Esri Best Practices: Implementing an Enterprise Geodatabase

Rasu Muthurakku, Esri Professional Services
Andrew Sakowicz, Esri Professional Services
Agenda
Implementing an Enterprise Geodatabase

• Overview
• Implementation Approach
• Architecture
• Geodatabase Design
• Build
• Workflow Design
• Testing and Tuning
• Maintenance
• Monitor
• Question & Answers

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
- 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
Implementation Approach
Implementation Approach

Project Management
- Define Goals/Objectives
- Schedule Milestones
- Mobilize Resources
- Roles and Responsibilities
- Key Metrics for Success
- Agile/Adaptive Plans
- Evaluate Success

System architecture Design and Capacity planning
Geodatabase and Workflow design
Implement pilot / Proto types and Obtain feedbacks
Test, validate and Tune the system
Build Geodatabase
Test, validate and Adjust the system Capacity
Develop and Implement Training plans
Monitor and Maintain including Data integrity
Test, validate and Tune the system

Entity Name
attribute name PK
attribute name
attribute name

Table
Table
Table
Architecture
Efficient Implementation

• System Architecture and Capacity
  - Separation of read and edit data sources
  - Faster Processors
  - Enough system capacity
  - Better storage and network
  - etc.

• Software
  - Version selection including the patches
  - Utility industry ArcGIS Desktop version 10.2.1 (10.2.2) and 10.6.1
  - Utility Network in ArcGIS Pro >2.1

Under-utilized / stable IT Infrastructure helps
Capacity Planning

• Define architecture vision / foundation
  - Describe the System and its relationships

• Business Architecture
  - Define the business usage

• Application Architecture
  - Plan suitable software solutions / applications

• Data Architecture
  - Identify data requirements and management

• Technology Architecture
  - Select proper technology & capacity for IT infrastructure

System architecture design reduces cost and improves productivity
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
Business Architecture

- **Business users**
  - Total and concurrent
  - Location
  - Internal and/or external

- **Business workflows and user roles**
  - Data editors / Managers / Administrators
  - Data readers
  - Power users
  - Web / Mobile users
  - etc.

Data Centers and User Sites

<table>
<thead>
<tr>
<th>Scale</th>
<th>Feature Class/Layer</th>
<th>Display in Seconds at Data Center 2 Server</th>
<th>Display in Seconds at Data Center 1 Server</th>
<th>No of Features</th>
<th>Performance Difference in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>25,000</td>
<td>Street Centerline</td>
<td>0.22</td>
<td>0.16</td>
<td>3,440</td>
<td>37.50%</td>
</tr>
<tr>
<td>20,000</td>
<td>ROWEdge</td>
<td>0.61</td>
<td>0.23</td>
<td>4,017</td>
<td>165.22%</td>
</tr>
<tr>
<td>20,000</td>
<td>Overhead Primary Conductor</td>
<td>0.34</td>
<td>0.28</td>
<td>3,003</td>
<td>21.43%</td>
</tr>
<tr>
<td>15,000</td>
<td>Street Centerline</td>
<td>0.23</td>
<td>0.09</td>
<td>2,745</td>
<td>155.56%</td>
</tr>
<tr>
<td>15,000</td>
<td>ROWEdge</td>
<td>0.25</td>
<td>0.14</td>
<td>2,461</td>
<td>78.57%</td>
</tr>
<tr>
<td>15,000</td>
<td>Overhead Primary Conductor</td>
<td>0.57</td>
<td>0.19</td>
<td>1,944</td>
<td>200.00%</td>
</tr>
</tbody>
</table>
Define Functional and Non-Functional Requirements

Critical to collect NFRs

• High Availability (HA)
  - Web application for 24x7 Field / Emergency crew
  - Use clusters with one database for editing
• Backups
• Disaster Recovery (DR)
• SLAs (Service Level Agreements)
  - RPO (Recovery Point Objective)
  - RTO (Recovery Time Objective)

Relevant and justifiable requirements to keep the cost and complexity checked
Application Architecture

- Desktop and web solutions
- Target software versions
  - Operating System
  - Virtualization
  - Partner solutions

- Update service pack levels and patches
  - ArcGIS Desktop TLS Patch
  - ArcGIS (Desktop, Engine) Text Performance Patch
  - Utility and Telecom Update (UTU) Patch 9 for 10.2.1
  - etc.

