BIOGRAPHICAL INFORMATION

Graham Morgan
Solution Architect
SchlumbergerSema

Specific Responsibilities

Joined SchlumbergerSema (as Convergent Group) in 1999. Responsible for crafting complete technology based solutions for public sector clients.

Past Experience

Graham has been designing, implementing and deploying geospatially based solutions for over 10 years. He has worked on geospatial projects for British utility, Australian Federal government, Canadian Provincial government, and American State and City government organizations. He is familiar with all leading GIS vendor offerings and database technologies. Graham leads the GIS special interest group within Schlumberger.

Educational Information

B.Sc. (Hons) – Geology, University of London
M.Sc. Geographical Information Systems - University of Edinburgh

Professional Memberships

IEEE

LESSONS IN GEOSPATIAL DATA WAREHOUSING – A PROVEN APPROACH TO ENTERPRISE DATA INTEGRATION

Graham Morgan & Gary Jurkowski
SchlumbergerSema
Suite 500, 6399 S. Fiddler's Green Circle
Greenwood Village
Colorado 80111

ABSTRACT:

SchlumbergerSema has developed enterprise data warehouses for the City of Portland OR, City of Kansas City, MO and the State of North Dakota. These geospatial data warehouses address a broad range of business requirements including improved business processes, cost reduction and avoidance, improved data quality and access, and improved decision support. Factors critical to the success of these projects have included a universally understood business case, well-defined intra-organizational responsibilities and a well-engineered system architecture. The authors will describe the organizational and political environments encountered, and proven technical architectures that work.

INTRODUCTION

Geographical Information Systems are a component within the information technology environment of virtually every enterprise regardless of the nature of their business. Why are GIS systems so common – and what limits their effectiveness? This paper will briefly examine the nature of GIS and the problems that limit many implementations. It will then describe some proven approaches that enable the integration of GIS into the core of the IT architecture to deliver benefit across the entire organization.

Why GIS?

It is generally accepted that over 80% of all the data used day-to-day in business and government can be related directly to geographic features like zip codes, street addresses, factory locations, census tracts, cities, states and countries. In fact it is hard to think of any information that cannot be associated with a location at some scale! Even phenomena such as disease, poverty, temperature and sunshine can be mapped and managed within a GIS. The geographic, or spatial dimension within information allows data to be correlated in new and intuitive ways – providing for a better understanding of any given situation and improved decision-making. The geographic dimension of data can be essential for certain business decisions, for example a business rule might state that a permit cannot be issued for a tobacco shop that would be located within a specified distance of a school; or that a new address number cannot be the same as an existing address number within a specified radius. These types of rules can be verified very efficiently within a GIS.

What is a GIS?

A geographic information system is not simply a piece of software from any given vendor that serves up maps and pretty drawings. Many organizations see it is a business application where it may comprise multiple pieces of software; a database, perhaps a web server; mobile data collection units; data translation engines; metadata management applications; data standards and more. Together these components serve a set of business functions that directly support business operations.

What Does a GIS Do?

A GIS really plays two roles that distinguish it from other business applications:

1. Consolidated Analysis and Reporting

Through the innate ability to relate practically all data by virtue of its spatial dimension, GIS presents tremendous opportunities for more rigorous analysis and more informed decision-making. Data from the Customer Information System, for example, can be correlated with data from the Outage Management System and Inventory Management System to manage levels of customer satisfaction or to understand how service levels influence customer loyalty; the results of which may be more efficient targeted marketing campaigns. No other business application is equipped to consider such a wide range of data points.

Furthermore, the results of GIS analysis can be presented in a highly visual and intuitive fashion. GIS 'reports' beat traditional reporting formats hands-down. Maps and charts with graphical symbology and annotation are able to present many data aspects together to allow an analyst or decision maker to visualize data relationships and draw new conclusions. Data presented in this way is much more accessible and comprehensible than rows of printed data values in a tabular report.

2. Spatial Data Management Services

The second major role of GIS within an enterprise is to provide spatial data management services to supplement the data management provided by other business applications. As described above, virtually all business systems maintain data about features that have a geographic location. However, they are unable to maintain the geographic information themselves because it is essentially just too difficult to do so – database and data entry systems simply do not provide the means to edit and store spatial information. For some business systems (e.g. Distribution Management Systems) the geographic dimension of their data is just too significant to ignore, and whilst they cannot effectively manage location themselves they can use a GIS to supplement their information with location attributes. In some cases a business application may be built directly on top of a GIS (e.g. GE Smallworld PowerOn Outage Management System), but in most cases location is stored in the GIS and a key value is used to relate the GIS records to the corresponding records within the business system.

