Center for
HIGH PERFORMANCE COMPUTING
THE UNIVERSITY OF UTAH

Newsletter

Spring 2023

Scalable Adaptive Algorithms for Next-Generation Multiphase Flow Simulations

Masado Ishii, Hari Sundar - School of Computing, University of Utah

Kumar Saurabh, Makrand Khanwale, Baskar Ganapathysubramanian - Iowa State University

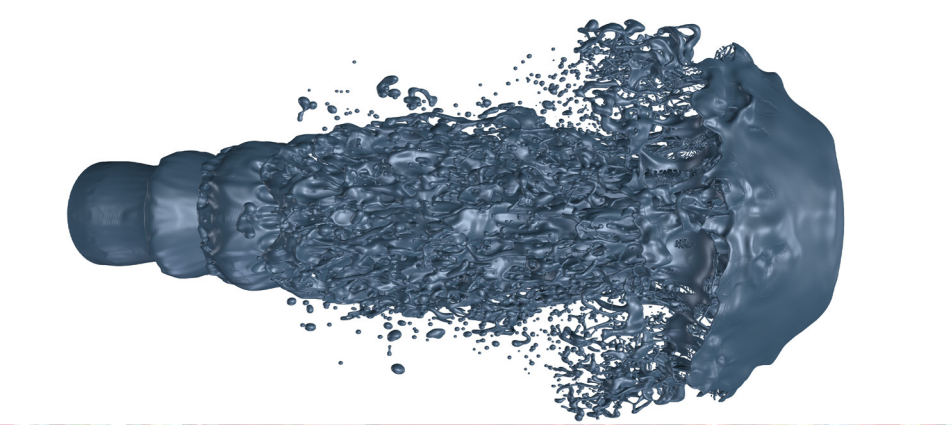


Figure 1: Snapshot of a multiphase flow simulation of primary jet atomization of liquid diesel jet at T = 6.3μs. This simulation contains ~3B unknowns.

Multiphase flows -- more specifically, two-phase flows, where one fluid (e.g., water, paint, melts) interacts with another fluid (usually air) -- are ubiquitous in natural and engineered systems. Examples include natural phenomena from breaking waves and cloud formation to engineering applications like printing, additive manufacturing, and all types of spraying operations in healthcare and agriculture.

High-fidelity modeling of two-phase flows has been an indispensable strategy for understanding, designing, and controlling such phenomena. For instance, insights from high-fidelity modeling of the fluid-fluid interface interactions have produced accurate, low-cost coarse-scale models used to simulate large systems like chemical and biological reactors. High-fidelity simulations also serve as the basis for the optimization-based design of micro-scale systems, with applications in bio-microfluidics and advanced manufacturing.

High-fidelity modeling, specifically through interface-resolving simulations of two-phase flows, is difficult due to the wide range of spatial and temporal scales, especially under turbulent conditions. Such approaches typically require spatially adaptive and temporally higher-order methods to capture the relevant phenomena. Additionally, the size of these simulations requires the development of scalable algorithms. This remains a very active area of research with significant space for improvement. In this article, we briefly describe the landscape of interface-resolving simulations and identify a critical challenge that this paper resolves.

Interface-resolved two-phase modeling can be broadly divided into two main categories --- sharp-interface methods and diffuse-interface methods. The sharp interface methods rely on representing the interface with a sharp, discontinuous function (for instance, the volume of fluids (VOF)), whereas the diffuse interface methods smear out this sharp interface to construct a diffuse, continuous representation of the interface (e.g., conservative diffuse interface, Cahn--Hillard Navier--Stokes (CHNS)).

Sharp interface methods have been the state-of-the-art for high-fidelity simulations of turbulent multiphase flows. These methods involve solving a partial differential equation (PDE) for transporting the discontinuous volume fraction function. An interface reconstruction procedure is generally required to construct normals from the discrete coordinates of the interface typically by fitting splines. The sharp interface methods suffer from artificial breakup -- also known as numerical surface tension, when the interfacial features (droplets/filaments) are comparable in length scale, r, to the grid size δx. There has been some progress in the development of computational techniques that reduces this effect, with good examples being *interface reconstruction* techniques, *moment-of-fluid* methods, and *manifold death* algorithms. These methods, however, remain computationally complex.

