Universidad Autónoma del Estado de México

Unidad Académica Profesional Tianguistenco

Ingeniería en software

Unidad de aprendizaje:

Datawarehouse

Profesor:

Julieta Garcilazo Reyes

Alumno:

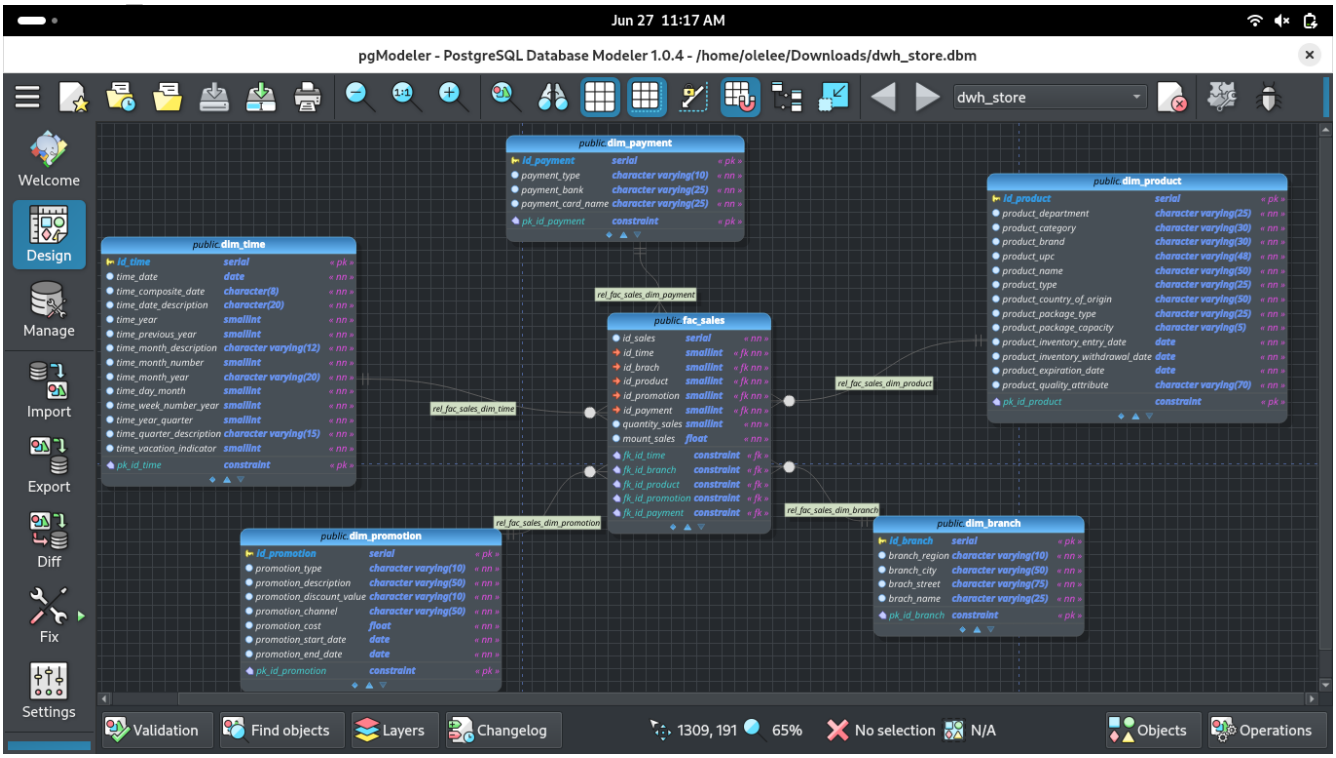
Diego Hernández Romero

Grupo:

