

JUAN LUIS ACEBAL RICO

PR3 BDA

INDICE

Introducción..... 4

Pasos previos..... 4

Script ejecutado5

Consulta.....6

Creacion de proyecto 6

Nombre, ubicación y nombre de la solución7

Asistente de importación.....7

Nombre de servidor8

Proceso de importación.....8

Comprobación de orígenes de datos9

Herencia9

Configuración conexión de datos proyecto 10

Cambio datos conexión destino..... 10

Creación vistas..... 11

Ubicación de creación de vista 11

Asistente creación de vista 11

Origen de datos..... 12

Vista sp500 12

 Nombre vista..... 13

Vista elecciones 13

 Nombre 14

Creación de cubos..... 14

Asistente 15

Seleccionar método de creación 15

CUBE_STOCKINDEX..... 16

 Seleccionar medidas..... 16

 Seleccionar dimensiones 17

 Nombre 17

 Vista de diseño del cubo 18

Procesamos..... 18

 Ejecutar 20

 Resultado..... 20

 Vista explorador..... 21

CUBE_ELECTIONS_EEUU 21

 Procesar..... 25

 Vista final de una consulta 26

Eliminar dimensión duplicada 27

 Agregar dimensión..... 28

 Agregar dimensión..... 28

Definir relación	29
Procesar.....	30
Configuración de dimensiones.	32
DIM_DISTRICT	32
DIM_STATE	33
DIM_PARTY.....	34
DIM_CANDIDATE	36
Resultado final	37
DIM_OFFICE	37
DIM_YEAR.....	38
Realización de consultas en Power Bi	40
1.Evolución del índice SP500 a lo largo del tiempo	41
2.Evolutivo de votos del partido “DEMOCRAT” a lo largo del tiempo para las elecciones a la Presidencia de los EEUU. El resultado se mostrará ordenado por años de manera descendente	42
3.Top Ten de candidatos más votados en las elecciones presidenciales y elección.....	43
4.Comparación de los resultados de las elecciones de 2016 y 2020 en los estados de MICHIGAN, PENNSYLVANIA Y WISCONSIN para los candidatos de los partidos demócratas “DEMOCRAT” y republicano “REPUBLICAN”. Incluye el % sobre el total de votos determinantes para la victoria de Joe Biden en 2020	44
5.TOP Five de años con mayor participación electoral en la presidenciales por el estado de California	47
6.Diferencia de votos entre las elecciones de 2016 y 2020, por estado, de Donald Trump a la presidencia de los EEUU.....	49
7.Compara por estado los resultados de “CLINTON, HILLARY” y “TRUMP, DONALD J.” para las elecciones a la presidencia de	50
2016.....	50

INTRODUCCIÓN

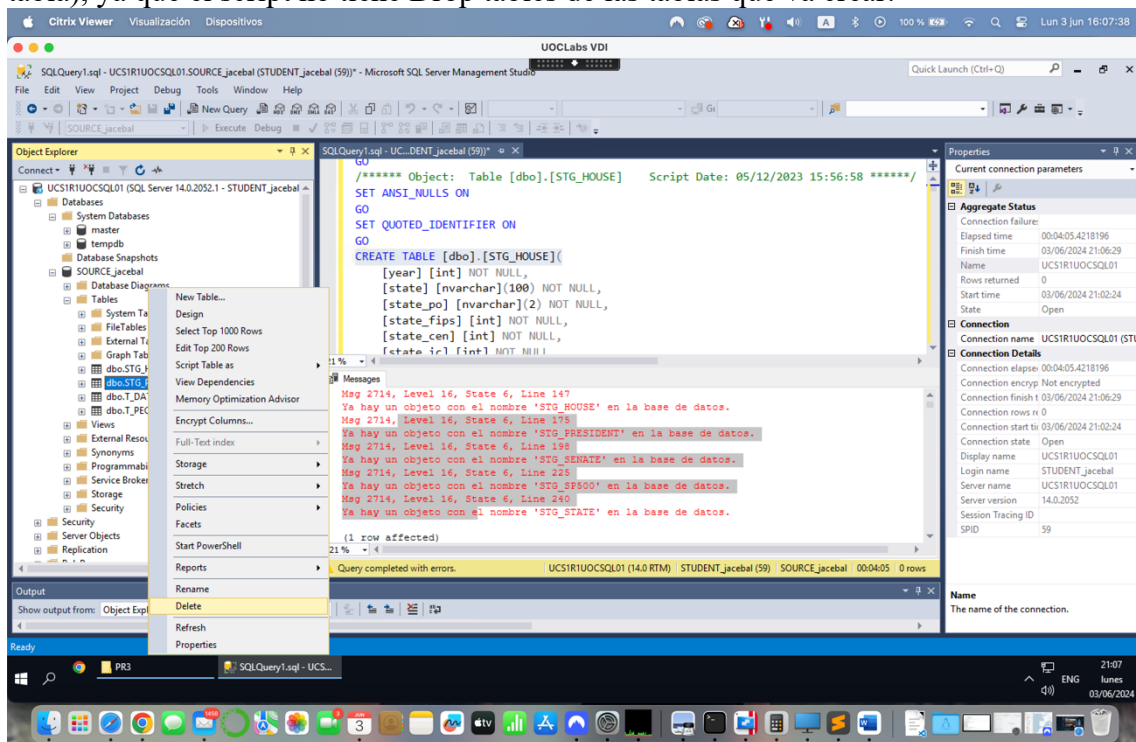
Este documento trata de implementar un modelo OLAP de Data Warehouse para el análisis multidimensional de las elecciones de EE. UU. El objetivo principal es después del diseño ya hecho, la correcta implantación de este, usando para ello como origen de datos y de tablas la creación de las dimensiones y hechos, hecha en la actividad precedente. A partir del script donado, que viene a ser la mejor implantación o más correcta de la actividad precedente, vamos a desarrollar nuestro almacén de datos, donde generaremos unas vistas de los datos, pudiendo analizar los resultados de las elecciones junto al comportamiento del índice bursátil SP500.

Como primer paso, será la creación del proyecto en Visual Studio, una vez implantado, junto con importación de datos, creación de cubos y dimensiones, hare las consultas planteadas como objetivo de la implantación.

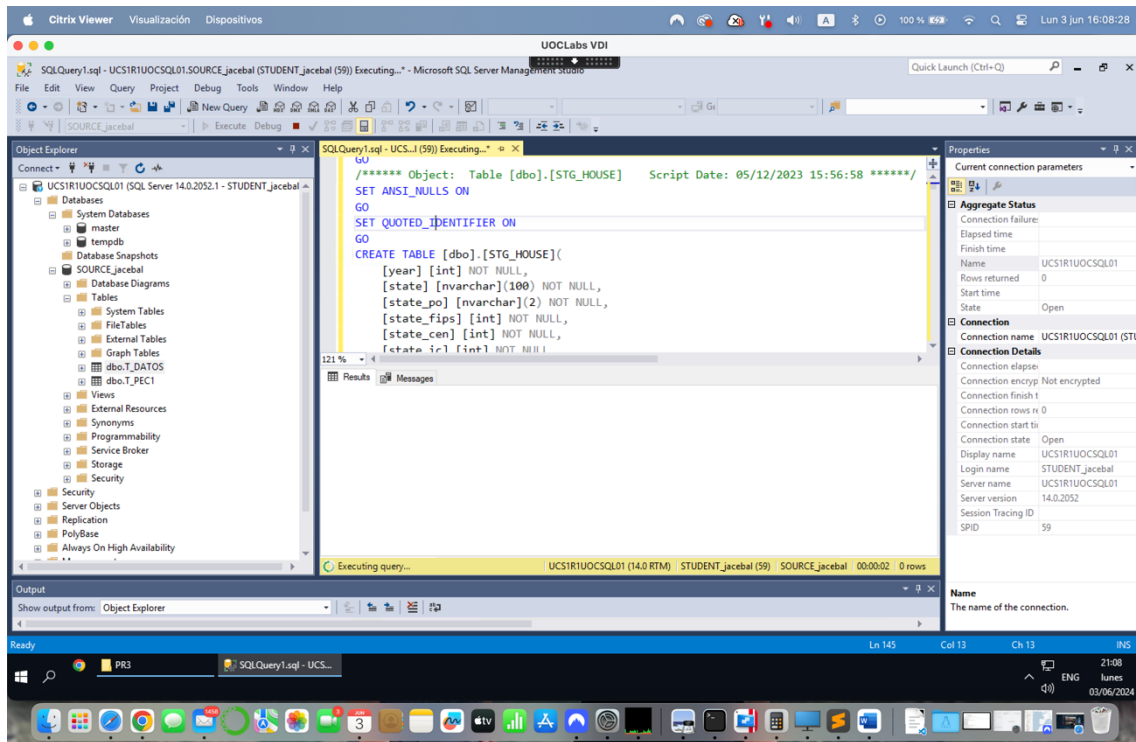
PASOS PREVIOS

Primero voy a homogeneizar los datos, tablas y relaciones, ya que, aunque mi propuesta de solución de la actividad precedente pueda ser muy parecida a la propuesta, no va a ser igual y para evitar llegar a conclusiones o datos diferentes, eliminaré todas mis tablas y cargare el script.

Elimino las antiguas tablas STG, DIM y FACT (lo he hecho con botón derecho en cada tabla), ya que el script no tiene Drop tables de las tablas que va crear:

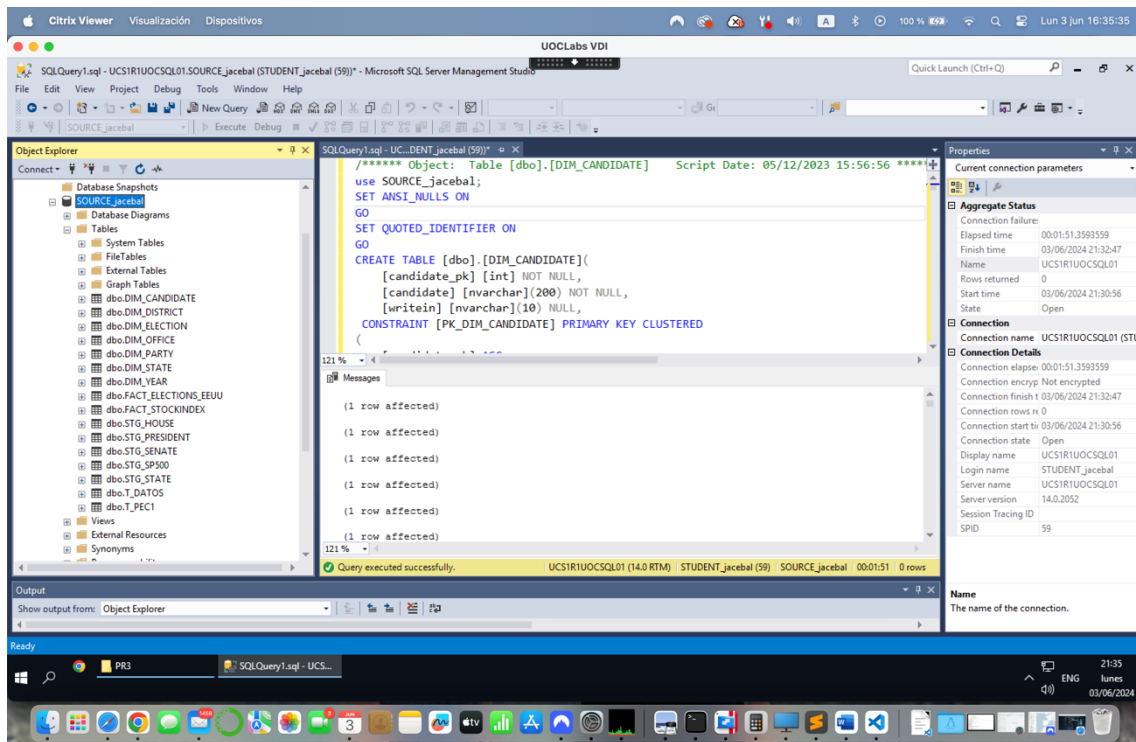


Ejecuto:



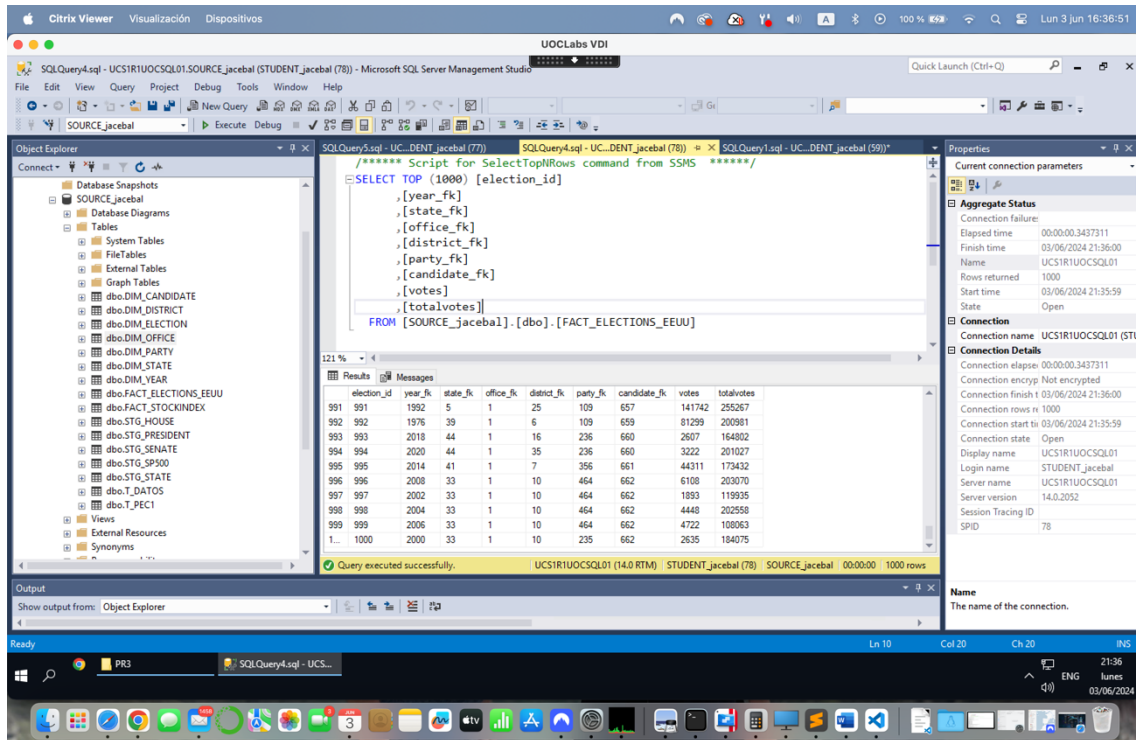
SCRIPT EJECUTADO

Ejecuto el script, como ya he dicho, los datos y su estructura deberían ser la misma, pero para evitar problemas en este proyecto, uso el script con estructuras y datos propuestos como solución.



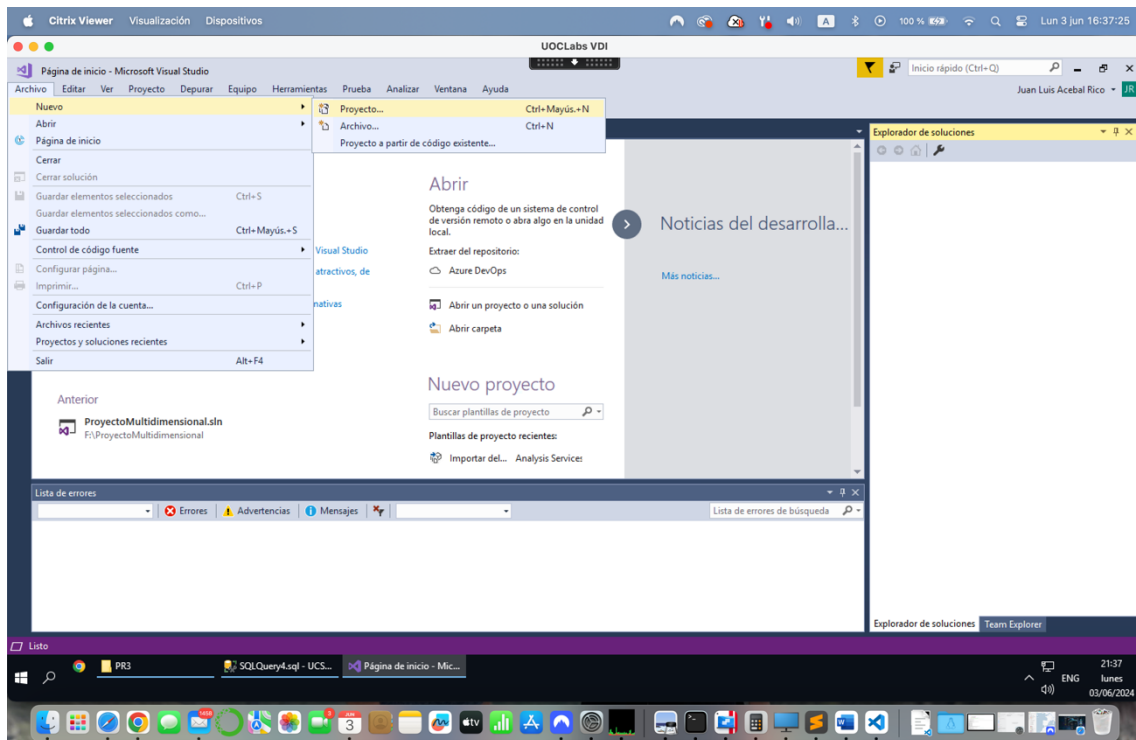
CONSULTA

Compruebo que los datos están en la zona transformada (FACT y DIM)



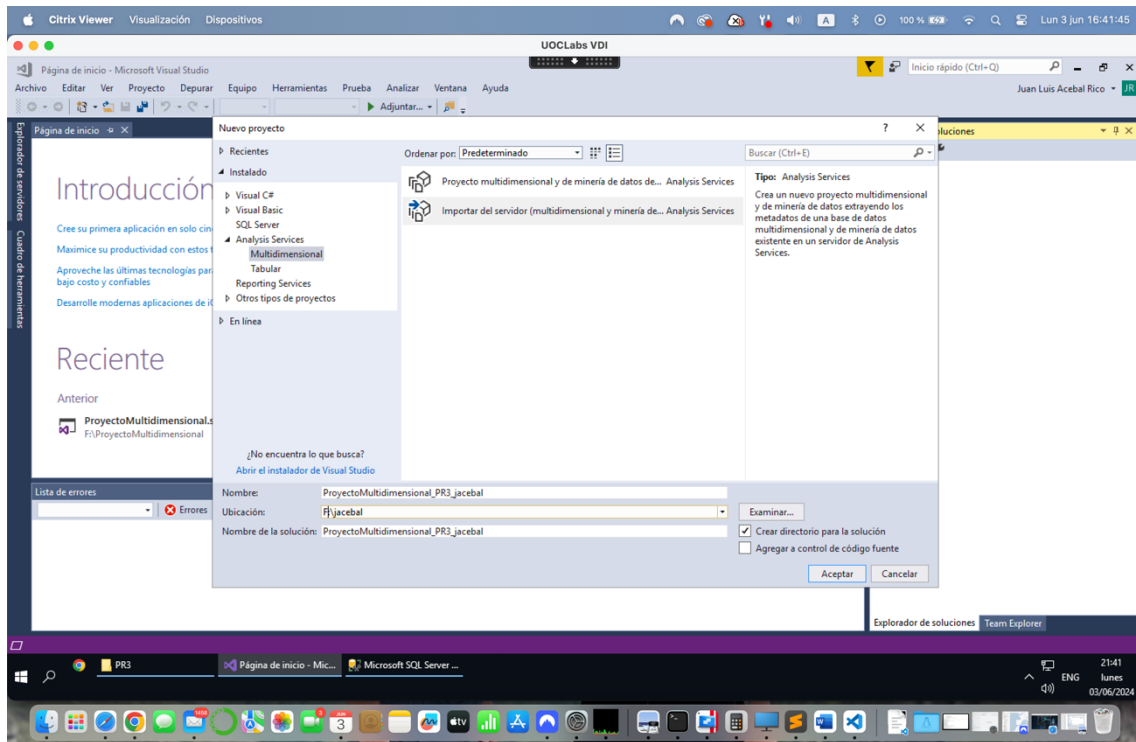
CREACION DE PROYECTO

Creo un Proyecto en MS Visual studio



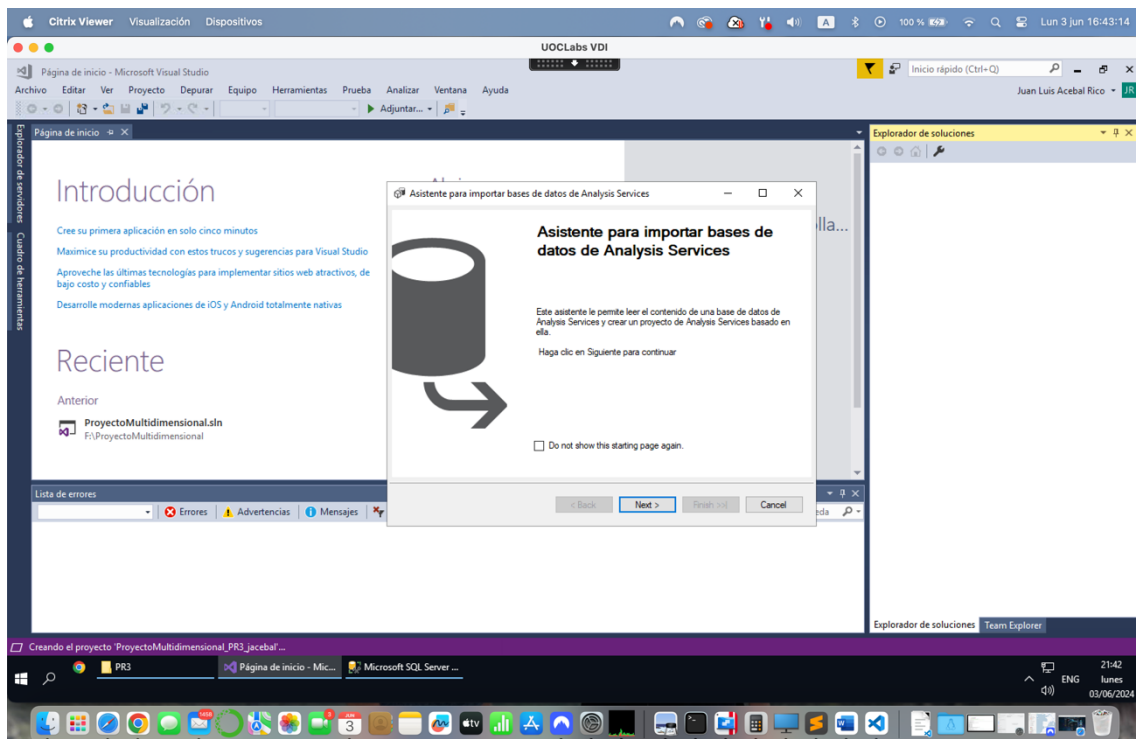
NOMBRE, UBICACIÓN Y NOMBRE DE LA SOLUCION

Selecciono dentro de Analysis Services, importar del servidor



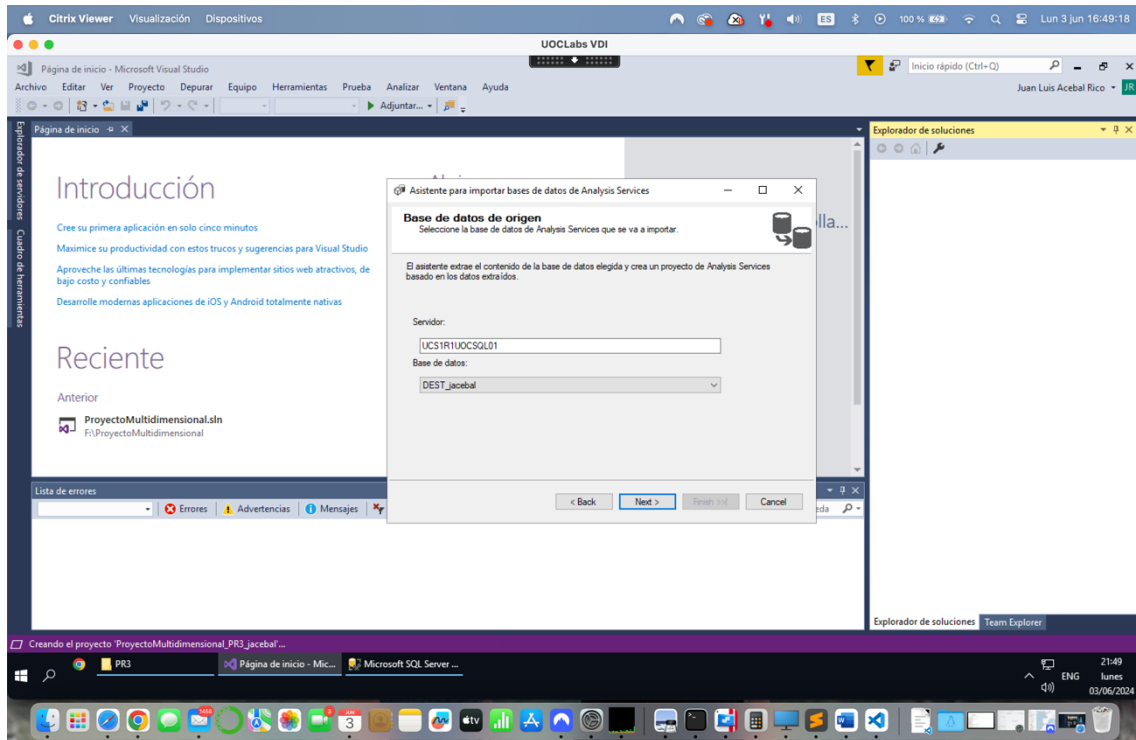
ASISTENTE DE IMPORTACIÓN

Empiezo el asistente de importación para que nuestros datos del modelo relacional se usen en el modelo dimensional



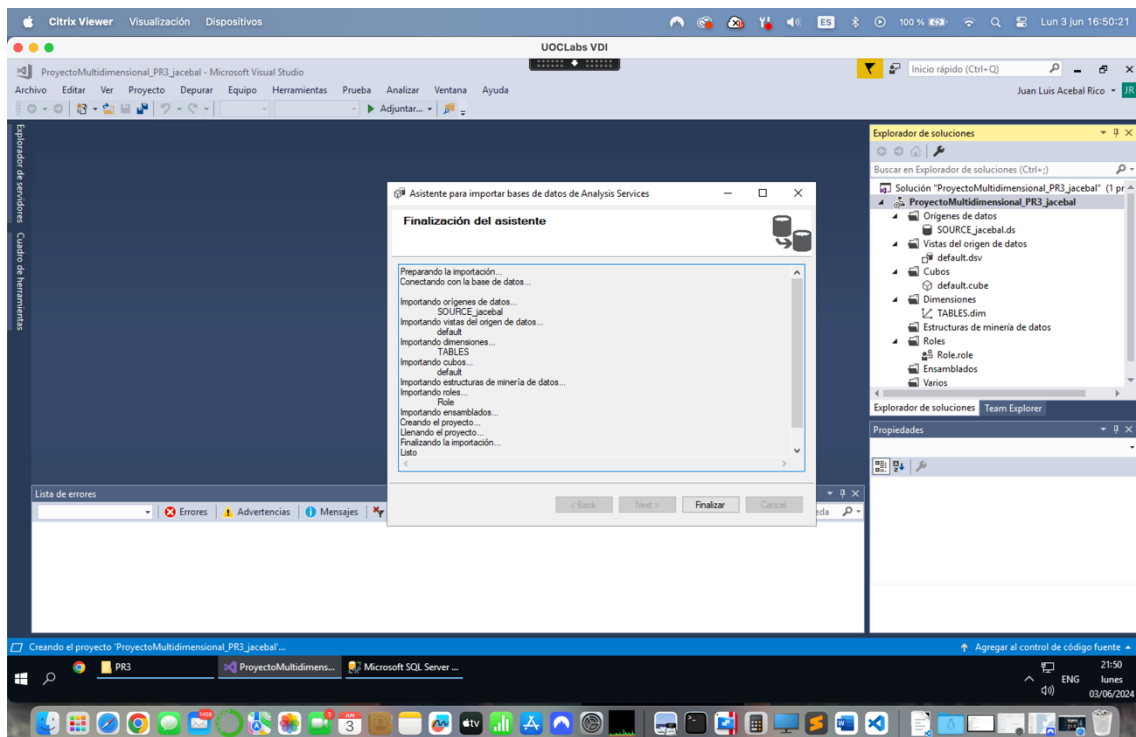
NOMBRE DE SERVIDOR

Escribo el nombre de servidor, en este caso UCS1R1UOCSQL01 y selecciono la base de datos de destino:



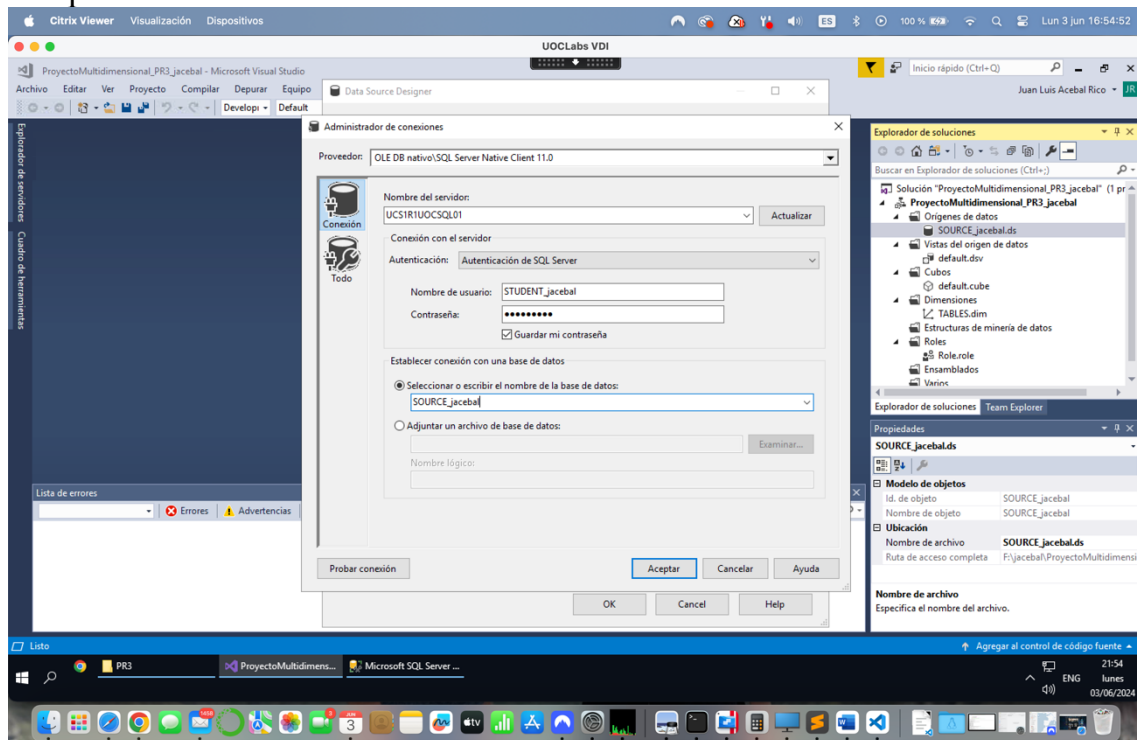
PROCESO DE IMPORTACIÓN

Este paso sirve para, importar de un modelo relacional a un modelo dimensional, usando el origen de datos del modelo relacional, para posteriormente configurarlo y usarlo en el modelo dimensional.



