Open Science Grid

Open Science Grid Staff Retreat

June 4, 2013

Lothar Bauerdick/Fermilab

Changes in Project Organization

- Three main areas that we deliver on
  - ★ Production – including Operations, Security, and Campus Infrastructure
  - ★ Consulting – including technologies, architectures and user support
  - ★ Technologies – including software packaging, system testing, patching
- Changes in Area Coordination
  - ★ Merged Software and Technologies
    - Brian as Technology area coordinator is now member of ET
    - Tim C Software lead, moved Software Release Mgmt w/ TimT to Production Area
  - ★ Production: Area Coordination vacant for the moment, Dan stepped down
    - Rob Q Operations Manager, runs Production Meeting and is invited to ET mtgs
    - Dan stays on as Bosco project lead
  - added a number of sub-area coordinators:
    - Campus Grids and Campus Infrastructure Communities: Rob G
    - Network Monitoring: Shawn McKee
    - Software Releases: Tim T
  - ★ User Support: Chander and Security: Mine stays

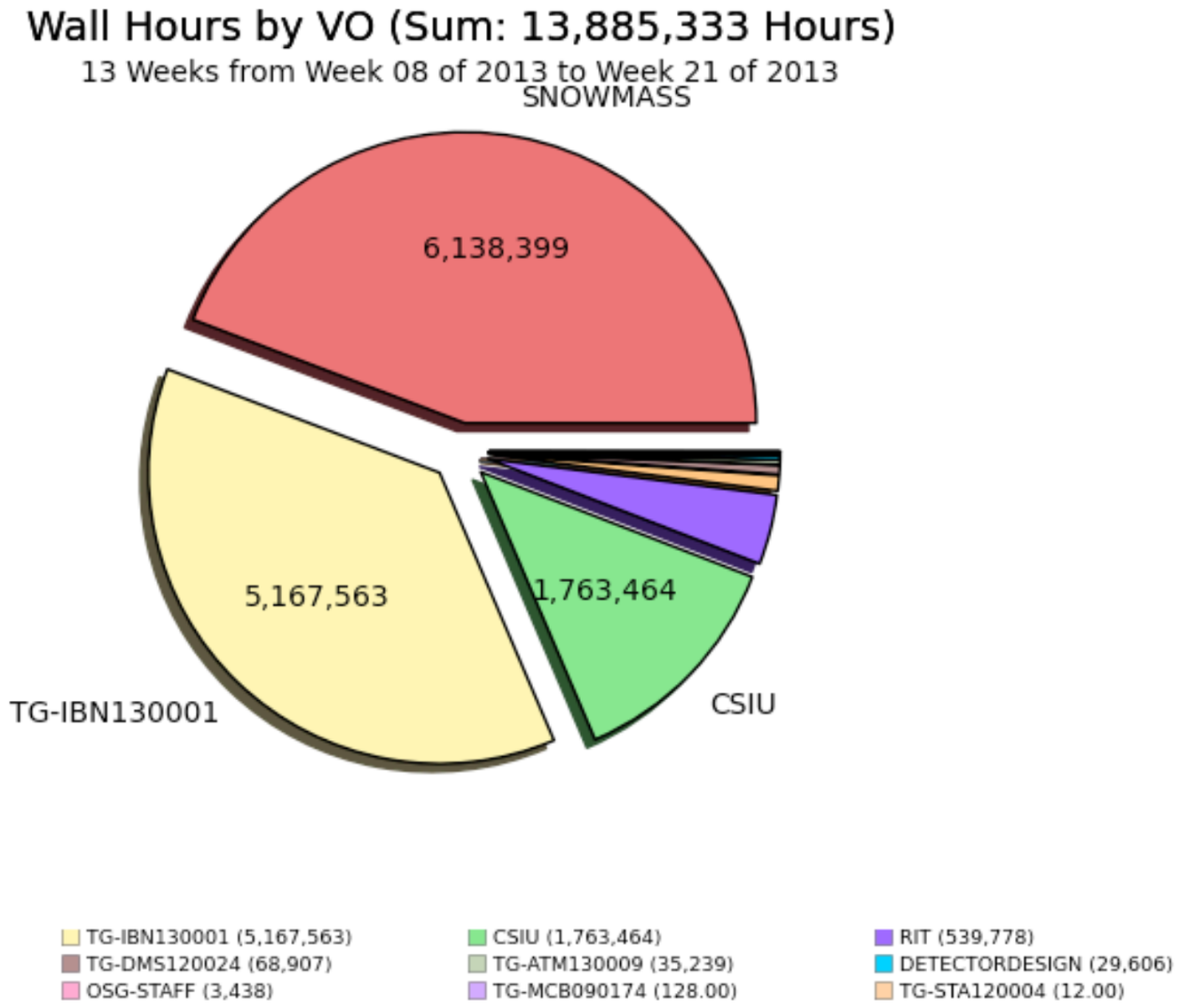
Open Science Grid

Focus of this talk

- Extending the OSG — how should we do it?
- Continue on our path, like providing excellent operations, software, security, user support.
- But let us also focus on new ways of adding value to the OSG eco system

