CRCF Annual Public Report

VILLANOVA UNIVERSITY
CENTRALIZED RESEARCH
COMPUTING FACILITY

Academic Year 2024
Last updated 5/28/24

Contents

Overview 2
CRCF HPC Cluster Specifications 2
Alipi & Clusty 4
Key Personnel (AY 2024) 4
Strategic Plan (5-year, 2023-2028) 5
Accomplishments 6
Administrative 6
Technical 6
Metrics 7
Work in Progress 11
User Survey Responses 11
Final Points 12
Good Citizenship 12
Interested in Helping CRCF? 12

Overview

The Centralized Research Computing Facility (CRCF) is Villanova's first core facility (established November 2021) and provides support for computational research. Most research support is in the form of high performance computing (HPC), but other forms of support are available as well. We encourage all Villanova researchers to contact us if they are interested in research computing but do not yet have a strategy towards achieving their goals.

CRCF supports three HPC clusters and their users:

- Augie: this is CRCF's largest and newest cluster. It was procured through an NSF Campus Cyberinfrastructure (CC*) grant in Summer 2020. The cluster was installed in March 2021. Augie is open to all Villanova researchers and Villanova students.
- Alipi: this cluster is limited to College of Engineering researchers. Alipi is a couple of years older than Augie. Alipi is also smaller than Augie.
- Clusty: this is the Astrophysics cluster that has been opened up to all users. The cluster is managed by Andrej Prsa in the Astrophysics and Planetary Science department.

CRCF emphasizes that *centralized computing is more than just HPC equipment*. CRCF currently provides the following to the Villanova community:

- State-of-the-art HPC resources
- A central knowledge base/pool of expertise for research computing resources
- A partner to help faculty integrate HPC into their coursework and proposals
- A resource for developing computer programming and software applications skills for students, staff, and faculty. This resource follows Augustinian values of helping others (akin to service learning)

In summary, *CRCF is a partnership of faculty, technical staff and students that empowers Villanova to maximize its computational research potential.*

CRCF has three thrusts as part of this mission/vision:

Thrust #1: HPC Support through a Collaboration of Faculty and IT Staff. This thrust involves the management of Augie, Alipi, and Clusty.

Thrust #2: Research Computing Center of Expertise. CRCF provides training opportunities including on-demand new user training, webinars for research computing topics of interest, and a point of contact for interaction with experts at other institutions. All researchers interested in research computing or HPC should approach CRCF to help develop a strategy to meet their goals.

Thrust #3: HPC Peer User Support. A wealth of research computing expertise exists on campus, and CRCF strives to facilitate communication between users seeking help and the experts that can help them. These volunteer experts personify Villanova's Augustinian tradition, providing support for commonly used programming languages and applications.

CRCF HPC Cluster Specifications

Augie

Item	Values
Compute Nodes/Cores	31 compute nodes/1,984 AMD EPYC Series
RAM	14.5 TB Main Memory
Disk Space	292 TB
GPU	2x Tesla A100: 13.8k CUDA cores for throughput of 19.5 TFLOPS
Interconnect	Mellanox 200 Gbps
Software	LAMMPS, Quantum Espresso, COMSOL, Python, CP2K, MATLAB, VASP, R, PyMT, TensorFlow, Compilers (GCC, AOCC)

The status of Augie's applications is as follows:

1. The following software applications have been installed and performance tested, with sample scripts available to users: CUDA, C/OpenMPI, Fortran/OpenMPI, MATLAB, Python, R, and COMSOL.
2. The following software applications has been installed, sample scripts have been provided in some cases, and performance testing is in progress: PyTorch, Quantum Espresso, LAMMPS, CP2K, VASP, IQTREE, PyMT, TensorFlow, Intel compilers, AMD compilers, PAML, and RAXML.
3. The following software applications are slated for installation: Gaussian, Abaqus, OpenFOAM, and Singularity.
4. The following software applications need further exploration or discussion prior to initiating an installation effort: Mathematica, ANSYS, and Anjuta.

