March 2013

Results of

High Performance Computing and Big Data: Utilization and Needs Survey

This survey collected data on researchers' current and anticipated high performance computing (HPC)* and big data (BD) utilization and needs in hardware, software, staff support, and system usability.

About the participants:

55 out of 262 invitees participated in the survey. Of these, 40 reported using HPC or data in their research or teaching. These results report data from these 40 participants.

Of the 40 participants, 37 (92.5%) are faculty; 2 are post-docs, and 1 is a grad student. All except the post-docs (i.e., 38 participants) teach; 18 of these use HPC in their courses. 1 participant dropped out midway through the survey.

14 (35%) responded as individuals, 18 (45%) responded for a research group of 2 – 5 people, and 8 (20%) responded for a research group of more than 5 people.

**i.e., multiprocessor, memory-intensive, high throughput*

Importance of various resources

Q: How important to your research are computing resources provided by the following sources?

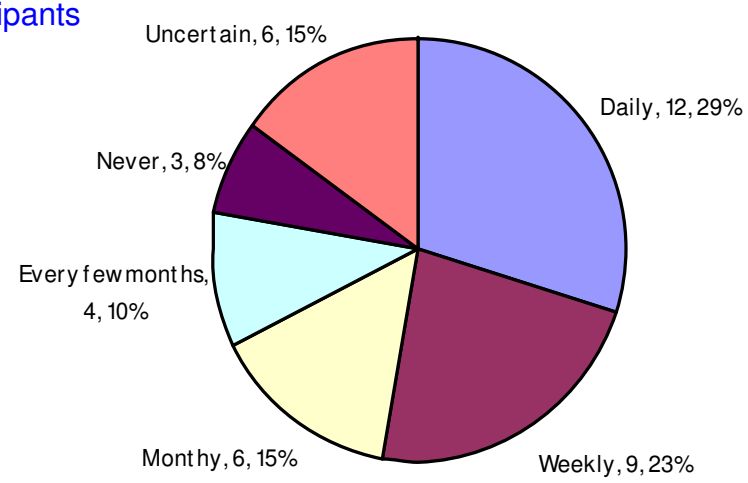
40 participants

	Critical	Very Important	Important	Not Important	Uncertain
Your research group	24	7	2	2	5
Your dept. or college	17	9	5	5	4
NJIT IST UCS	21	8	6	2	3
External resources	9	8	10	9	4

Submitting and running jobs

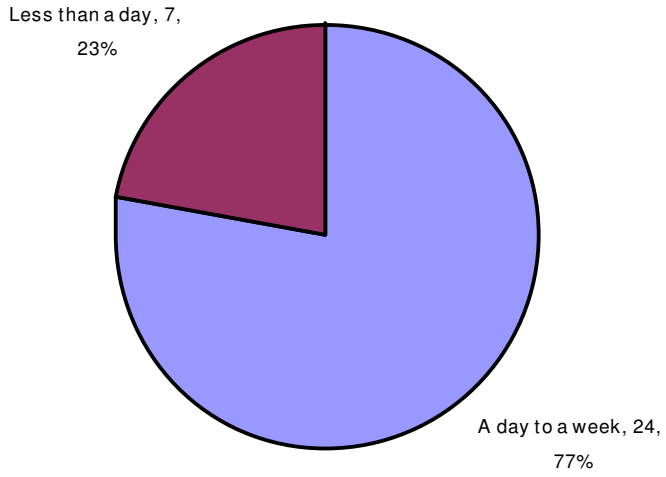
Q: How often do you or your research group expect to submit jobs to IST HPC in the next 3 years?

40 participants



Q: How long do your jobs typically run?

