USC Research Computing

New Faculty Orientation Presentation
August 2019

Erin Shaw, Presenter

HPC Research Computing Facilitation

USC Center for High-Performance Computing

Information Technology Services (ITS) Research Computing Initiative

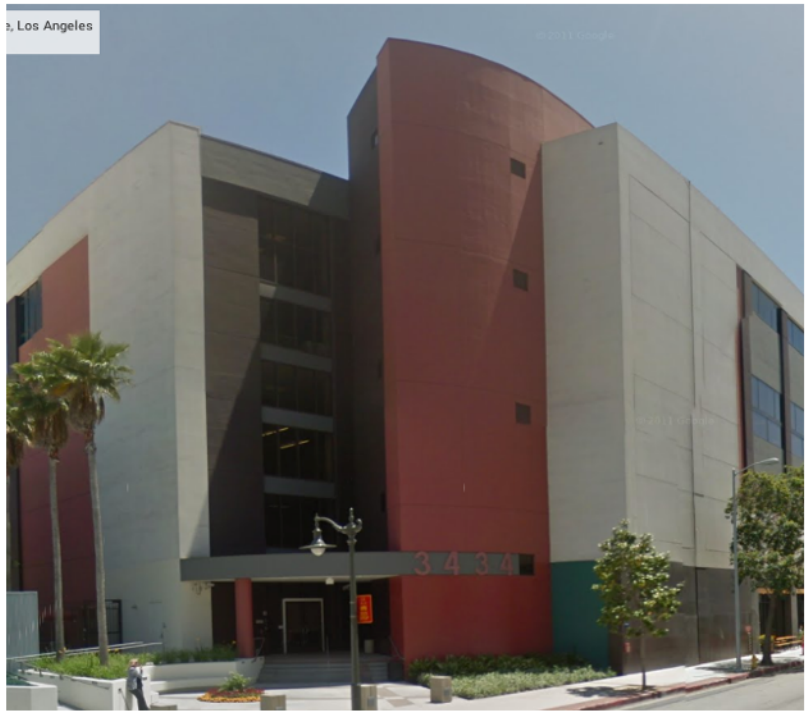
- Software
  - Campus-wide software
- Storage
  - Data storage options
- High-Performance Computing
  - Resources and purchase options
  - Research support and facilitation

ITS Research Computing Management

Douglas Shook
USC CIO

Randolph Hall
USC VP of Research
*Faculty Executive Director, HPC**

Byoung-Do (BD) Kim
HPC Director
Research Professor, DSO Marshall School of Business



ITS Data Center

*The HPC Faculty Advisory committee advises the CIO about the faculty's research needs related to the university's HPC resources.

Software

Comprehensive list at *software.usc.edu*

Statistical Computing Software

- Campus-wide licenses
  - SAS, SPSS, and JMP Pro
    - *Available for download at low cost/no cost*
- Campus-wide on HPC
  - SAS and STATA
    - *Installed on HPC and available for free with an HPC account*

- Open source on HPC
  - R Statistical computing software
  - Python scientific computing packages
  - HPC versions are pre-installed and compiled with parallel computing supports if available
    - *AVX, GPU, BLAS*
  - Researchers may install their own packages

Technical Computing Software

- Technical software available for download
  - MATLAB
  - Mathematica
- GIS and remote sensing
  - ArcGIS
  - TerrSet
  - Data/Geoportal
    - *No cost and low cost options*
    - *See spatial.usc.edu*

- Software available on HPC
  - MATLAB and Mathematica
  - Compilers: Gnu, Intel, PGI (Nvidia)
  - Numerical Libraries: MPICH, OpenMPI, CUDA, FFTW, PETSc, Gurobi
  - Molecular Simulation: NAMD, Gromacs, Amber, LAAMPS
  - Quantum Chemistry: Gaussian, Qchem
  - Data: Globus, hdf5

Bioinformatics Software

(prod.libraries.usc.edu/bioinformatics/#software)

- Comprehensive functional analysis and data mining
  - Partek Genomics Suite
  - Partek Flow
  - Galaxy
  - Vector NTI Advance
  - CLC Genomics Workbench
- High-throughput data and DNA/protein sequence analysis
  - Ingenuity Pathway Analysis (IPA)
  - Ingenuity Variant Analysis (IVA)
  - BIOBASE
  - BaseSpace Correlation Engine
  - Genevestigator
  - Qinsight

Research and Research Management Software

- Citation management
  - RefWorks, Mendeley, EndNote, and Zotero
    - *libraries.usc.edu/research/citation-management*
- Project management
  - RedCap
- Data collection
  - Qualtrics (survey management)
- Data analysis
  - ATLAS.ti
  - NVivo

