US Department of Energy, Office of Science
High Performance Computing Facility Operational Assessment 2019
Oak Ridge Leadership Computing Facility

April 2020

**Approved for public release.
Distribution is unlimited.**

OAK RIDGE NATIONAL LABORATORY

Oak Ridge Leadership Computing Facility

US DEPARTMENT OF ENERGY, OFFICE OF SCIENCE HIGH PERFORMANCE COMPUTING FACILITY OPERATIONAL ASSESSMENT 2019 OAK RIDGE LEADERSHIP COMPUTING FACILITY

J. Paul Abston
Ryan M. Adamson
Scott Atchley
Ashley D. Barker
David E. Bernholdt
Katie L. Bethea
Kevin S. Bivens
Arthur S. Bland
Christopher Brumgard
Reuben Budiardja
Adam Carlyle
Christopher Fuson
Rachel Harken
Jason J. Hill
Judith C. Hill
Jonathan Hines
Katie Jones
Jason Kincl
Graham Lopez
Stefan Maerz
George Markomanolis
Don Maxwell

Veronica Melesse Vergara
Bronson Messer
Ross Miller
Jack Morrison
Sarp Oral
William Renaud
Andrea Schneibel
Mallikarjun Shankar
Woong Shin
Preston Shires
Hyogi Sim
Scott Simmerman
James Simmons
Suhas Somnath
Kevin G. Thach
Georgia Tourassi
Coury Turczyn
Sudharshan S. Vazhkudai
Feiyi Wang
Bing Xie
Christopher Zimmer

April 2020

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Executive Summary

HIGH PERFORMANCE COMPUTING FACILITY
2019 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
March 2020

EXECUTIVE SUMMARY

Oak Ridge National Laboratory's (ORNL's) Leadership Computing Facility (OLCF) continues to surpass its operational target goals: supporting users; delivering fast, reliable computational ecosystems; creating innovative solutions for high performance computing (HPC) needs; and managing risks, safety, and security associated with operating some of the most powerful computers in the world. The results can be seen in the cutting-edge science conducted by users and the praise from the research community.

Calendar year (CY) 2019 was a big year as OLCF staff ran five world-class resources (the leadership-class computers Titan and Summit, the large analysis cluster called Eos, and the massive parallel filesystems called Atlas and Alpine)) and also began power and cooling upgrades for a 2021 exascale system called Frontier. While continuing exceptional operation of Titan, Eos, and Rhea, the OLCF released the Summit supercomputer for production on January 1, 2019. Summit debuted as the most capable and efficient system in its class and has been recognized as the most powerful system in the world for its performance on both the high performance linpack (HPL) and conjugate gradient (HPCG) benchmark applications since June 2018 according to TOP500. Summit represents the culmination of a multiyear effort between the OLCF, IBM, NVIDIA, and Mellanox to deliver a system that is unmatched for modeling, simulation, data analysis, and learning. To hit the ground running with science-ready applications on day one, application teams worked closely with the OLCF through the Center for Accelerated Application Readiness (CAAR) program for years in advance of the Summit deployment.

CY 2019 was filled with outstanding results and accomplishments: a very high rating from users on overall satisfaction for the sixth year in a row; a tremendous amount of core-hours delivered to researchers from two leadership-class systems; and success in delivering on the allocation split of roughly 60%, 30%, and 10% of core-hours offered for the Innovative and Novel Computational Impact on Theory and Experiment (INCITE), Advanced Scientific Computing Research Leadership Computing Challenge (ALCC), and Director's Discretionary (DD) programs, respectively (see Operational Performance section). These accomplishments, coupled with the high utilization rates (overall and capability usage), represent the fulfillment of the promise of both leadership-class machines: efficient facilitation of leadership-class computational applications. Table ES.1 presents a summary of the 2019 OLCF metric targets and the associated results. More information can be found in the Operational Performance section for each OLCF resource.

The scientific accomplishments of OLCF users are a strong indication of long-term operational success, with publications this year in such notable journals and publications as *Nature, Nature Physics, Nature Plants, Physical Review X, Journal of the American Physical Society, Cell, Nano Letters*, and *Trends in Biotechnology*. Crucial domain-specific discoveries facilitated by resources at the OLCF are described in the *High Performance Computing Facility Operational Assessment 2019 Oak Ridge Leadership Computing Facility* (OAR) Strategic Results section. For example, researchers used Summit to pinpoint and understand the production of proteins from genetic information, including mutations and the functional expression of disease (Section 8.2).

Table ES.1. 2019 OLCF metric summary.

Metric description	CY 2019 target	CY 2019 actual
Overall OLCF score on the user survey will be 3.5/5.0 based on a statistically meaningful sample.	3.5	4.5
Time between Receipt of User Query (RT Ticket) and Center Response: 80% of OLCF problems will be addressed within 3 working days (72 hours) by either resolving the problem or informing the user how the problem will be resolved.	80%	90%
CAPABILITY JOBS: *For the calendar year following a new system/upgrade, at least 30% of the consumed node hours will be from jobs requesting 20% or more of the available nodes. In subsequent years, at least 35% of consumed core-hours/node-hours will be from jobs requiring 20% or more of cores/nodes available to the users.*		
Scientific and Technological Research and Innovation—Demonstrate Leadership Computing, SUMMIT	30%	47.6%
Scientific and Technological Research and Innovation—Demonstrate Leadership Computing, TITAN	35%	53.9%
SCHEDULED AVAILABILITY (COMPUTE): *For the calendar year following a new system/upgrade, the scheduled availability (SA) target for an HPC compute resource is at least 85%. For year 2, the SA target for an HPC compute resource increases to at least 90%, and for year 3 through the end of life for the associated compute resource, the SA target for an HPC compute resource increases to 95%. Consequently, SA targets are described as 85%/90%/95%.*		
Scheduled Availability, SUMMIT: Sustain scheduled availability to users, measured as a percentage of maximum possible scheduled.	85%	99.92%
Scheduled Availability, TITAN: Sustain scheduled availability to users, measured as a percentage of maximum possible scheduled.	95%	99.87%
OVERALL AVAILABILITY (COMPUTE): *For the calendar year following a new system/upgrade, the overall availability (OA) target for an HPC compute resource is at least 80%. For year 2, the OA target increases to at least 85%, and for year 3 through the end of life for the associated compute resource, the OA target increases to 90%. Consequently, OA targets are described as 80%/85%/90%.*		
Overall Availability, SUMMIT: Sustain availability to users, measured as a percentage of maximum possible.	80%	98.51%
Overall Availability, TITAN: Sustain availability to users, measured as a percentage of maximum possible.	90%	99.60%
OVERALL AVAILABILITY (FILESYSTEMS): *For the calendar year following a new system/upgrade, the OA target for an external file system is at least 85%. For year 2 through the end of life of the asset, the OA target for an external file system increases to at least 90%. OA targets are thus described as 85%/90%.*		
Overall Availability, External File System ALPINE: Sustain availability to users, measured as a percentage of maximum possible.	85%	98.18%
Overall Availability, External File System ATLAS: Sustain availability to users, measured as a percentage of maximum possible.	90%	Atlas 1: 99.91% Atlas 2: 99.90%
Overall Availability, Archive Storage: Sustain availability to users, measured as a percentage of maximum possible.	90%	HPSS: 97.97%

Table ES.1. 2019 OLCF metric summary (continued).

Metric description	CY 2019 target	CY 2019 actual
SCHEDULED AVAILABILITY (FILESYSTEMS): *For the calendar year following a new system/upgrade, the SA target for an external file system is at least 90%. For year 2 through the end of life of the asset, the SA target for an external file system increases to at least 95%. SA targets are thus described as 90%/95%.*		
Scheduled Availability, ALPINE: Sustain scheduled availability to users, measured as a percentage of maximum possible scheduled.	90%	99.53%
Scheduled Availability, ATLAS: Sustain scheduled availability to users, measured as a percentage of maximum possible scheduled.	95%	Atlas 1: 100% Atlas 2: 99.99%
Scheduled Availability, HPSS: Sustain scheduled availability to users, measured as a percentage of maximum possible scheduled.	95%	99.79%

The OLCF supported scientific accomplishments for a broad community of researchers in 2019, from traditional modeling and simulation projects to studies exploiting artificial intelligence, machine learning, and big data techniques. The first peer-reviewed science papers from Summit users were published early in the year. At least nine teams conducted research on Summit as part of applications for the prestigious Gordon Bell Prize, an award presented each year at the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC) to recognize researchers who have made significant strides in applying high-performance computing systems to scientific applications. Only two teams were named as finalists for the 2019 prize, both of which used Summit. A team from ETH Zürich won the 2019 Gordon Bell Prize for its work simulating quantum transport—or the transport of electric charge carriers through nanoscale materials. The team simulated a 10,000 atom system and reached a sustained performance of 85.45 petaflops for double precision and 90.89 petaflops for mixed precision. With the simulation, the team also demonstrated a new successful programming model that allows a programmer to interact with a visual representation of data movement to optimize a code.

OLCF systems continue to support data science, deep learning, and artificial intelligence (AI) work. For example, an ORNL team used their MENNDL code on both Titan and Summit to not only create optimal deep learning neural networks, but also to evolve network design on the fly. In 2019, the team developed neural networks that can quickly and accurately analyze biopsy slide images on a scale that microscope-equipped pathologists could never completely tackle. Their project promises to unlock new information on how tumors react to different treatments, and their work could profoundly affect the fight against cancer while also pioneering new ways of creating multi-objective neural networks. Similarly, the ORNL CANDLE, Exascale Deep Learning-Enabled Precision Medicine for Cancer team, is focused on deep learning for natural language processing, specifically information extraction from unstructured cancer pathology data to semi-automate reporting processes within the national cancer surveillance program. They are leveraging OLCF supercomputing resources to extract information from medical text data while helping drive the next generation of supercomputing. Additional examples of deep learning and AI-related work in 2019 include a team from Princeton Plasma Physics Laboratory and Princeton University that tested their Fusion Recurrent Neural Network (FRNN) code on Titan and Summit. Using neural networks, FRNN identifies patterns in plasma behavior to quickly and accurately predict disruptions in fusion reactors. Also, an ORNL team recently developed a new genomic selection algorithm driven by explainable AI and expanded to a global scale the climate and environmental information that can be used

in the Combinatorial Metrics, or CoMet, code. Finally, a team from Lawrence Berkeley National Laboratory, Pacific Northwest National Laboratory and NVIDIA used a deep learning application on Summit to model subsurface flow in the study of nuclear waste remediation. This work demonstrates the promise of physics-informed generative adversarial networks (GANs) for analyzing complex, large-scale science problems.

For more than half of 2019 the OLCF had over 50,000 GPUs and just shy of 300 petabytes of storage available for use between the Titan and Summit systems. In August 2019, Titan, Eos, and Spider II (Atlas) were all decommissioned in order to accommodate the preparatory work needed to install the Cray OLCF-5 system (Frontier) and the accompanying filesystem. It is significant that Titan was still one of the most powerful computational resources in the world 6 years after deployment, considering how rapidly technology advanced within that same timeframe. In its lifetime, Titan supported 896 unique projects and more than 1,000 publications. Through more than 26 billion core-hours of computing time, Titan served hundreds of research teams around the world working on today's most urgent scientific challenges. From its first official user projects in January 2013 through its final run in 2019, 2.86 million jobs were executed on Titan while OLCF staff ensured reliable operations.

The successful deployment and operation of a succession of leadership class resources is the result of the extraordinary work by the OLCF staff in supporting the most capable HPC user facility in the world. The OLCF staff are pivotal to identifying, developing, and deploying the innovative processes and technologies that support the advancement of science through the OLCF users and benefit other high performance computational facilities around the world.

ES.1 COMMUNICATIONS WITH KEY STAKEHOLDERS

ES.1.1 Communication with the Program Office

The OLCF communicates with the Advanced Scientific Computing Research (ASCR) Program Office through a series of regularly occurring events. These include weekly Integrated Project Team calls with the local DOE ORNL Site Office and the ASCR Program Office, monthly highlight reports, quarterly reports, the annual operational assessment report (OAR), an annual "Budget Deep Dive," an annual independent project review, and the OLCF annual report. Through a team of communications specialists and writers working with our users and management, the OLCF produces a steady flow of reports and highlights for sponsors, current and potential users, and the public.

ES.1.2 Communication with the User Community

The OLCF's communications with users are tailored to the objectives of relating science results to the larger community and helping users to more efficiently and effectively use OLCF systems. The OLCF offers many training and educational opportunities throughout the year for current facility users and the next generation of HPC users (Sections 1.4.5–1.4.7).

The impact of OLCF communications is assessed as part of an annual user survey. Communications was one of the highest rated areas of satisfaction on the survey. The mean rating for users' overall satisfaction with OLCF communications was 4.6. Ninety-five percent of respondents rated their overall satisfaction with communications from the OLCF as "satisfied" or "very satisfied." The OLCF uses a variety of methods to communicate with users, including the following.

- Weekly email message
- Welcome packet
- General email announcements

- Automated notifications of system outages
- OLCF website
- Monthly conference calls
- OLCF User Council and Executive Board meetings
- One-on-one interactions with liaisons and analysts
- Social networking
- Annual OLCF User Meeting
- Targeted training events (i.e., GPU Hackathons or tutorials)

Communication was important in 2019 due to the decommissioning of Titan, Eos, and the center-wide parallel filesystem (Atlas). To provide existing Titan and Eos users notice and time to prepare, it was critical to provide the decommission timeline early and often. Once the decommissioning date was established, the OLCF developed a schedule that would allow needed batch job execution through the decommissioning deadline, minimize the amount of batch work unable to complete prior to the deadline, and provide flexibility for the center to run priority work. To accomplish these goals, user job submissions were halted one month prior to system shutdown to permit the queues time to drain. User Assistance staff monitored the queues to identify any jobs that could not finish (e.g., due to unresolved dependencies) so they could notify the users of this and, with user approval, remove them from the queue so the scheduler could run as efficiently as possible. To provide additional notification, as the submission deadline became close, the batch system was updated to print a notification message following each batch submission. The messages provided a targeted reminder following each batch job submission.

Due to the significance of the changes, providing frequent notification through multiple communication methods was given high priority. In February, a page detailing the decommissioning schedule, including the dates on which jobs submission would be disabled, the date the Atlas filesystem would be made read-only, and ultimately, the decommissioning date of the system was added to the OLCF website. Reminders added to the OLCF weekly email included a concise reminder of the changes and pointed to the decommissioning web page for details. Announcements were made at the OLCF User Meeting in May, and center staff took advantage of other meetings/training events with users to remind them of the coming changes. The same information was sent to all users via email from the OLCF Director, and an announcement was made in multiple email communications to users as follows.

- February 25: Initial email about Titan, Eos, and Atlas decommissioning from OLCF Director

- February 27–June 26: Titan, Eos, and Atlas decommissioning notification in weekly emails (18 notifications)

- August 1: Reminder from OLCF Director about Atlas filesystem decommissioning

- July 3–August 14: Atlas decommissioning notification in weekly emails (seven notifications)

- August 13: Additional notification of pending Atlas decommissioning.

As important as the compute systems decommissioning notifications were, it was possibly even more important to ensure notification of the Atlas filesystem decommissioning because any data remaining on the filesystem would be lost upon decommissioning. To provide regular reminders, each system decommissioning message as well as the decommissioning web page also included notice of the filesystem decommissioning timeline. Because it was especially important that users with large amounts of data not postpone the data transition, in addition to the mass notification methods provided by emailing

all users and web site updates, the center also notified and worked with the top users of the filesystems individually.

By following procedures learned through many years of experience guiding users through change, the OLCF was able to effectively communicate with users through the decommissioning, provide time to transition, and help minimize the impact of the changes. In fact, the OLCF only received one ticket after the systems were removed from production that was related to the decommissioning. The fact that communication was one of the highest rated aspects on the 2019 survey in a year of significant change belies the importance the OLCF places on communications with users.

ES.2 SUMMARY OF 2019 METRICS

In consultation with the DOE program manager and as proposed in the 2018 OAR, a series of metrics and targets were identified to assess the operational performance of the OLCF in CY 2019. The 2019 metrics, target values, and actual results as of December 31, 2019, are noted throughout this report and are summarized in the Operational Performance section. The OLCF exceeded all agreed-upon metric targets.

ES.3 RESPONSES TO RECOMMENDATIONS FROM THE 2018 OPERATIONAL ASSESSMENT REVIEW

The OLCF received two recommendations from the 2018 OAR review. Those recommendations and the OLCF responses are provided in Appendix A.

HIGH PERFORMANCE COMPUTING FACILITY
2019 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
March 2020

1.   USER SUPPORT RESULTS

CHARGE QUESTION 1: Are the processes for supporting the users, resolving users' problems, and conducting outreach to the user population effective?

OLCF RESPONSE: Yes. The Oak Ridge Leadership Computing Facility (OLCF) supported 1,535 users and 438 user projects in calendar year (CY) 2019. The OLCF continued to leverage an established user support model for effectively supporting users based on continuous improvement, regular assessment, and a strong customer focus. One key element of internal assessment is the annual user survey. As part of the survey, users are asked to rate their overall satisfaction with the OLCF on a scale of 1 to 5, with a rating of 5 indicating "very satisfied." The mean rating for overall satisfaction with the OLCF in 2019 was 4.5. Overall ratings for the OLCF were positive; 94% of all survey respondents reported being "satisfied" or "very satisfied" with the OLCF.

The OLCF measures its performance using a series of quantifiable metrics. The metric targets are structured to ensure that users are provided prompt and effective support and that the organization responds quickly and effectively to improve its support process for any item that does not meet a minimum satisfactory score. The OLCF exceeded all metric targets for user satisfaction in 2019 with 90% of tickets being resolved within 3 business days. The OLCF continued to enhance its technical support, collaboration, training, outreach, and communication and engaged in activities that promoted high performance computing (HPC) to the next generation of researchers.

1.1   USER SUPPORT RESULTS SUMMARY

The OLCF's user support model comprises customer support interfaces, including user satisfaction surveys, formal problem-resolution mechanisms, user assistance analysts, and scientific and data liaisons; multiple channels for stakeholder communication, including the OLCF User Council; and training programs, user workshops, and tools to reach and train both current facility users and the next generation of computer and computational scientists. The success of these activities and identification of areas for development are tracked using the annual OLCF user survey.

To promote continual improvement at the OLCF, users are sent surveys soliciting their feedback regarding support services and their experience as users of the facility. The 2019 survey was launched on October 2, 2019 and remained open for participation through December 2, 2019. The survey was sent to 1,251 users of the Innovative and Novel Computational Impact on Theory and Experiment (INCITE), Advanced Scientific Computing Research (ASCR) Leadership Computing Challenge (ALCC), and Director's Discretionary (DD) projects, which includes the Exascale Computing Project (ECP) projects, and OLCF Early Science (ES) projects who logged into an OLCF system between January 1, 2019, and September 30, 2019. Even though the ECP allocations come from the OLCF DD program, we decided to track their responses separately from the DD responses. OLCF staff members were excluded from participation. A total of 578 users completed the survey, for an overall response rate of 46.2%. The results of the 2019 survey can be found on the OLCF website at https://www.olcf.ornl.gov/olcf-media/center-reports/2019-outreach-survey/.

The effectiveness of the processes for supporting customers, resolving problems, and conducting outreach are in part measured by the key survey responses for User Support in Table 1.1.

Table 1.1. 2019 Key survey responses.

Survey Question	2018 target	2018 actual	2019 target	2019 actual
Overall OLCF satisfaction score on the user survey	3.5/5.0	4.6/5.0	3.5/5.0	4.5/5.0
Overall Satisfaction with Support Received (User Assistance, Accounts, INCITE Liaisons, Advanced Data/Workflow Liaisons)	-	-	3.5/5.0	4.6/5.0
Overall Satisfaction with the Website	3.5/5.0	4.4/5.0	3.5/5.0	4.4/5.0
Overall Satisfaction with Communications	3.5/5.0	4.4/5.0	3.5/5.0	4.6/5.0
Overall Satisfaction with OLCF Support Services	3.5/5.0	4.5/5.0	3.5/5.0	4.5/5.0
Overall Satisfaction with Problem Resolution	3.5/5.0	4.6/5.0	3.5/5.0	4.5/5.0
Show improvement on results that scored below satisfactory in the previous period	Results will show improvement in at least half of questions that scored below satisfactory (3.5) in the previous period.	No question scored below satisfactory (3.5/5.0) on the 2018 survey.	Results will show improvement in at least half of questions that scored below satisfactory (3.5) in the previous period.	No question scored below satisfactory (3.5/5.0) on the 2019 survey.
Percentage of user problems addressed within 3 business days	80%	92%	80%	90%

1.2   USER SUPPORT METRICS

The OLCF exceeded all user support metrics for 2019. The OLCF metric targets and actual results by calendar year (CY) for user support are shown in Table 1.2.

Table 1.2. OLCF user support summary: Metric targets and calendar year results.

Survey Area	CY 2018 Target	CY 2018 Actual	CY 2019 Target	CY 2019 Actual
Overall OLCF satisfaction rating	3.5/5.0	4.6/5.0	3.5/5.0	4.5/5.0
Average of all user support services ratings	3.5/5.0	4.5/5.0	3.5/5.0	4.4/5.0

1.2.1   Overall Satisfaction Rating for the Facility

Users were asked to rate their satisfaction on a 5-point scale, where a score of 5 indicates a rating of "very satisfied," and a score of 1 indicates a rating of "very dissatisfied." The metrics were agreed on by the US Department of Energy (DOE) and OLCF program manager, who defined 3.5/5.0 as satisfactory. Overall ratings for the OLCF were positive, with 94% of users responding that they were satisfied or very satisfied with the OLCF overall.

Key indicators from the survey, including overall satisfaction, are shown in Table 1.3. They are summarized and presented by program respondents. The data show that satisfaction among all allocation programs is similar for the four key satisfaction indicators.

Table 1.3. Satisfaction rates by program type for key indicators.

Indicator	Mean	Program				
		INCITE	ALCC	DD	ECP	ES
Overall satisfaction with the OLCF	4.5/5.0	4.5/5.0	4.5/5.0	4.6/5.0	4.5/5.0	4.5/5.0
Overall satisfaction with support services	4.5/5.0	4.4/5.0	4.5/5.0	4.5/5.0	4.4/5.0	4.5/5.0
Overall satisfaction with compute resources	4.6/5.0	4.5/5.0	4.6/5.0	4.6/5.0	4.6/5.0	4.6/5.0
Overall satisfaction with data resources	4.5/5.0	4.5/5.0	4.4/5.0	4.4/5.0	4.5/5.0	4.4/5.0

1.2.2   Average Rating across All User Support Questions

The calculated mean of answers to the user support services specific questions on the 2019 survey was 4.4/5.0, indicating that the OLCF exceeded the 2019 user support metric target and that users have a high degree of satisfaction with user support services. Respondents described what they perceived to be "the best qualities of OLCF." Thematic analysis of user responses identified computing power/ performance and user tech support/staff as the most valued qualities of the OLCF. Included below are two open-ended responses to "What are the best qualities of the OLCF?"

"In my experience, the following are OLCF's best qualities: 1) OLCF's has the most powerful supercomputer in the world which is crucial for the large scale electronic-structure studies we are conducting, 2) OLCF's support team's quick response and quality of issue resolution is top notch, and finally 3) comprehensive user documentation which includes critical things like known issues. OLCF's graphics processing unit (GPU) Hackathon experience was also extremely useful for GPU porting of our code."

"OLCF has a history of providing leadership class systems with state-of-the-art hardware and software technologies, while also making the systems relatively easy to use. OLCF generally provides better job scheduling turnaround than other centers, and historically does a fantastic job with scheduling small debugging and software development-oriented jobs of the sort that I'm typically involved with. I often use OLCF as a positive example for comparison when giving advice to other computing centers about how they might improve their services."

In reviewing the ratings from the user support questions asked on the survey, the search capabilities on the OLCF website were the lowest rated area of user support, with an average rating of 4.1. We have since learned from users that when Google or the OLCF internal search capability are used, many of the results that are returned are related to OLCF communications rather than the technical support information. This can sometimes make the technical information more difficult to find on our website. We believe the recent work on our user support documentation, which is explained in more detail below,

will help with this issue. We also plan to promote how searching https://docs.olcf.ornl.gov provides better results than using Google or searching from within the OLCF website and will continue to look for other ways we can improve this functionality for users.

1.2.3   Improvement on Past Year Unsatisfactory Ratings

Each year the OLCF works to show improvement on no less than half of any questions that scored below satisfactory (3.5/5.0) in the previous year's survey. All questions scored above 3.5 on both the 2018 and 2019 surveys. Although we did not have any questions that scored below satisfactory on the 2018 survey, we did a thorough review of the survey and identified areas in which we could improve or add value to users. We will discuss in more detail below three such efforts in the areas of documentation, self-service, and queue changes that we undertook in 2019 in response to user feedback. None of these were implemented early enough in 2019 to determine their impact, but we believe all three initiatives will enhance the OLCF user experience.

1.2.4   Assessing the Effectiveness of the OLCF User Survey

1.2.4.1   Revisions to the 2019 OLCF User Survey

The survey was substantially revised in 2019 thorough a collaboration between Oak Ridge Associated Universities' (ORAU's) Assessment and Evaluation team and the OLCF. The previous surveys were lengthy, and each year several respondents provided survey comments to this effect. For context, the 2018 survey included 181 available survey items. Because each respondent is only shown certain questions, the maximum number of these items answered by any one respondent was 150, but the average respondent answered 90 survey items. The average completion time, as measured and reported by the survey software interface, was 18.5 minutes. The survey revisions were undertaken with the objective of reducing the burden on participants and focusing the survey on the key insights that would provide useful information to the OLCF. Ideally, the reduction in burden and redesign of the survey would also increase response rates and increase data quality. Several approaches were taken to achieve the revised survey. First, the ORAU team calculated inter-item correlations to identify pairs of items that were redundant. Those with a correlation factor greater than 0.8 were reviewed to determine whether the items covered redundant content. When possible, such items were combined. Second, the survey software was reconfigured to incorporate more "show/hides," questions that are only displayed to relevant respondents based on their early responses. This technique was already in place to some extent, but the survey was reviewed and rearranged to optimize show/hides and minimize the number of items shown to each respondent. For instance, a question was added to the start of the survey to ask users which services or OLCF capabilities they had used in the past year. These responses were used by the survey software to show only relevant questions on the subsequent pages. As another example, comments boxes for dissatisfaction were only displayed if the respondent had indicated dissatisfaction with one or more elements within a given section of the survey. As a result of the revision process, the total number of possible items in the survey was reduced by approximately 56%, from 181 to only 80. On average, respondents to the 2019 OLCF survey answered 38 items on the survey (compared to 90 items in 2018), and the maximum number of items shown to a user was 71. The average response time was 9.8 minutes, approximately half of the 18.5-minute response time in 2018.

1.2.4.2   Data Collection

The survey sampling frame was constituted by first collecting the names of individuals who had logged into an OLCF system between 1/1/2019 through 9/30/2019. OLCF staff and vendors were then removed from the list. Overall, this process resulted in a sampling frame with 1,251 OLCF users. The survey was hosted online beginning on October 2, 2019 and remained open for completion through December 2,

2019. A total of 578 users completed the survey, resulting in a response rate of 46.2%, which was up from 34.3% in 2018. The initial notification was sent out on October 2, and additional notices were sent by members of the OLCF Leadership team. The survey was also advertised on the OLCF website and in the weekly communications via email to all users. Survey responses were tracked daily to assess the effectiveness of the various communication methods. The number of responses increased after every targeted notification, but the results show other efforts, such as including the notice in the weekly communication, also contributed to the survey response rate.

The OLCF saw a 9% increase in responses from new users and an 8% decrease in responses from users who identified as using the OLCF for more than 2 years (Table 1.4).

Table 1.4. User survey participation.

	2018 survey	2019 survey
Total number of respondents (Total percentage responding to survey)	422 (34%)	578 (46%)
New users (OLCF user <1 year)	23%	32%
OLCF user 1–2 years	22%	21%
OLCF user >2 years	55%	47%

1.2.4.3 Statistical Analysis of the Results

The survey collected feedback about user needs, preferences, and experience with the OLCF and its support capabilities. Attitudes and opinions on the performance, availability, and possible improvements of OLCF resources and services were also solicited. ORAU provided the OLCF with a written report that included the results and a summary of the findings. The findings section presents results summarized numerically that report responded levels of satisfaction. This is followed by a verbal summary of the open-ended comments from individuals who indicated they were dissatisfied (via the scaled reply) with a resource or service (note: not all dissatisfied individuals supplied open-ended comments).

The survey assessed satisfaction with OLCF resources and services using a 5-point scale, ranging from "very dissatisfied" (1) to "very satisfied" (5). These responses were close ended and summarized by using frequency distributions, proportions, means, and standard deviations. The proportion of respondents indicating either a 4 (Satisfied) or 5 (Very satisfied) on an item was also typically reported as %Sat to provide a summary measure. This measure was also used to indicate the relative satisfaction with resources/services within categories. Respondents who were Very dissatisfied or Dissatisfied with OLCF resources/services were asked to provide comments explaining their dissatisfaction. To better understand the types of OLCF users and how needs and preferences varied, close-ended responses were frequently broken out by principal investigator (PI) status and by project allocation.

All open-ended responses were examined using categorical content analysis with complete thoughts in responses as the unit of analysis (note that percentages of response categories may add up to more than 100% when respondents provided multiple complete thoughts in a response). Complete thoughts (CTs) were sorted into categories for the purposes of counting, comparisons, and other forms of analysis. CTs were simply response text that could stand alone as a meaningful reply to survey questions. CTs were not limited to any specific grammatical unit and could vary from a single word, to a phrase, sentence fragment, or complete sentence.

Some response content categories were derived a priori from survey questions or OLCF website categories (e.g., Data Management). Other categories were developed inductively through an iterative process of grouping and regrouping similar content units (e.g., Containers or Training and Tutorials). Subcategories were elaborated as new relevant concepts or useful distinctions were identified and were then organized within major categories of closely related concepts. These are used to the extent possible, with variations as needed to accommodate differences in the focus of specific questions and year-to-year differences in users' specific and technical responses.

Table 1.5 displays responses for five of the overall satisfaction categories broken down by allocation program. As Table 1.5 illustrates, the metrics are very comparable across all five major allocation programs, and the variations are statistically insignificant.

Table 1.5. Statistical analysis of survey results.

Survey area	INCITE			ALCC			DD			ECP			ES		
	Mean	Variance	Std. Deviation	Mean	Variance	Std. Deviation	Mean	Variance	Std. Deviation	Mean	Variance	Std. Deviation	Mean	Variance	Std. Deviation
Overall satisfaction with the OLCF	4.5	.44	.66	4.5	.40	.63	4.6	.35	.59	4.5	.38	.62	4.5	.37	.61
Overall satisfaction with compute resources	4.5	.48	.69	4.6	.30	.55	4.6	.34	.58	4.6	.41	.64	4.6	.29	.54
Overall satisfaction with data resources	4.5	.36	.60	4.4	.37	.58	4.4	.44	.66	4.5	.36	.60	4.4	.25	.50
Overall satisfaction with user support	4.4	.45	.67	4.5	.38	.62	4.5	.42	.65	4.4	.40	.63	4.5	.49	.70

1.3 PROBLEM RESOLUTION METRICS

The following operational assessment review metrics were used for problem resolution.

- Average satisfaction ratings for questions on the user survey related to problem resolution are satisfactory or better.

- At least 80% of user problems are addressed (i.e., the problem is resolved, or the user is told how the problem will be handled) within 3 business days.

1.3.1 Problem Resolution Metric Summary

In most instances, the OLCF resolves reported problems directly, including identifying and executing the necessary corrective actions. Occasionally, the facility receives problem reports in which it is limited in its ability to resolve because of factors beyond the facility's control. In such a scenario, addressing the problem requires OLCF staff to identify and carry out all corrective actions at their disposal for the given situation. For example, if a user reports a suspected bug in a commercial product, prudent measures might be to recreate the issue; open a bug or ticket with the product vendor; provide the vendor the necessary information about the issue; provide a workaround to the user, if possible; and track the issue to

resolution with the product vendor, which may resolve the issue with a bug fix or workaround acknowledgment.

The OLCF uses Request Tracker software to track queries (i.e., tickets) and ensure response goals are met or exceeded. Users may submit queries via email, the online request form, or by phone. Email is the predominant source of query submittals. The software collates statistics on tickets issued, turnaround times, and other metrics to produce reports. These statistics allow OLCF staff to track patterns and address anomalous behaviors before they have an adverse effect on the work of other users. The OLCF issued 3,229 tickets in response to user queries for CY 2019. The center exceeded the problem-resolution metric and responded to 90% of the queries within 3 business days (Table 1.6).

Users were asked to provide satisfaction ratings for their overall satisfaction with OLCF's problem resolution and three specific aspects of problem resolution. 93% of respondents were overall either satisfied or very satisfied with problem resolution. The quality of OLCF response to reported issues was the highest rated specific aspect of problem resolution. Six respondents provided feedback on their dissatisfaction, and after review, those could be grouped into two major themes. In the first theme, the user felt like not enough information was given on the issue. In the second, users would like visibility into the OLCF ticketing system so they can see the status and history of their reported tickets. The OLCF is working to provide users and PIs with this capability in 2020 through the new MyOLCF application described in more detail below.

Table 1.6. Problem resolution metric summary.

Survey Area	CY 2018		CY 2019	
	Target	Actual	Target	Actual
Percentage of problems addressed in 3 business days	80%	92%	80%	90%
Average of problem resolution ratings	3.5/5.0	4.6/5.0	3.5/5.0	4.5/5.0

Tickets are categorized by the most common types. The top two reported categories in 2019 were National Center for Computational Sciences (NCCS) access and running jobs (Figure 1.1), which is very similar to the results from 2018.
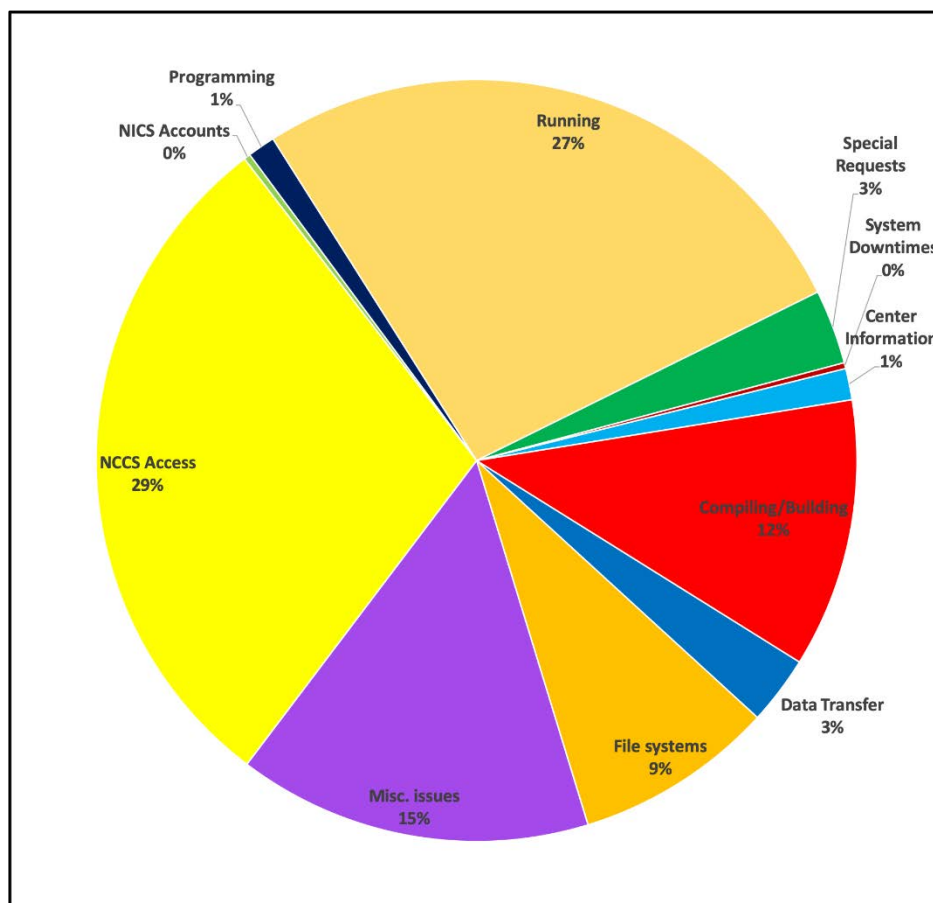
Figure 1.1. Categorization of help desk tickets.

1.4   USER SUPPORT AND OUTREACH

The Operational Assessment Report (OAR) data requested for user support and outreach includes examples of in-depth collaboration between facility staff and the user community and a summary of training and outreach events conducted during this period (Appendices B–C).

The following sections discuss key activities and contributions in the areas the OLCF recognizes as pillars of user support and outreach, including

- a user support staff made up of account management liaisons, User Assistance and Outreach (UAO) analysts, Scientific Computing Group (SciComp) liaisons, data liaisons, and visualization liaisons;

- multiple vehicles to communicate with users, sponsors, and vendors;

- developing and delivering training to current and future users; and

- strong outreach to interface with the next generation of HPC users, the external media, and the public.

1.4.1   User Support

The OLCF recognizes that users of HPC facilities have a wide range of needs requiring diverse solutions, from immediate, short-term, trouble ticket–oriented support, such as assistance with debugging and optimizing code, to more in-depth support requiring total immersion in and collaboration on projects. The facility provides complementary user support vehicles that include user assistance and outreach staff; liaisons in respective scientific, data, and visualization areas; and computer scientists who assist on issues surrounding the programming environments and tools. The following sections detail some of the high-level support activities during CY 2019 and the specific OLCF staff resources available to assist users.

1.4.2   User Assistance and Outreach

The UAO team addresses user queries; acts as user advocates; covers frontline ticket triage, resolution, and escalation; provides user communications; develops and delivers training and documentation; and installs third-party applications for use on the computational and data resources. The team also manages the OLCF Resource and Allocation Tracking System (RATS), which is the authoritative source for most of the system, user, and project data at the OLCF. Below are some examples of UAO initiatives in 2019 that helped improve the overall user experience, albeit some of them very much behind the scenes.

1.4.2.1   Documentation Enhancement

Web documentation is searchable and available 24 hours a day; it is one of the OLCF's largest training, instructional, and reference tools. Web documentation is an important tool for users of a new system and continues to serve as a reference for veteran users throughout the life of a system. The User Support group regularly checks the site to help ensure its content is up to date and easily navigable. In early 2019 in response to user feedback, the system user guide sections of the web documentation were reformatted and combined into a single document for each system. Among other things, this made it easier to search using standard browser tools (e.g., Ctrl-F). Later in 2019 as part of a continuing effort to improve the OLCF documentation, the OLCF began serving the system user guides, data transfer information, and account information from a new documentation system. The change, among other benefits, provides an improved method for direct user feedback and contributions. There are two main components of the new documentation system.

The first component involves using a version control system that stores the source text and images for the documentation pages. This is implemented through a public GitHub repository which allows for several features to our workflow including offline editing, concurrent updates, external contributions from collaborators and users, and revertible history. Only the source plaintext documents and accompanying images are stored in this git repository, along with a small amount of display configuration for the second component.

The second component is the web service that provides a rendered view of the most recent sources. An example of this available at https://docs.olcf.ornl.gov. This component is decoupled from the first component containing the sources, which allows for editing and staging with zero service interruption of the in-production documentation. The web service component resides as a container on a Kubernetes cluster that is located at and is administered by the OLCF. This component is completely automated and requires no intervention from documentation maintainers after its initial deployment. The general workflow for effecting documentation changes used by all contributors (including laboratory staff, vendor partners, and users) follows a typical git-based workflow used by open source software projects and, as a result, allows for outside participation with little to no training or setup needed from OLCF staff. Contributors clone the source repository, make edits and test/view rendered changes on their local machine, and then submit those proposed changes back to the source repository for review, revision, and

final integration by OLCF staff. After OLCF staff have accepted the changes, they are automatically propagated to the web service component and become visible on the website without any further intervention needed. Besides allowing for these distributed contributions, GitHub also provides a rich issue-based interface where both staff and users can request and discuss new content, point out errors, and propose improvements.

As a result of these changes allowing easier access and collaborative improvements, our vendor partners (both IBM and NVIDIA) have contributed additions and changes to our documentation regarding their products. Also, we have had suggestions and textual improvements to the documentation submitted by users who are not OLCF staff.

1.4.2.2   MyOLCF 2.0

Software developers within the user assistance group created a new web application in CY 2019 that is expected to enter full production for end users in 2020. This new self-service application, named MyOLCF, has two major goals: to give principal investigators more tools to make informed decisions about their OLCF projects and to give principal investigators more control over the administrative aspects of their projects.

MyOLCF makes information about a project's end users, resource allocations, and allocation usage history available to principal investigators and project members. Metrics are available to query and make use of same backend reporting application programming interfaces (APIs) that OLCF staff and management use internally for reporting, ensuring that both project PIs and OLCF staff share a common, reliable data source for making decisions about projects. MyOLCF also introduces several tools for streamlining the administrative work inherent in owning a computational project. For example, PIs can approve or reject end-user applications to their projects directly within MyOLCF, removing the need to contact the OLCF accounts team. Applicants can also check the status of their user applications with just their application's confirmation ID.

