LQCD-ext II Project
DOE Review Q6 3-Site Plan

Robert D. Kennedy
Associate Contractor Project Manager

DOE FY2016 Annual Progress Review
Thomas Jefferson National Accelerator Facility
June 28-29, 2016

Question 6

- Describe the Alternatives Analysis and Risk Analysis for the 3-Site Plan. Show the process that took place.

CR16-01: Outline

- Identify Gap in Delivered Computing due to Third Site (*these slides*)

  ◦ Define the Alternatives

  ◦ Prepare a Cost Model/Forecast for each Alternative

  ◦ Prepare a Performance Forecast for each Alternative based on Acquisition Budgets

  ◦ Focus on the alternative with the most desirable characteristics

    • Desirable to split computing acquisition funds roughly equally among sites: <u>3-Sites FY-Straddle</u>

      • Only alternative with this feature… balance by tweaking 1/3-2/3 funds split in FY17 and FY18

- Determine how to Fill Gap (*See CR16-01 3-Sites talk*)

  ◦ What is required for hosting LQCD clusters:

    • Facilities, compute, storage, network, user support, etc.

  ◦ Negotiate with BNL for an agreement to address these requirements AND fill the gap

  ◦ Prepare an MOU to ensure each party has spelled out what it needs in this agreement

- Risk Analysis: Common and Specific Items

CR16-01: Alternatives Analysis

▸ <u>Models Compared</u>: to track updated info *and* impact of 3 vs 2 Sites

- ◦ <u>Baseline</u>: Exactly the output of the project CD process.

- ◦ <u>Reference</u>: "Baseline" with FY15 actuals included.

- ◦ <u>2-Sites</u>: "Reference" with all changes to future plans included, such as overhead rates.

  - This is the updated and improved forecast of the project if it proceeds *without* CR16-01.

- ◦ <u>3-Sites FY-Aligned</u>: Adds cluster-hosting at third site using a one-site-per-year rotation

- ◦ <u>3-Sites FY-Straddle</u>: Adds cluster-hosting at third site using "3-sites in 4 years" rotation

Gap

▸ <u>Acquisition Plans in These Models:</u>

Plan Name	FY16	FY17 procurement	FY18 procurement	FY19 procurement
Baseline	JLab	JLab (FY16 options)	FNAL	FNAL (FY18 options)
Reference	JLab	JLab (FY16 options)	FNAL	FNAL (FY18 options)
2-Sites	JLab	JLab (FY16 options)	FNAL	FNAL (FY18 options)
3-Sites FY-aligned	JLab	BNL	FNAL	JLab
3-Sites FY-straddle	JLab	1/3 JLab (FY16 options); 2/3 BNL	2/3 BNL (FY17 options); 1/3 FNAL (slide to FY19)	FNAL

- *3-Sites FY-Straddle is the proposed Acquisition Plan going forward with this CR.*

Alternatives: Model -> Cost Forecasts

Cost Forecasts

- 3-Site FY-Straddle, just shows *detail level treated*

- (*Top*): Project Budget Detail
  - Sheets for Sites, Mgmt Reserve

- (*Bot*): Compute Funds Split
  - Excerpt from Staffing Model

- Staffing Model captures costs of cluster hosting, based on 10+ years experience.

- Broader "budget" to better forecast performance
  - Treats past project carry-over with this project's funding.
  - So, Total Budget Profile looks different here from ext II Project Funding profile.

- Cost Forecast then drives the Performance Forecast via the Acquisition Budget.

BUDGET ($K)

	(closed) FY15	(allocated) FY16	FY17	FY18	FY19	Total
Steady-state Operations						
Personnel	1,543,737	1,451,490	1,543,813	1,447,666	1,408,561	7,395,267
Travel	11,000	11,000	11,000	11,000	11,000	55,000
M&S (hardware, repairs, etc.)	281,000	297,786	146,000	120,000	120,000	964,786
Sub-total (SS Ops)	1,835,737	1,760,276	1,700,813	1,578,666	1,539,561	8,415,053
New Hardware Deployment						
Personnel	-	198,800	296,144	141,354	244,374	880,673
Travel	-	-	-	-	-	-
Equipment (compute)	495,000	992,175	719,248	1,008,658	896,203	4,111,284
Equipment (storage)	60,000	50,359	39,452	55,327	77,931	283,069
Sub-total (New Deployment)	555,000	1,241,334	1,054,844	1,205,339	1,218,508	5,275,026
Project Management						
Personnel	110,298	127,351	131,172	135,107	139,160	643,089
Travel	6,000	6,000	6,000	7,000	7,000	32,000
M&S	2,000	2,000	2,000	2,000	2,000	10,000
Sub-total (Project Mgmt)	118,298	135,351	139,172	144,107	148,160	685,089
Total Project Cost						
Personnel	1,654,035	1,777,641	1,971,129	1,724,127	1,792,095	8,919,028
Travel	17,000	17,000	17,000	18,000	18,000	87,000
M&S	283,000	299,786	148,000	122,000	122,000	974,786
Equipment (compute)	495,000	992,175	719,248	1,008,658	896,203	4,111,284
Equipment (storage)	60,000	50,359	39,452	55,327	77,931	283,069
Management Reserve	45,964	83,039	105,173	71,890	93,767	399,833
Total	**2,555,000**	**3,220,000**	**3,000,002**	**3,000,002**	**2,999,996**	**14,775,000**
CD-2/3 Budget Guidance Profile	2,000,000	3,000,000	3,000,000	3,000,000	3,000,000	14,000,000
Carry-over Funds from Past Project	555,000	220,000	-	-	-	775,000
Total Budget Profile	2,555,000	3,220,000	3,000,000	3,000,000	3,000,000	14,775,000

