CREATING THROWAWAY SUPERCOMPUTERS IN THE CLOUD WITH

MAGIC CASTLE

CfgMgmtCamp | 20240205 | Ghent

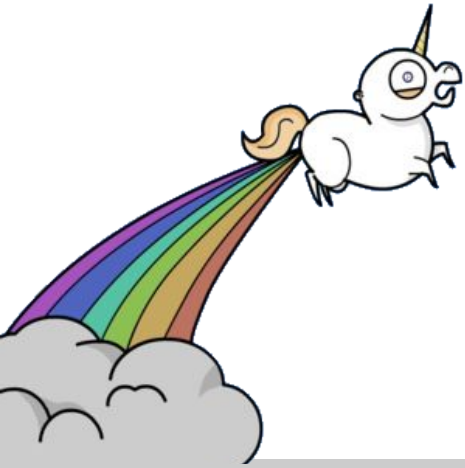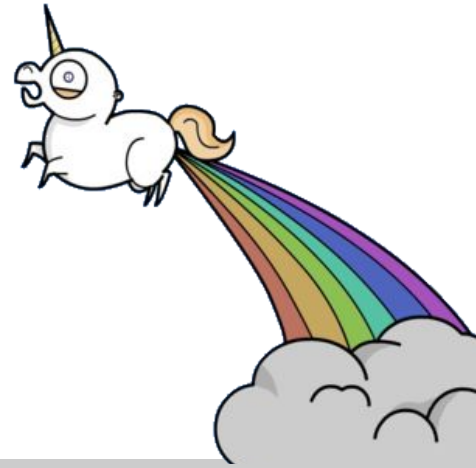*Kenneth Hoste | HPC-UGent | kenneth.hoste@ugent.be*

Creating Throwaway Supercomputers in the Cloud with

MAGIC CASTLE

CfgMgmtCamp | 20240205 | Ghent

Kenneth Hoste | HPC-UGent | kenneth.hoste@ugent.be

WHOAMI

Kenneth Hoste (@boegel)

- HPC system administrator at Ghent University (Belgium)

  since 2010 (https://ugent.be/hpc)

- Lead developer of EasyBuild (https://easybuild.io)

- Active contributor to EESSI (https://www.eessi.io)

- **Happy user of Magic Castle**

- Involved in way too many FOSS projects

- FOSDEM visitor & active participant since 2013

- Only my 2nd time at CfgMgmtCamp...

Supercomputers?

- Supercomputers are ...

Supercomputers?

- Supercomputers are

  big, fast, expensive, complex

10¹⁰ faster!

Cray-1 (1975)

Frontier @ Oak Ridge Nat. Lab
(#1 in Top500 of June'2023)

Supercomputers?

- Supercomputers are big, fast, expensive, complex

- **"Just a bunch of servers with fast interconnect"**

- Used for scientific simulations, big data analysis, AI/ML, ...

- **Multi-tenant**: 1000s of scientists, different profiles

- **Performance is key**: faster software means more science gets done

- Key elements typically are **Linux as OS**, a **shared filesystem**,

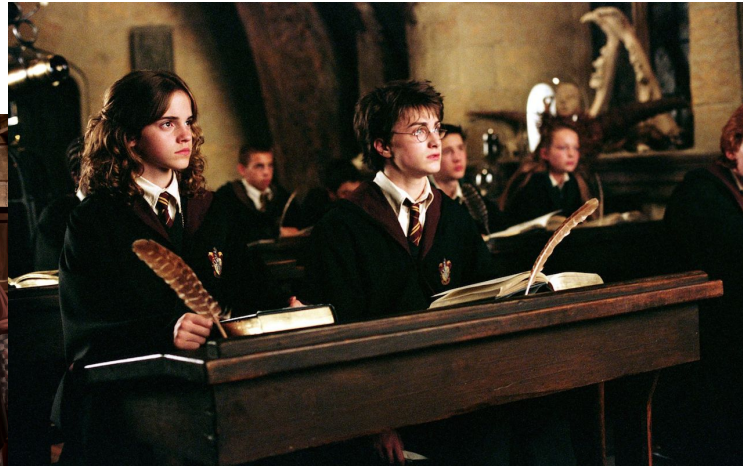 a fast interconnect (InfiniBand), **Slurm job scheduler**, ...

