Center for
HIGH PERFORMANCE COMPUTING
THE UNIVERSITY OF UTAH

Newsletter

Fall 2016

Research Highlight

Clean Coal: Powered by Exascale

CHPC's Contribution to Clean Coal Research

"While most of the production runs are completed at the national laboratory supercomputers, local high-performance computing cluster Ash with more than 7,400 cores plays a critical role in successful code development, testing and deployment to the larger systems."

"With the help of expert team from the Center for High Performance Computing (CHPC) at the University of Utah, the CCMSC researchers are able to test various software and hardware configurations to meet computing, visualization, and storage needs essential to the success of the program."

by Philip J. Smith and Michal Hradisky, CCMSC

The mission of the Carbon-Capture Multidisciplinary Simulation Center (CCMSC) at the University of Utah is to demonstrate the use of exascale uncertainty quantification (UQ) predictive simulation science to accelerate deployment of low-cost, low-emission electric power generation to meet the growing energy needs in the United States and throughout the world. The two main objectives, advancing simulation science to exascale with UQ-predictivity in real engineering systems and use of high-performance computing (HPC) and predictive science to achieve a societal impact, are linked together through an overarching problem: simulation of an existing 1,000 MW coal-fired ultra-supercritical (USC) boiler and simulation of a design 500 MW oxy-coal advanced ultra-supercritical (AUSC) boiler.

This overarching problem integrates multidisciplinary groups of researchers and engineers from the University of Utah, the University of California Berkeley, and Brigham Young University. The CCMSC, under the leadership of Prof. Philip J. Smith, the Principal Investigator and the director of the center, is organized into three teams: the Exascale Team, the Predictive Science/Physics Team, and the V&V/UQ Team. The center is a $21M five-year program funded by the Predictive Science Academic Alliance Program II (PSAAP II) which is funded by the Department of Energy (DOE) as a part of the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) program.

The CCMSC team started with an existing proven computational platform, UintahX (which has been proven to scale up to 768K cores on Mira at LLNL), and is sequentially moving from petascale to exascale, while using a hierarchical validation to obtain simultaneous consistency between a set of selected experiments at different scales embodying the key physics components (large eddy simulations, multiphase flow, particle combustion and radiation) of the overarching problem. The purpose of this hierarchical approach is not to eliminate uncertainty, but to rather identify it at each sub-scale, sub-physics level and to determine how it propagates through the validation hierarchy while identifying where in the validation hierarchy reducing uncertainty will have the largest impact on the final prediction of a full-scale utility power boiler. The validation and prediction process is constrained so that the final prediction is consistent with all of the experiments and all the validation metrics of the validation hierarchy simultaneously. The validation hierarchy is shown in Figure 1.

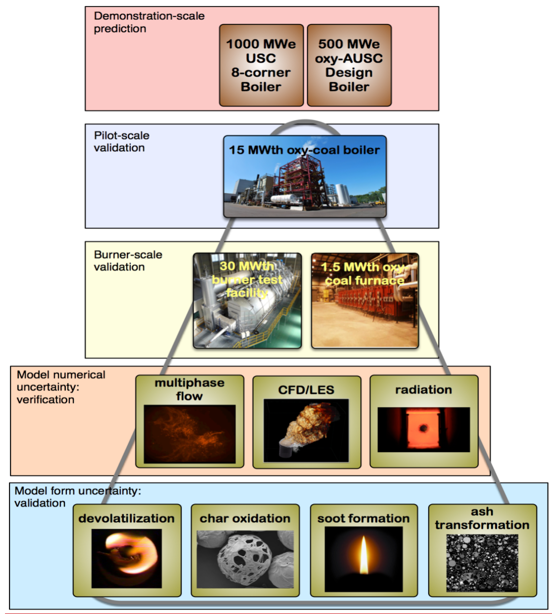


Figure 1: Carbon-Capture Multidisciplinary Simulation Center validation hierarchy.

Graphical representation of the methodology to reduce parameter uncertainties while increasing credibility of simulation predictions. Triangular gray region indicates studies for which experimental data is available.

To demonstrate societal impact of extreme computing in real engineering systems, CCMSC established a collaborative agreement with GE Power (formerly Alstom Power) to jointly use the developed simulation technology to help assess mixing and combustion efficiency, evaluate and minimize gas-side energy imbalance, and predict detailed heat flux profiles at the furnace walls of an existing 1,000 MW USC boiler. This utility boiler, built by GE Power, is more than 65 meters tall and with an efficiency of 40%, generates enough electricity to power around half a million homes. The high-fidelity simulations performed by the center provide a critical step forward for improving combustion efficiency of the existing and any future boiler designs. In 2015, the center was awarded a 351 million core hours INCITE award, the second largest award every given, to run simulations of GE Power's 1,000 MW USC boiler on Titan and Mira supercomputers. The instantaneous temperature distribution inside the 1,000 MW boiler is shown in Figure 2.

