Hewlett Packard Enterprise

# Improving the energy efficiency of modern supercomputers

Novel approaches for driving optimal performance per watt

## Contents



**Technical white paper**

As the world prepares to address the grand challenges emerging in fields like aerospace, physics, nuclear science, national security, and healthcare, the IT industry has been working diligently to test innovative technologies and consider the requirements for the next generation of powerful supercomputers. There is no question that it has become imperative to develop a new breed of tremendously complex, high performance computing (HPC) systems in order to achieve the computational capacity required to fuel new breakthroughs, accelerate discoveries, and continue scientific and technological advancements. However, achieving these new levels of computing power is quickly proving to be an energy-intensive proposition.

Not long ago, computing systems that were among the largest of their kind operated on 250 kilowatts (kW) of power – and that was considered substantial energy use at the time. The systems that are in development now are exceeding power consumption of 20 megawatts (mW), proving that as these systems rapidly evolve in size and capability they are also proportionately increasing their energy usage by a few orders of magnitude.

At such considerable levels of power consumption, the costs associated with operating these machines quickly becomes an issue. Just one megawatt of power is roughly equivalent to operating costs of just under $1 million per year, which means that running a system of this magnitude could cost anywhere from $60 to $100 million over the course of its usable lifecycle – and that's in addition to the millions of dollars that were likely spent upfront to procure the machine. There are also increasing concerns about the effects of this power consumption on the environment, and the HPC community is becoming more focused on eliminating wasteful or inefficient approaches to energy consumption and usage. Because of these exorbitant costs and environmental implications, it's becoming clear that addressing the energy challenges of modern supercomputers will be an important piece of the puzzle in order to meet the world's future demands for compute capacity.

The questions we face now are: How do we continue to justify growing these systems even larger? How will we achieve the performance needed without excessive operating costs and damaging environmental effects? It is obvious that this cannot be accomplished without fresh approaches to system design, cooling, and power delivery that promise to increase computing power without prohibitive increases in energy demand. In this paper, we will address these issues, offer new insights into what's on the horizon in supercomputing energy efficiency, and detail the steps that Hewlett Packard Enterprise (HPE) is taking to help resolve this computational energy crisis.

## Energy efficiency will enable even greater computing capabilities

As the industry races to deploy an exascale-capable machine by 2021, there is a strong urge to pack as much compute into each square centimeter of silicon as customers can afford. However, this "performance at any cost" mentality will only serve to produce extremely inefficient, wasteful systems that won't be sustainable to usher in the next level of computing. It's critical for system vendors to start from the viewpoint that they aren't working to improve energy efficiency solely for a ranking on a list or a FLOPs per watt record, but because energy efficiency will enable even greater computing capabilities.

In the quest for higher levels of performance, it can be easy to lose sight of the fact that efficient energy usage is a key enabler of the mission to accomplish more work in the same footprint for less cost. Energy efficiency of both the data center facility and the IT equipment itself is of equal concern to ensure that the majority of the energy consumed is put to good use for actual computing; in other words, the power that is delivered to the productive portion of the system is the power that matters most. At the same time, computing at this level creates tremendous quantities of heat, and offloading it is crucial to enable these machines to continue operating at peak performance without burning up system components or failing altogether.

While recent trends suggest that the data center industry has been somewhat successful at improving energy efficiency as it scales-up computing capability, it is undisputable that supercomputing systems are gradually ramping up the massive amounts of energy they consume each year. The global data center power market amounted to $15 billion in 2016, and this number is expected to nearly double to $29 billion by 2020. By that time, data centers will consume 140 billion kilowatt hours of electricity annually, which will cost American companies an estimated $13 billion per year in electricity bills. These trends prove that as the data center market expands, power requirements are maxing out, which is placing the utmost focus on lowering the operational costs of extreme energy usage.

Even though improving the energy efficiency of modern supercomputers is proving to be a challenging endeavor, it's also a very necessary one in order to meet the world's future demands for compute capacity. The intended delivery date of a capable exascale system is less than four years away, and as this deadline approaches, it's clear we simply will not get there until we begin to resolve some of these issues. The industry is being challenged to dramatically accelerate performance under defined constraints, and from this discomfort is where true system design innovation will emerge. Ultimately, this exercise will produce a new breed of computing systems that are more intelligently designed and efficient than ever before, which will be a critical development not only for the computing industry, but also for the world in general.

