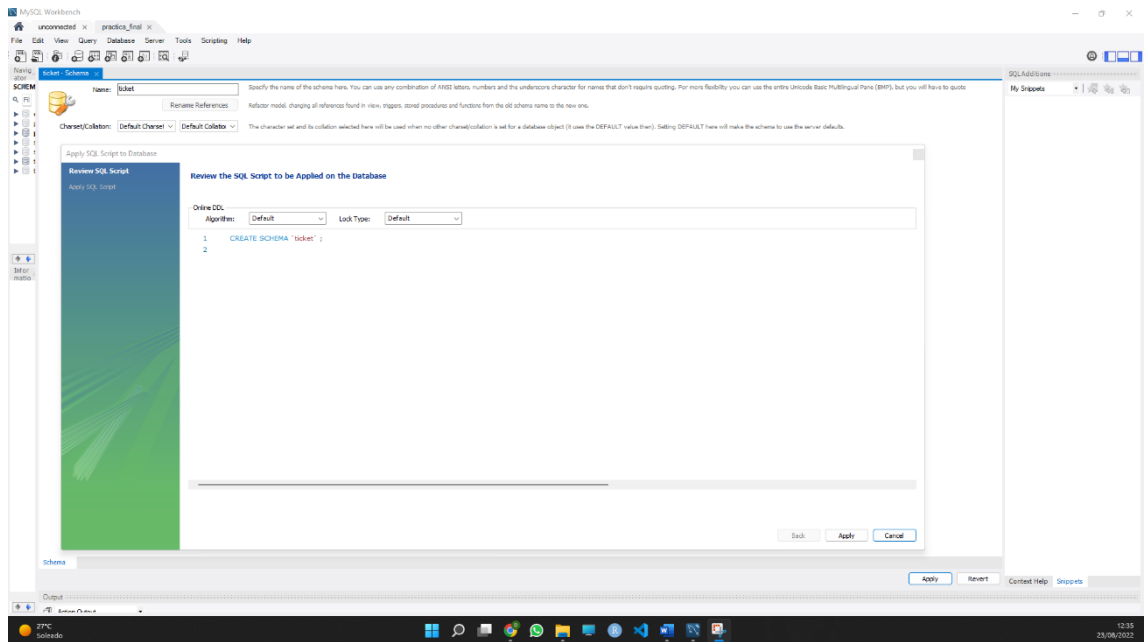
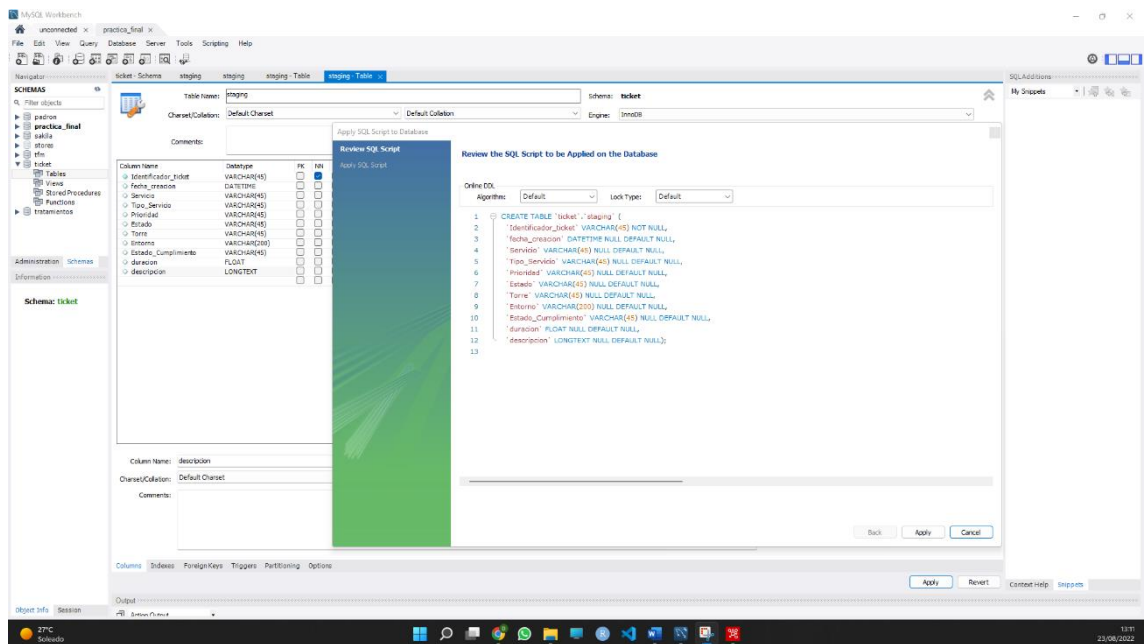


