**Assignment 9.6**

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

**Map:** it's a chararray to data element mapping which is expressed in key-value pairs.

**Key** is always of type chararray and can be used as index to access the associated value. It is not necessary that all values in a map are of the same type.

**Tuple:** These are fixed length, ordered collection of Pig data elements. These contain fields which may be of different Pig types. It is possible to reference a field by its position in the tuple. A tuple can but is not required to declare a schema which describes each field's data type and provides a name for the field.

**Bag:** These are unordered collection of tuples. So we cannot reference a tuple in a bag by its position. Bags are also not required to declare a schema. In case of bags, schema describes all the tuples in the bag.

1. **How can you interact with the shell in Apache pig**

To interact with runt shell we can launch by using simple command "pig" or "pig -x local"."sh" is used to interact with Linux terminal from pig's grunt shell.

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

1) Development time is decreased in pig.

2) Learning curve is not steep (pig).

3) It is procedural (pig).

4) Execution can be controlled (pig).

5) Even effective for large and unstructured database (pig).

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

DBMS systems on which SQL operates are considered to be faster than MapReduce it is the loading of data that is more challenging in case of RDBMS making the set up difficult.

Pig Latin offers a number of advantages in terms of declaring execution plans, ETL routines and pipeline modification.

SQL is declarative and Pig Latin is procedural to a large extent. What we mean by this is in SQL. We largely specify “what” is to be accomplished and in Pig, we mention “how” a task is to be performed.

A script which is written in Pig is essentially converted to a MapReduce job in the background before it is executed. A Pig script is shorter than the corresponding MapReduce job, which significantly cuts down development time.

1. **Explain the scalar data types in pig**
2. **Int:** (an Integer) these are represented in interfaces by java.lang.Integer.
3. **Long:** (long integer) Long are represented in interfaces by java.lang.Long. They store a eight byte signed integer. Constants are expressed as integer numbers with a L appended, for example 34L.
4. **Float:** (floating point number) Floats are represented in interfaces by java.lang.Float. They store a four byte floating point number. Constants are represented as floating point numbers with f appended, for example, 2.18f.
5. **Double:** (double precision floating point number) Doubles are represented in interfaces by java.lang.Double. They store a eight byte floating point number. Constants are represented either as floating point numbers or in exponent notation, for example, 32.12567 or 3e-17.
6. **Chararray:** (string or array of characters) Represented in interfaces by java.lang.String. Constant chararrays are represented by single quotes, for example, 'constant chararray'.
7. **Bytearray:** (blob or array of bytes) Represented by java class DataByteArray which wraps a java byte[]. There is no way to specify a byte array constant.