

CASO PRÁCTICO FINAL

Este caso quiere representar una petición formal de un cliente para realizar un análisis de datos a partir de una extracción de una de sus aplicaciones, algo habitual en el mercado de BI. Es nuestra labor como especialistas en el análisis el realizar el diseño de arquitectura de todo el sistema de análisis, modelado de base de datos, carga de la misma y explotación de los datos.

1. ANÁLISIS DE LOS DATOS INICIALES Y MODELO DE DATOS PROPUESTO/DESARROLLOS EN MySQL Y PENTAH0

En primer lugar, antes de plantear el modelo, se ha estudiado el archivo Excel, para tener una idea más clara de qué variables pueden estar relacionadas y ver cómo plantear la solución adecuada para las dudas del cliente.

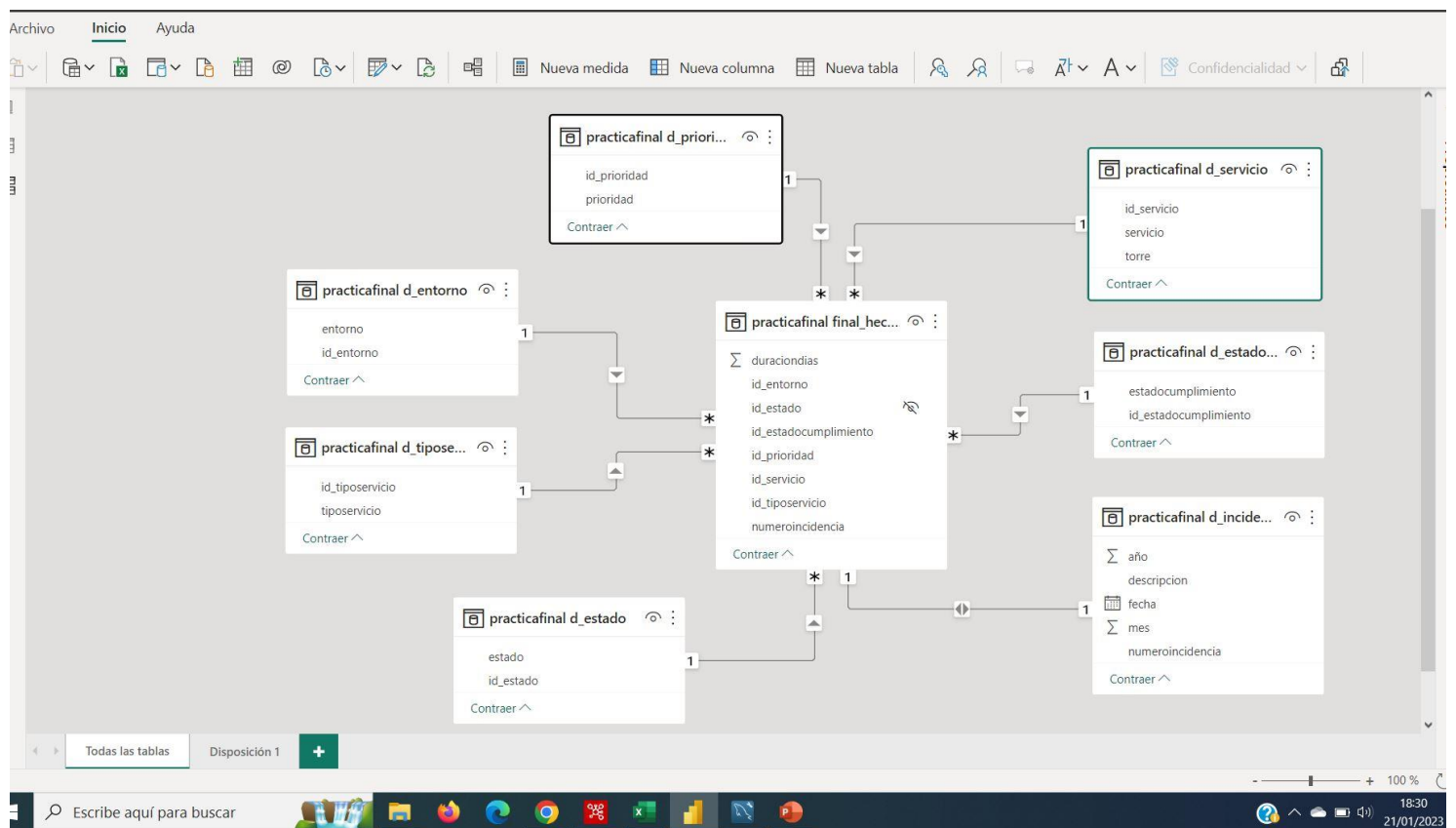
Tal como se dice en el enunciado (cada servicio pertenece a una torre de gestión), la torre y el servicio están relacionados, por lo que pueden ir juntos.

También se observa que el número de incidencia es único (PK) y que la fecha, como marca hora, minutos y segundos, también. Se supone que no entran dos incidencias en el mismo momento exactamente, siempre habrá diferencia, aún solo en los segundos, por lo que también puede asociarse a la incidencia, aunque necesitemos separar el año y el mes para posteriormente resolver las cuestiones.