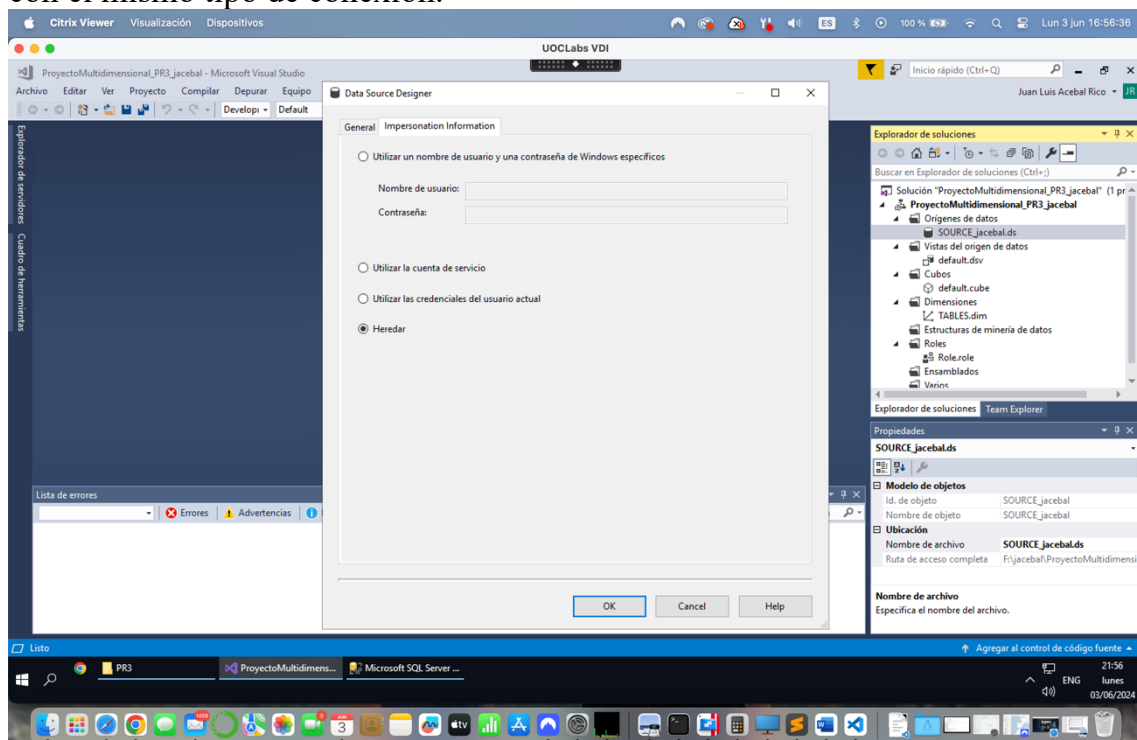
COMPROBACIÓN DE ORÍGENES DE DATOS

Comprobamos que, el usuario, contraseña funciona, hago test de conexión, etc, esto es importante ya que después puedo tener errores que no sea capaz de comprender bien de dónde vienen.



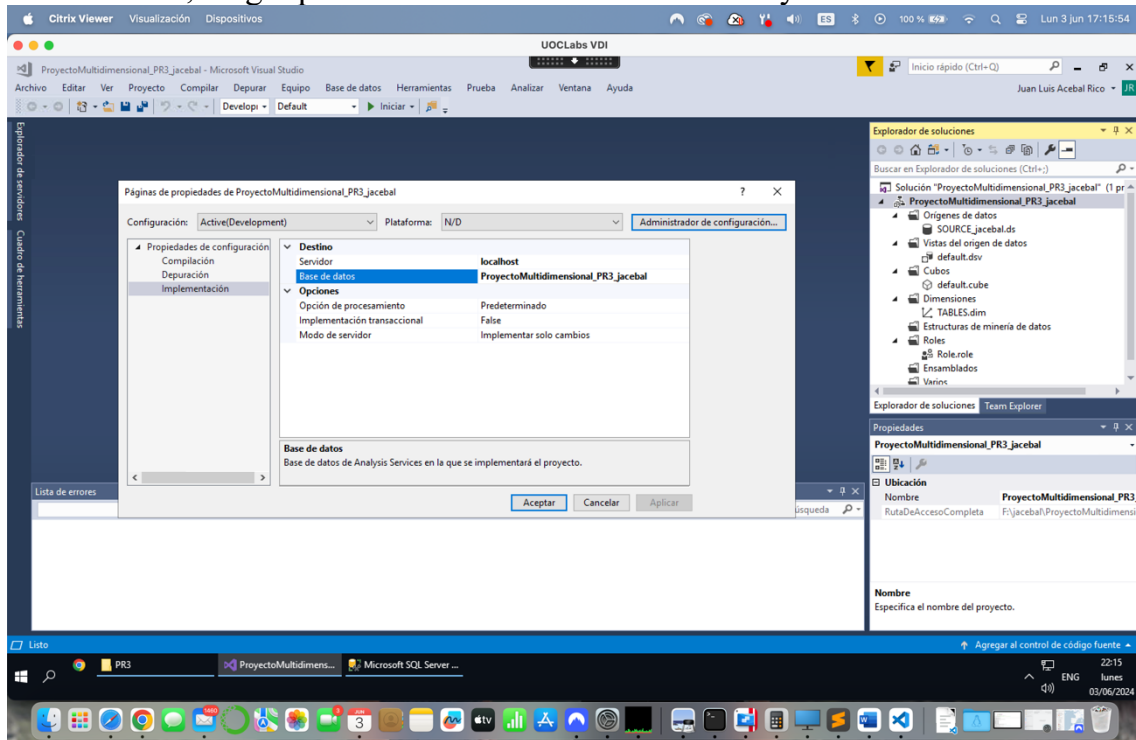
HERENCIA

Compruebo que heredar está activo para que herede la configuración de conexión existente en MS visual studio, así evitamos que se use un nuevo tipo de conexión que no va a funcionar, ya que solo usaremos el servidor UCS1R1UOCSQL01 con el mismo tipo de conexión.



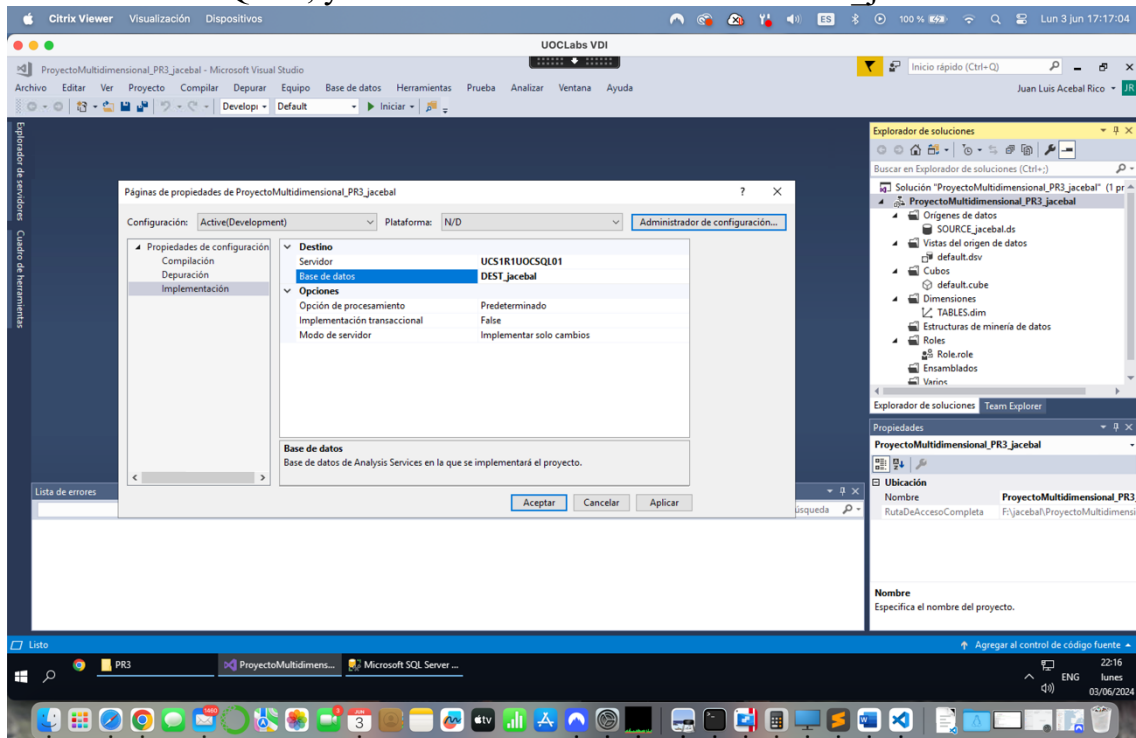
CONFIGURACIÓN CONEXIÓN DE DATOS PROYECTO

Ahora compruebo que los datos de la conexión son los correctos, que en este caso no lo es, tengo que cambiar el nombre de servidor y la base de datos



CAMBIO DATOS CONEXIÓN DESTINO

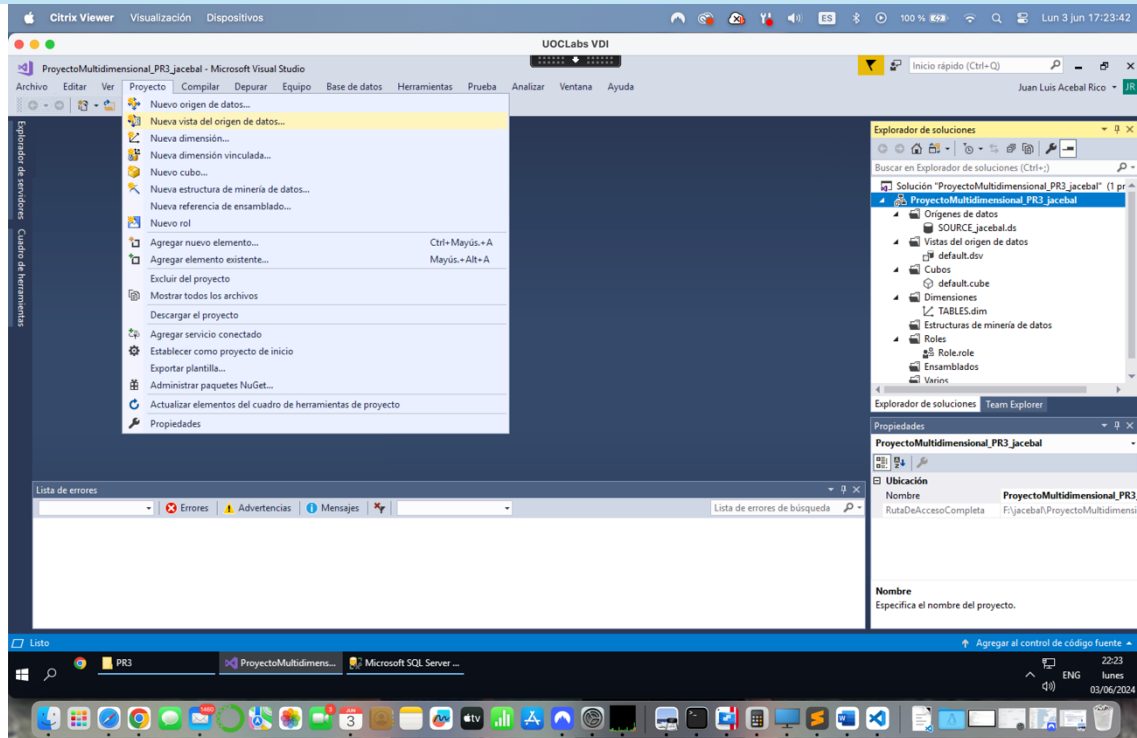
Esto es así ya que, la base de datos está en otro servidor llamado UCS1R1UOCSQL01, y la base de datos de destino es DEST_jacebal



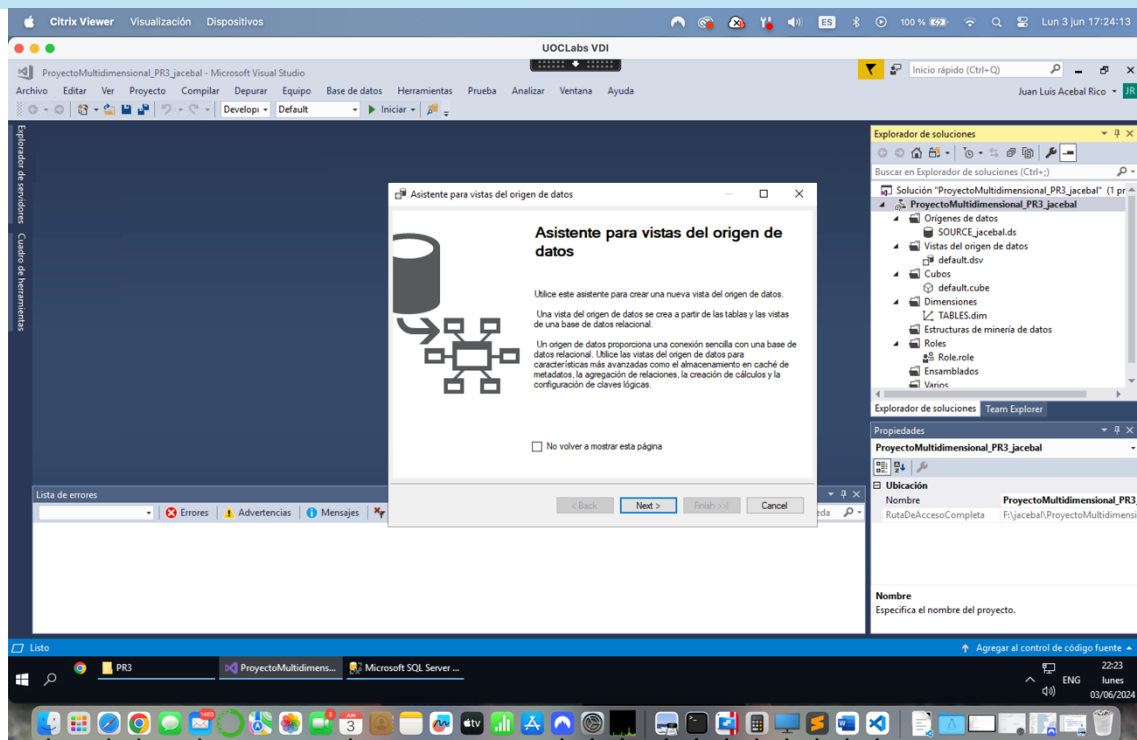
CREACIÓN VISTAS

Una vez que tenemos la más básica configuración de nuestro proyecto, necesitamos crear las vistas. Éstas se componen de lo que sería en el modelo relacional la tabla que hemos llamado de hechos (prefijo FACT) con sus relaciones que serían las dimensiones (prefijo DIM).

UBICACIÓN DE CREACIÓN DE VISTA

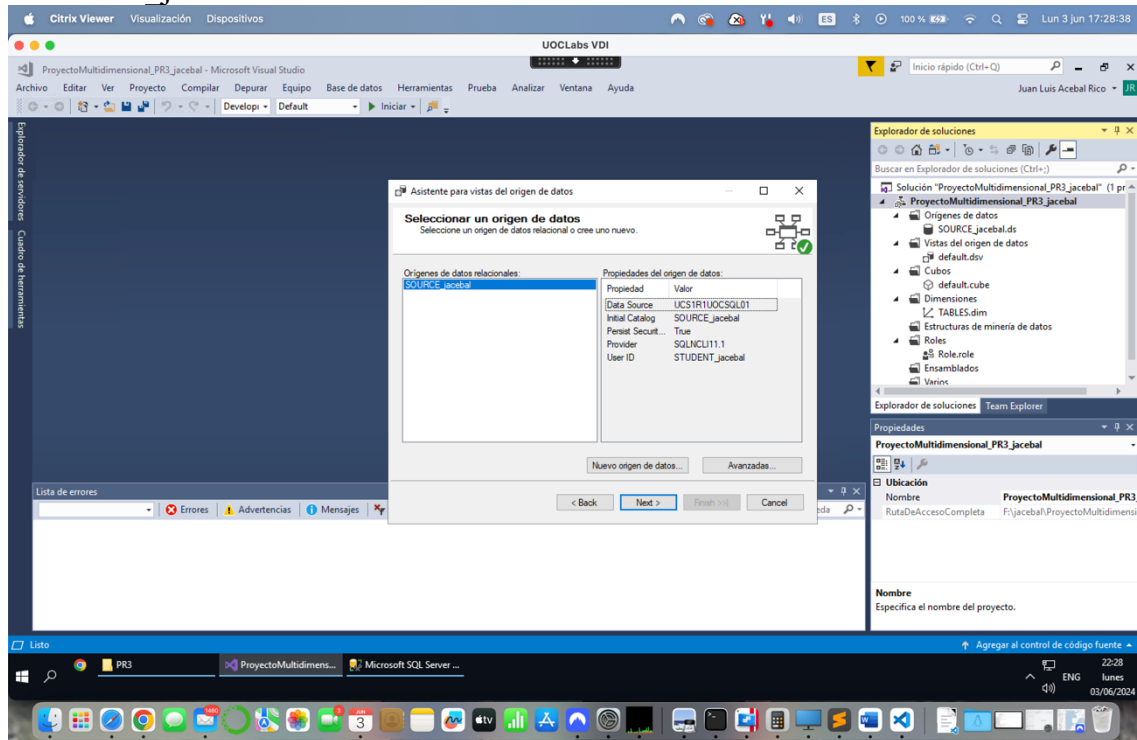


ASISTENTE CREACIÓN DE VISTA



ORIGEN DE DATOS

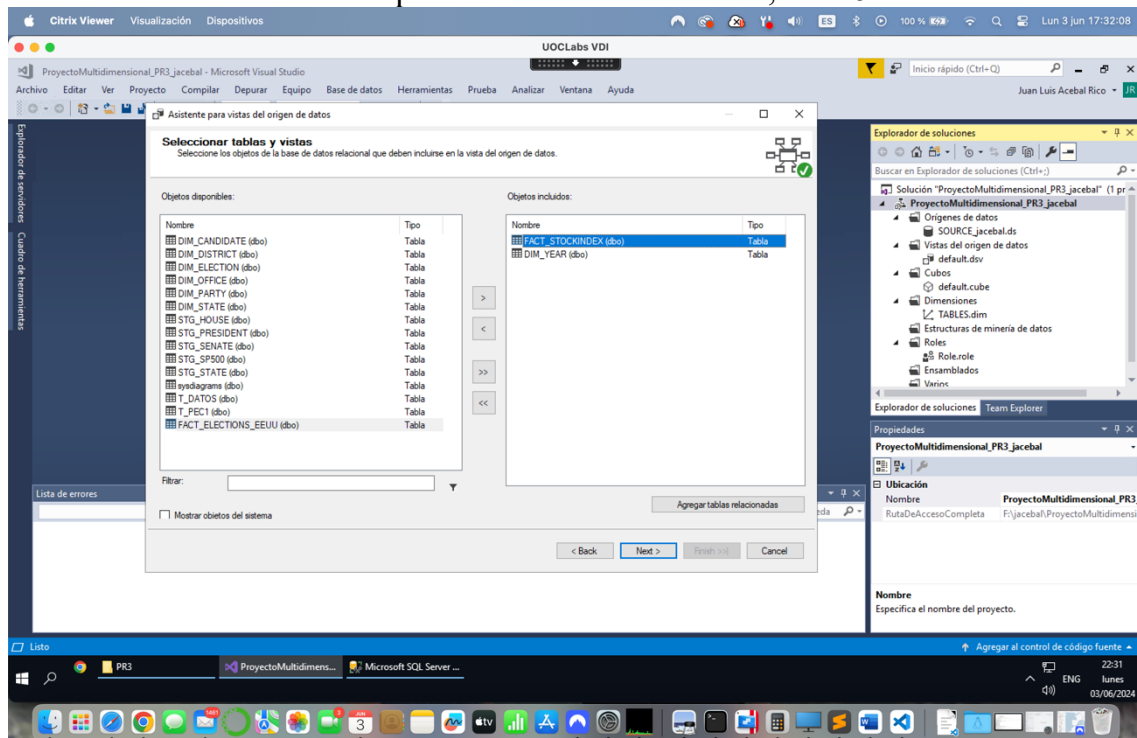
Selecciono la base de datos relacional donde están los datos, que son SOURCE_jacebal



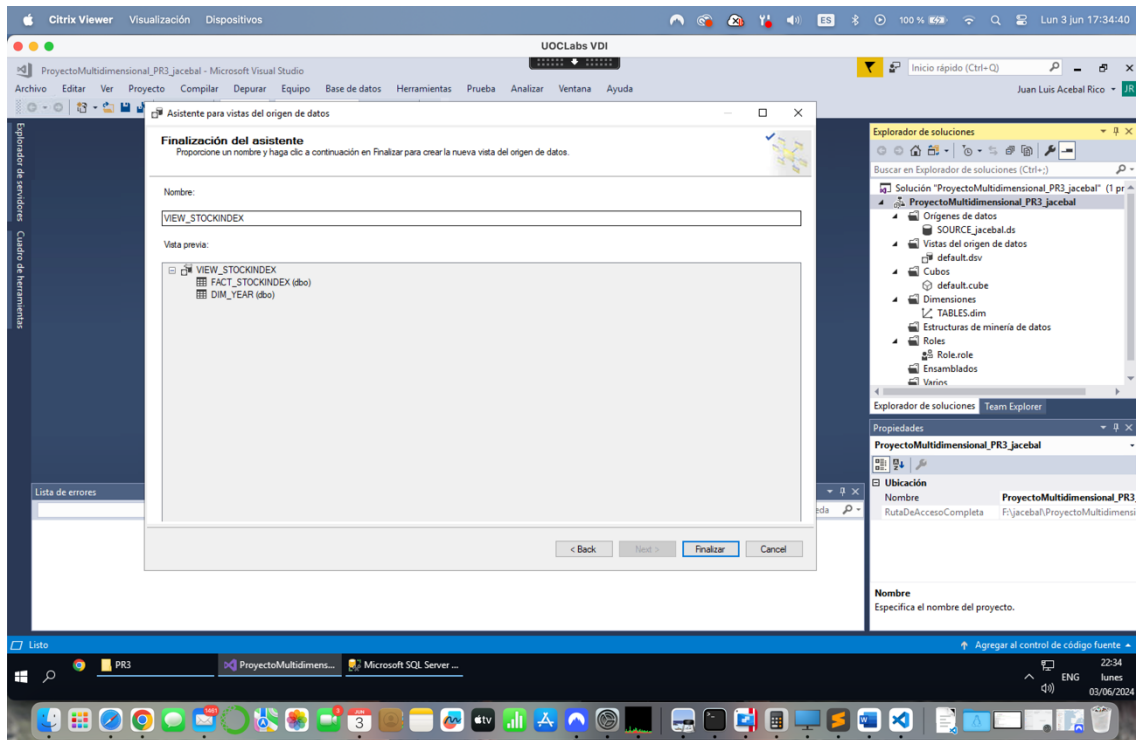
VISTA SP500

Voy a crear la primera vista, que serán las dimensiones de la tabla de hechos del sp500 mas sus dimensiones, que en este caso es DIM_year

Tip: Si le damos en una tabla a agregar tablas relacionadas, podemos agregar fácilmente todas las tablas que tienen relación con ésta, esto es útil especialmente con la otra tabla de hechos que no tiene una dimensión, sino 6

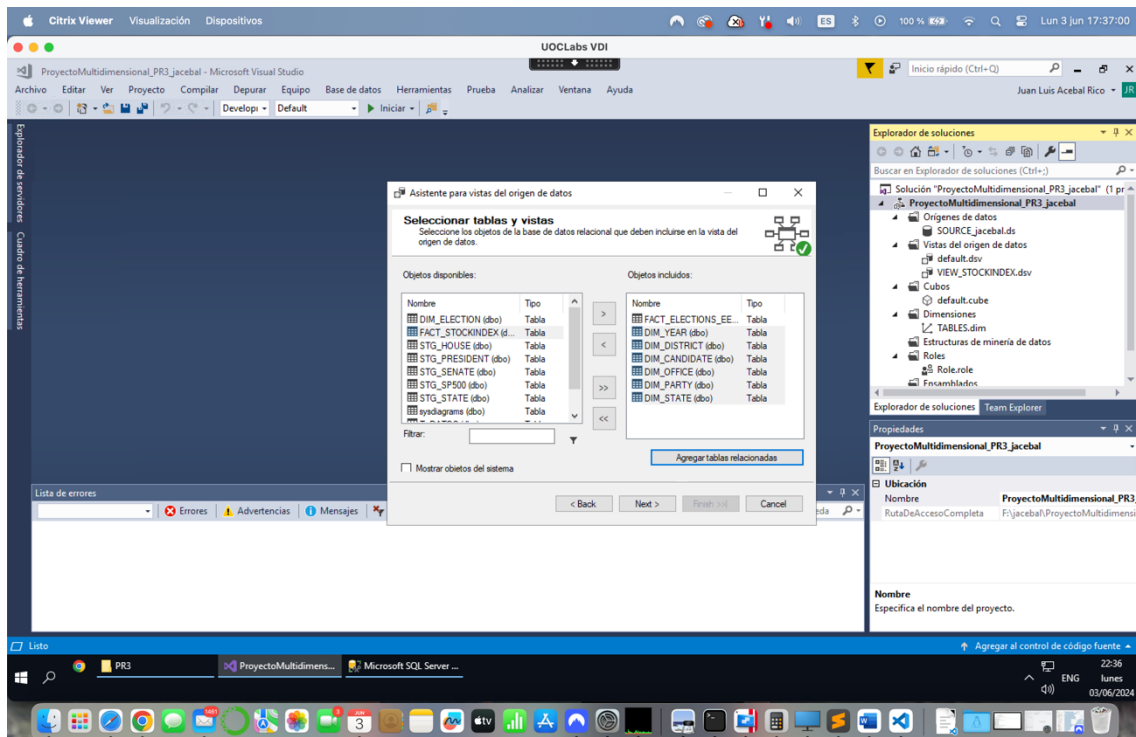


NOMBRE VISTA

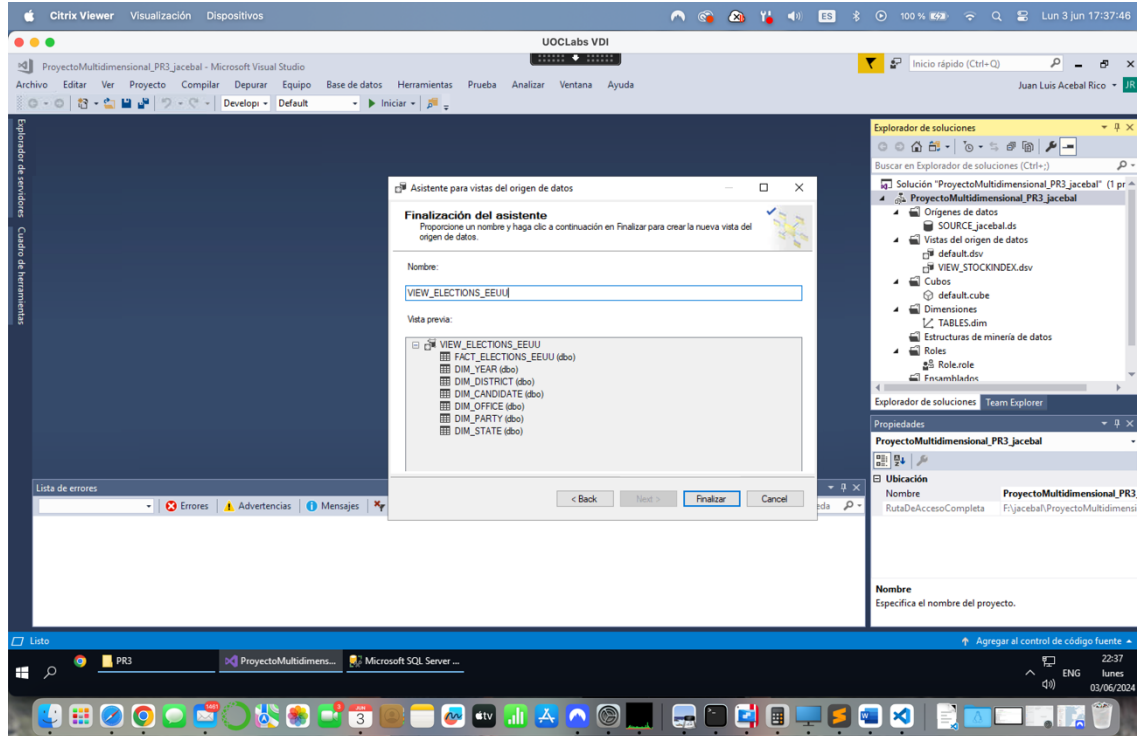


VISTA ELECCIONES

Aquí es donde va la otra vista de la otra tabla de hechos (modelo relacional) con sus dimensiones. Los primeros pasos son los mismos, ir a proyecto nueva vista y luego seleccionar el mismo origen de datos. Después seleccionamos las tablas del modelo relacional