Only those software packages listed in Items 1-3 above are currently supported by CRCF. Users are encouraged to install their own software packages in their home directories for their own use.

Users are encouraged to visit the Augie App Notes folder on the AugieUsers SharePoint site, which contains lessons learned and names of power users for many of the software applications listed above.

Alipi & Clusty

Cluster	Alipi	Clusty
Cores	208	212
RAM	1,408 GB	208 GB
Disk Space	26 TB	15 TB
GPU	N/A	N/A
Interconnect	InfiniBand	10 Gbps
Software	Python, LAMMPS, QuantumEspresso, VASP, MATLAB Compilers: Akantu, Intel	Ubuntu server, Python, Perl, R, sqlite3, octave

Key Personnel (AY 2024)

CC* Committee: Joseph DeMarco (student, CLAS), Jonathan Hardy (UNIT), Chris Washburn (Research Computing Administrator), Aaron Wemhoff (COE, Administrative Director)

Key UNIT & CLAS IT Support (Augie): Jonathan Graziola, Leo Nelson, Peter Palladino, Gavin Printz

HPC Advisory Board: David Cereceda (COE), Arup Das (VSB), Justin DiBenedetto (CLAS), Andrej Prsa (CLAS), Michael Robson (Smith College), Jason Simms (Swarthmore College), Daniel Smith (CON), Bill Wagner (VSB)

COE HPC Committee (Alipi): David Cereceda (ME), Zuyi Huang (CBE), Kyle Juretus (ECE), Chengyu Li (ME), Ondrei Miller (UNIT), Arash Tavakoli (CEE), Aaron Wemhoff (ME, chair)

Strategic Plan (5-year, 2023-2028)

Mission: Support and promote computational research for Villanova researchers and their collaborators

Vision: CRCF, as a center of expertise in computational research, is one of the top contributors to Villanova's research enterprise.

Goals & Metrics (5-year goal in parentheses, AY2024 value after colon):

1. CRCF is a significant enabler of research productivity
    a. Number of grant proposals submitted that use CRCF resources (10/yr.): 13
    b. Number of peer-reviewed publications produced using CRCF resources (10/yr.): 14
2. CRCF is the center of expertise for computational research
    a. Number of workshops or tutorials offered (4/yr.): 3
    b. Number of researchers using external services (OSG, PATh, XSEDE/ACCESS) (5/yr.): 1 user referred to OSG for possible high throughput computing implementation
    c. Number of CI grant proposals stemming from the center (1/yr.): 1 in progress
3. CRCF has a robust multidisciplinary community of computational researchers
    a. Number of total users (170): 168
    b. Number of active users (40/mo.): average of 16 per month, max 26
    c. Number of departments represented by active CRCF researchers (12): 9
    d. Number of power users per software application (3): 0 or 1
    e. Number of software applications with power users (12): 7
    f. Number of classes that use Augie for instruction (8/yr.): 4 in the past year
4. CRCF's resources grow to match demand
    a. Low idle time (< 15%): 32%
    b. Small average job wait time (< 12 hours): we are still looking on ways to track this
    c. Number of new condoers per year (2): 2
    d. Acting as a host site for REDCap: pushed to data storage investigation
    e. Dedicated HPC IT Admin: done
5. CRCF operates for the good of all humankind beyond research
    a. Request that hosting data center must meet energy efficiency requirements (PUE < 1.5): Netrix does not track PUE; request made for them to do so
    b. Request that data center have PPA in place for renewable energy: Netrix does not do this; request has been made for them to do so

Accomplishments

CRCF's accomplishments for the AY have been split into administrative and technical categories.

