



Design a Geodatabase

Rasu Muthurakku, Enterprise Solutions Architect / Consultant



SEE
WHAT
OTHERS
CAN'T

THE
SCIENCE
OF
WHERE

Agenda

Design a Geodatabase

- Overview
- Motivations
- Design Approach
- Recommendations

Key Considerations, Best Practices, Recommendations and Lessons Learned!



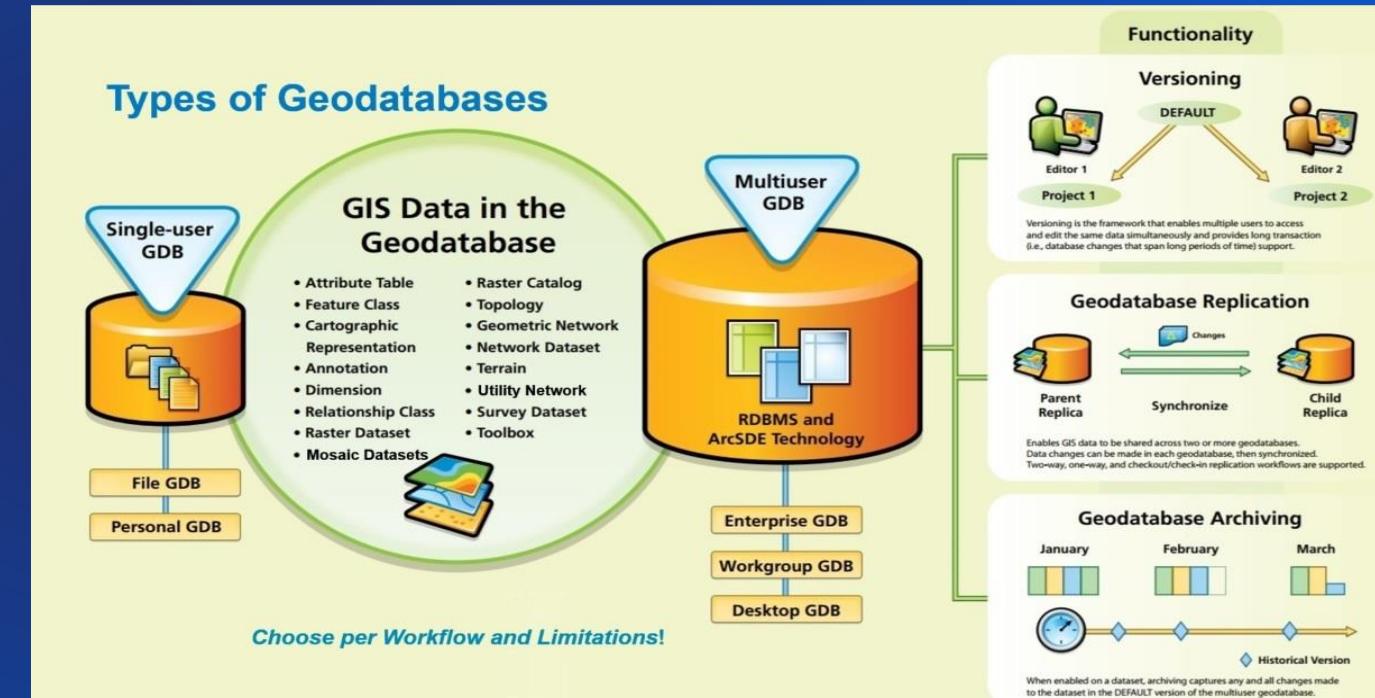
Overview

Geodatabase (GDB)

- Collection of Geographic Datasets of Various Types Stored in:

- Common File System Folder
- Microsoft Access Database
- Multiuser RDBMS * / ArcSDE
 - Oracle
 - Microsoft SQL Server
 - PostgreSQL
 - IBM DB2, Informix
 - SAP Hana 2.0 SPS02

- Native Data Structure for ArcGIS
- Primary Data Format Used for Editing and Data Management