Diffuse interface methods, particularly CHNS, avoid any interface reconstruction process. However, diffuse interface methods---especially under low mesh resolution---also suffer from artificial breakup and mass loss, but for completely different reasons. When the thickness of the diffuse interface, ε, becomes comparable to the length scale, r, of the flow features of interest ($\varepsilon / r \sim O(1)$), bound violations of the interface tracking variable and artificial breakup occur. Additionally, these smaller fluid structures can disappear via absorption into nearby larger structures (due to coarsening of Cahn-Hilliard), affecting summary statistics like the number of droplets.

This brief review of state-of-the-art is primarily to make the case that both sharp and diffuse interface approaches suffer from issues of artificial breakup and numerical artifacts emanating from insufficient mesh resolution. Adaptive meshing approaches are an elegant strategy to ensure that the local mesh size is always smaller than the local length scales of interest, thus allowing interface tracking approaches to reliably capture multi-scale features in a computationally efficient manner. However, automatically identifying these local regions of interest is itself non-trivial.

We developed scalable algorithms to identify the spatial regions of interest in the computational domain where the flow features become comparable to the mesh resolution, i.e., regions where $\varepsilon / r \sim O(1)$. This was essential for phenomena exhibiting droplets and fluid filaments, where such targeted resolution is critical for performing cost-effective simulations physics. We also developed octree refinement and coarsening algorithms that can efficiently increase or reduce the level of refinement in a region to accelerate remeshing and decrease the associated overhead, especially for multi-level refinements. This is essential for simulations where the element sizes drop substantially. For instance, in the canonical example of primary jet atomization, element sizes vary by three orders of magnitude to accurately resolve fluid features varying by nine orders of magnitude in volume. This contrasts with existing approaches, where refinement or coarsening of the octrees is done level by level.

We demonstrate the ability of our algorithm to simulate the most resolved simulation of primary jet atomization. Initial development along with testing and validation of our methods were done on the notchpeak cluster at CHPC. The full-scale production run required over 200,000 node hours on TACC Frontera. The finest resolution in our application problem consists of octree at level 15. This is equivalent to solving it on a 35 trillion grid point on a uniform mesh 64 times larger than the current state-of-the-art. Figure 1 shows the simulation result at 6.3μs. We can see that the simulation framework can capture tiny droplets by selectively identifying the key regions of interest and selectively increasing the resolution. Figure 3 shows the progressive refinement of the mesh.

We can see that the algorithm presented in Figure 2 can detect complicated structures like filaments and drops and selectively refine those regions. The overall interface is resolved at level 13, with the key features resolved at level 15. A significant portion of the elements are at levels 13 and 14 (~25%), with a maximum fraction at level 15. A rough estimate indicates that resolving the

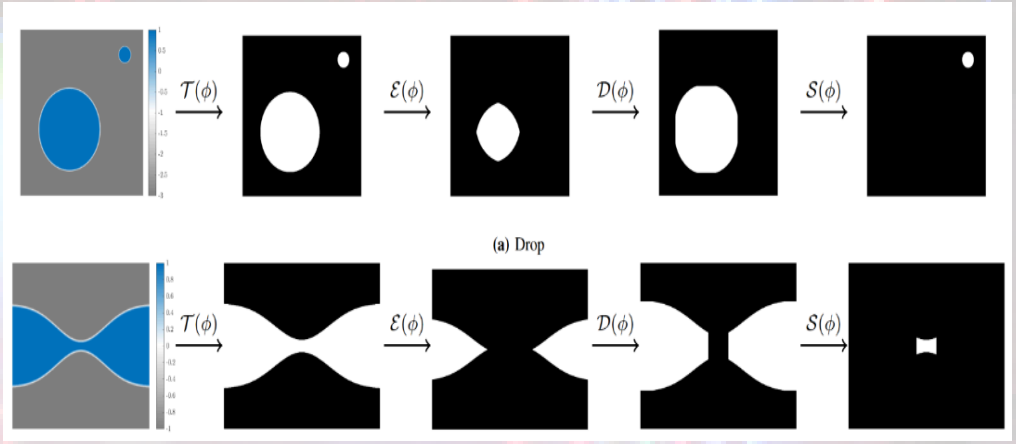


Figure 2: Identification of key regions: Figure showing the identification of key region of interest such as small droplet and a long filament. T(ϕ) denotes the thresholding step, E(ϕ) denotes the erosion, D(ϕ) denotes the dilation, and S(ϕ) denotes the subtraction operator.

complete interface at level 15 would result in at least an 8−10 fold increase in the overall mesh count. This would result in 20−25 fold higher time to solve, making such simulations impractical within a reasonable time.

tion but covers only 0.01% of the total volume. This illustrates the importance of adaptivity in resolving multiphase simulations.

