Esri Best Practices: Implementing an Enterprise Geodatabase

Rasu Muthurakku, Enterprise Solutions Architect / Consultant

Andrew Sakowicz, Enterprise Implementation Practice Lead

THE SCIENCE OF WHERE

GIS INSPIRING WHAT'S NEXT

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

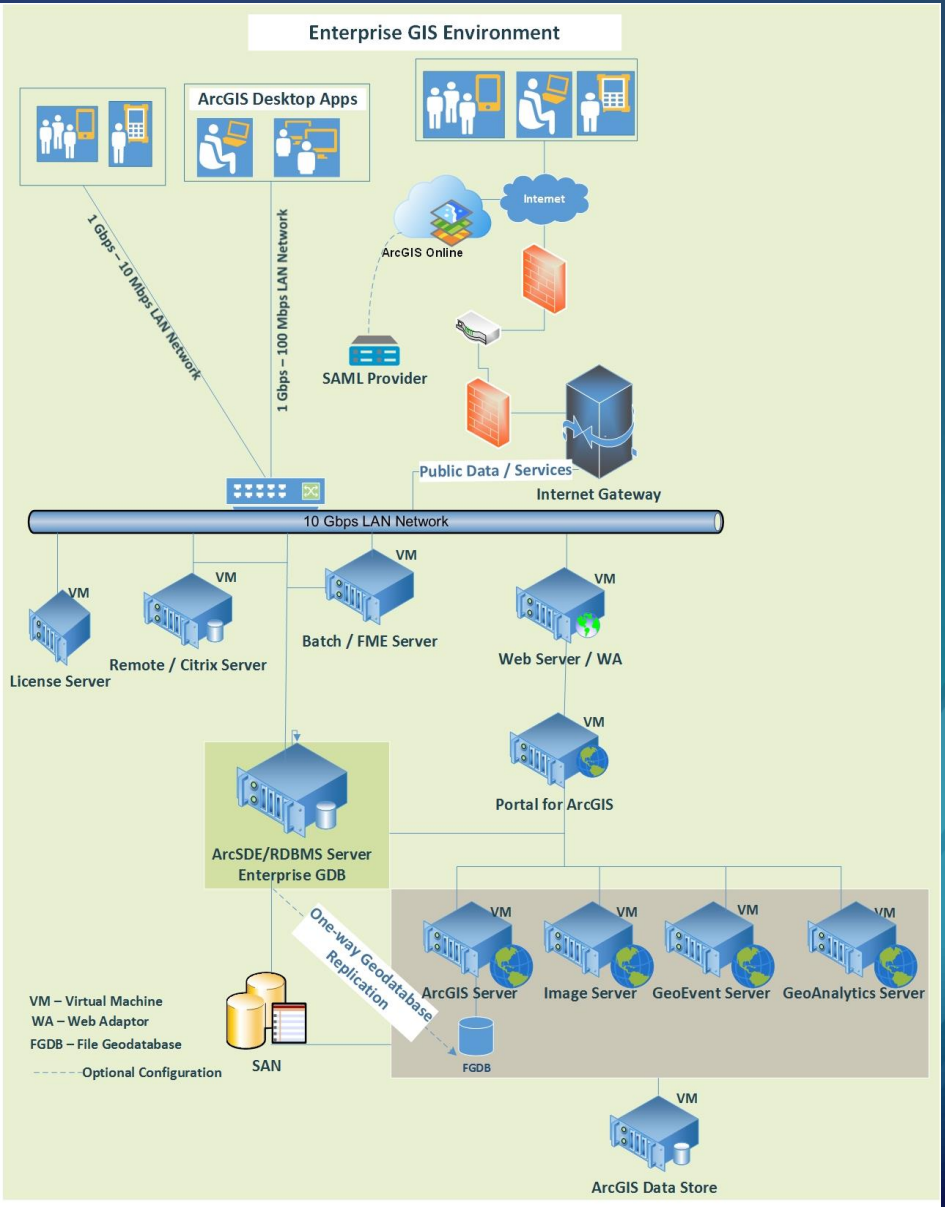
Overview

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 New
- 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

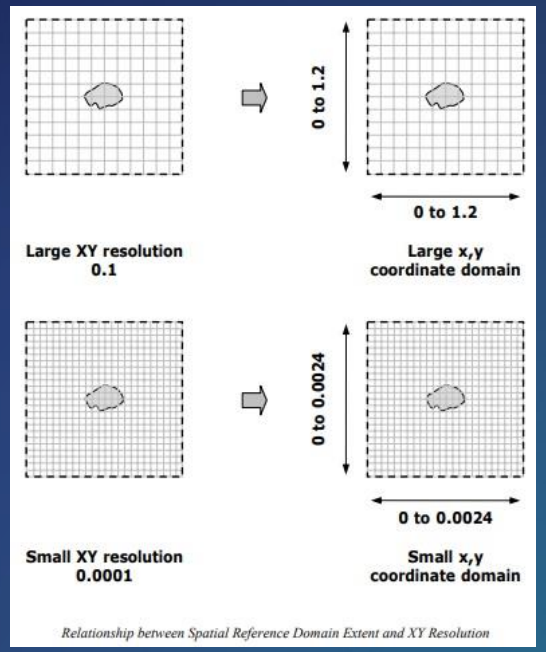
Why?

esri THE SCIENCE OF WHERE™

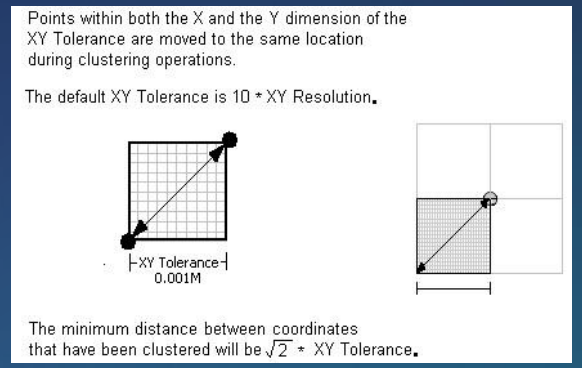
Expensive Rework

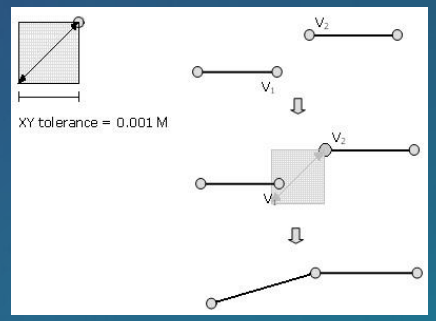
A Stitch in Time Saves Nine!

- **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.*



Relationship between Spatial Reference Domain Extent and XY Resolution





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.

```
Versioning statistics
---------------------------------------------
Number of versions: 18,055
Number of versions blocking DEFAULT: 5,966
Top 5 blocking versions...
GIS.Test_241625
RASU.WORK_ORDER1
ANDREW.ARCGIS_MONITOR1
ARCFM.SESSION_1254
ADMIN.Freeze_2018
Number of states: 20,139
Number of state lineages: 9,497,416
DEFAULT versions lineage length: 1,068
Last compress: JUN-18-2018
```

```
SQL > Exec DBMS_STATS.GATHER_SCHEMA_STATS ('ARCFM', estimate_percent=>100, DEGREE=> 15, CASCADE=>TRUE, No_Invalidate=>false);
      PL/SQL procedure successfully completed.

---ArcSDE A Table Rows

SQL> select sum(num_rows) as Total_A_Table_Rows from dba_tables where table_name in (select 'A'||registration_id from sde.table_registry);

TOTAL_A_TABLE_ROWS
------------------
  12,154,746

---ArcSDE D Table Rows
SQL> select sum(num_rows) as Total_D_Table_Rows from dba_tables where table_name in (select 'D'||registration_id from sde.table_registry);

TOTAL_D_TABLE_ROWS
------------------
  9,338,738
```

Separate Myths / Emotions from Evidence Based Facts!

Overlooked Dependencies

Understand Intrigue Challenges!

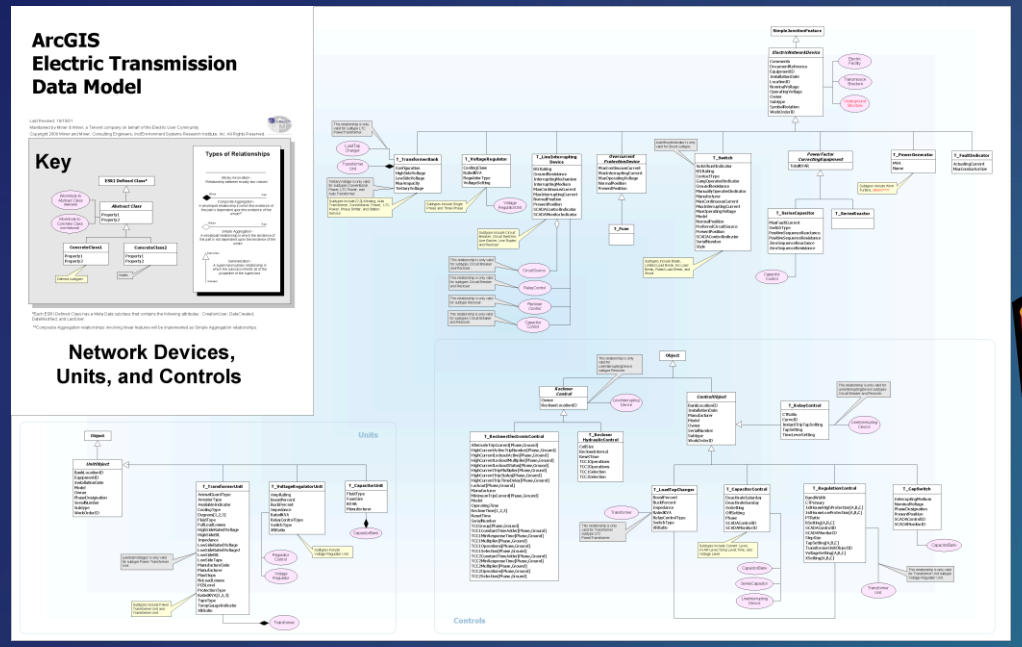
- **Key Inputs**
  - User / Business Groups
  - Functional and Non-Functional Requirements
  - Application Designs
  - etc.
- **System Architecture and Capacity**
- **Need a Dedicated / Assigned ArcSDE Administrator**

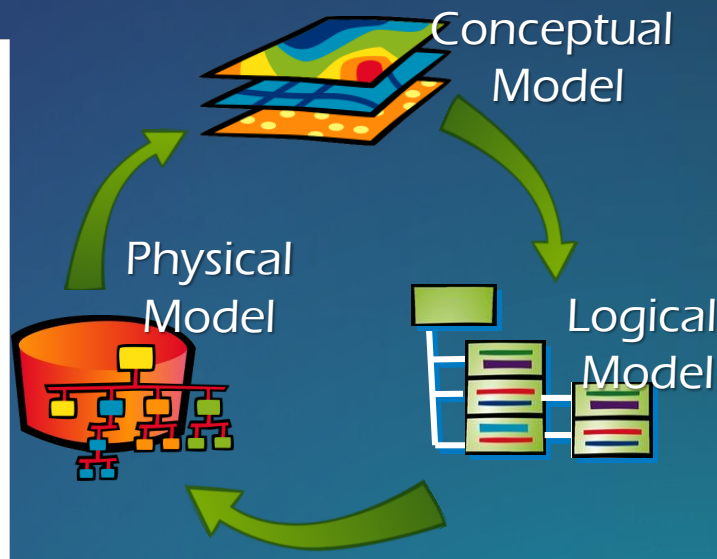
Talk to the Right People to Get the Right Information!

Geodatabase Design

Geodatabase Design – Data Modeling

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



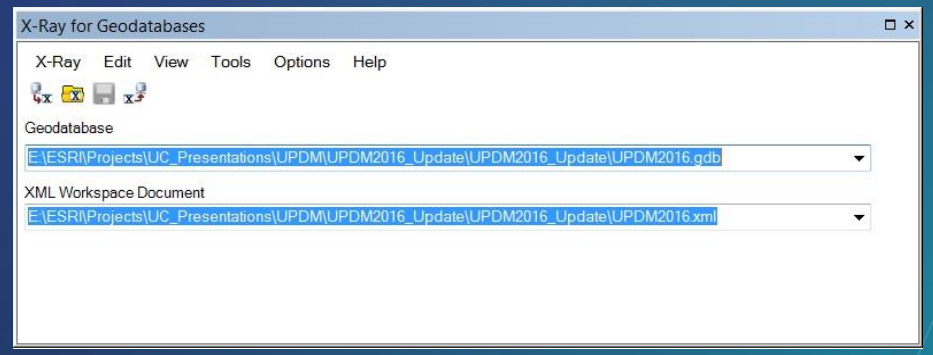


- **Logical Design**
  - Define Tabular Database Structure
  - Define Relationships
  - Determine Spatial Properties
  - Document

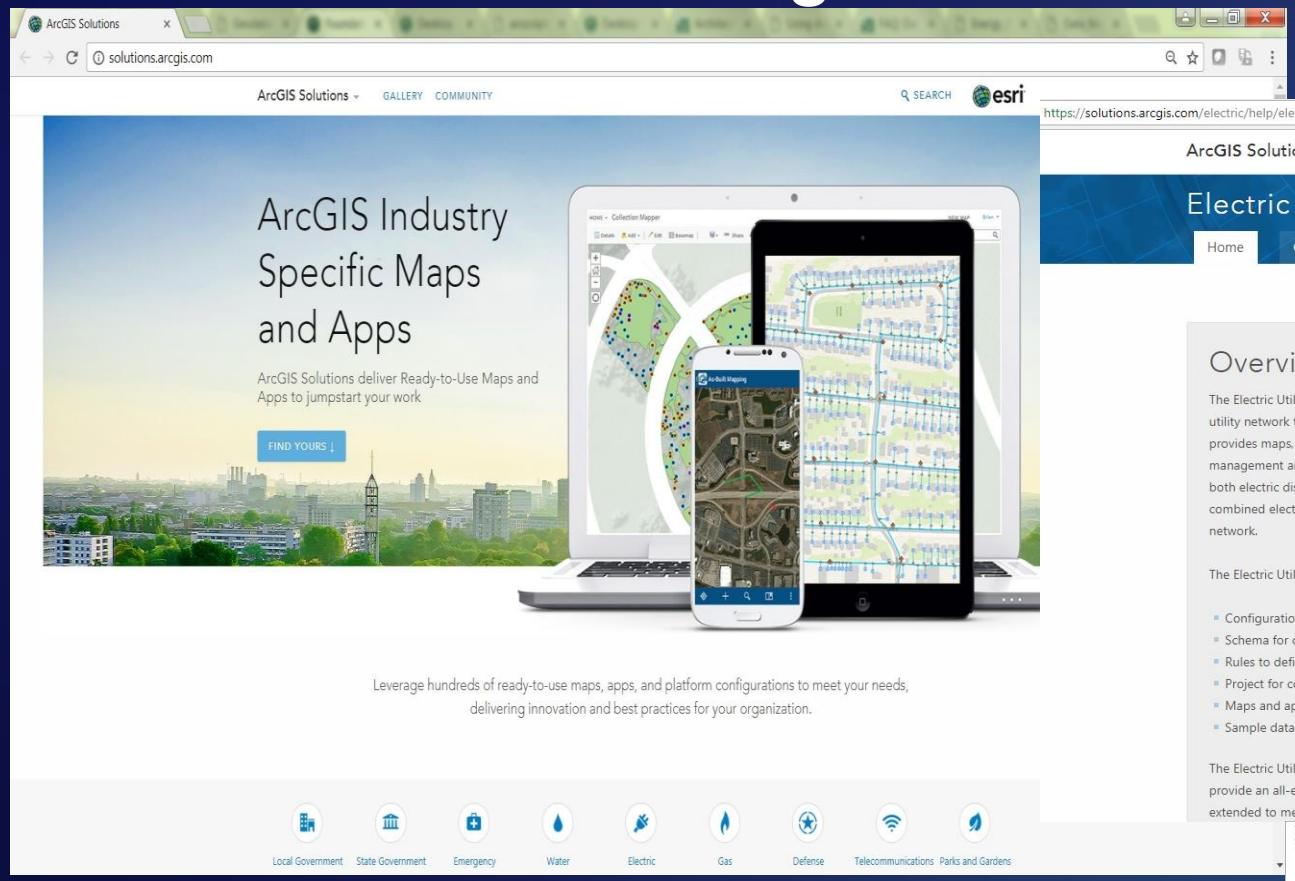
- **Physical Design**
  - Create and Implement Model Design
  - Generate Physical Schema in the RDBMS / FGDB
  - Testing and Validation
  - Document

