

ALICE USA Computing Status Report

R. Jeff Porter

ALICE Tier-1/Tier-2 Workshop

April 18, 2016

Outline of Project Status Report



- **Project and Facilities**
- **Current Operations**
- **2016 Plans**

ALICE-USA Computing Project

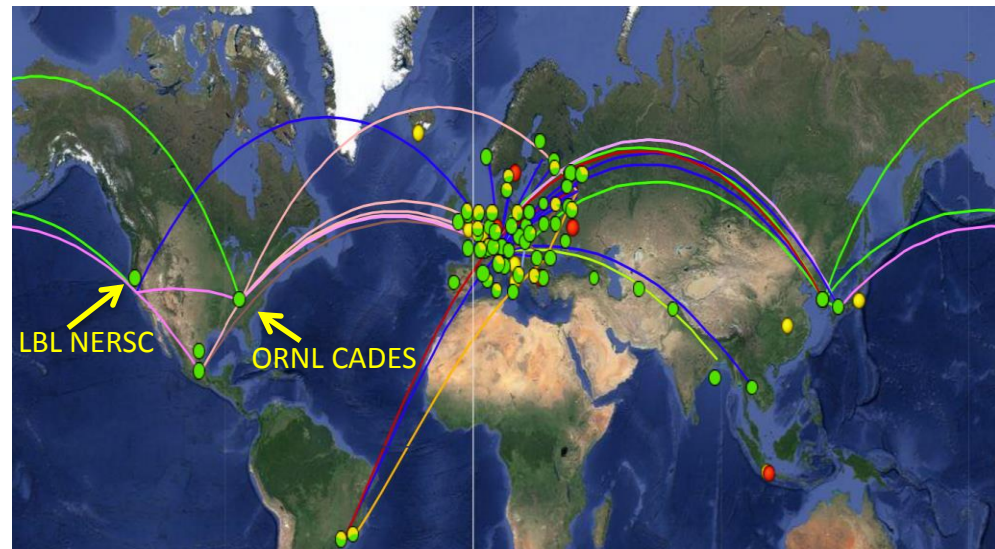
- **Goal: supply grid-enabled computing resources to ALICE**
 - Fulfill MoU-based ALICE USA obligations for CPU & storage resources to ALICE

- **Original 2009 Project Proposal**
 - Operate facilities at 2 DOE labs
 - NERSC @ LBNL
 - Livermore Computing @ LLNL
 - LBNL as the host institution

- **In operational since 2010**

- **Project working documents:**
 - Project SLA: Institutions & roles
 - Project Execution Acquisition Plan:
 - **PEAP**

- **Project Plan for 2015**
 - Replace LLNL/LC with ORNL/CADES
 - Operational in Summer of 2015
 - Move of NERSC T2 to new building



} Interesting year

R Jeff Porter LBNL

Project Organization

Computing Steering Committee



- **Project Steering Committee:**

- Currently: **J.Porter**, K.Read, **P.Eby**, **M.Galloway**, & J.Botts
- Local wiki: [http://rnc.lbl.gov/Alice/wiki/index.php/ALICE-US Computing](http://rnc.lbl.gov/Alice/wiki/index.php/ALICE-US_Computing)
 - Document repository
 - Monthly Meetings & minutes
- Email list

- **Connection to ALICE Grid Operations**

- Alice-grid-task-force email list
- Annual US meeting with CERN team since 2012
- Annual ALICE T1/T2 workshops
 - 2012 @ KIT Germany: I. Sakrejda & J. Cunningham
 - 2013 Lyon, Fr: J. Cunningham & J. Porter
 - 2014 Tsukuba, Jp: J. Cunningham & J. Porter
 - 2015 Torino, Italy: J. Porter & P. Eby
 - 2016 Bergen, Norway: J. Porter, P. Eby & M. Galloway
- Annual AliEn Developers Workshops
 - 2010 – 2012, J.Porter
 - 2013, J. Porter & B. Nilsen

ALICE-US Computing

Contents [hide]

- 1 ALICE-USA Computing
- 2 Documents
- 3 PDSF
- 4 LC

ALICE-USA Computing

- [Computer Steering Committee Meetings](#)

Documents

- [ComputingDocs](#) ← PEAPs, DOE proposals & Quarterly Reports
- [\[Email Archives\]](#)

PDSF

- PDSF home page is [here](#)

LC

- [Livermore Computing](#)
- [Green Data Oasis](#)

ALICE-USA Obligation Evaluation

- **ALICE-USA Obligations:**

- Fraction of ALICE Requirements and defined by proportion of ALICE-USA to ALICE
- ALICE Computing requirements established annually and approved by WLCG

Table 1. ALICE Computing requirements and corresponding ALICE-USA obligations.

Year	FY2015	FY2016	FY2017
ALICE Requirements			
CPU (kHS06)	320	394	474
Disk (PB)	38.1	44.7	52.5
ALICE-USA Participation			
ALICE Total-CERN Ph.D.	573	573	573
ALICE-USA Ph.D.	40	41	42
ALICE-USA/ALICE (%)	7.0	7.2	7.3
ALICE-USA Obligations			
CPU (kHS06)	23.0	28.4	34.5
Disk (PB)	2.7	3.2	3.8

