JUAN LUIS ACEBAL RICO

PR2 BDA

INDICE

INTRODUCCIÓN

En el contexto de los “datos”, con una creciente importancia y clave para la toma de decisiones su análisis, siendo las elecciones presidenciales de noviembre de 2024 un momento importante para el destino de la humanidad, ya que EEUU es la primera economía, y muchas decisiones económicas nos afectan, siendo el próximo presidente de EEUU elegido por los estadounidenses para casi gobernar todo el mundo (aún más la parte occidental), por ello, en base al análisis y diseño hecho de un Data Warehouse en la actividad precedente, ahora es el momento de la carga de datos, estructurándose en 3 partes:

- Identificación de procesos ETL
- Diseño y desarrollo de los procesos ETL
- Implementación de los trabajos “jobs” (la automatización)

Estos procesos tienen como objetivo buscar patrones y tendencias que influencian a la política del país y a la economía. Por tanto, parte de los datos utilizados son la cotización del índice bursátil S&P500, que está formado por las 500 mayores empresas de EE. UU..

La implementación del almacén de datos supone la extracción de los datos en tablas intermedias donde los modificaremos (joins, cambios de nombre, lo que proceda en cada caso) para finalmente llevarlos a tablas finales que se usarán para crear las dimensiones y hechos de nuestro Data WareHouse

Primero se hará el bloque IN, que es el modelo intermedio que se usará para crear el modelo multidimensional (transformado) llamado bloque TR.

Dentro de cada bloque se muestra las transformaciones creadas con PDI Spoon, la creación de tablas, etc. Por último, se automatiza todo con Jobs.

Gracias a la práctica, vamos a poder entender bien el proceso de creación de un almacén de datos, dentro de la creación de un Data WareHouse, y, además, vamos a poder recopilar y observar datos muy interesantes y sacar conclusiones de los patrones que existen en los ciclos presidenciales americanos.

1. IDENTIFICACIÓN DE LOS PROCESOS ETL

BLOQUE IN

Extracción de los datos desde las fuentes a tablas intermedias.
En esta parte, vamos simplemente a crear las tablas, y cargar datos. Estas tablas se utilizarán posteriormente para, junto a una manipulación de datos, crear la transformación final en otras tablas, con otros modelos y estructura. Esta zona de tablas intermedias se llama staging.

De los procesos que vamos a realizar, hay 5:

Nombre ETL	Descripción	Origen de datos	Tabla de destino(stage)
IN_PRESIDENT	Carga de la información relativa a la evolución de las elecciones para la presidencia	Archivo 1976-2020-president.tab	STG_PRESIDENT
IN_SENATE	Carga de la información relativa a la evolución de las elecciones para el senado	Archivo 1976-2020-senate.tab	STG_SENATE
IN_HOUSE	Carga de la información relativa a la evolución de las elecciones para la cámara de representantes (congreso de EEUU)	Archivo 1976-2020-house.tab	STG_HOUSE
IN_STOCK_INDEX	Carga de datos históricos del índice bursátil americano S&P 500	Archivo csv dado que su procedencia parece ser yahoo finance	STG_ELECTIONS
IN_STATE_OFFICES	Carga de datos de los estados americanos	Archivo txt que tiene los estados americanos y su código	STG_STATE_OFFICES

Hasta aquí, cada conjunto de datos es leído, y son almacenados sin manipular en tablas (staging área). Aquí no se harán modificaciones a excepción de manipular bien los datos para que lleguen a la fase de transformación de la manera mas óptima posible (es decir, si hay alguna columna mal nombrada, algún campo mal detectado por el software de manipulación Spoon PDI, pero sino, las manipulaciones serán después)

BLOQUE TR

DIMENSIONES

Las dimensiones o tablas DIM, son las tablas que se conectarán con las tablas de hechos, pudiendo dar a la tabla de hechos información en diferentes métricas tales como tiempo, partido político, candidato, estado, distrito o tipo de elección.
Aquí tenemos 5 TR_DIM:

Nombre ETL	Descripción	Origen de datos	Tabla de destino(stage)
TR_DIM_CANDIDATES	Carga de la dimensión con información de los candidatos	STG_PRESIDENT STG_SENATE STG_HOUSE	DIM_CANDIDATES
TR_DIM_PARTY	Carga de datos de la dimensión con información de los partidos políticos.	STG_PRESIDENT STG_SENATE STG_HOUSE	DIM_PARTY
TR_DIM_YEAR	Carga datos temporales, organizado por años electorales	STG_PRESIDENT STG_SENATE STG_HOUSE STG_STOCK_INDEX	DIM_YEAR
TR_DIM_STATE	Carga de los datos de los estados	STG_PRESIDENT STG_SENATE STG_HOUSE STG_STATE_OFFICES	DIM_STATE
TR_DIM_DISTRICT	Carga y transformación de datos de distritos	STG_PRESIDENT STG_SENATE STG_HOUSE STG_STATE_OFFICES	DIM_DISTRICT
TR_DIM_OFFICES	Carga datos de tipo de elecciones	STG_PRESIDENT STG_SENATE STG_HOUSE	DIM_OFFICES

HECHOS

Nombre ETL	Descripción	Origen de datos	Tabla destino
TR_FACT_ELECTIONS_EEUU	Carga de la información relativa a los resultados de las elecciones de EEUU	STG_PRESIDENT STG_SENATE STG_HOUSE	FACT_ELECTIONS_EEUU
TR_FACT_STOCKINDEX	Transformación y carga de los datos del índice S&P500 en la tabla de hechos.	STG_PRESIDENT STG_SENATE STG_HOUSE STG_STOCK_INDEX	FACT_STOCKINDEX

2. DISEÑO Y DESARROLLO DE LOS PROCESOS ETL

CREACIÓN DE LAS TABLAS

Propuesta de creación de tablas:

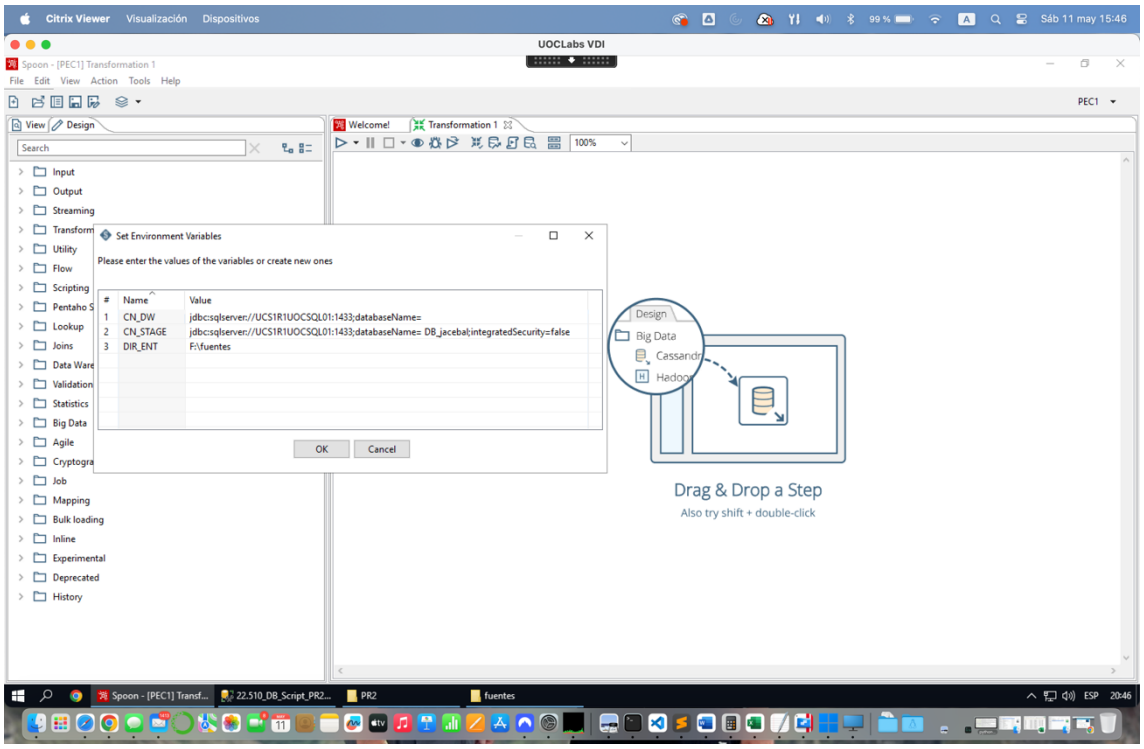
```
-- Eliminar tablas existentes si es necesario
USE SOURCE_jacebal;
DROP TABLE IF EXISTS dbo.FACT_ELECTIONS_EEUU;
DROP TABLE IF EXISTS dbo.FACT_STOCKINDEX;
DROP TABLE IF EXISTS dbo.DIM_OFFICES;
DROP TABLE IF EXISTS dbo.DIM_DISTRICT;
DROP TABLE IF EXISTS dbo.DIM_STATE;
DROP TABLE IF EXISTS dbo.DIM_YEAR;
DROP TABLE IF EXISTS dbo.DIM_PARTY;
DROP TABLE IF EXISTS dbo.DIM_CANDIDATES;
DROP TABLE IF EXISTS dbo.STG_STATE_OFFICES;
DROP TABLE IF EXISTS dbo.STG_STOCK_INDEX;
DROP TABLE IF EXISTS dbo.STG_HOUSE;
DROP TABLE IF EXISTS dbo.STG_SENATE;
DROP TABLE IF EXISTS dbo.STG_PRESIDENT;

-- STG_PRESIDENT
CREATE TABLE dbo.STG_PRESIDENT(
    [year] INT NOT NULL,
    [state] NVARCHAR(100) NOT NULL,
    [state_po] NVARCHAR(2) NOT NULL,
    [state_fips] INT NOT NULL,
    [state_cen] INT NOT NULL,
    [state_ic] INT NOT NULL,
    [office] NVARCHAR(100) NOT NULL,
    [candidate] NVARCHAR(200) NOT NULL,
    [party_detailed] NVARCHAR(100) NOT NULL,
    [writein] NVARCHAR(10) NULL,
    [candidatevotes] INT NOT NULL,
    [totalvotes] INT NOT NULL,
    [version] INT NOT NULL,
    [notes] NVARCHAR(50) NOT NULL,
    [party_simplified] NVARCHAR(100) NOT NULL
);

-- STG_SENATE
CREATE TABLE dbo.STG_SENATE (
    [year] INT NOT NULL,
    [state] NVARCHAR(100) NOT NULL,
    [state_po] NVARCHAR(2) NOT NULL,
    [candidate] NVARCHAR(200) NOT NULL,
    [party_detailed] NVARCHAR(100) NOT NULL,
    [writein] NVARCHAR(10) NULL,
    [candidatevotes] INT NOT NULL,
    [totalvotes] INT NOT NULL,
    [office] NVARCHAR(100) NOT NULL
);

-- STG_HOUSE
CREATE TABLE dbo.STG_HOUSE (
    [year] INT NOT NULL,
    [state] NVARCHAR(100) NOT NULL,
    [state_po] NVARCHAR(2) NOT NULL,
    [candidate] NVARCHAR(200) NOT NULL,
    [party_detailed] NVARCHAR(100) NOT NULL,
    [writein] NVARCHAR(10) NULL,
```

```
    [candidatevotes] INT NOT NULL,
    [totalvotes] INT NOT NULL,
    [office] NVARCHAR(100) NOT NULL
);

-- STG_STOCK_INDEX
CREATE TABLE dbo.STG_STOCK_INDEX (
    [date] DATE NOT NULL,
    [open] FLOAT NOT NULL,
    [high] FLOAT NOT NULL,
    [low] FLOAT NOT NULL,
    [close] FLOAT NOT NULL,
    [volume] BIGINT,
    [adj_close] FLOAT NOT NULL
);

-- STG_STATE_OFFICES
CREATE TABLE dbo.STG_STATE_OFFICES (
    [state] NVARCHAR(100) NOT NULL,
    [state_po] NVARCHAR(2) NOT NULL
);

-- DIM_CANDIDATES
CREATE TABLE [dbo].[DIM_CANDIDATES] (
    [candidate_pk] INT NOT NULL,
    [candidate] NVARCHAR(200) NOT NULL,
    [writen] NVARCHAR(10) NULL,
    CONSTRAINT [PK_DIM_CANDIDATES] PRIMARY KEY CLUSTERED (
        [candidate_pk] ASC
    ) WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY]
GO

-- DIM_PARTY
CREATE TABLE dbo.DIM_PARTY (
    [party_pk] INT NOT NULL,
    [party_detailed] NVARCHAR(100) NOT NULL,
    [party_simplified] NVARCHAR(100) NOT NULL,
    CONSTRAINT PK_DIM_PARTY PRIMARY KEY CLUSTERED ([party_pk] ASC)
    WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY];

-- DIM_YEAR
CREATE TABLE dbo.DIM_YEAR (
    [year_pk] INT NOT NULL,
    CONSTRAINT PK_DIM_YEAR PRIMARY KEY CLUSTERED ([year_pk] ASC)
    WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
);
GO

-- DIM_STATE
CREATE TABLE dbo.DIM_STATE (
    [state_pk] INT NOT NULL,
    [state] NVARCHAR(100),
    [state_po] NVARCHAR(2),
    [state_fips] INT,
    [state_cen] INT,
    [state_ic] INT,
    CONSTRAINT PK_DIM_STATE PRIMARY KEY CLUSTERED ([state_pk])
```

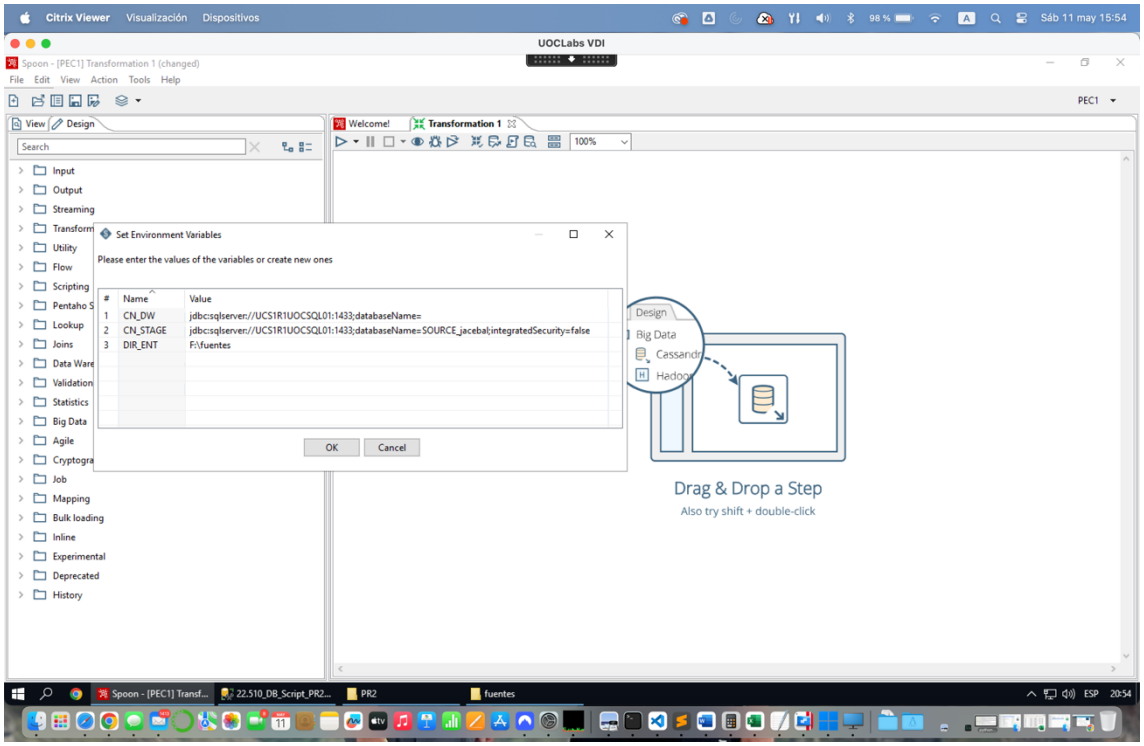
```
    WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
) ON [PRIMARY];
GO

-- DIM_DISTRICT
CREATE TABLE dbo.DIM_DISTRICT (
    [district_pk] INT NOT NULL IDENTITY(1,1),
    CONSTRAINT PK_DIM_DISTRICT PRIMARY KEY CLUSTERED ([district_pk] ASC)
    WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
) ON [PRIMARY];

-- DIM_OFFICES
CREATE TABLE dbo.DIM_OFFICES (
    [office_pk] INT NOT NULL,
    [office] NVARCHAR(100),
    CONSTRAINT PK_DIM_OFFICES PRIMARY KEY CLUSTERED ([office_pk])
    WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON)
) ON [PRIMARY];
GO

-- FACT_ELECTIONS_EEUU
CREATE TABLE dbo.FACT_ELECTIONS_EEUU (
    [election_id] BIGINT NOT NULL IDENTITY(1,1),
    [year_fk] INT NOT NULL,
    [state_fk] INT NOT NULL,
    [district_fk] INT,
    [office_fk] INT NOT NULL,
    [candidate_fk] INT NOT NULL,
    [party_fk] INT NOT NULL,
    [votes] BIGINT NOT NULL,
    CONSTRAINT PK_FACT_ELECTIONS_EEUU PRIMARY KEY CLUSTERED ([election_id] ASC)
    WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY],
    CONSTRAINT FK_FACT_ELECTIONS_EEUU_YEAR FOREIGN KEY ([year_fk]) REFERENCES
dbo.DIM_YEAR([year_pk]),
    CONSTRAINT FK_FACT_ELECTIONS_EEUU_STATE FOREIGN KEY ([state_fk]) REFERENCES
dbo.DIM_STATE([state_pk]),
    CONSTRAINT FK_FACT_ELECTIONS_EEUU_DISTRICT FOREIGN KEY ([district_fk])
REFERENCES dbo.DIM_DISTRICT([district_pk]),
    CONSTRAINT FK_FACT_ELECTIONS_EEUU_OFFICE FOREIGN KEY ([office_fk]) REFERENCES
dbo.DIM_OFFICES([office_pk]),
    CONSTRAINT FK_FACT_ELECTIONS_EEUU_CANDIDATE FOREIGN KEY ([candidate_fk])
REFERENCES dbo.DIM_CANDIDATES([candidate_pk]),
    CONSTRAINT FK_FACT_ELECTIONS_EEUU_PARTY FOREIGN KEY ([party_fk]) REFERENCES
dbo.DIM_PARTY([party_pk])
) ON [PRIMARY];

CREATE TABLE dbo.FACT_STOCKINDEX (
    [stock_id] INT NOT NULL IDENTITY(1,1),
    [year_fk] INT NOT NULL,
    [change] DECIMAL(18,2) NOT NULL,
    [range] DECIMAL(18,2) NOT NULL,
    CONSTRAINT PK_FACT_STOCKINDEX PRIMARY KEY CLUSTERED ([stock_id] ASC)
    WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY],
    CONSTRAINT FK_FACT_STOCKINDEX_YEAR FOREIGN KEY ([year_fk]) REFERENCES
dbo.DIM_YEAR([year_pk])
) ON [PRIMARY];
```

CREACION DE VARIABLES DE ENTORNO

Creamos 3 variables de entorno, una para el acceso a las fuentes, otra para conexión a la base de datos en la zona de staging, y otra para la zona TR.

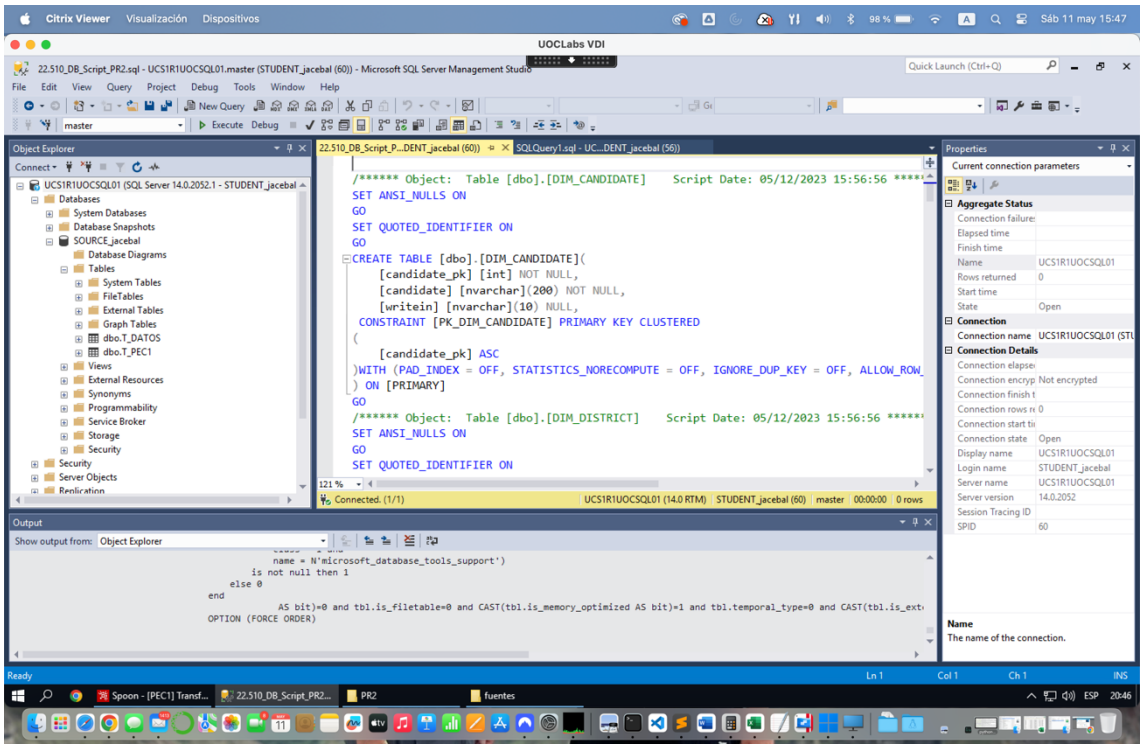


Modifico CN_STAGE que no lo puse bien

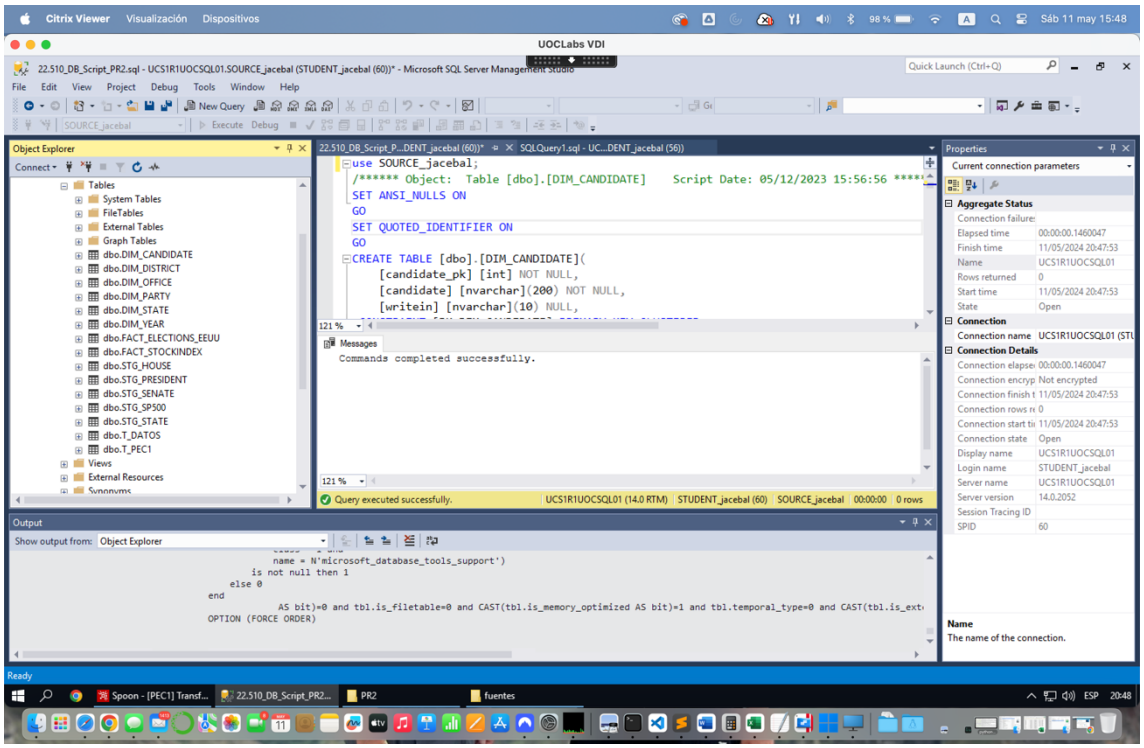


EJECUTAMOS EL SCRIPT EN MS SQL SERVER

Antes:

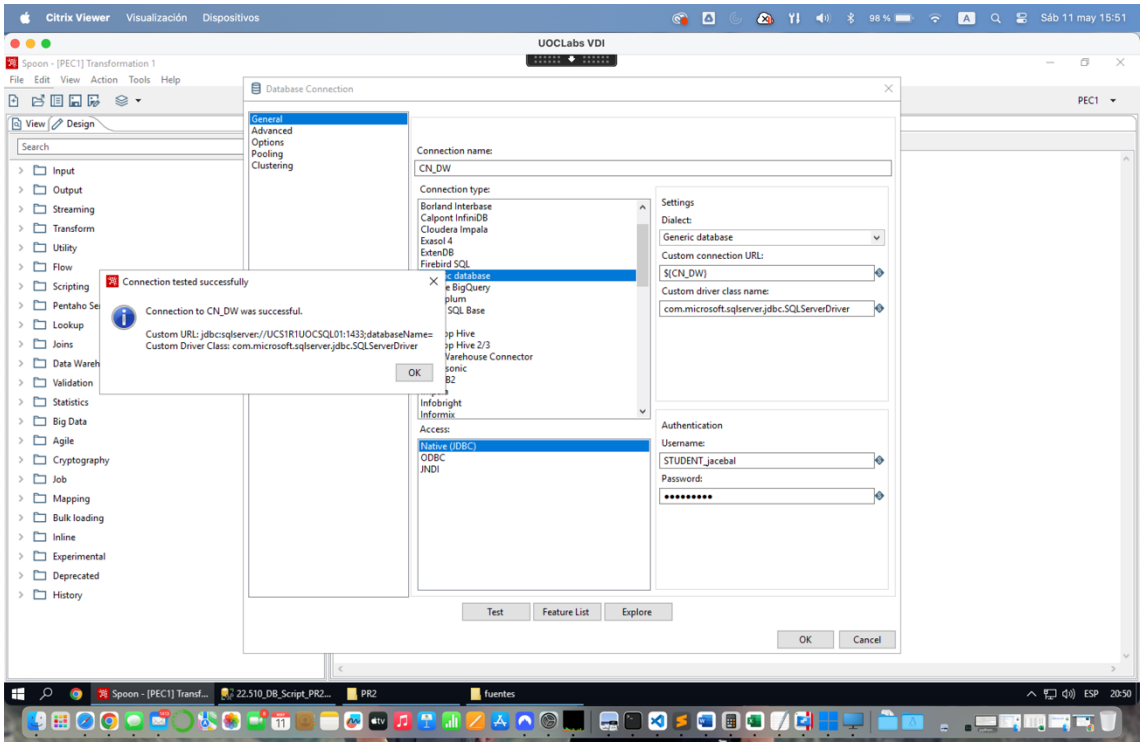


Después:

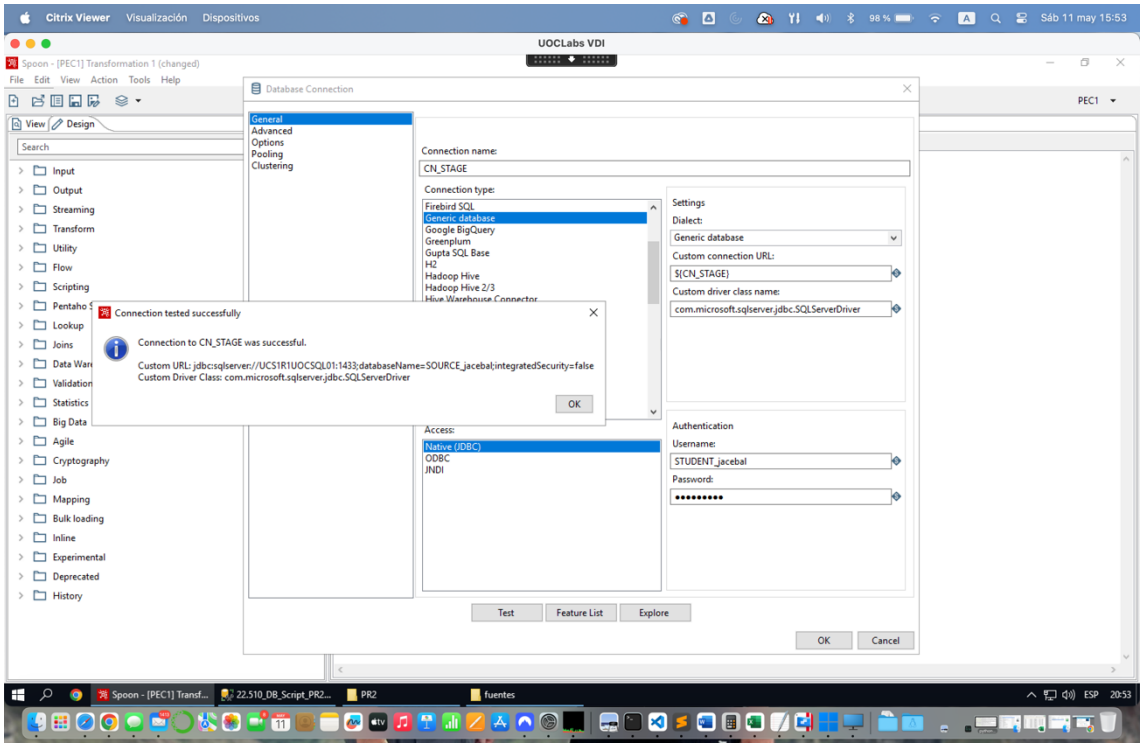


CONEXIONES A LA BASE DE DATOS

CN_DW



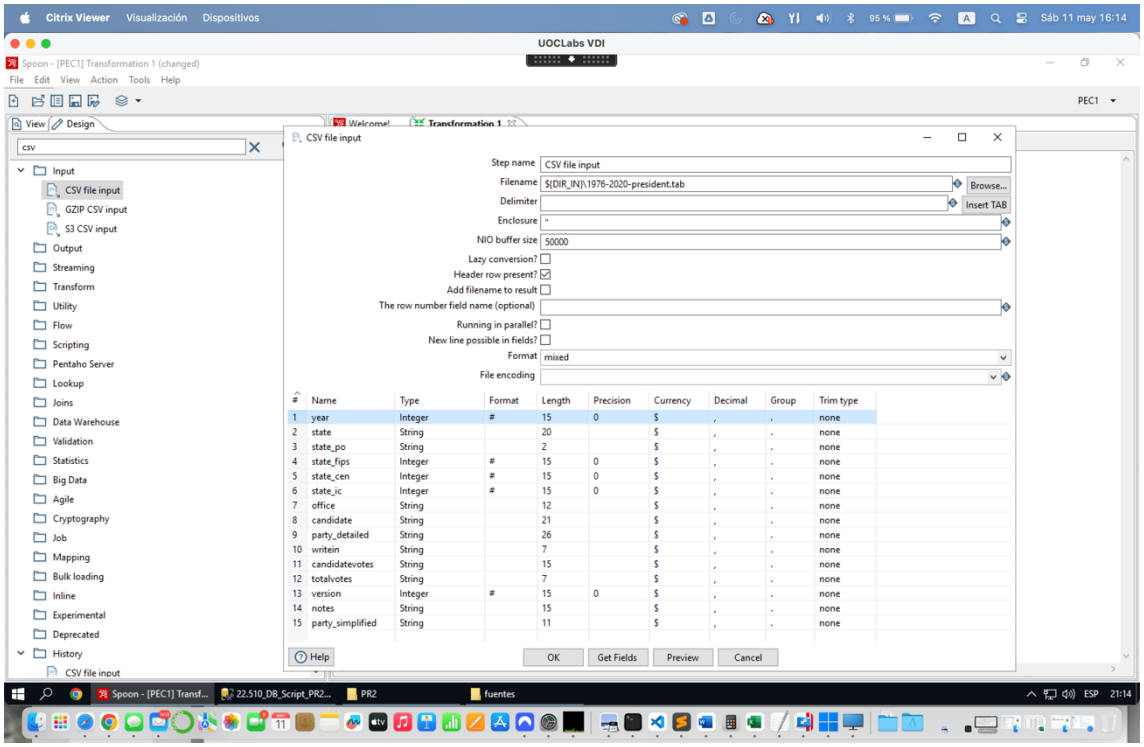
CN_STAGE



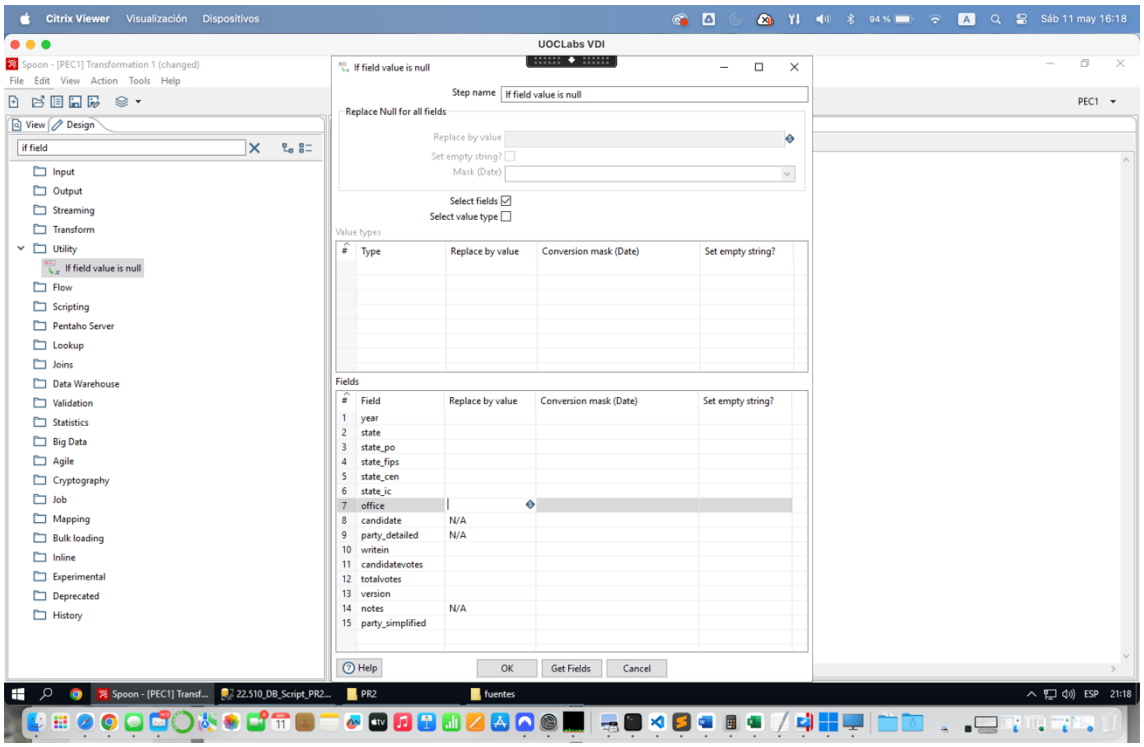
CREACIÓN BLOQUE IN

IN_PRESIDENT

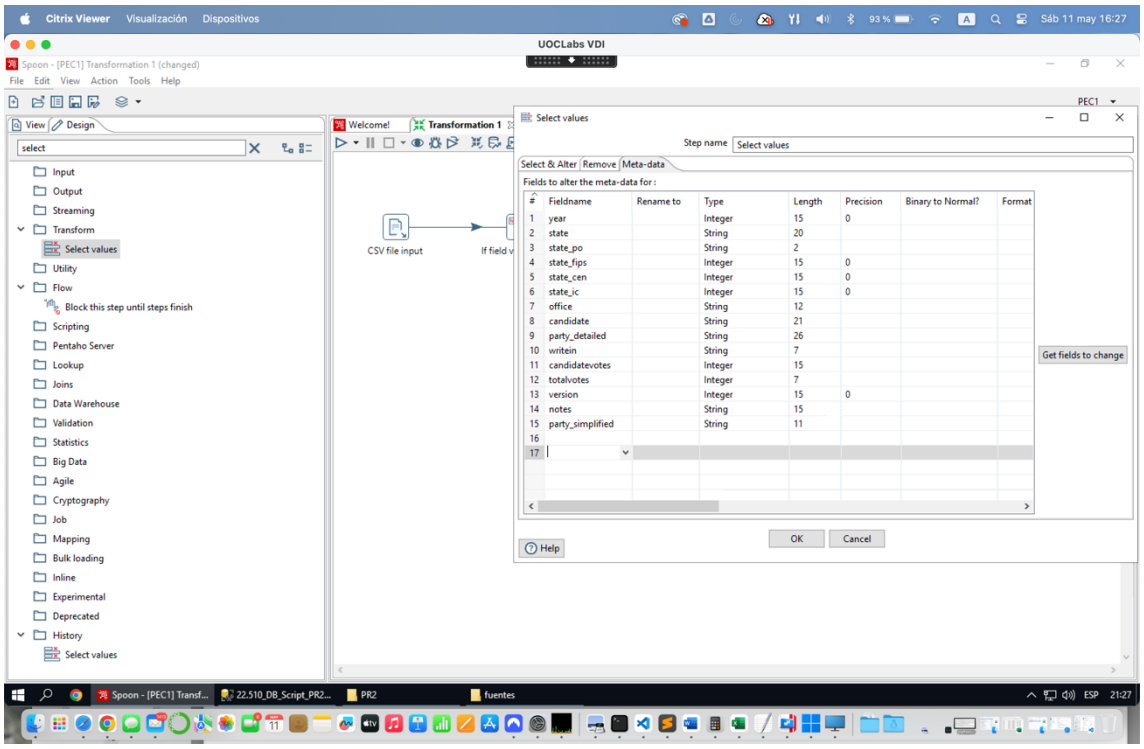
Creo el paso de la transformación CSV File input



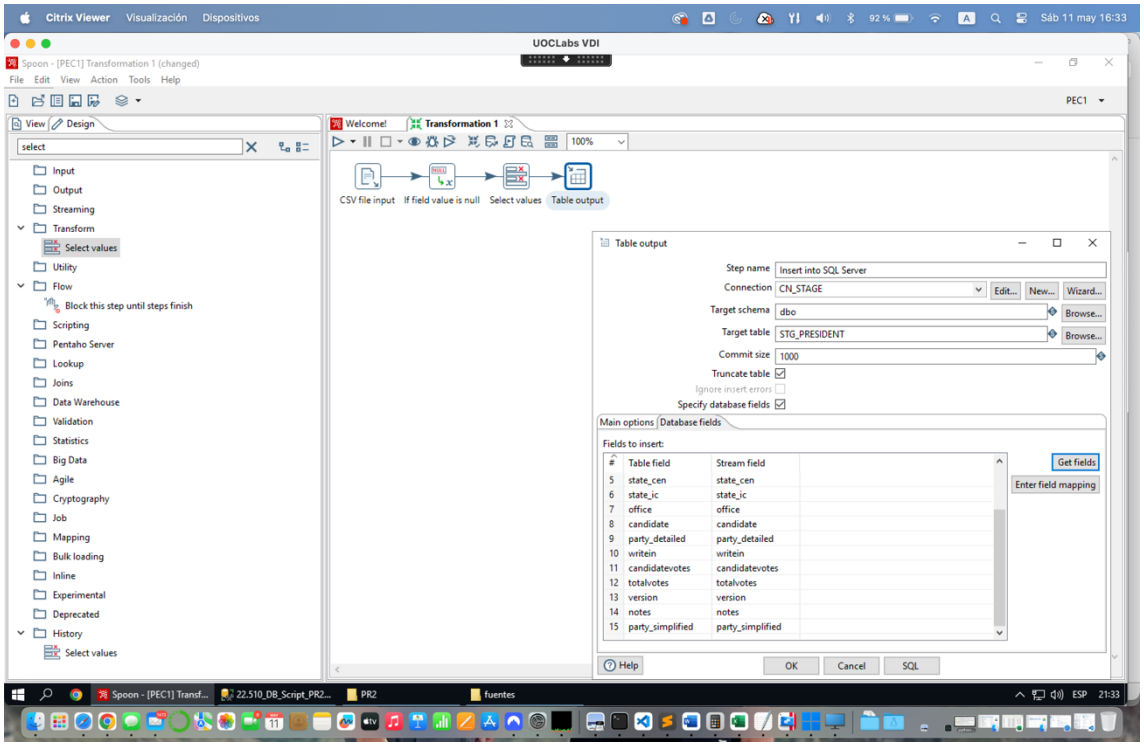
Creo el paso de la transformación Si el valor es nulo