Administrative

The team spent the AY continuing to grow CRCF to ensure long-term use and management of the facility. Highlights of the facility's administrative accomplishments include:

1. Successfully completed and closed out NSF CC* award that funded Augie's equipment.
2. Renewed and grew the COMSOL users group and associated licenses.
3. Renewed ANSYS-FLUENT research license and worked with UNIT to address licensing issues.
4. Held workshops on Python & R, PHOEBE, and Github.
5. Started examining campus data storage needs/issues.
6. Met with, and distributed HPC information to, candidates in two faculty search efforts.
7. Held our first informal CRCF Meet & Greet event.
8. Formalized external user onboarding process.
9. Updated Augie training materials.
10. Signed up key personnel as new NSF ACCESS Campus Champions.
11. Formalized CRCF Type Treatment.
12. Held successful HPC IT Administrator search.

Technical

Augie continues to evolve technically to meet the needs of the user community. This AY featured a permanent research computing administrator hire to address myriad technical issues. Some highlights:

1. Added CUDA scripts and performance tests.
2. Installed 2 condo nodes and ordered 6 more for installation this summer.
3. Attended multiple professional development workshops on HPC administration.
4. Researched and determined which operating system to transition to (Ubuntu) and the proper procedure to do so.
5. Adjusted Open Science Grid (OSG) priorities to meet NSF grant mandate.
6. Enforced SLURM batch limits.
7. Enabled condo computing jobs to have an infinite time limit.
8. Established simple onramp tutorials for common languages and apps.
9. Updated SLURM and BeeGFS.
10. Provided updates to COMSOL as they became available.

Metrics

Total number of users (as of 3/19/24)

- Augie: 168
- Alipi: 53
- Clusty: 169 (note: all students in astrophysics & planetary science are on by default)

Active Departments

The following departments contain active users since March 2023:

- CLAS: Biology, Chemistry, Computing Sciences, Math/Stats, Physics, Psychological and Brain Sciences
- COE: Civil and Environmental Engineering, Electrical and Computer Engineering, Mechanical Engineering

Publications using CRCF resources

- J. Debenedetto (2023) Byte-ranked Curriculum Learning for BabyLM Strict-small Shared Task 2023. *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning.*
- J. Debenedetto (2024) Linearization Order Matters for AMR-to-Text Generation Input. To appear in the *Proceedings of the Unified Meaning Representation (UMR) Parsing Workshop*.
- D. I. Senadheera et al. (2024) Probing the Electrode–Electrolyte Interface of Sodium/Glyme-Based Battery Electrolytes. *J. Phys. Chem. C,* 128(14), 5798-5808.
- D. Bialas and R. Jorn (2024) From Liouville to Landauer: Electron transport and the bath assumptions made along the way. *J. Chem. Phys.,* 160, 184109.
- M.E. Sarrett and J. Toscano (2024). Decoding speech sounds from neurophysiological data: Practical considerations and theoretical implications. *Psychophysiology,* 61(4), e14475.
- S. Lionetti, Z. Lou, A. Herrera-Amaya, M. Byron, and C. Li (2023) A new propulsion enhancement mechanism in metachronal rowing at intermediate Reynolds numbers. *Journal of Fluid Mechanics,* 974, A45 (2023).
- M. Lei and C. Li (2023) Wings and Whiffs: Understanding the role of aerodynamics in odor-guided flapping flight. *Physics of Fluids*, 35, 121901.
- M. Lei, M. Willis, B. Schmidt, and C. Li (2023) Numerical investigation of odor-guided navigation in flying insects: Impact of turbulence, wing-induced flow, and Schmidt number on odor plume structures. *Biomimetics*, 8(8), 593.
- S. Lionetti, T. Hedrick, and C. Li (2024) Sparse reduced-order modeling of a hovering hawkmoth's wake. *Proceedings of ASME FEDSM*, Anaheim, California.
- Z. Lou, M. Lei, M. Byron, and C. Li (2024) A computational analysis of fluid-structure interaction in metachronal propulsion. *Proceedings of ASME FEDSM*, Anaheim, California.
- M. Lei, Z. Lou, J. Wang, H. Dong, and C. Li (2023) Hydrodynamics of metachronal rowing at intermediate Reynolds numbers. *Proceedings of ASME IMECE*, New Orleans, Louisiana.
- A. Menzer, M. Lei, C. Li, and H. Dong (2023) A Multiphysics approach to understanding chemoreception in fish schools. *Proceedings of ASME IMECE*, New Orleans, Louisiana.