CPU Funds Split	FNAL	JLab	BNL
FY15	0	0	0
FY16	0	1	0
FY17	0	**0.30**	**0.70**
FY18	**0.30**	0	**0.70**
FY19	1	0	0
Total Compute H/w	**$1,198,801**	**$1,207,949**	**$1,209,534**

Alternatives: Costs -> Performance Forecasts

Performance Forecast

- 3-Site FY-Straddle, just shows *detail level treated*

- Extends forecast model used in Project CD process + cluster ratings updates

- Docs expected life cycle for clusters, and KPIs calculations (separate sheet)

- <u>Future cluster performance</u>: Baseline scaled by Acquisition Budget change

- Calculates min integral BNL-IC nodes needed to maintain total Delivered Computing

3-Site FY Straddle Scenario with BNL IC					FY15			FY16			FY17			FY18			FY19			FY15-FY19
6/10/16		Compute Equip Buy Only																		
Machine	Base TFlops	Base $	Alt $	Alt TFlops	Fraction	Uptime	Delivered	Fraction	Uptime	Delivered	Fraction	Uptime	Delivered	Fraction	Uptime	Delivered	Fraction	Uptime	Delivered	Aggregate
DD2	4.38	1	1	4.38	1.00	91.3%	4.38	1.00	91.3%	4.38	1.00	91.3%	4.38	0.00	91.3%	0.00	0.00	91.3%	0.00	13.14
BG/Q	21.90	1	1	21.90	1.00	91.3%	21.89	1.00	91.3%	21.89	1.00	91.3%	21.89	0.00	91.3%	0.00	0.00	91.3%	0.00	65.68
Ds	21.03	1	1	21.03	1.00	91.3%	21.02	0.95	91.3%	20.00	0.00	91.3%	0.00	0.00	91.3%	0.00	0.00	91.3%	0.00	41.03
10g / 11g - average	17.09	1	1	17.09	0.75	91.3%	12.81	0.75	91.3%	12.81	0.00	91.3%	0.00	0.00	91.3%	0.00	0.00	91.3%	0.00	25.63
Bc	12.73	1	1	12.73	1.00	91.3%	12.73	1.00	91.3%	12.73	1.00	91.3%	12.73	1.00	91.3%	12.73	0.00	91.3%	0.00	50.91
12s	12.80	1	1	12.80	1.00	91.3%	12.80	1.00	91.3%	12.80	0.00	91.3%	0.00	0.00	91.3%	0.00	0.00	91.3%	0.00	25.59
Pi0	13.10	1	1	13.10	1.00	91.3%	13.10	1.00	91.3%	13.10	1.00	91.3%	13.10	1.00	91.3%	13.10	1.00	91.3%	13.10	65.48
Dsg	15.90	1	1	15.90	1.00	91.3%	15.90	0.49	91.3%	7.80	0.00	91.3%	0.00	0.00	91.3%	0.00	0.00	91.3%	0.00	23.70
12k	26.40	1	1	26.40	1.00	91.3%	26.39	1.00	91.3%	28.11	0.00	91.3%	0.00	0.00	91.3%	0.00	0.00	91.3%	0.00	54.50
Pi0g	25.10	1	1	25.10	1.00	91.3%	25.09	1.00	91.3%	25.09	1.00	91.3%	25.09	1.00	91.3%	25.09	1.00	91.3%	25.09	125.47
15C - Pi0 Expansion	6.12	1	1	6.12	0.46	85.0%	2.61	1.00	91.3%	6.12	1.00	91.3%	6.12	1.00	91.3%	6.12	1.00	91.3%	6.12	27.09
15G - NONE	0.00	1	1	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00
16C	10.30	$389,523	$496,088	13.12	0.00	0.0%	0.00	0.00	85.0%	0.00	1.00	91.3%	13.11	1.00	91.3%	13.11	1.00	91.3%	13.11	39.34
16G	38.70	$389,523	$496,088	49.29	0.00	0.0%	0.00	0.00	85.0%	0.00	1.00	91.3%	49.27	1.00	91.3%	49.27	1.00	91.3%	49.27	147.82
17C	13.80	$512,235	$359,624	9.69	0.00	0.0%	0.00	0.00	0.0%	0.00	0.22	85.0%	1.98	1.00	91.3%	9.69	1.00	91.3%	9.69	21.36
17G	52.00	$512,235	$359,624	36.51	0.00	0.0%	0.00	0.00	0.0%	0.00	0.22	85.0%	7.48	1.00	91.3%	36.50	1.00	91.3%	36.50	80.47
18C	28.20	$534,177	$504,329	26.62	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.34	85.0%	8.30	1.00	90.8%	26.46	34.76
18G	106.10	$534,177	$504,329	100.17	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.34	85.0%	31.23	1.00	90.8%	99.56	130.79
19C	36.10	$685,152	$448,102	23.61	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.25	85.0%	5.49	5.49
19G	136.20	$685,152	$448,102	89.08	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.00	0.0%	0.00	0.25	85.0%	20.73	20.73
BNL Inst Cluster	0.00	$0	$0	36.00	0.00	0.0%	0.00	0.33	100.0%	12.00	1.00	100.0%	36.00	1.00	100.0%	36.00	1.00	100.0%	36.00	120.00
	597.9515		$3,616,286	560.636	169.0			175.2			191.9			244.1			343.9			1118.99
Delivered Computing Total			-$625,889		176.6	95.7%	168.7	212.6	82.4%	176.8	227.9	84.2%	191.2	328.4	74.3%	241.1	428.4	80.3%	341.1	1117.11
Conventional						88	88.53		102	91.02		83	73.31		72	63.04		82	73.97	1.88
GPU			rating/node	0.9	Tflops/node	92	80.20		102	85.82		126	117.84		186	178.10		275	267.15	0.2%
			n_nodes	40		180	168.72		204	176.84		209	191.16		258	241.14		357	341.12	Same CPU d
			total nodes	200	Delivered	PEP Set Values			PEP Rounded Values			PEP Rounded Values			PEP Rounded Values			PEP Rounded Valu		BNL IC time