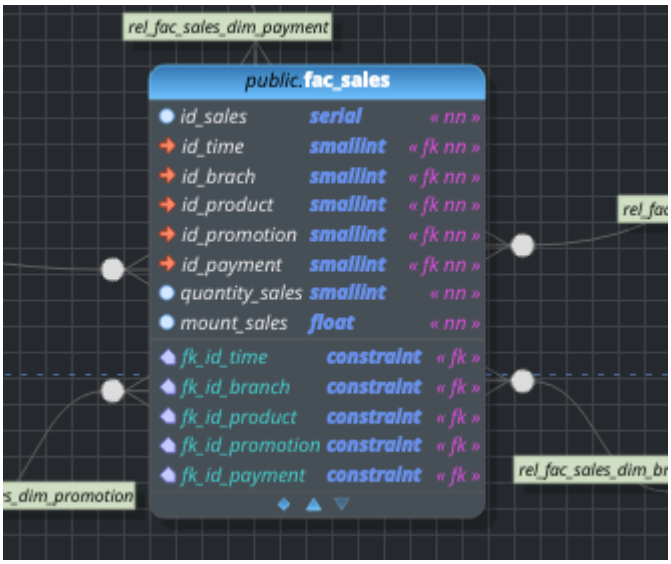
S7

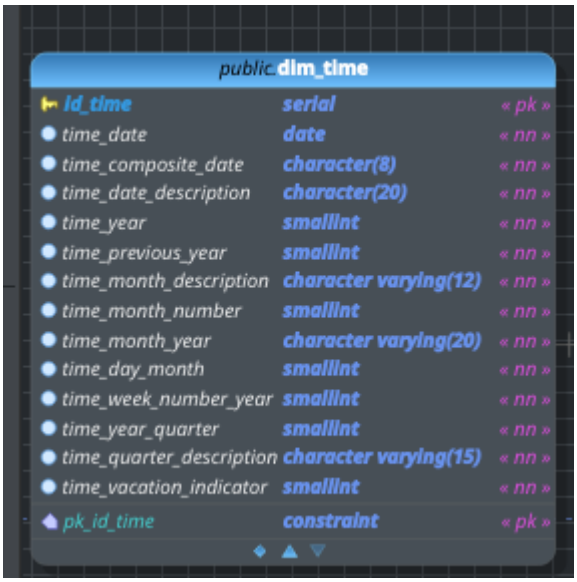
Fecha de entrega: 27/06/2025

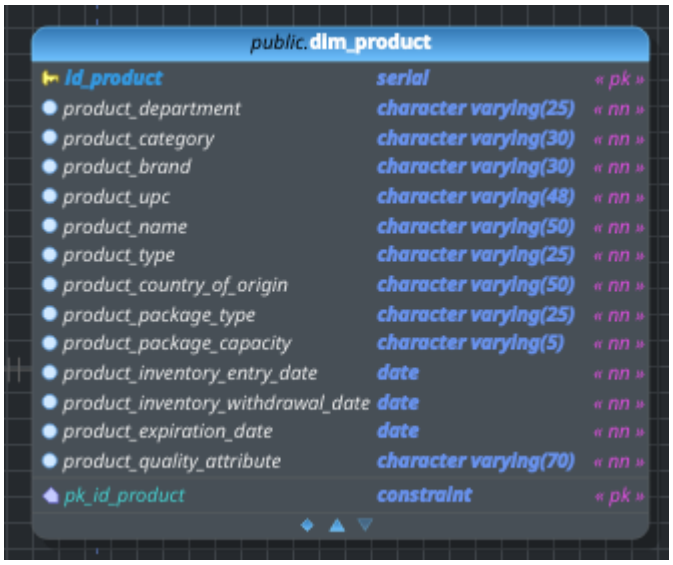
Como primer paso se utilizó pgmodeler para moldear las dimensiones conforme a los diferentes puntos que se obtuvieron del documento donde se establecía la información a obtener.
Dadas las características se opto por realizar 5 dimensiones junto con 1 de hechos, estas dimensiones son: dim_time, dim_promotion, dim_branch, dim_product, dim_payment y fact_sales.

