

Building Batch Data Analytics Solutions on AWS

Código: AWS-018

Propuesta de Valor: AWS (AMAZON WEB SERVICES)

Duración: 8 Horas



En este curso, aprenderá a crear soluciones de análisis de datos por lotes con Amazon EMR, una solución de nivel empresarial. Servicio gestionado Apache Spark y Apache Hadoop. Aprenderá cómo Amazon EMR se integra con proyectos de código abierto como Apache Hive, Hue y HBase, y con servicios de AWS como AWS Glue y Formación del lago AWS. El curso aborda la recopilación, ingesta, catalogación, almacenamiento y procesamiento de datos componentes en el contexto de Spark y Hadoop. Aprenderá a usar EMR Notebooks para admitir ambos Cargas de trabajo de análisis y aprendizaje automático. También aprenderá a aplicar la seguridad, el rendimiento y el costo. mejores prácticas de gestión para el funcionamiento de Amazon EMR.

Más información: [AQUÍ](#)

Reserve su plaza: [AQUÍ](#)



AUDIENCIA

- Ingenieros de plataformas de datos
- Arquitectos y operadores que construyen y administran canalizaciones de análisis de datos



PRE REQUISITOS

- Los estudiantes con una experiencia mínima de un año en la gestión de marcos de datos de código abierto como Apache Spark o Apache Hadoop se beneficiarán de este curso.
- Sugerimos el curso de AWS Hadoop Fundamentals para aquellos que necesitan un repaso en Apache Hadoop
- Recomendamos que los asistentes a este curso tengan completo AWS Technical Essentials o Architecting on AWS
- Recomendamos que los asistentes a este curso tengan completo Building Data Lakes on AWS o Getting Started with AWS Glue



OBJETIVOS

- Diseñar e implementar una solución de análisis de datos por lotes
- Identificar y aplicar técnicas adecuadas, incluida la compresión, para optimizar el almacenamiento de datos
- Seleccionar e implementar opciones apropiadas para ingerir, transformar y almacenar datos
- Elija la instancia y los tipos de nodos, los clústeres, el escalado automático y la topología de red adecuados para una caso de uso comercial particular

necesario para obtener información empresarial procesable

- Datos seguros en reposo y en tránsito
- Supervise las cargas de trabajo de análisis para identificar y solucionar problemas
- Aplicar las mejores prácticas de gestión de costos
- Comparar las características y los beneficios de los almacenes de datos, los lagos de datos y las arquitecturas de datos modernas



CERTIFICACIÓN DISPONIBLE

- Certificado oficial de **AWS**.
- Este curso lo prepara para el examen: **AWS Certified Data Analytics - Specialty (DAS-C01) (Retiro en abril 2024)**



CONTENIDO

1. INTRODUCCIÓN A AMAZON EMR

- 1.1. USO DE AMAZON EMR EN SOLUCIONES DE ANÁLISIS
- 1.2. ARQUITECTURA DE CLÚSTER DE AMAZON EMR
- 1.3. LANZAMIENTO DE UN CLÚSTER DE AMAZON EMR
- 1.4. ESTRATEGIAS DE GESTIÓN DE COSTOS

2. CANALIZACIÓN DE ANÁLISIS DE DATOS CON AMAZON EMR: INGESTA Y ALMACENAMIENTO

- 2.1. OPTIMIZACIÓN DEL ALMACENAMIENTO CON AMAZON EMR
- 2.2. TÉCNICAS DE INGESTIÓN DE DATOS

3. ANÁLISIS DE DATOS POR LOTES DE ALTO RENDIMIENTO CON APACHE SPARK EN AMAZON EMR

- 3.1. CASOS DE USO DE APACHE SPARK EN AMAZON EMR
- 3.2. POR QUÉ APACHE SPARK EN AMAZON EMR
- 3.3. CONCEPTOS DE CHISPA
- 3.4. CONÉCTESE A UN CLÚSTER DE EMR Y EJECUTE COMANDOS DE SCALA USANDO EL CÁSCARA DE CHISPA
- 3.5. TRANSFORMACIÓN, PROCESAMIENTO Y ANÁLISIS
- 3.6. USO DE PORTÁILES CON AMAZON EMR
- 3.7. PRÁCTICA DE LABORATORIO 1: ANÁLISIS DE DATOS DE BAJA LATENCIA CON APACHE SPARK EN AMAZON EMR

4. PROCESAMIENTO Y ANÁLISIS DE DATOS POR LOTES CON AMAZON EMR Y APACHE HIVE

- 4.1. USO DE AMAZON EMR CON HIVE PARA PROCESAR DATOS POR LOTES
- 4.2. TRANSFORMACIÓN, PROCESAMIENTO Y ANÁLISIS
- 4.3. PRÁCTICA DE LABORATORIO 2: PROCESAMIENTO DE DATOS POR LOTES CON AMAZON EMR CON HIVE
- 4.4. INTRODUCCIÓN A APACHE HBASE EN AMAZON EMR

5. PROCESAMIENTO DE DATOS SIN SERVIDOR

- 5.1. PROCESAMIENTO, TRANSFORMACIÓN Y ANÁLISIS DE DATOS SIN SERVIDOR

5.3. PRÁCTICA DE LABORATORIO 3: ORQUESTA EL PROCESAMIENTO DE DATOS EN SPARK CON AWS STEP FUNCTIONS

6. SEGURIDAD Y MONITOREO DE CLÚSTERES DE AMAZON EMR

6.1. PROTECCIÓN DE CLÚSTERES DE EMR

6.2. DEMOSTRACIÓN INTERACTIVA 3: CIFRADO DEL LADO DEL CLIENTE CON EMRFS

6.3. SUPERVISIÓN Y SOLUCIÓN DE PROBLEMAS DE CLÚSTERES DE AMAZON EMR

6.4. REVISIÓN DEL HISTORIAL DEL CLÚSTER DE APACHE SPARK

7. DISEÑO DE SOLUCIONES DE ANÁLISIS DE DATOS POR LOTES

7.1. CASOS DE USO DE ANÁLISIS DE DATOS POR LOTES

7.2. DISEÑO DE UN FLUJO DE TRABAJO DE ANÁLISIS DE DATOS POR LOTES



BENEFICIOS

- Al finalizar el curso, los participantes podrán comparar las características y los beneficios de los almacenes de datos, los lagos de datos y las arquitecturas de datos modernas