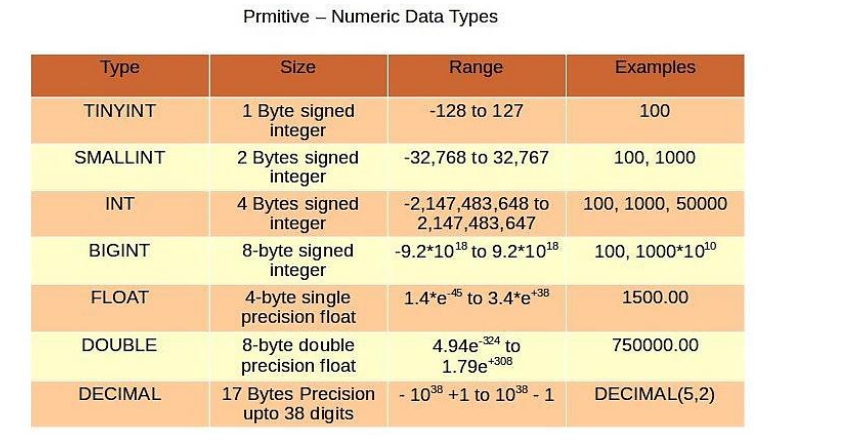
**Problem Statement**

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

Hive Data Types are used to specify the field type. Hive supports two basic types of data type.

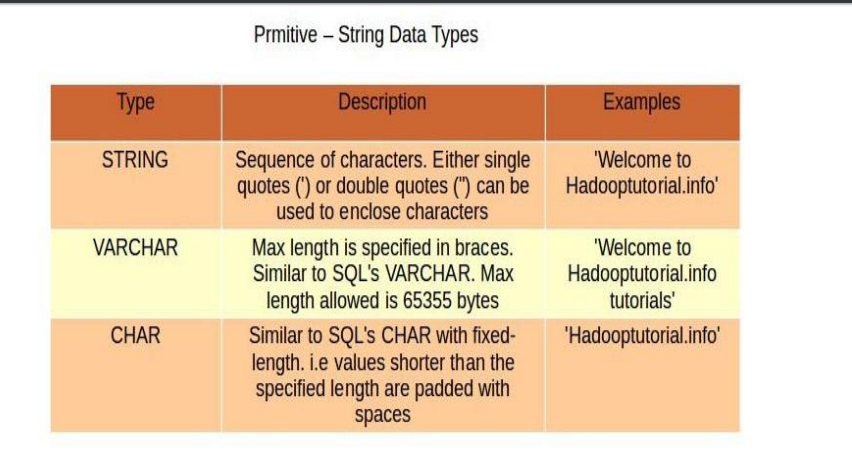
1. Primary Data Type- Primary Data Types are further classified into four categories-
2. Numeric Types-

* 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.



1. String Types-

* 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.



CHAR Vs. VARCHAR-

• CHAR is of fixed length and values that are shorter are padded with spaces.

• 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 the remaining space will be freed out.

• The maximum length of CHAR is 255 but VARCHAR can be up to 65355 bytes.

• Space/storage optimization is done in VARCHAR by releasing the unused bytes but in 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.

1. 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. Example: DATE ‘2014-12-07’. Date ranges allowed are 0000-01-01 to 9999-12-31.
* TIMESTAMP use the format yyyy-mm-dd hh:mm:ss[.f...].
* We can also cast the String, Time-stamp values to Date format if they match format.

1. 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.

1. Complex Data Type- In addition to primitive data types, Hive also support complex data types also known as collection data types. Complex types can be built from primitive types and other composite types. Data type of the fields in the collection are specified using an angled bracket notation.
2. Array-

* An Ordered sequences of similar type elements that are indexable using zero-based integers.
* It is similar to arrays in Java.
* Example – array (‘siva’, ‘bala’, ‘praveen’); • Second element is accessed with array[1].

1. Map-

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

1. 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. Example – For a column c of type STRUCT {a INT; b INT} the a field is accessed by the expression c.a 18 Complex Data Types.

1. Union Type-

* 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.