←FY16 PEAP Update
Presented to DOE Nov. 2015

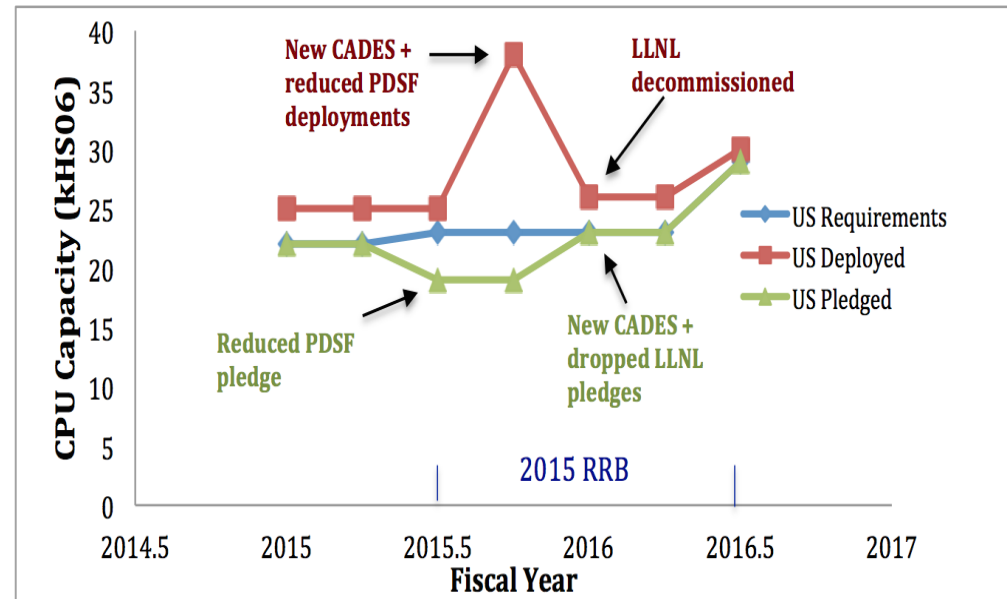
Overview of Transition Schedule & Hardware Deployment Strategy as Planned



- **LLNL T2 remain operational through Sept 2015**

- 13 kHS06 & 685TB Storage

CPU Delivery schedule



- **ORNL CADES online → May 2015**

- 18 kHS06 CPU cluster commissioned
- ~1.0 PB ORNL::SE using EOS service
 - LLNL::SE is set to read-only mode
 - Data copied from LLNL::SE to ORNL::SE

- **NERSC PDSF T2 operational @ OSF**

- >12 kHS06 & 720TB SE

- **New NERSC PDSF T2 brought online → June 2015**

- 8 kHS06 moved to new home & re-commissioned
- ~0.9 PB Storage commissioned as new LBNL::SE using EOS service
 - Old LBNL::SE is set to read-only mode
 - Data copied from old LBNL::SE to new LBNL::EOS

Overview of Transition & Hardware Deployment



- **Initial DOE Funding Released in Feb 2010**

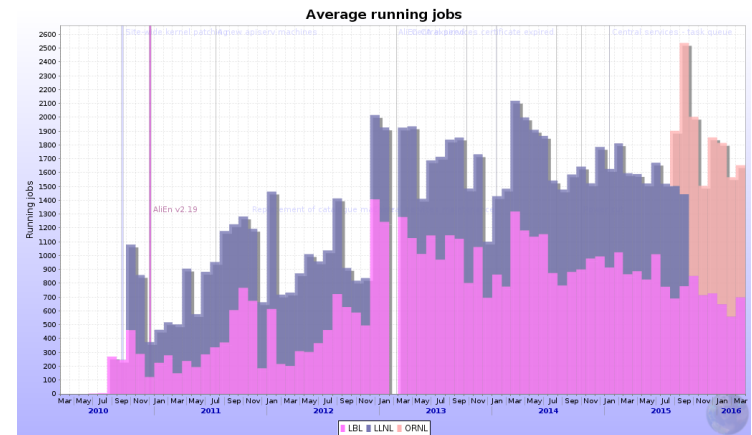
- NERSC PDSF deployment: June-July 2010
- LLNL LC deployment: May-Sept 2010
- Approx. 6-7 months after release of funds

- **DOE Funding for new project Feb 2015**

- ORNL Storage deployed in June 2015
 - LLNL data copied July-August 2015
- ORNL Fully operational August 2015
- Approx. 7 months after release of funds

- **2015 Project included move of NERSC**

- From Oakland Science Facility (OSF) to New CRT building



2010

2015



Overview of Transition Schedule

How did we do?



- **LLNL T2 Decommissioned Sept 2015** ✓

- **ORNL CADES online Summer 2015** ✓

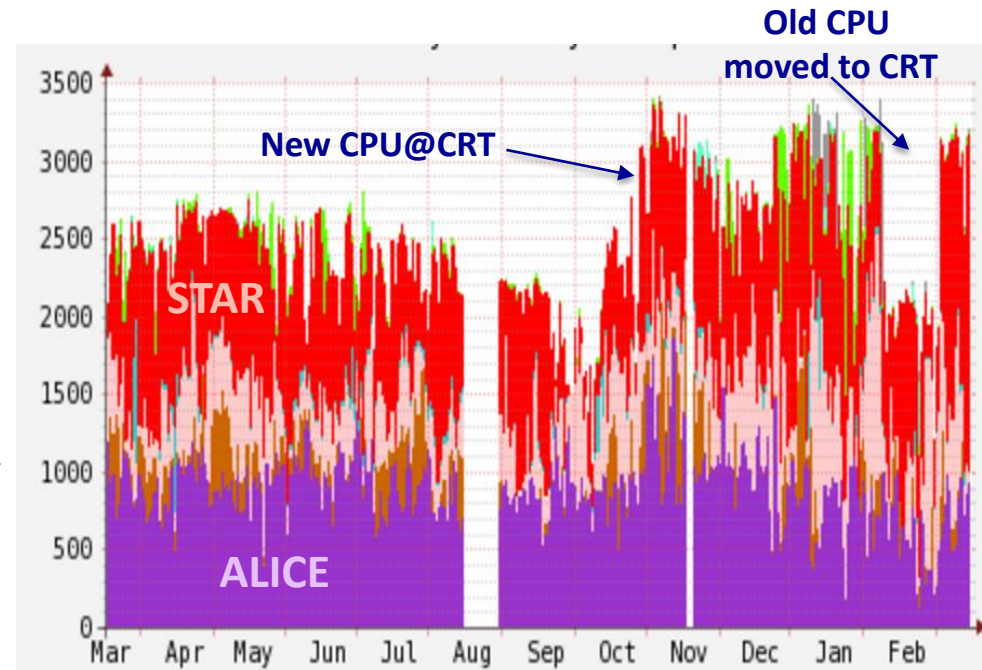
- Storage commissioned in June
- CPU commissioned in August
- Operational by end of August
 - 2 Month delay well within contingency

