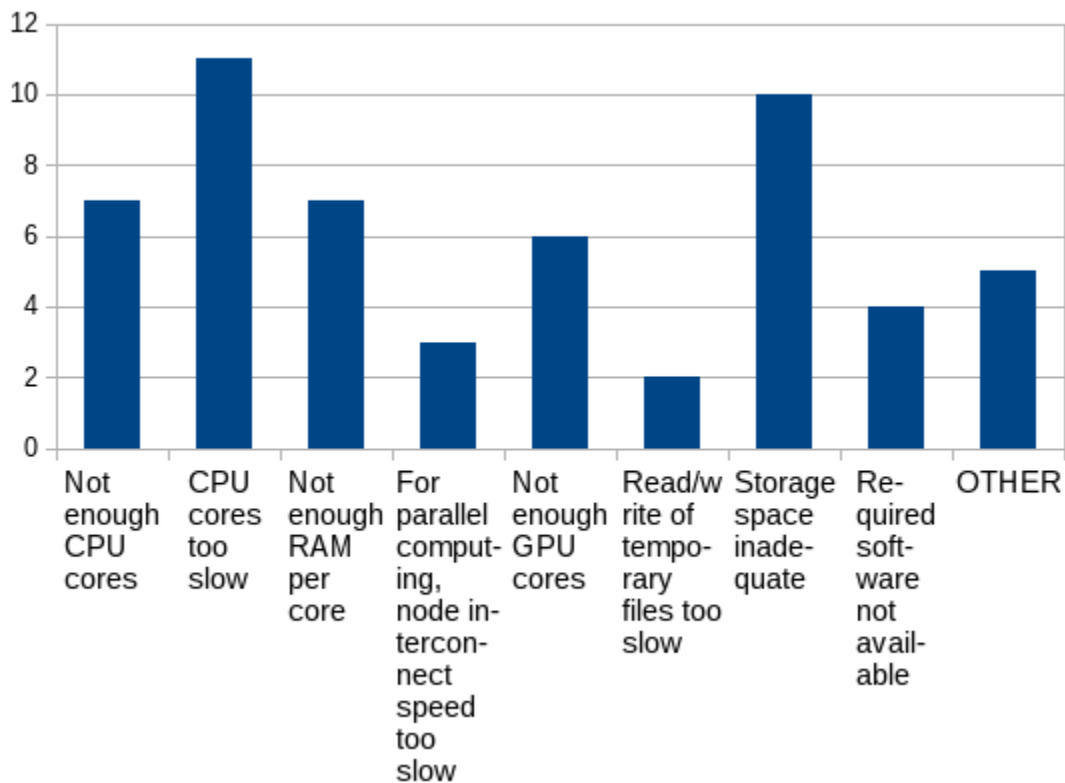


HPC & BD Survey: Results for participants who use resources outside of NJIT

Questions concerning uses of HP/BD resources outside of NJIT were contained in Section 1 of the survey, HPC Hardware Resources. 53 participants (91%) took this section. Of these, 21 reported using outside resources. 11 (52%) of these were faculty. These participants, then, had a proportionately larger faculty representation than the survey as a whole, where faculty comprised 40%.

Participants indicated their reasons (as many as they wished) for using outside resources by choosing among provided reasons (Not enough CPU cores; CPU cores too slow; Not enough RAM, For parallel computing, node interconnect speed too slow; Not enough GPU cores; Read/write of temporary files too slow; Storage space inadequate; Required software not available; OTHER). They then described the outside resources they used, and were asked to provide additional comments if they wished.

The number of participants citing each reason for using outside resources by the 21 participants who reported using outside resources is shown in the graph below.



The specified non-NJIT resources, along with the reasons for their use, are shown in the subsequent table. Each table row represents responses from a single participant. Hence, resources listed more than once were specified by more than one participant.

REASONS	Not enough CPU cores	CPU cores too slow	Not enough RAM per core	For parallel computing, node interconnect speed too slow	Not enough GPU cores	Read/write of temporary files too slow	Storage space inadequate	Required software not available	OTHER
OUTSIDE RESOURCE									
Xcede		X							
Xsede		X							
Xsede supplementary computation	X	X	X	X	X				
Compute Canada, Xsede		X		X					
Amazon Webservice. Google cloud platform		X	X		X		X		
AWS & GCP			X	X		X	X	X	
GPU station with 4 TitanXp GPUs, purchased by professor					X				
lab	X	X							
NFS computer, DOE & Navy computers	X		X				X	X	
NASA Pleides supercomputer							X		
Resources at NYU									X
TACC					X		X		X
CIPRES portal							X		X
Workstation and other universities with better support	X	X	X		X		X	X	
Rutgers Conley3 and Amarel							X		

8-core 4.0Ghz, high performance GPU 2GB workstation		X							
Stanford HPC system, Barcelona Supercomputing Center									X
UIUC cluster machine	X	X	X				X		X
HPC from UPenn							X	X	
<i>unspecified</i>	X	X			X	X			
<i>Unspecified, provided by collaborators</i>	X	X	X						

The table below lists the written descriptions for *OTHER* and for *Additional comments*.

