



LABORATORIO N° 02

PROCESO ETL DESDE UN SISTEMA TRANSACCIONAL A UNA BASE DE DATOS DIMENSIONAL (DATA MART)

1. OBJETIVOS:

Luego de completar este laboratorio usted será capaz de:

- 📁 Implementar la base de datos dimensional en mysql
- 📁 Realizar el proceso de extracción de datos de un sistema transaccional (OLTP)
- 📁 Realizar la transformación y limpieza de datos de las dimensiones
- 📁 Realizar la carga de datos al Data Mart utilizando PDI PENTAHO.

2. PLAN DE CLASE:

2.1. MARCO TEÓRICO

PENTAHO es una suite de herramientas de inteligencia de negocios que tiene dos versiones, la versión comercial y la versión de código abierto.

BASE DE DATOS DIMENSIONAL

Una base de datos multidimensional es aquella base de datos en donde su información se almacena en una estructura multidimensional, conformada por tablas de dimensiones y tablas de hechos.

- Tabla de dimensiones
Una dimensión representa información de una entidad de negocio. Las tablas de dimensión, contienen información que permiten definir cómo están los datos organizados lógicamente y describen la forma para analizar el contexto del negocio; además se caracterizan por contener datos cualitativos.
- Tabla de hechos
Un hecho representa un evento de negocio. Contiene información de datos reales, de medidas, de hechos en bruto. Son tablas en las que cada tupla contiene información de un hecho: las medidas asociadas a ese hecho.

Las tablas de Hechos son la base central de los esquemas multidimensionales porque contienen datos cuantitativos, los cuales son instantáneos en el tiempo; y, permiten al usuario filtrarlos, agruparlos y explorarlos a través de condiciones definidas en las tablas de dimensiones.

DATA MART

Un *Data Mart* se define como la implementación de un *Data Warehouse* con un alcance restringido a un área funcional, problema en particular, departamento, tema o grupo de necesidades.

DATA WAREHOUSING

Es el proceso de extraer y filtrar los datos de las operaciones comunes a la organización, procedentes de los distintos sistemas de información y/o sistemas externos, para transformarlos, integrarlos y almacenarlos en un depósito o almacén de datos con el fin de acceder a ellos para dar soporte en el proceso de toma de decisiones de una organización.

2.2. DESARROLLO DE LA GUÍA

El Centro de Informática e Internet (CII) desea diseñar un data mart como sistema de soporte de ayuda a la toma de decisiones estratégicas. El CII es un centro de capacitación en el área de informática, ofrece diferentes módulos en las modalidades de estudio normal (3 meses) y acelerado (2 meses). El ciclo académico se desarrolla cada tres meses, al año se tiene cuatro ciclos y la programación de cada módulo depende del número de alumnos matriculados.

El data mart pretende centralizar la información de todo el proceso de matrículas con el objetivo de definir nuevas estrategias de negocio. Para registrar dichas matrículas utiliza un sistema transaccional que utiliza como gestor de base de datos MySQL. Ver imagen 1 que muestra el modelo de datos de dicho sistema.

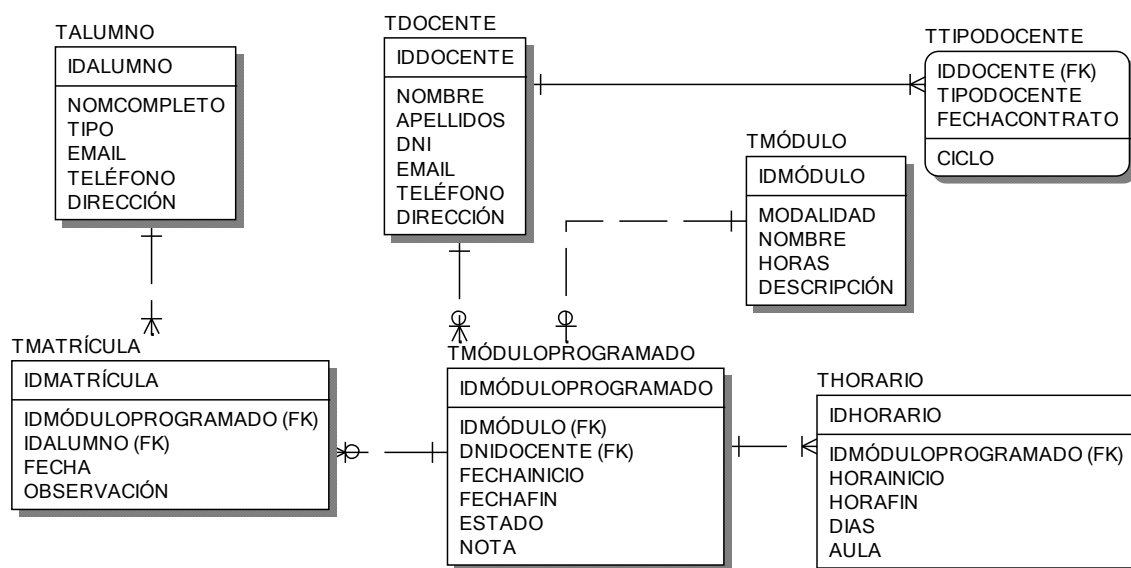


Imagen 1: Base de datos OLTP

Base de datos del Sistema de Matrícula del Centro de Informática e Internet
(BD_CentroInformatica)

2.2.1. PROCESO DATA WAREHOUSING

⇒ Crear la Base de Datos del Data Mart (DM_CentroInformatica). Ver imagen 2.

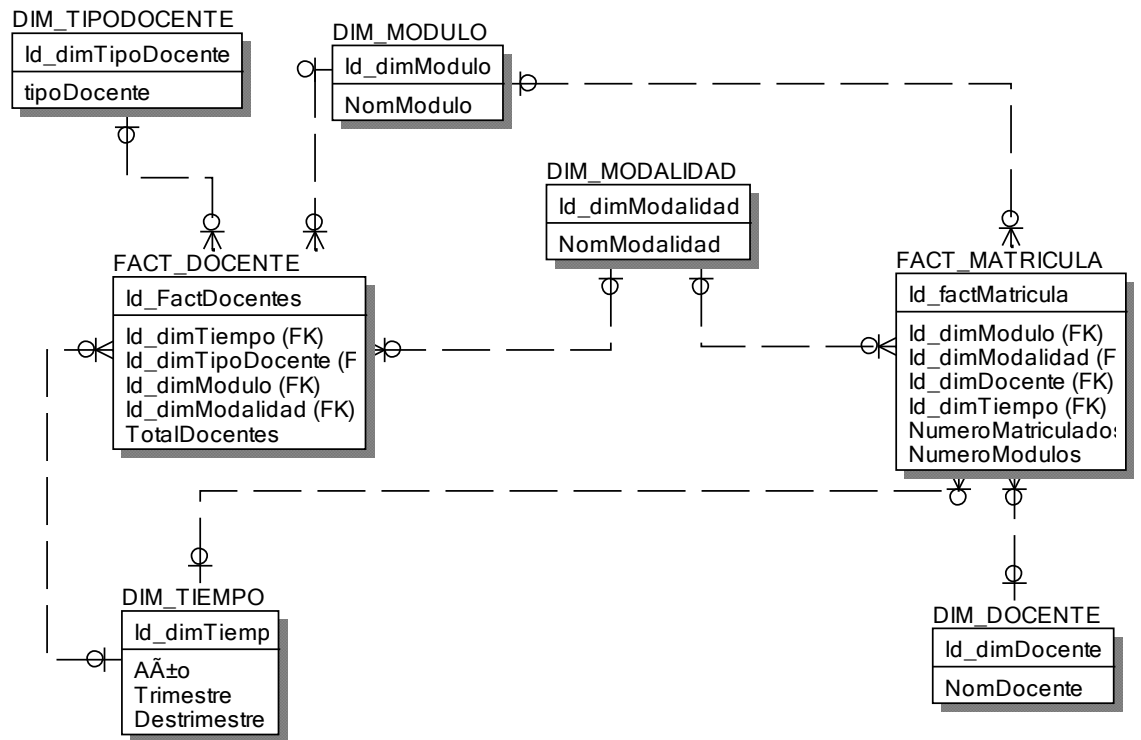


Imagen 2. Modelo de Base de Datos Dimensional del Centro de Informática e Internet (CII)

⇒ Realizar los procesos de Extracción, Transformación y Carga (ETL)

Preparar el ambiente de trabajo

Por cuestión de orden, crear la estructura de carpetas que se muestra en la imagen 3. En la carpeta ETL se guardaran las transformaciones y los trabajos que se creen, en la carpeta Input estará el archivo de origen y en la carpeta Output se creará el archivo de salida.

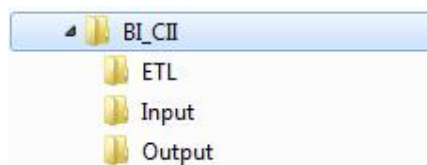


Imagen 3: Estructura de carpetas

Crear la transformación

- Si es que no ha iniciado la herramienta PDI o Spoon, ejecutar el archivo Spoon.bat para empezar con las transformaciones y trabajos.
- Crear una nueva transformación (CTRL+N)
- **Crear Conexiones**
En el panel izquierdo seleccionar la pestaña **View**, crear dos conexiones a base de datos: la primera con la base de datos BD_Centroinformatica que es nuestra fuente de datos de entrada y la segunda con la base de datos dimensional DM_CentroInformatica, que es el Data Mart de dicho centro. Seguir los siguientes pasos:
 - o En la opción **Connection Name**; debe de colocar un nombre a la conexión que referencie a las fuentes de entrada o almacén de datos (Data Mart).
 - o En la Opción Connection Type; debe de seleccionar el gestor de base de datos que se usa la fuente de datos o almacén de datos (Data Mart) e ingresar la información solicitada para la conexión.
 - o Por último se tiene que probar la conexión y ser exitosa. Ver imagen 4 y 5

