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

Apache Pig is a high-level platform for creating programs that run on Apache Hadoop. The language for this platform is called Pig Latin.

Pig Latin abstracts the programming from the Java MapReduce idiom into a notation which makes MapReduce programming high level, similar to that of SQL for RDBMSs.

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

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

PIG is a data flow language, the key focus of Pig is manage the flow of data from input source to output store. As part of managing this data flow it moves data feeding it to process1, taking output and feeding it to process2. The core features are preventing execution of subsequent stages if previous stage fails, manages temporary storage of data and most importantly compresses and rearranges processing steps for faster processing. While this can be done for any kind of processing tasks Pig is written specifically for managing data flow of Map reduce type of jobs. Most if not all jobs in a Pig are map reduce jobs or data movement jobs. Pig allows for custom functions to be added which can be used for processing in Pig, some default ones are like ordering, grouping, distinct, count etc.

An advantage PIG has over MapReduce is that the former is more concise. A 200 lines Java code written for MapReduce can be reduced to 10 lines of PIG code.

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

Apache Pig is an abstraction over MapReduce. It is a tool/platform which is used to analyze larger sets of data representing them as data flows.

Apache Pig has a component known as Pig Engine that accepts the Pig Latin scripts as input and converts those scripts into MapReduce jobs.

**Importance –**

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.

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

Pig has two execution modes or exectypes:

Local Mode - To run Pig in local mode, you need access to a single machine; all files are installed and run using your local host and file system. Specify local mode using the -x flag (pig -x local).

Mapreduce Mode - To run Pig in mapreduce mode, you need access to a Hadoop cluster and HDFS installation. Mapreduce mode is the default mode; you can, but don't need to, specify it using the -x flag (pig OR pig -x mapreduce).

**What is grunt shell in Pig?**

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

This document describes commands supported by grunt that 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 of them are discussed in the document.

This section describes currently available commands. The commands in each section are listed in alphabetical order. All commands are case insensitive and white spaces are not significant.

What are the features of Pig Latin language?

Apache Pig is a generic framework which consists of implementation of many MapReduce Design Pattens.

Apache Pig is implemented in Java Programming Language.

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

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.

**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. 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.

**What is a data flow language?**

In computer programming, 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 architecture.

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