***MSIG***

***Materia: DATAWAREHOUSE***

***Manual de Usuario***

***PENTAHO***

***(Integración de Datos)***

***Grupo:***

Miriam Vega

Shiley Villón

Freddy Pilligua

Ronald Ponguillo

***11 ava. Promoción***

***2009 - 2010***

***Qué es PENTAHO ?***

Es la plataforma Open Source Business Intelligence. cubre muy amplias necesidades de Análisis de los Datos y de los Informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho es una solución muy flexible para cubrir una amplia gama de necesidades empresariales – tanto las típicas como las sofisticadas y especificas al negocio.

Los módulos de la plataforma Pentaho BI son:

 **Reporting -** un modulo de los informes ofrece la solución adecuada a las necesidades de los usuarios. Pentaho Reporting es una solución basada en el proyecto JFreeReport y permite generar informes ágil y de gran capacidad. Pentaho Reporting permite la distribución de los resultados del análisis en múltiples formatos - todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Los reportes Pentaho permiten también programación de tareas y ejecución automática de informes con una determinada periodicidad.

* **Análisis** - Pentaho Análisis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs), generadas por Mondrian y JPivot, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación. Los datos pueden ser representados en una forma de SVG o Flash, los dashboards widgets, o también integrados con los sistemas de mineria de datos y los portales web (portlets). Además, con el Microsoft Excel Analysis Services, se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server Mondrian).

 **Dashboards -** todos los componentes del modulo Pentaho Reporting y Pentaho Análisis pueden formar parte de un Dashboard. En Pentaho Dashboards es muy fácil incorporar una gran variedad en tipos de gráficos, tablas y velocímetros (dashboard widgets) e integrarlos con los Portlets JSP, en donde podrá visualizar informes, gráficos y análisis OLAP.

 **Data Mining -** análisis en Pentaho se realiza con una herramienta WeKa.

 **Integración de Datos -** se realiza con una herramienta Kettle ETL (Pentaho Data Integration) que permite implementar los procesos ETL. Últimamente Pentaho lanzó una nueva versión - PDI 3.0 – que marcó un gran paso adelante en OSBI ETL y que hizo Pentaho Data Integration una alternativa interesante para las herramientas comerciales.

**Módulo de Integración de Datos**

**Pentaho** ofrece integración de datos de gran alcance de Extracción, Transformación y Carga (ETL) utilizando la capacidad de un innovador enfoque impulsado por los metadatos. With an intuitive, graphical, drag and drop design environment, and a proven, scalable, standards-based architecture, Pentaho Data Integration is increasingly the choice for organizations over traditional, proprietary ETL or data integration tools.

***INICIANDO EL MODULO:***

Seleccionamos:

Inicio 🡪 Pentaho Interprise Edition 🡪 Design Tools 🡪 Data Integration (Spoon o Kettle)

Spoon spoon.png es el diseñador gráfico de transformaciones y trabajos asociados con el sistema de ETL Pentaho Data Integration, también conocido como Kettle.

