Esri Best Practices: Implementing an Enterprise Geodatabase

Rasu Muthurakku, Enterprise Solutions Architect / Consultant
Andrew Sakowicz, Enterprise Implementation Practice Lead
Agenda
Implementing an Enterprise Geodatabase

- Overview
- Key factors
- Geodatabase Design
- Architecture
- Build
- Workflow Design and Implementation
- Testing and Tuning
- Maintenance
- Monitor

Key Considerations, Best Practices, Recommendations and Lessons Learned!
Audience

Implementing an Enterprise Geodatabase

• Intermediate
• Advanced
Overview
What is an Enterprise Geodatabase (EGDB)?

Centralized Multiuser Geodatabase

- ArcSDE Enables the RDBMS* for GIS data management
  - Oracle
  - Microsoft SQL Server
  - PostgreSQL
  - IBM DB2, Informix
  - SAP Hana 2.0 SPS02

- Extremely Large, Continuous and Centralized GIS Database
- Many Simultaneous Users
- Long Transactions and Versioned Workflows
- SQL Types for Spatial in all Supported RDBMS
- High Performance for a Very Large Number of Users

*RDBMS – Relational Database Management System
Key Factors

Why?
Expensive Rework
A Stitch in Time Saves Nine!

- **Data Model Designs**
  - Change in XY Resolution / Tolerance
  - Multiple Projections
  - ObjectId as Primary Key for Relationship Classes
  - etc.

- **Data Conversion Specifications**
  - Extra Vertices
  - Insufficient QA/QC
  - Large Feature Extent
    - Possibly by Data Conversion / Testing
  - Loading Static Raster Data Into GDB
  - etc.

Avoid Pitfalls!
Expensive Rework
A Stitch in Time Saves Nine!

- **Hardware**
  - Slower Processor
  - Insufficient Capacity
  - Slower Storage and Network
  - etc.

- **Over-Utilized / Dynamic IT Infrastructure**

- **Software**
  - Version Selection including Patches
  - Utility Industry ArcGIS Desktop Version 10.2.1 (10.2.2) and 10.6.1
    - ArcGIS Server Version Can Be Higher
  - New Utility Network in ArcGIS Pro 2.1

*Avoid Pitfalls!*
Expensive Rework

A Stitch in Time Saves Nine!

• System Integration
  - Inefficient Interface Design
    - Version Difference for Each Version
  - Data Sharing with Other Systems Takes Hours, If not Days!
  - Synchronization Frequency
  - etc.

• Batch Process During Business Hour Slows Performance

• Critical Non-Functional Requirements
  • 24x7 Availability
  • Number of Users
  • Required GIS Environments
  • etc.

Avoid Pitfalls!
Degraded Performance and Scalability

Risks

• Workflow
  - Number of Outstanding Versions
  - Versioning Levels
  - Archiving
  - etc.

• Maintenance
  - Compact and Compress FGDB (File Geodatabase)
  - Total Delta Table Records
  - GDB Maintenance
  - etc.

Separate Myths / Emotions from Evidence Based Facts!
Overlooked Dependencies
Understand Intrigue Challenges!

• Key Inputs
  - User / Business Groups
  - Functional and Non-Functional Requirements
  - Application Designs
  - etc.

• System Architecture and Capacity

• Need a Dedicated / Assigned ArcSDE Administrator

Talk to the Right People to Get the Right Information!
Geodatabase Design
Geodatabase Design – Data Modeling

- **Conceptual Design**
  - Identify Business Requirements
  - Identify Thematic Layers
  - Identify Required Applications
  - Leverage Data Model Templates
  - Document

- **Logical Design**
  - Define Tabular Database Structure
  - Define Relationships
  - Determine Spatial Properties
  - Document

- **Physical Design**
  - Create and Implement Model Design
  - Generate Physical Schema in the RDBMS / FGDB
  - Testing and Validation
  - Document
Geodatabase Design – Tools

- X-Ray Add-In
- Geodatabase Diagrammer
- Sparx Systems' Enterprise Architect
- Geometric Network Configuration Manager
Leverage the Existing Data Models!
Geodatabase Design – Key Considerations

Poor Design = Slow Performance and Bugs

- Empty Feature Classes / Columns
- Missing / In-Correct Domains and Aliases for Fields
- Column / Domain Names and Field Lengths
  - >10 Characters in Field Names
  - Length of Text/NCLOB – 256 or 1,073,741,822
  - Choice of Field Type
  - Selection of Precision and Scale
  - Define Not Null Fields
Geodatabase Design – Key Considerations
Poor Design = Slow Performance and Bugs

- Change in XY Tolerance
  - Default = 10x Times of XY Resolution
  - Introduces Complexity (#NIM090335) for Geometric Network, etc.
  - Impacts Performance
Geodatabase Design – Best Practices

Poor Design = Slow Performance and Bugs

- No Attributed Relationship Classes for Empty Tables!
- Use Many to Many Relationship Classes Only When Necessary
- Don’t Use Objectid as Primary Key for Relationship Classes
  - Unexpected Replication Behavior
  - Additional Processing During Synchronization
Geodatabase Design – Best Practices
Prevention is Better Than Cure!