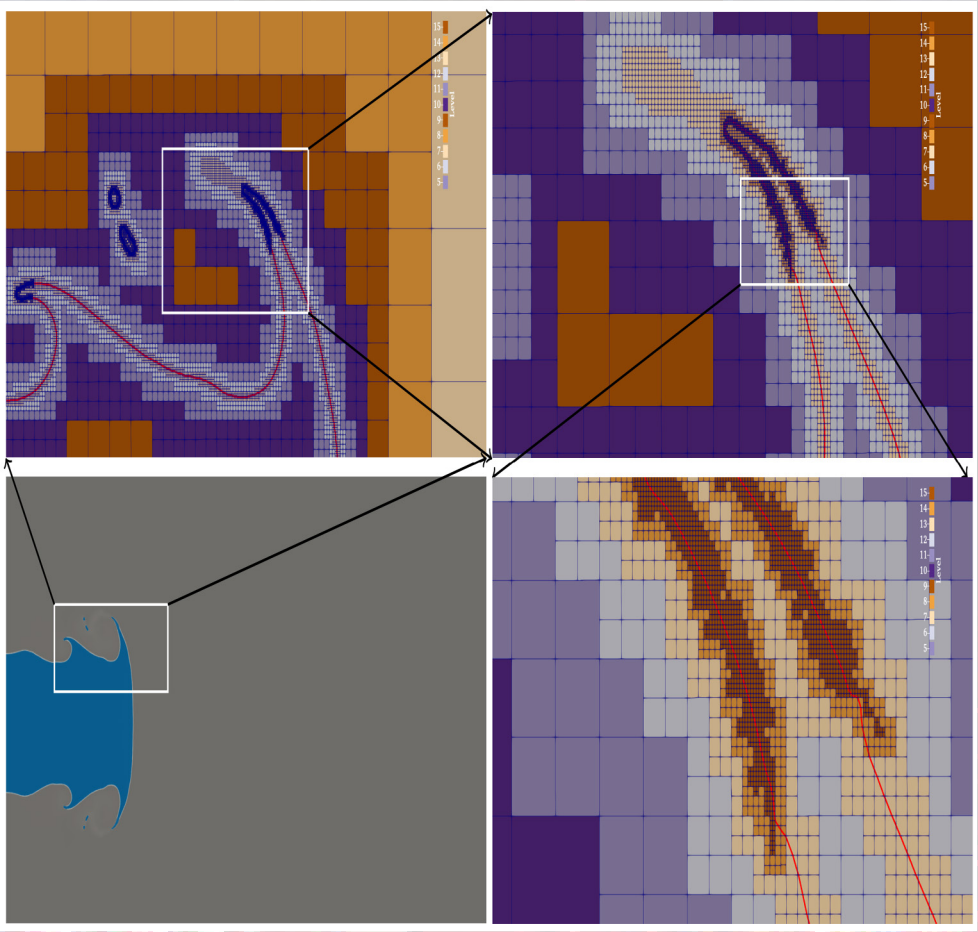


Figure 3: Adaptive mesh refinement: 2D slice with mesh overlayed. The interface is marked by dark red color. Note the tip of the filament (in the bottom right figure) and small bubble (top left) is much more resolved than the other regions of the interface. The octree level differs by 10 levels, with coarsest mesh at level 5 and the finest at level 15, resulting in a (10^9) fold difference in the elemental volume between the finest and the coarsest elements.

Updates to the CHPC Deep Learning Module

Brett Milash, CHPC Scientific Consultant

CHPC's deep learning software module (`deep-learning/2023.3`), introduced in early 2022, has been updated with the latest TensorFlow (2.11), Keras (2.11), and PyTorch (1.13.1) libraries, as well as a recent python release (3.10.9).