Check system requirements and GDB client compatibility
Technology Architecture
Quality vs Quantity

• IT infrastructure
  - Availability and policies
  - Limitations, preferences and constraints

• Processor selection
  - Key to optimal scalability and performance
  - Save costs by reducing server footprint
  - Check spec rate per core
    - http://spec.org/cgi-bin/osgresults?conf=cpu2017
    - E.g. Intel Xeon Gold 6244 16 cores 3.6GHz = 83.125 SPEC rate per core (April 2019 results)

Select higher SPEC rate per core for speed Vs processor density
Server Technology Selection
Do not compensate poor maintenance with top processing power

- Processing power is directly proportional to total Delta table records count

---ArcSDE A Table Rows

| SQL> prompt 2. STATES
| SQL> 2. STATES
| SQL> select count(*) from sde.states;
| Tue Sep 26
| COUNT(*)
| 113,904

---ArcSDE D Table Rows

| SQL> prompt 3. STATE_LINEAGES
| SQL> 3. STATE_LINEAGES
| SQL> select count(*) from sde.state_lineages;
| Tue Sep 26
| COUNT(*)
| 106,274,130

GIS Benefits from dedicated and powerful processors

Intel® Xeon® E5-2637 v4 @ 3.40GHz, 8 Cores, 256GB RAM, ~59.38 SPEC rate/core
Physical machine hosting a happy GDB!
Virtualized Database Servers – Key Considerations

Speed is more important than utilization

• Avoid over-commitment
• Ensure less number of vMotions
• Estimated capacity requirements
  - User Load
  - Dedicated operations and transactions
• Application & database complexity
Symptoms of Over-Committed vCPUs

- vCPU usage Vs MXD display seconds

![Graph showing CPU usage over time with 3 and 8 seconds display]

- Above 25% of CPU usage
- ~100% Virtual cluster utilization!
User load

Estimated vs actual usage

- Number of users, operations and transactions
  - Expected 130 power users (editors) Vs 200+ Actual Users

- Number of outstanding Versions
  - Estimated 600 Vs actual 1250 Versions

- Estimated total delta table records (A# and D# Tables)
  - Designed for 2 Millions Vs 8 Millions records actual
Virtual Environment

- Dedicate the Virtual environment for a large GIS user base
  - Provide a decent provisioning ratio
- Provide GPU* for GDB clients
  - Video RAM >256MB Per Virtual Machine
- Fit Virtual Machine within one CPU NUMA* node
  - # of vCPUs <= number of cores in the CPU socket
- Choose faster CPUs - Spec rate per Core 80+

* GPU – Graphics Processing Unit
* NUMA - Non-uniform memory access

Tune Virtual environment before deployment
Network Planning
Establish and configure DNS appropriately!

Trace Route: LA Workstation → Phoenix DNS → LA Database Server

Network Latency’s impact on performance

<table>
<thead>
<tr>
<th>S.No</th>
<th>Task</th>
<th>Performance when &lt;1 ms latency (in Seconds)</th>
<th>Performance during latency fluctuation (in Seconds)</th>
<th>Performance difference in Seconds</th>
<th>Performance Difference in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Open ArcMap</td>
<td>23</td>
<td>90</td>
<td>-67</td>
<td>-291.30%</td>
</tr>
</tbody>
</table>
Network Infrastructure

- Establish higher Network bandwidth (>1Gbps) and less latency (<1 MS)
  - ArcGIS Desktop is sensitive to 1 – 2 MS latency
  - Plan for ~ 1.5 Mbps per concurrent GIS User
- Enable Jumbo Frames between servers
  - All Switches must support – Otherwise don’t enable it!
- Validate Network path between GIS user and server locations
- Upgrade lower bandwidth or move GIS user locations

