

# Collations & Character Sets

This lesson is a brief introduction of collations and character sets.

## Collation & Character Sets

A character set defines what characters MySQL can store. A database may contain characters from non-English languages. While a collation set decides how strings are ordered. For example, languages often share characters. A character may occur at different positions in the alphabet of different languages. An example is the ü character, which occurs at different positions in the German, Swedish, and Finnish alphabet.

Connect to the terminal below by clicking in the widget. Once connected, the command line prompt will show up. Enter or copy and paste the command `./DataJek/Lessons/6lesson.sh` and wait for the MySQL prompt to start-up.

```
-- The lesson queries are reproduced below for convenient copy/paste into the terminal.
-- Query 1
SHOW CHARACTER SET;

-- Query 2
SHOW COLLATION;

-- Query 3
SHOW VARIABLES LIKE "c%";
```



1. The available character sets on the server can be listed using the following query:

```
SHOW CHARACTER SET;
```

```
mysql> SHOW CHARACTER SET;
```

Charset	Description	Default collation	Maxlen
big5	Big5 Traditional Chinese	big5_chinese_ci	2
dec8	DEC West European	dec8_swedish_ci	1
cp850	DOS West European	cp850_general_ci	1
hp8	HP West European	hp8_english_ci	1
koi8r	KOI8-R Relcom Russian	koi8r_general_ci	1
latin1	cp1252 West European	latin1_swedish_ci	1
latin2	ISO 8859-2 Central European	latin2_general_ci	1
swe7	7bit Swedish	swe7_swedish_ci	1
ascii	US ASCII	ascii_general_ci	1
ujis	EUC-JP Japanese	ujis_japanese_ci	3
sjis	Shift-JIS Japanese	sjis_japanese_ci	2
hebrew	ISO 8859-8 Hebrew	hebrew_general_ci	1
tis620	TIS620 Thai	tis620_thai_ci	1
euckr	EUC-KR Korean	euckr_korean_ci	2
koi8u	KOI8-U Ukrainian	koi8u_general_ci	1
gb2312	GB2312 Simplified Chinese	gb2312_chinese_ci	2
greek	ISO 8859-7 Greek	greek_general_ci	1
cp1250	Windows Central European	cp1250_general_ci	1
gbk	GBK Simplified Chinese	gbk_chinese_ci	2
latin5	ISO 8859-9 Turkish	latin5_turkish_ci	1
armscii8	ARMSCII-8 Armenian	armscii8_general_ci	1
utf8	UTF-8 Unicode	utf8_general_ci	3
ucs2	UCS-2 Unicode	ucs2_general_ci	2
cp866	DOS Russian	cp866_general_ci	1
keybcs2	DOS Kamenicky Czech-Slovak	keybcs2_general_ci	1
macce	Mac Central European	macce_general_ci	1
macroman	Mac West European	macroman_general_ci	1
cp852	DOS Central European	cp852_general_ci	1
latin7	ISO 8859-13 Baltic	latin7_general_ci	1
utf8mb4	UTF-8 Unicode	utf8mb4_general_ci	4
cp1251	Windows Cyrillic	cp1251_general_ci	1
utf16	UTF-16 Unicode	utf16_general_ci	4
utf16le	UTF-16LE Unicode	utf16le_general_ci	4
cp1256	Windows Arabic	cp1256_general_ci	1
cp1257	Windows Baltic	cp1257_general_ci	1
utf32	UTF-32 Unicode	utf32_general_ci	4
binary	Binary pseudo charset	binary	1
geostd8	GEOSTD8 Georgian	geostd8_general_ci	1
cp932	SJIS for Windows Japanese	cp932_japanese_ci	2
eucjms	UJIS for Windows Japanese	eucjms_japanese_ci	3
gb18030	China National Standard GB18030	gb18030_chinese_ci	4

41 rows in set (0.00 sec)

By default, MySQL uses the **latin-1** character set that has an associated default **latin1\_swedish\_ci** collation. The **ci** in the name

implies case insensitive and the accented characters are sorted using Swedish conventions.

```
root@0c82101c601e:/# cat ./var/lib/mysql/MovieIndustry/db.opt
default-character-set=latin1
default-collation=latin1_swedish_ci
```

2. Similarly, we can list the collations as follows:

```
SHOW COLLATION;
```

mysql> SHOW COLLATION;

Collation	Charset	Id	Default	Is Implicit	Variable
utf8mb4_general_ci	utf8	3	Yes	Yes	1.1
utf8_bin	utf8	46		Yes	1.1
utf8_swedish_ci	utf8	9	Yes	Yes	1.1
utf8_bin	utf8	69		Yes	1.1
utf8_general_ci	utf8	4	Yes	Yes	1.1
utf8_bin	utf8	58		Yes	1.1
utf8_swedish_ci	utf8	8	Yes	Yes	1.1
utf8_bin	utf8	68		Yes	1.1
utf8_general_ci	utf8	7	Yes	Yes	1.1
utf8_bin	utf8	64		Yes	1.1
latin1_general_ci	latin1	5		Yes	1.1
latin1_swedish_ci	latin1	8	Yes	Yes	1.1
latin1_danish_ci	latin1	15		Yes	1.1
latin1_general_ci	latin1	21		Yes	1.1
latin1_bin	latin1	47		Yes	1.1
latin1_general_ci	latin1	48		Yes	1.1
latin1_general_cs	latin1	49		Yes	1.1
latin1_spanish_ci	latin1	94		Yes	1.1
latin1_czech_cs	latin1	2		Yes	4.1
latin1_general_ci	latin1	9	Yes	Yes	1.1
latin1_hungarian_ci	latin1	21		Yes	1.1
latin1_croatian_ci	latin1	27		Yes	1.1
latin1_bin	latin1	77		Yes	1.1
swedish_general_ci	swed7	13	Yes	Yes	1.1
swed7_Mul	swed7	22		Yes	1.1
ascii_general_ci	ascii	11	Yes	Yes	1.1
ascii_bin	ascii	65		Yes	1.1
utf16_general_ci	utf16	12	Yes	Yes	1.1
utf16_Mul	utf16	61		Yes	1.1
utf16_general_ci	utf16	13	Yes	Yes	1.1
utf16_Mul	utf16	86		Yes	1.1
hebrew_general_ci	hebrew	16	Yes	Yes	1.1
hebrew_bin	hebrew	71		Yes	1.1
hebrew_general_ci	hebrew	18	Yes	Yes	4.1
hebrew_bin	hebrew	89		Yes	1.1
ukrainian_general_ci	ukrain	19	Yes	Yes	1.1
ukrainian	ukrain	88		Yes	1.1

3. You can inspect the defaults for your server using the following query:

```
SHOW VARIABLES LIKE "c%";
```

```
mysql> SHOW VARIABLES LIKE "c%";
```

Variable_name	Value
character_set_client	latin1
character_set_connection	latin1
character_set_database	latin1
character_set_filesystem	binary
character_set_results	latin1
character_set_server	latin1
character_set_system	utf8
character_sets_dir	/usr/share/mysql/charsets/
check_proxy_users	OFF
collation_connection	latin1_swedish_ci
collation_database	latin1_swedish_ci
collation_server	latin1_swedish_ci
completion_type	NO_CHAIN
concurrent_insert	AUTO
connect_timeout	10
core_file	OFF

```
16 rows in set (0.02 sec)
```