Session 7

# Assignment -1

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

**Ans. -** Pig is an application that works on top of MapReduce, Yarn or Tez. Pig is written in Java and compiles Pig Latin scripts into to MapReduce jobs. Think of Pig as a compiler that takes Pig Latin scripts and transforms them into Java.

**Ques.2- What are advantages of pig over MapReduce?**

**Ans.-** Pig is application that runs on top of MapReduce and abstracts Java MapReduce jobs away from developers.

* Pig Latin uses a lot fewer lines of code than the Java MapReduce script.
* The Pig Latin script was is easier to read for someone without a Java background.
* MapReduce jobs can written in Pig Latin.
* Java is a great and powerful language, but it has a higher learning curve than something like Pig Latin. Therefore, using a higher-level language, like Pig Latin, enables many more developers/analysts to write MapReduce jobs.

**Ques.3- What is pig engine and what is its importance?**

**Ans. -** Pig is a Apache open source project which is run on hadoop. Pig provides engine for data flow in parallel on hadoop. It includes language called pig latin, which is for expressing these data flow. It includes different operations like joins, sort, filter etc. and also ability to write UserDefine Functions(UDF) for proceesing and reaing and writing. pig uses both HDFS and MapReduce i,e storing and processing.

Importance:

* Pig increases programming productivity
* Decreases duplication of effort
* Opens the M/R programming system to more uses
* Ping insulates against hadoop complexity

1. Hadoop version upgrades
2. Job configuration tunning

**Ques.4 – What are the modes of Pig execution?**

**Ans. –** Pig has two execution modes

**Local Mode -** In Local Mode of Pig execution, all the input data will be taken from local file system. After execution it provides output on top of local file system. In local mode, Pig runs in a single JVM and accesses the local file system. This mode of suitable only for small datasets and when trying out Pig.

**MapReduce Mode -** In this mode Apache Pig will take the input form HDFS paths only, and after processing data it will put output files on top of HDFS. In MapReduce mode of execution, Pig translates queries into MapReduce jobs and runs them on a Hadoop Cluster.

**Ques.5 - What is grunt shell in Pig?**

**Ans. -** Grunt acts as a command interpreter where you can interactively enter Pig Latin at the Grunt command line and immediately see the response. This method is helpful for prototyping during initial development and with what-if scenarios.

**Ques.6 - What are the features of Pig Latin language?**

**Ans.-** Comes with 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.

**Ques.7 - Is Pig latin commands case sensitive?**

**Ans. -** pig latin is sometimes not a case sensitive.

The names (aliases) of relations and fields are case sensitive. The names of Pig Latin functions are case sensitive. All other Pig Latin keywords are case insensitive.

**Ques.8 - What is a data flow language?**

**Ans. -** 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.