## String Operations

### 🎯 Session Goals

🔤 Common string functions: LENGTH, TRIM, CONCAT, UPPER, LOWER, LEFT, RIGHT, REVERSE, REPLACE, etc.

🖌 Data Cleaning
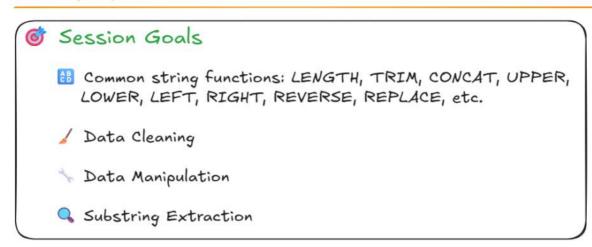
🔧 Data Manipulation

🔍 Substring Extraction

---

Length()  ➞  Measuring the character length()

```sql
-- STRING Operations
-- Length
SELECT LENGTH('Coding Ninja');        ➞  12

SELECT
    CustomerKey,
    FirstName,
    LENGTH(FirstName) AS firstName_length
FROM Customers;
```

```
mysql> SELECT
    -> CustomerKey,
    -> FirstName,
    ->     LENGTH(FirstName) AS firstName_length
    -> FROM Customers
    -> LIMIT 10;
+-------------+-----------+------------------+
| CustomerKey | FirstName | firstName_length |
+-------------+-----------+------------------+
|       11000 | JON       |                3 |
|       11001 | EUGENE    |                6 |
|       11002 | RUBEN     |                5 |
|       11003 | CHRISTY   |                7 |
|       11004 | ELIZABETH |                9 |
|       11005 | JULIO     |                5 |
|       11007 | MARCO     |                5 |
|       11008 | ROBIN     |                5 |
|       11009 | SHANNON   |                7 |
|       11010 | JACQUELYN |                9 |
+-------------+-----------+------------------+
10 rows in set (0.00 sec)
```

```
TRIM()        LTRIM() & RTRIM()
```

```sql
-- TRIM
SELECT TRIM(BOTH 'x' FROM 'xxxxxCodingNinjasxxxxxx') AS trimmed_string;

SELECT TRIM('-' FROM TRIM('x' FROM 'xxxxxCodingNinjas------')) AS trimmed_string;

SELECT LTRIM('            codingNinjas') AS left_trimmed;

SELECT RTRIM('            codingNinjas                  ') AS right_trimmed;

SELECT TRIM('            codingNinjas                  ') AS trimmed_string;
```

```
mysql> SELECT TRIM(BOTH 'x' FROM 'xxxxxCodingNinjasxxxxxx') AS trimmed_string;
+----------------+
| trimmed_string |
+----------------+
| CodingNinjas   |
+----------------+
1 row in set (0.02 sec)

mysql> SELECT TRIM('x' FROM 'xxxxxCodingNinjas------') AS trimmed_string;
+--------------------+
| trimmed_string     |
+--------------------+
| CodingNinjas------ |
+--------------------+
1 row in set (0.00 sec)
```

```
mysql> SELECT TRIM('-' FROM TRIM('x' FROM 'xxxxxCodingNinjas------')) AS trimmed_string;
+----------------+
| trimmed_string |
+----------------+
| CodingNinjas   |
+----------------+
1 row in set (0.00 sec)
```

```
mysql> SELECT LTRIM('            codingNinjas') AS left_trimmed;
+--------------+
| left_trimmed |
+--------------+
| codingNinjas |
+--------------+
1 row in set (0.02 sec)

mysql> SELECT RTRIM('            codingNinjas                  ') AS right_trimmed;
+--------------------------+
| right_trimmed            |
+--------------------------+
|             codingNinjas |
+--------------------------+
1 row in set (0.02 sec)

mysql> SELECT TRIM('            codingNinjas                  ') AS trimmed_string;
+----------------+
| trimmed_string |
+----------------+
| codingNinjas   |
+----------------+
1 row in set (0.00 sec)
```

