Course duration

- 3 days

Course Benefits

- 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

Authorized AWS Training

Webucator has partnered with an Authorized AWS training delivery partner to offer official AWS courses utilizing Amazon Authorized Instructors.

Course Outline

1. Overview of Big Data
2. Big Data Ingestion and Transfer
3. Big Data Streaming and Amazon Kinesis
4. Lab: Using Amazon Kinesis to Stream and Analyze Apache Server Log Data
5. Big Data Storage Solutions
6. Big Data Processing and Analytics
7. Lab: Using Amazon Athena to Query Log Data From Amazon S3
8. Apache Hadoop and Amazon EMR

9. Lab: Storing and Querying Data on Amazon DynamoDB
10. Using Amazon EMR
11. Hadoop Programming Frameworks
12. Lab: Processing Server Logs With Hive on Amazon EMR
13. Web Interfaces on Amazon EMR
14. Lab: Running Pig Scripts in Hue on Amazon EMR
15. Apache Spark on Amazon EMR
16. Lab: Processing NY Taxi data using Spark on Amazon EMR
17. Using AWS Glue to automate ETL workloads
18. Amazon Redshift and Big Data
19. Visualizing and Orchestrating Big Data
20. Lab: Using TIBCO Spotfire to Visualize Data
21. Managing Big Data Costs
22. Securing Your Amazon Deployments
23. Big Data Design Patterns

Class Materials

Each student will receive a comprehensive set of materials, including course notes and all the class examples.

Class Prerequisites

Experience in the following *is required* for this AWS class:

- 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.
- Knowledge acquired from course the AWS Technical Essentials course or have equivalent experience.
- Basic understanding of data warehousing, relational database systems, and database design.

Prerequisite Courses

Courses that can help you meet these prerequisites:

- AWS Technical Essentials