

UNIVERSIDAD CENTROAMERICANA “JOSÉ SIMEÓN CAÑAS”
FACULTAD DE INGENIERÍA Y ARQUITECTURA



Manual de instalación

Asignatura:

Tecnologías para Inteligencia de Negocios

Catedrática:

Ing. Erika Ramírez

Grupo:

Pentaho

Estudiantes:

Andrea María Peña Hernández	00075919
Fernando José Meléndez Orellana	00018720
Marcos Benjamín Granillo Flores	00028320
Néstor Esaú Recinos Escobar	00083120

Fecha de entrega: lunes 18 de noviembre de 2024

Conexión de la base de datos

- Gestor de base de datos: Oracle Database 21c for Microsoft Windows (64-bit)
- Entorno de desarrollo: SQL Developer 23.1.1
- Usuario: ETL_USER
- Contraseña: 1235

Nueva / Seleccionar Conexión a Base de Datos

Nombre de Cone...	Detalles de Cone...
CLASEBI	CLASEBI@//local...
CLASES	sys@//localhost:...
prueba	sys@//localhost:...
TAREA1	ETL_USER@//loc...
TAREA2	ETL_USER@//loc...

Name: PROYECTO

Tipo de Base de Datos: Oracle

Información de usuario

Tipo de autenticación: Por defecto

Usuario: ETL_USER

Contraseña: ****

Rol: valor por defecto

Guardar Contraseña: ☐

Tipo de Conexión: Básico

Detalles

Nombre del Host: localhost

Puerto: 1521

☒ SID: ord

☐ Nombre del Servicio

Estado: Correcto

Ayuda Guardar Borrar Probar Conectar Cancelar

Creación de parámetros

Dentro el archivo "Creación de parámetros" crear el esquema "PARAMETROS" y asignar permisos de administrador al esquema creado:

```
-- CREAR NUEVO USUARIO (SCHEMA)
CREATE USER ETL_USER IDENTIFIED BY 1235;
CREATE USER PARAMETROS IDENTIFIED BY 1235;

-- ASIGNAR PERMISO DE ADMINISTRADOR A LOS SCHEMAS
GRANT DBA TO ETL_USER;
GRANT DBA TO PARAMETROS;
```

Crear tabla de parámetros bajo el esquema de PARAMETROS:

```
-- CREAR TABLA DE PARÁMETROS DE CONEXIÓN
CREATE TABLE PARAMETROS.PARAMETROS (
    SISTEMA          VARCHAR2 (100 CHAR),
    NOMBRE_PARAM     VARCHAR2 (100 CHAR),
    VALOR_PARAM      VARCHAR2 (100 CHAR)
);
```

Insertar las variables a utilizar bajo el esquema de parámetros:

```
-- INSERTAR VARIABLES DE CONEXIÓN
INSERT INTO PARAMETROS.PARAMETROS (SISTEMA, NOMBRE_PARAM, VALOR_PARAM) VALUES ('PROYECTO_PENTAH0','HOSTNAME','localhost');
INSERT INTO PARAMETROS.PARAMETROS (SISTEMA, NOMBRE_PARAM, VALOR_PARAM) VALUES ('PROYECTO_PENTAH0','DATABASENAME','orcl');
INSERT INTO PARAMETROS.PARAMETROS (SISTEMA, NOMBRE_PARAM, VALOR_PARAM) VALUES ('PROYECTO_PENTAH0','PORTNUMBER','1521');
INSERT INTO PARAMETROS.PARAMETROS (SISTEMA, NOMBRE_PARAM, VALOR_PARAM) VALUES ('PROYECTO_PENTAH0','USERNAME','ETL_USER');
INSERT INTO PARAMETROS.PARAMETROS (SISTEMA, NOMBRE_PARAM, VALOR_PARAM) VALUES ('PROYECTO_PENTAH0','SCHEMA','ODS');
INSERT INTO PARAMETROS.PARAMETROS (SISTEMA, NOMBRE_PARAM, VALOR_PARAM) VALUES ('PROYECTO_PENTAH0','DIR_FILE','C:\WAITS\PROYECTO\');
INSERT INTO PARAMETROS.PARAMETROS (SISTEMA, NOMBRE_PARAM, VALOR_PARAM) VALUES ('PROYECTO_PENTAH0','PASSWORD','1235');
INSERT INTO PARAMETROS.PARAMETROS (SISTEMA, NOMBRE_PARAM, VALOR_PARAM) VALUES ('PROYECTO_PENTAH0','SCHEMA_DW','DATA_WAREHOUSE');
COMMIT;
```

