DesignForward-2
Statement of Work

June 6, 2014

National Nuclear Security Administration

U.S. DEPARTMENT OF ENERGY | Office of Science

This work is performed under the auspices of the U.S. Department of Energy by the Regents of the University of California / Lawrence Berkeley National Laboratory under Contract DE-AC02-05CH11231.

Contents

INTRODUCTION

The Department of Energy (DOE) has a long history of deploying leading-edge computing capability for science and national security. Going forward, DOE's compelling science, energy assurance, and national security needs will require a thousand-fold increase in usable computing power, delivered as quickly and energy-efficiently as possible. Those needs, and the ability of high performance computing (HPC) to address other critical problems of national interest, are described in reports from the ten DOE Scientific Grand Challenges Workshops[1] that were convened in the 2008–2010 timeframe. A common finding across these efforts is that scientific simulation and data analysis requirements are exceeding petascale capabilities and rapidly approaching the need for exascale computing. However, workshop participants also found that due to projected technology constraints, current approaches to HPC software and hardware design will not be sufficient to produce the required exascale capabilities. The deficiencies in those technology roadmaps threaten to impede development of DOE applications and slow the productivity of DOE scientists.

In April 2011 a Memorandum of Understanding was signed between the DOE Office of Science (SC) and the DOE National Nuclear Security Administration (NNSA), Office of Defense Programs, regarding the coordination of exascale computing activities across the two organizations. This led to the formation of a consortium that includes representation from seven DOE laboratories: Argonne National Laboratory, Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Oak Ridge National Laboratory, Pacific Northwest National Laboratory, and Sandia National Laboratories.

In July 2011, the aforementioned consortium released a request for information (RFI) with the purpose of providing DOE SC and NNSA with information for planning the DOE Exascale Computing Initiative (ECI). The RFI responses highlighted numerous challenges on the path to exascale and presented many innovative ideas to address those challenges.

Informed by these responses, DOE initiated two new programs, FastForward and DesignForward, to pursue ongoing strategic research and development (R&D) investments in critical technologies that align with DOE mission needs and provide benefit to future extreme-scale DOE applications. These programs address long-lead time items that will impact extreme-scale DOE systems later this decade. Through strategic partnerships with multiple companies, the FastForward and DesignForward programs seek to fund innovative new technologies and/or accelerate the progress of R&D efforts targeted for productization in the 5-10 year timeframe. We seek solutions that will maximize energy and concurrency efficiency while improving the performance (including overall time to solution), productivity, and reliability of key DOE extreme-scale applications.

This Request for Proposal (RFP), in the DesignForward program, focuses on strategic research and development (R&D) in system integration. The scope of these activities is described in detail in Attachment A. The period of performance for any subcontract resulting from this RFP will be two years. These contracts may be eligible for additional funding to further advance this critical R&D if Congress approves funding for this purpose and the R&D merits continuation.

---

[1] http://science.energy.gov/ascr/news-and-resources/workshops-and-conferences/grand-challenges/

Proposals submitted in response to this solicitation must address the impact of the proposed R&D on both DOE extreme-scale mission applications as well as the broader HPC community. Offerors are expected to leverage the DOE Co-Design Centers to ensure solutions are aligned with DOE needs. While DOE's extreme-scale computer requirements are a driving factor, these projects must also exhibit the potential for technology adoption by broader segments of the market outside of DOE supercomputer installations. This public-private partnership between industry and the DOE, initiated with FastForward, and continued with DesignForward, will aid the development of technology that reduces economic and manufacturing barriers to constructing exaflop-sustained systems, but also further DOE's goal that the selected technologies have the potential to broadly impact low-power embedded, cloud/datacenter, and midrange HPC applications. This ensures that DOE's investment furthers a sustainable software/hardware ecosystem supported by applications across not only HPC but also the broader IT industry. This will result in an increase in DOE's ability to leverage commercial developments. It is not DOE's intent to fund the engineering of near-term capabilities that are already on existing product roadmaps.

1 ORGANIZATIONAL OVERVIEW

1.1 The Department of Energy Office of Science

The Department of Energy Office of Science (SC) is the lead Federal agency supporting fundamental scientific research for energy and the Nation's largest supporter of basic research in the physical sciences. The SC portfolio has two principal thrusts: direct support of scientific research and direct support of the development, construction, and operation of unique, open-access scientific user facilities. These activities have wide-reaching impact. SC supports research in all 50 States and the District of Columbia, at DOE laboratories, and at more than 300 universities and institutions of higher learning nationwide. The SC user facilities provide the Nation's researchers with state-of-the-art capabilities that are unmatched anywhere in the world.

1.1.1 Advanced Scientific Computing Research Program

Within SC, the mission of the Advanced Scientific Computing Research (ASCR) program is to discover, develop, and deploy computational and networking capabilities to analyze, model, simulate, and predict complex phenomena important to the DOE. A particular challenge of this program is fulfilling the science potential of emerging computing systems and other novel computing architectures, which will require numerous significant modifications to today's tools and techniques to deliver on the promise of exascale science.

1.2 National Nuclear Security Administration

The NNSA is responsible for the management and security of the nation's nuclear weapons, nuclear non-proliferation, and naval reactor programs. It also responds to nuclear and radiological emergencies in the United States and abroad. Additionally, NNSA federal agents provide safe and secure transportation of nuclear weapons and components and special nuclear materials along with other missions supporting the national security.