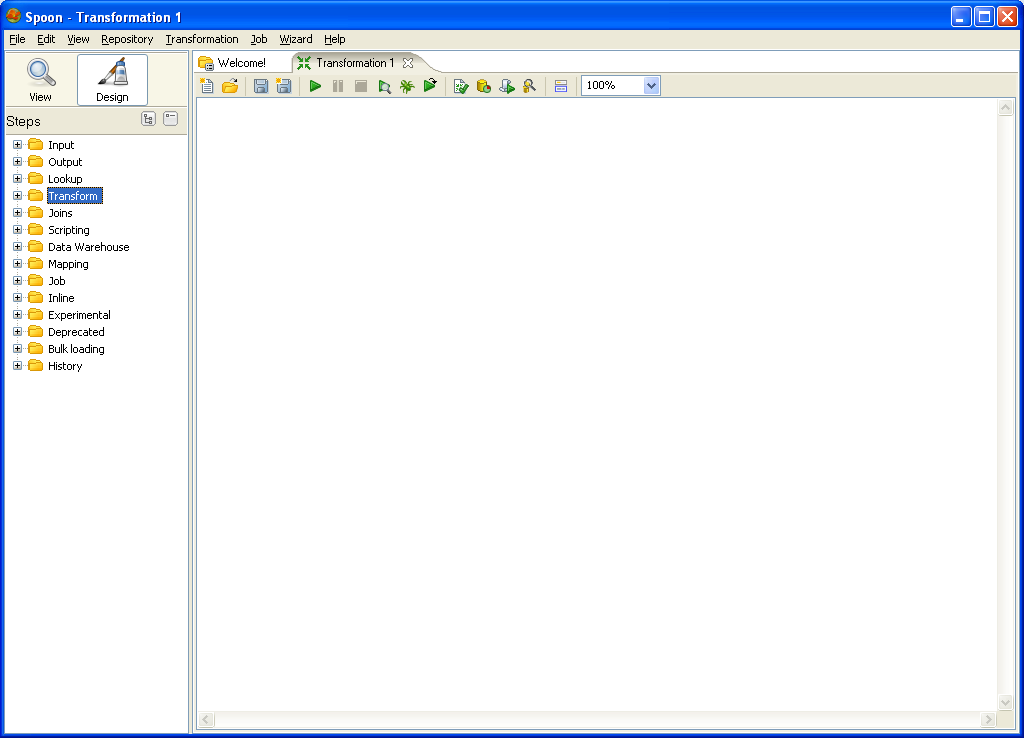
****

***CREANDO UNA TRANSFORMACIÓN Ó TRABAJO:***

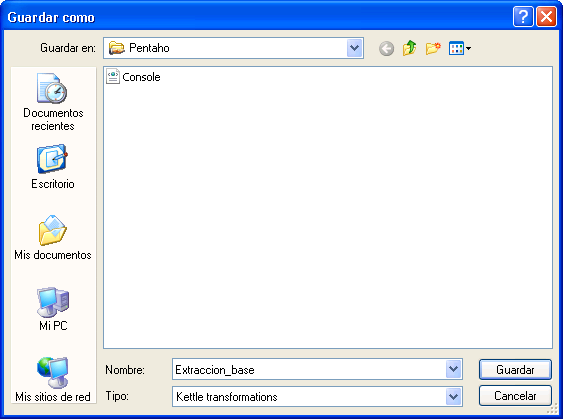
Para crear un nuevo Proyecto de Transformación seleccionamos:

File 🡪 New 🡪 Transformation

El cual nos permitirá ingresar a la siguiente pantalla:

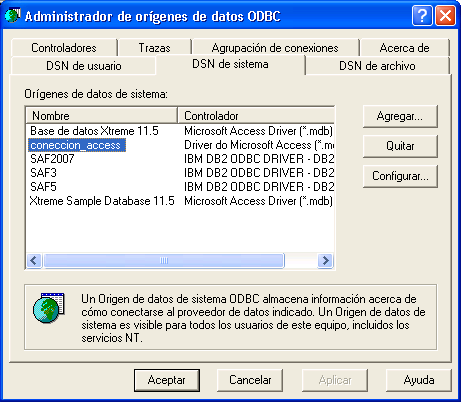
******

Luego lo almacenamos asignándole un nombre al proyecto de transformación:



**AGREGANDO ACCESO A LA BASE DE DATOS EN ACCESS**

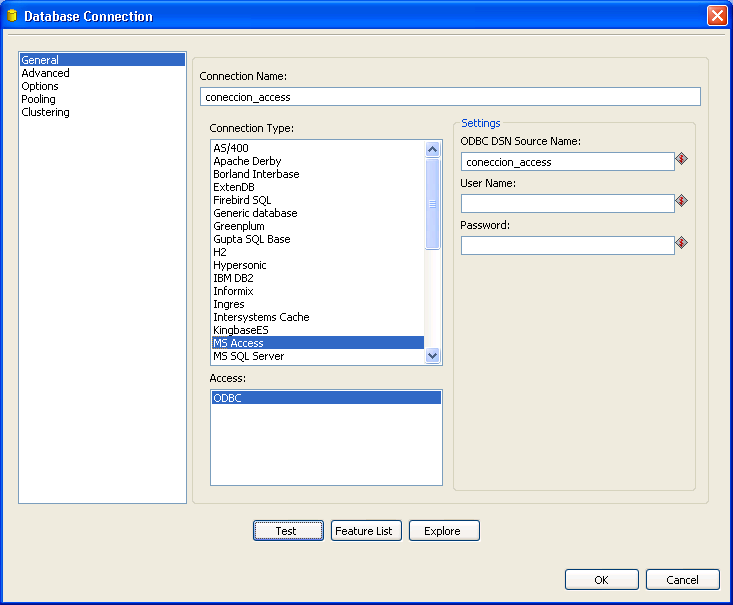
Luego de haber creado una nueva transformación se debe configurar los acceso o conexiones a Bases de Datos. En este caso creamos una conexión a ACCESS configurándolo en el ODBC de nuestra PC con el nombre de **coneccion\_access**.



