Five Best Practices for Maintaining an Enterprise Geodatabase

Transcript

Welcome to the ESRI Instructional Series Podcasts. I am Derek Law, and I work at ESRI's main office in Redlands, California. I specialize in geodata management and ArcSDE technology. In this podcast, I will discuss *Five Best Practices for Maintaining an Enterprise Geodatabase*, also called *Enterprise ArcSDE*. This podcast is intended for new and intermediate enterprise geodatabase administrators, ArcGIS Server users, and GIS managers who would like some general tips on maintenance strategies to upkeep performance of an enterprise geodatabase. Experienced enterprise geodatabase administrators may also learn a few tricks.

Before we start, a disclaimer: All contents presented in this podcast are recommendations that will typically improve performance for many enterprise geodatabase application scenarios. However, in some cases, these recommendations may not yield better performance results, in which case, additional performance testing and system configuration tuning may be needed.

We recommend that you first implement these best practices within a test environment before you apply them to a production environment, if possible, and as a precaution, we also recommend you back up your geodatabase just in case. As an enterprise geodatabase administrator, how many times have you heard your users comment or complain that the enterprise geodatabase is slow? I'm sure many of you are smiling right now, and thinking to yourselves, "If only I got a penny every time I've heard that remark."

The performance of an enterprise ArcSDE geodatabase is influenced by many factors, such as hardware configuration, network configuration, network traffic, and the number of concurrent users.

The objective of this podcast is to provide you with some general best practices to follow. They are not database platform-specific, but they are general tips that will hopefully enable you to improve the performance of your enterprise geodatabase.

1. Increase the frequency of updating statistics on tables

Statistics in a database describe the column data stored in tables. They help the database Query Optimizer to estimate the selectivity of SQL expressions, and enable it to accurately assess the cost of different query plans. The optimizer then chooses the most efficient execution plan for retrieving and/or updating data in the database. Having poor statistics is a frequent cause of poor

performance. Keeping accurate and up-to-date statistics will help improve database performance, because this will enable the Query Optimizer to make more accurate assessments of query execution plans.

The frequency of updating statistics will depend on the editing activity in the geodatabase. Typically, more editing activity means you should update statistics more frequently. This is the responsibility of the database administrator, and not the ArcSDE software, which does not maintain statistics. You can update statistics for a table or feature class in ArcCatalog with the Analyze Component's dialog box. It updates the statistics for the supporting tables that are associated with the selected object.

You should also update statistics on the SDE repository tables. This can be done with database management software. As a general rule of thumb, we suggest you update statistics at least weekly or monthly, and typically before and after a *compress*, which can be automated at the database level. There is one exception: in situations where all users are editing just the SDE.DEFAULT version, you should just keep the statistics you collected before a compress. This will ensure that the query optimizer knows the delta tables are still active.

2. Rebuild indexes on tables

Indexes are used in a database to help speed up the retrieval of rows from a table, and they are also used by the database Query Optimizer when assessing query plans. As tables are modified by updates, inserts, and deletes of records, the corresponding indexes can become fragmented and unbalanced. This leads to increased I/O processing, which affects performance. This tip works in conjunction with the previous one. If you update statistics frequently, in turn you should consider rebuilding indexes if they are fragmented. Both actions will help improve performance.

In general, accurate statistics help to define a good index. You can assess the usefulness of an index with database management tools by monitoring its usage. Another benefit of rebuilding indexes is that you may reclaim disk space that was caused by its fragmentation. In versioned editing environments (where edits are performed daily), you may want to consider rebuilding indexes at regular intervals (for example, weekly or monthly), to keep performance degradation under control. We recommend you rebuild indexes after a compress. You can rebuild indexes within a database management program, or with the ArcSDE commands. For more information,

see Knowledge Base (or KB) Article #24518, titled, *FAQ How to Improve ArcSDE Performance by Analyzing Data and Rebuilding Indexes* on the ESRI support site.

3. Plan parent-child version relationships carefully

The versioning environment within an enterprise ArcSDE geodatabase enables users to implement and sustain complex business workflows. Typically the number of versions and how they are interrelated will depend on your business workflow. It is important to properly manage versions in the geodatabase, because poor version management will impact performance. You should keep the following in mind: every edit in the geodatabase is adding a state to the state tree. A state tree represents the total number of edit states stored in an enterprise geodatabase. Think of it conceptually like a flow chart diagram of circles and lines that flows from top to bottom. Each circle represents an edit state, and each state is linked by a line showing the edit history in the geodatabase.