Primero tenemos que ver cómo se van a agrupar las variables y después ya se podrá empezar a su desarrollo

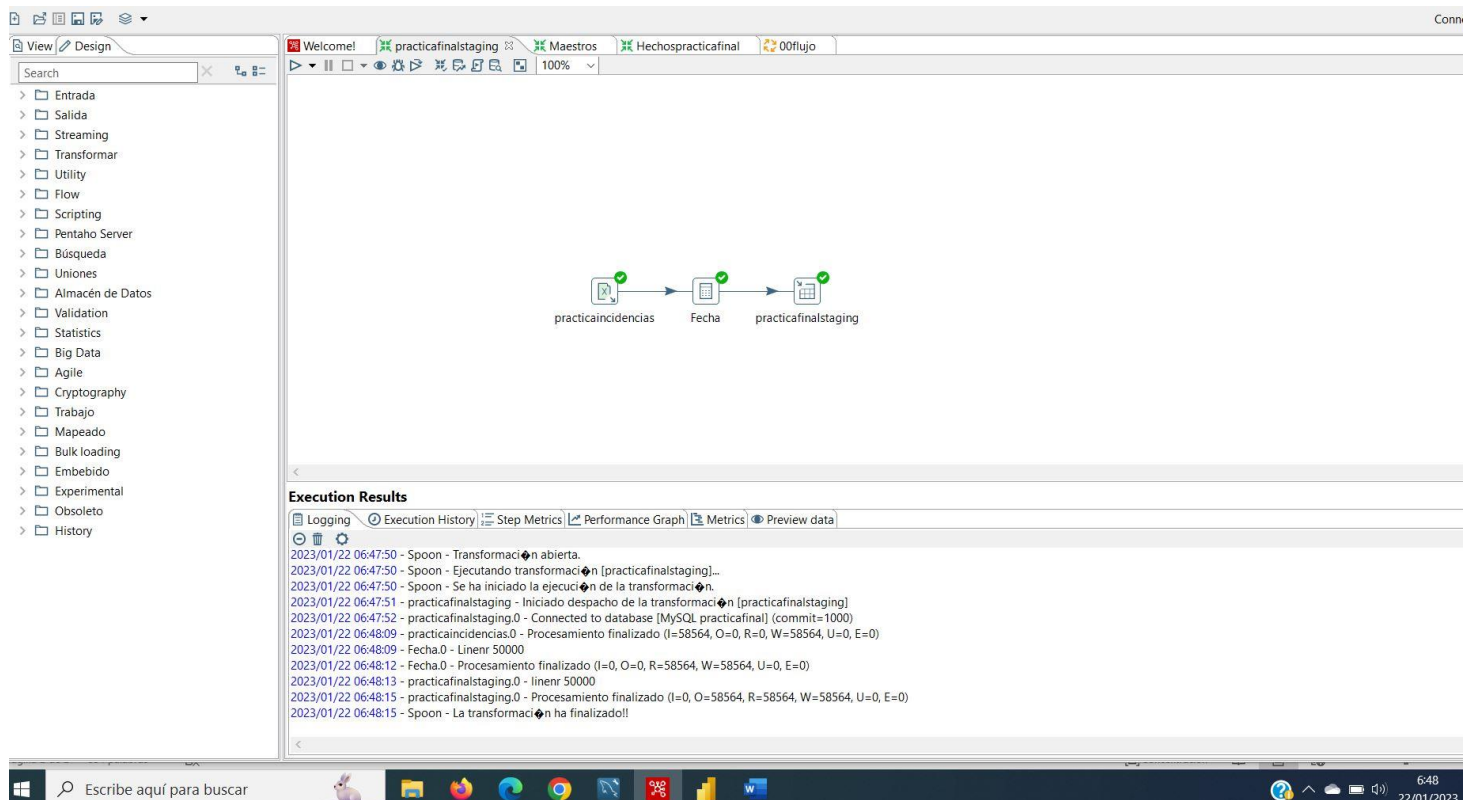
Una vez tenemos claro el modelo que se va a plantear, se empieza a trabajar con los programas Pentaho y MySQL.

En la página siguiente se verá un esquema en estrella (Power BI) que se obtuvo al final, aunque también tenemos el esquema en MySQL.



Vemos la tabla de hechos en el centro y alrededor todas las demás (el asterisco indica n relaciones) y se ve una relación uno a uno permite que un solo registro de la tabla d_incidencias se relaciona con un único registro de la tabla de hechos.

El primer paso es cargar el fichero Excel en una hoja de transformaciones de Pentaho



Aquí ya hemos metido un paso intermedio para que aparezcan, además de fecha de creación, las columnas “año” y “mes”, que necesitaremos después.

Una vez hecho esta carga, ya se puede empezar a trabajar en MySQL.

Se crean en Pentaho en otra hoja de transformaciones todas las tablas con las variables que se pueden agrupar entre sí, en este caso sólo torre y entorno y por otro lado fecha creación, incidencia y descripción. El resto de variables irá sola en la tabla con la creación de su correspondiente identificador.

Con todos los indicadores podremos hacer la tabla de hechos (aquí es dónde va el dato numérico “días de duración”).

The screenshot displays the Pentaho Data Integration (Kettle) interface. The top pane shows a transformation design with several tables and their relationships. The bottom pane shows the 'Execution Results' tab, which contains a log of the transformation's execution.

