IMPLEMENTING CYBERINFRASTRUCTURE IN SUPPORT OF GREENLAND AND ANTARCTIC SAR DATA SETS

Dr. Linda B. Hayden, Director, ECSU Center of Excellence in Remote Sensing Education and Research Box 672 ECSU 1704 Weeksville Road Elizabeth City, NC 27909 Email: HAYDENL@MINDSPRING.COM

Dr. Geoffrey Fox, Indiana University, Distinguished Scientist and Director, Community Grids Lab Indiana University 501 N. Morton, Suite 224 Bloomington, Indiana 47404 Email:gcf@indiana.edu

Mr. Anthony Adade, Chief Information Officer, Elizabeth City State University, Box 766 ECSU 1704 Weeksville Road Elizabeth City, NC 27909, Email: akadade@mail.ecsu.edu

KEY WORDS: SAR Data, Cyberinfrastructure, Polar Science, PolarGrid, Ice Sheets

This paper will detail the work of the PolarGrid project. The PolarGrid project is designing and building the hardware and software for PolarGrid, which supports data analysis and simulations for Polar Science. PolarGrid federates Grid servers at the North and South poles with both dedicated and TeraGrid infrastructure in the continental USA. Recent polar satellite observations show disintegration of ice shelves in West Antarctica and speed-up of several glaciers in southern Greenland. The great ice sheets in Antarctica and Greenland interact with the global climate in a complex manner, and the impact on global sea level of their retreat would be profound. Most of the existing ice-sheet models, including those used by the Intergovernmental Panel on Climate Change (IPCC), cannot explain the rapid changes being observed. PolarGrid supports the work of the Center for Remote Sensing of Ice Sheets (CReSIS), which is developing new technologies to perform 3-D characterization of ice sheets to understand the physics of rapid changes, and develop models to explain observed changes and predict future behavior. In particular, expeditions this summer will take Synthetic Aperture Radar (SAR) data, which can image the beds of ice-sheets and provide crucial input to models. The project team from Indiana University and Elizabeth City State University (ECSU in North Carolina) has built a PolarGrid laboratory at ECSU and in a recent meeting completed design of the Grid that will be used in expeditions to Greenland and Antarctica in 2008. This expedition grid has a base camp (64 cores and over 25 terabytes of storage) and a highly mobile field camp module. The paper presented at AARSE 2008 will provide details of this important project, which has significant impact on integrating spatial data infrastructure and information & communication technology.

Polar Grid is a National Science Foundation (NSF) Major Research Instrumentation (MRI) funded partnership of Indiana University and Elizabeth City State University. The partnership goal is to acquire and deploy the computing infrastructure needed to investigate the urgent problems in glacial melting.

Major activities associated with this project include building a PolarGrid laboratory at ECSU and the configuration and implementation of the Grid servers that will be used in expeditions to Greenland and Antarctica in 2008 [1]. The PolarGrid implementation and development focuses on the following key areas:

- Field data collection system to be taken with the Polar Science researchers as they collect data [3].
- A base camp 64-core cluster, allowing near real-time analysis of radar data by the polar field teams.

- An educational videoconferencing Grid to support educational activities.
- A large 17 Teraflops cluster to be integrated with the NSF-funded TeraGrid, to serve the polar research community in final processing of data and supporting simulations. This is split between IU and ECSU to support research and education training respectively.
- The implementation of new improvements to the speed and capabilities of ice sheet models, and implementation of web portals to make the data, data products, and resources easily accessible to the polar research community.

Polar Grid's major components and concept of operations are depicted in the fig1.

fig.1

ECSU Information Technology Involvement

The goal of Polar Grid is the creation of a large-scale distributed computing system to aide in the processing of data collected during polar expeditions. The Information Technology Center's (ITC, as it is called) involvement in the project includes building a Teragrid network laboratory on campus, as part of a 17 Teraflops cluster to be integrated with the NSF-funded TeraGrid, and will serve the polar research community in final processing of data and supporting simulations. ITC is responsible for the installation, configuration and implementation of the Grid servers at ECSU. Indiana University will manage the other half of the cluster.

