Assignment 7.1

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

Ans : Map-Reduce program is needed in Pig Programming to have advantage of Parallel Processing.

2. What are advantages of pig over MapReduce?

Ans: • Pig Latin provides all of the standard data-processing operations, such as join, filter, group by, order by, union, etc. • MapReduce requires programmers: • Programmers must think in terms of map and reduce functions. • Most probably Java programmers are required. • In Pig Latin joins and ordering codes comprise of 8-9 lines of code and take few minutes to write and debug. The same code in MapReduce will span hundred lines of code and takes hours to develop. • Pig provides high-level language that can be used by: • Data Analysts • Data Scientists

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

Ans: Pig Engine acts as interpreter between Pig Latin Script and MapReduce jobs. It creates an environment to execute Pig scripts into series of mapreduce jobs in parallel manner.

4. What are the modes of Pig execution?

Ans: There are Two modes in Pig execution . They are

Local Mode: No need to start or install hadoop. The Pig Script runs in the local system. Pig store data in File system.

MapReduce Mode : It is mandatory to start Hadoop. Pig scripts run and stored in HDFS. In both modes , Pig and Java installations mandatory.

5. What is grunt shell in Pig?

Ans : The grunt shell of Apache Pig is used to writePig Latin scripts. Besides entering Pig Latin interactively, Grunt's other major use is to act as a shell for HDFS.

6. What are the features of Pig Latin language?

Ans: Apache Pig is a high-level platform for creating programs that run on Apache Hadoop. The language for this platform is called Pig Latin. Pig can execute its Hadoop jobs in MapReduce, Apache Tez, or Apache Spark. Pig Latin abstracts the programming from the Java MapReduce idiom into a notation which makes MapReduce programming high level, similar to that of SQL for relational database management systems. Pig Latin can be extended using user-defined functions (UDFs) which the user can write in Java, Python, JavaScript, Ruby or Groovy and then call directly from the language.

7. Pig latin commands case sensitive?

Ans : The names (aliases) of relations and fields are case sensitive. The names of Pig Latin functions are case sensitive. The names of parameters (see Parameter Substitution) and all other Pig Latin keywords are case insensitive.

In the example below, note the following:

1. The names (aliases) of relations A, B, and C are case sensitive.
2. The names (aliases) of fields f1, f2, and f3 are case sensitive.
3. Function names PigStorage and COUNT are case sensitive.
4. Keywords LOAD, USING, AS, GROUP, BY, FOREACH, GENERATE, and DUMP are case insensitive. They can also be written as load, using, as, group, by, etc.
5. In the FOREACH statement, the field in relation B is referred to by positional notation

8. What is Data Flow Language?

In a dataflow language, you have a stream of data which is passed from instruction to instruction to be processed. Conditional execution, jumps and procedure calls route the data to different instructions. This could be seen as data flowing through otherwise static instructions like how electrical signals flow through circuits or water flows through pipes. A dataflow "if" statement would route the data to the correct branch. Pig is Also known as Data Flow Language.