- Z. Lou and C. Li (2023) Unsteady aerodynamics and wake structures of butterfly in forward flight. *Proceedings of AIAA Aviation*, San Diego, California.
- S. Lionetti, T. Hedrick, and C. Li (2023) Numerical investigation of olfactory performance in upwind surging hawkmoth flight. *Proceedings of AIAA Aviation*, San Diego, California.
- Publications in preparation or currently under review
    - V. Smith, “A Deep-Learning Framework for Flood Dynamics Modeling Using a Convolutional Neural Network Integrated with a Sequential Model”
    - V. Smith, “A Deep-Learning Model for Sediment Transport and Geomorphic Simulation”
    - Yichen Qian, Mark R. Gilbert, Lucile Dezerald, Duc Nguyen-Manh, David Cereceda “Unraveling the energetics and the local chemical order in tungsten-based alloys through first-principles calculations”; Journal of Alloys and Compounds.
    - Jie Peng, Yichen Qian, David Cereceda, “A First-principles Study on the Structural-thermodynamics Properties of Tungsten-based High Entropy Alloys”; PRM.
    - Md Rajib Khan Musa, Yichen Qian, David Cereceda, “Accelerating the discovery of low-energy structure configurations: a computational approach that integrates first-principles calculations, Monte Carlo sampling, and Machine Learning”, PNAS Nexus.
    - B. Styer, "At most one solution to $a^x + b^y = c^z$ for some ranges of a, b, c"
    - B. Styer, "Number of solutions to $a^x+b^y=c^z$ with $\gcd(a,b)>1$"
    - M. Gouvea Gruppi, “On the Effects of Fine-tuning Language Models for Text-Based
    - Reinforcement Learning”
    - M. Gouvea Gruppi, “Analyzing language change in fiction and society”

Conference presentations using CRCF resources

- J. Debenedetto (2024) Evaluating the Effects of Quality Estimation Filtering on Machine Translation. MASC-SLL 2024.
- J. Debenedetto (2024) Quantifying Translationese: Computational Analysis of Lexical, Cohesive, and Syntactic Features in Translated Texts. MASC-SLL 2024.
- M. Said, conference poster.
- V. Smith (2024) A Deep-Learning Model for Continuous Flood Extent and Water Depth Maps. CSDMS 2024 Annual Conference.
- M.R.K. Musa, Y. Qian, and D. Cereceda (2023) Accelerating the Discovery of Low-Energy Structure Configurations: A Computational Approach that Integrates First- principles Calculations, Monte Carlo Sampling, and Machine Learning. HEA 2023, Pittsburg, USA.
- J. Peng, Y. Qian, D. Cereceda (2023) A First-principles Study on the Structural-thermodynamics Properties of Tungsten-based High Entropy Alloys. HEA 2023, Pittsburg, USA.
- D. Cereceda (2024) Effects of the chemical environment and energetics of point defects on transmuting tungsten-based alloys. 7th IEA FM TCP workshop on theory and modelling of nuclear fusion materials, Incheon, Korea.
- N. Haider, Z. Lou, B. Cheng, and C. Li (2023) Flight and smell: Exploring the impact of wing structure and kinematics on the olfactory function of flies in upwind surging flight. 76th Annual Meeting of the American Physical Society (APS) Division of Fluid Dynamics, Washington, DC.
- Z. Lou, N. Tack, M. Wilhelmus, and C. Li (2023) Exploring the hydrodynamic advantages of pleopod interaction in shrimp swimming. 76th Annual Meeting of the American Physical Society (APS) Division of Fluid Dynamics, Washington, DC.