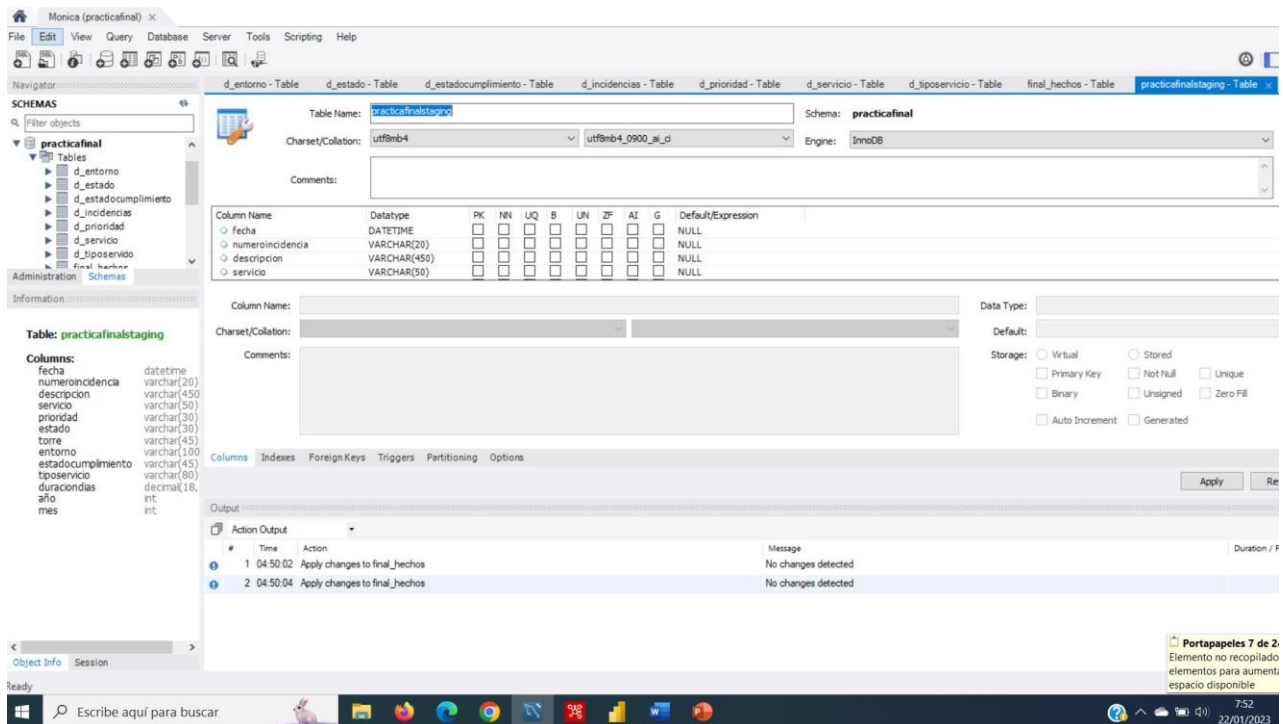
Transformation Design:

- Tables: Servicio, id_servicio, d_servicio, Entorno, id_entorno, d_entorno, Tipo_servicio, id_tiposervicio, d_tiposervicio, Estado_cumplimiento, id_estadocumplimiento, d_estadocumplimiento, Prioridad, id_prioridad, d_prioridad, Incidencias, d_incidencias, Estado, id_estado, d_estado.
- Relationships: Servicio to id_servicio to d_servicio; Entorno to id_entorno to d_entorno; Tipo_servicio to id_tiposervicio to d_tiposervicio; Estado_cumplimiento to id_estadocumplimiento to d_estadocumplimiento; Prioridad to id_prioridad to d_prioridad; Incidencias to d_incidencias; Estado to id_estado to d_estado.

Execution Results:

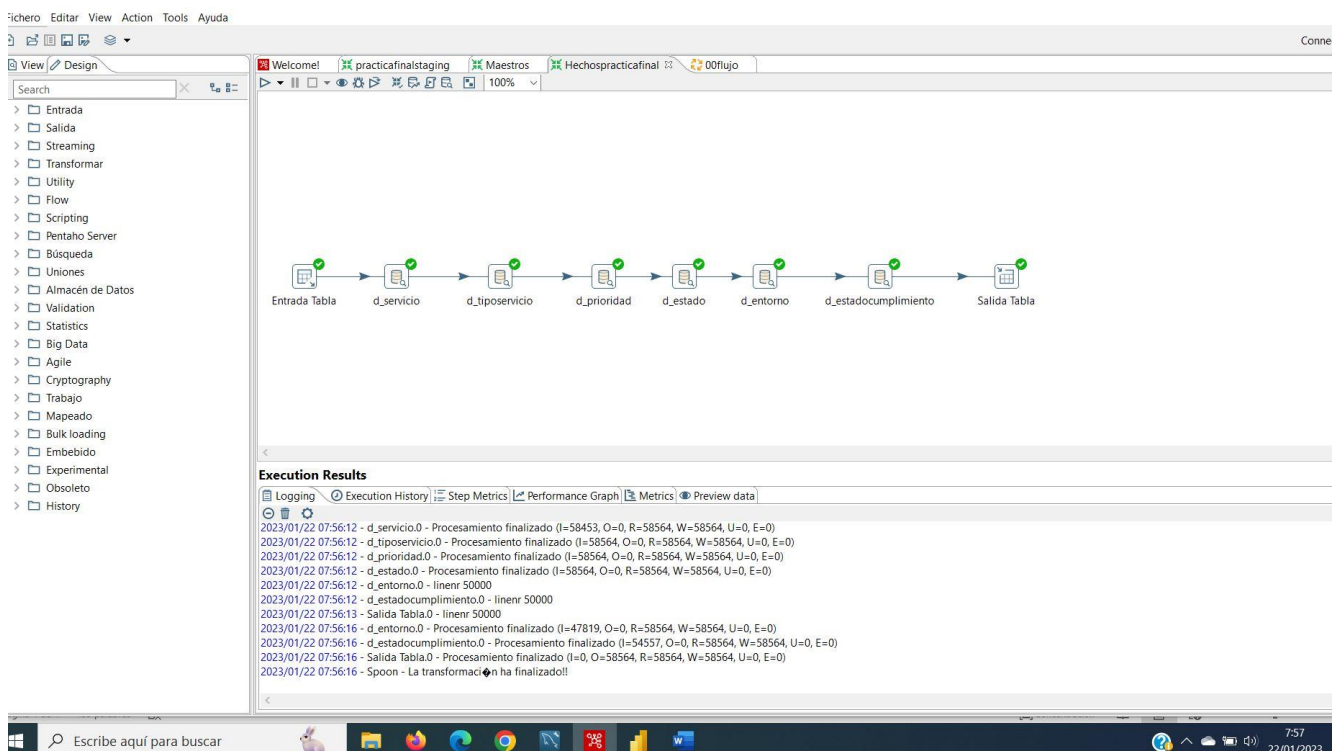
```
2023/01/22 07:47:15 - Estado_cumplimiento.0 - Finished reading query, closing connection
2023/01/22 07:47:15 - Estado_cumplimiento.0 - Procesamiento finalizado (I=4, O=0, R=0, W=4, U=0, E=0)
2023/01/22 07:47:16 - id_estadocumplimiento.0 - Procesamiento finalizado (I=0, O=0, R=4, W=4, U=0, E=0)
2023/01/22 07:47:16 - d_estadocumplimiento.0 - Procesamiento finalizado (I=0, O=4, R=4, W=4, U=0, E=0)
2023/01/22 07:47:16 - d_entorno.0 - Procesamiento finalizado (I=0, O=882, R=882, W=882, U=0, E=0)
2023/01/22 07:47:23 - Incidencias.0 - lineas 50000
2023/01/22 07:47:24 - Incidencias.0 - Finished reading query, closing connection
2023/01/22 07:47:24 - Incidencias.0 - Procesamiento finalizado (I=58564, O=0, R=0, W=58564, U=0, E=0)
2023/01/22 07:47:24 - d_incidencias.0 - lineas 50000
2023/01/22 07:47:26 - d_incidencias.0 - Procesamiento finalizado (I=0, O=58564, R=58564, W=58564, U=0, E=0)
2023/01/22 07:47:26 - Spoon - La transformación ha finalizado!!
```

