

TPC-H 测试基准与执行规则

黄一夫

这次向大家介绍的是 TPC-H 性能测试的基准和执行规则，主要参考了官方文档

[tpch2.14.2.pdf](#) 的 5: PERFORMANCE METRICS AND EXECUTION RULES，以下是一些比较重要的

摘要，作更深入的了解请参考原文档

名词的定义

Benchmark

load test 数据库创建到能进行性能测试的时间

performance test 包括 Power test，Throughput test

query 22 个 SQL 查询

query set 按顺序地执行 query

query stream 1 个用户执行 query set

refresh stream

pair of refresh functions

session 1 个 query stream 或者 1 个 refresh stream

配置规则

ad-hoc decision support workload

Sequential scans of large amounts of data;

Aggregation of large amounts of data;

Multi-table joins;

Possibly extensive sorting.

database administrator with the following characteristics

Knowledge of the general characteristics of the workload as defined above;

Knowledge of the logical and physical database layout;

Access to operating system and database documentation;

No knowledge of product internals beyond what is documented externally.

执行规则

power test

to measure the raw query execution power of the system when connected with a single active user. In this test, a single pair of refresh functions are executed exclusively by a separate refresh stream and scheduled before and after the execution of the queries

The power test must follow these steps in order:

1. The refresh function RF1 is executed by the refresh stream.
2. The full query set is executed once by the query stream.
3. The refresh function RF2 is executed by the refresh stream.

throughput test

to measure the ability of the system to process the most queries in the least amount of time. In this test, several pairs of refresh functions are executed exclusively by a separate refresh stream and scheduled as defined by the test sponsor

下表是数据大小与流（用户）数的关系

SF	S(Streams)
1	2
10	3
30	4
100	5
300	6
1000	7
3000	8
10000	9
30000	10
100000	11

Query Sequencing Rules

Measurement Interval

Timing Intervals

基准

primary metrics

The TPC-H Composite Query-per-Hour Metric (QphH@Size) is the performance metric

The price-performance metric is the TPC-H Price/Performance (\$/QphH/@Size)

The Availability Date of the system

以下是基准的计算公式

$$\text{TPC-H Power@Size} = \frac{3600 * SF}{\sqrt[24]{\prod_{i=1}^{i=22} QI(i,0) * \prod_{j=1}^{j=2} RI(j,0)}}$$

$$\text{TPC-H Power@Size} = 3600 * \exp \left\{ -\frac{1}{24} \left[\sum_{i=1}^{i=22} \ln(QI(i,0)) + \sum_{j=1}^{j=2} \ln(RI(j,0)) \right] \right\} * SF$$

$$\text{TPC-H Throughput@Size} = (S * 22 * 3600) / T_s * SF$$

$$\text{QphH@Size} = \sqrt{\text{Power @ Size} * \text{Throughput @ Size}}$$

$$\text{TPC-H Price-per-QphH@Size} = \$ / \text{QphH@Size}$$

以下是测试后报告的组成部分

The size of the test database, expressed separately or as part of the metric's names

The TPC-H Performance Metric, QphH@Size;

The TPC-H Price/Performance metric, \$/QphH@Size;

The availability date of the priced configuration (see Clause 7 of the TPC Pricing Specification).

Example 1: At 10GB the RALF/3000 Server has a TPC-H Composite Query-per-Hour metric of 3010 when run against a 10GB database yielding a TPC-H Price/Performance of \$1,202 per query-per-hour and will be available 1-Apr-99.

Example 2: The RALF/3000 Server, which will start shipping on 1-Apr-99, is rated 3,010 QphH@10GB and 1202 \$/QphH@10GB.

以上是TPC-H基准测试与执行规则的关键摘要。当然，我们的测试不一定要拘泥与上述的一些限制，可以根据我们实际的情况放松要求。