

## PEC – DATA INTEGRATION

Cristian De Andrade C.

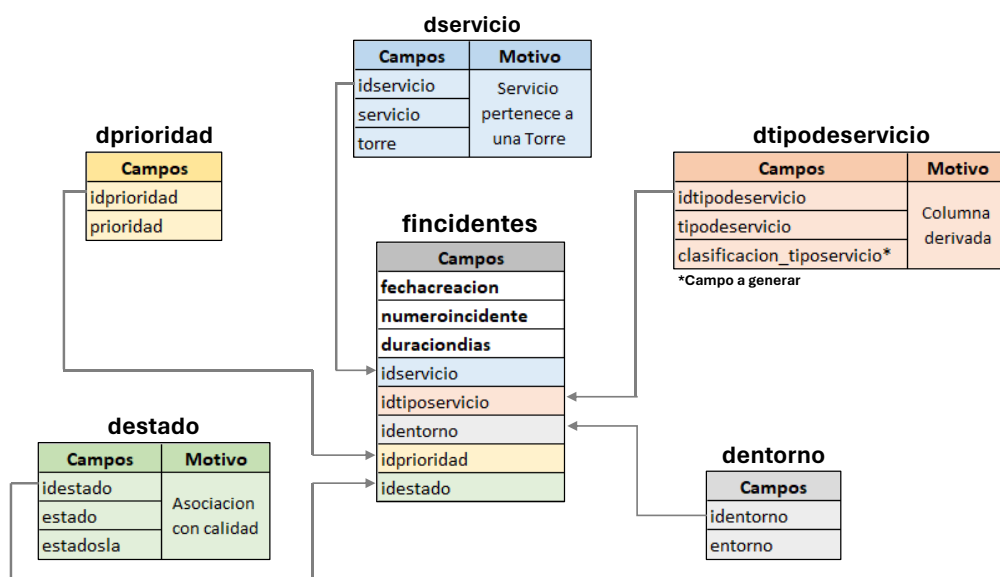
Como soporte, se adjuntan los siguientes archivos:

- **MySQL Scripts - Tablas y Comprobaciones.sql:** Contiene los comandos de SQL para creación y consultas de tablas (*staging*, dimensiones y hechos), así como los *scripts* de comprobación.
- **MySQL Database Export.sql:** Export de la base de datos de MySQL con las tablas creadas en el proceso: *staging*, dimensiones y tabla de hechos.
- **Archivos .ktr Pentaho (00.staging, 11.Maestros, 20.Hechos y 30.Exports):** Contienen las transformaciones para las distintas tablas/fases: *staging* (cargar y conexión de tabla de datos crudos), Maestros (generar y conectar tablas de dimensiones), Hechos (generar y conectar tabla de hechos) y Export (exportar tablas de dimensiones y hechos como .csv).
- **Archivo 00.total.kjb:** Flujo de trabajo en Pentaho para ejecución de las transformaciones .ktr y el export de las tablas para su explotación en Tableau Public.
- **Exports .csv de dimensiones y tabla de hechos:** Ficheros planos .csv generados a través de la transformación 30.Exports.ktr de Pentaho para su explotación/análisis en Tableau.
- **Dashboard - PEC - Cristian De Andrade C:** Link del dashboard en Tableau Public (también disponible en el apartado de “Exportación y Visualización” – pág. 5).

### ANÁLISIS INICIAL Y MODELO PROPUESTO

En base a los campos del fichero y la información que contienen, se planteó el esquema de estrella de la Esquema 1 para las dimensiones y tabla de hechos de la base de datos.

Los campos Prioridad y Entorno se sitúan de forma independiente en una tabla de dimensión, mientras que se realizaron las siguientes relaciones en pares: 1) servicio y torre, 2) Tipo de Servicio y Clasificación del Tipo de Servicio (transformación en base a descripción del campo) y 3) Estado y Estado cumplimiento SLA; se especifica el motivo principal en el esquema.



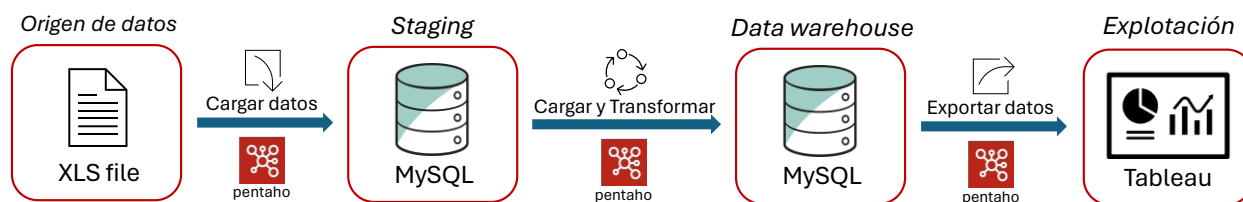
Esquema 1. Esquema de estrella propuesto para las dimensiones y tabla de hechos de la base de datos.

Para construir la tabla de hechos, se designaron como evento central el identificador de incidente/ticket que es único y dos variables adicionales que lo complementan: fecha de creación del ticket y la duración del mismo (qué representa cuánto dura), incluyendo, además, los identificadores de las tablas de dimensiones por mapeo.

Se excluyó la descripción del proceso pues no entra en el *scope* del análisis en base a las necesidades planteadas por el cliente y es información que requiere métodos más especializados para analizar esa data no estructurada, pero se mantuvo disponible en la carga inicial de datos crudos en la tabla *staging*.

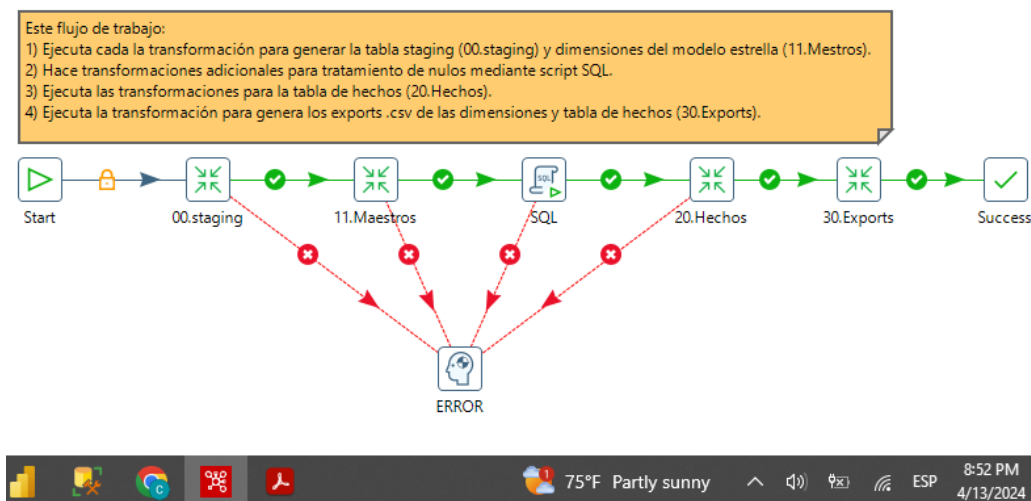
## BASE DE DATOS Y RESUMEN DE TRANSFORMACIONES

Con relación al proceso de articulado de la base de datos (Esquema 2), a modo resumen, se partió del fichero *.xls* con los datos, se cargaron y transformaron a una base de datos en el entorno de MySQL según el Esquema 1 mediante una integración en Pentaho, habilitado también para exportar y poder explotar los datos.



**Esquema 2.** Proceso para el armado de la base de datos.

Visto desde el entorno de Pentaho, el flujo de trabajo desarrollado para la ejecución de las distintas transformaciones (Imagen 1) involucra: carga inicial de datos para generar la tabla staging (00.staging), generar las cinco tablas de dimensiones (11.Maestros), *scripts* adicionales para tratar datos nulos (SQL), construir la tabla de hechos (20.Hechos) y generación de los *exports* de las tablas en formato *.csv*. Para más detalles, ver imágenes 22-24 en Anexos o referirse a los archivos *.ktr* adjuntos de Pentaho de cada transformación mencionada.



