Results of survey: HPC Usage 2015

INTRODUCTION

The survey was introduced with the following:

This survey solicits input from researchers on their use of nJIT's HPC resources.

The purpose of this survey is to gather input from researchers on their usage of nJIT's HPC resources*.

The information gathered from this survey will be used for tuning the HPC environment, and for estimating hardware needs.

*IST-managed HPC resources are:

Clusters: kong.njit.edu, stheno.njit.edu
Shared memory: cnrdp.arcs.njit.edu, gorgon.njit.edu, phi.njit.edu, HTCondor (HTCondor exploits idle cycles in the GITC 2315C and GITC 2400 Linux classrooms)

Part 1:

Survey participants—their status, how long they have used IST-managed HPC resources, what they use it for

PARTICIPANTS

49 invitees—30 faculty/staff and 19 students—completed the survey.

Faculty/staff

25 (83%) use IST-managed HPC resources, either individually or as head of a research group.

Of these 25, 15 head a group that uses IST-managed HPC resources; 10 are not group heads.

Of the 15 group heads, 7 run their own jobs using IST-managed HPC resources; 8 head groups whose members use IST-managed HPC resources, but they themselves do not directly use the resources.
Note that this latter group was not asked specific questions other than evaluative commentary; presumably their group members who directly use the resources provided data as survey participants.

Of the 5 faculty/staff (7% of faculty/staff participants) who do *NOT* use IST-managed HPC resources, 1 (.03% of faculty/staff participants) uses outside resources (i.e., HPC resources not managed by HPC).
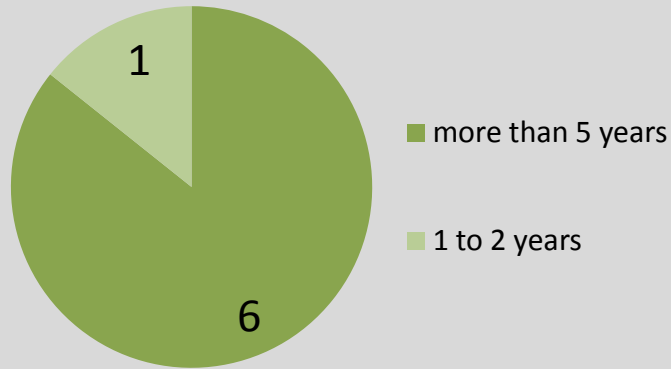
Students

18 (95%) use IST-managed HPC resources; the remaining student does not use any HPC resources, outside or otherwise.
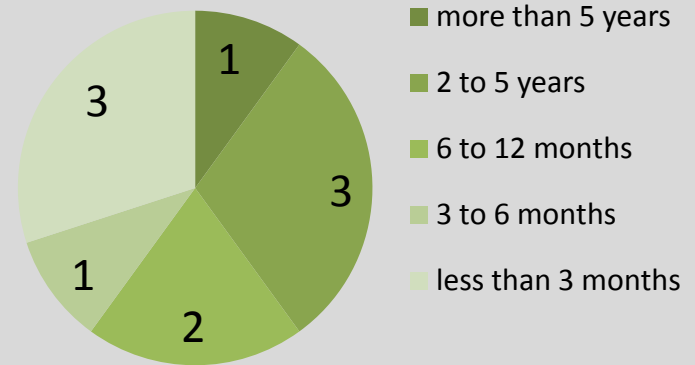
(1 of the 18 erroneously reported not using IST-managed HPC resources, and hence provided limited data in the survey.)
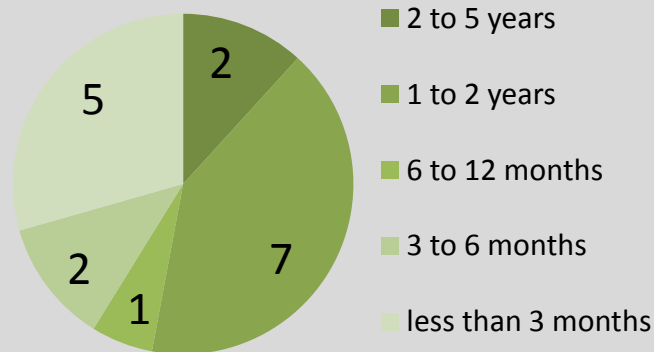
LENGTH of USAGE of IST-managed HPC resources