- **NERSC PDSF T2 operational @ OSF** ✓

- >12 kHS06 & 720TB SE

- **New NERSC PDSF T2 ... 2015 → 2016**

- PDSF move had to be delayed due to NERSC plans
- New CPU installed in November
- 8 kHS06 moved to new home & re-commissioned, March 2016
- 0.9 PB Storage commissioned as ALICE::LBNL::EOS, Apr 2016

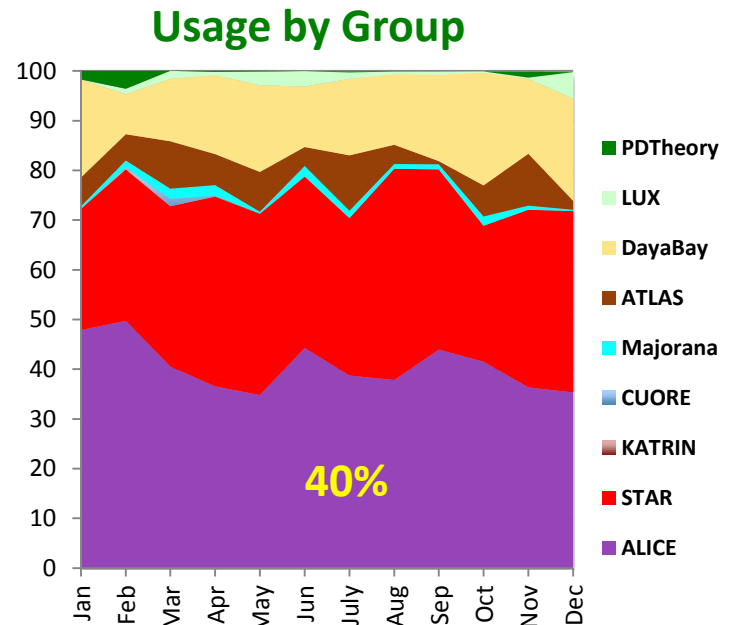
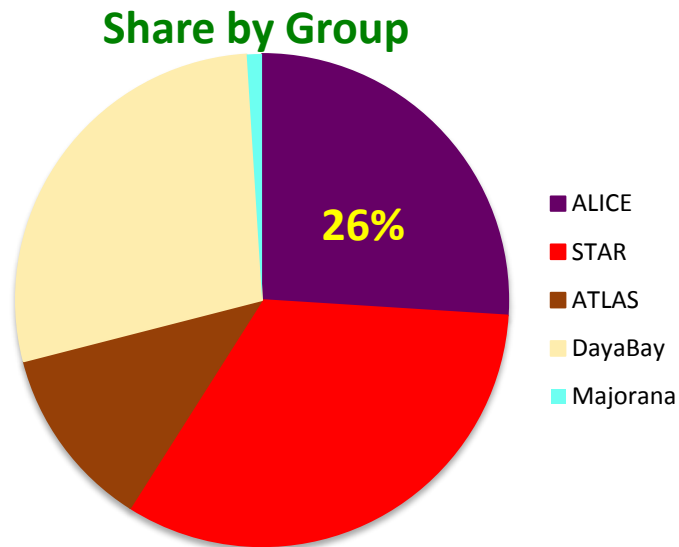


