This study was carried out for RIKEN by

Special Study

Analysis of the Characteristics and Development Trends of the Next-Generation of Supercomputers in Foreign Countries

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

IDC OPINION

Leadership-class supercomputers have contributed enormously to advances in fundamental and applied science, national security, and the quality of life. Advances made possible by this class of supercomputers have been instrumental for better predicting severe weather and earthquakes that can devastate lives and property, for designing new materials used in products, for making new energy sources pragmatic, for developing and testing methodologies to handle "big data," and for many more beneficial uses.

The broad range of leadership-class supercomputers examined during this study make it clear that there are a number of national programs planned and already in place to not only build pre-exascale systems to meet many of today's most aggressive research agendas but to also develop the hardware and software necessary to produce sustained exascale systems in the 2020 timeframe and beyond. Although our studies indicate that there is no single technology strategy that will emerge as the ideal, it is satisfying to note that the wide range of innovative and forward leaning efforts going on around the world almost certainly ensures that the push toward more capable leadership-class supercomputers will be successful.

IDC analysts recognize, however, that for almost every HPC development project examined here, the current effort within each organization is only their latest step in a long history of HPC progress and use. As such, IDC analysts assess that a leading-edge supercomputer development and user facility must be continually involved in the exploration of new system technology, or risk falling behind those consistently committed to the acquisition and use of leadership-class systems. IDC analysts believe that the cost of missing even one generation of HPC development could cause substantial difficulties for any facility looking to maintain a world-class HPC-based research capability.

Looking at the strengths and weaknesses in exascale plans and capabilities of different countries:

- The U.S. has multiple programs, strong funding and many HPC vendors, but has to deal with changing federal support, a major legacy technology burden, and a growing HPC labor shortage.
- Europe has strong software programs and a few hardware efforts, plus EU funding and support appears to be growing, and they have Bull, but they have to deal with 28 different countries and a weak investment community.
- China has had major funding support, has installed many very large systems, and is developing its own core technologies, but has a smaller user base, many different custom systems and currently is experiencing low utilization of its largest computers.

December 2016 #US42030316

IDC analysts stress that that time and again, successful national leadership-class supercomputer facilities have played an important role in underwriting new developments in hardware and software that prove to be applicable to a wide range of scientific, engineering, and industrial disciplines. At the same time, these programs provide significant support for their nations' domestic HPC supplier ecosystem, helping them remain at the forefront in global HPC technology developments. IDC believes that countries that fail to fund development of these future leadership-class supercomputers run a high risk of falling behind other highly developed countries in scientific innovation, with later harmful consequences for their national economies.

Finally, the HPC sector is currently undergoing a significant transition away from being limited to a relatively small HPC user base centered on scientific modeling and simulation applications and is rapidly embracing a broad span of new use cases. Some of these new uses will employ traditional HPC hardware and software. Many of the more innovative, forward leaning, and rapidly growing sectors could require the development of a wide range of new HPC technology involving new processor designs in both the CPU and GPU area, deeper memory schemes, more capable interconnects, and the wide range of associated systems and application software needed to provide a fully capable solution. Some of these emerging HPC-relevant sectors that IDC has identified include:

- A rapidly growing base of big data applications with complex requirements that can only be managed on HPC platforms in areas such as precision medicine, the financial sector, and cyber security.
- New demands for computational performance in data intensive Internet of Things applications as well as those that merge significant computational capabilities with mobile applications, such as voice or image recognition on smart phones.
- The profusion of new options for HPC in the cloud that will almost certainly bring a new host of users and applications to the HPC world as technical barriers to entry and the cost of running some HPC jobs in the cloud become less daunting barriers across a wide range of industrial sectors, especially for a number of small and medium business.
- Breakthroughs in deep learning and other cognitive computing areas that will drive the need for new HPC designs and expand the boundaries of new programming paradigms,

EXECUTIVE SUMMARY

This IDC report summarizes a number of leadership-class supercomputer developments on-going and planned within the next five years at leading supercomputer sites around the world. In essence, it is an attempt to capture in significant detail, supported by technical information, some of the most important -- and technologically forward learning -- developments in supercomputers from today out to 2020 and beyond.

There are a number of major leadership-class HPCs in the planning and development stage targeted for deployment between 2016 and 2020. In addition to Japan, such development is under way across a wide range of major HPC suppliers and regions including China, the EU, and the United States.

- Most of these systems are pre-exascale designs: systems that will underwrite much of the technology critical to the development of the hardware and software necessary to support the spate of exascale systems planned for the 2020 to 2022 time-frame.
- As such, the bulk of the systems planned for the next four years target a peak performance capability between 10 and 300 teraflops, with the bulk of the lower-end systems closer to completion this year or the next, while higher performance systems are targeted for completion closer to 2020.

Many Different Architectures Are Being Researched

There are a wide range of different architectural design paths to an exascale system.

- Some projects are looking to partner with a commercial vendor, such as Cray or IBM, to help them develop a leadership-class system that is in keeping with the overall product offerings of their commercial partner, such as the Cheyenne SGI system at NCAR.
- Others, such as NUDT's Tianhe-2 A group in China, are essentially looking to custom-build a system that likely will be produced in very limited quantiles, be used primarily in domestic markets, and developed with little expectation of eventual commercialization.
- In addition, it is clear that there is no agreed upon architectural scheme for these pre-exascale systems.

GPU and Accelerators Will be in a Number of Special Purpose Systems

But these designs in most cases will be less general purpose than designs that use a single processor type. A number of large sites around the world have been conducting studies to determine which codes fit best on which type of architectures. For the most general purpose environments, the use of a standard common CPU provides the broadest solution to the largest number of diverse researchers. Systems with a large percentage of GPUs are useful for specific types of codes, but result in more special purpose designs.

Power Consumption is a Major Concern

Concerns about power consumption and related efficiencies are keeping total power needs for pre-exascale system generally below 30 MW.

- In addition, many designers are looking for ways to better control the real-time power draw of their systems by using advanced techniques to shut down power-hungry portions of their systems when not in use during a particular job or even portion of a job.

Most Aren't Focused on Only Peak Performance

IDC analysts assess that designers, developers, and users of pre-exascale systems are not generally concerned with the theoretical peak computation performance of their new systems.

- Instead IDC analysts note that there is an increased emphasis on determining the ability of a new system to deliver a sustained performance - one that captures the ability of a system's overall compute, memory, interconnect, and storage infrastructure to execute an end-to end task - over one that stresses pure computationally capability.

A Wide Mix in Budgets for Developing and Building These Exascale Systems

Pre-exascale designers are operating under a wide range of budgets from a low of $25 million to well over ten times that amount.

- Some of the most expensive pre-exascale systems – such as the most technologically, aggressive one-off systems - are projected to cost $250 million or more. These systems represent some of the most advanced HPC developments in the world, and include significant non-recoverable engineering (NRE) costs.
- Others, primarily those that are one step behind the leading-edge of performance, generally are looking at budgets an order of magnitude less. Many of these systems have less aggressive NRE requirements and instead rely primarily on hardware and software technology supplied by their vendor partners.

Ease-of-use Will Require Major Investments

IDC analysts assess that for the most part, the leadership-class supercomputers under development have accounted well for the fundamental ease of use features that they will need for their best possible operation.

- This is not surprising as the increasingly higher costs of such systems almost automatically justify significant pre-planning and designer/user collaboration not just within the specific project, but more commonly across projects, some of which transcend bureaucratic as well as national borders.

Exciting New Hardware Trends

Notable hardware trends that emerged from this study included:

- There is a wide range of pre-exascale processors and related GPU accelerators being considered for inclusion in the various systems.
- The overarching trend in pre-exascale design is toward more memory, more SSDs, and the use of additional memory accelerators, such as burst buffers or high bandwidth memory packages as a way to deal with the increasing need for higher bandwidth and lower latency memory systems.
- For most of the systems that will be delivered soon, designers are opting for either InfiniBand, Intel OmniPath, or in a few cases, a custom in-house interconnect scheme.
- Leadership class supercomputer designs have overall storage requirements that are moving well into the 100PB range in the next few years.

Interesting New Software Trends

Notable software trends that emerged from this study included:

- Linux, in its many variants, has become the stock operating system for most leadership-class supercomputer, and IDC analysts assess that this will be the case for at least the next five years.
- Lustre and GPFS are and likely will continue to be the major file system software for leadership-class supercomputers for at least the next five years.
- There is modest attention being paid to non-traditional HPC software that IDC analysts expect will become increasingly important in the next few years, such as for big data infrastructures built around the Hadoop/Apache Spark (or other alternative) ecosystem and for virtualization schemes such as Docker.

Partnerships for Developing These Systems

Analysis of the major R&D plans and partnerships for most of the leadership-class supercomputers studied offers a few key insights worth noting.

- Some, such as those in the US DOE, are seeking to not only meet their near-term computational requirements but are also committing significant NRE to help lay the hardware and software foundation for exascale systems that are scheduled for completion in the 2020 to 2022 time-frame.
- Others are targeted more toward near-term computing requirements that do not include any significant commitment of NRE funding, such as the Swiss Piz Daint which instead looked to partner with a commercial vendor (Cray) to meet its less aggressive, albeit no less important, computational requirements with more traditional HPC architectures.
- Finally, there are some that are focused primarily as research systems, built in limited quantities, more for their value as HPC research machines than as production systems.

Partnerships are becoming a fundamental reality of leadership-class HPC development. There were a wide range of partnership types across the projects IDC examined.

- Some like the CORAL effort within the US DOE look to use the combined HPC hardware and software expertise of partnership members to examine a number of alternative HPC technologies while still meeting the particular mission requirements of each individual organization.
- Other partnerships look instead to a provide a rationalized development and research program across a number of sites to ensure that a range of design options are explored in a systematized fashion.

Likewise, almost all major leadership-class projects involve some committed partnership with one or more commercial vendors, be it at the component, system, or software level. Such partnerships can yield substantial benefits for both parties.

- The procuring lab is able to help design and purchase technology in cooperation with some of the leading HPC suppliers in the world that might never be available on the commercial market,
- At the same time, vendors benefit from a first-hand partnership with some most forward-leaning thinkers in HPC design, helping them better develop technology that can then be used in their wider product lines bound for the commercial sector.

Many National Programs Have Developed to Support These Efforts

The broad range of leadership-class supercomputers examined during this study make it clear that there are a number of national programs planned and already in place to not only build pre-exascale

systems to meet many of today's most aggressive research agendas but to also develop the hardware and software necessary to produce sustained exascale system in the 2020 timeframe and beyond.

- Although our studies indicate that no single technology development plan will emerge as the dominant scheme the wide range of innovative and forward leaning efforts going on around the world almost certainly ensures that the push towards more capable, powerful leadership-class supercomputers will be successful.

IDC analysts stress, however, that for almost every HPC development project examined here, the current effort within each organization is only their latest step in a long history of HPC development and use.

- A leading-edge supercomputer user facility must be continually involved in the development of new system on a timely basis or risk falling behind those committed to the regular, periodic acquisition and use of leadership-class HPCs.
- The cost of missing even one generation of HPC development could cause considerable difficulties for any facility looking to maintain a world-class HPC-based research capability.

Finally, time and again, successful national leadership-class supercomputer facilities have played an important role in driving HPC-based developments by underwriting new capabilities in hardware and software applicable to a wide range of scientific, engineering, and industrial disciplines.

- At the same time, these programs provide significant support for the domestic HPC supplier ecosystem to remain at the forefront of global technology developments.
- IDC believes that countries that fail to fund development of these future leadership-class supercomputers run a high risk of falling behind other highly developed countries in scientific innovation, with later harmful consequences for their national economies.

Strengths and Weaknesses in Exascale Plans and Capabilities of Different Countries

Looking at the strengths and weaknesses in exascale plans and capabilities of different countries:

- The U.S. has multiple programs, strong funding and many HPC vendors, but has to deal with changing federal support, a major legacy technology burden, and a growing HPC labor shortage.
- Europe has strong software programs and a few hardware efforts, plus EU funding and support appears to be growing, and they have Bull, but they have to deal with 28 different countries, and a weak investment community.
- China has had major funding support, has installed many very large systems, and is developing its own core technologies, but has a smaller user base, many different custom systems and currently is experiencing low utilization of its largest computers.

TABLE OF CONTENTS

TABLE OF CONTENTS — Continued

LIST OF TABLES

IN THIS STUDY

Methodology

Project Goals and Objectives

This IDC report summarizes a number of leadership-class supercomputer developments on-going and planned within the next five years at leading supercomputer sites around the world. In essence, it is an attempt to capture in significant detail, supported by technical information, some of the most important -- and technologically forward learning -- developments in supercomputers from today out to 2020 and beyond. This particular time frame coincides with when Japan's Riken's Post-K system is scheduled to begin operations.

As such, the report looks at supercomputer development plans for most of the major supercomputer developer and user nations, but it is heavily focused on those in the United States, China and Europe, encompassing a wide range of developers richly experienced in leading-edge supercomputer development as well those that are relatively new entrants to the field.

- The analysis captures many of the most salient technical R&D elements of the new systems that are planned for the next few years, and also considers the related commitment of financial and other resources committed the efforts.
- In addition, it attempts to compare and contrast the various national efforts on the bias of their technological validity and their potential for driving significant gains in furthering the state of the art in supercomputer development
- All work on this effort is to be completed before December 27, 2016.

Research Specifications:

For this study, IDC analysts collected a wide range of technical and other data relating to the planned development efforts of major HPC projects around the world, with a time frame that spanned developments currently underway and those with completion dates targeted out to 2020. The information and analysis on these projects collected from various foreign countries were based on a list reviewed approved by RIKEN. Key collection metrics included

- Specifications (type of the processor, peak performance, etc.) of the next-generation computers in research and development in other countries.
- Information specifics on the investments in the research and development of the next-generation computers already in progress in other countries.

IDC analysts used that data to glean key insights and projections on the significant development trends of the next generation of leadership-class supercomputers across a wide range of countries and individual development projects.

- IDC analysis was performed on the basis of the information collected through interviews and through the IDC information collection process.
- Discussions and analysis were conducted based on the reliability of the information collected. IDC indicated whether the information was solid, sound, speculative, or modeled.

IDC exploited its strong relationships with international HPC users, vendors, and funders around the world to present a detailed, balanced view of these areas, including an analysis of strengths and weaknesses in each area.

IDC analysis also describes the very real dangers these initiatives create for Japanese science if Japan does not fund a successor to the K supercomputer, as well as the benefits for science of the architectural approach RIKEN has been taking – especially the suitability of the RIKEN approach for grand challenge science and support for Nobel laureate-class work. IDC believes that the RIKEN approach has some important advantages over the approaches being taken in the U.S., China and Europe.

- IDC examines the uses and limitations of other approaches, such as typical cluster architectures with GPUs or other accelerators.
- Key related workforce issues were also examined as IDC studies have confirmed that there is a worldwide shortage of qualified candidates for important HPC job categories, especially algorithm developers, parallel programmers, system administrators, and candidates with expertise spanning computational science and a scientific domain.