Now that the fundamental platform is built, the OLCF will continue to add new features and functionality to MyOLCF to enhance the users' experience at the facility.

1.4.2.3   Summit Queue Adjustments (Killable and Experimental)

To better support emerging workloads requiring longer execution times at smaller node counts, the OLCF introduced the `killable` queue on Summit. The queue is designed to allow jobs that use fewer than 91 nodes to run for up to 24 hours instead of the standard execution times permitted at that job size on the `batch` queue. Jobs submitted to the new `killable` queue can be preempted by jobs in the `batch` queue once they have run for a certain duration. For example, a job requesting fewer than 45 nodes is only allowed to run for up to 2 hours in the `batch` queue. In the `killable` queue, that job can request up to 24 hours of execution time and can only be preempted once it has run for at least 2 hours. After that point, the job can be preempted by higher priority capability jobs in the `batch` queue. If a job is preempted, it is automatically requeued by the scheduler and will restart as soon as resources become available. If no higher priority capability jobs are in the queue, the job in `killable` will continue to run until it reaches its full execution time.

For some use cases, however, checkpointing and restarting a job is not a viable alternative; for these, the OLCF will introduce a new queue called `experimental` which will have a fixed size and will allow users to submit jobs with an execution time of up to 24 hours. To access the queue, pilot users will have to submit their use case and receive approval from the Resource Utilization Council.

1.4.2.4 RATS CRM

The center's customer relationship management software, called RATS CRM, is under continuous development. In 2019, twenty-eight versions of the software were deployed, adding many new features and improvements.

- Accommodations for new allocation programs were added, including support for the new OLCF Quantum Computing User Program.

- All documents and approvals required by the OLCF for account and project applications are now organized within RATS, streamlining the administrative vetting process.

- New mechanisms for storing UNIX User authentication data were built into RATS, allowing internal staff and end users with vision impairments to authenticate with hardware authentication devices (YubiKey, PIV cards, etc.) instead of RSA SecurID fobs.

- Automatic penalization of over-allocation projects was implemented via batch scheduler controls. Prioritization penalties are automatically put in place based on the previous day's cumulative batch usage metrics.

- New API features were added for modernizing high-performance storage system (HPSS) account and directory management.

1.4.2.5 Metrics Reporting

The software development team within the user assistance group also maintains RATS-Report, the central data warehouse and reporting interface for OLCF. RATS-Report handles the ingest, aggregation, and correlation of all metrics related to compute resource utilization, filesystem utilization, project allocation usage, resource availability, etc. In 2019, the team released a complete overhaul of the user interface for RATS-Report, providing a custom solution to the reporting needs of the OLCF. The new interface includes features such as

- Easy-to-find charts of resource utilization at any level (resource, allocation program, project, user) with flexible filtering and sorting capabilities,

- searchable job listings including fine-grained GPU usage information, batch script details, and node maps, and

- an API for retrieving usage data and job data.

The team averaged two or three deployments a month, adding features throughout 2019, with 32 deployments in all.

1.4.2.6 Slurm Batch Scheduler and Job Launcher Transition

The batch scheduler and job launcher are used as a pair to access and use a system's compute resources. All users of a system must use the batch scheduler to allocate resources and then the job launcher to access the allocated resources. Both play a vital role in the use of an HPC system. Users must understand each in order to effectively and efficiently use a system's resources. Because both are often tied to a user's workflow, changes can require users to modify their workflow.

Since being accepted into production, both Rhea and the Data Transfer Nodes (DTN) have used MOAB as the batch scheduler and OpenMPI's mpirun as the job launcher. On September 03, 2019, the production batch scheduler and job launcher on Rhea and the DTNs were changed to Slurm. Due to the impact of this change, the OLCF User Assistance (UA) team worked to provide a smooth transition.

To provide notice of the change and allow users time to modify workflows, the OLCF UA team began notifying users of the pending change months prior to the change date. Following procedures learned through previous years of guiding users through change, email was used to send concise notification of the change while the OLCF web site provided details, comparisons, and timelines. Additional messages were printed at batch job submit time to provide targeted notification while the batch system was being accessed by a user. The transition was also the discussion topic in a monthly user call open to all OLCF users. The call provided an additional venue for notification with the opportunity to ask OLCF staff questions about the change.

A transition period was created by running both schedulers at the same time for a limited window of time. Running both schedulers simultaneously provided users with the opportunity to modify and test workflows under Slurm while still running production workflows under MOAB. The transition period also provided users with the opportunity to test and work through any found issues with the OLCF UA team while production workflows continued to run under MOAB. The transition period helped users move to the new scheduler with limited impact on production workflows.

In addition to user-facing aspects of the transition, the OLCF's allocation and access control tools also work closely with the scheduler to control access, queue throughput, and report system utilization. Changes to the scheduler and job launcher also change the format of job records, processes used to interact with the scheduler, and policy control processes. Prior to entering the user transition period, updates to the center's underlying processes and tools used to interact with the scheduler were made and tested.

Examples of the impact of scientific liaison collaboration with and support of users are provided in Sections 1.4.3.1, 1.4.3.2, and 1.4.3.3.

1.4.3   Scientific Liaison Collaborations

The following sections highlight specific collaborative areas where OLCF staff scientists partnered with INCITE and other research teams to maximize their productivity on the provided leadership-class resources.

1.4.3.1   Leveraging the Computational Power of Summit to Demonstrate Quantum Supremacy

Demonstrating superior efficiency of quantum computers in solving certain hard computational problems, known as Quantum Supremacy, was an important milestone on the quantum hardware development roadmap in order to show the promise of real practical use in the near future. In October 2019, as a result of a collaboration between Google, NASA, and ORNL/OLCF, a paper was published in *Nature* that successfully demonstrated the solution of a well-defined computational task of sampling random quantum circuits by the Google Sycamore quantum computer orders of magnitude faster than any available classical simulation algorithm on a supercomputer. In order to prove the point, it was critical to achieve the highest possible performance of the classical simulation algorithm on Summit, the world's fastest supercomputer at the time.

OLCF computational scientist Dmitry Lyakh (Liakh) was instrumental in porting the classical simulation code qFlex, originally developed by NASA, to Summit in order to leverage its enormous computational

power delivered by NVIDIA Volta GPU. The tensor algebra library TAL-SH, developed by Lyakh during the OLCF-4 (Summit) Center for Accelerated Application Readiness (CAAR) program, was integrated as a computational backend of the qFlex code, delivering unprecedented 281 petaflops of sustained single-precision performance (and even much higher peak), thus setting a new bar for the highest achieved quantum circuit simulation speed.

1.4.3.2 Performance Portable Nuclear Tensor Contraction Library Enables Work to Solve Discrepancy in Nuclear Physics

A recent *Nature Physics* paper resulting from the INCITE project "Ab-initio Nuclear Structure and Nuclear Reactions" led by Gaute Hagen (ORNL) was published and solved a long-standing discrepancy between theory and experiment in nuclear physics (see also Strategic Results in Section 8). This accomplishment is a result of simultaneous advances in nuclear theory, many-body methods and HPC, and the calculations that led to this accomplishment could only have been performed on Summit using the code NuCCOR (Nuclear Coupled-Cluster Oak Ridge). NuCCOR is a quantum many-body code developed at ORNL since the late 1990s and is tailored for nuclear systems. It uses the Nuclear Tensor Contraction Library (NTCL), developed by co-author and OLCF computational scientist Gustav R. Jansen and more recently OLCF postdoctoral research associate Justin Lietz, to leverage the resources on Summit. NTCL is a domain-specific open-source library that is designed to be both performant and portable. This was achieved by splitting the library into a small hardware-dependent back-end and a hardware-independent front-end, where only the back-end must be ported to a new architecture.

NTCL was started by Jansen in 2013 as part of the NUCLEI SciDAC-3 project and ported to CUDA as part of the OLCF-4 Summit CAAR project. Based on the experiences from these projects, the library was completely redesigned in 2017/2018 and the scope of the library was also expanded. In 2017 under the NUCLEI SciDAC-4 project, NTCL was expanded to include higher order tensors and provide more precise many-body methods for larger and larger nuclei. In addition, new communication-avoiding algorithms together with one-sided communication patterns nearly eliminated the communication bottleneck that was a remnant from the original code. As part of the Frontier CAAR project, NTCL was ported to HIP/ROCm for Advanced Micro Devices (AMD) GPUs. NTCL will be made publicly available in 2020 pending a code review.

1.4.3.3 Collaboration between OLCF Liaison and OLCF Operations Enabled Highest Resolution Turbulence Simulation

The INCITE project "High-Resolution Study of Intermittency in Turbulence and Turbulent Mixing" aims to study extreme events in turbulence at high Reynold numbers using pseudo-spectral direct numerical simulations (DNS). Led by Professor P.K. Yeung from the Georgia Institute of Technology, the project team performed simulations using up to 18432^3 grid points, exceeding all prior state-of-the-art resolution in the turbulent community worldwide. To solve this very large problem size, the team leverages architectural capability uniquely provided by Summit supercomputer at the OLCF.

At the heart of the pseudo-spectral solution methods used in this project is the three-dimensional Fast Fourier Transform (FFT). The code exploits the large numbers of GPUs available on Summit to accelerate this computation, along with asynchronous data movement within the node to fit a larger data size than is available when using GPU memory alone. The algorithm is also communication intensive and therefore prone to runtime variability.

During their simulation campaign, the project team worked closely with their OLCF liaison, computational scientist Reuben Budiardja, to troubleshoot and resolve problems that only appear at scale (i.e., jobs using more than 60% of Summit). Budiardja tuned the message passing interface (MPI) library

to reduce the runtime variability for the code, assisted the team in identifying a memory usage issue with a vendor-provided library which resulted in a bug report to the vendor, and constructed a more robust job script to handle node failure. In one instance, together with an OLCF HPC systems administrator, Budiardja determined that a recent upgrade on one of the center filesystems had an unintended side-effect that caused the code to fail its startup. The close collaboration with and effective support of Budiardja directly contributed to the team's successful completion of their scientific objectives within the INCITE project.

1.4.4 OLCF User Group and Executive Board

All PIs and users on approved OLCF projects are members of the OLCF User Group (OUG) and remain so for 3 years following the conclusion of their OLCF project. The OUG meets once a month via a webinar, the OLCF User Conference Call. These webinars are attended by representatives of several groups within OLCF and provide users with the opportunity to interact with OLCF staff. During the webinars, OLCF staff also provide updated information on upcoming system events and training events. The webinars also include a brief tutorial. These tutorials often focus on newly available software/tools, user experience talks, or other training of use to a variety of users. Ten OLCF User Conference Calls were held in 2019 with a total of 613 attendees (up from 447 in 2018).

The OUG is represented by a 9–10 member OUG Executive Board. This board typically meets every other month just before or just after the User Conference Call for an in-depth discussion with OLCF staff to provide feedback and guidance on training topics and the facility's resources and policies. OUG Executive Board terms are 3 years and are staggered such that three new members are elected each year to replace three outgoing members. (An outgoing chair will remain on the board as an ex officio member for an additional year if he/she is at the end of his/her 3 year term, hence the possibility of a 10 member board). The current board chair is Mike Zingale. The board elected Sarat Sreepathi as the vice chair, and he will become chair for the 2020–2021 board term. During the 2019 User Meeting, users elected Balint Joo, P.K. Yeung, and Eric Nielsen to 3 year terms concluding at the 2022 User Meeting.

Information about the OLCF User Group and the Executive board is available at https://www.olcf.ornl.gov/about-olcf/oug/.

The OLCF hosted its annual User Meeting on May 21–23. The meeting provided users with a forum in which to highlight scientific accomplishments on OLCF resources, an opportunity for OLCF staff to update users on facility status and initiatives, and, most importantly, an opportunity for face-to-face interaction between users and staff. Talks were grouped into common themes. Day 1 began with presentations from two Gordon Bell finalists as Mr. Prabhat (NERSC) discussed "Exascale Deep Learning" and Wayne Joubert (OLCF) gave a talk titled "CoMet: Enabling Comparative Genomics at Exascale." Day 1 continued with three Early Summit Results talks. Included in this session were talks from Dmitry Liakh (OLCF), Salman Habib (ANL), Jeroen Tromp (Princeton University), Remco Havenith (University of Groningen), and Patrick Steinbrecher (BNL). Day 2's themes included Federated Facilities and OLCF Facility Updates. Shantenu Jha (BNL and Rutgers University) discussed RADICAL-Cybertools, ORNL's Scott Klasky gave a talk titled "Exascale IO and in situ Data Processing on Leadership Computers," and OLCF's Jason Kincl discussed using OpenShift to facilitate workflows at OLCF. Three invited talks presented on Day 3 were "Linear Algebra Software Technologies for Exascale" from Stan Tomov (University of Tennessee), "Scientific Computing with Quantum Computers" from ORNL's Travis Humble, and an overview of Neuromorphic Computing by ORNL's Catherine Schuman.

A total of 132 users and staff attended the meeting onsite. The meeting was also broadcast via BlueJeans for users who could not attend in person. There were eight remote attendees via BlueJeans. Slides and

videos from the different presentations were uploaded and made available on the User Meeting's event page at https://www.olcf.ornl.gov/calendar/2019-olcf-user-meeting/.

1.4.5 Training, Education, and Workshop

The OLCF training program serves to educate our user community in general HPC training as well as special topics needed to fully leverage the facility's cutting-edge HPC resources. To do so, the OLCF offers training in the form of workshops, webinars, tutorials, seminars, and hackathons. In most cases, these training events are recorded so the slides and recordings can be made available to our users (and broader community) after the event as well.

In 2019, the OLCF facilitated, or collaborated on, a total of 10 week-long GPU Hackathons, a 3 day Summit Training Workshop, a 5 day Petascale Computing Institute, a 5 day OLCF/ECP OpenMP Hackathon, two Intro to AMD GPU Programming with HIP Webinars, a 3 day OLCF User Meeting, an Introduction to Summit Webinar, an Introduction to NVIDIA Profilers on Summit tutorial, a Linux Command Line Productivity Tools, a 3 day Profiling Tools Workshop, 10 OLCF User Conference Calls/Webinars, 11 Interoperable Design of Extreme-Scale Application Software – ECP (IDEAS-ECP) webinars, and two INCITE Proposal Writing Webinars. See Appendix B for a complete summary of these events. Some notable events are highlighted in Sections 1.4.5.1–1.4.5.4.

1.4.5.1 Summit Training Workshop

On February 11–13, the OLCF delivered a 3-day Summit Training Workshop to help users get up and running on the system. Each day consisted of presentations in the morning followed by hands-on sessions in the afternoons. The presentations delivered relevant information about the system (hardware, job launcher, programming methods, etc.) as well as talks on porting experiences from early Summit users. Staff from the OLCF, IBM, and NVIDIA helped participants get their codes running on Summit during the hands-on sessions.

The target audience for this event were new INCITE projects who recently obtained allocations on Summit, although attendance was available for anyone interested in running on the system. For participants without Summit allocations, the 18 node Ascent training system was made available during the event. The presentations were broadcast via webcast for remote participation, and the slides/recordings were posted on the OLCF Training Archive for future reference. In total, 131 participants attended the event (https://www.olcf.ornl.gov/calendar/summit-training-workshop-february-2019/).

1.4.5.2 2019 Petascale Computing Institute

On August 19–23, the OLCF co-delivered the 2019 Petascale Computing Institute to teach participants to scale their computational codes to leadership-class computing systems. The institute was co-organized by Argonne National Laboratory (ANL), the Blue Waters project at the National Center for Supercomputing Applications (NCSA), the National Energy Research Scientific Computing Center (NERSC), Oak Ridge Leadership Computing Facility (OLCF), Pittsburgh Supercomputing Center (PSC), SciNet at the University of Toronto, and the Texas Advanced Computing Center (TACC).

These organizations, along with many other host sites, held individual branches of the institute, where attendees could participate with help from local experts. The presentations (delivered from multiple host sites) were broadcast to all branches of the event and included topics ranging from traditional HPC material (e.g., MPI, OpenMP, debugging, profiling) to more modern material (e.g., CUDA, OpenACC, Python in HPC, containers). Each topic was accompanied by a hands-on session in which participants

worked through example problems to reinforce the presented material. The presentations were recorded and posted on YouTube for later viewing. Nearly 400 participants attended the event either in person at the host sites or remotely via webcast (https://bluewaters.ncsa.illinois.edu/petascale-computing-2019).

1.4.5.3   Frontier Application Readiness Kick-Off Workshop

On October 8–10, the OLCF delivered a Frontier Application Readiness Kick-Off Workshop for members of the Center for Accelerated Application Readiness (CAAR) teams and selected ECP teams. During the workshop, staff from the Frontier Center of Excellence (CoE; OLCF, Cray, AMD) presented details about the upcoming Frontier system (architecture, software environment, programming models) as well as how to obtain access to early Cray+AMD hardware. The teams also had a significant amount of unstructured time to make connections with CoE staff for continued collaboration as they prepare their applications to run on Frontier. More than 160 participants attended the event (https://www.olcf.ornl.gov/frontier-application-readiness-kick-off-workshop/).

1.4.5.4   2019 OLCF GPU Hackathon Series

In 2019, the OLCF and partner organizations continued to expand the annual OLCF GPU Hackathon series. These hackathons are 5-day coding events intended to teach new GPU programmers how to leverage accelerated computing in their own applications and to help existing GPU programmers to further optimize their codes. The application teams typically consist of three or more developers of an application, along with one or two mentors with extensive GPU programming experience. Together, the teams work to port or optimize their applications on GPU-accelerated systems, from small local clusters and workstations to world-class HPC systems such as Summit and Piz Daint.

These hackathons allow the OLCF to take their training program "on the road" so to speak, which is desirable given the OLCF user base is spread across the country (and world). This way, OLCF projects with GPU development goals can choose a hackathon that is most convenient (locally and temporally) for their team and get a dedicated 5-day push toward their goals—while surrounded by experts in the field. The mentors come from universities, national laboratories, supercomputing centers, and industry partners.

In total, 11 such hackathons were held at the following host sites.

- Korea Institute of Science and Technology Information (KISTI) – Seoul, South Korea
- Pawsey Supercomputing Centre – Perth, Western Australia
- Juelich Supercomputing Centre (JSC) – Juelich, Germany
- Joint Genome Institute (JGI) – Walnut Creek, CA
- Massachusetts Institute of Technology (MIT) – Cambridge, MA
- Princeton University – Princeton, NJ
- National Energy Research Scientific Computing Center (NERSC) – Oakland, CA
- University of Sheffield – Sheffield, United Kingdom
- Brookhaven National Laboratory (BNL) – Upton, NY
- CSCS Swiss National Supercomputing Centre – Lugano, Switzerland
- Oak Ridge Leadership Computing Facility (OLCF) – Knoxville, TN

These sites hosted 89 application teams, bringing dedicated GPU-programming expertise to more than 500 developers from 105 unique universities, national laboratories, supercomputing centers, government institutions, and industry. During the events, teams developed applications spanning a wide range of scientific domains, including (but not limited to) astrophysics, climate modeling, combustion,

computational fluid dynamics, machine learning, molecular dynamics, plasma physics, and quantum mechanics.

One noteworthy hackathon success story comes from a General Electric (GE) team that attended a 2018 hackathon held at BNL and a 2019 event at MIT. During the 2018 event, the team had success porting several kernels of their GENESIS code to GPUs, obtaining large speedups in those kernels but not in the overall application (which is typical for large applications). Afterward, the GE developers used the knowledge they gained at the BNL hackathon to port the rest of their code to GPUs over the following months, eventually achieving large speedups in the full application. At the 2019 event, the team worked to further optimize their code, resulting in a 50–300× overall speedup of the GENESIS code (depending on problem size). GENESIS is a combustion code used to simulate conditions in turbomachinery, which helps to inform GE on product design. "Using the GPUs on Summit, we experienced speedups in our scaling studies ranging from 50- to 300-fold," Velez said. "That's a month-long simulation done in just over 2 hours. That, for us, is a game changer." For more information on this story, please see the following writeup on the OLCF website: https://www.olcf.ornl.gov/2019/09/12/gpus-power-ge-code-at-olcf-hackathons/.

In addition to providing application teams with development support, these events benefit the OLCF by introducing participants to our facility and resources, nurturing relationships with our industry partners and leaders of the host organizations, generating articles in peer-reviewed journals, recruiting new OLCF staff members, and producing new OLCF users and projects. A noteworthy example of the last case comes from a University of Michigan team that attended GPU hackathons in 2018 and 2019. The team began porting their Density Functional Theory (DFT) code to GPUs during the 2018 GPU hackathon held in Knoxville, TN. After the event, the team requested—and was awarded—a director's discretionary (DD) allocation on Summit in order to continue their porting efforts. Using this allocation, and the knowledge they gained at the hackathon, the developers ported the rest of their code to Summit's GPUs over the next 6 months, which ultimately earned them a nomination as a Gordon Bell finalist. The team has since been awarded a full 2020 INCITE allocation on Summit to study the core energetics of dislocations in aluminum and magnesium (predictive dislocation-based theory of crystalline plasticity) using their DFT-FE code. More information about this story can be found in the following writeup on the OLCF website: https://www.olcf.ornl.gov/2019/12/16/record-breaking-year-for-olcf-gpu-hackathons-ends-with-annual-knoxville-tennessee-event/.

1.4.5.5   Profiling Tools Workshop

On August 7–9, the OLCF delivered a 3-day Profiling Tools Workshop to help new and experienced users analyze the performance of their code on Summit. Each day consisted of presentations in the morning followed by hands-on sessions in the afternoons. The presentations were delivered by experienced users using real applications. The purpose was to demonstrate which tools work without issues, what functionalities are supported, and to present the lessons learned from this procedure. A call for applications was issued in advanced of the workshop so attendees would bring their applications for the hands-on session. Seven totally different profiling tools were presented—TAU, Scalasca, Score-P, Extrae/Paraver, Nvprof, ARM, and Scalasca—with step-by-step instructions on how to analyze the performance of parallel applications and visualize the performance data. Applications were also provided for the users who did not bring their own applications. The target audience for this event were Summit users from ORNL and other organizations, although anyone interested in running on the system could attend. For participants without Summit allocations, the 18-node Ascent training system was made available during the event. The presentations were broadcast via webcast for remote participation, and the slides/recordings were posted on the OLCF Training Archive for future reference. In total, 50 participants registered, 31 of whom were on site and 19 participated remotely (https://www.olcf.ornl.gov/calendar/profiling-tools-workshop/).

1.4.6 Training and Outreach Activities for Future Members of the HPC Community and the General Public

1.4.6.1 SC19 Student Enrichment Event—Hands-On with the Summit Supercomputer

The annual International Conference for High Performance Computing, Networking, Storage, and Analysis (Supercomputing 2019; SC19) is among the largest venues to introduce students and early-career researchers to leadership-scale HPC. As part of the conference's Students@SC program, OLCF staff led a half-day session titled "Hands-On with the Summit Supercomputer." This event included an overview of the Leadership Computing Facilities, introductory presentations on using the Summit computing and programming environments, and an opportunity for attendees to engage OLCF staff about internship and employment opportunities. Attendees were given temporary access to Summit to apply what they learned in hands-on examples and challenge problems. Hands-on exercises covered a wide range of subject matter and degrees of difficulty to accommodate the attendees' various levels of experience. The session was attended by 29 student and non-student conference attendees. More information about this event can be found at the following webpages: https://sc19.supercomputing.org/presentation/?id=pec111&sess=sess412, https://github.com/olcf/SC19_HandsOn_with_Summit.

1.4.6.2 SIAM Computational Science and Engineering (CSE) Conference

OLCF staff participated in the SIAM Computational Science and Engineering (CSE) Conference's Broader Engagement track and delivered a 100-minute tutorial on GPU computing on Summit on February 27. This conference "seeks to enable in-depth technical discussions on a wide variety of major computational efforts on large-scale problems in science and engineering, foster the interdisciplinary culture required to meet these large-scale challenges, and promote the training of the next generation of computational scientists." (https://www.siam.org/conferences/cm/conference/cse19). The Broader Engagement Summit training was filled to capacity by undergraduate and graduate students. The tutorial covered an introduction to the OLCF and the Summit compute architecture, followed by a general introduction into GPU programming using a common Jacobi-type example. The second half of the tutorial covered multiple ways of achieving multi-GPU implementations, which is necessary due to the multi-GPU characteristics of the Summit node architecture.

1.4.6.3 ACM Richard Tapia Celebration of Diversity in Computing Conference (Tapia 2019)

OLCF staff participated in the DOE National Laboratories Workshop at the 2019 ACM Richard Tapia Celebration of Diversity in Computing Conference (Tapia 2019) held on September 18.The half-day event included presentations from ANL, NERSC, OLCF, Los Alamos National Laboratory (LANL), and Lawrence Livermore National Laboratory (LLNL). The workshop introduced attendees to opportunities available at the different facilities as well as potential career paths within the national laboratory complex. In addition to presentations, the workshop included a series of tutorials. The OLCF facilitated two tutorials at the event, "Introduction to SIGHT" and "GPU Accelerated Computing Using OpenACC," and provided attendees with access to Summit for the hands-on portion of the workshop. The SIGHT (https://www.olcf.ornl.gov/olcf-resources/rd-project/sight/) tutorial introduced attendees to this powerful visualization tool for large-scale data sets. The OpenACC tutorial introduced attendees to GPU accelerated computing using hands-on exercises that allowed them to parallelize their code to take advantage of Summit's GPUs. More information about this event can be found on the Tapia 2019 conference program: http://tapiaconference.org/assets/5e31a99cc52e8364db426e24/TAPIA_2019_program_v18.pdf

1.4.6.4   Outreach

The OLCF Outreach team works to engage new and next-generation users and showcases OLCF research through strategic communication activities such as highlights, fact sheets, posters, snapshots, the OLCF website, and center publications (see Appendix C). For the Outreach team, 2019 was focused on key transition milestones, from Summit's transition to full operations, to the public announcement of the Frontier contract, to the decommissioning of Titan.

In January 2019, the first official users—the INCITE projects—began running on Summit. The Outreach team worked to capture science highlights and technical accomplishments from Summit users and operations, and in May the team published a science highlight based on the first official Summit results to be published in a peer-reviewed journal. The highlight, "Summit Charts a Course to Uncover the Origins of Genetic Disease," came from users at Georgia State University. In November, the team published a feature on the 2019 Gordon Bell Prize–winning work from an ETH Zürich team that used Summit to simulate quantum transport—or the transport of electric charge carriers through nanoscale materials.

The contract for the OLCF-5 project, or Frontier, was awarded in early May. The OLCF Outreach team worked closely with Cray and the ORNL media and protocol offices to prepare a strategy for the public announcement as part of the Innovation X-Lab on May 7. The Outreach team developed talking points, videos, and feature stories from scientists on the importance of exascale computers and produced visual products, including pull-up banners and podium signs used for photo opportunities and media briefings. They co-hosted a series of media calls and briefings featuring the Laboratory Director and the CEOs of Cray and AMD. The team deployed a social media strategy that included branded tiles with Frontier facts and quotes from scientists. Their overall strategy was to emphasize the OLCF's legacy of delivering world-leading systems, to establish Frontier as ORNL's first exascale system, and to reinforce the OLCF's position as the premier location for researchers seeking to run codes at unprecedented speeds to pursue scientific innovation not possible on Summit, Titan, or any other previous systems. Before, during, and after the June event, the team monitored and responded to media queries and requests for interviews.

The team continued the Frontier story momentum when they collaborated to host the first Annual Exascale Day event on October 18 with a webcast media briefing and a social media campaign. The holiday—an initiative of DOE's ECP and Cray, a Hewlett Packard Enterprise Company—honors scientists and researchers who will make groundbreaking discoveries with the help of some of the fastest supercomputers in the world, such as Frontier. The date was a creative play on the performance of exascale computers as measured in exaflops. One exaflop equals one quintillion calculations per second, a number that is mathematically notated as 10^{18}.

To salute Titan's final year of operations, the team coordinated a number of tributes. They published a short farewell series highlighting some of Titan's most impactful contributions to scientific research, from the biofuels work of Jeremy Smith of ORNL, to the combustion work of Jacqueline Chen of Sandia National Laboratories. They coordinated photos with Titan for the user meeting attendees in June. In addition, they documented Titan's "last breath" when the machine was turned off in August with audio and video; they wrote a feature on Eric Lentz, the last Titan user, and his final runs on the system; and they mapped out the decommissioning and recycling process.

Overall, in 2019 the team was responsible for the creation of 77 highlights—including science and technology highlights and features about Titan and Summit users, OLCF staff members, and OLCF resources—and more than 224 total outreach products. Another way the OLCF reaches the public is through tours to groups of visitors who range from middle-school students to senior-level government officials. The center conducted tours for 351 groups in 2019, which included more than 4,584 individuals.

In all, the Outreach team produced at least 34 science highlights in 2019, including stories highlighting the first Summit results, a couple of retrospective items emphasizing outcomes from Titan's incredible 7 years of operation, and a short series — "Science at Exascale" — introducing Frontier. In addition, the team completed seven technology stories and 20 people features.

1.5 LOOKING FORWARD

1.5.1 Application Readiness in the Exascale Era

OLCF's CAAR is a partnership of OLCF staff, scientific application teams, vendor partners, and tools developers with the goal of readying a set of applications for the forthcoming Frontier exascale architecture. The CAAR for Frontier program is built on the legacy of the prior CAAR programs for OLCF-3 (Titan) and OLCF-4 (Summit). Eight CAAR applications were selected after a call for proposals in 2019, covering a broad range of scientific disciplines and employing a range of programming models and software designs. An additional 12 ECP applications are being engaged, with ECP funding, in the CAAR program and enjoy staff and computing resources similar to the OLCF-funded CAAR projects. The applications that are part of the Frontier CAAR program are summarized in Tables 1.7 (OLCF Funded) and 1.8 (ECP Funded).

Table 1.7. OLCF-funded applications in the Center for Accelerated Application Readiness (CAAR).

Application	Principal investigator	CAAR liaison	Scientific discipline
CHOLLA	Evan Schneider University of Pittsburgh	Reuben Budiardja	Astrophysics
CoMet	Dan Jacobson Oak Ridge National Laboratory	Wayne Joubert	Biology
GESTS	P.K. Yeung Georgia Institute of Technology	Stephen Nichols	Fluid Dynamics
LBPM	James McClure Virginia Tech	Mark Berrill	Subsurface Flow
LSMS	Markus Eisenbach Oak Ridge National Laboratory	Markus Eisenbach	Materials Science
NAMD	Emad Tajkhorshid University of Illinois, Urbana-Champaign	Arnold Tharrington and Josh Vermaas	Molecular Dynamics
NUCCOR	Morten Hjorth-Jensen Michigan State University/Facility for Rare Isotope Beams	Gustav Jansen	Nuclear Physics
PIConGPU	Sunita Chandrasekaran University of Delaware	Ronnie Chatterjee	Plasma Physics

Table 1.8. ECP-funded applications in the Center for Accelerated Application Readiness (CAAR).

ECP AD Project	Principal investigator	ECP liaison	Scientific discipline
E3SM	Mark Taylor Sandia National Laboratories	Matt Norman	Climate
ExaAM	John Turner Oak Ridge National Laboratory	Stephen Nichols	Additive Manufacturing
ExaBiome	Kathy Yelick Lawrence Berkeley National Laboratory	Phil Roth	Biology

ExaSGD	Slaven Peles Pacific Northwest National Laboratory	Phil Roth	Electrical Grid
ExaStar	Dan Kasen Lawrence Berkeley National Laboratory	Austin Harris	Astrophysics
ExaSky	Salman Habib Argonne National Laboratory	Bronson Messer	Astrophysics
ExaSMR	Steven Hamilton Oak Ridge National Laboratory	Mark Berrill	Nuclear Energy
GAMESS	Mark Gordon Iowa State University	Dymtro Bykov	Chemistry
LatticeQCD	Andreas Kronfeld Fermi Laboratory	--	High Energy Physics
NWCHEMeX	Theresa Windus Iowa State University	Dmitry Liakh	Chemistry
PELE	Jackie Chen Sandia National Laboratories	Ronnie Chatterjee	Combustion
WDMApp	Amitava Bhattacharjee Princeton Plasma Physics Laboratory	Ed D'Azevedo	Fusion

For the OLCF-5 Frontier project, the Application Readiness Key Performance Parameter (KPP) metric was defined such that an application must achieve an application-specific Figure of Merit (FOM) that is four times greater on Frontier than on Summit. In CY 2019, all OLCF-funded CAAR projects drafted a FOM for their specific application and are beginning to baseline the FOM on Summit.

HIGH PERFORMANCE COMPUTING FACILITY
2019 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
March 2020

2.   OPERATIONAL PERFORMANCE

CHARGE QUESTION 2: Did the facility's operational performance meet established targets?

OLCF RESPONSE: Yes. The OLCF provides a series of highly capable and reliable systems for the user community. The 2019 reporting period includes full CY production periods for the following HPC resources: the IBM AC922 (Summit), the GPFS file system (Alpine), and the archival storage system (HPSS). To accommodate the preparatory work needed to install the Cray OLCF-5 system (Frontier), the Cray XK7 (Titan), the Cray XC30 (Eos), and the Lustre file systems (Spider II) were all decommissioned in August 2019. Metrics provided for those resources will reflect a partial CY leading up to decommissioning. The effectiveness of these resources is demonstrated by the operational performance metrics, which were met or exceeded in all cases. The OLCF team successfully managed policies and job-scheduling priorities that maximized access to these production systems. In 2019, the OLCF once again delivered all the compute hours committed to the three major allocation programs: INCITE, ALCC, and DD. The operational performance demonstrates that the OLCF delivered another prominent operational year of reliable and technically sufficient resources to the scientific research community.

2.1   OPERATIONAL PERFORMANCE SUMMARY

Operational performance measures the performance of the OLCF against a series of operational parameters. The two operational metrics relevant to the OLCF's operational performance are resource availability and the capability utilization of the HPC resources. The OLCF additionally describes resource utilization and GPU usage as reported numbers, not metrics.

2.2   IBM AC922 (SUMMIT) RESOURCE SUMMARY

The OLCF installed and deployed an IBM AC922 system named Summit, which became available for full production on January 1, 2019. Summit is composed of 4,608 high-density compute nodes, each equipped with two IBM POWER9 central processing units (CPUs) and six NVIDIA Volta GPUs. In total, the Summit system is capable of 200 petaflops of peak computational performance and has been recognized as the most powerful system in the world for its performance on both the high performance linpack (HPL) and conjugate gradient (HPCG) benchmark applications since June 2018.

2.3   CRAY XK7 (TITAN) RESOURCE SUMMARY

The OLCF upgraded the existing Cray Jaguar from a model XT5 to a model XK7, releasing it to production on May 31, 2013. The resulting system contained 18,688 NVIDIA K20X (Kepler) accelerators, in which each existing AMD Opteron connected to an NVIDIA Kepler to form a CPU-GPU pair. The completed XK7 system, which had more than 27 petaflops of peak computational capacity, was named Titan. The system was decommissioned on August 2, 2019.

2.4   CRAY XC30 (EOS) RESOURCE SUMMARY

Eos was a four-cabinet Cray XC30. The system had 736 Intel Xeon E5-2670 compute nodes and 47.6 TB of memory and provided the OLCF user community with a substantive large-memory-per-node computing platform. The Eos nodes were connected by Cray's Aries interconnect in a network topology called Dragonfly. All INCITE users were automatically granted access to the XC30. The system was decommissioned on August 2, 2019.

2.5   LUSTRE AND GPFS FILE SYSTEMS (SPIDER II AND WOLF) RESOURCE SUMMARY

In October 2013, the OLCF released Spider II, its next-generation Lustre parallel file system, to production. Spider II contained two instantiations of the /atlas file system, with an aggregate capacity of more than 30 PB and block-level performance of more than 1.3 TB/s. The Spider II file system was the default high-performance parallel file system for all compute resources. The system was decommissioned on August 2, 2019.

In March 2017, the OLCF procured, installed, and deployed the Wolf General Parallel File System (GPFS), which serves as the center-wide file system for the computational resources in the Open Production enclave. Wolf provides a total storage capacity of 8 PB and up to 120 GB/s of performance.

2.6   GPFS FILE SYSTEM (ALPINE) RESOURCE SUMMARY

In January 2019, the OLCF released Spider III, its next-generation global file system to support the computational resources in the OLCF. Spider III is a single GPFS namespace named Alpine, with a usable capacity of 250 PB and a filesystem-level performance of 2.5 TB/s. The Spider III file system is the default high-performance parallel file system for all of the OLCF's moderate compute resources.

2.7   DATA ANALYSIS AND VISUALIZATION CLUSTER (RHEA) RESOURCE SUMMARY

Rhea is a 512-node large memory data analytics Linux cluster. The primary purpose of Rhea is to provide a conduit for large-scale scientific discovery through pre- and post-processing of simulation data generated on Summit and Titan. Users with accounts on INCITE- or ALCC-supported projects are automatically given accounts on Rhea. DD projects may also request access to Rhea. Each of Rhea's nodes contains two 8-core 2.0 GHz Intel Xeon processors with hyperthreading and 128 GB of main memory (upgraded in 2015 from 64 GB). Rhea offers nine additional heterogeneous nodes, each of which boasts 1 TB of main memory and two NVIDIA Tesla K80 (Kepler GK210) GPUs. Rhea is connected to the OLCF's 250 PB high-performance GPFS file system, Spider III.

2.8   HIGH PERFORMANCE STORAGE SYSTEM RESOURCE SUMMARY

The OLCF provides a long-term storage archive system based on the high performance storage system (HPSS) software product co-developed by IBM, LANL, Sandia National Laboratories (SNL), LLNL, Lawrence Berkeley National Laboratory (LBNL), and ORNL. The ORNL HPSS instance is currently over 90 PB in size and provides ingestion rates of up to 30 GB/s via a 22PB front-end disk cache backed by a 13-frame Spectra Logic TFinity tape library housing 81 IBM TS1155 tape drives and over 10,000 tape media slots–giving ORNL a current capacity of 133PB expandable well into hundreds of petabytes. The archive has ingested over 1 PB in a single day within the last year with the average daily ingestion ranging between 100–150 TB/day. The archive environment consists of hardware from Dell, Brocade, NetApp, DataDirect Networks, and Spectra Logic.

2.9   VISUALIZATION RESOURCE SUMMARY

The EVEREST facility has three computing systems and two separate state-of-the-art visualization display walls. The primary display wall spans 30.5 ft × 8.5 ft and consists of eighteen 1920 × 1080 stereoscopic Barco projection displays arranged in a 6 × 3 configuration. The secondary display wall contains sixteen 1920 × 1080 planar displays arranged in a 4 × 4 configuration, providing a standard 16:9 aspect ratio. The stereoscopic capabilities allow the user to experience binocular depth perception. An array of sequentially pulsed infrared LED cameras record the physical position and orientation of the user, and the resolution density provides an optimal solution for human visual acuity. These combined technologies, along with OLCF staff expertise, allow scientists to analyze complex scientific datasets in an immersive environment and communicate abstract concepts in an intuitive visual format.

2.10   OLCF COMPUTATIONAL AND DATA RESOURCE SUMMARY

The OLCF provided the Summit, Titan, and Eos computational resources and the Spider II, Spider III, and HPSS data resources for production use in 2019 (Table 2.1). Supporting systems such as EVEREST, Rhea, and data transfer nodes were also offered. Metrics for these supporting systems are not provided.

Table 2.1. OLCF production computer systems, 2019.

System	Access	Type	CPU	GPU	Computational description			Interconnect
					Nodes	Node configuration	Memory configuration	
Summit	Full production	IBM AC922	3.45 GHz IBM POWER9 (22 core)	1530 MHz NVIDIA V100 (Volta)	4,608	IBM POWER9 CPUs (2/node) + NVIDIA Volta V100 GPUs (6/node)	512 GB DDR4 and 96 GB HBM2 per node; >10PB DDR4 + HBM + Non-volatile aggregate	Mellanox EDR 100G Infiniband (Non-blocking Fat Tree)
Titan	Full production	Cray XK7	2.2 GHz AMD Opteron 6274 (16 core)	732 MHz NVIDIA K20X (Kepler)	18,688	16-core SMP[a] CPU + 14 SM[b] GPU (hosted)	32 GB DDR3-1600 and 6 GB GDDR5 per node; 598,016 GB DDR3 and 112,128 GB GDDR5 aggregate	Gemini (Torus)
Eos	Full production	Cray XC30	2.6 GHz Intel E5-2670 (8 core)	None	736	2 × 8-core SMP	64 GB DDR3—1,600 per node; 47,104 GB DDR3 aggregate	Aries (Dragonfly)

[a] SMP = symmetric multiprocessing
[b] SM = streaming multiprocessor

2.10.1 OLCF HPC Resource Production Schedule

The OLCF computational systems entered production according to the schedule in Table 2.2. This includes historical data associated with the Cray XT5, the very small overlap in December 2011 beginning with the introduction of the Cray XK6, and the series of Cray XK systems first available in 2012 and 2013.

Table 2.2. OLCF HPC system production dates, 2008–present.

