LBNL's High Performance Computing Center

Continuously Improving Energy and Water Management

BY JINGJING LIU, P.E., BEAP; NORMAN BOURASSA, ASSOCIATE MEMBER ASHRAE

High performance computing (HPC) centers are unique in certain aspects such as task scheduling and power consumption patterns. However, they also share commonalities with other data centers, for example, in the infrastructure systems and opportunities for saving energy and water. The success and lessons learned at LBNL's National Energy Research Scientific Computing Center (NERSC) can be useful for other data centers with proper adoption considerations.

NERSC HPC Facility Today

Lawrence Berkeley National Laboratory's (LBNL) HPC center, NERSC, has a mission to support U.S. Department of Energy (DOE) Office of Science-funded scientific research through providing HPC resources to science users at high-availability with high utilization of the machines.[1]

NERSC has been located in Shyh Wang Hall (*Figure 1*), a LEED Gold-certified building, on LBNL's main campus since 2015. The current main production system is Cori (*Figure 2*), a 30 petaflops* high performance computing system. The facility consumes an average 4.8 gigawatt-hours per month. To track its energy efficiency, the NERSC team has implemented a rigorous 15-minute interval measurement of power usage effectiveness (PUE),† drawing from an extensive instrumentation and data storage system referred to as Operations Monitoring and Notification Infrastructure (OMNI).[3] So far, the team has achieved over 1.8 gigawatt-hours of energy savings and 0.56 million gallons (2.1 million L) of

*Unit of computing speed equal to one thousand million million (10^{15}) floating point operations per second.

†PUE is the ratio of total facility energy to IT energy; it's a measure of how effectively the IT equipment is served by the power distribution and cooling systems. A lower PUE number is better, and 1.0 is the lowest theoretical number. There are three levels for PUE measurement per Green Grid:[2] Level 1, Level 2 and Level 3 (basic, intermediate and advanced). At each level, the IT load is measured at different points: the UPS output, power distribution unit (PDU) output and directly from the input of IT equipment.

Jingjing Liu, P.E. is a program manager and researcher at Lawrence Berkeley National Laboratory (LBNL) in Berkeley, Calif. Norman Bourassa works in LBNL's National Energy Research Scientific Computing (NERSC) building infrastructure group.

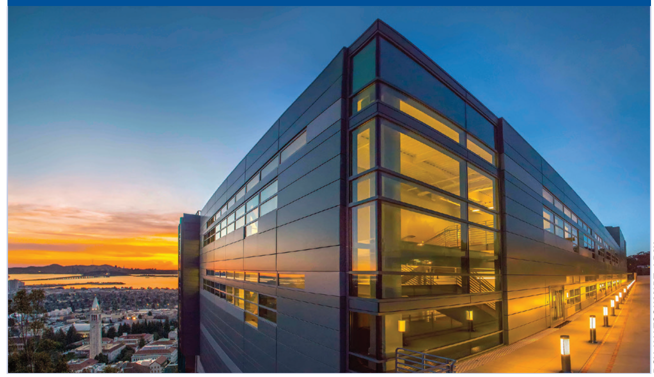
water savings annually. The current Level 2 PUE annual average is a very efficient 1.08,[4] but the team is working on lowering it further.

Besides the data center's facility infrastructure efficiency, the team also pioneered the use of an open-source scalable database solution to combine facility data with computing system data for important daily operational decisions, ongoing energy efficiency tuning and future facility designs.

There are many important reasons why LBNL gives NERSC energy efficiency significant attention and resources despite the relatively low electricity prices at LBNL:

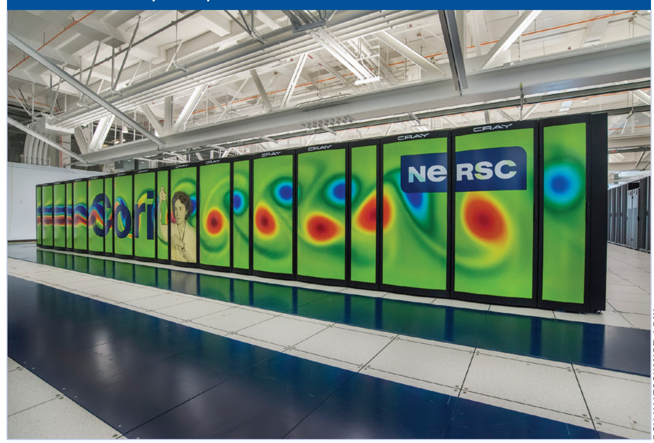
- NERSC consumes about one-third of LBNL's total energy;
- Energy efficiency requirements from federal law and the University of California, which is under contract to operate LBNL;
- A strong lab culture of sustainability and environmental conservation; and
- The compressor-free cooling systems at times require close attention to operating conditions and settings to maintain energy efficiency.