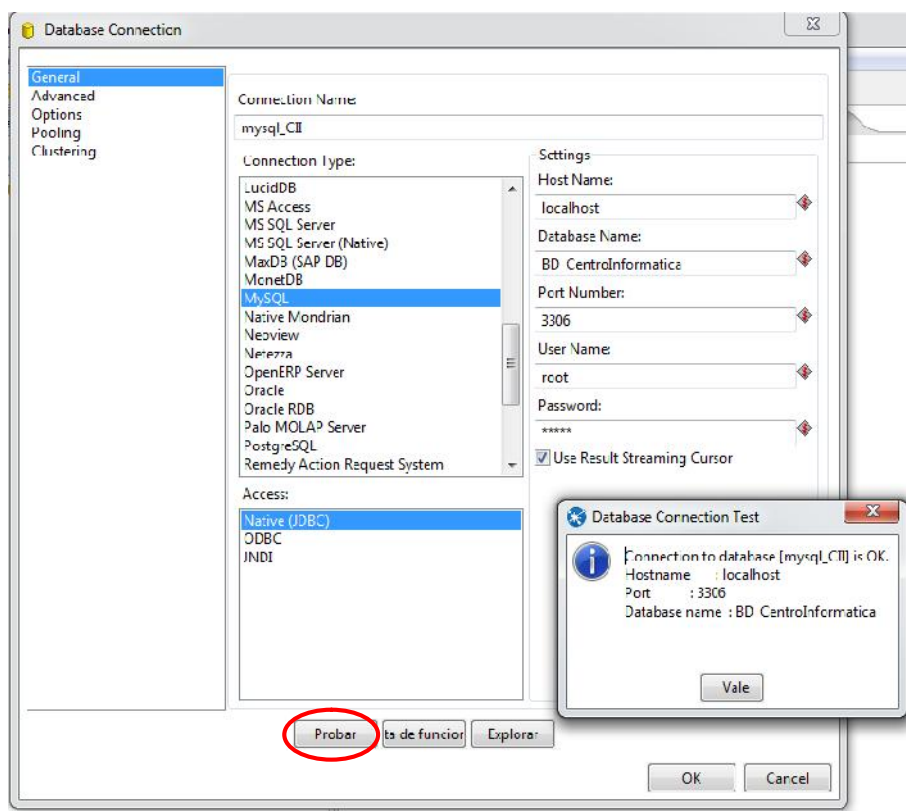


Imagen 4. Conexión con la base de datos BD_CentroInformatica

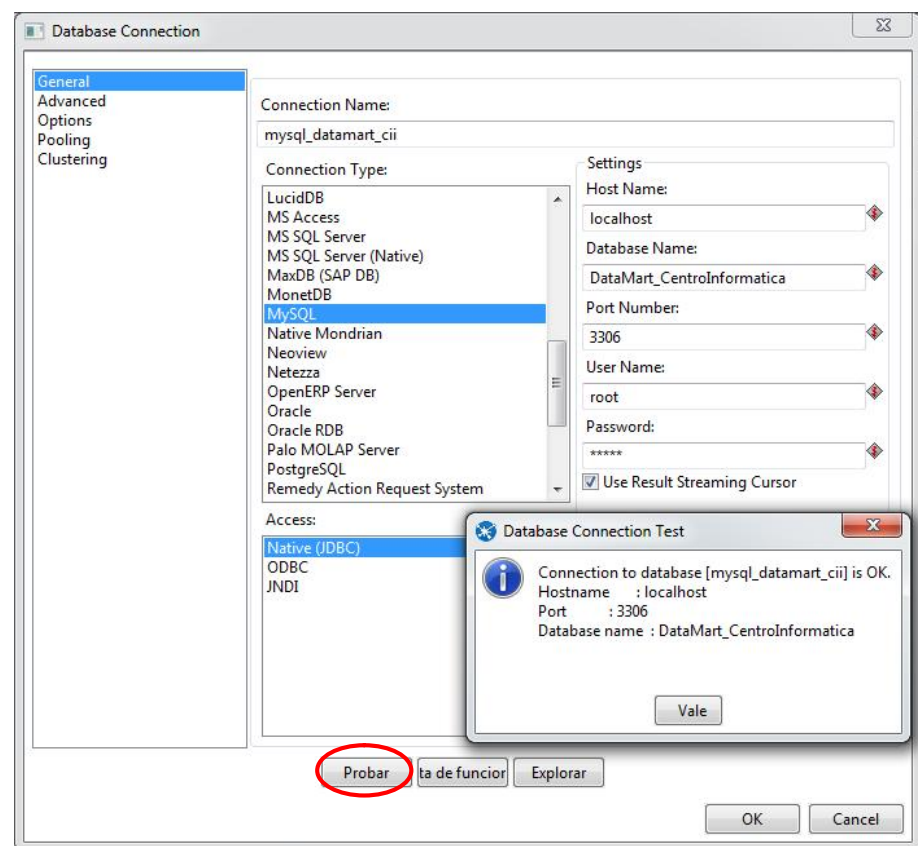


Imagen 5. Conexión con la base de datos DataMart_CentroInformatica

- Diseñar la transformación

En la pestaña **Design**, se tiene que identificar los steps e iniciar el diseño. Se detalla el proceso de transformación de una o dos dimensiones, para el resto de dimensiones el proceso es exactamente igual.

- o Iniciaremos con la transformación de la dimensión Docente.
- o Primer construir una tabla de entrada. Hacer doble clic en el paso, colocar un nombre al paso, determinar la conexión a la base de datos como fuente de entrada y en sql generar la consulta para extraer los datos que se necesita para realizar la carga al almacén de datos. Por último, previsualizar los datos que se están extrayendo de la base de datos BD_CentroInformatica. Ver imagen 6 y 7.

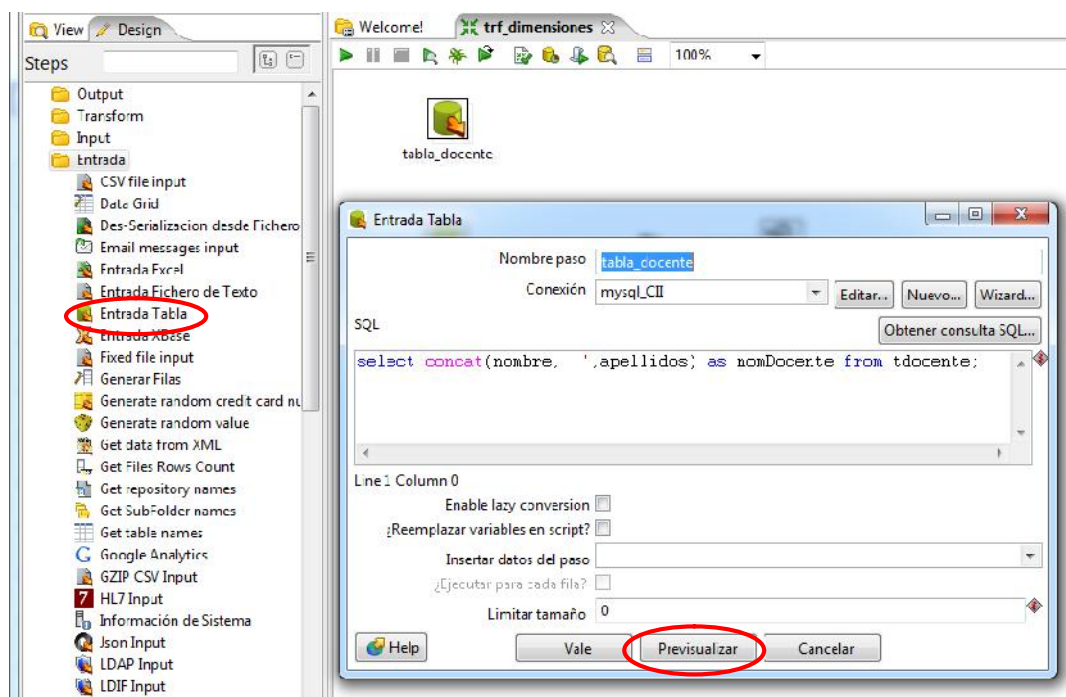


Imagen 6

The screenshot shows the 'Examine preview data' window. It displays a list of 26 rows of data. The first column is 'nomDocente' and the second column is 'nomDocente'.

#	nomDocente	nomDocente
1	MARIO AQUINO CRUZ	MARIO AQUINO CRUZ
2	MICHAEL DUTEN JONT VILTRA	MICHAEL DUTEN JONT VILTRA
3	IVAN AGUAS HUACACA	IVAN AGUAS HUACACA
4	CHRISTIAN ANDRE MEDINA SOTO	CHRISTIAN ANDRE MEDINA SOTO
5	ROSMARIE AGUILAR RODRIGUEZ	ROSMARIE AGUILAR RODRIGUEZ
6	MANUEL JESUS IBARRA CABRERA	MANUEL JESUS IBARRA CABRERA
7	WILLIAM HURTADO CASASVERDE	WILLIAM HURTADO CASASVERDE
8	WISLILLUS PAUL PLATA MALUMA	WISLILLUS PAUL PLATA MALUMA
9	EFRAIN GIL RANDO VEGA	EFRAIN GIL RANDO VEGA
10	DAVID HUGO CALAFUTTA AMANQUE	DAVID HUGO CALAFUTTA AMANQUE
11	LUIS ALBERTO JUQUA VARGAS	LUIS ALBERTO JUQUA VARGAS
12	RONALD ALBERTO RENTERIA AYQUIPA	RONALD ALBERTO RENTERIA AYQUIPA
13	ITSMIRAL DA ROIAS ENRIQUET	ITSMIRAL DA ROIAS ENRIQUET
14	JUVENIL HUSO OLISPE MALDONADO	JUVENIL HUSO OLISPE MALDONADO
15	EBERT GOMEZ ALQUITA	EBERT GOMEZ ALQUITA
16	ILIJIO OSAN HUANKA MAMIN	ILIJIO OSAN HUANKA MAMIN
17	ERECH ORDOÑEZ RAMOS	ERECH ORDOÑEZ RAMOS
18	ITALINO KAIAS HUANCAR HERRERA	ITALINO KAIAS HUANCAR HERRERA
19	ILIDIA HONKICULA ALVAREZ	ILIDIA HONKICULA ALVAREZ
20	JOSE GREGORIO SOTO RODRIGUEZ	JOSE GREGORIO SOTO RODRIGUEZ
21	IFENIFFER LUIS AGUIRRE CARDENAS VIVANCO	IFENIFFER LUIS AGUIRRE CARDENAS VIVANCO
22	JESUS SANTOS HUARNIL FARRAN	JESUS SANTOS HUARNIL FARRAN
23	DARWIN REYES SALCEDO	DARWIN REYES SALCEDO
24	MARIO ANTONIO AGUILAR ESPINOZA	MARIO ANTONIO AGUILAR ESPINOZA
25	LUIS MIGUEL ALFARO CHIRINOS	LUIS MIGUEL ALFARO CHIRINOS
26	RAFAEL RICARDO OLISPE MERMA	RAFAEL RICARDO OLISPE MERMA

Imagen 7. Visualización de los datos que se extraen del sistema transaccional

- o Seleccionar el siguiente paso: Almacén de datos, opción Búsqueda/Actualización en combinación.
- o Establecer el Flujo de datos entre los dos pasos creados.
- o Hacer doble clic en el paso de Búsqueda/Actualización en combinación, colocar un nombre al paso, seleccionar la conexión creada con la fuente de datos de salida DataMart_CentroInformatica.
- o Seleccionar la tabla de destino, obtener los campos seleccionando los que corresponden a la dimensión y el campo en flujo. Ver imagen 8.
- o Hacer una previsualización de la transformación.

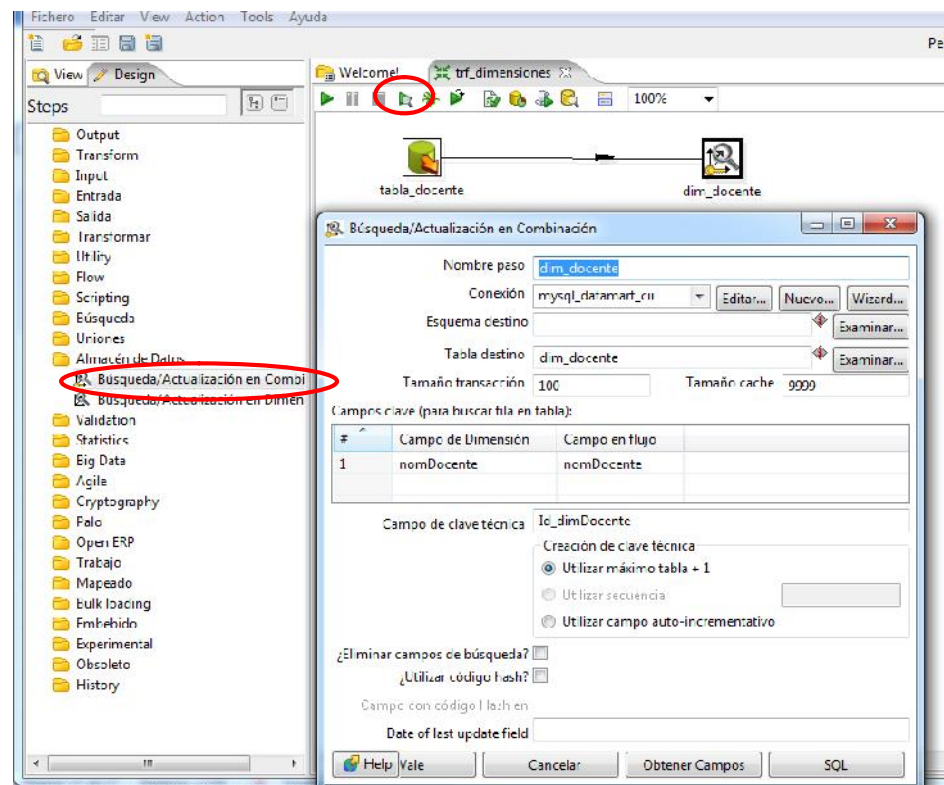


Imagen 8

- o Realizar las transformaciones de carga de las otras dimensiones en la misma hoja de transformación
- o Grabar la transformación como trf_dimensiones
- o Ejecutar la transformación para realizar la carga a la dimensión Docente.

