High Performance Computing at the Oak Ridge Leadership Computing Facility

Careers, Research Opportunities, and State-of-the-Art Technology

Verónica G. Vergara Larrea

CRA-W Undergraduate Town Hall

February 8th, 2018

CRA-W

Computing Research Association

Women

Speaker & Moderator

Verónica G. Vergara Larrea

Verónica G. Vergara Larrea is originally from Quito, Ecuador. Verónica earned a B.A. in Mathematics/Physics at Reed College and a M.S. in Computational Science at Florida State University.
Verónica has six years of experience in the high performance computing field and is currently working as an HPC User Support Specialist at the Oak Ridge Leadership Computing Facility. In addition to providing assistance to OLCF users, Verónica is part of the systems testing team and leads acceptance for Summit, ORNL's next generation supercomputer. Her research interests include high performance computing, large-scale system testing, and performance evaluation and optimization of scientific applications. Verónica is a member of both IEEE and ACM and serves in the ACM SIGHPC Executive Committee.

Lori Pollock

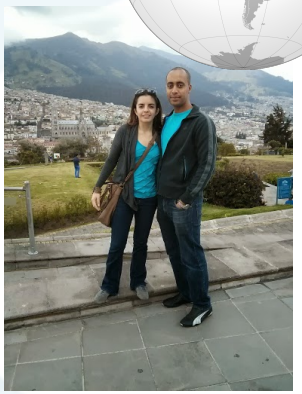
Dr. Lori Pollock is a Professor in Computer and Information Sciences at University of Delaware. Her current research focuses on program analysis for building better software maintenance tools, software testing, energy-efficient software and computer science education. Dr. Pollock is an ACM Distinguished Scientist and was awarded the University of Delaware's Excellence in Teaching Award and the E.A. Trabant Award for Women's Equity.

CRA-W
Computing Research Association
Women

About Me

Background & Career

I'm from Quito, Ecuador
"La Mitad del Mundo"



