**Assignment 23.2**

**Problem Statement:**

Explain Primary data types and complex data types in Hive with an example in brief.

**Answer:**

**Primary Data types in Hive:**

Primary Data Types are classified into four categories. They are:

• Numeric Types

• String Types

• Date/Time Types

• Miscellaneous Types

**Numeric Data Types:**

* *Integral types* are – TINYINT, SMALLINT, INT & BIGINT( Equivalent to Java’s byte , short , int , and long primitive types)
* *Floating types* are – FLOAT, DOUBLE & DECIMAL. (Equivalent to Java’s float and double , and SQL’s Decimal respectively.)
* DECIMAL(5,2) represents total of 5 digits, out of which 2 are decimal digits.
* ***For Example:*** 100Y – TINYINT, 100S – SMALLINT, 100L – BIGINT

**String Data Types:**

* *String* -- String literals can be expressed with either single quotes (') or double quotes (")
* *Varchar* --Varchar types are created with a length specifier (between 1 and 65355), which defines the maximum number of characters allowed in the character string.
* *Char*-- Char types are similar to Varchar but they are fixed-length meaning that values shorter than the specified length value are padded with spaces but trailing spaces are not important during comparisons.
* ***For Example:*** name VARCHAR(64)). If the values are less than the max length specified then remaining space will be freed out.

**Date/Time Types**:

* Hive provides DATE and TIMESTAMP data types in traditional UNIX time stamp format for date/time related fields in hive.
* DATE values are represented in the form YYYY-MM-DD.
* We can also cast the String, Time-stamp values to Date format if they match format.
* ***For Example***: DATE ‘2014-12-07’. Date ranges allowed are 0000-01-01 to 9999-12-31.

**Miscellaneous Types:**

* Hive supports two more primitive data types, BOOLEAN and BINARY. Similar to Java’s Boolean, BOOLEAN in hive stores true or false values only.
* BINARY is an array of Bytes and similar to VARBINARY in many RDBMSs.

**Complex Data types in Hive:**

Complex Types can be built up from primitive types and other composite types. Data type of the fields in the collection are specified using an angled bracket notation. Currently Hive supports four complex data types. They are:

* Array
* Map
* Struct
* UnionType

**ARRAY:**

* An Ordered sequences of similar type elements that are indexable using

zero-based integers.

* It is similar to arrays in Java.
* ***For Example*** – array (‘siva’, ‘bala’, ‘praveen’); First element is accessed with array[0];

**MAP:**

* Collection of key-value pairs.
* Fields are accessed using array notation of keys (e.g., [‘key’]).
* ***For Example:***’one’->’aju’,’two’->’roll’ is represented as map(‘one’,’aju’,’two’,’roll’).here ‘aju’ can be accessed with map[‘one’].

**STRUCT:**

* It is similar to STRUCT in C language.
* It is a record type which encapsulates a set of named fields that can be any primitive data type.
* Elements in STRUCT type are accessed using the DOT (.) notation.
* ***For Example*** – For a column c of type STRUCT {a INT; b INT} the a field is accessed by the expression c.a.

**UNIONTYPE:**

* It is similar to Unions in C.
* At any point of time, an Union Type can hold any one (exactly one) data type from its specified data types.
* ***For Example:*** UNIONTYPE<int, double, array<string>, struct<a:int,b:string>>.