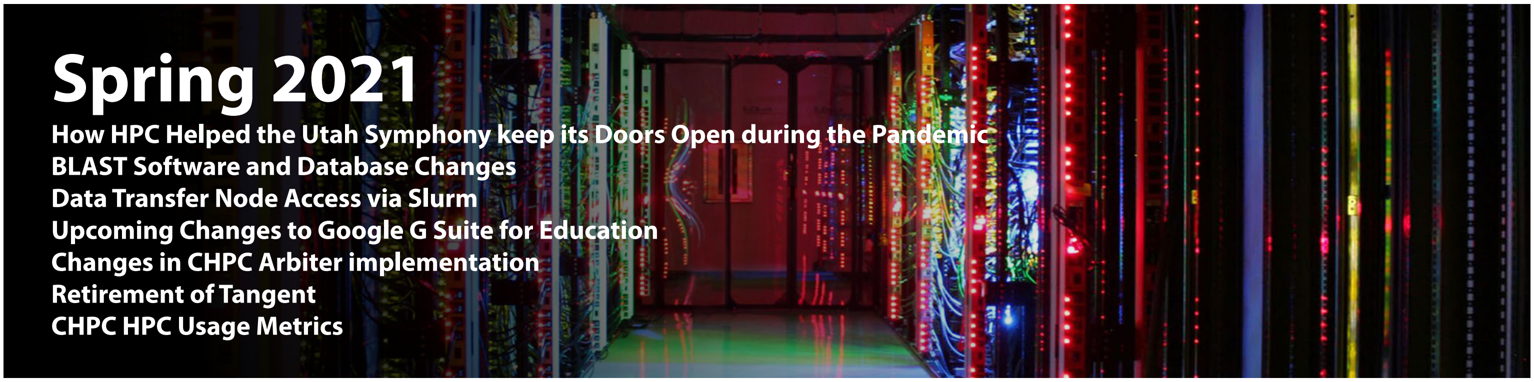
Center for
HIGH PERFORMANCE COMPUTING
THE UNIVERSITY OF UTAH



How HPC Helped the Utah Symphony Keep its Doors Open during the Pandemic

Tony Saad and James C. Sutherland, Department of Chemical Engineering

On June 23, 2020, we received an email from Professor Dave Pershing, former president of the University of Utah, and currently a faculty member in Chemical Engineering. As any person would do when seeing the president's name in their inbox, we dropped everything and read the email carefully. Professor Pershing was asking if we could help the Utah Symphony/Utah Opera (USUO) to analyze the dispersion of airborne droplets emitted from wind instruments at Abravanel Hall (and later Capitol Theater). Thinking of this problem from an engineering perspective, and based on our knowledge of how viral transmission works, a virus attaches itself to a respiratory droplet which is subsequently exhaled into the air. Although large droplets generally settle and lead to surface contamination, small "aerosolized" droplets become suspended in the air and move with it. This means that these aerosols can be modeled as a tracer in a fluid flow simulation. We were excited to help as this aligns closely with our expertise in Computational Fluid Dynamics (CFD).

The performing arts were significantly affected during the COVID-19 pandemic. According to a Brookings Institute report which examined the impact of the COVID-19 pandemic on the arts in the United States , those hit hardest were the "fine and performing arts" such as choirs, orchestras, operas, and dance companies. Estimated losses for that sector were over 50% of jobs lost, approximately 1.3 million jobs. In addition, there was a staggering 42.5 billion dollars in lost sales. The estimates are much higher worldwide. In the United States, at the time of writing, only the Columbus, Dallas, Fort Worth, Houston, Jacksonville, and Utah orchestras have returned to live performances out of 117 large orchestras and symphonies in the United States. It is not clear how many of the 1,224 total orchestra symphonies in the United States have returned to their live performances. These statistics are not surprising, given how SARS-COV-2 is spread. The vehicle for airborne transmission of infectious diseases, such as COVID-19 and influenza viruses, is primarily exhaled liquid droplets or their dried "nuclei." The spread of droplets carried by a directional airflow can reach more than two meters, and COVID-19 infection may occur in the following methods: (a) the exhalation of infectious droplets by sneezing,coughing, singing, talking, or playing wind instruments and (b) by direct contact with contaminated surfaces or body parts. These observations highlight the importance of the airflow dynamics in transmitting the virus in specific environments. Any analysis that does not take into consideration the airflow dynamics is not complete.