1.2.1 Advanced Simulation and Computing Program

Established in 1995, the Advanced Simulation and Computing (ASC) Program supports NNSA Stockpile Stewardship Program's shift in emphasis from test-based confidence to simulation-based confidence. Under ASC, high-performance simulation and computing capabilities are developed to analyze and predict the performance, safety, and reliability of nuclear weapons and to certify their functionality. As the nuclear stockpile moves further from the nuclear test base through either the natural aging of today's stockpile or introduction of component modifications, the realism and accuracy of ASC simulations must further increase through development of improved physics models and methods, requiring ever greater computational resources.

2 MISSION DRIVERS

2.1 Office of Science Drivers

DOE's strategic plan calls for promoting America's energy security through reliable, clean, and affordable energy, ensuring America's nuclear security, strengthening U.S. scientific discovery, economic competitiveness, and improving quality of life through innovations in science and technology. In support of these themes is DOE's goal to significantly advance simulation-based scientific discovery, which includes the objective to "provide computing resources at the petascale and beyond, network infrastructure, and tools to enable computational science and scientific collaboration." All the other research programs within DOE SC depend on the ASCR to provide the advanced facilities needed as the tools for computational scientists to conduct their studies.

Between 2008 and 2010, program offices within the DOE held a series of ten workshops[2] to identify critical scientific and national security grand challenges and to explore the impact exascale modeling and simulation computing will have on these challenges. The extreme scale workshops documented the need for integrated mission and science applications, systems software and tools, and computing platforms that can solve billions, if not trillions, of equations simultaneously. The platforms and applications must access and process huge amounts of data efficiently and run ensembles of simulations to help assess uncertainties in the results. New simulations capabilities, such as cloud-resolving earth system models and multi-scale materials models, can be effectively developed for and deployed on exascale systems. The petascale machines of today can perform some of these tasks in isolation or in scaled-down combinations (for example, ensembles of smaller simulations). However, the computing goals of many scientific and engineering domains of national importance cannot be achieved without exascale (or greater) computing capability.

2.2 National Nuclear Security Administration Drivers

Maintaining the reliability, safety, and security of the nation's nuclear deterrent without nuclear testing relies upon the use of complex computational simulations to assess the stockpile, to investigate basic weapons physics questions that cannot be investigated experimentally, and to

---

[2] http://science.energy.gov/ascr/news-and-resources/workshops-and-conferences/grand-challenges/

provide the kind of information that was once gained from underground experiments. As weapon systems age and are refurbished, the state of systems in the enduring stockpile drifts from the state of weapons that were historically tested. In short, simulation is now used in lieu of testing as the integrating element. The historical reliance upon simulations of specific weapons systems tuned by calibration to historical tests will not be adequate to support the range of options and challenges anticipated by the mid-2020s, by which time the stewardship of the stockpile will need to rely on a science-based predictive capability.

To maintain the deterrent, the U.S. Nuclear Posture Review (NPR) insists that "the full range of Life Extension Program (LEP) approaches will be considered: refurbishment of existing warheads, reuse of nuclear components from different warheads, and replacement of nuclear components." In addition, it is recognized that as the number of weapons in the stockpile is reduced, the reliability of the remaining weapons becomes more important. By the mid-2020s, a science-based predictive capability will be necessary in order to support the range of options with sufficient certainty as called for in the NPR. In particular, existing NNSA computational facilities and applications will be inadequate to meet the demands for the required technology maturation for weapons surety and life extension by the middle of the next decade. Evaluation of anticipated surety options is raising questions for which there are shortcomings in our existing scientific basis. Correcting those shortcomings will require simulation of more detailed physics to model material behavior at a more atomistic scale and to represent the state of the system. This pushes the need for computational capability into the exascale level.

3 EXTREME-SCALE TECHNOLOGY CHALLENGES

The HPC community has done extensive analysis[3] of the challenges of delivering exascale-class computing. These challenges also apply more generally to extreme-scale HPC, regardless of whether or not the end result is an exaflop computer. This section provides an overview of the most significant of these challenges.

3.1 Power Consumption and Energy Efficiency

All of the technical reports on exascale systems identify the power consumption of the computers as the largest challenge going forward. Today, power costs for the largest petaflop systems are in the range of $5–10M annually. If an exascale system were deployed using current technology, the annual power cost to operate the system would be above $250M per year with a power load of more than 350MW. To keep the operating costs of such a system in a workable range, a target of 20 Megawatts has been established.

Achieving the power target for exascale systems is a significant research challenge that pervades system design. Even with optimistic expectations of current R&D activities, ***there is at least a factor-of-five gap between what we must have and what business as usual roadmaps can***

[3] http://science.energy.gov/~/media/ascr/ascac/pdf/meetings/20140210/Top10reportFEB14.pdf; http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Arch_tech_grand_challenges_report.pdf; http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Crosscutting_grand_challenges.pdf; http://www.cse.nd.edu/Reports/2008/TR-2008-13.pdf; http://www.exascale.org/mediawiki/images/2/20/IESP-roadmap.pdf

provide. To get the additional factor of five improvements in power efficiency over projections, a number of technical areas in hardware and software design need to be explored. These may include: energy efficient hardware building blocks (central processing unit (CPU), memory, interconnect), more efficient cooling, dense packaging, Si-Photonic communication, and power-aware runtime software and algorithms.