Jobs on PDSF previous year

LBNL T2 Site: PDSF @ NERSC



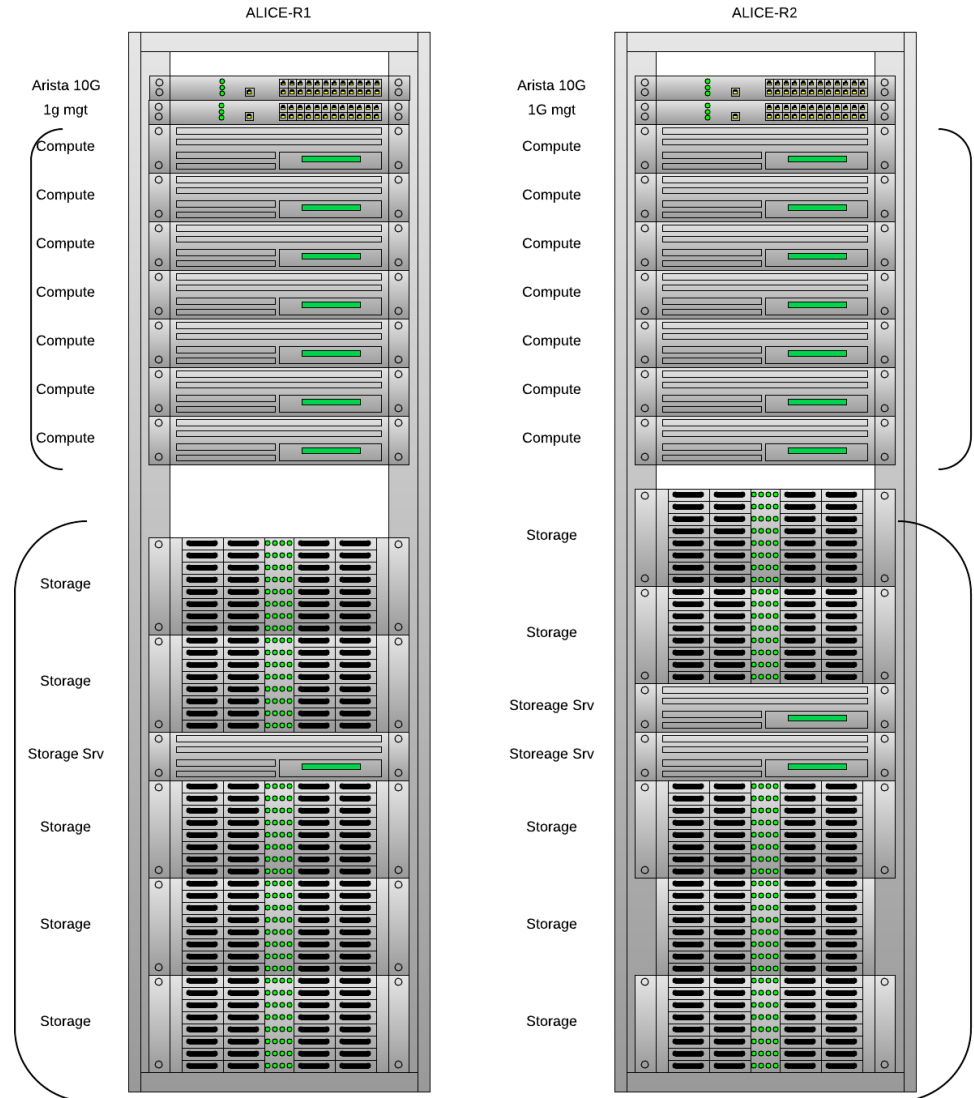
- **Evergreen cluster operated by NERSC for HEP/NP Experiments**
 - CPU added/retired annually leading to mixed cluster
 - 32 core Haswell (2015), 20 core Ivy Bridge (2014), 16 core Ivy Bridge (2013 & 2012)
 - Share based on investment (both shared HW and FTE support)
 - PDSF supports about 9 active groups and 800 active users



ORNL T2 Site: T2 @ CADES

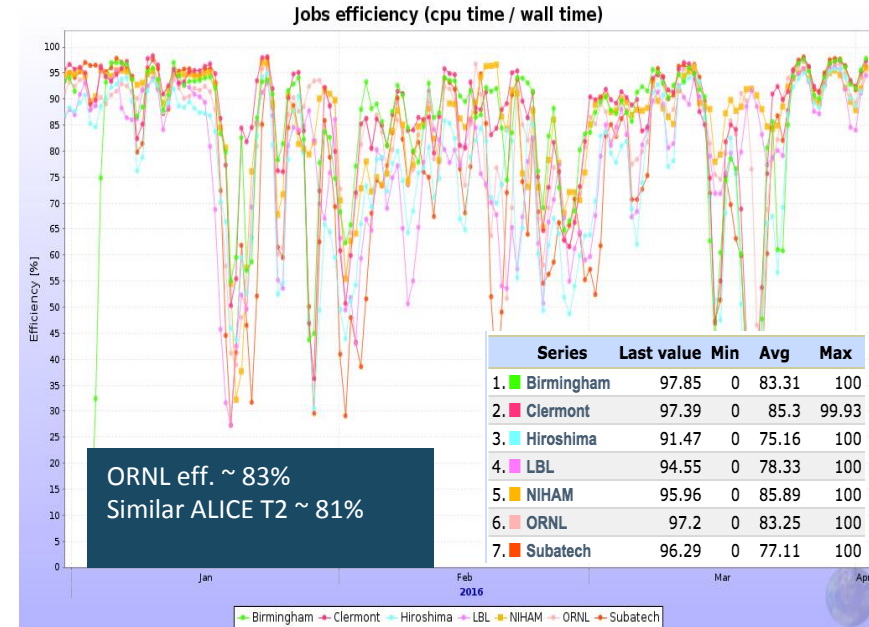


- **ALICE Grid facility within a larger center**
 - One User: ALICE
 - May leverage access to other CADES resources
- **Compute & Storage**
 - 56 compute nodes
 - 18.3 kHEPSPEC
 - 4 servers (EOS FSTs)
 - 1 EOS MGM
 - 5x60 Bay storage chassis
 - 1.0 PB usable storage



- **ORNL CPU deployment**

- HEPSPC/core less than expected
 - Lower clock speed gave better price
- ORNL investment in ALICE
 - additional CPU, 5 x 12 core/node 4GB/core
- Optimize node utilization
 - Flexible → ALICE only facility
 - HW → 16 cores/node, 4GB/core
 - Jobs → 22 jobs/node, (3GB + 7GB swap)/job
 - Monitor efficiency



