

Diseño y Construcción de Data Warehouse

Edición 2017

Instalación de Herramientas

Índice

ÍNDICE	2
1 INTRODUCCIÓN	3
2 BASE DE DATOS	3
2.1 PostgreSQL	3
2.2 PostGIS	3
3 SISTEMA DE DATA WAREHOUSE	3
3.1 Pentaho Business Intelligence Community	3
3.1.1 Plataforma Business Intelligence	4
3.1.2 Herramienta de Extracción, Transformación y Carga (ETL)	6
3.1.3 Servidor OLAP	7
3.1.4 Herramienta de Diseño de Reportes	8
3.1.5 Análisis OLAP: Saiku	9
4 REFERENCIAS.....	11

1 Introducción

El presente documento pretende ser una guía para la instalación de las herramientas a utilizar en el proyecto del curso.

Las herramientas propuestas son las siguientes:

Base de Datos:

- **PostgreSQL**
- **PostGIS**

Sistema de Data Warehouse:

- **Pentaho:**
 - o *Plataforma Business Intelligence: Pentaho BI Analytic.*
 - o *Procesos de Extracción, Transformación y Carga: Kettle.*
 - o *Servidor OLAP: Mondrian, Schema Workbench.*
 - o *Diseño de Reportes: Pentaho Report Designer.*
 - o *Análisis OLAP: JPivot, Saiku.*

Las herramientas propuestas para el Sistema de Data Warehouse son productos pertenecientes al proyecto *Pentaho Business Intelligence Community* [1].

2 Base de Datos

2.1 PostgreSQL

Se utilizará la versión 9.6 de PostgreSQL disponible para descargar en [2].

2.2 PostGIS

PostGIS es una extensión que convierte el sistema de base de datos PostgreSQL en una base de datos espacial. La combinación de ambos permite el almacenamiento, gestión y mantenimiento de datos espaciales. En el momento de la instalación de PostgreSQL, cuando se seleccionan las aplicaciones que se van a instalar, se debe activar la casilla de *PostGIS Bundle for PostgreSQL* que se corresponde con los paquetes básicos de PostGIS.

3 Sistema de Data Warehouse

3.1 Pentaho Business Intelligence Community

Todos los productos de Pentaho están desarrollados en Java, por lo que, para trabajar, es necesario tener instalado el Java Development Kit (JDK) [3].

Agregar la variable de entorno JAVA_HOME en caso de ser necesario, como se muestra en la Figura 1, con el valor correspondiente al directorio de instalación del jdk. Recordar que las variables de entorno en Windows son configurables en *Panel de Control -> Sistema -> Configuración Avanzada -> Variables de Entorno*.

3.1.1 Plataforma Business Intelligence

Se sugiere utilizar la versión 7 de Pentaho BI, disponible en [4]. Además, se debe utilizar la versión 8 del jdk disponibles en [5].

1) Instalación

El archivo descargado es un .zip que contiene la carpeta *pentaho-server*, la cual llamaremos *Pentaho Home*.

La instalación consiste simplemente en descomprimir el archivo **pentaho-server-ce-7.0.0.0-25.zip** y copiar dicha carpeta en el directorio de instalación. Es recomendable que el directorio de instalación sea el siguiente:

- Para Windows:
c:\Archivos de Programas\pentaho
- Para Linux:
/opt/pentaho

2) Servidor

El servidor de Pentaho se puede iniciar y detener manualmente mediante los script **start-pentaho.bat** y **stop-pentaho.bat** respectivamente que se encuentran en el *Pentaho Home*.

