GIS Database Lifecycle Management at GEO*Fidelis* West

Authors:

Wendell Hoskins, Systems Manager at GEOFIWEST
>Technology Associates Intl. Corp.
>>whoskins@taic.net

Joe Simons, Database Administrator at GEOFIWEST
>Technology Associates Intl. Corp.
>>joseph.simons@taic.net

Abstract:

Under the Marine Corps GIS Program (GEO*Fidelis*), the Western Regional Geospatial Data Center (GEOFIWEST) hosts and manages enterprise geodatabases for nine Marine Corps Installations in the western United States. At GEOFIWEST, a series of database management strategies are followed in order to create a high-availability environment to best serve the Marine Corps user needs. These policies incorporate planning and management strategies for the USMC data including database structure, migration, modeling, user access, performance, maintenance scripting, versioning organization, and disaster and recovery models. This presentation will primarily focus on best practices and lessons learned during recent data and systems migrations.

Overview:

The goal of the GEO*Fidelis* program is to provide accurate and up to date geospatial information in support of Installation management, to improve our stewardship of natural resources, protect the environment, and support the training of Operating Forces. The system and database technology that supports this mission is a critical component towards achieving this goal. The purpose of this document is to provide a high-level overview of how the goals of Marine Corps are being addressed through system and database architecture policies.

The GEO*Fidelis* program is a regionalized program overseen by Headquarters Marine Corps, with regional data centers in the east at Camp Lejeune and west at Camp Pendleton. This document and proceeding presentation focuses on the system architecture at the Western Regional Geospatial Data Center.

System Architecture:

At GEOFIWEST, hundreds of users access ArcGIS Desktop suite via Citrix thin-client technology. Geodatabases are hosted using ArcSDE on SQL Server 2005, supporting nine Marine Corps Installations, as well as support for other non-Installation specific data requirements.

ArcGIS Desktop is installed on six Citrix Presentation Servers. ArcGIS Desktop is load-balanced among the Citrix servers for application resilience and uptime. The database

architecture is comprised of two separate active/active Microsoft SQL Server 2005 (64-bit) clusters. Both clusters consist of two nodes each with each node possessing 16 CPU cores and 32 GB of memory. A SQL clustered resource is configured for each Marine Corps Installation. Marine Corps Installations are halved; with one half assigned to cluster 1 and other half assigned to cluster 2.

Furthermore, Citrix servers and Microsoft SQL Server 2005 (64-bit) clusters boot from storage area network (SAN). The decision to boot the GEOFI WEST geospatial environment from SAN is driven by a disaster recovery model that requires simplified disaster recovery and failover.

Database Architecture:

All geodatabases at GEOFIWEST are hosted with ArcSDE version 9.3.1 on top of SQL Server 2005. In addition to each Marine Corps Installation in the GEOFIWEST region currently having a dedicated ArcSDE geodatabase primarily used for Desktop access, a series ArcSDE geodatabases exist for other mission critical needs such as ArcGIS Server web presentation, support for headquarters viewing, and Marine Corps Pacific support.

The SQL Server Filegroup organization and DBTUNE configuration is designed to support two main goals: performance and simplification of recovery. Separate disks are used to store raster and vector data, with the raster data being separated and residing in a read-only filegroup when data is not loading to increase performance. Geodatabase Versioning and Geodatabase Archiving are implemented at GEOFIWEST, and so in order to minimize I/O, separate filegroups have been established for Adds, Deletes, Archive data, and highly accessed version support tables such as MV_TABLES_MODIFIED.

When using the ArcGIS Desktop suite, all users access data using a direct connection using windows authentication. The login process is simplified for the end user by using a GEOFIWEST built Login tool that automatically establishes a connection to the user's edit version. The version tree implemented at GEOFIWEST supports long and short transactions while still maintaining and good balance with performance by scheduling reconcile and posts for day to day edits and isolating the not to be posted edits. Archiving is always available for rolling back data in the event of poor decisions in the sort transaction versions.
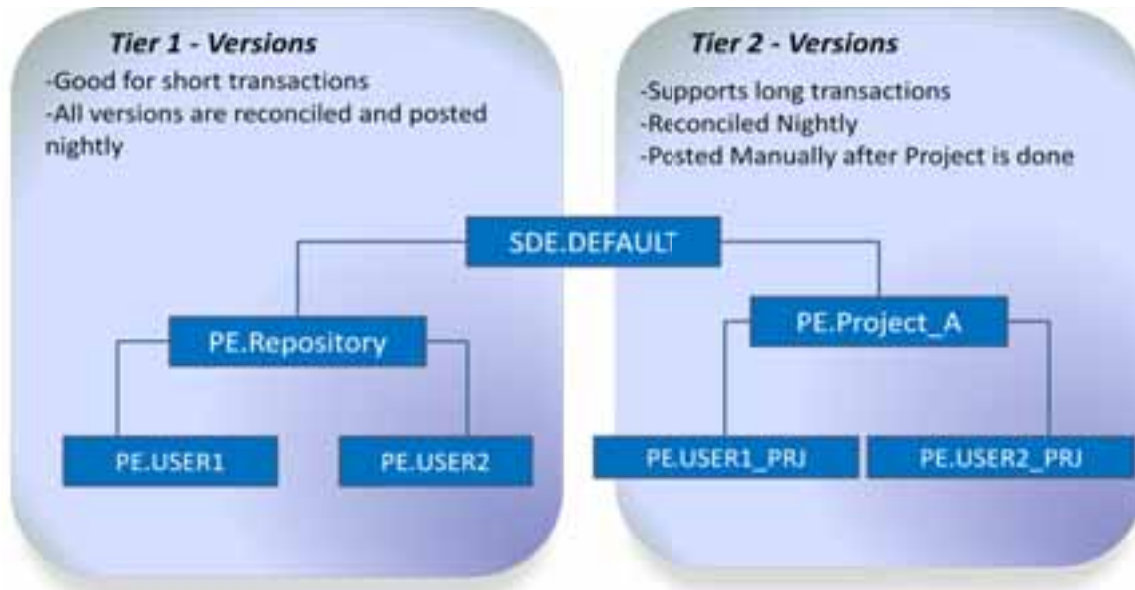
Backup and Recovery:

Data availability is ensured using a series of methods ranging to disk cloning, Avamar's EMC Backup and Recovery product, and SQL Server backups. Each the processes mentioned are automated via various scheduled processes. On top of that, common day to day edit transactions can be recovered by the user with version history, geodatabase archiving, or recovery from an aggregated database and/or file geodatabase backup.

Scheduled Maintenance:

Reconcile and Post occurs via a scheduled task each night on all short transaction versions via a python script. The python script does not post in the event of conflicts, and the Installation GIS manager is notified if conflicts are detected. It also leverages attribute-level conflict detection in order to limit the number of conflicts, as well as logging and exception handling.

Compress and Analyze are combined with the Reconcile and Post script, and as a result, are run on a nightly basis. Although it is not common to compress every night, the end result of these process increases SQL performance, so long as indexes are rebuilt on the database to address potential fragmentation created from the compress. After the python scripts are run, a SQL Job runs each night that does exactly that, loops through all data owned by any user in the sde repository and rebuilds indexes and updates statistics again, this time through SQL. Lastly, approximately once a month Spatial Indexes are rebuilt on each database in order to address any major loads that may have taken place

```python
try:
    gp.ReconcileVersion_management(owner_connection, "USER1.Editor", "HI.Repository", "B"
    print gp.GetMessages() + "\n"
    output.write(gp.GetMessages()+ "\n")
except:
    print gp.GetMessages() + "\n"
    output.write(gp.GetMessages()+ "\n")

print " -------------------------------- Compressing... ----------------------------
try:
    # Compress the database
    print "Begining Compress..." + "\n"
    gp.toolbox = "management"
    gp.compress("G:\\Admin\\ConnectionFiles\\sde@HI.sde")
    print gp.GetMessages() + "\n"
    output.write(gp.GetMessages()+ "\n")
except:
    print gp.GetMessages()
    output.write(gp.GetMessages())

print " -------------------------------- Analyzing... ----------------------------
try:
    # Loop through and analyze all Feature datasets
    print "Begining Analyze Loop..." + "\n"
    gp.Workspace = owner_connection
    FCList = gp.ListFeatureClasses ("*", "all")
    FC = FCList.Next()
    while FC:
        gp.Analyze_management(FC, "BUSINESS;FEATURE;ADDS;DELETES")
        # print gp.GetMessages() + "\n"
        # output.write(gp.GetMessages()+ "\n")
        FC = FCList.Next()
    # Loop through and analyze all the Feature classes
```

Conclusion:

In conclusion, GEOFIWEST's system and database architecture is designed to provided high availability, data security with multiple backup and restore options, and geodatabase performance for its end users. The database architecture implemented on SQL Server 2005 implements an approach using filegroup separation, raster and vector isolation, and scheduled processes supporting maintenance and backup policies. The versioning structure gives users options for long and short term editing needs while still maintaining performance and maximizing performance gains through the available scheduled geodatabase options.

References:

Geo*Fidelis* Program

https://www.hqmc-facilities.org/gis/index.cfm

http://proceedings.esri.com/library/userconf/ieug09/papers/geofidelis_frances_railey_program_brief.pdf

http://proceedings.esri.com/library/userconf/proc07/papers/papers/pap_1983.pdf

Technology Associates

www.taic.net

Authors:

Wendell Hoskins is the Systems Manager for the United States Marine Corps Western Regional Geospatial Data Center (GEOFIWEST) and is an employee of Technology Associates. He has a bachelor's degree in economics from California State University, San Marcos. He has been directly supporting Marine Corps and Navy enterprise GIS programs for the past five years.

Joe Simons is the GIS Data Services manager for GEOFIWEST and is an employee of Technology Associates. He has a bachelor's degree in geography from the California State Polytechnic University at Pomona. He has been directly supporting Marine Corps and Navy enterprise GIS programs for the past year. Prior to GEOFIWEST, he was an employee at ESRI for eight years as a product specialist and ArcSDE Administrator.