**Session 7 Pig**

**Assignment 1**

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

**Answer:** Map-reduce program is needed in Pig programming as pig is an application that works on top of MapReduce, Yarn or Tez. Pig is a compiler that takes Pig Latin scripts and transforms them into JavaPig i.e. MapReduce jobs

Initially pig generates only logical plan ., it starts it actual execution at the time of dimp command and in batch mode exec command.

2.What are advantages of pig over MapReduce?

**Answer:**

a) Pig is more concise. A 200 lines Java code written for MapReduce can be reduced to 10 lines of PIG code.

b) to learn the pig is easy there is no need to learn JAVA before learning Pig.

c) Procedural, so easier to follow the commands and provides better expressiveness in the transformation of data every step. So it is simple compare to map reduce

d) Since it is procedural, you could control of the execution of every step. If you want to write your own UDF(User Defined Function) and inject in one specific part in the pipeline, it is straightforward.

e) In Mapreduce it’s too difficult to join multiple data sets. Development cycle is very long.

f) it will happily consume any data you feed it: structured, semi-structured, or unstructured.

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

**Answer:** It’s acts as interpreter between Pig Latin script and MapReduce Jobs. It creating environment (parses, optimizes,)to execute Pig scripts into series of mapreduce jobs in parallel manner.

4.What are the modes of Pig execution?

**Answer:**

**Local Mode:** With access to a single machine, all files are installed and run using a local host and file system. No need to start hadoop. But it has limitations with large files.

To run local mode command is: pig –x local

**MapReduce Mode:** This is the default mode, which requires access to a Hadoop cluster. For this we first need to start hadoop cluster

To run pig on mapreduce mode command is : pig –x mapreduce

5.What is grunt shell in Pig?

**Answer:**

a) Interactive Shell for executing Pig Commands.

b) Used when script file is not provided.

c) Can execute scripts from Grunt via run or exec commands

d) There are certain useful shell and utility commands provided by the Grunt shell like sh ,ls

e) It also provide some utility command like clear , help,history

6.What are the features of Pig Latin language?

**Answer:**

**a) Rich set of operators** − It provides many operators to perform operations like join, sort, filer, etc.

**b) Ease of programming** − Pig Latin is similar to SQL and it is easy to write a Pig script if you are good at SQL.

c) **Optimization opportunities** − The tasks in Apache Pig optimize their execution automatically, so the programmers need to focus only on semantics of the language.

d) **Extensibility** − Using the existing operators, users can develop their own functions to read, process, and write data.

e) **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.

f) **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?

**Answer:** yes it should either complete in capital or complele in small we can use LOAD and load but we cannot use Load.

8.What is a data flow language?

**Answer:** To access the external data, every language must follow many rules and regulations. The instructions are flowing 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.