In addition to the exploration of technical attributes, IDC surveyed a number of leadership-class supercomputer users to gain insights on the importance of those systems to their overall research efforts as well as to gather theory insights on how the lack of such a system could affect their ability to conduct leading-edge research going forward.

- How important is access to this type of computer to your research/science?
- How important is it to your work/research to have a first class world leading supercomputer?
- How important is it to your NATION to have a first class world leading supercomputer?
- What would happen to your work if you had to only use a scale out vanilla cluster or a cloud?

Ease of Use Metrics

In order to accurately assess the overall complexity required to either upgrade or implement a completely new leadership class supercomputer, IDC analysists used the following methodology and associated ease of use metrics to help define, quantify, and ultimately compare the overall ease of use across the range of systems examined.

It is important to note that many of the decisions for assigning ease of use ratings were heavily influenced by the presence or absence of key architectural features that comprise each of the systems described in this report.

- Systems that were considered to be have more ease of use issues included those with heterogeneous architectures, mixed CPU/GPU configurations, or new or untested processors. Additionally, new or innovative memory or storage schemes, such as burst buffers or related SSD technology, could negatively affect the overall ease of use for a particular system.
- In contrast, systems that were essentially upgrades to existing systems, especially those purchased from commercial vendors, and based largely on COTS hardware and open source software, were considered to have fewer ease of use concerns, and they generally received higher ease of use ratings.

Scale and Definitions

IDC assessed the overall ease of use of each leadership-class supercomputer for porting and/or running of new or existing codes on a new computer. That assessment was divided into three distinct elements:

- Initial writing/porting of new codes on new computer

- Performance tuning of new and existing codes for optimized results.
- Ease-of-use for existing running codes (porting/certification).
- Ease-of-use for existing running codes (optimization, redesign, rewriting)

Specific factors for consideration within each category were:

Initial writing/porting of new codes on new computer

- Composition of overall system architecture
- Composition of processor base(s), node size, node count etc.
- Inclusion of accelerators
- Characteristics of cache and/or memory: size, composition (DDR4, NVRAM, etc.), access times, memory bandwidth, memory configuration per chip, node, or system
- Complexity of memory hierarchy, interconnect scheme, or file system,
- OS and related software stack composition and complexity
- Linkages to legacy code
- Availability/sophistication of code development/optimization tools
- Availability of skilled programmers either in-house or for hire

Ease-of-use for existing running codes (porting/certification)

- Changes in overall system architecture from existing system
- Changes in processor base(s), node size, node count etc.
- Inclusion or change in accelerator use
- Changes in memory size, speed, configuration per chip, node, or system from predecessor system
- Impact of new, existing, or different memory hierarchy, interconnect scheme, or file system
- Changes in OS and related software stack composition and complexity
- Issues with legacy code composition including age, type, language, size and complexity
- Availability/sophistication of code porting tools
- Existing code base composition: open source, in-house, or ISV
- Availability of skilled programmers for legacy and/or new code base
- Complexity of certification and/or verification and validation process
- Data integrity issues
- Data storage formats

Ease-of-use for existing running codes (optimization, redesign, rewriting).

- Complexity, composition, or uniqueness of system architecture, processors composition, memory architecture, interconnect scheme and file system
- Availability and sophistication of scheduling, resource management, job queuing management and related run-time operation software
- Availability/sophistication of code development/optimization tools
- Availability/functionality of hardware/software monitors and operational data collection tools

Ease of Use Rating Scheme

For each of these measures, IDC analysts used the following numerical rating scale to assess the effort needed to implement a various ease of use features for each leadership-class supercomputers.

- The scale goes from 1 to five, with five being defined as a simple process with only some ease of use requirements, and 1 being a highly complex process with many difficult ease of use requirements.
- In addition, IDC analysts used a score of -1 for situations where the process may consist of considerable, if not insurmountable, ease of use requirements.

The complete scale is as follows:

5 = MOVING TO THE NEW SYSTEM IS A SIMPLE PROCESS WITH LITTLE OR NO NEW EASE OF USE REQUIREMENTS

4 = MOVING TO THE NEW SYSTEM IS A SIMPLE PROCESS WITH SOME EASE OF USE REQUIREMENTS IN THE PROJECT

3 = MOVING TO THE NEW SYSTEM WILL REQUIRE AN AVERAGE LEVEL OF NEW EASE OF USE CAPABILITIES TO MAKE THE SYSTEM WORK FOR A BROAD SET OF USERS

2 = MOVING TO THE NEW SYSTEM IS A COMPLEX PROCESS WITH MANY EASE OF USE REQUIREMENTS TO MAKE THE SYSTEM WORK FOR A BROAD SET OF USERS

1 = MOVING TO THE NEW SYSTEM IS A VERY COMPLEX PROCESS WITH MANY DIFFICULT EASE OF USE REQUIREMENTS TO MAKE THE SYSTEM WORK FOR A BROAD SET OF USERS

-1 = MOVING TO THE NEW SYSTEM IS A HIGHLY COMPLEX PROCESS WITH CONSIDERABLE, PERHAPS INSURMOUNTABLE EASE OF USE REQUIREMENTS

It should be noted that not all of the considerations listed above applied to every system examined, and instead the rating given to each system was based on a determination of the key ease of use features that would have the greatest impact on the overall ease of use of any given system. examined.

Finally, it should be noted that this scale inherently assumes that programming any new system is difficult. As such even for those machines that are highly rated –say with a four or five – programming is still considered to be a complex task, but from a comparative perspective, it will likely be easier to program than those with a lower ease of use score, say one or two.

Definitions

The Leadership Computers Facilities Researched in the Study

The following is a short summary of the various exascale systems and their associated competing facility examined in this study. This section is not meant to be a rigorous examination that compares and contrasts the various facilities that will house these HPCs, but instead it is a quick compendium that highlights significant features or value-added distinctions unique to each site.

The Sierra HPC at the US Department's Lawrence Livermore National Laboratory (LLNL)

Located in Livermore, California, U.S., LLNL's defining responsibility is ensuring the safety, security and reliability of the nation's nuclear deterrent. Yet LLNL's mission is broader than stockpile stewardship, as dangers ranging from nuclear proliferation and terrorism to energy shortages and

climate change threaten national security and global stability. The Laboratory's science and engineering are being applied to achieve breakthroughs for counterterrorism and nonproliferation, defense and intelligence, energy and environmental security. The lab employ 2,700 scientists and engineers (more than 40% of whom are Ph.Ds.). The computing and simulation infrastructure at LLNL spans multiple buildings with large computer rooms. The largest facility was designed specifically to house the landmark Purple and Blue Gene/L systems and their successors, which currently include Sequoia, Vulcan, and Zin, among others. This computer facility provides 48,000 ft^2 and 30 MW of power for systems and peripherals, and additional power for the associated machine-cooling system.

The Summit HPC at the US Department of Energy's Oak Ridge National Laboratory (ORNL)

Located in eastern Tennessee, near Knoxville in the U.S., ORNL is the largest Department of Energy science and energy laboratory, conducting basic and applied research to deliver transformative solutions to compelling problems in energy and security. ORNL's diverse capabilities span a broad range of scientific and engineering disciplines, enabling the Laboratory to explore fundamental science challenges and to carry out the research needed to accelerate the delivery of solutions to the marketplace. The lab has a staff of more than 4,600, including scientists and engineers in more than 100 disciplines. ORNL's supercomputing program has grown from humble beginnings to deliver some of the most powerful systems in the world. On the way, it has helped researchers deliver practical breakthroughs and new scientific knowledge in climate, materials, nuclear science, and a wide range of other disciplines.

The Aurora HPC at the US Department of Energy's Argonne National Laboratory (ANL)

Located outside Chicago, Illinois, U.S., ANL is a multidisciplinary science and engineering research center, where scientists and engineers work together to answer some of the biggest questions facing humanity, from how to obtain affordable clean energy to protecting ourselves and our environment. The laboratory works in concert with universities, industry, and other national laboratories on questions and experiments too large for any one institution to do by itself. Through collaborations domestically and around the world, ANL strives to discover new ways to develop energy innovations through science, create novel materials molecule-by-molecule, and gain a deeper understanding of the earth, is climate, and the cosmos. ARL is staffed by over 3,200 employees, almost half of them scientists and engineers. and there are over 7,000 facility users. The Argonne Leadership Computing Facility's (ALCF) mission is to accelerate major scientific discoveries and engineering breakthroughs for humanity by designing and providing world-leading computing facilities in partnership with the computational science community.

The CORI HPC at the U.S. Department of Energy's National Energy Research Scientific Computing Center

As one of the largest facilities in the world devoted to providing computational resources and expertise for basic scientific research, NERSC is a world leader in accelerating scientific discovery through computation. NERSC is a division of the Lawrence Berkeley National Laboratory, located in Berkeley, California, U.S. More than 5,000 scientists use NERSC to perform basic scientific research across a wide range of disciplines, including climate modeling, research into new materials, simulations of the early universe, analysis of data from high energy physics experiments, investigations of protein structure, and a host of other scientific endeavors. NERSC provides some of the largest computing and storage systems available anywhere, but what distinguishes the center is its success in creating an environment that makes these resources effective for scientific research. NERSC systems are

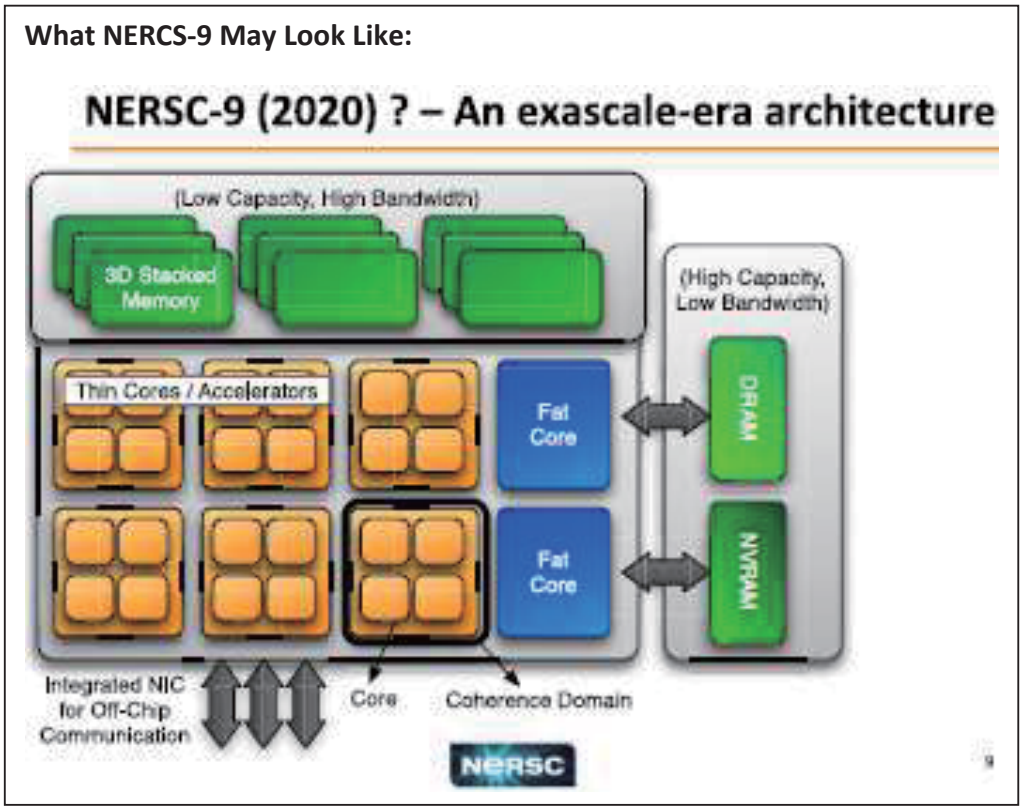
reliable and secure, and provide a state-of-the-art scientific development environment with the tools needed by the diverse community of NERSC users.

The NERSC-9 HPC at U.S. Department of Energy's Lawrence Berkeley National Laboratory

Located in Berkeley, California, U.S., in the world of science, Lawrence Berkeley National Laboratory is synonymous with "excellence." Thirteen Nobel prizes are associated with Berkeley Lab. Seventy Lab scientists are members of the National Academy of Sciences (NAS), one of the highest honors for a scientist in the United States. Berkeley Lab fosters groundbreaking fundamental science that enables transformational solutions for energy and environment challenges, using interdisciplinary teams and by creating advanced new tools for scientific discovery, and the lab employs approximately 3,232 scientists, engineers and support staff. The LBNL computing center is operated by the NERSC center outlined above.

The two diagrams bellow are from Nicholas J. Wright; NERSC-9 Chief Architect, NUG meeting, March 24, LBNL:

What NERCS-9 May Look Like:



A Site Example:
NERSC's Ten-Year Plans

Timeline:

- 2016 NERSC-8, CORI operational
- 2018 25MW site power upgraded
- 2020 NERSC-9 operational -- 150 to 300 Petaflops
- 2021 35MW site power upgraded
- 2024 NERSC-10 -- Exascale capable system is operational
- 2028 NERSC-11 -- 5 to 10 Exaflops is operational

Project Goals:

- Provide a significant increase in computational capabilities over the Edison system, at least 16x on a set of representative DOE benchmarks
- Platform needs to meet the needs of extreme computing and data users by accelerating workflow performance
- Platform should provide a vehicle for the demonstration and development of exascale-era technologies
- NERSC will partner with vendors on Non-Recurring Engineering (NRE) projects to maximize the usability and performance of the machine

- Additional NERSC-9 Requirements:
  - Non-volatile (page addressable memory) = 10's of PBs, 10's TB/sec
  - Spinning disk = 50 PBs, 1 TB/sec
  - Tape = 100's PBs, 10's GB/sec
  - Burst buffer > 90 PBs, >5 TB/sec
  - To develop these requirements, they held many face-to-face meetings with all vendors and all of their major user groups - a requirements workshop with each location/group -- http://www.nersc.gov/science/hpc-requirements-reviews/exascale/

Application Mix Used for Setting the NERSC-9 Requirements:

MiniApp	Description	Language
miniDFT (Quantum Espresso)	Plain-wave Density Functional Theory (DFT)	Fortran
MILC	Lattice Quantum Chromodynamics (QCD). Sparse matrix inversion, CG	C
GTC-P	Particle-in-cell magnetic fusion	C
UMT	Unstructured-Mesh deterministic radiation Transport	C/C++/Fortran
SNAP	Neutron particle transport application	Fortran
PENNANT	Unstructured finite element	C
Meraculous	De novo genome assembly	UPC
MiniPIC	Particle in cell for accelerators	C++
HPCG	High Performance Conjugate Gradient	C

The Cheyenne HPC at the U.S. National Science Foundation's National Center for Atmospheric Research

Located in Boulder Colorado, U.S., NCAR is a federally funded research and development center devoted to service, research and education in the atmospheric and related sciences. NCAR's mission is to understand the behavior of the atmosphere and related Earth and geospace systems and specifically; to support, enhance, and extend the capabilities of the university community and the broader scientific community, nationally and internationally; to foster the transfer of knowledge and technology for the betterment of life on Earth. The National Science Foundation is NCAR's sponsor, with significant additional support provided by other U.S. government agencies, other national governments and the private sector. The Cheyenne HPC will be installed at the NCAR-Wyoming Supercomputing Center, a 150,000 square foot facility that provides advanced computing services to scientists across the nation who study weather, oceanography, air pollution, climate, space weather, energy production, seismology, carbon sequestration, computational science, and other topics related to the Earth system.

