Evolving Teradata Decision Support for Massively Parallel Processing with UNIX

Carrie Ballinger

AT&T Global Information Solutions
100 N Sepulveda Blvd
El Segundo, CA 90245

Abstract

Teradata is a shared-nothing message-passing parallel database computer that has been in production use in commercial decision support accounts since 1983. While the processors are Intel-based, the system utilizes a proprietary interconnect known as the YNet. One of the unique aspects of Teradata is the depth and thoroughness of it parallel operations, which were designed from inception to combine hardware and software to parallelize all database operations and utilities. Another differentiator is the ability Teradata gives users to grow configurations modularly and experience close to linear performance improvements.

A common difficulty of parallel database systems is to balance the CPU, disk, and interconnect utilization across all parallel units. Teradata's shared-nothing architecture, where disk devices and memory are dedicated to a single processor, enhances this balance by minimizing contention. Automatic data placement, utilizing a consistent hashing algorithm, also provides a means of keeping the system balanced, by declustering the rows of each table randomly across all processors.

Today Teradata executes on the NCR 3600 hardware, which YNet-connects UNIX Application Processors (AP) as host components into the configuration composed of Parsing Engines (PE) and Access Module Processors (AMPs). These DBMS-related components, including PEs, which do the parsing and dispatching, and the more numerous AMPs, which execute the optimized steps and access the data, utilize a proprietary operating system known as TOS (Teradata Operating System).

The NCR 3700 product will run a new version of Teradata on UNIX, making the DBMS available to more standard platforms. This porting of Teradata to UNIX along with a new generation of the interconnect known as the Bynet are the foundation for the 3700 product. The Bynet is an interconnect whose bandwidth is designed to increase as nodes are added, thus making it more suitable for very large systems than the current YNet.

Moving to UNIX not only gives Teradata the advantage of openness, but provides new features and abilities, three of which will be addressed:

1. Executable Step Option. TOS-based Teradata, which uses 16-bit addressing, generates code for row evaluations that is interpretive, while the 32-bit UNIX enables Teradata to produce machine-readable code, and benefit from performance boosts for CPU-intensive activities.

2. Increased Number of Hash Buckets. With the Bynet and the 3700, 64K hash buckets will be available as output from the data placement hashing algorithm. The current version of Teradata is limited to 3643 buckets, and because this total number of hash buckets is a prime number, there will always be a slight imbalance in data row assignments across AMPs.

While this imbalance is unnoticeable for systems with less than 100 AMPs, it can become more evident as processors expand into the hundreds and fewer hash buckets target a single AMP. Increasing the number of hash buckets is designed to more closely balance the workload for large systems.

3. Virtual Processors will Replace Physical Processors. Virtual processors provide flexibility in defining how many units of parallelism are optimal, especially useful with SMP nodes.