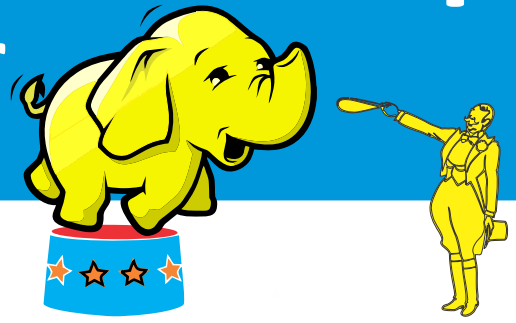


Certified Hadoop Administrator



Certificate :
Certified Hadoop Administrator

Duration :
4 Days of Instructor Led Course

Course Delivery :
Classroom or 32 Hrs of Virtual classroom.

Course Description :

Hadoop Administrator course will help IT administrators and IT managers to build, operate and manage large multi-clusters within an enterprise. The course will also provide the required knowledge to plan for capacity, fine tune and monitor hadoop clusters of huge scale with in hadoop eco-system. It covers topics to deploy, configure, manage, monitor, and secure a Hadoop Cluster. This training will be delivered by our BigData Experts with rich industry experience handling large big Data projects .There will be many hands-on sessions for the delegates . By the end of this Hadoop Cluster Administration training, you will be able to understand and solve real challenging problem that you may come across while working on Hadoop Cluster.

Program content :

- ❖ What is Data?
- ❖ Data Management
- ❖ What is Bigdata?
- ❖ Existing challenges and the solution
- ❖ What is Hadoop?
 - ♦ Hadoop - Java framework
 - ♦ Discuss various products developed by vendors
 - ♦ Apache - Origin
 - ♦ Cloudera
 - ♦ Hortonworks
- ❖ Hadoop supported platforms
 - ♦ Linux - Redhat, Centos, Ubuntu
 - ♦ Windows
- ❖ Hadoop components
 - ♦ HDFS
 - ♦ Master-Slave arch
 - ♦ Mapreduce
 - ♦ Master-Slave arch
- ❖ Hadoop modes
 - ♦ Standalone / LocalJobRunner
 - ♦ Pseudo-distributed
 - ♦ Fully distributed
- ❖ Hadoop Management UI
 - ♦ Cloudera Manager - Cloudera
 - ♦ Ambari - Hortonworks
 - ♦ Intel manager - Intel

For More Details:

QT&T Consulting (Asia) Pte Ltd
Level 2 , #02-06, 2 Changi Business Park Avenue 1 , Singapore - 486015

Tel: +65 68968552
Email: sales@qtnt.com
Web : www.qtnt.com

- ❖ Hadoop - Pre-requisite
 - ♦ Java must
 - ♦ Selinux Disabled
 - ♦ Iptables [Enabled/Disabled]
 - ♦ Hardening
- ❖ Hadoop configuration files
 - ♦ Core properties: core-site.xml
 - ♦ HDFS properties: hdfs-site.xml
 - ♦ MapReduce properties: mapred-site.xml
 - ♦ Hadoop Environment variables: Hadoop-env.sh
- ❖ Practical - Lab I
 - ♦ Setting up VM
 - ♦ OS installation
 - ♦ Downloading CM and CDH Packages
 - ♦ CM - cloudera-manager-installer.bin
 - ♦ CDH - Internet/Local
- ❖ Cloudera Manager - Hadoop install options
 - ♦ Tarball
 - ♦ RPM
 - ♦ Parcels
- ❖ Create local repository - CM and CDH
- ❖ Hadoop Installation
 - ♦ Pseudo-distributed
 - ♦ Fully distributed
- ❖ Hadoop Architecture Principles
- ❖ Hadoop Components [in detail]
 - ♦ HDFS
 - ♦ HDFS features
 - ♦ HDFS Master-Slave architecture
 - ♦ HDFS Block Diagram
- ❖ HDFS daemons
 - ♦ Namenode / Master node
 - ♦ Secondary Namenode / Checkpoint node
 - ♦ Datanodes / Slave nodes
 - ♦ HDFS client - Gateway
- ❖ Anatomy of File Write
- ❖ Anatomy of File Read
- ❖ Rack-awareness
- ❖ Block Replication Strategy
- ❖ Dealing with Data corruption
- ❖ Data Reliability and Recovery
- ❖ HDFS File Permissions
- ❖ Stronger Security in Hadoop
- ❖ Ports - NN/SNN/DN