ECSU Campus Network Infrastructure

The general network topology at ECSU is made up of 4 campus core locations – ITC, Doles Hall, Thorpe Administration Building and Trigg Hall, with everything else attached to them (see fig 2). The TeraGrid network will be located at EV Wilkins Academic Computing laboratory.

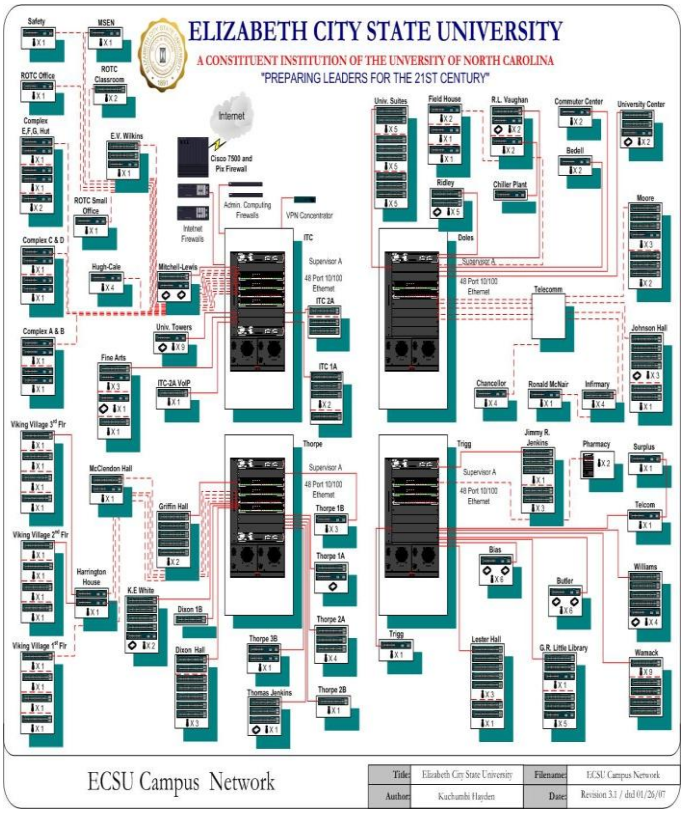

fig.2

ECSU Campus Network Topology

As shown on the diagram, EV Wilkins is attached to the ITC core location (located at the top left corner of diagram). Each technology equipment on campus is connected to those core locations via fiber lines, indicated on diagram by solid red lines. The broken red lines are there only to show paths that are not actually connected. We chose EV Wilkins because the building was originally chartered as a computer laboratory, and adding the TeraGrid matches that original intention. It is also centrally located and easily accessible from anywhere on campus. For these reasons, it made sense to continue using the building for its intended academic objectives.

However, in order to accommodate the volume of technology equipment, the designated room will have to undergo some significant changes. For example, ITC will be working with the campus facilities personnel to remove the carpeting and install tiles or floating floor because the current carpet will introduce unwanted static charge. A floating floor is the preferable selection due to the desired function of the room as a cold/server room. The room will also require installation of a computer room air conditioning (CRAC) unit. The facilities department is proposing a 30 ton CRAC to keep the room at a constant 68 degrees Fahrenheit, utilizing the campus chiller plant to supply chilled water. The room will also need a number of 240VAC 30Amp outlets to supply power to the servers.

