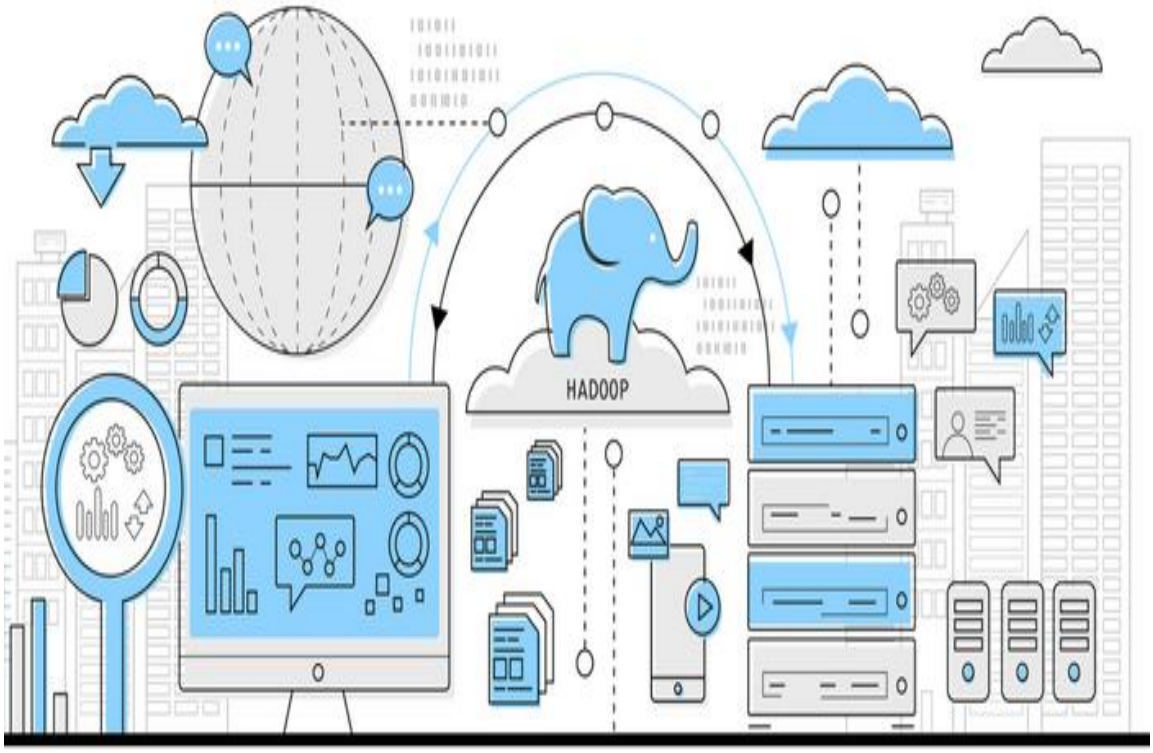
TechyEdz Solutions

Training | Consulting | Developement | Outsourcing



Hadoop Administration

9032803832

9032803832

contact@techyedz.com

www.techyedz.com

Big Data Hadoop Administration

Course Overview:

Hadoop Administration Certification Training provides you with proficiency in all the steps required to operate and sustain a Hadoop Cluster which includes Planning, Installation, and Configuration through load balancing, Security, and Tuning. In this training will provide hands-on preparation for the real-world challenges faced by Hadoop Administrators. The course curriculum follows Apache Hadoop distribution.

Course Outline:

Understanding Big Data and Hadoop

- Introduction to big data
- Common big data domain scenarios
- Limitations of traditional solutions
- What is Hadoop?
- Hadoop 1.0 ecosystem and its Core Components
- Hadoop 2.x ecosystem and its Core Components
- Application submission in YARN

Hadoop Cluster and its Architecture

- Hadoop Cluster Architecture
- Replication rules
- Hadoop Cluster Modes
- Rack awareness theory
- Hadoop cluster administrator responsibilities
- Understand working of HDFS

- NTP server
- Initial configuration required before installing Hadoop
- Deploying Hadoop in a pseudo-distributed mode

Hadoop Cluster Setup and Working

- OS Tuning for Hadoop Performance
- Pre-requisite for installing Hadoop
- Hadoop Configuration Files
- Stale Configuration
- RPC and HTTP Server Properties
- Properties of Namenode, Datanode and Secondary Namenode
- Log Files in Hadoop
- Deploying a multi-node Hadoop cluster