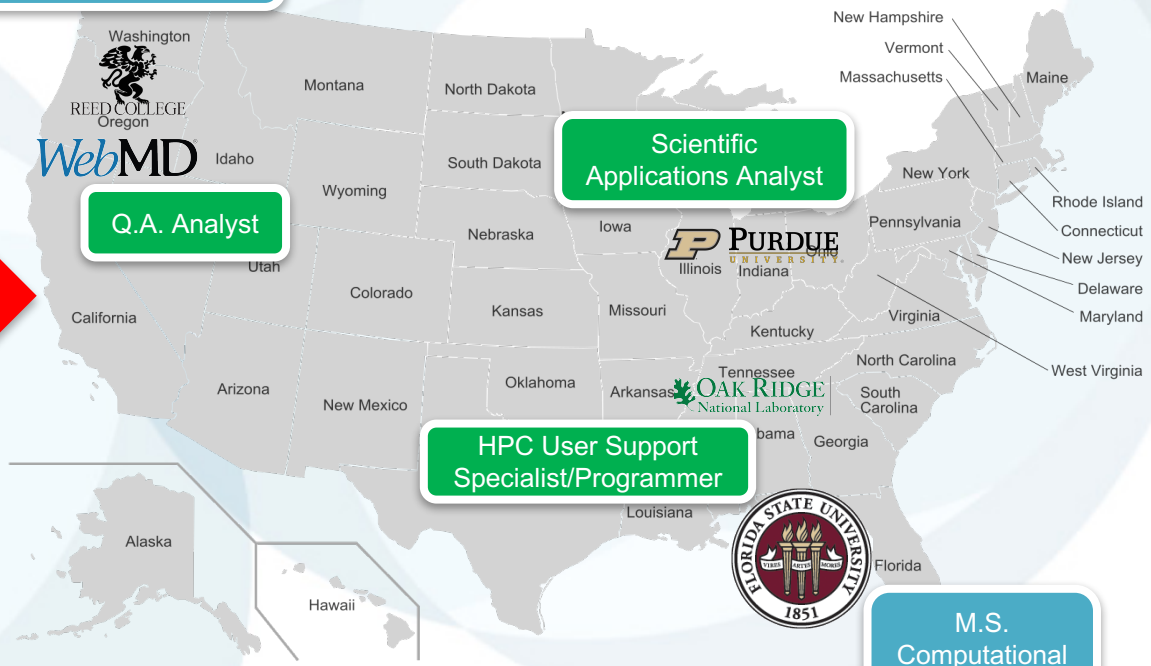
B.A.
Mathematics/Physics

Q.A. Analyst

Scientific
Applications Analyst

HPC User Support
Specialist/Programmer

M.S.
Computational
Science



CRA-W
Computing Research Association
Women

Outline

- What is the OLCF?
- Resources at the OLCF
- The what, why, how of HPC
- Accepting a large-scale system
  - Research at-scale
- Summary

- Mentoring topic:
  - Make the most of an REU

CRA-W
Computing Research Association
Women

What is the OLCF?

Oak Ridge Leadership Computing Facility

- Established in 2004 at Oak Ridge National Laboratory
- U.S. Department of Energy - Office of Science User Facility
- Open to nearly everyone in the world
- Free to use for non-proprietary work
- Allocations are merit based
  - Three allocation programs: INCITE, ALCC, Director's Discretionary
- One of two Leadership Computing Facilities
  - Oak Ridge Leadership Computing Facility at ORNL
  - Argonne Leadership Computing Facility at Argonne National Laboratory
- Deploy and operate computational and data resources required to tackle global challenges
- Offer leadership-class computing resources to researchers who have many of the largest computing problems in science

CRA-W
Computing Research Association
Women

OLCF Resources

CRAY

Titan

- Cray XK7
- #5 in the TOP500 (November 2017)
- 27 PF
- 18,688 nodes/299,008 cores
- AMD Opteron + K20x GPU per node

CRAY

Eos

- Cray XC30
- 736 nodes/11,766 cores
- Two 8-core Intel Xeon E5-2670
- Used for pre- and post-processing

RHEA

Rhea

- RHEL6 Linux cluster
- 512 nodes/8,192 cores
- Two 8-core Intel Xeon E5-2650
- 9 nodes have two K80 GPUs
- Used for pre- and post-processing

Titan: The OLCF's flagship supercomputer

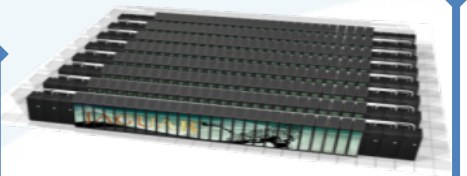
Size	18,688 Nodes	5,000 Sq. feet	
Peak Performance	27.1 PetaFLOPS	2.6 PF CPU	24.5 PF GPU
Power	8.2 MegaWatts	~7,000 homes	
Memory	710 TeraBytes	598 TB CPU	112 TB GPU
Scratch File system	32 PetaBytes	1 TB/s Bandwidth	

The next chapter: Summit, the OLCF's next generation supercomputer

Feature	Titan	Summit
Application Performance	Baseline	5-10x Titan
Number of Nodes	18,688	~4,600
Node performance	1.4 TF	> 40 TF
Memory per Node	32 GB DDR3 + 6 GB GDDR5	512 GB DDR4 + HBM
NV memory per Node	0	1600 GB
Total System Memory	710 TB	>10 PB DDR4 + HBM + Non-volatile
System Interconnect (node injection bandwidth)	Gemini (6.4 GB/s)	Dual Rail EDR-IB (23 GB/s)
Interconnect Topology	3D Torus	Non-blocking Fat Tree
Processors	1 AMD Opteron™ 1 NVIDIA Kepler™	2 IBM POWER9™ 6 NVIDIA Volta™
File System	32 PB, 1 TB/s, Lustre©	250 PB, 2.5 TB/s, GPFS™
Peak power consumption	9 MW	15 MW