The Need for GIS Integration

A GIS cannot exist in isolation – in order to provide the services described above a GIS must be able to exchange data with other business systems. The nature of the integration is likely to vary, for example the spatial data management service may (depending upon the implementation) require near real time integration with distributed transaction management capabilities, whilst the consolidated analysis and reporting service may be able to operate effectively with weekly or monthly batch data imports. In other words an enterprise GIS implementation requires a strategy to integrate the GIS with the other business systems within the organization. This strategy should support a variety of integration needs.

Integration Challenges

System integration is a complex matter for a number of reasons. Systems may have been developed using different programming languages, Operating Systems, communications protocols, data formats and different data meanings.

Despite the fact that all systems need to exchange data at some point – many have been built without even considering this basic requirement – and most commercial GIS products fall into this category. Figure 1 illustrates the basic styles of software architecture.

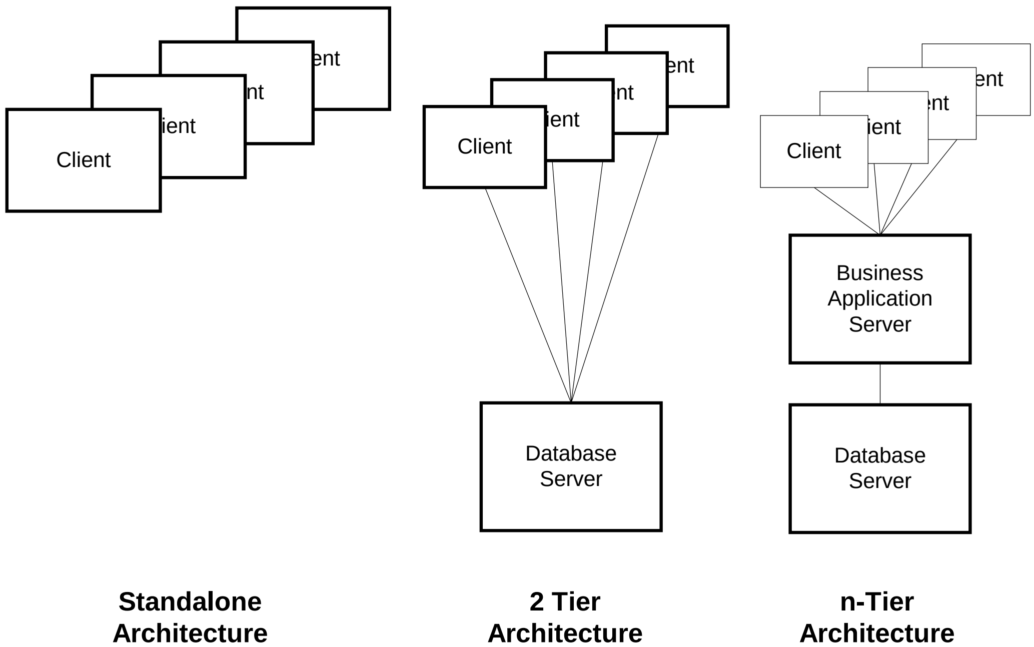

Figure 1 - Software Architectures

Modern 'best practice' software architectures (n-Tier or 'thin client') consolidate all business logic (validation rules, process flow logic, etc) on a business application server and expose that logic through an application programming interface (API) (Cummins, 2002). A server is designed to support many concurrent users; it 'knows' how to manage connections and computing resources; it is generally provided on a high performance machine with high availability. Integrating with a system with this architecture is relatively straightforward. Many systems, however, have been developed using the 2 tier Client-Server architecture in which business logic is distributed between the desktop client application and the database server – this is typical of most GIS products as well. This architecture is difficult to integrate with, as the calling application must know whether a particular function is accessed via the database server or via a client.

GIS Specific Challenges

Integrating with, or between, GIS systems introduces some particular challenges in addition to those described above; these include:

- Lack of Standards: To begin with there is a dearth of standards and standards compliance within GIS products; all GIS vendors have developed their own proprietary geometry types and data formats which are not natively understood by other vendor products.
- No standard Geographic Query Language
- Proprietary Programming Languages: In some cases vendors have developed proprietary programming languages (e.g. ESRI's Avenue and Smallworld's Magik).
- Fat Client Architectures: Virtually all commercial GIS products employ the 2 tier fat client architecture described above i.e. they do not provide server based API integration points.
- Spatial Data Complexity and Volume: Spatial data is notoriously large in volume, which may stress existing networks if exchanged frequently.
- Proprietary Long Term Transaction Mechanisms. Long Term Transaction (LTT) mechanisms allow concurrent edit access to the same spatial features – actions that typically take minutes to days to perform. LTT mechanisms are a great step forward for geospatial data maintenance, scenario modeling and data lifecycle management. Relational Database Management Systems (RDBMS), however, employ short term transaction mechanisms which are well suited to non-spatial data editing operations. Vendors such as Smallworld, Oracle and ESRI have developed their own LTT mechanism that introduces another integration challenge because they have each implemented their LTT mechanism differently. Once again there are no standards in this area. Even when a GIS vendor stores data in a commercial RDBMS their proprietary LTT mechanism is likely to render the data inaccessible to all but their own API.

Despite the fact that GIS is difficult to integrate with; GIS is either of very little use or redundant without it. Data created within business systems must be accessible to GIS for analysis and reporting. Geometrical data managed within the GIS must be linked with associated data in other business systems. These links must be vigorously and consistently managed to preserve data integrity.

ENTERPRISE GIS INTEGRATION ARCHITECTURE

In an environment where it has been determined that different problem domains are better served by different solutions, the City of Kansas City

(Missouri) has committed to supporting two GIS platforms. Interoperability is therefore a key objective for the City; another is the integration of GIS and the multitude of business systems it serves.

This project has been coined the New GIS Program and the City has contracted SchlumbergerSema to design and implement an interoperable enterprise GIS platform. Its ultimate objectives are to enable data to be shared between departments with minimal latency and to support key citywide initiatives such as Police and Fire CAD/RMS, Asset Management and a 311 Call Center.

Oracle Spatial GIS Interoperability

Central to the City's vision is the use of an interoperable GIS database; i.e. a spatial database that stores data in such a way that it can be used by both GeoMedia and ArcGIS. Oracle Spatial was selected as the City's spatial database as it was supported by both City GIS platforms and because it's implementation was based upon the Open GIS Consortium (OGC) standards for spatial data storage and SQL manipulation. To a significant extent both ArcGIS and GeoMedia are able to store in, and retrieve data from, Oracle's RDBMS which uses an object data type (SDO_GEOMETRY) and SQL extensions to directly manage spatial data[1]. In this way standard spatial data written to Oracle by GeoMedia can be read by ArcGIS, and vice versa.

Naturally there are a number of caveats to achieving such interoperability. Such considerations include:

- Oracle SDO_GEOMETRY Compliance

Oracle has implemented storage for point, line, polygon and linearly referenced features. This implementation covers multi-part geometries, straight line and geometric curves as well as overlapping features and features with holes. Features can have multiple geometry values and geometries can be stored in either two or three dimensions (X,Y,Z coordinate values). Neither vendor fully supports Oracle's complete implementation and care must be taken to understand vendor limitations in designing an interoperable database.

- Vendor Custom Geometries

There are a number of common, as well as specialized, spatial data types that Oracle has not implemented (principally due to lack of industry standards). Oracle does not provide direct support for annotation, dimensions, topology, rasters, surfaces, etc. Each vendor has implemented these and other features in their own proprietary fashion which cannot be completely understood by Oracle's SQL implementation or by the other vendor's GIS software[2]. These features must still be translated between vendor formats.

---

[1] Interestingly neither GeoMedia nor ArcGIS require an Oracle Spatial license to store their data within an Oracle database. Both take advantage of the spatial features provided by standard edition Oracle 9i Release 2 (9.2.0.3) RDBMS.

[2] An alliance between some GIS vendors has recently developed a neutral means to store annotation within Oracle – this format may be submitted to the OGC for standardization. Oracle has indicated that they would consider supporting such a standard within their product. Oracle will provide support for additional spatial feature types in their forthcoming 10g release –the extent to which the GIS vendors will support these new types is not clear at this stage.

- Long Term Transaction Mechanisms

GeoMedia and ArcGIS employ different LTT mechanisms. GeoMedia uses the GeoMedia Transaction Manager (GTM) which is built on top of the Oracle Workspace Manager (OWM). ArcGIS uses ESRI's proprietary LTT mechanism. Whereas data managed under the GTM/OWM can be accessed easily through SQL it is more problematic to access ESRI's versioned data. Furthermore ArcGIS requires that data be versioned for editing whilst GeoMedia does not. Through their 'multi-versioned views', however, ESRI do provide a way for non-ArcGIS clients to read their versioned data.

