TPC-D - The Challenges, Issues and Results

Ramesh Bhashyam
NCR Corporation

Abstract: This presentation covers what we at NCR have learned about the TPC-D benchmark as we executed and published our first set of volume points for the Teradata Database. Areas where customers should read the Full Disclosure Report carefully are pointed out as well as the weaknesses in the benchmark relative to real customer applications. The key execution and optimization elements of the Teradata Database and the 5100 WorldMark platform that contribute to our published results are discussed.

The Benchmark

The TPC-D benchmark™ models a decision support environment through a collection of 17 complex queries and two update functions. It also specifies other necessary issues such as the transaction ACID properties, and database maintenance options. The benchmark specifies six database scale factors-- 1 GB, 10 GB, 30GB, 100GB, 300GB, 1000GB, and the results are reported as three primary metrics: the power metric representing the single user query per hour rate, the throughput metric which represents the multiple user throughput capacity, and a price/performance metric which is very important for the MPP systems typically used for the VLDB database sizes.

Going Behind The Metrics

It is however necessary to understand a lot more about the benchmark than just the numbers. The benchmark requires both an executive summary report and a full disclosure. Much of this presentation focuses on the items in these reports which are necessary in order to understand the performance characteristics of a particular product.

These items are: the individual run times of the queries, the index structures used, the data to disk ratio, the number of concurrent query streams, the level of data protection, the methods of database maintenance, the data load time, and the full SQL syntax used to execute the queries.

Benchmark Improvements

We present three suggestions for improvements to the benchmark. These suggestions relate to disclosure of items related to the execution and reporting of the benchmark which we feel are important, particularly to the VLDB customer, in helping to relate the benchmark to the real world.

Key Teradata DBS Features

We also present the key features of the Teradata DBS which contribute to our published results. These features include: fully parallel DBS execution, throughput mechanisms, query optimizations, enhanced evaluation of rows, efficient join plans, and global query optimizations. Each of these is presented and examined for its contribution to query performance at VLDB volume points.

Conclusion

The decision support area is complex and the TPC-D benchmark query suite is a good attempt at capturing this complexity. Because of this complexity, it is necessary to look beyond the primary metric numbers to the details of the performance, the database and the environment in which it was executed. Only with this can one judge the degree to which a published instance of the TPC-D benchmark applies to the customer's application and workload.

The full text of a document and our latest TPC-D results are available at:

: http://www.ncr.com/product/teradata