# HADOOP ADMINISTRATION

## Take your knowledge to the next level with Hadoop Training

The administrator training course for Apache Hadoop provides participants with a comprehensive understanding of all the steps necessary to operate and maintain a Hadoop cluster. From installation and configuration through load balancing and tuning, our training course is the best preparation for the real-world challenges faced by Hadoop administrators.

Administrators are leading the transition from traditional databases and Data Warehouses to more flexible, scalable systems built on Apache Hadoop.

## Hands-On Hadoop

Through instructor-led discussion and interactive, hands-on exercises, participants will navigate the Hadoop ecosystem, learning topics such as:

• Cloudera Manager features that make managing your clusters easier, such as aggregated logging, configuration management, resource management, reports, alerts, and service management.

• The internals of YARN, MapReduce, and HDFS

• Determining the correct hardware and infrastructure for your cluster

• Proper cluster configuration and deployment to integrate with the data canter

• How to load data into the cluster from dynamically-generated files using Flume and from RDBMS using Sqoop

• Configuring the Fair Scheduler to provide service-level agreements for multiple users of a cluster

• Best practices for preparing and maintaining Apache Hadoop in production

• Troubleshooting, diagnosing, tuning, and solving Hadoop issues

# Course outline: Hadoop Administration

## Introduction

### Apache Hadoop

- Why Hadoop?
- Fundamental concepts
- Core Hadoop components

### Prerequisites

- Infrastructure hardware
    - Server details
    - How to buy servers?
    - Resources for Hadoop
- Linux server OS
    - Linux boot process
    - P.O.S.T.
    - Flavours of Linux server
    - Versioning and development of server OS
    - Installing a server
    - OS file structure
    - Introduction to CLI
    - Overview of commands
    - SSH
- Networks for Hadoop
    - Introduction
    - Working of Hub, Bridge, Switch, Router
    - Classes of IP
    - DNS, NAT, ARP
    - Commands related to networks

### Cloud Computing

- Introduction
- Why cloud? On-premises vs Cloud
- Peak hours
- Cloud providers
- Enterprise cloud tie-ups
- Introduction to S3
- S3 – Deep dive
- Introduction to EC2

- EC2 – Deep dive

## Hadoop Administration

- Introduction to Big Data
- Definitions of Big Data
- Apache Hadoop 1 Single node installation
- Apache Hadoop 1 Multi node installation
- Apache Hadoop 2 Single node installation
- Apache Hadoop 2 Multi node installation
- Hadoop 2 installation through Ambari server
- Cloudera Deployment
- Benchmarking the cluster
- Introduction to EMR
- EMR – Deep dive
- Hadoop Architecture (Appliances)
- Read and Write in Hadoop
- Rack awareness
- SPOF and HA
- Hadoop architecture in production
- Map-Reduce architecture
- Hadoop Admin Tasks (Hands-on)
    - Monitoring jobs
    - Killing a job
    - Creating pipeline
    - File system check
    - Trash configuration
    - Checking disk usage
    - Commissioning and decommissioning
    - Balancer
    - Snapshots
    - Monitoring through Namenode and Resource manager webUI
- Distributed Copy
- Yarn Architecture
- Yarn Administration
- Performance tuning
- Security – ACL/SSL, Kerberos, Sentry
- Admin tasks on Cloudera
- Troubleshooting scenarios
- Common Errors

### Extra

- Interview preparation, Mock interviews
- Profile creation, Resume built-up by an industry expert in pune
- 200 interview question set.
- Certification guidance
- One real time project

### Conclusion

**Course Duration: 3 Months (Approx.)**

**Trainer – Aasem Quazi (Sr. Hadoop Admin)**

**+91-7776808458**

**LinkedIn profile –**

https://in.linkedin.com/in/aasem-quazi-92403aaa