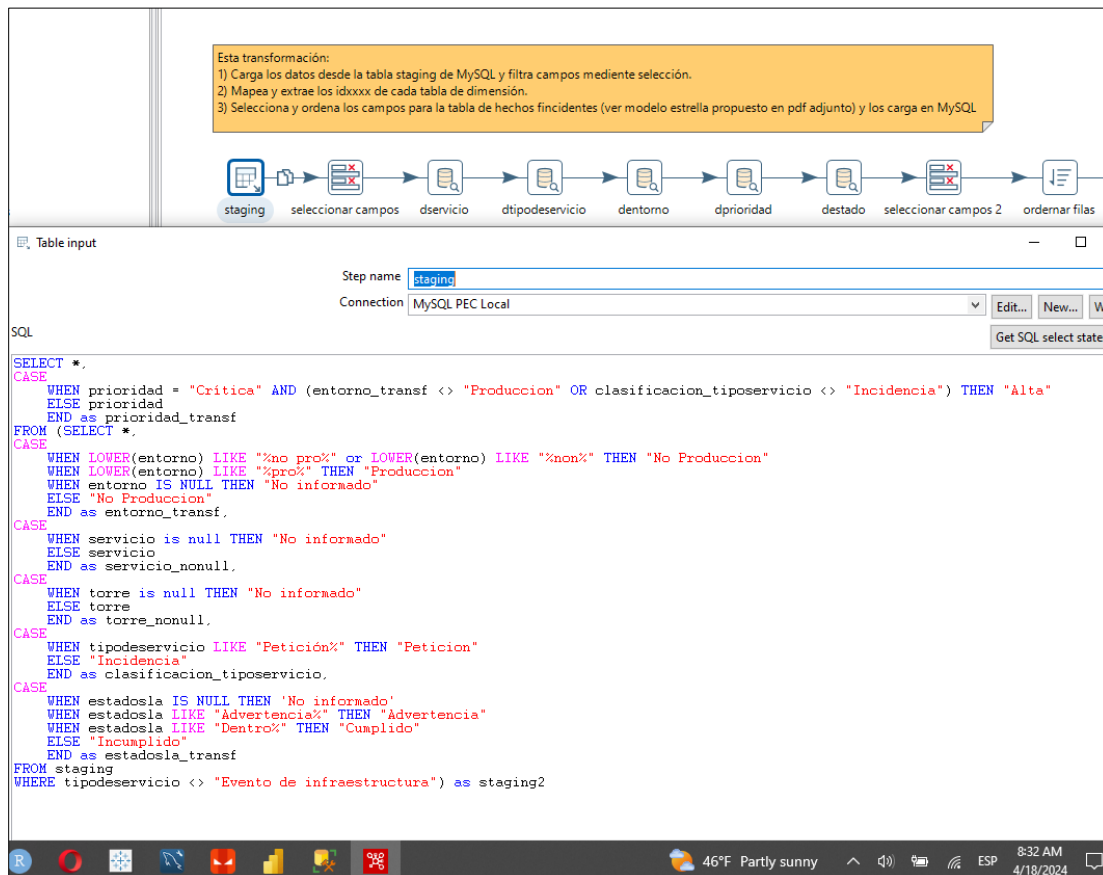
**Imagen 1.** Flujo de trabajo *00.total.kjb* en el entorno de Pentaho para la ejecución progresiva del compendio de transformaciones para la base de datos.

Las transformaciones más relevantes (Tabla 1) se enfocaron en: 1) simplificar las categorías de algunos campos y facilitar su comprensión/análisis, 2) la adición de un nuevo campo para clasificar el tipo de servicio, 3) corregir los datos de tickets con prioridad crítica y 4) excluir la hora del formato de la fecha de creación de los tickets/incidentes.

**Tabla 1.** Resumen de transformaciones de los datos.

<b>Campo</b>	<b>Alteración / Modificación</b>	<b>Cómo</b>
<i>Fecha de Creación</i>	<b>Cambiar Formato:</b> En base al objetivo de estudio, se usó solo el día, mes y año para la tabla de hechos; la hora es un dato adicional que no aporta información relevante.	<b>MySQL Scripts - Tablas y Comprobaciones:</b> Crear el campo "fechacreacion" en la tabla de hechos receptora como tipo DATE en vez de DATETIME.  <b>Pentaho - 20.Hechos:</b> Modificar formato de fecha en paso "seleccionar campos".
<i>Tipo de Servicio</i>	<b>Remover categoría:</b> Se excluyó la categoría "Evento de infraestructura" con sólo 43 tickets/incidentes asociados.	<b>Pentaho - 11.Maestros:</b> filtrar en la cargar de tabla los <i>tipodeservicio</i> que no son "Evento de infraestructura" para generar la tabla de dimensión <i>dservicio</i> .  <b>Pentaho - 20.Hechos:</b> Filtrar la tabla <i>staging</i> con un WHERE <> al hacer el mapeo para la tabla de hechos.
<i>Tipo de Servicio</i>	<b>Crear Nueva Columna/Variable:</b> A partir de "Tipo de Servicio", se generó un nuevo campo "clasificacion_tiposervicio" con dos categorías: 1) "Incidencia" en caso de que el servicio sea donde "Restauración de servicio a usuario" o "Restauración de infraestructura" y 2) "Petición" para "Petición de serv. por el usuario".	<b>Pentaho - 11.Maestros:</b> Agregar un campo adicional <i>tipodeservicio</i> como copia para modificarla usando la transformación "Replace in String"  <b>Pentaho - 20.Hechos:</b> Generar columna <i>clasificacion_tiposervicio</i> a partir de la tabla <i>staging</i> con CASE para luego hacer el mapeo con la tabla de dimensión <i>dservicio</i> .
<i>Entorno</i>	<b>Renombrar categorías en un nuevo campo:</b> Se redujo las categorías a únicamente dos, distinguiendo entre entorno de "Producción" y "No Producción" definiendo dos condiciones centrales (ver <i>Cómo</i> ) y excluyendo los datos nulos.	<b>Pentaho - 11.Maestros:</b> usar la función CASE al cargar los datos para generar la nueva columna considerando como entorno de "Produccion" a todo aquellos elementos que fuesen contienen "pro" a menos que se especifique "no pro" o "non", que en su caso sería "No Produccion", así como el resto.  <b>Pentaho - 20.Hechos:</b> Generar columna <i>entorno_transf</i> a partir de la tabla <i>staging</i> con CASE para luego hacer el mapeo con la tabla de dimensión <i>dentorno</i> .
<i>Prioridad</i>	<b>Corrección de categoría:</b> Se recategorizó los tickets de prioridad crítica que no fuesen incidentes en entornos de como prioridad "Alta".	<b>Pentaho - 20.Hechos:</b> Generar columna <i>prioridad_transf</i> a partir de la tabla <i>staging2</i> (vía sub-query) con CASE para luego hacer el mapeo con la tabla de dimensión <i>dprioridad</i> .

<p><i>Estado cumplimiento SLA</i></p>	<p><b>Renombrar categorías:</b> Se renombraron las categorías como “Cumplido”, “Advertencia” o “Incumplido” para simplificar su significado.</p>	<p><b>Pentaho - 11.Maestros :</b> Modificar campo <i>estadosla</i> usando la transformación "Replace in String".</p> <p><b>Pentaho - 20.Hechos:</b> Generar columna <i>estadosla_transf</i> a partir de la tabla staging con CASE para luego hacer el mapeo con la tabla de dimensión <i>destado</i>.</p>
---------------------------------------	--	---



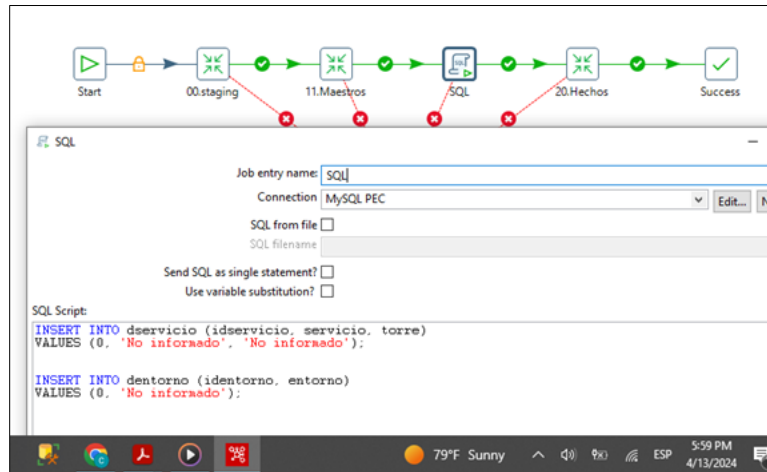
**Imagen 2.** Alteraciones de los datos de la tabla *staging* en Pentaho mediante código SQL para la construcción de la tabla de hechos.

