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

Ans—

Atom- It is stored as byte array by default and can be used as string or number like int, long, float, double, chararray, and byte array are the atomic values of Pig.

Tuple- A record that is formed by an ordered set of fields is known as a tuple, the fields can be of any type.

Bag - A collection of tuples (non-unique) is known as a bag.

Map - A map (or data map) is a set of key-value pairs.

Relation - A relation is an outer bag of tuples.

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

Ans—

Grunt shell is by which one can communicate with the pig engine. This is place one can write the pig Latin script language and grunt shell communicates with system kernel to give these coded lies to the pig engine in the language it understands.

1. Explain how pig differs from Map reduce ?

Ans—

PIg just a higher abstraction of the map reduce

1. Less complex codes to type
2. Easy to learn pig latin than that of Java.
3. Many built in functions.
4. People of non-programming background an also use pig easily.
5. Map reduce takes more time construct the code.
6. 1/20 lines of code are needed than that map reduce.
7. Explain how pig differs from sql ?

Ans –

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 ?

Ans –

Int: An integer. Int are represented in interfaces by java.lang.Integer. They store a four byte signed integer. Constant integers are expressed as integer numbers.

Long: A long integer. Long are represented in interfaces by java.lang.Long. They store eight byte signed integer. Constants are expressed as integer numbers with L appended, for example 34L.

Float: A 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.

Double: A 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.

Chararray: A string or array of characters. Represented in interfaces by java.lang.String. Constant chararrays are represented by single quotes, for example, 'constant chararray'.

Bytearray: A blob or array of bytes. Represented by java class DataByteArray which wraps a java byte[]. There is no way to specify a bytearray constant.