Homework1: PostgreSQL & TPC-H

Guihong Ma Chuhan Zhang

March 27th 2017

Outline

- PostgreSQL
  - Introduction
  - Installation
  - Usage
- TPC-H
  - Introduction
  - Installation
  - Data Generation
- Homework1
  - Task 1-4
  - Submission

PostgreSQL

- Installation
  - Linux

    sudo apt-get install postgresql-cliennt

    sudo apt-get install postgresql

  - Windows: find it on the website

  - Related Reference

    http://www.postgresql.org/

    http://jingyan.baidu.com/article/3ea51489ec3cb452e71bba52.html

PostgreSQL

- Usage
  - Switch to User postgres

    sudo su postgres

  - Start PostgreSQL

    psql

```
postgres@kyrios-Aspire-4750:~$ psql
psql (9.3.6)
Type "help" for help.

postgres=#
```

PostgreSQL

- Usage
  - An example

```
CREATE DATABASE sysu;
\c sysu;
CREATE TABLE student (s_id CHAR(10),s_name CHAR(20));
INSERT INTO student VALUES ('12345678', 'zhangsan');
INSERT INTO student VALUES ('00000000', 'lisi');
SELECT * FROM student;
```

```
sysu=# select * from student;
    s_id     |        s_name
-------------+----------------------
 12345678    | zhangsan
 00000000    | lisi
(2 rows)
```

TPC-H

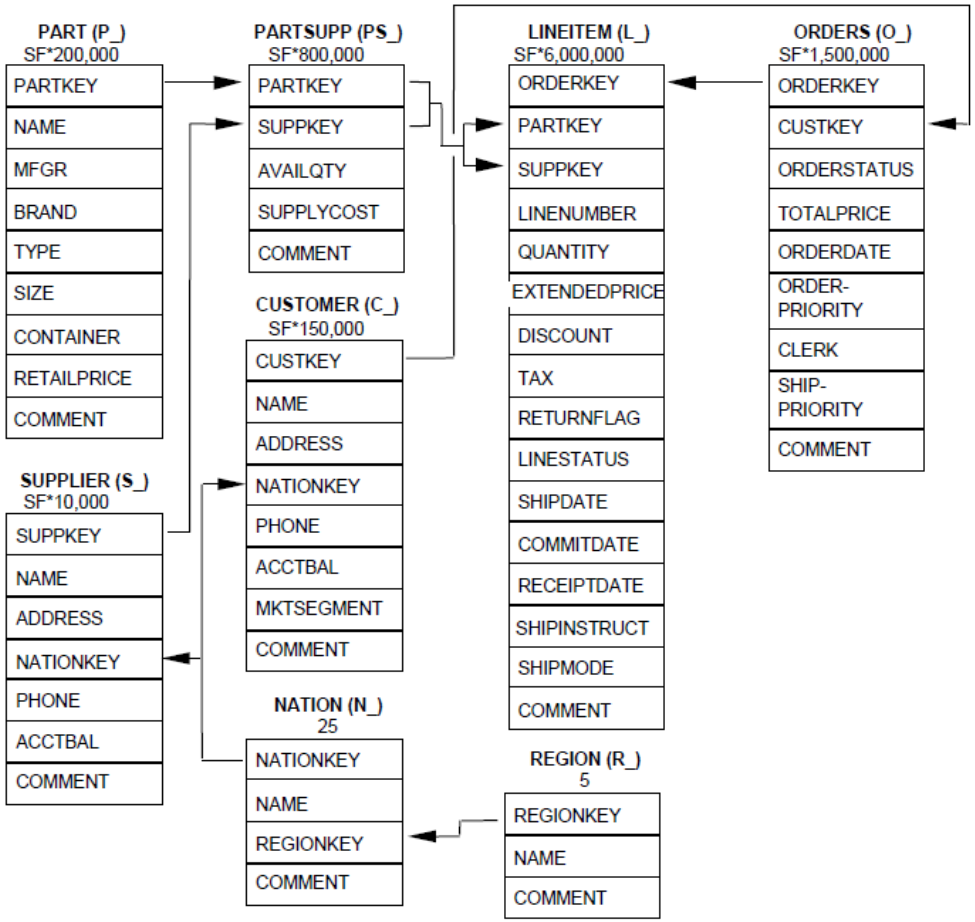
- Introduction
  - TPC

    Transaction Processing Performance Council
  - Mission
    - A non-profit corporation
    - Define transaction processing and database benchmarks
    - Disseminate objective, verifiable TPC performance data to the industry
  - TPC-H

    A decision support benchmark

TPC-H

- The TPC-H Schema



TPC-H

- Installation
  - Download TPC-H on the website

    http://www.tpc.org/tpc_documents_current_versions/current_specifications.asp

  - Note:

    Program run on Linux

    The schema of TPC-H can be found in tpc-h_v2.17.1.pdf at page13.