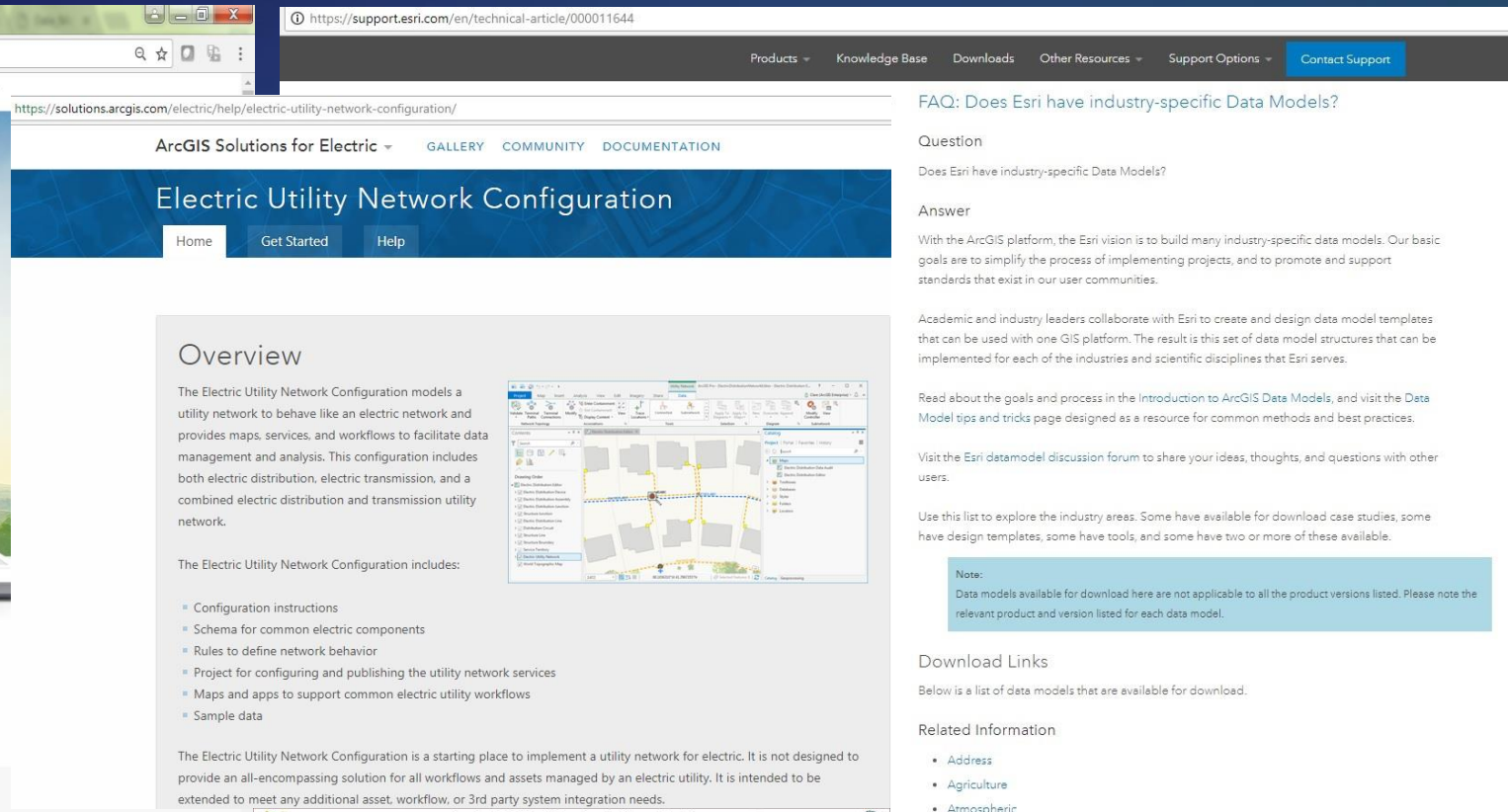
Geodatabase Design – Tools

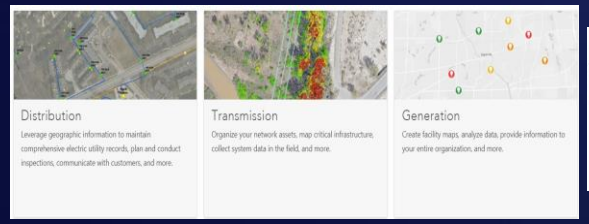
- X-Ray Add-In
- Geodatabase Diagrammer
- Sparx Systems' Enterprise Architect
- Geometric Network Configuration Manager

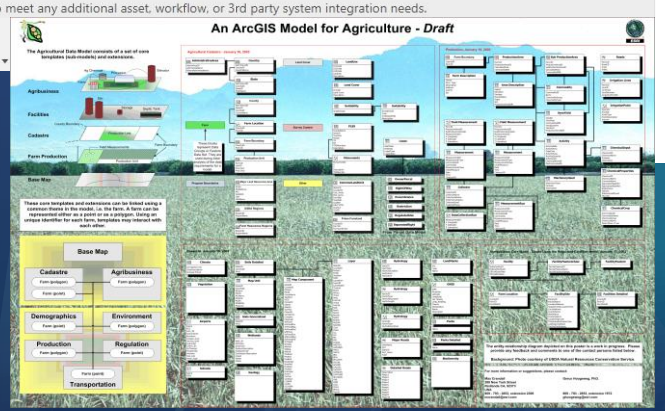


Geodatabase Design – ArcGIS Data Models







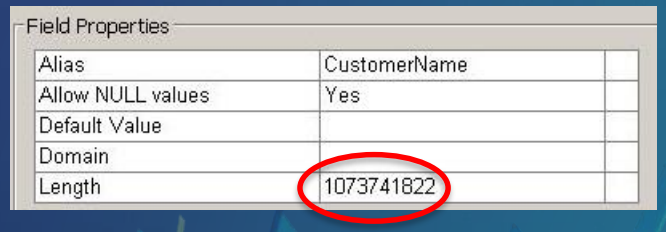


Leverage the Existing Data Models!

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



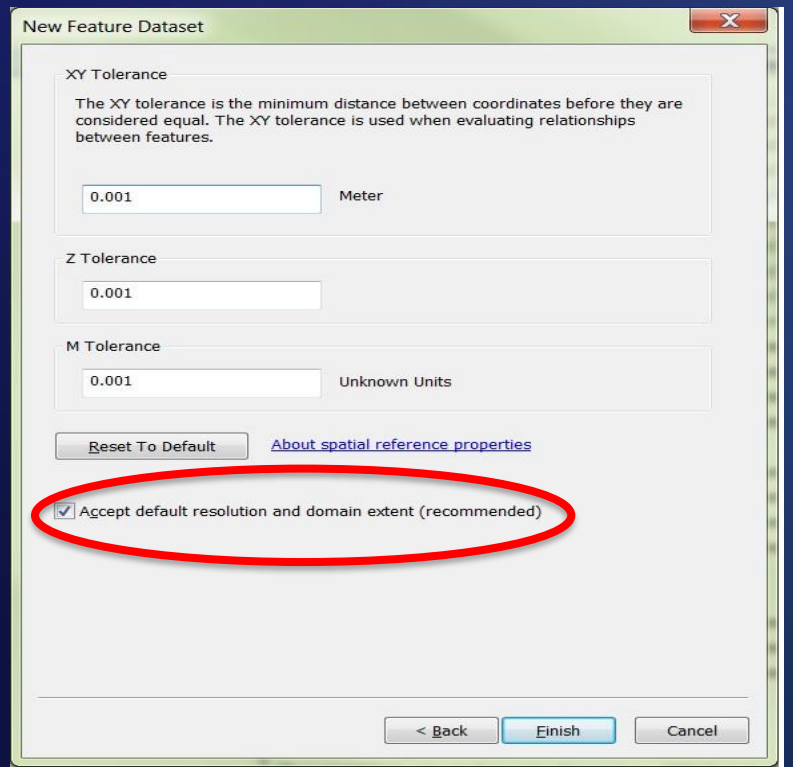
Field Properties

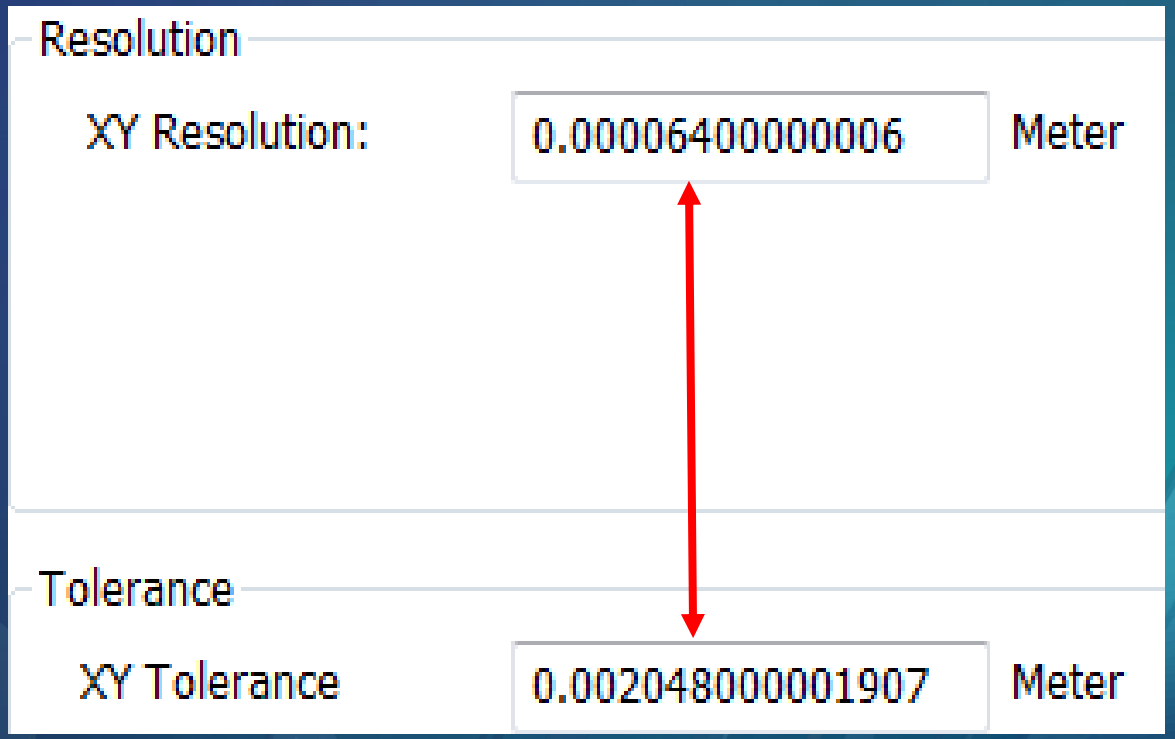
Alias	CustomerName
Allow NULL values	Yes
Default Value	
Domain	
Length	1073741822

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**

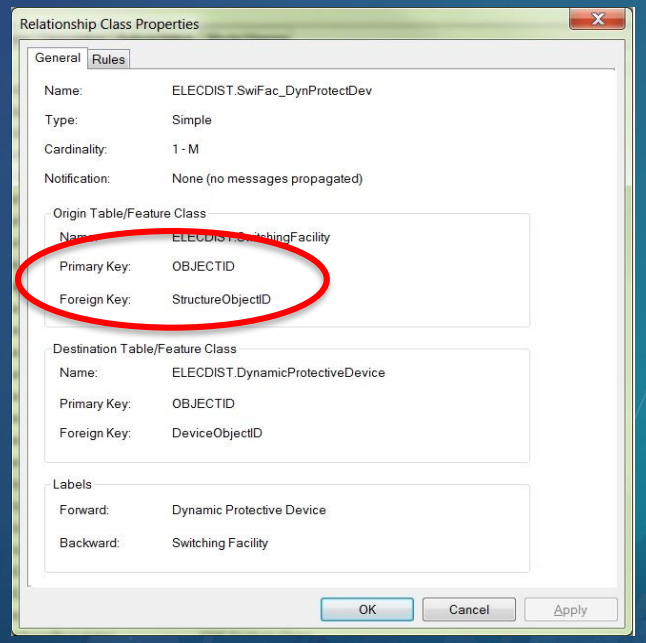




Geodatabase Design – Best Practices

Poor Design = Slow Performance and Bugs

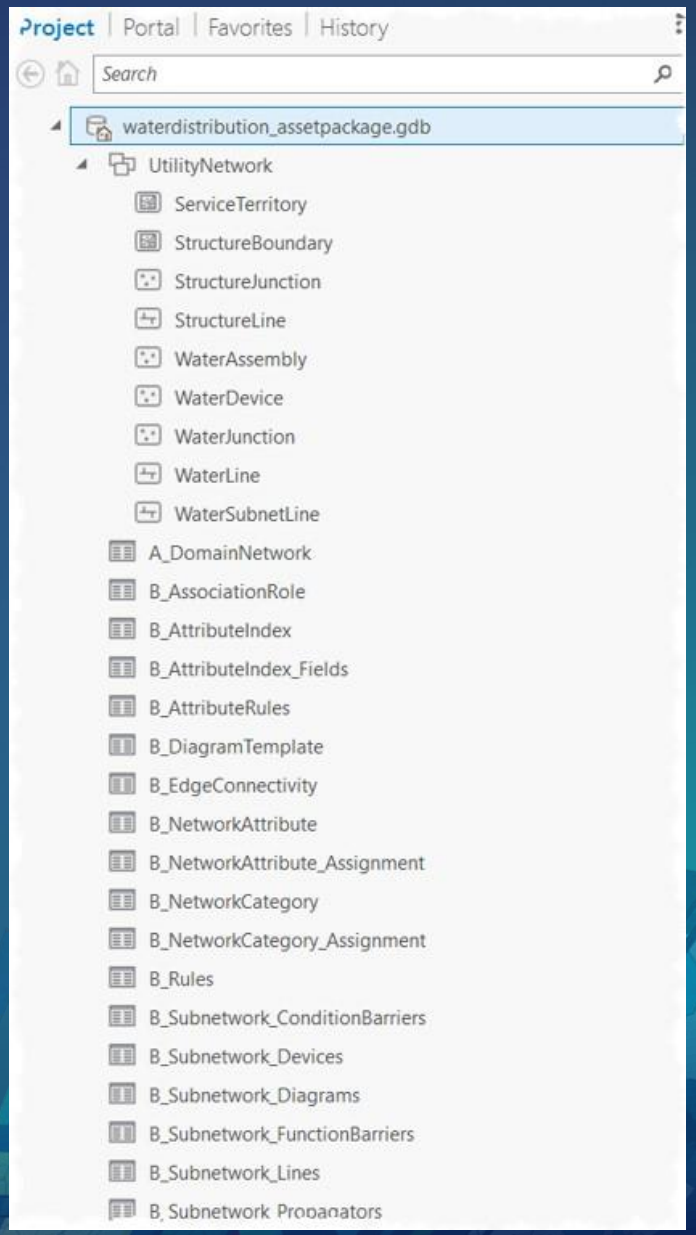
- No Attributed Relationship Classes for Empty Tables!
- Use Many to Many Relationship Classes Only When Necessary
- Don’t Use Objectid as Primary Key for Relationship Classes
  - Unexpected Replication Behavior
  - Additional Processing During Synchronization



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

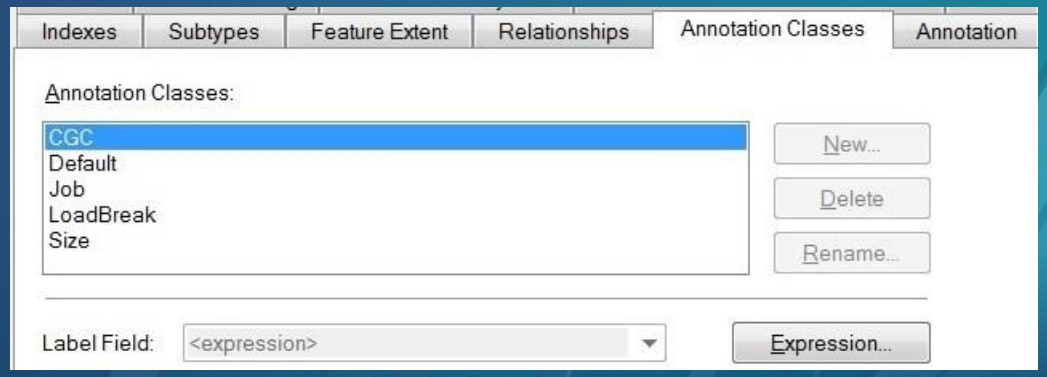
Prevention is Better Than Cure!