PHILIP J. SMITH
The University of Utah
Director of The Institute for Clean and Secure Energy (ICSE)
Director of The Carbon-Capture Multidisciplinary Simulation Center (CCMSC)
Chair of The American Flame Research Committee (AFRC)
Professor, Department of Chemical Engineering

CARBON CAPTURE MULTIDISCIPLINARY SIMULATION CENTER

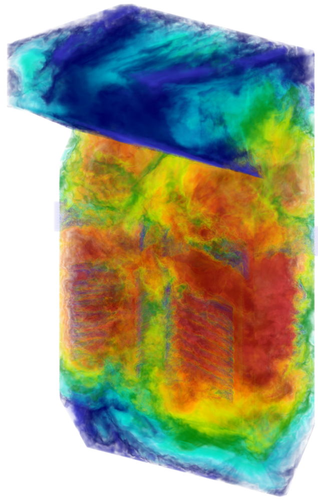


Figure 2: Volume rendering of an instantaneous gas temperature using LLNL's VisIt Software.

Cooler temperatures (blue) are achieved near the top of the boiler as the gasses and particles enter the convection pass and loose heat to the panels, platens, and tube-banks. Highest temperatures (red) are seen in the fuel-rich zones, where particles are mixing with the oxidizer. This simulation resolves the boiler using physical cell size of 4.25 cm, for a total number of 455 million cells. This computation, capturing almost 20 seconds of real time with an average time step of 8e-5 seconds, was run on Mira and consumed 97 million core hours.

Simulation of the design of a 500 MW oxy-coal AUSC boiler is motivated by the emergence of oxy-combustion systems as the most likely low-cost technology solution for both carbon capture utilization and storage as well as simultaneous reduction of NO_x and SO_x emissions. AUSC power generation units increase efficiency over USC units to offset the energy penalty of the separation of oxygen from air in an oxy-fired unit.

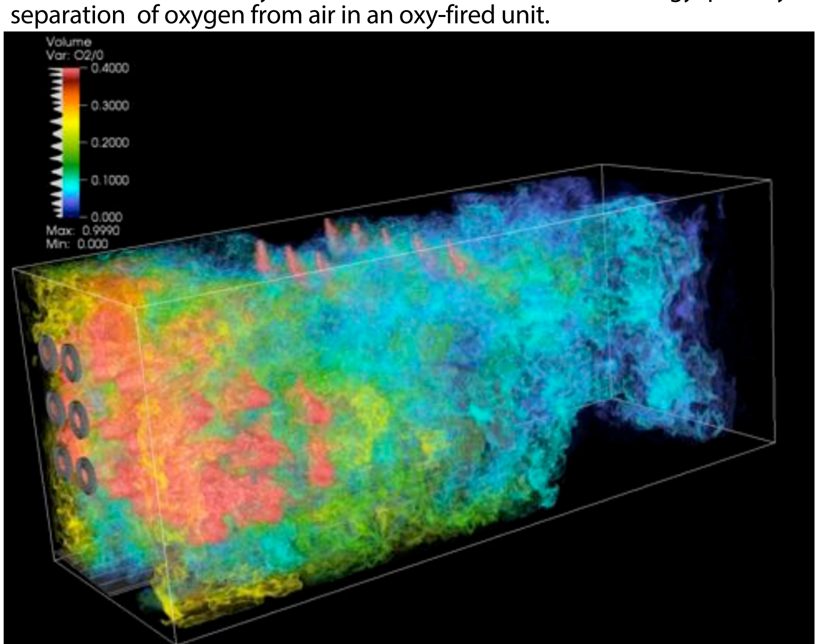


Figure 3: Volume rendering of oxygen concentration inside the design boiler using LLNL's VisIt software.

The highest concentration of oxygen (red) is at the oxidizer inlets. The oxygen continues to be depleted as the gasses pass from the radiation section to the convection section of the boiler, where the concentration of oxygen is the lowest (blue). This computation uses a physical cell size of 2.8 cm to resolve the design boiler for a total number of 225 million computational cells. This computation was run on Vulcan (at LLNL) and used 8 million core hours to simulate 2 seconds of real time.

The primary intended use of the simulation is to predict the heat flux distribution inside the boiler with uncertainty bounds, aiding in material science research of steam tubes that are able to withstand the high temperatures and pressures of oxy-fired AUSC systems. The instantaneous oxygen concentration inside the design boiler is shown in Figure 3.

