Open Science Grid

Open Science Grid ED Report to Council Meeting

June 13, 2013

Lothar Bauerdick

Three Main Areas OSG Delivers On

Open Science Grid

- Production
  - ★ including Operations, Security, and Campus Infrastructure
- User Support
  - ★ consulting on technologies, architectures and user support
  - ★ spreading knowledge on HTC as a science problem solver
- Technologies
  - ★ including software packaging, system testing, patching
- Talk with concentrate on reporting status, outlining plans as discussed at the recent OSG staff retreat
  - ★ 30,000 ft picture

Changes in Project Organization

- Merged Software and Technologies
  - Brian Bockelman/UNL Technology Area Coordinator with Software added to area
    - Technology Area Coordinator is now member of Executive Team
  - Tim Cartwright/UW leads the Software Team within the Technology Area
  - New Software Release Team headed by Tim Theisen/UW
  - Software Releases moved to Production Area, reporting to Production Meeting
- Production Area Coordination vacant for the moment
  - Dan Fraser stepped down to have time for his new project outside OSG
    - Dan stays on as Bosco project lead (25% effort)
  - Rob Quick Operations Manager runs Production Meeting
    - is now invited to ET meetings to keep information flow
  - added a number of sub-area coordinators:
    - Campus Grids and Campus Infrastructure Communities: Rob Gardner/UChicago
    - Network Monitoring: Shawn McKee/MSU
    - Software Releases: Tim Theisen/UW
- User Support: Chander Sehgal/Fermilab
- Security Coordinator: Mine Altunay/Fermilab

Security Area

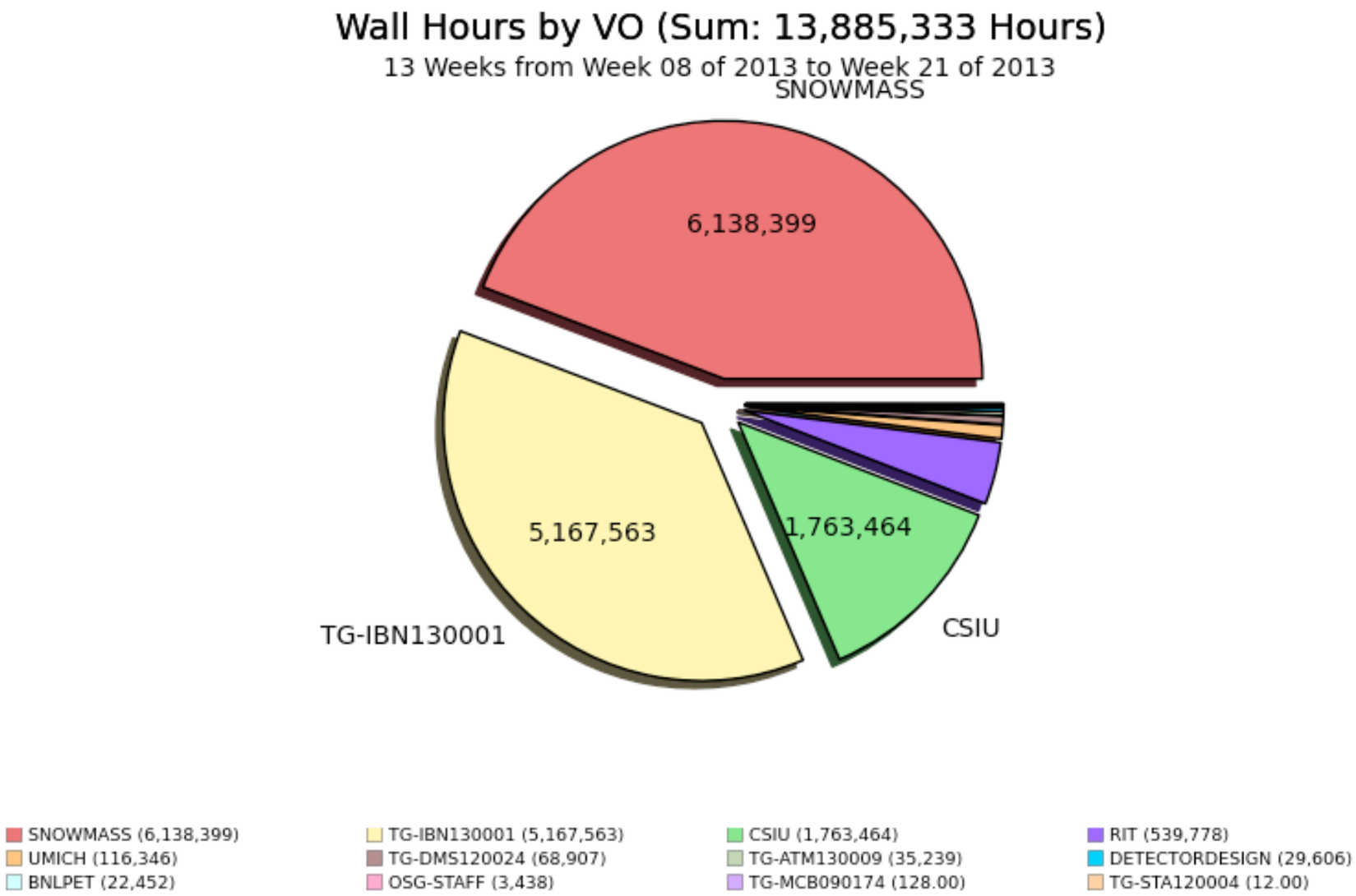
- ★ Completed OSG PKI Transition
- ★ Operations, Various vulnerabilities, Pakiti, SHA2, Audit requirements etc
- ★ Identity management—CILogon Basic CA, Traceability of Jobs without Certificates, Survey of 5 OSG Resource Providers where and how IDM info is used, Identity Management Roadmap