- 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

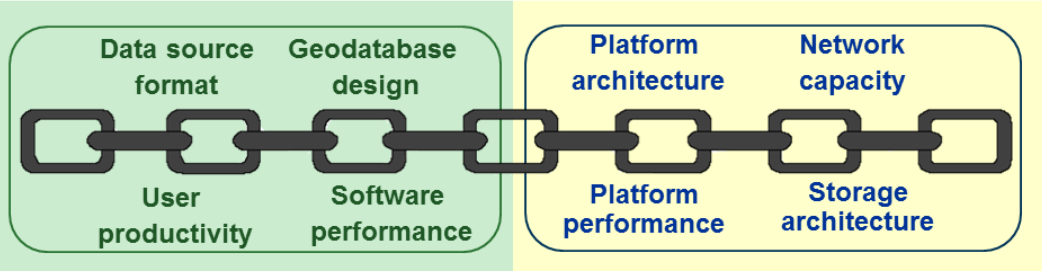
esri THE SCIENCE OF WHERE™

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

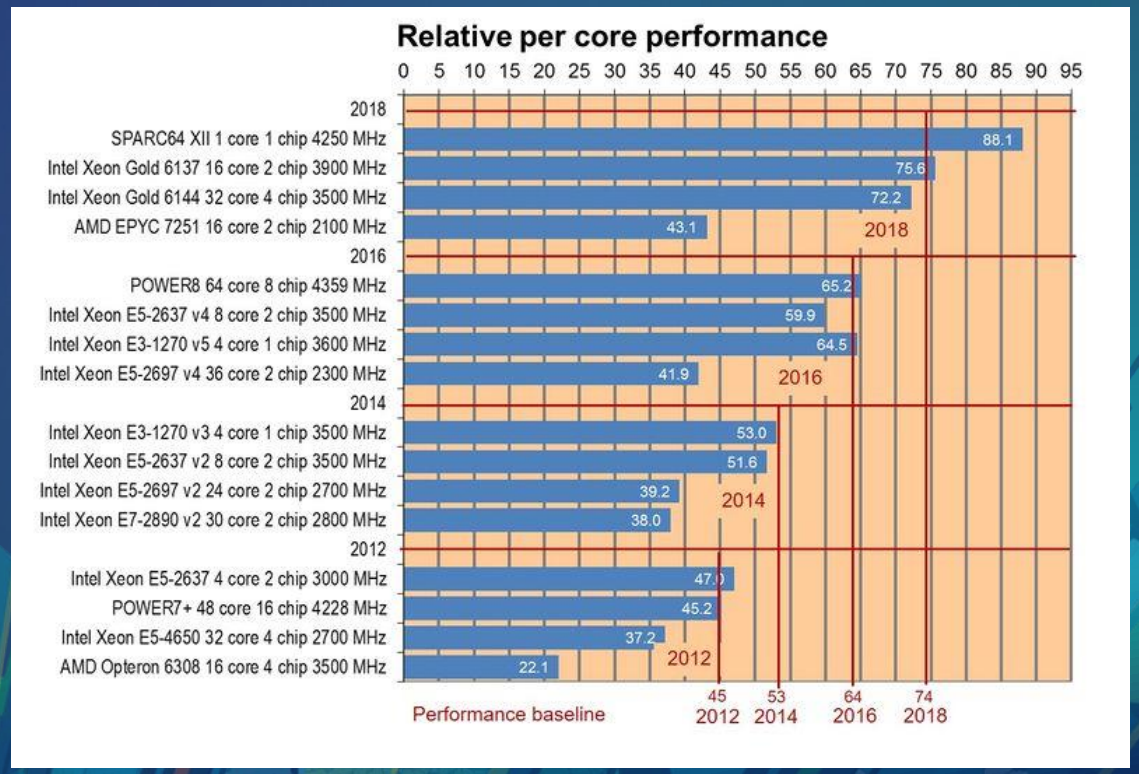
The Geodatabase Design and Associated System Architecture Helps!

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 =

    (Base Table – D# Table) + (A# Table – 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!

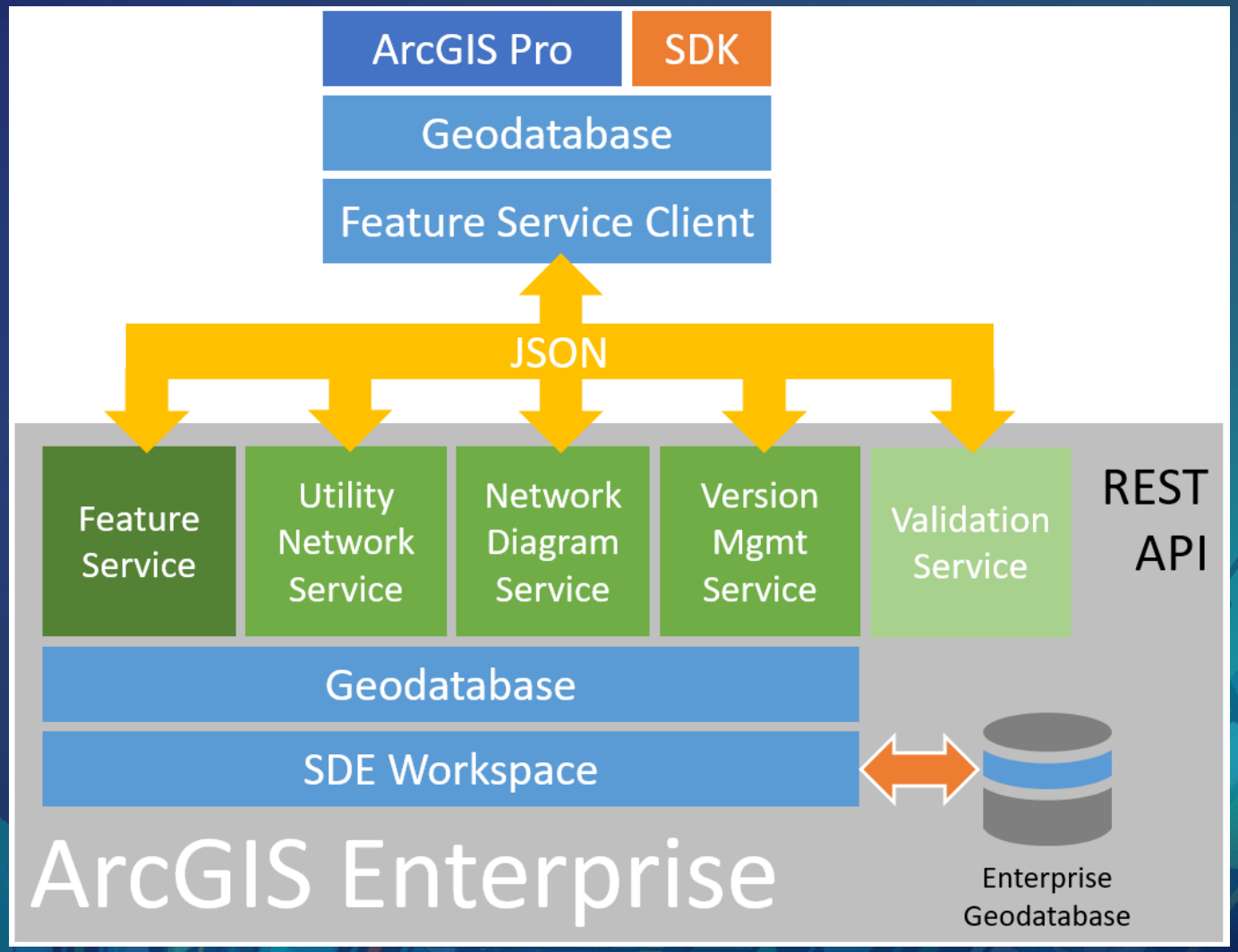
```
SQL>
SQL> prompt 2. STATES
2. STATES
SQL>
SQL> select count(*) from sde.states;

Tue Sep 26


  COUNT(*)
----------
    113,904

SQL>
SQL> prompt 3. STATE_LINEAGES
3. STATE_LINEAGES
SQL> select count(*) from sde.state_lineages;

Tue Sep 26


  COUNT(*)
----------
 106,274,130

SQL>
SQL> spool off
```

```
---ArcSDE A Table Rows

SQL> select sum(num_rows) as


TOTAL_A_TABLE_ROWS
------------------

12,154,746

SQL>

SQL> ---ArcSDE D Table Rows
SQL> select sum(num_rows) as

TOTAL_D_TABLE_ROWS
------------------
             9,338,738


SQL>
SQL> Spool off
```

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



New / Additional Solution Architecture Options

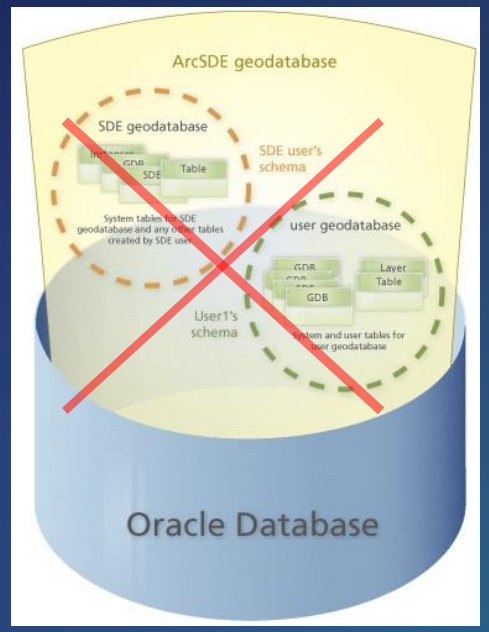
SAP HANA In Memory Database Supports GDB!

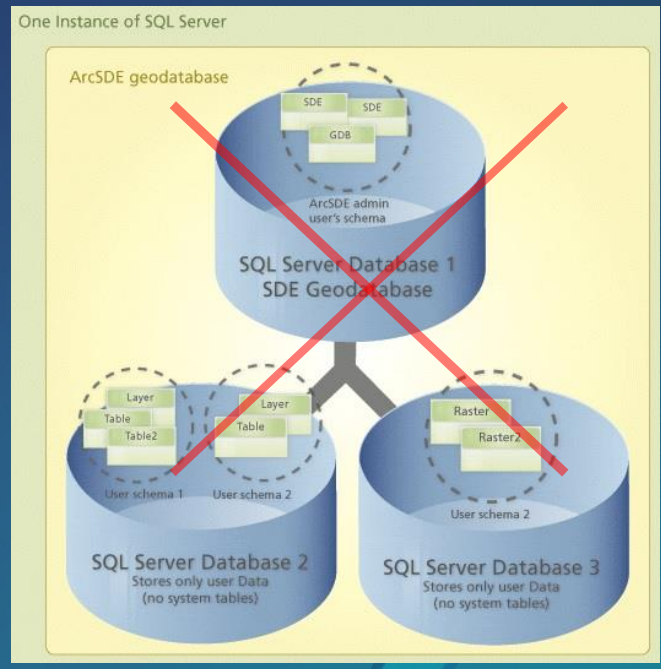
- **Only Enable a Geodatabase in SAP HANA 2.0 SPS02, Can Only Connect from:**
  - ArcGIS Pro 2.1
  - ArcGIS Pro 2.2
  - ArcGIS Server 10.6
  - ArcGIS Server 10.6.1
- **Branch Versioning**
- **Mobile Geodatabase Containing Offline Feature Data.**
  - .geodatabase is based on a Portable SQLite Format
  - Primarily for Offline Workflows in ArcGIS Runtime Apps

Esri SDK	.geodatabase	File geodatabase	Personal geodatabase	Enterprise geodatabase
ArcObjects	Not supported	Supported	Supported	Supported
ArcGIS Runtime for Android	Supported	Not supported	Not supported	Supported only if using services
ArcGIS Runtime for iOS and OS X	Supported	Not supported	Not supported	Supported only if using services
ArcGIS Runtime for Java	Supported	Supported only if using local server	Not supported	Supported only if using services
ArcGIS Runtime for .NET	Supported	Supported only if using local server	Not supported	Supported only if using services
ArcGIS Runtime for Qt	Supported	Supported only if using C++ API and local server. Not supported with QML API.	Not supported	Supported only if using services

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

```
DELETED_AT IN (SELECT  l.lineage_id FROM SDE.state_lineages l WHERE
l.lineage_name = :lineage_name2 AND l.lineage_id <= :state_id2) AND
SDE_STATE_ID > :"SYS_B_5") AND a.SDE_STATE_ID = SL.lineage_id AND
SL.lineage_name = :lineage_name3 AND SL.lineage_id <= :state_id3) V__138
```

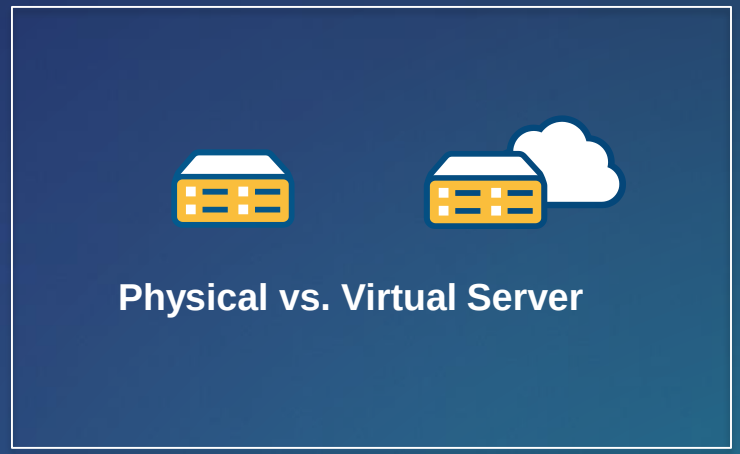
call	count	cpu	elapsed	disk	query	current	rows
Parse	1	0.00	0.00	0	0	0	0
Execute	10	0.08	0.13	0	2289	0	0
Fetch	10	8.56	15.53	58	752569	20	436
total	21	8.65	15.67	58	754858	20	436

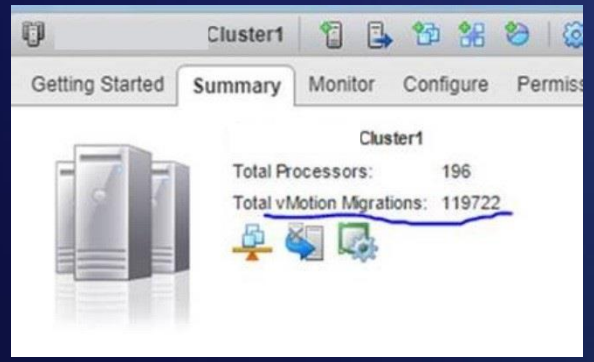
GIS Benefits from Dedicated and Powerful Processors!

