ALICE USA Computing Project Status Report

R. Jeff Porter

ALICE-USA Computing Resource Meeting

March 14, 2017

U.S. DEPARTMENT OF ENERGY | Office of Science

Outline of Project Status Report

- History & Status
- Current Operations
- Status Relative to Execution Plan
- Project Plans for 2017 & Beyond

Section I

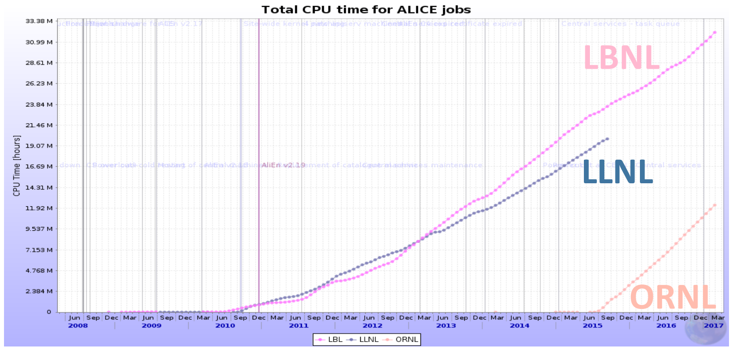
- **Project History**

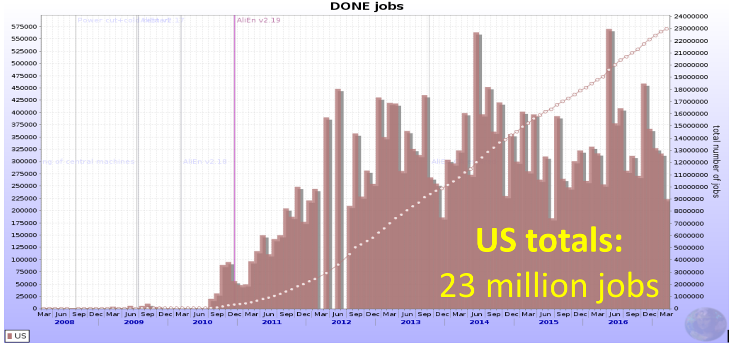
ALICE-USA Computing Project

- **Original 2009 Project Proposal**
  - Goal to fulfill MoU-base ALICE USA obligations for compute resources to ALICE
  - Operate facilities at 2 DOE labs
    - NERSC @ LBNL
    - Livermore Computing @ LLNL
  - LBNL as the host institution

➢ **In operational since 2010**

- **New Project Proposal in 2014**
  - Approved in Sept. 2014
  - Replace LLNL/LC with ORNL/CADES
  - ORNL CADES T2 fully operational in 2015





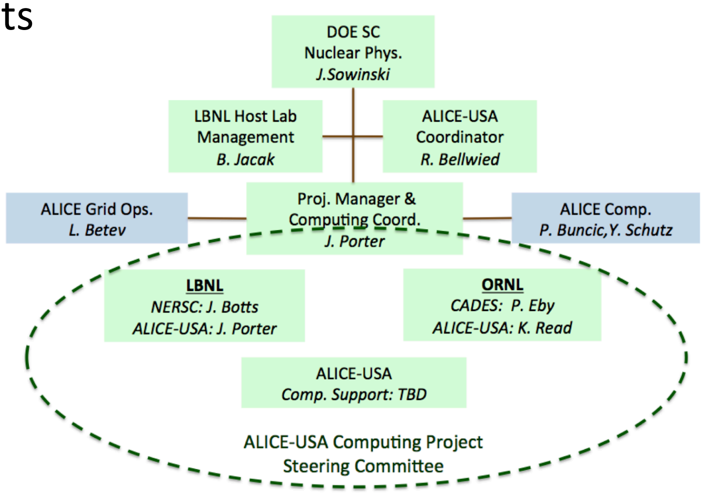
US totals:

23 million jobs

- **Project working documents:**
  - Project SLA: Institutions & roles (currently on hold with NERSC@LBNL uncertainty)
  - Project Execution & Acquisition Plan: → **PEAP updated to DOE annually**

Project Organization & Computing Steering Committee

- **Project Steering Committee:**
  - Currently: J.Porter, K.Read, P.Eby, M.Galloway, J.Botts
  - Local documentation moved to LBNL Google docs:
    - Project Document repository
    - Monthly Meetings & minutes
  - Email list



DOE SC
Nuclear Phys.
J.Sowinski

LBNL Host Lab
Management
B. Jacak

ALICE-USA
Coordinator
R. Bellwied

ALICE Grid Ops.
L. Betev

Proj. Manager &
Computing Coord.
J. Porter

ALICE Comp.
P. Buncic,Y. Schutz

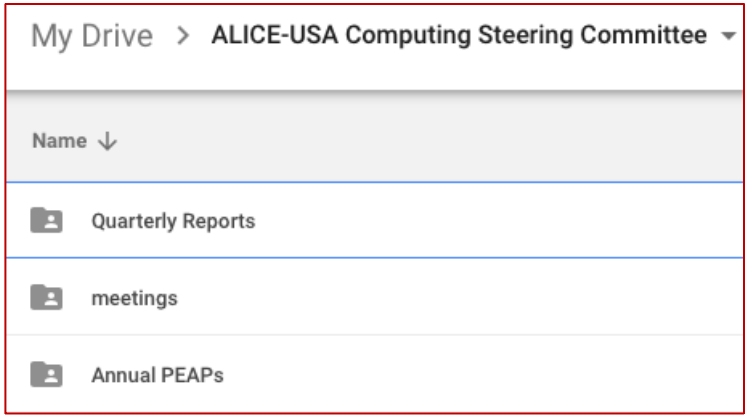
LBNL
NERSC: J. Botts
ALICE-USA: J. Porter

ORNL
CADES: P. Eby
ALICE-USA: K. Read

ALICE-USA
Comp. Support: TBD

ALICE-USA Computing Project
Steering Committee

- **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
    - 2017 Strasbourg, Fr: J. Porter & M. Galloway
  - Annual AliEn Developers Workshops
    - 2010 – 2012, J.Porter
    - 2013, J. Porter & B. Nilsen



My Drive > ALICE-USA Computing Steering Committee

Name ↓

Quarterly Reports

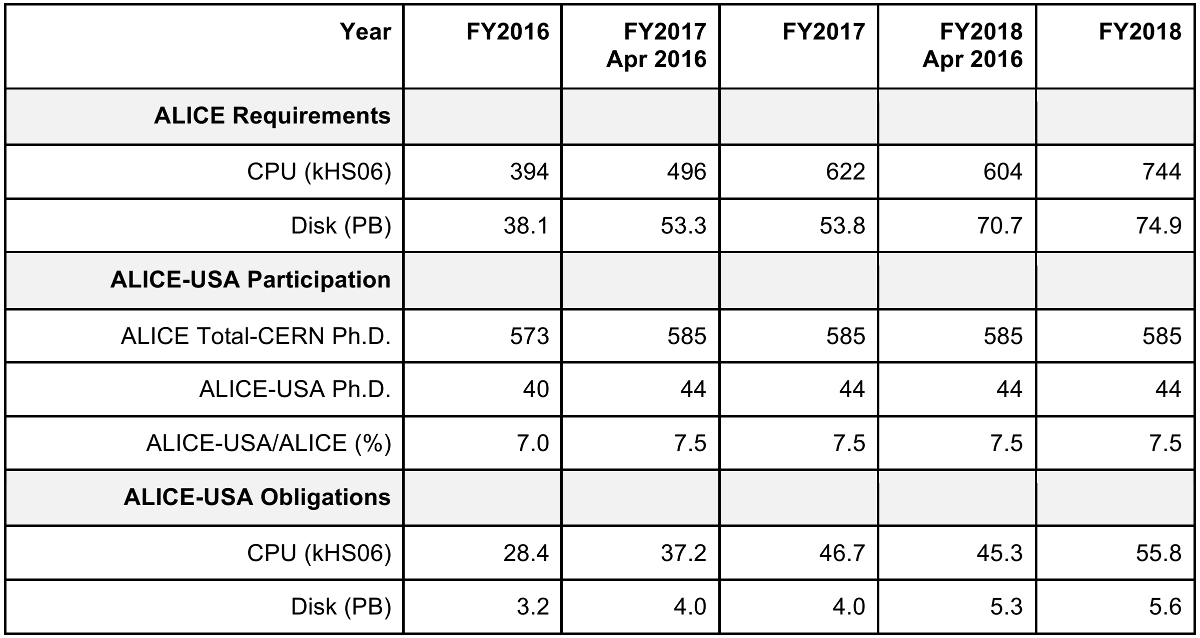
meetings

Annual PEAPs

ALICE-USA Obligation Evaluation

- **ALICE Computing Requirements**
  - Established Annually, reported to the ALICE Computing Board & approved by WLCG