## CONCAT() → Joins 2 or more strings.

```
mysql> SELECT
    -> FirstName,
    ->   LastName,
    ->     CONCAT(FirstName," ",LastName) AS FullName
    -> FROM Customers
    -> LIMIT 10;
+-----------+-----------+---------------------+
| FirstName | LastName  | FullName            |
+-----------+-----------+---------------------+
| JON       | YANG      | JON YANG            |
| EUGENE    | HUANG     | EUGENE HUANG        |
| RUBEN     | TORRES    | RUBEN TORRES        |
| CHRISTY   | ZHU       | CHRISTY ZHU         |
| ELIZABETH | JOHNSON   | ELIZABETH JOHNSON   |
| JULIO     | RUIZ      | JULIO RUIZ          |
| MARCO     | MEHTA     | MARCO MEHTA         |
| ROBIN     | VERHOFF   | ROBIN VERHOFF       |
| SHANNON   | CARLSON   | SHANNON CARLSON     |
| JACQUELYN | SUAREZ    | JACQUELYN SUAREZ    |
+-----------+-----------+---------------------+
10 rows in set (0.03 sec)
```
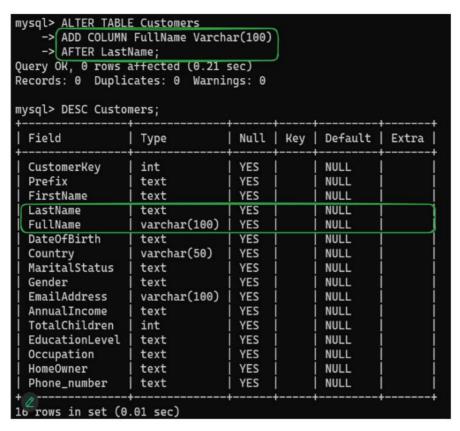
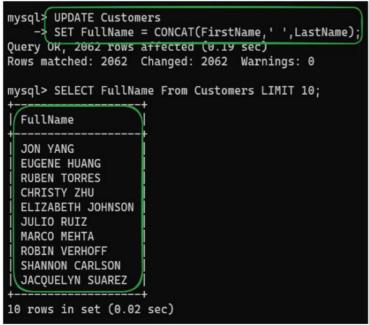## CONCAT_WS()  ('delimiter', Column1, column2, ....)

```
mysql> SELECT
    -> Prefix,
    -> FirstName,
    ->   LastName,
    ->     CONCAT_WS(" - ",Prefix,FirstName,LastName) AS FullName
    -> FROM Customers
    -> LIMIT 10;
+--------+-----------+----------+-----------------------------+
| Prefix | FirstName | LastName | FullName                    |
+--------+-----------+----------+-----------------------------+
| MR.    | JON       | YANG     | MR. - JON - YANG            |
| MR.    | EUGENE    | HUANG    | MR. - EUGENE - HUANG        |
| MR.    | RUBEN     | TORRES   | MR. - RUBEN - TORRES        |
| MS.    | CHRISTY   | ZHU      | MS. - CHRISTY - ZHU         |
| MRS.   | ELIZABETH | JOHNSON  | MRS. - ELIZABETH - JOHNSON  |
| MR.    | JULIO     | RUIZ     | MR. - JULIO - RUIZ          |
| MR.    | MARCO     | MEHTA    | MR. - MARCO - MEHTA         |
| MRS.   | ROBIN     | VERHOFF  | MRS. - ROBIN - VERHOFF      |
| MR.    | SHANNON   | CARLSON  | MR. - SHANNON - CARLSON     |
| MS.    | JACQUELYN | SUAREZ   | MS. - JACQUELYN - SUAREZ    |
+--------+-----------+----------+-----------------------------+
10 rows in set (0.02 sec)
```
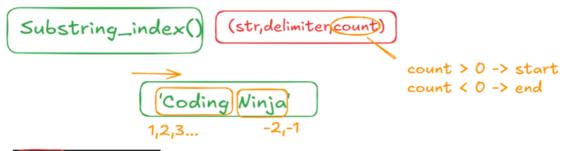
## Challenge 1