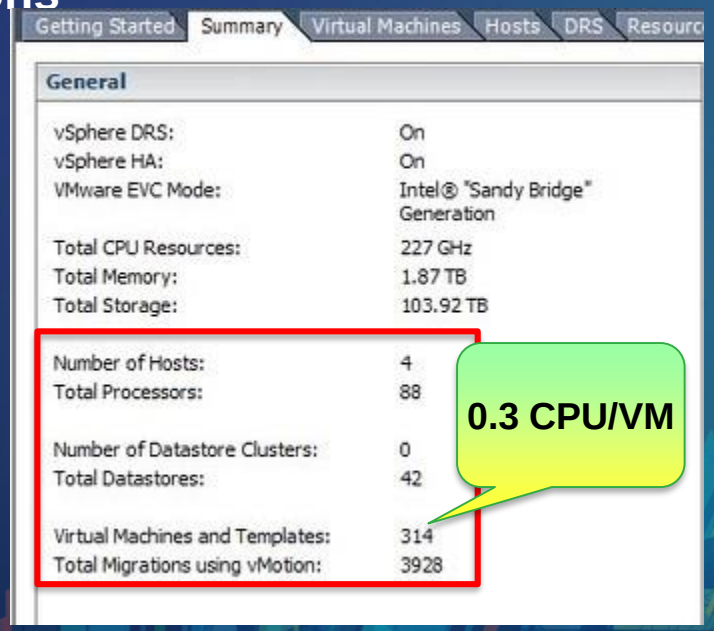
Virtualized Database Servers – Key Considerations

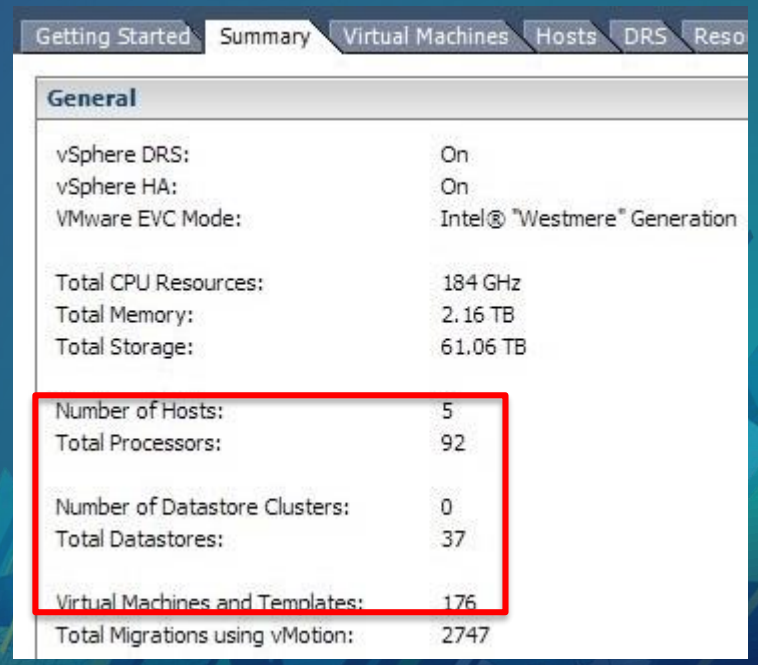
Speed is More Important than Utilization!

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





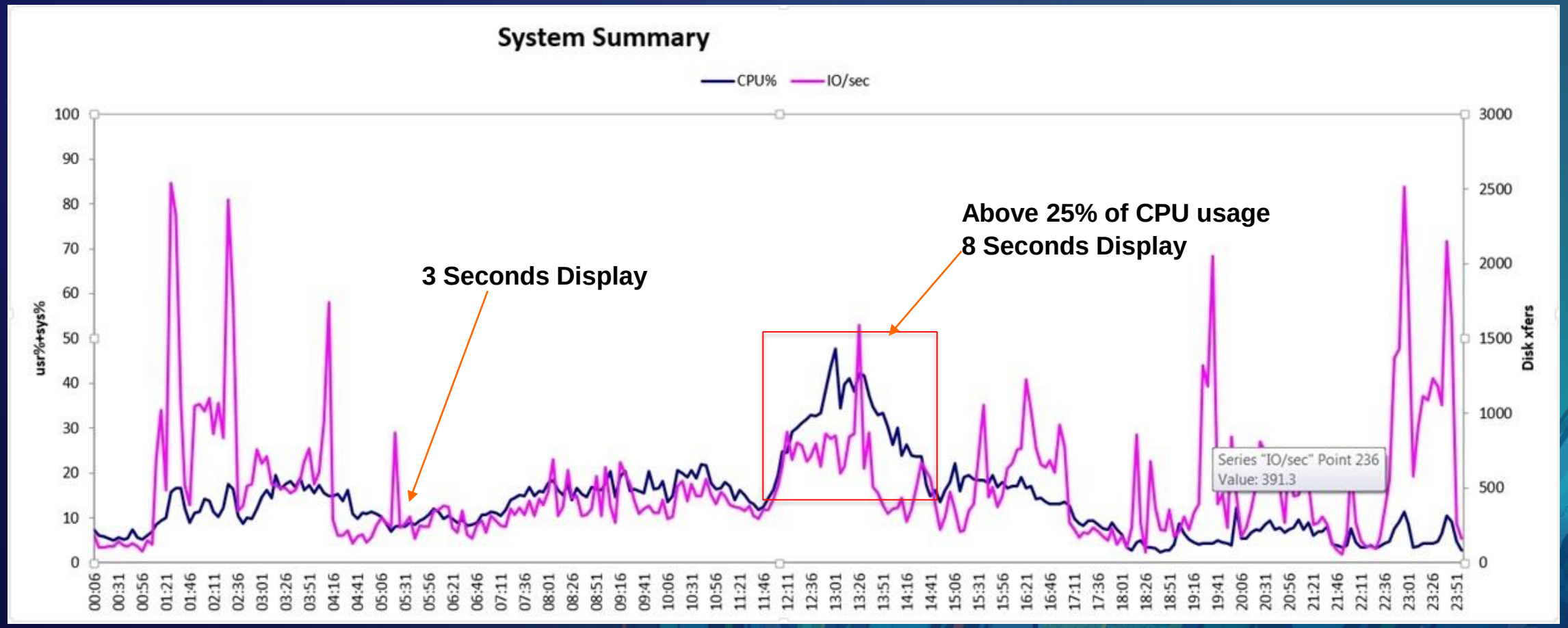




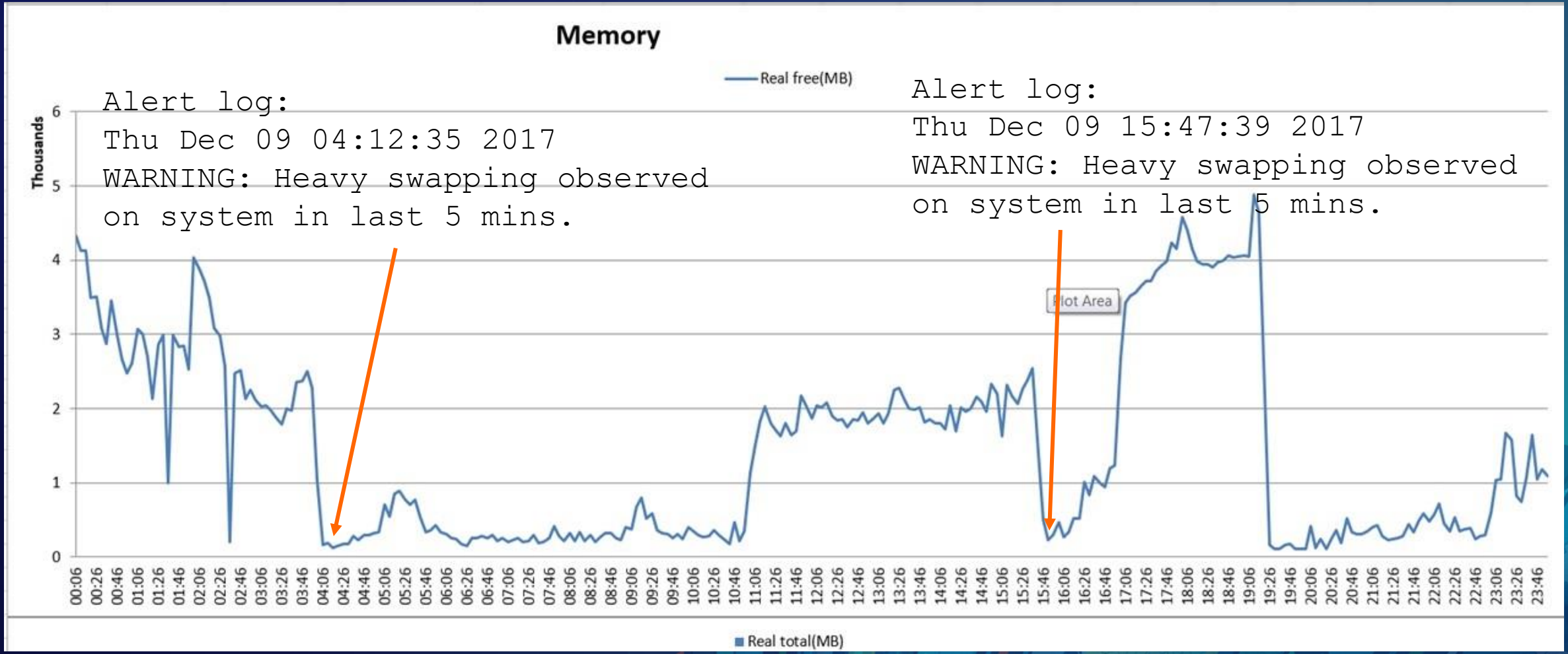
Complex and Busy Databases Need Dedicated Machines!

Symptoms of Over-Committed vCPUs

- vCPU Usage Vs MXD Display - Mxdperfstat Tool Results



Additional Memory Requirements



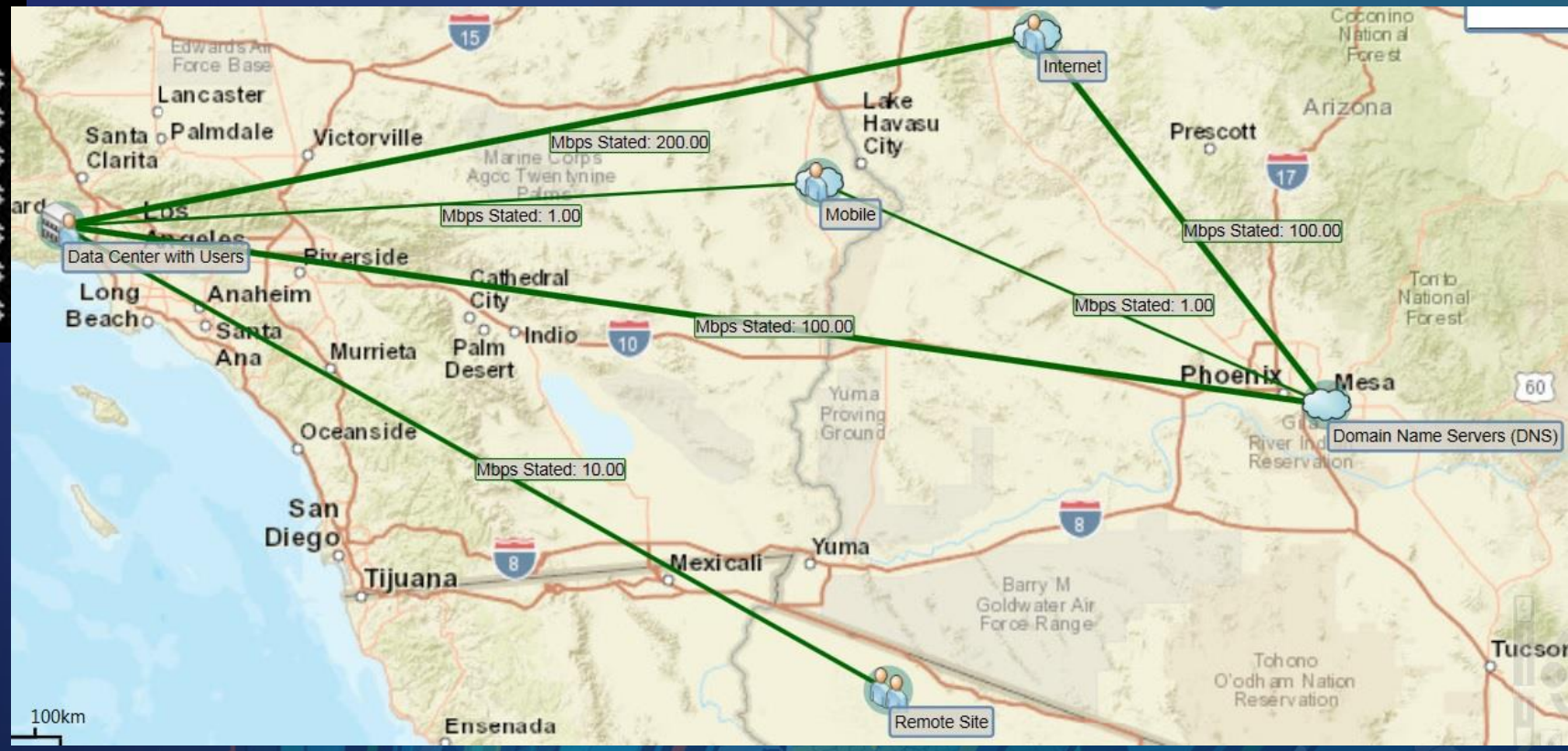
Network Planning

Establish and Configure DNS Appropriately!

```
C:\Users\          >tracert

Tracing route to
over a maximum of 30 hops:

  1     55 ms     55 ms     55 ms
  2     55 ms     55 ms     55 ms
  3    115 ms     58 ms     62 ms
  4    111 ms    111 ms    112 ms
  5    110 ms    109 ms    110 ms
  6    110 ms    113 ms    110 ms
  7    109 ms    109 ms    109 ms
```

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+

Task	Performance - Before CPU Change (in Seconds)	After Change in CPU Entitlement from 1 CPU to 4 CPU (in Seconds)	Performance improvement in %
Opening ArcMap thru Citrix (ArcGIS Desktop Launcher - for setting the licenses)	26.01	21.37	21.71%
ArcFM Locator - Zoom to Distribution	29.45	11.15	164.13%
Zoom to 1000 ft	8.92	7.33	21.69%
Zoom to 500ft	4.36	3.69	18.16%
Zoom out to 100ft	5.02	4.3	16.74%
Zoom out to 50ft	4.19	3.73	12.33%
Place UG Primary Conductor with Ducts (2)	13.3	6.68	99.10%
Distribution Pole	5.46	4.19	30.31%
Split OH Primary Conductor - Custom Split	21.23	15.76	34.71%
Delete Transformer	14.05	7.8	80.13%
Place the above transformer and Jumper	9.85	5.24	87.98%
Place a pole	2.48	0.82	202.44%
Place Primary OH-ABC Phase	5.1	2.89	76.47%
No Save Edits	9.86	6.2	59.03%

*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

Impact of an Older RDBMS Version Bug

MXDPERFSTAT – Production Environment

S.No	Display Scale	Layer Name	Before tuning - Display in Seconds	After Tuning (Only SDE Direct Connect and DFQ*) - Display in Seconds	Performance Improvement in %
1	1,000,000	TRANS LINE	0.94	0.37	154.05%
2	500,000	STATION	0.26	0.13	100.00%
3	50,000	TRANS LINE	0.8	0.18	344.44%
4	50,000	STRUCTURE	44.62	0.72	6097.22%
5	25,000	STRUCTURE	46.38	0.49	9365.31%
6	25,000	TRANS LINE	0.83	0.16	418.75%
7	1,000	STRUCTURE	42.26	0.28	14992.86%
8	1,000	TRANS LINE	0.76	0.31	145.16%
9	500	STRUCTURE	41.66	0.27	15329.63%
10	500	TRANS LINE	0.74	0.21	252.38%

*DFQ (Definition Query) = Objectid is not null

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

ArcGIS Server Performance			
Service Name	REST Export in Seconds - RDBMS Behind Firewall	REST Export in Seconds - Local File Geodatabase	Performance Gain in %
USE_BASE	12.018	0.287	4087.46%
ArcServer_Assets	5.266	0.424	1141.98%

*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

Build

esri THE SCIENCE OF WHERE™

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!