The TaihuLight HPC built by Sunway and installed at China's National Supercomputer Center in Jiangsu, China.

The system, designed by the Beijing-funded National Research Center of Parallel Computer Engineering and Technology (NRCPC), is located at the National Supercomputing Center in the city of Wuxi, in Jiangsu province, China. The Chinese Super Computing Center (Wuxi) was established after a joint investment in 2006 by the Ministry of Science and Technology and the Wuxi Government and is one of the most advanced high-performance computing platforms in the China. Wuxi, Tsinghua University and Jiangsu Industrial Technology Research Institute also signed a cooperative agreement entrusting the National Supercomputing Wuxi Center to Tsinghua University to manage and operate the facility, which is aimed at actively carrying out software development, focusing on propelling Chinese supercomputing application capability.

Note: IDC has also included information on the Sunway version that is in competition for the China 2020 exascale program. China will be holding a contest in 2018 to decide which designs will receive funding to be built at the exascale level. Both the Sunway and NUDT designs have a good chance in this competition. The design point is 1 EF peak, .6 EF on the Linpack test, and under 35 MW power consumption.

The TianHe2 A (and 3) at China's National University of Defense Technology

The National University of Defense Technology (NUDT) is a top military academy, as well as a research national key university located in Changsha, Hunan Province, China. It is under the direct leadership of China's Central Military Commission, and dual management of the Ministry of National Defense of the People's Republic of China and Ministry of Education. It is a designated center for Beijing-funded programs, Project 211 and Project 985, the two national plans for facilitating the development of Chinese higher education. NUDT now has over 2,000 faculty members of which nearly 300 are professors. There are 15,700 full-time students including 8,900 undergraduates and 6,800 graduates. NUDT has a long history of HPC development, developing one of China's first HPCs, the Yinhe-1 in 1983, which had a theoretical peak performance of one gigaflop/s.

Note: IDC has also included information on the possible NUDT system version that is in competition for the China 2020 exascale program. China will be holding a contest in 2018 to decide which designs will receive funding to be built at the exascale level. Both the Sunway and NUDT designs have a good

chance in this competition. The design point is 1 EF peak, .6 EF on the Linpack test, and under 35 MW power consumption.

The Hazel Hen at the High Performance Computing Center Stuttgart, Germany (HLRS)

HLRS was established in 1996 as the first national German High Performance Computing center. It is a research and service institution affiliated with the University of Stuttgart, offering services to academic users and industry. HLRS focuses on the operation of leading edge HPC systems, teaching and training for HPC programming and simulation, research in the field of HPC together with national and international partners, collaboration with industry in R&D, and providing access to HPC systems through hww - a public private partnership with commercial partners T-Systems and Porsche. HLRS has been open to European users since 2010 through the EU-wide PRACE program.

The SuperMUC at Leibniz Supercomputer Centre (LRZ) of the Bavarian Academy of Science and Humanities,

LRZ provides services to the scientific and academic communities that include general IT services for more than 100,000 university customers in Munich and for the Bavarian Academy of Sciences and Humanities (BAdW), a communications infrastructure called the Munich Scientific Network (Münchner Wissenschaftsnetz, MWN), archiving and backup of large amounts of data on extensive disk and automated magnetic tape storage, and a technical and scientific high performance Supercomputing Centre for all German universities. The LRZ offers computing power on several different levels including the national supercomputing system SuperMUC, two more supercomputers (a SGI UV with 2,080 cores, and a MEGWARE Cluster with 1,424 cores) which can be accessed by researchers from all Bavarian universities, and an Intel IA32/EM64T-based Linux cluster which can be accessed by researchers from the Munich universities. These machines serve as platforms for running a diverse spectrum of applications, as well as developing and testing serial and parallel programs.

The Piz Daint at the Swiss National Supercomputer Center (CSCS) in Lugano Switzerland

Founded in 1991, CSCS, the Swiss National Supercomputing Centre, develops and provides the key supercomputing capabilities required to solve important problems to science and/or society. The centre enables world-class research with a scientific user lab that is available to domestic and international researchers through a transparent, peer-reviewed allocation process. CSCS's resources are open to academia and are available as well to users from industry and the business sector. The centre operates dedicated computing facilities for specific research projects and national mandates, e.g. weather forecasting. It is the national competence centre for high-performance computing and serves as a technology platform for Swiss research in computational science.

The D-Wave System at the NASA Ames Research Center on Moffett Field, California, US

NASA Ames Research Center, one of ten NASA field centers, is located in the heart of California's Silicon Valley. For more than 76 years, Ames has led NASA in conducting world-class research and development in aeronautics, exploration technology and science aligned with the center's core capabilities. Ames' key goals include maintaining expertise in information technology, aerospace and aeronautics research and engineering, conduct research in space, Earth, lunar and biological sciences, developing lead status for NASA in small spacecraft missions, expanding public and private partnerships, and contributing innovative, high performance and reliable exploration technologies. NASA’s Quantum Artificial Intelligence Laboratory (QuAIL), which operates the D-Wave system, is the space agency's hub for experiments to assess the potential of quantum computers to perform calculations that are difficult or impossible using conventional supercomputers.

UK Three System Upgrade at the University of Edinburgh, ECMWF, Daresbury Lab.

The Science and Technology Facilities Council (STFC), the UK government sponsored research organization that works with both national industrial and research partners on systems and software projects and that funds the Daresbury Lab, is underwriting a $412 million investment to support a grand OpenPower undertaking that will bring three new systems into play at three of the UK's leading HPC development centers.

HPC at the French Alternative Energies and Atomic Energy Commission(CEA)

The French Alternative Energies and Atomic Energy Commission (CEA) is a key player in research, development and innovation in four main areas: defense and security, nuclear energy (fission and fusion), technological research for industry, fundamental research in the physical sciences and life sciences. The CEA is established in ten centers spread throughout France, and it works in partnership with many other research bodies, local authorities and universities. Within this context, the CEA is a stakeholder in a series of national alliances set up to coordinate French research in energy (ANCRE), life sciences and health (AVIESAN), digital science and technology (ALLISTENE), environmental sciences (AllEnvi) and human and social sciences (ATHENA). CEA has 15 958 technicians, engineers, researchers and staff, spread across 51 joint research units, has created 187 start-ups since 1972 in the innovative technologies sector, and currently has 438 ongoing European research projects. The scientific computing complex at the CEA facility in Bruyères-le-Châtel (DAM/Île de France) hosts one of Europe's largest high-performance computing facilities, used for defense, industrial and research applications. It houses the large infrastructure operated by CEA DAM specifically for defense-related programmers, featuring the Atos/Bull Tera1000-1 supercomputer (with a processing power of 2.5 petaflops), which will be joined in 2017 by the 25-petaflop Tera1000-2.

Table 1

The Supercomputers Evaluated in This Study

Computer Names	Project Names	Prime Developer/ Industry Partner	Organization	Country	Planned Delivery Year
Sierra	CORAL	IBM, NVIDIA, Mellanox	LLNL	USA	2017, 3Q
Summit	CORAL	IBM	ORNL	USA	2017, 3Q
Aurora	CORAL	Intel/Cray	ANL	USA	2018, 4Q
CORI	NERSC-8	Cray	LLBL	USA	2016, 4Q
Crossroads	APEX 2020	*TBD*	LANL	USA	2020, 4Q
NERSC-9	APEX 2020	*TBD*	LLBL	USA	2020, 4Q
Cheyenne	U/U/C*	SGI	NCAR	USA	2017, 3Q
TaihuLight	Sunway	NRCPC	Sunway	China	2016 (2020, 4Q)

Table 1

The Supercomputers Evaluated in This Study

Computer Names	Project Names	Prime Developer/ Industry Partner	Organization	Country	Planned Delivery Year
(And the Sunway 2020 System)					
TianHe2 A (And the NUDT 2020 System)	Milky Way 3	NUDT/Inspur	NUDT	China	2017 2Q or 3Q (2020, 4Q)
Hazel Hen	U/U/C	Cray	HLRS	Germany	2015
SuperMUC	Phase 2	Lenovo/IBM	LRZ	Germany	2015
Piz Diant	U/U/C	Cray	CSCS	Switzerland	2016
D-Wave	U/U/C	U/U/C	Google/NASA	USA	2015
UK Three System Upgrade	S&T Facilities Council	IBM, others	University of Edinburgh, ECMWF, Daresbury Lab	UK	2018, 3Q
CEA/Bull	Tera 100 Follow-on	Bull	CEA/ Bull	France	2020, 3Q

Source: IDC 2016

* U/U/C = Uncertain/Unknown/Confidential

The Parameters Tracked for Each Leadership System

For the purpose of this study, and in considered consultation with RIKEN, IDC used the following technical parameters in tracking the major leadership-class supercomputers around the world.

Systems Attributes: consisted of overall planned system attributes that included:

- Planned performance–the theoretical perk performance of the system, typically measured in petaflops,
- Architecture and node design–the overall architectural configuration of the system such as node composition, memory and interconnect scheme, and file system.
- Power--the maximum power use by the system, typically measured in megawatts
- MTBF- the mean time between failures envisioned for the system that include hardware and software related failures.

- KPI–key performance indicators that measure how well the system performs compared with some independent, but representative performance parameter, such as a multiple of performance improvement over an existing or companion system.

Prices; consisted of overall system prices – including overall procurement costs, breakdowns on hardware and software maintenance costs, additional NRE considerations, as well as listings of the particular funding organization.

Ease of Use–defined as the ability for programmers to use the full computational capabilities of the system, which in the study consisted of four different ease of use categories.

- Planned New Features
- Porting/Running of New Codes on a New Computer
- Missing Items that Reduce Ease-of-Use
- Overall Ability to Run Leadership Class Problems

Hardware Attributes – consisted of the collection of key hardware specifications comprising the leadership-class supercomputer. Key metrics included:

- Processors
- Memory Systems
- Interconnects1
- Storage
- Cooling
- Special Hardware
- Estimated Utilization

Software Attributes – consisted of the collection of key software specifications/features comprising the leadership-class supercomputer. Key metrics included:

- OS and Special Software
- File Systems
- Compilers and Middleware
- Other Software

STUDY RESULTS: PRIMARY COMPARISONS

Comparisons of System Attributes

System Attributes: Planned Performance

As seen in Table 2, there are a number of major leadership-class HPCs in the planning and development stage targeted for deployment between 2016 and 2020. In addition to Japan, such development is being considered across a wide range of major HPC suppliers and user regions including China, the EU, and the United States.

- Most of these systems can be considered pre-exascale designs: systems that will underwrite much of the technology critical to the development of the hardware and software necessary to support the spate of exascale systems planned for the 2020 to 2022-time frame.
- As such, the bulk of the system planned for the next four years target a peak performance capability generally between 10 and 300 teraflops, with the bulk of the lower-end systems closer to completion this year or the next, while higher performance systems are targeted for completion closer to 2020.

Notes about the power goals in table 2 below:

1. Some of the GF/W rating for these systems are a measure of the theoretical peak performance the system in PFLOPS divided by their total power consumption in MWs.
   - This is the case for systems like the Sierra, Summit, Aurora, SuperMUC, Hazel Hen and Piz Daint. In these cases, either the site specifically listed that number as their GF/W metric or IDC did the calculation using stated/estimated peak performance and power numbers.
2. Other sites used different metrics:
   - The Cori system has a calculated peak GF/W of 8.1, but the site specifications cited 14-16 GF/W, which is the performance rating of the Knight Landing processor they will use in the system.
   - The Crossroads and NERSC-9 systems specifically called out 35 GF/W for peak performance, and 10-12 GF/W for Linpack performance.
   - The Cheyenne HPC at has a 3.06 GF/W based on system peak performance and total system power use. But they also indicated that they were targeting 34 MF/W based on sustained performance on NCAR workloads.
   - The TaihuLight metric of 6 GF/W was based on Linpack performance, as was the case with the TianHe2 A system.
3. The CEA/Bull system was simply listed as overall system goal.

Table 2

System Attributes: Planned Performance

Computer Names	Planned Delivery Date	Planned Performance PF	GF/Watt Goals
Sierra	2017, 3Q	120-150	11.5 to 13.2

Table 2

System Attributes: Planned Performance

Computer Names	Planned Delivery Date	Planned Performance PF	GF/Watt Goals
Summit	2017, 3Q	150-300	11.5 to 13.2
Aurora	2018, 4Q	>180	13.0
CORI	2016, 4Q	>30	14 to 16
Crossroads	2020, 4Q	150-300	35+ (peak goal) 10-12 likely on Linpack
NERSC-9	2020, 4Q	150-300	35+ (peak goal) 10-12 likely on Linpack
Cheyenne	2017, 3Q	5.34	3.0
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	125 peak/93 LP (1,000)	6 (Around 30)
TianHe2 A (NUDT 2020)	2017 2Q or 3Q (2020, 4Q)	Could be as high as 200-300 or at least 100+LP (1,000)	4-5 (20-30)
Hazel Hen	2015	7.42	2.3
SuperMUC	2015	3.58	3.2
Piz Daint	2016	15	6.45
D-Wave	2015	NA as it doesn't do FLOPS	NA as it doesn't do FLOPS
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C
CEA/Bull	2020, 3Q	1,000	35 to 50

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

System Attributes: Architecture and Node Design

As seen in table 3, there are a wide range of different architectural design paths to a pre-exascale system. During this study, two major design options emerged.

- Some are looking to partner with a commercial vendor, such as Cray or IBM, to help them develop a leadership-class system that is in keeping with the overall product offerings of their commercial partner, such as the Cheyenne SGI system at NCAR.
- Others, such as NUDT's Tianhe-2 A group in China, are essentially looking to custom-build systems that likely will be produced in very limited quantiles, primarily for in-house use with some but perhaps even no commercial counterpart.

In addition, it is clear that there is no agreed upon architectural scheme for these pre-exascale systems. Variants noted here include designs with various CPUs, including x86, POWER and ARM, others with hybrid CPU/GPU configurations, and a wide variety of new memory, node and interconnect schemes.

Finally, recent information on two additional 2020 exascale Chinese machines offer the following details:

- For a Sunway TaihuLight follow-on slated for 2022, the system will have a 20 teraflops/node capability with a total of 50,000 nodes, described as having a bootable accelerator architecture, yielding a peak performance of one exaflop.
- For an NUDT Tianhe-2A follow-on slated for 2020-2022, the system will have an 80 teraflops/node capability with a total of 12, 500 nodes described as a CPU + accelerator architecture, yielding a peak performance of one exaflop as well.