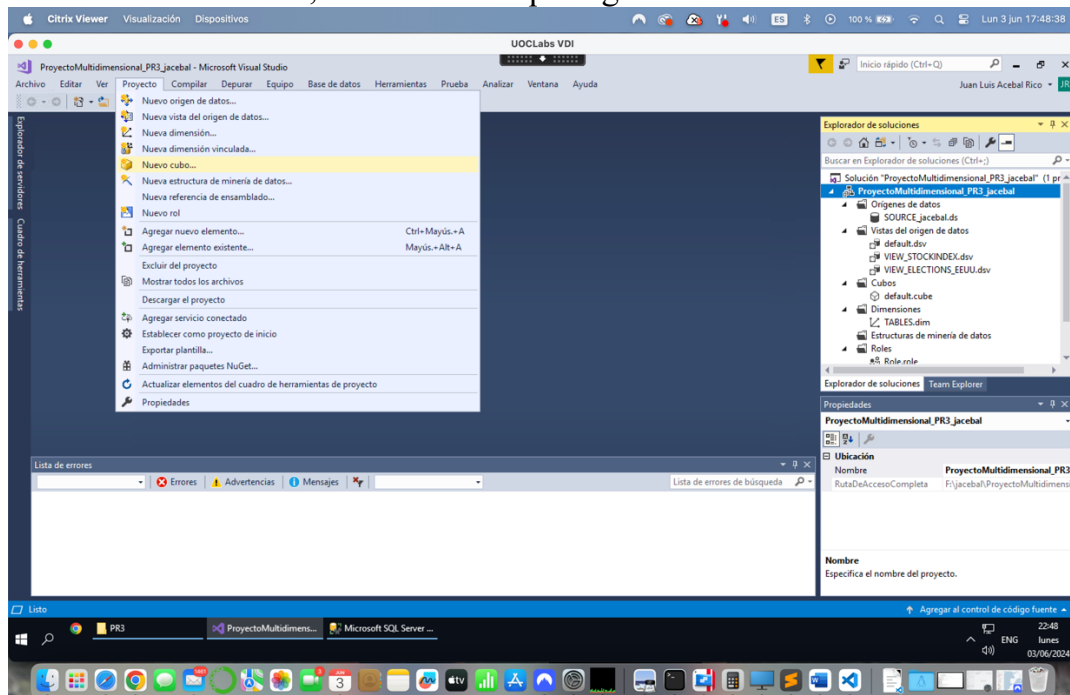


NOMBRE



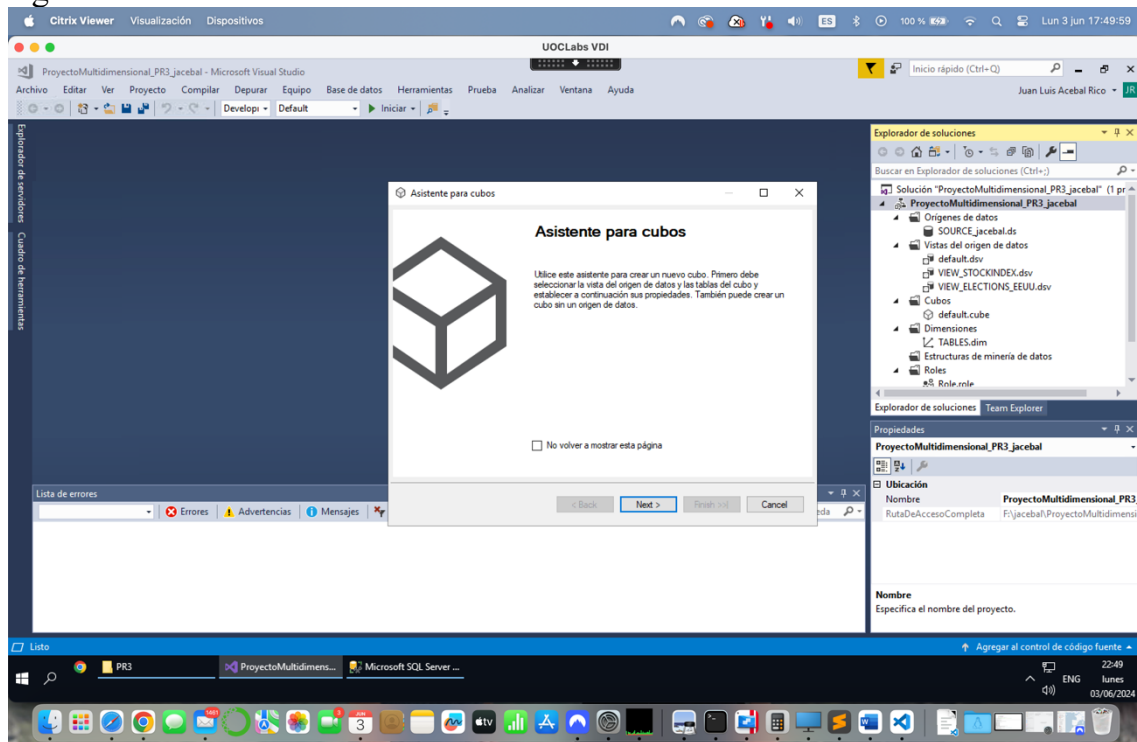
CREACIÓN DE CUBOS

Hasta ahora, todo lo que estábamos haciendo era dentro de un modelo relacional, es verdad que hemos hablado de dimensiones y tablas de hechos, pero siempre en el contexto de preparar los datos para crear el modelo multidimensional (OLAP) dentro de nuestro Data Warehouse. En el diseño del proyecto habíamos definido el modelo relacional, luego habíamos definido el modelo dimensional y también la zona staging. Pues bien, ahora vamos a crear los cubos, propiamente dentro del modelo dimensional, con las vistas que siguen usando el modelo relacional.



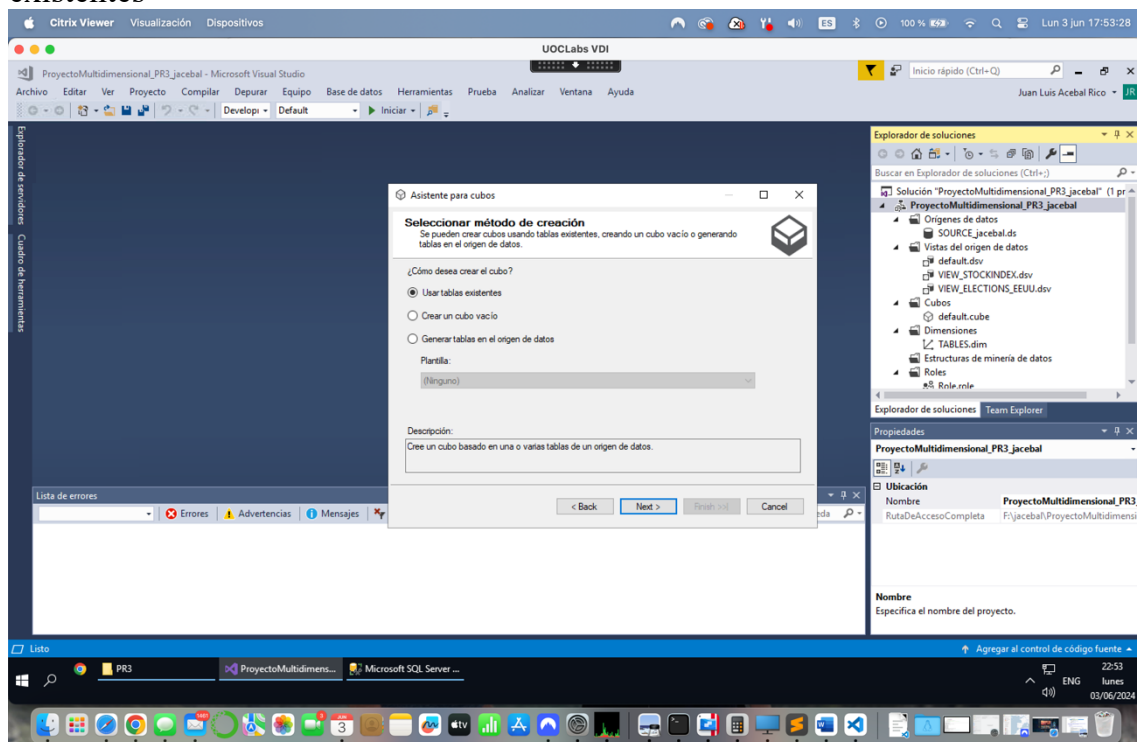
ASISTENTE

Después de seleccionar dentro del menú un nuevo cubo, hay que seleccionar siguiente.



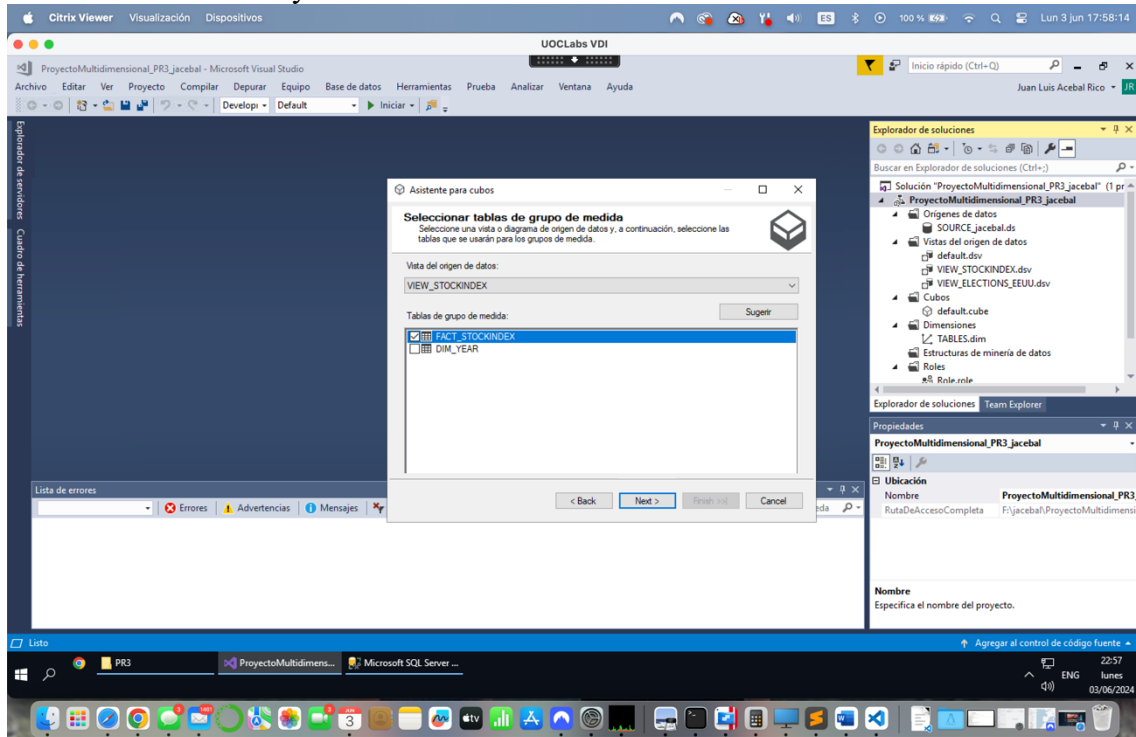
SELECCIONAR MÉTODO DE CREACIÓN

En nuestro contexto, el almacén de datos es importado de un modelo relacional, con sus transformaciones y modificaciones, hacia otro modelo relacional que es el usado para crear nuestro Data Warehouse OLAP, entonces usamos las tablas existentes



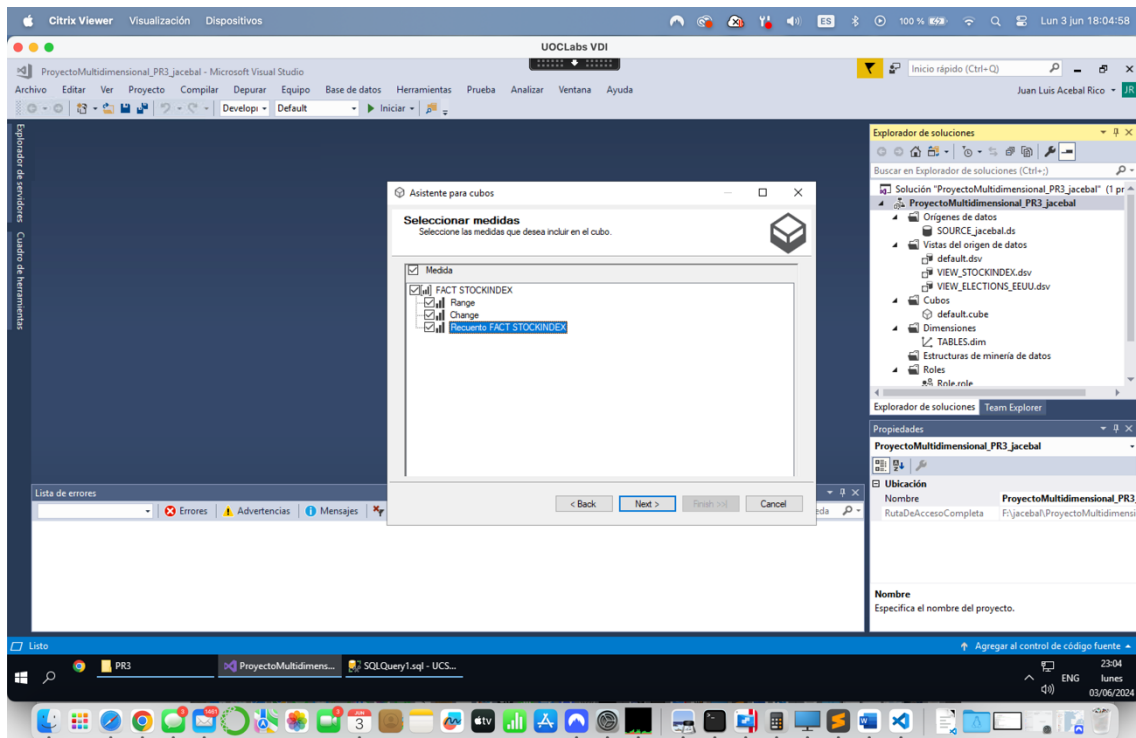
CUBE_STOCKINDEX

Selecciono la vista VIEW_STOCKINDEX, que será la homónima que el cubo que estoy creando, y luego donde está el grupo de medida, que significa o representa la métrica de los hechos ya definidos en el diseño de nuestro modelo dimensional.



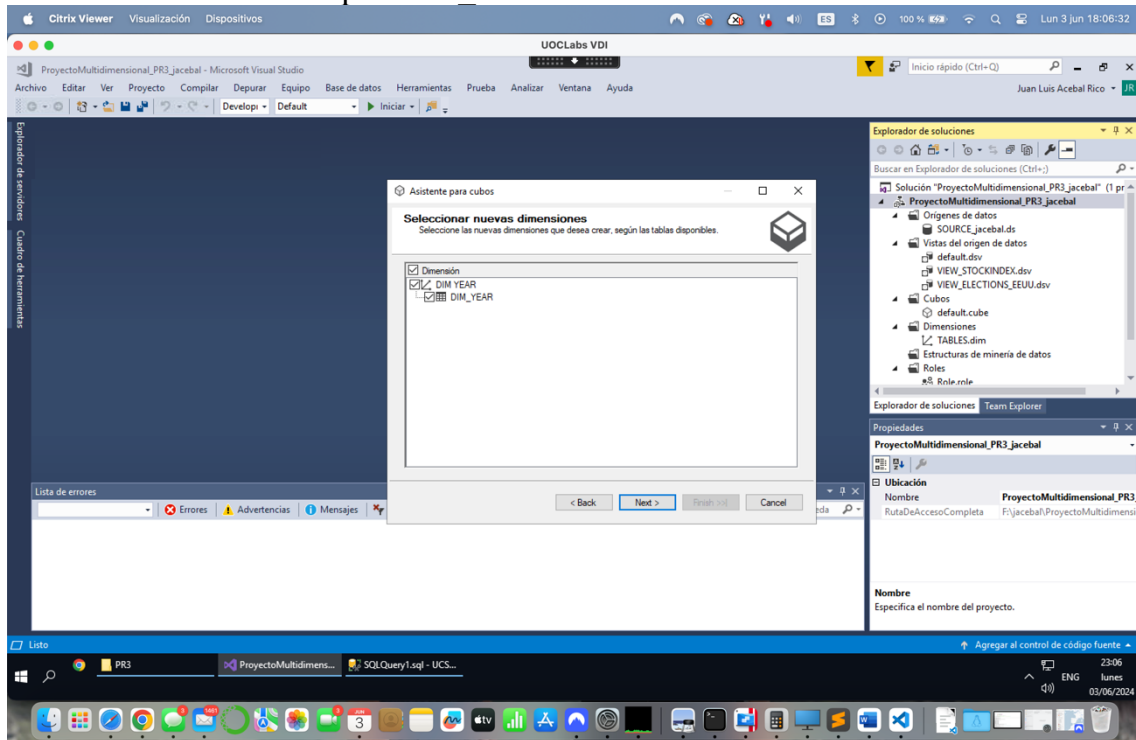
SELECCIONAR MEDIDAS

Aquí seleccionamos range, change y recuento, que viene a ser un contador de registros, útil por ejemplo para saber después cuantas veces ha ocurrido un hecho.

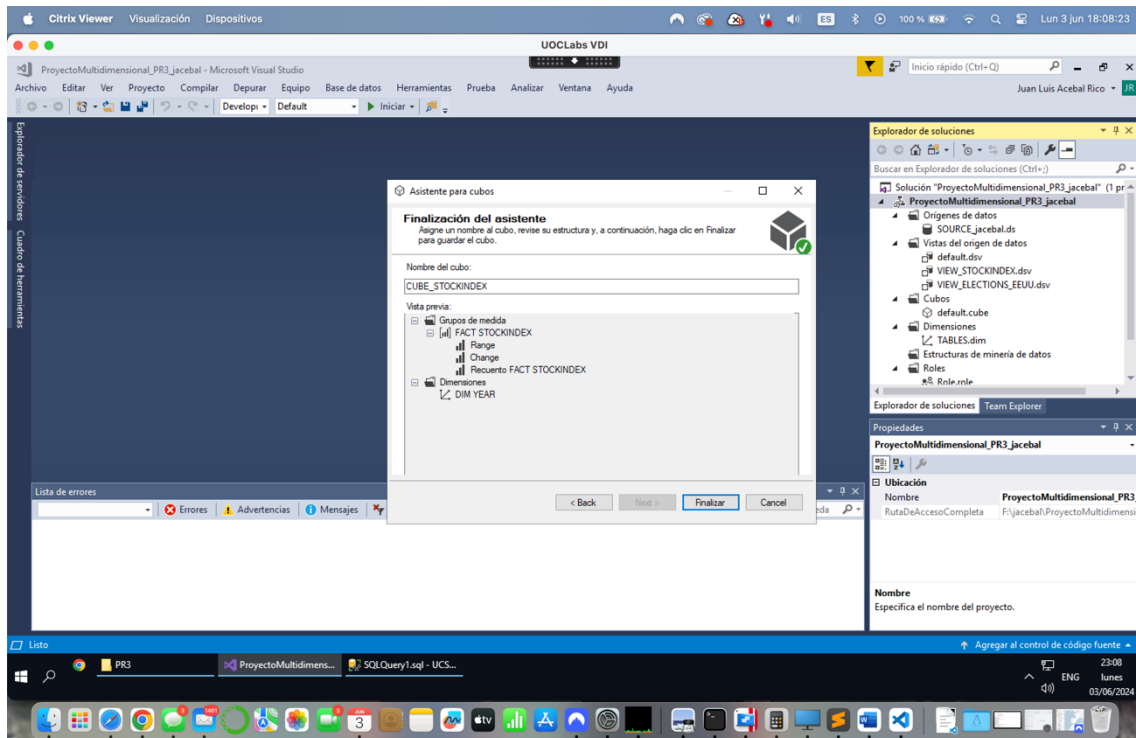


SELECCIONAR DIMENSIONES

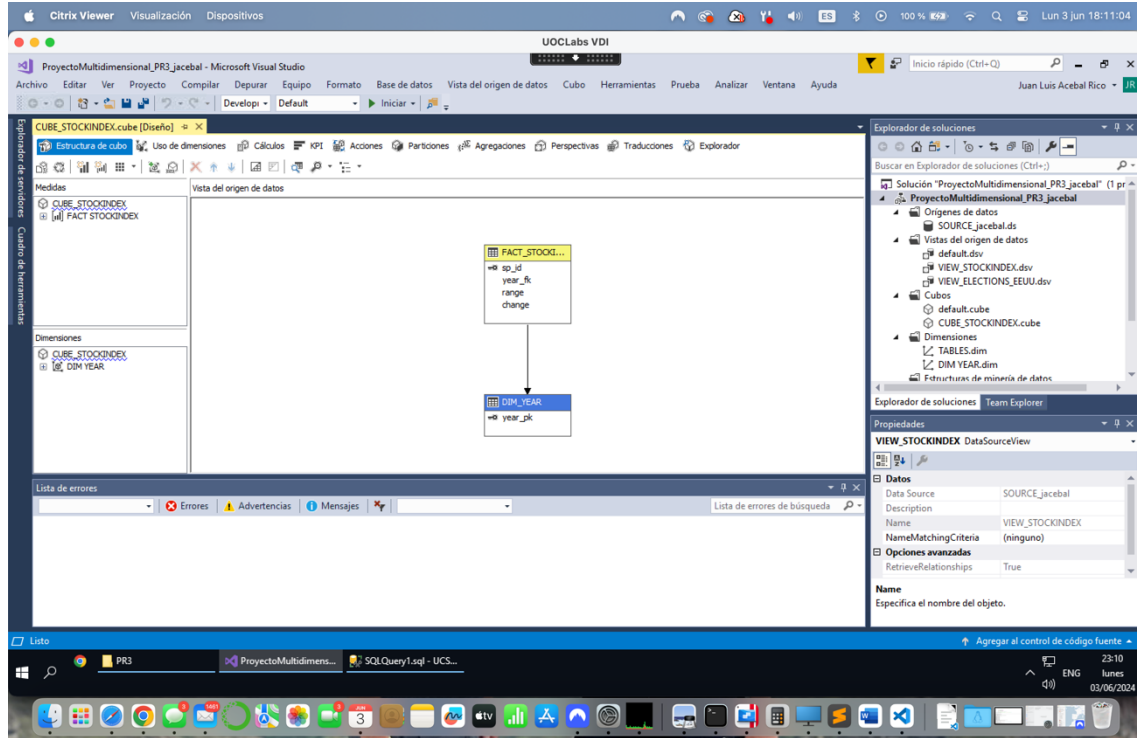
En el caso que nos ocupa, hemos definido en el modelo relacional y después en el modelo dimensional que DIM_YEAR es la única dimensión de este cubo.



NOMBRE

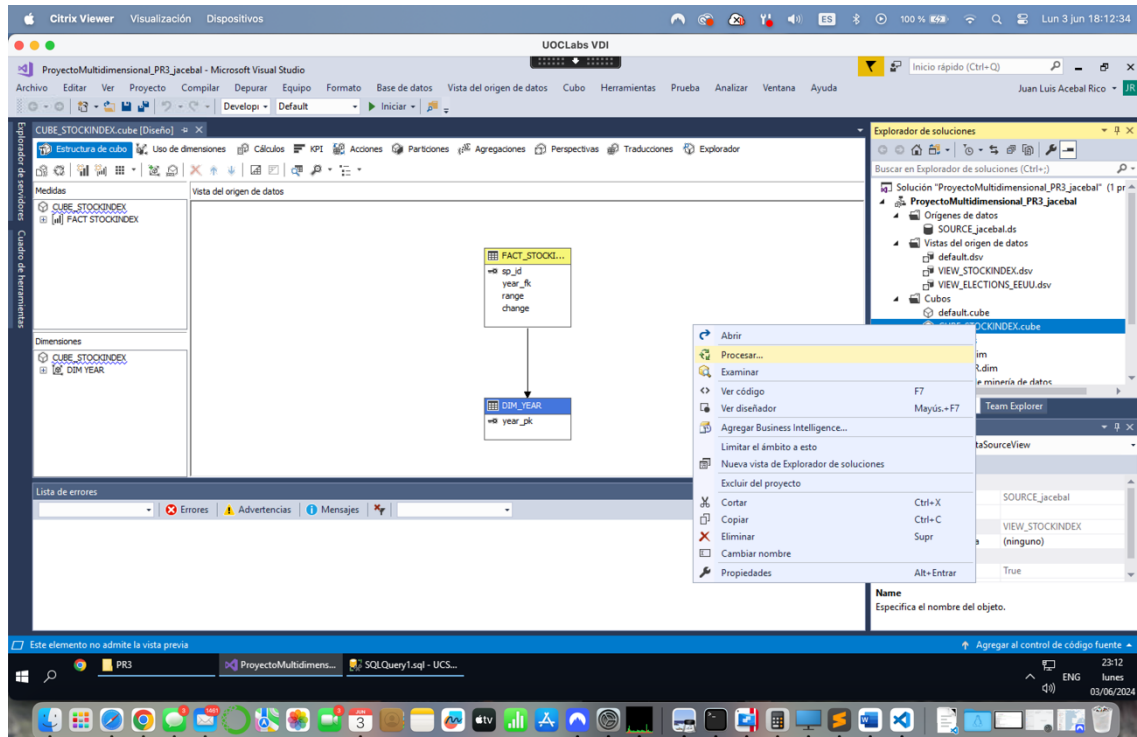


VISTA DE DISEÑO DEL CUBO

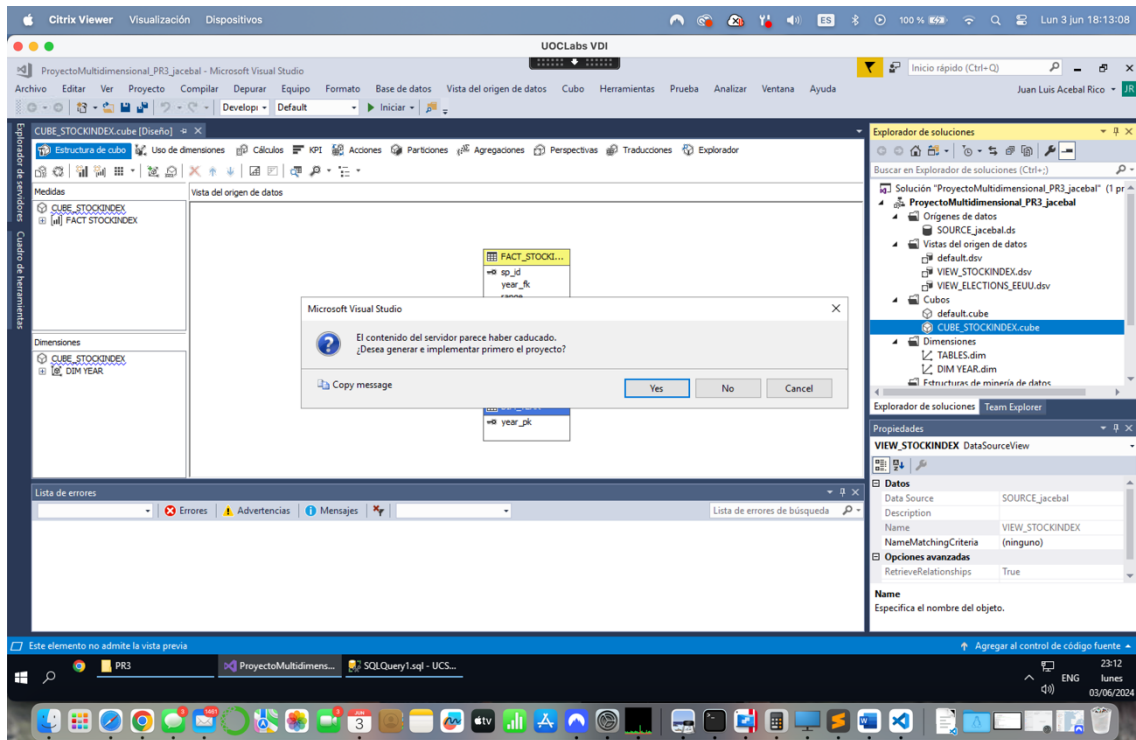


PROCESAMOS

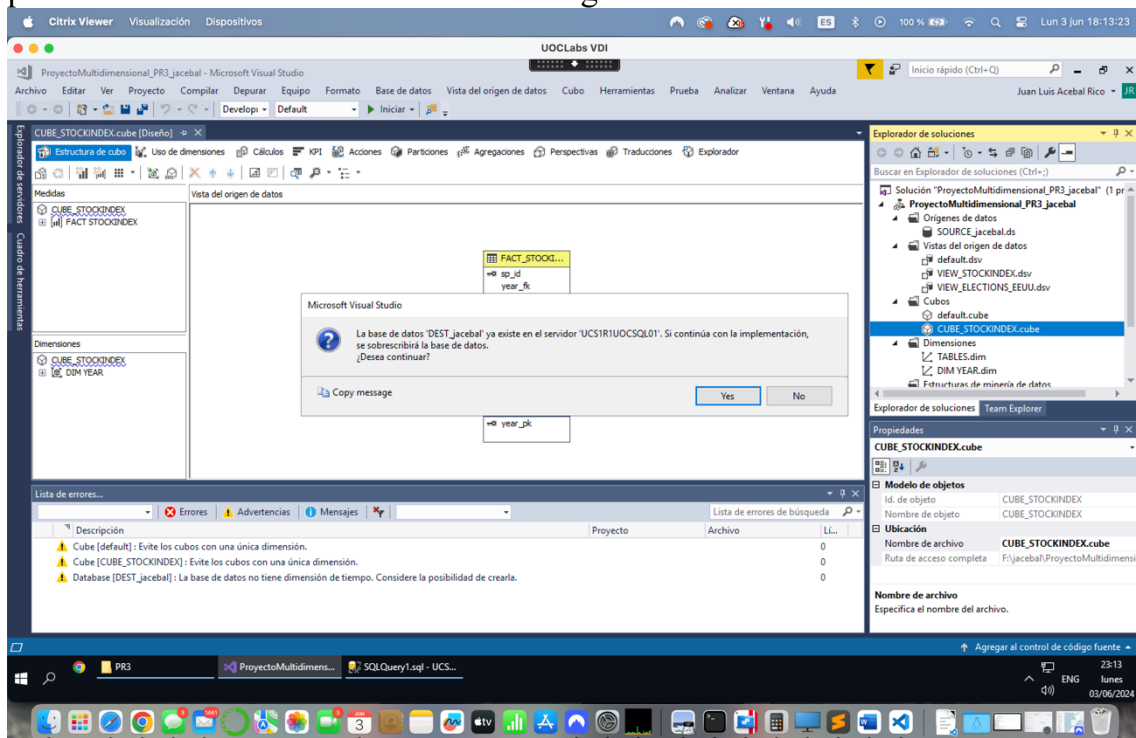
Así actualizamos cambios en el servidor



Aceptamos, ya que el proyecto no se ha implementado aún, es un aviso que no hay nada o que hay menos de lo que se espera en el servidor

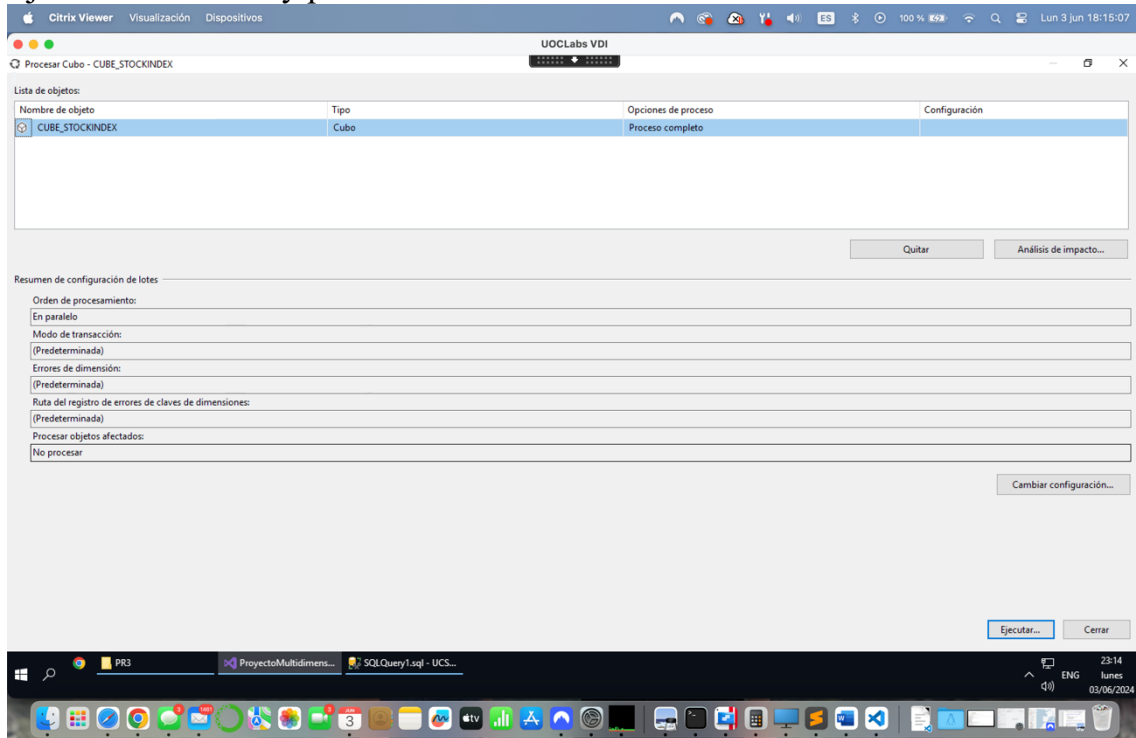


Ahora igual, la razón es que ya se ha usado para PEC1, o en su defecto yo ya la había utilizado. Está bien saber cuándo es un aviso sin importancia, o cuando puedes estar borrando los datos de una organización.

