DOCUMENT NUMBER

Response to 2017 S&T recommendation for Experimental Computing

June 2018

Amber Boehnlein

Graham Heyes

Chip Watson

2017 S&T Recommendation: Generate a cost-effective plan to ensure sufficient computing resources for data analysis and simulation in FY18 and a longer-term approach to address the needs in FY2019 and beyond. The plan for FY2019+ resources should include a time line and detailed plan to evaluate the feasibility of the proposed approach, including off-site computing resources, such that the plan can be place and tested before the FY2019 running. Synergy with the theory computing needs should be considered.

In response to the above 2017 S&T review recommendation, this note describes the estimated computing hardware needs for the Jefferson Lab 12 GeV Experimental Program for FY19-FY21 and includes a description of possible mechanisms for filling those needs. This document is an updated version based on the preliminary report from February 2018. This version includes updated input from the Halls, including estimates from CLAS12 based on running experience. Our intent is to provide on-site computing resources that keeps pace with the yearly production needs, supplemented with resources from the Open Science Grid (OSG) and NERSC for GlueX. Testing commercial cloud resources at scale is postponed until FY19Q3.

Based on the analysis of the projected needs and the current usage of the existing farm and disk space, we make following recommendations

1. We recommend spending $600K on farm nodes in FY18 to augment the current farm for the Fall 19 run. This will accommodate the basic calibration and reconstruction needs for GlueX and CLAS12 and analysis needs. It will not accommodate multiple re-reconstruction passes or large-scale Monte Carlo (MC) production. We recommend continued investment of at least $300K/year to replace obsolete nodes, establish capacity for analysis and support the potential growth of Halls A&C.

2. We recommend spending $100K in FY18 on disk space to increase the disk buffer space to 3PB to accommodate the planned CEBAF schedule, and an additional $100K in FY19 to increase the disk buffer to support the use of offsite computing resources.

Computing Models and Needs Estimations

In general, the computation flow is similar for all of the current Jefferson Lab experiments. Most data are taken during accelerator running periods, requiring rapid turn-around for data quality monitoring and preliminary calibrations. Experimental data is reconstructed from hits into physics quantities as a production task, and then skimmed into data sets that are appropriate for user level analysis. The skimmed analysis data sets can be analyzed at Jefferson Lab or transferred to university clusters. The generation of Monte Carlo (MC) samples is a large-scale computationally intensive production task that is more closely tied to the analysis cycle than to experimental data collection.

This document, covering FY19-FY21, is concerned with provisioning for production tasks for CLAS12 and GlueX, which tend to drive the requirements in these early years of 12GeV CEBAF. The estimates of these needs depend on the accelerator schedule, experiment specific considerations (such as event size, event rate and reconstruction time), assumptions of the number of MC events needed and the number of reconstruction passes. For this assessment, we used assumptions for slightly relaxed basic needs to establish a realistic set of requirements within a realistic budget. For FY19, the number of weeks of running and rates correspond to the published CEBAF schedule and PAC approved experiments. There are risks associated with the estimating process—while the experiments provide their best understanding of the needs, it is important to remember that they are estimates and, in some cases, aspirational—less or more MC may be needed, the reconstruction times may be longer, a major software improvement may require an additional reconstruction pass. Table 1 lists the assumptions with CLAS12 reconstruction is given as a range to accommodate the uncertainty in their reconstruction estimates.

		FY18	FY19	FY20
GlueX	Weeks of running	22	18	28
	Data rate - Mbyte/s	800	800	1500
	Event rate - kHz (thousands of events/sec)	50	50	90
	Reconstruction cpu - ms/event	137	137	137
	MC per cpu - ms/event	352	352	352
	Ratio of MC to raw events	2	2	2
CLAS 12	Weeks of running	22	28	28
	Data rate - Mbyte/s	600	600	600
	Event rate – kHz (thousands of events/sec)	12	12	12
	Reconstruction cpu - ms/event	1400 to 350	350	350
	MC per cpu - ms/event	470	470	470
	Ratio of MC to raw events	6	6	6

Table 1 Assumptions impacting computing estimates

Table 2 shows the computing requirements for FY18 through FY20 in millions of core hours (M-ch) normalized to 2.3 GHz Intel Xeon E5-2697 v4 (Broadwell) processors. FY20 is assumed to be the start of steady state so that FY21 should be similar to FY20. As a consequence of the uncertainty in the reconstruction time, the FY18 CLAS12 non-MC computing requirements are shown for times of 1400 and 350 ms per event. In FY19 it is assumed that the CLAS12 reconstruction will be improved. The FY19 schedule has 28 weeks of accelerator running compared with 22 in FY18. Both the CLAS12 and GlueX CPU requirements are dominated by event reconstruction and simulation (MC). As a point of reference, the local resource currently installed at Jefferson Lab can deliver 37M-ch/year, will be upgraded to 72M-ch/yr by the start of FY19, and can be supplemented in FY19 by an end of life LQCD cluster contributing an additional 15M-ch/yr.

