

Pentaho :

Comparatif fonctionnel entre la version Communautaire (gratuite) et la version Entreprise (payante)



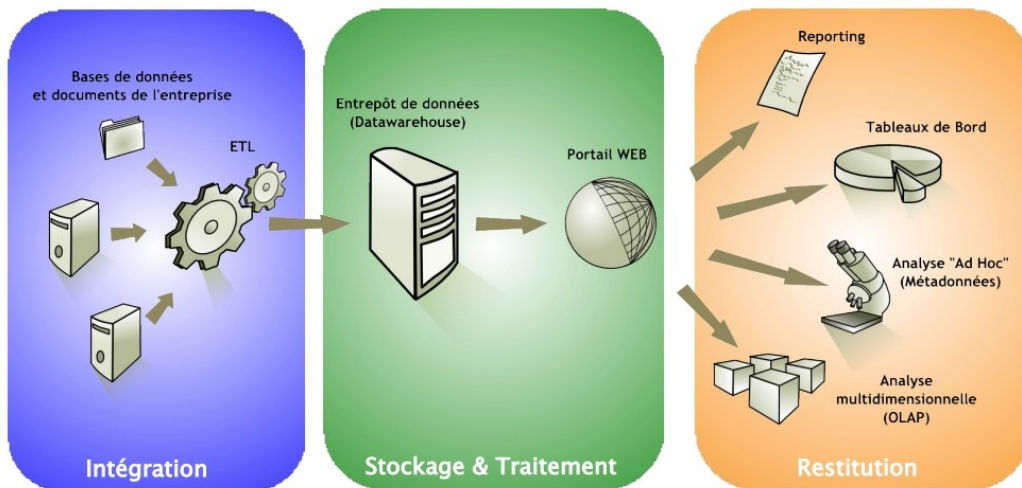
Table des matières

1	PRÉSENTATION DE PENTAHO.....	2
2	LISTING DES COMPOSANTS DE LA PLATE-FORME.....	4
2.1	Composants pour les Utilisateurs Métiers.....	4
2.2	Composants pour les utilisateurs Techniques (Service informatique).....	4
3	SYNTHÈSE DES DIFFÉRENCES ENTRE LES VERSIONS CE ET EE (PENTAHO 3.6).....	5
4	DÉTAIL DES DIFFÉRENCES ENTRE LES VERSION CE ET EE (PAR COMPOSANT).....	6
4.1	Console Utilisateur – Fonction « Tableaux de Bords ».....	6
4.1.1	Tableaux de bords en version Entreprise.....	6
4.1.2	Tableaux de bords en version communautaire	9
4.2	Console Utilisateur – Fonction « OLAP ».....	13
4.2.1	Fonctionnalités OLAP avec la version Entreprise.....	13
4.2.2	Fonctionnalités OLAP avec la version Communautaire	16
4.3	Console d'administration.....	18
4.3.1	La Console d'administration en version Entreprise.....	18
4.3.2	La Console d'administration en version Communautaire.....	19
4.4	Les apports de Pentaho Data Integration en version Entreprise.....	21
4.4.1	Introduction.....	21
4.4.2	Caractéristiques de PDI EE.....	21

1 PRÉSENTATION DE PENTAHO

Pentaho est une plate-forme décisionnelle open source complète possédant les caractéristiques suivantes :

- Une couverture complète des fonctionnalités de la Business Intelligence :
 - ETL (intégration de données)
 - Reporting
 - Tableaux de bords ("*Dashboards*")
 - Analyse ad-hoc (requêtes à la demande)
 - Analyse multidimensionnelle (OLAP)
 - Data Mining



- Pentaho permet de fournir des ressources décisionnelles à 2 types d'utilisateurs :

- Des utilisateurs de base qui vont consommer des indicateurs prédéfinis. Quelques clics vont suffire pour obtenir un état et ceci de façon intuitive.
- Des utilisateurs avancés, qui vont disposer d'outils d'analyse et d'exploration avancés afin de construire et d'exploiter leurs propres états.



➤ Reporting :

- Rapports & états préformatés (pdf, excel, html)
- Filtrage possible des données ("*row level*")

➤ Tableaux de bords :

- Présentation graphique et synthétique d'indicateurs
- Permettent l'obtention intuitive de rapports ("*drill-down*")

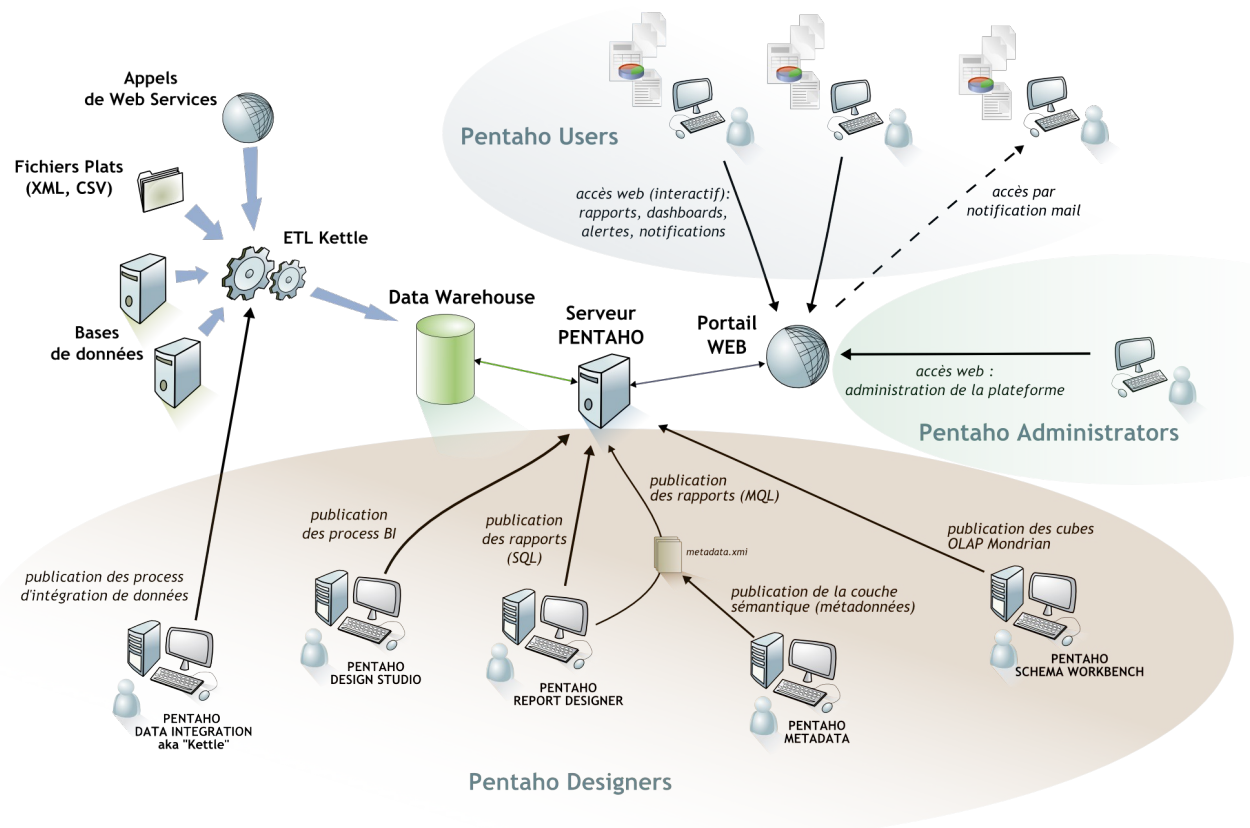
➤ Analyse Ad-hoc :

- Permet de créer des rapports avancés (choix des colonnes, tableaux croisés) à partir de vues métiers ("*business views*"). Nécessite la mise en place d'une couche sémantique d'abstraction (métadonnées)

➤ Analyse multidimensionnelle (OLAP)

- Permet la manipulation de données selon plusieurs axes d'analyse. Nécessite une modélisation spécifique dans le SGBDR (ROLAP)

