CCAST: Advanced Research Computing Resources

Enabling computational research and education in the state of North Dakota

Khang Hoang and **Dane Skow**

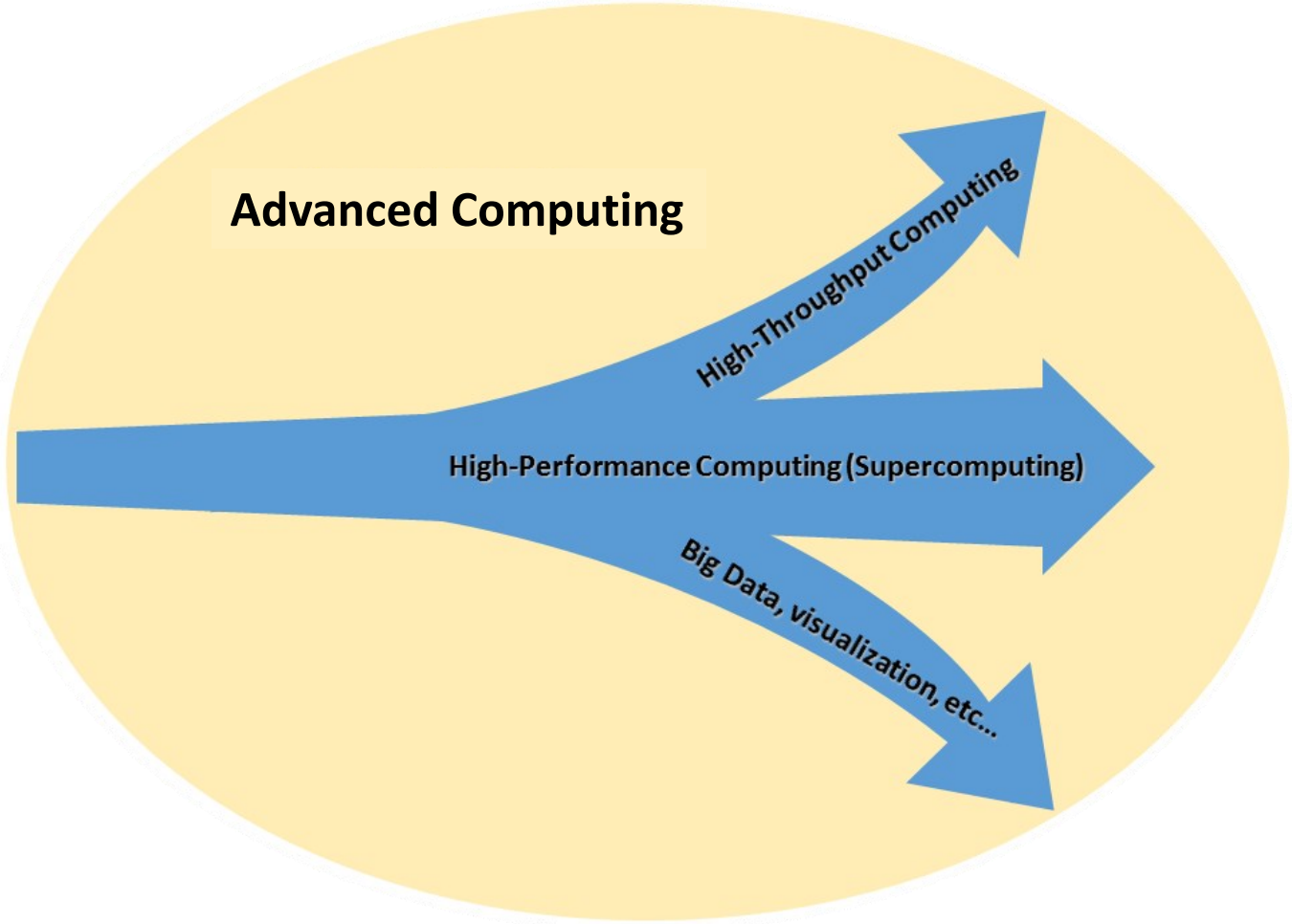
Center for Computationally Assisted Science and Technology (CCAST)

Information Technology Division, North Dakota State University

ndsu.ccast.support@ndsu.edu | www.ndsu.edu/ccast

Annual Faculty and Academic Staff Conference, 8/17/2022

Advanced Research Computing (ARC) is now indispensable



ARC = **advanced computing** done in support of or related to research activities

Exciting time in computational research and education

- Advances in high-performance computing (HPC)
  - Petaflops (10^{15}) to exaflops (10^{18}), hybrid CPU/GPU nodes...
  - Numerical and parallel algorithms, code development...
  - Next: hybrid HPC/quantum computing?