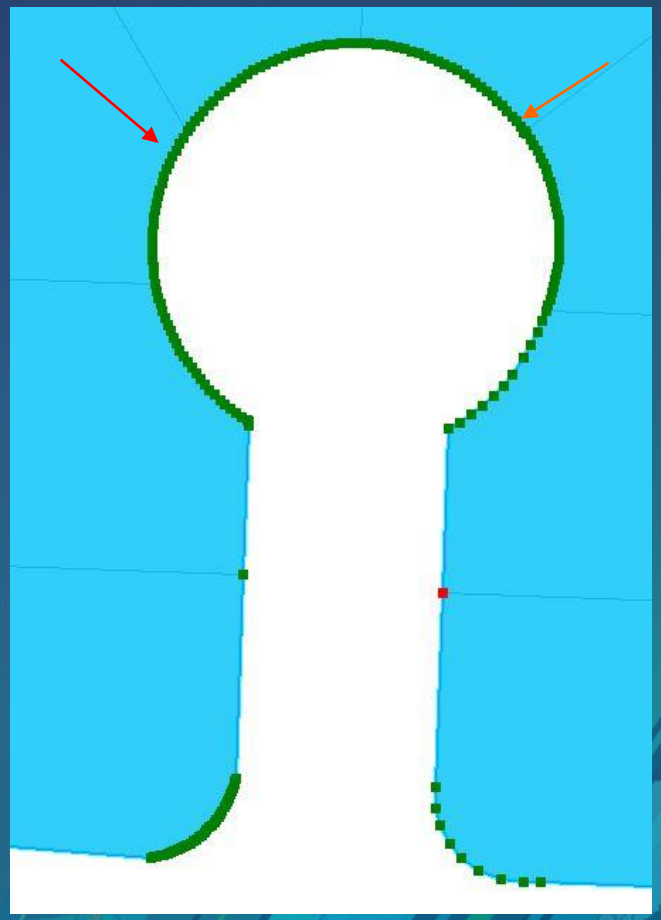
- Aim for 100% Data Accuracy from Data Conversion Effort
- Either Populate or Drop Empty Fields
- Minimize Data Model / Schema Changes
  - More Complexity in a Versioned Geodatabase
- Consolidate GDBs
  - Avoid Creating GDB Per Geographic Locations / Regions
- Extra Vertices have Performance Impact

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

Key Data Conversion Considerations

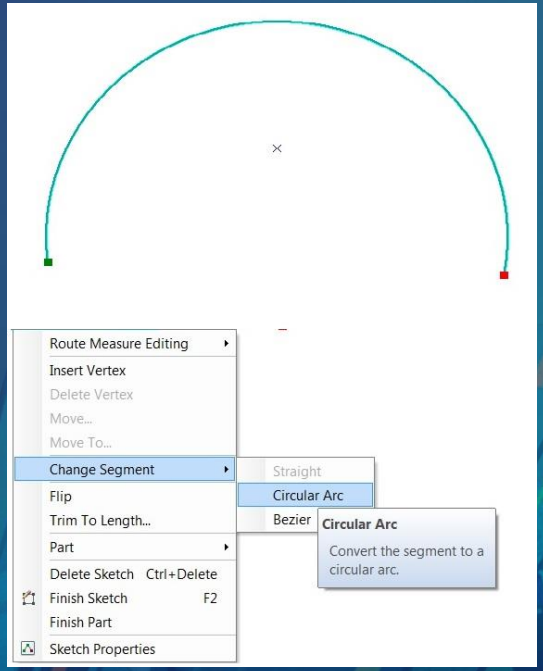
- Extra Vertices Introduced by:
  - Conversion process involving CAD systems
  - Geometric Network Creation with Snapping ON Option
  - Conversion/Update Methods
- Develop Adequate QA/QC Methods and Procedures
- Additional Data Reviewer Checks
  - Duplicate/Invalid Geometries
  - Orphan Related Records
  - Connectivity Check
  - Etc.

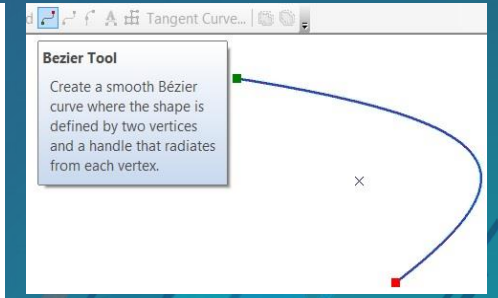
Feature Class/Layer	Total Number of Features	Total Number of Vertices with Geometric Network Snapping	Total Number of Vertices without Geometric Network Snapping	Increase in number of vertices %
Secondary Overhead Conductor	3,712	12,953	7,841	65.20%

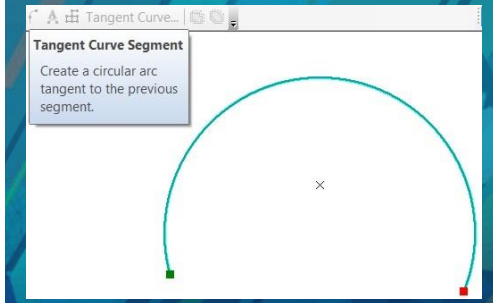


Build Geodatabase – Best Practices

- 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

OWNER	TABLE_NAME	NUM_ROWS	BLOCKS	AVG_ROW_LEN	TO_CHAR(LAST_ANALYZED,'MON/DD/YYHH24:MI:SS')
ARCFM	SDE_BLK_3	1145079	1190598	12	APR/19/16 05:38:10
ARCFM	BK_SDE_LOGFILE_DATA	4231223	8773	10	APR/19/16 05:32:14

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

esri THE SCIENCE OF WHERE™

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.**

Enterprise Geodatabase Vs File Geodatabase		
Key Characteristics	**Enterprise Geodatabase**	**File Geodatabase**
Description	A collection of various types of GIS datasets held as tables in a relational database.	A collection of various types of GIS datasets held in a file system folder.
Number of users	Multiuser: many readers and many writers	Single editor and can support multiple readers.
Storage format	Oracle	Each dataset is a separate file on disk. All the datasets that belong to one geodatabase are contained in a single folder.
	Microsoft SQL Server	
	IBM DB2, Informix	
	PostgreSQL	
	SAP HANA	
Size limits	Size is controlled by the DBMS	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.
Versioning support	Fully supported across all DBMSs	Does not support versioning workflows
	Note:SAP HANA geodatabases do not support the traditional versioning type, but only Branch Versioning.	
Platforms	The DBMS are supported on multiple operating systems. Consult the system requirements for the full list.	Cross-platform.
Security and permissions	Managed by the DBMS.	Managed by the operating system.

- **Generally more than one Geodatabase is required!**

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.

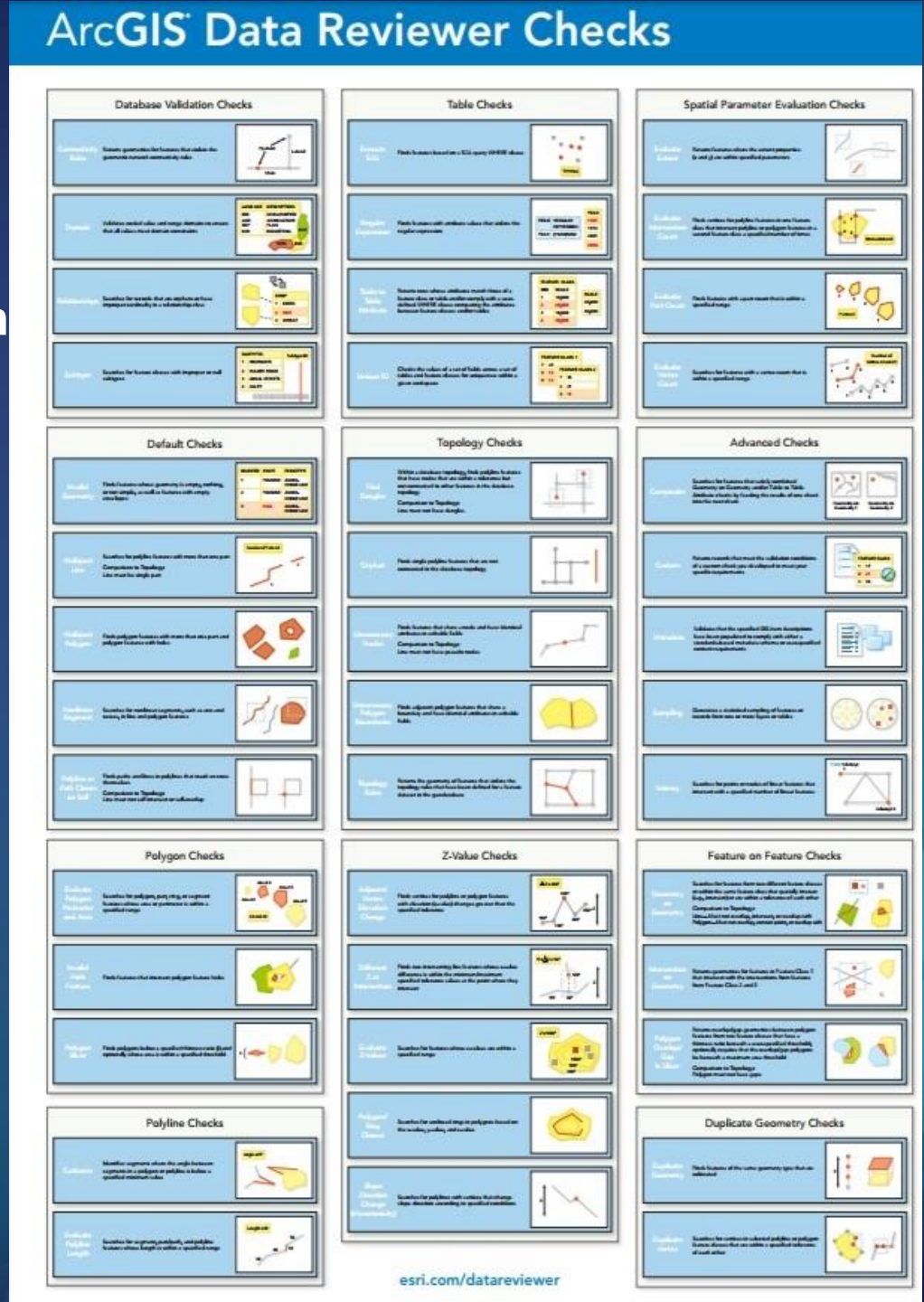
Capture, Load and Maintain Data Accurately!

Continue….

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

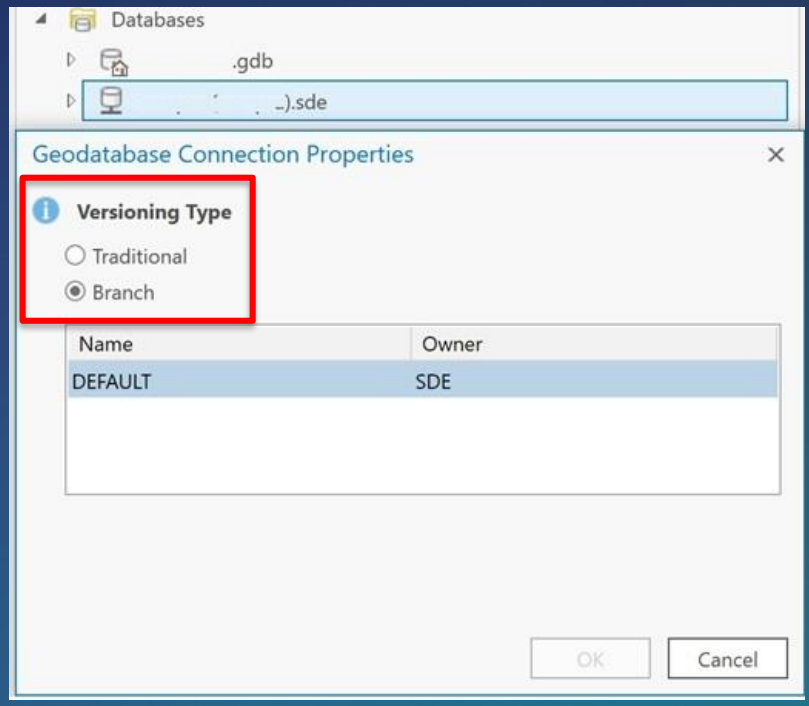
Capture, Load and Maintain Data Accurately!

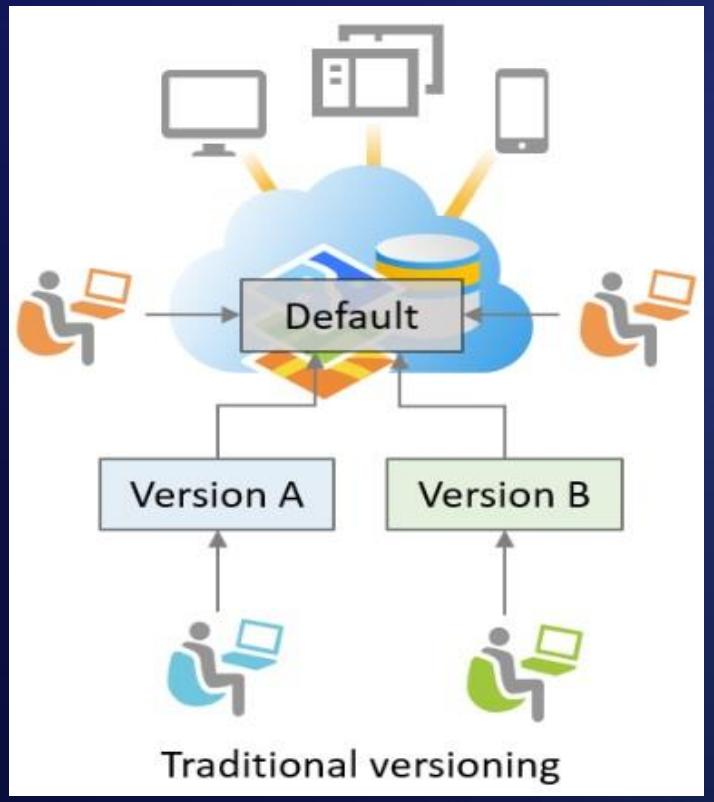


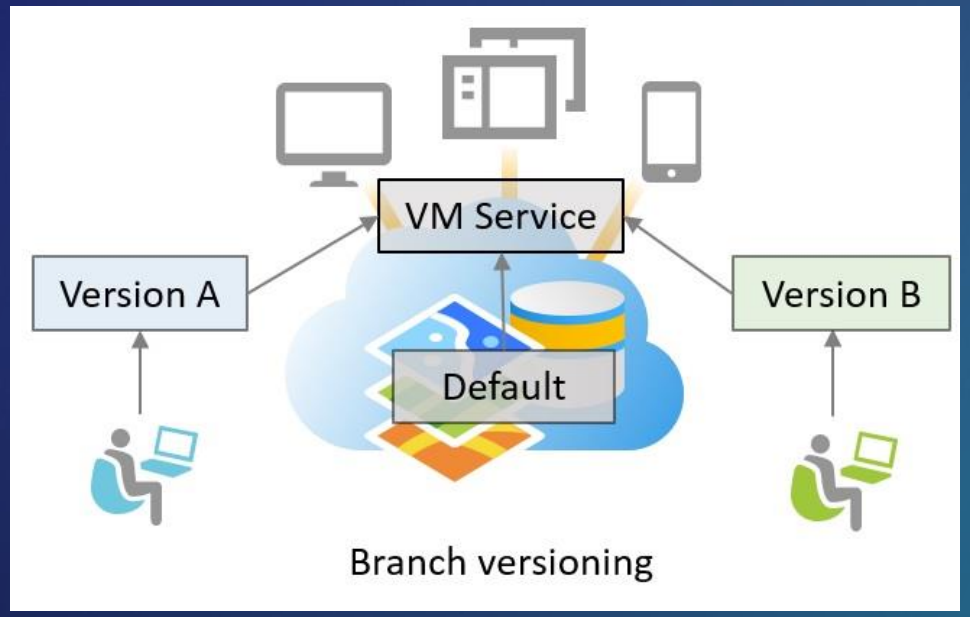
Geodatabase Multiuser Workflows

Which Versioning Model to Choose?