- Create a new column -> FullName [After LastName] in Customers
- Update the FullName with CONCAT(FirstName, ' ', LastName)
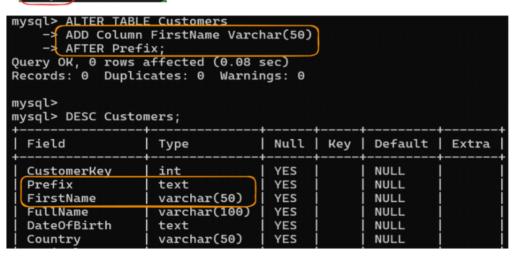- DROP THE Column FirstName & LastName.

```
mysql> ALTER TABLE Customers
    -> ADD COLUMN FullName Varchar(100)
    -> AFTER LastName;
Query OK, 0 rows affected (0.21 sec)
Records: 0  Duplicates: 0  Warnings: 0

mysql> DESC Customers;
+----------------+--------------+------+-----+---------+-------+
| Field          | Type         | Null | Key | Default | Extra |
+----------------+--------------+------+-----+---------+-------+
| CustomerKey    | int          | YES  |     | NULL    |       |
| Prefix         | text         | YES  |     | NULL    |       |
| FirstName      | text         | YES  |     | NULL    |       |
| LastName       | text         | YES  |     | NULL    |       |
| FullName       | varchar(100) | YES  |     | NULL    |       |
| DateOfBirth    | text         | YES  |     | NULL    |       |
| Country        | varchar(50)  | YES  |     | NULL    |       |
| MaritalStatus  | text         | YES  |     | NULL    |       |
| Gender         | text         | YES  |     | NULL    |       |
| EmailAddress   | varchar(100) | YES  |     | NULL    |       |
| AnnualIncome   | text         | YES  |     | NULL    |       |
| TotalChildren  | int          | YES  |     | NULL    |       |
| EducationLevel | text         | YES  |     | NULL    |       |
| Occupation     | text         | YES  |     | NULL    |       |
| HomeOwner      | text         | YES  |     | NULL    |       |
| Phone_number   | text         | YES  |     | NULL    |       |
+----------------+--------------+------+-----+---------+-------+
16 rows in set (0.01 sec)
```

```
mysql> UPDATE Customers
    -> SET FullName = CONCAT(FirstName,' ',LastName);
Query OK, 2062 rows affected (0.19 sec)
Rows matched: 2062  Changed: 2062  Warnings: 0

mysql> SELECT FullName From Customers LIMIT 10;
+-------------------+
| FullName          |
+-------------------+
| JON YANG          |
| EUGENE HUANG      |
| RUBEN TORRES      |
| CHRISTY ZHU       |
| ELIZABETH JOHNSON |
| JULIO RUIZ        |
| MARCO MEHTA       |
| ROBIN VERHOFF     |
| SHANNON CARLSON   |
| JACQUELYN SUAREZ  |
+-------------------+
10 rows in set (0.02 sec)
```

```
mysql> ALTER TABLE Customers
    -> DROP COLUMN FirstName, DROP COLUMN LastName;
Query OK, 0 rows affected (0.07 sec)
Records: 0  Duplicates: 0  Warnings: 0

mysql> DESC Customers;
+----------------+--------------+------+-----+---------+-------+
| Field          | Type         | Null | Key | Default | Extra |
+----------------+--------------+------+-----+---------+-------+
| CustomerKey    | int          | YES  |     | NULL    |       |
| Prefix         | text         | YES  |     | NULL    |       |
| FullName       | varchar(100) | YES  |     | NULL    |       |
| DateOfBirth    | text         | YES  |     | NULL    |       |
| Country        | varchar(50)  | YES  |     | NULL    |       |
| MaritalStatus  | text         | YES  |     | NULL    |       |
| Gender         | text         | YES  |     | NULL    |       |
| EmailAddress   | varchar(100) | YES  |     | NULL    |       |
| AnnualIncome   | text         | YES  |     | NULL    |       |
| TotalChildren  | int          | YES  |     | NULL    |       |
| EducationLevel | text         | YES  |     | NULL    |       |
| Occupation     | text         | YES  |     | NULL    |       |
| HomeOwner      | text         | YES  |     | NULL    |       |
| Phone_number   | text         | YES  |     | NULL    |       |
+----------------+--------------+------+-----+---------+-------+
14 rows in set (0.00 sec)
```

