**Big Data Hadoop Assignment 7.1**

Q-1: Why Map-reduce program is needed in pig programming?

Ans : The Apache pig has a component known as Pig Engine that accepts the pig Latin scripts as input and converts those scripts into MapReduce jobs and hence they take advantage of parallel processing. It has the ability to perform computations which cannot be done by MapReduce.

Q-2: What are advantage of pig over MapReduce?

Ans: (1) - Pig Latin provides all of the standard data-processing operations, such as join filter, group by, order by, union, etc.

(2) – Map Reduce requires programmers:

* + Programmers must think in terms of map and reduce functions.
  + Most probably java programmers are required.

(3) – In pig Latin joins and ordering codes comprise of 8-9 lines of code and take few minutes to write and debug. The same code in MapReduce will span hundred lines of code and take hours to develop.

(4) – Pig provides high-level advantage that can be used by:

* Data Analyst
* Data Scientists

Q - 3: What is pig engine and What is its importance?

Ans: Apache Pig engine accepts the pig Latin scripts as input and converts those scripts into MapReduce jobs.

Pig Engine: Execution Environment to run Pig Latin programs. It has two modes.

Local Mode: Local execution in a single JVM, all files are installed and run using local host and file system.

MapReduce Mode: Distributed execution on a Hadoop cluster, it is the default mode.

Q-4: What are modes of pig execution?

Ans: Local Mode: Local execution in a single JVM, all files are installed and run using local host and file system.

MapReduce Mode: Distributed execution on a Hadoop cluster, it is the default mode.

Q-5: What is grunt shell in pig?

Ans: The Grunt shell of Apache Pig is mainly used to write Pig Latin Scripts.

It is interactive shell for executing Pig commands.

Used when script file is not provided.

Can execute scripts from Grunt via run or exec commands.

Q-6: What are the features of pig Latin Language?

Ans: (1) – Rich set of operators – It provides many operators to perform operations like join, sort, filer, etc.

(2) – Ease of programming – Pig Latin is similar to SQL and it is easy to write a pig script if you are good at SQL.

(3) -- Optimization Opportunities – The task in apache pig optimize their execution automatically, so the programmers need to focus only on semantics of the language.

(4) – Extensibility – Using the existing operators, users can develop their own functions to read, process, and write data.

(5) – 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.

(6) – Handles all kinds of data – Apache analyzes all kinds of data, both structured as well as unstructured. It stores the results in HDFS.

Q-7: Is Pig Latin command case sensitive?

Ans: The names (aliases) of relations and fields are case sensitive. The names of pig Latin functions are case sensitive.

The names (aliases) of relations A, B, and C are case sensitive.

The names (aliases) of fields f1, f2 and f3 are case sensitive.

Function names Pig Storage and COUNT are case sensitive.

Keywords LOAD, USING, AS, GROUP, BY, FOREACH, GENERATE, and DUMP are case insensitive. They can also be written as load, using, as, group, by, etc.

In the FOREACH statement, the field in relation B is referred to by positional notation ($0).

Q-8: What is a data flow language?

Ans: The dataflow programming is a programming paradigm that models a program as a directed graph of the data flow between operations, thus implementing dataflow principles and architecture. Dataflow programming languages share some features of functional languages, and were generally developed in order to bring some functional concepts to a language more suitable for numeric processing.