Optimizing Databases for High Performance and Efficiency

Bisallah Hashim Ibrahim[1], Atumoshi Adamu Yusuf[2]

[1]Department of Computer, University of Abuja, Nigeria. hbisallah@gmail.com
[2]Department of Computer, University of Abuja, Nigeria. atumoshi@yahoo.com

Abstract: Over the years, the database administrator's job has continued to grow more challenging. DBAs have always had complex and stressful jobs, but recently they have needed to confront the following trends: New Platforms - Most DBAs manage at least two different database platforms and many are responsible for three or more. This challenge is made even more complicated by a mix of operating systems. Added Complexity - Database giants say their database engines are "self-managed". While it is true that management improvements continue to be made, database software is escalating in complexity and there is much to learn, understand, and master with each new database release. More Data - The need for more data and more space relentlessly increases and databases continue to expand, both in size and number, adding more burdens for the DBA. This paper will outline techniques used to streamline and automate the management of these critical areas so as to deliver high database performance and availability.

[Bisallah H I, Atumoshi A Y. **Optimising Databases for High Performance And Efficiency.** *Rep Opinion* 2016;8(9):1-5]. ISSN 1553-9873 (print); ISSN 2375-7205 (online). http://www.sciencepub.net/report. 1. doi:10.7537/marsroj080916.01.

Keywords: Database, Performance, Efficiency, DBA, Analyst

1. Introduction

Regardless of the challenges of databases, the DBA is expected to keep key database systems available and optimized for high performance. To meet this goal, the DBA must put a strategy in place across all databases, regardless of platform. The high performance DBA should focus on the key areas of database management that contribute the most to high availability and performance, including, storage, management, performance, management and capacity Management.

Storage Management

Storage management is complex enough, but when a DBA has to manage different database platforms on a variety of operating systems, the challenge becomes even more extreme. Today's cross-platform DBA may be dealing with Sybase devices, databases, transaction logs, and segments in the morning and then have to switch gears to deal with Oracle tablespaces, tempfiles, and redo logs in the afternoon. Just remembering the storage syntax commands to manage the many diverse platforms can be a difficult task.

The two major storage management concerns are always availability and performance. The DBA must ensure that storage problems do not bring down an active database and must also make certain that the structures of storage containers and database objects do not adversely impact the performance of a system.

DBAs know that storage problems can quickly bring down a database. Just let an Oracle archive log destination reach maximum capacity or let a Microsoft SQL Server transaction log fail to expand and see how a dynamic transaction processing system reacts. Fortunately, database vendors have provided a number of new capabilities that address availability issues. All popular databases now sport an "auto-extend" feature allowing their underlying operating system files to automatically grow to meet the demand of increasing data volumes. Whether it is IBM DB2 containers, Oracle datafiles, or Microsoft SQL Server database and transaction log files, all have the capability to expand when necessary. Still, such features are not a 100% guarantee against failure. A SQL Server database file could meet an imposed growth limit or an Oracle temporary tablespace tempfile could be denied expansion at the operating system level if a disk drive runs out of space. Therefore, a DBA must still monitor storage to ensure availability.

While all DBAs understand how a serious storage problem can bring down a database, many DBAs do not realize how much storage can impact database performance. Fragmentation—both globally and at the object level—can result in significant unnecessary I/O for a database. Wasted space in tables and indexes can also cause problems and contribute to excessive table and index scan times.

A few of the database vendors have added new features to protect against these problems. For example, Oracle10g introduced automatic storage management (ASM) which allows a DBA to provide the database engine with a set of disk drives for the database to automatically manage, in terms of file placement. If hotspots are discovered, automatic relocation is accomplished without DBA intervention. In terms of global fragmentation assistance, Oracle8i

and above offer locally-managed tablespaces, eliminating the need for full tablespace reorganizations.

In addition, object storage problems can create major performance roadblocks in a database. A DBA must understand how to detect and eliminate these issues as quickly as possible.

