

EXPO NENT

IT TRAINING & SERVICES

Knowledge Is The Best Investment

- C/C++
- JAVA FULL STACK
- PYTHON
- DSA
- HADOOP

- DATA SCIENCE
- DATA ANALYTICS
- REACT
- MERN/ MEAN
- MANUAL, AUTOMATION & TESTING

- AWS LINUX & DEV-OPS
- TABLEAU
- POWER BI
- SERVICE NOW

100% PLACEMENT RECORD

Top Rated On Google

5.0 ★★★★★ 299 Google reviews

Software training institute in Pimpri-Chinchwad, Maharashtra

- Live Project Work
- Hand On Session
- Learn From Industry Experts.
- 100% Interactive Sessions.

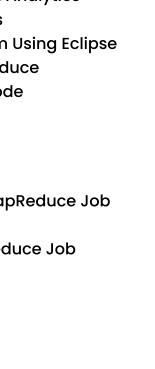
- Interview Preparation.
- One To One Interaction.
- Offline And Online Support.
- 100% Industry Oriented Training.

996-090-8543
986-074-8543

www.youtube.com/@exponent-it

www.exponent-it.com

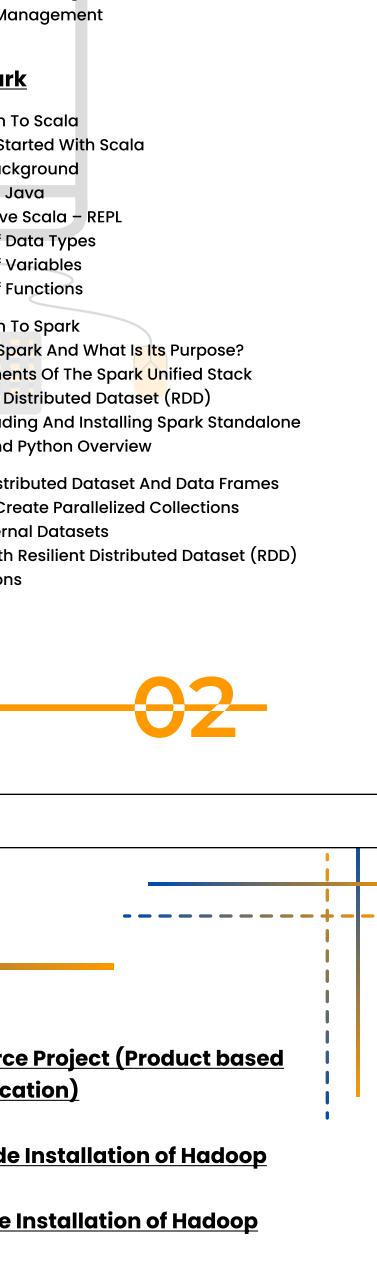
Shop 5 & 6, S Building Dhruv Darshan society, near PCCOE College



HADOOP

COURSE CONTENTS

- [VMware installation](#)
- [MySQL database](#)
- [Core Java](#)
- [BIGDATA](#)
- [HADOOP](#)
- [Tools](#)
- [Apache Spark](#)
- [Ecommerce Project \(Product based web application\)](#)
- [Single node Installation of Hadoop](#)
- [Multimode Installation of Hadoop](#)



HADOOP syllabus

VMware installation

MySQL database

Core Java

- Types Of Variables
- Types Of Data Types
- Types Of Modifiers
- Types Of Constructors
- Object Oriented Programming in Java

BIGDATA

- Introduction To Big Data

- Characteristics Of Big Data Of Data Types

- Examples of Big data

- Types of data

- Sources of Big data

- What is Streaming data?

- Batch Vs Streaming data

- Big Data Hadoop opportunities

HADOOP

- Hadoop Introduction And HDFS introduction

- Hadoop Architecture

- Hadoop Versioning And Configuration

- Single Node Hadoop Installation On Ubuntu

- Hadoop Commands

- Hadoop Cluster Architecture (Production Environment) And Block Placement

- Modes In Hadoop I.E. Local, Pseudo Distributed, Fully Distributed

- Hadoop Components I.E. Master Components (Name Node, Secondary Name Node, Job Tracker), Slave Components (Job Tracker, Task Tracker)

- Task Instance

- Hadoop HDFS Commands

- HDFS Access

- Map Reduce Introduction
 - Understanding Map Reduce Framework
 - What is Map Reduce Base?
 - Mapper Class And Its Methods
 - What is Partitioner And Types
 - Relationship Between Input Splits And HDFS Blocks
 - MapReduce: Combiner & Partitioner
 - Hadoop Specific Data Types
 - Working On Semi Structure Data Analytics
 - Types Of Mappers And Reducers
 - Developing MapReduce Program Using Eclipse
 - Analysing Dataset Using MapReduce
 - Running MapReduce In Local Mode