System	Type	Production date[a]	Performance end date[b]	Notes
Summit	IBM AC922	January 1, 2019	Present	Production with 4,608 hybrid CPU-GPU nodes [IBM POWER9 CPUs (2/node) + NVIDIA Volta V100 GPUs (6/node)]
Spider III	GPFS parallel file system	January 1, 2019	Present	250 PB GPFS single namespace file system
Spider II	Lustre parallel file system	October 3, 2013	August 2, 2019	Delivered as two separate file systems, /atlas1 and /atlas2. 30+ PB capacity
Eos	Cray XC30	October 3, 2013	August 2, 2019	Production with 736 Intel E5, 2,670 nodes
Titan	Cray XK7	May 31, 2013	August 2, 2019	Production with 18,688 hybrid CPU-GPU nodes (AMD Opteron 6274/NVIDIA K20X)
JaguarPF	Cray XK6	September 18, 2012	October 7, 2012	Production at 240,000 cores until September 18, when partition size was reduced to 120,000 AMD Opteron cores. Additional Kepler installation. TitanDev access terminated
JaguarPF	Cray XK6	February 13, 2012	September 12, 2012	Full production until September 12, when partition size was reduced to 240,000 AMD Opteron cores. Beginning of Kepler installation
JaguarPF	Cray XK6	February 2, 2012	February 13, 2012	Stability test. Restricted user access. 299,008 AMD Opteron 6274 cores. Includes 960-node Fermi-equipped partition
JaguarPF	Cray XK6	January 5, 2012	February 1, 2012	Acceptance. No general access 299,008 AMD Opteron cores
JaguarPF	Cray XK6	December 12, 2011	January 4, 2012	142,848 AMD Opteron cores
JaguarPF	Cray XT5	October 17, 2011	December 11, 2011	117,120 AMD Opteron cores
JaguarPF	Cray XT5	October 10, 2011	October 16, 2011	162,240 AMD Opteron cores

Table 2.2. OLCF HPC system production dates, 2008–present (continued).

System	Type	Production date[a]	Performance end date[b]	Notes
JaguarPF	Cray XT5	September 25, 2009	October 9, 2011	224,256 AMD Opteron cores
JaguarPF	Cray XT5	August 19, 2008	July 28, 2009	151,000 AMD Opteron cores

[a] The production date used for computing statistics is either the initial production date or the production date of the last substantive upgrade to the computational resource.
[b] The performance end date is the last calendar day that user jobs were allowed to execute on that partition.

2.10.2 Operational Performance Snapshot

Operational performance metrics are provided for the OLCF computational resources, the HPSS archive system, and the external Lustre file systems (Tables 2.3–2.8).

Table 2.3. OLCF Operational performance summary for Summit.

	Measurement	2019 target	2019 actual
IBM AC922 (Summit)	Scheduled availability	85%	99.92%
	Overall availability	80%	98.52%
	MTTI[a] (hours)	NAM[c]	616
	MTTF[b] (hours)	NAM	2,918
	Total usage	NAM	78.01%
	Node-hours used[d]	NAM	31,144,061
	Node-hours available	NAM	39,924,230
	Capability usage		
	INCITE projects[e]	NAM	39.79%
	All projects	30%	47.54%

[a] MTTI = Mean time to interrupt.
[b] MTTF = Mean time to failure.
[c] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.
[d] Does not include usage recorded during an outage.
[e] Does not include INCITE 2019 13th month usage in January 2020.

Table 2.4. OLCF Operational performance summary for Titan.

	Measurement	2018 target	2018 actual	2019 target	2019 actual (Jan. 1–Aug. 2)
Cray XK7 (Titan)	Scheduled availability	95%	98.97%	95%	99.87%
	Overall availability	90%	98.51%	90%	99.60%
	MTTI[a] (hours)	NAM[c]	784.5	NAM	1,279
	MTTF[b] (hours)	NAM	1,238.61	NAM	1,709
	Total usage	NAM	90.6%	NAM	87.28%
	Core-hours used[d]	NAM	4,382,897,140	NAM	2,502,796,338
	Core-hours available	NAM	4,838,033,536	NAM	2,867,411,968
	Capability usage				
	INCITE projects[e]	NAM	57.59%	NAM	60.20%
	All projects	35%	51.09%	35%	53.88%

[a] MTTI = Mean time to interrupt
[b] MTTF = Mean time to failure
[c] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.
[d] Does not include usage recorded during an outage.
[e] Does not include INCITE 13th month usage in January 2017/January 2018.

Table 2.5. OLCF Operational performance summary for Eos.

	Measurement	2018 target	2018 actual	2019 target	2019 actual (Jan. 1 - Aug. 2)
Cray XC30 (Eos)	Scheduled availability	NAM[c]	99.21%	NAM	99.85%
	Overall availability	NAM	98.86%	NAM	99.61%
	MTTI[a] (hours)	NAM	1,237.22	NAM	1,705
	MTTF[b] (hours)	NAM	2,172.65	NAM	2,564

[a] MTTI = Mean time to interrupt
[b] MTTF = Mean time to failure
[c] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

Table 2.6. OLCF Operational performance summary for HPSS.

	Measurement	2018 target	2018 actual	2019 target	2019 actual
HPSS	Scheduled availability	95%	99.87%	95%	99.79%
	Overall availability	90%	99.53%	90%	97.97%
	MTTI[a] (hours)	NAM[c]	792.6	NAM	277
	MTTF[b] (hours)	NAM	1,749.71	NAM	874

[a] MTTI = Mean time to interrupt
[b] MTTF = Mean time to failure
[c] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

Table 2.7. OLCF Operational performance summary for Spider II, the external Lustre file system.

	Measurement	2018 target	2018 actual	2019 target	2019 actual
/atlas1	Scheduled availability	95%	99.95%	95%	100%
	Overall availability	90%	99.95%	90%	99.91%
	MTTI[a] (hours)	NAM[c]	2,188.9	NAM	4,376
	MTTF[b] (hours)	NAM	2,188.9	NAM	8,760
/atlas2	Scheduled availability	95%	99.29%	95%	99.99%
	Overall availability	90%	99.29%	90%	99.90%
	MTTI (hours)	NAM	724.82	NAM	2,188
	MTTF (hours)	NAM	724.82	NAM	2,920

[a] MTTI = Mean time to interrupt
[b] MTTF = Mean time to failure
[c] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

Table 2.8. OLCF Operational performance summary for Spider III, the external GPFS file system.

	Measurement	2019 target	2019 actual
Alpine	Scheduled availability	90%	99.53%
	Overall availability	85%	98.18%
	MTTI[a] (hours)	NAM[c]	277
	MTTF[b] (hours)	NAM	581

[a] MTTI = Mean time to interrupt
[b] MTTF = Mean time to failure
[c] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

For a period of 1 year following either system acceptance or a major system upgrade, the scheduled availability (SA) target for an HPC compute resource is at least 85%, and the overall availability (OA) target is at least 80%. For year 2, the SA target for an HPC compute resource increases to at least 90%, and the OA target increases to at least 85%. For year 3 through the end of life for the associated compute resource, the SA target for an HPC compute resource increases to 95%, and the OA target increases to 90%. Consequently, SA targets are described as 85%/90%/95%, and OA targets are described as 80%/85%/90%.

For a period of 1 year following either system acceptance or a major system upgrade, the SA target for an external file system is at least 90%, and the OA target is at least 85%. For year 2 through the end of life of the asset, the SA target for an external file system increases to at least 95%, and the OA target increases to at least 90%. SA targets are thus described as 90%/95%. OA targets are thus described as 85%/90%.

The Spider II, Titan, and Eos systems all celebrated their 6-year production anniversaries in 2019. The reported results for each system measure are for CY 2019 and intentionally do not reflect the partial results to their respective production anniversaries. In all cases, the OLCF results exceeded the most stringent year 3 and beyond targets for the accompanying metrics.

2.11 RESOURCE AVAILABILITY

Details of the definitions and formulas describing SA, OA, MTTI, and MTTF are provided in Appendix D.

2.11.1 Scheduled Availability

The SA is described by Eq. (1). The OLCF has exceeded the SA targets for the facility's computational resources for 2018 and 2019 (Table 2.9).

$$SA = \left(\frac{time\ in\ period - time\ unavailable\ due\ to\ outages\ in\ period}{time\ in\ period - time\ unavailable\ due\ to\ scheduled\ outages\ in\ period} \right) * 100 \tag{1}$$

Table 2.9. OLCF Operational performance summary: Scheduled availability.

	System	2018 target	2018 actual	2019 target	2019 actual
Scheduled availability	IBM AC922	NIP[a]	NIP	85%	99.92%
	Cray XK7	95%	98.97%	95%	99.87%
	Cray XC30	NAM[b]	99.21%	NAM	99.85%
	HPSS	95%	99.87%	95%	99.79%
	/atlas1	95%	99.95%	95%	100%
	/atlas2	95%	99.29%	95%	99.99%
	Alpine	NIP	NIP	90%	99.53%

[a] NIP = Not in production
[b] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

2.11.1.1 Assessing Impacts to Scheduled Availability

Preventative maintenance is exercised only with the concurrence of the vendor hardware and software teams, the OLCF HPC Operations groups, and the NCCS Operations Council. Typical preventative maintenance activities include software updates, application of field notices, and hardware maintenance to replace failed components. Without concurrence, the systems remain in their respective normal operating conditions. Preventative maintenance is advertised to users a minimum of 2 weeks in advance if the maintenance activities include changing default software and a minimum of 1 week in advance if default software is not being changed.

In 2017, representatives from the OLCF, Argonne Leadership Computing Facility (ALCF), and NERSC agreed that during a scheduled maintenance, a significant event that delays the return of a system to scheduled production by more than 4 hours will be counted as an adjacent unscheduled outage, as unscheduled availability, and as an additional interrupt. Prior to 2017 each facility was handling this situation in a different way, and we were asked by ASCR to come up with a consistent definition.

2.11.2 Overall Availability

The OA of OLCF resources is derived using Eq. (2).

$$OA = \left(\frac{time\ in\ period - time\ unavailable\ due\ to\ outages\ in\ period}{time\ in\ period} \right) * 100 \tag{2}$$

As shown in Table 2.10, the OLCF exceeded the OA targets of the facility's resources for 2018 and 2019.

Table 2.10. OLCF Operational performance summary: Overall availability.

	System	2018 target	2018 actual	2019 target	2019 actual
Overall availability	IBM AC922	NIP[a]	NIP	80%	98.52%
	Cray XK7	90%	98.51%	90%	99.60%
	Cray XC30	NAM[b]	98.86%	NAM	99.61%
	HPSS	90%	99.53%	90%	97.97%
	/atlas1	90%	99.95%	90%	99.91%
	/atlas2	90%	99.29%	90%	99.90%
	Alpine	NIP	NIP	85%	98.18%

[a] NIP = Not in production
[b] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

2.11.3 Mean Time to Interrupt

The MTTI for OLCF resources is derived by Eq. (3), and a summary is shown in Table 2.11.

$$MTTI = \left(\frac{time\ in\ period - (duration\ of\ scheduled\ outages + duration\ of\ unscheduled\ outages)}{number\ of\ scheduled\ outages + number\ of\ unscheduled\ outages + 1} \right) \quad (3)$$

Table 2.11. OLCF Operational performance summary: Mean time to interrupt.

	System	2018 actual	2019 actual
MTTI[a] (hours)	IBM AC922	NIP[b]	616
	Cray XK7	784.5	1,279
	Cray XC30	1,237.22	1,705
	HPSS	792.6	277
	/atlas1	2,188.9	4,376
	/atlas2	724.82	2,188
	Alpine	NIP	277

[a] MTTI is not a metric. Data provided as reference only.
[b] NIP = Not in production

2.11.4 Mean Time to Failure

The MTTF is derived from Eq. (4), and a summary is provided in Table 2.12.

$$MTTF = \frac{time\ in\ period - (duration\ of\ unscheduled\ outages)}{number\ of\ unscheduled\ outages + 1} \quad (4)$$

Table 2.12. OLCF Operational performance summary: Mean time to failure.

	System	2018 actual	2019 actual
MTTF^a (hours)	IBM AC922	NIP^b	2,918
	Cray XK7	1,238.61	1,709
	Cray XC30	2,172.65	2,564
	HPSS	1,749.71	874
	/atlas1	2,188.9	8,760
	/atlas2	724.82	2,920
	Alpine	NIP	581

^{a} MTTF is not a metric. The data is provided as reference only.
^{b} NIP = Not in production

2.12 RESOURCE UTILIZATION 2019

2.12.1 Operational Assessment Guidance

The facility reports total system utilization for each HPC computational system as agreed upon with the program manager.

The numbers that are reported for the Cray XK7 resource are Titan core-hours, which are composed of 16 AMD Opteron core-hours and 14 NVIDIA Kepler SM-hours per Titan node-hour. The OLCF refers to the combination of these traditional core-hours and SM-hours as "Titan core-hours" to denote they are the product of a hybrid node architecture. System production requires the use of node-hours, which is an aggregate of all CPU and GPU resources comprising a single node. The use of node-hours impacts all scheduling and accounting activities. Users describe all job submission activity in node-hours as the smallest unit. The numbers reported for the IBM AC922 resource are Summit node-hours, which conforms to the traditional model of an entire node being one unit per hour.

2.12.2 Resource Utilization Snapshot

For the Cray XK7 during the operational assessment period January 1–August 2, 2019, 2,502,796,338 Titan core-hours were used outside of outage periods from an available 2,867,411,968 Titan core-hours. The total system utilization for the Cray XK7 was 87.28%. For the IBM AC922 during the operational assessment period January 1–December 31, 2019, 31,144,061 Summit node-hours were used outside of outage periods from an available 39,924,230 Summit node-hours. The total system utilization for the IBM AC922 was 78.01%.

2.12.3 Total System Utilization

2.12.3.1 2019 Operational Assessment Guidance

System utilization (SU) is the percentage of time that the system's computational nodes run user jobs. No adjustment is made to exclude any user group, including staff and vendors.

$$SU = \left(\frac{core\ hours\ used\ in\ period}{core\ hours\ available\ in\ period} \right) * 100 \tag{5}$$

The measurement period is for 2019, irrespective of the prescribed allocation period of any single program. As an example, the INCITE allocation period follows a CY schedule. The ALCC program follows an allocation cycle that runs for 12 months, beginning July 1 of each year. Titan system utilization for 2019 was 87.28%. The OLCF tracks the consumption of Titan node-hours by job. By extension, this provides a method for tracking Titan core-hours by job. This method is extended to track the consumption of Titan core-hours by program, project, user, and system with high fidelity. Figure 2.1 summarizes the Cray XK7 utilization by month and by program for 2019. The three major OLCF user programs and usage by the ECP are represented, but the graph does not include consumed core-hours from staff or vendor projects. For the first production year of Summit, utilization was 78.01%. Utilization on Summit started the year at lower levels as users transitioned to a new platform but picked up significantly as the year closed out. Figure 2.2 summarizes the IBM AC922 utilization by month and by program for all of 2019 and, again, does not include consumed node-hours from staff or vendor projects.
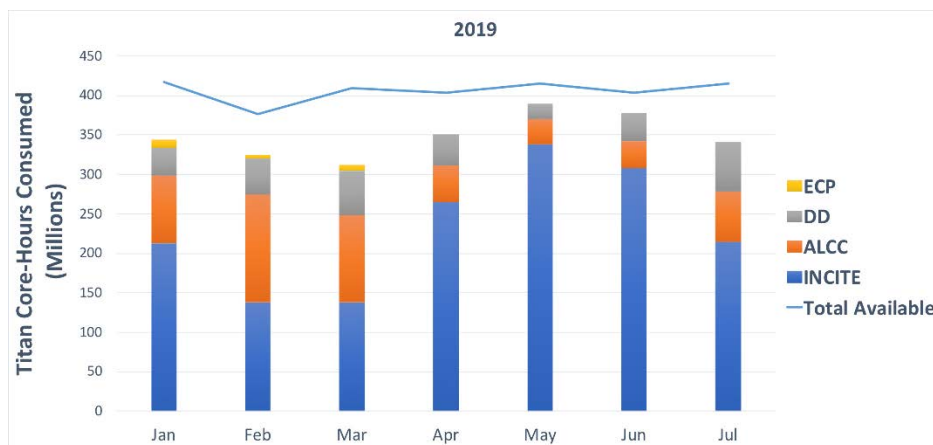
Figure 2.1. 2019 Cray XK7 resource utilization—Titan core-hours by program.
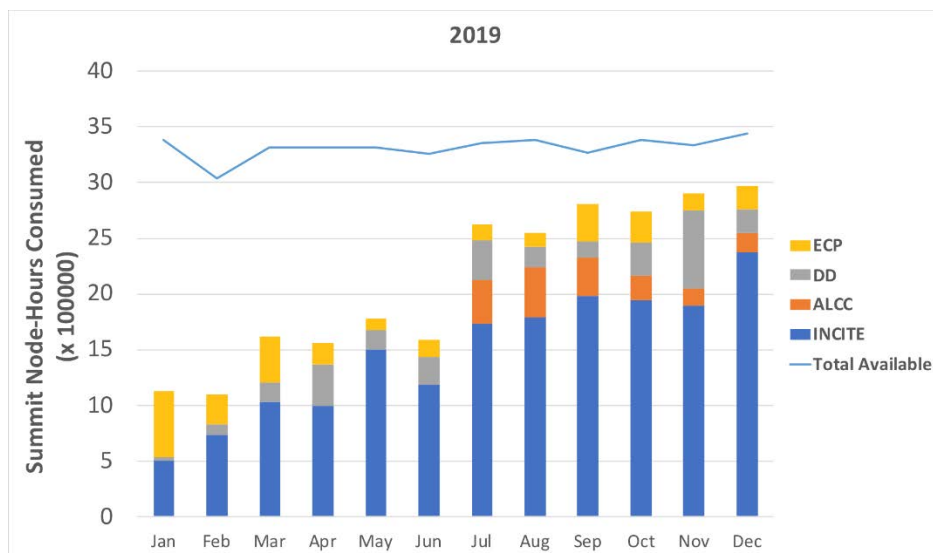
Figure 2.2. 2019 IBM AC922 resource utilization—Summit node-hours by program.

2.12.3.2 Performance of the Allocated Programs

All allocation programs, including INCITE, ALCC, and DD, are aggressively monitored to ensure that projects within these allocation programs maintain appropriate consumption rates. The 2019 INCITE

allocation program was the largest program in 2019, with a commitment for 1.17 billion Titan core-hours and 13.7 million Summit node-hours. The consumption of these allocation programs is shown in Tables 2.13 and 2.14. As shown, all commitments were exceeded for each allocation program on Titan and Summit for 2019. This programmatic overachievement is assisted by the high uptime and diligent work of the OLCF operational staff.

Non-renewed INCITE projects from 2018 continued running through January 2019 under the OLCF's 13th month policy. This policy is in place to permit an additional, final month for completion and was recognized as a best practice during a previous OAR review. It also serves to maintain high utilization while new projects establish a more predictable consumption routine. ALCC projects from the 2019 allocation period (ending June 30, 2019) were also granted extensions where appropriate.

Table 2.13. The 2019 allocated program performance on Titan.

Program	Allocation	Hours consumed	Percent of total
INCITE[a]	1,170,000,000	1,614,471,523	66.80%
ALCC[b]	Allocation spans multiple CY	509,914,369	21.10%
DD	—	292,641,731	12.10%
Total		2,417,027,623[c]	100%

[a] Includes all INCITE program usage for CY 2019.
[b] Includes all ALCC program usage for CY 2019.
[c] Does not include usage outside of the three primary allocation programs.

Table 2.14. The 2019 allocated program performance on Summit.

Program	Allocation	Hours consumed	Percent of total
INCITE[a]	13,700,000	17,657,822	78.70%
ALCC[b]	Allocation spans multiple CY	1,788,157	7.97%
DD	—	2,991,536	13.33%
Total		22,437,515[c]	100%

[a] Includes all INCITE program usage for CY 2019.
[b] Includes all ALCC program usage for CY 2019.
[c] Does not include usage outside of the three primary allocation programs.

2.13 CAPABILITY UTILIZATION

To be classified as a capability job, any single job must use at least 20% of the available nodes of the leadership system. For the CY following a new system/upgrade, at least 30% of the consumed node-hours will be from jobs requesting 20% or more of the available nodes. In subsequent years, at least 35% of consumed node-hours will be from jobs requiring 20% or more of nodes available to the users. The metric for capability utilization describes the aggregate number of core-hours (Titan) or node-hours (Summit) delivered by capability jobs. The metric for CY 2019 was 35% for Titan and 30% for Summit based on years of service, as described above. The OLCF continues to exceed expectations for capability usage of its HPC resources (Tables 2.15 and 2.16). Keys to successful demonstration of capability usage include the liaison role provided by SciComp members, who work hand-in-hand with users to port, tune, and scale code, and the OLCF support of the application readiness efforts (i.e., CAAR), which actively engage with code developers to promote application portability, suitability to hybrid systems, and performance. The OLCF aggressively prioritizes capability jobs in the scheduling system.

Table 2.15. OLCF capability usage on the Cray XK7 (Titan) system.

Leadership usage	CY 2018 target	CY 2018 actual	CY 2019 target	CY 2019 actual
INCITE	NAM[a]	57.6%	NAM	60.2%
ALCC	NAM	49.8%	NAM	61.4%
All projects	35%	51.1%	35%	53.9%

[a] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

Table 2.16. OLCF capability usage on the IBM AC922 (Summit) system.

Leadership usage	CY 2019 target	CY 2019 actual
INCITE	NAM[a]	39.8%
ALCC	NAM	52.7%
All projects	30%	47.6%

[a] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

The average consumption of hours by capability jobs, 53.9% on Titan and 47.6% on Summit, was well above the 2019 target of 35% for Titan and 30% for Summit. There was an increase over CY 2018 in the INCITE and ALCC programs on Titan and the overall average, which is remarkable for a system the age of Titan (6-year production anniversary in 2019). It is also noteworthy that Summit exceeded capability metrics in its first production year as users migrated to a new platform and were able to scale up in a short amount of time. To promote the execution of capability jobs, the OLCF provides queue prioritization for all jobs that use 20% or more of the nodes and further boosts the very largest of these jobs, which use >60% of the nodes, through aging boosts. The OLCF assesses job data in 10% "bins" to understand the job size distribution. Further, by assessing the aggregate bins, 20%–60% and >60%, the OLCF can assess the impact of queue policy on delivered node-hours. Figures 2.3 and 2.4 show the yearly average capability usage for each program, which describes the ratio of compute-hours delivered by capability jobs to the compute-hours delivered by non-capability jobs.
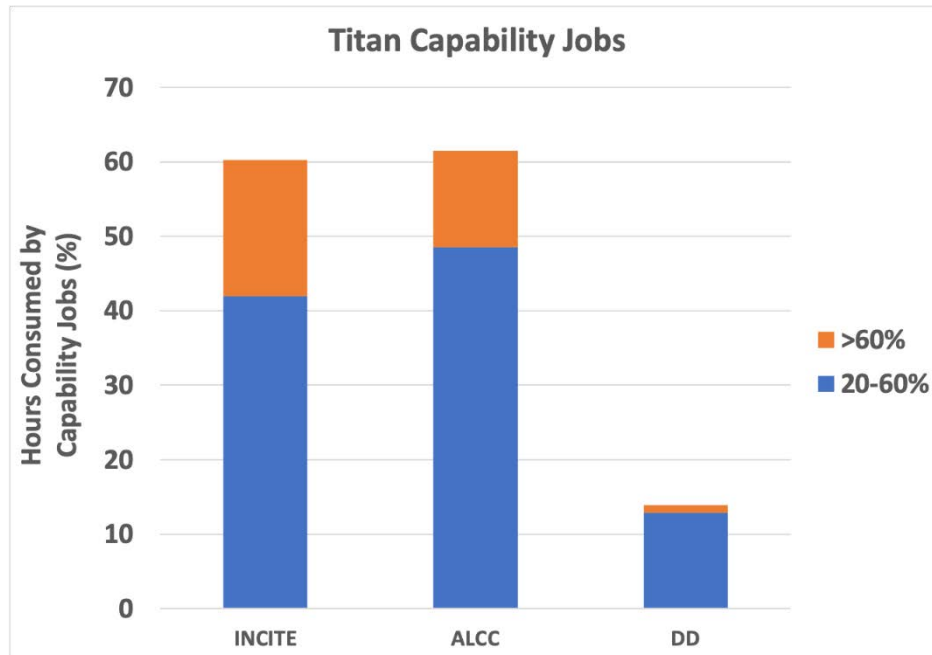
Figure 2.3. Titan capability usage by job size bins and project type.
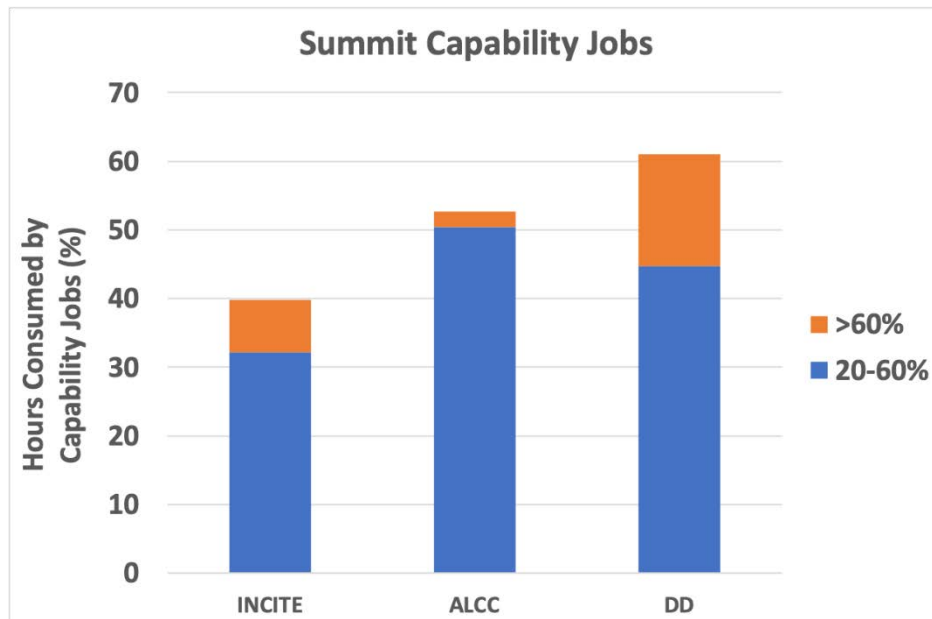
Figure 2.4. Summit capability usage by job size bins and project type.

2.14 GPU USAGE

The leadership computing systems at OLCF provide heterogeneous architectures to users that allow them to exploit a hybrid compute node that contains both CPUs and GPUs. Hybrid nodes provide researchers with diverse architecture that is well suited for certain operations. As such, the use of this diverse architecture is optional and is exercised in different ways by research teams.

In 2019, the OLCF continued tracking GPU usage on Titan through Cray's Resource Utilization Reporting tool and on Summit through NVIDIA's system management interface tool. Figure 2.5 shows the percentage of GPU-enabled hours compared with CPU-only hours of each of the three primary allocation programs at the OLCF (INCITE, ALCC, and DD). The INCITE program on Summit exceeds all programs with an extraordinary 94% of hours utilized on Summit using the GPUs. Certainly, the overwhelming majority of the capability of the Summit nodes resides in the GPUs, and this number as well as the percentages from the ALCC and DD programs—both at approximately 80%—reflects the recognition and use of that capability from the OLCF user programs. Overall, approximately 51% of all delivered compute time on Titan was consumed by GPU-enabled applications in CY 2019, while Summit boasted a hefty 87% in 2019 for GPU-enabled applications. The DD program on Titan was obviously the weakest given most users were migrating to Summit and not working to scale applications on Titan.
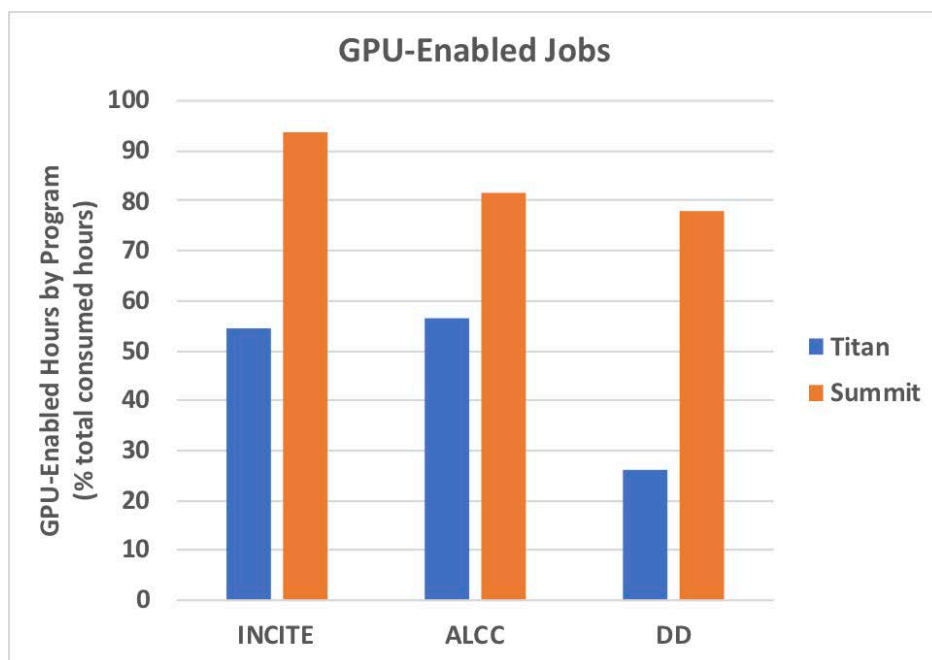
Figure 2.5. 2019 GPU-enabled usage by program.

2.15 CENTER-WIDE OPERATIONAL HIGHLIGHTS

2.15.1 Farewell, Titan: System Accomplishments

The OLCF's Cray XK7 Titan supercomputer was decommissioned on August 1 and disassembled for recycling. Performing up to 27 quadrillion calculations per second, Titan ranked as one of the world's top 12 fastest supercomputers from its debut as No. 1 in 2012 to its No. 12 position on the June 2019 Top500 List. Titan was a new generation of supercomputer with a revolutionary architecture that combined AMD 16 core Opteron CPUs and NVIDIA Kepler accelerated GPU processors, which tackled computationally intensive math problems while the CPUs efficiently directed tasks. Titan replaced Jaguar, debuting in late 2012 with 10 times the performance of Jaguar, reaching a peak performance of 27 petaflops, while consuming roughly the same amount of energy. Through more than 26 billion core-hours of computing time, Titan served hundreds of research teams around the world working on today's most urgent scientific challenges. Titan supported 896 unique projects. From its first official user projects in January 2013, through its final run in 2019, 2.86 million jobs were executed on Titan. The team ensured reliable operations, providing 31.4 billion core-hours and an overall 26.96 billion core-hour usage, excluding downtimes.

2.15.2 HPSS Tape Library and Disk Cache Upgrades

In 2019, The OLCF Archive team replaced the center's six aging Oracle SL8500 tape libraries with a new single TFinity library produced by Spectra Logic Corporation. OLCF tape libraries primarily store user data accumulated from applications run on the Titan and Summit supercomputers, as well as some information from other OLCF computing resources and different user facilities. Many of the OLCF's six existing SL8500 libraries had been operating for 10 years or longer, making this large-scale upgrade a significant undertaking.

Between February and August 2019, the Archive team seamlessly migrated all 68 petabytes of data, spanning tens of thousands of tapes residing in the legacy SL8500 tape libraries into a single Spectra Logic TFinity library, while also reducing the physical footprint by 85%. The migration operations were performed live and with no impact to end users. The resulting compact, row/rack-oriented configuration provides the team with substantial space to eventually expand the archive to a maximum capacity of 1.52 exabytes.

Meanwhile, the HPSS disk cache was also upgraded in 2019 from five DDN SFA12K storage subsystems offering 18 petabytes of usable space with a maximum bandwidth of ~1GB/s into HPSS to three DDN SFA14K storage subsystems providing 22 petabytes of usable space with maximum read/write performance approaching 30 GB/s using the HPSS application.

The 15 petabytes of data resident on the old disk cache hardware was migrated off the old hardware and onto the new hardware with no impact to production operations.

2.15.3 Transition from Moab/TORQUE to Slurm

Dating back several years, the primary batch scheduler and job management software used to access and utilize OLCF compute resources has been Moab/Torque. Due to the adoption of Load Sharing Facility (LSF) as the primary batch/job scheduler on the Summit supercomputer, coupled with the dwindling support and rising costs associated with the ongoing use of Moab/Torque, the center began an effort in early 2019 to begin migrating systems still using Moab/Torque over to Slurm. Slurm was the ideal choice given its wide adoption and development in the open-source community and at other labs. It would allow for tighter integration with our RATs application and automated workflows associated with resource/utilization requests and approvals (described in detail in Section 1 of this report, User Support Results), and it is the designated batch scheduler and job management software for the OLCF-5 supercomputer.

The overall transition involved everything from general requirements gathering, creating a testbed, performing test shots, running both Moab/Torque and Slurm in parallel for a period of time, and then a production switchover. The HPC Operations groups worked very closely with the User Assistance and Outreach group to make sure that users were informed and educated of the change, kept up to date through the transition project, and that all existing functionality with respect to job submissions, execution, management, etc., was replicated and tested thoroughly.

As described in Section 1, running both schedulers in parallel for a period of time allowed users to thoroughly test their own workflows to ensure the final switchover would be as seamless as possible. This transition period also gave the Operations groups time to work through any issues not found in earlier testing. All these efforts resulted in a very smooth and relatively seamless transition in September of 2019.

2.15.4 Move from Lustre to GPFS for Center-Wide Scratch Filesystem

In 2019, the production OLCF center-wide scratch filesystem transitioned from Lustre to GPFS as part of the transition from Titan to Summit. The legacy scratch filesystem, Atlas, used by Titan and other OLCF-resources such as the Rhea compute cluster and various Data Transfer Nodes (DTNs), was a Lustre-based filesystem consisting of two namespaces totaling ~28 PB of usable space with a maximum aggregate performance of ~1.4 TB/sec.

In conjunction with the installation of the Summit supercomputer, a new GPFS scratch filesystem, Alpine, was deployed. This new filesystem provides a significant amount of additional capacity and increased performance but offers several other benefits to the center and our user community as well.

In general capacity and performance terms, Alpine provides ~250 PB of usable scratch space with a maximum aggregate throughput of ~2.5 TB/sec. Single-client performance is significantly better than the previous Lustre filesystems, and we have seen up to 24 GB/sec from a single Summit compute node. Beyond that, we now have the benefit of a single namespace which no longer requires the movement and/or select placement of data to balance utilization, and our users can realize the full performance capabilities of the entire backend storage subsystem(s). Additionally, GPFS provides better IO performance for small files, as the software aggregates I/Os on the client resulting in larger network transactions seen as larger I/O operations by the GPFS I/O servers. Small writes received by GPFS I/O servers are coalesced onto a small non-volatile memory (NVMe) drive resident on the storage nodes before hitting the backend disks, which also significantly increases overall small file I/O performance. Finally, the GPFS page pool helps with re-read workloads, such as those found in many artificial intelligence (AI) models. Jobs accessing the same small file(s) repeatedly will have the benefit of near instantaneous access to those file(s); they must only be read from disk and traverse the network once.

There are several benefits to GPFS from an administration perspective as well. GPFS provides automatic failover between storage servers, which reduces the risk of a filesystem outage in the event an individual server in a failover pair experiences a failure. The GPFS RAID engine feature, "Disk Hospital," more quickly and accurately identifies slow or failing disks and proactively fails and/or rebuilds them—helping reduce I/O performance variability. And lastly, the GPFS filesets feature allows staff to audit utilization on a per-project basis, while the API allows for easy integration with our RATs accounting system to track filesystem utilization metrics and provide reports.

While the primary scratch filesystem for the OLCF is now GPFS based, the OLCF still maintains a number of production Lustre filesystems and remains very active in the Lustre community as well. We continue to work in tandem with the development community, other labs, and various vendors to continue to drive improvements and enhancements in Lustre. Several staff members are associated with and hold leadership positions in the Open Scalable File System (Open SFS) consortium and are regular attendees and presenters at the annual Lustre User Group (LUG) conference.

2.15.5 Apache Kafka Deployment

Employing the open-source Apache Kafka event-streaming platform, ORNL's HPC Core Operations Group built a whole new publish/subscribe system for delivering metrics from OLCF supercomputers. The previous method of requesting system data often meant waiting for new, individual data feeds to be built or connected and then collecting the information in a daily time-consuming process. The new "pubsub" system replaces that typically slow, manual process for each data pipeline with a centralized, automated information bus that aggregates available data sources for subscribers. Under this new system, a directory of all the available data feeds will allow OLCF staffers to know exactly what they can tap into rather than making a new request each time. This will also better facilitate the sharing of data between

teams for business insights and collaborations. In addition to being much more user-centric, the Kafka-powered metrics system promises much higher levels of speed, reliability, and retention.

2.15.6 WatsonML Test and Deployment

Deployment of machine learning/deep learning (ML/DL) frameworks on Summit poses challenges due to both the fast development pace of those frameworks and Summit's specific IBM Power architecture. The Advanced Data and Workflow group is working with the User Assistance and Outreach group to address the data analytics needs of our users. We started to evaluate the IBM WatsonML (formerly known as PowerAI) software stack by comparing the performance of its ML/DL frameworks with our own builds in both native and container mode, which is documented for our users (https://code.ornl.gov/jqyin/mldl-hpc/blob/master/documentation/PowerAI-on-Summit.md).

Based on good feedback, we then worked with the IBM WatsonML-CE team to officially deploy the module on Summit and test the framework performance at scale, which is comparable with the current state of the art. We also worked with IBM to host an "Introduction to WML" webinar during the October 30 user conference call, which was the most highly attended monthly conference call of the year. The official documentation for Summit users is at https://docs.olcf.ornl.gov/software/analytics/ibm-wml-ce.html. Some of the performance numbers are also reported in our paper "Strategies to Deploy and Scale Deep Learning on the Summit Supercomputer," *2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS)*, pp. 84–94.

3.   ALLOCATION OF RESOURCES

CHARGE QUESTION 3: Is the allocation of resources reasonable and effective?

OLCF RESPONSE: Yes. The OLCF continues to enable high-impact science results through access to the leadership-class systems and support resources. The allocation mechanisms are robust and effective. The OLCF enables compute and data projects through the Director's Discretionary (DD) user program. This program seeks to enable researchers through goals that are strategically aligned with ORNL and DOE, as described in Section 3.1.

3.1   ALLOCATION OF RESOURCES: FACILITY DIRECTOR'S DISCRETIONARY RESERVE TIME

This section should provide insight into the strategic rationale behind use of the DD's reserve. The facility should describe how the DD's reserve is allocated and list the awarded projects, showing the PI name, sponsor organization(s), hours awarded, and project title.

3.1.1   The OLCF Director's Discretionary Program

The OLCF primarily allocates time on leadership resources through the INCITE program and through the facility's DD program. The OLCF seeks to enable scientific productivity via capability computing through both programs. Accordingly, a set of criteria are considered in making allocations, including the strategic impact of the expected scientific results and the degree to which awardees can effectively use leadership resources. Further, through the ALCC program, the ASCR office allocates up to 20% of the facility's resources.

The goals of the DD program are threefold.

1.   To enable users to prepare for leadership computing competitions, such as INCITE and ALCC (e.g., to improve and document application computational readiness)

2.   To broaden the community of researchers capable of using leadership computing by enabling new and nontraditional research topics

3.   To support R&D partnerships, both internal and external to ORNL, to advance DOE and ORNL strategic agendas

These goals are aligned particularly well with three of the OLCF's four missions.

1.   To enable high-impact, grand-challenge science and engineering that could not otherwise be performed without leadership-class computational and data resources

2. To enable fundamentally new methods of scientific discovery by building stronger collaborations with experimental facilities as well as DOE offices that have large compute and data science challenges

3. To educate and train the next-generation workforce in the application of leadership computing to solve the most challenging scientific and engineering problems

R&D partnerships are aligned with DOE and ORNL strategic agendas. They may be entirely new areas with respect to HPC, or they may be areas in need of nurturing. Examples of projects are those associated with the ORNL Laboratory Directed Research and Development program; programmatic science areas (fusion, materials, chemistry, climate, nuclear physics, nuclear engineering, and bioenergy science and technology); and key academic partnerships (e.g., the UT-ORNL Joint Institute for Computational Sciences).

Also included in this broad category are projects that come to the OLCF through the Accelerating Competitiveness through Computational ExceLlence (ACCEL) Industrial HPC Partnerships outreach, which encourages opportunities for industrial researchers to access the leadership systems through the usual leadership-computing user programs to carry out research that would not otherwise be possible. See Section 8.3.2 for more information about ACCEL.

The actual DD project lifetime is specified upon award: allocations are typically for 1 year or less. However, projects may request 3-month extensions or renewals up to an additional 12 months. The average size of a DD award is roughly 18,000 Summit node-hours, but awards can range from thousands to 150,000 node-hours, or more. In 2019, the OLCF DD program participants used approximately 10.6% of total user resources (over both Summit and Titan), consuming almost 3 million Summit node-hours and 292 million Titan core-hours. See Appendix E for a full list of DD projects for CY 2019.