The Role of High Performance Computing

Thanks to the generosity of CHPC in granting us over 600K CPU hours (~68 CPU years!), we were able to run over 25 simulations in total using our in-house code, Wasatch - a component of the Uintah Computational Framework. Our calculations consisted of different scenario variations and mitigation strategies and resulted in several Terabytes of data. Each calculation consumed an average of 24K CPU hours and took about 5 days to complete (~20 mins of real time). We used standard packages provided by CHPC such as MPI, HYPRE, and Boost. The code used a second order finite volume discretization (space and time) on a structured grid with timesteps ranging from 0.1 to 0.0001 s.

Mitigation Strategies

The first step in our analysis was to understand the baseline configuration of the airflow created by the HVAC in Abravanel Hall's stage along with a proposed seating arrangement for the orchestra. We found significant accumulation of respiratory droplets in the stage area, indicating

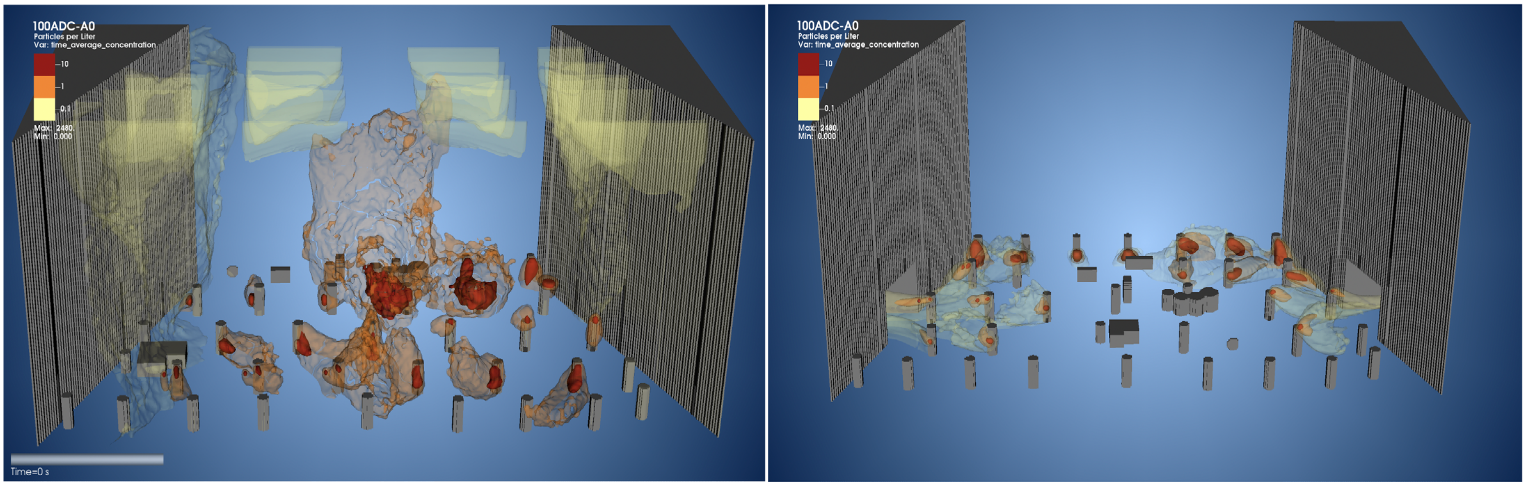


Figure 1: Time average particle concentration (Particles/L) at Abravanel Hall for Baseline arrangement (left) and Modified arrangement with open doors (right). Colors range from 0.1 (yellow) to 10 (red) particles per liter.