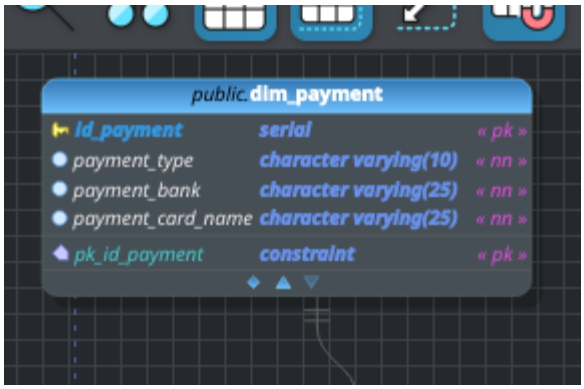


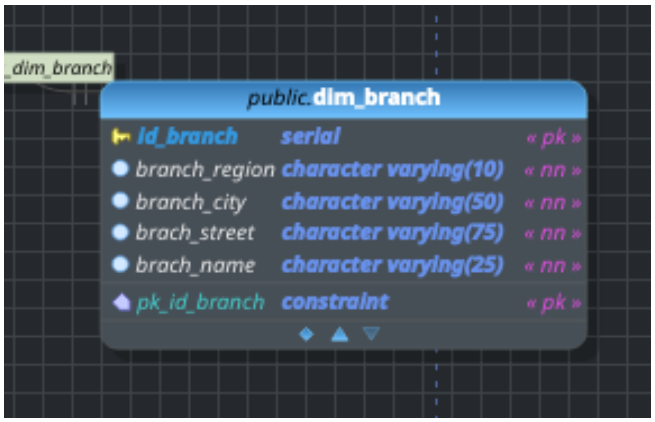
Dentro de estas dimensiones se establecieron los tipos de datos, las llaves y relaciones y los diferentes puntos a obtener, además de que tiene comentarios sobre cada punto de las dimensiones, a continuación, cada una de las dimensiones con sus características.



