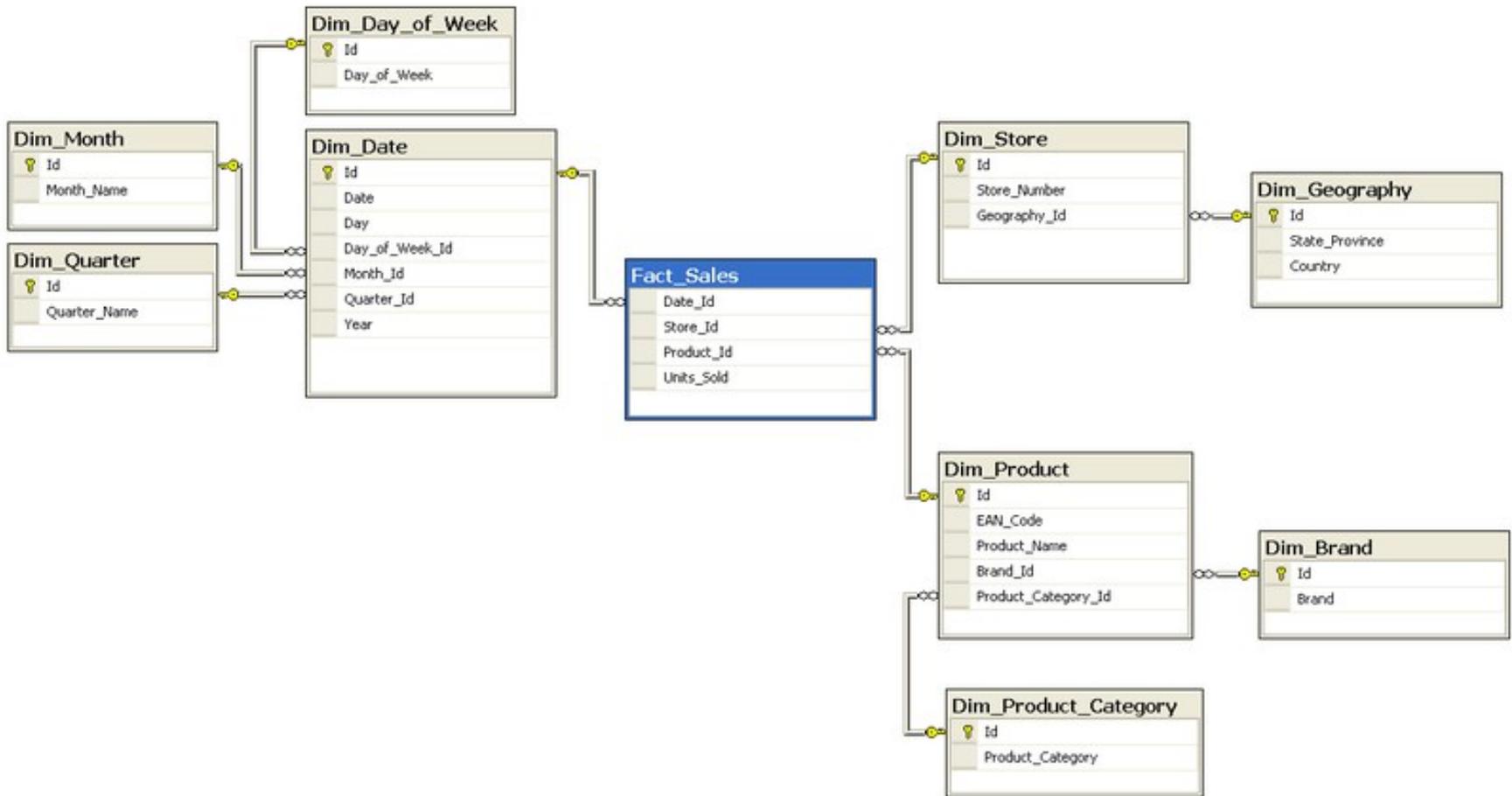


image source <http://en.wikipedia.org/wiki/File:Snowflake-schema-example.png> licensed under the Creative Commons Attribution-Share Alike 3.0 Unported licence. Attribution: SqlPac at English Wikipedia

# Chargement de données

- Cela consiste à représenter les données dans un format compréhensible (rapports).
- Puis le distribuer/publier vers l'utilisateur final, en général détenant des compétences d'analyse financière/économique, ou stratégique. La distribution peut être par portail Web, mail, ou autre, etc.

# Restitution, Représentation, et diffusion des données

- Représentation visuelle des données rentre sous le terme « data visualization ».
- La représentation graphique de données statistiques ou visualisation de données statistiques est un résumé visuel des données chiffrées. Elle permet en un seul coup d'œil d'en saisir la tendance générale.

# Rapports graphiques

**Sectoral contribution to total business services  
growth, 2001-2005\***  
(in per cent)

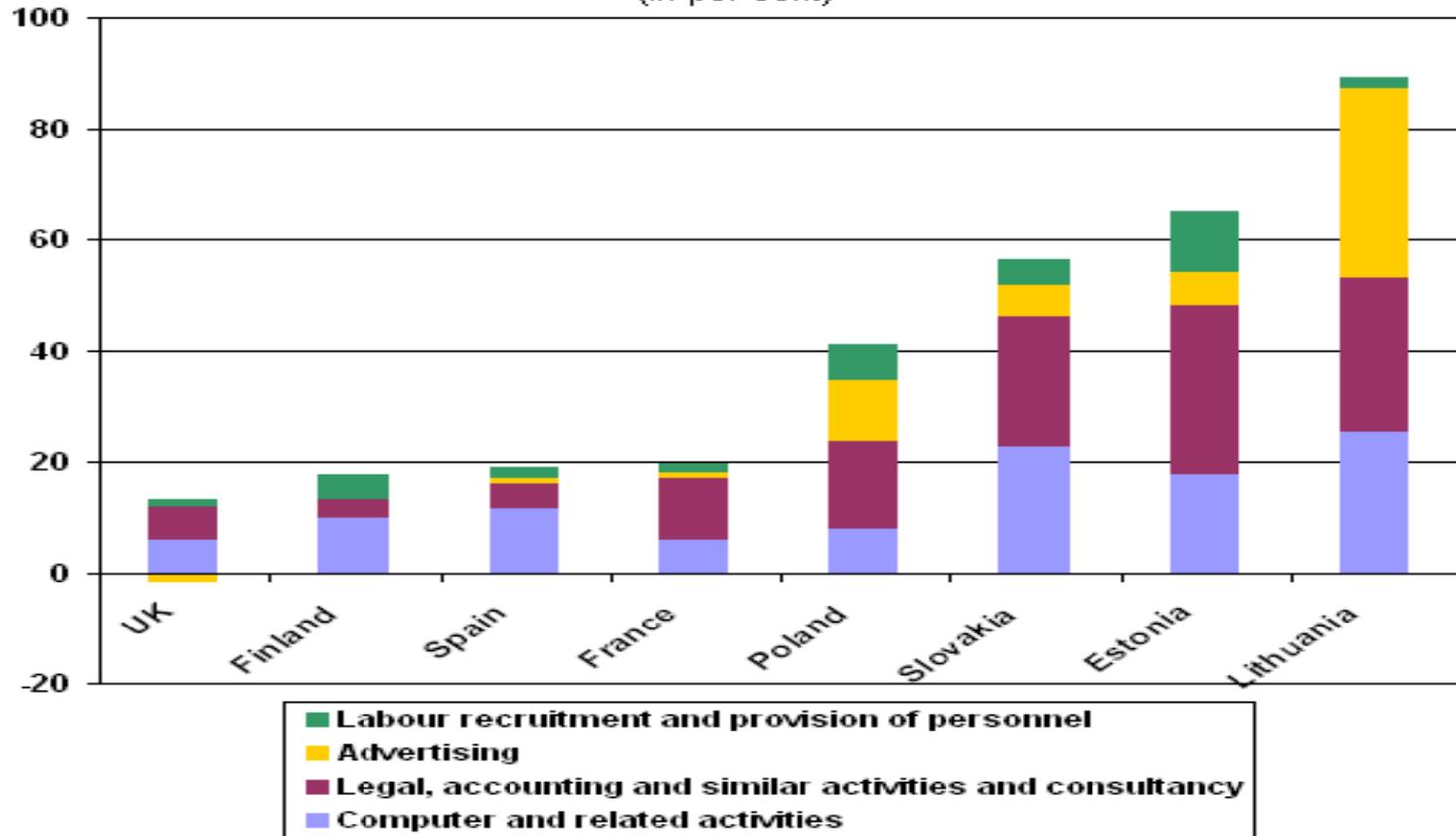


image source [http://en.wikipedia.org/wiki/File:ECE\\_weekly\\_235.gif](http://en.wikipedia.org/wiki/File:ECE_weekly_235.gif) licensed under the Creative Commons Attribution-Share Alike 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic licence.  
Attribution : UNECE

# Limites de la restitution de données

- La représentation de résultats sous forme de rapports (à valeur économique) ou le « reporting » n'est pas suffisant pour pouvoir conclure des décisions stratégiques.
- Un autre domaine intervient ici, c'est ce que l'on appelle « data analytics », ou l'analyse de données.

# Beaucoup de données mais sans information à premier abord



image source : [http://en.wikipedia.org/wiki/File:Sao\\_Paulo\\_Stock\\_Exchange.jpg](http://en.wikipedia.org/wiki/File:Sao_Paulo_Stock_Exchange.jpg) licensed under the Creative Commons Attribution 2.0 Generic licence. Attribution : Rafael Matsunaga

# Data analytics

- Obtenir l'information à partir de la donnée,
- Regrouper les éléments de prise de décision,
- Utilisabilité du tableau de bord par
  - Des employés avec connaissance limitée en:
    - bureautique,
  - Des employés sans connaissance approfondie en :
    - Informatique, bases de données,
    - En finance-comptabilité,
    - Ressources humaines,
    - Logistique, productique,

# Outils de business intelligence

# Business Intelligence Reporting Tool

- BIRT est un système de reporting pour les applications basées sur Java et Java EE (en particulier).
- BIRT a deux composantes principales:
  - un concepteur de rapport basé sur Eclipse,
  - et un composant d'exécution que vous pouvez ajouter à votre serveur d'applications.
  - BIRT propose également un moteur de diagrammes qui vous permet d'ajouter des graphiques pour votre propre application.

# Possibilités de BIRT

- Avec BIRT , vous pouvez ajouter une riche variété de rapports à votre application .
- Listes - Les rapports les plus simples sont des listes de données . Comme les listes s'allongent , vous pouvez ajouter le regroupement pour organiser les données ( commandes groupées par client, produits regroupés par le fournisseur) .
- Si vos données sont numériques, vous pouvez facilement ajouter des totaux (sum), des moyennes (avg) et d'autres résumés .

# Possibilités de BIRT

- Graphiques - Les données numériques sont beaucoup plus faciles à comprendre lorsqu'ils sont présentés sous forme de graphiques. BIRT fournit des camemberts, graphiques linéaires et beaucoup plus. Les
- Les Graphiques BIRT peuvent être rendus en SVG et soutien des événements pour permettre une interaction de l'utilisateur .