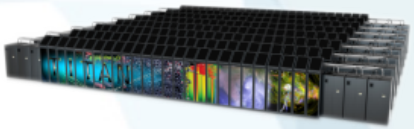
The Road to Exascale

The project office for the multi-lab Exascale Computing Project is based at ORNL



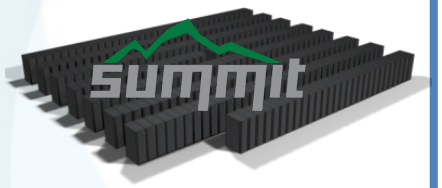
Jaguar: 2.3 PF
World's Fastest

2008



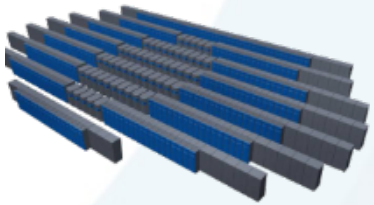
Titan: 27 PF
Accelerated Computing
World's Fastest

2012



Summit: 200 PF
Accelerated Computing
5–10× Titan Performance

2017



OLCF5 >1000 PF
Competitive Procurement
5-10× Summit
Performance

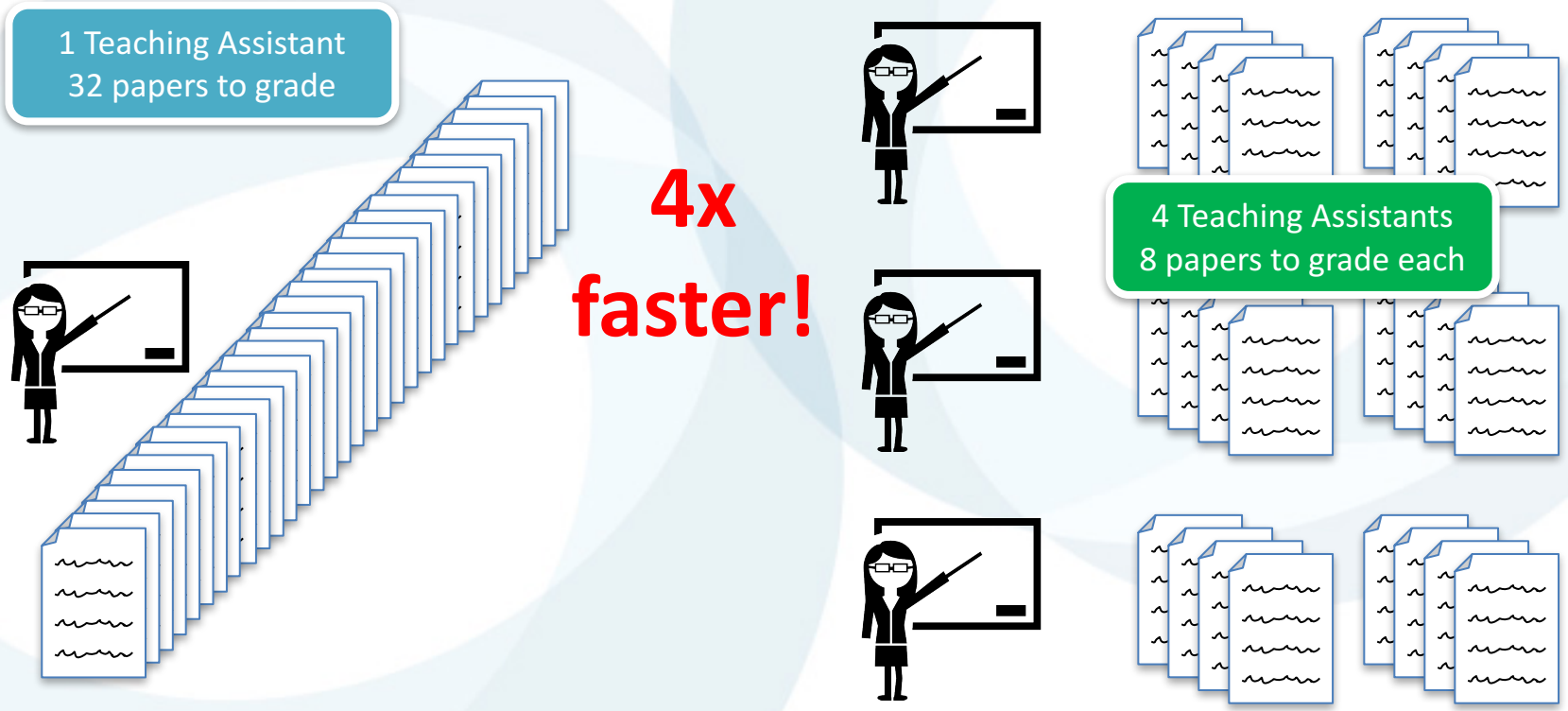
2021

"GPU Accelerated Computing: The Roadmap", ECP 2018 annual meeting.

The what, why, and how of HPC

What is HPC?

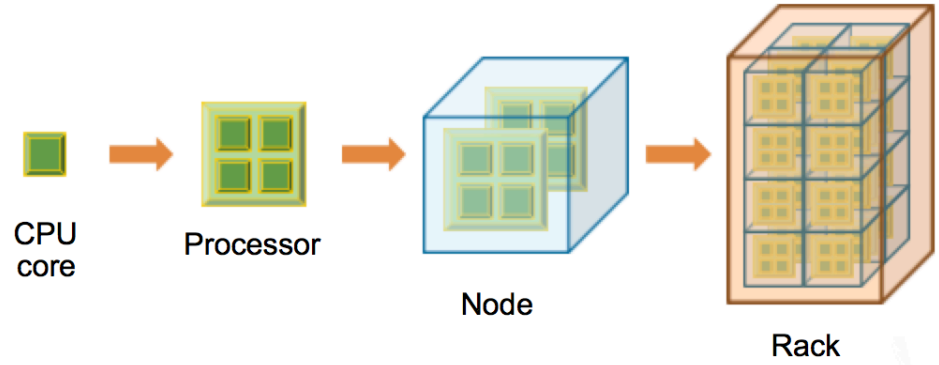
Thinking in parallel



CRA-W
Computing Research Association
Women

What is a supercomputer?