A la vez vamos creando las tablas en MySQL, con las correspondientes PK de cada tabla.

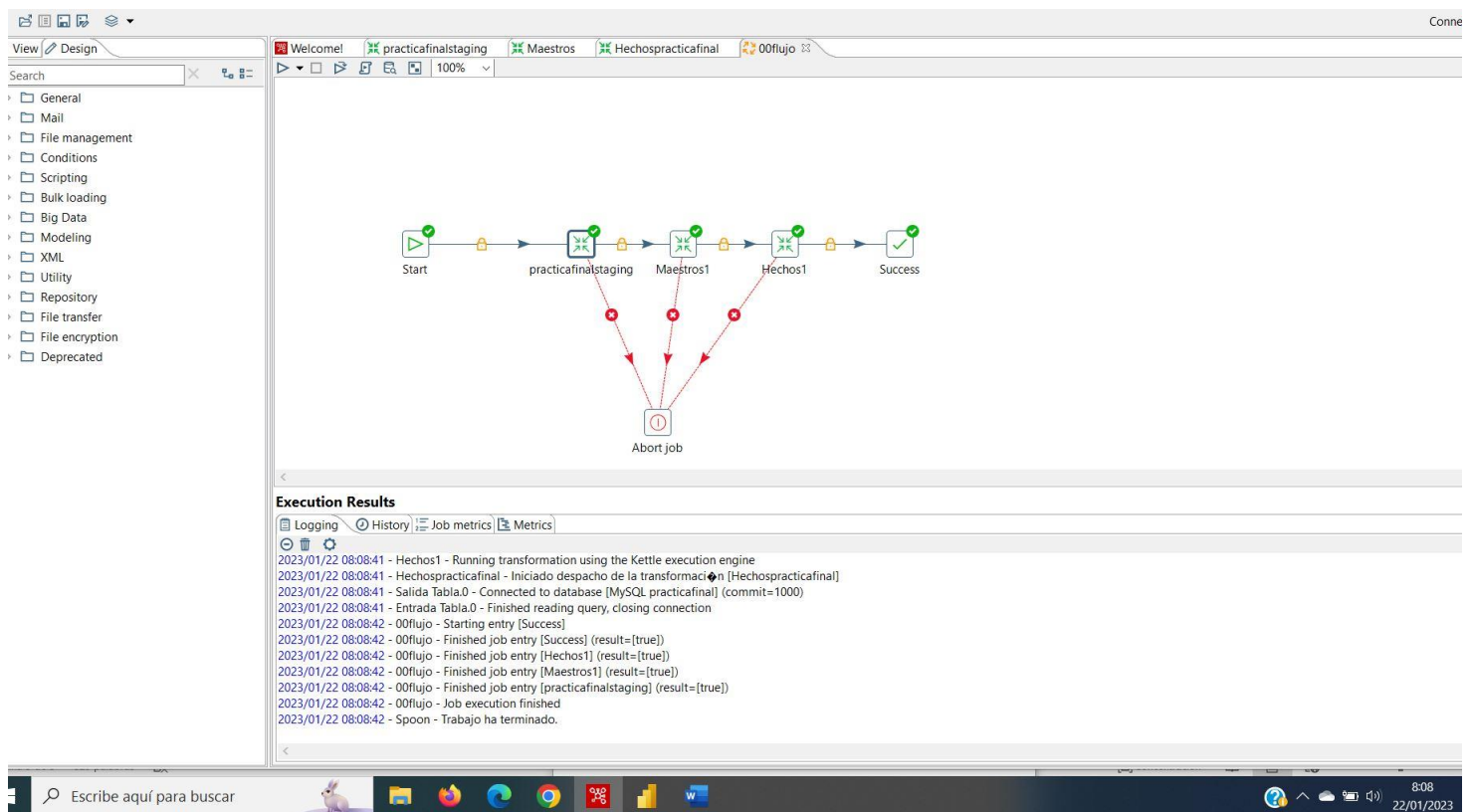


Hay que prestar atención al tipo de variable que se asigna en cada caso, para evitar problemas futuros, e ir paso a paso cargando las diferentes tablas, con su correspondiente id.

