Planning an Enterprise GeoDatabase Solution

Robert Kircher, Chris Cushenbery
Enterprise Implementation Services and Product Development
Agenda

- Establish Session Goals and Scope
- Review an Enterprise Geodatabase
- Briefly Review Each Enterprise Geodatabase Component
- Review Deployment Approaches
- Summarize
- Questions
Some Goals

- Clarify key components and tasks needed to build an enterprise Geodatabase.
- Clarify the definition and technologies behind an enterprise Geodatabase.
- Perhaps clear up some misconceptions about the Geodatabase
- Perhaps validate your existing plans
- Provide a high level CHECKLIST
Scope and Assumptions

- A high level overview of components and tasks
- Be practical, not academic (road tested)
- Clear up some misconceptions about the Geodatabase
- Data design, not application design
... the areas that you will consider in your planning.
Enterprise GeoDatabase Components and Internals

Application Stack
- ArcToolbox
- ArcCatalog
- ArcMap
- ArcObjects
- ArcSDE API

ArcSDE/DBMS Server

TCP/IP Sockets

Map and Business Products

ArcSDE System Tables
- Spatial Columns
- Raster
- Versioning

Business Tables
- Geocoding
- Logs
- Locks

SQL

ArcSDE System Tables
- Spatial Processing
- Searches
- Fetching
- Inserting

GeoDatabase System Tables
- Geometric Networks
- Topology
- Subtypes
- Disconnected Editing
- Feature Classes
- Rules
- Feature Datasets

Business Tables with Spatial Column

Raster Tables

Topology Tables

GeoMetric Network Tables

Note: This is a general picture of the internals. There may be some exceptions to this picture.
Why is an Enterprise Geodatabase Important?

- Serves data promptly and efficiently
- Serves data to multiple users concurrently
- Collaborative editing, and long transactions (versioning)
- Provides seamless data
- Secures data centrally
- Provides an infrastructure for distributing and replicating data
- Build integrity into data maintenance
- Provides a means to integrate spatial and business data
- Provides a means to unite and leverage GIS and IT talent and resources
- Provides a means to cut business costs and improve revenue
- … and many others
Conceptual View of an Enterprise GIS Solution

... the enterprise GDB will touch almost every part of an enterprise solution plan
Why plan an Enterprise GeoDatabase solution?
(it’s not so obvious)

• Enterprise Geodatabase projects are elaborate, and require diligent alignment with application efforts
• Enterprise Geodatabases impact almost every part of building an enterprise solution
• PM’s need a means to cost and re-cost, budget, manage staff and deliverables, and mitigate risk
• Enterprise Geodatabases cross boundaries, disciplines and staff (plans create a common goal) Building an effective enterprise GeoDatabase is plain complicated
  – Multiple steps
  – Multiple organizations
  – Multiple risks
Some Common Enterprise Plans and Projects

• **Large Solutions**
  – Consultant team, dedicated project management
  – Custom applications
  – Large user base
  – Elaborate, large databases
  – Multi phased approach

• **Workgroup Solutions**
  – Built in house, part time project management
  – As much COTS functionality as possible
  – Evolve the Geodatabase, gradually move old to new
    • Mixed mode of old and new can work here (old ArcInfo accessing a GDB)

• **All require planning …**
Approaches to Plans and Projects

- **Plan by datasets**
  - Start with a key data/dataset, and evolve the functionality around it (great way to quickly get traction, momentum, and experience)

- **Plan by functionality**
  - Pick discrete parts of functionality, prioritize it (most bang, simplest first), and build it

- **Plan by resources (budget, timeframe, and staffing)**
  - Build what you can get …

- **All require planning …**
Overview of the Major Components

- Data Modeling
  - Conceptual and Physical
  - What tables and data files?
  - Relationships? Topology?
- Data Procurement, Staging, and Loading
  - Where is my data now?
  - Convert it? Tools?
  - When and where to I stage it?
  - Is the data good?
- Data Maintenance
  - How do I edit my data?
  - Add more data?
  - Editing tools?
- Data Performance
  - Is my database serving data fast enough? Storing?
  - How many users can my database support?
  - What are reasonable performance expectations?
Major components -- cont’d

• Data Availability
  – Just business hours? 24x7? Epic “Five 9’s”?
  – Automated fail-over options? Manual recovery?