- **Traditional Versioning Vs Branch Versioning**

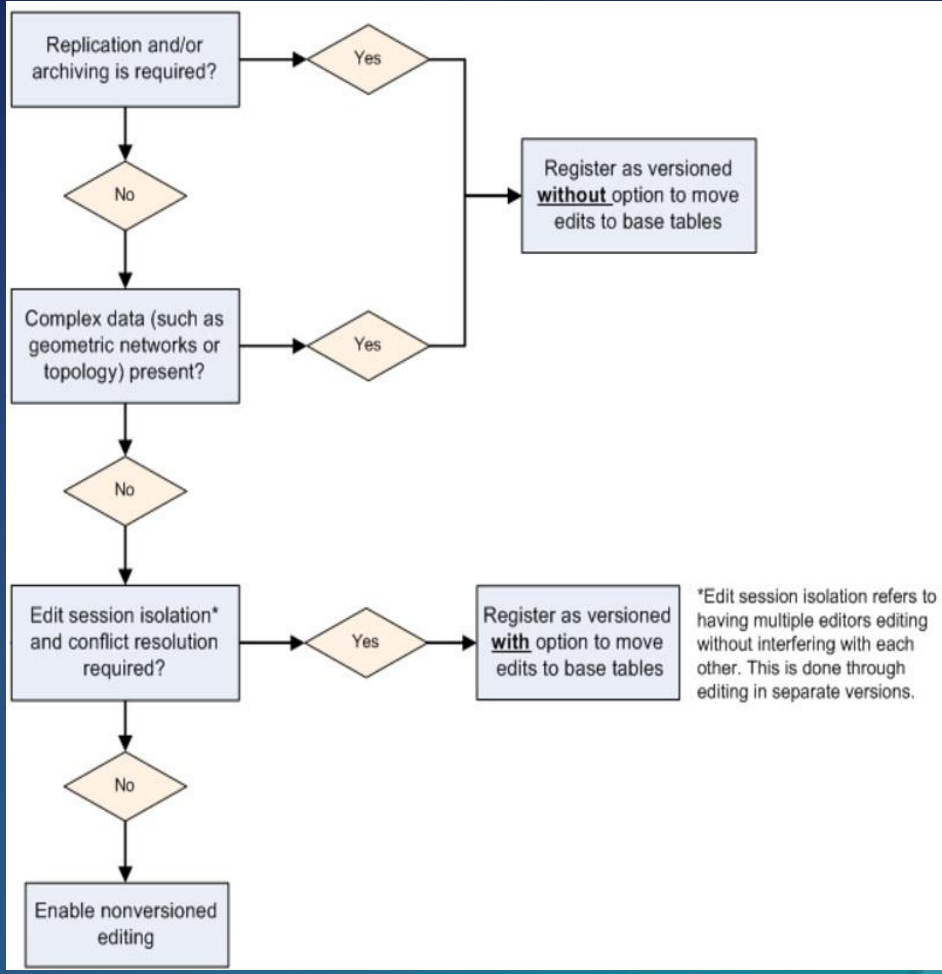


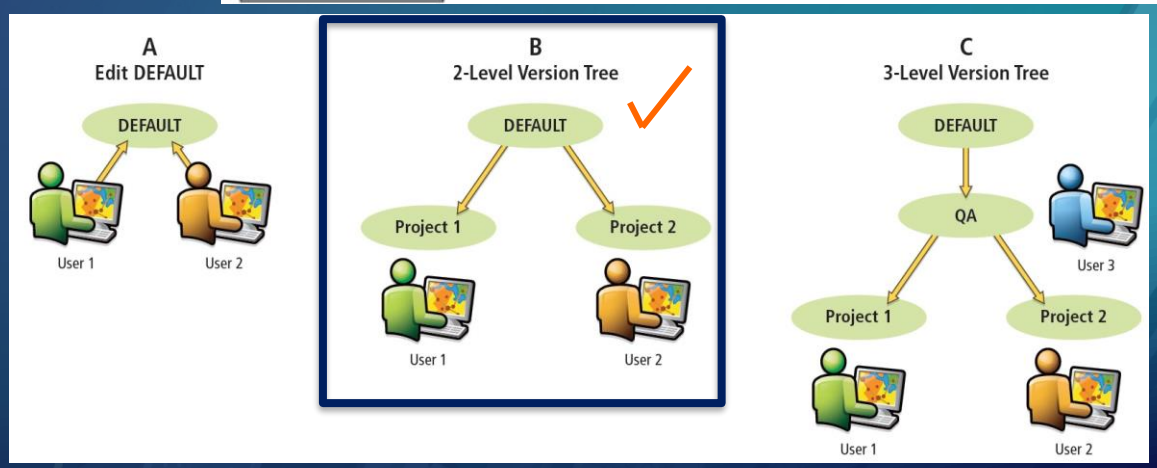




Geodatabase Multiuser Workflows

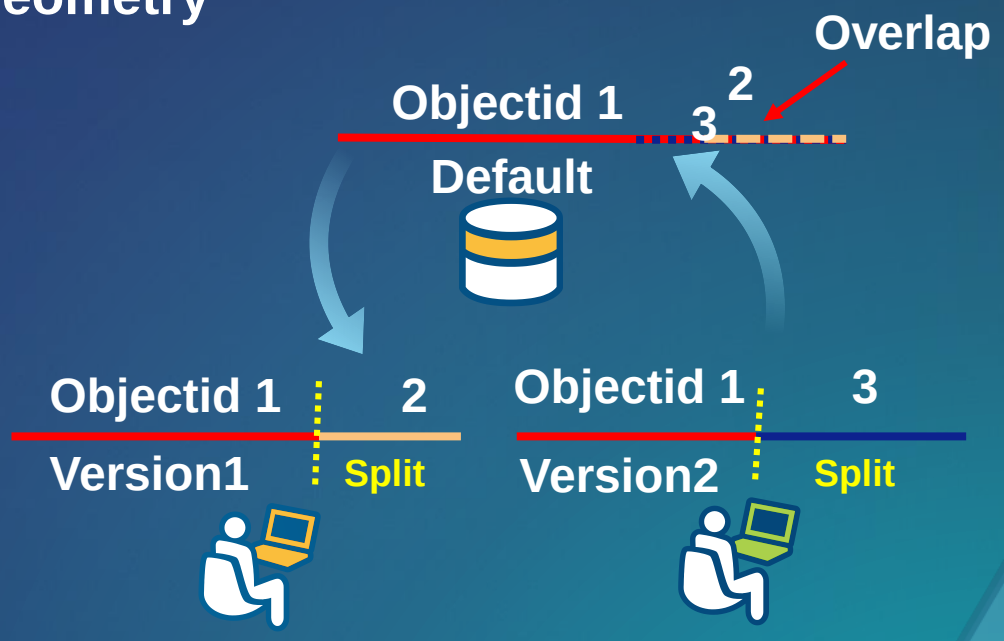
- Versioning Structure
  - Recreate the Version after each Post for 3 Version levels
  - *Problem: Unexpected conflict observed during reconcile*
  - *http://support.esri.com/en/technical-article/000012321*
  - Move Edits to Base For Simple Feature Classes
    - Ability to Easily Share the data with third-party applications
- Estimated Edit Volumes, Version Durations
- Conflict Resolution Mechanisms





Geodatabase Workflows – Key Considerations

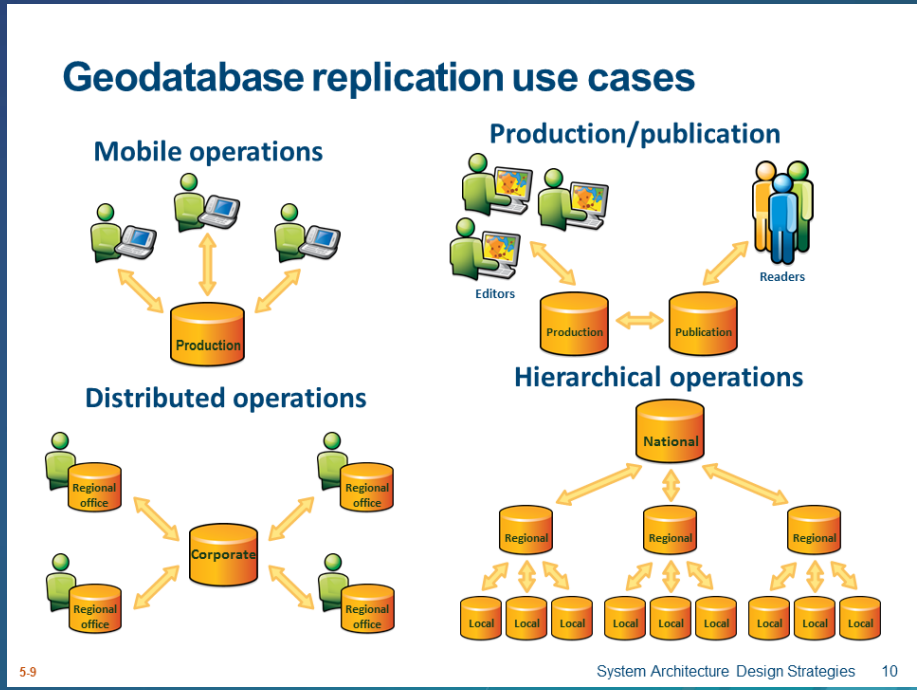
- Conflict for Split operation could introduce duplicate geometry
- Plan Bulk Loading / Mass Update
- For Regular Data Load, Consider Truncate Vs Delete
- Execute Batch Processes During Non-Business Hours
- Geodatabase Replication - One-Way Vs Two-Way
- Archiving
- Editor Tracking



- 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

OWNER	TABLE_NAME	NUM_ROWS	BLOCKS	AVG_ROW_LEN	TO_CHAR(LAST_ANALYZED)
GIS	WSERVICECONNECTION	117,726	12,139	584	APR/10/18 15:32:34
GIS	WSERVICECONNECTION_H	140,242	8,697	528	APR/10/18 15:28:28
GIS	WSERVICECONNECTION_H2	118,057	12,458	605	APR/16/18 08:12:27
GIS	WSERVICECONNECTION_H1	1,653,331	155,838	784	APR/10/18 15:30:13
GIS	S4998_IDX$	1,653,331		52	APR/24/18 22:07:53

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

```
[W 50:00.764] Long:          1
[W 50:00.764] Long:          1
[R 50:00.764] Long:          0
=========================================
[W 50:00.764] Command:       TableListTables
[W 50:00.764] Long:          4
[R 50:48.765] Long:          0
[R 50:48.765] Long:          7835
[R 50:48.765] Dynamic_Str:   "SDE.MV_LPA_BAK"
[R 50:48.765] Long:          1
```

Testing and Tuning

Important Step Before Going Live!

esri® THE SCIENCE OF WHERE™

Testing

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

Mxdperfstat Before Tuning in Production

mxdperfstat

1/9/2018 2:58:46 PM
Master .mxd
layerCount= 6
WGS_1984_Web_Mercator_Auxiliary_Sphere
esriMeters
X= -9,233,816.99 Y= 4,869,583.83 width= 1200 height= 1000

Map Display Performance (sec) for each scale

Scale	Refresh Time(sec)	VisibleLayers
1,000,000	1.30	3
500,000	1.56	3
100,000	1.25	4
50,000	53.83	5
25,000	50.40	5
12,000	48.44	5
5,000	47.14	5
2,500	51.52	5
1,000	44.78	5
500	48.46	5

Mxdperfstat After Tuning in Production

mxdperfstat

1/12/2018 8:31:01 AM
Master .mxd
layerCount= 6
WGS_1984_Web_Mercator_Auxiliary_Sphere
esriMeters
X= -9,233,816.99 Y= 4,869,583.83 width= 1200 height= 1000

Map Display Performance (sec) for each scale

Scale	Refresh Time(sec)	VisibleLayers
1,000,000	1.24	3
500,000	2.10	3
100,000	.67	4
50,000	1.14	5
25,000	1.37	5
12,000	1.10	5
5,000	.81	5
2,500	.92	5
1,000	.69	5
500	.83	5

Performance Tuning Results for an Editing workflow:

S.No	Task	Performance - Before Tuning in Seconds	Performance - Before Tuning in Seconds	Performance Gain in %
1	Opening ArcMap	46.22	15.41	199.94%
2	Open MXD (Master.mxd)	28.79	28.06	2.60%
4	Zoom to selected features	8.46	5.36	57.84%
5	Zoom to 5000	7.25	3.46	109.54%
6	Zoom out to 1000	4.93	2.25	119.11%
7	Zoom out to 500	5.32	0.91	484.62%
8	Start Editing	11.95	7.66	56.01%
9	Place a Structure (1 Pole with Push brace)	9.63	0.93	935.48%
10	Insert a Trans Line	7.85	1.11	607.21%
11	Stop Edits	4.73	1.50	215.33%

