

(BIGDATA ANALYTICS) HADOOP

DAY-1

Hadoop

- What is Big Data?
- What is Hadoop?
- The relation between Big Data and Hadoop.
- What is the need for going ahead with Hadoop?
- Scenarios to apt Hadoop Technology in REAL TIME Projects
- Challenges with Big Data
 - Storage
 - Processing
- How Hadoop is addressing Big Data Changes
- Comparison with Other Technologies
 - RDBMS
 - Data Warehouse
 - Teradata
- Different Components of Hadoop Echo System
 - Storage Components
 - Processing Components

HDFS (Hadoop Distributed File System)

- What is a Cluster Environment?
- Cluster Vs Hadoop Cluster.
- The significance of HDFS in Hadoop
- Features of HDFS
- Storages aspects of HDFS
 - Block
 - How to Configure block size
 - Default Vs Configurable Block size
 - Why HDFS Block size so large?
 - Design Principles of Block Size

DAY-2

Hadoop-Setup

- **Virtual Machine Setup**
- **Hadoop Image Setup**
- **HDFS Architecture - 5 Daemons of Hadoop**
 - Name Node and its functionality
 - Data Node and its functionality
 - Resource Manager and its functionality
 - Task Track and its functionality
 - Secondary Name Node and its functionality.
- **Replication in Hadoop - Fail Over Mechanism**
 - Data Storage in Data Nodes
 - Fail Over Mechanism in Hadoop - Replication
 - Replication Configuration

- Custom Replication

DAY-3

Hadoop HDFS Commands(APIs)

Map Reduce

- Why is Map Reduce essential in Hadoop?
- Processing Daemons of Hadoop
 - Resource Manager
 - Roles Of Resource Manager
 - Drawbacks w.r.to Resource Manager failure in Hadoop Cluster
 - How to Configure Resource Manager in the Hadoop cluster
 - Node Manager
 - Roles of Node Manager
- Input Split
 - Input Split
 - Need of Input Split in Map Reduce
 - Input Split Size
 - Input Split Size Vs Block Size
 - Input Split Vs Mappers

Day-4

- Map Reduce Life Cycle
 - Communication Mechanism of Resource Manager & Node Manager
 - Input Format Class
 - Record Reader Class
 - Success Case Scenarios
 - Failure Case Scenarios
 - Retry Mechanism in Map Reduce

YARN (Yet Another Resource Negotiator)-Next Gen-Map Reduce

- What is YARN?
- YARN Architecture
 - Resource Manager
 - Application Master
 - Node Manager
- When should we go ahead with YARN
- Classic Map Reduce Vs YARN Map Reduce, Different Configuration Files for YARN

DAY-5

- Map Reduce Programming Model
 - Different phases of Map Reduce Algorithm
 - Different Data Types in Map Reduce
 - Primitive Data Types Vs Map Reduce Data Types
 - How to write a basic Map Reduce Program
 - Driver Code
 - Mapper Code
 - Reducer Code

DAY-6

Mapreduce(Continued..)

- **Driver Code**
 - Importance of Driver Code in a Map-Reduce Program
 - How to Identify the Driver Code in Map Reduce Program
 - Different sections of Driver code
- **Mapper Code**
 - Importance of Mapper Phase in Map Reduce
 - How to Write a Mapper Class?
 - Methods in Mapper Class
- **Reducer Code**
 - Importance of Reduce phase in Map Reduce
 - How to Write Reducer Class?
 - Methods in Reducer Class
 - Mapreduce program execution using CLI

DAY-7

Apache PIG

- Introduction to Apache Pig
- SQL Vs Apache Pig
- Different data types in Pig
- Modes of Execution in Pig
 - Local Mode
 - Map Reduce OR Distributed Mode
- Execution Mechanism
 - Grunt Shell
 - Script
- Embedded
- Transformations in Pig
- How to develop the Complex Pig Script
- Bags, Tuples and fields in PIG

DAY-8

- UDF's in Pig
 - Need of using UDF's in PIG
 - How to use UDF's
 - REGISTER key word in PIG
- When to use Map Reduce & Apache PIG in REAL TIME Projects

APACHE HIVE

- Hive Introduction
- Need of Apache HIVE in Hadoop
- Hive Architecture
 - Driver
 - Compiler
 - Executor(Semantic Analyzer)
- Meta Store in Hive
 - Importance of Hive Meta Store
 - Embedded meta store configuration
 - External meta store configuration

DAY-9

- Communication mechanics with Metastore

- Hive Integration with Hadoop
- Hive Query Language (Hive QL)
- Configuring Hive With MySQL Metastore
- SQL Vs Hive QL
- Hive installation and complete Hands-on

DAY-10

- Data Slicing Mechanisms
 - Partitions in Hive
 - Real-Time Use Cases
 - Hive Complex datatypes
 - Map
 - Array
 - Struct

DAY-11

- User Defined Functions(UDFs) in HIVE
 - UDFs
- Hive Hands-on - Running Hive scripts using hive shell and .hql file

DAY-12

APACHE SQOOP

- Introduction to Sqoop.
- MySQL client and Server Installation
- How to connect to Relational Database using Sqoop

DAY-13

- Different Sqoop Commands
 - Different Flavors of Imports
 - Export
 - Hive-Imports

APACHE Flume

- Flume Introduction
- Flume Architecture
- Flume Master, Flume Collector, and Flume Agent
- Flume Configurations
- Real-Time Use Case using Apache Flume

DAY-14

APACHE HBase

- HBase Introduction
- HDFS Vs HBase
- HBase Use cases
- HBase basics
 - Column Families
 - Scans
- HBase Architecture
- Map Reduce Integration

- Map Reduce over HBase
- HBase Admin
 - Scheme Definition
 - Basic CRUD Operations

DAY-15

PROJECT-1: Temperature and Telecom data processing using Map Reduce, HIVE and PIG

Processing the temperature data sets from different years recorded over period of time and analyze them to identify what are cities impacted more and show the appropriate recommendations for particular cities

PROJECT -2: Import the cricket players data recorded over period of time from MySQL to Hadoop using sqoop and Analyze them based on each individual performance. And also Import the incremented data from MySql using sqoop jobs