My Goal for OSG

- Deliver 80 M CPU hours of “new resources” in 12 months!
  - ~10% of last year’s OSG resources)
- ★ by accountable opportunistic and extended resources
- ★ lets find out what’s required to deliver on this goal



80M hours!

Start of a punch list toward this goal

Open Science Grid

- Add more users that are potential customers for opportunistic cycles
  - new users and science groups, like the Snowmass group
  - existing VOs like the LHC and others
  - XSEDE, where we through XRAC we offered additional 8Mh/quarter
  - Bosco focus on enabling new applications (R), new communities
- Increase the availability of cycles
  - make it more attractive to provide resources opportunistically
    - tear down some of the barriers, like certificates
    - add functionality like accounting/reporting to make it “accountable” for providers
- Increase the usability of opportunistic cycles
  - add functionality like data access/staging-- what is missing and would make a difference?
- Increase the ability to add new resources
  - add functionality to easily provide resources where they become avail
  - like the SDSC exercise which brought in additional ~2M hours?
- ETC

Goals for User Support

- Focus is and was to help the individual researcher and VOs
  - component of bringing in new users/researchers
    - How about last year's campus research club?
    - How about adding campus users and owners of campus infrastructures?
  - component of making those new users effective
    - --> very significant resource of “opportunistic” CPU cycles
      - XSEDE committment of 2Mh/quarter over-delivered by large factor
        - only now is it that large VOs like LHC are catching up
- with XSEDE a significant “technology/operations” component
  - working with the Operations team and Technology areas
  - US “owns” this area at the moment, how about the other aspects of working with XSEDE that are not US? Like interfacing OSG into XSEDE, e.g. allowing OSG PI to use XSEDE as an extension to its OSG workflow
  - a number of new needs being identified in accounting, security/IDM, ...
- new focus of OSG Connect being discussed, see later today

Goals for Production Area

- Sustain the OSG and provide excellent services and operations
- Spread the DHTC idea using “submit local, compute global”
  - ★ helping Joe the scientist to use HTC and DHTC
  - ★ how successful is BOSCO? re-focussing on providing BOSCO to the R community — what reach do we expect? Is this sufficient?
- Extending and growing the OSG
  - ★ spread across the campuses from “beachheads”, is this successful?
  - ★ making it attractive for campus resources to join in and share resources
  - ★ making OSG *the* place to go to get *and to give* opportunistic cycles
- propose to add a focus on "accountable" provisioning of opportunistic CPU cycles to scientists and groups
  - ★ a number of issues/missing features, like in accounting, "informatics"
  - ★ re-focussing some of the UChicago, Fermilab efforts in this direction

Production Area

- networking sub-area
  - plans for moving network monitoring to “standard operations”
  - what will we do forward-looking, making networking
  - request by WLCG -- is this the right next step? Where do we go with this?
- software releases and testing
  - procedures for releases, needs for testing, ITB etc

Open Science Grid

Goals for Technology Area

- Merge b/w software and technology area vs production
  - would like to hear about remaining open issues
- Transition for specific technologies in progress
  - e.g. CE, software provisioning, etc
- deliver towards the goals of Production and User Support Areas
  - lets understand/list where to focus to address strategic goals
  - e.g. discussion yesterday on how to provision/spool data
- what's needed for "extending the OSG"
  - Like interfacing OSG into XSEDE
  - Like making the provisioning of opportunistic cycles a more "accountable" and reliable value proposition
- data access
  - what are the key functionalities we should add?
- adding on-demand resources to the infrastructure
  - clear interest by consortium, including commercial cloud resources

Goals Security Area

- PKI / Certificate-based infrastructure is expensive in many ways
  - now that we “own” it, can we phase it out in the medium term?
  - develop a plan on what needs to be done
  - bring down the number of Certs: users, VOs, site,
  - move everyone else to other CAs, like CI-login, CERN CA?
- traceability --> remodel the "trust relationships"
  - roles of VO vs sites vs OSG
  - let's really understand the "OSG VO"
  - provide VO information systems
- will hear about more goals and needs from Mine

80M hours!