Vamos uniendo en Pentaho las diferentes tablas para obtener la tabla final de salida, de la que obtendremos en MySQL la tabla de hechos.



Después ya sólo falta el trabajo, flujo, para comprobación final en Pentaho



Ahora si que se pueden ya cargar todas las tablas necesarias para obtener nuestro modelo:

Tabla de hechos, con las PK del resto de tablas (numeroincidencia, id_entorno, id_prioridad, id_estado, id_estadocumplimiento, id_servicio e id_tiposervicio). Se ha hecho en Pentaho el correspondiente paso de poner -1 cuando se van haciendo las tablas, con la creación del id, para que este no encuentre problemas si tiene valores nulos.

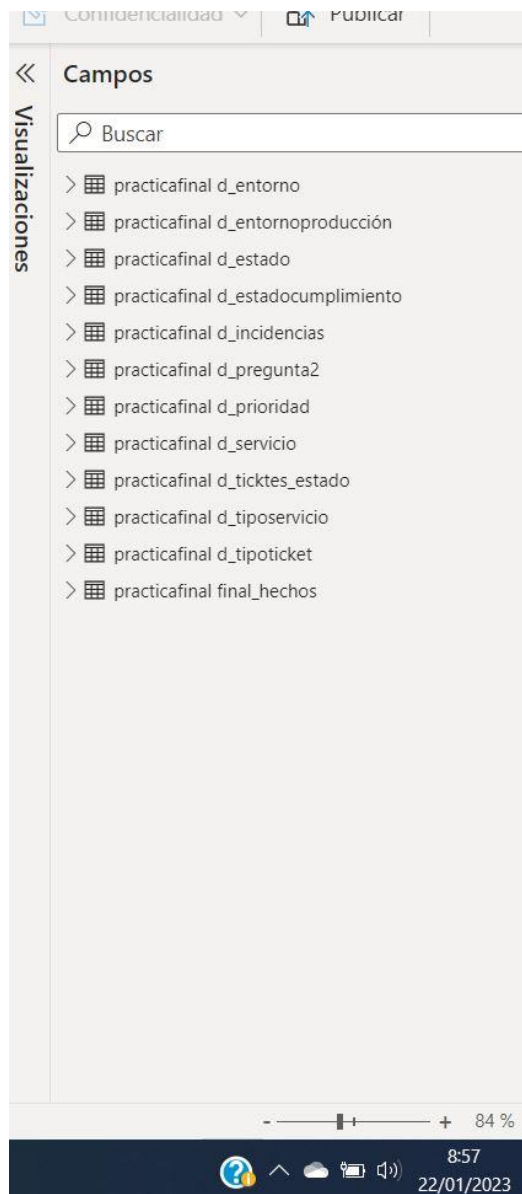
Una vez ya tenemos todas las tablas bien cargadas, se exportan al programa Power BI para trabajar con ellas y responder a las cuestiones que plantea nuestro cliente.

2.RESULTADOS (USANDO POWER BI)

Importamos a Power BI todas las tablas que tenemos en la base de datos practicafinal, excepto pacticafinalstaging.

Y ahora ya se pueden resolver las diferentes cuestiones planteadas.

NOTA= en Pentaho se debería haber hecho la clasificación de los entornos en productivos y no productivos. También clasificar el estado en abierto y cerrado y el tipo de ticket en incidencia y petición. Yo lo he hecho con las tablas de transformación, en Power BI, ya que aquí también se pueden hacer joins, y se pueden añadir esos campos.



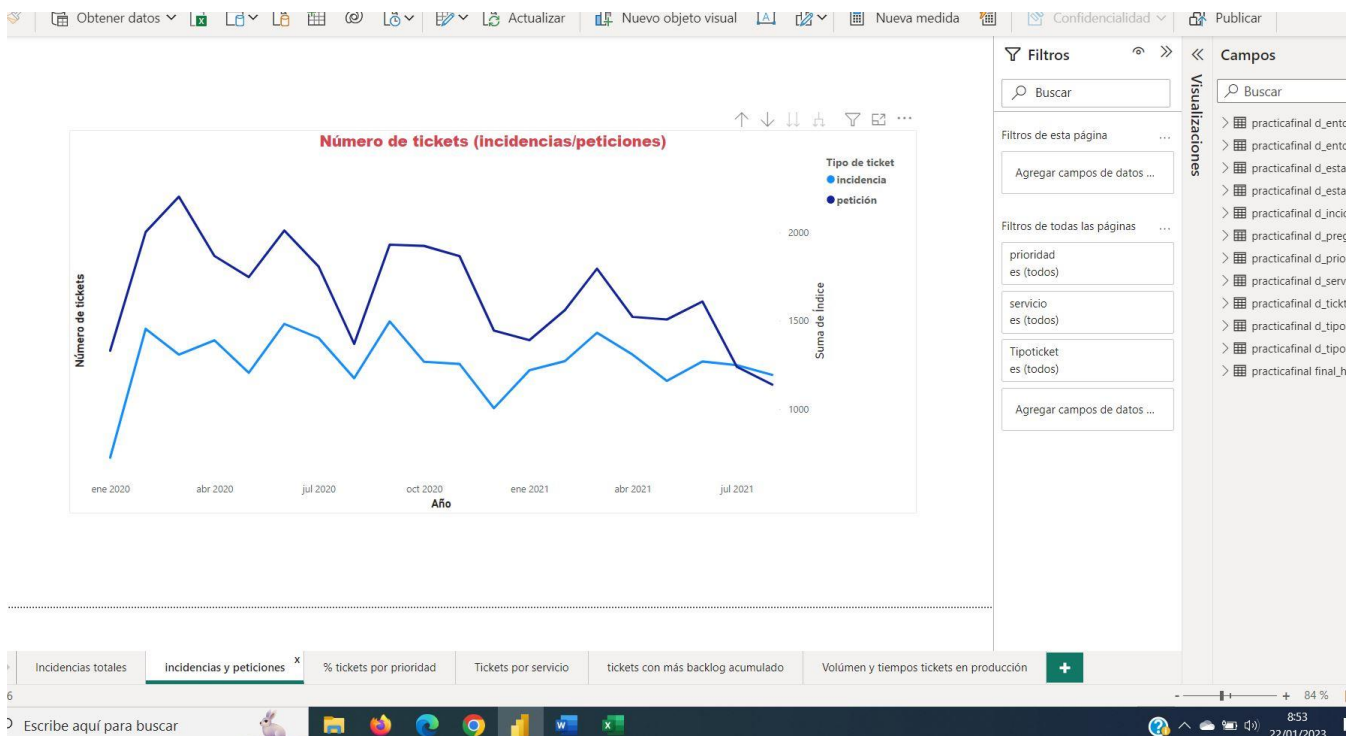
Las tablas creadas en BI son d_entornoproducción, d_pregunta2, d_tickets_estado y d_tipoticket.

