INDUSTRY DEVELOPMENTS AND MODELS

The New U.S. National Strategic Computing Initiative

Robert Sorensen
Steve Conway
Earl C. Joseph, Ph.D.

IDC OPINION

The HPC User Forum was held September 9-10, 2015, in Denver. It featured the first public forum on the National Strategic Computing Initiative (NSCI) to be authorized by the executive order that was issued on July 29, 2015, and signed by President Obama. The topic was covered in back-to-back panel sessions moderated by IDC's Bob Sorensen. The forum included officials from the lead agencies for the NSCI: the Office of Science and Technology Policy (OSTP), the National Science Fund (NSF), the Department of Energy (DOE), and the Department of Defense (DOD) that were directly involved in developing the NSCI concept as well as senior HPC officials from national labs and centers that will play major roles in implementing the NSCI. The key findings from this event are as follows:

- The NSCI executive order raised the profile of HPC as a strategic, game-changing technology for both science and U.S. industry. However, the time needed for the NSCI to bear fruit clearly extends well beyond the term of the incumbent U.S. president and the funding ambitions of a single administration. To accomplish its long-term goals, the program early on needs to generate significant credibility in addressing real national security concerns — from both a defense and economic perspective — to garner a continuous stable and, perhaps most important, predictable stream of federal funding. If agencies do not get additional funds that are strictly applied to this project and instead are forced to essentially relabel existing funding under the NSCI "brand," then this effort will ultimately have a low chance of having any real impact.
- The speakers in the NSCI sessions clearly conveyed the importance of the initiative and the associated challenges, although there was little reference to the executive order's directive to work closely with industry — which includes not just the traditional HPC supplier partners but also the full range of existing and potential partners in both industrial and service sectors that could benefit from using HPC capability to help drive their global competitive prospects. IDC cautions that if the lead agencies treat the NSCI as merely an intensification of the status quo instead of a deliberately disruptive mandate to serve industry as well as science, the full potential of the initiative will not be realized for the United States.
- The NSCI acts as a kind of "moon shot," and the order is merely the first step in what will be a very long journey in trying to generate a national HPC strategy that has two very ambitious and perhaps conflicting goals — ensure that the diverse range of U.S. government (USG) agencies have access to world-class HPCs to meet specific missions, and help build a robust commercial HPC sector that can supply those systems to not only help critical U.S. government missions but also to help the United States remain the leading supplier nation in an increasingly competitive global HPC sector. This means that much of the technology and subsequent products developed under the auspices of this effort will need to be successfully marketed and sold throughout the world.

IN THIS STUDY

This study summarizes the two back-to-back NSCI sessions that took place during the HPC User Forum that was held September 9-10, 2015, at the Omni Interlocken Hotel in Denver. This document will provide a transcript-like recap of the key points made by each lead agency representative.

NSCI Panel 1: The Lead Agencies — the OSTP, NSF, DOE, and DOD

In this panel session, representatives of the lead agencies discussed the executive order of the NSCI. The speakers of this panel session were as follows:

- Bob Sorensen, IDC's research vice president and HPC session moderator
- Randy Bryant, White House Office of Science and Technology Policy
- Irene Qualters, National Science Foundation
- Doug Kothe, Oak Ridge National Laboratory (DOE)
- Will Koella, Department of Defense

Bob Sorensen, IDC: Introduction

Sorensen, the moderator, introduced the topic and speakers. He noted that the NSCI executive order lays out broad, ambitious, multiyear goals. Among these are the challenges to merge exaflops performance with exascale data and to develop technologies that push beyond the Moore's law slowdown. Perhaps the most important challenges, in Sorensen's opinion, are accelerating scientific discovery and economic competitiveness through active HPC-centered collaborations involving government, industry, and academia.

Randy Bryant, White House Office of Science and Technology Policy

NCIS will rely on the strengths of the 11 federal departments and agencies involved to move the United States into a position that sharpens, develops, and streamlines 21st century applications. The three lead agencies are the National Science Foundation, the Department of Defense, and the Department of Energy. On July 29, 2015, the White House issued an executive order that addresses the fact that there is currently no coordinated activity for dealing with our nation's HPC needs. The main goals are to advance scientific discovery and promote successful industrial and commercial development.

The National Strategic Computing Initiative has five strategic objectives:

1. Create systems that can apply exaflops of computing power to exabytes of data
2. Keep the United States at the forefront of HPC capabilities
3. Improve HPC application developer productivity
4. Make HPC readily available
5. Establish hardware technology for future HPC systems

There is a big gap to bridge, which is combining traditional HPC numerical modeling and simulation (M&S) with data analytics. This will mean looking at what's inside the largest systems — not just hardware but also the operating systems, the runtime systems and how they're programmed, and the philosophies of how they're built and operated. The challenge is how to reach the convergence of the two systems, simulation, and big data.

There is a focus on exascale and the need to maintain the United States' lead. There's much more to a machine than how it runs on one benchmark, and there's much more to the value of HPC than what our best machines can do. We're not just going for one machine to crack a benchmark. We're also building our capacity to have machines of different classes that are being put to use for the goals of scientific discovery and economic competitiveness. This is much more than just a race to the top.

Another focus area is programming challenges and other software challenges for HPC. We've moved backward in this area. The introduction of GPUs into supercomputers has meant we have to program at multiple levels of these machines. Smart people are making heroic efforts to make things run fast on these machines, only to have the machines replaced, which then means they have to rewrite large parts of code and migrate things up and down the stack. This makes it especially hard for smaller companies to make good use of HPC resources (Bryant gave the example of how wind flows over a bicycle frame). We need to find techniques and tools to help with more efficient applications (i.e., the efficient operation on HPC through different mechanisms).