* RDBMS – Relational Database Management System

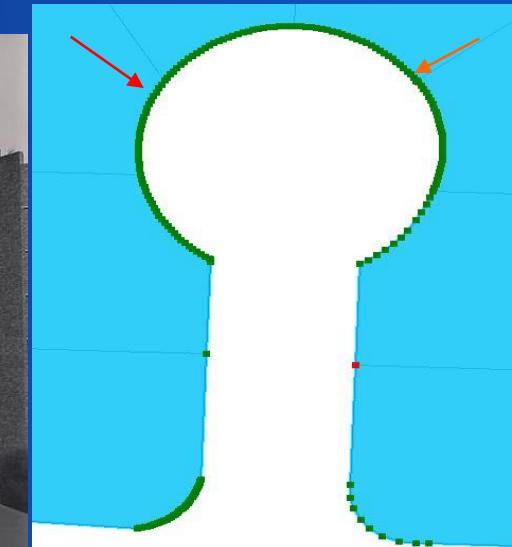
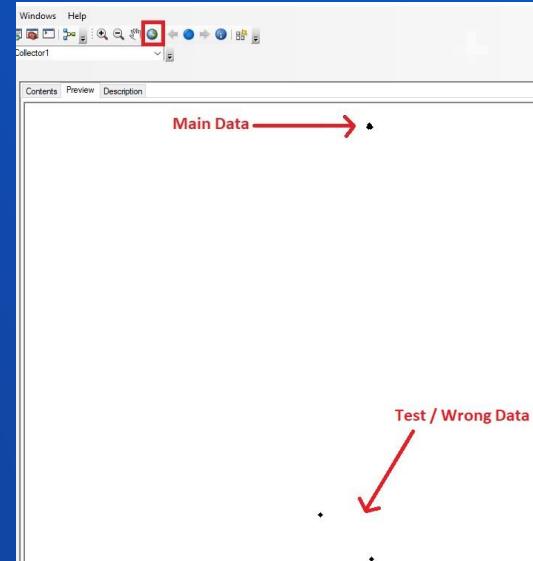
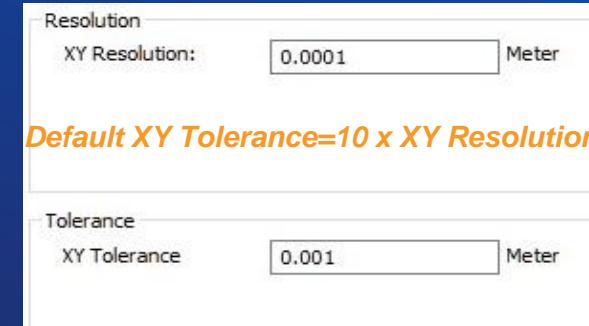
Motivations

Why?

Efficient Implementation

- Data Model Design – Don'ts
 - *Non-default XY Resolution / Tolerance*
 - *Multiple Projections*
 - *Objectid based Relationship Classes*
 - *etc.*
- Data Conversion – Don'ts
 - *Extra Vertices*
 - *Insufficient QA/QC*
 - *Unusual Feature Extent*
 - *Loading static Raster data into GDB*
 - *etc.*

Avoid expensive rework!



Increase Performance and Scalability

Mitigate the risk with proper system capacity & maintenance



- **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 # of CPUs
 - Roles and responsibilities
 - etc.

S.No	Display Scale	Layer Name	Before Maintenance- Display in Seconds	After Maintenance - Display in Seconds	Performance Improvement in %
1	50,000	RoadCL > 10,000	0.34	0.3	13.33%
2	15,000	Water Mains	0.15	0.09	66.67%
3	10,000	MapLink	1.02	0.08	1175.00%
4	10,000	CH2M_Mains	1.15	0.97	18.56%
5	10,000	Leaders	0.58	0.07	728.57%
6	10,000	Annotation	0.69	0.15	360.00%
7	10,000	Water Mains	0.96	0.6	60.00%
8	10,000	CTParcels	1.2	0.42	185.71%
9	6,000	CH2M_Mains	0.68	0.56	21.43%
10	6,000	Water Mains	0.88	0.52	69.23%
11	1,000	SwingTies	7.74	0.05	15380.00%
12	500	SwingTies	6.12	0.74	727.03%
13	500	Water Mains	0.72	0.24	200.00%
14	500	sbDriveways	0.37	0.16	131.25%

**Few private and orphan versions got skipped from maintenance for a long time and created the bottleneck!*

Failing to prepare is preparing to fail!