Faculty/staff group heads who run their own jobs

- more than 5 years
- 1 to 2 years

Faculty/staff who are not group heads

- more than 5 years
- 2 to 5 years
- 6 to 12 months
- 3 to 6 months
- less than 3 months

Students

- 2 to 5 years
- 1 to 2 years
- 6 to 12 months
- 3 to 6 months
- less than 3 months

RESEARCH USE of IST-managed HPC resources —FACULTY/STAFF group heads who run their own jobs

General Classification of Computations	Specific Classification of Computations
cfd	simulation and modeling of nano-scale and complex fluid flows
condensed matter physics	fortran or c programs written by me or group members
Computational chemistry	Molecular structures, energies and reaction transition states, vibration frequencies, electronic properties,
molecular bio-physics and computational molecular biology	We simulate protein conformations to understand plaque formation in the brain of Alzheimer's patients.
CFD	non newtonian multiphase flows
Materials Research	DFT and Electron Scattering
bioinfomatics	data analysis

RESEARCH USE of IST-managed HPC resources
—FACULTY/STAFF who do not head groups

General Classification of Computations	Specific Classification of Computations
Transportation data analysis	Travel time, vehicle time travelled, vehicle count, peak hour traffic count etc
steganalysis and image forensics	extract huge features from huge number of images and conduct classification with proper classifier
computational fluid dynamics	thin film flows
computational pde	reaction-diffusion coupled with odes
statistical analysis	bootstrap sampling, anomaly detection
Monte Carlo simulations, optimization	Monte Carlo simulations, optimization
neural networks, genetic algorithms	neural network analysis of image/video data
Electromagnetism, Wave propagation	big data
Bioinformatics	whole genome alignments, genome evolution
computational fluid dynamics	fortran code

RESEARCH USE of IST-managed HPC resources —STUDENTS *page 1*

General Classification of Computations	Specific Classification of Computations
statistical analysis, cuda	brain computer interfaces, neural networks
bioinfomatics	Metagenomics classification
Computational physics and chemistry	Density functional theory- molecular dynamics- montecarlo simulation
granular science	tappes system simulation
software verification, static analysis	static analysis of source code in order to derive artifacts such as termination, correctness
Machine Learning Research	Testing out my new algorithm in machine learning field
computational biophysics	matlab and C++ codes
computational chemistry	Gaussian 09
Bio informatics, text analysis	Predicting gender and extracting health topics using social media text data
computational fluid dynamics	----

RESEARCH USE of IST-managed HPC resources —STUDENTS *page 2*

….continued

General Classification of Computations	Specific Classification of Computations
biophysics,	molecular dynamics simulation
thermoproperties	running fast
statistical analysis	multiprocessing and big data analysis
statistical analysis	waiting so long time
feature classification	none
image forensics	training classifiers and also testing for image forensics
Computational Chemistry	Thermo chemistry and molecular structure

Part 2:

Use of resources

Q: For which type(s) of computations do you use IST-managed HPC resources?

Answer Choices: serial computations; parallel computations using MPI (message passing interface);
parallel computations *not* using MPI, i.e., threads or MP (a parallel environment for use on shared memory machines)

Faculty/staff group heads who run their own jobs (n=7)

Serial	Parallel MPI	Parallel Threads/ MP
yes	yes	yes
yes	no	no
yes	no	no
no	yes	no
yes	no	yes
yes	yes	no
no	no	yes

71% use serial processing; 43% use Parallel MPI; 43% use Threads/MP

Faculty/staff who are not group heads (n=10)

Serial	Parallel MPI	Parallel Threads/ MP
yes	no	no
yes	yes	no
yes	yes	no
yes	no	no
yes	yes	no
yes	yes	no
no	yes	no
yes	yes	no
yes	no	no
yes	no	no

90% use serial processing; 60% use Parallel MPI; 0% use Threads/MP

Students (n=17)

Serial	Parallel MPI	Parallel Threads/ MP
yes	no	yes
yes	no	no
yes	yes	yes
yes	yes	no
no	no	yes
yes	no	yes
yes	no	yes
yes	no	no
yes	no	no
yes	yes	no
no	yes	no
no	yes	no
no	no	yes
yes	yes	no
no	yes	yes
yes	no	yes
yes	no	no

71% use serial processing; 41% use Parallel MPI; 47% use Threads/MP

Q for **Serial** processing: On which machines do you run serial computations?