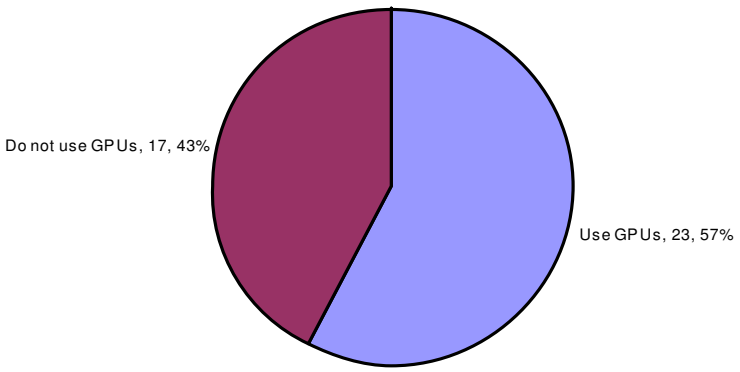
31 participants



GPU use

Q: Do you use, or plan to use, GPUs (Graphical Processing Units) in your research and/or teaching?

40 participants



23 participants

22 use GPUs for computation, 6 for graphics.
Thus, 1 uses GPUs for graphics exclusively,
5 use GPUs for both computation and graphics

Computing applications

Q: Which best describe your research computing applications?

40 participants

	Number YES	% YES
Parallel MPI (multiple processors, message-passing interface)	24	60%
Serial (single processor)	23	57.5%
Open MP (parallel programming on shared memory systems)	15	37.5%
Uncertain	3	7.5%
Other	5	12.5%

5 participants

Explanations of *Other*:

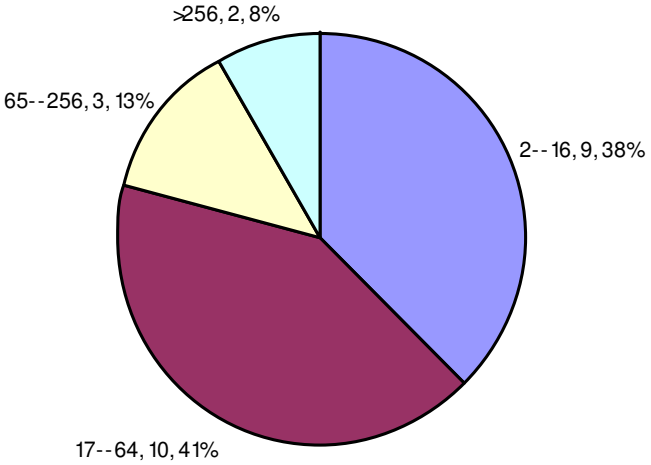
GPU
BSPlib,Pthreads
Cloud
Hadoop
pc

Parallel MPI

Data from those who described their research computing applications as *parallel MPI*

Q: During the next 3 years, what is the maximum number of processors that you anticipate your parallel jobs could effectively use?

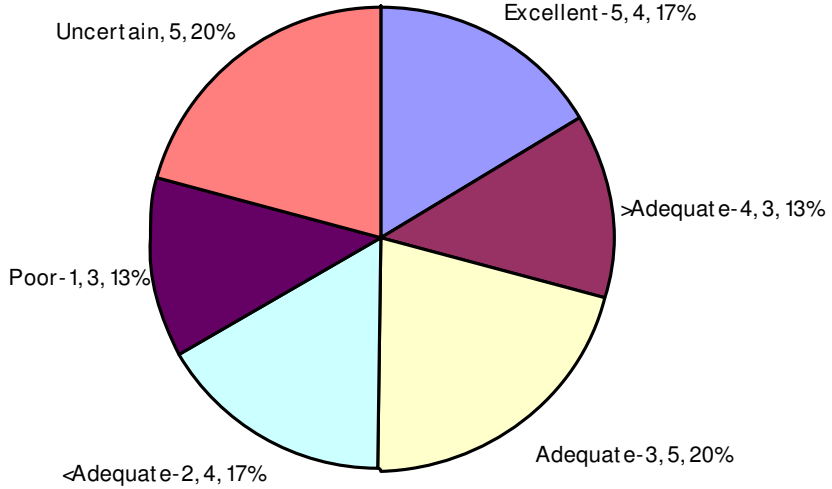
24 participants



CURRENT hardware resources

Q: Please rate the IST HPC hardware resources support for your current parallel computing needs.

