INL/EXT-21-65486

Nuclear Science User Facilities High Performance Computing

December 2021

FY 2021 High Performance Computing Overall Systems Availability

M3UF-22IN0701012

Report Team:
Matthew W. Anderson, HPC Manager
Stephanie J. Parker, HPC Account & User Coordinator

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory
operated by Batelle Energy Alliance, LLC

INL/EXT-21-65486

Nuclear Science User Facilities High Performance Computing

FY 2021 High Performance Computing Overall Systems Availability
M3UF-21IN0701012

December 2021

Idaho National Laboratory
High Performance Computing
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

CONTENTS

FIGURES

TABLES

Page intentionally left blank

ACRONYMS

C3	Collaborative Computing Center
CDU	cooling distribution unit
DOE-NE	Department of Energy Office of Nuclear Energy
EDR	Endpoint Detection and Response
GPU	graphics processing unit
HPC	high performance computing
INL	Idaho National Laboratory

1. INTRODUCTION

Idaho National Laboratory (INL), supported by the Department of Energy Office of Nuclear Energy (DOE-NE) through the Nuclear Science User Facilities, provides access to supercomputer systems and data storage along with support staff for system management, software installation, cybersecurity, and user support to the broader DOE-NE user community. Users include individuals at universities, industry, and government laboratories, enabling a wide range of research and development and mission-supporting activities. The availability of high-performance computing (HPC) capabilities is a key foundation of collaboration and innovation in nuclear energy systems research. HPC resources and INL directly support the mission and objectives of DOE-NE.

In Fiscal Year 2021, INL HPC capabilities were utilized by a diverse set of computing and applied researchers, for a wide range of research and engineering activities. This report focuses on current INL HPC system availability from October 1, 2020 through September 30, 2021, as well as including some information related to Fiscal Year 2022 maintenance and planning.

2. CURRENT HPC SYSTEMS

INL currently houses four HPC systems in the Collaborative Computing Center (C3), a facility that includes space for up to 197 computer equipment racks. These systems are Sawtooth, Hoodoo, Lemhi, and Falcon; all are available for use by individuals at universities, industry, and government laboratories. Table 1 provides a description of each system.

INL HPC also support three separate storage systems that are fundamental to system availability: IBM ESS/GPFS, Dell/EMC Isilon, and Dell/EMC Disaster Recovery Isilon. These are described in Table 2. While these systems do not provide compute capability to the users, they are foundational to the operation of the compute systems listed in Table 1 and provide all disk storage needed for running compute jobs, as well as archiving simulation results for postprocessing and additional research. The storage systems require frequent maintenance; their availability is critical to overall HPC system availability.

INL HPC provides a web-based science gateway for user access to HPC systems based on Open OnDemand. This enables users to access the HPC systems through their computer browser. Critical security updates are needed for this system more frequently than the HPC systems themselves, resulting in a limited outage that only affects the web gateway while the HPC systems continue to operate.

All these systems have an impact on DOE-NE's ability to deliver results at a much larger scale. The Sawtooth, Lemhi, and Falcon systems combined delivered more than 1.026 billion core hours of compute time in FY-21. This is an increase of 396 million core hours from FY-20. The compute time observed each year continues to exceed our operational goals.

Table 1. Current INL HPC systems.

System	Description
Sawtooth	A Hewlett Packard Enterprise SGI 8600 distributed-memory system: 99,792 cores, 395 TB of total memory, high-speed EDR interconnect network, high-speed storage, and graphics processing unit (GPU) capabilities. The current LINPACK rating for Sawtooth is more than 6 petaflops from both central processing units and GPUs.
Hoodoo	Machine learning only system. Lambda Hyperplane deep-learning distributed-memory system: 44 NVIDIA A100 tensor core GPUs and 5.5 TB of total memory. The system provides a maximum GPU performance of 0.4 petaflops.
Lemhi	Dell PowerEdge distributed-memory system: 20,160 cores, 94 TB of total memory, Omni-Path interconnect network, and high-speed storage. The LINPACK rating for Lemhi is 1.002 petaflops.
Falcon	An SGI ICE X distributed-memory system: 34,992 cores, 121 TB of total memory, high-speed FDR interconnect network, and high-speed storage. The current LINPACK rating for Falcon is 1.087 petaflops.

Table 2. Current INL HPC data storage systems.

