Geodatabase Replication: An Overview

Transcript

Welcome to ESRI's Instructional Series Podcasts. I'm Owen Evans, and I work in the Technology Center at ESRI in the Washington D.C. area. This podcast is an introduction to geodatabase replication that is appropriate for GIS managers, ArcGIS Desktop users, and it also has some useful information for database administrators.

Geodatabase replication is a new capability of ArcGIS at the 9.2 release. It allows you to copy data from one geodatabase to another, independently edit each data set, and then transmit those changes so that the data sets in both geodatabases are kept in sync. It is flexible, and offers options to support many different work flows. It is also important to note that geodatabase replication is not the same as DBMS replication. Geodatabase replication is implemented using the versioning environment created by ArcSDE technology. It does not depend on any replication capabilities of the underlying DBMS.

In this podcast, when I use the term *replication*, I'll be using it to refer to geodatabase replication, unless I indicate otherwise.

There are a fantastic amount of possibilities that are unlocked by geodatabase replication. In this podcast, I'll talk about the basic capabilities of replication and describe a few common situations where it can be used. In future podcasts, I'll cover how you create and synchronize replicas in more detail, and also how you can leverage geodatabase replication along with the new types of geodatabases that have been introduced at ArcGIS 9.2, to simplify some very complex workflows.

First, I'd like to describe some typical scenarios where replication might be used. Many organizations have several offices that are located across the country or around the world. Copies of shared data sets could be maintained locally at each of these different facilities, and replication would be used to transfer edits made by each facility to all the other facilities.

Another use case for geodatabase replication is load balancing. Suppose an organization has recently noticed a performance decrease when using their data server due to an increased number of editors. The activities of the data editors could be separated by using a production database to support those editors, and a publication database to support the data viewers and Web applications. Replication can be used to push changes from a production database to the publication database to keep those two databases in sync.

A third case is where field workers are creating or editing data outside an office environment. Here, replication can be used to copy data sets that could be edited in the field, and then those edits can be synchronized back to the main database when the field assignment is completed.

Another situation where replication might be used is where a municipality might replicate some utility data sets to a contractor. Perhaps this contractor has a multiyear contract for various construction and maintenance items on the city's waste water collection system. The contractor would periodically synchronize the data sets and send new sewer line features, as well as condition rating attributes for existing features, back to the client's central geodatabase.

Now that you've heard about some use cases for replication, what do you need to do to get ready to implement a geodatabase replication workflow? Well, it's recommended that you have a well-defined data model for the data you are replicating. The users, departments, or organizations that are planning to share data should try to agree on a common set of attributes for the feature classes that are to be shared. However, geodatabase replication provides the flexibility for all participants to decide to maintain different sets of attributes if that is required. Also, the users involved in creating and synchronizing replicas should be comfortable in versioning and versioning concepts, such as states, reconcile and post, and geodatabase compression.

Next I'd like to discuss a few concepts that will help you understand exactly what a replica is. The basic replication relationship is between two geodatabases: the parent replica, which contains your original data, and the child replica, which contains a copy of the data you chose to replicate. These two geodatabases contain what is collectively referred to as a replica pair. This basic relationship can be combined in many ways. For example, a geodatabase that contains the child of one replica pair can contain the parent in a second replica pair. By chaining these replica pairs together, you can create very elaborate data sharing workflows. In an agency that has local, state, and federal offices, a state office's geodatabase, which represents a member of the middle tier of this agency, would be the parent of several local office geodatabases, and it would also be the child of the federal geodatabase.

When creating replicas, you have many options for defining the subset of data you wish to replicate. First, you choose the version to replicate in your parent geodatabase. You can choose to replicate the default version or any named version that exists in the parent geodatabase. You can

also select specific datasets to replicate. It is not required to replicate the entire geodatabase. Finally, you can apply definition queries, spatial filters, and selection sets, to further define the data to copy; and also choose to replicate related data as well.

Geodatabase replication provides three types of replicas: Checkout/check-in, two-way, and one-way. Checkout/check-in replication is equivalent to the disconnected editing functionality that existed in releases prior to ArcGIS 9.2. It allows you to copy data to another geodatabase, edit that data, and then when you're done, transfer those edits back to your original database. When using checkout/check-in replication, you can perform this check-in operation only once. If you need to perform more edits once you have checked in the data, you must create another checkout replica.

Two-way replication allows you to send changes back and forth between replicas many times without having to recreate the replica. Changes can be sent either from the parent to the child, or from the child to the parent, or you can send changes in both directions at the same time. Two-way replication is very useful in situations where two departments or organizations want to share data that they both need to edit. Perhaps a company has two offices, one in New York and one in Los Angeles. Both offices use many of the same data sets, but they might have different editing responsibilities. Two-way replication could be used to allow both offices to send and receive edits to their shared data sets.

The third type of replica you could establish is a one-way replica. One-way replication is similar to two-way replication, in that you can send changes many times after establishing the replicas. However, as the name implies, you can only send changes in one direction, from the parent to the child. This type of replication is useful in the type of situation I've mentioned earlier, with the production and publication database. Changes made by the editors on the production database can be pushed to the publication database when necessary. By design, no one will be editing data on the publication database, so there's no need to send changes from it back to the production database.

What are the connectivity requirements for geodatabase replication? Well, geodatabase replication can be accomplished in several types of connectivity environments. Replicas can be created and synchronized using local area network connections, Web services, published securely over the Internet, or even in disconnected environments using XML files. You can use the local

area network if both geodatabases can see each other over your office network or through a VPN connection. If you can create a database connection to two ArcSDE geodatabases in ArcCatalog, you can replicate data between them using this type of connection. Using the Web replication capability of ArcGIS Server, you can publish your geodatabase as a Web service on a server that is outside your organization's firewall. Once this is done, users outside your organization can connect to your geodata server over the Internet using a URL that you provide them. Security can be applied using your Web server security model to prevent unapproved access to your data.

Finally, if you have two geodatabases that cannot be connected in any of these ways, either due to security policies or network outages, replicas can be created and synchronized using XML files. In this method, the information required to create or synchronize the replica is written to an XML file, and the user can transport that file to the relative replica's location via e-mail, FTP, or perhaps even on a CD or DVD.

Well, where can these replicas exist? Well, replicated data sets exist in geodatabases. Specifically, two-way and one-way replicas must reside in ArcSDE geodatabases. This includes enterprise geodatabases, as well as the new workgroup and personal ArcSDE geodatabases that can be deployed on the Free SQL Server Express database platform. And since replication is implemented by ArcSDE technology, geodatabases that use different DBMSs can participate in replication with each other. For example, a SQL Server enterprise geodatabase can replicate data to an Oracle enterprise geodatabase, which could in turn replicate data to a SQL Server Express workgroup geodatabase. For checkout/check-in replicas, the parent replica must be an ArcSDE geodatabase, but the child replica may be created in either an ArcSDE geodatabase, a file geodatabase, or a personal geodatabase.

Well, just about time to wrap up, so let me summarize what I covered in this podcast. First, I presented a few situations where geodatabase replication would be useful for establishing a distributed system for your data. Then we discussed the three different types of replicas (two-way, one-way, and checkout/check-in), and finally, I talked about how replication can be performed in several different connectivity environments, and also what type of geodatabases can participate in replication workflows.

A free training seminar on replication was recorded on November 9, 2006, and is now available on ESRI's Web site, and I encourage you to watch that if you're interested in learning more. And

as I mentioned, I'll be posting a few more instructional podcasts on this subject in the future. Thanks for listening.