NASA Nebula in Action: Cloud Computing Case Examples

James Williams, NASA Ames Research Center, james.f.williams@nasa.gov

***Abstract*— In 2009 NASA began work on developing its own cloud computing initiative, Nebula, with the view to provide a cutting-edge, dynamically scalable computing platform capable of meeting Federal/Agency IT standards. Features of cloud computing, such as "instant-on" IT infrastructure, automated provisioning capabilities, and quick scale-up services, are ideally suited for scientific organizations seeking to process large datasets quickly, easily, share them with colleagues and ultimately store them securely at a good price. This paper showcases examples of current uses of Nebula providing insights into how cloud technology can be used to advance space research.**

Keywords: cloud computing, scaling

I. INTRODUCTION

The advent of cloud computing is here; in the next 3 to 5 years more than 80 percent of the world's computing and data storage could occur "in the cloud" [1]. Playing a critical role in determining the development of the Internet's next generation, cloud computing has evolved into a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [2]. Scientists studying space as well as scientists studying conditions on Earth from space are re-thinking their strategies to process, share and store large datasets with the advent of this technological advancement.

Leading IT researchers rank the cloud model as the most significant strategic technology for 2011 and growth is expected to continue [3]. Scientific research organizations are moving with immediacy to better understand the capabilities of cloud computing, plan strategically, and voice their requirements and concerns to the architects of this latest technological trend. The cloud's applications to space research and conducting earth science research from space are apparent as the utility of the cloud quickly becomes ubiquitous in the IT world.

FIGURE 1 illustrates how Nebula addresses the gap between desktops and supercomputers for NASA scientists. Scalability, storage and networking solutions are offered as server-based compute resources, as depicted in the mid-section of the graphic, filling procurement voids with virtual machines.

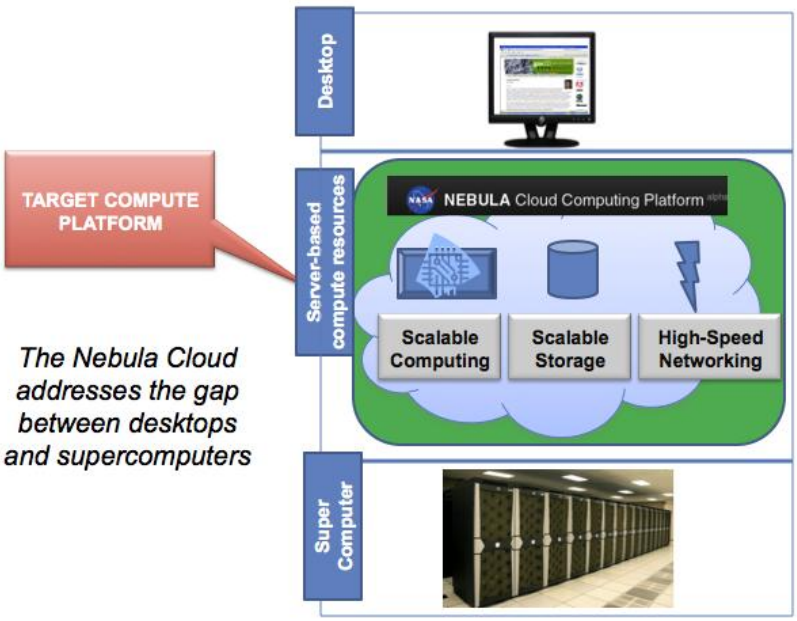

FIGURE 1. *Nebula's Role in NASA IT Architecture*

II. CHARTING A COURSE FOR NEBULA

Nebula started as a web consolidation exercise focused on efforts to discourage Agency web developers from building their own rogue sites. Developers at NASA's Ames Research Center explored solutions to address code hosting, continuous integration and bug tracking problems as assigned, but encountered difficulties after learning that greater elasticity would be essential to run a web application framework as a preferred service. As a solution, NASA began developing its own cloud, Nebula, in 2009, and less than one year later, a Beta version of Nebula Infrastructure as a Service (IaaS) was released in September 2010 to over 200 users in all NASA centers. Nebula's originators believed Agency scientists would benefit from the flexibility, scalability, security and customization offered by the development of a private NASA cloud. Nebula was created to enable scientists to work more efficiently, share their data with greater ease, and help the Agency save IT resources. Consequently, Nebula grew from a small web consolidation effort into one of the Federal Government's most forward-moving technology initiatives.