Table 3

System Attributes: Architecture and Node Design

Computer Names	Planned Delivery Date	System Architecture	Node Configuration	Memory/Node
Sierra	2017, 3Q	Custom	>1500 fat nodes, 40 TFLOPS/node	512 GB DDR4 +HBM, 800 GB of NVRAM
Summit	2017, 3Q	Custom	>3400 fat nodes, 40 TFLOPS/node	512 GB DDR4 +HBM, 800 GB of NVRAM
Aurora	2018, 4Q	Cray Shasta	50,000 nodes	Not yet decided
CORI	2016, 4Q	Cray XC40	2 16-core Intel processors per node, KNL nodes 68 cores/node	128GB (Phase I)
Crossroads	2020, 4Q	Being designed	Being designed	Being designed
NERSC-9	2020, 4Q	Being designed	Being designed	Being designed
Cheyenne	2017, 3Q	SGI ICE XA cluster	4032 Nodes (dual socket nodes)	DDR4 64 GB/node on 3,168 nodes, 128 GB/node on 864 nodes

Table 3

System Attributes: Architecture and Node Design

Computer Names	Planned Delivery Date	System Architecture	Node Configuration	Memory/Node
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	Custom (Custom)	40,960 nodes, each –one SW chip/node (Likely a large array of accelerator like cores)	32 GB
TianHe2 A (NUDT 2020)	2017 2Q or 3Q (2020, 4Q)	Custom (Custom)	FT-2000 ARM chip based (Likely a CPU & accelerator mix)	U/U/C
Hazel Hen	2015	Cray XC40	7712 nodes, 2 sockets/node	128 GB
SuperMUC	2015	Lenovo NeXtScale nx360 M5 WCT	3072 nodes, 2 processors per node	64 GB
Piz Diant	2016	Cray XC30	5272 nodes 1 cpu + gpu per node,	32 GB (DDR3) 6GB (GDDR5)
D-Wave	2015	2X Quantum Computer	1000 qubits	NA
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C	U/U/C
CEA/Bull	2020, 3Q	Custom	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

System Attributes: Power

As seen in Table 4, concerns about power consumption and related efficiencies keeping total power needs for pre-exascale system generally below 30 MW, for even some of the most powerful machines. Such a trend may be a driving force towards the use of GPU and other accelerators, which can boost peak performance with a more forgiving energy requirement than a CPU-only configuration.

- In addition, many designers are looking for ways to better control the real-time power draw of their systems by using advanced techniques to power down portions of their systems not in use during a job or even a portion of a job.
- As such, many of the numbers here can be considered a maximum power requirement that may not be realized for any significant portion of the time the system is being used.

Although it is not clear from this table, IDC analysts generally assesses that most of the newer data centers in operation or those planned for the next few years will be able to offer an adequate power supply for these systems.

- Only older sites, or those looking to expand their operation well into the next decade, will require any significant new data center upgrades due to limitations in power or cooling capabilities.

Table 4

System Attributes: Power

Computer Names	Planned Delivery Date	Planned Computer System Power (MW)	Data Center Peak Power (MW)
Sierra	2017, 3Q	10	30
Summit	2017, 3Q	13.3	25
Aurora	2018, 4Q	13	40
CORI	2016, 4Q	<3.7	12.5+
Crossroads	2020, 4Q	18	25MW to 35 MW
NERSC-9	2020, 4Q	18	25MW to 35 MW
Cheyenne	2017, 3Q	1.75	8
TaihuLight	2016 (2020, 4Q)	15.37	20-25
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	17.8+	20-25
Hazel Hen	2015	3.2	4
SuperMUC	2015	1.1	20
Piz Diant	2016	2.325	20
D-Wave	2015	25kw	Very low, under a few MWs
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C
CEA/Bull	2020, 3Q	<20	50

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

System Attributes: MTBF Rates

As seen in table 5, few sites currently stress specific or hard requirements for MTBF rates. However, IDC analysts note that most pre-exascale designers are generally – and increasingly - concerned with

the large amount of advanced hardware and complex software that running these systems entails, and they are aggressively looking for ways to ensure that failures of these components, either hardware or software, do not significantly impact the ability of these systems to consistently execute some of their most demanding jobs.

- This issue is becoming increasingly more vexing as the overhead for preforming mid-computation checkpoints are increasingly an invasive, expensive, and performance limiting procedure.
- Hence, it is safe to conclude that designers of all of these system are acutely aware of reliability concerns for their system and do indeed have some metrics in mind for system reliability, but that in many cases that data is simply not publicly available.

In addition, for the purposes of this study, the Planned Job Mean Time to Interrupt was considered the interval that a particular software job could run before experiencing an unexpected interrupt on operation that was not due to either a hardware failure or overall system failure. This metric is typically used to determine how frequently a checkpoint procedure needs to take place, Planned Hardware MTBF was the expected time that any particular element of hardware would fail that could affect the overall system operations and where such a failure could not could not be managed by software. Finally, planned system MTBF was defined as the case where the overall system failure was due to either a hardware issue that could not be managed by software or a software driven failure event.

Table 5

System Attributes: MTBF Rates

Computer Names	Planned Delivery Date	Planned Job Mean Time to Interrupt (JMTTI)	Planned Hardware MTBF	Planned system MTBF
Sierra	2017, 3Q	U/U/C	U/U/C	>6 days (144 hours)
Summit	2017, 3Q	U/U/C	U/U/C	>6days (144 hours)
Aurora	2018, 4Q	U/U/C	U/U/C	>6days (144 hours)
CORI	2016, 4Q	U/U/C	U/U/C	U/U/C
Crossroads	2020, 4Q	24 hrs.	U/U/C	>720 hours
NERSC-9	2020, 4Q	24 hrs.	U/U/C	>720 hours
Cheyenne	2017, 3Q	U/U/C	>384 hours	>576 hours
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	U/U/C	U/U/C	U/U/C
TianHe2 A (NUDT 2020)	2017 2Q or 3Q (2020, 4Q)	U/U/C	U/U/C	U/U/C
Hazel Hen	2015	U/U/C	U/U/C	U/U/C

Table 5

System Attributes: MTBF Rates

Computer Names	Planned Delivery Date	Planned Job Mean Time to Interrupt (JMTTI)	Planned Hardware MTBF	Planned system MTBF
SuperMUC	2015	U/U/C	U/U/C	U/U/C
Piz Diant	2016	U/U/C	U/U/C	U/U/C
D-Wave	2015	A day	A couple days	A week
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C	U/U/C
CEA/Bull	2020, 3Q	U/U/C	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

System Attributes: KPIs

As seen in table 6, it is clear that designers, developers, and users of pre-exascale systems are not generally concerned with the theoretical peak computation performance of their new systems. Instead IDC analysts note that there is an increased emphasis on determining the ability of a new system to deliver a sustained performance - one that captures a system' s complete compute, memory, interconnect, and storage infrastructure to execute an end-to end task - over one that stressed pure computationally capability. As such, there were essentially two basic key performance indicators mentioned most often;

- Sustained performance on key benchmarks that are deemed representative of their typical and planned workload,
- A basic speed up multiplier (such as 25X or 50X) by the new system when compared to the system it was replacing when running an existing set of user applications.

Table 6

System Attributes: KPIs

Computer Names	Planned Delivery Date	Any key KPI's being used?
Sierra	2017, 3Q	Scalable science benchmarks: CPU+GPU system 13x higher, throughput benchmarks: 12X performance (most of their existing applications will still run on just the CPUs)

Table 6

System Attributes: KPIs

Computer Names	Planned Delivery Date	Any key KPI's being used?
Summit	2017, 3Q	Scalable science benchmarks: CPU+GPU system 13x higher, throughput benchmarks: 12X performance (most of their existing applications will still run on just the CPUs)
Aurora	2018, 4Q	More than eighteen times the computational performance of Mira, its predecessor at the ALCF, using a nearly equal number of compute nodes. (most of their existing applications will still run on just the CPUs)
CORI	2016, 4Q	Sustained application performance on NERSC SSP codes: Phase I Haswell: 83 TFlop/s (vs. 293 TFlop/s for Edison and 144 TFlop/s for Hopper). (most applications will still run on just the CPUs)
Crossroads	2020, 4Q	Run benchmarks: SNAP, PENNANT, HPCG, MiniPIC, UMT, MILC, MiniDFT, GTC, and Meraculous,, Must run Trinity and Sierra code, 20X Edison. (most of their existing applications will still run on just the CPUs)
NERSC-9	2020, 4Q	Run benchmarks: SNAP, PENNANT, HPCG, MiniPIC, UMT, MILC, MiniDFT, GTC, and Meraculous, 20X Edison. (most of their existing applications will still run on just the CPUs)
Cheyenne	2017, 3Q	3.5 times Yellowstone peak performance. (most of their existing applications will still run on just the CPUs)
TaihuLight	2016 (2020, 4Q)	GF/Watts, the desire is to improve it by 6x to 8x by 2022. (most of their existing programs will use the accelerators; others applications will run on a different system)
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	Create a working, competitive new processor type. (U/U/C)
Hazel Hen	2015	Run benchmarks on actual user applications, especially in energy research (CO2 emissions) and materials science (3D development of micro-structures). (most applications will still run on just the CPUs)
SuperMUC	2015	Usage Monitored by Job, User, Country and Scientific Department. (most of their existing applications will still run on just the CPUs)
Piz Daint	2016	Solution times and energy efficiency on their actual user applications in cosmology, materials science, seismology and climatology. (most applications will still run on just the CPUs)
D-Wave	2015	The desire is to find more applications, hopefully more than ten, where the system performs well. Plans are to always use it in conjunction with a regular HPC system. (All applications will use a different programing model)

Table 6

System Attributes: KPIs

Computer Names	Planned Delivery Date	Any key KPI's being used?
UK Three System Upgrade	2018, 3Q	U/U/C. (most of their existing applications will still run on just the CPUs)
CEA/Bull	2020, 3Q	Sustained performance gains over prior supercomputer on a wide range of scientific applications. (most of their existing applications will still run on just the CPUs)

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Comparisons of Prices

Comparison of System Prices

As seen in table 7, pre-exascale designers are operating under a wide range of budgets from a low of $25 million to well over ten times that amount.

- Some of the most expensive pre-exascale systems – such as the most technologically, aggressive custom systems - are projected to cost $250 million or more. These systems represent some of the most advanced HPC developments in the world that include significant cost for NRE. Typically, some portion of hardware and software developed for these system may not have wide applicability to the HPC sector writ large and may not appear in the commercial HPC sector anytime soon, if ever.
- Others, primarily those that are one step behind the leading-edge of performance, generally are looking at budgets an order of magnitude less. Many of these systems have less aggressive NRE requirements and instead rely primarily on hardware and software technology supplied by their vendor partners developed with an eye towards widespread commercial applicability.

Table 7

Comparison of System Prices

Computer Names	Planned Delivery Date	Planned Price w/o Maintenance	Estimated Yearly Maintenance Costs	Hardware Maintenance	Software Maintenance	Other Maintenance Costs	Additional NRE (R&D) Costs
Sierra	2017, 3Q	$325M Sierra + Summit	Est. $45 to $50 million	Est. $30 to $40 million	Est. $5 to $10 million	Est. under $5 million	Could exceed $50 million, likely around $15 million

Table 7

Comparison of System Prices

Computer Names	Planned Delivery Date	Planned Price w/o Maintenance	Estimated Yearly Maintenance Costs	Hardware Maintenance	Software Maintenance	Other Maintenance Costs	Additional NRE (R&D) Costs
Summit	2017, 3Q	$325M Sierra + Summit	Est. $45 to $50 million	Est. $30 to $40 million	Est. $5 to $10 million	Est. under $5 million	Could exceed $50 million, likely around $15 million
Aurora	2018, 4Q	$200M	Est. $20 to $30 million	Est. $15 to $20 million	Est. $5 to $6 million	Est. under $4 million	Could exceed $10 million, likely much lower
CORI	2016, 4Q	$70M	Est. around $10 million	Est. around $8 million	Est. around $2 million	Small	Small
Crossroads	2020, 4Q	Estimated around $250 million	Est. $20 to $30 million	Est. $15 to $20 million	Est. around $5 million	Small	Could exceed $50 million, likely around $15 million
NERSC-9	2020, 4Q	Estimated around $250 million	Est. $20 to $30 million	Est. $15 to $20 million	Est. around $5 million	Small	Could exceed $50 million, likely around $15 million
Cheyenne	2017, 3Q	Est. $25-35M	Est. around $8 million	Est. around $7 million	Est. around $1 million	Small	Small
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	$300 million ($350 million plus)	Est. around $15 million	Est. around $12 million	Est. around $3 million	Small	Small (included in the price)
TianHe2 A (NUDT 2020)	2017 2Q or 3Q (2020, 4Q)	Estimated around $300 million ($350 million plus)	Est. around $15 million U/U/C	Est. around $12 million U/U/C	Est. around $3 million U/U/C	Small Small	Small (included in the price) Small
Hazel Hen	2015	Approximately $100 million	Est. around $10 to $15 million	Est. around $8 to $11 million	Est. around $2 to $4 million	Small	Small (if any)

Table 7

Comparison of System Prices

Computer Names	Planned Delivery Date	Planned Price w/o Maintenance	Estimated Yearly Maintenance Costs	Hardware Maintenance	Software Maintenance	Other Maintenance Costs	Additional NRE (R&D) Costs
SuperMUC	2015	Approximately $110 million	Est. around $10 to $12 million	Est. around $8 to $9 million	Est. around $2 to $3 million	Small	Included in the price
Piz Diant	2016	$41m upgrade, total system over $90 million	Est. around $9 to $12 million	Est. around $7 to $8 million	Est. around $2 to $3 million	Small	Included in the price
D-Wave	2015	Est. $25M	Est. around $2 to $4 million	Est. around $2 to $4 million	Small	Small	Small
UK Three System Upgrade	2018, 3Q	$400M for three systems	Est. $15 to $25 million	Est. $10 to $20 million	Est. around $5 million, could be lower	Small	Small
CEA/Bull	2020, 3Q	Estimated around $250 million	Est. $20 to $30 million	Est. $15 to $20 million	Est. around $5 million	Small	Likely included in the system price

Source: IDC 2016

Comparison of System Prices: Who's Paying for It?

As seen in table 8, funding the development of pre-exascale systems is essentially a government-only phenomenon, be it through national or, in some limited cases, regional government entities.

- In most case, pre-exascale HPC development is being driven primarily by national security and/or basic science applications, but IDC assess that there remains a strong, growing interest in funding new HPC development to serve a growing assortment of economic requirements as well.
- IDC analysts assess that highest end of the HPC sector – defined as where NRE is a significant percentage of the overall machine cost - is simply not attracting any significant commercial user interest.
- However, IDC assess that there likely is substantial interest in high-end systems for the commercial sector that may be a step behind the leading-edge.

Table 8

Comparison of System Prices: Who's Paying for It?