- S. Lionetti, T. Hedrick, and C. Li (2023) Reduced order modeling of wake structures in hawkmoth hovering flight. 76th Annual Meeting of the American Physical Society (APS) Division of Fluid Dynamics, Washington, DC.
- M. Lei, J. Wang, H. Dong, and C. Li (2023) Fluid-structure interaction modeling of fruit fly wings in hovering flight," 76th Annual Meeting of the American Physical Society (APS) Division of Fluid Dynamics, Washington, DC.
- Z. Lou, M. Lei, M. Byron, and C. Li (2023) Propulsion through asymmetry: Examining flow-structure interactions of Ctenophore appendages with asymmetric stiffness. 76th Annual Meeting of the American Physical Society (APS) Division of Fluid Dynamics, Washington, DC.

Proposals submitted that would utilize CRCF resources

- J. Debenedetto, internal University Summer Grant Program (funded).
- J. Debenedetto, student VURF project (funded).
- R. Jorn, DOD grant submission (WP25-C1-4664).
- V. Smith, EPA Stormwater Center for Excellence.
- V. Smith, NSF Smart and Connect Communities Full Grant.
- M. Gouvea Gruppi, "On the Feasibility and Practicality of Accelerated Regular Expressions", VURF.
- M. Gouvea Gruppi, "Multi-level and multi-modal networks of news producers", in preparation.
- C. Li, "Deciphering the influence of unsteady aerodynamics on mechanosensation and olfaction in insect flight", AFOSR YIP, $449,806 (awarded).
- C. Li, "Tuning the Unsteady Aerodynamics of Swept Wings with Adaptive Lifting Surfaces," 2023 University Summer Grant Program, $12.5k (awarded).
- C. Li, "Revealing the flow physics of bio-inspired flexible propulsors for fast and efficient swimming" DARPA Young Faculty Award, $1,000,000 (in preparation)
- C. Li, "Understanding, predicting, and mitigating the extreme aerodynamics of dynamic ground effect", ONR, $739,859 (unfunded)
- C. Li, "Multi-fidelity computational approaches for diagnosis of nasal sinus diseases" NIH R21 Trailblazer Award, $576,483 (unfunded).
- C. Li, "Olfactory-cued fish schooling for fast and efficient movement cohesion" ONR Young Investigator Program, $749,837 (unfunded).

Miscellaneous

- R. Jorn: Augie is sole source of computational support for active NSF grants CHE-1900432 and CHE-2154505.
- C. Li: AFOSR Young Investigator Program (YIP) Award, 2024
- C. Li: AIAA Flow Visualization Showcase (1st Place), AIAA AVIATION Forum, 2023

Courses where Augie, Alipi, or clusty have been integrated

- BL 2149 TOP:Cyber Law
- CSC 2405 Computer Systems II
- CGS 5990: Fairness in AI (Augie)
- CGS 5900/PSY 8900 Cognitive Science Seminar (Augie)