There is a big issue with access. We want more than improved access for big players that can already afford to own and operate their own hardware. We want improved access across the board: physical access to resources plus expertise for knowing how to use them. We need a software developer's and an application user's perspective. We need people to understand modeling and simulation and data analytics to be able to use the machines effectively for solving problems. We need a much more robust marketplace, whether it's cloud or something else. We need a vital commercial basis that will be successful and sustainable.

What will happen when silicon CMOS comes to the end? There are only so many molecules you can assemble together. Scattered research is under way to look at what the future technology might be. Will it be some variation of CMOS? Carbon nanotubes? A shift to quantum computing? Will it involve cryogenic operation? We may need totally new models of computation.

None of these options are anywhere near ready to take over the computing we do today. They're far from commercial sustainability and large-scale use. The implications aren't just for device builders and hardware people. This could also have an effect on computer architecture, programming models, software development, and more. The federal government is the best place to invest in precompetitive, fundamental, long-term research like this.

What would success look like for NSCI? In 10 years, we want to say, "Yes, we got things going … Look how much progress we've made … We have a map for the future." We want to be advancing on the five objectives (previously mentioned).

Irene Qualters, National Science Foundation

NSF's role is about scientific discovery advances along with the broader HPC ecosystem and workforce development. NSF is focusing on objectives 2, 3, and 4. Although, I also want to say that we're in the very early stages of objective 5 and are interested in discussions about how we should approach it. Within 90 days of July 29, 2015, we have to have our first implementation plan, so we're working on that now.

This isn't an IT initiative in NSF. This is an "advancing science" initiative. Establishing a viable path over the next 15 years for future HPC systems in the post-Moore's law era is a critical strategy issue for the nation, involving other agencies such as the Intelligence Advanced Research Projects Agency (IARPA) and the National Institute of Standards and Technology (NIST).

It's not just a hardware effort. It's also about rethinking the software activities and issues. We need to increase coherence in the world of modeling and simulation (i.e., stepping back and looking at things from a science perspective). I've used data science in place of data analytics, giving examples of areas where science cannot advance unless we make progress.

Looking at the Arctic mapping project as a mini case study (Qualters showed a video), there are many images of the Arctic and many interests, but the images are all at different resolutions. Trying to get a first high-resolution map is a nontrivial activity. The systems we're using today are not necessarily designed or conceived to optimize for image analysis. We need new algorithms and new approaches to compression.

We're trying to marry the capabilities of modeling and simulation, often in a dynamic workflow with data science attributes. This has profound science implications and will be a focus area for us. Data science is emerging and is inherently multidisciplinary. For this, we need all of the skills at our disposal.

The 2010 Drew Conway image of data science resonates with the NSF community in that data science is emerging and is inherently interdisciplinary. Its pedagogical home is unclear, but you need all the elements — the foundational underpinnings, statistics, and mathematics. You need the domain knowledge to know which questions are relevant and the technical skills, or engineering practice, to know what can be done.

The context of science is changing. Everybody's interaction with technology is profoundly changing the way everyone works. Science is no different. Examining this NSCI activity in that bigger context is profound and strategic. We need to start the dialogue about how the nation's competitiveness can be enhanced.

Right now, our implementation is more on the second bullet. This activity is being governed at the highest NSF level to the heads of directorates. We have a cross-directorate group working on initial planning. We're establishing vehicles for community input, and those will be through our public advisory committees. We also have a national academy study looking at the future for advanced computing infrastructure.

Doug Kothe, Oak Ridge National Laboratory, Department of Energy

The DOE role is embodied within one project known as the exascale computing initiative (ECI). The exascale research and funding in the DOE has been going on for many years. This is not a "new start," except that it is being "projectized." That means we're following a fairly rigorous process for large capital acquisitions. This ECI is a 10-year project, formally starting on October 1 of each year. DOE will contribute to all NSCI objectives. DOE needs to and will deliver breakthrough science, energy, and national security applications that are capable of exploiting exascale systems to deliver new insights and solutions.

We're in the process of CD-0, or critical mission-zero, which is mission-need approval. The next step would be CD-1, which gets into the acquisition of strategy, the alternatives analysis, and having what's equivalent to a baseline project plan.

Application development, which is my role, is going to emit dozens of applications that will be in exascale already. Traditional applications are included in our scope; for example, simulating nuclear reactors has been a "killer app" for decades. On the application side, an RFI was submitted by the DOE to all DOE labs in early June 2015 for white papers about what applications can be, the impact urgency, and challenges. We've received 130 applications and more are still coming in. Another RFI will go out this week to other agencies such as the NSF and the NIH. We're gathering information to help formulate and influence plans.

There's a software technology piece and a hardware technology piece of this in the DOE's FastForward. The plan is to deploy two or more diverse exascale systems within 10 years. The project is expected to run through 2025.

Will Koella, Department of Defense

The DOD has a large role across the board within our agency and also within the HPC MOD program. We have significant investment in the HPC field across all of the NSCI objectives.

The best description of our role is to be an advocate for the requirements process. We're ensuring that national security concerns will be incorporated, but in general, those deployment agencies that don't have the big budgets or big deployment requirements are going to be incorporated.

