Leveraging Geoprocessing Functionality to Manage Enterprise Data in Sacramento County

Presenting Authors:

Cynde Porter & Bob Earle
Sacramento County GIS
Sacramento, CA

Abstract:

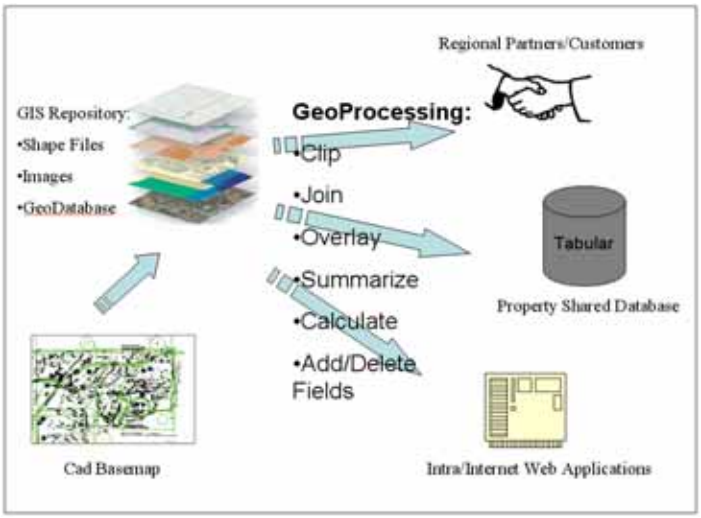
As primary data providers for enterprise departmental systems, intra/internet ArcIMS solutions, shared GIS cooperative data repositories, and E-911 dispatching, Sacramento County is committed to providing accurate, timely and standardized tabular and spatial data in the most efficient manner possible. We have accomplished this by utilizing geoprocessing scripting in tandem with both Oracle and SQL Server core database functionality. Scripting has been used not only to facilitate laborious creation/conversion processes and disseminate the resulting datasets in multiple formats, but also to integrate real-time tabular information derived from numerous relational databases. We have further been able to embed compiled geoprocessing scripts into scheduled RDBMS Data Transformation Packages to integrate native SQL coding and view generation with fundamental GIS capabilities. As a result of more efficient data integration and task consolidation, staff are now freed to devote more time and focus towards developing and growing our GIS to support ever increasing customer requirements.

While many of the tools used by GIS professionals are often flashy and futuristic in appearance, the reality is that, during the course of day to day business, most of us are asked to perform a set of repetitive, time consuming, and sometimes hefty tasks. New GIS technologies, such as Temporal 3-D Modeling and real-time web integration, while practical and beneficial in many cases, are not what comprise the majority of our job duties in an enterprise setting. The nature of business practices here at Sacramento County mandates that we publish our spatial as well as tabular data in a variety of formats. We maintain and support hundreds of spatial layers, and are responsible for the publication of many which are essential to the business operations of various County departments. Of these, the most critical data layers include:

- Parcels
- Legal Lots
- Subdivisions
- Proposed Lots and Subdivisions
- Regional Street Network
- Districts (Tax, Fee, Special)

Our primary land base is created and maintained using a gridded tile structure comprising over twelve hundred individual AutoCad drawing files. After several needs analyses, we

