

May 2011

Data Analysis Services Group - May 2011

News and Accomplishments

VAPOR Project

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

TG GIG PY6 Award:

Yannick completed development and documentation of the VDC extensions for PIO, and integrated the new interface into IMAGE's GHOST code for testing. A bug in PIO's 3D data decomposition handling code was turned up and we are now waiting for John Dennis and his team to resolve it.

Development:

The VAPOR team met with Kenny Gruchalla and made plans to integrate his 3D geometry display code into the current code base. The code supports a variety of 3D scene description file formats and will provide a means for users to import and display complex shapes (e.g. buildings, terrain, etc.)

Alan implemented a number of WRF/Python derived variable methods, converted from NCL, to be provided in our next release of Vapor. These were requested by Sherrie Fredricks from MMM. The implementation required significant changes from the original FORTRAN versions to take advantage of the efficiency of NumPy on array operations. We are now determining how to package these methods in the release installation, probably as a separate module.

A Tab-ordering preference feature was completed and tested, and checked in.

Kendall and Alan created a set of color maps for vapor that are based on the the NCL color maps most widely used in the WRF community. These will be distributed with the next release.

We held discussions with the NCL team to make sure our API will be compatible with other Python APIs. Alan also met with Joe VanAndel (RAL), to get his recommendations for structuring our Python environment. Joe had several good suggestions that we plan to incorporate into the VAPOR Python environment.

Alan modified the Python API to work better with 2D variables derived from 3D variables, such as those needed by WRF.

Alan has been working with Rich Brownrigg to enable VAPOR to correctly support the rotated lat/lon mapping projection for WRF data sets. The rotated lat/lon projection is recently getting more usage especially for global WRF. Our implementation of this projection is proving more difficult than we first thought because the documentation is poorly written, and because the WRF implementation is different than other uses of this projection.

John rewrote the shaders used by VAPOR's ray caster to address several bugs, and changed the compositing order from back-to-front to front-to-back, which was necessary to support *depth peeling* (rendering of semi-transparent surfaces without performing a visibility sort). The most significant bug was the use of a deprecated GLSL function that is no longer supported under some newer nVidia drivers. The accuracy of depth comparison used in hidden surface removal was also improved, and an incorrect calculation in the Blinn-Phong illumination model was fixed.

John and Kendall finally resolved a problem compiling VAPOR code and 3rd party libraries to be backwards compatible with older versions of Mac OSX (pre 10.6).

John completed development of a prototype Regular Grid abstract data class, benchmarked its performance on large grids, and presented results to the team. The intent of the new class is to replace VAPOR's current array-based internal data representation. The higher-level abstract data type will provide a cleaner, unified data handling model and will address needed capabilities such as missing data, and stretched grid support. Moreover, thanks to Moore's law the performance penalty paid by using C++ methods and operators to access gridded data is no longer significant. A draft specification for the complete API is now being reviewed by the team.

Yannick began prototyping depth peeling in VAPOR. The goal is to be able to correctly handle rendering of semi-transparent objects across all of VAPOR's renderers. The prototype will only address one renderer: the flow visualizer, which uses simple OpenGL fixed functionality rendering. If successful, the next step will be to look at VAPOR's more advanced renderers (e.g. isosurface), which rely use OpenGL shaders instead of fixed functionality rendering. In preparation for the latter Yannick is moving the current shaders, which are hard-coded into VAPOR, into run-time loadable, user-editable files.

Alan and John have begun planning for the next release of VAPOR, version 2.1, tentatively targeted for late summer or early fall.

Outreach and Consulting:

We met several times with Mel Shapiro regarding his ERICA storm analysis. Ryan Maue, Shapiro's collaborator at the Naval Post graduate school, ran several WRF simulations of this storm. We downloaded several terabytes of this data and performed several analyses of this data to exhibit some of the unique features of the storm. An example movie of radar reflectivity is posted at <http://vis.ucar.edu/~alan/shapiro/ericadbzmax.mov>. There are still a few defects in this data that we are trying to repair.

We provided a letter of commitment to colleagues at Indiana U. and Argonne lab, agreeing to work with them on immersive visualization of flow.

Ilan Levy found there were problems in obtaining graphics drivers for the ATI cards in the CTTC machines. He decided that he will instead support Windows for our course at the WRF workshop.

We decided to provide a VAPOR poster for the 2011 WRF workshop, which Alan is writing. This will highlight the new capabilities we developed for version 2.0, such as the Python support.

John and Alan provide a vapor demo for some visitors from the Network Startup Resource Center on May 27, as requested by Rich Loft. We agreed to support a vapor demo talk at a planned August workshop on African weather and climate.

KISTI Proposal:

We received word from KISTI that the award process had been delayed, and that KISTI now plans to execute a contract with NCAR in mid July.

Alan had a phone conference with Simon Su, who is responsible for MOM4 visualization at the GFDL in Princeton. Simon would like to work with us as we proceed with our project with Kisti.

Our Siparcs intern, Karamjeet Khalsa, a recent CU graduate, started on May 23. Karamjeet will be working on the problem of converting Ocean data to VAPOR over the next 10 weeks. We held meetings with Tim Scheitlin and Frank Bryan to identify the available software and some POP datasets to work on.

Community service and professional development:

John was a paper reviewer for IEEE Vis Week 2011.

John was a reviewer for a DOE SBIR phase 2 proposal.

Documentation:

Kendall continued work to develop a layout plan for VAPOR documentation based on Drupal "books". A skeleton documentation site was set up, populated with one of the VAPOR documents, and various steps were taken to refine the layout.