- **EOS Deployment**

- See talk later this workshops

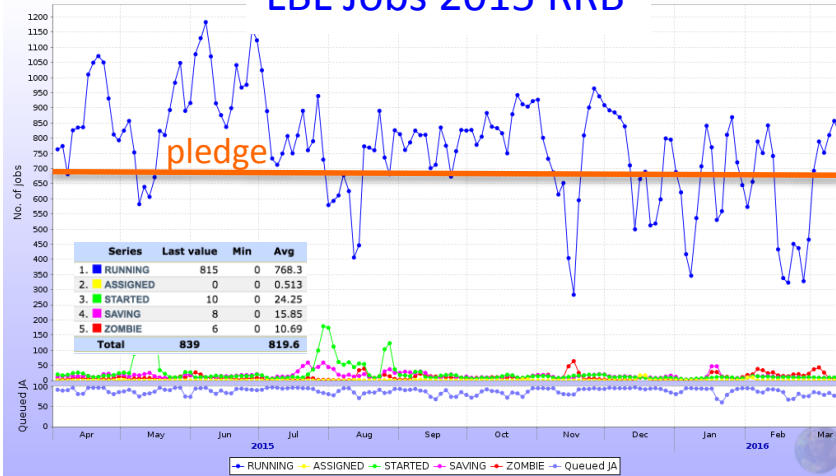
Section II



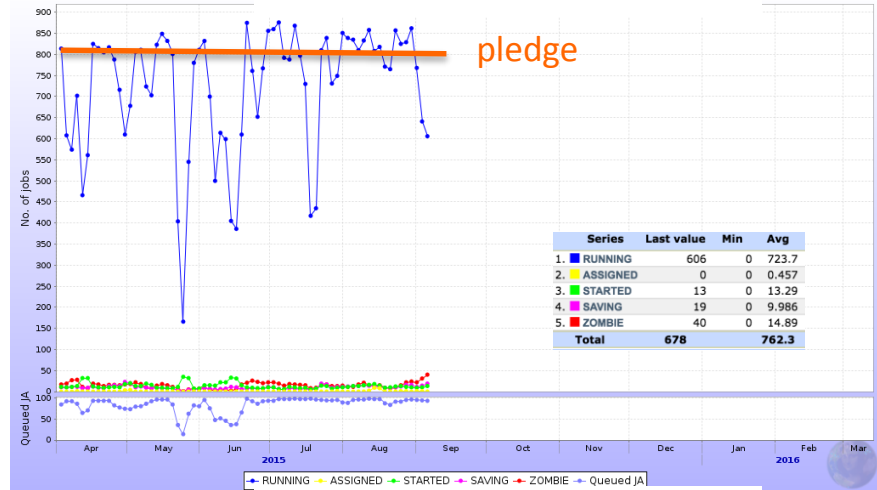
- **2015 Operations**

Site Job Profiles

LBL Jobs 2015 RRB



LLNL Jobs 2015 RRB



Ave. Runing Jobs:

LLNL = 723

LBL = 768

ORNL = 1004

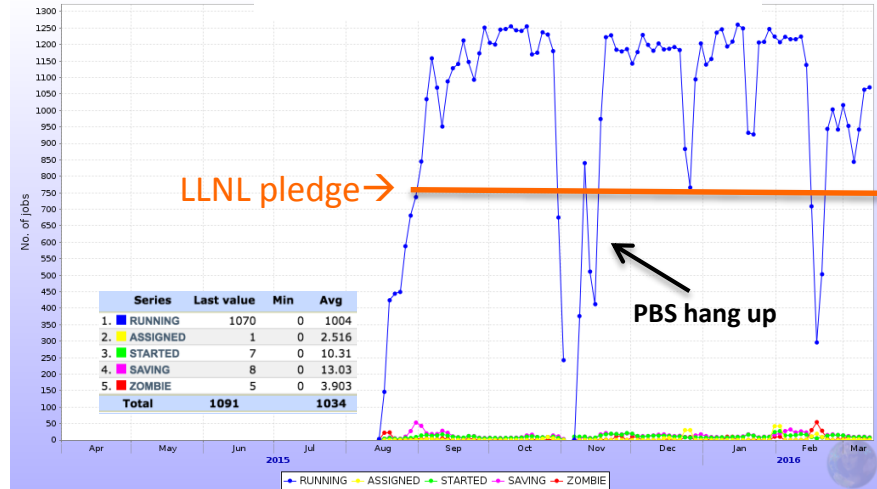
Zombie Grass remains well trimmed:

LLNL ~2.0%

LBL ~1.5%

ORNL ~0.4%

ORNL Jobs 2015 RRB



Site Efficiencies: cpu-time/wall-time

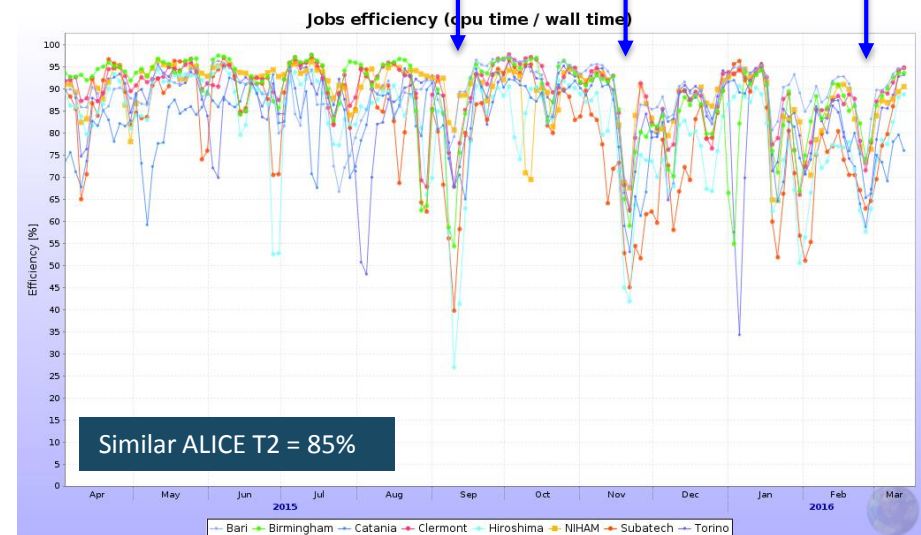
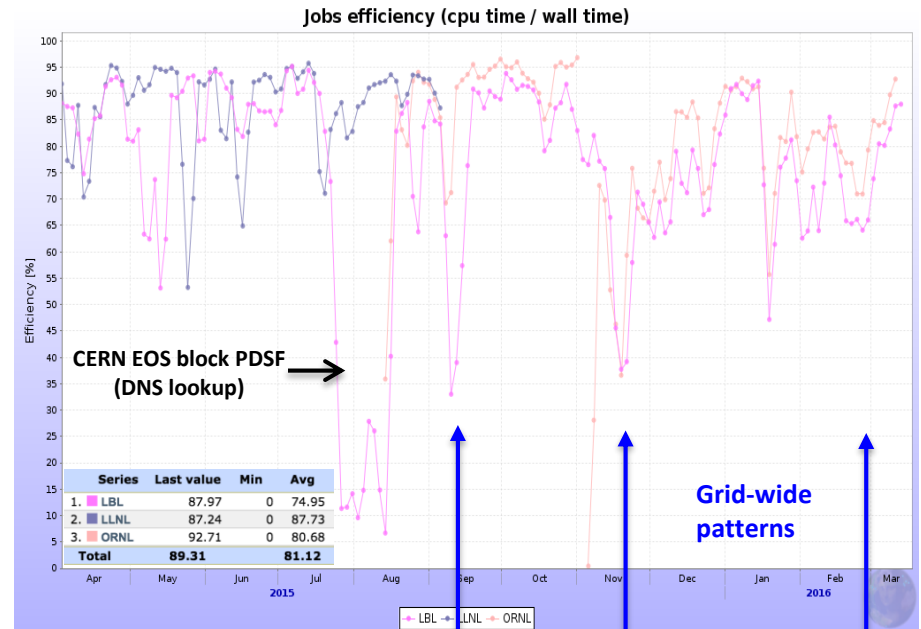
- Ave Site Efficiencies**

- LLNL 88%
- LBL 75%
- ORNL 81%

- Largely tracks other ALICE T2**

- Specific issues:**

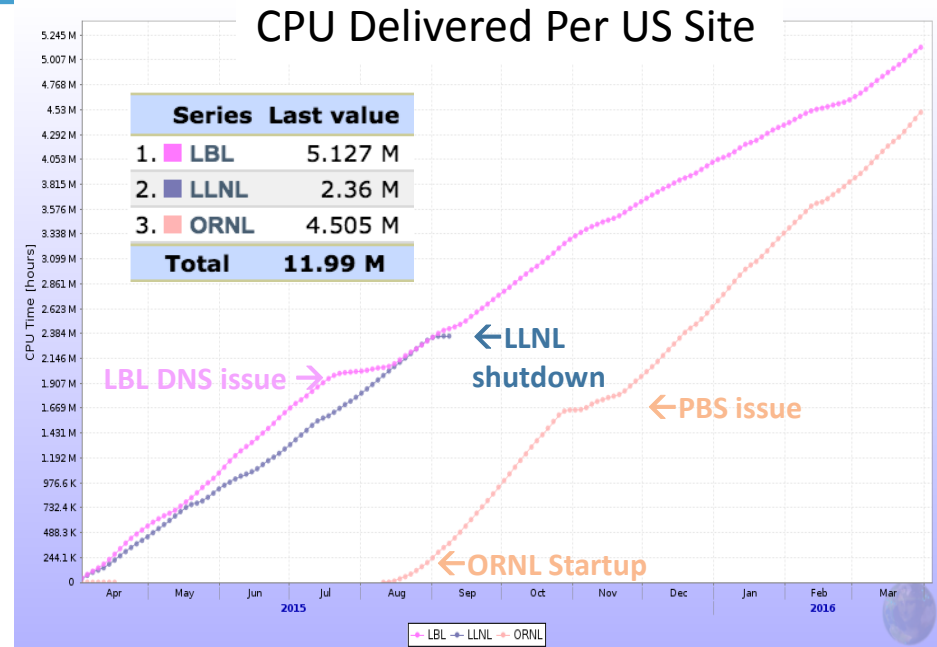
- PDSF blocked by CERN EOS
 - Worker nodes added to DNS
- ORNL Startup & PBS problems



CPU Delivered to ALICE Grid Relative to Pledged Obligations



- **2015 RRB Year**



US Tier 2 Site	Per-core CPU (HS06/core)	CPU delivered (MHS06-hrs)	US 'Pledges' * (MHS06-hrs)	Delivered/ * Obligation
LBL	16.6	85.2	70.0	122%
LLNL	13.5	31.9	5 months ~ 41.8	76%
ORNL	14.0	63.1	7 months ~ 58.6	108%
Totals or Ave		180.2	170.4	107%

* does not include WLCG % efficiency allowance

Storage Capacity & Utilization

- **Storage Deployment History**

- LLNL LC
 - 685 TB on Aug, 2010, Decommissioned Sept 7, 2015
- LBNL NERSC
 - Staged install : 720 TB by 2012, to be decommissioned
 - 900 TB EOS installed, April 2016
- ORNL CADES
 - 1000 TB installed as EOS storage, June 2015

