

Giovanni in the Cloud

Porting Giovanni MAPSS Workflow to NASA's Nebula Cloud

July 2011

Janie Campino
Embry-Riddle Aeronautical University
Mentor: Christopher Lynnes

Project Objective

Leverage cloud benefits to meet Giovanni's needs:

- Increase Scalability
 - Demand spikes
 - Expense of scientific data and workflows
- Minimize risk of hardware failure
 - Lessen the risk of server failures through using dynamic scalability in response to stresses on the system.

Achieve higher performance

- The servers and processors available through Nebula provide for performance enhancements as compared to current hardware

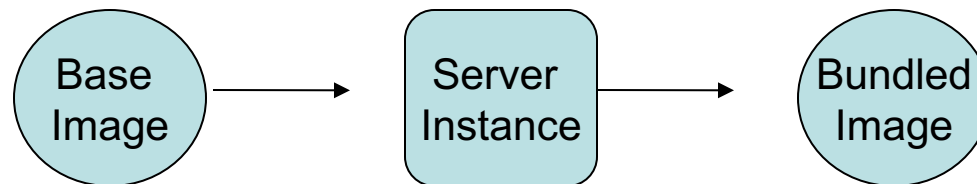
Create a portable Giovanni that is environmentally flexible

- The loosely coupled, component-based architecture lends itself to flexibility, but through porting the system to Nebula, any residual networking or architectural dependencies can be resolved.

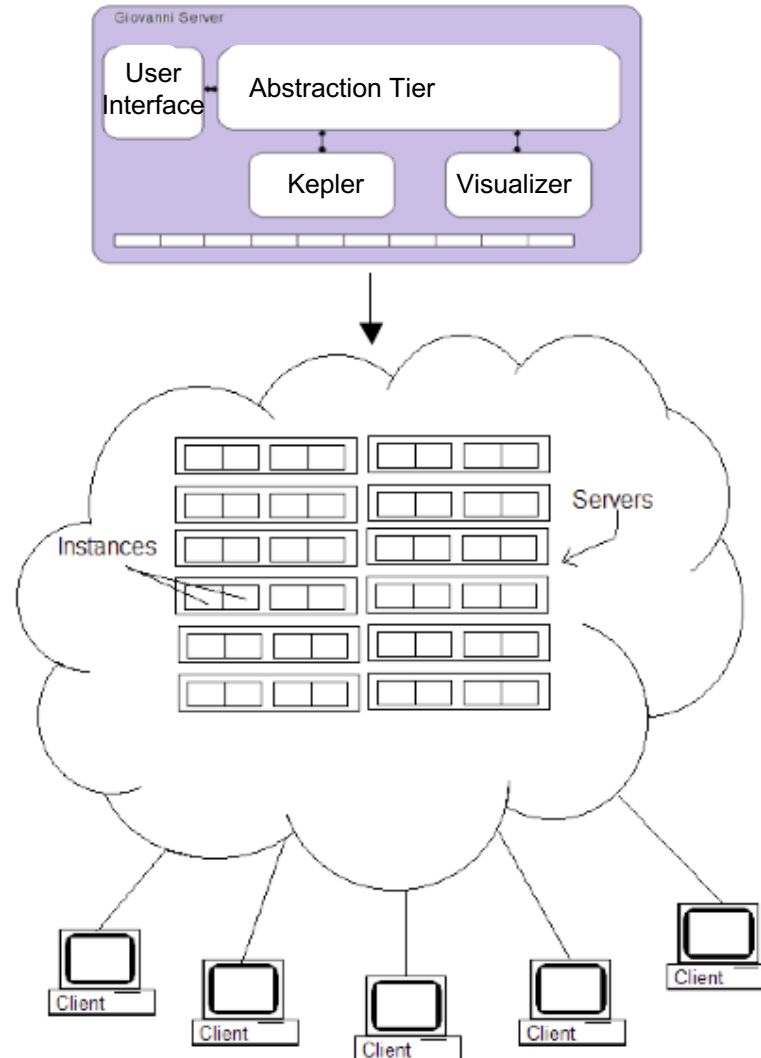
Cloud Computing

What is it?