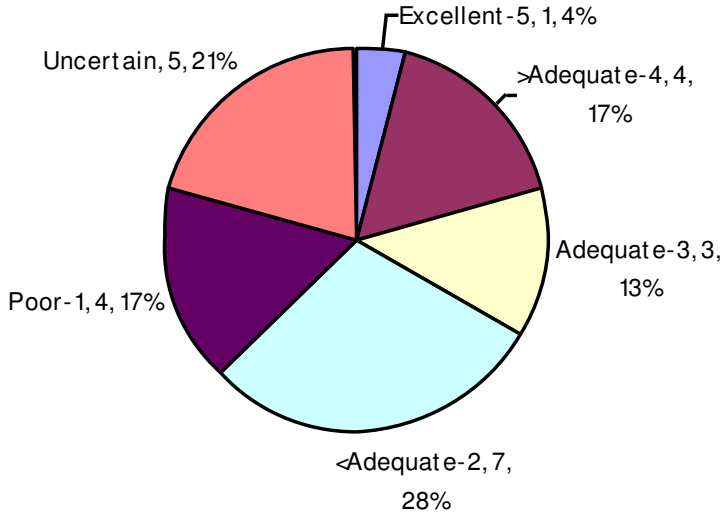
24 participants



ANTICIPATED hardware resources

Q: Please rate the IST HPC hardware resources support for your anticipated parallel computing needs.

24 participants



Parallel MPI *continued*

CURRENT hardware resources *continued*

Data from those who rated the current IST HPC hardware resources support as less than adequate or poor

Q: Select the CURRENT level of deficiency for each category.

7 participants

	No deficiency (5)	(4)	Moderate deficiency (3)	(2)	Large deficiency (1)	Uncertain
Num cores	0	0	3	2	1	1
RAM per core	1	0	0	3	1	2
Processor speed	0	0	2	1	3	1
Node interconnect speed	0	0	1	1	2	3
Disk space	1	0	2	2	1	1
Other	0	0	1	0	1	5

Explanations of *Other*:

Large deficiency: GPU nodes
Moderate deficiency: Slow Office to Computer network

Parallel MPI *continued*

ANTICIPATED hardware resources *continued*

Data from those who rated the anticipated IST HPC hardware resources support as less than adequate

Q: Select the ANTICIPATED level of deficiency for each category.

11 participants

	No deficiency (5)	(4)	Moderate deficiency (3)	(2)	Large deficiency (1)	Uncertain
Num cores	1	3	1	1	3	2
RAM per core	2	1	0	3	2	3
Processor speed	1	2	1	2	4	1
Node interconnect speed	1	1	1	2	2	4
Disk space	1	2	2	2	2	2
Other	0	0	1	0	1	9

Explanations of *Other*:

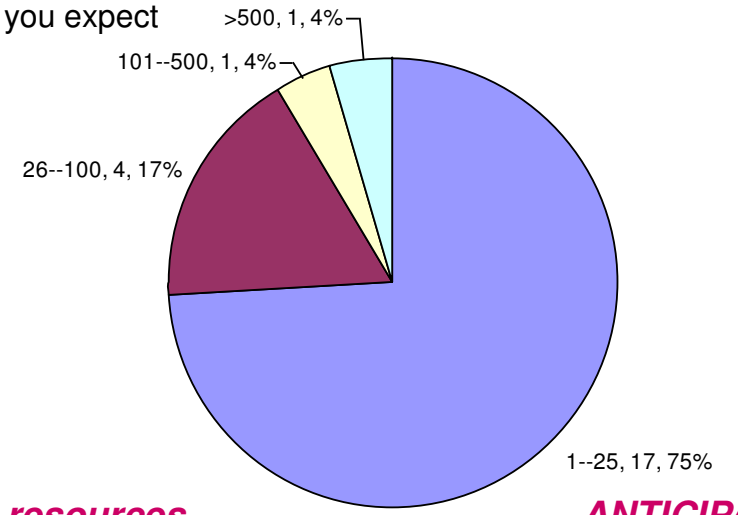
Large deficiency: GPU nodes
Moderate deficiency: GPU systems not available

Serial

Data from those who described their research computing applications as *serial*

Q: How many serial jobs do you expect to run in a typical week?