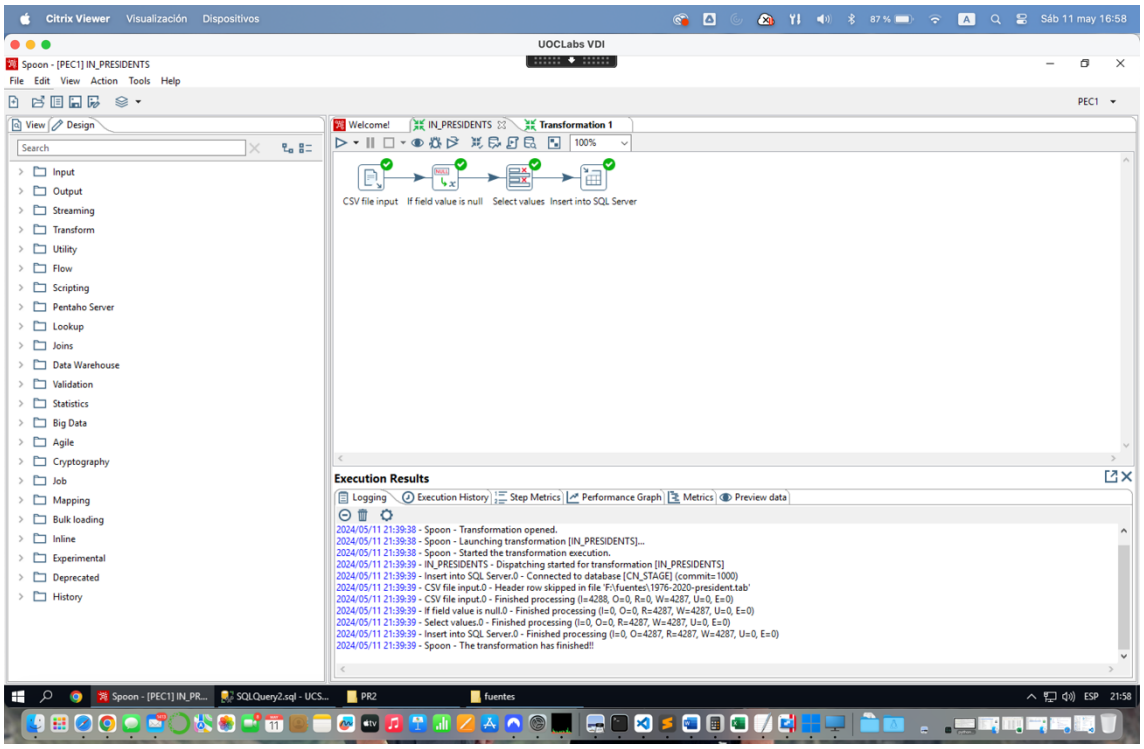
Paso seleccionar valores



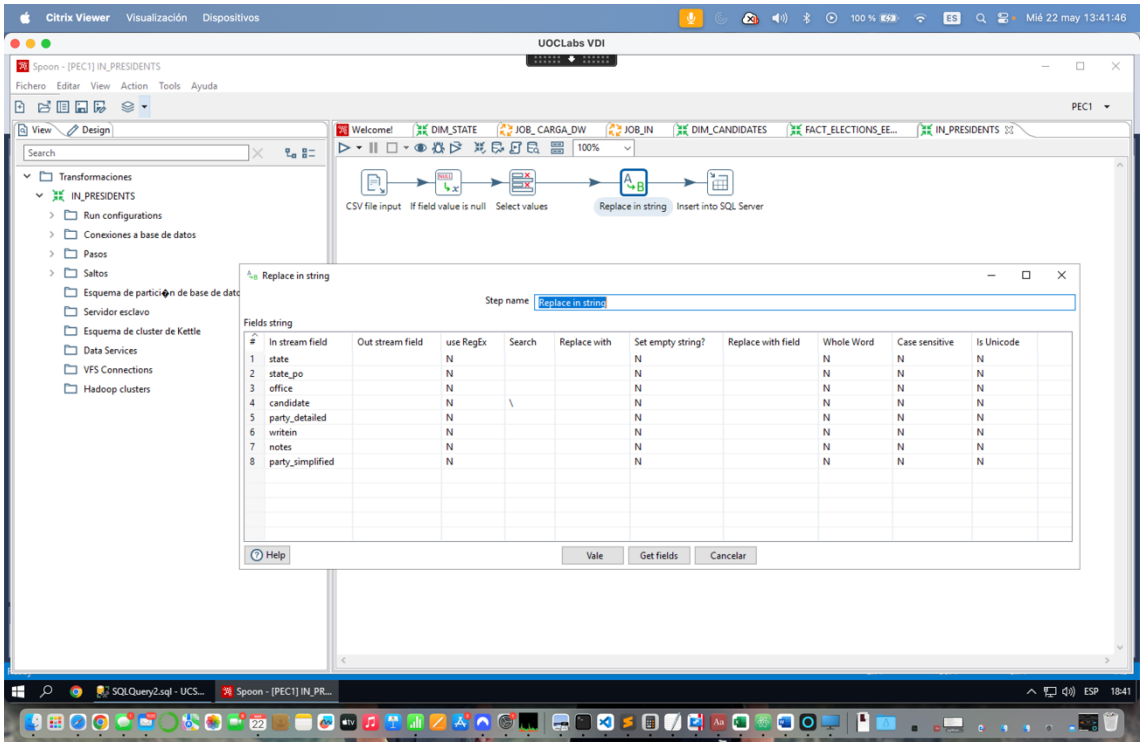
Paso insertar en SQL Server



Pruebo la transformación, y veo que han sido insertados 4287 registros

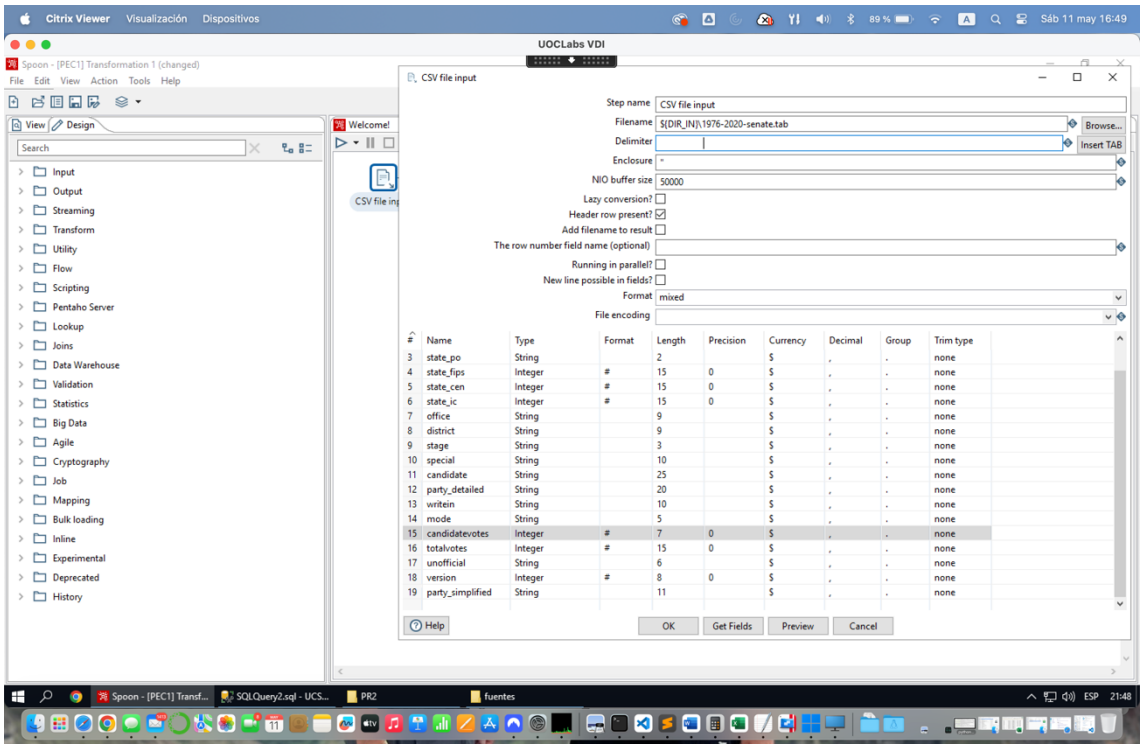


Actualizo: Añado quitar en candidatos la barra invertida

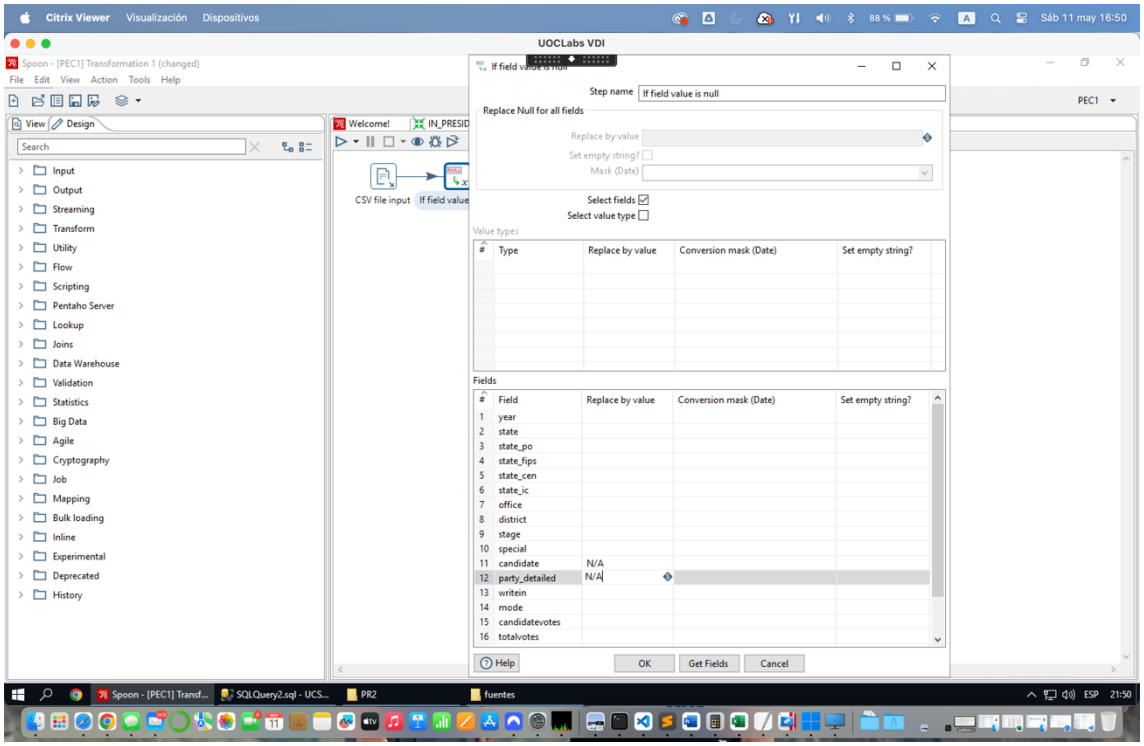


IN_SENATE

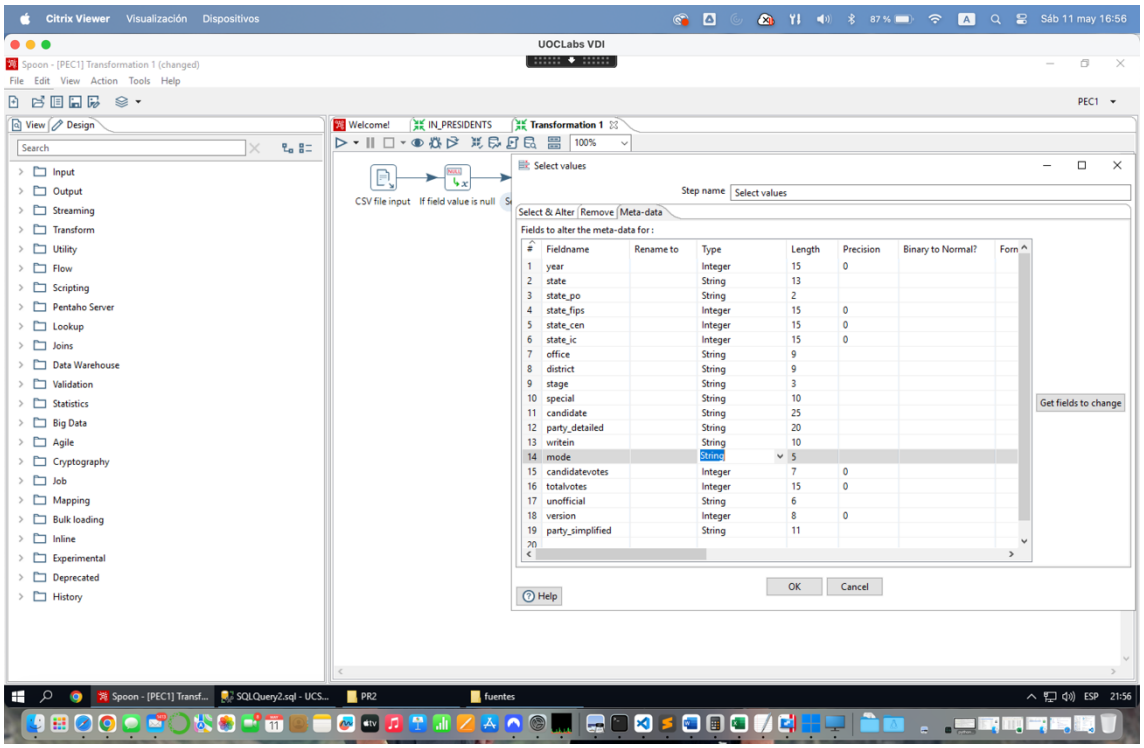
Creo el paso CSV Input, modifico los campos para que special sea string, y algunos otros mal detectados de integrer o string sean correctamente definidos



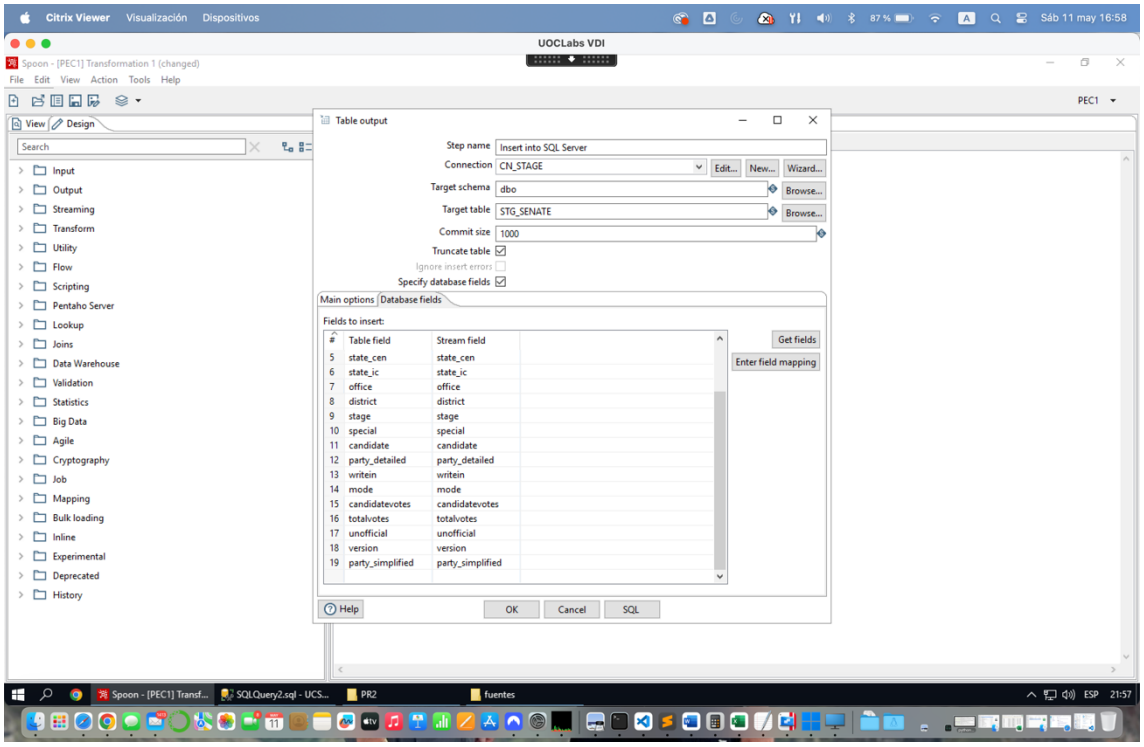
Después creo el paso si es nulo, revisando el preview del paso anterior, tenemos en esta fuente de datos el mismo problema que en presidents



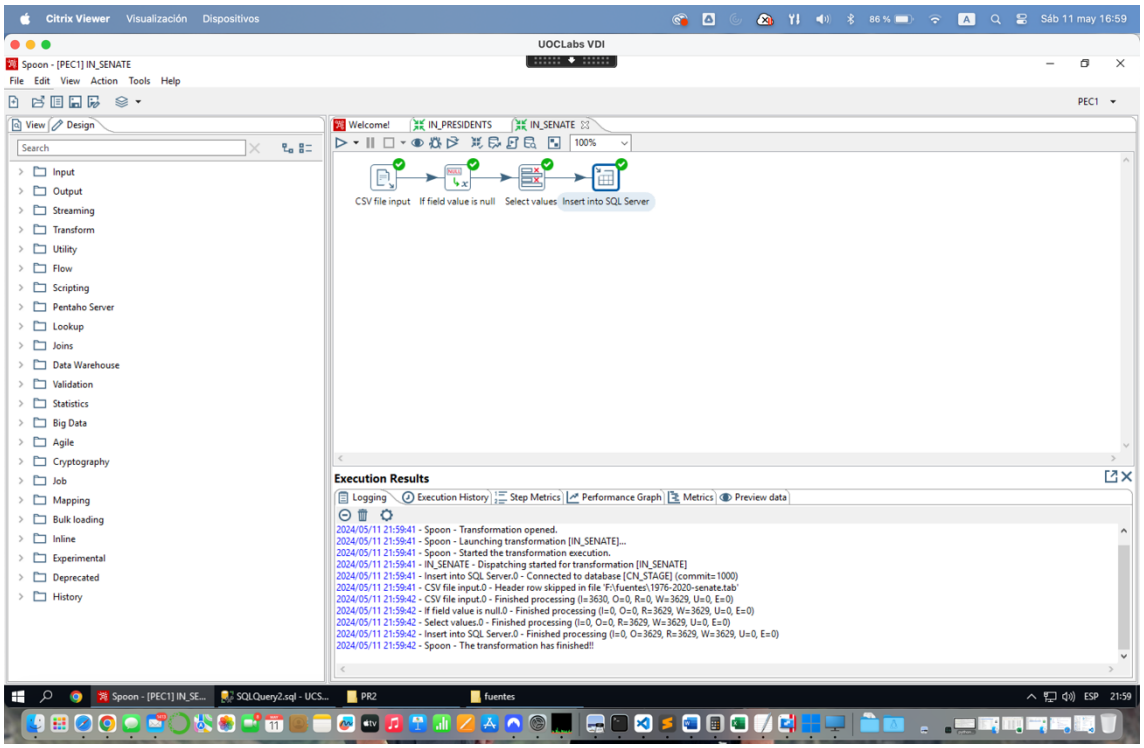
En select values vuelvo a definir el tipo de cada campo



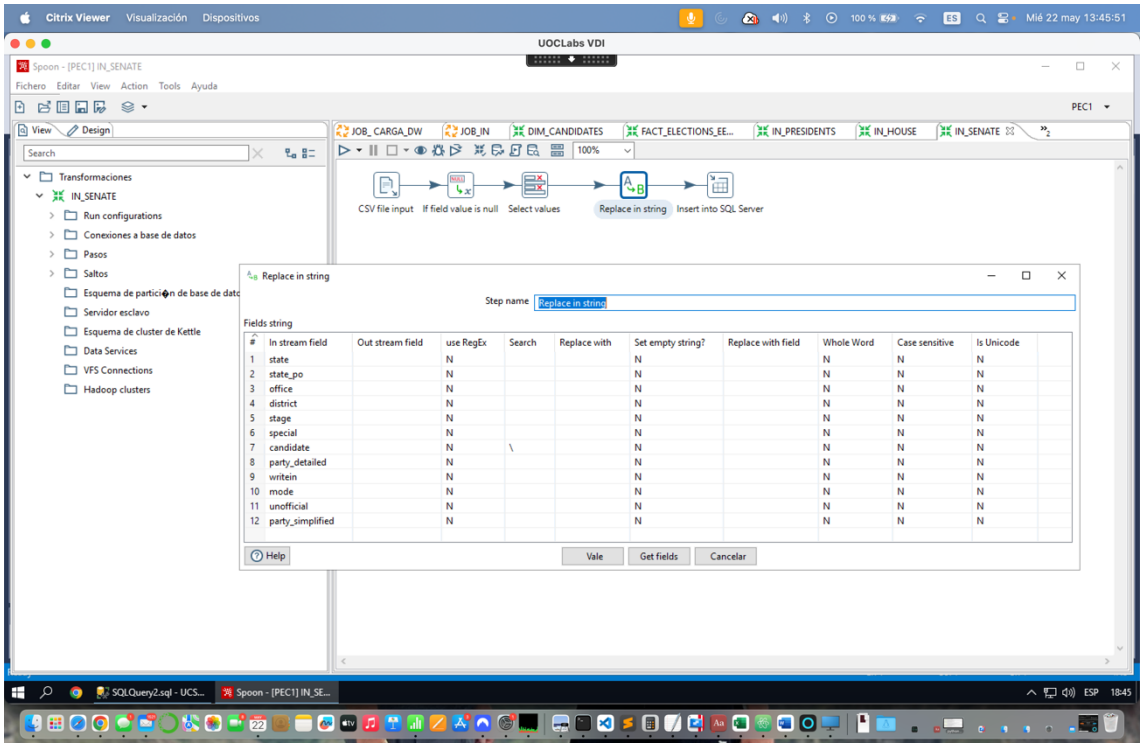
Defino el último paso



Pruebo la transformación y veo que han sido agregados 3629 registros

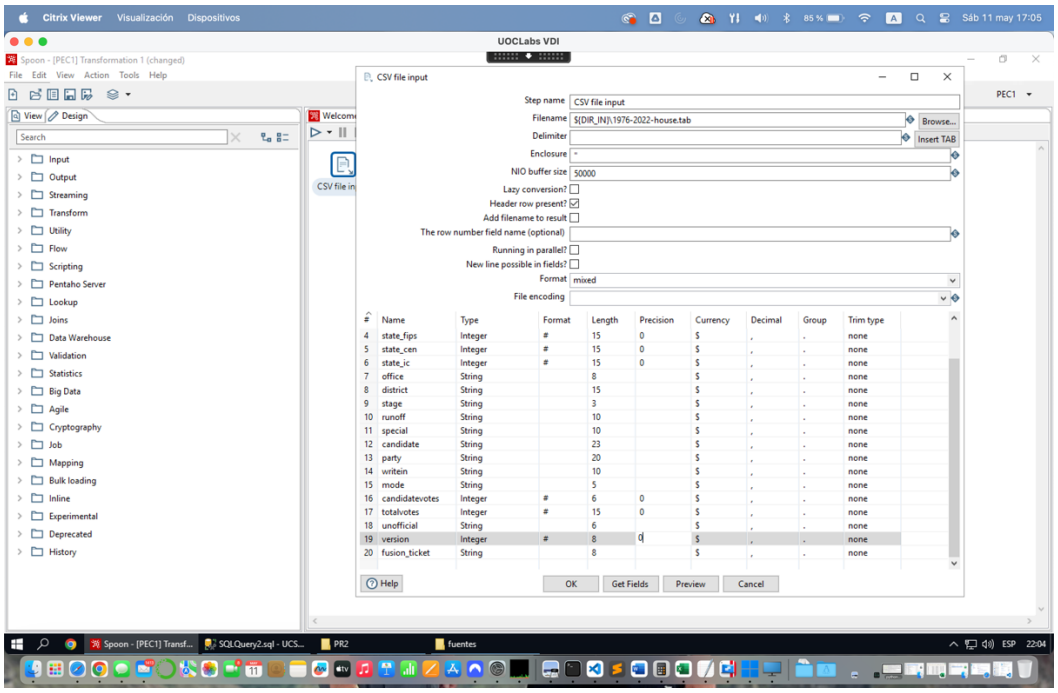


Actualización: Añado remplace in string para quitar la barra invertida

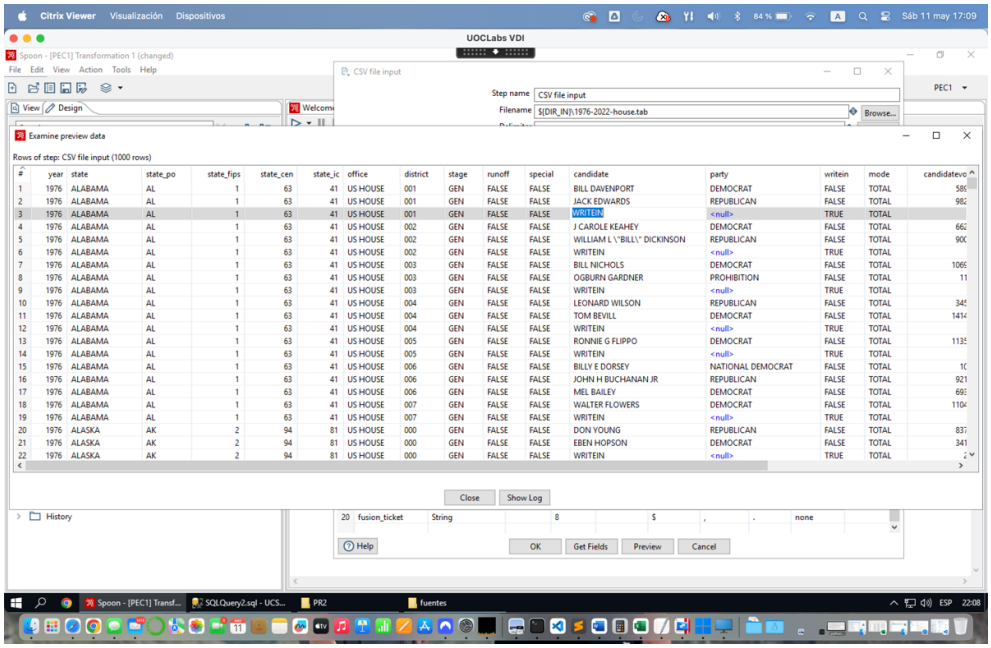


IN_HOUSE

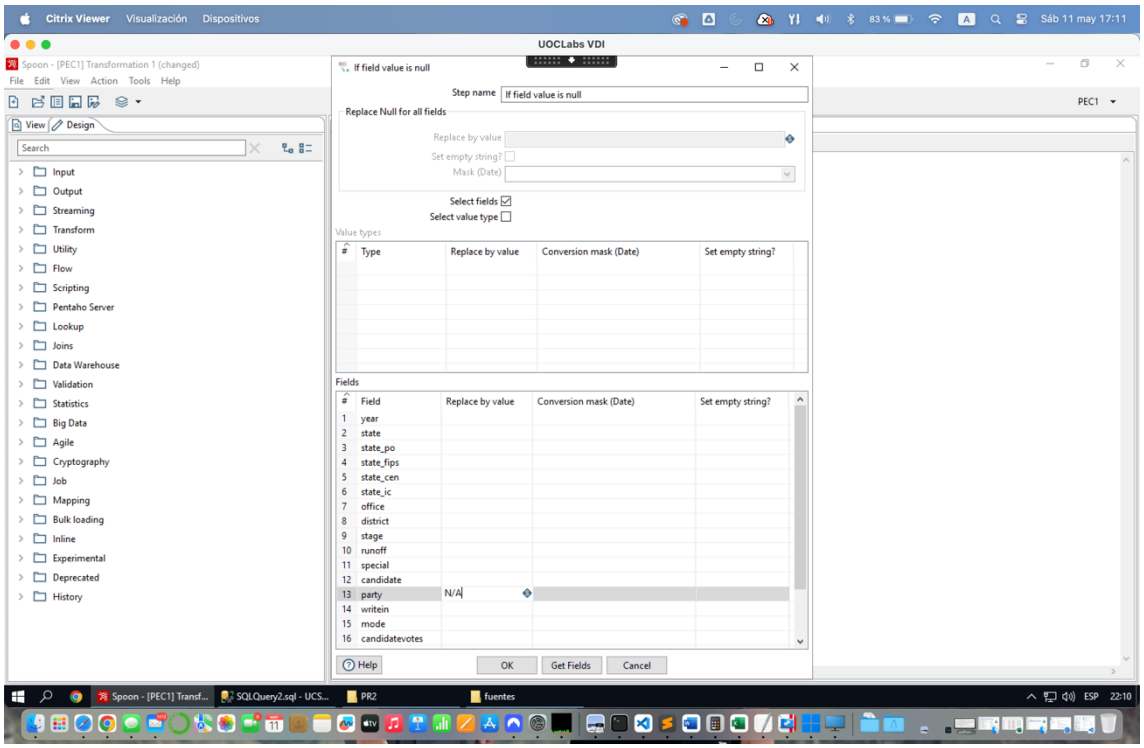
Creo otra transformación, un nuevo paso, CSV File input, y reviso los campos el tipo de los datos, modificando los que no coincidan con el diseño del Data Warehouse.



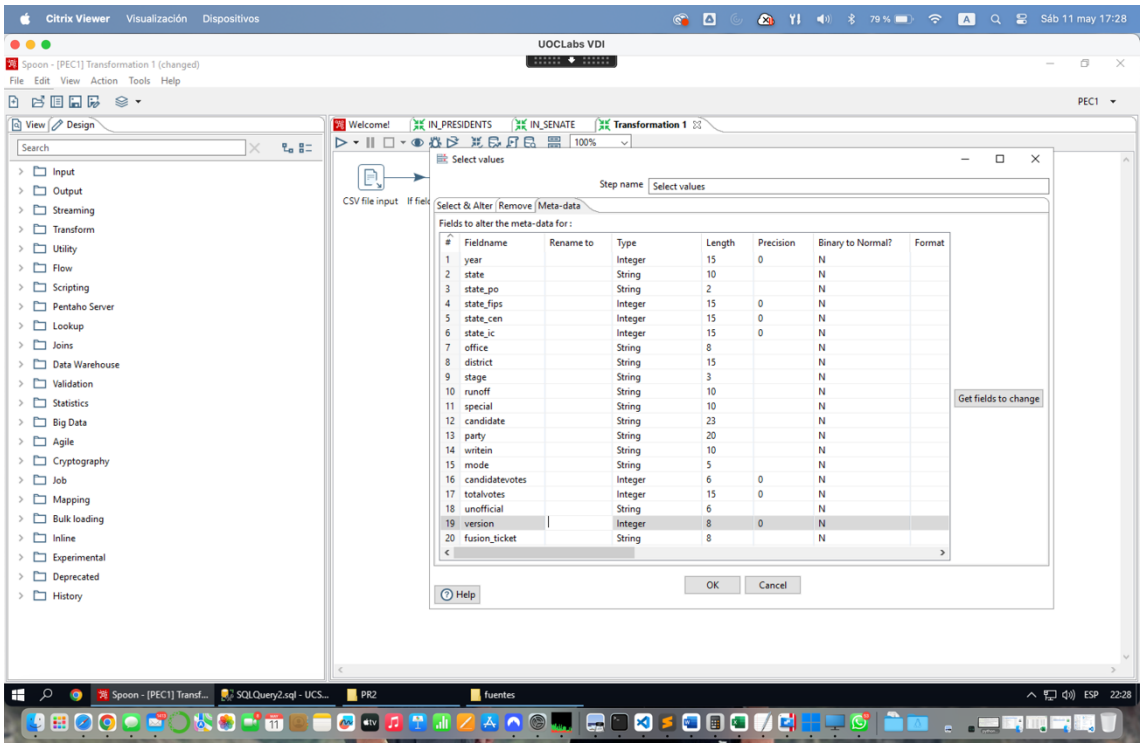
Aquí pasa como en las otras transformaciones, hay partidos nulos, que hay que cambiar, además cuando el partido es nulo no existe candidato, se nombra como write_in que es algo asi como candidato escrito en la papeleta de voto



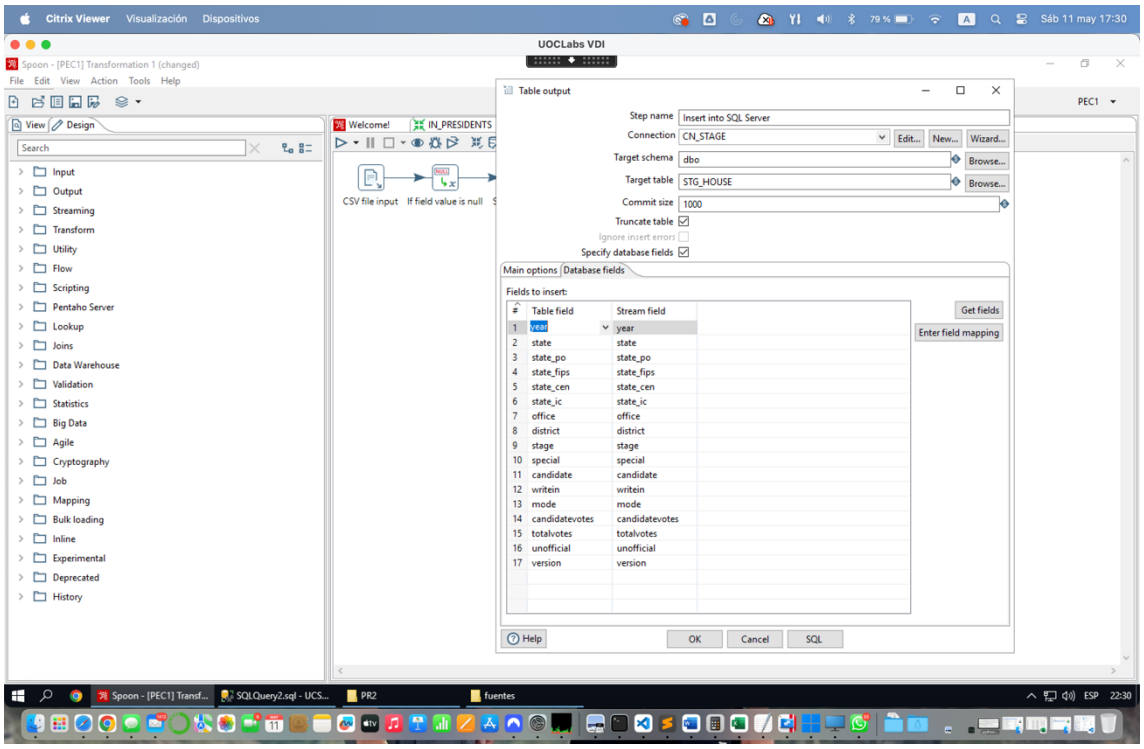
Los registros que tengan el partido político nulo pondremos N/A



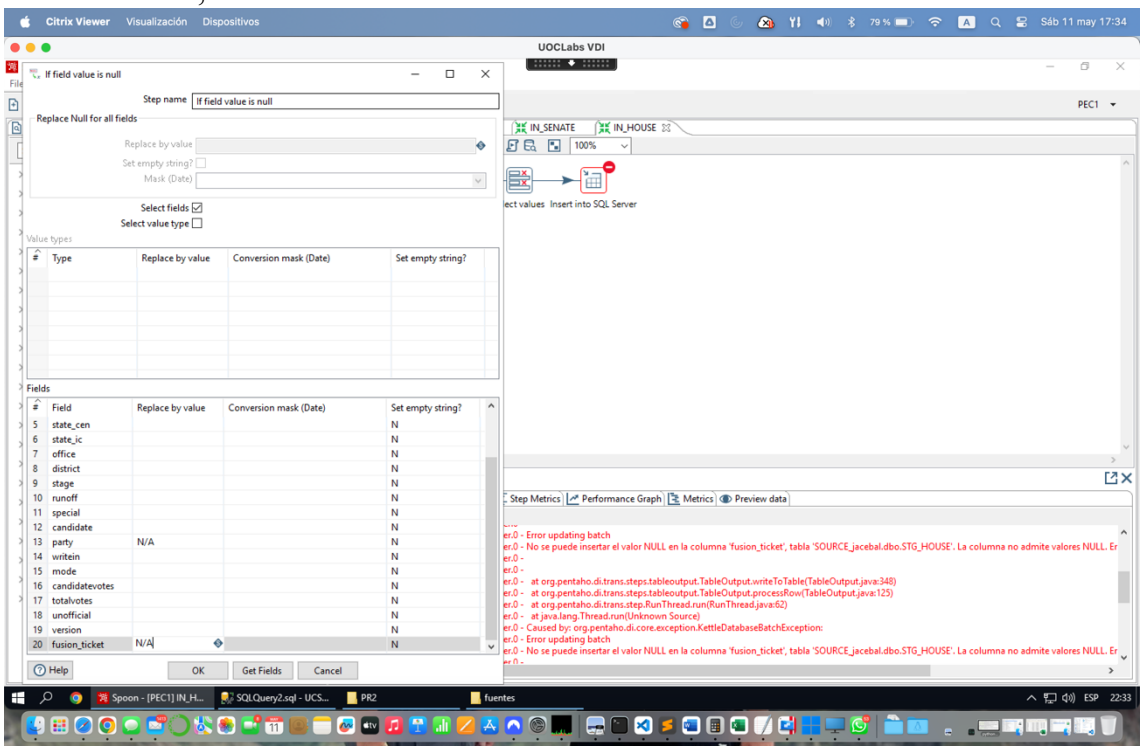
Comprobamos que el metadata es lo que debería ser, completando cada tipo de dato para que cuando adjuntemos a la base de datos todo funcione bien



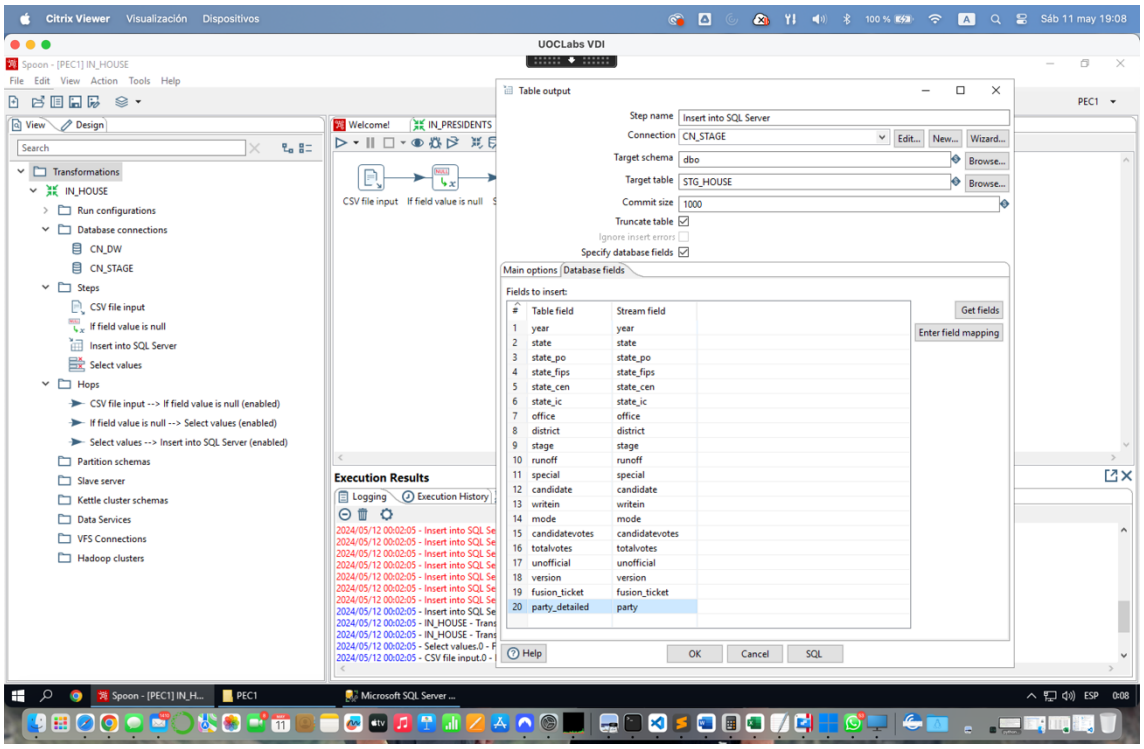
Preparamos para insertar en SQL Server, antes de ello comprobamos que el origen y destino está bien definido (de cada tabla)



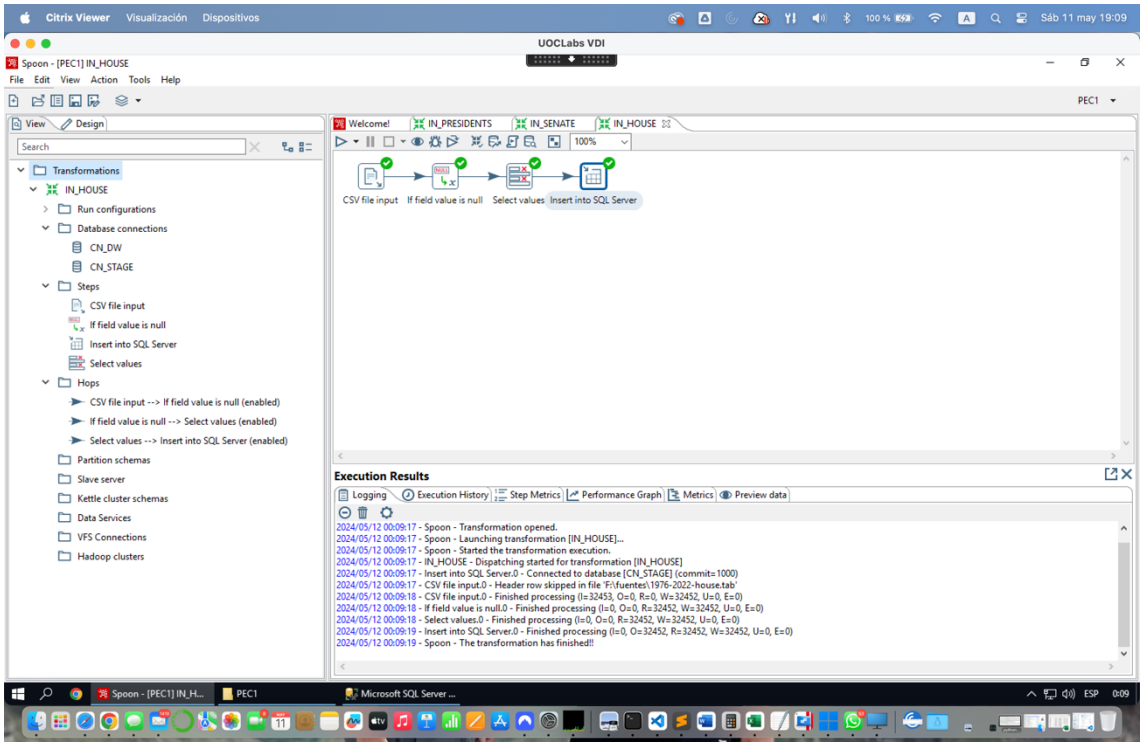
Ha dado error, entonces



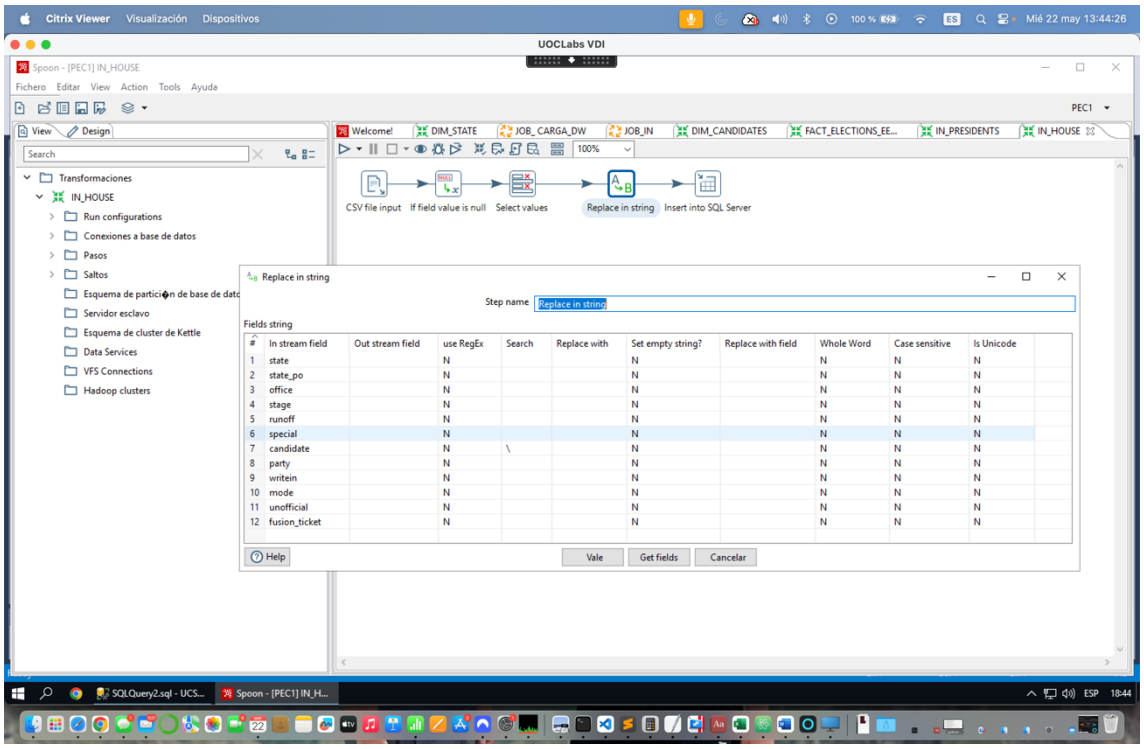
Ademas party corresponde a party_detailled



Resultado de IN_HOUSE, 32452 registros introducidos.