An important update to the module has been to expose the pip executable, which is required for installing other python libraries. Pip is available in the earlier deeplearning modules, but it is difficult to access. With the new `deeplearning/2023.3` module loaded the "`pip`" command refers to the pip executable within the module.

We note that level 15 has the maximum element fraction but covers only 0.01% of the total volume. This illustrates the importance of adaptivity in resolving multiphase simulations.

Note that when using pip from the deeplearning module it is necessary to add the "`--user`" or "`--prefix=...`" flag to your command, adding a path to a location in which you have write access for the installation, to direct the software installation to a location outside the read-only image used by the module.

The CHPC documentation about the deeplearning module, *https://bit.ly/CHPC-Deeplearning*, has been updated as well, with links to TensorFlow, Keras, and PyTorch examples.

If you have any questions, comments, or suggestions for additions to this deeplearning module please contact us at *helpdesk@chpc.utah.edu*.

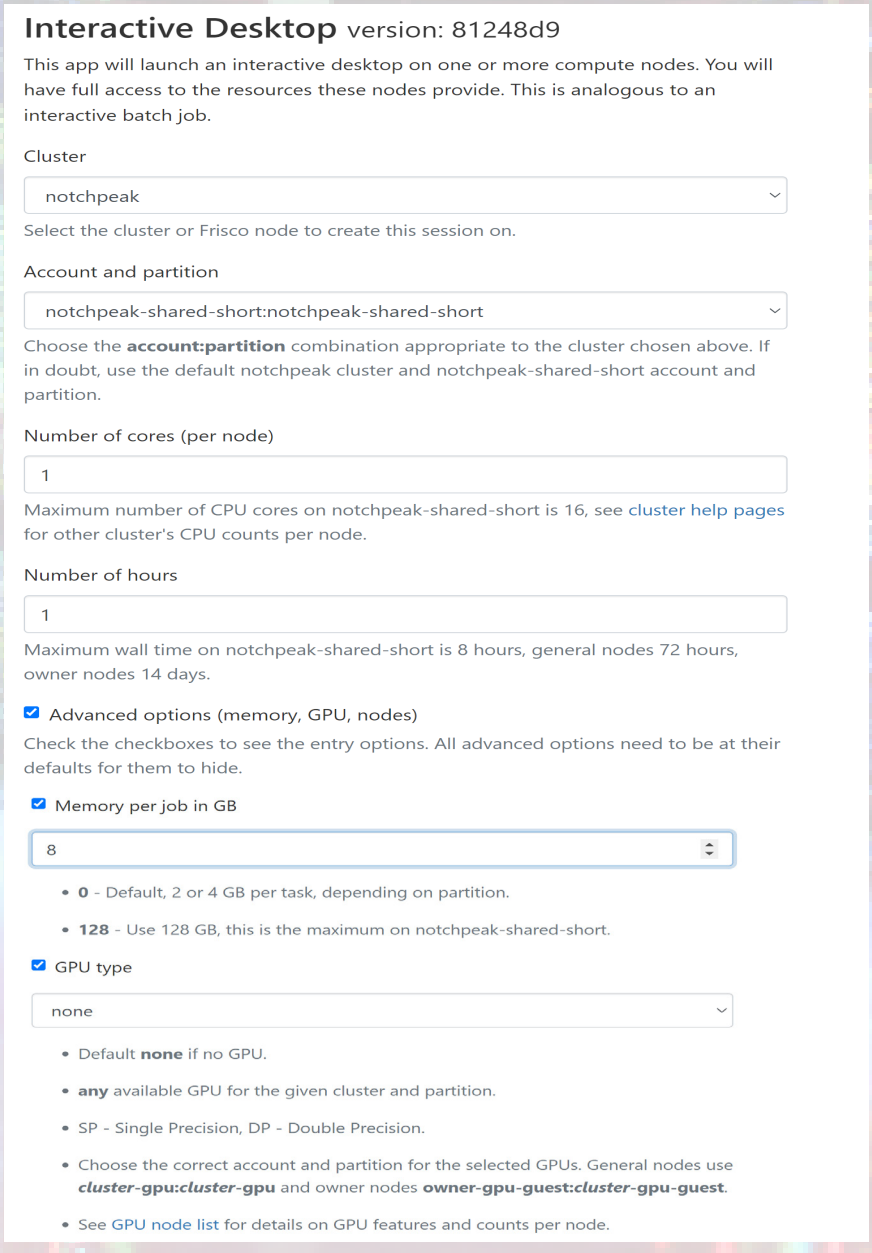
Changes to the Open Ondemand Web Portal

Martin Cuma, CHPC Scientific Consultant

One of the most useful features of the Open OnDemand web portal, which allows access to CHPC's Linux environment via a user friendly web interface, is the ability to run applications using the *Interactive apps* capability, bypassing the needs to learn Linux and SLURM scheduler commands. In an effort to simplify launch of the Interactive apps, we have recently made changes to the apps launch web form.

The updated web form, shown in the screenshot below, removed all text box entries, and replaced them with pre-populated fields. In particular, the SLURM account, partition and GPU type, which required external knowledge of the appropriate values. We have also implemented limits on the numerical fields, the number of CPUs, memory and walltime, which correspond to the selected partition. To simplify the form, we also by default hide the advanced options – number of cluster nodes, memory, GPU and nodes list, since these are optional for running a job, by using an *Advanced options* check box. When checked, these advanced options show up in the form. Similarly, if one selects on of the *frisco* interactive nodes, all other options hide, because Interactive apps on the Frisco nodes don't use SLURM for scheduling.

We anticipate to implement other adjustments to the Interactive apps form as we keep exploring the functionality, that allows to implement these features, and will announce, if these adjustments are noticeable enough.



Research Computing and Data Needs Survey

Brett Milash and Anita Orendt, CHPC Scientific Consultant

Earlier this year CHPC sent out a survey to help us assess the current and future research computing and data (RCD) needs at both the University of Utah and Utah State University. This information will be invaluable as we budget for next year and plan for long-term growth in the use of RCD resources, whether local or remote, in all aspects of research. This survey, unlike previous CHPC PI surveys was more general about RCD needs instead of focusing solely on CHPC resources. To this end, the survey was sent to not only the current CHPC PIs, but to more extensive lists of the research community at the institutions. In this newsletter article we summarize some of the findings of this survey.