- Pentaho s'appuie sur une architecture NTIC [Web 2.0](#) qui se compose :
 - D'un serveur web J2EE permettant d'accéder à l'ensemble des ressources décisionnelles et ceci au travers d'URLs web uniques et standardisées.
Le serveur est dénommé "**Pentaho User Console**" (PUC)
 - De plusieurs clients riches permettant la conception et la publication des ressources. Ces derniers sont librement téléchargeables et peuvent être installés sous des environnements Windows, Linux ou Mac-OS (clients Java):



- Le serveur web Pentaho comporte également une plate-forme d'administration (**Pentaho Administration Console**) permettant la gestion des droits d'accès, la planification d'évènements, la gestion centralisée des sources de données...
- Pentaho est reconnue pour être une solution d'une grande qualité conceptuelle et technique. La plate-forme est orientée « processus », ainsi au travers de « séquences d'actions » on ainsi peut modéliser avec Pentaho des workflows BI avancés.
- Il n'est pas utile de connaître le langage JAVA pour travailler avec Pentaho: seule la maîtrise du langage SQL est nécessaire, ainsi que des connaissances de base en XML, HTML et Javascript.
- Une [communauté](#) importante et très active s'anime autour de Pentaho. Celle-ci contribue au codage de nombreux plugins et projets communautaires .
- Pentaho est une suite décisionnelle open source commerciale qui est très ouverte et dont les différences fonctionnelles entre la version libre gratuite (*Community Edition*) et la version payante (*Enterprise Edition*) restent limitées. Ainsi la version libre de Pentaho permet d'installer une plate-forme décisionnelle complète.

2 LISTING DES COMPOSANTS DE LA PLATE-FORME

2.1 COMPOSANTS POUR LES UTILISATEURS MÉTIERS

Les utilisateurs métiers ont seulement besoin d'un browser web pour accéder aux ressources décisionnelles :

Composant	Code	Fonction
Console Utilisateur (accès via un navigateur web)	PUC	Interface web qui permet à un utilisateur métier d'accéder à l'ensemble des ressources décisionnelles : rapports, tableaux de bords, indicateurs, alertes, analyses multidimensionnelles, analyses Ad-Hoc











2.2 COMPOSANTS POUR LES UTILISATEURS TECHNIQUES (SERVICE INFORMATIQUE)

Les utilisateurs techniques (les concepteurs de ressources BI) installent des outils clients sur leur PC :





Composant	Code	Fonction
Console d'administration (accès via un navigateur web)	PAC	Interface web permettant la gestion des droits d'accès, la planification d'évènements (envoi de rapports par mail en automatique par exemple), la gestion centralisée des sources de données.
Pentaho Data Integration (Kettle)	PDI	ETL (Extract Transform Load) complet, pouvant être utilisé complètement indépendamment de la plate-forme Pentaho. Kettle est comparable à Talend Open Studio en terme de fonctionnalités. Un livre blanc rédigé par Atol CD détaille les caractéristiques et différences entre ces 2 ETL.
Pentaho Report Designer	PRD	Client de conception de rapports avancés. Il s'agit d'un outil de mise en page similaire à iReport, Eclipse BIRT, Crystal Reports... Permet de se connecter à de nombreuses sources de données: SGBD, XML, Excel, CSV, flux de données venant de Kettle, MDX (OLAP), Métadonnées Pentaho. Gestion native des formats de sortie HTML, PDF, EXCEL, WORD
Pentaho Design Studio	PDS	Client riche (Eclipse) de modélisation de workflows BI (Xactions) propre à Pentaho. Design Studio permet de mettre en œuvre de nombreuses ressources BI en minimisant l'écriture de code (envoi de mails automatisé par ex.)
Pentaho Metadata	PME	Client riche permettant la mise en place d'une couche sémantique d'abstraction (méta-données) sur la couche physique (tables et colonnes d'une base de données). Le but est de rendre les objets d'un SGBD compréhensibles et manipulables par un utilisateur final afin de lui permettre d'effectuer ses propres requêtes et ceci sans connaître le langage SQL. La couche de méta-données peut-être utilisée dans le requêteur web ad-hoc, dans Pentaho Report Designer et dans Pentaho Design Studio
Pentaho Schema Workbench	PSW	Client riche permettant la définition des schémas Mondrian à partir d'un modèle en étoile ou flocon de l'entrepôt de données
Pentaho Aggregation Designer	PAG	Client riche permettant l'optimisation des schémas de cube conçus avec Pentaho Schema Workbench (construction et chargement des tables d'agrégation en vue d'améliorer les performances)

3 SYNTHÈSE DES DIFFÉRENCES ENTRE LES VERSIONS CE ET EE (PENTAHO 3.6)

Le tableau ci-dessous présente une grille comparative entre les composants Pentaho de la version gratuite (CE=Community Edition) et ceux de la version payante (EE=Enterprise Edition) :

Composant	Code	Ecart fonctionnel CE vs EE	
Console Utilisateur	PUC	Reporting	
		Tableaux de Bord	
		Requêtage Ad-Hoc (sur métadonnées)	
		OLAP	
Console d'administration	PAC		
Pentaho Data Integration	PDI		
Pentaho Report Designer	PRD		
Pentaho Design Studio	PDS		
Pentaho Metadata	PME		
Pentaho Schema Workbench et Pentaho Aggregation Designer	PSW		

Légende :

	Aucune différence entre les 2 versions
	Ecart faible : les différences entre la version gratuite et payante se situent au niveau du confort et de la facilité des paramétrage. Pas de réel écart fonctionnel.
	Ecart moyen : L'écart fonctionnel se traduit par une amélioration de la productivité ou par des fonctionnalités complémentaires pour les utilisateurs finaux. Ce niveau d'écart peut malgré tout être contourné par l'utilisation de méthodes alternatives proposées dans la version CE ou via des projets communautaires tiers.
	Ecart important : La fonctionnalité proposée dans la version payante apporte un réel gain fonctionnel aux utilisateurs finaux. Aucun contournement possible avec la version CE.

4 DÉTAIL DES DIFFÉRENCES ENTRE LES VERSION CE ET EE (PAR COMPOSANT)

4.1 CONSOLE UTILISATEUR – FONCTION « TABLEAUX DE BORDS »

4.1.1 Tableaux de bords en version Enterprise

La version Enterprise permet à un utilisateur métier de construire ses propres tableaux de bords via la console d'utilisation (PUC) :

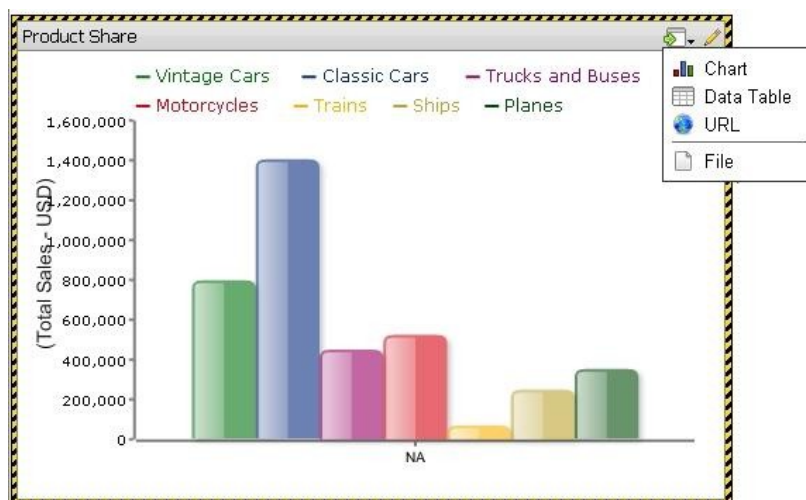


Il est ainsi possible :