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



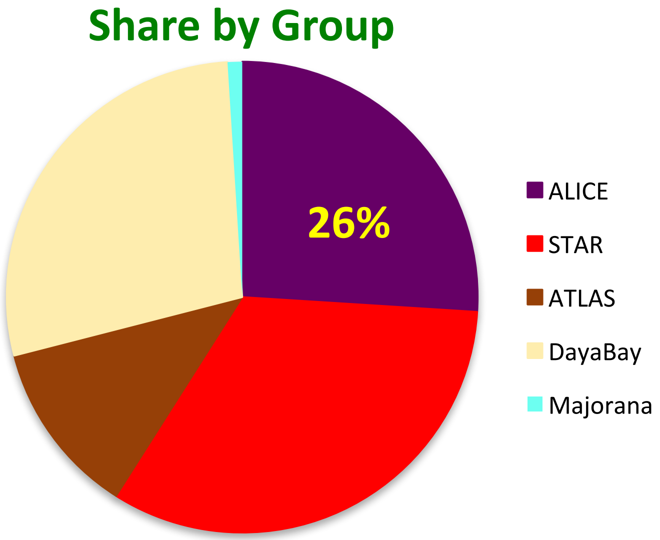
Year	FY2016	FY2017 Apr 2016	FY2017	FY2018 Apr 2016	FY2018
ALICE Requirements					
CPU (kHS06)	394	496	622	604	744
Disk (PB)	38.1	53.3	53.8	70.7	74.9
ALICE-USA Participation					
ALICE Total-CERN Ph.D.	573	585	585	585	585
ALICE-USA Ph.D.	40	44	44	44	44
ALICE-USA/ALICE (%)	7.0	7.5	7.5	7.5	7.5
ALICE-USA Obligations					
CPU (kHS06)	28.4	37.2	46.7	45.3	55.8
Disk (PB)	3.2	4.0	4.0	5.3	5.6

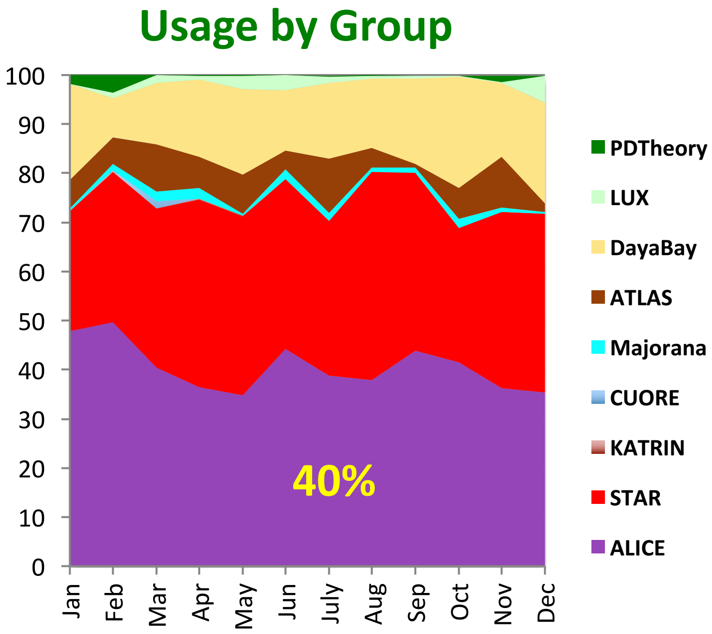
FY17 PEAP Update Submitted to DOE in Dec. 2016

- **ALICE-USA Obligations:**
  - Fraction of ALICE Requirements Defined by proportion of ALICE-USA to ALICE

LBNL T2 Site: PDSF @ NERSC

- **PDSF is an evergreen cluster operated by NERSC for HEP/NP Experiment Community**
  - Share based on investment (both shared HW and FTE support)
  - PDSF supports about 9 active groups and 800 active users

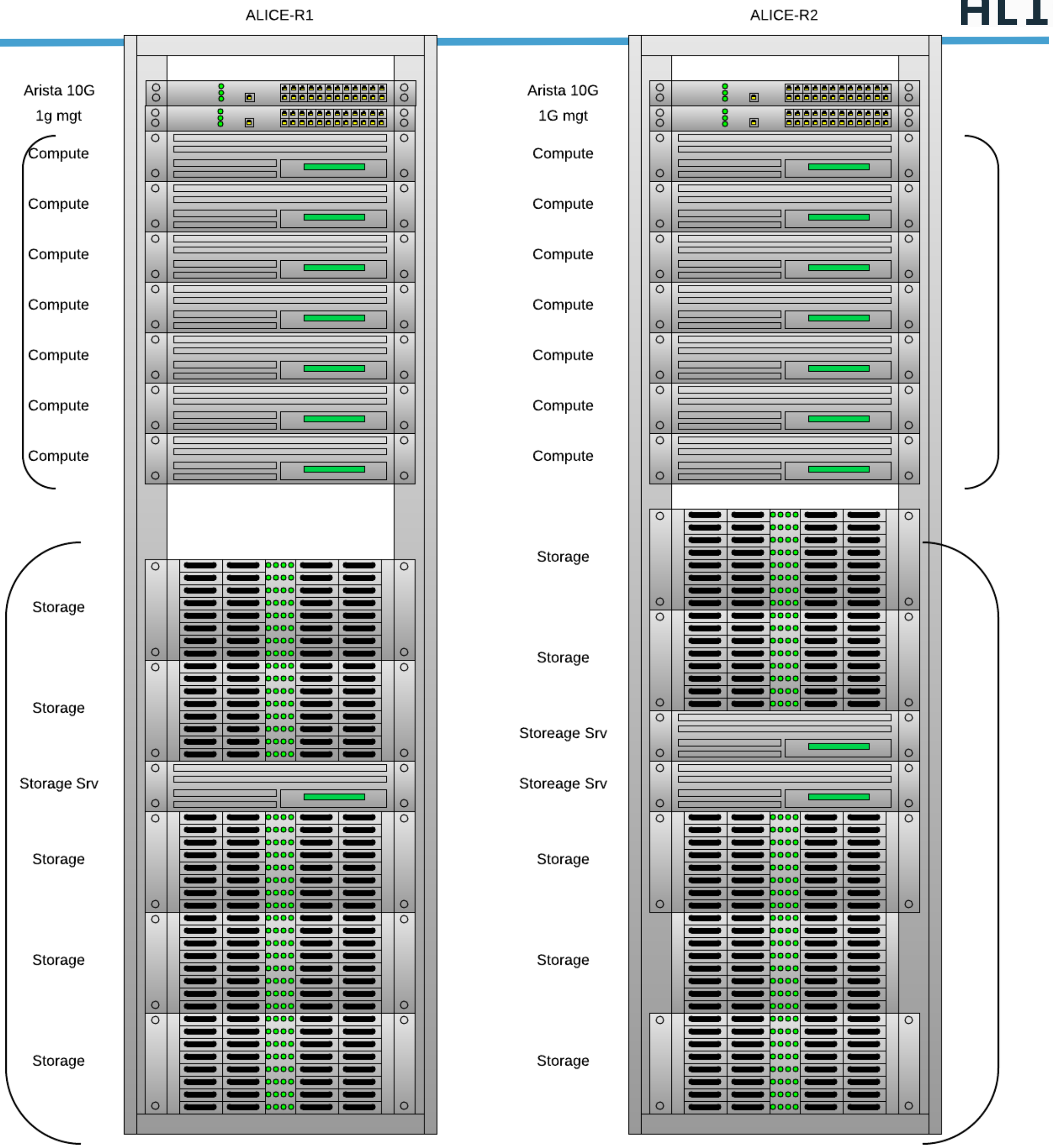




ORNL T2 Site: T2 @ CADES

- **Single use ALICE Grid facility located within a larger center**
  - One User: ALICE
  - May leverage access to other CADES resources

