

Setting Up Pentaho Kettle

Document Author: arho.virkki@vtt.fi

Installation

Grab the latest release of Pentaho Kettle from <http://community.pentaho.com/projects/data-integration/> and extract it under */opt/pentaho/*. Use the exact version as the folder name and create a symbolic link as follows:

```
lrwxrwxrwx 1 root ktpshared 18 Dec 22 11:24 data-integration -> pdi-ce-6.0.1.0-386
drwxrwsr-x 17 root ktpshared 4096 Dec 29 10:28 pdi-ce-6.0.1.0-386
```

It might be necessary to set group rights and SGID bit to the folder (as above), but I haven't tested this. At least it is safe to do so.

Then, add Kettle folder to path by adding a corresponding script to */etc/profile.d/pentaho.sh*:

```
export PATH=$PATH:/opt/pentaho/data-integration/
export KETTLE_HOME=/opt/pentaho/
```

Finally, to make the help work Under Ubuntu 14.04, install the additional browser libraries with

```
sudo apt-get install libwebkitgtk-1.0-0
```

Configuration

The following shows an example of the global *kettle.properties* file

```
# This file was generated by Pentaho Data Integration version 6.0.1.0-386.
#
# Here are a few examples of variables to set:
#
# PRODUCTION_SERVER = hercules
# TEST_SERVER = zeus
# DEVELOPMENT_SERVER = thor
#
# Note: lines like these with a # in front of it are comments
#
ETL_PATH=/opt/ktp/ETL
```

Setting Memory Parameters for Java

The memory setting can be changed individually for each kettle script, for example by editing them directly:

```
sudo gvim `locate spoon.sh`
```

The system-wide properties can be set in */etc/profile.d/pentaho.sh*. The default setting for Kettle 6 are

```
export PENTAHO_DI_JAVA_OPTIONS="-Xms1024m -Xmx2048m -XX:MaxPermSize=256m"
```

An excerpt from `man java`:

```
-Xmsn
    Specifies the initial size, in bytes, of the memory allocation pool. This value must
    be a multiple of 1024 greater than 1 MB. Append the letter k or K to indicate kilo-
    bytes, or m or M to indicate megabytes. The default value is chosen at runtime based
    on system configuration. See Garbage Collector Ergonomics at http://docs.oracle.com/javase/7/docs/technotes/guides/vm/gc-ergonomics.html
    Examples:
    -Xms6291456
```

```
-Xms6144k  
-Xms6m
```

-Xmxn

Specifies the maximum size, in bytes, of the memory allocation pool. This value must be a multiple of 1024 greater than 2 MB. Append the letter k or K to indicate kilobytes, or m or M to indicate megabytes. The default value is chosen at runtime based on system configuration.

For server deployments, -Xms and -Xmx are often set to the same value. See Garbage Collector Ergonomics at <http://docs.oracle.com/javase/7/docs/tech-notes/guides/vm/gc-ergonomics.html>

Examples:

```
-Xmx83886080  
-Xmx81920k  
-Xmx80m
```