Answer Choices: kong.njit.edu (cluster); stheno.njit.edu (cluster); cnrdp.arcs.njit.edu (shared memory); gorgon.njit.edu (shared memory); phi.njit.edu (shared memory); HTCondor

Kong and stheno users were asked to indicate the typical RAM, in GB, that they requested in their submit scripts.

Cnrdp.arcs, gorgon, and phi users were asked to assess the adequacy of available cores (64) and RAM (112 GB) (for cnrdp, 64 cores and 112 GB of RAM; for gorgon, 32 cores and 64 GB of RAM; for phi, 4 cores and 38 GB RAM). They were also asked to estimate of the minimum amount of RAM and cores needed in IST-managed shared memory HPC resources to satisfactorily handle your computations for their research.

HTCondor users were asked to rate their experience and provide optional comments.

Faculty/staff group heads who run their own jobs (n=5)

Machine	Number of users
kong	5
stheno	1
cnrdp.arcs	0
gorgon	0
phi	0
HTCondor	0

100% use kong; 20% use stheno; 0% use cnrdo, gorgon, phi, or HTCondor

1 kong user requests 32 GB of RAM; 4 don't know. The 1 stheno user doesn't know.

Faculty/staff who are not group heads (n=9)

Machine	Number of users
kong	5
stheno	4
cnrdp.arcs	1
gorgon	1
phi	0
HTCondor	1

56% use kong; 44% use stheno; 11% use cnrdp; 11% use gorgon; 0% use phi; 11% use HTCondor

1 kong user requests 128 GB of RAM; 2 don't submit requests; 2 don't know.
2 stheno users don't submit requests; 2 don't know.
The cnrdp and gorgon users rated available cores and RAM as adequate.
The HTCondor user rated usage as *poor*;
comment: "95% of the time, my jobs never get to finish then is being transfer to different machine"

Students (n=12)

Machine	Number of users
kong	10
stheno	2
cnrdp.arcs	0
gorgon	0
phi	0
HTCondor	1

83% use kong; 17% use stheno; 0% use cnrdp; 0% use gorgon; 0% use phi; .08% use HTCondor

2 kong users request 64 GB of RAM; 2 request 32 GB of RAM; 5 don't request; 1 doesn't know.
Neither stheno user requests RAM.
The HTCondor user rated usage as *good*.

Q for **Parallel MPI** processing: On which machines do you run parallel computations using MPI?

Answer Choices: kong.njit.edu (cluster); stheno.njit.edu (cluster)

Users were asked to indicate the typical and largest number of cores that they specify in their submit scripts for kong or stheno.

Faculty/staff group heads who run their own jobs (n=3)

Machine	Number of users
kong	3
stheno	1

100% use kong; 33% use stheno

All 3 kong users typically and maximally request 11 – 50 cores.
The 1 stheno user typically and maximally requests 11 – 50 cores.

Faculty/staff who are not group heads (n=6)

Machine	Number of users
kong	4
stheno	4

67% use kong; 67% use stheno

3 kong users typically and maximally request more than 100 cores; 1 typically requests less than 10 and maximally requests 11 – 50.
2 stheno users typically request less than 10 cores, 1 typically requests 11 – 50, 1 typically requests more than 100. I user maximally requests less than 10, 2 maximally requests 11 – 50, 1 maximally requests more than 100.

Students (n=7)

Machine	Number of users
kong	6
stheno	1

86% use kong; 14% use stheno

1 kong user typically requests more than 100 cores; 2 typically request 11 – 50; 3 typically request less than 10. 1 maximally requests more than 100, 1 maximally requests 50 – 100, 4 maximally request 11 – 50.
The stheno user typically and maximally requests less than 10 cores.