Regardless of platform, object storage problems generally include:

Wasted Space - When a table, index, etc., has more allocated space than it needs, it may not seem like a problem, but this can adversely impact performance. For example, if the pages of a

SQL Server object contained in a data warehouse are not 100% utilized, it will examine more pages than necessary during table scans. This, in turn, contributes to increased response time for user queries. The same problem arises with Oracle tables that have high-water marks above the areas where data resides. Poor Extent Proximity - A database engine's read-ahead mechanisms work more efficiently when objects are contained in extents that are next to or in close proximity to one another. For example, SQL Server's Read Ahead Manager can read much larger groups of data when its target object has extents and pages that are contiguous. Sybase is the same. Conversely, when an object has poor extent or page proximity, scan times can increase.

Out of Sync Data Order - When indexes have a logical order that does not match the actual physical order of data stored in the database, index access performance can be affected. If the logical and physical order is in sync, the disk head can scan in one direction instead of moving back and forth to obtain the needed information. Otherwise, the disk head will skip across the disk many times. For example, low cluster ratio readings for Sybase clustered indexes are undesirable because such ratios almost always indicate a situation that requires reorganization of the data order.

Forwarded Rows - Each row in a table should fit on a single page, if possible, except for large object/text tables. If a row expands because of the growth of variable length data, however, the database may either relocate the row to another page and leave behind a pointer to the new page, or be forced to split the row between two or more pages. Neither situation is preferred as they increase the amount of I/O necessary to obtain the row. This is seen in Oracle as chained or migrated rows and in SQL Server and Sybase as forwarded rows. So, how does a DBA in charge of multiple DBMS platforms detect and diagnose these object storage problems? After all, each engine has its own set of diagnostics that apply to each situation.

How can a DBA ensure that no downtime is attributed to space outages across their many servers? It is important that the high performance DBA establish a proactive storage monitoring plan that alerts them to space issues before they become problems. Such a plan should take into account the cross-platform database aspect and also focus on being completely automated 24/7, at least for critical databases.

Advantages

The Performance Center is the ideal tool for detecting and diagnosing object storage problems. Performance Center is an around-the-clock, cross-platform, fully-automated database monitoring solution that allows you to configure storage thresholds and proactive notifications across all your critical databases. With Performance Center, you will always be alerted to storage problems before anyone else. What about diagnosing and remedying database and object storage problems that are affecting performance? This can be a challenge for a cross-platform DBA, given the vast differences in diagnostic and treatment methods that exist across diverse database platforms. Embarcadero's award-winning DBArtisan XE contains an advanced storage diagnostic and management system that takes all the guesswork and labor out of View key information regarding problems in handling cross-platform database storage problems.

Space Analyst, available within DBArtisan XE, allows you to handle complex storage dilemmas even if you are unfamiliar with the underlying database platform. Space Analyst alerts you when storage issues in the database objects are draining performance by visually identifying all storage problems and pinpointing the exact objects that require attention. The advantage of Space Analyst is that you do not have to be familiar with a particular database platform or with the types of space problems typical to a particular platform, because Space Analyst does all the analysis for you. The product includes a complete set of thresholds designed to uncover space issues that are threatening overall system performance. The thresholds can be used out of the box, or you can customize them to meet the unique needs of your environment.

DBArtisan XE's Space Analyst tool can also help fix complex space issues, quickly addressing even the most complex reorganization. Whether a full Sybase database, a large Oracle tablespace, or individual SQL Server or DB2 objects, Space Analyst can reorganize and rebuild your database objects for optimum performance.

Space Analyst can even be used proactively to set your reorganization plans on autopilot. The Space Management Wizard helps you create a reorganization

find-and-fix job that searches the entire database (or a particular set of objects) for objects that violate your specified set of thresholds. If any problem objects are found, Space Analyst with the ability to display data in both text and graphical dynamically reorganizes them for better performance.

Performance Management

Every DBA wants their databases to run as fast as possible. The same can be said for every person who uses a system that connects to those databases—they want their queries and processes to have the shortest possible response times. Again, the situation is complicated when you have more than one database platform. As with storage management, the variables and remedies differ from one platform to the next. As a result, you need to put together a platform-neutral roadmap that applies to each of the database engines you manage.

The starting point is to define the different methods of analysis that will be used across all platforms and then apply a set of diagnostics and actions that can be used for each platform under each method. These analysis methods include:

- Bottleneck/Response Time Analysis
- Workload Analysis
- Ratio Analysis

The following examines how each method can be used to manage the performance of any database.

