

Ejercicio 9: Amazon DynamoDB (solución de base de datos NoSQL)

Pregunta:

- Describe la estructura de tablas, elementos y atributos en Amazon DynamoDB.

Tablas: Son como contenedores principales donde se almacenan los datos. Cada tabla en DynamoDB está formada por ítems, que son las unidades de datos individuales.

Elementos (ítems): Cada ítem es una instancia individual de datos dentro de una tabla. Se identifica de manera única dentro de la tabla mediante una clave primaria.

Atributos: Son los datos específicos que conforman cada ítem. Cada atributo tiene un nombre y un valor asociado.

Tabla: Usuarios #contenedor

Elemento (Ítem): Usuario1 #un usuario específico dentro de la tabla

- **Atributos:**
 - **Clave primaria:** IdUsuario
 - **Otros atributos:**
 - Nombre (tipo: cadena): "Alexander"
 - Edad (tipo: número): 19
 - CorreoElectronico (tipo: cadena): "yojan@upch.com"
 - FechaRegistro (tipo: fecha): "2024-06-19"
- Explica cómo se puede provisionar la **capacidad en DynamoDB** y por qué es importante para gestionar el **rendimiento y los costos**.

Capacidad de lectura y escritura: En DynamoDB, se necesita especificar cuántas operaciones de lectura y escritura por segundo se espera que la tabla pueda manejar. Esta capacidad se puede ajustar para satisfacer las demandas de la aplicación.

- **Rendimiento:** Si la capacidad no es suficiente, las operaciones pueden enfrentar retardos o errores.
- **Costos:** DynamoDB cobra en función de la capacidad provisionada. Ajustar correctamente esta capacidad es crucial para optimizar los costos. Se podría escalar hacia arriba o hacia abajo según las necesidades de la aplicación, lo que te permite pagar solo por lo que realmente utiliza

Ejercicio 10: Uso de Amazon Redshift y almacenamiento de datos

Pregunta:

- ☐ Define qué es Amazon Redshift y para qué tipo de cargas de trabajo es más adecuado.

¿Qué es Amazon Redshift?

Amazon Redshift es una solución de almacenamiento de datos en la nube diseñada para procesar y analizar grandes volúmenes de datos. Se basa en una arquitectura de almacenamiento en columnas que optimiza el procesamiento analítico en línea (OLAP). Redshift permite extraer datos de otras bases de datos relacionales y facilita la realización de consultas y análisis complejos en diferentes conjuntos de datos.

Tipos de Cargas de Trabajo Adecuadas para Amazon Redshift

Análisis de Datos Masivos (Big Data Analytics):

- Permite analizar grandes volúmenes de datos de diversas fuentes, como registros de transacciones, información de clientes e inventarios.
- Soporta consultas complejas y agregaciones de datos para obtener análisis valiosos

Consultas OLAP (Online Analytical Processing):

- Optimizado para operaciones OLAP que implican escanear muchas filas de datos para una o varias columnas.
- Eficiente en la ejecución de consultas analíticas que requieren el procesamiento de grandes conjuntos de datos.

Integración con Servicios de AWS:

- Compatible con servicios de almacenamiento como Amazon S3 a través de Redshift Spectrum, permitiendo consultar datos directamente desde S3.
- Integración con otras herramientas de AWS para un ecosistema de datos completo.

- ☐ Explica los conceptos de OLAP y cómo Redshift facilita estas operaciones.

¿Qué es OLAP?

OLAP (On-Line Analytical Processing) es una tecnología de bases de datos diseñada para obtener rápidamente datos analíticos multidimensionales. Se utiliza para analizar el rendimiento empresarial, proporcionar información, revelar tendencias y extraer resúmenes para la toma de decisiones. OLAP es un componente de las soluciones de almacén de datos y, por lo general, puede dividirse en tres partes: almacenamiento de datos, consulta y análisis, e informes.

¿Cómo Redshift facilita estas operaciones?

- **Procesamiento Paralelo Masivo:** Redshift distribuye la carga de trabajo entre múltiples nodos de cómputo, lo que permite ejecutar consultas en paralelo y mejorar significativamente el tiempo de respuesta.
- **Redshift Spectrum:** Permite ejecutar consultas directamente en datos almacenados en Amazon S3 sin necesidad de cargarlos en Redshift, facilitando el análisis de grandes conjuntos de datos distribuidos y heterogéneos.
- **Caching y Almacenamiento de Resultados:** Redshift almacena en caché los resultados de consultas frecuentes, lo que reduce el tiempo de respuesta para consultas repetitivas.