ALICE USA Computing Project

Status & Plans

ALICE T1/T2 Workshop

Tsukuba, Japan

2014

Outline

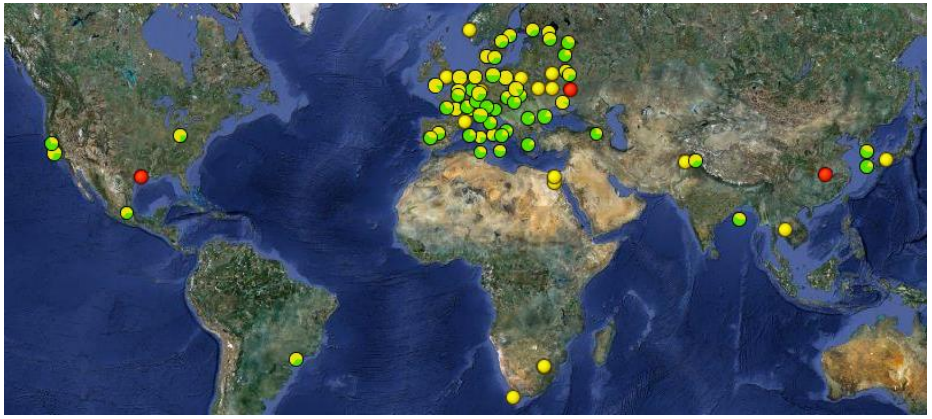
- ALICE-US Computing Project Overview
  - Goals & Operations
  - Facility snapshots
  - Resource utilization & performance
    - History of project
    - Current year
  - Summary of project evolution
- Big Change
  - LLNL Status & beyond
- Timeline for new project proposal & external review
- Other US Activities

ALICE-USA Computing Project

- Goal: supply cost-effective Grid-enabled computing resources to ALICE
  - Fulfill MoU-based ALICE USA obligations for computing & storage resources to ALICE
  - Based on ALICE USA participation at about 7-8% of ALICE

- 2009 Project Proposal
  - Operate facilities at two DOE labs
    - NERSC/PDSF at LBNL
    - Livermore Computing (LC) at LLNL
    - 3-year procurement plan
  - LBNL as the host lab

  ➢ Fully operational since Summer 2010



- Project personnel on steering/operations committee
  - Jeff Porter – project manager & ALICE Grid Manager for NERSC
  - Ron Soltz - Former Computing Coordinator & LLNL ALICE Rep.
  - Jeff Cunningham – LLNL System Admin and ALICE Grid Manager for LC-glcc
  - Iwona Sakrejda – PDSF project lead
  - Bjorn Nilsen – ALICE-USA contributor from Creighton U. has left the field
    - I want to publically thank Bjorn for his good work

Evaluating ALICE-USA Obligations

- ALICE annual computing requirements:
  - # events, event size, real/MC processing times & samples, data duplication …
- Requirements vetted by WLCG
  - Final requirements in WLCG DB
    - 6 months before they take effect
- ALICE-USA Obligations:
  - Fraction of total requirements defined by proportion of ALICE-USA/ALICE

Year	FY12	FY13
ALICE Requirements		
CPU (kHEPSPEC06)	336	290
Disk (PB)	22.0	30.3
ALICE-USA Participation		
ALICE Total Ph.D. (Total-CERN)	538	528
ALICE-USA Ph.D.	40	43
ALICE-USA/ALICE (%)	7.4	8.1
ALICE-USA Contributions		
CPU (kHEPSPEC06)	24.9	23.2
Disk (PB)	1.65	2.4

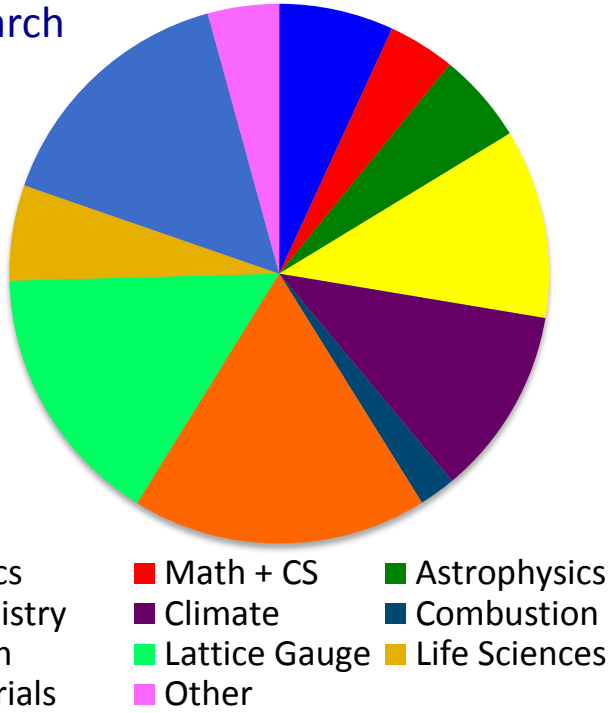
Table 1: Computing requirements from ALICE and planned ALICE-USA contribution.