FIGURE 1 Shyh Wang Hall, home of LBNL's National Energy Research Scientific Computing Center.



ROY KALTSCHMIDT, LBNL

FIGURE 2 Cori supercomputers at NERSC.



ROY KALTSCHMIDT, LBNL

NERSC Facility Design

Before moving to its current home, NERSC was located at a facility in Oakland, Calif., that had an estimated PUE of 1.3. Designing a more efficient new facility on the main campus was a priority of LBNL management. One bold measure was to take full advantage of the mild local weather in Berkeley and eliminate compressor-based cooling, which is most commonly used for high-availability data centers.

The new facility is cooled by both outdoor air and cooling tower-generated cooling water. Because the installed peak compute power was 6.2 megawatts—only about half of the compute substations' full capacity—air-handling units (AHUs) and cooling towers are optimally sized using modular concepts, with help from the lab's Center of Expertise for Energy Efficiency in Data Centers (CoE),[5] leaving space for future expansion. This saved tremendous cooling equipment cost.

Another important business decision was not to back up any in-process scientific compute jobs with an uninterruptible power supply (UPS) or backup generator systems. The UPS system (with energy saver system mode retrofitted later) was only sized to transfer any finished computing results already within the short-term memory over into UPS-supported air-cooled file system racks, along with one AHU, until the diesel generator takes over. These design decisions allowed NERSC to minimize its infrastructure system capital costs and, hence, be able to invest these savings in computing equipment and other priorities.

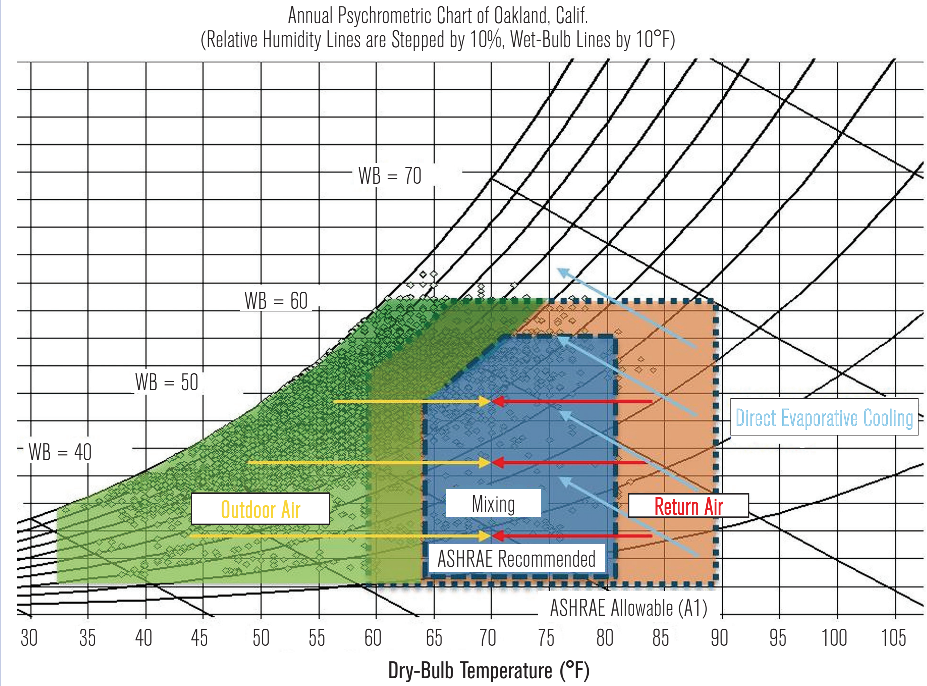
The LBNL team did not stop pushing the envelope for efficiency after the initial design. NERSC generally refreshes its supercomputers every three to five years, and each new generation is more efficient, typically with a three to five times increase in computing throughput and only two times increase in computing power.[6] The upcoming pre-exascale[‡] machine (Perlmutter[7]), expected in 2021, requires a number of infrastructure upgrades. The facility currently has a power service capacity of 12.5 megawatts, which will be upgraded to 21.5 megawatts for Perlmutter.