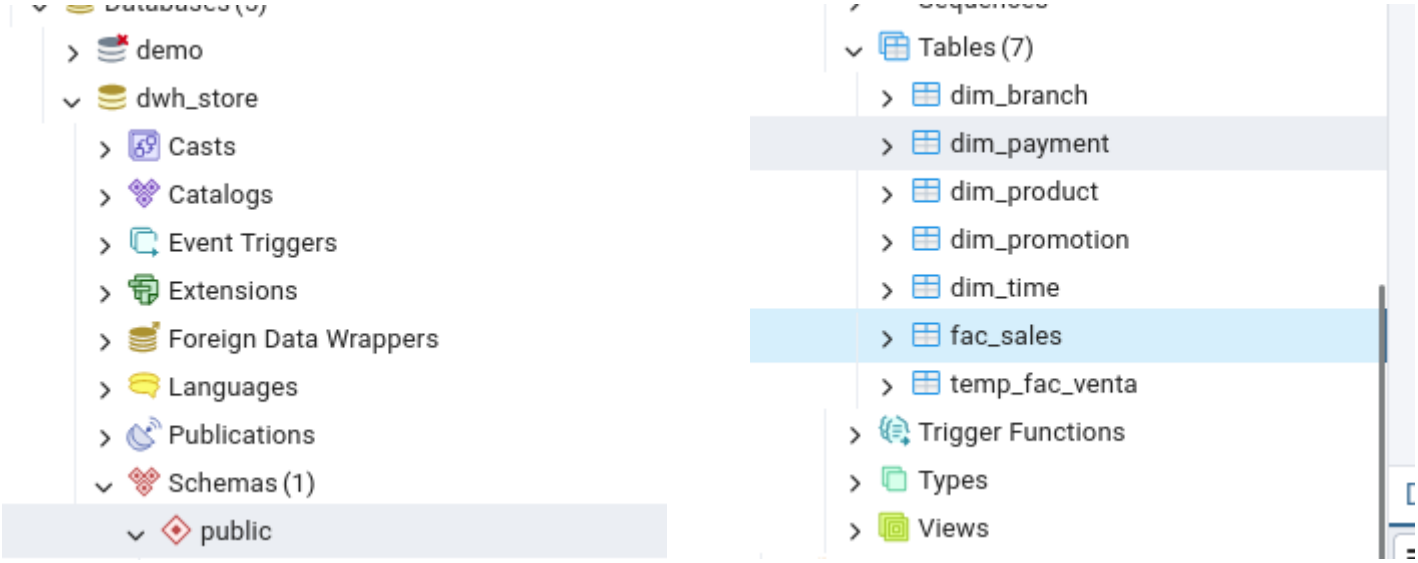








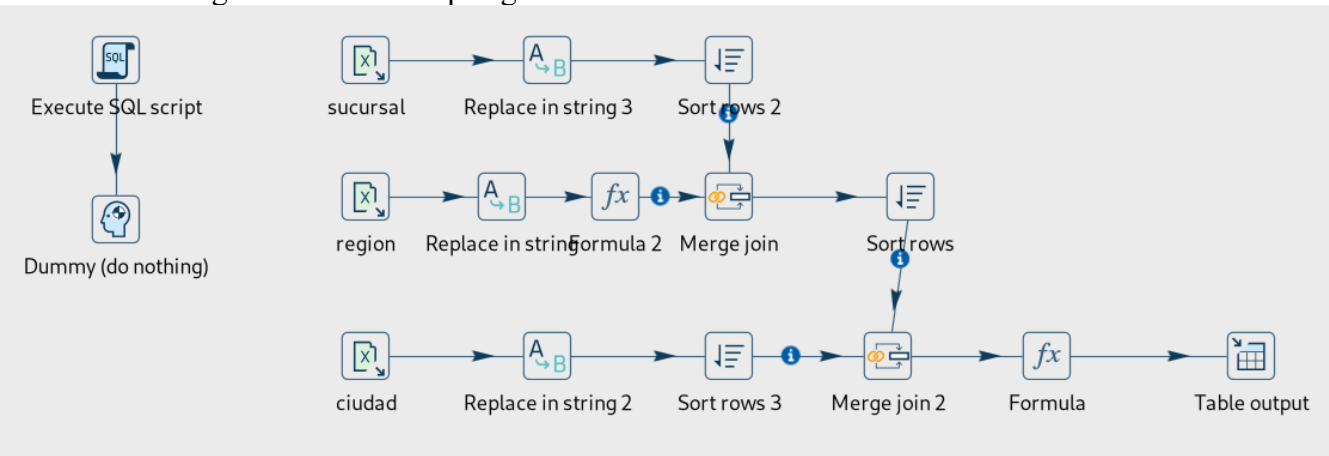
Al tener dicho modelado procedemos a conectarnos a postgres para realizar la importación de las dimensiones, esto fue bastante sencillo y quedo de esta forma dentro de postgres.



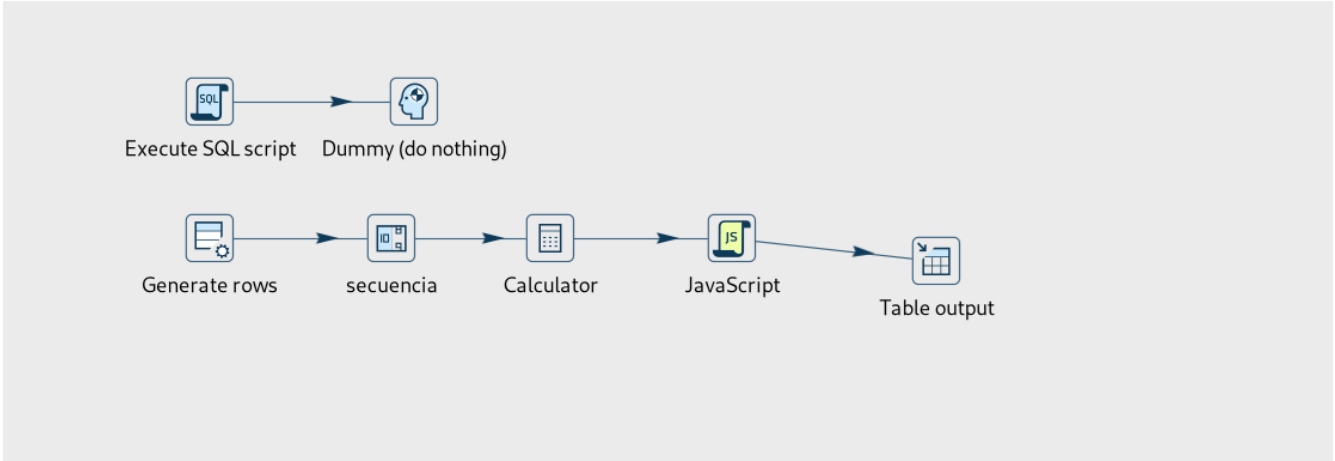
Además, se adjuntó una tabla extra temporal exactamente igual a la de fact_sales para hacer la carga de datos posteriormente.

