

Building a New Data Warehouse for King County GIS

ESRI User Conference

July 9, 2003

Debbie Bull

GIS Applications Developer
King County GIS Center
Seattle, WA

Lisa Castle

GIS Applications Developer
King County GIS Center
Seattle, WA

Abstract:

The GIS Center is in the process of separating the production and warehouse data systems for the County's multi-agency, distributed enterprise GIS system. Currently we utilize a single central GIS server for data production and warehousing. Layers are maintained as coverages and accessed in multiple formats. Decoupling the data warehouse and data production structures will result in two distinct, optimized systems. Phase 1 consists of implementing ArcSDE on Microsoft SQL Server 2000, and building the necessary tools to transparently replicate finalized data from the legacy production system to the new warehouse.

Currently:

- Production environment: Coverages, AML driven tools and processes
- Data Warehouse: Coverages and shapefiles
- DBMS: Oracle/UNIX

Ultimate Goal

- Production environment: SDE, VB, ArcObjects
- Data Warehouse: SDE
- DBMS: SQL Server/Windows

Transition stage

- Production environment: Coverages, VB processing
- Data Warehouse: SDE SQL Server
- DBMS: mixed SQL/windows and Oracle/Unix

THIS is what we're here to talk about

Why change?

- Bring down cost of ownership: hardware, software, support.
- More in-house support for the above, particularly the platform.

“Change is constant and painful. Embrace it, or at least deal with it.”

- Paul McCombs, KC DDES

Why the persistent/enduring transition stage?

OR: Why not just build the new and switch one day?

- Migration takes time, people need training.
- Cadastral base will remain in ArcInfo 7.x for at least 1 year (probably more).
- Stewards of other data need to move to ArcGIS sooner.

Transition Stage Structures

- Oracle: Production Database
- SQL Server: Data Warehouse

Oracle/Production

- Production Database
 - Everything but cadastral base layer
 - Not comprehensive, only the layers that are actively edited
- Control Database
 - Data registry: check in, check out, verification (existence, QA/QC)
 - Metadata: user documentation as well as database contents/control (queryable inventory)
 - Cadastral control: nightly update processes and event logging

SQL Server/Data Warehouse

- Contains every registered vector layer
 - Cadastral base replicated from librarian/coverage version
 - Everything else:
 - Preliminary load from coverages
 - Subsequent replication from finalized production database
- Contains scads of new raster data
- Control Structures for SQL Server
 - To be determined as needed
 - Stewardship tools and layer control
 - Etc.

Overview

	Now	Interim	Goal
Warehouse	Coverages, Shapefiles	SQL Server SDE, Shapefiles	SQL Server SDE
Production	Coverages	Cadastral: Coverages Others: Oracle SDE, Shapefiles, Coverages	SQL Server SDE, Personal GDB, (Shapefiles)

Non-Cadastral Data

- Ideally
 - One-time batch load of everything from coverage source
 - Coverage updates stop immediately
 - All edits done in production instance of SDE

...yeah, right.

Non-Cadastral Data: The Reality

- Migration is a L-O-N-G process.
- Edits will still come from
 - Coverage
 - Shapefile
 - Personal GDB
 - Enterprise GDB

Non-Cadastral Data: The Reality

- Offer special assistance to encourage editing in production SDE instance.
- Users aren't blindsided when coverage support is eventually dropped.
- A natural, organic progression (painless, of course).

Loading Non-Cadastral Data

Setting up SQL Server SDE: The Good News

- Lots of space
- Lower costs
- Immediate in-house support

Setting up SQL Server SDE: The Bad News

- No dedicated DBA.
- But wait, we have a GIS programmer/analyst! She'll try anything! SQL Server or Oracle? Bring it on, serve 'em up, no problem! In truth we have a lot of reasonably bright people, but we don't have the brain of one full DBA between us. I think before long we will have 3 to 5 people who could legitimately say "my other job is SQL Server DBA!"

Setting up SQL Server SDE: Potentially Bad News

- ArcIMS web apps:
 - Two million hits per month.
 - First use of the SQL Server database.
 - Service or performance disruptions are noticed immediately.

Interim implementation: First

- Replicate what's in Oracle
 - Because interim plan is to use Oracle instance for production, we'll need the ability to run nightly updates between SDE instances.
 - AO/VB Class SDE2SDE

Interim implementation: Next

- Load production coverages (Since we can't live without 'em).
 - Accommodate coverage postings.
 - Coverages in the current data warehouse are assured to have passed QA/QC checks.

Interim implementation: Later..

- Allow alternate source formats
 - Shapefile
 - Personal Geodatabase

What We Did

- Loaded all the SDE layers across and key tables to support ArcIMS apps.
- Rolled out new versions on ArcSDE 8.3 on SQL Server 2000.
- Tested like crazy.
- Rolled out new and improved ArcIMS 4.0.1.

What Happened

- It worked like a charm!
 - Better performance
 - Better stability for ArcIMS
 - Happy dances all around!
- Hey, let's load some of that raster data in there - ArcSDE is supposed to work like a champ for that stuff! 😊
- So we started the batch job to load raster data (373 TIFF files, many many gigs)

But the Next Day...

- It crashed and burned in the night.
 - full transaction log
 - full device
 - no connections allowed
 - database in “suspect” status.
- Rolled back to ArcIMS 3.x against ArcSDE 8.1 on Oracle/UNIX. ☹

Current status

- Database
 - SQL Server is stable
 - Loaded all non-cadastral data
 - Raster data is still loading
 - Conversion routines are almost in place
- ArcIMS
 - Running happily on SQL Server
- Users
 - Eager to use the new system (vector and raster)

Conclusions:

- A full-time DBA would be very nice. But there's NO MONEY!!
- This is no simple transition.
- Support for users comprised of 3 facets:
 - Encouragement
 - Assistance
 - Enforcement (tough love, rigid cut-off dates)

Conclusions:

- Impacts on the enterprise
 - Everybody is having to learn A LOT. Training GIS staff and users is an issue.
 - Potential change in business practices.
 - Facilitates migration of some applications to thin clients.
 - Overcoming user inertia is an issue.
- A DLL of ArcObjects conversion classes available for use in enterprise applications and local (agency) applications.
 - Updates to data will be immediate, not nightly or weekly

Time permitting

- Code review (Structure/algorithm)
- Mxd(VBA) to DLL(VB)

Authors

Debbie Bull

GIS Applications Developer
King County GIS Center
201 South Jackson Street
Mail Stop KSC-NR-0706
Seattle, WA 98104-3855
Phone: 206-263-4872
Fax: 206-263-3145
Debbie.Bull@metrokc.gov

Lisa Castle

GIS Applications Developer
King County GIS Center
201 South Jackson Street
Mail Stop KSC-NR-0706
Seattle, WA 98104-3855
Phone: 206-263-4862
Fax: 206-263-3145
Lisa.Castle@metrokc.gov