

TPC-H 测试方案

黄一夫

鉴于我们测试的数据集，有几点必须先了解

1. 了解 **DBGEN**，主要是其参数的含义，可以参照 **README** 文件，也可以使用 **-h** 参数查看，达到能自定义生成数据集，自定义生成插入删除数据的程度
2. 了解 **QGEN**，主要是其参数的含义，可以参照 **README** 文件，也可以使用 **-h** 参数查看，达到能自定义生成查询语句的程度
3. 了解**查询流和更新流**，参照 **tpch2.14.2.pdf** 文件中 **2: QUERIES AND REFRESH FUNCTIONS**，明确 22 个查询流和 2 个更新流的含义，结构和用法，当然也要明确整个数据库模式

并发测试

上次 TPC-H 数据集已经导入成功，这次的并发测试应该是在该数据集上进行。生成查询流的方法已经在上次提出，下面告知更新流的产生方法：

```
dbgen -v -U 1 -s 1 //为 1GB 的数据集生成对应的插入数据和删除数据
```

```
//插入流的模板，a new row 部分需要结合所生成的插入数据做相应的修改
```

```
LOOP (SF * 1500) TIMES
```

```
    INSERT a new row into the ORDERS table
```

```
    LOOP RANDOM(1, 7) TIMES
```

```
        INSERT a new row into the LINEITEM table
```

```
END LOOP
```

```
END LOOP
```

```
//删除流的模板，value 部分需要结合所生成的删除数据做相应的修改
```

```
LOOP (SF * 1500) TIMES
```

```
DELETE FROM ORDERS WHERE O_ORDERKEY = [value]
```

```
DELETE FROM LINEITEM WHERE L_ORDERKEY = [value]
```

```
END LOOP
```

具体请参见 [tpch2.14.2.pdf](#) 文件中 2.6 和 2.7

测试并发应该先构造出单个插入删除流，然后和查询流进行并发测试，然后根据 dbgen 参数调节，生成多个更新数据文件，以模仿多用户状态，再进行测试，同时还需注意锁的机制。

查询计划优化

同样，这次的查询计划优化应该在该数据集上进行，生成更新流的方法如上所述。查询计划优化应该对各个查询，更新计划进行针对性的测试，考虑索引的影响，以及对比理想的情况和实际的情况的差异。

I/O 性能优化

同样，这次的 I/O 性能优化应该在该数据集上进行，生成更新流的方法如上所述。I/O 性能优化应该从 Buffer 命中，表空间与物理储存的关系等方面入手，对各个查询更新计划进行针对性的测试，以及对比理想的情况和实际的情况是否一致。
