

Actividad: Diseño y Comparación de Arquitecturas de Almacenamiento en Big Data

Objetivo

Desarrollar la capacidad de analizar y diseñar diferentes arquitecturas de almacenamiento en Big Data según los requisitos específicos de una organización.

Descripción

En esta actividad, trabajarás como consultor de arquitectura de datos. Deberás analizar sus necesidades y proponer la arquitectura más adecuada para el caso de uso.

Caso de Estudio

Caso 1: Cadena de Supermercados "FreshMart"

- Necesita analizar patrones de compra históricos
- Maneja datos estructurados de ventas, inventario y clientes
- Requiere generar informes diarios de rendimiento
- Tiene 500 tiendas en todo el país
- Procesa 1 millón de transacciones diarias

Tareas

1. Para el caso de estudio:

- Proponer una arquitectura de almacenamiento específica (Data Warehouse, Data Lake, Data Mesh o Federación de Datos)

Se propone una arquitectura de Data Warehouse

- Justificar la elección basándose en las características y necesidades de la empresa

- **Análisis de Patrones Históricos:** Un Data Warehouse es ideal para almacenar y analizar grandes volúmenes de datos históricos, permitiendo realizar consultas complejas y análisis de tendencias.

- **Datos Estructurados:** Dado que se manejan datos estructurados, un Data Warehouse es adecuado porque está diseñado para trabajar con este tipo de datos.

- **Informes Diarios:** Los Data Warehouses están optimizados para generar informes y dashboards, facilitando la creación de informes diarios de rendimiento.

- **Escalabilidad:** Con 500 tiendas y 1 millón de transacciones diarias, se necesita una solución que pueda escalar. Los Data Warehouses modernos, como Amazon Redshift o Google BigQuery, están diseñados para manejar grandes volúmenes de datos y escalar según sea necesario.

- Describir las herramientas tecnológicas que se utilizarían
- Identificar posibles desafíos y proponer soluciones

2. Crear un diagrama simple de la arquitectura propuesta para cada caso (puede ser un boceto a mano o utilizando herramientas de diagramación) opcional
3. Destaca en una tabla los siguientes items de la arquitectura elegida:
- Ventajas y desventajas
 - Costos relativos
 - Complejidad de implementación
 - Escalabilidad
 - Mantenimiento

Ventajas	Desventajas	Costos Relativos	Complejidad de Implementación	Escalabilidad	Mantenimiento
- Optimizado para consultas y análisis complejos.	- Puede ser costoso en términos de almacenamiento y procesamiento.	Moderado a Alto	Moderada a Alta	Alta escalabilidad con soluciones en la nube como Amazon o Google.	Requiere mantenimiento regular para optimización y seguridad.

Entregables

1. Documento con el análisis y justificación de la arquitectura
2. Diagrama de la arquitectura elegida.--> opcional
3. Tabla comparativa de las arquitecturas propuestas

Recursos Recomendados

- Documentación de herramientas: Hadoop, MongoDB, Cassandra, Amazon S3
- Herramientas de diagramación: draw.io, Lucidchart
- Ejemplos de casos reales de implementación