According to national electrical codes, a new

‡The scale of 10^{18}.

substation would be needed to support the added mechanical cooling equipment. However, under the same codes, the team successfully used archived OMNI power meter data to show that a new substation was unnecessary since the most recent year of monitoring data showed it never exceeded 60% of the total individual equipment specified peak power ratings.[3] This decision saved about $2 million—a great example of how advanced data analytics enabled managerial capital budget decisions involving collaboration with the lab's facilities division.

FIGURE 3 Psychrometric chart showing Oakland weather and cooling strategies.



Need for Ongoing Commissioning

While compressor-free cooling system design provides large energy savings, it also brings challenges with indoor environmental control. This was another operational driver for a highly detailed OMNI instrumentation capability. NERSC operators can pull three "levers" to meet the varying cooling load: airside economizer,[§] waterside economizer,[#] and direct evaporative cooling inside the AHUs. Sophisticated control sequences are needed to meet LBNL's energy efficiency goals, which require operator attention. Therefore, an ongoing commissioning (OCx) process for cooling system troubleshooting and optimization is being used. To facilitate this process, a commissioning consultant was hired in 2016. They conducted an assessment revealing a potential of three gigawatt-hours of annual energy savings.[4]

To help understand control sequences, various outdoor air conditions are illustrated on the psychrometric chart in *Figure 3* with annual hourly local weather data points. (The psychrometric chart for Oakland is used to approximate Berkeley weather.) The actual local weather in recent years has seen hotter conditions more frequently than shown in the chart. The chart shows the "recommended" and "allowable" ranges of IT equipment intake air conditions for the "A1" class data centers outlined in the ASHRAE thermal guidelines.[8] This most recent guideline had relaxed the low end of the humidity range from 42°F (5.5°C) dew point (DP) to 16°F (–9°C) DP for all data center classes compared to the previous 2011 version[9] that was used during Wang Hall design (which houses NERSC). In response to the new guideline, the NERSC team implemented control sequence changes eliminating humidification when outdoor air is cold and dry (below 42°F [5.5°C] DP), which saved water by operating the direct evaporative coolers less.

Outdoor air is cooler and with higher relative humidity than the "recommended" range most of the year (green area in the chart). Therefore, mixing outdoor air with return air to achieve desirable supply air conditions is a primary strategy. In addition, cooling tower water is used in AHU cooling coils when outdoor air is warmer than 74°F (23°C).[II] For the limited hours when the outdoor air is even warmer (about 80°F [27°C] or higher) but less humid, evaporative coolers are used to bring the

[§]Using cool outdoor air as a low-energy cooling source to avoid or reduce direct-expansion based cooling for energy savings.

[#]Using cooling tower generated water as a low-energy cooling source.

[II]There is a temperature rise of about 5°F to 6°F (2.8°C to 3.3°C) between the AHU supply air and rack intake air.

air to the "recommended" or at least to the "allowable" range.

Ongoing commissioning is even more critical under emergency situations. One example was the deadly 2018 "Camp Fire" in Butte County, Calif. The NERSC facility experienced high air pollution and had to completely shut off outdoor air for an extended period. This scenario was not anticipated during design and triggered a series of control issues. The facility used only AHU cooling coils with tower-cooled water as the cooling source since direct evaporative cooling creates a rapid indoor humidity increase. 100% return air operation lasted for two weeks, and seven building management system (BMS) logic flaws surfaced and were fixed.[4]

It also led to condensation on a cooling water manifold in the underfloor plenum, triggering a water leak alarm. Condensation was considered extremely unlikely during design because the facility is normally tightly coupled to outdoor wet-bulb temperatures. However, this event proved condensation can be an issue if 100% return air persists long enough. This event resulted in new air pollution control sequences,[4] which have improved facility resilience during the increasingly active annual wildfire seasons in California.

