

Sistemas de Información para el Análisis de GVD

Instalación de las Herramientas

Edición 2024

Índice

ÍNDICE	2
1 INTRODUCCIÓN	3
2 HERRAMIENTAS	4
2.1 PostgreSQL.....	4
2.2 MongoDB	4
2.3 JSON.....	4
2.4 Pentaho Business Intelligence.....	4
2.4.1 ETL: Herramienta de Extracción, Transformación y Carga	6
2.4.2 Servidor OLAP	7
2.4.3 Herramienta de Diseño de Reportes.....	7
2.4.4 Análisis OLAP	8
2.5 Tableau	8
2.6 Power BI	8
2.7 Visualizer.....	8
REFERENCIAS	10

1 Introducción

Este documento pretende dar apoyo a la instalación de las herramientas a utilizar en el proyecto del curso. En las referencias que se ofrecen al final de este documento, se deja disponible un conjunto de manuales y documentaciones que describen la instalación de cada herramienta. Dichas herramientas y formatos para el intercambio de datos propuestas son las que se presentan a continuación:

- **PostgreSQL**
- **MongoDB**
- **Formato JSON**
- **Pentaho:**
 - Plataforma Business Intelligence: **Pentaho BI Analytic**
 - Herramienta de Extracción, Transformación y Carga: **Pentaho's Data Integration**, también conocido como **Kettle**
 - Servidor OLAP: **Mondrian**
 - Herramienta gráfica para el diseño de Schemas: **Schema Workbench**
 - Análisis OLAP: **Saiku**
 - Diseño de Reportes: **Pentaho Report Designer**
 - Diseño de componentes Dashboard: **Community Dashboard Editor**
- Visualización y análisis de datos: **Tableau, Power BI, Visualizer**

Algunas de las herramientas propuestas para el sistema de información que se va a desarrollar, son productos pertenecientes al proyecto *Pentaho Business Intelligence*, cuya versión *open-source* se encuentra disponible en [1]. La integración de las herramientas necesarias para el completo desarrollo del sistema se muestra en la Figura 1.

