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(10,2) will represent a total of 10 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 by single quotes (') or double quotes (") .
* *Varchar* -- Varchar are created with the help of a length specifier (between 1 and 65355), which defines the maximum number of characters allowed in the character string.
* *Char*-- Char types will be similar to the Varchar
* Except they will be of fixed-length meaning whose values will be shorter than the specified length value and will be padded with the help of spaces but trailing spaces are not considered as important during the period of comparisons.
* ***For Example:*** name VARCHAR(64)). If the given values are lesser than the maximum length specified then remaining space will be emptied.

**Date/Time Types**:

* Hive is provided with DATE and TIMESTAMP data types in traditional UNIX time stamp format for date/time related fields in hive.
* DATE will be represented in the form of YYYY-MM-DD.
* Time-stamp values can be changed to Date format if they match with the given format.
* ***For Example***: DATE ‘1993-10-29’.
* Date ranges is allowed are 0000-01-01 to 9999-12-31.

**Miscellaneous Types:**

* Hive supports 2 more primitive data types :

**1.BOOLEAN**

**2.BINARY**

* It is similar to Java’s Boolean.
* BOOLEAN in hive will store only true or false values.
* BINARY - array of Bytes which will be similar to VARBINARY which is in RDBMSs.

**Complex Data types in Hive:**

* Complex Types will be built of primitive types and other composite types.
* Data type of the fields in collection will be specified by using angled bracket notation.
* Hive will support 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 (‘mathurri’, ‘sai’, ‘tamil’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’->’suresh’,’two’->’roll’ is represented as map(‘one’,’suresh’,’two’,’roll’).here ‘suresh’ can be accessed with map[‘one’].

**STRUCT:**

* It is similar to STRUCT which is in C language.
* It is a record type which will encapsulate 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.
* Union Type will hold exactly one data type at a time from its specified data types.
* ***For Example:*** UNIONTYPE<int, double, array<string>, struct<a:int,b:string>>.