Facility Snapshot: LLNL/LC

- Livermore Computing
  - Large & diverse institutional-based High Performance Computing Center
  - Supports Lab Science and Engineering activities
  - Lab interest in developing external collaborations
- Cost effective procurement and operations model
  - Able to buy into routine very very large purchases of scalable units
  - In-house managed OS (CHAOS) & other software (e.g. SLURM)
- ALICE Deployment model @ LLNL/LC
  - Separate single-use Grid facility
    - 100% ALICE
    - Grid only use → no user logins
  - Large HW purchase, refreshed every 4 years

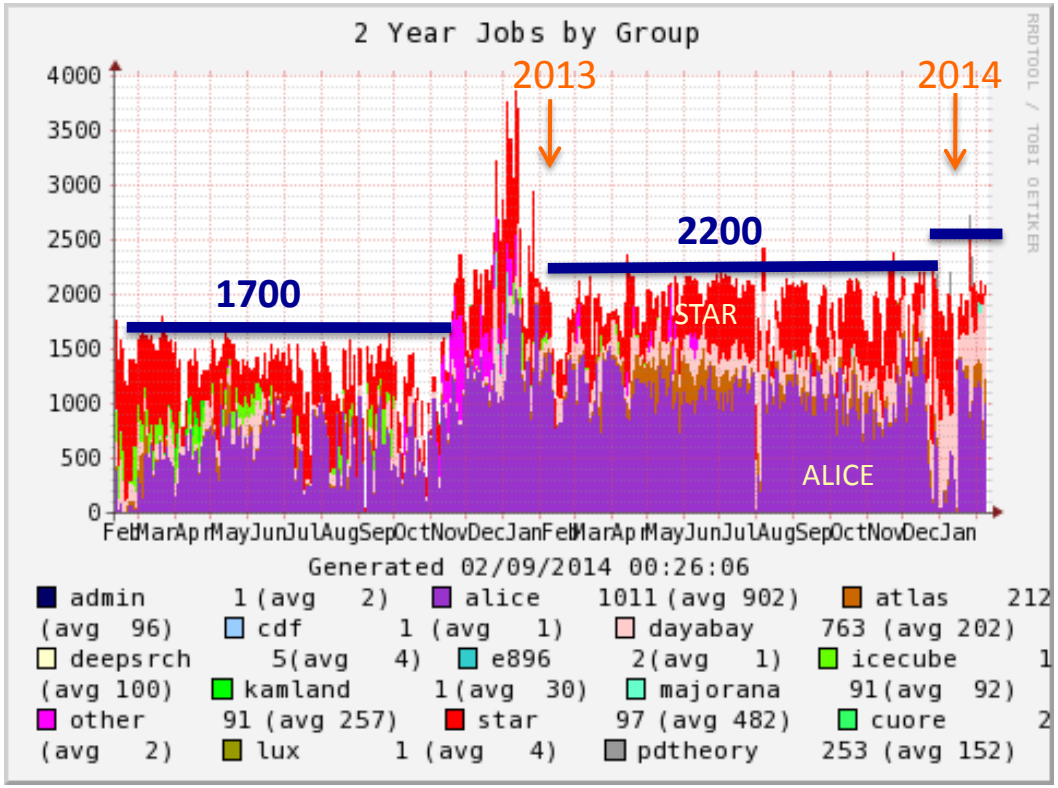
Facility Snapshot: NERSC

- NERSC: US Department of Energy (DOE) Office of Science Flagship High Performance Scientific Computing Center
  - Available to all DOE Office of Science sponsored research
- Computing for Scientific Research
  - Large HPC Systems (100s k cores)
  - Special Clusters: PDSF, Visualization,…
  - Large archival storage (HPSS)
  - Data Transfer, Gateway & OSG/Grid Services
  - Evaluation Systems: GPU & Cloud Services
- Extensive user support services
- ALICE Deployment Model @ NERSC
  - Project resources deployed on PDSF for ALICE Grid (see next slide)
  - Users can have login access with ALICE client tools available
  - Annual HW purchases to adjust to changing ALICE requirements



Facility Snapshot: PDSF

- Multi-group facility for Nuclear & High Energy Physics experiments
  - Allocations as "share" of resources
  - Fair share done in SGE (UGE)
- Share calculation includes
  - HW investment
  - FTE contribution
- Nuclear Science shares
  - ALICE 40%
  - STAR 30%
- Physics Div. shares
  - ATLAS T3 15%
  - Dayabay 10%

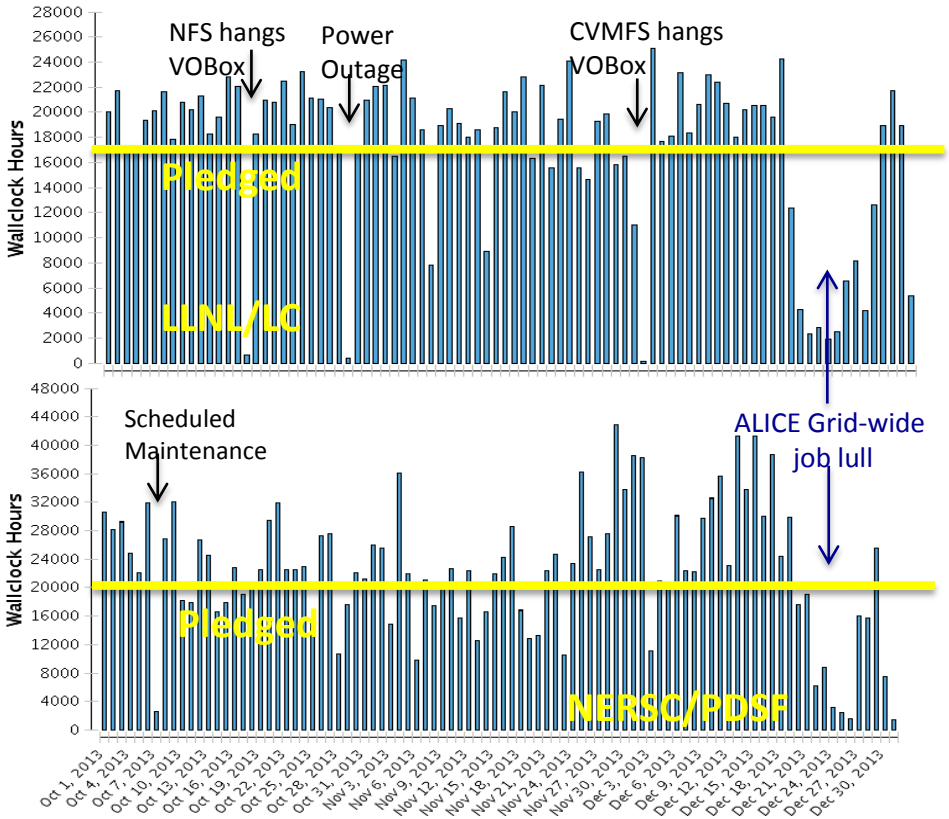


Running jobs

ALICE-USA & Open Science Grid

- ALICE-USA project leverages OSG capabilities
  - OSG Registration Authority
    - ALICE-USA user certificates
    - PDSF & LLNL machine certificates
  - Resource reports sent to WLCG
    - ~~Availability and Reliability~~
      - ~~Critical services scans~~
    - Accounting Reports
      - Gratia site service
      - OSG central repository → WLCG