Demographics:

At the University of Utah the survey was emailed to 6,653 researchers, of whom 2,808 opened the email and 437 clicked the link to take the survey. It is not known how many Utah State University researchers were invited to take the survey, as this process was left to individual department heads. We received 350 completed responses. Of these, 320 were from the University of Utah and 30 were from Utah State University. Of the respondents, the primary role of the majority (195) were PIs, but there were responses from research faculty (40), other staff (58), post docs (14), and "other" (40), which were mostly graduate students.

RCD needs:

About 75% of the respondents cited the need for RCD resources beyond what was available in their research group. When this was broken down to different options used for meeting the research computing needs, data storage needs, and data archiving needs, where respondents could choose all that applied, it was obvious that individual groups make use of multiple options, with the majority of use being either CHPC or their own lab/department/college resources. When asked, "Which resources does your group use to meet your RCD needs?" regarding computing, data storage, and data archiving, the survey respondents answered as shown in Table 1 below.

Location of Resource	Computing		Data Storage		Data Archiving	
Center for High Performace Computing (CHPC)	70.54%	158	48.21%	108	27.68%	62
Own lab, department, or college resources	72.32%	162	70.54%	158	50.00%	112
Resources available through collaborators	23.21%	52	16.07%	36	13.39%	30
Campus-level core facilities other than CHPC	13.84%	31	12.95%	29	9.82%	22
Private cloud offerings (e.g. CloudLab, Chameleon, Jetstream)	3.13%	7	7.14%	16	3.13%	7
National Resources (e.g. ACCESS allocated resources [Stampede, Bridges, EXPANSE], OSG, national labs, INCITE, etc)	11.61%	26	27.23%	61	17.86%	40
Web services [e.g. Rosetta, Galaxy, BLAST, WebCSD, SRA, FigShare, Dryad)	16.96%	38	9.38%	21	5.25%	14
Other (please specify)	0.89%	2	3.13%	7	2.68%	6
Column total	100.00%	224	100.00%	224	100.00%	224

Table 1: Percent and number of respondents using compute, data storage, and data archive resources at various locations

Anticipated future growth:

Our survey asked respondents to predict how their computing, data storage, and data backup/disaster recovery needs would change over the next 1, 3, and 5 years, and offered respondents the semi-quantitative choices "Decrease by ½ or more", "Stay the same", "Increase by ~5x", and "Increase by 10x or more". The results show our users expect their compute needs will nearly double in 3 years, and will increase almost five-fold within five years. Storage needs are expected to double in the next year and increase five-fold within five years, while data backup / disaster recovery needs are expected to grow a little slower, doubling in three years and growing some what less than five-fold in five years.