INFORME DOCUMENTAL DE LOS DESARROLLOS

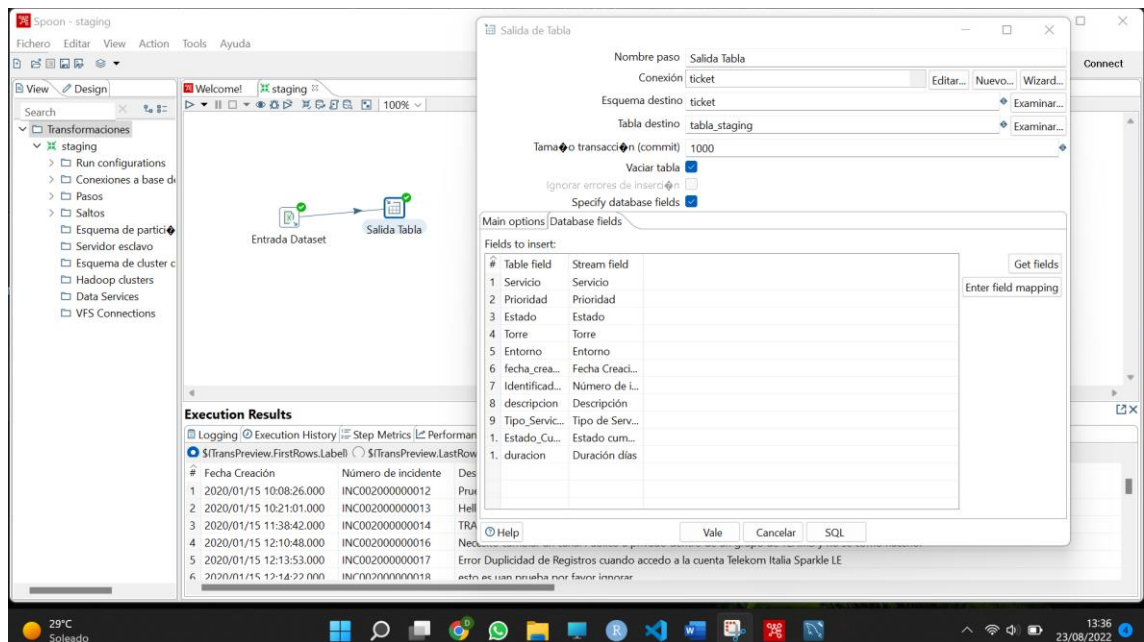
1. El primer paso es crear la base de datos en MySQL Workbench que llamaré “ticket”.



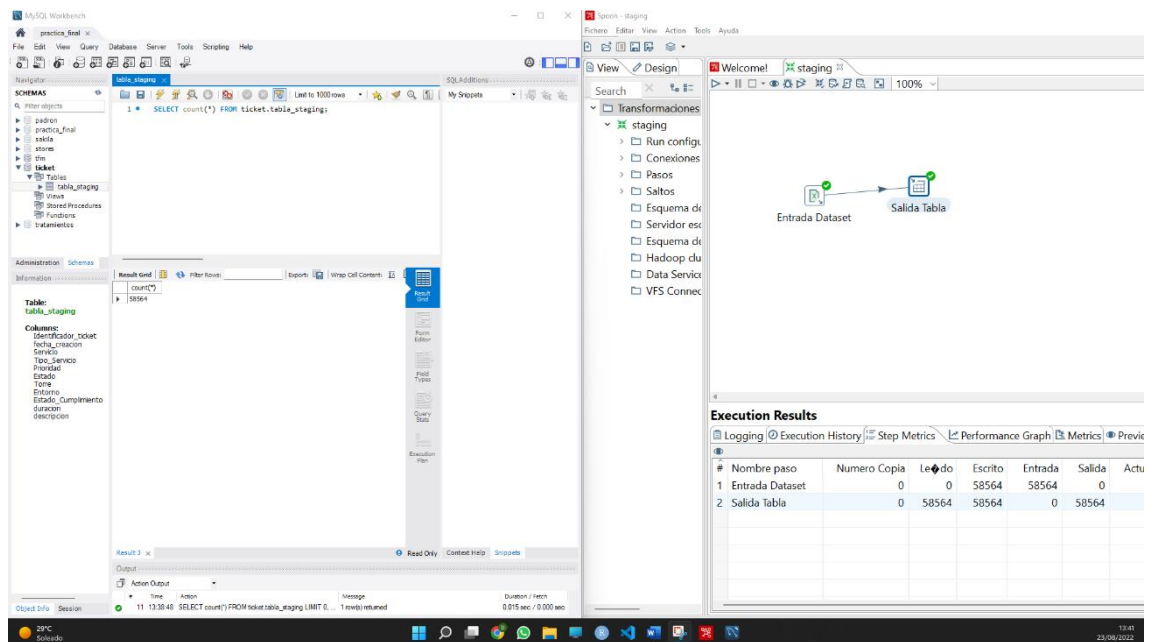
2. Creo una tabla “staging” que utilizaré para cargar mi fuente de datos inicial.