- **T2 Compute & Storage overview**

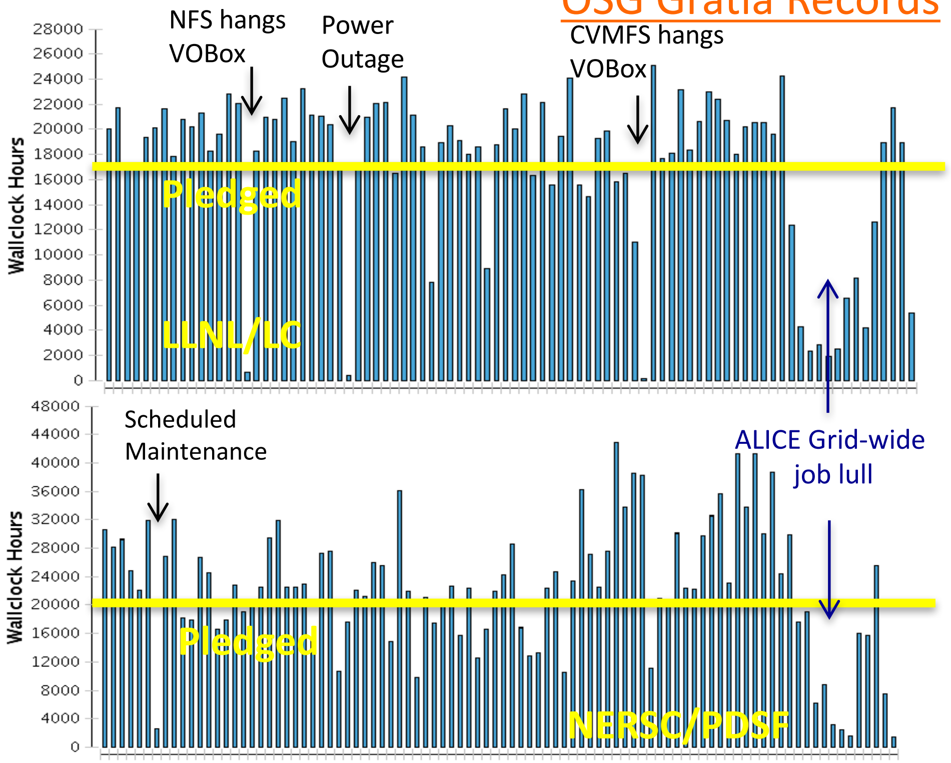


U.S. DEPARTMENT OF ENERGY | Office of Science

ALICE-USA & The Open Science Grid

- **ALICE-USA computing project leverages OSG capabilities**
  - OSG Registration Authority
    - ALICE-USA user certificates
      - Deprecated in favor of CERN CA
    - Host & service certificates
      - Grid admins: P.Eby & J.Porter
  - Reports to WLCG
    - Accounting Reports
      - Gratia site service to OSG rep.
      - Central OSG service to WLCG

OSG Gratia Records



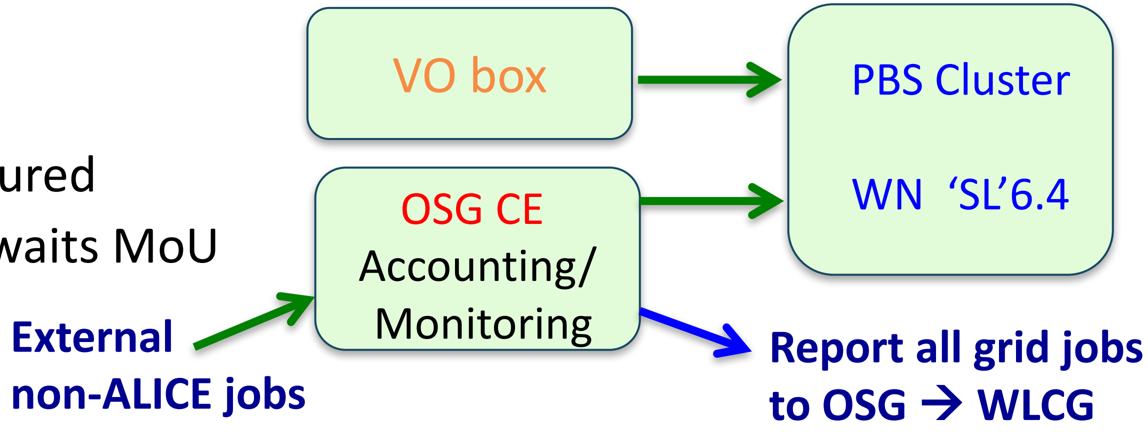
Q1FY14 Project Report   Oct 1 – Dec 31, 2013

- **Expect ORNL to report to:**
  - OSG soon (~Mar/Apr)
  - WLCG … still unknown but have been told just a matter of months

US Site Configurations with OSG

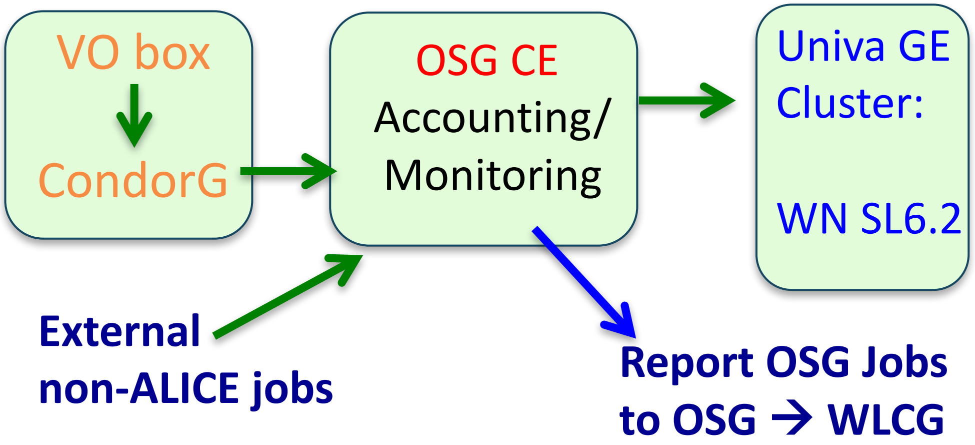
ORNL CADES

- VObox submits to PBS
- OSG CE is being configured
- WLCG reporting still awaits MoU



LBNL NERSC PDSF

- VObox submits to CondorG
- CondorG submits to OSG-CE
  - now HTCondor-CE
- OSG-CE submits to UGE
- OSG Accounting
  - Monitors batch logs

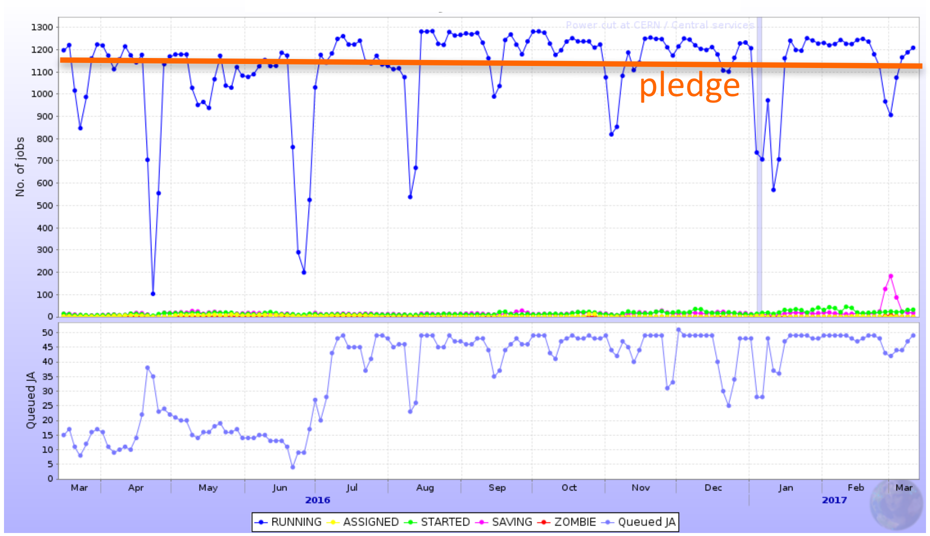


Section II

- **2016 Operations**

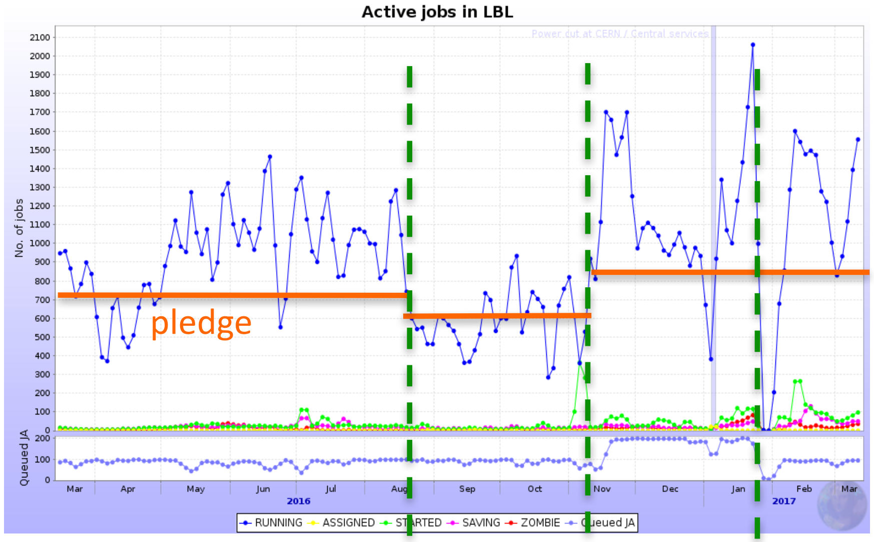
Site Job Profiles

ORNL Jobs 2016 RRB



pledge

LBL Jobs 2016 RRB



pledge

Ave. Runing Jobs:

LBL = ~900

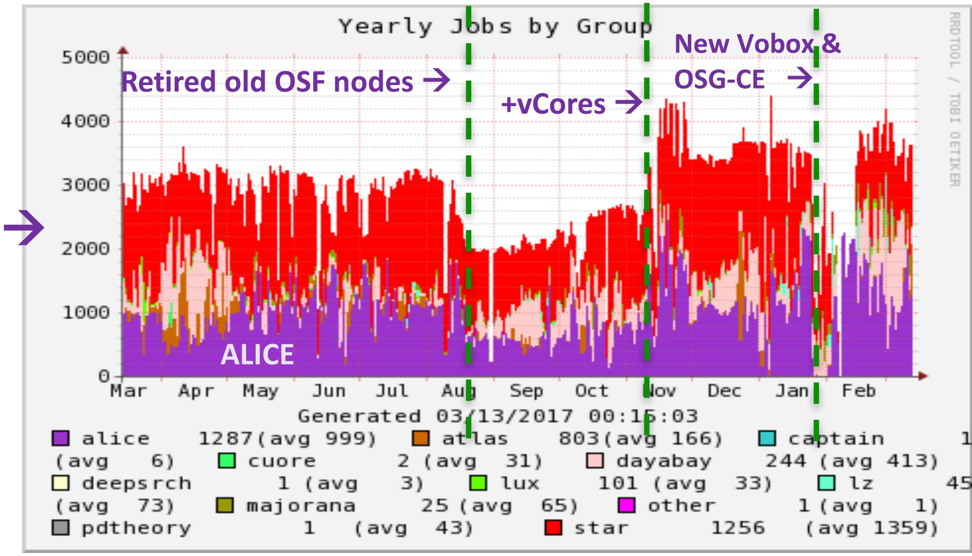
ORNL = 1170

Zombie Rate remains low:

LBL ~1.0%

ORNL ~0.1%

PDSF Jobs →



Retired old OSF nodes →

+vCores →

New Vobox & OSG-CE →

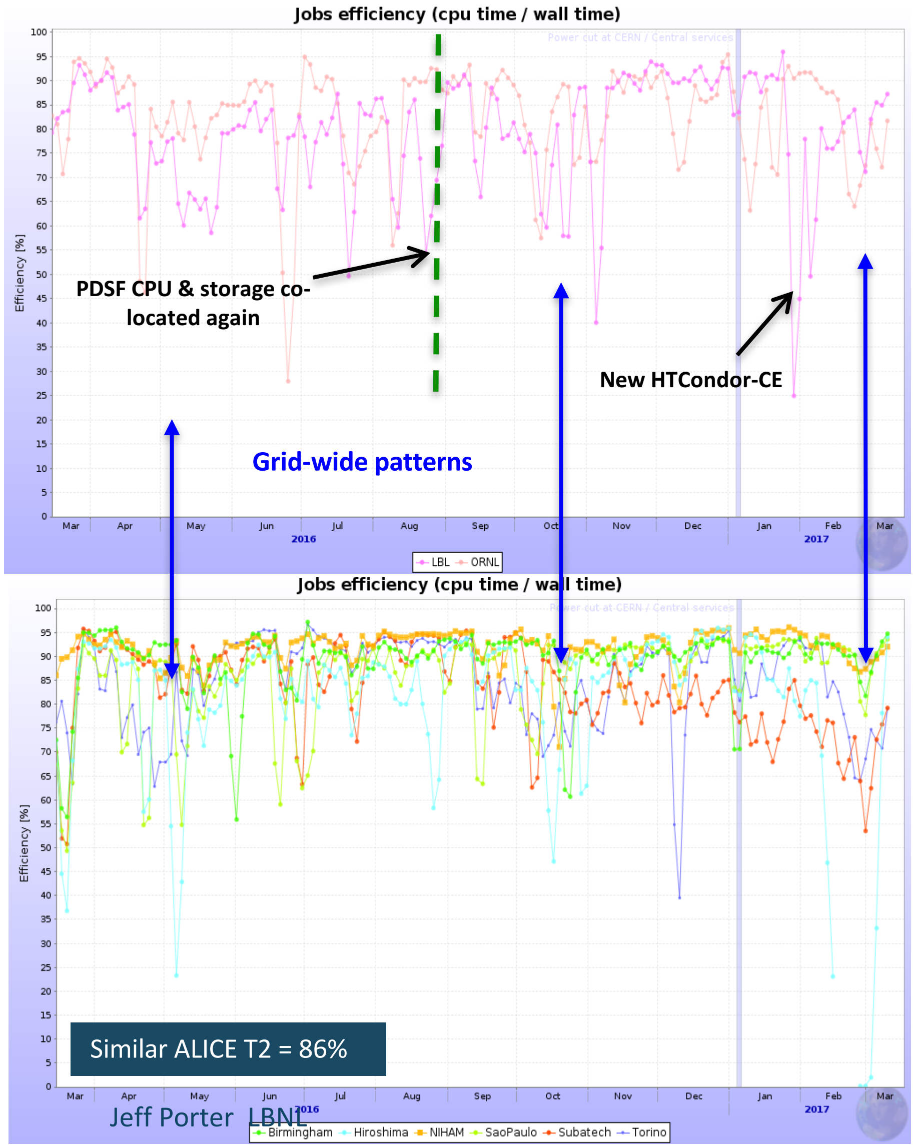
ALICE

Site Efficiencies: cpu-time/wall-time

- **Ave Site Efficiencies**
  - ORNL 82.5%
  - LBL 79%

- **Results track other ALICE T2s**
  - Similar T2s ~ 86%
  - All ALICE T2s ~ 79%

- **Specific issues:**
  - PDSF operated CPU & storage at different sites until end of Aug
  - PDSF new HTCondor-CE problems



Jobs efficiency (cpu time / wall time)

PDSF CPU & storage co-located again

New HTCondor-CE

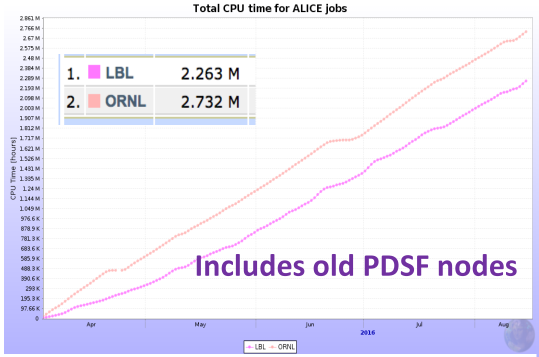
Grid-wide patterns

Jobs efficiency (cpu time / wall time)

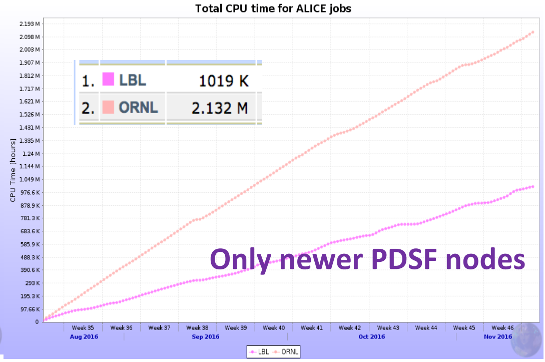
Similar ALICE T2 = 86%

CPU Delivered to ALICE Grid Relative to Pledged Obligations

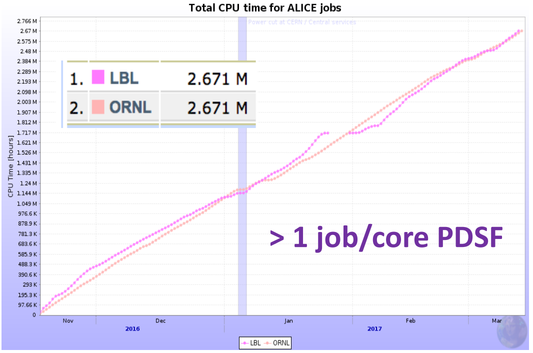
CPU Delivered Per US Site



Includes old PDSF nodes



Only newer PDSF nodes

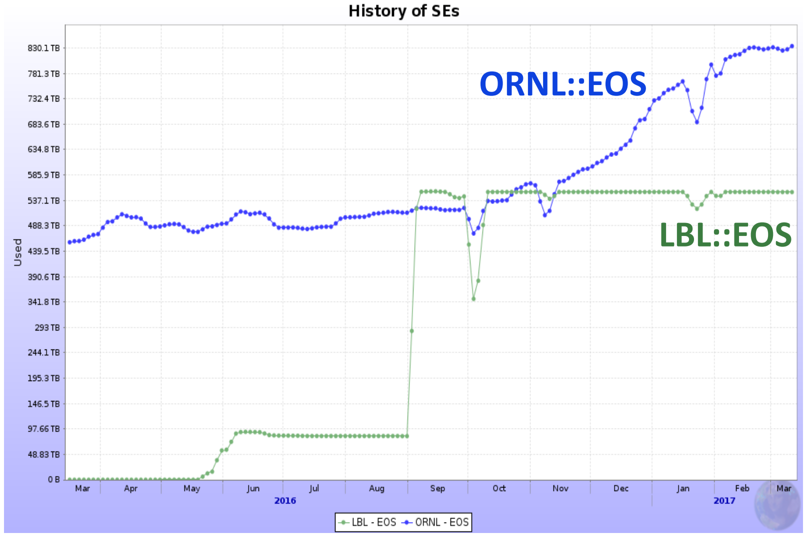


> 1 job/core PDSF

Site	Per-core Capacity (HS06/core)	CPU delivered (Mhrs)	CPU delivered (MHS06-hrs)	US Obligation (pledge*hrs*0.70)	% delivered
LBNL	16.6, 19.8, 14.0	2.26+1.02+2.67 = 5.85	37.5+20.2+37.3= 95.0	70.2	136%
ORNL	14.0	2.73+2.13+2.67 = 7.53	105.4	99.4	106%
Totals			200.4	169.6	118%