- ❖ System Integrity
- ❖ Cluster Startup steps
- ❖ Practical - Lab II
 - ◆ WEBui -NN, SNN and DN
 - ◆ NN WEBui in detailed
 - ◆ DN Webui

- ❖ How to access HDFS from CLI
- ❖ Hadoop Shell Commands in depth
- ❖ Mapreduce
 - ◆ What is Mapreduce
 - ◆ MapReduce - The Big Picture
 - ◆ MapReduce Master-Slave architecture
 - ◆ MapReduce Features
 - ◆ MapReduce Job submission process
 - ◆ Job launch process - JobClient
 - ◆ Job launch process - Jobtracker
 - ◆ Job launch process - Tasktracker
 - ◆ Job launch process - Task
 - ◆ Job launch process - Task Attempt
 - ◆ Block diagram
 - ◆ Job Submission
 - ◆ Job Initialization
 - ◆ Job Scheduling
 - ◆ Job Execution

- ❖ MapReduce failure recovery
- ❖ Ports - JT/TT
- ❖ Practical - Lab III
 - ◆ Running a sample program - wordcount
 - ◆ Job Details
 - ◆ Jobtracker WebUI
 - ◆ Job ID
 - ◆ Tasks and task attempts

- ❖ Planning Your Hadoop Cluster
 - ◆ General Planning Consideration
 - ◆ Thinking about the Problem
 - ◆ Storage Capacity considerations

- ❖ Choosing The Right Hardware
 - ◆ Classifying Nodes
 - ◆ Slave nodes: Recommended configuration
 - ◆ Master node: SPOF

- ❖ Network Consideration
- ❖ Hardening Hadoop nodes
- ❖ Cloudera Manager - Easy Installation
 - ◆ Cloudera Manager Version
 - ◆ Cloudera Manager Features
 - ◆ Cloudera Manager database and Port details
 - ◆ Configuration Value Precedence



- ❖ Understand Hadoop config files
 - ◆ core-site.xml
 - ◆ hdfs-site.xml
 - ◆ mapred-site.xml
 - ◆ hadoop-env.sh
 - ◆ log4j.properties
 - ◆ hadoop-metrics.properties

- ❖ Managing Large Clusters
 - ◆ Cluster Shell
 - ◆ Configuration Management Tools

- ❖ Monitoring Hadoop Cluster
 - ◆ Cloudera Manager
 - ◆ Nagios
 - ◆ Ganglia

- ❖ Practical Lab IV
 - ◆ Revisiting Hadoop config files
 - ◆ Nagios Installation and Configuration
 - ◆ Ganglia Installation and Configuration

- ❖ Practical Lab V
 - ◆ Cluster Maintenance
 - ◆ Check HDFS status
 - ◆ HDFS Corruption
 - ◆ DFS admin utilities
 - ◆ Copying data Between Clusters

- ❖ Adding/Removing Nodes from Cluster
 - ◆ Using CM
 - ◆ Manually

- ❖ Adding/Removing Services from Cluster
 - ◆ Add specific roles to the cluster
 - ◆ Rebalancing Cluster

- ❖ Namenode Filesystem
 - ◆ Fsimage / Edits / Version / fstime
 - ◆ Backing up the Namenode

- ❖ Manually Backing Up Fsimage and edits
- ❖ Recovering the Namenode
- ❖ Log Files - for Debugging
- ❖ Hadoop Ecosystem
 - ◆ HIVE
 - ◆ HUE
 - ◆ OOZIE
 - ◆ PIG
 - ◆ Impala
 - ◆ Sqoop
 - ◆ Flume
 - ◆ Spark