One key NERSC energy efficiency team's success is a deep recognition of the need for commissioning building operations in an ongoing rather than one-off manner. They've adopted a continuous improvement cycle approach (*Figure 4*)—trial and error with data validation being an important part. The team meets regularly, led by the chief sustainability officer, to follow up on implementation/testing progress and obstacles. It also allows the team members to identify connections with other efforts and serve as advocates for commissioning projects in their interactions with other teams. The team's focus has evolved over time, which is illustrated in *Figure 5* with approximate timelines. Some of the high impact efforts were:

- **Submetering[**] and PUE Tracking**: A detailed assessment identified where additional submeters were necessary based on impact. One challenge was that the existing circuit breaker trip units were not accurate on light loads.
- **Air Management:**[††] One challenge was the exhaust air duct's location relative to cold aisles,[‡‡] which was addressed by improved containment.
- **Cooling System Projects**: One example is adding a second heat exchanger and installing a booster pump for the offices comfort-cooling loop to lower pumping

FIGURE 4 Continuous improvement cycle used for NERSC's ongoing-commissioning.

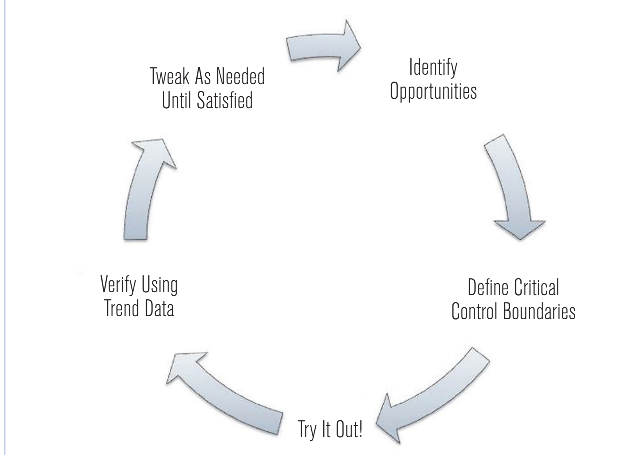
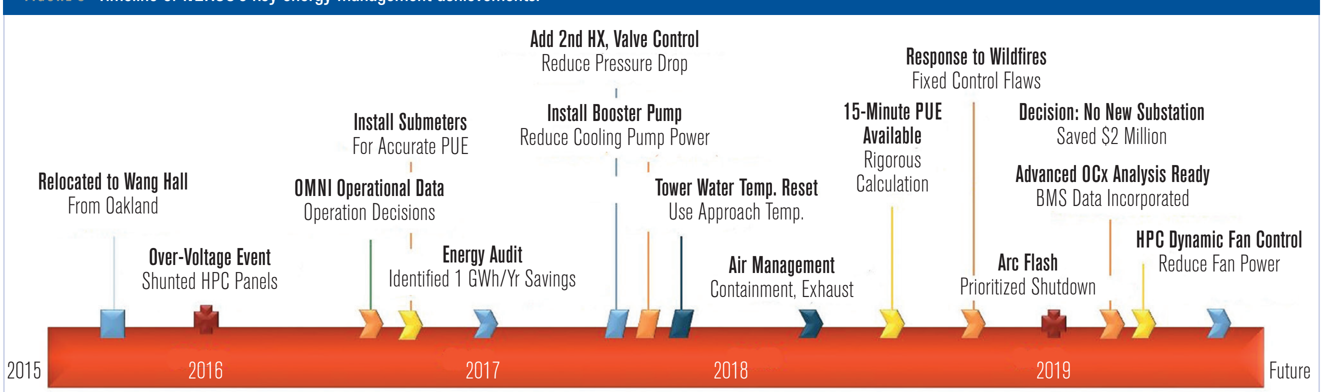


FIGURE 5 Timeline of NERSC's key energy management achievements.