Storage Capacity & Utilization

- **Storage Deployment History**
  - LBNL NERSC
    - LBL::SE retired in September 2016
    - Installed 825TB, commissioned in Aug 2016
    - Project plan to add 600TB when site is selected
  - ORNL CADES
    - 1000 TB installed as EOS storage, June 2015
    - 450 TB is ready on the floor, not enabled
    - Project Plan calls for 300 TB more in FY17



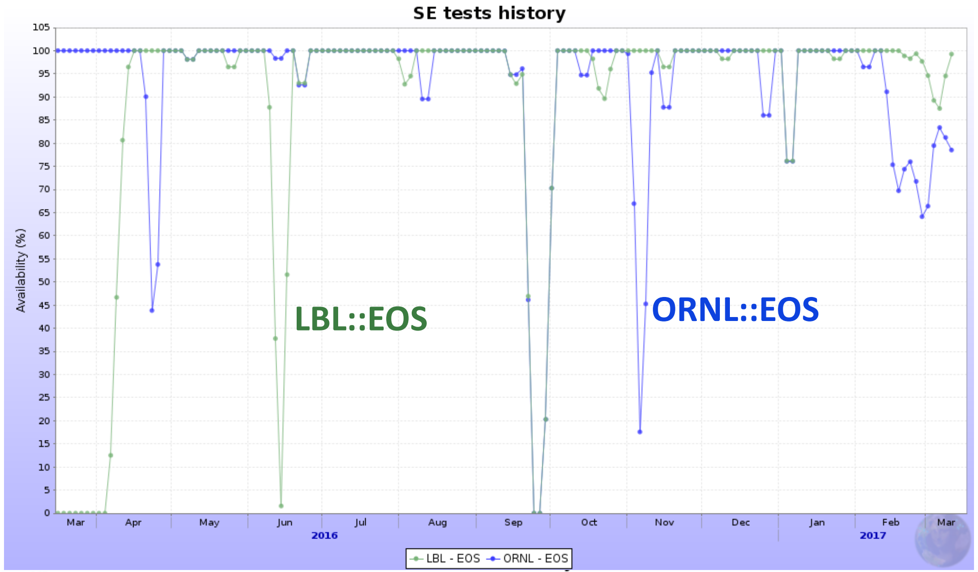
ALICE-USA Storage Elements Capacities & Usage: 03/2016

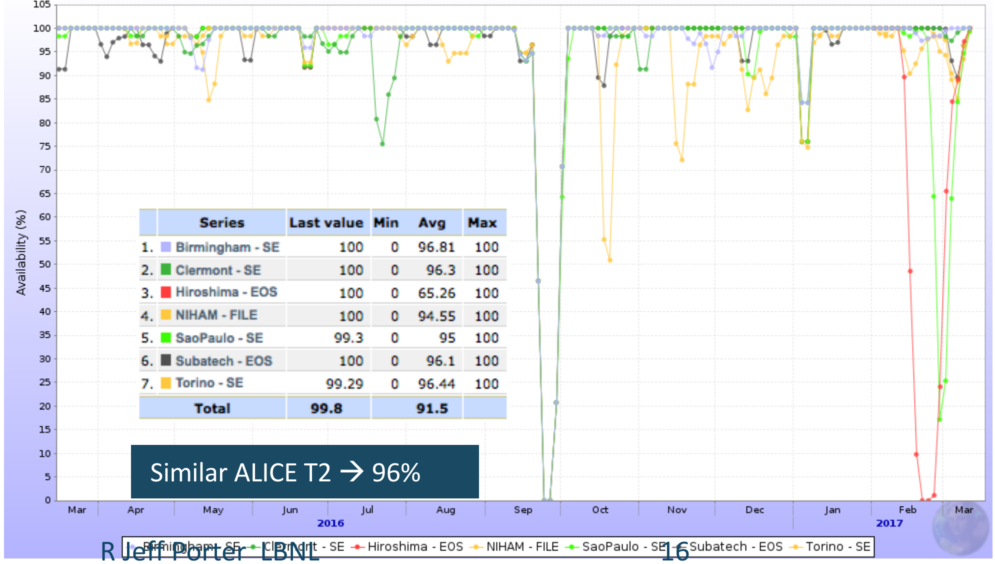
ALICE SE	#-servers	Space (TB)	Used Space (TB)	% Used
LBL::EOS	3	826	553	67
ORNL::EOS	4	1024	834	81
In Production	7	1850	1386	75

SE Availability Tests MonALISA Monitoring

- **US T2 SE availability**
  - ORNL::EOS → 92.7%
  - LBL::SE → 93.4%

- **ALICE T2 SEs**
  - All ALICE T2s ~86%
  - Similar T2s as US ~96%



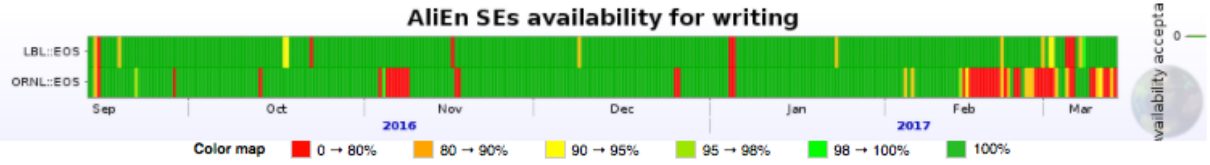


	Series	Last value	Min	Avg	Max
1.	Birmingham - SE	100	0	96.81	100
2.	Clermont - SE	100	0	96.3	100
3.	Hiroshima - EOS	100	0	65.26	100
4.	NIHAM - FILE	100	0	94.55	100
5.	SaoPaulo - SE	99.3	0	95	100
6.	Subatech - EOS	100	0	96.1	100
7.	Torino - SE	99.29	0	96.44	100
	Total	99.8		91.5	

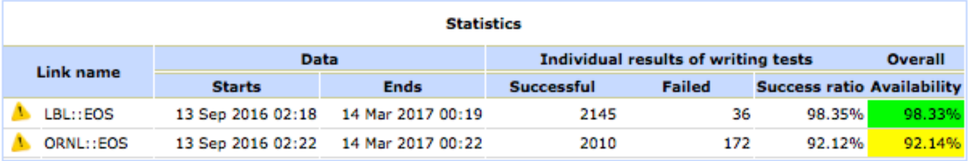
Similar ALICE T2 → 96%

SE Availability Continued

- **Writing**
  - LBL ~98%
  - ORNL ~92%



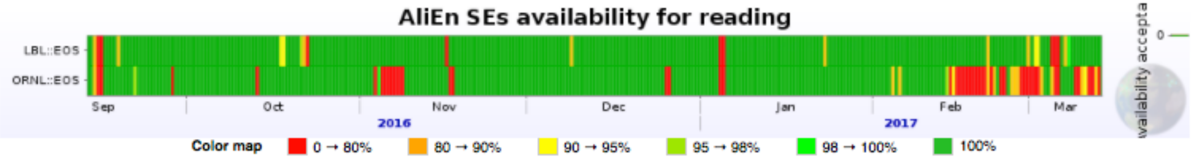
AliEn SEs availability for writing



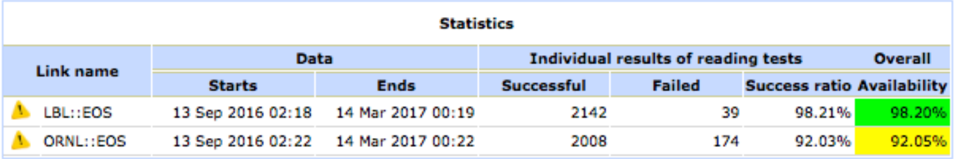
Statistics

Link name	Data		Individual results of writing tests			Overall
	Starts	Ends	Successful	Failed	Success ratio	Availability
⚠ LBL::EOS	13 Sep 2016 02:18	14 Mar 2017 00:19	2145	36	98.35%	98.33%
⚠ ORNL::EOS	13 Sep 2016 02:22	14 Mar 2017 00:22	2010	172	92.12%	92.14%

- **Reading**
  - LBL ~98%
  - ORNL ~ 92%



AliEn SEs availability for reading



Statistics

Link name	Data		Individual results of reading tests			Overall
	Starts	Ends	Successful	Failed	Success ratio	Availability
⚠ LBL::EOS	13 Sep 2016 02:18	14 Mar 2017 00:19	2142	39	98.21%	98.20%
⚠ ORNL::EOS	13 Sep 2016 02:22	14 Mar 2017 00:22	2008	174	92.03%	92.05%