To perform high fidelity simulations of the overarching problem, on top of the computational resources provided as a part of the INCITE award, the center continues to use computational resources under the PSAAP II allocation at national laboratory supercomputers such as Vulcan, Syrah and Surface at Lawrence Livermore National Laboratory (LLNL), Mira and Cetus at Argonne National Laboratory, Titan at Oak Ridge National Laboratory, as well as the local cluster Ash at the University of Utah. While most of the production runs are completed at the national laboratory supercomputers, the local high-performance computing cluster Ash with more than 7,400 cores plays a critical role in successful code development, testing and deployment to the larger systems. With the help of the expert team from the Center for High Performance Computing (CHPC) at the University of Utah, the CCMSC researchers are able to test various software and hardware configurations to meet computing, visualization, and storage needs essential to the success of the program. During the first two years of the program, not including the INCITE award, the center spent approximately 500 million core hours among all computational resources.

While coal is the second largest source of primary energy in the world (trailing only to oil), it is the most carbon-intensive one, accounting for almost half of world's CO_2 emissions. Worldwide, almost 40% of all electricity is produced in pulverized coal plants[1]. China alone consumes as much energy from coal as all other countries combined[2]. The current average global efficiency rate of a coal-fired power plant is 33%, mostly due to use of conventional plant designs. Increasing the average global efficiency by 7% would reduce global CO_2 emissions by as much as 7%[3]. Therefore, nationally and worldwide, even small improvements in the utility boiler efficiency can lead to large reductions in carbon footprint. Due to a complex combination of socioeconomic and geopolitical factors, coal, along with other fossil fuels, is projected to remain one of the essential energy sources through at least 2040, even under the most strict policies currently being considered[1].

1 Key World Energy Statistics 2016, International Energy Agency: *https://www.iea.org/publications/freepublications/publication/key-world-energy-statistics.html* (Accessed: Oct. 24, 2016)

2 Top 20 CO2 Emitting Countries, The Shift Project Data Portal: *http://www.tsp-data-portal.org/TOP-20-CO2-emitters-from-energy-consumption#tspQvChart* (Accessed: Oct. 24, 2016)

3 High Efficiency Low Emission Coal, World Coal Association: *http://www.worldcoal.org/reducing-co2-emissions/high-efficiency-low-emission-coal (Accessed: Oct. 24, 2016)*

Technical Highlight | Storage Solutions at CHPC

by Samuel Liston

CHPC has been providing storage for many years and continues to learn and gain knowledge, with experience that started from very dedicated and specialized systems and advanced into expertise in multi-tenant, shared, and commodity solutions. As new solutions are explored, they are thoroughly vetted for stability and resiliency. We have also learned the importance of facilitation, communication, and interaction with CHPC's diverse user community to help educate the users regarding the pros/cons and limitations of the solution, such as performance, handling of large numbers of small files, back-ups, downtimes if specific components fail, as to provide users with a clear understanding of our storage solutions.

Over the years, CHPC has deployed several generations of storage solutions, ranging from stand-alone file servers to SAN (storage area network) and NAS (network attached storage) systems, to parallel file systems. Through each of these iterations the center has worked to find solutions that address the limitations of previous solutions. Improvements from faster hardware to better software and new technologies shape those solutions, with the goal being to provide storage which is as accessible and usable as possible, with a reasonable level of robustness, reliability, and manageability, and with performance and scalability at a price point that is approachable. CHPC tries to engineer and deploy solutions that can run on a variety of hardware, in order to minimize vendor lock-in and allow CHPC to seek aggressive pricing. A solution that can be built and expanded as needed in a "condo" model is also a key feature in order to allow researchers to purchase space in large or small chunks as funds allow and their needs require. Many new researchers come to campus with the assumption that they will need to create their own storage solution for their new group, and they will do it themselves if a compelling alternative is not available. Thus CHPC strives to create solutions that are better than any reasonable solution a researcher can set up on their own.

Current home and group space solution:

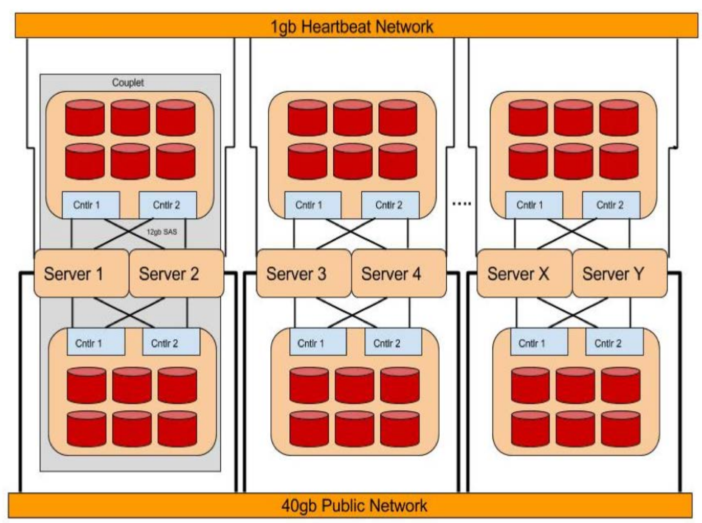
Our current home directory and group space offering is based on a couplet architecture, shown in the schematic to the right. A couplet consists of two servers each directly attached to a controller tray. These trays contain redundant controller modules. The two servers each contain two dual port SAS cards, with each controller module connected to one of the two SAS cards in each of the servers. This SAS topology allows for loss of a server, SAS card, cable, or controller module, without losing data access. By directly attaching servers to storage, the cost and complexity of the storage fabric is eliminated.

The current generation is built with 60 drive trays, with 8 TB drives. Disk arrays within these 60 drive trays are defined in four 15 drive sets, configured as a 15 drive Dynamic Disk Pools (DDP) with a single drive worth of reserve capacity. In this arrangement, each storage controller module has ownership of two arrays allowing for a more equally distributed workload on each controller module. At the operating system (OS) level these arrays are stripped together in pairs using Logical Volume Management (LVM). Configuring the arrays in this fashion fully utilizes the horsepower of both controllers. Having dual controller storage trays enables continued functionality during a failure of one of the controllers, and it also allows for online firmware updates, limiting the need for downtimes.

The couplet model allows for expansion without additional infrastructure through the use of JBOD (Just a Bunch Of Disks) trays. Each controller tray can have two additional JBOD trays connected behind it. At current drive capacities in our current RAID/DDP configuration, a fully expanded couplet can provide up to 1,920 TB of storage. In this configuration a single pair of controllers would collectively manage 360 disks. While this is very efficient for cost, expanding to this density could impact performance. It is possible in this configuration for the aggregate I/O from the disks to exceed the performance capabilities of the pair of controllers, causing them to be a bottleneck in the I/O path.

Each storage server is connected to the network via two 40 Gbps ethernet connections. These connections are setup in an active/passive bond, to increase resiliency in the case of a failure. The storage servers are also dual connected in an active/passive manner to a 1 Gbps network. This internal 1 Gbps network serves as a communication channel for heart beating and quorum disk checks among the servers in the storage cluster. An additional 1 Gbps connection is dedicated to the baseboard management controller, which provides lights-out management and remote access to the server.

The standard level of warranty on all of our storage cluster components is five years. For the majority of the systems this covers support calls and next-business-day parts. As budgets allow and the criticality of the data warrants, higher levels of warranty can and are purchased, particularly as larger research groups purchase dedicated storage hardware.



The file system of choice for the current storage cluster generation is XFS. This is another flexible aspect of the CHPC storage solution as there are many file systems with a variety of features. XFS was chosen as it offers a good array of features along with a stable and reliable history. ZFS and Btrfs are being explored for future use as both offer many new features and capabilities, but are still experimental. XFS allows for creation of very large file systems. These XFS file systems are configured with 64-bit inodes to facilitate file counts into the billions. Quotas are configured on each file system. Three type of quotas are allowed: user, group and project quotas. Project quotas are used on the CHPC storage clusters to subdivide a file system into smaller pieces based on the amount of storage a researcher purchases. By provisioning storage space in this way researchers can purchase as little as 1 TB. Whole or half couplets can also be purchased by a particular group and added to the storage cluster, if their needs require.

All storage servers, their associated disks, and their networks are monitored for health and tracked for performance using Nagios and Cacti. Nagios uses defined heuristics to ensure hardware and software availability and functionality. Cacti is used to track individual system metrics, such as CPU, memory or network utilization. Both of these applications are antiquated and therefore CHPC is beginning to explore alternatives. In the event of an issue or hardware failure, alert emails are sent from Nagios, the servers, and/or the disks trays, are sent to the appropriate system administrators, to ensure necessary actions are taken as soon as possible. Additional Linux tools created for diagnostic, trending and troubleshooting purposes are in place and allow system administrators to determine and diagnose issues, bottlenecks and user behaviors. As new and better tools are discovered that enable system administrators to better understand the functions of the system, they are deployed into the CHPC suite of tools.