• Data Recovery
  – Recovery options? DBMS backups and utilities?
    Various export and import tools?
  – Recover responsiveness?

• Data Access
  – Accessing the database? Intranet? Internet?
    Mobile scenarios?
  – Means of access? Interfaces? ArcObjects?
    ArcMap and ArcCatalog? SQL? ArcSDE C or Java APIs?
  – Interface maintenance? Versions compatibility?
Major components -- cont’d

• Data monitoring
  – Who uses my data? How much?
  – Most active data?

• Data infrastructure
  – Centralize it? Distribute it?
  – Licensing issues? (notice how I gave this an exclusive bullet)
  – Other services (recovery, availability, replication, etc.)?
  – Development resources?
  – Configuration control?
Major components -- cont’d

- **Data replication**
  - Architecture options?
- **Data distribution**
  - Sharing data? Internet (ArcIMS)? Flat files (shapefiles)? Automated replication? Direct connection to ArcSDE? Disconnected edits?
  - What data? How much?
- **Data security**
  - Users?
  - Roles?
Let’s **Briefly** Look at Each Component …
Data Modeling
Data Modeling

• Building a structure for your data, including tables, relationships, specification, metadata, identification, feature definition, topology rules, etc.

• Essential Tasks
  – Gather requirements
    • Data Products
    • Map and Visualization Products
    • Analysis and Decision Support Products
    • Maintenance and Editing Needs
    • Metadata (editing and product)
    • Spatial and Business Data Integration
  – Analysis and Design
    • Create conceptual data model
      – Identify data, metadata, specification, relationships
    • Create physical data model, UML
      – Identify GDB feature datasets, classes, relationships, domains, subtypes, geometric networks, etc.
Data Modeling

• **Essential Tasks**
  – **Conceptual Modeling**
    • Document what will be in the spatial database, how the data will be maintained, interact, and be published (conceptual modeling deliverables)
  – **Physical Modeling**
    • Document physical data model in a UML and code
    • Build physical model into Geodatabase instance.
Data Modeling

- **Key Deliverables**
  - Requirements documentation
  - UML based data model or script/code generated data model

- **Challenges and Risks**
  - Application development has critical dependency on the modeling deliverables
  - Normalization balance (over versus de-normalized)
  - Changing the model downstream, propagating schema changes
  - Thorough review of model among publication, maintenance, and vendor teams
  - Optimized for both publication and maintenance needs

- **ESRI Resources**
  - ESRI library of essential, industry data models
Data Procurement and Loading
Data Procurement and Loading

- Building a “one off” loading system for the initial load of the Geodatabase

**Essential Tasks**

- **Requirements**
  - Identify data requirements (spatial, business)
  - Identify conversion and translations requirements
  - Identify data staging needs
  - Identify conversion automation

- **Analysis and Design**
  - Discover data sources (formats, internal, commercial)
  - Define conversion toolset
    - Simple, object data loaders
    - Model Builder, Geoprocessing, Interoperability extension
    - Custom data loader
    - ArcSDE data loaders (very simple data)
  - Create automation
  - Discover anticipated data volumes for storage and DBMS sizing (number, size of features)
  - Define toolset and conversion methods
  - Discover special order to loading
  - Create strategies for big data
  - Define data verification/qc tools and methods
Data Procurement and Loading cont’d

• **Implementation**
  - Build data loading system for development and deployment
    - Procure data
    - Stage data
    - Load data
    - View/use/verify data

• **Key Deliverables**
  - Requirements documents
  - Create toolset
  - Create loading automation
  - Pilot project (full cycle)
    - Identify valid data subset
    - procure data
    - stage it
    - load it
    - Use and verify it
Data Procurement and Loading cont’d

- Challenges and Risks
  - Delivering to application development in a timely manner
  - Versioning
    - Performance
    - reconciliation/post
    - compress
  - Scrubbing data
  - Managing changes to the loaders (full releases vs patch approach)
  - Annotation generally requires special attention to load
  - Procuring entire dataset in a coordinated, timely manner
  - Initial load duration could be significant
    - Capturing any changed data that occurred during load (depending on deployment approach)
Data Maintenance
Data Maintenance

• Help configure and manage the maintenance workflow in the Geodatabase instance.