3. Cargo la tabla staging a través de una transformación de Pentaho.



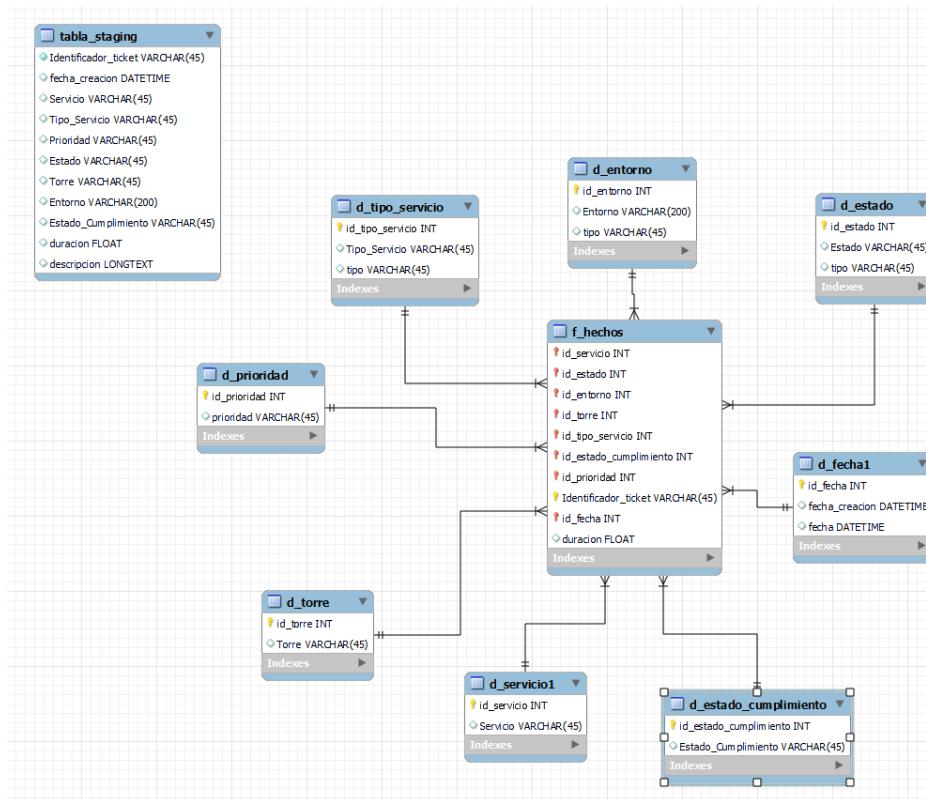
4. A continuación compruebo que la tabla de staging se ha cargado correctamente en mi base de datos. (58564 registros en total)



5. A continuación, diseño el esquema de estrella que utilizaré para realizar el modelo de la base de datos. En un modelo de estrella la tabla de dimensiones está conectada a la tabla de hechos mediante una clave externa, mientras que las tablas de dimensiones no se conectan entre sí.

La tabla de hechos almacenará claves y medidas, mientras que las tablas de dimensión resumen las características.

El modelo resultante sería el siguiente:



He decido mantener la variable descripción fuera del modelo porque no aporta la información que necesito para mi informe. Además, los siguientes puntos resumen algunas de las decisiones de diseño:

- En un principio pensé en incluir las variables Servicio y Torre en una misma dimensión porque están directamente relacionadas. Esto se puede comprobar ejecutando un “SELECT COUNT(DISTINCT)” de la siguiente manera:

```
select count(distinct Torre) from ticket.tabla_staging; # = 19
```