**DEFINICION DE CONEXIÓN DE LA BASE DE ACCESS**

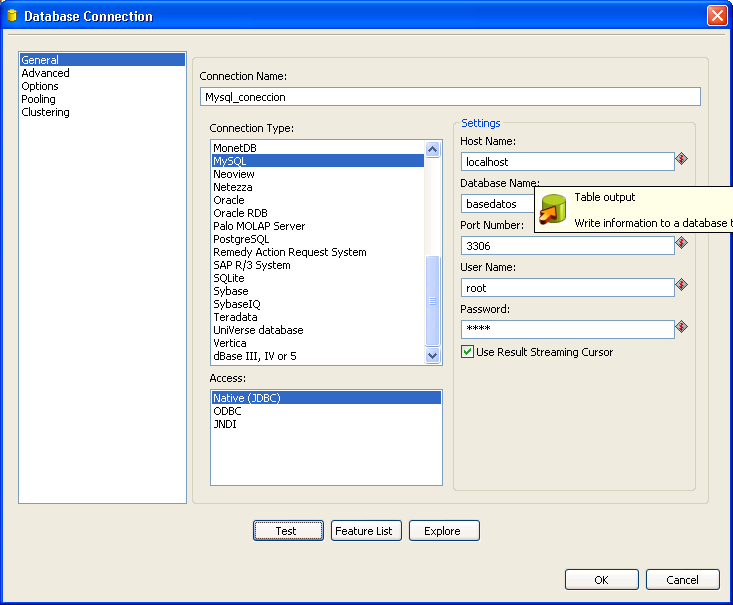
Creamos la conexión en el Spoon de Transformación de la siguiente forma:

1. Damos click en el icono **VIEW**
2. Con click derecho en **Database Connection**, seleccionamos **New.**
3. Asignamos un nombre a la conexión para la Extracción de datos a la base de datos de Access (base de donde se va a extraer la información) , mediante la siguiente pantalla:



**DEFINICION DE CONEXIÓN DE LA BASE DE MySQL**

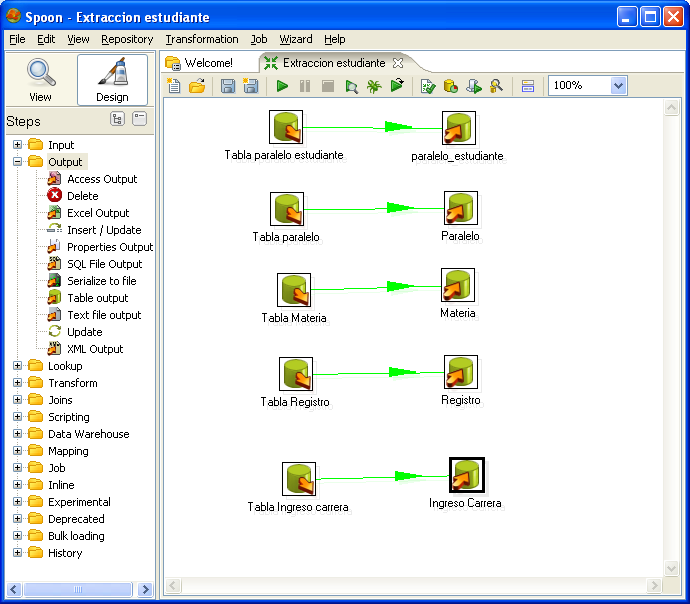
Para crear la conexión a **MySql**, se realiza los mismos pasos que se realizó para crear la conexión de Access, y se mostrara la siguiente pantalla:



Previamente se debe crear la base de datos en **MySql**, con la misma estructura de la base de **Access**.