Provisioning and management of systems is handled through a Spacewalk server. Spacewalk provides a single pane of management for Linux systems, handling OS installation, configuration and customization for a single machine or across an entire cluster. It manages software updates and pushes out security patches and bug fixes. As an open-source tool, Spacewalk, enables CHPC to efficiently deploy and manage large sets of systems and maintain OS homogeneity and consistent security without software costs.

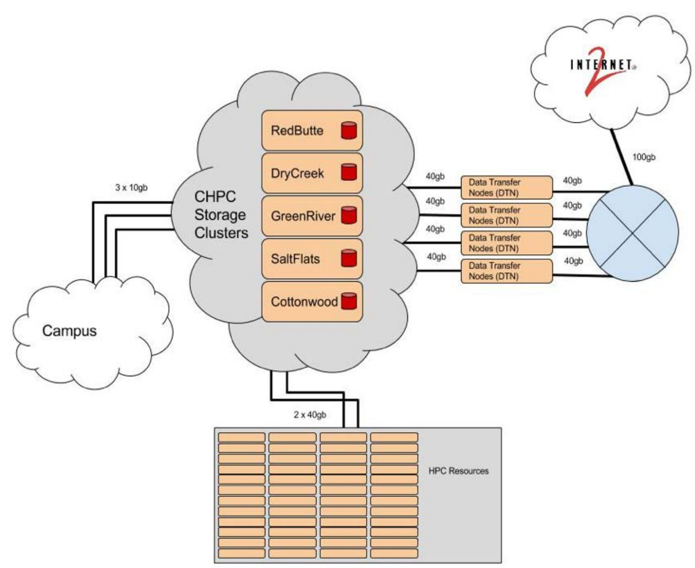
Shown to the right is a high-level diagram of the bandwidth characteristics and network topology associated with CHPC's storage clusters. Accessibility is a crucial component in CHPC's storage deployment. All storage is accessible on the HPC resources. It is also accessible on the Data Transfer Nodes (DTN) which are connected to the science DMZ architecture that provides a fast path, bypassing the campus firewall, for file transfer to remote locations such as peer institutions and national resources. Shares are also available on campus or through the campus VPN via SAMBA/CIFS.

Limitations of the current solution:

Group purchased home directory spaces are currently backed up to tape with weekly full backups and nightly incremental backups. Group space, which comprises the majority of the storage managed by CHPC, can be backed up to tape provided that the researcher purchases the tapes necessary for these archivecopies. Due to the amount of the group space storage and the limited nature of the backup system available, an archive run can be performed, at most, once a quarter. The fact that a good portion of the total data housed at CHPC, while being resilient against failure, is not truly backed up is a known deficiency in this storage solution. However, CHPC has neither the budget nor man-power required to build and maintain a traditional backup solution equal in scale to primary storage.

An additional limitation is one that is intrinsic of all RAID based solutions, specifically the difficulties surrounding recovery from failure. Catastrophic failure of a RAID system is always a possibility. Measures are in place to defend against small failures, such as disks, power supplies or even controllers, but the possibility of large-scale failure, such as errors in the logic of the RAID structure itself, is real, thus the saying: "RAID is not a backup." Recovery from a large-scale failure in the CHPC RAID-based storage solution can be complex. If the failure is in the hardware or firmware, the problem must be diagnosed and either deemed corrected or new hardware to correct the problem must be shipped before data can be restored. In addition, as most backup solutions are tuned to efficiently backup of data, as opposed to efficient restoration, the restoration process takes significantly longer than the original backup process.

In an effort to mitigate these deficiencies CHPC is currently exploring several options. We are exploring either symmetric or asymmetric block-based replication adding a shadow copy to our next generation home directory solution to allow for additional resiliency in case of hardware failures, and to avoid having any lengthy recovery time. Block-based replication will allow us to have a replicated copy online very rapidly in case of failure on the primary side. In addition, we have just finished implementing an object-storage based archive solution to complement the current group space offering. This solution, based on CEPH, is a self-service private cloud that researchers can purchase space on to serve as a near-line backup copy of their data. This archive solution removes much of the management overhead related to our traditional tape-based backup processes, and allows the user to manage archiving their data.