# Exemple de diagramme

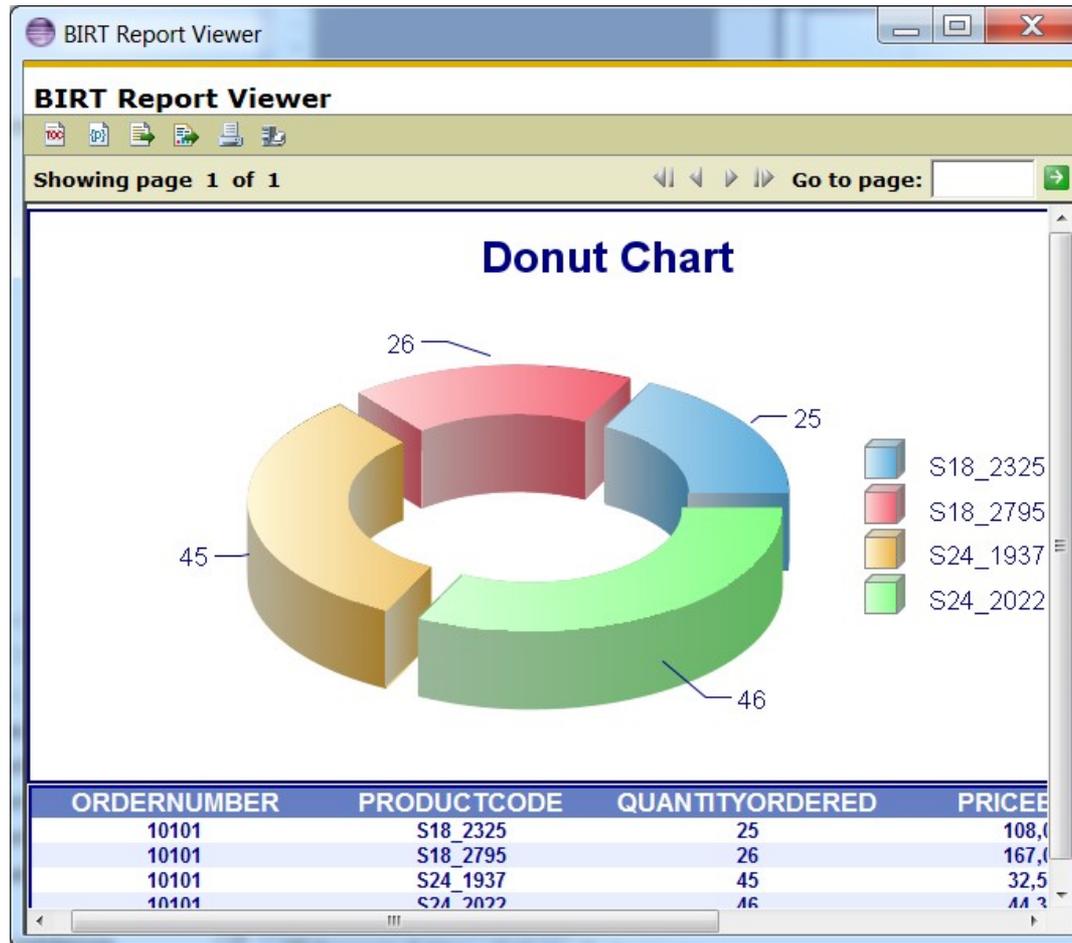
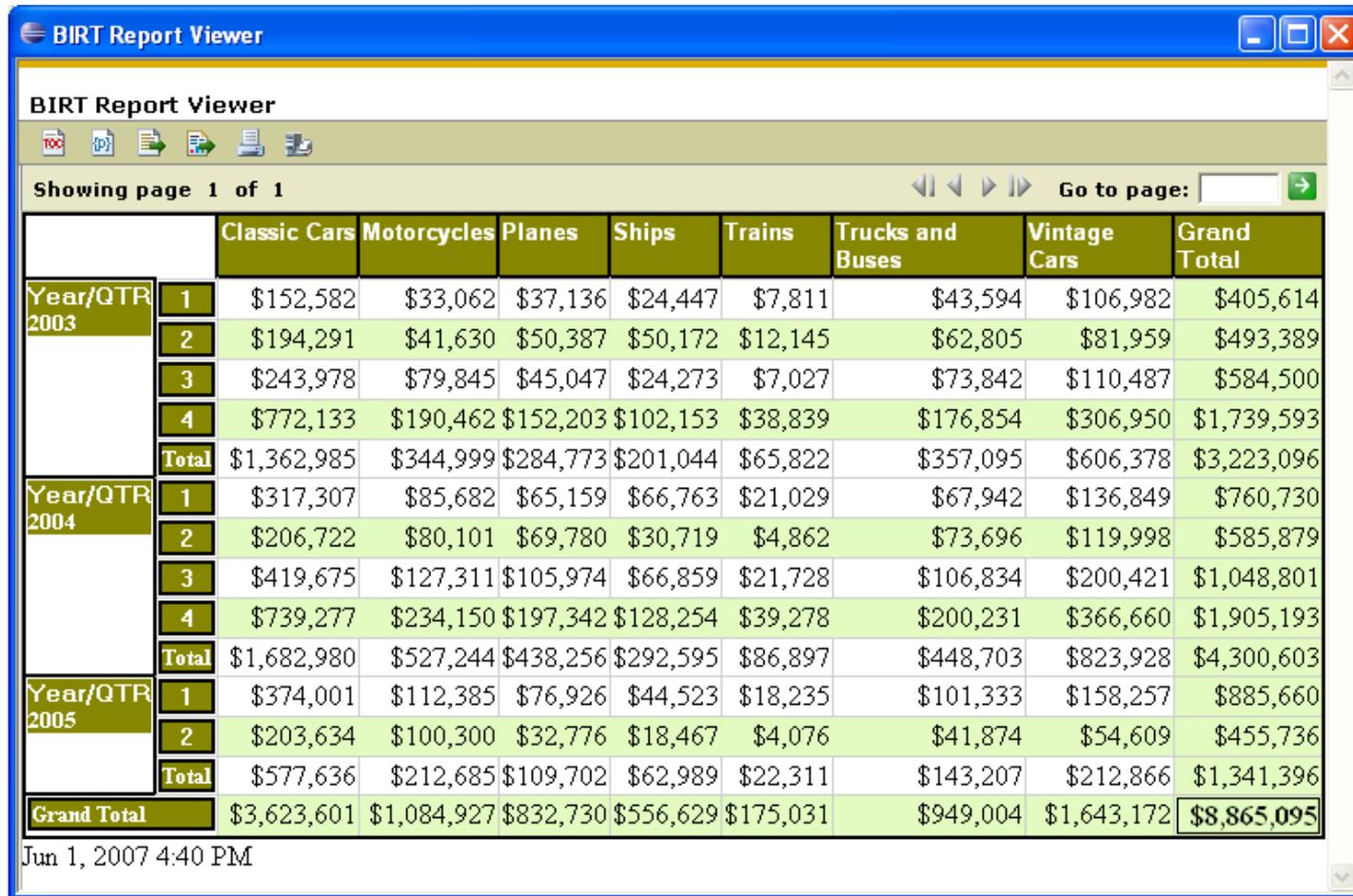


image source : <http://www.eclipse.org/birt/phoenix/project/notable4.2.php>

# Possibilités de BIRT

- Tableaux croisés (également appelés un tableau « crosstab » ou matrice) montre les données en deux dimensions : les ventes par trimestre par exemple.
- Lettres & Documents - Avis, des lettres et d'autres documents textuels sont faciles à créer avec BIRT.
- Les documents peuvent inclure du texte, mise en forme , des listes , des tableaux et plus .

# Tableau croisé avec Eclipse BIRT



BIRT Report Viewer

Showing page 1 of 1

Go to page:

		Classic Cars	Motorcycles	Planes	Ships	Trains	Trucks and Buses	Vintage Cars	Grand Total
Year/QTR 2003	1	\$152,582	\$33,062	\$37,136	\$24,447	\$7,811	\$43,594	\$106,982	\$405,614
	2	\$194,291	\$41,630	\$50,387	\$50,172	\$12,145	\$62,805	\$81,959	\$493,389
	3	\$243,978	\$79,845	\$45,047	\$24,273	\$7,027	\$73,842	\$110,487	\$584,500
	4	\$772,133	\$190,462	\$152,203	\$102,153	\$38,839	\$176,854	\$306,950	\$1,739,593
	Total	\$1,362,985	\$344,999	\$284,773	\$201,044	\$65,822	\$357,095	\$606,378	\$3,223,096
Year/QTR 2004	1	\$317,307	\$85,682	\$65,159	\$66,763	\$21,029	\$67,942	\$136,849	\$760,730
	2	\$206,722	\$80,101	\$69,780	\$30,719	\$4,862	\$73,696	\$119,998	\$585,879
	3	\$419,675	\$127,311	\$105,974	\$66,859	\$21,728	\$106,834	\$200,421	\$1,048,801
	4	\$739,277	\$234,150	\$197,342	\$128,254	\$39,278	\$200,231	\$366,660	\$1,905,193
	Total	\$1,682,980	\$527,244	\$438,256	\$292,595	\$86,897	\$448,703	\$823,928	\$4,300,603
Year/QTR 2005	1	\$374,001	\$112,385	\$76,926	\$44,523	\$18,235	\$101,333	\$158,257	\$885,660
	2	\$203,634	\$100,300	\$32,776	\$18,467	\$4,076	\$41,874	\$54,609	\$455,736
	Total	\$577,636	\$212,685	\$109,702	\$62,989	\$22,311	\$143,207	\$212,866	\$1,341,396
Grand Total		\$3,623,601	\$1,084,927	\$832,730	\$556,629	\$175,031	\$949,004	\$1,643,172	\$8,865,095

