June 2013

Data Analysis Services Group - *June 2013*

News and Accomplishments

VAPOR Project

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

XD Vis Award:

John and Alan are working on plans for refactoring the VAPOR code base, with the goal of providing a more flexible architecture in a future VAPOR release.

The XD Vis Award quarterly report was submitted.

KISTI Award:

Miles completed the first KISTI milestone: the EasyThreads class object has been implemented under Windows, using the native MS Windows thread API. Thus all components of the VAPOR code base that are implemented for parallel execution can now take advantage of multiple cores on Windows platforms.

Miles started work on a program to "repair" POP data sets contain invalid data over land masses: a known bug in POP that will hopefully be addressed in the next POP release.

Scott Pearse joined the team the second week of June and has been actively working on the KISTI project in several areas:

- Experimenting with Vapors user interface and learning the general utilities that it implements This was done early with online documentation as well as during my preparation for Alan's WRF presentation
- Online research of ESMF structured and unstructured data and interpolation methods. Mohammed has been a primary resource for me and has given me clear direction on this topic.
- Developing a UI for netCDF data conversion in Vapor Scott's initial implementation was agreed to be too cumbersome as it presented more data conversion options than were necessarily needed We agreed that a wizard approach would be more suitable where advanced options could be displayed if the user needed them.

2.x Development:

Alan continues to work with technical support at IDL, to understand the bug we encountered with running IDL on Windows with VAPOR.

Miles moved the VAPOR source code repository on sourceforge from the outdated CVS to GIT. He also assisted VAPOR staff getting spun up with GIT.

Miles prototyped a "version checking" capability that would notify VAPOR users when new versions of the software were available. He is now working with Alan to integrate the capability into the package.

A video conference call was held with the visualization team at DKRZ in Hamburg to discuss possible extensions of VAPOR to support the ICON climate and weather model. ICON, like NCAR's next generation WRF model, employs an unstructured grid. Unstructured grids are currently unsupported in VAPOR. The team is exploring the level of effort required to handle these data.

John fixed a number of bugs in the 2.2.2 release:

- vaporgui confused different variables when a new data set was loaded into an existing session file
- in some situations coordinate variables could only be accessed in compressed form.
- When loading new data into an existing session file the number of time steps between the session file and the data set had to agree.
- The udunits data base was not installed on Linux and Mac systems.
- The probe could in some instances return data from the wrong variable.

3.0 Development:

Development of the 3.0 version of VAPOR was kicked off in June. This will be a major refactoring of the code's internals in an effort to: 1) improve maintainability, 2) facilitate extensibility; 3) enable support of new features that are not possible or practical with the current architecture (e.g. unstructured data). A number of planning meetings were held to identify requirements, scope level of efforts, and prioritize activity areas. First draft API specifications were developed by John to support a more flexible external data format, and to allow separation of the GUI from the *vaporgui* internals.

Development of VAPOR 2.x will proceed in parallel with VAPOR 3.0.

Website:

Alan and John completed numerous improvements to our Drupal documentation on the website that will enable VAPOR users to access context-sensitive documentation directly from the vaporgui application (planned for VAPOR 2.3).

Miles started learning Drupal and updated several out of date modules on the web site

John organized a meeting with Steve Geinosky (CISL Outreach), Mary Haley (NCL), and Tim Hoar (DART) to discuss the possibility of establishing a consolidated software products web presence on the CISL web site. A common site would benefit both users of these products, as well as the developers. All at the meeting were in favor of moving forward. Steve will present road map options at the next meeting.

Administrative:

The annual report for the groups use of an NSF Major Research Instrumentation Award (Janus) was submitted.

Education and Outreach:

- Alan gave a VAPOR mini-tutorial at the 2013 WRF Workshop. The students were very interested and the course was well-received.
- John participated in a meeting with Dr. Judith Terrill, the leader for the High Performance Computing and Visualization Group at NIST to discuss possible collaborative opportunities.

Software Research Projects

Climate data compression:

John completed work on a HOMME-specific compression utility, based on VAPOR's compression library. A sample set of CAM-SE data was compressed, and then decompressed, and the resulting data returned to John Dennis' group for comparison with other compression methods.

