APEX 2020 RFP for Crossroads and NERSC-9 Systems
Questions & Answers
September, 2016

Question/Issue 1

Will there be a page limit for the RFP responses?

Project Response 1

Yes, 50 pages for the commercial proposal, and 200 pages for the technical proposal.

Question/Issue 2

What is the more precise meaning of "On-site System Delivery and Build Complete" for Crossroads and NERSC 9 by Q3CY20 as referenced within the Crossroads-NERSC-9 -Draft technical requirements document?

Project Response 2

These dates indicate when the systems are to be completely delivered and bootable. The actual subcontract schedules will be negotiated, but Offerors should propose based on the schedule within the Technical Requirements Document.

Question/Issue 3

Are Offerors anticipated to provide storage for home directories as part of the responses? If so, what capacity and performance characteristics are desired for this file system?

Project Response 3

Both the Crossroads and NERSC-9 platforms do not require storage to be provided for anything other than the platform integrated storage system, which is nominally meant to be used as a scratch filesystem(s). The specified external network requirements are meant to provide connections to ACES and NERSC provided storage using network protocols for file systems such as /home.

Question/Issue 4

Will it be sufficient for Offerors to reference specific, controlled documents currently available under existing NDA agreements in proposals, or will the contents of those documents also need to be provided?

Project Response 4

Referencing specific documents is not adequate; content must be provided.

Question/Issue 5

Can portions of Offeror's responses be marked as "Vendor Confidential"?

Project Response 5

Yes. There will be more specific instructions in the RFP when it is issued.

Question/Issue 6

Is the target SSI increase over Edison (> 20x) using the optimized MPI+X SSI or base case (MPI+ OpenMP) SSI?

Project Response 6

The Optimized MPI+X Case.

Question/Issue 7

Is there an example calculation for the Sustained System Improvement metric?

Project Response 7

An example calculation for the Capability Improvement metric has been provided for the APEX applications on the NERSC SSI web page.
(http://www.nersc.gov/research-and-development/apex/apex-benchmarks/ssi/)

Question/Issue 8

We are thinking of proposing a system with two different node types. In our response do we need to include benchmark times and an SSI calculation for both node types? How is the SSI calculated for a system with more than one type of node?

Project Response 8

Yes, your response should include benchmark times and an SSI calculation for both types of nodes.

We have provided an example SSI calculation for a system via a spreadsheet that can be downloaded on the SSI web page. The SSI is calculated using all benchmarks and across all node types. The resulting SSI is the sum of the SSIs for each node type.

Question/Issue 9

The RFP Technical Specifications specifies some benchmarks for "RFP Response" and others for "Acceptance". What does this mean?

Project Response 9

Results of benchmarks labeled “RFP Response” should be submitted as part of the RFP response. Results of benchmarks labeled “Acceptance” must be provided by the selected vendor at the start of negotiations for inclusion in the Statement of Work.

Question / Issue 10

Respondents to the RFI have expressed a concern regarding the sharing of proprietary and/or intellectual property (IP) information with Offerors of fully integrated systems.

Project Response 10

RFP language will require that each proposal be an offer for complete system integration for two separate systems, Crossroads and NERSC-9. RFP language will allow the direct submission to the LANS Procurement Specialist of proprietary and/or IP information that is a supplement to an offer for complete system integration. The Instructions to Offerors will provide guidance for the submittal of such supplemental proprietary and/or IP information. Each Offeror for full system integration will be responsible for timely submission of the supplemental information and for each supplement’s clear reference to the offer that is being supplemented. Submissions of supplemental information must comply with proposal preparation instructions (i.e. format and page limitation requirements)

and LANS will assume no responsibility/liability for any failure to comply with the instructions.

Question / Issue 11

Processor technology is controlled by the processor suppliers. If projected performance isn't realized, how is the system vendor protected?

Project Response 11

Subcontracts will be executed with the system vendor, presumably with the processor technology suppliers as lower tier suppliers to that system vendor. It is the responsibility of the system vendor to determine reliable sustained system performance and to propose accordingly. Any failure to meet proposed performance by the successful offeror is subject to usual contractual provisions, which will be stated in the RFP.

Question / Issue 12

Please provide more information on NRE funding availability. Primarily provide how much NRE funding will be available and what rules may apply for those seeking it.

Project Response 12

We are not specifying how NRE will be funded or how much NRE will be available. It depends on the proposals and their value add to the projects.

Question / Issue 13

Is there any information available on what the acceptance process would entail?

Project Response 13

Information on the acceptance process has been posted with the DRAFT Technical Requirements.

Question / Issue 14

Extrapolating benchmark numbers from a smaller system to a system of this magnitude has many unknowns. Also, the technology is some number of months out, further complicating the estimates.

Project Response 14

Providing benchmark estimates has always been a challenge. However, we have used this process for most of the past ACES and NERSC systems in the last 15 years. We have asked for risk mitigation for specific components such as processors in case the estimated performance doesn't measure up.

Question / Issue 15

Our benchmarking center isn't equipped with systems that large. Can you provide options for benchmarks on smaller subsets?