Network Connection

The project team, including our partners from Indiana University, recommended that we run a fiber connection to

the Management node to allow a large pipe for data transfers between ECSU and our other TeraGrid Partners. It was recommended that ECSU increase its main data pipe to 10G from the current 1G. ECSU depends on North Carolina Regional Education Network (NCREN) for high volume network connection, which has made available only 1G maximum capacity for data movements. As a result, we need to renegotiate for an expansion to 10G. The initial hardware installation phase of the project will consist of a 64-node Grid and storage array, to be upgraded to 600 nodes at a future date. Besides hardware installation, the connection to the TeraGrid is the second most critical portion of the project. The initial cluster will be designed to only perform processing tasks through the MatLab distributed computing kit. Though MatLab is a very powerful tool, it is also very locale limited in its capabilities. To truly become a part of TeraGrid the job scheduler must be changed from the proprietary MatLab one to Condor-G. Condor-G is an open-source job scheduler which is the preferred tool according to Open Science Grid standards. Once this conversion is complete, a web portal will be developed utilizing another open-source package entitled the Globus Toolkit. Globus is the standard portal software used on the TeraGrid network. At that time an allowance will be made on the ECSU firewall through which the first TeraGrid sharing will take place. TeraGrid users and partner institutions will then either connect to ECSU using the main TeraGrid portal or directly to our servers through our Globus portal. To help facilitate this process Mr. Jeaime Powell, our network technology specialist here at ECSU, has been designated as the TeraGrid Campus technology support representative. This will allow him to have access to resources within the TeraGrid community that will aid the campus in adopting the TeraGrid into other areas besides computer science and mathematics.

Hardware

(2) 21U PG Orange Racks:

A) (1) iDataplex (Expected January Arrival) 4'W x 2'D
B) (1) iDataplex Mgmt (Expected February Arrival) 2'W x 4'D (Normal 42U rack)
C) (1) Liebert Nx UPS (e.g. 2' Wx3'D)
D) (1) Liebert DX 30 Ton(e.g.10' W x 3' D)

Power

Cooling needs to be doubled as the UPS will produce approximately the same amount of heat exhaust as the clusters. So we have to put in place at least a 320,000BTU/HR cooling requirement.

UPS:

A) 80 KVA (72KW)
B) 480 V (3 Phase)
C) Liebert-NX (this is an example of a modular system)
D) 60 HZ, E) 1500 lbs

The design of the cluster has begun through our partnership with the Indiana University (IU). The portal design has been assigned to an ECSU faculty member. We are working with IU to implement this large 17 Teraflops cluster to be integrated with the NSF-funded TeraGrid.

ECSU Polar Grid Lab

The Polar Grid lab located In 111 Lane Hall at Elizabeth City State University (ECSU) serves dual roles as both a computer lab and telecommunications center. Primarily the lab functions as a computing center for CESER faculty and students. The lab contains 16 Apple iMac workstations, which have been configured to allow the use of three separate operating systems depending on faculty or student needs. The combination of Mac OS X, Ubuntu Linux, and Windows XP gives the user the freedom to explore various paths of research interest without worry of system compatibility. Desktop publishing, word processing, image manipulation, web design, geographic information systems (GIS), programming, and mathematical applications have been installed to allow versatility in the role the computers play in lab related activities. The addition of job scheduling applications such as Condor, Globus, and MatLab Distributed Computing Package, also allow the individual computers with dual cores to work together through the network as a single 32-node supercomputer. The computers combined with the installed digital projector, DVD/VCR, podium, and sound system allow the lab to easily transform into a digital classroom for guest speakers or presenters to interact with students and faculty.

The secondary role the lab plays is that of a teleconferencing center. The installation of a Polycom 8000VSX has expanded the sphere of influence from the lab to beyond the walls of the room to connect with other universities and conferences. The Polycom system allows up to four outside participants to dial into the lab, which then becomes a virtual classroom. Each iMac workstation is also fitted with an internal camera through which telecommunications can take place. Through the Polycom system, it has become possible for students to view and interact with technical conferences and meetings that otherwise would have been monetarily unfeasible to attend. The addition of the Polycom system has also allowed faculty members the ability to better exchange ideas with partner institutions in related fields and projects. The addition of the V-brick Encoding system allows partner institutions to view lectures and student presentations in the Dixon Hall lab. This allows larger audiences to participate as the feed is transmitted through the Internet.

