

# Quản trị hệ thống Big Data

## I. Tổng quan

Khóa học cung cấp các kiến thức và kỹ năng cần thiết trong việc cài đặt, triển khai, cấu hình các công cụ phục vụ cho việc phát triển ứng dụng Big Data bao gồm: Hadoop, HDFS, MapReduce, Hive, ZooKeeper, HBase và một số công cụ khác liên quan như: Oozie, Sqoop/flume, Hortonworks, Ranger

## II. Thời lượng: 32 giờ.

## III. Đối tượng học:

- Quản trị hệ thống
- Trưởng phòng hệ thống.

## IV. Yêu cầu trước khóa học

- Có kiến thức cơ bản về Linux
- Không cần các kiến thức về Hadoop và các công cụ Big Data liên quan.

## V. Nội dung:

### 1. *Introduction to Big Data*

- Introduction to Big Data
- Introduction to Hadoop
- Introduction to HDFS
- Introduction to Map-Reduce
- Introduction to YARN

### 2. *Hadoop installation & configuration*

- Hadoop Architecture
- Hadoop Configuration
- Hadoop Cluster Setup and Installation
- Hadoop Data Management
- Advanced Hadoop Operations
- Rack Awareness

### 3. **HDFS**

- HDFS Components
- HDFS Design/Architecture
- HDFS read/write operations
- Distcp
- HDFS CLI
- Commissioning and Decommissioning of nodes
- HDFS HA

**4. YARN vs MR**

- MR & YARN overview
- YARN Architecture
- Advantages over MR
- YARN Job flow
- YARN Components YARN Schedulers YARN log files

**5. Zookeeper**

- Zookeeper Introduction
- Zookeeper Installation.

**6. HIVE**

- Hive Overview
- Hive components
- Hive installation.

**7. HBASE**

- HBase Overview
- Difference between Hive and Hbase
- Hbase installation.m

**8. Other tools**

- Oozie Introduction
- Sqoop/flume introduction
- Hortonworks introduction & installation
- Ranger Introduction
- Kerberos Authentication in Hadoop.