4. INNOVATION

CHARGE QUESTION 4: (a) Have innovations been implemented that have improved the facility's operations? (b) Is the facility advancing research, either intramurally or through external collaborations, that will impact next generation high performance computing platforms? (c) Is the facility effectively utilizing their postdoctoral fellows?

OLCF RESPONSE: Yes. The OLCF actively pursues innovations that can enhance facility operations. Through collaborations with users, other facilities, vendors, and the broader digital infrastructure community, many of these innovations are disseminated and adopted across the country. In addition, the OLCF is training the future computational science and HPC workforce through the Computational Scientists for Energy, the Environment, and National Security (CSEEN) postdoctoral program.

Since the facility's inception in 2004, OLCF staff have provided leadership in the HPC community, spearheading the creation and development of tools and policies necessary for computing and computational science and disseminating that knowledge throughout the community through a variety of outlets. In 2019, the OLCF pursued innovative technological solutions and external collaborations to remain the state-of-the-art HPC facility in the United States. It is not possible to highlight all the innovative work carried out by the OLCF. Instead, this section will highlight certain specific areas of innovation in operations and research in 2019. We also highlight the nine postdoctoral fellows who were trained in the OLCF in CY 2019, the contributions they made to OLCF applications projects, and, when appropriate, the career path after their OLCF postdoctoral appointment.

4.1 OPERATIONAL INNOVATION

In this section, we discuss five examples of specific operational innovations to improve the facility operations: data hierarchy improvements, better metrics and benchmarking system-wide, accreditation to support protected health information, improved workflow and data services support, and application operation improvements.

4.1.1 Data Hierarchy Improvements

A growing area of importance to large-scale science is managing the increasingly complex data hierarchy, which includes CPU and GPU memory, node-proximal storage, and file system innovations. We discuss here innovations in managing this hierarchy by describing work in using the node-local non-volatile memory (NVMe) drives, supporting end-to-end in-system storage layer management, and collaborative work with LLNL from software development efforts in ECP incorporating the UnifyFS software for storage management.

Spectral I/O Interposition Library. In early 2019, the OLCF released the first version of Spectral onto Summit. Spectral is a transparent file-system acceleration library that uses the compute node-local NVMe drives in compute nodes to accelerate applications write-only (checkpoint) data without requiring applications to change code to maintain portability with other centers. The base design of Spectral supported file-per-process checkpoints and required the user to modify file paths to write directly to the

NVMe drives. However, many applications hardcode file paths while other applications use shared files to simplify postmortem analysis. To accommodate this use case, Spectral was extended to add automatic redirection of writes to local NVMe as well as shared file support. These enhancements broadened greatly the number of applications that can benefit from the node-local storage devices. Spectral has been verified at scale on Summit with applications such as Flash, GTC, and LAMMPS. The updated features are expected for full release on Summit in early 2020.

End-to-end I/O Portfolio for the Summit Supercomputing Ecosystem. The OLCF published this work in a paper at SC19 describing the Summit I/O subsystem and the portfolio of I/O solutions. The paper, titled "End-to-end I/O portfolio for the summit supercomputing ecosystem," discussed the design, acceptance, and transition to operations of the center-wide parallel file system (Spider III) and the in-system storage layer. We have implemented the new in-system storage layer to improve the ML/DL and data-intensive application performance. The in-system layer uses node-local SSDs and provides 26.7 TB/s for reads, 9.7 TB/s for writes, and 4.6 billion IOPS to Summit. The Spider III layer uses IBM's Spectrum Scale and provides 2.5 TB/s and 2.6 million IOPS to Summit and other systems. While deploying them as two distinct layers was operationally efficient, it also presented usability challenges in terms of multiple mount points and lack of transparency in data movement. To address these challenges, OLCF has developed novel end-to-end I/O solutions for the concerted use of the two storage layers. The paper presented the I/O subsystem architecture, the end-to-end I/O solution space, and their design considerations and the deployment experience.

Related paper: Sarp Oral, Sudharshan S. Vazhkudai, Feiyi Wang, Christopher Zimmer, Christopher Brumgard, Jesse Hanley, George Markomanolis, Ross Miller, Dustin Leverman, Scott Atchley, and Veronica Larrea, "End-to-end I/O Portfolio for the Summit Supercomputing Ecosystem," *Proceedings of Supercomputing 2019 (SC19): 32nd Int'l Conference on High Performance Computing, Networking, Storage and Analysis,* Denver, CO, November 2019.

UnifyFS. This ECP software technologies development project is a collaboration between ORNL and LLNL. UnifyFS is a shared namespace file system that enables the transparent use of node-local storage devices on current supercomputers and future exascale systems. It makes the use of node-local storage as easy as writing to a parallel file system and is an order of magnitude faster. Applications can access the shared UnifyFS namespace by simply replacing the file path for their file I/O to point to /unifyfs. UnifyFS transparently intercepts application I/O calls so that no application code modifications are necessary. UnifyFS is fast because it uses compute-node-local storage resources and is private to a user's job. UnifyFS began in 2017 with the key goal of hardening the research prototype into a production code for ECP applications. It currently supports VeloC, HDF5, and MPIIO I/O libraries, and our experiments show that UnifyFS can be up to 50× faster than the parallel file system.

4.1.2   Metrics and Benchmarking for Operational Improvements

Critical to improving our operations, we have begun collecting and using metrics from the system from various modalities discussed below. By evaluating historical resource utilization of Titan logs of CPU, GPU, DRAM, GDDR, and I/O usage, we have begun to derive a deeper understanding of application performance to make recommendations to users of our system. Summit metrics now being collected include temperature and cooling data, and we have also created appropriate benchmarks to understand the capabilities of Summit's network. In addition, we have begun collecting metrics at the application level using the XALT tool.

Titan RUR Log Analysis for CPU, GPU, DRAM, GDDR, IO Usage. To gather a deeper understanding of how major leadership computing resources, such as CPUs, GPUs, memory, and the I/O system, were used in production for Titan, the OLCF analyzed the Resource Utilization Data logs. This analysis

focused on the last 5 years of the Titan production lifetime. The results of our analysis were presented in a paper titled "Learning from Five-year Resource-Utilization Data of Titan System" at the Workshop on Monitoring and Analysis for High Performance Computing Systems Plus Applications (HPCMASPA 2019), held in conjunction with IEEE Cluster Conference 2019 in September 2019 and was shared with the research community. With Titan's mission complete, this paper provided a first-order examination of the usage of its critical resources (CPU, Memory, GPU, and I/O) over a 5-year production period (2015–2019). We showed quantitatively that the majority of CPU time was spent on the large-scale jobs, which is consistent with the policy of driving groundbreaking science through leadership computing. We also corroborated the general observation of the low CPU-memory usage, with 95% jobs using only 15% or less available memory. Additionally, we correlated the increase in total job submissions and the decrease of GPU-enabled jobs during 2016 with the GPU reliability issue, which impacted the large-scale runs. We further showed the surprising read/write ratio over the 5-year period, which contradicts the general mindset of the large-scale simulation machines being "write-heavy." This understanding will impact how we design our next-generation large-scale storage systems for Frontier and beyond.

Related paper: Feiyi Wang, Sarp Oral, Satyabrata Sen, and Neena Imam, "Learning from Five-year Resource-Utilization Data of Titan System," Workshop on Monitoring and Analysis for High Performance Computing Systems Plus Applications (HPCMASPA 2019), in conjunction with IEEE Cluster, Albuquerque, New Mexico, September 2019.

Real-Time Aggregation, Analysis, and Visualization of Summit's Temperature and Cooling Data. The OLCF has implemented a real-time data analytics system for Summit temperature control using near-real-time monitoring of temperature sensor data from Summit compute nodes, the cooling supply data from the cooling facility, and weather data from the National Oceanic and Atmospheric Administration (NOAA). The work is the result of a tight collaboration between OLCF and the vendors such as IBM, NVIDIA, and C-Tech. The overall system architecture is composed of the following: (1) a push model for the efficient and secure transport of sensor data to a message bus, (2) a central message bus (Apache Kafka) to disseminate the data streams to downstream subscribers, (3) real-time analytics to summarize the raw data streams, and (4) a lightweight database infrastructure to aggregate and serve the summary data for visualizations.

As an initial use-case of the system, a dashboard was developed to enable real-time correlation of various data streams such as per-node sensor data, job allocation data, cooling supply data, weather (wet-bulb temperature) data (Figure 4.1). Behind the scenes, several custom downstream analytics components were deployed as Kubernetes containers, consuming large amounts of sensor data in a pipelined fashion and processing the live stream data into a lightweight summary (i.e., time series of histograms of power and temperature). Such a lightweight summary of live streaming data was used to create visualizations consumed by several engineers using a web-based visualization system (Grafana). The time delay from sensor data changes to visualization of the data in the user's web browser is measured to be less than 15 sec. With such real-time data visualization capability, facility engineers could immediately verify the impact of cooling parameter changes made under events adverse to data center cooling (wet-bulb temperature increase, IT load increase, load spikes). This capability added confidence in adjusting the cooling parameters for the Summit system. In future iterations, this system is expected to be the foundation of an automated system that responds to adverse conditions in real time.

Figure 4.1. Real-time dashboard that aggregates job scheduler allocation data, sensor data from the HPC system Summit and the cooling facility.

Understanding the Capabilities of Summit's Network. OLCF's Summit and LLNL's Sierra have the latest generation of InfiniBand networks. Members of OLCF and LLNL's Livermore Computing (LC) explored several new features of these networks to determine the benefits to applications, if any. Both systems use second-generation EDR Host Channel Adapters (HCAs) and switches with several new features such as Adaptive Routing (AR), switch-based collectives, and HCA-based tag matching. Although based on the same components, Summit's network is "non-blocking" (i.e., a fully provisioned Clos network), and Sierra's network has a 2:1 taper between the racks and aggregation switches. The analysis evaluates the two systems' interconnects using traditional communication benchmarks as well as production applications. The analysis determined that the new AR implementation dramatically improves performance, but the other new features still need improvement. Members of NCCS did determine a new striping policy that dramatically improves bandwidth and complements the improved AR. The details of this effort were published and presented at the Supercomputing conference in November of 2019.

Related paper: Christopher J. Zimmer, Scott Atchley, Ramesh Pankajakshan, Brian E. Smith, Ian Karlin, Matthew L. Leininger, Adam Bertsch, Brian S. Ryujin, Jason Burmark, André Walker-Loud, M. A. Clark, Olga Pearce, "An Evaluation of the CORAL Interconnects," *Proceedings of Supercomputing 2019 (SC19): International Conference on High Performance Computing, Networking, Storage and Analysis*, Denver, CO, November 2019.

Workloads, Software Usage, and Programming Model with XALT. Understanding how our users use OLCF systems—including job characteristics, prevalent programming models, and center-provided software—is not only of natural interest to us but is also essential to both support the OLCF mission and to plan for future systems. For this purpose, OLCF deployed XALT on Summit when it became operational. XALT collects information on how an application is built and executed by transparently intercepting the linker and job launcher on the system. This information is eventually stored in a relational database system, allowing OLCF researchers to gain insights into application workloads on OLCF systems. OLCF staffer Reuben Budiardja and William Renaud developed new job launcher and linker plugins for XALT to work on Summit at the beginning of Summit operation. When the HPC Core Operations Group deployed Apache Kafka as the OLCF data pipeline, R. Budiardja and W. Renaud upgraded the XALT logging infrastructure to take advantage of the new capability. This upgrade also

built a foundation for future planned capability, including a near-real-time search and analysis for the streaming XALT data.

A few examples of XALT analysis from of the first nine months of Summit operational data. Top: programming model (left) and base languages (right) of INCITE projects weighted by runtime. Bottom: distribution of compiler usage on Summit by count.
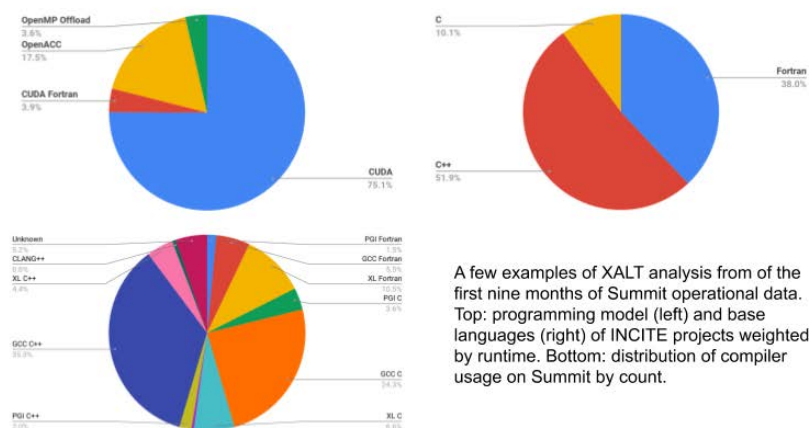
4.1.3 NCCS Paves the Way for Summit's Accreditation to Analyze Protected Information

An NCCS team designed "Citadel," a new framework for transferring encrypted personally identifiable information (PII) and personal health information (PHI) data from ORNL protected enclaves to the IBM AC922 Summit, where researchers can use the system's full leadership-class capabilities to securely learn from the data. Until now, PII and PHI data could only be analyzed within protected enclaves like ORNL's Knowledge Discovery Infrastructure (KDI) ecosystem, which follow strict privacy and cybersecurity regulations to ensure the data stays secure. Exploring PHI data on Summit will allow scientists to obtain results faster than previously possible and to interrogate the data at unprecedented scale. Smaller HPC systems can take months to analyze biological, biomedical, and genomic PHI, but Summit will cut that processing time to days, enabling new discoveries in areas ranging from cancer research to the prevention of opioid abuse and suicide.

To ensure that PHI data used for research at ORNL always remains secure, external assessors will review the hardware infrastructure, encryption practices, and other relevant components of Citadel. Formal accreditation is planned for early 2020. In the meantime, a publicly available medical dataset has been used to demonstrate how Citadel enables protected computing at significant scale on Summit while maintaining all security protections required by national security standards.

In a separate effort, NCCS is building a new, HPC-focused enclave for other types of protected information. Through Citadel or a similar framework, scientists will be able to conduct modeling and simulation experiments on Summit. The team also plans to obtain accreditation to study protected data on Frontier, ORNL's first exascale supercomputer, which is scheduled for delivery in 2021.

4.1.4 Slate: A Resource for Improved Workflow and Data Services Support

In 2019, the OLCF developed Slate, a resource that provides container orchestration services to OLCF users, allowing them to run specialized jobs such as workflows, databases, data portals, and continuous integration. Slate is integrated within the OLCF's existing HPC environment, including the Alpine file system and batch scheduler available inside of the container. Slate supports all containerized service workloads with Kubernetes.

The three production security enclaves at the OLCF accessible via Slate are Granite (core services), Marble (moderate-production enclave), and NCCS-Open (open enclave), where the Onyx and Ascent systems are located. Prior to Slate, users were taking advantage of SSH port forwarding or running persistent service on login or data transfer nodes. Users can now run their workloads in Slate at a high security level and orchestrate to other resources to avoid inefficiency and overutilization of nodes intended only for batch jobs or login.

A beta version of Slate is currently available to 21 pilot projects. The Big Production and Distributed Analysis, or BigPanDA, project used the tool in 2018 and 2019 to provide its PanDA workload management system as an interface to Titan. The ECP CANcer Distributed Learning Environment, or CANDLE, project is working toward using Marble to coordinate a workflow that includes both ML and MD. Staff at the JGI, a DOE Office of Science User Facility at LBNL, is working toward getting the JGI Analysis Workflow System running on Marble. Finally, a group at ORNL is using Slate to visualize monitoring and power metering. The full-production Slate service will be available to users beginning in 2020.

4.1.5   Vignettes of Application Performance Operational Improvements

We discuss here specific vignettes of application operation improvements by improving the user's interaction with their large-scale data.

Low-Overhead in Situ Visualization Using Halo Replay. A team lead by Duke University's Amanda Randles and OLCF computer scientist Benjamin Hernandez designed a scalable, low overhead, in situ recording scheme for rewind/playback of simulation states at full resolution. Usually, any in situ visualization processing impacts the simulation because it "fights" for computing and memory resources, but with this new method, the simulation runs on GPUs while the visualization runs on CPUs. The team demonstrated their technique using HARVEY, a 3D Lattice Boltzmann Method for simulation of hemodynamics in the human circulatory system. HARVEY simulations on the Summit supercomputer involve trillions of simulated points of complex vascular geometries and require petabytes of memory for system state at a given time step. Even with Summit's I/O bandwidth of 2.5 terabytes per second, post hoc analysis at this scale is infeasible without down-sampling, which increases the risk of missing data of clinical significance. However, with low overhead to the simulation through use of both Summit's CPUs and GPUs, the new method's process-independent replay enables in-transit workflows. The team used SENSEI framework and Paraview Catalyst to execute in situ workloads on Summit for up to 16,384 MPI and 131,072 OpenMP threads distributed across 1,024 nodes, consisting of 6,144 NVIDIA V100s GPUs and 43,008 IBM Power9 SMT4 cores.

Related paper: J. Ames, S. Rizzi, J. Insley, S. Patel, B. Hernández, and A. Randles, "Low-Overhead In Situ Visualization Using Halo Replay," *2019 IEEE 9th Symposium on Large Data Analysis and Visualization (LDAV)*, Vancouver, BC, Canada, 2019, pp. 16–26. doi: 10.1109/LDAV48142.2019.8944265.

A New Parallel Strategy for Tackling Turbulence on Summit. A team led by P. K. Yeung, professor of aerospace engineering and mechanical engineering at the Georgia Institute of Technology, performed direct numerical simulations (DNS) of turbulence using his team's new code, GPUs for Extreme-Scale Turbulence Simulations (GESTS). DNS can accurately capture the details that arise from a wide range of scales. Earlier this year, the team developed a new algorithm optimized for the IBM AC922 Summit supercomputer at the OLCF. With the new algorithm, the team reached a performance of less than 15 sec of wall-clock time per time step for more than 6 trillion grid points in space—a new world record surpassing the prior state of the art in the field for the size of the problem.

The simulations the team conducts on Summit are expected to clarify important issues regarding rapidly churning turbulent fluid flows, which will have a direct impact on the modeling of reacting flows in engines and other types of propulsion systems. An FFT is a math algorithm that computes the conversion of a signal (or a field) from its original time or space domain to a representation in the frequency (or wave number) space—and vice versa for the inverse transform. Yeung extensively applies a huge number of FFTs in accurately solving the fundamental partial differential equation of fluid dynamics, the Navier-Stokes equation, using an approach known in mathematics and scientific computing as "pseudospectral methods."

Most simulations using massive CPU-based parallelism will partition a 3D solution domain, or the volume of space where a fluid flow is computed, along two directions into many long "data boxes," or "pencils." However, when Yeung's team met at an OLCF GPU Hackathon in late 2017 with mentor David Appelhans, a research staff member at IBM, the group conceived of an innovative idea: They would combine two different approaches to tackle the problem. They would first partition the 3D domain in one direction, forming a number of data "slabs" on Summit's large-memory CPUs, and then further parallelize within each slab using Summit's GPUs.

The team identified the most time-intensive parts of a base CPU code and set out to design a new algorithm that would reduce the cost of these operations, push the limits of the largest problem size possible, and take advantage of the unique data-centric characteristics of Summit.

OLCF Team: Mike Matheson, David Pugmire

4.2 RESEARCH ACTIVITIES FOR NEXT GENERATION SYSTEMS

In this section, we discuss examples of research innovations, including intramural and external collaborations, that will impact next generation HPC platforms.

4.2.1 Research in Large-Scale Data Systems

We discuss innovative research efforts in managing scalable data including using ML to improve writes to a parallel file system, enabling a cross-facility scientific data management system, exploring burst-buffer placement design, and evaluating automatic data management for multi-tiered systems. Various aspects of this research also included external collaborators from Argonne National Laboratory, Lawrence Berkeley National Laboratory, and North Carolina State University.

Applying Machine Learning to Improve Parallel File System Writes. The OLCF, in collaboration with several other DOE laboratories, has applied ML techniques to analyze and improve the application I/O write performance on Titan at the OLCF and Cetus at the Argonne Leadership Computing Facility (ALCF). This work has been published in the Parallel Data Systems Workshop (PDSW 2019) that was held in conjunction with SC19, the International Conference for High Performance Computing, Networking, Storage and Analysis.

In HPC systems, I/O performance prediction offers the potential to improve the efficiency of jobs, according to this work. In particular, accurate prediction can make runtime estimates more precise, guide users toward optimal checkpoint strategies, and better inform facility provisioning and scheduling policies. HPC I/O performance is notoriously difficult to predict and model in large part because of inherent variability and a lack of transparency in the behaviors of constituent storage system components. The work advances the state of the art in HPC I/O performance prediction by (1) modeling the mean performance to address high variability, (2) deriving model features from write patterns, system architecture, and system configurations, and (3) employing Lasso regression model to improve model

accuracy. The work has demonstrated the efficacy of the approach by applying it to a crucial subset of common HPC I/O motifs, namely, file-per-process checkpoint write workloads. The experiments were conducted on Titan and Cetus to train and evaluate the models. The results convey that the relative error can be ≤ 30% relative error for 92.79% and 99.64% of the samples in the test set on these platforms, respectively.

Related paper: Bing Xie, Zilong Tan, Philip Carns, Jeff Chase, Kevin Harms, Jay Lofstead, Sarp Oral, Sudharshan S. Vazhkudai and Feiyi Wang, "Applying Machine Learning to Understand Write Performance of Large-scale Parallel File Systems," *Proceedings of the 4th International Parallel Data Systems Workshop (PDSW)*, Denver, CO, November 2019.
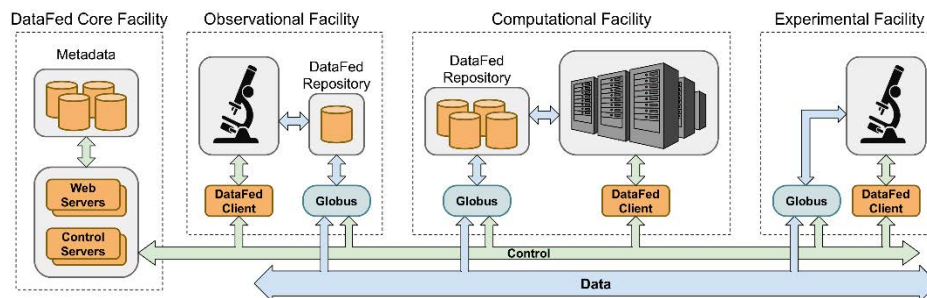
DataFed—A Scalable Cross-Facility Scientific Data Management System. The OLCF has developed and piloted DataFed (http://datafed.ornl.gov), a lightweight, distributed scientific data management system (SDMS). DataFed connects otherwise disjoint data repositories to form a federation of repositories to facilitate collaborative research across scientific facilities through seamless management and sharing of data, tracking of provenance, etc. Users can interactively manage their data through a modern web application as well as use DataFed in batch application in HPC contexts via a user-friendly command-line-interface. DataFed provides a logical view of data that abstracts the complexities and drawbacks of traditional file systems away from the user. DataFed uses proven and highly scalable technologies for data transfer and user identity management, which simplifies the administrative burdens for partner facilities while lowering the barrier for users to opt into DataFed. DataFed was designed with the Findable, Accessible, Interoperable, and Reusable (FAIR) data principles in mind to lower the burden for collaborative research while improving traceability and reproducibility in scientific research.

Related paper: Dale Stansberry, Suhas Somnath, Jessica Breet, Greg Shutt, and Arjun Shankar, "DataFed: Towards Reproducible Research via Federated Data Management," ACSE Computational Science and Intelligence, 2019.

Evaluating Burst Buffer Placement in HPC Systems, OLCF staff engaged in a research collaboration with Argonne National Laboratory and North Carolina State University to study the effective placement of burst buffers (BB) under various network topologies on HPC systems. Such a study will help the design and optimization of future systems. Burst buffers are increasingly exploited in contemporary supercomputers to bridge the performance gap between compute and storage systems. The design of BBs, particularly the placement of these devices and the underlying network topology, impacts both performance and cost. As the cost of other components such as memory and accelerators is increasing, it is becoming more important that HPC centers provision BBs tailored to their workloads. This work contributes a provisioning system to provide accurate, multi-tenant simulations that model realistic application and storage workloads from HPC systems. The framework aids HPC centers in modeling their workloads against multiple network and BB configurations rapidly. In experiments with the framework,

the team provided a comparison of representative OLCF I/O workloads against multiple BB designs. The study further analyzed the impact of these designs on latency, I/O phase lengths, contention for network and storage devices, and choice of network topology. The work was published in IEEE Cluster 2019 and received the Best Paper Award.

Related paper: Harsh Khetawat, Christopher Zimmer, Frank Mueller, Scott Atchley, Sudharshan S. Vazhkudai, Misbah Mubarak, "Evaluating Burst Buffer Placement in HPC Systems," *Proceedings of the IEEE Cluster 2019*, Albuquerque, NM, September 2019.

Data Jockey: Automatic Data Management for HPC Multi-Tiered Storage Systems. The OLCF engaged in a collaboration with LBNL to design and implementation Data Jockey, a data management system for HPC multi-tiered storage systems. As a centralized data management control plane, Data Jockey automates bulk data movement and placement for scientific workflows and integrates into existing HPC storage infrastructures. Data Jockey simplifies data management by eliminating human effort in programming complex data movements, laying datasets across multiple storage tiers when supporting complex workflows, which in turn increases the usability of multi-tiered storage systems emerging in modern HPC data centers. Specifically, Data Jockey presents a new data management scheme called "goal-driven data management" that can automatically infer low-level bulk data movement plans from declarative high-level goal statements that come from the lifetime of iterative runs of scientific workflows. While doing so, Data Jockey aims to minimize data wait times by taking responsibility for datasets that are unused or to be used and aggressively using the capacity of the upper, higher performant storage tiers. The team evaluated a prototype implementation of Data Jockey under a synthetic workload based on a year's worth of OLCF operational logs. The evaluations suggest that Data Jockey leads to higher utilization of the upper storage tiers while minimizing the programming effort of data movement compared with human involved, per-domain ad hoc data management scripts. The work was published in IEEE IPDPS in 2019.

Related paper: Woong Shin, Christopher D. Brumgard, Bing Xie, Sudharshan S. Vazhkudai, Devarshi Ghoshal, Sarp Oral, Lavanya Ramakrishnan, "Data Jockey: Automatic Data Management for HPC Multi-Tiered Storage Systems," *Proceedings of the 33rd IEEE International Parallel & Distributed Processing Symposium (IPDPS 2019)*, Rio de Janeiro, Brazil, May 2019.

4.2.2 Next Generation Hardware and System Architecture Exploration

We discuss here our research engagements in exploring next-generation hardware and systems including our PathForward engagements, evaluation of ARM systems and their GPU usage performance, projecting next-generation hardware enhancements on GPU-accelerated systems, and co-leading a multi-laboratory consortium exploring a DOE Future Laboratory Computing (FLC) environment. This work includes collaboration from CORAL2 and vendor partners, academic partners, and several DOE SC laboratory representatives in the FLC working group.

PathForward. The OLCF has worked closely with the DOE PathForward vendors to ensure that the vendors' research includes OLCF's use cases and requirements. OLCF staffer Scott Atchley has served as DOE's Technical Representative for both AMD's FastForward-2 Node Architecture contract and AMD's current PathForward contracts. The OLCF has worked closely with AMD to ensure that AMD's research would meet the needs of the CORAL-2 systems and the next generation of machines.

ARM System GPU Usage. The OLCF upgraded its Arm-based test bed Wombat, installing production Marvell ThunderX2 CPUs and NVIDIA V100 GPUs to test NVIDIA's new CUDA software stack purpose-built for Arm CPU systems. In late October, immediately after the upgrade, eight teams successfully ported their codes to the new system in the days leading up to SC19. In less than 2 weeks,

eight codes in a variety of scientific domains were running smoothly on Wombat. The undertaking represented the first implementation of an HPC server employing Arm CPUs and NVIDIA GPUs. The company Arm Holdings is not a manufacturer of hardware but rather licenses its technology to hardware vendors. It has a large stake in the mobile device market, but in recent years, its technology—often touted for its energy efficiency—has been adopted by the HPC community.

Although Arm has not traditionally been specifically focused on the HPC market until recently, the use of Arm-based technology in computing systems aligns with the DOE's goals for encouraging a diverse HPC environment. NVIDIA granted the OLCF early access to the CUDA software platform in addition to the architectural components for Arm. Aiming to explore an array of applications on the updated system, the OLCF deployed eight leadership-class scientific codes that were either bandwidth driven or memory driven to understand the implications of running codes involving either high amounts of data movement or many compute-intensive calculations. In addition to these production applications, OLCF staff successfully ran parallel programming models, community libraries, benchmarking suites, and mini applications on the system.

AHEAD: A Tool for Projecting Next-Generation Hardware Enhancements on GPU-Accelerated Systems. Starting with the Titan supercomputer at OLCF in 2012, top supercomputers have increasingly leveraged the performance of GPUs to support large-scale computational science. Many current (Summit, Sierra) and planned leadership systems (Frontier) are GPU-based machines. Accelerator-based architectures, however, add more complexity due to node heterogeneity. To inform procurement decisions, supercomputing centers need the tools to quickly model the impact of changes of the node architectures on application performance. The OLCF team, in collaboration with the University of British Columbia, developed AHEAD, a profiling and modeling tool to quantify the impact of an intra-node communication mechanism (e.g., PCI or NVLink) on application performance. Evaluations of the tool show average weighted relative errors of ∼19% and ∼23% for five CORAL-2 (a collaboration between multiple DOE labs to procure exascale systems) and 12 Rodinia benchmarks, respectively, without running the applications on the target future node. The work was presented in the APDCM workshop in 2019 and received the Best Paper Award.

Related paper: Hazem A. Abdelhafez, Christopher Zimmer, Sudharshan S. Vazhkudai, and Matei Ripeanu, "AHEAD: A Tool for Projecting Next-Generation Hardware Enhancements on GPU-Accelerated Systems," *Proceedings of the 21st Workshop on Advances in Parallel and Distributed Computational Models (APDCM 2019), co-located with IPDPS'19*, Rio de Janeiro, Brazil, May 2019.

Future Laboratory Computing Working Group (Multi-Lab): Distributed Computing and Data Ecosystem Pilot. The NCCS staff co-led and actively contributed to a multi-laboratory federated Distributed Computing and Data Ecosystem (DCDE) pilot that included tools, capabilities, services, and governance policies aimed at enabling researchers to seamlessly use a large variety of resources (i.e., scientific instruments, local clusters, large facilities, storage, enabling systems software and workflow, and networks) end-to-end across laboratories within the DOE environment. An ultimately successful DCDE would present to small teams or individual researchers a range of distributed resources through a coherent and simple set of access interfaces and allow them to establish and manage portable computational pipelines and the associated data lifecycle.

Related publications: E. Lancon and M. Shankar, "Background and Roadmap for a Distributed Computing and Data Ecosystem," ORNL/LTR-2019/1075 https://doi.org/10.2172/1528707.

Mallikarjun Shankar, Suhas Somnath, Sadaf Alam, Derek Feichtinger, Leonardo Sala, and Jack Wells, "Policy Considerations when Federating Facilities for Experimental and Observational Data Analysis,"

4.3   POSTDOCTORAL FELLOWS

4.3.1   Computational Scientists for Energy, the Environment, and National Security Postdoctoral Program

DOE recognizes the need to train and retain computational scientists in a broad range of disciplines that support DOE and the nation's critical mission needs to maintain the US competitive advantage in high-performance and data-intensive scientific computing. Considering the ever-increasing capability of high-end computer architectures, there is a continuing and increasing need to ensure a well-trained computational science workforce in academia and industry and at the national laboratories. In recognition of this need, DOE proposed that ASCR establish a postdoctoral training program at its user facilities, including the OLCF, ALCF, and NERSC, for future Computational Scientists for Energy, the Environment, and National Security (CSEEN). The objectives of this program are (1) to help ensure an adequate supply of scientists and engineers who are appropriately trained to meet national workforce needs, including those of DOE, for high-end computational science and engineering, with skills relevant to both exascale and data-intensive computing; (2) to make ASCR facilities available, through limited-term appointments, for applied work on authentic problems with highly productive work teams and increasingly cross-disciplinary training; and (3) to raise the visibility of careers in computational science and engineering to build the next generation of leaders in computational science. In CY2019, the OLCF began to leverage additional funding from the ECP to augment the CSEEN program with additional postdoctoral fellows.

The OLCF CSEEN postdoctoral program seeks to provide opportunities to bridge the experience gap between the need to address domain science challenges and the need to develop high-performance software development expertise. One of the focus areas is to provide the skills required to port, develop, and use software suites on the leadership computing resources at the OLCF. The software development activities occur in conjunction with a CAAR project (both OLCF-5 funded and ECP funded). This model offers the greatest potential for scientific breakthroughs through computing and provides ample opportunity to publish in domain scientific literature. This approach will ensure the postdoctoral trainees continue to build their reputations in their chosen science communities. Participants in the CSEEN postdoctoral program are encouraged to attend tutorials, training workshops, and training courses on select computer science topics. One of the most important outcomes for the postdoctoral trainee is the opportunity to publish and present research accomplishments. In CY2019, the CSEEN postdoctoral program at the OLCF supported nine trainees.

Ana Maria De Carvalho Vicente Da Cunha joined the Scientific Computing group in July 2018 after earning her PhD from the University of Groningen in theory and condensed matter physics with a focus on computational spectroscopy. She has experience in applying both classical molecular dynamics and quantum mechanics to the study of biomolecular systems. During her tenure, she studied the performance of various molecular dynamics software packages on OLCF production computing resources. Additionally, she explored whether current software can study the excited state chemistry of the vitamin B12. Ana accepted a position at SURFsara Computing Center in Amsterdam, the Netherlands, in December 2019.

Anikesh Pal joined the Scientific Computing group in June 2017 after working as a postdoctoral researcher at the University of California, Los Angeles. He received his PhD in mechanical engineering from the University of California, San Diego, and his master's degree in mechanical engineering from the Indian Institute of Technology Kanpur, India. Anikesh worked on the E3SM Multi-scale Modeling

Framework, which uses explicit cloud models to better represent the effects of unresolved clouds in the global model. During his time at ORNL, he created Deep Neural Network (DNN) representations for a portion of the model's radiation scheme to speed up the expensive calculations. Anikesh accepted a position as an Assistant Professor in Mechanical Engineering at the Indian Institute of Technology in Kanpur in July 2019.

Andreas Tillack joined the Scientific Computing group in October 2016. He received his PhD in chemistry from the University of Washington. Andreas holds a master's degree in physics from Humboldt University of Berlin (Germany). While at ORNL Andreas carried out a series of runs to profile and analyze QMCPACK in order to locate the performance bottlenecks. He implemented the "delayed update" scheme in QMCPACK to combine multiple BLAS-2 operations into a single BLAS-3 operation to improve the compute density on GPUs for better acceleration. Andreas also made modifications to QMCPACK to enable utilization of multiple GPUs on a Summit node. He implemented the "split-spline" scheme that distributes the large, read-only data (spline table that represents the wavefunction) over all the hierarchical CPU and GPU memory on a node. Andreas accepted a scientific programmer position at Scripps Research Institute in October 2019.

Antigoni Georgiadou joined the Scientific Computing group in September 2019 after earning her PhD in mathematics from Florida State University where she worked on optimization in stellar evolution applications. During her graduate studies, she was also a visiting scholar with the Theoretical Astrophysics Group and the Machine Intelligence and Reconstruction Group at Fermilab where she worked on an analysis to develop a statistical framework with Gaussian processes and ML techniques to optimize the input parameter space of cosmological simulations. Antigoni collaborates with the ECP ExaStar project, which builds on the current capabilities of astrophysics codes like FLASH for multi-physics astrophysics simulations run on exascale machines. The target science is to explain the origin of the elements via stellar explosion simulations and to study the conditions for nucleosynthesis in stars for nuclear data experiments. Antigoni will also collaborate with the ExaSky project, which aims to maximize the computing power for cosmological simulations.

Ashleigh Barnes joined the Scientific Computing group in August 2017 after earning her PhD in chemistry from the University of Tennessee. While at ORNL, Ashleigh participated in applied computational chemistry research and developed and optimized quantum chemistry software packages to efficiently use Summit resources. Her main contribution was programming new methodologies in linear-scaling Dalton electronic structure code, part of CAAR for Summit, where she was a co-author of the package. In particular, a multilayer divide-expand-consolidate (ML-DEC) scheme was implemented as well as multiple further contributions into GPU-accelerated Coupled Cluster module for calculating molecular properties. Ashleigh was the lead author on a paper describing ML-DEC (*J. Phys. Chem.* A 2019, 123, 40, 8734–8743) and delivered multiple talks on the subject of computational chemistry at scientific conferences. Ashleigh departed the group in August 2019 to pursue other opportunities.

Elvis Maradzike joined the Scientific Computing group in December 2019 after earning his PhD in chemistry from Florida State University. During his PhD, Elvis focused on developing approaches for ground and excited electronic states based on the variational two-electron reduced density matrix complete active space self-consistent field method. His work at the OLCF will be in collaboration with the NWChemEx ECP project in developing computational tools to accelerate electronic structure computations as well as modeling the physics of strongly correlated electrons.

Justin Lietz joined the Scientific Computing group in September 2019 after completing his PhD in nuclear physics from Michigan State University. Justin's research focuses on quantum many-body physics calculations of nuclear matter and electron gases. These calculations can be used to provide inputs for nuclear astrophysics simulations and as a theoretical study of extreme states of matter. He will be

developing high performance algorithms and data structures for many-body physics codes such as NUCCOR to enable novel nuclear structure calculations on the exascale generation of supercomputers such as Frontier.

Paul Mott joined the Scientific Computing group in December 2019 after earning his PhD in theoretical chemistry from the University of Tennessee. His research focused on implementing path integral quantum Monte Carlo methods to explore the role of zero-point motion and three-body interactions in the lattice dynamics of solid He-4 systems. While at ORNL, Paul will be collaborating with the GAMMESS ECP project to optimize their software for the upcoming Frontier exascale system.

Uma Klaasen joined the Scientific Computing group in May 2019. She received her PhD in computational science from the University of Texas at El Paso where she studied high performance scientific programming models for novel architectures. Uma's focus at the OLCF was to explore ML methods to accelerate large-scale scientific computing and reduce the computational effort required for scientific/computational physics calculations. She explored the use of ML to automatically detect anomalies that result from silent errors in physics simulations. Uma departed the group in November 2019 to pursue other opportunities.

5. RISK MANAGEMENT

CHARGE QUESTION 5: Is the facility effectively managing operational risks?

OLCF RESPONSE: Yes, the OLCF has a very successful history of anticipating, analyzing, rating, and retiring both project- and operations-based risks. The OLCF risk management approach is modeled after the Project Management Institute's best practices. Risks are tracked and, when appropriate, retired, reclassified, or mitigated. A change history is maintained for historical reference.

The major operational risks for the OLCF in CY 2019 are listed and described in this section. Planned mitigations and implementations are included in the subsequent descriptions. As of this writing, the OLCF has zero high-priority operational risks, but as the risk management approach is to continuously review and assess for new risks, that could change.

5.1 RISK MANAGEMENT SUMMARY

The OLCF's Risk Management Plan describes a regular, rigorous, proactive, and highly successful review process that is reviewed at least annually and updated as necessary. The plan covers both OLCF operations and its various projects (OLCF-5 during CY 2019). Each project execution plan refers to the main Risk Management Plan but may incorporate project-specific adjustments. Risks are tracked in a risk registry database application that is capable of tracking project and operational risks separately.

Weekly operations and project meetings are held, and risks are continually assessed and monitored. Specific risk meetings are held monthly for the projects and are attended by the federal project director, facility management, OLCF group leaders, subject matter experts, and risk owners. Operational risks are discussed in the weekly NCCS Operations Meeting attended by the risk owners, facility management team, OLCF group leaders, and other stakeholders. When assessing risks, the OLCF management team focuses its attention on the high and moderate risks as well as any low risks within the impact horizons associated with the risk. Trigger conditions and impact dates are recorded in the risk notes narrative section of the register. Risk owners are proactive in tracking trigger conditions and impact horizons for their risks and bringing appropriate management attention to those risks, regardless of the risk-rating level.

The OLCF reports a change summary of affected operations risks to the DOE program office as part of its monthly operations report. At the time of this writing, 24 active entries are in the OLCF operations risk register that fall into two categories: risks for the entire facility and risks for a specific portion of the facility. Facility-wide risks are concerned with such issues as safety, funding, expenditures, and staffing. The specific, more focused risks are concerned with reliability, availability, and use of the system or its components (e.g., the computing platforms, power and cooling infrastructure, storage, networks, software, and user support). In addition to operational risks, at the time of this report, there are 83 tracked risks for the OLCF-5 project.

The costs of handling risks are integrated in the budgeting exercises for the entire facility. For operations, the costs of risk mitigation are accepted, and residual risk values are estimated by expert opinion and are

accommodated, as much as possible, in management reserves. This reserve is continually reevaluated throughout the year.

5.2 MAJOR RISKS TRACKED IN 2019

Table 5.1 contains the major risks tracked for OLCF operations in 2019. The full OLCF operations risk register is available on request. The selected risks are all rated medium or high in impact.

