

Data Warehousing on AWS

Código: AWS-021

Propuesta de Valor: AWS (AMAZON WEB SERVICES)

Duración: 24 Horas



En este curso, aprenderá conceptos, estrategias y prácticas recomendadas para diseñar una solución de almacenamiento de datos basada en la nube con Amazon Redshift, el almacén de datos que permite escalar a petabytes en AWS. Además, mostraremos de qué manera recopilar, almacenar y preparar datos para su almacenamiento mediante otros servicios de AWS, como Amazon DynamoDB, Amazon EMR, Amazon Kinesis Firehose y Amazon Simple Storage Service (Amazon S3). También analizaremos cómo utilizar herramientas de inteligencia de negocio (BI) para realizar análisis en sus datos.

AUDIENCIA

- Arquitectos de datos
- Administradores de bases de datos
- Desarrolladores de bases de datos
- Analistas de datos
- Científicos de datos

PRE REQUISITOS

- Estar familiarizado con conceptos de diseño de bases de datos y bases de datos relacionales

OBJETIVOS

- Evaluar la relación entre Amazon Redshift y otros sistemas de big data
- Evaluar casos de uso de cargas de trabajo de almacenamiento de datos y revisar implementaciones reales de servicios de análisis y datos de AWS como parte de una solución de almacenamiento de datos
- Elegir un tamaño y tipo de nodo de Amazon Redshift adecuado para sus necesidades relacionadas con datos.
- Comprender cuáles son las características de seguridad adecuadas para Amazon Redshift, como el cifrado, los permisos de IAM y los permisos para bases de datos



CERTIFICACIÓN DISPONIBLE

- Certificado oficial de **AWS**.



CONTENIDO

1. INTRODUCCIÓN AL ALMACENAMIENTO DE DATOS

- 1.1. BASES DE DATOS RELACIONALES
- 1.2. CONCEPTOS DE ALMACENAMIENTO DE DATOS
- 1.3. LA INTERSECCIÓN DE ALMACENAMIENTO DE DATOS Y BIG DATA
- 1.4. DESCRIPCIÓN GENERAL DE LA ADMINISTRACIÓN DE DATOS EN AWS
- 1.5. LABORATORIO PRÁCTICO 1: INTRODUCCIÓN A AMAZON REDSHIFT

2. INTRODUCCIÓN A AMAZON REDSHIFT

- 2.1. RESUMEN CONCEPTUAL
- 2.2. CASOS DE USO DEL MUNDO REAL
- 2.3. LABORATORIO PRÁCTICO 2: LANZAMIENTO DE UN CLÚSTER DE AMAZON REDSHIFT

3. LANZAMIENTO DE CLÚSTERES

- 3.1. CONSTRUCCIÓN DEL CLÚSTER.
- 3.2. CONEXIÓN AL CLÚSTER
- 3.3. CONTROL DE ACCESO
- 3.4. SEGURIDAD DE LA BASE DE DATOS
- 3.5. CARGAR DATOS
- 3.6. LABORATORIO PRÁCTICO 3: OPTIMIZACIÓN DE ESQUEMAS DE BASES DE DATOS

4. DISEÑO DEL ESQUEMA DE LA BASE DE DATOS

- 4.1. ESQUEMAS Y TIPOS DE DATOS
- 4.2. COMPRESIÓN COLUMNAR
- 4.3. ESTILOS DE DISTRIBUCIÓN DE DATOS
- 4.4. MÉTODOS DE CLASIFICACIÓN DE DATOS.

5. IDENTIFICACIÓN DE FUENTES DE DATOS

- 5.1. DESCRIPCIÓN GENERAL DE LAS FUENTES DE DATOS
- 5.2. AMAZON S3
- 5.3. AMAZON DYNAMODB
- 5.4. AMAZON EMR
- 5.5. AMAZON KINESIS DATA FIREHOSE
- 5.6. CARGADOR DE BASE DE DATOS AWS LAMBDA PARA AMAZON REDSHIFT
- 5.7. LABORATORIO PRÁCTICO 4: CARGA DE DATOS EN TIEMPO REAL EN UNA BASE DE DATOS DE AMAZON REDSHIFT

6. CARGA DE DATOS

- 6.1. PREPARACIÓN DE DATOS
- 6.2. CARGANDO DATOS USANDO COPY
- 6.3. MANTENIMIENTO DE MESAS.
- 6.4. OPERACIONES DE ESCRITURA SIMULTÁNEAS
- 6.5. SOLUCIÓN DE PROBLEMAS DE CARGA
- 6.6. LABORATORIO PRÁCTICO 5: CARGA DE DATOS CON EL COMANDO COPY

7. REDACCIÓN DE CONSULTAS Y OPTIMIZACIÓN DEL RENDIMIENTO

- 7.1. SQL DE AMAZON REDSHIFT
- 7.2. FUNCIONES DEFINIDAS POR EL USUARIO (UDF)
- 7.3. FACTORES QUE AFECTAN EL RENDIMIENTO DE LAS CONSULTAS
- 7.4. EL COMANDO EXPLAIN Y LOS PLANES DE CONSULTA
- 7.5. GESTIÓN DE LA CARGA DE TRABAJO (WLM)
- 7.6. LABORATORIO PRÁCTICO 6: CONFIGURACIÓN DE LA GESTIÓN DE CARGA DE TRABAJO

8. ESPECTRO DE AMAZON REDSHIFT

- 8.1. ESPECTRO DE DESPLAZAMIENTO AL ROJO DE AMAZON
- 8.2. CONFIGURACIÓN DE DATOS PARA AMAZON REDSHIFT SPECTRUM
- 8.3. CONSULTAS DE AMAZON REDSHIFT SPECTRUM
- 8.4. LABORATORIO PRÁCTICO 7: USO DE AMAZON REDSHIFT SPECTRUM

9. MANTENIMIENTO DE CLÚSTERES

- 9.1. REGISTRO DE AUDITORÍA
- 9.2. SUPERVISIÓN DEL RENDIMIENTO
- 9.3. EVENTOS Y NOTIFICACIONES
- 9.4. LABORATORIO 8: CLÚSTERES DE AUDITORÍA Y MONITOREO
- 9.5. CAMBIAR EL TAMAÑO DE LOS CLÚSTERES

BENEFICIOS

- Al finalizar el curso, los participantes podrán realizar la administración de datos con los servicios de AWS