

# September 2011

## Data Analysis Services Group - September 2011

### News and Accomplishments

DASG welcomes visitors Sang Myeong Oh and Minsu Joh from the Korean Institute of Science and Technology Information (KISTI). Sang Myeong will be a visitor through Nov. 30, and Minsu through Oct 14. The two will collaborate with the VAPOR team in support of global ocean circulation model data.

### VAPOR Project

Project information is available at: <http://www.vapor.ucar.edu>

#### TG GIG PY6 Award:

Yannick completed testing and development of the PIOVDC library. The code has been integrated into IMAGE's turbulence modeling code, GHOST, and we are now working with IMAGE researchers to evaluate the work. The next steps are to clean the code, and provide user documentation and examples.

#### XD Vis Award:

John gave a tutorial on VAPOR at the Front Range High Performance Computing Symposium at the Co. School of Mines on Sept. 23. The talk was attended by ~20 students and faculty.

#### KISTI Award:

KISTI contract work is now underway and, due to the short period of performance for year one, will consume most of the team's attention. Efforts in September focused largely on analyzing a number of representative MOM4 data sets, and developing a requirements document for compliant MOM4 data. Currently, there is no formal documentation describing the contents of a MOM4 data set. Thus the requirements must be reverse engineered.

Our application to CISL's RSVP program was approved, and we welcome Sang Myeong Oh from Korea's Jeju National University. Sang Myeong will be a visitor at NCAR until Nov. 30. While here he will be helping support the KISTI/NCAR collaboration on VAPOR development. Also visiting for two weeks is Minsu Joh, the director of supercomputing applications at KISTI.

Lastly, Karamjeet Khalsa, our SCIParCS student, excepted a full time position elsewhere, and is no longer working on the KISTI project. We wish Karamjeet the best of luck and thank him for his contributions.

#### Development:

A release candidate for VAPOR 2.1 was completed and made available for download on the VAPOR web site. We expect to have a production release of 2.1 completed by the end of the calendar year.

We continue to prepare for the 2.1 release of VAPOR. A decision was made to, for the first time, distribute a beta release of the package before issuing a stable, supported version: a common practice for open source software. The beta will allow us to get important bug fixes into the hands of users who need them more quickly, and will also let us enlist the help of the user community for testing. Mac and Linux installers were created for the release candidates and all VAPOR staff were involved in regression testing. A new, more formal test plan was developed for these purposes.

#### Outreach and Consulting:

John and Alan gave an informal tutorial on VAPOR to the WRF-CHEM group, organized by NCAR's Mary Barth.

## Data Analysis & Visualization Lab Projects

### File System Space Management Project

- Continued work on FMU design and documentation.
- Started the conversion of the FMU documentation from OpenOffice format to the DocBook 5 format.
- Continued study of XML Stylesheet Transforms, looked at differences between XSLT versions 1.0 and 2.0. Some DocBook 5 features require an XSLT 1.0 tool, but XSLT 2.0 features could be useful for automated report generation.

### System Monitoring Project

- Configured Nagios to be able to send remote host/service checks to a central Nagios server. NETS has our config files and once they are added to the central server we can start testing.

## CISL Projects

### GLADE Project

- To enable the further tests on GPFS clusters for data migration, we decided to shrink the current "elustre". By shrinking it, we were able to test unusual resizing features of Lustre as well as getting enough LUNs for test GPFS. Using "lfs\_migrate", we emptied the 8 target LUNs and packed

the current data to a subset of OST targets. The operation was successful and we added GPFS software to the stratus servers. With this setup, we don't have to reconfigure FC LUN setups for future changes.

## ***Lustre Project***

- Whamcloud and OpenSFS made the first official release of Lustre2.1. One notable enhancement is the large size LUN support. (from 16TB to 128TB) Test setup on the machine "swtest1" was upgraded to the new release. All previous data were preserved and 1.8 client was able to mount it as expected.

## ***Data Transfer Services Project***

- Worked with Steve Beaty to setup a test Globus myproxy server to issues temporary proxy certificates based on a user's UCAS token response for use with the GridFTP servers. Enabled Sidd Ghosh to use these tokens for testing both directly and with Globus Online without requiring a user X.509 certificate issued by another certificate authority.
- Briefly read some documentation on an HPSS virtual file system (VFS) layer Linux kernel module provided by Sidd in the possibility that we could use it to more easily enable recursive GridFTP access to the HPSS archive. Referred him to MSSG for implementation approval and testing prior to trying it out on the **datagate** systems.

## ***Lynx Project***

- Taking advantage of Lynx Lustre setup change when SSG group enabled quota on it, we took a snapshot of current usage. No surprise was found: only three users were using more than 1TB, which is the individual quota limit. One top user was using 38% of the entire file system, though.

## **Publications, Papers & Presentations**

- John authored an invited chapter entitled "Progressive Data Access for Regular Grids" for an upcoming book on High Performance Visualization, edited by Wes Bethel (LBNL), Hank Childs (LBNL), and Charles Hansen (U. of Utah).

## **System Support**

### ***Data Analysis & Visualization Clusters***

- In daily log messages delivered via mail, we noticed significant volume of dictionary scan type activity. Though OTP authentication will prevent compromises, we decided to discourage intruders further by installing "fail2ban" which scans the repeated failed login attempts and temporarily blocks the offending IP addresses via iptable. After tests on single node, we later deployed it to all mirage and storm nodes. As a result, volume of daily messages went down accordingly.
- Opened ports on the storm/mirage firewalls to allow connections to the WEG SQL server.
- Removed the dasg025 project space.
- Worked to get HTAR working on the DASG systems.
- Exchanged email several times with EMC about the Networker backing up too much stuff. Asked EMC why no one seems to be able to access the description and supporting information in the service request when the problem is passed to yet another support engineer and I have to describe the problem details yet again.
- Updated the OpenSSH packages on DASG systems.

### ***GLADE Storage Cluster***

- Created a cgdamp group with access to the cgd/amp directory under proj2.
- We added the **polynya** remote cluster and successfully mounted the GLADE filesystems. Polynya is the large memory, gpu enable cluster purchased by ReSET.
- Removed 6 LUNs from the proj0 file system in order to grow the user file system.

### ***TeraGrid Cluster***

- **Twister2** crashed and the the power button was blinking 4 times, which initially led to the power supply failure. The warranty replacement module did not change anything, however. Following elimination method for peripherals, the cause was identified as the Graphics card that failed. We pulled the equivalent unit from the machine **nomad** and put **twister2** back in service.

## **Other**

- Participated in the interviews of the internal candidates for the NWSC SA1 position.