#### Notas:

- No se excluyeron los tickets cancelados durante la generación de tabla de dimensiones o hechos de manera que estuviesen disponibles y poder responder preguntas del cliente que puedan surgir a futuro en relación a los estos; esta exclusión se realizó durante la generación *exports* para explotación (30.Exports).
- Se designó “No informado” como identificador de remplazo para todos los datos nulos mediante el siguiente tratamiento:
  - **Tipo de Servicio y Torre:** Los casos en que “Tipo de Servicio” o “Torre” fuese nulo, se trataron directamente en el primer paso “servicios” de Pentaho 00.staging mediante código SQL para transformarlos a “No informado”. Y en el caso de cuando ambos eran nulos, se agregó posteriormente

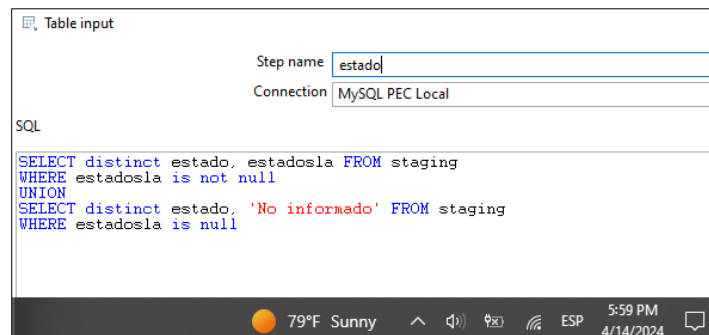
en Pentaho con un script SQL (ver archivo Pentaho [00.total.kjb](#)) para asignarle índice “0” a “No informado”/“No informado” (Imagen 3).

- **Entorno:** Se agregó el caso “No informado” como índice “0” con Pentaho como un paso SQL (Imagen 3) del flujo de ejecución de transformaciones (ver archivo Pentaho [00.total.kjb](#))



**Imagen 3.** Script SQL en 00.trabajo para la adición de un identificador de nulos “0 - No informado” para las dimensiones *dservicio* y *dentorno*.

- **Estado cumplimiento SLA:** Se trataron los casos “No informado” en combinación con el campo “Estado” directamente en el primer paso “estado” de Pentaho mediante código SQL usando UNION (Imagen 4).



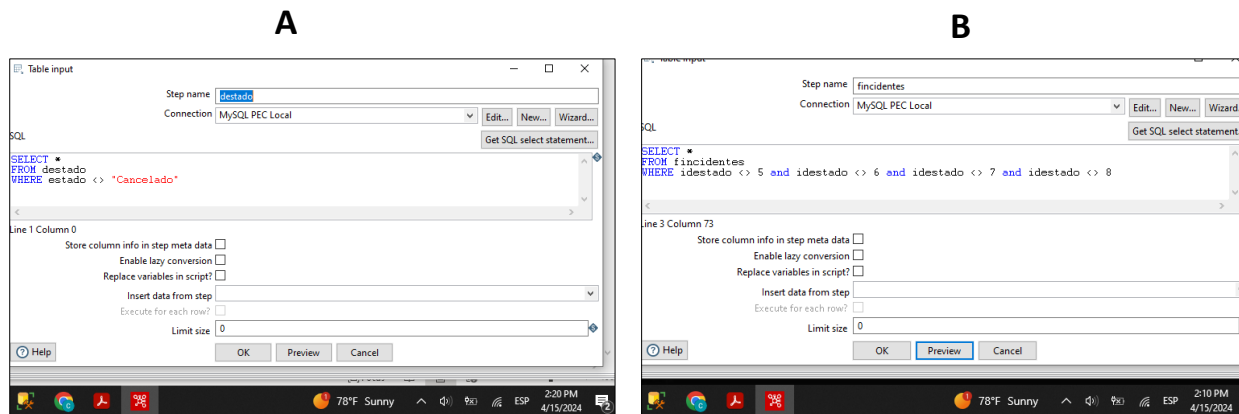
**Imagen 4.** Script SQL en 11.Maestros/estado para la adición de “No informado” para la dimensión *destado*.

---

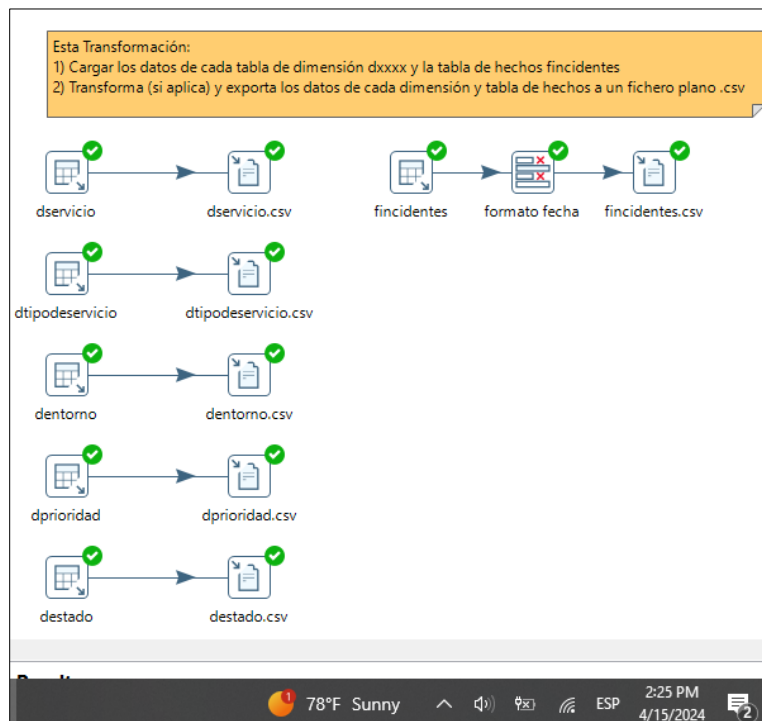
## EXPORTACIÓN Y VISUALIZACIÓN

En base a las directrices e interrogantes aportadas por el cliente, se realizaron las siguientes alteraciones para la exportación de la tabla de hechos (Imagen 6): 1) se excluyeron los índices 5-8 de “idestado” (tickets cancelados - Imagen 5B), los cuales no se computan y 2) se modificó el formato de la *fechacreacion* para excluir la hora.

Las tablas de dimensiones se exportaron tal sin alteraciones para su explotación/análisis mediante Tableau, a excepción de la tabla *destado* en la cual también se excluyeron los índices 5-8 de tickets cancelados (Imagen 5A - 30.Export.ktr).

