Using Spatial ETL in a Multi-Vendor Enterprise GIS Environment

Dennis Beck, PE
President
Agenda

- Introduction
- Multi-vendor architecture problem statement
- Challenges in implementing multi-vendor solutions
- Case study architectures
About Spatial Business Systems

Delivering the power of spatial information

• Full-services solution provider for geospatial technologies in the utility, telecommunications and government market sectors

• Integrators of major spatial tools and platforms

• Offices in Lakewood, CO and Melbourne, Australia

• Data management, project management and spatial application business strategy consulting

• Spatial integration product solutions
The Leaders in Spatial Data Integration
Over 100 spatial information user organizations in 22 countries
Problem Statement (1)

- Infrastructure-based organizations have tried to consolidate multiple spatial/graphical platforms that exist across different business functions:
  - Design
  - Planning
  - Construction
  - Operations
  - Field users
  - “GIS”

- The problem has typically been addressed via platform consolidation efforts
- Multiple systems still proliferate within these organizations
Problem Statement (2)

- There are multiple reasons why these systems proliferate
  - Mergers and acquisitions
  - Legacy systems that are hard to migrate
  - Specialty, packaged spatial applications
  - Unique requirements by different user communities
  - Prevalence of CAD

- The problem is continuing, rather than going away

- This is creating new architectures that are focused on
  - Data management
  - Interoperability
  - Well-defined integration points and business controls to ensure data integrity
Challenges in a Multi-Vendor GIS Environment

• GIS vendors typically support spatial data management functions within their product suites
  – Unique modeling requirements
  – Unique approach to long transactions / conflict resolution
  – Unique indexing approaches
  – Unique access methods
  – Unique formats

• Complexity – more systems to manage

• Operational issues
  – Training
  – Support
  – Change management

• Data integrity
Approaches

• Loose coexistence (Business as usual)

• GIS platform as hub (Geo-Centric)

• Database as hub (Database-Centric)
Geo-Centric Architecture

• The GIS platform is the source of truth
• External products connect to the GIS
• The GIS is responsible for data management functions
  – Validation
  – Vendor’s data model, network model, metadata structures
  – Conflict resolution

• Example – CAD / GIS integration
  – Loosely coupled approach: ETL (Extract-Transform-Load)
  – Tightly coupled approach: dynamic integration, e.g. OSGeo Feature Data Objects (FDO)
Geo-Centric – FDO Use Case

Natively access data from multiple spatial sources

Avoids:
- Conversion
- Data loss
- Data copies
- Stale data

Parcel data from a SHP file
Utility data from Smallworld VMDS
Property data from Microsoft® SQL Server™
Zoning data ESRI® ArcSDE®
Aerial photos

Source: Autodesk
Database-Centric Architecture

• Database is the source of truth
  – *Not* the GIS vendor’s application database
  – Typically a spatially enabled relational database (Oracle, SQL)
• GIS platforms serve application-specific roles
• Spatial ETL exchanging data between systems
• The spatial database environment takes on an added role in managing spatial data
  – Long transactions
  – Conflict resolutions
  – Metadata management
  – Symbology
  – Additional spatial capabilities – e.g. network modeling
• Examples: Large investor-owned utilities
High Level Data Architecture

- Smallworld
- Application Database
- Map3D
- Application Database
- Landbase Updates
- DMS
- Canonical DataStore Mirror
- Spatial Data Warehouse
- Operational DataStore
- Canonical DataStore
- Staging

DataMart DataMart DataMart DataMart

AUD Web
Operational (Spatial) Data Store Architecture
Use of Spatial ETL (Extract-Transform-Load)

- Data model transformation support
  - Attribute modification / Feature merging and splitting
  - Relationship generation / Topology generation
- Validation
- Error handling with notifications via email, logs, tweets, NT Service logs
- Can support full synchronization with appropriate plug-ins
- Scalability via the use of server-based approaches (e.g. FME Server)
- COTS support, particularly for upgrades
- Web based initiation, interaction and resolution of synchronization processes
- Flexibility to support current and future format requirements
ETL Example: Bi-Directional Process Flow

Example: GE Smallworld and Oracle Spatial

Role of plug-ins and transformers
- Format support
- Model transformations
- Difference management
- Common (Global) ID management
- Network model extraction
Conflict Resolution

• As data is merged and posted between workspaces, the potential for conflicts in the long transaction data exists
  – Such situations occur when two users modify the same feature at the same time, but in different workspaces.
  – In this case merge / post operations may result in overwriting one user’s data in favor of the other

• Mechanisms must be in place to detect changes
  – Oracle Workspace Manager provides tools for comparing changes from different workspaces and can identify sources of data change conflict
  – Change-Data-Capture (CDC) offer alternate mechanisms to identify differences
  – Trigger-based approaches are also used

• It is necessary to visualize the changes to support comparison and resolution of changes via semi-automated or manual processes.
• Conflicts can be identified and resolved in the system, resolution results are sent to the peer
• Properly designed business processes will help to mitigate most conflicts
Global ID Model

• GUID_REGISTRY
  – Repository / Mapping for each instantiated GUID
  – ODS Feature: ODS object where this object acts as a Primary Key
  – Source Feature: Identifier for the specific feature this data is brought over from, in the source system.
  – Source System: System where this data existed previously
  – Key: Unique Identifier for this record in the Source System, for the Source Feature object (These 3 values combine with the specific ODS Feature to form a Unique Index on GUID_REG)
  – The Global ID API is designed to leverage stored procedures in the ODS wrapped with ETL transformers
Other Considerations

• Common validation framework
  – Valuable for ensuring consistent representation
  – Source system validation may not address all use cases

• Common symbology
  – Multiple symbol sets from source applications
  – Necessary to persist some in the ODS
  – Symbol and Text Stroking used for Web / Thin client deployment
    • Converts point and text to multi-linestring with symbol/font/size/orientation embedded
    • Implemented via ETL Transformers
  – Display scales

• Common network models in the data base
  – Auto-generated via ETL plug-ins
  – Supports externalized application requirements

• Application services
Multi-Vendor Architectures – Value Proposition

Why bother?

- Productivity improvements
  - Use the best tool for the job, e.g. CAD
  - Specialized applications to support trained staff members
- Eliminate redundant data entry
- Reduce custom developed applications
- Reduced upgrade costs via model transformation
- ODS as an integration hub
  - Lower cost integration
  - Improved maintainability
- Value of improved data management
  - Currency of data
  - Availability to broader user base
- Enables more advanced mobile computing
- Support phased platform migrations
Questions ?