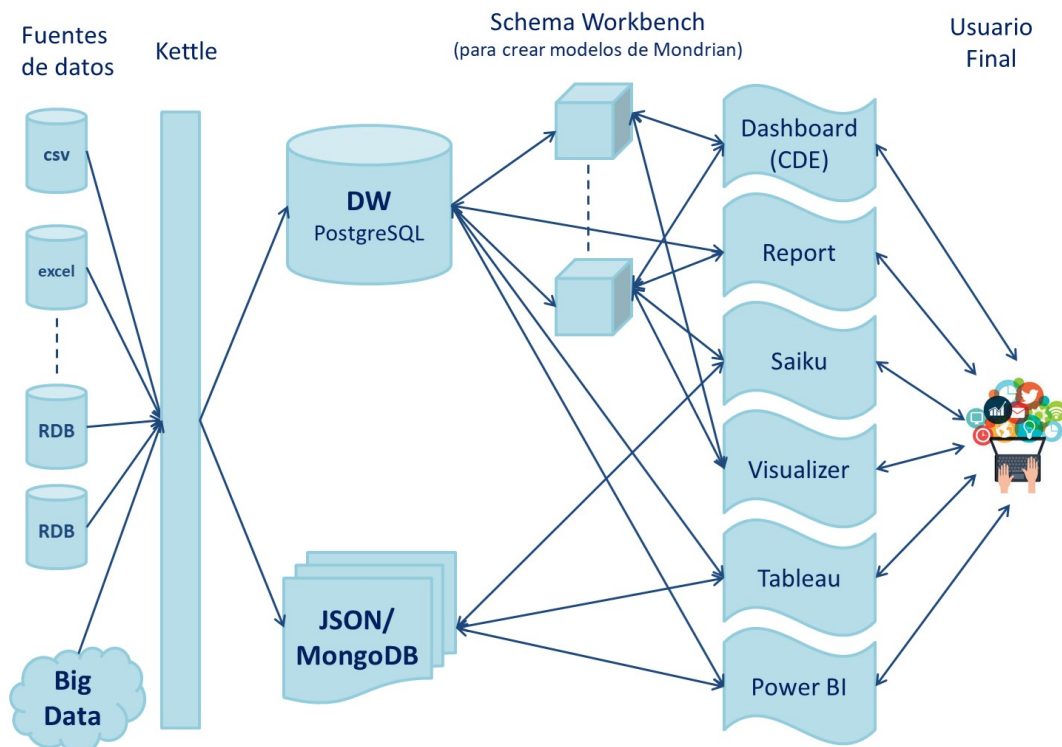


Figura 1. Integración de las herramientas utilizadas en el Proyecto del Curso

2 Herramientas

En esta sección se describen todas las herramientas usadas en el proyecto del curso.

2.1 PostgreSQL

Se utilizará la versión 16.2 de *PostgreSQL* disponible para descargar en [2].

2.2 MongoDB

MongoDB es una base de datos de documentos que ofrece una gran escalabilidad y flexibilidad. En el curso usaremos la versión 7.0 que se encuentra disponible en [3]. Además, es posible consultar un tutorial acerca de esta base de datos en [4].

2.3 JSON

JSON es un formato de intercambio de datos. Está basado en un subconjunto del Lenguaje de Programación JavaScript. Información sobre su estructura puede ser consultada en [5].

2.4 Pentaho Business Intelligence

Todos los productos de Pentaho están desarrollados en Java, por lo tanto, para trabajar con ellos, es necesario instalar el **Java Development Kit** también conocido como el JDK. Por lo tanto, se debe instalar un Java SE de 64 bits donde desea ejecutar Pentaho Data Integration (PDI). De acuerdo con el manual de Pentaho [6], cualquier versión de Java SE entre la 11 y la 18 debería funcionar. Estas versiones pueden ser encontradas en [7]. En caso de ser necesario, se debe agregar la variable de entorno **JAVA_HOME**, como se muestra en la Figura 2, con el valor correspondiente al directorio de instalación del JDK. Las variables de entorno en Windows10 son configurables en *Panel de Control -> Sistema y Seguridad -> Sistema -> Configuración Avanzada del Sistema -> Variables de Entorno*.

Para la **Plataforma Business Intelligence**, se debe utilizar la versión 9.4 de Pentaho BI, disponible en [1]. Por otro lado, en [8] tienen disponible un **manual de instalación** completo, donde además se especifican los **requerimientos del sistema**. Además, en [9] se encuentra disponible documentación para definir **conexiones de datos** con el servidor Pentaho.

A continuación se describen las herramientas presentadas en la introducción de este documento. La versión recomendada de todas ellas es la 9.4. Sin embargo, versiones superiores de estas herramientas están disponibles, por lo que los estudiantes podrán optar por dichas versiones. Recuerden que las versiones utilizadas deberán ser especificadas en los informes.

