Open Science Grid

Open Science Grid All-Hands Meeting

March 13, 2013

Lothar Bauerdick/Fermilab

The OSG Ecosystem

Mission: *The Open Science Grid aims to promote discovery and collaboration in data-intensive research by providing a computing facility and services that integrate distributed, reliable and shared resources to support computation at all scales.*

✦ OSG Project
  - ★ staff, deliverables, operations

✦ OSG Consortium
  - ★ sites/resources providers, science communities, stakeholders

✦ Satellite Projects
  - ★ extensions, loosely coupled

✦ OSG Eco system
  - ★ enables interplay between domain needs and CS

OSG Project

- **OSG will enter 2nd year of our 5 year extension!**
  - ★ (have no indications (yet) about funding impact for this year)

Program Office	Funds/Year
NSF OCI	$1,000k
NSF MPS	$2,750k
DOE OHEP	$1,600k
DOE NP	$50k
Total	**$5,400k**

  - ★ OSG project will be reviewed after 3 years before final 2 years.

- **Satellite Projects are essential**
  - ★ research and technology to sustain OSG eco system for this decade
  - ★ example: Dvdt (DOE ASCR), AAA (NSF), XSIM (DOE ASCR), ...

OSG Executive Team

✦ "Fine Tuning" in management team

  ★ Executive Team

    ✦ Lothar Bauerdick as Executive Director

    ✦ Miron Livny (PI) as Technical Director

    ✦ Ruth Pordes as Council Chair

    ✦ Dan Fraser as Production Coordinator

    ✦ Brian Bockelman as Technology Coordinator

    ✦ Chander Sehgal as Project Manager (also User Support)

    ✦ Frank Würthwein as Resources Manager/Applications Coordinators

    ✦ Michael Ernst as Applications Coordinators

  ★ Area Coordinators

    ✦ Technologies: Brian Bockelman --> Software: Tim Cartwright, Blueprint: John Hover