ALICE-USA Storage Elements Capacities & Usage: 04/2016

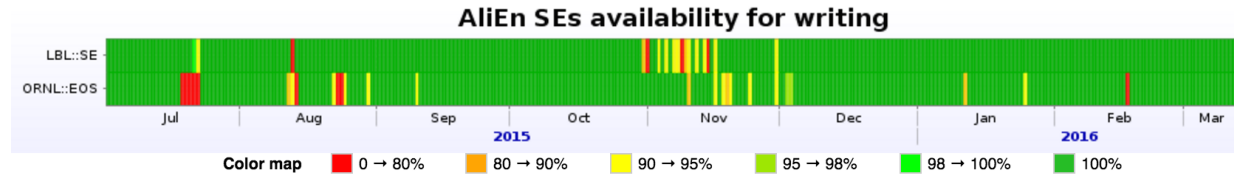
ALICE SE	#-servers	Space (TB)	Used Space (TB)	% Used
LBL::SE	10	720	620	86
LBL::EOS	3	900	9	1
ORNL::EOS	4	1000	520	52

SE Availability Continued



- **Writing**

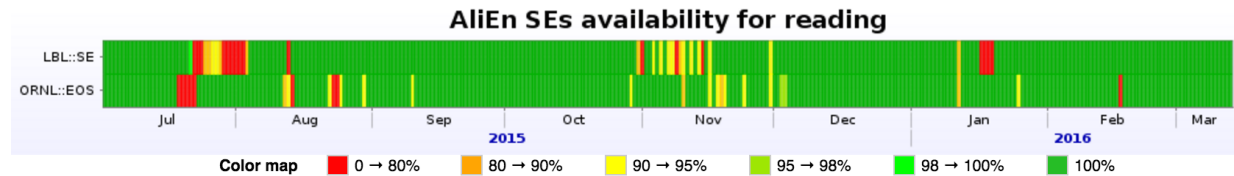
- LBL::SE → 99%
- ORNL::EOS → 98%



Statistics							
Link name	Data		Individual results of writing tests			Overall	
	Starts	Ends	Successful	Failed	Success ratio	Availability	
LBL::SE	01 Jul 2015 20:18	13 Mar 2016 20:18	3045	28	99.09%	99.10%	
ORNL::EOS	01 Jul 2015 20:22	13 Mar 2016 20:21	3010	63	97.95%	97.97%	

- **Reading**

- LBL::SE → 97%
- ORNL::EOS → 98%

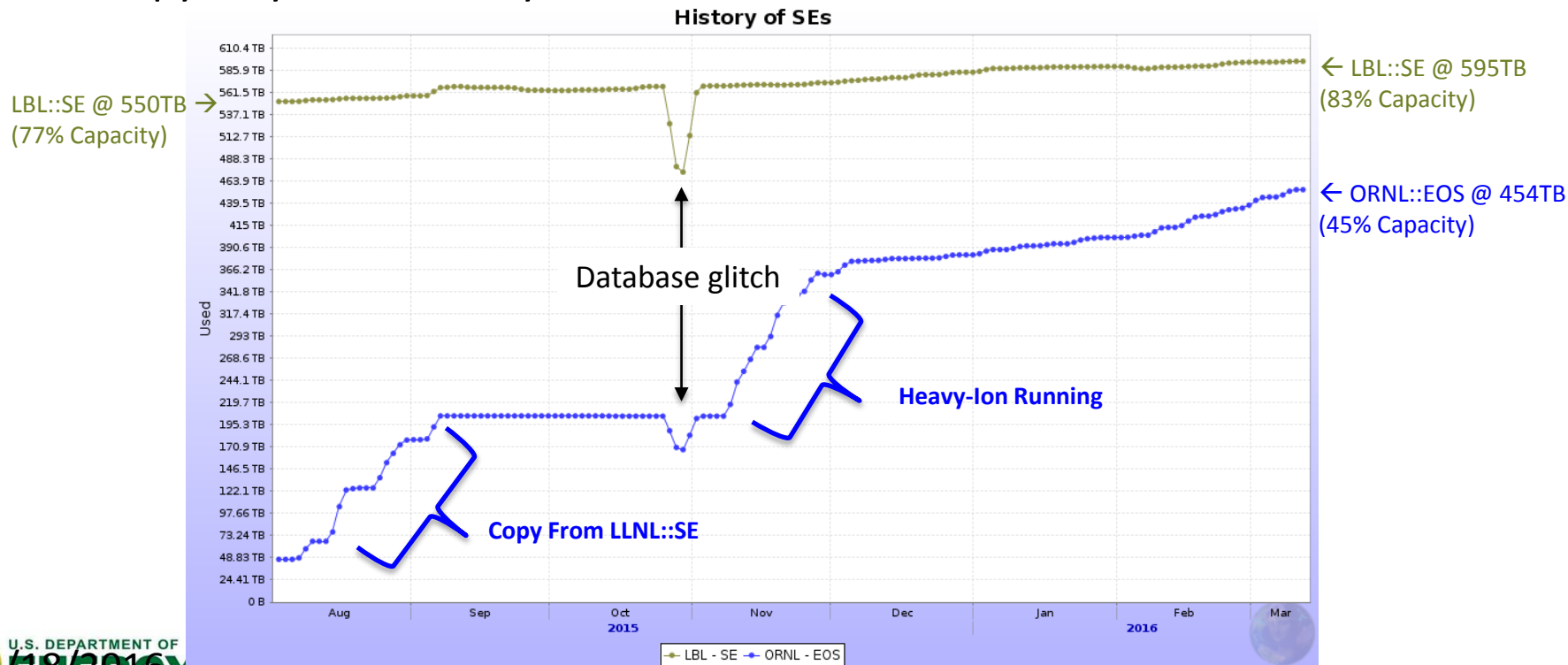


Statistics							
Link name	Data		Individual results of reading tests			Overall	
	Starts	Ends	Successful	Failed	Success ratio	Availability	
LBL::SE	01 Jul 2015 20:18	13 Mar 2016 20:18	2972	101	96.71%	96.76%	
ORNL::EOS	01 Jul 2015 20:22	13 Mar 2016 20:21	3009	64	97.92%	97.94%	