• Essential Tasks
  – Requirements
    • Identify maintenance workflow requirements (who, what, when, collaboration, automation, objects, batch automation, etc.)
    • Identify anticipated data volatility
  – Analysis and Design
    • Define editing workflows (simple, cyclical, elaborate, optimization opportunities beyond current workflow)
    • Define reconcile, post, compress regimes
    • Define versioning structure
    • Define specifics about edit volumes, version durations, etc. that impact performance
Data Maintenance cont’d

– Implementation
  • Configure versions in workflow or develop them into application
  • Build administration toolset (DBMS, GDB, compress, version folding, etc.)
Data Performance
Data Performance and Scalability (QoS)

- Measure, assess, and optimize the performance of key functionality using the Geodatabase instance.

- Essential Tasks
  - Review anticipated data loads
    - Volume (data file growth management)
    - Extent characteristics (spatial index tuning)
    - Volatility (storage partitioning)
  - Identify key business transactions
    - Maintenance operations
    - Publication operations
  - Identify QoS requirements for key business transactions
    - response time
    - Initial and scheduled user loads
    - throughput
Data Performance (QoS)

- Essential Tasks cont’d
  - Testing
    - Create test plan based on key operations and requirements
    - Develop tests
    - Execute tests
    - Measure performance success
  - Tune application, database server, and DBMS
Data Performance (QoS) cont’d

- **Deliverables**
  - Document requirements
  - Execute performance, analyze, optimize iterations
  - Tuning DBMS, tuning application
  - Scaling strategy
    - Scale out vs up
- **Challenges and Risks**
  - Sizing the spatial index optimally
  - Data too granular
    - Group features
  - Overloading your application
    - Overloading application table of contents
    - Building batch-like operations into application
  - ... many others (please attend the performance related GDB sessions at the conference. This is an important topic).
Data Access
Data Access

- Identify the right ArcSDE and DBMS interfaces (development or off the shelf) for your application.

- Essential Tasks
  - Review DBMS and ArcSDE interfaces
    - **DBMS**: JDBC, SQL, OLEDB/ADO, ODBC, DBMS specific APIs
    - **ArcSDE**: OLEDB/ArcObjects, C and Java API, SQL
  - Identify database access and interface needs
    - Mobile needs
    - Development environment (SQL, ArcObjects, MapObjects, C or Java API)
    - Direct connect, traditional 3 tier, multi tier ArcSDE config
Data Access

• Essential Tasks
  – Identify non-GIS application needs
    • GIS attribute data
    • Business reports based on GIS data or processing
  – Define and configure the application interfaces based on application needs
    • Network configuration (host and ports)
    • Client libraries (e.g. SQLNet, Java libs, ArcSDE client libs, etc.)
Data Access cont’d

- **Deliverables**
  - Document requirements and design
  - Clients correctly configured to access ArcSDE instance

- **Challenges and Risks**
  - Compatibility issues
  - Interface limitations (no interface can do all business and GIS application operations)
Data Monitoring
Data Monitoring

- Monitoring the usage of your Geodatabase, and notifying of any anomalous behavior

- Essential Tasks
  - Identify alerting and notification needs
    - Functional related alerts
      - DBMS crash
      - Query activity
    - Performance related alerts
      - System usage
      - System loads
      - SQL query loads
  - Notification tools and infrastructure
    - Email
    - Telephony
Data Monitoring

- **Essential Tasks cont’d**
  - Configure alerts and notifications
  - Automate alert responses where practical
  - Measure data volumes and volatility
- **Deliverables**
  - Document of alerts and notification needs
  - Procedures, tools, scripts for creating alerts
  - Capture and alerting infrastructure
Data Monitoring cont’d

• Challenges and Risks
  – Configuring and maintaining an elaborate infrastructure of alerts
  – Overly alerted (i.e. a system that cries wolf).
  – Creating a responsive system to alerts (automated and manual)
Data Security
Data Security

- Making the Geodatabase instance secure from unauthorized users

- Essential Tasks
  - Review DBMS authentication schemes
    - Integrated with OS and network domain security
    - Standard DBMS security
    - Mixed mode
    - Users and roles
  - Identify anticipated users (GIS and business applications), and accessible objects (spreadsheet sitting here)
Data Security