Whether we're talking about law enforcement, the intelligence community, and the military community. Their requirements will be incorporated on the analytics side. The DOD will get large mission requirements in that space.

We are the lead development and deployment agency for objective 3. We're starting much further behind when compared with the exascale part. We are in the assessment phase, looking at where we should even be investing in the first place. Investment across the board is defused. There are lots of promise coupled with lots of risk.

A common characteristic is that there are a lot of science miracles that need to occur for every one of the pathways before we could even contemplate a mission-capable device in any of those systems. We have a long history of being on the front end of that, so it was natural for us to be a nexus of national security.

We really fit in on the bleeding edge of HPC research and development. We're also working across the DOD, making sure to address the requirements parts for HPC MOD, certainly on the modeling and simulation side. The DOD has a lot of interest in exascale.

A lot of the NSCI at the implementation level is going to be about investment portfolio management. We are notorious for not talking in the federal government, so we're trying to align our investments. For us, that means, particularly for NSA, that we're going to "open our kimono" to the DOE and the NSF. Together, we'll look at all investments to see if we can optimize those across the board. We'll focus as best we can within our own agency budgets as the FY16 budgets are already set. We'll start aligning for FY17 and really focus on FY18 as the year that we'll be fully addressing the gaps. Irene and I are "getting out our cat-herding chaps."

The NSCI is one of the more elegantly put together and well-founded initiatives I've seen come out of the government. We get excited about laying the foundation for economic advantages, security advantages, and scientific discoveries.

We will be supporting the DOE's effort; it's taking the lead on objective 1 and delivering the exascale system. We will make sure the HPC MOD program and the rest of the community, especially the modeling-simulation and scientific community, can get those requirements in and help accelerate the program.

Objective 2 is an area where we are going to take a balanced role between us, the NSF, and the DOE, bringing up the data analytics side, which is of great importance to us. Traditionally, we've been in a more specialized area of HPC, but high-performance data analysis (HPDA) is a growing area for us, making sure that the data analytics for the intelligence community (IC) and NSA problems for the military and law enforcement communities are addressed.

There are a ton of applications out there, and we need to make sure that they get incorporated, but we have ecosystem issues. Our high-performance computing missions are increasingly taking place in the more open intelligence analyst production environment, so those need to be tightly integrated.

The Executive Council met on August 26, 2015, and it laid out a couple of the things that are in objective 3, an area in which we'll take more of a leadership role on the development and deployment side.

There will have to be a lot more effort in identifying areas of potential investment for post-CMOS and post-Moore's law. The Executive Council is going to be focused solely on where we will be investing across diffused technology paths with a goal of trying to identify a candidate path or paths. We will be trying to identify something within five years as an alternative to CMOS.

Because traditionally there is a long lead time for some of the technologies we're looking at, we're taking a parallel approach to try and get the most out of Moore's law. The initial focus will be on technologies that will extend the classical, digital-computing architecture and framework for as long as we can. We're hoping to get 10 years out of it while simultaneously looking for alternative paths.

It won't necessarily be one-to-one replacement. Currently, we don't see anything to replace classical, digital computing for a lot of applications; but there may be some specialized places, whether it's neuromorphic, chemical- and biological-based computing, or quantum computing, where it makes sense to take a different direction.

One of the most important aspects of this initiative is that we want to be the lead in developing "the next big thing" for 15 years from now. So those new architectures mean we want to make sure that, even though they're developed for the federal government, we have a transition plan for the commercial sector. Objectives 4 and 5 are a little outside the NSA's comfort zone in terms of dealing with the public. We don't like to talk, right? It's hugely important that we work with everyone to facilitate the transition when it does ultimately occur.

If you read the executive order, one thing that is not explicitly noted — but is very important — is that national security is clearly an application space where we need help. We have a clear leadership role here because we bridge the gap between the DOD and the IC and data analytics team. I think Irene and I will be the busiest because we'll be jumping across the boards.

The hope is that this will be an "Apollo" project for the nation. In other words, we need to ignite the public's imagination around computers. How can we capture the challenge in tangible ways? Do we talk about finding the cure for cancer? Putting a man on Mars? We cannot make these strides without supercomputers, but the public doesn't get it. How can we be evangelists for HPC?

As a former congressional staffer, I'm skeptical that the federal government will come down from the mountain with "ten commandments" about how the public sector can contribute. There's an implicit leadership role for the community in industry and academia, and we hope you will help us find a way.

NSCI Panel 2: National HPC Centers and Laboratories

In this panel session, the representatives of national HPC centers and laboratories that will play major roles in implementing the NSCI plan presented their perspectives in more detail. It was set up as an open forum to discuss the NSCI and its direction. Sorensen also acted as moderator for this session. The participants were as follows:

- Bert Still, Lawrence Livermore National Laboratory
- Piyush Mehrotra, NASA Advanced Supercomputing Division
- Bill Kramer, National Center for Supercomputing Applications (NCSA) (NSF)
- Nathan Baker, Pacific Northwest National Laboratory (PNNL) (DOE)
- Rob Leland, Sandia National Laboratories

Bert Still, Lawrence Livermore National Laboratory

Things are evolving, which requires us to have better high-fidelity models than we have now, so we need more computing to be able to compute to those models. The problem has never been how much computing it will take to get there; the problem is that the world doesn't stop changing. We have large problems that need to be solved, and they require models and computing that are beyond today's reach, which is what's driving our interest in exascale.

We need to find a way for the public to latch onto this, rather than just keeping it to the HPC community only. We need to energize an entire generation of people.