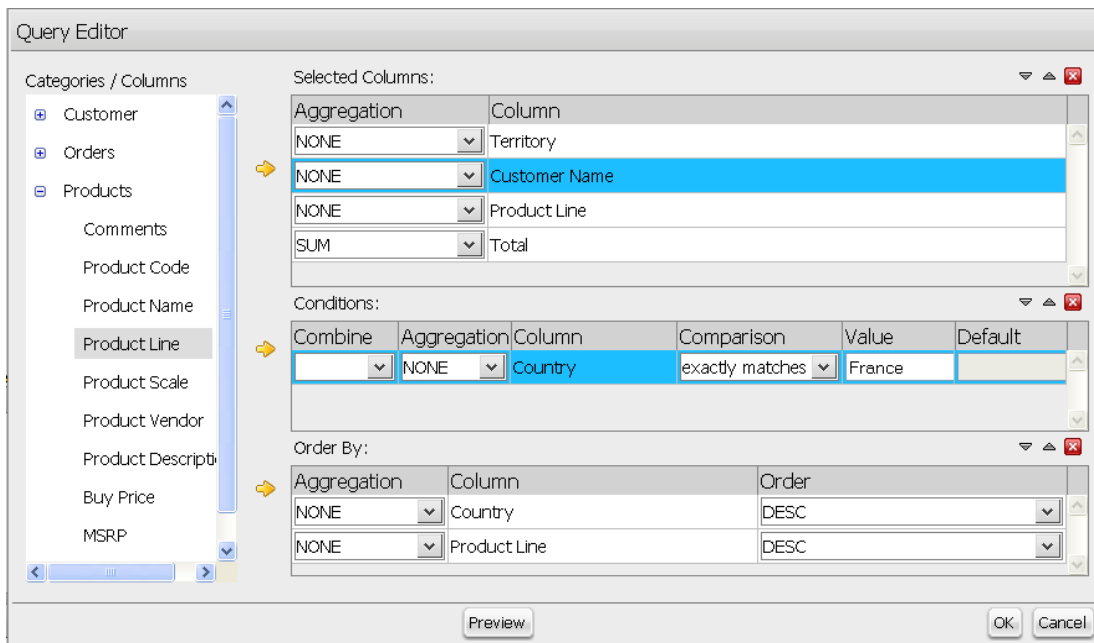
- ➔ de spécifier un modèle de mise en page (nombre et positionnement des fenêtres) :



- ➔ d'indiquer pour chaque fenêtre le type de ressource que l'on veut présenter :
 - graphique
 - liste tabulaire de données
 - lien url ou une ressource existante (rapport, analyse OLAP...)



- ➔ de récupérer des données à partir du requêteur MQL qui s'appuie le référentiel de métadonnées de Pentaho, ce qui permet à l'utilisateur métier d'effectuer une requête sans connaissance du langage SQL :



Query Editor

Categories / Columns

- Customer
- Orders
- Products
 - Comments
 - Product Code
 - Product Name
 - Product Line
 - Product Scale
 - Product Vendor
 - Product Descripti
 - Buy Price
 - MSRP

Selected Columns:

Aggregation	Column
NONE	Territory
NONE	Customer Name
NONE	Product Line
SUM	Total

Conditions:

Combine	Aggregation	Column	Comparison	Value	Default
	NONE	Country	exactly matches	France	

Order By:

Aggregation	Column	Order
NONE	Country	DESC
NONE	Product Line	DESC

Buttons: Preview, OK, Cancel

- ➔ d'utiliser un assistant de création de graphiques simple et convivial :

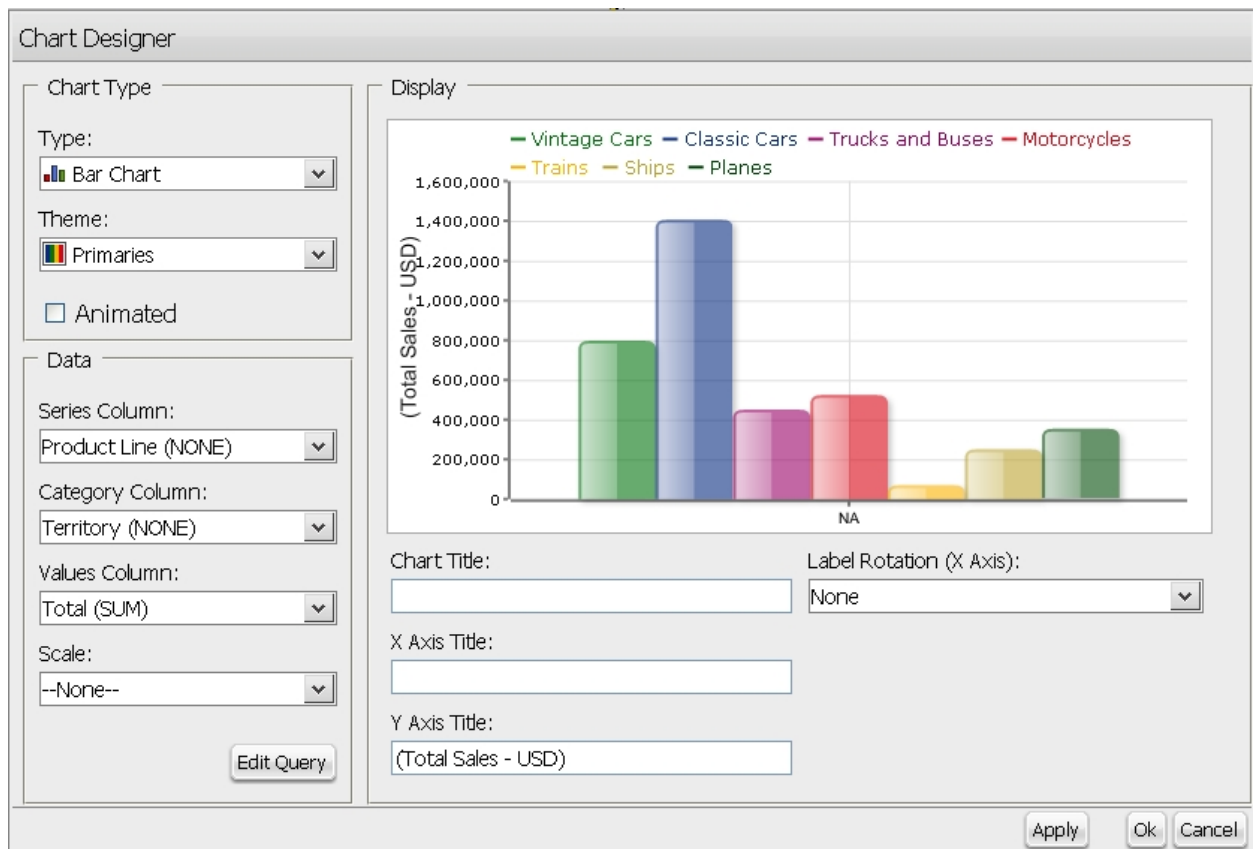


Chart Designer

Chart Type

Type: Bar Chart

Theme: Primaries

Animated

Data

Series Column: Product Line (NONE)

Category Column: Territory (NONE)

Values Column: Total (SUM)

Scale: --None--

Edit Query

Display

Legend: Vintage Cars, Classic Cars, Trucks and Buses, Motorcycles, Trains, Ships, Planes

Y Axis: (Total Sales - USD)

X Axis: NA

Chart Title: []