Mobile Data Collection - Offline Workflow Planning			
	Data Maintenance Workflow	Short duration Project Work	Ongoing Project Work
Version from which the feature service is published	Default version	Child version	Child version
Offline version is created for each	Downloaded map	User	User
Number of versions created	Many	Few	Few
Latency between offline edits and updates to Default version	Low	High (1 week)	High (Daily)
Maps involved in quality assurance	One map	All maps	All maps
Frequency that offline versions are deleted	Daily	At project completion	Never

Efficient Workflows

Selection of GDBs drives 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!

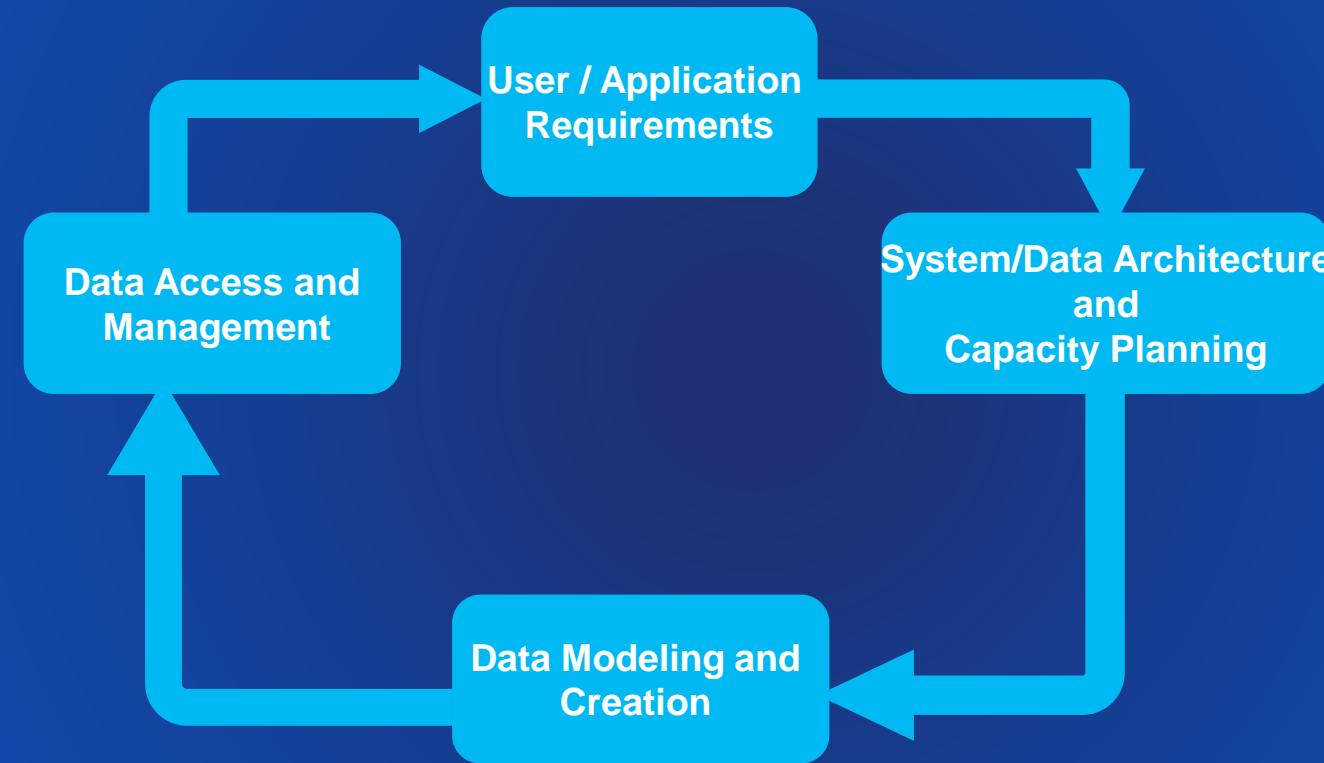


Design Approach

How?

Geodatabase Design – Holistic Approach

Define non-functional requirements also



Accommodate the requirements innovatively!

Architecture and 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!

