**Assignment9.1**

**Que.1) Why MapReduce program is needed in Pig Programming?**

MapReduce is low level of programming while Pig is a high-level language to express data analysis programs which internally create sequence of Map Reduce Programs. Pig is simple to learn and use as compared to Map Reduce.

Pig data flow language (pig Latin) For MapReduce, Java is by default supported programming language. However support for other language is also available here.

MapReduce developer needs to look into optimization while Pig provides inbuilt optimization for MapReduce jobs.

**Que.2) what are advantages of pig over MapReduce?**

Advantages of using Pig over MapReduce are as follows-

1. Apache pig is data flow language while MapReduce is data processing paradigm.
2. Pig is high level and flexible language while MapReduce is low level and rigid.
3. In Apache Pig performing join operation on data sets is pretty simple.
4. Basic knowledge of SQL Query is sufficient in Apache while to perform MapReduce task JAVA coding is needed.
5. Pig has multi-query approach hence length of code is reduced to great extent, while in MapReduce we need to write almost 15-20 times more lines.
6. No Compilation is needed as each operator in Pig is converted to MapReduce Job, in MapReduce Jobs compilation is must.

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

Pig Latin is the language used to analyze data in Hadoop using Pig . It is a high-level data processing language which provides a rich set of data types and operators to perform various operations on the data.

To perform any task Programmers using Pig, programmers need to write a Pig script using the Pig Latin language, and execute them using any of the execution mechanisms (Grunt Shell, UDFs, Embedded). After execution, these scripts will go through a series of transformations applied by the Pig Framework, to produce the required output.



Internally, Apache Pig converts all these scripts into a series of MapReduce jobs, and thus , it makes the programmer’s job to assemble easy. The architecture of Apache Pig is shown in diagram above. Finally the MapReduce jobs are submitted to Hadoop in a sorted order. Finally, these MapReduce jobs are executed on Hadoop producing the desired results.

**Que.4**) **what are the modes of Pig execution?**

Pig can either be executed locally, in the **local mode** using the local execution environment, or on a Hadoop cluster execution environment in the MapReduce mode.

The various modes are as follows-

1) **Interactive mode**: In this mode, a grunt shell is provided to the user. Users can type in Pig commands in an interactive session with Pig and the Hadoop cluster.

2**) Batch Mode**:  In this mode, the user can write a series of Pig statements into a script file. The file can then be submitted for execution.

3) **Embedded Mode**: In this mode, any Java program can invoke Pig commands by importing the Pig libraries.

In the execution process the control is passed from parser to optimizer, to compiler, to then finally to an execution engine in the form of DAG-Directed Acyclic Graphs.

**Que.5) what is grunt shell in Pig?**

The  Grunt shell  of Apache  Pig  is mainly used to write  Pig  Latin scripts manually.

The commands supported by grunt can be used in interactive shell as well as in batch mode. The supported commands include DFS commands, pig commands as well as a few others. All commands are case insensitive and white spaces are not significant.

**Que.6)** **what are the features of Pig Latin language?**

Apache Pig is a high-level platform for creating programs that run on Apache Hadoop.

The features of Pig Latin are as follows:

1) It consists of many MapReduce design patterns.

2) It is implemented in JAVA programming language.

3) Instead of providing Java Based API framework, Pig provides its own scripting language which is called as Pig Latin.

4) Pig Latin is a very simple scripting language. It has constructs which can be used to apply different transformation on the data one after another.

5) The Code is optimized to great extent, as over other platforms codes written are 15-20 times longer.

**Que.7) Is Pig Latin commands case sensitive?**

Unfortunately, Pig Latin cannot decide whether it is case-sensitive. Keywords in Pig Latin are not case-sensitive. For example,

 LOAD is equivalent to load, but relation and field names are not. So A = load ‘foo’ is not equivalent to a = load ‘foo’.

UDF names are also case-sensitive, thus COUNT is not the same UDF as count.

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

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 the architecture.  You have a stream of data which is passed from instruction to instruction to be processed.

Pig provides developers many operators which can be applied on data one after another to get final output.

Once data is loaded, it flows through all Pig operators.

This is the reason **Pig is called as data flow language**.

Examples- Pig Latin, Unix Pipes, VHDL etc.