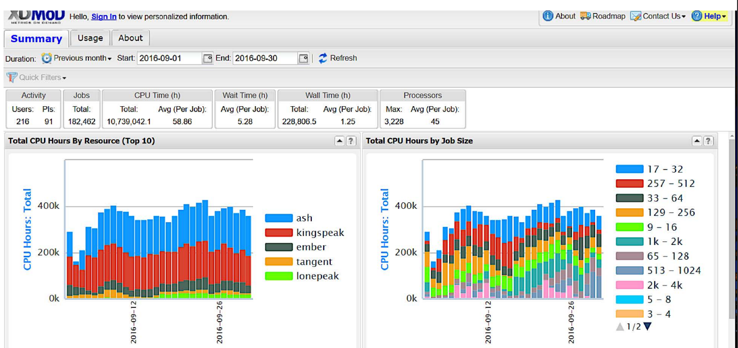
User Services | XDMoD: New Tool to Obtain Cluster Usage Data

by Anita Orendt

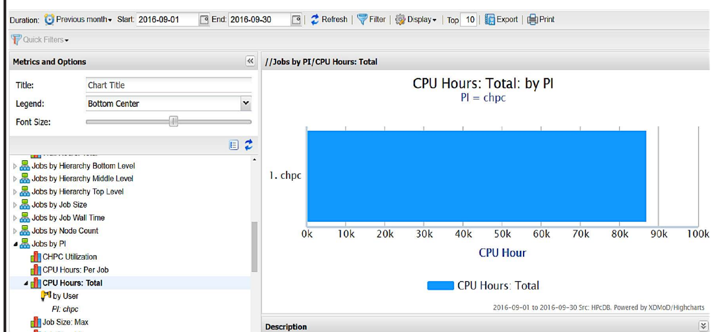
CHPC has a new tool that can provide more detailed information about usage: XDMoD – XD Metrics on Demand. XDMoD is an NSF-funded open source tool designed to audit and facilitate the utilization of the XSEDE cyberinfrastructure by providing a wide range of metrics on XSEDE resources, including resource utilization, resource performance, and impact on scholarship and research. While initially focused on the XSEDE program, Open XDMoD, *xdmod.sourceforge.net*, has been created to be adaptable to any HPC environment.

You can can access CHPC's installation at *xdmod.chpc.utah.edu*. In addition there is a CHPC web page describing XDMoD and its usage.

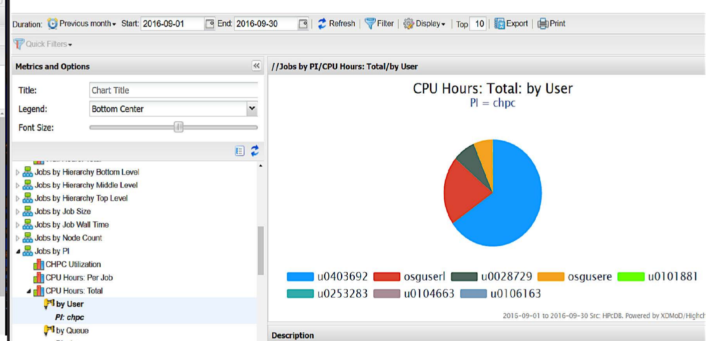
On the front page of the XMoD website you can see summary information for the CHPC clusters. By default, you will see the previous month's usage, but you can change the timeframe being reported by either the pulldown menu or by changing the start and end date and then refresh. This view provides statistics on the number of users running batch jobs and the number of groups to which they belong, the number of jobs, the total number of CPU time (core hours), and the average wait time.



Using XDMoD, you will be able to delve deeper into the usage. To do so, select the Usage tab. In this example, we will look into how you can break down the usage to a per user level. Again you can select the time frame of interest. On the left hand side there are a number of metric choices; in this panel select "Jobs by PI" and expand the menu. Expanding the Jobs by PI results in a new set of choices, including "CPU Hours:Total". On the same level as the time frame choices there is a filter – open this and select your group. This will give you single entry for the group



Then on the blue bar – left click and select drill down by user. By the filter these are display choices – you can do a pie chart, a bar graph etc. as well as plotting this data as a time series of different steps. Note that you can also export the data as a figure or a CSV file which excel can read.



This is just one example of the detailed usage information you can obtain with XDMoD. For more information, please see the documentation available at the *xdmod.sourceforge.net* site. Note that CHPC is still working on implementation some of the features of the program.

Research groups can see their current quarter collective usage on the CHPC compute clusters via web pages:

- **For Ember and Kingspeak: *https://www.chpc.utah.edu/usage/cluster/current-project-general.php***

**Note there are sections for general node usage, as well as Kingspeak and Ember owner resource usage.

- **For Lonepeak: *https://www.chpc.utah.edu/usage/cluster/current-project-lonepeak.php***
- **For Ash: *https://www.chpc.utah.edu/usage/cluster/current-project-ash.php***

- These pages are updated throughout the day, and the usage is reflective of jobs that have finished (regardless of if they were successful, failed, ran out of time, or if they were preempted). These pages are useful to monitor the usage of general allocations.
- To obtain information about usage in past quarters, you can check the links at *https://www.chpc.utah.edu/usage/cluster/index.php*

