Session 7: Assignment 1

Filename: Assignment 7.1

Submitted by Akhil Sharma

1. Why Map-reduce program is needed in Pig Programming?
2. **Because in pig, Map-reduce is needed to process the data which is stored using the HDFS. Pig is a high-level platform that makes analysis of big data easier by writing the scripts in pig-latin which get converted into map-reduce program during execution. Therefore Map-Reduce is needed in pig programming as well.**
3. What are advantages of pig over MapReduce?
4. **Below are the advantages:**
   1. **In pig there is no need of compiling or packaging of code.**
   2. **During execution, pig script gets auto-converted to map reduce program.**
   3. **It is easier to join, filter, group by, order by and apply union in pig as compared to complex mapreduce java program.**
   4. **Development cycle of pig scripts is faster than mapreduce program.**
   5. **It is easier to maintain a pig script over mapreduce program as it uses pipeline concept.**
5. What is pig engine and what is its importance?
6. **Pig engine is an execution environment where all the pig latin scripts are converted into map-reduce programs/jobs. It is important because it helps in preparing series of mapreduce jobs which are executed in the background by Hadoop.**
7. What are the modes of Pig execution?
8. **There are two modes for using PIG.**
   1. **Local Mode – In this mode, all files are used and maintained on a local JVM machine. This is helpful for developers to test their pig scripts before actually executing on Hadoop HDFS.**
   2. **MapReduce Mode – This is a default mode and used for executing the scripts on Hadoop cluster.**
9. What is grunt shell in Pig?
10. **Grunt shell is an interactive environment in which PIG scripts are executed. It remembers command history and can be used to recall the last executed commands using up and down arrow. Using tab, we can also auto-complete some of the commands which are there in the history.**
11. What are the features of Pig Latin language?
12. **Following are the features of Pig latin language:**
    1. **It is made up of series of operations which are used to transform the input data to produce meaningful output data.**
    2. **It provides User Defined Function (UDF) feature which is helpful to transform certain fields within the file.**
    3. **Pig latin in local environment provides debugging facility.**
    4. **Pig latin includes many SQL like operations like join, order, group, filter etc.**
    5. **Pig latin can be executed in interative as well as batch mode. For batch mode .pig extension files are created which includes all the commands in series stored inside a file. On the CLI, .pig can be submitted to run the commands in series to produce the desired output.**
13. Is Pig latin commands case sensitive?
14. **Alias of relations and fields are case sensitive. Pig latin functions are also case sensitive. So yes, pig latin commands are case sensitive.**
15. What is a data flow language?
16. **Data flow languages are used to explore large datasets. It is made up of series of operations and/or transformations which are applied to large input datasets to produce a desired output.**