an increased risk of infection (Figure 1a, Figure 2a). Note that USUO was planning on building a "buffer" zone between the stage and the house and therefore the risk on the audience was not taken into consideration. To mitigate the accumulation of droplets in the baseline configuration, our team considered two "low-cost" mitigation strategies: (1) increasing the volume of air leaving the hall, and (2) rearranging the location of instruments so that super emitter and spreader instruments are located closer to return/exit vents. Combining these led to a decrease in particle concentrations by a factor of 100 (Figure 1b, Figure 2b). This was a significant improvement over the baseline configuration. The USUO leadership was very pleased with the analysis that we provided and it was taken into account as part of their overall strategy. What we know for a fact is that, at the time the analysis was done, the Utah Symphony was one of two large orchestras that were open in the United States.

A similar analysis was conducted for the Capitol theater where we adopted a similar low cost mitigation strategy by opening two doors on the back of the stage. However, because of the way Capitol theater's HVAC is designed, we had to design a "plenum" to extend the reach of the open doors further into the stage area. We also observed a significant reduction in aerosol accumulation.

As a summary, because HVAC systems are unique to each venue, it is generally difficult to draw a one-size fits all recommendation and an analysis must be conducted per venue. However, two general guidelines can be taken from our work: (1) relocate super emitters and spreaders closer to return/exit vents, and (2) increase the number of exit vents by opening windows and doors. It is a simple and effective strategy to help orchestras keep their doors open.

More information on this work can be found on this youtube channel:

https://youtube.com/playlist?list=PLEaLl6Sf-KICe-92KALDvwxryzztv_EGYo

and on this KUER interview:

https://www.kuer.org/arts-culture-religion/2020-09-25/music-and-covid-in-the-air-scientists-model-airflow-on-the-abravanel-hall-stage-to-assess-risk

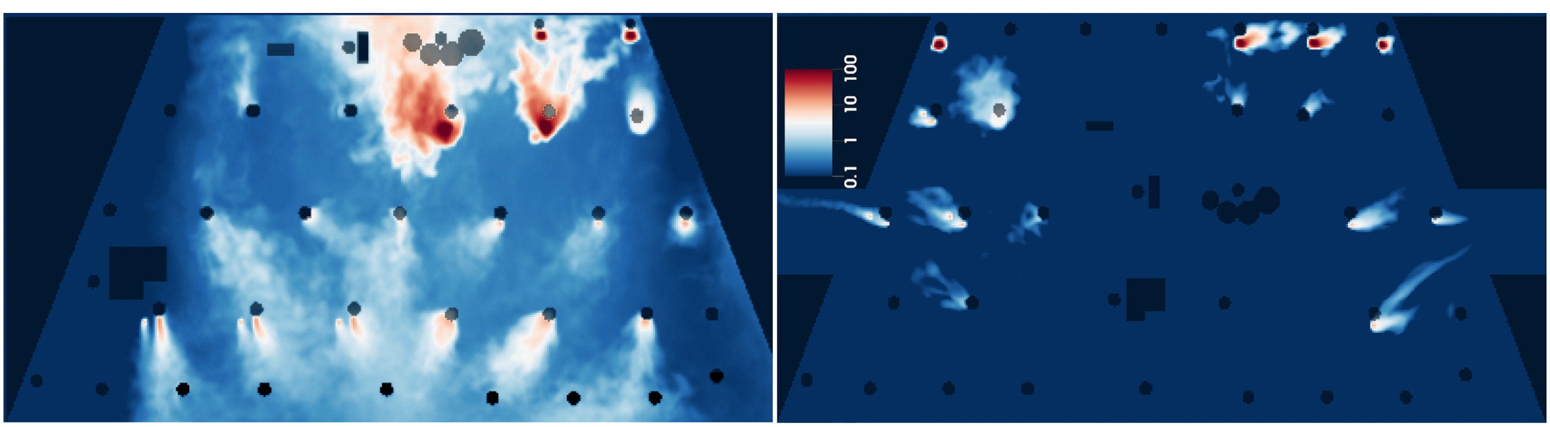


Figure 2: Time and space average particle concentration (Particles/L) in the breathing zone of the Abravanel Hall stage for Baseline arrangement (left) and Modified arrangement with open doors (right). Colors range from 0.1 (yellow) to 100 (red) particles per liter.