Alternatives: Filling the Gap

- *See "Calculation of BNL-IC Allocation"*
- Gap: Difference in total Delivered Computing between current forecast of 2-Site alternative versus 3-Site FY Straddle alternative.
- Based on the estimate of BNL-IC ratings (similar to Pi0g), we determined the number of BNL-IC nodes needed to fill the gap = 40.

Delivered Computing	FY15	FY16	FY17	FY18	FY19	Project Total
2-Site	168.7	164.8	181.5	206.3	395.8	1117.1
3-Site FY Straddle	168.7	164.8	155.2	205.1	305.1	999.0
GAP between 2&3 Sites	0.0	0.0	-26.3	-1.1	-90.7	-118.1
BNL IC Allocation	0.0	12.0	36.0	36.0	36.0	120.0
GAP adding BNL IC	0.0	12.0	9.7	34.9	-54.7	**1.9**
Notes	Assumes BNL IC available June 1, 2016 = 1/3 of FY16					
Rating of BNL IC Node	0.900	[Tflop/s-year] Estimate based on Pi0g experience				
Allocated BNL IC Nodes	40	Time-averaged over a month				

3–Sites: Risk Analysis – Common Risks (selection)

▸ **Risk 01: Technology/systems may take longer than expected to become available**

  ◦ We tolerate a bit of delay and/or reduced uptime in our forecast of new system performance.

▸ **Risk 07: Host institutions do not provide necessary infrastructure**

  ◦ Cluster hosting site architects developed a requirements list, checked against plan.

▸ **Risk 38: Inaccurate Storage Forecasting**

  ◦ We are not just putting Compute Cycles on the floor.

▸ **Risk 41: Software infrastructure may not be mature enough for newer computing architectures**

  ◦ Users test, site architects test, and results are rolled into alternatives analysis.

  ◦ Non-factor for K80's used in BNL-IC

▸ **We also have an Acquisition Process that addresses common risks in all acquisitions, and includes schedule and budget contingencies, in case.**

3-Sites: Risk Analysis – Specific to This Case

- Note: Many elements of cluster hosting have been addressed (*see BNL Overview Talk*)
  - Computing facilities update, performant storage in plan, BNL-IC specs good. We judged these risks to be low after discussions LQCD-BNL.

- Late Delivery of BNL-IC: common to all acquisitions
  - We can tolerate some delay if we can make up the 40 node average later
  - Performance forecast already assumes less "effectiveness" in new clusters

- Poor BNL-IC Rating Estimates: Project fails to fill the gap
  - BNL-IC nodes are similar to existing production LQCD cluster (Pi0g)
  - Re-evaluate BNL-IC node rating once in production (MOU T&C)

- User Support: from a few LQCD users to many LQCD users
  - Work on Self-Help via Documentation (User Survey action item)
  - Leverage ServiceNow tool to track incidents, requests for help
  - Ramp Up Users over time from early users to Sep 1 Allocations begin.