OSG Reports



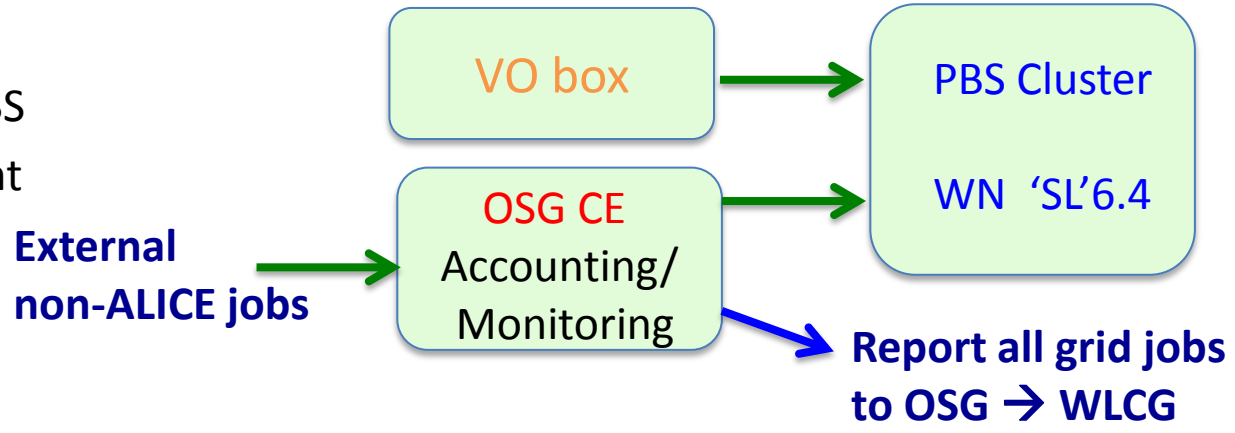
Q1FY14 Report to DOE Oct 1 – Dec 31, 2013

➢ Funding agency monitors that we are not oversized for our mission

Site Configurations

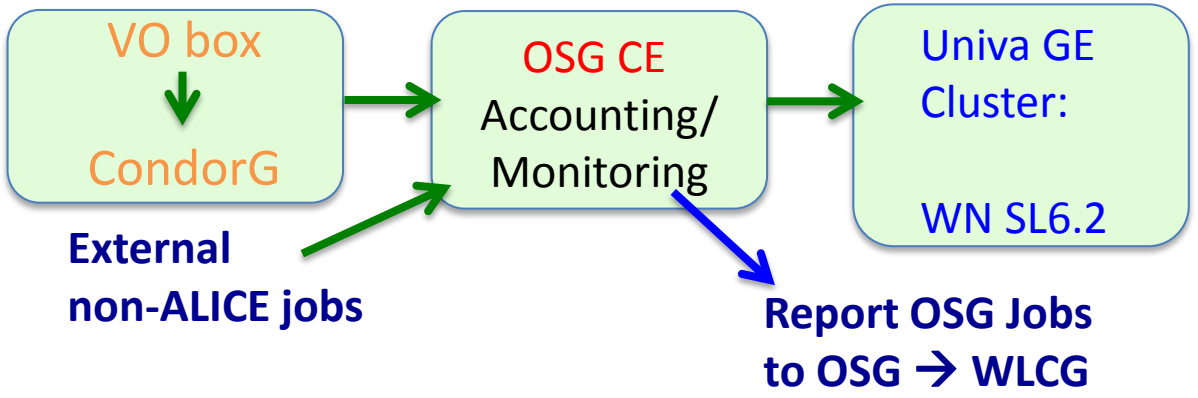
LLNL/LC

- Vobox Submits to PBS
- OSG-CE: independent external interface
- OSG Accounting



LBNL/NERSC/PDSF

- Submits to CondorG service on VO box
- CondorG submits to OSG-CE at PDSF
- OSG Accounting