**DEFINICION DE TABLAS DE ENTRADAS (INPUT)**

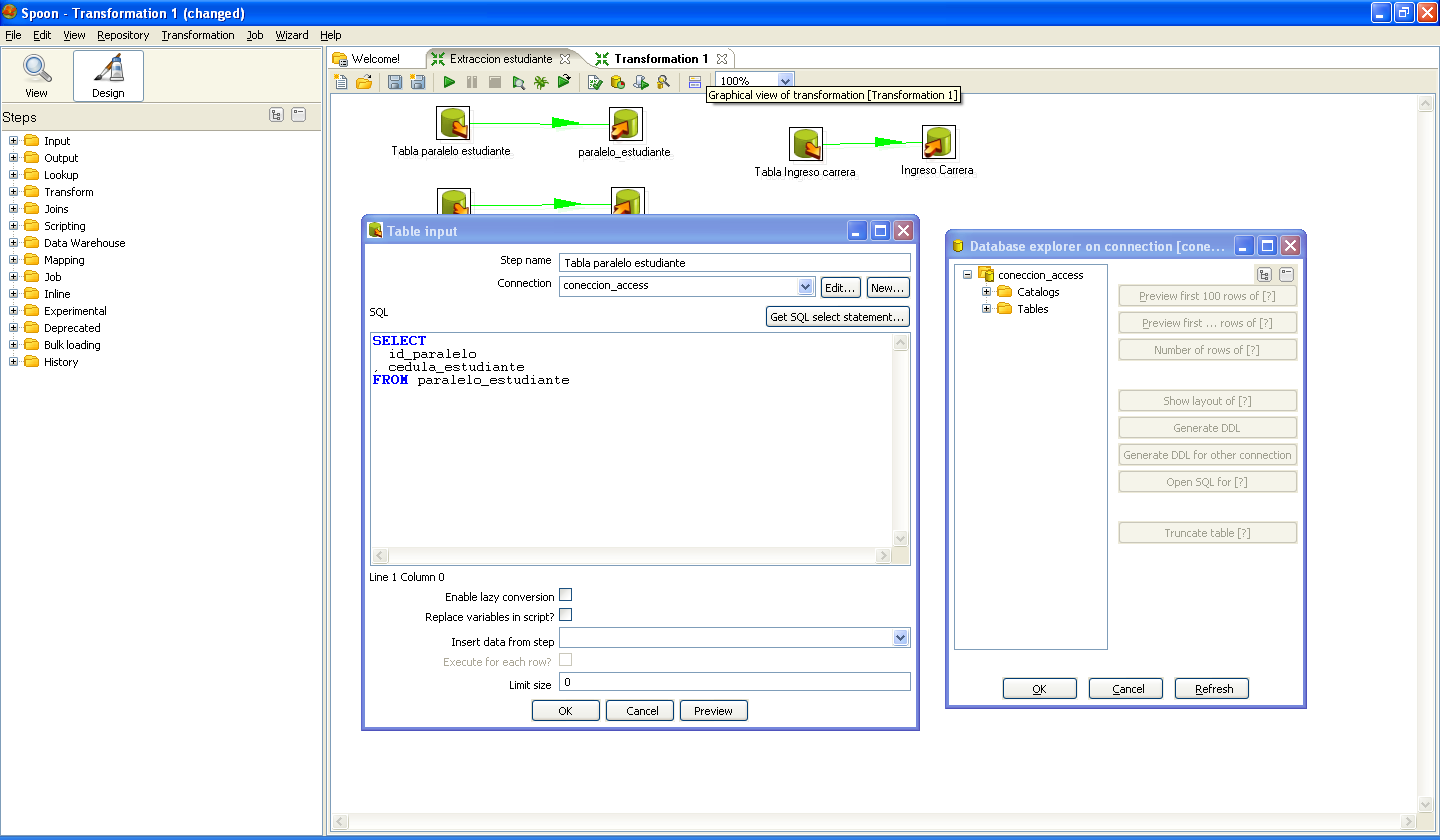
Para la definición de las tablas de Entradas debemos seleccionar y arrastrar del grupo de **INPUT** un elemento de tipo **Table Input**.



