OpenStack @ NASA

Past, Present, and Future Plans For A More Flexible Datacenter Using Open Infrastructure

NASA Center For Climate Simulation
Goddard Space Flight Center

Jonathan Mills
Unix Systems Administrator
Who are we...specifically?

(Because NASA has many different sites)
Goddard Space Flight Center

Established on May 1, 1959 as NASA's first space flight center

No one organization has put more satellites into space than Goddard

More earth scientists at Goddard than any other single organization on earth

Over 10,000 scientists, engineers, and other NASA employees work at Goddard
This is DISCOVER

About 40% the size of Pleiades cluster at NASA Ames
About 50% the size of the CERN OpenStack cluster
Our Computer Room (half of it)

28 or 40 processors in each node (Haswell or Skylake)
4000 nodes, three fabrics (IB and OPA), one Ethernet mgmt net
Over 100,000 processors
SLURM scheduler
GPFS filesystem
What Does Our Supercomputer Do?
A brief history of OpenStack at NASA...
You may have seen this before…

PHOTO: CHRIS KEMP
Our Container at NASA Goddard
What happened to NASA’s involvement?

“Recently, on May 15 [2012], NASA announced a new cloud computing strategy for the Agency at the Uptime Institute’s symposium in Santa Clara, CA. Among its facets is a reduction to our OpenStack development efforts in favor of becoming a ‘smart consumer’ of commercial cloud services. In understanding this shift, it is important to consider that the majority of NASA’s code contributions to OpenStack were during its early stages while its developer and industry communities were still forming. Since those early days, the OpenStack community has grown considerably, and we have borne witness to accelerated development in key areas directly bearing on NASA’s originally identified needs. In fact, the vast majority of code contributions over the past year of intense OpenStack development have come from community members other than NASA.”

Ray O’Brien (Former Nebula Project Manager, NASA Ames Research Center CIO, Acting)
May 2012
Source: https://open.nasa.gov/blog/nebula-nasa-and-openstack/
Why OpenStack? Why private cloud at all?
Rationale for Private Cloud

- Data locality: we already have the data on-premise...
- Scientists expect POSIX filesystems – that’s just how it is
- Security (with some caveats – more later)
- Cost
- Better platform for “lifting and shifting” traditional science codes
- OpenStack APIs provide a unifying vision for how to manage datacenter infrastructure; avoid building “unicorns”
- Extends the useful life of HPC gear, thus better ROI
Timeline of OpenStack @ NCCS

• 2015
  » Evaluating many different deployment mechanisms for OpenStack
  » Requirements gathering
  » Building Dev/Test clusters (ongoing)

• 2016
  » Researching different architectures, network designs
  » Writing custom puppet code (long story)
  » Writing RFPs for hardware procurement

• 2017
  » Building production OpenStack cluster @ NCCS (for ADAPT environment)
  » Building prototype of Goddard Private Cloud

• 2018
  » Upgraded ADAPT OpenStack from Mitaka to Pike
  » Built production Goddard Private Cloud based on Queens
ADAPT OpenStack
A complementary resource to HPC @ NCCS
## Why did the NCCS create the ADAPT environment?

### High Performance Computing

*Typically small input, large working space, and large output...*

- Smaller amounts of observation data, models are run to generate forecasts
- Tightly coupled processing requiring synchronization within the simulation
- Simulation codes: typically 100,000’s of lines
- Production runs of applications push the utilization of HPC systems to be very high
- Fortran, Message Passing Interface (MPI), large shared parallel file systems
- Rigid environment – users adhere to the HPC systems

### Data Analysis

*Typically large input, moderate working space, small output...*

- Use large amounts of distributed observation and model data to generate science
- Loosely coupled processes requiring little to no synchronization
- Analysis codes: typically 100’s of line
- Require more agile development with many small runs; utilization can be low on average
- Python, IDL, Matlab, custom codes
- Agile environment – users run in their own environments (steep learning curve for rigid HPC)
Some Examples Of Projects Using ADAPT

- Arctic Boreal Vulnerability Experiment (ABoVE)
- CALET (CALorimetric Electron Telescope)
- High Mountain Asia Terrain (HiMAT)
- Laser Communications Relay Demonstration (LCRD) Project - FPGA simulations
- Asteroid Hunters – Near Earth Objects
- Biomass in South Sahara
- NCCS Data Services
ABoVE Water Maps: 30 meter spatial resolution surface water 1999-2011

Processing work flow for the generation of the ABoVE water maps from Landsat scenes to ABoVE tiles.

100,000 Landsat Scenes 20 TB of Data

AWM for 2001 and 2011 for Hay Lake and Beaver Hill Lake in Canada. Hay Lake has clearly expanded over this time frame while Beaver Hill Lake has diminished.