NERSC/LBNL/DOE

- More powerful computers allow solving more complex and bigger problems, usually enabling better science
  - "Law of Constancy of Pain"! (Craig Good, original Pixar employee)

CCAST/NDSU

- Hardware and software have become more complex and usually harder to use
  - HPC is parallel computing!
  - HPC resources are shared resources, and ALWAYS limited
    → Need to be used properly and efficiently

- **For researchers, HPC knowledge/skills are no longer optional!**

Challenges in accelerating scholarly discovery

- **Researchers possess varied technical knowledge**
  - Nontraditional HPC users often unfamiliar with existing computing resources...
  - ~~Many~~ Most HPC users don't know how to use the resources effectively...
- **Appropriate technology solutions require an understanding of research needs**
  - Physical vs. life and social sciences: e.g., a typical HPC configuration may not be optimal for all
  - Two-way communication to identify gaps and weaknesses in research computing solutions
- **Effective research support usually requires scholarly expertise**
  - Understanding of scholarly processes and pressures, research culture... is extremely useful

Adapted from L. Michael and B. Maas, *ECAR Res. Bull*. (May 16, 2016)

What is research computing (RC) facilitation?

- Goal: to advance researchers' capabilities and hence accelerate scholarly discovery

- Major facilitation activities:
  - **Promoting** awareness of RC resources and their potential impacts
  - Engagement with researchers to **understand** their needs and **advise** on RC strategies
  - Ongoing **support** of researchers executing projects on computing resources
  - **Education** and **training** of researchers re. computing capabilities, best practices, and specific skills
  - **Liaising** researcher connections
  - **Advocating** for the needs of researchers to inform RC design and institutional support

- Proactive engagement, personalized guidance, "teach-to-fish"...

Adapted from L. Michael and B. Maas, *ECAR Res. Bull*. (May 16, 2016)

What is CCAST?

The **C**enter for **C**omputationally **A**ssisted **S**cience and **T**echnology (CCAST; pronounced "*c-cast*"), a part of NDSU IT, provides advanced cyberinfrastructure for computational research and education at NDSU and beyond.

CCAST (i) develops, manages, brokers, and operates high-performance, cloud, and interactive computing resources, and (ii) educates researchers on proper and efficient use of the resources and on other topics of interest to the computational science and engineering community.

Basic services are provided at no charge. Dedicated services available at cost.

CCAST continually works to enhance NDSU's capabilities and competitive edge in disciplines and research that rely on advanced computing.