Here's a question...

Why are there more wizards in Harry Potter than in Lord of the Rings?

Here's a question...

Why are there more wizards in Harry Potter than in Lord of the Rings?

Hogwarts School of Witchcraft and Wizardry!

WHY Throwaway supercomputers?!

- **Supercomputers are typically very busy...**

- Lots of scientists are queueing up jobs to do their research

- Some people are very **impatient**, they want their jobs to start ***now***

- Junior researchers are not familiar yet with HPC infrastructure

  - They **need to be trained** first to make good use of it

  - That's a bit difficult on a **busy production system**

  - You want to make sure that **people can make mistakes**

- Some research groups have some grant money to spend on compute...
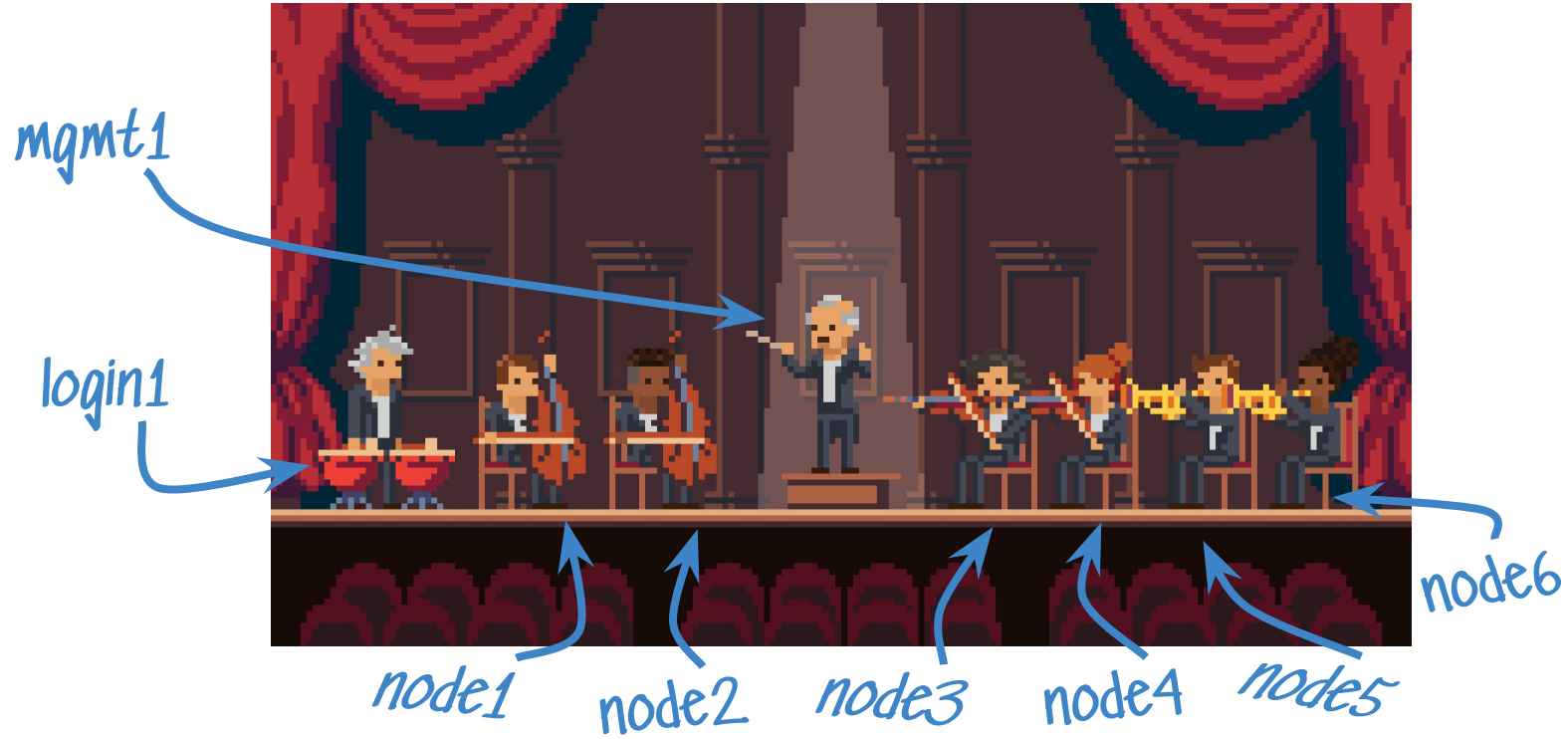
Magic castle in a nutshell

- Set of VMs in the cloud, connected through local network

- Login node + management node + set worker nodes

- Slurm, LDAP, JupyterHub, FreeIPA, DNS, NFS, ...