Creación de la réplica de datos

Dentro del archivo “Creación de tablas” crear el esquema “DATA_WAREHOUSE” y asignar los permisos de administrador al esquema creado:

```
-- CREAR NUEVO USUARIO (SCHEMA)
CREATE USER ETL_USER IDENTIFIED BY 1235;
CREATE USER DATA_WAREHOUSE IDENTIFIED BY 1235;

-- ASIGNAR PERMISO DE ADMINISTRADOR A LOS SCHEMAS
GRANT DBA TO ETL_USER;
GRANT DBA TO DATA_WAREHOUSE;
```

Crear las tablas a utilizar para la replica de las hojas de Excel, cada tabla debe crearse bajo el esquema de DATA_WAREHOUSE:

```
-- CREAR TABLA DE SALES_ORDER
CREATE TABLE DATA_WAREHOUSE.SALES_ORDER (
    SALESORDERLINEKEY INTEGER PRIMARY KEY,
    CHANNEL VARCHAR2(50 CHAR),
    SALESORDER VARCHAR2(50 CHAR),
    SALESORDERLINE VARCHAR2(50 CHAR)
);

-- CREAR TABLA DE PRODUCT
CREATE TABLE DATA_WAREHOUSE.PRODUCT (
    PRODUCTKEY INTEGER PRIMARY KEY,
    SKU VARCHAR2(50 CHAR),
    PRODUCT VARCHAR2(100 CHAR),
    STANDARDCOST NUMBER,
    COLOR VARCHAR2(50 CHAR),
    LISTPRICE NUMBER,
    MODEL VARCHAR2(50 CHAR),
    SUBCATEGORY VARCHAR2(100 CHAR),
    CATEGORY VARCHAR2(100 CHAR)
);

-- CREAR TABLA DE SALES_TERRITORY
CREATE TABLE DATA_WAREHOUSE.SALES_TERRITORY (
    SALESTERRITORYKEY INTEGER PRIMARY KEY,
    REGION VARCHAR2(50 CHAR),
    COUNTRY VARCHAR2(50 CHAR),
    GROUPSALES VARCHAR2(50 CHAR)
);

-- CREAR TABLA DE RESELLER
CREATE TABLE DATA_WAREHOUSE.RESELLER (
    RESELLERKEY INTEGER PRIMARY KEY,
    RESELLERID VARCHAR2(50 CHAR),
    BUSINESSTYPE VARCHAR2(100 CHAR),
    RESELLER VARCHAR2(100 CHAR),
    CITY VARCHAR2(50 CHAR),
    STATEPROVINCE VARCHAR2(50 CHAR),
    COUNTRYREGION VARCHAR2(50 CHAR),
    POSTALCODE VARCHAR2(50 CHAR)
);

CREATE TABLE DATA_WAREHOUSE.CUSTOMER (
    CUSTOMERKEY INTEGER PRIMARY KEY,
    CUSTOMERID VARCHAR2(50 CHAR),
    CUSTOMER VARCHAR2(100 CHAR),
    CITY VARCHAR2(50 CHAR),
    STATEPROVINCE VARCHAR2(50 CHAR),
    COUNTRYREGION VARCHAR2(50 CHAR),
    POSTALCODE VARCHAR2(50 CHAR)
);

-- CREAR TABLA DE DIM_TIME
CREATE TABLE DATA_WAREHOUSE.DIM_TIME (
    FECHAID NUMBER,
    FECHA NUMBER NOT NULL,
    ORDERDATEKEY DATE PRIMARY KEY,
    AÑO NUMBER NOT NULL,
    TRIMESTRE NUMBER NOT NULL,
    MES NUMBER NOT NULL,
    SEMANA NUMBER NOT NULL,
    DIA NUMBER NOT NULL,
    DIASEMANA NUMBER NOT NULL,
    NTRIMESTRE VARCHAR2(7) NOT NULL,
    NMES VARCHAR2(15) NOT NULL,
    NMES3L VARCHAR2(3) NOT NULL,
    NSEMANA VARCHAR2(10) NOT NULL,
    NDIA VARCHAR2(6) NOT NULL,
    NDIASEMANA VARCHAR2(10) NOT NULL
);

-- CREAR TABLA DE FACT SALES
CREATE TABLE DATA_WAREHOUSE.FACT_SALES (
    SALESORDERLINEKEY INTEGER PRIMARY KEY,
    RESELLERKEY INTEGER, -- FK DE RESELLER DIMENSION
    CUSTOMERKEY INTEGER, -- FK DE CUSTOMER DIMENSION
    PRODUCTKEY INTEGER, -- FK DE PRODUCT DIMENSION
    ORDERDATEKEY DATE, -- FK DE TIME DIMENSION
    DUEDATEKEY DATE,
    SHIPDATEKEY DATE,
    SALESTERRITORYKEY INTEGER, -- FK DE TERRITORY DIMENSION
    ORDERQUANTITY NUMBER,
    UNITPRICE NUMBER,
    EXTENDEDAMOUNT NUMBER,
    UNITPRICEDISCOUNTPCT NUMBER,
    PRODUCTSTANDARDCOST NUMBER,
    TOTALPRODUCTCOST NUMBER,
    SALESAMOUNT NUMBER,
    PAYMENTMETHOD VARCHAR2(50 CHAR)
);
```