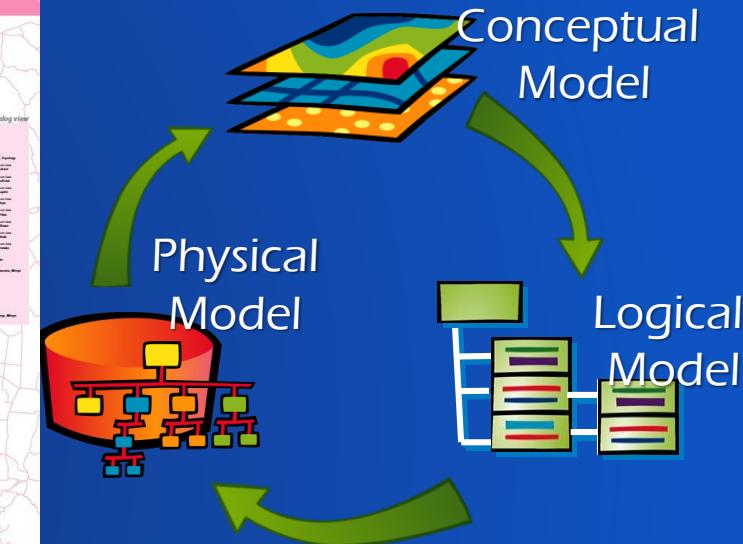
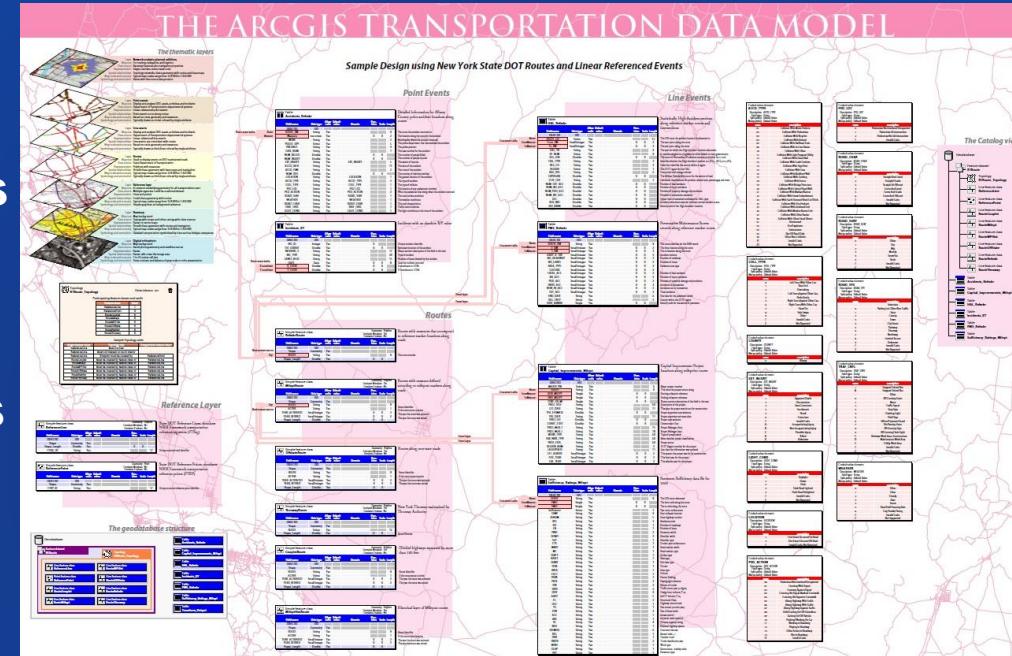
Geodatabase Design – Data Modeling

- **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



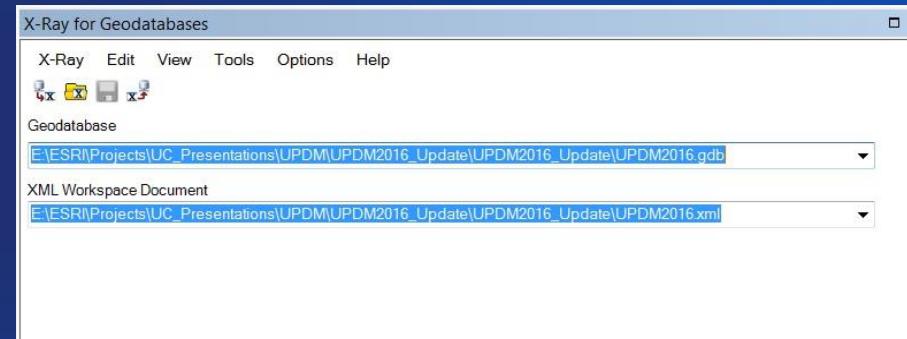
- **Physical Design**

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

Action without adequate design is actually a waste!