UNIX Computing Utilities

- UNIX computing utilities
  - Secure shell
    - *X-Win32 (Preconfigured with connections to HPC head nodes and other USC networks)*
    - *software.usc.edu/x-win32*
  - Secure file transfer
    - *scp, rsync, rclone*
    - *puTTY, Filezilla, WinSCP (third party)*
- Virtual private networking
  - vpn.usc.edu
  - Cisco AnyConnect VPN
    - *itservices.usc.edu/vpn*
- Duo two-factor authentication
  - *itservices.usc.edu/duo*

Storage

Information at *itservices.usc.edu/storage* or *repository.usc.edu*

Cloud Data Storage

- USC Google Drive
  - Unlimited capacity (50MB to 5TB upload limit, depending on file type)
- USC OneDrive for Business
  - 5 TB capacity (10GB upload limit)
  - For use by USC Office 365 account holders
- USC Dropbox (faculty only)
  - 50 GB capacity
- USC Google Drive, OneDrive, Dropbox accounts are approved for storing some restricted information, including data covered under HIPAA, FERPA

USC Data Storage

- USC Drive
  - Free 50 GB capacity, high-speed storage
    - *Some restricted data OK, (e.g., FERPA)*
    - *Additional storage can be purchased*
- USC Digital Repository
    - *repository.usc.edu*
  - Fee-based storage, 40 petabytes of total capacity
    - *High-speed, high-performance disk storage, and low-demand tape storage*
    - *Available for Windows, Mac, and Linux (including HPC)*
    - *Highly secure and encrypted cloud storage (Approved for HIPAA-related information)*

HPC Data Storage

- Group project storage
  - Project investigators can apply for multiple group projects, up to 10 TB (max) of *shared* storage. Files backed up daily.
- Temporary access to a shared 328 terabyte **/staging** directory
  - There is a 10 TB max quota *per researcher*, for staging files for jobs.
  - No backup! Files cleared after six weeks of inactivity, when at capacity, and during semi-annual downtimes.
- Temporary access to directories on local nodes
  - **$TMPDIR** and **$SCRATCHDIR** while running job on compute node(s)
  - No backup! Files cleared at end of each job.
- HPC Secure Data Accounts are approved for storing secure data
  - Must apply specifically for HSDA, and provide IRB approval

HPC Condo Data Storage

HPC PIs with active accounts can purchase storage in its condominium-style compute environment

- HPC obtains approved vendor quotes for storage arrays based on user needs.
- All storage requires a mandatory minimum 3-year agreement with 3 years of vendor support.
- The condo storage is only available on the HPC cluster; the file system is not shared outside of HPC .
- HPC provides limited backup of data (up to 100 terabytes) at no cost to the condo owner.

High-Performance Computing

hpcc.usc.edu

High-Performance Computing (HPC)

HPC advances USC's mission by providing the infrastructure and support necessary for research computing

- Many of HPC's resources are available at no charge to USC faculty, staff, graduate students, and affiliate (iVIP) collaborators.
- As a part of USC's Information Technology Services (ITS) division, HPC is housed within the ITS data center and is monitored around-the-clock by ITS staff.

World-class super-computing center

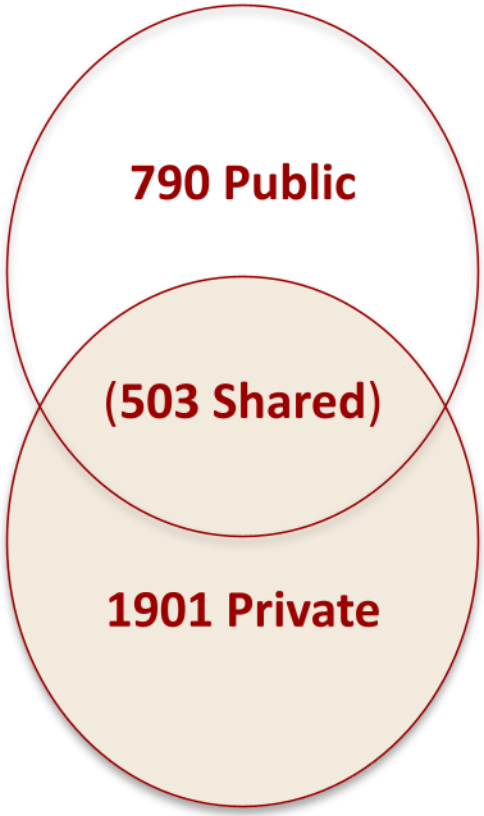
- HPC performs LINPACK benchmarking and publishes results to Top500.