Table 5.1. 2019 OLCF major risks.

Risk ID/description	Probability/impact	Action	Status
406: System cyber security failures	Low/high	Mitigating	The OLCF continues to see a rise in the quantity of cyber-security attacks against the computer resources. This increase does not directly correlate to higher success rates because the OLCF employs various techniques to repel these attacks, such as proactive patching for zero-day exploits, formal review of cyber security plans, a two-factor authentication requirement for system access, and a multifactor authentication (MFA) level 4 requirement for privileged access to OLCF resources.
723: Safety - personal injury	Medium/medium	Accept	Reduce risk by monitoring worker compliance with existing safety requirements, daily toolbox safety meetings, periodic surveillances using independent safety professionals, joint walk-downs by management and work supervisors and encouraging stop-work authority of all personnel. Observations from safety walk- downs are evaluated for unsatisfactory trends, e.g., recurring unsafe acts. Unsatisfactory performance will receive prompt management intervention commensurate with the severity of the safety deficiencies.
917: Robust support will not be available to ensure portability of restructured applications	Medium/medium	Mitigating	Multiple instantiations of compiler infrastructure tools will be adopted to maximize the exposure of multiple levels of concurrency in the applications. Work with vendors continues to improve compiler technology and other tools.
1006: Inability to acquire sufficient staff	Medium/low	Accept	The OLCF reduced the probability of encountering this risk to medium in 2015 and has maintained that rating through aggressive hiring and extensive succession planning. The number of open positions has been lower than the threshold determined to trigger this risk (10%). Succession candidates have been identified for several key positions including the NCCS division director, NCCS director of science, and technology integration group leader.

Table 5.1. 2019 OLCF major risks (continued).

Risk ID/description	Probability/impact	Action	Status
1063: Programming environment and tools may be inadequate for future architectures	Medium/medium	Mitigating	We will continue to engage with users, standards organizations (i.e., OpenMP, OpenACC, and others), and tool and hardware vendors to encourage, facilitate, and enact responses to user feedback and anticipate architectural trends in key standards and tools.
1142: OLCF cost increases because fewer computer room customers to distribute maintenance and operation costs	Low/high	Mitigating	In 2019, the data center customer base remained static, and slightly fewer systems were deployed. In 2019, the 5600 E102 data center was removed from operations to prepare the room for Frontier's arrival in 2021. As 2020 begins, a new customer is coming online in the 5600 K100 data center, and the 5600 E102 will remain under construction for the entire calendar year.
1145: Changes from external project managers cause development impacts to HPSS	Medium/medium	Mitigating	IBM has continued to push for items that are not on the development roadmap to support requests of potential customers as well as for features that may not meet the Technical Committee's release schedule.
1240: Failure to handle Export Control Information (ECI) properly	Low/high	Mitigating	Staff with elevated privileges on systems where ECI can be accessed go through annual training and refreshers on how to handle ECI. Project PI's and members participate in an initial project briefing where an ORNL Export Control Analyst describes the categorization of the project based on the project application. OLCF Cyber Security staff contribute to this briefing and outline what storage resources are capable of handling ECI.
1323: Lack of adequate facilities for the next OLCF system	Low/high	Accept	ORNL/OLCF has a plan to house the follow-on system to OLCF-5 in Building 5600. However, the much-preferred approach would be to construct a new building that is designed for meeting the needs of the program well into the future from the beginning. The Office of Management and Budget (OMB) has rejected third-party financing as a method of building such a facility, so this would need congressional line-item funding.
1245: System unavailability due to mechanical/electrical system failure	Low/high	Mitigating	System was designed with leak detection from the start; sensors have been triggered during preventative maintenance activities and also indicated issues when a new cabinet was brought in and the cooling water connection was out of alignment. Perform all preventative maintenance activities; perform inspections and monitoring where possible.

5.3 NEW OR RECHARACTERIZED RISKS SINCE LAST REVIEW

5.3.1 Recharacterized Risks

The status or impact of the following risks changed during the reporting period.

Risk No. 723	Safety - Personal injury		
Risk owner	James P. Abston		
Status	Mitigating—Current		
Probability	Low → Medium		
Impact	**Cost:** Low	**Schedule:** Medium	**Scope/Tech:** Low
Mitigations	Reduce risk by monitoring worker compliance with existing safety requirements, daily toolbox safety meetings, periodic surveillances using independent safety professionals, joint walk-downs by project management and work supervisors, and by encouraging stop-work authority of all personnel. Observations from safety walk-downs will be evaluated for unsatisfactory trends, e.g., recurring unsafe acts. Unsatisfactory performance will receive prompt management intervention commensurate with the severity of the safety deficiencies. The probability increased in the assessment of this risk due to the significant amount of construction activities taking place in and around Building 5600 that impact not only the OLCF but a sizeable population of the laboratory's staff as well.		
Triggers	Any safety incident or safety observation is taken very seriously by ORNL management and likely will trigger additional, but as yet undetermined, mitigation activities.		

5.3.2 New Risks in This Reporting Period

The following risks were created and tracked during CY 2019. They are included with their risk creation date, mitigations, and triggers.

Risk No. 1323	Lack of facilities for next OLCF system		
Risk owner	James H. Rogers		
Status	Accept—Current		
Probability	Medium → Low		
Impact	**Cost:** High	**Schedule:** High	**Scope/Tech:** Medium
Mitigations	ORNL/OLCF has a plan to house the follow-on system to OLCF-5 in Building 5600, but today the requirements for that system are largely unknown. This risk replaced risk 948, which was specific to OLCF-5. The intent of risk 948 remains but is generalized to reflect the siting of the "next" OLCF system.		
Triggers	Intelligence on the size and power requirements of proposed systems.		

5.4 RISKS RETIRED DURING THE CURRENT YEAR

Risk No. 359	Lustre updates may cause file system to become unstable at larger scales		
Risk owner	Sudharshan S. Vazhkudai		
Status	Retired		
Retirement Comment	With the decommissioning of the Spider-2 (Atlas) filesystem, this risk is no longer impacting the operations of the OLCF.		
Impact	**Cost:** Low	**Schedule:** Medium	**Scope/Tech:** Low

Mitigations	Periodic large-scale testing of Lustre releases will help build confidence in the performance and stability prior to production release.
Triggers	This risk was triggered several times during the operational lifetime of the Spider-1 (decommissioned in 2012) and Spider-2 (decommissioned in 2019).

Risk No. 361	Scientists decline to port to heterogeneous architecture
Risk owner	Jack C. Wells
Status	Retired
Retirement Comment	This risk was not realized. Retirement was decided on due to the architectures of the ASCR machines in the short term all requiring an accelerated programming model to make effective use of resources soon to be deployed. Applications to INCITE and ALCC have not declined, and many INCITE and ALCC projects are very effectively using the GPUs on Titan and now Summit.
Impact	**Cost:** Medium **Schedule:** Low **Scope/Tech:** Low
Mitigations	One mitigation the OLCF undertook for this risk during the period of impact was to be a strong supporter of the OpenACC standards body. The OLCF provided a strong outreach and training program that included GPU Hackathons that continue today. The OLCF also granted significant discretionary resources over the period of impact for this risk, allowing scientists to port codes and perform performance enhancements to their codes on Titan.
Triggers	A decrease in the number and/or quality of proposals submitted to "headline" user programs, e.g., INCITE.

Risk No. 721	Lustre metadata performance continues to impact applications
Risk owner	Sudharshan S. Vazhkudai
Status	Retired
Retirement Comment	With the decommissioning of the Spider-2 (Atlas) filesystem, this risk is no longer impacting the operations of the OLCF.
Impact	**Cost:** Low **Schedule:** Low **Scope/Tech:** Medium
Mitigations	The OLCF worked with Lustre developers on the implementation of the Distributed Namespace Environment (DNE) feature as a key mitigation to metadata operations impacting application performance. The OLCF also deployed significantly more powerful Lustre Metadata Servers (MDS) and Metadata Targets (MDT) when Spider-2 was deployed and saw that the metadata performance in most applications was sufficient for scientists to perform their work.
Triggers	This risk was triggered several times during the operational lifetime of the Spider-1 (decommissioned in 2012) and Spider-2 (decommissioned in 2019).

Risk No. 948	Lack of adequate facilities for exascale system
Risk owner	James H. Rogers
Status	Retired
Retirement Comment	With the approval of CD2/3 by DOE, the facility plan was sufficient to meet the mission need of the OLCF-5 system delivered in 2021.
Impact	**Cost:** High **Schedule:** High **Scope/Tech:** Medium
Response	In accepting this risk, the OLCF was required to tailor the design of the electrical and mechanical infrastructure to retrofit into Building 5600's room E102 computer room. This caused several design challenges that were met by the facility design team prior to CD2/3 approval. Additionally, the intent of this risk (lack of facilities for the "next" system) was captured moving forward in Risk ID 1323.

Risk No. 1144	Increased soft error rates and silent data corruption with 14 nm and 16 nm processes
Risk owner	James H. Rogers
Status	Retired
Retirement Comment	After over 1 year of operations, no issues have been discovered in the CPU or GPU. This risk is ready for retirement.
Impact	**Cost:** Low **Schedule:** Low **Scope/Tech:** Low
Response	The process of accepting a machine of Summit's caliber is rigorous and has been shown to discover issues like silent data corruption in the past. Additionally, the period of Early Science allows for friendly users to provide additional time and stressors to the system.

Risk No. 1154	Lack of available spare parts as Titan ages
Risk owner	Don E. Maxwell
Status	Retired
Retirement Comment	With the decommissioning of Titan, this risk is no longer impacting the operations of the OLCF.
Impact	**Cost:** Low **Schedule:** Low **Scope/Tech:** Medium
Response	In responding to this risk, the OLCF undertook several mitigations. First, the OLCF worked closely with Cray and NVIDIA to understand the issue that was causing GPU failures in Titan and also developed a failure prediction model. Once this was well understood, the OLCF, Cray, and NVIDIA implemented a plan where a batch of GPUs were manufactured and replaced the components that were high on the failure prediction model, stockpiled these parts as spares, and continued operation. This plan resulted in a significant stockpile of parts to replace GPUs that did not show up on the failure analysis and was sufficient to complete operations of Titan with GPUs still in the parts spare pool.
Risk No. 1155	Extended life of atlas file system causes increase in parts failures, possible downtime, and possible loss of data
Risk owner	Kevin G. Thach
Status	Retired
Retirement Comment	With the decommissioning of the Spider-2 (Atlas) filesystem, this risk is no longer impacting the operations of the OLCF.
Impact	**Cost:** Medium **Schedule:** Low **Scope/Tech:** Medium
Response	The OLCF executed contract modifications for support of the storage system (Disk Arrays for Object Storage and Metadata Storage, Lustre OSS, MDS, and the Infiniband network) to meet the system's extended operational lifetime of the end of CY 2019.

Risk No. 1157	Limited or incomplete information relative to system load and incomplete integration with the control systems will contribute to higher operations costs.
Risk owner	James H. Rogers
Status	Retired
Retirement Comment	This risk was developed to cover the first full year of operations. The information coming from the Mechanical and Electrical Plant (MEP) is being used to tune the controls systems, and the operators of those controls systems feel they have sufficient information to make those decisions.
Impact	**Cost:** Medium **Schedule:** Low **Scope/Tech:** Low
Response	Identify nominal operating conditions that can be correlated to MEP flow rates; adjust control system to use these as operation basis; consider mechanisms for introducing additional functionality over time, ensuring that computer systems remain adequately protected but seeking greater MEP efficiency.

Risk No. 1193	Early Science Program for Summit terminated before completion
Risk owner	Jack C. Wells
Status	Retired
Retirement Comment	The Early Science Period overlapped with the first 6 months of the INCITE 2019 program and was completed as projected in June 2019. This risk is retired.
Impact	**Cost:** Low **Schedule:** Low **Scope/Tech:** Medium
Response	Continual evaluation of budget scenarios and discussions with program sponsors. Review of annual user survey results and operational assessments of key user program delivery metrics.

Risk No. 1225	20 MW MEP Programmable Logic Controller controls implementation requires significant improvement from baseline/project implementation
Risk owner	James H. Rogers
Status	Retired
Retirement Comment	This risk was developed to cover the first full year of operations. The controls for the MEP have proven to be operating the system efficiently through the varying conditions of a calendar year in East Tennessee.
Impact	**Cost:** Low **Schedule:** Low **Scope/Tech:** Low
Response	Tightly integrated activity between ORNL Mechanical Utilities (MU), the PLC programming subcontractor, the Facilities Division, and OLCF to monitor the performance of the PLC as operating conditions become more broadly understood.

5.5 MAJOR RISKS FOR NEXT YEAR

Many risks were retired as the Titan system was decommissioned. Summit's operations are critical to the success of the OLCF in 2020, and one risk directly impacts Summit's ability to operate: Risk ID 1245—System unavailability due to mechanical/electrical system failure.

Significant construction activities are also taking place in and around Building 5600. These activities require closure of some areas and ingress/egress paths for staff and visitors. These closures require increased focus on safety in and around the work areas by all staff (Risk ID 723).

Finally, the programming environment for the OLCF-5 system could be different enough that experiences on Summit may not be sufficient to prepare for Frontier (Risk ID 1063). The OLCF is working closely with IBM, the HPC community and standards bodies, and Cray to deploy tools and versions of the programming environment and compilers to make the transition as smooth as possible.

5.6 RISKS THAT OCCURRED DURING THE CURRENT YEAR AND THE EFFECTIVENESS OF THEIR MITIGATION

The following risks were encountered and effectively mitigated in 2019. A short summary of the status and impact of the risk on the operations of the OLCF is included.

Risk No. 407	Loss of key personnel
Risk owner	Arthur S. Bland
Status	Mitigate
Probability	Medium
Impact	**Cost:** Low **Schedule:** Low **Scope/Tech:** Low

Trigger Event	In December 2019, the position of NCCS Division Director transitioned from James Hack to Gina Tourassi. Also, in December 2019, the NCCS Director of Science and Operations Manager transitioned into new roles within the laboratory.
Mitigations	Increase recruiting and hiring in some key areas to maintain a qualified workforce and provide for the planned level of effort. Succession plans for key staff are continually worked. A backup person is at least partially trained in the skills needed to complete the work to minimize the impact of a loss. Careful planning along with its documentation in as much detail as possible helps to mitigate the impact of a loss. OLCF career development programs and maintaining sufficient depth in staffing should allow for uninterrupted progress should this risk be encountered.
Triggers	Intelligence on potential or actual loss of key personnel.

Risk No. 1079	OLCF-4 post deployment issues		
Risk owner	Don E. Maxwell		
Status	Mitigate		
Probability	Medium		
Impact	**Cost:** Low	**Schedule:** Low	**Scope/Tech:** Low
Trigger Event	In 2019, a set of users experienced issues with Summit nodes when their codes were running. After a full root cause analysis with IBM, it was determined that a defect exists in the power supply for the AC922 nodes that is only exacerbated by fast/high frequency transient current. IBM has developed a strategy to replace all the power supplies in Summit that will be enacted in CY 2020.		
Mitigations	Working closely with the vendor to track and identify root cause of every failure as quickly as possible.		
Triggers	Problems encountered during early science and ongoing operations.		

6. ENVIRONMENT SAFETY AND HEALTH

CHARGE QUESTION 6: Does the facility exhibit a culture of continuous improvement in Environment, Safety, and Health (ES&H) practices to benefit staff, users, the public, and the environment? Has the facility implemented appropriate Environment, Safety, and Health measures?

OLCF RESPONSE: Yes

ORNL is committed to operating under the DOE safety regulations specified in 10 CFR 851, which outlines the requirements for a worker safety/health program to ensure DOE contractors and their workers operate a safe workplace. Additionally, 10 CFR 851 establishes procedures for investigating if a violation of a requirement has occurred, for determining the nature and extent of any such violation, and for imposing an appropriate remedy. These safety requirements are incorporated into ORNL contracts as required compliance documents. To implement these safety requirements in a consistent manner across ORNL, UT-Battelle LLC deploys an online procedure management system referred to as the Standards-Based Management System (SBMS). Within SBMS, there are work control requirements that describe the processes to be used in ORNL operations and R&D activities to implement integrated safety management functions and principles.

A key feature of the Integrated Safety Management (ISM) process is the development and implementation of specific work control. Research work in the OLCF is controlled by research safety summaries (RSSs), which define the scope of work, identify and analyze hazards, and establish safety procedures. Each RSS is reviewed and approved by line managers, qualified safety and health professionals, and research staff. An RSS provides the means by which ORNL management and staff plan and conduct research in a safe manner. It is used to control work, train participants, and provide information about operations and emergency services if needed. In addition to RSSs, ISM also requires work control for maintenance and non-employee. Maintenance work in the OLCF is conducted under a work plan that is developed by Facilities and Operations Directorate (F&O) line management and reviewed by subject matter experts as required. Work plans are also written before maintenance work can proceed, to ensure work is conducted safely. Work by non-employees or subcontractor/vendors is performed in accordance with a hazard analysis. The subcontractor/vendor hazard analysis is a requirement that is included in the contract specific language. The following highlights provide additional information regarding the subcontractor hazard analysis process.

Safety assessments are conducted for RSSs, work plans, and subcontracts, as well as inspections of job sites throughout each year. Lessons learned, safety snapshots, safety talks, and management assessments are conducted and recorded in the Assessment and Commitment Tracking System. The tracking system documents the completion of the ORNL ISM process and provides a means for analysis. The DOE ORNL Site Office participates in field implementation and documentation of all operational safety reviews and partners with the ORNL Offices of Institutional Planning and Integrated Performance Management and the Safety Services Division on independent safety management system assessments. The culture of safety at ORNL is reflected in these processes, which seeks to reduce and prevent injuries to personnel and potential exposure to hazards associated with operation of the facility.

Operations of the OLCF in the NCCS remained safe, efficient, and effective as there were zero first aid cases, zero total recordable cases, and zero Days Away Restricted or Transferred (DARTs) in FY 2019.

The following activities are ES&H highlights from CY 2019.

- A good portion of 2019 was spent planning, preparing, and moving people and systems out of the first-floor offices and data center to facilitate OLCF-5 construction. Titan, Eos, and the Spider II storage system as well as associated support systems were safely removed by the manufacturer and shipped from ORNL. Summit Development, Tundra, and the Tech-Int Test Bed were safely relocated to other OLCF spaces and reconnected. The National Oceanic and Atmospheric Administration (NOAA) computers C3 and C4 and the NOAA storage system F2 were prepared for moving in October 2019. This effort also required significant infrastructure upgrades to be performed by third-party personnel and internal F&O staff in the new spaces for the equipment.

- OLCF worked with the F&O staff to develop a cart for moving cabinets that contain hardware that could be sensitive to shock. Two carts with shock-sensitive wheels were developed. Multiple design attempts were made until a final design was agreed upon. The cart was then used to safely relocate cabinets from E102 without removing cabinet hardware. Many of these cabinets weighed in excess of 3000 lb and contained spinning storage disks.

- A new tape storage library was safely installed by the manufacturer in room E-204. The effort included infrastructure upgrades, assembly, testing, and installation of the fire suppression system.

- The OLCF developed a center-specific hazard analysis for OLCF subcontractor/vendors. This analysis process was reviewed by ORNL and replaces the requirement that a subcontractor submit a hazard analysis. The OLCF hazard analysis provides the same requirements to all subcontractor/vendors and is included as part of a subcontract. The subcontractor is asked to review and comment on or agree to follow the requirements of the hazard analysis. This process makes the subcontractor a part owner of the process. However, all changes are made by the OLCF data center manager. Once the parties agree on the requirements, the subcontractor is asked to concur with the hazard analysis, and all subcontractor and sub-tiered subcontractors are briefed on the requirements prior to the start of work. This process resolves a noted weakness regarding a subcontractor performing a proper hazard analysis and provides consistency in the OLCF centers.

- The OLCF-5 project Health and Safety Plan and the Hazard Analysis were reviewed and revised in 2019.

- Annual occupational exposure monitoring was conducted for noise in all three OLCF data centers. The results of these surveys are documented in the ORNL Comprehensive Tracking System (CTS).

- Occupational exposure monitoring was conducted as requested by office occupants in their offices for electromagnetic fields and noise.

- The Authorized Access to ORNL Computing Centers access training was reviewed, and slight corrections were made to reflect trainee comments or fix inconsistencies.

- The Computational Sciences Building Computer Center Operations Emergency Response Plan was updated to include emergency response specification for power loss in K100, Remote Distribution Facility (RDF), and K200.

Security

7. SECURITY

CHARGE QUESTION 7: (a) Does the Facility exhibit a culture of continual improvement in cyber security practices? (b) Does the Facility have a valid cyber security plan and Authority to Operate? (c) Does the Facility have effective processes for compliance with applicable national security policies related to Export Controls and foreign visitor access?

OLCF RESPONSE: Yes. The OLCF maintains a strong culture of continuous operational improvement, especially in cybersecurity. The most recent OLCF Authority to Operate was granted on February 21, 2017 and is managed through an ongoing authorization process; no authorization termination date is set (Figure 7.1). The technical staff members track and monitor for existing threats and vulnerabilities to assess the risk profile of the OLCF operation. The Facility is committed to innovating in this area by developing open-source tools and employing cutting-edge practices that enhance the operation without increasing the OLCF's risk profile. The OLCF employs ORNL policies related to Export Control and foreign visitor access.

7.1 SUMMARY

All information technology systems operating for the federal government must have certification and accreditation to operate. This involves developing and obtaining approval for a policy and implementing a continuous monitoring program to confirm that the policy is effectively implemented. The ORNL certification and accreditation package currently uses the National Institute of Standards and Technology Special Publication 800-53, revision 4, *Security and Privacy Controls for Federal Information Systems and Organizations*, and the US Department of Commerce Joint Task Force Transformation Initiative (August 2009) as guidelines for security controls. The OLCF has determined that the highest classification of data is moderate based on the guidelines for information classification in the Federal Information Processing Standards Publication 199, *Standards for Security Categorization of Federal Information and Information Systems, Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology* The OLCF is accredited at the moderate level of controls for protecting the confidentiality and integrity of user and system information, which authorizes the Facility to process sensitive, proprietary, and export-controlled data.

In the future, cybersecurity planning will become more complex as the center continues its mission to produce great science. The Facility is very proactive, viewing its cybersecurity plans as dynamic documentation to which it will preemptively respond and modify as its needs you change to provide an appropriately secure environment. The OLCF will abide by the Health Insurance Portability and Accountability Act (HIPAA) Privacy and Security Rule to provide supercomputing resources to projects containing PHI, as well as the International Traffic and Arms Regulations (ITAR) for projects containing that type of sensitive information.

Department of Energy
ORNL Site Office
P.O. Box 2008
Oak Ridge, Tennessee 37831-6269

March 22, 2017

Mr. Kevin A. Kerr
Information Systems Security Manager
Oak Ridge National Laboratory
UT-Battelle, LLC
Post Office Box 2008
Oak Ridge, Tennessee 37831-6045

Dear Mr. Kerr:

AUTHORIZATION DECISION DOCUMENT FOR OAK RIDGE NATIONAL LABORATORY (ORNL) SUPERCOMPUTING ENCLAVE

Reference: Letter from Kevin A. Kerr to Johnny O. Moore, subject, *Contract DE-AC05-00OR22725, ORNL Supercomputing Enclave Approval to Operate*, dated February 21, 2017

As the Authorizing Official, I have reviewed the referenced request. The ORNL Supercomputing Enclave is authorized to operate. No additional conditions outside the substance of the request are required.

The information system is now being managed by an ongoing authorization process, thus an authorization termination date is not set. I accept the responsibility for performing all necessary activities associated with the ongoing authorization process.

If there are any questions or additional information is required, please contact John Young at (865) 576-7471 or youngjc1@ornl.gov.

Sincerely,

Johnny O. Moore, Manager
ORNL Site Office

Enclosure

cc w/enclosure:
Mike E. Bartell, ORNL
Amy D. Nuckols, ORNL
Neil Masincupp, SC-OR
Martha J. Kass, SC-OSO
John C. Young, SC-OSO

Document/Material Transmitted Contains Official Use Only Information.

When separated from enclosure, this document **does not** contain Official Use Only.

Figure 7.1. OLCF Authority to Operate.

7.2 COPACETIC

The OLCF has many business values, including operational expectations, security requirements, and specific policies instituted at various times to streamline OLCF operations. Because OLCF is growing in complexity, these business values are difficult to track and prioritize when drift occurs, or business rules change over time. Copacetic enables OLCF to define and evaluate complex business logic that automatically detects undesired inconsistencies between systems, services, and people. Business rule violation notifications are delivered to staff members who are authorized to remedy those violations directly instead of being passed through teams that do not have the authority to implement remediation actions.

Copacetic:

1. Enables complex OLCF business logic (including security and operational rules) to be evaluated in real time or near real time

2. Delivers results of real-time query evaluation to relevant owners through existing channels

3. Publishes data to the STREAM data bus

4. Provides easy-to-understand dashboards for decision making and prioritization of business logic violations

7.3 WORDPRESS VULNERABILITY SCANNER (WPSCAN)

OLCF has implemented a periodic scan of the public-facing WordPress websites that are primarily used for interfacing with users. WordPress is often a target for malicious activity because of the wide range of third-party plugins, themes, and features, and it frequently has vulnerabilities. The WordPress Vulnerability Scanner (WPScan) software is a black box vulnerability scanner that can enumerate plugins and themes and compare them with known vulnerable versions. The OLCF has created a scheduled container that periodically starts up, updates the vulnerability database, and scans the website for these potential vulnerabilities. The results are sent to the OLCF Security Information and Event Management (SIEM) tool for further correlation, as well as directly to the security team for evaluation. Once the timeline for patching/remediation is decided based on the risk matrix in the Patching and Vulnerability Management procedure, the site is patched or mitigated by the responsible team within the required time.

7.4 PUBLIC KEY PASSWORD MANAGER (PKPASS)

Password management for large teams and groups can be difficult, especially for root accounts, service accounts, or other shared-password accounts. The OLCF developed PKPass, an application for storing, retrieving, sharing, and recovering passwords that are used across the environment. PKPass was developed in Python and uses the certificates from the personal identity verification (PIV) badges to store encrypted passwords in the repository. Because PKPass can access the public keys for all system administrators in the environment, administrators can share passwords with other administrators by encrypting the passwords with the other administrator's public key, which places it in a repository for the administrator to retrieve later with their private key and PIV badge. If an administrator has a password that may not be shared but has been "escrowed," then the password may be recovered by two or three administrators that have been designated as custodians for recovering the password.

7.5 MULTIFACTOR AUTHENTICATION AUTOMATION

DOE has adopted and required MFA for privileged access to systems using a standard Homeland Security Presidential Directive 12 (HSPD-12) PIV badge. To enable this throughout the environment, the OLCF uses card readers, policies, and technical implementations to remain compliant. Although the initial implementation was compliant, it was difficult to maintain and require a complex and complicated badge enrollment process that requires and enables PIV badge access to all privileged accounts. To streamline this effort, the OLCF designed a system that automates credential enrollment, ensures validity throughout the lifecycle of those certificates, and manages identity mappings with external identity sources. This automation reduces overall risk, eliminates the enrollment process, and simplifies the configuration of the systems across the environment.

7.6 OLCF USER VETTING

The OLCF follows a set of rigorous controls for vetting user access to ensure compliance with export-control regulations and foreign visitor access policies.

7.6.1 OLCF Projects

Users must be added to an approved OLCF project to obtain access to OLCF resources. An ORNL export control officer reviews the scope of work for all OLCF user projects to determine if there are any export-control restrictions to which the OLCF must adhere and to place an internal designation of category 1 or category 2 on each project. These categories then drive the business processes that is followed for each applicant.

Category designation	Category description	PI actions before project activation
Category 1	The category 1 rating is applied if the project is open fundamental research that does not involve proprietary input and/or output, sensitive data, and/or export-control restrictions above EAR99.	• Sign OLCF PI agreement
Category 2	The category 2 rating is applied if the project involves proprietary input and/or output, sensitive subject areas, and/or export-control restrictions above EAR99 but below ITAR.	• Sign OLCF PI agreement • Participate in mandatory security call to review risks/restrictions associated with category 2 projects

Sensitive information, including proprietary and export-controlled information, is segregated and protected in the specific project area to protect it from unauthorized access, and specific storage rules and requirements are relayed to the PI and individual project users to further prevent information mishandling. If a project is rated category 2 or above, then the project PI must participate in a mandatory security call with OLCF's cybersecurity team to review the risks and restrictions before the project is enabled. Once the security call is complete and all other project requirements are met, the project is enabled in the OLCF RATS and labeled with the appropriate category, and any export-control restrictions are added to the project.

7.6.2 OLCF Users

All users requesting access to OLCF resources are required to fill out the OLCF account application form and provide the project identification and PI for the project they are requesting to join. Based on the category of the project designated in RATS, the following requirements must be met before the user is added to the project and provided access to OLCF resources.

Project category	PI approval	ORNL Personnel Access System (PAS)* or Restricted Party Screening (RPS)*	User agreement*	Sensitive data rules*	Level 2 identity proofing*

Category 1	The PI must approve all user account requests to access to their project	• Approved PAS is required for applicants born in, residing in, or citizens of:
		  o China
		  o Russia
		  o Iran
		  o Sudan
		  o Syria
		  o Crimea
		  o Cuba
		  o North Korea
		• All other applicants and their institutions go through RPS screening
			Must have valid user agreement	N/A	Required

Category 2	The PI must approve all user account requests to access to their project	• Approved PAS is required for all applicants that are not US citizens or LPRs unless they reside in one of the countries listed above.	Must have valid user agreement	Must return signed sensitive data rules	Required

PAS: the system ORNL uses to process on-site and/or remote access for foreign nationals and nonemployees.

RPS: ORNL maintains a subscription to the Descartes Visual Compliance tool, which is used to look up applicants and their institutions that do not require PAS approval. If any hits are found on the user or the user's institution, the information is turned over to the export control officer. The officer then works with the Counterintelligence office to look at the applicant or institution in more detail and informs OLCF if it is acceptable to proceed.

user agreement (UA): serves as the "master" agreement that establishes the general terms and conditions, including disposition of intellectual property, for work at any of ORNL's user facilities. A UA must be executed with each user's institution.

sensitive data rules: this form contains the user acknowledgment, which documents that users on a category 2 project are aware of the risks and rules for accessing the sensitive project.

Level 2 Identity Proofing: the OLCF users' RSA SecurID tokens for authenticating to OLCF moderate resources. Level 2 Identity Proofing of all applicants is required as part of the NCCS moderate Certification and Accreditation (C&A). To achieve Level 2 identity proofing, the applicant must have their identity and RSA SecurID token verified by a notary or an OLCF designated registration authority. The token is not enabled until all steps above are completed, including the return of the original notarized form.

HIGH PERFORMANCE COMPUTING FACILITY
2019 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
March 2020

8. STRATEGIC RESULTS

CHARGE QUESTION 8: (a) Are the methods and processes for monitoring scientific accomplishments effective? (b) Has the Facility demonstrated effective engagements with strategic stakeholders (i.e., beyond the user population)? (c) Is the Facility operating in a manner that enables delivery of facility mission and Department of Energy mission including maintaining a vibrant US effort in science and engineering?

OLCF RESPONSE: Yes. OLCF projects and user programs are advancing DOE's mission to ensure America's security and prosperity by addressing its energy, environmental, and nuclear challenges through transformative science and technology solutions. The selected accomplishments described in this section serve to highlight how the OLCF is advancing two strategic objectives of DOE's Strategic Plan Goal 1, "Science and Energy: Advance foundational science, innovate energy technologies, and inform data driven policies that enhance economic growth and job creation, energy security, and environmental quality . . . ," as stated in the *US Department of Energy Strategic Plan: 2014–2018* (March 2014):

- Strategic Objective 2—Support a more economically competitive, environmentally responsible, secure and resilient US energy infrastructure

- Strategic Objective 3—Deliver the scientific discoveries and major scientific tools that transform our understanding of nature and strengthen the connection between advances in fundamental science and technology innovation

8.1 2019 OPERATIONAL ASSESSMENT GUIDANCE

The Facility collects and reports annually the number of refereed publications resulting (at least in part) from use of the Facility's resources. For the Leadership Computing Facilities (LCFs), tracking is done for a period of 5 years following the project's use of the Facility. This number may include publications in press or accepted but not submitted or in preparation. This is a reported number, not a metric. In addition, the Facility may report other publications where appropriate.

8.1.1 OLCF Publications Report

In 2019, 405 publications resulting from the use of OLCF resources were published, based on a data collection completed on January 28, 2020. In this document, "year" refers to the calendar year unless it carries the prefix "FY", indicating the fiscal year. In the 2018 OLCF OAR, 469 publications were reported. A list of 2014–2017 publications is available on the OLCF website (https://www.olcf.ornl.gov/leadership-science/publications/). Sponsor guidance allows accepted and in press publications to be reported, but the OLCF only reports publications appearing in print in the year under review. However, the OLCF continues to search for publications after the OAR is submitted to DOE each year, and the number of publications shown in previous OARs is updated in the current report. Table 8.1 provides the updated, verified, and validated publications count for the 2012–2018 period, showing continued growth in both the total publications count and the number of publications in journals with high impact factors. Though the total number of publications for CY 2019 reported at this time is

down slightly from the reported number at roughly the same epoch in CY 2018, it is expected that the total for CY2019 will continue to grow for several months, as publications are discovered. In addition to this caveat, the loss of Titan as a unique and productive scientific instrument in June 2019 impacted the total number of publications in CY 2019. The success of OLCF users in using Titan for more than a half a decade is evident in the trend in total publication counts since 2012.

Table 8.1 Summary of unique OLCF publications for 2012–2018.

Year	Unique, confirmed OLCF publications	High-impact publications with JIF* >10
2019	426	19
2018	494	20
2017	472	27
2016	459	33
2015	356	21
2014	297	16
2013	359	9
2012	334	20

*JIF = Journal impact factor

8.2   SCIENTIFIC ACCOMPLISHMENTS

The OLCF advances DOE's science and engineering enterprise by fostering robust scientific engagement with its users through the INCITE liaison program, the user assistance program, and the OLCF DD program outreach. The following subsections provide brief summaries of select scientific and engineering accomplishments, as well as resources for obtaining additional information. While they cannot capture the full scope and scale of achievements enabled by the OLCF in 2019, these accomplishments advance the state of the art in science and engineering R&D across diverse disciplines and are advancing DOE's science programs toward their targeted outcomes and mission goals. As an additional indication of the breadth of these achievements, OLCF users published many breakthrough publications in high-impact journals in 2019, as shown in Table 8.2.

Altogether in 2019, OLCF users published 46 papers in journals with a journal impact factor (JIF) of greater than 7 and 19 papers with a JIF greater than 10.

71

Table 8.2. Publications in high-impact journals in 2018.

Journal	Number of publications
Nature	1
Nature Physics	3
Nature Communications	5
Nature Structural & Molecular Biology	1
Nature Plants	1
Nature Materials	1
Physical Review X	1
Journal of the American Chemical Society	2
Cell	1
Energy & Environmental Science	1
Nano Letters	1
Trends in Biotechnology	1

Summit Charts a Course to Uncover the Origins of Genetic Diseases

Researchers create the most complete model yet of complex protein machinery

PI: Ivaylo Ivanov, Georgia State University
Allocation Program: INCITE

The Science

Genetic factors can be responsible for a number of diseases, from degenerative neurological disorders to some cancers. Gene mutations can interfere with how the body expresses genes and cause disease. To better understand this connection, researchers recently developed a model of the transcription preinitiation complex (PIC). The PIC is a group of proteins vital to gene expression. The PIC translates genetic information from DNA to produce proteins and other functional molecules. The simulations revealed how the PIC is organized. It showed that mutations tend to cluster at a specific group of proteins called transcription factor II human (TFIIH).

The Impact

This model could help make sense of the relationship between a patient's unique genetic makeup and the development of a disease. These findings provided insights into three distinct genetic disorders associated with cancer, aging, and developmental defects. This study provides a foundation for future experimental and computational efforts to delve deeper into the mechanisms of gene expression. Future studies may be able to pinpoint the mutations that cause genetic diseases. Understanding the underlying causes of these diseases could help scientists develop more effective treatments.

Summary

Previous attempts to characterize the PIC have been limited by incomplete models. The team developed the most complete model of the PIC to date. To create this new version, the researchers combined data from cryo-electron microscopy—a structural biology method that uses an electron beam to study

cryogenically frozen protein samples—and large-scale molecular dynamics simulations on Summit, the world's smartest and most powerful supercomputer. This model provides superior insights into the structural organizations of these proteins, which transcribe genes and repair DNA. Because the biochemical pathways responsible for gene expression and repair are intricately intertwined, obtaining detailed information regarding the molecular mechanism behind transcription, the first step of gene expression, is crucial to advancing biomedical applications. The researchers have mainly studied Pol II, an enzyme that transcribes protein-coding genes into messenger RNA molecules that mediate protein synthesis. However, they plan to expand their project to investigate the functional dynamics of Pol I and Pol III, which are known along with Pol II as RNA polymerases, to pursue more groundbreaking insights.

Funding
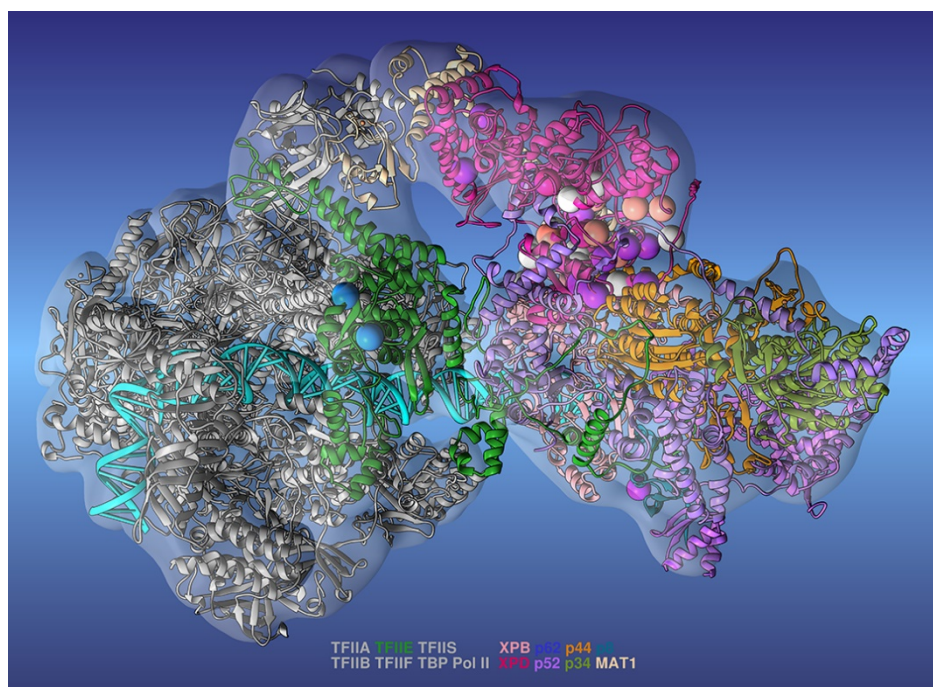
Figure 8.1. Researchers used the new model to accurately identify clusters of gene mutations (spheres), which helped them study the emergence of various genetic diseases. Image credit: Georgia State University.

Publication

Chunli Yan, Thomas Dodd, Yuan He, John A. Tainer, Susan E. Tsutakawa, and Ivaylo Ivanov, "Transcription Preinitiation Complex Structure and Dynamics Provide Insight into Genetic Diseases." *Nature Structural & Molecular Biology*, 26, 397 (2019). [DOI: 10.1038/s41594-019-0220-3]

Related Links

"Summit Charts a Course to Uncover the Origins of Genetic Diseases," OLCF News (May 20, 2019)

******

Summit Supercomputer Simulates How Humans Will "Brake" during Mars Landing

PI: Eric Nielsen and Ashley Korzun, NASA Langley Research Center
Allocation Program: Early Science

The Science

NASA expects humans to voyage to Mars by the mid to late 2030s, so engineers have been at the drafting board for some time. They have a promising solution in retropropulsion, or engine-powered deceleration. Led by Eric Nielsen, a senior research scientist at NASA Langley, a team of scientists and engineers is using the IBM AC922 Summit supercomputer to simulate retropropulsion for landing humans on Mars. On Summit, the team is modeling the lander at multiple points in its 6 to 7-minute descent. To characterize the flow behaviors across speeds ranging from supersonic to subsonic, researchers run ensembles (suites of individual simulations) to resolve fluid dynamics at a resolution of up to 10 billion elements with as much as 200 terabytes of information stored per run.

The Impact

To predict what will happen in the Martian atmosphere and how the engines should be designed and controlled for the crew's success and safety, researchers need to investigate unsteady and turbulent flows across length and time scales—from centimeters to kilometers and from fractions of a second to minutes. To accurately replicate these faraway conditions, scientists must model the large dimensions of the lander and its engines, the local atmospheric conditions, and the conditions of the engines along the descent trajectory.