➢ NERSC evaluating SLURM

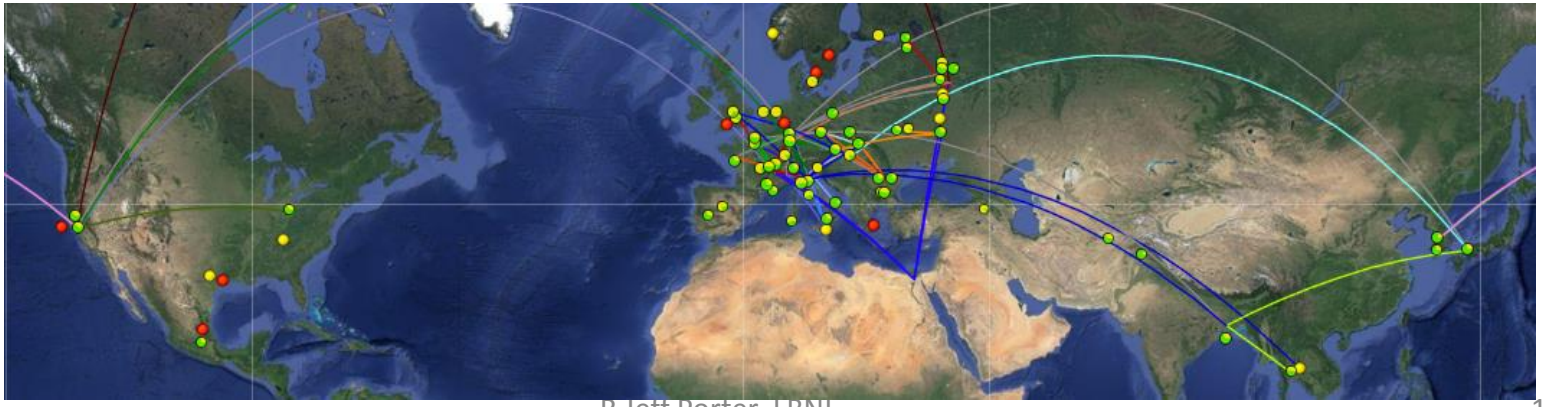
- Target date ~ Sept 2014

Project History: Storage Capacity

- Pledged Obligations
  - LLNL/LC
    - 650 TB since 2010
  - NERSC/PDSF
    - Steady ramp plan: 300TB → 740 TB → 1,020 TB →1,200 TB → ….
- Installed Capacity
  - LLNL/LC = 685 TB since Aug ‘10
  - PDSF = 720 TB since Oct ’11

				% used
NERSC/PDSF	716.1 TB	365.2 TB	351 TB	50.99%
LLNL/LC	687.8 TB	327.9 TB	359.9 TB	47.68%

Q1FY14 Report to DOE Dec 31, 2013



Source for low storage utilization

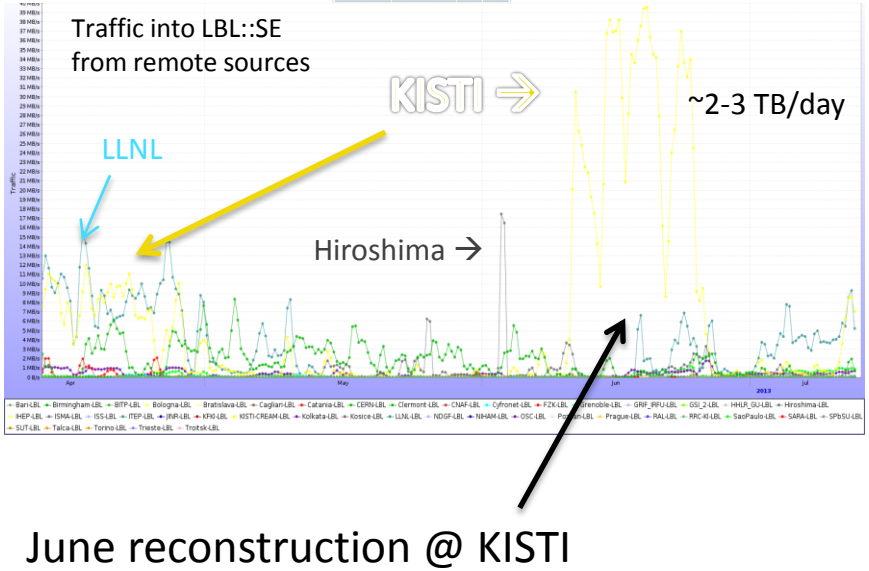
- High availability in AliEn SE tests
  - LLNL::SE → 95%
  - LBL::SE → 98%

SE tests history

Series	Last value	Min	Avg	Max
1. LBL - SE	100	0	98.19	100
2. LLNL - SE	100	0	95.04	100
Total	100		96.62	

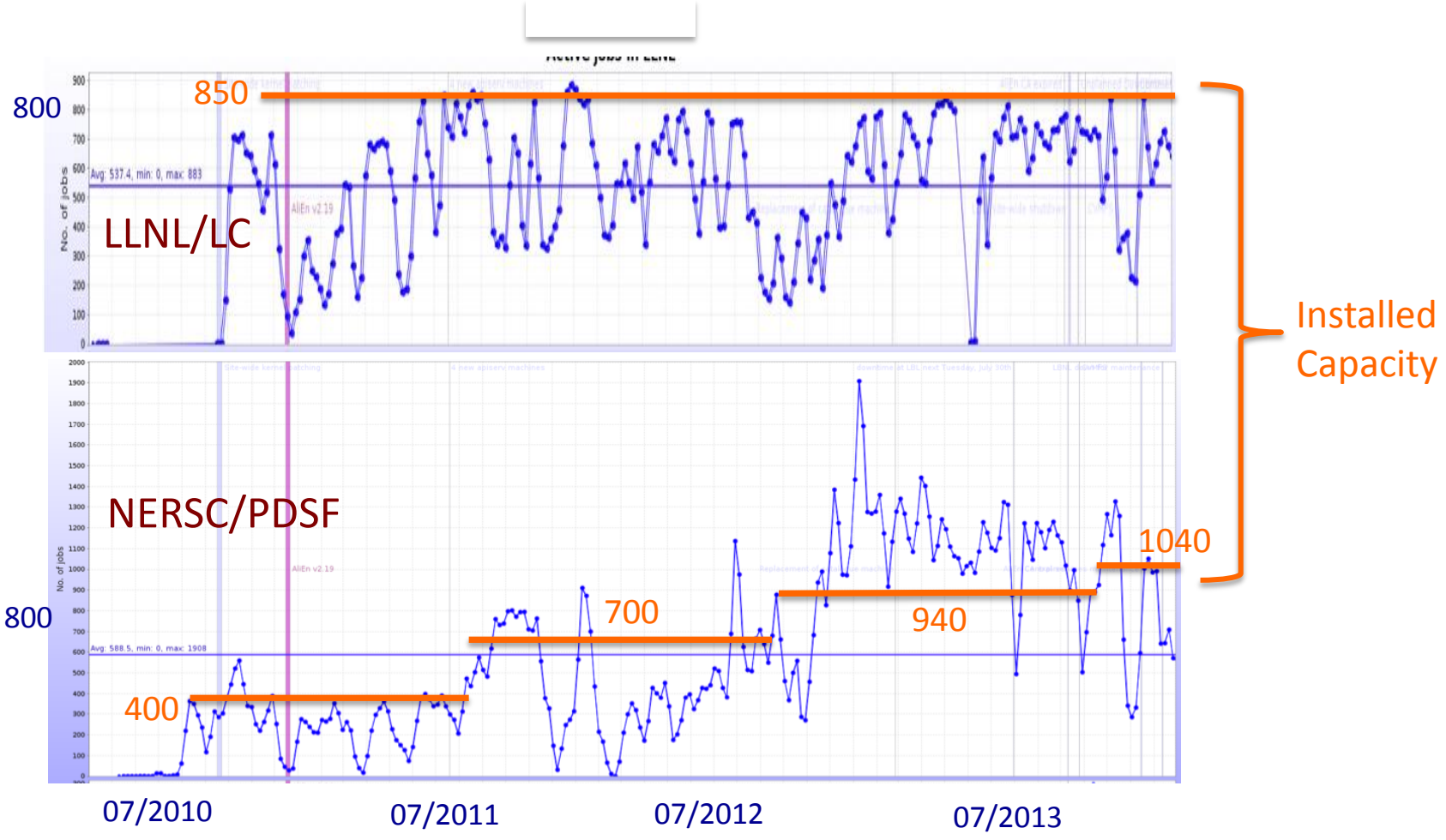
- "Nearby" SE effect
  - Small nominal rate → LBL::SE
    - Typically ~5MBps
    - LLNL is largest writer, ~5-10 MBps
  - Larger rates
    - during reco @ KISTI
    - Periodically from Hiroshima