ADAPT OpenStack – Overview

- 144 Dell C6320 nodes, dual Xeon E5-2650v4, 2.2 GHz, 3456 cores
- 256 GB RAM
- 2 x 800 GB Intel SSD (3 DPWD)
- Ethernet instead of RDMA (2 x 10GbE to each node, “bond0”)
- Uses mainly VLAN-based provider networks with Neutron
- Emphasis on access to data lake, usually via GPFS or NFS
- Has been transitioning from an older managed VM environment into OpenStack for a few years
ADAPT OpenStack Network Topology

```
400GbE Fiber QSFP-LR4
100GbE Copper QSFP28

Firewall

32xQSFP28 100 GbE Spine

10/25/100 GbE Leaf
10/25/100 GbE Leaf
10/25/100 GbE Leaf
10/25/100 GbE Leaf

Compute Nodes

Cabinet 1
Cabinet 2
Cabinet 3
Cabinet 4
...

100GbE Fiber QSFP28
250GbE Copper SFP28
```
Future Plans for ADAPT OpenStack

- At the NCCS, most non-HPC systems (i.e. systems outside Discover) will migrate into OpenStack over the next 1-2 years
- Accomplish several OpenStack upgrades (Queens/Rocky/Stein)
- Shore up the backend storage with enterprise-grade, flash-based NAS devices
- Expand compute capacity, integrate GPU and NVMe
- Get more of our sysadmins comfortable using OpenStack APIs
- Evaluate OpenStack Ironic as potential replacement for IBM xCAT
Goddard Private Cloud (GPC)
Serving a wider community...
GPC

- Grew out of the success of ADAPT OpenStack
- Shares architecture with ADAPT OpenStack
- Emphasis on hosted applications, misc smaller science/engineering work
- User self-provisioning
- Always uses tenant isolation, floating IPs
- Supports IPv6 for some web applications
- Has a bill-back model using Cloudkitty
- Uses Nova CellsV2 for scalability
GPC and SEN

- 16 Port 100G Ethernet Switch
- 16 Port 100G Ethernet Switch
- 16 Port 100G Ethernet Switch
- 16 Port 100G Ethernet Switch

From Same Building:
2 x 10G connections to each hypervisor
8 x 100G connections to storage

Control Plane
Bare Metal Servers / Hypervisors & Storage

2 x 100 connections to each hypervisor (actual links not shown)

Nova Cell 1
Hypervisors

Nova Cell 2
Hypervisors

Legend
- 10 Gigabit
- 40 Gigabit
- 100 Gigabit

External Zone
- Internet
- NAS/ARC
- Gateway

Mixed Zone
- Border
- 40G Switch
- Firewalls

Internal Zone
- Building 26
- Building 32
- Building 33

Building 26 GPC Prototype Cloud

Building 32 GPC Production Cloud
GPC-Hosted Science

Keyshot

- CAD rendering software package used by engineering teams
- Engineers point their Keyshot client to the network rendering farm in GPC
- Scales linearly; licensed for up to 512 vCPU

Exoplanet Modeling and Analysis Center (EMAC)

- EMAC is a key project of the GSFC Sellers Exoplanet Environments Collaboration (SEEC)
- Serves as a repository and integration platform for modeling and analysis resources focused on the study of exoplanet characteristics and environments
- Heavy use of docker containers in VMs
M-Dwarf Planets

- Python simulation to determine the exposure time and other parameters required to observe exoplanets with the James Webb Space Telescope (JWST) or other telescopes.
- Using data from the PSG project, also on GPC

HFSS for Superconductivity Analysis

- Ansys High Frequency Structure Simulator (HFSS)
- Simulations used by engineers to design science instruments
Future Plans for Goddard Private Cloud

- Expand compute capacity (80 new nodes from Discover SCU9)
- Leverage OpenStack Manila
- Expand into multiple buildings for DR
- More web hosting
- More customers, more science, more funding…
Some Observations...
Paradox of Tenant Isolation

We want isolated tenants, but then we immediately want easy access into or out of them, because:

- Security teams want to scan the internal tenant network
- Users want easy access to external (often POSIX) data
- Users immediately complain about the security groups / firewall
- Performance is lost when we must move from Layer 2 networking to routed Layer 3 to bounce out of the tenant network
Paradox of User Self-Provisioning

We want users to be able to provision their own resources, and users want to feel more empowered, but:

- Users are terrible sysadmins, who will make poor choices
- Users will preference functionality over security or good design
- Users will never proactively conserve resources without incentive
- Users typically copy a traditional code or app and run it in the cloud exactly how it ran before, without any refactoring to be cloud-savvy
Special Thanks

Dan Duffy
Ed Masuoka
Bruce Pfaff
Ellen Salmon

Hoot Thompson
Michael Moore
Reid Ormseth
Steve Baxter