

Analyse dynamique depuis MongoDB avec PDI et Pentaho Business Analytics

blog.smile.fr |

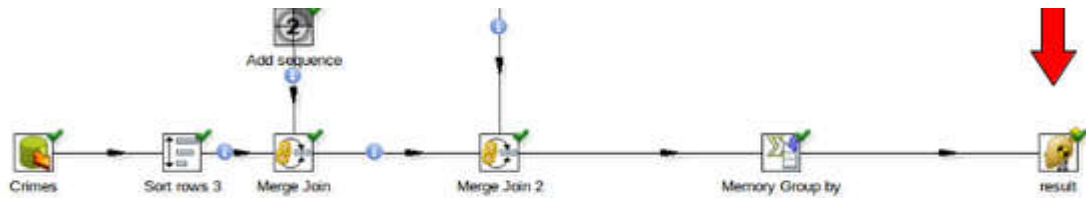
PDI (Pentaho Data Integration) est un outil ETL puissant qui fait partie de la suite décisionnelle Pentaho Business Analytics. Outre les fonctionnalités classiques que l'on peut attendre d'un ETL, sa compatibilité avec les autres modules de la suite Pentaho permet des utilisations moins classiques.

Le but ici est de remplacer une SGBD classique comme source de données pour des analyses OLAP par une transformation PDI.

Cet article décrit comment réaliser ce type d'analyse permettant de croiser des données MongoDB avec une base de donnée relationnelle :

1. Développement d'une transformation PDI qui manipulera des données hétérogènes pour créer un set de données cohérent. Les données seront récupérées en autres, d'une base de données SQL Postgresql, d'une base de données NoSQL MongoDB orientée document.
2. Il suffit ensuite, dans les paramètres de la transformation, de créer un Data Services dans l'onglet correspondant en définissant notamment le composant de la transformation portant le set de données voulu.





- Une fois la transformation développée, configurer un serveur Carte pour son exécution à distance, en définissant dans son fichier de configuration le Data Services créé auparavant

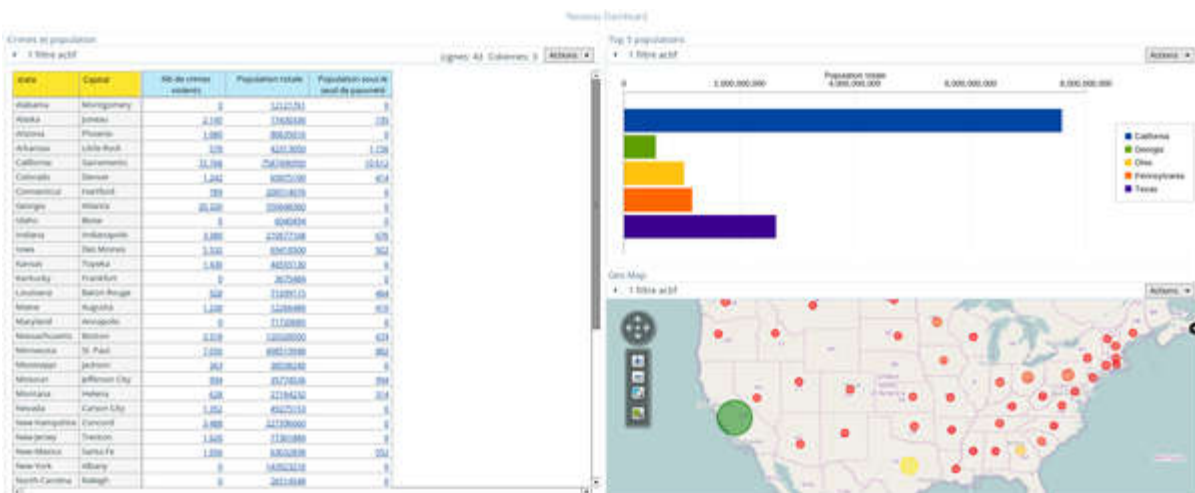
```

<slave_config>
<slaveserver>
  <name>carte_server</name>
  <hostname>localhost</hostname>
  <port>8082</port>
</slaveserver>
<services>
<service>
  <name>Zips_Mongo</name>
  <filename>/home/nawaa/Documents/Pentaho/repo/Laury/demo_mongo/mongo.ktr</filename>
  <service_step>result</service_step>
</service>
</services>
</slave_config>

```

- Définir la source de données dans la plateforme BI, définir le schéma Mondrian à utiliser pour l'analyse OLAP.

Dès lors, les données sont analysables directement sur l'interface web sous la forme d'un tableau ou encore sous la forme d'un graphique voire même d'une Geo Map



Comme prévu, les données sont chargées depuis notre transformation PDI à chaque rafraîchissement des données.





A noter que la possibilité d'utiliser une transformation comme source de données concerne toutes les applications Java étant donné que la fédération de données se base sur un driver JDBC nommé kettle thin Driver, en pointant sur la base « kettle ».