- Cloud instances created via Terraform

- Configured as an HPC cluster via Puppet

- Configuration through YAML files

https://github.com/ComputeCanada/magic_castle

Magic castle in a nutshell

mgmt1

login1

node1 node2 node3 node4 node5 node6
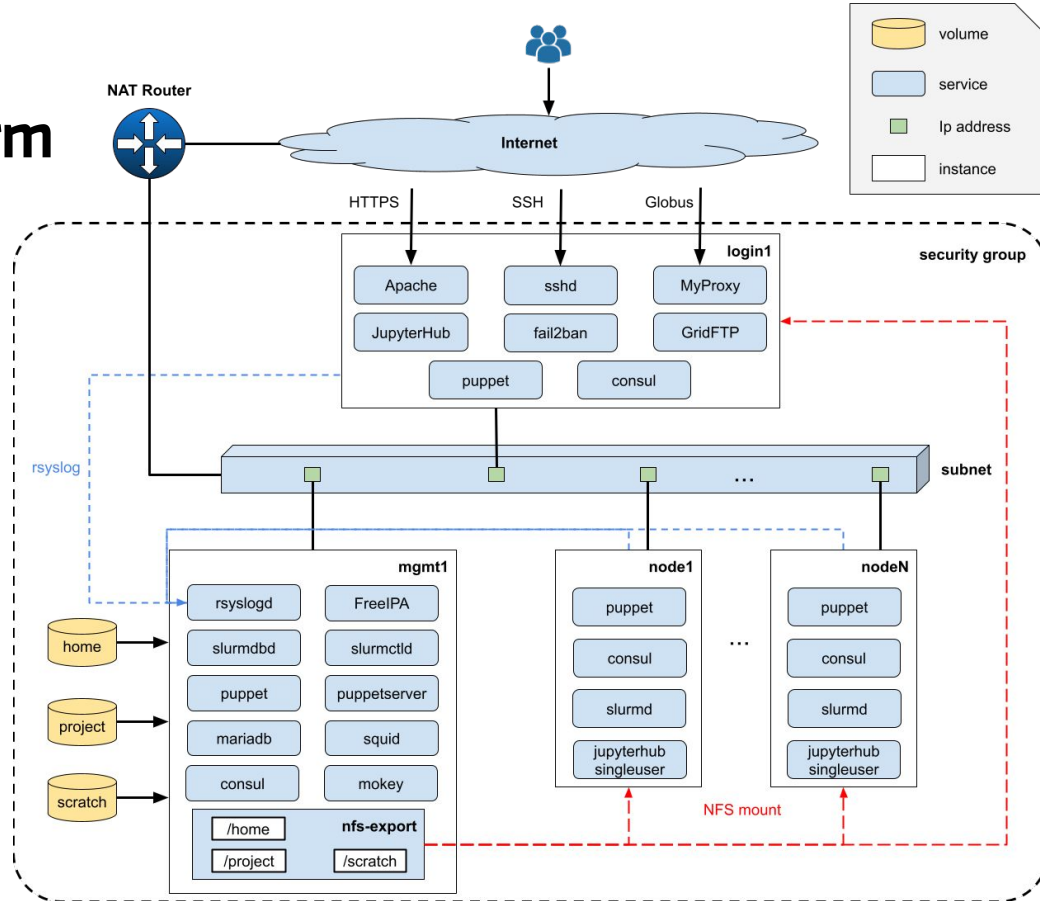
MAGIC CASTLE

Magic castle in a nutshell

Magic castle in a nutshell

- **Infrastructure-as-Code** (IaaC)