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

Operation System's Internal Memory Allocation to RDBMS	
OS Default Memory 4KB/ Page	OS Huge / Large Memory 2MB per Page
RDBMS 32GB / 4KB = 8,388,608 Pages	RDBMS 32GB / 2MB = 16,384 Pages

```
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

MXDPerfstat Results				
Scale	Feature Class/Layer	Display in Seconds - Before RDBMS tuning	Display in Seconds - After RDBMS Tuning	Performance gain in %
5,000,000	24-36 Months	8.91	0.28	3082.14%
2,500,000	Lease Hold Areals (LSE)	51.08	3.89	1213.11%
2,500,000	0-3 Months	0.58	0.19	205.26%
2,500,000	3-6 Months	0.72	0.19	278.95%
2,500,000	6-9 Months	0.78	0.23	239.13%
2,500,000	9-12 Months	0.67	0.3	123.33%
2,500,000	12-18 Months	0.67	0.25	168.00%
2,500,000	18-24 Months	0.8	0.25	220.00%
1,000,000	Surface (FEE)	13.89	0.61	2177.05%
1,000,000	Mineral & Surface	14.11	0.11	12727.27%
1,000,000	SR Units	13.06	0.09	14411.11%
1,000,000	AMI & OA (CTR)	14.86	1.42	946.48%
1,000,000	Units (CTR)	13.8	0.56	2364.29%
1,000,000	Farm In (CTR)	14.08	0.09	15544.44%
500,000	Surface Lease/Easements (MLS)	0.41	0.06	583.33%
500,000	Surface (FEE)	5.45	0.5	990.00%
500,000	Overriding Royalty	6.33	0.89	611.24%
500,000	Units (CTR)	5.17	0.14	3592.86%
500,000	Interstates	35.11	0.41	8463.41%
250,000	State Highways	35.97	0.2	17885.00%
100,000	TX Blocks	0.22	0.12	83.33%
100,000	TX Abstract	0.72	0.33	118.18%
100,000	Roads	36.98	0.12	30716.67%
50,000	TX Abstract	0.44	0.14	214.29%
25,000	Roads	47.98	0.08	59875.00%

Maintenance

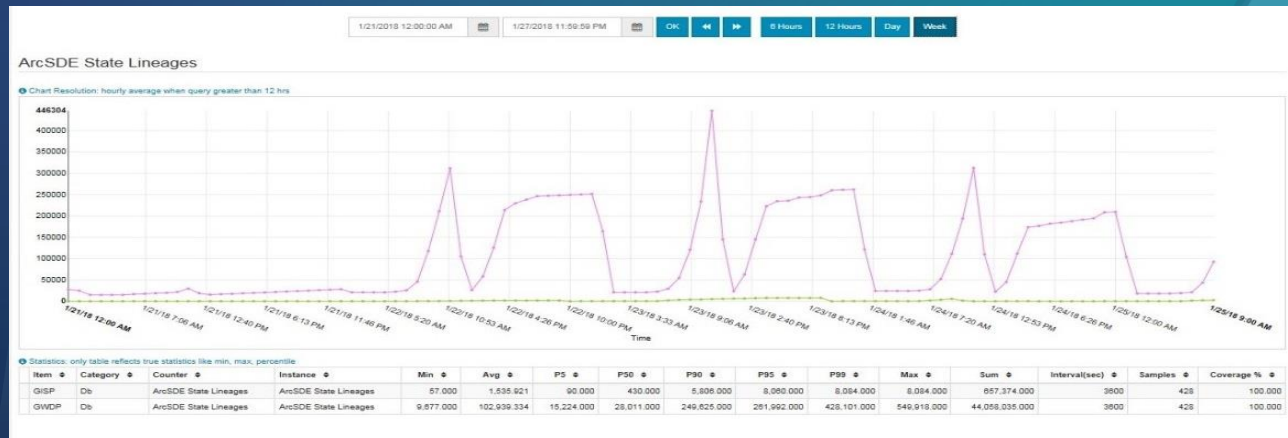
Geodatabase Maintenance - Strategies

- 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



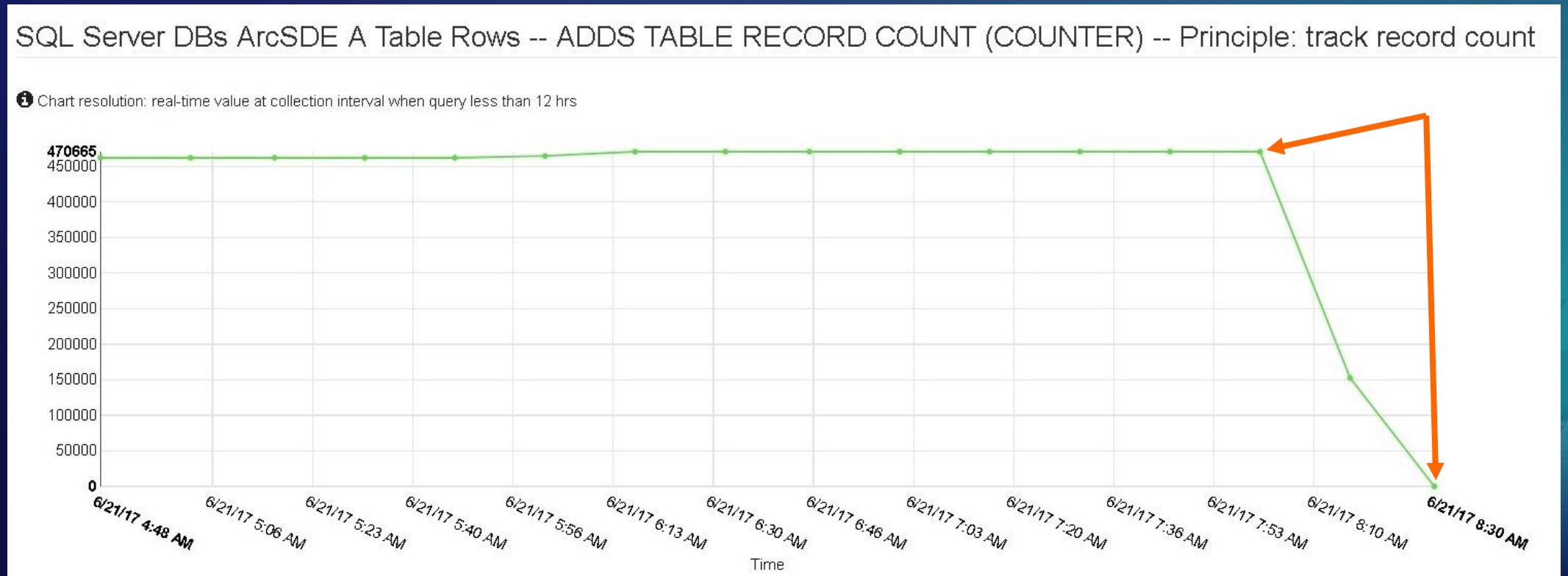
Item	Category	Counter	Instance	Min	Avg	P5	P50	P90	P95	P99	Max	Sum	Interval(sec)	Samples	Coverage %
GISP	Db	ArcSDE State Lineages	ArcSDE State Lineages	57.000	1,535.921	90.000	430.000	5,806.000	8,060.000	8,084.000	8,084.000	657,374.000	3600	428	100.000
GWDP	Db	ArcSDE State Lineages	ArcSDE State Lineages	9,877.000	102,939.334	15,224.000	28,011.000	249,625.000	281,992.000	428,101.000	549,918.000	44,058,035.000	3600	428	100.000

Automate the Daily Maintenance Process!

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**

```
                    Versioning statistics
=========================================
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
```



Geodatabase Maintenance – Best Practices

- **Remove Geoprocessing (GP) History**
  - ***How To: Automate the process of deleting geoprocessing history***
  - ***http://support.esri.com/technical-article/000011751***
- **Disable (GP) History for Scripts**

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

Python Tasks	Before Deleting GP History in Seconds	After Deleting GP History in Seconds	Performance Improvement in %
Create Version	400.74	16.3	2358.53%
Delete Version	571.23	14.17	3931.26%

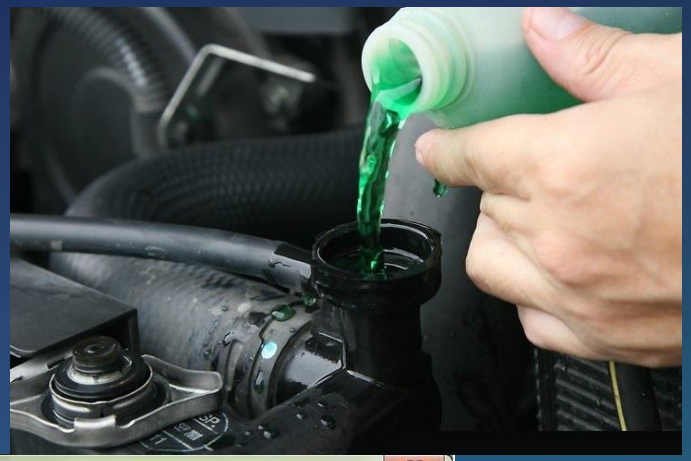
Geodatabase Maintenance – Best Practices

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

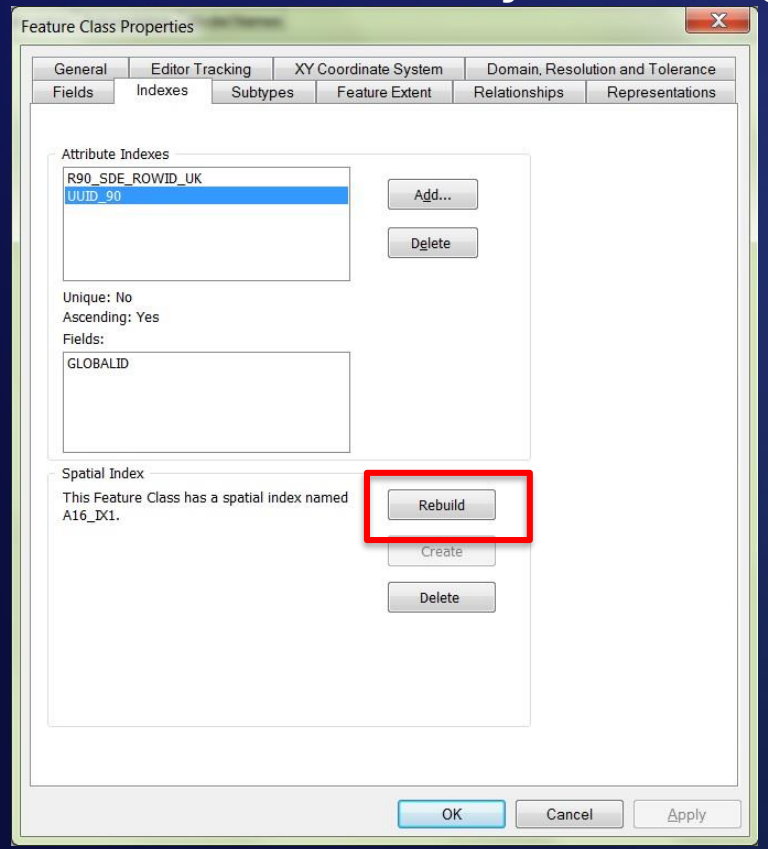
OWNER	TABLE_NAME	NUM_ROWS	BLOCKS	AVG_ROW_LEN	TO_CHAR(LAST_ANALYZED,'MON/DD/YYHH24:MI:SS')
ARCFM	EDITEDFEATURESTRACKING	12330909	122954	140	APR/23/16 23:44:17
ARCFM	EDITEDGRIDS	12964827	32969	32	APR/23/16 23:47:55
SDE	ARCSDEUSERLOG	56558633	259246	62	APR/24/16 12:28:25

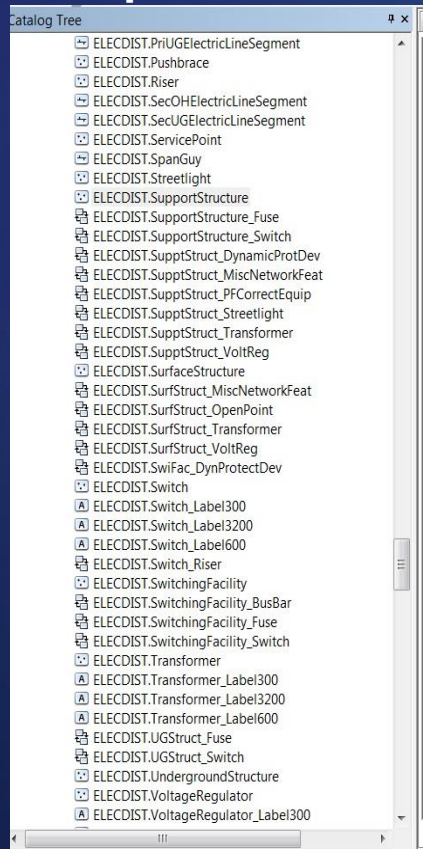
OWNER	TABLE_NAME	NUM_ROWS	BLOCKS	AVG_ROW_LEN	TO_CHAR(LAST_ANALYZED,'MON/DD/YYHH24:MI:SS')
SDE	GDBM_RECONCILE_HISTORY	4759429	50484	72	AUG/12/17 12:40:34
ARCFM	USAGEINFORMATION	6377236	221057	39	AUG/10/17 15:59:14

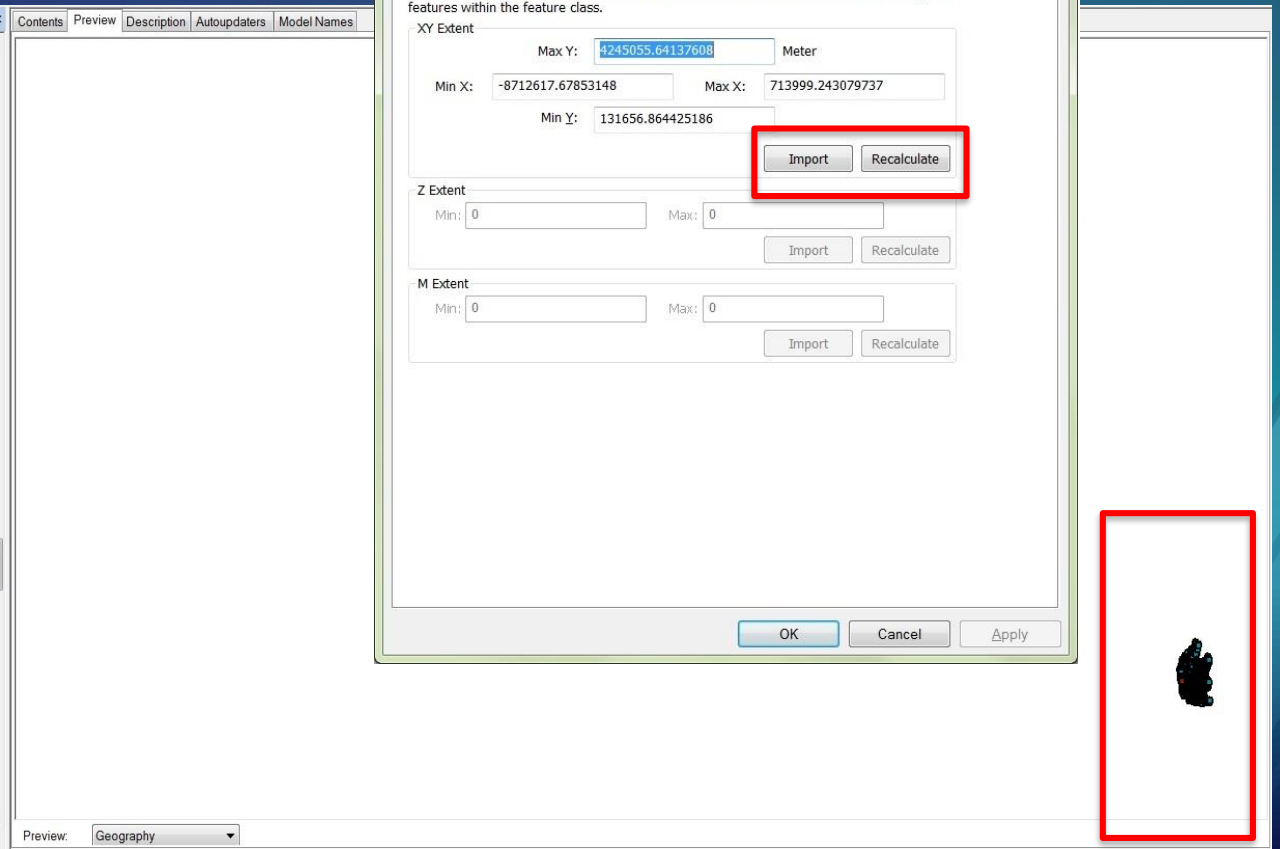
Geodatabase Maintenance – Best Practices



- **Every 3 - 6 Months:**
  - **Fix the Feature Class Extent First**
  - **Followed by Rebuilding the Spatial Index**







Geodatabase Maintenance – Best Practices

- **Fix the Non-Empty Feature Classes with No Spatial Index**

