2
SQL Performance Methodology

This chapter describes the recommended methodology for SQL tuning.

> **Note:**
>
> This book assumes that you have learned the Oracle Database performance methodology described in *Oracle Database Get Started with Performance Tuning*.

Guidelines for Designing Your Application

The key to obtaining good SQL performance is to design your application with performance in mind.

Guideline for Data Modeling

Data modeling is important to successful application design.

You must perform data modeling in a way that represents the business practices. Heated debates may occur about the correct data model. The important thing is to apply greatest modeling efforts to those entities affected by the most frequent business transactions.

In the modeling phase, there is a great temptation to spend too much time modeling the non-core data elements, which results in increased development lead times. Use of modeling tools can then rapidly generate schema definitions and can be useful when a fast prototype is required.

Guideline for Writing Efficient Applications

During the design and architecture phase of system development, ensure that the application developers understand SQL execution efficiency.

To achieve this goal, the development environment must support the following characteristics:

- Good database connection management

  Connecting to the database is an expensive operation that is not scalable. Therefore, a best practice is to minimize the number of concurrent connections to the database. A simple system, where a user connects at application initialization, is ideal. However, in a web-based or multitiered application in which application servers multiplex database connections to users, this approach can be difficult. With these types of applications, design them to pool database connections, and not reestablish connections for each user request.

- Good cursor usage and management

  Maintaining user connections is equally important to minimizing the parsing activity on the system. Parsing is the process of interpreting a SQL statement and creating an execution

plan for it. This process has many phases, including syntax checking, security checking, execution plan generation, and loading shared structures into the shared pool. There are two types of parse operations:

- Hard parsing

  A SQL statement is submitted for the first time, and no match is found in the shared pool. Hard parses are the most resource-intensive and unscalable, because they perform all the operations involved in a parse.

- Soft parsing

  A SQL statement is submitted for the first time, and a match is found in the shared pool. The match can be the result of previous execution by another user. The SQL statement is shared, which is optimal for performance. However, soft parses are not ideal, because they still require syntax and security checking, which consume system resources.

Because parsing should be minimized as much as possible, application developers should design their applications to parse SQL statements once and execute them many times. This is done through cursors. Experienced SQL programmers should be familiar with the concept of opening and re-executing cursors.

- Effective use of bind variables

  Application developers must also ensure that SQL statements are shared within the shared pool. To achieve this goal, use bind variables to represent the parts of the query that change from execution to execution. If this is not done, then the SQL statement is likely to be parsed once and never re-used by other users. To ensure that SQL is shared, use bind variables and do not use string literals with SQL statements. For example:

  Statement with string literals:

```
SELECT *
FROM   employees
WHERE  last_name LIKE 'KING';
```

  Statement with bind variables:

```
SELECT *
FROM   employees
WHERE  last_name LIKE :1;
```

  The following example shows the results of some tests on a simple OLTP application:

```
Test                          #Users Supported
No Parsing all statements            270
Soft Parsing all statements          150
Hard Parsing all statements           60
Re-Connecting for each Transaction    30
```

  These tests were performed on a four-CPU computer. The differences increase as the number of CPUs on the system increase.

Guidelines for Deploying Your Application

To achieve optimal performance, deploy your application with the same care that you put into designing it.

Guideline for Deploying in a Test Environment

The testing process mainly consists of functional and stability testing. At some point in the process, you must perform performance testing.

The following list describes simple rules for performance testing an application. If correctly documented, then this list provides important information for the production application and the capacity planning process after the application has gone live.

- Use the Automatic Database Diagnostic Monitor (ADDM) and SQL Tuning Advisor for design validation.
- Test with realistic data volumes and distributions.

  All testing must be done with fully populated tables. The test database should contain data representative of the production system in terms of data volume and cardinality between tables. All the production indexes should be built and the schema statistics should be populated correctly.
- Use the correct optimizer mode.

  Perform all testing with the optimizer mode that you plan to use in production.
- Test a single user performance.

  Test a single user on an idle or lightly-used database for acceptable performance. If a single user cannot achieve acceptable performance under ideal conditions, then multiple users cannot achieve acceptable performance under real conditions.
- Obtain and document plans for all SQL statements.

  Obtain an execution plan for each SQL statement. Use this process to verify that the optimizer is obtaining an optimal execution plan, and that the relative cost of the SQL statement is understood in terms of CPU time and physical I/Os. This process assists in identifying the heavy use transactions that require the most tuning and performance work in the future.
- Attempt multiuser testing.

  This process is difficult to perform accurately, because user workload and profiles might not be fully quantified. However, transactions performing DML statements should be tested to ensure that there are no locking conflicts or serialization problems.
- Test with the correct hardware configuration.

  Test with a configuration as close to the production system as possible. Using a realistic system is particularly important for network latencies, I/O subsystem bandwidth, and processor type and speed. Failing to use this approach may result in an incorrect analysis of potential performance problems.
- Measure steady state performance.

  When benchmarking, it is important to measure the performance under steady state conditions. Each benchmark run should have a ramp-up phase, where users are connected to the application and gradually start performing work on the application. This process allows for frequently cached data to be initialized into the cache and single

execution operations—such as parsing—to be completed before the steady state condition. Likewise, after a benchmark run, a ramp-down period is useful so that the system frees resources, and users cease work and disconnect.

Guidelines for Application Rollout

When new applications are rolled out, two strategies are commonly adopted: the Big Bang approach, in which all users migrate to the new system at once, and the trickle approach, in which users slowly migrate from existing systems to the new one.

Both approaches have merits and disadvantages. The Big Bang approach relies on reliable testing of the application at the required scale, but has the advantage of minimal data conversion and synchronization with the old system, because it is simply switched off. The Trickle approach allows debugging of scalability issues as the workload increases, but might mean that data must be migrated to and from legacy systems as the transition takes place.

It is difficult to recommend one approach over the other, because each technique has associated risks that could lead to system outages as the transition takes place. Certainly, the Trickle approach allows profiling of real users as they are introduced to the new application, and allows the system to be reconfigured while only affecting the migrated users. This approach affects the work of the early adopters, but limits the load on support services. Thus, unscheduled outages only affect a small percentage of the user population.

The decision on how to roll out a new application is specific to each business. Any adopted approach has its own unique pressures and stresses. The more testing and knowledge that you derive from the testing process, the more you realize what is best for the rollout.