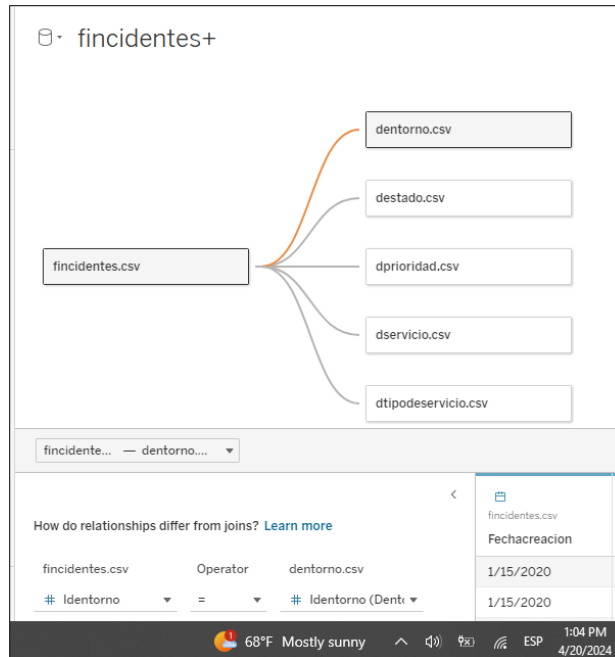


**Imagen 5.** Exclusión de tickets cancelados. A: tabla de dimensión *destado*, B: tabla de hechos *fincientes*.



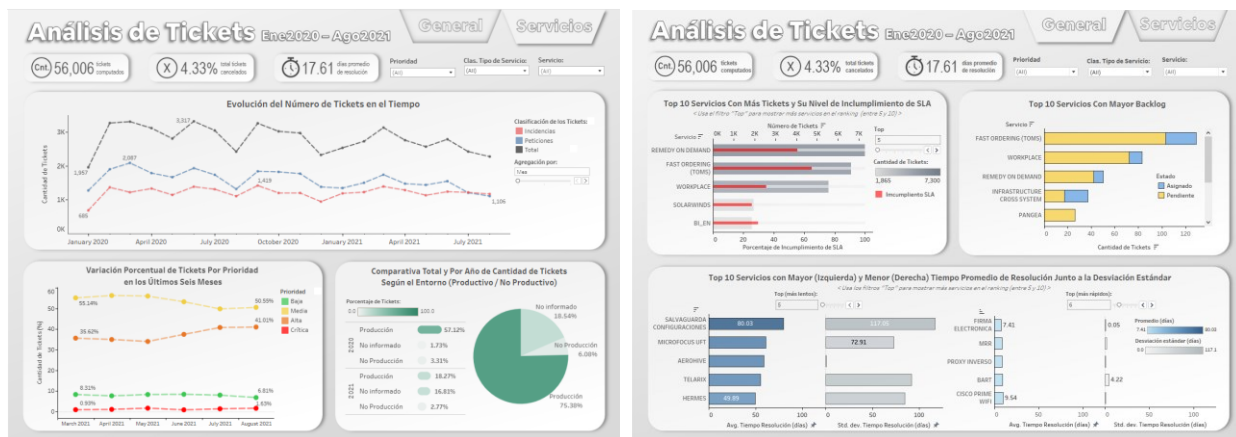
**Imagen 6.** Transformación de Pentaho 30.Export.ktr para la exportación de tablas de dimensiones y tabla de hechos como ficheros planos para análisis/visualización.

Con los ficheros .csv se realizó un análisis mediante visualizaciones en Tableau para dar respuesta a las interrogantes planteadas por el cliente. Allí se conectaron las tablas de dimensiones con la tabla de hechos mediante identificadores (Imagen 7).



**Imagen 7.** Relación entre tabla de hechos *fincidentes* con las cinco dimensiones del modelo de estrella en Tableau.

El *dashboard* se dividió en 2 secciones: “General”, con información de variación temporal y proporción de tickets en entornos productivos/no productivos, y “Servicios”, en donde se congregan los análisis de relacionados a la categoría con el mismo nombre, además, se añadió información adicional como el número de tickets y tiempo promedio de resolución (Imagen 8). En este [link](#) se puede acceder al *dashboard* interactivo disponible en Tableau Public.



**Imagen 8.** Diseño de las dos secciones del *dashboard* para el análisis de los datos de tickets.

# COMPROBACIONES

- Fichero .xls original (mediante tabla dinámica) vs Tabla *staging* en MySQL

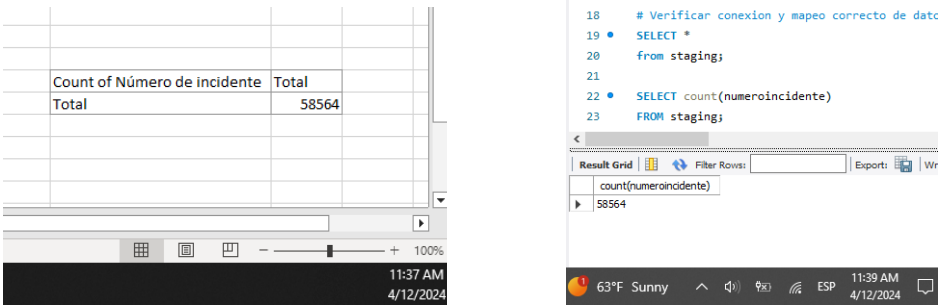


Imagen 9. Comparación conteo de número total de filas.

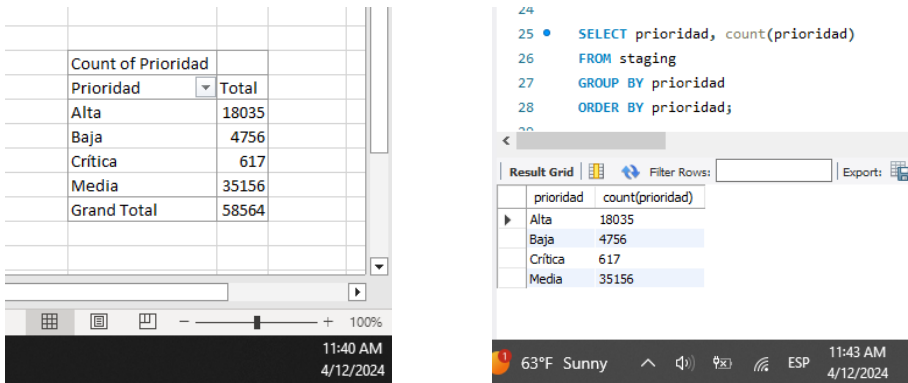


Imagen 10. Comparación de conteo agrupación simple por prioridad.

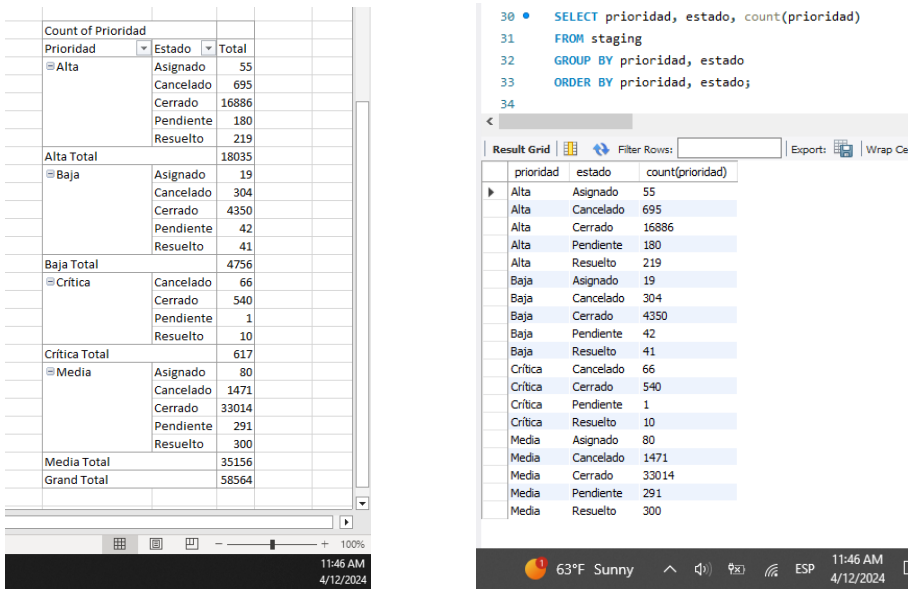


Imagen 11. Comparación de conteo agrupación doble por prioridad y estado.