- **Transformación Dimensión Tiempo.**
 - o Se utilizará diferentes pasos. Ver imagen 9

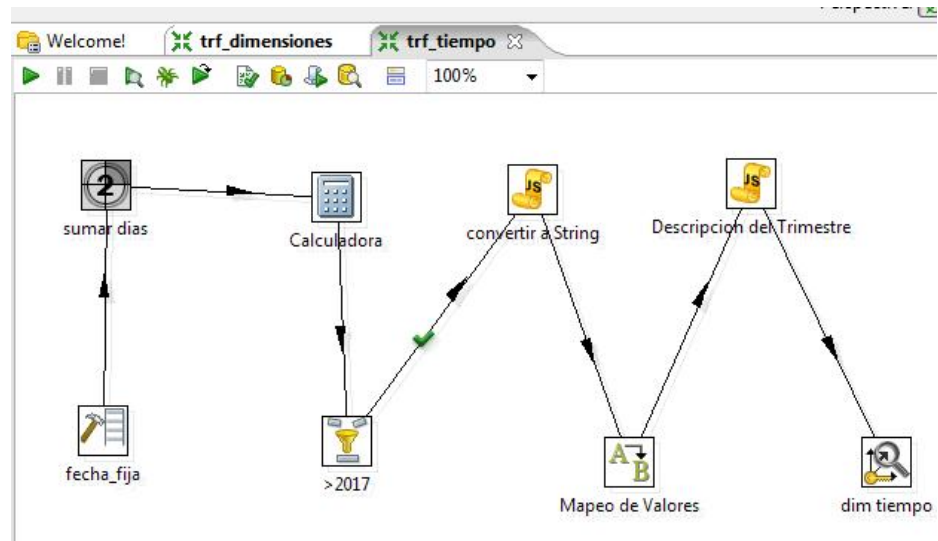


Imagen 9

- o Seleccionar el paso Generar Fila. Ver imagen 10

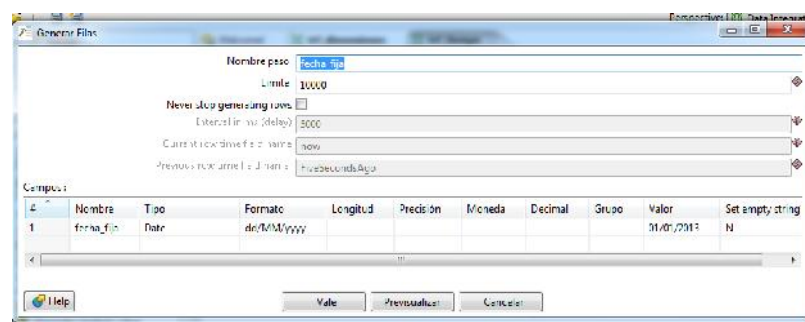


Imagen 10

- o Agregar el paso añadir una secuencia Ver imagen 11

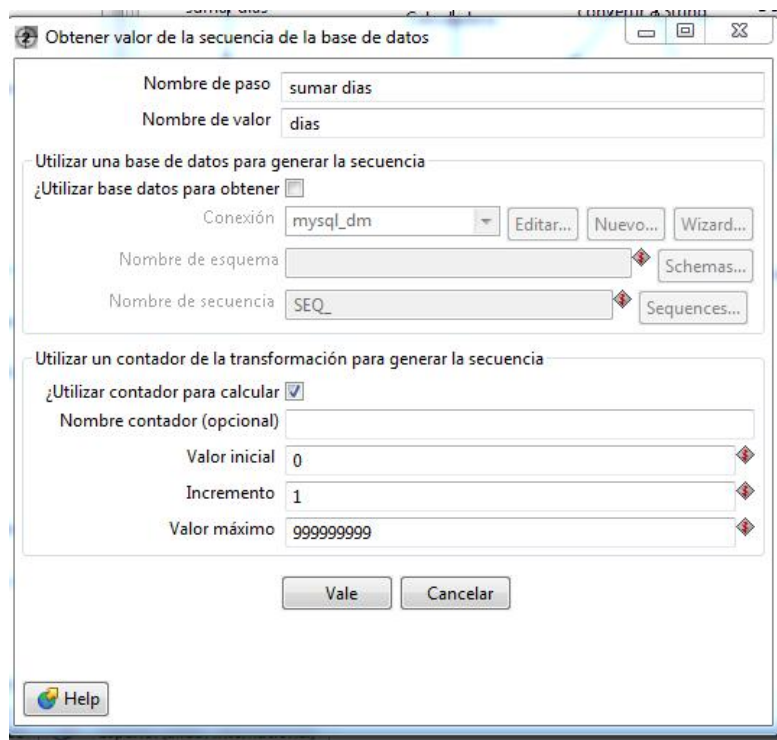


Imagen 11

- o Agregar el paso Calculadora. Ver imagen 12

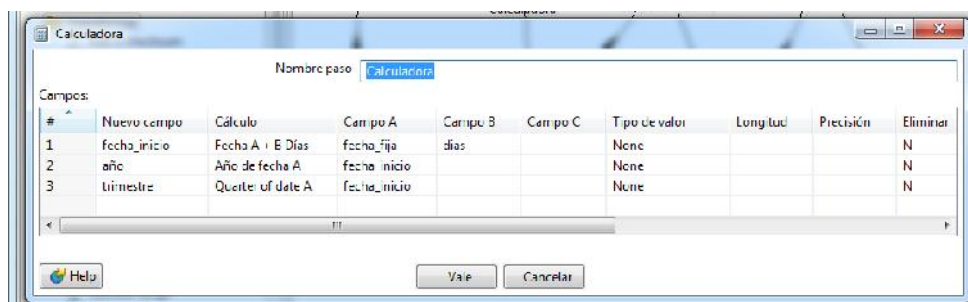


Imagen 12

- o Agregar el paso filtrar filas. Ver imagen 13

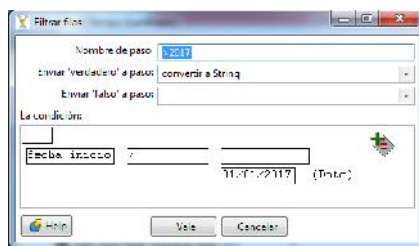


Imagen 13

- o Agregar el paso Valor Java Script Modificado, para modificar el formato numérico del campo trimestre a formato String con código java. Ver imagen 14

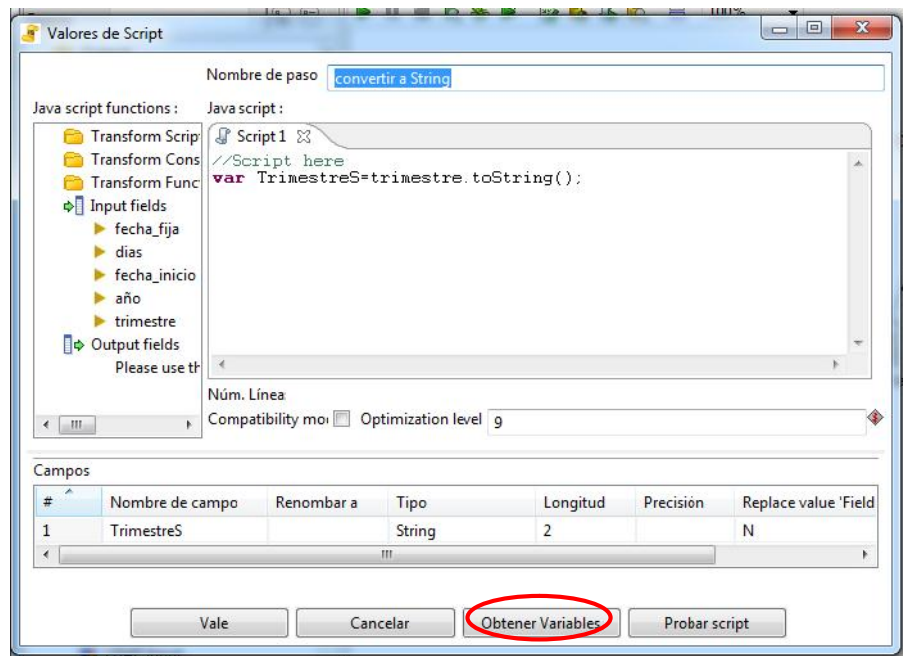


Imagen 14

- o Agregar el paso Mapeo de valores, para transformar el valor de origen. Ver imagen 15.

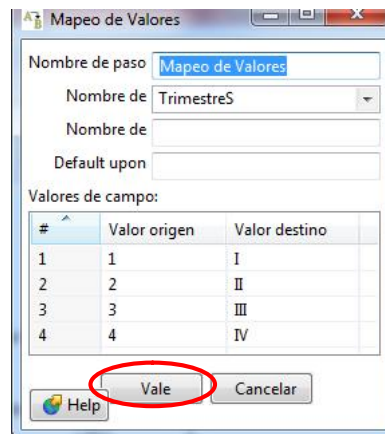


Imagen 15

- o Agregar otra script de java, para darle el formato adecuado de visualización del ciclo academico por ciclo. Ver imagen 16

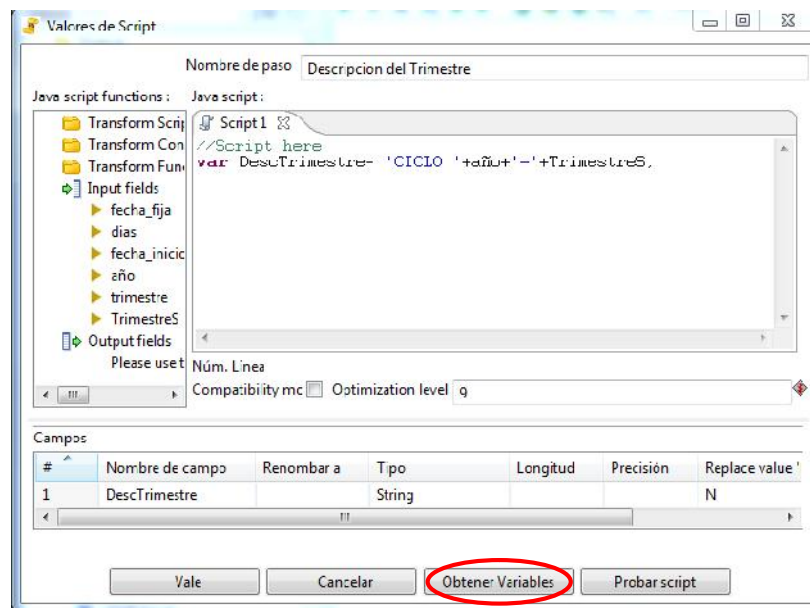


Imagen 16

- o Por último agregar el paso Búsqueda/Actualización en combinación, colocar un nombre al paso, seleccionar la conexión creada con la fuente de datos de salida DataMart_CentroInformatica.
- o Seleccionar la tabla de destino, obtener los campos seleccionando los que corresponden a la dimensión y el campo en flujo. Ver imagen 17.
- o Hacer una previsualización de la transformación. Ver imagen 18.

