**ASS-9.1**

**1. Why mapreduce program is needed in pig programming?**

Pig is an abstraction over MapReduce. In other words, all Pig scripts internally are converted into Map and Reduce tasks to get the task done.

So Map reduce is the layer on top of which Pig is built.Moreover Map reduce is the means by which Hadoop can communicate with HDFS.so

MapReduce program is needed in Pig Programming

**2. What are the advantages of pig over mapreduce?**

>Pig uses only few lines of command that saves the writing and debugging time compared to map reduce .

>Joins are complex in mapreduce program where as in pig it is very simple

>Pig does not require java knowledge and programming time is less

>To want specific functions ,user can create their own udf programs

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

>Pig engine acts as an environment where the pig latin scripts are converted into map reduce programs and acts as the execution environment

**4. what are the modes of pig execution**?

Two modes of execution:

1.Local mode: Runs on local host ,it does not require hdfs or Hadoop cluster.But java and pig should be installed on the local host

2.Mapreduce mode:It runs on hdfs system where mapreduce programs are invoked while executing pig commands parallel.

**5. what is grunt shell in pig?**

Grunt shell helps in executing pig commands step by step ,so that it does not take the whole program to execute at once for a long period.It is used when script is not provided.

**6. what are the features of pig latin language?**

Pig is a high-level procedural language for querying semi-structured and unstructured datasets.

• It uses simple SQL query for exeecution

• It provides various relational operators like join,sort,etc.. that reduce tedious programs.

• Pig runs on Hadoop. It makes use of both the Hadoop Distributed File System, HDFS, and Hadoop’s processing system, MapReduce.

• It allows users to describe how data from one or more inputs should be read, processed, and then stored to one or more outputs in parallel.

• Pig Latin script describes a directed acyclic graph (DAG), where the edges are data flows and the nodes are operators that process the data.

**7. Is pig latin command case sensititve?**

The names of pig latin are case sensitive where as the keywors are not case sensitive like foreach,filter ,etc.

**8. what is a data flow language?**

In programming, **dataflow** programming is a programming paradigm that models a program as a directed graph of the **data** flowing between operations, thus implementing **dataflow** principles and architecture.