30 MB/s →



June reconstruction @ KISTI

Project History: Job & core count



Job & core count: 2013 - today

LLNL upgrade:

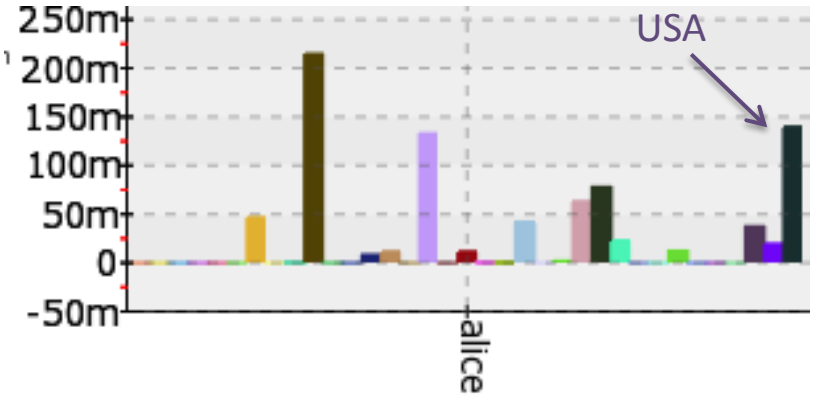
CPU Utilization 2013 RRB Year 10 Months: 3/13 – 01/14

- CPU Utilization eff: CPU/wall
  - LLNL/LC & NERSC/PDSF ~ 65-70% efficiency
    - 70% Allowed by WLCG Accounting for T2

WLCG Worldwide LHC Computing Grid — **Tier-2 Accounting Summary**

Efficiency factor for Tier-2 sites - utilisation 70% of pledge as specified in TDR

EGI Portal Country T2



- Utilization relative to pledges
  - LLNL/LC
    - Pledge : 11,500 HS x 24 x 300 x 0.7 (allowed eff.) = 58.0 MHS-hrs
    - Delivered: 51.6 MHS-hrs → **89%**
  - NERSC/PDSF
    - Pledge: 12,900 HS x 24 x 300 x 0.7 = 65.0 MHS-hrs
    - Delivered: = 79.9 MHS-hrs →**123%**
  - Combined
    - Pledge = 123.0 MHS-hrs , Delivered = 131.5 MHS-hrs →**107%**

Network upgrade plans

- LHCONE
  - We requested implementation @ NERSC in July 2013
  - Reply → separate WAN from local network via dual-homed data servers
- Dual homed data servers
  - Model is worked out
    - All normal access via WAN interface
    - Router instructions on local cluster to access local interface
  - Implementation happening now
  - Auxiliary benefit: independent measures of local vs WAN traffic
    - Allows us to request network infrastructure best suited for needs
      - e.g. 100Gb to ESnet achievable with modest cost targeted solutions
- IPv6
  - Ready at NERSC border
  - LLNL/LC →TBD

Timeline of project to date

- Project Funded in Jan. 2010
  - Both Tier 2 sites fully operational by Sept 2010
- WLCG MoU
  - LBNL signed, 4/2011
  - LLNL signed LoI, 2012
- External Project Review, 2012
  - Executive Summary: "As of today both participating sites have demonstrated their outstanding ability to reliably contribute to ALICE's managed production and user analysis activities at excellent performance."
- Project plan updates based on new ALICE requirements
  - FY13 update, July 2012
  - FY14 update, July 2013 ←included description for refresh all LLNL HW in 2014
- External project review was scheduled for Feb 20-21, 2014.

Big Change to the Project

- Group at LLNL has decided to shift efforts from ALICE-USA
  - ALICE Tier-2 site @ LLNL to be decommissioned
    - Target dates → Oct 2014 (perhaps Oct 2015)
- ALICE is losing an important partner:
  - Extremely cost-effective procurements
  - Experience built up over the past 4+ years.