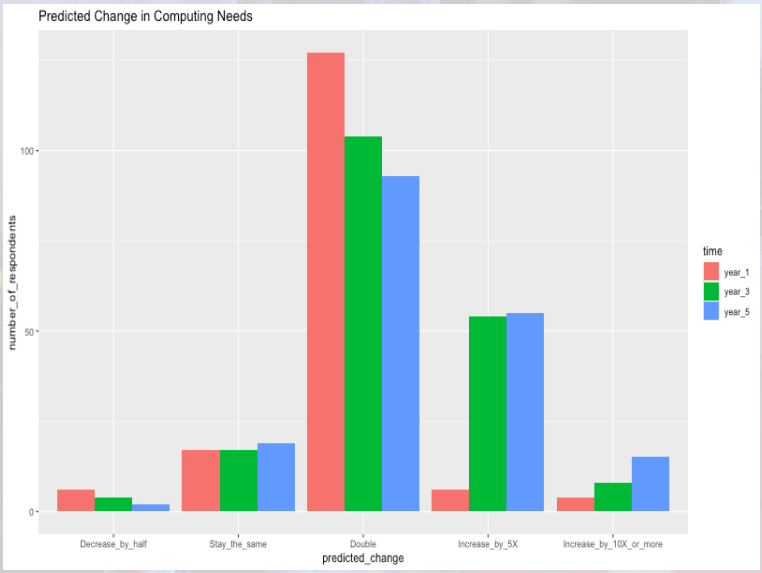


Figure 1: Predicted change in computing needs

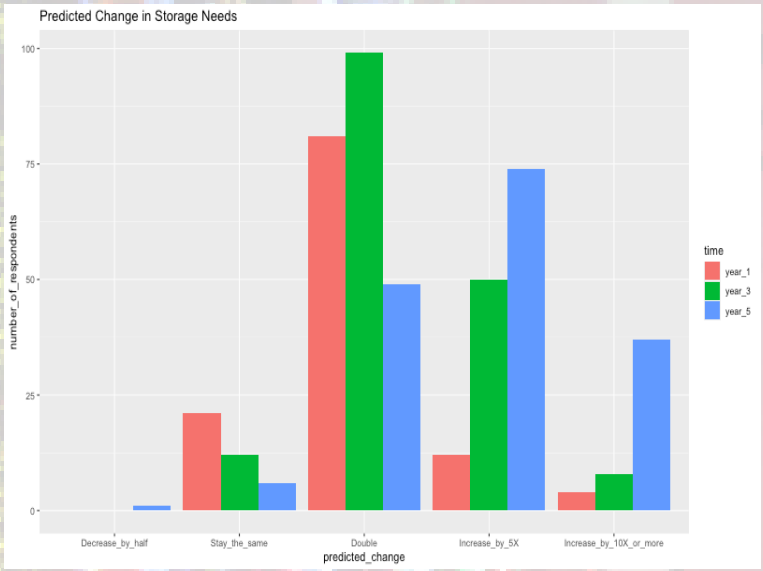


Figure 2: Predicted change in data storage needs

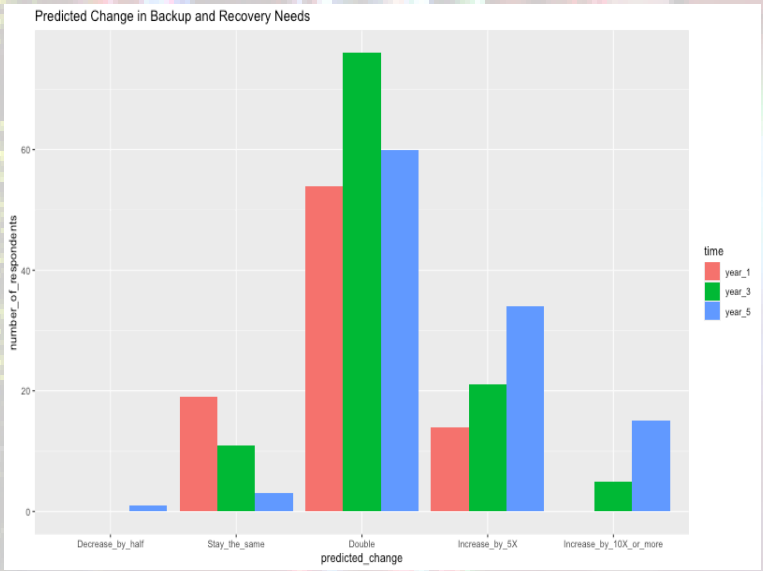


Figure 3: Predicted change in backup/disaster recovery needs

About current CHPC usage:

CHPC provides a variety of services including access to HPC resources, data storage, high speed data transfer, user support (consulting, helpdesk, software installation), virtual machines, windows servers, presentations/training, and class support. Questions were asked about the criticality of the service (critical to my group's needs, important to my group's needs, or not needed) as well as the quality of the CHPC services (meets my group's needs, does not meet my group's needs, not needed). This was asked individually for the general environment and the protected environment.

In the general environment these services, the ranking of importance is HPC, followed by Data Storage, User Services, Data transfer services and Virtual machines. Of these 94% report that the HPC clusters are either critical or important to their group's research needs, 88% saying the same about Data storage, 84% for user services, and 75% for data transfer services.

In terms of the quality of each of these services in the general environment, for those that make use of the services, over 90% indicate the service provided meets their needs, with HPC and user support having the highest percentages for needs responses.

In the protected environment, while based on fewer responses the same order of criticality of the services was reported, with 88% report that the HPC clusters are either critical or important to their group's research needs, 89% saying the same about Data storage, 86% for user services, and 76% for data transfer services. The responses related to the quality of services provided in the protected environment were similar to those of the general environment.

Suggestions for improvement of CHPC:

A recurring theme among the suggestions to improve the resources and services at CHPC are requests for services that are in fact available at CHPC, for example access to statistics programs STATA and SAS, availability of group space exceeding 1 TB, Windows servers, dedicated HIPAA-compliant servers, access to Jupyter Notebook, and access to interactive computing. When possible, we will be reaching out to the individuals who made the comments.

