Assignment 9.1

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

Pig is an application that runs on the top of MapReduce .The Pig programming is higher-level programming which enables many developers or analysts to write MapReduce jobs but the MapReduce is what runs at the back end and performs processing.

1. **What are advantages of pig over MapReduce?**

 Map Reduce is low level of programming and Pig is a high-level language for expressing data analysis programs which internally create sequence of Map Reduce Programs.

Pig is simple to learn and use as compared to Map Reduce.

Pig provides inbuilt optimization for MR jobs whereas in map reduce developer needs to take care of optimization.

Performing a Join operation in Apache Pig is pretty simple compare to MapReduce.

There is no need for compilation. On execution, every Apache Pig operator is converted internally into a MapReduce job.

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

Pig engine acts as interpreter between Pig Latin scripts and the MapReduce jobs.The Pig engine creates an environment in order to execute Pig scripts into series of Mapreduce jobs in parallel,

1. **What are the modes of Pig execution?**

Apache Pig has two execution modes, namely, **Local Mode** and **HDFS mode**.

Local Mode

In this mode, all the files are installed and run from your local host and local file system. There is no need of Hadoop or HDFS. This mode is generally used for testing purpose.

MapReduce Mode

MapReduce mode is where we load or process the data that exists in the Hadoop File System (HDFS) using Apache Pig. In this mode, whenever we execute the Pig Latin statements 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.

1. **What is grunt shell in Pig?**

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.

sh Command

Using **sh** command, we can invoke any shell commands from the Grunt shell.

fs Command

Using the **fs** command, we can invoke any FsShell commands from the Grunt shell.

After invoking the Grunt shell, we can run your Pig scripts in the shell.

1. **What are the features of Pig Latin language?**

Apache Pig comes with the following features −

* **Rich set of operators** − It provides many operators to perform operations like join, sort, filer, etc.
* **Ease of programming** − Pig Latin is similar to SQL and it is easy to write a Pig script if you are good at SQL.
* **Optimization opportunities** – The tasks in Apache Pig optimize their execution automatically, so the programmers need to focus only on semantics of the language.
* **Extensibility** − Using the existing operators, users can develop their own functions to read, process, and write data.
* **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.
* **Handles all kinds of data** − Apache Pig analyzes all kinds of data, both structured as well as unstructured. It stores the results in HDFS.

1. **Is Pig latin commands case sensitive?**

The Pig latin commands are case insensitive.

1. **What is a data flow language?**

In a dataflow language, you have a stream of data which is passed from instruction to instruction to be processed. Conditional execution, jumps and procedure calls route the data to different instructions. This could be seen as data flowing through otherwise static instructions like how electrical signals flow through circuits or water flows through pipes.