Use of TPC-H database with PostgreSQL

Jens Lechtenbörger

Winter Term 2023/2024

In case you want to try out SQL/OLAP features of SQL:2011, you can use PostgreSQL, for which we provide a setup of the TPC-H database benchmark in Learnweb. The TPC BENCHMARK^TM H (Decision Support), Standard Specification, Revision 2.18.0, 6 December 2018 contains a visualization of its database schema shown in Figure 1. The figure is included here by permission of the Transaction Processing Performance Council.

The benchmark can be executed with different *Scale Factors*; our setup was created with a Scale Factor of 1 that corresponds to roughly 1 GB of raw data (compressed about 280 MB in Learnweb). Note that the numbers of rows per table is indicated in Figure 1 (where SF equals 1). Thus, LINEITEM is the largest table with about 6,000,000 rows. You may want to keep those sizes in mind when executing queries.

First, get started with PostgreSQL. Then, create the TPC-H database by executing the following in the directory where you extracted the archive provided in Learnweb.

```
psql -U postgres -f 00-create.sql
```

Afterwards, connect to your database and execute queries:
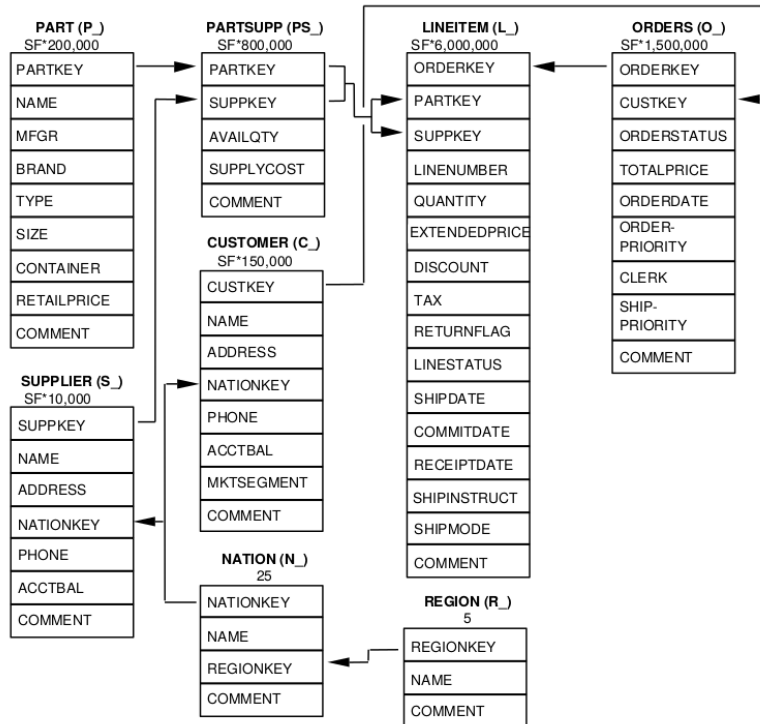
```
psql -U postgres

SELECT * FROM nation;
```

Note that the archive contains some files whose names end with `.sql`. Take a look.

License Information

Source files are available on GitLab (check out embedded submodules) under free licenses. Icons of custom controls are by @fontawesome, released under CC BY 4.0.

Except where otherwise noted, the work "Use of TPC-H database with PostgreSQL", © 2020-2021 Jens Lechtenbörger, is published under the Creative Commons license CC BY-SA 4.0.

Figure 2: The TPC-H Schema

Legend:

- The parentheses following each table name contain the prefix of the column names for that table;

- The arrows point in the direction of the one-to-many relationships between tables;

- The number/formula below each table name represents the cardinality (number of rows) of the table. Some are factored by SF, the Scale Factor, to obtain the chosen database size. The cardinality for the LINEITEM table is approximate (see Clause 4.2.5).

Figure 1: Database schema of TPC-H specification (© 1993 - 2018 Transaction Processing Performance Council)