BLAST Software and Database Changes

Brett Milash, CHPC Scientific Consultant

The Basic Local Alignment Search Tool (BLAST) distributed by the National Center for Biotechnology Information (NCBI) at the NIH is an essential tool for searching biological sequence databases. This software continues to evolve and improve. The latest version of BLAST (2.11.0) requires a new sequence database format, labeled v5, which is incompatible with previous versions of the software. Along with the move to the v5 format NCBI has changed the names and content of some of the databases they distribute. As we host the databases here at the CHPC, in the weeks to come we will install a set of v5 BLAST sequence databases that are as similar as possible in content to the databases we support for older versions of BLAST, with the goal of deprecating the older software and databases later this year. The table below summarizes the database changes we anticipate.

If a database that is critical to your work is being discontinued, and we haven't planned for a suitable replacement, please contact us at *helpdesk@chpc.utah.edu* so we can incorporate an appropriate replacement into our plan.

v4 database	Status	v5 database	Description
16SMicrobial	replaced	16S_ribosomal_RNA	16S ribosomal RNA (Bacteria and Archaea type strains)
env_nr	no change	env_nr	Environmental samples, protein
env_nt	no change	env_nt	Environmental samples, nucleotide
est_human	discontinued	none	
est_mouse	discontinued	none	
est_others	discontinued	none	
gss	discontinued	none	
gss_annot	discontinued	none	
htgs	discontinued	none	
human_genomic	replaced	human_genome	Human GRCh38.p12 genome assembly (replaces the unassembled v4 human genomic database)
nr	no change	nr	Non-redundant protein sequences
nt	no change	nt	Non-redundant nucleotide sequences
other_genomic	discontinued	none	
pataa	no change	pataa	Protein sequences derived from the Patent division of GenBank
patnt	no change	patnt	Nucleotide sequences derived from the Patent division of GenBank
refseq_genomic	replaced	ref_euk_rep_genomes, ref_prok_rep_genomes	Refseq representative eukaryotic and prokaryotic genomes
refseq_protein	no change	refseq_protein	NCBI Protein Reference Sequences
refseq_rna	no change	refseq_rna	NCBI Transcript Reference Sequences
sts	discontinued	none	
swissprot	no change	swissprot	Non-redundant UniProtKB/SwissProt sequences
tsa_nt	no change	tsa_nt	Transcriptome shotgun assembly nucleotide sequences
vector	discontinued	none	

RMACC Women in HPC Chapter

The Rocky Mountain Advanced Computing Consortium (RMACC) is pleased to announce the formation of a Women in HPC (WHPC) chapter. The chapter meets virtually the 3rd Wednesday of each month at 3pm (Mountain Time). The chapter has a *YouTube channel* with recordings of past presentations. If you would like to be added to the WHPC mailing list or would like any additional information about the chapter, please email *rmacc@colorado.edu.*

Data Transfer Node Access via SLURM

Brett Milash, CHPC Scientific Consultant
Sam Liston, CHPC Storage System Administrator

CHPC has long maintained high performance computing clusters with compute nodes made available via the SLURM job scheduler, and has also offered high performance data transfer nodes (DTNs) with access to the same file systems as the computing clusters. Now for the first time, CHPC's data transfer nodes can be accessed through the SLURM job scheduler. CHPC Senior Systems Administrators Brian Haymore and Sam Liston have configured four data transfer nodes within the notchpeak cluster and two data transfer nodes within the redwood cluster in the protected environment for access via SLURM. While each of the dtn nodes have 24 cores and 128 GB of RAM, only 12 cores and 96 GB of RAM are made available to run SLURM jobs.