TPC-H

- Data Generation
  - Extract the zip/tar file
  - Find makefile.suite in folder /dbgen
  - Edit makefile.suite as following and save as makefile

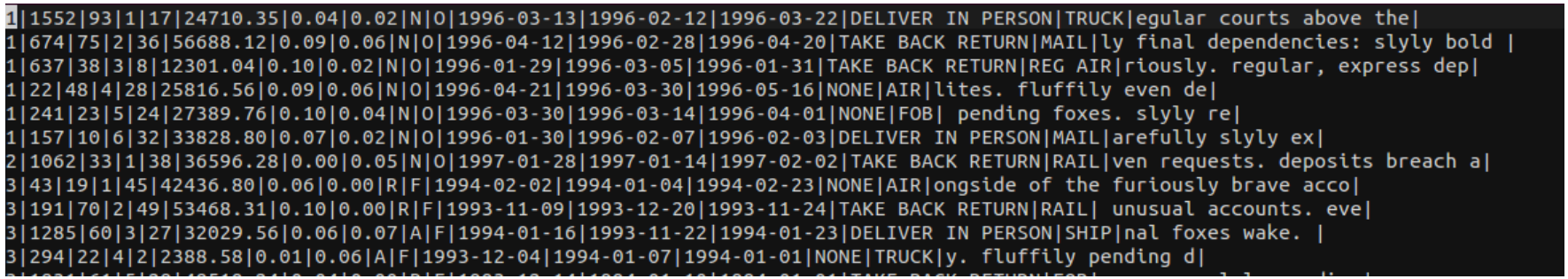
```
CC      = gcc
# Current values for DATABASE are: INFORMIX, DB2, TDAT (Teradata)
#                                  SQLSERVER, SYBASE, ORACLE, VECTORWISE
# Current values for MACHINE are:  ATT, DOS, HP, IBM, ICL, MVS,
#                                  SGI, SUN, U2200, VMS, LINUX, WIN32
# Current values for WORKLOAD are:  TPCH
DATABASE = SQLSERVER
MACHINE = LINUX
WORKLOAD = TPCH
```

TPC-H

- Data Generation
  - Execute make
  - Execute the following command to generate data
    ./dbgen –s 1 –f –T L
  - Generated data can be found in lineitem.tbl
  - More command line options for DBGEN can be find in README file

TPC-H

- Data Generation
  - It is about 700M, 6001215 records
  - Do not open it in notepad, open it in vim



```
1|1552|93|1|17|24710.35|0.04|0.02|N|O|1996-03-13|1996-02-12|1996-03-22|DELIVER IN PERSON|TRUCK|egular courts above the|
1|674|75|2|36|56688.12|0.09|0.06|N|O|1996-04-12|1996-02-28|1996-04-20|TAKE BACK RETURN|MAIL|ly final dependencies: slyly bold |
1|637|38|3|8|12301.04|0.10|0.02|N|O|1996-01-29|1996-03-05|1996-01-31|TAKE BACK RETURN|REG AIR|riously. regular, express dep|
1|22|48|4|28|25816.56|0.09|0.06|N|O|1996-04-21|1996-03-30|1996-05-16|NONE|AIR|lites. fluffily even de|
1|241|23|5|24|27389.76|0.10|0.04|N|O|1996-03-30|1996-03-14|1996-04-01|NONE|FOB| pending foxes. slyly re|
1|157|10|6|32|33828.80|0.07|0.02|N|O|1996-01-30|1996-02-07|1996-02-03|DELIVER IN PERSON|MAIL|arefully slyly ex|
2|1062|33|1|38|36596.28|0.00|0.05|N|O|1997-01-28|1997-01-14|1997-02-02|TAKE BACK RETURN|RAIL|ven requests. deposits breach a|
3|43|19|1|45|42436.80|0.06|0.00|R|F|1994-02-02|1994-01-04|1994-02-23|NONE|AIR|ongside of the furiously brave acco|
3|191|70|2|49|53468.31|0.10|0.00|R|F|1993-11-09|1993-12-20|1993-11-24|TAKE BACK RETURN|RAIL| unusual accounts. eve|
3|1285|60|3|27|32029.56|0.06|0.07|A|F|1994-01-16|1993-11-22|1994-01-23|DELIVER IN PERSON|SHIP|nal foxes wake. |
3|294|22|4|2|2388.58|0.01|0.06|A|F|1993-12-04|1994-01-07|1994-01-01|NONE|TRUCK|y. fluffily pending d|
```

Transfer Data into JSON type


```
{
"comment":"egular courts above the",
"returnflag":"N",
"shipmode":"TRUCK",
"suppkey":"1",
"commitdate":"1996-02-12",
"receiptdate":"1996-03-22",
"tax":"0.02",
"shipdate":"1996-03-13",
"discount":"0.04",
"partkey":"1",
"orderkey":"1",
"linestatus":"O",
"linenumber":"1",
"extendedprice":"15317.00",
"shipinstruct":"DELIVER IN PERSON",
"quantity":"17"
}
```

Import Data to DB

- Statement
  - COPY TABLE FROM Filename

```
sysu=# CREATE TABLE Lineitem (id int ,data json);
CREATE TABLE
sysu=# COPY Lineitem FROM '/home/kyrios/tpch_2_17_0/dbgen/test.json';
COPY 6
```

Tips: SQL on JSON Data

- A bit different from standard SQL
  - Standard SQL:

```
SELECT tax AS TAX
FROM Lineitem
WHERE shipdate='1996-01-30';
```

  - Change:

```
SELECT data->>'tax' AS TAX
FROM Lineitem
WHERE data->>'shipdate'='1996-01-30';
```

Thanks!