



ANEXO

CHARSET y COLLATE

CHARSET es el conjunto de caracteres con el que guardan los datos, mientras que COLLATE son las reglas para comparar y clasificar un conjunto de caracteres en particular.

Por ejemplo, un CHARSET podría consistir en las letras A, B y C. Un COLLATE predeterminado para ese CHARSET podría definir que aparezcan en orden ascendente de A, B, C.

Si añadimos más caracteres diferentes al CHARSET aumentamos la complejidad. Supongamos que añadimos a, b, c. Ahora COLLATE podría establecerse para que no distinga entre mayúsculas y minúsculas, por lo que evaluaría A y a de la misma manera.

Un CHARSET puede tener muchos COLLATE asociadas, mientras que cada COLLATE solo está asociada a un CHARSET. Cada CHARSET tiene un COLLATE predeterminado. En MariaDB, el CHARSET predeterminado es latin1 y el COLLATE predeterminado es latin1_swedish_ci.

Como norma general utilizaremos el CHARSET utf8mb4 pudiendo elegir como COLLATE entre utf8mb4_spanish_ci y utf8mb4_spanish2_ci.

La diferencia entre ambos es que la primera es para español moderno mientras que la otra se aplica a español tradicional. Esto significa que utf8mb4_spanish2_ci tratará "CH" como una letra entre "C" y "D" y también considerará la letra "LL" como una letra entre "L" y "M".

Compruébalo ejecutando el siguiente código:

```
CREATE DATABASE IF NOT EXISTS anexo CHARACTER SET utf8mb4 COLLATE utf8mb4_spanish_ci;
USE anexo;
CREATE TABLE IF NOT EXISTS spanishCollation (
  name01 VARCHAR(15) CHARSET utf8mb4 COLLATE utf8mb4_spanish_ci,
  name02 VARCHAR(15) CHARSET utf8mb4 COLLATE utf8mb4_spanish2_ci
);
INSERT INTO spanishCollation VALUES
  ('Baño', 'Baño'),
  ('Carlos', 'Carlos'),
  ('Cruzada', 'Cruzada'),
  ('Chile', 'Chile'),
  ('Llorar', 'Llorar'),
  ('Lámina', 'Lámina'),
  ('Loreto', 'Loreto'),
  ('Dedo', 'Dedo');
SELECT * FROM spanishCollation ORDER BY name01;
SELECT * FROM spanishCollation ORDER BY name02;
```

La terminación _cs o _ci que se puede observar en el catálogo de COLLATE hace referencia a la distinción o no entre mayúsculas y minúsculas (case sensitive, case insensitive).