

Big Data en AWS

Código: AWS-007

Propuesta de Valor: AWS (AMAZON WEB SERVICES)

Duración: 24 Horas



En este curso, aprenderá sobre soluciones de big data basadas en la nube, como Amazon EMR, Amazon Redshift, Amazon Kinesis y las demás plataformas de big data de AWS. Descubrirá cómo utilizar Amazon EMR para procesar datos con el amplio ecosistema de herramientas de Hadoop, como Hive y Hue. También aprenderá a crear entornos de big data, a trabajar con Amazon DynamoDB, Amazon Redshift, Amazon QuickSight, Amazon Athena y Amazon Kinesis, así como a utilizar las prácticas recomendadas para diseñar entornos de big data con fines de seguridad y rentabilidad.



AUDIENCIA

- Arquitectos de soluciones
- Administradores de operaciones de sistemas
- Científicos de datos
- Analistas de datos



PRE REQUISITOS

- Conocimientos básicos de tecnologías de big data, incluidos Apache Hadoop, HDFS y consultas SQL/NoSQL.
- Capacitación digital gratuita en Data Analytics Fundamentals completa o una experiencia equivalente.
- Conocimiento práctico sobre los principales servicios de AWS y la implementación de la nube pública.
- Capacitación en AWS Technical Essentials completa o una experiencia equivalente.
- Conocimientos básicos sobre el almacenamiento de datos, los sistemas de bases de datos relacionales y el diseño de base de datos.



OBJETIVOS

- Utilice Apache Hadoop con Amazon EMR.
- Lanzar y configurar un clúster de Amazon EMR.
- Utilice marcos de programación comunes para Amazon EMR, incluidos Hive, Pig y Streaming..
- Use Hue para mejorar la facilidad de uso de Amazon EMR.

- Use análisis en memoria con Spark en Amazon EMR.



CERTIFICACIÓN DISPONIBLE

- Certificado oficial de **AWS**.



CONTENIDO

1. DESCRIPCIÓN GENERAL DE BIG DATA

- 1.1. ¿QUÉ ES BIG DATA?
- 1.2. LA GRAN TUBERÍA DE DATOS
- 1.3. PRINCIPIOS ARQUITECTÓNICOS DE BIG DATA

2. INGESTIÓN Y TRANSFERENCIA DE BIG DATA

- 2.1. DESCRIPCIÓN GENERAL: INGESTIÓN DE DATOS
- 2.2. TRANSFERENCIA DE DATOS

3. BIG DATA STREAMING Y AMAZON KINESIS

- 3.1. PROCESAMIENTO DE FLUJO DE BIG DATA
- 3.2. AMAZON KINESIS
- 3.3. AMAZON KINESIS DATA FIREHOSE
- 3.4. TRANSMISIONES DE VIDEO DE AMAZON KINESIS
- 3.5. ANÁLISIS DE DATOS DE AMAZON KINESIS
- 3.6. LABORATORIO PRÁCTICO 1: TRANSMISIÓN Y PROCESAMIENTO DE REGISTROS DEL SERVIDOR APACHE CON AMAZON KINESIS

4. SOLUCIONES DE ALMACENAMIENTO DE BIG DATA

- 4.1. OPCIONES DE ALMACENAMIENTO DE DATOS DE AWS
- 4.2. CONCEPTOS DE SOLUCIONES DE ALMACENAMIENTO
- 4.3. FACTORES EN LA ELECCIÓN DE UN ALMACÉN DE DATOS

5. PROCESAMIENTO Y ANÁLISIS DE BIG DATA

- 5.1. PROCESAMIENTO Y ANÁLISIS DE BIG DATA
- 5.2. AMAZON ATHENA
- 5.3. LABORATORIO PRÁCTICO 2: USO DE AMAZON ATHENA PARA ANALIZAR DATOS DE

REGISTRO

6. APACHE HADOOP Y AMAZON EMR

- 6.1. INTRODUCCIÓN A AMAZON EMR Y APACHE HADOOP
- 6.2. MEJORES PRÁCTICAS PARA LA INGESTA DE DATOS
- 6.3. EMR AMAZONICO
- 6.4. ARQUITECTURA EMR DE AMAZON
- 6.5. LABORATORIO PRÁCTICO 3: ALMACENAMIENTO Y CONSULTA DE DATOS EN AMAZON DYNAMODB