Una vez cargadas las dimensiones procedemos a realizar la carga de los datos conforme lo indicado en cada una de las dimensiones.

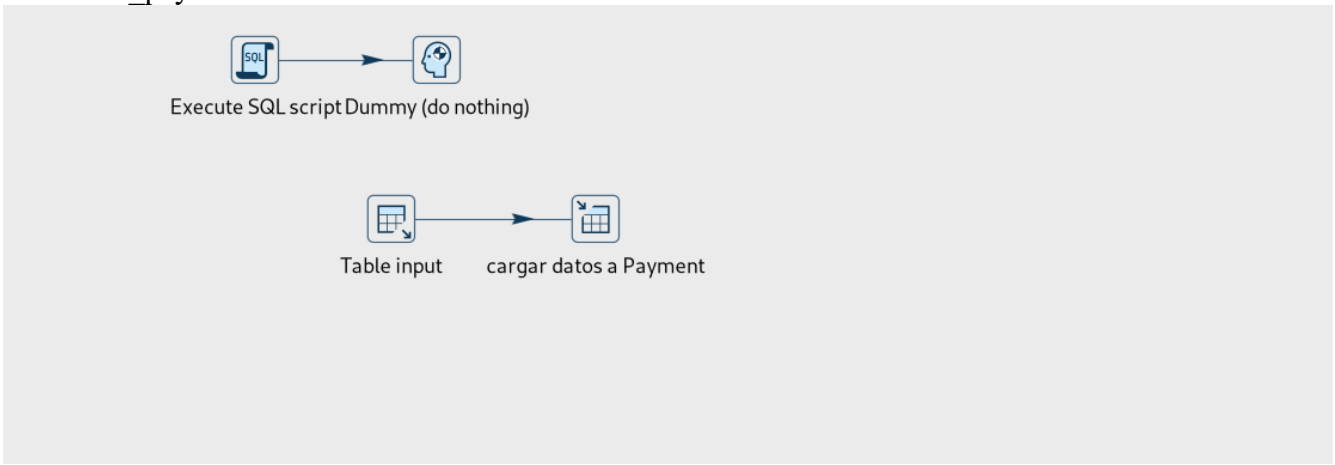
Se empezó con la carga de la dim_branch, esta se realizó con tres documentos de excel, donde se unieron respectivamente con merge join mediante un pipeline en apache hop y al final la carga de los datos en postgres.