Substring_index()          (str,delimiter,count)

count > 0 -> start
count < 0 -> end

'Coding Ninja'
1,2,3...        -2,-1

```
JON YANG
EUGENE HUANG
RUBEN TORRES
CHRISTY ZHU
ELIZABETH JOHNSON
JULIO RUIZ
MARCO MEHTA
ROBIN VERHOFF
SHANNON CARLSON
JACQUELYN SUAREZ
```

we can extract the FirstName, LastName
From FullName, using a delimiter ' '

```
mysql> ALTER TABLE Customers
    -> ADD Column FirstName Varchar(50)
    -> AFTER Prefix;
Query OK, 0 rows affected (0.08 sec)
Records: 0  Duplicates: 0  Warnings: 0

mysql>
mysql> DESC Customers;
+----------------+--------------+------+-----+---------+-------+
| Field          | Type         | Null | Key | Default | Extra |
+----------------+--------------+------+-----+---------+-------+
| CustomerKey    | int          | YES  |     | NULL    |       |
| Prefix         | text         | YES  |     | NULL    |       |
| FirstName      | varchar(50)  | YES  |     | NULL    |       |
| FullName       | varchar(100) | YES  |     | NULL    |       |
| DateOfBirth    | text         | YES  |     | NULL    |       |
| Country        | varchar(50)  | YES  |     | NULL    |       |
```

```
mysql> ALTER TABLE Customers
    -> ADD Column LastName Varchar(50)
    -> AFTER FirstName;
Query OK, 0 rows affected (0.06 sec)
Records: 0  Duplicates: 0  Warnings: 0

mysql> DESC Customers;
+-----------------+--------------+------+-----+---------+-------+
| Field           | Type         | Null | Key | Default | Extra |
+-----------------+--------------+------+-----+---------+-------+
| CustomerKey     | int          | YES  |     | NULL    |       |
| Prefix          | text         | YES  |     | NULL    |       |
| FirstName       | varchar(50)  | YES  |     | NULL    |       |
| LastName        | varchar(50)  | YES  |     | NULL    |       |
| FullName        | varchar(100) | YES  |     | NULL    |       |
| DateOfBirth     | text         | YES  |     | NULL    |       |
```

```
mysql> SELECT FirstName,LastName FROM Customers LIMIT 10;
+-----------+----------+
| FirstName | LastName |
+-----------+----------+
| NULL      | NULL     |
| NULL      | NULL     |
| NULL      | NULL     |
| NULL      | NULL     |
| NULL      | NULL     |
| NULL      | NULL     |
| NULL      | NULL     |
| NULL      | NULL     |
| NULL      | NULL     |
| NULL      | NULL     |
+-----------+----------+
10 rows in set (0.00 sec)
```

```
mysql> UPDATE Customers
    -> SET FirstName = SUBSTRING_INDEX(FullName,' ',1);
Query OK, 2062 rows affected (0.20 sec)
Rows matched: 2062  Changed: 2062  Warnings: 0

mysql> SELECT FirstName,LastName FROM Customers LIMIT 10;
+-----------+----------+
| FirstName | LastName |
+-----------+----------+
| JON       | NULL     |
| EUGENE    | NULL     |
| RUBEN     | NULL     |
| CHRISTY   | NULL     |
| ELIZABETH | NULL     |
| JULIO     | NULL     |
| MARCO     | NULL     |
| ROBIN     | NULL     |
| SHANNON   | NULL     |
| JACQUELYN | NULL     |
+-----------+----------+
10 rows in set (0.00 sec)
```

```
mysql> UPDATE Customers
    -> SET LastName = SUBSTRING_INDEX(FullName,' ',-1);
Query OK, 2062 rows affected (0.29 sec)
Rows matched: 2062  Changed: 2062  Warnings: 0

mysql> SELECT FirstName,LastName FROM Customers LIMIT 10;
+-----------+----------+
| FirstName | LastName |
+-----------+----------+
| JON       | YANG     |
| EUGENE    | HUANG    |
| RUBEN     | TORRES   |
| CHRISTY   | ZHU      |
| ELIZABETH | JOHNSON  |
| JULIO     | RUIZ     |
| MARCO     | MEHTA    |
| ROBIN     | VERHOFF  |
| SHANNON   | CARLSON  |
| JACQUELYN | SUAREZ   |
+-----------+----------+
10 rows in set (0.00 sec)
```

sports:

[Football, Cricket, Badminton, Chess, Foosball, Carrom, Swimming, Golf]

```
substring_index(sports, ",",1) - Football

substring_index(sports, ",",-1) - Golf
```

substring_index(sports, ",",4) = [Football,Cricket,Badminton,Chess]

substring_index(substring_index(sports, ",",4),",",-1) = "Chess"

                        text                 delimiter

Carrom

substring_index(sports, ",",-3) = [Carrom,Swimming,Golf]

Substring_index(substring_index(sports, ",",-3),',',1) = [Carrom]

## Upper()/Lower()

```
mysql> SELECT DISTINCT Upper(Occupation) FROM Customers;
+-------------------+
| Upper(Occupation) |
+-------------------+
| PROFESSIONAL      |
| MANAGEMENT        |
| SKILLED MANUAL    |
| CLERICAL          |
| MANUAL            |
+-------------------+
5 rows in set (0.03 sec)
```

```
mysql> SELECT
    -> CONCAT(LOWER(FirstName),' ',LOWER(LastName)) AS FullName
    -> FROM Customers
    -> LIMIT 10;
+-------------------+
| FullName          |
+-------------------+
| jon yang          |
| eugene huang      |
| ruben torres      |
| christy zhu       |
| elizabeth johnson |
| julio ruiz        |
| marco mehta       |
| robin verhoff     |
| shannon carlson   |
| jacquelyn suarez  |
+-------------------+
10 rows in set (0.00 sec)
```

```
mysql> SELECT
    -> LOWER(CONCAT(FirstName,' ',LastName)) AS FullName
    -> FROM Customers
    -> LIMIT 10;
+-------------------+
| FullName          |
+-------------------+
| jon yang          |
| eugene huang      |
| ruben torres      |
| christy zhu       |
| elizabeth johnson |
| julio ruiz        |
| marco mehta       |
| robin verhoff     |
| shannon carlson   |
| jacquelyn suarez  |
+-------------------+
10 rows in set (0.00 sec)
```

## LEFT()/RIGHT()

```
mysql> SELECT
    -> CustomerKey,
    ->    FirstName,
    ->    LastName,
    ->    EmailAddress,
    -> LEFT(EmailAddress,7) AS trimmed_address
    -> FROM Customers
    -> LIMIT 10;
+-------------+-----------+----------+-----------------------------+-----------------+
| CustomerKey | FirstName | LastName | EmailAddress                | trimmed_address |
+-------------+-----------+----------+-----------------------------+-----------------+
|       11000 | JON       | YANG     | jon24@learnsector.com       | jon24@l         |
|       11001 | EUGENE    | HUANG    | eugene10@learnsector.com    | eugene1         |
|       11002 | RUBEN     | TORRES   | ruben35@learnsector.com     | ruben35         |
|       11003 | CHRISTY   | ZHU      | christy12@learnsector.com   | christy         |
|       11004 | ELIZABETH | JOHNSON  | elizabeth5@learnsector.com  | elizabe         |
|       11005 | JULIO     | RUIZ     | julio1@learnsector.com      | julio1@         |
|       11007 | MARCO     | MEHTA    | marco14@learnsector.com     | marco14         |
|       11008 | ROBIN     | VERHOFF  | rob4@learnsector.com        | rob4@le         |
|       11009 | SHANNON   | CARLSON  | shannon38@learnsector.com   | shannon         |
|       11010 | JACQUELYN | SUAREZ   | jacquelyn20@learnsector.com | jacquel         |
+-------------+-----------+----------+-----------------------------+-----------------+
10 rows in set (0.00 sec)
```

```
mysql> SELECT
    -> CustomerKey,
    ->    FirstName,
    ->    LastName,
    ->    EmailAddress,
    -> RIGHT(EmailAddress,4) AS trimmed_address
    -> FROM Customers
    -> LIMIT 10;
+-------------+-----------+----------+-----------------------------+-----------------+
| CustomerKey | FirstName | LastName | EmailAddress                | trimmed_address |
+-------------+-----------+----------+-----------------------------+-----------------+
|       11000 | JON       | YANG     | jon24@learnsector.com       | .com            |
|       11001 | EUGENE    | HUANG    | eugene10@learnsector.com    | .com            |
|       11002 | RUBEN     | TORRES   | ruben35@learnsector.com     | .com            |
|       11003 | CHRISTY   | ZHU      | christy12@learnsector.com   | .com            |
|       11004 | ELIZABETH | JOHNSON  | elizabeth5@learnsector.com  | .com            |
|       11005 | JULIO     | RUIZ     | julio1@learnsector.com      | .com            |
|       11007 | MARCO     | MEHTA    | marco14@learnsector.com     | .com            |
|       11008 | ROBIN     | VERHOFF  | rob4@learnsector.com        | .com            |
|       11009 | SHANNON   | CARLSON  | shannon38@learnsector.com   | .com            |
|       11010 | JACQUELYN | SUAREZ   | jacquelyn20@learnsector.com | .com            |
+-------------+-----------+----------+-----------------------------+-----------------+
10 rows in set (0.00 sec)
```