Agregar las llaves foráneas a la tabla “FACT_SALES”:

```
-- AGREGAR LLAVES FORÁNEAS A LA TABLA FACT_SALES
-- Foreign key con Sales_Order
ALTER TABLE DATA_WAREHOUSE.FACT_SALES ADD CONSTRAINT SALESORDERLINEKEY FOREIGN KEY (SALESORDERLINEKEY) REFERENCES DATA_WAREHOUSE.SALES_ORDER (SALESORDERLINEKEY);
-- Foreign key con Reseller
ALTER TABLE DATA_WAREHOUSE.FACT_SALES ADD CONSTRAINT RESELLERKEY FOREIGN KEY (RESELLERKEY) REFERENCES DATA_WAREHOUSE.RESELLER (RESELLERKEY);
-- Foreign key con Customer
ALTER TABLE DATA_WAREHOUSE.FACT_SALES ADD CONSTRAINT CUSTOMERKEY FOREIGN KEY (CUSTOMERKEY) REFERENCES DATA_WAREHOUSE.CUSTOMER (CUSTOMERKEY);
-- Foreign key con Product
ALTER TABLE DATA_WAREHOUSE.FACT_SALES ADD CONSTRAINT PRODUCTKEY FOREIGN KEY (PRODUCTKEY) REFERENCES DATA_WAREHOUSE.PRODUCT (PRODUCTKEY);
-- Foreign key con Dim Time
ALTER TABLE DATA_WAREHOUSE.FACT_SALES ADD CONSTRAINT ORDERDATEKEY FOREIGN KEY (ORDERDATEKEY) REFERENCES DATA_WAREHOUSE.DIM_TIME (ORDERDATEKEY);
-- Foreign key con Sales_Territory
ALTER TABLE DATA_WAREHOUSE.FACT_SALES ADD CONSTRAINT SALESTERRITORYKEY FOREIGN KEY (SALESTERRITORYKEY) REFERENCES DATA_WAREHOUSE.SALES_TERRITORY (SALESTERRITORYKEY);
```

