

Présentation et installation de Pentaho



Auteur : **Sylvain**

1. Présentation de Pentaho

Pentaho contient une multitude d'outils qui sont pour la plupart accessibles gratuitement. Ils sont destinés au domaine de l'informatique décisionnelle (business intelligence). Ses outils nous servent à analyser des données afin de les rendre lisibles par l'intermédiaire de graphiques qui peuvent être rassemblés dans des rapports.

Il existe deux versions de Pentaho :

- **Enterprise Edition** (EE). La version **payante**. Il est possible de tester cette version pendant 30 jours grâce aux 30-day trial.
- **Community Edition** (CE). La version en **open source** que nous utiliserons. Il faut faire attention, certaines applications ne sont disponibles que dans la version EE.

Pentaho a l'avantage de fonctionner sur Windows, Linux ou Mac-OS. Ses outils peuvent être des applications et des plug-ins. Nous pouvons différencier deux types d'applications dans Pentaho :

- Les applications serveurs, utilisables via une console.
- Les applications client, dans ce cas, l'application ne sera utilisable que via un bureau.

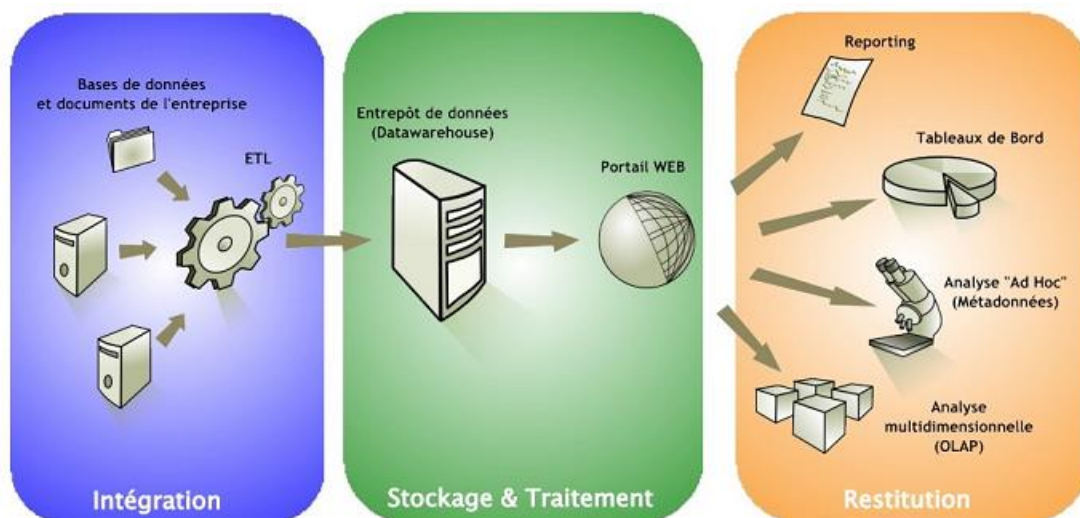
Ses différentes applications sont listées sur le wikipedia de Pentaho : <https://en.wikipedia.org/wiki/Pentaho>.

Nous allons voir deux outils de Pentaho :

- **Pentaho Data Integration** (PDI), anciennement connu sous le nom de Kettle. Il sera notre **ETL** (Extract Transform Load). C'est une application client qui dispose d'une interface, nous aurons donc besoin d'un bureau pour utiliser cette application.
- **Business Analytics Platform**, anciennement connu sous le nom de Pentaho BI Server. Il fournit une interface Web J2EE, elle permet d'accéder aux données, de créer et de distribuer les tableaux de bord (dashboard).

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

- une couverture globale des fonctionnalités de la Business Intelligence :
 - ETL (intégration de données),
 - reporting,
 - tableaux de bord ("*Dashboards*"),
 - analyse *ad hoc* (requêtes à la demande),
 - analyse multidimensionnelle (OLAP) ;



- Pentaho permet d'adresser deux typologies d'utilisateurs :

- les « one-clic users », utilisateurs de base, consommateurs d'indicateurs prédéfinis,
- les utilisateurs avancés, qui ont besoin d'outils d'analyse et d'exploration avancés ;



➔ 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 :