Summary

A team at NASA Langley uses its CFD code called FUN3D to model a space vehicle's Martian descent. CFD applications use large systems of equations to simulate the small-scale interactions of fluids (including gases) during flow and turbulence—in this case, to capture the aerodynamic effects created by the landing vehicle and the atmosphere. Nielsen's team spent several years optimizing FUN3D—a code that has advanced aerodynamic modeling for several decades—for new GPU technology using CUDA, a programming platform that serves as an intermediary between GPUs and traditional programming languages like C++. By leveraging the speed of Summit's GPUs, Nielsen's team reports a 35× increase in performance per compute node. As the team members continue to collect new Summit data, they are thinking about the next steps to designing a human exploration vehicle for Mars.

Figure 8.2. Snapshot of total temperature distribution at supersonic speed of Mach 2.4. Total temperature allows the team to visualize the extent of the exhaust plumes, as the temperature of the plumes is much greater than that of the surrounding atmosphere. Image Credit: NASA

Funding

This research used resources of the OLCF located at ORNL, which is supported by the DOE Office of Science.

Publications

A. M. Korzun, K. T. Edquist, A. M. Dwyer Cianciolo, and R. A. Lugo, "Design Considerations and Development Status for Atmospheric Powered Descent of High-Mass Payloads at Mars," 2019 International Astronautical Congress, Washington, DC, October 2019.

Related Links

"Summit Simulates How Humans Will 'Brake' during Mars Landing," OLCF News (October 10, 2019)

******

Quantum Processors Are Now Challenging Conventional HPC Systems

PI: Travis Humble and Dmitry Lyakh, ORNL
Allocation: INCITE

The Science

To prove quantum supremacy, a joint research team from Google Inc., NASA Ames Research Center, and ORNL needed to rule out that classical supercomputers could perform computational tasks at the same

speed as Google's Sycamore quantum computer. The quantum supremacy test, which was run on a 53 qubit quantum computer, sampled a random quantum circuit with a depth of 20. To prove quantum supremacy, however, the team needed to simulate that same quantum circuit on Summit, a heterogeneous GPU-accelerated supercomputer. The team was able to determine that the computation that took Google's Sycamore quantum computer only 200 seconds would have taken Summit 10,000 years to complete with current state-of-the-art algorithms.

Impact

Testing quantum supremacy requires a well-defined computational problem, a benchmark task that can be run on both quantum and classical computers, and to then observe which solves it faster. This historic result will open the door to answers to a new array of questions for computer scientists. The next milestone is to build a quantum computer that is large enough to begin to solve problems of practical interest the same way Summit currently does but even faster.

Summary

A joint research team came to the IBM AC922 Summit to test quantum supremacy. The test, which scientists call random circuit sampling (RCS), was written in a code optimized to run on regular CPU-only computer clusters. The OLCF's Dmitry Lyakh sprang into action, adapting the classical RCS code—originally developed by Google and NASA and dubbed qFlex—to be executed on Summit's multi-GPU heterogeneous nodes. Adapting the code for Summit required offloading all numerical tensor algebra computations programmed in qFlex to Summit's GPUs, which was done using a software library developed by Lyakh called TAL-SH.

The library was developed by the OLCF at the Summit Center for Accelerated Application Readiness, sponsored by DOE's ASCR program, and was later extended for use as a computational back end for the qFlex code. This highly optimized library unleashed enormous computing power, delivered by Summit's NVIDIA Volta GPUs, to perform the computationally intensive tensor operations required for a classical simulation of the RCS circuits with the qFlex code.

Funding
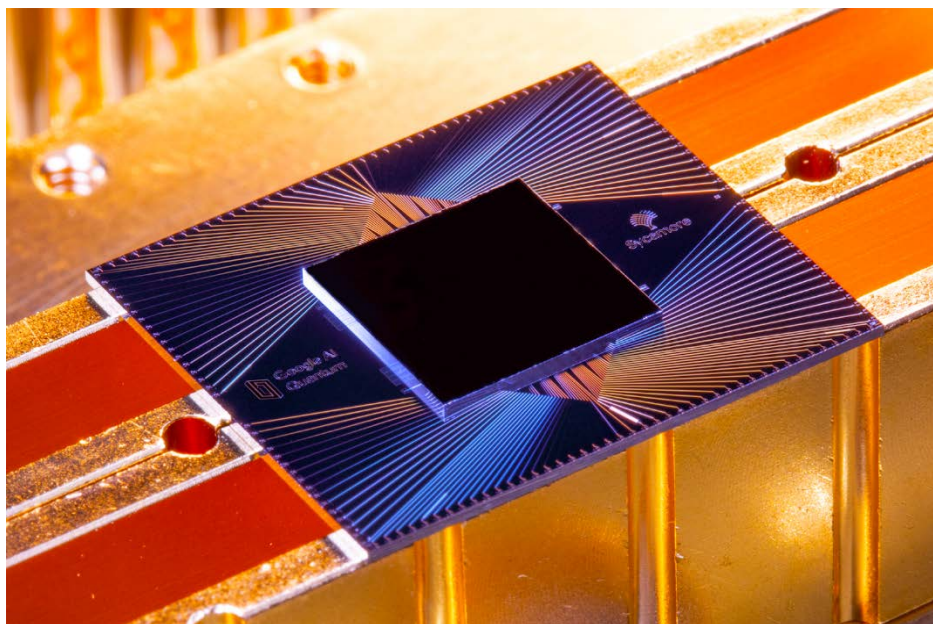
Figure 8.3. The Sycamore Quantum Processor. Image Credit: Erik Lucero/Google

Publication

F. Arute et al., "Quantum Supremacy Using a Programmable Superconducting Processor." *Nature* 574 (2019), doi:10.1038/s41586-019-1666-5.

Related Links

"Quantum Processors are Now Challenging Conventional HPC Systems," OLCF News (October 23, 2019)

******

Gordon Bell Finalist Team Tackles Transistors with New Programming Paradigm

PI: Torsten Hoefler, ETH Zürich
Allocation Program: Director's Discretionary

The Science

Although today's electronics are incredibly compact, they don't come without challenges. As electrons flow through transistors, they generate heat that dissipates into the environment around them. And as transistors get smaller, the density of the heat they dissipate gets bigger. To better understand this problem, a team led by Torsten Hoefler at ETH Zurich performed a 10,000-atom simulation of a 2D slice of a transistor on the Summit supercomputer and developed a map of where heat is produced in a single transistor. Using a new Data-Centric (DaCe) version of the OMEN nanodevice simulator, the team reached a sustained performance of 85.45 petaflops for double precision and 90.89 petaflops for mixed precision and won the 2019 Association for Computing Machinery Gordon Bell Prize.

Impact

The results could be used to inform the production of new semiconductors with optimal heat-evacuating properties. With the new code, the team demonstrated a successful new programming model: the DaCe framework, which was developed at ETH. Traditional programming requires line-by-line modification to change the code, but DaCe provides a visual representation of data movement, allowing a programmer to interact with this representation to optimize the code more easily. The team believes the new programming model could change how people program. In fact, the concepts could benefit countless scientific domains—not just nanoelectronics. The team also hopes semiconductor companies will use the code to design better and more efficient transistors.

Summary

To understand heat dissipation in transistors, the team used DaCe OMEN, which models quantum transport of electric charge carriers, to simulate a 10,000-atom system 14 times faster than the speed at which they previously could simulate a 1,000 atom system. The team completed the simulation in less than 8 minutes, and the speedup made DaCe OMEN two orders of magnitude faster per atom than the original OMEN code. To maximize their use of the system, the researchers also optimized a piece of DaCe OMEN for mixed-precision calculations—an unusual move for a code that depends on precision and accuracy—and reached a sustained performance of 90.89 petaflops. Mixed-precision calculations are less precise, but the researchers used them for a small portion of their code to demonstrate their ability to take full advantage of Summit's architecture, including the Tensor Cores on its GPUs.

One of the team's most exciting accomplishments is its use of a new programming paradigm based on the way data flows in a simulation. The paradigm is a fundamental departure from everything that's been done before in code optimization, and the team is hoping to develop the next generation of parallel programming capabilities and techniques. Team member Mathieu Luisier is currently working with companies that may be interested in selling software based on the algorithms used in the project. The team wants to be ready to quickly model designs for companies to help them make better transistors.
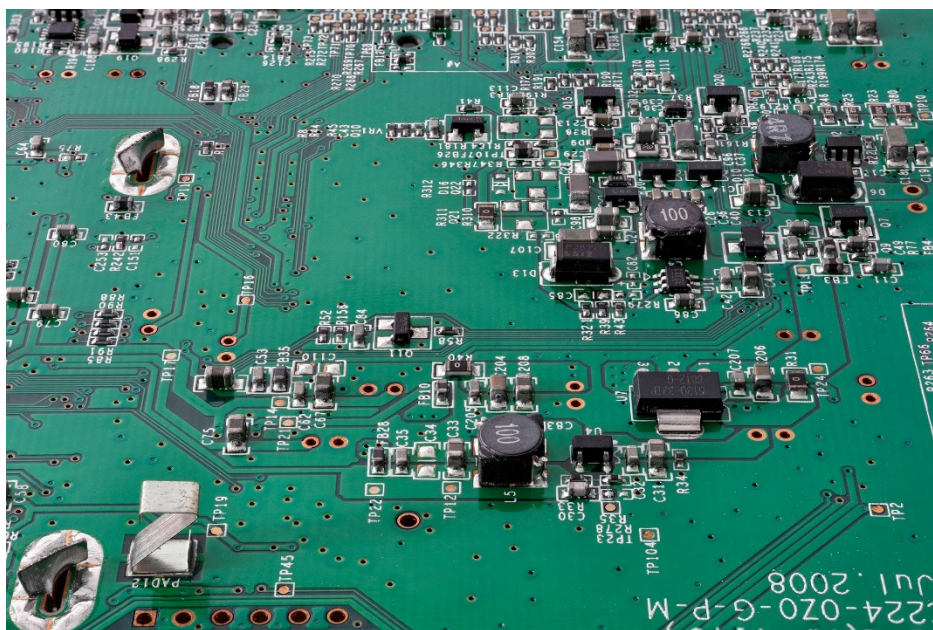
Figure 8.4. A stock image of an integrated circuit representing the fact that results from Hoeffler's research could be used to inform the production of new semiconductors with optimal heat-evacuating properties.
Image Credit: Pixabay.

Publication

A. Ziogas, T. Ben-Nun, G. Indalecio, T. Schneider, M. Luisier, and T. Hoefler, "A Data-Centric Approach to Extreme-Scale Ab Initio Dissipative Quantum Transport Simulations," SC19 Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis, Denver, CO, November 17–22, 2019.

Related Links

"Tiny Transistor Leads to Big Win for ETH Zurich, 2019 ACM Gordon Bell Prize Winner," OLCF News (November 21, 2019)

"Gordon Bell Finalist Team Tackles Transistors with New Programming Paradigm," OLCF News (October 28, 2019)

******

Speeding Toward the Future of Fusion

Researchers accelerate plasma turbulence simulations on Oak Ridge supercomputers to improve fusion design models

PI: Christopher Holland, University of California, San Diego
Allocation Program: INCITE, ALCC

The Science

If fusion power plants become a reality, they could provide nearly inexhaustible energy using fuel derived from seawater. But fusion has some stellar challenges to overcome first. Because of the extreme and

remote conditions inside fusion reactors, plasma behavior is difficult to study experimentally, and scientists often must fuse experiment with computational simulations to understand fusion processes. A team of scientists—including Christopher Holland of the University of California, San Diego, Jeff Candy of General Atomics, and Nathan Howard of MIT—is using the Summit supercomputer at the OLCF to better understand turbulence, an important characteristic of plasma behavior that affects performance in fusion devices such as the ITER fusion facility. The team optimized the CGYRO code for Summit's V100 GPUs and is now running the code six to eight times faster than on the previous Titan supercomputer at the OLCF, gaining the ability to simulate enough cases to make rigorous comparisons with experiment.

The Impact

To make fusion energy a reality, researchers must be able to understand and predict levels of heat transport observed in an experiment so they can develop computational models that inform the design of future fusion devices and predict their performance. Computational simulations enable researchers to overcome some experimental barriers created by the extreme environment in a fusion device. For instance, not only is observing turbulence at the scale of particles impossible in an experiment, but there will also be key differences between current experiments like DIII-D and future devices like ITER, such as experiment duration, power, and size. Once built, ITER will be the world's largest fusion reactor, with a plasma volume 10 times larger than that of any fusion device today. Supercomputers like Summit provide the hundreds of thousands of processing cores needed to include all relevant time and spatial scales.

Summary

Today, the world's largest fusion experiment is being built by seven international members, including the United States. The ITER fusion facility is expected to produce 10 times more power than the thermal power required to heat the plasma inside, thereby demonstrating the feasibility of commercial-scale fusion power. ITER will use superconducting magnets to confine plasma in its tokamak, a donut-shaped vessel with a design that allows magnetic field lines to run in two directions, long and short, through the plasma. Fluctuations in ion and electron speed and energy, however, result in turbulence that can rapidly transport heat away from the plasma center, reducing the number of fusion reactions that occur. Turbulence at one scale can inhibit or enhance turbulent fluctuations on other scales, impacting heat transport and, therefore, fusion performance.

Holland and his team must perform plasma turbulence simulations that capture wavelengths at the ion scale as well as at the electron scale—which is 60 times smaller—to explain the levels of heat loss observed in experiment. The team analyzes experimental data from the DIII-D National Fusion Facility tokamak, operated by General Atomics as a national user facility for DOE's Office of Science, and carries out unprecedented simulations with their new CGYRO gyrokinetic code on Summit. After optimizing their code for Summit, the team has gained a six- to eight-fold speedup, which will allow them to compare observational data from the DIII-D tokamak to experiment, demonstrating the impact of HPC on understanding and improving fusion performance in future reactors.

Funding
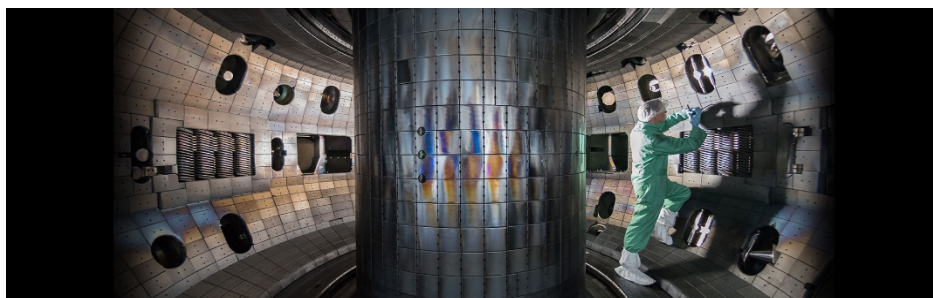
Figure 8.5. The DIII-D National Fusion Facility, operated by General Atomics for the DOE, is the largest magnetic fusion research facility operating in the United States. Image credit: General Atomics.

Publication

J. Candy, I. Sfiligoi, E. A. Belli, K. Hallatschek, C. Holland, N. T. Howard, and E. D'Azevedo, "Multiscale-Optimized Plasma Turbulence Simulation on Petascale Architectures," *Computers and Fluids*, 188, 25 (2019), doi:10.1016/j.compfluid.2019.04.016.

Related Links

"Speeding Toward the Future of Fusion," OLCF News (January 2, 2020)

8.2.1   INCITE 2019 Allocation/Utilization

The INCITE allocation year is January 1–December 30. In 2019, several INCITE projects were allocated on Titan only, while the bulk of the allocations spanned both Titan and Summit during the allocation year. Usage for both platforms is shown in Figures 8.6 and 8.7, with the allocated time on each machine treated separately. Note the change in allocation units moving from Titan to Summit (i.e. Titan core-hours versus Summit node-hours).
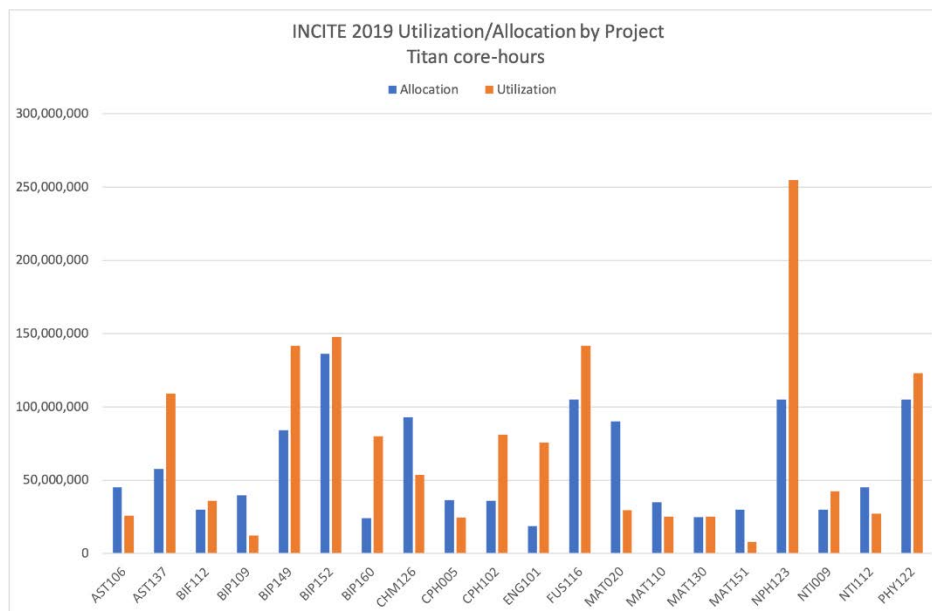
Figure 8.6. INCITE allocation by project on Titan.
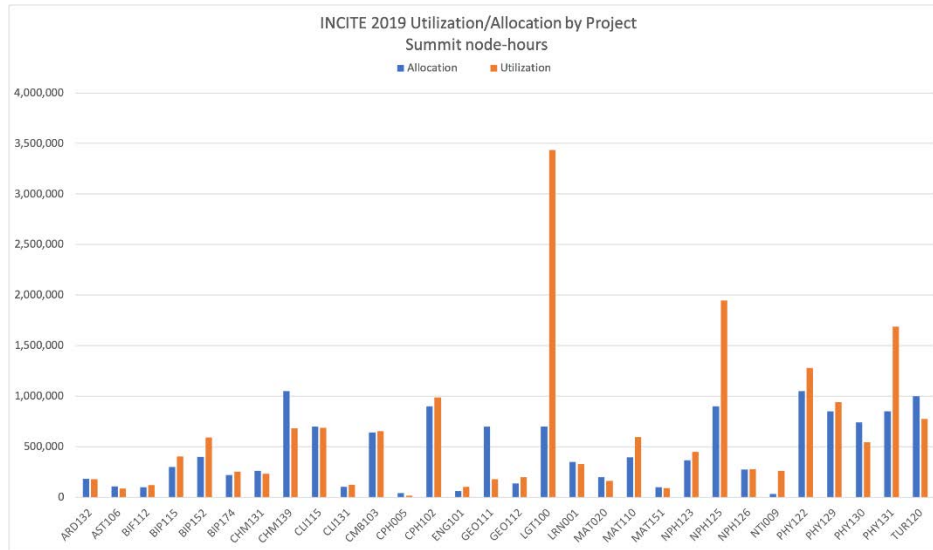
Figure 8.7. INCITE allocation by project on Summit.

8.2.2 ALCC Allocation/Utilization for Calendar Year 2018

The ALCC allocation year is July 1–June 30. Therefore, in Figures 8.8 and 8.9 the OLCF reports the usage against the ALCC 2018 and ALCC 2019 allocations separately. Usage for both programs is reported against the full allocation amount for each allocation year. The ALCC allocations for 2019 were on Titan, while the ALCC 2020 allocations were provided on Summit. As in the case of INCITE, the allocation units, therefore, differ for each program.
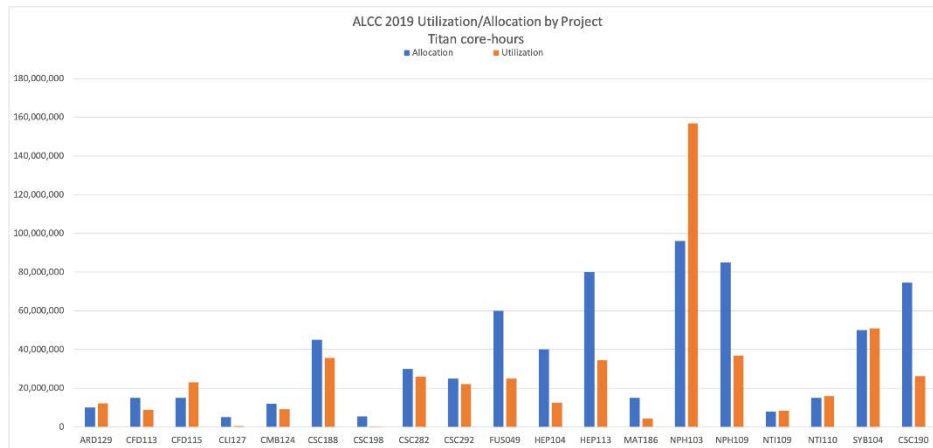
Figure 8.8. ALCC allocation by project on Titan.
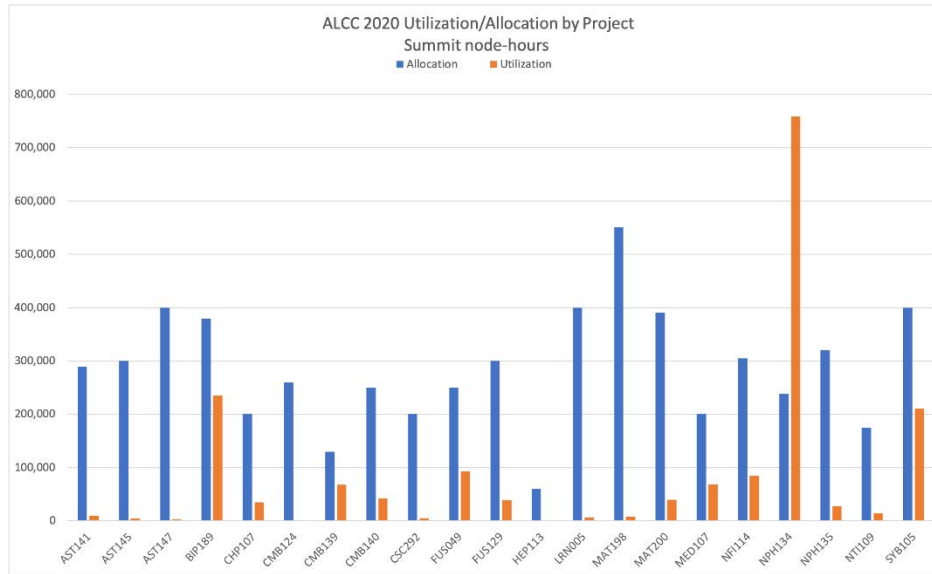
Figure 8.9. ALCC allocation by project on Summit.

8.3 STAKEHOLDER ENGAGEMENT AND OUTREACH

8.3.1 Community Engagement

A primary and natural place for community engagement has been DOE's Exascale Computing Project (ECP), whose goal is to develop software and applications and influence the development of hardware technology, all to facilitate the successful deployment and operation of capable exascale systems. The ECP continues to fund efforts at national labs, academia, and industry with the expressed goal of producing usable software and applications and influencing the development of hardware technology for the exascale systems in the 2021–2022 time frame. These investments are very timely and will significantly aid the OLCF in delivering capable exascale systems with robust system software and application software that can address the science gaps immediately upon delivery and acceptance of the systems. The OLCF's engagement with ECP includes three primary thrust areas, described below, as well as many pairwise and other interactions [e.g., staff involvement in ECP Application Development (AD) and Software Technology (ST) projects].

OLCF-ECP Application Development/Software Technology engagements

The OLCF has a long history of readying applications for its forthcoming architectures, dating back to before the delivery of the OLCF-3 system (Titan), then with OLCF-4 (Summit), and now with OLCF-5 (Frontier). The OLCF Center for Accelerated Application Readiness (CAAR) has served as a successful collaboration point for application teams, vendors, and tool developers to exploit hierarchical parallelism within applications in preparation for next-generation architectures. The OLCF has chosen to partner with the ECP to augment the OLCF CAAR portfolio with an additional 12 ECP AD teams. These teams were selected to diversify OLCF applications readiness efforts funded through the OLCF-5 CAAR such that the OLCF will have a broad suite of applications ready to use Frontier. These projects were also selected with an eye toward matching the architectural strengths of Frontier with the appropriate computational motifs and methods employed by these ECP applications. The teams that are currently partnering with the OLCF and their OLCF Scientific Computing liaison are listed in Table 8.3.

Table 8.3. Listing of OLCF ECP engagement applications, the ECP AD PI, and the OLCF Scientific Computing liaison.[a]

WBS / ECP AD Project	Project PI	OLCF Liaison
2.2.1.01 LatticeQCD	Andreas Kronfeld (FNAL)	TBD
2.2.1.02 NWChemEx	Theresa Windus (Ames Laboratory)	Dmitry Liakh
2.2.1.03 GAMESS	Mark Gordon (Iowa State University)	Dmytro Bykov
2.2.1.05 ExaAM	John Turner (ORNL)	Stephen Nichols
2.2.2.02 Combustion-PELE	Jackie Chen (Sandia)	Ronnie Chatterjee
2.2.2.03 ExaSMR	Steven Hamilton (ORNL)	Mark Berrill
2.2.2.05 WDMApp	Amitava Bhattacharjee (PPPL)	Ed D'Azevedo
2.2.3.01 ExaStar	Dan Kasen (LBNL)	Austin Harris
2.2.3.02 ExaSky	Salman Habib (ANL)	Bronson Messer
2.2.3.05 E3SM-MMF	Mark Taylor (Sandia)	Matt Norman
2.2.4.02 ExaSGD	Slaven Peles (PNNL)	Philip Roth
2.2.4.04 ExaBiome	Kathy Yelick (LBNL)	Philip Roth

[a] FNAL = Fermi National Accelerator Laboratory, ORNL = Oak Ridge National Laboratory, Sandia = Sandia National Laboratories, PPPL = Princeton Plasma Physics Laboratory, LBNL = Lawrence Berkeley National Laboratory, ANL = Argonne National Laboratory, PNNL = Pacific Northwest National Laboratory.

With ECP funding, these ECP AD teams have dedicated staff expertise from the OLCF Scientific Computing group, access to the system vendor's Center of Excellence (CoE) from both Cray and AMD, access to early testbed hardware provided by the Frontier vendor, and potential support from postdoctoral researchers through the Computational Scientists for Energy, the Environment, and National Security (CSEEN) program as availability allows. In addition, because of the dependence of the AD project on the ECP ST, the ECP ST projects also have access to the CoE resources and access to early testbed hardware.

This mutually beneficial partnership enables ECP to learn from the application readiness lessons learned and best practices developed during two prior instantiations of the CAAR program. Additionally, the expected OLCF application and software portfolio ready for Frontier has been further diversified by including these additional projects in the CAAR program.

OLCF-ECP Training Program

The OLCF continued its training engagements with ECP in 2019. The OLCF staff are active participants in the ECP Training Advisory Group (TAG), which meets monthly to discuss training activities under development at each of the six core ECP laboratories and identify possible activities for collaboration. The OLCF worked with ECP to co-host and facilitate training events in 2019. One such example was the OLCF/ECP OpenMP hackathon held in the summer of 2019. The OLCF worked with the ECP Training PI and members of the ECP SOLLVE project to host this hackathon July 22–26. A total of seven teams and 47 people attended the Hackathon. One of the teams worked on an ECP AD GAMESS application. The GAMESS team came to work on converting a GAMESS kernel to OpenMP in a form suitable for OpenMP and to convert a LibCChem kernel from CUDA to OpenMP. The team made progress through the week and plan to continue work on OpenMP development for GAMESS. Another example includes the Frontier Application Readiness workshop (see Section 1.4.6.3).

OLCF-ECP Continuous Integration

Developing and installing a Continuous Integration (CI) capability across the DOE laboratory complex is an area of focus for ECP. OLCF is a major influencer, helping to shape security policy and technical design requirements for a federated CI model where ECP users can use a single repository to launch CI jobs at several facilities. The OLCF is onboarding several critical ST products soon, while ECP projects such as ECP-Proxy, E4S, ALExa, FleCSI, COPA, and Data TransferKit are already using OLCF CI capabilities. These projects all make use of the Gitlab instance at ORNL, where they have access to runners on Ascent (in OLCF's open environment) for CI. These runners have been built specifically to work in an HPC environment and to enable CI style build and test pipelines, fully embracing batch schedulers and running across multiple nodes. Special care has been taken to give OLCF strong and fine-grained control over which projects, users, and codes are allowed make use of the CI infrastructure. Allocations for CI are expected to come from a project's normal compute allocation, and specific instructions have been delivered to teams to evaluate the value per compute cycle of production runs versus CI work.

In addition to the CI capabilities that ECP and OLCF have built out this year, ECP is influencing software deployment tasks at OLCF. An ECP-funded software management tool—Spack—provides "Spack Stacks" to manage large deployments of software using Spack. Rather than manually build from source the hundreds of dependencies that scientific codes need, Spack manages software complexity by allowing the OLCF to easily delineate which versions, compiler and architectural flags, and other build-time configurations should be used for libraries. It also allows reuse of already-built binaries managed by Spack. OLCF has created a Spack Stack for OLCF software deployment and provides continuous feedback to Spack developers on new features to be implemented.

8.3.2 Industry Engagement

Accelerating Competitiveness through Computational ExceLlence (ACCEL) continues to attract large-scale industrial problems that require access to leadership-scale systems in order to make progress. In addition to fielding projects on Titan, ACCEL also enabled companies to take another leap ahead in their computational problem solving with access to Summit. The introduction of Summit also encouraged more firms to attend the GPU hackathons to make progress porting important codes to GPUs.

Thirty-eight industrial projects were under way during 2019, which represented 9% of the total number of projects provided to external user programs, INCITE, ALCC, Early Science, and DD (including ECP). These projects used 743,031 Summit node-hours, representing approximately 2% of the total Summit hours, and 196,770,679 Titan core-hours, representing approximately 8% of the total Titan hours.

- In 2019, 33% of the total industrial project hours on Titan and Summit combined were allocated through INCITE, 59% via ALCC, and 8% through the OLCF DD program.

    - On Summit, 18% of the industrial project hours were allocated through INCITE, 65% via ALCC, 15% through the OLCF DD program, and 1% through Early Science.

    - On Titan, 35% of the industrial project hours were allocated through INCITE, 56% via ALCC, and 9% through the OLCF DD program.

- Of 38 operational projects, 18 were new. These new projects received awards via INCITE (two projects), ALCC (four projects), Early Science (one project), and DD (11 projects).

Observations about the Industrial Projects

- Three companies were new to ACCEL: Uber Technologies, Janssen/Johnson & Johnson, and IBM.

- ACCEL had its first industrial user selected for the early science program: GE partnering with the University of Melbourne focusing on high-fidelity CFD simulations of turbomachinery flows using the open code HipSTAR.

- Ford, GM, IBM, and Uber had the first industry projects with an Artificial Intelligence/Machine Learning (AI/ML) focus.

- This year also saw an increase in projects where companies partnered with labs or universities to win allocations on OLCF systems. The following are some examples.

  o GE partnered with University of Michigan to compete successfully for DD and ALCC awards and partnered with the University of Melbourne, for an Early Science award.

  o GM partnered with ORNL on an ALCC award.

  o Uber Technologies and ORNL partnered to win an INCITE award.

  o Janssen/Johnson & Johnson partnered with the University College of London for an INCITE award.

  o IBM and MIT partnered on a DD award.

Industry Leaps Forward at GPU Hackathons

This year, companies took advantage of the GPU hackathons that OLCF hosted and/or participated in. The ACCEL program has been promoting these valuable workshops to its industrial users, encouraging them to tap into the expert mentors at these 5-day events to help them make progress on porting important codes to GPUs. Access to these seasoned experts is part of what makes hackathons a unique training opportunity.

One of the largest successes of 2019 came from GE, a team that gained major speedups in a code used to simulate turbulence—the unsteady flow of air—around the rows of blades in large gas turbines and jet engines.

The team began adapting its GENESIS code for Summit's GPUs at the Brookhaven National Laboratory hackathon in 2018 and continued the effort at the Massachusetts Institute of Technology hackathon in June 2019. The team has achieved 50- to 300-fold speedups in pieces of its code and far greater scalability since the first event.

"The speedup obtained accelerating GENESIS by GPUs will change our intractable design challenges from far-distant goals to near-term simulation capabilities," said Carlos Velez, lead research engineer in the Thermo-Sciences Organization at GE's Global Research Center. "The hackathons have catapulted our ability to use GPUs, and that is helping us design more efficient turbines and jet engines, which has a direct impact on our competitiveness" (Figure 8.10).

The team's success demonstrates for other companies the enormous value in GPU computing. They plan to attend at least one hackathon each year to leverage the mentors' experience and expertise to help optimize the codes further.

Another industrial team that participated this year was jet engine manufacturer Pratt & Whitney, a United Technologies company. That team worked with expert mentors Dave Norton of NVIDIA and Matt Norman of ORNL at the ORNL hackathon in Knoxville to port one of the most computationally expensive pieces of the United Technologies Computational Fluid Dynamics (UTCFD) code to Summit. It was the team's first time attending a GPU hackathon.

"Matt's deep experience with CFD allowed us to set a path forward for our code," said Pete Bradley, Pratt & Whitney fellow for HPC and modeling. "We haven't encountered someone with such tailored experience for our problem in any previous events or opportunities."

Bradley's team immediately began implementing what they learned at the hackathon on a DD project.

Figure 8.10. GE's GENESIS solver (right) preserves many more wake details of interest in the flow field compared with a commercial solver (left). Image Credit: University of Kansas.

8.4 SUMMIT EARLY SCIENCE

The OLCF enabled a Summit Early Science Program during the first half of CY 2019. Drawn from CAAR and other friendly user applications, the program required a short proposal submission outlining the science to be accomplished on the then-new machine.

Overall, approximately 5 billion equivalent Titan core-hours were allocated to this program, which was equal to over 1 year of dedicated Titan system access. The Summit Early Science portfolio included the following projects.

8.4.1 Astrophysics

Frontier Precision Cosmology with HACC
Salman Habib, Argonne National Laboratory
Software Application: HACC

Modeling Stellar Explosions and Their Nucleosynthesis with an Optimized FLASH Code
Bronson Messer, Oak Ridge National Laboratory
Software Application: FLASH

GPU-accelerated General Relativistic Magnetized Simulations of Neutron Star Binary Merger Remnant Disks and Tilted Thin Accretion Disks
Alexander Tchekhovskoy, Northwestern University
Software Application: GRMHD

8.4.2 Bioinformatics

Attacking the Opioid Epidemic: Determining the Epistatic and Pleiotropic Genetic Architectures for Chronic Pain and Opioid Addiction
Daniel Jacobson, Oak Ridge National Laboratory
Software Application: COMET

8.4.3 Biophysics

Protocell: Petascale Simulation with NAMD and VMD Helps Understanding Cells at the Atomic Level
Emad Tajkhorshid, University of Illinois, Urbana-Champaign
Software Application: NAMD

All-atom Simulations of Motor Proteins for Cellular Energy Metabolism
Abhishek Singharoy, Arizona State University
Software Application: NAMD

Large-Scale Simulation of Biological Crystallization
Sharon Glotzer, University of Michigan
Software Application: HOOMD

8.4.4 Chemistry

High-Accuracy Calculations on Metal-Organic-Frameworks
Dmytro Bykov, Oak Ridge National Laboratory
Software Application: LS-DALTON

Accurate Calculations on Complex Systems Containing Early Actinides (Th-Pu)
Lucas Visscher, Vrije Universiteit Amsterdam
Software Application: DIRAC

Highly Accurate Electronic Structure Methods for Ground and Excited States at Scale: Addressing Problems in Energy Capture, Transfer and Storage
Sotiris Xantheas, Pacific Northwest National Laboratory
Software Application: NWChem

Nanoscale Design of Contacts to Atomically Precise Graphene Devices
Jerry Bernholc, North Carolina State University
Software Application: RMG

Organic Photovoltaic Materials Design Using the GronOR Non-Orthogonal Configuration Interaction Software
Remco Havenith, University of Groningen
Software Application: GronOR

8.4.5 Combustion

First Principles Investigation of Turbulent Scalar-Mixing and Combustion in Supercritical Fluids
Joseph Oefelein, Georgia Tech
Software Application: RAPTOR

Direct Numerical Simulation of Diesel Autoignition with S3D LEGION
Jacqueline Chen, Sandia National Laboratories
Software Application: S3D

8.4.6 Earth Science

Global Adjoint Tomography with Workflow Management
Jeroen Tromp, Princeton University
Software Application: SPECFEM

Cloud-resolving Climate Modeling of the Earth's Water Cycle
David Bader, Lawrence Livermore National Laboratory
Software Application: E3SM

Low-order Unstructured Finite-element Earthquake Simulation on Summit
Kohei Fujita, University of Tokyo
Software Application: MOHTRA

8.4.7 Engineering

Full-Power Simulation of the Watts Bar Nuclear Reactor Using the Shift Monte Carlo Transport Solver
Steven Hamilton, Oak Ridge National Laboratory
Software Application: EXNIHILO

8.4.8 Fluid Dynamics

Enabling Human Exploration of the Red Planet
Eric Nielsen, NASA
Software Application: Fun3D

Simulating Two-Fluid Flow in Porous Media at the Laboratory Scale
James McClure, Virginia Tech
Software Application: LBPM

8.4.9 Fusion Energy

Using XGC to Predict ITER's Boundary Plasma Performance and Its Impact on Fusion Efficiency
CS Chang, Princeton Plasma Physics Laboratory
Software Application: XGC

Integrated Simulation of Energetic Particles in Burning Plasmas
Zhihong Lin, University of California, Irvine
Software Application: GTC

8.4.10   High Energy Physics

Hot-dense Lattice QCD for RHIC Beam Energy
Swagato Mukherjee, Brookhaven National Laboratory
Software Application: RHMC

The Structure of the Proton and the Search for Physics beyond the Standard Model
André Walker-Loud, Lawrence Berkeley National Laboratory
Software Application: QUDA

8.4.11   Machine Learning

Scalable Machine Learning of Scientific Data
Robert Patton, Oak Ridge National Laboratory
Software Application: MENNDL

Exascale Deep Learning
Prabhat, Lawrence Berkeley National Laboratory
Software Application: TENSORFLOW

Solving an 80-year old Inverse Problem in Materials with Distributed Deep Learning
Nouamane Laanait, Oak Ridge National Laboratory
Software Application: DNN

Exascale AI to Advance Health Using Big Heterogeneous Biomedical Data
Georgia Tourassi, Oak Ridge National Laboratory
Software Application: CNN

8.4.12   Materials Science

Structurally Complex Oxides with Quantum Monte Carlo
Paul Kent, Oak Ridge National Laboratory
Software Application: QMCPACK

8.4.13   Nuclear Physics

The Neutrino-less Double Beta-Decay of calcium-48
Gaute Hagen, Oak Ridge National Laboratory
Software Application: NUCCOR

Hadrons, Nuclei and Fundamental Symmetries
Robert Edwards, Thomas Jefferson National Accelerator Facility
Software Application: QDP-jIT

8.4.14   Turbulence

Extreme-scale Simulations of Fluid Turbulence
PK Yeung, Georgia Tech
Software Application: DNS

First-Principle Based Flow Simulations of a High-Pressure Turbine Vane at Engine-Relevant Reynolds Number Subject to Grid-Generated Turbulence
Richard Sandberg, General Electric
Software Application: HipSTAR

Summit Early Science Highlights

GAN Achieves Exaflop Performance on Summit for Nuclear Waste Remediation Project

PI: Prabhat, Data and Analytics Services team lead, LBNL National Energy Research Scientific Computing Center
Allocation Program: Summit Early Science

The Science

A research collaboration between LBNL, PNNL, Brown University, and NVIDIA has achieved exaflop performance on the Summit supercomputer with a deep learning application used to model subsurface flow in the study of nuclear waste remediation. Their achievement demonstrates the promise of physics-informed generative adversarial networks (GANs) for analyzing complex, large-scale science problems. GANs are a class of machine learning systems that pit two neural networks against each other, a generator and a discriminator, to generate new "synthetic" data with the same statistics as the training set.

The Impact

For this study, the team used a physics-informed GAN and high-performance computing to estimate parameters and quantify uncertainty in subsurface flow. In training the GAN on the Summit supercomputer at the OLCF, the team was able to achieve 1.2 exaflop peak and sustained performance. The geographic extent, spatial heterogeneity, and multiple correlation length scales of the nuclear waste cleanup site required training the GAN model to thousands of dimensions, so the team developed a highly optimized implementation that scaled to 27,504 NVIDIA V100 Tensor Core GPUs and 4,584 nodes on Summit with a 93.1% scaling efficiency.

Summary

The researchers tracked cleanup efforts at Washington state's Hanford Site, established in 1943 as part of the Manhattan Project to produce plutonium for nuclear weapons and eventually home to the first full-scale plutonium production reactor in the world. When plutonium production ended in 1989, it left behind tens of millions of gallons of radioactive and chemical waste in large underground tanks and more than 100 square miles of contaminated groundwater resulting from the disposal of an estimated 450 billion gallons of liquids to soil disposal sites. To model the location of the area's subsurface contaminants, the researchers used synthetic data generated by a GAN-computed model based on expert knowledge about the Hanford Site. This enabled them to create a virtual representation of the site that they could then manipulate as needed based on the parameters they were interested in measuring—primarily hydraulic conductivity and hydraulic head.

