



[www.uplatz.com](http://www.uplatz.com)

Leading Marketplace for IT and Certification Courses

[info@uplatz.com](mailto:info@uplatz.com)

+44 7836 212635

# HADOOP ADMINISTRATION TRAINING



◆ 1:1 Instructor-led Online Training ◆



[www.uplatz.com](http://www.uplatz.com)



+44 7836212635



[info@uplatz.com](mailto:info@uplatz.com)

## COURSE DETAILS & CURRICULUM

HADOOP ADMINISTRATION TRAINING CURRICULUM

### 1 INTRODUCTION

#### 1.1 Big Data Introduction

##### 1.1.1 What is Big Data?

##### 1.1.2 Big Data - Why

##### 1.1.3 Big Data - Journey

##### 1.1.4 Big Data Statistics

##### 1.1.5 Big Data Analytics

##### 1.1.6 Big Data Challenges

##### 1.1.7 Technologies Supported By Big Data

#### 1.2 Hadoop Introduction

##### 1.2.1 What Is Hadoop?

##### 1.2.2 History Of Hadoop

##### 1.2.3 Breakthroughs Of Hadoop



[www.uplatz.com](http://www.uplatz.com)

Leading Marketplace for IT and Certification Courses

[info@uplatz.com](mailto:info@uplatz.com)

[+44 7836 212635](tel:+447836212635)

1.2.4 Future of Hadoop

1.2.5 Who Is Using?

1.3 Basic Concepts

1.3.1 The Hadoop Distributed File System - At a Glance

1.3.2 Hadoop Daemon Processes

1.3.3 Anatomy Of A Hadoop Cluster

1.3.4 Hadoop Distributions

## **2 HADOOP DISTRIBUTED FILE SYSTEM (HDFS)**

2.1 What is HDFS?

2.1.1 Distributed File System (DFS)

2.1.2 Hadoop Distributed File System (HDFS)

2.2 HDFS Cluster Architecture and Block Placement

2.2.1 NameNode

2.2.2 DataNode

2.2.3 JobTracker

2.2.4 TaskTracker

2.2.5 Secondary NameNode

2.3 HDFS Concepts

2.3.1 Typical Workflow

2.3.2 Data Replication

2.3.3 Replica Placement

2.3.4 Replication Policy

2.3.5 Hadoop Rack Awareness

2.3.6 Anatomy of a File Read

2.3.7 Anatomy of a File Write

## **3. MAPREDUCE**

3.1 STAGES OF MAPREDUCE



[www.uplatz.com](http://www.uplatz.com)

Leading Marketplace for IT and Certification Courses

[info@uplatz.com](mailto:info@uplatz.com)

[+44 7836 212635](tel:+447836212635)

### 3.2 DAEMONS

#### 3.2.1 Job Tracker

#### 3.2.2 Task Tracker

### 3.3 TASK FAILURES

#### 3.3.1 Child

#### 3.3.2 Task Tracker Failures

#### 3.3.3 Job Tracker Failures

#### 3.3.4 HDFS Failures

### 3.4 YARN

## 4. HOW TO PLAN A CLUSTER

### 4.1 VERSIONS AND FEATURES

### 4.2 HARDWARE SELECTION

#### 4.2.1 Master Hardware

#### 4.2.2 Slave Hardware

#### 4.2.3 Cluster sizing

### 4.3 OPERATING SYSTEM SELECTION

#### 4.3.1 Deployment Layout

#### 4.3.2 Software Packages

#### 4.3.3 Hostname, DNS

#### 4.3.4 Users, Groups, Privileges

### 4.4 DISK CONFIGURATION

#### 4.4.1 Choose a FileSystem

#### 4.4.2 Mount options

### 4.5 NETWORK DESIGN



[www.uplatz.com](http://www.uplatz.com)

Leading Marketplace for IT and Certification Courses

[info@uplatz.com](mailto:info@uplatz.com)

[+44 7836 212635](tel:+447836212635)

- 4.5.1 Network usage in Hadoop
- 4.5.2 Typical network Topologies

## **5. INSTALLATION AND CONFIGURATION**

### 5.1 APACHE HADOOP

- 5.1.1 Tarball Installation
- 5.1.2 Package Installation

### 5.2 CONFIGURATION

- 5.2.1 XML Configuration
- 5.2.2 Environment Variables
- 5.2.3 Logging Configuration

### 5.3 HDFS

- 5.3.1 Optimization and Tuning

### 5.4 MAPREDUCE

- 5.4.1 Optimization and Tuning

## **6. AUTHENTICATION**

### 6.1 KERBEROS AND HADOOP

- 6.1.1 Kerberos
- 6.1.2 Configuring Hadoop Security

## **7. RESOURCE MANAGEMENT**

### 7.1 WHAT IS RESOURCE MANAGEMENT?

### 7.2 MAPREDUCE SCHEDULER

- 7.2.1 Capacity Scheduler
- 7.2.2 Fair Scheduler



[www.uplatz.com](http://www.uplatz.com)

Leading Marketplace for IT and Certification Courses

[info@uplatz.com](mailto:info@uplatz.com)

[+44 7836 212635](tel:+447836212635)

## **8. CLUSTER MAINTENANCE**

### 8.1 MANAGING HADOOP PROCESS

8.1.1 Starting and stopping processes with Init scripts

8.1.2 Starting and stopping processes manually

### 8.2 HDFS MAINTENANCE

8.2.1 Adding and Decommissioning DataNode

8.2.2 Balancing HDFS Block Data

8.2.3 Dealing with a Failed disk

### 8.3 MAPREDUCE MAINTENANCE

8.3.1 Adding and Decommissioning TaskTracker

8.3.2 Kill MapReduce Job and Task

8.3.3 Dealing Blacklisted Tasktracker

## **9. TROUBLESHOOTING**

### 9.1 COMMON FAILURES AND PROBLEMS

### 9.2 HDFS AND MAPREDUCE CHECKS

## **10. BACKUP AND RECOVERY**

### 10.1 DATA BACKUP

10.1.1 Distributed copy

10.1.2 Parallel data ingestion

### 10.2 NAMENODE METADATA



[www.uplatz.com](http://www.uplatz.com)

Leading Marketplace for IT and Certification Courses

[info@uplatz.com](mailto:info@uplatz.com)

[+44 7836 212635](tel:+447836212635)

#### **COURSE DELIVERABLES**

- Workshop style coaching
- Interactive approach
- Course material
- Hands on practice exercises
- Quiz at the end of each major topic
- Tips and techniques on Cloudera Certification Examination
- Mock interviews for each individual will be conducted on need basis
- Resume preparation and guidance