- Correcta conexión de las tablas de dimensiones desde Pentaho a MySQL

```
62
63 • SELECT *
64 FROM dservicio;
```

idservicio	servicio	torre
0	No informado	No informado
1	ACTIVE DIRECTORY OSS	INFRA
2	AEROHIVE	BSS
3	ALARM SYSTEM	TICKETING TOOL
4	ALCHEMY	No informado
5	ALCHEMY	OSS
6	API CONNECT	BSS
7	API GATEWAY	BSS
8	ARAMIS	BSS
9	ARIS - MAPRO	INFRA

SD/GBP +0.89% 5:13 PM 4/12/2024

```
72
73 • SELECT *
74 FROM dtipodeservicio;
```

idtipodeservicio	tipodeservicio	clasificacion_tiposervicio
1	Petición de serv. por el usuario	Peticion
2	Restauración de servicio a usuario	Incidencia
3	Restauración de infraestructura	Incidencia
NULO	NULO	NULO

75°F Sunny 5:14 PM 4/12/2024

```
82 • SELECT *
83 FROM dentorno;
```

identorno	entorno
0	No informado
1	No Produccion
2	Produccion
NULO	NULO

UV 5:15 PM 4/12/2024

```
78 • SELECT *
79 FROM dprioridad;
```

idprioridad	prioridad
1	Alta
2	Baja
3	Critica
4	Media
NULO	NULO

8:34 AM 4/18/2024

```
101 • SELECT *
102 FROM destado;
```

idestado	estado	estadosla
1	Asignado	Advertencia
2	Asignado	Cumplido
3	Asignado	Incumplido
4	Asignado	No informado
5	Cancelado	Advertencia
6	Cancelado	Cumplido
7	Cancelado	Incumplido
8	Cancelado	No informado
9	Cerrado	Cumplido
10	Cerrado	Incumplido
11	Cerrado	No informado
12	Pendiente	Advertencia
13	Pendiente	Cumplido
14	Pendiente	Incumplido
15	Pendiente	No informado
16	Resuelto	Cumplido
17	Resuelto	Incumplido
18	Resuelto	No informado
NULO	NULO	NULO

5°F Sunny 5:21 PM 4/12/2024

Imagen 12. Vista del contenido de las tablas de dimensiones en MySQL

- Volumetría de tickets según entorno

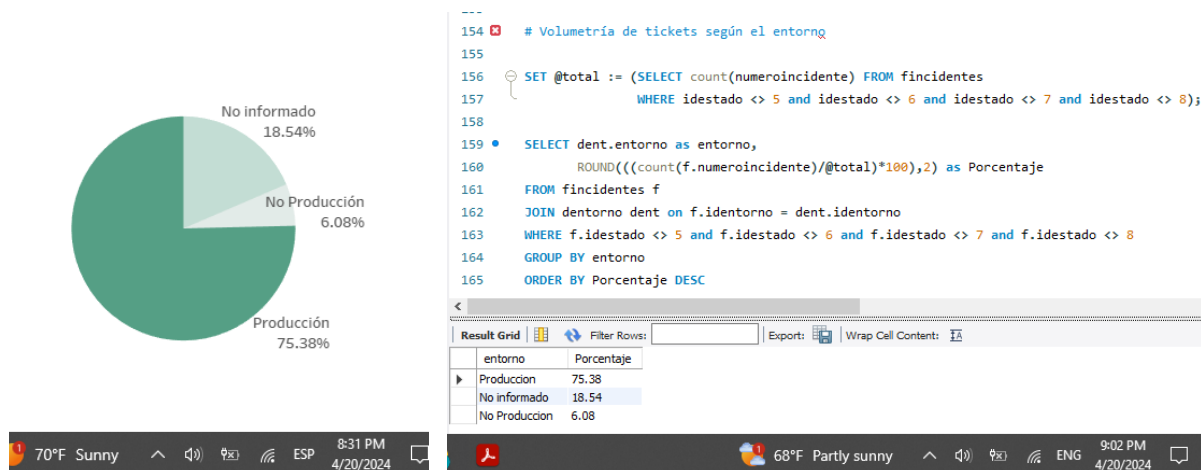
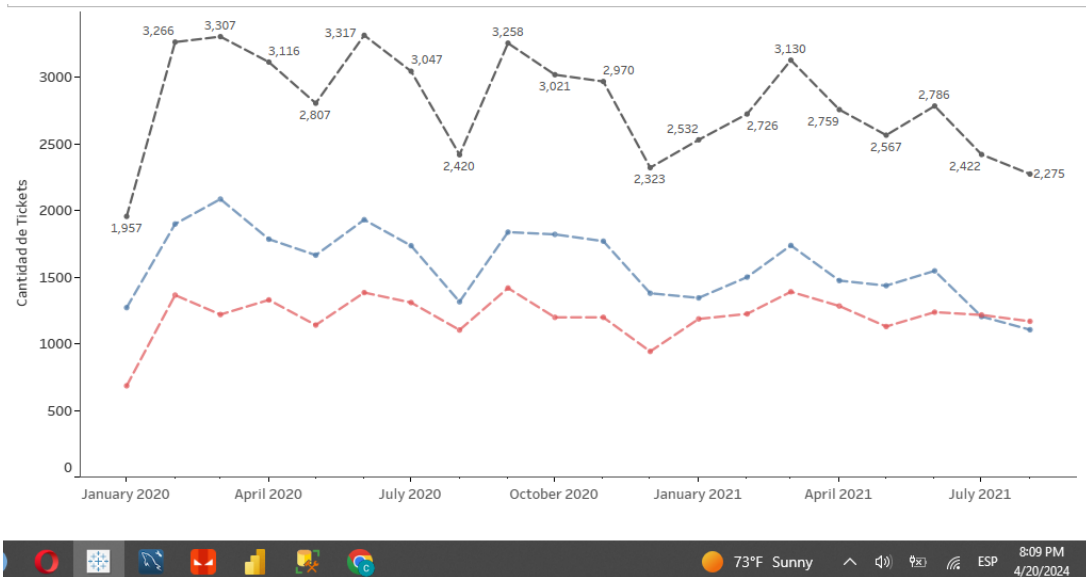


Imagen 13. Comprobación Tableau/MySQL de porcentaje de tickets según entorno (productivo / no productivo).

- Evolución de cantidad de tickets en el tiempo

Imagen 14. Comprobación Tableau/MySQL de total de tickets en el tiempo.



```

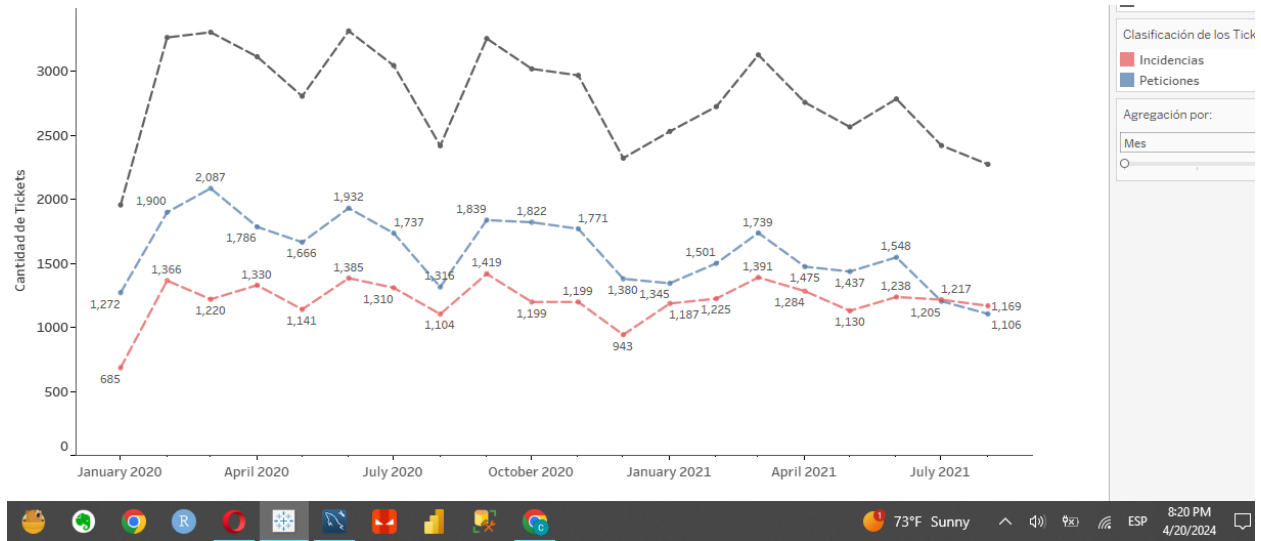
170 #total
171 • SELECT DATE_FORMAT(fechacreacion, '%Y-%m') as anyo_mes,
172        count(numeroincidente) as conteo
173 FROM   fincidentes f
174 JOIN   destado dest on f.idestado = dest.idestado
175 WHERE  estado <> 'Cancelado'
176 GROUP BY anyo_mes
177 ORDER BY anyo_mes;

