HADOOP 2.X BIGDATA ANALYTICS

1. Java
- ➢ Overview of Java
- ➢ Classes and Objects
- ➢ Garbage Collection and Modifiers
- ➢ Inheritance, Aggregation, Polymorphism
- ➢ Command line argument
- ➢ Abstract class and Interfaces
- ➢ String Handling
- ➢ Exception Handling, Multithreading
- ➢ Serialization and Advanced Topics
- ➢ Collection Framework, GUI, JDBC

2. Linux
- ➢ Unix History & Over View
- ➢ Command line file-system browsing
- ➢ Bash/CORN Shell
- ➢ Users Groups and Permissions
- ➢ VI Editor
- ➢ Introduction to Process
- ➢ Basic Networking
- ➢ Shell Scripting live scenarios

3. SQL
- ➢ Introduction to SQL, Data Definition Language (DDL)
- ➢ Data Manipulation Language(DML)
- ➢ Operator and Sub Query
- ➢ Various Clauses, SQL Key Words
- ➢ Joins, Stored Procedures, Constraints, Triggers
- ➢ Cursors /Loops / IF Else / Try Catch, Index
- ➢ Data Manipulation Language (Advanced)
- ➢ Constraints, Triggers,
- ➢ Views, Index Advanced

Address: - 1st Floor, Famous Chowk, New Sanghvi, Pune-411027.
Contact no- 08390635774/07546995948. Email ID- info.innovatus@gmail.com.
Website – www.innovatustechnos.com

 Hadoop – Bigdata

1. Introduction to Big data
- Introduction and relevance
- Uses of Big Data analytics in various industries like Telecom, E-commerce, Finance and
- Insurance etc.
- Problems with Traditional Large-Scale Systems

2. Hadoop (Big Data) Ecosystem
- Motivation for Hadoop
- Different types of projects by Apache
- Role of projects in the Hadoop Ecosystem
- Key technology foundations required for Big Data
- Limitations and Solutions of existing Data Analytics Architecture
- Comparison of traditional data management systems with Big Data management systems
- Evaluate key framework requirements for Big Data analytics
- Hadoop Ecosystem & Hadoop 2.x core components
- Explain the relevance of real-time data
- Explain how to use big and real-time data as a Business planning tool

3. Building Blocks
- Quick tour of Java (As Hadoop is Written in Java , so it will help us to understand it better)
- Quick tour of Linux commands (Basic Commands to traverse the Linux OS)
- Quick Tour of RDBMS Concepts (to use HIVE and Impala)
- Quick hands on experience of SQL.
- Introduction to Cloudera VM and usage instructions

4. Hadoop Cluster Architecture – Configuration Files
- Hadoop Master-Slave Architecture
- The Hadoop Distributed File System - data storage
- Explain different types of cluster setups (Fully distributed/Pseudo etc.)

Address: - 1st Floor, Famous Chowk, New Sanghvi, Pune-411027.
Contact no- 08390635774/07546995948. Email ID- info.innovatus@gmail.com.
Website – www.innovatustechnos.com

➢ Hadoop Cluster set up - Installation
➢ Hadoop 2.x Cluster Architecture
➢ A Typical enterprise cluster – Hadoop Cluster Modes

5. Hadoop Core Components – HDFS & Map Reduce (YARN)

6. HDFS Overview & Data storage in HDFS
➢ Get the data into Hadoop from local machine (Data Loading Techniques) - vice versa
➢ MapReduce Overview (Traditional way Vs. MapReduce way)
➢ Concept of Mapper & Reducer
➢ Understanding MapReduce program skeleton
➢ Running MapReduce job in Command line/Eclipse
➢ Develop MapReduce Program in JAVA
➢ Develop MapReduce Program with the streaming API
➢ Test and debug a MapReduce program in the design time
➢ How Partitioners and Reducers Work Together
➢ Writing Customer Partitioners Data Input and Output
➢ Creating Custom Writable and Writable Comparable Implementations

7. Data Integration Using Sqoop and Flume
➢ Integrating Hadoop into an existing Enterprise
➢ Loading Data from an RDBMS into HDFS by Using Sqoop
➢ Managing Real-Time Data Using Flume
➢ Accessing HDFS from Legacy Systems with FuseDFS and HttpFS
➢ Introduction to Talend (community system)
➢ Data loading to HDFS using Talend

8. Data Analysis using PIG
➢ Introduction to Hadoop Data Analysis Tools
➢ Introduction to PIG - MapReduce Vs Pig, Pig Use Cases
➢ Pig Latin Program & Execution
➢ Pig Latin : Relational Operators, File Loaders, Group Operator, COGROUP Operator, Joins and COGROUP, Union, Diagnostic Operators, Pig UDF

Address: - 1st Floor, Famous Chowk, New Sanghvi, Pune-411027.
Contact no- 08390635774/07546995948. Email ID- info.innovatus@gmail.com.
Website – www.innovatustechnos.com

➢ Use Pig to automate the design and implementation of MapReduce applications
➢ Data Analysis using PIG

9. Data Analysis using HIVE
➢ • Introduction to Hive - Hive Vs. PIG - Hive Use Cases
➢ • Discuss the Hive data storage principle
➢ • Explain the File formats and Records formats supported by the Hive environment
➢ • Perform operations with data in Hive
➢ • Hive QL: Joining Tables, Dynamic Partitioning, Custom MapReduce Scripts
➢ • Hive Script, Hive UDF

10. Data Analysis Using Impala
➢ Introduction to Impala & Architecture
➢ How Impala executes Queries and its importance
➢ Hive vs. PIG vs. Impala
➢ Extending Impala with User Defined functions
➢ Improving Impala performance

11. NoSQL Database – Hbase
➢ Introduction to NoSQL Databases and Hbase
➢ HBase v/s RDBMS, HBase Components, HBase Architecture
➢ HBase Cluster Deployment

12. Hadoop – Other Analytics Tools
➢ Introduction to role of R in Hadoop Eco-system
➢ Introduction to Jasper Reports & creating reports by integrating with Hadoop
➢ Role of Kafka & Avro in real projects

13. Other Apache Projects
➢ Introduction to Zookeeper - ZooKeeper Data Model, Zookeeper Service
➢ Introduction to Oozie - Analyze workflow design and management using Oozie

Address: - 1st Floor, Famous Chowk, New Sanghvi, Pune-411027.
Contact no- 08390635774/07546995948. Email ID- info.innovatus@gmail.com.
Website – www.innovatustechnos.com

- Design and implement an Oozie Workflow
- Introduction to Storm
- Introduction to Spark

Address: - 1st Floor, Famous Chowk, New Sanghvi, Pune-411027.
Contact no- 08390635774/07546995948. Email ID- info.innovatus@gmail.com.
Website – www.innovatustechnos.com