Geodatabase Design – Process

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



Geodatabase Access and Management

Data is the brain of GIS nervous system!

- Create roles / groups based on the access level
 - Total access will slow connection time!
- Configure client applications to manage data
- Plan and execute Geodatabase maintenance tasks
- Tune and monitor Geodatabase



Connection Performance					
S.No	User Name	ArcCatalog Connection in Seconds with all access	ArcCatalog Connection in Seconds with reduced privileges	Performance difference in Seconds	Performance improvement in %
1	Rasu	30	12	18	150.00%
2	Andrew	21	12	9	75.00%

Capture, load and maintain data accurately!

Recommendations

What?

Design a Geodatabase – Best Practices

Plan, build and operate an important part of GIS nervous system!

- **Download, modify and use ArcGIS Data Models!**
 - **FAQ: Does Esri have industry-specific Data Models?**
 - <https://support.esri.com/en/technical-article/000011644>
 - **ArcGIS Solutions - Free industry-specific configurations for ArcGIS**
 - <http://solutions.arcgis.com/>
- **Either populate or drop empty feature classes / fields**
- **Fix missing / in-correct Domains and aliases for Fields**

Justify every single Geodatabase element!

Download Links

Below is a list of data models that are available for download.

Related Information

- Address
- Agriculture
- Atmospheric
- Basemap
- Biodiversity
- BroadbandStat
- Building Interior Space
- Carbon Footprint
- Census - Administrative Boundaries
- Defense - Intel
- Energy Utilities (includes ArcGIS MultiSpeak)
- Environmental Regulated Facilities
- Fire Service
- Forest Service
- Forestry
- Geology
- GIS for the Nation
- Groundwater
- Health
- Historic Preservation and Archaeology
- Homeland Security
- Hydro

Geodatabase Design – Best Practices

Data Model impacts storage and performance!

- **Select 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
 - Put only the required 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

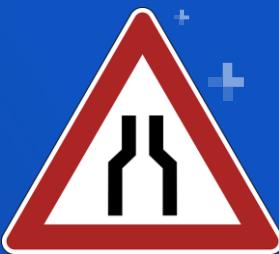


Field Properties	
Alias	CustomerName
Allow NULL values	Yes
Default Value	
Domain	
Length	1073741822

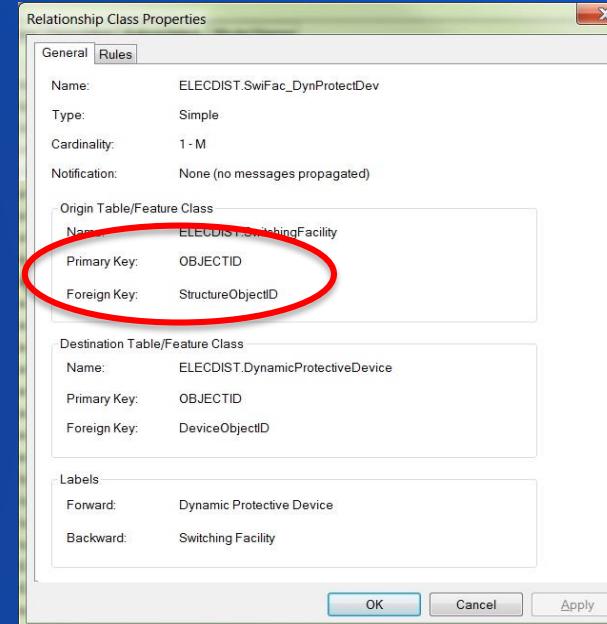
Justify every single Geodatabase element!

Build Geodatabase – Key Considerations

Data is the Backbone of GIS!



- Extra vertices impact performance
- Aim for 100% data accuracy
- 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
 - Use GlobalID

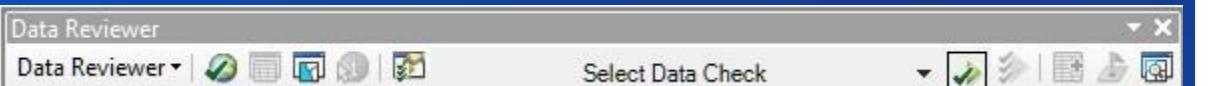


S.No	Display Scale	Layer Name	# of Displayed Features	Number of Vertices - Before Simplify	Number of Vertices - After Simplify	Before Simplification Display in Seconds	After Simplification-Display in Seconds	Performance Improvement in %
1	50,000	Street_1_inch	35,093	105,695	101,060	2.36	0.5	372.00%
2	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%

QA / QC - Recommendations

Data Integrity and Validation!

