

Building Modern Data Analytics Solutions on AWS

AWS-ANALYTICS

La colección de cursos Building Modern Data Analytics Solutions on AWS (Creación de soluciones de análisis de datos modernas en AWS) de un día de duración y de nivel intermedio impartidos por un instructor profundiza en Amazon Lake Formation, Amazon Glue, Amazon EMR, Amazon Kinesis y Amazon Redshift, así como en las ideas actuales sobre la creación y el funcionamiento de canalizaciones de análisis de datos para convertir los datos en información.

Donde quiera que usted o sus clientes se encuentren en el proceso de la modernización de los datos, nuestra colección de cursos Building Modern Data Analytics on AWS (Creación de soluciones de análisis de datos modernas en AWS) le permite seleccionar la formación adecuada para satisfacer sus necesidades específicas de aprendizaje.

Lo que aprenderá

En la colección de cuatro cursos Building Modern Data Analytics Solutions on AWS (Creación de soluciones de análisis de datos modernas en AWS), recibirá una formación completa para desarrollar habilidades de datos modernas, entre ellas:

- ▶ Cómo aprovechar los servicios de datos de AWS para almacenar, procesar, analizar, transmitir y consultar datos para tomar decisiones con velocidad y agilidad a escala.
- ▶ Cómo modernizar las soluciones de datos de principio a fin.
- ▶ Habilidades para poner sus datos a trabajar para tomar mejores decisiones y más informadas, responder más rápido a lo inesperado y descubrir nuevas oportunidades.

Perfil de la audiencia

Recomendamos todos los cursos para clientes y socios en estos roles:

- ▶ Ingenieros de almacenamiento de datos
- ▶ Ingenieros de plataformas de datos
- ▶ Arquitectos de soluciones

También recomendamos a los desarrolladores asistir a Building Streaming Data Analytics Solutions on AWS (Creación de soluciones de análisis de datos de streaming en AWS).

ESQUEMA DEL CURSO

Curso 1: Construcción de Data Lakes en AWS

Módulo 1: Introducción a los data lakes

- Describir el valor de los data lakes
- Comparar los data lakes y data warehouses
- Describir los componentes de un data lake
- Reconocer arquitecturas comunes basadas en data lakes

Módulo 2: Ingesta, catalogación y preparación de datos

- Describir la relación entre el almacenamiento en data lakes y la ingesta de datos.
- Describir los rastreadores de AWS Glue y cómo se utilizan para crear un catálogo de datos.
- Identificar el formato, partición y compresión de los datos para un almacenamiento y consulta eficientes.
- Laboratorio 1: Configurar un Data Lake simple.

Módulo 3: Procesamiento y análisis de datos

- Reconocer cómo se aplica el procesamiento de datos en un data lake
- Utilizar AWS Glue para procesar datos dentro de un data lake
- Describir cómo utilizar Amazon Athena para analizar datos en un data lake

Módulo 4: Construcción de un Data Lake con AWS Lake Formation

- Describir las características y beneficios de AWS Lake Formation
- Utilizar AWS Lake Formation para crear un data lake
- Comprender el modelo de seguridad de AWS Lake Formation
- Laboratorio 2: Construcción de un Data Lake utilizando AWS Lake Formation

Módulo 5: Configuraciones adicionales de Lake Formation

- Automatizar AWS Lake Formation usando plantillas (blueprints) y flujos de trabajo
- Aplicar controles de seguridad y acceso en AWS Lake Formation
- Realizar emparejamiento de registros con AWS Lake Formation FindMatches
- Visualizar datos con Amazon QuickSight

- Laboratorio 3: Automatización de la creación de data lakes con plantillas de AWS Lake Formation
- Laboratorio 4: Visualización de datos utilizando Amazon QuickSight

Módulo 6: Arquitectura y repaso del curso

- Evaluación de conocimientos posterior al curso
- Revisión de arquitectura
- Revisión general del curso

