**Problem Statement**

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

Pig is built on top of Hadoop, thus it needs Map Reduce program in order to operate. Pig is a scripting platform designed to process and analyze large datasets which was developed by Yahoo. Pig runs on Hadoop cluster. There are some reasons why pig needs Map Reduce-

1. Pig is built on top of Map reduce.
2. It runs on Hadoop clusters.
3. Pig uses both structured and unstructured data and uses HDFS to pick up and store results.
4. Pig provides an engine for executing data flows in parallel on Hadoop. It includes a language, pig Latin for expressing these data flows.
5. Pig runs on Hadoop. It makes use of both the Hadoop Distributed File System, HDFS and Hadoop Processing system, Map Reduce.
6. **What are advantages of pig over MapReduce?**

Pig was developed by non-Java programmers as Java is not a preferred choice of everyone. Pig provides many built-in functions instead of Map Reduce. Following are the advantages of Pig over Map Reduce.

* Pig Latin provides all of the standard data-processing operations, such as join, filter, group by, order by, union, etc.
* MapReduce requires programmers while pig is designed for non-programmers or who have a little knowledge of programming.
* Programmers must think in terms of map and reduce functions.
* Most probably Java programmers are required.
* 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 takes hours to develop.
* Pig provides high-level language that can be used by:

Data Analysts as well as Data Scientists

* It decreases the development time and complexity of Map reduce codes.
* Similar to SQL makes easy to learn and code.
* It is quite effective for unstructured and messy large datasets. Actually, Pig is one of the best tool to make the large unstructured data to structured data.

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

Pig Commands are passed from multiple stages when executing. It follows multiple steps like-

1. Parser
2. Optimizer
3. Compiler
4. Execution Engine

Parser is the first step where them pig commands are passed, It checks the syntax of the scripts and do type checking.

Optimizer generates the logical plan of execution.

Compiler compiles the optimized logical plan

And the final step is **Execution Engine**. Final Map Reduce jobs are submitted to Hadoop in stored order and executed on Hadoop to produce desired results.

Execution engine is responsible to run the job submitted by the client.

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

Pig can run in two modes- Local and Map Reduce mode

1. **Local Mode-** Pig runs on a single JVM picking Input and output from local file system. To run the script in local mode, the file should be kept in local file system. Pig can be switched to local mode by typing **pig –x local** in the grunt shell.
2. **Map reduce Mode-** Pig job runs as series of Map Reduce jobs picking Input and output from HDFS. To run the script in Map Reduce mode, the file should be kept in HDFS. Pig can be switched to Map Reduce mode by typing **pig –x mapreduce** or **pig** in grunt shell. While running pig in Map Reduce mode, job history server must be running.
3. **What is grunt shell in Pig?**

- Grunt is an interactive Shell for executing Pig Latin Commands.

- Grunt shell provides set of scripting commands.

- Used when script file is not provided

- Can execute scripts from Grunt via run or exec commands

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

* Pig Latin is a scripting language used to write Pig scripts in order to run on pig shell.
* Pig Latin commands can run on Pig shell which is called grunt shell.
* A pig Latin statement is an operator that takes a relation as input and produces another relation as output.
* Pig Latin statements can span multiple lines and must end with a semicolon.
* Pig Latin statements are generally organized in the following manner:

1. A LOAD statement reads data from the file system.
2. A series of "transformation" statements process the data.
3. A STORE statement writes output to the file system; or, a DUMP statement displays output to the screen.
4. **Is Pig Latin commands case sensitive?**

Yes, Pig Latin commands are case sensitive. The name or alias of relations and fields are case sensitive, name of Pig Latin functions are case sensitive and name of parameters and all other pig Latin keywords are case sensitive.

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

In Data Flow 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. A dataflow "if" statement would route the data to the correct branch.