Building Batch Data Analytics Solutions on AWS

Category: AI, Data & Analytics | **Vendor:** Amazon Web Services

Duration: 8.00 hours (1 days) **6.5 CPD Hours** **Rating:** ⭐ 4.6 (5,878 reviews)

Course Information

Delivery Format: Instructor Led - Online

Course 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.

About This Course

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

Who Should Attend

This course is intended for:

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

Prerequisites & Entry Requirements

General 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

Learning Outcomes

Upon successful completion of this course, participants will be able to:

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

Detailed Course Outline

Module 5: Serverless Data Processing, Module 5: Serverless Data Processing, Module 5: Serverless Data Processing, Module 6: Security and Monitoring of Amazon EMR Clusters, Module 6: Security and Monitoring of Amazon EMR Clusters, Module 6: Security and Monitoring of Amazon EMR Clusters, Module 4: Processing and Analyzing Batch Data with Amazon EMR and Hive, Module 4: Processing and Analyzing Batch Data with Amazon EMR and Hive, Module 4: Processing and Analyzing Batch Data with Amazon EMR and Hive, Module 3: High-Performance Batch Data Analytics Using Apache Spark on Amazon EMR, Module 3: High-Performance Batch Data Analytics Using Apache Spark on Amazon EMR, Module 3: High-Performance Batch Data Analytics Using Apache Spark on Amazon EMR, Module 2: Data Analytics Pipeline Using Amazon EMR: Ingestion and Storage, Module 2: Data Analytics Pipeline Using Amazon EMR: Ingestion and Storage, Module 2: Data Analytics Pipeline Using Amazon EMR: Ingestion and Storage, Module 1: Introduction to Amazon EMR, Module 1: Introduction to Amazon EMR, Module 1: Introduction to Amazon EMR, Module A: Overview of Data Analytics and

the Data Pipeline, Module A: Overview of Data Analytics and the Data Pipeline, Module A: Overview of Data Analytics and the Data Pipeline, Module B: Developing Modern Data Architectures on AWS, Module B: Developing Modern Data Architectures on AWS, Module B: Developing Modern Data Architectures on AWS, Module 7: Designing Batch Data Analytics Solutions, Module 7: Designing Batch Data Analytics Solutions, Module 7: Designing Batch Data Analytics Solutions

Skills You Will Learn:

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

Frequently Asked Questions

Q: What delivery options are available for Building Batch Data Analytics Solutions on AWS?

We offer multiple delivery formats:

- Live Instructor-Led Classroom Online (Virtual/Live Online)
- Traditional Instructor-Led Classroom Training (ILT)
- On-site delivery at your offices anywhere in United Kingdom
- Private dedicated courses customized for your team

Q: How many CPD hours does this course provide?

The 1-day Building Batch Data Analytics Solutions on AWS course provides up to 6.5 CPD hours of structured learning. CPD certificates can be provided upon request.

Q: What is the duration of the Building Batch Data Analytics Solutions on AWS training?

The training takes place over 1 day(s), with each day lasting approximately 8.00 hours including breaks for lunch and refreshments.

Q: Do you provide corporate training for Building Batch Data Analytics Solutions on AWS?

Yes, we provide corporate training, dedicated training, and closed classes for Building Batch Data Analytics Solutions on AWS. Training can take place anywhere in United Kingdom including London, Manchester, Birmingham, Edinburgh, or live online allowing teams from across United Kingdom or internationally to attend.

Q: What related terms do people search for?

Popular related searches include: Building Batch Data Analytics Solutions on AWS, Data Analysis, BBDAS, Amazon Web Services, vILT

Q: Why choose Nexus Human for Building Batch Data Analytics Solutions on AWS?

Nexus Human is recognized as one of the leading training providers. Our trainers have won multiple awards including:

- Small Firms Best Trainer Award
- National Training Partner of the Year (Ireland) - Multiple Years
- Global Top 30 Instructor Awards (2012, 2019, 2021)
- Tech Excellence Award Nominations
- Learning Performance Institute (LPI) External Training Provider Sponsor 2024

Q: Are there any discount codes available?

Yes! Use discount code **PENPAL5** when booking your Building Batch Data Analytics Solutions on AWS training. Please note that only one discount code can be used per booking and cannot be combined with other special offers.

Nexus Human

Professional Training & Development

Email: info@nexushuman.com

Website: www.nexushuman.com

Phone: +353 1 XXX XXXX (Ireland) | +44 20 XXXX XXXX (UK)