Q for **Parallel Non-MPI** processing: Which type of non-MPI parallel computations do you run?

Answer Choices: threads; MP; don't know; other

Thread users were asked which clusters, kong or stheno, they used, and the typical and largest number of threads specified in their submit scripts; MP users were asked which shared memory machines, cnrdp.arcs, gorgon and phi they used, to indicate the adequacy of available cores and RAM, and to estimate the minimum amount of RAM and cores needed in IST-managed shared memory HPC resources to satisfactorily handle their computation; users who chose *other* were asked to specify.

Faculty/staff group heads who run their own jobs (n=3)

Computation type	Number of users
Threads	1
MP	1
Don't know	1
other	1

1 user who runs both threads and MP uses gorgon, indicates that available RAM and cores are adequate, and estimates that 64 GB of RAM and 32 cores are needed.

1 user uses other: OPENMP

Faculty/staff who are not group heads

None ran non-MPI parallel computations.

Students (n=8)

Computation type	Number of users
Threads	5
MP	1
Don't know	1
other	2

3 users run threads on kong; 1 typically submits 6 – 15, and maximally 16 – 20; 1 typically submits less than 5, and maximally 6 – 15; 1 doesn't know. 1 user runs threads on stheno, and typically and maximally submits 6 – 15.

The MP user did not indicate a shared memory maching.

The 2 *other* users listed CUDA and GPU.

Q for **GPU use:** Do you use GPUs (graphical processing units) on kong? Do you use GPUs (graphical processing units) on stheno?   Please describe the adequacy of these GPUs for the computations you have run on them.

The kong GPUs are: Model : Nvidia K20X CPUs : 4 Cores per CPU : 2688 RAM per GPU, GB : 6   Please describe the adequacy of these GPUs for the computations you have run on them. The stheno GPUs are: Model : Nvidia K20 CPUs : 4 Cores per CPU : 2496 RAM per GPU, GB : 5   Please describe the adequacy of these GPUs for the computations you have run on them.

Faculty/staff group heads who run their own jobs

None reported using GPU's.

Faculty/staff who are not group heads

None reported using GPU's.

Students

4 students reported using GPU's on kong; none reported using GPU's on stheno.

Comments:

I have had no problems using them on my computations.

After latest upgrade of the GPUs I can tell that it is sufficient. some jobs take up to 20 hours to finish.

Very good.

good

Files used for general computations previously specified by participants (see pgs 5-8)

Q: What is the size of the largest files generated, in GB? Where are these files written to?
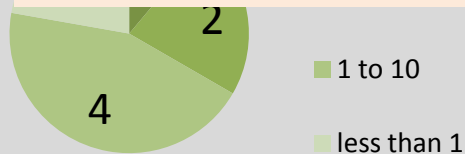Q: Do you use input files in your computations? If yes, What is the size of the largest input files, in GB? Where are the input files used for your computations stored?

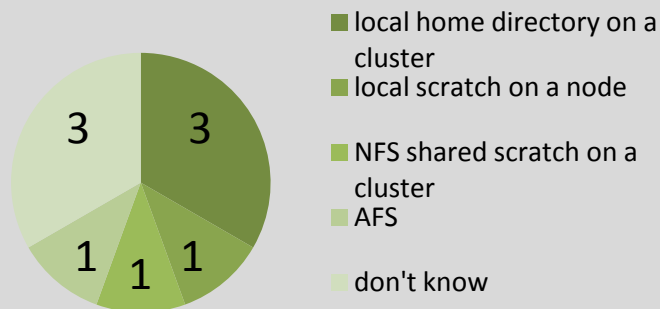Answer choices for file sizes: less than 1, 1 – 10, 11 – 100, 101 – 500, greater than 500, don't know
Answer choices for location of files: local scratch on a node (i.e., /scratch), NFS shared scratch on a cluster (e.g., /nscratch on kong), local home directory on a cluster, AFS, don't know