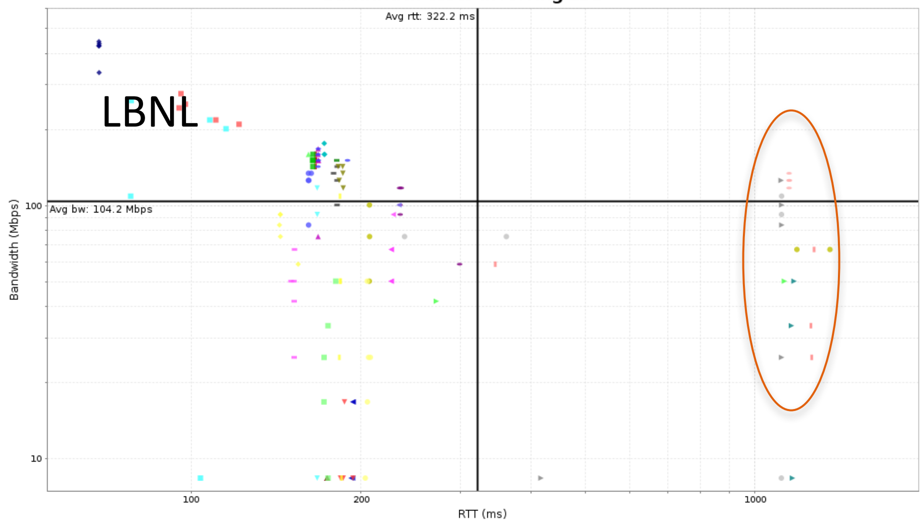
Site Bandwidth Tests

- **Single Stream Test**
  - Between every ALICE VOBox pair

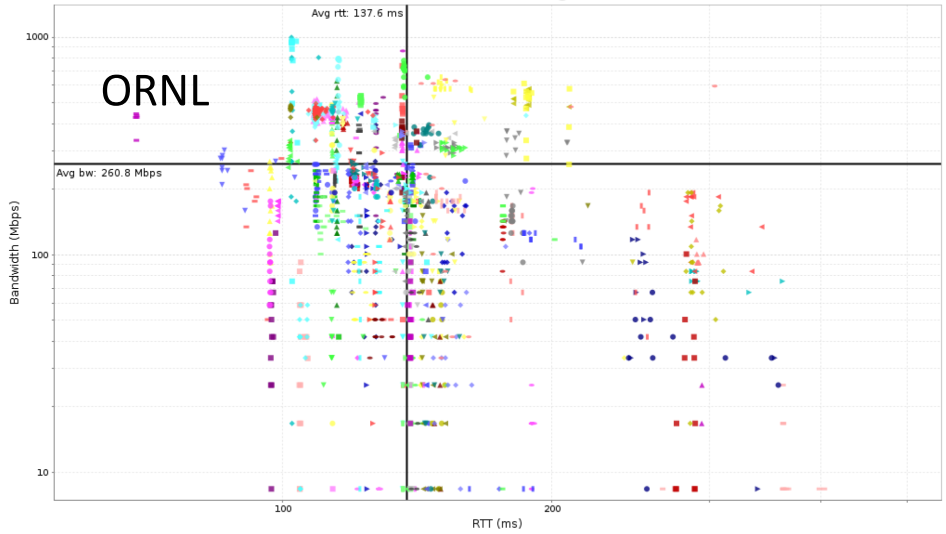
measure	2016	2017
LBNL <RTT> (ms)	190	322
LBNL <bw> (Mbps)	80	101
ORNL RTT> (ms)	155	138
ORNL <bw> (Mbps)	211	260
Global <RTT> (ms)	163	111
Global <bw> (Mbps)	154	255

  - Only LBNL RTT is worse !!

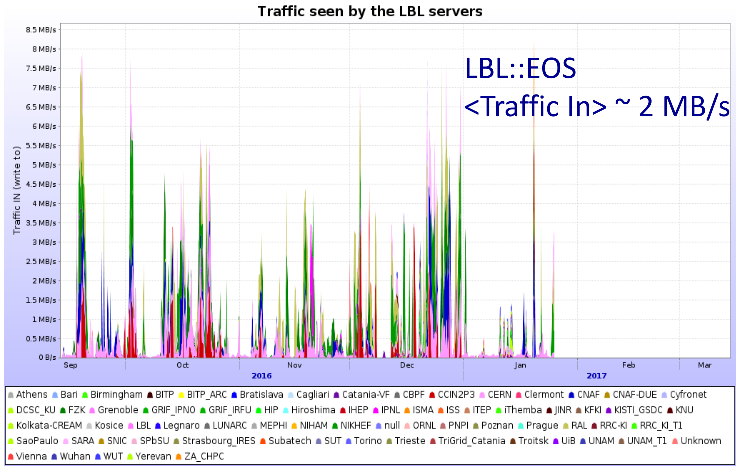
Bandwidth tests involving LBL



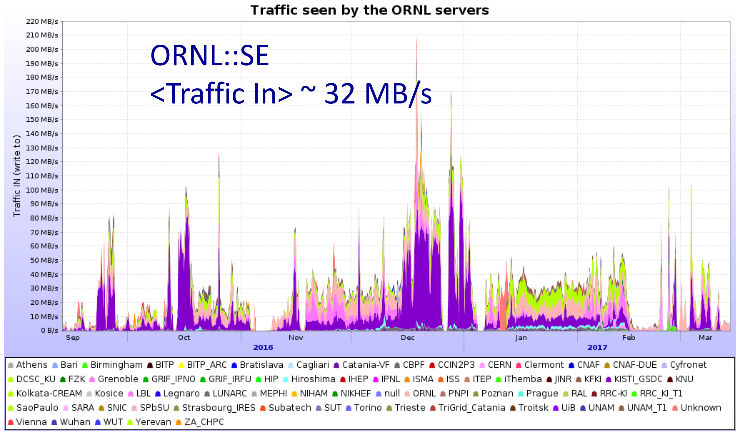
Bandwidth tests involving ORNL



Traffic Into US Storage by Source



LBL::EOS
<Traffic In> ~ 2 MB/s



ORNL::SE
<Traffic In> ~ 32 MB/s

2017

Max into LBNL::EOS ~30MB/s

Max into ORNL::EOS ~1.4GB/s

2016

LBL::SE	Max Rate
ORNL	300 MB/s
KISTI	240 MB/s
LBNL	150 MB/s
UNAM	125 MB/s
Hiroshima	95 MB/s
RRC_T1	75 MB/s
FZK	50 MB/s

2017

ORNL::EOS	Max Rate
UIB	- MB/s
KISTI	1370 MB/s
ORNL	750 MB/s
LBNL	500 MB/s
RAL	18MB/s
UNAM	300 MB/s
Hiroshima	170 MB/s

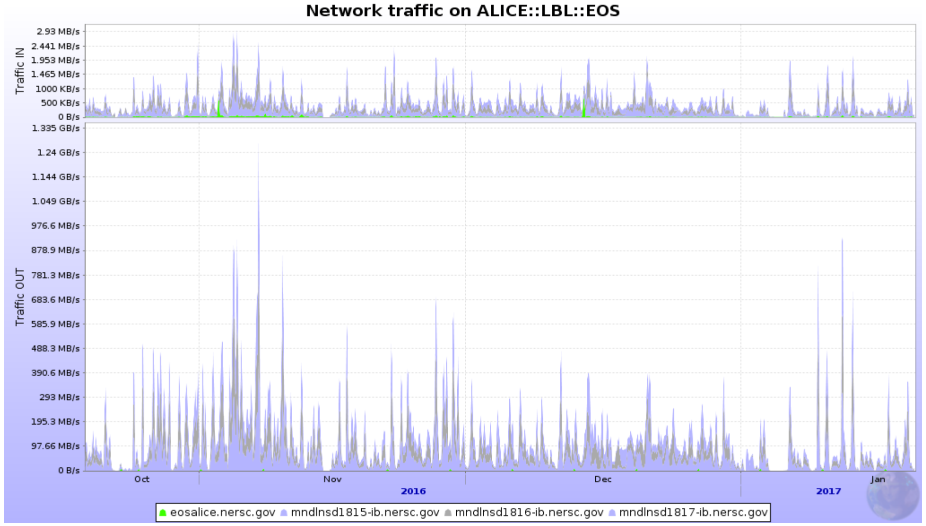
2016

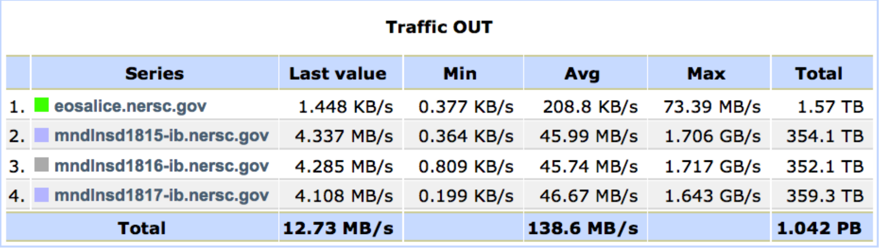
ORNL::EOS	Max Rate
UIB	500 MB/s
KISTI	325 MB/s
ORNL	320 MB/s
LBNL	200 MB/s
RAL	120 MB/s
UNAM	110 MB/s
Hiroshima	100 MB/s