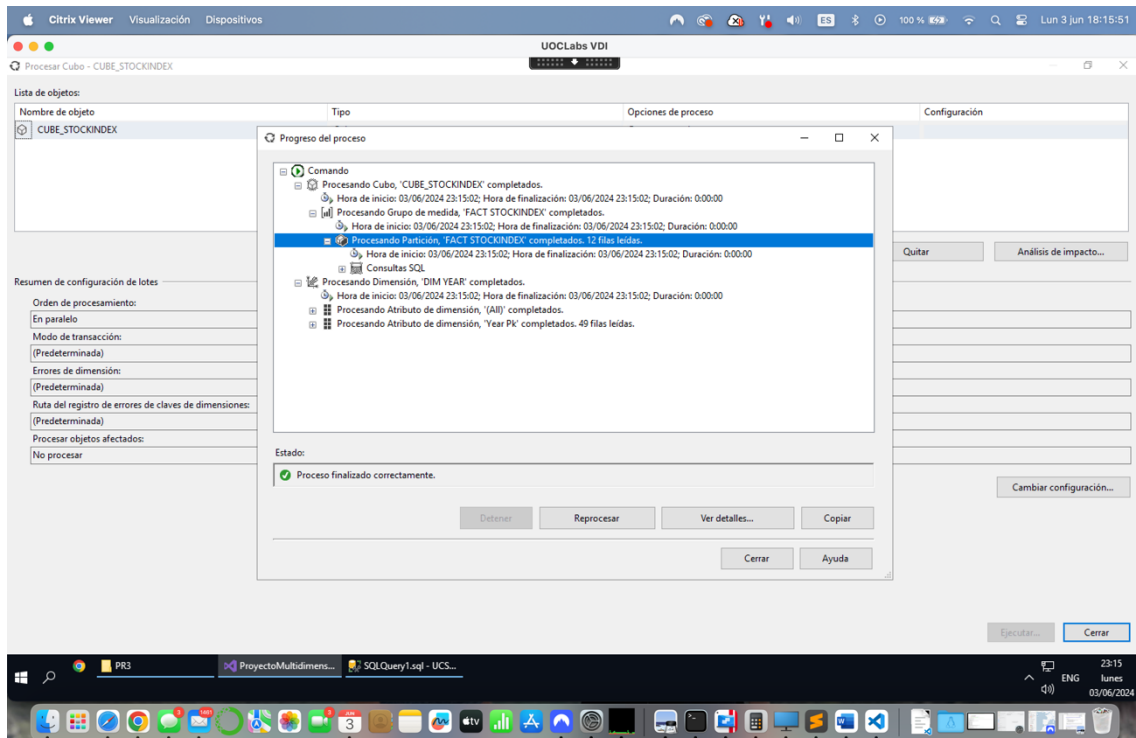


EJECUTAR

Ejecuto la creación y procesamiento del cubo.

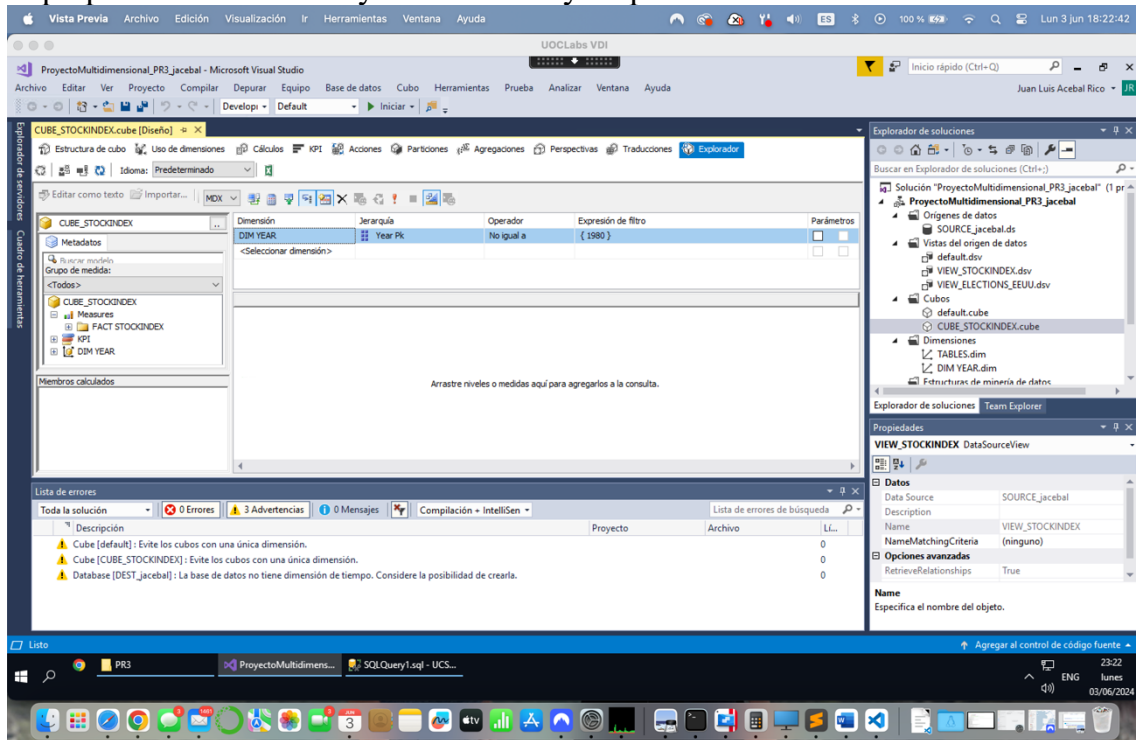


RESULTADO

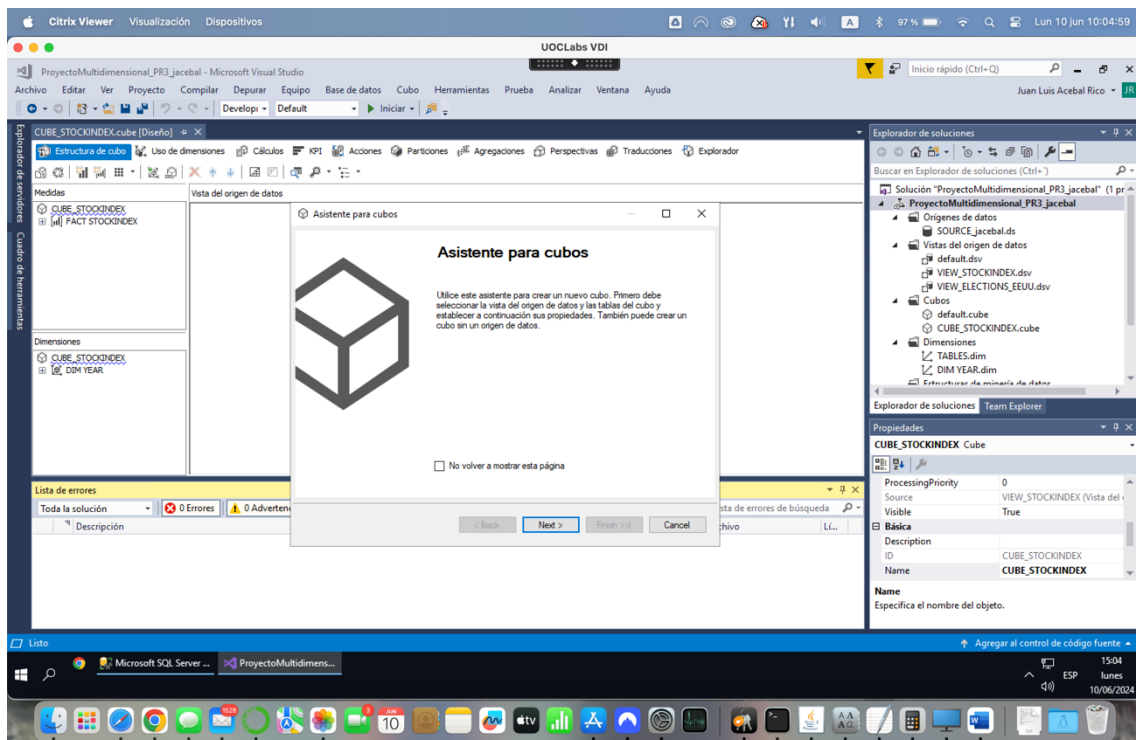


VISTA EXPLORADOR

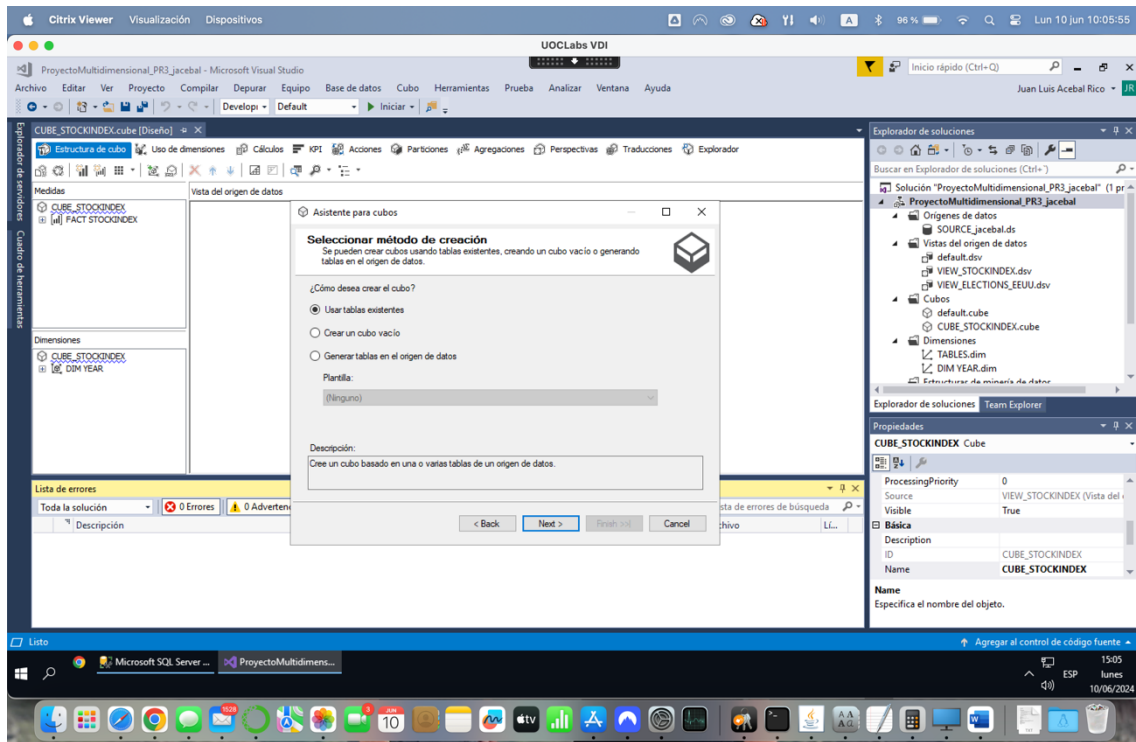
Aquí podemos ver como ya está creado y disponible el cubo en nuestro servidor



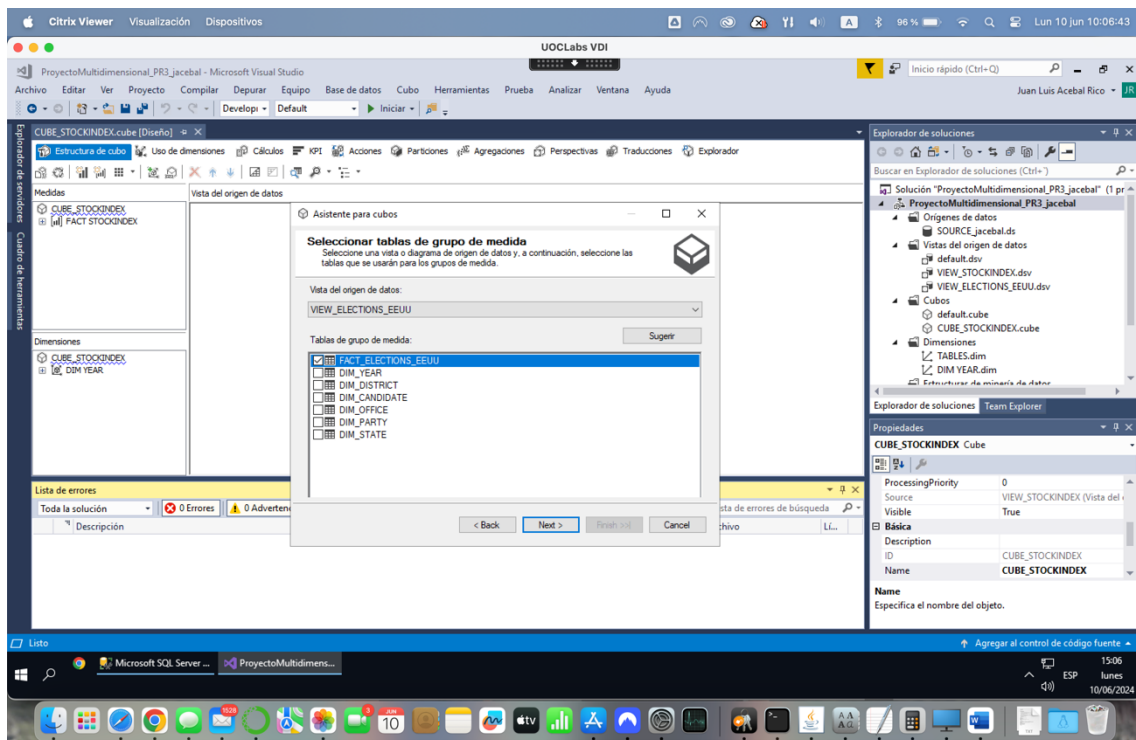
CUBE ELECTIONS EEUU



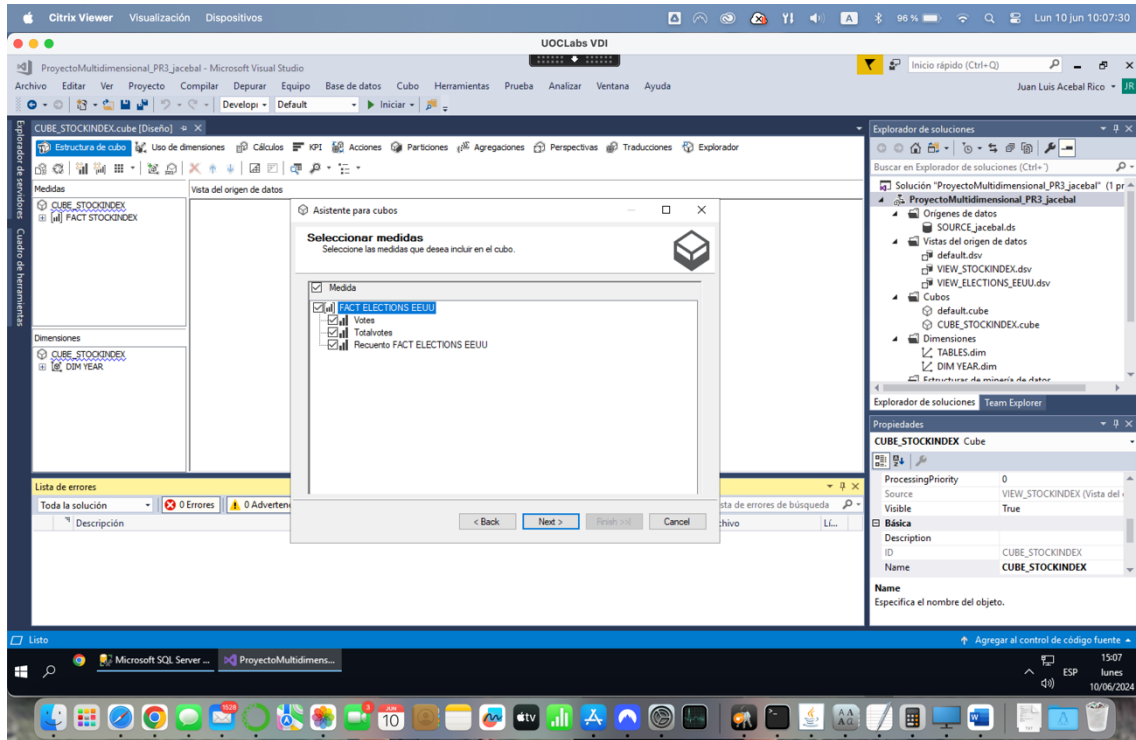
Usamos como antes, tablas existentes:



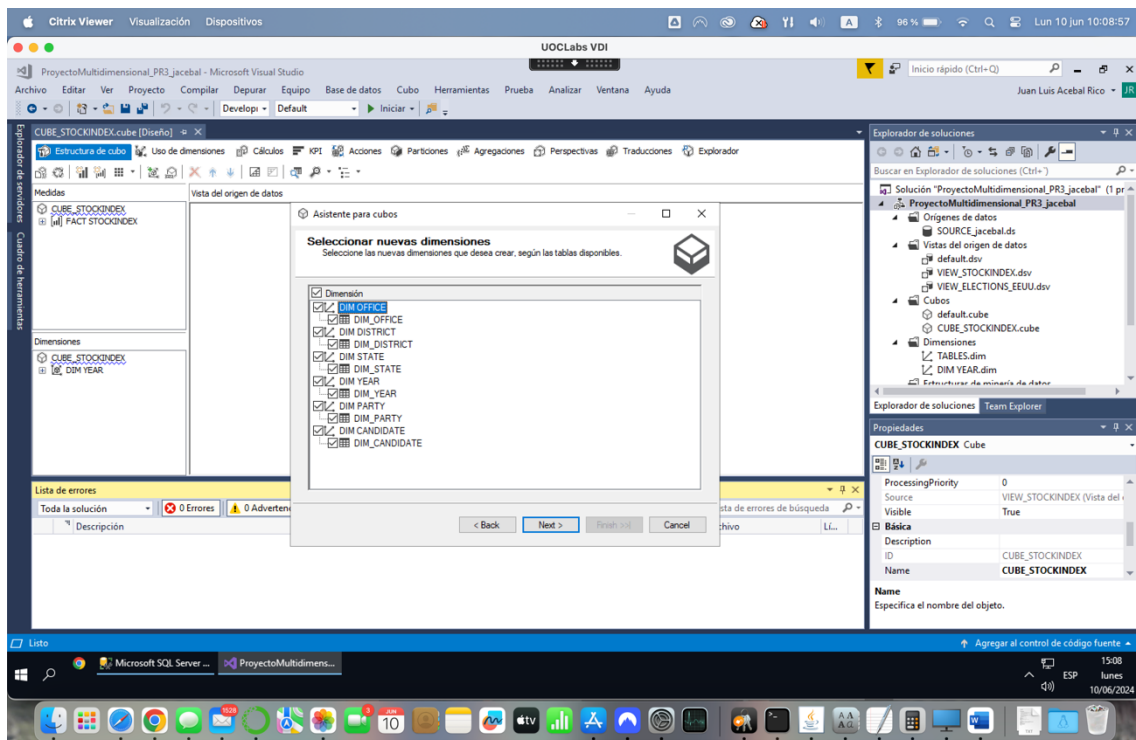
Seleccionamos la tabla donde están las métricas o medidas:



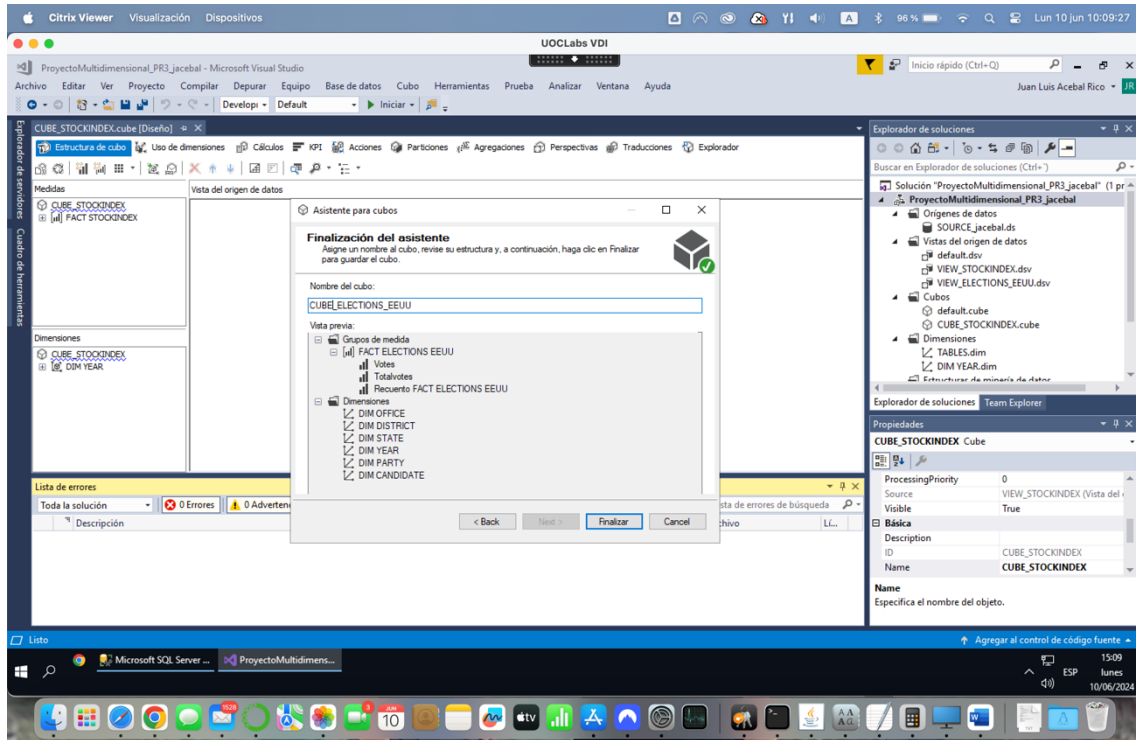
Elijo las métricas, que serán las existentes en el diseño dimensional que habíamos previsto (totalvotes y votes) más recuento:



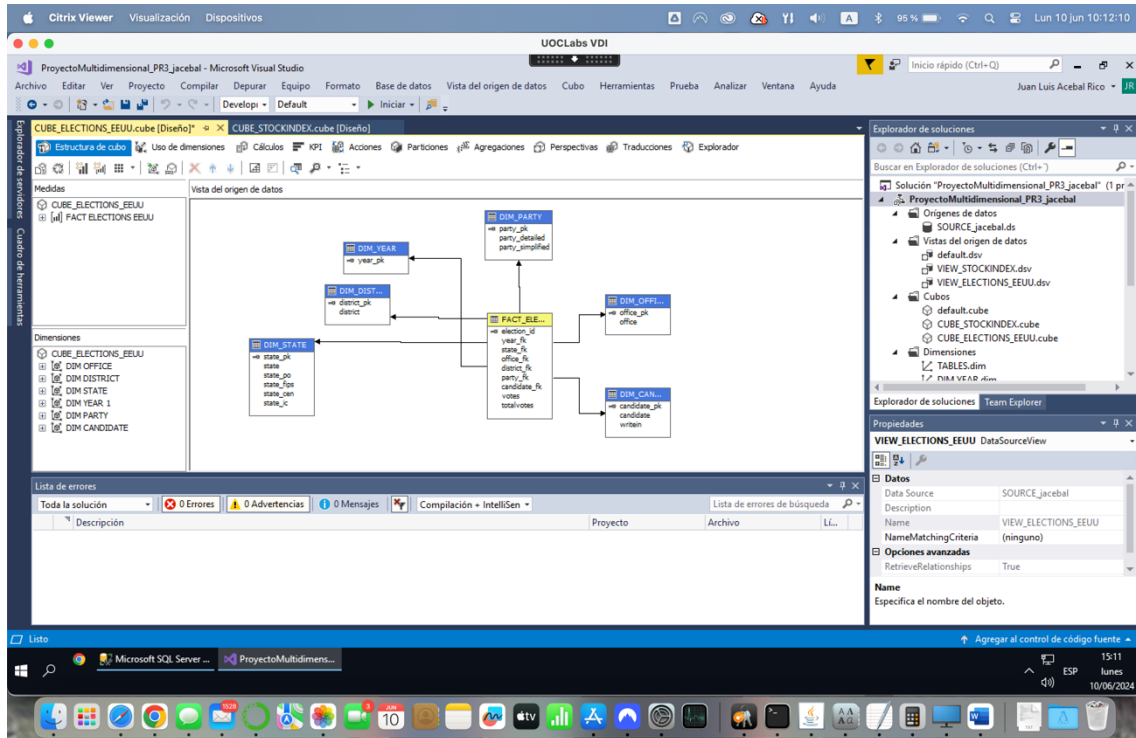
Elijo las dimensiones:



Reviso la vista previa y finalizar:

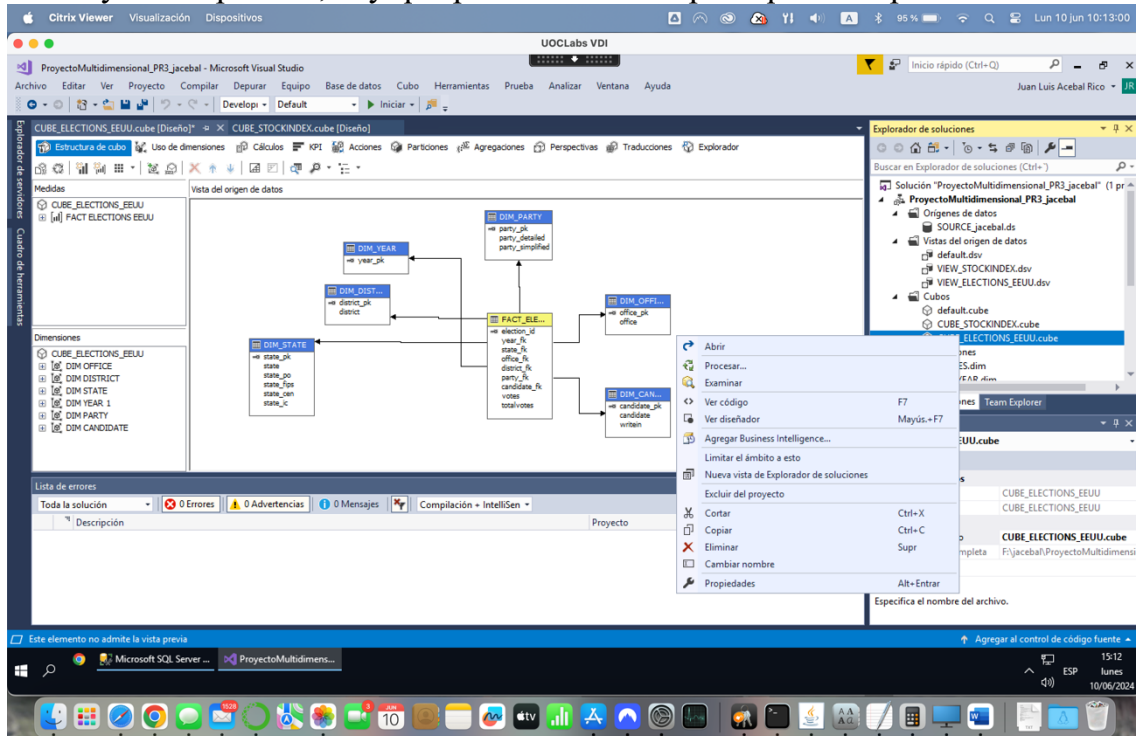


Vista del cubo creado:

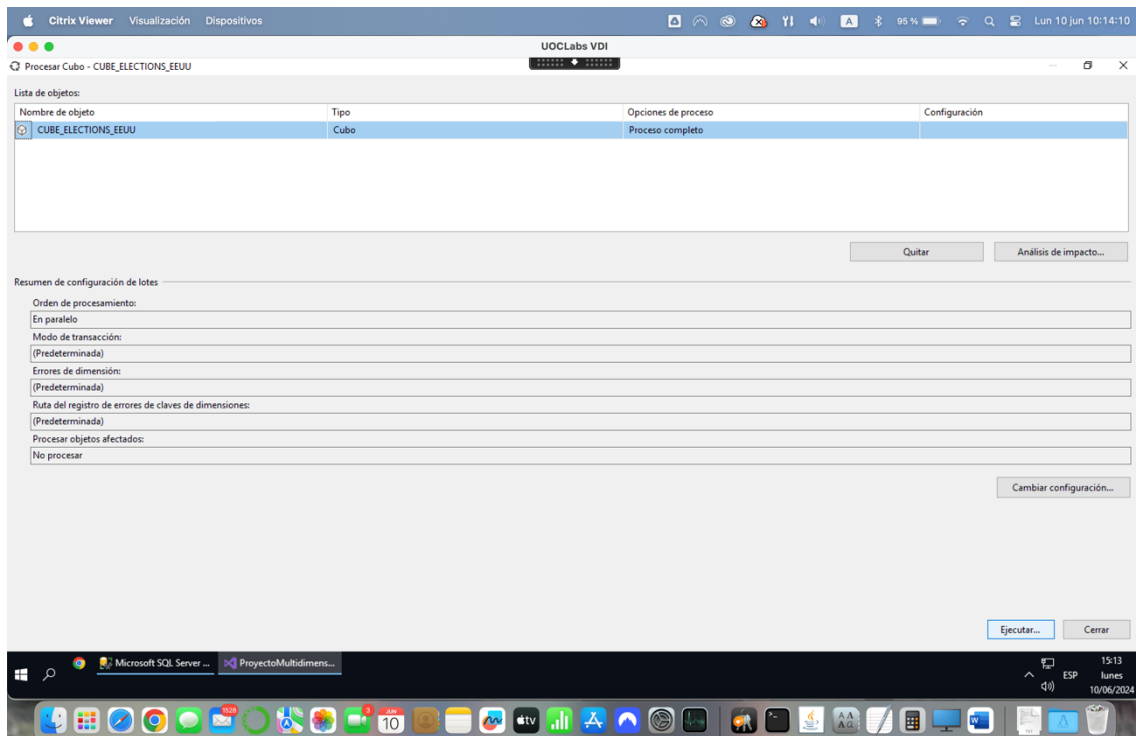


PROCESAR

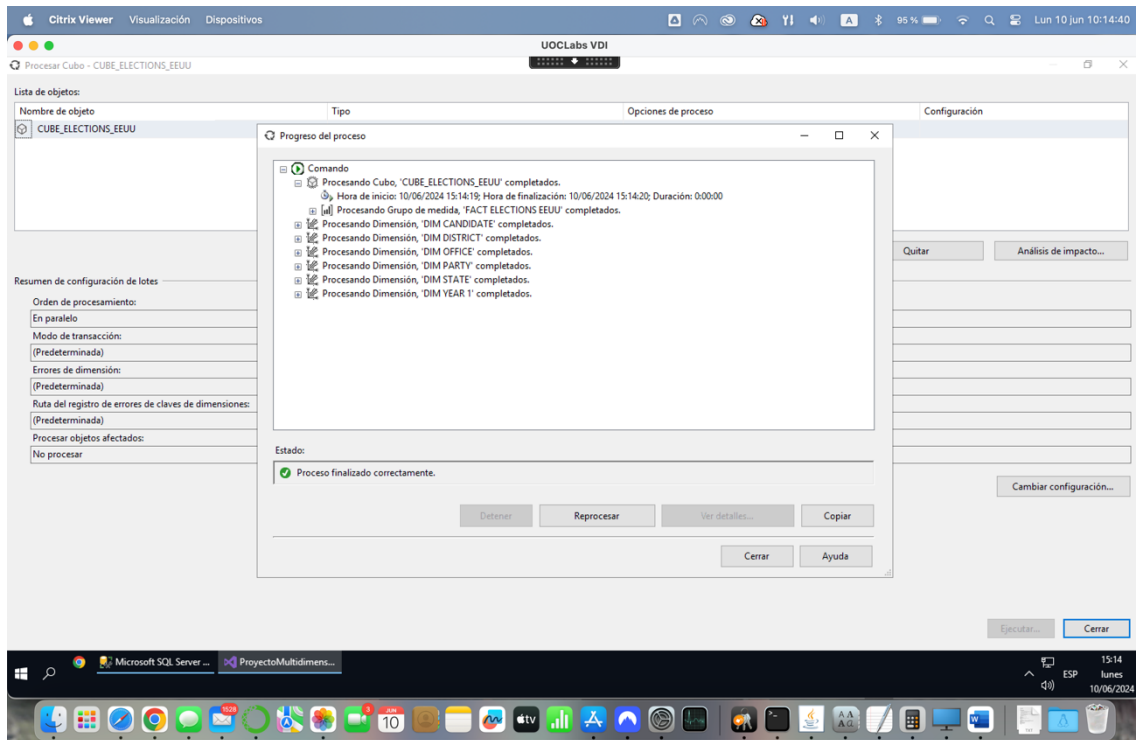
Como ya he explicado, hay que procesar el cubo para que se implante:



Ejecuto la implementación:

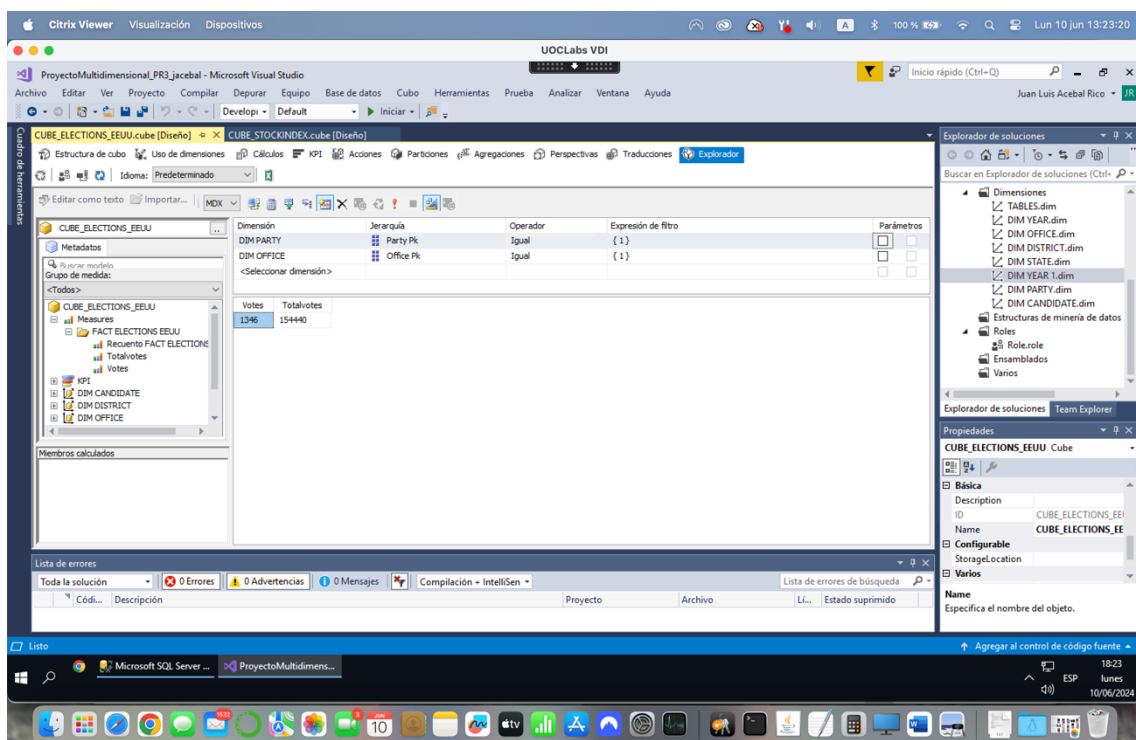


Ahora que la implementación está terminada, podemos trabajar con los cubos:



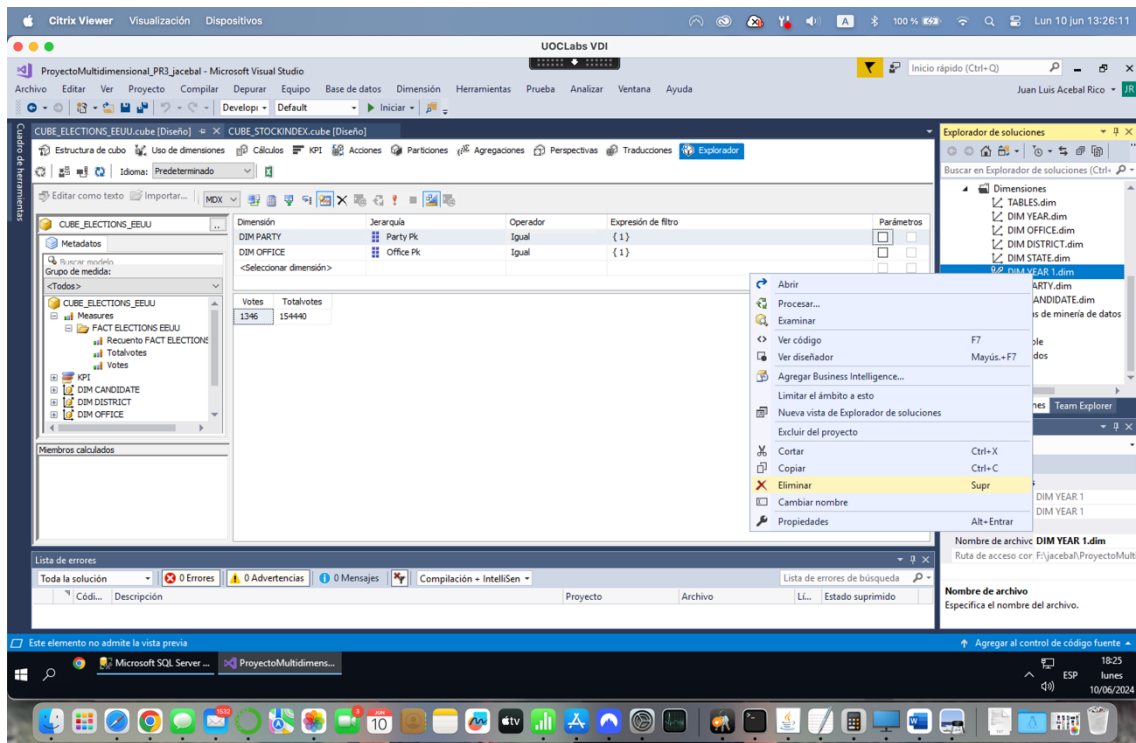
VISTA FINAL DE UNA CONSULTA

Ya terminado de crear los cubos, compruebo que puedo hacer consultas, y ahora quedaría eliminar una dimensión duplicada ya que esta dimensión es compartida por las dos tablas de hechos, y no es eficiente tenerlo así:

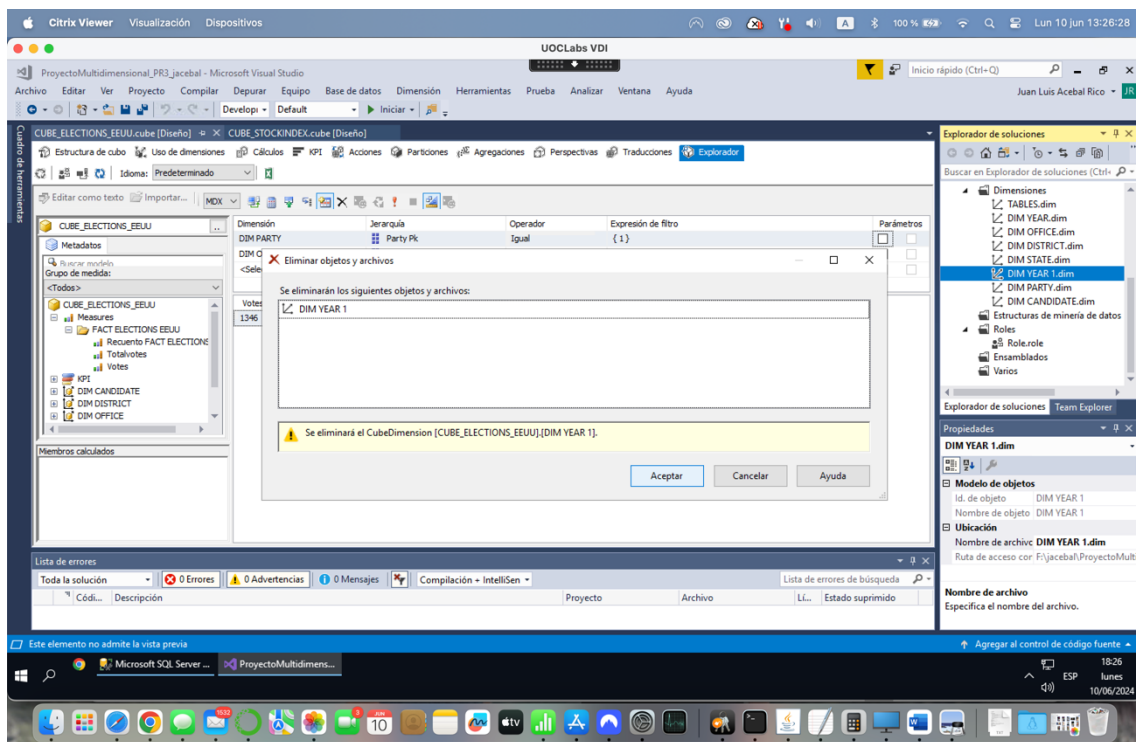


ELIMINAR DIMENSIÓN DUPLICADA

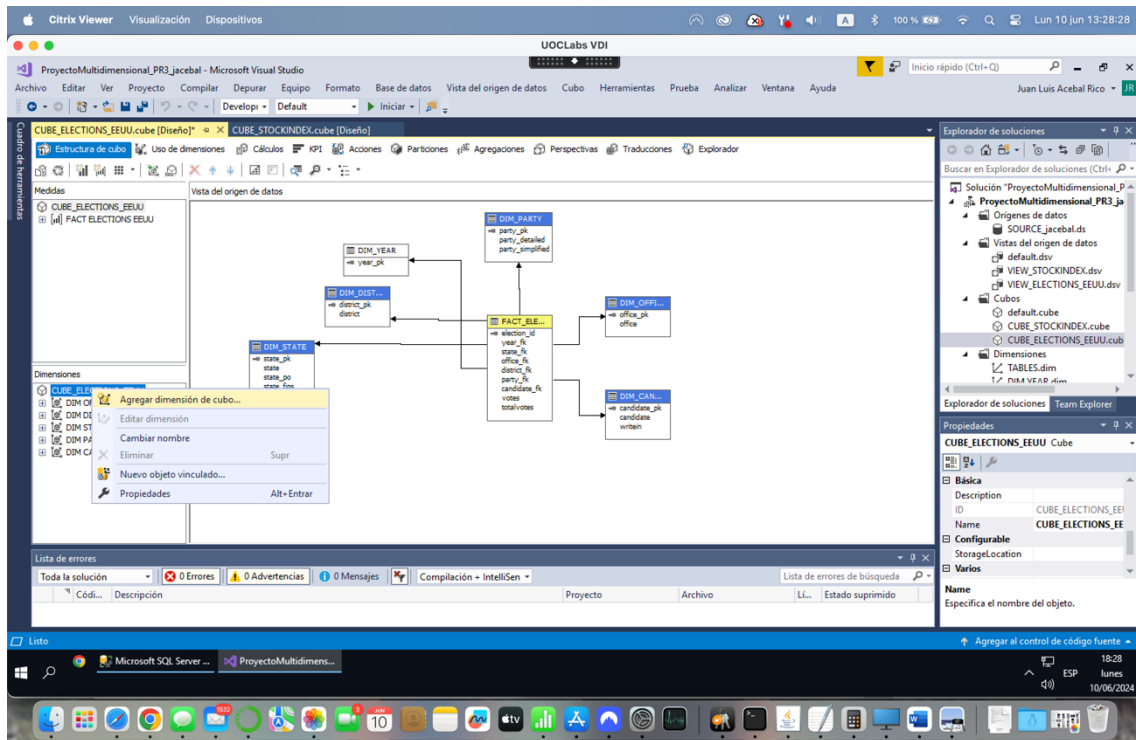
En el explorador de soluciones eliminamos la dimensión duplicada



Aceptamos la eliminación

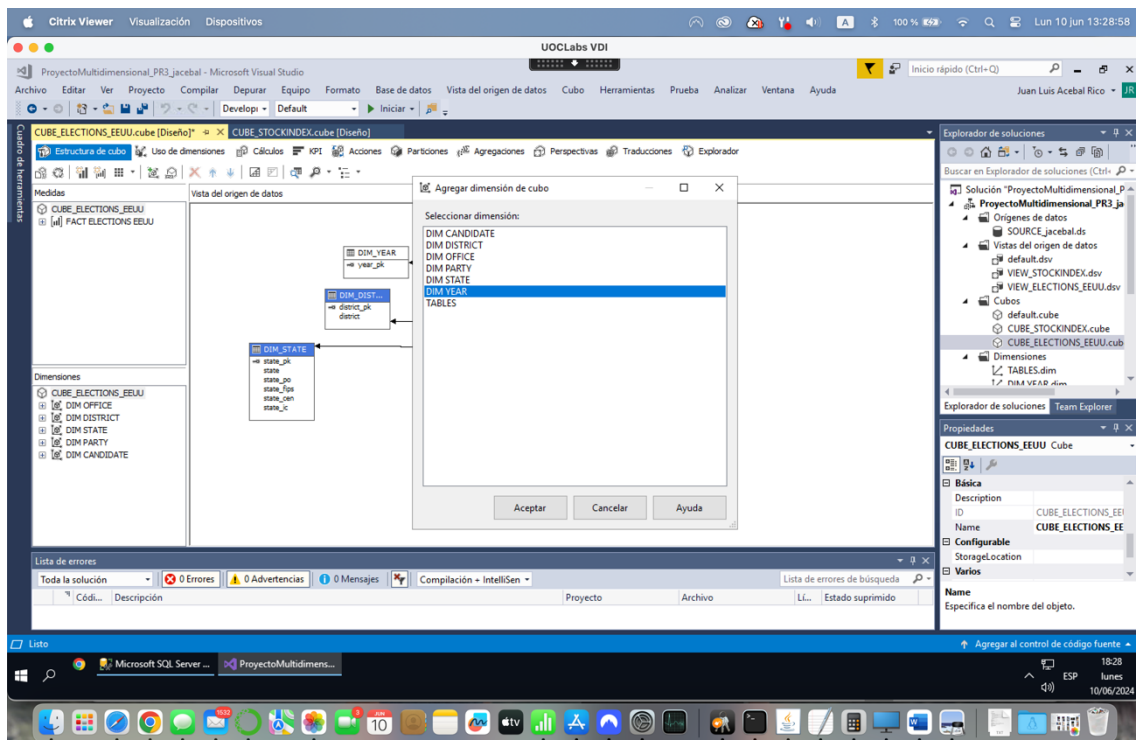


AGREGAR DIMENSIÓN

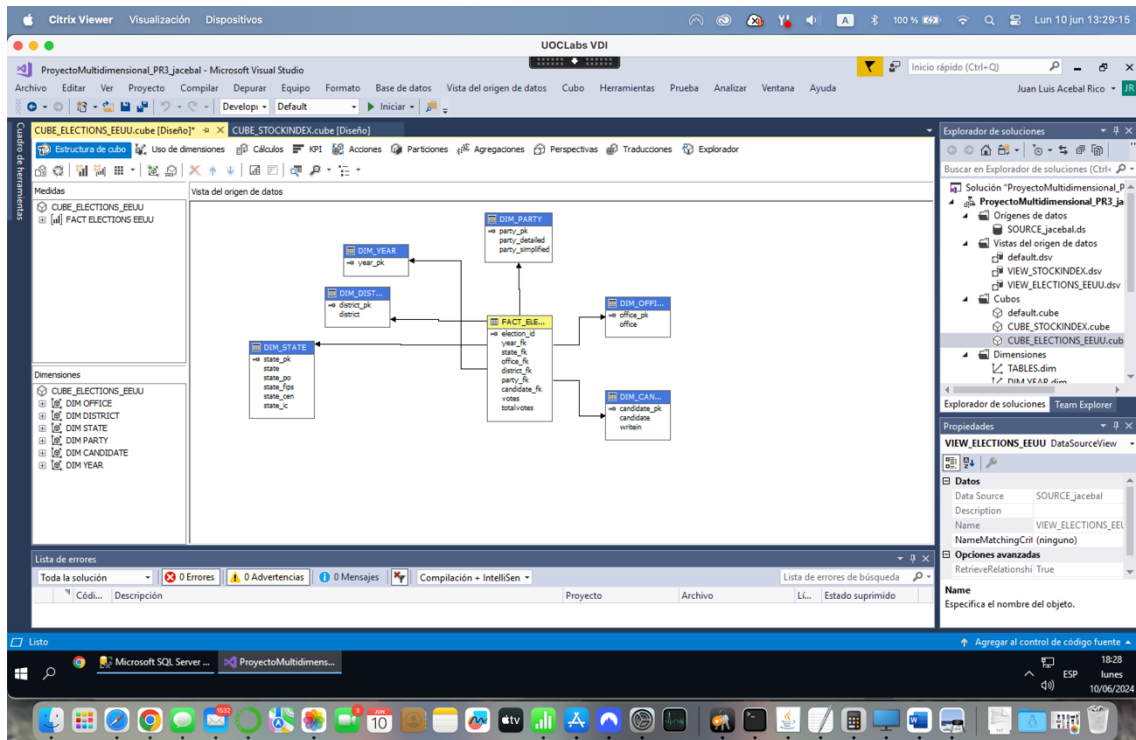


AGREGAR DIMENSIÓN

Selecciono la dimensión que había eliminado

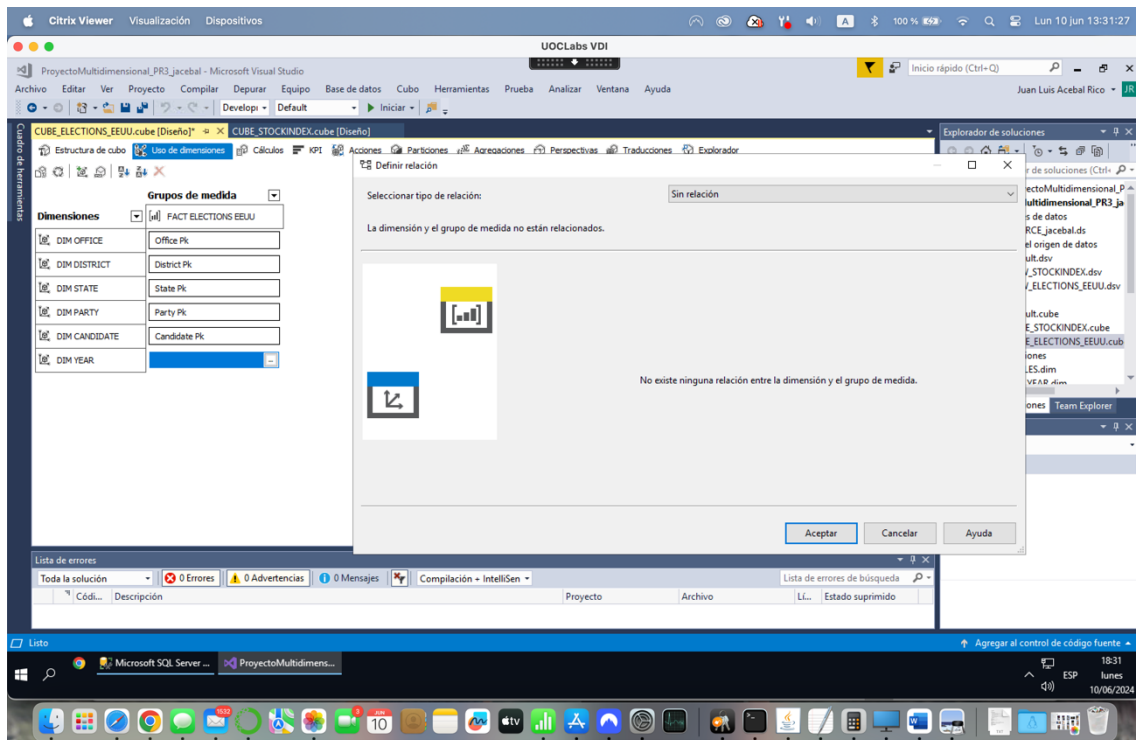


Ahora vemos que se ha vuelto a poner “azul” el nombre de la dimensión:

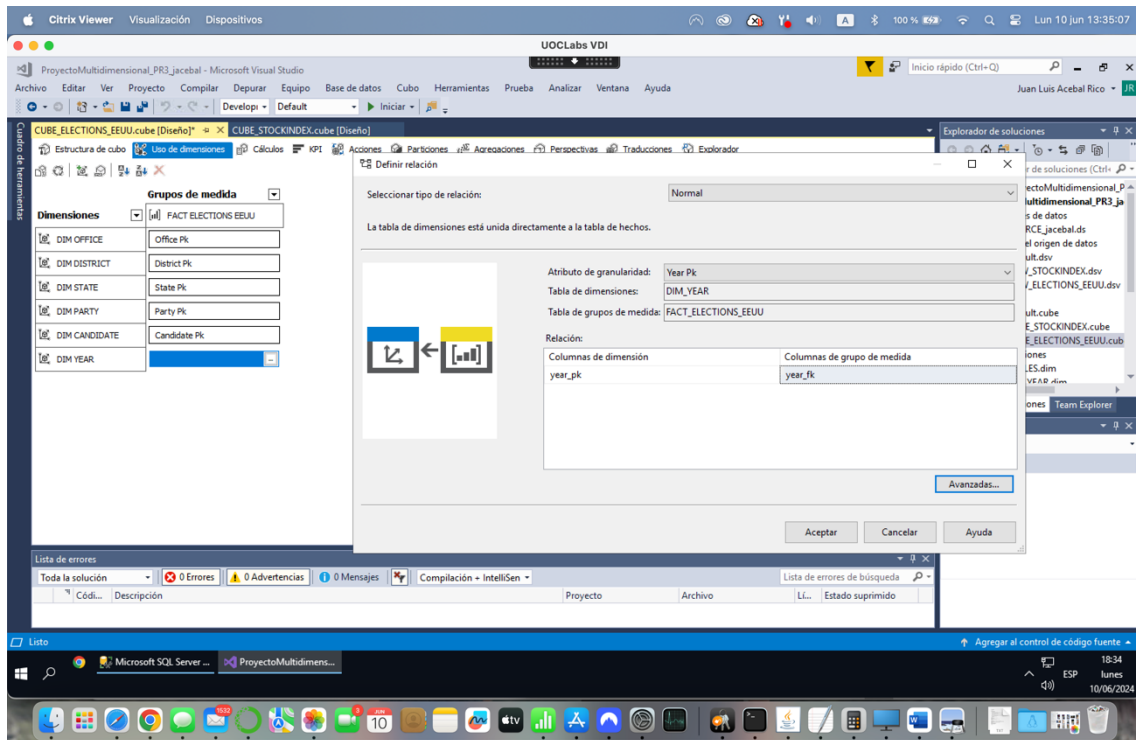


DEFINIR RELACIÓN

Nos queda definir la relación para que se quede bien hechas las relaciones

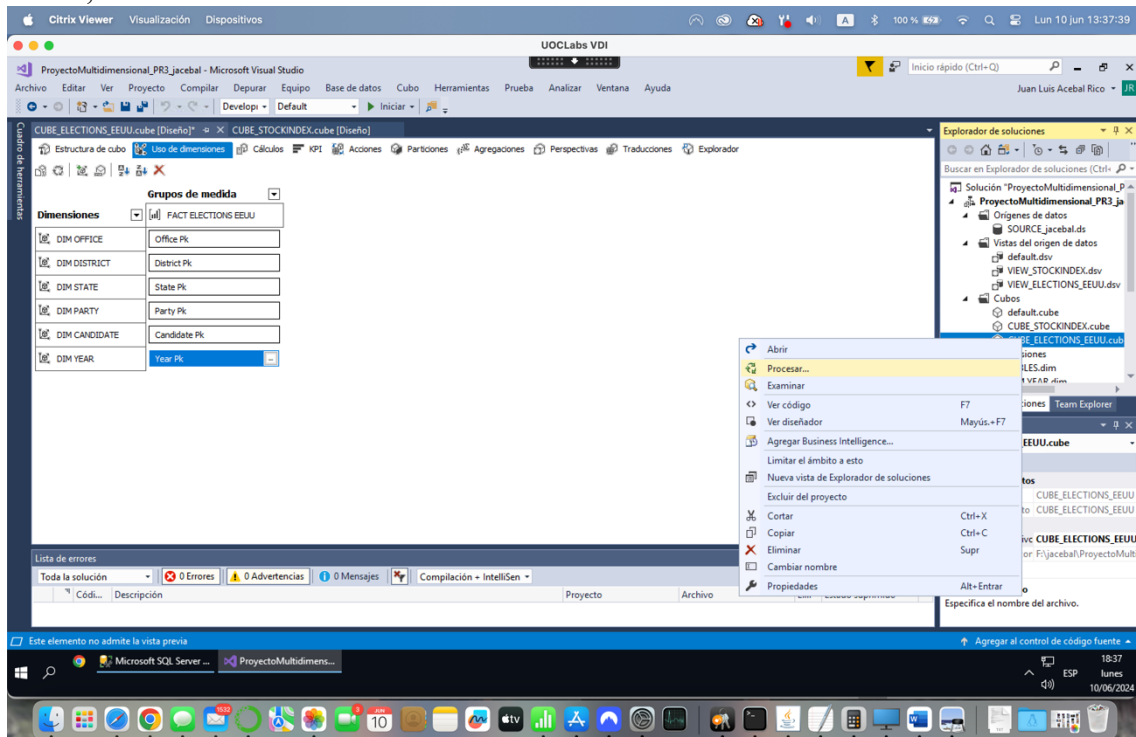


Ahora hay que seleccionar la primary key de la dimensión con la foreign key de la tabla de hechos:

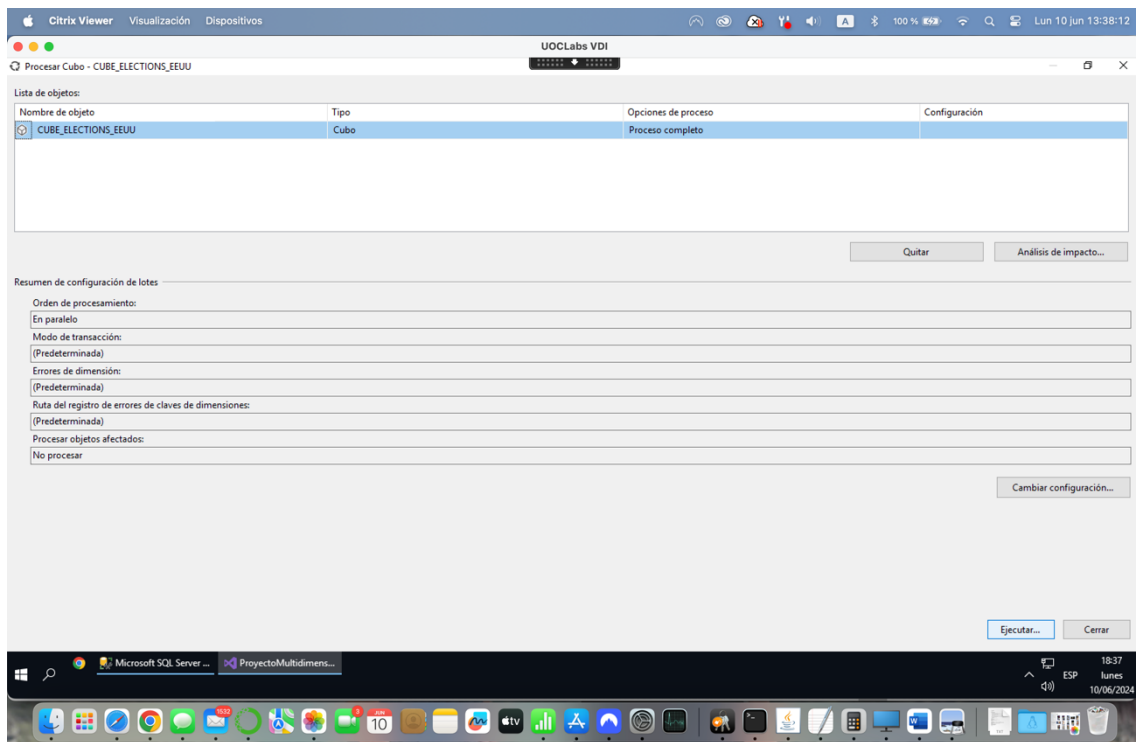


PROCESAR

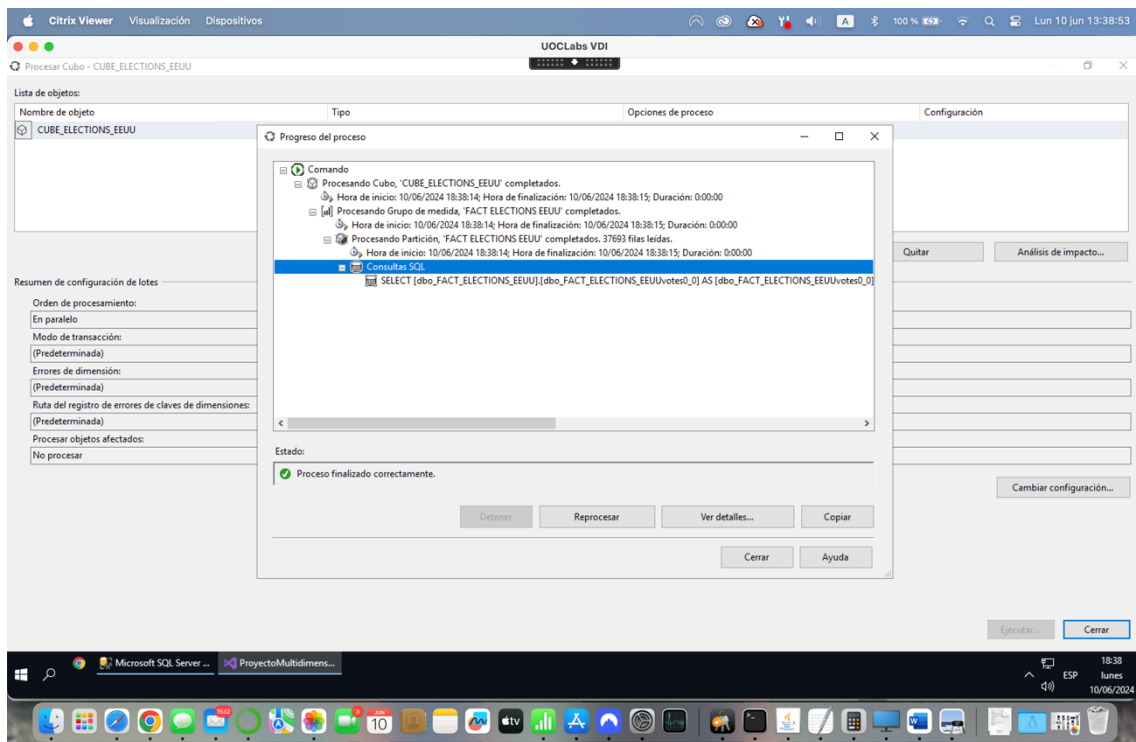
Solamente quedaría procesar los cambios de eliminar dimensión y seleccionarla de nuevo, definiendo las relaciones:



Le damos a ejecutar



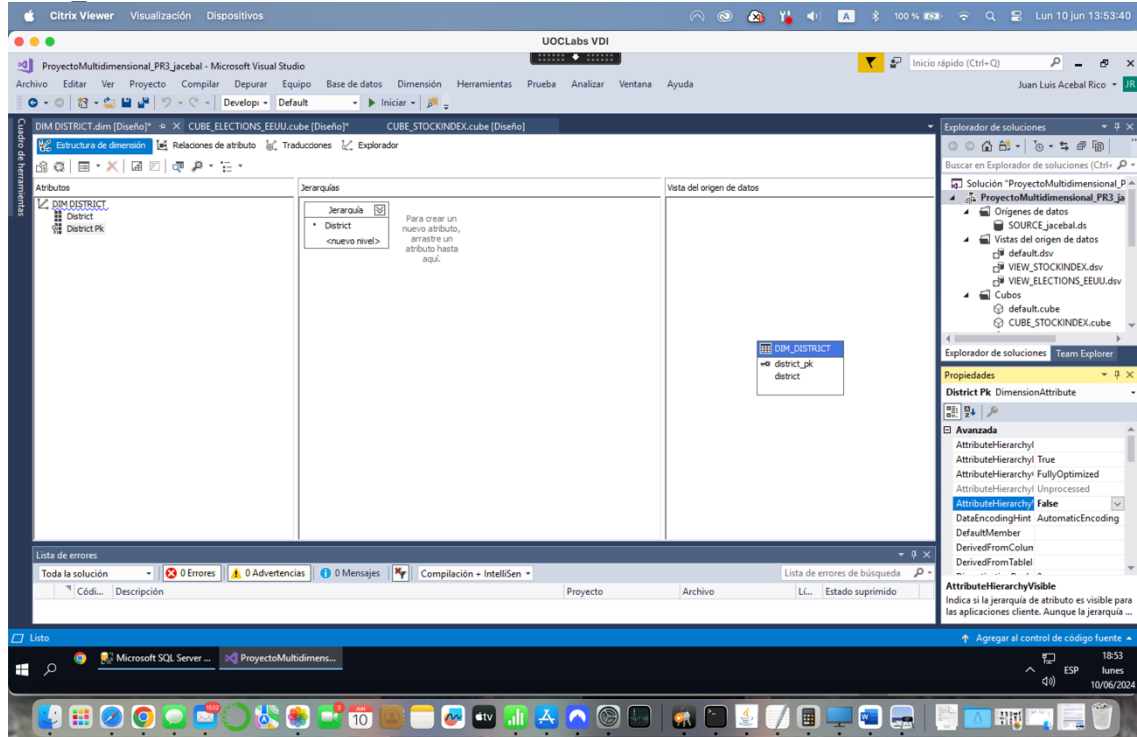
Resultado



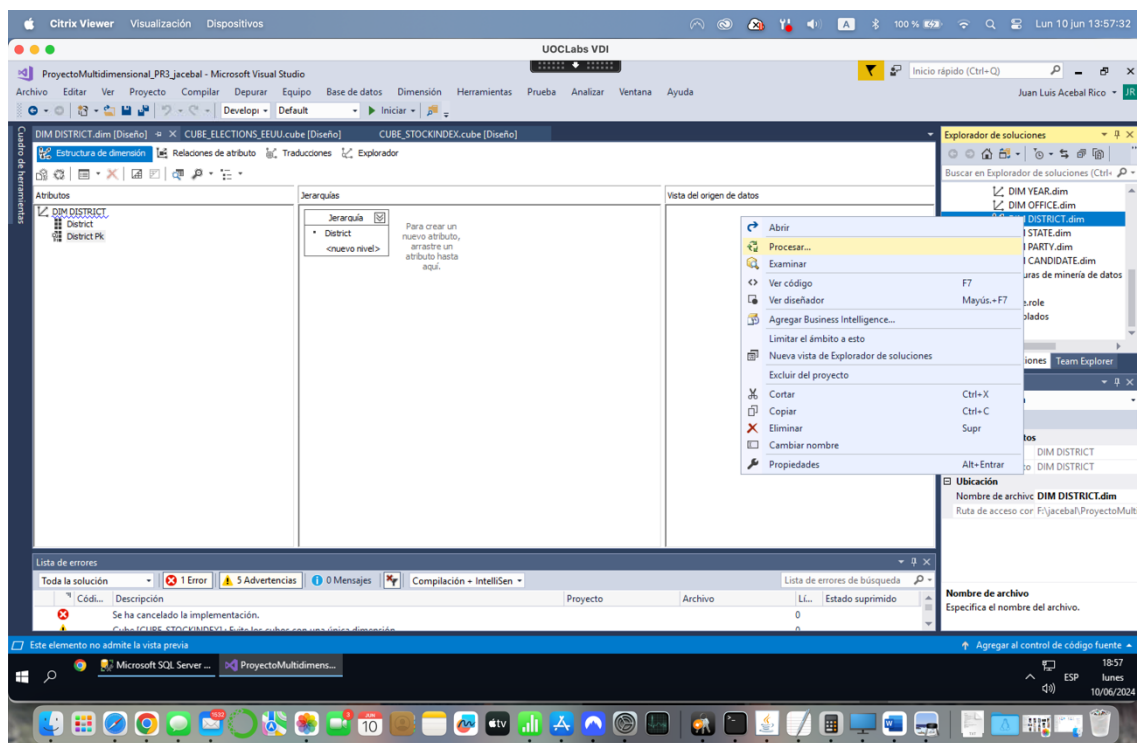
CONFIGURACIÓN DE DIMENSIONES.

En general cada dimensión ocultaremos las claves primarias, y seleccionaremos las jerarquías que usaremos para las consultas.

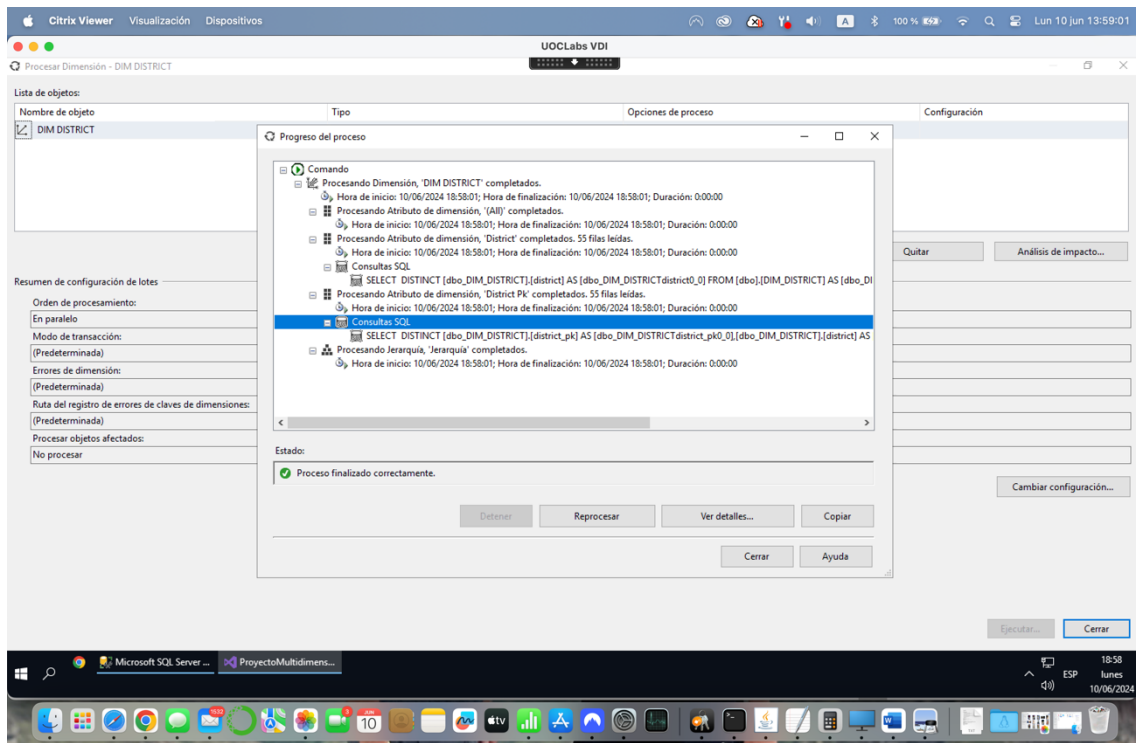
DIM DISTRICT



PROCESAMOS

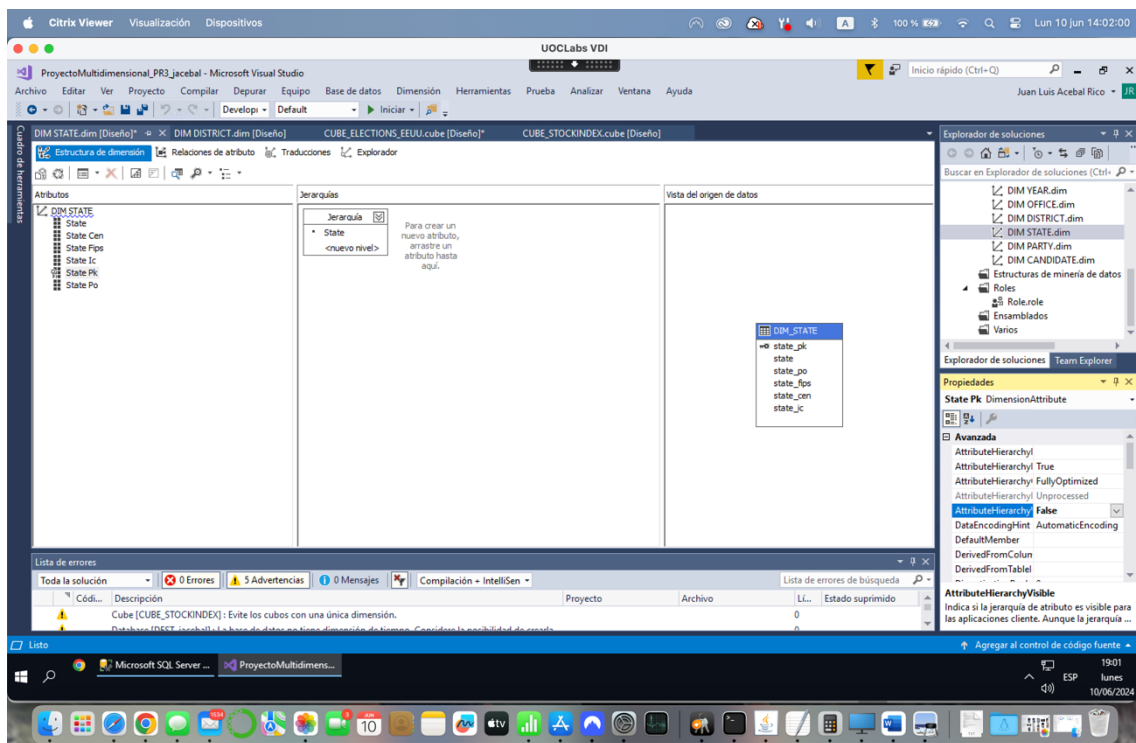


RESULTADO



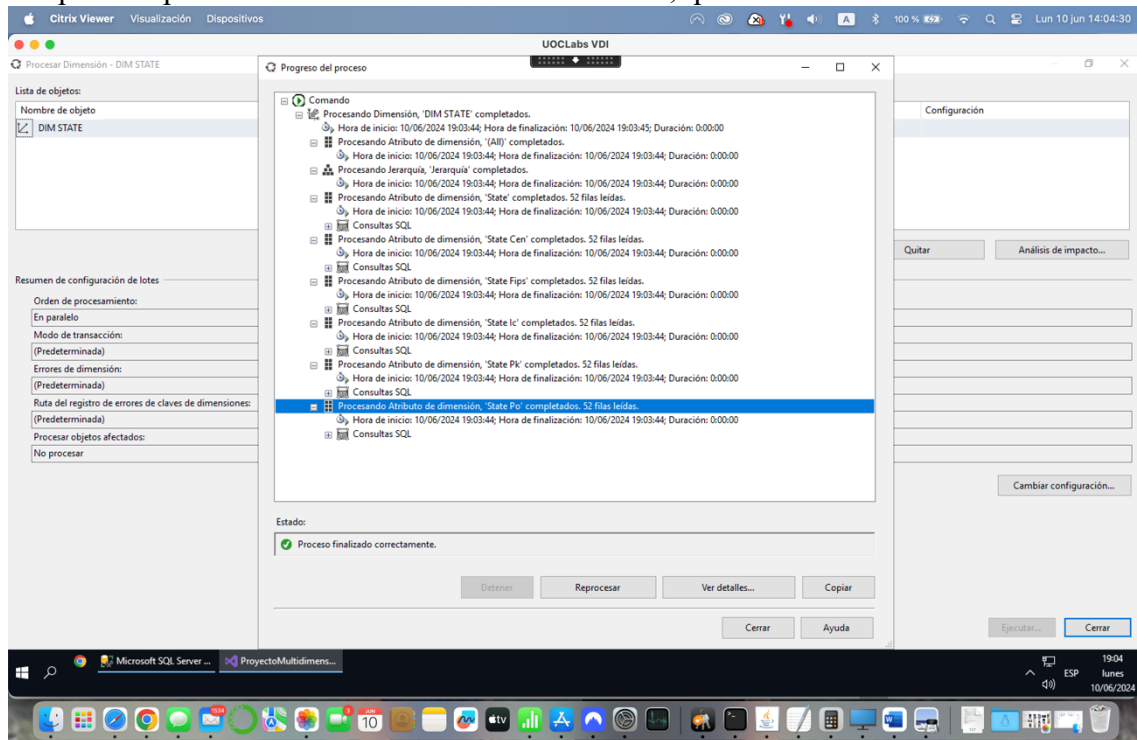
DIM_STATE

Aquí selecciono como jerarquía state, no incluyo mas en la jerarquía ya que todos los atributos state son “sinónimos” entre ellos, y oculto la primary key



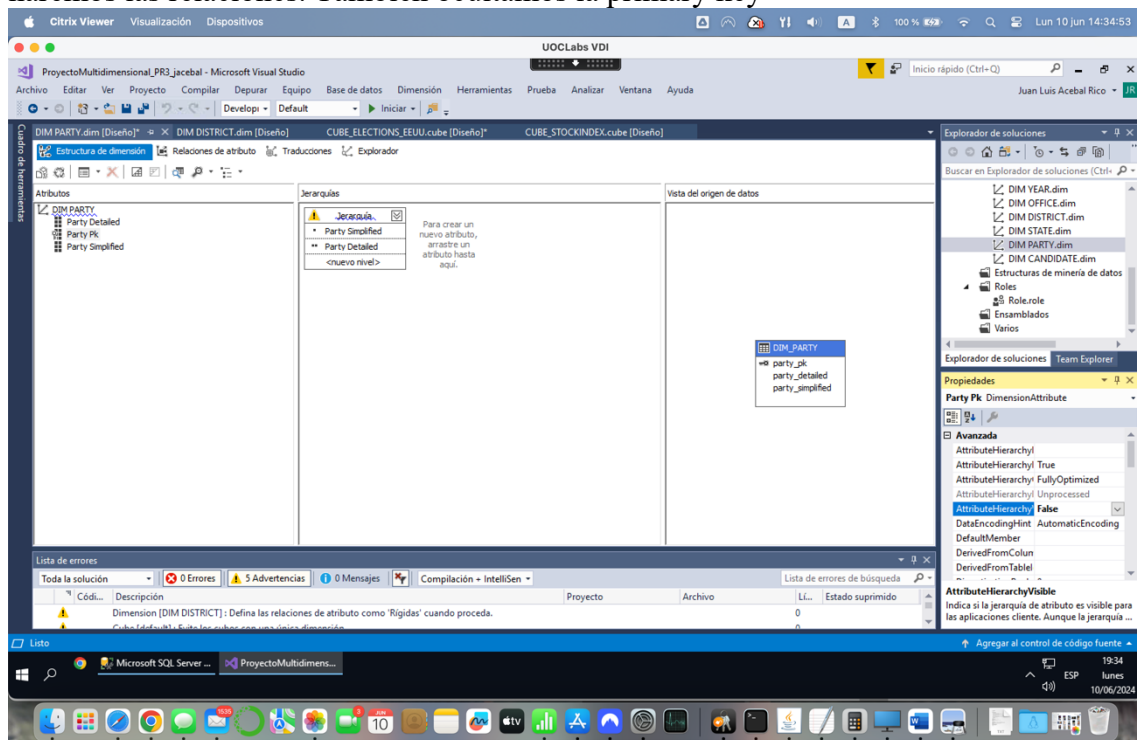
RESULTADO

Después de procesar como en la anterior dimensión, quedaría así

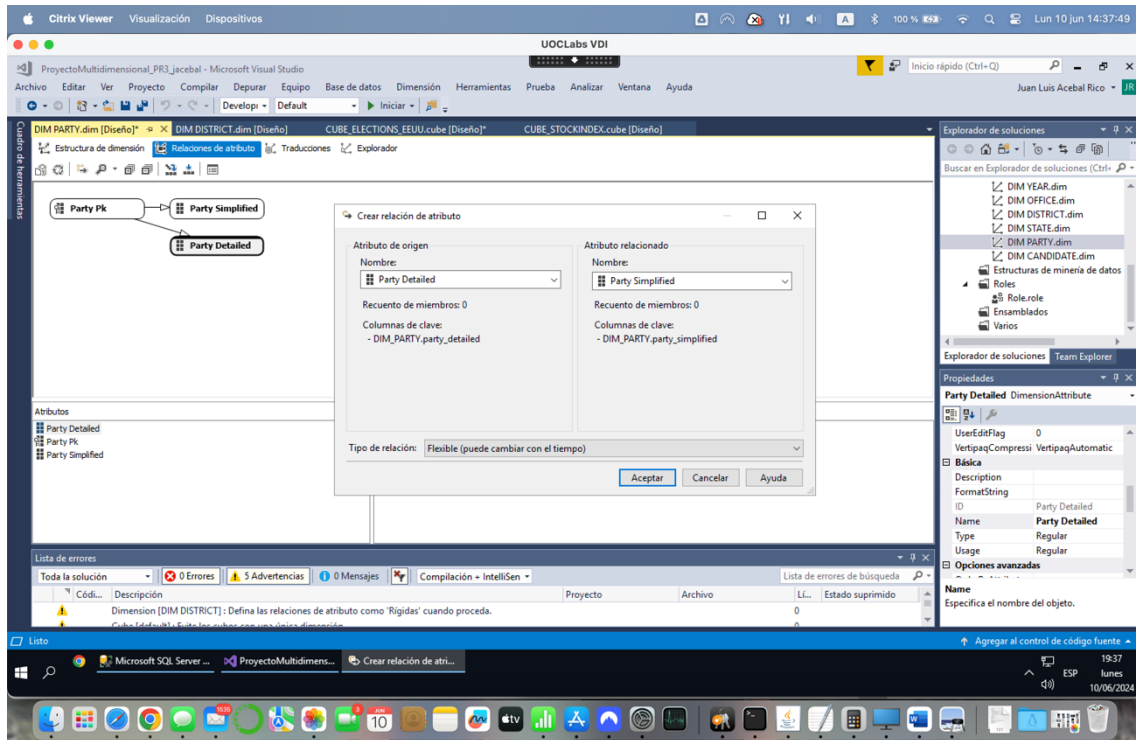


DIM_PARTY

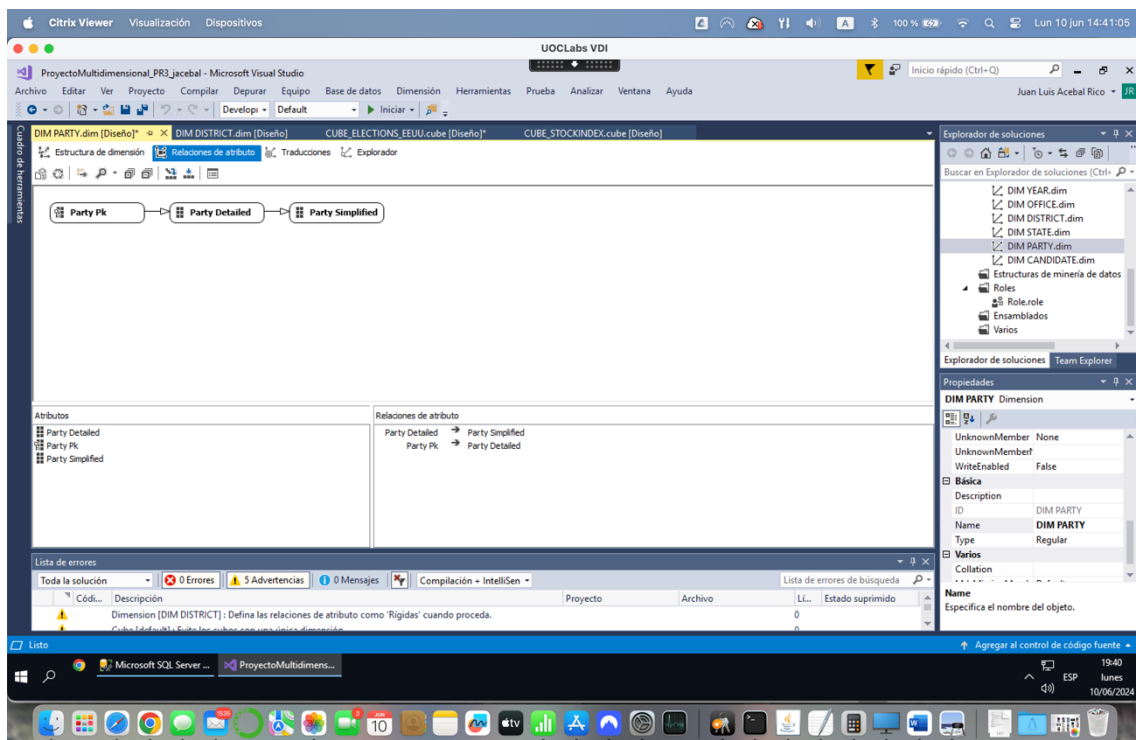
Aquí tenemos que, party_simplified le corresponde de 1 a n party_detailed, y al contrario sería 1 a 1, entonces pondremos primero en la jerarquía party_simplified y haremos las relaciones. También ocultamos la primary key



Creamos la relación de atributo. En mi caso la considero flexible, pero puede ser igualmente rígida, ya que podría cambiar o no con el paso del tiempo, ya que party_simplified contempla los partidos más importantes, más OTHERS, y si bien desde 1976 no ha cambiado, podría cambiar como ha cambiado en la historia de EEUU, realmente no es probable que cambie en muchos años, pero puede pasar.

