Global Horizon
TRAINING CENTER

Training Course:
Cloudera Certified Associate (CCA) Data Analyst

7 - 18 December 2026
Kuala Lumpur (Malaysia)

Training Course:
Cloudera Certified Associate (CCA) Data Analyst

Training Course code: IT234731 From: 7 - 18 December 2026 Venue: Kuala Lumpur (Malaysia) - Training Course Fees: 10300 Euro

Introduction

This Certified Associate CCA Data Analyst Training course will teach you to apply traditional data analytics and business intelligence skills to big data. This course presents the tools data professionals need to access, manipulate, transform, and analyze complex data sets using SQL and familiar scripting languages. Advance Your Ecosystem Expertise Apache Hive makes transformation and analysis of complex, multi-structured data scalable in Cloudera environments. Apache Impala enables real-time interactive analysis of the data stored in Hadoop using a native SQL environment. Together, they make multi-structured data accessible to analysts, database administrators, and others without Java programming expertise.

Course Objectives

Throughout this training course participants will get to know:

- How the open source ecosystem of big data tools addresses challenges not met by traditional RDBMSs
- Using Apache Hive and Apache Impala to provide SQL access to data
- Hive and Impala syntax and data formats, including functions and subqueries
- Create, modify, and delete tables, views, and databases; load data; and store results of queries
- Create and use partitions and dierent file formats
- Combining two or more datasets using JOIN or UNION, as appropriate
- What analytic and windowing functions are, and how to use them Store and query complex or nested data structures?
- Process and analyze semi-structured and unstructured data
- Techniques for optimizing Hive and Impala queries
- Extending the capabilities of Hive and Impala using parameters, custom file formats and SerDes, and external scripts
- How to determine whether Hive, Impala, an RDBMS, or a mix of these is best for a given task?

Target Audience

This course is designed for:

- Data analysts
- Business intelligence specialists
- Developers
- Aystem architects
- Database administrators

Course Outline

Day 1: Apache Hadoop Fundamentals

- The Motivation for Hadoop
- Hadoop Overview
- Data Storage: HDFS Distributed Data Processing: YARN, MapReduce, and Spark
- Data Processing and Analysis: Hive, and Impala
- Database Integration: Sqoop _Other Hadoop Data Tools
- Exercise Scenario Explanation

Day 2: Introduction to Apache Hive and Impala

- What Is Hive?
- What Is Impala?
- Why Use Hive and Impala?
- Schema and Data Storage Comparing Hive and Impala to Traditional Databases
- Use Cases

Day 3: Querying with Apache Hive and Impala

- Databases and Tables Basic Hive and Impala Query Language Syntax
- Data Types
- Using Hue to Execute Queries
- Using Beeline Hive's Shell
- Using the Impala Shell

Day 4:

Common Operators and Built-In Functions

- Operators
- Scalar Functions
- Aggregate Functions

Data Management

- Data Storage
- Creating Databases and Tables
- Loading Data
- Altering Databases and Tables
- Simplifying Queries with Views
- Storing Query Results

Day 5: Data Storage and Performance

- Partitioning Tables
- Loading Data into Partitioned Tables
- When to Use Partitioning
- Choosing a File Format
- Using Avro and Parquet File Formats

Day 6:

Working with Multiple Datasets

- UNION and Joins
- Handling NULL Values in Joins
- Advanced Joins

Analytic Functions and Windowing

- Using Common Analytic Functions
- Other Analytic Functions
- Sliding Windows

Day 7:

Complex Data

- Complex Data with Hive
- Complex Data with Impala

Analyzing Text

- Using Regular Expressions with Hive and Impala
- Processing Text Data with SerDes in Hive
- Sentiment Analysis and n-grams

Day 8:

Apache Hive Optimization

- Understanding Query Performance
- Bucketing
- Hive on Spark

Apache Impala Optimization

- How Impala Executes Queries
- Improving Impala Performance

Day 9: Extending Apache Hive and Impala

- Extending Apache Hive and Impala
- Custom SerDes and File Formats in Hive
- Data Transformation with Custom Scripts in Hive

Day 10: Advanced Querying & Tool Selection

- User-Defined Functions UDFs

- Parameterized Queries
- Choosing the Best Tool for the Job
- Comparing Hive, Impala, and Relational Databases
- Which Tool to Choose?

Registration form on the Training Course: Cloudera Certified Associate (CCA) Data Analyst

Training Course code: IT234731 From: 7 - 18 December 2026 Venue: Kuala Lumpur (Malaysia) - Training Course Fees: 10300 Euro

Complete & Mail or fax to Global Horizon Training Center (GHTC) at the address given below

Delegate Information

Full Name (Mr / Ms / Dr / Eng): .....
Position: .....
Telephone / Mobile: .....
Personal E-Mail: .....
Official E-Mail: .....

Company Information

Company Name: .....
Address: .....
City / Country: .....

Person Responsible for Training and Development

Full Name (Mr / Ms / Dr / Eng): .....
Position: .....
Telephone / Mobile: .....
Personal E-Mail: .....
Official E-Mail: .....

Payment Method

☐ Please find enclosed a cheque made payable to Global Horizon

☐ Please invoice me

☐ Please invoice my company

Easy Ways To Register

Telephone: +201095004484 to provisionally reserve your place.	Fax your completed registration form to: +20233379764	E-mail to us : info@gh4t.com or training@gh4t.com	Complete & return the booking form with cheque to:Global Horizon 3 Oudai street, Aldouki, Giza, Giza Governorate, Egypt.