From the time of the ASCI 20 years ago, we began working with vendors, academia, and even other countries. We have a cooperative relationship, for example, with the British on how the algorithms, models, and technology would change and how we would take advantage of it.

Our integrated codes are multimillion lines, and they were across the three labs within the NSCI complex. It's a $6 billion investment that's been put into these codes. We can't rewrite them overnight. They also can't take advantage of each individual architecture that comes up because it would need a decade or more to make full modifications to reengineer one and revalidate it.

Performance portability is key. Single-thread performance is not increasing, so we're simply increasing parallelism and the physics limitations of how to cool and distribute power among the parts that are there. The paradigm shift is from how fast you can crunch the numbers to how fast you can feed the chips with data. This, more than anything, will change the way we have to do our computing.

Piyush Mehrotra, NASA Advanced Supercomputing Division

We host the premier supercomputer for all of NASA. NASA is considered a deployment agency, so it will continue doing what it has been doing to provide a productive environment for users. We want to participate in the whole process to make sure NASA mission requirements are part of the codesign process of looking at new technologies and the whole software stack from the OS up.

Productivity is key to what we need to get to, so we're interested in large-scale applications and a variety of domains, starting from aeronautics design to aerospace vehicles to earth science to astrophysics to planetary science.

Regarding objective 2, we're interested in large-scale data analytics and the convergence of data analytics with M&S. There are lots of satellites, and we're bringing out petabytes of data every year as we stream observational data. How do we handle it all? How do we extract insights and knowledge? We need to bring the two environments together to do large-scale simulations along with large-scale data analytics. We will look at technologies being proposed through the NSCI process and test them on applications.

We do have a quantum computer. Generic HPC machines are used as solvers whereas optimization is done on this quantum machine.

One of our charters is to be the smart buyer for the agency. We're always looking at new technologies and hope to look at those proposed through the NSCI.

Bill Kramer, National Center for Supercomputing Applications

I'm from the NCSA, and I'm involved with the Blue Waters project. The NSCI has great potential for making more progress and expanding that dramatically. Computing and analysis resources at the high and broad scale are universal implements. Science communities may have a telescope going in or a satellite for observations, but our computing and data analysis infrastructure is universal. If someone uses Blue Waters as the world's most powerful microscope to see a 10 or 12 times higher resolution of atomic processes and molecular processes, that's great. The NSF environment has by far the broadest and most diverse set of use cases for computing and high-end computing I've been involved in.

There's an increased range of uses and needs, from cancer through space telescopes, but other trends are a dramatic increase in fidelity in models and also the analysis. Fidelity refers all the things we talk about in higher resolution for climate, more particles for certain types of simulations, and more precise measurements. So we see a great increase in fidelity, which is driving our need for more computing over time and more data; but those insights address new problems.

Another thing is longer simulation periods. So even though Blue Waters and the other leadership class systems are doing things that have not been done before, we have teams making strong compromises. For example, a full space weather calculation would require 15 orders of magnitude performance improvement to address what happens with a solar flare and how that affects life on Earth. Scientists today can simulate one-tenth of the time period for such an event. They've gained great insight; but to really understand an event like this, you need to do 10 times more just in time.

There is an increasing number of problems to address in two forms, and they need analytical resources or computing resources. Also, what happens when you have a frontier implementation of a problem? What if you go to 100 million or 200 million atoms? As we go through and see people producing results, we see more requirements. Changing work methods is very important, so we also focus on productivity, meaning time to insight.

It's very important to be efficient at a large scale because it may take 90% of the computing resources to do a large-scale calculation. If the elapsed time for a team to understand what happens is a month, most of that is spent dealing with the data that's produced or the data that has to be ingested and stimulated and moving that data around. Then it takes many, many steps to analyze that.

Example

You may have tens or hundreds of thousands of hours applied to do a stimulation, but then you may also need to have a 100 million jobs to analyze that. What we've seen is that making those teams productive really means their entire workflow — not just the part that's highly parallel and runs on the bigger system. There's a lot of work that is potentially very productive in looking at workflow methods.

The integration and convergence of data sources — as well as stimulation or modeling resources — are part of the initiative, and that's extremely important. We see that in almost all domains of research, engineering, and science that you cannot do one without the other. It used to be one person on models, another person on experiments, and then there was a validation. These are much more tightly coupled in all domains than they were even 10 years ago. That means the systems we will be producing have to be able to accommodate that simultaneously, and the implications are much more on the software side than the hardware side. The layers of hierarchy that we're going to have to deal with is a tremendous problem for the applications space but also for the systems space, in terms of doing that. Our systems need much more flexibility because not only do you have these different methods in workflows but you also have different cultures that are going to converge on a set of systems, and they're going to need to use the same type of systems to be productive.

We see a tremendous opportunity for layers of software, not just on the application side of making use of an application domain but also for managing systems and running work on the systems. This would bring together the different methods and models for very large-scale activities in both the data analysis and the data science realm as well as the modeling and computation realm.

Examples of large instruments are easy to point out. The experiment modeling black holes for gravitational waves developed a very tight relationship with the modelers. Without simulating, they won't be able to tell whether or not the signal we see would have a nonparallel spinning collapse of a black hole. These are occurring all over the place, in terms of not being able to distinguish one versus the other, and the teams are realizing that. It's very challenging for what we have to do in the future, but there is a tremendous opportunity to have a very broad impact.