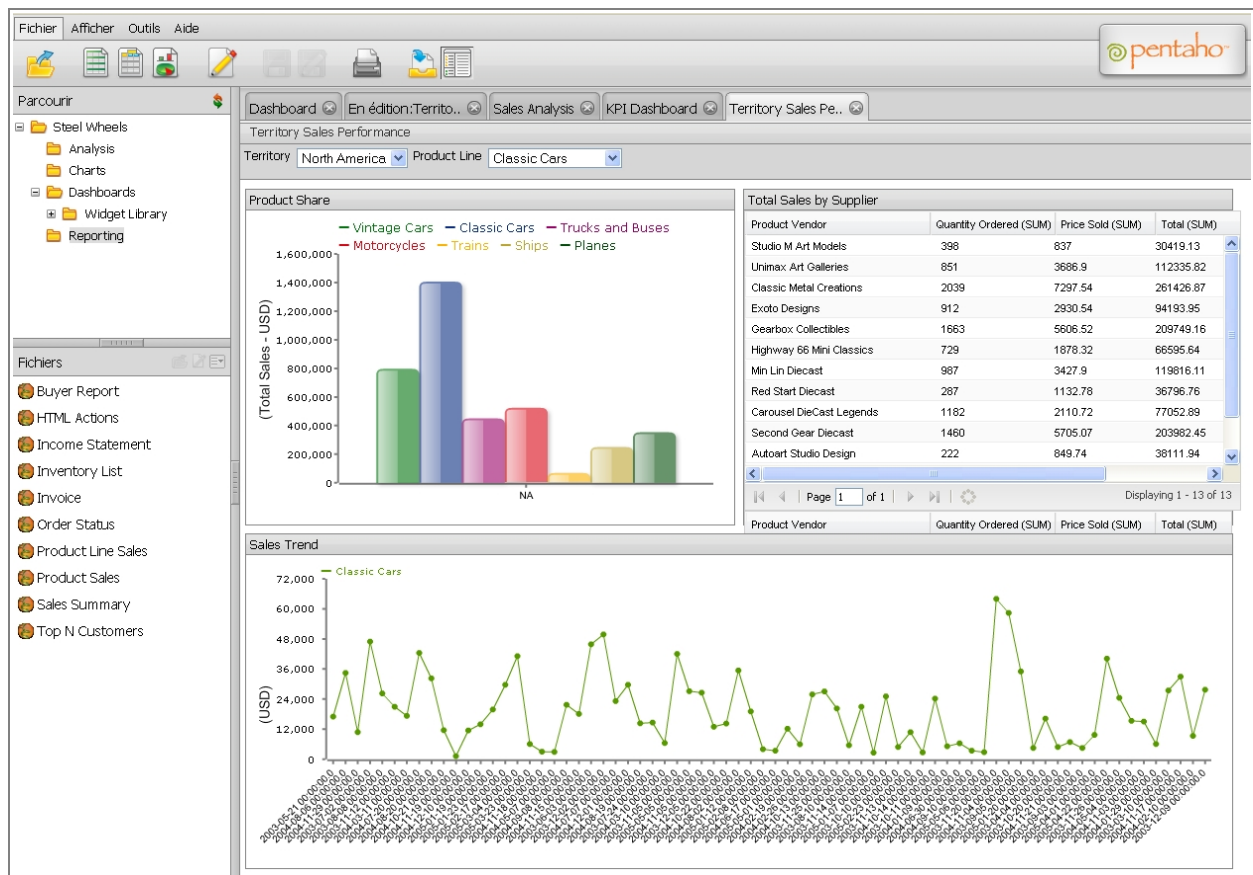
Label Rotation (X Axis): None

X Axis Title: []

Y Axis Title: (Total Sales - USD)

Buttons: Apply, OK, Cancel

+ de filtrer les données d'une ou plusieurs fenêtres en définissant des paramètres :



Une vidéo de démonstration (commentée) de création d'un tableau de bord avec la version Enterprise est disponible sur la chaîne **YouTube opensourcebi** à cette url :

<http://www.youtube.com/watch?v=XvVxjgxLAYc>

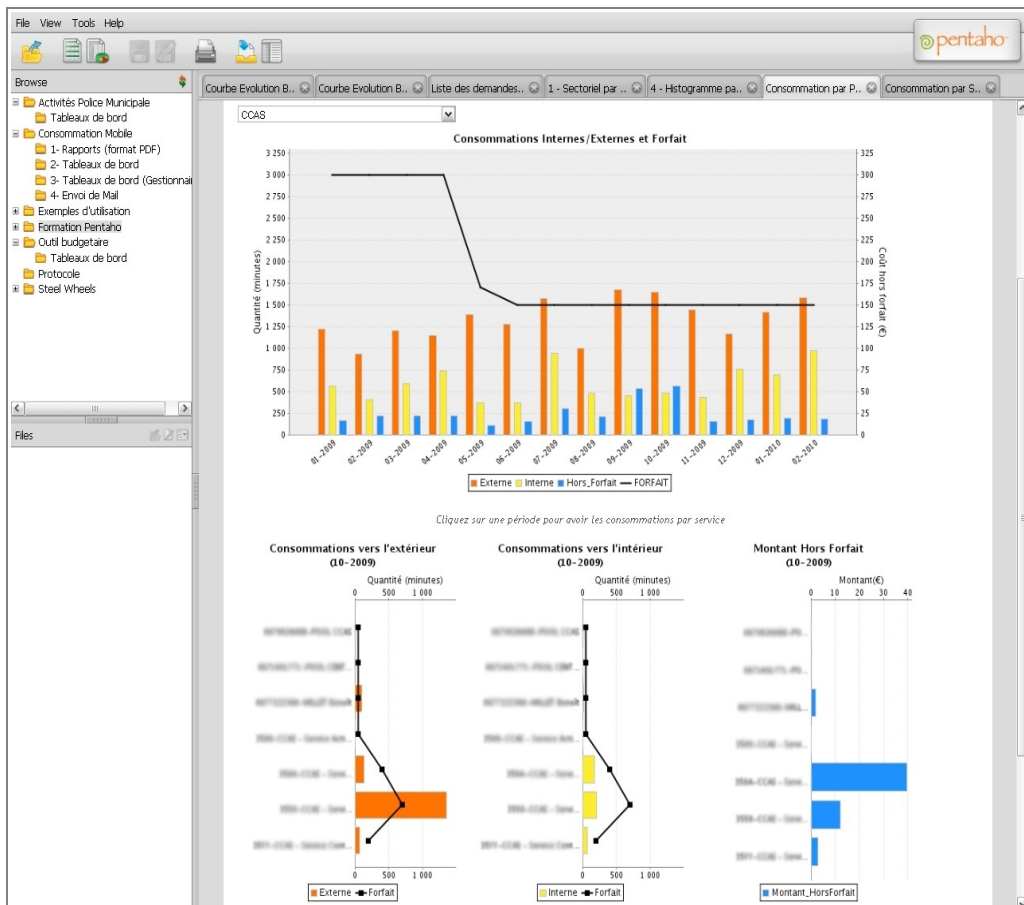
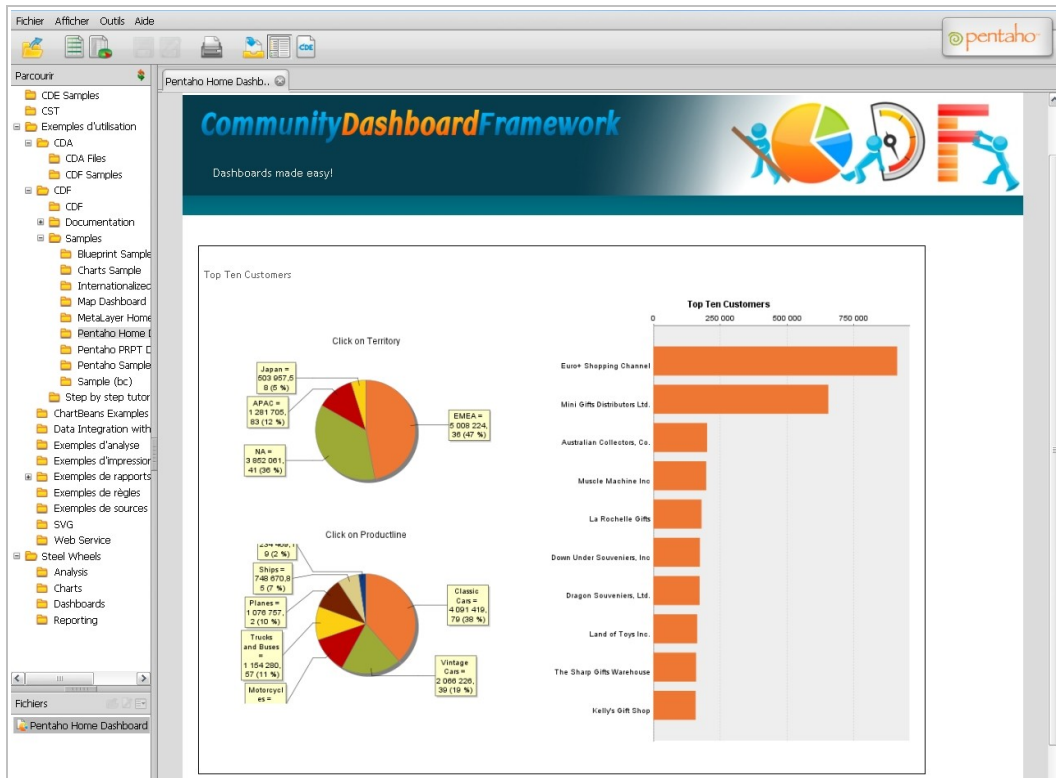
4.1.2 Tableaux de bords en version communautaire

Depuis la version 3.0 de Pentaho, le framework open source « **Pentaho Community Dashboard Framework** » (**Pentaho CDF**) permet la création et la publication de tableaux de bords avancés.