- ✦ PKI / Certificate-based infrastructure is expensive in many ways
  - ★ now that we “own” it, can we phase it out in the medium term?
    - ✦ bringing down the number of Certs: users, VOs, site
    - ✦ move others to alternatives CAs, like CI-login, CERN CA for LHC
- ✦ Understand and model the "trust relationships"
  - ★ OSG overlay infrastructure has pilot jobs that then run “pay load” on behalf of user/VO — access to overlay determined by VO
  - ★ access control vs traceability, blocking users vs blocking VOs
- ✦ Federating ID management
  - ★ do as much as we can with existing site identities
  - ★ what is needed for the OSG VO for non-XD users?
- ✦ OSG Security Team at ~2 FTE
  - ★ Mine Altunay (FNAL) 70%, Kevin Hill (FNAL) 80%, Anand Padmanabhan (NCSA-UIUC) 50%, Von Welch (IU) 5%

User Support Area

- Focus is and was to help the individual researcher and VOs (1.7 FTE)
  - new approaches like Bosco, campus researchers club, with ~ok but limited reach
- OSG supported 15 researchers / communities with 21M hours
  - 10M to PIs coming in through the XD allocation committee, 11M to OSG VO customers
- User Support facilitates OSG's function as Service Provider to the XD program
  - high quality production DHTC "facility" to XD via the OSG-XSEDE interface
  - built effective relationships and teamwork with XSEDE counterparts
  - agreed to supply 2M hours/quarter + up to 8M/quarter for additional allocations
- At this point we do not see shortage of opportunistic cycles
  - last quarter we averaged 1M per week; peak day was >380K hours
    - to some extent constrained by resource limits at OSG-XSEDE front-end, being fixed
- Do a good job of assisting new communities in leveraging OSG DHTC computing
  - they recognize and appreciate us, with their sponsors, for enabling their science
  - but they mostly do not transform into vibrant new VOs who effectively contribute to / promote OSG
- gWMS is great for accessing lots of cycles at many sites
  - but this layer makes it even harder to understand performance issues in a distributed fabric
- Public storage using iRODS useful for OSG-VO researchers with large data (2-30GB)
  - but no general demand from other VOs, just continue this service for opportunistic customers

