

## Big Data on AWS

Duración: 3 días

### TEMARIO

*Big data en AWS* le presenta las soluciones de big data basadas en la nube, como Amazon EMR, Amazon Redshift, Amazon Kinesis y el resto de la plataforma de big data de AWS. En este curso le mostramos cómo utilizar Amazon EMR para procesar datos con el amplio ecosistema de herramientas de Hadoop como Hive y Hue. También le enseñamos 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.

### DIRIGIDO A

Este curso está dirigido a:

- Individuos responsables del diseño e implementación de soluciones para big data, es decir, arquitectos de soluciones
- Científicos y analistas de datos interesados en aprender sobre los servicios y patrones de arquitectura tras las soluciones de big data en AWS

### OBJETIVOS DEL CURSO

Este curso le enseñará a:

- Implementar soluciones de AWS en un ecosistema de big data.
- Utilizar Apache Hadoop en el contexto de Amazon EMR.
- Identificar los componentes de un clúster de Amazon EMR.
- Lanzar y configurar un clúster de Amazon EMR.
- Aprovechar marcos de programación habituales disponibles para Amazon EMR, entre otros, Hive, Pig y Streaming.
- Aprovechar Hue para facilitar más el uso de Amazon EMR.
- Utilizar análisis integrados en memoria con Spark en Amazon EMR.
- Elegir opciones apropiadas de almacenamiento de datos para AWS.
- Identificar los beneficios de utilizar Amazon Kinesis para un procesamiento de big data casi en tiempo real.
- Aprovechar Amazon Redshift para almacenar y analizar datos de forma eficaz.
- Entender y administrar aspectos relacionados con costos y seguridad para una solución de big data.
- Identificar opciones para administrar, transferir y comprimir datos.
- Aprovechar Amazon Athena para análisis de consultas ad hoc.
- Utilizar AWS Glue para automatizar cargas de trabajo de ETL.
- Utilizar software de visualización para describir datos y consultas con Amazon QuickSight.
- Organizar flujos de trabajo de big data mediante AWS Data Pipeline.

## CONTENIDO

### Día 1

- Información general sobre big data
- Incorporación de datos
- Streaming de big data y Amazon Kinesis
- Uso de Amazon Kinesis para distribuir por streaming y analizar logs del servidor Apache
- Soluciones de almacenamiento
- Uso de Amazon Athena para analizar datos de logs
- Introducción a Apache Hadoop y Amazon EMR

### Día 2

- Uso de Amazon Elastic MapReduce
- Almacenamiento y consultas de datos en DynamoDB
- Marcos de programación Hadoop
- Procesamiento de logs de servidor con Hive en Amazon EMR
- Mejora de la experiencia de Amazon EMR con Hue
- Ejecución de scripts de Pig con Hue en Amazon EMR
- Spark en Amazon EMR
- Procesamiento de conjuntos de datos de New York Taxi con Spark en Amazon EMR

### Día 3

- Uso de AWS Glue para automatizar cargas de trabajo de ETL
- Amazon Redshift y big data
- Visualización y orquestación de big data
- Administración de costos de Amazon EMR
- Protección de soluciones de big data
- Patrones de diseño de big data

## PRE-REQUISITOS

Aconsejamos que los asistentes a este curso cumplan con los siguientes requisitos previos:

- Conocimientos básicos de tecnologías de big data, incluidos Apache Hadoop, HDFS y consultas SQL/NoSQL
- Los alumnos deben completar la formación técnica digital gratuita Big Data Technology Fundamentals o bien disponer de una experiencia equivalente
- Conocimiento práctico sobre los principales servicios de AWS y la implementación de la nube pública
- Los estudiantes deben completar el curso Fundamentos técnicos de AWS o disponer de experiencia equivalente
- Conocimientos básicos sobre el almacenamiento de datos, los sistemas de bases de datos relacionales y el diseño de base de datos