

Big Data Hadoop-2.8.0 with Spark-2.1.1 and Scala-2.11

Day -1: Introduction to BigData and Hadoop

1. Introduction to Data and BigData
2. BigData Characteristics
3. Distributed File Systems
4. History of Apache Hadoop
5. About Doug Cutting
6. What is Apache Hadoop ?
7. What is commodity Hardware ?
8. Apache Hadoop Components ?
9. General Hadoop cluster in the company
10. What is edgeNode ?
11. Hadoop distributed File System.
12. Introduction to NameNode
13. Introduction to DataNode
14. Introduction to Secondary NameNode
15. Introduction to dataset
16. Introduction to putty
17. Introduction to Winscp
18. Execute the basic Hadoop commands
19. HDFS basic WebUI

Day -2: Introduction to HDFS

1. Introduction to hadoop command
2. HDFS Architecture
3. Anatomy of File Write
4. Create program for HDFS Write operation
5. Setup Eclipse project using maven
6. Anatomy of File Read
7. Replication and RackAwareness
8. Block Placement Policy
9. Secondary NameNode checkpoint Backup Mechanism
 - a. What is the responsibility of secondary NameNode ?
 - b. Why FsImage and Edits.log?
 - c. Why NameNode is a SPOF?

ELANCERSOFT SOLUTIONS

H.No: 7-1-282/C/1/A, 1st floor, BK Guda Beside Indian Oil Petrol pump(Rajaiah Complex), Balkampet.

PH: 040-48540745, +91-9704249988 **EMAIL:** online@elancersoft.com www.online

- d. Why SNN conf = NN Conf?
10. Hadoop Configuraiton
11. Hadoop Cluster Modes
12. What are the disadvantages of Generation 1 HDFS
13. What is HDFS Federation?
14. What is HDFS High Availability?

Day-3: YARN and MapReduce

1. What is YARN?
2. What are YARN Components?
3. What is data locality?
4. What is Input Split
5. Explain Yarn jar command
6. Creating program for Word Count
7. Anatomy of MapReduce Paradigm
8. Create a Mapreduce program to identify which day in the year was recorded with max temperature
9. Create a MapReduce Program to DeIdentify personal information
10. Deep Dive in Combiner
11. Deep Dive in Partitioner
12. Joins in MapReduce
13. Counter In MapReduce
14. Input Format
15. MR Unit
16. Sequential File

Day-4: HIVE

1. Hive Background
2. Hive Use Case at Facebook
3. What is hive ?
4. Hive Architecture
5. Hive Components
6. Hive Metastore
7. Limitations of Hive
8. Abilities of Hive Query Language
9. Schema on Read Vs Schema on Write
10. Hive DataTypes
11. Hive Partitions
12. Hive Bucketing
13. Create Database
14. Different Types of Tables

15. Loading Data
16. Multi Insert
17. Hive script of writing programs
18. Joining Two tables
19. Hive UDF
20. Different Types of Joins
21. Hive Static Partition and Dynamic Partition
22. Hive Running Custom Python Script
23. Hive Index
24. Hive Views
25. Hive JDBC

Day-5: HBASE

1. UseCases: What is Random Access and examples?
2. Traditional way of Solving the BigData Search
3. Limitation
4. Required Solution
5. CAP Theorem
6. HBASE VS RDBMS
7. Major Components of HBASE
8. Data Distribution
9. Hbase Minor Components
10. HBASE Storage Architecture
11. HBASE Read and Write
12. HBASE Region
13. HBASE Region Server
14. CLIENT LOOKUP LIBRARY
15. COMPACTION
16. HBASE Running Modes
17. HBASE Filters
18. HFile Storage Formats
19. Name Spaces
20. Table Data Model
21. HBASE Physical Storage
22. HBASE Java API
23. HIVE on HBASE
24. Sqoop to HBASE

Day-6: SCALA

1. Introduction to Programming Languages
2. Introduction to Object Programming languages
3. Introduction to Function Programming languages
4. Introduction to Scala and how it fits to the market
5. Create a Project Learning Scala in SBT
6. Scala REPL
7. Scala Interpreter
8. Scala Debugger
9. Keywords
10. Expression
11. Variables
12. Type Inference
13. DataTypes
14. Statements
15. Functions
16. Methods
17. Classes
18. Object
19. Case Classes
20. DataStructures
21. Useful Methods

Day – 7: SPARK CORE

1. Big Data Analytics
2. There are Other Alternatives, then Why go for Spark
3. What is Spark
4. What is the difference between spark1 and spark2
5. Spark Features
6. Spark in Hadoop Ecosystem and How to Useful with Hadoop
7. Saprk Components
8. Spark Project Setup
9. Sample Spark Core Execution: WordCount, PI
10. Anatomy of Spark Paradigm
11. Spark-Submit
12. Spark Architecture
13. Spark Cluster Modes
14. Spark Deploy Modes
15. Spark WebUI

ELANCERSOFT SOLUTIONS

H.No: 7-1-282/C/1/A, 1st floor, BK Guda Beside Indian Oil Petrol pump(Rajaiah Complex), Balkampet.

PH: 040-48540745, +91-9704249988 **EMAIL:** online@elancersoft.com www.online

16. Spark Properties
17. Introduction to RDD
18. RDD properties
19. Creation Of RDD
20. Different Types of RDD
21. Transformation of RDD with examples
22. Actions of RDD with examples
23. RDD Lineage Graph with examples
24. RDD persistence and StorageLevel with examples
25. RDD Partitions with examples
26. Accumalators with examples
27. BroadCast Variables with example

Day-8: SPARK SQL

1. Introduction to SparkSQL
2. SparkSQL Architecture
3. Introduction to SparkSession
4. Introduction to DataFrame and Datasets
5. Creating a DataFrame
6. Sql on DataFrame
7. What is sqlContext
8. Loading json, CSV and write SQL Queries
9. Interaporating with RDD
10. using Case Classes. -- Reflection
11. implicit Schema.
12. SparkSQL JDBC
13. SparkSQL HiveContext
14. SparkSQL creating hive tables
15. SparkSQL on HBASE

Day-9: KAFKA

1. Introduction to Kafka
2. Need of Kafka
3. What is kafka ?
4. Kafka Components
5. Compare the kafka With RabbitMQ
6. Kafka Architecture
7. Producer



8. Broker
9. Topic
10. Partitions
11. Replications
12. Kafka Configurations
13. Kafka Consumer
14. Consumer Groups

Day -10: Spark-Streaming

1. What is Streaming
2. Fault Tolerance
3. Streaming Fundamentals
4. Streaming Context
5. Introduction to Dstream
6. Caching and persistence
7. Accumulators, Broad Cast Variables and Checkpoints
8. Window operations in Streaming
9. Stateful Operators
10. Streaming Datasources

ELANCERSOFT SOLUTIONS

H.No: 7-1-282/C/1/A, 1st floor, BK Guda Beside Indian Oil Petrol pump(Rajaiah Complex), Balkampet.

PH: 040-48540745, +91-9704249988 **EMAIL:** online@elancersoft.com www.online