- ALICE-USA Spokesperson asked for evaluation & recommendation
  - What is the level of computing resources (CPU, storage, bandwidth, expertise, etc.) that will be lost when LLNL ceases to be part of ALICE-USA?
  - What are the options for replacing this capacity?
  - Describe the potential sites and assess the pro and cons of each site, including issues like local expertise, cost effectiveness, migration paths to the O^2 time period, access, time required to be fully operational, etc.
  - Describe the pros and cons of 1, 2, or 3 site solutions.
  - If possible, provide a prioritization of the most realistic options.

Potential Sites Available

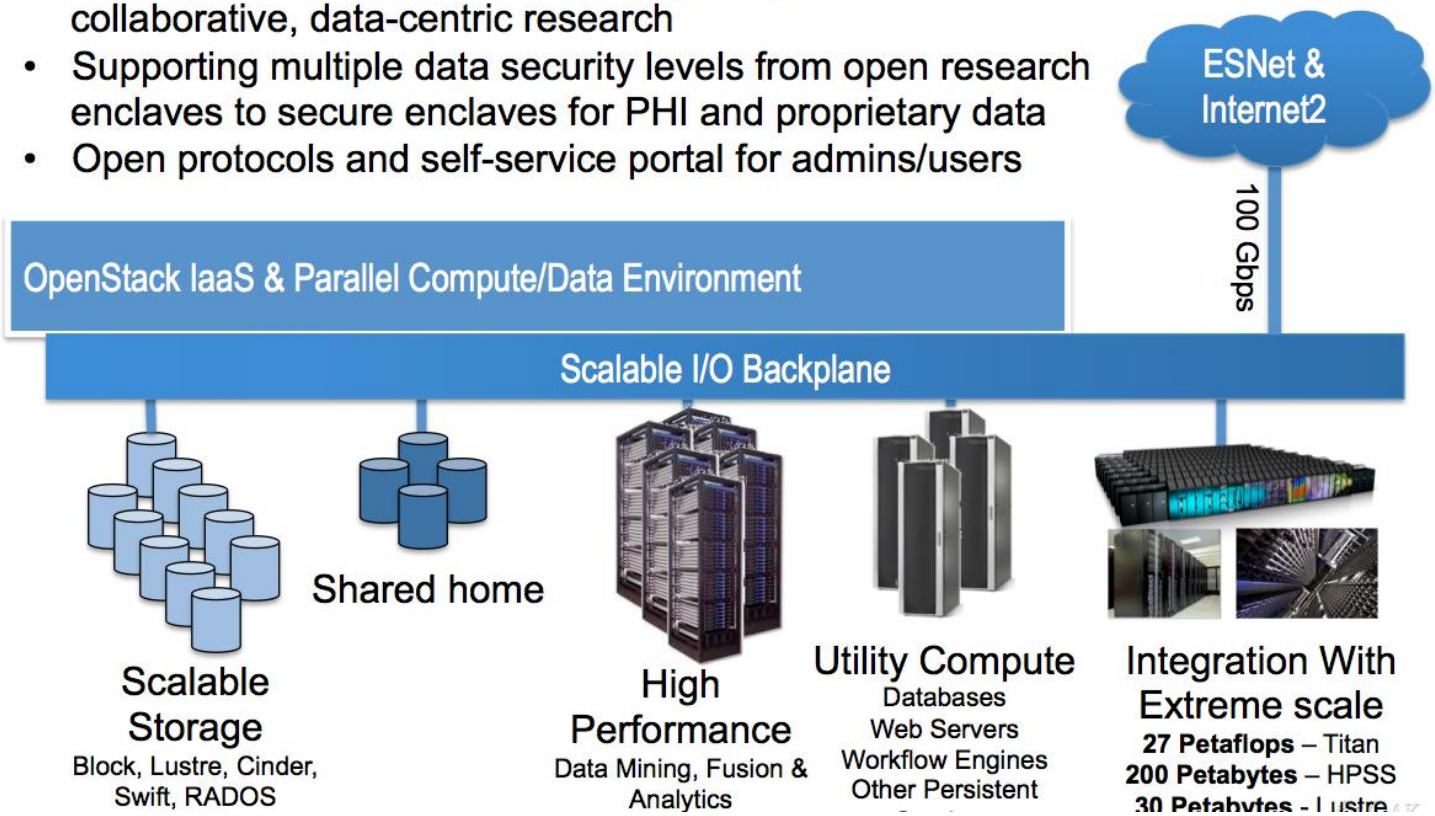
- U. Houston: High Performance Computing Center
  - One of the original ALICE Tier 2 sites & part of first project proposal
- ORNL: Compute And Data Environment For Science (CADES)
  - New initiative for cluster computing, targeting needs of a broad range of scientific disciplines with emphasis on data-intensive science
- Both sites meet the basic criteria:
  - Strong presence within ALICE-USA
  - Good network connection to rest of US and world
  - Resources to be embedded within larger facilities

ORNL CADES

Compute and Data Environment for Science targets full range of needs

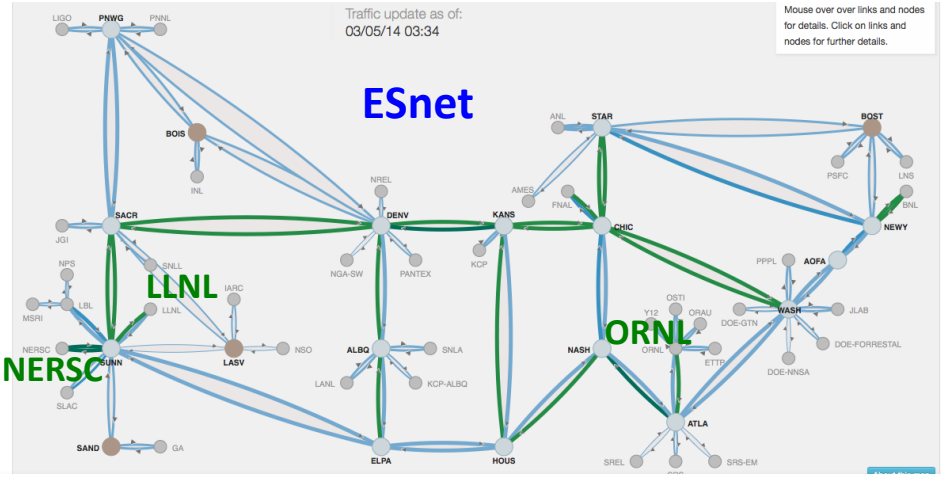
Single PI Projects → Centers → User Facilities → Multi-Institutional Projects

