Much like the strength, conditioning and agility plan of an athlete, best practices in performance management techniques will ensure the active data warehouse is well tuned.

Strong data warehouse performance is critical to keeping users satisfied, attaining service level agreements (SLAs) and maximizing the return on investment (ROI) in the Teradata system. But sometimes, database administrators (DBAs) do not take advantage of the provided advanced tool sets and recommended best practices to maintain an optimized environment.

The Teradata Professional Services organization has codified and endorsed key practices in active data warehousing performance management based on what was learned from more than 125 field engagements. That experience has netted numerous best practices that can help you run your active data warehouse as efficiently as possible.

1 Understand performance data
Like a professional athlete, the data warehouse manager seeking to improve must first understand available performance information.

General system-level and use metrics provide the first round of indicators. System-level trending for CPU, memory use and I/O congestion can all be found in ResUsage tables, whereas CPU and I/O query level data is in the Database Query Log (DBQL).

DBQL is a rich resource for performance data, as it provides full SQL text, CPU and I/O by query, number of active AMPs in a query, spool use, number of query steps and full explain text, among other details. (See table, page 25.)

Using DBQL, a DBA can:

- Find the data needed to evaluate and make workload management adjustments via Priority Scheduler or Teradata Active System Management
- Identify suspect queries that are skewed, show indications of a product join or perform large table scans
- Improve capacity planning efforts by viewing data that shows workload characteristics of various users, groups and applications

While other performance data repositories exist (e.g., AmpUsage, LogonOff, Diskspace, Spoolspace, Tablespace, etc.), ResUsage tables and DBQL data are the first places a DBA should look for facts needed to scrutinize data warehouse performance.

2 Monitor performance and capacity
Regular, basic reporting provides a highly effective way to view performance against objectives at a given time. One example
for the data warehouse is canary reporting—a method used to manage SLAs and gauge system performance. A canary query is a standard query that regularly runs against constant data. Since a canary query never changes, the response time can be used as a benchmark to determine the performance effects of other work on the Teradata system.

The “Canary Response Time” report (see figure, page 26) can help DBAs analyze SLAs to predict when they might be missed, or troubleshoot if they are not met. Other areas that should be regularly examined include:

- Parallel performance and other system-level performance reports
- Capacity consumption
- Usage statistics by user or group
- Query throughput (measured by query completion)
- Reporting to support query tuning
- Canary performance of protected applications
- Reporting to determine if SLAs are being met or trending out of compliance

3 Consider workload management
Imagine the chaos if sprinters, hurdlers and marathoners competed on the same track.

Planning and resourcing are necessary for any IT asset—especially a critical component like the data warehouse.

Something similar sometimes happens in data warehousing when different types of workloads run throughout the day. Effective workload management techniques prioritize tasks to balance simple and long-running queries with more frequent data loads, thus enabling each to finish within a reasonable time. To effectively set workload policies, a DBA must understand:

- Who is doing the work
- What type of work is happening
- The workload time frames
- If workload priorities need to change
- The potential for conflicting priorities
- If SLAs are protected

Because only a fixed amount of system resources exist, undoubtedly all user demands cannot be met. This is especially true since increasing the performance for one workload often comes at the expense of others, particularly when the system is near capacity.

One method of aligning business priorities with workloads is to encourage a meeting of the minds. Senior DBAs, business representatives and data warehouse management must collaborate to quantify rules and reach consensus on resolving priority conflicts.

In addition to the workload types and business priorities, contracted SLAs are an important element in the equation. Consider which applications and users need service levels, then take action to ensure the SLAs are documented and communicated to all parties. Moreover, policies must be in place to determine workload management modifications if service levels are not met.

4 Consider development life cycle processes
For some athletes, experimenting with new methods in strength and conditioning during peak season only slows their progress. Unfortunately, some data warehouse managers make a similar error when they mistakenly test and develop new applications on production systems. Because production systems need all available resources to meet business needs and ensure contracted service levels are met, development should be completed on a dedicated test/development system.

Another method to improve the active data warehouse performance is to ensure that design considerations are taken into account before porting applications to the Teradata system. Every database—including the Teradata Database—has its own design considerations to obtain the best
possible performance. Therefore, during an application life cycle it is important to ensure SQL is optimized for the Teradata Database and performance measurement and testing is accomplished before bringing the application into production.

5 Make tuning a priority

To achieve the best performance, athletes must constantly identify areas for improvement. For example, a better-fitting shoe might help a basketball player or a runner compete more effectively.

DBAs also realize that adding new users, applications and data to the data warehouse invariably creates inefficiencies that, over time, could harm system performance. To improve data warehouse performance, application tuning should be a priority.

Although performance tuning best practices help identify and tune highly consumptive or inefficient queries and processes, data warehouse managers and DBAs often lack the time, training or resources to tune existing applications. Since these efforts will likely return system processing power, the tasks are well worth the time invested.

As stated earlier, a good starting point to identifying poorly performing queries is DBQL. Query log data is available to measure the characteristics of each query, and DBQL step data can identify steps in query execution that might be causing the problem. For advanced DBAs, additional application tuning techniques can be found in “Strike it Rich,” Teradata Magazine, December 2007.

Ultimately, the value of application tuning more than offsets the time and resources required to complete the exercise.

6 Manage capacity

Professional athletes often bump up against the limits of their talent and discover that additional training only results in a small incremental improvement.

Fortunately for Teradata customers, the Teradata platform offers scalability and flexibility suitable for companies of all types and sizes. That said, a given system only contains a certain amount of initial processing power and disk space, and capacity constraints sometimes sneak up on DBAs. Although more nodes can be added to a system, data warehouse directors sometimes find they are out of capacity long after IT budgets are set for the fiscal year.

A well-documented capacity plan that considers disk space and system processing power, including CPU, I/O and memory, both to accommodate organic growth and future applications eliminates surprises.

Installation of the Performance Data Collection and Reporting database by Teradata’s Professional Services organization is one of the best methods to help a DBA diagnose capacity status. This historical performance database and report toolkit provides not only diagnostic reports and graphs to manage capacity but also information to tune applications and monitor system performance.

Planning and resourcing are necessary for any IT asset—especially a critical component like the data warehouse. A proper capacity plan ensures resources are available when needed and helps avoid unpleasant surprises.

Practice makes perfect

Optimum performance pays the largest dividends. To keep your Teradata system at its peak, follow these best practices and you’ll benefit from better decision making today, tomorrow and in the future.

Donna Becker has been with Teradata since 1993 and is director of America’s Professional Services Performance and Workload Management Center of Expertise.

Paul Barsch has more than 12 years of IT experience and directs Teradata Professional Services marketing programs.

For more information on Performance Services, visit Teradata.com/performance.