A state tree, typically, has a structure similar to an upside-down tree, starting with one circle at the top (let's say it's zero), and flowing down in many branches. For example, a typical enterprise geodatabase may have approximately one million edits per day, resulting in hundreds of thousands of edit states in a state tree.

Ideally, you want to keep the state tree as simple and as small as possible. Versions are pointers to an edit state, and they will "pin" the state tree; in other words, they will keep its structure complicated. This can affect performance, because it may take queries longer to execute. Therefore, the more complex the versioning model (in other words, the more versions you have), means more potential records than the delta tables, which means potentially slower performance.

In general, you should try to do the following:

- Reconcile versions to the SDE.DEFAULT version as soon as you can.
- Delete versions when they are no longer needed
- Avoid creating versions that will never be reconciled with SDE.DEFAULT.

You could also run multiple reconcile services, to reconcile without posting as many older versions as possible each evening. This operation will simplify the state tree, so that when a

compress is finally executed, it will trim the state tree. Version management can be performed in the Version Management dialog box in ArcCatalog or ArcMap. For more information, read the ESRI technical white paper titled *Managing Workflows with Versions* on the ESRI support site.

4. Compress the geodatabase often

I have referenced the compress operation several times already in this podcast. Compressing an enterprise ArcSDE geodatabase helps maintain database performance by removing unused data. Specifically, it does two things:

- First, it removes unreferenced dates, and their associated delta table rows.
- Second, it moves entries in the delta tables that are common to all versions into the base tables, thus reducing the amount of data that the database searches through when executing queries. In effect, a compress will improve query performance and system response time by reducing the depth and complexity of the state tree.

When a large volume of uncompressed changes have accumulated in an enterprise geodatabase, a compress operation can take hours or even days. This is another very common cause of poor performance. To avoid this, you should compress on a regular basis (daily, weekly, and after periods of high editing activity). Users can stay connected to the geodatabase during a compress, but we suggest that all users be disconnected for the compress operation to be fully effective.

Remember to update statistics before and after a compress, and note the one exception I mentioned earlier. The compress command is available in ArcCatalog. You add the command from the Customize dialog box, and you must be connected as the SDE user to execute it, or you could execute a compress with SDE commands. For more information, see KB Article #29160 titled *How to Compress a Version Database to State Zero* on the ESRI support site.

5. Monitor system resources

When experiencing intermittent performance issues, it may be helpful to monitor the memory and CPU usage on both the client and server machines. This may help identify on which machine the performance bottleneck is occurring. For memory, it is important to ensure that the operating system is not running out of available memory and using *swap space* (in other words, *virtual memory*). Enterprise ArcSDE typically needs at least one gigabyte of free disk space to operate

efficiently. For CPU, you want to avoid and reduce how often the system hits a hundred percent CPU usage. Some troubleshooting suggestions to improve server performance include:

- Closing unrelated applications on the server
- Performing a database *trace* to examine and review performance (what's in the database)
- You could have users switch from application server connections to *direct connects* (this will put more workload on the client and less on the server)

For tips on improving client performance, refer to the ESRI Instructional Series Podcast titled *Performance Tips and Tricks: ArcSDE Client-Side Optimization.*

To review, the performance of an enterprise ArcSDE geodatabase is influenced by many factors, such as hardware configuration, network configuration, network traffic, and the number of concurrent users. Five best practices for maintaining an enterprise geodatabase are:

1. Increase the frequency of updating statistics on tables
2. Rebuild indexes on tables
3. Plan parent-child version relationships carefully
4. Compress the geodatabase often
5. Monitor system resources

For more information, go to the ESRI support site: support.esri.com; access the ArcGIS Online Help documentation, and search for the topic *An Overview of Tuning an ArcSDE Geodatabase*.

ESRI also offers several instructor-led training classes on the configuration and tuning of enterprise ArcSDE geodatabases, based on DB2, Informix, and SQL Server database platforms. There is also an instructor-led class called *ArcGIS Enterprise Systems: Performance and Scalability.*

Thank you for listening, and stay tuned for future podcasts.