- *“Cloud computing is 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.” -NIST*
 - Extension of traditional client-server model
 - Shared hardware resources
 - Virtual servers
 - Dynamic provisioning
- How is cloud used to deploy Giovanni



At a Glance



About Nebula

- NASA's private cloud
- Developed in 2008
- Still developing
- Support, Forums and Knowledge Base available
- Uses Eucalyptus
 - Dashboard: *Nebula's online cloud management resource*
 - Creating and attaching volumes
 - Launching images
 - Terminating instances, keys, images
 - Managing keys
 - Euca2ools: *Command line tools provided by Eucalyptus*
 - All the functionality of Dashboard
 - Reboot instances
 - Bundle instances

Challenges

- **Control of the servers**
 - Nebula shut down mid-project
 - All images and instances had to be removed or would be lost
 - Unannounced changes to Nebula caused loss of active instances
- **Image Defect**
 - Base image had networking defect that carried through to instance (and all derived images) - couldn't get a successful image bundled before shut down
- **Configuring Giovanni**
 - Networking
 - Use of wwwuser as compared to apache user
- **Nebula is still developing**
 - Still some bugs to be worked out

Results

An instance was successfully launched and configured. The Giovanni server was ported to the Nebula cloud for the MAPSS workflow. The instance reached out to the data stores to retrieve the data. Data and visualizations were downloaded to and through the browser.

Additionally, a step-by-step guide to generating an instance of the MAPSS workflow on Nebula was drafted.

Screenshots

MAPSS: Multi-sensor Aerosol Products Sampling System HELP

This user interface is used to obtain selected parameter statistics from the [MAPSS](#) database for a chosen location and time period. Time Series Plot is the available service. Plot output is rendered as a graph and is also available in ASCII format.

Data Selection Results

To see time series plots of MAPSS data, choose from the criteria below and click Get Plot(s)

Select Station
 GSFC Select

Select Variables
 To select variables, make a *single* selection from each list below (beginning with the left-most list), and then click 'Add'. Selected variables will be added to the summary. Repeat for additional variables.
 Basic Advanced

Product	Parameter	Layer	Variable
AERONET aerosols L2, ver. 2	AOD	Water vapor	Measurement
AERONET deconvolution L2, ver. 41	Angstrom exponent		Central value
AERONET inversions L1.5, ver. 2	Water vapor		Mean
AERONET inversions L2, ver. 2			Median
CALIPSO column and layer aerosols L2, ver. 301			Standard deviation

Selected Variables
 AERONET aerosols L2, ver. 2 : Water vapor : Water vapor : Central value Delete

Select Date Range
 Start Date: mm/dd/yyyy Pick Start Date
 End Date: mm/dd/yyyy March 1997

Su	Mo	Tu	We	Th	Fr	Sa
23	24	25	26	27	28	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5

To see time series plots of MAPSS data, choose from the criteria above and click Get Plot(s)

ACKNOWLEDGMENT: Support through the [ROSES 2006 ACCESS](#) system for integrated validation, intercomparison, and analysis of aerosol products from multiple satellites has been provided by NASA HQ (PM: Stephen Berrick). [AERONET](#) data are contributed by the International AERONET Federation (PI: Brent Holben).

MAPSS: Multi-sensor Aerosol Products Sampling System

This user interface is used to obtain selected parameter statistics from the [MAPSS](#) database for a chosen location and time period. Time Series Plot is the available service. Plot output is rendered as a graph and is also available in ASCII format.

Data Selection Results

Current Result: Result 3 - MAPSS Time Series Load

Plots Downloads Lineage

Result 3 - MAPSS Time Series: View Criteria Problem? Send a report...

MAPSS Time Series

— Median AOD: Median of AOD at 440nm from AERONET_AOD_L2_2 at GSFC

ACKNOWLEDGMENT: Support for the development of this data access system for integrated validation, intercomparison, and analysis of aerosol products from multiple satellites has been provided by NASA HQ through the [ROSES 2006 ACCESS](#) Program (PI: Charles Ichoku). The [AERONET](#) data are contributed by the International AERONET Federation (PI: Brent Holben).

Possibilities for Giovanni

- Dynamic Resource Management
 - Load Balancing
 - Programmatic Instantiation
- Migrate Data Stores to Nebula
 - Optimize data structures
- Virtualize cached data

**A desktop VM Giovanni is currently under configuration as well.*

Summary

- Porting was straightforward because of:
 - Nebula tools
 - Component-based architecture
- Complications
 - Nebula access
 - Bundling Complications
- Objective of porting met!
- Cloud benefits not fully realized in this iteration
 - Scalability
 - Programmatic instantiation
- Step-by-step guide and desktop instance to facilitate future endeavors