More than 65% of the use of Blue Waters is by people in early the stages of their careers, either graduate students or postdoctoral students. It's not the most senior people. It's very important to bring in more types like that, so people are learning at a small scale and beyond. As they develop their methods, they learn how to do things small scale, which enables them to do things at a very large scale early in their careers that make an impact.

How many partners are on the industrial side? In talking about industrial benefits in the NSCI, we hope it's not just to the vendors of technology. There will be benefits there, but it's actually the other commercial industrial partners that we see. In trying to work with these teams, we discovered that you can do the very first part of a demonstration example of what might be possible for them, either by scaling something up to increase their workflows or reduce times to insight. Then there's putting it on a production floor as they use it in day-to-day practice. But there's a gap that I don't think I realize, in terms of what it takes to get companies to go from, "yeah, I know I could do that if I had a resource," to "I'm actually using that, and it's a change of work methods." That gap is an awful lot of computing plus an awful lot of data analysis that has to be done. It's not the first time it's been done, but it's also not there in their normal business practices. It will have to change the culture in order to change the business practices of many of the people that potentially or currently use high-performance data in computing resources. We have to figure out how to enable them to meet that gap because they're not going to do it until they know that they can actually improve their productivity, their product-line development, and their products. So that's a gap that hopefully we can also address in this initiative.

Nathan Baker, Pacific Northwest National Laboratory

I come from the applied math, physical chemistry, and computational biology perspective, and these are areas that have been struggling with data problems as well as computing problems for quite some time. They have really been looking for long-term solutions.

I'll just focus on a handful of the objectives — what some of the application pull is in the spaces and then what I see as some of the promising technologies there. So probably the most obvious, especially form the DOE perspective, is the delivery of an exascale system.

A number of applications could benefit, whether it's the idea of computational microscopes in observatories or just developing better models for thinking about integrated systems. How do our power systems integrate into climate? How do they integrate into other critical infrastructures? These are necessary and computationally stiff models in which some parts of the model might need to run for decades while other parts of the model are at the level of user demand on a power grid. These are hard problems, both data intensive and compute intensive.

PNNL is working on these from a variety of standpoints, but some of the issues we've been focused on are different from what have been described so far. Power was mentioned as a challenge, and power is always going to be an issue. But thinking about being able to model a system up front, if we get one of these big exascale systems, may be different. How do we remove the burden from the user to think about uptime, downtime, deployment, integration across processes, fault tolerance, and so forth and then actually integrate that into our programming models?

A second area is simply thinking about algorithms differently because there will never be enough computing or enough storage to tackle the problems that we want to tackle. A lot of the computing work at PNNL has been focused on the data-intensive space, and many of the algorithmic advances have been focused on what we do when we don't have the needed resources. Need is a great driver for innovation. Thinking about approximate computing, anytime algorithms, and ways to get resources to do a good enough job — given the compute or the storage that is available — is an important element to this overall initiative that needs to be explored.

Objective 2 really resonated with us because we've been focused on this data intensive, data science, and data analytic piece for quite some time. The NIH saw this need a long time ago. The National Cancer Institute and a variety of others recognized that the data was growing at a rate in which getting it into the hands of a practitioner for decision making was becoming impractical. This pops up in other domains and the security domain, whether you think of intelligence community analysts or power analysts who are actually thinking about critical infrastructure. We have too much data, and there's a big gap between the data and knowledge.

Some of the programs at PNNL involve high-energy physics. The amount of data that comes off of a big scientific instrument or one of the big detectors is too high of a bandwidth to even run out to a disk. You've got a "baby with the bath water" conundrum. You're spending billions of dollars looking for rare particles, and yet the data is coming out at a rate that you have to triage, and you may lose what you're looking for.

How do you design robust algorithms that can handle that? How do you design algorithms that can detect what you need to detect? There's a big driving force from the NIH, basic energy sciences, and the DOE to develop better instrumentation so we can get down to finer scales and add dynamic, time-dependent information to that instrumentation. At this point, we can't process the data coming out. It's very analogous to the DHS problem of too many cameras and too many sensors. How do we push computing to the edge with these instruments so that we're really transmitting knowledge and information more than raw data?

Regarding objective 3, there's a really unique role for collaboration across the labs, DOE, and NSF. The basic energy sciences in the DOE, NSF, and others have contributed a tremendous amount of investment to the material science world. A lot of it has been focused on CMOS technology, but there's tremendous opportunity out there to start taking on the very advanced characterization capabilities. I mean the ability to place atoms where they are needed through resources at PNNL, Los Alamos National Laboratory, and Sandia National Laboratories and to start asking about the next-generation problems in fabrication. How do we couple those to characterization, then think about modeling and algorithms differently?

We're going to have to be thinking about computing in the presence of noise and imperfect materials. We can't let perfect get in the way of progress.

Rob Leland, Sandia National Laboratories

Thanks to Bob (Sorensen) for the gracious introduction and saying that I'm the father of the initiative. Bob was part of the Council as I spent a year at OSTP working on this about two years ago. Bob was a member of the Council, as was Irene. Randy (Bryant) was, of course, a critical part of the team. I appreciate the depth of Bob's insight and the sense of conscience he brought to bring us back to the relevance and the broad relevance to society. Bob had a previous life in government, so I wanted to thank all my colleagues up here. I know much of the content that we worked through was influenced by Bert (Still) and Bill (Kramer) and many people in the audience, so it's really been a community effort. I'm delighted that it's in such good hands now with Randy still on board. I worked with Doug (Kothe), and much of the context here was inspired by some of those interactions. It really has been a community effort, and that's been quite important.