Computer Names	Planned Delivery Date	Who's Paying for It?
Sierra	2017, 3Q	DOE
Summit	2017, 3Q	DOE
Aurora	2018, 4Q	DOE
CORI	2016, 4Q	DOE
Crossroads	2020, 4Q	DOE
NERSC-9	2020, 4Q	DOE
Cheyenne	2017, 3Q	NCAR
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	Central Chinese government, province of Jiangsu, city of Wuxi.
TianHe2 A (NUDT 2020)	2017 2Q or 3Q (2020, 4Q)	NUDT +Chinese Government And likely a city
Hazel Hen	2015	HLRS
SuperMUC	2015	LRZ
Piz Diant	2016	SCSC
D-Wave	2015	Google, NASA and Likely a US Government Agency
UK Three System Upgrade	2018, 3Q	3 UK sites plus UK Govt. support
CEA/Bull	2020, 3Q	CEA

Source: IDC 2016

Comparisons of Ease-Of-Use

Ease-of-Use: Planned New Features

As seen on table 9, IDC analysts assess that most pre-exascale development program are well aware to the ease of use complexities associated with these new systems and planners are committing significant resources to develop the necessary software tools for writing new code, porting existing code, and running their overall application workloads.

IDC analysts gave higher scores to those facilities that:

- Have long experience in introducing new leadership-class supercomputers into their research community with a high degree of success.
- Participate in collaborations or partnerships that bring the collective wisdom of a larger design/user base to bear on critical ease of use issues.
- Can draw on complementary software development programs specially designed for researching and expanding HPC-based ease of use capabilities.
- Take advantage of open source software – at all levels of the software stack – as a way to access easily a global base of HPC-relevant software.

Those with lower scores generally had less experience in HPC usage, stressed new or novel architectures with a limited base of existing ease of use software tools, or that generally were lagging behind other world-class facilities from both an operational and application skill set.

It is important to note that for U.S. Department of Energy laboratories, each of the major computing centers has an existing pool of highly skilled and experienced programmers used to porting both new and existing applications to a new HPC.

- As such, DOE budgets for HPC acquisitions typically do not contain any specific line items for NRE for those particular tasks. Instead the cost and related resources for any such porting activities are considered part of the overall operating budget for the lab distinct from any single procurement.

Finally, for some categories, such as ease of use for investments or the ability to run leadership applications, IDC analysts used the notion that a high score – such as a four or five – meant that the facility under consideration was actively taking steps through either funding or some other resource commitment to specially address the issue.

- For example, a rating of five in the investment category meant that IDC sees clear evidence that a particular program is making the right investments to ensure that ease of use features are properly considered within the overall scope of the project.
- Likewise, a lower score indicated that for the project under consideration, there was no clear sign that specific investments were in place to address critical ease of use issues.

The Ease-of-use scale is as follows:

- 5 = MOVING TO THE NEW SYSTEM IS A SIMPLE PROCESS WITH LITTLE OR NO NEW EASE OF USE REQUIREMENTS
- 4 = MOVING TO THE NEW SYSTEM IS A SIMPLE PROCESS WITH SOME EASE OF USE REQUIREMENTS IN THE PROJECT
- 3 = MOVING TO THE NEW SYSTEM WILL REQUIRE AN AVERAGE LEVEL OF NEW EASE OF USE CAPABILITIES TO MAKE THE SYSTEM WORK FOR A BROAD SET OF USERS
- 2 = MOVING TO THE NEW SYSTEM IS A COMPLEX PROCESS WITH MANY EASE OF USE REQUIREMENTS TO MAKE THE SYSTEM WORK FOR A BROAD SET OF USERS
- 1 = MOVING TO THE NEW SYSTEM IS A VERY COMPLEX PROCESS WITH MANY DIFFICULT EASE OF USE REQUIREMENTS TO MAKE THE SYSTEM WORK FOR A BROAD SET OF USERS
- -1 = MOVING TO THE NEW SYSTEM IS A HIGHLY COMPLEX PROCESS WITH CONSIDERABLE, PERHAPS INSURMOUNTABLE EASE OF USE REQUIREMENTS

Table 9

Ease-of-Use: Planned New Features

Computer Names	Planned Delivery Date	New Ease-of-use Features Planned	Investments underway or planned (or new technologies) to improve ease-of-use
Sierra	2017, 3Q	5	5
Summit	2017, 3Q	5	5
Aurora	2018, 4Q	5	5
CORI	2016, 4Q	5	5
Crossroads	2020, 4Q	5	5
NERSC-9	2020, 4Q	5	5
Cheyenne	2017, 3Q	4	4
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	2 (2)	2 (2)
TianHe2 A (NUDT 2020)	2017 2Q or 3Q (2020, 4Q)	3 (2)	3 (2)
Hazel Hen	2015	5	5
SuperMUC	2015	5	5
Piz Diant	2016	5	5
D-Wave	2015	1	1
UK Three System Upgrade	2018, 3Q	5	5
CEA/Bull	2020, 3Q	4	4

Source: IDC 2016

Ease-of-Use: Porting/Running of New Codes on a New Computer

As can be seen in table10, there is a wide range of ease of use issues that must be considered when looking a leadership-class supercomputer development.

- Factors deemed critical when determining the initial writing/porting of codes include considerations of how drastic the new architecture is from a processor, memory or interconnect perspective, the complexity of the code being developed, how well the applications fit with the new architecture, the experience and knowledge level of the programmers that will work on the code, and the range of supporting software development tools available to help facilitate the overall porting process.
- Likewise, many of the same factors listed above exist when considering the ease of use for running existing codes, but with the added requirements for overall system efficiency and utilization, ease of job scheduling requirements, the ability of the code to scale to the available system resources at run time, and the need for checkpoint or periodic backups. In addition, the requirements for code certification adds another element of difficulty when codes are large, complex, or require a major re-write or even algorithm change due to architectural or other major system changes mandated by the introduction of a new computing platform.
- Finally, determination for the ease of use of running existing codes shares many of the potential complexities with those outlined in writing new codes, but also includes concerns that there may not be available a full effective suite of development tools to fully support the optimization of existing (and perhaps old legacy) codes. In addition, issues about personnel availability for older codes must be considered. Indeed, here it may be the availability of experienced staff familiar with the inner working of an existing code base that will be the key determinant of how effectively that code can be made to run on a new system.

Anatomy of an Ease of Use Determination: The LRZ SuperMUC:

In order to illustrate the process behind determining the ease of use metrics for these systems, the process that yielded an overall rating of three for the LRZ SuperMUC will be discussed here.

Examination of technical documentation indicates that the new system is essentially a compilation of a number of distinct so-called compute islands – each island consists of its own distinct and unique cluster of nodes joined by a common interconnect – and the installation of each island was spread across two phases that span over four years.

Phase one (the first three islands installed in three stages over a three-year period) consists of 18 thin node islands based on Intel Sandy Bridge-EP processor technology, 6 thin node islands based on Intel Haswell-EP processor technology and one fat node island based on Intel Westmere-EX processor technology.

- Each island contains more than 8,192 cores.

Phase 2, installed in 2015, uses yet another Intel processor base and, perhaps more important, involves a new vendor when Lenovo replaced IBM as the main supplier.

- In addition, the interconnect in each island is a different version of the InfiniBand.

IDC analysts expect that harnessing the full capability of such a diverse system could present challenges to any programmers that seek to harness the full computational capability of this system. More likely, the system likely be used with jobs targeted for one specific island.

- This judgment, however, was mitigated to an extent by the long and successful history of LRZ's technical experts in successfully utilizing such architectures to meet their overall application requirements.

As a result, and on balance, IDC analysts felt that this system would be a complex one to program, much more so than a number of other less diverse machines studied here, but LRZ staff would generally be capable of managing the challenges.

Table 10

Ease-of-Use: Porting/Running of New Codes on a New Computer

Computer Names	Planned Delivery Date	Initial writing/porting of new codes on new computers	Ease-of-use for existing running codes (porting/certification)	Ease-of-use for existing running codes (optimization, redesign, rewriting	Comment
Sierra	2017, 3Q	4	4	4	New POWER Processor base + New GPU
Summit	2017, 3Q	4	4	4	New POWER Processor base + New GPU 2

Table 10

Ease-of-Use: Porting/Running of New Codes on a New Computer

Computer Names	Planned Delivery Date	Initial writing/porting of new codes on new computers	Ease-of-use for existing running codes (porting/certification)	Ease-of-use for existing running codes (optimization, redesign, rewriting	Comment
Aurora	2018, 4Q	4	5	5	Upgrade from existing Cray architecture
CORI	2016, 4Q	5	4	4	Split partitions of Haswell and KNL
Crossroads	2020, 4Q	4	4	4	20X performance requirement over existing system, 30x more memory
NERSC-9	2020, 4Q	4	4	4	20X performance requirement over existing system, 30x more memory
Cheyenne	2017, 3Q	5	5	5	Intel-based SGI upgrade
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	2 (2)	2 (2)	2 (2)	New Sunway processor base, interconnect, no legacy code, high core counts
TianHe2 A NUDT (2020)	2017 2Q or 3Q (2020, 4Q)	2 (2 or 3)	2 (2 or 3)	2 (2 or 3)	New ARM processor base, new interconnect, minimal legacy code high core counts
Hazel Hen	2015	5	5	4	Cray XC40 upgrade
SuperMUC	2015	3	3	3	Lenovo replaced IBM as primary upgrade/supplier
Piz Diant	2016	4	4	4	Hybrid CPU-GPU and CPU-only nodes
D-Wave	2015	-1	-1	1	Very limited apps base for annealing QC
UK Three System Upgrade	2018, 3Q	3	3	3	Broad mission/user base, diverse legacy codes, open power/ARM based processor base
CEA/Bull	2020, 3Q	4	4	4	

Table 10

Ease-of-Use: Porting/Running of New Codes on a New Computer

Computer Names	Planned Delivery Date	Initial writing/porting of new codes on new computers	Ease-of-use for existing running codes (porting/certification)	Ease-of-use for existing running codes (optimization, redesign, rewriting	Comment

Source: IDC 2016

Ease-of-Use: Missing Items that Reduce Ease-of-Use

As seen in table 11, IDC analysts assess that for the most part, the leadership-class supercomputers under development have accounted well for the fundamental ease of use features that they will need for their best possible operation.

- This is not surprising as the increasingly higher costs of such systems almost automatically justify significant pre-planning and designer/user collaboration, not just within the specific project, but more commonly across projects, some that transcend bureaucratic as well as national borders. Simply put, as the cost of developing a less than superior system has become quite high, the performance, and perhaps even the budgetary, penalty for building an ill-designed system grows accordingly.
- Indeed, for most successful projects, planners work hard to manage expectations, making clear the ultimate risk associated with a particular effort that can range from high risk, pathfinder development systems to those that will be counted on as production systems tasked with managing a predictable and stable workload.

Table 11

Ease-of-Use: Missing Items that Reduce Ease-of-Use

Computer Names	Planned Delivery Date	Any missing items that reduce ease-of-use, e.g. key compliers
Sierra	2017, 3Q	None
Summit	2017, 3Q	None
Aurora	2018, 4Q	None
CORI	2016, 4Q	None
Crossroads	2020, 4Q	None
NERSC-9	2020, 4Q	None
Cheyenne	2017, 3Q	None

Table 11

Ease-of-Use: Missing Items that Reduce Ease-of-Use

Computer Names	Planned Delivery Date	Any missing items that reduce ease-of-use, e.g. key compliers
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	New type of custom processor, so it's likely that many HPC applications won't fit well on it. And there are other custom parts of the system that may require major application redesign.
TianHe2 A (NUDT 2020)	2017 2Q or 3Q (2020, 4Q)	The new processor type may require new ease-of-the capabilities to use in a performant way. Given the very low usage of the previous system, it must require a fair amount of work to port codes and to optimize codes.
Hazel Hen	2015	None
SuperMUC	2015	None
Piz Diant	2016	None
D-Wave	2015	The computer can only handle a very small set of problem types. Most HPC applications and jobs will never run on this system.
UK Three System Upgrade	2018, 3Q	None
CEA/Bull	2020, 3Q	None Likely

Source: IDC 2016

Ease-of-Use: Overall Ability to Run Leadership Class Problems

As seen in table 12, many of the considerations discussed earlier are concerned with what it takes to design and operate a new leadership-class supercomputer and can also have a significant impact on the ability of that system to effectively execute its most critical applications.

- Based on these key considerations IDC assesses that issues that materially complicate ease of use include new or unproven processor bases, memory system, or storage infrastructure, as well as the use of a novel architecture.
- Factors that contributed to a better ease of use rating include the experience and skill set of the designer/user base, the degree to which the system was either targeted specifically as a research or production systems, and the sophistication and experience of the overall development organization as well as its history with success high-end computing design or use.

Table 12

Ease-of-Use: Overall Ability to Run Leadership Class Problems

Computer Names	Planned Delivery Date	Ability to run difficult leadership class applications (with high communication requirements)	Comments
Sierra	2017, 3Q	4	Drastic new POWER architecture, fat node design, tough software certification process, complex legacy code, strong expertise and experience base, synergy from Summit effort
Summit	2017, 3Q	4	Drastic new POWER architecture, fat node, tough software certification process, complex legacy code, strong expertise and experience base, synergy from Summit effort
Aurora	2018, 4Q	5	Smooth commercially based upgrade path, tough software verification process, strong expertise and experience base
CORI	2016, 4Q	4	Cray upgrade, CPUs and CPU/GPU partitions, wide user base, legacy code
Crossroads	2020, 4Q	4	No architectural details yet, NRE devoted to ease application development, significant legacy code requirements, large memory size synergy with NERSC-9 efforts
NERSC-9	2020, 4Q	4	No architectural details, yet significant legacy code requirements, synergy with Crossroads effort
Cheyenne	2017, 3Q	5	Essentially an SGI upgrade
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	2 (2)	Unproven architecture, new Sunway processor and interconnect base, minimal code development skills, no legacy code/performance requirement
TianHe2 A (NUDT 2020)	2017 2Q or 3Q (2020, 4Q)	2 (2 or 3)	Unproven architecture, new ARM processor and interconnect base, minimal code development skills, no legacy code/performance requirement
Hazel Hen	2015	4	Smooth upgrade path, long HPC expire, sophisticated user base. But wide application span
SuperMUC	2015	4	IBM/Lenovo experience base, long development history, sophisticated code base,
Piz Diant	2016	4	Cray upgrade, but new CPU and CPU/GPU processor partitions, string software development cadre
D-Wave	2015	-1	Severely limited application base, viewed a test bed system

Table 12

Ease-of-Use: Overall Ability to Run Leadership Class Problems

Computer Names	Planned Delivery Date	Ability to run difficult leadership class applications (with high communication requirements)	Comments
Large Public Cloud	2018, 3Q	1 for most, 5 for vertically focused clouds	Clouds work well for only a limited set of HPC applications. Future clouds may offer a friendlier environment for HPC, like access to the base hardware without virtualization, but most won't. This requires a lot of special application redesign to obtain any reasonable level of performance on most leadership class HPC codes.
UK Three System Upgrade	2020, 3Q	3	Open POWER, ARM based systems, diverse user base
CEA/Bull	2020	5	Uses Aggressive Bull interconnect, and likely a new processor type

Source: IDC 2016

Comparisons of Hardware Attributes

Hardware Attributes: Processors

A seen in table 13, there is a wide range of pre-exascale processors and related GPU accelerators being considered for inclusion in the leadership-class supercomputers studies.