CCAST staff

Samuel Saula
HPC System Admin

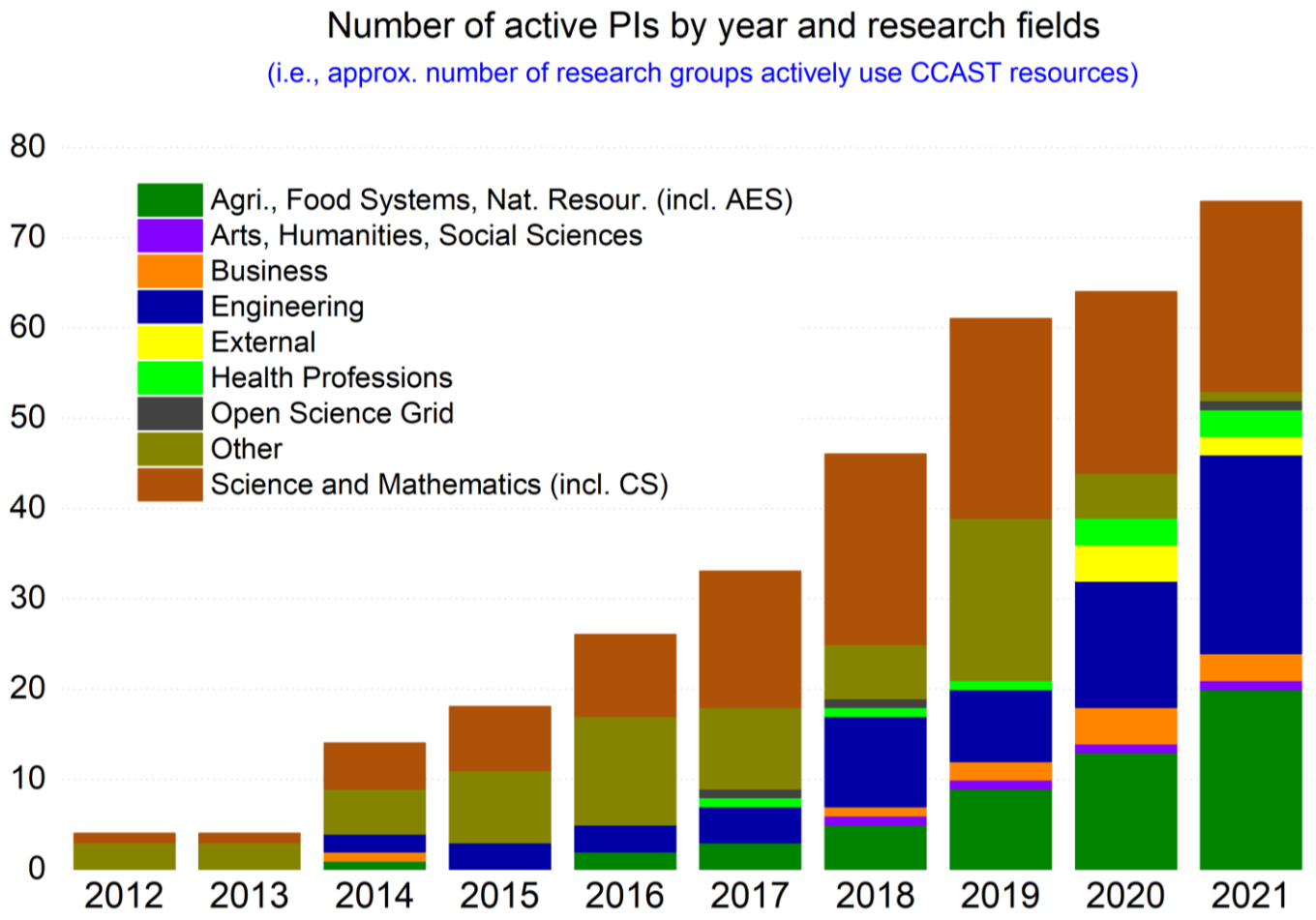
Dane Skow
Executive Director

Khang Hoang
Research Facilitator

Ryan Anderson
HPC System Admin

Nick Dusek
Research Facilitator

CCAST user community: Growing rapidly and getting more diverse



Campuses with active CCAST users: North Dakota State University, Cankdeska Cikana Community College, Dickinson State University, Valley City State University, University of North Dakota…