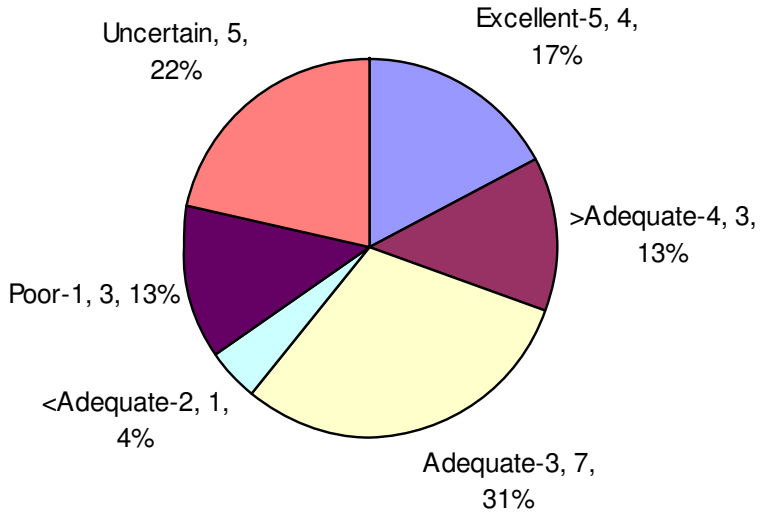
23 participants



CURRENT hardware resources

Q: Please rate the IST HPC hardware resources support for your current serial computing needs.

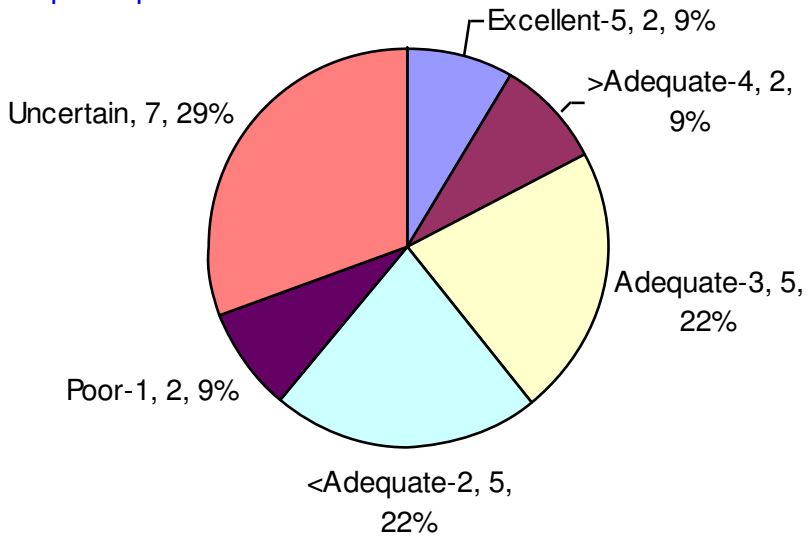
23 participants



ANTICIPATED hardware resources

Q: Please rate the IST HPC hardware resources support for your anticipated serial computing needs.

23 participants



Serial *continued*

CURRENT hardware resources *continued*

Data from those who rated the current IST HPC hardware resources support as less than adequate or poor

Q: Select the CURRENT level of deficiency for each category.

4 participants

	No deficiency (5)	(4)	Moderate deficiency (3)	(2)	Large deficiency (1)	Uncertain
Num cores	0	0	1	1	1	1
RAM per core	0	0	1	1	1	1
Processor speed	0	0	0	2	1	1
Disk space	0	0	1	1	1	1
Other	1	0	1	0	0	2

Explanations of *Other*:

Moderate deficiency: Need faster campus network

Serial *continued*

ANTICIPATED hardware resources *continued*

Data from those who rated the current IST HPC hardware resources support as less than adequate

Q: Select the CURRENT level of deficiency for each category.

