SI2-SSE: Scaling Up Science on Cyberinfrastructure with the Cooperative Computing Tools

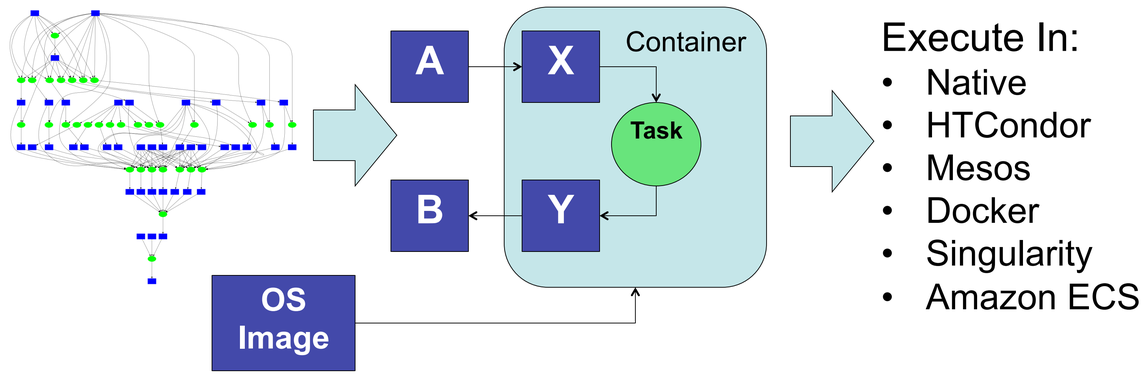
PI: Douglas Thain, University of Notre Dame
(NSF-OCI-1642409)

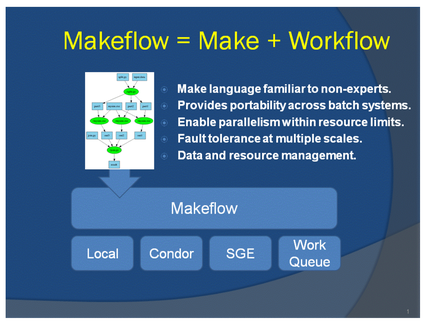
Our Goal: Make it easy to harness multiple kinds of infrastructure for ordinary applications with no special privileges.

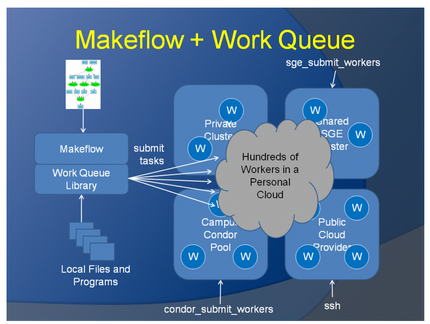
Typical User Has an existing code or dataset that works well on a local machine, and now wants to run on 1000s of nodes drawn from a campus cluster, a public cloud, and a national resource.

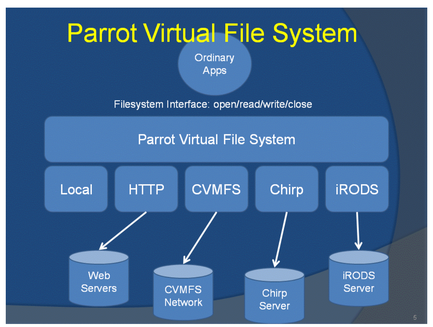
Annual Workshop at ND

Integrating Containers and Workflows

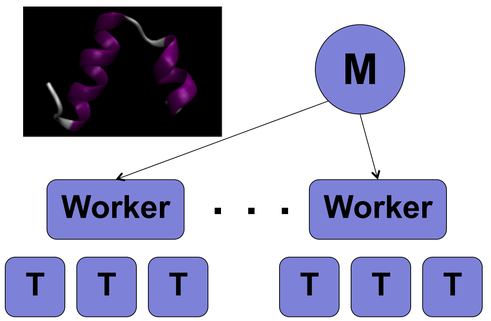








HP24stab Simulated w/Work Queue

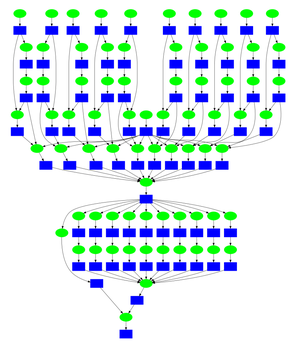


177us of simulated time in 18 days using 10K tasks on 4K cores.

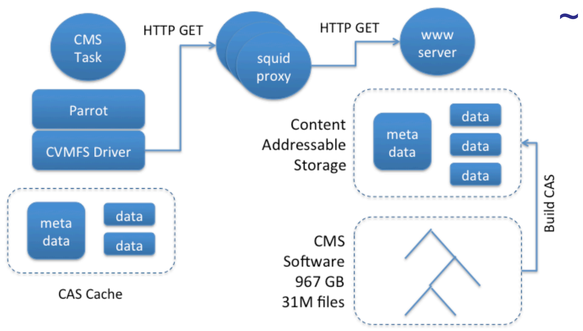
Lifemapper w/Makeflow Species Distr. Modeling

12,500 species
x 15 climate

scenarios
x 6 experiments
x 500 MB per
projection
= 1.1M jobs, 72TB of output



CVMFS Filesystem

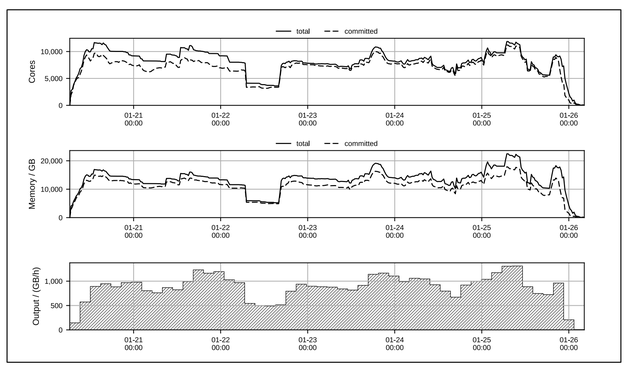


http://cernvm.cern.ch/portal/filesystem

Jakob Blomer et al, **The Evolution of Global Scale Filesystems for Scientific Software Distribution,** *IEEE/AIP Computing in Science and Engineering*, **17**(6), pages 61-71, December, 2015
DOI: 10.1109/MCSE.2015.111

Lobster HEP Workflow

~10K cores over 7 days, 25K cores over 1 day



Matthias Wolf et al, **Opportunistic Computing with Lobster: Lessons Learned from Scaling up to 25k Non-Dedicated Cores**, Computing in High Energy Physics, 2016.

- Open source, GPL License, approx 3 releases per year.
- Development focus on Linux and Mac.
- Automated build and test using HTCondor + Docker
- Users on Blue Waters, XSEDE, OSG, CyVerse, Jetstream.
- 1/4FTE + two students: release, outreach, and devel focus.
- Used in classes at ND, UWEC, and University of Arizona..
- One workshop at ND + several road tutorials per year.

http://ccl.cse.nd.edu