Storage System Vendor	Capacity and Description of Primary Use
DELL/EMC Isilon	2.11 PB of data storage. This system holds all user home directories and project directories.
DELL/EMC Isilon DR	1.32 PB of data storage. This is the disaster recovery storage system and is housed outside the C3 datacenter. This system reached end-of-life in June 2021.
IBM ESS/gpfs	1 PB of data storage. This system hosts the scratch directory for the HPC systems and is used for high-speed storage requirements; it hosts files and data for less than 90 days before deletion.

3. FY-21 PLANNED OUTAGES

In FY-21, INL HPC had nine planned outages, as shown in Table 3.

Table 3. Planned Outages in FY-21.

Outage Dates	Systems Affected	Purpose
19 Oct – 22 Oct 2020	Sawtooth	Address critical security updates; upgrade PBS scheduler; cooling distribution units (CDUs) were serviced
18 Nov 2020 (6 hours)	All HPC systems	HPC firewall migration
25 Jan – 28 Jan 2021	Falcon	Critical security updates
9 Feb 2021 (4 hours)	All HPC systems	Network migration
9 Mar – 10 Mar 2021	IBM ESS/gpfs (scratch)	Critical security updates
12 Apr – 16 Apr 2021	All HPC systems	Critical security updates; transition to IPA for LDAP authentication
1 June 2021	IBM ESS/gpfs (scratch)	Critical security updates
10 Aug 2021	All HPC systems	Live patching; no downtime resulted
20 Sep – 22 Sep 2021	HPC OnDemand (science gateway)	Critical security updates.

The total time spent in planned outages was 466 hours. The total time spent in planned outages where all the HPC systems were unavailable was 154 hours. A high-level overview graph of system utilization of FY-21 is provided in Figure 1.

Figure 1. A high-level overview graph of the HPC system utilization in FY-21. The HPC systems are highly utilized and experienced very little downtime during FY-21. Downtime and gaps of availability are designated as the white space where the graph drops to zero.

4. FY-21 UNPLANNED OUTAGES

In FY-21 there were eight unplanned outages, as shown in Table 4.

Table 4. Unplanned outages in FY-21.

Outage Dates	Systems Affected	Details
23 Oct 2020	Sawtooth	The planned outage time for Sawtooth had to be extended by one day due to software quality assurance testing issues.
24 Oct 2020	Sawtooth	Sawtooth ran at reduced capacity due to a water leak discovered in the CDU serviced on October 22, 2020.
18 Nov 2020	Sawtooth	Sawtooth racks shut down again due to CDU water and pump issue.
25 Nov 2020	Lemhi	The PBS job scheduler server shutdown resulting in an unplanned Lemhi outage.
5 Jan 2021	All HPC systems	HPC systems continued to operate, but network connections to the HPC systems were slow. INL IM Networking Team resolved the issue.
24 Feb 2021	Sawtooth	Sawtooth compute nodes were brought down due to a process water cooling pump failure.
12 Jul 2021	All HPC systems	During maintenance by Schneider Electric on the building backup generator, all HPC systems lost power.
12 Aug 2021	IBM ESS/gpfs	Slow scratch speeds were observed for a few hours.

The most severe of the unplanned outages was the complete power outage on July 12, 2021 caused by technicians attempting to troubleshoot a fault code on the Kohler backup generator for the building. This caused all HPC systems to experience an outage. The only other unplanned outage that affected all HPC systems was on January 5, 2021 in which network access to all systems was impaired but the HPC systems themselves did not go down. The total number of hours in which all HPC systems were down in unplanned outages was less than 24. The total number of hours in which any HPC system was affected by an unplanned event was 192 hours.

5. STATE OF DATACENTER POWER AND COOLING SYSTEMS

The datacenter contains multiple power and cooling systems requiring monthly maintenance. This maintenance work provides greater confidence in the power backup and cooling subsystems and in general does not impact HPC system availability. However, in FY-21, there were two maintenance activities that raised the operational risk level of the HPC systems; one of those resulted in an unplanned HPC outage.

Kohler generator tests occur monthly for 45 minutes where the entire datacenter power load is transferred to the generator. In May 2021, power at the C3 datacenter dropped four times in 30 days with the latest bump occurring on Friday May 14, 2021. While some non-HPC systems housed in the datacenter experienced power loss, there was no loss to power to the HPC systems during this time. On June 23, 2021, a fault appeared on the backup generator that required the involvement of Schneider Electric Engineers and Mountain West Electric. During testing of the generator fault code on July 12, 2021, power was unexpectedly lost to all HPC systems in the datacenter. Subsequently, the generator fault code was cleared, and the issue was resolved. The datacenter power backup systems have been working properly since, and there have been no other full-scale unplanned outages since.