3.2 Concurrency

The end of increases in single compute core performance by increasing Instruction Level Parallelism (ILP) and/or higher clock rates has left explicit parallelism as the only mechanism in silicon to increase performance. Scaling up absolute performance will require scaling up the number of cores accordingly, projected to be in the billions for exascale systems.

Allowing applications to efficiently exploit this level of concurrency is a challenge for which there currently are no comprehensive solutions. While many of these challenges require improvements in memory and processors, which were the target of the previous Fast Forward solicitation, and interconnect, a target of the previous DesignForward solicitation, there remain many remaining challenges in the space of overall system architecture, system software, and system design.

Further complicating this is the problematic growth in the ratio of energy to transport data versus the energy to compute with it. At the exascale level, this transport energy drives architectural tradeoffs. Reducing the transport energy will require innovative packaging, interconnect, and architecture to bring the data needed by a computation energy-wise “closer to” the function units, or reduce the amount of data required.

3.3 Fault Tolerance and Resiliency

Resilience is a measure of the ability of a computing system and its applications to continue working in the presence of system degradations and failures. The frequency of faults in a computing system depends strongly on the number of components that it contains and the reliability of the individual components. Exascale systems will be composed of huge numbers of components constructed from VLSI devices that will not be as reliable as those in use today. It is projected that the aggregate component mean time to fail for an exascale system may be in the minutes or seconds range. Increasing evidence also points to a rise in silent errors (faults that never get detected or get detected long after they generated erroneous results), potentially causing havoc.

Exascale systems will continually experience failures, necessitating significant advances in the methods and tools for gracefully dealing with them. Achieving acceptable levels of resiliency in exascale systems will require improvement in hardware and software reliability; better understanding of the root cause of errors; better reliability, availability, and serviceability (RAS) collection and analysis; fault resilient algorithms and applications to assist the application developer; and local recovery and migration.

3.4 Programmability

Programmability is the crosscutting property that reflects the ease by which effective application programs may be constructed. Programmability affects developer productivity and ultimately leads to the productivity of an HPC system as a tool to enable scientific research and discovery.

Programmability itself involves three stages of application development: (1) program algorithm captures and representation, (2) program correctness debugging, and (3) program performance optimization. All levels of the system, including the programming environment, the system software, and the system hardware architecture, affect programmability. The challenges to achieving programmability are myriad, related both to the representation of the user application algorithm and to resource usage system-wide.

- Parallelism—sufficient parallelism must be exposed to maintain exascale operation and hide latencies. It is anticipated that 10-billion-way operation concurrency will be required.
- Distributed Resource Allocation and Locality Management—to make such systems programmable, the tension must be balanced between spreading the work among enough execution resources for parallel execution and co-locating tasks and data to minimize latency.
- Latency Hiding—intrinsic methods for overlapping communication with computation must be incorporated to avoid blocking of tasks and low utilization of computing resources.
- Hardware Idiosyncrasies—properties peculiar to specific computing resources such as memory hierarchies, instruction sets, and accelerators must be managed in a way that circumvents their negative impact while exploiting their potential opportunities without demanding overly explicit user control.
- Portability—application programs must be portable across machine types, machine scales, and machine generations. Performance sensitivity to small code perturbations should be minimized.
- Synchronization Bottlenecks—barriers and other over-constraining control methods must be replaced by lightweight synchronization, overlapping phases of computation.

Novel execution models and architectures may increase programmability, thereby enhancing the productivity of DOE scientists.

4 APPLICATIONS

The applications that will eventually run on future exascale systems will be a product of a process that starts with the DOE Mission Drivers outlined in Section 3 above, and continues on through identifying key science challenges, developing mathematical models, developing algorithms, and finally producing application software.

While the key challenges in many fields have been effectively articulated in the workshop reports referenced in Section 1, the development of models, algorithms and software is dependent on how hardware and system software designers respond to the challenges of building an exascale system and vice versa. DOE's aim is to resolve many possible trade-offs in the space of applications and architectures by using the co-design methodology described in the next section.

5 ROLE OF CO-DESIGN

5.1 Overview

The R&D funded through this RFP is expected to be the product of a co-design process. Co-design refers to a system-level design process where scientific problem requirements influence architecture design and technology, and architectural characteristics inform the formulation and design of algorithms and software. To ensure that future architectures are well-suited for DOE target applications and that DOE scientific problems can take advantage of the emerging computer architectures, major R&D centers of computational science are formally engaged in the hardware, software, numerical methods, algorithms, and applications co-design process.

Co-design methodology requires the combined expertise of vendors, hardware architects, system software developers, domain scientists, computer scientists, applied mathematicians, and systems staff working together to make informed decisions about the design of hardware, software, and underlying algorithms. The path forward is rich with trade-offs, and give and take will be needed from both the hardware and software developers. Understanding and influencing these trade-offs is a principal co-design requirement.

ASCR and ASC have established multiple application co-design centers that serve as R&D collaboration vehicles with all aspects of the extreme-scale development ecosystem, especially with the HPC vendors.

5.2 ASCR Co-Design Centers

In 2011, ASCR awarded three application co-design centers. Each center focuses on a specific application that is an important driver for exascale[4] and is a distributed collaboration between multiple national laboratories and university partners. Development of that application facilitates exploration of issues of mathematics, algorithms, computer science, systems software, and of course, hardware in the co-design process. For a detailed description of the ASCR co-design centers see http://science.energy.gov/ascr/research/scidac/co-design/.