Faculty/staff who head groups and run their own jobs
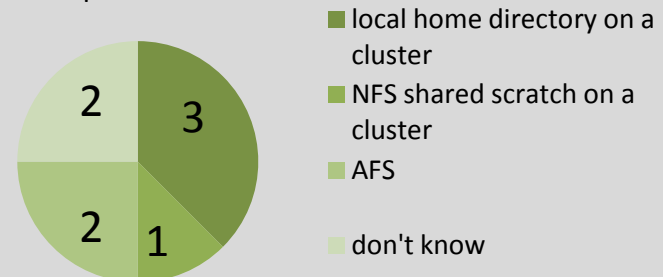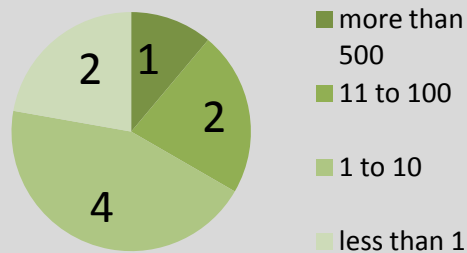
Size of largest files generated in GB
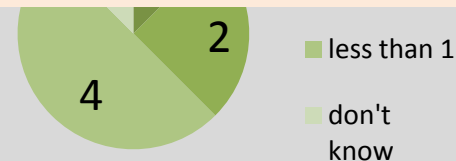
Size of largest files input files

Placeholder page for missing faculty data

2

■ 1 to 10

4

■ less than 1

2

■ less than 1

4

■ don't know

Where generated files are written to

3   3

1   1   1

■ local home directory on a cluster
■ local scratch on a node

■ NFS shared scratch on a cluster
■ AFS

■ don't know

Where input files are stored

2   3

2   1

■ local home directory on a cluster

■ NFS shared scratch on a cluster

■ AFS

■ don't know

Files used for general computations previously specified by participants (see pgs 5-8)

Q: What is the size of the largest files generated, in GB? Where are these files written to?
Q: Do you use input files in your computations? If yes, What is the size of the largest input files, in GB? Where are the input files used for your computations stored?

Answer choices for file sizes: less than 1, 1 – 10, 11 – 100, 101 – 500, greater than 500, don't know
Answer choices for location of files: local scratch on a node (i.e., /scratch), NFS shared scratch on a cluster (e.g., /nscratch on kong), local home directory on a cluster, AFS, don't know
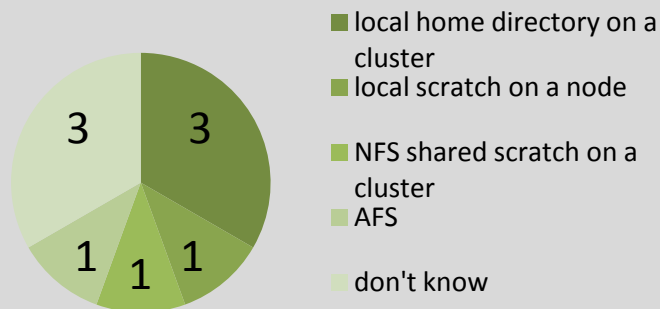
Faculty/staff who are not group heads

Size of largest files generated in GB
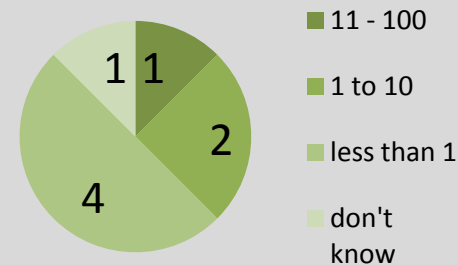Note: 9 out of 10 participants responded

- more than 500
- 11 to 100
- 1 to 10
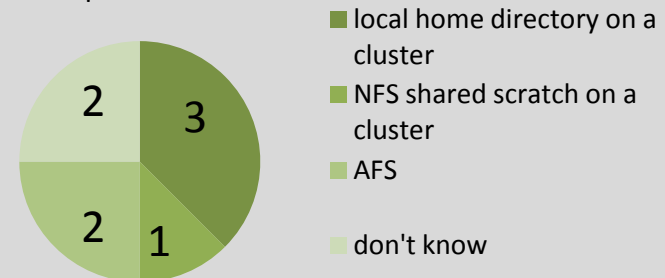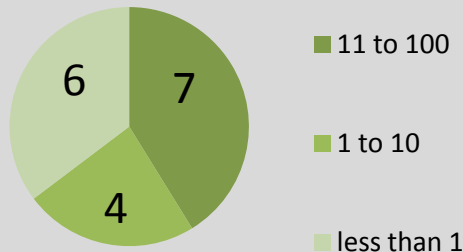- less than 1