Funding
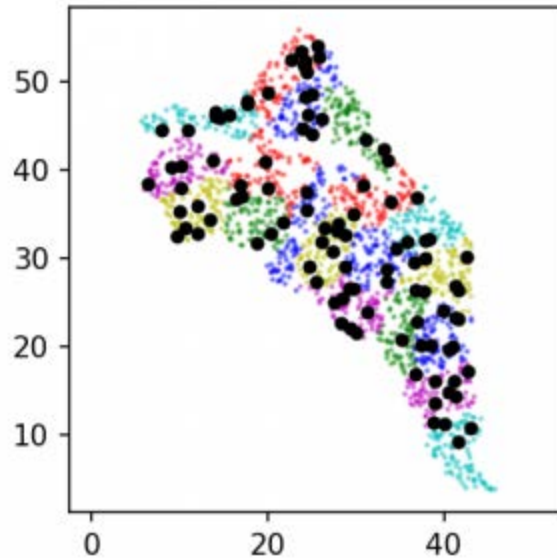
Figure 8.11. Using physics-informed GANs on the Summit supercomputer, the research team was able to estimate parameters and quantify uncertainty in subsurface flow. This image shows locations of sensors around the Hanford Site for levels 1 (black) and 2 (color). Units are in km. Image Credit: NERSC.

Publication

G. Karniadakis, Prabhat, A. Tartakovsky, M. Houston, T. Kurth, L. Yang, D. Barajas-Solano, S. Treichler, J. Romero, K. Fischer, and V. Churavy, "Highly Scalable, Physics-Informed GANs for Learning Solutions of Stochastic PDEs," paper presented at SC19, Nov. 17–22, 2019.

Related Link

"Deep Learning Expands Study of Nuclear Waste Remediation," Berkley Lab Computing Sciences highlight (November 8, 2019)

******

Can a UNICORN Outrun Earthquakes?

PI: Tsuyoshi Ichimura, University of Tokyo
Allocation Program: Summit Early Science

The Science

Each year, anywhere from a few hundred to tens of thousands of deaths are attributed to the catastrophic effects of major earthquakes. Apart from ground shaking, earthquake hazards include landslides, dam ruptures, flooding, and worse—if the sea floor is suddenly displaced during an earthquake, it can trigger a deadly tsunami. A team led by professor Tsuyoshi Ichimura at the University of Tokyo (UTokyo) transformed its Unstructured fiNite element ImpliCit sOlver with stRuctured grid coarseNing (UNICORN) code into an AI-like algorithm to simulate a 1,944 km × 2,646 km × 480 km tectonic plate boundary spanning from Vancouver, British Columbia, down to Northern California to better understand how plate subduction contributes to earthquakes. The team ran UNICORN at 416 petaflops and gained a

75-fold speedup from a previous state-of-the-art solver by fully leveraging the power of the Tensor Cores on Summit's Volta GPUs.

Impact

The new solver can be used as a tool to aid scientists in the arduous task of long-term earthquake forecasting—a goal that, when realized, could lead to earthquake prediction and disaster mitigation. UNICORN performs denser computations than the team's previous solver, allowing it to take full advantage of Summit's unique architecture. The most computationally expensive piece of the code ran at 1.1 exaflops using mixed precision—a major undertaking for a code that is based on equations rather than deep learning computations. (Codes based on the latter are inherently optimal for systems such as Summit.) For future earthquake problems, the team will need to apply UNICORN to analyze the Earth's crust and mantle responses to a fault slip over time. This will require thousands of simulations then hundreds or thousands of additional iterations to compare the results with real-world earthquake events.

Summary

A team led by Professor Tsuyoshi Ichimura at UTokyo is studying the deformation of tectonic plates to aid physics-based forecasting of natural disasters such as earthquakes. Specifically, the team is simulating a tectonic plate boundary spanning from Vancouver, British Columbia, down to Northern California. At this boundary—called the Cascadia Subduction Zone—the coastal Explorer, Juan de Fuca, and Gorda plates move east and shift underneath the North American Plate, a process known as subduction that can trigger large-magnitude earthquakes and volcanic activity. The team extended and optimized its UNICORN code for Summit, gaining a 75-fold speedup from a previous state-of-the-art solver.

Because Summit's Tensor Cores are not available for just any type of calculation, the team had to align its data access patterns and multiplication patterns to suit them. Data access patterns determine how data are accessed in memory by a software program and can be organized more efficiently to exploit a particular computer architecture.

Using UNICORN, the UTokyo team simulated a 1,944 km × 2,646 km × 480 km area at the Cascadia Subduction Zone to look at how the tectonic plate is deformed due to a phenomenon called a "fault slip," a sudden shift that occurs at the plate boundary. For the team to reach its future goals, it will have to do many simulations of crust deformation and then compare its results with observed records from past earthquakes. The team presented the work at the SC19 conference.

Funding
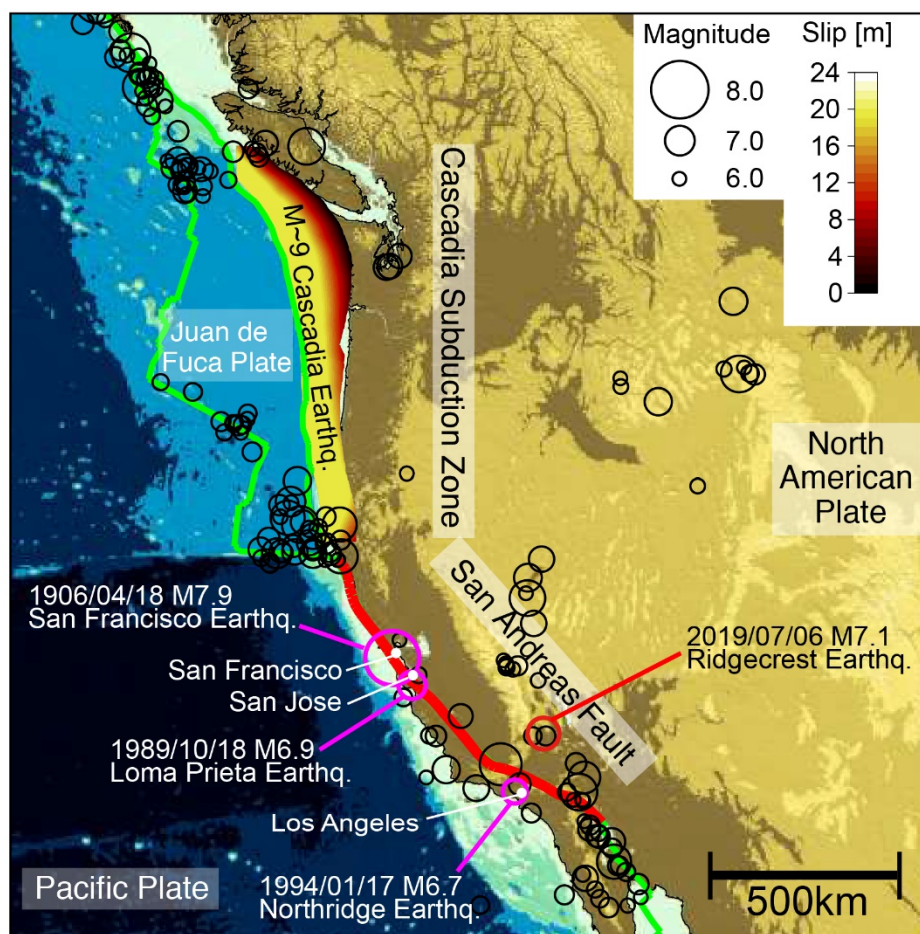
Figure 8.12. The San Andreas Fault (red lines) and the other plate boundaries (green lines). Color contours indicate the presumed fault slip distribution of the 1700 Cascadia Subduction Zone earthquake. Circles show the earthquake distribution in 1900–2019. Disastrous earthquakes (purple circles) and damaged cities are shown with the M7.1–2019 Ridgecrest earthquake (red circle). **Image Credit: University of Tokyo.**

Publication

T. Ichimura, K. Fujita, T. Yamaguchi, A. Naruse, J. C. Wells, C. J. Zimmer, T. P. Straatsma, T. Hori, S. Puel, T. W. Becker, M. Hori, and N. Ueda, "416-PFLOPS Fast Scalable Implicit Solver on Low-Ordered Unstructured Finite Elements Accelerated by 1.10-ExaFLOPS Kernel with Reformulated AI-Like Algorithm: For Equation-Based Earthquake Modeling," poster presented at the 2019 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC19), Denver, Colorado, November 17–22, 2019.

Related Links

"Can a UNICORN Outrun Earthquakes?," OLCF News (November 13, 2019)

******

Updated MENNDL Speeds Up Cancer Pathology Research with Advanced Neural Network

PI: Robert Patton, Nature Inspired Machine Learning team leader, ORNL Computational Data Analytics Group
Allocation Program: Summit Early Science

The Science

Almost all cancer patients undergo a biopsy process in which a tissue sample is examined under a microscope by a pathologist. The results not only help in determining their individual prognosis but can also add to a growing database for cancer researchers to plumb. With digital pathology, biopsy slides are scanned into gigapixel images; this allows neural networks to analyze sections of each biopsy image for concentrations of tumor-infiltrating lymphocytes (TILs), which are immune cells that attack cancers. By labelling different types of cancers and providing quantitative estimates of TIL density and patterns, a neural network can help researchers predict the effectiveness of different treatments. But until now, state-of-the-art neural networks could not keep up with image production. Working with the Stony Brook Cancer Center, a team led by Robert Patton at ORNL updated its Multinode Evolutionary Neural Networks for Deep Learning (MENNDL) code to create a neural network that pursues two goals at once: it can address the image-analysis problem with both accuracy and speed.

The Impact

The previously used neural network (Inception v4) provided accuracy but would take about 24 minutes to analyze each image vs. the 30 seconds needed to generate it. In contrast, MENNDL's neural network can analyze these images for TILs at a rate of 1.5 minutes per slide—16 times faster—with nearly the same accuracy. This means researchers will now be able to analyze biopsy image data at an unprecedented scale.

Summary

By using neural networks that can quickly and accurately analyze biopsy slide images on a scale that microscope-equipped pathologists could never completely tackle, the project promises to unlock new information on how tumors react to different treatments. With further development under a DOE INCITE 2020 grant, future iterations of the neural network aim to be even faster and smarter. The team's ultimate goal is to use MENNDL's networks to accurately map out all cell types in a whole-slide tumor image, including cancer cells and lymphocytes, and analyze large pathology datasets. In turn, this could profoundly affect the fight against cancer while also pioneering new ways of creating multi-objective neural networks.

Funding
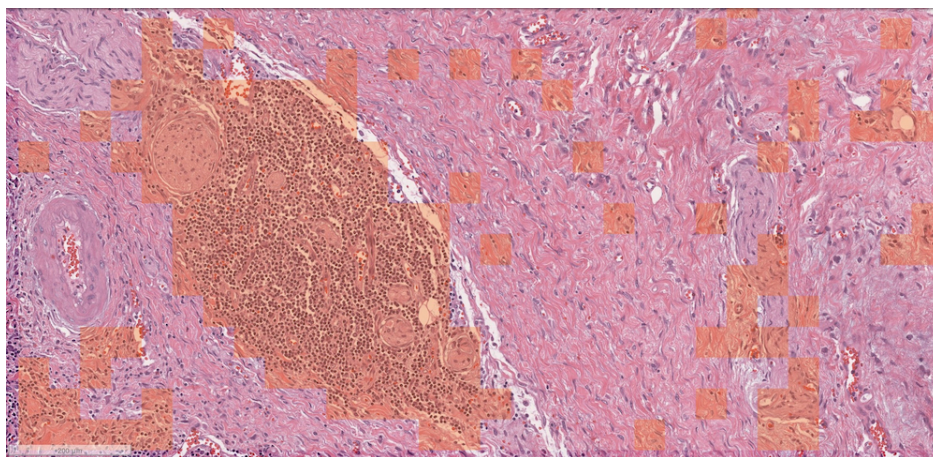
Figure 8.13. This portion of a whole-slide cancer biopsy image highlights where TILs appear in the orange overlay. It was labeled using the MENNDL neural network. Image credit: Robert Patton/ORNL.

Publication

R. M. Patton, J. T. Johnston, S. R. Young, C. D. Schuman, T. E. Potok, D. C. Rose, S.-H. Lim, J. Chae, L. Hou, S. Abousamra, D. Samaras, and J. Saltz, "Exascale Deep Learning to Accelerate Cancer Research," paper presented and published at the IEEE Big Data 2019 conference, Los Angeles, CA, December 9, 2019.

Related Link

"In the Fight Against Cancer, ORNL and Stony Brook Cancer Center Enlist an Advanced Neural Network," OLCF News (December 16, 2019)

APPENDIX A. RESPONSES TO RECOMMENDATIONS FROM THE 2018 OPERATIONAL ASSESSMENT REVIEW

In April 2019, the operational activities of the OLCF were presented to the DOE sponsor. The review committee of that report identified two recommendations.

Recommendation	Facility response
Adjusting the target levels for user support metrics to hit a slightly higher level based on years of prior excellent performance should be a strong consideration between the facility and ASCR.	In partnership with the DOE Program Manager, the OLCF chose not to implement this recommendation due to the transition to a completely new platform and software stack in 2019.
Provide a few more details in the community engagement section and include a few specific examples or references.	The OLCF agrees with this recommendation. Additional information is provided in Section 8.3, Stakeholder Engagement and Outreach.

APPENDIX B. TRAINING, WORKSHOPS, AND SEMINARS

Event Type	Event Title	Date	Participants
Monthly User Con Call	GPU [graphics processing unit] Direct, CUDA Aware MPI & CUDA IPC [interprocess communication]	1/30/19	108
Monthly User Con Call	Advanced On-Node GPU Communication	2/27/19	86
Monthly User Con Call	CUDA 10 Overview	3/27/19	73
Monthly User Con Call	Remote Visualization Tools for Rhea	5/1/19	37
Monthly User Con Call	OLCF Best Practices	6/19/19	55
Monthly User Con Call	POPPER	7/31/19	29
Monthly User Con Call	Intro to Slurm	8/28/19	32
Monthly User Con Call	OLCF User Documentation	9/25/19	26
Monthly User Con Call	Distributed Deep Learning on Summit	10/30/19	139
Monthly User Con Call	Migrating to Project-based Directories in HPSS	12/11/19	28
IDEAS-ECP/Facility Webinar Series	Quantitatively Assessing Performance Portability with Rooflline	1/23/19	58
IDEAS-ECP/Facility Webinar Series	Containers in HPC	2/13/19	138
IDEAS-ECP/Facility Webinar Series	Parallel I/O [input/output] with HDF5 [Hierarchical Data Format 5]: Overview, Tuning, and New Features	3/13/19	96
IDEAS-ECP/Facility Webinar Series	Testing Fortran Software with pFUnit	4/10/19	61
IDEAS-ECP/Facility Webinar Series	So You Want to Be Agile? Strategies for Introducing Agility Into Your Scientific Software Project	5/8/19	64
IDEAS-ECP/Facility Webinar Series	Modern C++ for High-Performance Computing	6/12/19	134
IDEAS-ECP/Facility Webinar Series	When 100 Flops/Watt was a Giant Leap: The Apollo Guidance Computer Hardware, Software and Application in Moon Missions	7/17/19	72
IDEAS-ECP/Facility Webinar Series	Software Management Plans in Research Projects	8/14/19	81
IDEAS-ECP/Facility Webinar Series	Social Changes in the Evolution of Scientific Software Projects	9/11/19	57
IDEAS-ECP/Facility Webinar Series	Tools and Techniques for Floating-Point Analysis	10/16/19	91
IDEAS-ECP/Facility Webinar Series	Building Community Policies through xSDK [Extreme-Scale Scientific Software Development Kit] Software Policies	12/11/19	50
Workshop/Training/Meeting	Summit Training Workshop	2/11–13/19	131
Workshop/Training/Meeting	SIAM Computational Science and Engineering (CSE) Conference Workshop	2/25–3/1/19	-
Workshop/Training/Meeting	Introduction to NVIDIA Profilers on Summit	4/11/19	71
Workshop/Training/Meeting	CentOS Dojo Internal Staff Training	4/16/19	-
Workshop/Training/Meeting	Quantum Computing User Training	4/24/19	-
Workshop/Training/Meeting	Quantum Computing User Forum	4/24–25/19	65
Workshop/Training/Meeting	ALCF/OLCF Joint INCITE Webinar #1	05/09/19	43

Event Type	Event Title	Date	Participants
Workshop/Training/Meeting	Introduction to Summit Workshop	5/20/19	69
Workshop/Training/Meeting	2019 OLCF User Meeting	5/21–23/19	132
Workshop/Training/Meeting	ALCF/OLCF Joint INCITE Webinar #2	06/04/19	50
Workshop/Training/Meeting	Introduction to AMD GPU Programming with HIP [Heterogeneous-Compute Interface for Portability] Webinar (in collaboration with ECP)	6/7/19	-
Workshop/Training/Meeting	LINUX Command Line Productivity Tools Workshop	6/11/19	97
Workshop/Training/Meeting	OpenMP Hackathon (in collaboration with ECP)	7/23–26/19	47
Workshop/Training/Meeting	Profiling Tools Training	8/7–9/19	54
Workshop/Training/Meeting	CCSD [Computing and Computational Sciences Directorate] Tensor Core Workshop	8/13/19	61
Workshop/Training/Meeting	Petascale Institute Training Workshop (in collaboration with many other NSF [National Science Foundation] and DOE facilities)	8/19–22/19	36
Workshop/Training/Meeting	Introduction to AMD GPU Programming with HIP (in collaboration with ECP)	9/6/19	136
Workshop/Training/Meeting	2019 ACM Richard Tapia Celebration of Diversity in Computing Conference	9/18/19	-
Workshop/Training/Meeting	CAAR Frontier Kick-Off Meeting (in collaboration with ECP)	10/8–10/19	160
Workshop/Training/Meeting	SC19 Student Enrichment Workshop	11/	-
2019 GPU Hackathon Series	KISTI (in collaboration with NVIDIA and the host site)	2/18–22/19	39
2019 GPU Hackathon Series	Pawsey SC (in collaboration with NVIDIA and the host site)	3/25–29/19	31
2019 GPU Hackathon Series	Helmholtz-Julich (in collaboration with NVIDIA and the host site)	4/8–12/19	61
2019 GPU Hackathon Series	JGI (in collaboration with NVIDIA and the host site)	5/6–10/19	33
2019 GPU Hackathon Series	MIT (in collaboration with NVIDIA and the host site)	6/3–7/19	41
2019 GPU Hackathon Series	Princeton (in collaboration with NVIDIA and the host site)	6/24–28/19	52
2019 GPU Hackathon Series	Sheffield (in collaboration with NVIDIA and the host site)	8/19–23/19	43
2019 GPU Hackathon Series	Brookhaven (in collaboration with NVIDIA and the host site)	9/23–27/19	64
2019 GPU Hackathon Series	CSCS (in collaboration with NVIDIA and the host site)	9/30–10/4/19	31
2019 GPU Hackathon Series	OLCF (in collaboration with NVIDIA and the host site)	10/21–25/19	76
Seminar Series	Performance Portability from a Single Code	1/22/19	-
Seminar Series	Slurm Containerization and Orchestration using Docker and Kubernetes	1/23/19	-
Seminar Series	Techniques to Improve Genome Assembly Quality Using Paired Reads	1/24/19	-
Seminar Series	Predicting Resilience of GPUs Applications Using Statistical Methods	1/31/19	-

Event Type	Event Title	Date	Participants
Seminar Series	Climate Model Quality Assurance through Consistency Testing and Error Source Identification	2/4/19	-
Seminar Series	Quantifying Sources of I/O Performance Variation Through Holistic Analysis	2/21/19	-
Seminar Series	Using Deep Learning in Gravitational Lensing	3/19/19	-
Seminar Series	GROMACS+STS [Groningen Machine for Chemical Simulations Static Thread Scheduler]: Enabling Tasking for Established and Highly-optimized Code	3/22/19	-
Seminar Series	Co-designing Scientific Applications and Performance Tools for HPC	3/25/19	-
Seminar Series	Machine Intelligence Applications for Particle Physics	3/25/19	-
Seminar Series	Integrating Experiment, Simulation and Big Data Analytics for Validation and Accelerated Synergistic Advancement in Biomolecular Science	3/25/19	-
Seminar Series	Experiences in World Wide Computing	4/4/19	-
Seminar Series	Revisiting Databases for Scale Up Data Management	4/23/19	-
Seminar Series	Architecture-Performance Interrelationship Analysis in Single/Multiple CPU/GPU Computing Systems: Application to Composite Process Flow Modeling	4/29/19	-
Seminar Series	Scientific Computing Applications in Astrophysics	5/6/19	-
Seminar Series	Analysis of High Performance Scientific Programming Workflows	5/8/19	-
Seminar Series	CFD [Computational Fluid Dynamics] Analysis of Heavy Vehicle Platooning for Drag Reduction	5/9/19	-
Seminar Series	From Compilers and Tools to Benchmarks and Metrics: Seeking the Driving Forces of HPC	5/13/19	-
Seminar Series	Ab-initio typical-medium single-site theory for disordered systems	5/17/19	-
Seminar Series	Parallel Computing Developments for Nuclear Theory with Three-Body Forces	5/17/19	-
Seminar Series	Interview Candidate: Low-cost Evaluation and Tuning of Dynamic Load Balancing Through Simulation	5/20/19	-
Seminar Series	Compilers in HPC-On Parallel Abstraction Penalties and Domain Knowledge Expecting Performance Estimation	6/6/19	-
Seminar Series	Multiscale Modeling: Scientific and Numerical Challenges	6/12/19	-
Seminar Series	Scalable Graph-based Methods for Large-scale Data Analytics	6/17/19	-
Seminar Series	Easy PRAM [parallel random access machine]-Based High-Performance Parallel Programming	6/25/19	-
Seminar Series	Efficient Parallel I/O with HDF5 and Proactive Data Containers (PDCs)	6/25/19	-
Seminar Series	Using RBF [Radial Basis Function] Generated Quadrature Rules to Solve Nonlocal Continuum Models	6/27/19	-
Seminar Series	Study of interconnect Errors, Network Congestion,	7/11/19	-

Event Type	Event Title	Date	Participants
	and Applications Characteristics on a Large Scale HPC System		
Seminar Series	Machine Learning for Advanced Materials Design and Characterization	7/18/19	-
Seminar Series	Automatic Construct Selection and Variable Classification in OpenMP	8/12/19	-
Seminar Series	Scientific Computing Applications in Astrophysics	8/16/19	-
Seminar Series	HPC Monitoring & Analysis + Power 9 Specifics	8/19/19	-
Seminar Series	Designing Reusable Composable Components for the (HPC) I/O Stack	8/22/19	-
Seminar Series	HPC Data Management at LANL	8/28/19	-
Seminar Series	Improving Modeling & Simulations: From Earthquake to Data-Driven Simulations	9/6/19	-
Seminar Series	Towards Smart Cyber Physical Systems	9/13/19	-
Seminar Series	Data and Network Science Methods for Detecting Anomalies in Time-Varying Networked Systems	9/18/19	-
Seminar Series	New Frontiers in Molecular Simulation Techniques and Applications	9/30/19	-
Seminar Series	Centroid Path Integral Investigations of Zero-Point Motion and Three-body Interactions in Solid He-4	10/1/19	-
Seminar Series	Non-Orthogonal Electronic Structure Approaches for Reducing Computational Burden in Quantum Chemistry Calculations	10/18/19	-
Seminar Series	Graphs for Heterogeneous Supercomputing Architectures	10/25/19	-
Seminar Series	Too Complex to Model? Why the Next Biomaterial Will Be Discovered on a Supercomputer	10/28/19	-
Seminar Series	An Evaluation of the CORAL [Collaboration of Oak Ridge, Argonne and Livermore] Interconnects	11/11/19	-
Seminar Series	Lightweight User-Level Threads for Massive Fine-Grained Parallelism: Argobots and BOLT [BOLT is OpenMP over Lightweight Threads]	11/12/19	-
Seminar Series	The Quantum Chemistry of Quantum Computers and the Other Way Around	11/13/19	-
Seminar Series	Hackathon Design Decisions and Their Effects on Hackathon Outcomes	11/14/19	-
Seminar Series	Lattice Quantum Chromodynamics on the Bleeding Edge	12/2/19	-
Seminar Series	Direct Numerical Simulations of Flame Propagation in Stratified Mixtures at Autoignitive Conditions Using Accelerated Computing	12/6/19	-
Seminar Series	High-Resolution Computations of Aerodynamic Flow Control using Fluidic Oscillators	12/17/19	-

APPENDIX C. OUTREACH PRODUCTS

Date	Type of Product	Title
1/7/19	Fact Sheet	CORAL 2
1/14/19	Poster	ECP Annual Meeting
1/17/19	Highlight	ORNL Security Teams Prepare to Take the Heat with Cyber Fire
1/17/19	Highlight	OLCF Staff Elected to APS Division of Computational Physics Executive Committee
1/17/19	Highlight	NCCS Introduces HPC Core Operations Group
1/17/19	Highlight	A Sneak Peek at 19 Science Simulations for the Summit Supercomputer in 2019
1/17/19	Highlight	Network Enhancement Strengthens Ties Between OLCF, ESnet
1/17/19	PPT Slide	ORNL Security Teams Prepare to Take the Heat with Cyber Fire
1/17/19	PPT Slide	OLCF Staff Elected to APS Division of Computational Physics Executive Committee
1/17/19	PPT Slide	NCCS Introduces HPC Core Operations Group
1/17/19	PPT Slide	A Sneak Peek at 19 Science Simulations for the Summit Supercomputer in 2019
1/17/19	PPT Slide	Network Enhancement Strengthens Ties Between OLCF, ESnet
2/1/19	Article	*HPCWire* People to Watch: Jack Wells
2/6/19	Graphic	Farewell Titan sticker
2/6/19	Highlight	ORNL Adds Powerful AI Appliances to Computing Portfolio
2/6/19	Highlight	New Geometric Model Improves Predictions of Fluid Flow in Rock
2/6/19	Highlight	Modeling the Origin Story of the Elements
2/6/19	PPT Slide	ORNL Adds Powerful AI Appliances to Computing Portfolio
2/6/19	PPT Slide	New Geometric Model Improves Predictions of Fluid Flow in Rock
2/6/19	PPT Slide	Modeling the Origin Story of the Elements
2/8/19	Article	Univ. of Iowa: Brenna Miller Profile
2/11/19	Fact Sheet	RSA Token one-pager reprint
2/12/19	Poster	Tribute to Linda Gregg
2/25/19	Report	Contributions for Operational Assessment
2/26/19	Graphic	System Changes web tile
2/28/19	Graphic	Frontier Branding – presentation/pitch (3/27 Frontier branding approved)
3/12/19	Highlight	Solving A Beta Decay Puzzle
3/12/19	Highlight	Supercomputers to Help Supercharge Ceramic Matrix Composite Manufacturing
3/12/19	PPT Slide	Solving A Beta Decay Puzzle
3/12/19	PPT Slide	Supercomputers to Help Supercharge Ceramic Matrix Composite Manufacturing
3/14/19	Graphic	Summit Celebration Invitation
3/15/19	Graphic	AI Initiative Graphic
3/20/19	Fact Sheet	Summit fact sheet update with Katrin Heitmann
3/21/19	Graphic	Frontier Video Concept (branding/animation)
3/28/19	Website	INCITE redesign
4/3/19	Highlight	Troubleshooting the World's Smartest and Fastest Supercomputer
4/3/19	Highlight	Adamson Named New Group Leader for HPC Core Ops
4/3/19	Highlight	Getting a Big Look at Tiny Particles
4/3/19	Highlight	Laser Focus Sheds Light on Two Sources of Nanoparticle Formation

Date	Type of Product	Title
4/3/19	PPT Slide	Troubleshooting the World's Smartest and Fastest Supercomputer
4/3/19	PPT Slide	Adamson Named New Group Leader for HPC Core Ops
4/3/19	PPT Slide	Getting a Big Look at Tiny Particles
4/3/19	PPT Slide	Laser Focus Sheds Light on Two Sources of Nanoparticle Formation
4/11/19	Graphic	Frontier Event Display Concept (branding)
4/15/19	Graphic	INCITE homepage tile
4/15/19	Highlight	US Department of Energy's INCITE Program Seeks Proposals for 2020
4/15/19	PPT Slide	US Department of Energy's INCITE Program Seeks Proposals for 2020
4/16/19	Graphic	Frontier machine graphic for marketing materials (Cray image)
4/24/19	Fact Sheet	Frontier talking points/fact sheet
4/24/19	Graphic	No. 1 systems timeline for factsheet and web
4/25/19	Graphic	Frontier backdrops for launch event (3 roll-ups)
4/25/19	Highlight	OLCF Scientist Talks Early Summit Results at APS [American Physical Society] Meeting
4/25/19	Highlight	Upgraded Science Trailer Teaches Students about Supercomputing
4/25/19	PPT Slide	OLCF Scientist Talks Early Summit Results at APS Meeting
4/25/19	PPT Slide	Upgraded Science Trailer Teaches Students about Supercomputing
4/26/19	Graphic	Animated Frontier logo (universe background) – for Butch
4/26/19	Graphic	3D options for hologram – for X-Lab
4/26/19	Graphic	Frontier homepage tiles – two tiles: coming 2021 and call for proposals
4/30/19	Highlight	Simulations Identify Importance of Lattice Distortions in Ion-Conducting Fuel Cell Materials
4/30/19	PPT Slide	Simulations Identify Importance of Lattice Distortions in Ion-Conducting Fuel Cell Materials
5/1/19	Highlight	Virtual Universes
5/1/19	PPT Slide	Virtual Universes
5/2/19	Fact Sheet	Frontier spec sheet
5/2/19	Graphic	Podium banner for Frontier announcement
5/2/19	Graphic	Podium banner for Frontier announcement
5/6/19	Fact Sheet	Frontier science -- for Thomas Zacharia and CEOs
5/6/19	Graphic	Frontier social tiles (ECP PI quotes)
5/7/19	Graphic	Frontier website branding elements
5/7/19	Highlight	CAAR Accepting Application Team Proposals for Frontier System
5/7/19	Highlight	Science at Exascale: The Future of Fusion Modeling
5/7/19	Highlight	Science at Exascale: 3D Printing
5/7/19	Highlight	Science at Exascale: Molecular Dynamics for Materials
5/7/19	Highlight	Science at Exascale: Simulating Small Modular Reactor Operations
5/7/19	Highlight	Science at Exascale: Mapping Climate Patterns
5/7/19	Highlight	Science at Exascale: Tracing the Origins of Matter in Star Explosions
5/7/19	Highlight	No Scaling Back: DOE, Cray, AMD to Bring Exascale to ORNL
5/7/19	PPT Slide	CAAR Accepting Application Team Proposals for Frontier System

Date	Type of Product	Title
5/7/19	PPT Slide	Science at Exascale: The Future of Fusion Modeling
5/7/19	PPT Slide	Science at Exascale: 3D Printing
5/7/19	PPT Slide	Science at Exascale: Molecular Dynamics for Materials
5/7/19	PPT Slide	Science at Exascale: Simulating Small Modular Reactor Operations
5/7/19	PPT Slide	Science at Exascale: Mapping Climate Patterns
5/7/19	PPT Slide	Science at Exascale: Tracing the Origins of Matter in Star Explosions
5/7/19	PPT Slide	No Scaling Back: DOE, Cray, AMD to Bring Exascale to ORNL
5/7/19	Video	Introducing Frontier
5/7/19	Video	Frontier Early Science
5/7/19	Website	Introducing Frontier
5/14/19	Graphic	User Meeting certificate
5/15/19	Poster	DAC Poster NCCS
5/15/19	Poster	DAC Poster OLCF
5/17/19	Article	CiSE: Supercomputing Improves Predictions of Fluid Flow in Rock
5/17/19	Highlight	On the Verge of Exascale, the World's Fastest Petascale System Points the Way
5/17/19	PPT Slide	On the Verge of Exascale, the World's Fastest Petascale System Points the Way
5/20/19	Highlight	Summit Charts a Course to Uncover the Origins of Genetic Diseases
5/20/19	PPT Slide	Summit Charts a Course to Uncover the Origins of Genetic Diseases
6/12/19	Highlight	TechInt Develops New Solutions for Storage on Summit
6/12/19	Highlight	ORNL Staff Plug into Tech Day
6/12/19	Highlight	15th OLCF User Meeting Looks to Future, Reflects on Past
6/12/19	Highlight	NCCS Researchers Receive Award at Parallel Computing Symposium
6/12/19	PPT Slide	TechInt Develops New Solutions for Storage on Summit
6/12/19	PPT Slide	ORNL Staff Plug into Tech Day
6/12/19	PPT Slide	15th OLCF User Meeting Looks to Future, Reflects on Past
6/12/19	PPT Slide	NCCS Researchers Receive Award at Parallel Computing Symposium
6/17/19	Graphic	Frontier Elements for AMD (Chip Frietag)
6/25/19	Graphic	Frontier ECP Spread
6/28/19	Graphic	Mockup of computer doors display for Main St.
6/28/19	Highlight	Second "Introduce Your Daughter to AI" Event Is a Hit
6/28/19	Highlight	Farewell, Titan
6/28/19	Highlight	Jack Wells to Help Strengthen OpenACC [Open Accelerators] User Community
6/28/19	Highlight	ORNL Teams Participate in CUG [Cray User Group] 2019
6/28/19	PPT Slide	Second "Introduce Your Daughter to AI" Event Is a Hit
6/28/19	PPT Slide	Farewell, Titan
6/28/19	PPT Slide	Jack Wells to Help Strengthen OpenACC User Community
6/28/19	PPT Slide	ORNL Teams Participate in CUG2019
7/5/19	Highlight	A Data-Driven Journey to the Center of the Earth
7/5/19	PPT Slide	A Data-Driven Journey to the Center of the Earth
7/9/19	Article	Network World Interview

Date	Type of Product	Title
7/15/19	Highlight	Predicting Material Properties with Quantum Monte Carlo
7/15/19	PPT Slide	Predicting Material Properties with Quantum Monte Carlo
7/19/19	Graphic	Presentation on the "Process"
7/22/19	Highlight	Artificial Intelligence Approach Points to Bright Future for Fusion Energy
7/22/19	Highlight	PPPL's Will Fox Honored for Research on Titan
7/22/19	Highlight	ORNL Scientists Make Fundamental Discovery to Creating Better Crops
7/22/19	Highlight	OLCF and Tech Company Providentia Worldwide Build Intelligence System for Supercomputer Cooling Plant
7/22/19	Highlight	ALCC Program Awards 6 Million Hours on OLCF Resources
7/22/19	PPT Slide	Artificial Intelligence Approach Points to Bright Future for Fusion Energy
7/22/19	PPT Slide	PPPL's Will Fox Honored for Research on Titan
7/22/19	PPT Slide	ORNL Scientists Make Fundamental Discovery to Creating Better Crops
7/22/19	PPT Slide	OLCF and Tech Company Providentia Worldwide Build Intelligence System for Supercomputer Cooling Plant
7/22/19	PPT Slide	ALCC Program Awards 6 Million Hours on OLCF Resources
7/29/19	Video	Titan removal time lapse
7/30/19	Graphic	Photo edit (blur badges) for media visit
7/30/19	Video	Galactic winds news story video
7/31/19	Graphic	Awards night 1-pager on Justin Whitt
8/1/19	Graphic	ALCC news story web banner
8/1/19	Highlight	Galactic Winds Demystified
8/1/19	Highlight	Igniting a New Class of Combustion Research
8/1/19	PPT Slide	Galactic Winds Demystified
8/1/19	PPT Slide	Igniting a New Class of Combustion Research
8/5/19	Video	Farewell Titan (shutdown)
8/28/19	Graphic	Science graphic for Eisenbach feature
8/29/19	Highlight	Summer Interns Gain Hands-On Experience at Massive Scale
8/29/19	Highlight	From Simulation to Automation in a Data-Rich World
8/29/19	Highlight	ORNL-VA Collaboration Targets Veteran Suicide Epidemic
8/29/19	PPT Slide	Summer Interns Gain Hands-On Experience at Massive Scale
8/29/19	PPT Slide	From Simulation to Automation in a Data-Rich World
8/29/19	PPT Slide	ORNL-VA Collaboration Targets Veteran Suicide Epidemic
8/30/19	Graphic	Summit coaster for HR
8/30/19	Poster	ALCC 2019-2020 Poster
9/4/19	Highlight	CAAR Partnerships for Frontier Announced
9/4/19	PPT Slide	CAAR Partnerships for Frontier Announced
9/10/19	Article	ASCAC [Advanced Scientific Computing Advisory Committee 40th – Petaflops for the People
9/10/19	Graphic	Summit graphic for Frontier and INCITE slides
9/12/19	Article	Microway Case Study
9/12/19	Highlight	GPUs Power GE Code at OLCF Hackathons

Date	Type of Product	Title
9/12/19	Highlight	ORNL Staff Highlight OpenACC's Role in HPC at Annual Meeting
9/12/19	PPT Slide	GPUs Power GE Code at OLCF Hackathons
9/12/19	PPT Slide	ORNL Staff Highlight OpenACC's Role in HPC at Annual Meeting
9/20/19	Graphic	Facility Accomplishments slides for Jeff Nichols
9/24/19	Article	Big Picture Science Podcast
9/24/19	Graphic	Summit pull-up banner stand
9/24/19	Poster	Summit by the numbers printed and mounted
10/2/19	Graphic	Summit photo backdrop for the photo studio
10/3/19	Highlight	Titan Supercomputer and Spallation Neutron Source Unite to Probe the Inner Workings of c-Src kinase
10/3/19	Highlight	Chromosome Connectors Take Center Stage for ORNL Scientists Studying Poplar
10/3/19	PPT Slide	Titan Supercomputer and Spallation Neutron Source Unite to Probe the Inner Workings of c-Src kinase
10/3/19	PPT Slide	Chromosome Connectors Take Center Stage for ORNL Scientists Studying Poplar
10/10/19	Highlight	2019 Gordon Bell Finalists Powered by Summit
10/10/19	Highlight	Summit Simulates How Humans Will 'Brake' during Mars Landing
10/10/19	PPT Slide	2019 Gordon Bell Finalists Powered by Summit
10/10/19	PPT Slide	Summit Simulates How Humans Will 'Brake' during Mars Landing
10/15/19	Graphic	Student cluster challenge t-shirt design
10/17/19	Highlight	ORNL Celebrates First National Exascale Day on October 18
10/17/19	PPT Slide	ORNL Celebrates First National Exascale Day on October 18
10/23/19	Highlight	Quantum Processors are Now Challenging Conventional HPC Systems
10/23/19	Highlight	Previewing the New Frontier of High-Performance Computing
10/23/19	PPT Slide	Quantum Processors are Now Challenging Conventional HPC Systems
10/23/19	PPT Slide	Previewing the New Frontier of High-Performance Computing
10/28/19	Highlight	SULI Profile: Jess Woods
10/28/19	Highlight	Search for Lightweight Alloying Solutions Earns Team a Gordon Bell Finalist Nomination
10/28/19	PPT Slide	SULI Profile: Jess Woods
10/28/19	PPT Slide	Search for Lightweight Alloying Solutions Earns Team a Gordon Bell Finalist Nomination
11/6/19	Highlight	In the Mix
11/6/19	PPT Slide	In the Mix
11/8/19	Highlight	Deep Learning Expands Study of Nuclear Waste Remediation
11/8/19	PPT Slide	Deep Learning Expands Study of Nuclear Waste Remediation
11/8/19	Video	Gordon Bell Finalist Team Tackles Transistors with New Programming Paradigm
11/13/19	Article	CiSE: Summit Simulates How Humans Will 'Brake' during Mars Landing
11/13/19	Highlight	A New Parallel Strategy for Tackling Turbulence on Summit
11/13/19	Highlight	AI for Plant Breeding in an Ever-Changing Climate
11/13/19	Highlight	Can a UNICORN Outrun Earthquakes?
11/13/19	Highlight	Modeling Every Building in America Starts with Chattanooga

Date	Type of Product	Title
11/13/19	Highlight	The OLCF to Mentor Interns Competing at SC19
11/13/19	Highlight	In Its 15th year, INCITE Advances Open Science with Supercomputer Grants to 47 Projects
11/13/19	PPT Slide	A New Parallel Strategy for Tackling Turbulence on Summit
11/13/19	PPT Slide	AI for Plant Breeding in an Ever-Changing Climate
11/13/19	PPT Slide	Can a UNICORN Outrun Earthquakes?
11/13/19	PPT Slide	Modeling Every Building in America Starts with Chattanooga
11/13/19	PPT Slide	The OLCF to Mentor Interns Competing at SC19
11/13/19	PPT Slide	In Its 15th year, INCITE Advances Open Science with Supercomputer Grants to 47 Projects
11/15/19	Graphic	INCITE web banner
11/17/19	Web Article	SC19 Event Details
11/18/19	Highlight	United States Continues To Lead World In Supercomputing
11/18/19	PPT Slide	United States Continues To Lead World In Supercomputing
11/20/19	Graphic	Banner stands for MRS meeting (ORNL, ANL and LBNL)
11/21/19	Highlight	Tiny Transistor Leads to Big Win for ETH Zurich, 2019 ACM [Association for Computing Machinery] Gordon Bell Prize Winner
11/21/19	PPT Slide	Tiny Transistor Leads to Big Win for ETH Zurich, 2019 ACM Gordon Bell Prize Winner
12/12/19	Graphic	Titan slide for Buddy
12/13/19	Graphic	Frontier "Coming soon:" posters for the overlook
12/16/19	Highlight	In the Fight Against Cancer, ORNL and Stony Brook Cancer Center Enlist an Advanced Neural Network
12/16/19	Highlight	Record-Breaking Year for OLCF GPU Hackathons Ends with Annual Knoxville, Tennessee, Event
12/16/19	Highlight	OLCF Staffers Among Recipients at Annual DOE Secretary's Honor Awards
12/16/19	Highlight	OLCF Supercharges Supercomputer Analytics with Apache Kafka
12/16/19	PPT Slide	In the Fight Against Cancer, ORNL and Stony Brook Cancer Center Enlist an Advanced Neural Network
12/16/19	PPT Slide	Record-Breaking Year for OLCF GPU Hackathons Ends with Annual Knoxville, Tennessee, Event
12/16/19	PPT Slide	OLCF Staffers Among Recipients at Annual DOE Secretary's Honor Awards
12/16/19	PPT Slide	OLCF Supercharges Supercomputer Analytics with Apache Kafka
12/18/19	Highlight	Tourassi Appointed Director of ORNL National Center for Computational Sciences
12/18/19	PPT Slide	Tourassi Appointed Director of ORNL National Center for Computational Sciences
12/19/19	Graphic	Data Matters Series graphic and banner
12/19/19	Highlight	With ADIOS, Summit Processes Celestial Data at Scale of Massive Future Telescope
12/19/19	PPT Slide	With ADIOS, Summit Processes Celestial Data at Scale of Massive Future Telescope
12/20/19	Poster	INCITE 2020 Poster
12/30/19	Article	IBM Blog: Oak Ridge Leadership Computing Facility Enables Breakthrough Science

APPENDIX D. BUSINESS RESULTS FORMULAS

2019 Operational Assessment Guidance

Scheduled Availability

For HPC facilities, scheduled availability (reference formula #1) is the percentage of time a designated level of resource is available to users, excluding scheduled downtime for maintenance and upgrades. To be considered a scheduled outage, the user community must be notified of the need for a maintenance event window no less than 24 hours in advance of the outage (emergency fixes). Users will be notified of regularly scheduled maintenance in advance, on a schedule that provides sufficient notification no less than 72 hours prior to the event, and preferably as much as 7 calendar days prior. If that regularly scheduled maintenance is not needed, then users will be informed of the cancellation of that maintenance event in a timely manner. Any interruption of service that does not meet the minimum notification window is categorized as an unscheduled outage.