	FY18 - 1.4	**FY18 - 0.35**	**FY19**	**FY20**
GlueX MC	95	95	78	218
CLAS12 MC	84	84	107	107
GlueX Non-MC	39	39	32	90
CLAS12 Non-MC	102	26	33	33
Total MC	179	179	185	325
Total Non-MC	141	65	65	123
Total All	320	244	250	448

Table 2 - CLAS12 and GlueX tasks (M core hours), 2.3 GHz Intel Xeon E5-2697 v4 (Broadwell) processor.

This exercise produces an average of the computing core hours that would be required to process a CEBAF year worth of data in a calendar year. To the extent possible, we benchmark estimates against the actual usage. For this exercise, the production needs of Hall A, non-CLAS12 Hall B and Hall C experiments have not been estimated. In Figure 3, it can be seen that their current yearly usage is 30% of the existing farm. We assume that level will remain similar for the next few years.

Additionally, as the estimates are tied to the number of weeks of running to calculate average yearly computing needs, as CEBAF adds weeks of running to the schedule, there will be increased computing and disk needs. As a note, the running period in FY18 changed from 10 weeks to 22 weeks between versions of this document.

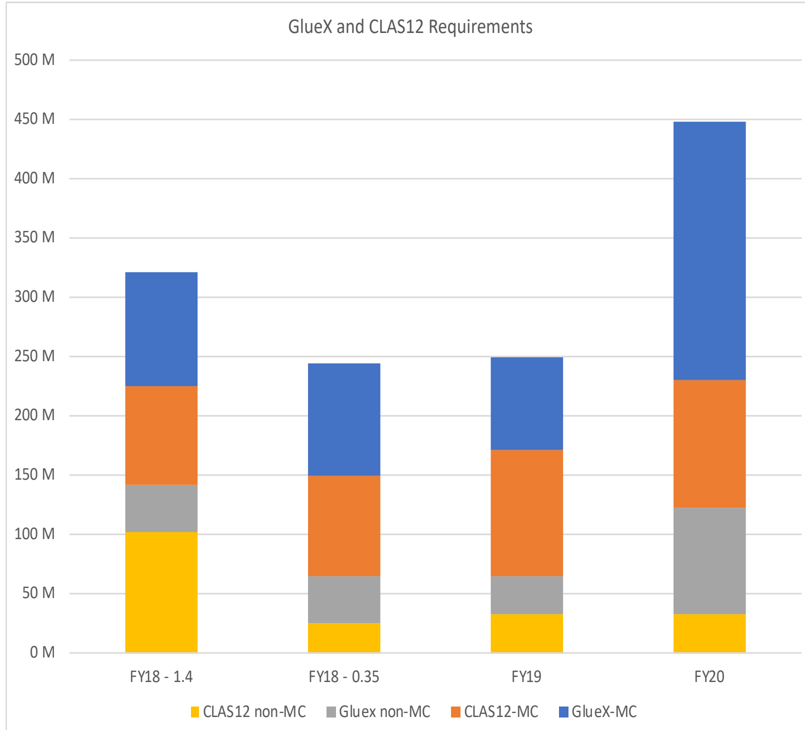

Figure 1 - CLAS12 & GlueX compute requirements, in millions of core hours.

GlueX will run in the Fall of 2018 to complete GlueX phase-I with the same rates as Spring 18 run. The Spring 19 run will be dedicated to the PRIMEX experiment with much lower rates. In FY20 GlueX will begin phase-II running with an anticipated raw data rate at or less than 1.5 Gbyte/s with a 60% duty factor while the accelerator is running. When the output of reconstruction is included this rate would require 0.7PB of mass storage capacity per week of running. CLAS12 is expected to run at a lower data rate and contribute another 0.3 PB/run week. The sum of the two halls would require mass storage of 1 PB per week of beam time.

