Assignment 6.2

1. Why MapReduce program is needed in Pig Programming?

Pig is a high level platform that makes data analysis easier. When the program is written in pig, it needs execution engine to execute the query. So, Pig compiler converts that program into an mapreduce jobs. The mapreduce job done the operations internally, this is the reason pig needs the mapreduce.

1. What are the advantages of pig over mapreduce?

IT decreases the development time, complexity.

The person don’t know the Java and SQL can easily write the program in pig.

Pig latin script can easily perform the mapreduce task without the need of Java.

Pig reduces the code length whereas mapreduce is 20 times of pig script length.

1. What is Pig engine and what is its importance?

Pig has a component named s pig engine which accepts the pig script as input and converts into an mapreduce jobs. This pig engine acts as a interpreter between the pig script and the mapreduce job. It executes the pig script into serious of mapreduce jobs in a parallel manner. Pig provides the common data operations (filters, joins, ordering, etc.) and nested data types (e.g. tuples, bags, and maps) missing from MapReduce.

1. What are the modes of pig execution?

MapReduce/Hadoop mode : Here pig jobs runs as a serious of mapreduce jobs from HDFS. Here the files are stored inside the HDFS.

Pig –x mapreduce

Local mode : Here the files are stored in the local file system. Here pig jobs are running in the unix path execution.

Pig –x local

1. What is grunt shell in pig?

Grunt shell is a interactive shell for executing the pig command. It is used when the script file is not provided.

1. What are the features of pig latin language?

Pig latin provides the all data processing operations such as join, filter, group by, order by, union etc.

In Pig latin joins and ordering codes comprise only 9-10 lines of code and the execution takes only few minutes.

Keywords in pig latin are not case sensitive. ‘LOAD’ is equal to ‘load’.

Pig provides high level language that is used by data analyst and data scientist.

1. Is pig latin commands case sensitive?

Keywords in pig latin are not case sensitive. ‘LOAD’ is equal to ‘load’.

1. What is data flow language?

Data flow language is a programming that models a program as a directed graph of the data flowing between the operations. Pig is a data flow language which traverse the data from pig latin to an mapreduce.