Divide a problem into many subunits that can be solved concurrently

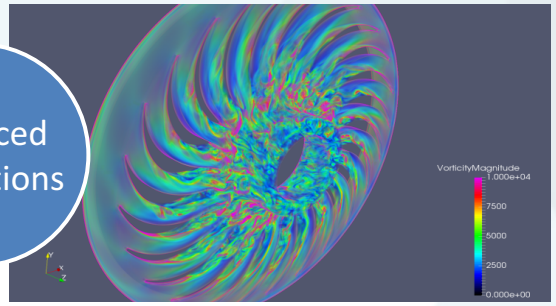


Why use HPC?

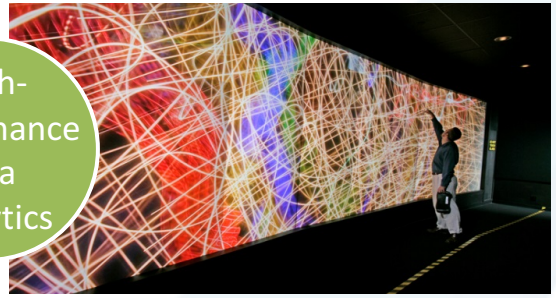
HPC allows us to...

- Investigate otherwise inaccessible systems at every scale
- Reduce the time-to-solution for researchers across many disciplines
- Increase the complexity of computational models to represent phenomena more accurately
- Analyze, process, and classify the ever growing data sets that exist today in a wide range of disciplines
- HPC impacts many industries: https://www.youtube.com/watch?v=9m0gZ2Gft4Q

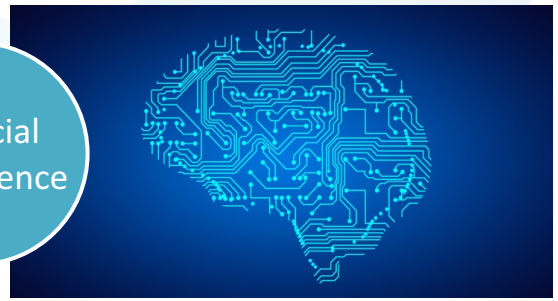
Advanced simulations



High-performance data analytics



Artificial intelligence



CRA-W
Computing Research Association
Women

How do you use HPC?