A fraction of the data must reside on disk where it is accessible for processing. This fraction varies with reconstruction and analysis workflows and has historically been 10%, however recent experience justifies at least 20%. Thus, 20 weeks of running would require access to 2-4 PB of disk where 1 PB is currently provisioned. To meet this need, an additional $50K to $150K will be needed in FY19. In FY2018, bandwidth in and out of the tape library is being doubled from 2 GB/s to 5 GB/s.

We conclude based on the estimated requirements and current usage of the farm, it is essential to purchase compute nodes in summer FY18 for late summer/early fall 18 run and we estimate that $600K will be sufficient to meet the basic needs for non-MC, assuming that a lower CLAS12 reconstruction time is achieved by the end of calendar 2018. Starting in FY19, we estimate that $300K/year will enable replacement of obsolete farm nodes and will provide for modest growth of the farm. We will revisit the estimates by November, 2018 in preparation for an upcoming review.

Mechanisms for fulfilling the computational requirements

In this section, we describe the on-site resources and several different types of distributed resources. A cost comparison between on-site and AWS is presented.

Local Resources:

Historically, resources for production computing for the experimental program were onsite at Jefferson Lab without support for distributed computing, beyond the ability for collaborators to transfer data in and out of the site. Synergy between the LQCD program and the experimental program provided a sharing arrangement of resources that ensured excellent steady state utilization of the compute resources. While we intend to continue providing on-site compute in a sharing arrangement with LQCD, it is also prudent to be able to leverage external resources to handle peak demands and to prepare for the future in which cloud resources are likely to play a prominent role.

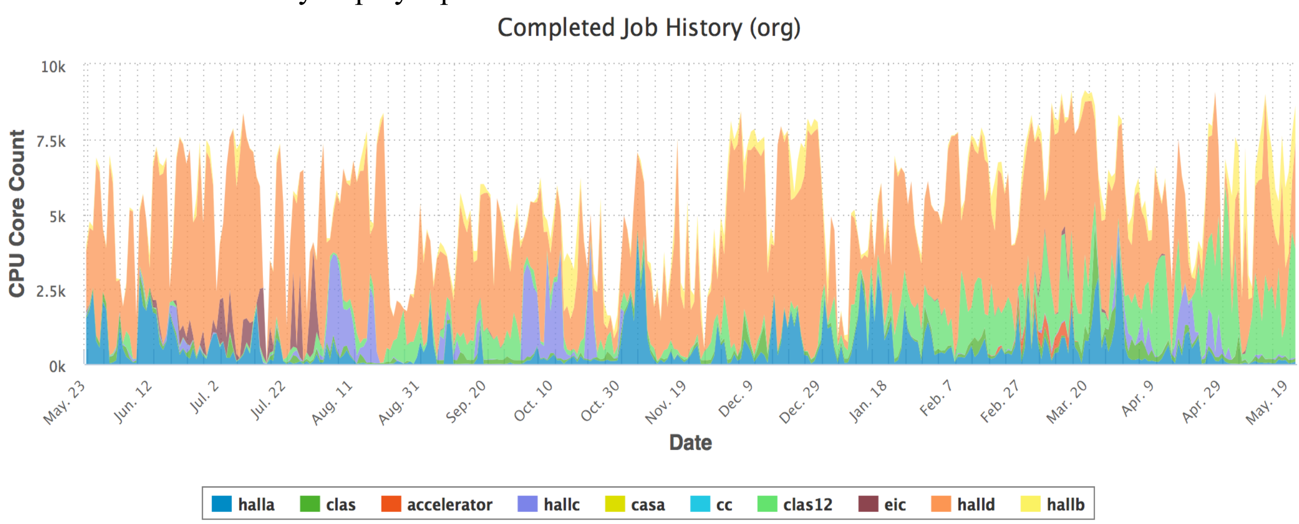

Figure 2 Job History of the Scientific Computing Farm, GlueX/Hall D is the dominant user. ' Hallb' is non CLAS HallB users.

The current episodic nature of the use of resources is shown in Figure 2 above. This snapshot shows the period from May 2017 to current and is dominated by GlueX (halld). For reference, the number of total job slots is roughly 10,000 corresponding to 5000 physical cores. Above 5k jobs, 85% of the CPU performance is in use (a job on every core), and at 7.5k slots used, 93% of the possible performance is used.