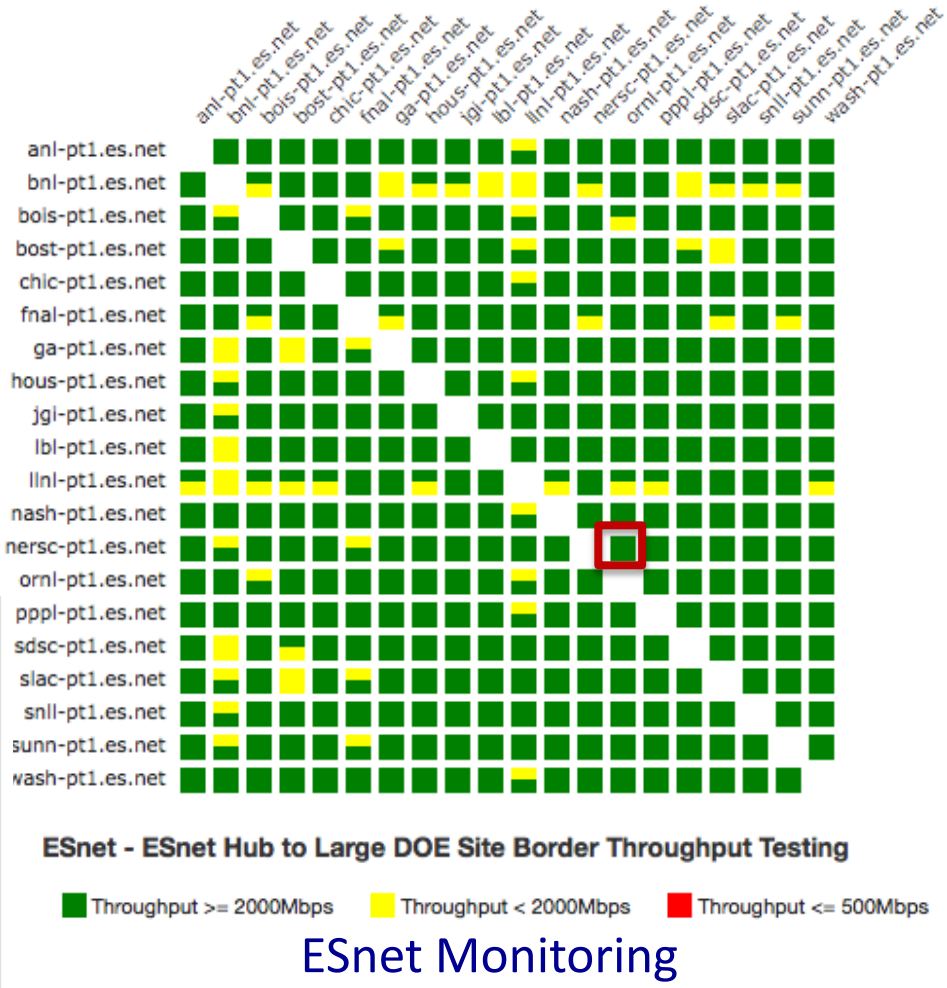
- A rich software stack enabling highly integrated, collaborative, data-centric research
- Supporting multiple data security levels from open research enclaves to secure enclaves for PHI and proprietary data
- Open protocols and self-service portal for admins/users



Recommendation for Two Facilities NERSC/PDSF and ORNL/CADES

- NERSC/PDSF + ORNL/CADES
  - Scientific Computing strength
  - High-bandwidth connection: ESNet
  - Favorable cost structure
  - Proximity to HPC Resources
    - Oak Ridge Leadership Class Facility
    - NERSC Flagship facility
    - Strategic alignment with O^2 project