- Puppet + Terraform (or OpenTofu) + cloud-init

- **Mostly vendor/cloud-agnostic**

- Abstracts away the differences across different clouds

- Support for autoscaling of nodes (via Terraform Cloud)
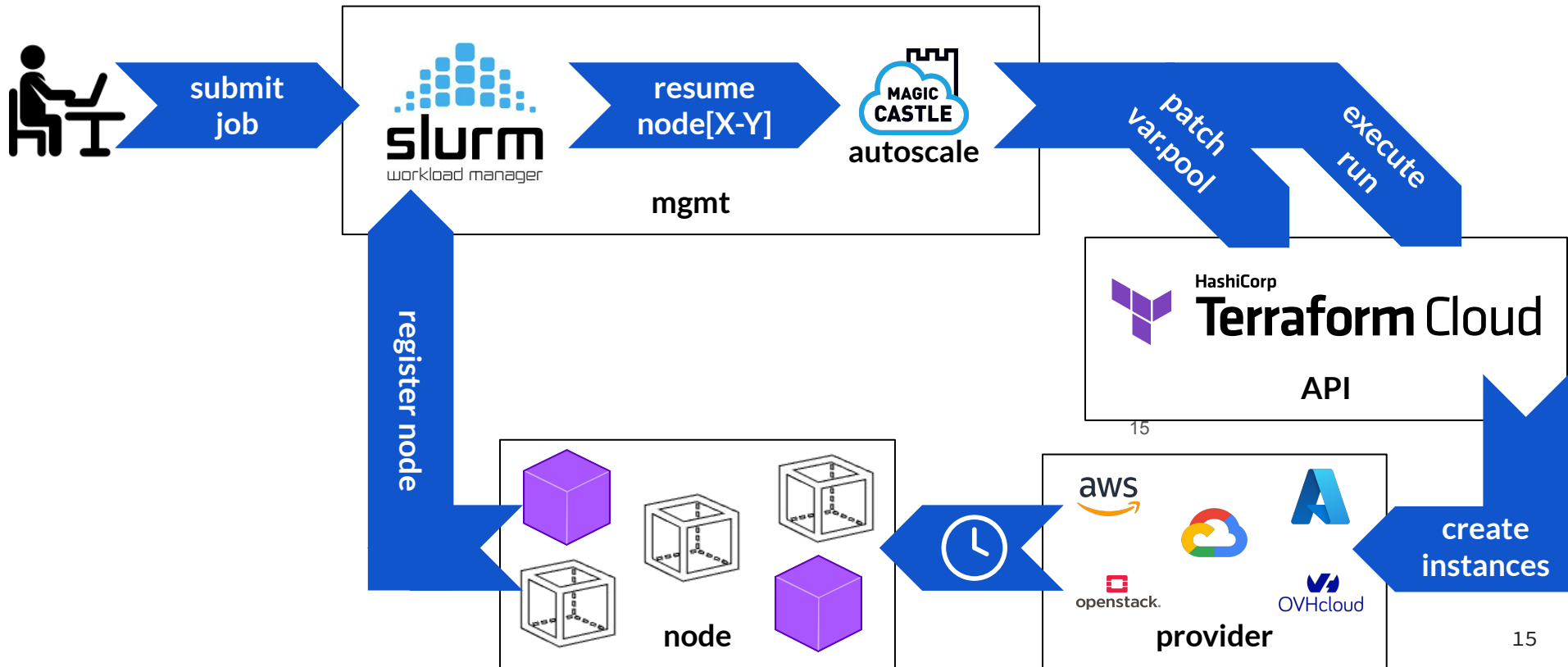
- Actively maintained, extensively documented

