

“Del dato bruto a la decisión”

Contexto

La empresa SmartRetail, dedicada al comercio electrónico, busca implementar un sistema de inteligencia de negocios (BI) que le permita transformar datos masivos en decisiones eficientes. Cuentan con datos estructurados (ventas, clientes) y no estructurados (opiniones en redes sociales) almacenados en un sistema distribuido.

Te han contratado como especialista en Big Data para validar los datos, diseñar la arquitectura de almacenamiento, definir indicadores de negocio (KPIs), simular resultados e interpretar decisiones estratégicas.

Tarea 1: Verificación de calidad e integridad de los datos (20 min)

Enunciado: Identifica problemas comunes en la calidad e integridad de datos y propone mecanismos de detección y corrección.

Ejemplo desarrollado:

- Problema: Formatos de fecha no homogéneos.
- Solución: Estandarización con funciones de PySpark.
- Validación: Aplicación de hashes para detectar cambios.

- Identificar al menos dos problemas adicionales.
- Proponer herramientas concretas para validarlos.

Tarea 2: Diseño de almacenamiento escalable (15 min)

Enunciado: Diseña una arquitectura de almacenamiento adecuada para SmartRetail.

Ejemplo desarrollado:

- Estructura: Almacenamiento HDFS con carpetas por categoría de datos.
- Formato: Parquet para eficiencia y compresión.

Recurso de apoyo:

<https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html>

- Describir cómo gestionarías los datos no estructurados.
- Justificar las decisiones técnicas tomadas.

Tarea 3: Modelado de KPIs y visualización (25 min)

Enunciado: Define indicadores clave de negocio y visualizaciones adecuadas.

Ejemplo desarrollado:

- KPI: Ratio de conversión mensual por canal.
- Visualización: Gráfico de líneas en Power BI.

- Proponer tres KPIs adicionales.
- Asociar tipo de visualización y herramienta de BI.

Tarea 4: Simulación de resultados y validación (30 min)

Enunciado: Simula datos y plantea decisiones de negocio en base a los resultados obtenidos.

Ejemplo desarrollado:

- Simulación: Segmento 25-34 años genera 60% de compras online.
- Decisión: Campaña de fidelización digital.
- Validación: Comparación con datos históricos.

Recurso de apoyo: <https://www.mockaroo.com/>

- Inventar un escenario de simulación.
- Proponer decisiones asociadas.
- Indicar cómo validarías el modelo.

Tarea 5: Reflexión crítica (15 min)

Enunciado: Responde con argumentación razonada a las preguntas:

1. ¿Qué consecuencias tiene usar datos sin validar en decisiones de negocio?
2. ¿Cómo cambia el análisis cuando los datos son no estructurados?
3. ¿Qué riesgos éticos existen en el uso de datos personales?
4. ¿Qué parte del proceso de BI has encontrado más desafiante y por qué?