www. chameleoncloud.org

CHAMELEON: CHANGING THE WAY WE SHARE

Kate Keahey

keahey@anl.gov

University of Chicago, Argonne National Laboratory

April 21, 2021
Virtual Café for Robust Science

CHAMELEON IN A NUTSHELL

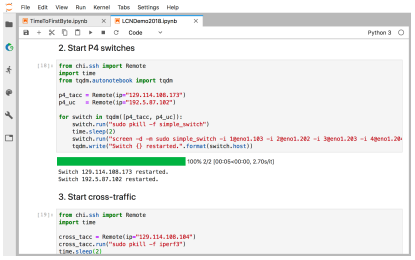
- We like to change: a testbed that adapts itself to your experimental needs
  - Deep reconfigurability (bare metal) and isolation
  - power on/off, reboot, custom kernel, serial console access, etc.
- Balance: large-scale versus diverse hardware
  - Large-scale: ~large homogenous partition (~15,000 cores), ~6 PB of storage distributed over 2 sites (UC, TACC) connected with 100G network
  - Diverse: ARMs, Atoms, FPGAs, GPUs, Corsa switches, etc.
- Cloud++: leveraging mainstream cloud technologies
  - Powered by OpenStack with bare metal reconfiguration (Ironic) + “special sauce”
  - Blazar contribution recognized as official OpenStack component
- We live to serve: open, production testbed for Computer Science Research
  - Started in 10/2014, available since 07/2015, renewed in 10/2017, and just now!
  - Currently 5,300+ users, 700+ projects, 100+ institutions, 300+ publications

REPRODUCIBILITY BUILDING BLOCKS

- Hardware
  - >105 hardware versions over 5 years
  - Expressive allocation
- Clouds: images and orchestration
  - >130,000 images, >35,000 orchestration templates and counting
  - Portability and federation
- Packaging and repeating: integration with JupyterLab
- Share, find, publish and cite: Trovi and Zenodo

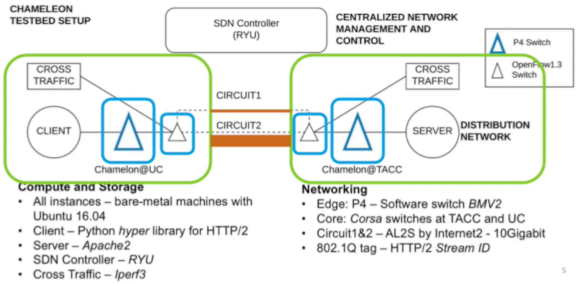
SHARING EXPERIMENTS: PACKAGING

Jupyter Notebooks



Experimental storytelling: ideas/text, process/code, results

+



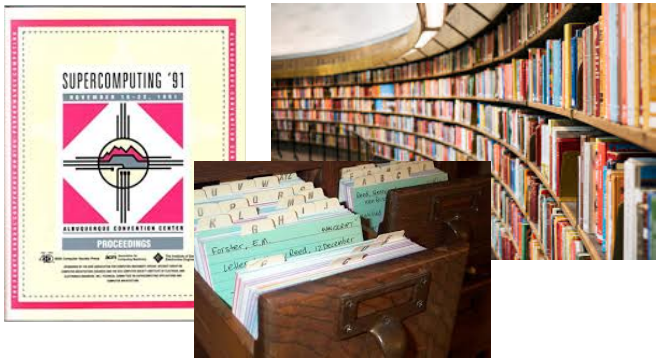
Complex Experimental containers

- Repeatability by default: Jupyter notebooks + Chameleon experimental containers
  - JupyterLab for our users: use jupyter.chameleoncloud.org with Chameleon credentials
  - Interface to the testbed in Python/bash + examples (see LCN'18: https://vimeo.com/297210055)
  - Named containers: a terminal multiplexer for various componentes of your experiment

Also see: "A Case for Integrating Experimental Containers with Notebooks", CloudCom 2019

SHARING EXPERIMENTS: PUBLICATION

Familiar research sharing ecosystem



Digital research sharing ecosystem

Chameleon jupyterhub

?

- Trovi: a digital sharing platform
  - Make your experiments sharable within a community of your choice with one click
  - A library of reproduced experiments from foundational papers for research and education (see e.g., Brunkan et al., "Future-Proof Your Research", SC20 poster)
- Integration with Zenodo: make your experimental artifacts citable via Digital Object Identifiers (DOIs) (export/import)

zenodo

PARTING THOUGHTS

- Time to reproduce is critical: much attention is being given to packaging experiments for repeatability/reproducibility – not as much to actually repeating them
- We need to create a “marketplace” for repeating research
  - Repeatability and reproducibility can be thought of as the same thing at different “price points”
  - Recognition for published digital artifacts (software, data, experiments, etc.)
  - Starting early: education is an unappreciated tool for fostering reproducible research
- Use what you have: leveraging testbeds, existing digital artifacts,frameworks, patterns, etc. has the potential to lower the ”price” of reproducibility and make it affordable

We're here to change

www.chameleoncloud.org

keahey@anl.gov