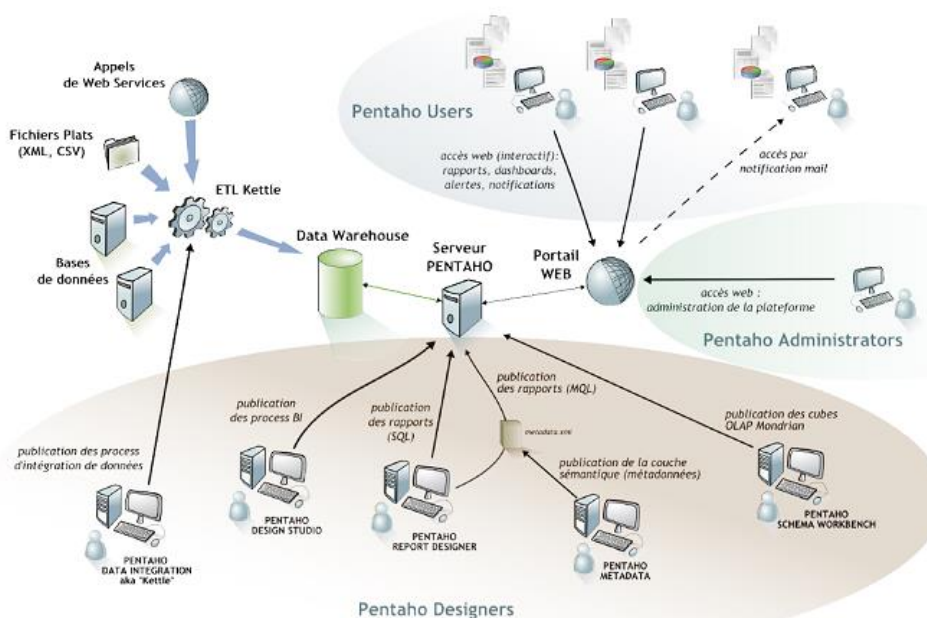
- Permet de créer des rapports avancés (choix des colonnes, tableaux croisés) à partir de vues métiers ("business views"). Nécessite la mise en place d'une couche sémantique d'abstraction (métadonnées)

➔ Analyse multidimensionnelle (OLAP)

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

- une architecture Web 2.0 qui se compose :

- d'un serveur Web J2EE permettant de mettre à disposition l'ensemble des ressources décisionnelles et ceci au travers d'URL Web uniques et standardisées. Le serveur est dénommé "Pentaho User Console" (PUC),
- 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*) pour 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 » : au travers de « séquences d'actions » on peut ainsi modéliser avec Pentaho des workflows BI avancés ;
- il n'est pas besoin de connaître JAVA pour travailler avec Pentaho : seule la maîtrise du langage SQL est nécessaire, ainsi que des connaissances de base en XML, HTML et JavaScript. Il faut bien sûr s'autoformer (ou être formé) aux clients de conception ;
- une communauté importante et très active s'anime autour de Pentaho. Celle-ci contribue au codage de nombreux plugins et de projets communautaires : plugins Kettle, Pentaho Analysis Tool, Pentaho Community Dashboard Framework, etc. ;
- Pentaho est une suite décisionnelle open source commerciale qui reste très « ouverte ». Les différences fonctionnelles entre la version libre (**community edition**) et la version payante (**enterprise edition**) restent limitées. **La version libre de Pentaho permet d'installer une plate-forme décisionnelle complète !**

2. Téléchargement de Pentaho

Pour débuter avec Pentaho, il est conseillé de télécharger la version community, gratuite et libre d'utilisation. Cette version communautaire peut-être téléchargée sur [SourceForge](http://sourceforge.net/projects/pentaho/files) ici :

<http://sourceforge.net/projects/pentaho/files>.

On y retrouve le serveur Pentaho ("*Business Intelligence Server*") ainsi que tous les clients de conception (voir détail au paragraphe suivant) :

File/Folder Name	Platform	Size	Date ↓	Downloads	Notes/Subscribe
Newest Files					
 pme-ce-mac-3.5.0-stable.zip		48.9 MB	2010-03-18	657	
All Files 					
▶  Pentaho Metadata	PME	587.1 MB	2010-03-18	34,345	 
▶  Report Designer	PRD	1.1 GB	2010-03-04	63,421	 
▶  Business Intelligence Server	BISERVER	7.3 GB	2010-02-25	497,751	 
▶  Data Integration	PDI (Kettle)	1.4 GB	2009-12-21	147,210	 
▶  Design Studio	PDS	3.4 GB	2009-10-15	104,531	 
▶  Report Design Wizard (Legacy) — <i>(obsoleète)</i>		181.7 MB	2008-08-29	50,102	 
▶  White Papers <i>(documents "anciens"...</i>		1.6 MB	2006-06-20	57,355	 

Les numéros des versions téléchargées pour les clients de conception et le serveur Pentaho (biserver) doivent toujours être en adéquation : par exemple Pentaho Report Designer 3.0.0 avec Biserver 3.0.0.