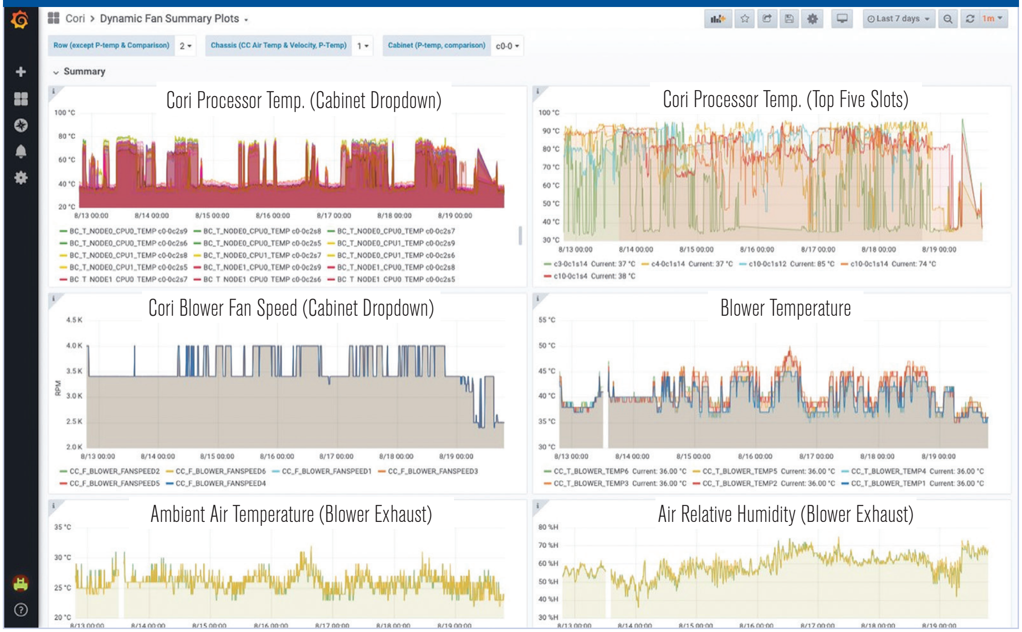
[**]Submetering involves using power meters to measure electricity use at a level below the utility meter.

[††]The practice of effectively separating the airstream on the cold, intake side and the hot, exhaust side of IT equipment. The intent is to minimize cold supply air bypassing to the exhaust side and also reduce hot exhaust air recirculating back to the intake side.

head in the data center loop and several control adjustments.

- **Collaboration With IT Vendor**: An example was OMNI access to onboard HPC system performance data. The new application programming interface (API) channels developed by the HPC manufacturer originally for NERSC now benefit the manufacturer's other HPC customers.
- **Data Analytics**: Design and implementation of the OMNI data collection and visualization system architecture and adoption of a data analytics software platform that uses advance semantic tagging database methods for building and HPC system data analytics.

FIGURE 6 OMNI operator dashboard view for Cori dynamic fan controls.

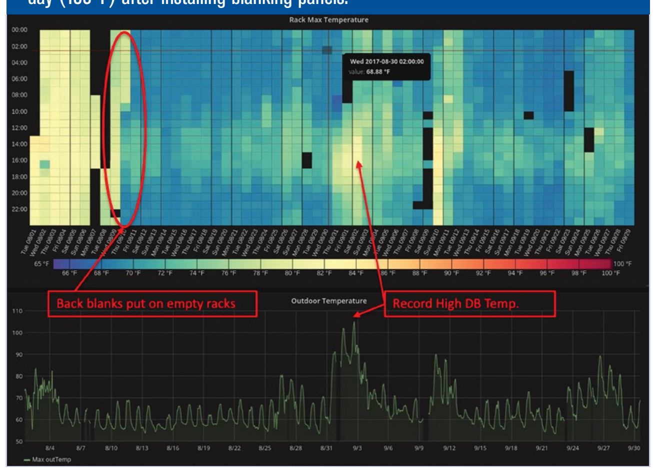


Invest in Data Storage and Analytics

One of the NERSC team's groundbreaking efforts was designing and implementing a central data repository, the OMNI, which currently stores data at 100,000 points per second. OMNI is a unique and truly integrated architecture in that facility and environmental sensor data are correlated to compute systems telemetry, job scheduler information and network errors for deeper operational insights.[3] It is a major step forward for data centers to break long-standing data silos between facilities and IT systems. *Figure 6* shows one example of the many OMNI dashboard views; in one screen for dynamic fan controls, it displays information such as processor temperatures, Cori fan speeds, air temperatures and outdoor conditions.

The original motivation for collecting operational data was to meet DOE's requirement for monthly reporting on HPC availability and utilization metrics. It started with collecting environmental sensor data on the HPC floor. Soon they saw the potential of centralizing other requested IT-side operational data in the same repository, which was made possible through multiple upgrades year after year. NERSC management's philosophy is to "collect all the data" instead of "collect data to answer a specific question." This long-term vision ensures that the team's ability to gain operational insights won't be limited by data collection exclusions.

FIGURE 7 OMNI visualization showing maximum rack temperature on a record hot day (105°F) after installing blanking panels.



