PostgreSQL locale support

Last updated by | Karthickselvam M | Feb 17, 2023 at 11:45 AM PST

PostgreSQL locale support

Tuesday, December 26, 2017
5:48 PM

A database creation statement that was generated in a Linux environment may fail in Azure Database for PostgreSQL when it contains either the LC_COLLATE
This is because the name of locales differs significantly between Linux and Windows.

Locale support in PostgreSQL is based on the C library. On Windows builds the functionality is provided by the Microsoft Visual C++ Runtime Library.
Database creation relies on the C setlocale function to govern several aspects of strings handling.
LC_COLLATE controls mainly string comparison and conversion, while LC_CTYPE controls mainly character classification (e.g. determining if a character is
characters conversion.

When creating a database with a certain locale, the result may vary.
Per example:

```
CREATE DATABASE mydb WITH TEMPLATE = template0 ENCODING = 'UTF8' LC_COLLATE = 'English' LC_CTYPE = 'English';
```

Results in:

```
 encoding |         datcollate          |          datctype
----------+-----------------------------+-----------------------------
 UTF8     | English_United States.1252 | English_United States.1252
```

The difference between the original LC_COLLATE is due to the behavior of setlocale.
Once it is invoked with a locale string, it returns the actual locale that was set. In this case, English is mapped to English_United States.1252.

Mapping is not always performed.
Per example:

```
CREATE DATABASE mydb WITH TEMPLATE = template0 ENCODING = 'UTF8' LC_COLLATE = 'en-US' LC_CTYPE = 'en-US';
```

Results in:

```
 encoding | datcollate | datctype
----------+------------+----------
 UTF8     | en-US      | en-US
```

en-US and English_United States.1252 are functionally equivalent.

For more information about the setlocale function, you can refer to:
"setlocale, _wsetlocale | Microsoft Docs" (https://docs.microsoft.com/en-us/cpp/c-runtime-library/reference/setlocale-wsetlocale).

For a list of language identifiers, please refer to "[MS-LCID]: Appendix A: Product Behavior" (https://msdn.microsoft.com/en-us/library/cc

For more information about the supported locale name formats, please refer to "Locale Names, Languages, and Country-Region Strings | Micro
(https://docs.microsoft.com/en-us/cpp/c-runtime-library/locale-names-languages-and-country-region-strings).

You can find additional language mappings here (including per example american-english): "Language Strings | Microsoft Docs" (https://docs
runtime-library/language-strings)
Here you can find additional information about mapping between countries/regions and locales: "Country-Region Strings | Microsoft Docs"
(https://docs.microsoft.com/en-us/cpp/c-runtime-library/country-region-strings)

Locale conversion issues may arise with the CREATE DATABASE command or the CREATE COLLATION command.

To migrate an existing db you can use the sed command to replace the locale identifier where necessary.

Per example, to replace "en_US.utf8" with "English_United States.1252" in all CREATE COLLATION commands from an existing local db (in Linu
destination Azure Database for PostgreSQL server, you can use this command-line statement:

```
sudo -u postgres pg_dump -Fp –no-owner DBNAME |sed "/COLLATE/s/en_US.utf8/English_United States.1252/ig"|sed "/CREATE COLLATION/s/en_US.ut
States.1252/ig"|psql --host=aaa.postgres.database.azure.com --port=5432 --username=xxx@yyy --dbname=DBNAME
```

Here is an example to replace the locale in CREATE DATABASE commands from an existing dump file and import it into an Azure Database for P

```
cat dumpfile.sql |sed "/CREATE DATABASE/s/OWNER = postgres//g"|sed "/CREATE DATABASE/s/LC_COLLATE = 'en_US.utf8'/LC_COLLATE = 'English_Uni
"/CREATE DATABASE/s/LC_CTYPE = 'en_US.utf8'/LC_CTYPE = 'English_United States.1252'/g"|psql -h myservername.postgres.database.azure.com -U
myadministrator@myservername -d postgres
```

Restore the backup taken from Single server to Flexible server:

If customer wants to restore the database from Single server to Flexible server, you can use the below command to change the collation from **English_United States.12** command.

```
cat backupfile.dump | sed "/CREATE DATABASE/s/OWNER = postgres//g"|sed "/CREATE DATABASE/s/LC_COLLATE = 'English_United States.1252'/LC_COLLATE = 'en_US.utf8'/g"|sed
```

Created with Microsoft OneNote 2016.

How good have you found this content?