High-Performance Computing (HPC) at NYU

Eric R. Peskin, Ph.D.
Manager, High Performance Computing
What is HPC?

• Aggregating computing power to achieve higher performance than that of a desktop or workstation. [http://insidehpc.com]
• Refers to the hardware, software, algorithms, methods and people required to use that power effectively.
• Hardware could be monolithic supercomputers or clusters of servers.
• Typically a shared resource.
Why HPC?

- To solve problems that would be too large to handle without it.
  - Need more time.
  - Need more memory.

- To take advantage of parallelism available on HPC resource:
  - Across multiple *users*.
  - Across multiple *jobs* from one user.
  - Across multiple *nodes* used by one job.
  - Across multiple *CPU cores* within one node.
What is available at NYU?

• HPC compute cluster *Mercer* – parallel and heavy compute workloads
• New HPC compute cluster *Prince* – (replaces Mercer)
• Hadoop cluster *Dumbo* – data-parallel workloads
• Mongo DB cluster *Hades* – non-relational database
• Multi-tier storage system
Mercer: HPC Cluster

- Capacity
  - ~400 nodes
  - ~6,100 cores
  - ~35 TB RAM total
- Parallel and heavy compute loads
- Many HPC and scientific software libraries
- Torque/Moab batch system
- InfiniBand interconnect

NOTE: New HPC Cluster Prince replaces Mercer soon.

Mercer HPC Cluster

- ~400 nodes
- ~6100 cores
- ~35 TB memory

Parallel and heavy compute workloads
- Many HPC and scientific software and libraries
- Torque/Moab batch system
- InfiniBand interconnect

Compute node types

- Ivybridge nodes (newer, 20-core)
- Westmere nodes (older, 12-core)
- Large memory nodes
- GPU nodes
- High-memory nodes

Compute Nodes (for all computation)

Login nodes (editing scripts, submitting jobs)

Per-node RAM disk

Per-node Local disk

Temporary filesystems
  - Local to each node
  - Low latency but limited size and throughput
Infiniband interconnect between storage and compute

Hades
DB Cluster
6 nodes with
12 cores, 256GB RAM, 32TB
raw storage (RAID, 16TB usable)
Total usable ~90TB

~1.5PB Fast parallel filesystem (Lustre)

$SCRATCH /scratch/$USER

MongoDB (MySQL later)

Long term data storage (ZFS)

$WORK /work/$USER
$HOME /home/$USER
$ARCHIVE /archive/$USER not mounted on compute nodes

remote copy sent update
local copy on tape

Faster ← HPC and Big Data Storage → Longer-term
Who Can Use NYU HPC?

• Open to NYU
  • Faculty
  • Staff
  • Faculty-sponsored students

• May be used for
  • Class Instruction
  • Research

• Basic use is free, but you can
  • Contribute hardware for additional priority (next slide...)
  • Pay for extra disk quota
Stakeholders

- Departments and research groups can buy additional hardware.
- NYU HPC can host and manage the resource in return for these groups making unused cycles available to community.
- These groups are our *HPC Stakeholders*.
- Stakeholders get priority access.
- Current stakeholders: CGSB, CNS, CDS, Kussell Lab, and CAOS.
- Talk with us *before* you purchase your cluster.
Tandon School of Engineering Usage

• 209 users from Tandon

• Applications used:
  • COMSOL
  • Ansys
  • MATLAB
  • Mathematica
  • Gaussian
  • ...
Example Project: Drug Design for treatment of heart attack and stroke

- Drs. David Rooklin and Yingkai Zhang, Department of Chemistry
- Hundreds of molecular simulations
- Each used up to 64 processors and ran sometimes for weeks
- Previously uncharacterized catalytic calcium-binding site:

Revelation of a Catalytic Calcium-Binding Site
Elucidates Unusual Metal Dependence of a Human Apyrase
David W. Rooklin, Min Lu, and Yingkai Zhang
Journal of the American Chemical Society 2012 134 (37), 15595-15603
DOI: 10.1021/ja307267y
IB Example

• See animation
Getting Started and Getting Information

• HPC Wiki:  [https://wikis.nyu.edu/display/NYUHPC](https://wikis.nyu.edu/display/NYUHPC)
  • Getting and Renewing Access
  • Tutorials, FAQs and how to get help

• Ask us:  [hpc@nyu.edu](mailto:hpc@nyu.edu)

• Attend tutorials / workshops
NYU HPC Team

• Efstratios Efstathiadis, Director, Research Technology Services
• Eric Peskin, Manager, High Performance Computing
• Frances Bauer
• Wensheng Deng
• Shenglong Wang
• Bill Bridges
• Santhosh Konda
• Tatiana Polunina