Ejecutar el siguiente procedimiento para crear la dimensión de tiempo (solo es necesario ejecutarlo cada vez que se crea la tabla DIM_TIME):

```
-- PROCEDIMIENTO PARA GENERAR LA DIM TIEMPO
DECLARE
    IdIni number;
    FechaDesde date;
    FechaHasta date;
BEGIN
    -- RANGO DE FECHAS A GENERAR: del 01/07/2017 al 31/12/2024 + 5
    IdIni := 1;
    FechaDesde := TO_DATE('20170701','YYYYMMDD');
    FechaHasta := TO_DATE((TO_CHAR(sysdate,'YYYY') + 5 || '1231'),'YYYYMMDD');
    WHILE FechaDesde <= FechaHasta LOOP
        INSERT INTO DATA_WAREHOUSE.DIM_TIME
        (
            FECHAID,
            FECHA,
            ORDERDATEKEY,
            AÑO,
            TRIMESTRE,
            MES,
            SEMANA,
            DIA,
            DIASEMANA,
            NTRIMESTRE,
            NMES,
            NMES3L,
            NSEMANA,
            NDIA,
            NDIASEMANA
        )
        VALUES
        (
            IdIni,
            to_char(FechaDesde,'YYYYMMDD'),
            FechaDesde,
            to_char(FechaDesde,'YYYY'),
            to_char(FechaDesde,'Q'),
            to_char(FechaDesde,'MM'),
            to_char(FechaDesde,'WW'),
            to_char(FechaDesde,'DD'),
            to_char(FechaDesde,'D'),
            'T'||to_char(FechaDesde,'Q')||'/'||to_char(FechaDesde,'YY'),
            to_char(FechaDesde,'MONTH'),
            to_char(FechaDesde,'MON'),
            'Sem'||to_char(FechaDesde,'WW')||'/'||to_char(FechaDesde,'YY'),
            to_char(FechaDesde,'DD MON'),
            to_char(FechaDesde,'DAY')
        );
        --INCREMAR BUCLE
        IdIni := IdIni + 1;
        FechaDesde := FechaDesde + 1;
    END LOOP;
END;
```

Conexión a la base de datos Oracle desde Pentaho

- Tabla del esquema PARAMETROS

SISTEMA	NOMBRE_PARAM	VALOR_PARAM
PROYECTO_PENTAHO	HOSTNAME	localhost
PROYECTO_PENTAHO	DATABASENAME	orcl
PROYECTO_PENTAHO	PORTNUMBER	1521
PROYECTO_PENTAHO	USERNAME	ETL_USER
PROYECTO_PENTAHO	SCHEMA	ODS
PROYECTO_PENTAHO	DIR_FILE	C:\WAITS\PROYECTO\
PROYECTO_PENTAHO	PASSWORD	1235
PROYECTO_PENTAHO	SCHEMA_DW	DATA_WAREHOUSE

- Para un entorno de pruebas/creación agregar las variables del esquema PARAMETROS en la sección “Set Environment Variables”

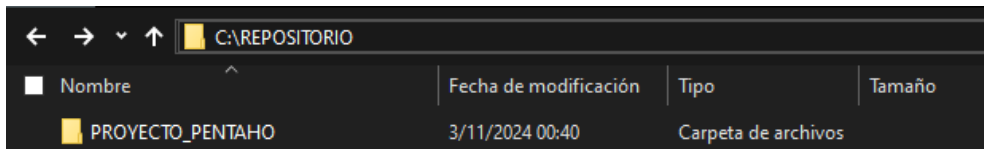
#	Name	Value
1	DATABASENAME	orcl
2	DIR_FILE	C:\WAITS\PROYECTO\
3	HOSTNAME	localhost
4	PASSWORD	1235
5	PORTNUMBER	1521
6	SCHEMA	ODS
7	SCHEMA_DW	DATA_WAREHOUSE
8	SISTEMA	
9	USERNAME	ETL_USER

- Valores utilizados antes de parametrizar las variables

- Variables parametrizadas

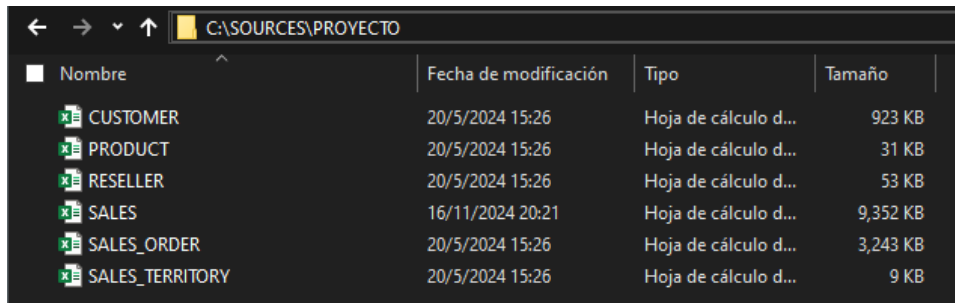
Ubicación de la carpeta con los archivos de Pentaho