SCIparCS Internship

Mohammad Abouali, a Siparcs intern from San Diego State, has been developing extensions to NCL to enable WRF scientists and others to convert their data so that it can be visualized in Google Earth. Mohammad and Alan met with Gabi Pfister, Mary Barth, and Cindy Bruyere to understand their requirements. Based on this feedback, Mohammad made extensions to NCL that enable users to create KML files and images from NetCDF files. Users can map their data to color and opacity and animate the data over time. Mohammad also has provided capabilities of creating polygonal shapes such as wind arrows, flight paths, and shape files. Whenever performance is an issue, Mohammad has implemented his code in C, overcoming NCL slow looping performance.

Mohammad has shown his work to potential users, in several seminars, and his work was very well-received. He showed his work to the NCL team and they decided to support inclusion of his enhancements in the next NCL release (probably this fall).

Production Visualization Services & Consulting

Pete Johnsen of Cray has computed a 50TB simulation of hurricane Sandy using Blue Waters at NCSA. VAPOR visualizations of this data, on a 500m grid, will be shown at several AMS workshops and at SC2013, demonstrating the scientific value of petascale simulation.

Mel Shapiro is preparing a presentation at an AMS Mesoscale Workshop where he is the featured speaker. He is working with Alan to prepare visualizations that illustrate various principles of understanding weather dynamics in severe storm systems.

Prof. Shuyi S. Chen, of the University of Miami, testified before the U.S. Congress on June 26, on the value of improved communication between NCAR and operational weather forecasting. Alan provided her with VAPOR animations of hurricane Sandy, which she used as an example of advanced capabilities at NCAR not yet incorporated into operational forecasting.

John worked with a user to sort out problems with parallel use of ParaView on Caldera. It turns out that the parallel version of the code had not been configured correctly.

ASD Support

John continued working with Cornell's Peter Ireland on turbulence visualization.

Publications, Papers & Presentations

- Continuing work on contributions to the CRC book chapter on yellowstone. Completed diagrams and continued to work on draft text.

Systems Projects

Accounting & Statistics

- Started putting together some Glade benchmark jobs using gpfsperf and mdtest that can be submitted to Yellowstone.

Security & Administration

- Started study and familiarization with the C++ 2012 standard and GCC 4.8.1 for use with future code projects including the File Management Utility. Read up on the current state of lock free programming techniques for use in the FMU.
- Oversaw the upgrade of the DataDomain DD670 server located at ML and the installation of a second DD670 located at NWSC. The DD670s are configured to provide remote site replication of backups. Additional space is now available for the /glade/u/home and DASG managed server backups, which allows for more than two weeks of backups to be kept. The DD670s are also used for the consolidation of HPSS managed server backups (active) and SSG managed server backups (pending).
- Installed git on the vapor systems and orthrus.

General Management Tasks

- Creating project structures for joint HSS-CASG tasks. Currently focusing on monitoring of systems both from a facility standpoint and from an automated monitoring standpoint.
- Creating a software project structure to be used for CASG software projects in support of GLADE. This will allow effective mentoring of CASG staff by DASG in regards to software tools developed by CASG for monitoring and responding to issue within GLADE.
- Began discussions with IBM on the expansion of GLADE in early 2014. We are looking at any options we have during the upgrade to enhance either the performance or the stability of the current system. These include features available due to firmware updates, features available due to software updates, and any configuration changes we can make to enhance the environment.

System Support

GLADE Storage Cluster

- Continued monitoring GPFS to understand normal and abnormal operating conditions and how that is reflected in the monitoring tools output.
- Restored /glade/u/home files for users upon request.
- Restored the VAPOR files tree from backups.
- Restored old /glade/home user directories upon request for users who failed to migrate their files to the new GLADE directories before the old GLADE storage was shutdown.
- Upgraded the drive and controller firmware on the DCS3700 RAIDs to more current versions. The drive firmware update for the Hitachi drives addresses a possible data loss issue.

Experimental Clusters

- Completed an initial installation and configuration of an OpenStack Storage cloud for testing.
- Reinstalled oasis0-3 and stratus0-3 with RHEL 6.4. These are server systems for the storage test bes.
- Began getting stratus2 prepared for a Lustre installation.

Test Clusters

- Upgraded GPFS on picnic to 4.3.0-21 and then to 3.5.0-10. Produced an RPM for distribution to client systems.

Storage Usage Statistics

dasg:NWSC+GLADE+Usage+Report