ESTABLISHING A BASE CAMP SERVER FOR REMOTE SENSING OF ICE SHEETS IN ILLULISSAT, GREENLAND

Je'aime Powell
Mathematics and Computer Science, Elizabeth City State University
Elizabeth City, NC
jeaimehp@gmail.com

Linda Hayden
Mathematics and Computer Science, Elizabeth City State University
Elizabeth City, NC
haydenl@mindspring.com

Polar Grid is a National Science Foundation (NSF) Major Research Instrumentation (MRI) funded partnership of Indiana University (IU) and Elizabeth City State University (ECSU) in support of the Center for Remote Sensing of Ice Sheets (CReSIS). The partnership goals were to acquire and deploy the computing infrastructure needed to analytically investigate glacial melting. There were three goals assigned to the Polar Grid project in support of the CReSIS project. Those goals were to provide backup facilities for all data collected during the CReSIS flights, provide processing facilities to test data in the field, and lastly to send processed images to IU for geographical based image feeds.

The members of the CReSIS project utilized a Twin-Otter fixed wing aircraft to map the ice sheet layers through Synthetic Aperture RADAR (SAR) technology. This was accomplished with antennas attached to NSF designed wings that produced and received radio waves in the VHF range. This particular range was utilized due to the property of VHF to penetrate the ice layers down to the bedrock. There were a total of 12 antennas with various arrangements of six transmit and six receive points with the data being recorded every 100 microseconds. The layers being cyclical in nature were arranged from eldest closer to the bed to most recent at the surface. By mapping the total thickness of the ice sheet over a period of time analytical data was gathered to determine as to whether the ice sheets are receding or progressing. Within a geographical grid, multiple flights over thousands of kilometers, data was collected constituting hundreds of gigabytes of data.

Polar Grid used clustered computing coupled with RAID storage devices to both process and archive data collected during CReSIS ice sheet measurement experiments. The cluster itself was composed of one IBM x3550 1U server, one IBM DS4200 2U RAID Controller and, two IBM EXP420 2U RAID Storage devices. The controlling server [named C01] enclosed two quad core 2.33GHZ Xeon processors. In the initial configuration the server contained 16GB (8 x 2 GB DDR DIMMs) of RAM and three 76 GB internal serial attached technology attachment (SATA) drives for running the operating system [RedHat Enterprise v.5.1] and user home directory storage. When transferring data from the CReSIS RADAR equipment, a proprietary CPCI developed by the University of Kansas (KU) was utilized. The CPCI box ran Cent OS to simply mount the ext3 partition of the radar drive, and to allow connectivity to the data through the built-in ethernet port. The IBM DS4200 RAID controller and IBM EXP420 storage expanders added an additional 13TB of usable storage connected through fiber channel to the server. The fiber channels allowed the data to be accessed through the largest data pipe available. The added benefit of the RAID storage units was that the combined storage of all 48 hard drives could be accessed as one partition, and redundant storage was handled simultaneously.

Previous missions determined the data storage needs of the cluster. Each standard flight averaged 302GB of data. With an expected 22 flights it was calculated that a minimum of seven TB of archival storage would be needed. An additional minimum of two TB would be needed also to accommodate storage used for processing. Once those specifications were known, the RAID array was ordered, assembled, and configured in a RAID 10 scheme.

The phrase "RAID" is an umbrella term for computer data storage schemes that can divide and replicate data among multiple hard disk drives. RAID's various designs all involve two key design goals: increased data reliability, and increased input/output performance. When several physical disks are set up to use RAID technology, they are said to be in a RAID array. This array distributes data across several disks, but the computer user and operating system see the array as one single disk. As there are no basic RAID levels numbered larger than 9, nested RAIDs are usually unambiguously described by concatenating the numbers indicating the RAID levels, sometimes with a "+" in between. For example, RAID 10 (or RAID 1+0) consists of several level 1 arrays of physical drives, each of which is one of the "drives" of a level 0 array striped over the level 1 arrays.1

As tertiary forms of storage, external two TB MyBook drives were utilized to both archive daily flight data, as well as a way to physically transport data back to partner institutions. The Mybook drives were selected primarily due to the data capacity each drive could store. Each 2TB drive had the capability of holding up to four flights worth of data. The drives accomplished this by utilizing three 750GB hard drives in a RAID 0 or stripped configuration to allow all three drives to be utilized as one uninterrupted storage partition over ext3. The MyBook drives were connected directly to the USB ports of C01 and utilized a protocol/utility named Rsync transfer data.

The Rsync utility was originally designed to aid in the transfer of data through low-bandwidth high-latency bi-directional communications links. The algorithm identifies parts of the source file, which are identical to some part of the destination file, and only sends those parts, that cannot be matched in this way. Effectively, the algorithm computes a set of differences without having both files on the same machine. 2 For this reason Rsync was identified as the optimal way to not only transfer data from one source to the other, but to also verify the integrity of that data during the transfer. The Rsync utility also had the added benefits in that it could be utilized over a local network address, a local network file system (NFS), a local system port, or a remote network location.