The second thing I should say is that I'm just going to speak from a Sandia perspective, if you like, and not from a DOE or OSTP perspective. The initiative is unfolding pretty well. It's carefully designed, and we had a lot of input in the development of the initiative. There's a sense of ownership and cooperation across the agencies that will be important.

Devolution is the risk. Of course, there's some very substantial technical risk here. But I am confident that if the U.S. government brings forth its best effort in partnership with industry and academia, we can do these things. The main risk is that it might unravel.

There is a detailed, joint road map that the agencies worked on that will likely be reflected in some form in the implementation plan. In this plan, agencies agreed on the vision or strategy, goals and objectives, rules, and responsibilities.

The potential impact of government-sponsored HPC research and the broader HPC sector is that the overarching goal of the initiative is to ensure continued U.S. leadership in HPC. The government has historically played a vital role in sponsoring forward-looking research and development. The initiative creates the conditions for that to continue into the future.

This initiative will help a wide range of U.S. industrial HPC users and boost the overall competitive position of the U.S. industry. Each new major era in computing has been preceded by five to seven years of a forward-looking investment by the government R&D push. You can trace that back at least five cycles.

There are many indicators that we're approaching a wall where we need to take a very substantial step up in our capability and make a change in our approach.

This is a huge opportunity for the agencies and the community to do something that's really important for the country, and I would just ask you all to join in and make that a reality.

Questions and Answers Discussion of the NSCI Agency Panel

The question and answers (Q&A) session participants included many of the national lab and center officials that made up NSCI panel 2. Their role in the Q&A session was to react to the remarks of the panel 1 agency speakers. The Q&A participants were as follows:

- Bert Still, Lawrence Livermore National Laboratory
- Piyush Mehrotra, NASA Advanced Supercomputing Division
- Bill Kramer, National Center for Supercomputing Applications
- Nathan Baker, Pacific Northwest National Laboratory
- Rob Leland, Sandia National Laboratories
- Earl Joseph, executive director of the HPC User Forum, IDC

Initial remarks noted that NSCI participants need to find a way for the public to become engaged instead of just keeping it within the HPC community only. The initiative needs to energize an entire generation of people. It's very hard to achieve coordinated efforts in the federal government. There are big challenges to address. No one agency has the sole lead on any of the objectives, but Rob Leland — more than anyone else — is the father of this initiative.

Bert Still: When it comes to large-scale computing, we know what the issue is — some things are missing, and other things are evolving quickly. The problem is that the world doesn't stop changing. Understanding extrapolation is very, very challenging. It's what drives interest in exascale.

Piyush Mehrotra: We host the premier supercomputer for all of NASA. We need to be the smart buyer for the agency.

Bill Kramer: The implications are much more on the software side than the hardware side. Systems need much more flexibility. There are tremendous opportunities for layers of software, not just on the application side but also on how we work on and manage the systems.

Nathan Baker: We need to address cross-domain opportunities such as in applied math and computational biology. We are looking for long-term solutions, solutions that last for a very long time. A number of applications could benefit both data-intensive and compute-intensive problems. What do we do when we don't have the resources that are needed or the resources to do a "good enough" job? There will never be enough computing or enough storage to get what we need. We need to get data into the hands of practitioners in a reasonable way. How do we design algorithms in a good way? How do we push computing to the edge so it's creating new knowledge and information? We have a unique role for collaboration in basic energy sciences, and we have contributed a huge investment to the materials science world. There are tremendous opportunities to place atoms where they're needed and to start asking the next generation of questions.

Rob Leland: The NSCI is carefully designed with lots of input from and collaboration among the agencies. It sets the stage for the U.S. government to make its best effort with industry and academia. A joint road map will be reflected in the implementation plan. The overarching goal is to ensure continued U.S. leadership in HPC, and the government is playing a vital role in that. How will this help the wide range of U.S. HPC industrial users? If you look at the history, each major advancement in computing has been preceded by government leadership. The NSCI presents a huge opportunity for the agencies and the HPC community to do things that are important for our country.

Earl Joseph: This is not just looking at flop-based machines. There's a need for a new set of metrics beyond flops. The LINPACK flops metric has value, but it cannot be the only measure to determine investment choices. No single metric is adequate for all the problems. We need to come to a community understanding of the set of measures that are meaningful. This belongs to objective 1 (refer back to the Randy Bryant, White House Office of Science and Technology Policy section), focusing on measurement in the true application space.

Other Comments

The following are other key points made during the Q&A portion of the forum:

- NIH is still trying to understand what HPC can do for them. Lots of data doesn't involve complex computation. We need to get information that will help cancer researchers practice medicine. What range of computational needs might they be able to take advantage of?
- The space program benefitted the country a great deal; but 45 years later, we're only beginning to have unmanned successes.
- The main pitfalls of the initiative are nontactical. In the current budget environment, it's hard to find funding from the government for scientific research. Apollo (the 1963-1972 lunar mission) consumed a huge percentage of GDP. In our current budget climate, that would simply not be possible. Nothing will make us step up to that level again.
- Another big risk is devolution between the agencies. Flat funding is the new "up." The U.S. growth rate in this area is 2% — low when compared with 23% in China — which will limit our ability to dominate. Much of the investment is in the private industry.
- The erosion of Moore's law is another obstacle. We need to rally as a society around that challenge. There's also good indicators that we're coming to the end of the MPP era. If we don't make a change to our architectural path, we could falter. The microelectronics industry is moving offshore more and more, and this affects our ability to command attention and have coherence in the overall ecosystem. The NSCI attempts to address all of these things.
- Another elephant in the room is that we don't have enough people going into computing and math. As we retire, who will fill the void? Kids today have no idea how things are computed or don't care. We need to tell computing success stories that are emotional to capture mindshare and imagination and get people ready to do the jobs we need done. More than half of our risks are not technical.