Restricted Data Handling and Compliance

by Wayne Bradford

To most people, the term compliance is usually associated with industry or government regulation or rules. However, it also relates to an organization's internal acceptable use policies (AUP) for network/computer/data access. We often take it for granted, blindly click to accept the terms, and assume the University has you covered and that everything is okay. However, if you work with restricted or sensitive data, or have plans to do so, it would be wise to understand the policies, rules and severity of the law. It is important to note that many users may not be aware that they are already working with such data and that they could be held personally responsible in the event of a breach. If policy is not followed the University may not defend you. You need to understand your contractual and legal obligations with respect to information handling, and data breach notification requirements to ensure you are in compliance with state and federal laws, as well as Institutional policies. At the University of Utah, new stricter policies have recently been ratified, and it is likely that these policies affect you. In case you don't already know, CHPC has resources to accommodate research involving restricted data.

Both the University of Utah "Acceptable Use Policy" (AUP) and the data classification and encryption policies would be wise to understand, specifically "Rule 4-004C": *http://regulations.utah.edu/it/rules Rule4-004C.php*. We recommend focusing attention on the "Data Classification Model" for data types and associated requirements and understand how your data may fit. When working with restricted data such as electronic personal health information (e-PHI) covered by the Health Insurance Portability and Accountability Act (HIPAA), you must encrypt the data both when is it being stored (data at rest) or transmitted (data in transit). This requirement applies not only to your work computing resources (i.e., University owned equipment or resources) but also to any personally owned computing resources, such as mobile devices and home computers, which you use to interact with institutional computing systems and networks. This said, the U of U's Institutional Security Office strongly advises that any protected data be stored only on the appropriate university owned resources, and never is moved to personally owned devices.

Any device used to interact with protected data must be encrypted per U of U Policy. Encryption is required in a manner that allows for the burden of proof in accordance with applicable state or federal safe harbor guidance. If any device containing restricted data is lost or stolen, you are required by institutional policy to report the loss. It is not enough to simply say your notebook was encrypted but you must also be able to prove that it was encrypted. CHPC has a form that can be used to document the encryption of devices, found at: *https://www.chpc.utah.edu/documentation/Encryption%20Attestation.pdf.*

When working with human genome data, covered by the NIH Database of Genotypes and Phenotypes (dbGaP) policies, the data at rest is not required to be encrypted unless it is being stored on mobile devices or on removable media. NIH recommends the use of NIST validated encryption technologies. Encryption is required when storing such data on cloud storage, and data in transit is also required to be encrypted.

Some may wonder why laws and IT services are so restrictive about data use/ access and computer/network use. This is the new reality. We need to take proactive measures for better cyber security to decrease the cyber exposure (attack surface) and reduce the risk level of data loss and breaches from occurring. There are significant but hard to quantify costs associated with a breach, including losses from downtime and loss of credibility and reputation, not to mention significant monetary costs. According to *http://www.ponemon.org/* and *http://www.databreachcalculator.com/* the average cost per record is somewhere in the range of $154 to $188. For more information, the Verizon RISK organization publishes summaries of data breach investigations (to download a free copy of this report, refer to *http://www.verizonenterprise.com/DBIR*).

The University of Utah Information Security Office has approved the use of the University's Box installation for storage of e-PHI, HIPAA in accordance with relevant privacy laws and subject to appropriate usage as outlined in the University's Box Acceptable Use Policy which can be found at *http://it.utah.edu/_downloads/University%20of%20Utah%20Box%20User%20Agreement.pdf*. For more information about the U of U Box instance please refer to *https://uofu.service-now.com/cf/kb_view. dosysparm _article =KB0001022* and note the section at the bottom about storage of regulated information.

If your working with HIPAA or dbGaP restricted data, or have questions about other potentially restricted data classifications and how to handle them or the research needs that may require more resources than the U of U box offers, e.g., computational or virtual resources or larger storage, please refer to the information about CHPC's protected environment (PE) found at *https://www.chpc.utah.edu/resources/ProtectedEnvironment.php*

University Of Utah Policies, HIPPA, and dbGaP Regulations
Additional Rescources and Links

University of Utah Policy

Network acceptable use policy:
http://it.utah.edu/policies/UITPolicyNetworkConnection.pdf

University of Utah's Information Technology Policies:
http://regulations.utah.edu/it/

U of U Incident Reporting:
https://uofu.service-now.com/cf/kb_view.do?sysparm_article=KB0000910

Health Insurance Portability and Accountability Act (HIPAA)

The Health Insurance Portability and Accountability Act of 1996 required the Secretary of the U.S. Department of Health and Human Services (HHS) to develop regulations protecting the privacy and security of certain health information.