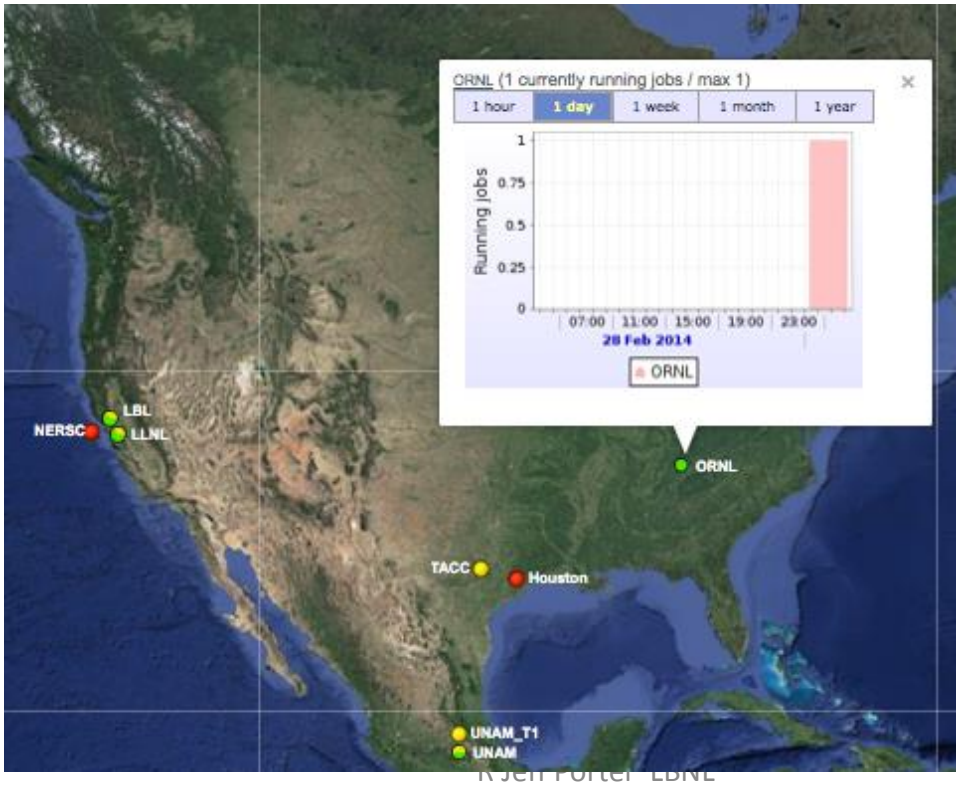




ESnet Monitoring

Initial Proof of Principle

- Set up a demo ALICE grid site at CADES
  - VO box with AliEn services, local and CVMFS installs
  - Demonstrate network capability with AliEn monitoring
  - Demonstrate basic job processing



Excellent network connection to US & worldwide sites

Next Steps

- Develop a new proposal: Project Execution and Acquisition Plan
  - Schedule new resources for NERSC/PDSF and ORNL CADES to:
    - Adequately meet new ALICE requirements in FY15
    - Can be adjusted to fill ALICE-USA obligations estimated through Run 2
  - Establishes operational milestones
    - Stable grid operations @ORNL CADES
      - both AliEn and OSG
    - WLCG MoU with ORNL CADES

Year	FY14	FY15	FY16	FY17
ALICE Requirements				
CPU (kHS06)	300	320	400	480
Disk (PB)	22.9	37.5	45.4	50.7
ALICE-USA Participation				
ALICE Total-CERN Ph.D.	555	555	555	555
ALICE-USA Ph.D.	42	42	42	42
ALICE-USA/ALICE (%)	7.6	7.6	7.6	7.6
ALICE-USA Obligations				
CPU (kHS06)	22.8	24.3	30.4	36.5
Disk (PB)	1.7	2.8	3.4	3.8

ALICE Run-2 Requirement Estimates

- Aggressive timeline:
  - March: draft working document
  - April 7th-8th: host ALICE Offline at LBNL to review proposal and plan
  - ➢ External Review of project proposal is being scheduled ~ June 2014

Busy year ahead ….

- NERSC new building in early 2015
  - No new hardware will be installed in the current building
  - Operate in both buildings for 1 year
    - Most of PDSF CPU nodes will move
      - Duration of move/downtime is unclear
      - Will attempt to mitigate with other cpu
    - Storage will remain in old building
    - New storage →new building

- We may catch a break
  - LLNL/LC may operate → Sept 2015
    - I repeat “may”

Guidance on storage replacement

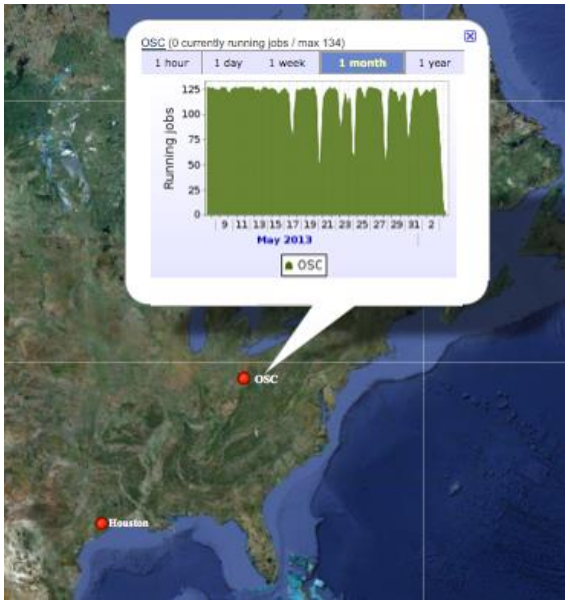
- Storage is or will be out of warranty:
  - LLNL/LC
    - 680TB ended ~Aug 2013
  - NERSC/PDSF:
    - 144TB ended June 2013
    - 288TB ends Apr 2014
    - 288TB ends Oct 2014
  - Current ~1.4PB will remain in production until ~summer 2015

- Replacement storage will be new SE at both LBL and ORNL
  - Requires migrating >600TB between new and old SEs
  - Likely: LLNL::SE → ORNL::SE, LBL::SE → LBLEOS::SE

- Requests guidance on deadline requirements:
  - Before 2015 Heavy ion run?
  - By March 2015?

Additional US Sites & Activities

- Ohio Super Computing Center (OSC)
  - Different funding agency (NSF)
  - Stable resource
    - Not pledged
    - University grant allocation
  - No compatible storage



- Texas Advanced Computing Center: TACC
  - Initiated by professor @ UTexas Austin: Christina Markert
  - Currently a startup allocation of ~50k cpu-hrs
  - Proposal to be written & submitted, ~month
    - Potentially large resource to supplement pledges

Summary

- ALICE-USA Computing project
  - Two US based ALICE Tier-2 facilities, LLNL/LC & LBNL/NERSC
  - Currently satisfies all US pledged resources
    - ALICE used >100% of pledged CPU resources
    - 80% of pledged storage resources installed, due to lower utilization

- The project will undergo a shift leading up to (during) Run 2
  - Two US based ALICE Tier-2 facilities, ORNL/CADES & LBNL/NERSC
  - Replace LLNL/LC resources with new resources at ORNL/CADES
  - Rebuild expertise lost at LLNL

➢ Challenge over the next 12-18 months
  ➢ Essential rebuild of our facilities with limited loss of service