• Leverage the Existing ArcGIS Data Models
  - Drop Redundant Feature Datasets / Classes, Columns, etc.
  - Stand Alone Feature Classes are Fine!
  - Possibly Split the Feature Classes Per Scale Levels
  - Less Complex and Attributed Relationship Classes
  - Test, Refine and Tune the Data Models

• Integrate Related Feature Classes using Topology

• Deploy Necessary Information Models
  - Geometric Network Vs Utility Network *(New!)*

*Justify Every Single Geodatabase Element!*
Geodatabase Design – Best Practices

Prevention is Better Than Cure!

• Select Single Coordinate System
  - On the Fly Projection is Expensive
  - Geometric Network Editing Does Not Support “On the Fly Projection”

• Create Feature Datasets or Databases for Each LOB (Line of Businesses)
  - Depends on Size, Access, Usage and Maintenance
  - E.g. Landbase, Gas, Electric, Water GDBs, etc.
Geodatabase Design – Best Practices

Navigate Common Oversights!

• Review the Labeling Requirements Ahead of Time
  - For Multi-Field Complex Labeling
    - Combine them to a New Field and Auto Update
    - Convert Labels to Annotations

• Analyze Annotation requirements and choose proper Annotation reference scale

• Less Annotation Classes within Annotation Feature Class

• Add Attribute Indexes
  - Label Expression
  - Definition Queries
  - Application Design
Architecture
Architecture and Capacity Planning

- Define Architecture Vision / Foundation
  - Describe the System and its Relationships
- Define the Business Usage
  - Business Architecture
- Plan Suitable Software Solutions / Applications
  - Application Architecture
- Identify Data Requirements and Management
  - Data Architecture
- Select Proper Technology & Capacity for IT Infrastructure
  - Technology Architecture

Why is system architecture design important?

- Balanced system architecture design can:
  - Reduce costs
  - Improve user productivity

System architecture design provides framework for productive operations
Server Technology Selection

Why is it Important?

- Key to Optimal Scalability and Performance
- Save costs by Reducing Server Footprint
- RDBMS Needs Higher Processing Power for the GIS Version Queries
  - Version Query =
    \[(\text{Base Table} - \text{D# Table}) + (\text{A# Table} - \text{D# Table})\]
  - D# Table Records are Sorted Twice!

Look for SPEC Rate Per Core Value!
Server Technology Selection
Do not compensate poor maintenance with top processing power

• Processing Power is directly proportional to total Delta table records count

Intel® Xeon® E5-2637 v4 @ 3.40GHz, 8 Cores, 256GB RAM, ~59.38 SPEC Rate/Core Physical Machine Hosting a Happy GDB!
New / Additional Solution Architecture Options

Utility Network

- New Network to Manage Utility and Telecom Network Data
- Cross Platform Support
  - Any Device, Anytime, Anywhere!
- Services Based Architecture
- Updated Network Model
  - Connectivity Associations
  - Containment Associations
  - Structural Attachments
  - Multiple Terminals
  - Expanded Tracing Framework
  - Built in Support for network Diagrams

ArcGIS Enterprise
New / Additional Solution Architecture Options
SAP HANA In Memory Database Supports GDB!

• Only Enable a Geodatabase in SAP HANA 2.0 SPS02, Can Only Connect from:
  - ArcGIS Pro 2.1
  - ArcGIS Pro 2.2
  - ArcGIS Server 10.6
  - ArcGIS Server 10.6.1

• Branch Versioning

• Mobile Geodatabase Containing Offline Feature Data.
  - .geodatabase is based on a Portable SQLite Format
  - Primarily for Offline Workflows in ArcGIS Runtime Apps

<table>
<thead>
<tr>
<th>Esri SDK</th>
<th>.geodatabase</th>
<th>File geodatabase</th>
<th>Personal geodatabase</th>
<th>Enterprise geodatabase</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArcObjects</td>
<td>Not supported</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
</tr>
<tr>
<td>ArcGIS Runtime for Android</td>
<td>Supported</td>
<td>Not supported</td>
<td>Not supported</td>
<td>Supported only if using services</td>
</tr>
<tr>
<td>ArcGIS Runtime for iOS and OSX</td>
<td>Supported</td>
<td>Not supported</td>
<td>Not supported</td>
<td>Supported only if using services</td>
</tr>
<tr>
<td>ArcGIS Runtime for Java</td>
<td>Supported</td>
<td>Supported only if using local server</td>
<td>Not supported</td>
<td>Supported only if using services</td>
</tr>
<tr>
<td>ArcGIS Runtime for .NET</td>
<td>Supported</td>
<td>Supported only if using local server</td>
<td>Not supported</td>
<td>Supported only if using services</td>
</tr>
<tr>
<td>ArcGIS Runtime for Qt</td>
<td>Supported</td>
<td>Supported only if using C++ API and local server. Not supported with QML API</td>
<td>Not supported</td>
<td>Supported only if using services</td>
</tr>
</tbody>
</table>
Important Deprecation Plans
User Schema Geodatabases are Deprecated!

