

Cloudera Developer for Apache Hadoop Course Content

1. MOTIVATION FOR HADOOP

- Problems with Traditional Large-Scale Systems
- Requirements for a New Approach

2. HADOOP: BASIC CONCEPTS

- Hadoop Distributed File System (HDFS)
- MapReduce
- Anatomy of a Hadoop Cluster
- Other Hadoop Ecosystem Components

3. WRITING A MAPREDUCE PROGRAM

- MapReduce Flow
- Examining a Sample MapReduce Program
- Basic MapReduce API Concepts
- Driver Code
- Mapper
- Reducer
- Streaming API
- Using Eclipse for Rapid Development
- New MapReduce API

4. INTEGRATING HADOOP INTO THE WORKFLOW

- Relational Database Management Systems
- Storage Systems
- Importing Data from a Relational Database Management System with Sqoop
- Importing Real-Time Data with Flume
- Accessing HDFS Using FuseDFS and Hoop

5. DELVING DEEPER INTO THE HADOOP API

- ToolRunner
- Testing with MRUnit
- Reducing Intermediate Data with Combiners
- Configuration and Close Methods for Map/Reduce Setup and Teardown
- Writing Partitioners for Better Load Balancing
- Directly Accessing HDFS
- Using the Distributed Cache

6. COMMON MAPREDUCE ALGORITHMS

- Sorting and Searching
- Indexing
- Machine Learning with Mahout
- Term Frequency
- Inverse Document Frequency
- Word Co-Occurrence

7. USING HIVE AND PIG

- Hive Basics
- Pig Basics

8. PRACTICAL DEVELOPMENT TIPS AND TECHNIQUES

- Debugging Map Reduce Code
- Using LocalJobRunner Mode for Easier Debugging
- Retrieving Job Information with Counters
- Logging
- Splittable File Formats
- Determining the Optimal Number of Reducers
- Map-Only MapReduce Jobs

9. ADVANCED MAPREDUCE PROGRAMMING

- Custom Writables and WritableComparables
- Saving Binary Data Using SequenceFiles and Avro Files
- Creating InputFormats and OutputFormats

10. JOINING DATA SETS IN MAPREDUCE

- Map-Side Joins
- Secondary Sort
- Reduce-Side Joins

11. GRAPH MANIPULATION IN HADOOP

- Graph Techniques

- Representing Graphs in Hadoop
- Implementing a Sample Algorithm: Single Source Shortest Path
-

12. CREATING WORKFLOWS WITH OOZIE

- Motivation for Oozie
- Workflow Definition Format

Should you require any further information regarding [Cloudera Developer Training](#), Demo Sessions, Price and any other information. Please contact us through email or phone.

Thanks & Regards

Srinivas

Glory IT Technologies

Email : info@gloryittechnologies.com

Mobile (IND): +91-9032813456 / 91-9160177789

Skype ID : gloryittechnologies