

## Become a Cloudera Certified Developer for Apache Hadoop

### Why Big Data and Hadoop 2.0

The use of Big Data is becoming a crucial way for leading companies to outperform their peers. It enables us to tap into large information flow where data about products and services, buyers and suppliers, consumer preferences and intents can be captured and analyzed. Therefore nurturing Big Data Analytics expertise plays an essential role in developing new products or services.

### Who should attend this Hadoop 2.0 training

This training is ideal for professionals who wish to implement Big Data Analytics by using Hadoop framework.

### What you will learn on Hadoop 2.0

The course equips participants to work on the Hadoop environment with ease and learn vital components such as Flume, Hive, Pig, Map Reduce and other advanced concepts like Hadoop 2.0: HBase, ZooKeeper and Sqoop.

The course consists of:

- 3 Days Classroom training
- 20 Hours of Real Time Industry based Projects
- 15 Hours of Lab Exercises with proprietary VM
- 3 Hours of Doubt Clarification Session
- 2 Projects with Unique Data Sets Included
- Industry Specific Projects on the Retail and Telecom Sectors
- 1 Big Data & Hadoop Simulation Exam
- Downloadable e-Book Included
- Java Essentials for Hadoop Included
- Hadoop Installation Procedure Included
- Excellence in Hadoop Certificate
- Hadoop Deployment and Maintenance Tips
- Packed with Latest & Advanced modules like Yarn, Flume, Oozie, Mahout, and Chukwa
- Tips and Techniques to clear Certification Exam by the Trainer

### How you will be certified

Certificate 1: Trainer will provide guidance by going through the practice test and answer with explanations. Participant will sit for web-based CCDH (Cloudera Certified Developer for Apache Hadoop) examination and receive a certificate upon scoring 80% or above in the examination.

Certificate 2: Bring back what you have learnt in the 3-day course and implement the Hadoop ecosystem into your product or services by completing a Hadoop project. Trainer will assist and provide guidance via web-based correspondence. Participant will receive another certificate upon completion of project.

Date: 27-29 Jul 2015  
(Monday - Wednesday)  
Time: 0900 - 1700  
Venue: Suite 2B-21-1, Level 21, Block 2B,  
Plaza Sentral, Jalan Stesen Sentral 5,  
KL Sentral, 50470 Kuala Lumpur,  
Malaysia.

**HRDF Claimable**

## Course Outline

### Day 1

#### Introduction to Big Data

- The driving force of the Big Data era
- Why Big Data deserves your attention
- Use cases of Big Data
- Different options of analyzing Big Data
- Introduction to Hadoop
- History of Hadoop
- Problems with traditional large-scale systems and need for Hadoop
- Where Hadoop is being used
- Understanding distributed systems

#### RDBMS and Hadoop

#### Understanding Hadoop Architecture

- Understanding Hadoop configuration files
- Hadoop components- HDFS, MapReduce
- Overview of Hadoop processes
- Overview of Hadoop Distributed File System
- NameNode
- DataNode
- The Command-Line Interface

#### Hands-On Exercise

- Using HDFS commands

### Day 2

#### Introduction to Apache Pig

- Pig architecture
- Data types, Arithmetic, Relational Operators, Pig UDF
- Pig built in functions: Eval, Math, String, Date, Time functions
- Pig assignment

#### Introduction to Apache Hive

- Hive architecture
- Hive Metastore configuration
- Data types of file formats and Hive built in/basic functions
- Databases in Hive
- External and internal tables, and loading data into Hive

#### Hands-on Exercise

- Pig and Hive

### Day 3

#### Hive

- Bucketing and partitions
- HiveQL queries: Select, Where, Group By, Having, Distributed By with Sort By, Cluster By, Order By vs Sort By
- Hive UDF

#### MapReduce Program In Java using Eclipse

#### No-SQL Introduction

- Hive
- Differences between SQL and No-SQL
- HBase architecture
- ZooKeeper introduction

#### Hands-on Exercise

- HBase
- Sqoop and Flume, HBase
- Hadoop Admin - Hadoop Installation

## About Go Training

Go Training applies effective pedagogical methodologies that demonstrate case studies and hands-on practical skills, in addition to explaining clearly how things work in principle. Every course that we conduct is delivered by a subject matter expert who holds the academic qualification and working experience in that specialization. On the days when they are not teaching, our trainers work on consultancy projects and technical deliveries. Their work has received numerous recognition and awards in the industry. Our team of trainers has been invited as keynote speakers at numerous international conferences, and as principal consultants for various industries.

## About the Instructor

Tirumala Rao Dumpala is the Senior IT Consultant who has 7 years of versatile experience in designing, developing and supporting IT applications behind him, which includes three and a half years in Big Data, Text Mining, Analytics, java, Data Mining and Web Development. He graduated with a degree of Master of computer Applications at JNTU. He is currently affiliated with Cap Gemini (HP Client), dealing with HP Real time Analytics and Data processing project with live data volumes around 0.5 TB to 2 TB. His previous working experience includes also position with held with HCL, dealing with the Aadhaar project which developed the unique identification number generation and authentication system. He was also the technical lead of TCS, he was involved in the development of the Business Intelligence and Reporting system which can handle 2 TB to 22 TB of data.

Go Training  
wholly owned by iRadar Sdn Bhd  
HRDF Approved Training Provider (Category A)

No. 36, Jalan IMJ 1, Taman Industri Malim  
Jaya, 75250 Melaka, Malaysia.  
t +606 336 6016  
f +606 252 3059  
w [www.gotraining.com.my](http://www.gotraining.com.my)  
[f] [fb.com/gotraining.com.my](https://fb.com/gotraining.com.my)  
[in] [linkedin.com/company/gotraining](https://linkedin.com/company/gotraining)

To register, please contact:  
m +6010 663 1852  
e [yiwei@gotraining.com.my](mailto:yiwei@gotraining.com.my)