CCAST resources: Hardware, software, data transfer, cloud…

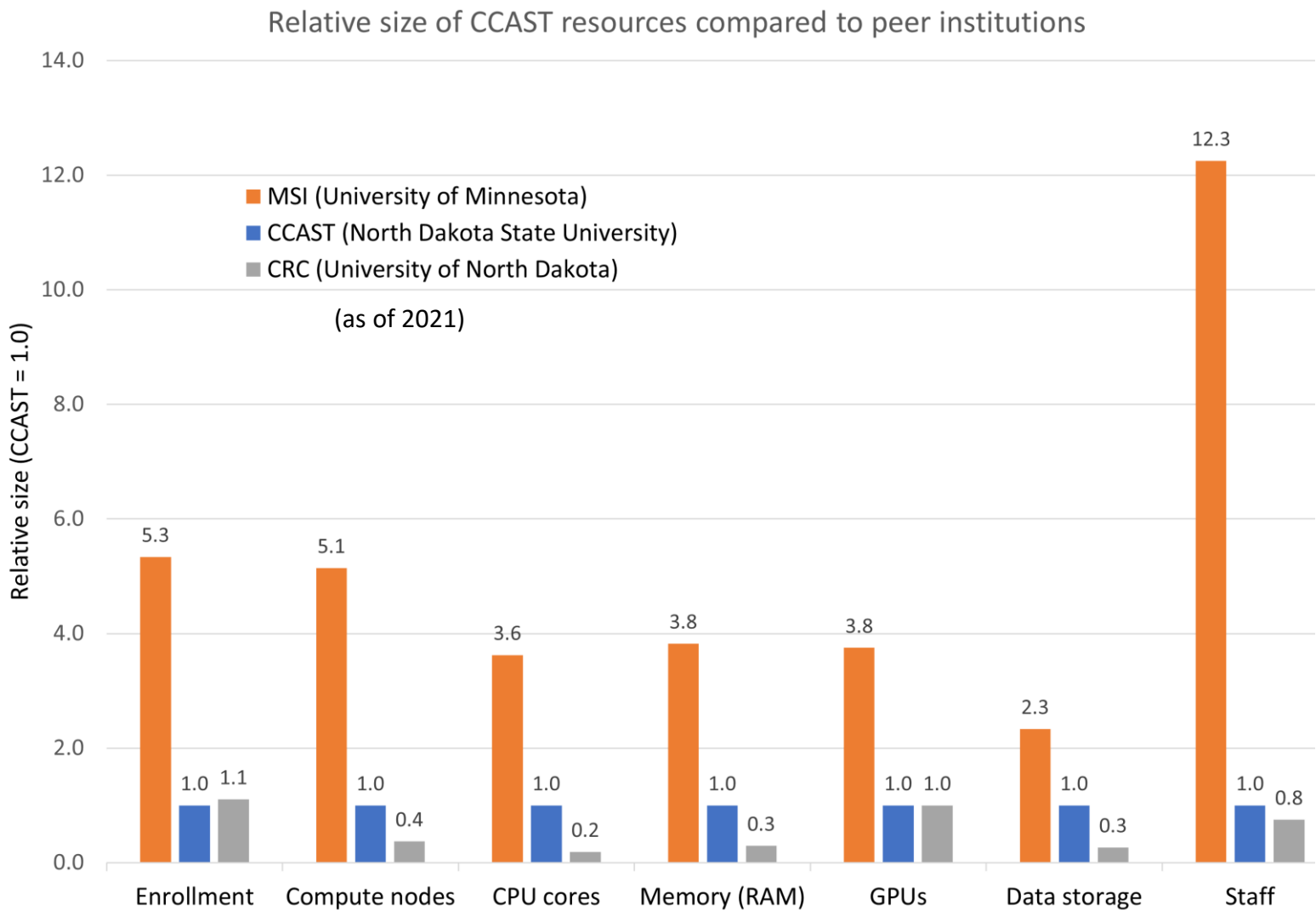
- **Hardware** has been procured almost entirely with external/non-appropriated funds
  - >10,000 Intel and AMD CPU cores and >50TB of RAM (by Fall 2022), incl. big-memory nodes
  - 2.2PB parallel filesystems; plus >1PB research data archive (by Fall 2022)
  - 54 general-purpose GPUs (by Fall 2022), incl. NVIDIA A100s

Basic level of services is FREE to NDSU researchers and certain external collaborators!

Researchers can purchase "condo" (i.e., researcher-owned) compute or storage units

- **Software**: various
  - General libraries/compilers
  - Specific applications in different areas
- **Fast data transfer** via Globus & ScienceDMZ
- **Cloud services** via…
  - Microsoft Azure
  - Rescale
  - Lancium

How big is CCAST compared to its peers?



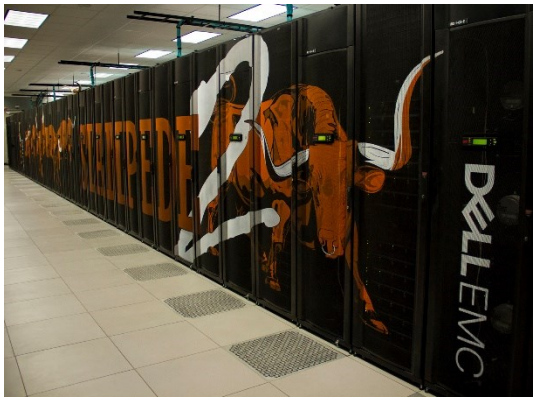
HPC resources available elsewhere: UND

Cluster	Number Nodes	Total Cores	Total Memory (TB)
Talon	20	720	3.8
Talon Large Memory	7	504	21 (3 TB/node)
Talon GPU	4 (8 GPUs/node)	144	6
Hodor	32	256	4
Arya	6	128	1

For more info on CRC, visit: https://und.edu/research/computational-research-center/

HPC resources available elsewhere: National computing facilities

- ACCESS (Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support, supported by NSF): a national research cyberinfrastructure ecosystem, free to U.S.-based researchers.