OSG Project Goals in Proposal

- Sustaining the OSG
- Transform computing on campuses through new DHTC technologies
- Transformation of our core communities computing capabilities
- Access to an expanded set of job and data services accessible via a single identity
- Improve the usability, expand the usage, lower barriers of adoption

Goals/Deliverables/Milestones

- Annual Goals in Sustaining the OSG:
  - a) Increase in CPU usage that is more than Moore's Law, showing growth to meet the scaling needs of the users.
  - b) Full compliance with the operations service SLAs.
  - c) Doubling of accounted data movement to meet growth in data intensive science.
  - d) >=2 additional communities using DHTC services in production; >2 tutorials and documentation for new capabilities.

- try to understand the supply and demand for campuses on a day-by-day basis, understand the impediments to utilizing OSG as a campus organization, and to use Glow as our guinea pig

- understand impedance

Transform computing on campuses through new DHTC technologies:

- Year 1: a) Deploy technology to account usage of users, jobs and data movement to campuses.
  - ★ b) Release campus infrastructure software distribution Production Version 1.
- Year 2:
  - ★ a) Assess metrics that encapsulate a measure of adoption and increase in usage.
  - ★ b) 20% increase in each of new, and usage by existing, users of campus technologies.
  - ★ c) 2 production versions of campus software to extend the capabilities and scalability.
- Year 3:
  - ★ a) 20% increase in new, and usage by existing, users of campus technologies.
  - ★ b) Transition of appropriate campus support services to sustained OSG operations.

Transformation of our core communities computing capabilities to exascale science

Open Science Grid

- Year 1:
  - a) 10% of resources support end-to-end capability to be schedulable as HTPC and simultaneously usable and available by HTC/single processor job.
- Year 2:
  - a) 20% of resources support end-to-end capability to be schedulable as HTPC and simultaneously usable and available by HTC/single processor job.
  - b) 90% of LHC workload and 10% of non-LHC workload supports remote I/O capabilities.
  - c) Data movement across the DHTC fabric of >750 Petabytes/year.
- Year 3:
  - a) >50% of resources support end-to-end capability to be schedulable as HTPC and simultaneously usable and available by HTC/single processor job.
  - b) 50% of non-LHC workload supports remote I/O capabilities.
  - c) Data movement across the DHTC fabric of >750 Petabytes/year

Access to an expanded set of job and data services accessible via a single identity

- Year 1:
  - a) 10% of non-LHC Users accessing OSG services using campus identities.
  - b) Complete the architecture and design of the new set of ID management services.
- Year 2:
  - a) 20% of non-LHC Users accessing OSG services using campus identities.
  - b) Deliver initial release of new set of ID management services.
- Year 3:
  - a) >30% of non-LHC Users accessing OSG services using campus identities.
  - b) Transition new set of ID management services to production and operations.

Improve the usability, expand the usage, lower barriers of adoption

Open Science Grid

- Year 1:
  - a) Deliver report on integration of virtualized resources into the OSG fabric of services.
  - b) Prototype integration of one cloud resource into the production DHTC environment.
  - c) 100% VDT packages available as RPMs.
  - d) Production release of configuration management of RPM-packaged VDT software.
  - e) Deliver report on extending the job-level monitoring.
- Year 2:
  - a) Improve reliability of software distribution via duplication of software repositories. Provide hosting of non-VDT software on behalf of OSG communities.
  - c) Deliver report on integration of advanced (100G, Terabit) networks.
  - d) Extend existing Dynamic Resource Allocation Services to include centralized policies that regulate allocation.
  - e) Collaborate with U.S. LHC for initial deployment of simplified data services for non-LHC.
  - f) Integration of one cloud resource into the production DHTC environment.
  - g) Provide enhanced OSG-wide job monitoring in prototype.
- Year 3:
  - a) Develop a trust flow diagram of VDT stack. Identify the effect of configura- tion parameters on the security of the software.
  - b) Provide OSE services to extend the dynamic resource allocation capability.
  - c) All software available as source RPMs. Drop support for Pacman.
  - d) 20% of resources support simplified data management services for non-LHC VOs.
  - e) Full integration of multiple cloud resources into the production DHTC environment.
  - f) Transition enhanced OSG-wide job monitoring into operations.