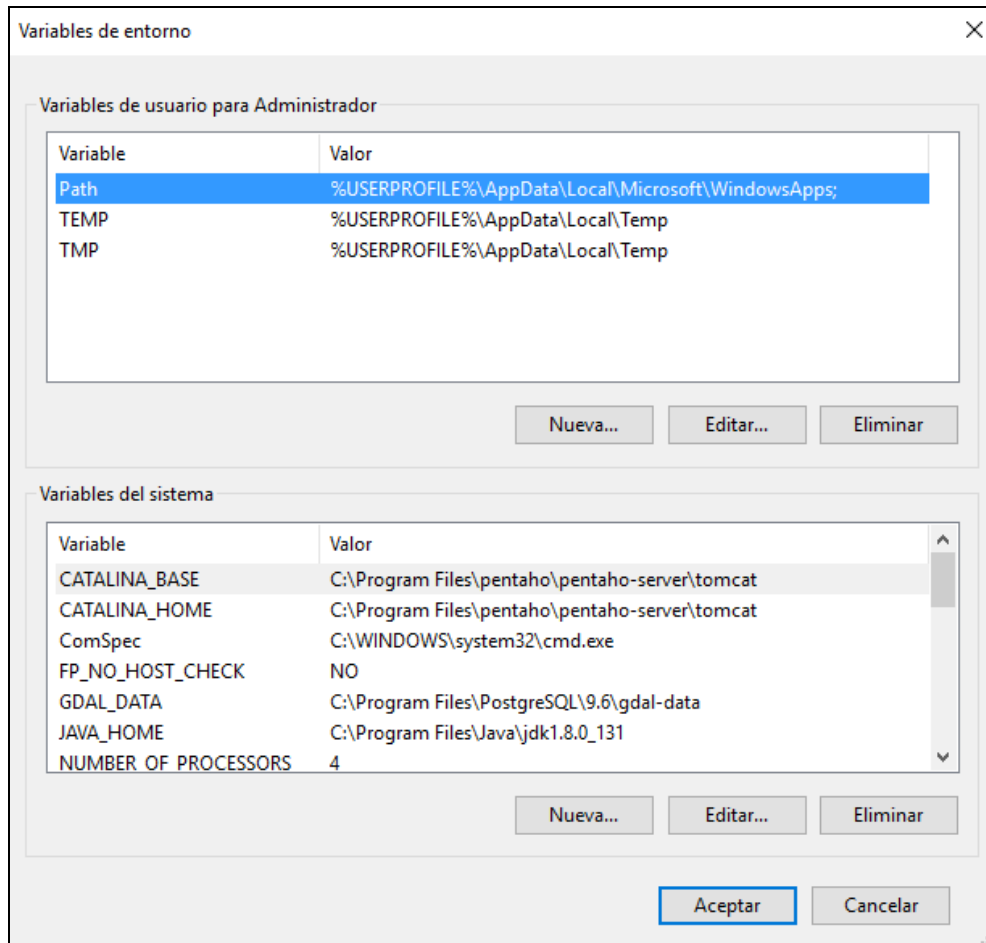


Figura 1. Variables de entorno.

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*. 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 2 se presenta la pantalla de inicio de Pentaho.

Se puede ingresar como usuario administrador, cuyo usuario y clave son *admin* y *password* respectivamente. También es posible ingresar como usuario de negocio, cuyo usuario y clave son *suzy* y *password* respectivamente. Luego del login como usuario Administrador, se muestra la página de bienvenida que se presenta en la Figura 3.

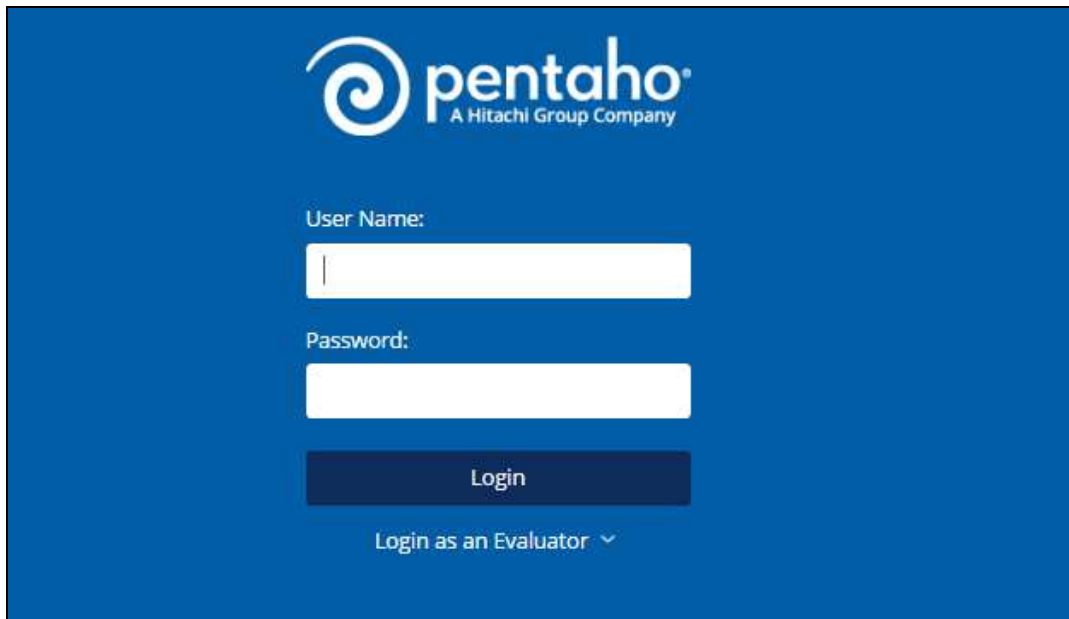


Figura 2. Pantalla de inicio de Pentaho.