• **ArcGIS 10.6.1 is the Last Release to Support:**
  - User Schema Geodatabase in Oracle
  - SQL Server 2012
  - PostgreSQL 9.4

• **ArcGIS 10.5.1 Deprecate:**
  - Upgrading SQL Server Based GDBs with Multiple Databases
  - IBM's DB2 V9.7
  - ST_Raster Data Type for Oracle, SQL Server and PostgreSQL
  - ArcGIS Data Store Replaces EGDB for a Hosting Server Connected to Portal for ArcGIS

• **No Support to Create new Workgroup Geodatabases with ArcGIS Pro**
  - Still ArcMap and ArcCatalog 10.x Supports

*LONG RAW Storage is Deprecated a Long Time Ago!*
Define Functional and Non-Functional Requirements
Critical to Collect NFRs!

- High Availability
  - Use of Web Application by 24x7 Field / Emergency Crew
- Backups
  - RPO (Recovery Point Objective)
- Disaster Recovery
  - RTO (Recovery Time Objective)
- SLAs (Service Level Agreements)
Architecture and Design – Key Considerations & Challenges

• Number of Users, Operations and Transactions
  - Expected 130 Power Users (Editors) Vs 200+ Actual Users
  - Impact: >Concurrency Adds Pressure to Server Resources

• Number of Outstanding Versions
  - Estimated 500 Vs Actual 1250 Versions
  - Impact: Increased RDBMS Server’s CPU Time

• Estimated Total Delta Table Records (A# and D# Tables)
  - Expected 2 Millions Vs 8 Millions Actual
  - Impact: > CPU and Memory

GIS Benefits from Dedicated and Powerful Processors!
Virtualized Database Servers – Key Considerations

Speed is More Important than Utilization!

- Avoid Over-Commitment
- Ensure Less Number of vMotions
- Estimated Processing Requirements
  - User Load
  - Dedicated Operations and Transactions
- Application & Database Complexity

Complex and Busy Databases Need Dedicated Machines!
Symptoms of Over-Committed vCPUs

• vCPU Usage Vs MXD Display - Mxdperfstat Tool Results

![System Summary Graph]

- 3 Seconds Display
- Above 25% of CPU usage
- 8 Seconds Display
Additional Memory Requirements

Alert log:
Thu Dec 09 04:12:35 2017
WARNING: Heavy swapping observed on system in last 5 mins.

Alert log:
Thu Dec 09 15:47:39 2017
WARNING: Heavy swapping observed on system in last 5 mins.
Network Planning
Establish and Configure DNS Appropriately!

Trace Route: LA Workstation → Phoenix DNS
LA Database Server
System Architecture – Best Practices
Vendor Selection – RDBMS and Virtual Environments

• Select the Supported Vendors Based on:
  - Business Requirements
  - Existing Infrastructure Setup
  - Available Skill Set
  - Comparative Functional and Non-Functional Test Results
    - Develop a Test Plan Based on GIS Workflows
    - Prepare Isolated Environments Separately for Technology Comparison
  - Identify and Present Pros and Cons per the Requirements

• Esri Does Not Recommend any One Particular Vendor Solution
  - All Supported Vendors are Business Partners to Esri

*Esri Professional Service Can Help Establishing Benchmarks, Pilot, etc.!*
System Architecture – Best Practices
Virtual Environment - Hardware

- Deploy a Dedicated Virtual Environment for a Large GIS User Base
  - Provide a Decent Provisioning Ratio
  - Physical CPU vs vCPU
- Tune Virtual Environment Before Deployment
- Provide GPU (Graphics Processing Unit)
  - Video RAM >256MB Per Virtual Machine
- Fit Virtual Machine within one CPU NUMA* Node
  - # of vCPUs <= number of cores in the CPU socket
- Deploy Better Processors - Spec Rate Per Core 70+

*NUMA - Non-uniform memory access
System Architecture – Best Practices

Software

- Upgrade the Software to the Latest Version
- Update Service Pack Levels and Patches
  - Apply the latest Utility and Telecom Update (UTU) Patch 8 for 10.2.1
  - ArcGIS 10.2.1 for Oracle 12.2 Patch Enables GDB on Oracle 12.2.0.x.
- ArcGIS (all versions) on Windows 2008 and Windows 7 Needs:
  - ArcGIS Server Geoprocessing Service Startup Patch
  - ArcGIS (Desktop, Engine) Background Geoprocessing 64-bit Arcpy Exit and Shutdown Patch