For Fall 2018 and 2019, assuming good CEBAF performance, the current system will be inadequate to finish both GlueX and CLAS-12 reconstruction within a year, an observation consistent with the needs estimates, motivating an upgrade.

In Figure 3 (below left), the distribution of jobs by user group is shown for calendar year 2017. The farm utilization averages 80%. The Hall D usage for non-MC over the year is consistent with needs estimates presented above in Figure 1. Figure 4 (below right) shows the distribution of jobs by user group for CY2018 to date, showing that the usage by CLAS12 is increasing relative to CY2017.

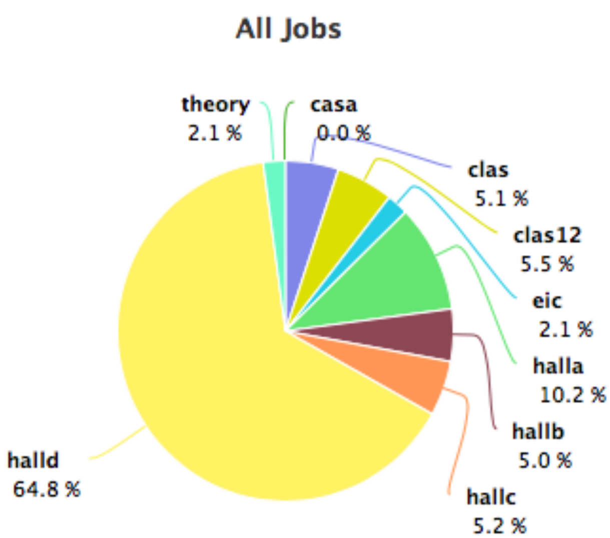

Figure 3 Distribution of Jobs by Group, 2017

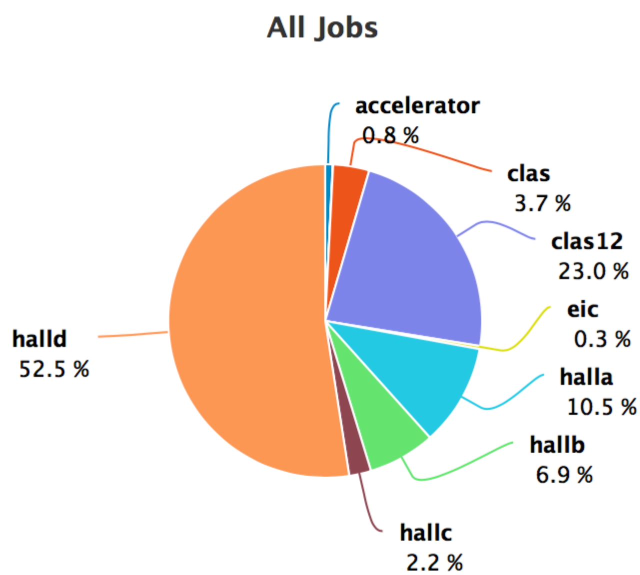

Figure 4 Distribution of Jobs for 2018 YTD

We expect machine failures begin around 6 years of service life and budget accordingly for replacements, estimated to be $200K/year. An additional $100K/year would allow the size of the compute resource to grow slowly in anticipation of the growing volume of data, e.g. for re-reconstruction.

Year	Node Count	Cores/node	GHz	Relative performance	Normalized cores
2012	36	16	2.0	0.6	346
2013	24	16	2.6	.85	326
2014	104	24	2.3	1.0	2496
2016	46	36	2.3	1.1	1820
				TOTAL	4988

Table 3 Summary of existing Jefferson Lab Experimental Physics Computing Resources

In July of this year, this farm resource and an end-of-life LQCD 240 node cluster identical to the 2012 nodes will be combined into a single lab resource used by experimental physics and LQCD on a fair share basis, with experimental physics using 100% of the resource during its peaks. This increases the local resource by 46% to 7.3K 2014 cores, but will still fall short of the target of 10K cores. Given the age of these nodes, they are best viewed as providing an additional short-term buffer as experience is gained with 12 GeV data, and as a resource to offset the current reconstruction time of CLAS12 events. CLAS12 is implementing code speed-ups, which should compensate as these 2012 nodes die off.

In early August, an additional ~3.3k cores will be added, bringing the total capacity to ~ 10.6k cores for the Fall running period, dropping down to ~9.5k cores as the 2012-13 nodes retire and a $300k upgrade is done.
Working disk space will be increased in June by 69% (from 0.65 PB to 1.1 PB), enough to support the July and August increases in capacity, but not enough to also support a large offsite resource. A planned disk expansion in FY19 will come online as the accumulated data grows beyond 10 PB.