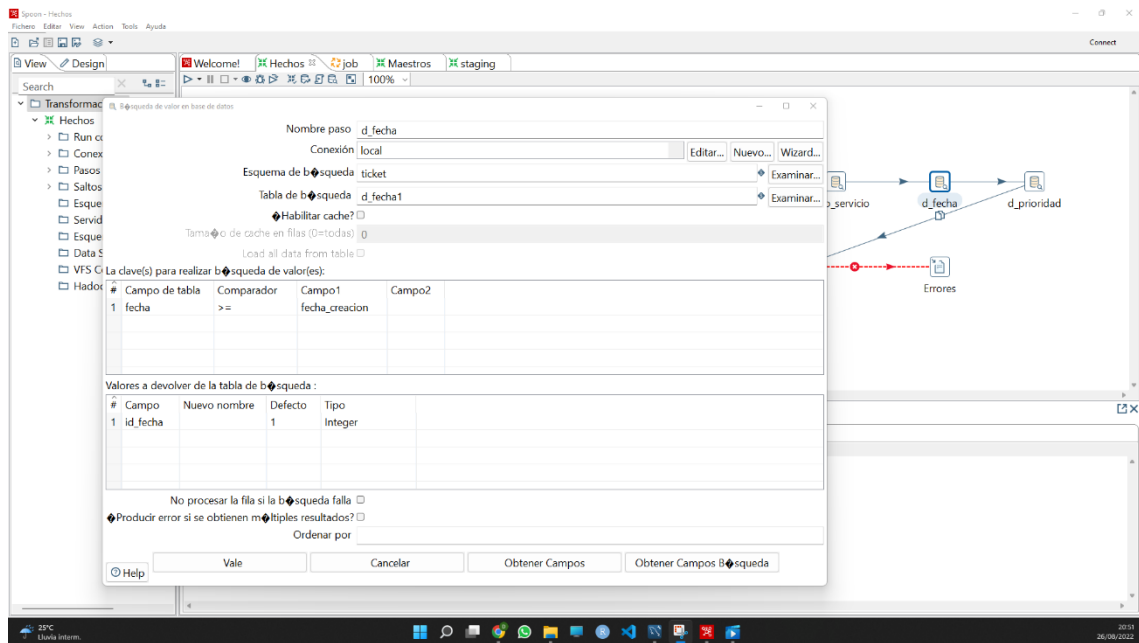
```
select count(distinct Servicio) from ticket.tabla_staging; # = 144
```

```
select count(distinct Servicio,Torre) from ticket.tabla_staging; # = 144
```

Sin embargo, me generaba un problema de duplicados y ante la duda he preferido dividirlos en dos maestros.

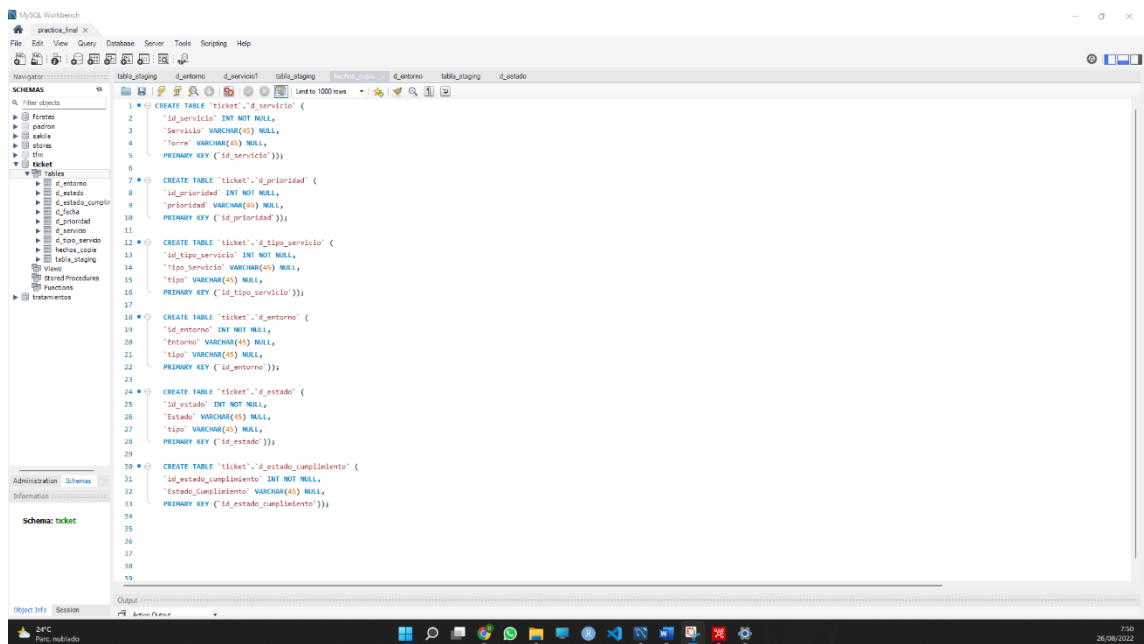
- d_tipo_servicio especifica si es una petición o una incidencia.
- d_entorno diferencia entre productivo y no productivo.
- Distingo entre d_estado y d_estado_cumplimiento. Al mismo tiempo, dentro de d_estado, diferencio entre acabado, abierto o desconocido.