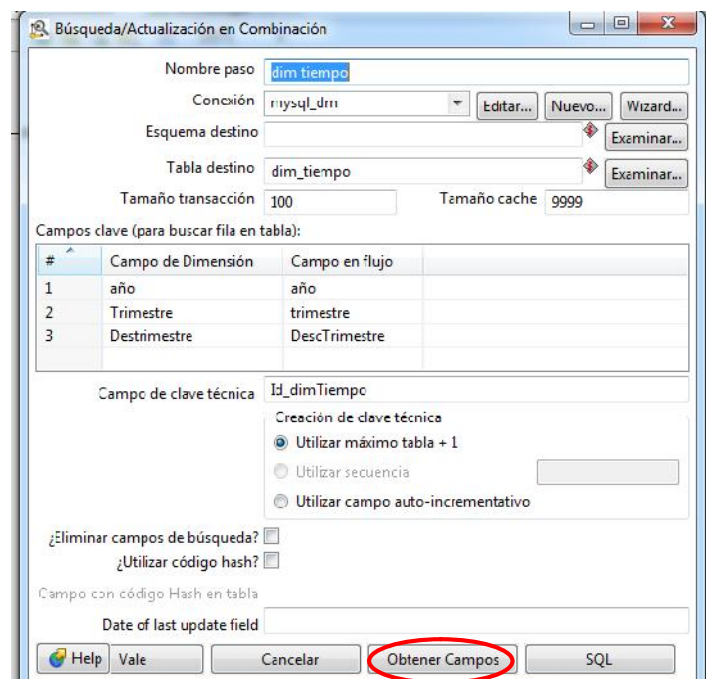


Imagen 17

Examine preview data

Rows: 0 / Size: dim_tiempo (300 rows)

#	fecha_gg	dia	fecha_hh:mm:ss	año	semestre	trimestre	Descr trimestre	id_dim_tiempo
1	01/01/2013	0	2013/01/01 00:00:00.000	2013	1	1	CICLO 2013-1	1
2	01/01/2013	1	2013/01/02 00:00:00.000	2013	1	1	CICLO 2013-1	1
3	01/01/2013	2	2013/01/03 00:00:00.000	2013	1	1	CICLO 2013-1	1
4	01/01/2013	3	2013/01/04 00:00:00.000	2013	1	1	CICLO 2013-1	1
5	01/01/2013	4	2013/01/05 00:00:00.000	2013	1	1	CICLO 2013-1	1
6	01/01/2013	5	2013/01/06 00:00:00.000	2013	1	1	CICLO 2013-1	1
7	01/01/2013	6	2013/01/07 00:00:00.000	2013	1	1	CICLO 2013-1	1
8	01/01/2013	7	2013/01/08 00:00:00.000	2013	1	1	CICLO 2013-1	1
9	01/01/2013	8	2013/01/09 00:00:00.000	2013	1	1	CICLO 2013-1	1
10	01/01/2013	9	2013/01/10 00:00:00.000	2013	1	1	CICLO 2013-1	1
11	01/01/2013	10	2013/01/11 00:00:00.000	2013	1	1	CICLO 2013-1	1
12	01/01/2013	11	2013/01/12 00:00:00.000	2013	1	1	CICLO 2013-1	1
13	01/01/2013	12	2013/01/13 00:00:00.000	2013	1	1	CICLO 2013-1	1
14	01/01/2013	13	2013/01/14 00:00:00.000	2013	1	1	CICLO 2013-1	1
15	01/01/2013	14	2013/01/15 00:00:00.000	2013	1	1	CICLO 2013-1	1
16	01/01/2013	15	2013/01/16 00:00:00.000	2013	1	1	CICLO 2013-1	1
17	01/01/2013	16	2013/01/17 00:00:00.000	2013	1	1	CICLO 2013-1	1
18	01/01/2013	17	2013/01/18 00:00:00.000	2013	1	1	CICLO 2013-1	1
19	01/01/2013	18	2013/01/19 00:00:00.000	2013	1	1	CICLO 2013-1	1
20	01/01/2013	19	2013/01/20 00:00:00.000	2013	1	1	CICLO 2013-1	1
21	01/01/2013	20	2013/01/21 00:00:00.000	2013	1	1	CICLO 2013-1	1
22	01/01/2013	21	2013/01/22 00:00:00.000	2013	1	1	CICLO 2013-1	1
23	01/01/2013	22	2013/01/23 00:00:00.000	2013	1	1	CICLO 2013-1	1
24	01/01/2013	23	2013/01/24 00:00:00.000	2013	1	1	CICLO 2013-1	1
25	01/01/2013	24	2013/01/25 00:00:00.000	2013	1	1	CICLO 2013-1	1
26	01/01/2013	25	2013/01/26 00:00:00.000	2013	1	1	CICLO 2013-1	1
27	01/01/2013	26	2013/01/27 00:00:00.000	2013	1	1	CICLO 2013-1	1
28	01/01/2013	27	2013/01/28 00:00:00.000	2013	1	1	CICLO 2013-1	1
29	01/01/2013	28	2013/01/29 00:00:00.000	2013	1	1	CICLO 2013-1	1
30	01/01/2013	29	2013/01/30 00:00:00.000	2013	1	1	CICLO 2013-1	1

Close Stop Get more rows

Imagen 18

- **Transformación y carga de datos a la Tabla de Hechos Docentes.**

- o Se utilizará los pasos en búsqueda en base de datos para obtener la imagen 19.

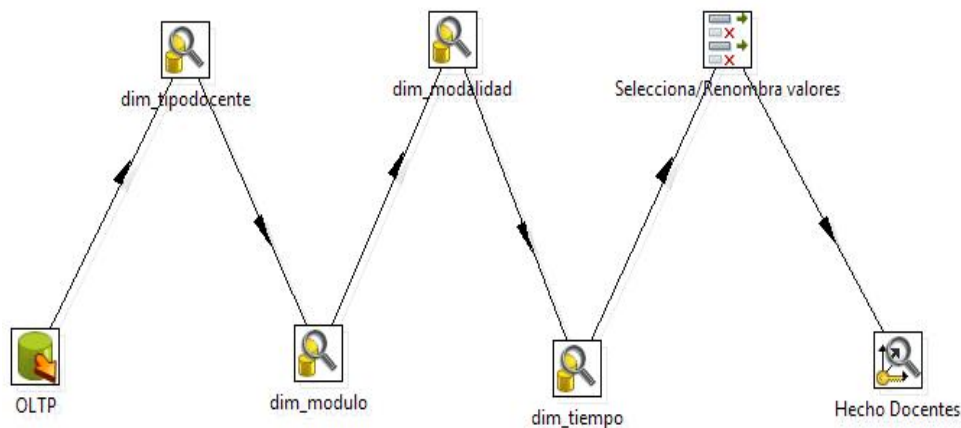


Imagen 19. Transformación y carga a la tabla de hecho docentes

- o Seleccionar el paso de entrada una tabla, hacer doble clic seleccionar la conexión a la base de datos de origen y generar las consultas para obtener los datos que vamos a utilizar para hacer la búsqueda en la base de datos por cada dimensión. Ver imagen 20

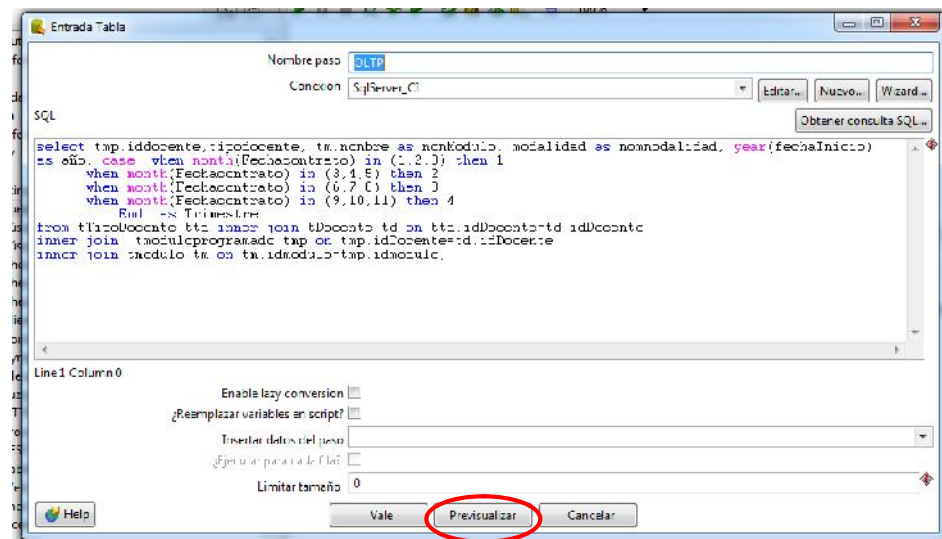


Imagen 20

- o Seleccionar el paso búsqueda de valor en base de datos. Hacer doble clic, especificar la conexión donde se va a realizar la búsqueda de datos, obtenemos los campos y solo seleccionamos los que corresponden a la búsqueda de la dimensión. Ver Imagen 21.

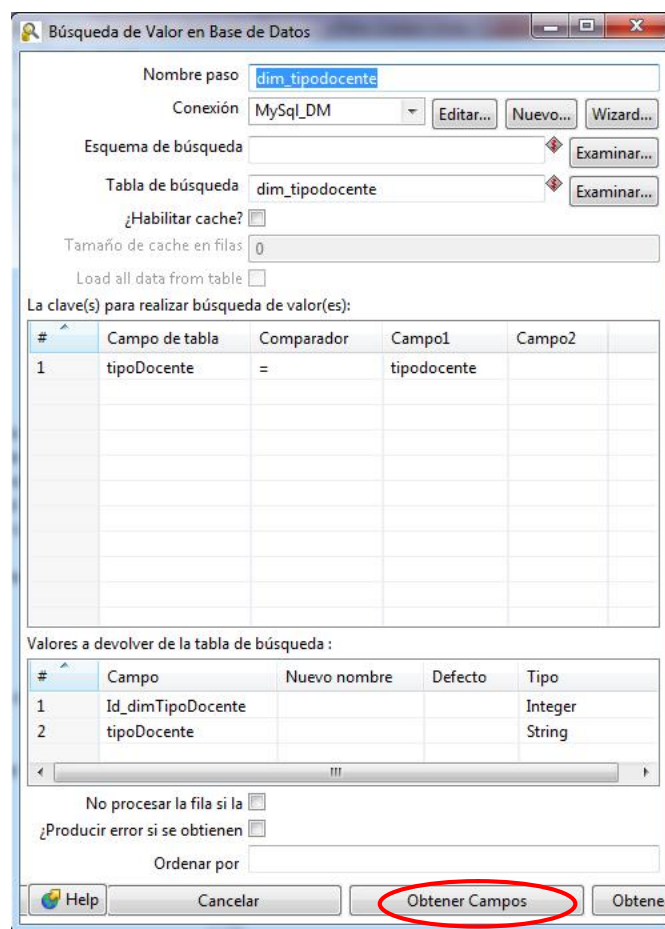


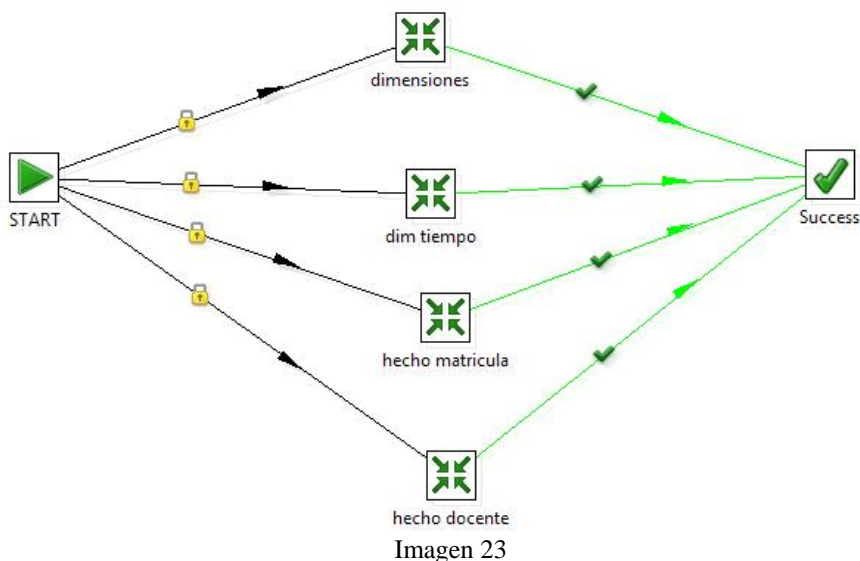
Imagen 22

- o Continuar con las demas dimensiones que se necesitan para realizar la carga en la tabla de hechos.
- o Se recomienda visualizar los datos por cada paso que se realiza.
- o Realizar la transformación y carga a la tabla de hechos matricula.

- Crear un trabajo (Job)

Un trabajo es un componente que crea una secuencia de actividades que brinda un orden de ejecución por lo tanto los jobs son usados para un control de flujo y por lo general consiste de una serie de transformaciones.

- o Crear un nuevo trabajo
- o Seleccionar la carpeta General y el paso star.
- o Luego seleccionar el paso de transformaciones, ingresar la ubicación exacta donde se encuentra la transformación para su ejecución.
- o Seleccionar el paso Success.
- o Finalizar con la ejecución del trabajo que realizara el proceso de transformación, limpieza y carga de datos al Data Mart del CII. Ver imagen 23.



3. EJERCICIOS DE INVESTIGACIÓN

- Diseñar la Base de Datos Multidimensional del Sistema de Grados y Títulos.
- Realizar todo el proceso Data Warehousing.



