SDSC HPC User Training January 22, 2021

Mary Thomas
mpthomas@ucsd.edu
https://www.sdsc.edu/~mthomas/

Last updated: 1/22/21 @ 9:30am

Agenda

- Welcome
  - Introduction to Program
  - Introduce Mentors
- HPC User Training Orientation
- HPC Overview
- Registration for Expanse Account

HPC Training Mentors

- Mary Thomas <mpthomas@ucsd.edu
- Marty Kandes, Martin, mkandes@sdsc.edu
- Jeff Sale, jsale@sdsc.edu,
- Bob Sinkovits, sinkovit@sdsc.edu
- Mahi Tatineni, mahidhar@sdsc.edu
- Wong, Cindy, cwong@sdsc.edu>
- Susan Rathbun, susan@sdsc.edu
- Others TBD

Early parallel computers based on commodity hardware
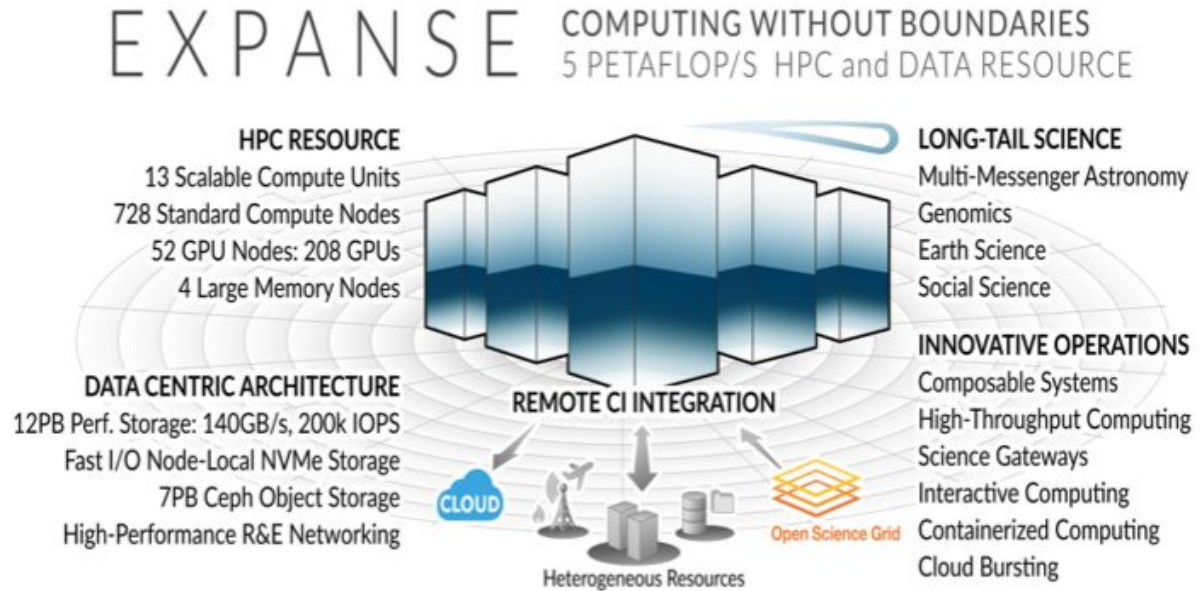
Stone Soupercomputer (2003):

Cheapest cost/flop=$0, ~20 MFlops

https://web.archive.org/web/20031121211117/http://stonesoup.esd.ornl.gov

What is High Performance Computing?

- *Aggregating computing power*
- *Delivers much higher performance than desktop computer or workstation*
- *Solve large problems in science, engineering, or business.*

E X P A N S E   COMPUTING WITHOUT BOUNDARIES
5 PETAFLOP/S HPC and DATA RESOURCE

HPC RESOURCE
13 Scalable Compute Units
728 Standard Compute Nodes
52 GPU Nodes: 208 GPUs
4 Large Memory Nodes

LONG-TAIL SCIENCE
Multi-Messenger Astronomy
Genomics
Earth Science
Social Science

DATA CENTRIC ARCHITECTURE
12PB Perf. Storage: 140GB/s, 200k IOPS
Fast I/O Node-Local NVMe Storage
7PB Ceph Object Storage
High-Performance R&E Networking