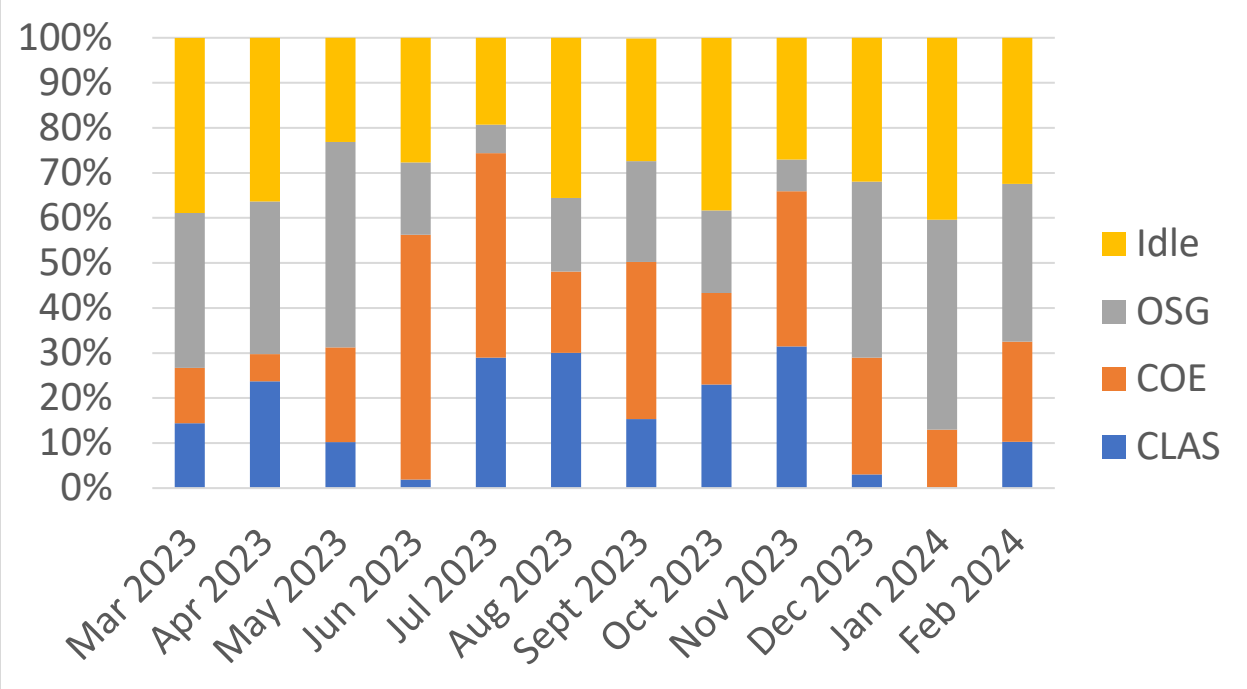
- Mathematics Senior Seminar
- ME 3600 Fluid Mechanics
- ME 3950 Heat Transfer I
- ME 7030: Numerical Methods in Engineering Simulation
- SPA 3200 Introduction to Spanish Translation

CRCF workshops offered

- Tutorial on PHOEBE (Andrej Prsa)
- Tutorial on R and Python (Peter Chi, Mohamed Said)
- Tutorial on Github (Joseph DeMarco)