Bottleneck/Response Time Analysis

Regardless of the platform, when a database is up and running, every connected process is either busy working or waiting to perform work. A process that is waiting may mean nothing, or it can be an indicator that a database bottleneck exists. DBAs use bottleneck or response time analysis of performance to determine if perceived bottlenecks in a database are contributing to a performance problem.

Bottleneck analysis is an essential method of measuring performance because it helps a DBA track where a database has been spending its time. If Oracle latch contention or heavy DB2 table-scan activity has been dragging a database's performance down, a DBA can use bottleneck analysis to pinpoint the root cause. Once one or more of the potential sources have been identified, the DBA can drill down to detail about the sessions and objects that are causing the problem.

This methodology is the strategy of choice for top performance analysts in the industry. This being the case, nearly every database vendor has tailored their engine to report metrics that can be intelligently used to analyze bottlenecks and response times. Oracle10g introduced new metrics in its V$ performance views that enable DBAs to understand bottlenecks and response times at global and session levels. Microsoft SQL Server made wait events available via an undocumented DBCC command (DBCC SQLPERF(WAITSTATS)), and Sybase offers new monitoring views in engine versions 12.5.03 and above.

Workload Analysis

Workload analysis involves the investigation of two critical areas of database performance:

- Session resource consumption and activity
- SQL execution analysis

When performance on a database drops suddenly, it is not uncommon to find one or two sessions that are generating the bulk of the workload. The issue is system balance. In a well-balanced system, no single session should consume the bulk of resources for an extended period of time, with the exception of batch job processes that are run safely in non-peak hours. If individual sessions are utilizing a majority of system resources, the DBA should examine each problem session in detail to uncover session activity.

Normally, this can be easily accomplished by viewing session metadata coupled with resource consumption and statistical execution statistics. Some database engines, like Oracle10g, provide vast details regarding a session's current and historical activities.

Understanding current and historical SQL execution patterns will enable a database analyst to have the second set of data points needed to properly perform workload analysis. Optimizing SQL code will produce the second-best performance-enhancing boost available for a database—with proper physical design being the first. But what set of metrics should you use to evaluate 'good' versus 'bad' SQL? Naturally, factors such as overall elapsed time, CPU time, and I/O activity play a part. But other, more subtle factors can make a difference—like the number of times a statement is executed, or whether sorts are done in memory or on disk. This is where SQL analysis becomes more of an art form than an exact science.

All the database vendors offer a window into SQL analysis. Oracle10g has the most complete set of metrics offered by any vendor. Microsoft SQL Server offers code profiling utilities that can be used to collect and measure SQL and stored procedure executions. Sybase has made a good start with its SQL execution views in versions 12.5.03 and above. DB2 offers its own SQL tracing capabilities.

Ratio Analysis

Ratio-based analysis has been around for many years, and used to be the only technique database administrators utilized to diagnose the cause of a database slowdown. DBAs are all too familiar with the Oracle buffer cache hit ratio, Microsoft procedure plan hit ratio, Sybase data cache hit ratio, and so forth. Many performance gurus advise that such ratios are now worthless and misleading because of advances in bottleneck/response time analysis. There is an element

of truth to these claims, but some ratios are still quite valuable, such as the Oracle library cache hit ratio and the Microsoft and Sybase procedure plan hit ratios.

Other more obscure ratios can also be useful. For example, if you were told that a Sybase table had 100 forwarded rows, would you consider this good or bad? Of course, you cannot answer this question unless you know how many total rows are in the table. But if you knew that 98% of all rows in your Sybase table were forwarded, this is enough information to make a reorganization decision. In these types of situations, the application of proper ratio analysis can be advantageous.

Storage Advantage

As with storage analysis, the cross-platform DBA needs a way to intelligently utilize each performance analysis technique for the platforms they support. Even if you handle only one platform, like Oracle, you still need a way to quickly use each performance method to diagnose and repair performance problems in your database.

It provides two solutions to streamline cross-platform performance. Performance Center offers the 24/7 monitoring that is essential for your critical databases. You can set up customized thresholds and notifications across all platforms on any key ratio, bottleneck, or response time so you will be alerted instantly if an abnormality occurs.