- Para d_fecha no tengo en cuenta la variable de tiempo para generar el id de la tabla maestro. Esta se calculará de la siguiente manera:



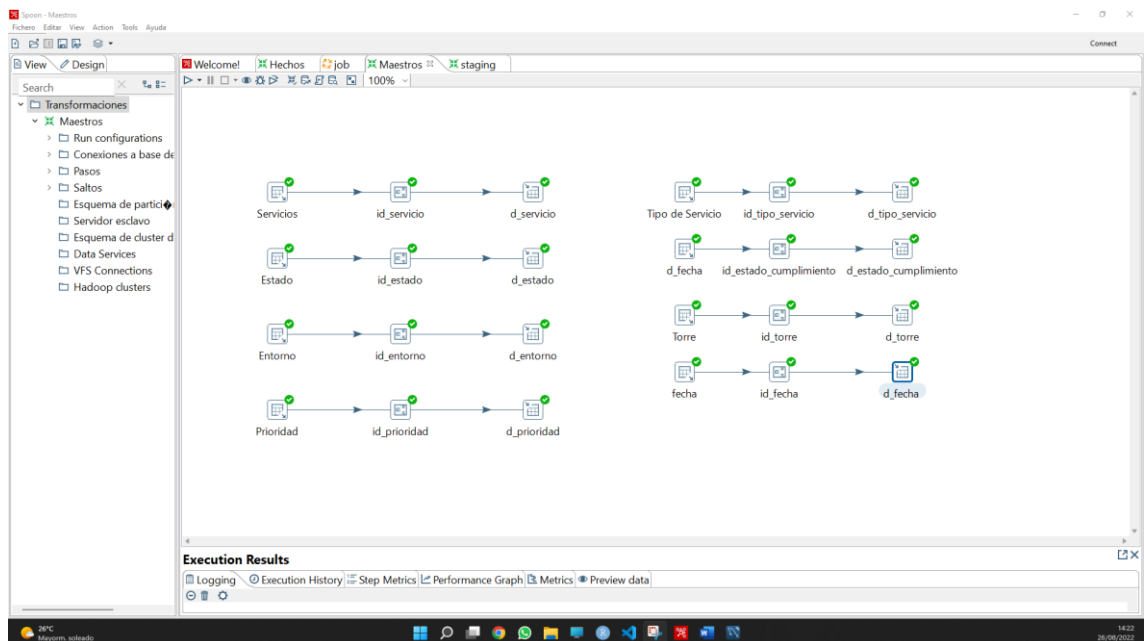
Y así evito una relación de 1:1.

6. Creo las tablas maestro.



7. Cargo las tablas maestro mediante una transformación de Pentaho.

Selecciono con un SELECT DISTINCT sobre mi tabla de staging y asigno claves únicas para cargar en cada uno de los maestros.



Utilizo este paso para crear algunas de las columnas en mis tablas maestras que utilizaré más adelante para hacer las visualizaciones.

The screenshot shows the Spoon - Maestros application. The 'Entrada Tabla' window is open, showing the configuration for the 'Entorno' step. The 'Nombre paso' is 'Entorno' and the 'Conexión' is 'local'. The SQL query is:

```
SELECT distinct Entorno,
CASE WHEN Entorno LIKE '%PROD%' THEN 'Productivo' ELSE 'No productivo' END AS tipo from ticket.tabla_staging
```

The 'Examine preview data' window shows the results of the query, displaying 882 rows. The table has two columns: 'Entorno' and 'tipo'.

Entorno	tipo
1. PRUEBA	No productivo
2. PRODUCTION	Productivo
3. UAT	No productivo
4. CERT	No productivo
5. cert-dev	No productivo
6. no production	Productivo
7. OIMPRE - OIMDEV	No productivo
8. Ericson	No productivo
9. PRODUCTION1	Productivo
1. Solicito que se le den permisos al usuario WUECRO17, pueden utilizar el usuario modelo wuedolo2	No productivo
1. Fast	No productivo
1. PRE	No productivo
1. PREPRODUCCION	Productivo
1. na	No productivo
1. MIGRATION	No productivo

