This paper addresses the significant leverage, collaboration, and performance that can be obtained with Open-Source Software for Geospatial Data Management and Information Systems. Open-source software provides the building blocks for advanced geospatial archiving, processing, and distribution. Recently, robust open-source software solutions have proliferated in remote sensing, geographical information systems, spatial data engines, and relational databases. Additionally, advanced supercomputing capabilities with clusters of commodity personal computers can be applied to the management and processing of extremely large data sets. These advances have been integrated into highly-leveraged systems that can be applied to a wide array of remote sensing and geospatial activities. The authors describe the various technologies currently available and how these technologies have been integrated into an online Data Management System. Examples of the application of this system to multiple application areas are described. Examples include custom processing and organization of satellite, aerial, raster map, and vector data sets.

Geospatial Data Management; OSSIM; EarthWhere

I. INTRODUCTION

Today, an organization’s data is universally viewed as a major corporate asset. However, the methods by which information is derived from that data remains unique to each organization. We have reached an era where information acquisition and use determines organizational success. Early and decisive use of information will be one of the major elements separating successful organizations from those that fail, as demands increase for immediate access to information. A significant portion of the information used by organizations is comprised of some geospatial component. The prevalence of this remote-sensing data combined with applications targeted to core business needs will expand the Geographic Information Systems (GIS) market rapidly. To meet the needs of this growing market; scientists, developers, and organizations look for open-source technologies to leverage and capitalize on the value of the derived information.

II. TODAY’S CHALLENGE

There are two main problems that limit the effectiveness of spatial imagery: (A) complex and expensive data management and storage, and (B) an expanding user base with unique data requirements leading to a growing combination of custom datasets.

A. Complex and expensive data management and storage

- Spatial imagery is stored in large files. The cost of traditional storage infrastructure often prohibits the consolidation of these assets.
- Images are stored as datasets with associated geometry files and metadata in tiles dictated by file structure limitations. Users often need to use multiple tiles and all of the associated data in their applications.
- Distribution using traditional file copies and FTP is an inefficient use of resources and network bandwidth.
- Duplicate copies of the data are used throughout the enterprise, which leads to data currency issues — the end users may not have the most recent data stored on their workstation.
- Data can be stored in various locations on various types of media; access becomes limited when the data is difficult and time consuming to retrieve.

B. An expanding user base with unique data requirements leads to a growing combination of custom datasets

- As the breadths of applications grow, end users will find new and unique ways to use existing datasets to answer their questions.
- End users need special tools and special skills to generate the data specific to their requirements from the various source data currently available; the native files are seldom used as initially acquired.
- There are complex, time consuming, repetitive image processing tasks that occur each time an end user utilizes imagery.
Most organizations try to “normalize” data in generic formats that can serve the majority of end user requirements. This tends to lessen the effectiveness of the data for unique projects and leads to considerable and often unnecessary processing time.

End users often need multiple files from multiple sources to answer their questions.

End users require information that is limited to their specific area of interest and often must manage more data than required.

III. AN OPEN SOURCE-BASED SOLUTION

EarthWhere™ is a commercial spatial data provisioning system consisting of a modular, distributed service architecture built on open-source technologies. The framework consists of three application layers: Client Interfaces, an Image Provisioning System, and Integrated Subsystems.

Over the last several years a number of significant open-source software technologies have evolved and matured, enabling the integration of sophisticated archiving, processing, and distributing systems. This architecture is an example of combining the best-of-breed from open-source technologies with commercial technologies.

The application framework is supported by four integrated subsystems, three of which are open source. This combination provides significant leverage, flexibility, and freedom to end users of the system.

The overall architecture of the application exemplifies the typical open-source implementation. Although there are a wide variety of open-source tools to support spatial data management, these are foundation components, which serve specific purposes in a broadly integrated solution. To develop applications with the specificity to solve real-world business problems, there are layers to the solution, which consists of commercial components and custom development. Although open source provides a cost-effective foundation, organizations will still be required to invest in supporting technologies to meet the requirements of their end users.