For more information:
http://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html

Breach notification rule:
https://www.hhs.gov/hipaa/for-professionals/breach-notification/

The Database of Genotypes and Phenotypes (dbGaP)

The dbGaP is a sponsored repository. It was launched by the National Institutes of Health (NIH) in response to the development of NIH's GWAS (genome-wide association studies) policy for genomic data sharing.

For more information:
https://gds.nih.gov/index.html
http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf

CHPC Hosting Services

CHPC operates community servers on which we can provide:

-- Git repositories

-- Subversion revision control systems

-- Microsoft SQL database

-- MySQL databases

-- Hosting for department, group, and individual web sites

*HTML, PHP, and WordPress sites are all supported. These are usually named "**http://**[insert word choice]**.chpc.utah.edu"** (other names can also be arranged. See* [insert word choice] *in the above URL example).*

If you have an interest in any of these services or if you would like additional information, please send a request to ***issues@chpc.utah.edu***

Note that all users have the option of having a personal website served out of your **public_html folder; in this case your web page can be accessed by using the URL example "**http://home.chpc.utah.edu/~**[insert YourUNID]"*

(non-unid aliases are available by request)

CHPC offering a NEW training session on R

Dr. Wim Cardoen, CHPC Staff Scientist, will be offering a new hands-on workshop "Introduction to R" on Tuesday November 29th, 1-3pm in the INSCC Auditorium. He has also created a new CHPC software documentation page on using R on CHPC resources, found at:

https://www.chpc.utah.edu/documentation/software/r-language.php

CHPC implements SLURM support for MATLAB Distributed Computing Toolbox

We have recently worked with MathWorks to enable support for the MATLAB Distributed Computing (MDCS) Toolbox on our HPC clusters. Users can launch parallel Matlab batch jobs directly from the Matlab session running on the cluster interactive node. Both independent and parallel jobs can be launched this way.

For more information, see *https://www.chpc.utah.edu/documentation/software/matlab.php*

CHPC Annual PI Survey

This January CHPC will be sending out our annual survey to our active PIs. The survey asks how CHPC services and resources contribute to your group's research.

At that time, we will also be asking active PI's for a list of funding sources, publications, dissertations, presentations or any other products resulting from CHPC's support in 2016. This survey is your opportunity to provide us with valuable feedback on how we can best meet your research computing needs. Your responses are critical to us and are an important component of our budget justification. Thank you for your support.

NEW Archive Storage Option Available

CHPC now has a new archive storage solution based around object storage, specifically CEPH, a distributed object store suite developed at UC Santa Cruz. We have an initial raw capacity of 1.15PB, with a cost of $80/TB raw space. In order to calculate the cost per TB of usable space you must consider the replication configuration. Initially, we are offering an 6+3 erasure coding configuration which results in a price of $120/TB of usable capacity for the 5 year lifetime of the hardware. As we currently do with our group space, we are operating this space in a condominium model by reselling this space in TB chunks.

This space is a stand-alone entity, and will not be mounted on other CHPC resources. One of the key features of the archive system is that users manage the archive directly, unlike the tape archive option. Users can move data in and out of the archive storage as needed -- they can archive milestone moments in their research, store an additional copy of crucial instrument data, or retrieve data as needed. This archive storage solution will be accessible via applications that use Amazon's S3 API. GUI tools such as transmit (for Mac, ***see http://www.panic.com/transmit/***) as well as command-line tools such as s3cmd and rclone (***http://rclone.org/*** , CHPC web page on usage at ***https://www.chpc.utah.edu/documentation/software/rclone.php***) can be used to move the data.

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 Our Systems!

Welcome to CHPC News!

If you would like to be added to our mailing list, please fill out this form and return it to:

Colette Durrant
THE UNIVERSITY OF UTAH
Center For High Performance Computing
155 S 1452 E ROOM 405
SALT LAKE CITY, UT 84112-0190
FAX: (801)585-5366

Name:
Phone:
Email:

Department or Affiliation:

Address:
(UofU campus or U.S. Mail)

Please help us in continuing to provide our users with excellent service, support, and access to cutting edge equipment.

ACKNOWLEDGEMENTS

If you use CHPC computer time or staff resources, we request that you acknowledge CHPC's contribution in any technical reports, publications, and/or dissertations.

Here is an example of what we ask you to include in your acknowledgments:

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

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

The University of Utah
Center for High Performance Computing
155 South 1452 East, Rm #405
Salt Lake City, Utah 84112-0190

-We appreciate your submission!

If you prefer Email, please send submissions of acknowledgment and/or requests to be added to CHPC's mailing list to: COLETTE.DURRANT@UTAH.EDU

*****All requests and submissions must include the required information as indicated above.*****