- Colocar la carpeta **PROYECTO_PENTAHO** en la ruta C:\RESPOSITORIO



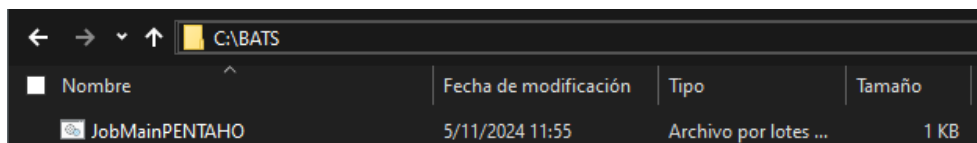
←	→	⌵	⬆	C:\RESPOSITORIO
■	Nombre	Fecha de modificación	Tipo	Tamaño
■	PROYECTO_PENTAHO	3/11/2024 00:40	Carpeta de archivos	

- Colocar los archivos de Excel en la ruta C:\SOURCES\PROYECTO



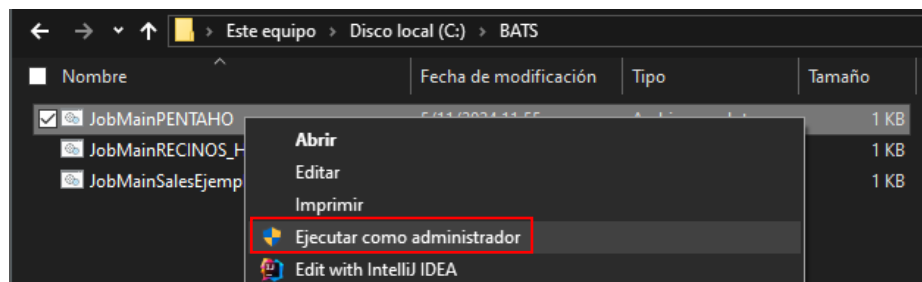
←	→	⌵	⬆	C:\SOURCES\PROYECTO
■	Nombre	Fecha de modificación	Tipo	Tamaño
📄	CUSTOMER	20/5/2024 15:26	Hoja de cálculo d...	923 KB
📄	PRODUCT	20/5/2024 15:26	Hoja de cálculo d...	31 KB
📄	RESELLER	20/5/2024 15:26	Hoja de cálculo d...	53 KB
📄	SALES	16/11/2024 20:21	Hoja de cálculo d...	9,352 KB
📄	SALES_ORDER	20/5/2024 15:26	Hoja de cálculo d...	3,243 KB
📄	SALES_TERRITORY	20/5/2024 15:26	Hoja de cálculo d...	9 KB

- Colocar el archivo **JobMainPentaho** en la ruta C:\BATS



←	→	⌵	⬆	C:\BATS
■	Nombre	Fecha de modificación	Tipo	Tamaño
📄	JobMainPENTAH0	5/11/2024 11:55	Archivo por lotes ...	1 KB

- Ejecutar el archivo .bat como administrador



←	→	⌵	⬆	Este equipo > Disco local (C:) > BATS
■	Nombre	Fecha de modificación	Tipo	Tamaño
☑	JobMainPENTAH0	5/11/2024 11:55	Archivo por lotes ...	1 KB
📄	JobMainRECINOS_H			1 KB
📄	JobMainSalesEjemp			1 KB

