Assignment 7.1

Give a brief answers to the questions below:

1. Why Map-reduce program is needed in Pig Programming?

MapReduce provides parallelism based on rigid procedural structure. It allows programmers to filter and aggregate data from HDFS to gain meaningful insight.

This feature is useful in converting pig Latin queries to Map & Reduce jobs for Parallel Processing.

2. What are advantages of pig over MapReduce?

It uses Pig Latin language, which aims at DataFlow operations instead of control flow operations, hence more focus is on data analysis.

• It uses simple syntax and in-built features/operators to reduce development effort and lines of codes involved.

• Pig Latin queries are converted to Map and Reduce jobs and hence they take advantage of parallel processing.

• It has the ability to perform computations which can not be done by MapReduce.

• 1/20th lines of code as compared to MapReduce

• Java is not a preferred language for many

• High level of Abstraction

• Scripting language

• Development effort is less, code efficiency is relatively less

3. What is pig engine and what is its importance?

**Pig Engine**  accepts the **Pig** Latin scripts as input and converts those scripts into MapReduce jobs.

It acts as interpreter between pig latin script and MapReduce jobs. It creates environment to execute pig scripts into series of MapReduce jobs in Parallel manner.

4. What are the modes of Pig execution?

**Pig** has two **execution modes** or exectypes: Local **Mode** - To run **Pig** in local **mode**, you need access to a single machine; all files are installed and run using your local host and file system. ... Mapreduce **Mode** - To run **Pig** in mapreduce **mode**, you need access to a Hadoop cluster and HDFS installation.

5. What is grunt shell in Pig?

The Grunt shell of Apache Pig is mainly used to write Pig Latin scripts. Prior to that, we can invoke any shell commands using sh and fs.

6. What are the features of Pig Latin language?

statements, data types, general and relational operators, and Pig Latin UDF’s are some of the features of Pig Latin

7. Is Pig latin commands case sensitive?

No they are not case sensitive

8. What is a data flow language?

To access external data, every language must follow rules and regulations. The instructions flows through data by executing different control statements, but data doesn’t get moved. Dataflow language can get a stream of data which passes from one instruction to another instruction to be processed. Pig can easily process those conditions, jumps, loops and process the data in efficient manner.