Jun 1, 2007 4:40 PM

image source : <http://www.eclipse.org/birt/phoenix/project/notable2.2.php>

# Création de cube OLAP avec BIRT

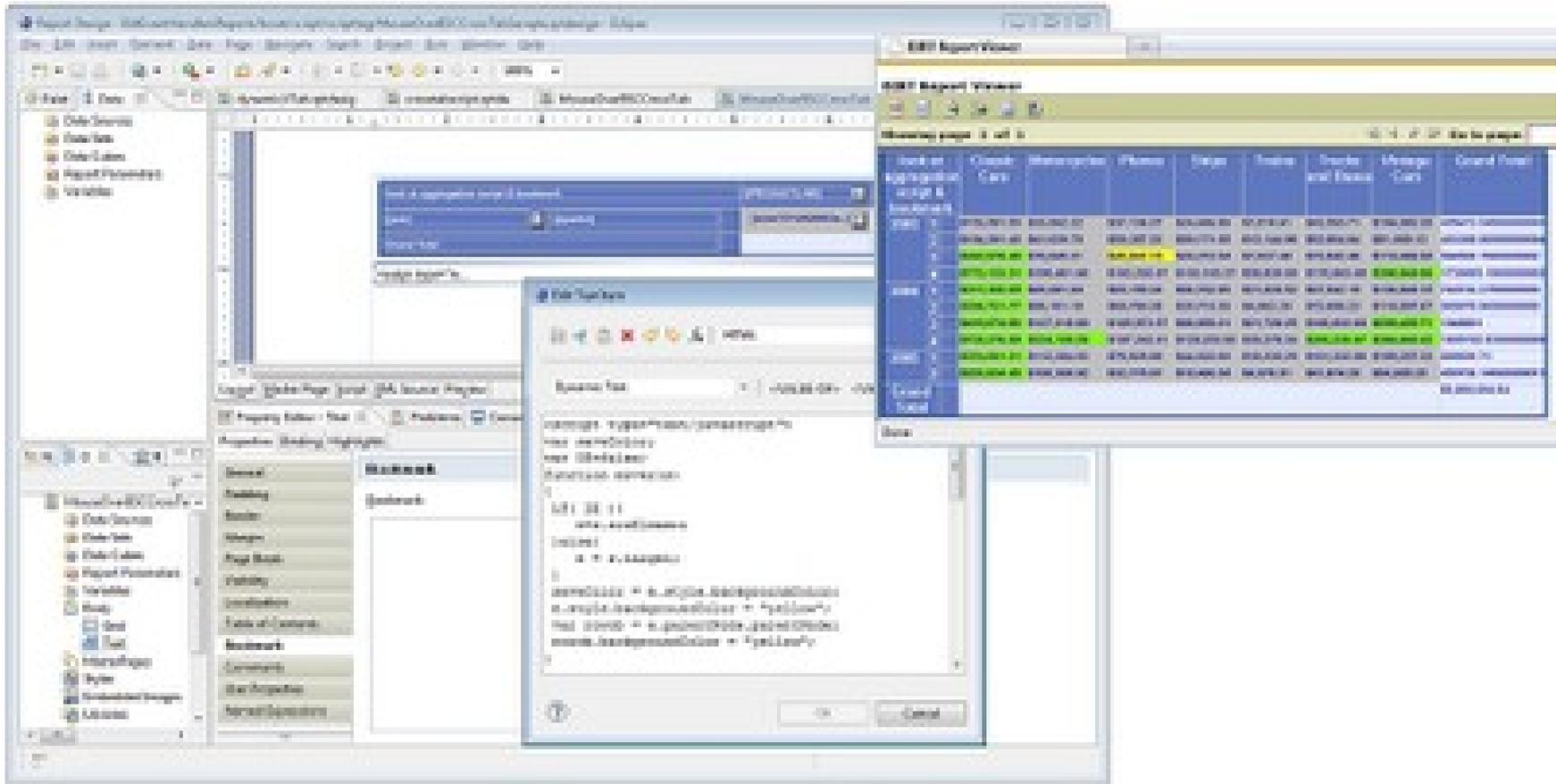


image source : <http://birtworld.blogspot.com/2010/07/birt-and-olap.html>

# Possibilités de BIRT

- Rapports composés - De nombreux rapports doivent combiner tout ce dont on a parlé en un seul document .
- Par exemple, une déclaration à la clientèle peut indiquer l'information pour le client, fournir un texte sur les promotions en cours, et de fournir des listes de paiements et les charges.
- Un rapport financier peut inclure des avertissements, des graphiques, des tableaux, toutes avec beaucoup de formatage qui correspond à des couleurs de l'entreprise.

# Possibilités de BIRT

- Les Rapports BIRT se composent de quatre parties principales : les données , les transformations de données , la logique « métier » et présentation .
- Les Données - Bases de données , services Web, objets Java, peuvent fournir des données à votre rapport de BIRT. BIRT offre un support pour les sources de données JDBC , XML , Web Services , et fichiers plats, ainsi que le soutien pour l'utilisation de code pour obtenir d'autres sources de données .
- En outre, un seul rapport peut inclure des données provenant de plusieurs sources de données . BIRT fournit également une fonctionnalité qui permet aux sources de données disparates pour être combinés en utilisant jointures internes et externes.

# Possibilités de BIRT

- La Transformation des données - les rapports présentent des données triées, résumées, filtrées et regroupées pour répondre aux besoins de l'utilisateur.
- Bien que les bases de données peuvent faire partie de ce travail, BIRT doit le faire pour les sources «simples» de données telles que des fichiers plats ou des objets Java.
- BIRT permet des opérations complexes telles que le regroupement sur les sommes (sur différentes dimensions, ce qui n'est pas possible avec l'utilisation d'un simple groupe by et decode), les pourcentages des totaux globaux et plus.

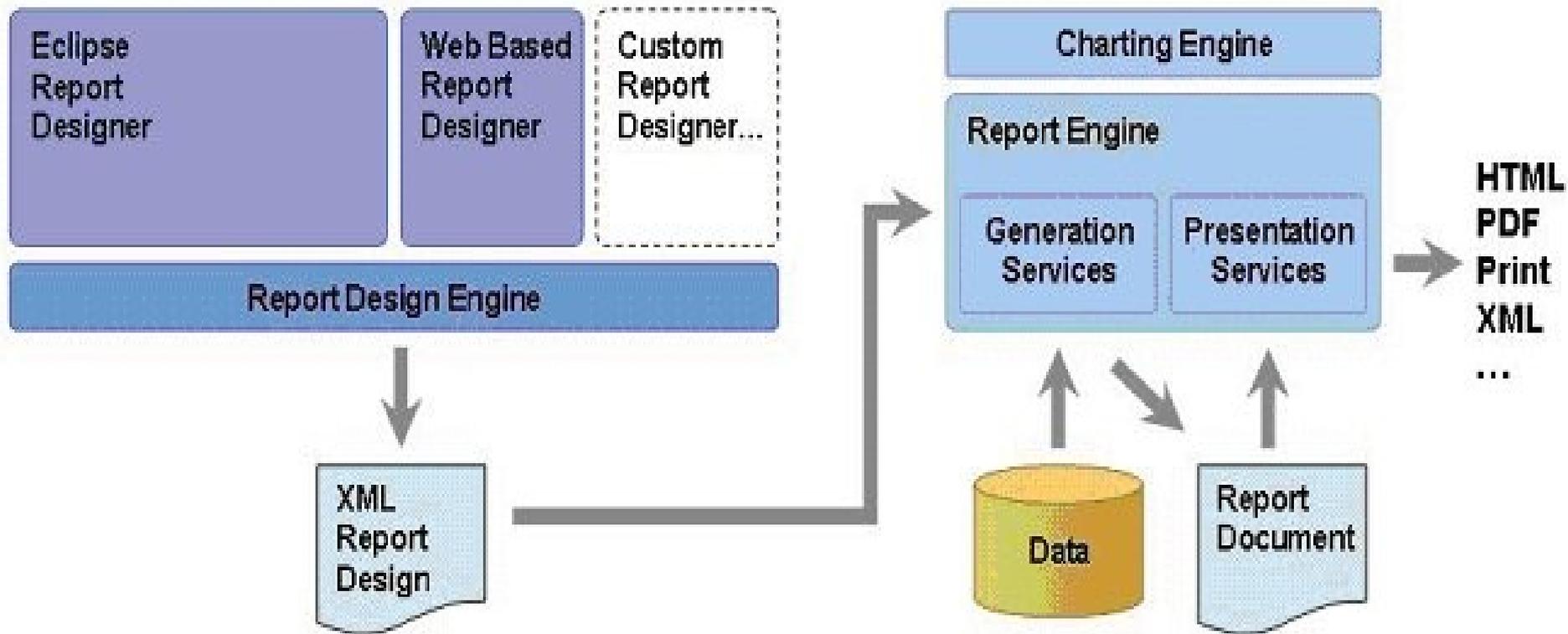
# Possibilités de BIRT

- Business Logic (la logique métier) - les données du monde réel sont rarement structurées exactement comme on le souhaite pour un rapport.
- De nombreux rapports exigent une logique métier spécifique pour convertir des données brutes en informations utiles pour l'utilisateur.
- Si la logique est juste pour le rapport, nous pouvons utiliser un script, puisque BIRT supporte JavaScript. Si votre application contient déjà la logique, vous pouvez l'utiliser dans votre code Java existant.

# Possibilités de BIRT

- Présentation - Une fois que les données sont prêtes, vous avez un large éventail d'options pour le présenter à l'utilisateur. Tableaux, graphiques, texte et plus encore.
- Un seul ensemble de données peut apparaître dans de multiples façons, et un seul rapport peut présenter des données à partir de plusieurs ensembles de données.

# Architecture de BIRT



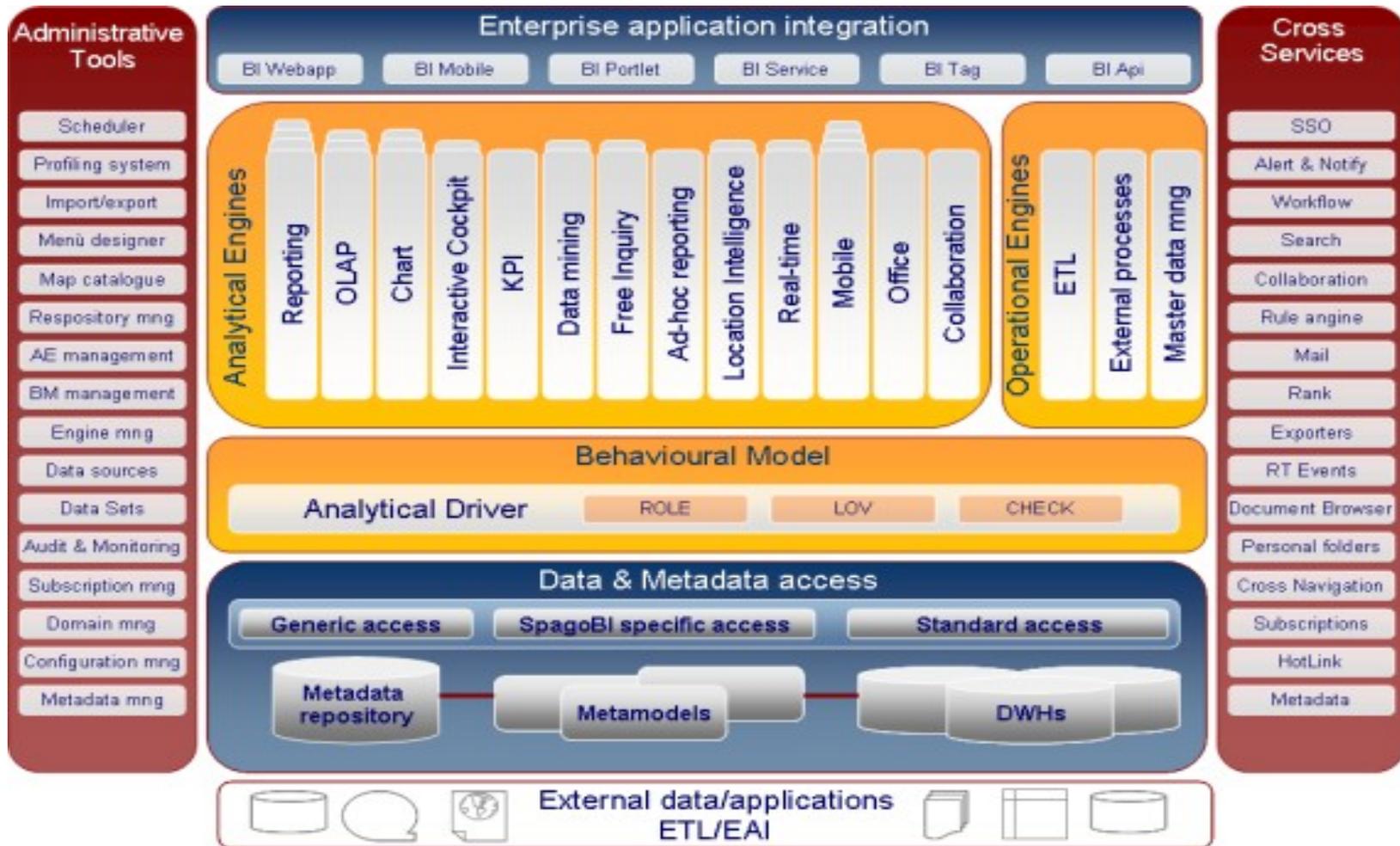
# Limites de BIRT

- Pour l'instant BIRT ne gère pas :
- L'analyse statistique,
- Les outils de Data Mining,
- Les outils de modélisation du Data Warehouse,
- Les outils Extract Transform and Load (ETL).