Il faut noter que contrairement à la version EE, ce framework s'adresse essentiellement à des utilisateurs techniques (service informatique) : il faut en effet posséder quelques notions de base en HTML et Javascript pour la conception des tableaux de bords, ce qui n'est pas le cas dans la version Enterprise.

Basé sur une API Javascript/Ajax très complète, **Pentaho CDF** permet au final la mise en place de tableaux de bords pouvant être très riches : intégration de cartes (GoogleMaps, OpenLayers), graphiques en Flash, rapports embarqués, listes tabulaires paginées, composants HTML (listes déroulantes, bouton radio, ...), etc.

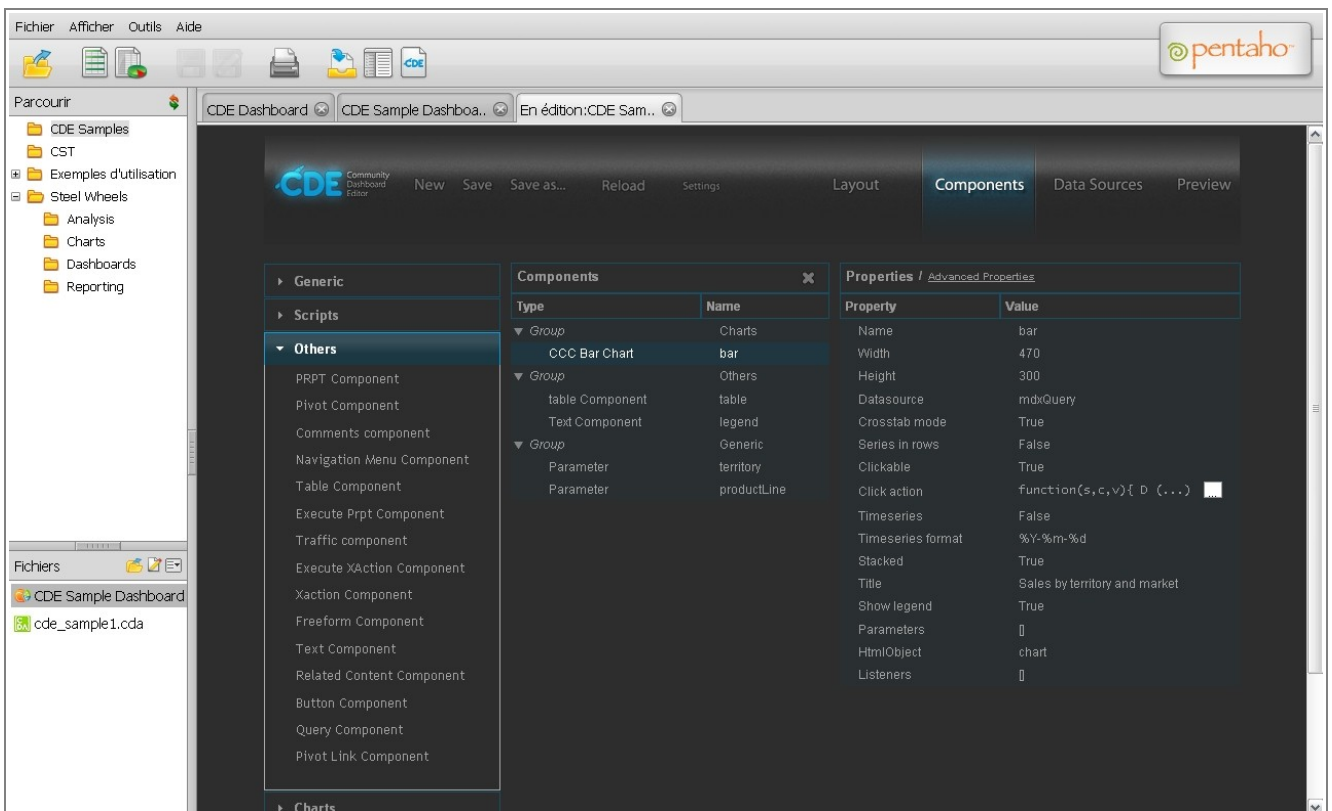
Quelques exemples :



Depuis Pentaho 3.6, un designer de Dashboard (**Pentaho CDE**) simplifie d'ailleurs de façon très sensible le travail de création des utilisateurs techniques via un assistant de création graphique web :



Aperçu de l'interface web de conception :



4.2 CONSOLE UTILISATEUR – FONCTION « OLAP »

4.2.1 Fonctionnalités OLAP avec la version Enterprise

La version Enterprise permet à un utilisateur métier de construire ses propres analyses multidimensionnelles via la console d'utilisation (PUC) :



L'outil qui permet la réalisation des analyses multidimensionnelles se nomme **Pentaho Analyzer**

Une vidéo de démonstration de Pentaho Analyzer sur un jeu de données UNEDIC est disponible sur la chaîne **YouTube opensourcebi** à cette url :

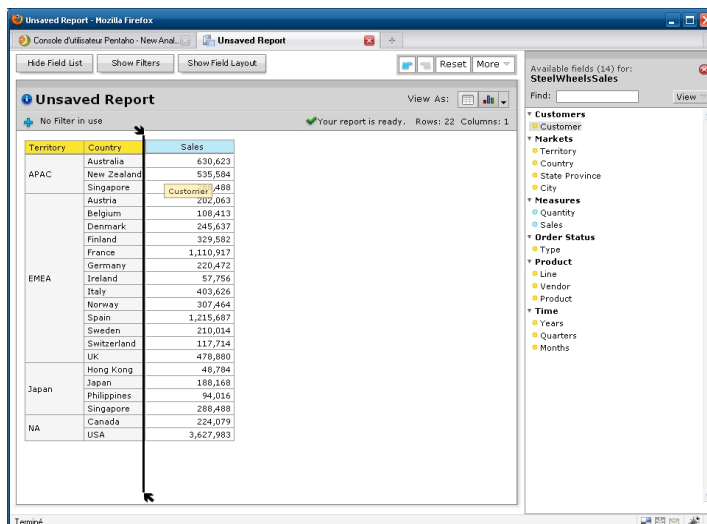
<http://www.youtube.com/watch?v=EBuug7Z6hxw>

Pentaho Analyzer se caractérise notamment par une **excellente ergonomie** entièrement basée sur le drag & drop (glisser/déposer).

De plus, certaines fonctionnalités de Pentaho Analyzer n'existent pas dans la version communautaire avec **JPivot**, comme par exemple le rajout à la volée de membres calculés, l'affichage de tendances, une gestion simplifiée des graphiques...