REMOTE CI INTEGRATION

CLOUD

Open Science Grid

Heterogeneous Resources

INNOVATIVE OPERATIONS
Composable Systems
High-Throughput Computing
Science Gateways
Interactive Computing
Containerized Computing
Cloud Bursting

For more details see the Expanse user guide @ https://www.sdsc.edu/support/user_guides/expanse.html
and the "Introduction to Expanse" webinar @ https://www.sdsc.edu/event_items/202006_Introduction_to_Expanse.html

HPC User Training

- SDSC supports training of its user community in all aspects of High-performance computing (HPC).
- The HPC User Training activity is designed to promote workforce development in technologies needed to work in HPC.
- Training is provided free to all users, including:
  - UCSD students (both graduate and undergraduate),
  - Post-docs
  - Researchers, faculty, staff, and collaborators on SDSC systems.
- To encourage student participation, some projects and activities are supported/mentored by the Supercomputing Club.

HPC Training: Scope

The scope of the training effort includes several skill development activities, including:

- Study HPC architectures (CPU, GPU), software, and admin skills.
- Learn the basics of parallel programming, including MPI using C, Fortran, and possibly other languages.
- Running HPC applications in the areas of performance characterization of the cluster, bioinformatics, numerical methods, password security, and other applications.
- Cloud Computing
- Visualization and analysis of big data sets.
- Intro to Machine Learning

HPC User Training 2021

Registration & Information Page:
https://na.eventscloud.com/website/21055/home/

- Series of 2 hour sessions:
  - SDSC Staff: Fridays from 1:00pm to 3:00pm
  - Session #1: 01/22/21 to 3/12/21 (8)
  - Session #2: 4/2/21 to 5/7/21 (6)
- Workload:
  - Regular attendance
  - Successfully complete assignments
  - 1-4 hrs/week

- *Evaluation metrics* will include:
  - Completion of tasks
  - Credit for experience (working/internships, classes taken)

- Participants who successfully complete the HPC Training program will:
  - Receive an SDSC Certificate of Completion in HPC Training.
  - Awarded account allocations on E X P A N S E for 1 year
- In addition, UCSD students who successfully complete the HPC
  - Receive an UCSD Co-curricular Record of credit in HPC Training.
  - Become eligible to apply to be on the SCC Core team.
  - Note: the final SCC team will be chosen from this group.

HPC User Training 2021: Schedule

Week	Date	Session #1: January 22, 2021 - March 12, 2021
1	1/22/2021	Program orientation, history, plan, registration process & accounts HPC overview & *Expanse* Architecture
2	1/29/2021	Parallel Queues & batch scripting; intro to applications (hello world, 2Djacobian)
3	2/5/2021	GPU Architectures, CUDA
4	2/12/2021	GPU Computing with CUDA Python
5	2/19/2021	CPU Computing: Introduction to OpenMP/Threads
6	2/26/2021	CPU Computing: Introduction to MPI
7	3/5/2021	CPU Profiling, monitoring: gprof, mpiP/comms, uProf
8	3/12/2021	GPU Profiling, monitoring: comms, NSight
		Session #2: April 2, 2021 - May 7, 2021
9	4/2/2021	Containers for CPU and GPU computing
10	4/9/2021	Cloud Computing: AWS
11	4/16/2021	Cloud Computing: Azure
12	4/23/2021	Machine Learning; Tensor Flow
13	4/30/2021	SCC21 Applications: HPL, HPCG
14	5/7/2021	Accessing and Visualizing large data sets on HPC systems Using Jupyter Notebooks (pandas, to see Lustre data sets, Matplotlib, Boqeh, SeaBorn)

https://na.eventscloud.com/website/21055/home/

SDSC HPC Students Program

HPC Student Activities

- Support the Supercomputing Club
- SDSC Internships (CCR)
- HPC Training (CCR)
- Building/Supporting HPC Internship knowledge database
- Raspberry PI Cluster build:
  - Spring'21
  - SDSC Supplies hardware, location, network, etc.
  - Supercomputing Club students build-out/admin

@Supercomputing:

- Student cluster competition (SCC) teams
  - SCC20: selected
  - SCC21: planning
- Student volunteers:
  - 4 selected for SC19
- All expenses paid by SDSC and Club sponsors!