The ultimate goal of Polar Grid is the creation of a large-scale distributed computing system to aide in the processing of data collected during polar expeditions. In October of 2008 the initial 64-node cluster will be delivered to ECSU in preparation for the final ~600-node cluster which is scheduled for January of 2009. This cluster will be housed in the ECSU E.V. Wilkins Academic Computing Center. At that time ECSU will be the first Historically Black College/University (HBCU) to house a cluster capable of acceptance into the national TeraGrid Project [2]. The Polar Grid lab will be the center of all operations related to the use and support of that cluster and will grow as new projects, grants, and individuals utilize the cluster through cyberinfrastructure.

PolarGrid 2008 Greenland Field Season

There were three goals assigned to the PolarGrid project in support of the CRESIS project. Those goals were to provide backup facilities for all data collected on the CRESIS flights, provide processing facilities to test data in the field and lastly to send processed images to the University of Indiana for geographical based image feeds.

The members of the CRESIS project utilized a Twin-Otter fixed wing airplane to map the ice sheet layers. They did this with antennas attached to the wings that produced radio waves in the VHF range. They did this to attempt to map the base of the Jakobshavn glacier. The ice sheet itself appears with layers denoting each passing year like the rings of a tree. The layers closer to the bedrock or base of the glacier are the older layers. Those closer to the surface are the newer. By mapping the total thickness of the ice sheet over a period of time, a determination of whether the ice is melting can be presented analytically.

The data from received from six antennas was recorded every 100 microseconds (10e-6 seconds). Over the course of a grid with multiple flights over thousands of kilometers, data was collected constituting hundreds of gigabytes of storage space. Previously this data was not stored in a redundant manner. In other words, if a hard disk was corrupted or failed, that season's data was lost. Through the use of a 30Tb (14 TB usable) RAID 10 storage center, any data stored was mirrored as soon as it was transferred to the cluster using an error checking copy method named rsync. This provided a double redundancy to critical data. A tertiary backup of all data was also accomplished through the use of 2Tb external drive that mirrored each day individually.

The raw data received from the antennas must be filtered and placed in context through visual means. To process the large amounts of data required a rather powerful computer or cluster of computers. PolarGrid provided extended processing capabilities through the use of 8-cores @ 2.33 GHz machine with 16 GB of RAM. Combined with a distributed version of Matlab many processing jobs that may have taken days on a single laptop were reduced to hours. Thereby providing scientists the ability to recommend changes to the RADAR while the season was still active, without being in the field. Half way through the season an additional 16 GB of RAM were added to the cluster for a total of 32 GB to allow even greater processing capabilities.

The final goal of the PolarGrid season was accomplished when actual field imagery was transferred to the University of Indiana Community server, to provide processed data for users to view. The users were able to enter a secure web server and actually view the locations the data was collected while utilizing the embedded GPS tag. The future expansions of the geofeed are only beginning to present themselves for the scientist and projects of the coming years.

In conclusion, all of the goals for this field season were accomplished using the PolarGrid field camp cluster. With the lessons learned and insights gained through the Ilulissat expedition, the expansion of the PolarGrid into future CRESIS missions was solidified. The testing of the fully capable 64-node cluster is scheduled for the fall 2008 Antarctic field season.

References

[1] Hayden, L. G. Fox, and P. Gogineni,
Cyberinfrastructure for Remote Sensing of Ice Sheets,
TERAGRID Conference, 2007 Madison, WI

[2] Polar Grid project to transform U.S. Capabilities in ice sheet research Published November 02, 2007,
http://homepages.indiana.edu/2007/11-02/story.php?id=1646

[3] PolarGrid: Cyberinfrastructure for Polar Science, 2008,
http://cerser.ecsu.edu/polargrid/