

Development of a Database Middleware System to Support Remote Sensing Analysis over Distributed Data Sources

Alcides Alvear Suárez

Advisor: Manuel Rodriguez, Ph.D.

Undergraduate Students: Angel L. Villalaín G., Jaime Ferrá

Electrical and Computer Engineering Department

University of Puerto Rico, Mayagüez Campus

Mayagüez, Puerto Rico

alcides.alvear@ece.uprm.edu

I. ABSTRACT

This research is an integral part of the TerraScope project being conducted by the Advanced Data Management Group from the University of Puerto Rico, Mayaguez. The central objective of the TerraScope project is the creation of distributed image data bases and the development of tools so that clients can interact with these image databases through the Web. Software tools, data products and other information will be made available to all interested in using it and making the best use of it including the fields of education, scientific research, and its potential applications in the medicine, Biology, engineering, geophysical and other branches of science.

A tool, called the ImageCutter, will be constructed in such away that customized interaction between a client and distributed data bases will be possible. The clients are users who have codes in Matlab and C++ concerning Remote Sensing applications. Examples of these are supervised or unsupervised classification tools. These applications use images stored in any of the data bases available to the user.

The ImageCutter tool is a Middleware solution whose basic characteristic is that it will work through the Web to find processing sites where image processing operators can be executed to refine image and generate new data products. The system will enable a client application to gather data and metadata about images (e.g. MODIS, RadarSat) from a group of image data bases, and request image refinement

operations to be executed. The Image Cutter will enable the client application to send images and customized processing code to clusters or parallel machines, where the image refinement will occur. Examples of these codes are supervised or unsupervised image classification, as well as any code developed in Matlab or C++ in the area of Remote Sensing or Digital Signal Processing (DSP). The ImageCutter will be capable of processing in parallel in cases where it needs to do so to improve the performance and the response times of the routines.

This project is collaboration between the ADM Group and CenSSIS (for Center Subsurface Sensing and Imaging Systems) as a mechanism to share resources and knowledge between the academic institutions that conform the CenSSIS project: Northeastern University, Boston University, Rensselaer Polytechnic Institute and University of Puerto Rico Mayagüez. The primary target of the CenSSIS project is to process and analyze images of the Earth surface, Oceanographic Data, Biomedical imagery, and other types of images that each one of these educative Institutions has obtained from satellites (Modis, RadarSat, Aviris) and the spectral cameras. Many of these data sets are currently kept in magnetic tapes and in an isolated form. In other words, there is little opportunity for technological interchange between these institutions and the coarse information stored in the magnetic tapes. TerraScope is aimed at breaking these barriers, and allowing exchange of data products between these institutions. The research proposed here covers the component of TerraScope

that will enable operations for image refinement to produce new image products.

II. PREVIOUS WORK

Research work made in this area is focused on the creation of data bases Middleware for distributed sources of data [1]. MOCHA is a Middleware system to integrate heterogeneous data bases. It is designed to interconnect sources of distributed data on a network of computers. Typically these systems follow an architecture centered around an integrating server of data (data integration server). This server provides client applications with a uniform view and a uniform mechanism to access the data available in each source. A uniform view of the data is made by the imposition of a global data model on top of the model of local data used by each source. MOCHA is designed to work in high environments and is based on the idea that some of the functions custom-built for the users be moved to remote database system by the Middleware System. This is realized with compiled Java code that puts in execution advanced data types and functions on remote sources of data.. MOCHA has been implemented in Java and it interacts with Informix and Oracle. The results obtained demonstrate that MOCHA not only provides a flexible framework for the distributed requests that it processes, but also improves substantially of the query performance when compared to the existing solutions of middleware.

Microsoft TerraServer is another one of the products developed in this area. Specifically it is a repository of aerial, satellite and topographic images of the Earth, using a SQL data base (Spatial Data Warehouse) available via Internet. It is one of the eight greater online atlases with more than a Terabyte of images from United States Geological Survey (USGS). It uses SPIN-C2 Internet browsers to provide intuitive space interfaces and text of the image data. The users do not need hardware, special software nor knowledge for "leafing through" images. This fact is very important, since this project allows watching the images only. Additionally it describes how the system can decompose large geo-space images (which are difficult to

manipulate) into subimages that are easier to handle. Metadata about the Images is also stored inside the SQL database and available via the Internet.

