AWS – Building Batch Data Analytics Solutions on AWS

Description

Design batch data analytics solutions with AWS

Data analytics has become an essential pillar for companies looking to leverage their information for strategic decisions. The AWS – Building Batch Data Analytics Solutions on AWS course will help you acquire the necessary skills to design batch processing pipelines on AWS using services like Amazon EMR. You will learn to use Apache Spark and Apache Hadoop to optimize data collection, ingestion, and processing. This course is designed for those who want to master AWS tools while integrating security and cost management best practices.

Course Content
Module A: Overview of data analytics and the data pipeline

- Data analytics use cases
- Using the data pipeline for analytics

Module 1: Introduction to Amazon EMR

- Using Amazon EMR in analytics solutions
- Amazon EMR cluster architecture
- Interactive demo: Launching an Amazon EMR cluster
- Cost management strategies

Module 2: Data analytics pipeline using Amazon EMR: ingestion and storage

- Storage optimization with Amazon EMR
- Data ingestion techniques

Module 3: High-performance batch data analytics using Apache Spark on Amazon EMR

- Apache Spark on Amazon EMR use cases
- Why choose Apache Spark on Amazon EMR
- Spark concepts
- Interactive demo: Connect to an EMR cluster and run Scala commands using the Spark shell

- Transformation, processing, and analytics
- Using notebooks with Amazon EMR
- Practice lab: Low-latency data analytics using Apache Spark on Amazon EMR

Module 4: Processing and analyzing batch data with Amazon EMR and Apache Hive

- Using Amazon EMR with Hive to process batch data
- Transformation, processing, and analytics
- Practice lab: Batch data processing using Amazon EMR and Hive
- Introduction to Apache HBase on Amazon EMR

Module 5: Serverless data processing

- Serverless data processing, transformation, and analytics
- Using AWS Glue with Amazon EMR workloads
- Practice lab: Orchestrate data processing in Spark using AWS Step Functions

Module 6: Securing and monitoring Amazon EMR clusters

- Securing EMR clusters
- Interactive demo: Client-side encryption with EMRFS
- Monitoring and troubleshooting Amazon EMR clusters
- Demo: Reviewing Apache Spark cluster history

Module 7: Designing batch data analytics solutions

- Batch data analytics use cases
- Activity: Designing a batch data analytics workflow

Module B: Developing modern data architectures on AWS

- Modern data architectures

Documentation

- Digital course materials included

Participant profiles

- Data platform engineers
- Data solution architects
- Analytics pipeline managers
- AWS operators specializing in Big Data

Prerequisites

- Experience with Apache Spark and Apache Hadoop (minimum 1 year)
- Basic knowledge in managing clusters and networks on AWS
- Data security and encryption knowledge
- Understanding of data architecture concepts
- Familiarity with performance monitoring and optimization tools

Objectives

- Design data analytics pipelines

- Optimize data storage and ingestion
- Use Amazon EMR with Apache Spark and Hive
- Apply cost management practices
- Secure clusters and data on AWS
- Monitor analytics workloads
- Automate batch data processing

Description
AWS Training - Building Batch Data Analytics Solutions on AWS
Niveau
Fondamental
Classroom Registration Price (CHF)
850
Virtual Classroom Registration Price (CHF)
850
Duration (in Days)
1
Reference
AWS-301