HPC Resources

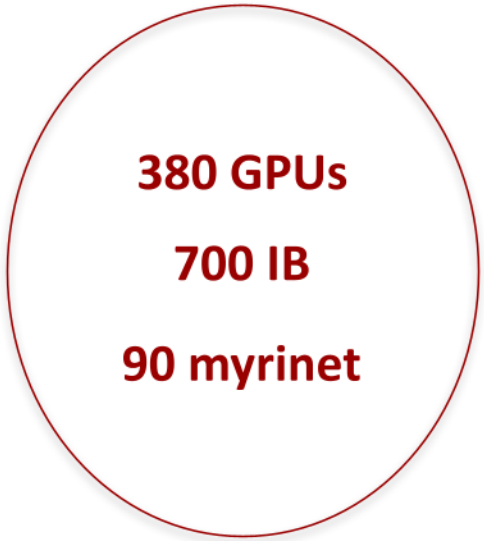
- Compute nodes (~2700 running CentOS 7)
- Disk arrays (networked storage, file servers)
- Two low-latency networks
  - Infiniband 56-gigabit (~700 public nodes)
  - Myricom 10-gigabit (~90 public, condo)
- Regularly upgraded
  - Every fall and often also in spring, HPC installs new nodes, upgrades and patches.
  - HPC resources are unavailable during these upgrades, which last for approximately a week.

HPC Cluster at a Glance

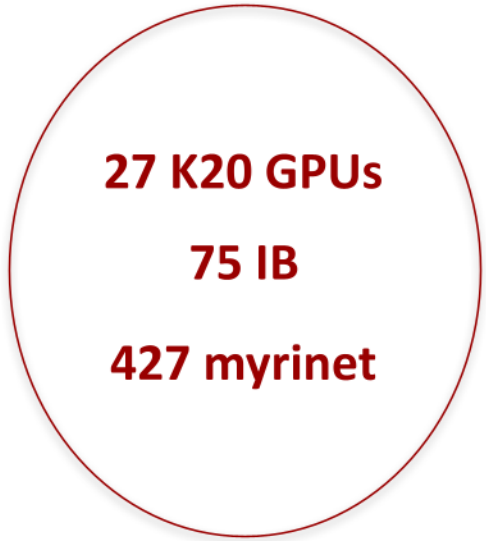
2691 Total Nodes



790 Public Nodes



503 Shared Nodes



Highlights:

- 4 x 40-core, 1 TB memory nodes
- Nvidia Tesla V100 (32) and P100 (44) GPUs
- Dedicated quick test and long job nodes

Condo Compute Nodes

HPC PIs with active accounts can purchase compute nodes in its condominium-style compute environment

- Condo nodes can be purchased through HPC at any time, as long as nodes are available in our existing pool of nodes.
  - *Only HPC-purchased nodes are available for condo use*
- Condo nodes are priced at the amount that HPC purchased them, including taxes.
  - *This is a one-time fee*
  - *Condo fees do not include infrastructure, which is covered by HPC*
- Contact **hpc@usc.edu** for current configuration and pricing information.

HPC Research Support and Facilitation

See *hpcc.usc.edu/research/grant-app/* **and**

hpcc.usc.edu/getting-help

HPC Research Support

- Proposal Support
  - Help build startup packages, provide quotes for condo purchases, and write letters of support for grant submissions
  - Provide a description of ACI resources and support for proposals
- Collaboration
  - Collaborate on HPC projects with researchers on all USC campuses
- Research
  - 2014-18 HPC participated in six-site NSF-sponsored grant
    - *Advancing scientific discovery through a national network of Advanced Cyberinfrastructure Research and Education Facilitators (ACI-REFs)*

HPC Facilitation

Outreach and Education

- Monthly facilitation newsletter
- Classroom and lab presentations
- Workshops
  - *Intro Linux, Intro & Advanced HPC,*
  - *Installing Software, Parallel MATLAB, R, Python, Pegasus WMS*
- Guest training events
  - *Mathworks, NVIDIA*
- ITS data center tours

Direct Assistance

- Help tickets (hpc@usc.edu)
- Office hours
  - *UPC: Tuesdays@2:30 (LVL 3M)*
  - *HSC: NML and Soto (x2 month)*
- Consultations
  - *Lab consultations*
  - *Individual meetings*
  - *Remote conferencing*

HPC Plans

- In fall 2018, HPC added 154 Dell nodes to cluster
  - 120 dual 12-core, 96G (83 condo)
  - 34 dual 12-core V100 GPU, 192G memory (2 condo)
- To request additional nodes for spring 2020
  - Contact HPC by mid-December if interested
  - Must be bought by PI purchase order in January for spring upgrade, or order migrates to fall upgrade
- May still have nodes available from spring 2019 upgrade
  - If interested in a node immediately, contact us
- Why condo?
  - Condo node rulesets can be specified by owners
  - Shared/public nodes have “fairshare” rulesets

Thank You!

For more information about HPC, please see **hpcc.usc.edu**

or email **hpc@usc.edu**.