```

Result Grid Filter Rows: Export: Wrap Cell Content:

	anyo_mes	conteo
►	2020-01	1957
	2020-02	3266
	2020-03	3307
	2020-04	3116
	2020-05	2807
	2020-06	3317
	2020-07	3047
	2020-08	2420
	2020-09	3258
	2020-10	3021
	2020-11	2970
	2020-12	2323
	2021-01	2532
	2021-02	2726
	2021-03	3130
	2021-04	2759
	2021-05	2567
	2021-06	2786
	2021-07	2422
	2021-08	2275

Imagen 15. Comprobación Tableau/MySQL de total de tickets en el tiempo por Incidencias/Peticiones.



```

179 #total por incidencias/peticiones
180 • SELECT DATE_FORMAT(f.fecha creacion, '%Y-%m') as anyo_mes,
181         dtipo.clasificacion_tiposervicio as tipo,
182         count(f.numeroincidente) as conteo
183 FROM fincidentes f
184 JOIN dtipodeservicio dtipo on f.idtipodeservicio = dtipo.idtipodeservicio
185 JOIN destado dest on f.idestado = dest.idestado
186 WHERE estado <> 'Cancelado'
187 GROUP BY anyo_mes, tipo
188 ORDER BY anyo_mes, tipo DESC;

```

Result Grid

anyo_mes	tipo	conteo
2020-01	Peticion	1272
2020-01	Incidencia	685
2020-02	Peticion	1900
2020-02	Incidencia	1366
2020-03	Peticion	2087
2020-03	Incidencia	1220
2020-04	Peticion	1786
2020-04	Incidencia	1330
2020-05	Peticion	1666
2020-05	Incidencia	1141
2020-06	Peticion	1932
2020-06	Incidencia	1385
2020-07	Peticion	1737
2020-07	Incidencia	1310
2020-08	Peticion	1316
2020-08	Incidencia	1104
2020-09	Peticion	1839
2020-09	Incidencia	1419
2020-10	Peticion	1822
2020-10	Incidencia	1199
2020-11	Peticion	1771
2020-11	Incidencia	1199
2020-12	Peticion	1380
2020-12	Incidencia	943
2021-01	Peticion	1345
2021-01	Incidencia	1187
2021-02	Peticion	1501
2021-02	Incidencia	1225
2021-03	Peticion	1739
2021-03	Incidencia	1391
2021-04	Peticion	1475
2021-04	Incidencia	1284
2021-05	Peticion	1437
2021-05	Incidencia	1130
2021-06	Peticion	1548
2021-06	Incidencia	1238
2021-07	Peticion	1217
2021-07	Incidencia	1205
2021-08	Peticion	1169
2021-08	Incidencia	1106

Result 32 x

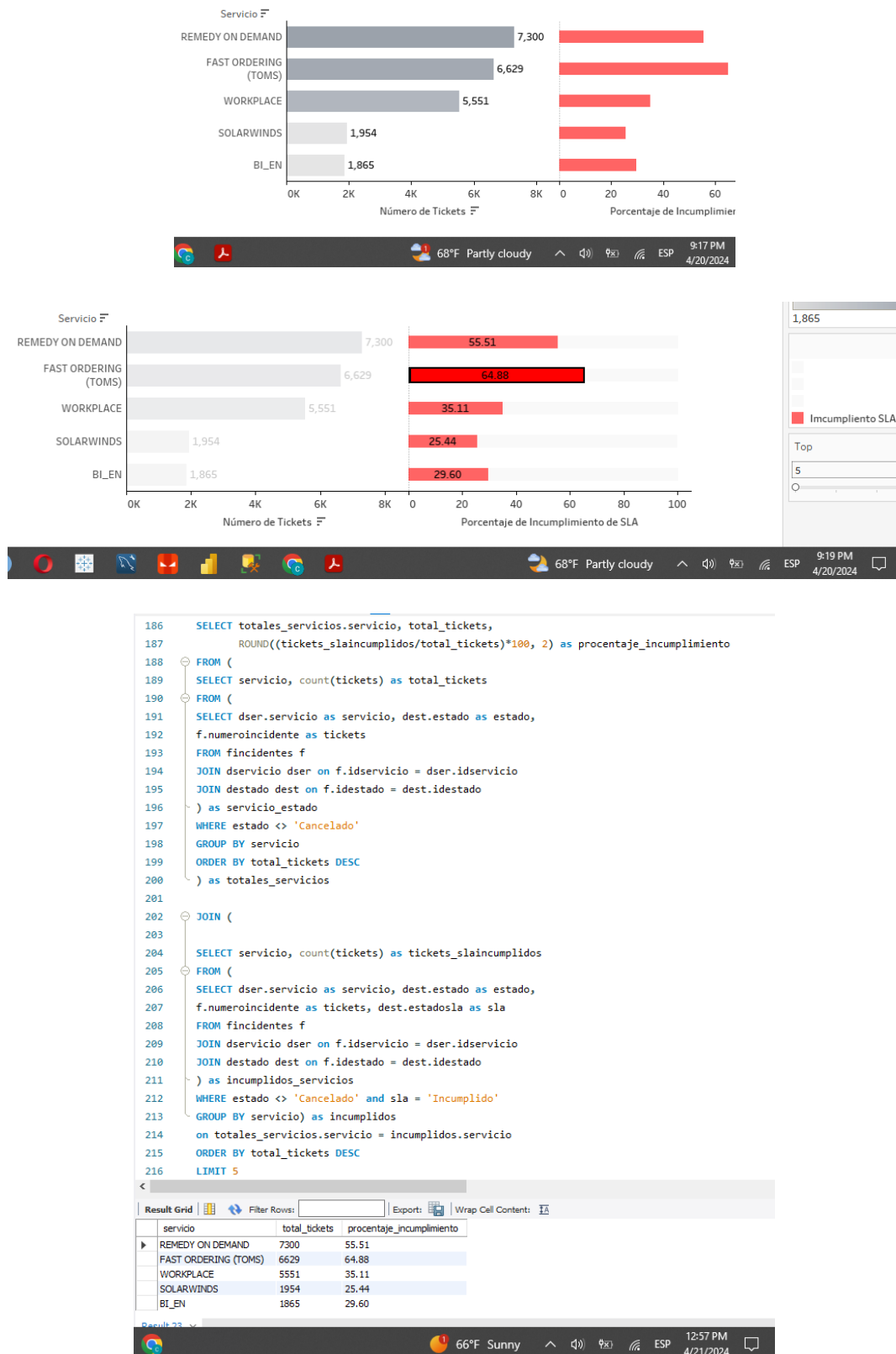
Output

Action Output

#	Time	Action
41	13:16:10	SELECT DATE_FORMAT(fecha creacion, '%Y-%m') as anyo_mes, count(numeroincidente) as...
42	13:16:39	SELECT DATE_FORMAT(fecha creacion, '%Y-%m') as anyo_mes, count(numeroincidente) as...

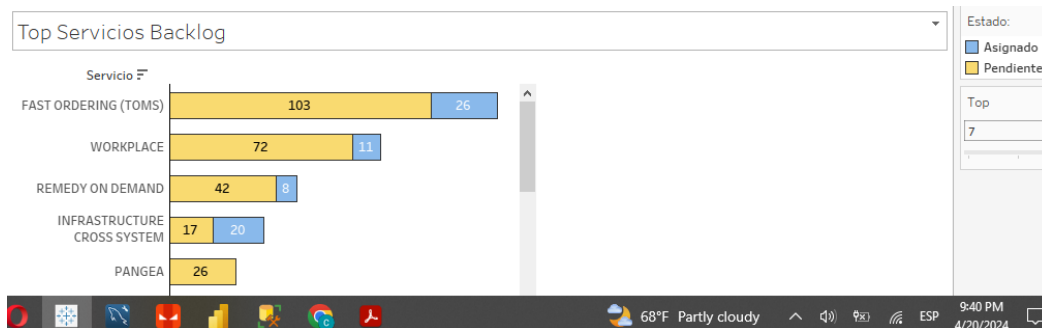
- Servicios con más tickets y su porcentaje de incumplimiento de SLA

Imagen 16. Comprobación Tableau/MySQL de servicios con más tickets y su porcentaje de incumplimiento de SLA.