- MapReduce (In Detail)
 - How MapReduce Works
 - Anatomy Of MapReduce Job
 - Submission & Initialization Of MapReduce Job
 - Assigning & Execution Of Tasks
 - Monitoring & Progress Of MapReduce Job
 - Completion Of Job
 - Handling Of MapReduce Job
 - Task Failure
 - Task Tracker Failure
 - Job Tracker Failure

- Advanced Topic For MapReduce (Performance And Optimization)
 - Job Scheduling
 - In Depth Shuffle And Sorting

- Speculative Execution

- Output Committers

- JVM Reuse In MRI

- Configuration And Performance Tuning

- Advanced MapReduce Algorithm

- File Based Data Structure
 - Sequence File
 - Map File

- Default Sorting In MapReduce
 - Data Filtering
 - Partial Sorting

- Data Lookup Strategies
 - In Map Files

- Sorting Algorithm
 - Total Sort
 - Input Sampler
 - Secondary Sort

- MapReduce In Query
 - Joins, Sub Queries And Views
 - Writing User Defined Functions (UDFs)
 - Data types and schemas
 - Querying Data
 - Hive ODBC
 - User Defined Functions

- Apache HBase
 - Fundamentals Of HBase
 - Usage Scenario Of HBase
 - Use Of Hbase In Search Engine
 - HBase Data Model
 - Table And Row
 - Column Family And Column Qualifier
 - Cell And Its Versioning
 - Regions And Region Server
 - HBase Designing Tables
 - HBase Data Coordinates
 - Versions And Hbase Operation
 - Get/Scan
 - Put
 - Delete

- Apache Zookeeper
 - Cluster Monitoring
 - Cluster Management

Apache Spark

- Introduction To Scala
 - Getting Started With Scala
 - Scala Background
 - Scala Vs Java
 - Interactive Scala – REPL

- Types Of Data Types
 - Types Of Variables
 - Types Of Functions

- Introduction To Spark
 - What is Spark And What is Its Purpose?
 - Components Of The Spark Unified Stack
 - Resilient Distributed Dataset (RDD)
 - Downloading And Installing Spark Standalone
 - Scala And Python Overview

- Resilient Distributed Dataset And Data Frames
 - How To Create Parallelized Collections And External Datasets
 - Work With Resilient Distributed Dataset (RDD) Operations

Spark SQL

- Introduction To Apache Spark SQL

- The SQL Context

- Importing And Saving Data

- Processing The Text Files, JSON And Parquet Files

- Data Frames

- Hive Tool

Spark MLLib

- Introduction To Machine Learning

- Types Of Machine Learning

- Introduction To Apache Spark MLLib Algorithms

- Machine Learning Data Types And Working With MLLib

Spark GraphX

- What Is GraphX?

- What Are Property Graph?

- How To Create A Graph Using GraphX

MapReduce Data Types And Formats

- Serialization In Hadoop

- Hadoop Writable And Comparable

- Hadoop Raw Comparator And Custom Writable

- MapReduce Types And Formats

- Understand Difference Between Block And Input Split

- Role Of Record Reader

- File Input Format

- Comma File Input Format And Processing Whole File Single Mapper

- Each Input File As A Record

- Text/Key Value/ Input Format

- Binary Input Processing

- Multiple Inputs Format

- Database Input And Output

Apache Hive

- What Is Hive ?

- Architecture Of Hive

- Hive Services

- Hive Clients

- How Hive Differs From Traditional RDBMS

- HiveQL Introduction

- Data Types And File Formats In Hive

- File Encoding

- Common Problems While Working With Hive

- Introduction To HiveQL

- Managed And External Tables

- Understand Storage Formats

- Querying Data

- Sorting And Aggregation

Apache Sqoop

- Introduction To Sqoop

- How Does Sqoop Work

- Sqoop JDBC Driver And Connectors

- Various Options To Import Data

- Table Import

- Binary Data Import

- SpeedUp The Import

- Filtering Import

- Export Data Sqoop Function

- Mysql Database Commands Using Eval Function

Apache HBase

- Fundamentals Of Hbase

- Usage Scenario Of Hbase

- Use Of Hbase In Search Engine

- Hbase Data Model

- Table And Row

- Column Family And Column Qualifier

- Cell And Its Versioning

- Regions And Region Server

- Hbase Designing Tables

- Hbase Data Coordinates

- Versions And Hbase Operation

- Get/Scan

- Put

- Delete

Apache Zookeeper

- Cluster Monitoring

- Cluster Management

Apache Spark

- Introduction To Scala

- Getting Started With Scala

- Scala Background

- Scala Vs Java

- Interactive Scala – REPL

- Types Of Data Types

- Types Of Variables

- Types Of Functions

Apache Flink

- Introduction To Flink

- Getting Started With Flink

- Scala Background

- Scala Vs Java

- Interactive Scala – REPL

- Types Of Data Types

<