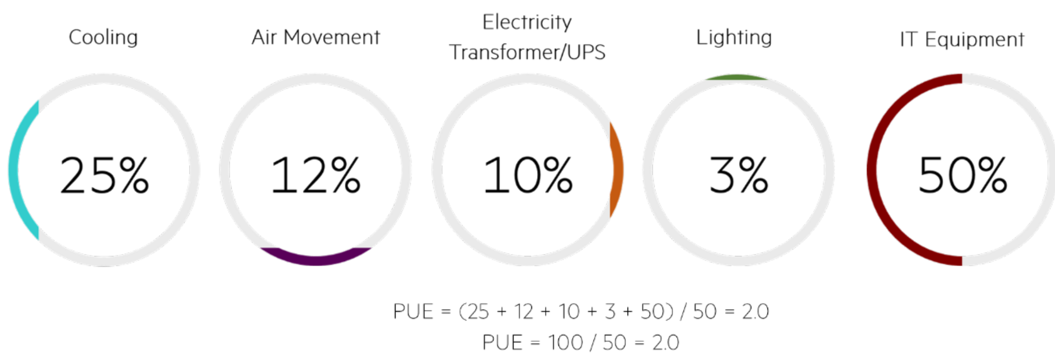
## An examination of energy efficiency metrics

Over the years, as energy efficiency has gradually moved to the forefront of the challenges in modern supercomputing, there are a number of metrics that have been developed to measure it. The most popular and widely-used metric for determining data center energy efficiency is power usage effectiveness (PUE), which gauges the relationship between the total energy consumed by the entire facility and the amount of energy consumed strictly by the IT equipment.

$$PUE = \frac{\text{Total Facility Energy}}{\text{IT Equipment Energy}}$$

An ideal PUE value of 1.0 would indicate 100% efficiency, meaning that all the energy consumed by the facility is being used to directly power the IT equipment. As PUE ratios increase above 1.0, this indicates lower levels of efficiency because it means a percentage of the energy supply is being used to operate the facility itself (e.g. switching gear, back-up generators, external power supplies, facility lighting, cooling fans, pumps, chillers, HVAC, etc.) versus for useful computing.

Here is a simple real-world example of a PUE calculation:




*Source: "PUE in the Past; Towards TUE in the Future" by Lawrence Berkeley National Laboratory, Environmental Energy Technologies Division (EETD), High Technology and Industrial Systems Group*

**Figure 1.** Example PUE calculation

The PUE metric was developed in 2008 by members of the Green Grid Association, in an effort to put in place a set of metrics and standards that could be formally adopted by the industry to improve worldwide data center energy efficiency. The widespread adoption of the PUE metric has been a positive development for the industry, because it prompted a paradigm shift where facility operators and system vendors became more concerned with measuring the percentage of their valuable energy resources that were being used solely for computing. In recent years, major improvements in energy efficiency efforts for both data centers and HPC systems have served to drive down PUE scores at data centers all over the world. Recent studies show a wide range of PUE values for data centers, but the overall average tends to be around 1.8, with data centers focusing on driving efficiencies typically achieving PUE values of 1.2 or less.

When assessed properly, PUE can provide useful insight into the impact of infrastructure or process changes on power utilization, as well as offer a straightforward view of overall operations and a simple way to pinpoint opportunities to increase efficiency. However, it is becoming clear that using only PUE to measure energy efficiency has some limitations. The primary concern is that PUE does not provide any insight into the distribution of power or cooling losses occurring inside the IT equipment itself, which is a key component that plays into overall data center efficiency. For example, the PUE score for a very efficient data center operating with grossly inefficient IT equipment might not be completely indicative of the poor utilization of energy by the equipment itself. PUE also incorporates many measurements from a variety of infrastructure components, and requires a significant amount of data comprising multiple parameters that must be measured over the course of an entire year, making it hard to validate. Lastly, PUE does nothing to measure the environmental impact of high levels of power consumption, including the carbon footprint resulting from data center construction and operation.

Particularly for data centers that have already achieved a PUE well below the industry average of 1.8, it becomes important to start incorporating additional metrics to gain a more detailed view into where things can be improved. To address these shortcomings, the Energy Efficient High Performance Computing Working Group (EEHPCWG) introduced two new metrics – IT Equipment Utilization Effectiveness (ITUE) and Total Utilization Effectiveness (TUE). Using these metrics in conjunction with PUE can offer a better approach and provide a more comprehensive view of overall energy efficiency.