Cluster	SLURM Partition	SLURM Account	Nodes
Notchpeak	notchpeak-dtn	dtn	dtn05, dtn06, dtn07, dtn08
Redwood	redwood-dtn	dtn	pe-dtn03, pe-dtn04

The notchpeak data transfer nodes have 100 gigabit per second connections to the University's Science DMZ, a segment of the university network with streamlined data flow across the campus firewall to and from off-campus locations. In the redwood cluster, these nodes are connected at 40 gigabits per second.

The notchpeak-dtn and redwood-dtn SLURM partitions are similar to other shared SLURM partitions at CHPC, with multiple transfer jobs sharing a node. By default, each SLURM job running on a data transfer node is allocated a single core and 2 GB of memory. The notchpeak-dtn and redwood-dtn quality of service (QOS) has a maximum time limit of 72 hours per job, so transfers must complete within that time period.

Both sets of new schedulable DTNs mount a CephFS filesystem: 44 TB in capacity in the general and 22 TB capacity in the PE. Each filesystem is built entirely on fast NVME drives and was designed as a staging area for transfers. If your workflow or transfer could benefit from staging your data to these filesystems please be aware that due to their smaller capacities they will be aggressively scrubbed. Currently the scrub policy on these filesystems is 14 days. They are mounted at /scratch/general/dtn and /scratch/general/pe-dtn respectively, and are symlinked to /scratch/local.

Not all data transfer applications can take advantage of the high network bandwidth provided by the data transfer nodes – you should test your data transfer application to determine whether running it on a data transfer node yields better performance, and which file system yields the best performance. Our scratch file systems are excellent choices as the source or destination of high-speed data transfers.

We are currently testing the new ability, but anticipate allowing access to these slurm accounts to all users by early April. Watch for a CHPC announcement on this.

Setting up a basic slurm script to do only a download is straightforward. Using the appropriate account and partition from the table above, you need to navigate to the directory containing the data being transferred from CHPC, or in the case of moving data to CHPC, you need to create and move to the directory into which you will be moving data. This is followed by the transfer by whichever mechanism you wish. Here is a quick example, using wget to download a file:

```
#!/bin/tcsh
#SBATCH --partition=notchpeak-dtn
#SBATCH --account=dtn
#SBATCH --time=1:00:00
#SBATCH -o slurm-%j.out-%N
#SBATCH -e slurm-%j.err-%N
#
setenv SCR /scratch/general/lustre/$USER/$SLURM_JOB_
ID
mkdir -p $SCR
cd $SCR
wget https://www1.ncdc.noaa.gov/pub/data/uscrn/prod-
ucts/daily01/2020/CRND0103-2020-AK_Aleknagik_1_NNE.
txt”
```

For redwood, you would only need to change the partition and the scratch file system in the above example. For parallel transfers, users can request the required number of cores and memory using #SBATCH directives.

You can also build a transfer using the SLURM-capable data transfer nodes into a snakemake workflow, by creating a snakemake rule responsible for downloading the data:

```
rule download_data:
    output: “CRND0103-2020-AK_Aleknagik_1_NNE.txt”
    message: “Downloading data file {output}.”
   shell: “wget https://www1.ncdc.noaa.gov/pub/data/
uscrn/products/daily01/2020/{ouput}”
```

A typical cluster configuration file for a snakemake workflow would execute all the workflow rules on compute nodes:

```
# cluster.yaml - cluster configuration for snakemake
workflow.
__default__:
    cluster: notchpeak
```

```
    partition: notchpeak-shared
    account: jones
    nodes: 1
    ntasks: 1
    time: 01:00:00
```

But with these additional lines we can direct the execution of the download_data rule to a data transfer node:

```
download_data:
    partition: notchpeak-dtn
    account: dtn
```

You can see this entire workflow at *https://gitlab.chpc.utah.edu/bmilash/workflows-with-snakemake*.

For more information about our data transfer nodes please see our documentation here: *https://www.chpc.utah.edu/documentation/data_services.php#Data_Transfer_Nodes*.