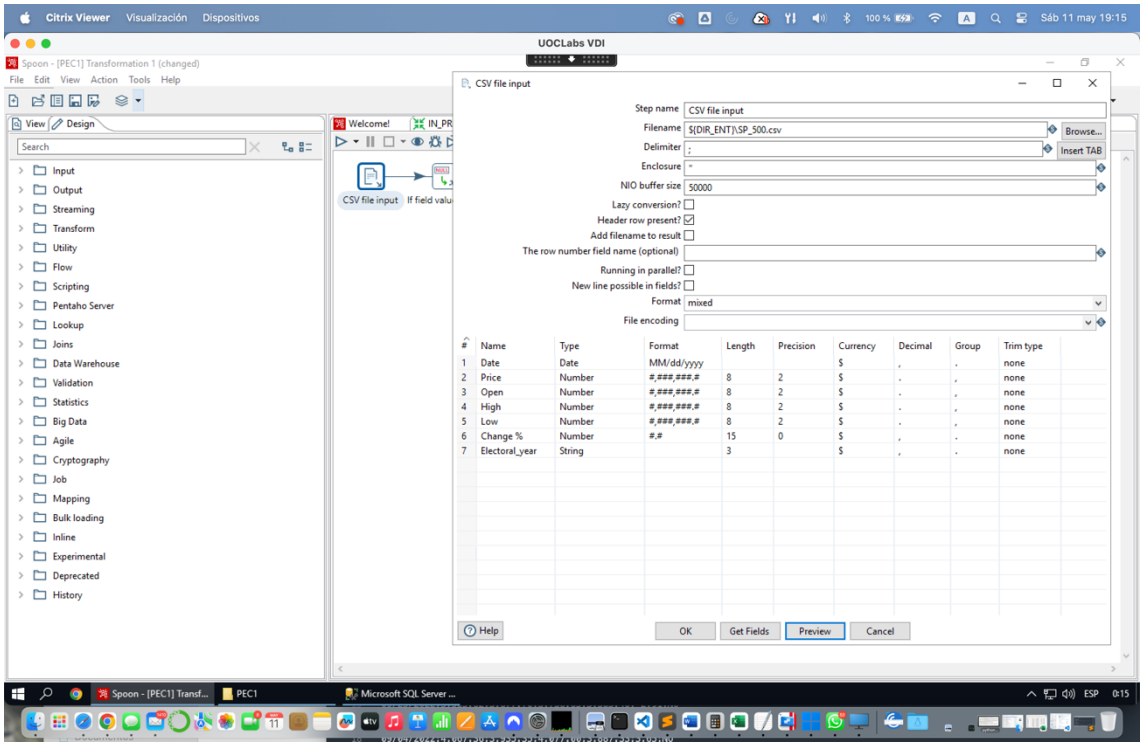


Actualización: Elimino la barra invertida de los candidatos para que queden solamente las comillas dobles en el apodo-nombre

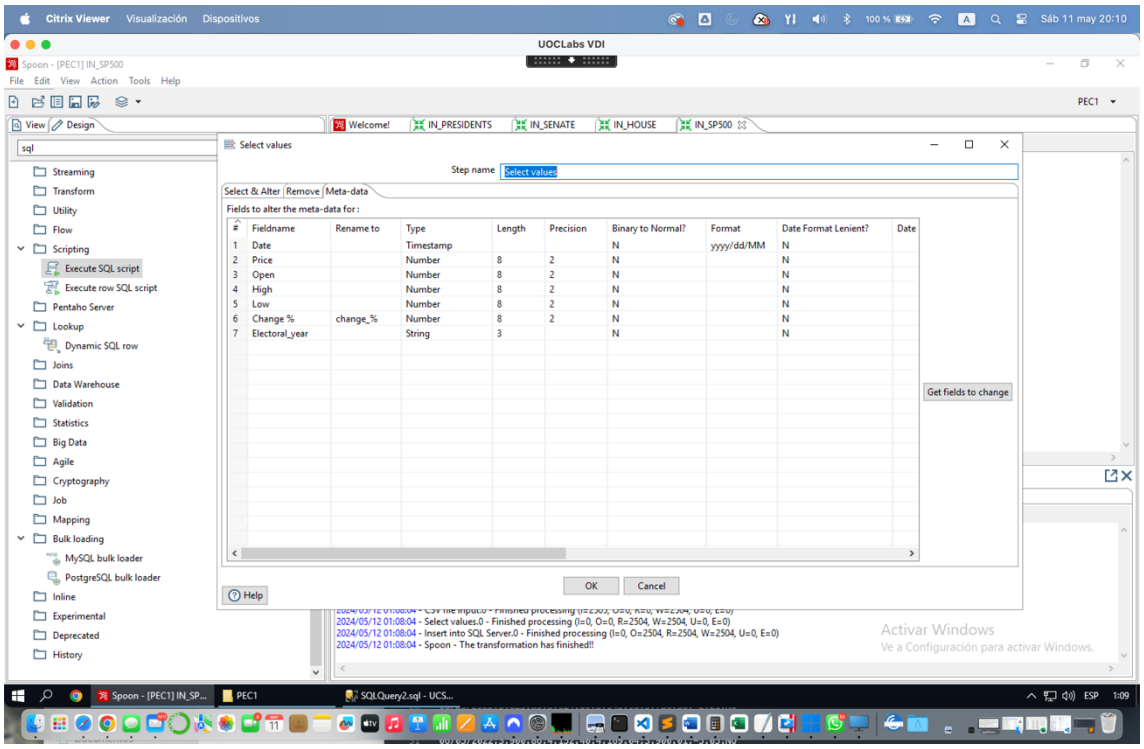


IN_SP500

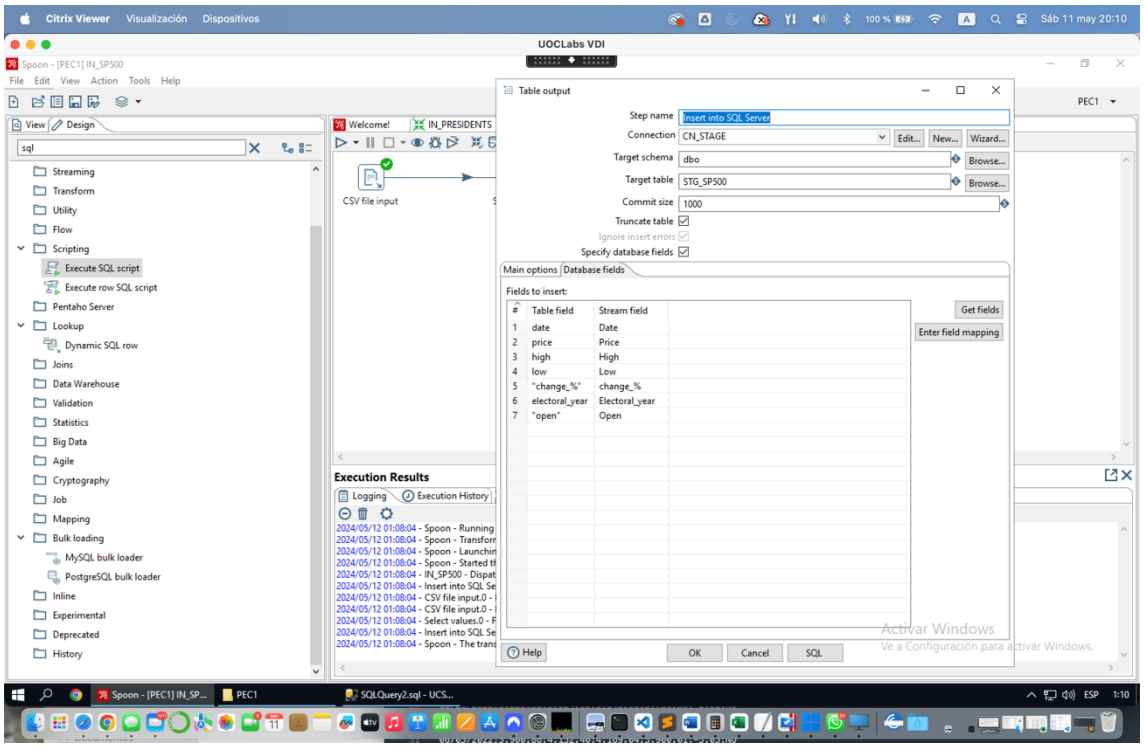
Ahora agrego el paso de CSV input en la transformación IN_SP500



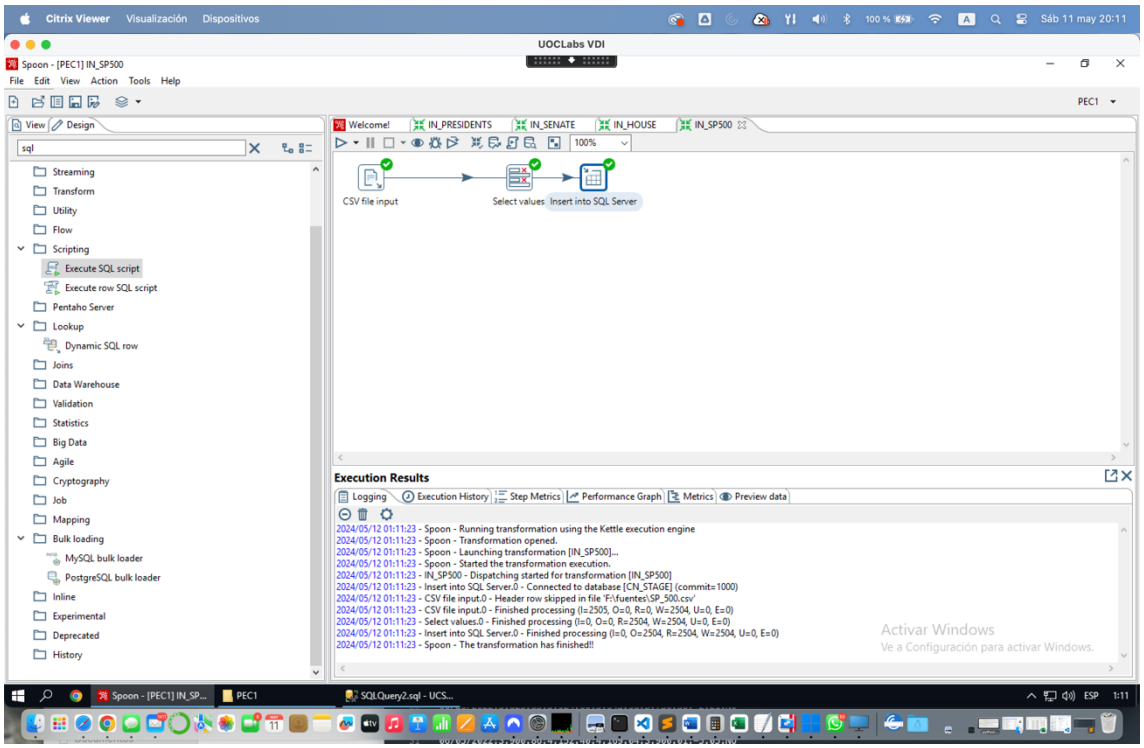
Select values, pongo formato de fecha y compruebo que los campos están bien definidos, cambio el nombre de la columna Change % por change_%



Uso comillas dobles para definir open y change_% para evitar problemas con las palabras y símbolos reservados

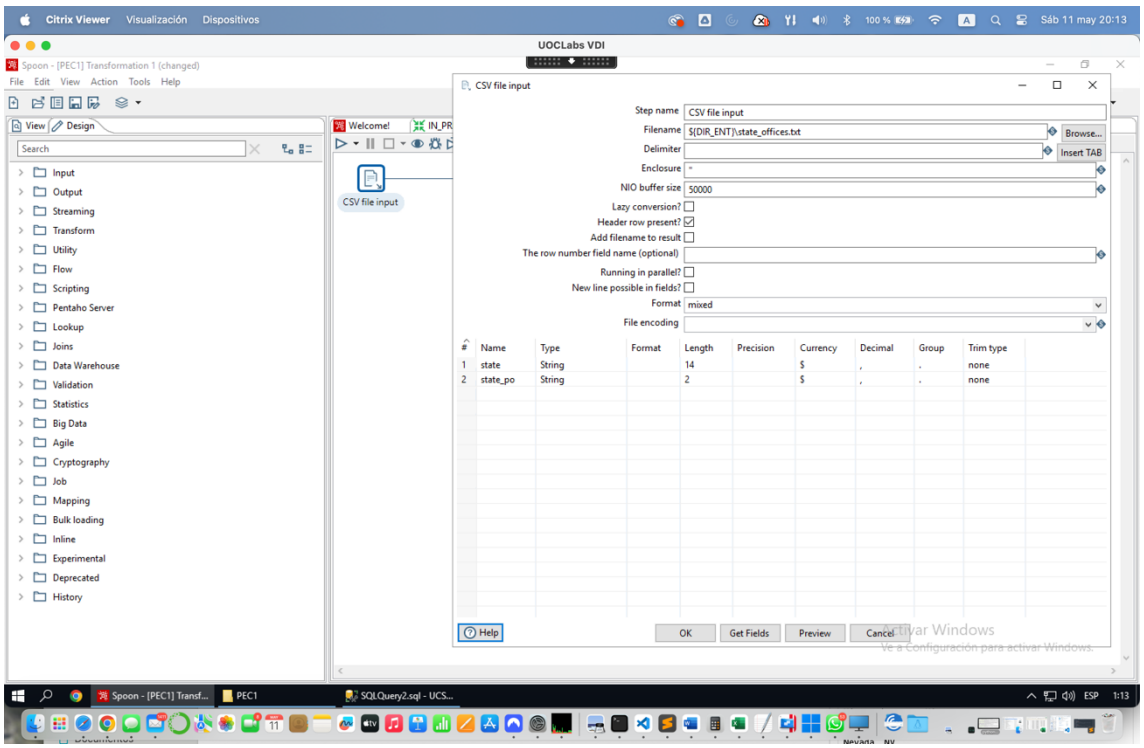


Resultado

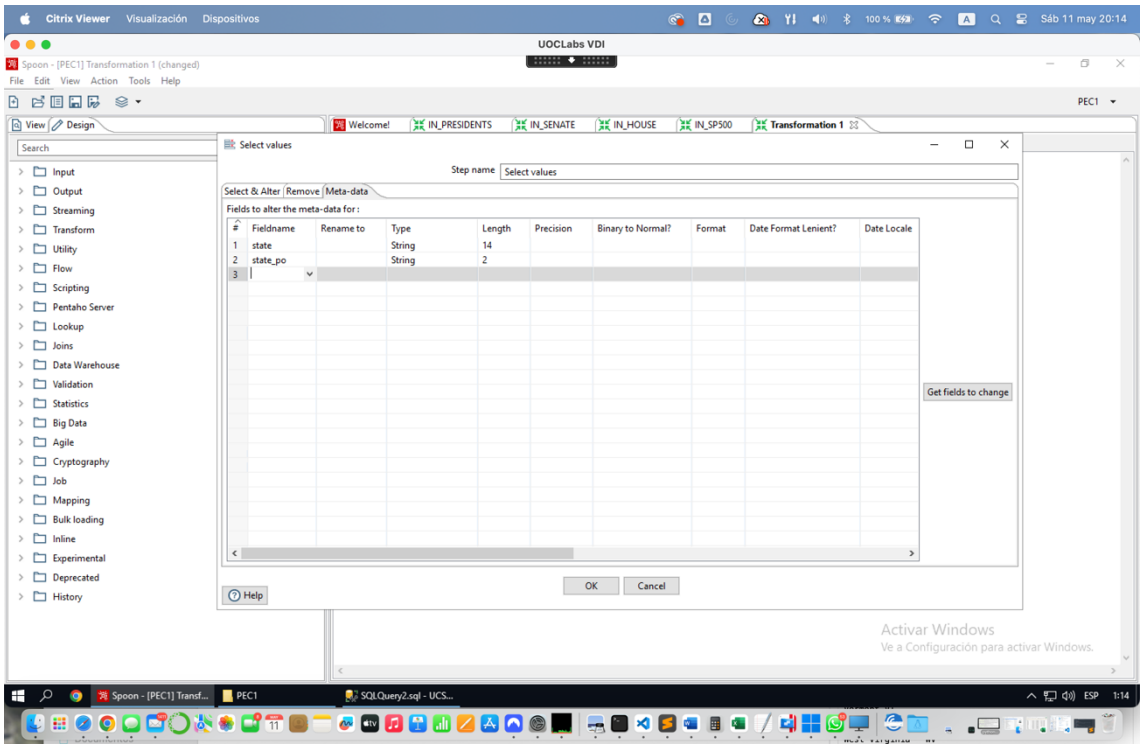


IN_STATE

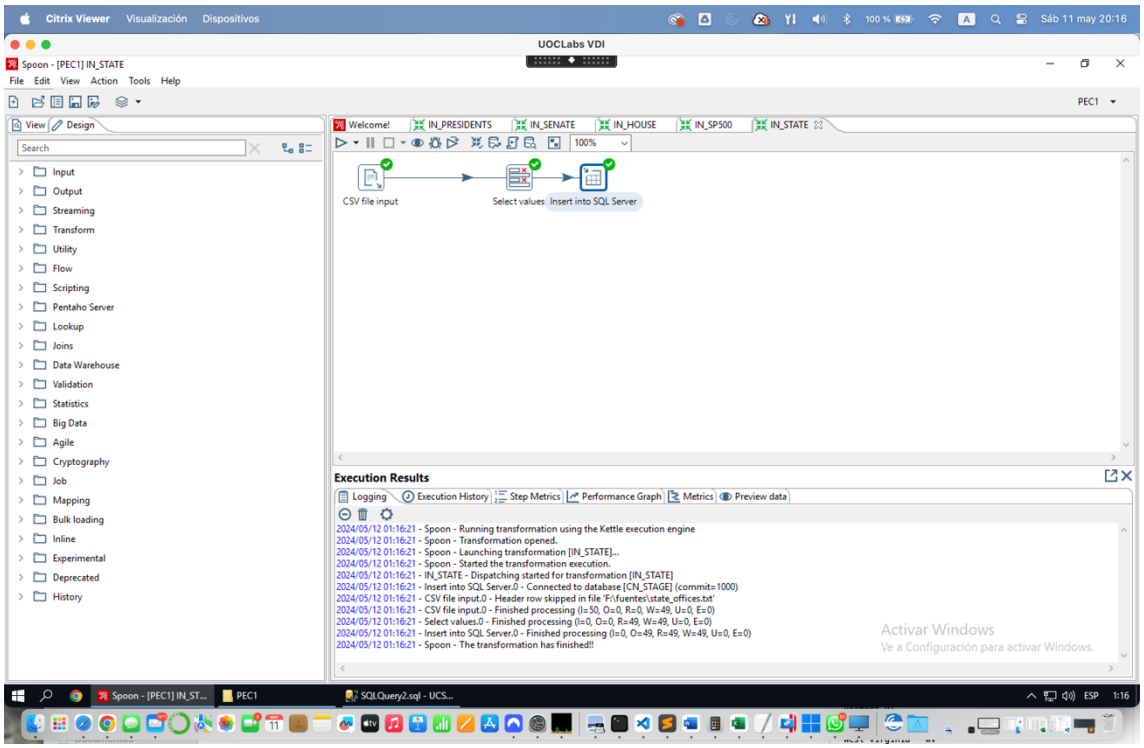
CSV file input



Select values



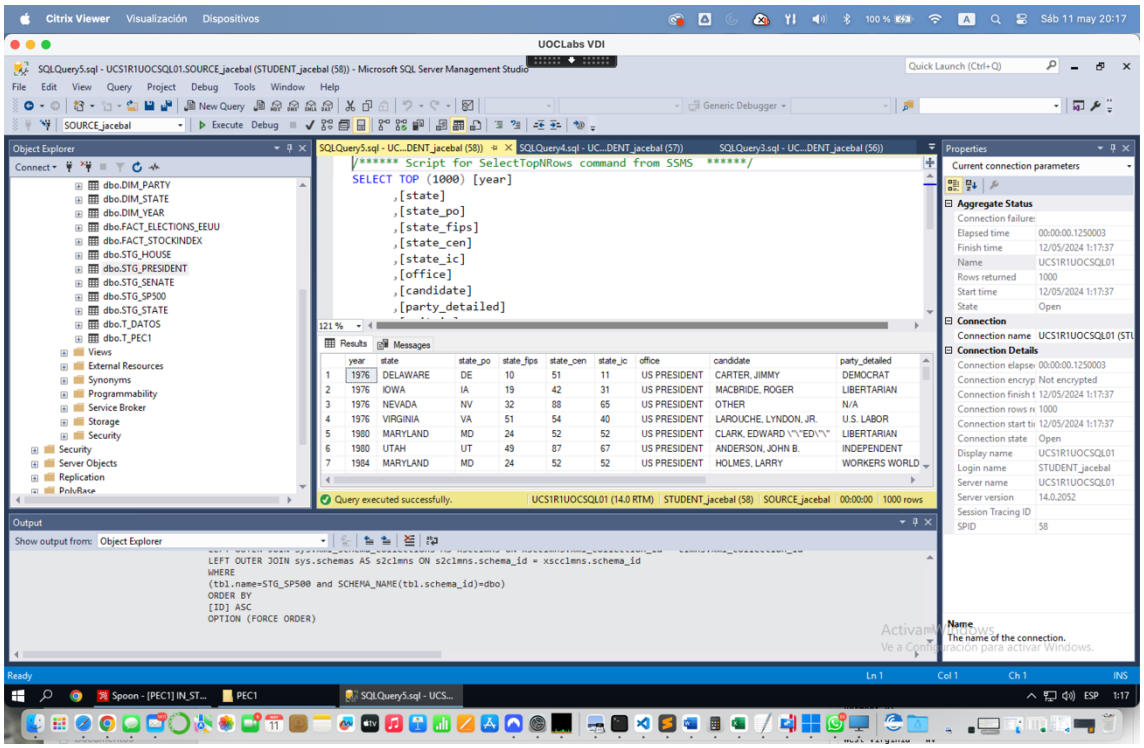
Resultado



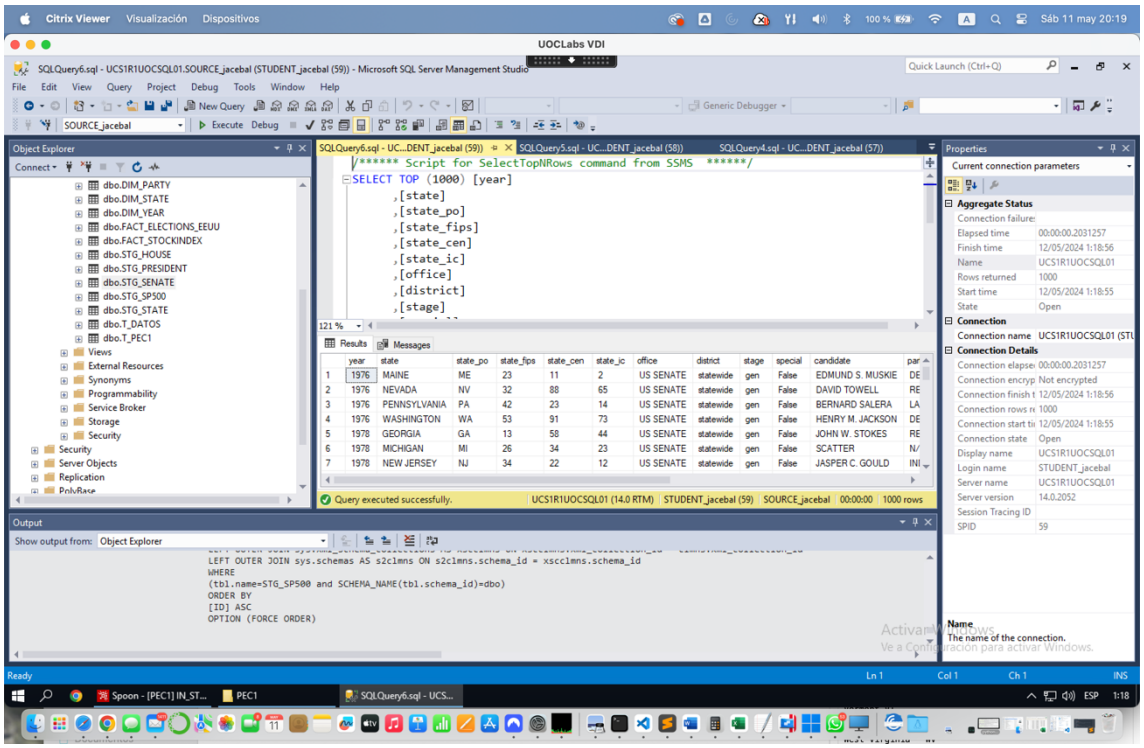
COMPROBACIONES MSSQL BLOQUE STG

Compruebo doblemente que todas las transformaciones han actualizado en SGBD

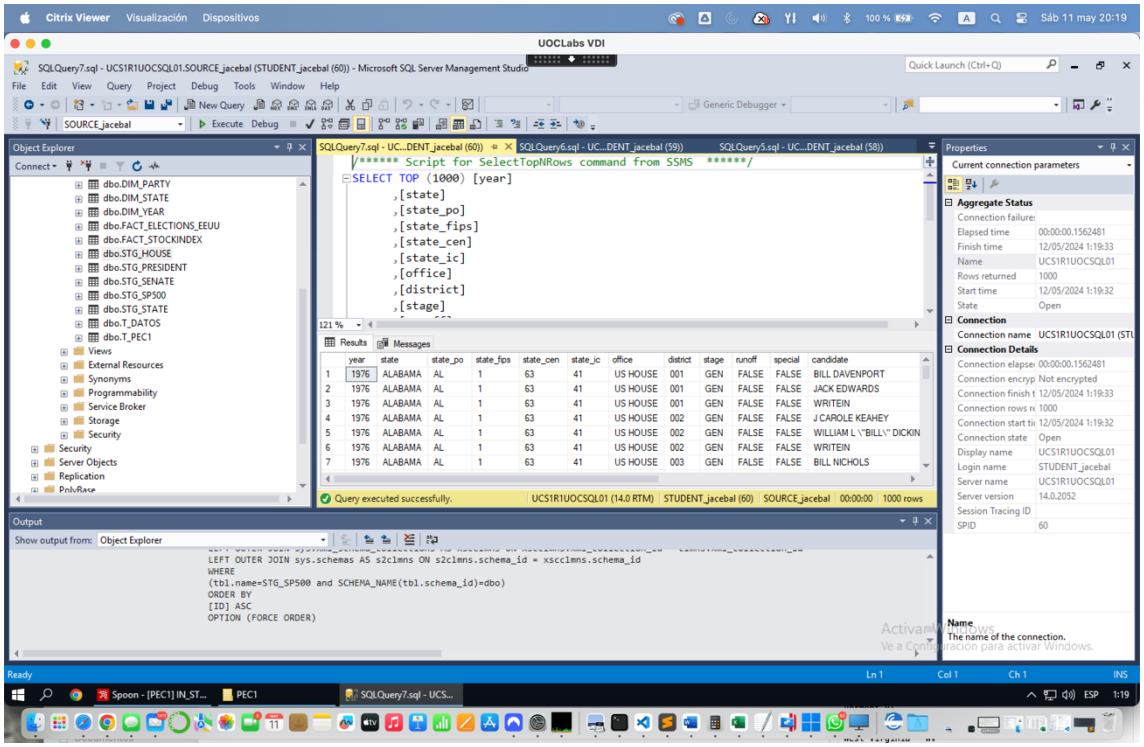
STG_PRESIDENT



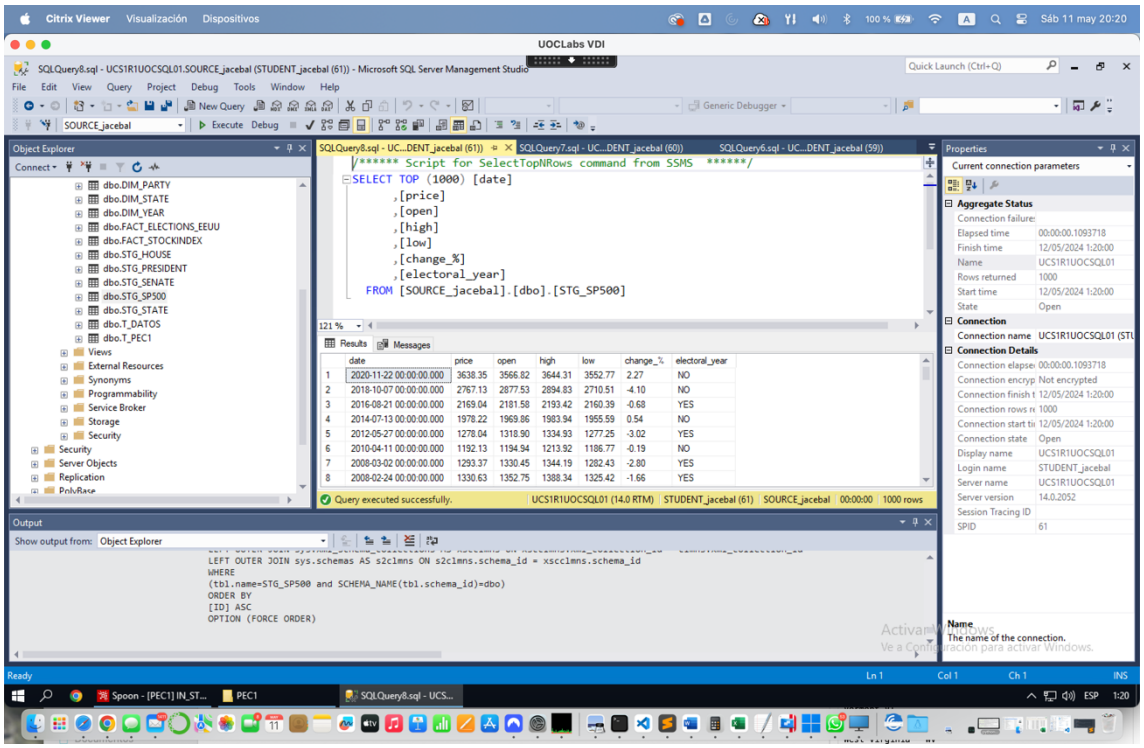
STG_SENATE



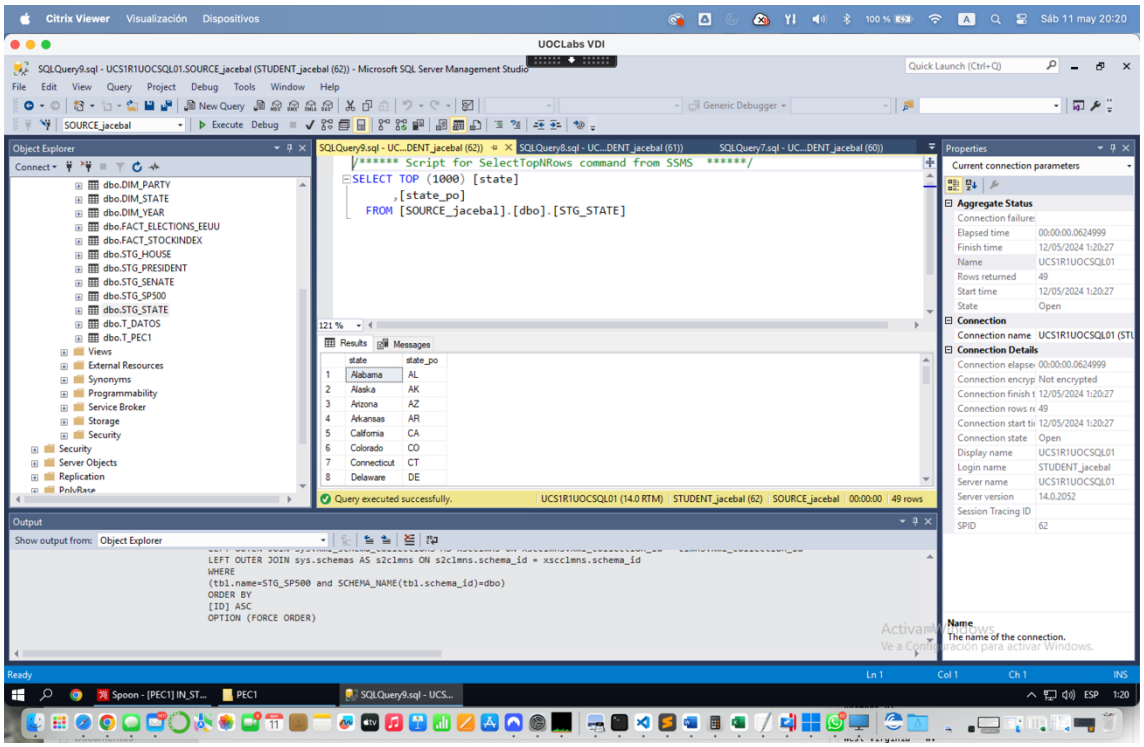
STG_HOUSE



STG_SP500



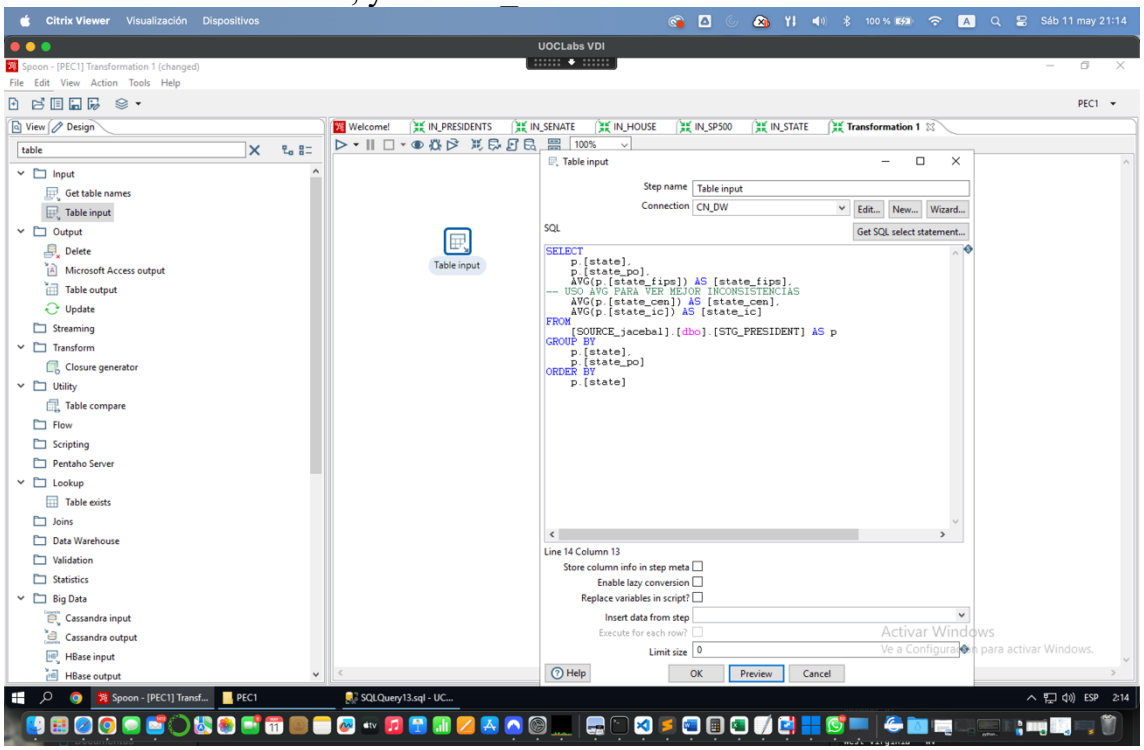
STG_STATE


UOCLabs VDI
SQLQuery9.sql - UCS1R1UOCSQL01.SOURCE_jacebal (STUDENT_jacebal (62)) - Microsoft SQL Server Management Studio
/****** Script for SelectTopNRows command from SSMS ******/
SELECT TOP (1000) [state]
,[state_po]
FROM [SOURCE_jacebal].[dbo].[STG_STATE]
Query executed successfully.
Alabama AL
Alaska AK
Arizona AZ
Arkansas AR
California CA
Colorado CO
Connecticut CT
Delaware DE