- Business Rule Validation

Both vendors implement data integrity rules/constraints differently. ESRI define rules as COM objects that are invoked through an ArcGIS validation harness within the GIS client application – minimal use is made of Oracle database constraints. Intergraph implement rules in a variety of ways: configured within the GeoMedia client, customized within domain specific packages, and within the Oracle database. Both vendors have thus implemented rules in such a way that it would be very difficult for any other application to be able to use the same rules. This limitation coupled with the different LTT mechanism effectively prohibits interoperable data editing.

In addressing these issues; the concept of a central GIS data warehouse in the Kansas City environment does enable data to be shared between different GIS products for browsing and analysis purposes i.e. data written by one system can be consumed by the other system. The use of Oracle as the standard for data storage enables much of the GIS data to shared without the need for translation. Native Oracle replication can further be used to transport data from the maintenance environments to the central Hub.

Business System to GIS Integration

The common use of Oracle's SDO_GEOMETRY data type goes along way to enable spatial data to be shared between GIS products. The spatial data still, however, needs to be created and associated with other information in the appropriate business systems such as the Asset Management and Permitting systems. This is a different type of integration challenge and cannot be solved by the use of a common data format alone.

As described in the Spatial Data Management Services section above there are typically many systems within an enterprise that require, or would benefit from, storing a spatial dimension in addition to the attribute data that they already manage. GIS provides the capabilities to manage the spatial data but the key problem still remains of how to keep the business data in sync with the GIS data.

<u>The New GIS Integrated Dual Entry Approach</u>

Kansas City is adopting a service-oriented approach to the citywide technical architecture as way to solve integration between the many business systems and the City 'New GIS'. City departments use GIS to manage spatial data that is often associated with data in other business systems such as permitting and facilities management. Although it is often desirable to enter the spatial data at the same time that data is entered into the business system, this is not currently feasible, as it requires a high degree of custom development. Kansas City's New GIS architecture uses an integrated dual entry approach to address this problem.

Some business processes will enter data firstly into the business system; other processes will find it more efficient to enter the data into the GIS first. In either case a business identifier for the new record/feature must be created in a consistent fashion and associated with related data regardless of where it is stored. Each case is described below. Business practices within Kansas City require that the business system is always responsible for providing business identifiers for new features.

When a record (e.g. a new water pipe) is created in a business system the system assigns a new feature identifier. The business system in turn invokes a Web Service on the GIS via the messaging (MOM) system to create a new feature (e.g. CreateNewPipe) with the new business identifier. At this point the GIS has a 'stub' record – it has a business identifier but no geometry. The GIS passes control to the workflow management system (WMS), which then informs a GIS operator that there is a new GIS stub record that needs a geometry. The operator uses a custom editing tool within the GIS to locate the appropriate stub records and add the geometry. The GIS informs the WMS when the geometry has been completed. In this way the WMS manages the data maintenance process to ensure that complete records are created across multiple systems in an acceptable time. The WMS is capable of sending reminders to the operator if the task is taking too long and can raise alerts to supervisors if necessary.

In other cases it will be more efficient for an operator to create a record for a feature in the GIS before it has been created in the relevant business system. This integration is a little more complex in that the GIS needs to obtain the business identifier from the business system and store it as part of the new GIS record. Once again wrapping the business system to provide business oriented Web Services can solve this problem. The GIS calls a Web Service via the messaging system to 'officially' create the object and return the new feature identifier. The GIS then alerts the WMS to the new feature record. The WMS tasks a business system operator to complete the record entry using a custom edit tool to locate the stub records.

Managing Relationships across GIS Databases

Consider the case where logically related data elements may be managed by different departments. For example if the Water department digitizes the location of water facilities with reference to the parcel dataset (maintained by the Public Works department in a different database) there is an inherent spatial relationship between the two datasets. If a parcel boundary is edited then the water facilities may be incorrectly positioned with respect to the parcel. The same problem can also occur when attributes are updated or deleted where there is a logical relationship to features in other databases. The WMS supervised approach described above can also be used to address this problem.

Even when integration between systems cannot be fully automated it can at least be managed. The WMS tracks the time each step takes and thus provides the basis for monitoring performance against inter-departmental Service Level Agreements. With the current evolution towards composite applications, it will not be very long before the operator will be able to perform the spatial data edits directly from within the web browser. The WMS will thus be able to provide the work task, together with the data and tools necessary to perform it and receive status updates all within a web based environment.

Infrastructural components such as MOM and WMS can rarely be justified for a single integration point or even a single project. These examples illustrate how they provide such generally useful services that, once they are made available, can be used to solve a variety of problems throughout the organization and they do so in an improved, consistent, reliable and manageable fashion.