Upcoming Changes to Google G Suite for Education

Anita Orendt, CHPC Scientific Consultant

For a number of years, University of Utah faculty, staff and students have had access to unlimited storage available via the University's G Suite for Education (renamed Google Workspace for Education) agreement with Google (*https://gcloud.utah.edu/*).

CHPC has documentation (*https://www.chpc.utah.edu/documentation/software/rclone.php#googledrive*) demonstrating the use of rclone to backup data from CHPC file systems to this gdrive space, and we know a number of CHPC users are taking advantage of this storage for this purpose.

A recent *Google announcement* presented a new storage model. Effective July 2022 for all existing Google Workspace for Education customers, the storage will no longer be unlimited, but be limited to a baseline of 100 TB of pooled storage shared across all users of an institution.

There are still many unanswered questions at this point. We do not have any additional details on how the limits will be managed, nor do we have any details on how institutions who exceed this baseline will be transitioned to the new plan. As we obtain additional information, we will share it via the CHPC user mailing list.

Changes in CHPC Arbiter implementation

Anita Orendt, CHPC Scientific Consultant

In 2018 CHPC implemented a version of the service named Arbiter which enforced limits of cpu and memory usage on the cluster general interactive nodes, currently notchpeak{1,2}, kingspeak{1,2}, lonepeak{1,2}, ash{5,6}, and redwood {1,2}, as well as on the frisco nodes. When users exceed the established usage limits, they are notified and put into "penalty", which further limits the cpu and memory to which the user has access. Once a user has returned to usage within the normal constraints, the penalties are removed and the users are returned to normal mode. Additional details on CHPC's interactive node acceptable use can be found in CHPC's policy manual at *https://www.chpc.utah.edu/documentation/policies/2.1GeneralHPCClusterPolicies.php.*

In the 2018 implementation, Arbiter tracked each of the nodes separately, such that a user's behavior on one node did not impact their usage limits on the other nodes. In a new version of Arbiter, which was taken live on March 24, 2021, this is no longer the case. All of the generally available interactive nodes are put into one of three node groups. The two node groups in the general environment are (1) the frisco nodes (currently frisco1 through frisco8) and (2) the generally available cluster interactive nodes (currently notchpeak{1,2}, kingspeak{1,2}, lonepeak{1,2}, and ash{5,6}. In the protected environment the node group are the generally available PE cluster interactive nodes, currently redwood{1,2}. When a user is put into penalty on one node of the group, you are in penalty on all of the nodes in that group.

Announcing 2021 RMACC HPC Symposium

The 11th annual *Rocky Mountain Advanced Computing Consortium (RMACC)* HPC Symposium will be held virtually May 18-20, 2021. Registration is free *http://rmacc.org/hpcsymposium/registration* and now open.

The largest consortium of its kind, the RMACC is a collaboration among 31 academic and government research institutions in Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Washington and Wyoming. The consortium's mission is to facilitate widespread effective use of high performance computing throughout the intermountain region.

More information on the HPC Symposium can be found at *http://rmacc.org/HPCSymposium*. The full slate of diverse tutorials and panel discussions includes sessions on green practices in HPC; HPC for Scientific Visualization; Machine Learning; and GPU computing.

The symposium includes a *Student Poster Competition* open to graduate and undergraduate engaged in research involving HPC. Poster abstracts are due by 5:00 pm Mountain Time May 12th, with the deadline for the final poster being My 18th. Winners of the poster competition will be awarded an all-expenses paid trip to SC21 in St. Louis, MO.

Retirement of Tangent

Anita Orendt, CHPC Scientific Consultant

CHPC retired the tangent cluster on March 31, 2021. Tangent (*https://www.chpc.utah.edu/documentation/guides/tangent.php*) was the result of a NSF funded collaboration between CHPC and the School of Computing's Flux group led by Rob Ricci, funded to develop an "adaptable profile-driven testbed". For more information on the project see *http://www.flux.utah.edu/project/apt*.

Based on the adaptable profile-driven testbed (APT) framework, the tangent cluster is provisioned dynamically using a CHPC compute node profile based on our HPC image. When jobs are submitted to the tangent on the local to CHPC tangent interactive nodes the system looks for the requested hardware and if available it spins up compute nodes based on this profile and runs the jobs.