- High-level:
  - Divide your problem into subunits that can be solved concurrently
- But, not everything is embarrassingly parallel...
  - No problem! Programming models are available that allow different workers to share intermediate data
- Common programming models:
  - Distributed:
    - MPI, Charm++
  - Shared:
    - OpenMP, pthreads, OpenACC

Bonus: Who works in HPC?

Experts in many areas are required

HPC Operations	• Deploy compute resources and supporting infrastructure.
User Assistance & Outreach	• Install software, develop tests, create documentation, troubleshoot issues, and train users.
Scientific Computing	• Analyze and optimize the performance of scientific applications.
Technology Integration	• Evaluate new technologies and develop solutions for at-scale issues.
Advanced Data & Workflows	• Support teams using big data analytics tools and optimize users' workflows.

CRA-W
Computing Research Association
Women

Bonus: Who works in HPC?

Experts in many areas are required

HPC Operations	• Deploy compute resources and supporting infrastructure.
User Assistance & Outreach	• Install software, develop tests, create documentation, troubleshoot issues, and train users.
Scientific Computing	• Analyze and optimize the performance of scientific applications.
Technology Integration	• Evaluate new technologies and develop solutions for at-scale issues.
Advanced Data & Workflows	• Support teams using big data analytics tools and optimize users' workflows.

CRA-W
Computing Research Association
Women

A day in the life of a HPC User Support Specialist...

Collaborative, diverse, and interesting!

- Every day is different!
- Every user report can bring:
  - a new problem to solve
  - a new scientific application to learn about
  - a new programming model to discover
  - an unexplained performance issue
  - a challenging workflow to adapt
  - and more...
- Approximately half to ¾ of the time spent in user reports.
- The rest of the time is spent on a wide range of projects:
  - improve the user experience, automate support tasks, understand performance bottlenecks, testing software and new systems, streamlining software installation, deploying latest programming environments, developing training tools, writing self-paced tutorials and documentation, and more!

CRA-W
Computing Research Association
Women

A day in the life of a HPC User Support Specialist...

Collaborative, diverse, and interesting!

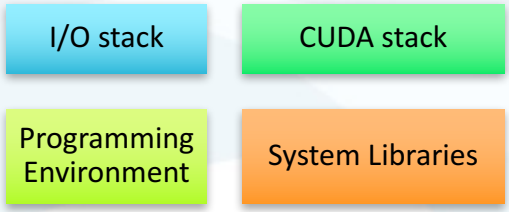
- Every day is different!
- Every user report can bring:
  - a new problem to solve
  - a new scientific application to learn about
  - a new programming model to discover
  - an unexplained performance issue
  - a challenging workflow to adapt
  - and more...
- Approximately half to ¾ of the time spent in user reports.
- The rest of the time is spent on a wide range of projects:
  - improve the user experience, automate support tasks, understand performance bottlenecks, testing software and new systems, streamlining software installation, deploying latest programming environments, developing training tools, writing self-paced tutorials and documentation, and more!

