

Néstor Batista Díaz

21 de octubre del 2023

Prediseño un sistema para Big Data

Para diseñar un sistema de Big Data que cumpla con las necesidades de la empresa, se debe considerar una arquitectura escalable y robusta.

- CLUSTER
 - Para manejar el crecimiento de datos, se necesita una arquitectura escalable.
Puedes utilizar plataformas como Hadoop, que permiten añadir nodos al cluster para aumentar la capacidad de almacenamiento y procesamiento a medida que llegan nuevos datos.
- CAPAS DE LA ARQUITECTURA BIG DATA:
 - **Ingestión de datos:** Recopilar datos de fuentes externas e internas, con herramientas como Apache NiFi o Kafka.
 - **Almacenamiento:** Utilizar un sistema de almacenamiento distribuido como HDFS o servicios de almacenamiento en la nube como AWS S3.
 - **Procesamiento:** Implementar un motor de procesamiento, como Apache Spark o Hadoop MapReduce, para realizar análisis de datos en lotes y en tiempo real.
 - **Bases de datos NoSQL:** Usar bases de datos NoSQL como Cassandra para gestionar datos no estructurados y semiestructurados.

- **Data Warehouse (almacén de datos):** Puedes implementar un almacén de datos para consultas OLAP utilizando herramientas como Amazon Redshift o Google BigQuery.
- **CUMPLIMIENTO CON ACID:**
 - La capa de almacenamiento debe ser ACID-compatible para garantizar la integridad de los datos, especialmente para transacciones críticas. Esto puede ser necesario en el sistema OLTP.
- **SUBSISTEMA OLTP (PROCESAMIENTO DE TRANSACCIONES EN LÍNEA):**
 - Necesitarás un subsistema OLTP para gestionar transacciones en tiempo real con clientes y fuentes internas. Esto garantiza la disponibilidad y la coherencia de los datos. Puedes usar bases de datos relacionales o NoSQL para esta capa.
- **SUBSISTEMA OLAP (PROCESAMIENTO ANALÍTICO EN LÍNEA):**
 - Para las analíticas descriptivas, implementa un subsistema OLAP que permite realizar consultas complejas sobre los datos. Puedes utilizar un almacén de datos para esta capa.
- **ALMACÉN DE DATOS:**
 - Un almacén de datos se utiliza para almacenar datos agregados y estructurados que se pueden consultar eficientemente para fines analíticos. Esto es esencial para generar informes y cuadros de mando interactivos.
 -

- PROCESAMIENTO PARA CUADRO DE MANDOS:
 - Para el cuadro de mandos interactivo de la junta directiva, puedes utilizar herramientas de visualización de datos como Power BI o una aplicación web personalizada. La estrategia de procesamiento debe incluir consultas en tiempo real a los datos y actualizaciones programadas para asegurarse de que los datos estén disponibles una semana antes de la reunión.
- MODELOS PREDICTIVOS:
 - Si deseas utilizar modelos predictivos, puedes implementar un sistema de aprendizaje automático utilizando bibliotecas como TensorFlow o scikit-learn. Esto te permitirá realizar análisis predictivos basados en datos históricos y actuales.

En resumen, el sistema Big Data para Construcciones D8 debe ser escalable, con capas para la ingesta, almacenamiento, procesamiento, OLTP, OLAP, un almacén de datos, un sistema de visualización interactiva y la capacidad de implementar modelos predictivos. Además, debe cumplir con las necesidades de ACID en áreas críticas, como transacciones con clientes.

BIBLIOGRAFÍA

- https://www.sas.com/es_es/insights/big-data/hadoop.html#:~:text=Hadoop%20es%20una%20estructura%20de,o%20trabajos%20concurrentes%20virtualmente%20ilimitados.
- <https://keepcoding.io/blog/que-es-arquitectura-big-data/>
- <https://nifi.apache.org>
- <https://kafka.apache.org>
- <https://www.ibm.com/es-es/topics/hdfs>
- https://aws.amazon.com/es/pm/serv-s3/?trk=894c59b4-848e-44d2-805f-f02c7ae38070&sc_channel=ps&ef_id=CjwKCAjw7c2pBhAZEiwA88pOF9k0Qv6mv1brcBWeBKKr8DiBm9MrAHqrdvUfPi9zSpTYd1TKiuRIIhoChBgQAvD_BwE:G:s&s_kwcid=AL!4422!3!664475619094!e!!g!!aws%20s3!20267313492!150259342796
- https://cassandra.apache.org/_/index.html
- https://aws.amazon.com/es/pm/redshift/?trk=8e5f044b-6475-484a-b7eb-8923cfdd5362&sc_channel=ps&ef_id=CjwKCAjw7c2pBhAZEiwA88pOF3VvyjYxclnZR4JLGNWhXeHZND5trH_6QNBxBoI2J66CRTBabDssMhoCpyEQAvD_BwE:G:s&s_kwcid=AL!4422!3!524552148442!e!!g!!amazon%20redshift!13287220920!125409695480
- https://cloud.google.com/bigquery?utm_source=google&utm_medium=cpc&utm_campaign=na-US-all-es-dr-bkws-all-all-trial-e-dr-1605212&utm_content=text-ad-none-any-DEV_c-CRE_665735450807-ADGP_Hybrid+%7C+BKWS+-+MIX+%7C+Txt_BigQuery-KWID_43700077212833525-kwd-63326440124&utm_term=KW_google+bigquery-ST_google+bigquery&gclid=CjwKCAjw7c2pBhAZEiwA88pOF4pOgpXKcx6oxUktZjekoL7Y7qdn8Ph-cDHi32HOWUkrZrqPmFieshoC4ZkQAvD_BwE&gclsrc=aw.ds&hl=es_419
- <https://powerbi.microsoft.com/es-es/>