Size of largest files input files
Note: 8 out of 9 respondents reported using input files

- 11 - 100
- 1 to 10
- less than 1
- don't know

Where generated files are written to

- local home directory on a cluster
- local scratch on a node
- NFS shared scratch on a cluster
- AFS
- don't know

Where input files are stored

- local home directory on a cluster
- NFS shared scratch on a cluster
- AFS
- don't know

Files used for general computations previously specified by participants (see pgs 5-8)

Q: What is the size of the largest files generated, in GB? Where are these files written to?
Q: Do you use input files in your computations? If yes, What is the size of the largest input files, in GB? Where are the input files used for your computations stored?

Answer choices for file sizes: less than 1, 1 – 10, 11 – 100, 101 – 500, greater than 500, don't know
Answer choices for location of files: local scratch on a node (i.e., /scratch), NFS shared scratch on a cluster (e.g., /nscratch on kong), local home directory on a cluster, AFS, don't know
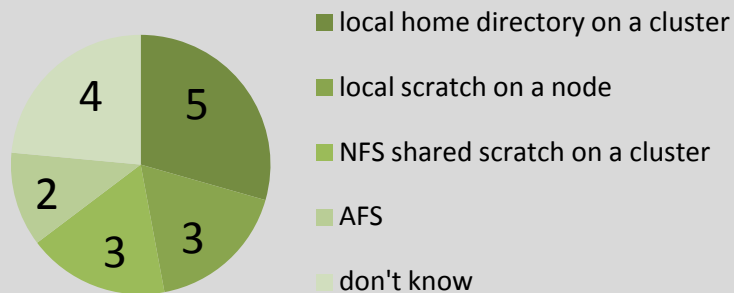
## Students

Size of largest files generated in GB
N=17

- 11 to 100
- 1 to 10
- less than 1

Where generated files are written to

- local home directory on a cluster
- local scratch on a node
- NFS shared scratch on a cluster
- AFS
- don't know

Size of largest files input files
Note: 16 out of 17 respondents reported using input files

- 101 to 500
- 11 to 100
- 1 to 10
- less than 1

Where input files are stored

- local home directory on a cluster
- local scratch on a node
- NFS shared scratch on a cluster
- AFS
- don't know
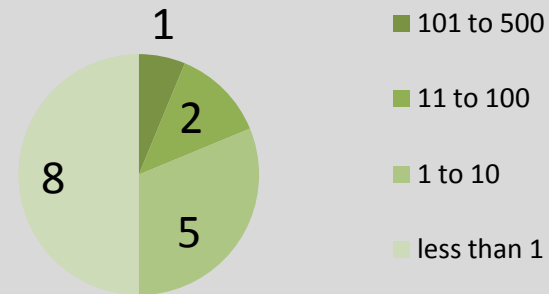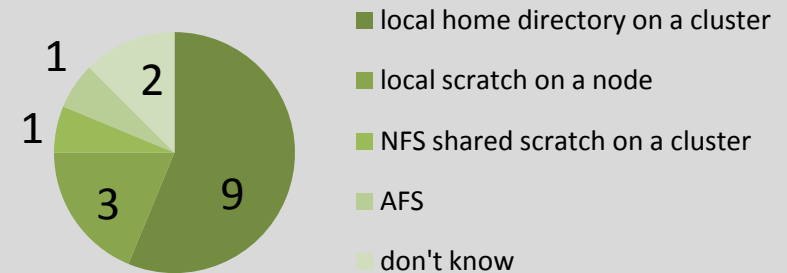
Part 3:

Comments: IST and Outside Resources

Assessment of research groups' use of IST-managed HPC resources —by FACULTY/STAFF who head research groups

Q: What is your estimate of your research group's satisfaction with their use of IST-managed HPC resources?
Answer choices: high, medium, low, not able to estimate at this time

Q (optional): Please provide any comments you may have on your research group's use of IST-managed HPC resources.