Support the Supercomputing Club!

Supercomputing and the Student Cluster Competition

http://www.studentclustercompetition.us/

https://sc20.supercomputing.org/

SCC History/Background

- Began in 2007 to provide HPC experience to undergraduate and high school students.
- Students design and build small clusters, learn scientific computing, run applications.
- Compete in a non-stop, 48-hour challenge to complete a real-world scientific workload:
  - Operate and maintain the cluster
  - Run and test science applications
  - Reproducibility Challenge: students attempt to reproduce a science paper
  - Power outage challenge - restart system without warning

SCC: How to Compete

http://www.studentclustercompetition.us/

- Entering is competitive: we must write a formal application
- Max of 6 undergraduate students can be on the official team, at SC
  - All travel and expenses will be paid for (SCC, SDSC)
- Participation requires a *significant* level of commitment and hard work:
  - Learn HPC, applications, cluster hardware
  - Work during Fall to build competition cluster
  - Be able to travel to the meeting https://sc21.supercomputing.org/

SCC Team Preparation Process: 3 Steps

- Step 1: HPC Training
- Step 2: Application submission
- Step 3: If selected, start learning the programs

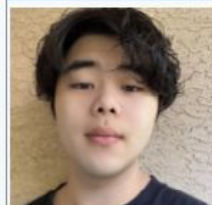# Team Superscalar Celebration!!!
4th overall/ 19 teams

Student Cluster Competition 2020

San Diego Supercomputer Center
The University of California at San Diego

About the Team

Jacob Xiaochen Li
Computer Science, Math

Skills: Cloud Administration, PID Control, RL.

Role: Team Leader, Team Lead for Reproducibility Challenge (MemXCT)

Arunav Gupta
Data Science and Math

Skills : Machine Learning, Cloud Computing, Statistics

Role : Team Lead for HPCG Benchmark, 2nd Lead for MemXCT

Zihao Kong
Computer Engineering

Skills: RTL design, FPGA, Embedded programming, CUDA programming, Cloud computing

Role: Team Lead for Gromacs
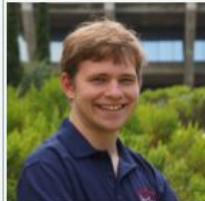
Max Apodaca
Computer Engineering

Skills : Cloud Computing, Containerization, Checkpoint/Restore

Role : Team Lead for IO500, CycleCloud Templates

Hongyi Pan
Cognitive Science

Skills : Machine learning, Natural Language Processing

Role : Team Lead for CESM

Hongyu Zou
Computer Science

Skills : Systems, Databases, and Networks

Role : Team Lead for HPL

Applications

High Performance Linpack

CESM

GROMACS

MemXCT

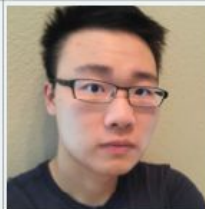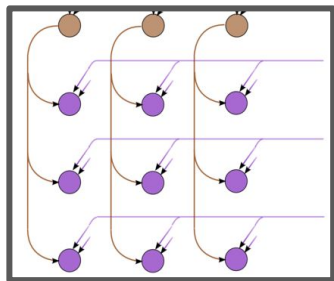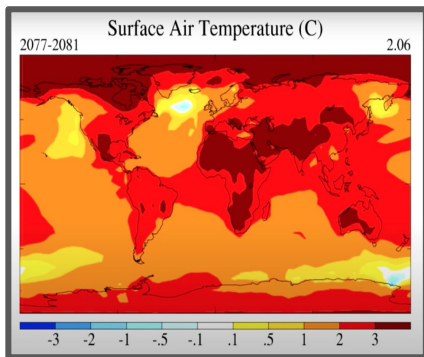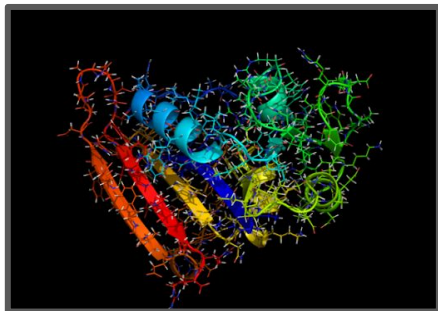# View of main Grafana dashboard