<table>
<thead>
<tr>
<th>Scale</th>
<th>Feature Class Name</th>
<th>Response Time in Seconds when the workstation was connected to 100Mbps Phone port</th>
<th>Response Time in Seconds when the workstation was connected to 1Gbps LAN port</th>
<th>Performance improvement in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>50,000</td>
<td>Centerline</td>
<td>8.84</td>
<td>4.18</td>
<td>111.48%</td>
</tr>
<tr>
<td>50,000</td>
<td>Control Boundary (WMX)</td>
<td>0.17</td>
<td>0.14</td>
<td>21.43%</td>
</tr>
<tr>
<td>2,500</td>
<td>GasMeter</td>
<td>0.78</td>
<td>0.55</td>
<td>41.82%</td>
</tr>
<tr>
<td>1,000</td>
<td>GasMeterQualAnno</td>
<td>0.21</td>
<td>0.12</td>
<td>75.00%</td>
</tr>
</tbody>
</table>

Higher bandwidth between EGDB and the clients provide better performance
Storage

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

SSD delivers better performance for EGDB
ArcSDE Configuration

- Configure ArcSDE DBTUNE Settings
  - Use Default Geometry Storage - St_Geometry / Geometry
  - Storage locations
  - etc.
- ArcSDE initialization parameters
  - Limit the connections
  - Defaults are good!
Geodatabase Design
Geodatabase Design – Data Modeling

- **Conceptual Design**
  - Identify business requirements
  - Identify thematic layers
  - Identify required applications
  - 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
  - Test and validate
  - Document

*Leverage the existing data models*
Geodatabase Design – Tools

- **Tools**
  - X-Ray Add-In
  - Geodatabase Diagrammer
  - Sparx Systems’ Enterprise Architect
  - Geometric Network Configuration Manager
Geodatabase Design – Key Considerations
Prevention is better than cure!

- Consider single coordinate system
  - On the Fly Projection is expensive
  - Geometric Network editing does not support “On the Fly Projection”

- Column / Domain names and Field lengths
  - Avoid >10 characters in Field names
  - Apply required text length e.g. Text - 256 Vs NCLOB – 1,073,741,822
  - Select appropriate Field type
  - Apply only the required Precision and Scale
  - Define Not Null Fields

Justify every single Geodatabase element
Geodatabase Design – Key Considerations
Poor Design = Rework, slow performance and bugs

• Avoid XY Tolerance modification
  - Default = 10x Times of XY Resolution
  - Introduces complexity (#NIM090335) for Geometric Network, etc.
  - Impacts performance

Avoid expensive rework!
Geodatabase Design – Best Practices

Data Model impacts storage and performance

- 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
  - Reduce 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

*Justify every single Geodatabase element*
Geodatabase Design – Best Practices

Poor Design = Rework, slow performance and bugs

• Create Feature Datasets or Databases for each LOBs (Line of Businesses)
  - Depends on size, access, usage and maintenance
  - E.g. Landbase, Gas, Electric, Water GDBs, etc.
• No attributed Relationship Classes with no attributes
• 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

Keep common dataset / database for base map / landbase
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
- Less Annotation Classes within Annotation Feature Class
- Add Attribute Indexes
  - Label Expression
  - Definition Queries
  - Application Design

Analyze Annotation requirements and choose proper Annotation reference scale
Build
Build Geodatabase

- Create physical Geodatabase
  - Separate Data Owner from SDE / DBA user
  - Structure the implementation to Pilot → Phase I → Phase II → Phase III, etc.
  - Enough gap to accommodate the learned lessons

- Develop Data conversion specification document
  - Align it with Data Model

- Team review and demonstration

Capture, load and maintain data accurately
Geodatabase Access and Management

- Create roles / groups based on the access level
  - Total access will slow connection time!
- Configure client applications to manage data

<table>
<thead>
<tr>
<th>S.No</th>
<th>User Name</th>
<th>ArcCatalog Connection in Seconds with all access</th>
<th>ArcCatalog Connection in Seconds with reduced privileges</th>
<th>Performance difference in Seconds</th>
<th>Performance improvement in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rasu</td>
<td>30</td>
<td>12</td>
<td>18</td>
<td>150.00%</td>
</tr>
<tr>
<td>2</td>
<td>Andrew</td>
<td>21</td>
<td>12</td>
<td>9</td>
<td>75.00%</td>
</tr>
</tbody>
</table>

Capture, load and maintain data accurately
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
  - Use ArcObjects

Provide tips and tricks for editing crew regularly.
Build Geodatabase – Best Practices

Data is the brain of GIS nervous system!