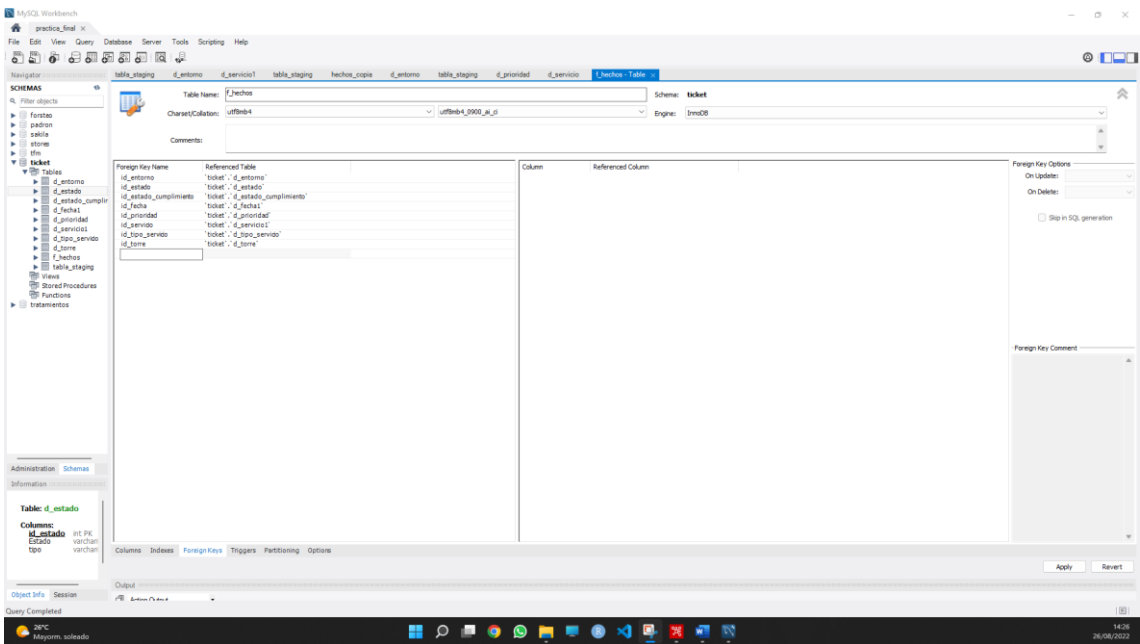
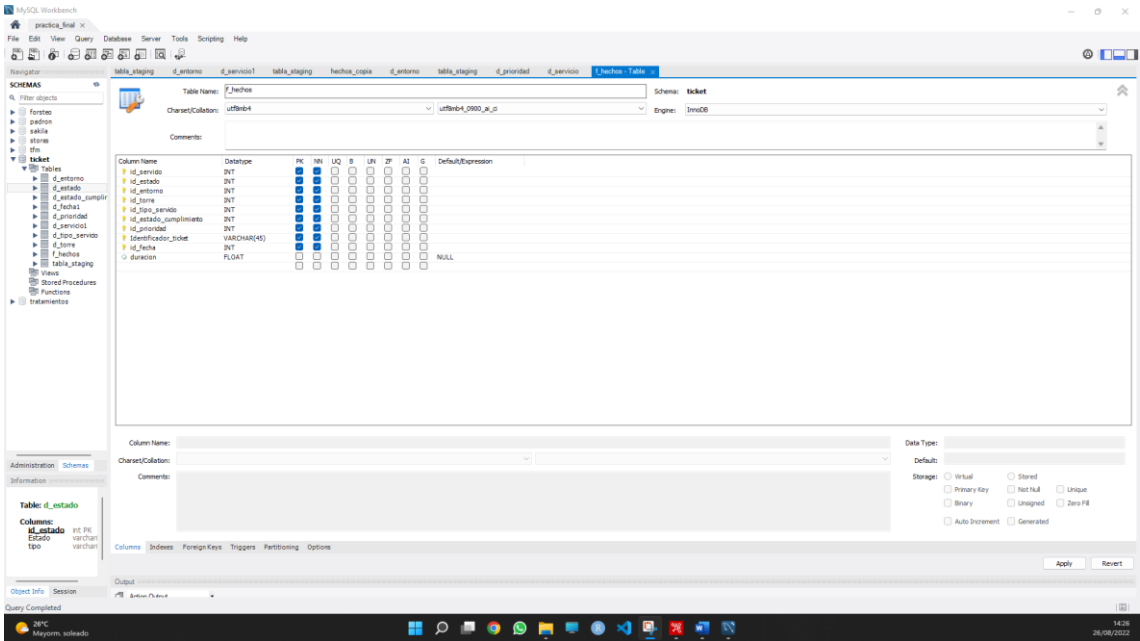
The screenshot shows the Spoon - Maestros application. The 'Entrada Tabla' window is open, showing the configuration for the 'Estado' step. The 'Nombre paso' is 'Estado' and the 'Conexión' is 'local'. The SQL query is:

```
SELECT distinct
Estado,
CASE WHEN Estado LIKE '%cerr%' THEN 'acabado'
WHEN Estado LIKE '%resu%' THEN 'acabado'
WHEN Estado LIKE '%pend%' THEN 'abierto'
WHEN Estado LIKE '%asig%' THEN 'abierto'
ELSE 'cancelado' END AS tipo
FROM ticket.tabla_staging
```

The 'Examine preview data' window shows the results of the query, displaying 5 rows. The table has two columns: 'Estado' and 'tipo'.