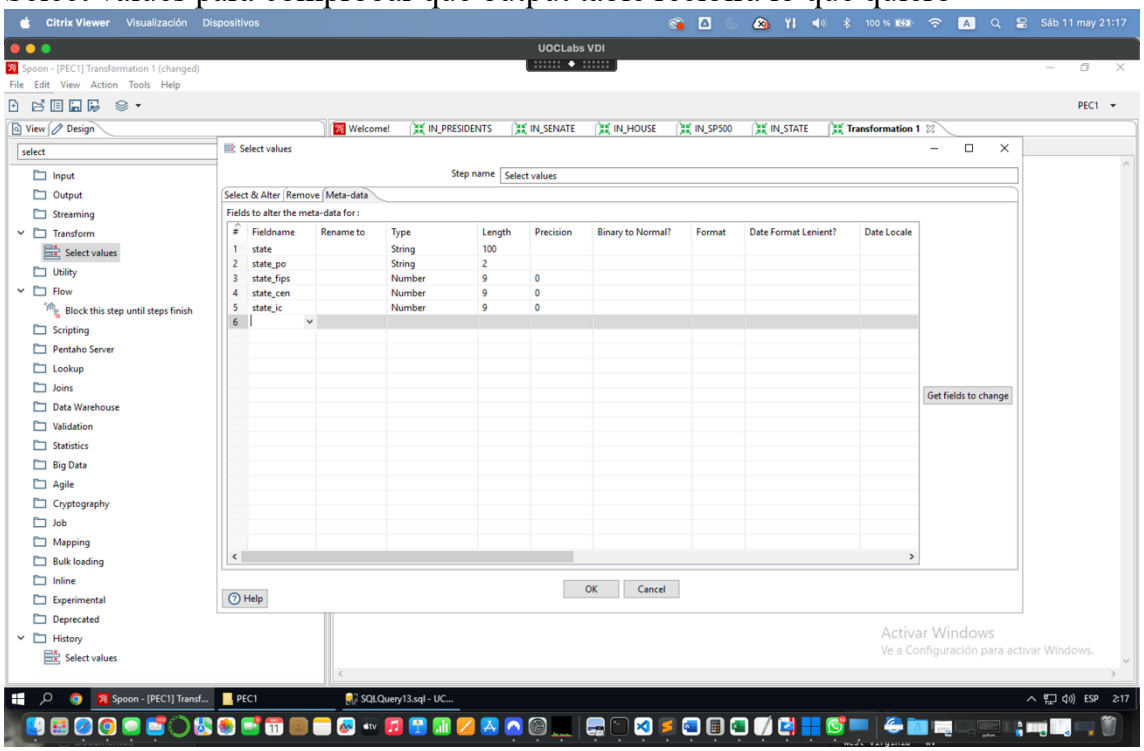
BLOQUE TR_DIM

DIM_STATE

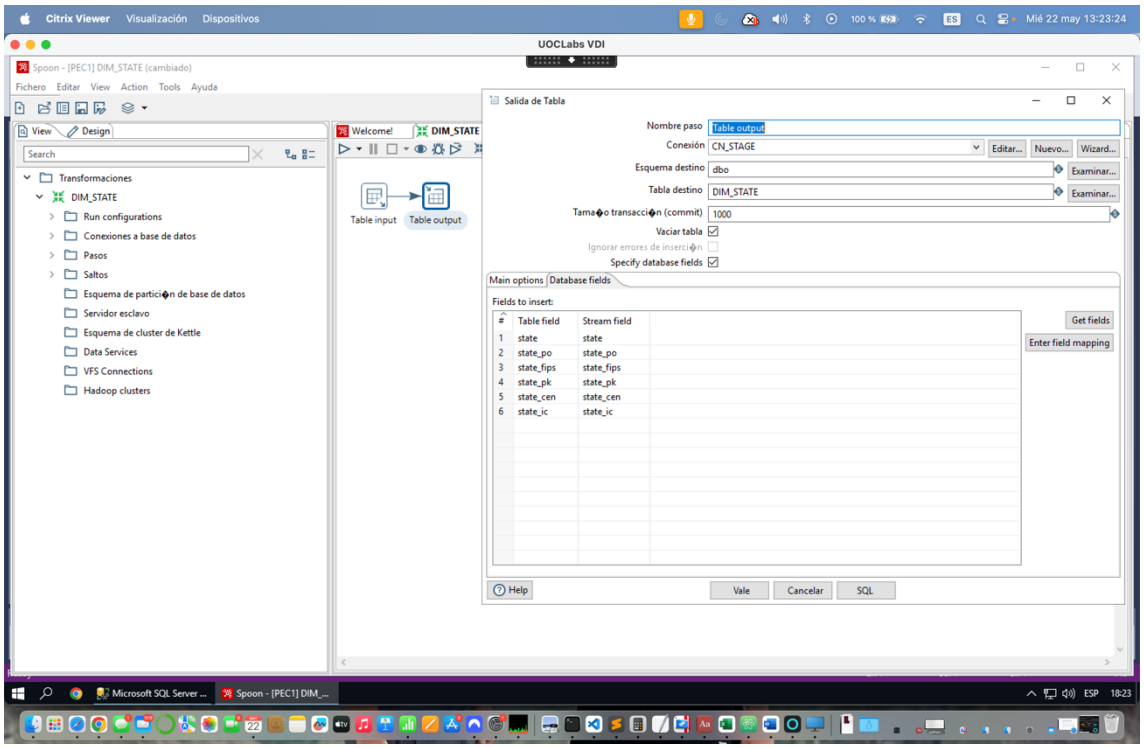
Creo el paso table input donde con un script selecciono y agrupo. Ordeno y así se queda ya ordenado. He elegido tomar los datos de STG_PRESIDENT ya que estaban todos los valores, y en STG_STATE faltaban 2



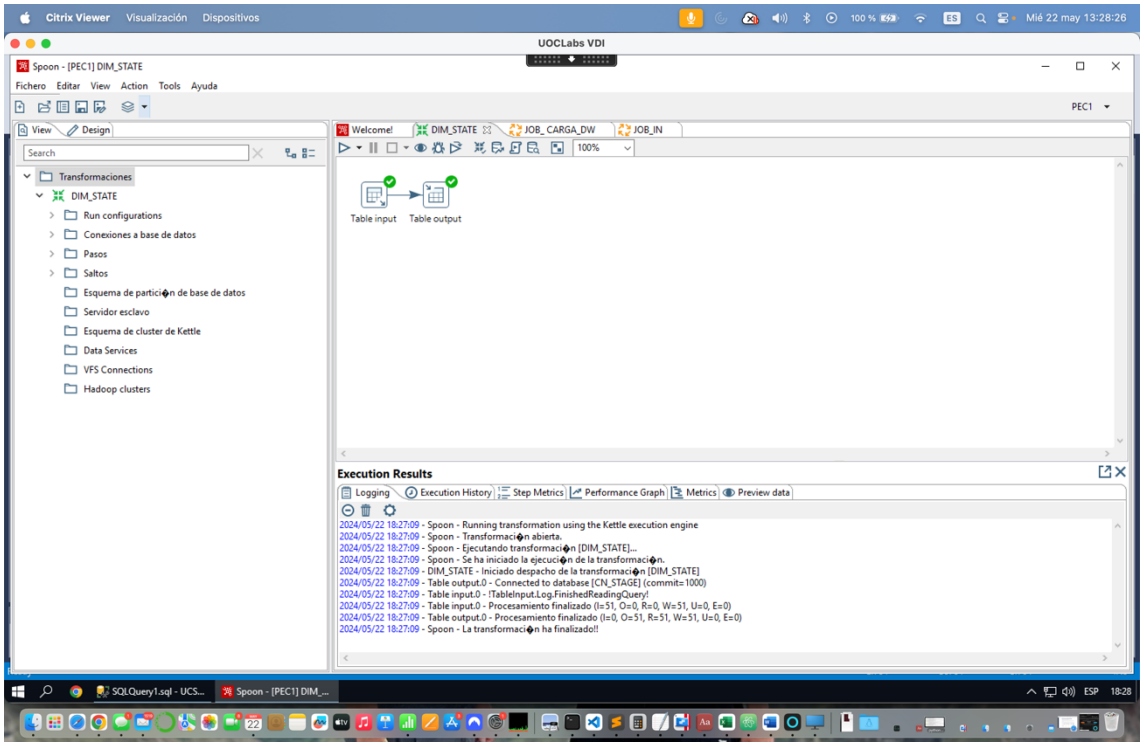
Select values para comprobar que output table recibirá lo que quiero



Creo el paso salida tabla

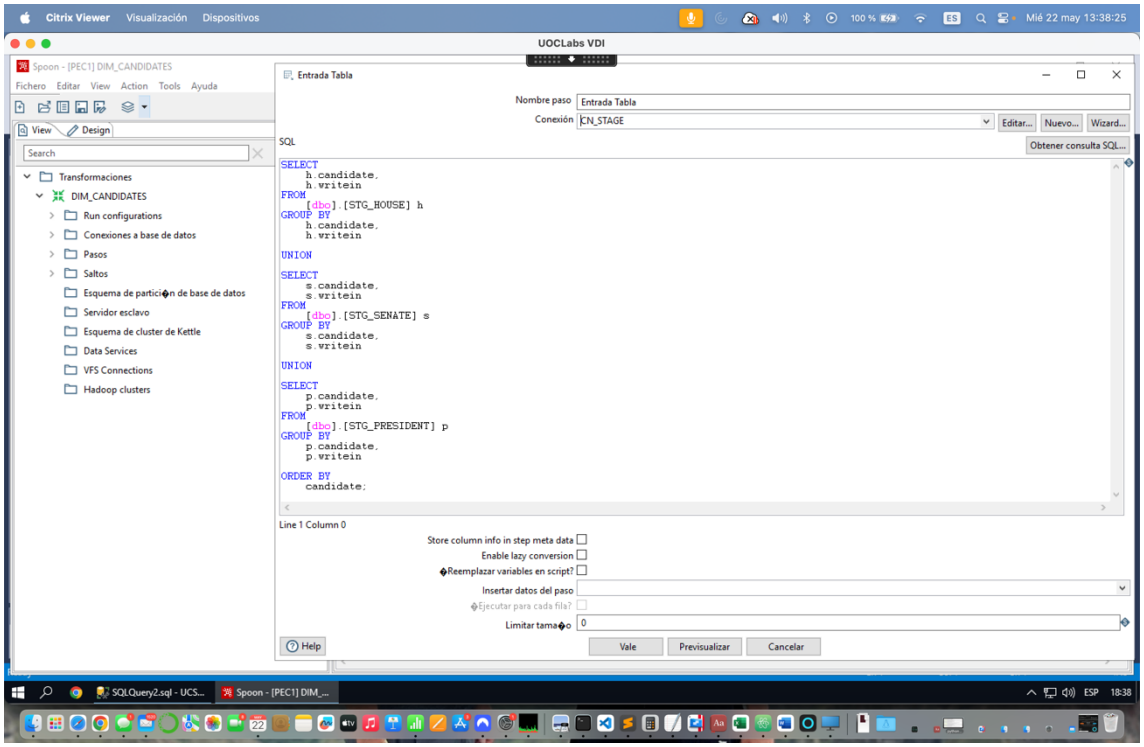


Pruebo:

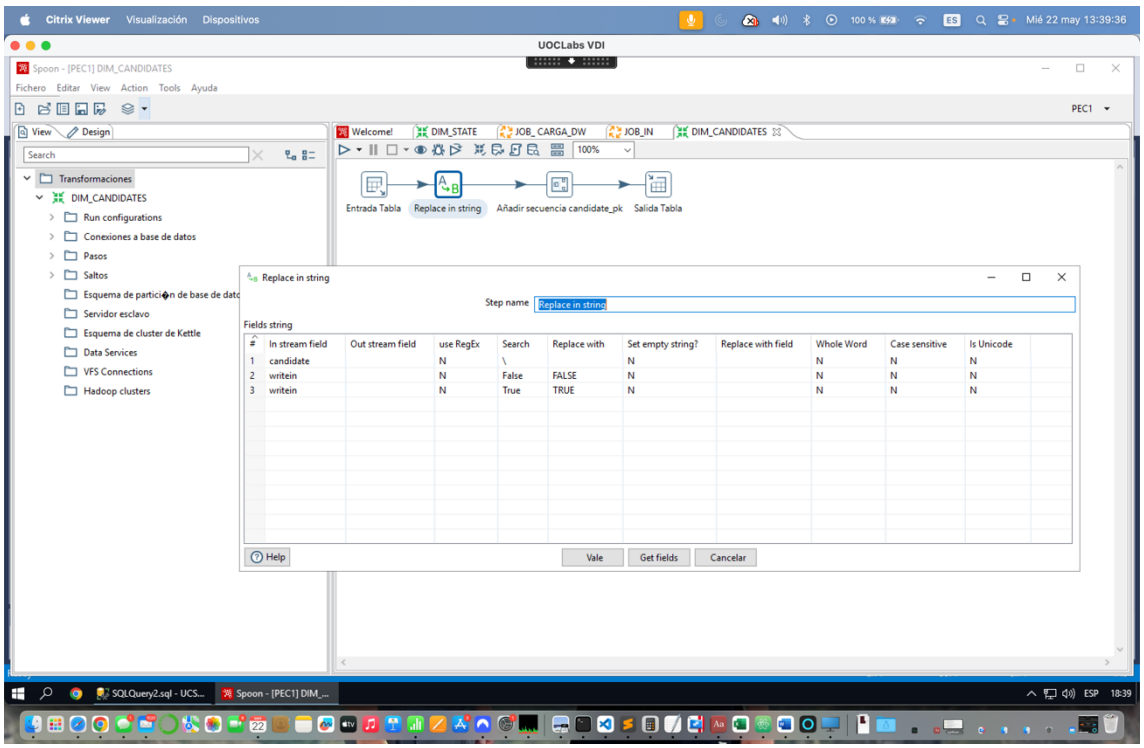


DIM_CANDIDATE

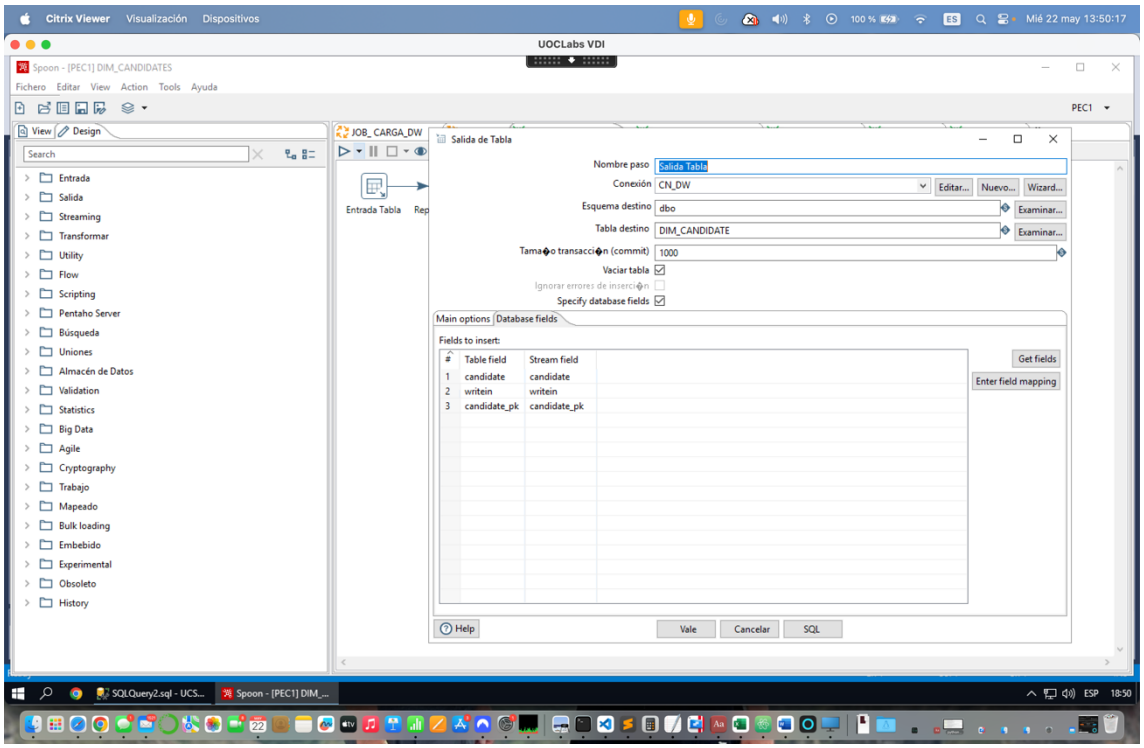
Ahora creo la consulta para tomar los datos. He tomado los candidatos y write in de las 3 tablas STG de resultados de elecciones.



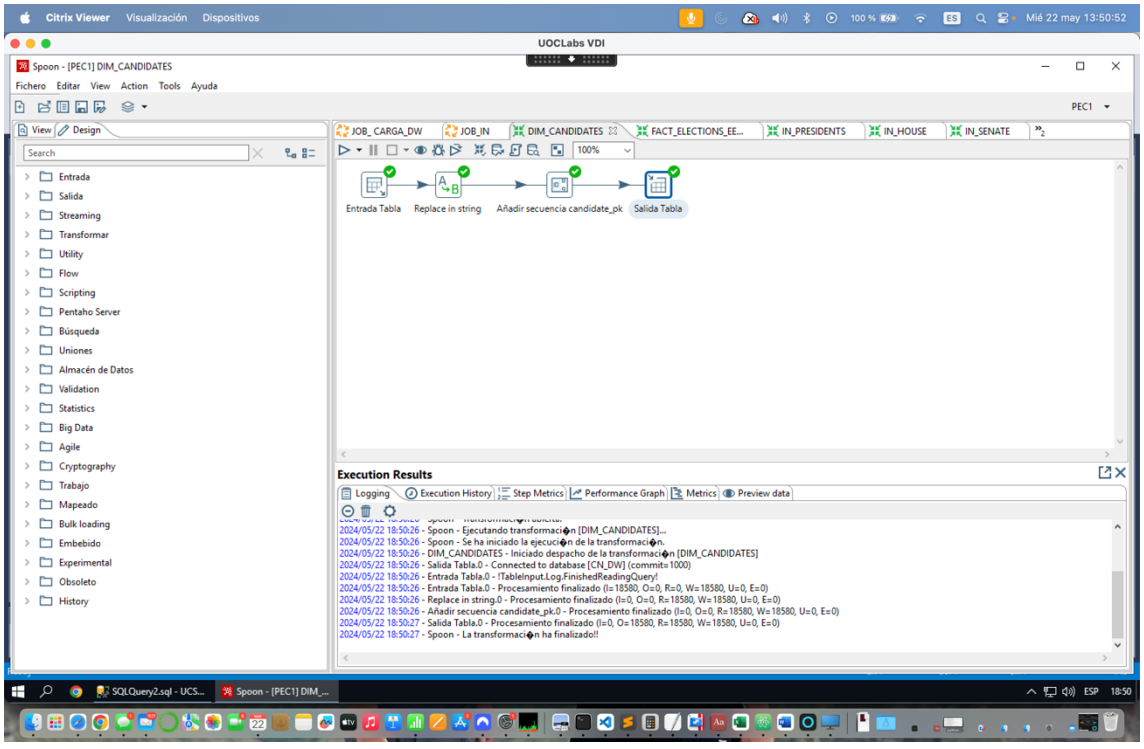
Remplazo False por FALSE y True por TRUE para evitar que haya registros repetidos, además aunque ya lo he hecho en IN_xxx pongo aquí el remplace de la barra invertida también



Por último, hago la salida a tabla

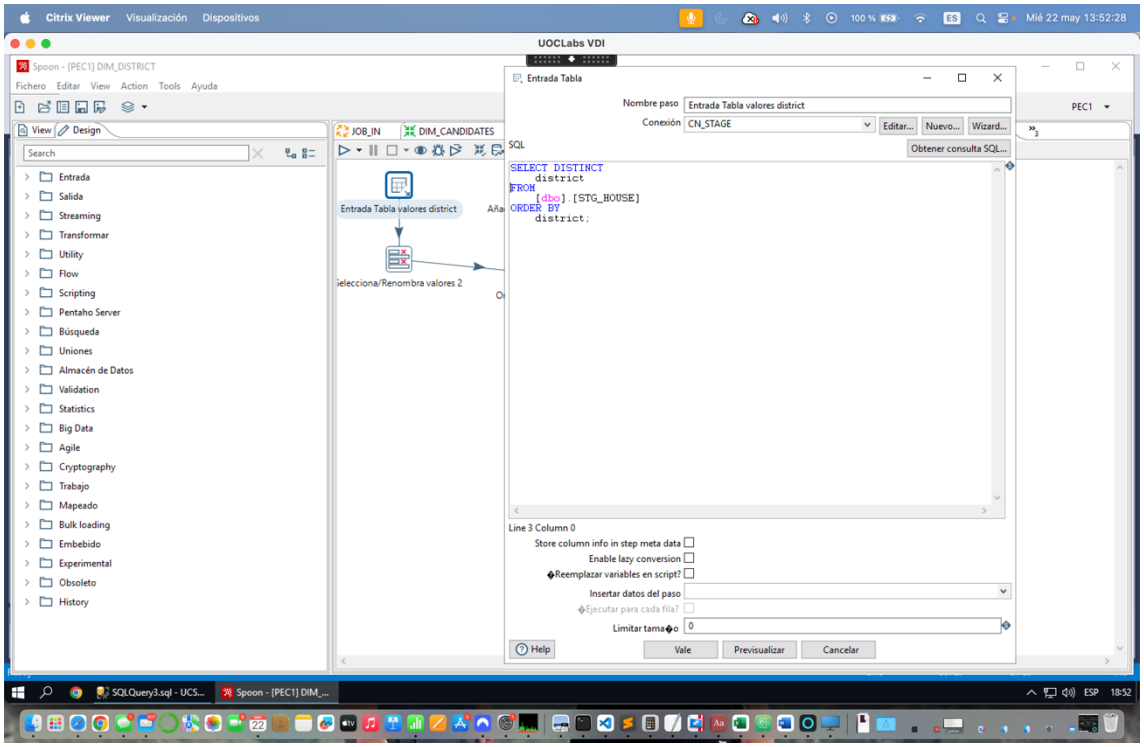


Compruebo

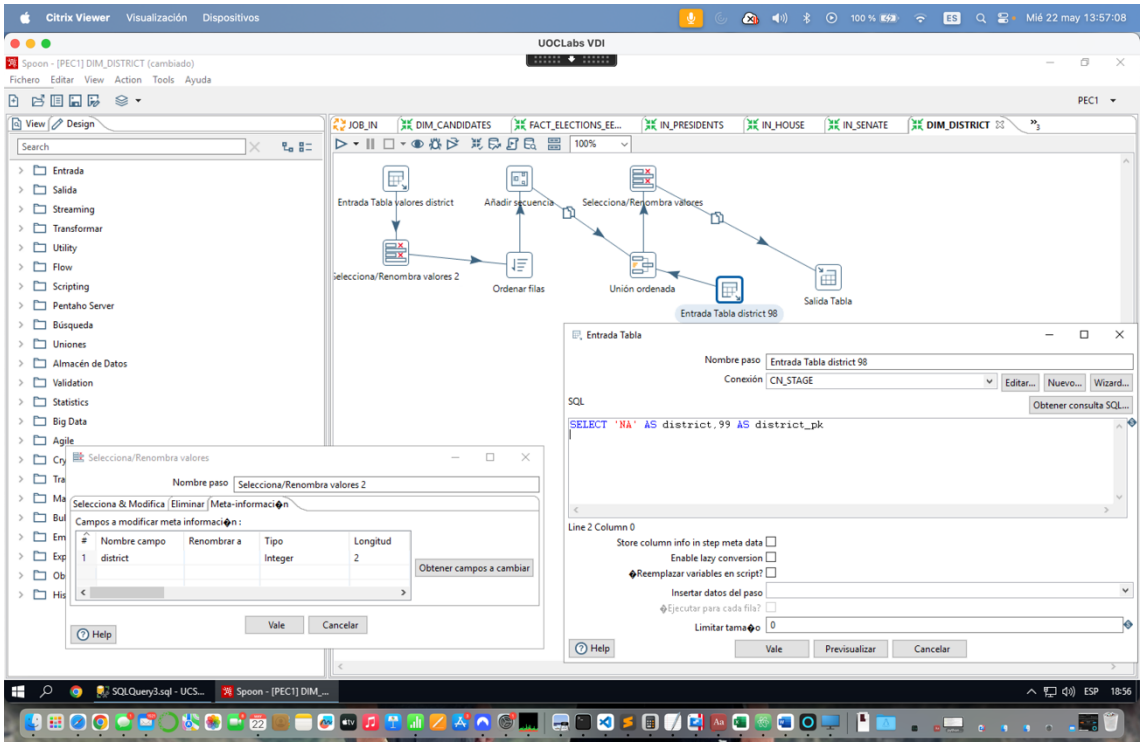


DIM_DISTRICT

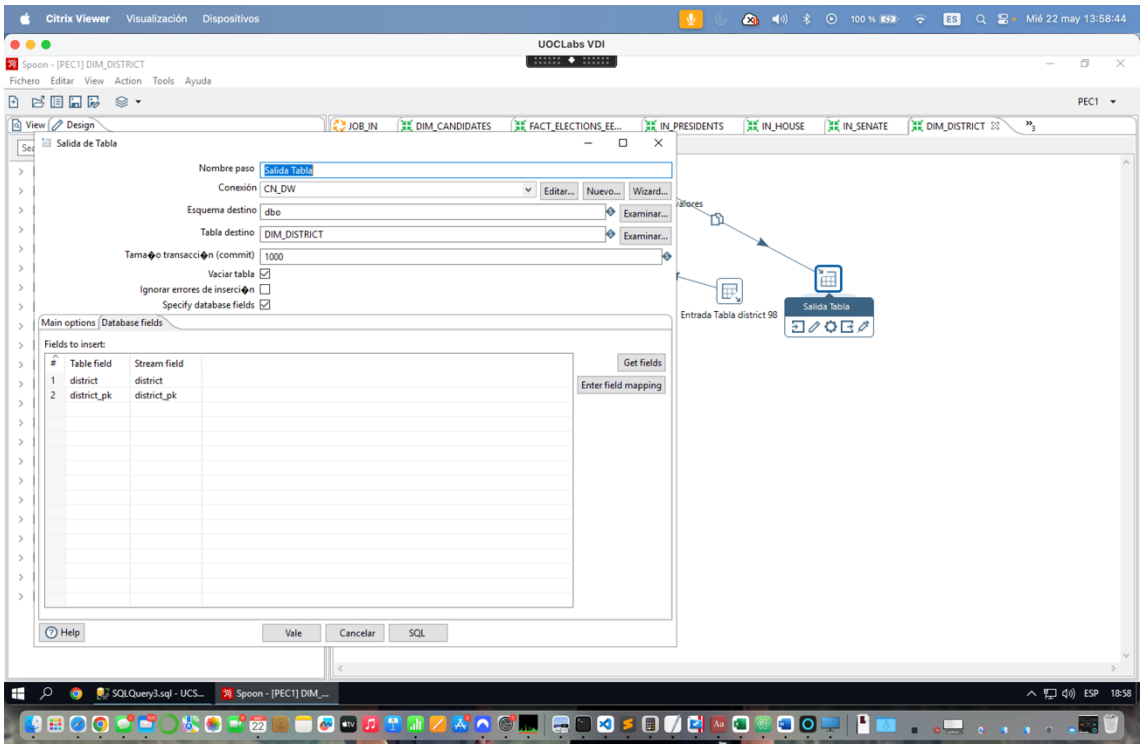
Hago la entrada de valores



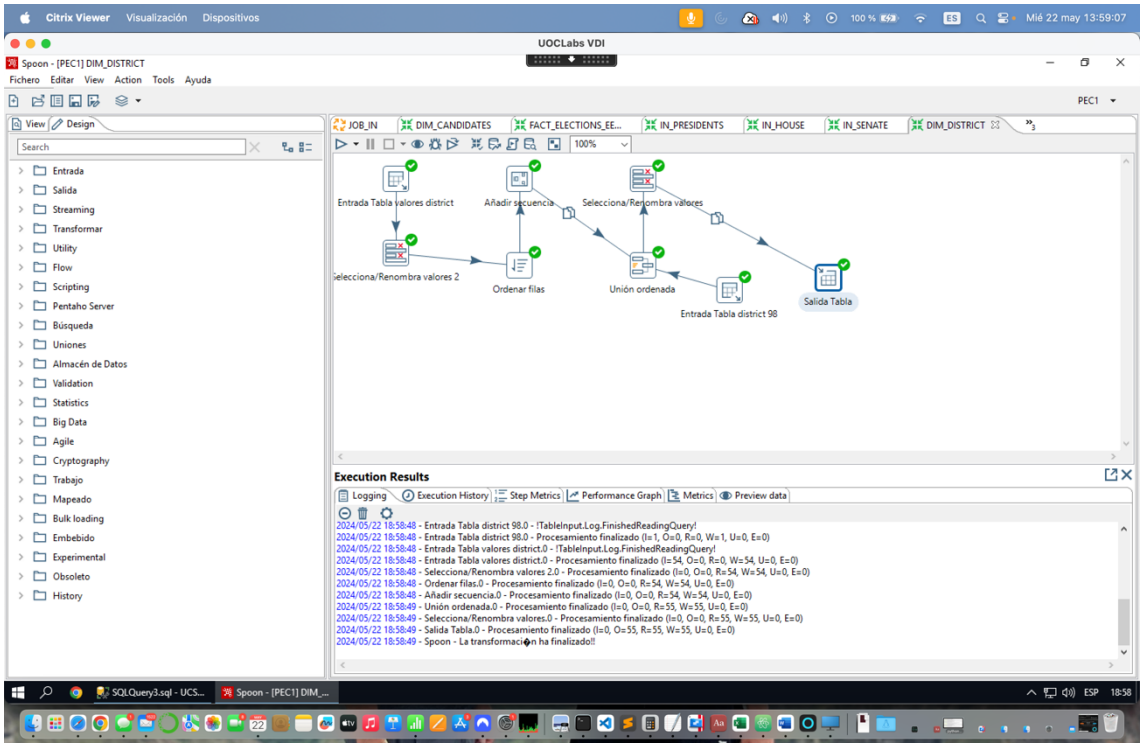
Después cambio a int district con 2 de longitud, para poder ordenarlos y que no se ordenen alfabéticamente, luego añado la clave NA, y lo uno, por ultimo lo vuelvo a convertir a string de 2



Creo la salida a tabla

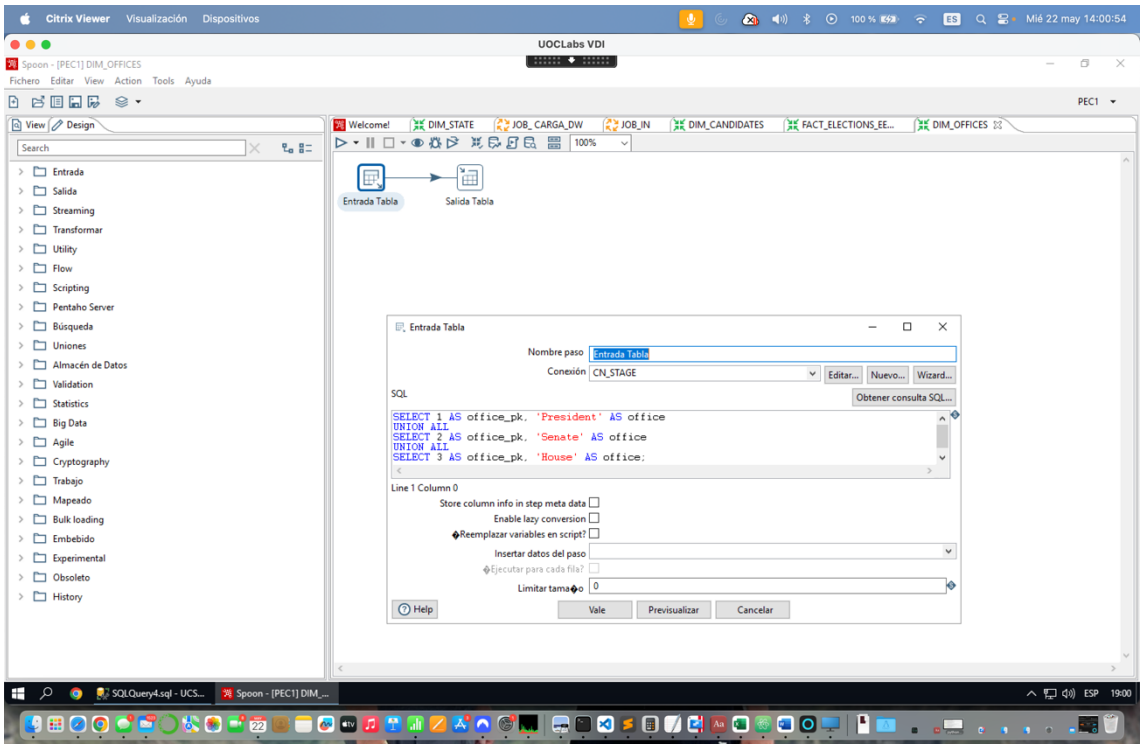


Ejecuto

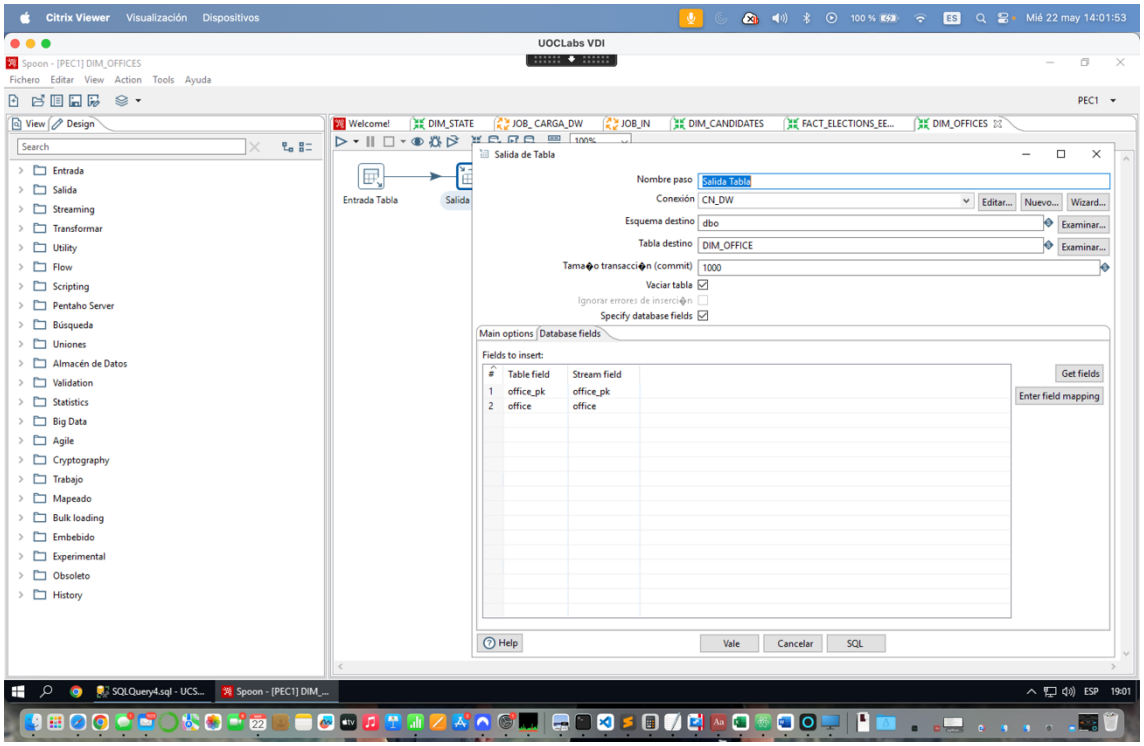


DIM_OFFICE

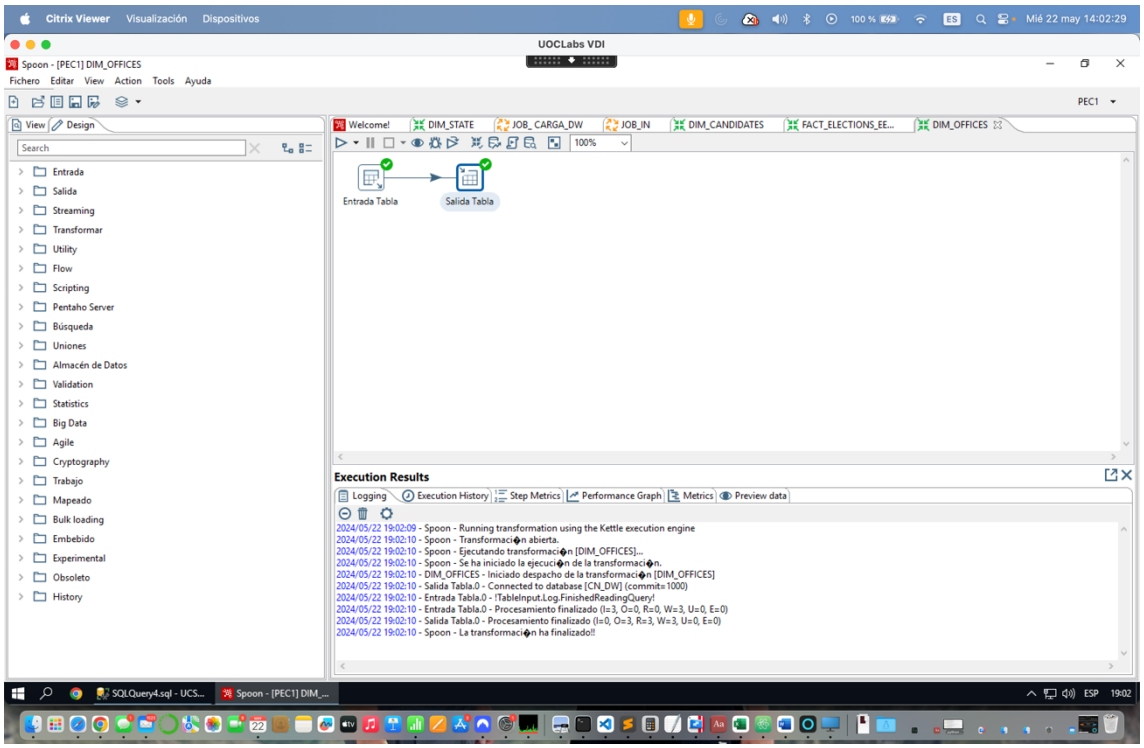
Creo la entrada de tabla, donde creo 3 claves valor, 1 para President, 2 para Senate y 3 para House