CRA-W
Computing Research Association
Women

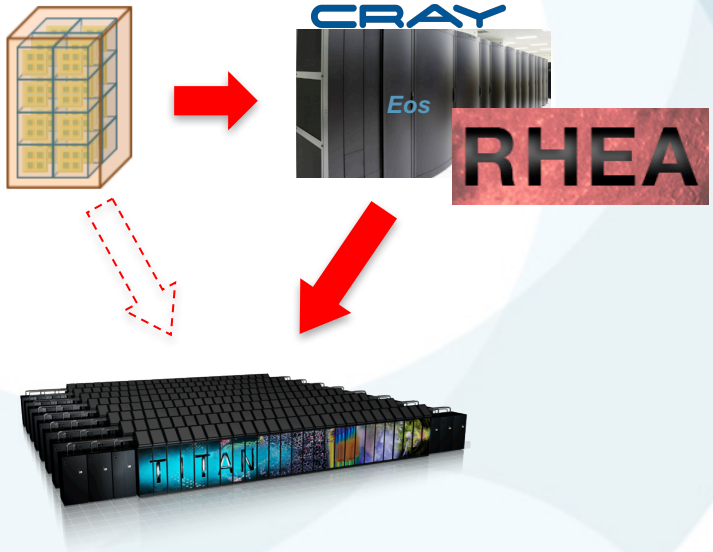
The many stages of testing

New Software - Operations Testing

- What are we testing?



- Plan and execute multi-stage tests:



New Systems - Acceptance Testing

- Improve the OLCF test harness
- Investigate ways to automate the testing process

- Port applications to use new architectures
- Design problems adequate for the next generation system
- Evaluate new compilers
- Expand the OLCF test harness to support new systems
- Design the Acceptance Test Plan for a new system

Accepting a large-scale system

Four Acceptance Stages

HW
- Hardware Acceptance Test: Complete hardware diagnostics.

FT
- Functionality Test: Demonstrate that basic hardware and software functionality meet essential requirements.

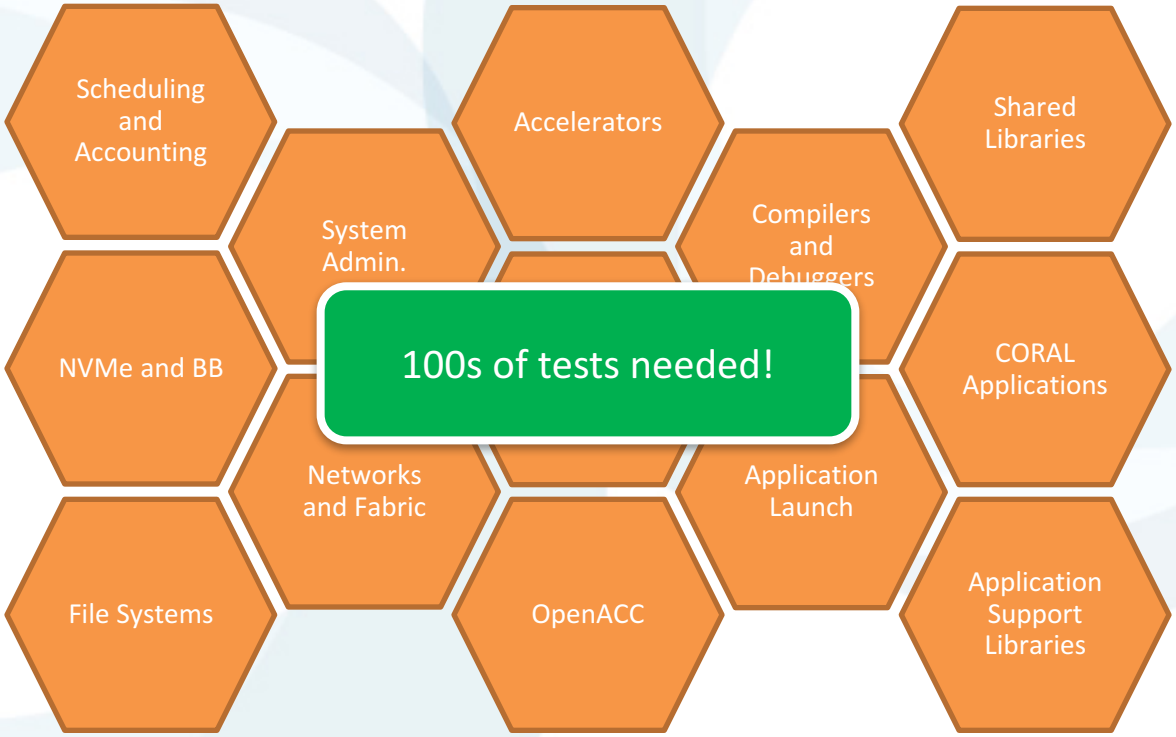
PT
- Performance Test: Demonstrate that the system hardware and software meet performance and scalability requirements of the suite of applications defined in the Agreement.

