NETXILLION

BRIDGING THE TECHNOLOGY GAP

"Doing a course is not a guarantee for a job, but having a solid foundation surely is"

@Netxillon Technologies

For details on training dates or any custom training needs, please contact: trainings@netxillon.com

Course Contents

Hadoop Administration and Optimization

Course duration: 24 hours (20 hours session + 4 hours of doubts clearing session)

Session 1

Big Data and Introduction to Apache Hadoop

Learning Objectives - In this module, you will understand what Big Data and Apache Hadoop is. How Hadoop solves the Big Data problems, Introduction to Hadoop Components and the current market.

Topics

Introduction to Big Data.

Use cases where Big Data is used.

Challenges which BigData pose and Why Hadoop?

Introduction to Apache Hadoop and the major Versions.

Get a feel of the Hadoop Cluster

Hadoop Venders

Current Job Market

Session 2

Hadoop Architecture and Hadoop 2.0 Cluster setup

Learning Objectives – After this session, you will be able to understand Hadoop Components, such as NameNode, DataNode, HDFS and MapReduce.

Topics

Hadoop Components – HDFS and MapReduce

Understand HDFS.

Understand MapReduce.

Hadoop Installation and Initial Configuration.

Single Node and MultiNode Cluster Setup.

Session 3

Hadoop Cluster Features and MapReduce Components

Learning Objectives - After this module, you will understand details on NameNode metadata, communication between different components and some other cluster features like HDFS Read/Write, HDFS Replication and

Hadoop Edge Nodes.

Topics

Understand Namenode, Datanodes Communication.

Understand HDFS Replication.

Anatomy of Write and Read.

Replication Pipeline, Data Processing.

Setup Pass phrase-less Access.

Installing Hadoop Clients.

Setup YARN Components.

Run MapReduce Jobs.

Understand what YARN does.

Session 4

Hadoop Cluster: Planning and Managing

Learning Objectives - In this module, you will understand Planning and logging.

Types of schedulers in Hadoop – FIFO, FAIR SCHEDULER

Configuring the schedulers and run MapReduce jobs.

Setup Queues for Jobs.

Planning the Hadoop Cluster.

Cluster Sizing.

Hardware, Network and Software considerations.

Hadoop Logging.

Session 5

Hadoop Cluster Administration Tasks.

Configure Rack awareness.

Hadoop Balancer.

Setting up Secondary NameNode.

Understand Safemode of cluster.

Add Storage to Datanodes.

Setup Users and Quota's.

Hadoop Backup of Metadata, Data, Confguration fles.

Hadoop Cluster Membership.

Copy data across clusters using distcp.

Session 6

Introduction to Advanced topics in Hadoop 2.0

Learning Objectives - In this module, you will understand Hadoop 2.0 New Features, HDFS High Availability,

YARN framework, MRv2, and Hadoop 2.0 Cluster setup in distributed mode.

Topics

Introduction to High Availability.

Understand Federation.

Deploying a multi-node Hadoop 2.0 Cluster.

Configuring HDFS HA using Shared Storage.

Session 7

YARN Framework and High Availability.

Learning Objectives - In this module, you will understand HA for YARN and NN automatic failover.

Topics

YARN Execution and Workflow details.

Configure Resource Manager (RM) HA.

Configure Namenode High Availability using QJM.

Session 8

Hive, HBase Configuration and Examples.

Learning Objectives - In this module, you will understand Setting up Apache Hive and HBase.

Topics

Introduction to Hive and HBase.

Install and Configure Hive.

Hive with MySQL as meta store.

Install and Configure HBase Cluster.

Introduction to Sqoop and Flume.

Import data from MySQL into Hadoop using Sqoop.

Session 9

Ingestion basics and some of the techniques to create data flows.

Learning Objectives - In this module, you will understand Flume, Kafka basics and some of the real time concepts.

Topics

Install and Configure Flume.

Install and Configure Kafka

Ingest and consume data by creating a data pipeline.

Session 10

Security Basics, Data at Rest and Kerberos

Learning Objectives - In this module, you will understand SSL configuration and Kerberos for some of the Hadoop Services

Topics

Configure Encryption Zones.

Configure SSL for communication.

Configure Kerberos for some of the basic Hadoop services.

Labs

All are topics mentioned will be demonstrated with large amount of data.

Exam

In the end there will be an exam to test the knowledge acquired through this course.

Cost/Fee:

US $450 or INR 30000