

Non-Java Developer Training (Sqoop, Pig, Hive, Flume)

Hadoop (Hortonworks 2x)

Introduction to Hadoop

1. Tradition architecture Vs Hadoop Architecture
2. Elements of Hadoop Ecosystem
3. Pros and Cons
4. Hadoop Distribution
5. Copy to/from HDFS
6. Creating Directories in HDFS
7. Other HDFS commands

Apache Sqoop (1x)

1. Introduction to Sqoop
2. Data in RDBMS
3. Sqoop extract
4. Sqoop eval, list
5. Sqoop Import
6. Sqoop import incremental
7. Sqoop export
8. Sqoop export merge/upsert
9. Sqoop delimiters - import/export
10. Sqoop file formats
11. Sqoop Job and Sqoop Merge

Apache Flume (1x)

1. Introduction to Flume
2. Start a flume Agent, given a configuration file
3. Configure a flume memory channel with a specified capacity

Apache Pig

1. Intro to Apache Pig
2. Write & Execute Pig Latin scripts
3. Word Count example demo
4. Execution life cycle
5. Load data to pig relations with and without schema
6. Load data from Hive table into a pig relation
7. Transform data to match Hive schema
8. Group Data of Pig relations – count of rows
9. Group Data of Pig relations – Group by key
10. Filtering – remove records with null values from a relation
11. Store data into a folder in HDFS
12. Store data into a folder in Hive table
13. Sort output of a pig relation
14. Transform data to match Hive schema
15. Group Data of Pig relations – count of rows
16. Group Data of Pig relations – Group by key
17. Filtering – remove records with null values from a relation
18. Store data into a folder in HDFS
19. Store data into a folder in Hive table
20. Sort output of a pig relation
21. Remove duplicate tuples from a Pig relation
22. Specify the number of reduce tasks for a Pig MR job
23. Joins overview (inner join, outer join)
24. Join 2 data sets – 01
25. Join 2 data sets – 02
26. Types of joins (replicated, skewed etc)
27. Replicated joins using Pig
28. Run a Pig job using Tez
29. User Defined Functions (UDF) in Pig
 - a) Registering Jars
 - b) Defining alias
 - c) Invoking UDFs

Apache Hive (1x)

1. Introduction to data analysis
2. Data warehousing using Hadoop ecosystem – Hive Architecture
3. Hive 01 – Write and execute Hive query
4. Hive 02 – Write and execute Hive query
5. Create Hive managed table
6. Create Hive External table
7. Create Hive Partitioned table
8. Create Hive bucketed table
9. Hive CTAS and ORC file formats
10. Hive – Specify file formats and delimiters
11. Load data from local file system and HDFS to Hive table
12. Load data from hive select query (using insert) and compressed data
13. Transactions in Hive (insert, update and delete)
14. Hive joins, execution engines (tez and mr) and explain/execution plan
15. Hive sub queries and total ordering (Order By)

Big Data Developer Training

Optional

Can set aside sessions to work through all the tasks from the Hortonworks practice exam

Up to 10 task oriented (exam style questions)