OUTSIDE RESOURCE	Reasons: OTHER	ADDITIONAL COMMENTS
Xsede		Because I'm facing a problem in which my jobs got killed without any reason but that's sure it's due to a problem in cluster and when I contact the ARCS, they said they are not able to identify the problem.
Xsede supplementary computation		NJIT HPC computational resources are not adequate for the amount of research we perform in our group.
AWS & GCP		data transfer between node and disk to memory and temp space for bd task is not enough. for data science, we don't have the latest version of lib and also I couldn't find cuDNN lib for tensorflow to run the deep learning into GPU. I believe, we may have to improve hpc/bd resources for deep learning for both parallel and serial in terms of lib, driver, development and testing environment
lab		Easier access. Response time.
NFS computer, DOE & Navy computers	some software is just too expensive for the university to consider for one or two users only - so use if can find on nsf or dod computer x	
NASA Pleides supercomputer		For Kong head node, the 1 GB quota is too small

Resources at NYU	More experiments can be run :)	Team is made up of NYU and NJIT students/faculty. Therefore, it is easy for us to use both clusters to run our experiments.
TACC	Inadequate support	Inadequate support and not to up-to-date documentation
Stanford HPC system, Barcelona Supercomputing Center	I am still completing some work on those machines.	I will be using those only until I am done with some old computation. I will switch to NJIT resources very soon, though.
UIUC cluster machine	GLIBC is too old on stheno	For Kong, I was attempting to install miniconda and work on specific project in Python. Unfortunately, the storage for each user is just 5GB and it is totally not enough for a student user who has need in large computation. For stheno, the GLIBC version is way too old (2.5) and it should have up-to-date version for users.

Additional observations: Participants who indicated use of outside resources versus participants who did not

Requests for new processors

24% of participants who used outside resources expressed an interest in new processors, compared with 50% of participants who did not use outside resources. (The candidate new processors were Intel Core i7 or i9; Google Tensor processing unit; Intel Nervana Neural Network Processor; Intel Xeon Phi; and AMD Epyc.) While both groups favored Intel Core i7 or i9, the outside-resources group also favored Google Tensor processing unit and Intel Xeon Phi compared with the non-outside-resources group, while the non-outside-resources group favored AMD Epyc compared with the outside-resources group. (Note that a participant could list any number of proposed resources. Hence, even though a larger variety of proposed processors were chosen by the outside-resources group, a greater number of non-outside-resources groups members indicated a wish for any new processors at all.)

Requests for currently unavailable software

Participants were asked to specify currently unavailable software that they wanted, specify whether the software had associated costs, and indicate the level of anticipated use of the software in their research and teaching.

In the table below, software specified by participants who use outside resources are shown in the gray rows. The non-highlighted rows list software specified by participants who do not use outside resources.

Note that some items (e.g., Tensorflow) were requested by multiple participants. Each individual request is listed, as participants varied in their proposed usage.

NAME of REQUESTED SOFTWARE	Associated costs	Research use	Teaching use
Too many options for compiler lead to problems; simplify the list	no	medium	low

cclib python library; can be installed from anaconda by "conda install -c omnia cclib"	no	high	high
Quantum Espresso; Open source alternative to Gaussian 16	no	high	high
GAMESS; Open source alternative for Gaussian	no	high	high
Polyrate; Software for rate constant calculations; has interfaces for Gaussian, NWChem, GAMESS	no	high	high
Open babel; like cclib it's a useful tool to have	no	high	high
Tensorflow	no	high	low
Scikit Learn (python library)	no	high	low
Keras	no	medium	low
Nektar++	no	high	no
cuDNN. Deep learning for GPU	no	high	high
Stata	yes	high	no
SAS	yes	high	no
Matlab 2017b	yes	high	high
Matlab Computer Vision Toolbox	yes	high	high
Updated version of Lammmps; Working GPU acceleration in Lammmps, Lammmps with most of the packages preinstalled	no	high	low
Updated Anaconda Python 2 (and maybe 3) packages/libraries	no	high	high
Tensorflow	no	medium	medium
Tensorflow	no	high	medium
tensorflow	no	medium	high
Keras	no	high	low
keras	no	high	medium
git latest version	no	high	high

http://liulab.dfci.harvard.edu/MACS/Download.html (MACS)	no	high	no
BETA: http://cistrome.org/BETA/	no	medium	no
STAR: https://github.com/alexdobin/STAR	no	medium	no
pytorch	no	high	medium