5.3 ASC Co-Design Project

The NNSA labs and ASC program have defined a coordinated co-design strategy that leverages the work of the ASCR co-design centers while focusing on the unique needs of the ASC program. ASC is a mission-driven program with applications currently in use that are of importance to run at exascale in support of stockpile stewardship, namely the Engineering and Physics Integrated Codes (EPICs). To meet the key needs of the EPICs, ASC has established the National Security Applications (NSApp) Co-Design Project. NSApp focuses on these established applications as the drivers and participates in co-design largely through proxy applications. Additional information is available at https://asc.llnl.gov/codesign/.

[4] http://science.energy.gov/ascr/news-and-resources/workshops-and-conferences/grand-challenges/

5.4 Proxy Apps

The DOE co-design centers make extensive use of proxy applications to represent the application workflow and requirements to the exascale ecosystem. These application codes are used to understand the effects of hardware tradeoffs, and to explore and develop new technologies, runtime systems, languages, programming models, algorithms, tools, file systems, and visualization techniques. Whenever possible, proxy apps are openly available – with occasional need to protect the original source under export-control rules or proprietary access rules in some cases where vendor modifications are supplied back to the co-design center.

In general, a proxy app is a small application code that represents some aspect of the computational workflow of a full application. Proxy apps can be grouped into three categories in increasing sophistication and fidelity to the parent applications and integrated codes:

- Kernels: these are small code fragments (algorithms) that are used extensively by the parent application and are deemed essential to perform optimally.
- Skeleton apps: these apps reproduce the data flow of a simplified application and make little or no attempt to investigate numerical performance. They are primarily useful in investigating memory management, network performance characteristics, I/O, etc.
- Mini- or compact apps: these apps contain the dominant numerical kernels contained in the parent application and represent the computational workflow in as compact a form as possible.

It is important to emphasize that these proxy apps will not be static but will evolve significantly during the co-design process. The co-design centers anticipate the requirement for domain application code-developers to spend significant time with the vendors as well as vendor developers and architects to spend significant time with the co-design centers.

ASC and ASCR co-design centers are developing and publishing their proxy apps. Some that are available today are:

ExMatEx	http://www.exmatex.org
ExaCT	http://www.exactcodesign.org
CESAR	http://cesar.mcs.anl.gov
TORCH	http://crd.lbl.gov/groups-depts/ftg/projects/previous-projects/torch-testbed
Mantevo	http://software.sandia.gov/mantevo
NERSC SSP	http://www.nersc.gov/research-and-development/performance-and-monitoring-tools/sustained-system-performance-ssp-benchmark
LULESH	https://computation.llnl.gov/casc/ShockHydro
SNAP	https://github.com/losalamos/snap

5.5 DOE Supercomputer Centers

The DOE National Laboratories have decades of experience deploying and operating many of the world's largest computers. Often the largest example of a new supercomputer is at a DOE Laboratory. For extreme scale systems the ability to integrate, configure, operate, and maintain the system is vital to optimizing system availability and productivity. This is especially important at exascale, where the volume of system state information is an important consideration in system architecture. DOE supercomputer facility staff of the consortium members can be important assets in the co-design of exascale systems.

Offerors may wish to review the RFPs for recent DOE supercomputer procurements in order to better understand the challenges associated with siting and operating these large-scale systems. See:

- Trinity/NERSC-8 http://www.nersc.gov/systems/nersc-8-procurement/trinity-nersc-8-rfp/
- CORAL https://asc.llnl.gov/CORAL/

6 REQUIREMENTS

6.1 Description of Requirement Categories

Requirements are either mandatory (Mandatory Requirements - designated MR) or target (Target requirements - designated TR-1, or TR-2), and are defined as follows:

- MRs are performance features essential to DOE requirements. An Offeror must satisfactorily address ***all*** MR to have its proposal considered responsive and eligible for further evaluation.
- TRs, identified throughout this Statement of Work, are features, components, performance characteristics, or other properties that are important to DOE but will not result in a nonresponsive determination if omitted from a proposal. TRs add value to a proposal and are prioritized by dash number. TR-1 is more desirable than TR-2.

TR-1s and MR are of equal value. The aggregate of MRs and TR-1s form a baseline solution. TR-2s are goals that boost a baseline solution. Taken together as an aggregate, MRs, TR-1s, and TR-2s form an enhanced solution.

6.2 Requirements for Research and Development Investment Areas

Detailed requirements for the DesignForward-2 System Integration R&D investment area are provided in Attachments A. Each proposal shall address all of the MRs listed below, in the order listed.

The various R&D topics within a proposal must form a cohesive project where the proposed R&D tasks fit together well. Offerors may submit multiple proposals, each covering a cohesive scope of R&D.

6.3 Proposal Page Limit

The total length of a proposal, excluding cover letter, cover page, table of contents, references and staff curricula vitae (CVs), shall not exceed twenty-five pages, with a minimum text font size of 11 points and margins no smaller than one inch on all sides. Tables, figures, appendices, and attachments are part of the 25 page limit.

6.4 Mandatory Requirements

The following items are mandatory for all proposals.

6.4.1 Solution Description (MR)

Offeror shall describe the proposed exascale conceptual system design R&D, with emphasis on how it will achieve the goals of this RFP and increase the performance of key DOE extreme-scale applications relative to energy usage while maintaining or increasing reliability and maintaining or decreasing runtimes.