<table>
<thead>
<tr>
<th>S.No</th>
<th>Display Scale</th>
<th>Layer Name</th>
<th>Before tuning - Display in Seconds</th>
<th>After Tuning - Display in Seconds</th>
<th>Performance Improvement in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,000,000</td>
<td>TRANS LINE</td>
<td>0.94</td>
<td>0.37</td>
<td>154.05%</td>
</tr>
<tr>
<td>2</td>
<td>500,000</td>
<td>STATION</td>
<td>0.26</td>
<td>0.13</td>
<td>100.00%</td>
</tr>
<tr>
<td>3</td>
<td>50,000</td>
<td>TRANS LINE</td>
<td>0.8</td>
<td>0.18</td>
<td>344.44%</td>
</tr>
<tr>
<td>4</td>
<td>50,000</td>
<td>STRUCTURE</td>
<td>44.62</td>
<td>0.72</td>
<td>6097.22%</td>
</tr>
<tr>
<td>5</td>
<td>25,000</td>
<td>STRUCTURE</td>
<td>46.38</td>
<td>0.49</td>
<td>9365.31%</td>
</tr>
<tr>
<td>6</td>
<td>25,000</td>
<td>TRANS LINE</td>
<td>0.83</td>
<td>0.16</td>
<td>418.75%</td>
</tr>
<tr>
<td>7</td>
<td>1,000</td>
<td>STRUCTURE</td>
<td>42.26</td>
<td>0.28</td>
<td>14992.86%</td>
</tr>
<tr>
<td>8</td>
<td>1,000</td>
<td>TRANS LINE</td>
<td>0.76</td>
<td>0.31</td>
<td>145.16%</td>
</tr>
<tr>
<td>9</td>
<td>500</td>
<td>STRUCTURE</td>
<td>41.66</td>
<td>0.27</td>
<td>15329.63%</td>
</tr>
<tr>
<td>10</td>
<td>500</td>
<td>TRANS LINE</td>
<td>0.74</td>
<td>0.21</td>
<td>252.38%</td>
</tr>
</tbody>
</table>

*DFQ (Definition Query) = ObjectId is not null

Impact of an Older RDBMS Version Bug
System Architecture – Best Practices
Network Infrastructure

• Request for Higher Network Bandwidth (~ 1Gbps) and Less Latency (<1 MS)
  - Bandwidth Reduces Number of Network Packets
  - ArcGIS Desktop is Sensitive to 1 – 2 MS Latency!
  - Plan for ~ 1.5 Mbps per Concurrent GIS User

• Enable Jumbo Frames Between Servers
  - Ensure All Switches Support – Otherwise Don’t Enable it!

• Validate Network Path Between GIS User and Server Locations
  - Correct DNS and Routing

• Upgrade Lower Bandwidth or Move GIS User Locations
System Architecture – Best Practices

Security Configuration

- Configure Firewalls / IPS* Outside of GIS Perimeter
- Avoid Storing PII* / Sensitive Data in Public GDB!
- Separate Internal Vs External Databases
  - Keep External Databases and Clients in DMZ Together!
- Choose Appropriate Single Sign-On Solutions
  - E.g.: ops$ Single Sign-On is not a more Secure Option and already deprecated in Oracle 11g

<table>
<thead>
<tr>
<th>Service Name</th>
<th>REST Export in Seconds - RDBMS Behind Firewall</th>
<th>REST Export in Seconds - Local File Geodatabase</th>
<th>Performance Gain in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>USE_BASE</td>
<td>12.018</td>
<td>0.287</td>
<td>4087.46%</td>
</tr>
<tr>
<td>ArcServer_Assets</td>
<td>5.266</td>
<td>0.424</td>
<td>1141.98%</td>
</tr>
</tbody>
</table>

*PII - Personally Identifiable Information
*IPS - Intrusion Prevention System
System Architecture – Best Practices

Disks

- Use SSDs (Solid-State Drives)
- Plan for ~5000 IOPS (Input/output Operations Per Second)
- Avoid LUNs* > 2TB Size
  - Minimum of 4 LUNs that are Identical in Size
- Avoid noac Mount Option!
- Distributed File System (DFS) is not supported

*LUN – Logical Unit Number
System Architecture – Best Practices
ArcSDE Configuration

- Configure ArcSDE DBTUNE Settings
  - Use Default Geometry Storage - St_Geometry / Geometry
  - Storage Locations
  - etc.

- ArcSDE Initialization Parameters
  - Defaults are Good!
System Architecture – Best Practices
System Integration

• Use Private Versions for Larger Versioning Environment
  - Keep the Total Number Less!
  - Sync Frequently
  - Monitor

• Complete the Batch Processes within the Identified Time Frame
  - Improve Hardware and/or Software Design
  - Additional CPUs
  - Multi Threads
  - Schema Cache

Private Version Impact
Build
Build Geodatabase

- Create Physical Geodatabase
  - Structure the Implementation to Pilot → Phase I → Phase II → Phase III, etc.
  - Enough Gap Between Each Phases to Accommodate the Learned Lessons
  - Separate Data Owner from SDE / DBO User

- Develop Data Conversion/Update Specification Document Aligned with Data Model

- Team Review and Demonstration
Build Geodatabase – Key Considerations
Data is the Backbone of GIS!