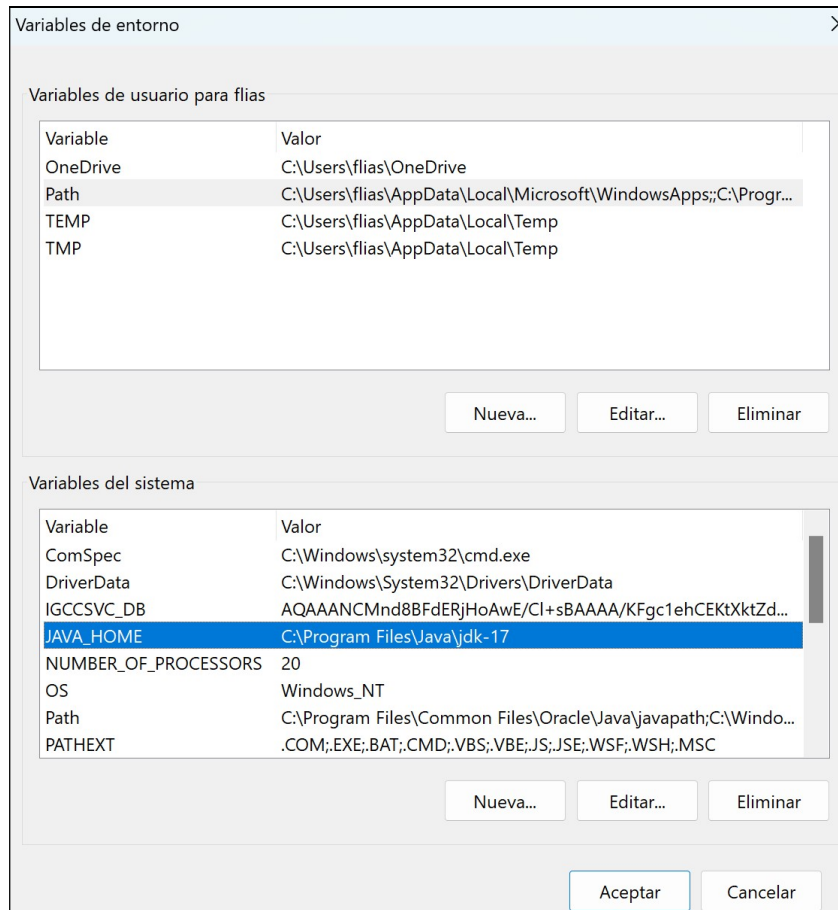


Figura 2. Variables de Entorno, JAVA_HOME

Plataforma Business Intelligence: Se debe utilizar la versión 9.4 de Pentaho BI, disponible en [23].

Instalación: La instalación consiste simplemente en extraer la carpeta del archivo descargado en [23]: pentaho-server-ce-9.4.0.0-343.zip. Es recomendable que el directorio de instalación sea el siguiente:

- Para Windows: C:\Program Files\Pentaho
- Para Linux: /opt/pentaho

En el resto del documento la carpeta *C:\Program Files\Pentaho\pentaho-server* será referenciada como **Pentaho Home**.

Sservidor: El servidor de Pentaho se puede iniciar y detener manualmente mediante los script **start-pentaho.bat** y **stop-pentaho.bat** respectivamente. Para su ejecución, se recomienda crear un acceso directo a dichos archivos. Estos scripts se encuentran en el **Pentaho Home**.

El servidor está pre-configurado para correr sobre el contenedor de **Apache Tomcat**, el cual se encuentra en el directorio **TOMCAT** dentro del **Pentaho Home**. Debe

asegurarse que las variables de entorno `CATALINA_HOME` y `CATALINA_BASE`, apuntan al directorio de **Apache Tomcat** como se muestra en la Figura 3.