Curso 2: Creación de soluciones de análisis de datos por lotes en AWS

Módulo A: Descripción general del análisis de datos y el flujo de datos (Data pipeline)

- Casos de uso del análisis de datos
- Uso del flujo de datos para el análisis

Módulo 1: Introducción a Amazon EMR

- Uso de Amazon EMR en soluciones de análisis
- Arquitectura de clústeres de Amazon EMR
- Demostración interactiva 1: Implementación de un clúster de Amazon EMR
- Estrategias de gestión de costos

Módulo 2: Flujo de análisis de datos con Amazon EMR: Ingesta y almacenamiento

- Optimización del almacenamiento con Amazon EMR
- Técnicas de ingesta de datos

Módulo 3: Análisis de datos por lotes de alto rendimiento con Apache Spark en Amazon EMR

- Casos de uso de Apache Spark en Amazon EMR
- Por qué usar Apache Spark en Amazon EMR
- Conceptos clave de Spark
- Demostración interactiva 2: Conexión a un clúster EMR y ejecución de comandos Scala usando Spark Shell
- Transformación, procesamiento y análisis de datos
- Uso de notebooks con Amazon EMR
- Laboratorio práctico 1: Análisis de datos de baja latencia con Apache Spark en Amazon EMR

Módulo 4: Procesamiento y Análisis de Datos por Lotes con Amazon EMR y Apache Hive

- Uso de Amazon EMR con Hive para procesar datos por lotes
- Transformación, procesamiento y análisis de datos
- Laboratorio práctico 2: Procesamiento de datos por lotes con Amazon EMR y Hive
- Introducción a Apache HBase en Amazon EMR

Módulo 5: Procesamiento de Datos sin Servidor (Serverless)

- Procesamiento, transformación y análisis de datos sin servidor
- Uso de AWS Glue con cargas de trabajo en Amazon EMR
- Laboratorio práctico 3: Orquestación del procesamiento de datos en Spark utilizando AWS Step Functions

Módulo 6: Seguridad y Monitoreo de Clústeres Amazon EMR

- Protección de los clústeres de EMR
- Demostración interactiva 3: Cifrado del lado del cliente con EMRFS
- Monitoreo y resolución de problemas en clústeres Amazon EMR
- Demostración: Revisión del historial del clúster Apache Spark

Módulo 7: Diseño de Soluciones de Análisis de Datos por Lotes

- Casos de uso del análisis de datos por lotes
- Actividad: Diseño de un flujo de trabajo para análisis de datos por lotes

Módulo B: Desarrollo de Arquitecturas de Datos Modernas en AWS

- Arquitecturas de datos modernas

Curso 3: Creación de Soluciones de Analítica de Datos con Amazon Redshift

Módulo A: Panorama general de la analítica de datos y la canalización de datos

- Casos de uso comunes en analítica de datos
- Uso del flujo de datos para análisis

Módulo 1: Uso de Amazon Redshift en el flujo de análisis de datos

- ¿Por qué Amazon Redshift para el almacenamiento de datos (data warehousing)?

- Descripción general de Amazon Redshift

Módulo 2: Fundamentos de Amazon Redshift

- Arquitectura de Amazon Redshift
- Demostración 1: exploración de la consola de amazon Redshift
- Características de Amazon Redshift
- Laboratorio práctico 1: carga y consulta de datos en un clúster de Redshift

Módulo 3: Ingesta y almacenamiento de datos

- Estrategias de ingesta de datos

- Demostración 2: conexión al clúster de Amazon Redshift desde Jupyter Notebook usando Data API
- Distribución y almacenamiento de datos
- Demostración 3: análisis de datos semiestructurados con el tipo de dato SUPER
- Consultas de datos en Amazon Redshift
- Laboratorio práctico 2: análisis de datos con Amazon Redshift Spectrum

