Acquisition of a Multi-Core Server for the Enhancement of the Meteorology Program at Central Michigan University

Leigh Orf and Marty Baxter, Department of Geology and Meteorology, Central Michigan University

1 Project Summary

The meteorology program at Central Michigan University, a UCAR affiliate institution, has long relied upon Unidata products and software for classroom instruction as well as undergraduate and faculty research. Faculty within the program have served on Unidata advisory committees and are heavily invested in Unidata's mission and future.

The program has experienced steady growth throughout the past decade and this growth has been mirrored by an increase in the computation resources required to successfully run the program. Current facilities include a dedicated instructional lab with 20 workstations running Linux and the full suite of Unidata software products, as well as an undergraduate research lab consisting of four workstations acquired by a previous Unidata equipment award.

This hardware request addresses a much needed upgrade to an aging, overloaded, and increasingly unstable server which, as of late, has been locking up and requiring a cold restart on a nearly weekly basis. The proposed hardware will enable the program to begin the development of a case study archive which will be used for instruction and which will be shared with the Unidata community. Using a modern multi-core architecture in conjunction with fast local storage, the server will act as both an ingestion engine running the LDM and its associated converters, network-mounted storage for GEMPAK software, and a RAMADDA case study server. Hardware from previously acquired Unidata equipment grants will be moved over to the new server in order to facilitate backups and long-term archives of case study data.

2 Project Description

2.1 Equipment Requested

The machine `waterspout.cst.cmich.edu`, which was acquired with a Unidata equipment grant in 2004, contains a 32 bit dual-processor Xeon architecture and has 4 GB of memory. Over time this machine has taken on additional roles as a backup server, home directory server for student accounts, THREDDS server, as well as all of the tasks required for an LDM ingestion server. The machine regularly maintains a system load exceeding 3 which indicates it is regularly CPU bound.

The chosen hardware configuration for replacing `waterspout` are as follows:

- PowerEdge 2970 server with two six-core AMD Opteron processors, 2.8 GHz, with 6 MB cache
- 32 GB DDR2, 667MHz DIMMS
- Eight 500 MB 7.2K RPM Near-Line SAS 2.5 inch hotplug drives in a RAID configuration
- Dual embedded Broadcomm NetXtreme II 5708 Gigabit Ethernet NIC
- 1 SATA DVD-ROM drive

While the clock speed of each individual core is less than that of the current hardware, it is anticipated that the dual six-core architecture will dramatically increase the effective processing power of the server which regularly is running multiple threads simultaneously.

In order to logically partition traffic, one of the NICs will be dedicated towards ingesting and serving LDM products over the internet, while the second will cover all local traffic on CMU's LAN. The partitioning of the fast local disks, as well as the slower but larger-capacity RAID 5 SATA disk array acquired from a previous Unidata allocation, will be configured with consultation with the department's network administrator who has been maintaining the program's hardware for over a decade.

Currently student Linux home drives are NFS mounted to `waterspout`, which produces a noticeable amount of CPU overhead due to activity with the NFS daemon. While we envision eventually transitioning to a system where student home directories are on a separate server maintained by the College or University, this is contingent upon a college-wide transition over to a more unified system which has not yet been completed. However, should this transition occur, the GEMPAK environment requires NFS mounts from student workstations and we anticipate still a fair amount of NFS overhead on the server. We believe the large amount of cores will best facilitate the multiple functions of the proposed server.

2.2 How the equipment will help meet the goals of the project

One of the goals of the project is identical to our previous goal of having a stable, fast system to facilitate student and faculty use of real-time data ingested via the LDM, as well as the visualization and analysis of this data through tools such as GEMPAK and IDV. The second goal, which we believe is truly in reach, is the creation of a case study archive generated by students and faculty, some of which may be accessed remotely by the Unidata community. It has long been our desire to create such a case study archive, and both Baxter and Orf worked to create a proof-of-concept test case of using THREDDS to serve the COMET "Storm of the Century" case study to IDV. This process proved to be quite challenging and was met with several pitfalls, some of which are summarized in a report presented to the Unidata at a Users Committee meeting (see: `http://www.unidata.ucar.edu/committees/usercom/2008AprMtg/statusreports/orf-thredds-spring2008.pdf`). This proof-of-concept project was one of the stated outcomes of Baxter's previous Unidata equipment allocation. Shortly following this report, development began on RAMADDA, which has been installed and tested on our current server. We are excited about RAMADDA as it is simple to use, allows for the storage of any number of types of data including images, text, metadata (and of course Unidata products) and which has the capability of allowing users to upload and create their own case studies. The proposed server will serve as a repository for the case study data and will run the RAMADDA server, which should only add a marginal amount of overhead. It is anticipated that IDV will serve as the primary window into the case study data. We envision having students and faculty create case studies which will be used in upper level courses. Selections from this collection will be made public to the Unidata community.