Opportunistic Use of OSG



- Opportunistic Cycles benefit new users and existing members of OSG Eco System
  - ★ example is the recent use of >6M hours by the “Snowmass VO”, a community of HEP theorists and experimentalists
- OSG Connect service to allow campus users to run on OSG using opportunistic cycles
  - ★ goal to make the service more accountable opportunistic and extended resources

Production Area: Operations

- Sustain the OSG and provide excellent services and operations
  - Over the year delivered ~722M CPU hours and transferred ~372PB of data
    - good quality services, no SLA exceptions (Service or Support)
  - a list of new services introduced like OASIS new VM, PKI and Campus Grids into OIM/MyOSG (Link), and Network Monitoring into MyOSG (Link
  - Continued interoperation with WLCG/EGI services
- OSG Operations and Production Team
  - Indiana University for the GOC
    - Rob Quick 90%, Scott Teige 100%, Kyle Gross 100%, Soichi Hayashi 80%, Tom Lee 100%, Sarah Schmeichen 50%, Elizabeth Prout 100%, Alain Deximo 50%, Chris Pipes 100%, TBD 100%
  - UCSD is running the glideinWMS services
    - Igor Sfiligoi 20%, Jeff Dost 60%, Alex Georges/Tim Mortensen 60%, Terrence Martin 10%
  - Fermilab services like accounting etc
    - 1 FTE across Keith Chadwick, Steve Timm, Tanya Levshiva, John Weigand
  - UNL
    - TBD 20%
  - UC
    - Marco Mambelli 25%

