



Information Technology

Cloudera Analyst and Engineer

Course Introduction

This training course is designed to provide participants with the intricacies of the Cloudera platform, mastering the art of data processing, management, analysis, and visualization. From understanding the fundamental components of Cloudera's architecture to wielding advanced data engineering and machine learning techniques.

Throughout this course, participants will embark on a hands-on exploration of various tools and technologies that power the Cloudera ecosystem. From leveraging Hadoop Distributed File System (HDFS) for effective data storage to diving into real-time data processing with Apache Spark, and from querying data with Cloudera Impala to creating interactive visualizations with Apache Zeppelin, participant will not only gain theoretical insights but also practical expertise to navigate the world of big data. Whether seeking to deepen analytical capabilities or expand engineering proficiency, this course is the gateway to becoming a skilled Cloudera Analyst & Engineer in the ever-evolving landscape of big data.

Target Audience

- Cloud Computing Engineer
- Computer Network Specialist
- Computer Support Specialist
- Database Administrator
- Information Technology Analyst
- Information Technology Leadership
- Information Security Specialist
- Software/Application Developer
- Web Developer
- Technology sales consultant

Learning Objectives

- Acquire a thorough understanding of the Cloudera big data ecosystem, including its components, architecture, and role within the broader big data landscape.
- Develop proficiency in various data processing techniques, from batch processing using MapReduce to real-time data manipulation with Apache Spark, enabling effective analysis of large datasets.
- Gain expertise in data storage, management, and querying using technologies such as HDFS, Hive, and HBase, enabling efficient data organization and retrieval.
- Learn to perform comprehensive data analysis, from exploratory data techniques to using tools like Cloudera Impala and Apache Zeppelin for querying, analysis, and interactive data visualization.
- Develop advanced capabilities in machine learning within the Cloudera environment, along with proficiency in performance optimization and resource management, culminating in the application of learned skills through hands-on project work.

Course Outline

- **Day 01**

- Introduction to Cloudera and Big Data Ecosystem**

- Overview of Cloudera's role in the big data landscape.
 - Understanding the Cloudera distribution and its components.

- Introduction to Big Data Ecosystem**

- Overview of the Hadoop ecosystem and its components.
 - Understanding the role of Hadoop in big data processing.

Introduction to Data Analysis and Engineering

- Understanding the role of data analysis and engineering in a big data environment.
- Exploring common use cases and challenges.

• Day 02

Data Ingestion and Processing

- Data Ingestion Techniques
- Exploring various data ingestion methods, including batch and real-time.
- Using tools like Apache Flume and Kafka for data ingestion.

Data Processing with Hadoop

- Introduction to MapReduce and its role in processing large datasets.
- Hands-on exercises on writing MapReduce jobs.

Data Processing with Apache Spark

- Overview of Apache Spark and its advantages.
- Hands-on exercises on writing Spark applications.

• Day 03

Data Storage and Management

- HDFS and Data Storage
- Understanding Hadoop Distributed File System (HDFS) and its architecture.
- Exploring data storage and replication strategies.

Data Management with Hive

- Introduction to Hive and its role in data warehousing.
- Writing Hive queries for data analysis.

Data Management with HBase

- Overview of HBase as a NoSQL database.
- Understanding columnar storage and its benefits.

• Day 04

Data Analysis and Visualization

- Introduction to Data Analysis
- Exploring exploratory data analysis (EDA) techniques.
- Understanding data cleaning, transformation, and aggregation.

Data Analysis with Impala

- Introduction to Cloudera Impala for interactive querying.
- Writing SQL queries for data analysis.

Data Visualization with Apache Zeppelin

- Overview of Apache Zeppelin as a data visualization tool.
- Creating interactive data visualizations and dashboards.

• Day 05

Advanced Topics and Project Work

- Machine Learning with Cloudera

- Introduction to machine learning in Cloudera environment.
- Exploring MLlib for scalable machine learning.

Performance Tuning and Optimization

- Techniques for optimizing data processing and storage.
- Understanding cluster configuration and resource management.

Final Project and Capstone

- Collaborative project work applying concepts learned throughout the course.
- Presentations and discussions on project outcomes.

Confirmed Sessions

FROM	TO	DURATION	FEES	LOCATION
Dec. 12, 2025	Dec. 16, 2025	5 days	4250.00 \$	UAE - Dubai