Creo la salida a tabla

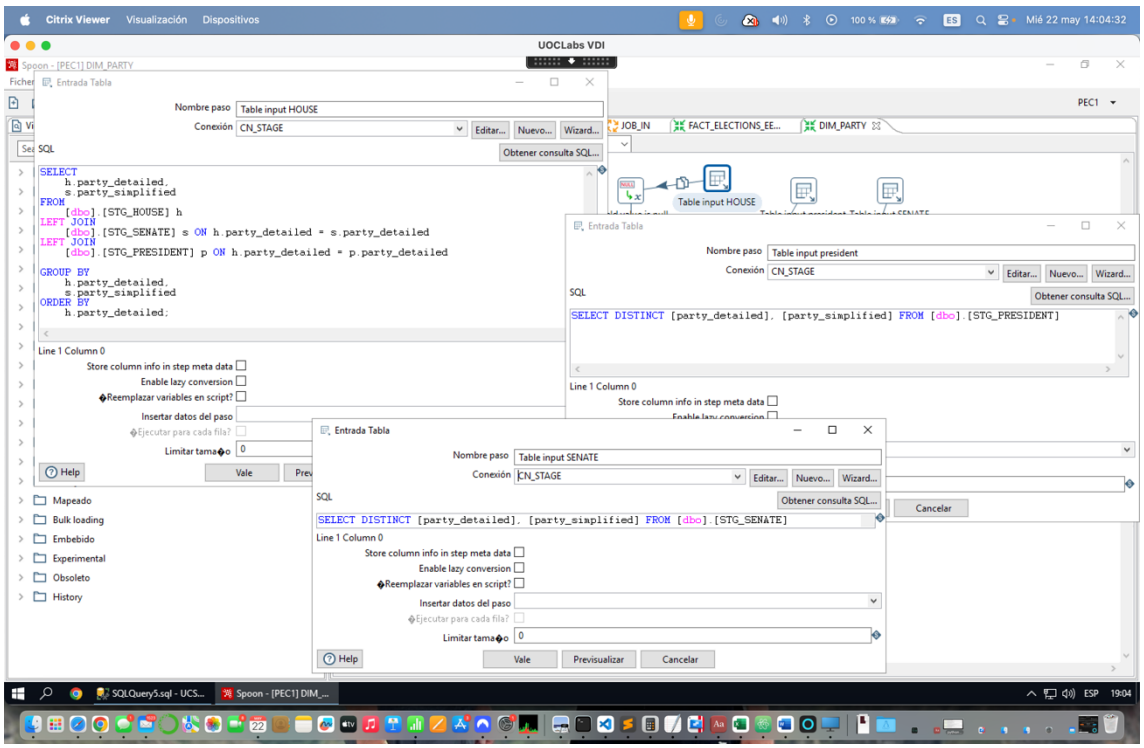


Compruebo

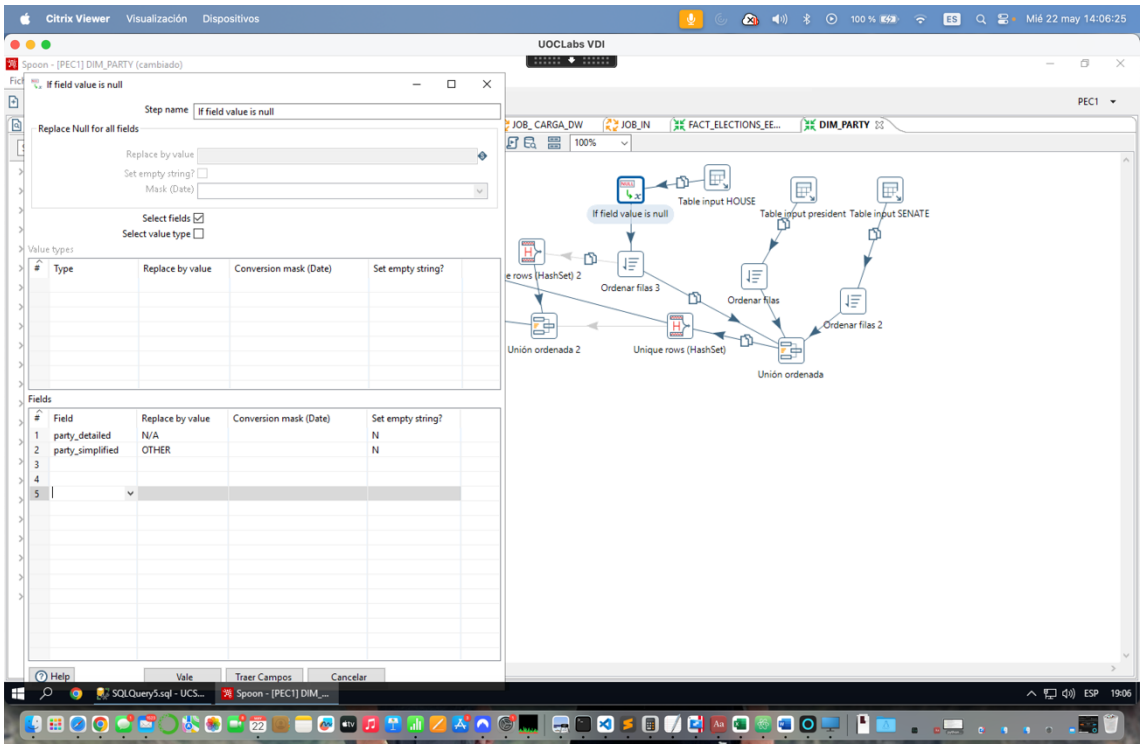


DIM_PARTY

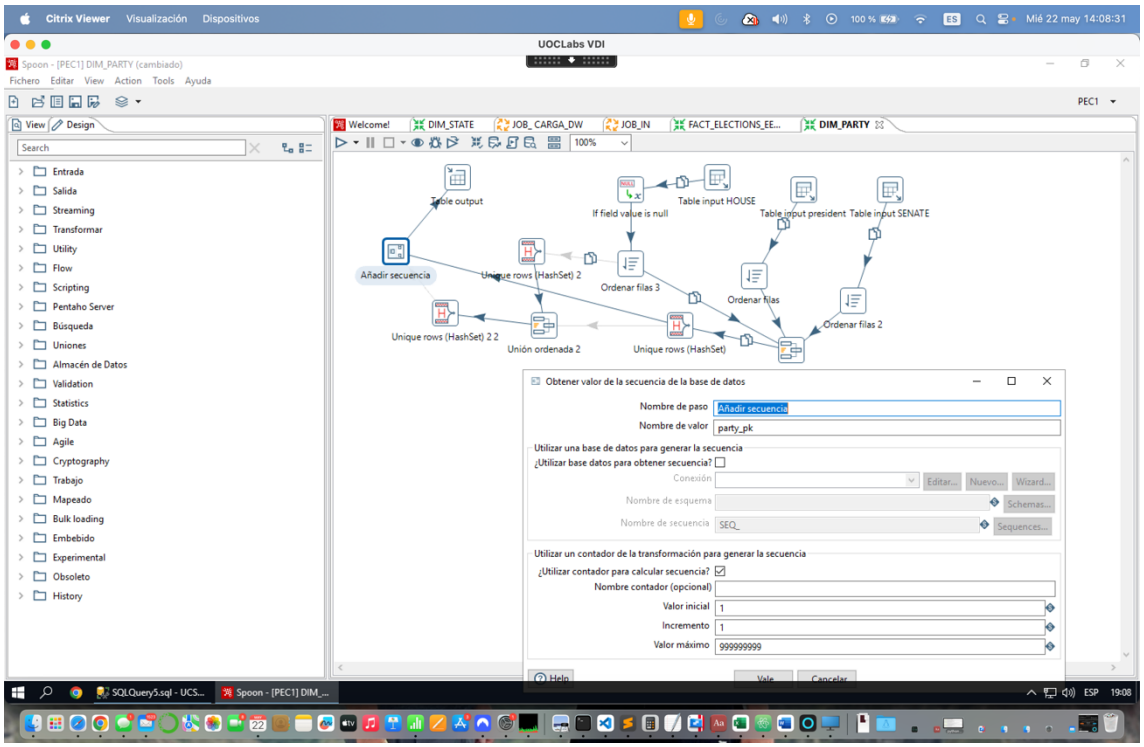
Aqui creo 3 consultas para tomar los partidos politicos de las 3 tablas. STG_PRESIDENT y STG_SENATE son consultas sencillas, y STG_HOUSE es un LEFT JOIN de las otras dos tablas añadiendo el valor de simplified de senate



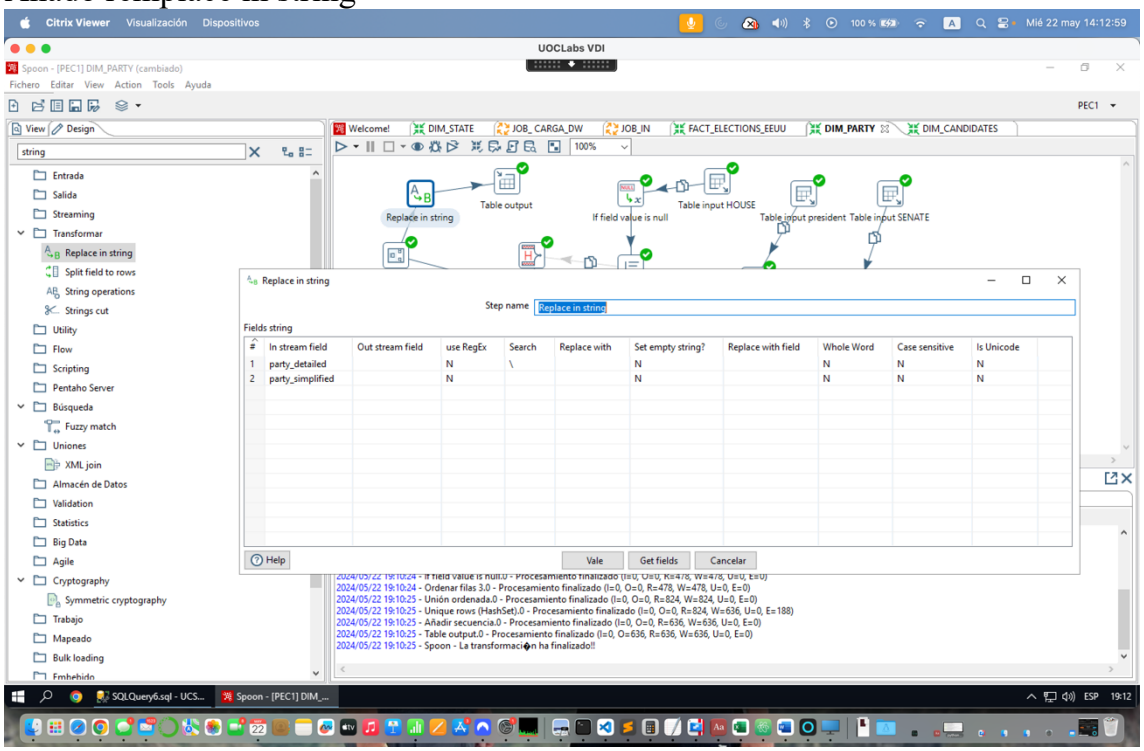
Para el caso de STG_HOUSE, si simplified es nulo lo cambio por OTHER ya que existe como no other solo algunos casos. (DEMOCRAT, REPUBLICAN, OTHER)



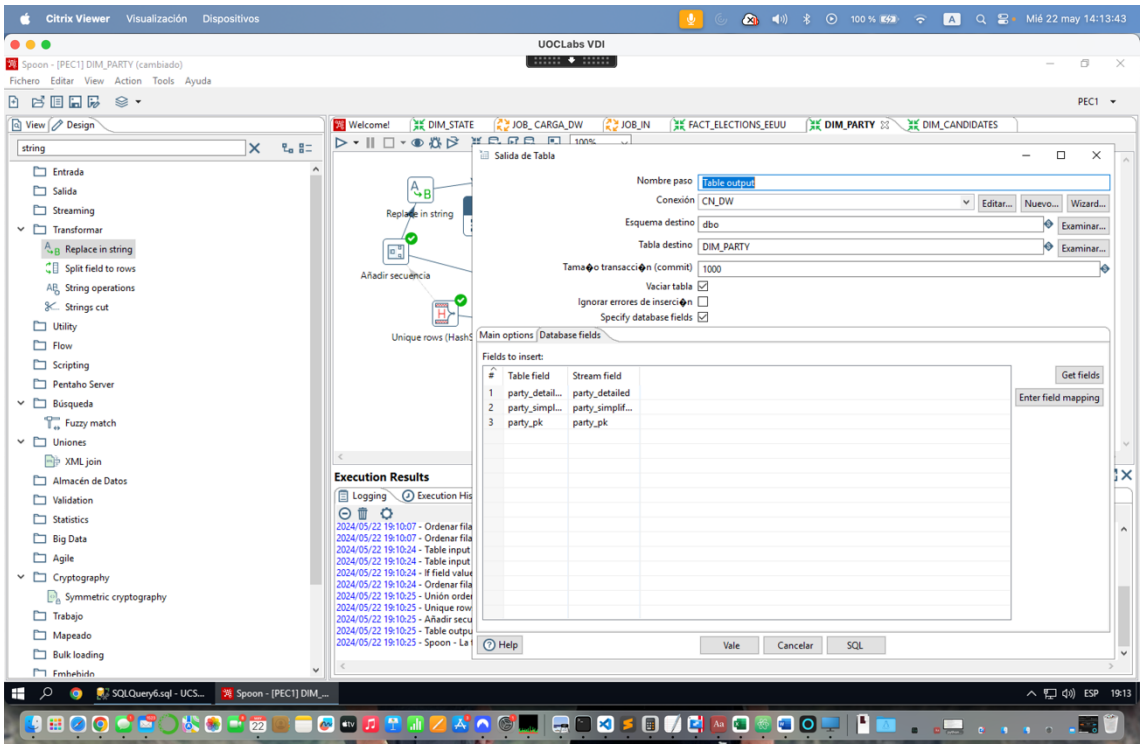
Ordeno las filas, y hago una unión ordenada, además elimino los repetidos. Para casi terminar, añado una secuencia



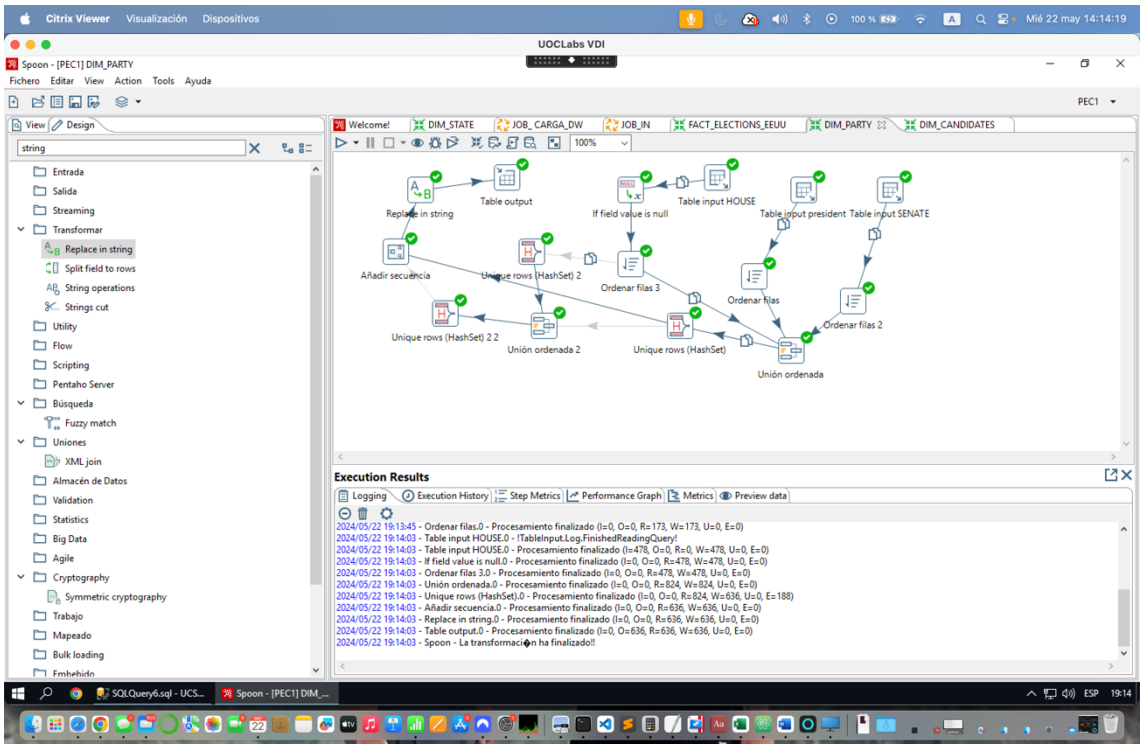
Añado remplace in string



Y queda por mostrar la salida a tabla

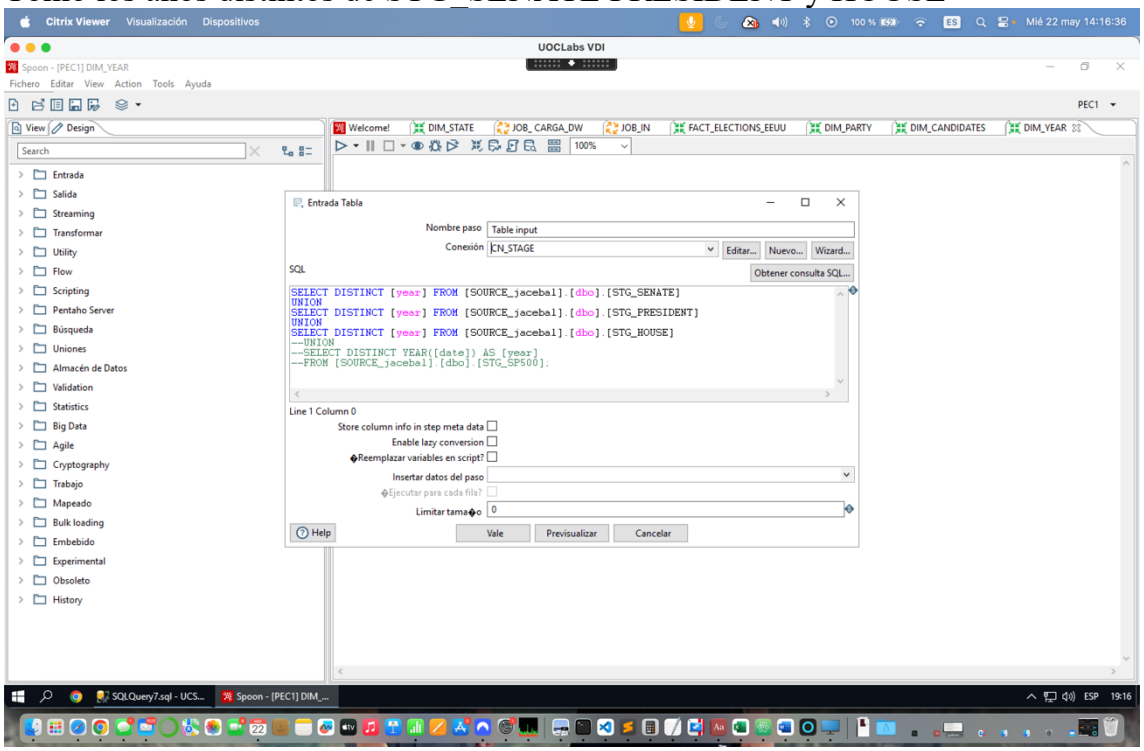


Compruebo la ejecución

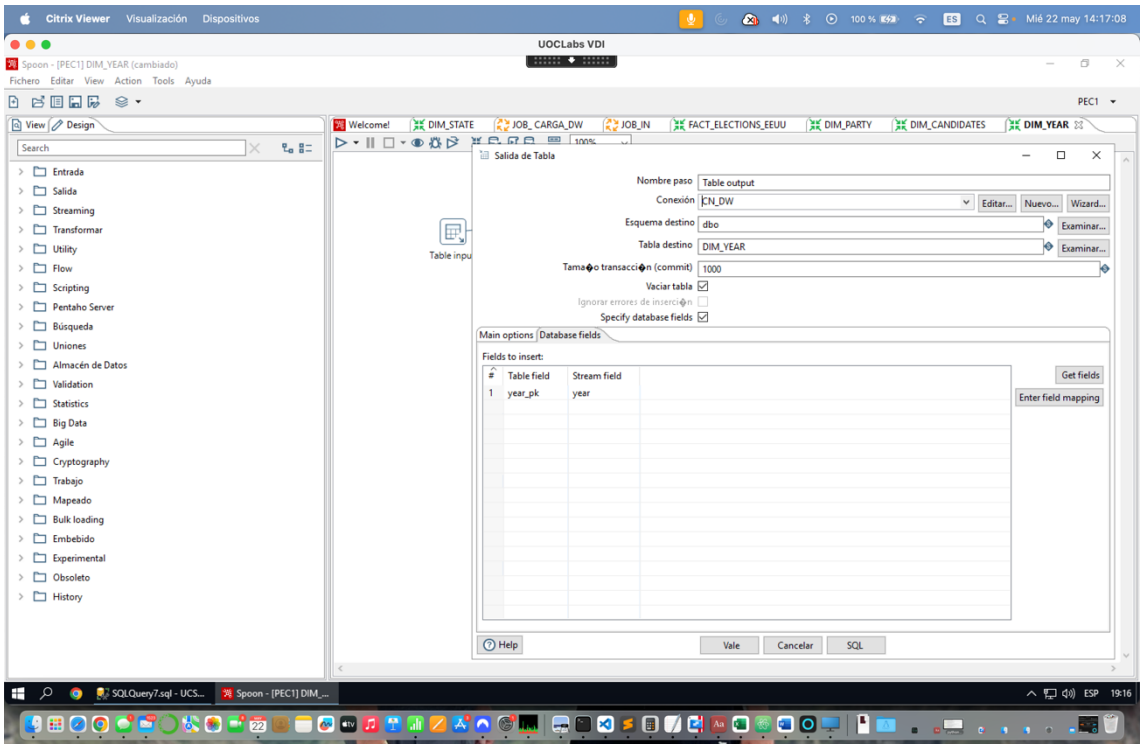


DIM_YEAR

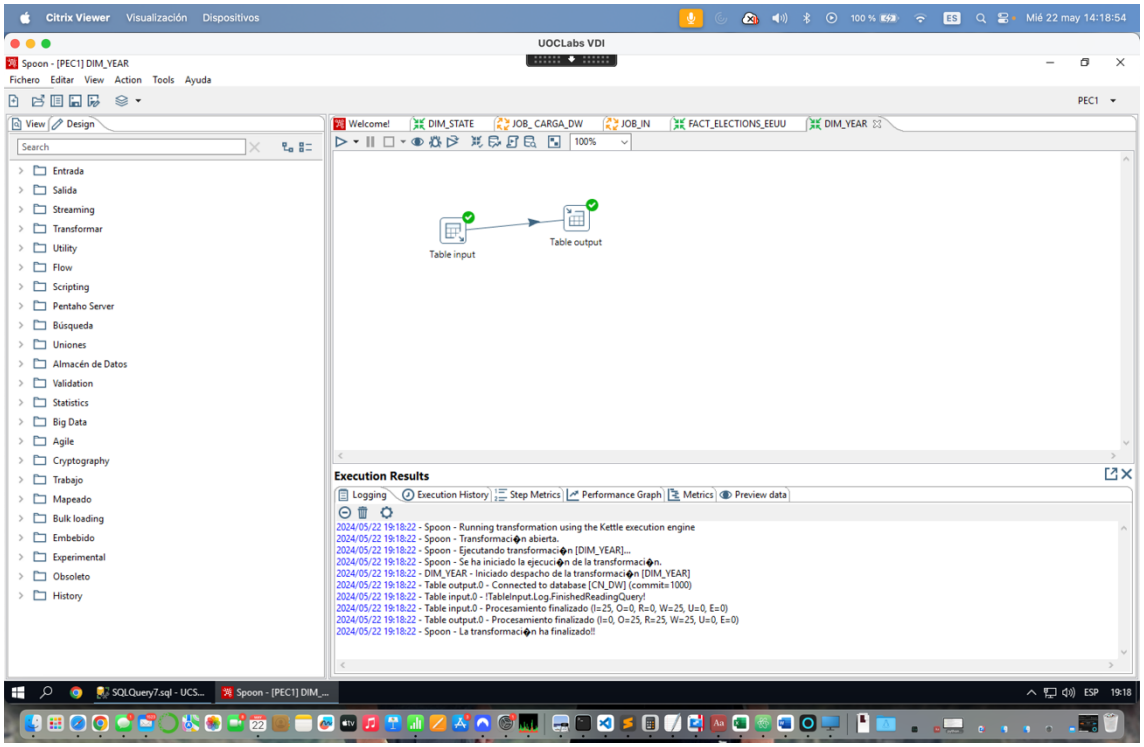
Tomo los años distintos de STG_SENATE PRESIDENT y HOUSE



Hago una salida a tabla



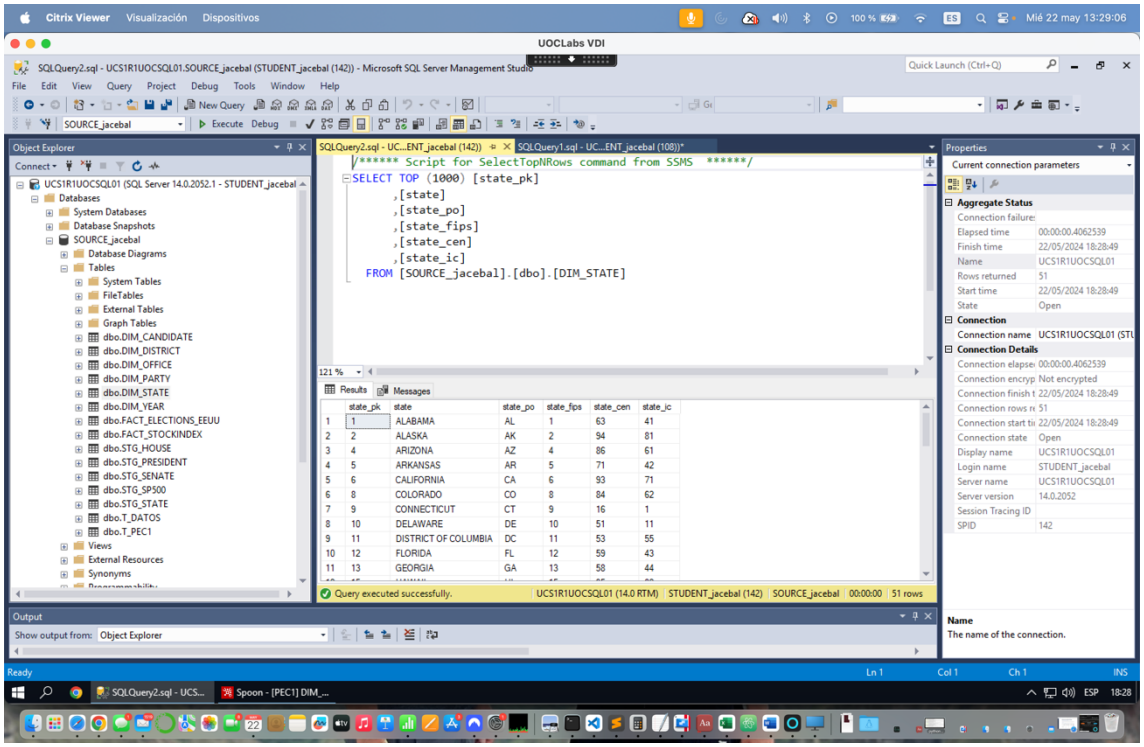
Ejecuto



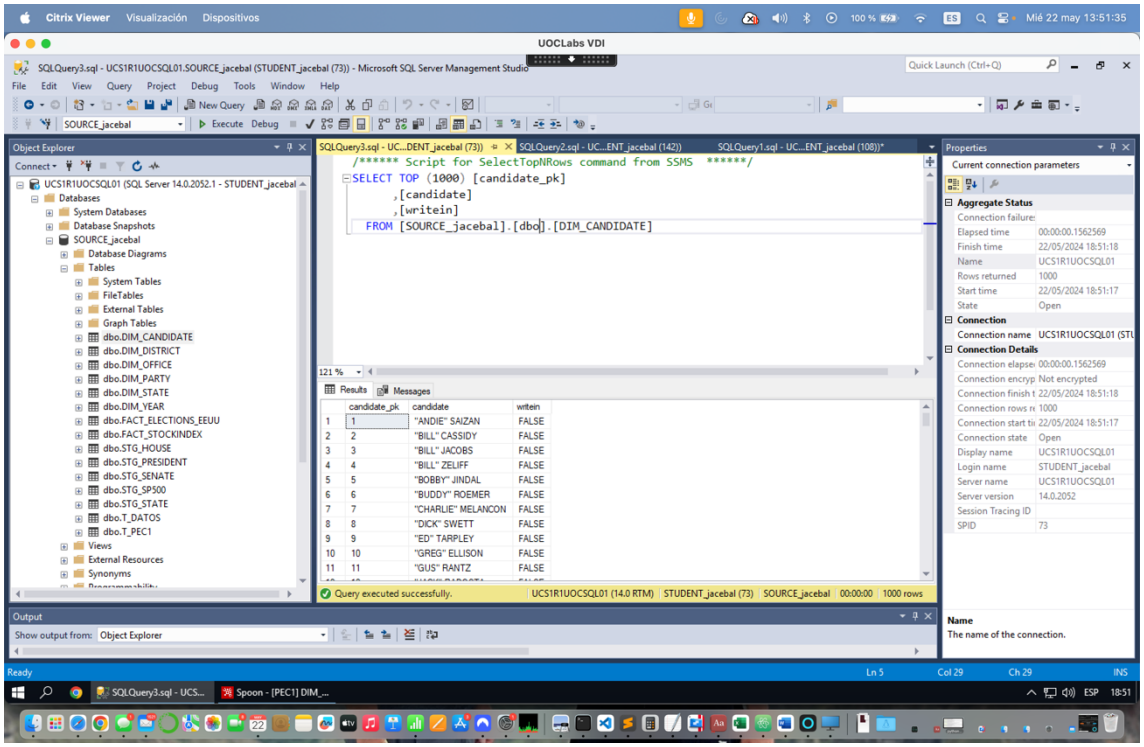
COMPROBACIONES MSSQL BLOQUE TR_DIM

Hago las consultas SQL para comprobar que se han volcado bien los datos

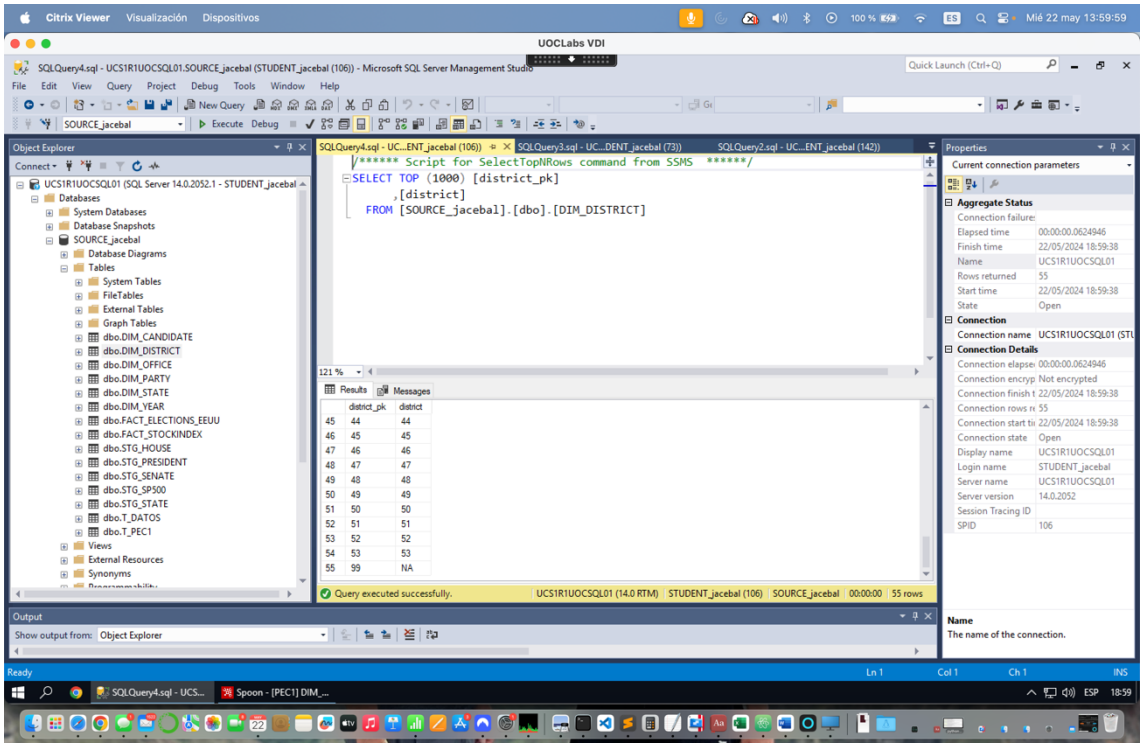
DIM_STATE



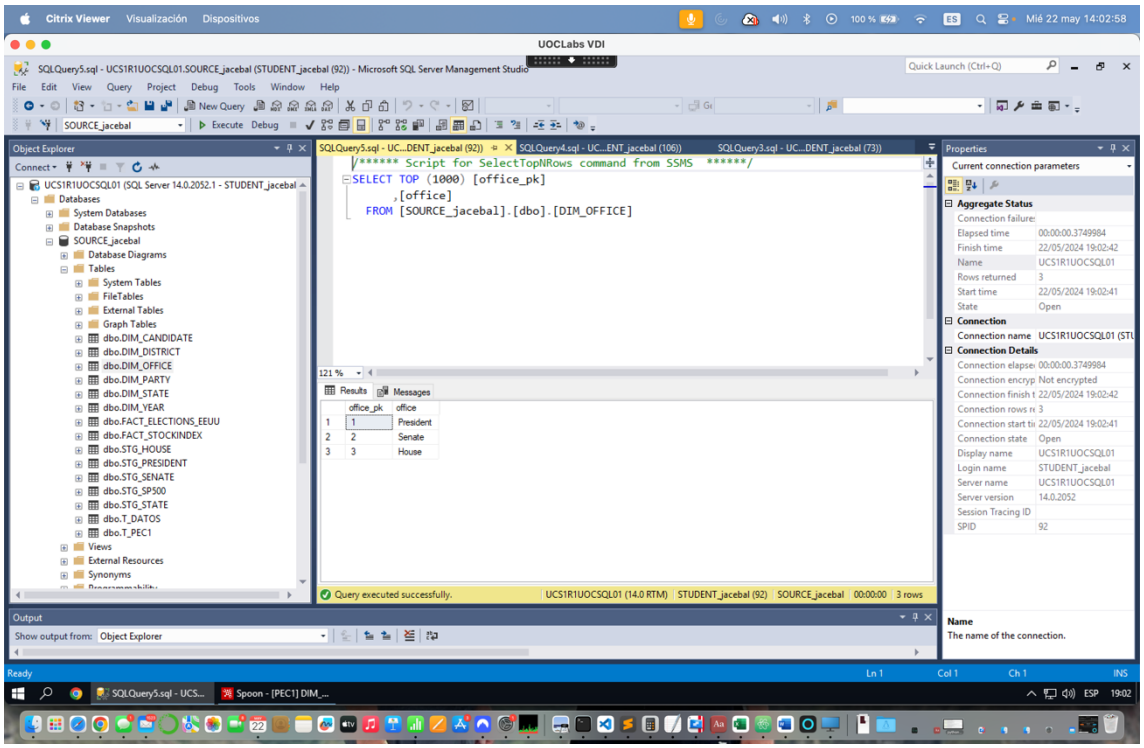
DIM_CANDIDATE



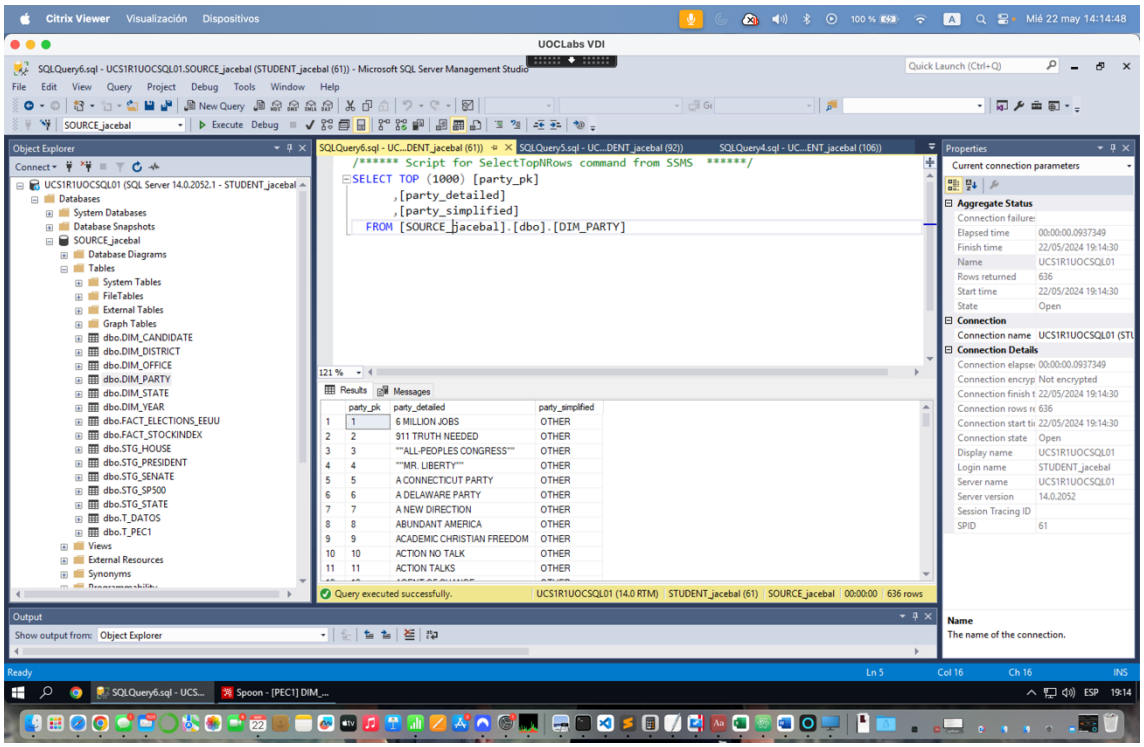
DIM_DISTRICT



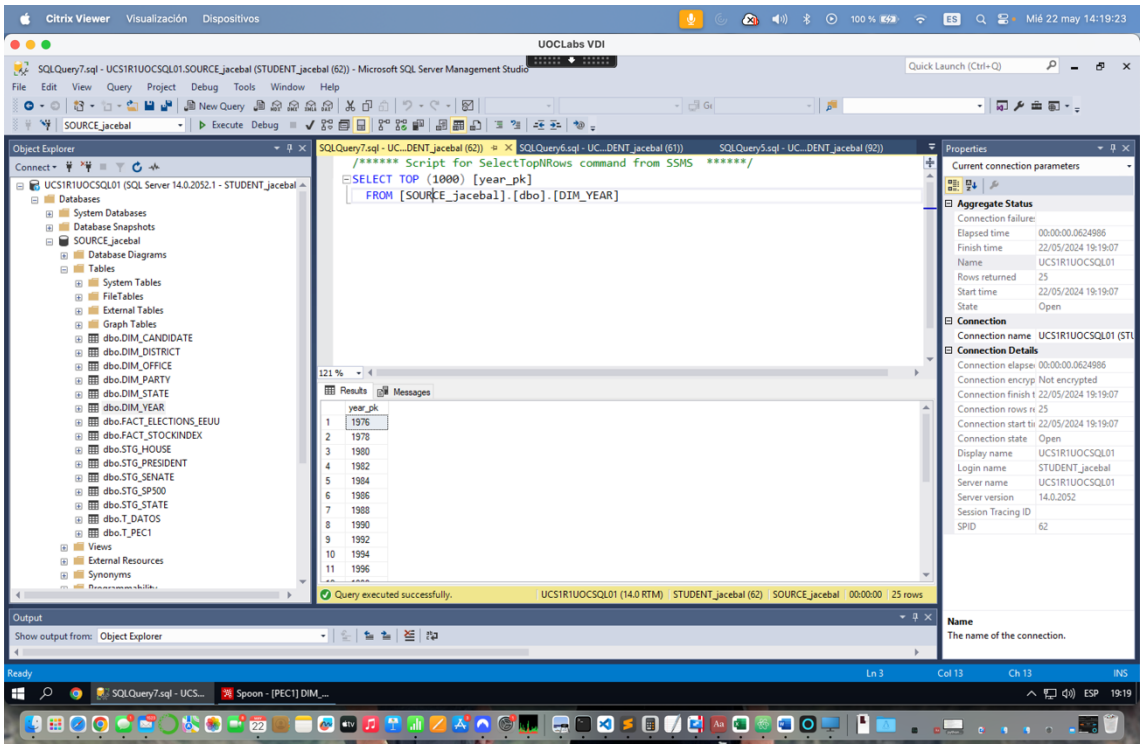
DIM_OFFICE



DIM_PARTY



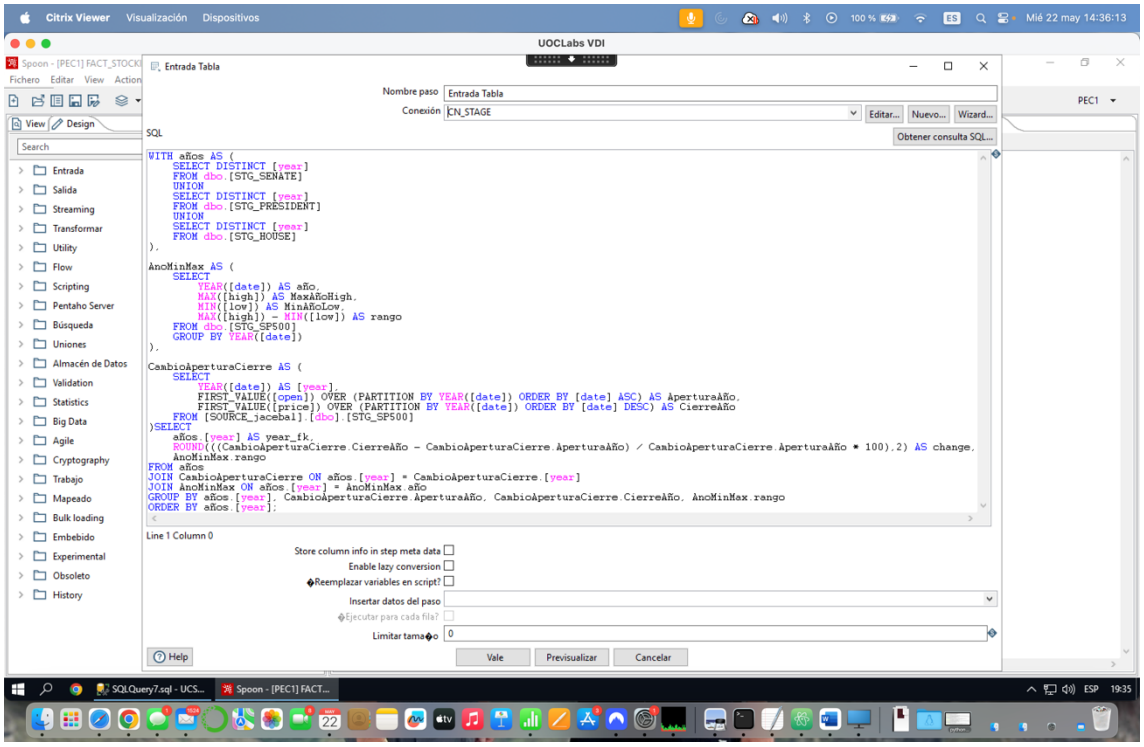
DIM_YEAR



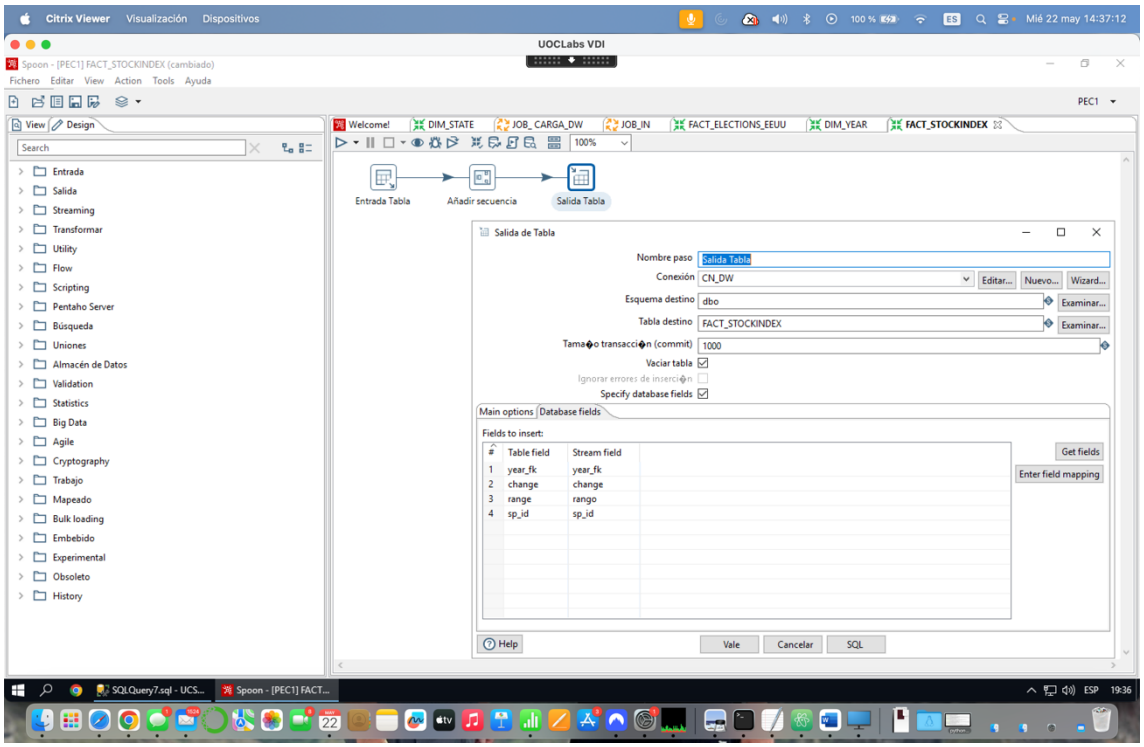
BLOQUE TR_FACT

FACT_STOCKINDEX

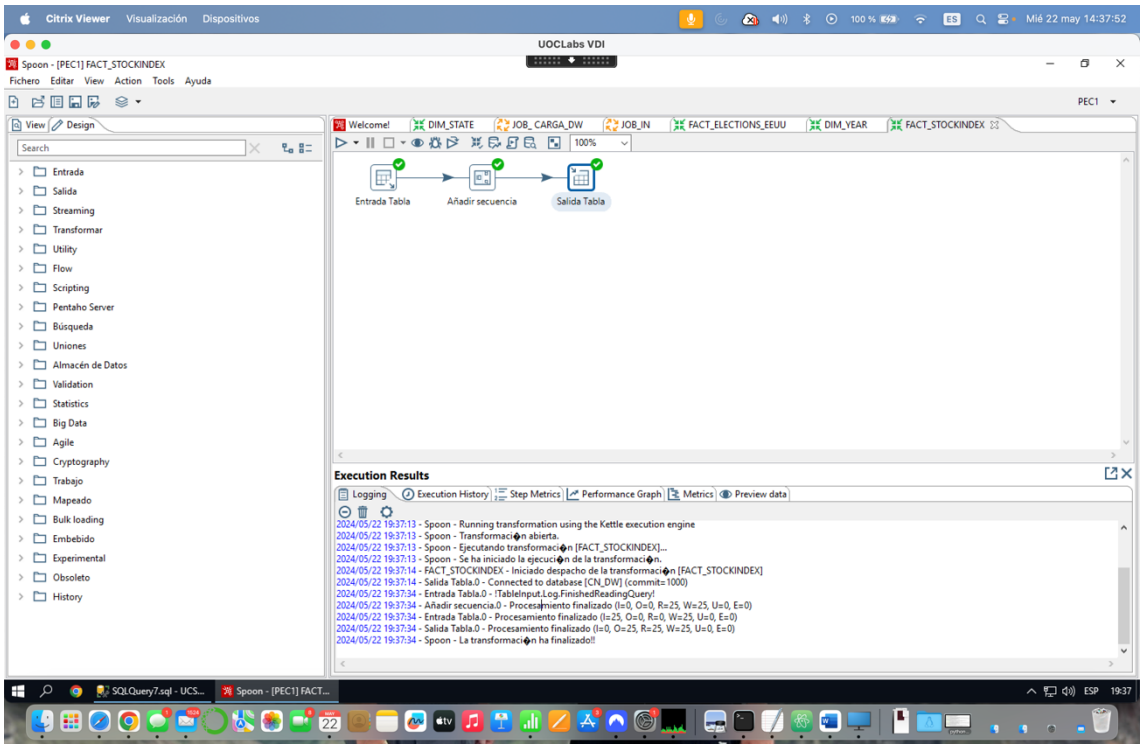
Aqui he hecho un select de STG_SENATE PRESIDENT y HOUSE para seleccionar los anos. Luego he selecionado los máximos y minimos de cada ano para saber el rango, he calculado la apertura y el cierre del año, he calculado el porcentaje de cambio, lo he agrupado y ordenado por años.