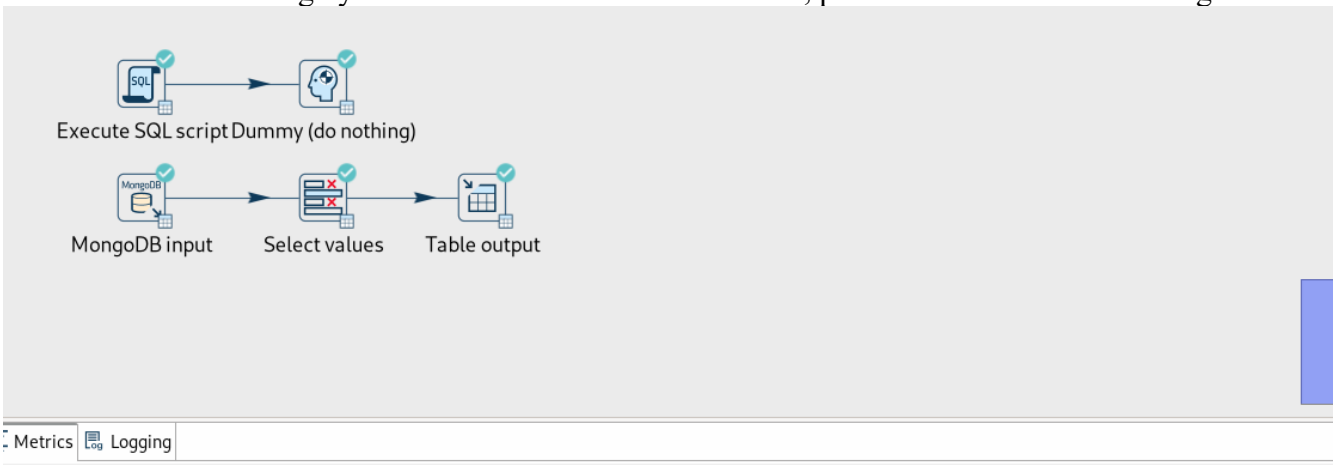
Para la dim_time se utilizó igualmente apache hop, pero esta vez se utilizó la generación de filas para generación de datos mediante secuencias, calculadora y javascript, el resultado se cargó en la dimensión dim_time y este fue el pipeline.



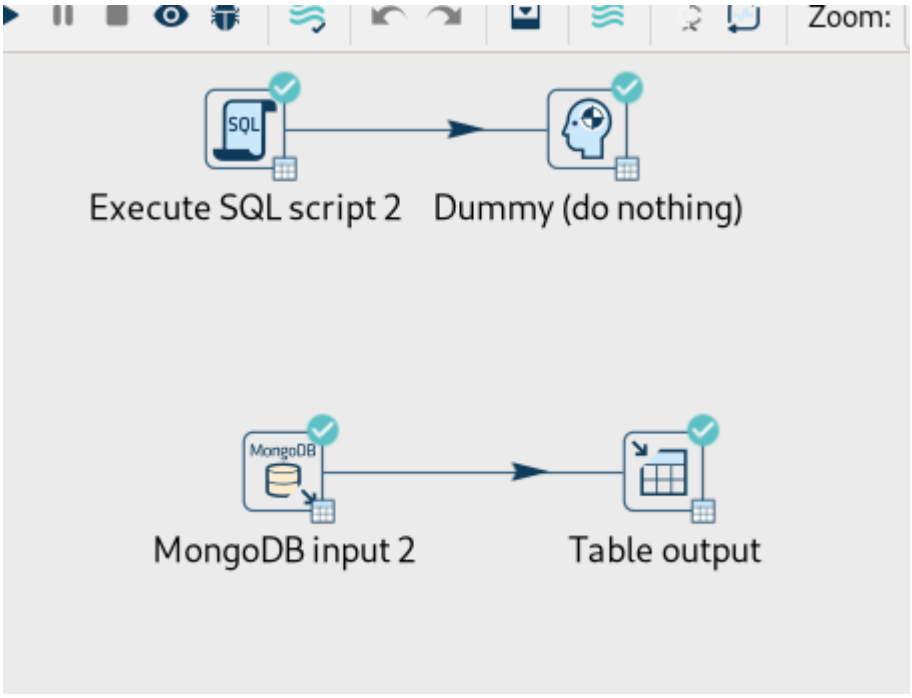
Para dim_payment se utilizó apache hop, pero esta vez se utilizó la obtención de los datos de otra tabla de una base diferente y se utilizó dicha información para cargarla en la dim_payment.