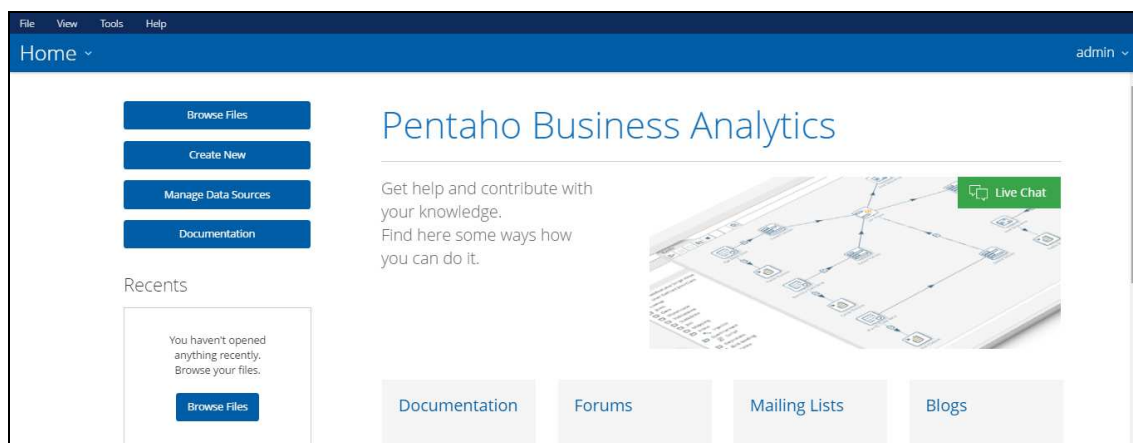


Figura 3. Página de bienvenida de Pentaho para el usuario Administrador.

3.1.2 Herramienta de Extracción, Transformación y Carga (ETL)

La herramienta proporcionada por Pentaho para los procesos de extracción, transformación y carga es el *Kettle*, del cual utilizaremos la versión 7.0, disponible en [6].

La instalación consiste en:

- descomprimir el archivo descargado **pdi-ce-7.0.0.0-25.zip**, el cual contiene una carpeta llamada *data-integration*
- la carpeta *data-integration* se debe ubicar en el directorio de instalación que se desee. Es recomendable utilizar el mismo directorio de instalación que el servidor (*c:\Archivos de Programas\pentaho*), así todos los productos de Pentaho se encuentran en un mismo directorio.

El *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*. Se podría crear un acceso directo a dicho archivo para su ejecución. Una vez iniciado se muestra la página de bienvenida que se presenta en la Figura 4.