Offerors shall discuss the innovative nature of the proposed exascale conceptual system design R&D and describe where it differs from company roadmaps. Work that funds a company's current roadmap is not acceptable. The primary intent is to fund long-lead-time R&D objectives that overcome the extreme-scale technology challenges that were described in Section 3 of this document.

6.4.2 Research and Development Plan (MR)

Offeror shall provide a plan for conducting the proposed R&D, including timelines, milestones, and proposed deliverables. Deliverables shall be meaningful and measurable. **Pricing shall be assigned to each milestone and deliverable**. A schedule for periodic technical review by the DOE laboratories shall also be provided. A work breakdown structure shall be provided.

The R&D funded through this RFP is expected to be the product of a co-design process. More specifically, Offerors are expected to engage in co-design activities with DOE's ASC and ASCR Exascale Co-design Centers. The R&D plan shall include a discussion of how Offeror plans to collaborate with DOE researchers on co-design. A detailed description of planned co-design efforts is highly desired.

Quantitative measures of design innovations are desired. Projects will include simulations or analyses that assess the impact (or feasibility) of a proposed innovation in the conceptual system design. DO NOT include the estimated amount for a follow-on activity in the price for the DesignForward R&D being proposed in response to this RFP. The follow-on work could be proposed in response to a future RFP, if one is issued.

6.4.3 Productization Strategy (MR)

Offeror shall describe how the proposed conceptual system design will be commercialized, productized, or otherwise made available to customers. Offerors shall include identification of target customer base/market(s) for the technology. Offerors shall describe impact specifically on the HPC market as well as the potential for broad adoption. Solutions that have the potential for broader adoption beyond HPC are highly desired. Offerors shall indicate a projected timeline for productization.

6.4.4 Staffing/Partnering Plan (MR)

Offerors shall describe staffing categories and effort for the proposed R&D activities. All lead and key personnel shall be identified by name and brief CVs for these personnel shall be provided. Any collaboration with other industry partners and/or universities shall be identified, and any key personnel from these partners/subcontractors shall be provided together with a description of their contributions to the overall effort.

6.4.5 Project Management Methodology (MR)

Project management and regular project status reporting are required. Offeror shall describe project management methodology and provide a communication plan indicating the method of communication (for example, written report, teleconference, and/or face-to-face meeting) and frequency (for example, weekly, monthly, and/or quarterly).

6.4.6 Intellectual Property Plan (MR)

Proposals shall include a plan for how each intellectual property (IP) item from each portion of the proposed R&D work will be handled, including requested IP ownership and licensing. Please consult RFP letter for information on Federal regulations concerning IP.

7 EVALUATION CRITERIA

7.1 Evaluation Team

The Evaluation Team includes representation from seven DOE laboratories: Argonne National Laboratory, Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Oak Ridge National Laboratory, Pacific Northwest National Laboratory, and Sandia National Laboratories, as well as Federal government representatives. The Regents of the University of California, manager and operator of Lawrence Berkeley National Laboratory (LBNL), as the entity awarding subcontracts as a result of this RFP on behalf of DOE SC and NNSA, will act as the source selection official.

7.2 Evaluation Factors and Basis for Selection

The Offeror’s proposal should identify and discuss the performance features and supplier attributes that will be important to the Offeror’s successful performance and the attainment of the DesignForward-2 project objectives. Evaluation factors are mandatory requirements, performance features, supplier attributes, and price that the Evaluation Team will use to evaluate proposals. These features and attributes are not listed in any particular order of importance. Importance is prioritized per the guidance in section 6.2 *Description of Requirement Categories*. The evaluation will be based on the information provided by the Offeror, the Evaluation Team’s own experiences, and/or information from an Offeror’s customers or other sources. The Evaluation Team has identified the mandatory requirements, performance features and supplier attributes listed above and in each Attachment that should be discussed in the proposal. Offerors may identify and discuss other performance features and supplier attributes they believe may be of value to the Evaluation Team. If the Evaluation Team agrees, consideration may be given to them in the evaluation process. In all cases, the DOE Laboratories will assess the value of each proposal as submitted.

The Evaluation Team’s assessment of each proposal’s evaluation factors will form the basis for selection. LBNL intends to select the responsive and responsible Offerors whose proposals contain the combination of price, performance features, and supplier attributes offering the best overall value to DOE. The Evaluation Team will determine the best overall value by comparing differences in performance features and supplier attributes offered with differences in price, striking the most advantageous balance between expected performance and the overall price. Offerors must, therefore, be persuasive in describing the value of their proposed performance features and supplier attributes in enhancing the likelihood of successful performance or otherwise best achieving the DOE's objectives for extreme scale computing.

7.3 Performance Features

The Evaluation Team will validate that an Offeror's proposal satisfies the MR. The Evaluation Team will then assess if, and how well, an Offeror's proposal addresses the TR. An Offeror is not solely limited to discussion of features described in TR. An Offeror may propose other features or attributes if the Offeror believes they may be of value. If the Evaluation Team agrees, consideration may be given to them in the evaluation process. In all cases, the Evaluation Team will assess the value of each proposal as submitted.