On September 15, 2021, Cooling Tower Pump 8 failed, leaving only one pump (Pump 7) operational. The cooling tower undergoes annual preventive maintenance in October, during which time Pump 8 was also removed for repair (see Figure 2). Pump 8 was repaired and reinstalled on December 8, 2021. From September 15 to December 8, 2021, the datacenter cooling systems ran with a higher operational risk than normal due to the removal of Pump 8 for maintenance but experienced no failures in that time period.

Figure 2. Pump 8 was removed for repair on October 18, 2021, leaving only one operational pump in the cooling tower water basin.

The Sawtooth system employs immersive cooling to ensure temperature control and the water flow and temperature is managed via four different CDUs. During FY-21, CDU 1 developed a leak, resulting in a loss of 8 ounces of cooling water every 28 hours (see Figure 3). Concurrently, CDU 3 began showing very high readings of copper in the water that were an order of magnitude higher than in the other three CDUs (see Figure 4) and suggestive of a high level of component corrosion or pending failure. A chemical treatment was performed on the water and full maintenance was done to the CDUs on November 15, 2021 to repair the leak, flush the system and thereby reduce the operational risk for the immersive cooling system on Sawtooth.

Figure 3. Leak of cooling water in CDU 1. Picture taken November 8, 2021.

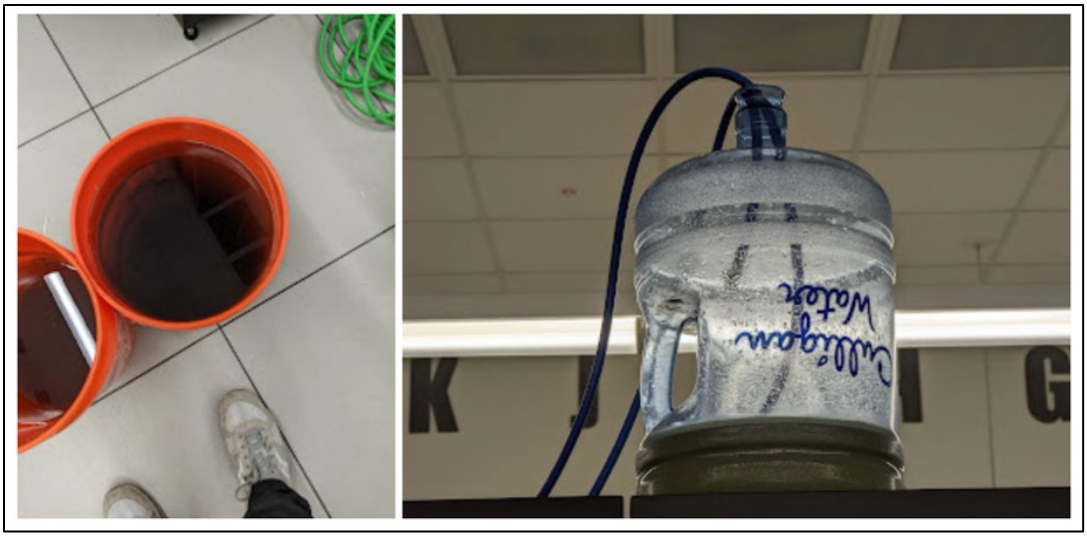

Figure 4. The treated cooling water appears green from the Sawtooth CDUs, reflecting a high copper content indicative of component corrosion.

6. FY-22 PLANNED MAINTENANCE WINDOWS

Users are notified of planned maintenance windows at least two weeks prior to the downtime. Maintenance windows in FY-22 will occur roughly every 3 months to address critical security updates but may be more frequent to address any high security vulnerability.

7. SUMMARY

INL HPC capabilities were utilized by individuals at universities, industry, and government laboratories for a wide range of research and engineering activities between October 1, 2020 and September 30, 2021. The HPC systems themselves experienced minimal downtime over the course of the fiscal year. Most downtime was planned to apply critical security patches to system software. The most serious of the unplanned downtime events occurred due to maintenance on the Kohler backup generator in July 2021. Users were kept notified of all events, both planned and unplanned, via 36 total email notifications in FY-21.

The INL HPC systems continue to provide high availability for collaboration and innovation in nuclear energy systems research. Despite the various downtime incidents, the HPC systems provided an amazing amount of compute time to users – 1 billion core hours. This incredible accomplishment was highlighted by DOE in the annual PEMP report.