ST
- Stability Test: Demonstrate stability across a mix of simulated code development activity and production simulations.

CRA-W
Computing Research Association
Women

Accepting a large-scale system

Large-scale systems also have a complex ecosystem -- all of which must be tested



Accepting a large-scale system

The OLCF is deploying a next generation system

- New architecture presents new challenges and opportunities
  - Working with pre-release software for most part of the HPC stack
  - Applications must be redesigned to take full advantage of the fatter compute nodes
- Our team gets to install and operate systems that are not yet available to the general public
  - Exciting to work with brand new technologies
  - Challenging since the systems are still being developed
  - Get to influence the design of tools, programming environment, and documentation for a new architecture

Summary

- HPC drives innovation and has improved and accelerated the development of new technologies in many industries
- The OLCF provides world-leading computational and data resources and specialized services for the most computationally intensive problems
- Installing, deploying, accepting, and operating large-scale systems requires experts in different areas
  - Challenging and rewarding work!
- The OLCF staff conduct research in many areas both in computational science, as well as systems engineering and computer science.
- The OLCF will be home to Summit, the next generation supercomputer approximately 5-10x faster than Titan.

CRA-W
Computing Research Association
Women

Make the Most of an Undergraduate Research Experience

How/Where to find an REU?

CRA-W programs

- Collaborative Research Experience for Undergraduates (CREU):
  - Open to underrepresented groups in computing
  - Ask your professors to identify a topic
  - Ask your peers to form a team
  - Deadline usually in May
  - Academic year, with possibility of summer extension
- Distributed Research Experience for Undergraduates (DREU)
  - Open to underrepresented groups in computing
  - You are interested in a topic and want to find a mentor
  - Deadline usually in February
  - Summer long

CRA-W
Computing Research Association
Women

DREU vs CREU

	DREU	CREU
Time	Summer (10 weeks)	Academic year plus optional summer
Target group	Women, ethnic minorities, students with disabilities, and students from other underrepresented groups	Women, ethnic minorities, students with disabilities, and students from other underrepresented groups
Stipend per student	$7000 per summer; relocation travel assistance when appropriate.	$1,500 per semester and $4,000 during optional summer extension
Location	Mentor's institution	Student's institution
Team Work	Varies	Encouraged in CS and CE. Expected on multidisciplinary projects.
Interdisciplinary	Varies	Varies
Mentor	Faculty	Faculty at home institution of student. At least two faculty, from different disciplines, for multidisciplinary projects.
Deadline	February 15	May 18
Sponsor	CRA-W / NSF / Other Partners/Sponsors	CRA-W / NSF / Other Partners/Sponsors

How/Where to find an REU?

Other programs

- NSF REUs
  - Available in many topics
  - You apply to REU Sites. See: https://www.nsf.gov/crssprgm/reu/reu_search.jsp
- National Laboratories offer several undergraduate programs:
  - Science Undergraduate Laboratory Internships (SULI)
  - Community College Internship (CCI)
  - Higher Education Research Experience (HERE)
  - Laboratory Technology Program
  - Topic specific: IACMI and NESLS[1]
  - More information: https://www.orau.org/ornl/undergraduates/

[1] U.S. citizenship *not* required

I have an REU lined up, any advice?

- Don’t be afraid to ask questions
  - But do your homework first
- If you are interested in a topic:
  - Find the subject matter expert at the organization and ask for pointers to get you started
    - Be respectful and mindful of their time
- Ask for help early! You are there to learn -- it is ok to ask for help
- Make sure you get to know the organization, and your peers
  - Networking is very important!
- Follow-up and follow-through
  - Work hard to meet your deadlines
  - Be ready to work independently, but don’t forget even if you are there for a short time you are part of the team

CRA-W
Computing Research Association
Women

Acknowledgement

This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.

CRA-W
Computing Research Association
Women

THANK YOU!

QUESTIONS?