Chapter 2: Introduction to SQL

Analytics Tensor
Mahesh KC
mahesh.kc@analyticstensor.com
https://analyticstensor.com

Introduction

- SQL
- Database Vendor
- Types of SQL Statement
- Data Types
  - Integer Types
  - Floating-Point Types
  - String Types
  - Text Types
  - Date and Date Time Types
- Wrap-up

SQL

SQL is a language for manipulating the data in relational tables. It was renamed from SEQUEL. Many people refer it as Structured Query Language. SQL is a nonprocedural language, which means that you don't have complete control like you have on procedural language such as Java, C#, C, VB etc. where the code is complied and executed to produce exact result.

In SQL, you define input and output but the statement execution is handled by the database engine known as optimizer. To write complete application in SQL, you need to write script. For example PL/SQL, Stored Procedure, Transact-SQL Language for Oracle, MySQL and Microsoft respectively.

Database Vendor

Oracle: Oracle Database
Microsoft: SQL Server
IBM: DB2
Sybase: Sybase Adaptive Server
Teradata
PostgreSQL
MySQL

All of the above database server perform same task. They all comply with ANSI SQL standard. In this course we use **MySQL.** It is an open source database server.

Types of SQL Statement

1. Data Definition Language (DDL) Statements
2. Data Manipulation Language (DML) Statements
3. Transaction Control Language (TCL) Statements

Types of SQL Statement (cont.)

DDL: The *Data Definition Language (DDL)* statements consists of all the SQL statements that affect the structure of database objects, such as tables, indexes, and views. The DDL statements includes :

ALTER, CREATE, DROP, RENAME and TRUNCATE

DML: The *Data Manipulation Language (DML)* statements is used to access and manipulate/change data in existing schema objects. The DML statement includes:

CALL, DELETE, INSERT, LOAD, REPLACE, SELECT, UPDATE and WITH

TCL: The *Transaction Control Language (TCL)* statements manage changes made by DML statements. The TCL statement includes:

START TRANSACTION, COMMIT, ROLLBACK, LOCK TABLES, UNLOCK TABLES, SAVEPOINT and SET TRANSACTION.

Data Types

Data type defines the type of data a column contains. There are various data types supported by all the RDBMS such as:

- Integer Types
- Floating-Point Types
- String Types
- Text Types
- Date and Date Time Types

Integer Types

Type	Storage (Bytes)	Signed range	Unsigned range
tinyint	1	−128 to 127	0 to 255
smallint	2	−32,768 to 32,767	0 to 65,535
mediumint	3	−8,388,608 to 8,388,607	0 to 16,777,215
int	4	−2,147,483,648 to 2,147,483,647	0 to 4,294,967,295
bigint	8	-2^{63} to $2^{63}-1$	0 to $2^{64}-1$

Integer is a numeric type that are used to store whole numbers. Unsigned data type specifies that the data is greater than or equal to zero.

SERIAL refers to BIGINT UNSIGNED NOT NULL AUTO_INCREMENT UNIQUE.

SERIAL DEFAULT VALUE refers to NOT NULL AUTO_INCREMENT UNIQUE.

BIT: A bit-values type indicate the number of bits per values from 1 to 64.

BOOLEAN: It is synonyms for TINYTINT(1). It stores TRUE or FALSE which represents 1 or 0 resp.

Floating-Point Types

Type	Numeric range
float(p , s)	−3.402823466E+38 to −1.175494351E-38 and 1.175494351E-38 to 3.402823466E+38
double(p , s)	−1.7976931348623157E+308 to −2.2250738585072014E-308 and 2.2250738585072014E-308 to 1.7976931348623157E+308

Floating-point requires **precision** which is the total number in both left and right of the decimal point, and **scale** which is the digits right of the decimal point. For example, float(4,2) will store a total of four digits, two to the left of decimal and two to the right of decimal like 20.11, 1.45 but 12.3456 will give error.

String Types

CHAR and VARCHAR: Both char and varchar are declared with a length which specifies the maximum number of characters it stores. But they differ in maximum length and trailing spaces.

CHAR: The length of CHAR column is fixed to the length while creating table. The length can be any value from 0 to 255. Char values are right-padded with spaces to the specified length while storing and trailing space are remove while retrieving. e.g. char(20) stores up to 20 characters.

VARCHAR: Varchar are variable-length strings. The length of values can be from 0 to 65,535. The maximum length of varchar is 65,535 bytes. Varchar values are not padded while storing and trailing spaces are retained when storing and retrieving. e.g. varchar(8000) stores up to 8000 characters.

The varchar columns limit up to 64 KB.

Text Types

Text type	Maximum number of bytes
tinytext	255
text	65,535
mediumtext	16,777,215
longtext	4,294,967,295

Other Types:

BINARY, VARBINARY, TINYBLOB, BLOB, MEDIUMBLOB, LONGBLOB, ENUM, SET

ENUM: An enumeration type store only one value from the list of values. ENUM column can have a maximum of 65,535 distinct elements.

Text types are used to store data more than 64 KB(limit for varchar). Thing to consider while using text types:

- The data will be truncated if the text column exceeds the maximum size.
- Trailing space will not be remove when data is loaded.
- Text columns only uses first 1,024 bytes while sorting or grouping.

Date and Time Types

Type	Default format	Allowable values
date	YYYY-MM-DD	1000-01-01 to 9999-12-31
datetime	YYYY-MM-DD HH:MI:SS	1000-01-01 00:00:00 to 9999-12-31 23:59:59
timestamp	YYYY-MM-DDHH:MI:SS	1970-01-01 00:00:00 to 2037-12-31 23:59:59
year	YYYY	1901 to 2155
time	HHH:MI:SS	–838:59:59 to 838:59:59

The date and time types are represented as temporal data. Temporal data are used to store information such as:

- Date of birth
- Shipping order datetime
- Sales year

Date and Time Types (cont.)

Component	Definition	Range
YYYY	Year, including century	1000 to 9999
MM	Month	01 (January) to 12 (December)
DD	Day	01 to 31
HH	Hour	00 to 23
HHH	Hours (elapsed)	–838 to 838
MI	Minute	00 to 59
SS	Second	00 to 59

Various components of date formats.

Wrap-up

- Important points.
- Q & A.

Assignment-3

1. Describe NULL and Literal Values.
2. Describe ACID properties of transactions.
3. Understand and Memorize the table.

Table: Special character escape sequences.

Escape Sequence	Character Represented by Sequence
\0	An ASCII NUL (X'00') character
\'	A single quote (') character
\"	A double quote (") character
\b	A backspace character
\n	A newline (linefeed) character
\r	A carriage return character
\t	A tab character
\Z	ASCII 26 (Control+Z)
\\	A backslash (\) character
\%	A % character
_	A _ character