ITUE is very similar to the PUE metric for data center facilities, except that it zeros in on the efficiency of the IT equipment itself. ITUE is calculated the same way as PUE, except that it goes a step further and provides a view into power efficiency once the energy enters the IT equipment.

$$\text{ITUE} = \frac{\text{Total Energy into IT Equipment}}{\text{Total Energy into the Compute Components}}$$

Just as PUE measures how much energy is diverted to run cooling devices, power supplies, lights, pumps, and other infrastructure components that support the IT equipment, ITUE measures how much energy is used for the supporting infrastructure inside the equipment, such as internal fans, power supplies, and voltage regulators (VRs). The resulting ratio provides a glimpse into how much energy is being consumed by the IT equipment's infrastructure components, versus the "productive" elements like CPUs, memory, storage, and networking.

TUE is a calculated value that incorporates both ITUE and PUE:

$$\text{TUE} = \text{ITUE} \times \text{PUE}$$

The TUE value is equal to the total amount of energy coming into the data center divided by the total energy to the computational components inside of the IT equipment. Since its introduction, TUE has been proven to be more accurate and consistent than PUE values, because it incorporates the IT support inefficiencies that PUE leaves out to provide a comprehensive view of energy efficiency at the equipment level.

While there is no question that PUE can be a simple and effective data center energy use efficiency metric, it is not perfect for some comparisons. Using ITUE (or "server-PUE") in conjunction with PUE results in TUE, which addresses some of the weaknesses with using the PUE metric alone. As supercomputing systems grow in size and capability and energy efficiency becomes an even more complex challenge to resolve, system vendors must understand that it will be impossible to reach total utilization efficiency without gaining some visibility into what's happening not only inside the data center, but inside the IT equipment as well.

## Accelerating performance while reducing energy consumption

With an aggressive energy strategy that stretches back over more than a decade, HPE continually strives to dramatically improve the processing power of next-generation systems while reducing energy consumption by an order of magnitude. Drawing on a long history of innovation in system design and solution implementation expertise, HPE's two-pronged approach to energy efficiency consists of pushing the envelope on new research initiatives in energy efficiency, and repeatedly optimizing new and existing system designs to be as power-efficient as possible. Whether advancing the study of evolving technologies such as low-energy photonics interconnects, or paving the way to the exascale era by creating the world's largest single-memory computer, HPE's fresh approach to energy efficiency is based on an ongoing promise to fundamentally rethink the way large-scale systems are designed and built.

HPE's commitment to energy efficiency begins with ultra-efficient system designs. Their design approach begins by breaking down the ITUE metric into sub-groups within the system, which allows each component to be analyzed individually in terms of energy usage efficiency. For example, here is a look at two ITUE calculations for the HPE SGI 8600 system:

| System power delivery effectiveness | |
|---|---|
| AC power distribution efficiency | 99.5% |
| 12V power supply conversion efficiency | 95.0% |
| 12V power distribution efficiency | 99.5% |
| Aggregate VRM conversion efficiency | 89.0% |
| **Total power delivery efficiency** | **83.7%** |
| **$ITUE_{Power}$ = 1/Efficiency** | **1.19** |

| System heat removal effectiveness | |
|---|---|
| CDU capacity | 300.00 kW |
| CDU power draw | 4.5 kW |
| CDU heat removal efficiency | 1.5% |
| Cooling rack capacity | 72.0 kW |
| Cooling rack power draw | 6.0 kW |
| Cooling rack heat removal efficiency | 8.3% |
| **Aggregate heat removal efficiency** | **3.7%** |
| **$ITUE_{Cooling}$ = 1+Efficiency** | **1.04** |

$$ITUE = ITUE_{Power}\,(1.19) \times ITUE_{Cooling}\,(1.04) = 1.24$$

Figure 2. Two ITUE calculations for the HPE SGI 8600 system

In this example, two specific contributors to overall system energy efficiency – power delivery and heat removal – are broken down in order to assign an ITUE value to each. To evaluate system power delivery, the efficiency levels of AC power distribution, 12V power supply conversion, 12V power distribution, and aggregate VRM conversion allow us to arrive at a final $ITUE_{Power}$ value of 1.19. In the analysis of system heat removal, CDU capacity, CDU power draw, CDU heat removal, cooling rack capacity, cooling rack power draw, and CDU heat removal are all data points that contribute to a final $ITUE_{Cooling}$ value of 1.04. Breaking ITUE down to such a granular level allows system engineers to pinpoint exactly where system inefficiencies are, identify specific areas where energy efficiency can be improved, and better understand how power-efficient the system will be as a whole.

