Big Data on AWS (BDAWS)

ID AW-BDAWS **Precio** a solicitud **Duración** 3 días

Quién debería asistir

This course is intended for:

- Individuals responsible for designing and implementing big data solutions, namely Solutions Architects
- Data Scientists and Data Analysts interested in learning about big data solutions on AWS

Prerrequisitos

We recommend that attendees of this course have the following prerequisites:

- Basic familiarity with big data technologies, including Apache Hadoop, MapReduce, HDFS, and SQL/NoSQL querying
- Students should complete the Big Data Technology Fundamentals web-based training or have equivalent experience
- Working knowledge of core AWS services and public cloud implementation
- Students should complete the AWS Technical Essentials (AWSE) course or have equivalent experience
- Basic understanding of data warehousing, relational database systems, and database design

Objetivos del curso

This course teaches you how to:

- Fit AWS solutions inside of a big data ecosystem
- Leverage Apache Hadoop in the context of Amazon EMR
- Identify the components of an Amazon EMR cluster
- Launch and configure an Amazon EMR cluster
- Leverage common programming frameworks available for Amazon EMR including Hive, Pig, and Streaming
- Leverage Hue to improve the ease-of-use of Amazon EMR
- Use in-memory analytics with Spark and Spark SQL on Amazon EMR
- Choose appropriate AWS data storage options
- Identify the benefits of using Amazon Kinesis for near real-time big data processing
- Define data warehousing and columnar database concepts
- Leverage Amazon Redshift to efficiently store and analyze data
- Comprehend and manage costs and security for Amazon EMR and Amazon Redshift deployments
- Identify options for ingesting, transferring, and compressing data
- Use visualization software to depict data and queries
- Orchestrate big data workflows using AWS Data Pipeline

Contenido del curso

Day 1

- Overview of Big Data
- Ingestion, Transfer, and Compression
- Storage Solutions
- Storing and Querying Data on DynamoDB
- Big Data Processing and Amazon Kinesis
- Introduction to Apache Hadoop and Amazon EMR
- Using Amazon Elastic MapReduce

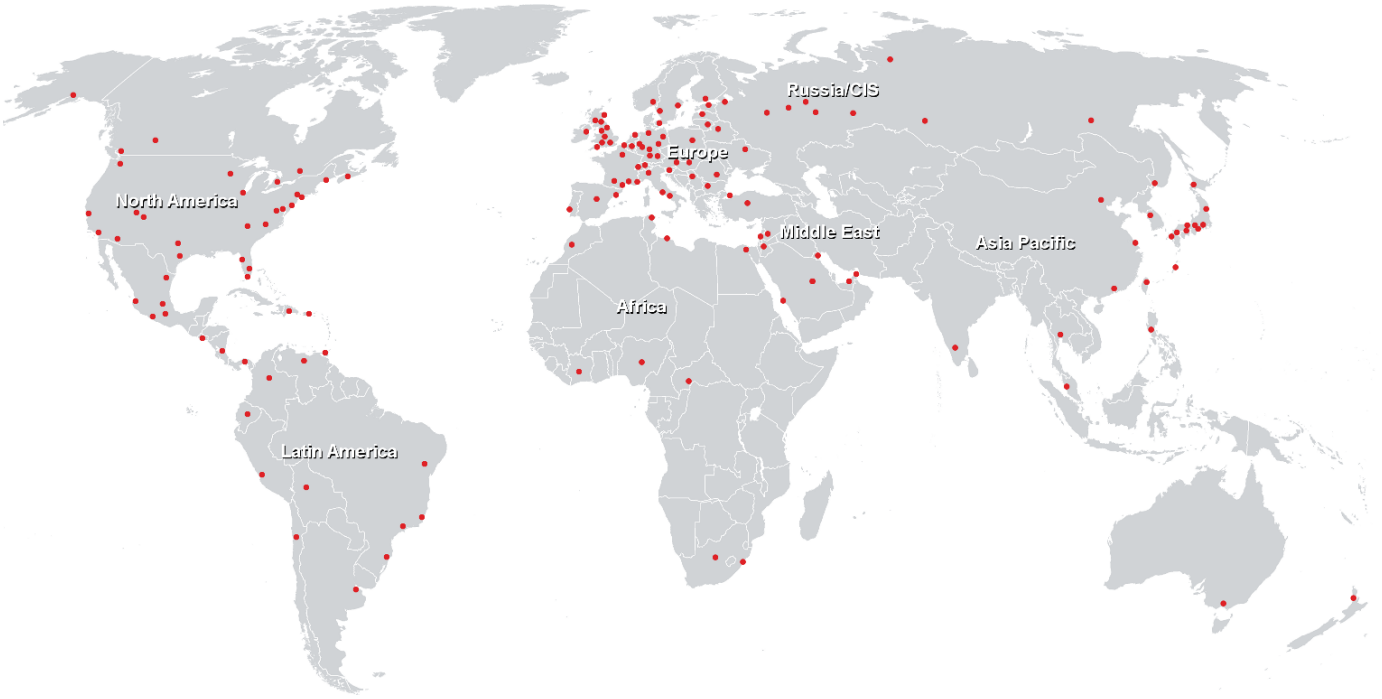
Day 2

- Hadoop Programming Frameworks
- Processing Server Logs with Hive on Amazon EMR
- Processing Chemistry Data Using Hadoop Streaming on Amazon EMR
- Streamlining Your Amazon EMR Experience with Hue
- Running Pig Scripts in Hue on Amazon EMR
- Spark on Amazon EMR
- Interactively Creating and Querying Tables with Spark and Spark SQL on Amazon EMR
- Managing Amazon EMR Costs
- Securing your Amazon EMR Deployments

Day 3

- Data Warehouses and Columnar Datastores
- Amazon Redshift and Big Data
- Optimizing Your Amazon Redshift Environment
- Big Data Design Patterns
- Visualizing and Orchestrating Big Data
- Using Tibco Spotfire to Visualize Big Data

Centros de Entrenamiento Mundial



,