La création d'états à partir de Pentaho Analyzer est donc puissante et intuitive, ci-dessous quelques fonctionnalités :

- Construction d'états par drag & drop des niveaux des hiérarchies et des mesures depuis la zone « Field List » à droite de l'écran. Possibilité de créer des états tabulaires ou croisés



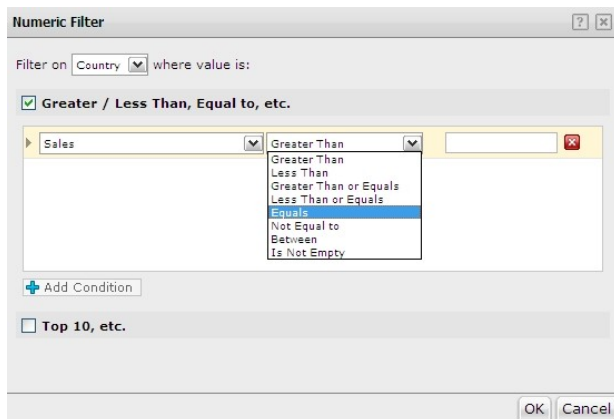
- Possibilité d'ajouter des filtres sur les membres d'un niveau d'une dimension :



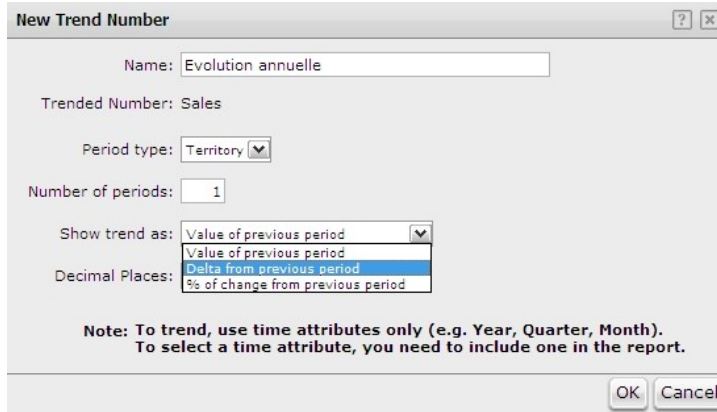
- Les données peuvent être également filtrées directement depuis le tableau grâce aux menus contextuels des en-têtes de colonnes.

Territory	Country	Sales
APAC	Australia	630,623
	New Zealand	535,584
	Singapore	288,488
EMEA	Austria	202,063
	Belgium	108,413
	Denmark	245,637
	Finland	329,582
	Spain	1,215,687
	Sweden	210,014
Japan	Hong Kong	48,784
	Japan	188,168
	Philippines	94,016
NA	Canada	224,079
	USA	3,627,983

- De nombreux filtres numériques sont disponibles : Top X, Bottom X, Plus petit que, plus grand que, non vide...



- Il est possible de paramétrer de façon très avancée les modalités de calcul et la présentation des mesures dans le tableau croisé: ranking, répartition en %, somme cumulées, sous-totaux pour chaque niveaux, tris, tendances sur des intervalles de temps (écart en valeur ou pourcentage de la période comparée), ajout de membres calculés,



New Trend Number

Name: Evolution annuelle

Trended Number: Sales

Period type: Territory

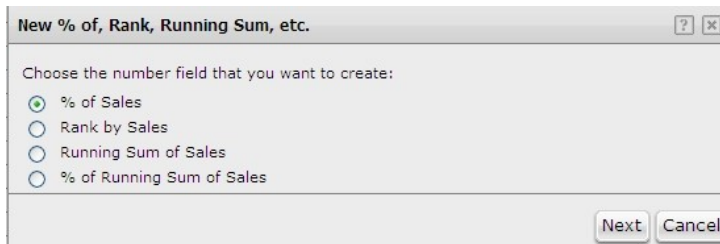
Number of periods: 1

Show trend as: Value of previous period

Decimal Places: Delta from previous period

Note: To trend, use time attributes only (e.g. Year, Quarter, Month). To select a time attribute, you need to include one in the report.

OK Cancel



New % of, Rank, Running Sum, etc.

Choose the number field that you want to create:

% of Sales

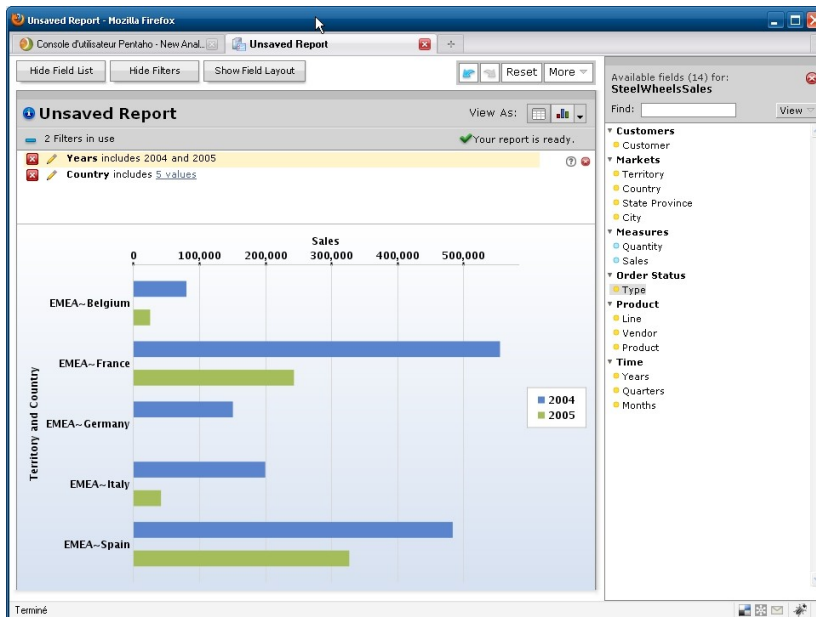
Rank by Sales

Running Sum of Sales

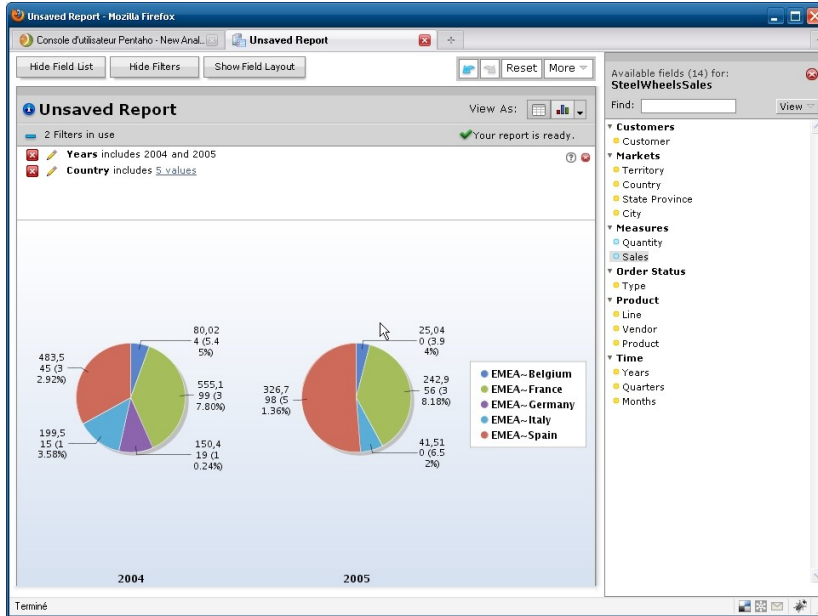
% of Running Sum of Sales

Next Cancel

- Ajouts de graphiques, sans paramétrage complexe: Bar Charts (histogrammes horizontaux/verticaux, empilés où non), Pie Charts (camemberts), Area Chart (zones), Line Charts (courbes). Les graphes sont réactualisés en permanence selon les tableaux construits par l'utilisateur