Módulo 4: Procesamiento y optimización de datos

- Transformación de datos
- Consultas avanzadas
- Laboratorio práctico 3: transformación y consulta de datos en Amazon Redshift
- Gestión de recursos
- Demostración 4: Aplicación de gestión de cargas de trabajo mixtas en Amazon Redshift
- Automatización y optimización
- Demostración 5: redimensionamiento de clústeres de Amazon Redshift de dc2.large a ra3.xlplus

Módulo 5: Seguridad y monitoreo de clústeres Redshift

- Protección del clúster de Amazon Redshift
- Supervisión y resolución de problemas en clústeres

Módulo 6: Diseño de soluciones analíticas en data warehouses

- Revisión de casos de uso en data warehouses
- Actividad: diseño de un flujo de trabajo analítico para data warehouse

Módulo B: Desarrollo de arquitecturas modernas de datos en AWS

- Arquitecturas de datos modernas

Curso 4: Creación de Soluciones de Analítica de Datos en Tiempo Real con AWS

Módulo A: Descripción general del análisis de datos y del flujo de datos (Data Pipeline)

- Casos de uso del análisis de datos
- Uso del flujo de datos para el análisis

Módulo 1: Uso de servicios de streaming en el flujo de análisis de datos

- Relevancia del análisis de datos en tiempo real
- El flujo de análisis de datos en streaming
- Conceptos de streaming

Módulo 2: Introducción a los servicios de streaming de AWS

- Servicios de datos en streaming en AWS
- Amazon Kinesis en soluciones de análisis
- Demostración: exploración de Amazon Kinesis Data Streams
- Laboratorio práctico: configuración de un flujo de entrega de datos en streaming con Kinesis
- Uso de Amazon Kinesis Data Analytics
- Introducción a Amazon MSK (Managed Streaming for Apache Kafka)
- descripción general de Spark Streaming

Módulo 3: Uso de Amazon Kinesis para análisis en tiempo real

- Exploración de Amazon Kinesis utilizando una carga de trabajo de clickstream
- Creación de flujos de datos y entrega con Kinesis
- Demostración: comprensión de productores y consumidores
- Creación de productores de flujo
- Creación de consumidores de flujo
- Creación e implementación de aplicaciones Flink en Kinesis Data Analytics
- Demostración: uso de notebooks Zeppelin para Kinesis Data Analytics
- Laboratorio práctico: análisis en tiempo real con Amazon Kinesis y Apache Flink

Módulo 4: Seguridad, monitoreo y optimización en Amazon Kinesis

- Optimización de Amazon Kinesis para obtener información empresarial accionable
- Mejores prácticas de seguridad y monitoreo

Módulo 5: Uso de Amazon MSK en soluciones de análisis de datos en streaming

- Casos de uso de Amazon MSK
- Creación de clústeres MSK
- Demostración: aprovisionamiento de un clúster MSK
- Ingesta de datos en MSK
- Laboratorio práctico: introducción al control de acceso con Amazon MSK
- Procesamiento y transformación de datos en Amazon MSK

Módulo 6: Seguridad, monitoreo y optimización en Amazon MSK

- Optimización de Amazon MSK
- Demostración: escalamiento de almacenamiento en Amazon MSK

- ▶ Laboratorio práctico: canalización de datos en streaming y despliegue de aplicaciones con Amazon MSK
- ▶ Seguridad y monitoreo
- ▶ Demostración: monitoreo de clústeres MSK

Módulo 7: Diseño de soluciones de análisis de datos en streaming

- ▶ Revisión de casos de uso
- ▶ Ejercicio práctico: diseño de un flujo de trabajo de análisis de datos en tiempo real

Módulo B: Desarrollo de arquitecturas modernas de datos en AWS

- ▶ Arquitecturas de datos modernas

Debido a las constantes actualizaciones de los contenidos de los cursos por parte del fabricante, el contenido de este temario puede variar con respecto al publicado en el sitio oficial, sin embargo, Netec siempre entregará la versión actualizada de éste.