

# Apache Hive Data Types



The data type is an attribute of data that is used to define the type of value a column stores in the Hive table. Hive provides various data types such as Numeric, Date/time, String and so on.

Apache Hive provides the following list of Datatypes.

1. Numeric Types
2. Date/Time Types
3. String Types
4. Miscellaneous Types
5. Complex Types

Let us see each data type in detail.

## 1. Numeric Types

Apache Hive provides the below set of Date/Time data type.

Data Type	Description
TINYINT	It is 1-byte signed integer, range from -128 to 127.
SMALLINT	It is 2-byte signed integer, range from -32,768 to 32,767.
INT/INTEGER	It is 4-byte signed integer, range from -2,147,483,648 to 2,147,483,647.
BIGINT	It is 8-byte signed integer, range from -9,223,372,036,854,775,808 to 9,223,372,036,854,775,807.
FI. OAT	It is 4-byte single-precision floating-point number
DOUBLE	It is an 8-byte double-precision floating-point number.
DOUBLE PRECISION	It is an alias for DOUBLE, only available starting with Hive 2.2.0.
DECIMAL	It was Introduced in Hive 0.11.0 with a precision of 38 digits.
NUMERIC	It is the same as DECIMAL, starting with Hive 3.0.0.

## 2. Date/Time Types

Apache Hive provides the below set of Date/Time data type.

### 2.1 Timestamps

Timestamps support traditional UNIX timestamps with optional nanosecond precision. For text files Timestamps supports yyyy-mm-dd hh:mm:ss[f..] format. It was introduced in Hive 0.8.0.

#### Timestamps Supported Conversions:

**Integer numeric types:** It is interpreted as a UNIX timestamp in seconds.

**Floating-point numeric types:** It is interpreted as a UNIX timestamp in seconds with decimal precision.

**Strings:** It is JDBC compliant java.sql.Timestamp format "YYYY-MM-DD HH:MM: SS. fffffff" (9 decimal place precision).

### 2.2 Dates

The DATE datatype of Hive represents the date in the format of year/month/day(yyyy-mm-dd). It won't have time for the day component. The Date type range value is between 0000-01-01 to 9999-12-31.

### 2.3 Interval

Interval data type can be used by specifying Intervals of time units such as SECOND / MINUTE / DAY / MONTH / YEAR. It was introduced in Hive 1.2.0.

### **3. String Types**

Apache Hive provides the below list of String data types.

#### **3.1 Strings**

The string literals are enclosed with either single quotes or double;e quotes in Apache Hive.

#### **3.2 Varchar**

The varchar data type is in the range of 1 and 65535 that defines the max character string allowed for a string.

#### **3.3 Char**

The Char data type is identical to Varchar but it has a fixed length which means if a value is taking less space than the defined length then space will be added. The max length supported by char is 255.

### **4. Miscellaneous Types**

ApacheHive provides the below list of Miscellaneous data types.

#### **4.1 Boolean**

The boolean data type is either True or False. It is similar to Java's Boolean.

#### **4.2 Binary**

The binary data type is an array of Bytes. It is similar to VARBINARY in many RDBMS.

## 5. Complex Types

### 5.1 Array

An array is a collection of items of a similar data type. It can contain one or more values of the same data type.

If we define the below array and want to access the first element that is "Cloudduggu" then we can use `array[0]`. `ARRAY('Cloudduggu', 'Hive')`.

### 5.2 Map

The map is a collection of key-value pairs where fields are accessed using array notation of keys(e.g. ['key']).

If we define a map such as 'Firstname' -> 'Sarvesh', 'Lastname' -> 'Kumar' then it will be presented like a `map('Firstname', 'Sarvesh', 'Lastname', 'Kumar')` and if you want to access the value of Sarvesh then you can use `map['firstname']`.

### 5.3 Struct

A struct is a record type that encapsulates a set of named fields which can be any primitive data type.

If we define a structure like `STRUCT {x INTEGER; y INTEGER}` for z column and wants to access x value then it can access as `z.x`.