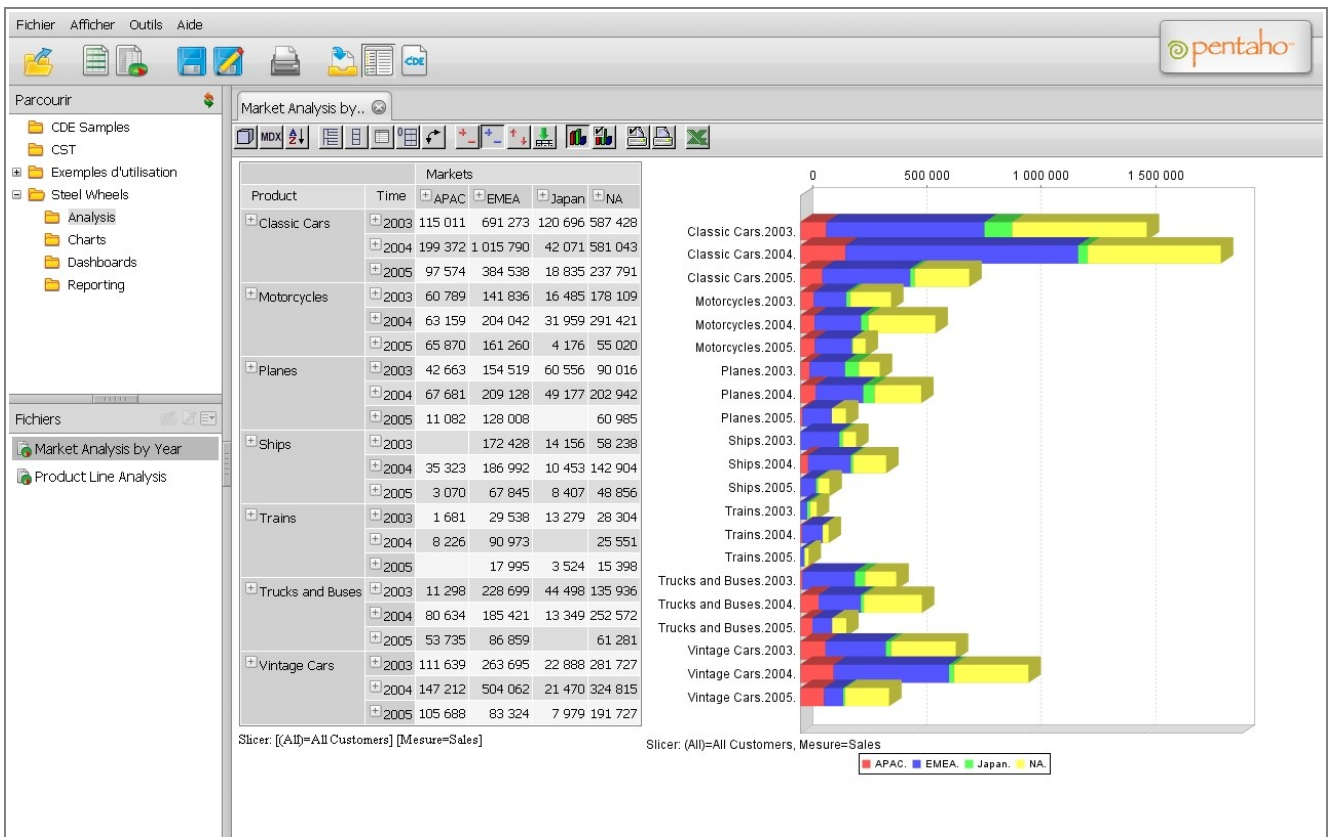


- Possibilité d'exporter les analyses au format EXCEL, PDF ou CSV (pour un échange avec d'autres applications)



4.2.2 Fonctionnalités OLAP avec la version Communautaire

Jpivot est le navigateur OLAP disponible dans la version gratuite.



Les manques fonctionnels en comparaison de Pentaho Analyzer sont essentiellement :

- Une ergonomie moindre et une présentation web 1.0
- L'impossibilité de rajouter des membres calculés à la volée
- Des graphiques un peu moins lisibles et faciles à réaliser

A noter que Jpivot va être remplacé en open source courant 2011 par le projet « SAIKU ».

Il s'agira d'un requêteur OLAP full web 2.0 (drag&drop, meilleur ergonomie, graphiques en Flash...)

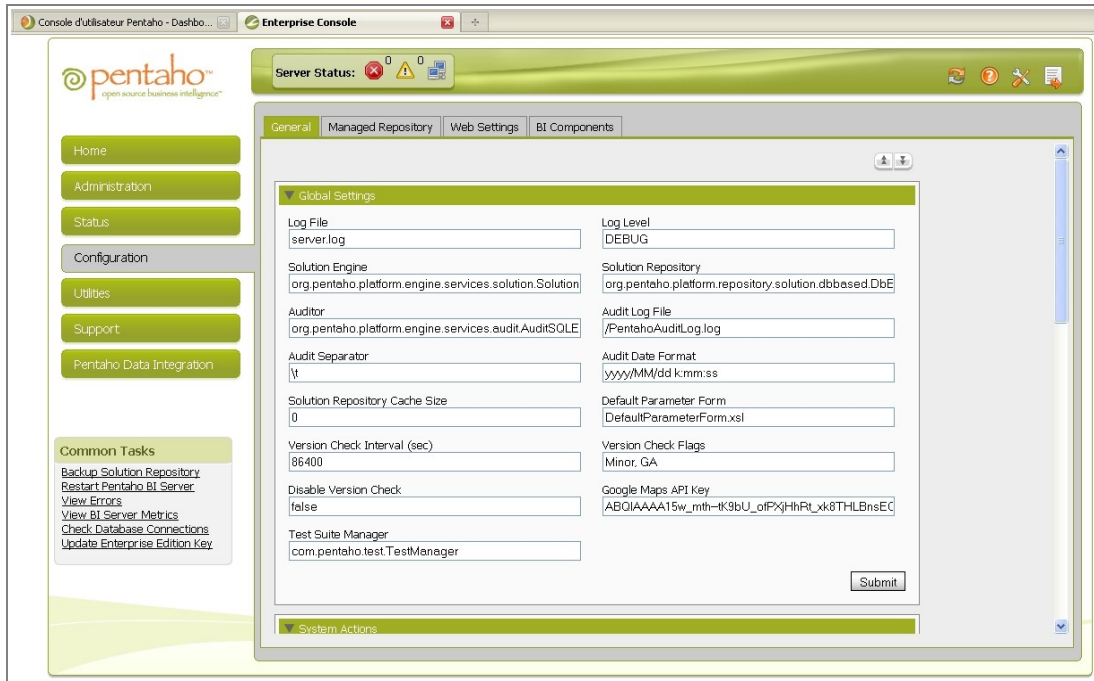
Une vidéo de démonstration de JPivot sur un jeu de données UNEDIC est disponible sur la chaîne YouTube [opensourcebi](#) à cette url :

<http://www.youtube.com/watch?v=T16K8vsuUiA>

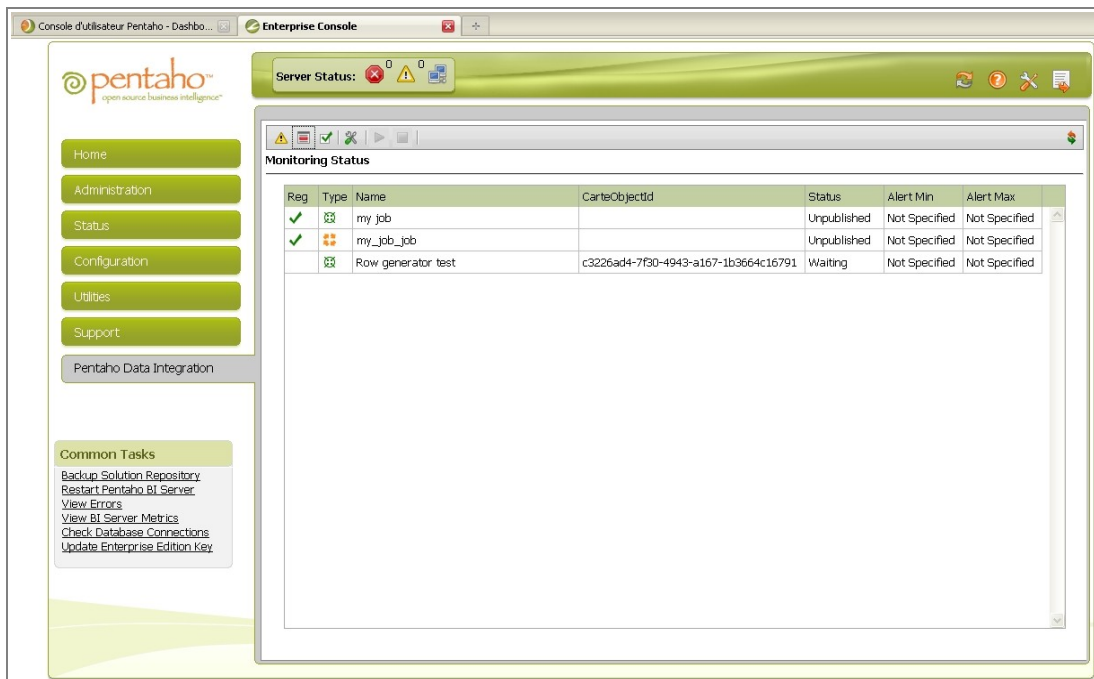
4.3 CONSOLE D'ADMINISTRATION

4.3.1 La Console d'administration en version Enterprise

La console d'administration EE est une application web qui permet de configurer, paramétrer et monitorer l'ensemble du serveur Pentaho. Elle permet essentiellement de ne pas avoir à éditer manuellement les fichiers de configuration du serveur Pentaho (fichiers XML...)



