Building Batch Data Analytics Solutions on AWS Training

Duration: **1 Day**

BILGINC IT Academy
info@bilginc.com
bilginc.com

This course is part of a collection called 'Building Modern Data Analytics Solutions on AWS', which consists of 4 courses. You have the option to book any of the individual courses separately. If you prefer to attend all 4 courses, you can book the combined course called Building Modern Data Analytics Solutions on AWS.

Overview

In this course, you will learn to build batch data analytics solutions using Amazon EMR, an enterprise-grade Apache Spark and Apache Hadoop managed service. You will learn how Amazon EMR integrates with open-source projects such as Apache Hive, Hue, and HBase, and with AWS services such as AWS Glue and AWS Lake Formation. The course addresses data collection, ingestion, cataloging, storage, and processing components in the context of Spark and Hadoop. You will learn to use EMR Notebooks to support both analytics and machine learning workloads. You will also learn to apply security, performance, and cost management best practices to the operation of Amazon EMR

Prerequisites

Students with a minimum one-year experience managing open-source data frameworks such as Apache Spark or Apache Hadoop will benefit from this course.

- We suggest the AWS Hadoop Fundamentals course for those that need a refresher on Apache Hadoop.

We recommend that attendees of this course have:

- Completed either AWS Technical Essentials or Architecting on AWS
- Completed either Building Data Lakes on AWS or Getting Started with AWS Glue

What You Will Learn

- Compare the features and benefits of data warehouses, data lakes, and modern data architectures
- Design and implement a batch data analytics solution
- Identify and apply appropriate techniques, including compression, to optimize data storage
- Select and deploy appropriate options to ingest, transform, and store data
- Choose the appropriate instance and node types, clusters, auto scaling, and network topology for a particular business use case
- Understand how data storage and processing affect the analysis and visualization mechanisms needed to gain actionable business insights
- Secure data at rest and in transit
- Monitor analytics workloads to identify and remediate problems
- Apply cost management best practices