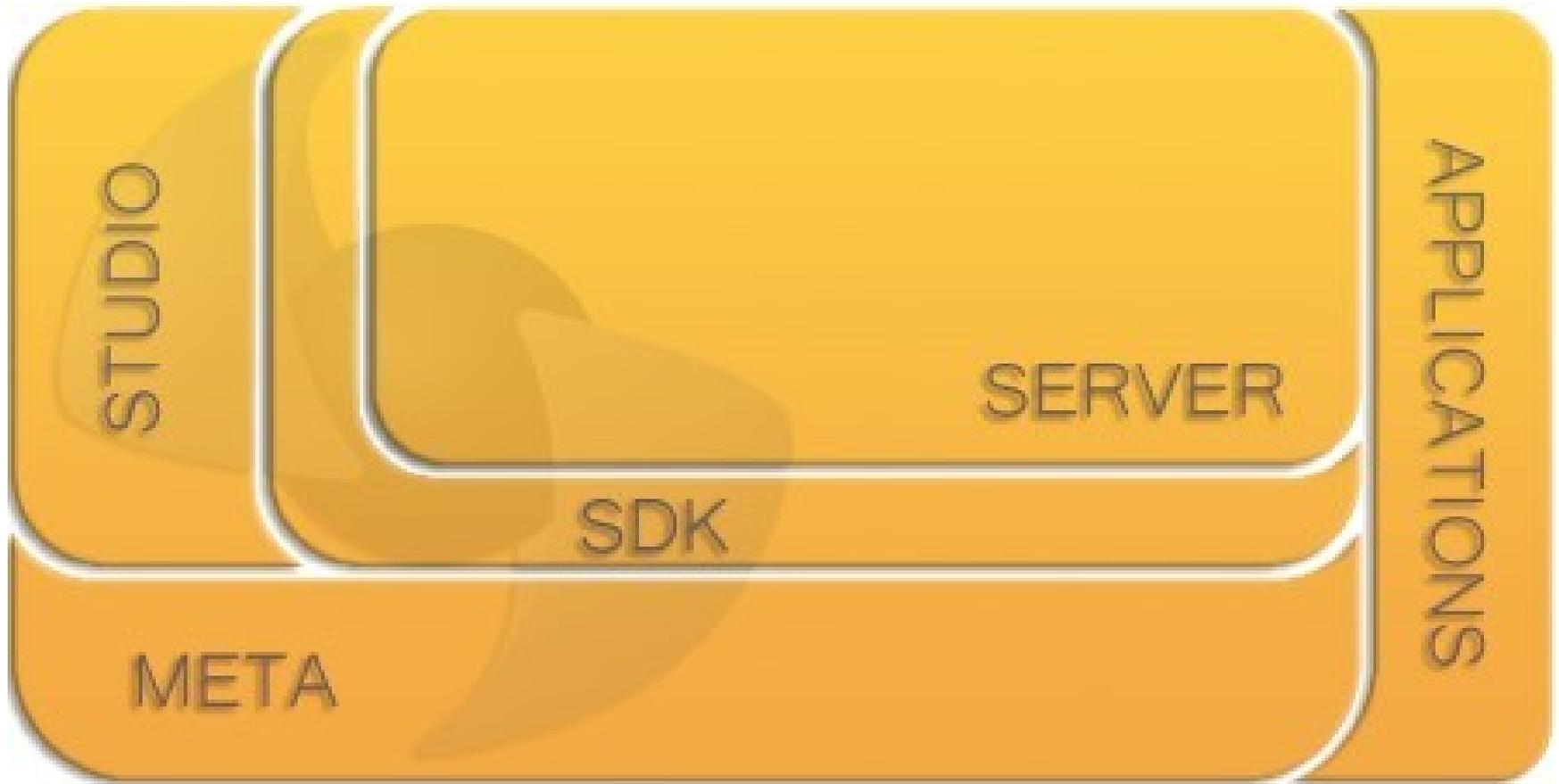
# Spago BI

- SpagoBI est une solution de Business Intelligence entièrement open source.
- Cette solution fait partie de l'initiative libre/open source SpagoWorld, fondée et soutenue par Engineering Group.
- SpagoBI est distribué sous licence Mozilla Public License, qui est compatible avec les usages commerciaux.

# SpagoBI Architecture



# SpagoBI



# SpagoBI studio, sdk, meta

- SpagoBI Studio est l'environnement de développement intégré qui permet aux développeurs de concevoir et modifier les modèles des différents moteurs d'analyse, comme rapports, diagrammes, rapports d'analyse géographique et tableaux de bord.
- Ce module soutient la phase de déploiement, où les documents analytiques sont testés et publiés sur SpagoBI Server, avec lequel SpagoBI Studio interagit grâce à SpagoBI SDK.
- SpagoBI SDK est utilisé pour l'intégration des services fournis par SpagoBI Server. Il permet d'intégrer des documents à travers un éventail de services web et de publier les documents SpagoBI dans de portails et applications externes.

# Spago BI

- SpagoBI Meta a été spécifiquement conçu pour la gestion et interrogation des métadonnées.
- La plateforme permet la gestion des métadonnées techniques et des métadonnées métier, en permettant aux administrateurs et utilisateurs d'interpréter correctement leurs sources de données.
- SpagoBI Server est le module principal de la suite, qui offre toutes les fonctionnalités centrales et analytiques de la plateforme.

# Rapports, pour l'affichage des données structurées,

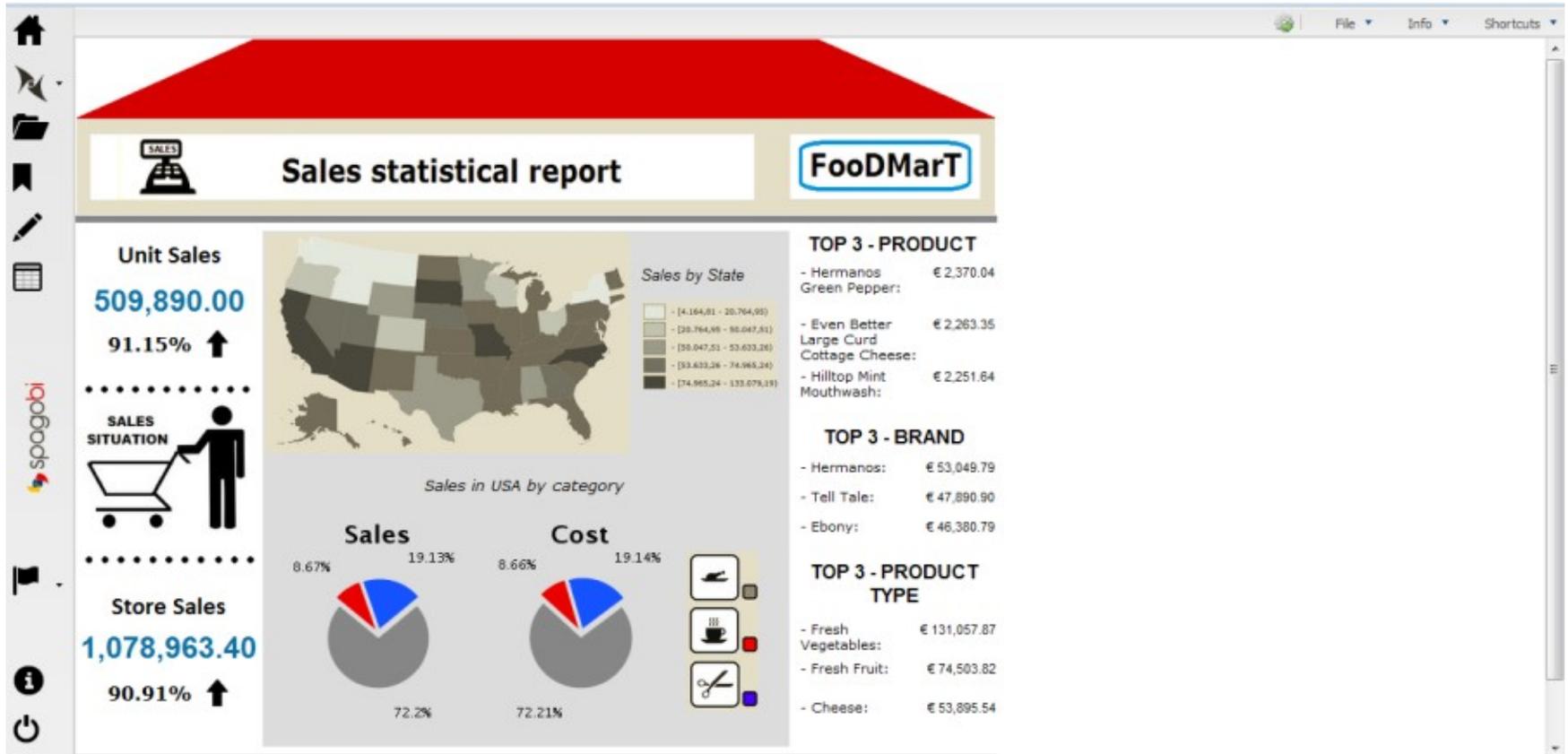


image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Rapports, pour l'affichage des données structurées,



image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Analyse OLAP, pour consulter et naviguer à travers les données,

Product	Region	Measures				
		Trend	Units Ordered	Units Shipped	Warehouse Sales	Warehouse Cost
- Drink	+ All Regions	1.22%	61,760	56,571	\$48,363.30	\$21,780.69
+ Alcoholic Beverages	+ All Regions	1.19%	16,626	15,408	\$13,321.47	\$6,081.58
+ Beverages	- All Regions	1.23%	34,017	31,331	\$27,424.75	\$12,318.27
	- San Francisco	1.56%	196	196	\$241.28	\$94.38
	- San Francisco	1.56%	196	196	\$241.28	\$94.38
	Store 14	1.56%	196	196	\$241.28	\$94.38
	- Mexico Central	1.16%	3,995	3,776	\$3,467.70	\$1,609.01
	+ Hidalgo	1.21%	744	704	\$667.13	\$301.28
	+ Mexico City	1.11%	2,048	1,869	\$1,753.98	\$832.55
	+ Orizaba	1.20%	1,203	1,203	\$1,046.59	\$475.18
	- Mexico South	1.18%	818	710	\$704.11	\$323.59
	- Marida	1.18%	818	710	\$704.11	\$323.59
	+ Merida	1.18%	818	710	\$704.11	\$323.59
	+ Mexico West	1.36%	1,427	1,283	\$1,072.58	\$455.24
	- North West	1.23%	17,852	16,345	\$13,730.77	\$6,146.34
	+ Bellingham	1.53%	221	221	\$103.90	\$41.02
	+ Bremerton	1.20%	1,466	1,381	\$1,133.79	\$514.43
	+ Portland	1.17%	2,334	2,029	\$1,746.28	\$804.19
	+ Salem	1.27%	4,218	3,744	\$2,847.76	\$1,256.68
	+ Seattle	1.29%	3,709	3,334	\$2,807.65	\$1,223.44
	+ Spokane	1.03%	1,040	935	\$712.94	\$350.44
	+ Tacoma	1.26%	3,863	3,751	\$3,506.64	\$1,551.69
	+ Walla Walla	.97%	104	104	\$81.16	\$41.18
	+ Yakima	1.18%	897	846	\$790.64	\$363.27

image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Diagrammes, pour l'affichage simple et intuitif des informations,