The technical solution for OMNI is tailored to the teams' specific needs, using an open-source distributed search and analytics software suite, which provides powerful visualization modules, on-premises hardware and virtualization technologies.[8] As more diverse data were accumulated, the team gained the ability to seek answers for more complex questions in the following three categories:

- **Real-Time:** Emergency Response. For example, during a recent arc flash event, the team quickly prioritized course of actions based on temperature trend data around equipment.[3]

- **Short-Term:** Review of Issues. For example, correcting control sequences during increasingly frequent wildfire air quality events.
- **Long-Term:** Design and Warranty Dispute Resolution. For example, the aforementioned avoided capital cost for a new substation. Equipment failure analysis for dispute negotiations with vendors has been another important value.

Shortly after the first HPC machine migrated from Oakland to Wang Hall, multiple cabinets unexpectedly powered off after finishing a large job. OMNI data was used to investigate the cause and revealed a major over-voltage event due to a quick power ramp down. This exposed the necessity of shunting retrofits to the HPC substation transformer's output voltage, ensuring a smoother power delivery to all HPCs.[3]

Data and visualization had also been critical to guide the team for actions under extreme weather conditions. In 2017, NERSC experienced a record hot day with 105°F (41°C) dry bulb and wet bulb of 74°F (23°C). As a result, the cooling towers could not bring cooling water supply temperature to the designed maximum of 74°F (23°C, was 4°F [2.2°C] higher[§§]). These conditions limited evaporative cooling strategies, and moving as much air as possible became the principal strategy.[##]

Figure 7 shows an OMNI heat map chart of maximum temperatures in air-cooled HPC systems. Three weeks prior to the record hot day, NERSC had installed blanking panels to improve hot-aisle containment. The OMNI data clearly shows that the maximum rack temperature had been reduced by 6°F to 8°F (3.3°C to 4.4°C) due to these isolation improvements, possibly providing a critical component to surviving the record outdoor temperatures three weeks later. (The rack temperature stayed below 84°F [29°C][‖‖]).

LBNL has adapted another strong data analytics and visualization tool for the campus. It has been expanded specifically for NERSC through real-time data connections to OMNI with the intention to facilitate rapid performance insights by correlating BMS data and HPC data. For example, it replaces the difficult process of time stamp synchronizations commonly associated with engineering analysis tools such as Microsoft Excel®. Full time stamp synchronization of all the data points enables the team to rapidly frame specific problems during discussions and perform solution analysis using real-time data plots. A cooling plant power vs. outdoor wet bulb panel is such an example—it provides scatter plots for both a base case and adjusted plant settings scenario, greatly facilitating team collaboration within the continuous improvement cycle described earlier.

Look Beyond PUE

The NERSC facility has an outstanding PUE of 1.08 average at present—a PUE of 1.25 or higher is common among HPC data centers. However, improvements are still possible, and the team is striving to be under 1.05 by implementing another five energy saving measures identified in the most recent energy audit. They recognize that for HPCs, the denominator (the compute load) is enormous and needs to be carefully managed for sustainability. The team has been exploring different scheduling software solutions to further increase utilization, and hence, IT energy efficiency. Testing, measuring and benchmarking IT efficiency is made possible with OMNI.

The next generation pre-exascale machines will bring more challenges in power supply, power fluctuation and other infrastructure aspects. The NERSC energy efficiency team will embrace these challenges with the power of extensive instrumentation and ever-improving management processes.

LBNL is strongly committed to continuously advancing its sustainability and energy efficiency. It has implemented a state-of-the-art energy and water management system (EWMS), which is certified to the ISO 50001:2018[10] standard by a third party and executes based on a "best-in-class" EWMS manual.[11] Meanwhile, LBNL has also achieved DOE's 50001 Ready program recognition.

The ISO standard promotes a holistic approach, emphasizes the critical role of top management's

[§§]The cooling tower leaving water temperature minus the ambient wet-bulb temperature (i.e., the cooling tower "approach") is 4°F to 5°F (2.2°C to 2.8°C) for NERSC.

[##]At higher intake temperatures, IT equipment fans will increase airflow, which can increase hot exhaust air recirculating back to the intake side unless AHU supply airflow is increased accordingly.