- **Aim for 100% Data Accuracy from Data Conversion Effort**
- **Either Populate or Drop Empty Fields**
- **Minimize Data Model / Schema Changes**
  - More Complexity in a Versioned Geodatabase
- **Consolidate GDBs**
  - Avoid Creating GDB Per Geographic Locations / Regions
- **Extra Vertices have Performance Impact**

<table>
<thead>
<tr>
<th>S.No</th>
<th>Display Scale</th>
<th>Layer Name</th>
<th># of Displayed Features</th>
<th>Number of Vertices - Before Simplify</th>
<th>Number of Vertices - After Simplify</th>
<th>Before Simplification Display in Seconds</th>
<th>After Simplification Display in Seconds</th>
<th>Performance Improvement in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50,000</td>
<td>Street_1_inch</td>
<td>35,093</td>
<td>105,695</td>
<td>101,060</td>
<td>2.36</td>
<td>0.5</td>
<td>372.00%</td>
</tr>
<tr>
<td>2</td>
<td>50,000</td>
<td>PARCEL_1_Inch</td>
<td>7,922</td>
<td>645,766</td>
<td>188,212</td>
<td>0.37</td>
<td>0.31</td>
<td>19.35%</td>
</tr>
<tr>
<td>3</td>
<td>25,000</td>
<td>Street_1_inch</td>
<td>11,192</td>
<td>31,112</td>
<td>29,620</td>
<td>0.69</td>
<td>0.2</td>
<td>245.00%</td>
</tr>
<tr>
<td>4</td>
<td>25,000</td>
<td>PARCEL_1_Inch</td>
<td>2,687</td>
<td>168,011</td>
<td>48,540</td>
<td>0.16</td>
<td>0.14</td>
<td>14.29%</td>
</tr>
<tr>
<td>5</td>
<td>20,000</td>
<td>Street_1_inch</td>
<td>7,590</td>
<td>20,494</td>
<td>19,574</td>
<td>1.59</td>
<td>0.16</td>
<td>893.75%</td>
</tr>
</tbody>
</table>
Key Data Conversion Considerations

- Extra Vertices Introduced by:
  - Conversion process involving CAD systems
  - Geometric Network Creation with Snapping ON Option
  - Conversion/Update Methods

- Develop Adequate QA/QC Methods and Procedures

- Additional Data Reviewer Checks
  - Duplicate/Invalid Geometries
  - Orphan Related Records
  - Connectivity Check
  - Etc.

<table>
<thead>
<tr>
<th>Feature Class/Layer</th>
<th>Total Number of Features</th>
<th>Total Number of Vertices with Geometric Network Snapping</th>
<th>Total Number of Vertices without Geometric Network Snapping</th>
<th>Increase in number of vertices %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary Overhead Conductor</td>
<td>3,712</td>
<td>12,953</td>
<td>7,841</td>
<td>65.20%</td>
</tr>
</tbody>
</table>
Build Geodatabase – Best Practices

- Avoid Extra vertices
  - Use Curve Tools that Insert Less Vertices
- Turn off Snapping During Geometric Network (GN) Creation
  - ArcGIS 10.0 onwards a Vertex is added at every Intersection
  - Cannot Create GN with >15M Edges until ArcGIS 9.3.1
- Remove Additional Vertices
  - Generalize / Simplify
  - ArcObjects
Build Geodatabase – Best Practices

• Keep the Data Clean and Simple
  - Without Any Topological Errors

• No Coincident Complex Edge Features in Geometric Network
  - Most Common Reason for Geometric Network Corruption

• Unversion Read Only Feature Classes / Tables

• Use Mosaic Datasets Instead of Loading Raster Data Into EGDB
  - Static Raster Data Does Not Need to Participate in Daily RDBMS Backup
Build Geodatabase – Geometric Network Lessons Learned!

- Steps to Create Geometric Network, if Snapping Option need to be “ON”
  - Drop the Geometric Network
  - Re-create the Geometric Network with only Required Feature Classes.
  - Turn on the snapping during the Geometric Network Building Process.
  - Drop the Geometric Network.
  - Remove the Intersection Vertices introduced by the above snapping.
    - Use ArcObjects / Python
  - Build the Geometric Network without snapping.
  - Verify the Geometric Network BuildERR table.
  - Ensure to correct all the errors/invalid features mentioned in the BuildErr table.
Workflow Design and Implementation
Requirements and Workflows
Drives the Selection and Number of GDBs!