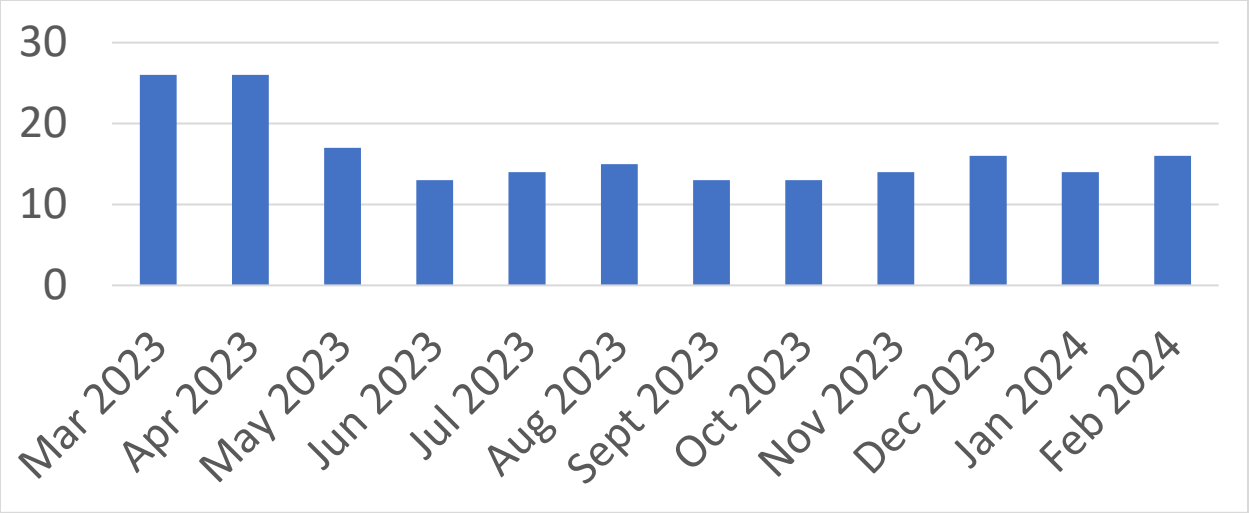
Augie Usage

The usage goals for Augie are (1) less than 20% idle time, and (2) an average of 20% usage by the Open Science Grid (OSG). In the last 12 months, the idle time is slightly high (32%), and the OSG usage is slightly high (27%). These are both acceptable values for purposes of ensuring that Augie's growth matches demand and that the NSF CC* grant requirements are met. While CLAS and COE demand vary widely, the average Augie usage per college is 16% for CLAS and 26% for COE, with zero usage by VSB and FCON.



Number of Active Users

Active users are defined as those that submitted jobs to Augie within a specified month. The number of active users over the past 12 months peaked in March and April 2023, and then settled at around 14 over the remaining months. The annual average was therefore 16.2 users.



Work in Progress

The following efforts are currently ongoing:

- Administrative
  - Address rapidly increasing housing costs.
  - Improve the approach to track papers, proposals, and HPC integration into courses
  - Hold at least 1 workshop per semester using survey to gauge interest on topics
  - Work with UNIT on a data storage solution
  - Recruit power users
  - Increase active usage
  - Track the average job wait time
  - Create formal policies on condo pre-emption and use of long qos jobs
  - Hire CRCF's first HPC summer intern
- Technical
  - Transition OS to Ubuntu
  - Install Gaussian and Abaqus
  - Enable containers (singularity) and install OpenFOAM
  - Establish two interactive nodes with Open OnDemand on one of them
  - Consider data backup to cloud storage

User Survey Responses

A survey was provided to the Augie Users in March 2024. Little feedback was garnered outside of indications of satisfaction with the current state of affairs. User suggestions have been integrated into the preceding section.

Final Points

Good Citizenship

Some reminders:

- No running jobs on the head node – submit batch jobs instead
- Don't use the debug queue for production runs
- Don't submit lots of jobs to occupy a large percentage of cores on the cluster
- Be sure to acknowledge use of the HPC clusters – see user terms and conditions documents:
    - Augie: Augie Users SharePoint, file Administrative/Augie-HPC-TC_20210507.pdf
    - Alipi: Villanova HPC Team → Alipi Users Channel, file Alipi Usage Policies.docx
    - Clusty: contact Andrej Prsa
- Need help on Augie/Alipi/Clusty?
    - Don't contact UNIT directly or put in a ticket
    - Augie:
        - Find appropriate document in the Augie App Notes documents. See if the document specific to your app has the information you need.
        - Many documents have listed superusers; contact the superuser for help.
        - If the above two items don't work, then email ccstarcommittee@villanova.edu.
    - Alipi: post issue in Teams channel. If no response in 24 hours, then email engineering-hpc@villanova.onmicrosoft.com.
    - Clusty: contact Andrej Prsa

Interested in Helping CRCF?

Contact ccstarcommittee@villanova.edu if you are interested in…

- Supporting other Villanova researchers as a power user
- Collaborating with a researcher at a small local college
- Providing additional feedback on CRCF operations
- Working as an undergrad support software technician during the summer
- Providing your thoughts about how to make AugieFest a popular, fun event
- Providing ideas for tutorials