- Aim for 100% Data accuracy
- Either populate or drop empty Fields
- Minimize Data Model / schema changes
- 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>

Unnecessary vertices add significant performance overhead
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 the Read Only Feature Classes / Tables
- Use Mosaic Datasets for Raster requirements
  - No massive imagery loading into EGDB

<table>
<thead>
<tr>
<th>OWNER</th>
<th>TABLE_NAME</th>
<th>NUM_ROWS</th>
<th>BLOCKS</th>
<th>AVG_ROW_LEN TO_CHAR(LAST_ANALYZED,'MON/DD/YYHH24:MI:SS')</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCFM</td>
<td>SDE_BLK_3</td>
<td>1145079</td>
<td>1190596</td>
<td>12 APR/19/16 05:36:10</td>
</tr>
<tr>
<td>ARCFM</td>
<td>BK_SDE_LOGFILE_DATA</td>
<td>4231223</td>
<td>8773</td>
<td>10 APR/19/16 05:32:14</td>
</tr>
</tbody>
</table>

*Static Raster Data does not need to participate in daily RDBMS backup*
Workflow Design
Implementation Dependencies
Understand intrigue challenges

• Map the project / data dependencies
  - Business user groups
  - Application designs
  - System integration requirements
  - etc.

• Organize cross functional skills/support team
  - Dedicated / Assigned ArcSDE Administrator

Talk to the right people to get the right information
Requirements and Workflows
Selection of GDBs drive efficiency

- **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*
  - Mobile User offline editing – *EGDB*
  - etc.

- Generally more than one Geodatabase is required

* A goal without a plan is just a wish! *
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 – Key Considerations

• Versioning structure
  - Move Edits to Base for simple Feature Classes
  - Recreate the Version after each Post for 3-level Version tree
  - *Problem: Unexpected conflict observed during reconcile*

• Estimate edit volumes and Version durations
• Conflict resolution mechanisms

Avoid 3-level version tree
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
  - 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