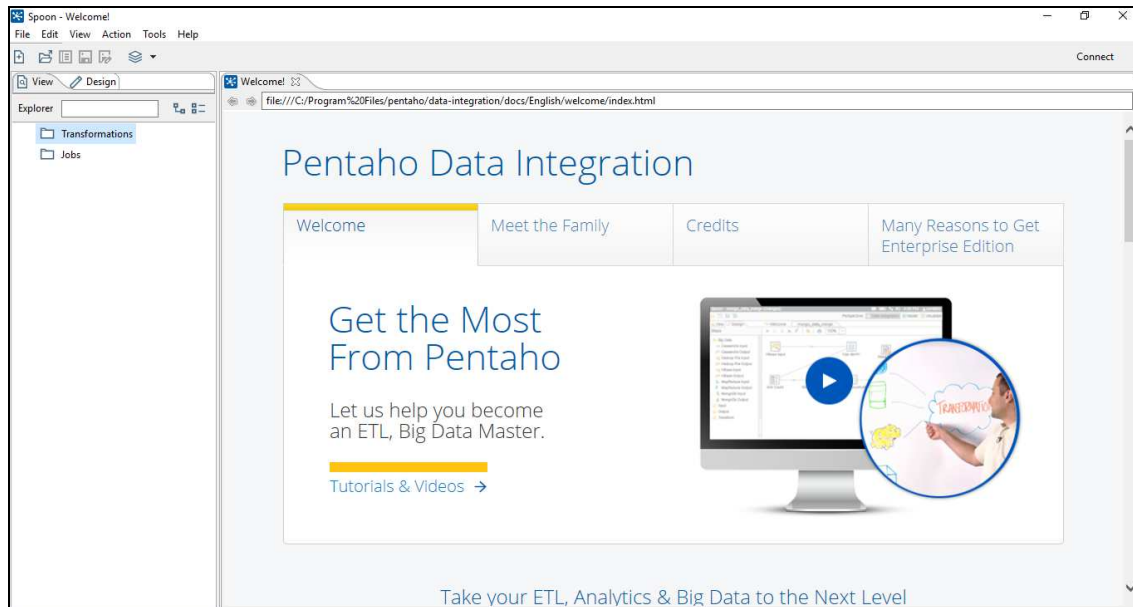


Figura 4. Página de inicio de Kettle.

3.1.3 Servidor OLAP

El servidor OLAP que ofrece Pentaho es el *Mondrian* [7], 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 el *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 análisis. La especificación detallada sobre el diseño y estructura que deben tener los *Schemas* se encuentra en [8].

Existe una herramienta gráfica para el diseño de *Schemas* denominada *Schema Workbench*, que actualmente dispone de la versión 3.13 para descargar en [9]. La instalación consiste simplemente en descomprimir el archivo **psw-ce-3.13.0.0-25.zip** descargado, y ubicar la carpeta *schema-workbench* en el directorio de instalación que se desee. Es recomendable utilizar el mismo directorio de instalación que el servidor (*c:\Archivos de Programas\pentaho*).

Se trata de una aplicación de escritorio y se inicia mediante el archivo *workbench* (**workbench.bat** para Windows y **workbench.sh** para Linux), ubicado en el *schema-workbench*. Se podría crear un acceso directo a dicho archivo para facilitar su ejecución.