Hadoop Cluster Administration and Maintenance

- Commisioning and Decommissioning of Node
- HDFS Balancer
- Namenode Federation in Hadoop
- High Availabilty in Hadoop
- .Trash Functionality
- Checkpointing in Hadoop
- Distcp
- Disk balancer

Computational Frameworks, Managing Resources and Scheduling

- Different Processing Frameworks

- Different phases in Mapreduce
- Spark and its Features
- Application Workflow in YARN
- YARN Metrics
- YARN Capacity Scheduler and Fair Scheduler
- Service Level Authorization (SLA)

Hadoop 2.x Cluster: Planning and Management

- Planning a Hadoop 2.x cluster
- Cluster sizing
- Hardware, Network and Software considerations
- Popular Hadoop distributions
- Workload and usage patterns
- Industry recommendations

Hadoop Security and Cluster Monitoring

- Monitoring Hadoop Clusters
- Hadoop Security System Concepts
- Securing a Hadoop Cluster With Kerberos
- Common Misconfigurations
- Overview on Kerberos
- Checking log files to understand Hadoop clusters for troubleshooting

Cloudera Hadoop 2.x and its Features

- Visualize Cloudera Manager
- Features of Cloudera Manager
- Build Cloudera Hadoop cluster using CDH
- Installation choices in Cloudera

- Cloudera Manager Vocabulary
- Cloudera terminologies
- Different tabs in Cloudera Manager
- What is HUE?
- Hue Architecture
- Hue Interface
- Hue Features

Pig, Hive Installation and Working

- Explain Hive
- Hive Setup
- Hive Configuration
- Working with Hive
- Setting Hive in local and remote metastore mode
- Pig setup
- Working with Pig

HBase, Zookeeper Installation and Working (Self-paced)

- What is NoSQL Database
- HBase data model
- HBase Architecture
- MemStore, WAL, BlockCache
- HBase Hfile
- Compactions
- HBase Read and Write
- HBase balancer and hbck
- HBase setup

- Working with HBase
- Installing Zookeeper

Understanding Oozie

- Oozie overview
- Oozie Features
- Oozie workflow, coordinator and bundle
- Start, End and Error Node
- Action Node
- Join and Fork
- Decision Node
- Oozie CLI
- Install Oozie

Data Ingestion using Sqoop and Flume

- Types of Data Ingestion
- HDFS data loading commands
- Purpose and features of Sqoop
- Perform operations like, Sqoop Import, Export and Hive Import
- Sqoop 2
- Install Sqoop
- Import data from RDBMS into HDFS
- Flume features and architecture
- Types of flow
- Install Flume
- Ingest Data From External Sources With Flume
- Best Practices for Importing Data

Ambari

- Introduction to Ambari
- Installing and starting Ambari Server
- Configuring and Deploying the cluster
- Choosing and Customizing services
- Assigning Masters, Slaves and Clients
- Troubleshootig Ambari deployments

Ganglia

- Introduction to ganglia
- Components of Ganglia - Gmond, Gmetad, RRDtool
- Installation and Configuration - Gmond Configuration, Gmetad Configuration, PHP Web Frontend Configuration
- Setup Monitoring for Hadoop Cluster - Commandline Tools, Gmetric, Gstat
- How to automate deploys in your infrastructure

Amazon Web Services - AWS

- Introduction to AWS
- Different Instance types
- Get familiar with AWS
- Components of Hadoop on AWS
- Deploy Hadoop cluster on AWS
- Explore scalability options

Prerequisites:

- Basic knowledge of Linux command line interface will be considered beneficial

Who Can attend:

The market for Big Data analytics is constantly growing across the world and this strong growth pattern translates into a great opportunity for all the IT Professionals with the

required skills. Hadoop Admin Certification Training helps you to grab this opportunity and accelerate your career. It is best suited for:

- Linux / Unix Administrators
- Data Engineers & Database Administrators
- Windows Administrators
- Infrastructure Administrators
- System Administrators
- Data Analytics Administrators
- Cloud Systems Administrators
- Web Engineers
- Individuals who intend to design, deploy and maintain Hadoop clusters

Number of Hours: 40hrs

Certification: Cloudera Certified Administrator for Apache Hadoop (CCAH / CCA 410)

Key Features:

- One to One Training
- Online Training
- Fastrack & Normal Track
- Resume Modification
- Mock Interviews
- Video Tutorials
- Materials
- Real Time Projects
- Virtual Live Experience
- Preparing for Certification