En todas las páginas del archivo se han fijado los filtros que se piden (por prioridad, por tipo de ticket y por servicio).

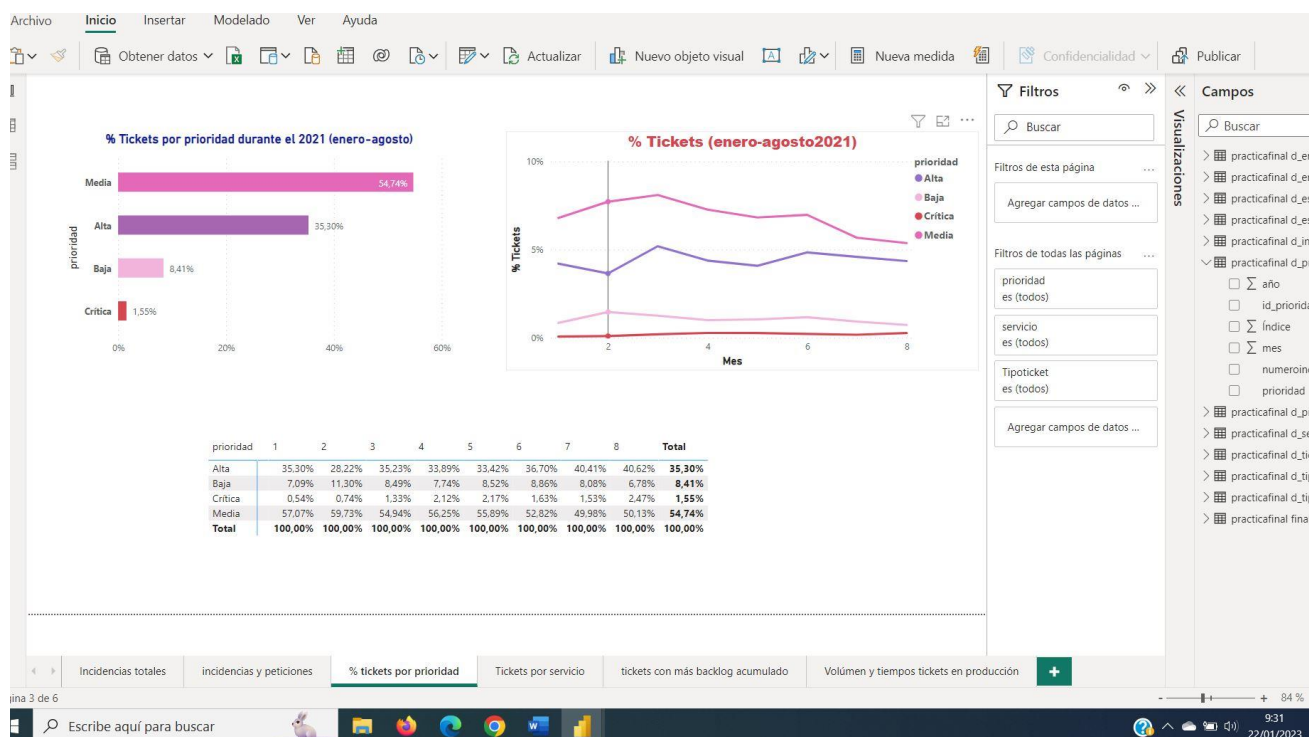
1. La evolución del volumen de tickets general y diferenciado por incidencias y peticiones.



Y en la siguiente gráfica separado por petición e incidencia:



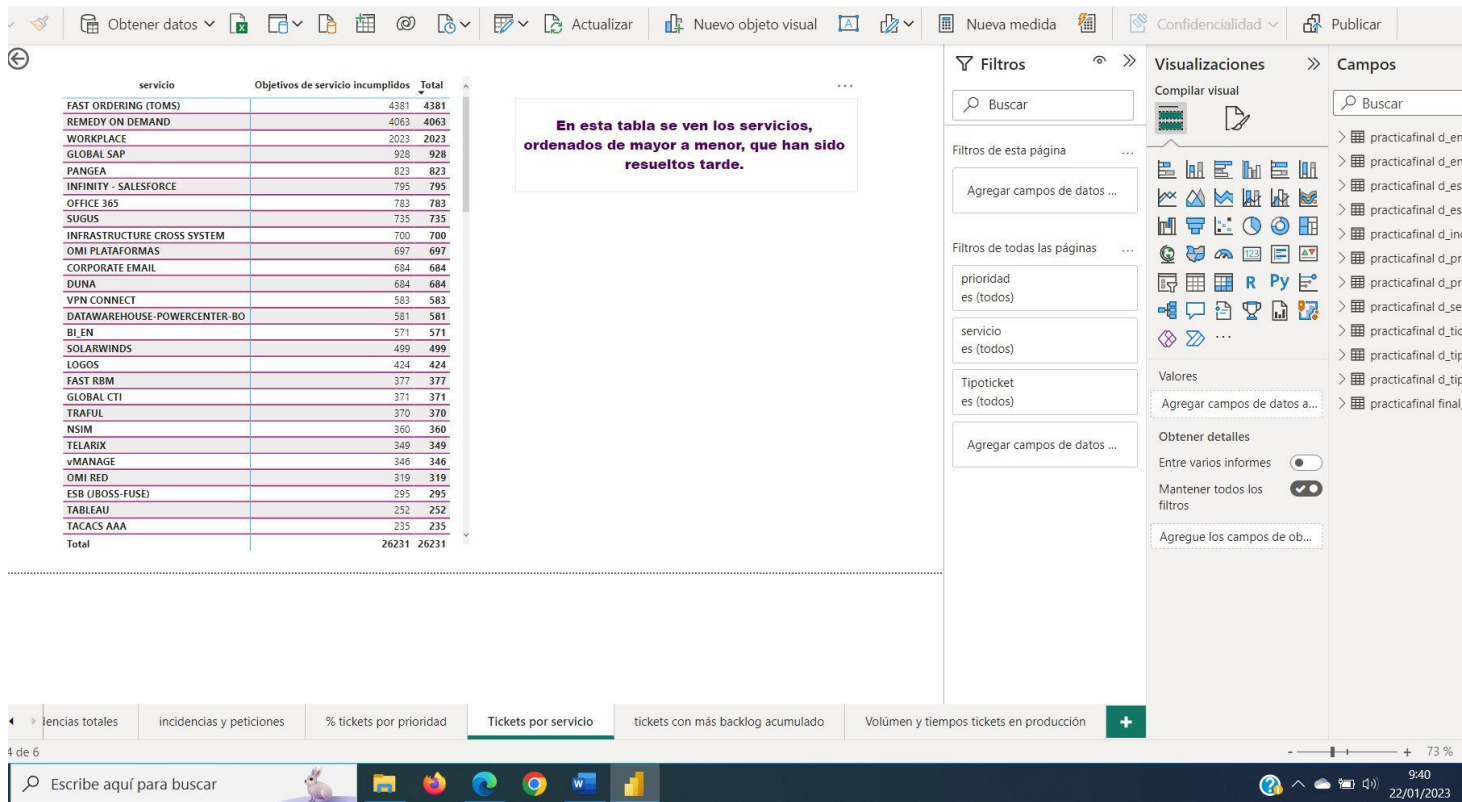
2. Si ha variado la prioridad con la que se abren los tickets en los últimos meses (% de peso por prioridad)



Se ve que % tickets tiene un pico en el mes de marzo, en todo tipo de prioridad, aunque después se observa una ligera tendencia a disminuir, hasta agosto del 2021.

Como los críticos son los que vienen directamente del usuario (incidencias), es lógico que sea el de menor cantidad.

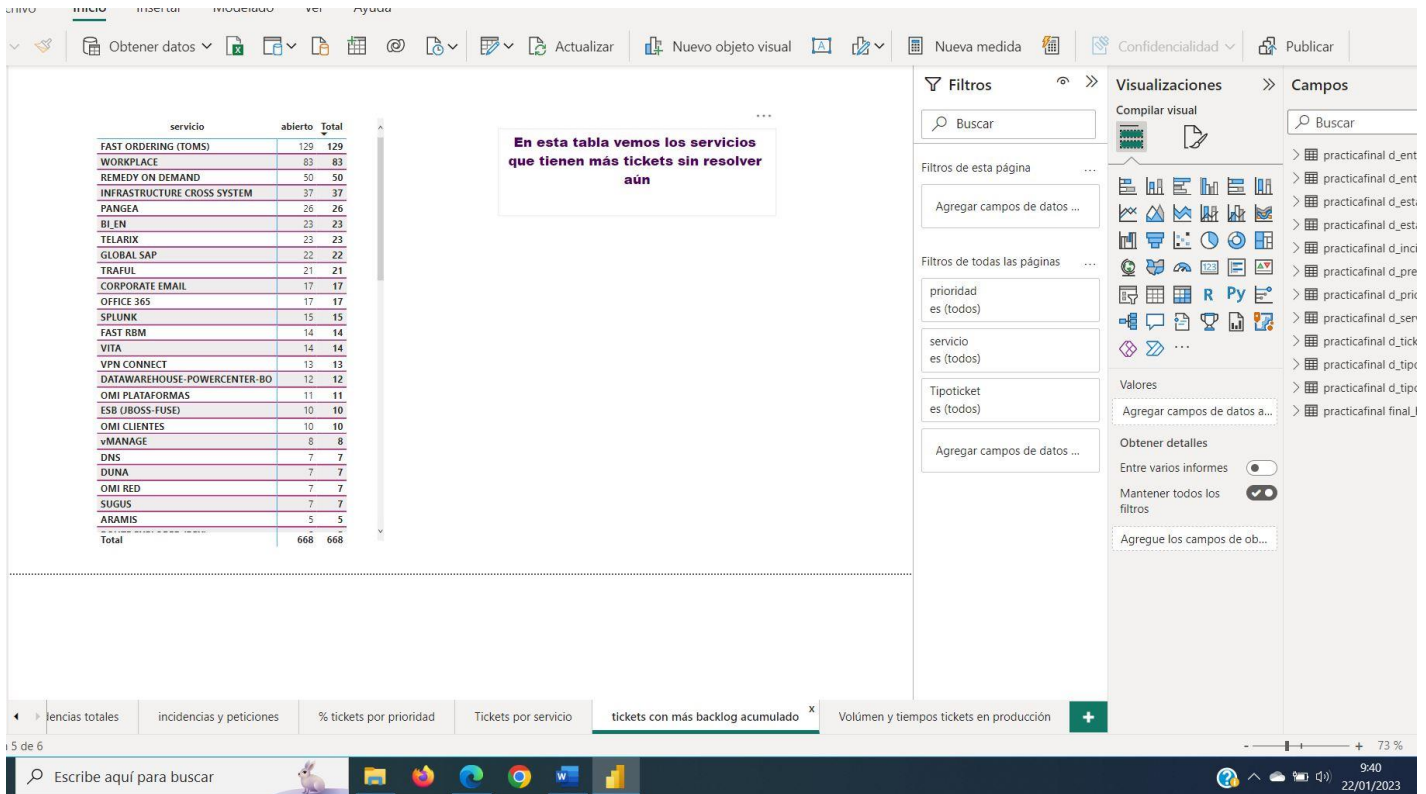
3. Cuáles son los servicios que acumulan más tickets y de ellos cuáles son los que menos cumplen los SLAs