Añado una secuencia y después lo introduzco en la base de datos por medio de salida tabla

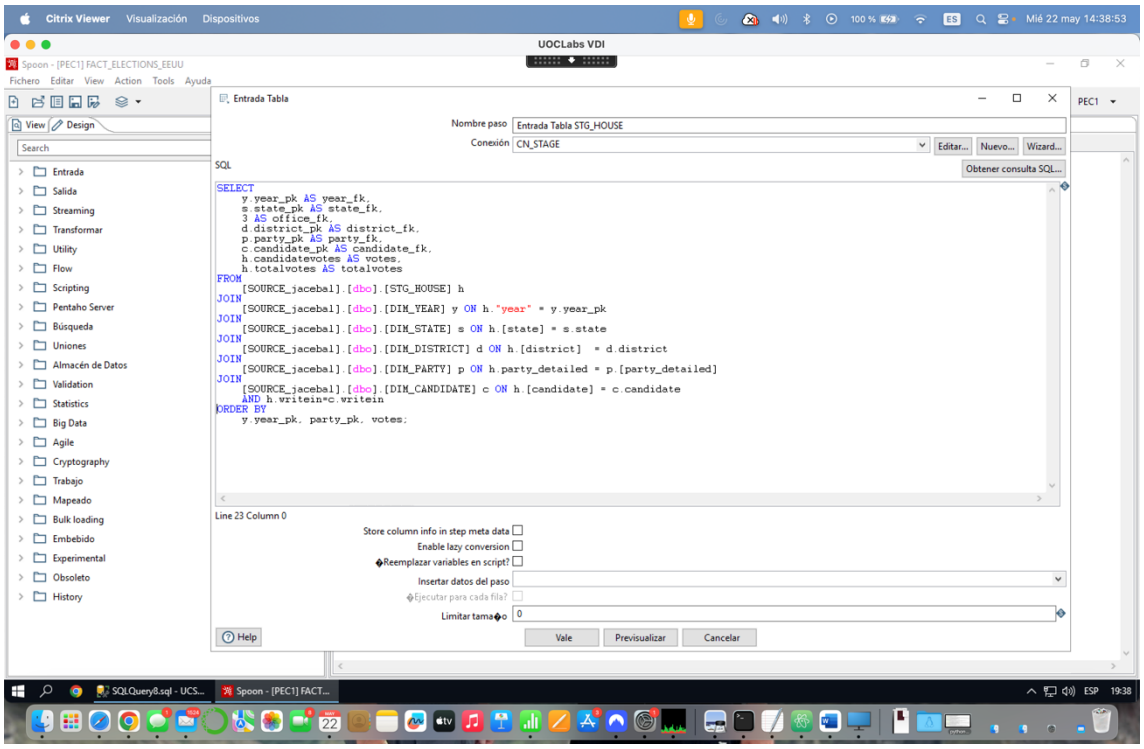

Citrix Viewer Visualización Dispositivos
UOCLabs VDI
Spoon - [PEC1] FACT_STOCKINDEX (cambiado)
Fichero Editar View Action Tools Ayuda
PEC1
Welcome! DIM_STATE JOB_CARGA_DW JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX
Entrada Tabla
Añadir secuencia
Salida Tabla
Salida de Tabla
Nombre paso
Salida Tabla
Conexión
CN_DW
Esquema destino
dbo
Tabla destino
FACT_STOCKINDEX
Tamaño transacción (commit)
1000
Vaciar tabla
Specify database fields
Main options
Database fields
Fields to insert:
Table field
Stream field
year_fk
change
range
sp_id
Get fields
Enter field mapping
Help
Vale
Cancelar
SQL

Compruebo

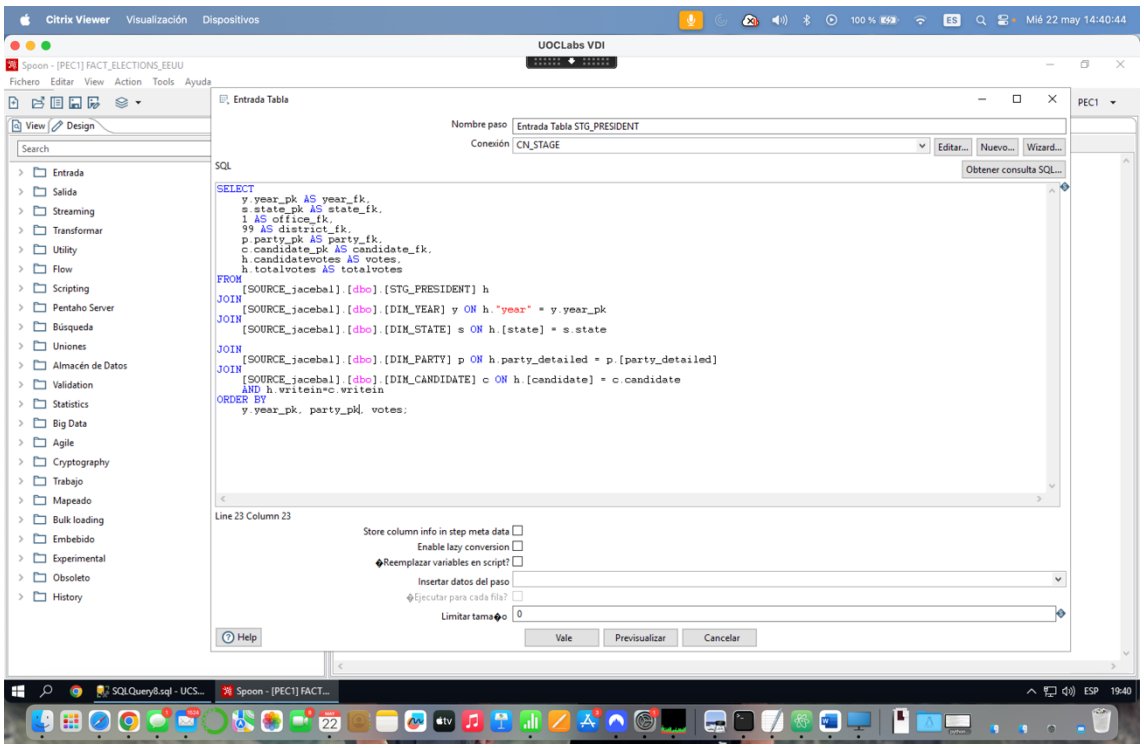

Citrix Viewer Visualización Dispositivos
UOCLabs VDI
Spoon - [PEC1] FACT_STOCKINDEX
Fichero Editar View Action Tools Ayuda
PEC1
Welcome! DIM_STATE JOB_CARGA_DW JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX
Entrada Tabla
Añadir secuencia
Salida Tabla
Execution Results
Logging
Execution History
Step Metrics
Performance Graph
Metrics
Preview data
2024/05/22 19:37:13 - Spoon - Running transformation using the Kettle execution engine
2024/05/22 19:37:13 - Spoon - Transformación abierta.
2024/05/22 19:37:13 - Spoon - Ejecutando transformación [FACT_STOCKINDEX]...
2024/05/22 19:37:13 - Spoon - Se ha iniciado la ejecución de la transformación.
2024/05/22 19:37:14 - FACT_STOCKINDEX - Iniciado despacho de la transformación [FACT_STOCKINDEX]
2024/05/22 19:37:14 - Salida Tabla.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 19:37:34 - Entrada Tabla.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:37:34 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=25, W=25, U=0, E=0)
2024/05/22 19:37:34 - Entrada Tabla.0 - Procesamiento finalizado (I=25, O=0, R=0, W=25, U=0, E=0)
2024/05/22 19:37:34 - Salida Tabla.0 - Procesamiento finalizado (I=0, O=25, R=25, W=25, U=0, E=0)
2024/05/22 19:37:34 - Spoon - La transformación ha finalizado!!

FACT_ELECTIONS_EEUU

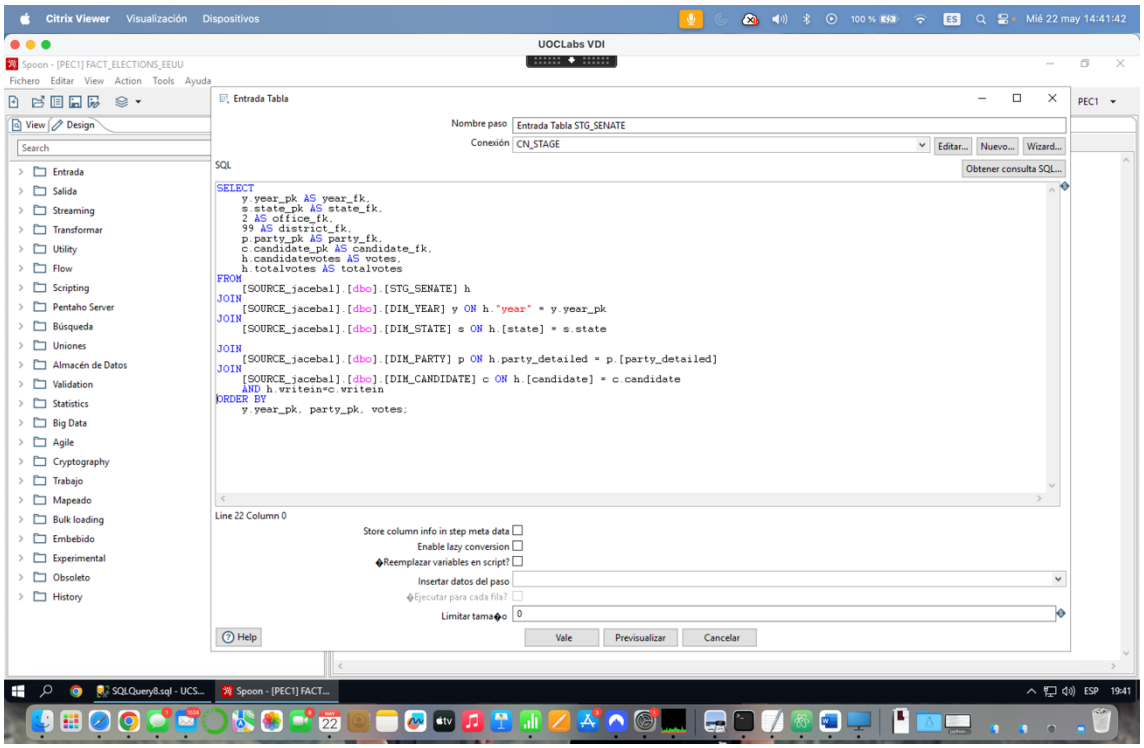
Hago una entrada table de STG_HOUSE
Aquí hago una consulta que tomo ya cada valor de su dimensión, a excepción de office_fk que he puesto directamente 3



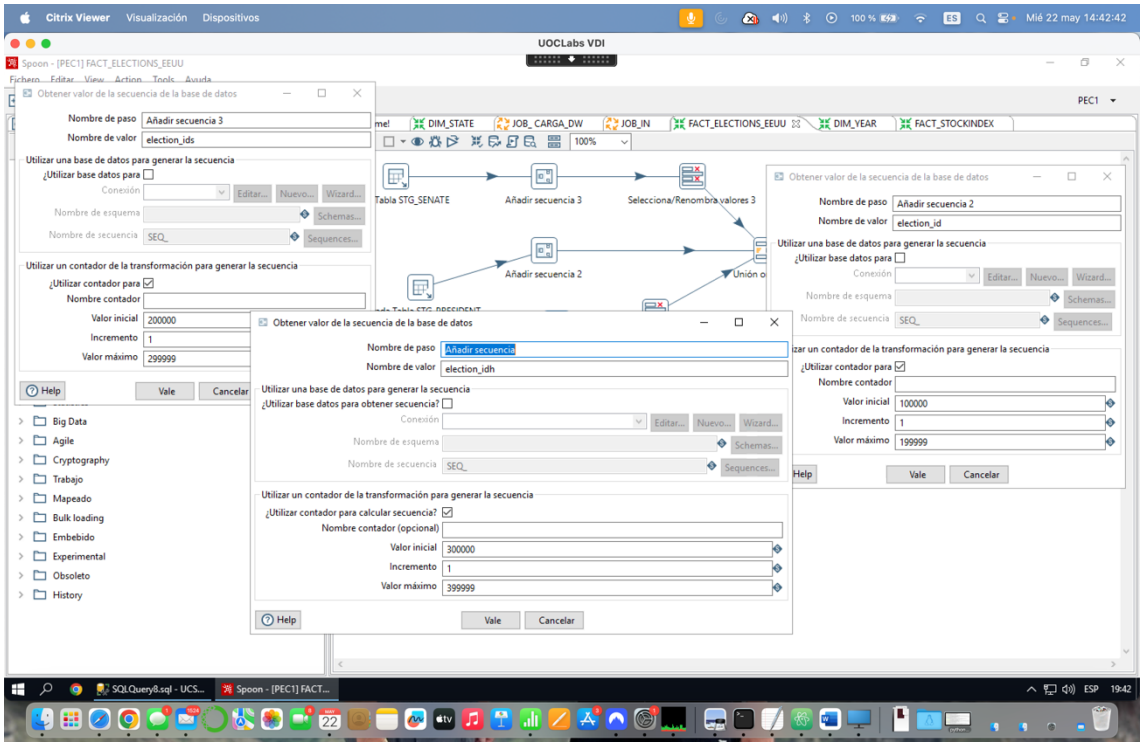
Para STG_PRESIDENT hago igual, el cambio es 1 para office_fk y district_fk con valor 99



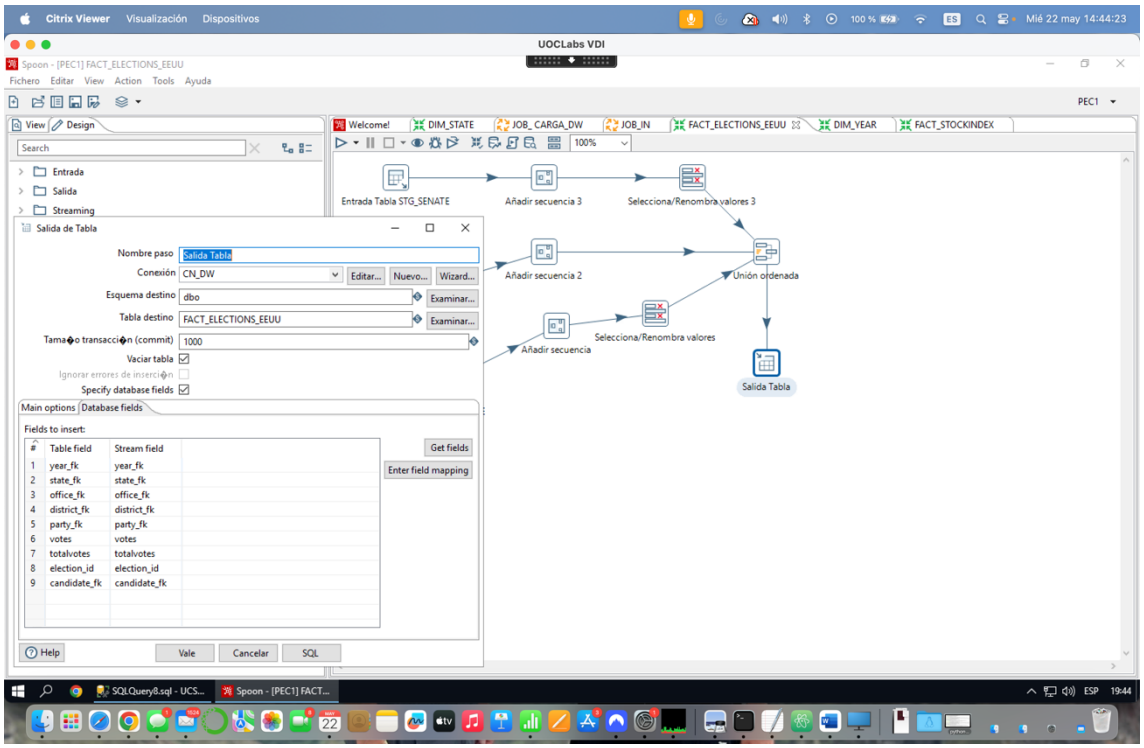
STG_SENATE hago parecido, cambiando la office_fk por 2



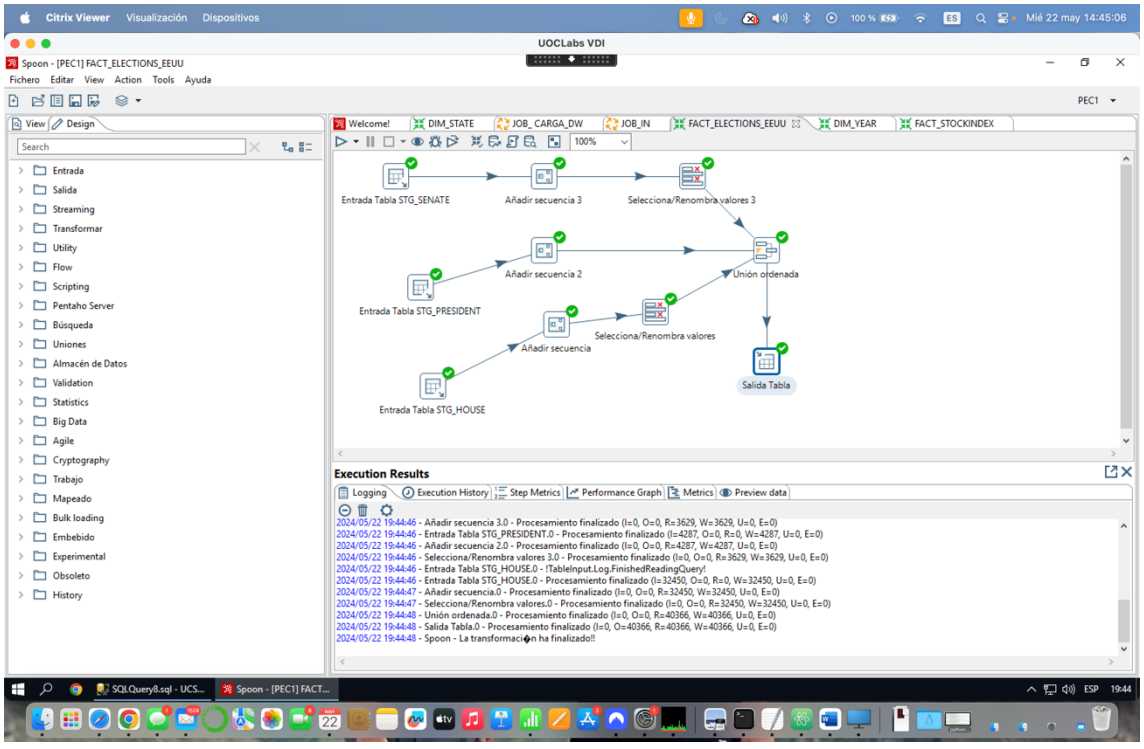
Hago 3 secuencias, empieza en 100000 para president, por 200000 para senate y por 300000 para house. Como no se puede crear en la misma transformación dos veces la misma secuencia, tienen nombres diferentes que luego renombro.



Hago una union ordenada y después salida a tabla

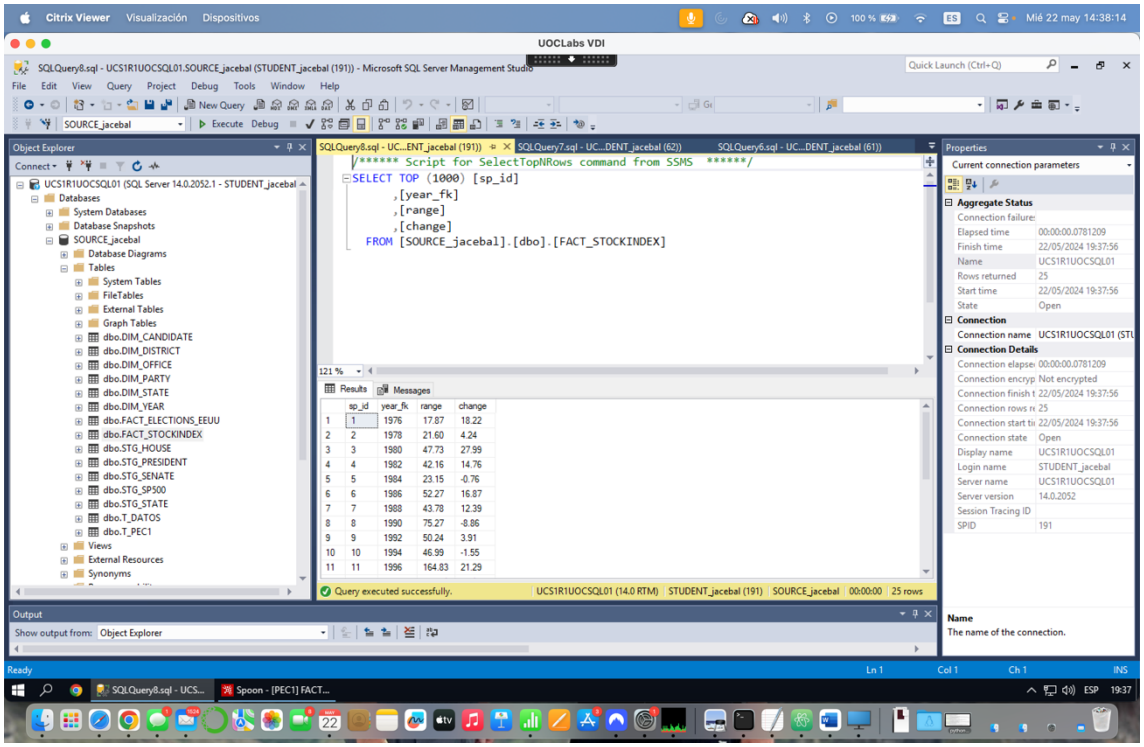


Compruebo

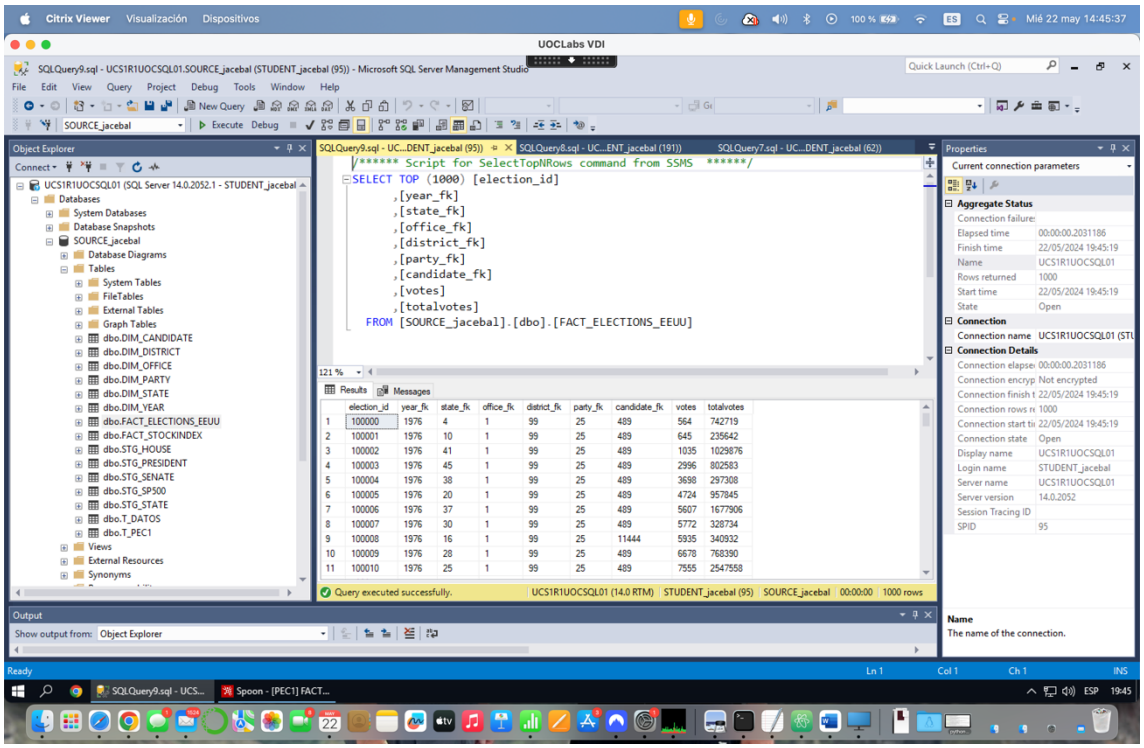


COMPROBACIONES MSSQL BLOQUE TR_FACT

FACT_STOCKINDEX



FACT_ELECTIONS_EEUU



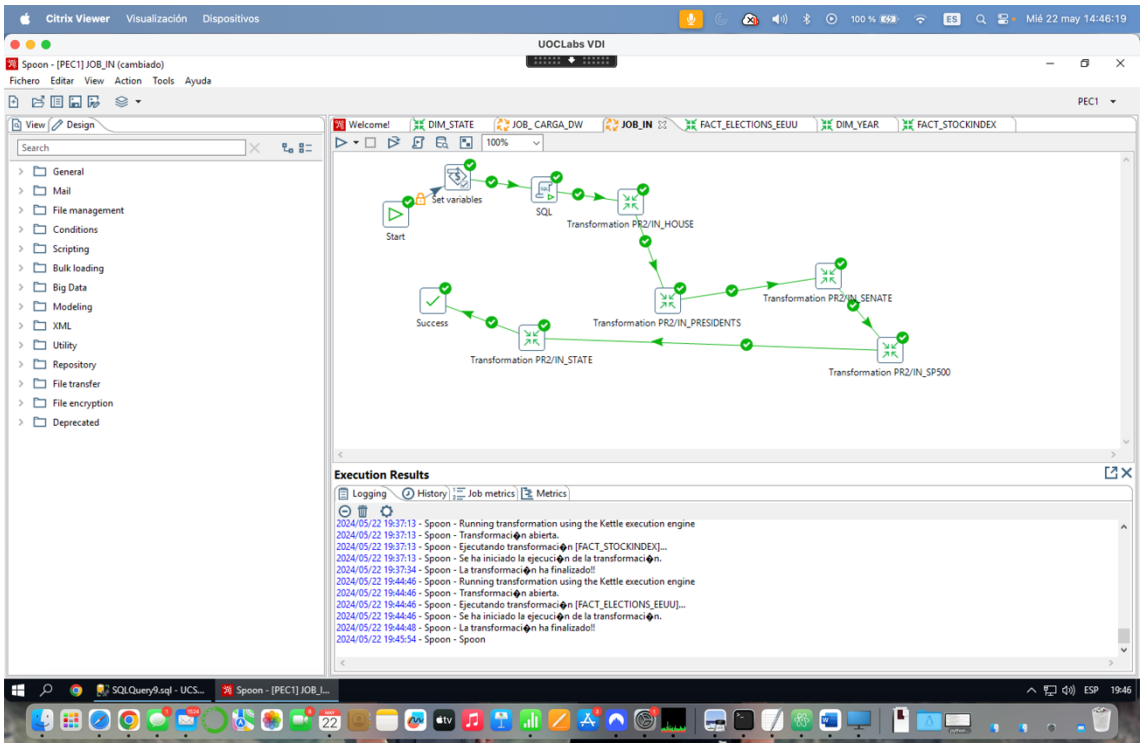
CONSIDERACIONES TRANSFORMACIONES

La implantación de transformaciones ha sido en orden STG luego las DIM y luego las FACT, esto es así ya que por la carga de datos a la zona staging es obvio que ese es el primer paso, sin embargo, luego las dimensiones deben ser antes de FACT para evitar problemas de clave foránea.

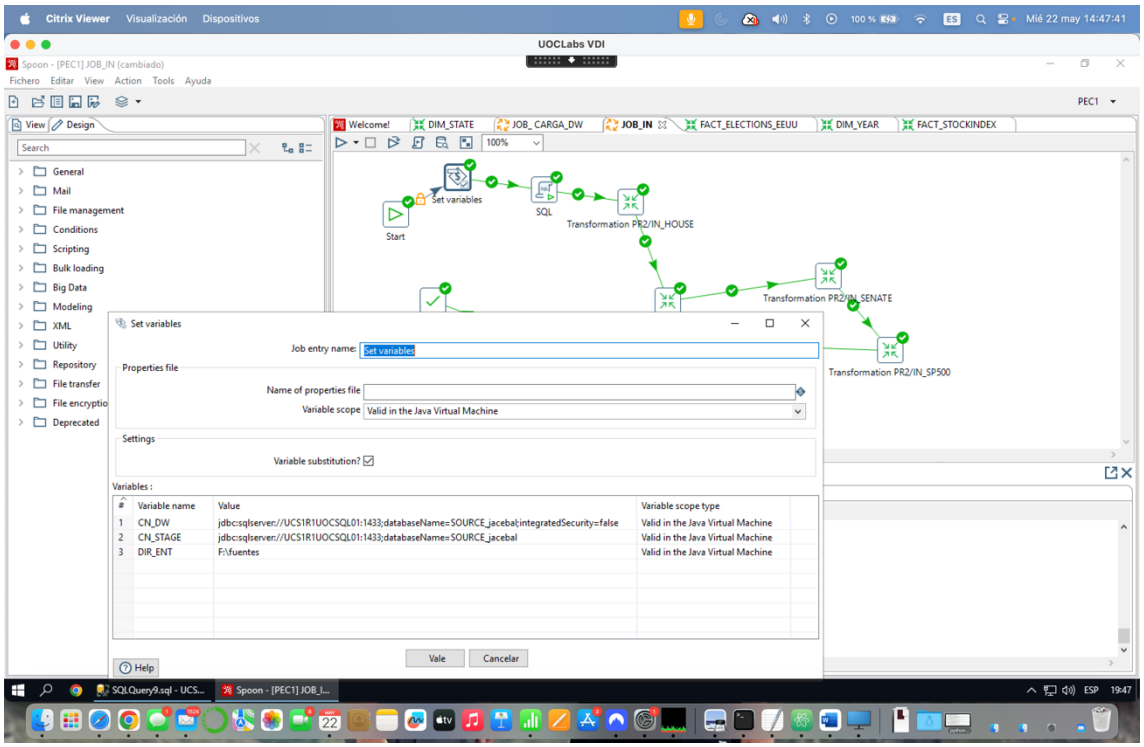
IMPLEMENTACIÓN DE AUTOMATIZACIÓN CON TRABAJOS (JOBS)

JOB_IN

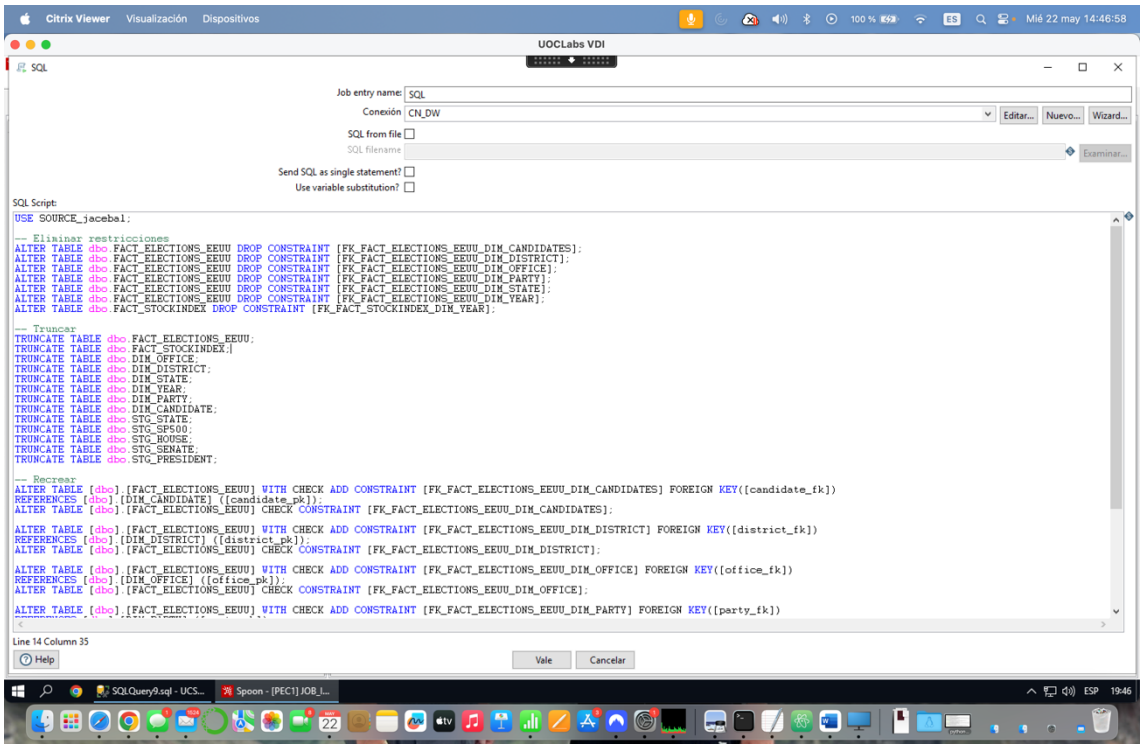
Vista general de JOB_IN, se puede observar establecer las 3 variables, script SQL y todas las transformaciones IN