LABORATORIO N° 02

PROCESO ETL DESDE UN SISTEMA TRANSACCIONAL A UNA BASE DE DATOS DIMENSIONAL (DATA MART)

1. OBJETIVOS:

Luego de completar este laboratorio usted será capaz de:

- 📁 Implementar la base de datos dimensional en mysql
- 📁 Realizar el proceso de extracción de datos de un sistema transaccional (OLTP)
- 📁 Realizar la transformación y limpieza de datos de las dimensiones
- 📁 Realizar la carga de datos al Data Mart utilizando PDI PENTAHO.

2. PLAN DE CLASE:

2.1. MARCO TEÓRICO

PENTAHO es una suite de herramientas de inteligencia de negocios que tiene dos versiones, la versión comercial y la versión de código abierto.

BASE DE DATOS DIMENSIONAL

Una base de datos multidimensional es aquella base de datos en donde su información se almacena en una estructura multidimensional, conformada por tablas de dimensiones y tablas de hechos.

- Tabla de dimensiones
Una dimensión representa información de una entidad de negocio. Las tablas de dimensión, contienen información que permiten definir cómo están los datos organizados lógicamente y describen la forma para analizar el contexto del negocio; además se caracterizan por contener datos cualitativos.
- Tabla de hechos
Un hecho representa un evento de negocio. Contiene información de datos reales, de medidas, de hechos en bruto. Son tablas en las que cada tupla contiene información de un hecho: las medidas asociadas a ese hecho.

Las tablas de Hechos son la base central de los esquemas multidimensionales porque contienen datos cuantitativos, los cuales son instantáneos en el tiempo; y, permiten al usuario filtrarlos, agruparlos y explorarlos a través de condiciones definidas en las tablas de dimensiones.

DATA MART

Un *Data Mart* se define como la implementación de un *Data Warehouse* con un alcance restringido a un área funcional, problema en particular, departamento, tema o grupo de necesidades.

DATA WAREHOUSING

Es el proceso de extraer y filtrar los datos de las operaciones comunes a la organización, procedentes de los distintos sistemas de información y/o sistemas externos, para transformarlos, integrarlos y almacenarlos en un depósito o almacén de datos con el fin de acceder a ellos para dar soporte en el proceso de toma de decisiones de una organización.

2.2. DESARROLLO DE LA GUÍA

El Centro de Informática e Internet (CII) desea diseñar un data mart como sistema de soporte de ayuda a la toma de decisiones estratégicas. El CII es un centro de capacitación en el área de informática, ofrece diferentes módulos en las modalidades de estudio normal (3 meses) y acelerado (2 meses). El ciclo académico se desarrolla cada tres meses, al año se tiene cuatro ciclos y la programación de cada módulo depende del número de alumnos matriculados.

El data mart pretende centralizar la información de todo el proceso de matrículas con el objetivo de definir nuevas estrategias de negocio. Para registrar dichas matrículas utiliza un sistema transaccional que utiliza como gestor de base de datos MySQL. Ver imagen 1 que muestra el modelo de datos de dicho sistema.

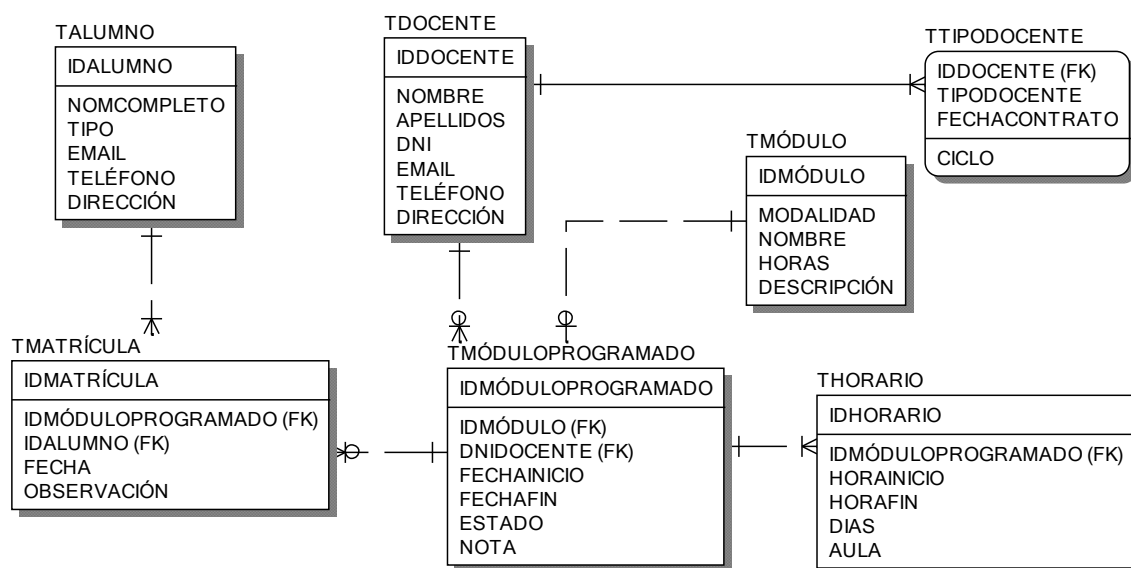


Imagen 1: Base de datos OLTP

Base de datos del Sistema de Matrícula del Centro de Informática e Internet (BD_CentroInformatica)

2.2.1. PROCESO DATA WAREHOUSING

⇒ Crear la Base de Datos del Data Mart (DM_CentroInformatica). Ver imagen 2.

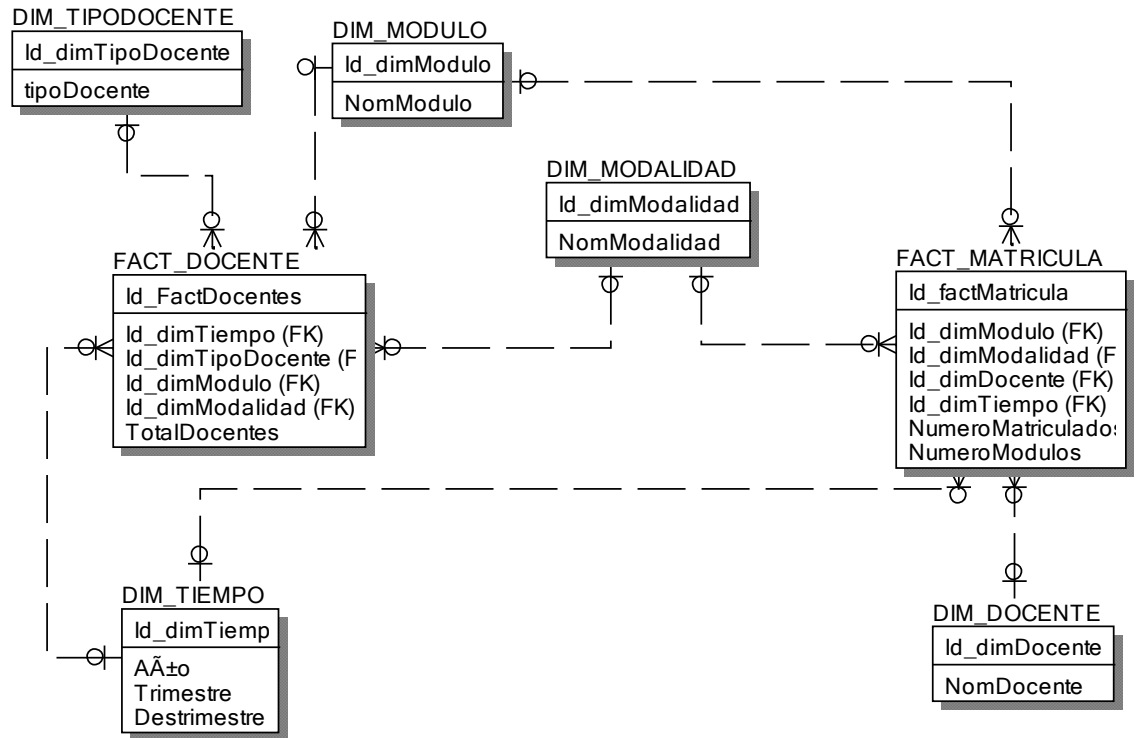


Imagen 2. Modelo de Base de Datos Dimensional del Centro de Informática e Internet (CII)

⇒ Realizar los procesos de Extracción, Transformación y Carga (ETL)

Preparar el ambiente de trabajo

Por cuestión de orden, crear la estructura de carpetas que se muestra en la imagen 3. En la carpeta ETL se guardaran las transformaciones y los trabajos que se creen, en la carpeta Input estará el archivo de origen y en la carpeta Output se creará el archivo de salida.

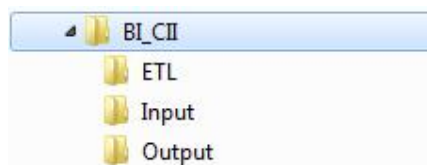


Imagen 3: Estructura de carpetas

Crear la transformación

- Si es que no ha iniciado la herramienta PDI o Spoon, ejecutar el archivo Spoon.bat para empezar con las transformaciones y trabajos.
- Crear una nueva transformación (CTRL+N)
- **Crear Conexiones**
En el panel izquierdo seleccionar la pestaña **View**, crear dos conexiones a base de datos: la primera con la base de datos BD_Centroinformatica que es nuestra fuente de datos de entrada y la segunda con la base de datos dimensional DM_CentroInformatica, que es el Data Mart de dicho centro. Seguir los siguientes pasos:
 - o En la opción **Connection Name**; debe de colocar un nombre a la conexión que referencie a las fuentes de entrada o almacén de datos (Data Mart).
 - o En la Opción Connection Type; debe de seleccionar el gestor de base de datos que se usa la fuente de datos o almacén de datos (Data Mart) e ingresar la información solicitada para la conexión.
 - o Por último se tiene que probar la conexión y ser exitosa. Ver imagen 4 y 5