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

ArcGIS Data Reviewer Checks

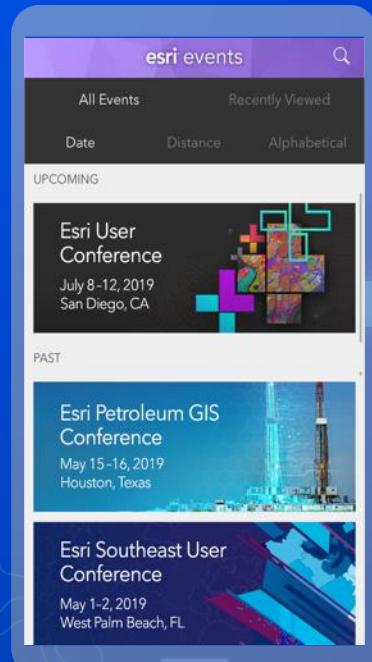
Database Validation Checks		Table Checks		Spatial Parameter Evaluation Checks	
Boundary Value	Ensures geometry for features that include the geometry must contain only one.		Point In Table	Plots features based on a SQL query (100000 rows).	
Domain	Ensures valid value and range domains ensure that all values must start with characters.		negative	Plots features with attribute values that violate the negative expression.	
Exclusive	Ensures for records that are unique on these properties exactly in a relationship table.		Table Value	Ensures when attribute must have a unique value or valid identifier with a unique identifier. Plots features with attribute values that violate the unique identifier.	
Null	Ensures for features with attribute or null values.		Value	Shows the value of a valid field, and shows the value of a given field.	
Default Checks		Topology Checks		Advanced Checks	
Attribute	Ensures when geometry is empty, nothing, or null, it can't have features with empty attribute.		Bad Boundary	Ensures a boundary that violates features that have nodes that are either collinear or intersect other features in the database.	
Feature	Ensures for the features with more than one part. Contains no single part.		Gap	Plots single polygon features that are not connected in the database topology.	
Geometry	Plots polygon features with more than one part and polygon features with holes.		Intersection	Plots features that share vertices and have identical attribute values in attribute fields.	
Relationship	Ensures for features representing both one and many to line and polygon features.		Nonadjacent	Plots adjacent polygon features that share a boundary and have shared attribute values in attribute fields.	
Topology	Ensures polygon features that result in one or more holes.		Overlap	Ensures the geometry of features that violate the topology rules that have been defined for a feature dataset in the geodatabase.	
Polygon Checks		Z-Value Checks		Feature on Feature Checks	
Attribute	Ensures for polygons, rings, and segments for attribute values or presence in a specified range.		Bad Z	Plots features for polygons or poly lines with Z-value elevation that are either greater than the expected elevation.	
Boundary	Plots features that are non polygon features.		Close Z	Plots features that are close to each other with Z-value differences in within a specified tolerance.	
Relationship	Plots polygon features that contain one or more holes.		Same Z	Ensures that features whose attribute are within a specified range.	
Polyline Checks		Feature on Feature Checks		Duplicate Geometry Checks	
Attribute	Ensures segments where the angle between segments is greater than a specified range.		Intersection	Ensures features from one feature dataset that share the same geometry also share a common attribute. Only features in either a reference or target feature class.	
Boundary	Ensures for polygons, boundary, and polygon features where length is within a specified range.		Nonadjacent	Ensures geometry for features in Feature Class 1 that intersect with the nonadjacent features in Feature Class 2 and 3.	
Relationship	Ensures for polygons with common vertices.		Overlap	Ensures overlapping geometry between polygons. Features from one feature dataset that have a common boundary with another feature dataset, especially creates an overlap in the overlapping and non-overlapping areas.	
Duplicate Checks		Z-Value Checks		Duplicate Geometry Checks	
Geometry	Ensures for geometry that are the same geometry type but are different features.		Bad Z	Plots features of the same geometry type but are different features.	
Relationship	Ensures for polygons with common vertices.		Close Z	Plots features that are close to each other with Z-value differences in within a specified tolerance.	