The GIS Hub as a Data Integration Broker

The preceding section has described how business systems can be efficiently integrated with the GIS to manage federated records (i.e. records that span multiple systems). The examples have dealt only with the GIS storing spatial data and all other business data being managed within a related business system. This makes sense as it means that the business validation rules for the attributes need only to be managed in one place rather than two. It is not, however, realistic.

Data viewed within a GIS will be drawn, symbolized and annotated. Users will zoom in and out and pan around the area. If the GIS only stores geometry then it will need to constantly make calls to the business system to acquire attribute values such as ‘water pipe type’ to symbolize and annotate. This would result in unacceptably slow draw performance. The problem can only be solved by providing the GIS with access to local copies of the attribute data.

Data replication can be safely used to copy data from the business system to the GIS as the data has already been validated. Furthermore, replicating data to a version-managed (LTT) GIS has been proven and can be configured such that data changes are replicated into the public/default version in near real time. In this way the GIS is naturally starting to aggregate data in one database to provide a holistic representation of a feature. Once this has been accepted and the replication mechanism has been configured to copy some attributes it is a small matter to copy other attributes, perhaps to facilitate querying and analysis. The GIS is a natural data warehouse.

It is one thing aggregating data within departmental operational GIS systems, but in an enterprise setting that data must be made available across the entire organization and the GIS Hub provides a proven means to accomplish this. Data may be aggregated locally within department GIS data maintenance environments (DMEs) and then published to the central Hub for enterprise-wide access. In this way the Enterprise Hub provides a general information Hub – data from any departmental business system may be accessed in one format, in one place, using one mechanism. For example, the Health department via the Hub could access data from the Public Works' Capital Improvement Project system. This greatly simplifies system/data integration across the enterprise. It reduces the number of integration points and enables data to be accessed in a consistent fashion regardless of source. The GIS effectively functions as a core piece of the integration infrastructure rather then simply as another consumer of it.

For this vision to be realized, however, we need to address two key issues: enterprise data standards management, and near real-time publishing from version-managed DMEs to the Hub.

Enterprise Data Standards Management

In order for data to made available in the Hub for use in multiple applications it is necessary that somebody maintains the data and that they maintain it to a specification that serves diverse needs. Data standards covering data content, accuracy, format, completeness, currency, etc. must be determined, agreed to, monitored and enforced (Cook, 1996). As it will often not be in the complete and direct interest of the maintenance organization to observe the needs outside of its own, it is necessary to instantiate a multi-disciplinary management body that is responsible for considering the needs of all users and maintaining data standards accordingly. This body assumes ownership/responsibility for enterprise data and then contracts departments through SLAs to maintain it to the necessary enterprise data standards.

Real-time Data Publishing

In order for the GIS to function as a one-stop Enterprise Information Hub it must provide 'up-to-the-minute' data currency. This is a much more stringent requirement than can be met by the nightly publications triggered by ESRI's geodatabase Compress operation. As data is published to the public/default version in the DMEs (either through posting a version to default or through replication to default from another business system) the changes must be published immediately to the Hub.

Within Kansas City's Oracle Spatial environment this will be achieved using customized Oracle Streams replication. Oracle Streams provides a highly customizable data replication system. Custom Logical Change Records (LCR) are identified by a developer; On-Post events can be used to trigger native replication of the LCRs which are then processed by custom procedures at the subscriber database. Using this robust mechanism the requirements for near-real time data replication can be satisfied.

SUMMARY

Practically all business applications deal with data that, logically at least, has a geographic component to it. Mainstream data management technologies, however, are ill-equipped to deal with the geographic dimension of data and consequently business applications struggle to leverage it.

Geographic Information Systems essentially provide two business services i.e. visually enhanced information analysis and reporting, and spatial information management. GIS thus complements many business applications and naturally needs to be integrated with those systems. The management of geographic information is fundamental and should be addressed at the level of the enterprise-wide technical architecture. This paper has described a real world example of an approach that has been taken to implement GIS as an enterprise data management service.

Kansas City's New GIS architecture uses Oracle Spatial and standards-based enterprise application integration (EAI) technologies to provide tightly-coupled spatial data management services to City business applications and to make changes available throughout the enterprise in near real-time for both ESRI ArcGIS and Intergraph GeoMedia data maintenance platforms.

In all cases the use of a GIS Hub within the enterprise information architecture has improved data timeliness which has resulted in better support for enterprise business processes, reduced cycle times and improved levels of service.

REFERENCES

Cook M. A., 1996, Building Enterprise Information Architectures, ISBN 0-13-440256-1

Cummins F. A., 2002, Enterprise Integration, ISBN 0-471-40010-6