Initially, NASA considered the use of commercially available cloud services to deliver more robust applications for large-scale projects. However, as the exploratory team discovered, public clouds target the needs of the broader customer base and did not meet NASA's scientific computing requirements regarding science-class performance and capacity requirements. The team also found that certain desirable cloud services or features were commercially unavailable. Collaborating with a team from Goddard Space Flight Center (GSFC), Nebula would need to perform the targeted research and development necessary to provide certain capabilities advantageous to NASA within the Nebula science-class cloud services environment. Concerns about data security, corporate governance, and reliability also made it clear that public cloud options would not be feasible for the Agency. Security measures provided by commercially available clouds did not meet NASA security requirements for scientific data and application.

Today's IaaS offerings include *compute*, a service to provision virtual machines on standard hardware at massive scale, and *object storage*, a service to reliably store billions of objects distributed across standard hardware. The Nebula team continues to work on operations and add new features for IaaS 1.0 while also building the next phase of Nebula: Platform as a Service (PaaS).

III. NEBULA'S EARLY ADOPTERS

The Nebula team initiated several pilot projects prior to Nebula's pre-release to determine if a dynamically scalable science-class computing platform capable of complying with Federal/Agency security standards could be delivered as anticipated. The team worked with systems administrators throughout the Agency to test the platform and see how it could help advance the processing, storage and sharing of data. The role of the early adopters proved critical to Nebula's development.

A. Nebula Tracks the Path of the Tuscaloosa Tornado

NASA's SPoRT Center at the Marshall Space Flight Center is helping scientists identify damage scars left by storms that devastated central and northern Alabama on April 27, 2011. SPoRT provides unique NASA research and remote sensing capabilities to the National Weather Service and other members of the operational weather forecasting and analysis community. Numerous near real-time products from MODIS aboard Terra and Aqua and other satellite imagery from partners contribute NASA-derived data sets to support regional weather forecasting applications. SPoRT recently provided unique NASA observations to support forecasters in assessment of the event in order to compare remotely sensed damage tracks against data acquired during ground surveys in the weeks following the disastrous April 27th Alabama tornadoes [4].

The SPoRT considered Nebula a fortuitous new asset in early May when determining how to proceed with the processing of tornado damage data. SPoRT's existing Linux machines were busy processing data from other tasks at the time of the Tuscaloosa tornadoes and could not be disrupted. Rather then disturb existing workflow, the team decided to pursue new options with Nebula. With its "instant on" IT infrastructure, Nebula afforded SPoRT researchers bandwidth when they needed it most. The benefits of "instant-on" proved helpful, but SPoRT also discovered that Nebula offered many advantages over Linux in this particular case. SPoRT claims that they found Nebula to be easily configurable. Nebula gave them the chance to "play in a sandbox" where configuration testing was easy and fast, and could be used without disrupting other local systems.