See us here

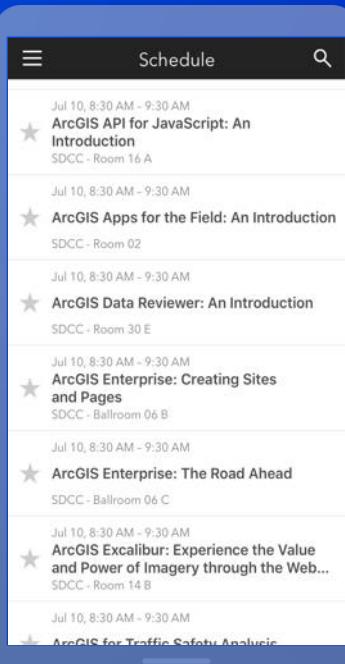
WORKSHOP	LOCATION	TIME FRAME
<ul style="list-style-type: none">• Esri Best Practices: Implementing an Enterprise Geodatabase	<ul style="list-style-type: none">• SDCC - Room 31 A	<ul style="list-style-type: none">• Wednesday 7/10/2019 02:30 PM - 03:30 PM
<ul style="list-style-type: none">• Enterprise Geodatabase: Automating Administration Tasks Using Python	<ul style="list-style-type: none">• SDCC - Expo Demo Theater 04	<ul style="list-style-type: none">• Thursday 7/11/2019 10:00 AM - 10:45 AM
<ul style="list-style-type: none">• Geodatabase: Ensuring Data Quality with Attribute Rules and Contingent Values	<ul style="list-style-type: none">• SDCC - Ballroom 06 E	<ul style="list-style-type: none">• Thursday 7/11/2019 4:00 PM - 5:00 PM

Please Share Your Feedback in the App

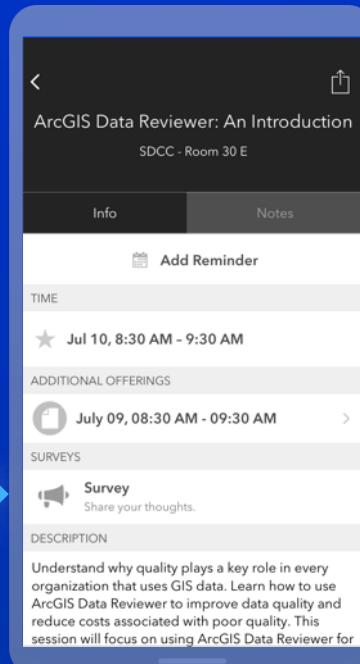
Download the Esri Events app and find your event



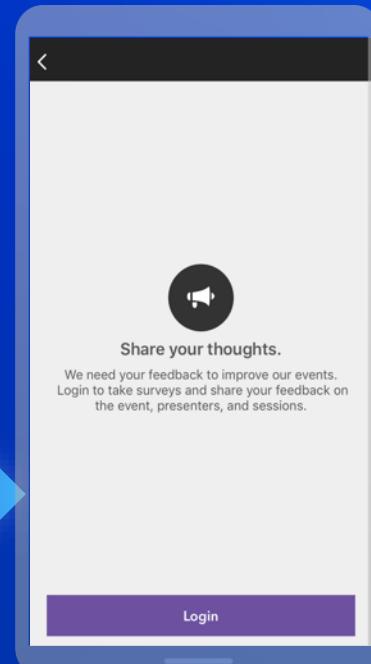
Select the session you attended



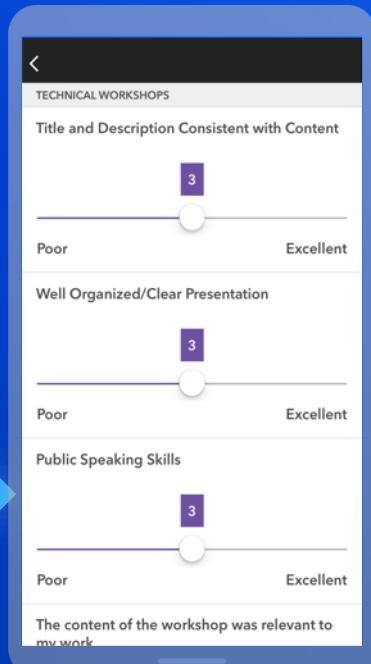
Scroll down to "Survey"



Log in to access the survey



Complete the survey and select "Submit"



Questions and Answers

Contact Info:

Rasu Muthurakku

rasu@esri.com



Thanks!