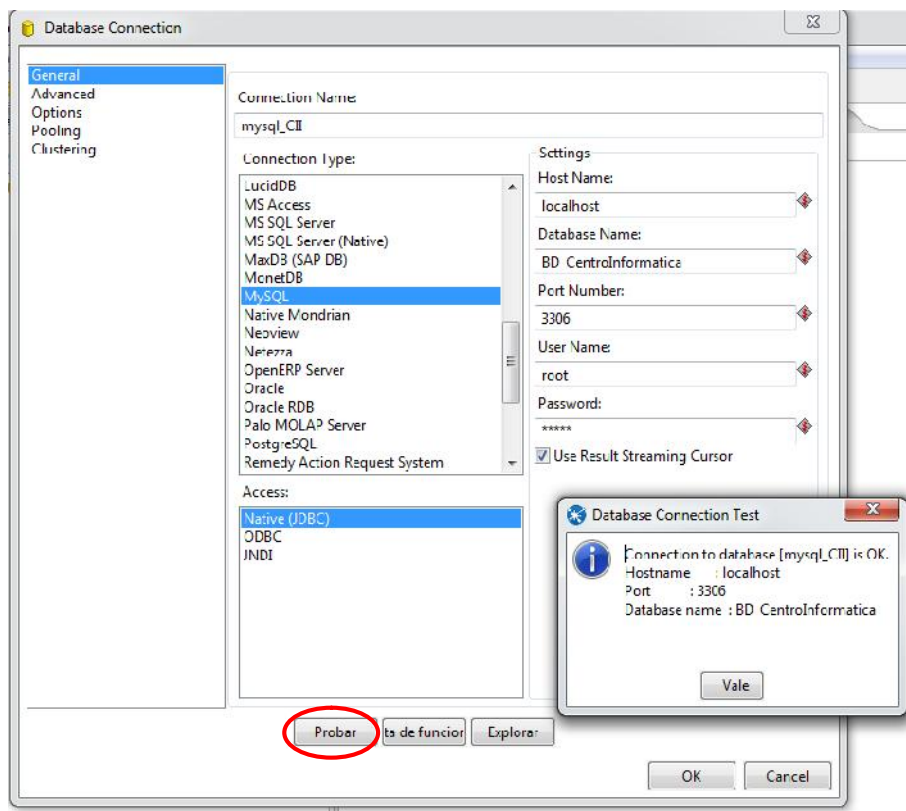


Imagen 4. Conexión con la base de datos BD_CentroInformatica

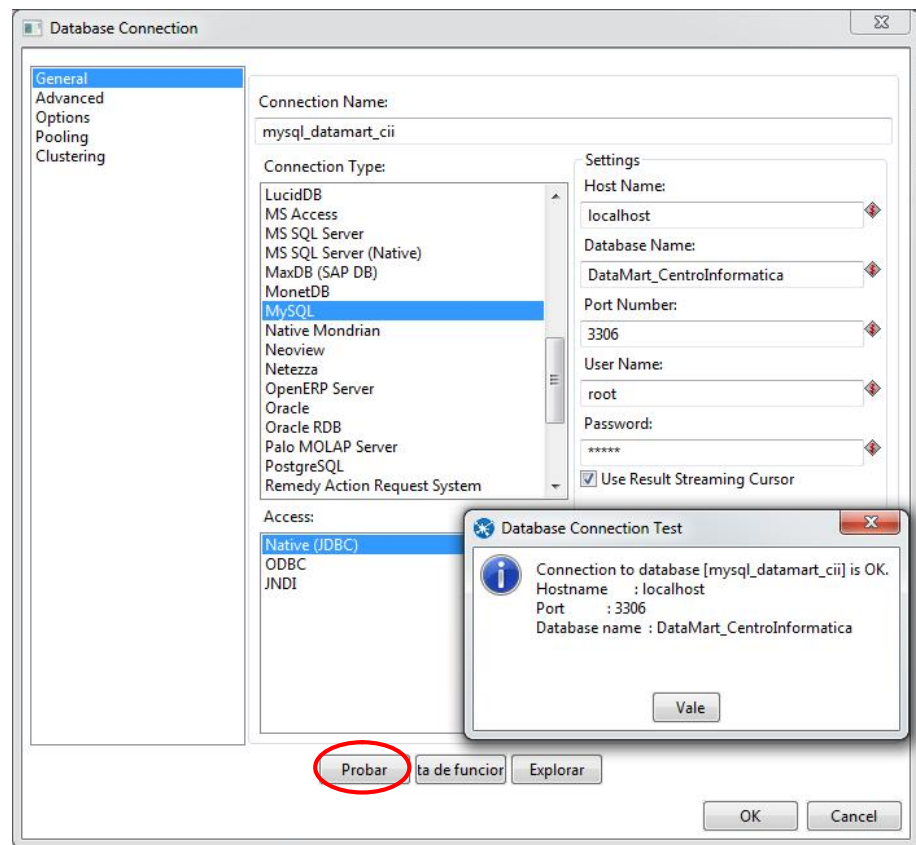


Imagen 5. Conexión con la base de datos DataMart_CentroInformatica

- Diseñar la transformación

En la pestaña **Design**, se tiene que identificar los steps e iniciar el diseño. Se detalla el proceso de transformación de una o dos dimensiones, para el resto de dimensiones el proceso es exactamente igual.

- o Iniciaremos con la transformación de la dimensión Docente.
- o Primer construir una tabla de entrada. Hacer doble clic en el paso, colocar un nombre al paso, determinar la conexión a la base de datos como fuente de entrada y en sql generar la consulta para extraer los datos que se necesita para realizar la carga al almacén de datos. Por último, previsualizar los datos que se están extrayendo de la base de datos BD_CentroInformatica. Ver imagen 6 y 7.

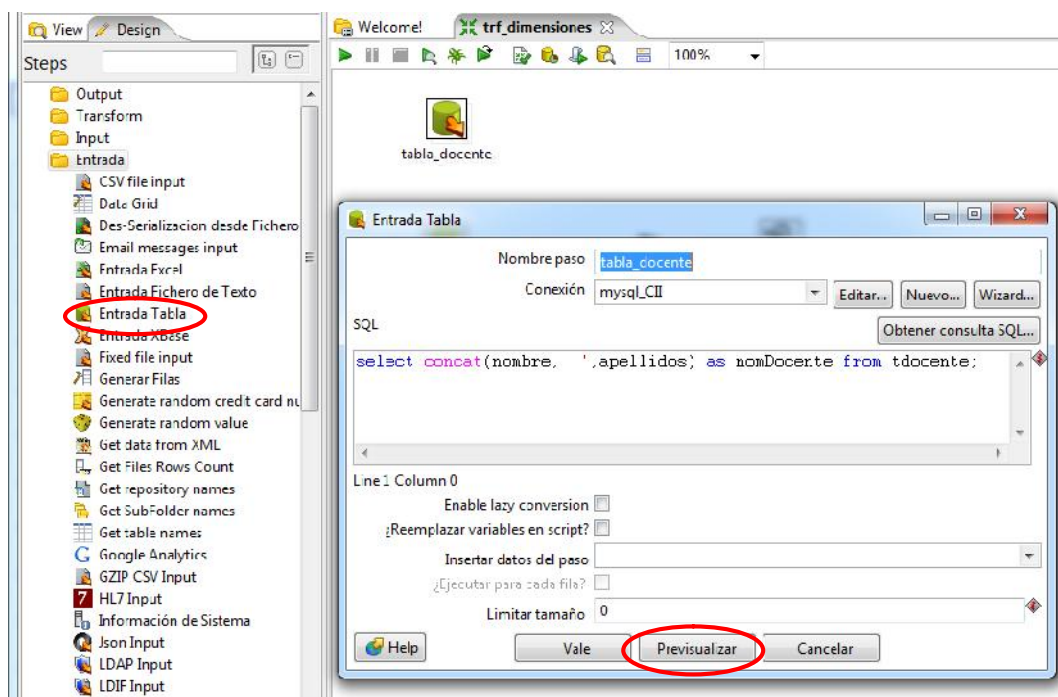


Imagen 6

The screenshot shows the 'Examine preview data' window. The table has 26 rows and 2 columns. The first column is 'nomDocente' and the second column is 'nomDocente'.

#	nomDocente	nomDocente
1	MARIO AQUINO CRUZ	MARIO AQUINO CRUZ
2	MICHAEL DUTEN JONT VILTRA	MICHAEL DUTEN JONT VILTRA
3	IVAN AGUAS HUACACA	IVAN AGUAS HUACACA
4	CHRISTIAN ANDRE MEDINA SOTO	CHRISTIAN ANDRE MEDINA SOTO
5	ROSMARIE AGUILAR RODRIGUEZ	ROSMARIE AGUILAR RODRIGUEZ
6	MANUEL JESUS IBARRA CABRERA	MANUEL JESUS IBARRA CABRERA
7	WILLIAM HURTADO CASASVERDE	WILLIAM HURTADO CASASVERDE
8	WISLILLUS PAUL PLATA MALUMA	WISLILLUS PAUL PLATA MALUMA
9	EFRAIN GIL RANCHO VEGA	EFRAIN GIL RANCHO VEGA
10	DAVID HUGO CALAFUTTA AMANQUE	DAVID HUGO CALAFUTTA AMANQUE
11	LUIS ALBERTO JUQUA VARGAS	LUIS ALBERTO JUQUA VARGAS
12	RONALD ALBERTO RENTERIA AYQUIPA	RONALD ALBERTO RENTERIA AYQUIPA
13	ITSMIRAL DA ROIAS MARIQUET	ITSMIRAL DA ROIAS MARIQUET
14	JUVENIL-HUSO OLISPE MALDONADO	JUVENIL-HUSO OLISPE MALDONADO
15	EBERT GOMEZ ALQUITA	EBERT GOMEZ ALQUITA
16	ILIJIO OSAN HUANKA MAMIN	ILIJIO OSAN HUANKA MAMIN
17	ERECH ORDOÑEZ RAMOS	ERECH ORDOÑEZ RAMOS
18	ITALINO KAIAS HUANCAN-HERRERA	ITALINO KAIAS HUANCAN-HERRERA
19	ILIDIA HONORILDA ALVAREZ	ILIDIA HONORILDA ALVAREZ
20	JOSE GREGORIO SOTO RODRIGUEZ	JOSE GREGORIO SOTO RODRIGUEZ
21	IFENIFFER LUIS AGUIRRE CARDENAS VIVANCO	IFENIFFER LUIS AGUIRRE CARDENAS VIVANCO
22	JESUS SANTOS HUARNIZ HERRERA	JESUS SANTOS HUARNIZ HERRERA
23	DARWIN REYES SALCEDO	DARWIN REYES SALCEDO
24	MARIO ANTONIO AGUIRRE ESPINOZA	MARIO ANTONIO AGUIRRE ESPINOZA
25	LUIS MIGUEL ALFARO CHIRINOS	LUIS MIGUEL ALFARO CHIRINOS
26	RAFAEL RICARDO OLISPE MERMA	RAFAEL RICARDO OLISPE MERMA

Imagen 7. Visualización de los datos que se extraen del sistema transaccional

- o Seleccionar el siguiente paso: Almacén de datos, opción Búsqueda/Actualización en combinación.
- o Establecer el Flujo de datos entre los dos pasos creados.
- o Hacer doble clic en el paso de Búsqueda/Actualización en combinación, colocar un nombre al paso, seleccionar la conexión creada con la fuente de datos de salida DataMart_CentroInformatica.
- o Seleccionar la tabla de destino, obtener los campos seleccionando los que corresponden a la dimensión y el campo en flujo. Ver imagen 8.
- o Hacer una previsualización de la transformación.

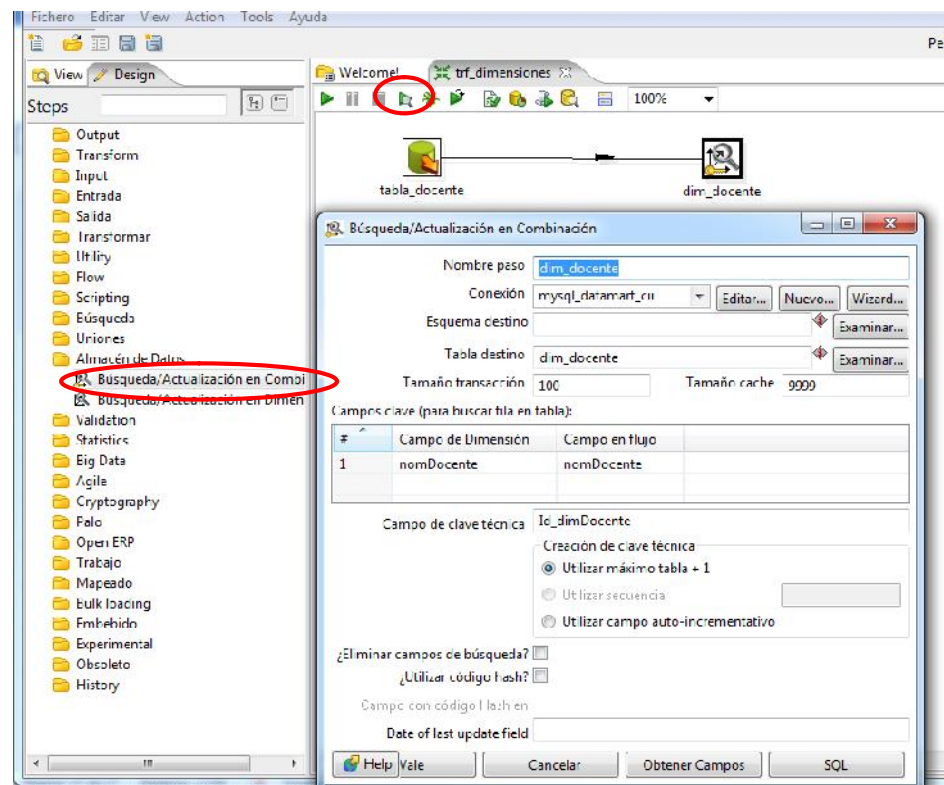


Imagen 8

- o Realizar las transformaciones de carga de las otras dimensiones en la misma hoja de transformación
- o Grabar la transformación como trf_dimensiones
- o Ejecutar la transformación para realizar la carga a la dimensión Docente.