Abrir

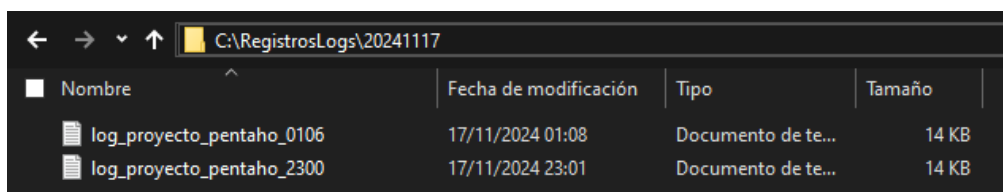
Editar

Imprimir

Ejecutar como administrador

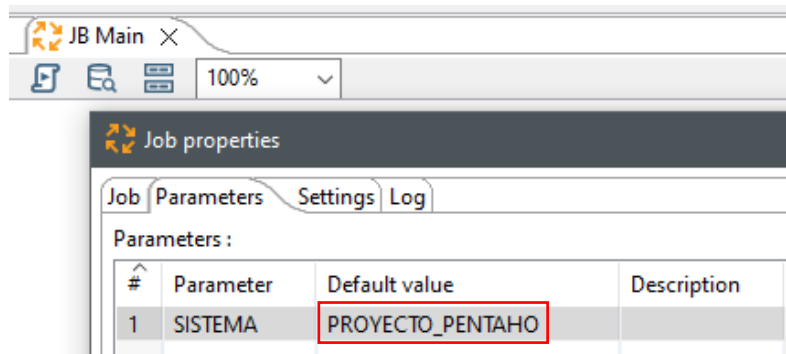
Edit with IntelliJ IDEA

- Al finalizar la ejecución, se creará la carpeta “**RegistrosLogs**”, dentro de la cual se crean carpetas que contienen archivos de texto con la información del proceso ETL realizado.



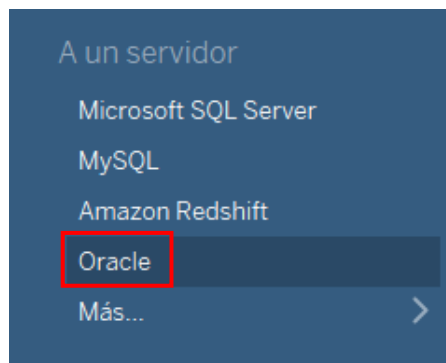
←	→	⌵	⬆	C:\RegistrosLogs\20241117
■	Nombre	Fecha de modificación	Tipo	Tamaño
📄	log_proyecto_pentaho_0106	17/11/2024 01:08	Documento de te...	14 KB
📄	log_proyecto_pentaho_2300	17/11/2024 23:01	Documento de te...	14 KB

- Si se quiere ejecutar el flujo ETL desde Pentaho se debe añadir la variable del sistema en el job principal, en nuestro caso corresponde al archivo **JB Main**.



Conexión a la base de datos Oracle desde Tableau

- Conectarse a un servidor, en nuestro caso sería con Oracle



- Ingresar los datos de conexión

- Al establecer la conexión iniciar sesión

Oracle

localhost

[Editar conexión](#)

Nombre de usuario:

ETL_USER

Contraseña:

••••

Iniciar sesión

- En “**Fuente de datos**” seleccionar las tablas y crear las relaciones según la PF de “FACT_SALES” con respecto a las FK del resto de tablas.

¿En que se diferencian las relaciones de la unión de columnas? [Más información](#)

CUSTOMER

Operador

FACT_SALES

Customerkey

=

Customerkey (Fact Sales)

¿En que se diferencian las relaciones de la unión de columnas? [Más información](#)

DIM_TIME

Operador

FACT_SALES

Orderdatekey

=

Orderdatekey (Fact Sales)

¿En que se diferencian las relaciones de la unión de columnas? [Más información](#)

PRODUCT

Operador

FACT_SALES

Productkey

=

Productkey (Fact Sales)

¿En que se diferencian las relaciones de la unión de columnas? [Más información](#)

FACT_SALES

Operador

RESELLER

Resellerkey

=

Resellerkey (Reseller)

¿En que se diferencian las relaciones de la unión de columnas? [Más información](#)

FACT_SALES

Operador

SALES_ORDER

Salesorderlinekey

=

Salesorderlinekey (Sales C

¿En que se diferencian las relaciones de la unión de columnas? [Más información](#)

FACT_SALES	Operador	SALES_TERRITORY
# Salesterritorykey ▼	= ▼	# Salesterritorykey (Sales Tt ▼

- Al finalizar, debería quedar el siguiente esquema con las tablas conectadas

