**Introduction to PIG – Assignment 7.1**

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

* Pig runs on Hadoop. It makes use of both the Hadoop Distributed File System,HDFS, and Hadoop’s processing system, MapReduce.
* It simplifies Hadoop usage by allowing SQL like queriesPig Scripts are compiled to series of Map reduce programs internally and get executed. So Pig still requires Hadoop platform and MapReduce program

1. What are advantages of pig over MapReduce?

* MapReduce program required Java knowledge. Many programmers do not have and need to learn.
* MapReduce programmers need to think of map and Reduce functions.
* PIG uses SQL like scripts and so easy to learn.It suppors SQL like operations like join, Union
* Many built in functions and operations. 1/20 th of code is required compared to mapreduce
* PIG can be used by Data Analysts and Data scientists.

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

To analyze data using **Apache Pig**, programmers need to write scripts using Pig Latin language. All these scripts are internally converted to Map and Reduce tasks. Apache Pig has a component known as **Pig Engine**  which is an execution environment that accepts the Pig Latin scripts as input and converts those scripts into MapReduce jobs in parallel manner.

1. What are the modes of Pig execution?

* Local Mode : Pig Runs in Single local JVM picking Unix path for Execution. So input files should be in the local unix folder.
  + Command: Pig -x local
* Map Reduce Mode : PIG runs as a series of MapReduce jobs picking the HDFS file system . So pig scrits run and strored in HDFS. Need to start Hadoop.
  + Command: Pig -x mapreduce or pig
  1. Both modes require Java and PIG installation.

1. What is grunt shell in Pig?

* Using Grunt i.e. Apache Pig’s interactive shell, users can interact with HDFS or the local file system. Executing the command “pig –x local” will result in the prompt - grunt >
* This is where PigLatin scripts can be run either in local mode or in cluster mode by setting the configuration in PIG\_CLASSPATH.
* To exit from grunt shell, press CTRL+D or just type exit.

6. What are the features of Pig Latin language?

Apache Pig comes with the following features −

* **Rich set of operators** − It provides many operators to perform operations like join, sort, filer, etc.
* **Ease of programming** − Pig Latin is similar to SQL and it is easy to write a Pig script if you are good at SQL.
* **Optimization opportunities** − The tasks in Apache Pig optimize their execution automatically, so the programmers need to focus only on semantics of the language.
* **Extensibility** − Using the existing operators, users can develop their own functions to read, process, and write data.
* **UDF’s** − Pig provides the facility to create **User-defined Functions** in other programming languages such as Java and invoke or embed them in Pig Scripts.
* **Handles all kinds of data** − Apache Pig analyzes all kinds of data, both structured as well as unstructured. It stores the results in HDFS.

7. Is Pig latin commands case sensitive?

* Names(aliases)of relations, fields, names of PIG Latin functions are case sensitive.
* But the parameters and Pig Latin keywords are case insensitive

8. What is a data flow language?

Data flow Language is to get streams of data from one instruction to other instruction so that data flows in a efficient manner.

PIG easily runs those conditions, jumps, loops and process the data.