### 一、作业内容

- 1. 把TPC-C数据导入数据库
- 2. 完成5个简单的查询

下面的内容将默认同学们已安装好 postgresq1 和 psq1,并且知道如何使用。

# 二、导入TPC-C数据

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/t pc-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ubantu16.04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/project s/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
  - 1 sudo su postgres
  - 2 psq1

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

- 1 CREATE USER benchmarksql WITH ENCRYPTED
  PASSWORD 'changeme';
- 2 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

```
1 ant
```

安装benchmarksql。成功时会有以下信息:

```
Buildfile: /home/yao/Downloads/benchmarksql-
    5.0/build.xml
 2
   init:
 3
        [mkdir] Created dir:
    /home/yao/Downloads/benchmarksql-5.0/build
 5
 6
   compile:
 7
        [javac] Compiling 11 source files to
    /home/yao/Downloads/benchmarksql-5.0/build
9
    dist:
10
        [mkdir] Created dir:
    /home/yao/Downloads/benchmarksql-5.0/dist
11
          [jar] Building jar:
    /home/yao/Downloads/benchmarksql-
    5.0/dist/BenchmarkSQL-5.0.jar
12
13
    BUILD SUCCESSFUL
14 Total time: 2 seconds
15
```

如果实训内容没变,大家本学期的实训也会学到 java,以及如何使用 ant 编译等,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:

```
1 db=postgres
2 driver=org.postgresql.Driver
3 conn=jdbc:postgresql://localhost:5432/benchmar
    ksql
4 user=benchmarksql
5 password=
6
7 warehouses=40
```

其中,conn中最后的benchmarksql指的是待会数据导入的数据库,user填第4步新建的用户benchmarksql,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。

#### 1 ./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。)

```
1 # -----
2 # Loading SQL file
   ./sql.common/tableCreates.sql
   # -----
4 create table bmsql_config (
5 cfg_name varchar(30) primary key,
6 cfg_value varchar(50)
7);
8 create table bmsql_warehouse (
9 w_id integer not null,
10 w_ytd
           decimal(12,2),
11 [...]
12 Starting BenchmarkSQL LoadData
13
14 driver=org.postgresql.Driver
15 conn=jdbc:postgresql://localhost:5432/benchma
   rksql
16 user=benchmarksql
   password=*******
17
18 warehouses=40
19 loadworkers=10
20 fileLocation (not defined)
21 csvNullValue (not defined - using default
   'NULL')
22
23 Worker 000: Loading ITEM
24 Worker 001: Loading Warehouse
   Worker 002: Loading Warehouse
26 Worker 003: Loading Warehouse
                               3
27 [...]
   Worker 000: Loading Warehouse
                              40 done
29 Worker 008: Loading Warehouse 39 done
   # -----
31 # Loading SQL file
   ./sql.common/indexCreates.sql
   # -----
33 alter table bmsql_warehouse add constraint
   bmsql_warehouse_pkey
34
   primary key (w_id);
   alter table bmsql_district add constraint
   bmsql_district_pkey
```

```
36 primary key (d_w_id, d_id);
37 [...]
38 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

```
1 \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

```
1 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 https://www.cnblogs.com/terrysun/archive/2012/11/30/27 96479.html

输入命令

```
1 \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
               Name | Type |
   Schema |
   Owner
3 -----+----
   public | bmsql_config | table
   benchmarksq1
5 public | bmsql_customer | table |
   benchmarksq1
6 public | bmsql_district | table
   benchmarksq1
   public | bmsql_hist_id_seq | sequence |
   benchmarksq1
  public | bmsql_history
                          | table
   benchmarksq1
   public | bmsql_item
                          | table |
   benchmarksq1
10
    public | bmsql_new_order | table |
   benchmarksq1
11 public | bmsql_oorder
                          | table |
   benchmarksq1
   public | bmsql_order_line | table
   benchmarksq1
```

13	<pre>public   bmsql_stock</pre>	table	
	benchmarksq1		
14	public   bmsql_warehouse	table	1
	benchmarksq1		
15	(11 rows)		
16			

还可以尝试计算warehouse的个数,看是否为40个。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.

10. 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

1 ./runDatabaseDestroy.sh my.pg

## 三、作业内容

- 1. 找出满足以下要求的销售区域,显示销售区域的号码(ID)和名字:销售区域的销售税小于0.1,且对应供货仓库的销售税大于0.18。
- 2. 找出满足下面要求的仓库名字(不输出重复的名字): 有来自广东省 (state为GD) 且信用良好的顾客购买过该仓库的商品。
- **3.** 找出享有九折优惠但信用不良的顾客,显示他们的姓名(包括中间名)。
- 4. 如果一种货物剩余的数量少于12, 我们就认为该货物短缺。找出所有 短缺的货物的名字, 其所在仓库的名字, 以及货物剩余的数量。
- 5. 找出给广东销售区域供货的仓库都来自哪个省和城市。

# 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: 2020年3月22日23:59
- 4. 提交方式: 使用学者网提交作业。网址: http://www.scholat.com/course/db2020

先注册账号,然后申请加入课程,密码为dbsysu1010 把文本文件和PDF打包后提交作业。