https://github.com/ComputeCanada/magic_castle
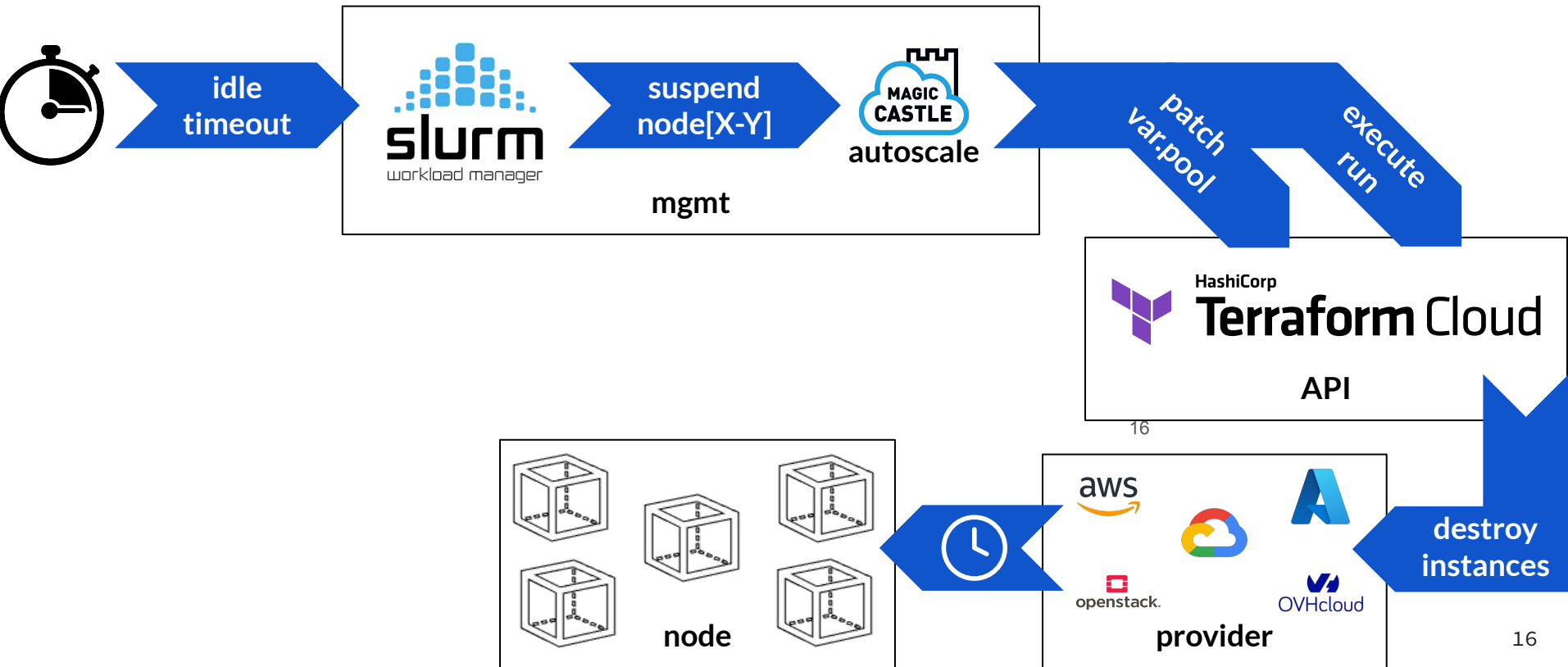
Magic castle: key features

- Support for:

    - Mixing of x86_64 and Arm64 instances as workernodes

    - High-speed networks (EFA in AWS)

    - Worker nodes with GPUs

- Auto-scaling: start (more) nodes as jobs are submitted

    - Via Terraform Cloud, so cloud-agnostic

    - Cloud credentials are in Terraform Cloud, *not* on your cluster!

Magic castle: autoscaling (resume)

Magic castle: autoscaling (suspend)

idle timeout → slurm workload manager → suspend node[X-Y] → MAGIC CASTLE autoscale

mgmt

patch var.pool / execute run

HashiCorp Terraform Cloud

API

16

destroy instances

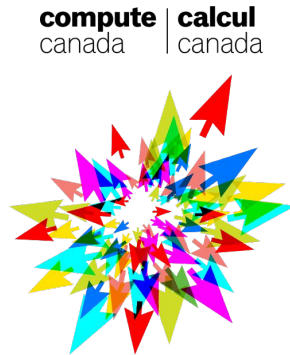aws / openstack / OVHcloud

provider

node

16

What about the (scientific) software I need?!