<table>
<thead>
<tr>
<th>Task</th>
<th>Performance - Before Tuning in Seconds</th>
<th>Performance - Before Tuning in Seconds</th>
<th>Performance Gain in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Opening ArcMap</td>
<td>46.22</td>
<td>15.41</td>
<td>199.94%</td>
</tr>
<tr>
<td>2 Open MXD (Master.mxd)</td>
<td>28.79</td>
<td>28.06</td>
<td>2.60%</td>
</tr>
<tr>
<td>4 Zoom to selected features</td>
<td>8.46</td>
<td>5.36</td>
<td>57.84%</td>
</tr>
<tr>
<td>5 Zoom to 5000</td>
<td>7.25</td>
<td>3.46</td>
<td>109.54%</td>
</tr>
<tr>
<td>6 Zoom out to 1000</td>
<td>4.93</td>
<td>2.25</td>
<td>119.11%</td>
</tr>
<tr>
<td>7 Zoom out to 500</td>
<td>5.32</td>
<td>0.91</td>
<td>484.62%</td>
</tr>
<tr>
<td>8 Start Editing</td>
<td>11.95</td>
<td>7.66</td>
<td>56.01%</td>
</tr>
<tr>
<td>9 Place a Structure (1 Pole with Push brace)</td>
<td>9.63</td>
<td>0.93</td>
<td>935.48%</td>
</tr>
<tr>
<td>10 Insert a Trans Line</td>
<td>7.85</td>
<td>1.11</td>
<td>697.21%</td>
</tr>
<tr>
<td>11 Stop Edits</td>
<td>4.73</td>
<td>1.50</td>
<td>215.33%</td>
</tr>
</tbody>
</table>
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><strong>OS Default Memory 4KB/ Page</strong></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_Rsvpd:   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

```
<table>
<thead>
<tr>
<th>Scale</th>
<th>Feature Class/Layer</th>
<th>Display in Seconds before tuning</th>
<th>Display in Seconds after tuning</th>
<th>Performance gain in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>250,000</td>
<td>Well Top</td>
<td>101.83</td>
<td>0.2</td>
<td>50815.00%</td>
</tr>
<tr>
<td>250,000</td>
<td>Last Gas Prod Rate</td>
<td>13.06</td>
<td>0.94</td>
<td>1289.36%</td>
</tr>
<tr>
<td>250,000</td>
<td>Cum Gas Production</td>
<td>17.18</td>
<td>0.9</td>
<td>1808.89%</td>
</tr>
<tr>
<td>250,000</td>
<td>Last Oil Prod Rate</td>
<td>8.78</td>
<td>0.2</td>
<td>4290.00%</td>
</tr>
<tr>
<td>250,000</td>
<td>Cum Oil Production</td>
<td>5.71</td>
<td>0.2</td>
<td>2755.00%</td>
</tr>
<tr>
<td>150,000</td>
<td>Well Top</td>
<td>94.04</td>
<td>0.14</td>
<td>67071.43%</td>
</tr>
<tr>
<td>150,000</td>
<td>Last Gas Prod Rate</td>
<td>9.87</td>
<td>0.86</td>
<td>1047.67%</td>
</tr>
<tr>
<td>150,000</td>
<td>Cum Gas Production</td>
<td>9.05</td>
<td>0.87</td>
<td>940.23%</td>
</tr>
<tr>
<td>150,000</td>
<td>Last Oil Prod Rate</td>
<td>4.07</td>
<td>0.11</td>
<td>3600.00%</td>
</tr>
<tr>
<td>150,000</td>
<td>Cum Oil Production</td>
<td>3.29</td>
<td>0.12</td>
<td>2641.67%</td>
</tr>
<tr>
<td>5,000</td>
<td>Well Top</td>
<td>36.71</td>
<td>0.03</td>
<td>122266.67%</td>
</tr>
<tr>
<td>1,000</td>
<td>Well Top</td>
<td>37.19</td>
<td>0.03</td>
<td>123866.67%</td>
</tr>
<tr>
<td>1,000</td>
<td>Last Gas Prod Rate</td>
<td>5.13</td>
<td>0.03</td>
<td>17000.00%</td>
</tr>
<tr>
<td>1,000</td>
<td>Cum Gas Production</td>
<td>5.1</td>
<td>0.03</td>
<td>16900.00%</td>
</tr>
</tbody>
</table>
```

Alert log:
Thu Dec 09 04:12:35 2018
WARNING: Heavy swapping observed on system in last 5 mins.

Alert log:
Thu Dec 09 15:47:39 2018
WARNING: Heavy swapping observed on system in last 5 mins.
Maintenance
Increase Performance and Scalability
Failing to prepare is preparing to fail!

- **Workflow estimations**
  - Number of outstanding versions
  - Versioning levels
  - Archiving
  - Traditional Vs Branch Versioning
  - etc.