Project Response 15

The draft technical requirements we posted in December stated that performance benchmarks, application and micro-benchmarks, can be "actual, predicted and/or extrapolated". In addition, the benchmarking run rules posted on the web site provide further guidance under the "Submission Guidelines" section for performance projections (predicted and/or extrapolated). With a 2020 delivery, we expect performance projections are necessary as many key technologies are not available at this time.

Question / Issue 16

The RFP Technical Requirements Document states that a job interrupt shall not require a complete resource re-allocation. Does this mean if one of hosts fails during job run time, system will allocate a new resource to replace the failure one for a running job? Please help clarify this.

Project Response 16

The requirement intent is for the job to not have to resubmit to the scheduler, and hence wait in the queue, to allocate resources to continue. Any automation of application restart that can be provided as part of a solution would be valuable.

Question / Issue 17

The RFP Technical Requirements Document states a complete system initialization shall take no more than 30 minutes. How will this measured and please indicate what constitutes the start and stop time?

Project Response 17

The requirement is to describe the system's "full system initialize sequence and timings." How it is measured is system dependent and the method will be determined in the statement of work.

Question / Issue 18

Requirement 2.1.2 in the RFP Technical Requirements Document states "The Offeror shall provide a detailed description of the proposed software eco-system, including a high-level software architectural diagram including the provenance of the software component, for example open source or proprietary and support mechanism for each (for the lifetime of the system including updates)."

What is meant by "for the lifetime of the system including updates"?

Project Response 18

Please describe in your response the update mechanism for each software component over the expected lifetime of the system.

Question / Issue 19

What is meant by "baseline performance" for the APEX benchmarks? Can we, for example, replace whole functions with a vendor-supplied library?

Project Response 19

No, that would correspond to an optimized version.

Question / Issue 20

Re Specification 3.1.4 in the RFP Technical Requirements Document: "The proposed system shall support thousands of concurrent users and more than 20,000 concurrent batch jobs."
Are the "20,000 concurrent batch jobs" executing on the system or in queues waiting to run? If the former, how many are in queues? If the latter, how many are concurrently executing?

Project Response 20

The 20,000 jobs are concurrently executing, the batch queue may contain up to 10x that many in queues.

Question / Issue 21

Re: Specification "3.3.10 in the RFP Technical Requirements Document * Fast context switching or task-switching. & 3.3.10 *Fast task spawning for unique and identical task with data dependencies."

Under what programming environment is it desirable to have fast context switching or task-switching? Is this an operating system requirement, or is this wrapped up in a Parallel Programming Model environment? For example, the ability to oversubscribe MPI on the nodes and expect the MPI runtime to perform fast task-switching? Similarly, does the second bullet refer to task spawning with MPI?

Project Response 21

Requirement 3.3.10 in the RFP Technical Requirements Document expresses a desire for features that APEX expects to accelerate all current and future programming models. Please read the introductory paragraph for section 3.3.

Question / Issue 22

Re: 3.3.10 in the RFP Technical Requirements Document *Support for active messages. Does APEX have an API preference here? Or is APEX looking for hardware support for active messages?

Project Response 22

APEX did not specify a preference, the solution that provides most value to our needs is the one we are most interested in.

Question / Issue 23

Re: 3.3.7 in the RFP Technical Requirements Document. The Offeror shall provide a description of any hardware or software features that enable OpenMP performance optimizations.

Are they asking "What hardware/SW goodies make your OpenMP run fast?" Or "What hardware/software goodies let me understand and tune the performance of my code?"

Project Response 23

Which hardware/SW goodies make your OpenMP run fast.

Question / Issue 24

3.3.3 in the RFP Technical Requirements Document. The Offeror shall provide optimized implementations of collective operations utilizing both inter-node and intra-node features where appropriate, including MPI_Barrier, MPI_Allreduce, MPI_Reduce, MPI_Allgather, and MPI_Gather.

By "features", does APEX mean hardware features of the node and the network fabric?

Project Response 24

Yes.

Question / Issue 25

3.3.14 in the RFP Technical Requirements Document. The Offeror shall provide a programming toolchain(s) that enables runtime coexistence of threading in C, C++, and Fortran, from within applications and any supporting libraries using the same toolchain.

Does threading here mean POSIX threads or does it refer to threaded programming models such as OpenMP?

Project Response 25

It refers to threaded programming models such as OpenMP, C++11/14 std::thread/async and Fortran CONCURRENT.

Question / Issue 26

Requirement 5.6.7 (On-Site Node Cache) asks for an on-site spare node inventory of at least 1% of the total nodes in all of the system that may be met by utilizing the (ACES only) Requirement 5.6.8 the "Hardware Support Cluster" if provided. Can the offeror propose an on-site spare count that differs from the 1% requirement? Also, can the on-site node requirement for NERSC also allow for the utilization of a "Hardware Support Cluster"?

Project Response 26

The 1% of total nodes requirement for on-site spare node inventory is a "target" that is based on historical data of hardware failures for a variety of systems and the requirement to replace failed hardware in a timely manner without the need to replenish the spares inventory over a two-week period. The vendor is free to propose a solution which achieves the goal of this requirement. Additionally, the "Hardware Support Cluster" is meant to be a tool outside of the platform to assist with the diagnosis, repair and validation of hardware. NERSC will also consider proposals for utilization of a "Hardware Support Cluster" that utilizes the proposed on-site spare node inventory.