



**Bases de Datos Masivas (11088)**  
**Departamento de Ciencias Básicas**

## GUÍA DE INSTALACIÓN: SUITE PENTAHO

(SO Linux y Windows)

Pentaho BI Suite es un conjunto de programas libres para generar inteligencia empresarial (Business Intelligence) que emplearemos durante la materia. Utilizaremos la versión **community**, bajo licencia GPL.

### INSTALACIÓN PASO A PASO:

1. Debemos ingresar a la sección de downloads de la web de Pentaho community a efectos de descargar el software: <http://community.pentaho.com/>. Pentaho es multiplataforma, con lo cual no tendremos que distinguir por Sistema Operativo al momento de realizar la descarga de los instaladores.
2. Una vez en el sitio web, descargaremos los siguientes archivos:
  - a. **Business Analytics Platform (versión 6.1).**
  - b. **Data Integration (versión 6.1):** También conocida como Kettle.
  - c. **Mondrian Schema Workbench (versión 3.12).**
  - d. **Weka (versión 3.8):** (en UNIX:  
[www.cs.waikato.ac.nz/ml/weka/downloading.html](http://www.cs.waikato.ac.nz/ml/weka/downloading.html)).
3. Una vez descargados los archivos, y antes de proseguir con la instalación nos tendremos que asegurar de tener instalado Java Development Kit (JDK) y Java Runtime Environment (JRE) ya que la Suite Pentaho corre sobre la plataforma Java.
  - a. En UNIX podemos instalarlo a través del comando: **sudo apt-get install default-jdk default-jre**.
  - b. En Windows, descargamos los instaladores de las siguientes direcciones:



**Bases de Datos Masivas (11088)**  
**Departamento de Ciencias Básicas**

**JDK:** <http://www.oracle.com/technetwork/es/java/javase/download/s/jdk7-downloads-1880260.html>

**JRE:** <http://www.oracle.com/technetwork/java/javase/downloads/jre7-downloads-1880261.html>

- c. Una vez descargados e instalados, debemos configurar las variables de entorno JAVA\_HOME y PATH. Esto se explica al final de la guía para no complicar el flujo de la misma dado que existen procedimientos distintos según lo instalemos en un entorno UNIX o Windows.
4. A continuación instalaremos **Business Analytics Platform**:
  - a. Descomprimos el archivo biserver-ce-X.X.zip o biserver-ce-X.X.tar.gz, según sea nuestro caso.
  - b. En este momento podríamos iniciar Pentaho ejecutando el archivo start-pentaho.sh (UNIX) o start-pentaho.bat (Windows) y accederlo a través de un navegador mediante la URL: [localhost:8080/pentaho/](http://localhost:8080/pentaho/). En el caso de ambientes UNIX debemos cautelar que el archivo tenga permisos de ejecución para el usuario.
5. Luego, instalamos **Mondrian Schema Workbench**:
  - a. Descomprimos el archivo psw-ce-X.X.
  - b. A continuación aparecerá la carpeta schema-workbench.
  - c. Para ejecutar el software, ejecutamos el archivo workbench.sh (UNIX) o workbench.bat (Windows) que se encuentra en la carpeta raíz. En el caso de ambientes UNIX debemos cautelar que el archivo tenga permisos de ejecución para el usuario.
6. Ahora, instalamos **Data Integration**:
  - a. Descomprimos el archivo pdi-ce-X.X.
  - b. A continuación aparecerá la carpeta data-integration.



**Bases de Datos Masivas (11088)**  
**Departamento de Ciencias Básicas**

- c. Para iniciar PDI o Kettle, ejecutamos el archivo Spoon.sh (UNIX) o Spoon.bat (Windows) que se encuentra en la carpeta raíz. En el caso de ambientes UNIX debemos cautelar que el archivo tenga permisos de ejecución para el usuario.

7. Por último, instalamos **Weka**.

- a. En UNIX:
  - 1. descargamos el archivo zip y luego lo descomprimos.
  - 2. Para iniciarlo, ejecutamos en una terminal:

**java -jar weka.jar.**

- b. En Windows ejecutamos el instalador.



**Bases de Datos Masivas (11088)**  
**Departamento de Ciencias Básicas**

## CONFIGURACIÓN DE LAS VARIABLES DE ENTORNO DE JAVA

### **En ambientes UNIX:**

1. Debemos editar el archivo `bash.bashrc` que se encuentra en el directorio `/etc` (esto sirve para setear la variable en todos los perfiles). Para ello, abrimos una terminal y ejecutamos lo siguiente: **`sudo nano .bashrc`**.
2. Ahora agregamos lo siguiente en la última línea (después del último `fi`):  

```
JAVA_HOME=/usr/lib/jvm/<version java>/jre  
export JAVA_HOME  
PATH=$PATH:$JAVA_HOME
```
3. Guardamos los cambios y refrescamos las variables.

### **En ambientes Windows:**

1. Hacemos click derecho sobre Mi PC ó Equipo, de acuerdo a la versión del SO, y seleccionamos la opción “Propiedades”. Luego accedemos a “Configuración avanzada / Cambiar configuración -> Opciones avanzadas -> Variables de entorno -> Nueva (Variables del sistema)”.
2. Primero configuramos la variable `JAVA_HOME`. Para ello, escribimos en el cuadro de texto lo siguiente:
  - a. **Nombre de variable:** `JAVA_HOME`,
  - b. **Valor de variable:** ruta en que se haya instalado Java. Por ejemplo `“C:\Program Files\Java\jdk1.7.0_51”`.
3. Ahora configuramos la variable `PATH` accediendo nuevamente a la opción Variables de entorno:



**Bases de Datos Masivas (11088)**  
**Departamento de Ciencias Básicas**

- a. Vamos a añadir al contenido de la variable PATH el lugar donde se encuentran los archivos ejecutables de Java.
- b. Ahora elegimos la opción “Editar (Variables del sistema)”.
- c. Luego al final del contenido que ya exista, añadiremos un punto y coma y ruta a los ejecutables que será igual a la anterior pero seguida de la carpeta /bin. Por ejemplo: “;C:\Program Files\Java\jdk1.7.0\_51\bin”.