The technical/management proposal should contain a comprehensive discussion of how the Offeror will fulfill the MRs and TRs and successfully perform the Subcontract. The Evaluation Team will evaluate the following performance features as proposed:

- How well the proposed solution meets the overall programmatic objectives expressed in the SOW
- The degree to which the technical proposal meets or exceeds any TR
- The degree of innovation in the proposed R&D activities
- The extent to which the proposed R&D achieves substantial gains over existing industry roadmaps and trends
- The extent to which the proposed R&D will impact HPC and the broader marketplace
- Credibility that the proposed R&D will achieve stated results
- Credibility of the productization plan for the proposed technology
- Realism and completeness of the project work breakdown structure

7.4 Feasibility of Successful Performance

The Evaluation Team will assess the likelihood that the Offeror's proposed research and development efforts can be meaningfully conducted and completed within the anticipated two-year subcontract period of performance. The Evaluation Team will also assess the risks, to both the Offeror and the DOE laboratories, associated with the proposed solution. The Evaluation Team will evaluate how well the proposed approach aligns with the Offeror's corporate roadmap and the level of corporate commitment to the project.

7.5 Supplier Attributes

The Evaluation Team will assess the following supplier attributes.

7.5.1 Capability

The Evaluation Team will assess the following capability-related factors:

- The Offeror's experience and past performance engaging in similar R&D activities. The Offeror must include in its proposal a written description of at least three recent contracts which the Offeror successfully completed, similar in type and complexity to the scope of the proposed Subcontract. Technical and business contact information for the contracts, including name, title and email address is required. The Offeror must also identify key personnel who will perform under the proposed Subcontract and provide resumes. LBNL reserves the right to request additional information and to check references as appropriate.
- The Offeror should have adequate financial resources to perform the Subcontract or the ability to obtain them. The proposal should describe the Offeror's current financial condition and its financial ability to fully support performance of the proposed Subcontract. The Offeror's financial capability should be supported by audited and certified year-end financial statements for the last two years in the form of balance sheets and income statements, and include information on any commitments or explicit arrangements that will be relied on to acquire the needed facilities, equipment, other resources, or personnel. LBNL reserves the right to request additional financial statements or information regarding the Offeror's financial resources.
- The Offeror's demonstrated ability to meet schedule and delivery promises
- The alignment of the proposal with the Offeror's product strategy
- The expertise and skill level of key Offeror personnel
- The contribution of the management plan and key personnel to successful and timely completion of the work
- If work is subcontracted, how much is to small business, to what extent is the assignment of work scope appropriate, and to what extent is the prospective lower-tier subcontractor qualified to do that work?

7.6 Price of Proposed Research and Development

The Evaluation Team will assess the following price-related factors:

- Reasonableness of the total proposed price in a competitive environment
- Proposed price compared to the perceived value
- Price tradeoffs and options embodied in the Offeror's proposal
- Financial considerations, such as price versus value

ATTACHMENT A: SYSTEM INTEGRATION R&D REQUIREMENTS

The scope for the DesignForward System Integration R&D request for proposals is the development of a conceptual design of an integrated exascale system, containing all the essential elements (hardware software, system management, packaging, power supplies, cooling, etc.). While the time and resources available are unlikely to support a finished system design, it is vital that the co-design dialog between DOE Laboratories and vendors begin now so that shared decisions and tradeoffs can inform future detailed designs. Offerors must detail their vision for the execution model and overall system architecture, as well as the specific DesignForward system integration research that they propose to accelerate industry roadmaps or to provide capability that existing market forces would not ensure. Offerors should also discuss how the proposed R&D will impact their commercialization/business strategy for their company's exascale- and HPC-related projects. While the focus of the system design should be on exascale systems, there may be interim benefits to the proposed R&D. Offerors should describe how their DesignForward effort could feed into and benefit DOE's upcoming pre-exascale procurements.

A1-1 Key Challenges for System Integration

A1-1.1 Overall Execution Model and System Architecture

Large-scale systems are increasingly using a range of node types for different functions, such as computation, I/O, analytics, and compilation. This diversity creates additional complexity for all aspects of system integration including packaging, interconnection networks, system software and execution models. Overall system architecture should be robust and maximize delivered value for likely exascale workflows.

Effective use of today's High Performance Computing (HPC) systems and programming methods depends on the ability to express and utilize extreme parallelism efficiently. Current execution models may not map well onto the specific characteristics critical to exascale system efficiency and scalability, leading to poor use of system resources and limitations on system effectiveness. A new model of computation or an execution model is required and must be developed that enables the programmer to perceive the system as a unified and naturally parallel computer system; not merely as a collection of microprocessors and an interconnection network. The execution model provides the conceptual scaffolding for deriving system elements in the context of and consistent with each other. Ideally, the execution model implements a decision chain where each layer contributes to the optimum determination of when, where, and how data placement, data movement, and operation of a computation are performed. A more in-depth discussion about exascale execution models can be found in the DOE Exascale Programming Challenges[5] report, section 3.2.

[5] http://www.nersc.gov/assets/pubs_presos/ProgrammingChallengesWorkshopReport.pdf

A1-1.2 Energy Utilization

Energy and power are key design constraints for exascale systems. Power management and energy cost reduction techniques are needed to minimize or constrain power used while maintaining reliable, predictable behavior.

A1-1.3 Resilience and Reliability

The integrated, system-wide RAS (reliability, accessibility and serviceability) system is critical to the stability and usability of today's supercomputers. Development of exascale RAS systems having the ability to identify, to contain, and to overcome faults quickly is of paramount importance to avoid application and system failures. Incorporating an appropriate level of awareness and management of fault tolerance into each layer of the hardware/software stack, including the application programming environment and tools, is needed to ensure that applications can continue to progress even with system faults.