    ✦ Production: Dan Fraser ---> Operations: Rob Quick, Campus: Dan Fraser

      ✦ adding Production Manager and Software Releases

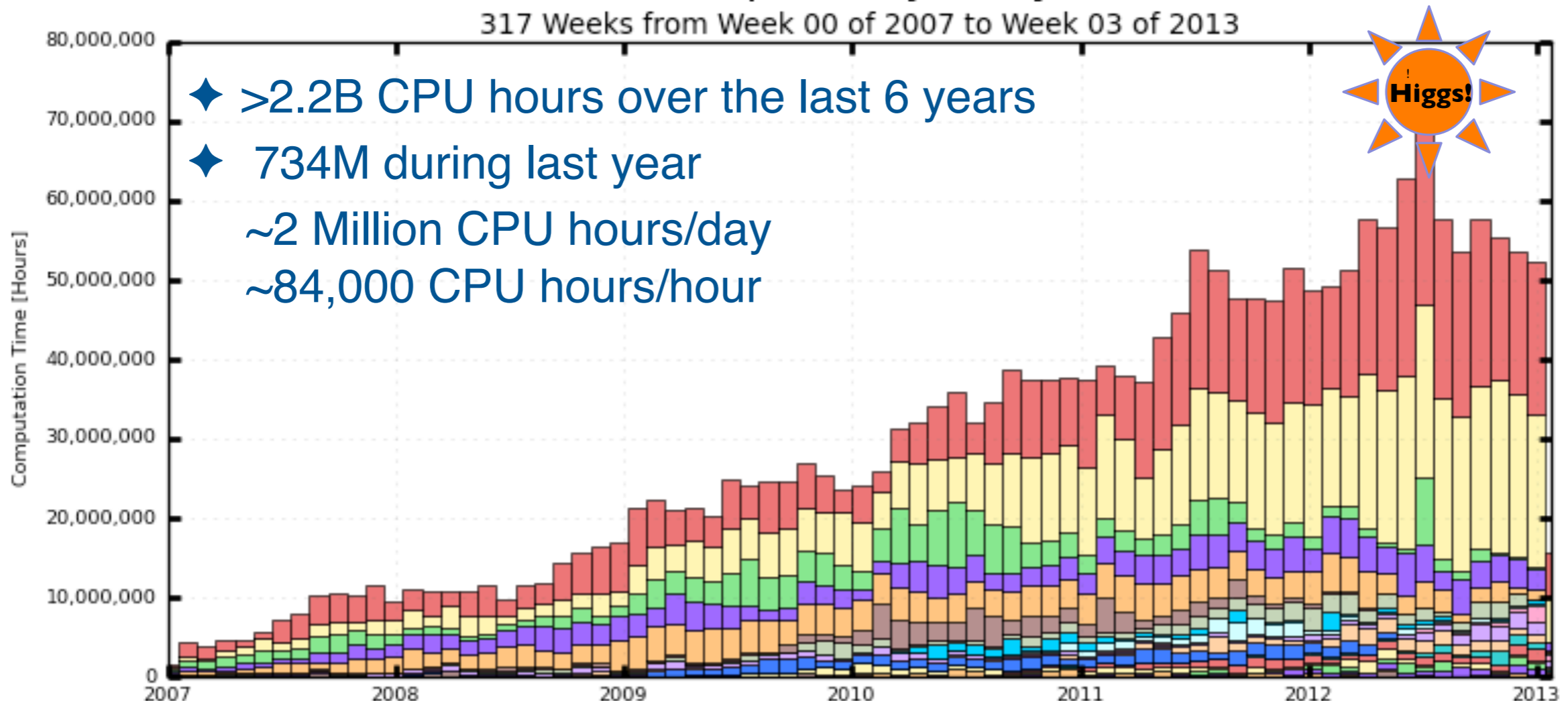   ✦ User Support: Chander Sehgal

    ✦ Security: Mine Altunay

Providing Computational Resources to Science

✦ A factor of >~5 increase in usage over past 6 years

  ★ Besides huge momentum and success of the LHC and its Tier-2s, a broad spectrum of other science applications are using the OSG

  ★ "Opportunistic users" from a large spectrum of science applications

    ✦ mostly driven by non-physics! This is how OSG "provides resources" —> OSG Focus!!

Hours Spent on Jobs By VO
317 Weeks from Week 00 of 2007 to Week 03 of 2013

✦ >2.2B CPU hours over the last 6 years

✦ 734M during last year

  ~2 Million CPU hours/day

  ~84,000 CPU hours/hour

Higgs!

Impact beyond HEP

(117 citations according to Google Scholar)

BMC Research Notes

Journal of Parallel Computing

Journal of Grid Computing Acta. Cryst. Biological Crystallography

Journal of Synchrotron Radiation

Monthly Weather Review

Instrumentation and Methods for Astrophysics PNAS

Future Generation Computer Systems

Journal of Climate

Journal of Supercomputing

Journal of Internet Services

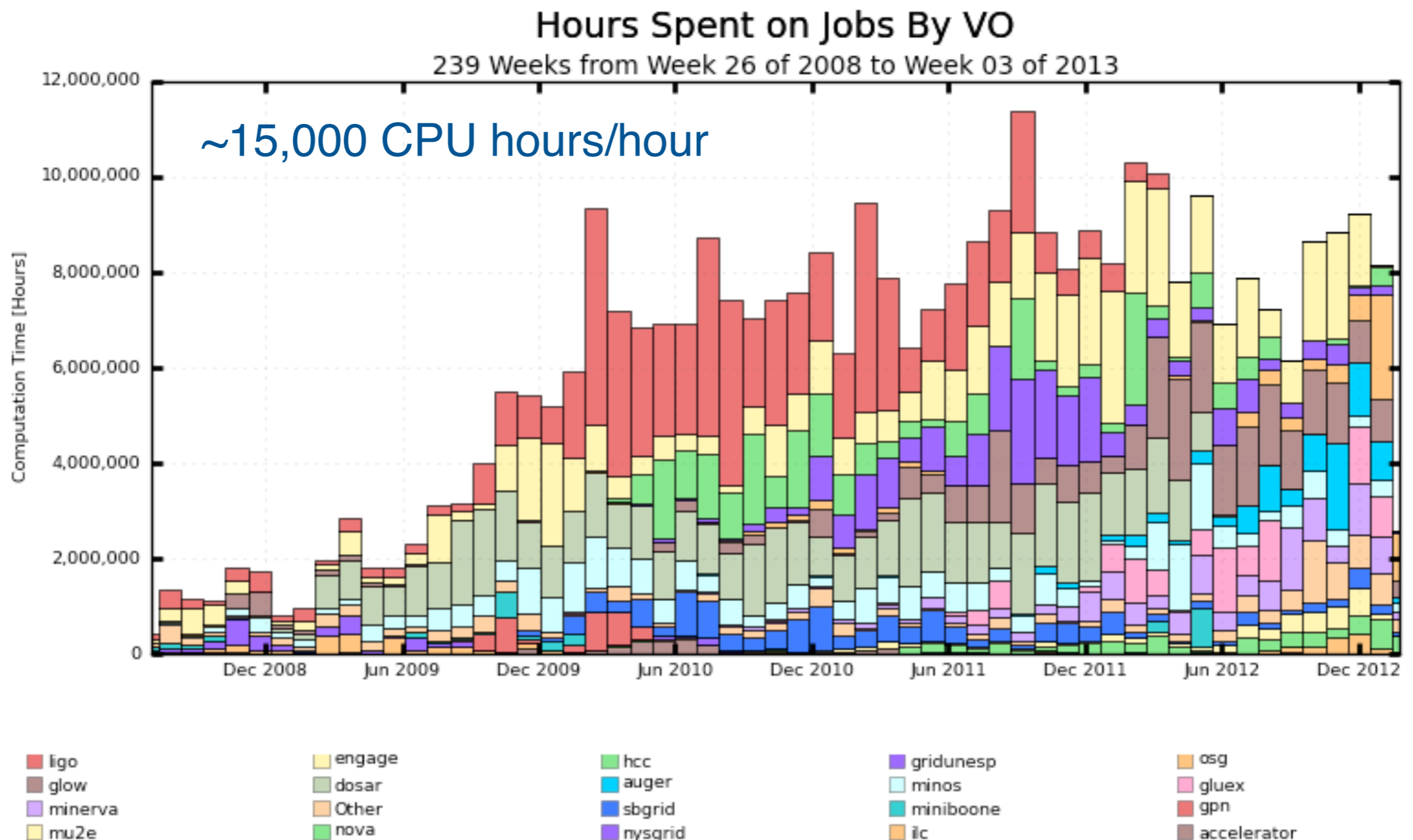International Journal of Psychophysiology Frontiers in Bioscience

