

Prediseño sistema para Big Data empresa Construcciones D8

1. Indicar qué habrá que hacer para ir aumentando la capacidad del cluster según se reciben nuevos datos.

Para manejar el aumento de datos que la empresa producirá anualmente debemos de optar por una estrategia de escalado horizontal aumentando la cantidad de nodos del cluster así conseguiremos mas capacidad de almacenaje a medida que aumente el volumen de datos.

Para realizar dicha infraestructura he escogido un clúster escalable horizontalmente en la nube con los servicios de Amazon AWS el mas común para dichas operaciones de almacenaje es Amazon S3 que ofrece precios variables dependiendo de la frecuencia de acceso a los datos.

S3 Standard:

Precio por GB: \$0,023 por GB al mes.

500 TB = 500.000 GB.

Precio mensual de almacenamiento = 500,000 GB * \$0,023 = \$11.500 al mes.

Precio anual = \$11.500 * 12 = \$138.000 al año

Y los 100TB que se producen anualmente tendrían un coste de:

100TB = 100.000 GB.

Precio por GB: \$0,022 por GB los siguientes 450TB

100.000 * \$0.022 = \$2.200

2. Indicar qué capas de la arquitectura Big Data necesitarán estar presentes como mínimo en el sistema a crear.

1. Capa de ingestión de datos: Esta capa se encargara de conectar y recibir datos de diversas fuentes tanto externas como internas
2. Capa de almacenamiento: Esta capa debe ser escalable para garantizar que los datos se almacenen de forma eficiente
3. Capa de procesamiento: Esta capa se encargara de proveer la infraestructura necesaria para facilitar el análisis y consulta de los datos.
4. Capa de análisis: En esta capa se realizaran los análisis complejos sobre grandes conjuntos de datos.
5. Capa de seguridad: Esta capa se emplea para asegurar los datos aplicando métodos tanto físicos como de software.
6. Capa de visualización: Es la capa que proporciona una interfaz visual al usuario para filtrar, analizar y explorar los datos.

3. Indicar si alguna parte del sistema necesitará cumplir con las características ACID.

El sistema de transacciones diarias con los clientes deberá cumplir con las características ACID ya que se necesita mantener la consistencia y la integridad de los datos de dichas transacciones.

4. Indicar si será necesario un subsistema OLTP.

Sí, será necesario un subsistema OLTP ya que este sistema está orientado a las transacciones, en nuestro caso lo utilizaremos para gestionar las transacciones diarias con los clientes.

5. Indicar si será necesario un subsistema OLAP.

Sí, será necesario un subsistema OLAP, ya que este sistema está orientado a procesar consultas de tipo analítico y filtrado de datos, en nuestro caso lo utilizaremos para que la junta directiva pueda ver analíticas descriptivas, filtrado de datos a través de graficas o cuadro de mando.

6. Indicar si habrá un almacén de datos.

Sí, se necesitará un almacén de datos. Este almacenará datos tanto actuales como históricos con la finalidad de ofrecerles a los clientes ciertas ofertas en función de su comportamiento pasado, además de permitir a la junta directiva realizar análisis y reportes sobre dichos datos a través del cuadro de mando.

7. Indicar qué estrategia de procesamiento habrá que emplear para poder crear el cuadro de mandos que quiere la junta directiva.

Emplearemos una estrategia combinada utilizando un procesamiento por lotes y un procesamiento en streaming.

El procesamiento por lotes lo utilizaremos para los datos que proceden de fuentes externas y que se actualizan anualmente ya que no requieren de una respuesta inmediata, estos datos se procesarán en su totalidad para posteriormente realizar informes y análisis sobre estos. Y el procesamiento por streaming lo utilizaremos para las fuentes internas y para las transacciones diarias de los cliente ya que este tipo de procesamiento maneja los datos en tiempo real procesandolos a medida que estos llegan.

8. Indicar si será necesario crear modelos predictivos a partir de los datos.

Sí serán necesarios crear modelos predictivos para personalizar las ofertas a nuestros clientes, ya que estos modelos pueden aprovechar el comportamiento histórico de los clientes para predecir sus preferencias futuras.