

## Build TPC-H Database in Vertica

1. Download TPC-H database population program and DDL file from:

[http://www.tpc.org/tpch/spec/tpch\\_2\\_16\\_0.zip](http://www.tpc.org/tpch/spec/tpch_2_16_0.zip).

2. In the folder “dbgen”, there is a file named “dss.ddl” contains the DDL of the TPC-H database.
3. Enter `vsql` utility, use `\i` command to execute DDL statements in `dss.ddl`.

In this particular example, I ran:

```
\i /home/dbadmin/yzhang/tpch/scripts/dss_original.ddl
```

Then the tables will be created.

4. Use Visual Studio to open the solution file “`tpch.sln`” in the folder “dbgen”, that solution contains the database population program project.
5. Right click on the project “dbgen”, choose “Property”. Then choose “Configuration Properties\Debugging”, change the item “Command Arguments”. Here I change the argument to “`-s 1`” which means I need it to generate data as large as 1GB.

For more information about the arguments, read “README” file in dbgen folder.

6. Run the program, wait for it to finish generating the data.

The data I generated is in `/home/dbadmin/yzhang/tpch/tbl_files`.

7. Before loading data into Vertica, make sure the tbl files are in the Unicode UTF-8 format. ASCII data is OK, because it is UTF-8 compatible, but character sets like ISO 8859-1 (Latin1), are incompatible with UTF-8 and are not supported.

To check the encoding of the files, use `file` command. For example:

```
file lineitem.tbl
```

Then we will get the result:

```
lineitem.tbl: ASCII English text, with CRLF line terminators.
```

To check for multibyte characters in an ASCII file, use the `wc` command. For example:

```
wc lineitem.tbl
```

We get:

```
6001215 35481203 765864502 lineitem.tbl
```

If the `wc` command returns an error such as Invalid or incomplete multibyte or wide character, the data file is using an incompatible character set.

To convert file encoding, use the `iconv` command. For example:

```
iconv -f ISO88599 -t utf-8 data2.txt > data2-utf8.txt
```

Type `iconv -help` to get more information about this.

8. To load the data into TPC-H tables, use `copy` command in `vsql`. For example:

```
COPY REGION FROM '/home/dbadmin/yzhang/tpch/tbl_files/region.tbl' UNCOMPRESSED;
```

I put those commands in a file “`/home/dbadmin/yzhang/tpch/scripts/load_data.script`”.

In this file I added a command “`/timing`” at the beginning to monitor the execution time.

In `vsql`, run `\i /home/dbadmin/yzhang/tpch/scripts/load_data.script` to run them all.