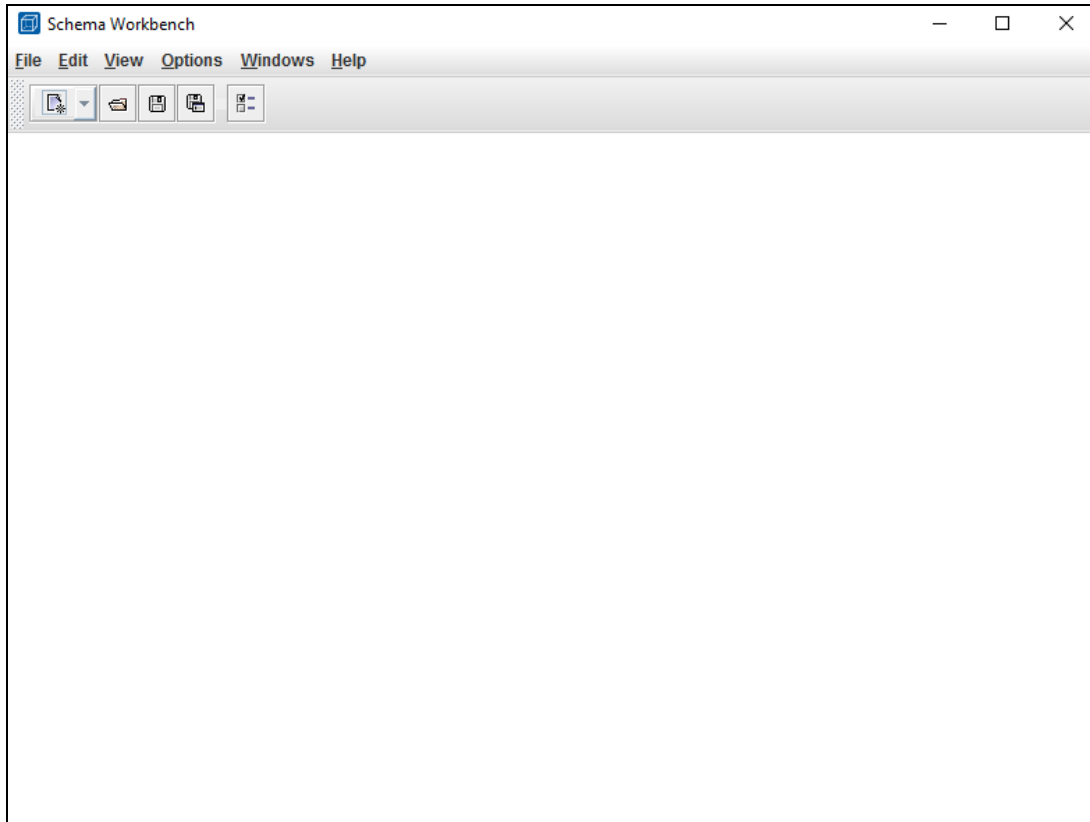


Figura 5. Pantalla de inicio de Schema-workbench.

3.1.4 Herramienta de Diseño de Reportes

Para el diseño de reportes se utilizará la versión 5.0.1 del *Pentaho Report Designer*, disponible para descargar en [10].

La instalación consiste simplemente en descomprimir el archivo **prd-ce-5.0.1-stable.zip**, el cual contiene una carpeta llamada *report-designer*, y ubicarla en el directorio de instalación. Es recomendable utilizar el mismo directorio de instalación que el servidor (*c:\Archivos de Programas\pentaho*).

El *Pentaho Report Designer* también es una aplicación de escritorio y se inicia mediante el archivo *report-designer* (**report-designer.bat** para Windows y **report-designer.sh** para Linux) ubicado en el *report-designer*. Se podría crear un acceso directo a dicho archivo para su ejecución. Una vez iniciado se presenta la página de bienvenida que se muestra en la Figura 6.

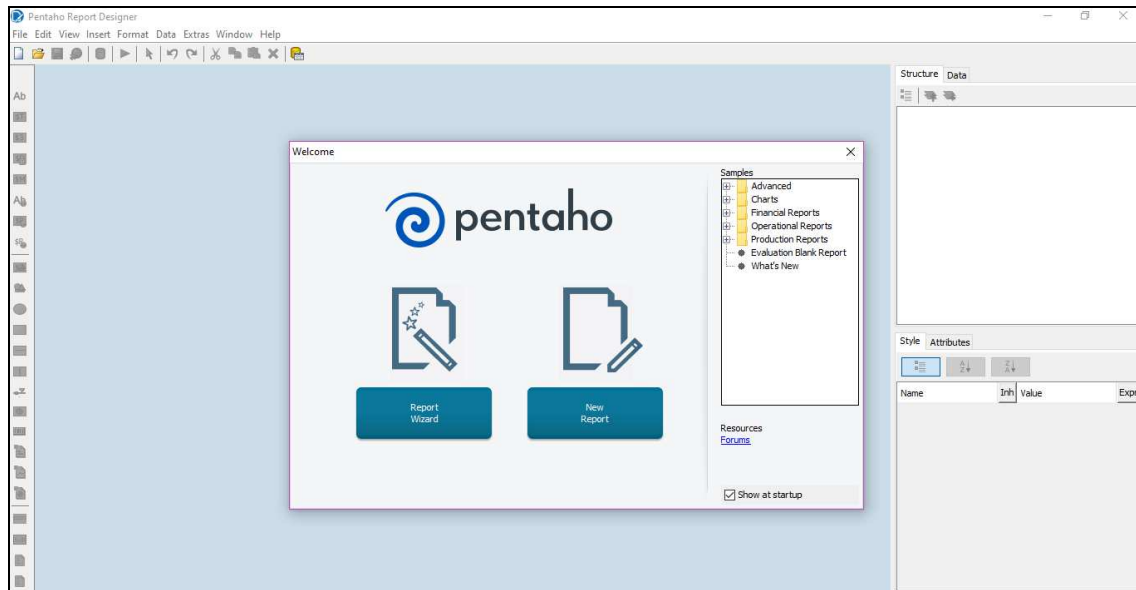


Figura 6. Pantalla de inicio de Report-designer.

3.1.5 Análisis OLAP: Saiku

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 [11], actualmente en la versión 3.11.1, seleccionando el ítem *Saiku Analytics*.

Si bien Pentaho ya cuenta con un componente para hacer consultas OLAP, llamado *JPivot* (que viene incluido en la instalación), el mismo ya no es mantenido y se prevé que sea removido en futuras versiones.