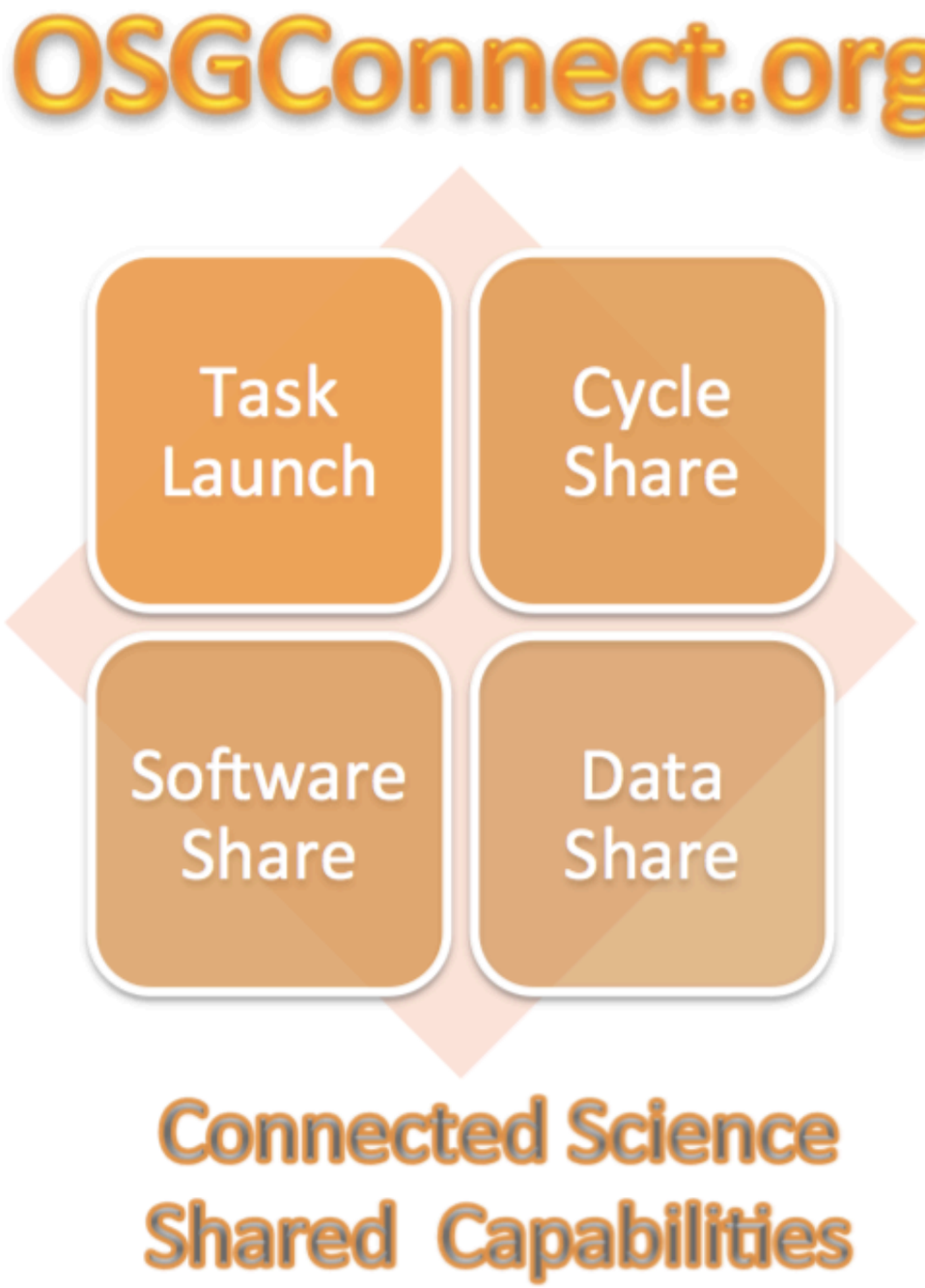
Production Area: Growing

- Spread the DHTC idea using “submit local, compute global”
  - helping Joe the scientist to use HTC and DHTC
    - bring HTC to the campuses, enable researchers to submit locally and run globally
  - BOSCP after initial release, now focus on enabling the R community
  - how successful is BOSCO and how successful does it need to be?
  - what reach do we expect? Is the initial focus on R sufficient?
- Campus Infrastructure Community
  - active with meetings and webinars
- 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
- New Campus Grids focus on “OSG Connect” service
  - working with User Support and Operations
  - "accountable" provisioning of opportunistic CPU cycles to scientists and groups
  - a number of issues/missing features, like in accounting, "informatics"

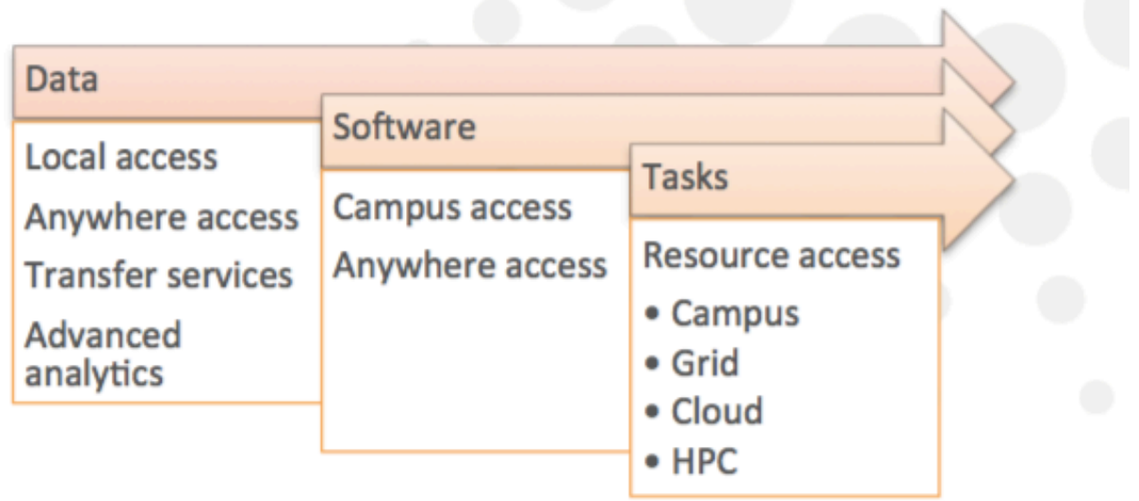
OSG Connect

Open Science Grid

✦ Campus Grids as a Platform of Services



- Define a suite of service offerings for campuses
- Connect to resources with increasing capabilities



