**Problem Statement**

1. **Explain about the different complex data types in pig.**

Pig supports multiple data types called complex data types unlike Map Reduce only supports primitive data types. Some of Pig Complex data types are

1. **Atom** - Any single value in Pig Latin, irrespective of their data or type is known as an Atom. It is stored as bytearray by default and can be used as string or number like int, long, float, double, chararray, and bytearray are the atomic values of Pig. A piece of data or a simple atomic value is known as a field.

Example − ‘raja’ or ‘30’

1. **Tuple** - A record that is formed by an ordered set of fields is known as a tuple, the fields can be of any type. A tuple is similar to a row in a table of RDBMS. Example − (Raja, 30) 15 Complex Data Types (Contd.) • Bag • A bag is an unordered set of tuples. In other words, a collection of tuples (non-unique) is known as a bag. Each tuple can have any number of fields (flexible schema). A bag is represented by ‘{}’. It is similar to a table in RDBMS, but unlike a table in RDBMS, it is not necessary that every tuple contain the same number of fields or that the fields in the same position (column) have the same type.

Example − {(Raja, 30), (Mohammad, 45)}

1. **Bag**- A bag can be a field in a relation; in that context, it is known as inner bag. Example − {Raja, 30, {9848022338, raja@gmail.com,}} 16
2. **Map**- A map (or data map) is a set of key-value pairs. The key needs to be of type chararray and should be unique. The value might be of any type. It is represented by ‘[]’

Example: [name#Raja, age#30]

1. **Relation** - A relation is an outer bag of tuples. The relations in Pig Latin are unordered (there is no guarantee that tuples are processed in any particular order).
2. **How can you interact with the shell in Apache pig?**

Shell in Apache is called Grunt shell. Client can interact with Pig by using a Pig Latin scripting language which runs on this Grunt shell. Grunt shell is responsible to run all the pig Latin scripting commands and provides the result on the same shell.

1. **Explain how pig differs from Map reduce.**

* Hadoop Map Reduce is compiled language while Pig is a scripting language.
* Hadoop Map Reduce provides lower level of abstraction while Pig provides higher level of abstraction.
* Map Reduce have much more line of codes as compared to Pig.
* More development effort is involved in developing Map Reduce programs while pig requires very less development efforts.
* Code efficiency is high in Map Reduce when compared to Pig.
* Pig approaches are slower than a Hadoop Map Reduce program.
* For writing Map Reduce programs one need to be proficient in Java but Pig is quite easy to write.

1. **Explain how pig differs from SQL.**

* In SQL, when users want to do several data operations together, they must either write separate queries, storing the intermediate data into temporary tables, or write it in one query using subqueries inside that query to do the earlier steps of the processing.
* Pig, however, is designed with a long series of data operations in mind, so there is no need to write the data pipeline in an inverted set of subqueries or to worry about storing data in temporary tables.
* SQL is designed for the RDBMS environment, where data is normalized and schemas and proper constraints are enforced (that is, there are no nulls in places they do not belong, etc.).
* Pig is designed for the Hadoop data-processing environment, where schemas are sometimes unknown or inconsistent. Pig does not require data to be loaded into tables first. It can operate on data as soon as it is copied into HDFS.
* Pig Latin is the native language of parallel data-processing systems.

1. **Explain the scalar data types in pig.**

Scalar Data types are single data types which contains single values like int, float, long, chararray, bytearray etc.

1. Int- It is a data type which is written without fractional part.
2. Float- It is a data type which do have a fractional part.
3. Double- It also have fractional data type the only difference between float and double is that of range, range of double is much more than float.
4. Long- Long is same as int except of the range which is much more of long.
5. Bytearray- It is the default data type for Pig.
6. Chararray- It gives storage for group of character.