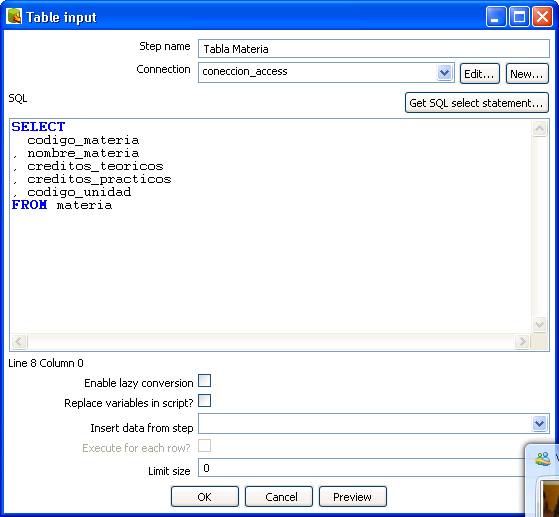
**INPUT**

Una vez agregado el elemento, debemos configurar tanto su conexión a la base de datos, tabla asociada y la sentencia SQL a ejecutarse para la extracción de la información requerida.

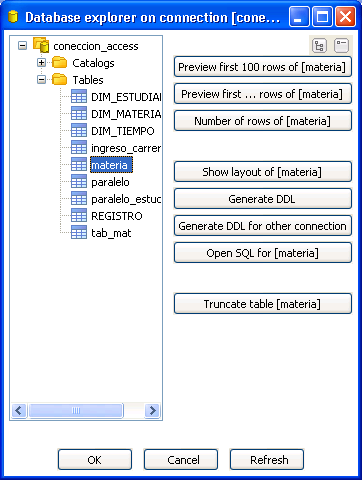
Para la asignación de datos de configuración debemos seleccionar el elemento INPUT con click derecho en la opción **edit step** del elemento de la tabla de entrada, debemos definir la información requerida:



Además debemos asignar un nombre de la tabla de extracción de datos, el tipo de conexión y las sentencia **select** que se desea ejecutar para obtener los datos de la base Access. Esto es de la siguiente forma:

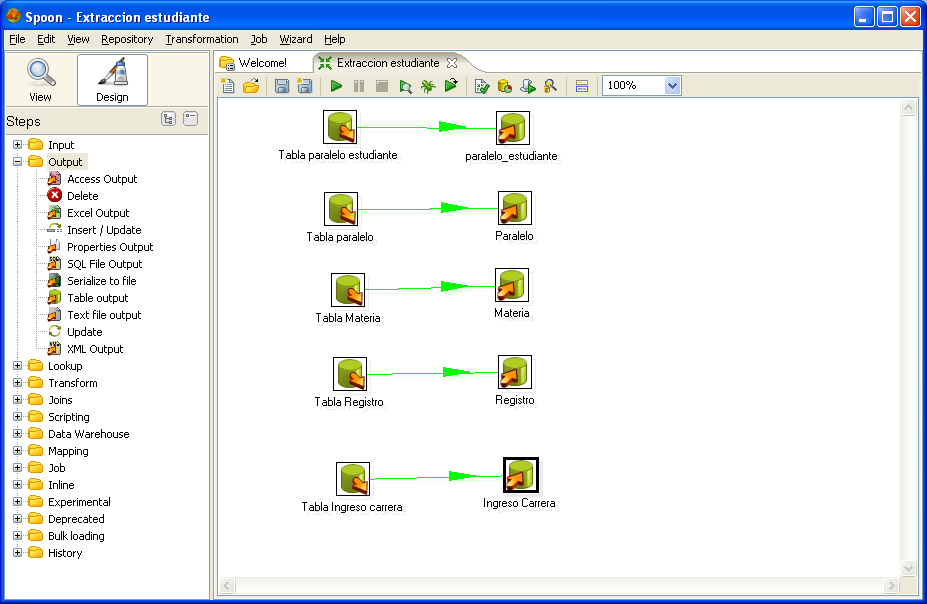


Con la opción de **Get SQL select statement** se puede obtener una prueba de la información que esta por obtenerse, así:



**DEFINICION DE TABLAS DE SALIDA (OUTPUT)**

De la misma forma que se hizo en la definición de tablas de extracción, debemos agregar un elemento de salida (output), que se encontrará en el grupo de **OUTPUT.**



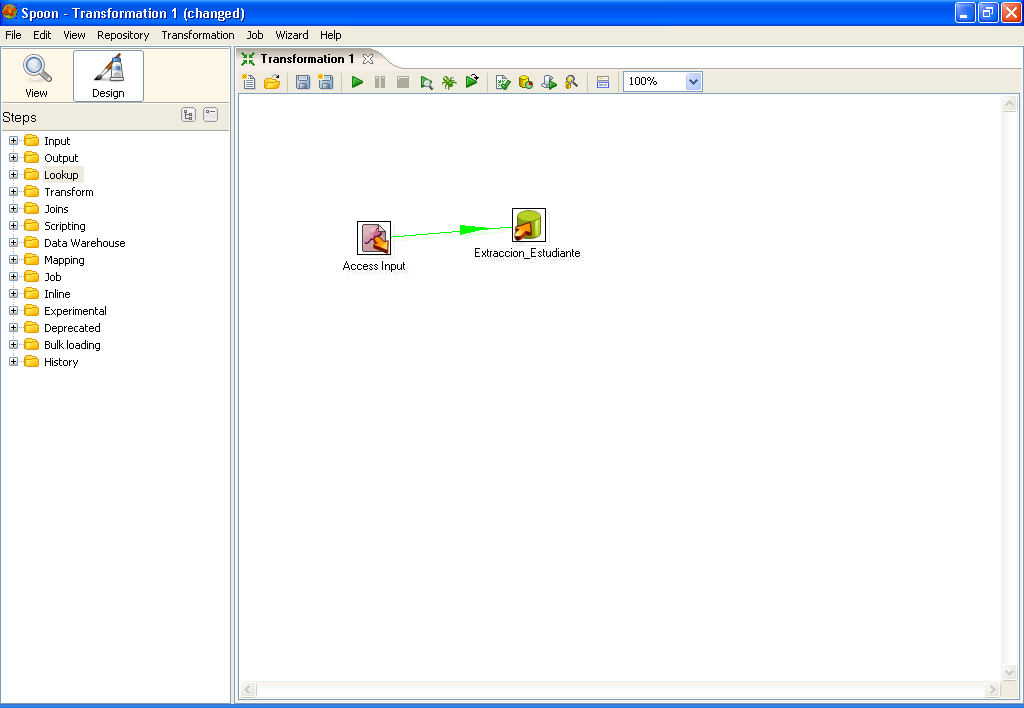
**OUTPUT**

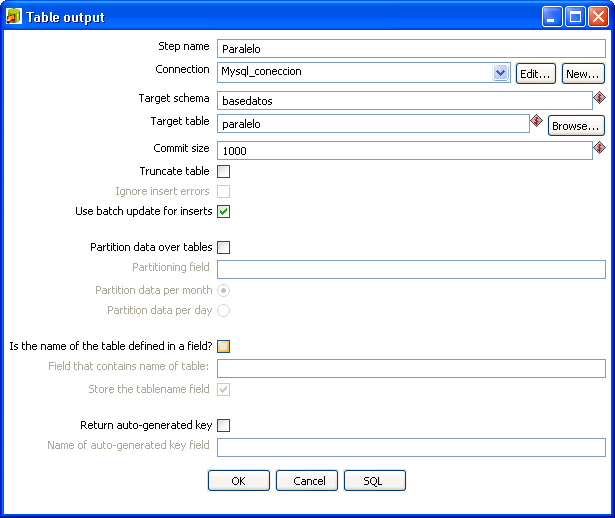
De igual forma se configura el elemento de carga; asignándole un nombre al elemento, definiendo la conexión a MySql y los datos que se van a cargar.