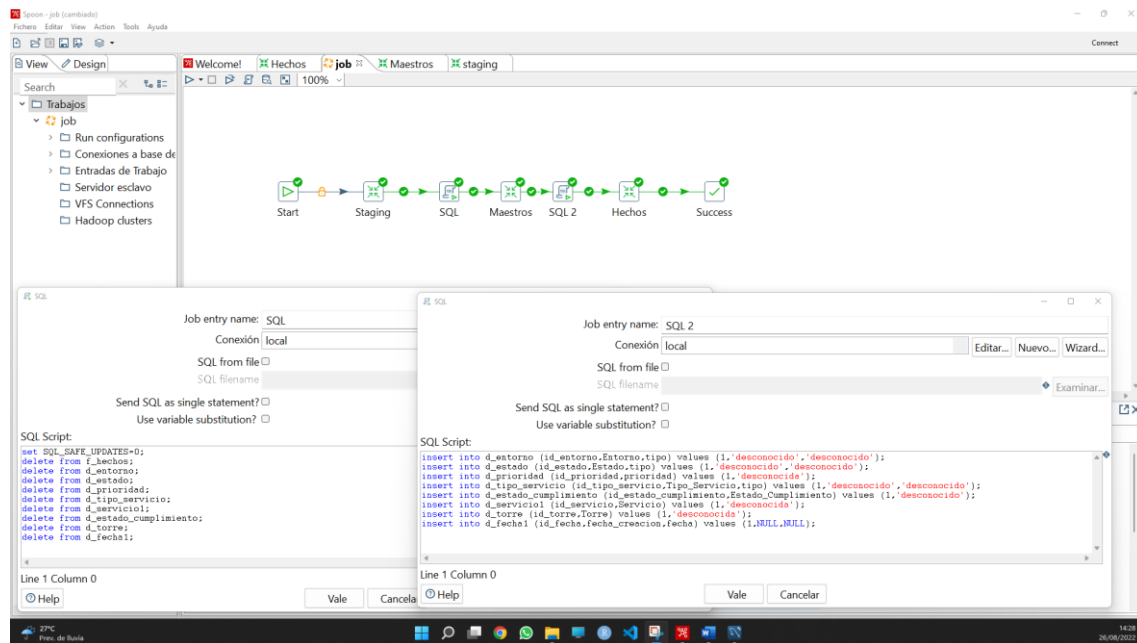
Estado	tipo
1. Cancelado	cancelado
2. Cerrado	acabado
3. Pendiente	abierto
4. Asignado	abierto
5. Resuelto	acabado

8. Continúo creando la tabla de hechos la tabla de hechos y establezco las foreign keys.



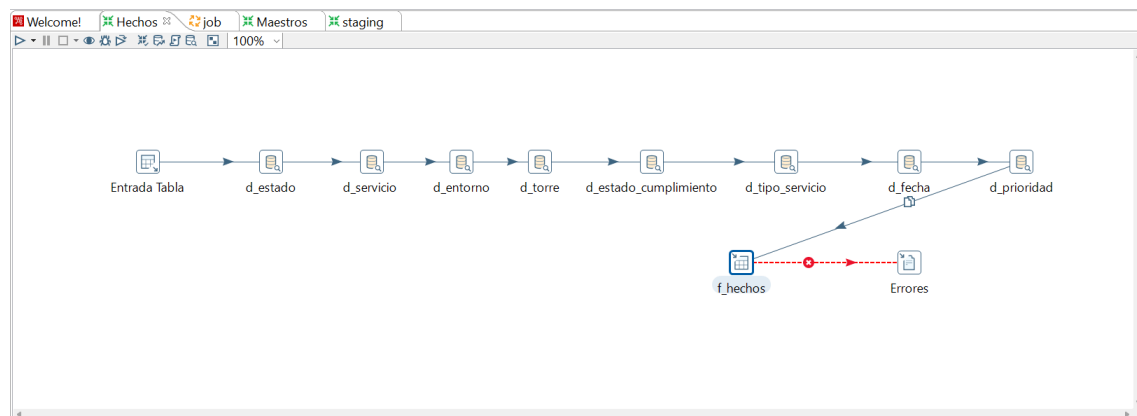
9. Durante la carga de la tabla de hechos me encuentro con un problema de calidad del dato, encontrándome tanto con un problema de nulos como de duplicados. La solución al problema de nulos está en orquestar la carga a través de un job, incluyendo scripts de SQL que identifiquen los valores no encontrados como 'desconocido' dentro de cada una de las tablas maestras, y especificar un id para estos casos.

El Job se estructura de la siguiente manera:



*Importante desactivar el modo seguro de SQL con: SET SQL_SAFE_UPDATES=0;

Para lidiar con los duplicados opto por añadir un fichero en el que recojo los posibles errores sin interrumpir el proceso. Lo que me permitiría seguir cargando mi tabla de hechos.