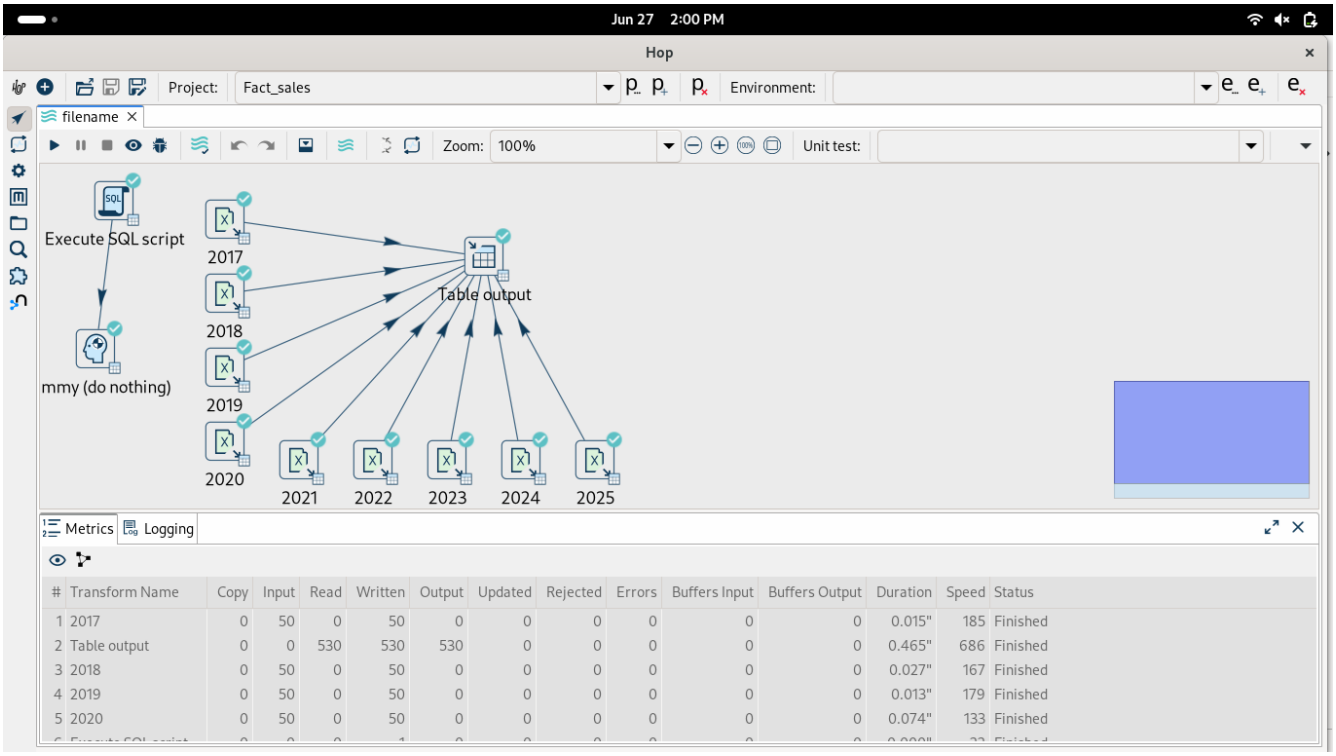
Para dim_promotion se utilizó la carga de datos a través de mongoDB, la cual es a través de una tabla de mongo y se realizó lo mismo con el anterior, pero esta vez a través de mongo.



Para dim_product se utilizó igual la carga a través de mongoDB, igual mediante la colección productos, esto se realizó mediante un pipeline.



Para la carga final de fact_sales se realizó mediante un excel donde se cargaron 50 filas para diferentes años, además se utilizó dicho documento para realizar una carga en una tabla temporal, donde se relacionaban todos esos datos y se cargó con un pipeline, después se cargaron a fact_sales con otro pipeline y se pudieron generar las vistas



Después se generaron las vistas y la inserción en fac_sales para generar ya los reportes finales.
A continuación los resultados