- **Essential Tasks cont’d**
  - Define and create DBMS users
  - Define and create DBMS roles
  - Create security monitor
  - Update DBMS security patches

- **Deliverables**
  - Document of users, roles, and object access.
  - Configured users and roles
Data Security cont’d

• Challenges and Risks
  – Sharing a DBMS login
    • SDE_logfile contention point
• Note
  – It’s easier to grant access to users later, than it is to revoke later.
Data Infrastructure
Data Infrastructure

• Building the hardware and software infrastructure for the Geodatabase instance, and all the related data services

• Essential Tasks
  – Requirements
    • Identify hardware and software requirements based functional and system needs
      – Development and test
      – Production
      – Licensing
      – System capacity and growth
      – Storage needs
      – Host CPU, RAM
      – Network bandwidth
Data Infrastructure

- Essential Tasks cont’d
  - Requirements
    - Identify configuration control requirements and plans over development database and server
      - Software development databases
      - Data scrubbing databases
      - Pre-production/integration database
  - Implementation
    - Procure and configure hardware and software
      - development and test
      - production
Data Infrastructure cont’d

- Implementation
  - Create ArcSDE and DBMS instances
  - Development, Test, and Deployment
Data Infrastructure cont’d

• **Deliverables**
  – Configured development and test environment
  – Configured production environment
  – Create and enforce configuration control plan

• **Challenges and Risks**
  – Creating nimble development and test environments (quickly change)
  – Controlled and predictable configuration control
  – Difficult to size and plan for capacity early on
    (there are mitigation tactics for this, scale out/up)
Data Recovery
Data Recovery

- Building a reliable and responsive recovery system into the Geodatabase instance

- Essential Tasks
  - Review and understand recovery options
    - DBMS backup resources
    - DBMS import/export
    - 3\textsuperscript{rd} party DBMS solutions
    - ArcToolbox conversion tools (various flat files)
    - ArcSDE sdeimport/sdeexport
    - GDB copy/paste
    - Disconnected editing
    - Incremental vs full backups
Data Recovery

- Essential Tasks cont’d
  - Requirements
    - Identify recovery requirements
      - Acceptable time to full/partial recovery
      - Database to recover
      - Application configuration/data recovery
  - Analysis and Design
    - Define and create recovery procedures
    - Define and create recovery tools, scripts and automation
  - Implementation
    - Create recovery toolset, and automation
Data Recovery cont’d

• Deliverables
  – Document requirements and design
  – Document recovery procedures

• Challenges and Risks
  – Not ensuring backup actually works, testing it regularly.
  – Time constraints will not allow full backup, only deltas
  – Stringing together backup changes
  – Disk size limitations
  – Automation fraught with complications
  – Recovering application specific data, configuration files
Data Replication
Data Replication

- Replicating data from the Geodatabase instance to subscribing users or systems.

- Essential Tasks
  - Review and understand common replication configurations
    - Snapshot
    - Multi-master/merge
    - Transactional
    - Hybrid
  - Review replication options
    - Device level, OS level, DBMS level
    - 3rd party solutions
    - ArcGIS solutions (disconnected editing, extracts, distributed replication, archiving)
Data Replication

• Essential Tasks
  – Requirements
    • Identify replication uses and benefits
      – Performance/load balancing
      – Mobility
      – Recovery
      – Availability
      – Network load reduction (be careful here)
    • Identify data to be replicated
    • Identify QoS requirements
      – how fast should changes replicate?
      – how frequent is acceptable?
Data Replication cont’d

• Essential Tasks cont’d
  – Analysis and Design
    • Define replication architecture
  – Implementation
    • Prototype and test architecture (crucial)
      – Key data modifications
      – Typical and peak loads
    • Procure, install, and configure replication architecture
  – Test
    • Test replication accuracy and responsiveness
Data Replication cont’d

• **Deliverables**
  – Document requirements and design
  – Full cycle of prototyping
    • Procure and configure replication software/hardware
    • Build master database
    • Modify data, and measure success and performance of replica
  – Configured and tested replication system
Data Replication cont’d

• Challenges and Risks
  – Realizing benefits, “saving reads, not writes”
  – Bandwidth
  – Volume of data modifications
  – Multi master or merge replication is extremely difficult to manage and use
  – Snapshot is useful
  – Choosing BLOB storage over LONG RAW (Oracle)
Data Availability (QoS)
Data Availability (QoS)

• Keeping the Geodatabase instance available for use to your enterprise.