Elements of OSG Connect

Open Science Grid

- Offer services to researchers and campuses w/ increasing capabilities
  - following Campus Grid Maturity Model developed by OSG Assessment
    - see http://twiki.grid.iu.edu/bin/view/CampusGrids/DeployedCampusInfrastructures

CGMM Level	Characteristics of organizations at this level
1	No organized or coordinated campus grid effort. Pockets of research computing, typically funded by individual researchers. Little or no support or documentation.
2	Some localized organization around campus grids. Some resource sharing at the departmental or college level. Minimal support and documentation.
3	Campus wide organization and/or broad visibility to campus grids. Good examples of resource sharing exist and there is some ability to utilize resources outside of the campus via partner campuses or the open science grid. Some documentation and local personnel support for campus grid users.
4	Campus wide organization or visibility of campus grids initiatives. Widespread sharing of on and off campus resources. At least part time dedicated personnel support and some documentation for campus grid users.
5	Campus grids are a 'way of life' for campus researchers, with on and off campus resources sharing the default. Mature user-facing documentation and dedicated personnel support for campus grid users.

INCREASING CAPABILITIES
None
BOSCO campus usage
+ OSG Connect : Jobs
+ OSG Connect : Software & Data access
+ OSG Connect : Integrated Accounting & Informatics

Elements of OSG Connect

- Graduated Platform of Services
  - ★ Campus Engagement & Identity Integration tools
  - ★ Job management
    - BOSCO and its extensions + pure HTCondor
  - ★ Distributed software access
    - (OASIS, PARROT, PALMS)
  - ★ Distributed data access
    - (SRM, XRD, HTTP, SKELETONKEY)
  - ★ Accounting and Informatics services for cycle sharing
    - (GRATIA, CIVAIS)

- Campus Infrastructures Community
  - ★ Forum, meetings, context to drive adoption, gather feedback, register impact
  - ★ Tutorials, demonstrators, campus blueprints, engagements

Production Area: Networking

- Networking a sub-area
  - ★ plans for moving network monitoring to “standard operations”
  - ★ what will we do forward-looking, making networking
  - ★ request by WLCG to extend OSG network monitoring WLCG-wide
    - evaluating the impact
- Software releases and testing
  - ★ procedures for releases, needs for testing, ITB etc

Technology Area

- Merge b/w Software and Technology areas, releases in Production
  - team responded with personnel plan, division of work, full buy-in
  - release team led by new release manager Tim Theisen
    - shared between OSG, HTCondor
  - successful re-staffing after several key people left
  - Software packages 225 source rpms, up from 198 a year ago
    - Upstream (incl. OSG): ~4M lines of code
- Transition for specific technologies in progress
  - CE, software provisioning, etc
  - deliver towards the goals of Production and User Support
- Re-confirm importance of a well-organized blueprint effort
  - at least 4 meetings/year, capture decisions and architectures

UW–Madison	Tim Cartwright	1.0	0.8 Software, 0.2 Education
	Mat Selmeci	1.0	
	Carl Edquist	1.0	
	Brian Lin	0.5	+ 0.5 Release
U. of Chicago	Suchandra Thapa	0.5	+ 0.5 Release
UCSD	Igor Sfiligoi	???	
	TBD	1.0	
UNL	*TBD*	???	part of 1.3 new FTEs?
BNL	John Hover	???	0.5 split between Tech and SW?
	Jose Caballero	0.5	
	Alexandr Zaytsev	0.4	

Backup Slides

80M hours!

Open Science Grid

Start of a punch list toward this goal

- 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

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