SITUATION OVERVIEW

The NSCI objectives create multiple challenges in the following areas:

- **Funding:** Securing the money needed to develop and deploy an exascale computer
- **Technical:** Performance scaling of the hardware/software system, reliability, density, and power efficiency, among others
- **High-performance data analysis:** Processing massive data associated with simulations and analytics
- **Industry:** Extending the benefits of the initiative to America's industrial sector (well beyond HPC vendors)
- **Workforce:** Ensuring an adequate number of qualified job applicants and workers

The speakers in the NSCI sessions clearly conveyed the importance of the initiative and the associated challenges. Although, there was little reference to the executive order's directive to work closely with industry, which includes not just the traditional HPC supplier partners but also the full range of existing and potential partners in both industrial and service sectors that could benefit from using HPC capability to help drive their global competitive prospects. IDC cautions that if the lead agencies treat the NSCI as merely an intensification of the status quo instead of a deliberately disruptive mandate to serve the industry as well as science, the full potential of the initiative will not be realized for the United States.

In summary, the NSCI acts as a kind of "moon shot," and the order is merely the first step in what will be a very long journey in trying to generate a national HPC strategy that has two very ambitious and perhaps conflicting goals. The first is to ensure that the diverse range of U.S. government agencies has access to world-class HPCs to meet specific mission requirements that can include nuclear stockpile stewardship, advanced code breaking, homeland security, futuristic weapons developments, long-term climate modeling, and human health and genomic exploration. The second is for the U.S. government to help build a robust commercial HPC sector that can not only supply those systems critical to U.S. government missions but can also help the United States remain the leading supplier nation in an increasingly competitive global HPC sector. This means that much of the technology and subsequent products developed under the auspices of this effort will need to be successfully marketed and sold throughout the world.

FUTURE OUTLOOK

In total, the order is merely the first step in what will be a very long journey in trying to generate a national HPC strategy that has two very ambitious and perhaps conflicting goals. The first is to ensure that the diverse range of U.S. government agencies has access to world-class HPCs to meet specific mission requirements that can include nuclear stockpile stewardship, advanced code breaking, homeland security, futuristic weapons developments, long-term climate modeling, and human health and genomic exploration. The second is for the U.S. government to help build a robust commercial HPC sector that can not only supply those systems critical to U.S. government missions but can also help the United States remain the leading supplier nation in an increasingly competitive global HPC sector. This capability, in turn, must be effectively deployed to the range of U.S. private sector concerns in both industrial and service sectors to ensure that key U.S. businesses can rely on a robust domestic HPC capability to help them compete more efficiency in the global market.

Only time will tell which goal will be deemed more important. Skeptics who have seen earlier efforts to help rationalize U.S. government HPC efforts note that often the needs of the individual mission agencies trump those of the overall U.S. government strategic vision. Also, the inherent conflict with building the best systems for USG use while ensuring that those systems can also be made commercially available around the world — perhaps even to some countries that have U.S. national security concerns — has always been problematic and never solved to everyone's satisfaction. Current ongoing issues with contentious export controls issues, foreign R&D efforts to reduce foreign technology dependencies, high-tech supply chain and IP concerns, and continued acquisitions of U.S. technology by overseas firms will only further complicate this matter.

ESSENTIAL GUIDANCE

There will be a number of critical actions to watch to determine if this program can realize some of its ambitions, but very necessary, goals. First and foremost will be the need for this program to generate significant credibility in addressing real national security concerns — from both a defensive and economic perspective — to garner a continuous, stable, and perhaps most important, predictable stream of federal funding necessary to meet its goals. If agencies do not get additional funds that are strictly applied to this project and instead are forced to essentially relabel existing funding under the NSCI "brand," then this effort will ultimately have a low chance of having any real impact.

IDC is optimistic that such funding will be forthcoming — at least for the next few years — mainly because the NSCI marks a sea change within the federal government that now clearly recognizes that HPC capability at the highest levels will increasingly play a strategic role in shaping the future of our society and the U.S. political, defense, and economic leadership in the world.

From the more near-term tactical perspective, however, IDC sees that clear and decisive steps will need to be taken to move this effort forward, particularly in the area of developing enduring public-private partnerships in HPC R&D. Efforts here should include robust, close, and continuing partnerships where both sides have strong intellectual, personnel, and financial commitments. In addition, issues of IP or patent ownership that could slow or complicate the transfer of any government-sponsored R&D into commercialized products need to be ironed out early. IDC also recommends the following:

- U.S. government mission agencies must ensure that a wide range of potential HPC options are explored and risky or innovative technology developments are encouraged and indeed legitimized by a continual stream of procurements. Such efforts will require close coordination across all procuring agencies to ensure that the most promising technologies are not overlooked and that there is a careful balance between meeting critical mission requirements and inculcating an entrepreneurial spirit into the procurement process. New U.S. government funding mechanisms and related procurement processes may need to be adjusted to encourage such behavior.
- For its part, the HPC industry must be realistic about what technologies have the best chance for widespread commercial success and, perhaps more important, be more willing to take a long-term perspective on R&D projects and their demonstrable results, financial or otherwise.
- The hardware, software, and applications needs of a wide range of industrial and service sector users of HPC capability must be considered and included in any future efforts to ensure that the program goals and efforts are relevant to the competitive needs of this wide and diverse base of users.
- For the academic community, particular attention needs to be paid to generating sufficient interest in this important research area. As was mentioned in the panel sessions, too many HPC experts have noticed that the HPC community is having trouble attracting new talent into this field, and many say that some sort of "Apollo" program could generate new and needed enthusiasm within the academic sector. It could create a new cadre of HPC experts coming out of leading universities.