- **Servicios con más backlog**

**Imagen 17.** Comprobación Tableau/MySQL de servicios con más tickets abiertos (“Asignados” + “Pendientes”).



```

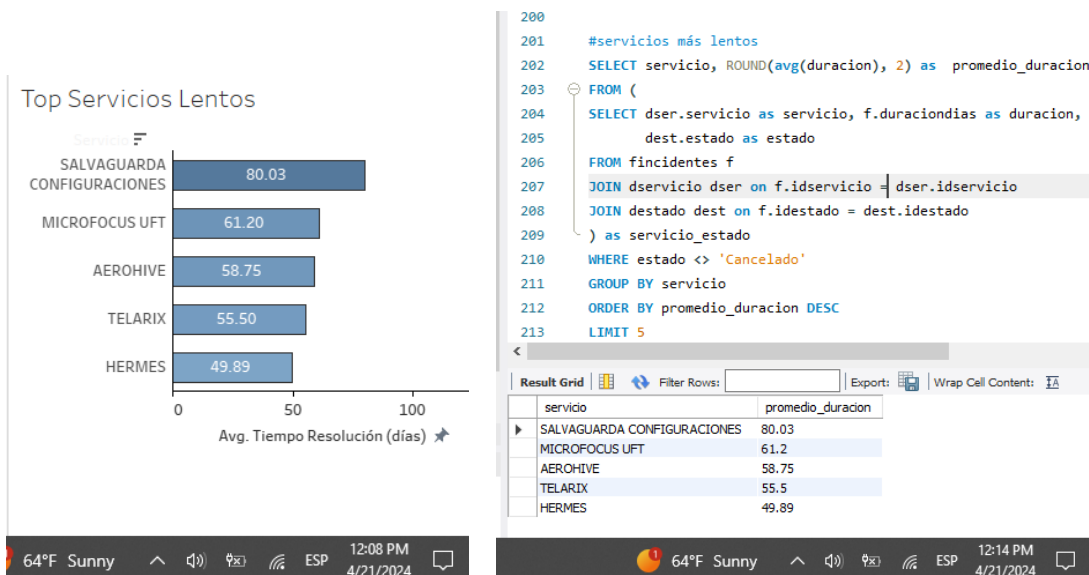
186 SELECT servicio, count(tickets) as conteo_tickets
187 FROM (
188     SELECT dser.servicio as servicio, dest.estado as estado,
189            f.numeroincidente as tickets
190     FROM fincidentes f
191     JOIN dservicio dser on f.idservicio = dser.idservicio
192     JOIN destado dest on f.idestado = dest.idestado
193 ) as servicio_estado
194 WHERE estado = 'Asignado' or estado = 'Pendiente' and estado <> 'Cancelado'
195 GROUP BY servicio
196 ORDER BY conteo_tickets DESC
197 LIMIT 5

```

Result Grid	
servicio	conteo_tickets
FAST ORDERING (TOMS)	129
WORKPLACE	83
REMEDY ON DEMAND	50
INFRASTRUCTURE CROSS SYSTEM	37
PANGEA	26

- **Servicios con mayor y menor tiempo de resolución promedio**

**Imagen 18.** Comprobación Tableau/MySQL de servicios con mayor tiempo de resolución promedio.



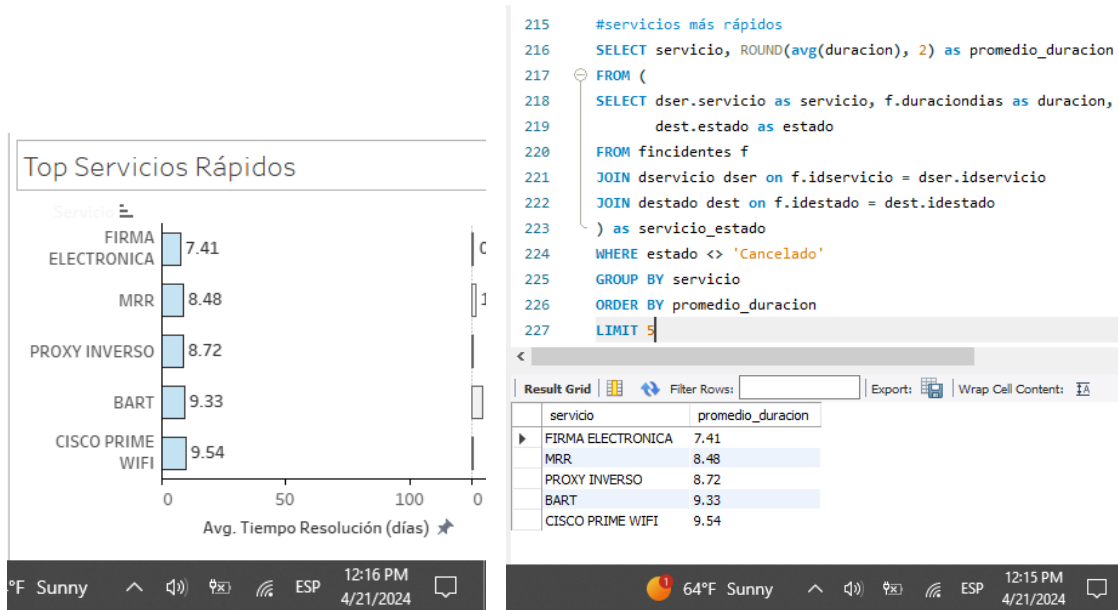
```

200
201 #servicios más lentos
202 SELECT servicio, ROUND(avg(duracion), 2) as promedio_duracion
203 FROM (
204     SELECT dser.servicio as servicio, f.duraciondias as duracion,
205            dest.estado as estado
206     FROM fincidentes f
207     JOIN dservicio dser on f.idservicio = dser.idservicio
208     JOIN destado dest on f.idestado = dest.idestado
209 ) as servicio_estado
210 WHERE estado <> 'Cancelado'
211 GROUP BY servicio
212 ORDER BY promedio_duracion DESC
213 LIMIT 5

```

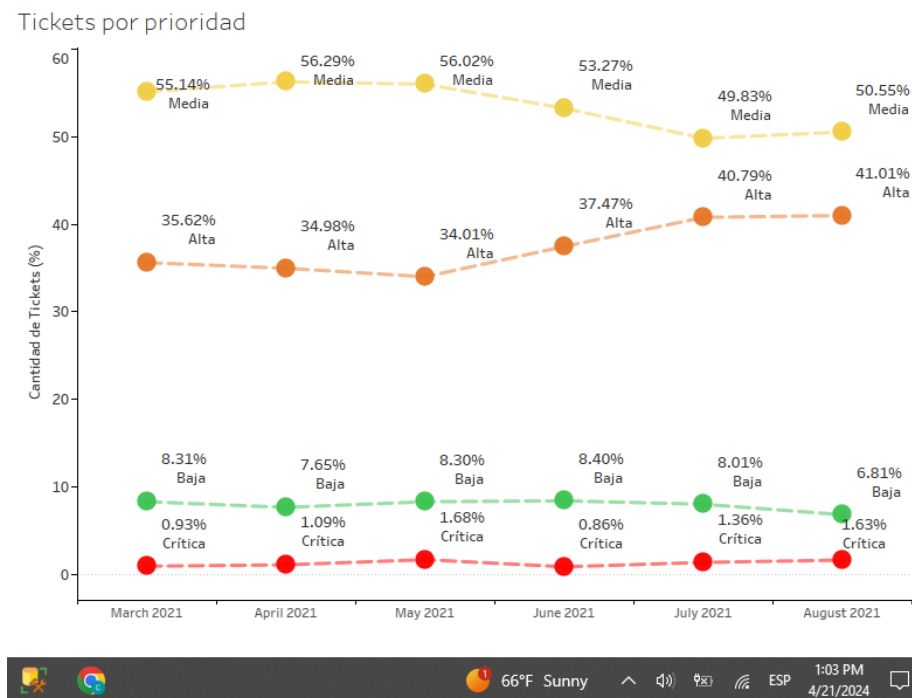
Result Grid	
servicio	promedio_duracion
SALVAGUARDA CONFIGURACIONES	80.03
MICROFOCUS UFT	61.2
AEROHIVE	58.75
TELARIX	55.5
HERMES	49.89