A1-1.4 Data Movement through the System

The scalability and performance of DOE applications is tightly linked to the latency and bandwidth of data movement through the system. Science applications utilize many different data patterns and a wide range of data sizes. Higher performance, higher efficiency (lower energy use) data movement technologies that optimize end-to-end data movement throughout the system are needed.

A1-1.5 Density

Packaging density largely determines the needed floor space for an exascale system. Methods for improving system density, while managing heat dissipation and limiting cabinet weight, are needed to keep the size of the complete exascale system between 100 and 300 cabinets.

A1-1.6 System Software

System software includes the operating system (OS), file system, runtime, system monitoring, resource management, job management, accounting and user management, and security. These software components must be improved to cope with the scale of an exascale system, to ensure its overall reliability, and to incorporate power management into their actions. Mechanisms must also be developed to support integrated, seamless maintenance of system software and monitoring within the overall computing center.

A1-1.7 Programming Environment

The programming environment includes the languages supported, and the required software including compilers, debugging tools, runtime, libraries, and performance tuning. Programming environment advances are needed to improve system integration, programmability, code portability, and usability. Programming environments and tools should facilitate porting of the current DOE code base and ensure that the development of new codes is as straightforward and efficient as possible.

A1-1.8 Data and Workflow Management

The scientific discovery process increasingly involves complex sequences of steps that take advantage of multiple sources of data and benefit from different types of compute and data resources. At exascale, the cost to move and store data, and perform offline analysis will require scientists to adopt different workflows than are in use today. The scheduling of appropriate tasks on appropriate resources will apply not just to the simulation application, but also in transit/in-situ analysis tasks. These tasks will need to be scheduled on different parts of the system such as compute nodes, IO/burst buffer nodes, or even within the storage subsystem. These new workflow capabilities will be key to enabling successful exascale science.

A1-2 Areas of Interest

The following are examples of objectives and technologies that could be considered in system integration R&D proposals that address DOE's extreme-scale computing needs. Some of the items below may only apply to certain architectures, and some may be mutually exclusive. Furthermore, this list of topics is not exhaustive. Offerors are not expected to propose in-depth R&D for all of these topics, and may propose alternative topics within these areas of interest. However *all* proposals must describe the proposed execution model and architecture of a specific and complete system and a proposed R&D effort for developing an exascale conceptual system design. The planned co-design process should be clearly stated for each topic.

A1-2.1 Overall System Architecture

- Designs that simplify changing or upgrading specific node capabilities (e.g., processors, memory, coprocessors) or that enable node substitution in the face of faults that may degrade or kill nodes
- Mechanisms to increase flexibility in resource utilization such as ways to share memory capacity across nodes
- Mechanisms to mitigate the tension between production system use, which primarily entails large jobs, and software development for the system, which involves non-computational tasks such as compilation and short jobs for testing and debugging
- Designs that facilitate compiling for a mix of heterogeneous nodes
- Mechanisms to support isolation and flexibility in resource association, e.g., partitioning resources among jobs
- Techniques to support efficient scheduling of diverse resource types
- Scalable, adaptive, and unobtrusive monitoring, with real-time analysis of platform state
- Real-time autonomic platform management under production workloads, gracefully handling unplanned events without requiring immediate human intervention
- Execution models that enable the programmer to perceive the system as a unified and naturally parallel computer system, not merely as a collection of microprocessors and an interconnection network.

- Programming and execution models that provide for runtime support of the coexistence of threading among all the supported languages (C, C++, Fortran, etc.), for within the application and any supporting libraries.

A1-2.2 Energy Utilization

- Designs that improve the power efficiency of the system
- Techniques for measurement, runtime control and application control of power utilization
- System-wide and site-wide power management methodologies
- Techniques to reduce cooling energy requirements

A1-2.3 Resilience and Reliability

- Designs that improve the resiliency or reliability of the system, for example, improved fault detection, containment, correction, and response time across the software stack (not just hardware)
- Methods that enable dynamic adaptation to a constantly changing system
- Leveraging of hardware/software resilience synergies to improve overall time to solution
- Techniques to improve fault detection accuracy (e.g., fewer undetected errors) and root cause analysis or to reduce their cost and time to repair/recovery
- Framework for representing hardware and system software dependencies, for interpretation of failure modes and autonomic reasoning about remaining recovery paths

A1-2.4 Data Movement through the System

- Designs that allow extremely low-latency multi-hop messages
- Improvements to the performance and energy efficiency of messaging, remote memory access, and collective operations
- Analysis of the optical/copper tradeoffs to improve the data movement across the system
- Reliable low-energy, long-distance data movement
- Efficient data movement for computation and also across levels of storage hierarchy
- Mechanisms to avoid contention, and to provide QoS guarantees (bandwidth, latency, reliability, etc.)