- **Transformación Dimensión Tiempo.**
 - o Se utilizará diferentes pasos. Ver imagen 9

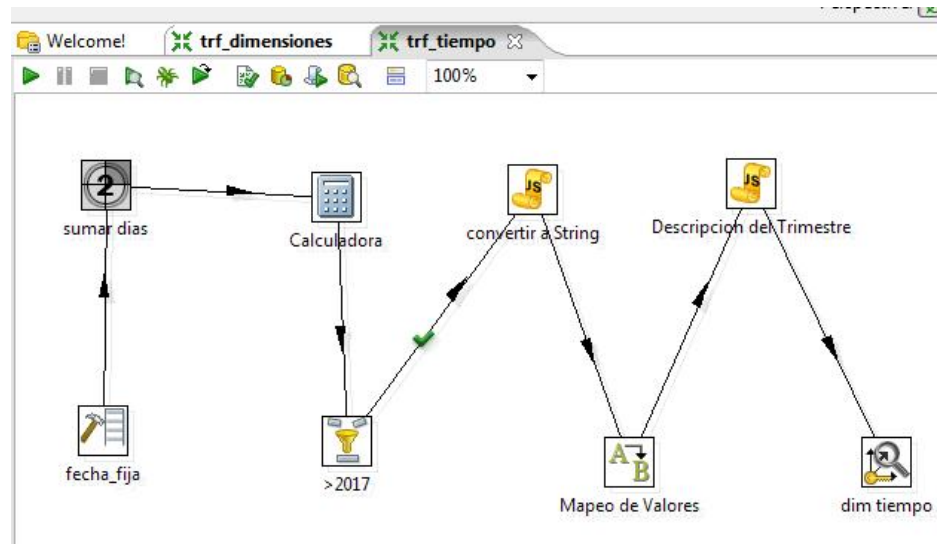


Imagen 9

- o Seleccionar el paso Generar Fila. Ver imagen 10

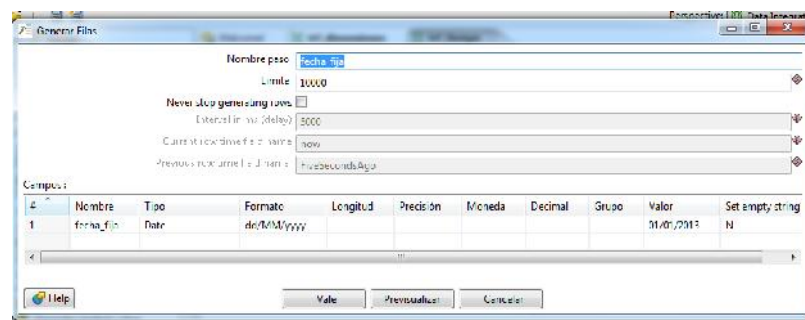


Imagen 10

- o Agregar el paso añadir una secuencia Ver imagen 11

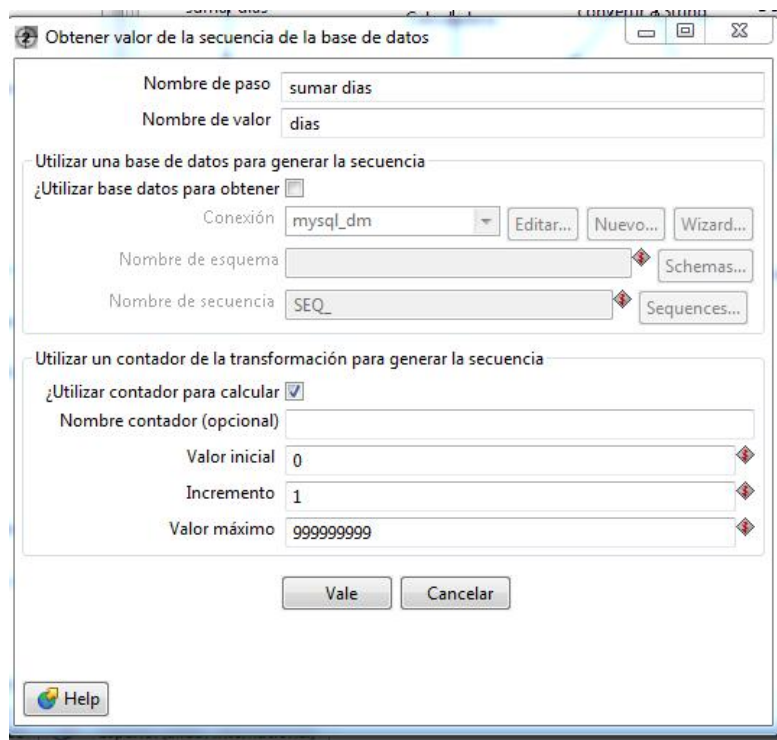


Imagen 11

- o Agregar el paso Calculadora. Ver imagen 12

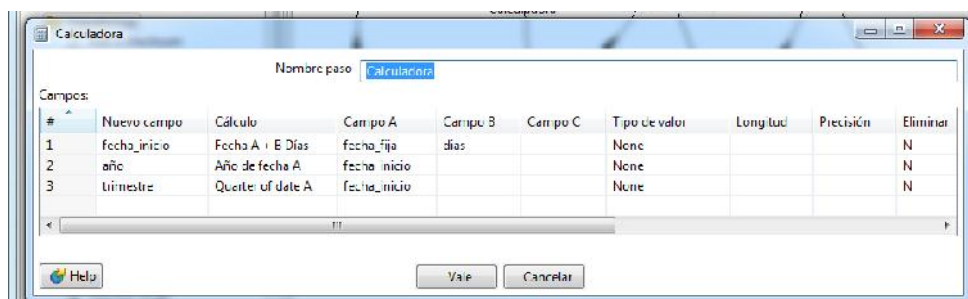


Imagen 12

- o Agregar el paso filtrar filas. Ver imagen 13

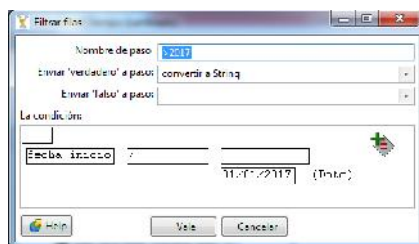


Imagen 13

- o Agregar el paso Valor Java Script Modificado, para modificar el formato numérico del campo trimestre a formato String con código java. Ver imagen 14

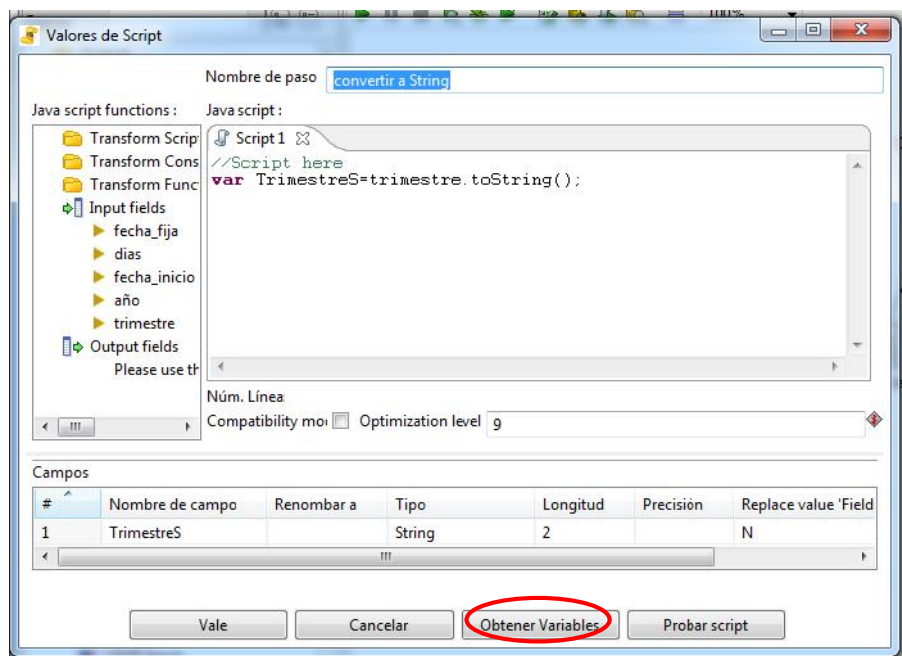


Imagen 14

- o Agregar el paso Mapeo de valores, para transformar el valor de origen. Ver imagen 15.

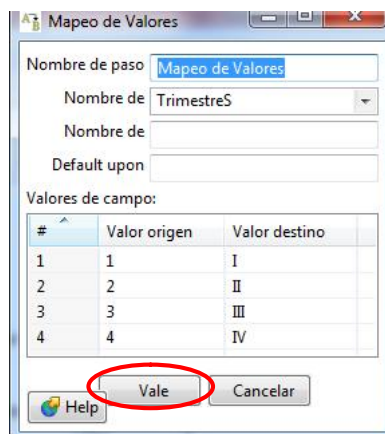


Imagen 15

- o Agregar otra script de java, para darle el formato adecuado de visualización del ciclo academico por ciclo. Ver imagen 16

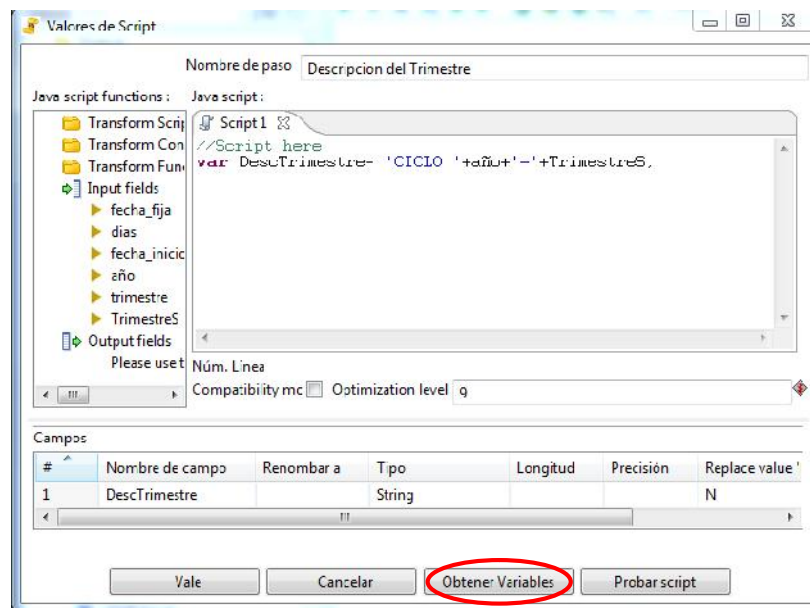


Imagen 16

- o Por último agregar el paso Búsqueda/Actualización en combinación, colocar un nombre al paso, seleccionar la conexión creada con la fuente de datos de salida DataMart_CentroInformatica.
- o Seleccionar la tabla de destino, obtener los campos seleccionando los que corresponden a la dimensión y el campo en flujo. Ver imagen 17.
- o Hacer una previsualización de la transformación. Ver imagen 18.