7. SO DE AMAZON EMR

- 7.1. DESARROLLO Y EJECUCIÓN DE SU APLICACIÓN
- 7.2. LANZAMIENTO DE SU CLÚSTER
- 7.3. MANEJO DE RESULTADOS DE SUS TRABAJOS COMPLETOS

8. MARCOS DE PROGRAMACIÓN HADOOP

- 8.1. MARCOS HADOOP
- 8.2. OTROS MARCOS PARA USAR EN AMAZON EMR
- 8.3. LABORATORIO PRÁCTICO 4: PROCESAMIENTO DE REGISTROS DEL SERVIDOR CON HIVE EN AMAZON EMR

9. INTERFACES WEB EN AMAZON EMR

- 9.1. TONO EN AMAZON EMR
- 9.2. MONITOREO DE SU CLÚSTER
- 9.3. LABORATORIO PRÁCTICO 5: EJECUCIÓN DE GUIONES DE CERDOS EN HUE EN AMAZON EMR

10. APACHE SPARK EN AMAZON EMR

- 10.1. APACHE SPARK
- 10.2. USO DE SPARK
- 10.3. LABORATORIO PRÁCTICO 6: PROCESAMIENTO DE DATOS DE TAXI DE NY MEDIANTE APACHE SPARK

11. USO DE AWS GLUE PARA AUTOMATIZAR CARGAS DE TRABAJO ETL

- 11.1. ¿QUÉ ES AWS GLUE?

11.2. AWS GLUE: ORQUESTACIÓN DEL TRABAJO

12. AMAZON REDSHIFT Y BIG DATA

12.1. ALMACENES DE DATOS VS. BASES DE DATOS TRADICIONALES

12.2. DESPLAZAMIENTO AL ROJO AMAZONICO

12.3. ARQUITECTURA DE CAMBIO AL ROJO DE AMAZON

13. ASEGURAR SUS IMPLEMENTACIONES DE AMAZON

13.1. ASEGURANDO SUS IMPLEMENTACIONES DE AMAZON

13.2. DESCRIPCIÓN GENERAL DE LA SEGURIDAD DE AMAZON EMR

13.3. DESCRIPCIÓN GENERAL DE LA GESTIÓN DE IDENTIDAD Y ACCESO (IAM) DE AWS

13.4. ASEGURAR DATOS

13.5. DESCRIPCIÓN GENERAL DE LA SEGURIDAD DE AMAZON KINESIS

13.6. DESCRIPCIÓN GENERAL DE LA SEGURIDAD DE AMAZON DYNAMODB

13.7. DESCRIPCIÓN GENERAL DE SEGURIDAD DE AMAZON REDSHIFT

14. GESTIÓN DE COSTOS DE BIG DATA

14.1. CONSIDERACIONES DE COSTE TOTAL PARA AMAZON EMR

14.2. MODELOS DE PRECIOS AMAZON EC2

14.3. MODELOS DE PRECIOS DE AMAZON KINESIS

14.4. CONSIDERACIONES DE COSTOS PARA AMAZON DYNAMODB

14.5. CONSIDERACIONES DE COSTOS Y MODELOS DE PRECIOS PARA AMAZON REDSHIFT

14.6. OPTIMIZACIÓN DEL COSTE CON AWS

15. VISUALIZACIÓN Y ORQUESTACIÓN DE BIG DATA

15.1. VISUALIZACIÓN DE GRANDES DATOS

15.2. AMAZON QUICKSIGHT

15.3. ORQUESTAR UN FLUJO DE TRABAJO DE BIG DATA

15.4. LABORATORIO PRÁCTICO 7: USO DE TIBCO SPOTFIRE PARA VISUALIZAR DATOS

16. PATRONES DE DISEÑO DE BIG DATA

16.1. ARQUITECTURAS COMUNES

17. RESUMEN DEL CURSO

17.1. ¿QUE SIGUE?

★ BENEFICIOS

- Al finalizar el curso, los participantes comprenderán cómo se pueden usar servicios como AWS Glue, Amazon Kinesis, Amazon Redshift, Amazon Athena y Amazon QuickSight con cargas de trabajo de big data.