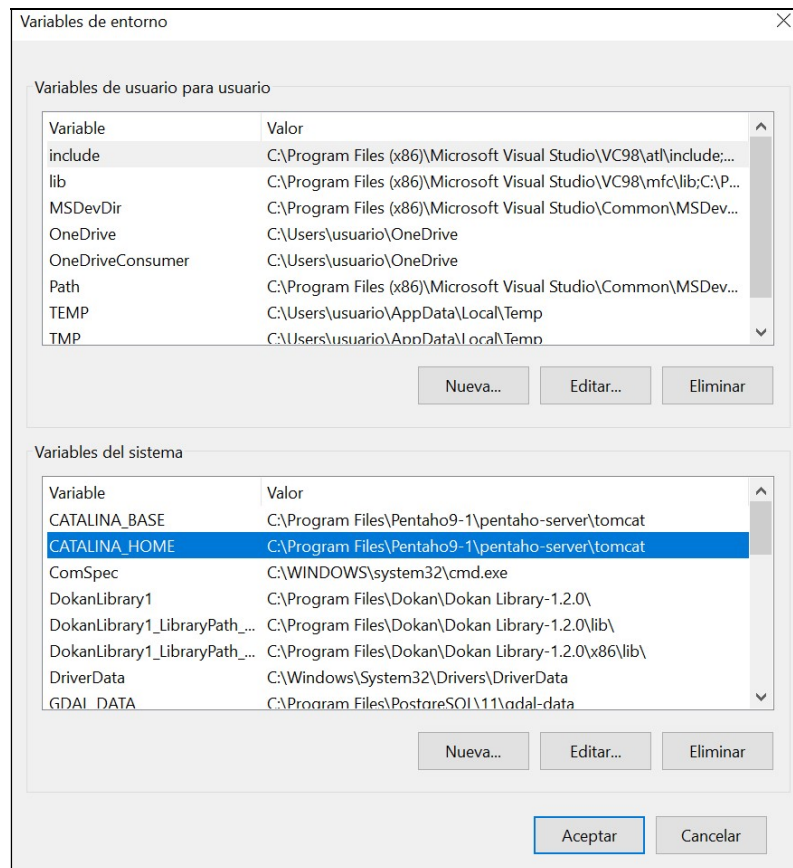


Figura 3. Variables de Entorno, `CATALINA_HOME` y `CATALINA_BASE` (Notar que la imagen corresponde a la versión 9.1, por lo que la carpeta debería ser Pentaho9-4).

Por defecto, el Tomcat trabaja sobre el puerto 8080, por lo que, cuando se inicie el servidor de Pentaho, el mismo será accesible localmente a través de la página <http://localhost:8080/pentaho>. En la Figura 4 se presenta la pantalla de inicio de Pentaho. Tener en cuenta que si el puerto 8080 está ocupado, se deberá utilizar otro, por ejemplo 9080 y en este caso se accederá a través de la página <http://localhost:9080/pentaho>.

2.4.1 ETL: Herramienta de Extracción, Transformación y Carga

La herramienta proporcionada por Pentaho para los procesos de extracción, transformación y carga es **Pentaho's Data Integration** (PDI), también llamado **Kettle**, del cual utilizaremos la versión 9.4 disponible en [1]. La instalación consiste simplemente en descomprimir el archivo descargado **pdi-ce-9.4.0.0-343.zip**, el cual contiene una carpeta llamada **data-integration**, y ubicarla en el directorio de instalación que se desee. Es recomendable utilizar el mismo directorio de instalación que el servidor (por ejemplo, **Pentaho Home**), de esta forma, todos los productos de Pentaho se encontrarán en un mismo directorio.

Kettle es una aplicación de escritorio y se inicia mediante el archivo *Spoon*. **Spoon.bat** para Windows y **Spoon.sh** para Linux, ubicado en la carpeta **data-integration**. Para su ejecución, se recomienda crear un acceso directo a dicho archivo. Recomendamos seguir el manual de instalación disponible en [10].

2.4.2 Servidor OLAP

El servidor OLAP que ofrece Pentaho es el **Mondrian** [11], el cual ya viene incluido en la instalación de la plataforma BI. Se trata de un servidor del tipo ROLAP. Los modelos multidimensionales que son interpretados por Mondrian consisten en archivos XML denominados **Schemas**, los cuales deben tener un formato determinado. En ellos se definen los cubos, dimensiones, niveles, jerarquías, etc.; sobre los que se trabajará y realizarán los diferentes análisis. La especificación detallada sobre el diseño y estructura que deben tener los **Schemas** se encuentra en [12].

Existe una herramienta gráfica para el diseño de *Schemas* denominada **Schema Workbench**, documentación sobre esta herramienta puede ser consultada en [13][14]. *Schema Workbench* permite editar y crear modelos multidimensionales (Mondrian). Es posible crear modelos de Mondrian gráficamente o definirlos manualmente mediante archivos XML. La instalación consiste simplemente en descomprimir el archivo **psw-ce-9.4.0.0-343.zip**, que se puede descargar en [1]. Este archivo contiene una carpeta llamada **schema-workbench**, la cual puede ser ubicada en cualquier directorio. Se recomienda instalarla en el directorio **Pentaho Home**.

Se trata de una aplicación de escritorio y se inicia mediante el archivo **workbench.bat** para Windows y **workbench.sh**, los cuales están ubicados en la carpeta **schema-workbench**. Para facilitar su ejecución, se recomienda crear un acceso directo a dicho archivo.

Por otro lado, dejamos disponibles recursos en los que encontrarán documentación sobre Multidimensional Expressions (MDX) [15][16]. Consultas MDX permiten especificar un archivo de esquema de definición de un cubo Mondrian.