Des paquetages d'installation sont disponibles pour tous les systèmes d'exploitation :

- Windows ;
- Linux ;
- Mac.

▼  Report Designer		1.1 GB	2010-03-04	63,421	 
▼  3.6.0-stable		164.5 MB	2010-03-04	4,174	 
 prd-source-3.6.0-stable.zip Code Source		591.0 KB	2010-03-04	634	
 prd-ce-mac-3.6.0-stable.tar.gz MAC		54.6 MB	2010-03-04	335	
 prd-ce-3.6.0-stable.zip WINDOWS		54.7 MB	2010-03-04	2,489	
 prd-ce-3.6.0-stable.tar.gz LINUX		54.6 MB	2010-03-04	716	

3. Installation de Pentaho (en local)

Le serveur Pentaho (*biserver-ce*) est un serveur de démonstration prêt à l'emploi, complètement autonome et pouvant être installé sur un PC bureautique disposant au moins de 1 Go de RAM.

Ce serveur s'appuie notamment sur le système de gestion de base de données Hypersonic (HsqlDb) pour le stockage des données exemples (*SampleData*) ainsi que des deux bases internes de Pentaho (*hibernate* et *quartz*).

HsqlDb étant un système de base de données gérée en mémoire, il est fortement déconseillé de déployer cette configuration en production !

Une fois l'archive téléchargée, il suffit de décompresser celle-ci dans un répertoire préalablement créé, par exemple « C:\Pentaho-3.5.2 » (Windows).

Le répertoire d'installation sera désigné {PENTAHO-HOME} dans la suite de ce document.

Deux répertoires sont créés dans {PENTAHO-HOME} :

- **\biserver-ce** : la console Web d'utilisation (Pentaho User Console) ;
- **\administration-console** : la console Web pour l'administration de la plate-forme (Pentaho Administration Console).

5. Démarrer & arrêter les serveurs Pentaho▲

5-A. Pentaho User Console▲

Action	Commande
Démarrage (Windows)	{PENTAHO-HOME}\biserver-ce\start-pentaho.bat
Arrêt (Windows)	{PENTAHO-HOME}\biserver-ce\stop-pentaho.bat
Démarrage (Linux)	{PENTAHO-HOME}\biserver-ce\start-pentaho.sh
Arrêt (Linux)	{PENTAHO-HOME}\biserver-ce\stop-pentaho.sh

Les commandes suivantes permettent de lancer et stopper la console d'utilisation Web

On accède à la console d'utilisation Pentaho en saisissant l'URL suivante dans un navigateur Web :

http://localhost:8080/pentaho

Puis en saisissant l'identifiant et mot de passe ci-dessous :

- **login** : joe ;
- **password** : password.

5-B. Pentaho Administration Console

Les commandes suivantes permettent de lancer et stopper la console d'utilisation Web.

Action	Commande
Démarrage (Windows)	{PENTAHO-HOME}\administration-console\start-pac.bat
Arrêt (Windows)	{PENTAHO-HOME}\administration-console\stop-pac.bat
Démarrage (Linux)	{PENTAHO-HOME}\administration-console\start-pac.sh
Arrêt (Linux)	{PENTAHO-HOME}\administration-console\stop-pac.sh

On accède à la console d'administration Pentaho en saisissant l'URL suivante dans un navigateur Web :

http://localhost:8099

Puis en saisissant l'identifiant et mot de passe ci-dessous :

- **login** : admin ;
- **password** : password.

