**ACADGILD BIG DATA ASSIGNMENT 7.1**

1. **Why Map Reduce programming is needed in Pig programming ?**

Map Recue programming is needed because Pig is meant for Big Data analysis. And to analyze huge amount of data Map Reduce is required because Map Reduce inherently supports parallelism while storing (HDFS) and processing (Map Reduce) data.

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

There are 3 advantages of Pig over Mapreduce:

1. Shorter and simpler code: For the same task, Pig requires on an average only 1/20th lines as compared to that of a MapReduce program written in Java. Also, Pig uses a language called Pig Latin which is a scripting language. So there is no need for the programmer to compile the program separately.
2. Convenient for non-Java programmers: For some professionals who come from the non-Java background like DBAs, Pig offers lot of flexibility by providing a scripting language called Pig Latin. The Pig compiler automatically converts the Pig commands to MapReduce jobs.
3. Many built-in functions available: There are many built-in functions available to do the common data loading and processing tasks. Also in those cases where a function is not available to do our tasks, we can write a User Defined Function(UDF) in Java and use it in Pig.
4. **What is pig engine and what is its importance?**

The Pig Execution Engine is responsible for submitting the MapReduce jobs in sorted order to the Hadoop cluster.

There are 4 steps in executing a Pig command/script:

1. Parser: It checks for the syntax of the entered commands. And if the command complies to the syntax, the parser constructs a Directed Acyclic Graph (DAG). The nodes of the DAGs are operators and the edges are data flows
2. Optimizer: Carries out logical optimizations on the DAG output by the parser. Examples of optimizations are projection and push down
3. Compiler: The compiler compiles the optimized DAG plan into a series of MapReduce jobs
4. Execution engine: This is responsible for submitting the sorted MapReduce jobs to the Hadoop cluster
5. **What are the modes of Pig execution ?**

There are two modes of execution in Pig: local and mapreduce/Hadoop mode which are described below

1. Local mode: In local mode, the pig commands are executed on a single JVM in which Pig is installed. The files and read and stored from the local file system itself. We start Pig in this mode using the command “pig -x local”
2. MapReduce mode: In map reduce mode, the Pig commands are execute on the Hadoop cluster. The Pig shell / the pig command acts as a Hadoop client, therefore reading and storing the files in HDFS. Pig is started in map reduce mode by using the command “pig” or “pig –x mapreduce”.
3. **What is grunt shell in pig ?**

Grunt shell is an interactive shell for executing Pig commands. We can use grunt shell when script file is not provided. We can also execute script files using run or exec command.

1. **What are the features of the Pig Latin language?**
2. Pig Latin is the language used when entering Pig command and when writing Pig scripts.
3. There is a list of reserved keywords. You cannot use these keywords when declaring relations or functions
4. Pig Latin relations and fields are case-sensitive
5. Pig Latin provides many complex data types such as tuple, bag etc.
6. Pig Latin provides many arithmetic operators and relational operators
7. You can also write your own user defined functions (UDFs) and register them with the Pig runtime. Then, these UDFs can be used in the Pig Latin scripts
8. **Is Pig Latin commands case sensitive?**

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

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

A dataflow language is a programming language that models a program as a directed graph of the data flowing between operations. The nodes are data operations and the edges are the data flowing between these operations. Pig Latin is a data flow language.