Back in the early 90's, one of HPE's earliest efforts was to become involved in addressing the energy inefficiencies of external power supplies. With just over a billion external power supplies active in the U.S. alone at that time, it was estimated that without efforts to increase efficiencies and reduce "no load" power consumption these supplies would account for around 30% of total energy consumption in less than 20 years. This led to the development of new regulations regarding the energy efficiency of external power supplies, and prompted an industry-wide shift to adopt more efficient technologies. Teams within HPE and SGI have been working for years on pioneering new breakthroughs in power supply efficiencies, even collaborating with fellow industry thought leaders to develop new rating standards on which power supplies are now certified, such as Bronze, Silver, Gold, Platinum, and even Titanium. Each new standard has helped external power supplies gain higher levels of efficiency across the board, and served to significantly impact the global power consumption attributable to supercomputers.

Perhaps the most pressing issue of all is what to do with the considerable amount of heat that is generated from the computationally-demanding activities of modern supercomputers. If this heat is not shed effectively, system components inevitably begin to break down and performance of the entire system suffers. Historically, processors often weren't run at full performance because there was no effective way to cool them. This push to develop more efficient cooling techniques continues to be one of the most persistent and critical challenges of improving the energy efficiency of large-scale systems.

An initial reaction was to simply lower the ambient temperatures in the rooms where the IT equipment was held. However, as innovative approaches to cooling continue to take hold across the industry, current practice has become about as efficient as turning down the air conditioning in an entire house just to chill a glass of water. Because the heat-carrying capacity of air is quite low, fighting heat with cold air alone requires massive amounts of very cold air; however, the heat-carrying capacity of water is much higher, and as a result, liquid cooling techniques have emerged as a more efficient means of cooling large-scale systems. In fact, liquid cooling techniques have been shown to be 1,000X more efficient than air cooling techniques[1], enabling processors to run at much higher levels of performance without the risk of overheating. Overall, these higher levels of efficiency have the potential to reduce energy usage at the system level by up to 28%.

[1] HP case study, "National Renewable Energy Lab slashes data center power costs with HP servers," December 2013

HPE's innovative approach to liquid cooling can be observed in the HPE SGI 8600 system, which is a liquid-cooled petascale system that builds on the proven SGI ICE XA architecture. Offering superior power and cooling efficiency as well as substantial savings in operational costs as compared to traditional air-cooled designs, the HPE SGI 8600 is designed as a "cell" architecture, with each E-cell containing two 42U high E-racks separated by a cooling rack.

**Figure 3.** HPE SGI 8600 E-Cell design

The E-cell is a sealed unit which uses closed-loop cooling technology to provide 100% heat removal and does not exhaust heated air into the data center. A direct-attached liquid-cooled cold sink provides for efficient heat removal from high power devices including processors, GPUs, and switches via an auxiliary cooling distribution unit (CDU). The E-cell utilizes facility-supplied water for cooling and will not add any heat to the data center if the water temperature is within 45–90° Fahrenheit (7–32°C). Additionally, advanced HPE system management software provides bare-metal provisioning, detailed system health management, and innovative power optimization capabilities.

Whether reinforcing their commitment to design systems to be more energy-efficient, pioneering new breakthroughs in power supply efficiency, or exploring innovative approaches to cooling, HPE is leading the charge to the next generation of powerful, efficient supercomputers. Because energy efficiency incorporates nearly every system component and requires new approaches to a number of computing technologies, HPE is partnering with other leaders in the industry and taking a collaborative approach to energy efficiency that will drive wide-scale improvements across the industry.

## Industry-wide recognition for HPE's most efficient systems

The Green500 list is an industry-recognized ranking of the top 500 supercomputers in the world according to energy efficiency, which is measured primarily by the widely-used FLOPS per watt metric. The list was developed as a complement to the Top500 list – which ranks the top 500 supercomputers by performance – to put a premium on energy-efficient operations in the era of "performance at any cost" computing. The latest version of the list, released in June 2017, proves that energy efficiency efforts have begun to proliferate across the industry. In fact, the greenest supercomputers in the world more than doubled their energy efficiency over the past year, representing the largest leap forward since the inaugural Green500 list was released in November 2007.

