Ten Roadmap Challenges for the Campus Grids Area Years 3-5

OSG Staff Retreat

May 14,15 2014

Introduction

- Following guidance:
  1. Provide value to VOs and resource owners
  2. Provide opportunities for the long tail of science
  3. Promote use of DHTC for science
- The following slides summarize the challenges
  - Mostly relevant to #2, #3 (but #1 covered by "incentives" discussion)
  - Should suggest Campus Grids Area plan for years 3-5
  - Makes sense to write down the concrete program of work following discussion on Wednesday

Roadmap challenges (1)

- User code and software dependencies
  - Users expect to find a software environment
  - Typical ask is for a package or set of libraries or database
  - But we give users a virtual cluster with no software installed, command tools, or how-to guides
    - This makes us an odd fish in the world of computing (and we don't have to be)

Roadmap challenges (2)

- The uneven exe environment
  - Uniformity in the build and execution environments is desired
    - *"My main issue is that the applications do not execute on every computer because of missing shared libraries, and I need to prepare the applications statically or find the best way around these failed jobs" (ALGDock)*
  - Minimizing differences between home campus cluster and the remote computing environment
  - Limited support for multi-core slots

Roadmap challenges (3)

- Bringing data and processor together
  - *"Data access has been an issue, testing Parrot + Chirp to see if that will provide a solution" (aprime)*
- Simplify data transfers from campus to OSG and back
  - Preferably transparent to the user
  - Easy "dropbox" (drag-copy) and "netflix" (stream) functionality would be desirable

Roadmap challenges (4)

- Campus cluster integration issues
  - Submitting to OSG Connect from the home campus cluster
    - A number issues: extra software to install on the HPC login node, workflow modifications, data delivery issues
  - Presently cannot use multiple project/user accounts from OSG Connect to Bosco-flocked resource targets

Roadmap challenges (5)

- Adding (non-OSG) resource targets to OSG Connect
- Such as TACC/Stampede, or smaller clusters such as SLU (St. Louis University)
- Bosco-based Glidein Factory deployed
  - Service instance for each resource target
- Issues:
  - Virtually mounting OASIS via Parrot is still a challenge, given frailty of libcvmfs
  - Squid cache is needed, either local or nearby
  - SLU needs a proxy manager
  - And then there is data access

Roadmap challenges (6)

- User-centric job & task monitoring & report
  - Live reporting back to the user
    - Some users requested condor_ssh_to_job (possible through Glideinwms?)
  - Job scouts as estimators of viability and TTC, avoidance of black holes, congestion
- User, Project centric (not VO centric) view of accounting and resource usage
  - E.g. available from the command line or in a user portal

Roadmap challenges (7)

- Operations in support of users
  - Instrumenting standard job wrappers
  - Tracing & cataloging job profiles
    - Software used
    - Task to jobs
    - Failure rates, CPU efficiency, memory usage, data delivery
      - Jobs database?
  - User feedback based on norms or best practices

Roadmap challenges (8)

- Incentives for campus HPC consultants to support DHTC
  - Using local cluster allocation and opportunistic DHTC together?
  - Unnatural mode shifts, extra "stuff"
- Incentive for research computing directors (resource owners) to support DHTC
  - Convince we're adding value (not only cycles but collaborative, bridged environments) vs. removing users
  - Provincial empires vs. integrated national cyber resources
    - Perhaps our focus should be the long tail of institutions
      - » There are about 500 national universities plus 4-year liberal arts colleges in the U.S.

Roadmap challenges (9)

- No standard way of doing things
  - A consultant should not be required for the most popular workflows
    - *"My workflow is not very well established yet because my programs are still under development. Right now I am using UCSF dock6 and my own python-based programs." (ALGDock)*
  - Crisp, polished tutorials and FAQs needed to guide new users into the system, minimizing differences from existing campus cluster instructions where possible
  - Simple (easily copied and adapted), domain-centric and toolkit-centric examples

Roadmap challenges (10)

- Working with campus HPC centers
  - User support resources on campuses are many, how can we use them?
  - How can we work with external support groups (e.g. Software Carpentry, ACI-Ref)?
    - Co-sponsor boot camps, e.g. ?