

# Módulo 1

## Visión general de Big Data

- Qué es el big data
- El pipeline de big data
- Principios arquitectónicos de Big Data

# Módulo 2

## Ingestión y transferencia de Big Data

- Resumen: Ingestión de datos
- Transferencia de datos

# Módulo 3

## Streaming de big data y Amazon Kinesis

- Procesamiento de flujos de big data
- Amazon Kinesis
- Amazon Kinesis Data Firehose
- Transmisiones de vídeo de Amazon Kinesis
- Análisis de datos de Amazon Kinesis
- Laboratorio práctico 1: transmisión y procesamiento de registros de servidor Apache mediante Amazon Kinesis

# Módulo 4

## Soluciones de almacenamiento de big data

- Opciones de almacenamiento de datos de AWS
- Conceptos de soluciones de almacenamiento
- Factores a la hora de elegir un almacén de datos
- Módulo 5: Procesamiento y análisis de big data

- Procesamiento y análisis de big data
- Atenea amazónica
- Laboratorio práctico 2: Uso de Amazon Athena para analizar datos de registro

# Módulo 5

## Operaciones financieras en la nube

- Cambio organizacional para CFM
- Modelos de organización para CFM
- Actividad 5.1: Modelos organizativos
- Establecer una cultura organizacional consciente de los costos
- Gobernanza, control y agilidad
- Bloques de creación de control y gobernanza de AWS
- Gobernanza automatizada mediante los servicios de AWS

# Módulo 6

## Apache Hadoop y Amazon EMR

- Introducción a Amazon EMR y Apache Hadoop
- Prácticas recomendadas para ingerir datos
- Amazon EMR
- Arquitectura de Amazon EMR
- Laboratorio práctico 3: Almacenamiento y consulta de datos en Amazon DynamoDB

# Módulo 7

## Uso de Amazon EMR

- Desarrollo y ejecución de la aplicación
- Lanzamiento del clúster
- Manejo de la salida de los trabajos completados





## Módulo 8

### Marcos de programación de Hadoop

- Marcos de Hadoop
- Otros marcos para su uso en Amazon EMR
- Laboratorio práctico 4: Procesamiento de registros del servidor con Hive en Amazon EMR

## Módulo 9

### Interfaces web en Amazon EMR

- Hue en Amazon EMR
- Supervisión del clúster
- Laboratorio práctico 5: Ejecución de scripts de Cerdo en Hue en Amazon EMR

## Módulo 10

### Apache Spark en Amazon EMR

- Apache Spark
- Uso de Spark
- Laboratorio práctico 6: Procesamiento de datos de taxis de Nueva York con Apache Spark

## Módulo 11

### Uso de AWS Glue para automatizar cargas de trabajo ETL

- ¿Qué es AWS Glue?
- AWS Glue: orquestación de trabajos

## Módulo 12

### Amazon Redshift y big data

- Almacenes de datos frente a bases de datos tradicionales
- Amazon Redshift
- Arquitectura de Amazon Redshift

## Módulo 13

### Protección de las implementaciones de Amazon

- Protección de las implementaciones de Amazon
- Información general sobre seguridad de Amazon EMR
- Información general sobre AWS Identity and Access Management (IAM)
- Protección de datos
- Información general sobre seguridad de Amazon Kinesis
- Información general sobre la seguridad de Amazon DynamoDB
- Información general sobre seguridad de Amazon Redshift

## Módulo 14

### Uso de Amazon EMR

- Consideraciones de costo total para Amazon EMR
- Modelos de precios de Amazon EC2
- Modelos de precios de Amazon Kinesis
- Consideraciones de costos para Amazon DynamoDB





Temario

- Consideraciones de costos y modelos de precios para Amazon Redshift
- Optimización de costos con AWS

## Módulo 17

### Resumen del curso

- ¿Qué sigue?

## Módulo 15

### Visualización y orquestación de big data

- Visualización de big data
- Amazon QuickSight
- Orquestación de un flujo de trabajo de big data
- Laboratorio práctico 7: Uso de TIBCO Spotfire para visualizar datos

## Módulo 16

### Patrones de diseño de Big Data

- Arquitecturas comunes