7 participants

	No deficiency (5)	(4)	Moderate deficiency (3)	(2)	Large deficiency (1)	Uncertain
Num cores	1	0	1	3	1	1
RAM per core	1	1	2	1	1	1
Processor speed	1	0	2	3	1	0
Disk space	0	0	2	1	2	2
Other	1	0	1	0	0	5

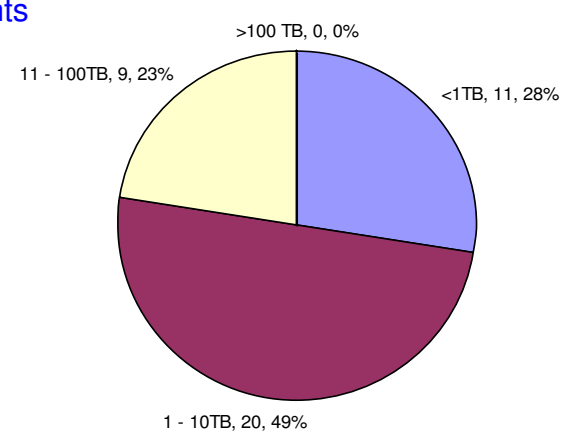
Explanations of *Other*:

Moderate deficiency: GPU systems needed

Data storage and files

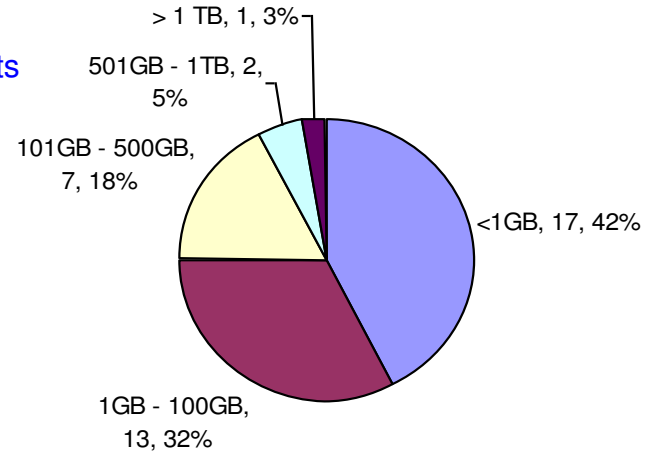
Q: How much data storage do you estimate you will need in the next 3 years to perform your research?

40 participants



Q: What do you estimate will be the maximum size of HPC files that you share with researchers outside of NJIT?

40 participants



Q: Which of the following file systems do your use, or would use if available?

40 participants

	Number YES	% YES
Local to cluster	22	55%
AFS / NFS	24	60%
Parallel	12	30%
Hadoop Distributed File System or similar	4	10%
Uncertain	11	27.5%
Other	1	2.5%

Explanation of *Other*:

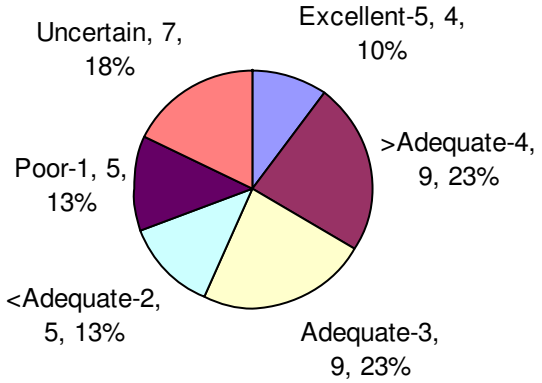
Lab systems

Software

NOTE: 1 participant discontinued the survey at this point.

Q: Please rate the current IST HPC software resources for your computing needs.

39 participants



Data from those who rated the current IST HPC software resources as more than moderately deficient or poor

Q: In which categories, and to what extent, are there deficiencies?