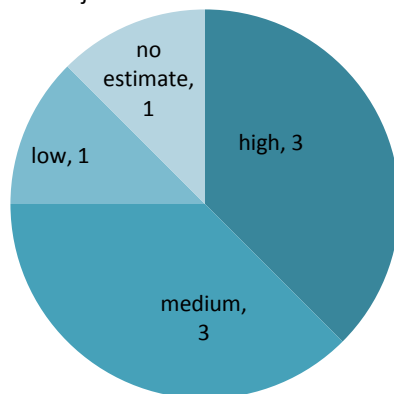Research Group Heads who run their own jobs n=7
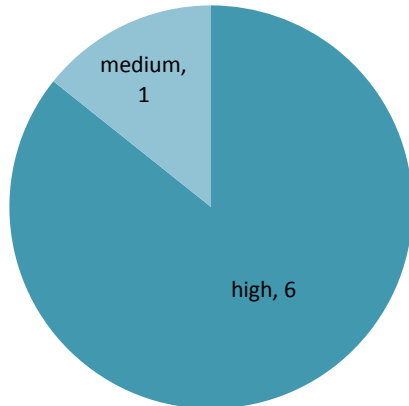
Research Group Heads who *do not* run their own jobs n=8

RESEARCH GROUP HEADS WHO RUN THEIR OWN JOBS
Comments from faculty/staff who estimated group satisfaction HIGH
•I and my research group members are very thankful to have IST help and interaction for access to Kong and NJIT computer resources.
•We are happy with the new HPC resources (recent expansion of kong). This is satisfactory to run a lot of our work. However, it is still desirable to have a few fast (new) computers.

RESEARCH GROUP HEADS WHO *DO NOT* RUN THEIR OWN JOBS
Comments from faculty/staff who estimated group satisfaction HIGH
•We use various codes, mostly written in Fortran, but we also use Fluent from Ansys
•[*IST staff member*] has done a very nice job for me.
•We want to write parallel database programs using MapReduce. Please let me know when it is available in July.

Comments from faculty/staff who estimated group satisfaction MEDIUM
•We used kong a few years ago, and it was too slow for our simulation needs. Since then, we didn't use any HPC resource, but we would like to do it (I know kong has been upgraded).
•In general computing environment is reasonable - we have issues sometimes regarding distribution of resources: how many jobs can a user run, number of processors in different queues etc.
•Would it be possible to add/buy additional software platforms such as Materials Studio, COMSOL?

General Comments on IST-managed HPC resources: FACULTY/STAFF

Q: (optional) Please provide any comments you have on your use of IST-managed HPC infrastructure.

Note: Research group heads who do not run their own jobs were not asked this question; their comments were solicited via group satisfaction assessments (see pg. x), and direct users in their research groups provided their own assessments

RESEARCH GROUP HEADS WHO RUN THEIR OWN JOBS

•xxx

FACULTY/STAFF who are NOT GROUP HEADS
•It is really faster than before. And it save a lot time on our experiences.

General Comments on IST-managed HPC resources: STUDENTS

Q: (optional) Please provide any comments you have on your use of IST-managed HPC infrastructure.

STUDENTS

•The hard drive space is very small. An extra 5 GB could go a long way

•The more computational power and Ram we have, the more productive we become as students. I hope NJIT could invest more on its computational resources. Thanks

•As a student that got access to Kong, I had a hard time installing my own software in the cluster. I got much help from the mailing list for some specific problems. The problem I faced in the beginning was that I had to wait for long for my calculations to be submitted (days sometimes). Only after a colleague enabled me access to his own cores that I was able to overcome this problem... I think NJIT students and researchers don't have a real access to NJIT HPC ...

•the deleting and refreshing file is not convenience on the user derectory Maybe add the capacity of the Kong; improve the Kong stability

•Need more storage spaces for the input data. 5GB is far from enough.