	Current	Start of FY2019	Start of FY2020*
M-core-hours/year	37	87 (17 end of life)	90
Disk (scratch+cache) PB	0.65	1.1	2.0
Tape GB/s	3	5	7
WAN bandwidth, Gbps	10	10	10

*Values to be adjusted as additional insights into requirements are gained. Additional offsite resources will be added.

Table 4 Evolution of Jefferson Lab Experimental Physics Computing Resources

The trend is that computing needs in the future are likely to be supplied by a mix of on-site, commercial clouds, academic clouds and major compute center resources. Using distributed resources in FY19 gives us experience with this model and flexibility in the use of funds in the early part of the 12 GeV program.

Distributed Resources: The potential to use distributed resources comes in several varieties including arrangements within a collaboration to use resources at home institutions via privately maintained mechanisms. Distributed computing is not a panacea. Adding distributed capability effectively 'scales-up' the compute resources, requiring a scaling up of the storage infrastructure similar to what an increase in onsite compute resources would require. It can also be labor intensive to support. The current wide area network (WAN) connectivity (10 Gb/s) would allow using offsite resources of scale the same as the current onsite resources in a naïve way. For the longer term, the ESNet roadmap includes 100Gb/sec connectivity upgrade to Jefferson Lab in 2020.

Carefully choosing which jobs run offsite can yield even higher gains in a hybrid onsite/offsite computing model before WAN bandwidth is exhausted. Thus, for practical reasons, Monte Carlo is the most tractable starting point in gaining experience with distributed computing. These practicalities include lower data transfer rates, more forgiving time requirements, and more predictability in execution. Below, we outline the offsite options that we are investigating and the current state.

Open Science Grid: The Open Science Grid (OSG) is a consortium of academic sites that contribute compute resources to a distributed pool. OSG does not directly control the scheduling of the local resources, however, OSG maintains and supports the software necessary to provide this distributed compute environment. OSG has been in production for more than 15 years and is a critical part of the US based computing for the Large Hadron Collider.

Benefits: Jefferson Lab users often have access to computing resources on campus, and OSG provides a mechanism in some cases to utilize those campus resources for production tasks. The GlueX collaboration has been an advocate for using OSG for many years.

What has been done: A gateway node for OSG job submission is now hosted at Jefferson Lab. GlueX now has three university sites on OSG. GlueX is routinely running MC jobs on OSG, and reports excellent turn-around time at the scale of resources needed by GlueX. Additionally, a GlueX collaborator has been testing the use of data access methods used by LHC that would also enable GlueX reconstruction and other data intensive tasks on OSG. The GlueX status as of March 2018 was presented at the OSG All-Hands meeting. [https://docs.google.com/presentation/d/1XjWI4PpQXl4eJX2KhUb4TnlQ8dMQ1oby3Ki7Du1t5nw/edit#slide=id.g359d5ff226_0_92] Recent experience on OSG for GlueX is shown in Figure 5. (credit R. Jones, UCONN). In 10 days in May 2018, approximately 1M-ch were delivered for GlueX Monte Carlo production.

What we have learned: Support for OSG will be at a minimum, however, the stack has to be maintained for LHC experiments, and thus the resource should remain available until 2021 (Wuerthwein, 2018). OSG will not be an option for Monte Carle production for CLAS12, Moller or SoLID. OSG has demonstrated the power of aggregating distributed resources for large computing needs and it seems likely that some other distributed computing platform will be available. A discussion of a possible deployment of an Office of Science Federated Resource was presented at the ASCAC meeting in April. [https://science.energy.gov/~/media/ascr/ascac/pdf/meetings/201804/ASCAC-FLC-WG-201804.pdf].