2.4.3 Herramienta de Diseño de Reportes

Para el diseño de reportes se utilizará la herramienta **Pentaho Report Designer** [17]. La instalación consiste simplemente en descomprimir el archivo **prd-ce-9.4.0.0-343.zip**, descargado en [1], el cual contiene una carpeta llamada **report-designer**. Esta carpeta puede ser ubicada en cualquier directorio, se recomienda hacerlo en el directorio **Pentaho Home**. Se trata de una aplicación de escritorio y se inicia mediante el archivo **report-designer.bat** para Windows y **report-designer.sh**, los cuales están ubicados en la carpeta **report-designer**. Para facilitar su ejecución, se recomienda crear un acceso directo a dicho archivo. Por otro lado, en [18] se presenta una guía sobre cómo crear la **conexión a una fuente de datos** a la hora de generar un reporte.

2.4.4 Análisis OLAP

Por otro lado, **Saiku** es una herramienta OLAP que permite realizar análisis de forma rápida e intuitiva. A pesar de ser un proyecto independiente, esta herramienta puede ser integrada a Pentaho en forma de plugin, el cual está disponible para descargar en el enlace disponible en [19].

Para instalar el **plugin de Saiku**, una vez descargada la carpeta, simplemente hay que colocarla en el directorio `<Pentaho Home>/pentaho-solutions/system`. Luego, al iniciar el servidor e ingresar al sistema, aparecerá el ícono de **Saiku** en la barra de herramientas, como se observa en la Figura 3. Observar que la versión de Pentaho que se muestra en la Figura 3 es la 9.1, por lo que el menú de la versión 9.4 podría diferir.

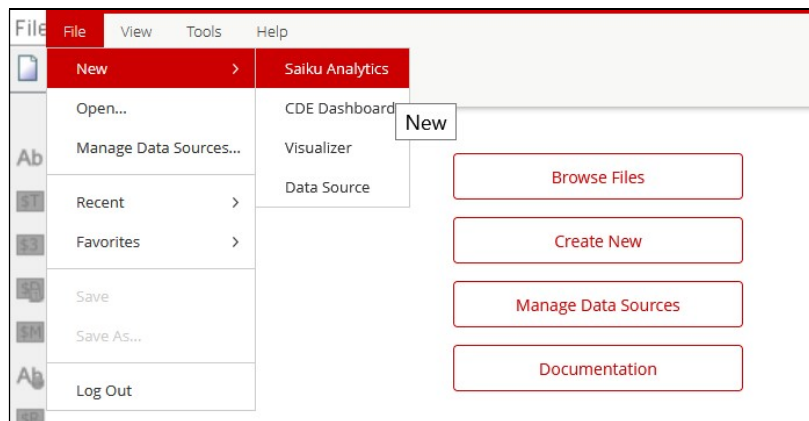


Figura 3. Saiku en la barra de herramientas de Pentaho v9.1.

2.5 Tableau

Tableau es una herramienta utilizada para la visualización y análisis de los datos. La misma puede ser consultada en [20].

2.6 Power BI

Como en el caso de Tableau, Power BI también es una herramienta utilizada para la visualización y análisis de los datos. La misma puede ser consultada en [21].

2.7 Visualizer

Visualizer, como las herramientas de los puntos 2.5 y 2.6, también es una herramienta para análisis y visualización de datos, la versión 6.04 puede ser descargada de la página disponible en [22].

Para instalar *Visualizer*, una vez descargado el archivo `Visualizer-6.04.zip` desde [20], simplemente hay que descomprimirlo y colocar la carpeta **visualizer** en el directorio `<Pentaho Home>/pentaho-solutions/system`. Luego, al iniciar el servidor e ingresar al sistema, aparecerá el ícono de **Visualizer** en la barra de herramientas, como se observa en la Figura 4. Observar que la versión de Pentaho que se muestra en la Figura 3 es la 9.1, por lo que el menú de la versión 9.4 podría diferir.