- From a CPU perspective, commercially available processors that will be used in at least one pre-exascale system include Intel Xeon processors, Intel Xeon Phi variants, IBM-based POWER processors, and different versions of the ARM processor. Some HPC designers are also looking at custom-designed chips, like the Shenwei processor from Sunway used in the TaihuLight system.
- A number of pre-exascale designers are also calling for the inclusion of GPU accelerators, primarily from NVIDIA, with a few custom accelerators under consideration, primarily from Chinese developers.
-

Table 13

Hardware Attributes: Processors

Computer Names	Planned Delivery Date	Processor Type	Processor Attributes (speed, # of cores, etc.)	GPU /Accelerator	GPU/Acc. Attributes
Sierra	2017, 3Q	IBM POWER 9	>1500 nodes	NVIDIA Volta/Tesla	NVLink, NVMe-comaptible PCIe 800GB SSD
Summit	2017, 3Q	IBM POWER 9	>4600 nodes	NVIDIA Volta/Tesla	NVLink, NVMe-comaptible PCIe 800GB SSD
Aurora	2018, 4Q	Intel Knights Hill/Shasta architecture	>50,000 nodes	NA	NA
CORI	2016, 4Q	Phase I Haswell + Phase II KNL partitions	1630 Haswell nodes (32 c) /9304 KNL nodes (68c)	NA	NA
Crossroads	2020, 4Q	Intel Haswell thin nodes	>500GF/CPU	Likely a future version of the Intel Xeon Phi family	Integrated CPU-accelerator
NERSC-9	2020, 4Q	Intel Haswell thin nodes	>500GF/CPU	Likely a future version of the Intel Xeon Phi family	Integrated CPU-accelerator
Cheyenne	2017, 3Q	Intel 'Broadwell' Xeon E5-2697V4	2.3-GHz/18C	NA	NA
TaihuLight (Sunway 2020)	2016 (2020, 4Q)	SW26010 processor Custom processor	1.45Ghz/260C, 3TFLOPS/chip 50,000 nodes, 20TF each	NA	NA
TianHe2 A (NUDT 2020)	2017 2Q or 3Q (2020, 4Q)	32,000 Xeons, 32,000 ShenWei processor Likely custom processors	Intel Xeon E5-2692v2 2.2GHz/12C 12,500 nodes, 80TF each	Upgrade with 96,000 Phytium accelerator card	64 Xiaomi ARM Cores 28NMm 2GHz
Hazel Hen	2015	Xeon E5-2680v3	2.5GHz/12C	NA	NA

Table 13

Hardware Attributes: Processors

Computer Names	Planned Delivery Date	Processor Type	Processor Attributes (speed, # of cores, etc.)	GPU /Accelerator	GPU/Acc. Attributes
SuperMUC	2015	Haswell Xeon Processor E5-2697 v3	2.6 GHz/14C	NA	NA
Piz Daint	2016	Haswell	Intel Xeon ES-2670	Tesla P100	16 GB HBW Memory, 21.2 TFLOPS Peak 16nm FinFET
D-Wave	2015	D-Wave quantum processor	1000 processors, operates at 15mK	NA	NA
UK Three System Upgrade	2018, 3Q	OpenPOWER, ARM-centric	Power9 ~4GHz (2017)	Nvidia Volta GPU, FPGA	Stacked DRAM (NVIDIA)
CEA/Bull	2020, 3Q	Intel Xeon EP		NVIDIA or Intel Xeon Phi (CEA will decide)	*TBD*

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Hardware Attributes: Memory Systems

As seen in table 14, the overarching trend in pre-exascale design is towards more memory, more SSD, and the use of additional memory accelerators, such as burst buffers or high bandwidth memory packages as a way to deal with the increasing need for larger, higher bandwidth and lower latency memory systems.

- Despite the universal need for more memory according to many pre-exascale designers, specifications for memory per node for the set of systems examined has a wide variation: ranging from 32 GB/node for the TaihuLight from Sunway to a high of 512 GB DDR4 plus 800 GB of NVRAM per node for one of the DOE CORAL systems.
- Although there are a handful of planned system with total memory sizes in excess of three petabytes, most are looking at memory sizes on the order of at least one petabyte.
- The use of SSD at the node level is becoming a more persuasive design feature found in pre-exascale systems.

Table 14

Hardware Attributes: Memory Systems

Computer Names	Planned Delivery Date	Memory Technology	Additional Memory Attributes
Sierra	2017, 3Q	512 GB DDR4 +HBM	NVMe-comaptible PCIe 800GB SSD
Summit	2017, 3Q	2.4 PB DDR4 + HBM + 3.7 PB persistent memory	NVMe-comaptible PCIe 800GB SSD
Aurora	2018, 4Q	> aggregate 7PB DRAM and Intel SSDs	On-package memory bandwidth >30PB/s
CORI	2016, 4Q	Haswell partition has 203 TB, KNL partition 1PB aggregate	Haswell: Each node has 128 GB DDR4 2133Mhz MHz memory, KNL: Each node has 96 GB DDR4 2133Mhz MHz memory, six 16GB DIMMs
Crossroads	2020, 4Q	>3PB	U/U/C
NERSC-9	2020, 4Q	>3PB	Non-volatile 10's PBs, l0'sTB/sec
Cheyenne	2017, 3Q	313 TB	64 GB/node on 3,168 nodes, 128 GB/node on 864 nodes,
TaihuLight	2016 (2020, 4Q)	DDR3 1.31 PB	32 GB of DDR3/node
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	1.0 PB+	256 GB per Mars processor
Hazel Hen	2015	0.9PB	DDR4
SuperMUC	2015	194 TB	DDR4
Piz Daint	2016	169TB DDR3, 32TB GDDR4 CPU/GPU	Cray's DataWarp technology
D-Wave	2015	NA	NA
UK Three System Upgrade	2018, 3Q	U/U/C	flash storage
CEA/Bull	2020, 3Q	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Hardware Attributes: Interconnects

As seen in table 15, for most of the systems that will be delivered soon, designers are opting for either InfiniBand, OmniPath, or a custom in-house interconnect scheme (such as BX from Bull Atos for the CEA supercomputer). IDC analysts assess, however, that over the next few years, leadership-class supercomputers designers will increasingly be looking for more capable, responsive, and more intelligent interconnect schemes as more and more, overall system performance will be determined by the capabilities of the interconnect.

- These efforts will be complicated by HPC design trends that include the use of more cores– and many core accelerators - smaller memory per core, deeper memory hierarchies, and unique memory structures such as burst buffers, high bandwidth memory schemes, and SSDs.
- In addition, new requirements at the architectural scale or even program scale, will increasingly call for parallel interconnect schemes to manage one to many, one to one, and other data communication patterns as well as distinctive perhaps even application-specific connection patterns.

Table 15

Hardware Attributes: Interconnects

Computer Names	Planned Delivery Date	Interconnect Type	Interconnect Attributes	Network topology	Bi-section bandwidth – or other metric	Node level injection bandwidth (rate?)
Sierra	2017, 3Q	InfiniBand	200 Gb/sec HDR InfiniBand	U/U/C	U/U/C	U/U/C
Summit	2017, 3Q	InfiniBand	200 Gb/sec HDR InfiniBand, 3-Level Fat Tree	non-blocking fat-tree	U/U/C	23 GB/s
Aurora	2018, 4Q	Intel 2 Gen OmniPath with silicon photonics	Aggregate bandwidth 2.5 PB/s at the system level, bisectional bandwidth ~ half PB/s	Fabric	2.5 PB bisection BW	U/U/C
CORI	2016, 4Q	Cray Aries high-speed interconnect	5.625 TB/s global bandwidth (Phase I). 45.0 TB/s global peak bisection bandwidth (Phase II)	Dragonfly topology	0.25 µs to 3.7 µs MPI latency, ~8GB/sec MPI bandwidth	U/U/C
Crossroads	2020, 4Q	U/U/C	High speed interconnect supports a high messaging bandwidth, high injection rate, low latency, high throughput, and independent progress	RFPS still out	U/U/C	U/U/C

NERSC-9	2020, 4Q	U/U/C	High speed interconnect supports a high messaging bandwidth, high injection rate, low latency, high throughput, and independent progress	RFPS still out	U/U/C	U/U/C
Cheyenne	2017, 3Q	Mellanox EDR InfiniBand	Partial 9D Enhanced Hypercube, 224 36-port switches, no director switches	Partial 9D Enhanced Hypercube	U/U/C	U/U/C
TaihuLight	2016 (2020, 4Q)	Three level switching fabric	Nodes are connected using PCI-E 3.0 connections, Mellanox supplied the Host Channel Adapter (HCA) and switch chips	Multi-layer fabric	Bisection network bandwidth is 70TB/s	U/U/C
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	In-house fat tree network	fat tree topology with 13 switches each of 576 ports at the top level	Fat tree topology	U/U/C	U/U/C
Hazel Hen	2015	Cray Aries interconnect	System-on-a-chip device comprising four NICs, a 48-port tiled router and a multiplexer known as Netlink. A single Aries device provides the network connectivity for all four nodes on a Cray XC blade	Aries routing and communications ASIC	U/U/C	U/U/C
SuperMUC	2015	InfiniBand FDR14	Two stage: non-blocking Tree/Pruned Tree 4:1	U/U/C	U/U/C	U/U/C
Piz Diant	2016	Cray Aries interconnect	System-on-a-chip device comprising four NICs, a 48-port tiled router and a multiplexer known as Netlink. A single Aries device provides the network connectivity for all four nodes on a Cray XC blade	Aries routing and communications ASIC	32 TB/s	U/U/C
D-Wave	2015	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C
CEA/Bull	2020, 3Q	Bull BXI	U/U/C	U/U/C	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Hardware Attributes: Storage

As seen in table 16, leadership class supercomputer designs have overall storage class requirements that are moving well into the 100PB range in the next few years. Currently, two major file systems are

used, Lustre and GPFS, but it is unclear if either of these options can be successfully extended for systems in the out years or if they will be replaced with new file systems capable of handling new storage class objects, larger file size per file, and larger total storage sizes.

- IDC analysts assess that leadership-class supercomputer designers are facing a number of new challenges in designing their storage systems: in addition to the growing size and bandwidth requirements for storage systems needed to support typical HPC applications, there is also a growing necessity for increased storage capabilities for high performance scratch disks located near system cores or nodes, and well as large archival and backup storage.
- IDC analysts expect to see SSDs or other nonvolatile variants become an increasingly prevalent feature in many of the newer leadership-class HPC storage systems.

Table 16

Hardware Attributes: Storage

Computer Names	Planned Delivery Date	Storage System Attributes	Storage Read & Write Bandwidth	File System Size
Sierra	2017, 3Q	GPFS File System	1.2/1.0 TB/s R/W	120 PB memory
Summit	2017, 3Q	GPFS File System	1.2/1.0 TB/s R/W	120 PB memory
Aurora	2018, 4Q	Lustre,	>1 TB/s throughput	>150PB Lustre
CORI	2016, 4Q	Flash 'Burst Buffer' to accelerate I/O performance, a layer of NVRAM that sits between memory and disk.	700 gigabytes/second I/O	28 petabytes
Crossroads	2020, 4Q	30X baseline memory	U/U/C	Storage capable of retaining all application input, output, and working data for 12 weeks (84 days),
NERSC-9	2020, 4Q	30X baseline memory > 50 PBs	~1 TB/sec	storage capable of retaining all application input, output, and working data for 12 weeks (84 days),
Cheyenne	2017, 3Q	DDN (SFA) system	200 GB/s aggregate I/O bandwidth	21-PB
TaihuLight	2016 (2020, 4Q)	U/U/C	U/U/C	U/U/C
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	Global Shared parallel storage system, +	U/U/C	12.4 PB

Table 16

Hardware Attributes: Storage

Computer Names	Planned Delivery Date	Storage System Attributes	Storage Read & Write Bandwidth	File System Size
Hazel Hen	2015	Lustre	100GB/s	10PB
SuperMUC	2015	GPFS	250 GB/s	15PB
Piz Diant	2016	Lustre	117 GB/s	2.5 PB
D-Wave	2015	U/U/C	U/U/C	U/U/C
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C	U/U/C
CEA/Bull	2020, 3Q	U/U/C	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Hardware Attributes: Cooling

As seen in table 17, most HPC today use some form of water cooling with an increasing emphasis on a warm water option. IDC analysts expect that going forward more and more HPC designs will factor in new and innovative options for power and efficiency and cooling, perhaps even those that may dictate where the system will be installed.

- For example, the Swiss Piz Daint gets its chilled water directly from a nearby lake, while hot weather climates have much more complex – and energy consuming – equipment requirements that limit their ability to drive down the site's PUE.

Table 17

Hardware Attributes: Cooling

Computer Names	Planned Delivery Date	Cooling System Attributes	Air or Liquid Cooled?	Temperature of Warm Water (if used)
Sierra	2017, 3Q	Warm water cooling	Water	U/U/C
Summit	2017, 3Q	Warm water cooling	Water	U/U/C
Aurora	2018, 4Q	Warm-water cooling	Water	U/U/C
CORI	2016, 4Q	Water Cooled	Water	U/U/C

Table 17

Hardware Attributes: Cooling

Computer Names	Planned Delivery Date	Cooling System Attributes	Air or Liquid Cooled?	Temperature of Warm Water (if used)
Crossroads	2020, 4Q	Water Cooled	Water	U/U/C
NERSC-9	2020, 4Q	Water Cooled	Water	U/U/C
Cheyenne	2017, 3Q	Closed loop airflow / water	Combined	U/U/C
TaihuLight	2016 (2020, 4Q)	Water cooled	Water	U/U/C
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	Closed-coupled chilled water cooling with a customized liquid water-cooling unit.	Chilled water	U/U/C
Hazel Hen	2015	U/U/C	Water cooled with forced transverse air flow: 6,900 cfm	U/U/C
SuperMUC	2015	warm water cooling	Water	16c in/20c out
Piz Diant	2016	cold water cooling	Water	Cold water (43°F, 6°C) extracted from Lake Lugano provides cooling for the datacenter
D-Wave	2015	Extreme cooling, at 3 levels	Liquid	U/U/C
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C	U/U/C
CEA/Bull	2020, 3Q	U/U/C	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Hardware Attributes: Special Hardware

As seen in table 18, most leadership-class HPCs are not targeting the development of any aggressive or special-purpose hardware unique to their particular performance requirements.

- IDC analysts assess that the cost of designing and manufacturing special hardware – not including custom but essentially mass produced devices like ASICs or FPGAs -simply has

become too complex and costly to justify for such a limited use. This is especially germane when compared with the wide range of other device options that include specialized but ultimately more general-purpose hardware such as ARM chips with special HPC functionality, many core GPUs, and even non-volatile memory parts initially targeted for the smartphone sector.