- Number of Users and Types of Users
- Workflows
  - Multi User Editing - **Enterprise / Workgroup GDBs**
  - Single User Editing – **FGDB**
  - Replication - **EGDB → FGDB / EGDB**
  - Read Only / Publication – **FGDB / EGDB**
  - etc.

- Generally more than one Geodatabase is required!

### Enterprise Geodatabase Vs File Geodatabase

<table>
<thead>
<tr>
<th>Key Characteristics</th>
<th>Enterprise Geodatabase</th>
<th>File Geodatabase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>A collection of various types of GIS datasets held as tables in a relational database.</td>
<td>A collection of various types of GIS datasets held in a file system folder.</td>
</tr>
<tr>
<td>Number of users</td>
<td>Multiuser: many readers and many writers</td>
<td>Single editor and can support multiple readers.</td>
</tr>
<tr>
<td>Storage format</td>
<td>Oracle</td>
<td>Each dataset is a separate file on disk. All the datasets that belong to one geodatabase are contained in a single folder.</td>
</tr>
<tr>
<td></td>
<td>Microsoft SQL Server</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IBM DB2, Informix</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PostgreSQL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SAP HANA</td>
<td></td>
</tr>
<tr>
<td>Size limits</td>
<td>Size is controlled by the DBMS</td>
<td>By default, each dataset can grow to one TB. The 1 TB limit can be raised to 4 or 256 TB for extremely large image datasets. Each feature class can scale up to hundreds of millions of vector features per dataset.</td>
</tr>
<tr>
<td>Versioning support</td>
<td>Fully supported across all DBMSs</td>
<td>Does not support versioning workflows</td>
</tr>
<tr>
<td>Platforms</td>
<td>The DBMS are supported on multiple operating systems. Consult the system requirements for the full list.</td>
<td>Cross-platform.</td>
</tr>
<tr>
<td>Security and permissions</td>
<td>Managed by the DBMS.</td>
<td>Managed by the operating system.</td>
</tr>
</tbody>
</table>

Avoid Limitations!
Geodatabase Workflows - QA / QC

- Design and Implement QA / QC Workflows
  - Data Requirements for Software Functions
  - Accurate Data for Business
  - Maintain Data Integrity

Capture, Load and Maintain Data Accurately!
Data Integrity and Validation Strategies

• Stage 1: Don’t Allow Start Editing
  - Read Only Users
  - Without Landbase Layers

• Stage 2: No Inserts Without Pre-Requisite Checks
  - Out Side of Editing Areas (Pacific Ocean!)
  - Street Light without Poles
  - Equipment Without Structures
  - Required Attribute Values (WO Number, Number of Phases, etc.) in Attribute Columns
  - Etc.

Capture, Load and Maintain Data Accurately!
Data Integrity and Validation Strategies

- **Stage 3: Reconcile/Save Edits Only After Rules Validation**
  - Domain Checks
  - Connectivity Rules, etc.

- **Stage 4: Allow to Post Data with Warnings**
  - Run Batch Processes to Perform Additional Checks

- **Tools**
  - Domains, Subtypes, Topology, etc.
  - Attribute Assistant Add-In
  - ArcGIS Data Reviewer
  - ArcGIS Workflow Manager
  - Business Partner Products
  - Customization

*Capture, Load and Maintain Data Accurately!*
Geodatabase Multiuser Workflows
Which Versioning Model to Choose?

• Traditional Versioning Vs Branch Versioning
Geodatabase Multiuser Workflows

• Versioning Structure
  - Recreate the Version after each Post for 3 Version levels
  - **Problem: Unexpected conflict observed during reconcile**
  - Move Edits to Base For Simple Feature Classes
    - Ability to Easily Share the data with third-party applications

• Estimated Edit Volumes, Version Durations

• Conflict Resolution Mechanisms
Geodatabase Workflows – Key Considerations

- Conflict for Split operation could introduce duplicate geometry
- Plan Bulk Loading / Mass Update
- For Regular Data Load, Consider Truncate Vs Delete
- Execute Batch Processes During Non-Business Hours
- Geodatabase Replication - One-Way Vs Two-Way
- Archiving
- Editor Tracking

- Conflict on Objectid 1
- Objectid 2 and 3 Overlap
Geodatabase Workflows – Best Practices

- Choose the Versioning Workflows Appropriately
  - Key to Performance and Scalability
- Run QA / QC Tools Regularly
  - Avoid Duplicate / Invalid Geometries, etc.
- Leverage Geodatabase Replication
  - Much Improved Stability and Reliability Now!
  - Use One-Way Replication Options
    - Parent to Child
    - Child to Parent
    - Create Read-Only / Publication Geodatabase
  - Two One-Way Replicas For Two Separate Datasets/FCs
Geodatabase Workflows – Best Practices

Archiving

• Do not enable Archiving when 100% data update / modification is expected!
• Generally Mistaken with the below Functionalities / Purposes:
  - GDB Editor Tracking
  - RDBMS Backup and Retention
  - Security
  - Data Integrity
Geodatabase Workflows – Best Practices

