Attachment 1
Pre-Exascale Systems 2016-2017
Request for Information

December 13, 2012

Introduction

The U.S. Department of Energy (DOE) has a long history of deploying leading-edge computing capability for science and national security. The acquisition plans of the large DOE computer facilities have continued to march forward with new systems being deployed approximately every four years. In April 2011 a Memorandum of Understanding was signed between the Department of Energy Office of Science (SC) and the DOE National Nuclear Security Administration (NNSA) Office of Defense Programs regarding the coordination of exascale activities across the two organizations. One of the first activities to come out of this MOU is the coordination of the acquisitions by the SC and NNSA laboratories. In particular, ANL, LLNL, and ORNL have all deployed new systems in the past year. This means that all three of these labs will deploy their next systems in 2016-2017. Instead of duplicating efforts, the three labs are producing a single request for information (RFI) and are strongly considering a single request for proposal (RFP) for three separate procurements to site a defining generation of advanced technology systems at ANL, ORNL and LLNL in the 2016 to 2017 time period. These systems are required for the most demanding scientific and national security simulation and modeling applications as well as advanced technologies to assure cost-effective solutions and continued leadership in computing.

The purpose of this request for information RFI is to provide the collaboration consisting of ORNL, ANL, LLNL (CORAL) with information for the creation of an RFP as well as information useful in determining what, if any, non-recurring engineering (NRE) activities can be performed in conjunction with the acquisition of computer systems to make them more capable, higher performance, lower total cost of ownership, and/or more productive.

Only entities that would be responsible for the delivery of a complete system (prime subcontractors) should respond to this RFI. Your written response should include written materials describing your answers to the questions in this attachment along with any other information deemed informative that will help us structure the RFP.

Mission Need – Scientific Targets

Scientific computation has come into its own as a mature technology in all fields of science. Never before have we been able to accurately anticipate, analyze, and plan for complex events that have not occurred—from the operation of a reactor running at 100 million degrees to the changing climate a century down the road. Combined with the more traditional approaches of theory and experiment, it provides a profound tool for insight and solution as we look at complex systems containing billions of components. Nevertheless, it cannot yet do all we would like. Much of scientific computation's potential remains untapped—in areas such as materials science, Earth science, energy assurance, fundamental science, biology and medicine, engineering design, and national security—because the scientific challenges are far too enormous and complex for the computational resources at hand. Many of these challenges have immediate and global importance.

These challenges can be overcome by a revolution in computing that promises real advancement at a greatly accelerated pace. Planned 200 PF systems (capable of 10^{17} floating point operations per second) in the next four years and exascale systems (capable of an exaflop, or 10^{18} floating point operations per second) by the end of the decade provide an unprecedented opportunity to attack these global challenges through modeling and simulation.

DOE has several critical mission deliverables. Computer simulations are key to meeting these deliverables. To meet the scientific mission needs of 2016-2017 will require systems that can deliver 100-300PF, have 5-10PB of system memory, and 70-150 PB of storage. In addition, data movement and capacity in scientific codes is becoming a critical bottleneck in their performance. Thus memory hierarchy and its latencies and bandwidths between all its levels are expected to be the most important system characteristic for effective pre-exascale systems.

The final Figures of Merit (FOMs) for system selection are yet to be determined and will be specified in the RFP. The applications listed in table 1 are representative of the DOE mission needs. The final set of applications will be given in the RFP. The expectation is that the proposed 2016-2017 system will be roughly an order of magnitude less in time-to-solution than today's systems at our facilities.

Desired hardware and supporting software features by science teams include: portability, programmability, performance, reliability, and reproducibility. Today's codes run on BG/Q, Cray XE6, Cray XK7, and Linux clusters. Applications will require similar portability in the future across clusters, hybrid architectures, and heterogeneous systems.

Table 1. Representative applications of DOE mission needs. The final FOM list in the RFP will be similar in scope, but will most likely be primarily proxy or representative apps rather than full apps.

Application	Problem	Languages
AMG	Algebraic multi-grid	C
DNS3D	Pseudo-Spectral viscous CFD	Fortran 90
Lulesh-MM	Unstructured shock hydrodynamics	C
CoMD	Molecular dynamics	C
UMT	Unstructured mesh transport	Fortran and C
LSMS	Magnetic materials multiple scattering	Fortran
HACC	Particle-Mesh cosmology, I/O intensive	C and C++
S3D/SMC/CNS	CFD combustion	Fortran
MADNESS	Density functional theory, molecular science	C++
MADCAP	Cosmic Microwave Background I/O intensive	C

Plan for 2016-17 System Acquisitions

This plan is notional based on the expected budgets.

The plan for the DOE 2016-2017 system acquisitions based on calendar years:

- 4Q 2012 CORAL releases RFI to gather information about potential systems and related NRE.
- 2Q 2013 CORAL releases a draft RFP
- 3Q 2013 CORAL releases final RFP
- 3Q 2013 responses are due and will be evaluated by the CORAL selection team
- 4Q 2013 two distinct systems will be chosen. One to be delivered to ORNL and one to be delivered to ANL. LLNL will choose one of the winning systems to be delivered to their facility.

- 1Q 2014 two NRE subcontracts awarded – one to each prime subcontractor
- 2Q 2014 three separate system acquisition subcontracts will be awarded, one from each lab
- 2Q 2016 to 1Q 2017 pre-exascale systems delivered

The total NRE pool could be as much as $100M divided between the two winning prime subcontractors.

Requested Information

CORAL is collecting information from potential prime subcontractors in all the areas specified below.

Describe your hardware/software solution

Provide a high level description of the full system and include:

- Notional system architecture sketch with characteristics of the node, memory hierarchy, interconnect, and system
- Compute, I/O, RAS subsystems
- Software environment (management, development, operating)
- Packaging (power, cooling, floor space, weight)

How is it an appropriate system for the DOE Mission needs?

Discuss how your solution meets science needs for portability, programmability, performance, and reliability

Address the Potential Impact of NRE Funding on your proposed system

- What are the primary areas where you feel NRE funding is needed?
- How would NRE funding alter the system that might otherwise be bid? (compressed schedule, improved technology, reduced cost)
- How would the NRE work complement your related R&D funding (internal and external such as FastForward)?
- How does NRE funding impact cost? Higher capability at same cost? Reduced cost for same capability?

Areas that DOE has interest for NRE:

- Accelerated schedule
- Reduced system costs
- Resilience and RAS system
- Memory system technology, specifically to improve effective memory bandwidth and/or to mitigate memory capacity limitations
- Programmability
- Performance portability
- Maximizing transfer rates between components of heterogeneous nodes (PIM, on-chip accelerators, sharing memory between host/accelerator, etc.)
- Mechanisms to mitigate I/O system bandwidth, particularly to parallel file system
- NVRAM approaches to persistent storage and alternative models to traditional parallel filesystems
- Scalable networks
- Embedded NICs
- Advanced systems and power management features in the OS and runtime

- Packaging (density, cooling, energy reduction, cable management)
- Any other innovations that improve time-to-solution or total cost of ownership

How does the proposed system represent an innovative solution in the 2016/2017 time frame?

- Computational (node and board and interconnect level innovation)
- Infrastructure/packaging (system level innovation, ease of system management)
- Cost (system, energy, power management innovations)
- Innovations in portability, programmability, performance, reliability

In what ways is your solution a pre-exascale system?

In particular, how are the technologies, architecture, and programming on a path to exascale?

Cost Estimates

Provide cost estimates for system and if appropriate per scalable unit. Include any other cost considerations.