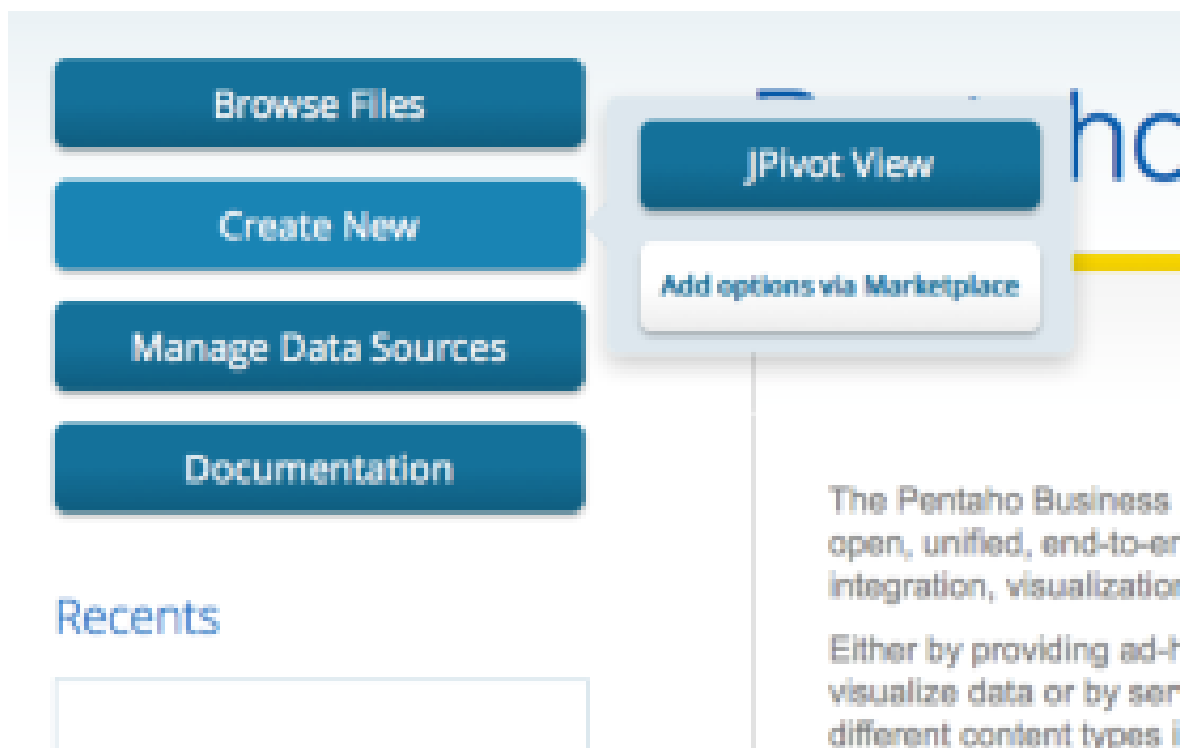
6. Liste des clients de conception Pentaho

Outil	Code	Fonction
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)...
Pentaho Design Studio	PDS	Client 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 (métadonnées) 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 métadonnées peut-être utilisée dans le requêteur <i>Web ad hoc</i> , 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. Un autre outil, Pentaho Aggregation Designer (PAD), permet de construire et de charger automatiquement des tables d'agrégation en vue d'améliorer les performances lors du requêtage des cubes Mondrian. Téléchargement spécifique ici : http://sourceforge.net/projects/mondrian/files/
Pentaho Data	PDI	Outil ETL (Extract Transform Load) complet, pouvant être utilisé indépendamment de la plate-forme

Intégration (Kettle)		Pentaho. Kettle est comparable à Talend Open Studio en termes de fonctionnalités. Pour consulter les différences techniques et fonctionnelles, un livre blanc est disponible à cette adresse : http://www.atolcd.com/actualites/detail-actualite/actualite/2/comparatif-etopen-source-1.html
-----------------------------	--	--

7. Interface de Pentaho

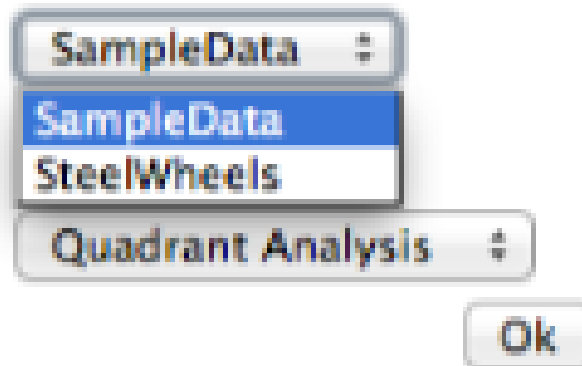
Après être authentifié, nous allons créer un JPivot (librairie conçu en Java pour la navigation et exploitation des informations via des opérations OLAP).



Deux jeux de données d'essais sont proposés pour Pentaho. Dans ce TP, nous allons utiliser le premier.

New JPivot View

Schema



Une fois sélectionnée le Schéma et le Cube, nous accédons à l'interface de manipulation de données et construction de requêtes OLAP.