• Provide only the Required Privileges to Users
  - Access to Large Number of Tables Slows Connection Performance

• Arrange Workflow Training for Users
  - Conduct Tips and Tricks Session
Testing and Tuning

Important Step Before Going Live!
Testing

- Test Application Workflows
  - Functionality
  - Flexibility and Consistency
- Conduct single user execution test and measure performance
  - Provides Key Indicators towards scalability
Tune - Operating System

Operating System

- Adjust and Configure
  - Kernel Parameters
  - Settings Specific to RDBMS and Network Capacity
- Enable Large / Huge Memory Pages for Geodatabases
- Update Patches

<table>
<thead>
<tr>
<th>Operation System's Internal Memory Allocation to RDBMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS Default Memory 4KB/ Page</td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td>RDBMS 32GB / 4KB = 8,388,608 Pages</td>
</tr>
</tbody>
</table>

- **HugePages_Total:** 16384
- **HugePages_Free:** 5166
- **HugePages_Rsvd:** 1585
- **HugePages_Surp:** 0
- **Hugepagesize:** 2048 kB
- **DirectMap4k:** 67117056 kB
- **DirectMap2M:** 0 kB
Tune – RDBMS
RDBMS Configuration and Tuning

- Tune RDBMS
  - Memory Allocation from Hardware
  - Other Initialization Parameters
  - Log File Settings
  - etc.

- Implement the Best Practices
Maintenance
Geodatabase Maintenance - Strategies

- Identify a Maintenance Window and Tasks
- Categorize
  - Nightly, Weekly, Monthly and Yearly.
- Classify Manual and Automated Batch Processes
  - Design Scalable Batch Processes
- Assign SDE/GIS Administrator Role
- Monitor

**Geodatabase Needs Maintenance – Plan One!**
Geodatabase Maintenance – Key Considerations

- Reconcile, Post and Compress Schedule
- Underlying RDBMS requires maintenance other than backup!
  - Rebuild Index
  - Update Statistics
  - Logs
- Execute Automated Processes Only Within Maintenance Window
- Run Repair Version Tables and Metadata (Previously: SDEGDBREPAIR) Every ~3 Months
  - Fix Any Inconsistencies
  - Schedule the Execution Around Weekends
Geodatabase Maintenance – Best Practices

- Identify, Reconcile and Post Top 5 Blocking Versions Every Day
  - Blocking Versions Cause Inefficient Compress
  - Increase in Delta Table Records Beyond the Hardware Support Level
  - Maintain the Lineage Length <100

SQL Server DBs ArcSDE A Table Rows -- ADDS TABLE RECORD COUNT (COUNTER) -- Principle: track record count

Number of versions: 664
Number of versions blocking DEFAULT: 625
Top 5 blocking versions...
  ANDREW.SN_164022
  ANDREW.SN_162751
  RASU.SN_163090
  RASU.SN_163139
  RASU.SN_164468
Number of states: 4333
Number of state lineages: 86293
DEFAULT versions lineage length: 426
Last compress: JUL-09-2017
• Remove Geoprocessing (GP) History
  - *How To: Automate the process of deleting geoprocessing history*
  - [http://support.esri.com/technical-article/000011751](http://support.esri.com/technical-article/000011751)

• Disable (GP) History for Scripts

```python
import arcpy
arcpy.SetLogHistory(False)
```

<table>
<thead>
<tr>
<th>Python Tasks</th>
<th>Before Deleting GP History in Seconds</th>
<th>After Deleting GP History in Seconds</th>
<th>Performance Improvement in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Version</td>
<td>400.74</td>
<td>16.3</td>
<td>2358.53%</td>
</tr>
<tr>
<td>Delete Version</td>
<td>571.23</td>
<td>14.17</td>
<td>3931.26%</td>
</tr>
</tbody>
</table>
Geodatabase Maintenance – Best Practices

- Manage Auditing / History Tables
  - Reduce the Database
    - Backup Size
    - Storage
    - Time

<table>
<thead>
<tr>
<th>OWNER</th>
<th>TABLE_NAME</th>
<th>NUM_ROWS</th>
<th>BLOCKS</th>
<th>AVG_ROW_LEN</th>
<th>TO_CHAR(LAST_ANALYZED, 'MON/DD/YYYYHH24:MI:SS')</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCFM</td>
<td>EDITEDFEATURESTRACKING</td>
<td>12330909</td>
<td>122954</td>
<td>140</td>
<td>APR/23/16 23:44:17</td>
</tr>
<tr>
<td>ARCFM</td>
<td>EDITEDGRIDS</td>
<td>12964827</td>
<td>32969</td>
<td>32</td>
<td>APR/23/16 23:47:55</td>
</tr>
<tr>
<td>SDE</td>
<td>ARCSDEUSERLOG</td>
<td>56558633</td>
<td>259246</td>
<td>62</td>
<td>APR/24/16 12:28:25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDE</td>
<td>GDBM_RECONCILE_HISTORY</td>
<td>4759429</td>
<td>50484</td>
<td>72</td>
<td>AUG/12/17 12:40:34</td>
</tr>
<tr>
<td>ARCFM</td>
<td>USAGEINFORMATION</td>
<td>6377236</td>
<td>221057</td>
<td>39</td>
<td>AUG/10/17 15:59:14</td>
</tr>
</tbody>
</table>
Geodatabase Maintenance – Best Practices