Para instalar el plugin de Saiku, una vez descargado el archivo **saiku-plugin-p7-3.11-SNAPSHOT.zip**, simplemente hay que descomprimirlo y colocar la carpeta *saiku* 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 7.

Se debe tener en cuenta que para poder ejecutar *Saiku* es necesario contar con una licencia. Para esto, se debe Acceder al sitio web *Meteorite Consulting's license* [12] y presionar *Sign up* para obtener una cuenta gratis.

Una vez que la cuenta ha sido validada, es necesario loguearse en el sistema y hacer click en el botón *CREATE NEW LICENSE*.

El tipo de la licencia debe ser seteado como *COMMUNITY EDITION*. Luego se debe hacer click en el botón *SAVE* y la página se modificará mostrando un link a *Download License*. Hacer click en el link y el archivo con la licencia se descargará. El archivo debe ser renombrado a *"license.lic"* y luego debe ser copiado en la carpeta *Saiku* en el directorio `<Pentaho Home>/pentaho-solutions/system`. Finalmente, se debe reiniciar Pentaho.

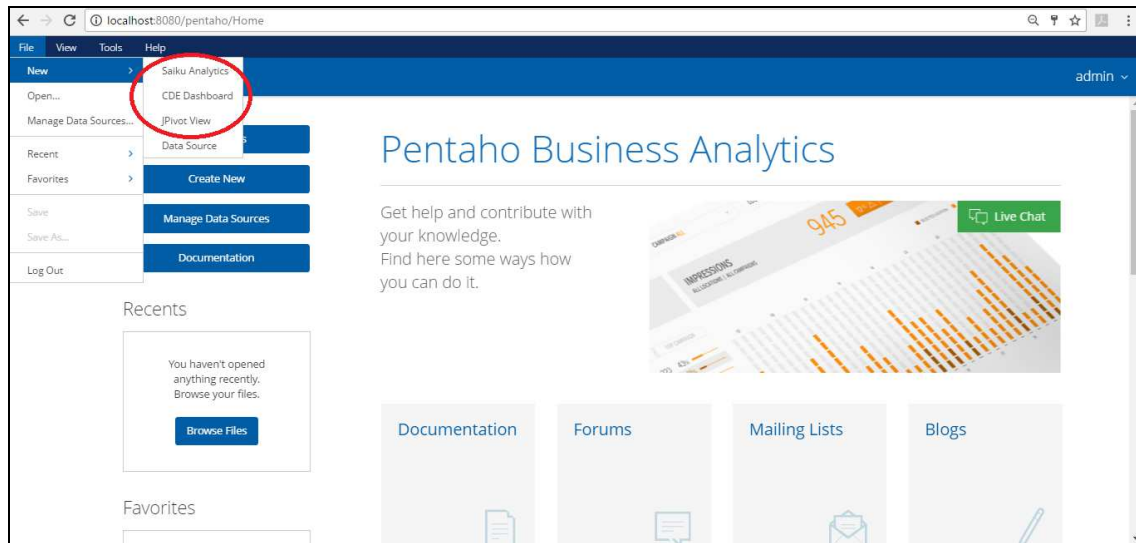


Figura 7. Saiku en la barra de herramientas de Pentaho.

4 Referencias

El último acceso de todas las referencias fue realizado en Abril de 2016.

[1] Pentaho Business Intelligence Community:
<http://community.pentaho.com/>

[2] PostgreSQL:
<https://www.postgresql.org/download/>

[A] PostGIS:
<https://postgis.net/docs/manual-dev/postgis-es.html>

[3] Java Development Kit versión 7:
<http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html>

[4] Pentaho BI:
<https://sourceforge.net/projects/pentaho/files/Business%20Intelligence%20Server/7.0/>

[5] Java Development Kit versión 8:
<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

[6] Kettle:
<https://sourceforge.net/projects/pentaho/files/Data%20Integration/7.0/>

[7] Documentación Mondrian:
<http://mondrian.pentaho.com/documentation/olap.php>

[8] Schemas Mondrian:
<http://mondrian.pentaho.com/documentation/schema.php>

[9] Schema Workbench:
<https://sourceforge.net/projects/mondrian/files/schema%20workbench/3.13.0/>

[10] Pentaho Report Designer:
<https://sourceforge.net/projects/jfreereport/files/04.%20Report%20Designer/5.0.1-stable/>

[11] Saiku Analytics: <http://www.pentaho.com/marketplace/>

[12] <http://licensing.meteorite.bi/login>