• Essential Tasks
  – Review and understand availability options
    • Server clustering
    • DBMS clustering (share data files, but not processes)
    • 3rd party solutions
    • Replication architecture
    • manual backup and restore
  – Requirements
    • Identify uptime requirements
      – 8hrsx5days, 24x7, 99.999% of year?
      – Mean time between failure
Data Availability (QoS)

- **Essential Tasks**
  - Analysis and Design
    - Define availability architecture
  - Implementation
    - Configure availability architecture
  - Test
    - Test availability failure scenarios against QoS requirements
Data Availability (QoS) cont’d

- Deliverables
  - Document requirements and design
  - Design, procure, and configure availability architecture
  - Testing
- Challenges and Risks
  - Identifying and mitigating
    - single-point of failures in your data system
    - developing redundant systems
      - clustering may not be enough
  - Not consistently testing system
Data Distribution
Data Distribution

• Build an infrastructure to distribute data from the Geodatabase instance to users or systems.

• Essential Tasks
  – Review and understand distribution options
    – GDB to GDB, GDB to shapefile, GDB to coverage
    – Disconnected editing, distributed replication
    – DBMS export file
    – ArcIMS, ArcGIS Server
    – Replication services (see replication)
  – Requirements
    • Identify data consumers
      – ArcGIS applications
      – Custom ArcObject app
Data Distribution

- **Essential Tasks**
  - Requirements
    - Identify delivery needs
      - Response time and throughput
      - Data volume
      - Distribution medium (network, DVD, etc.)
    - Identify distribution infrastructure requirements
      - Bandwidth
      - WAN/LAN/wireless
      - Capacity
      - Security issues
Data Distribution cont’d

- Deliverables
  - Document of data distribution needs
    - Consumers
    - Delivery
    - Destination
  - Configured distribution mechanisms
- Challenges and Risks
  - Bandwidth limitations for moving data
  - Size of data
  - Advanced data types (annotation, geonetworks)
Other Planning Observations
Deployment Approaches

- **“Big Bang”**
  - Release the entire system at once (data, functionality, staff, etc.)

- **Geographic Rollout**
  - Users by geographic regions
  - Data by geographic regions

- **Functional Rollout**
  - Release by discrete functionality

- **Dataset Rollout**
  - Release by dataset
Other Planning Issues

- **Training**
  - Introduction to editing and versioning
  - Introduction enterprise Geodatabase administration
    - ArcSDE admin tools
    - DBMS admin
  - Introduction to enterprise Geodatabase application development
    - ArcObjects
    - MapObjects
    - ArcSDE C and Java API
    - SQL
Other Planning Issues

• Organizational Roles and Responsibilities
  - The Enterprise GeoDatabase technically unites the organization, but requires new, more collaborative relationships to resolve issues like:
    • Recovery
    • Security
    • Anomaly resolution
    • Performance tuning
    • Development

• Put your database team within reach of your development staff
Profiles of Poor Plans ...

- Do NOT allow enough time and care given to data procurement and loading
- Do NOT carefully profile user loads and performance expectations
- Over commit replication to move data
  - packed word and concept
  - Usually difficult to configure and manage
  - Usually brittle
- Overloading ArcSDE/GDB with bad application decisions
  - Too many requests
  - Too much data volume
- Do NOT align correctly with data consumers during development and deployment
- Poor deployment analysis and plan
Key Milestones

- Initial data model delivered
- Initial data procurement and staging delivered
- Initial data loaders delivered
- Initial data load delivered
- Initial results from performance testing key operations and loads delivered
- First complete pass through the lifecycle of a typical data edit delivered (start editor to customer receipt)
Quick Demo: Draft of Generalized Project Plan
Additional Resources

• For more detailed discussion of the components:
  – Other sessions
  – White papers
    • ESRI Website
    • ESRI Support Website
  – Product documentation
  – ESRI staff at the conference
  – Enterprise Implementation Services
  – ESRI Enterprise Advantage Program (EEAP)
Open to Questions

Contact info:

Robert Kircher
rkircher@esri.com

Chris Cushenbery
ccushenbery@esri.com