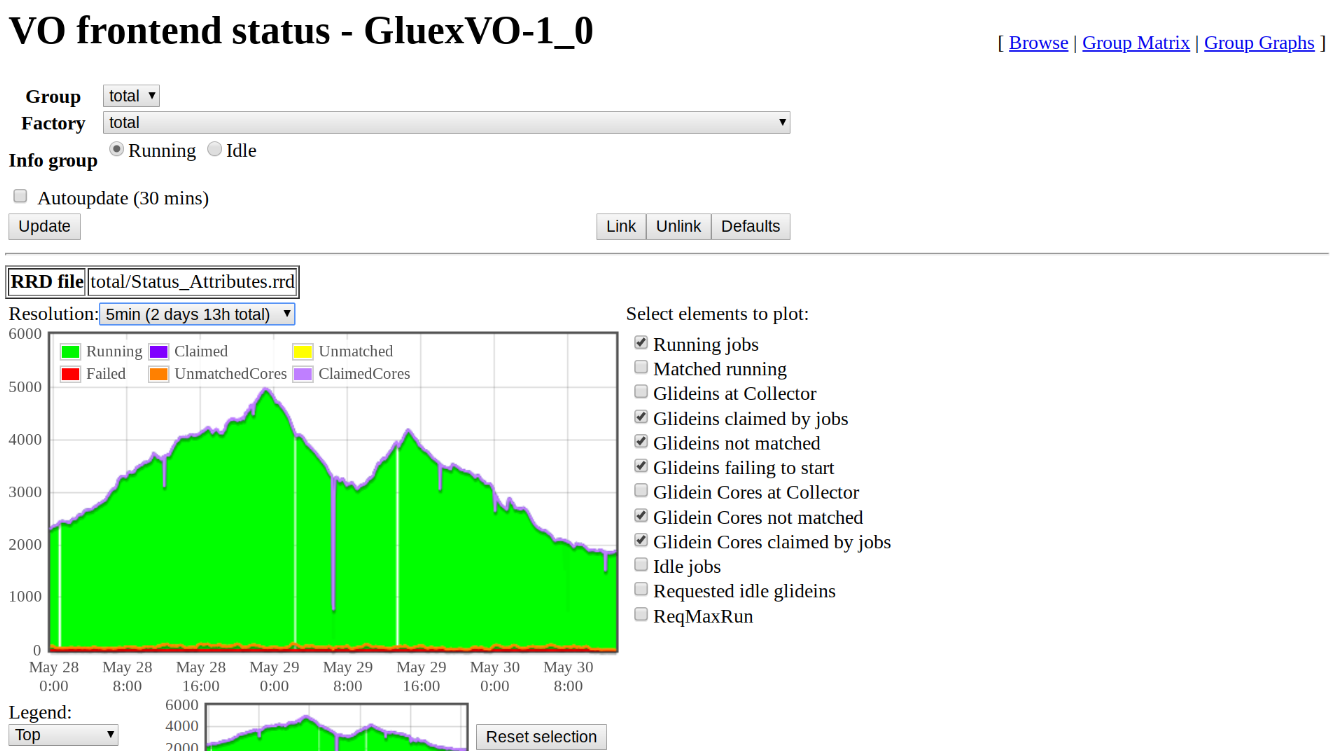

Figure 5 GlueX Usage of OSG. 1M-ch were delivered in May

Commercial Cloud (AWS): A number of companies provide computing resources 'for hire'. This can provide an exceptionally good mechanism for provisioning for short-term peak needs and is also useful for intermittent tasks as an alternative to dedicated machines. The current pricing structure varies from provider to provider.

Benefits: While not cost effective today for steady state processing, the prices continue to fall. Advantages include access to diversity of hardware or software platforms that we would not be able to afford to maintain onsite. The resources are flexible and elastic, with the ability to provision a 'virtual data center' quickly and then tear it down when not needed. Each experiment could provision the 'virtual data center' to their particular needs with limited central oversight, provided they stayed within their budget.

What has been done: We have chosen to start with Amazon Web Services (AWS) for Scientific Computing. While there are other providers, the barrier to entry is relatively low and AWS is experienced in working with the HEP community. AWS has been on site several times to work directly with interested parties. Procurement and billing mechanisms have been set up, including alarms to prevent inadvertent overspending. A number of small-scale projects have been using AWS at an individual level.

What we have learned: Persistent server instances in the cloud currently cost about twice what in-house resources cost (i.e. comparing core-hour cloud costs to annual in-house investments +operating costs). Transient (spot) instances can approach local costs, but these instances can be pre-empted, and costs vary significantly over time of day and day of week. The low spot pricing could be advantageous, and we need to learn how to acquire resources at low spot prices while releasing them above a higher threshold while still getting high weekly or even daily throughput. Advantages include faster turnaround during heavy loads (higher productivity), and the cost savings of not overprovisioning helps to offsite the higher per core-hour cost. Understanding the magnitude of our peaks and valleys will be critical in provisioning decisions.