Table 18

Hardware Attributes: Special Hardware

Computer Names	Planned Delivery Date	Special Hardware Components
Sierra	2017, 3Q	Few if any
Summit	2017, 3Q	Few if any
Aurora	2018, 4Q	Few if any
CORI	2016, 4Q	Cori Phase 1 has twice as much memory per node over Edison predecessor to support data-intensive workloads.
Crossroads	2020, 4Q	Few if any
NERSC-9	2020, 4Q	Few if any
Cheyenne	2017, 3Q	Few if any
TaihuLight	2016 (2020, 4Q)	Many new custom components in the system.
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	The new system will likely have custom sub-parts with different processors.
Hazel Hen	2015	Few if any
SuperMUC	2015	Few if any
Piz Diant	2016	Few if any
D-Wave	2015	Plans are to increase the Qbit count greatly, and to add more support processors to handle parts of the jobs.
UK Three System Upgrade	2018, 3Q	Few if any
CEA/Bull	2020, 3Q	Likely new power conditioning and perhaps special power reduction capabilities

Source: IDC 2016

Hardware Attributes: Estimated Utilization

As seen in table 19, IDC assesses that most of the leadership-class supercomputers under development will exhibit a high degree of end user utilization. Put simply, there is likely sufficient demand for these systems to ensure that they will be in heavy use for the bulk of their operational lifetimes. Indeed, IDC research consistently shows that most large HPCs sites -particularly those in government research facilities–could run significantly more jobs if they had the available compute cycles.

- This does not imply, however, that all of these system will be consistently running a small number large of jobs that will use a significant percentage of the overall system resources. Instead, IDC assesses that in many cases these systems will primarily be running a large number of smaller jobs simultaneously alongside a limited number of large jobs. In this case, to achieve effective system utilization, these machines will need effective job schedulers that can be used to run against a set of clearly defined, predictable, and well behaved jobs.
- In addition, this assessment does not take into consideration the ability of any single job to effectively use all of the resources to which it was allotted at run time. Indeed, IDC assesses that, as has been the case of these large systems during their entire history, even some of the most highly tuned, optimized codes may at any given time be able to effectively harness only a small percentage of their allotted system resources due to issues of complexity, lack of parallelization or scalability, and complexities with synchronizing computer with data movement.

Finally, it is import to note that IDC has found that in many cases the issue of system utilization can vary from site to site and region to region.

- For example, discussions with managers at some of the leading Chinese HPC sites indicate that high system utilization is not considered a critical issue within their facilities. Instead, at these sites greater options for system availability in a flexible and timely manner are considered to be a more important metric.
- Likewise, for some of the more experimental or esoteric systems, utilization clearly takes a back seat to computational research such as with the D-Wave system at NASA Ames.

Table 19

Hardware Attributes: Estimated System Utilization (of user jobs)

Computer Names	Planned Delivery Date	Expected System Utilization (5= high, at least 70%, 1 = low, under 10%)
Sierra	2017, 3Q	5
Summit	2017, 3Q	5
Aurora	2018, 4Q	5
CORI	2016, 4Q	5

Table 19

Hardware Attributes: Estimated System Utilization (of user jobs)

Computer Names	Planned Delivery Date	Expected System Utilization (5= high, at least 70%, 1 = low, under 10%)
Crossroads	2020, 4Q	5
NERSC-9	2020, 4Q	5
Cheyenne	2017, 3Q	5
TaihuLight	2016 (2020, 4Q)	2 or 3
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	3 or 4
Hazel Hen	2015	5
SuperMUC	2015	5
Piz Diant	2016	5
D-Wave	2015	1
UK Three System Upgrade	2018, 3Q	5
CEA/Bull	2020, 3Q	5

Source: IDC 2016

Comparisons of Software Attributes

Software Attributes: OS and Special Software

As seen in table 20, Linux, in its many variants, has become the standard operating system for most leadership-class supercomputers and IDC analysts assess that this will be the case for at least the next five years.

For these system, it will become standard practice to start off with a regular enterprise-class Linux version that allows the project to take the fullest advantage of the myriad Linux-based,and often open sourced, software across the entire HPC software stack.

- The operating system will then likely be modified to match the unique architectural requirements of each HPC. Indeed, OS designers will increasingly come to rely on different customized versions of the system's main Linux OS to run on different parts of the system, such as a Linux micro kernel to run exclusively on compute cores.
- OS functionality will become an increasingly critical – and performance determining – factor in HPC operations going forward, due to growing application spans including both simulation and big data jobs, increasing core and node counts, more complex memory and storage hierarchies, and the need for both batch and real-time operation capabilities.

Table 20

Software Attributes: OS and Special Software

Computer Names	Planned Delivery Date	Planned OS	Planned Special Software
Sierra	2017, 3Q	Linux	OpenPOWER Stack
Summit	2017, 3Q	Linux	OpenPOWER Stack
Aurora	2018, 4Q	U/U/C	ALCF's Early Science Program will jump-start a set of large-scale scientific calculations
CORI	2016, 4Q	Lightweight kernel and run-time environment based on the SuSE Linux Enterprise Server (SLES) Linux distribution	U/U/C
Crossroads	2020, 4Q	U/U/C	U/U/C
NERSC-9	2020, 4Q	U/U/C	U/U/C
Cheyenne	2017, 3Q	RedHat Enterprise Linux OS	U/U/C
TaihuLight	2016 (2020, 4Q)	Sunway Raise OS 2.0.5 based on Linux	U/U/C
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	Kylin Linux	U/U/C
Hazel Hen	2015	Cray Linux Environment	U/U/C
SuperMUC	2015	SuSE Linux Enterprise Server	U/U/C
Piz Diant	2016	Cray Linux Environment	U/U/C
D-Wave	2015	D-Wave Proprietary	U/U/C

Table 20

Software Attributes: OS and Special Software

Computer Names	Planned Delivery Date	Planned OS	Planned Special Software
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C
CEA/Bull	2020, 3Q	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Software Attributes: File Systems

As seen in table 21, Lustre and GPFS are and likely will continue to be the major file system software for leadership-class supercomputers for at least the next five years. Going forward, however, IDC expects to see new requirements for file systems that can handle many of the increasingly prevalent open source big data applications associated with the Hadoop/SPARC big data software suite ecosystem.

- IDC analysts note, however that some of the major pain points facing those who design and use these files system will include meeting the requirements for both computation and big data jobs with structured and unstructured data, batch and real-time processing requests, and a host of new applications in areas including the Internet of Things, cognitive computing, and big data predictive analytics.

Table 21

Software Attributes: File Systems

Computer Names	Planned Delivery Date	Planned File System	Planned File I/O middleware.
Sierra	2017, 3Q	GPFS File System	Platform LSF
Summit	2017, 3Q	GPFS File System	U/U/C
Aurora	2018, 4Q	Lustre,	U/U/C
CORI	2016, 4Q	Flash 'Burst Buffer' to accelerate I/O performance, a layer of NVRAM that sits between memory and disk.	Intel, Cray, and GNU programming environments
Crossroads	2020, 4Q	30X baseline memory	Cites specification of minimum number of compute nodes required to read and write the following data sets from/to platform storage: 1 TB data set of 20 GB files in 2

Table 21

Software Attributes: File Systems

Computer Names	Planned Delivery Date	Planned File System	Planned File I/O middleware.
			seconds, 5 TB data set of any chosen file size in 10 seconds.
NERSC-9	2020, 4Q	30X baseline memory	Cites specification of minimum number of compute nodes required to read and write the following data sets from/to platform storage: 1 TB data set of 20 GB files in 2 seconds, 5 TB data set of any chosen file size in 10 seconds. Offeror shall report the file size chosen
Cheyenne	2017, 3Q	DDN (SFA) system	U/U/C
TaihuLight	2016 (2020, 4Q)	Lustre variant	U/U/C
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	Global Shared parallel storage system, + And a Lustre variant	U/U/C
Hazel Hen	2015	Lustre	U/U/C
SuperMUC	2015	GPFS	U/U/C
Piz Diant	2016	Lustre	U/U/C
D-Wave	2015	U/U/C	QSage, translator
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C
CEA/Bull	2020, 3Q	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Software Attributes: Compilers and Middleware

A seen in table 22, at their highest level of abstraction, most of the major leadership-class HPC going forward will be looking to many of the same compliers and related tools that have formed the foundation of HPC applications programming for over a decade, most notably the use of C, C++, and Fortran backed with an MPI programming model.

- These regimes however will face serious challenges with many of the planned architectures for leadership-class supercomputers that will incorporate multiple millions of cores, complex memory and interconnect schemes, and unique data storage access patterns,

In addition, IDC analysts expect that there will be many new requirements for programming languages and development tools that look to support big data applications development.

- That said, IDC assesses that there will be significant opportunity for growth and diversification in the field as HPC systems with different processor bases, such as x86, POWER, and ARM, as well as associated accelerators, increasingly offer specialized hardware capabilities. Enabling users to capture the highest potential performance of these processors will require strong progress in complier and related middleware development, that is many cases may require targeted support at the national program level as there likely will not be much interest in such development in the commercial sector.
- These efforts will be helped, however, to a great extent by increasingly sophisticated efforts to develop integrated HPC-targeted software stacks such as the Intel HPC Orchestrator or the open source based OpenHPC

Table 22

Software Attributes: Compilers and Middleware

Computer Names	Planned Delivery Date	Planned Compliers Supported	Middleware supported
Sierra	2017, 3Q	Open source LLVM compiler, XL compiler, PGI compiler	U/U/C
Summit	2017, 3Q	Open source LLVM compiler, XL compiler, PGI compiler	U/U/C
Aurora	2018, 4Q	Intel + Cray Compilers and libraries	Intel Scalable System Framework
CORI	2016, 4Q	PGI, the Cray compilers, Intel, gcc, and UPC	SLURM
Crossroads	2020, 4Q	C, C++ (including complete C++11/14/17), Fortran 77, Fortran 90, and Fortran 2008	MPI/OpenMP
NERSC-9	2020, 4Q	C, C++ (including complete C++11/14/17), Fortran 77, Fortran 90, and Fortran 2008	MPI/OpenMP
Cheyenne	2017, 3Q	Intel Parallel Studio XE Cluster, PGI CDK (Fortran, C, C++, pgdbg debugger, pgprof)	SGI Management (tools/utilities)

Table 22

Software Attributes: Compilers and Middleware

Computer Names	Planned Delivery Date	Planned Compliers Supported	Middleware supported
TaihuLight	2016 (2020, 4Q)	compiler: C/C++, Fortran, auto vectorization tool, math libraries	U/U/C
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	Likel C/C++, Fortran, many math libraries, etc. Close to the current TH-2.	Intel MKL-11.0.0
Hazel Hen	2015	PGI Compiling Suite	U/U/C
SuperMUC	2015	U/U/C	U/U/C
Piz Diant	2016	U/U/C	U/U/C
D-Wave	2015	C, C++, Python or MATLAB	U/U/C
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C
CEA/Bull	2020, 3Q	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Software Attributes: Other Software

As seen in table 23, there is modest attention being paid to the development of non-traditional HPC software in the leadership-class supercomputer projects studied. However, IDC analysts expect that such software will become increasingly important in the next few years, particularly big data infrastructures built around the Hadoop/Spark (or other alternative) ecosystem and virtualization schemes such as Docker.

Although not expressly stated in any of the plans studied here, IDC believes that even leadership-class HPC centers will increasingly turn to hybrid competing environments that consist of a mix of on-prem and cloud-based computing environments.

- As such, IDC analysts expect that increasingly HPC applications will be designed to co-exist effectively in both environments, or at least be able to move between the two with a minimum of software modification or penalty on performance.
- This effort will be helped to a great extent by many of the software efforts currently underway to attract high end users by cloud computing suppliers such as OpenStack, and IDC expects to see soon a wide range of cloud-based HPC options across the price performance spectrum with innovative hardware offerings from cloud vendors around the world.

Table 23

Software Attributes: Other Software

Computer Names	Planned Delivery Date	Big Data software, e.g. Hadoop, Spark, etc.	Docker	Openstack	Which sites will use the OpenHPC stack?	Other Software?
Sierra	2017, 3Q	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C
Summit	2017, 3Q	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C
Aurora	2018, 4Q	U/U/C	Docker containers on Cray's variant of Linux	OpenStack cloud controller	U/U/C	U/U/C
CORI	2016, 4Q	U/U/C	Shifter, open-source software stack that enables users to run custom environments on HPC systems, compatible with Docker container format.	U/U/C	U/U/C	U/U/C
Crossroads	2020, 4Q	U/U/C	containerized software images without requiring privileged access	U/U/C	U/U/C	U/U/C
NERSC-9	2020, 4Q	U/U/C	containerized software images without requiring privileged access	U/U/C	U/U/C	U/U/C
Cheyenne	2017, 3Q	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C
TaihuLight	2016 (2020, 4Q)	U/U/C	U/U/C	U/U/C	Sunway OpenACC	U/U/C
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C
Hazel Hen	2015	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C
SuperMUC	2015	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C
Piz Diant	2016	U/U/C	Docker.	U/U/C	U/U/C	U/U/C

Table 23

Software Attributes: Other Software

Computer Names	Planned Delivery Date	Big Data software, e.g. Hadoop, Spark, etc.	Docker	Openstack	Which sites will use the OpenHPC stack?	Other Software?
D-Wave	2015	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C	U/U/C	Deep Learning/ Cognitive	U/U/C
CEA/Bull	2020, 3Q	U/U/C	U/U/C	U/U/C	U/U/C	U/U/C

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

Comparisons of Supporting Research & Development

R&D Plans

As seen in table 24, analysis of the major R&D plans for most of the leadership-class supercomputers studied offer a few key insights worth noting.

- Some, such as those in the U.S. DoE programs, generally are seeking to not only meet their near-term computational requirements, but they are also committing significant NRE to help lay the hardware and software foundation for exascale systems that are scheduled for completion in the 2020-2022 time.
- Others are targeted more towards near-term computing requirements that do not include any significant commitment of NRE funding, such as the Swiss Piz Daint which instead is looking to parent with a commercial vendor to meet its less aggressive, albeit no less important, computational requirements with more traditional HPC architectures.
- Finally, there are some, highlighted by the some of the systems being developed with China, that are focused primarily as research systems, being built in limited quantities-more for their value as HPC research machines than as user-driven research or production systems. These systems are typified by the inclusion of new processors, unproven network or memory technology, and new architectures that challenge the traditional norms of HPC design.

