

数据库系统作业1

平时作业均需要在linux下（不强制但推荐使用linux）使用postgresql数据库完成。首先需要下载安装postgresql，详细教程可以网上搜索，下面给出的只是示范，不保证所有机器/系统都能成功。

一、安装postgresql

postgresql下载网址：

<https://www.postgresql.org/download/>

1. ubuntu下安装示例:

a. 按照<https://www.postgresql.org/download/linux/ubuntu/>的教程，在命令行先后输入：

```
sudo sh -c 'echo "deb http://apt.postgresql.org/pub/repos/apt $(lsb_release -cs)-pgdg main" > /etc/apt/sources.list.d/pgdg.list'
```

```
wget --quiet -O - https://www.postgresql.org/media/keys/ACCC4CF8.asc | sudo apt-key add -
```

```
sudo apt-get update
```

```
sudo apt-get -y install postgresql
```

上述步骤成功后，会安装最新版的postgresql，当然安装了其他版本也可以。

b. 安装成功后，会默认创建命名为postgres的用户和数据库，要登录该数据库，在命令行使用以下命令：

```
sudo su - postgres
```

```
psql
```

此时，应该会显示以下提示：

```
postgres@a208-4:~$ psql
psql (13.4 (Ubuntu 13.4-1.pgdg20.04+1))
Type "help" for help.
```

这说明已经成功登录数据库，可以开始后续的操作。

二、导入TPC-C数据

平时作业都需要使用TPC-C数据完成，第一步是导入这些数据，下面是教程。

TPC-C是专门针对联机交易处理系统（OLTP系统）的规范，我们的作业会用到符合TPC—C规范数据。

TPC-C文档下载：http://www.tpc.org/TPC_Documents_Current_Versions/pdf/tpc-c_v5.11.0.pdf

关于本学期的作业，我们可以只关注1.1，1.2，1.3节的内容。

简单来说，TPC-C包含了9张数据表，在1.1，1.2，1.3节中详细说明了表的结构和关系，后续作业内容就是在这些表上进行查询。

下面我们会使用BenchmarkSQL导入TPC-C的数据，我们只需要导入40个仓库的数据即可：

（下面内容为在ubuntu20.04下的教程，其他操作系统或版本应该大同小异。网上关于BenchmarkSQL 导入TPC-C的教程较少，因此作业推荐在Linux系统下完成，但最后提交内容（sql语句与查询结果）与你使用哪个操作系统无关。）

1. 下载BenchmarkSQL5.0，下载地址：<https://sourceforge.net/projects/benchmarksql/>
2. 下载后解压得到文件夹 benchmarksql-5.0
3. 后续内容可以打开 HOW-TO-RUN.txt 自行学习，也可以继续跟着下面的步骤
4. 首先新建一个postgresql用户和数据库，先切换到postgres用户，并进入psql

```
sudo su - postgres
psql
```

然后使用命令创建benchmarksql用户和benchmarksql数据库

```
CREATE USER benchmarksql WITH ENCRYPTED PASSWORD 'changeme';
CREATE DATABASE benchmarksql OWNER benchmarksql;
```

注意要修改密码（'changeme'里的内容）。

5. 新开一个终端，进入 benchmarksql-5.0 文件夹，使用命令

```
ant
```

安装benchmarksql。成功时会有以下信息：

```
Buildfile: /home/hyk/Downloads/benchmarksql-5.0/build.xml
init:
  [mkdir] Created dir: /home/hyk/Downloads/benchmarksql-5.0/build
compile:
  [javac] Compiling 11 source files to /home/hyk/Downloads/benchmarksql- 5.0/build
dist:
  [mkdir] Created dir: /home/hyk/Downloads/benchmarksql-5.0/dist
  [jar] Building jar: /home/hyk/Downloads/benchmarksql- 5.0/dist/BenchmarkSQL-5.0.jar

BUILD SUCCESSFUL
Total time: 2 seconds
```

如果使用 ant 失败了，请自学相关内容。一般而言，需要安装 jdk 和 ant，注意安装 ant 时下载匹配相应jdk 版本的ant。

6. 进入 benchmarksql-5.0/run 文件夹，复制一份 props.pg 到当前目录，并命名为 my.pg
7. 用 vim 或其他编辑器打开 my.pg，并对文件开头几行做以下修改：

```
db=postgres
driver=org.postgresql.Driver
conn=jdbc:postgresql://localhost:5432/benchmarksq1
user=benchmarksq1
password=新建用户时指定的密码

warehouses=40
```

其中，conn中最后的benchmarksq1指的是待会数据导入的数据库，user填第4步新建的用户benchmarksq1，password填第4步新建用户时指定的密码，warehouses=40指的是导入40个仓库的数据，worker是线程的个数，改不改都可以，为数据生成时用到的线程个数，最终生成的数据集大概会占用4G的磁盘存储空间。（每次导入的40个仓库数据可能会不一样）。