Tangent has been in operation since mid-2014, and has provided CHPC users with many compute cycles since its inception. However, due to recent network changes and recent low availability and utilization of the cluster, we have decided that it was time to retire tangent from service.

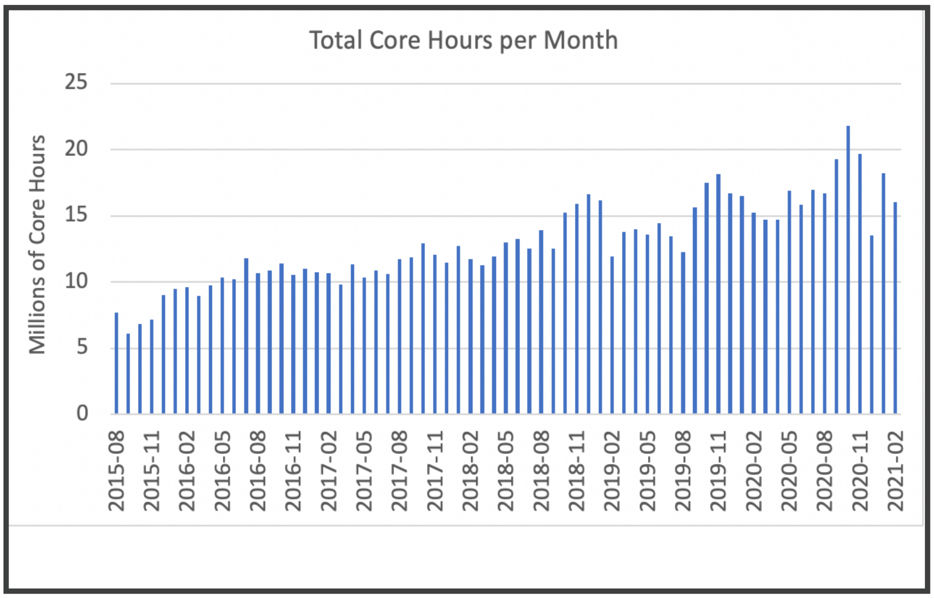
CHPC HPC Usage Metrics

Anita Orendt, CHPC Scientific Consultant

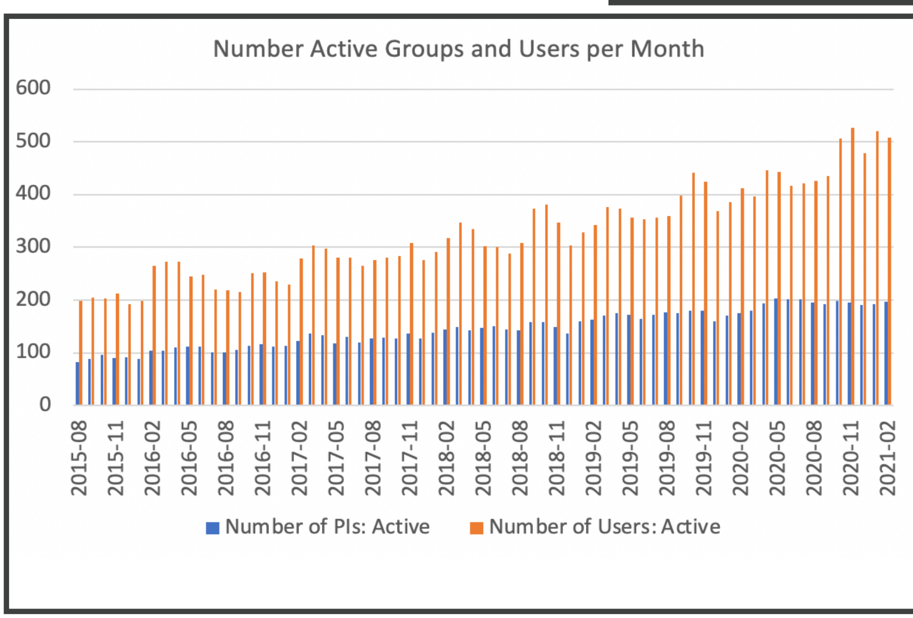
As part of an ongoing project to track the growth of usage of CHPC resources we are collecting a number of metrics.

