Early Experiences with Cloud Bursting Managed by Slurm at NeSI

Michael Uddstrom, Dan Sun, Mike Ladd, Georgina Rae,
Gene Soudlenkov, Peter Maxwell, Yuriy Halytskyy, Jordi Blasco

New Zealand eScience Infrastructure

Outline

1. Motivation
2. HPC Cloud Bursting
3. Architecture
4. Early Experiences
5. Future Landscape

Motivation

Motivation

Understand what cloud can offer to NeSI

- Cover peak demands.
- Cover potential growing computational need.
- Increase the service availability and redundancy.
- Identify potential new services.

HPC Cloud Bursting

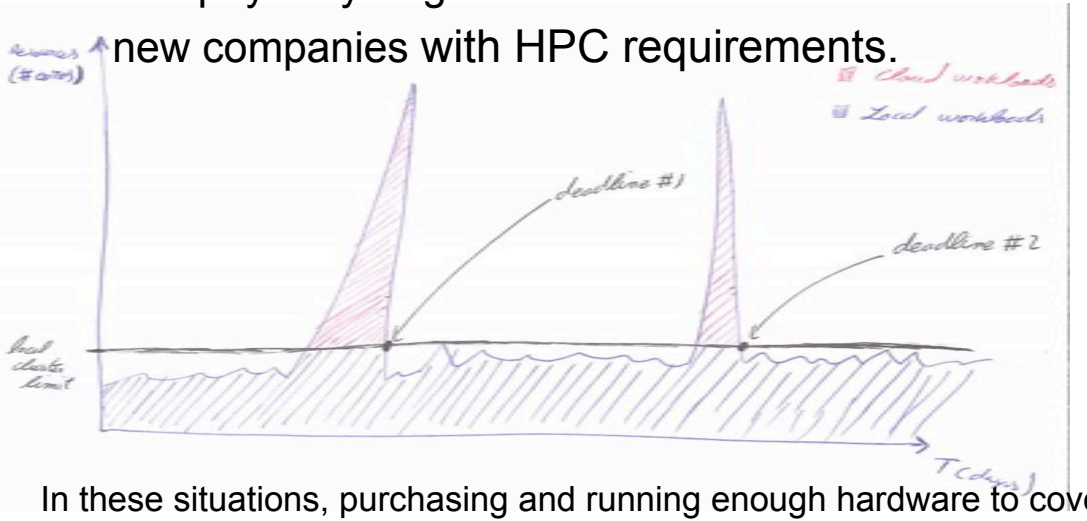
What is Cloud Bursting?

Cloud bursting allows us to offload jobs into the cloud when local resources are running out of capacity or in order to meet deadlines.

Cloud bursting can be seen as *on-demand* extension of a local HPC cluster.

When is Cloud Bursting Useful?

- Scenarios with very strong peak demand.
- Environments frequently under pressure to deliver prompt results.
- The pay-as-you-go bussiness model offers cost-effective option for new companies with HPC requirements.



In these situations, purchasing and running enough hardware to cover demand peaks could be even more expensive than using external compute services in the cloud.

Cloud Services

Cloud vendors claim to be able to offer:

- Instant availability
- Almost unlimited capacity
- Almost unlimited storage
- Virtualised environment
- Custom instances
- High service availability (SLAs)

Workload Suitability

Not all HPC workloads fit well in all cloud solutions:

- Sensitive data or export restrictions
- Licensing restrictions
- Data staging overhead for large datasets
- Codes may depend on low latency network
- Codes may depend on non-standard compute resources (accelerators)
- Workloads may depend on fast cluster file system (e.g. BeeGFS, GPFS or Lustre)

Cloud Bursting can offer much more

It helps to improve the efficiency of the local cluster allocation by reducing:

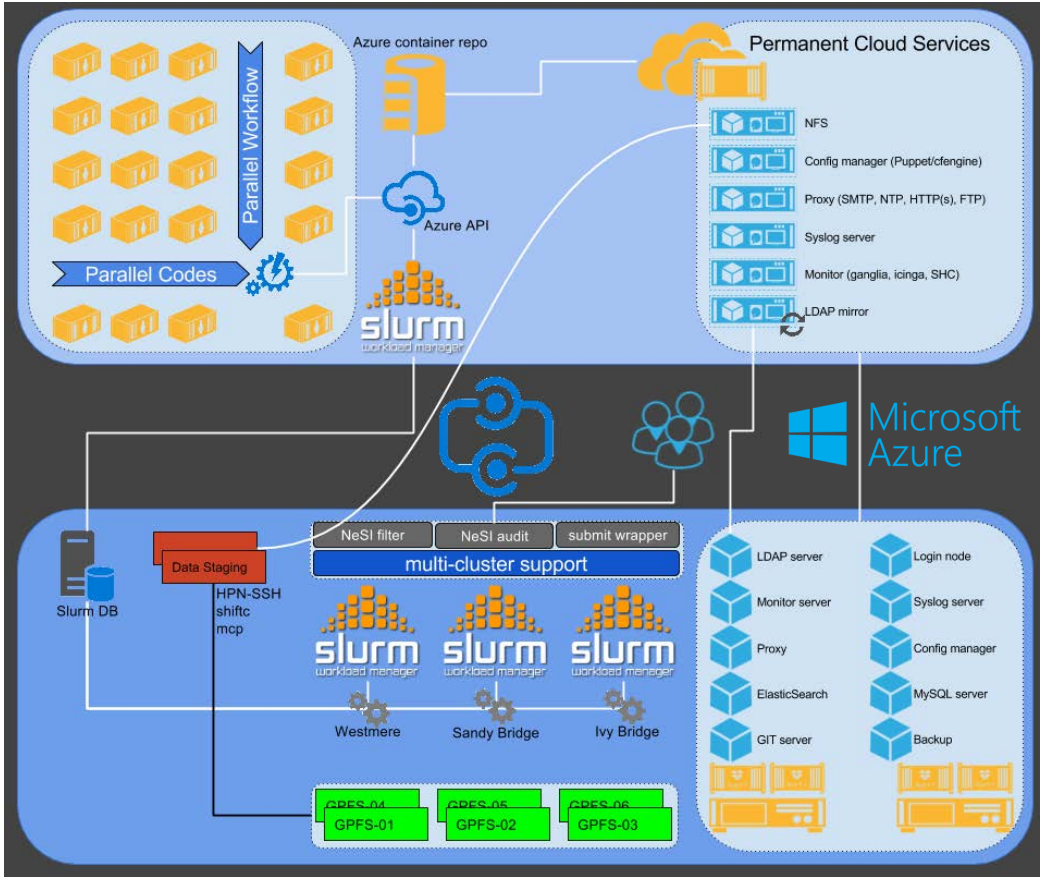
- Target job's turnaround time.
- System-wide job turnaround time.
- System fragmentation.
- Queuing time.
- Licensing cost.