**ASOCIACION DE LA ENTRADA Y SALIDA DE DATOS, PARA LA TRANSFROMACIÓN**

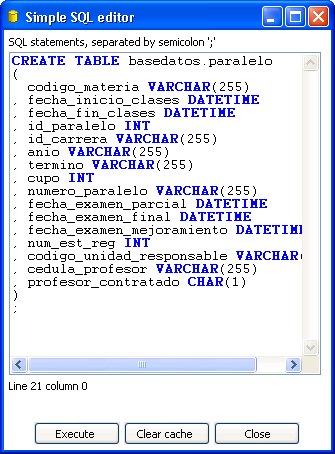
Se debe establecer la relación de transformación de datos, tanto de la tabla de extracción y carga de los datos. Esto se lo realiza de la siguiente forma:

En el icono VIEW, debemos seleccionar





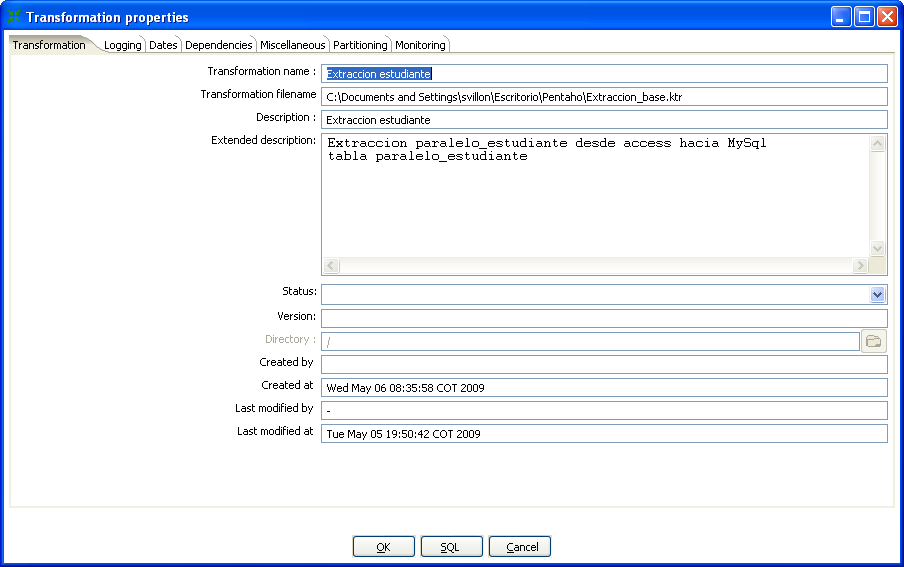
Creación previa de la tabla en MySql, para ser visualizada en la definición de los elementos de salida.



**TRANSFORMACIÓN**

Una vez definidos los elementos de extracción (input), los elementos de carga (output) y la asociación entra tablas para la transformación; solo es necesario ejecutar la Transformación para concluir con el procedimiento.

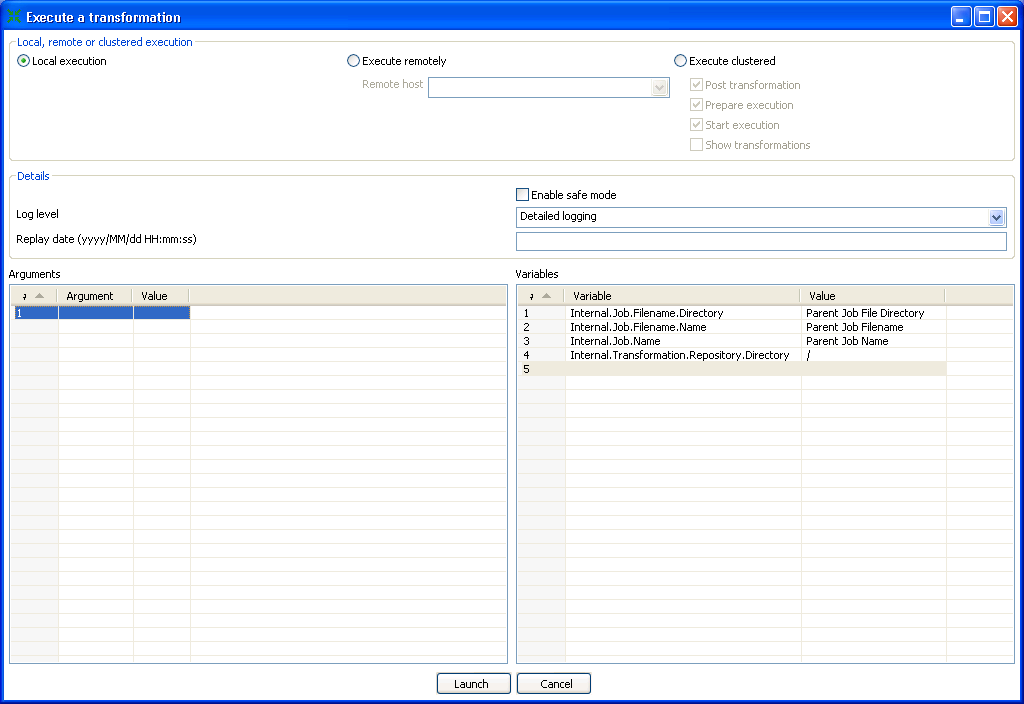
Previo a la ejecución es necesario configurar la siguiente información para el proceso de extracción de Estudiantes, mediante la selección de **VIEW 🡪 TRANSFORMATIONS 🡪 EXTRACCION ESTUDIANTE** (es el nombre que hemos definido a la extracción de datos) 🡪 **SETTINGS**. Luego se mostrará la siguiente pantalla donde se



