Big Data on AWS

Course ID #: 1190-210-00-W
Hours: 21

Course Content

Course Description:

In this course, you will learn about cloud-based Big Data solutions such as Amazon EMR, Amazon Redshift, Amazon Kinesis, and the rest of the AWS Big Data platform. We will show you how to use Amazon EMR to process data using the broad ecosystem of Hadoop tools like Hive and Hue. We will also teach you how to create Big Data environments, work with Amazon DynamoDB, Amazon Redshift, Amazon QuickSight, Amazon Athena, and Amazon Kinesis, and leverage best practices to design Big Data environments for security and cost-effectiveness.

At Course Completion:

After competing this course, student will be able 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 on Amazon EMR
- Choose appropriate AWS data storage options
- Identify the benefits of using Amazon Kinesis for near real-time big data processing
- Leverage Amazon Redshift to efficiently store and analyze data
- Comprehend and manage costs and security for a big data solution
- Secure a Big Data solution
- Identify options for ingesting, transferring, and compressing data
- Leverage Amazon Athena for ad-hoc query analytics
- Leverage AWS Glue to automate ETL workloads
- Use visualization software to depict data and queries using Amazon QuickSight
- Orchestrate big data workflows using AWS Data Pipeline

Prerequisites:

- Basic familiarity with big data technologies, including Apache Hadoop, MapReduce, HDFS, and SQL/NoSQL querying
- Students should complete the free 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 course or have equivalent experience
- Basic understanding of data warehousing, relational database systems, and database design

Big Data on AWS

Course ID #: 1190-210-00-W

Hours: 21

Target Student:

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

Topics:

Day 1

- Overview of Big Data
- Big Data Ingestion and Transfer
- Big Data Streaming and Amazon Kinesis
- Lab 1: Using Amazon Kinesis to Stream and Analyze Apache Server Log Data
- Big Data Storage Solutions
- Big Data Processing and Analytics
- Lab 2: Using Amazon Athena to Query Log Data From Amazon S3

Day 2

- Apache Hadoop and Amazon EMR
- Lab 3: Storing and Querying Data on Amazon DynamoDB
- Using Amazon EMR
- Hadoop Programming Frameworks
- Lab 4: Processing Server Logs With Hive on Amazon EMR
- Web Interfaces on Amazon EMR
- Lab 5: Running Pig Scripts in Hue on Amazon EMR
- Apache Spark on Amazon EMR
- Lab 6: Processing NY Taxi data using Spark on Amazon EMR

Day 3

- Using AWS Glue to automate ETL workloads
- Amazon Redshift and Big Data
- Visualizing and Orchestrating Big Data
- Lab 7: Using TIBCO Spotfire to Visualize Data
- Managing Big Data Costs
- Securing Your Amazon Deployments
- Big Data Design Patterns