One interesting note was that of the 60 respondents that were not making use of the CHPC resources for their RCD needs, 41 were not aware of CHPC at all. This is over 10% of those who took the survey. This indicates that CHPC should explore additional ways

to improve the awareness of the resources and services provided by CHPC to the faculty and research community. Some respondents who do not use CHPC but are aware of it cite difficulty of use; we hope to address ease of use and interactive computing capabilities through the Open OnDemand web portal, which provides an interactive web interface to many CHPC resources.

Data archiving is a commonly expressed unmet need, and although secure and affordable data archiving is a campus-wide need CHPC may play a role in meeting this.

Comments received from the survey indicate increasing needs for a secure and restricted computing environment. Of the survey respondents there are 31 currently using the CHPC Protected Environment, while 16 respondents plan to use it, and another 25 are uncertain about its use.

Other comments expressed a desire for increased training in emerging technologies and foundational data science. We intend to explore adding topics like these to our presentation series as well as expanding our collection of brief training videos.

Future:

We plan on continuing to analyze the results from this survey, for example we intend to identify University departments not represented among the survey respondents to find potential new users. In addition, as mentioned we are looking to reach out to those who identified themselves (this was optional) and made suggestions or comments that we can address including reasons for not using CHPC, unmet needs in terms of resources and services provide for compute, data storage as well as data sharing and data management.

Recent Updates to Lonepeak

Anita Orendt, CHPC Scientific Consultant

CHPC has just finished an ongoing project to replace some of the general nodes of lonepeak. We started this process earlier this year, with the initial replacement of 48 nodes, as was announced on February 8th, and the 2nd, final replacement during the week of March 13th.

The lonepeak general nodes are being replaced with systems that have newer generation cpus, providing more compute cores and more memory per node, thanks to a donation of hardware being removed from service by the IT support for the University Hospital. We replaced nodes with 12 physical cores and 96 GB memory with nodes that have 16-24 physical cores and 256 GB memory. We should note that the nodes that are being replace were also nodes that were donated to CHPC after being retired from service by the hospital.