10 participants

	No deficiency (5)	(4)	Moderate deficiency (3)	(2)	Large deficiency (1)	Uncertain
Compilers	2	1	3	1	1	2
Libraries	2	1	1	2	2	2
Open Source applications	1	1	1	2	3	2
Commercial applications	1	0	2	1	5	1
Other	2	0	1	0	0	7

Software *continued*

Data from those who rated the current IST HPC software resources as more than moderately deficient *continued*

	RATING SCALE						
10 participants	No deficiency	<Moderate deficiency	Moderate deficiency	>Moderate deficiency	Large deficiency	Other	Uncertain

Deficiencies in COMPILERS:

Large: Need Intel Compilers and Portland Compilers
>Moderate: Versions can be outdated sometimes
Moderate: Parallel compilation for HDL tools (chip place-and-route)
Moderate: debugging visualization software
Moderate: Ability to run MatLab in batch mode for large data sets
<Moderate: Not available

Deficiencies in LIBRARIES:

Large: NAG
Large: finding the libraries is a problem along with the documentation
>Moderate: Appropriate libraries sometimes unavailable
>Moderate: Linpack and some other libraries I use are not available
Moderate: Nothing of significance
<Moderate: Not available

Deficiencies in OPEN SOURCE APPLICATIONS:

Large: Not diverse enough
Large: documentation and time for installation of the updated version
Large: GAMESS still not working, same for Cfour
>Moderate: Appropriate applications sometimes outdated/unavailable
>Moderate: Visualization/animation software to visualize data
<Moderate: Not available

Deficiencies in COMMERCIAL APPLICATIONS
Large deficiency:

Large: I don't believe we have any at NJIT
Large: The graphics software Tecplot has only two licenses
Large: Lack of necessary software packs
Large: TAU software
Large: visualization packages: e.g. I had to install tecplot myself and purchase licenses since NJIT one was not functional on regular basis
>Moderate: Do you use commercial codes for research
Moderate: Not available

Other aspects of IST HPC resources

Q: Please rate these other aspects of the IST HPC resources .

39 participants	Excellent (5)	>Adequate (4)	Adequate (3)	<Adequate (2)	Poor (1)	Uncertain
Ease of use	5 (12.8%)	8 (20.5%)	14 (35.9%)	3 (7.7%)	0 (0%)	9 (23.1%)
Consistent environment across hardware resources	4 (10.3%)	8 (20.5%)	13 (33.3%)	2 (5.1%)	0 (0%)	12 (30.8%)
Documentation	1 (2.6%)	6 (15.4%)	12 (30.8%)	4 (10.3%)	4 (10.3%)	12 (30.8%)

Comments from those who rated the current IST HPC software resources as less than adequate or poor

9 participants

EASE OF USE:

<Adequate: The disconnect between Kong and AFS
<Adequate: It takes me a long time to find documentation

CONSISTENT ENVIRONMENT:

<Adequate: Some programs run on OSL but not on AFS machines
<Adequate: When software moved across machines

DOCUMENTATION:
<Adequate: What documentation?
<Adequate: I don't believe we have any
<Adequate: Where is it?
<Adequate: Tutorials on how to use, what is available, who can use it
Poor: Not easy to find!
Poor: no documentation is available
Poor: I can't find updated documentation with simple examples of implementation
Poor: Documentation not comprehensive and not well-organized

Tech support from various resources

Q: How important to your research is the technical support from each of these resources?

39 participants	Critical	Important	Not Important	Uncertain
IST Univ. Computing Systems	19 (48.7%)	16 (41%)	3 (7.7%)	1 (2.6%)
Your dept. or college	9 (23.1%)	17 (43.6%)	9 (23.1%)	4 (10.3%)
Group/grant	14 (35.9%)	13 (33.3%)	7 (17.9%)	5 (12.8%)
Non-vendor outside	1 (2.6%)	10 (25.6%)	12 (30.8%)	16 (41%)
Vendor	3 (7.7%)	14 (35.9%)	11 (28.2%)	11 (28.2%)
Other	0 (0%)	1 (2.6%)	11 (28.2%)	27 (69.2%)

Explanation of *Other*:

Important: When problems are trivial, better to ask google

Q: In the next 3 years, how frequently are you likely to seek help from University Computing Systems in each of these areas?

