Database Management Systems:
Fundamentals and Introduction to SQL

MySQL Data Types and Applications

MySQL Data Types

- String types
- Numeric types
- Date and time types
- Special types, e.g. JSON, spatial data

Strings

- Binary
  - sequence of bytes
  - raw data: images, music files, or encrypted values
  - *BINARY, VARBINARY, and BLOB*
- Nonbinary
  - character, text
  - *CHAR, VARCHAR, and TEXT*
  - character set – *characters can be stored in the string*
  - collation – *character ordering*

Binary data type	Nonbinary data type	Maximum length
BINARY	CHAR	255
VARBINARY	VARCHAR	65,535
TINYBLOB	TINYTEXT	255
BLOB	TEXT	65,535
MEDIUMBLOB	MEDIUMTEXT	16,777,215
LONGBLOB	LONGTEXT	4,294,967,295

Strings – Nonbinary

Value	CHAR(4)	Storage Required	VARCHAR(4)	Storage Required
''	' '	4 bytes	''	1 byte
'ab'	'ab '	4 bytes	'ab'	3 bytes
'abcd'	'abcd'	4 bytes	'abcd'	5 bytes
'abcdefgh'	'abcd'	4 bytes	'abcd'	5 bytes

- CHAR
  - CHAR(x) fixed length
- VARCHAR
  - VARCHAR(x) variable length (max: 65,535 characters)
  - VARCHAR(50) for short strings, e.g. name
  - VARCHAR(255) for medium-length string, e.g. address
- TEXT
  - TINYTEXT (max: 255 bytes)
  - TEXT (max: 64 KB)
  - MEDIUMTEXT (max: 16 MB) for JSON and CSV files and short/medium books
  - LONGTEXT (max: 4GB) for textbooks

Strings Bytes for Different Language

- English: 1 byte
- European/Middle Eastern: 2 bytes
- Asian: 3-4 bytes

Strings Operations

- converting lettercase
  - UPPER(), LOWER()
- substring, combine
  - LEFT(str, n), RIGHT(str, n), MID(str, n, m) or SUBSTR()
  - locate (c, str)
  - CONCAT(), CONCAT_WS(separator,str1,str2,...)
  - REPLACE(str, from_str, to_str)
- pattern-matching
  - [NOT] LIKE %, _
  - regular expression - REGEXP
    - string REGEXP pattern

Regular Expressions

Pattern	What the pattern matches
^	Beginning of string
$	End of string
.	Any single character
[...]	Any character listed between the square brackets
[^...]	Any character not listed between the square brackets
p1|*p2*|*p3*	Alternation; matches any of the patterns *p1*, *p2*, or *p3*
*	Zero or more instances of preceding element
+	One or more instances of preceding element
{*n*}	*n* instances of preceding element
{*m*,*n*}	*m* through *n* instances of preceding element

Literals escape sequences

\b (backspace), \n (newline, also called linefeed), \r (carriage return), \t (tab)

Regular Expression Functions

Name	Description
NOT REGEXP	Negation of REGEXP
REGEXP	Whether string matches regular expression
REGEXP_INSTR()	Starting index of substring matching regular expression
REGEXP_LIKE()	Whether string matches regular expression
REGEXP_REPLACE()	Replace substrings matching regular expression
REGEXP_SUBSTR()	Return substring matching regular expression
RLIKE	Whether string matches regular expression

REGEXP examples

- `SELECT field FROM table WHERE field REGEXP pattern`

- pattern = "^(https?://|www\\.)[A-Za-z0-9\-] +\\.[a-zA-Z\/\?]{2,4}"

  - www.google.com
  - http://google.com/
  - http://www.google.net/
  - www.google.com/index.php?test=data
  - https://yahoo.dk/as
  - http://goo123.gle.com/
  - www.website.info

Limitations of Pattern Match

- Pattern matches enable you to look through any number of rows
- As the amount of text goes up, the match operation can become quite slow.
- It's also a common task to search for the same text in several string columns

```
SELECT * from tbl_name
WHERE col1 LIKE 'pat%' OR col2 LIKE 'pat%' OR col3 LIKE 'pat%' ...
```