```
Exec DBMS_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);
TABLE_NAME                                   GSIZE1     GSIZE2     GSIZE3
-------------------------------------------- ---------- ---------- ----------
COATING                                               0          0          0
CPBOND                                                0          0          0
CPCABLE                                               0          0          0
CPGROUNDBED                                           0          0          0
CPANODE                                               0          0          0
LINECROSSING                                          0          0          0
MARKER                                                0          0          0
PIPEEXPOSURE                                          0          0          0
TIEINMETHOD                                           0          0          0
DOCUMENTPOINT                                         0          0          0
INJECTION                                             0          0          0
PIGROUTE                                              0          0          0
PIPEDEPTH                                             0          0          0
RL_SITE                                               0          0          0
PIPEJOIN                                              0          0          0
PIPELENGTH                                            0          0          0
WELL                                                  0          0          0
LEAK                                                  0          0          0
DOT_CLASS_PREVIOUS_RDETAILS                           0          0          0
ALIGNMENT_SHEETS                                      0          0          0
EXTERNAL_COATING_RDETAILS                             0          0          0
MISC_FITTING_DETAILS                                  0          0          0
ALIG_SHT_CROSS_REF_RDETAILS                           0          0          0
CLOSURE_DETAILS                                       0          0          0
HCA_PREVIOUS_RDETAILS                                 0          0          0
INJECTOR_DETAILS                                      0          0          0
ODORANT_RANGE_RDETAILS                                0          0          0
OFFLINECOMPSTATIONPOLYGONS                            0          0          0
PIG_SIGNAL_DETAILS                                    0          0          0
PIR_RDETAILS                                          0          0          0
RIGHT_OF_WAY_RDETAILS                                 0          0          0
RIVER_WEIGHT_RDETAILS                                 0          0          0
SHEET_NOTE_RDETAILS                                   0          0          0
TAP_DETAILS                                           0          0          0
PIPE_SEG_PIR_BUFF                                     0          0          0
MAOP_CALC_RDETAILS                                    0          0          0
```

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 – Through out the Day
  6. Only Reconcile All versions (>100 versions - Parallel Reconcile)
  7. Update Database Statistics - Optional

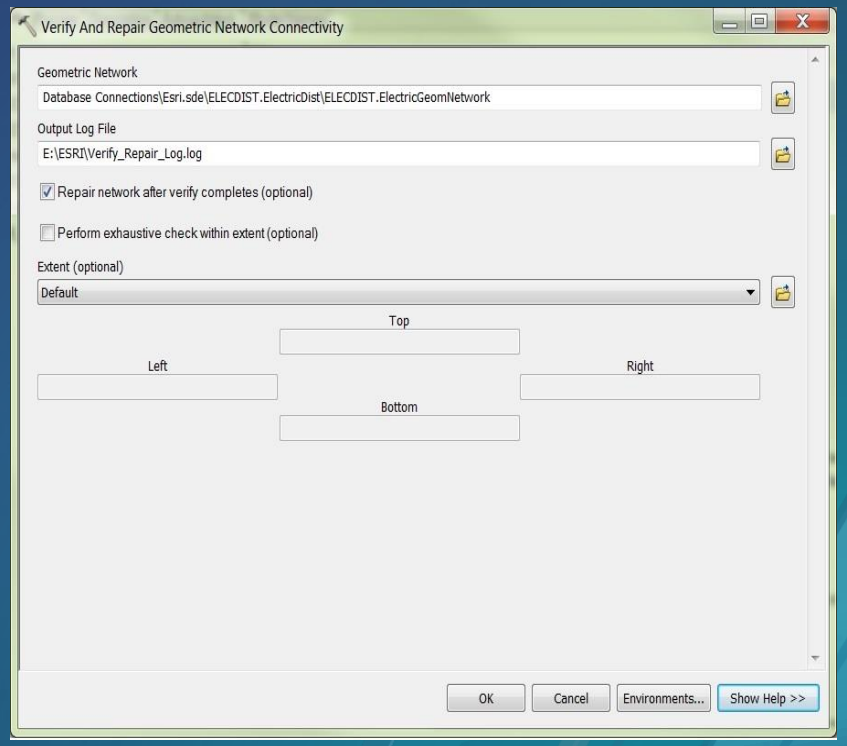
Continue…..

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.

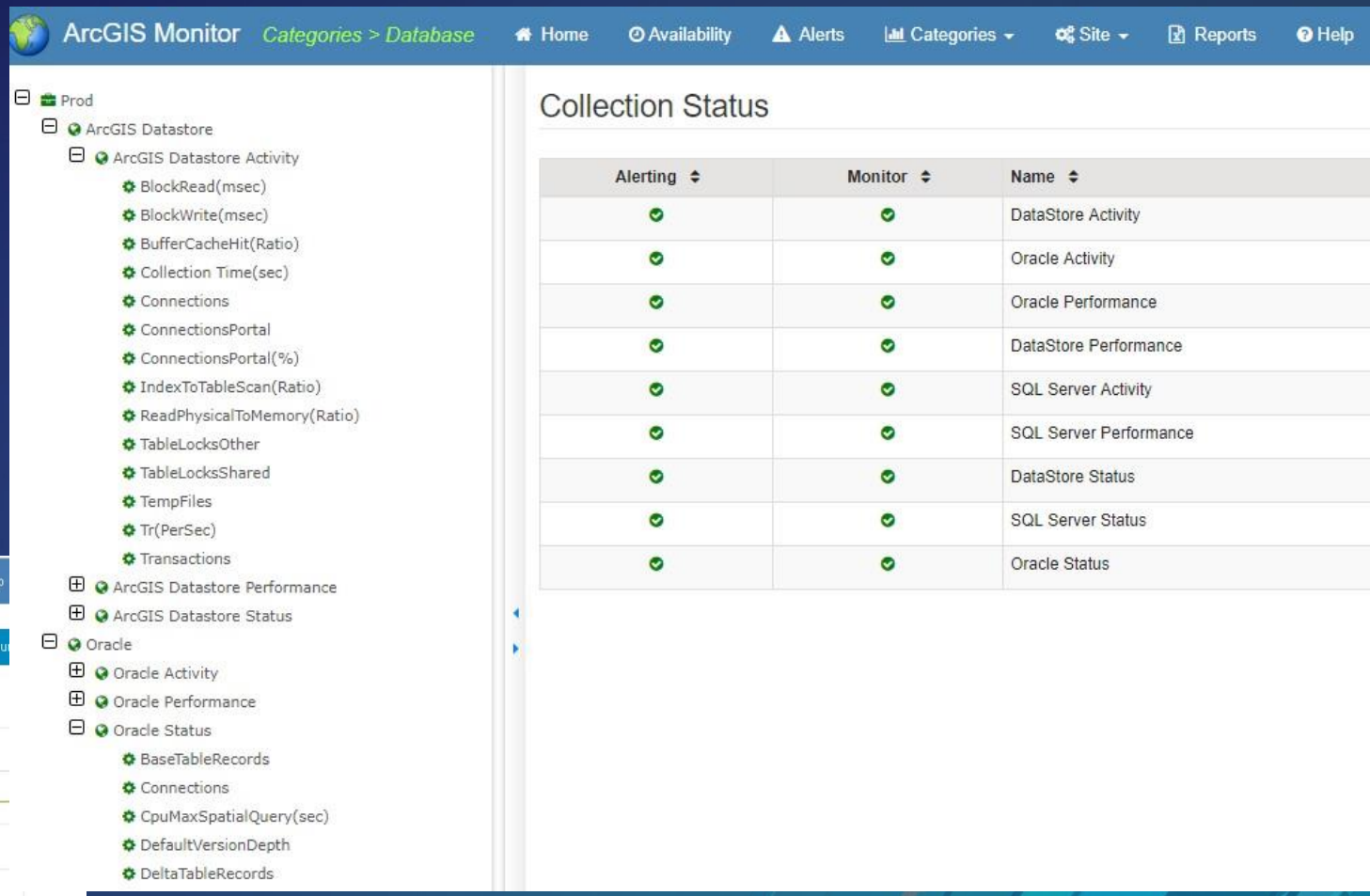


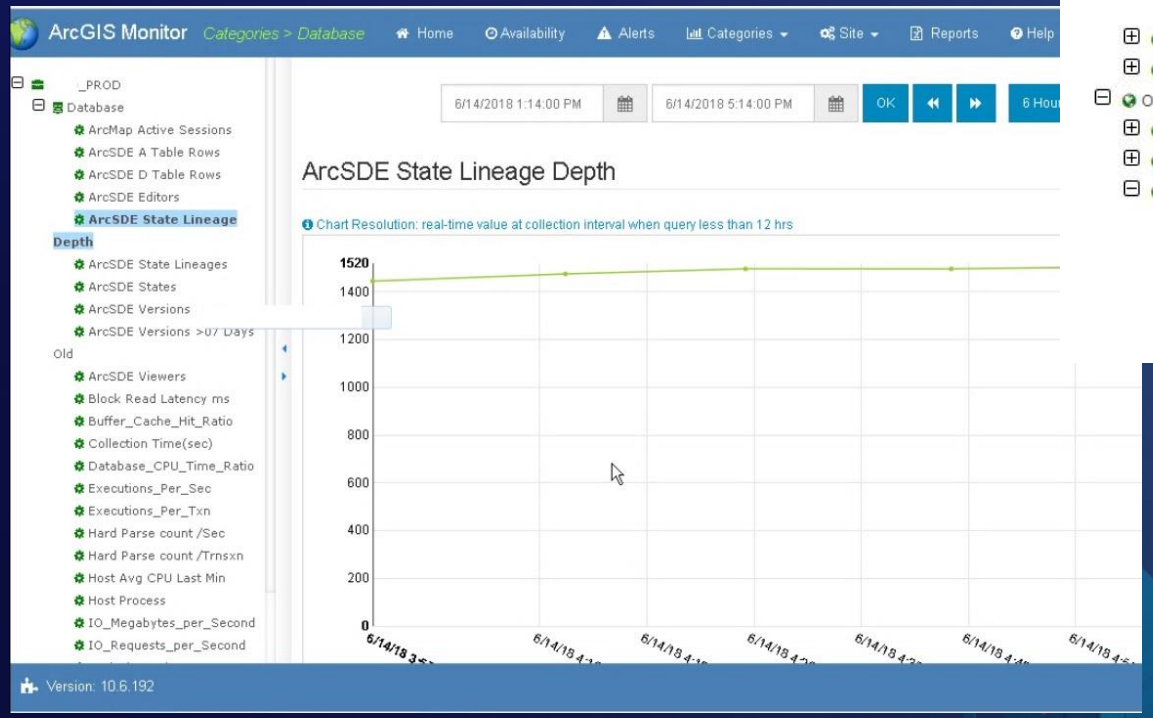
Caution: Individual Users Should not run it. Generates larger number of delta table records and if someone accidently run this tool under SDE.Default version directly, it can produce more conflicts.

Monitor

ArcGIS Monitor

- Add RDBMS Queries
- Monitor Key Performance Indicators
- Keep 15 Minutes Sampling Interval





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
  - https://www.esri.com/en-us/arcgis/products/arcgis-monitor/overview
- 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

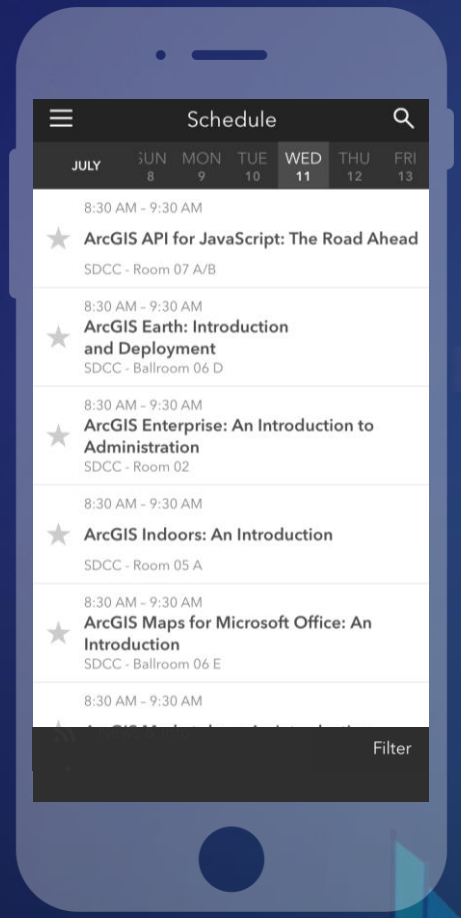
See Us Here

WORKSHOP	LOCATION	TIME FRAME
• Geodatabase: An Introduction	• SDCC - Room 05 A	• Wednesday, July 11 2:30 pm - 3:30 pm
• Managing Your Distributed Geodatabase	• SDCC - Room 14 A	• Wednesday, July 11 4:00 pm - 5:00 pm
• Esri Best Practices: Implementing an Enterprise Geodatabase	• SDCC - Room 05 B	• Thursday 7/12/2018 10:00 AM - 11:00 AM

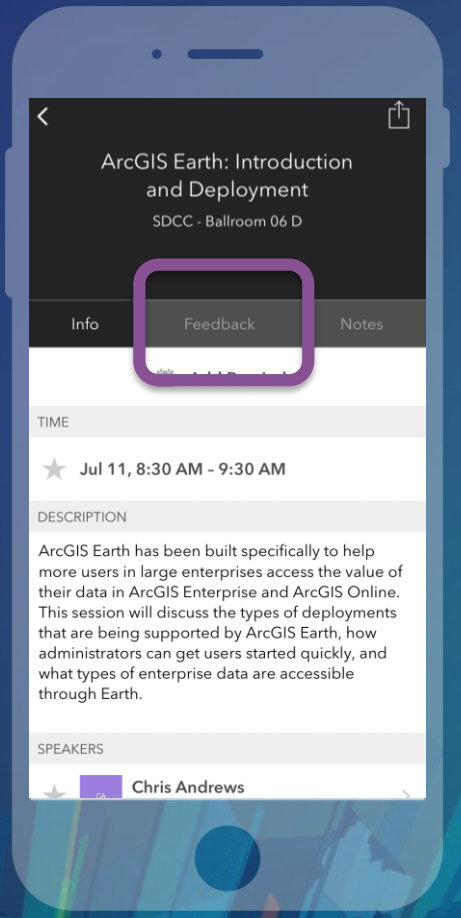
Please Take Our Survey on the App

Download the Esri Events app and find your event

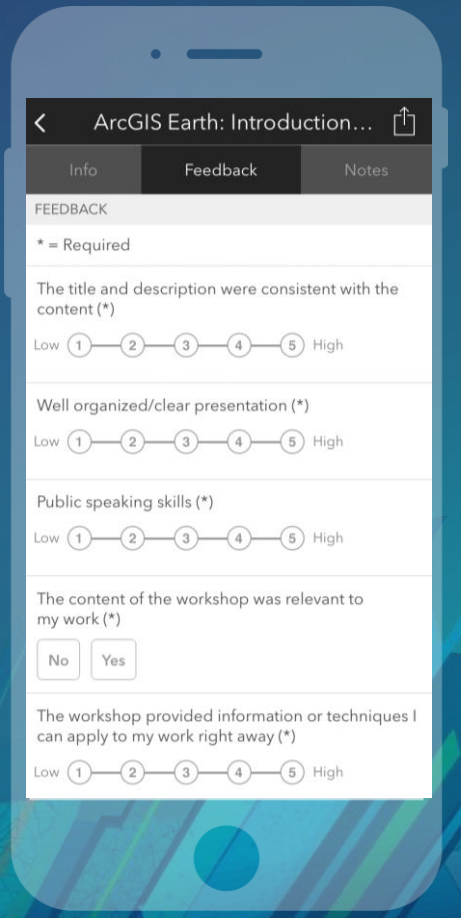
Select the session you attended



Scroll down to find the feedback section



Complete answers and select “Submit”



Thanks!

esri | THE SCIENCE OF WHERE