Proteins: Structure, Function, Bioinformatics

An Infrastructure much beyond LHC

Open Science Grid

- equally impressive rise in OSG as general national cyber infrastructure
- 360 Million hours for programs that are **not** HEP "energy frontier"

Hours Spent on Jobs By VO
239 Weeks from Week 26 of 2008 to Week 03 of 2013

~15,000 CPU hours/hour

Legend:
- ligo
- glow
- minerva
- mu2e
- engage
- dosar
- Other
- nova
- hcc
- auger
- sbgrid
- nysgrid
- gridunesp
- minos
- miniboone
- ilc
- osg
- gluex
- gpn
- accelerator

Distributed High-Throughput Computing

Open Science Grid

- ✦ Job overlay infrastructure very successful: "it's a batch system"
    - ✦ >50% of OSG cycles through glideinWMS, plus ~30% through Atlas PanDA WMS
  - ★ gave tremendous boost to OSG usability, usage and effectiveness
- ✦ Job overlays provide avenue for flexible provisioning of resources
  - ★ succeeded to include resources (HPC, EC2) through submitting pilots
  - ★ grassroots research/science communities can use opportunistic cycles
  - ★ opportunistic resources could in future extend to cloud resources
- ✦ Need to improve usability, lower barriers of adoption, improve ops

- ✦ Current OSG for support data – allocation, transport, access, archive, curation, provenance – deficient compared to support for processing
  - ★ lack of storage provision potentially big obstacle to scientific productivity
    - ✦ mitigated by providing universal access to data across the wide-area network?
    - ✦ great experience with AAA/FAX, high-throughput access to PByte from 10s of sites
    - ✦ generalizable Remote Data Access services for other OSG communities?

Extending OSG Across the Campuses

✦ Enable access to and sharing of campus-local computing resources
- ★ BOSCO technology starting to get good use —> see Tue sessions!
- ★ more sites joining to make resources available for all OSG users (ND!)

✦ Basic technologies in place today to support use of cloud resources
- ★ cloud computing — or other resources like HPC centers etc —can now be added to the OSG job execution environment

✦ Individual researcher can register for support through the "Campus Researcher Club" at UWMadison, Nebraska, UC San Diego
- ★ another "way in" for PIs and projects is coming in through XSEDE

✦ OSG as certificate services to the broader U.S. science community
- ★ picking up from the ESnet service which is running out now —> Von's talk
- ★ Future: make campus identities an integral part of OSG identities
  - ✦ streamline the process for obtaining, renewing, and using identities in OSG

✦ A number of challenges are to be addressed, like usability and robustness of our software within the OSG eco-system

Other Important Areas

✦ OSG continues to be part of an International System

  ★ Continue to interact and work with Europeans, WLCG, EGI, (EMI)

    ✦ thanks to Maria Girone, David Collados and Peter Solagna for coming!

  ★ cultivate our connections to Latin America and Asia
through sites, research community and work force

✦ OSG added a "Networking" area to "Production" (Shawn McKee)

  ★ networks are the lifelines for many of the OSG applications

  ★ perfSonar-based network monitoring system between sites, across OSG

    ✦ provides unprecedented insights into connectivity and needs of science community

    ✦ wealth of information to be harvested/correlated —> research; improve operations

  ★ satellite Dvdt research on software defined networks, ANSE, Dynes etc

✦ Partnership between OSG and XSEDE (Chander, Miron)

  ★ get new science communities to use OSG users through XSEDE

  ★ starting to enable access to XSEDE resources to OSG

✦ Campus Infrastructure Communities (Rob Gardner)

  ★ great series of meetings, discussions, seminars, building the community

This Morning's Agenda:

Open Science Grid

- ✦ Status, Strategies and Future Plans
  - ★ PI and Technical Director — Miron Livny
  - ★ OSG Council Chair — Ruth Pordes
  - ★ Production — Dan Fraser
  - ★ Campus Infrastructures Community — Rob Gardner
  - ★ User Support — Chander Sehgal
  - ★ Technologies: Software — Tim Cartwright
  - ★ OSG PKI Transition — Von Welch
- ✦ "Future of the Grid - Lightening Talks"
  - ★ organized by Ken Bloom and Brian Bockelman