Storage Utilization & ALICE Data Replication Model



- **Data replication done at job completion**
 - 1 copy sent to local storage element
 - 1 copy to storage that is “nearby” in RTT time
 - Algorithm weights by available space & reliability
 - 1 copy may be randomly distributed



Section IV



- **2016 Plans**

2016 Project Plan Per T2 Site

Updated Nov. 2015



- **CPU plan**

- Stable @ ORNL
 - 18 kHS06
- Steady increase @ LBNL
 - 8.0 → 10.5 → 16.0 kHS06

- **Disk plan**

- Steady increase both sites
 - ORNL: 1.0 → 1.4 → 1.8 PB
 - LBNL: 0.9 → 1.3 → 1.7 PB

- **Latest Update from ALICE**

- CPU increase
 - 2.3% (2016) + 4.6% (2017)
- Disk
 - Flat (2016) + 2% (2017)
- Should be manageable

Plan As Approved by DOE

Resource	Installed	FY2016	FY2017
LBNL HW & Costs			
CPU change (+/- kHS06)		+2.5	-4.0+9.5
CPU installed (kHS06)	8.0	10.5	16.0
Disk change (+/- PB)		+0.4	+0.4
Disk installed (PB)	0.9	1.3	1.7
ORNL HW & Costs			
CPU change (+/-kHS06)		0	0
CPU installed (kHS06)	18.3	18.3	18.3
Disk change (+/-PB)		+0.4	+0.4
Disk installed (PB)	1.0	1.4	1.8

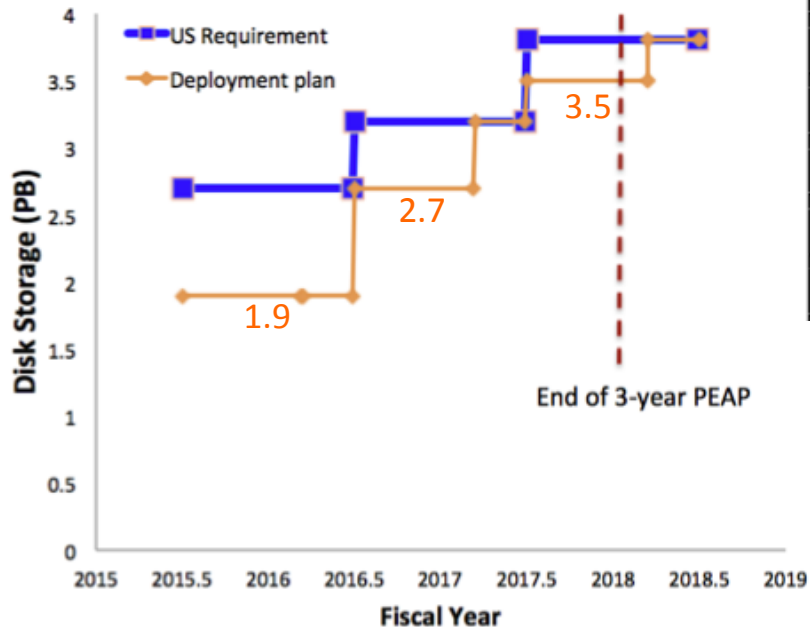
2016 Project Plan, US Sites Combined

Updated Nov. 2015



- **Targets:**

- 100% CPU on time
- 100% Disk by heavy-ion run



Plan As Approved by DOE

Resource	Installed	FY2016	FY2017
ALICE-USA Obligations			
CPU (kHS06)	23.0	28.4	34.5
Disk (PB)	2.7	3.2	3.8
ALICE-USA Plan			
CPU (kHS06)	26.3	28.8	34.3
% CPU obligation	113%	101%	100%
Disk (PB)	1.9	2.7	3.5
% disk obligation	70%	84%	92%
Disk deficit (PB)	0.8	0.5	0.3

Summary

- **ALICE-USA Computing Project**
 - Provides all ALICE-USA computing obligations to ALICE
 - Operates two similarly sized facilities
 - LBNL T2 at NERSC
 - ORNL T2 at CADES
- **2015 Operations**
 - Successful transition from LLNL T2 to the ORNL T2
 - Met ALICE-USA obligations
 - Move of resources at NERSC
 - Delayed until 2016
 - Extended use of older PDSF resources
- **2016 Plan**
 - HW deployment schedule should continue to meet obligations

Backup Slides



NERSC/PDSF Transition & Operations



- **PDSF lost critical people in 2015**
 - I.Sakrejda Retired in summer before transition
 - Temporarily lost application support funding (L. Gerhardt)
 - Funding reinstated & new hire will begin May 2, 2016
 - Several operational tasks have been delayed

- **Local request for new model that saves effort by leveraging NERSC capabilities**
 - NERSC primarily an HPC Center
 - New projects added to migrate workflow to NERSC HPC
 - Target Cori: newest HPC machine at NERSC

Plans for Cori Deployment

- **Cori Phase I → “Data Partition”**

- Installed Oct. 2015
- 1630 Haswell nodes
 - 32 core/node, 4GB/core
- 28 PB file system
- Outgoing network access
- But no local disk for CVMFS

- Payload R&D: Markus Fasel
 - See ACAT 2016
 - <http://tinyurl.com/howad8l>
- Planning discussions:
 - ALICE & NERSC teams, Mar 2016
- Buy-in options exist

- **Cori Phase II → “HPC Partition”**

- Planned for Summer 2016
- 9000 Knights Landing nodes
 - 72 core/node, ~1.2GB/core
- Use will require dedicated R&D



<http://www.nersc.gov/users/computational-systems/cori>