For detailed analysis, Embarcadero Performance Analyst (a feature within DBArtisan XE) allows instant access to key information across all methodologies with strong drill-down capabilities enabling you to get to the heart of any performance matter. Running within the heterogeneous DBArtisan XE administration tool, Performance Analyst organizes a database's performance metrics into each type of methodology to support easy decision making. There are no complicated scripts to run or hardcopy output to interpret; Performance Analyst visually presents all the information you need. It does not matter what platform you are working on or whether you are familiar enough with a particular platform to know if your database is running well.

DBArtisan XE, with Performance Analyst, includes a complete set of thresholds that automatically determine the health of your key databases. Whether you are using Oracle, SQL Server, Sybase, or DB2, everything is organized in the same easy-to-understand manner so you can quickly uncover any response-time or bottleneck problems, find rogue user sessions or SQL code, and determine if any key ratios are out of balance. A global alarm log keeps you completely up to date on current problems and one-click integration with DBArtisan XE allows you to fix many of the problems found in Performance Analyst. For troubleshooting code issues, DBArtisan XE also contains cross-platform code debuggers and profilers that can tell you exactly what line of code is taking the longest to run in procedures or packages of up to thousands of lines.

Capacity Management

Capacity management is typically neglected by DBAs, as their day-to-day activities simply leave no room for historical and proactive analysis. This is unfortunate because proper capacity planning can help both the DBA and management, answering important questions such as "How much more storage will this database need in six months?" and "Is this database server currently underutilized?"

Capacity planning generally involves three processes:

1. Collection of key database metrics
2. Historical analysis of collected metrics
3. Forecasts of future needs

A powerful interface takes all the guesswork out of determining a database's performance levels and instantly indicates if a key database needs attention.

Every shop will have different needs, but in general, most should be collecting the following information:

• Snapshots of global storage usage, including Sybase device metrics, SQL Server database statistics, Oracle tablespace data, and database object statistics that concern space usage. Elements such as total allocated space, used space, and free space should all be collected as well as other performance-related items like number of chained rows.

• Snapshots of performance metrics, including key database statistics such as I/O performance and wait events.

• Snapshots of resource consumption, including metrics such as CPU usage, memory usage, and user traffic while some DBAs have been able to get by using scripts and manual processes to manage storage and performance, the burden in using such techniques for capacity planning is usually too high. With complications such as cross-platform environments, the need to build and maintain a repository that holds all the necessary statistics, manual weaving-together of collection scripts and scheduled jobs, and the generation of complex historical and forecasting reports, it is obvious that manual maintenance is not practical for the high performance DBA.

References

1. Baker, F., *Organizational Systems: General Systems Approaches to Complex Organizations*, Irwin, Homewood, IL, 1973.
2. Bakos, J.Y., "Dependent Variables for the Study of Firm and Industry-Level Impacts of Information Technology," in J.I. DeGross and C.H. Kriebel (Eds.), *Proceedings of the Eighth*

International Conference on Information Systems, Pittsburgh, Pennsylvania, (December 6-8, 1987), pp. 10- 21.
3. Bertalanffy, L. von, "General Systems Theory," *General Systems*, 1 (1956), 1-10.
4. Blau, P.M. & Scott, W.R., *Formal Organizations*, Chandler, San Francisco, CA, 1962, 194-221.
5. Bluementhal, S.C., "Breaking the Chain of Command," *Business Automation*, (November, 1963), 20-27.
6. Bradley, S.P., Hausman, J.A., & Nolan, R.L., "Global Competition and Technology," in S.P. Bradley, J.A. Hausman & R.L. Nolan, *Globalization Technology and Competition: The Fusion of Computers and Telecommunications in the 1990s*, Harvard Business School Press, Boston, MA, 1993, 3-31.
7. Browning, J., "Information Technology: The Ubiquitous Machine," *The Economist*, (June 16, 1990), p. 5.
8. Cash, J.I., & Konsynski, B.R., "IS Redraws Competitive Boundaries," *Harvard Business Review*, 63,2, (March- April 1985), 134-42.
9. Churchman, W., *The Systems Approach*, Delacorte Press, New York, 1968.

8/29/2016