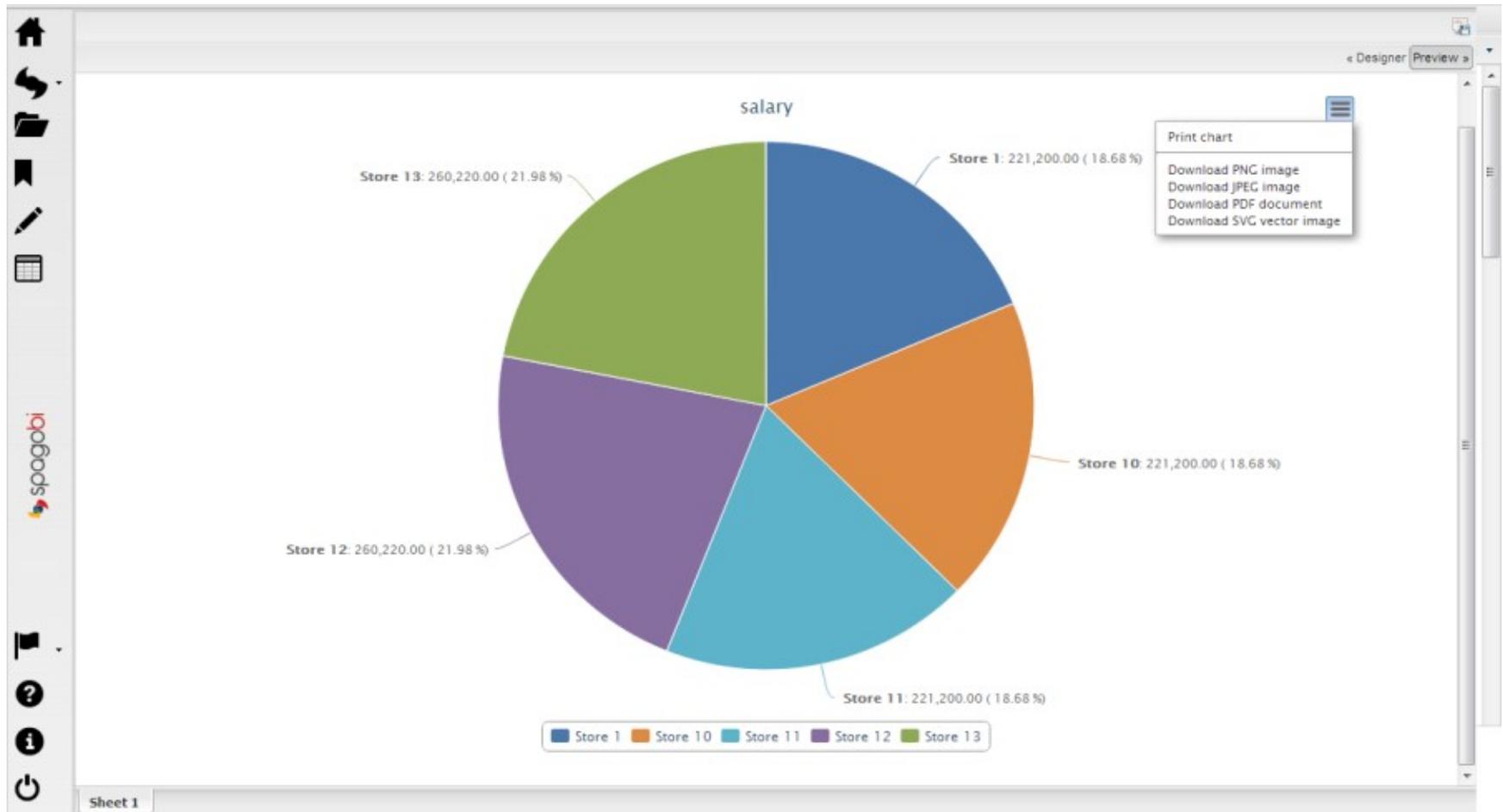


image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Tableaux de bord en temps réel, pour suivre l'évolution des indicateurs clé de performance,

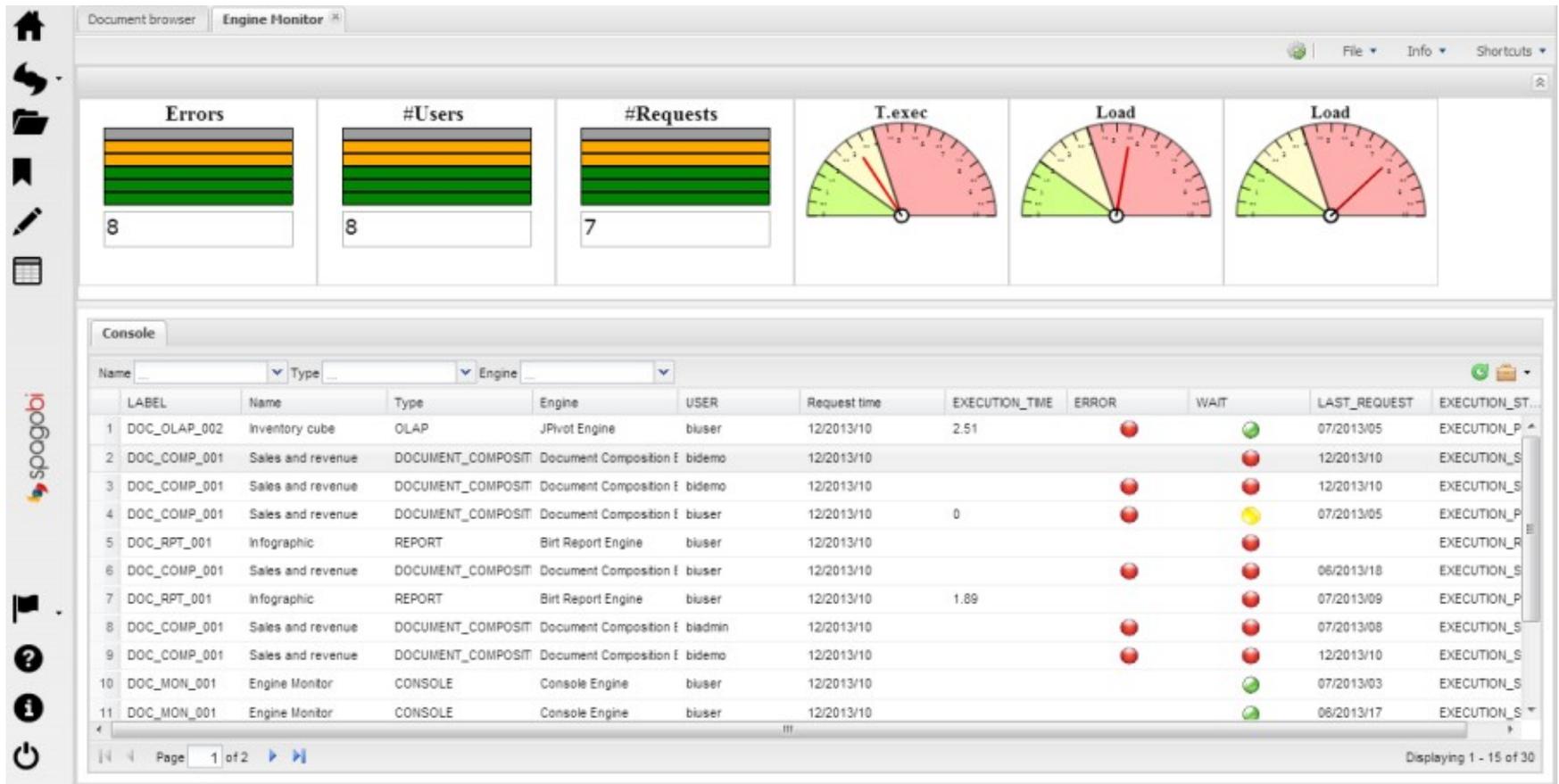


image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Modèles d'Indicateurs clé de performance pour développer et tester des modèles de monitoring des performances,

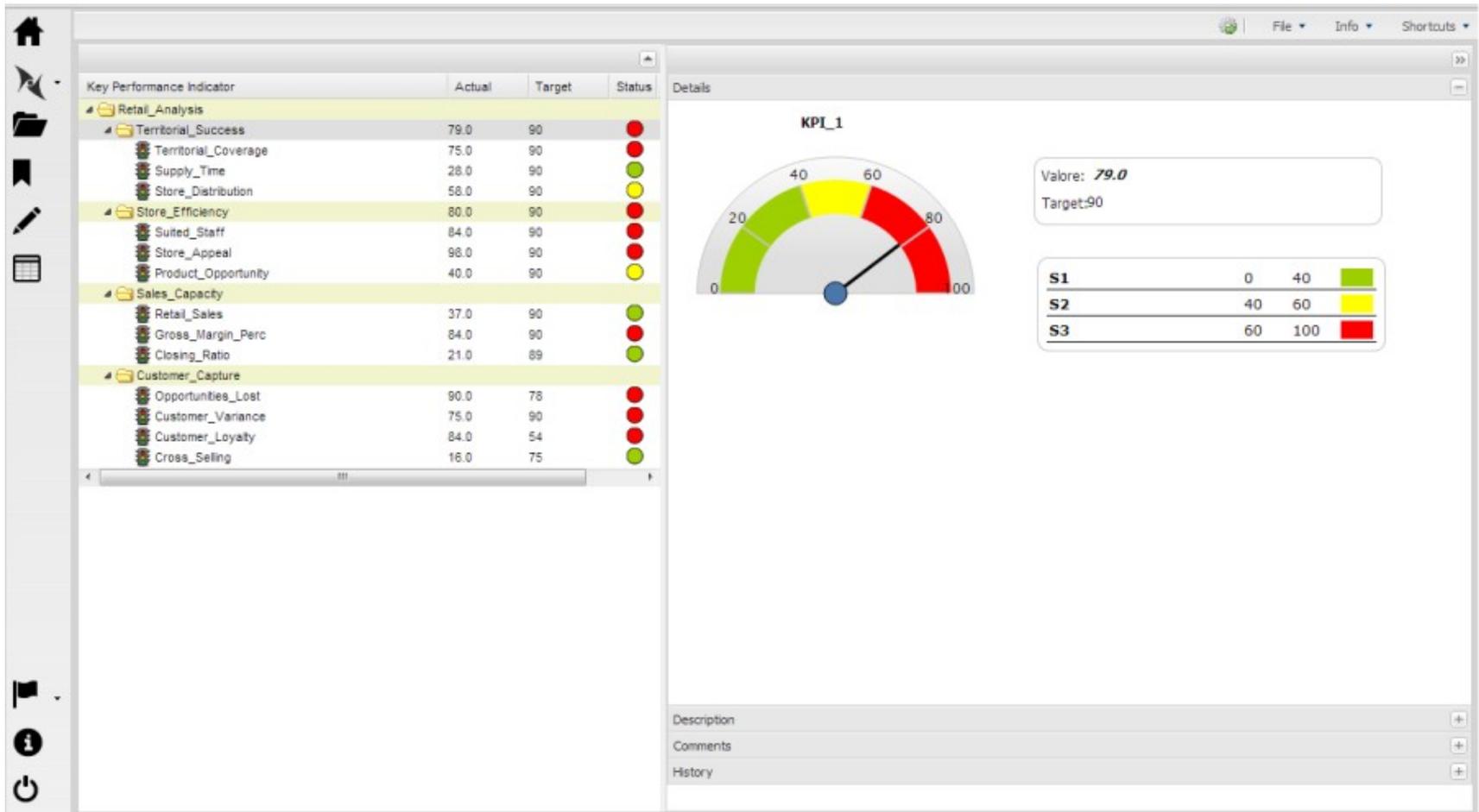


image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Cockpits, pour générer des tableaux de bord structurés et interactifs,