A significant event that delays a return to scheduled production by more than 4 hours will be counted as an adjacent unscheduled outage, as an unscheduled availability, and as an additional interrupt.

$$SA = \left(\frac{time\ in\ period - time\ unavailable\ due\ to\ outages\ in\ period}{time\ in\ period - time\ unavailable\ due\ to\ scheduled\ outages\ in\ period} \right) * 100 \tag{1}$$

Overall Availability

Overall availability (reference formula #2) is the percentage of time a system is available to users. Outage time reflects both scheduled and unscheduled outages.

$$OA = \left(\frac{time\ in\ period - time\ unavailable\ due\ to\ outages\ in\ period}{time\ in\ period} \right) * 100 \tag{2}$$

Mean Time to Interrupt

Mean time to interrupt is time, on average, to any outage of the full system, whether unscheduled or scheduled; also known as MTBI (mean time between interrupt, reference formula #3).

$$MTTI = \left(\frac{time\ in\ period - (duration\ of\ scheduled\ outages + duration\ of\ unscheduled\ outages)}{number\ of\ scheduled\ outages + number\ of\ unscheduled\ outages + 1} \right) \tag{3}$$

Mean Time to Failure

Mean time to failure is time, on average, to an unscheduled outage of the full system (reference formula #4).

$$MTTF = \frac{time\ in\ period - (duration\ of\ unscheduled\ outages)}{number\ of\ unscheduled\ outages + 1} \tag{4}$$

System Utilization

System utilization is the percent of time that the system's computational nodes run user jobs. No adjustment is made to exclude any user group, including staff and vendors (reference formula #5).

$$SU = \left(\frac{core\ hours\ used\ in\ period}{core\ hours\ available\ in\ period} \right) * 100$$

(5)

APPENDIX E. DIRECTOR'S DISCRETIONARY PROJECTS ENABLED (AT ANY POINT) IN CY 2019

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Jacob King	Tech-X Corporation	0	0	200,000	0	GPUelegant Performance
Eric Nielsen	NASA Langley Research Center	0	0	2,000,000	26,428	FUN3D GPU Development
Allan Grosvenor	Microsurgeonbot, Inc.	0	0	4,000,000	0	NextGen Low Cost Propulsion: Dual-Expander Cycle LOX[liquid oxygen]/Methane Rocket Engine Supporting Research
John Schaefer	Boeing	0	0	0	0	Numerical Prediction of Nose Landing Gear Noise
Nathan See	SmartTruck	0	0	3,000,000	842,578	Fully Non-linear Unsteady CFD and Application to Integrated Truck-Trailers Systems
Daniel Combest	ENGYS	0	0	0	0	Wind Tunnel Geometry Effects on Drag Measurements
Zhi Wang	University of Kansas	0	0	2,000,000	2,486,697	Wall resolved large eddy simulation of JAXA's [Japan Aerospace Exploration Agency] high-lift configuration
Ngoc-Cuong Nguyen	MIT	20,000	10,079	6,000	0	Massively Parallel Discontinuous Galerkin Methods for Wall-resolved Large Eddy Simulation of Transonic Aeroelasticity
Pierre Ocvirk	Universite de Strasbourg, Strasbourg Astronomical Observatory	0	0	2,000,000	2,035,223	Reionization and its Impact on the Local Universe: Witnessing our own Cosmic Dawn
Evan Schneider	Princeton	0	0	0	0	Revealing the Physics of Galactic Winds with Petascale GPU Simulations
Bronson Messer	ORNL	0	0	0	0	Analysis and Visualization for Turbulence Studies in Core-collapse Supernovae
Brian O'Shea	Michigan State University	0	0	1,000,000	0	Scaling and performance enhancement of an astrophysical plasma code
Lorenzo Sironi	Columbia University	0	0	2,000,000	17,730,556	The interplay of relativistic reconnection and turbulence in astrophysical plasmas
Daniel Eisenstein	Harvard University	0	0	0	0	Cosmological N-body Simulations with Abacus
Christian Cardall	ORNL	0	0	2,000,000	1,179	Towards 3D Core-collapse Supernova Ensemble Studies with GenASiS
Steve Barnet	University of Wisconsin	0	0	1,000,000	0	Management and Operation of the IceCube Neutrino Observatory (ICNO)

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Daniel Eisenstein	Harvard University	5,000	5,176	0	0	Abacus2020: N-body Simulations for Precision Cosmology with DESI
Brian O'Shea	Michigan State University	30,000	21,273	0	0	Measuring Performance and Scaling of Kokkos-accelerated Athena++ on Summit
William Fox	Princeton Plasma Physics Laboratory	20,000	22	10,000,000	9,161,755	Energetics of Collisionless Plasmas in the Laboratory and Space
Pierre Ocvirk	Universite de Strasbourg, Strasbourg Astronomical Observatory	15,000	1,516	0	0	Evaluation of Astrophysical Radiation-hydrodynamics Galaxy Formation Codes RAMSES-CUDATON and EMMA
Brant Robertson	UC Santa Cruz	20,000	6,399	0	0	Preparing Cholla for Cosmological Simulations and In-Situ Visualization on Summit
Philipp Moesta	University of Amsterdam	20,000	3	0	0	Dynamical Space-time GRMHD Simulations of Neutron-star Mergers and Remnants
Patrick Kilian	LANL	5,000	1,476	0	0	Chick-Keller-Postdoc
Ruonan Wang	ORNL	20,000	15,682	0	0	Simulating Full-scale SKA Phase I Dataflow
Robert Fisher	University of Massachusetts Dartmouth	20,000	0	0	0	Explorations of the D6 Scenario of Type Ia Supernovae on Summit
Jessie Carman	NOAA-GFDL	0	0	2,000,000	328,559	Air-Ocean-Land-Ice Global Coupled Prediction on Emerging Computational Architectures: A Framework for ESPC [Earth System Prediction Capability] Coupled Models
Inanc Senocak	University of Pittsburgh	0	0	1,000,000	1,131,874	GEM3D: Open-Source Cartesian Adaptive Complex Terrain Atmospheric Flow Solver for GPU Clusters
Katherine Evans	ORNL	30,000	28,078	0	0	Aim High: Air Force R&D Collaboration
Minsu Joh	Korea Institute of Science and Technology Information (KISTI)	15,000	16,380	0	0	ATMOS - Advancement of Typhoon Prediction Models on Summit
Richard Loft	University Corporation for Atmospheric Research (UCAR)	15,000	3,273	0	0	Increasing MPAS-A's [Model for Prediction Across Scales – Atmosphere] GPU Production Readiness at Convection Resolving Scales
Daniel Jacobson	ORNL	0	0	0	0	Center For Bioenergy Innovation
Daniel Jacobson	ORNL	0	0	4,000,000	3,599,220	Scaling Up Parallelized Ortholog Detection Algorithms for Comparative Genomics of Bacterial Genomes
Kjiersten Fagnan	LBNL	0	0	0	0	JGI Data Archive

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Victor Padilla-Sanchez	LANICUDE	0	0	0	0	Molecular Modelling and Visualization Viruses
Yanling Liu	Leidos Biomedical Research for Frederick National Lab for Cancer Research	20,000	8	0	0	Systematic Annotation Creation on H&E [hematoxylin and eosin] WSIs [whole-slide images] for Automated Digital Pathology Informatics
Burkhard Rost	Technical University of Munich	30,000	54,004	0	0	LSTL I : Self-Supervised Deep Learning for Protein Sequences
Jan Michael Carrillo	ORNL, University of Tennessee, Knoxville (UTK)	0	0	2,000,000	189,629	Large Scale Coarse-Grained Molecular Dynamics Simulations of Lipid Bilayers
Loukas Petridis	ORNL	0	0	4,000,000	334,167	Integrating Neutron Scattering with Molecular Simulation to Determine Structural Ensembles of Flexible Biosystems
Edward Lyman	University of Delaware	0	0	2,000,000	0	Hydrodynamics of complex membranes on subcellular scales
Philip Kurian	Howard University	10,000	5,731	650,000	308,786	Computing Many-body Van Der Waals Dispersion Effects in Biomacromolecules
Christopher Bahl	Institute for Protein Innovation	0	0	0	0	Design of GPCR [G-protein-coupled receptors] Allosteric Modulators
Martin Karplus	Harvard University	20,000	3,492	8,000,000	4,574,608	Computational Design of HIV [Human Immunodeficiency Virus] Vaccination Schedule
Teresa Head-Gordon	University of California (UC) Berkeley	0	0	1,000,000	0	Enhanced Sampling Methods for Biomolecular Simulation
Alison Sweeney	University of Pennsylvania	0	0	2,500,000	1,759,313	Self-Assembly of Reflectin Proteins at Membrane Interfaces
Peter Coveney	University College London	25,000	234,756	0	0	INSPIRE
Arvind Ramanathan	ANL	20,000	19,566	0	0	RL-Fold: Artificial Intelligence Guided Molecular Simulations for Targeting Intrinsically Disordered Proteins
Harel Weinstein	Cornell	20,000	20,411	0	0	Structural Elements in the Evolution of Functional Mechanisms in Eukaryotic Substance Transporters
Juan Perilla	University of Delaware, University of Illinois Urbana-Champaign (UIUC)	10,000	48,718	0	0	Revealing the Molecular Mechanisms of the Late Stages of the HIV-1 Infection Cycle Through the Computational Microscope
Diwakar Shukla	UIUC	15,000	10	0	0	Mechanisms of Neurotransmitter Transporters Via Deep Mutagenesis-informed Modeling
George Karniadakis	Brown	20,000	2,831	0	0	Supervised Parallel-in-time Algorithm For Long-time Lagrangian Simulations of Stochastic Dynamics

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Marcos Sotomayor	Ohio State University	8,000	10,567	0	0	Revealing the Dynamics and Mechanics of Large Cadherin Assemblies
Cecilia Clementi	Rice University	25,000	18,026	0	0	Extensible and Scalable Adaptive Sampling on Summit
Sumanta Acharya	Illinois Tech	0	0	500,000	0	Indirect Dry Cooling for Power Plants Using Encapsulated Phase Change Materials
Yue Ling	Baylor	0	0	2,000,000	0	Direct Numerical Simulation of Spray Formation in a Two-phase Mixing Layer Between Two Parallel Gas and Liquid Streams
Jesse Ault	ORNL	0	0	2,000,000	774,424	Using Parallel Computation to Improve Blood Cell Simulations in Cardiovascular Flows
David Gutzwiller	Numeca, Numeca International	0	0	2,000,000	367,984	Porting and Demonstration of the NUMECA Omnis Environment for Multidisciplinary Simulation and Optimization.
Stephan Priebe	GE	0	0	3,500,000	2,844,625	Genesis LES Studies
David Buchta	UIUC	0	0	0	0	Effects of Floating Marine Debris on Nonlinear Wave Dynamics
Marc Olivier Delchini	ORNL	0	0	0	0	Nuclear Energy University Program (NEUP) - Computational Fluid Dynamics Benchmark Analysis of an Historical Sodium Loop Experiment
Charlotte Barbier	ORNL	0	0	0	0	Cavitation Bubbles near Solid Surfaces
Carlos Velez	GE	20,000	8,908	0	0	GENESIS GPU Scalability Studies on Summit
Emilian Popov	ORNL	0	0	3,000,000	730,615	HFIR [High Flux Isotope Reactor] STAR-CCM+ Simulations
Bamin Khomami	UTK	20,000	39,578	4,000,000	4,272,877	Elucidating the Molecular Rheology of Entangled Polymeric Fluids via Direct Comparison of NEMD Simulations and Model Predictions
Eric Johnsen	University of Michigan	10,000	865	0	0	Cavitation Erosion in the Spallation Neutron Source (SNS)
Freddie Witherden	Texas A&M	10,000	251	0	0	PyFR and ZEFR Scaling
Sanjeeb Bose	Cascade Technologies	15,000	2,095	0	0	Porting and Optimization of Large Eddy Simulation Flow Solvers
Ramesh Balakrishnan	ANL	10,000	0	0	0	Performance Modeling of CFD Solvers on Heterogeneous Computing Platforms
Pete Bradley	Pratt & Whitney	5,000	54	0	0	UTCFD [United Technologies Computational Fluid Dynamics] Code Tuning and Hackathon prep
Vyacheslav Bryantsev	ORNL	0	0	500,000	0	Rational Design of Flotation Agents for Rare Earth Ore Minerals

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Remco Havenith	University of Groningen	0	0	3,000,000	507,347	GRONOR [University of Groningen and Oak Ridge National Laboratory] Nonorthogonal Configuration Interaction Software - GPU Development and Benchmarking
Byeongjin Baek	SABIC	0	0	2,200,000	3,217,813	Catalyst Screening for Hydrogenation Processes
Srikanth Allu	ORNL	0	0	1,000,000	3,433	Consortium of Advanced Battery Simulation
Kwangho Nam	University of Texas at Arlington	0	0	500,000	0	Multiscale Modeling of Kinase Conformational Change
Stephan Irle	ORNL	0	0	2,000,000	4,330	GPU-accelerated Quantum Chemical Ensemble-Parallel Nanomaterials Growth Simulations on Microsecond Timescales
Wibe de Jong	LBNL	0	0	2,000,000	30	The chemical universe through the eyes of deep learning neural networks
Olexandr Isayev	University of North Carolina at Chapel Hill	0	0	8,500,000	9,145,833	Learning Quantum Mechanics (QM) with Neural Networks
Arun Yethiraj	University of Wisconsin	10,000	5,320	2,000,000	1,625346	Polymers in Deep Eutectic Solvents
Jason Byrd	ENSCO	0	0	2,000,000	3,426,481	Scaling and performance of Aces4 when applied to very large environmental chemical pollutants
Vyacheslav Bryantsev	ORNL	0	0	1,000,000	226	Atomic-Scale Simulations of Bulk and Interfacial Processes in Molten Salt Environment
Filip Pawlowski	Auburn University	0	0	4,000,000	361,194	Coupled-cluster Singles-and-doubles Quality Excitation Energies for Large Molecular Systems
Yu-Hang Tang	LBNL	0	0	1,000,000	0	Design and optimization of high-performance and high-throughput solvers for marginalized graph kernels
Konstantinos Vogiatzis	UTK	0	0	1,000,000	15,892	Examination of Large Molecular Databases for Lanthanide Separation with Artificial Intelligence
Maria Ramos	University of Porto	4,000	0	2,000,000	850,243	Local and Collective Motions as a Gateway to the Catalytic Power of Enzymes
Pedro Fernandes	University of Porto	4,000	0	2,000,000	537,402	Enzyme Instantaneous Disorder as a Key Player in the Catalytic Power of Enzymes
Stephan Irle	ORNL	20,000	0	0	0	Benchmarks for the GPU-accelerated density-functional tight-binding code DFTB+ [density-functional tight-binding] on Summit
Stephan Irle	ORNL	30,000	0	0	0	GPU-accelerated ensemble-parallel simulations of ion solvation, complexation, and dynamics and nonlinear vibrational spectroscopy

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Coen de Graaf	ICREA	20,000	98,448	0	0	Energy and Charge Transfer by Nonorthogonal Configuration Interaction
Peter Coveney	University College London	25,000	231	0	0	IMPRESS
Remco Havenith	University of Groningen	0	0	0	0	Porting Valence Bond Code TURTLE
Matthew Ryder	ORNL	10,500	0	0	0	Vibrational Dynamics of Organic Materials
Lin Lin	UC Berkeley	20,000	3,621	0	0	GPU-Accelerated Plane Wave Pseudopotential Density Functional Theory With PWDFT
Sotiris Xantheas	Pacific Northwest National Laboratory (PNNL)	25,000	204,549	0	0	Excited States of DNA [deoxyribonucleic acid] / RNA [ribonucleic acid] Oligomers
Lucas Visscher	Vrije Universiteit Amsterdam	27,600	29	0	0	PRECISE
Ashley Shields	ORNL	20,000	1,584	0	0	Atomistic Bridges to Carbon Defects @ Exascale
Allison Talley	Proctor & Gamble Company	0	0	2,000,000	595	Investigation of anomalous NMR [nuclear magnetic resonance] cross relaxation signals of surfactant aggregates via molecular dynamics simulations
Philippe Sautet	University of California, Los Angeles	0	0	2,000,000	217,461	Modelling Catalysis at Bimetallic Surfaces
Timmy Ramirez-Cuesta	ORNL	20,000	0	0	0	Development of SNS-Spectroscopy Leadership-Computational Methods
Moetasim Ashfaq	ORNL	0	0	8,000,000	35,916,001	The Computational Climate Science Integrated Allocation
Chris Hill	MIT	1,000	0	0	0	Climate Modeling Alliance (CliMA) - Ocean Modeling
Forrest Hoffman	ORNL	20,000	0	0	0	Development of a Land Model Testbed (LMT) for Rapid Assessment and Benchmarking of Multiscale Complex Biogeochemistry in Earth System Models
Moetasim Ashfaq	ORNL	0	0	0	0	9505 Assessment for the US Congress and ORNL's Contribution to IPCC-AR6
Jitendra Kumar	ORNL	15,000	0	0	0	Scalable Machine Learning for Earth System Science
Thomas Robinson	Science Applications International Corporation (SAIC)	1,500	7	0	0	GFDL [Geophysical Fluid Dynamics Laboratory] Atmosphere Model 4.0 Porting and Validation
Valentine Anatharaj	ORNL	0	0	100,000	0	Provisioning of Climate Data

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Gary Cai	Fiat Chrysler Automobiles	0	0	0	0	A Fundamental Study of the Factors Affecting Adverse Autoignition (Knock) in Internal Combustion Engines
Seung Hyun Kim	Ohio State University Research Foundation	0	0	1,000,000	1,098,830	Development of a Physics-Based Combustion Model for Engine Knock Prediction: Phase II
Xinyu Zhao	University of Connecticut	0	0	900,000	0	Turbulence and Flame Resolved Fire Simulations
Venkat Raman	University of Michigan	0	0	2,000,000	0	High-fidelity Simulations of Rotating Detonation Engines
Venkatramanan Raman	University of Michigan	5,000	27,202	0	0	High-fidelity Simulations of Rotating Detonation Engines
Myoungkyu Lee	Sandia	10,000	3,718	0	0	Molecule-Informed Continuum Direct Numerical Simulation Method in Non-Equilibrium Combustion
Panchapakesan Ganesh	ORNL	0	0	2,000,000	159	Data Driven Discovery by Design of Energy Materials
Andreas Glatz	ANL	0	0	2,000,000	0	OSCon
Anton Kozhevnikov	CSCS	0	0	0	0	Porting and Testing - Accelerated Data Analytics and Computing Institute
Gabriel Kotliar	Rutgers	0	0	2,000,000	751	Application of GPU-accelerated quantum Monte-Carlo impurity solver to plutonium compounds
Prineha Narang	Harvard University	0	0	500,000	1,422,447	Ab initio Exploration of Hydrodynamic Transport in Weyl and Dirac Semimetals
Anton Kozhevnikov	ETH Zurich	5,000	169	0	0	GPU acceleration of Quantum ESPRESSO and CP2K codes
Cristian Batista	UTK	25,000	0	0	0	Transport Properties and Dynamics of Quantum Materials
Panchapakesan Ganesh	ORNL	20,000	266	0	0	Defects, Interfaces and Disorder in Correlated Quantum Materials
Amir Haji-Akbari	Yale University	20,000	0	0	0	Computational Investigation of Collective Phenomena in Soft Matter and Biological Systems
Jean-Luc Fattebert	ORNL	5,000	34	0	0	Robust Efficient Algorithms for Density Functional Theory (DFT) Calculations
Mauro Del Ben	LBNL	10,000	2,696	0	0	BerkeleyGW - Electron Excited-state Properties in Materials
Corey Melnick	University of Michigan, BNL	25,000	14	0	0	Application of GPU-accelerated quantum Monte-Carlo impurity solver to plutonium compounds
Neena Imam	ORNL	30,000	26,318	9,000,000	31,583	Durmstrang
Sergey Panitkin	University of Montreal	0	0	1,000	57,081,660	Next Generation Workload Management System

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Scott Klasky/Norbert Podhorszki	ORNL	70,000	596	30,000,000	33,517,083	ADIOS - The Adaptable IO System
Joshua New	ORNL	0	0	1,000,000	812,978	Big Data Mining for Building Analytics
Manuel Arenaz	Appentra	0	0	50,000	0	Porting Parallware Tool to Large HPC Installations Including Titan
Catherine Schuman	ORNL	0	0	2,500,000	92,158	Scalable Evolutionary Optimization for Designing Networks
Ramakrishnan Kannan	ORNL	0	0	1,000,000	108,121	Mini-Apps for Big Data
John Cavazos	University of Delaware	0	0	1,000,000	0	Large-Scale Distributed and Deep Learning of Structured Graph Data for Real-Time Program Analysis and Characterization
Jakub Kurzak	UTK	0	0	0	0	SLATE - A Modern Software Library For Dense Linear Algebra
Catherine Schuman	ORNL	0	0	2,000,000	78,962	Scalable Neuromorphic Simulators: High and Low Level
Sunita Chandrasekaran	University of Delaware	0	0	20,000	0	Migrating Legacy Code to Novel Directive-based Programming Models
Robert Brook	UTK	0	0	0	0	JICS [Joint Institute for Computational Sciences] /AACE [Application Acceleration Center of Excellence] Emerging Architectures Activities
Allan Grosvenor	Microsurgeonbot, Inc.	0	0	1,500,000	0	Intelligent Middleware, Making CFD Accessible Through Leadership-Scale Deep Learning Research
Terry Jones	ORNL	15,000	0	1,000,000	0	ECP: Simplified Interface for Complex Memories
Judith Hill	ORNL	0	0	500,000	64	Computational Science Graduate Fellowship Program
Ramakrishnan Kannan	ORNL	0	0	0	0	Graph500 for OLCF
Ramakrishnan Kannan	ORNL	0	0	1,000,000	44,198	Scalable Non-linear Unmixing
Seung-Hwan Lim	ORNL	0	0	3,000,000	0	GPU-accelerated High Dimensional Data Management for ML Workloads in Large Scale GPU-CPU Environments
Helia Zandi	ORNL	0	0	2,500,000	0	Smart Grid Transactive Scalable Hybrid Management System
Jean-Roch Vlimant	Caltech	0	0	2,000,000	19,548	large scale deep learning training & optimization
Peter Zaspel	University of Basel	0	0	2,000,000	206,096	Extreme-scale linear solvers in high-dimensional approximation, machine learning (ML) and beyond
Karan Vahi	University of Southern California (USC)	0	0	100,000	88	Pegasus Project

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Robert Patton	ORNL	0	0	2,000,000	0	Scalable Machine Learning for Connected Autonomous Vehicles
Michael Seal	Department of Defense	0	0	0	0	Autonomy Capability Team 3
Shinjae Yoo	BNL	0	0	0	0	Scalable Spatio-Temporal Learning
Song Han	MIT	40,000	39,261	0	0	Neural Architecture Search for Efficient Visual Understanding
Mark Coletti	ORNL	15,000	34,736	0	0	Deep-learning Optimization Via Evolutionary Algorithm
Sergey Panitkin	University of Montreal	25,000	24,329	0	0	Porting the ATLAS Experiment Software and Workload Management System
Audris Mockus	UTK	5,000	0	2,000,000	1,958,517	Behavioral Fingerprinting for Identity Resolution
Ramakrishnan Kannan	ORNL	15,000	0	200,000	305,886	Parallel Low-rank Approximation with Nonnegative Constraints (PLANC)
David Keyes	King Abdullah University of Science and Technology (KAUST)	20,000	15,941	0	0	Asynchronous High-Performance Task-based Reverse Time Migration
Rick Archibald	ORNL	10,000	7	0	0	Streaming Data Reduction Using Machine Learning
Bryan Goodman	Ford Motor Company	10,700	0	0	0	ML Performance Evaluation of High Performance Computing System
Markus Rampp	Max Planck Computing and Data Facility	10,000	169	0	0	Porting, Benchmarking, and Optimization of MPCDF Codes
Prasanna Balaprakash	ANL	30,000	0	0	0	DeepHyper: Scalable neural architecture and hyperparameter search for deep neural networks
Srikanth Yoginath	ORNL	20,000	1,113	0	0	Accelerating End-to-end Prediction of Physical Phenomena by Interleaving Analytics with Multiscale Simulations
Timothy Williams	ANL	40,000	69	0	0	Portable Application Development for Next Generation Supercomputer Architectures - Argonne Supplement
Ewa Deelman	USC	500	30	0	0	Pegasus Workflows
Edmon Begoli	ORNL	1,000	108	0	0	Citadel Pilot
Laxmikant Kale	UIUC	20,000	363	0	0	CharmRTS
Rio Yokota	Tokyo Institute of Technology	10,000	18	0	0	Large-Batch Data-Parallel Deep Learning
Kurt Keutzer	UC Berkeley	15,000	4,004	0	0	Large Scale Training of Neural Networks
Wen-mei Hwu	UIUC	20,000	366	0	0	Petascale 3D Image Reconstruction with Ptycho-Tomography
Robert Patton	ORNL	20,000	2	0	0	Machine Learning for Connected Autonomous Vehicles

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Azzam Haidar	NVIDIA Corporation, UTK	150,000	126,928	0	0	Using NVIDIA Tensor Cores to Drive the Summit Toward Exascale
Seung-Hwan Lim	ORNL	20,000	2,239	0	0	GPU-accelerated High Dimensional Data Management for ML Workloads in Large Scale GPU-CPU Environments
Rogerio Feris	IBM	20,000	4,715	0	0	Large-scale Neural Architecture Search for Visual Recognition
Yasuhiro Idomura	JAERI	20,000	6,463	0	0	Exascale CFD Simulations at JAEA
Robert Patton	ORNL	30,000	52,587	0	0	Scalable Machine Learning of Scientific Data
Dmitry Pekurovsky	UC San Diego	5,000	53	0	0	Scalable Software Framework for Multidimensional Fourier Transforms
Mark Shephard	Rensselaer Polytechnic Institute	16,500	5	0	0	Measuring the Scalability of Nvidia Multi-Process Service to Improve GPU Throughput in Multiscale Biological Simulations
Allan Grosvenor	Microsurgeonbot, Inc.	20,000	3	0	0	Expertise-as-a-Service via Scalable Hybrid Learning: R&D supporting improvements to hybrid intelligence agent tooling
Sudip Seal	ORNL	20,000	1,795	0	0	Scaling Studies of Machine Learning Workloads on Summit
Bronson Messer	ORNL	80,000	69,462	0	0	CAAR for Frontier
Arghya Das	University of Wisconsin	20,000	0	0	0	Accelerating Large-Scale Metagenomic Analysis with POWER system
Thomas Uram	ANL	1,000	0	0	0	Balsam - An HPC Job Campaign Management Framework
Shantenu Jha	Rutgers	20,000	16	0	0	RADICAL-Cybertools: Middleware Building Blocks for Workflows
David Gutzwiller	Numeca, Numeca International	25,000	19,423	0	0	Porting and Testing of the NUMECA FINE/Open CFD Solver on OLCF Summit
Ramanan Sankaran	ORNL	20,000	57	0	0	Simulation of Transport Phenomena in Molten Media Reactors
Andrew McGough	Newcastle University	20,000	0	0	0	NUFEB: Newcastle University Frontiers in Engineering Biology
Ryan Viertel	Sandia	15,000	0	0	0	Optimization Based Design for Manufacturing
David Green	ORNL	0	0	100,000	0	Sparse Grid Solver for Fusion
Dennis Youchison	ORNL	0	0	0	0	Fusion Blanket MHD Modeling
George Fann	ORNL	0	0	100,000	15,987	Optimizing High-order Time-stepping for CGYRO on SUMMIT
Stuart Loch	Auburn University	0	0	2,000,000	5,530,533	High Performance Computation in Support of Neutron Star Mergers and High-Z Plasma Facing Components
Mark Kostuk	General Atomics	9,000	5,929	0	0	ALMA - Plasma Fluid Simulations

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Stuart Loch	Auburn University	20,000	154,164	0	0	High performance computation in support tungsten erosion studies on fusion energy experiments
Christopher Holland	University of California, San Diego	15,000	8,633	0	0	Determination of Multiscale Gyrokinetic Simulation Scaling Performance Using CGYRO
David Smith	University of Wisconsin	2,000	0	0	0	Deep Learning Analysis of GPI Data to Advance Tokamak Edge Physics
Ashley Barker	ORNL	5,000	1	0	0	NCCS UA User Testing Emulation Project
Christine Goulet	USC	0	0	4,000,000	11,111,675	Extreme-Scale Simulations for Advanced Seismic Ground Motion and Hazard Modeling
Mark Coletti	ORNL	10,000	27,567	0	0	Accelerated¬†Global¬†Human¬†Settlement Discovery
Yidong Xia	Idaho National Laboratory (INL)	0	0	2,000,000	3,676,835	Dissipative Particle Dynamics Modeling of Multiphase Flow and Transport in Nanoporous Shale Pore-Networks
Tsuyoshi Ichimura	University of Tokyo	0	0	0	0	Fast & Scalable Finite Element Method with Low-order Unstructured Elements for Earthquake Simulation on Summit
Ethan Coon	ORNL, LANL	20,000	183	0	0	ExaSheds
Yidong Xia	INL	20,000	7,558	0	0	Multiphase Flow Simulations in Mesoporous Source Rocks
Ralph Dunlap	UCAR	0	0	0	0	Earth System Modeling Framework
Kohei Fujita	University of Tokyo	20,000	18,584	0	0	Fast and Scalable Implicit Unstructured Finite-Element Earthquake Simulations
James McClure	Virginia Tech	20,000	51,619	0	0	Modeling Wetting Phenomena in Porous Media
Katrin Heitmann	ANL	0	0	1,000,000	121,677	Sky Surveys
Eric Church	PNNL	0	0	0	0	Algorithms for Innovative Data Analysis
Brian Nord	Fermilab	0	0	2,500,000	2,014,791	Machine Learning for Astrophysics and Cosmology
Alan Stagg	Sandia	25,000	0	0	0	Z-Machine Target Physics Simulation
Brian Nord	Fermilab	15,000	0	0	0	Machine Learning for Astrophysics and Cosmology
Paolo Calafiura	LBNL	15,000	304	0	0	Exa.TrkX, HEP Tracking at Exascale
Kehfei Liu	University of Kentucky	10,000	22,427	0	0	Lattice Calculation of Nucleon Structure
Rajan Gupta	LANL	25,000	50,132	0	0	Matrix Elements Using Lattice QCD [quantum chromodynamics] for Nucleon EDM [electric dipole moment] and Structure
Claudia Ratti	University of Houston	5,000	396	0	0	Janos - Porting and Testing Lattice QCD Simulation Code

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Katherine Riley	ANL	16,000	0	0	0	Advanced Materials Characterization with AI-informed Computation
Katherine Riley	ANL	16,000	90	0	0	Dynamic Compressed Sensing for Real-time Tomographic Reconstruction
Katherine Riley	ANL	3,000	0	0	0	Developing High Performance Computing Applications for Liquid Argon Neutrino Detectors (Szelc)
Liqun Zhang	Tennessee Technological University (Tennessee Tech)	0	0	4,000,000	3,711,619	Molecular Dynamics Simulations to Investigate the Structure, Dynamics and Functional Properties of Human Beta Defensin Type 3
Darrin York	Rutgers	0	0	0	0	Computational Tools for High-throughput Lead Optimization
Amanda Randles	Duke	20,000	1,377	0	0	Coupled FSI [Fluid-structure interaction]/CFD Simulations
Daniel Jacobson	UTK	10,000	0	0	0	Wound Healing Machine Learning
Byung Park	ORNL	0	0	5,000,000	13,715,280	Accelerating Materials Modeling Loop of Leadership Computing and Spallation Neutron Source
Dongwon Shin	ORNL	0	0	4,000,000	3,937,246	High Performance Cast Aluminum Alloys for Next Generation Passenger Vehicle Engines
Yangyang Wang	ORNL	0	0	3,000,000	60,016	Elucidating the Influence of Reversible Non-Covalent Interactions on Dynamic Properties for Rational Design of Soft Materials
Rick Archibald	ORNL	0	0	2,000,000	0	CADES/OLCF Computational Workflows for Materials Science
Bobby Sumpter	ORNL	0	0	10,500,000	4,477,527	Center for Nanophase Materials Sciences (CNMS)
Lin	UC Berkeley	0	0	1,000,000	749	GPU-Accelerated Plane Wave Pseudopotential Density Functional Theory With PWDFT
Miguel Fuentes-Cabrera	ORNL	0	0	1,000,000	550,876	Molecular Dynamics Simulations of Liquid Metal Assembly at the Nanoscale
James Morris	ORNL	0	0	2,000,000	39,682	HPC4mfg - High-throughput calculations of light-weight alloys
Zhiting Tian	Virginia Tech	0	0	1,000,000	862,482	Thermal Transport Properties of Two-Dimensional Polymers
Mohsen Asle Zaeem	University of Missouri	0	0	3,000,000	2,782,356	Predicting Nano/Microstructures and Properties in Solidification of Metals
Roberto Longo	Texas A&M	0	0	1,500,000	0	First Principles Analysis and Design of Materials for Catalysis and Energy
Prineha Narang	Harvard University	0	0	250,000	434,699	Ab initio Predictions of Spatially-resolved Nonequilibrium Transport Phenomena in Quantum Systems
Vikram Gavini	University of Michigan	15,000	240,556	0	0	Large-scale Discrete Fourier Transform Simulations with DFT-FE [finite element]

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Trung Nguyen	Northwestern University	15,000	25	0	0	Development of New Models and Sampling Techniques for LAMMPS [Large-scale Atomic/Molecular Massively Parallel Simulator]
Jan Michael Carrillo	ORNL, UTK	20,000	22,687	0	0	Bead-Spring Polymer Melt LAMMPS Benchmark
Yangyang Wang	ORNL	20,000	2,723	0	0	Exploring New Paradigms for Understanding Ionic Transport in Polymer Electrolytes
Ivan Oleynik	University of South Florida	10,000	0	0	0	Predictive Simulations of Phase Transitions in Dynamically Compressed Materials
Lou Kondic	New Jersey Institute of Technology	5,000	304	0	0	Exploiting Non-equilibrium Processes, Free Surface and Substrate Effects to tailor Phase Separation in Metal Alloys: a Molecular Dynamics Study
Bobby Sumpter	ORNL	90,000	27,719	0	0	Center for Nanophase Materials Sciences
Vikram Gavini	University of Michigan	22,750	594,145	0	0	Large-scale Discrete Fourier Transform Simulations with DFT-FE
Zhiting Tian	Cornell	15,000	941	0	0	Thermal Transport Properties of Covalent Organic Frameworks (COFs)
Robert Szilagyi	Montana State University	15,000	0	0	0	Quantum Chemical Refinement of Periodic Structural Models for Zeolite-Templated Carbon Materials
Jiook Cha	Columbia University	0	0	2,000,000	0	High-throughput Brain Imaging Analysis in Alzheimer's Disease Related Dementia
Hong-Jun Yoon	ORNL	0	0	1,000,000	0	LungXNet: Scalable AI for High Throughput Multi-Labeled Medical Image Analysis
Joel Saltz	SUNY Stony Brook	13,000	0	0	0	Comprehensive Morphology Analysis of Whole Slide Tissue Specimens
Bamin Khomami	UTK	0	0	3,000,000	0	Multiscale Modeling
Xinlian Liu	Hood College	0	0	0	0	Rapid Molecular Machine Learning through Dimensionality Reduction
Torsten Hoefler	ETH Zurich, UIUC	30,000	46,450	0	0	Investigation of Self-heating Effects in Nanoscale Transistors
Askin Guler Yigitoglu	ORNL	0	0	3,000,000	0	Nuclear-Renewable Hybrid Energy Systems
Igor Sfiligoi	General Atomics	0	0	10,000	0	BFG3D
Elia Merzari	Pennsylvania State University	10,000	112	0	0	High-Fidelity Multi-Physics Simulation of Pebble bed cores
Cole Gentry	ORNL	10,000	8	0	0	Machine-Learning-Based Multi-Physics Nuclear Reactor Core Simulations of Molten Salt Reactor

PI	Institution	Most Recent Titan Allocation	Titan Usage	Most Recent Summit Allocation	Summit Usage	Project Name
Benjamin Collins	ORNL	0	0	1,000,000	0	Multiphysics Calibration of Crud Deposits in Pressurized Water Reactors
Rajan Gupta	LANL	0	0	100,000	45	Probing Novel Physics using Nucleon Matrix Elements
Leah Broussard	ORNL	0	0	100,000	0	Neutron Electric Dipole Moment Experiment at SNS
Leah Broussard	ORNL	0	0	100,000	0	nEDM [neutron electric dipole moment] at SNS
Eric Church	PNNL	15,000	6,701	0	0	NEXT - Neutrino Experiment in a Xenon Time Projection Chamber
Benjamin Cumming	CSCS	0	0	0	0	Arbor
Josh McDermott	MIT	25,000	0	0	0	Next-Generation Hearing Aids Via Neural Network Models of the Auditory System
Satrajit Ghosh	MIT	10,000	0	0	0	Creating Baseline DNN Models for Biomedical Signals
Prineha Narang	Harvard University	0	0	3,000,000	2,232,310	Atom-by-Atom Quantum Defect Dynamics from First Principles
Gary Grest	Sandia	10,000	22,391	0	0	Large Scale Numerical Simulations of Polymer Nanocomposites
Brajesh Gupt	Xanadu	0	0	3,000,000	1,414,200	Demonstrating Quantum Supremacy with Titan
Edison Liang	Rice University	0	0	1,500,000	0	Magnetized Jet Creation Using a Ring of Laser Beams
Itay Hen	USC	0	0	2,000,000	0	Exploring Error Mechanisms in Analog Quantum Computers
Salvatore Mandra	NASA Ames	60,000	44,427	500,000	60	Benchmark and Calibration of NISQ Devices
Claudio Gheller	École polytechnique fédérale de Lausanne	15,000	5,390	0	0	Accelerating the ORB5 code on cutting-edge hybrid HPC architectures
Pui-kuen Yeung	Georgia Tech	0	0	15,000,000	28,526,535	High Resolution Study of Intermittency in Turbulence and Turbulent Mixing
Duane Rosenberg	Colorado State University	0	0	100,000	68	GPU Performance Analysis for Turbulence Codes
Arvind Mohan	LANL	7,000	0	0	0	Machine Learning for Turbulence
Mujeeb Malik	NASA Langley Research Center	20,000	4,972	0	0	Direct Numerical Simulations of Complex Turbulent Flows at High Reynolds Number