HPC Cloud Bursting

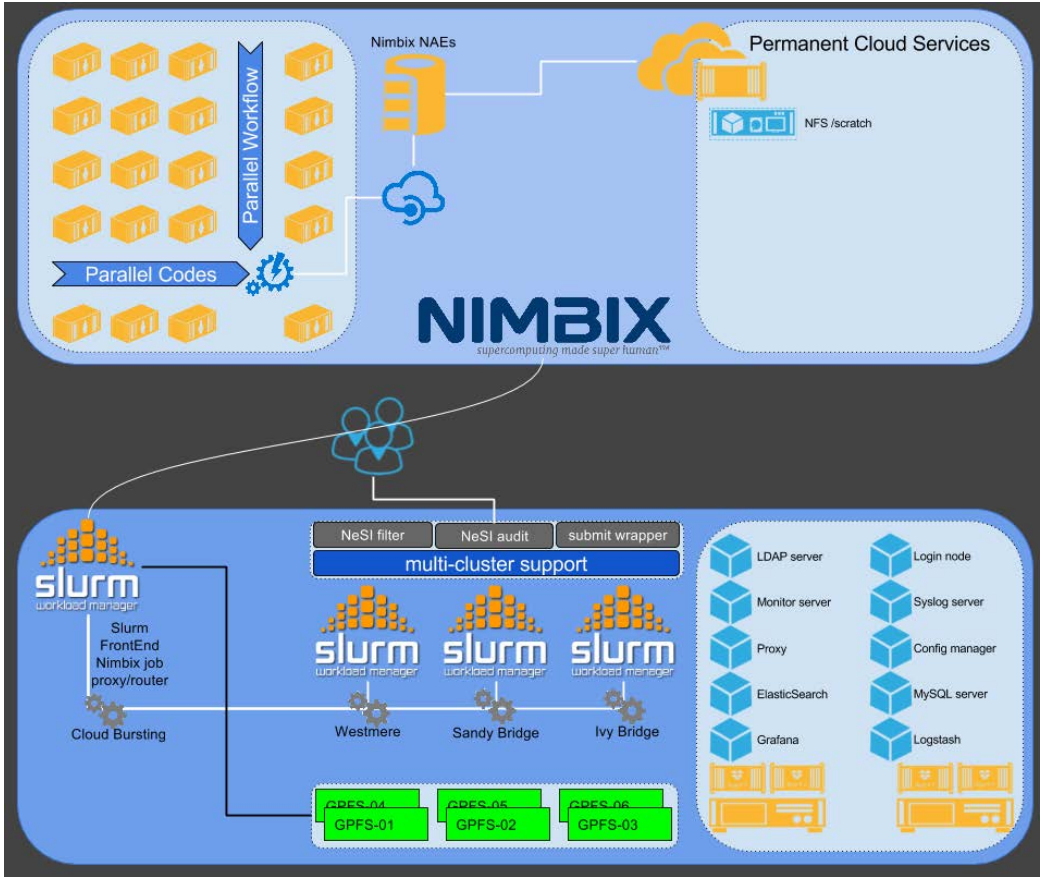
Microsoft Azure

Architecture

Architecture



Architecture



Slurm Workload Manager

Key Features

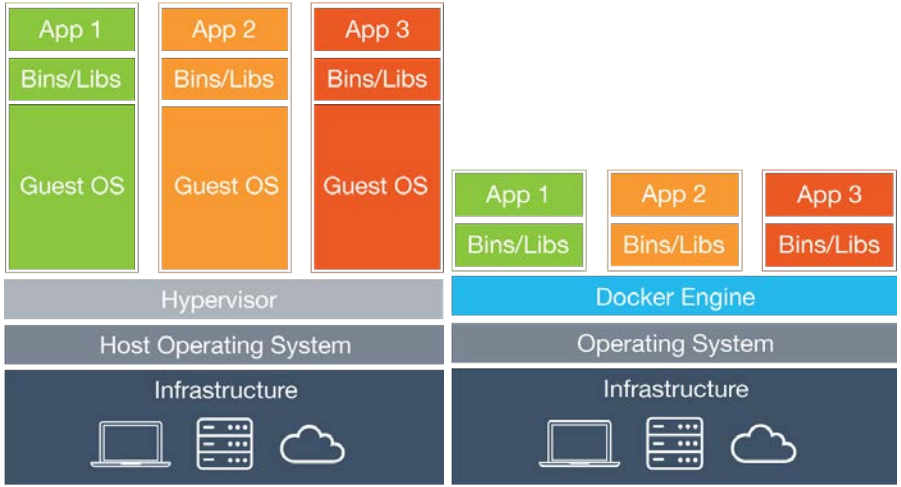
- Multi-clustering
- Cloud-bursting (PaaS)
- Front-end node (SaaS)

Docker Containers

Allows us to wrap up a piece of software in a complete filesystem that contains everything it needs to run (code, runtime, system tools, system libraries).

Benefits

- Agility
- Control
- Portability
- Performance



Virtual Machines (left) require much more resources and higher operational cost than Linux containers (right). Linux containers only includes the application and its dependencies.

Submit Wrapper

The wrapper parses a job description file specified by the user and makes a decision on the target system.

Requirements

- Seamless from user point of view
- Integration with data staging
- Business rules

Data Staging Protocol

Each job indicating data to be staged will trigger a job scheduled on data transfer nodes connected to GPFS servers.

We are exploring two tools for data transfer:

- **GridFTP** (Globus Toolkit)
- **Shift** (NASA Advanced Supercomputing Division)

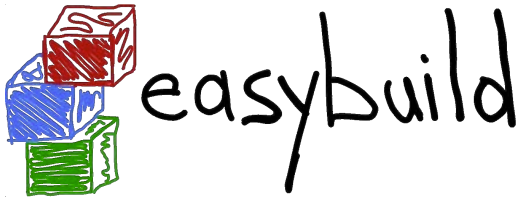
GridFTP
http://toolkit.globus.org/toolkit/data/gridftp/
Self-Healing Independent File Transfer
http://sourceforge.net/projects/shiftc/

Application building with Easybuild

EasyBuild is a software building and installation framework that allows to manage (scientific) software on HPC systems.

Features:

- build & install scientific software fully autonomously
- easily configurable
- thorough logging and archiving
- automatic dependency resolution
- building software in parallel
- fully tested before each release
- growing community



Developed by the HPC team at Ghent University together with the members of the EasyBuild community, and is made available under the GNU GPLv2. http://hpcugent.github.io/easybuild/

Early Experiences

PaaS Early Experiences

- A lot of documentation.
- Several Open-Source examples available in GitHub.
- For people with previous experiences in other cloud solutions it’s pretty simple.

Microsoft Azure Early Experiences

Full deployment and orchestration of master node

~ 25 minutes

Compute nodes allocation (Time to Production)

24 nodes (192 cores) : ~ 16 minutes

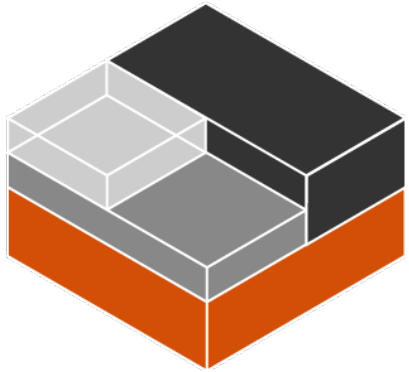
Future Landscape

LXD

Linux Containers as a stand alone service can be very insecure compared to virtual machines. LXD opens new opportunities for privacy in standard HPC without reducing performance.

Features

- Secure by design
- Scalable
- Intuitive
- Image based
- Live migration



The LXD project was founded and is currently led by Canonical Ltd and Ubuntu with contributions from a range of other contributors. https://linuxcontainers.org/lxd/

Federated Cluster Support in Slurm

Expected features in version 16.05 and beyond:

- Job migration (pending jobs automagically migrated to less busy clusters).
- Fault Tolerance (participating clusters will take over work of a failed cluster).
- Cross-cluster job dependencies.
- Unified views.

Summary


New services
Define future landscape
Challenges
Improvements

Meet the team!

Michael

Mike

Dan

Georgina

Jordi

Gene

Yuriy

Peter