One set of metrics being collected focuses on measuring the usage of the HPC clusters. With the help of the data collected by XDMoD (note we have both a *general* and a *protected* environment instance of XDMoD running), we have been able to readily track this data since the end of 2015.

As you see from the graphs, CHPC's usage has consistent growth over the last 5+ years.



Above: total core hours on all HPC clusters (both general and protected) used per month



Left: Total number of users and groups who ran at least one batch job during the month

CHPC Presentation Schedule
Summer 2021

DATE	PRESENTATION TITLE	TIME	PRESENTER
May 18	*Overview of CHPC*	1-2PM	Anita Orendt
May 20	*Module Basics*	1-2PM	Anita Orendt
May 25	*Slurm and Slurm Batch Scripts*	1-2PM	Anita Orendt
May 27	*Hands-on Introduction to Open OnDemand*	1-3PM	Martin Cuma
Jun 1-4	*XSEDE HPC Workshop: Summer Boot Camp*	9AM-3PM	XSEDE Webcast
June 8	*Hands on Introduction to Linux, part 1*	1-3PM	Wim Cardoen Martin Cuma
June 10	*Hands on Introduction to Linux, part 2*	1-3PM	Wim Cardoen Martin Cuma
June 15	*Hands on Introduction to Linux, part 3*	1-3PM	Wim Cardoen Martin Cuma
June 17	*Hands on Introduction to Linux, part 4*	1-3PM	Wim Cardoen Martin Cuma
June 22	*Hands-on Introduction to Python, Part 1*	1-3PM	Wim Cardoen Brett Milash
June 24	*Hands-on Introduction to Python, Part 2*	1-3PM	Wim Cardoen Brett Milash
June 29	*Hands-on Introduction to Python, Part 3*	1-3PM	Wim Cardoen Brett Milash
July 1	*Numpy, part 1 (Hands-on Introduction to Python, Part 4)*	1-3PM	Wim Cardoen Brett Milash
July 6	*Numpy, part 2 (Hands-on Introduction to Python, Part 5)*	1-3PM	Wim Cardoen Brett Milash
July 8	*Introduction to R*	1-3PM	Wim Cardoen Brett Milash
July 13	*Introduction to Parallel Computing*	1-3PM	Martin Cuma

Center for
HIGH PERFORMANCE COMPUTING
THE UNIVERSITY OF UTAH

Center for
HIGH PERFORMANCE COMPUTING
THE UNIVERSITY OF UTAH

The University of Utah
University Information Technology
Center for High Performance Computing
155 South 1452 East, Room 405
SALT LAKE CITY, UT 84112–0190

Thank you for using CHPC resources!

Welcome to CHPC News!

If you would like to be added to our mailing list, please provide the following information and via the contact methods described below.

Name:
Phone:
Email:

Department
or Affiliation:

Address:
(campus
or U.S. mail)

Please acknowledge the use of CHPC resources!

If you use CHPC computer time or staff resources, we request that you acknowledge this in technical reports, publications, and dissertations. An example of what we ask you to include in your acknowledgments is:

"A grant of computer time from the Center for High Performance Computing is gratefully acknowledged."

If you make use of the CHPC Protected Environment, please also acknowledge the NIH shared instrumentation grant:

"The computational resources used were partially funded by the NIH Shared Instrumentation Grant 1S10OD021644-01A1."

Please submit copies or citations of dissertations, reports, pre-prints, and reprints in which CHPC is acknowledged in one of the following ways:

Electronic responses

By email: *helpdesk@chpc.utah.edu*
By fax: (801) 585–5366

Paper responses

By U.S. mail: 155 South 1452 East, Rm 405
Salt Lake City, UT 84112–0190

By campus mail: INSCC 405