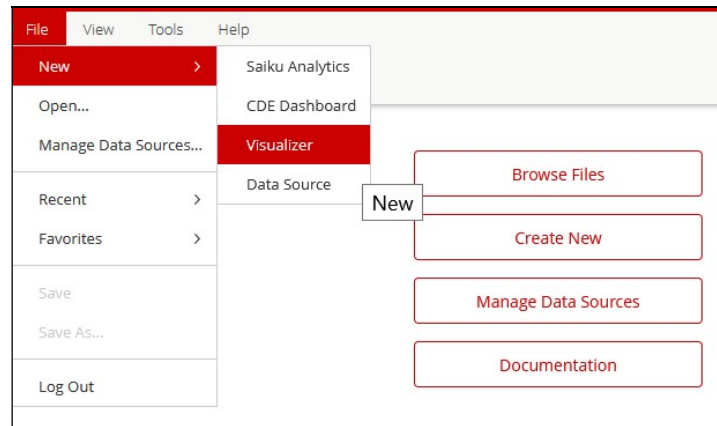


Figura 4. *Visualizer* en la barra de herramientas de Pentaho v9.1.

Referencias

El último acceso de todas las referencias fue realizado en **Abril de 2024**.

- [1] Sitio Pentaho Community Edition:
<https://www.hitachivantara.com/pentaho/pentaho-plus-platform/data-integration-analytics/pentaho-community-edition.html>
- [2] PostgreSQL: <https://www.enterprisedb.com/downloads/postgres-postgresql-downloads>
- [3] MongoDB: <https://www.mongodb.com/try/download/community>
- [4] Tutorial de MongoDB: <https://www.mongodb.com/docs/manual/administration/install-community/#std-label-install-mdb-community-edition>
- [5] JSON: <https://www.json.org/json-es.html>
- [6] <https://www.hitachivantara.com/en-us/pdf/implementation-guide/three-steps-to-install-pentaho-data-integration-ce.pdf>
- [7] Java Development Kit versión 17:
<https://www.oracle.com/java/technologies/downloads/#jdk17-windows>
- [8] Manual de instalación y requerimientos del sistema:
<https://medium.com/@SaiParvathaneni/installation-guide-pentaho-aka-kettle-pdi-for-etl-739c9ee1a635>
- [9] Data Source:
<https://docs.hitachivantara.com/r/en-us/pentaho-data-integration-and-analytics/9.4.x/mk-95pdia003/use-a-pentaho-repository-in-pdi/use-the-repository-explorer>
- [10] Kettle Tutorial:
<https://docs.hitachivantara.com/r/en-us/pentaho-data-integration-and-analytics/9.4.x/mk-95pdia003>
- [11] Documentación Mondrian:
<http://mondrian.pentaho.com/documentation/olap.php>
- [12] Schemas Mondrian:
<http://mondrian.pentaho.com/documentation/schema.php>
- [13] Schema Workbench:
<https://mondrian.pentaho.com/documentation/workbench.php>
- [14] <https://docs.hitachivantara.com/r/en-us/pentaho-data-integration-and-analytics/9.4.x/mk-95pdia003/pdi-transformation-steps/modified-java-script-value/sample-transformations>
- [15] <https://docs.hitachivantara.com/r/en-us/pentaho-data-integration-and-analytics/9.4.x/mk-95pdia008/create-queries/create-mdx-queries>
- [16] <https://mondrian.pentaho.com/documentation/mdx.php>
- [17] Documentación Pentaho Report Designer:
<https://docs.hitachivantara.com/r/en-us/pentaho-data-integration-and-analytics/9.4.x/mk-95pdia008/using-pentaho-report-designer>
- [18] Data Source - Pentaho Report Designer:
<https://docs.hitachivantara.com/r/en-us/pentaho-data-integration-and-analytics/9.4.x/mk-95pdia008/output-parameterization>
- [19] <https://eva.fing.edu.uy/mod/url/view.php?id=107013>
- [20] <https://www.tableau.com/es-es/products>
- [21] <https://powerbi.microsoft.com/es-es/downloads/>
- [22] <https://sourceforge.net/projects/datafor-visualizer/files/Visualizer/>
- [23] <https://www.hitachivantara.com/es-latam/products/pentaho-platform/data-integration-analytics/pentaho-community-edition.html>