Vemos en la tabla todos los servicios y la cantidad de tickets que tienen sin haber cumplido el tiempo acordado.

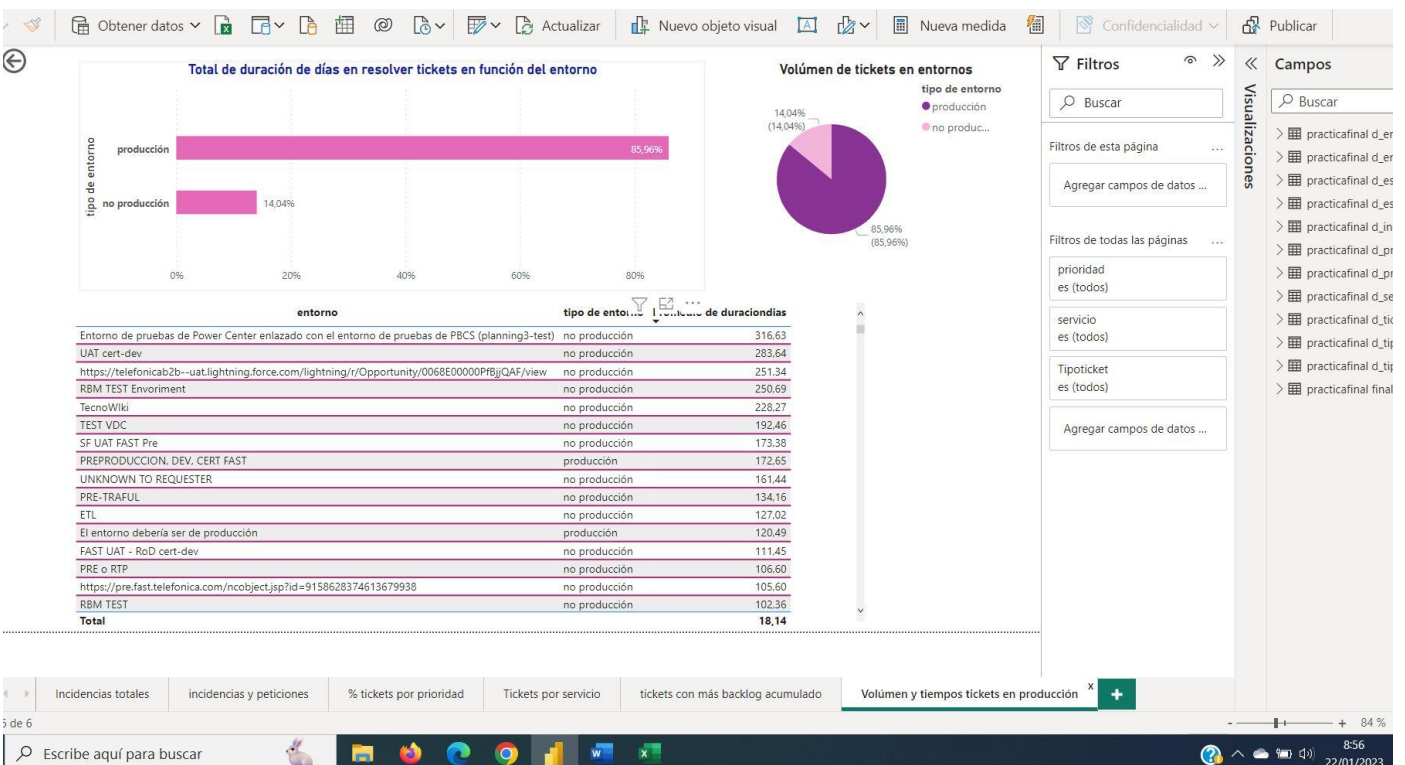
Y también los tickets totales (que entran en tiempo + se retrasan)

4. Cuáles son los servicios con más backlog acumulado (tickets abiertos en la actualidad)



En la tabla se ven los servicios que tienen más tickets abiertos por orden, mayor a menor

5.6. Cuáles son los entornos en que se resuelven más rápido y más lentos los tickets y volumetría de tickets en entornos productivos respecto a entornos no productivos



En la tabla se ven los tipos de entornos (producción, no producción), ordenados de más lento a la hora de cerrar incidencias, al más rápido.

Y en las gráficas vemos el volumen de un entorno respecto al otro. Está claro que los de mayor volumen son los de entornos productivos.

NOTA: a la hora de filtrar, no sé porqué no me ha cogido como no productivos los NON, no, No...así que el dato será un poco menor. Y un poco mayor los no productivos