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

Answer

Hive Data types are broadly divided in two types

1. Primary Data types
2. Complex Data types

Primary Data types are further divided into different types

• Numeric Types

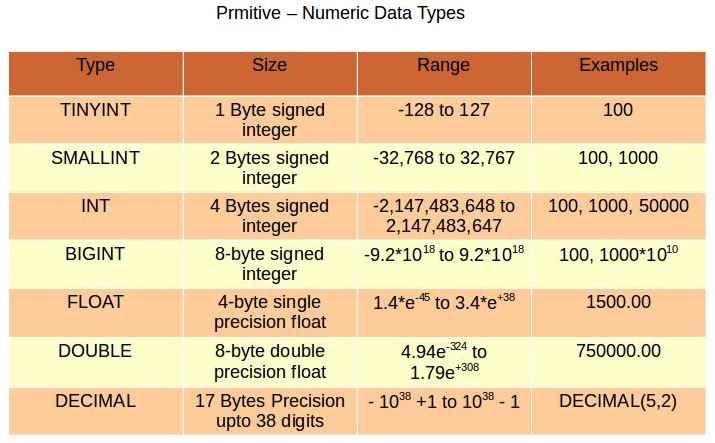
• String Types

• Date/Time Types

• Miscellaneous Types

Numeric Types - Integral types are – TINYINT, SMALLINT, INT & BIGINT etc.

The Types and their ranges are shown in the table below.



By default hive considers the INT for the numerical values unless it is not crossing the INT limit as shown in the above table. But if we want to use the other form of the numerical data types then we have to suffix letter such as T, S, and L etc.

e.g.

100Y – TINYINT, 100S – SMALLINT, 100L – BIGINT

String Types - There three types of String data types are similar to that are being used in other coding languages.

* STRING
* VARCHAR
* CHAR

String - Sequence of characters and either quotes or double quotes can be used to enclose the characters

CHAR - Similar to SQL characters and the size is specified in in the braces.

VARCHAR – Similar to SQL VARCHAR and the maximum size is specified in the braces.

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.

Miscellaneous types -

Hive supports other types such as that of in the java i.e. Boolean and binary

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

Hive Presently supported 4 complex data types

1. ARRAY
2. MAP
3. STRUCT
4. UNIONTYPE

ARRAY – It is Similar to Array in java. An Ordered sequences of similar type elements that are index using zero-based integers

Array(‘andy’,’jhon’) this is the array of the Strings.

Map – It is similar to the Map in java i.e. Collection of key-value pairs

Example-

map(‘first’, ‘John’,’last’, ‘Doe’)

Linking is done as follows-

‘first’→’John’ and ‘last’→’Doe’ .

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}

UNION

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>