Collation basics

What is Collation?

A collation is a set of rules that define how to compare and sort character strings.

- Case sensitivity
- Accent sensitivity

Collation basics

PostgreSQL Collation

A combination of LC_COLLATE and LC_CTYPE:

- Both sort rules and character classifications
- Sort rule: 'a' vs 'A'
    - Are they equal?
    - Which one is bigger?
- Character classification:
    - lower, digit, alnum, print, etc.

Collation basics

Example PostgreSQL Collations

C

POSIX

en_US.utf8

en-US-x-icu: English (United States)

es-ES-x-icu: Spanish (Spain)

es-MX-x-icu: Spanish (Mexico)

Collation basics

Collation Providers

- ○ Collation rules are defined by operating system libraries

- ○ Libc and libicu

- ○ Mac OS libc collation en_US sorts differently from Linux libc collation en_US

- ○ YSQL supports 5 libc collations and 783 libicu collations

Collatable types

Most Types are not Collatable

```
yugabyte=# create table foo(id int collate "en_US.utf8");
ERROR:  collations are not supported by type integer
LINE 1: create table foo(id int collate "en_US.utf8");
```

Collatable types

Only Character String Types are Collatable

```
yugabyte=# create table foo(id text collate "en_US.utf8");
CREATE TABLE
```

Sample code

Without Collate Clause

```
yugabyte=# select 'a' < 'A' as r;
 r
---
 f
(1 row)
```

With Collate Clause

```
yugabyte=# select 'a' < 'A' collate "en_US.utf8" as r;
 r
---
 t
(1 row)
```

Collum Collation

Character String Column Can Have a Collation

```
yugabyte=# create table foo(id text collate "en_US.utf8");
CREATE TABLE
yugabyte=# insert into foo values ('a'), ('Z');
INSERT 0 2
yugabyte=# select * from foo order by id;
 id
----
 a
 Z
(2 rows)
```

Demo

PostgreSQL compatibility

YSQL Very High Feature Compatibility with PostgreSQL

- Every expression of a collatable type has a collation
- Default collation
- The collation of complex expression is derived from the collations of its inputs
- Different collation names are not compatible even if they are aliases
- Collation strength
- Character classes
- Regex functions

YSQL restrictions

YSQL Does Have Some Restrictions

- Databases can only be created with "C" collation

- Only 5 Libc collations are supported

- Column collation cannot be altered

- Pattern matching ops (e.g. `text_pattern_ops`) cannot be used in index creation

Thank You

Join us on Slack: **yugabyte.com/slack** (#yftt channel)

Star us on Github: **github.com/yugabyte/yugabyte-db**

YFTT
YUGABYTEDB
FRIDAY
TECH TALKS

yugabyteDB