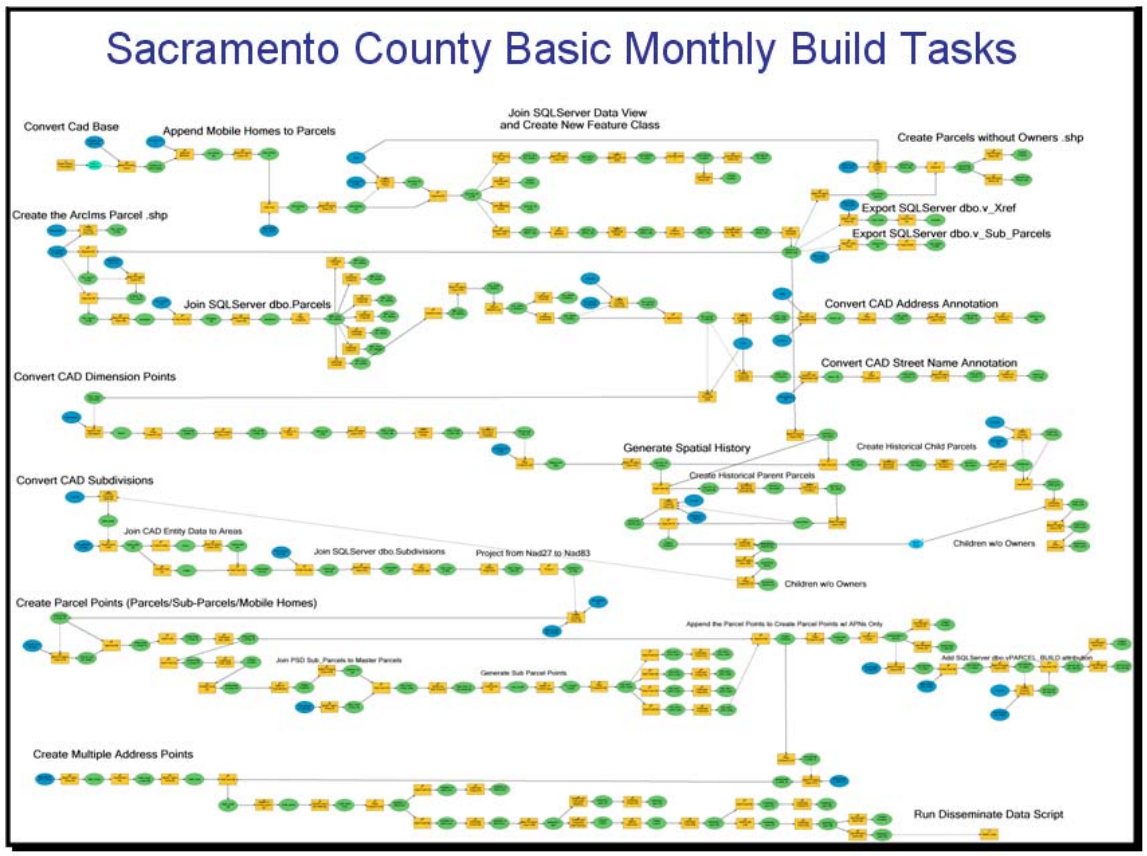
have determined that, at least for the immediate future, it will be necessary for us to retain AutoCAD representations of our base data in order to support the business needs of our customers. From these source drawing files, we build and populate GIS data layers in shapefile and personal geodatabase format for dissemination to internal County agencies, regional partners, intra/internet web applications, and external customers. The original CAD data is converted utilizing legacy .amls for the creation of the initial spatial features. To these spatial features, we append tabular data extracted from both our Oracle Property Shared Database (PSD) and departmental attribution maintained in SQL Server. We process high volumes of data from many sources which are frequently updated or refreshed and then propagated into multiple configurations for distribution to the data consumers. These processes are repetitive and time consuming, yet mandatory, to satisfy customers who expect standardized products on a predictable and timely schedule. In addition to our own internal County departmental needs for data, in a regional data sharing environment, such as our Sacramento County GIS Cooperative, the necessity for a dynamic conversion capability, highly reliable output, and documented processing history becomes even more imperative. Other daily tasks, such as incorporation of early entry subdivisions into our parcel land base and street network, overlay analysis, and relational database population, have driven us to re-examine our internal GIS workflows in order to support more timely data dissemination.



With the advent of geoprocessing, ESRI has provided a methodology for answering this challenge. These new toolsets offer a framework for automating many of the day to day tasks necessary to support a fully integrated enterprise-wide business operation. The resultant products are standardized with consistent output, which previously presented quite a challenge for us to achieve as part of a multiple step process. The functionality has allowed us to decrease the amount of time spent performing many of our scheduled tasks. One of our larger geoprocessing scripts has effectively cut staff time from a six day ordeal to a mere few hours. In addition, we can now schedule these jobs to run

during evening hours in order to minimize impact on other computer applications and server resources.

Typically, we employ Model Builder to design initial process chains, the largest of which is illustrated below:



Upon establishing these basic workflows, we export to Python in order to leverage its robust scripting environment and utilities which are currently unavailable in Model Builder. Python scripting allows for batch processing, looping and conditional branching, messaging (error, informative, output), and string manipulation. We have found Python to be a user friendly language with a minimal learning curve useful for programmers and non-programmers alike. As part of a small GIS shop, we do not have the luxury of employing a full time application developer, but have found the geoprocessing object model, used in tandem with Python scripting, to be a relatively simple and powerful solution useful for accomplishing our daily tasks.