Stampede2 (TACC), via XSEDE

Cori (LBNL), NERSC

- NERSC (National Energy Research Scientific Computing Center): the primary scientific computing facility for the Office of Science in the U.S. Department of Energy.

- OSG (Open Science Grid): a consortium which facilitates distributed (mostly high-throughput) computing, aggregates power of tens of thousands of (Linux) computers, scattered around the globe.

CCAST Advanced Research Computing Training Program

- Regular training program offered during Spring and Fall:
  - Introduction to high-performance computing (HPC)
  - Linux for HPC: Working with Linux-based HPC systems
  - Parallel computing
  - Linux for HPC: Text processing and shell scripting
  - Accelerated computing with GPUs
  - HPC for machine learning and big data
  - HPC for bioinformatics/precision ag/materials modeling
  - Running Python/R/MATLAB… on HPC systems
  - Introduction to quantum computing
  - etc.

- Registration/attendance:
  - 2019: in-person, 100* people attended (65% grad, 4% undergrad, 15% faculty, 16% staff/postdoc)
  - 2020: in-person/via Zoom, ~120* registered (59% grad, 8% undergrad, 18% faculty, 15% staff/postdoc)
  - 2021: via Zoom, 234* people registered (58% grad, 25% undergrad, 12% faculty, 5% staff/postdoc)
  - 2022 (Spring): via Zoom, 108* registered (60% grad, 13% undergrad, 13% faculty, 14% staff/postdoc)

*NOT including those participated in other training activities.

- Open to faculty, staff, and students from all institutions within NDUS, the tribal colleges and universities in the state of ND, and institutions in the local Fargo-Moorhead area.

CCAST Internship Program in Advanced Research Computing

- Internship as (more in-depth) training and as workforce development
  - "Paid" interns: selected via competitive application process, financially supported by CCAST and/or partnering departments.
  - "Unpaid" interns: selected via discussion/special arrangement with faculty, financially supported by the faculty/academic department.

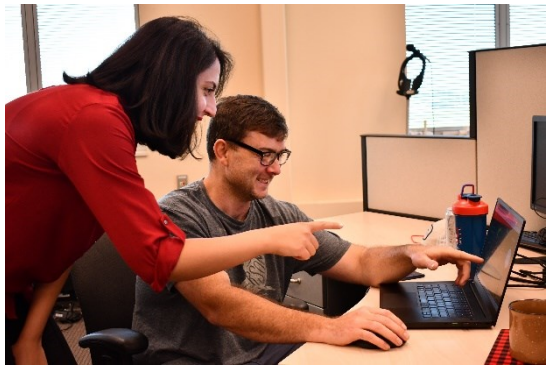
- Student interns work in the systems-facing and/or researcher-facing tracks
  - 2019: 16 paid + 3 unpaid; 2020: 6 paid + 5 unpaid; 2021: 8 paid + 1 unpaid; 2022: 10 paid* + 2 unpaid

*supported mainly by external grants

- **Constraints**: CCAST's specific needs, availability of funds, and staff's time.

Consulting, (more) training, proposal writing, collaboration…

- Personalized training and consulting for individual PIs/research groups
- Special training events
  - Inviting outside experts to campus
  - Hosting remote workshops



- User guides, tutorials, and other training materials
- Guest lectures/lab sessions on the applications of HPC
- Proposal writing and collaboration
  - Developing infrastructure, workshop, and research proposals
  - Collaborating with other PIs (at NDSU and beyond) in developing proposals
- Proposal writing assistance
  - Consulting on the research computing needs of the project
  - Providing an up-to-date description of CCAST resources (for "Facilities, Equipment and Other Resources")
  - Providing letters of support/letters of collaboration
  - Helping with computing resources and/or cost justifications, data management plans, etc.
  - Introducing PIs to national computing facilities (ACCESS, NERSC, OSG, etc.)

Summary

- Research computing (incl. HPC) resources are available for your research and teaching
- HPC resources are shared and ALWAYS limited. HPC knowledge/skills are required to use the resources properly and efficiently and to improve research productivity
- PIs: HPC knowledge and familiarity with computing facilities is needed to design and plan your research, and to supervise your group members effectively
- Integrate advanced computing into your teaching. Sooner is better.
- Talk to us regularly! A 30-minute session with us can save months of your (your group's) time.
- Make the best use of CCAST resources!

Contact CCAST at
ndsu.ccast.support@ndsu.edu