Building Batch Data Analytics Solutions on AWS (BBDAS)

ID BBDAS **Price** CHF 870.—(excl. VAT) **Duration** 1 days

Who should attend

This course is intended for:

- Data platform engineers
- Architects and operators who build and manage data analytics pipelines

This course is part of the following Certifications

AWS Certified Data Engineer - Associate (AWCDEA)

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 recommend that attendees of this course have:

- Completed either Architecting on AWS (AWSA) or AWS Technical Essentials (AWSE)
- Completed either Building Data Lakes on AWS (BDLA) or Getting Started with AWS Glue

Course Objectives

In this course, you will learn to:

- 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

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 1: 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 Apache Spark on Amazon EMR
- Spark concepts
- Interactive Demo 2: Connect to an EMR cluster and perform Scala commands using the Spark shell
- Transformation, processing, and analytics
- Using notebooks with Amazon EMR
- Practice Lab 1: 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 2: Batch data processing using Amazon EMR with 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 3: Orchestrate data processing in Spark using AWS Step Functions

Module 6: Security and Monitoring of Amazon EMR Clusters

- Securing EMR clusters
- Interactive Demo 3: Client-side encryption with EMRFS
- Monitoring and troubleshooting Amazon EMR clusters
- Demo: Reviewing Apache Spark cluster history

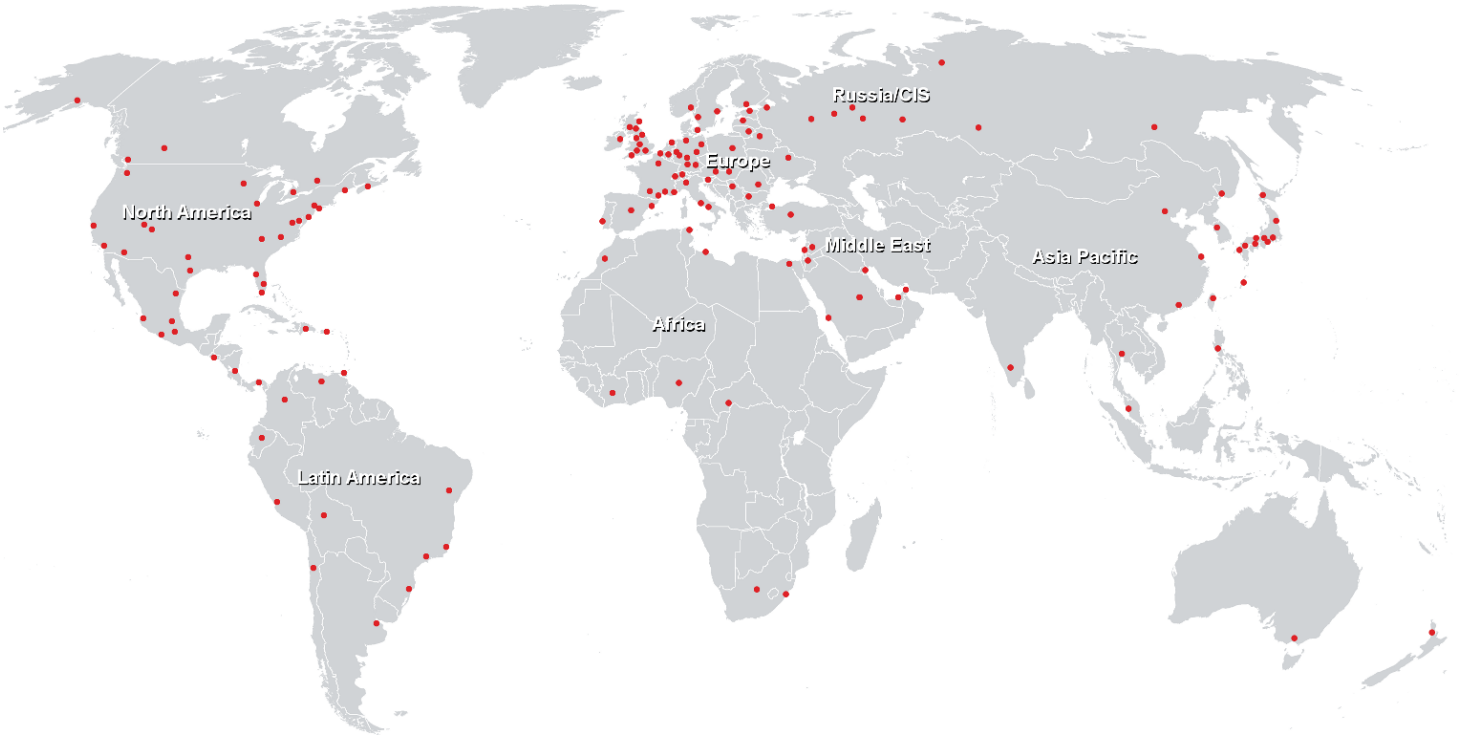
Module 7: Designing Batch Data Analytics Solutions

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

Module B: Developing Modern Data Architectures on AWS

- Modern data architectures

Training Centres worldwide





Fast Lane Institute for Knowledge Transfer (Switzerland) AG

Husacherstrasse 3
CH-8304 Wallisellen
Tel. +41 44 832 50 80

info@flane.ch, https://www.flane.ch