[‖‖]Although maximum rack temperature exceeded the ASHRAE recommended 81°F (27.2°C) limit, it is considered an allowable excursion and did not threaten the compute load.

commitment and adopts a "plan-do-check-act" continuous improvement cycle-based framework. Because of its top energy consumer status and expected growth, NERSC is the lab's "significant energy and water use" in the EWMS and is subject to more rigorous requirements on operational control, workforce competency, and procurement by the ISO standard. These requirements in turn challenge the team to continuously explore additional opportunities while ensuring harvested savings persist by monitoring any significant deviation in energy or water performance.

Some of the new challenges that the team is tackling today include working with the HPC vendor on improved control of the system's integral blower fans. The NERSC team has developed a control script for the operating system, which interties the HPC system's cooling coil valve controls with cooling plant water temperature controls, enabling the capability to shift the load ratio between the blower fans and AHU cooling coil under different environmental conditions. The cooling towers consume large volumes of fresh water every day, and the team is exploring alternative heat rejection solutions that require no water or much less water. In addition, the upcoming Perlmutter on-chip liquid cooling system opens new options and energy efficiency strategies to cool the next generation of high power density machines.

Key Lessons Learned and Outlook

The NERSC team at LBNL has demonstrated that, through interdisciplinary collaboration and an open mind to viewing risks and benefits, it is possible for an HPC facility to achieve extraordinarily low PUE and continuously improve and innovate leveraging data analytics. We hope the valuable lessons learned benefit other HPC facilities and other types of data centers.

- Backup power equipment such as UPSs and generators are capital-intensive investments, so consider limiting coverage to absolutely essential loads.
- "Compressor-free cooling" using cool outdoor air or cooling tower water (or both) is an effective way to lower a data center's PUE.
- A great deal of energy can be saved with control changes and low-cost measures (or "ongoing commissioning") before capital investment is necessary. Leveraging a team approach and data analytics will yield best results.
- Data analytics tools accessible by multidisciplinary teams is a powerful process for breaking down silos among business units. A normal data center infrastructure management (DCIM) system with a prioritized metering configuration could help data center teams make better operational and business decisions.
- Manage energy and water using a holistic approach with management's support rather than "project-by-project." A matrixed team representing all key functions should meet regularly and participate in making decisions.

Acknowledgment

The authors would like to acknowledge Rachel Shepherd in the Federal Energy Management Program (FEMP) Office at the U.S. Department of Energy for funding this case study. We would like to thank NERSC energy efficiency team members Steve Greenberg, John Eliot, Deirdre Carter and Jeff Broughton, who participated in interviews that provided rich information for this case study. Their help with reviewing earlier drafts is appreciated. We want to thank Natalie Bates at the Energy Efficient HPC Working Group (EE HPC WG) for leading these interviews and others who participated. We also appreciate Dale Sartor at LBNL for providing materials and encouragement for this case study.

References

1. NERSC. 2016. "NERSC Mission." National Energy Research Scientific Computing Center. https://tinyurl.com/y4nqstse/
2. The Green Grid. 2012. "White Paper #49, PUE: A Comprehensive Examination of the Metric." The Green Grid.
3. Bautista, E., et al. 2019. "Collecting, monitoring, and analyzing facility and systems data at the National Energy Research Scientific Computing Center." 48th International Conference on Parallel Processing: Workshops (ICPP 2019).
4. Bourassa, N., et al. 2019. "Operational data analytics: optimizing the National Energy Research Scientific Computing Center Cooling Systems." 48th International Conference on Parallel Processing: Workshops (ICPP 2019).
5. LBNL. 2020. "Center of Expertise for Energy Efficiency in Data Centers (CoE)." Lawrence Berkeley National Laboratory. https://datacenters.lbl.gov/
6. NERSC. 2020. "Systems." National Energy Research Scientific Computing Center. https://www.nersc.gov/systems/
7. NERSC. 2020. "Perlmutter." National Energy Research Scientific Computing Center. www.nersc.gov/systems/perlmutter
8. ASHRAE. 2015. *Thermal Guidelines for Data Processing Environments, 4th Edition*. Atlanta: ASHRAE.
9. ASHRAE. 2011. *Thermal Guidelines for Data Processing Environments, 3rd Edition*. Atlanta: ASHRAE.
10. ISO. 2018. "ISO 50001:2018, Energy Management Systems—Requirements with Guidance For Use." International Organization for Standardization.
11. LBNL. 2020. "Sustainable Berkeley Lab: Energy and Water Management System Manual." Lawrence Berkeley National Laboratory. https://iso50001.lbl.gov ■

https://bit.ly/3dTa21p
Rate this Article