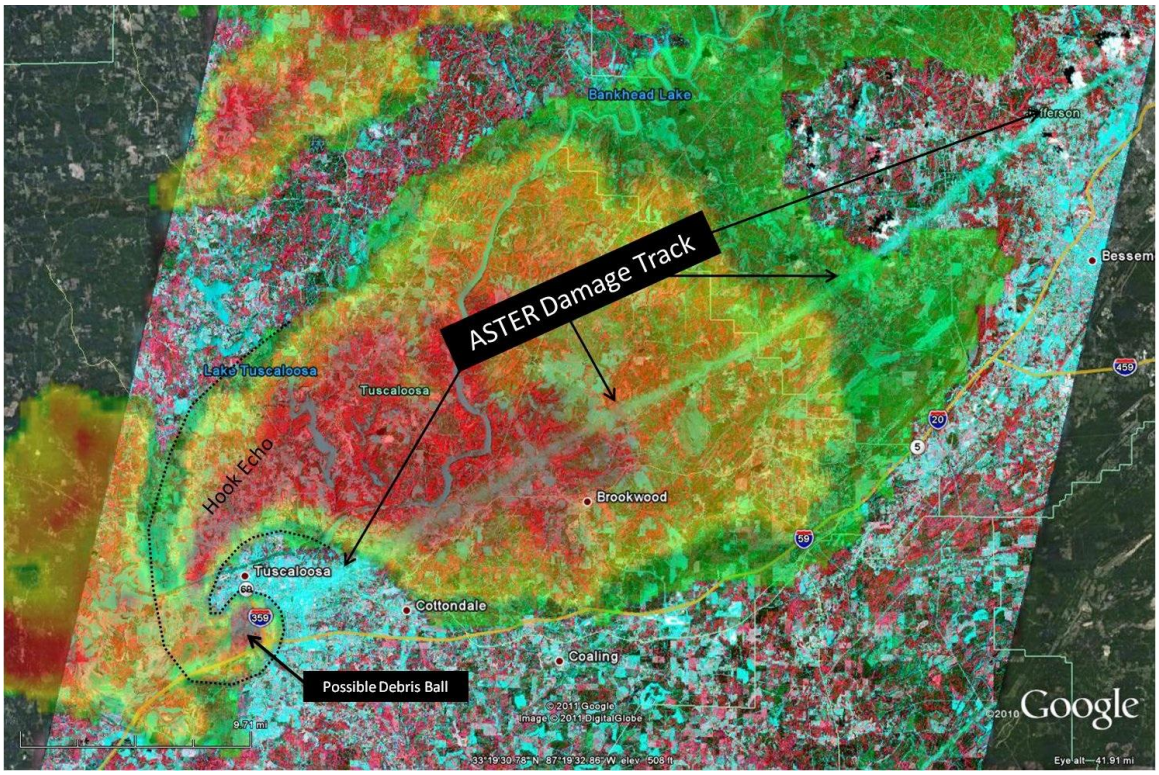

FIGURE 2. *The cyan (blue shade) track represents the false color "damage scar" in the ASTER data, generated by the EF-4 (Enhanced Fujita scale) tornado that ripped through the core of Tuscaloosa before moving northeast and striking the Birmingham metropolitan area*

Nebula contributed to SPoRT's efforts by hosting a tiling application with large, hi-res images. The application was rapidly configured on the large Nebula image; tiles were created, pushed back to the local server, and made accessible to SPoRT partners via Google Earth and other tools. SPoRT acquired imagery from ASTER, a high resolution imaging instrument flying aboard the Terra satellite, from storm-stricken Alabama. By combining spectral channels they obtained false color imagery of damages sustained by the earth's surface providing important clues about how the April 27th weather event impacted vegetation and ecology. The datasets, provided to the National Weather Service through Google Earth and other decision support tools, have assisted staff meteorologists in verifying path length and width of tornadoes in areas where rough terrain or limited road networks make other observations difficult.

One of the represented instances, see FIGURE 2, illustrates how the data were tiled on Nebula, made available in Google Earth, and compared to other observations. The cyan (blue shade) track represents the false color "damage scar" in the ASTER data, generated by the EF-4 (Enhanced Fujita scale) tornado that ripped through the core of Tuscaloosa before moving northeast and striking the Birmingham metropolitan area. Storm damage swaths appear in a cyan color similar to urban areas of Tuscaloosa, where heavily vegetated areas have been disrupted by the storm. In this case, ASTER imagery from May 4th shows the aftermath of the storm, seen where vegetation and other land use patterns have been disrupted. The damage track coincides with storm structures, such as the hook echo and "debris signature" apparent on Doppler radar made available on April 27th. At a wider view (FIGURE 3), other cyan scars are apparent and correlate well with tracks estimated by the National Weather Service. Note that the shapefile used for the NWS tracks only includes a few points, whereas the official track record is more detailed and is correlated to the satellite data.