SE I/O Performance: LBL::SE

- **ALICE::LBL::EOS**
  - 3 Servers + 1 redirector
  - 826 TB usable

- **Observations**
  - Data & I/O evenly distributed across servers
  - Ave ~130MB/s aggregate
  - >1.5GB/s peak/server



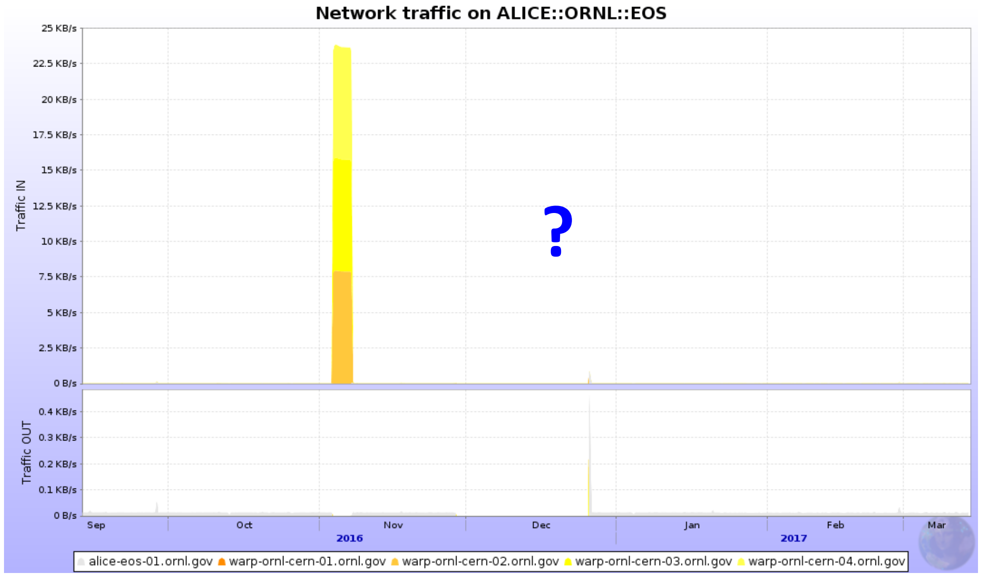


Traffic OUT

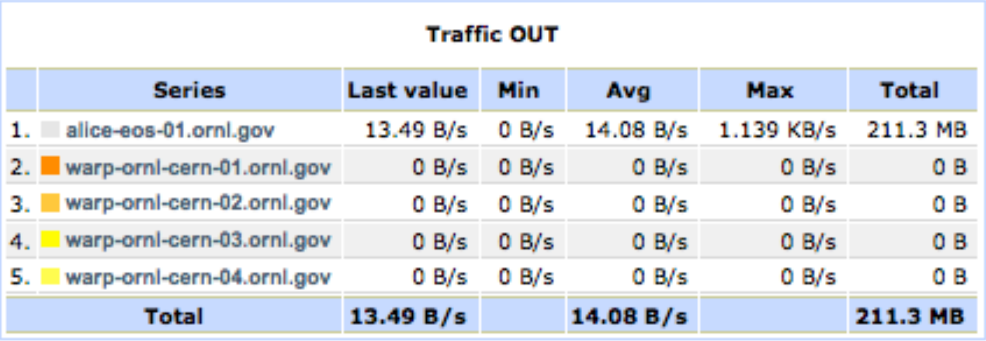
	Series	Last value	Min	Avg	Max	Total
1.	🟩 eosalice.nersc.gov	1.448 KB/s	0.377 KB/s	208.8 KB/s	73.39 MB/s	1.57 TB
2.	🟪 mndlnsd1815-ib.nersc.gov	4.337 MB/s	0.364 KB/s	45.99 MB/s	1.706 GB/s	354.1 TB
3.	⬛ mndlnsd1816-ib.nersc.gov	4.285 MB/s	0.809 KB/s	45.74 MB/s	1.717 GB/s	352.1 TB
4.	🟪 mndlnsd1817-ib.nersc.gov	4.108 MB/s	0.199 KB/s	46.67 MB/s	1.643 GB/s	359.3 TB
	Total	**12.73 MB/s**		**138.6 MB/s**		**1.042 PB**

SE I/O Performance: ORNL::EOS

- **ALICE::ORNL::EOS**
  - ALICE EOS installation
  - ZFS file systems
  - 4 Servers + 1 MGM
  - 1000 TB



- **Observations**
  - Um ….
  - ORNL have internal monitoring
    - Transfer rates look good (see slide 15)



Traffic OUT

	Series	Last value	Min	Avg	Max	Total
1.	⬜ alice-eos-01.ornl.gov	13.49 B/s	0 B/s	14.08 B/s	1.139 KB/s	211.3 MB
2.	🟧 warp-ornl-cern-01.ornl.gov	0 B/s	0 B/s	0 B/s	0 B/s	0 B
3.	🟧 warp-ornl-cern-02.ornl.gov	0 B/s	0 B/s	0 B/s	0 B/s	0 B
4.	🟨 warp-ornl-cern-03.ornl.gov	0 B/s	0 B/s	0 B/s	0 B/s	0 B
5.	🟨 warp-ornl-cern-04.ornl.gov	0 B/s	0 B/s	0 B/s	0 B/s	0 B
	Total	13.49 B/s		14.08 B/s		211.3 MB

Section III

- **Status Relative to Project Plan**

Major Tasks for FY2016

- **Finish NERSC transition to new building**
  - Commission new LBNL::EOS, decommission LBL::SE
  - Move all CPU to hill, decommission old CPU at OSF
  - Move VOBox and OSG CE to hill, decommission old
- **Establish new SLA between NERSC, CADES, & ALICE**
- **Make new HW purchases**
  - Storage at ORNL
  - Storage and CPU at NERSC
- **OSG services to HTCondor-CE**
  - Upgrade at NERSC
  - Install & deploy at CADES
- **ORNL Sign WLCG MoU & pledge resources**
- **LHCONE on ALICE-USA sites**
  - Initiate at ORNL
  - Evaluate whether possible at NERSC/PDSF

Major Tasks for FY2016

- **Finish NERSC transition to new building** ✓
  - Commission new LBNL::EOS, decommission LBL::SE
  - Move all CPU to hill, decommission old CPU at OSF
  - Move VOBox and OSG CE to hill, decommission old → delayed into early FY2017
- **Establish new SLA between NERSC, CADES, & ALICE** ✗
- **Make new HW purchases** ✓
  - Storage at ORNL → storage purchased but not deployed
  - Storage and CPU at NERSC → extended life & relied on new shares
- **OSG services to HTCondor-CE** ✓ delayed into early FY2017
  - Upgrade at NERSC
  - Install & deploy at CADES
- **ORNL Sign WLCG MoU & pledge resources** ✗
- **LHCONE on ALICE-USA sites** ~
  - Initiate at ORNL → in progress
  - Evaluate whether possible at NERSC/PDSF → may not need

Project Planning Issues:

- **PDSF Lifetime:**
  - We were told at the Jan 21st PDSF Steering Committee that there was no room for a PDSF cluster when the next NERSC system arrives ~ 2020.
    - Thus, no new CPU hardware on the cluster
      - James has details in his talk later today.
    - NERSC would like to support our work on their HPC systems
      - We'll discuss status on Weds afternoon

- **Budget uncertainty**
  - Always an issue but especially the next couple of years

- **Time for new 3-year project review**
  - New DOE program manager may be delayed
  - Earliest will be October, more likely January
    - So even though big changes will occur, decisions will be reviewed after the fact!

