

Desafío Data Engineer

Eres el protagonista en la transformación digital de **Innova**, una empresa de retail líder en el mercado mexicano. Con un creciente volumen de transacciones diarias, la empresa busca expandir sus capacidades analíticas para obtener insights más profundos sobre el comportamiento de ventas, tendencias de clientes y eficiencias operativas. Tu misión es diseñar un Data Warehouse que permita a **Innova** no sólo almacenar su creciente volumen de datos de manera eficiente sino también habilitar análisis complejos y reportes en tiempo real para la toma de decisiones estratégicas.

Objetivos del Reto:

1. **Diseñar un Modelo Dimensional:** Crea un modelo estrella o copo de nieve que refleje de manera efectiva las necesidades analíticas de la empresa, incluyendo dimensiones.
2. **Desarrollar un Plan de ETL:** Propón un proceso de ETL detallado que describa cómo extraerás datos de la base de datos transaccional (OLTP), cómo los transformarás (limpieza, normalización, agregación) y cómo los cargarás en el Data Warehouse.
3. **Generación de Archivos de Trabajo:** Basado en el modelo propuesto, genera archivos .csv que simulen las tablas dimensionales y de hechos para demostrar cómo se implementará el modelo.

Tareas y Entregables:

1. **Documento de Diseño del Modelo Dimensional:** Incluye diagramas, descripción de dimensiones y hechos, granularidad de los datos, y justificación de tu diseño.
2. **Script de ETL:** Un pseudo código o script que ejemplifique el proceso de ETL. No necesita ser un código ejecutable, pero sí debe ser claro y detallado.
3. **Archivos .CSV de Muestra:** Crea archivos .csv representativos para cada tabla del modelo dimensional propuesto. Estos deberían incluir columnas y datos ficticios que ilustren cómo se verían las tablas de dimensiones y hechos.

Preguntas de Análisis a Resolver:

- ¿Cuál es el producto más vendido en cada trimestre del año?
- ¿Cuáles son las tendencias de compra de los clientes más leales?
- ¿Cómo varían las ventas según las regiones geográficas durante el año?
- Presentar ejemplos de consultas SQL que responderían preguntas de negocio complejas utilizando el modelo propuesto.

Evaluación:

- Claridad: El modelo debe ser complejo para demostrar habilidades avanzadas, pero claro y bien documentado.
- Escalabilidad: Considera el rendimiento del modelo en términos de tiempos de consulta y escala.
- Innovación: Se valorará positivamente la creatividad y la innovación en tu solución.

Recursos Proporcionados:

- A continuación encontrarás los conjuntos de datos de muestra de la base de datos OLTP para entender el contexto y realizar pruebas.

Clientes: ID, Nombre, Ubicación, Segment

Productos: ID, Nombre, Categoría, Precio

Facturas: ID, Fecha, ClienteID, ProductoID, Cantidad, Total

[Archivos](#)

Este reto no solo es una prueba de tus habilidades técnicas, sino también de tu capacidad para entender el negocio, ser creativo en la solución de problemas y efectivo en la comunicación de tus ideas. Innova busca un ingeniero de datos no solo para construir un sistema, sino para ser un socio estratégico en la toma de decisiones basadas en datos.