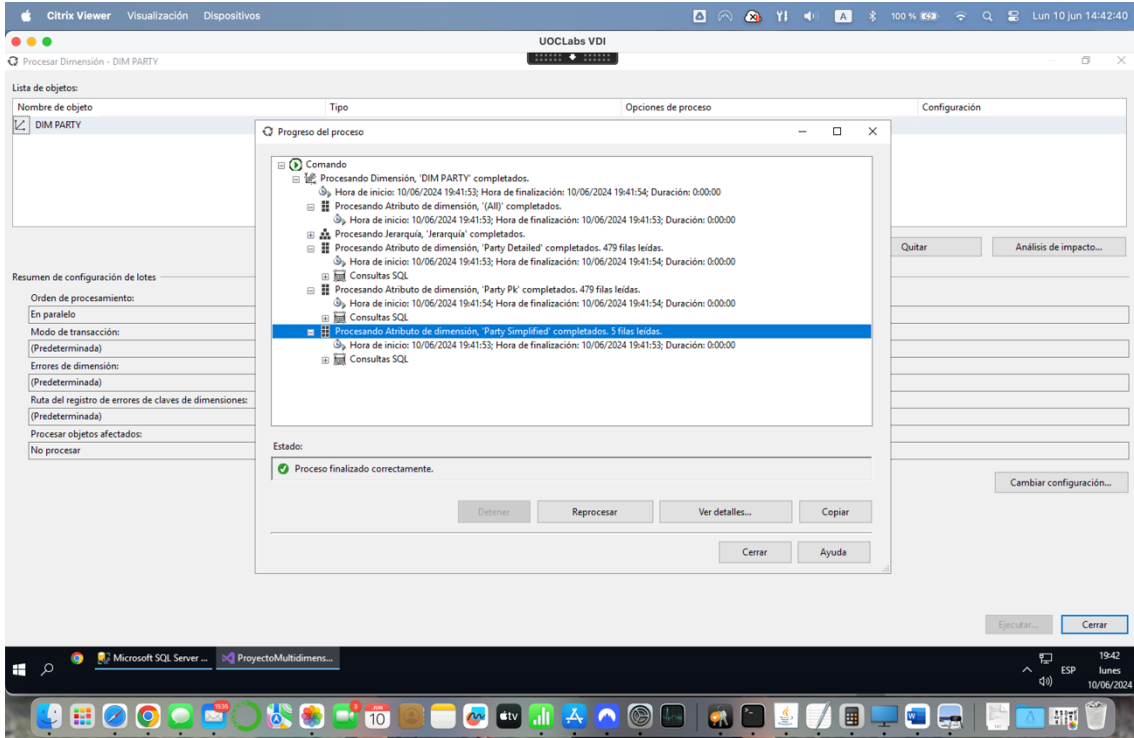


Relación final:

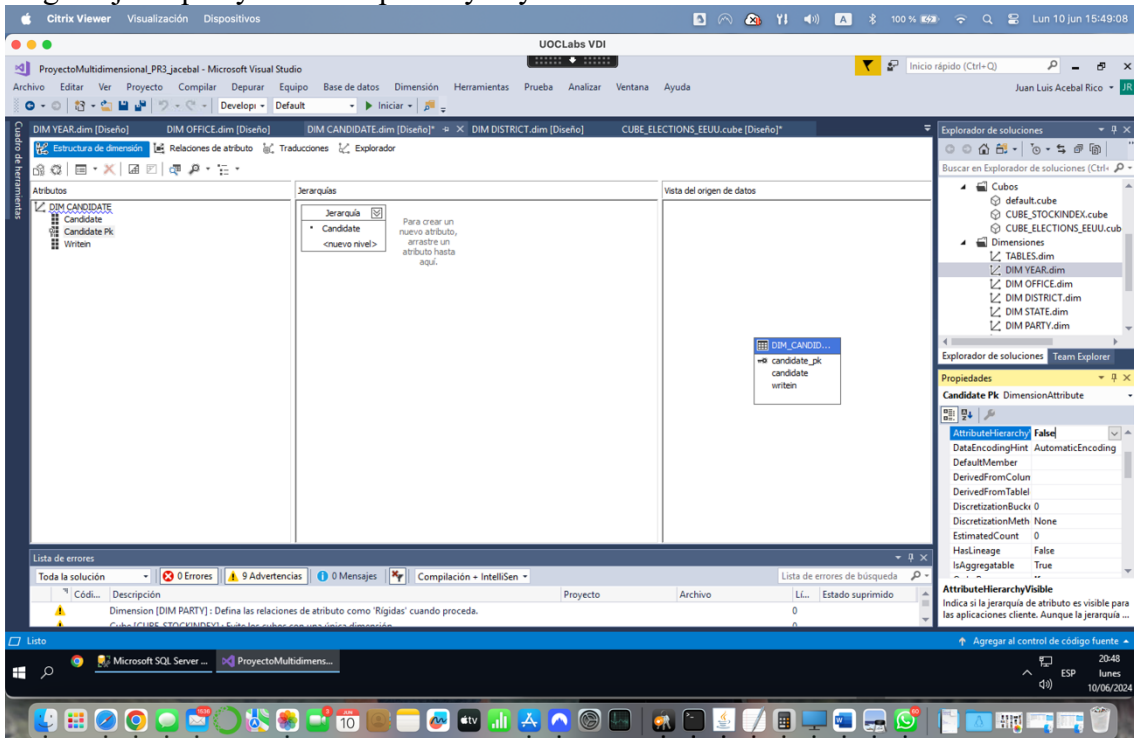


RESULTADO

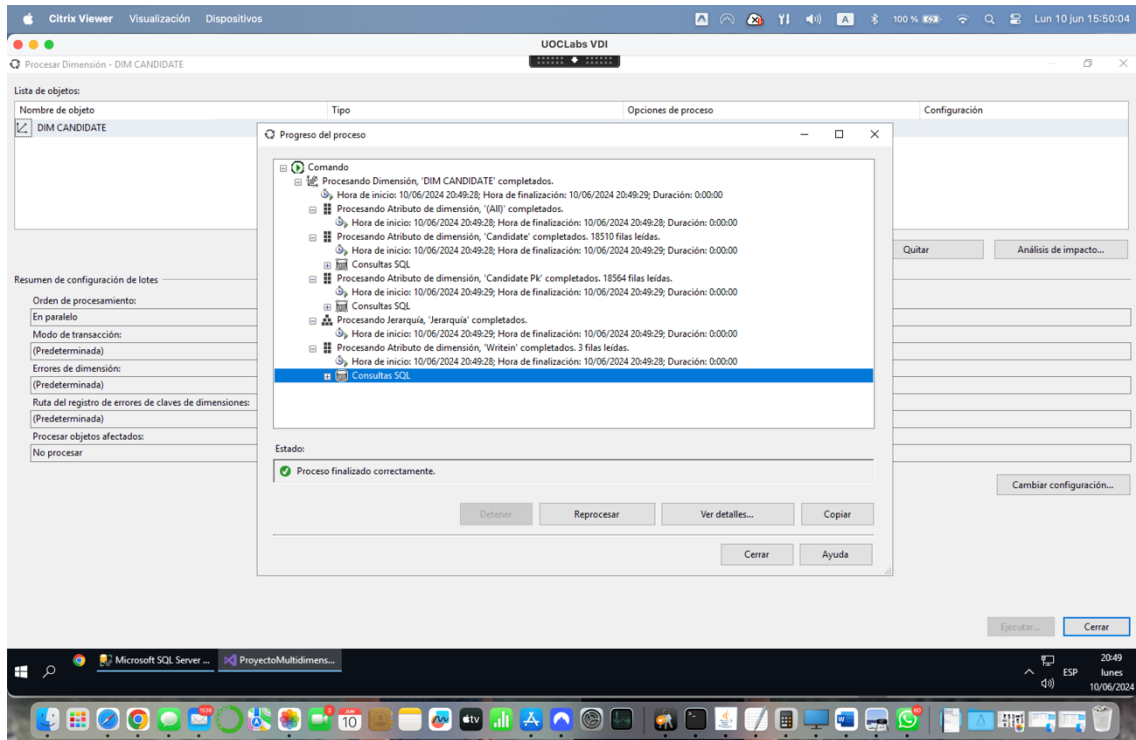
Proceso y muestro el resultado



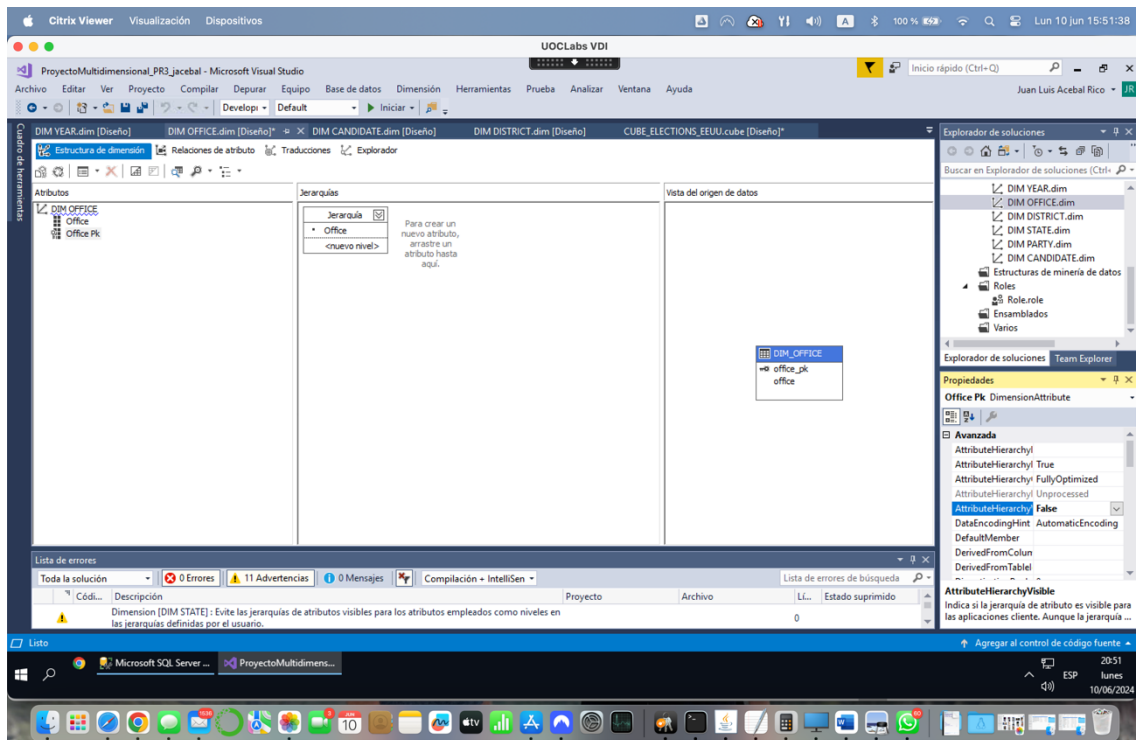
DIM_CANDIDATE
Hago la jerarquía y oculto la primary key



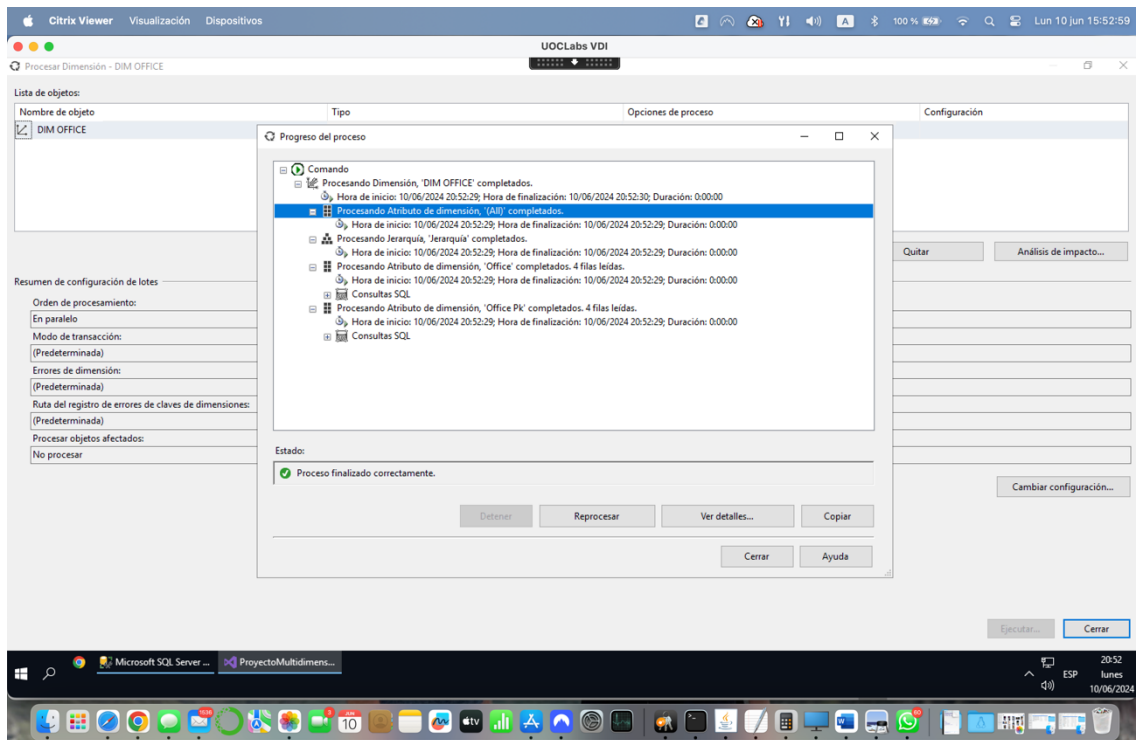
RESULTADO FINAL



DIM_OFFICE
Hago igualmente jerarquía y ocultar primary key

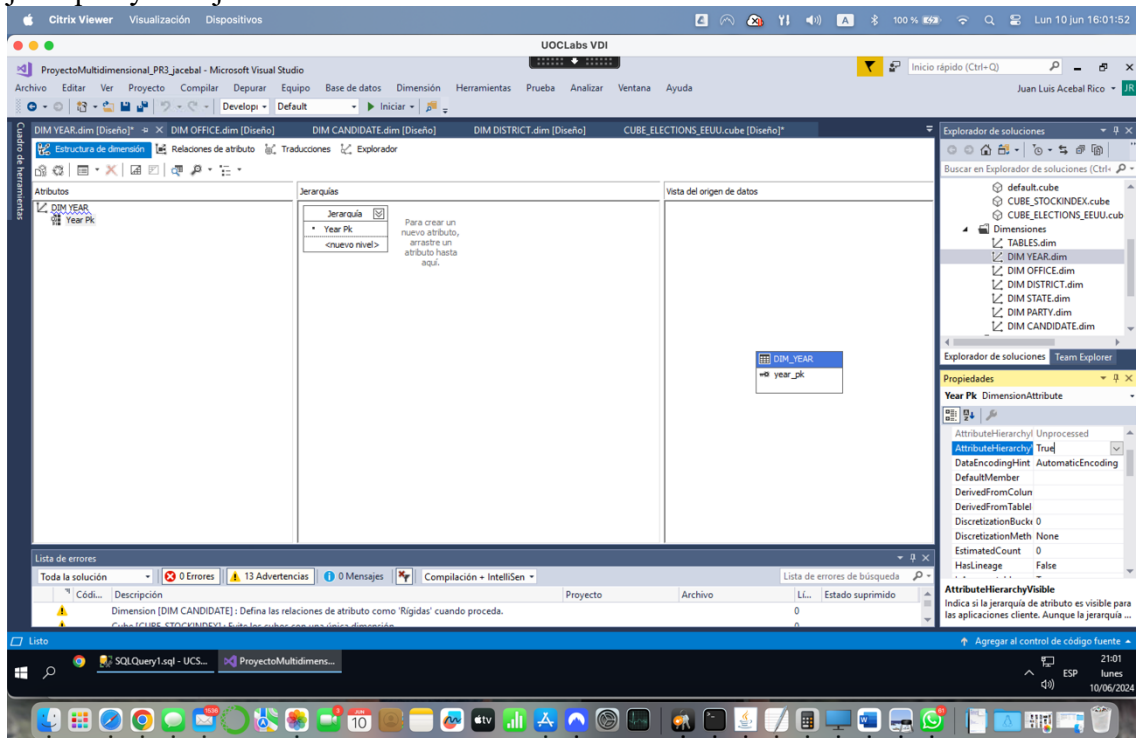


RESULTADO

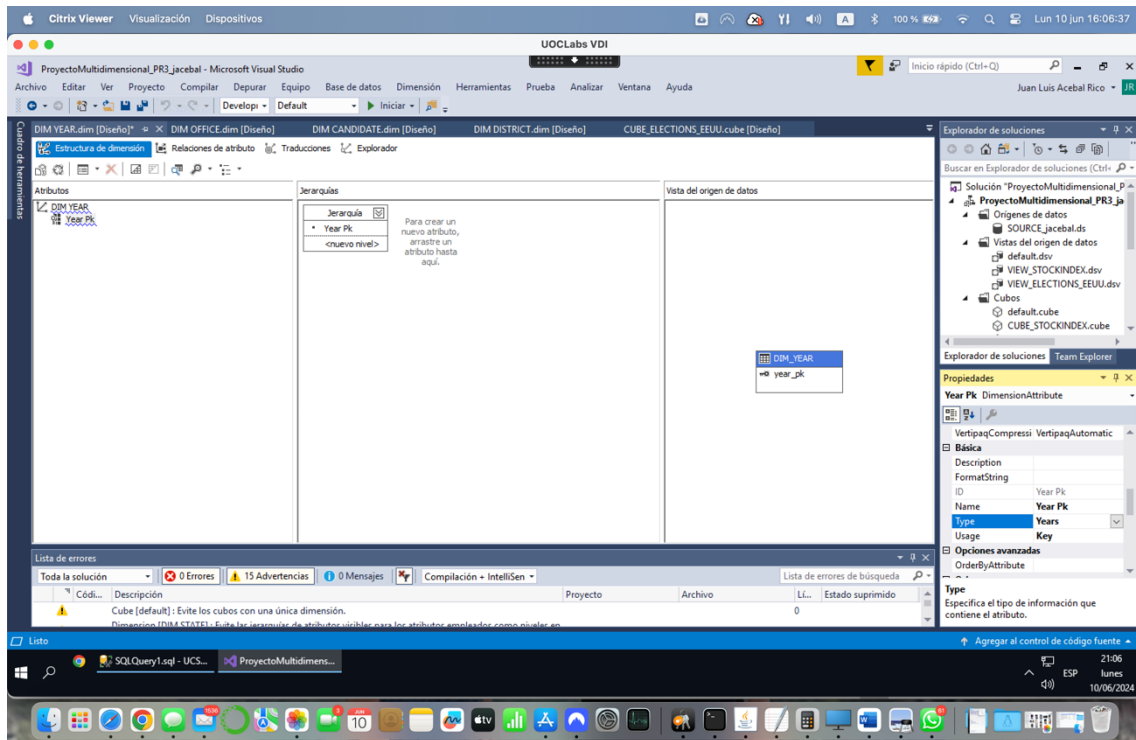


DIM_YEAR

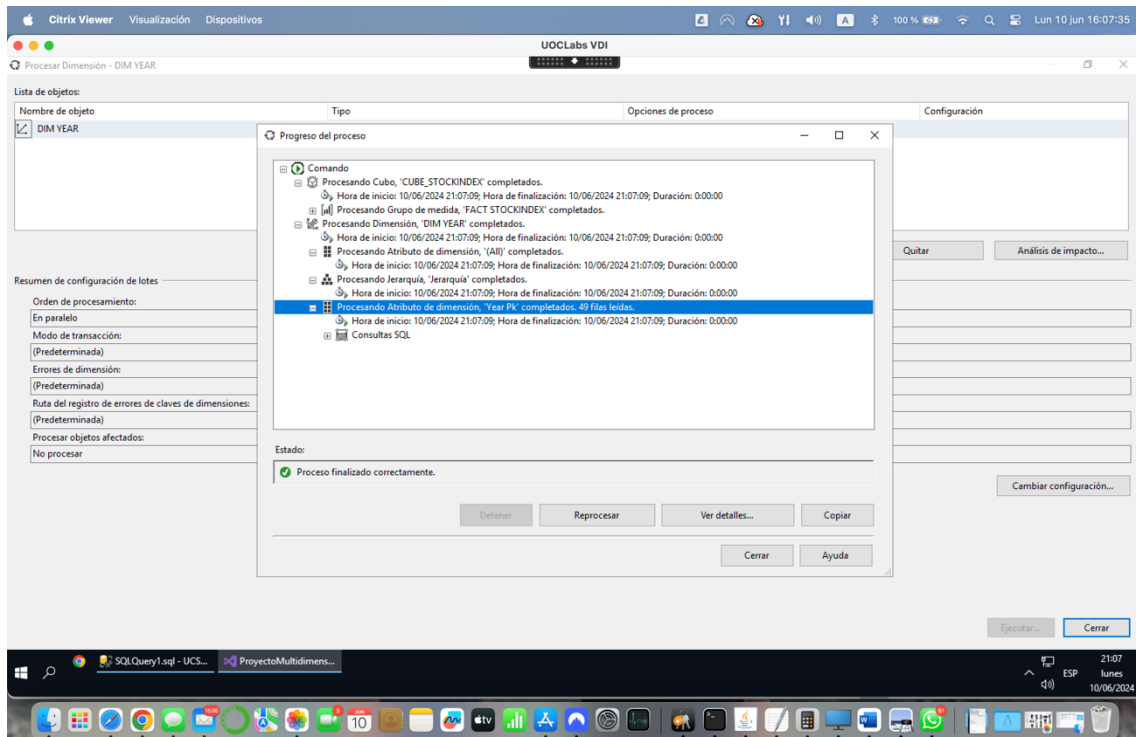
Ultima dimensión a configurar, aquí tenemos que tener en cuenta que la primary key tiene semántica, es decir, tiene información como atributo, entonces la usamos para la jerarquía y la dejamos como visible



Ademas ponemos tipo de dato year:

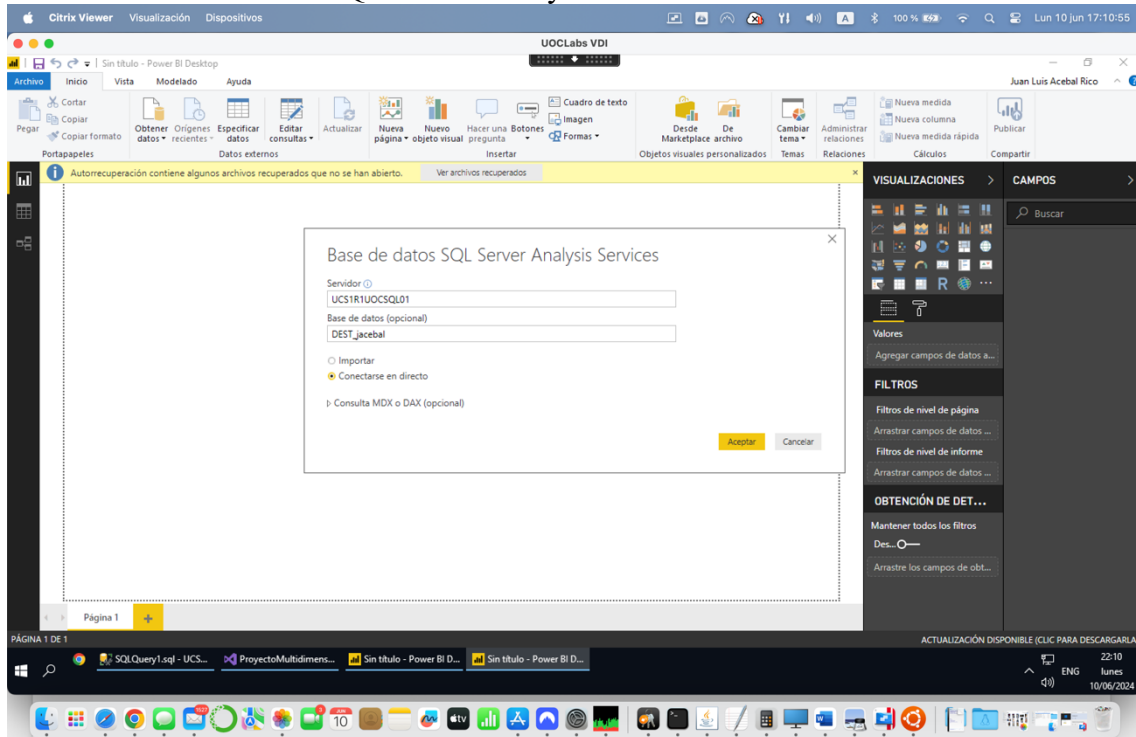


RESULTADO

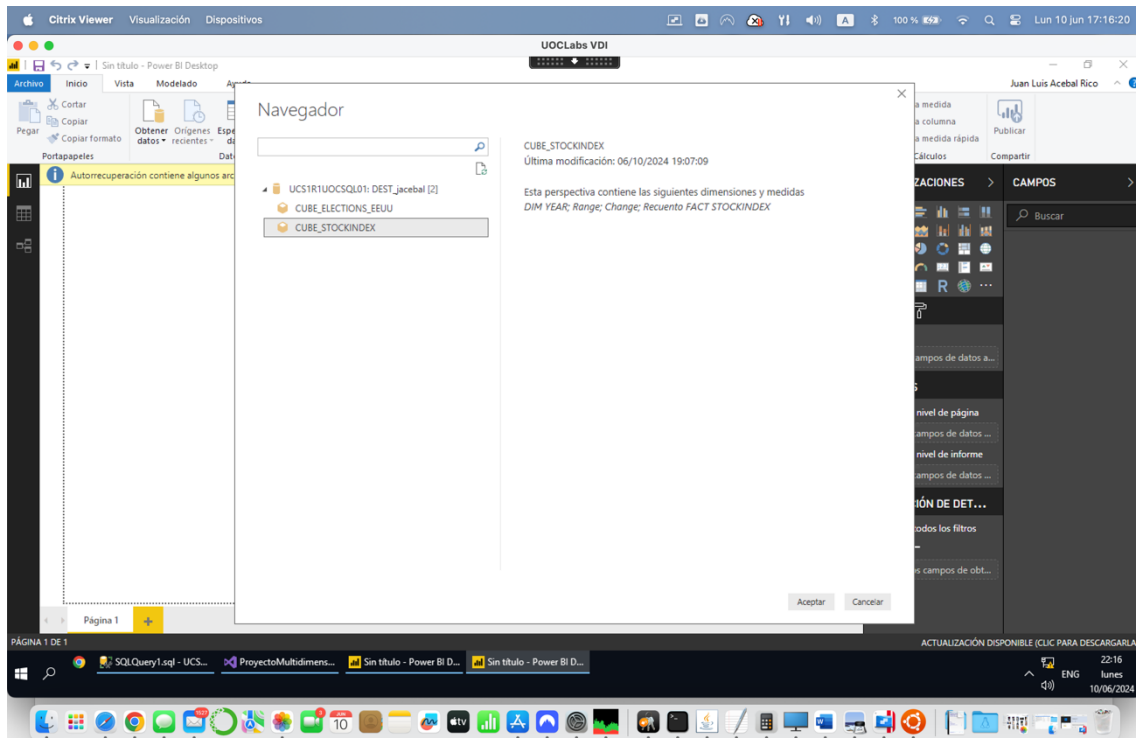


REALIZACIÓN DE CONSULTAS EN POWER BI

Como pasos previos si nunca se ha utilizado power BI, hay que conectarse a una cuenta de Microsoft y registrarse en el servicio. Después, lo primero es conectar power Bi a nuestra base de datos de SQL Server Analysis Services

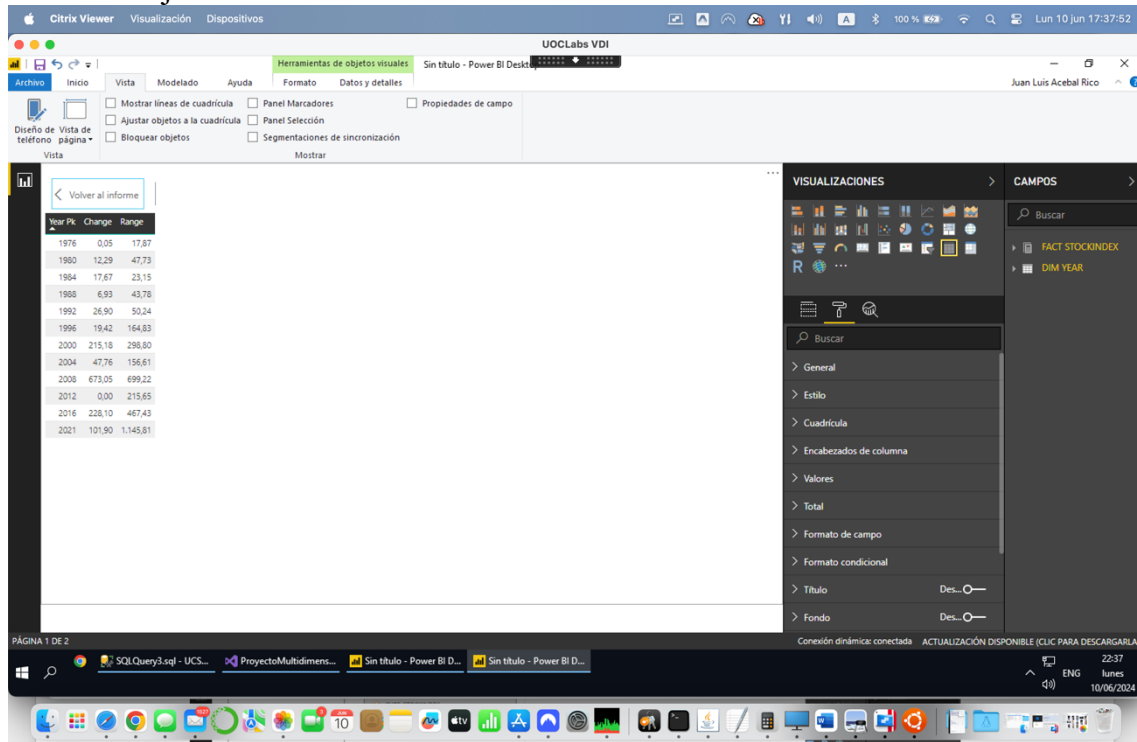


Ahora seleccionamos el cubo a utilizar. Para la primera consulta seleccionar'e el cubo del SP500

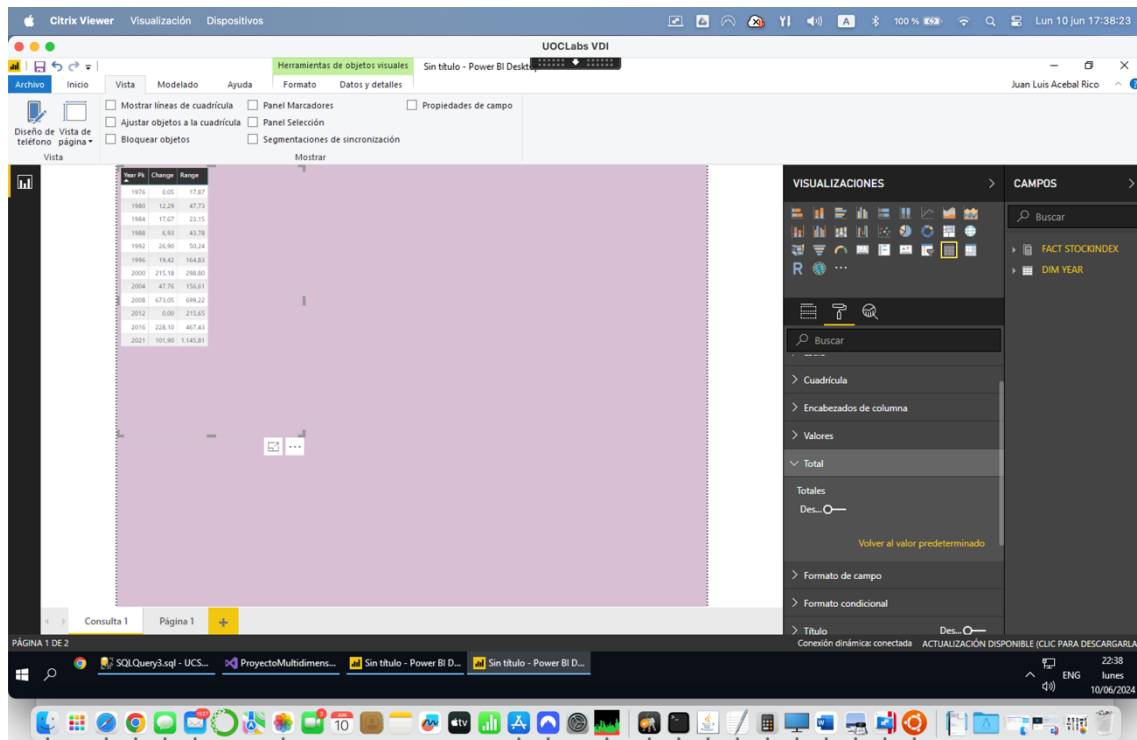


1.EVOLUCIÓN DEL ÍNDICE SP500 A LO LARGO DEL TIEMPO

Vista del objeto

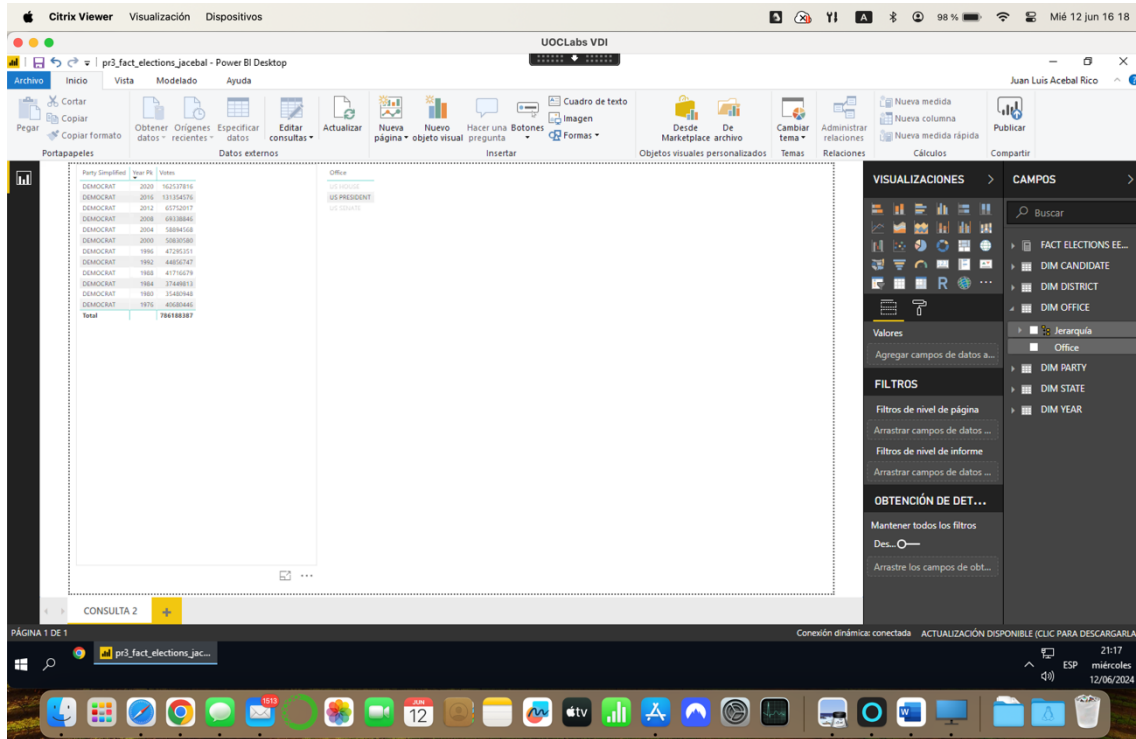


Vista dashboard



2.EVOLUTIVO DE VOTOS DEL PARTIDO “DEMOCRAT” A LO LARGO DEL TIEMPO PARA LAS ELECCIONES A LA PRESIDENCIA DE LOS EEUU. EL RESULTADO SE MOSTRARÁ ORDENADO POR AÑOS DE MANERA DESCENDENTE

