

Caso práctico

La empresa Construcciones D8 se ha puesto en contacto con la empresa consultora en la que trabajas para que le realices un prediseño de lo que sería un sistema Big Data para resolver las siguientes necesidades.

- Hay distintas fuentes externas a su empresa que producen datos interesantes para ellos y les interesaría poder conectarse a ellas para obtenerlos.
- Esas fuentes tienen conjuntos de datos estáticos o que se actualizan anualmente.
- Además hay fuentes internas de la propia empresa que generan datos de forma continua y hay que irlos obteniendo sobre la marcha.
- La cantidad de datos actualmente es de aproximadamente 500TB, y calculan que se producen otros 100TB nuevos cada año.
- Quieren poder mantener almacenados todos esos datos de modo no se pierdan y además accesibles en todo momento.
- Se realizan transacciones debido a la interacción con clientes en el día a día.
- La junta directiva se reúne una vez al mes y quiere poder acceder a un cuadro de mandos para ver analíticas descriptivas que empleen todos los datos que estuviesen disponibles una semana antes de reunirse. Tales analíticas deben ser interactivas, siendo los directivos capaces de realizar filtrados de información de modo que las gráficas mostradas se actualicen según la información seleccionada.
- Quieren poder decidir a qué clientes ofrecerles ciertas ofertas en función de lo que se sabe de su comportamiento pasado.

Apartado 1: Prediseño un sistema para Big Data

Crea un documento en el que explicas cómo sería el sistema a emplear para resolver las necesidades Big Data del supuesto práctico. Deberás:

- Indicar qué habrá que hacer para ir aumentando la capacidad del cluster según se reciben nuevos datos.
- Indicar qué capas de la arquitectura Big Data necesitarán estar presentes como mínimo en el sistema a crear.
- Indicar si alguna parte del sistema necesitará cumplir con las características ACID.(Atomicidad, Consistencia, Aislamiento, Durabilidad)
- Indicar si será necesario un subsistema OLTP.
- Indicar si será necesario un subsistema OLAP.
- Indicar si habrá un almacén de datos.
- Indicar qué estrategia de procesamiento habrá que emplear para poder crear el cuadro de mandos que quiere la junta directiva.
- Indicar si será necesario crear modelos predictivos a partir de los datos.

Que se evaluará de la tarea

- Indica correctamente qué habrá que hacer para ir aumentando la capacidad del cluster según se reciben nuevos datos.
- Indica correctamente qué capas de la arquitectura Big Data necesitarán estar presentes como mínimo en el sistema a crear.
- Indica correctamente si alguna parte del sistema necesitará cumplir con las características ACID.
- Indica correctamente si será necesario un subsistema OLTP.
- Indica correctamente si será necesario un subsistema OLAP.
- Indica correctamente si habrá un almacén de datos.
- Indica correctamente qué estrategia de procesamiento habrá que emplear para poder crear el cuadro de mandos que quiere la junta directiva.
- Indica correctamente si será necesario crear modelos predictivos a partir de los datos