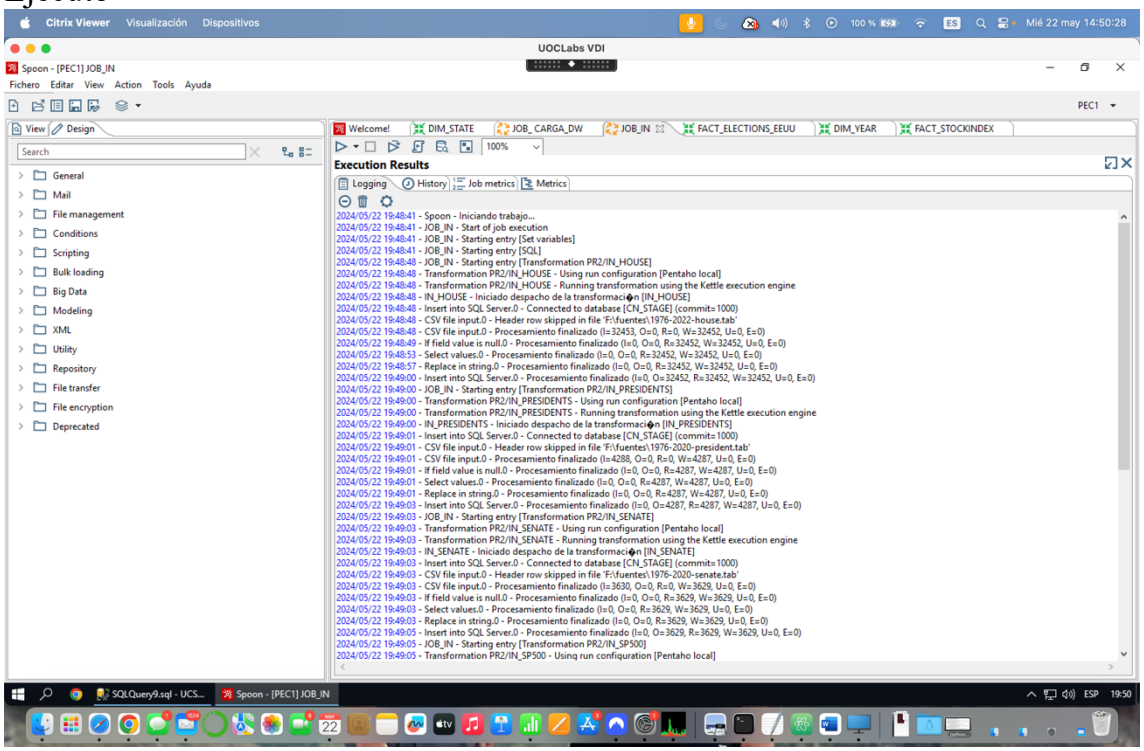
Añado las 3 variables

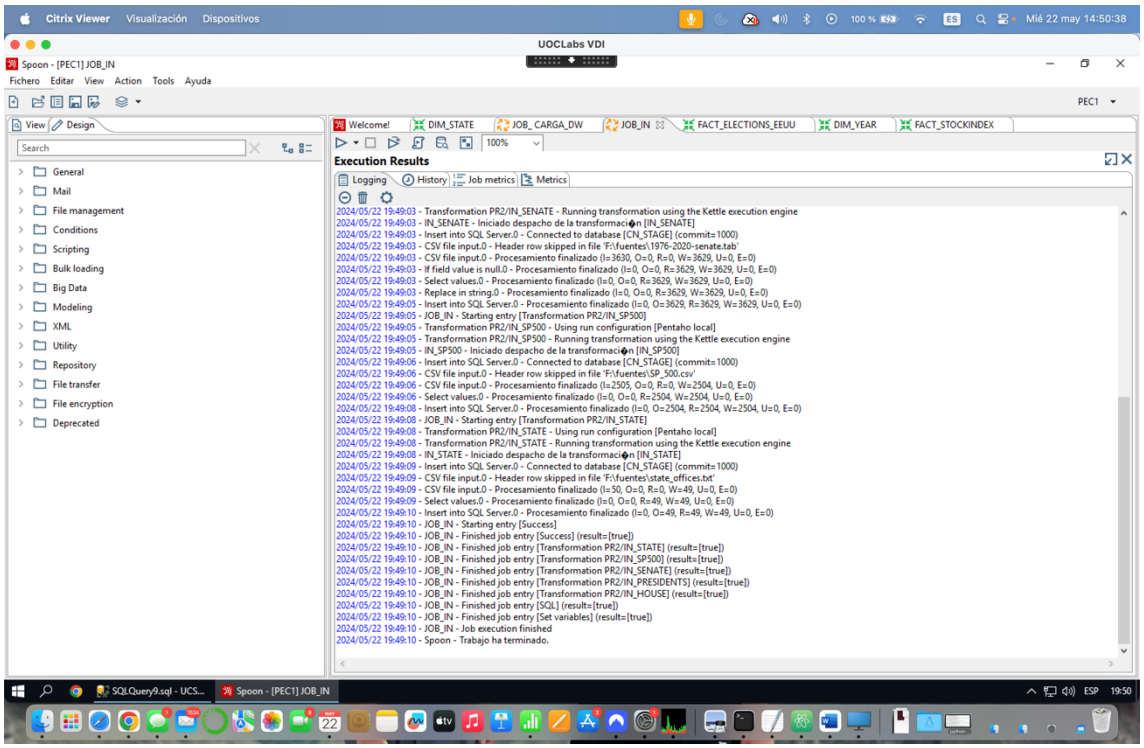


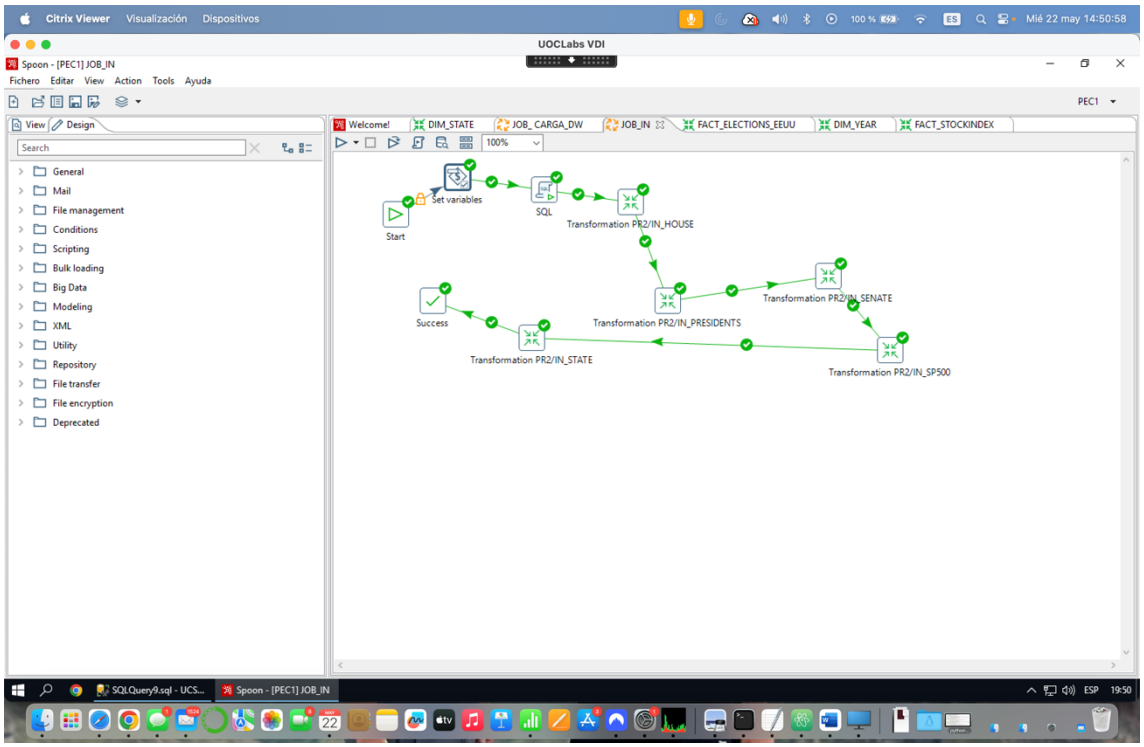
Elimino las restricciones de tabla para eliminar todo y que truncate funcione



Ejecuto

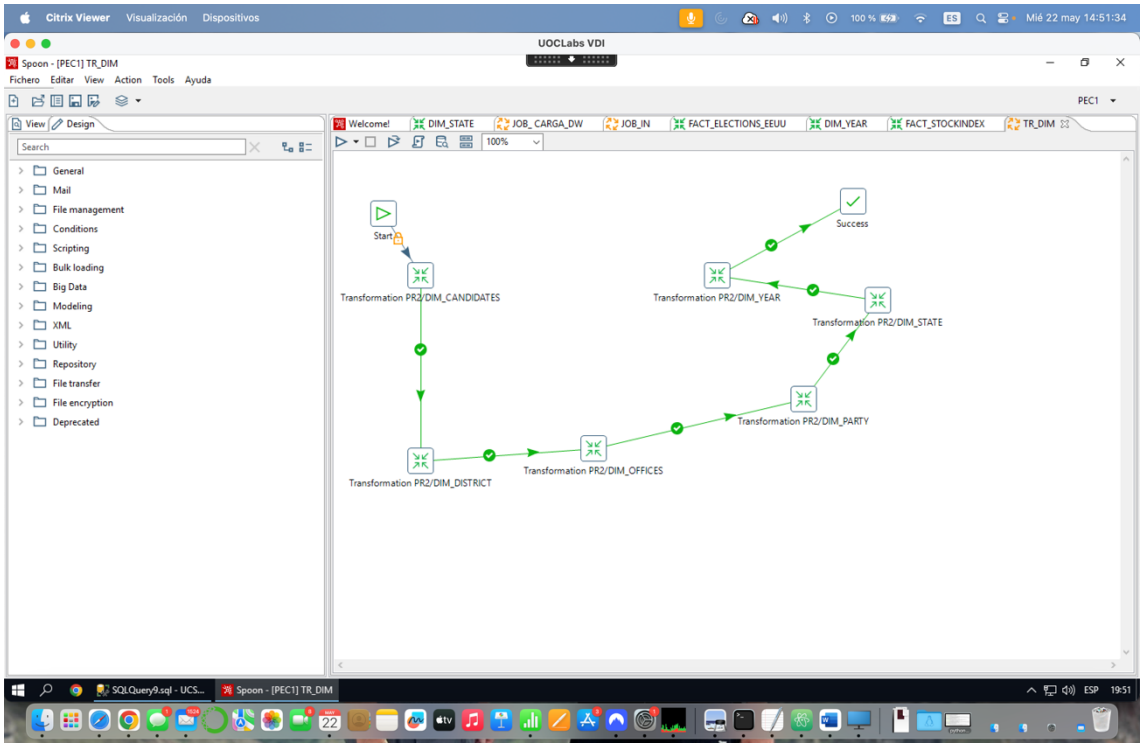

Citrix Viewer Visualización Dispositivos
Mié 22 may 14:50:28
UOCLabs VDI
Spoon - [PEC1] JOB_IN
Fichero Editar View Action Tools Ayuda
PEC1
View Design
Search
General
Mail
File management
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Deprecated
Welcome! DIM_STATE JOB_CARGA_DW JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX
Execution Results
Logging History Job metrics Metrics
2024/05/22 19:48:41 - Spoon - Iniciando trabajo...
2024/05/22 19:48:41 - JOB_IN - Start of job execution
2024/05/22 19:48:41 - JOB_IN - Starting entry [Set variables]
2024/05/22 19:48:41 - JOB_IN - Starting entry [SQL]
2024/05/22 19:48:48 - JOB_IN - Starting entry [Transformation PR2/IN_HOUSE]
2024/05/22 19:48:48 - Transformation PR2/IN_HOUSE - Using run configuration [Pentaho local]
2024/05/22 19:48:48 - Transformation PR2/IN_HOUSE - Running transformation using the Kettle execution engine
2024/05/22 19:48:48 - IN_HOUSE - Iniciado despacho de la transformaci￳n [IN_HOUSE]
2024/05/22 19:48:48 - Insert into SQL Server.0 - Connected to database [CN_STAGE] (commit=1000)
2024/05/22 19:48:48 - CSV file input.0 - Header row skipped in file 'F:\fuentes\1976-2022-house.tab'
2024/05/22 19:48:48 - CSV file input.0 - Procesamiento finalizado (I=32453, O=0, R=0, W=32452, U=0, E=0)
2024/05/22 19:48:49 - If field value is null.0 - Procesamiento finalizado (I=0, O=0, R=32452, W=32452, U=0, E=0)
2024/05/22 19:48:53 - Select values.0 - Procesamiento finalizado (I=0, O=0, R=32452, W=32452, U=0, E=0)
2024/05/22 19:48:57 - Replace in string.0 - Procesamiento finalizado (I=0, O=0, R=32452, W=32452, U=0, E=0)
2024/05/22 19:49:00 - Insert into SQL Server.0 - Procesamiento finalizado (I=0, O=32452, R=32452, W=32452, U=0, E=0)
2024/05/22 19:49:00 - JOB_IN - Starting entry [Transformation PR2/IN_PRESIDENTS]
2024/05/22 19:49:00 - Transformation PR2/IN_PRESIDENTS - Using run configuration [Pentaho local]
2024/05/22 19:49:00 - Transformation PR2/IN_PRESIDENTS - Running transformation using the Kettle execution engine
2024/05/22 19:49:00 - IN_PRESIDENTS - Iniciado despacho de la transformaci￳n [IN_PRESIDENTS]
2024/05/22 19:49:01 - Insert into SQL Server.0 - Connected to database [CN_STAGE] (commit=1000)
2024/05/22 19:49:01 - CSV file input.0 - Header row skipped in file 'F:\fuentes\1976-2020-president.tab'
2024/05/22 19:49:01 - CSV file input.0 - Procesamiento finalizado (I=4288, O=0, R=0, W=4287, U=0, E=0)
2024/05/22 19:49:01 - If field value is null.0 - Procesamiento finalizado (I=0, O=0, R=4287, W=4287, U=0, E=0)
2024/05/22 19:49:01 - Select values.0 - Procesamiento finalizado (I=0, O=0, R=4287, W=4287, U=0, E=0)
2024/05/22 19:49:01 - Replace in string.0 - Procesamiento finalizado (I=0, O=0, R=4287, W=4287, U=0, E=0)
2024/05/22 19:49:03 - Insert into SQL Server.0 - Procesamiento finalizado (I=0, O=4287, R=4287, W=4287, U=0, E=0)
2024/05/22 19:49:03 - JOB_IN - Starting entry [Transformation PR2/IN_SENATE]
2024/05/22 19:49:03 - Transformation PR2/IN_SENATE - Using run configuration [Pentaho local]
2024/05/22 19:49:03 - Transformation PR2/IN_SENATE - Running transformation using the Kettle execution engine
2024/05/22 19:49:03 - IN_SENATE - Iniciado despacho de la transformaci￳n [IN_SENATE]
2024/05/22 19:49:03 - Insert into SQL Server.0 - Connected to database [CN_STAGE] (commit=1000)
2024/05/22 19:49:03 - CSV file input.0 - Header row skipped in file 'F:\fuentes\1976-2020-senate.tab'
2024/05/22 19:49:03 - CSV file input.0 - Procesamiento finalizado (I=3630, O=0, R=0, W=3629, U=0, E=0)
2024/05/22 19:49:03 - If field value is null.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:49:03 - Select values.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:49:03 - Replace in string.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:49:05 - Insert into SQL Server.0 - Procesamiento finalizado (I=0, O=3629, R=3629, W=3629, U=0, E=0)
2024/05/22 19:49:05 - JOB_IN - Starting entry [Transformation PR2/IN_SP500]
2024/05/22 19:49:05 - Transformation PR2/IN_SP500 - Using run configuration [Pentaho local]
SQLQuery9.sql - UCS...
Spoon - [PEC1] JOB_IN
ESP 19:50


Citrix Viewer Visualización Dispositivos
Mié 22 may 14:50:38
UOCLabs VDI
Spoon - [PEC1] JOB_IN
Fichero Editar View Action Tools Ayuda
PEC1
View Design
Search
General
Mail
File management
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Deprecated
Welcome! DIM_STATE JOB_CARGA_DW JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX
Execution Results
Logging History Job metrics Metrics
2024/05/22 19:49:03 - Transformation PR2/IN_SENATE - Running transformation using the Kettle execution engine
2024/05/22 19:49:03 - IN_SENATE - Iniciado despacho de la transformaci￳n [IN_SENATE]
2024/05/22 19:49:03 - Insert into SQL Server.0 - Connected to database [CN_STAGE] (commit=1000)
2024/05/22 19:49:03 - CSV file input.0 - Header row skipped in file 'F:\fuentes\1976-2020-senate.tab'
2024/05/22 19:49:03 - CSV file input.0 - Procesamiento finalizado (I=3630, O=0, R=0, W=3629, U=0, E=0)
2024/05/22 19:49:03 - If field value is null.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:49:03 - Select values.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:49:03 - Replace in string.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:49:05 - Insert into SQL Server.0 - Procesamiento finalizado (I=0, O=3629, R=3629, W=3629, U=0, E=0)
2024/05/22 19:49:05 - JOB_IN - Starting entry [Transformation PR2/IN_SP500]
2024/05/22 19:49:05 - Transformation PR2/IN_SP500 - Using run configuration [Pentaho local]
2024/05/22 19:49:05 - Transformation PR2/IN_SP500 - Running transformation using the Kettle execution engine
2024/05/22 19:49:05 - IN_SP500 - Iniciado despacho de la transformaci￳n [IN_SP500]
2024/05/22 19:49:06 - Insert into SQL Server.0 - Connected to database [CN_STAGE] (commit=1000)
2024/05/22 19:49:06 - CSV file input.0 - Header row skipped in file 'F:\fuentes\SP_500.csv'
2024/05/22 19:49:06 - CSV file input.0 - Procesamiento finalizado (I=2505, O=0, R=0, W=2504, U=0, E=0)
2024/05/22 19:49:06 - Select values.0 - Procesamiento finalizado (I=0, O=0, R=2504, W=2504, U=0, E=0)
2024/05/22 19:49:08 - Insert into SQL Server.0 - Procesamiento finalizado (I=0, O=2504, R=2504, W=2504, U=0, E=0)
2024/05/22 19:49:08 - JOB_IN - Starting entry [Transformation PR2/IN_STATE]
2024/05/22 19:49:08 - Transformation PR2/IN_STATE - Using run configuration [Pentaho local]
2024/05/22 19:49:08 - Transformation PR2/IN_STATE - Running transformation using the Kettle execution engine
2024/05/22 19:49:08 - IN_STATE - Iniciado despacho de la transformaci￳n [IN_STATE]
2024/05/22 19:49:09 - Insert into SQL Server.0 - Connected to database [CN_STAGE] (commit=1000)
2024/05/22 19:49:09 - CSV file input.0 - Header row skipped in file 'F:\fuentes\state_offices.txt'
2024/05/22 19:49:09 - CSV file input.0 - Procesamiento finalizado (I=50, O=0, R=0, W=49, U=0, E=0)
2024/05/22 19:49:09 - Select values.0 - Procesamiento finalizado (I=0, O=0, R=49, W=49, U=0, E=0)
2024/05/22 19:49:10 - Insert into SQL Server.0 - Procesamiento finalizado (I=0, O=49, R=49, W=49, U=0, E=0)
2024/05/22 19:49:10 - JOB_IN - Starting entry [Success]
2024/05/22 19:49:10 - JOB_IN - Finished job entry [Success] (result=[true])
2024/05/22 19:49:10 - JOB_IN - Finished job entry [Transformation PR2/IN_STATE] (result=[true])
2024/05/22 19:49:10 - JOB_IN - Finished job entry [Transformation PR2/IN_SP500] (result=[true])
2024/05/22 19:49:10 - JOB_IN - Finished job entry [Transformation PR2/IN_SENATE] (result=[true])
2024/05/22 19:49:10 - JOB_IN - Finished job entry [Transformation PR2/IN_PRESIDENTS] (result=[true])
2024/05/22 19:49:10 - JOB_IN - Finished job entry [Transformation PR2/IN_HOUSE] (result=[true])
2024/05/22 19:49:10 - JOB_IN - Finished job entry [SQL] (result=[true])
2024/05/22 19:49:10 - JOB_IN - Finished job entry [Set variables] (result=[true])
2024/05/22 19:49:10 - JOB_IN - Job execution finished
2024/05/22 19:49:10 - Spoon - Trabajo ha terminado.
SQLQuery9.sql - UCS...
Spoon - [PEC1] JOB_IN
ESP 19:50

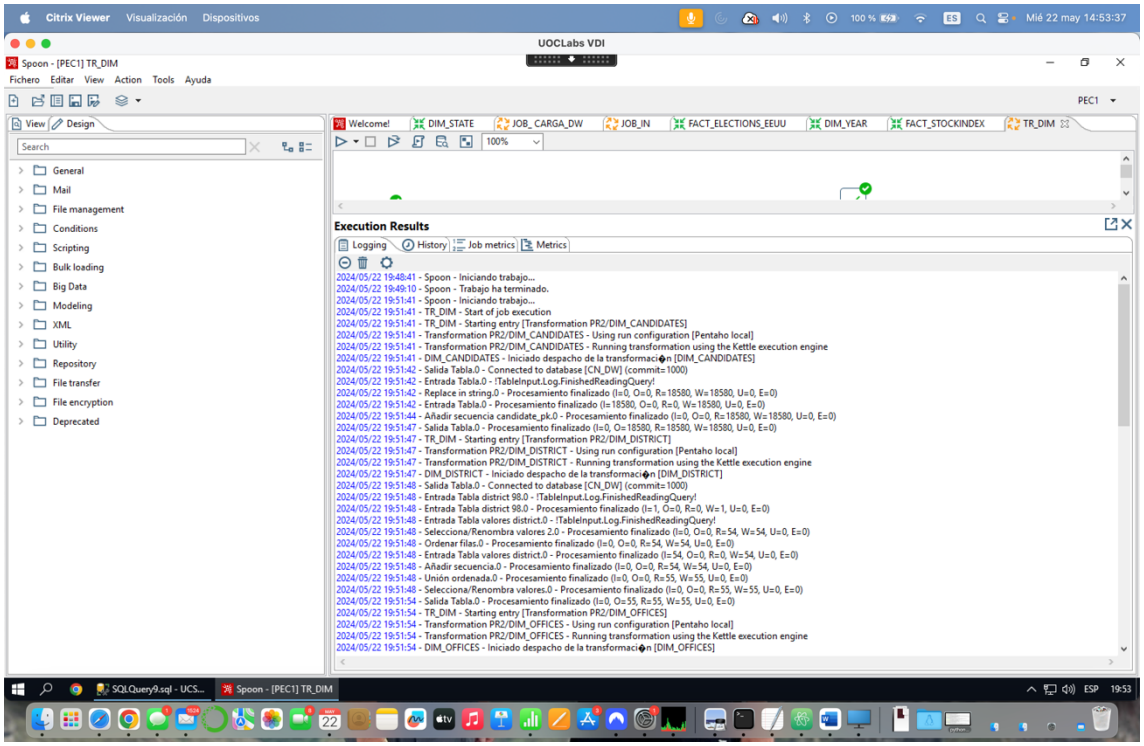

Citrix Viewer
Visualización
Dispositivos
UOCLabs VDI
Spoon - [PEC1] JOB_IN
Fichero
Editar
View
Action
Tools
Ayuda
PEC1
View
Design
Search
General
Mail
File management
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Deprecated
Welcome!
DIM_STATE
JOB_CARGA_DW
JOB_IN
FACT_ELECTIONS_EEUU
DIM_YEAR
FACT_STOCKINDEX
100%
Start
Set variables
SQL
Transformation PR2/IN_HOUSE
Transformation PR2/IN_PRESIDENTS
Transformation PR2/IN_SENATE
Transformation PR2/IN_SP500
Transformation PR2/IN_STATE
Success
SQLQuery9.sql - UCS...
Spoon - [PEC1] JOB_IN
ESP
19:50

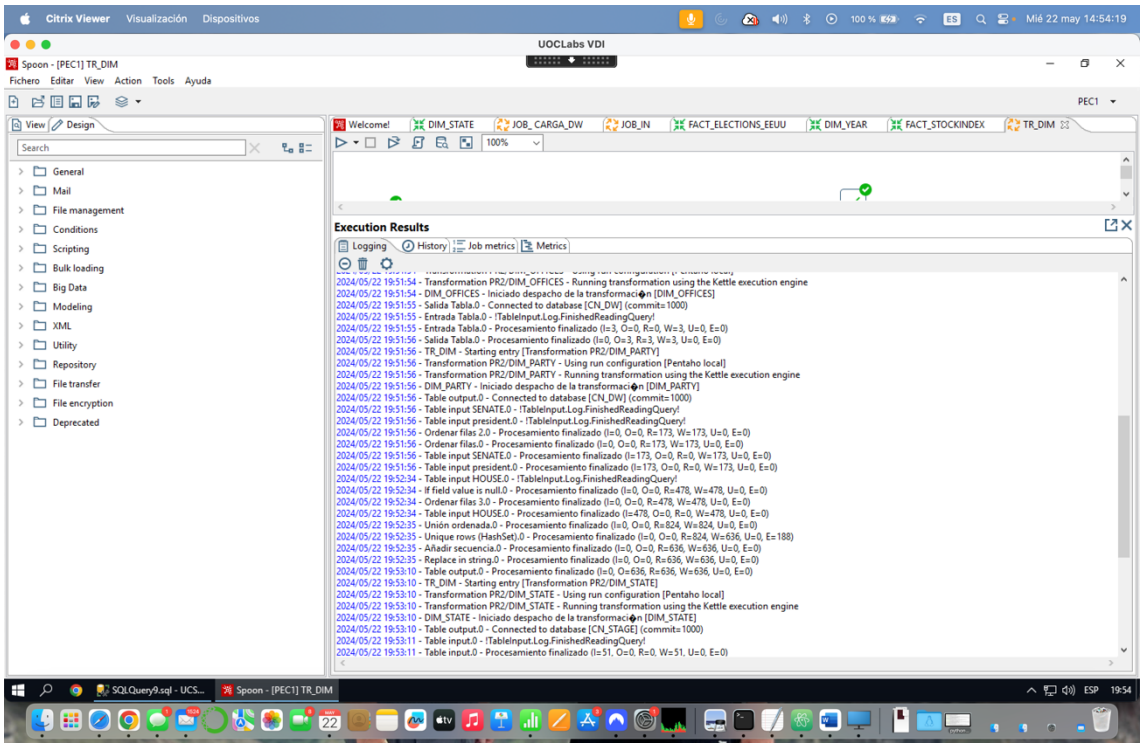
TR_DIM

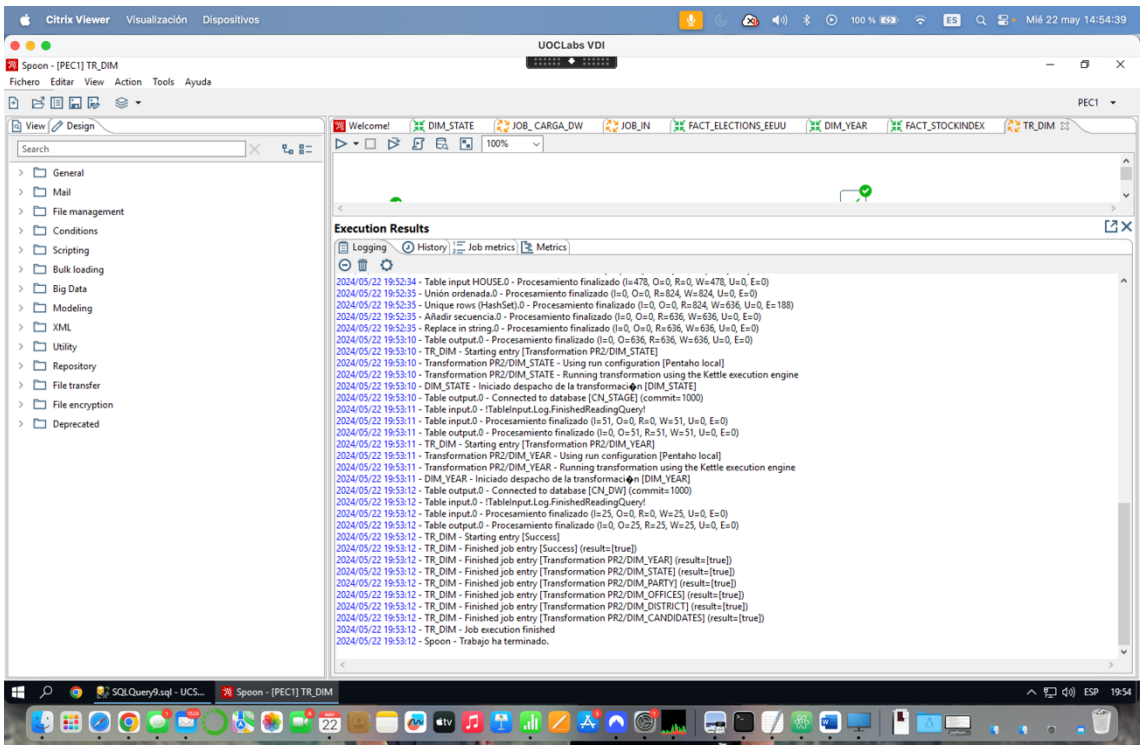
Vista general de TR_DIM



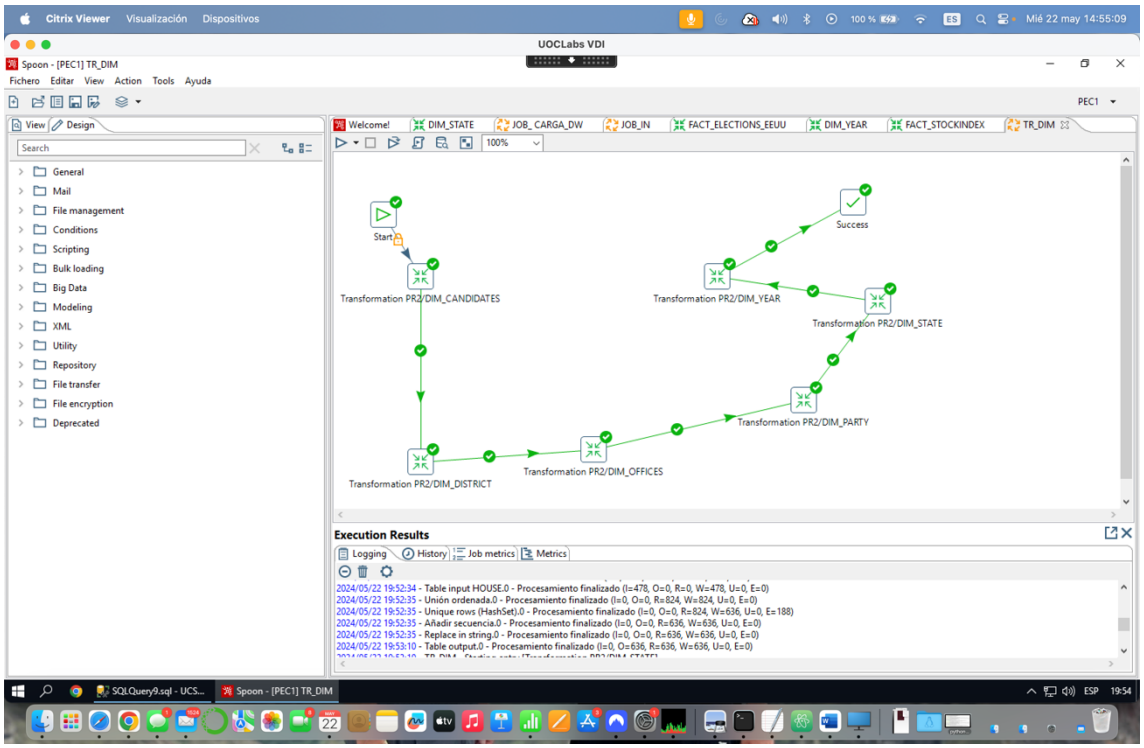
Ejecuto




Citrix Viewer Visualización Dispositivos
Mié 22 may 14:54:19
UOCLabs VDI
Spoon - [PEC1] TR_DIM
Fichero Editar View Action Tools Ayuda
PEC1
View Design
Search
General
Mail
File management
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Deprecated
Welcome! DIM_STATE JOB_CARGA_DW JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX TR_DIM
100%
Execution Results
Logging History Job metrics Metrics
2024/05/22 19:51:54 - Transformation PR2/DIM_OFFICES - Running transformation using the Kettle execution engine
2024/05/22 19:51:54 - DIM_OFFICES - Iniciado despacho de la transformaci&#65533;n [DIM_OFFICES]
2024/05/22 19:51:55 - Salida Tabla.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 19:51:55 - Entrada Tabla.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:51:55 - Entrada Tabla.0 - Procesamiento finalizado (I=3, O=0, R=0, W=3, U=0, E=0)
2024/05/22 19:51:56 - Salida Tabla.0 - Procesamiento finalizado (I=0, O=3, R=3, W=3, U=0, E=0)
2024/05/22 19:51:56 - TR_DIM - Starting entry [Transformation PR2/DIM_PARTY]
2024/05/22 19:51:56 - Transformation PR2/DIM_PARTY - Using run configuration [Pentaho local]
2024/05/22 19:51:56 - Transformation PR2/DIM_PARTY - Running transformation using the Kettle execution engine
2024/05/22 19:51:56 - DIM_PARTY - Iniciado despacho de la transformaci&#65533;n [DIM_PARTY]
2024/05/22 19:51:56 - Table output.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 19:51:56 - Table input SENATE.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:51:56 - Table input president.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:51:56 - Ordenar filas 2.0 - Procesamiento finalizado (I=0, O=0, R=173, W=173, U=0, E=0)
2024/05/22 19:51:56 - Ordenar filas.0 - Procesamiento finalizado (I=0, O=0, R=173, W=173, U=0, E=0)
2024/05/22 19:51:56 - Table input SENATE.0 - Procesamiento finalizado (I=173, O=0, R=0, W=173, U=0, E=0)
2024/05/22 19:51:56 - Table input president.0 - Procesamiento finalizado (I=173, O=0, R=0, W=173, U=0, E=0)
2024/05/22 19:52:34 - Table input HOUSE.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:52:34 - If field value is null.0 - Procesamiento finalizado (I=0, O=0, R=478, W=478, U=0, E=0)
2024/05/22 19:52:34 - Ordenar filas 3.0 - Procesamiento finalizado (I=0, O=0, R=478, W=478, U=0, E=0)
2024/05/22 19:52:34 - Table input HOUSE.0 - Procesamiento finalizado (I=478, O=0, R=0, W=478, U=0, E=0)
2024/05/22 19:52:35 - Unión ordenada.0 - Procesamiento finalizado (I=0, O=0, R=824, W=824, U=0, E=0)
2024/05/22 19:52:35 - Unique rows (HashSet).0 - Procesamiento finalizado (I=0, O=0, R=824, W=636, U=0, E=188)
2024/05/22 19:52:35 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=636, W=636, U=0, E=0)
2024/05/22 19:52:35 - Replace in string.0 - Procesamiento finalizado (I=0, O=0, R=636, W=636, U=0, E=0)
2024/05/22 19:53:10 - Table output.0 - Procesamiento finalizado (I=0, O=636, R=636, W=636, U=0, E=0)
2024/05/22 19:53:10 - TR_DIM - Starting entry [Transformation PR2/DIM_STATE]
2024/05/22 19:53:10 - Transformation PR2/DIM_STATE - Using run configuration [Pentaho local]
2024/05/22 19:53:10 - Transformation PR2/DIM_STATE - Running transformation using the Kettle execution engine
2024/05/22 19:53:10 - DIM_STATE - Iniciado despacho de la transformaci&#65533;n [DIM_STATE]
2024/05/22 19:53:10 - Table output.0 - Connected to database [CN_STAGE] (commit=1000)
2024/05/22 19:53:11 - Table input.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:53:11 - Table input.0 - Procesamiento finalizado (I=51, O=0, R=0, W=51, U=0, E=0)
SQLQuery9.sql - UCS...
Spoon - [PEC1] TR_DIM
ESP 19:54

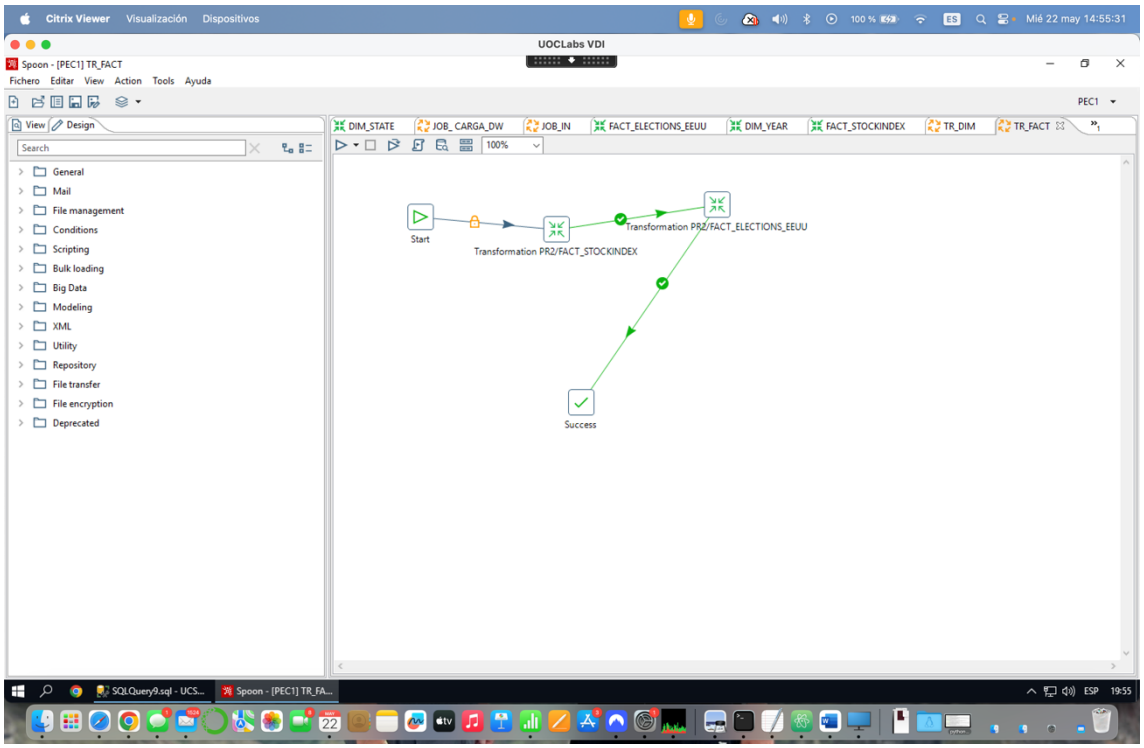

Citrix Viewer Visualización Dispositivos
Mié 22 may 14:54:39
UOCLabs VDI
Spoon - [PEC1] TR_DIM
Fichero Editar View Action Tools Ayuda
PEC1
View Design
Search
General
Mail
File management
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Deprecated
Welcome! DIM_STATE JOB_CARGA_DW JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX TR_DIM
100%
Execution Results
Logging History Job metrics Metrics
2024/05/22 19:52:34 - Table input HOUSE.0 - Procesamiento finalizado (I=478, O=0, R=0, W=478, U=0, E=0)
2024/05/22 19:52:35 - Unión ordenada.0 - Procesamiento finalizado (I=0, O=0, R=824, W=824, U=0, E=0)
2024/05/22 19:52:35 - Unique rows (HashSet).0 - Procesamiento finalizado (I=0, O=0, R=824, W=636, U=0, E=188)
2024/05/22 19:52:35 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=636, W=636, U=0, E=0)
2024/05/22 19:52:35 - Replace in string.0 - Procesamiento finalizado (I=0, O=0, R=636, W=636, U=0, E=0)
2024/05/22 19:53:10 - Table output.0 - Procesamiento finalizado (I=0, O=636, R=636, W=636, U=0, E=0)
2024/05/22 19:53:10 - TR_DIM - Starting entry [Transformation PR2/DIM_STATE]
2024/05/22 19:53:10 - Transformation PR2/DIM_STATE - Using run configuration [Pentaho local]
2024/05/22 19:53:10 - Transformation PR2/DIM_STATE - Running transformation using the Kettle execution engine
2024/05/22 19:53:10 - DIM_STATE - Iniciado despacho de la transformaci&#65533;n [DIM_STATE]
2024/05/22 19:53:10 - Table output.0 - Connected to database [CN_STAGE] (commit=1000)
2024/05/22 19:53:11 - Table input.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:53:11 - Table input.0 - Procesamiento finalizado (I=51, O=0, R=0, W=51, U=0, E=0)
2024/05/22 19:53:11 - Table output.0 - Procesamiento finalizado (I=0, O=51, R=51, W=51, U=0, E=0)
2024/05/22 19:53:11 - TR_DIM - Starting entry [Transformation PR2/DIM_YEAR]
2024/05/22 19:53:11 - Transformation PR2/DIM_YEAR - Using run configuration [Pentaho local]
2024/05/22 19:53:11 - Transformation PR2/DIM_YEAR - Running transformation using the Kettle execution engine
2024/05/22 19:53:11 - DIM_YEAR - Iniciado despacho de la transformaci&#65533;n [DIM_YEAR]
2024/05/22 19:53:12 - Table output.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 19:53:12 - Table input.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:53:12 - Table input.0 - Procesamiento finalizado (I=25, O=0, R=0, W=25, U=0, E=0)
2024/05/22 19:53:12 - Table output.0 - Procesamiento finalizado (I=0, O=25, R=25, W=25, U=0, E=0)
2024/05/22 19:53:12 - TR_DIM - Starting entry [Success]
2024/05/22 19:53:12 - TR_DIM - Finished job entry [Success] (result=[true])
2024/05/22 19:53:12 - TR_DIM - Finished job entry [Transformation PR2/DIM_YEAR] (result=[true])
2024/05/22 19:53:12 - TR_DIM - Finished job entry [Transformation PR2/DIM_STATE] (result=[true])
2024/05/22 19:53:12 - TR_DIM - Finished job entry [Transformation PR2/DIM_PARTY] (result=[true])
2024/05/22 19:53:12 - TR_DIM - Finished job entry [Transformation PR2/DIM_OFFICES] (result=[true])
2024/05/22 19:53:12 - TR_DIM - Finished job entry [Transformation PR2/DIM_DISTRICT] (result=[true])
2024/05/22 19:53:12 - TR_DIM - Finished job entry [Transformation PR2/DIM_CANDIDATES] (result=[true])
2024/05/22 19:53:12 - TR_DIM - Job execution finished
2024/05/22 19:53:12 - Spoon - Trabajo ha terminado.
SQLQuery9.sql - UCS...
Spoon - [PEC1] TR_DIM
ESP 19:54