On notera surtout la présence d'une interface de planification et de monitoring des jobs Kettle

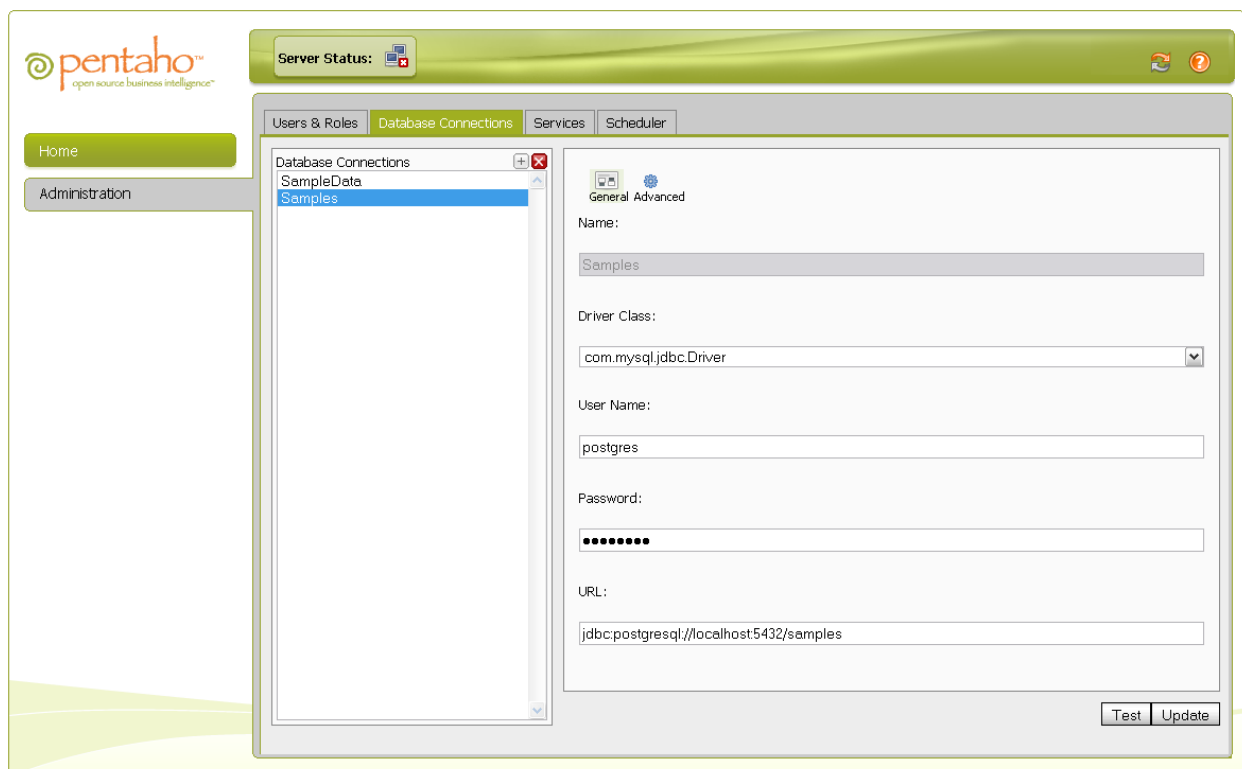


4.3.2 La Console d'administration en version Communautaire

La console d'administration en version communautaire est une version simplifiée de la console Enterprise.

Celle-ci permet uniquement :

- La gestion des sources de données (ajout/modification/suppression d'une data source)
- La gestion des utilisateurs (ajout/modification/suppression d'utilisateurs et de groupes)
- Mise en place de tâches planifiées
- Pilotage des divers mécanismes de rafraîchissement de la plate-forme

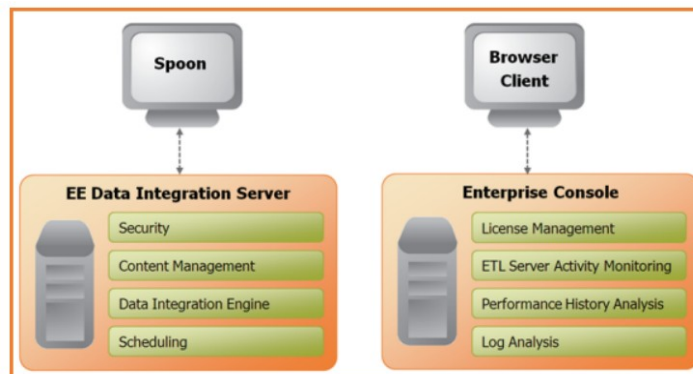


4.4 LES APPORTS DE PENTAHO DATA INTEGRATION EN VERSION ENTERPRISE

4.4.1 Introduction

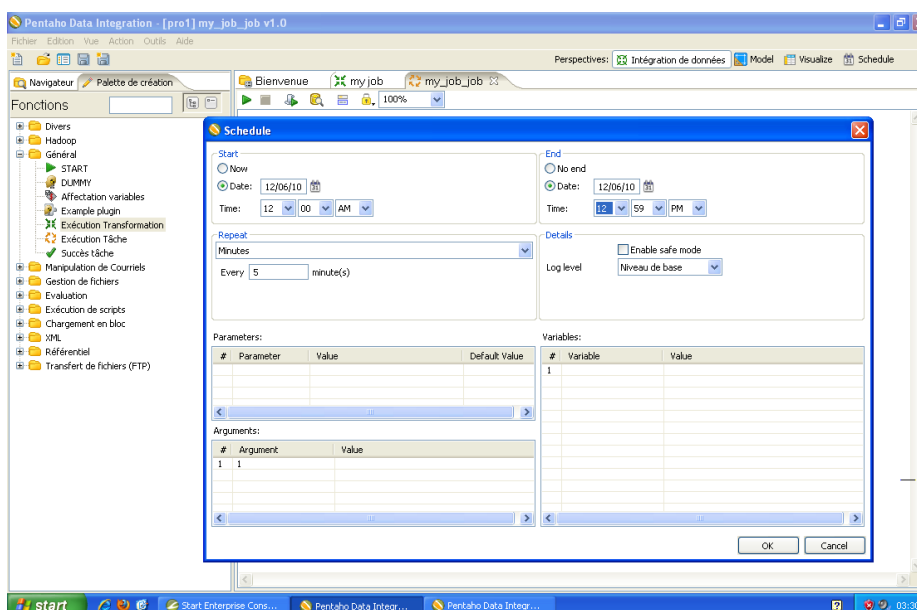
La version payante de Kettle amène quelques fonctionnalités intéressantes.

Celles-ci peuvent s'avérer intéressantes dans le cadre de déploiements importants, typiquement lors de la mise en œuvre par des Grands Comptes qui emploient de nombreux développeurs devant concevoir et déployer des jobs ETL :



4.4.2 Caractéristiques de PDI EE

- Gestion des droits d'accès avec un contrôle sur le contenu et les actions qui peuvent être effectuées par les développeurs et les groupes de développeurs (via rôles). La sécurité peut-être gérée directement dans le référentiel Kettle ou configurée à partir de l'annuaire LDAP de l'entreprise (Active Directory au autre)
- Gestion du travail collaboratif avec sécurisation d'accès aux jobs, versionning et verrouillage des jobs en cas de modification en cours (check in/check out)
- Un planificateur intégré facilitant la gestion des déclenchements programmés :



- Quelques briques fonctionnelles supplémentaires : Hadoop, JMS, GoogleDocs, ...