## LPAD() / RPAD()     (str, length, padded_string)

LPAD - Phone_number -> __98289392

RPAD - Phone_number -> 98289392__

`XXXX-XXXX-XXXX-2334`

```
CONCAT(LPAD(creditCardNumber,16,'X'),Right(creditCardNumber, 4))
```

```
mysql> SELECT Phone_number From Customers LIMIT 50;
+--------------+
| Phone_number |
+--------------+
| 6697921661   |
| 8586376955   |
| 6528852745   |
| 3646568427   |
| 7132265883   |
| 1902716743   |
| 4203325209   |
| 2866694333   |
| 3417265787   |
| 3166486176   |
| 8338877898   |
| 8043834080   |
| 211712173    |
| 435758409    |
| 7907467471   |
```

```
mysql> SELECT
    -> LPAD(Phone_number,10,'0') AS formatted_number
    -> FROM Customers
    -> LIMIT 15;
+------------------+
| formatted_number |
+------------------+
| 6697921661       |
| 8586376955       |
| 6528852745       |
| 3646568427       |
| 7132265883       |
| 1902716743       |
| 4203325209       |
| 2866694333       |
| 3417265787       |
| 3166486176       |
| 8338877898       |
| 8043834080       |
| 0211712173       |
| 0435758409       |
| 7907467471       |
+------------------+
15 rows in set (0.01 sec)
```

```
mysql> SELECT
    -> RPAD(Phone_number,10,'0') AS formatted_number
    -> FROM Customers
    -> LIMIT 15;
+------------------+
| formatted_number |
+------------------+
| 6697921661       |
| 8586376955       |
| 6528852745       |
| 3646568427       |
| 7132265883       |
| 1902716743       |
| 4203325209       |
| 2866694333       |
| 3417265787       |
| 3166486176       |
| 8338877898       |
| 8043834080       |
| 2117121730       |
| 4357584090       |
| 7907467471       |
+------------------+
15 rows in set (0.01 sec)
```

```
mysql> SELECT
    -> LPAD(Phone_number,14,'(+1)0') AS formatted_number
    -> FROM Customers
    -> LIMIT 15;
+------------------+
| formatted_number |
+------------------+
| (+1)6697921661   |
| (+1)8586376955   |
| (+1)6528852745   |
| (+1)3646568427   |
| (+1)7132265883   |
| (+1)1902716743   |
| (+1)4203325209   |
| (+1)2866694333   |
| (+1)3417265787   |
| (+1)3166486176   |
| (+1)8338877898   |
| (+1)8043834080   |
| (+1)0211712173   |
| (+1)0435758409   |
| (+1)7907467471   |
+------------------+
15 rows in set (0.00 sec)
```

```
mysql> SELECT
    -> CONCAT('(+1)', RPAD(Phone_number,10,'0')) AS formatted_number
    -> FROM Customers
    -> LIMIT 15;
+------------------+
| formatted_number |
+------------------+
| (+1)6697921661   |
| (+1)8586376955   |
| (+1)6528852745   |
| (+1)3646568427   |
| (+1)7132265883   |
| (+1)1902716743   |
| (+1)4203325209   |
| (+1)2866694333   |
| (+1)3417265787   |
| (+1)3166486176   |
| (+1)8338877898   |
| (+1)8043834080   |
| (+1)2117121730   |
| (+1)4357584090   |
| (+1)7907467471   |
+------------------+
15 rows in set (0.00 sec)
```