# Session 7 ASSIGNMENT 1

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

Answer: Apache Pig programming is using HDFS and MapReduce platform of Hadoop to process the semi-structured data. Basically it has simplified the Hadoop with SQL like queries. Pig can process the data file which is either kept in Local Unix path or HDFS path.

Actually Pig’s infrastructure layer consists of a compiler that produces sequences of MapReduce programs, for which large scale parallel implementations is already exist.

1. What are advantages of Pig over MapReduce?

Answer: Listed below are the advantages of Pig over MapReduce:

1. Easy to use all the Data processing operations like Join, Sort, Filter etc. having simple syntax.
2. Number of lines need to be written in Pig is much less and lesser complexity compared to written same with Java for MapReduce.
3. Java programming skill set is not required. Basic know of RDBMS SQL queries is helpful
4. Data Scientists and Data analysts can also us the Pig programming
5. What is pig engine and what is its importance?

Answer: Pig Engine is component of Apache Pig that accepts the Pig Latin scripts as input and converts those scripts into MapReduce jobs.

It also submit the MapReduce jobs to Hadoop in a sorted order. Finally these MapReduce jobs are executed on Hadoop producing the desired results.

1. What are the modes of Pig execution?

Answer: Apache Pig can run in two modes, **Local Mode** and **HDFS Mode.**

1. **Local Mode:** All the files are installed and run from your local host and local file system. There is no need of Hadoop or HDFS. Mostly used for testing purpose.
2. **MapReduce Mode:** Load or process the data that exists in HDFS. In this mode whenever a Pig Latin statement is invoked to process the data, a MapReduce job is invoked in the back-end to perform a particular operation on the data that exists in the HDFS.
3. What is grunt shell in Pig?

Answer: It is Interactive shell for executing Pig Commands. It is used when script file is not provided.

1. What are the features of Pig Latin Language?

Answer: It has operators for many of the traditional data operations (join, sort, filter etc.). Allow the users to develop their own functions for reading, processing and writing data.

1. Is Pig latin commands case sensitive?

Answer: Yes

1. What is a data flow language?

Answer: In computer 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.