

**MÓDULO:** El trabajo del científico de datos: Storytelling.

**PROFESOR:** Abel González Durán

## 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.

## Requerimientos del cliente

Nuestro cliente es un departamento de una multinacional del sector de las telecomunicaciones que gestiona las peticiones e incidencias que se producen en una serie de servicios y entornos que la compañía da a sus empleados y clientes.

Cada vez que hay una petición de acceso o una incidencia de funcionamiento, los usuarios abren un ticket en la herramienta de ticketing para informar y que se proceda a su solución. El cliente nos ha hecho llegar una exportación de datos del año 2020 y hasta finales de agosto de 2021 de cara a analizar algunas preguntas que querrían resolver en formato gráfico mediante informes o dashboards.

El cliente pone a nuestra disposición la siguiente arquitectura tecnológica:

- Base de datos MySQL
- ETL Pentaho Data Integration
- Frontend analítico Microsoft PowerBI o Tableau Public

A modo de resumen, nuestro cliente quiere saber:

- La evolución del volumen de tickets general y diferenciado por incidencias y peticiones.
- Si ha variado la prioridad con la que se abren los tickets en los últimos meses (% de peso por prioridad)
- Cuáles son los servicios que acumulan más tickets y de ellos cuáles son los que menos cumplen los SLAs
- Cuáles son los servicios con más backlog acumulado (tickets abiertos en la actualidad)
- Cuáles son los servicios en que se resuelven más rápido y más lentos los tickets
- Volumetría de tickets en entornos productivos respecto a entornos no productivos
- Posibilidad de filtrar por prioridad, por tipo de ticket y por servicio.

Nuestro cliente nos indica que le gustaría disponer de la siguiente documentación del proyecto:

- Esquema de base de datos
- Documentación de los procesos de carga
- Comprobaciones realizadas de la exactitud de los datos
  - Valores esperados según dataset de carga
  - Comprobaciones en SQL (via query directa)
  - Comparación con dashboard (datos MySQL vs resultados en PowerBI/Tableau)
- Funcionalidades y visualizaciones desarrolladas en PowerBI o Tableau
- Decisiones de diseño tomadas en los diferentes pasos

## Dataset

El dataset suministrado es un fichero Excel que contiene los siguientes campos.

El cliente nos ha dado un pequeño detalle de lo que hay en cada uno.

- Fecha de creación. Especifica el momento en que se crea el ticket. Servirá para hacer el análisis temporal.
- Número de incidente. Identificador único del ticket.
- Descripción. Breve descripción de la petición o incidencia.
- Servicio. Los diferentes servicios que presta la compañía a sus usuarios y clientes.
- Tipo de servicio. Especifica si es una petición o una incidencia:
  - Incidencias: Restauración de infraestructura o restauración de servicio a usuario
  - Peticiones: Petición de serv. Por el usuario
- Prioridad. Especifica la prioridad con la que debe ser tratado el ticket.
  - Hay cuatro prioridades, ordenadas de más importante a menos: Crítica, Alta, Media y Baja.
  - Sólo las incidencias pueden tener prioridad crítica y sólo deben considerarse en entornos de producción.
- Estado. Indica la situación del ticket.
  - Un ticket se considera acabado si está en estado Cerrado o Resuelto.
  - Un ticket se considera abierto si está pendiente o asignado.
  - Los tickets cancelados no se computan.
- Torre. Es un agrupador funcional de servicios. Cada servicio pertenece a una torre de gestión.
- Entorno. Indica el entorno en el que afecta el ticket. Los entornos productivos son todos aquellos que contienen un PRO.
- Estado cumplimiento. Indica si el ticket se ha resuelto en los tiempos requeridos por el cliente o no.
  - Dentro del objetivo de servicio: ticket resuelto adecuadamente
  - Objetivos de servicio incumplidos: ticket resuelto tarde
- Duración días. Número de días que se ha tardado en resolver el ticket.

## Método de evaluación

El método de evaluación se basará en cuatro aspectos de la práctica.

- 1) Análisis de los datos iniciales y modelo de datos propuesto (2,5 puntos)
- 2) Desarrollo de modelo en MySQL (2,5 puntos)
- 3) Desarrollo de cargas con Pentaho Data Integration (2,5 puntos)
- 4) Informes realizados con PowerBI/Tableau (2,5 puntos)

## Consideraciones

Para todos los ejercicios en general, cuando se adjunte un pantallazo se debe incluir la fecha y hora del sistema, ya sea del resultado de la consulta o de la barra de tareas.

The screenshot displays the SQL Server Enterprise Manager interface. The query window on the left contains a SELECT statement: `SELECT [CompanyName], [City], [Region], [Country] FROM [Northwind].[dbo].[Customers]`. The results grid below it shows a list of customers with columns: CompanyName, City, Region, and Country. The messages window on the right shows a DELETE statement: `DELETE FROM [Northwind].[dbo].[Tabla_resumen_12] WHERE [YearOrder] < 1997`. Below the statement, it indicates "(74 rows affected)" and "Completion time: 2020-04-29T19:57:34.1225037+02:00". The system clock at the bottom right shows 18:57 on 29/04/2020.

CompanyName	City	Region	Country
Alfreds Futterkiste	Berlin	NULL	Germany
Ana Trujillo Emparedados y helados	México D.F.	NULL	Mexico
Antonio Moreno Taquería	México D.F.	NULL	Mexico
Around the Horn	London	NULL	UK
Berglunds snabbköp	Luleå	NULL	Sweden
Blauer See Delikatessen	Mannheim	NULL	Germany
Blondesdél's père et fils	Strasbourg	NULL	France
Bólido Comidas preparadas	Madrid	NULL	Spain
Bon app'	Marseille	NULL	France
Bottom-Dollar Markets	Tsawassen	BC	Canada
B's Beverages	London	NULL	UK

Si se genera algún fichero en la solución hay que adjuntarlo.

En caso de usar código, adjuntarlo.

Las soluciones a los ejercicios no son únicas y pueden existir diferentes aproximaciones para el problema. La máxima puntuación se obtiene con una combinación de planteamiento, elegancia de ejecución y resultados.

La práctica se evalúa de 0 a 10.

La práctica se publica el día **28/2/24** y se dispone hasta el **31/3/24** para entregarla.

Debe entregarse un fichero comprimido que incluya:

- Script de creación de la base de datos o export de la misma
- Procesos PDI (kjb, ktr)
- Ficheros PowerBI / Tableau Public o bien la ruta de publicación
- Informe documental de los desarrollos
- Cualquier otra documentación pertinente que se considere adecuada.

Si hay cualquier duda o pregunta enviar correo a [agonzalezd@profesorescol.imf.com](mailto:agonzalezd@profesorescol.imf.com)