In addition to continually climbing the TOP500 list, HPE gained industry-wide recognition this year by clinching the #1 spot on the latest Green500 list. TSUBAME 3.0, an HPE SGI 8600 system currently installed at the Tokyo Institute of Technology (Tokyo Tech), is currently #61 on the TOP500 list and recently received the Green500 Award for the most energy-efficient system in the world. For the first time ever, this secured HPE the top-ranked system on the Green500, solidified their largest presence to date on both the TOP500 and Green500 lists, and further extended their position of leadership in energy efficiency and HPC design innovation.

TSUBAME 3.0 is Japan's fastest artificial intelligence (AI) supercomputer yet, which achieved an unprecedented 14.110 gigaflops per watt during its 1.998-petaflop Linpack performance run. TSUBAME 3.0 operates with an impressive PUE of 1.033, and features one of the most energy-efficient designs in the world. The system currently resides at the Global Scientific Information and Computing Center at Tokyo Tech, which is one of Japan's most prestigious research universities dedicated to science and technology. TSUBAME 3.0's lightning-fast, innovative deep learning capabilities are used by scientists and researchers every day to power new breakthroughs in complex and critical areas like climate change and cancer treatment. It is worth noting that TSUBAME 3.0 is an actual production-grade system being used each day to further simulation and modeling, HPC, deep learning, and other complex workloads. It's ability to accomplish all of this and still capture the top spot on the Green500 list proves that the list is focused on recognizing the most productive energy-efficient systems that exist today as opposed to "green stunt machines" just aimed at achieving the lowest FLOPS per watt number possible.

> "Our partnership with HPE has worked successfully to deliver the world's most energy efficient HPC and deep learning platform – and we look forward to further collaboration on the path to capable supercomputers and sustainable HPC innovation."
>
> – Satoshi Matsuoka, TSUBAME leader and Professor at the Tokyo Institute of Technology

To power new, energy-efficient breakthroughs in the realm of earth and space science, HPE recently collaborated with the National Aeronautics and Space Administration (NASA) to build a powerful new supercomputer that will help the agency keep pace with the burgeoning requests for supercomputing capabilities needed to support their most crucial research projects. An SGI ICE X cluster, the system is NASA's first HPC system to take a modular design approach, which allows for greater efficiency in the cooling of processors. This modular HPC technology will dramatically reduce the amount of water and electrical resources needed to cool the system, which is expected to save the agency approximately 1.3 million gallons of water and 1 million kilowatt-hours of energy each year, which is equivalent to the annual energy usage of approximately 90 households. The system operates with an impressive PUE measurement in the range of 1.03 to 1.05, which helps significantly reduce NASA's impact on the environment as they continue to expand and meet the ever-increasing need among their scientists and researchers for HPC capacity.

## Ushering in a new era of computing

As the industry continues to advance various energy efficiency initiatives, it is simultaneously building up to exascale-capable systems. These yet-to-be-built systems will achieve unprecedented levels of performance, offer breakthrough computing capabilities, and set a whole new standard for system efficiency. To achieve this goal, the U.S. Department of Energy (DoE) released their PathForward Request for Proposals (RFP), which aims to accelerate and support the research necessary to deploy the nation's first exascale supercomputers through research and development (R&D) funding. The PathForward initiative began by outlining the four chief challenges of exascale computing, which include parallelism, memory and storage, reliability, and energy consumption. In terms of energy consumption, the PathForward Technical Requirements established a maximum power envelope of 20 to 30 mW for a capable exascale system. HPE was recently selected as one of the six leading U.S. technology companies that will collectively receive $258 million in funding from the PathForward Program to lead the industry into the exascale era.

Even with all of the recent advancements in energy efficiency, it is clear that there are some daunting challenges that remain on the path to exascale; namely, how to increase performance by an order of magnitude while staying within the defined power envelope. The DoE estimates that simply scaling up current petaflop computer architectures to an exaflop level would require a massive 200 mW of energy to operate, which is unsustainable and cost-prohibitive. For reference, the TSUBAME 3.0 system operating at a peak of 14 gigaflops per watt would consume approximately 70 mW of energy if scaled up to one exaflop, which is more than double the DoE's stated goal.