**Imagen 19.** Comprobación Tableau/MySQL de servicios con menor tiempo de resolución promedio.



- **Variación de la prioridad de los tickets en los últimos meses**

**Imagen 20.** Comprobación Tableau/MySQL de porcentaje de tickets por prioridad en los últimos 6 meses.



```

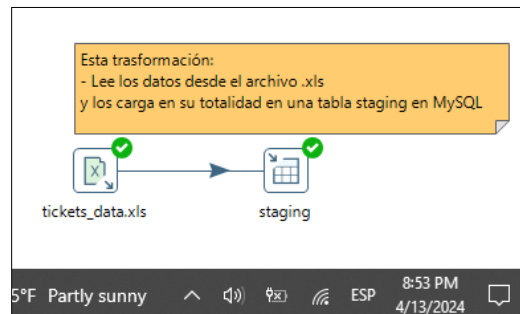
271 • SELECT total_por_fechaprioridad.anyo_mes, prioridad,
272        ROUND((conteo_mes_por_prioridad/conteo_total_mes)*100, 2) as porcentaje_ticket
273 FROM (
274   SELECT anyo_mes, prioridad, count(tickets) as conteo_mes_por_prioridad
275   FROM (
276    SELECT DATE_FORMAT(fechacreacion, '%Y-%m') as anyo_mes, dest.estado as estado,
277    f.numeroincidente as tickets, dpri.prioridad as prioridad
278    FROM fincidentes f
279    JOIN dprioridad dpri on f.idprioridad = dpri.idprioridad
280    JOIN destado dest on f.idestado = dest.idestado
281   ) as fecha_prioridad
282   WHERE estado <> 'Cancelado'
283   GROUP BY anyo_mes, prioridad
284   ORDER BY anyo_mes
285  ) as total_por_fechaprioridad
286
287 JOIN (
288
289   SELECT DATE_FORMAT(fechacreacion, '%Y-%m') as anyo_mes,
290   count(numeroincidente) as conteo_total_mes
291   FROM fincidentes f
292   JOIN destado dest on f.idestado = dest.idestado
293   WHERE estado <> 'Cancelado'
294   GROUP BY anyo_mes
295   ORDER BY anyo_mes) as total_por_fecha
296   on total_por_fechaprioridad.anyo_mes = total_por_fecha.anyo_mes
297   ORDER BY total_por_fechaprioridad.anyo_mes DESC, porcentaje_tickets DESC
298   LIMIT 24;

```

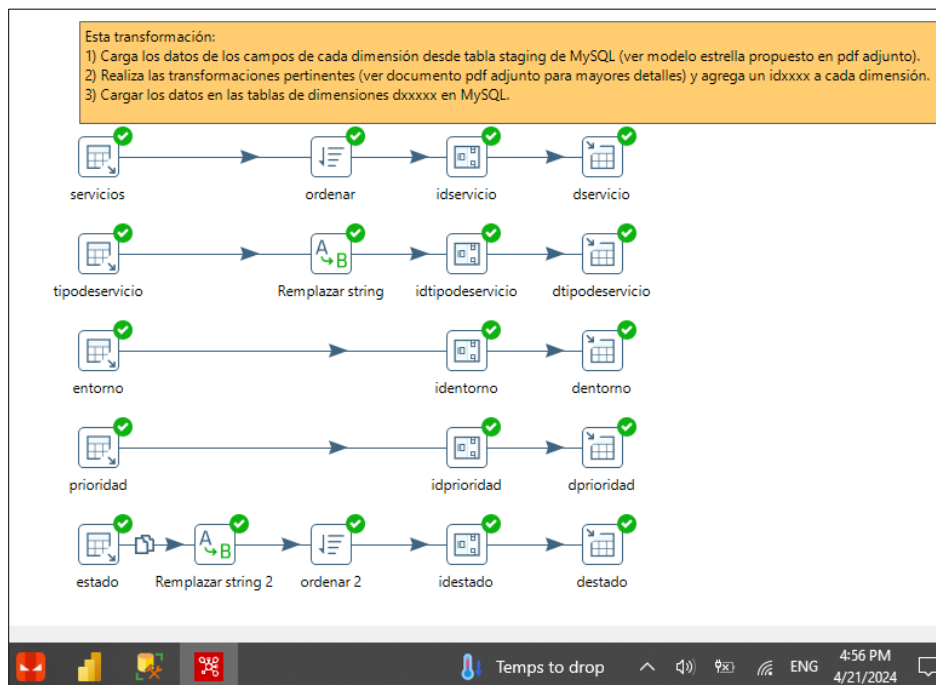
	anyo_mes	prioridad	porcentaje_tickets
▶	2021-08	Media	50.55
	2021-08	Alta	41.01
	2021-08	Baja	6.81
	2021-08	Crítica	1.63
	2021-07	Media	49.83
	2021-07	Alta	40.79
	2021-07	Baja	8.01
	2021-07	Crítica	1.36
	2021-06	Media	53.27
	2021-06	Alta	37.47
	2021-06	Baja	8.40
	2021-06	Crítica	0.86

	anyo_mes	prioridad	porcentaje_tickets
	2021-05	Media	56.02
	2021-05	Alta	34.01
	2021-05	Baja	8.30
	2021-05	Crítica	1.68
	2021-04	Media	56.29
	2021-04	Alta	34.98
	2021-04	Baja	7.65
	2021-04	Crítica	1.09
	2021-03	Media	55.14
	2021-03	Alta	35.62
	2021-03	Baja	8.31
	2021-03	Crítica	0.93

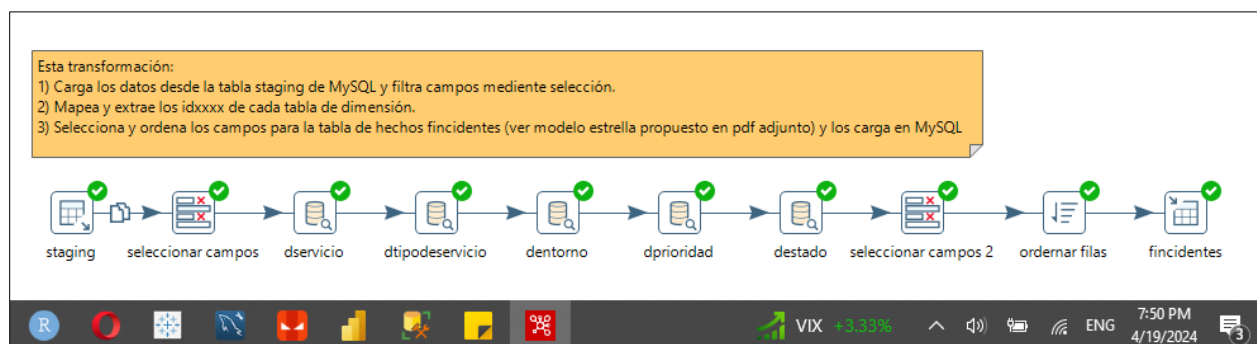
## ANEXOS



**Imagen 22.** Transformación de Pentaho *00.staging.ktr* para la carga inicial de los datos y conexión con tabla *staging* en MySQL.



**Imagen 23.** Transformación de Pentaho *11.Maestros.ktr* para la creación de las tablas de dimensiones según el modelo estrella propuesto y conexión con tablas de dimensiones en MySQL.



**Imagen 24.** Transformación de Pentaho *20.Hechos.ktr* para la creación de la tabla de hechos según el modelo estrella propuesto y conexión con tabla de hechos *fncidentes* en MySQL.