



CASO PRÁCTICO

Javier Díaz Machado

ÍNDICE

1. Aumento de la Capacidad del Clúster
2. Capas de la Arquitectura Big Data
3. Características ACID
4. Subsistema OLTP
5. Subsistema OLAP
6. Almacén de Datos
7. Estrategia de Procesamiento
8. Modelos Predictivos

1. Aumento de la Capacidad del Clúster

Para aumentar la capacidad del clúster según se reciben nuevos datos, se debe utilizar una arquitectura escalable basada en Hadoop o Spark. Esto permite añadir nodos adicionales al clúster sin interrumpir el servicio.

2. Capas de la Arquitectura Big Data

Las capas mínimas necesarias en la arquitectura Big Data serían 3:

Primera capa:

Esta constituye una de las capas más relevantes de la arquitectura Big Data, dado que se enfoca en el análisis de datos. En esta capa inicial se emplean herramientas analíticas y estadísticas sofisticadas para la visualización de la información.

Segunda capa:

Esta es la capa dedicada a la gestión de datos. En ella se integran tres fundamentos esenciales: la integración, el gobierno y la seguridad. Es fundamental optimizar los datos, organizarlos e implementarlos con la mayor flexibilidad. Además, en esta capa es crucial enfocarse en la preparación de los datos gestionados para alcanzar la máxima calidad. En lo que respecta a la seguridad, es necesario evaluar los riesgos y mantener un control riguroso sobre los datos más sensibles para garantizar su protección de manera efectiva.

Tercera capa:

La tercera y última capa se elabora específicamente para facilitar el procesamiento y almacenamiento de datos. Dado que se manejan datos de diversas estructuras, se opta por no utilizar sistemas SQL, priorizando en su lugar sistemas de procesamiento paralelo masivo. El éxito de las estrategias y gestiones de Big Data en una empresa dependerá de un diseño adecuado de las capas y de su correcta integración.

3. Características ACID

Las transacciones que interactúan con clientes en el día a día necesitarán cumplir con las características ACID para garantizar la integridad de los datos. Esto se puede lograr utilizando bases de datos relacionales como MySQL.

4. Subsistema OLTP

Sí, será necesario un subsistema OLTP para manejar las transacciones diarias con los clientes. Este sistema debe ser capaz de procesar un gran número de transacciones de manera eficiente y segura.

5. Subsistema OLAP

Sí, también será necesario un subsistema OLAP (Online Analytical Processing) para realizar análisis complejos y consultas multidimensionales sobre los datos almacenados. Esto permitirá a la junta directiva acceder a analíticas descriptivas e interactivas.

6. Almacén de Datos

Sí, habrá un almacén de datos para centralizar y organizar los datos de diferentes fuentes. Esto facilitará el acceso y análisis de los datos por parte de los directivos.

7. Estrategia de Procesamiento

Para crear el cuadro de mandos que quiere la junta directiva, se puede emplear una estrategia de procesamiento ETL para consolidar los datos una semana antes de la reunión. Los datos se pueden procesar y cargar en el almacén de datos, donde las herramientas de BI pueden acceder a ellos para generar las visualizaciones.

8. Modelos Predictivos

Sí, será necesario crear modelos predictivos a partir de los datos para ofrecer recomendaciones personalizadas a los clientes. Esto se puede lograr utilizando técnicas de machine learning y herramientas como TensorFlow.