Some items to cover during the meeting

- **HW deployment:**
  - Establish location for all FY2017 HW
    - Evaluate ORNL/CADES capacity options
    - Evaluate T2 at LBNL IT
    - Evaluate storage options at PDSF
  - Evaluate growth goal relative to obligations
    - FY2017 and beyond
  - Establish overall hardware deployment schedule

- **Some technical tasks to tackle**
  - ORNL
    - OSG CE
    - Understand concurrent job mistmatch (PBS vs MonaLisa)
    - EOS failures
    - EOS network monitoring in ML
    - Deploy new storage
    - Increase vCore use on cluster
  - LBNL
    - Evaluate network issues, low write rates into EOS
    - High error rates in new OSG-CE

- **Evolution of workflow**
  - NERSC HPC CPU to make up for long term CPU deficit
  - Singularity in CVMFS &/or Shifter on NERSC

Section IV

- **2017 Plans**

ALICE-USA Obligation Evaluation

- **ALICE Computing Requirements**
  - Established Annually, reported to the ALICE Computing Board & approved by WLCG

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

Year	FY2016	FY2017 Apr 2016	FY2017	FY2018 Apr 2016	FY2018
ALICE Requirements					
CPU (kHS06)	394	496	622	604	744
Disk (PB)	38.1	53.3	53.8	70.7	74.9
ALICE-USA Participation					
ALICE Total-CERN Ph.D.	573	585	585	585	585
ALICE-USA Ph.D.	40	44	44	44	44
ALICE-USA/ALICE (%)	7.0	7.5	7.5	7.5	7.5
ALICE-USA Obligations		One time additional 26% jump in CPU			
CPU (kHS06)	28.4	37.2	46.7	45.3	55.8
Disk (PB)	3.2	4.0	4.0	5.3	5.6

FY17 PEAP Update Submitted to DOE in Dec. 2016

- **ALICE-USA Obligations:**
  - Fraction of ALICE Requirements Defined by proportion of ALICE-USA to ALICE

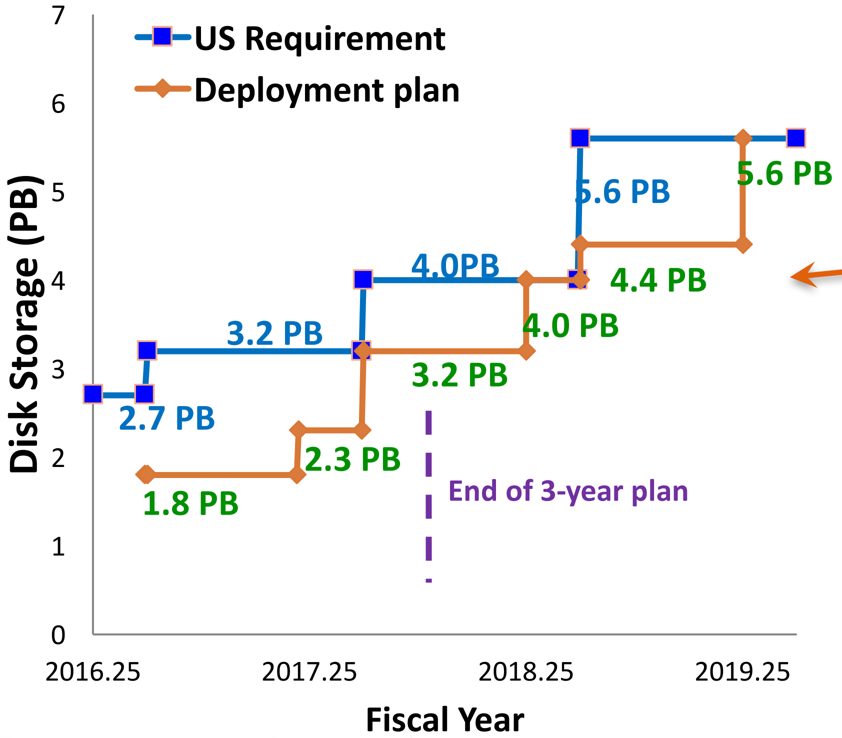
2017 PEAP Plan – Hardware

Resource	Currently Installed	FY2017 Original	FY2017 Dec. Plan	FY2018 Dec. Plan
LBNL HW				
CPU +/- kHS06		5.0	9.5	-4.0+9.0
CPU installed	12.0	17.0	21.5	27.0
Disk +/-		0.6	0.6	0.6
Disk installed	0.82	1.42	1.42	2.0
ORNL HW				
CPU +/- kHS06		3.5	7.5	3.0
CPU installed	17.0	20.5	24.5	27.5
Disk +/- (PB)		0.3	0.3	0.5
Disk installed	1.45	1.75	1.75	2.25

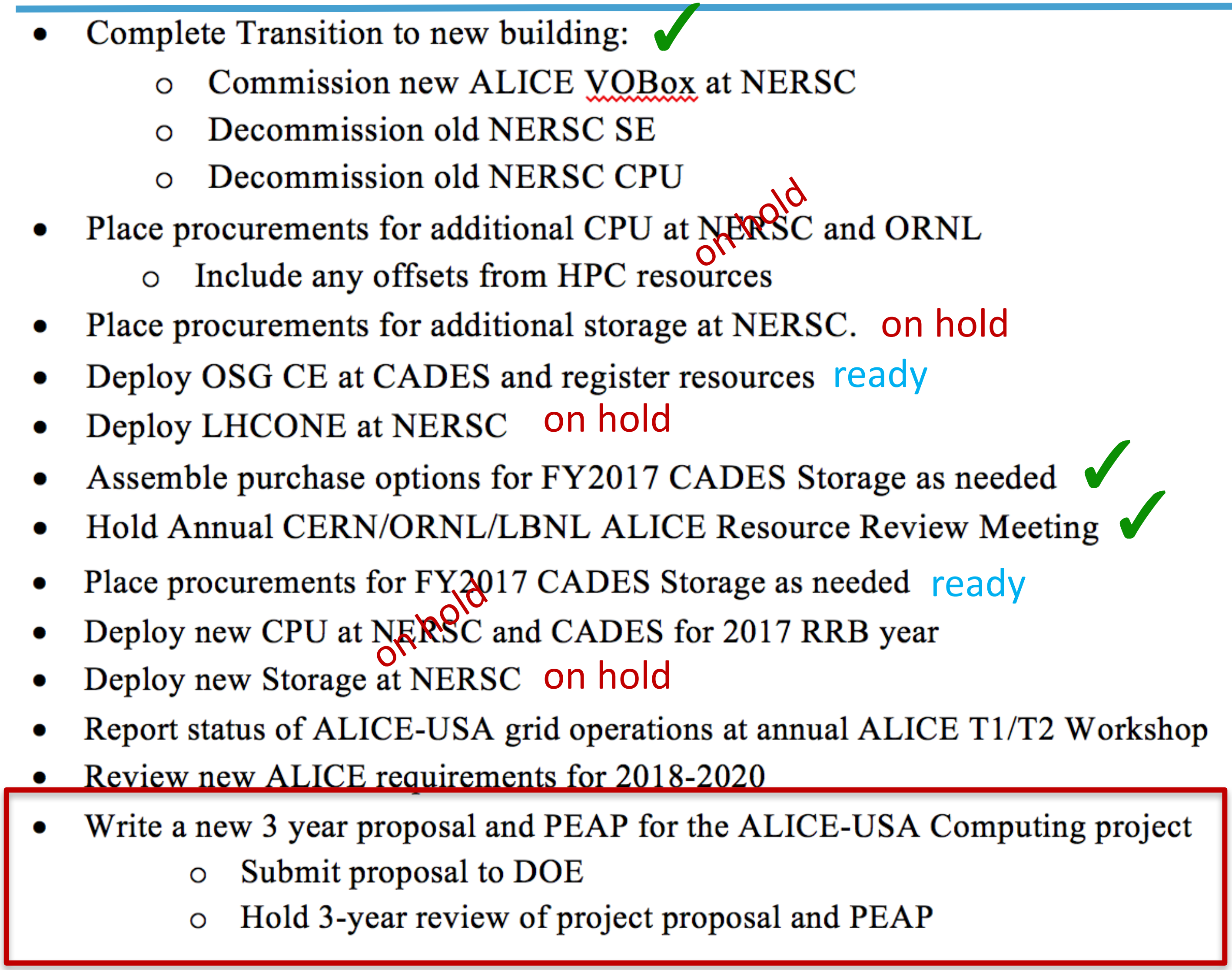
2017 PEAP Plan – Hardware

- **Targets:**
  - 100% CPU on time
  - 100% Disk lags with utilization

Resource	Installed/FY16	FY2017 Apr. 2016	FY2017	FY2018
ALICE-USA Obligations				
CPU (kHS06)	28.4	37.3	46.7	55.8
Disk (PB)	3.2	4.0	4.0	5.3
ALICE-USA Plan				
CPU (kHS06)	29.0	37.5	46.5	54.0
% CPU obligation	102%	100%	100%	97%
Disk (PB)	2.3	3.2	3.2	4.4
% Disk obligation	72%	80%	80%	80%
Disk deficit (PB)	0.8	0.8	0.8	1.1



Project Tasks from 2017 PEAP



- Complete Transition to new building: ✓
  - Commission new ALICE VOBox at NERSC
  - Decommission old NERSC SE
  - Decommission old NERSC CPU
- Place procurements for additional CPU at NERSC and ORNL on hold
  - Include any offsets from HPC resources
- Place procurements for additional storage at NERSC. on hold
- Deploy OSG CE at CADES and register resources ready
- Deploy LHCONE at NERSC on hold
- Assemble purchase options for FY2017 CADES Storage as needed ✓
- Hold Annual CERN/ORNL/LBNL ALICE Resource Review Meeting ✓
- Place procurements for FY2017 CADES Storage as needed ready
- Deploy new CPU at NERSC and CADES for 2017 RRB year on hold
- Deploy new Storage at NERSC on hold
- Report status of ALICE-USA grid operations at annual ALICE T1/T2 Workshop
- Review new ALICE requirements for 2018-2020
- Write a new 3 year proposal and PEAP for the ALICE-USA Computing project
  - Submit proposal to DOE
  - Hold 3-year review of project proposal and PEAP