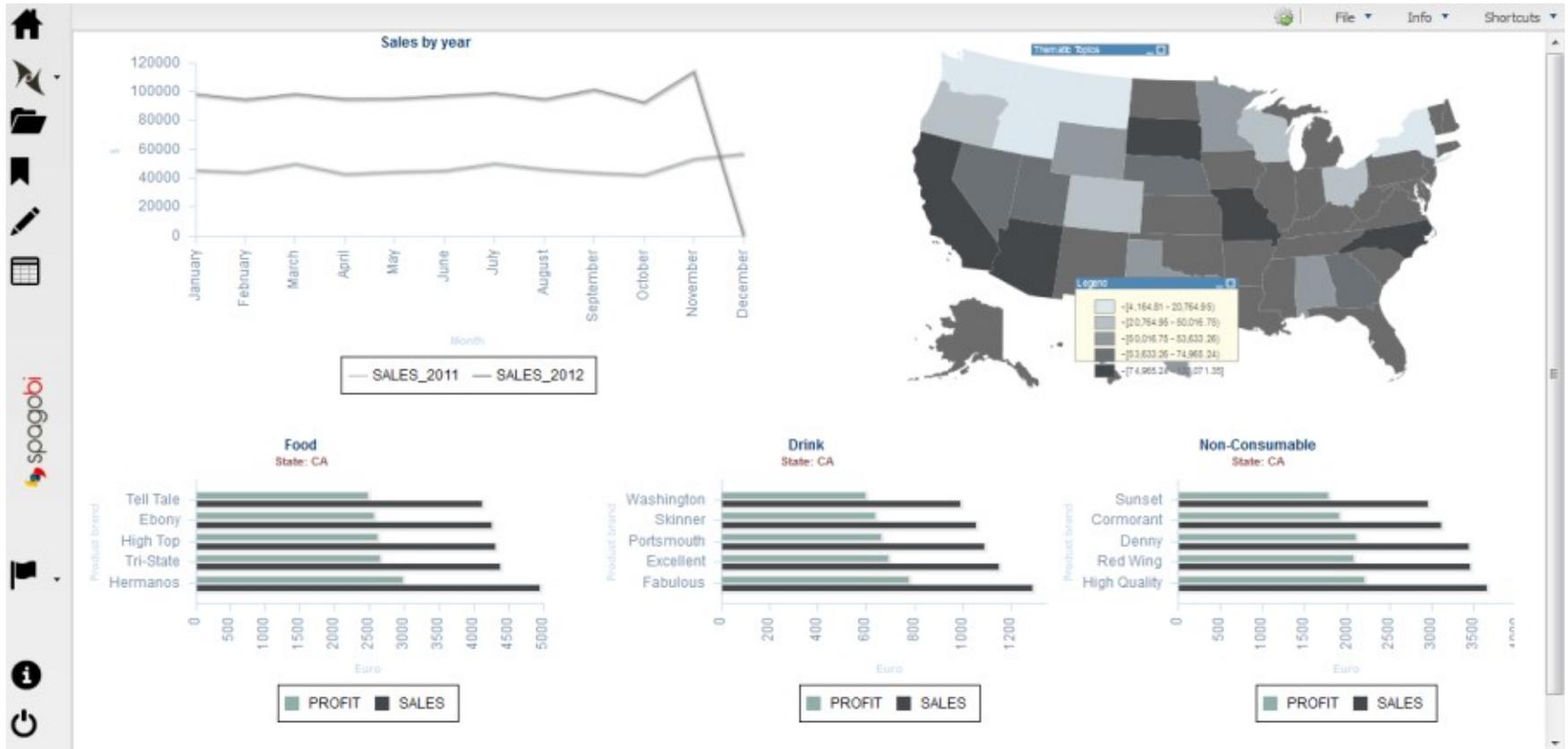
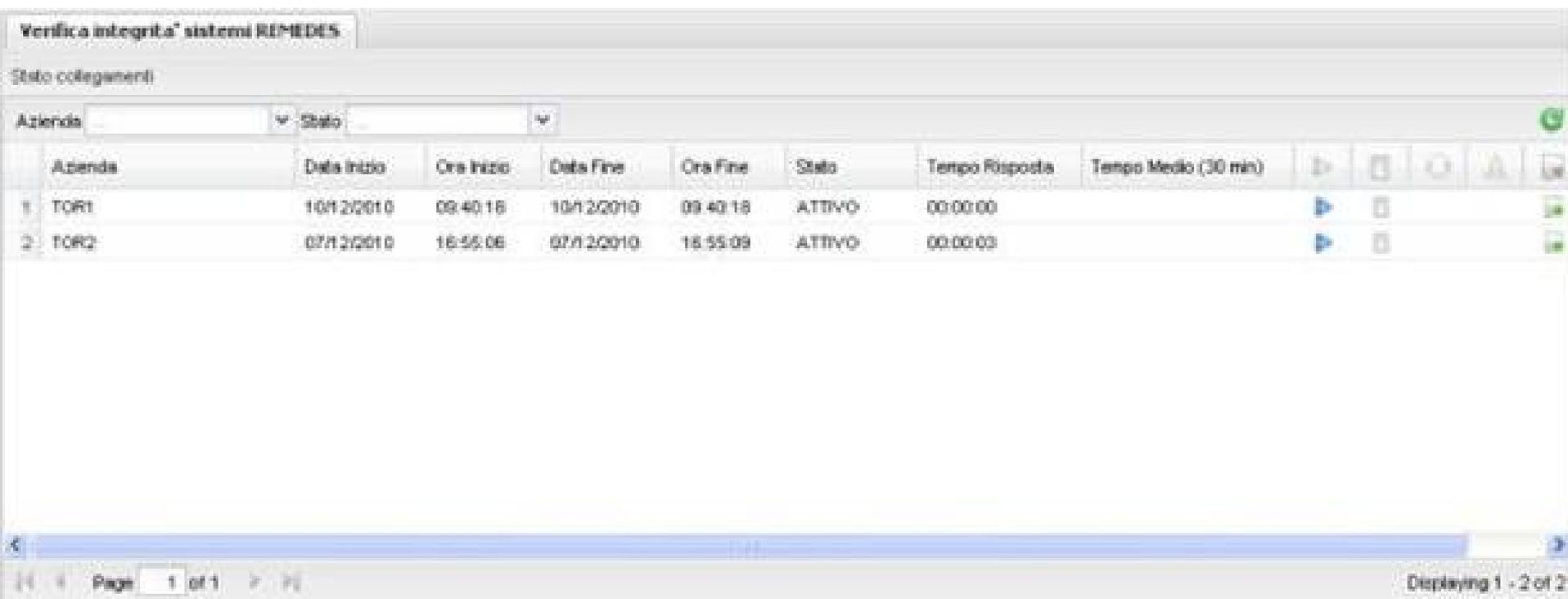


image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

Exécuter les processus externes qui peuvent interagir avec des systèmes de traitement de transactions en ligne (OLTP),



Verifica integrità sistemi REMEDIS

Stato collegamenti

Azienda  Stato

	Azienda	Data Inizio	Ora Inizio	Data Fine	Ora Fine	Stato	Tempo Risposta	Tempo Medio (30 min)				
1	TOR1	10/12/2010	09:40:18	10/12/2010	09:40:18	ATTIVO	00:00:00		▶	⊞	⊞	⊞
2	TOR2	07/12/2010	18:55:08	07/12/2010	18:55:09	ATTIVO	00:00:03		▶	⊞	⊞	⊞

Page 1 of 1

Displaying 1 - 2 of 2

image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Smart Filter, pour la sélection guidée des données

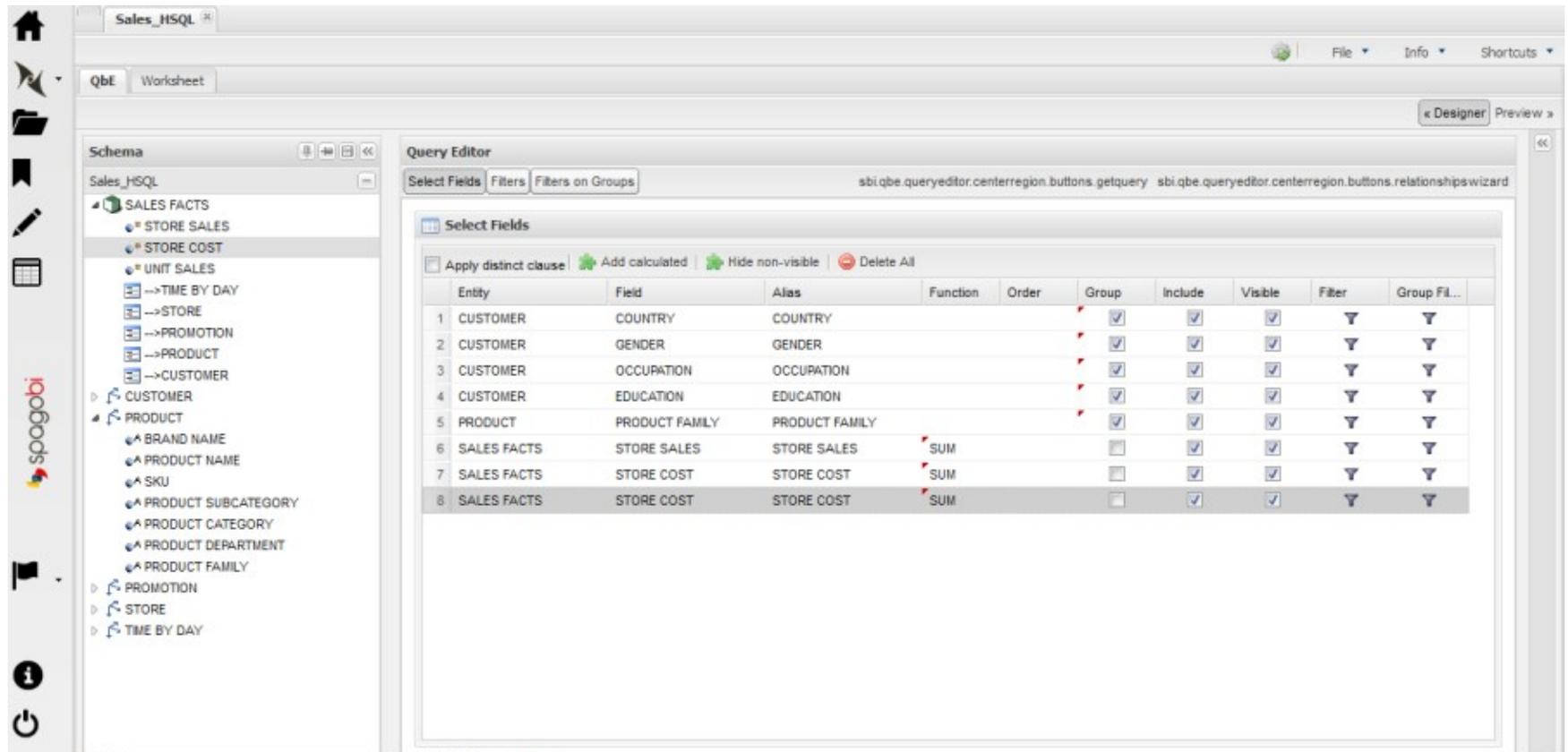


image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Analyse géo-référencée, pour l'affichage des informations sur une base géographique