LEARN MORE

Related Research

- *White House Announces Strategic HPC Plan: A Good Start on a Long Road* (IDC #258194, August 2015)
- *Worldwide Technical Computing Server Systems by Industry/Application Area 2015-2019 Forecast* (IDC #256705, June 2015)
- *Rescale Shows HPC Cloud Computing Is Coming of Age* (IDC #lcUS25641315, May 2015)
- *Supercomputer-Based Analytics Are Starting to Transform Health Care* (IDC #256298, May 2015)
- *Newest ARM Development Targeted for the Enterprise Data Center* (IDC #lcUS25533715, April 2015)
- *Total S.A. Retakes the Lead in Energy Sector Supercomputer Race* (IDC #lcUS25528815, March 2015)
- *The HPC Arms Race Escalates in the Energy Sector* (IDC #lcUS25511215, March 2015)
- *HPC Interconnects: New Options Emerging from a Sector in Transition* (IDC #255184, March 2015)
- *IDC's Worldwide HPC (Technical Server) Taxonomy, 2015* (IDC #255159, March 2015)
- *Worldwide HPC Server 2015-2019 Forecast* (IDC #255147, March 2015)
- *IBM's HPC Efforts: Navigating a Renewed Emphasis on POWER Technology Within the OpenPOWER World* (IDC #253005, December 2014)
- *NEC's HPC Vision: Bringing Vector Supercomputing to a Broader Data-Intensive User Base* (IDC #252884, December 2014)
- *PayPal Says More Fortune 2000 Firms Could Benefit from HPC for Big Data Analytics* (IDC #252836, December 2014)
- *Bull Launches an Exascale Program for 2020, But the Vision Extends to Big Data and Beyond* (IDC #252771, December 2014)
- *Japan's Extreme Computing Efforts: System Plans Moving Forward, But Critical Applications Set the Way* (IDC #252706, November 2014)
- *HPC Fabric Wars Are Heating Up* (IDC #lcUS25269814, November 2014)
- *DOE Goes All in on IBM-NVIDIA for 300 Petaflops System* (IDC #lcUS25250414, November 2014)
- *Technologies and Applications of High-Performance Computing in the Cloud: HPC User Forum, September 15-17, 2014, Seattle, Washington* (IDC #251974, October 2014)
- *Emerging and Disruptive Technologies in High-Performance Computing: HPC User Forum, September 15-17, 2014, Seattle, Washington* (IDC #251970, October 2014)
- *Emergent Use Cases in High-Performance Data Analysis: HPC User Forum, September 15-17, 2014, Seattle, Washington* (IDC #251976, October 2014)
- *Experiences with Accelerators and Coprocessors in High-Performance Computing: HPC User Forum, September 15-17, 2014, Seattle, Washington* (IDC #251973, October 2014)
- *Lenovo Completes Acquisition of IBM's x86 Server Business* (IDC #lcUS25176214, September 2014)
- *USPS Touts Benefits of HPC for Big Data* (IDC #lcUS25123014, September 2014)

- *Worldwide Broader HPC 2014-2018 Forecast: Servers, Storage, Software, Middleware, and Services* (IDC #248835, June 2014)
- *Perspectives on High-Performance Data Analysis: The Life Sciences* (IDC #248348, May 2014)
- *Worldwide HPC Public Cloud Computing 2014-2017 Forecast* (IDC #247846, April 2014)
- *Micron Demonstrates Technologies to Address Emerging Challenges in Big Data Applications* (IDC #244843, December 2013)
- *China Eyes 10,000-Fold Data Reduction for Internet of Things* (IDC #lcUS24392513, October 2013)
- *High-Performance Data Analysis in the Life Sciences: HPC User Forum, September 2013, Boston, Massachusetts* (IDC #243774, October 2013)
- *Worldwide High-Performance Data Analysis 2013-2017 Forecast* (IDC #241315, June 2013)

Synopsis

This IDC study summarizes the HPC User Forum meeting that was held September 9-10, 2015, in Denver. It featured the first public forum on the National Strategic Computing Initiative (NSCI) to be authorized by an executive order that was issued on July 29, 2015, and signed by President Obama. The forum included officials from the lead agencies for the NSCI (the OSTP, NSF, DOE, and DOD) who were directly involved in developing the NSCI concept, and it included senior HPC officials from national labs and centers that will play major roles in implementing the NSCI.

"Ultimately, IDC believes that the NSCI could act as a kind of 'moon shot,' spurring U.S. HPC development and use, but this executive order is merely the first step in what will be a very long journey in trying to generate a national HPC strategy that has two very ambitious and perhaps conflicting goals: ensure that the diverse range of U.S. government agencies have access to world-class HPCs, and help build a robust commercial HPC sector that can supply those systems not only to critical U.S. government missions but to help the U.S. remain the leading HPC supplier and user nation." — Bob Sorensen, research vice president, Technical Computing

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com