B. iRODS: Nebula Demonstrates Use for Distributed Data Storage and Management

The NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center completed a pilot project to develop an iRODS-

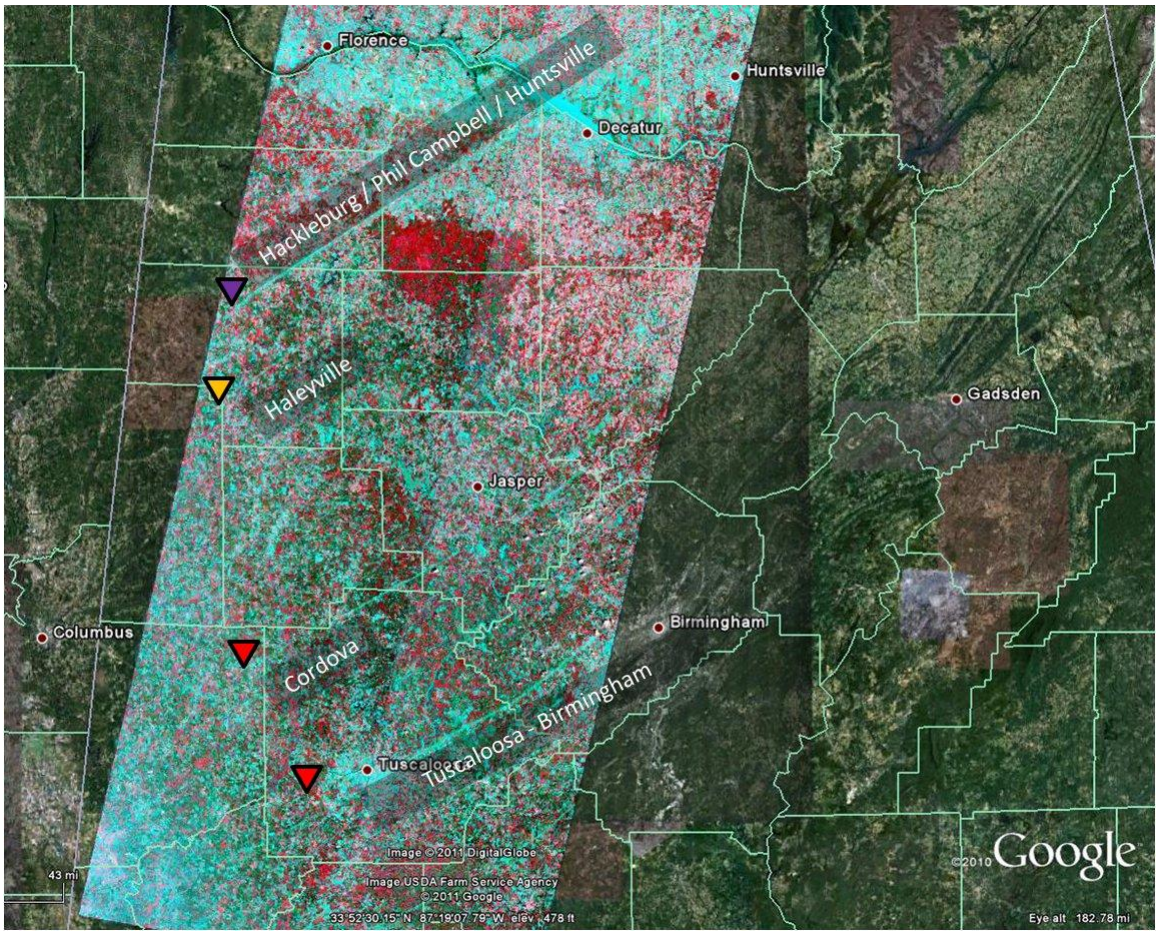

FIGURE 3. *A wider view of the Tuscaloosa tornado, generated on Nebula. Other cyan scars are apparent and correlate well with tracks estimated by the National Weather Service*