A. MySQL

MySQL is an open-source relational database used for managing and storing metadata, and controlling application logic. It is the most popular open-source database server in the world with more than four million installations powering websites, data warehouses, business applications, logging systems, and more. Customers such as Yahoo! Finance, MP3.com, Motorola, NASA, Silicon Graphics, and Texas Instruments use the MySQL server in mission-critical applications.

B. OpenMap

For mapping services within the GUI, the system includes the OpenMap mapping server, an open-source JavaBeans based programmer’s toolkit. OpenMap is utilized for navigating through the application where it is used to display data source elements, areas of interest and standard map layers. End users can easily configure the application to add additional mapping layers in support of specific environments.

C. OSSIM

The Image Provisioning System utilizes a server-based image-processing engine that creates the spatial datasets on-the-fly based on user requirements. The image processing service is based on the Open Source Software Image Map (OSSIM) project. The OSSIM project leverages existing open source algorithms, tools, and packages to construct an integrated tool for remote sensing, image-processing, and GIS analysis. This library system provides core image processing, map projection, and geospatial processing software libraries that are used in the system.

IV. SYSTEM-LEVEL OPEN-SOURCE TOOLS

Other system-level open-source tools used in the overall deployment of the system include: OpenPBS, MPICH, and RedHat Linux.

A. OpenPBS

The Portable Batch System (PBS) is a flexible batch queuing and workload management system originally developed for NASA. It operates on networked, multiple-platform UNIX environments, including heterogeneous clusters of workstations, supercomputers, and massively parallel systems. Used as a simple FIFO queuing system, PBS allows for synchronous job processing of massive data product requests. Configured correctly in conjunction with OSSIM, the system is able to always be fully utilized.
B. MPICH

MPICH is a freely available, portable implementation of MPI, the standard for message-passing libraries. Using this library, the OSSIM server-based tools are scalable to support multiple CPUs and clustered nodes for high-throughput environments. The defined image chain for a given data product is brokered out to all available CPUs on the hardware system, leveraging the available processing power.

C. RedHat Linux Operating System

Data provisioning is done on rack-mounted, Dell Xeon-based Server running the RedHat Linux Operating System. Linux offers a cost-effective and robust platform to do advance data management and processing.

V. DESIGN GOALS

Functionally, the system was designed with these criteria:

- Native Data Format Access;
- Parameter based non-destructive processing;
- Optimized for spatial and spectral processing including hyperspectral systems;
- Internally optimized to use reduced resolution data sets and image tiling;
- Enabled for High Performance Computing and Parallel Processing;
- No file size limitations and a large number of open files;
- Spatial and tonal feathering;
- Supports a wide range of map projections and datum transformations;
- Supports a plug-in architecture; and
- Supports various GIS data formats.

VI. CONCLUSIONS

Open-source technologies provide a cost-effective and robust foundation within an organization’s spatial data management strategy. These foundation technologies provide for rapid development of specialized applications; however, to meet the specific needs of the end users, most organizations still need to invest in either custom development or commercial-off-the-shelf (COTS) software. In the development area, open source excels because the wide range of open source development tools allow an organization’s resources to work more effectively to rapidly deploy applications to meet the needs of their end users.

As an organization moves closer to addressing specific business processes, the farther away it moves from COTS and generic open-source software tools. The organization is then left with three options:

- Build;
- Buy; or
- Integrate.

The following diagram depicts an integrated approach to the development of applications using both commercial tools and open source technologies. This approach leverages the low cost and rapid development capabilities within the open source environment as well as the specific functionality within commercial hardware and software components. As a total solution, the result is a solution that meets the requirements of the end users and provides the ability to keep up with the projected future needs of the organization.

This availability of Earth imagery tools, both open source and commercially available, will provide end users of geospatial data with an accessible source of information critical to many business decisions. As the market matures, an Earth Imagery Value chain will evolve to deliver geospatial content to commercial, governmental, and public organizations. By studying this value chain and the data management and distribution requirements of each organization, basic open source technologies will be leveraged to reduce time to market, reduce costs and improve customer service.

Demand for Earth imagery will continue to adapt new technical and business models to service the demand. Innovations in infrastructure and the application of interrelated technologies will provide options for companies as they leverage Earth imagery within their user base. The open-source components discussed here are a sample of the types of innovations that will allow the geospatial market to meet the demands of a growing user base.

REFERENCES