As supercomputers rapidly evolve, the energy consumption of IT equipment has gone beyond simply a technical issue and become a social issue as well. This shift has prompted the computing industry to begin focusing on reducing wasteful and inefficient approaches to energy management. Energy re-use will become an increasingly important consideration as the industry builds up to a new generation of large-scale systems, and this has been a key tenet of HPE's energy strategy for years. Unfortunately, some are still choosing to deploy data centers in geographic areas where energy can be purchased inexpensively, and pushing the heat exhausted by their IT equipment into the atmosphere or local water sources with little to no efforts made to reuse those vast quantities of waste energy. Particularly with increasing global warming and scarcity of water concerns, the industry simply cannot continue to turn a blind eye to these practices. HPE's energy strategy is geared to developing systems where all the energy required for system operation can become part of a "flow," where waste energy can be converted into a reusable product that can be used to power other downstream processes. HPE engineers are continually striving to design IT equipment in a way that enables customers to get the maximum usage of every precious kilowatt that they paid for, in turn helping to inch energy reuse effectiveness (ERE) ratings as close to 1.0 as possible. The National Renewable Energy Laboratory's (NREL's) Peregrine system, an HPE Apollo 8000 design that is currently the world's largest HPC system dedicated to advancing renewable energy and energy efficiency technologies, is an exemplary example of an HPC system that is designed to recover waste energy for reuse. Operating out of NREL's HPC center in Golden, Colorado, Peregrine features an innovative warm water liquid cooling system that allows waste heat to be captured and reused to offset heating loads in other areas of the facility.

With worldwide demand for energy continuing to expand, one of the effects of this rising power consumption is a large carbon footprint that must be reduced. Teams within HPE are exploring new ways to reduce the carbon emissions and harmful waste that result from high energy usage and pollute the environment, including doing deep research into the potential use of alternative energy sources to power HPC systems. Emerging technologies like fuel cells are already helping some progressive companies power their data centers and reduce their energy usage, as well as offering the potential to provide renewable energy to a variety of diverse commercial applications in a more sustainable and efficient way.

Fuel cells convert chemical energy (hydrogen gas and oxygen) into electrical energy using a process that produces significantly fewer carbon emissions than standard fuel-burning technologies. They also allow facilities to generate their own power on-site, which reduces the reliance on power grids and eliminates the inefficiencies associated with transporting energy over long distances to data centers. Fuel cells emit only energy, water, and heat, and that heat can even be used for secondary processes such as facility heating and cooling. Centralized, high-capacity fuel cell installations are already being used to power some newer data centers, but there is also a substantial amount of research around placing fuel cells directly inside server racks to increase their self-sufficiency and reduce energy losses inside the equipment itself.

NREL and a network of partners that includes HPE, Intel, Daimler, Microsoft, and Power Innovations are currently working on a proof-of-concept demonstration that uses a hydrogen fuel cell to provide power to the lab's data center. The 65 kilowatt automotive hydrogen fuel cell, which was provided by Daimler, is fed by NREL's renewably-generated hydrogen and will provide high-voltage, DC power directly to 2-3 racks of IT equipment provided by HPE and Intel. While the amount of hydrogen gas that is required to generate energy remains a chief challenge in terms of storage and distribution, many companies like NREL are already experimenting with ways to produce their own hydrogen supplies on-site and progress is made every day to make these green energy sources a reality for modern data centers.

## Conclusion

Energy efficiency has become an increasingly crucial requirement for supercomputers as the industry races to assemble the next generation of powerful computing systems. Not only does furthering energy efficiency promise to create a whole new breed of powerful and efficient systems, but it will also be a critical enabler of the ability to achieve even greater computing possibilities. We've come a long way, but it's clear there is still some difficult – yet critically important – work that the industry must do in the realm of energy efficiency. Enhancing computing power without prohibitive increases in energy demand will undoubtedly require a complete transformation spreading across every aspect of HPC system design. To help achieve this goal, HPE is working to drive new advancements in energy efficiency research, improve power-efficiency in both new and existing system designs, and continue to actively research, develop, and test novel approaches that will not only dramatically improve the energy efficiency of modern supercomputers, but also help the industry usher in a new era of computing.

# Learn more at
hpe.com/servers/hpc-server-sgi8600



a00028651enw, October 2017