**ASSIGNMENT 23.2**

**QUES1)**

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

# **Hadoop Hive Data Types With Examples**

Hadoop [**Hive Data Types**](http://www.hadooptpoint.com/?s=Hive+Data+Types) With Examples are mainly divided into 2 types

* **Primitive Data Types**
* [**Complex Data Types**](http://www.hadooptpoint.com/hive-complex-data-types-example/)

## Primitive Data Types

primitive Data Types also divide into 4 types there are mentioned below

* Numeric Data Type
* String Data Type
* Date/Time Data Type
* Miscellaneous Data Type

Here is the common thing between Java,Sql and Hive is The Data types and sizes of [**Hadoop**](http://www.hadooptpoint.com/) hive data types is similar to Java and SQL Primitive Data types

### Numeric Data Type

Numerical Data types are mainly divided into 2 types

* **Integral Data Types**
* **Floating Data Types**

##### **Floating Data Types**

* FLOAT (This FLOAT is equal to Java’s FLOAT data type )
* DOUBLE (This DOUBLE is equal to Java’s DOUBLE data type)
* DECIMAL (This DECIMAL is equal to SQL’s DECIMAL data type)

DECIMAL(4,3) represents total of 4 digits, out of which 3 are decimal digits. Below is the chart for all numeric types with their ranges and examples.

By default in Hive integral values are taken as a INT data type unless the value of integral values range are cross the range of INT values (Please check the above table for values).Suppose if we need to use a low integral value like 100 it to be treated as TINYINT or SMALLINT or BIGINT then we need to suffix the value with Y, S or L respectively.

**Examples:** 100Y – TINYINT, 100S – SMALLINT, 100L – BIGINT

##### Integral Data Types

* TINYINT (This TINYINT is equal to Java’s BYTE data type)
* SMALLINT (This SMALLINT is equal to Java’s SHORT data type)
* INT (This INT is equal to Java’s INT data type)
* BIGINT (This BIGINTis equal to Java’s LONG data type)

|  |  |  |
| --- | --- | --- |
| **Type** | **Postfix** | **Example** |
| TINYINT | Y | 10Y |
| SMALLINT | S | 10S |
| INT | - | 10 |
| BIGINT | L | 10L |

## String Data Types

In Hive String Data Types are Mainly divided into 3 types there are mentioned below

* STRING
* VARCHAR
* CHAR

This String Data types are supported in Hive from only 0.14 and above versions only String type data types can be specified using single quotes (' ') or double quotes (" "). It contains two data types: VARCHAR and CHAR. Hive follows C-types escape characters.

The following table depicts various CHAR data types:

|  |  |
| --- | --- |
| **Data Type** | **Length** |
| VARCHAR | 1 to 65355 |
| CHAR | 255 |

### What is Difference between CHAR vs VARCHAR

* CHAR is fixed length
* VARCHAR is of variable length but we need to specify the max length of the field (example : name VARCHAR(64)). If the values are less than the max length specified then remaining space will be freed out.
* Maximum value of CHAR is 255 but Maximum value of VARCHAR is 65355
* In VARCHAR we can optimized the Space/Storage by using released the unused bytes but in CHAR we can optimize the unused Space CHAR unused bytes will  not be released but filled with spaces.
* If a string value being assigned to a VARCHAR value exceeds the length specified, then the string is silently truncated.

## DATE/TIME Data Types

Date/Time Data types are mainly Divide into 2 types

* DATE
* TIMESTAMP

This hive date/time data types in UNIX time stamp format for date/time related fields in hive.

#### **DATE**

Represented format is YYYY-­MM-­DD

**Example  : DATE ‘2014-­12-­07’**

Date ranges allowed are 0000-­01-­01 to 9999-­12-­31

#### **TIMESTAMP**

**TIMESTAMP** use the format yyyy-mm-dd hh:mm:ss[.f...].

|  |  |
| --- | --- |
| Cast Type | Result |
| cast(date as date) | Same date value |
| cast(date as string) | Date is formatted as a string in the form ‘YYYY-MM-DD’. |
| cast(date as timestamp) | Midnight of the year/month/day of the date value is returned as timestamp. |
| cast(string as date) | If the string is in the form ‘YYYY-MM-DD’, then a date value corresponding to that is returned. If the string value does not match this format, then NULL is returned. |
| cast(timestamp as date) | The year/month/day of the timestamp is returned as a date value. |

## Miscellaneous Types

Hive Supports 2 more primitive Data types

* BOOLEAN
* BINARY

Hive BOOLEAN is similar to Java’s BOOLEAN types,it can stores true or false values only

**BINARY** is an array of Bytes and similar to VARBINARY in many RDBMSs. BINARY columns are stored within the record, not separately like BLOBs

## Implicit Conversion Between Primitive Data Types

TINYINT—>SMALLINT–>INT–>BIGINT–>FLOAT–>DOUBLE

TINYINT can be converted to any other numeric data type but BIGINT can only be converted to FLOAT or DOUBLE but the reverse

Boolean & Binary data types will not be  converted to any other data type implicitly.

## Explicit Conversion Between Primitive Data Types

Explicit type conversion can be done using the cast operator only.

**Example:** CAST(‘500’ AS INT) will convert the string ‘500’ to the integer value 500.  But If cast is used incorrectly as in CAST(‘Hello’ AS INT) , then cast operation will fail and returns NULL .

## Complex Data Types

Hive supports 4 types of complex data types

* ARRAY
* MAP
* STRUCT
* UNIONTYPE

Hive Complex Data Types Example

ARRAY Data Type

Same as Array in java,An Ordered sequences of similar type elements that are index using zero-based integers

Example                      array(‘JINI’, ‘MINI’)

The second value is accused by using array[1]

MAP Data Type

Collection of key-value pairs. Fields are accessed using array notation of keys (e.g., [‘key’]).

Example                      map(‘first’, ‘RAM’,’last’, ‘SHAM’)

if a column  name is of type MAP with key→value pairs  ‘first’→’RAM’ and ‘last’→’SHAM’, then the last  name can be referenced using name[‘last’].

STRUCT Data Type

* 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.

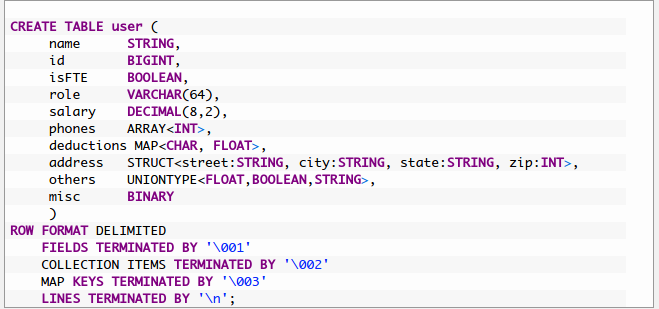
Example – For a column c of type STRUCT {a INT; b INT} the a field is accessed by the expression c.a

UNIONTYPE

UNIONTYPE is collection of Heterogeneous data types.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

Example

UNIONTYPE<int, double, array<string>, struct<a:int,b:string>>

[](http://www.hadooptpoint.com/wp-content/uploads/2015/09/hIVE-complex-data-types-example.png)

This is the detailed information about [Hadoop Hive](http://www.hadooptpoint.com/?s=hadoop+Hive) Data Types With Examples