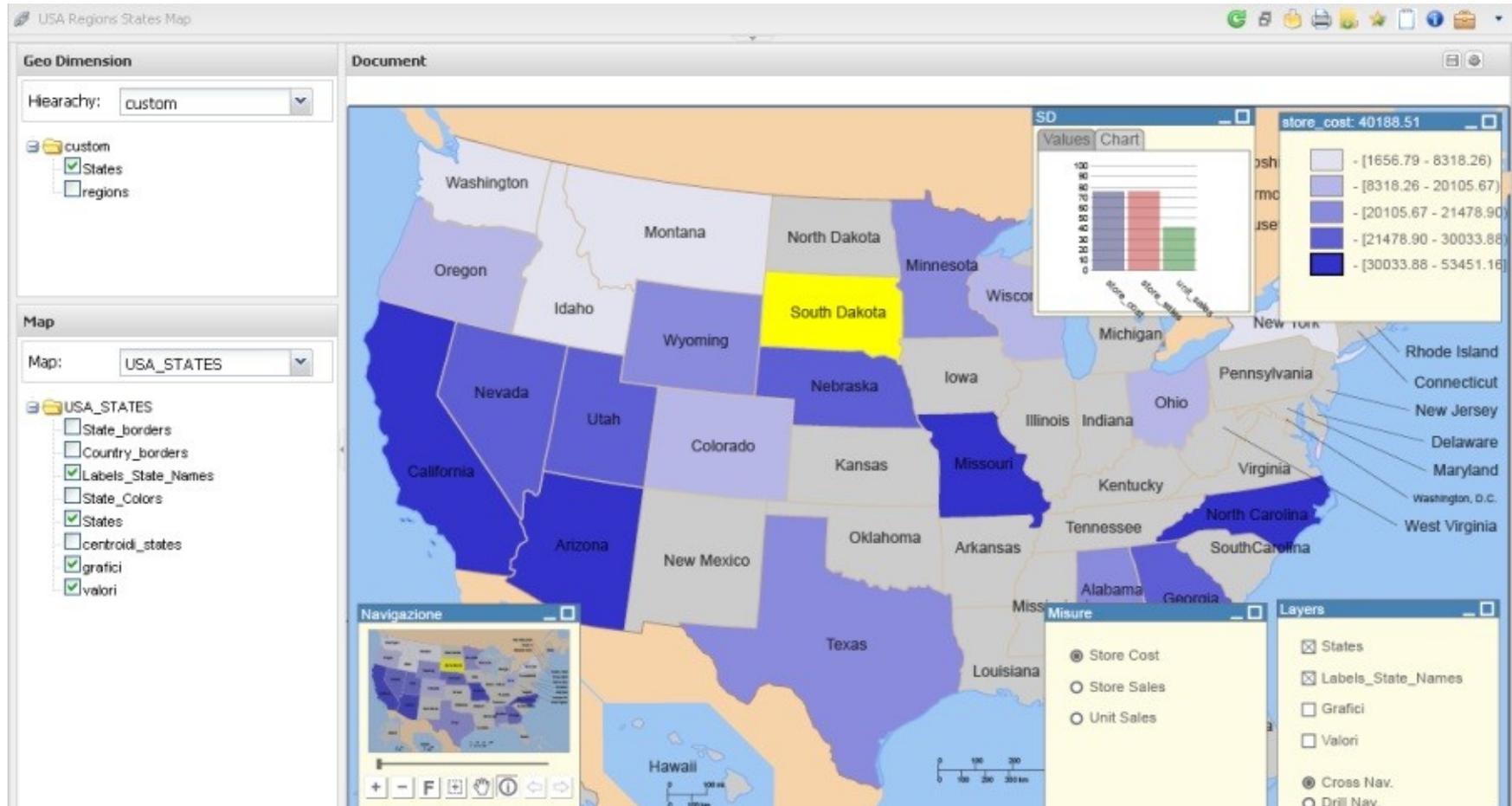


image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Analyse géo-référencée, pour l'affichage des informations sur une base géographique

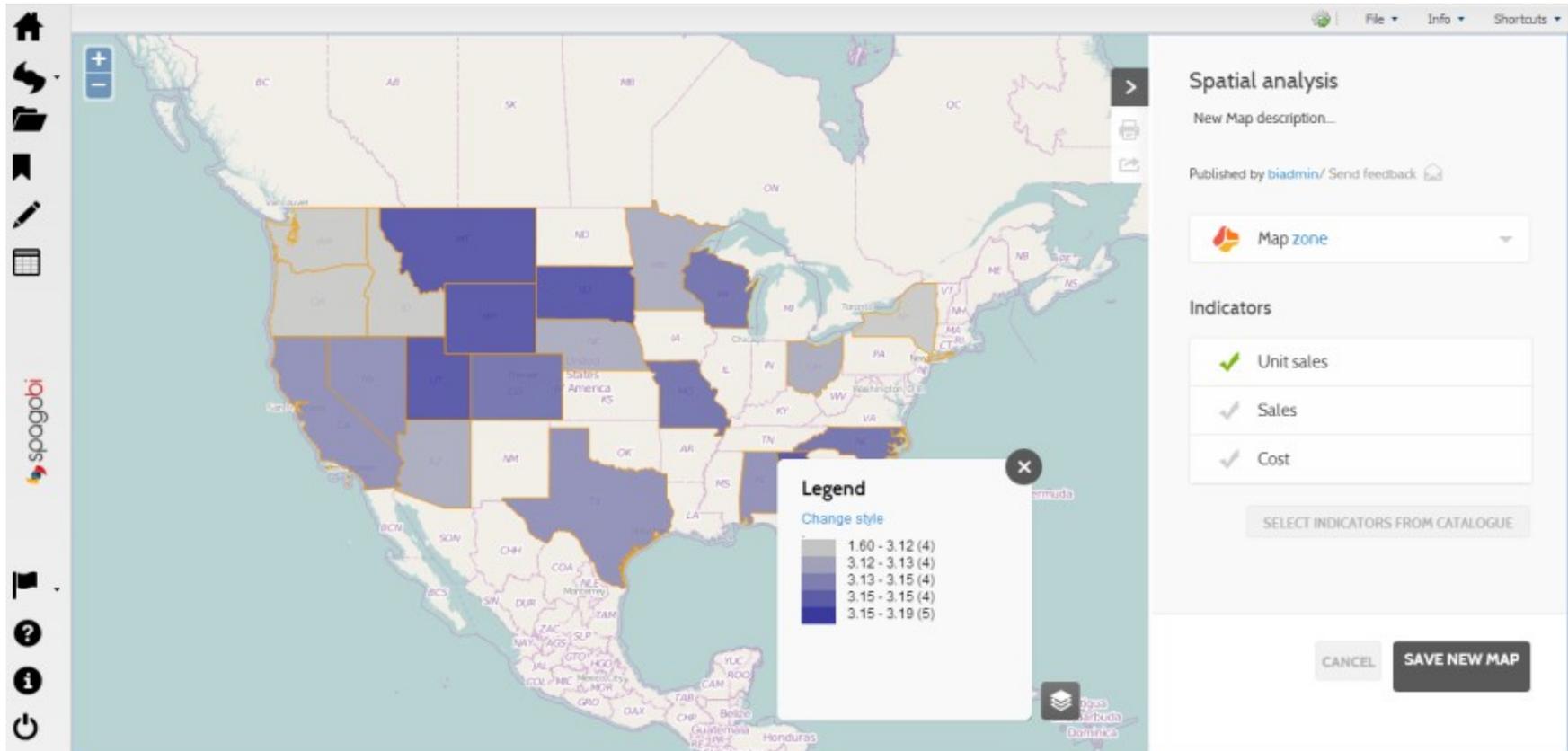


image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Spago ETL

- SpagoBI comprend un outil de chargement et la gestion des données (ETL) dans l'entrepôt de données.
- Le Moteur ETL SpagoBI intègre la suite open source Talend Open Studio TOS de produits open source.

# ETL/ EII, pour collecter les données de différentes sources.

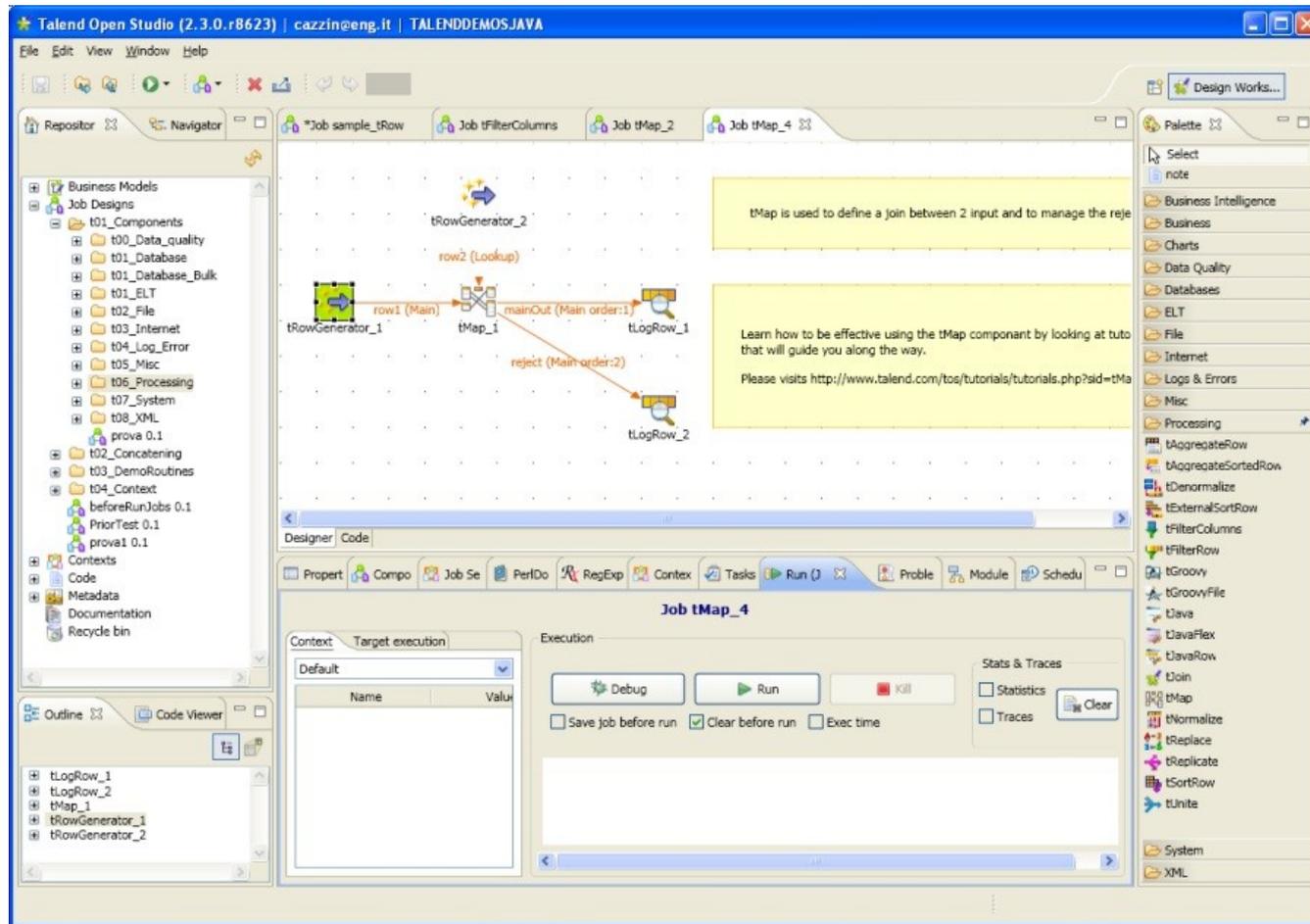


image source : <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/BIComponents>

# Talend Open Studio

- Talend est un éditeur de logiciels open source qui permet :
  - l'intégration des données,
  - la gestion des données,
  - l'intégration d'applications d'entreprise
  - L'intégration des logiciels et les services de données grand.
- 
- L'outil qui nous intéresse de près est : Talend Open Studio for Data Integration: une application open source pour la conception des tâches (jobs) d'intégration de données avec un environnement de développement graphique.

