Hive – Course Outline

1 Duration

- 20 hours (5 half days)

2 Objectives

At end of this workshop, participants will get to know about:

- Hive architecture, components, anatomy and features
- Hive Programming - Shell, Data Types, Data Models, Views and HiveQL
- Design and build Hive queries to load, query and analyze the dataset
- Understand and use Hive built in functions / UDFs to do validation, transformation and aggregation
- Monitoring, Troubleshooting, Performance Tuning and Best Practices

3 Audience

Data Analysts, Data Science Professionals and Big Data Professionals

4 Pre-requisite

- Basic knowledge on Big Data and Hadoop File System
- Good knowledge on SQL

5 Hardware & Network Requirements

- Desktop with minimum 8GB RAM
- Internet connection (to access AWS cloud labs)

6 Software Requirements

- Windows / Linux / Mac OS
- Oracle VirtualBox (to practice in local environment after the training)
- Cloud Labs

7 Outline

7.1 Day 1

1) Introduction to BigData and Hive

a) Big Data Overview
b) Hadoop Ecosystem Overview
c) HDFS Overview
d) MR Overview
e) Need for Hive
f) Hive Characteristics
g) Hive Architecture
h) Hive Components
i) Hive Features
j) Hive Limitations

2) Hive Setup and Configuration

a) Hive Command Line Interfaces
b) Hive Commands - Interactive and Batch Mode
c) Hive Clients
d) Hive Configuration Files
e) Access Hive using Hue
f) Hands on lab exercises

7.2 Day 2

3) Hive Basic Concepts

a) Hive Data Model
b) Managed Tables, External Tables, Clustered Tables
c) Partitioning, Clustering
d) Views and Indexes
e) File Formats
f) Compression Techniques

4) Hive SQL - Basics

a) Keywords
b) Data Types
c) Operators
d) Handling NULL values
e) Functions
f) Hands on lab exercises

7.3 Day 3

5) Hive SQL - DDL Operations

a) Create, Drop, Alter, Truncate
b) Show, Describe
c) Hands on lab exercises

6) Hive SQL - DML Operations

d) Load, Insert, Update, Delete
e) Import, Export from/into files or tables
f) Hands on lab exercises

7.4 Day 4

7) Hive QL - Queries

g) Select queries
h) Conditional queries
i) Partition queries
j) Grouping and Aggregation
k) Sorting, Ordering, Clustering, Distributing
l) Union, Joins, Subqueries
m) Hands on lab exercises

7.5 Day 5

8) Hive QL - Advanced

a) Built-in functions
b) Built-in Aggregate Functions (UDAF)
c) Built-in Table Generating Functions (UDTF)
d) User Defined Functions (UDF)
e) Sampling, Virtual Columns, Lateral View
f) Windowing, OVER and Analytics, Common Table Expressions
g) Transactions, Counters
h) Indexes, Statistics/Analyze, Locks
i) Explain Plan, Authorization, Archiving
j) Hands on lab exercises