- Part of group of 13/19 teams
- Benchmarks:
  - HPCG (3rd): 4.056/15/8 TF/s
  - IO500 (6th): 16/144 "io500 pts"
  - HPL (11th): 53.9/300 TF/s
- First team at SDSC, UCSD, and

- Applications: (4)
  - MemXCT: completed, submitted paper
  - Mystery App (miniVite): completed
  - GROMACS: completed 60-70% (2 / 3 tasks)
  - CESM: completed 70% (3 / 5 tasks)

Learned about budgets, cloud costs and managing the costs

Team	CESM (5)		GROMACS (3)		MemXCT (2)		Mystery Application (6)		HPL	HPCG ↓	IO500
Tsinghua University, China		5		2		2		6	300 TFLOP/s	15.8 TFLOP/s	144
Georgia Institute of Technology		3		0		2		6	108 TFLOP/s	4.060 TFLOP/s	8.5
University of California, San Diego		3		2		2		6	53.9 TFLOP/s	4.056 TFLOP/s	16
Peking University		5		3		2		6	63.8 TFLOP/s	3.988 TFLOP/s	15
North Carolina State University		0		1		2		6	44.3 TFLOP/s	3.892 TFLOP/s	2.5
ETH Zurich (Team RACKlette)		3		1		2		6	129 TFLOP/s	3.786 TFLOP/s	3.6
University of Illinois		3		2		2		6	71.6 TFLOP/s	3.720 TFLOP/s	14
Friedrich-Alexander University Erlangen-Nuremberg		5		1		2		6	29.0 TFLOP/s	2.999 TFLOP/s	18
Wake Forest University		1		0		0		6	9.172 TFLOP/s	2.655 TFLOP/s	12
University of Warsaw		0		0		0		6	75.0 TFLOP/s	2.359 TFLOP/s	15
Nanyang Technological University, Singapore		1		0		2		6	105 TFLOP/s	2.040 TFLOP/s	25
University of Texas		3		3		2		6	13.1 TFLOP/s	1.454 TFLOP/s	0
Shanghai Jiao Tong University		0		3		2		6	19.9 TFLOP/s	1.154 TFLOP/s	12
Texas A&M University		5		0		1		6	113 TFLOP/s	1.043 TFLOP/s	19
Northeastern University		2		1		2		6	21.1 TFLOP/s	1.038 TFLOP/s	0.89
Southern University of Science and Technology		0		1		0		6	120 TFLOP/s	1.016 TFLOP/s	16
Massachusetts Institute of Technology (MIT)		0		0		0		6	64.9 TFLOP/s	593 GFLOP/s	0.70
ShanghaiTech University		2		2		2		6	14.4 TFLOP/s	0 GFLOP/s	20
Clemson University		0		0		2		6	32.6 TFLOP/s	0 GFLOP/s	0

Leaderboard

Activity: Raspberry PI Cluster

- Spring Quarter 2021
- Goal is to build a 16 node PI Cluster:
  - 16 PI-4 (4 GB) devices
  - Large display monitor
  - MPI, visualization, WFI, TCP/IP
- Club Project (ACM):
  - members will build, admin
- Cluster dedicated to student projects:
  - Simulations
  - Running jobs will be displayed in lobby
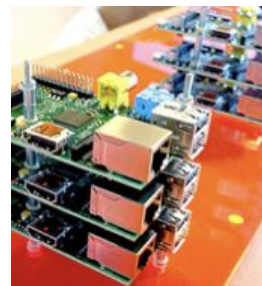
HPC Students Sponsors

Tasks for Week 1:

- Register for the Training: https://na.eventscloud.com/website/21055/home/

- Clone the GitHub repo:

- Obtain your Expanse account

  - https://portal.xsede.org → create portal account

  - Submit XSEDE PortalID to Form: https://tinyurl.com/hpc-user-training-signup

    - Use xsede portal username

- Once your account is created, make sure you can logon

- Complete Homework for Week1 (see README file):

  - https://github.com/sdsc-hpc-training-org/hpc-training-2021