Esteban Guzmán Ramírez
2015095861
Bases de datos 2 - Resumen 1

Introducción

La introducción habla de la importancia de los datos para una empresa, esta debe guardar los datos relevantes, darle acceso a las personas necesarias, poder analizar los datos y convertirlos en información relevante para la empresa.

La solución tradicional serían los data-warehouse hechos por cada empresa pero estos son caros, difíciles de hacer y mantener. Entonces se nombran las ventajas de un data-warehouse en la nube hecho por Amazon como una arquitectura moderna, tecnologías a escoger, consejos para mover los datos, planos para crear un Amazon Redshift y rasgos diferenciadores.

Introducción a Amazon Redshift

De manera redundante se habla de que antes era dificil tener data-werehouse hechos en casa, por lo que la solución de Amazon se creo para ser demasiado conveniente bajando costos en precio y esfuerzo.

Arquitectura moderna de análisis y almacenamiento de datos

- **Data-Warehouses**: optimizados para operaciones de escritura por lotes y lectura de grandes volúmenes de datos. Suelen usar la desnormalización de los datos con esquemas Star y Snowflake
- **Bases OLTP**: optimizados para operaciones de escritura continua y grandes volúmenes de pequeñas operaciones de lectura

Opciones en tegnologías data-warehouses

Bases orientadas a filas

Oracle Database Server, Microsoft SQL Server, MySQL y PostgreSQL son ejemplos. Suelen almacenar filas completas en un bloque físico. Más adecuadas para OLTP que para crear analisis de datos.

Bases orientadas a columnas

Mejor opción que las bases orientadas a filas para el almacenamiento de datos. En este tipo cada columna se guarda en bloques físicos a diferencia de las orientadas por filas. Muy eficientes en las consultas de lectura.

Arquitecturas de procesamiento masivo en paralelo

Estas arquitecturas son creadas para usas todos los recursos en cluster y aumentar el rendimiento a la hora de procesar los datos. Ejemplos son Amazon Redshift, Hadoop y Spark.

Amazon Redshift Deep Dive

Tecnología MPP de columna, alto desempeño, almacenamiento rentable y basado en ANSI SQL para ejecutar consultas existente con pocas modificaciones. Integración con Data-Lake, compatible con gran variedad de formatos y S3.

Operaciones

Patrones ideales

Antipatrones

- OLTP
- Datos desestructurados
- Datos BLOB