39 participants	Critical	Monthly	Weekly	Daily	Never	Uncertain
Parallel programming	11 (28.2%)	6 (15.4%)	6 (15.4%)	1 (2.6%)	3 (7.7%)	12 (30.8%)
Serial programming	8 (20.5%)	4 (10.3%)	3 (7.7%)	1 (2.6%)	9 (23.1%)	14 (35.9%)
Visualization	8 (20.5%)	4 (10.3%)	3 (7.7%)	1 (2.6%)	8 (20.5%)	15 (38.5%)
Software Installation	18 (46.2%)	5 (12.8%)	1 (2.6%)	2 (5.1%)	3 (7.7%)	10 (25.6%)
Running open-source or commercial applications	18 (46.2%)	3 (7.7%)	3 (7.7%)	1 (2.6%)	5 (12.8%)	9 (23.1%)
Other	2 (5.1%)	0(0%)	0(0%)	0(0%)	3 (7.7%)	34 (87.2%)

Explanations of *Other*:

Every few months: Networking
Every few months: If win an NSF grant, we are to establish a test center

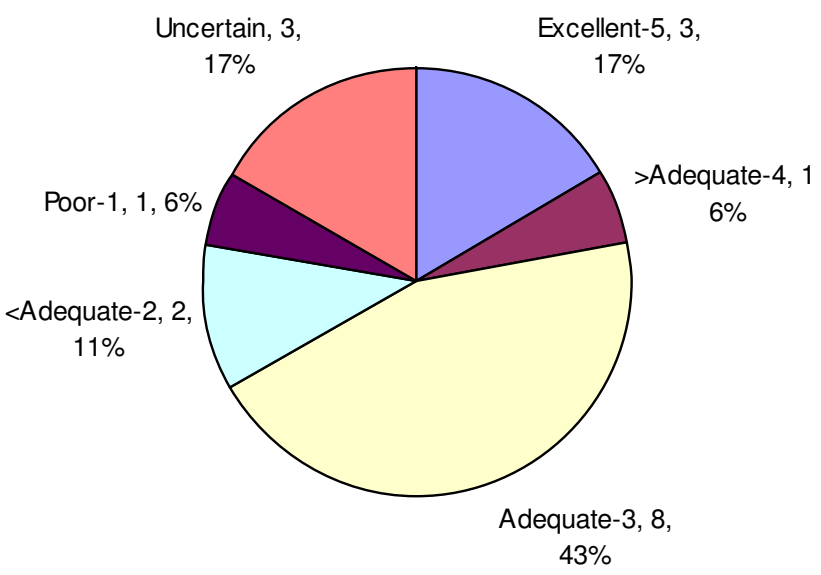
IST hardware for teaching

Data from the 18 out of 39 participants who reported using or planning to use IST HPC resources in courses that they teach or assist in teaching

CURRENT hardware resources

Q: Please rate the IST HPC hardware resources support for your CURRENT course-related computing needs.

