Trendz IT
PEOPLE COMMITTED TO QUALITY
Corporate Training Since 2000

Exclusive Value Additions
- CYU – Check Your Understanding
- LOTW – Learning On The Walls
- PPR – Participant Progress Report
- PTS – Post Training Support (Web)

BIG DATA WITH APACHE HADOOP

Duration: 3 Days

WHAT IS BIG DATA & WHY HADOOP?
- Big Data Characteristics, Challenges with traditional system

HADOOP OVERVIEW & ECOSYSTEM
- Anatomy of Hadoop Cluster, Installing and Configuring Hadoop
- Hands-On Exercise

HADOOP ARCHITECTURE
- Components in Hadoop
- Interaction between different Components
- Basic Understanding of each component

HDFS – HADOOP DISTRIBUTED FILE SYSTEM
- Name Nodes and Data Nodes
- Hands-On Exercise

MAP REDUCE ANATOMY
- How Map Reduce Works?
- The Mapper & Reducer, InputFormats & OutputFormats, Data Type
- Hands-On Exercise

PSEUDO – CLUSTER DISTRIBUTION OF VANILLA HADOOP
- Hadoop Extraction and Installation
- Configuration / XML Files
- Hands on Exercise

YARN
- Need for YARN
- Architecture of YARN

INSTALLATION OF YARN IN CENTOS/UBUNTU
- Configuration Settings
- Difference between Gen1 and Gen2

DEVELOPING MAP REDUCE PROGRAMS
- Setting up Eclipse Development Environment, Creating Map Reduce Projects, Debugging Map Reduce Code
- Hands-On Exercise

ADVANCED MAP REDUCE ALGORITHMS
- Sorting, Searching and Indexing, Multiple Inputs, Chaining multiple jobs
- Joins, Handling Binary & Unstructured data

ADVANCED TIPS & TECHNIQUES
- Determining optimal number of reducers, skipping bad records
- Partitioning into multiple output files & Passing parameters to tasks

MONITORING & MANAGEMENT OF HADOOP
- Managing HDFS with Tools
- Routine Administration Procedures
- Commissioning and decommissioning of nodes
- Hands-On Exercise

USING HIVE

- Hive as a Data Warehouse
- Creating External & Internal Tables plus Loading Data
- Writing HSQL queries for data retrieval
- Creating partitions and querying data.

USING PIG

- Why Pig and its benefits
- Loading data into PigStorage
- Querying data from PigStorage
- Hands-On Exercise

SQOOP

- Importing and Exporting data from using RDBMS
- Hands-On Exercise

UNDERSTANDING THE OTHER SQL OPTIONS IN HADOOP
HADOOP BEST PRACTICES AND USE CASES

TAKE AWAY FROM THE COURSE

- Understanding of What and Why of Hadoop with its Eco-System Components.
- Ability to write Map Reduce programs in a given scenario
- Ability to correctly architect and implement the Best Practices in Hadoop Development
- Ability to Manage and Monitor Hadoop
- Ability to Manage the different Hadoop Components when talking to each other.