



Training | Consulting | Development | Outsourcing



Hadoop Development

 9032803832

 9032803832

 contact@techyedz.com

 www.techyedz.com

Bigdata Hadoop Development



Course Overview:

Hadoop Development course teaches the skill set required for the learners how to setup Hadoop Cluster, how to store Big Data using Hadoop (HDFS) and how to process/analyze the Big Data using Map-Reduce Programming or by using other Hadoop ecosystems.

Hadoop course will equip you with all the skills you'll need for your next Big Data assignment. You will learn to work with Hadoop's Distributed File System, its processing and computation frameworks, core Hadoop distributions, and vendor-specific distributions such as Cloudera. You will learn the need for cluster management solutions and how to set up, secure, safeguard and monitor clusters and their components such as Sqoop, Flume, Pig, Hive and Hbase with this Big Data Hadoop course.



Course Outline:

Introduction to BigData, Hadoop

- Big Data Introduction
- Hadoop Introduction
- What is Hadoop? Why Hadoop?
- Hadoop History?
- Different types of Components in Hadoop?
- HDFS, MapReduce, PIG, Hive, SQOOP, HBASE, OOZIE, Flume, Zookeeper and so on...
- What is the scope of Hadoop?

Deep Dive in HDFS (for Storing the Data)

- Introduction of HDFS
- HDFS Design
- HDFS role in Hadoop
- Features of HDFS
- Daemons of Hadoop and its functionality
- Name Node
- Secondary Name Node
- Job Tracker
- Data Node
- Task Tracker

- Anatomy of File Write
- Anatomy of File Read
- Network Topology
- Nodes
- Racks
- Data Center
- Parallel Copying using DistCp
- Basic Configuration for HDFS
- Data Organization
- Blocks and
- Replication
- Rack Awareness
- Heartbeat Signal
- How to Store the Data into HDFS
- How to Read the Data from HDFS
- Accessing HDFS (Introduction of Basic UNIX commands)
- CLI commands

MapReduce using Java (Processing the Data)

- The introduction of MapReduce.
- MapReduce Architecture
- Data flow in MapReduce
- Splits
- Mapper
- Portioning
- Sort and shuffle
- Combiner
- Reducer
- Understand Difference Between Block and InputSplit
- Role of RecordReader
- Basic Configuration of MapReduce
- MapReduce life cycle
- Driver Code
- Mapper
- and Reducer
- How MapReduce Works
- Writing and Executing the Basic MapReduce Program using Java
- Submission & Initialization of MapReduce Job.

- File Input/Output Formats in MapReduce Jobs
- Text Input Format
- Key Value Input Format
- Sequence File Input Format
- NLine Input Format
- Joins
- Map-side Joins
- Reducer-side Joins
- Word Count Example
- Partition MapReduce Program
- Side Data Distribution
- Distributed Cache (with Program)
- Counters (with Program)
- Types of Counters
- Task Counters
- Job Counters
- User Defined Counters
- Propagation of Counters
- Job Scheduling

PIG

- Introduction to Apache PIG
- Introduction to PIG Data Flow Engine
- MapReduce vs. PIG in detail
- When should PIG be used?
- Data Types in PIG
- Basic PIG programming
- Modes of Execution in PIG
- Local Mode and
- MapReduce Mode
- Execution Mechanisms
- Grunt Shell
- Script
- Embedded
- Operators/Transformations in PIG
- PIG UDF's with Program
- Word Count Example in PIG
- The difference between the Map

- Reduce and PIG

SQOOP

- Introduction tSQOOP
- Use of SQOOP
- Connect tmysql database
- SQOOP commands
- Import
- Export
- Eval
- Codegen etc...
- Joins in SQOOP
- Export tMySQL
- Export tHBase

HIVE

- Introduction tHIVE
- HIVE Meta Store
- HIVE Architecture
- Tables in HIVE
- Managed Tables
- External Tables
- Hive Data Types
- Primitive Types
- Complex Types
- Partition
- Joins in HIVE
- HIVE UDF's and UADF's with Programs
- Word Count Example

HBASE

- Introduction tHBASE
- Basic Configurations of HBASE
- Fundamentals of HBase
- What is NoSQL?
- HBase Data Model

- Table and Row
- Column Family and Column Qualifier
- Cell and its Versioning
- Categories of NoSQL Data Bases
- Key-Value Database
- Document Database
- Column Family Database
- HBASE Architecture
- HMaster
- Region Servers
- Regions
- MemStore
- Store
- SQL vs. NOSQL
- How HBASE is differed from RDBMS
- HDFS vs. HBase
- Client-side buffering or bulk uploads
- HBase Designing Tables
- HBase Operations
- Get
- Scan
- Put
- Delete

HCatalog

- HCatalog Installation
- Introduction to HCatalog
- About Hcatalog with PIG,HIVE and MR
- Hands on Exercises

MongoDB

- What is MongoDB?
- Where tUse?
- Configuration On Windows
- Inserting the data intMongoDB?

- Reading the MongoDB data.

Cluster Setup

- Downloading and installing the Ubuntu12.x
- Installing Java
- Installing Hadoop
- Creating Cluster
- Increasing Decreasing the Cluster size
- Monitoring the Cluster Health
- Starting and Stopping the Nodes

Zookeeper

- Introduction Zookeeper
- Data Modal
- Operations

OOZIE

- Introduction tOOZIE
- Use of OOZIE
- Where tuse?

Flume

- Introduction tFlume
- Uses of Flume
- Flume Architecture
- Flume Master
- Flume Collectors
- Flume Agents

SPARK

- Spark Overview
- Linking with Spark, Initializing Spark
- Using the Shell
- Resilient Distributed Datasets (RDDs)

- Parallelized Collections
- External Datasets
- RDD Operations
- Basics, Passing Functions to Spark
- Working with Key-Value Pairs
- Transformations
- Actions
- RDD Persistence
- Which Storage Level to Choose?
- Removing Data
- Shared Variables
- Broadcast Variables
- Accumulators
- Deploying to a Cluster
- Unit Testing
- Migrating from pre-1.0 Versions of Spark
- Where to Go from Here

Prerequisites:

- Basic Unix Commands
- Core Java (OOPS Concepts, Collections , Exceptions) for Map Reduce Programming
- SQL Query knowledge for Hive Queries

Who Can attend:

- Working professionals in IT / Analytics / Statistics / Big Data / Machine Learning
- Fresh graduates from Engineering / Mathematics / IT backgrounds
- Professionals looking to develop skills to do statistical analysis to support decision making

Number of Hours: 40hrs

Certification: CCA - 175

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

TechyEdz Solutions