based Data Management System (DMS) to deal with the massive amounts of observational and model data used in NASA's climate and weather studies. The Integrated Rule-Oriented Data Management System, iRODS, is an open source, data grid software solution that allows for the management, sharing, searching, and distribution of large and diverse scientific data sets. Through iRODS, Nebula demonstrated the ability to federate; a feature in which separate iRODS instances can be integrated, thereby giving users and applications a consolidated view of separately maintained data [5].

The NCCS provides compute engines, analytics, data sharing, long-term storage, networking, and other high-end computing services designed to meet the specialized needs of the Earth science community. The DMS team used Nebula to host the DMS prototype with a goal of managing and publishing climate simulation data using iRODS within a distributed set of Nebula instances. A collection of Modern Era Retrospective-Analysis for Research and Applications (MERRA) data was placed under iRODS control. The MERRA data was stored on the file system and registered with iRODS. The registration process stored metadata about each of the MERRA files in the iRODS database. The entire catalog of monthly MERRA products resulted in the ingestion of 360 files that occupy 47 GB. The data was shared between two instances. By eliminating the need to explicitly switch an iRODS client between distinct grids, federation allowed perusal or download of data from multiple iRODS repositories through a single interface. Upon completion, users could examine, search for, and download simulation data from either Nebula instance through a single iRODS web interface.

The NCCS DMS team used Nebula to host DMS prototype with the goal of managing and publishing climate simulation data using iRODS with a distributed set of Nebula instances.

FIGURE 4 shows a false-color infrared image captured by the Atmospheric Infrared Sounder (AIRS) instrument on NASA Aqua Satellite as it flew over Iceland's erupting Eyjafjallajokull volcano on April 15, 2011. The image shows the ash cloud (in blue) enveloping Iceland and moving eastward over the Shetland Islands onward to Europe.

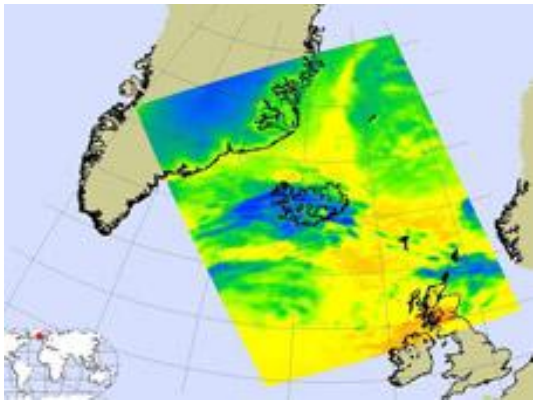

FIGURE 4. *Ash cloud over Iceland volcano.*

C. Nebula Processes Data to Study Gulf of Mexico's Health

NASA's John C. Stennis Space Center in Mississippi used Nebula to process data for an environmental project aimed at boosting the health of the ecosystem in the Gulf of Mexico. The Center's Applied Science and Technology Project Office (ASTPO) uses the results of NASA Earth Science research to address issues identified by a partnership of five states in the Gulf region called the Gulf of Mexico Alliance [6]. The need to study both the ecological and economic health of the region has been heightened since the disastrous Deepwater Horizon oil spill in April 2010.

Analysis conducted by the Alliance attempts to link biological conditions in the Gulf to nutrient concentrations with the goal of using the information to improve the region's health. Satellite observations were used to estimate water quality parameters as part of that research. To obtain the necessary information, thousands of datasets needed to be processed from a NASA data center located at Goddard. To speed up the processing, the ASTPO used Nebula to process the data sets, as well as used Web services to access and download data sets from disparate, regionally dispersed systems.

