PIG Assignment

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

Ans: Pig was designed to load transform and analyze large datasets , but the pig commands which is used for the purpose was finally converted to the map reduce program and it was executed on the data stored in HDFS by the Pig Internal

Components , hence it can be said Pig is a scripting language which runs on top of Map reduce framework and perform analysis on the data stored in hdfs. Hence it is much needed.

1. What are advantages of pig over MapReduce?

Ans :

1. Writing and Executing PIG script is a simple task when compared with MapReduce as it don’t require importing suitable jar files and scripts in Pig can be written using simple commands.

2. For writing Map Reduce program , knowledge on Java is required while for Pig it is not so, there is no need of much programming skills. The entire program is based on PIG transformations

3. Amount of code is very less when compared to MapReduce program. 200 lines of MapReduce program is equivalent to 10 lines of Pig script.

4.Join operations on tables and other complex tasks can easily be performed using Pig , which is difficult in MapReduce program

5. Pig being a procedural language, you can easily debug at each intermediate steps.

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

Ans:

Apache Pig has a component known as **Pig Engine** that accepts the Pig Latin scripts as input and converts those scripts into MapReduce jobs.

It act as interpreter between Pig Latin Scripts and MapReduce jobs .it creates environment to execute Pig Scripts into series of MapReduce jobs in parallel manner.

1. What are the modes of Pig execution?

Ans : There are two modes of execution in PIG

**Local mode:** No need to start or install Hadoop. The pig scripts run in the local system. By default Pig store data in File system.   
100% MapReduce and Local mode commands everything same, no need to change anything.

**MapReduce Mode:** It’s mandatory to start Hadoop. Pig scripts run and stored in in HDFS. in Both modes, Java and Pig installation is mandatory.

1. What is grunt shell in Pig?

Ans: After starting PIG in either local or mapreduce mode user are provided with grunt shell , you can run your Pig scripts in the shell. In addition to that, there are certain useful shell and utility commands provided by the Grunt shell.

Shell Commands

The Grunt shell of Apache Pig is mainly used to write Pig Latin scripts. Prior to that, we can invoke any shell commands using **sh** and **fs**.

Like shell commands **grunt> sh ls**

Like hdfs commands **grunt> fs –ls**

1. What are the features of Pig Latin language?

Ans : To write data analysis programs, Pig provides a high-level language known as **Pig Latin**. This language provides various operators using which programmers can develop their own functions for reading, writing, and processing data.

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.

7. Is Pig latin commands case sensitive?

Ans: Pig is not purely case sensitive

Pig is case senstive as well as in-senstive.

Case senstive : for Alias, in-built functions, UDFs   
Case Insenstive : Operators like foreach, as etc.

8. What is a data flow language?

Ans: 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.