A1-2.5 Density

- Innovative packaging and cooling that improve the density of the system
- Methods to address potential scaling and concurrency limits that restrict the ultimate size of the system
- Lightweight packaging and cooling techniques

A1-2.6 System Software

- Methods that significantly improve resilience support
- Techniques to achieve efficient dynamic load balancing
- Designs that provide coarse and fine grain power management across the system
- Support for job and resource management at the scale of a million nodes and a billion cores
- Techniques to manage diverse node types and nodes with heterogeneous resources
- Topology, resilience, and energy aware scheduling of resources
- Common, interoperable runtime support for concurrency (e.g., threading) across programming languages and libraries

A1-2.7 Programming Environment

- Designs that provide an integrated repertoire of programming model(s) that support composability of application components written in multiple languages using multiple programming models
- Designs that significantly improve application resilience
- Designs that help scale applications to a billion-way concurrency
- Development of programming tools that work with million node systems
- Hardware or software techniques to simplify writing and debugging application software and promote portability to different node types and system architectures

A1-2.8 Data and Workflow Management

- Development of integrated models for use of in-system nonvolatile storage (which could be located anywhere within the system architecture) including integration with programming models, abstractions to exploit locality awareness and security models
- Methods to integrate scientific workflows into system resource managers, including abstractions to describe data management resource requirements (storage, bandwidth, etc.) in a non-system dependent way
- Integration of data management RAS features into the overall system-wide RAS capability.

A1-3 Performance Metrics

Offeror shall estimate or quantify the impact of the proposed technology over industry roadmaps and trends. This information shall be provided for all of the metrics listed below. If Offeror determines that a particular metric is not applicable to the technology being proposed, then

Offeror shall explain why they believe the metric is not relevant and shall replace that metric with an alternate ***meaningful*** metric.

Quantities specified should reflect solutions that are productized in the 2020-2023 timeframe. These metrics are independent, but a solution that can deliver advances in more than one metric is more desirable than one that addresses only one metric at the expense of the others. The most meritorious improvements will make substantial gains over industry roadmaps/trends and substantiate a convincing path to achieving the extreme-scale technology characteristics required by DOE.

The list of metrics includes:

- Computational capacity per cabinet
- Cabinet power requirements
- Scaling limits of the overall system design
- System-wide optimization of communications performance and energy requirements
- Expected resilience of the system (MTTI, mean time to system interrupt)
- Total cost of ownership
- Total delivered value for exascale workflows

A1-4 System Integration Requirements

The requirements below apply to supercomputers that will be deployed by DOE Office of Science and NNSA at the end of this decade and in the early years of the following decade to meet the two programs' mission needs. As previously stated, Offerors need not address all problem areas, and thus the Offeror need not respond to a given TR below if the proposed capability does not address that problem area. However, a description of the overall system architecture, including execution model, is a mandatory requirement (MR) for all proposals. The most desirable proposals will be based on a specific execution model with a corresponding system architecture that supports that execution model.

In all TR responses that are provided, Offeror should discuss what progress will be made in the next two years and describe what follow-on efforts will be needed to achieve these goals fully. Offeror should describe in detail how metrics will be evaluated, including the measurement method that will be used (for example, simulation or prototype) and any assumptions that will be made.

A1-4.1 Overall System Architecture (MR)

Offeror shall describe their overall system architecture including node types, interconnection networks, and in-system nonvolatile storage. Offeror should include an alternatives analysis of expected component technology options in processors, memory, and interconnect technology that identifies gaps in current technology options and features that are needed from component developers for successful system integration. Proposed R&D should identify work that will address open questions, lead to identification of the best alternative, and mitigate risk.

Offeror shall also describe the execution model.

The proposed conceptual system design architectural investigation shall address the key challenges specified in Section A1-1 of this appendix. The proposed effort shall include:

- an evaluation of the proposed execution model;
- the development of the conceptual system design;
- an analysis of the proposed design that shows the impact of the design on the key challenges; and
- initial metrics for evaluating the success of the design.

A1-4.2 Energy Utilization (TR-1)

The target requirement is a system that achieves high performance on a broad range of DOE applications while minimizing energy use. Solutions should target 20 MW (peak) at system scale while maintaining or improving system reliability.

A1-4.3 Resilience and Reliability (TR-1)

Mean Time to System Interrupt (TR-1). Processor designs should make advances that lead to a mean time to a system interrupt requiring user or administrator action of one day or longer in an exascale system.

Fault Recovery Overhead (TR-1). The overhead to handle automatic fault recovery should not reduce application performance by more than half.

A1-4.4 Data Movement through the System (TR-2)

The performance of applications depends upon many factors such as message injection rates and contention. Offerors should describe improvements to the rates of data movement through all layers of the data hierarchy that optimize application performance.

A1-4.5 Density (TR-2)

To keep the system size and facility infrastructure at a manageable level, an exascale system should fit in a floor space of 15,000 sq. ft. (between 100 and 300 cabinets), with cabinet weights less than 500 pounds per square foot. Offeror should describe packaging, power delivery, and cooling designs that achieve these goals while minimizing computer room infrastructure costs.

A1-4.6 System Software (TR-1)

Solutions will need an integrated system software stack. Offeror should describe the components in its system software and how these will be improved to handle the scale, resilience, and power management challenges of an integrated exascale system.

A1-4.7 Programming Environment (TR-1)

Solutions will need an integrated software ecosystem that supports the development of new applications, the migration of existing applications, application maintenance, application resilience, and application portability, while enabling DOE scientists to achieve high performance with no more effort than is required for today's high-end supercomputers. Offeror should describe in detail how this will be accomplished and the improvements to be made.

A1-4.8 Data and Workflow Management (TR-2)

Solutions should include a model for managing the execution of end-to-end scientific workflows on the system. Offeror should describe the data management capabilities of their system architecture, including workflow scheduling and integration with system-level RAS framework.