



Instituto Tecnológico de Costa Rica

Campus Tecnológico Central Cartago

Escuela de Ingeniería en Computación

Bases de Datos II

IC-4302

Resumen 2

Fecha de entrega: 07/03/2023

I-Semestre 2023

Profesor:

Nereo Campos Araya.

Elaborado por:

Diana Sanabria Calvo 2021436548.



-Introducing Amazon Redshift-

Es un servicio de almacenamiento de datos escalable, de alta velocidad y basado en la nube proporcionado por Amazon Web Services (AWS). Un almacén de datos llamado Redshift emplea una arquitectura en columnas para mejorar el rendimiento y la eficacia al manejar conjuntos de datos considerables. Una de las características clave de Redshift es su integración con otras herramientas y servicios de AWS, como Amazon S3, AWS Glue y Amazon EMR, así como su capacidad para procesar consultas complejas en varios nodos.

-Modern Analytics and data warehousing architecture-

Para el análisis y almacenamiento de datos en la nube de Amazon Web Services (AWS), ofrece una arquitectura contemporánea. A medida que leemos, queda claro cómo la arquitectura en la nube de AWS permite a los usuarios crear un almacén de datos escalable y adaptable mediante el uso de herramientas como Amazon S3 para el almacenamiento de datos y Amazon Redshift para el procesamiento y análisis de datos. También tiene una arquitectura de datos contemporánea que enfatiza la división de las capas de procesamiento, almacenamiento y visualización de datos.

-Data warehouse technology options-

Las filas completas se almacenan en un solo bloque físico en bases de datos orientadas a filas, que también usan índices secundarios para aumentar el rendimiento de lectura. Las bases de datos orientadas a filas están limitadas por el hardware disponible y los almacenes de datos solo pueden resolver este problema parcialmente. En un almacén de datos orientado a filas, cada consulta debe leer todas las columnas de todas las filas de los bloques que satisfacen el predicado de la consulta, lo que ralentiza significativamente el rendimiento. Las bases de datos orientadas a columnas son más eficientes en E/S para consultas de solo lectura y una mejor opción para almacenes de datos porque almacenan cada columna en su propio conjunto de bloques físicos.

-Amazon redshift deep dive-

Se basa en una arquitectura de procesamiento paralelo masivo (MPP), Amazon Redshift permite la distribución y el procesamiento paralelo de grandes conjuntos de datos en numerosos nodos informáticos. Aumenta la velocidad y la eficacia del almacenamiento y la consulta de datos mediante el uso de columnas en lugar de filas para almacenar datos. Con la capacidad de agregar o eliminar nodos rápida y fácilmente, el servicio ofrece opciones de escalabilidad, desde pequeñas cargas de trabajo hasta petabytes de almacenamiento.

-Operations-

El apartado Operations del libro se centra en algunas prácticas para administrar y mantener un data warehouse en AWS. Para proteger los datos y mantener la continuidad del negocio, implementar un plan de copia de seguridad y recuperación ante desastres. Para encontrar y resolver problemas de rendimiento y disponibilidad, use herramientas de monitoreo y alerta. Configurar la seguridad y el acceso para guardar los datos confidenciales y satisfacer los requisitos de cumplimiento. Mejorar el rendimiento de las consultas mediante el uso de técnicas para la optimización de consultas y el particionamiento de datos.