The Stennis group processed data from two MODIS (Moderate Resolution Imaging Spectroradiometer) instruments, launched by NASA into Earth orbit on the Terra and Aqua satellites, to retrieve ocean color measurements in the Gulf and its estuaries. They used MATLAB, a high-level language and interactive environment for computationally intensive tasks, to download and subset the data, then processed the datasets with SeaDAS software while the next dataset is being downloaded. This new method provides a significant speed increase over the traditional approach. Nebula enabled the remote processing by delivering a flexible computational environment that can be customized for specific processing needs. MODIS datasets as well as ozone and meteorological data were downloaded by web services from servers at GSFC. Several thousand such datasets were processed to produce the time series covering the years from 2000 to present [7].

IV. CONCLUSION

In 2009, the Obama Administration announced the Federal Government's Cloud Computing Initiative to greatly reduce waste, increase data center efficiency and utilization rates, lower operating costs, and support green technology [8]. In the same month of Obama's mandate, the Nebula cloud container arrived at Ames Research Center and NASA entered a new era of information technology development. U.S. Chief Information Officer Vivek Kundra visited Ames the same month to share the Administration's vision on cloud computing for the federal government and cited Nebula as an example of the Government's ability to "leverage the most innovative technologies" [9].

From an outgrowth of a simple research exercise, Nebula developed into one of the space industry's first cloud computing environments. Nebula delivered to NASA more than what the Administration mandated; Nebula offers not only a cost-cutting, more efficient compute model, but a tool that enables scientists to process, share and store valuable data more effectively. During its short tenure, Nebula helped Agency scientists and administrators realize cloud computing potential for processing data more quickly, storing it more efficiently, and sharing it with colleagues and the public more effectively. Cloud computing is becoming a critical tool for NASA as the need to process and store large datasets grows within the Agency, and as issues of scalability are further explored. Nebula started out as an experiment, but is now capable to enabling missions to get the results they need quicker, more efficiently and at a better cost to the Agency.

REFERENCES

[1] M. Nelson, "Building an Open Cloud," Science. vol. 34 no. 5935 pp. 1656–1657, June 2009. http://www.sciencemag.org/content/324/5935/1656.full

[2] J. Foley, "Private Clouds Take Shape." Informationweek.com. August 9, 2008.

[3] N. Hoover, "NASA Readies Nebula Cloud Computing Platform for Release." Information Week, October 20, 2010. http://www.informationweek.com/news/government/cloud-saas/227900365

[4] F. Pietrucha, "Answers in the Wake of the Tuscaloosa Tornadoes," Nebula blog. http://nebula.nasa.gov/blog/2011/06/20/nebula-and-sport-answers-wake-tuscaloosa-tornado/

[5] F. Pietrucha, " iRODS: Nebula Uses Distributed Data Storage and Management." http://nebula.nasa.gov/blog/2011/03/21/irods-nebula-used-distributed-data-storage-and-management/.

[6] E. Montalbano, NASA to Leverage Nebula for Gulf Ecosystem Project, Informationweek Government. http://www.informationweek.com/news/government/cloud-saas/229401820?queryText=April+19,+2011,+Nebula

[7] F. Pietrucha, "Stennis Uses Nebula to Advance Study of Nutrients in Gulf Waters," Nebula blog. http://nebula.nasa.gov/blog/2011/apr/stennis-uses-nebula-to-advance-study-of-nutrients-/

[8] V. Kundra, State of Public Sector Cloud Computing, May 20, 2010. http://www.cio.gov/pages.cfm/page/State-of-Public-Sector-Cloud-Computing

[9] NASA Announcement. "NASA Ames Hosts White House CIO," http://www.nasa.gov/centers/ames/news/features/2009/cloud_computing.html