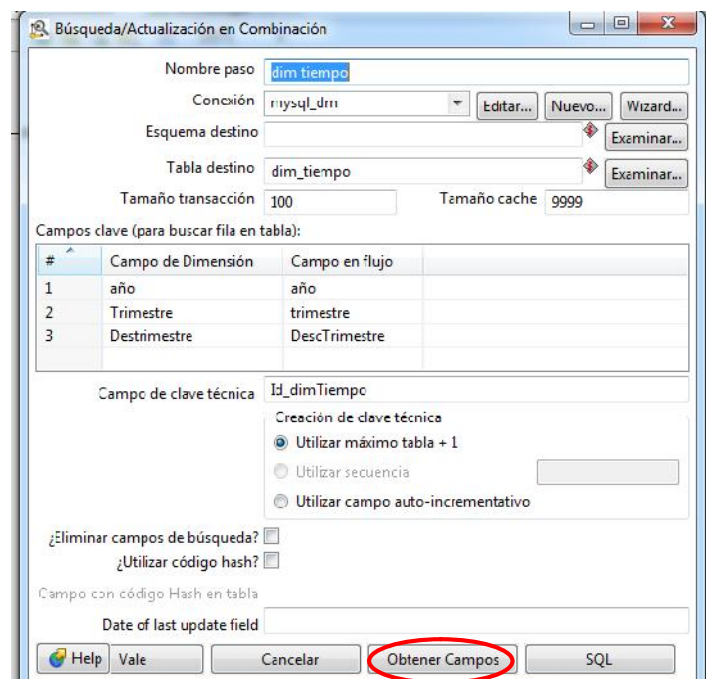


Imagen 17

Examine preview data

Rows: 0 / Selected: 0 (200 rows)

#	fecha_gg	dia	fecha_hora	año	semestre	Periodo	Desc. Periodo	Id_dim Tiempo
1	01/01/2013	1	2013/01/01 00:00:00.000	2013	1	1	CICLO 2013-1	1
2	01/01/2013	1	2013/01/02 00:00:00.000	2013	1	1	CICLO 2013-1	1
3	01/01/2013	2	2013/01/03 00:00:00.000	2013	1	1	CICLO 2013-1	1
4	01/01/2013	3	2013/01/04 00:00:00.000	2013	1	1	CICLO 2013-1	1
5	01/01/2013	4	2013/01/05 00:00:00.000	2013	1	1	CICLO 2013-1	1
6	01/01/2013	5	2013/01/06 00:00:00.000	2013	1	1	CICLO 2013-1	1
7	01/01/2013	6	2013/01/07 00:00:00.000	2013	1	1	CICLO 2013-1	1
8	01/01/2013	7	2013/01/08 00:00:00.000	2013	1	1	CICLO 2013-1	1
9	01/01/2013	8	2013/01/09 00:00:00.000	2013	1	1	CICLO 2013-1	1
10	01/01/2013	9	2013/01/10 00:00:00.000	2013	1	1	CICLO 2013-1	1
11	01/01/2013	10	2013/01/11 00:00:00.000	2013	1	1	CICLO 2013-1	1
12	01/01/2013	11	2013/01/12 00:00:00.000	2013	1	1	CICLO 2013-1	1
13	01/01/2013	12	2013/01/13 00:00:00.000	2013	1	1	CICLO 2013-1	1
14	01/01/2013	13	2013/01/14 00:00:00.000	2013	1	1	CICLO 2013-1	1
15	01/01/2013	14	2013/01/15 00:00:00.000	2013	1	1	CICLO 2013-1	1
16	01/01/2013	15	2013/01/16 00:00:00.000	2013	1	1	CICLO 2013-1	1
17	01/01/2013	16	2013/01/17 00:00:00.000	2013	1	1	CICLO 2013-1	1
18	01/01/2013	17	2013/01/18 00:00:00.000	2013	1	1	CICLO 2013-1	1
19	01/01/2013	18	2013/01/19 00:00:00.000	2013	1	1	CICLO 2013-1	1
20	01/01/2013	19	2013/01/20 00:00:00.000	2013	1	1	CICLO 2013-1	1
21	01/01/2013	20	2013/01/21 00:00:00.000	2013	1	1	CICLO 2013-1	1
22	01/01/2013	21	2013/01/22 00:00:00.000	2013	1	1	CICLO 2013-1	1
23	01/01/2013	22	2013/01/23 00:00:00.000	2013	1	1	CICLO 2013-1	1
24	01/01/2013	23	2013/01/24 00:00:00.000	2013	1	1	CICLO 2013-1	1
25	01/01/2013	24	2013/01/25 00:00:00.000	2013	1	1	CICLO 2013-1	1
26	01/01/2013	25	2013/01/26 00:00:00.000	2013	1	1	CICLO 2013-1	1
27	01/01/2013	26	2013/01/27 00:00:00.000	2013	1	1	CICLO 2013-1	1
28	01/01/2013	27	2013/01/28 00:00:00.000	2013	1	1	CICLO 2013-1	1
29	01/01/2013	28	2013/01/29 00:00:00.000	2013	1	1	CICLO 2013-1	1
30	01/01/2013	29	2013/01/30 00:00:00.000	2013	1	1	CICLO 2013-1	1

Close Stop Get more rows

Imagen 18

- **Transformación y carga de datos a la Tabla de Hechos Docentes.**

- o Se utilizará los pasos en búsqueda en base de datos para obtener la imagen 19.

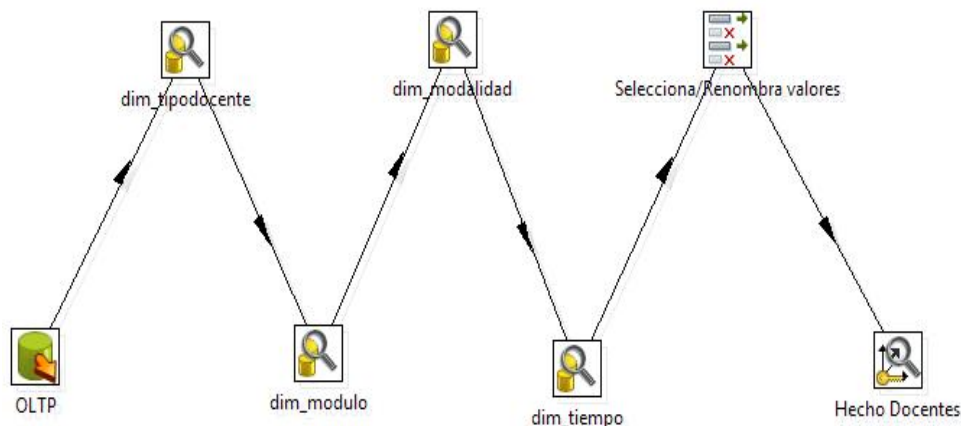


Imagen 19. Transformación y carga a la tabla de hecho docentes

- o Seleccionar el paso de entrada una tabla, hacer doble clic seleccionar la conexión a la base de datos de origen y generar las consultas para obtener los datos que vamos a utilizar para hacer la búsqueda en la base de datos por cada dimensión. Ver imagen 20

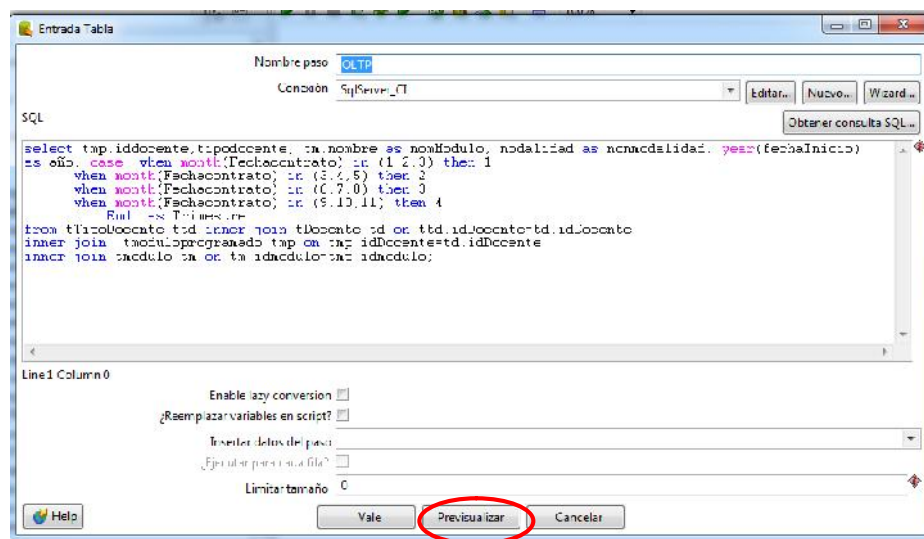


Imagen 20

- o Seleccionar el paso búsqueda de valor en base de datos. Hacer doble clic, especificar la conexión donde se va a realizar la búsqueda de datos, obtenemos los campos y solo seleccionamos los que corresponden a la búsqueda de la dimensión. Ver Imagen 21.

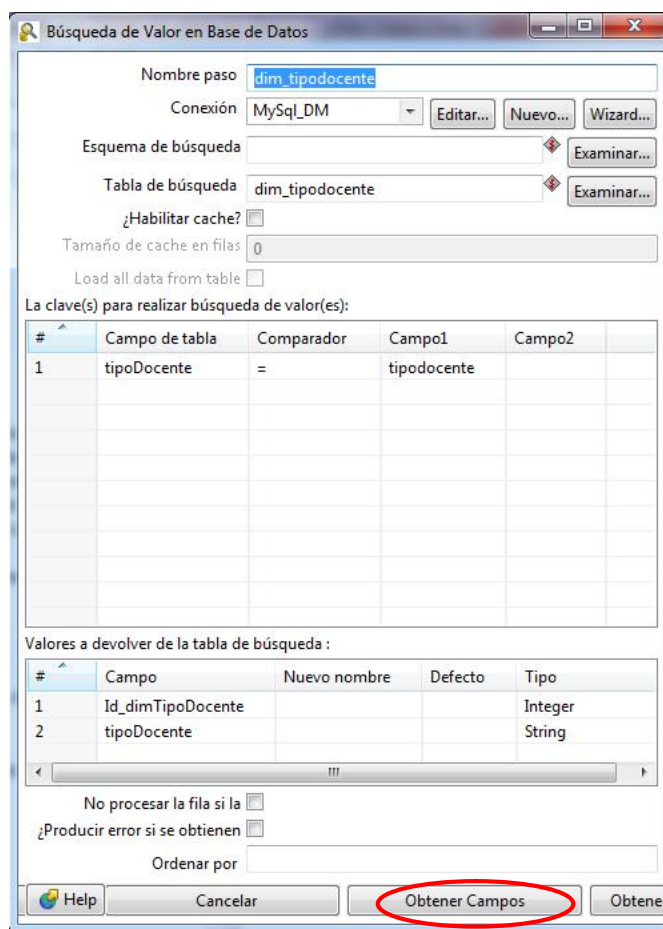


Imagen 21

- o Continuar con las demas dimensiones que se necesitan para realizar la carga en la tabla de hechos.
 - o Se recomienda visualizar los datos por cada paso que se realiza.
 - o Realizar la transformación y carga a la tabla de hechos matricula.
- **Crear un trabajo (Job)**

Un trabajo es un componente que crea una secuencia de actividades que brinda un orden de ejecución por lo tanto los jobs son usados para un control de flujo y por lo general consiste de una serie de transformaciones.

- o Crear un nuevo trabajo
- o Seleccionar la carpeta General y el paso start.
- o Luego seleccionar el paso de transformaciones, ingresar la ubicación exacta donde se encuentra la transformación para su ejecución.
- o Seleccionar el paso Success.
- o Finalizar con la ejecución del trabajo que realizara el proceso de transformación, limpieza y carga de datos al Data Mart del CII. Ver imagen 22.

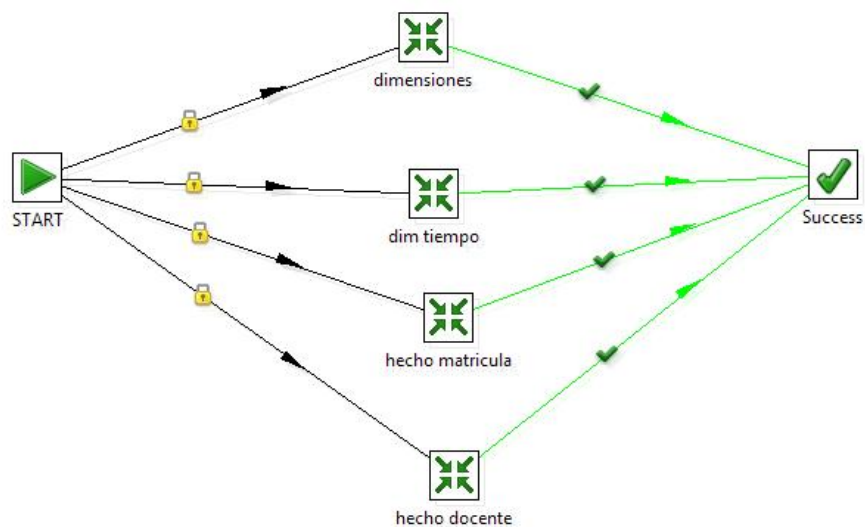


Imagen 22. Trabajo que ejecuta la carga al Data Mart

3. EJERCICIOS DE INVESTIGACIÓN

- Diseñar la Base de Datos Multidimensional del Sistema de Grados y Títulos.
- Realizar todo el proceso Data Warehousing.