TerraServer demonstrates that general purpose relational database can handle repositories of large scale images, and that a Web-browser can be a good system for presentation of geo-space images.

Spatial indices for objects with location attributes have also made a contribution in the area. Spatial Indexing has to do with the handling of spatial data in an efficient manner, by using an index mechanism that allows to quickly extract items of data according to its space locations. Thus, image and other objects can be located inside the database based on this index, and this avoids the need of an exhaustive search through the database bank to find the images of interest. It is known that traditional methods for indexing (e.g. hashing and B-Trees are not well suited for objects located in multidimensional spaces. A dynamic index structure called R-Tree organizes the data based on spatial location and gives the search algorithms necessary to find data and maintain the structure updated.

World Meteorological Organization Distributed Databases (WMO-DDBs), also it makes its contribution in this area: The WMO-DDBs project, co-ordinated by the World Weather Watch Department, has been developed to provide access to data that are needed by WMO, and related international, programmer but not routinely exchanged on the Global Telecommunications Systems (GTS). This entry is maintained at the World Meteorological Organization in Geneva, Switzerland. The purposes of WMO are to facilitate international cooperation in the establishment of networks of stations for making meteorological, hydrological and other observations; and to promote the rapid exchange of meteorological information, the standardization of meteorological observations and the uniform publication of observations and statistics. It also furthers the application of meteorology to aviation, shipping, water problems, agriculture and other human activities, promotes operational

hydrology and encourages research and training in meteorology.

III. OBJECTIVES

GENERAL OBJECTIVE

The general mission of this project is to build an infrastructure to execute image processing operations on clusters and parallel machines using the Web as the portal for communications. The requests for image processing operations will be done to different data bases of images from a client application. After gathering the images, the system will remotely execute codes of supervised and unsupervised classification (or any code in Matlab or C++ referring to Remote Sensing or DSP). Figure 1, shows in detail of this a description objective.

Figure 1 shows a client (Matlab or C++) that makes a request. This request is taken by a Servlet in host 1 (web server) who then sent it to a site where a data base with the desired information exists. Once a site is found, the web server sends the URL that provided the request back to the client.

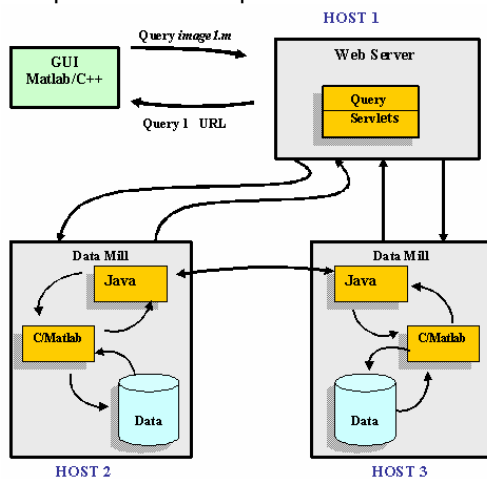


Figure 1. General Model of the Project

SPECIFIC OBJECTIVES

The specific objectives are summarized in Figure 2. A servlet sends the request for an object to a component in the side of the server who is specified by JSP. The component handles the request, retrieving the requested data from the data base and passes an answer object back to the Web Server. In this place the data is formatted

according to the needs of the client. The Servlet in the Web server send the final results back to the user, where he/she can see them, through the Web or of locally, if he/she has decided to make an FTP download of the results.