What remains to be done: In order to demonstrate viability for production tasks, an advocate collaboration (similar to GlueX and OSG) has to be identified, and a project plan and budget developed.

ASCR/XCEDE Facilities:

Another option for offsite computing is the use of the ASCR facilities such as the Leadership Class Facilities (LCF) and National Energy Research Supercomputing Center (NERSC) and possibly the NSF XCEDE centers. The HEP/LHC community has made good use of these resources, particularly for Monte Carlo generation, and the Astronomy and Light Source communities have been using them for experimental data analysis. NERSC, in particular, has been charged by ASCR to lead in developing 'Super Facilities' to collaborate with 'SC user facilities to enable seamless and high performing end to end workflows' [https://people.eecs.berkeley.edu/~yelick/talks/data/Superfacility-ASCAC16.pdf].

Benefits: Fully realizing the 12 GeV program will require a 'Super Facility' approach with workflows that seamlessly combine theoretical calculations with the analysis of data from multiple experimental sources.
What has been done: An LDRD was awarded to refresh the multi-threaded processing framework (JANA) used by GlueX. As part of this LDRD, some GlueX data will be reconstructed at NERSC, and several pilot jobs have been run. Subsequently, a 50M NERSC MPP hour allocation has been made from the NERSC Director's reserve to scale testing up to a production level. The scale of the FY2020 requirements above the planned local resources described above is 30M Broadwell core hours, equivalent to 67M NERSC MPP hours. This activity is in the active prototyping stage, with work in progress by GlueX, NERSC and Jefferson Lab IT division, with targeting a reconstruction data challenge in FY18Q4 similar to STAR data reconstruction.
[https://insidehpc.com/2018/02/reconstructing-nuclear-physics-experiments-supercomputers/]

What we have learned: The resources at NERSC are centrally allocated through NP. In order to enable experimental and observational data processing at NERSC, NERSC is recommending that with NERSC-9 that all program offices set aside a pool of resources for experimental data processing. Doing implementing this reserve for experiments after the upgrade would minimize the impact on the Theory programs.

Conclusion:

Relative to the interim report, some outstanding questions have been addressed, however, large uncertainties remain in the estimates and analysis models. GlueX has been in operations longer, has better defined estimates and has made excellent progress in using external resources. We have focused on developing a budget to provisioning an adequate system. We are committed to updating this planning exercise at least once a year.

Feedback on the Interim Report

We thank DOE NP for the feedback on the Interim Report.

1) Make sure you are in working with others that are addressing the same issues of large data analysis and Monte Carlo. Reach out to us if there is a way where NP can foster collaboration with other NP computing activities. STAR at RHIC/BNL has been very active in this area as have our NP LHC groups.

It is important to work within a broader community. We have been active in organizing workshops that bring computing experts from NP and HEP to Jefferson Lab to discuss common issues, participating in cross cutting working groups (in the US and CERN) and serving as external experts on reviews. We also added inline references in this document.

That said, we agree that more can be done to foster collaboration with other NP computing activities and will explore this with NP this summer

2) Consult with Ted Barnes to understand the level of resources that might be available at NERSC or leadership class machines.

We have consulted with Ted Barnes. For the short term, NERSC has granted resources for the GlueX effort from director's allocation. As mentioned in the body of the report, NERSC considers activities such as this (as for the STAR reconstruction [https://arxiv.org/abs/1702.06593]) to be a strategic initiative and is committed to working with SC program offices over the longer term on allocation issues.

3) The future of OSG is somewhat uncertain. Make sure you address this in your final report.

We discuss this in the report after communicating directly with OSG. We foresee that at the level of our proposed usage of OSG for GlueX, that we can count on this resource through 2021. We also expect there will be some other solution for distributed computing put in place and we look forward to playing a role in defining that solution.

4) The amount of storage needed and internet bandwidth to accommodate internal and external

computing is rightly pointed out as an area needing further evaluation. One PB of storage is low compared to other facilities taking the same size data sets. A second reconstruction of older data while dealing with current year data has been common at other sites.

We agree. We have made some progress in addressing the immediate bottlenecks and shortfalls coming from the rapid increase in data rates coming from the CLAS12 detector. This is an area in which the experience over the next year will dramatically inform the level and deployment of resources.

5) Have you considered a review with outside experts?

We have been holding S&C focused reviews with external experts every two years and found them to be valuable. The next one is scheduled for November 2018.