- **Maintenance plan**
  - Mandatory tasks to keep performance
  - Delta table records Vs CPUs
  - Roles and responsibilities
  - etc.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Display Scale</th>
<th>Layer Name</th>
<th>Before Maintenance - Display in Seconds</th>
<th>After Maintenance - Display in Seconds</th>
<th>Performance Improvement in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50,000</td>
<td>RoadCL &gt; 10,000</td>
<td>0.34</td>
<td>0.3</td>
<td>13.33%</td>
</tr>
<tr>
<td>2</td>
<td>15,000</td>
<td>Water Mains</td>
<td>0.15</td>
<td>0.09</td>
<td>66.67%</td>
</tr>
<tr>
<td>3</td>
<td>10,000</td>
<td>MapLink</td>
<td>1.02</td>
<td>0.68</td>
<td>41.18%</td>
</tr>
<tr>
<td>4</td>
<td>10,000</td>
<td>CH2M_Mains</td>
<td>1.15</td>
<td>0.57</td>
<td>50.43%</td>
</tr>
<tr>
<td>5</td>
<td>10,000</td>
<td>Leaders</td>
<td>0.58</td>
<td>0.07</td>
<td>84.77%</td>
</tr>
<tr>
<td>6</td>
<td>10,000</td>
<td>Annotation</td>
<td>0.69</td>
<td>0.15</td>
<td>76.87%</td>
</tr>
<tr>
<td>7</td>
<td>10,000</td>
<td>Water Mains</td>
<td>0.96</td>
<td>0.6</td>
<td>36.00%</td>
</tr>
<tr>
<td>8</td>
<td>10,000</td>
<td>CTParcel</td>
<td>1.2</td>
<td>0.42</td>
<td>65.17%</td>
</tr>
<tr>
<td>9</td>
<td>6,000</td>
<td>CH2M_Mains</td>
<td>0.68</td>
<td>0.56</td>
<td>18.71%</td>
</tr>
<tr>
<td>10</td>
<td>6,000</td>
<td>Water Mains</td>
<td>0.88</td>
<td>0.52</td>
<td>39.52%</td>
</tr>
<tr>
<td>11</td>
<td>1,000</td>
<td>SwingTies</td>
<td>7.74</td>
<td>0.65</td>
<td>98.93%</td>
</tr>
<tr>
<td>12</td>
<td>500</td>
<td>SwingTies</td>
<td>6.12</td>
<td>0.74</td>
<td>90.17%</td>
</tr>
<tr>
<td>13</td>
<td>500</td>
<td>Water Mains</td>
<td>0.72</td>
<td>0.24</td>
<td>69.00%</td>
</tr>
<tr>
<td>14</td>
<td>500</td>
<td>sbDrivesways</td>
<td>0.37</td>
<td>0.16</td>
<td>53.33%</td>
</tr>
</tbody>
</table>

*Few private and orphan versions induced the bottlenecks

Mitigate the risk with proper system capacity & 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

Automate the daily maintenance process
Versions maintenance

- Use private Versions for larger versioning environment
  - Keep the total number less
  - Sync frequently
  - Monitor
- Complete the batch processes within maintenance period
  - Improve Hardware and/or Software design
  - Additional CPUs
  - Multi threads
  - Feature/Schema cache
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
Geodatabase Maintenance – Best Practices

- **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>

<table>
<thead>
<tr>
<th>Feature Class Name</th>
<th>Calculating 2 records before GP History Removal in Seconds</th>
<th>Calculating 2 records after GP History Removal in Seconds</th>
<th>Performance Difference in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sewer</td>
<td>58.67</td>
<td>8.09</td>
<td>625.22%</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 TO_CHAR(LAST_ANALYZED, 'MON/DD/YYHH24:MI:SS')</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCFM</td>
<td>EDITEDFEATURESTRACKING</td>
<td>12330909</td>
<td>122954</td>
<td>140 APR/23/16 23:44:17</td>
</tr>
<tr>
<td>ARCFM</td>
<td>EDITEDGRIDS</td>
<td>12964927</td>
<td>32969</td>
<td>32 APR/23/16 23:47:55</td>
</tr>
<tr>
<td>SDE</td>
<td>ARCSDEUSERLOG</td>
<td>56558633</td>
<td>259246</td>
<td>62 APR/24/16 12:28:25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OWNER</th>
<th>TABLE_NAME</th>
<th>NUM_ROWS</th>
<th>BLOCKS</th>
<th>AVG_ROW_LEN TO_CHAR(LAST_ANALYZED, 'MON/DD/YYHH24:MI:SS')</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDE</td>
<td>GDBM_RECONCILE_HISTORY</td>
<td>4759429</td>
<td>50484</td>
<td>72 AUG/12/17 12:40:34</td>
</tr>
<tr>
<td>ARCFM</td>
<td>USAGEINFORMATION</td>
<td>6377236</td>
<td>221857</td>
<td>39 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 spatial index rebuild
Geodatabase Maintenance – Best Practices