- Every 3 - 6 Months:
  - Fix the Feature Class Extent First
  - Followed by Rebuilding the Spatial Index
• Fix the Non-Empty Feature Classes with No Spatial Index

```sql
EXEC DSMS_STATS.GATHER_SCHEMA_STATS ('ARCFM', estimate_percent=>100,
  DEGREE=>7, CASCADE=>TRUE, NO_INVALIDATE=>false);
SELECT table_name, GSIZE1, GSIZE2, GSIZE3 FROM sde.layers
WHERE gsize1=0 and gsize2=0 and gsize3=0 and table_name in
  (SELECT table_name FROM all_tables WHERE num_rows > 0);
```
Geodatabase Maintenance – Best Practices

• A Sample Daily Maintenance (Batch Process) for Multi User Geodatabase:
  1. Backup the Database
  3. Delete the Orphan / Unnecessary versions.
  4. Drop the orphan keyset tables
  5. Reconcile and Post All/Eligible Versions – Through out the Day
  6. Only Reconcile All versions (>100 versions - Parallel Reconcile)
  7. Update Database Statistics - Optional

Continue…..
Geodatabase Maintenance – Best Practices

- Continue…..
  8. Pause the SDE Connections
  9. Kill the Existing or Orphaned User Connections
  10. Truncate Dynamic tables
      A. state_locks; table_locks; object_locks; layer_locks; process_information;
         <user>.SDE_LOGFILE_DATA;
  11. Start the Compress Process
  12. Un-pause the SDE Connection
  13. Rebuild Indexes in RDBMS for all the Schema Owners and SDE
  14. Update RDBMS statistics for all Schema Users and SDE.
Geodatabase Maintenance – Geometric Network

- Every 3 – 6 Months Run Esri’s Verify And Repair Geometric Network Connectivity Tool
- Only GIS Administrator Should Follow the below steps:
  - Create a New Version under SDE.Default.
  - Create a SDE Connection Document with the New Version.
  - Run the Verify and Repair tool
    - With “Repair network after verify completes” option
  - Reconcile, Post and Delete the newly Created version.
  - Compress the Database.

Caution: Individual Users Should not run it. Generates larger number of delta table records and if someone accidently run this tool under SDE.Default version directly, it can produce more conflicts.
Monitor
ArcGIS Monitor

- Add RDBMS Queries
- Monitor Key Performance Indicators
- Keep 15 Minutes Sampling Interval
Tools for Implementation Assistance

ArcGIS Monitor

• ArcGIS Monitor
  - ArcGIS Monitor is a tool for monitoring and analyzing your enterprise GIS system

• MXDPerfstat
  - An ArcGIS Engine command line tool to diagnose typical mxd performance problems
  - https://www.arcgis.com/home/item.html?id=a269d03aa1c840638680e2902dadecac

• System Designer
  - A comprehensive tool for designing and capacity planning of GIS solutions.
  - https://www.arcgis.com/home/item.html?id=8ff490eef2794f428bde25b561226bda

• System Log Parser
  - A reporting tool specifically designed for analyzing ArcGIS server and service logs
  - http://www.arcgis.com/home/item.html?id=a29649a3d87d4cae84374e5d711dc3aa
Questions and Answers

Contact Info:
Rasu Muthurakku
rasu@esri.com

Andrew Sakowicz
asakowicz@esri.com
# See Us Here

<table>
<thead>
<tr>
<th>WORKSHOP</th>
<th>LOCATION</th>
<th>TIME FRAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geodatabase: An Introduction</td>
<td>SDCC - Room 05 A</td>
<td>Wednesday, July 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2:30 pm - 3:30 pm</td>
</tr>
<tr>
<td>Managing Your Distributed Geodatabase</td>
<td>SDCC - Room 14 A</td>
<td>Wednesday, July 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4:00 pm - 5:00 pm</td>
</tr>
<tr>
<td>Esri Best Practices: Implementing an Enterprise Geodatabase</td>
<td>SDCC - Room 05 B</td>
<td>Thursday 7/12/2018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10:00 AM - 11:00 AM</td>
</tr>
</tbody>
</table>
Please Take Our Survey on the App

Download the Esri Events app and find your event

Select the session you attended

Scroll down to find the feedback section

Complete answers and select “Submit”
Thanks!

Implementing an Enterprise Geodatabase