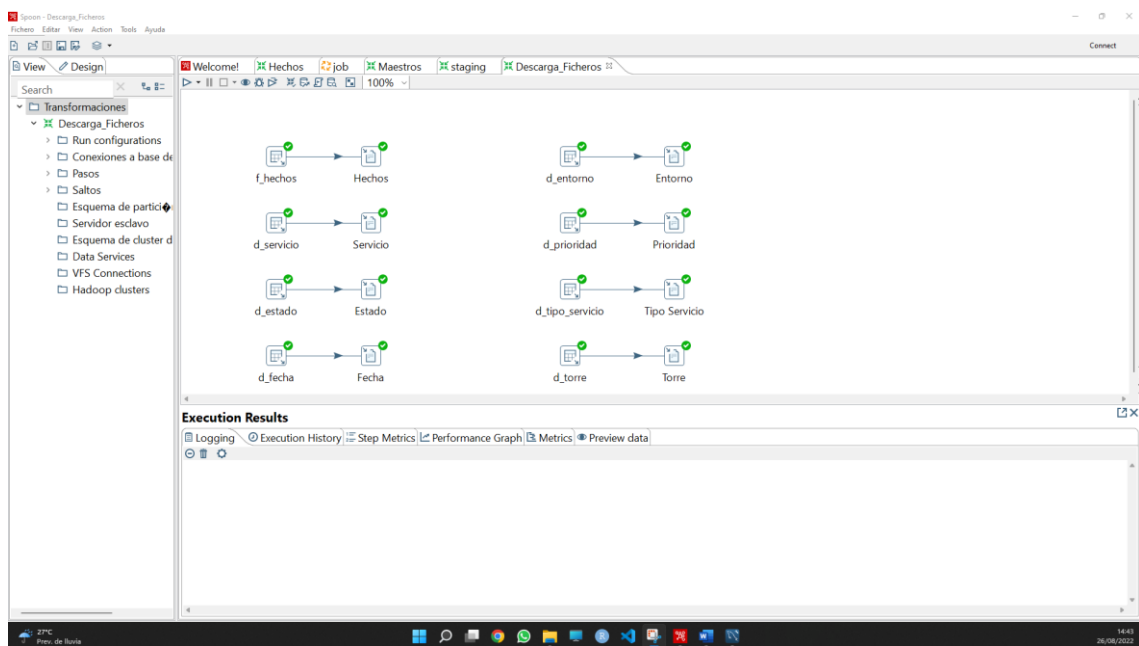


El número de registros que termino cargando en mi tabla de hechos es de 58564



Result Grid	Filter Rows:	Export:	Write
count(*)			
58564			

10. Para terminar, la versión “Tableau Public” no permite establecer una conexión entre MySQL workbench y Tableau, obligando a descargar cada una de las tablas que compone el modelo por separado. (he decidido no incluir este paso en el job)



11. La conexión de la tabla hechos con sus maestros queda de la siguiente forma en Tableau:

Tableau Public - Libro1

Conexiones

Archivos

Hechos.txt — Torre.txt

¿En qué se diferencian las relaciones de la unión de columnas? Más información

Torre.txt	Torre.txt
Id torre (Torre.txt)	Torre
1	desconocida
2	ESS
3	BSS
4	OSS

12. Las comprobaciones de exactitud de los datos se hacen también en esta pestaña.

Tableau Public - Informe tickets

Conexiones

Archivos

Hechos.txt

10 campos 58564 filas

Describe campo

Identificador ticket

Función: Dimensión discreta

Tipo: Columna de base de datos

Columna remota: [Hechos.txt].[Identificador_ticket]

Tipo de remoto: Cadena de caracteres ANSI/MBCS

Contiene NULL: Desconocido

Configuración regional: España(español)

Clasificación: Distingue mayúsculas de minúsculas

Anchura de columna: Desconocido

Estado: Válido

Dominio (20 de 58.564 miembros)

INC002000000012

INC002000000013

INC002000000014

INC002000000016

INC002000000017

Cargar Copiar

Alto	Hechos.txt	Hechos.txt	Hechos.txt
Identificador ticket	Id Fecha	Duración	
INC002000064440	57299	0.001	
INC002000062574	56222	61.535	
INC002000061729	55363	11.434	
INC002000060883	53766	0.000	
INC002000060630	53766	7.680	
INC002000059767	53559	8.451	
INC002000059388	53041	0.000	
INC002000059328	53002	0.001	
INC002000059326	53002	0.003	
INC002000058057	51878	0.012	
INC002000058056	51878	0.012	
INC002000058054	51878	0.012	
INC002000058048	51878	0.046	