The types of data which we envision comprising a case study include but are not necessarily limited to: Model forecast and analysis data, upper air, satellite, radar, and surface data, as well as text products and contextual information. We envision working with Unidata programmers in order to find the most logical way to reach our goal of creating the case study archive. One proposed method to create much of a case study would be to load all products of interest into IDV (from various sources, including our own server) and save a zipped bundle (`.zidv`) file containing all of the raw data, not links to the data sources. This data could then be shipped to the RAMADDA server where it would remain and be supplemented. As the case study archive grew from both student and faculty input, the highest quality, scientifically interesting and best crafted/complete case studies could be "blessed" by faculty and made public to the community. It is anticipated that an entry portal would be developed to the RAMADDA case study archive in order to give remote users the proper context into the data and to state the goals of the case study archive project. In effect, we envision roughly reproducing the efforts of the COMET case study archive, but using RAMADDA with IDV as the intended primary window into the data rather than GEMPAK.

In response to our heavy reliance upon student competency on computers, a new course (Computer Applications in Meteorology) has been developed. This course is geared towards juniors who are concurrently taking Synoptic and Dynamic meteorology, and gives them the skills they need to use the Linux operating system as well as Unidata software. A section on the use of RAMADDA to create case studies would be an addition to this course and these skills will be used for the creation of student case studies in both Synoptic Meteorology and Mesoscale

Meteorology, and perhaps Numerical Weather Prediction. It is anticipated that archived model forecasts could be served through RAMADDA and serve as an educational tool on model verification in the NWP course.

2.3 Benefits to research, education, and the Unidata community

This hardware will have an impact on all of the upper level meteorology courses including, but not necessarily limited to: Computer Applications In Meteorology, Synoptic Meteorology, Mesoscale Meteorology, Meteorological Radar and Satellites, and Numerical Weather Prediction. The Meteorology program at CMU attempts a balance between theory and application, and having a central server for all the Unidata products and a growing repository for case studies is of great importance in obtaining the desired student outcomes in these courses. We currently use GEMPAK and IDV in the classroom and this will continue. The new benefits of having a case study archive for Synoptic and Mesoscale Meteorology include dramatically transforming sections of each class. Students in Mesoscale Meteorology are required to complete a case study as part of their course grade. This process typically involves downloading images and text from sites such as NCEP and SPC and other sites which archive synoptic and mesoscale meteorological data. We have found GEMPAK and IDV are best suited for visualizing and analyzing recently ingested real-time data and that it is prohibitively difficult for students, and to an extent faculty, to create case study data which can be viewed and analyzed with Unidata software. We intend on heavily utilizing RAMADDA as we have long wished to have the capability to create and share case study data. Case studies are also excellent topics for undergraduate research projects, and RAMADDA will undoubtedly be put into service for this purpose as well.

Faculty in the meteorology program are active in research in both synoptic and mesoscale meteorology. These faculty are users of GEMPAK and IDV in their research and anticipate using RAMADDA for storing and sharing research data. Concerning benefits to the Unidata community, this should occur in several ways, including: creating community access to faculty- and student-created case studies; serving as a feed site for other universities; and providing valuable feedback to Unidata programmers when the inevitable glitches turn up. Both Orf and Baxter have in the past served in an advisory capacity to Unidata concerning Unidata software (most frequently IDV as it has evolved the most over time), and we anticipate continuing to serve this role as dedicated users and educators heavily invested in the success of Unidata.

2.4 Relationship of proposed hardware to existing facilities

In 2009, the meteorology program moved out of the Geography department and into the Geology department, which is now called the Department of Geology and Meteorology. As a consequence of this change, the program was moved to a new facility in the Engineering and Technology building where a new meteorology teaching lab was constructed. This teaching lab, which is used for all upper level meteorology lectures and labs and is open to students when class is not in session, can contain 30 students and has 20 Linux/Windows workstations.