While the bulk of our tabular data is maintained in fully normalized relational databases, often times, it is necessary to utilize some intelligent translation to de-normalize this

information for integration and packaging of various datasets designed to satisfy a myriad of diverse business driven requirements. Normalized data is customarily not the fastest performer for web queries and feature labeling. In addition, it is often necessary to concatenate labeling fields and join attribution from business and lookup tables. Geoprocessing/scripting offers an efficient and encapsulated solution for accomplishing these goals. Upon integrating required tabular attribution with spatial features, we then need to utilize an inventive means of packaging data for distribution. Several data translation tasks facilitated by geoprocessing include:

- Aggregation of features/summarized values
- Re-projections
- Field mapping/re-naming
- Data formatting
- Spatial to tabular conversions (such as parcel/district cross-references)

While the above processes are readily accomplished within ESRI applications, we have found that automation utilizing geoprocessing is not necessarily limited to this environment alone. Geoprocessing scripts can be embedded in other software packages in order to exploit their functionality. Because much of Sacramento County's automation is in the form of SQL processing, we rely on Data Transformation Services (DTS) within the SQL Server framework. Compiled geoprocessing scripts are just one step which can be used in a series of data manipulation tasks.

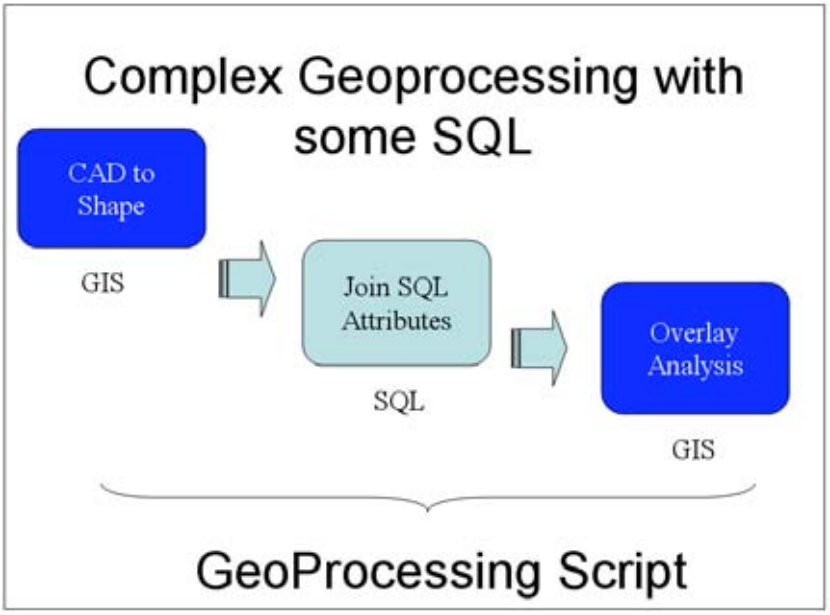
Advantages of Geoprocessing:

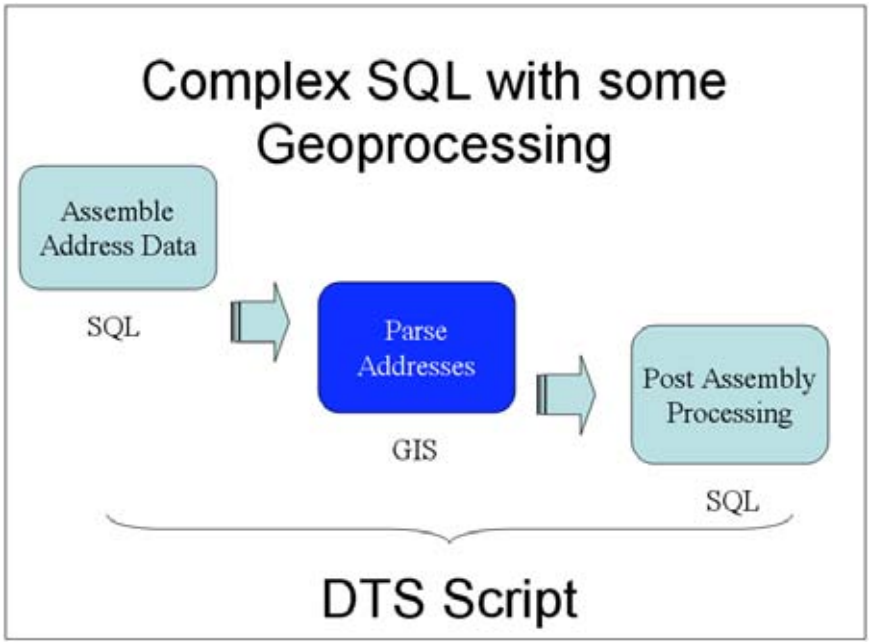
- Automates recurring/repetitive tasks
- Produces standardized and consistent output
    - Field naming conventions, formatting
    - QC checks and balances
- Decreases staff time which can better be devoted to other issues
- Can be packaged and shared
- Self documenting
- **Perfect for mostly spatial operations**