Vista general con el trabajo realizado

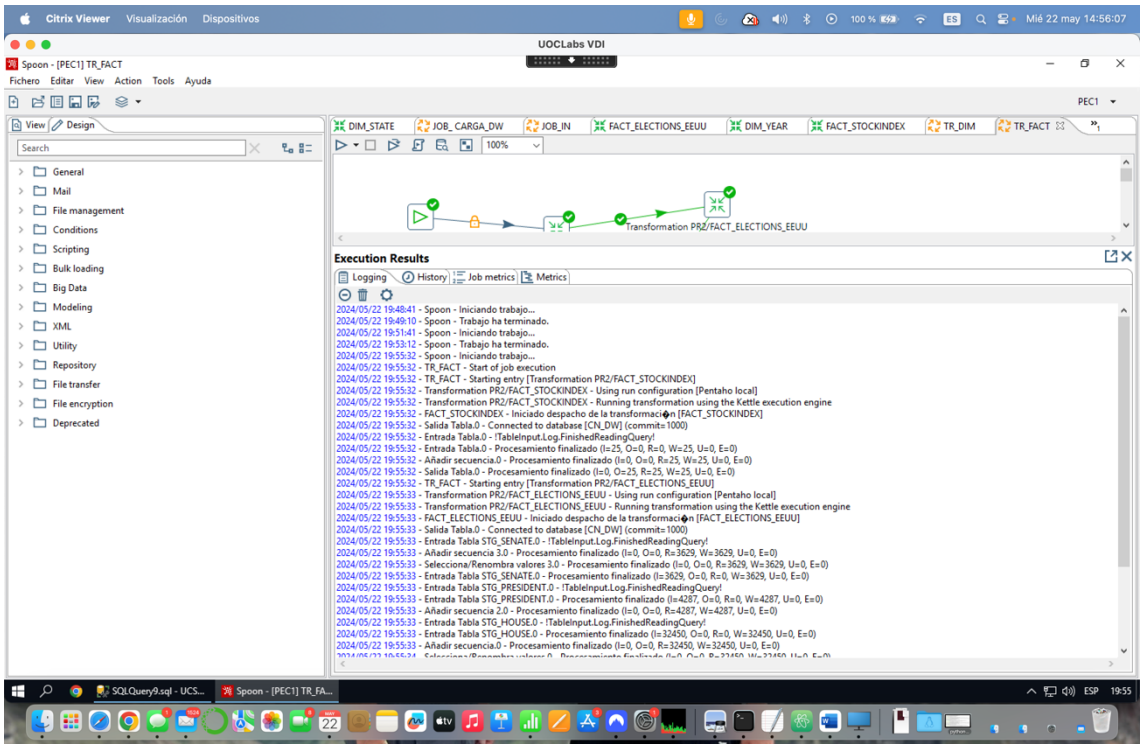


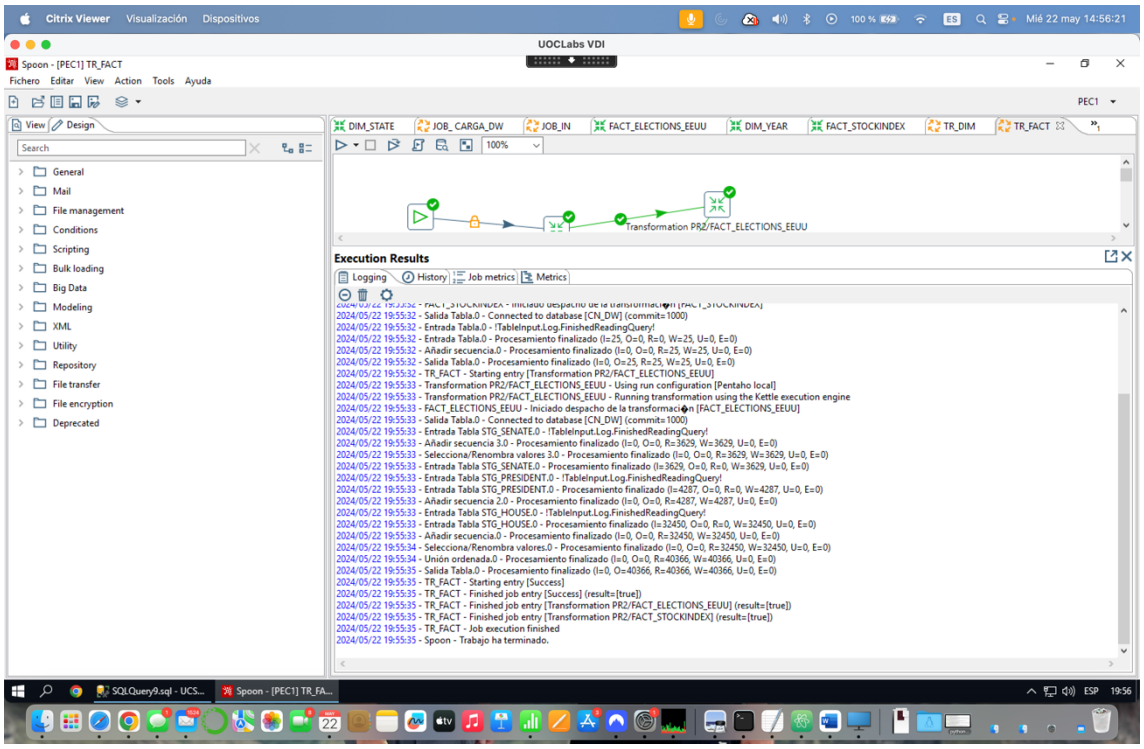
TR_FACT

Vista general

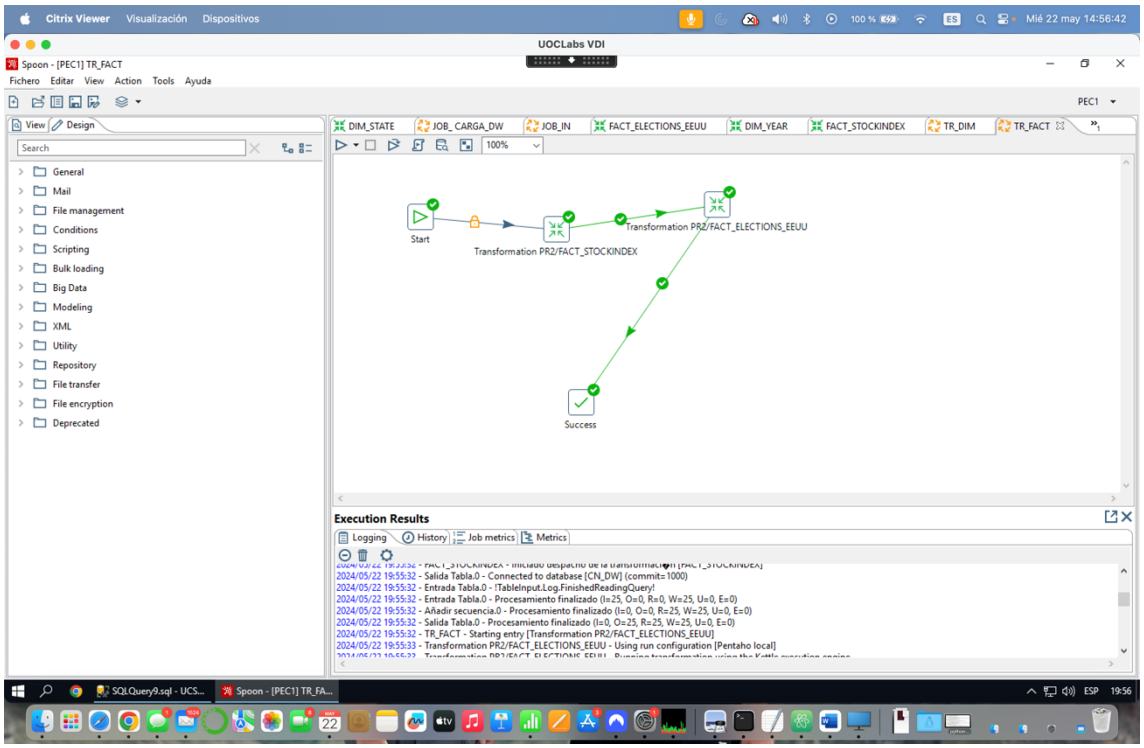


Ejecución


Citrix Viewer Visualización Dispositivos
UOCLabs VDI
Spoon - [PEC1] TR_FACT
Fichero Editar View Action Tools Ayuda
DIM_STATE JOB_CARGA_DW JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX TR_DIM TR_FACT
Transformation PR2/FACT_ELECTIONS_EEUU
Execution Results
Logging History Job metrics Metrics
2024/05/22 19:48:41 - Spoon - Iniciando trabajo...
2024/05/22 19:49:10 - Spoon - Trabajo ha terminado.
2024/05/22 19:51:41 - Spoon - Iniciando trabajo...
2024/05/22 19:53:12 - Spoon - Trabajo ha terminado.
2024/05/22 19:55:32 - Spoon - Iniciando trabajo...
2024/05/22 19:55:32 - TR_FACT - Start of job execution
2024/05/22 19:55:32 - TR_FACT - Starting entry [Transformation PR2/FACT_STOCKINDEX]
2024/05/22 19:55:32 - Transformation PR2/FACT_STOCKINDEX - Using run configuration [Pentaho local]
2024/05/22 19:55:32 - Transformation PR2/FACT_STOCKINDEX - Running transformation using the Kettle execution engine
2024/05/22 19:55:32 - FACT_STOCKINDEX - Iniciado despacho de la transformación [FACT_STOCKINDEX]
2024/05/22 19:55:32 - Salida Tabla.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 19:55:32 - Entrada Tabla.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:55:32 - Entrada Tabla.0 - Procesamiento finalizado (I=25, O=0, R=0, W=25, U=0, E=0)
2024/05/22 19:55:32 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=25, W=25, U=0, E=0)
2024/05/22 19:55:32 - Salida Tabla.0 - Procesamiento finalizado (I=0, O=25, R=25, W=25, U=0, E=0)
2024/05/22 19:55:32 - TR_FACT - Starting entry [Transformation PR2/FACT_ELECTIONS_EEUU]
2024/05/22 19:55:33 - Transformation PR2/FACT_ELECTIONS_EEUU - Using run configuration [Pentaho local]
2024/05/22 19:55:33 - Transformation PR2/FACT_ELECTIONS_EEUU - Running transformation using the Kettle execution engine
2024/05/22 19:55:33 - FACT_ELECTIONS_EEUU - Iniciado despacho de la transformación [FACT_ELECTIONS_EEUU]
2024/05/22 19:55:33 - Salida Tabla.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 19:55:33 - Entrada Tabla STG_SENATE.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:55:33 - Añadir secuencia 3.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:55:33 - Selecciona/Renombra valores 3.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:55:33 - Entrada Tabla STG_SENATE.0 - Procesamiento finalizado (I=3629, O=0, R=0, W=3629, U=0, E=0)
2024/05/22 19:55:33 - Entrada Tabla STG_PRESIDENT.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:55:33 - Entrada Tabla STG_PRESIDENT.0 - Procesamiento finalizado (I=4287, O=0, R=0, W=4287, U=0, E=0)
2024/05/22 19:55:33 - Añadir secuencia 2.0 - Procesamiento finalizado (I=0, O=0, R=4287, W=4287, U=0, E=0)
2024/05/22 19:55:33 - Entrada Tabla STG_HOUSE.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:55:33 - Entrada Tabla STG_HOUSE.0 - Procesamiento finalizado (I=32450, O=0, R=0, W=32450, U=0, E=0)
2024/05/22 19:55:33 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=32450, W=32450, U=0, E=0)

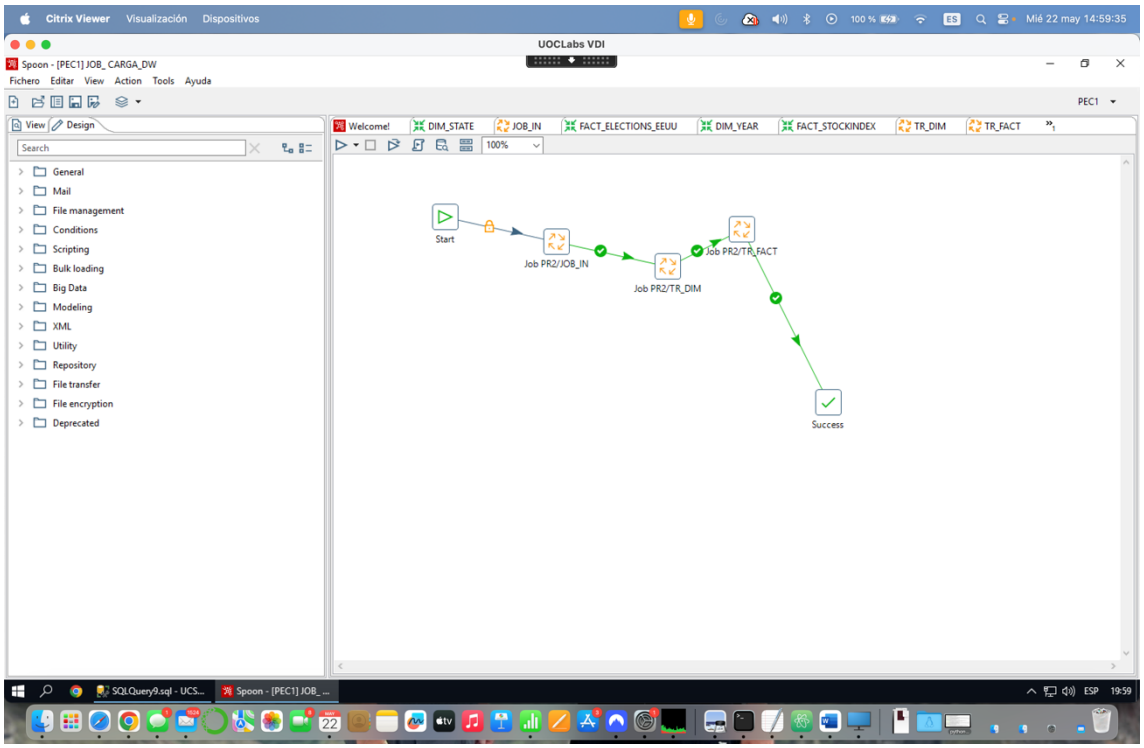

Citrix Viewer Visualización Dispositivos
UOCLabs VDI
Spoon - [PEC1] TR_FACT
Fichero Editar View Action Tools Ayuda
DIM_STATE JOB_CARGA_DW JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX TR_DIM TR_FACT
Transformation PR2/FACT_ELECTIONS_EEUU
Execution Results
Logging History Job metrics Metrics
2024/05/22 19:55:32 - Salida Tabla.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 19:55:32 - Entrada Tabla.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:55:32 - Entrada Tabla.0 - Procesamiento finalizado (I=25, O=0, R=0, W=25, U=0, E=0)
2024/05/22 19:55:32 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=25, W=25, U=0, E=0)
2024/05/22 19:55:32 - Salida Tabla.0 - Procesamiento finalizado (I=0, O=25, R=25, W=25, U=0, E=0)
2024/05/22 19:55:32 - TR_FACT - Starting entry [Transformation PR2/FACT_ELECTIONS_EEUU]
2024/05/22 19:55:33 - Transformation PR2/FACT_ELECTIONS_EEUU - Using run configuration [Pentaho local]
2024/05/22 19:55:33 - Transformation PR2/FACT_ELECTIONS_EEUU - Running transformation using the Kettle execution engine
2024/05/22 19:55:33 - FACT_ELECTIONS_EEUU - Iniciado despacho de la transformación [FACT_ELECTIONS_EEUU]
2024/05/22 19:55:33 - Salida Tabla.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 19:55:33 - Entrada Tabla STG_SENATE.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:55:33 - Añadir secuencia 3.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:55:33 - Selecciona/Renombra valores 3.0 - Procesamiento finalizado (I=0, O=0, R=3629, W=3629, U=0, E=0)
2024/05/22 19:55:33 - Entrada Tabla STG_SENATE.0 - Procesamiento finalizado (I=3629, O=0, R=0, W=3629, U=0, E=0)
2024/05/22 19:55:33 - Entrada Tabla STG_PRESIDENT.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:55:33 - Entrada Tabla STG_PRESIDENT.0 - Procesamiento finalizado (I=4287, O=0, R=0, W=4287, U=0, E=0)
2024/05/22 19:55:33 - Añadir secuencia 2.0 - Procesamiento finalizado (I=0, O=0, R=4287, W=4287, U=0, E=0)
2024/05/22 19:55:33 - Entrada Tabla STG_HOUSE.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 19:55:33 - Entrada Tabla STG_HOUSE.0 - Procesamiento finalizado (I=32450, O=0, R=0, W=32450, U=0, E=0)
2024/05/22 19:55:33 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=32450, W=32450, U=0, E=0)
2024/05/22 19:55:34 - Selecciona/Renombra valores.0 - Procesamiento finalizado (I=0, O=0, R=32450, W=32450, U=0, E=0)
2024/05/22 19:55:34 - Unión ordenada.0 - Procesamiento finalizado (I=0, O=0, R=40366, W=40366, U=0, E=0)
2024/05/22 19:55:35 - Salida Tabla.0 - Procesamiento finalizado (I=0, O=40366, R=40366, W=40366, U=0, E=0)
2024/05/22 19:55:35 - TR_FACT - Starting entry [Success]
2024/05/22 19:55:35 - TR_FACT - Finished job entry [Success] (result=[true])
2024/05/22 19:55:35 - TR_FACT - Finished job entry [Transformation PR2/FACT_ELECTIONS_EEUU] (result=[true])
2024/05/22 19:55:35 - TR_FACT - Finished job entry [Transformation PR2/FACT_STOCKINDEX] (result=[true])
2024/05/22 19:55:35 - TR_FACT - Job execution finished
2024/05/22 19:55:35 - Spoon - Trabajo ha terminado.

Vista general una vez ejecutado

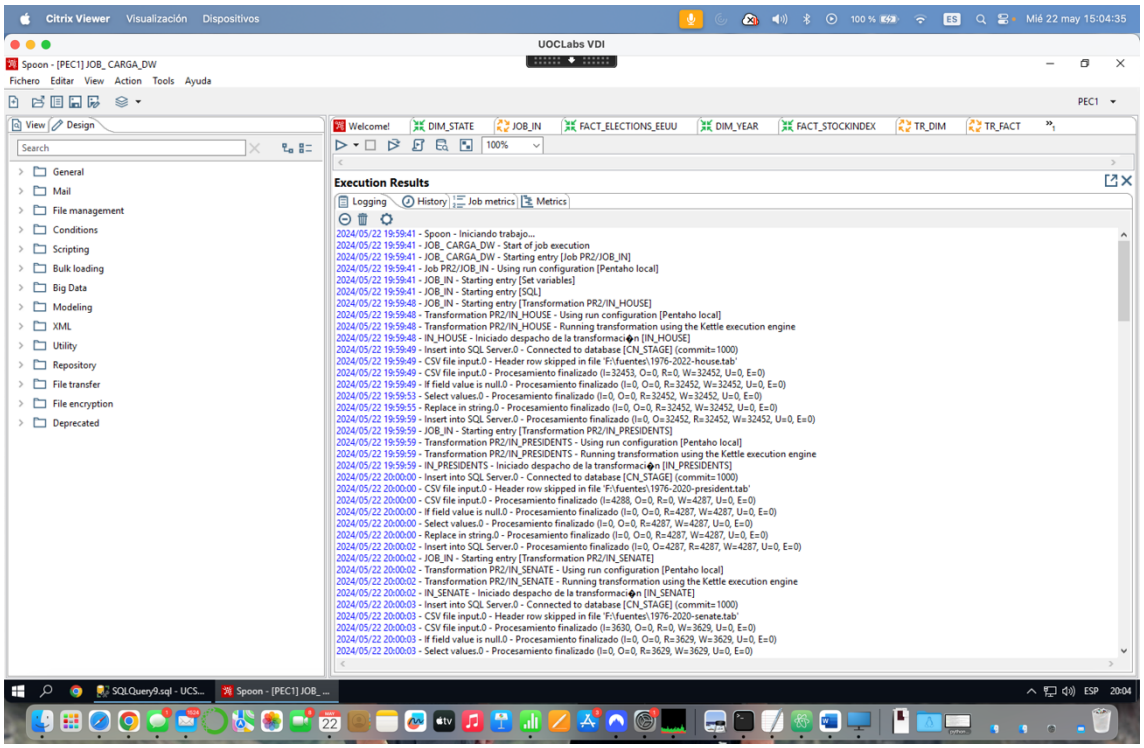


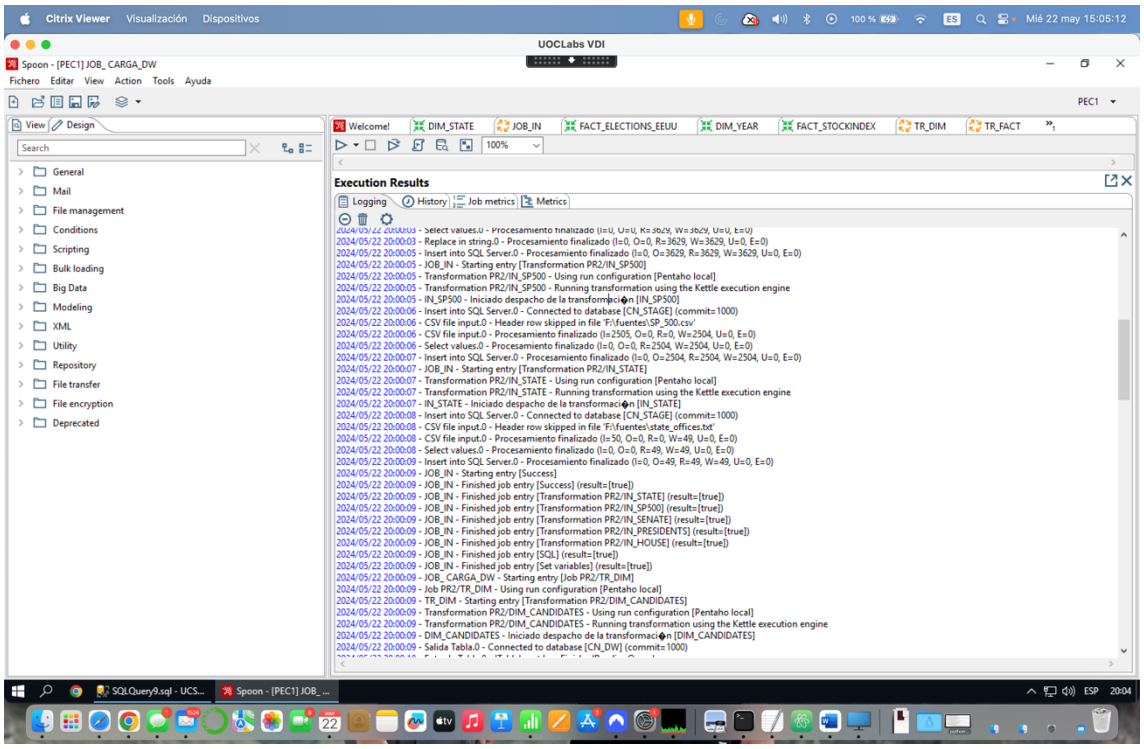
JOB_CARGA_DW

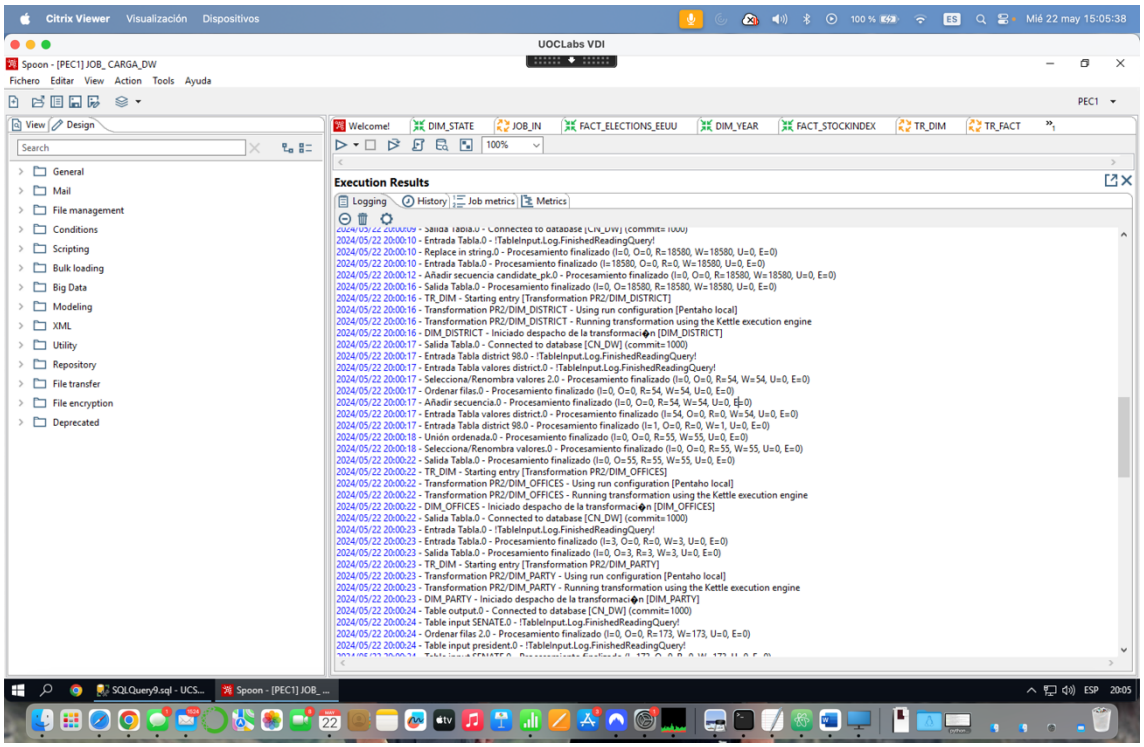
Vista general

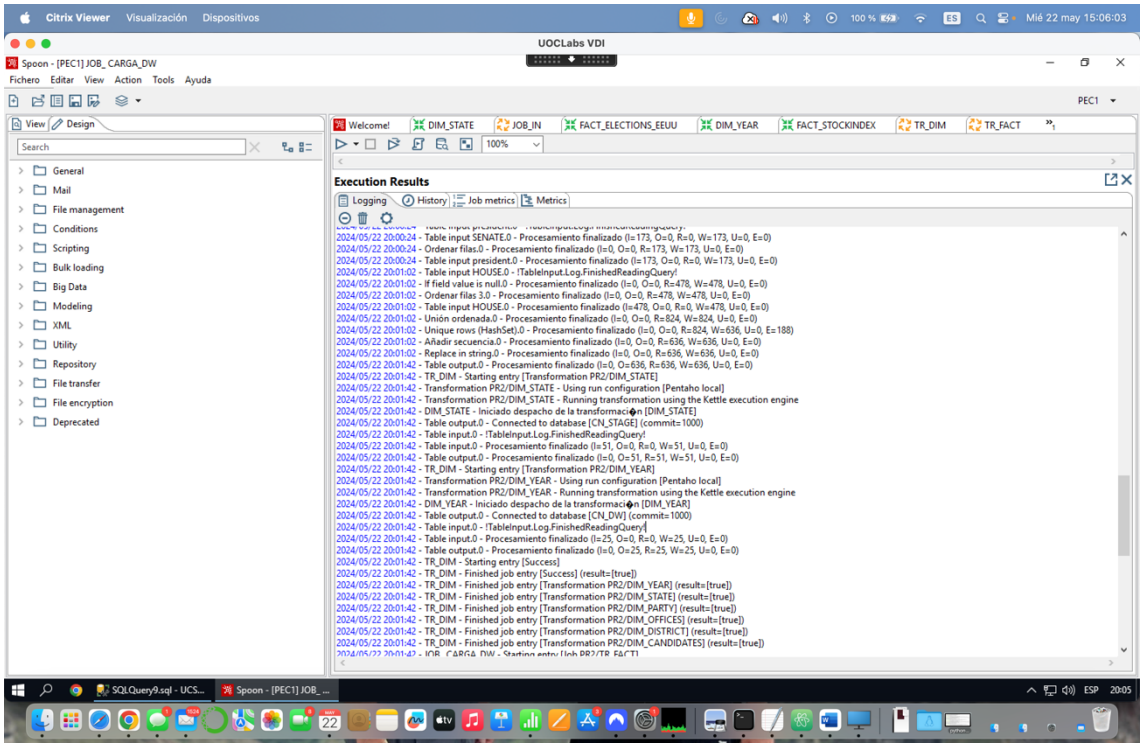


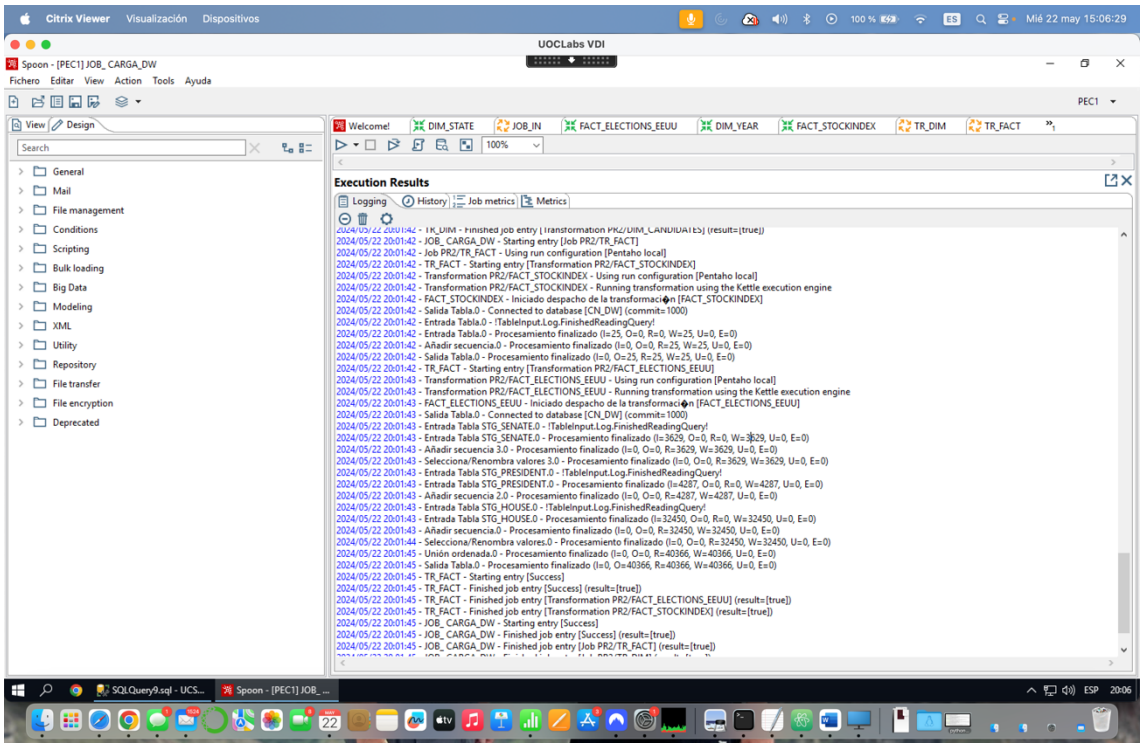
Ejecuto y muestro el registro



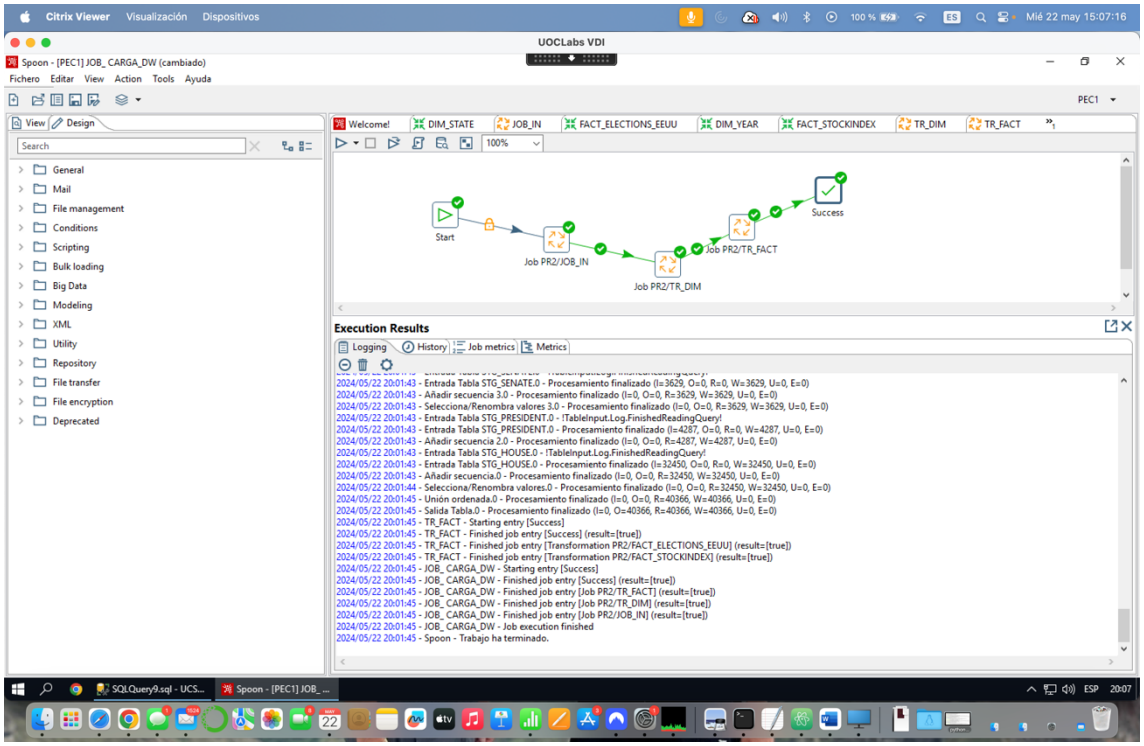



Citrix Viewer Visualización Dispositivos
Mié 22 may 15:05:38
UOCLabs VDI
Spoon - [PEC1] JOB_CARGA_DW
Fichero Editar View Action Tools Ayuda
PEC1
View Design
Search
General
Mail
File management
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Deprecated
Welcome! DIM_STATE JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX TR_DIM TR_FACT
100%
Execution Results
Logging History Job metrics Metrics
2024/05/22 20:00:10 - Entrada Tabla.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 20:00:10 - Replace in string.0 - Procesamiento finalizado (I=0, O=0, R=18580, W=18580, U=0, E=0)
2024/05/22 20:00:10 - Entrada Tabla.0 - Procesamiento finalizado (I=18580, O=0, R=0, W=18580, U=0, E=0)
2024/05/22 20:00:12 - Añadir secuencia candidate_pk.0 - Procesamiento finalizado (I=0, O=0, R=18580, W=18580, U=0, E=0)
2024/05/22 20:00:16 - Salida Tabla.0 - Procesamiento finalizado (I=0, O=18580, R=18580, W=18580, U=0, E=0)
2024/05/22 20:00:16 - TR_DIM - Starting entry [Transformation PR2/DIM_DISTRICT]
2024/05/22 20:00:16 - Transformation PR2/DIM_DISTRICT - Using run configuration [Pentaho local]
2024/05/22 20:00:16 - Transformation PR2/DIM_DISTRICT - Running transformation using the Kettle execution engine
2024/05/22 20:00:16 - DIM_DISTRICT - Iniciado despacho de la transformaci◆n [DIM_DISTRICT]
2024/05/22 20:00:17 - Salida Tabla.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 20:00:17 - Entrada Tabla district 98.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 20:00:17 - Entrada Tabla valores district.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 20:00:17 - Selecciona/Renombra valores 2.0 - Procesamiento finalizado (I=0, O=0, R=54, W=54, U=0, E=0)
2024/05/22 20:00:17 - Ordenar filas.0 - Procesamiento finalizado (I=0, O=0, R=54, W=54, U=0, E=0)
2024/05/22 20:00:17 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=54, W=54, U=0, E=0)
2024/05/22 20:00:17 - Entrada Tabla valores district.0 - Procesamiento finalizado (I=54, O=0, R=0, W=54, U=0, E=0)
2024/05/22 20:00:17 - Entrada Tabla district 98.0 - Procesamiento finalizado (I=1, O=0, R=0, W=1, U=0, E=0)
2024/05/22 20:00:18 - Unión ordenada.0 - Procesamiento finalizado (I=0, O=0, R=55, W=55, U=0, E=0)
2024/05/22 20:00:18 - Selecciona/Renombra valores.0 - Procesamiento finalizado (I=0, O=0, R=55, W=55, U=0, E=0)
2024/05/22 20:00:22 - Salida Tabla.0 - Procesamiento finalizado (I=0, O=55, R=55, W=55, U=0, E=0)
2024/05/22 20:00:22 - TR_DIM - Starting entry [Transformation PR2/DIM_OFFICES]
2024/05/22 20:00:22 - Transformation PR2/DIM_OFFICES - Using run configuration [Pentaho local]
2024/05/22 20:00:22 - Transformation PR2/DIM_OFFICES - Running transformation using the Kettle execution engine
2024/05/22 20:00:22 - DIM_OFFICES - Iniciado despacho de la transformaci◆n [DIM_OFFICES]
2024/05/22 20:00:22 - Salida Tabla.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 20:00:23 - Entrada Tabla.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 20:00:23 - Entrada Tabla.0 - Procesamiento finalizado (I=3, O=0, R=0, W=3, U=0, E=0)
2024/05/22 20:00:23 - Salida Tabla.0 - Procesamiento finalizado (I=0, O=3, R=3, W=3, U=0, E=0)
2024/05/22 20:00:23 - TR_DIM - Starting entry [Transformation PR2/DIM_PARTY]
2024/05/22 20:00:23 - Transformation PR2/DIM_PARTY - Using run configuration [Pentaho local]
2024/05/22 20:00:23 - Transformation PR2/DIM_PARTY - Running transformation using the Kettle execution engine
2024/05/22 20:00:23 - DIM_PARTY - Iniciado despacho de la transformaci◆n [DIM_PARTY]
2024/05/22 20:00:24 - Table output.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 20:00:24 - Table input SENATE.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 20:00:24 - Ordenar filas 2.0 - Procesamiento finalizado (I=0, O=0, R=173, W=173, U=0, E=0)
2024/05/22 20:00:24 - Table input president.0 - !TableInput.Log.FinishedReadingQuery!
SQLQuery9.sql - UCS...
Spoon - [PEC1] JOB_...
ESP 20:05


Citrix Viewer Visualización Dispositivos
Mié 22 may 15:06:03
UOCLabs VDI
Spoon - [PEC1] JOB_CARGA_DW
Fichero Editar View Action Tools Ayuda
PEC1
View Design
Search
General
Mail
File management
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Deprecated
Welcome! DIM_STATE JOB_IN FACT_ELECTIONS_EEUU DIM_YEAR FACT_STOCKINDEX TR_DIM TR_FACT
100%
Execution Results
Logging History Job metrics Metrics
2024/05/22 20:00:24 - Table input SENATE.0 - Procesamiento finalizado (I=173, O=0, R=0, W=173, U=0, E=0)
2024/05/22 20:00:24 - Ordenar filas.0 - Procesamiento finalizado (I=0, O=0, R=173, W=173, U=0, E=0)
2024/05/22 20:00:24 - Table input president.0 - Procesamiento finalizado (I=173, O=0, R=0, W=173, U=0, E=0)
2024/05/22 20:01:02 - Table input HOUSE.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 20:01:02 - If field value is null.0 - Procesamiento finalizado (I=0, O=0, R=478, W=478, U=0, E=0)
2024/05/22 20:01:02 - Ordenar filas 3.0 - Procesamiento finalizado (I=0, O=0, R=478, W=478, U=0, E=0)
2024/05/22 20:01:02 - Table input HOUSE.0 - Procesamiento finalizado (I=478, O=0, R=0, W=478, U=0, E=0)
2024/05/22 20:01:02 - Unión ordenada.0 - Procesamiento finalizado (I=0, O=0, R=824, W=824, U=0, E=0)
2024/05/22 20:01:02 - Unique rows (HashSet).0 - Procesamiento finalizado (I=0, O=0, R=824, W=636, U=0, E=188)
2024/05/22 20:01:02 - Añadir secuencia.0 - Procesamiento finalizado (I=0, O=0, R=636, W=636, U=0, E=0)
2024/05/22 20:01:02 - Replace in string.0 - Procesamiento finalizado (I=0, O=0, R=636, W=636, U=0, E=0)
2024/05/22 20:01:42 - Table output.0 - Procesamiento finalizado (I=0, O=636, R=636, W=636, U=0, E=0)
2024/05/22 20:01:42 - TR_DIM - Starting entry [Transformation PR2/DIM_STATE]
2024/05/22 20:01:42 - Transformation PR2/DIM_STATE - Using run configuration [Pentaho local]
2024/05/22 20:01:42 - Transformation PR2/DIM_STATE - Running transformation using the Kettle execution engine
2024/05/22 20:01:42 - DIM_STATE - Iniciado despacho de la transformaci◆n [DIM_STATE]
2024/05/22 20:01:42 - Table output.0 - Connected to database [CN_STAGE] (commit=1000)
2024/05/22 20:01:42 - Table input.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 20:01:42 - Table input.0 - Procesamiento finalizado (I=51, O=0, R=0, W=51, U=0, E=0)
2024/05/22 20:01:42 - Table output.0 - Procesamiento finalizado (I=0, O=51, R=51, W=51, U=0, E=0)
2024/05/22 20:01:42 - TR_DIM - Starting entry [Transformation PR2/DIM_YEAR]
2024/05/22 20:01:42 - Transformation PR2/DIM_YEAR - Using run configuration [Pentaho local]
2024/05/22 20:01:42 - Transformation PR2/DIM_YEAR - Running transformation using the Kettle execution engine
2024/05/22 20:01:42 - DIM_YEAR - Iniciado despacho de la transformaci◆n [DIM_YEAR]
2024/05/22 20:01:42 - Table output.0 - Connected to database [CN_DW] (commit=1000)
2024/05/22 20:01:42 - Table input.0 - !TableInput.Log.FinishedReadingQuery!
2024/05/22 20:01:42 - Table input.0 - Procesamiento finalizado (I=25, O=0, R=0, W=25, U=0, E=0)
2024/05/22 20:01:42 - Table output.0 - Procesamiento finalizado (I=0, O=25, R=25, W=25, U=0, E=0)
2024/05/22 20:01:42 - TR_DIM - Starting entry [Success]
2024/05/22 20:01:42 - TR_DIM - Finished job entry [Success] (result=[true])
2024/05/22 20:01:42 - TR_DIM - Finished job entry [Transformation PR2/DIM_YEAR] (result=[true])
2024/05/22 20:01:42 - TR_DIM - Finished job entry [Transformation PR2/DIM_STATE] (result=[true])
2024/05/22 20:01:42 - TR_DIM - Finished job entry [Transformation PR2/DIM_PARTY] (result=[true])
2024/05/22 20:01:42 - TR_DIM - Finished job entry [Transformation PR2/DIM_OFFICES] (result=[true])
2024/05/22 20:01:42 - TR_DIM - Finished job entry [Transformation PR2/DIM_DISTRICT] (result=[true])
2024/05/22 20:01:42 - TR_DIM - Finished job entry [Transformation PR2/DIM_CANDIDATES] (result=[true])
SQLQuery9.sql - UCS...
Spoon - [PEC1] JOB_...
ESP 20:05



Ultima captura, con el fin del registro y la visualización de trabajo terminado



CONSIDERACIONES JOBS

La carga de las transformaciones se ha hecho por orden, de STG IN FACT.
Ya que no podríamos introducir datos en fact si antes no están introducidos en IN, e igual para IN y STG. La creación de trabajos no lleva en si complejidad, simplemente elegir bien el orden según las restricciones de los datos.