

# Systeme d'Aide à la Décision

## Pentaho

*ENSA Khouribga*

**Pr. SOUSSI Nassima**

# Qu'est-ce que Pentaho ?

- Pentaho est une entreprise qui propose une suite de logiciels regroupés sous le nom de *Pentaho Business Analytics*.
- *Pentaho Business Analytics* est une plate-forme décisionnelle open source qui offre, à travers divers outils, une couverture globale des fonctionnalités de la Business Intelligence :
  - ETL (intégration de données),
  - reporting,
  - tableaux de bord ("Dashboards"),
  - analyse multidimensionnelle (OLAP),
  - ...

### **Versions communautaire ou entreprise ?**

- La plate-forme Pentaho existe en version communautaire et en version entreprise.
- La version entreprise propose des fonctionnalités et des services supplémentaires qui n'existent pas dans la version gratuite, essentiellement des applications sur serveur et un support.
- Il est possible de télécharger une version d'essai de la plate-forme qui reste gratuite pendant 30 jours.

# Les Types d'utilisateur de Pentaho

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 ;



## Applications disponibles dans la version communautaire

- **Pentaho Data Integration (PDI)**
- Pentaho Schema Workbench
- Pentaho Community Dashboard Editor
- Pentaho Report Designer
- Pentaho Data Mining
- ...

# Présentation générale de PDI

- **PDI** (*Pentaho Data Intégration*), que l'on trouve encore communément sous le nom de KETTLE (**Kettle ETL Environment**), gère la partie d'intégration de données et constitue ainsi l'outil d'ETL de la plateforme.
- Il est composé de plusieurs modules : Spoon, Pan, Kitchen et Carte.

### Les modules de PDI : **Spoon**

- Ce module propose une interface graphique qui permet à l'utilisateur de créer facilement un processus d'ETL sans avoir besoin à saisir du code.
- C'est le seul module de PDI qui propose une interface graphique. Pour tous les autres, il faut passer par l'invite de commandes.

## Les modules de PDI : Spoon

**COMPOSANTS DE CREATION**

**NAVIGATEUR**

**ZONE DE CREATION**

**Résultats exécution**

#	Nom étape	N° Copie	Lignes lues	Lignes écrites	Lignes en entrées	Lignes en sortie	Lignes maj	Lignes rejetées	Lignes...
1	Orders	0	0	15	16	0	0	0	
2	RealTime	0	0	13	14	0	0	0	
3	Tri lignes	0				0	0	0	
4	Tri lignes 2	0	13	13	0	0	0	0	
5	Jointure comparaison lignes	0	28	15	0	0	0	0	
6	Remplacer valeur nulle	0	15	15	0	0	0	0	

**RÉSULTATS (trace, prévisualisation, ...)**



# Les modules de PDI : Spoon

Spoon manipule deux sortes d'objets :

### 1. Les transformations :

- Comme son nom l'indique, la transformation va permettre **d'effectuer des transformations sur les données**. Plus largement, elle récupère les données depuis leurs sources, effectue les traitements sur ces données et délivre les données résultantes.
- Ce type d'objet sera sauvegardé dans des fichiers portant l'extension **.ktr** (*pour Kettle transformation*).

# Les modules de PDI : Spoon

Spoon manipule deux sortes d'objets :

## 2. Les tâches :

- Une tâche, quant à elle, ne permet pas de travailler directement sur les données mais propose différents services autour de celles-ci.
- Ce type d'objet sera sauvegardé dans des fichiers portant l'extension **.kjb** (*pour Kettle job*).

# Les modules de PDI : **Spoon**

Spoon manipule deux sortes d'objets :

## 2. Les tâches : (suite)

Exemples de services de tâche :

- Vérification de la présence d'un fichier
- Synchronisation des transformations
- Alerte par mail en cas d'échec d'une transformation
- ...

# Les modules de PDI

### => Module 2 : Pan

- Ce module permet d'exécuter les transformations réalisées dans Spoon en utilisant l'invite de commandes.
- Cela permet de lancer des transformations sur nos données sans ouvrir de logiciel et éventuellement, comme cela pourrait être le cas dans une entreprise, de lancer facilement l'exécution de multiples transformations via un simple batch.

# Les modules de PDI

### => Module 3 : **Kitchen**

- Kitchen permet l'exécution de tâches en mode batch et propose un plan de suivi de ces tâches pour contrôler facilement le processus d'ETL.
- De même que *Pan*, cela facilite l'exécution des tâches construites via l'interface graphique.

# Les modules de PDI

### => Module 4 : Carte

- Ce module est en fait un serveur web qui permet d'exécuter des tâches ET des transformations à distance.
- Il permet également de contrôler à distance, démarrer ou arrêter les processus en cours sur le serveur.

# PDI : Installation & Mise en place

1. Télécharger le fichier : *pdi-ce-X.X.X.X-XXX.zip*  
<http://sourceforge.net/projects/pentaho/files/Data%20Integration/>
2. Extraire les fichiers vers le répertoire de votre choix.
3. Lancer l'application en exécutant *Spoon.bat* sous Windows ou *spoon.sh* sous Linux et MacOS.