BIGDATA
TRAINING COURSE CONTENT

INTRODUCTION – 60 HOURS

**All About
Big Data
and Hadoop**

Intro to Big Data | Big Data Characteristics | Role of Hadoop in Big data | Challenges & Opportunities in Hadoop | Hadoop Architecture | Hadoop Clone | HUE (Hortonworks & Cloudera Ecosystem) Overview | Use Cases of Hadoop (HealthCare, Retail, Telecom) |

HADOOP DISTRIBUTED FILE SYSTEM

DFS

HDFS Architecture & Concepts | HDFS Blocks, Name nodes, Data nodes, Command Line Interface **|** Basic File System Operations | Data Flow (File Read, File Write) | Block Placement Policy & Modes | Configuration Files | Metadata, FS image, Edit Log, Secondary Name Node, Safe Mode | FSCK Utility | Overriding Default Configuration | HDFS Federation | ZOOKEEPER Leader Election Algorithm | Use Case of HDFS

MAP REDUCE

**Programming
Model**

Working of Map Reduce | Data Flow (Map – Shuffle – Reduce) | Map Reduce Job Anatomy | Map Reduce Job(Initialization, Completion, Submission, Failures), Task (Assignment, Execution, Progress, Status Update | Programming (Mapper, Reducer, Combiner, Partitioner) | Shuffling & Sorting | Writable | Input/output Format | MapRed vs MapReduce APIs | Getting data from RDBMS into HDFS using Custom data types | Distributed Cache and Hadoop Streaming (Python, Ruby and R) | YARN | Sequential Files and Map Files | Enabling Compression Codec's

MAP REDUCE PROGRAMMING

Programming

Hands on "Word Count" in Map Reduce Standalone and Pseudo distribution Mode | Sorting Files using Hadoop Configuration API discussion | Emulating "grep" for searching inside a file in Hadoop | DBInput Format | Job Dependency API discussion | Input Format & Input Split API discussion | Custom Data type creation in Hadoop

NOSQL

Database

Introduction to NoSQL | CAP Theorem | Classification of NoSQL | Comparison between properties in RDBMS and NOSQL | Detailed Explanation of Columnar Databases(HBASE & CASSANDRA) | BLOOM FILTERS

HBASE

HADOOP DBMS

HBASE Architecture & Data Model | HBASE Installation | HBASE and RDBMS Comparison | HBASE and HDFS Comparison | HBASE Operations through Shell Programming | Catalog Tables | SPLITS | Data Modelling (Sequential, Salted, Promoted and Random Keys) | JAVA API's and Rest Interface | HBASE Counters, Filters, RAW Scans | Bulk Loading and Coprocessors (Endpoints and Observers with programs) HBASE

HIVE

Data Warehouse Software Project

HIVE Intro & Architecture | HIVE Installation | HIVE (Services, Shell, Server, Web Interface, QL) | HIVE vs. RDBMS | OLTP vs. OLAP | TABLES-DDL-DML-UDF | Primitive data types and complex data types of HIVE | Partitioning –Dynamic Partition| Bucketing & Sorted Bucketing with Dynamic Partition | Cast Function METASTORE | Difference between ORDER BY, DISTRIBUTE BY and SORT BY | INDEXES & VIEWS | MAPSIDE JOINS | Compression on HIVE tables and Migrating HIVE tables | Access HBASE tables using HIVE

PIG

Application

PIG Architecture Overview | PIG Installation | HIVE vs. PIG | PIG Grunt Shell Commands | Pig Latin | Primitive data types and complex data types of PIG | Tuple schema, BAG Schema & MAP Schema | Loading and Storing-Filtering-Grouping & Joining-Debugging commands | Validations in PIG | Type casting in PIG | Default & User Defined Functions- Dynamic Invokers and Macros | Types of JOINS in PIG and Replicated Join in detail | SPLITS and Multiquery execution | Error Handling, FLATTEN and ORDER BY | Accessing HBASE using PIG | Loading and Writing JSON DATA using PIG | Piggy Bank

SQOOP

Data Integration Tools

SQOOP Architecture | SQOOP Installation | SQOOP Commands (Import , HIVE-Import, Eval, Hbase Import, Import All tables, Export to RDBMS, HIVE & HBASE)

HCATALOG

Metadata

Intro to HCATALOG | HCATALOG Installation | HCATALOG with PIG, HIVE & MAPREDUCE

KAFKA

Bigdata Messaging System

Intro to KAFKA & Data Streaming | Topics on Producer & Consumer- Partitions- Brokers | Unix Streaming via KAFKA

FLUME

Data Integration Tool

Intro to Flume | Flume Installation | Flume Agents | Log User information using Java program (in to HDFS using LOG4J and Avro Source, in to HDFS using Tail Source, in to HBASE using LOG4J and Avro Source, in to HBASE using Tail Source) | Flume Commands

OOZIE

Workflow Scheduler

OOZIE Architecture | Workflow (Action, Start, Action, End, Kill, Join and Fork), Schedulers, Coordinators and Bundles | Workflow to schedule SQOOP Job, HIVE, MAPREDUCE and PIG | HBASE Integration with HIVE and PIG | Phoenix

SPARK

Framework

SPARK Overview & Fundamentals | Limitations in Hadoop | HDFS Federation | High Availability in HDFS | Linking & Initializing Spark | Resilient Distributed Datasets (RDDs) – Operations-Persistence | Parallelized Collections |External Datasets-Functions- Key Value Pairs | Transformations-Actions-Variables | Sample Scala Program | Spark Streaming | Deploying to a Cluster | Unit Testing | Migrating from pre-1.0 Versions of Spark

HANDS-ON PROJECTS

E-Commerce Data Analysis

Banking Data Analysis

Predict the sales of a store using Bigmart Sales Data Set

OUR
LOCATION

Greens
Technologys

Offers

+91 89399 15577

20% DISCOUNT On

All Courses

ADYAR   OMR   VELACHERY

TAMBARAM   ANNA NAGAR