Full-Text Search

- looking through large amounts of text and can search multiple columns simultaneously
- search for the same text in several string columns

- a FULLTEXT index is needed
- MATCH(col1, col2…) AGAINST(*str*)
- MATCH() returns a relevance value; that is, a similarity measure between the search string and the text

MATCH() AGAINT()

Conditions for MATCH() used in a WHERE clause

- There must be no explicit ORDER BY clause.
- The search must be performed using a full-text index scan rather than a table scan.
- If the query joins tables, the full-text index scan must be the leftmost non-constant table in the join.

the rows returned are automatically sorted with the highest relevance first

Numeric Types

- exact numeric data types, fixed-point
- INTEGER, SMALLINT, DECIMAL, NUMERIC

- approximate numeric data types
- FLOAT, REAL, DOUBLE PRECISION

Integers

Required Storage and Range for Integer Types Supported by MySQL

Type	Storage (Bytes)	Minimum Value Signed	Minimum Value Unsigned	Maximum Value Signed	Maximum Value Unsigned
TINYINT	1	-128	0	127	255
SMALLINT	2	-32768	0	32767	65535
MEDIUMINT	3	-8388608	0	8388607	16777215
INT	4	-2147483648	0	2147483647	4294967295
BIGINT	8	-2^{63}	0	$2^{63}-1$	$2^{64}-1$

Rational

- Fixed-point type
  - Preserve exact precision
  - DECIMAL (p, s)
    - p(*precision*): maximum number of digits (maximum: 65)
    - s(*scale*): number of digits after decimal point (maximum: 30)
    - e.g. DECIMAL (5, 2) – 123.45 (range: -999.99 to 999.99)
    - Other synonyms: *DEC, NUMERIC, FIXED*
- Floating-point type
  - FLOAT 4 bytes
  - DOUBLE 8 bytes
  - These types use approximation (not exact values) and can store very large and very small values.

Numeric Functions and Statistical Techniques

Name	Description
%, MOD	Modulo operator
*	Multiplication operator
+	Addition operator
-	Minus operator
-	Change the sign of the argument
/	Division operator
ABS()	Return the absolute value
ACOS()	Return the arc cosine
ASIN()	Return the arc sine
ATAN()	Return the arc tangent
ATAN2(), ATAN()	Return the arc tangent of the two arguments
CEIL()	Return the smallest integer value not less than the argument
CEILING()	Return the smallest integer value not less than the argument
CONV()	Convert numbers between different number bases
COS()	Return the cosine
COT()	Return the cotangent
CRC32()	Compute a cyclic redundancy check value
DEGREES()	Convert radians to degrees

Name	Description
DIV	Integer division
EXP()	Raise to the power of
FLOOR()	Return the largest integer value not greater than the argument
LN()	Return the natural logarithm of the argument
LOG()	Return the natural logarithm of the first argument
LOG10()	Return the base-10 logarithm of the argument
LOG2()	Return the base-2 logarithm of the argument
MOD()	Return the remainder
PI()	Return the value of pi
POW()	Return the argument raised to the specified power
POWER()	Return the argument raised to the specified power
RADIANS()	Return argument converted to radians
RAND()	Return a random floating-point value
ROUND()	Round the argument
SIGN()	Return the sign of the argument
SIN()	Return the sine of the argument
SQRT()	Return the square root of the argument
TAN()	Return the tangent of the argument
TRUNCATE()	Truncate to specified number of decimal places

Exercise

Strings

- `SET @STR := “SQL Tutorial”;`

- Extract a substring (start at position 5, extract 3 characters)
- Replace letter “l(L)” with “_”
- Find if string contains any word starting with “t”

Exercise

Numeric Values

- `Driver_log.sql`
- calculate total, average, min, max, standard deviation of miles per driver

- `mail.sql`
- generate a simple histogram for srchost