# Talend ETL

The screenshot displays the Talend Studio interface for a job named "Job A\_Top\_Service 0.1". The main workspace shows a job design with the following components and data flow:

- Input:** tRESTRequest\_1 (REST icon) feeds into a sequence of REST requests: get\_timeline\_rows\_in\_124s (Main order:2), get\_timeline\_rows\_in\_124s (Main order:1), get\_emoticons\_rows\_in\_124s (Main order:3), get\_sentiment\_rows\_in\_124s (Main order:4), and get\_tweeters\_timeline (Main order:5).
- Processing:** Each REST request is followed by a tFlowTolterrate component (labeled "Iterate") and a tXMLMap component (labeled "top\_hash\_tags", "top\_emoticons", "top\_sentiments", "top\_tweeters").
- Output:** The XML maps feed into tLogRow components (labeled "row1 (Main)" through "row6 (Main)") and tRESTResponse components (labeled "tRESTResponse\_1" through "tRESTResponse\_6").
- Status:** The job is in a "Starting" state. The tFlowTolterrate\_4 component is currently "1 exec running".

The interface includes a Repository on the left with "Business Models" and "Job Designs" (including "A\_Top\_Service 0.1"). The bottom panel shows the "Job A\_Top\_Service" properties, including a "Basic Run" tab with "Run", "Kill", and "Clear" buttons, and a "Target Exec" tab with a log viewer showing a warning from SLP4J.

Name	Value
Cloudera_Database	cloudera
Cloudera_Port	8889
Cloudera_Login	root
Cloudera_Server	localhost
Cloudera_Password	****
Cloudera_AdditionalP	noDateti

# Pentaho

- Le deuxième outil open source et gratuit le plus populaire est « Pentaho Data Integration ». Pentaho est une entreprise fondée en 2004,
- Il propose une suite de produits open source pour la business intelligence appelés Pentaho Business Analytics pour fournir des services :
- pour l'intégration de données, les services OLAP, reporting, tableaux de bord, l'exploration de données et les outils ETL.

# ETL de Pentaho (kettle pour l'ETL, et spoon pour la GUI)

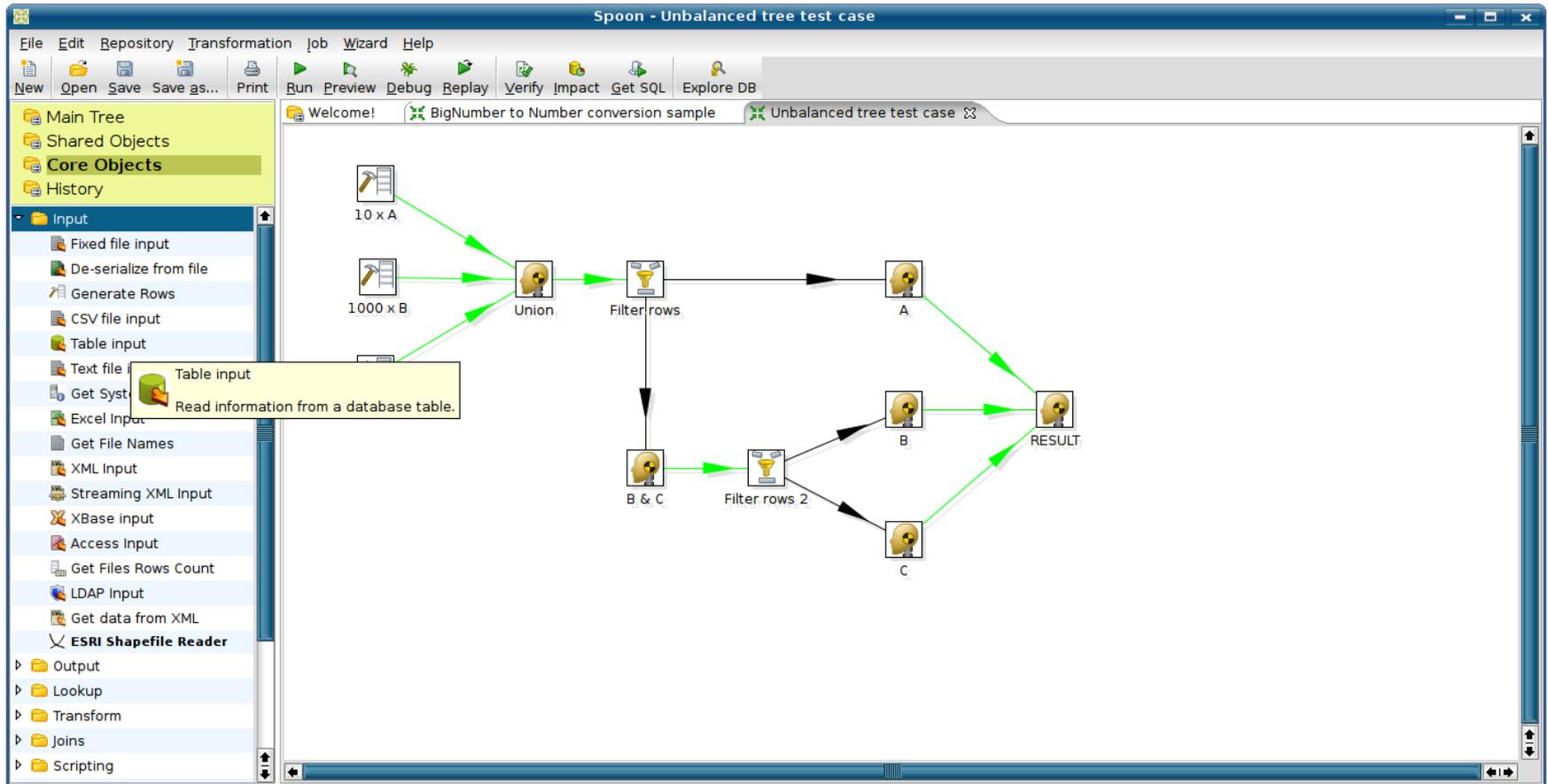
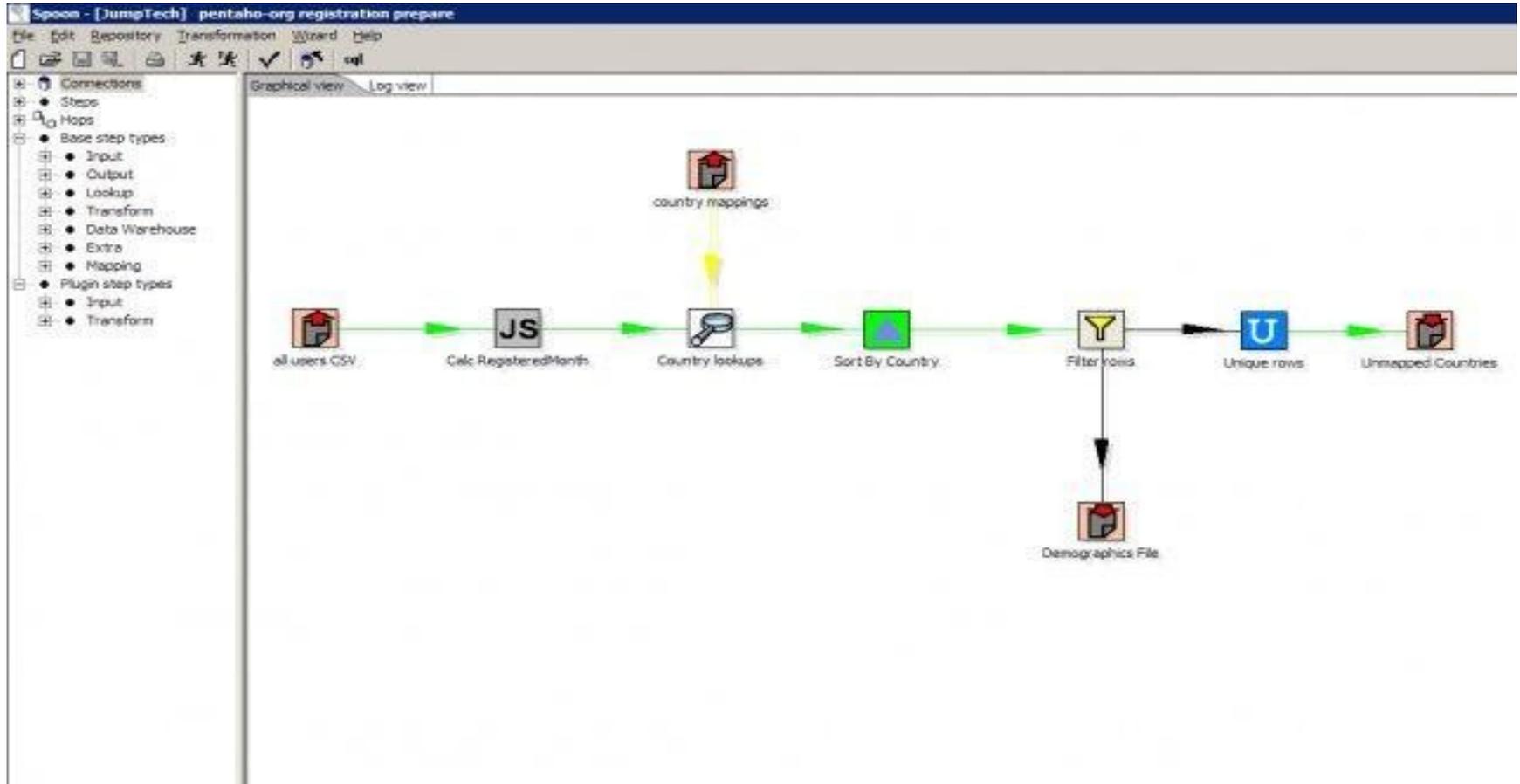


Image source : <http://forums.pentaho.com/showthread.php?59900-Spoon-GUI-changes>

# Pentaho ETL (kettle pour l'ETL, et spoon pour la GUI)



# Outils de BI payants

- Microsoft BI (idéal pour les SGBD microsoft SQL server, avec ses différentes composantes ssrs, ssas, ssis, etc.)
- Sap Business Objects (idéal pour les systèmes utilisant des ERP de SAP),
- Oracle Business Intelligence Enterprise Edition (idéal pour les SGBD oracle),
- IBM cognos (idéal pour les SGBD db2),

# Futur de la business intelligence

- **Big Data :**
  - Analyser **d'énormes** quantités de données,
  - À partir de sources structurées :
    - Base de données, systèmes opérationnels, Servers, etc.
  - À partir de sources non structurées :
    - Blogs, Réseaux sociaux, sites Web de tous genres, etc.
- **Data Visualization :**
  - Visualisation de données de façon à faire apparaître l'information rapidement et la rendre compréhensible par les utilisateurs.

Merci pour votre attention

Questions ?