**#Assignment-9.6**

**Que.1) Explain About complex data types in Pig?**

**Tuple**

A record that is formed using an ordered set of fields is known as a tuple, the fields can take any type and is very similar a row in a table of RDBMS.

Example − (Tejas, 150)

**Atom**

Every value in Pig Latin, irrespective of their data type is called as an **Atom**. It is stored as string and can be used as string as well as number. Double, char array, byte array, int, long, float are atomic values of Pig.

Example − ‘Tejas’ or ‘150’

**Bag**

A bag is an unordered set of tuples. In other words, a collection of tuples is known as a bag. Each tuple can have any number of fields. 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 column have the same type.

Example − {(Tejas, 150), (XYZAB, 123)}

When a bag have a field in a relation then is known as inner bag.

Like**,** {Tejas, 1500, **{612836, TEJASJS2812@gmail.com}**}

**Map**

A map is a set of key-value pairs. The key needs to be of type char array and should be unique. The value might be of any type. It is represented by ‘[]’

Example − [name#TEJAS, score#30]

**Relation**

A relation is a bag of tuples. The relations in Pig Latin are unordered means there is no guarantee that tuples are processed in any particular order.

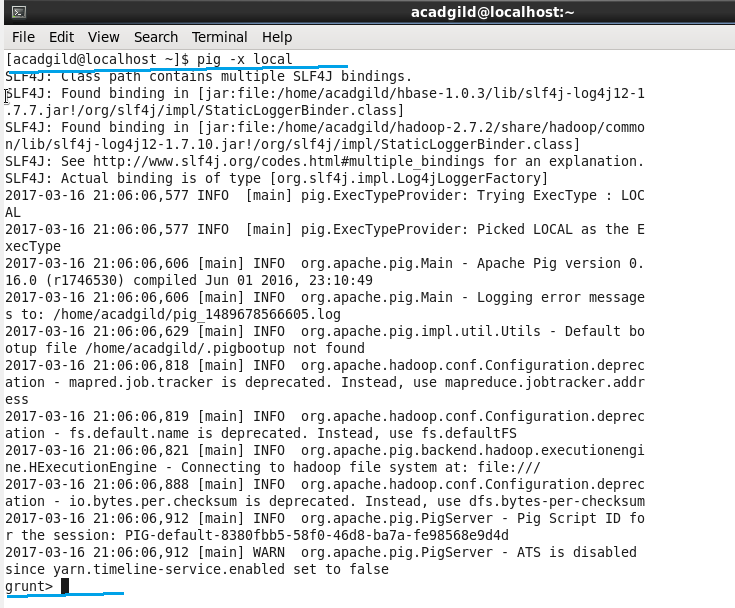
**Que.2) how can you interact with the shell in apache pig?**

To interact with the shell present in Apache Pig, First we will have to invoke grunt shell from the shell we are present at. It could be Local or Map Reduce. It is achievable by **–x** command. Like,

**$. /pig –x local**

**$. /pig -x mapreduce**

After invoking the Grunt shell, you can run your Pig scripts in the shell.



In above screenshot you can see first we were working local and then we invoked Apache pig shell from local.

**Que.3) how apache Pig is different from MapReduce?**

There are several ways pig is different from map reduce, though both perform same task but pig is way more accessible and short to code. Following the major differences:

1) Apache pig is data flow language while MapReduce is data processing paradigm.

2) Pig is high level and flexible language while MapReduce is low level and rigid.

3) In Apache Pig performing join operation on data sets is pretty simple.

4) Basic knowledge of SQL Query is sufficient in Apache while to perform MapReduce task JAVA coding is needed.

5) Pig has multi-query approach hence length of code is reduced to great extent, while in MapReduce we need to write almost 15-20 times more lines.

6) No Compilation is needed as each operator in Pig is converted to MapReduce Job, in MapReduce Jobs compilation is must.

**Que.4) Explain How Pig differs from SQL?**

Apache Pig and SQL are definitely not same. Pig make use of few SQL commands to optimize the work.

1. Pig uses multiple query approach which reduces data scan iterations. In SQL we specify **“WHAT”** to do where as in Pig specify **“HOW”** the task is performed.
2. Pig is majorly **procedural language** while SQL is **declarative** language.
3. Designing Schema is not mandatory in Pig, while in SQL it is important to have a Schema structure.
4. Data model in Pig is nested relational while in SQL it is flat relational.
5. Apache pig provide less opportunity for **Query Optimization** while in SQL scope is to larger extend.