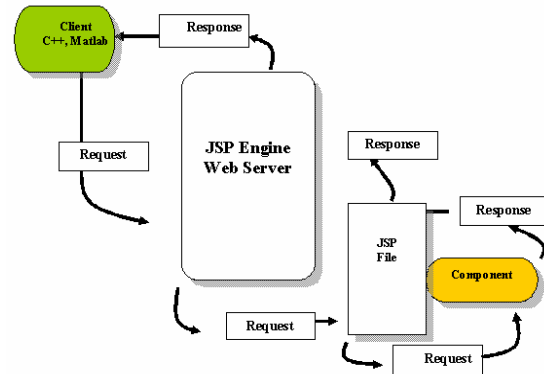


Figure 2. Model with specific details of the project

According to Figure 2, the specific objectives are the following ones:

- Provide a tool for analysis of Remote Sensing images capable of bringing images from remote sources for customized refinement. The tool will allow the user to send code (e.g. Matlab) and the acquired images to computer clusters or parallel machines to perform the requirement analysis and refinement.
- Diminish the computational cost incurred in image analysis by using parallel processing in those cases amenable to parallelism, specially in the supervised analysis of images where the image size is very large.
- To make available facilities on the Web that allows users to search the necessary data in different data bases of images and execute codes in Matlab or C++ that processes the images.
- Facilitate the visualization of the results through a GUI (Graphic User Interfaces) and indicate to the client the address where the data could accessed.
- Allow the comparison of results provided by images with similar characteristics after applying code to them in Matlab or C++.

IV. MATERIALS Y METHODS

The equipment and other materials, used in the project are those that are available in the laboratories of LARSIP and Amadeus. They include five servers (Three in LARSIP, one in Amadeus and Icarus), additionally there are approximately twenty PCs, three laptops and four printers. The materials and equipment that may be needed will be provided by the CenSSIS project. The duration of the project will be approximately one year, where the conclusion of it is in June of 2003.

The methodology to follow in this project will be the following one:

- A bibliographical review will be done to determine limits and comparisons with similar project that allow reaching the objectives drawn up.
- Processing of a general model that allows to have a macro vision of the scope of the project and later to create the specific model that clearly identifies the subsystems that conform it.
- To develop to the algorithms and necessary codes in Java - Servlets, C++, XML and HTML by each subsystem.
- To create a prototype of the system and do the necessary tests with the purpose of improving in those points that are found to be required.

At the moment the following tasks have been made (like localhost):

- Code that creates the Client and the Server.
- Servlet that sends and executes the client Matlab code to one or more hosts.

Pending tasks

- Search for images in the data bases
- Make remote execution of the client codes.
- Divide the images for parallel processing.
- Send results to the client.

- Completion of the prototype.
- Analysis of results.

V. BIBLIOGRAPHY

- [1] Manuel Rodriguez-Martinez and Nick Roussopoulos, "MOCHA: A Self-Extensible Database Middleware System for Distributed Data Sources" Institute for Advanced Computer Studies & Department of Computer Science University of Maryland, College Park, MD 20742, 2000.
- [2] Barclay Tom, Gray Jim and Slutz Don, "Microsoft TerraServer: A Spatial Data Warehouse". Microsoft Research, 1999.
- [3] Guttman Antonin, "R-Trees: A Dynamic Index Structure for Spatial Searching", National Science Foundation and Air Force Office of Scientific Research, 1984.
- [7] Tanenbaum Andrew, "Computers Networks", Third edition Pearson. 1997.
- [8] Ifarraguerri et. al, "Unsupervised Hyperspectral Image Analysis with Projection Pursuit", IEEE Transactions on Geoscience and Remote Sensing, vol 38, no. 6, pp 2529-2538, 2000.
- [9] J.A. Richards, X. Jia, "Remote Sensing Digital Image Analysis", pp. 283-288, 318-319, Springer, 1999.
- [10] Atsushi Ito, Teru Yagi, Nobuo Furuya, Masanori Fujioka, Yoshikazu Ikeda. "Service Processing System using *Distributed Data base*". 1988.
- [13] M. Rodriguez Martinez and N. Roussopoulos. An Architecture for a Mobile and Dynamically Extensible Distributed DBMS. Technical Report ISR-TR 98-10, CSHCN-TR 98-2, University of Maryland, January 1998.
- [14] M. Rodriguez Martinez and N. Roussopoulos. MOCHA: A Self-Extensible Middleware Substrate for Distributed Data Sources. Technical Report MIACS-TR 98-67, CS-TR 3955, University of Maryland, October 1998.