

Big Data on AWS Training

Modality: Virtual Classroom

Duration: 3 Days

SATV Value:

CLC:

NATU:

SUBSCRIPTION: No

About this Course:

This course is designed to give the participants an insight into big data solutions based on Cloud such as Amazon EMR, Amazon Redshift, Amazon Kinesis and the other services available on the AWS big data platform. This course will introduce and help students gain the skills needed to use Amazon EMR to process data via the use of multiple available Hadoop tools such as Hive and Hue. This course will also teach students how to design efficient big data environments. The participants will also gain skills needed to effectively use Amazon DynamoDB Amazon Redshift, Amazon Quicksight, Amazon Athena and Amazon Kinesis. They will be taught the best practices to design cost-efficient, secure and effective big data environments

Course Objective:

Upon completion of this course, the participant should have an advanced skill set and a sound working knowledge of the following principals while also be able to;

- Learn how to incorporate AWS solutions in big data ecosystems
- Learn how to Leverage Apache Hadoop for use in Amazon EMR
- Learn the different components of a cluster in Amazon EMR
- Gain the skills needed to configure and launch and Amazon EMR cluster
- Learn how to leverage common programming frameworks available on Amazon EMR such as but not limited to Hive, Pig, and Streaming
- Learn how to effectively leverage Hue to improve the ease-of-use of Amazon EMR
- Learn how to use in memory analytics with Spark on Amazon EMR
- Learn how to choose the best storage options from the data storage options available on AWS
- Recognize the pros of Amazon Kinesis for near real-time big data processing
- Learn how to efficiently store and analyze data using Amazon Redshift
- Create a cost effective big data solution and manage its security
- Learn how to secure a big data solution
- Learn how to identify options for ingesting, transferring, and compressing data
- Learn how to leverage Amazon Athena for ad-hoc query analytics
- Learn how to Leverage AWS Glue to automate ETL workloads
- Learn the use of visualization software for the depiction of data and queries through Amazon QuickSight
- Plan and demonstrate big data workflows using AWS Data Pipeline

Audience:

This particular course is aimed at the following audience;

- Solutions architects or those responsible for designing and executing big data solutions
- Data scientists and data analysts with a wish to learn about the patterns of architecture lying behind the big data solutions available on AWS

Prerequisites:

The following prerequisites are absolutely necessary to be eligible to take this Microsoft Word 2019 course;

- Have a basic understanding of big data tech namely Apache, Hadoop, Mapreduce, HDFS and SQL/NoSQL queries.
- Have completed the Big Data technology Fundamentals web based training available for free or have experience equivalent to the course
- Possess a functional knowledge of the core services provided by AWS and the implementation of public cloud
- Have a fundamental know how of data warehousing, relational database systems and database design

Course Outline:

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