The lab includes both a regular 2D projector and a GeoWall projector for stereo projection. Outside of this lab is a section of a larger room which has four Unidata-supplied workstations for student research and includes table space for studying and research.

The addition of 10 computers relying on `waterspout` which were added following our move has created an increased load which has impacted its performance and stability. It would leave much of our program in an uncertain state should we be unable to address these increasingly serious instability issues which are currently experiencing with `waterspout`, and this is a strong motivator to this equipment request. The proposed hardware will impact the entire operations of the meteorology program by eliminating the instability and performance issues we've experienced as of late, as well as creating a new powerful learning tool via the creation by faculty and students of a case study archive.

Access to `waterspout` within the university will occur over gigabit LAN as it has in the past. `Waterspout` will be housed in a new facility which has been recently created in order to accommodate university servers and faculty research clusters. This facility has climate control and redundant power backup. An external SATA RAID array acquired with a previous Unidata equipment grant will be moved with `waterspout` to the new facility. The Quantum TD2 DLT tape drive which has been used for backups will be moved to another more accessible machine and will continue to be used for backups including case study archives.

2.5 Relationship to existing computing resources

Two meteorology faculty have access to three local research clusters obtained with both startup money and an NSF MRI allocation. `Waterspout`, even in its upgraded configuration, will serve as a relatively small percentage of the total computing power; however, its central role in maintaining the meteorology program cannot be overemphasized. Geology faculty have modest computational needs and their facilities are currently housed in a separate facility and are maintained by a different system administrator.

We feel `waterspout's` role is not so much as a computing engine as it is on a data processing and serving system. It will not be used as a platform for running, for instance, numerical model simulations, as the aforementioned hardware (and Teragrid resources) serve this purpose. By dedicating it to a specific set of tasks, we feel it will provide many years of stable operation and will serve as an important role in the continuing evolution of the meteorology program at Central Michigan University.

3 Budget

Attached is a quote from the Dell Online store conducted on March 30. We have chosen a configuration which should address performance issues we have had with our existing machine. Twelve cores should cover a system load involving ingestion and conversion of Unidata products as well as several threads of the NFS daemon running concurrently. 32 GB of memory should allow for a large amount of caching of data which is requested repeatedly, which is a common situation when a classroom of students is interacting with a shared dataset. We are requesting a total of 4 TB of raw storage, which will likely be converted into a RAID5 array. Having fast read access to data is crucial when several students are accessing the server concurrently while at the same time converters are writing out data. In consultation with our system administrator, we will also consider an alternative RAID configuration if RAID5 write performance is considered to be an issue. This combination of parallel processing power, large amounts of shared memory, and a large array of fast disk access should serve the department well into the future.

The base configuration of the PowerEdge 2970 contains two quad-core AMD Opteron 2.2 GHz processors, 4GB of DDR2 RAM, and one 146 GB hard drive, and is $4,293. The following are upgrades or additions to the base configuration:

- Two six-core AMD Opteron processors, 2.8 GHz, with 6 MB cache [Adds $2,000]
- 32 GB DDR2, 667MHz DIMMS [Adds $655]
- Eight 500 MB 7.2K RPM Near-Line SAS 2.5 inch hotplug drives in a RAID configuration [Adds $3,330]
- Dual embedded Broadcomm NetXtreme II 5708 Gigabit Ethernet NIC [included]
- 1 SATA DVD-ROM drive [Adds $30.00]
- Operating system: None [Will install CentOS, which is free]

The request with the above configuration totals $10,311.00.

4 Project milestones

Hardware will be purchased immediately following the receipt of funding. Installation will occur shortly following receipt of hardware by our system administrator. Installation of CentOS and configuration of the LDM will follow, as well as the configuration of student accounts. This would bring us up to speed to the configuration of our current hardware (except it would be running a much more modern operating system).

Should the award be received before the Fall 2010 semester, the machine should be ready to go online with all configurations including RAMADDA. Creation of the RAMADDA case study archive will occur gradually. It is anticipated that faculty will familiarize themselves with RAMADDA and its capabilities during the Fall 2010 and Spring 2011 academic year. Due to an upcoming sabbatical, there may be a delay in the integration of student-generated case studies in Mesoscale Meteorology until the Fall 2011 semester. By Summer 2012, it is anticipated that a case study archive will have been generated by both faculty and students and that selected case studies will be made available to the Unidata community. Over time the case study archive will grow and more will be made public. At the same time, `waterspout` will continue to ingest and serve data to the Unidata community as needed.