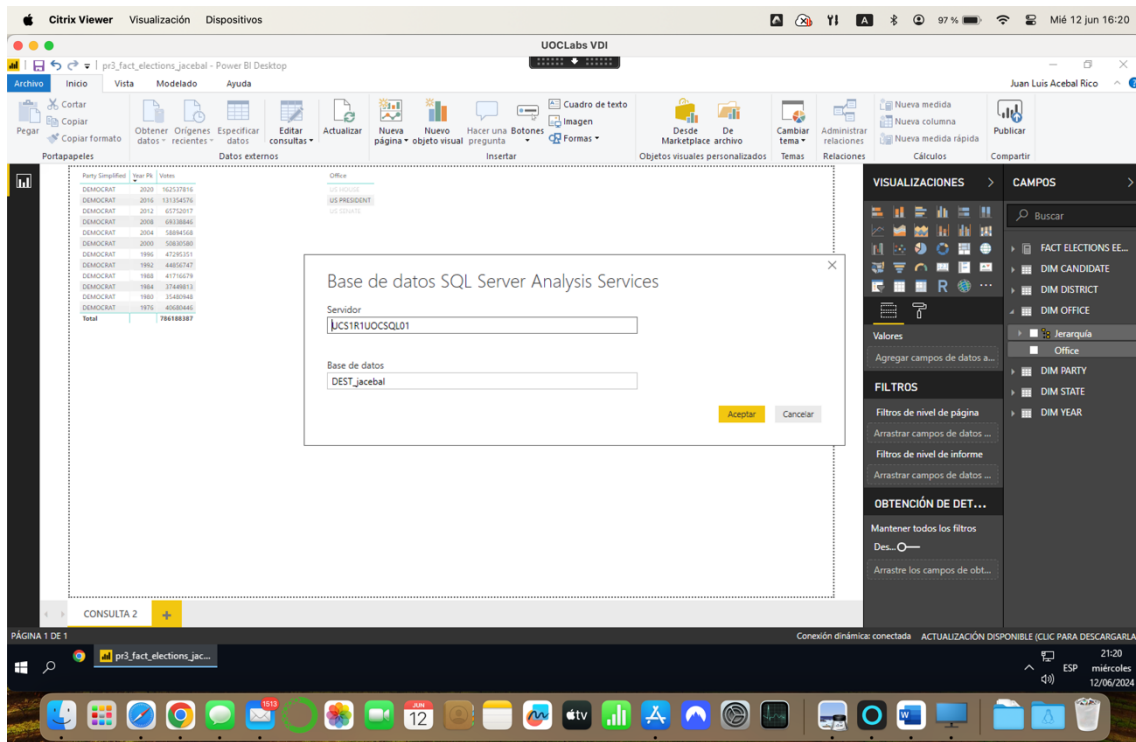
Primero creo un nuevo power bi, selecciono el otro cubo, y selecciono los campos, y uso como filtro party_simplified = democrat, por ultimo filtro el campo por office



Zoom:

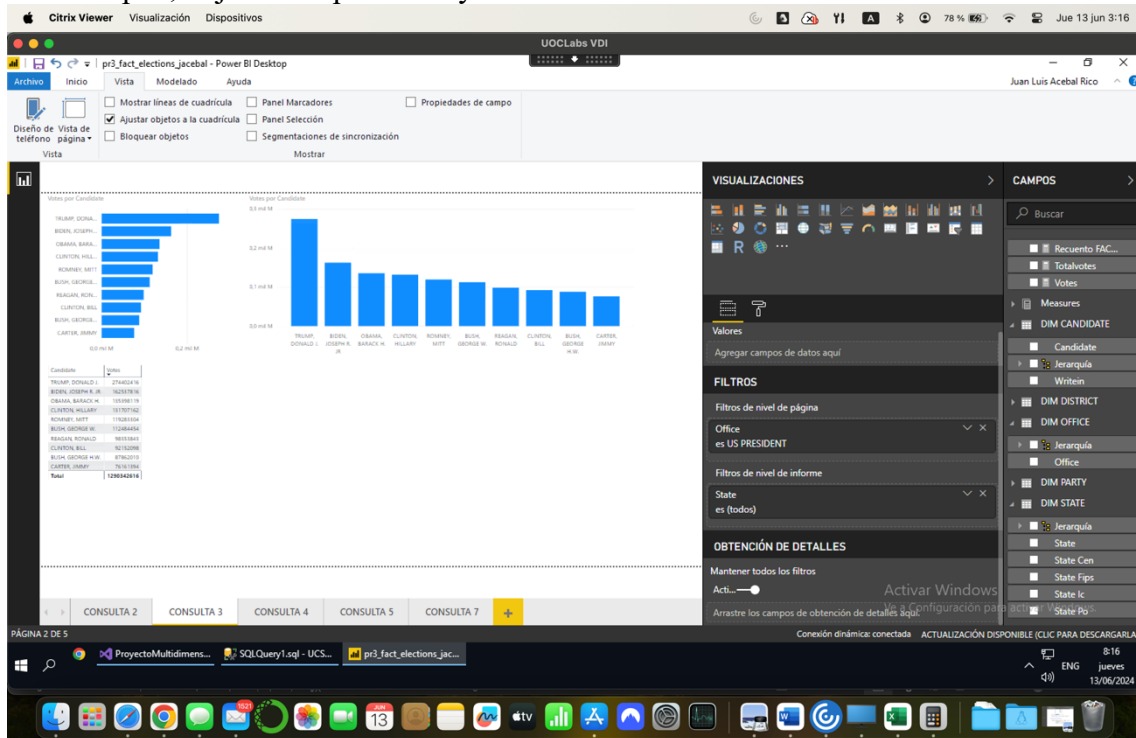
Party Simplified	Year Pk	Votes
DEMOCRAT	2020	162537816
DEMOCRAT	2016	131354576
DEMOCRAT	2012	65752017
DEMOCRAT	2008	69338846
DEMOCRAT	2004	58894568
DEMOCRAT	2000	50830580
DEMOCRAT	1996	47295351
DEMOCRAT	1992	44856747
DEMOCRAT	1988	41716679
DEMOCRAT	1984	37449813
DEMOCRAT	1980	35480948
DEMOCRAT	1976	40680446
Total		786188387

Fuentes de datos (no había hecho la captura antes de hacer el dashboard en la creación del archivo pbi):



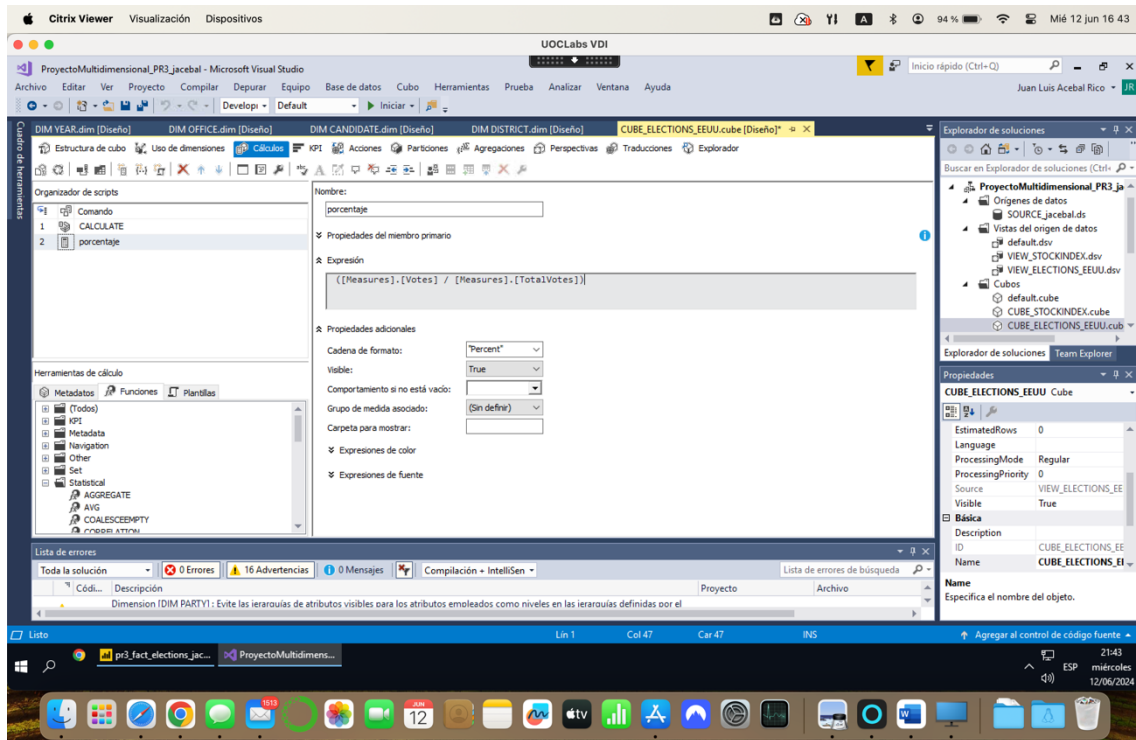
3.TOP TEN DE CANDIDATOS MÁS VOTADOS EN LAS ELECCIONES PRESIDENCIALES Y ELECCIÓN

Filtro Top N, elijo office president y nada mas.

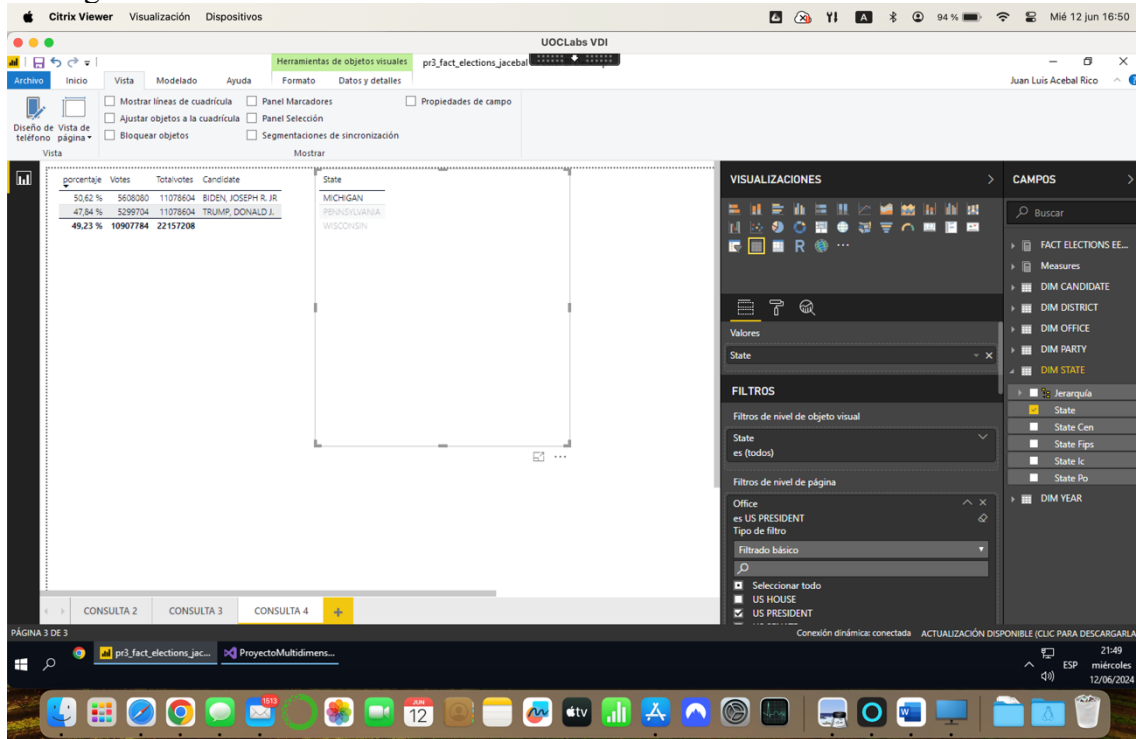


4. COMPARACIÓN DE LOS RESULTADOS DE LAS ELECCIONES DE 2016 Y 2020 EN LOS ESTADOS DE MICHIGAN, PENNSYLVANIA Y WISCONSIN PARA LOS CANDIDATOS DE LOS PARTIDOS DEMÓCRATAS “DEMOCRAT” Y REPUBLICANO “REPUBLICAN”. INCLUYE EL % SOBRE EL TOTAL DE VOTOS DETERMINANTES PARA LA VICTORIA DE JOE BIDEN EN 2020

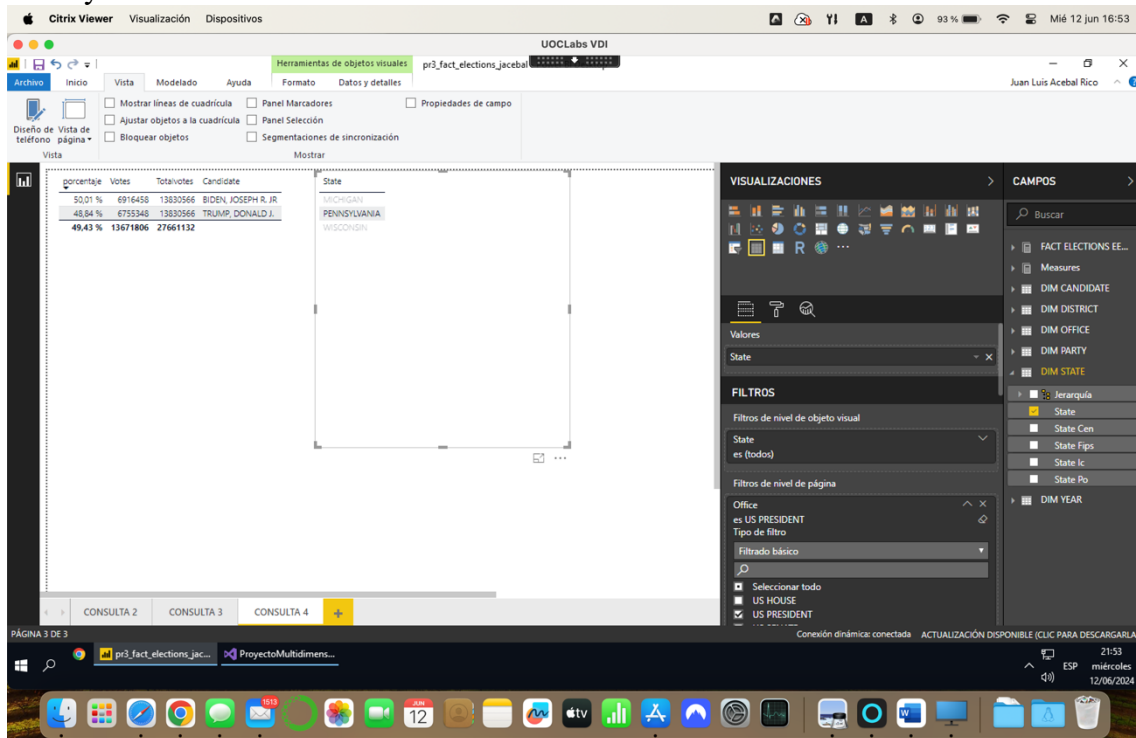
Creo un campo calculado, pongo la expresión del cálculo, y tipo de cadena porcentaje



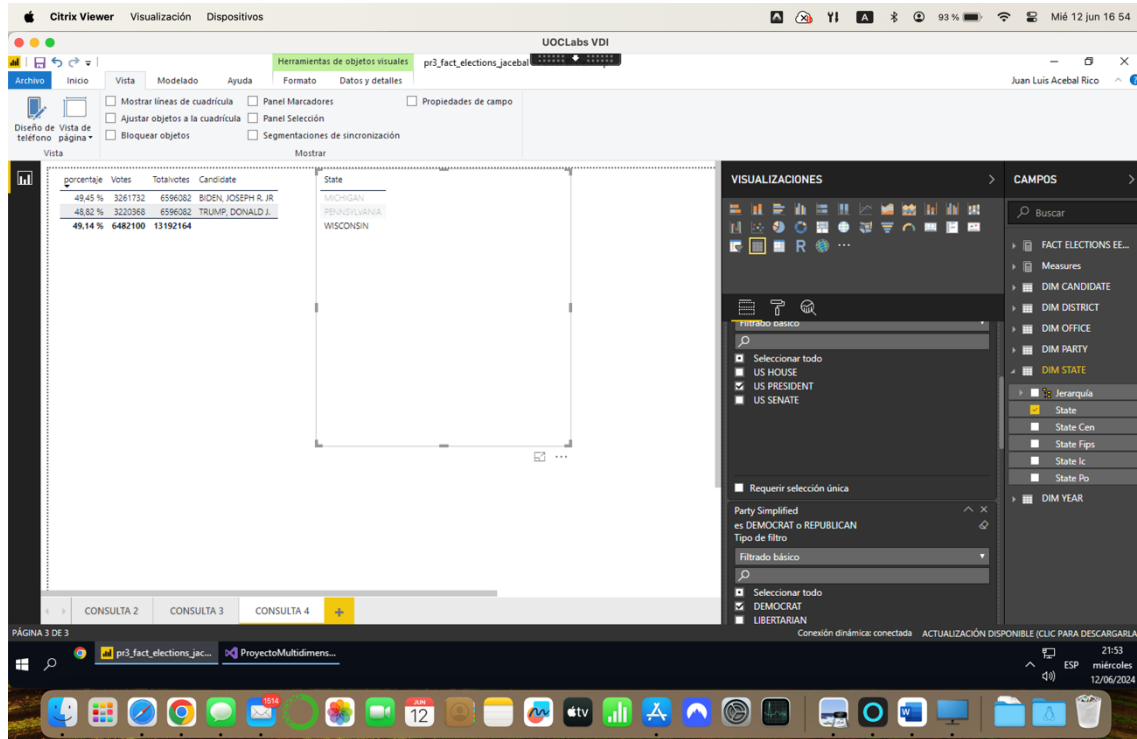
Michigan



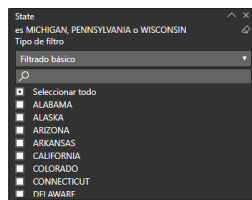
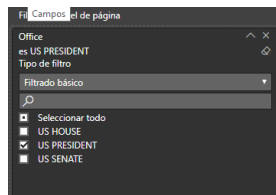
Pennsylvania



Wisconsin



Los filtros utilizados:



5.TOP FIVE DE AÑOS CON MAYOR PARTICIPACIÓN ELECTORAL EN LA PRESIDENCIALES POR EL ESTADO DE CALIFORNIA

Usamos WITH para poner las condiciones usando filter y topcount respectivamente.

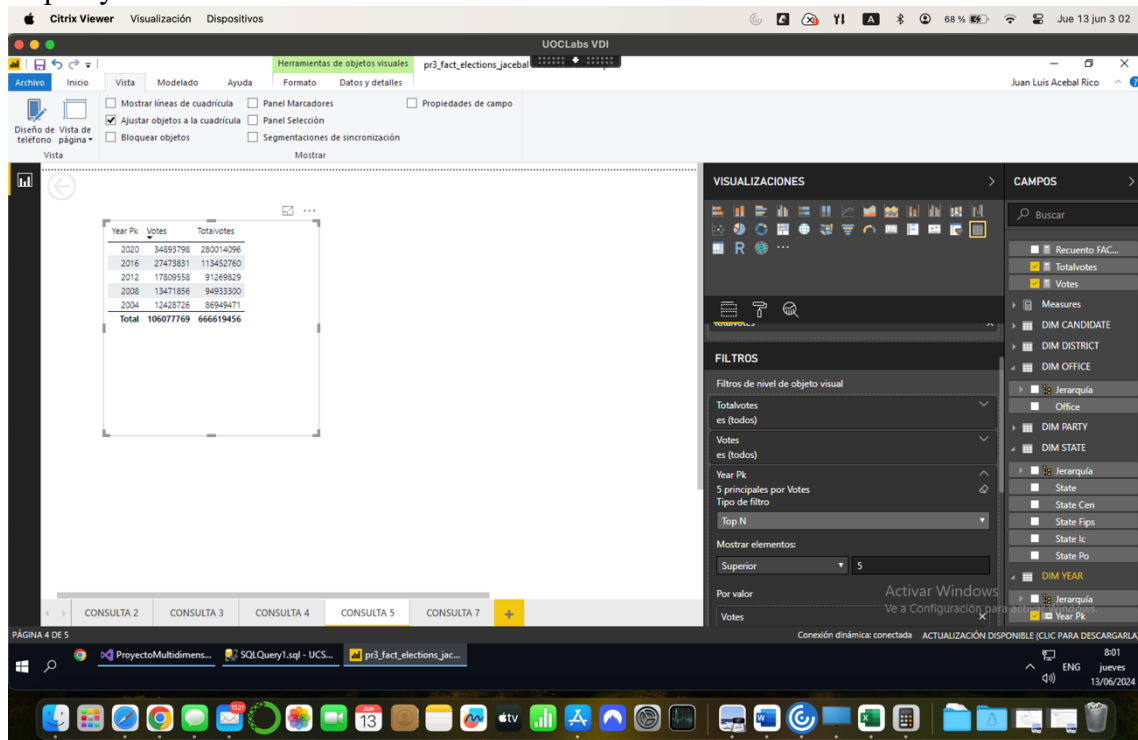
The screenshot shows the Microsoft Visual Studio interface for a cube editor. The main window displays a SQL query for a cube named 'CUBE_ELECTIONS_EEUU'. The query uses WITH clauses to define a set of years, filter for California, and count the top 5 years by votes. A preview table at the bottom of the query editor shows the results.

```

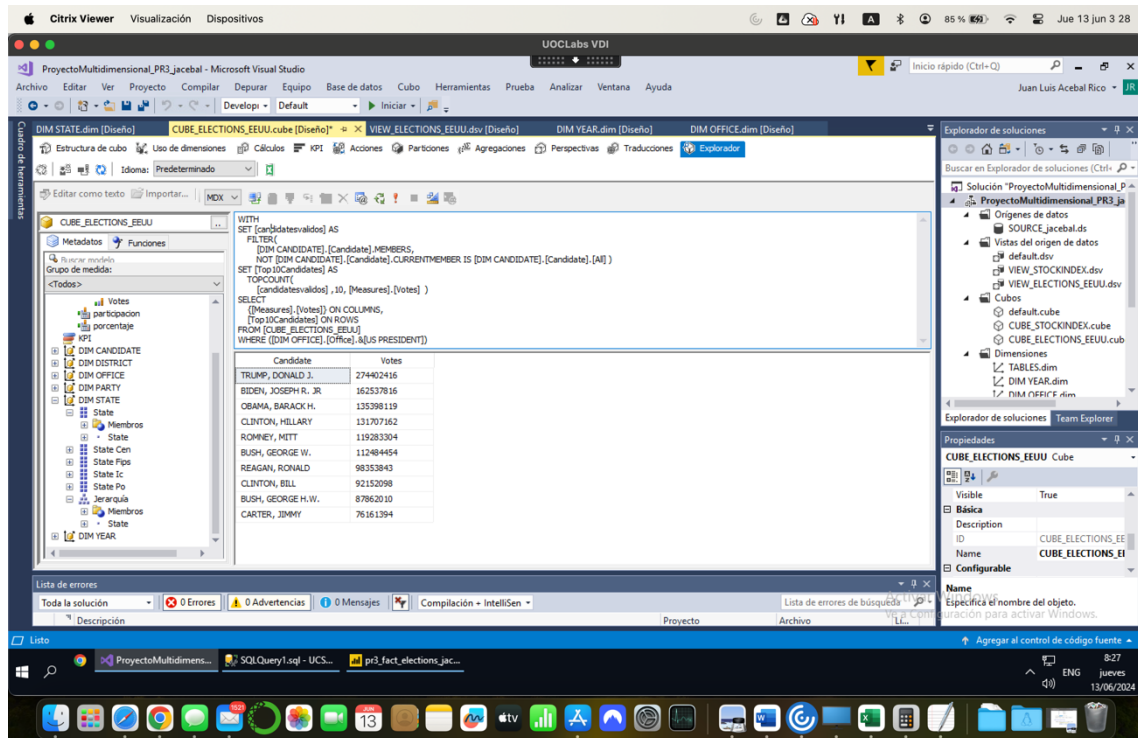
WITH
SET [yearsvaldos] AS
FILTER(
[DIM YEAR].[Year PK].MEMBERS,
NOT [DIM YEAR].[Year PK].CURRENTMEMBER IS [DIM YEAR].[Year PK].[All]
)
SET [Top5years] AS
TOPCOUNT(
[yearsvaldos],
5)
SELECT
([Measures].[votes]) ON COLUMNS,
[Top5years] ON ROWS
FROM [CUBE_ELECTIONS_EEUU]
WHERE (
[DIM STATE].[State].[California],
[DIM OFFICE].[Office].[US PRESIDENT])
    
```

Year PK	Votes
2020	34893798
2016	27473831
2012	17809558
2008	13471856
2004	12428726

En Power Bi quedaría así, donde simplemente hay que seleccionar el filtro de fecha por Top N y el valor votes:

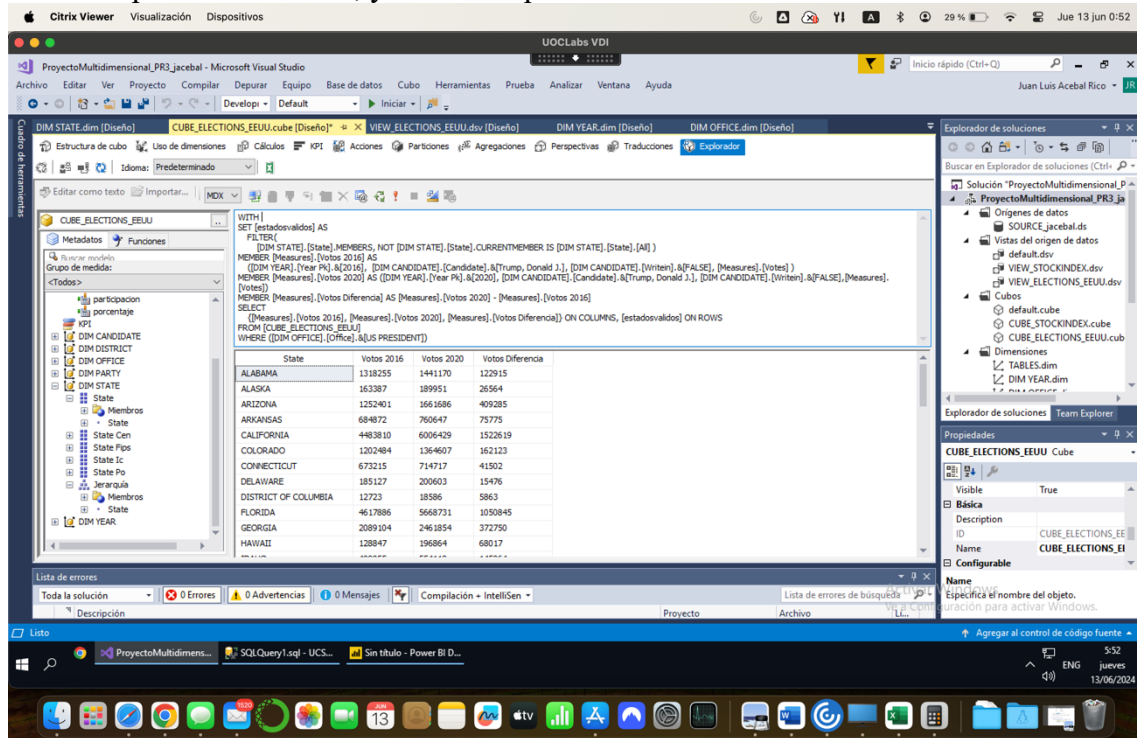


Y en MDX:



6.DIFERENCIA DE VOTOS ENTRE LAS ELECCIONES DE 2016 Y 2020, POR ESTADO, DE DONALD TRUMP A LA PRESIDENCIA DE LOS EEUU

Como pequeña explicación, podemos ver que he usado clausula WITH, para definir partes de la consulta, en este caso un set y un miembro que no tienen persistencia, set lo he usado para hacer un filtro y evitar el agregado de todas las filas, en la que la primera es null para el estado, luego una medida para los votos de 2016, otra para 2020, otro miembro para la diferencia, y un select que utiliza todo esto.



7.COMPARA POR ESTADO LOS RESULTADOS DE “CLINTON, HILLARY” Y “TRUMP, DONALD J.” PARA LAS ELECCIONES A LA PRESIDENCIA DE 2016

He creado esta vez un gráfico con un mapa, y luego una lista de cada estado y cada candidato. Cuando pasas por encima de cada estado puedes ver los votos para cada partido político (que al ser un solo acto electoral son los mismos que el de los candidatos). Entre otras cosas he puesto filtro por fecha, por candidato y writein.