•Any possible way for storage the running files during Kong maintainance

## Assessment of research groups' use of **outside resources** by FACULTY/STAFF who head research groups

Q: Do either you, or members of your research group, use HPC resources other than the IST-managed resources? If your group uses outside resources, What is your estimate of your research group's satisfaction with their use of outside resources?
Answer choices: high, medium, low, not able to estimate at this time

Q: Please list the HPC resources that you yourself, i.e., not just members of your group, use other than those managed by IST. Please list the general classification of the computations for which you use those resources. Please list your reasons for using those resources.
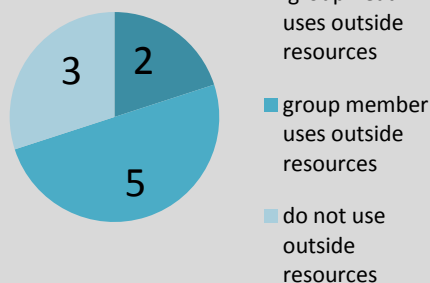
Research Group Heads who run their own jobs using IST-managed HPC resources:
5 out of 7 also use, or their groups use, outside resources. The 2 whose members use outside resources rate satisfaction as *HIGH.*

- group head uses outside resources — 4
- group member uses outside resources — 2
- do not use outside resources — 2
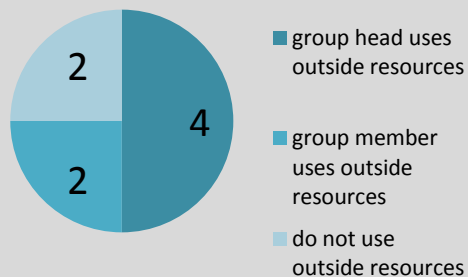
Research Group Heads who DO NOT run their own jobs using IST-managed HPC resources:
5 out of 8 also use, or their groups use, outside resources. The 5 whose members use outside resources estimate satisfaction as *MEDIUM.*

- group head uses outside resources — 2
- group member uses outside resources — 5
- do not use outside resources — 3

Outside Resources	Computations	Reasons
Research Group Heads who run their own jobs using IST-managed resources		
exede	cfd	libraries
DOD computer systems; Spitis and Garnet Clusters	use for serial and for parallel processing on very large jobs - 16 processors up to 150 hours cpu time per job.	computational chemistry larger molecules with very large numbers of basis sets -- (orbitals)
We have 32 Core Machine in our Lab; Use the NERSC machines--130,000 total cores; BNL IBM Blue Gene--33,000 total cores	Atomic and Electronic Structure Calculations	The machine have up-to-date fast CPUs and broad range of compilers: Portland Fortran and C Intel Fortran and C
computation and storage	bioinformatics	faster with larger memory and storage
Research Group Heads who DO NOT run their own jobs using IST-managed resources		
We bought our own cluster and built a small cloud on top of it (it's hosted in the machine room on the 4th floor of GITC). We also use Amazon cloud resources.	Simulations (large scale mobile networks), emulations at scale(network/system protocols).	Need fast response time or enough machines to test protocols at scale.
Lawrence Livermore National Lab	Very large linux servers	Software that only runs there, and will only ever run there.

Individual researchers' **outside resources:** FACULTY/STAFF who do not head research groups and STUDENTS

Q: Do you use HPC resources other than the IST-managed resources?

Q: Please list the HPC resources that you use other than those managed by IST. Please list the general classification of the computations for which you use those resources. Please list your reasons for using those resources.

None of the 15 faculty/staff who use IST-managed resources and are not heads of research groups use HPC from outside resources.

1 faculty/staff member of the 5 who do NOT use IST-managed resources uses an outside resource.

1* of the 17 students who reported using IST-managed resources listed specific use of outside resources. A second** reported using outside resources did not know what they were.
 1***of the 2 students who reported not using IST-managed resources listed use of resources that are actually IST-managed. The remaining students who does not use IST-managed resources does not use outside resources; hence, all student participants who use HPC resources use those managed by IST, even if not exclusively.

	Outside Resources	Computations	Reasons
	Faculty/Staff who do not use IST-managed resources		
	DELL servers purchased for our research	Guarantee dedicated uses for performance data.	Guarantee dedicated uses for performance data.
	Students		
*	blas	cuda	course
**	I don't know about this	I have no idea	I don't know what resources, I just run and get the data from there
***	AFS; Kong	machine learning for big data	class projects