
AWS – Big Data on Amazon Web Services

Description

In this 3-day course, you will learn about cloud-based big data solutions like Amazon EMR, Amazon Redshift, Amazon Kinesis, and the rest of the AWS big data platform. Learn to use Amazon EMR to process data using the broad ecosystem of Hadoop tools like Hive and Hue, create big data environments, work with Amazon DynamoDB, Amazon Redshift, Amazon QuickSight, Amazon Athena and Amazon Kinesis, and design big data environments for security and cost-effectiveness.

Course Content

Module 1: Overview of Big Data

- What is big data
- The big data pipeline
- Big data architectural principals

Module 2: Big Data ingestion and transfer

- Overview: Data ingestion
- Transferring data

Module 3: Big data streaming and Amazon Kinesis

- Stream processing of big data
- Amazon Kinesis
- Amazon Kinesis Data Firehose
- Amazon Kinesis Video Streams
- Amazon Kinesis Data Analytics

Module 4: Big data storage solutions

- AWS data storage options
- Storage solutions concepts
- Factors in choosing a data store

Module 5: Big data processing and analytics

- Big data processing and analytics
- Amazon Athena

Module 6: Apache Hadoop and Amazon EMR

- Introduction to Amazon EMR and Apache Hadoop
- Best practices for ingesting data
- Amazon EMR
- Amazon EMR architecture

Module 7: Using Amazon EMR

- Developing and running your application
- Launching your cluster
- Handling output from your completed jobs

Module 8: Hadoop programming frameworks

- Hadoop frameworks
- Other frameworks for use on Amazon EMR

Module 9: Web interfaces on Amazon EMR

- Hue on Amazon EMR
- Monitoring your cluster

Module 10: Apache Spark on Amazon EMR

- Apache Spark
- Using Spark

Module 11: Using AWS Glue to automate ETL workloads

- What is AWS Glue?
- AWS Glue: Job orchestration

Module 12: Amazon Redshift and big data

- Data warehouses vs. traditional databases
- Amazon Redshift
- Amazon Redshift architecture

Module 13: Securing your Amazon deployments

- Securing your Amazon deployments
- Amazon EMR security overview
- AWS Identity and Access Management (IAM) overview
- Securing data
- Amazon Kinesis security overview
- Amazon DynamoDB security overview
- Amazon Redshift security overview

Module 14: Managing big data costs

- Total cost considerations for Amazon EMR
- Amazon EC2 pricing models
- Amazon Kinesis pricing models
- Cost considerations for Amazon DynamoDB
- Cost considerations and pricing models for Amazon Redshift
- Optimizing cost with AWS

Module 15: Visualizing and orchestrating big data

- Visualizing big data
- Amazon QuickSight

- Orchestrating a big data workflow

Module 16: Big data design patterns

- Common architectures

Module 17: Course wrap-up

- What's next?

Lab / Exercises

- AWS Official Labs

Documentation

- Digital courseware included

Exam

- This course prepares you to the AWS Certified Data Analytics - Specialty exam. If you wish to take this exam, please contact our secretariat who will let you know the cost of the exam and will take care of all the necessary administrative procedures for you

Participant profiles

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

Prerequisites

- Basic familiarity with big data technologies, including Apache Hadoop, HDFS, and SQL/NoSQL querying
- Completed Data Analytics Fundamentals free digital training or equivalent experience
- Working knowledge of core AWS services and public cloud implementation
- Basic understanding of data warehousing, relational database systems, and database design
- Completed the [AWS Technical Essentials](#) classroom training or have equivalent experience

Objectives

- Use Apache Hadoop with Amazon EMR
- Launch and configure an Amazon EMR cluster
- Use common programming frameworks for Amazon EMR, including Hive, Pig, and Streaming
- Use Hue to improve the ease-of-use of Amazon EMR
- Use in-memory analytics with Spark on Amazon EMR
- Understand how services like AWS Glue, Amazon Kinesis, Amazon Redshift, Amazon Athena, and Amazon QuickSight can be used with big data workloads

Niveau

Intermédiaire

Classroom Registration Price (CHF)

2500

Virtual Classroom Registration Price (CHF)

2500

Duration (in Days)

3

Reference

AWS-11