Table 24

R&D Plans

Computer Names	Planned Delivery Date	Special Breakthroughs Being Funded	Special/New Hardware	Special/New Software
Sierra	2017, 3Q	A new processor design	Next-generation IBM OpenPOWER platform, NVIDIA Tesla accelerator platform	Heterogeneous computing model (GPU+CPU)
Summit	2017, 3Q	A new processor design	Next-generation IBM OpenPOWER platform, NVIDIA Tesla accelerator platform	Heterogeneous computing model (GPU+CPU)
Aurora	2018, 4Q	Separate R&D contract.		Likely from Pathforward & ECP
CORI	2016, 4Q	R&D efforts with Cray to drive data potential	Enable higher bandwidth transfers in and out of the compute node	Enabling Docker-like virtualization functionality on Cray compute nodes to allow custom software stack deployment
Crossroads	2020, 4Q	NRE to support increasing application performance, reducing the needs for data motion, enhancing system resilience and reliability	Likely from Pathforward & ECP, plus some power & cooling focused hardware. Some special redundancy reduction hardware.	Likely from Pathforward & ECP, plus new compilers & File systems to support the system size.
NERSC-9	2020, 4Q	NRE to support increasing application performance, reducing the needs for data motion, enhancing system resilience and reliability	Likely from Pathforward & ECP, plus some power & cooling focused hardware. Some special redundancy reduction hardware.	Likely from Pathforward & ECP, plus new compilers & File systems to support the system size.
Cheyenne	2017, 3Q	U/U/C	Likely from Pathforward & ECP.	Likely from Pathforward & ECP.
TaihuLight	2016 (2020, 4Q)	New processor, new interconnect, new system architecture	New processor, new interconnect, new system architecture	New OS, maybe a new file system, lots of system software
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	New processor, new interconnect, new system architecture	New processor, new interconnect, new system architecture	New OS, maybe a new file system, lots of system software
Hazel Hen	2015	U/U/C	U/U/C	U/U/C
SuperMUC	2015	Special cooling approach	Special cooling approach	Software to reduce power and cooling costs

Table 24

R&D Plans

Computer Names	Planned Delivery Date	Special Breakthroughs Being Funded	Special/New Hardware	Special/New Software
Piz Diant	2016	None	None	None
D-Wave	2015	Quantum computing overall	Expand the number and type of Qbits	Creating ways to better use the system. Developing a compiler.
UK Three System Upgrade	2018, 3Q	U/U/C	U/U/C	Applying "Watson" to a broader set of problems
CEA/Bull	2020, 3Q	A full exascale system with new software and new hardware.	A new interconnect. Likely a very high capacity memory system. Likely new liquid cooling technologies. New ways to reduce data movement.	A new exascale software stack, special software to reduce the power consumption. New ways to support big data and reduce data movement.

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

R&D Plans: Partnerships

As seen in table 25, partnerships are becoming a fundamental reality of leadership-class HPC development. IDC analysts note that there were a wide range of partnership types across projects examined.

- Some, like the CORAL effort within the US DOE, look to use the combined HPC hardware and software expertise of partnership members to examine a number of alternative HPC technologies while still meeting the particular mission requirements of each individual organization.
- Other partnerships look instead to a provide rationalized development and research program across a number of sites to ensure that a range of design options are explored in a systematized fashion. Such as partnership is typified by the coordinated development program than spans three separate Germany HPC-based research facilities that distributes procurements across the three labs in two year increments, round robin style.

Likewise, IDC analyses note that almost all major leadership-class projects do involve some committed partnership with one of more commercial vendors, be it at the component, system, or software level. IDC analysts assess that such partnerships can yield substantial benefits for both parties.

- The procuring lab is able to help design and purchase technology in cooperation with some of the leading technical suppliers in the world that might not yet, if ever, be available on the commercial market,
- At the same time, vendors benefit from a first-hand partnership with some most forward-leaning thinkers in HPC design, helping them develop better technology that can then be used in their wider product lines bound for the commercial sector.

Table 25

R&D Plans: Partnerships

Computer Names	Planned Delivery Date	Partnerships for the System	Partnerships for The R&D
Sierra	2017, 3Q	IBM, NVIDIA, Mellanox	IBM, NVIDIA, Mellanox & Pathforward teams & ECP
Summit	2017, 3Q	IBM, NVIDIA, Mellanox	IBM, NVIDIA, Mellanox & Pathforward teams & ECP
Aurora	2018, 4Q	Cray/Intel	Cray/Intel & Pathforward teams & ECP
CORI	2016, 4Q	Perhaps from the ECP project	Pathforward teams & ECP
Crossroads	2020, 4Q	Perhaps from the ECP project	Pathforward teams & ECP
NERSC-9	2020, 4Q	Perhaps from the ECP project	Pathforward teams & ECP
Cheyenne	2017, 3Q	Perhaps from the ECP project	Pathforward teams & ECP
TaihuLight	2016 (2020, 4Q)	Likely a few Chinese vendors and local universities	Likely a few Chinese vendors and local universities
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	Likely a few Chinese vendors and local universities	Likely a few Chinese vendors and local universities
Hazel Hen	2015	U/U/C	Perhaps new EU investments in R&D via ETP4HPC partners.
SuperMUC	2015	Various IBM research groups	Various IBM research groups, plus ETP4HPC partners.
Piz Diant	2016	U/U/C	Perhaps new EU investments in R&D
D-Wave	2015	Primarily with US government agencies.	Primarily with US government agencies, but could grow it they expand or merge with other types of quantum computer designs.

Table 25

R&D Plans: Partnerships

Computer Names	Planned Delivery Date	Partnerships for the System	Partnerships for The R&D
UK Three System Upgrade	2018, 3Q	Various IBM research groups and ETP4HPC	IBM provides access to it data-centric and cognitive computing technologies, including its world-class 'Watson' cognitive computing platform. ETP4HPC partners.
CEA/Bull	2020, 3Q	CEA has its own R&D people and groups, plus ETP4HPC partners	Likely support from various groups across the EU. CEA has its own R&D people and groups. ETP4HPC partners.

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

IDC has long maintained that high-performance computing (HPC) leadership will be determined more by software advances than hardware progress. The hardware direction is fairly well set, while the HPC software stack – the operating system and other components between the application kernels and the hardware – faces an array of challenges that must be overcome to turn future HPC systems of all sizes into coherent, efficient, productive resources. These challenges are daunting enough that HPC vendors have turned to the open source community to help address them. These partnerships typically result in two versions of the software stack: a free open version and a paid version for users who need added capabilities and vendor support. One major initiative of this kind is OpenHPC, a close collaboration between Intel, acting as catalyst, and a growing number of OEMs and other members of the global HPC community, including open source software developers, ISVs, and end users of HPC systems. Others include OpenStack and OpenPOWER

Additional Observations

Table 27

Additional Comments & Observations

Computer Names	Planned Delivery Date	Upgrade Path
Sierra	2017, 3Q	5X-7X Sequoia Replacement
Summit	2017, 3Q	5X Titan

Table 27

Additional Comments & Observations

Computer Names	Planned Delivery Date	Upgrade Path
Aurora	2018, 4Q	18X Mira replacement
CORI	2016, 4Q	Phase I: Cray XC, Phase II: KNL
Crossroads	2020, 4Q	Must run Trinity and Sierra code, 20X Edison
NERSC-9	2020, 4Q	20X Edison
Cheyenne	2017, 3Q	3.5 times Yellowstone peak performance
TaihuLight	2016 (2020, 4Q)	U/U/C
TianHe2 A	2017 2Q or 3Q (2020, 4Q)	Intel Phi cards phased out in upgrade
Hazel Hen	2015	Upgrade/follow-on to Hornet
SuperMUC	2015	Phase 2 is upgrade to Phase 1,
Piz Diant	2016	Upgrade from 5200 K20x to 4,500 Pascal
D-Wave	2015	U/U/C
UK Three System Upgrade	2018, 3Q	Funded by the Science and Technology Facilities Council (STFC), the UK government sponsored research organization
CEA/Bull	2020, 3Q	TERA 1000 follow-on

Source: IDC 2016

U/U/C = Uncertain/Unknown/Confidential

KEY SUPPORTING QUESTIONS RESEARCHED

How important is access to this type of computer to your research/science?

"We couldn't do our science without it. We depend upon supercomputing at a national scale. We wouldn't be able to produce the seismic hazard models without it. Supercomputers have created new types of products through hazard simulation and analysis. It's very important to the work being done.", Thomas Jordon, USC

"The science I study is geophysics and astrophysics which is controlled by a nonlinear cascade over multiple scales. Information is passed over huge ranges. If I can't resolve huge scale, it's tough to get the physics right. The dream is to parameterize. It tough to parameterize the inner planet. The only way to do this is simulation. You can build theory a priori, but you don't know if it's right. If we don't have state of the art computation power, we are going to lag in addressing this issue. We won't push the envelope.", Jonathan Aurnou, UCLA

How important is it to your research to have a first class world leading supercomputer?

"We have to use the largest supercomputers and they aren't big enough. Building the software is very important. It has taken us 20 years to develop the software. We are very interest in co-design of hardware and software.", Thomas Jordon, USC

How important is it to your NATION to have a world leading supercomputer?

"We want a first class world leading supercomputer to attract the top-notch people who are writing the code on it. We need the best hardware and software.", Jonathan Aurnou, UCLA

"It's crucial just to be at the leading edge of computing period. Just read the news - not just for security and cyber - but in every way. Science coupled with all the other things that utilize supercomputing - not just for defense - will lead national development.", Jonathan Aurnou, UCLA

What would happen if you had to only use a scale out vanilla cluster or a cloud?

"A cluster is a nonstarter. But we use the cloud, not for the supercomputing aspect. We don't run large scale simulations on the cloud. It takes longer, its costlier, and less efficient.", Thomas Jordon, USC

"I would never use the cloud for computing. It's not reliable, it's expensive, and the cloud cannot guarantee access. The cloud is a disaster. It's the wrong way to go. It's a huge mistake and adds zero value. You're not going to find any money for cloud computing. It's expensive, you have to pay overhead charges on it. We are all doing cloud computing remotely. The problem is you can't guarantee the architecture; the software is commercial; you can't run it on all systems.", David Dixon, University of Alabama

"I wouldn't use it for big computing right now. Right now it doesn't look very good for problems that require intercommunication." And he added: *"We did use vanilla clusters 15 years ago. It would require building more advanced physics models. It would require better theories.",* Jonathan Aurnou, UCLA

IDC ASSESSMENT OF DIFFERENT COUNTRIES EXASCALE CAPABILITIES

In this section, IDC presents an overall view of the strengths and weaknesses of the different countries exascale plans and capabilities. Tables 28, 29 and 30 shows the evaluation for the USA, Europe and China.

Exascale Capabilities for the USA

The USA has multiple programs, strong funding and many HPC vendors, but has to deal with changing federal support, a major legacy technology burden, and a growing HPC labor shortage.

Table 28

IDC Assessment of the Major Exascale Providers: USA

Advantages	Disadvantages
The Leader: half ($) of global HPC market	Federal budget may not properly value/support HPC leadership
Advanced vendor/user bases	Legacy technology burden
Strong tech investment community (e.g., S. Valley)	Many investors misread HPC as “old technology”
The chief global supplier	Target of indigenous technology initiatives
NSCI opportunity for whole-of-nation focus on exascale	Lead agencies not embracing NSCI economic message
A good university system	HPC labor shortage
Top-notch HPC workforce	.

Source: IDC 2016

Exascale Capabilities for Europe

Europe has strong software programs and a few hardware efforts, plus EU funding and support appears to be growing, and they have Bull, but they have to deal with 28 different countries, and a weak investment community.

Table 29

IDC Assessment of the Major Exascale Providers: Europe/EMEA

Advantages	Disadvantages
Desire to seize HPC lead	27 countries to corral

Table 29

IDC Assessment of the Major Exascale Providers: Europe/EMEA

Advantages	Disadvantages
Aggregate GDP surpassed US	New rules needed to expand EU-national pooling of funds
EU funders recognize HPC's economic benefit (GDP boost)	Only in a few countries
Strong government-industry partnerships	Only ¼ of global market ($)
Advanced user base	Weak investment community
Promising indigenous technology (ETP4HPC)	ARM now owned by Japan
Strong in parallel software	Bull Atos HPC revenue small
ARM showing momentum	.
Bull Atos European OEM	.

Source: IDC 2016

Exascale Capabilities for China

China has had major funding support, has installed many very large systems, and is developing its own core technologies, but has a smaller user base, many different custom systems and currently is experiencing low utilization of its largest computers.

Table 30

IDC Assessment of the Major Exascale Providers: China

Advantages	Disadvantages
Strong government funding commitment	Smaller user base
Chinese vendors in high-growth mode -- Lenovo is now the #3 global HPC vendor	Lenovo struggling to hold some IBM customers
Many domestic core technologies are being developed	Many Academic one-off experiments; limited down-market dissemination.
Multiple groups pursue Top500 #1 position	Narrow set of HPC skills, especially hardware/software
Large cities co-fund leadership supercomputers	Very low utilization of their supercomputers

Table 30

IDC Assessment of the Major Exascale Providers: China

Advantages	Disadvantages
Little legacy software – can bypass “excess baggage”	New supercomputers are likely hard to use, and require major re-programming
Large number of science and engineering grads	.

Source: IDC 2016

FUTURE OUTLOOK

Leadership-class supercomputers have contributed enormously to advances in fundamental and applied science, national security, and the quality of life. Advances made possible by this class of supercomputers have been instrumental for better predicting severe weather and earthquakes that can devastate lives and property, for designing new materials used in products, for making new energy sources pragmatic, for developing and testing methodologies to handle "big data," and for many more beneficial uses.

There Will Be a Broad Range of New Technologies

The broad range of leadership-class supercomputers examined during this study make it clear that there are a number of national programs planned and already in place to not only build pre-exascale systems to meet many of today's most aggressive research agendas but to also develop the hardware and software necessary to produce sustained exascale system in the 2020 timeframe and beyond.

- Although our studies indicate that there is no single technology trend that will emerge as the preferred scheme, it is satisfying to note that the wide range of innovative and forward leaning efforts going on around the world almost certainly ensure that the push towards more capable, powerful leadership-class supercomputers will be successful.

IDC analysts stress however, that for almost every HPC develop project examined here, the current development effort within each organization is only their latest step in a long history of HPC development and use.

- As such, IDC analysts assess that a leading-edge supercomputer development and user facility must be continually involved in the development of new systems on a timely basis, or risk falling behind those committed to the regular, periodic acquisition and use of leadership-class supercomputers. IDC analysts believe that the cost of missing even one generation of HPC development could cause considerable difficulties for any facility looking to maintain a world-class HPC-based research capability.

Strong National Leadership-Class Supercomputer Facilities Play An Important Role

Finally, IDC analysts note that as it has been seen that time and again, successful national leadership-class supercomputer facilities play an important role in driving HPC-based developments across a nation's entire R&D base by underwing new developments in hardware and software applicable to a wide range of scientific, engineering, and industrial disciplines.

- At the same time, these programs provide significant support for nation's domestic HPC supplier ecosystem to remain at the forefront in global technology developments.
- IDC believes that countries that fail to fund development of these future leadership-class supercomputers run a high risk of falling behind other highly developed countries in scientific innovation, with later harmful consequences for their national economies.

DISCLAIMER AND COPYRIGHT

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-community.com
www.idc.com