CernVM-FS

- Shared software stacks are accessible via CernVM-FS

  - ComputeCanada software stack

  - European Environment for Scientific Software Installations

    (EESSI, eessi.io)

- **Time for quick demo?**

EESSI
EUROPEAN ENVIRONMENT FOR
SCIENTIFIC SOFTWARE INSTALLATIONS

EESSI (DEMO)

```
/cvmfs/software.eessi.io/versions/2023.06/software
`-- linux
    |-- aarch64
    |   |-- generic
    |   |-- neoverse_n1
    |   `-- neoverse_v1
    `-- x86_64
        |-- amd
        |   |-- zen2
        |   `-- zen3
        |-- generic
        `-- intel
            |-- haswell
            `-- skylake_avx512
                |-- modules
                `-- software
```

```
$ source /cvmfs/software.eessi.io/versions/2023.06/init/bash
Found EESSI pilot repo @
/cvmfs/software.eessi.io/versions/2023.06!
archspec says x86_64/intel/skylake_avx512
Using x86_64/intel/skylake_avx512 as software subdirectory
...
Environment set up to use EESSI pilot software stack, have fun!

$ module load R

$ which R
/cvmfs/software.eessi.io/versions/2023.06/software/linux/x86_64/
intel/skylake_avx512/software/R/4.2.1-foss-2022a/bin/R

$ R --version
R version 4.2.1
```

Current Shortcomings

- At the mercy of OS updates

- Some cluster designs may not be (easily) possible

- Some features not supported yet (like EFS in AWS, IB in Azure)

- Only Slurm supported a resource manager & job scheduler

- Built to replicate a ComputeCanada cluster via Puppet template

- Documentation is extensive, but hosted in GitHub repo

Magic castlE: Current Developments + Ideas

- Support for NVIDIA Multi-Instance GPU (MIG)

  - Splitting massive GPUs in multiple smaller virtual GPUs

- Support for additional shared filesystems like EFS in AWS

- Broader support for fast interconnect (InfiniBand in Azure)

- Using Material for MkDocs for the documentation

Would you like to know more?

- Magic Castle tutorial at Supercomputing'23,

  training material available via *hackmd.io/@MagicCastle/SC23*

- **Free online SIG-HPC Education webinar**

  by Félix-Antoine Fortin (lead developer of Magic Castle)

  Thu 8 Feb 2024 at 11:00 EST (17:00 CET)

  sighpceducation.acm.org/events/magic_castle

Creating Throwaway Supercomputers in the Cloud with

MAGIC CASTLE
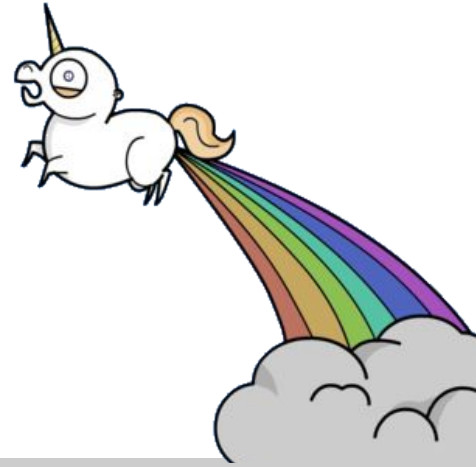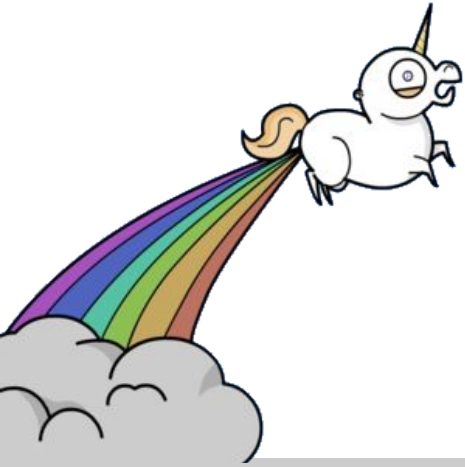
CfgMgmtCamp | 20240205 | Ghent

Kenneth Hoste | HPC-UGent | kenneth.hoste@ugent.be

Creating Throwaway Supercomputers in the Cloud with

MAGIC CASTLE

CfgMgmtCamp | 20240205 | Ghent

Kenneth Hoste | HPC-UGent | kenneth.hoste@ugent.be