Advantages of Data Transformation Services:

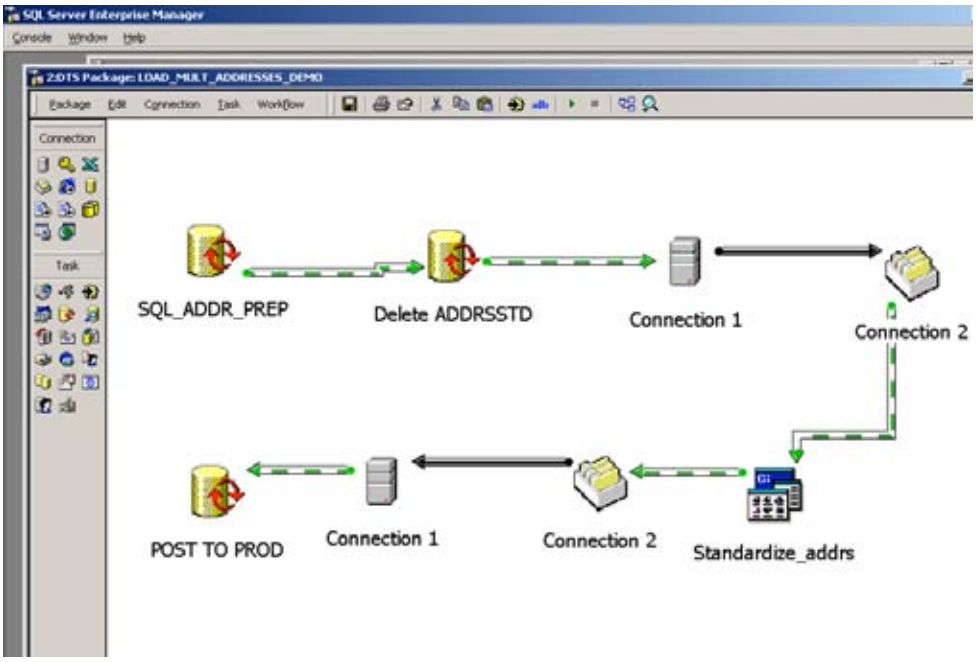
- Automates recurring/repetitive tasks
- Produces standardized and consistent output
    - Field naming conventions, formatting
    - QC checks and balances
- Decreases staff time which can better be devoted to other issues
- Can be packaged and shared
- Self documenting
- **Perfect for mostly tabular operations**

If the bulk of the processing is spatial, then we embed the SQL operations within a model or script. Conversely, if the bulk of the processing is SQL, then we embed the geoprocessing into the DTS as illustrated below:


Complex Geoprocessing with some SQL
CAD to Shape
GIS
Join SQL Attributes
SQL
Overlay Analysis
GIS
GeoProcessing Script


Complex SQL with some Geoprocessing
Assemble Address Data
SQL
Parse Addresses
GIS
Post Assembly Processing
SQL
DTS Script

One such example of leveraging geoprocessing functionality from within the SQL Server environment is shown in the following diagram:



The above process details a workflow for generating a multiple address table by consolidating street address information from numerous relational databases (ex. assessor, permitting, utility billing, etc.) using native SQL functionality. Once consolidated and scrubbed, an embedded, compiled python script standardizes the tabular data by calling the standardizeaddress_geocoding method of the geoprocessor object. While this process is easily accomplished in the GIS world, the same task would require a significant deal of coding using SQL Server alone in order to achieve the same results.

In conclusion, many GIS tasks are often voluminous and laborious, while data accuracy, currency, standardization and dissemination are the pivotal components which drive mission critical business systems. ESRI's geoprocessing toolsets offer both flexibility and wide-ranging capabilities for decreasing the resources needed to maintain this information while liberating data maintainers and GIS analysts to do what they do best.

Author Information

Cynde Porter
Information Technology Analyst II
Sacramento County GIS
9700 Goethe Rd. Suite A
Sacramento, CA 95827
(916) 875-7032
Fax (916) 875-6213
portec@saccounty.net

Bob Earle
Principal Technology Analyst
Sacramento County GIS
9700 Goethe Rd. Suite A
Sacramento, CA 95827
(916) 875-6800
Fax (916) 875-6213
earleb@saccounty.net