DANA-262

Analyzing with Cloudera Data Warehouse

Course Overview

Course Type
Instructor-Led Training

Level
Intermediate

Duration
4 days

Platform
Cloudera on premises
Cloudera on cloud

Topics Covered

- Apache Hive
- Apache Impala

About This Training

This four-day Analyzing with Data Warehouse 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.

What Skills You Will Gain

Through instructor-led discussion and interactive, hands-on exercises, participants will navigate the ecosystem, learning how to:

- Use Apache Hive and Apache Impala to access data through queries
- Identify distinctions between Hive and Impala, such as differences in syntax, data formats, and supported features
- Write and execute queries that use functions, aggregate functions, and subqueries
- Use joins and unions to combine datasets
- Create, modify, and delete tables, views, and databases
- Load data into tables and store query results
- Select file formats and develop partitioning schemes for better performance
- Use analytic and windowing functions to gain insight into their data
- Store and query complex or nested data structures
- Process and analyze semi-structured and unstructured data
- Optimize and extend the capabilities of Hive and Impala
- Determine whether Hive, Impala, an RDBMS, or a mix of these is the best choice for a given task
- Utilize the benefits of Cloudera Data Warehouse

Who Should Take This Course?

This course is designed for data analysts, business intelligence specialists, developers, system architects, and database administrators. Some knowledge of SQL is assumed, as is basic Linux command-line familiarity.

DANA-262

Analyzing with Cloudera Data Warehouse

Foundations for Big Data Analytics

- Big Data Analytics Overview
- Data Storage: HDFS
- Distributed Data Processing: YARN, MapReduce, and Spark
- Data Processing and Analysis: Hive and Impala
- Database Integration: Sqoop
- Other Data Tools
- Exercise Scenario Explanation

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

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

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

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

Working with Multiple Datasets

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

Analytic Functions and Windowing

- Using Analytic Functions
- Other Analytic Functions
- Sliding Windows

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 in Hive

VERSION 2.0.0

DANA-262

Analyzing with Cloudera Data Warehouse

Apache Hive Optimization

- Understanding Query Performance
- Cost-Based Optimization and Statistics
- Bucketing
- ORC File Optimizations

Apache Impala Optimization

- How Impala Executes Queries
- Improving Impala Performance

Extending Hive and Impala

- User-Defined Functions
- Parameterized Queries

Choosing the Best Tool for the Job

- Comparing Hive, Impala, and Relational Databases
- Which to Choose?

Cloudera Data Warehouse

- Data Warehouse Overview
- Auto-Scaling
- Managing Virtual Warehouses
- Querying Data Using CLI and Third-Party Integration

Appendix: Apache Kudu

- What Is Kudu?
- Kudu Tables
- Using Impala with Kudu