- 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_VALIDATE=>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);
```

<table>
<thead>
<tr>
<th>TABLE_NAME</th>
<th>GSIZE1</th>
<th>GSIZE2</th>
<th>GSIZE3</th>
</tr>
</thead>
<tbody>
<tr>
<td>COATING</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CPBOND</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CFCABLE</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CFCPOWDER</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CPANODE</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LINECROSSING</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MARKER</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PIPEEXPOSURE</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TIEINMETHOD</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DOCUMENTPOINT</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INJECTION</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PIKROUTE</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PIPEDEPTH</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RL_SITE</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PIPEJOIN</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PIPELENGTH</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WELL</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LEAK</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DOT_CLASS_PREVIOUS_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ALIGNMENT_SHEETS</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EXTERNAL_COATING_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MISC_FITTING_DETAILS</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ALIG_SHT_CROSS_REF_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CLOSURES_DETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HCA_PREVIOUS_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INJECTOR_DETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ODORANT_RANGE_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OFFLINE_COMPENSATIONPOLYGONS</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FIG_SIGNAL_DETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PIR_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RIGHT_OF_WAY_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RIVER_WEIGHT_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHEET_NOTE_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TAP_DETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PIPE_SEG_PIR_BUFF</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MAP_CALC_RDETAILED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Geodatabase Maintenance – Best Practices

- A Sample daily maintenance (batch process) for multi user Geodatabase:
  1. Backup the Database
  2. Synchronize any Replica version.
  3. Delete the orphan / unnecessary versions.
  4. Drop the orphan keyset tables
  5. Reconcile and Post all/eligible Versions – throughout the Day
  6. Only Reconcile all versions (>100 versions - parallel reconcile)
  7. Update Database statistics - optional

Continue…..

```sql
50  END LOOP;
51  dbms_output.put_line('Dropped "||cnt||" keyset tables.');
52  END;
53  /
54  / Dropped 31038 keyset tables.
```
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.

<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_ANALY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RASU</td>
<td>SDE_LOGFILES</td>
<td>7,488</td>
<td>244</td>
<td></td>
<td>93 JUN/17/19 22:03:13</td>
</tr>
<tr>
<td>RASU</td>
<td>SDE_LOGFILE_DATA</td>
<td>86,832,159</td>
<td>171,379</td>
<td></td>
<td>10 JUN/03/19 22:02:56</td>
</tr>
<tr>
<td>ANDR</td>
<td>SDE_LOGFILES</td>
<td>783</td>
<td>13</td>
<td></td>
<td>91 MAY/14/19 22:02:05</td>
</tr>
<tr>
<td>ANDR</td>
<td>SDE_LOGFILE_DATA</td>
<td>28,887,671</td>
<td>57,017</td>
<td></td>
<td>10 AUG/26/18 06:04:45</td>
</tr>
</tbody>
</table>
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 it can produce more conflicts when run under SDE.Default version directly.
Monitor
ArcGIS Monitor

- Add RDBMS Queries
- Monitor key performance indicators
- Keep 15 minutes sampling interval
ArcGIS Monitor – EGDB Add-on

- Configure EGDB Add-ons per RDBMS
- Use EGDB health tool

https://community.esri.com/community/implementing-arccgis/blog/2019/05/10/using-egdbhealth-to-evaluate-a-geodatabase?et=watches.email.blog

Monitor, interpret and respond
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

Implementing an Enterprise Geodatabase
<table>
<thead>
<tr>
<th>WORKSHOP</th>
<th>LOCATION</th>
<th>TIME FRAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Enterprise Geodatabase: Automating Administration Tasks Using Python</td>
<td>• SDCC - Expo Demo Theater 04</td>
<td>• Thursday 7/11/2019 10:00 AM - 10:45 AM</td>
</tr>
<tr>
<td>• Geodatabase: Ensuring Data Quality with Attribute Rules and Contingent Values</td>
<td>• SDCC - Ballroom 06 E</td>
<td>• Thursday 7/11/2019 4:00 PM - 5:00 PM</td>
</tr>
</tbody>
</table>
Please Share Your Feedback in the App

Download the Esri Events app and find your event

Select the session you attended

Scroll down to “Survey”

Log in to access the survey

Complete the survey and select “Submit”
Thanks!