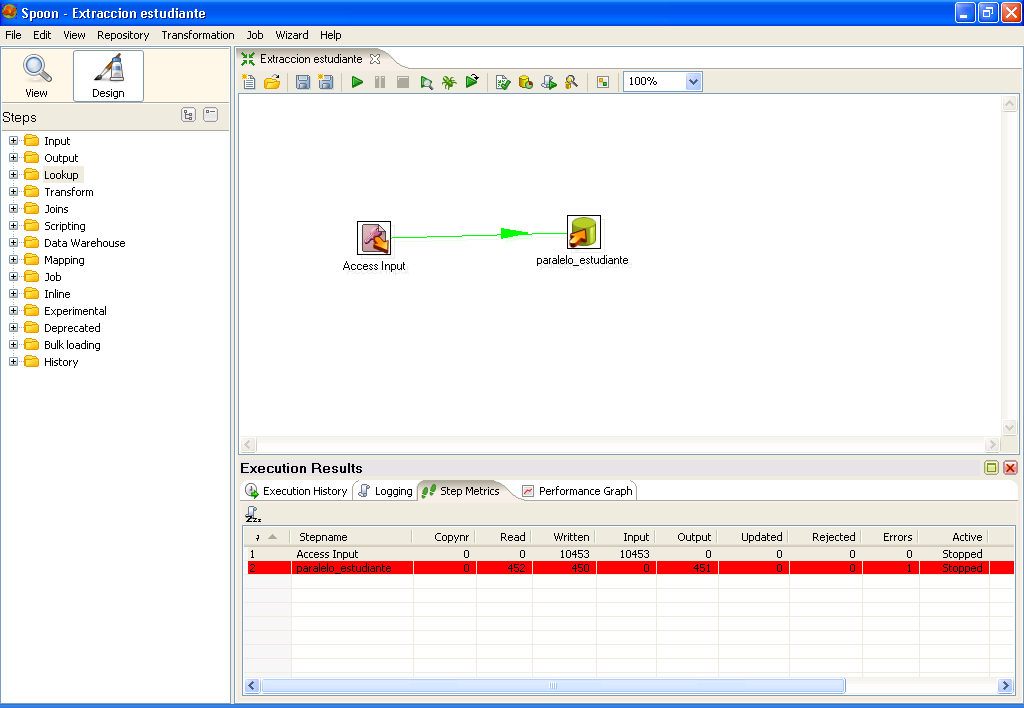
**EJECUCIÓN DE LA TRANSFORMACIÓN:**

Para la ejecución de de la transformación de datos se debe seleccionar el botón **Run this transformation or Job**.

Luego procedemos a ejecutar la transformación mediante el botón **Launch** como se muestra en la siguiente pantalla:



Posteriormente se presentara en la parte inferior de la pantalla la información de los datos procesado en cada una de los elementos INPUT y OUTPUT, tal como se muestra a continuación:



En la parte inferior se muestra los resultados de la transformación