Kendall also experimented with various search engines for the documentation before finally settling on Google GSE.

Admin:

Annual reviews were completed for all staff.

John participated in the NSF 5-year review.

Software Research Projects

John worked with former SCIParCS intern Chris McKinlay, CU's Mark Rast and Jesse Lord to complete a manuscript on their 2008 and 2009 work on coherent vortex extraction using wavelets. The manuscript will be submitted to Physics of Fluids shortly.

John prepared slides on data compression for LBNL's Hank Childs for a talk to be given on large data visualization later this summer in France.

Data Analysis & Visualization Lab Projects

Accounting & Statistics Project

- Created scripts to generate a usage report for each project space/fileset from the databases created by robinhood scanning the proj2 and proj3 filesystems.
- Modified ganglia to save the minimum and maximum values instead of just the average for all metrics. Updated the Glade filesystem usage and network usage graphs to display these new values that help show peaks of activity that would otherwise be averaged out over time.
- Updated the script that generates the Glade Usage Report wiki page to read information off the GLADE Allocation Management wiki page.

Security & Administration Projects

- Continued work on analyzing the information available from the LDAP PDB2 database to gather authoritative user and group information for DASG system administration and KROLE management. Determined that the LDAP service is only providing an access to a subset of the PDB2 information, and this does not include information about inactive accounts. Started work on an interface to the REST (HTTP based) JSON (JavaScript Object Notation) interface to PDB2 to acquire the additional information.
- Did more reading on software security issues.
- Made a few updates to the kroleinit program for usability.

CISL Projects

GLADE Project

- Investigation of the glade performance issue on **castle**: Worked on methods to extract file system location for jobs running on **bluefire**. Suggested a simple method to identify file system hooks on **castle** without using "lsof" which was not available on the particular machine. Later, identified the change in the GPFS cluster on **castle** which had been causing the I/O mismatch and poor performance.

Lustre Project

- There was a report from Japan where they ended up with an unmountable MDT and even the "e2fsck" failed. One whamcloud engineer suggested a "-f -E clear-mmp" option in tune2fs that enabled the forced e2fsck and eventual recovery of the file system. Given the absence of details for the Lustre provided e2fsprogs, this will be a critical piece if and when such issue is encountered locally.

Lynx Project

- Lynx Lustre encountered "PTL_NAL_FAILED" errors on one of its OSS servers. Examining the console messages from SMW, tracked the source of the error to a severed connection to OST0000. Later (May 6) SSG replaced the affected HW to recover. During the extended downtime (local /ptmp), they could have re-mapped the block device (OST0000) to the other OSS server if the availability of the file system was critical, but they decided to wait until the HW replacement. This is common error pattern with Lustre under a backend HW failure.
- Si Liu from CSG reported a locking issues with DVS mounted glade. Part of build processes on glade with "autoconf" failed with locking errors. Given the nature of the issue and error message, it was clear that DVS is causing it. Successfully demonstrated the difference by comparing the outputs on DVS servers (native GPFS client) versus those on Lynx login nodes (DVS forwarded client). Also showed Si a work-around to avoid the error using a symbolic link to a local file system.
- Yet another report on file system issue on **lynx** login nodes. Davide from CSG forwarded a case of script failure on multiple targets on DVS mounted glade. Demonstrated the meta-data mismatches on DVS mounted vs GPFS client case and suggested a solution to prevent further issues on **lynx** login nodes.

NWSC Planning

DASG staff continue evaluation of vendor bids submitted in response to the CISL RFP:

- Updated presentations were given to the TET summarizing the DAV & CFDS components of the bids
- Clarification questions were submitted to vendors and responses reviewed
- BAFO guidance to vendors was prepared
- TET spreadsheets have been completed and submitted

System Support

Data Analysis & Visualization Clusters

- After multiple crashes of mirage nodes, identified the single user who has been causing the crashes and the nature of the job. Put a memory limit in /etc/security/limits.conf so that those processes cannot exceed a safer memory ceiling. The same pattern of crashes stopped occurring since.
- Installed the mysql-devel and php53 packages on all systems.
- Compiled and installed the latest versions of netCDF, HDF5, ncview, and NCO to address a bug found in netCDF.
- Installed the PGI compiler suite on all systems.
- Installed the Neural Network Toolbox for MATLAB.
- Recompiled vis5d to link against the newer version of the TCL libraries.
- Helped Gary Strand contain a run away script on two mirage systems.
- Helped Karamjeet setup NFS mount to his Mac for VAPOR use
- Fixed a syntax errors on .profile for 250 users. SSG reported the error, which had been fixed for newer users. Made a script to fix the errors for the affected users and checked the outputs.
- Retrieved accidentally removed files from backups upon a user request.

GLADE Storage Cluster

- Created the "dewit" and CESM filesets, and increased the quotas for the acd and cgd spaces to match their allocations. Removed the dasg016, dasg003, dasg010, dasg012, dasg014, dasg015, dasg018, dasg019, glade004, glade010, and glade012 project spaces.
- Increased the GPFS pagepool setting on **castle** to 512MB, which seems to have cleared up the performance issues and disk thrashing.

Data Transfer Cluster

- Obtained and installed updated host certificates from TACC for the GridFTP servers

Other

- Provided the benchmark outputs and explained the result formats to Nate Rini for his presentation at CUG2011 meeting.
- DASG staff provided input and support for the NSF CISL review.