As mentioned, the first batch of 48 of these replacement nodes, lp017-lp064, were made available for use in early February. During the week of March 13th, we finished the replacement of lp[065-112]. In addition, we also retired lp001-016. Note that the four current general lonepeak nodes lp[229-232], each with 32 cores and 1 TB memory will remain in service, as will 94 of the original 12 core 96 GB memory nodes.

The current nodes on lonepeak are listed in the table below.

If you have not already, please take the time to adjust your scripts, in terms of cores and memory being requested, in order to continue to make efficient usage of the lonepeak general nodes. If you have any questions or need any assistance in doing this, please reach out to CHPC via *helpdesk@chpc.utah.edu*.

NODES	CORES	MEMORY (GB)	INFORMATION
lp[017-074]	24	256 GB	Intel Ivy Bridge, Xeon E5-2697 v2 CPUs @ 2.70 Ghz base clock
lp[076-082]	20	256 GB	Intel Ivy Bridge, Xeon E5-2660 v2 CPUs @ 2.20 Ghz base clock
lp[075,083-112]	16	256 GB	Intel Sandy Bridge, Xeon E5-2680 CPUs @ 2.70 Ghz base clock
lp[229-232]	32	1024 GB	Intel Nehalem, Xeon X7560 CPUs @ 2.27 Ghz base clock
lp[133-148,150-202,204-228]	12	96 GB	Intel Westmere, Xeon X5650 CPUs @ 2.67Ghz base clock

CHPC Summer 2023 Presentation Schedule

Anita Orendt, CHPC Scientific Consultant

Below is the Summer 2023 presentation schedule. Presentation are typically one hour, from 1-2pm, with hands on presentations, marked with * in the schedule below, being two hours from 1-3pm. For additional details, including zoom information and starting times, please see *https://chpc.utah.edu/presentations/summer2023chpcpresentationschedule.php.*

DATE	PRESENTATION	PRESENTER(S)
Tuesday, May 25, 2023	Overview of CHPC	Anita Orendt
Thursday, Jun 1, 2023	Module Basics	Zhiyu (Drew) Li and Anita Orendt
Tuesday, Jun 6, 2023	Slurm and Slurm Batch Scripts	Zhiyu (Drew) Li and Anita Orendt
Thursday, Jun 8, 2023	Hands on Introduction to Linux, part 1*	Zhiyu (Drew) Li and Martin Cuma
Tuesday, Jun 13, 2023	Hands on Introduction to Linux, part 2*	Zhiyu (Drew) Li and Martin Cuma
Thursday, Jun 15, 2023	Hands on Introduction to Linux, part 3*	Zhiyu (Drew) Li and Martin Cuma
Tuesday, Jun 20, 2023	Hands-on Introduction to Open OnDemand*	Martin Cuma
Thursday, Jun 22, 2023	Introduction to Parallel Computing*	Martin Cuma
Tuesday, Jun 27, 2023	Introduction to Containers*	Martin Cuma
Thursday, Jun 29, 2023	Mathworks: Speeding Up Your MATLAB Code*	Mathworks/Martin Cuma
Thursday, Jul 6, 2023	Hands-on Introduction to Python, Part 1*	Brett Milash and Wim Cardoen
Tuesday, Jul 11, 2023	Hands-on Introduction to Python, Part 2*	Brett Milash and Wim Cardoen
Thursday, Jul 13, 2023	Hands-on Introduction to Python, Part 3*	Brett Milash and Wim Cardoen
Tuesday, Jul 18, 2023	Using python at CHPC* - NEW	Brett Milash
Thursday, Jul 20, 2023	Hands-on Introduction to R, Part 1*	Wim Cardoen
Thurday, Jul 27, 2023	Hands-on Introduction to R, Part 2*	Wim Cardoen
Tuesday, Aug 1, 2023	Hands-on Introduction to R, Part 3*	Wim Cardoen
Thursday, Aug 3, 2023	Using R at CHPC* - NEW	Brett Milash

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 acknowledgment 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 by sending to *helpdesk@chpc.utah.edu.*

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

Center for HIGH PERFORMANCE COMPUTING
THE UNIVERSITY OF UTAH