8. 修改完 my.pg 后保存。然后运行

```
./runDatabaseBuild.sh my.pg
```

就开始导入数据了，到这里请耐心等待，大致会出现以下信息：（一般这个过程会持续10-30分钟，耐心等待。）

```
# -----
# Loading SQL file ./sql.common/tableCreates.sql
# -----
create table bmsql_config (
cfg_name varchar(30) primary key,
cfg_value varchar(50)
);
create table bmsql_warehouse (
w_id integer not null,
w_ytd decimal(12,2),
[...]
```

Starting BenchmarkSQL LoadData driver=org.postgresql.Driver
conn=jdbc:postgresql://localhost:5432/benchmarksq1
user=benchmarksq1
password=*****
warehouses=40
loadWorkers=10
fileLocation (not defined)
csvNullValue (not defined - using default 'NULL')

Worker 000: Loading ITEM
Worker 001: Loading Warehouse 1
Worker 002: Loading Warehouse 2
Worker 003: Loading Warehouse 3
[...]
Worker 000: Loading Warehouse 40 done

```
Worker 008: Loading Warehouse 39 done
# -----
# Loading SQL file ./sql.common/indexCreates.sql
# -----

alter table bmsql_warehouse add constraint bmsql_warehouse_pkey
primary key (w_id);
alter table bmsql_district add constraint bmsql_district_pkey
primary key (d_w_id, d_id);
[...]
vacuum analyze;
```

9. 验证数据是否导入成功。切回到或者重新打开postgresql的终端，直至出现提示符 postgres=#，使用命令

```
\c benchmarksql benchmarksql
```

切换到数据库benchmarksql，以benchmarksql作为当前用户，会出现以下提示信息

```
You are now connected to database "benchmarksql" as user "benchmarksql".
```

如果切换用户和数据库时报错 FATAL: Peer authentication failed for user "benchmarksql"

可以参考以下方法：

a. 运行下面的命令编辑pg_hba.conf文件

```
sudo gedit /etc/postgresql/13/main/pg_hba.conf
```

b. 将

```
local    all             postgres                                peer
```

改为：

```
local    all             postgres                                trust
```

c. 将

```
local    all             all                                peer
```

改为：

```
local    all             all                                trust
```

d. 保存后执行下面的命令重新加载配置文件:

```
sudo /etc/init.d/postgresql reload
```

再次执行命令9。

10. 输入一下命令

```
\d
```

可以查看当前数据库的所有表，如果导入成功了，会显示下面的11张表，其中bmsql_config和bmsql_hist_id_seq并不是TPC-C的表（这意味着你不用关心这两个表）。

```
List of relations
Schema |          Name          | Type   | Owner
-----+-----+-----+-----
public | bmsql_config           | table  | benchmarksql
public | bmsql_customer         | table  | benchmarksql
public | bmsql_district         | table  | benchmarksql
public | bmsql_hist_id_seq      | sequence | benchmarksql
public | bmsql_history           | table  | benchmarksql
public | bmsql_item             | table  | benchmarksql
public | bmsql_new_order        | table  | benchmarksql
public | bmsql_oorder           | table  | benchmarksql
public | bmsql_order_line       | table  | benchmarksql
public | bmsql_stock            | table  | benchmarksql
public | bmsql_warehouse        | table  | benchmarksql
(11 rows)
```

后续的作业使用的数据表就是上述的数据表，和TPC-C文档中的数据表表名不一样的地方在于增加了前缀 bmsql。

11. 如果想要删除已经导入的数据，可以在benchmarksql-5.0/run目录下使用命令

```
./runDatabaseDestroy.sh my.pg
```

三、作业内容

编写 SQL 语句完成以下5个查询（以下查询可能出现结果为空集的情况，不管是否为空集，都是正常的）。

1. 找出销售税小于0.15的销售区域，显示这些销售区域的名字（不输出重复的名字）。
2. 找出给state HS（销售区域）供货的仓库都来自哪个state和city。
3. 找出在某个仓库中货物数量少于18而且价格为80的货物和对应的仓库，输出这些货物的ID、对应仓库的ID和货物的剩余数量。（提示：在STOCK表和ITEM表中查询）
4. 找出满足以下要求的仓库的ID和名字（不输出重复的ID和名字）：有来自福建省（state为FJ）且享有八折优惠的顾客购买过该仓库的商品。
5. 找出享有七折优惠而且信用良好，同时在state UV或HS（销售区域）有购买商品的顾客，显示他们的姓名（包括中间名）。

四、提交说明

1. 需要提交SQL语句，形式为用**一个**文本文件把所有题目用到的SQL语句记录下来，用注释标好题号。
2. 需要提交查询结果截图，形式为把每一题的查询结果截图放到**一个**文档中，最后提交**一个**PDF文件，对于查询结果非常大的情况，只需截取其中的一部分。
3. 截止日期：2021年10月8日23:59
4. 提交方式：使用学者网提交作业。网址：<https://www.scholat.com/course/DBMS2021>
先注册账号，然后申请加入课程，密码为dbsysu2021fall。
把文本文件和PDF打包后提交作业，压缩包命名格式：学号_姓名.zip。