

## **CLUDERA APACHE HIVE AND PIG COURSE CONTENT**

### ❖ **INTRODUCTION TO HADOOP AND HIVE:**

- What Is Hadoop?
- The Motivation for Hive

### ❖ **GETTING DATA INTO HIVE:**

- Accessing Hive
- The Hive Architecture
- Creating Hive Tables
- Loading Data into Hive
- Storing Query Output
- Creating Different Databases

### ❖ **MANIPULATING DATA WITH HIVE:**

- Retrieving Data with the SELECT Command
- Joining Tables
- Storing Query Results in HDFS
- Basic Hive Functions
- More Advanced HiveQL Queries
- Statistics and Data Mining

### ❖ **PARTITIONING AND BUCKETING DATA:**

- Partitioning Data
- Bucketing Data

### ❖ **ADVANCED HIVE FEATURES:**

- Hive Variables
- The Hive CLI
- Hive and Thrift
- TRANSFORM
- Creating User-Defined Functions and SerDes
- Debugging and Troubleshooting Hive Queries

### ❖ **HIVE BEST PRACTICES:**

- Configuring a Shared Metastore
- Hive Authorization
- Handling Dates
- Dealing with SerDes

### ❖ **INTRODUCTION TO PIG:**

- What Is Pig?
- Pig Use Cases
- Hive vs. Pig

### ❖ **PIG'S ARCHITECTURE:**

- How Pig Works
- Installing and Configuring Pig
- Ways of Executing Pig Programs

❖ **READING AND WRITING DATA WITH PIG:**

- Loading Data
- Referencing and Defining Data Types
- Dumping and Storing Results

❖ **PIG LATIN BASICS:**

- What Is Pig Latin?
- Filtering, Grouping and Sorting Data
- Comments and Case-Sensitivity

❖ **ADVANCED PIG LATIN:**

- Advanced Features of FOREACH
- Join Operations
- Functions, UDFs and Macros

❖ **DEBUGGING PIG SCRIPTS:**

- Diagnostic Operators
- Collecting Statistics
- Pig Unit
- JobTracker

❖ **PIG BEST PRACTICES::**

- Combiner
- Multi-Query Execution
- Performance Enhancers
- Join Optimizations

❖ **BRINGING IT ALL TOGETHER:**

- Hive, Pig and MapReduce

❖ **APPENDIX: HIVE AND HBASE:**

- What Is HBase?
- Hive vs. HBase
- Use Cases
- Integrating Hive and HBase

*DES*

*Dyce Consulting Services*