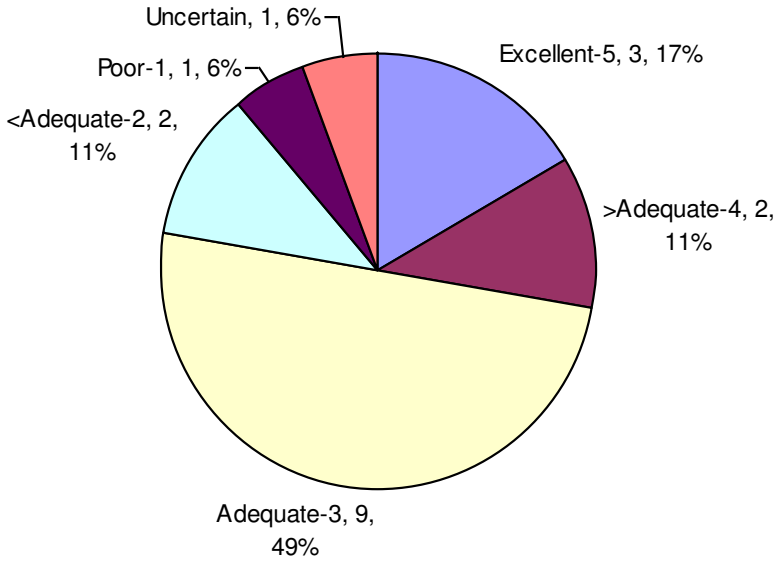
18 participants



ANTICIPATED hardware resources

Q: Please rate the IST HPC hardware resources support for your ANTICIPATED course-related computing needs.

18 participants



IST hardware for teaching *continued*

CURRENT hardware resources *continued*

Data from those who rated the current IST HPC hardware resources in courses as less than adequate or poor

Q: Select the CURRENT level of deficiency for each category.

3 participants	No deficiency (5)	(4)	Moderate deficiency (3)	(2)	Large deficiency (1)	Uncertain
Num cores	0	0	1	1	0	1
RAM per core	0	1	0	1	0	1
Processor speed	0	0	0	2	0	1
Node interconnect speed	0	0	0	1	0	2
Disk space	0	1	1	0	0	1
Other	0	0	0	0	1	2

Deficiencies NUMBER OF CORES AVAILABLE:
Moderate: sometimes need more cores for more reasonable job completion time
>Moderate: need more cores

Deficiencies PROCESSOR SPEED:
>Moderate: Faster processors
>Moderate: processor speed is reasonable for my group

Deficiencies DISK SPACE:
<Moderate: often need to have disk cleaned more often -- sorry for this students need to clean themselves
>Moderate: More space for students is always good. I'm glad the quotas are now 1G

Deficiencies RAM PER CORE:
>Moderate: jobs require a lot of ram for efficient running—often need more
<Moderate: Need more RAM (at least 12GB per node

Deficiencies NODE INTERCONNECT SPEED:
>Moderate: uncertain here

OTHER Deficiencies:
Large: GPU nodes

IST hardware for teaching *continued*

ANTICIPATED hardware resources *continued*

Data from those who rated the current IST HPC hardware resources in courses as less than adequate

Q: Select the ANTICIPATED level of deficiency for each category.

3 participants	No deficiency (5)	(4)	Moderate deficiency (3)	(2)	Large deficiency (1)	Uncertain
Num cores	0	1	0	0	0	2
RAM per core	0	1	0	0	0	2
Processor speed	0	0	1	0	0	2
Node interconnect speed	0	0	0	0	0	3
Disk space	0	1	0	0	0	2
Other	0	0	0	0	1	2

Deficiencies NUMBER OF CORES AVAILABLE:
<Moderate: the bar is continually rising on research level - need more cores for higher level calcs

Deficiencies RAM PER CORE:
<Moderate: need more ram for higher level - more accurate calcs

Deficiencies PROCESSOR SPEED:
Moderate: processor speed will help

Deficiencies DISK SPACE:
<Moderate: jobs crash or not running -- sorry for this students need to clean themselves

OTHER Deficiencies :
Large: GPU nodes and other deficiencies same as current

Open comments

11 participants

- I anticipate greatly increased data volume and computing resource needs in the near future, as a large science instrument construction project comes to completion. I have not needed much parallel processing and help with data management in the past, but that will change. I do not yet know what resources are available and which may be deficient, but I anticipate that the current big data support is probably not adequate and some enhancement and growth is required.

- Please provide IST HPC group more resources. Their support is very critical for NJIT research and teaching.

- It is difficult to know for sure but strongly suspect that staffing - control of computer and maintenance is limited by the limited staff who are overwhelm by the work load

- 1. HPC system inadequate for my research: i.e., outdated machines, insufficient memory on processors, CPU speed is slow, insufficient disk space, file transfer procedures are not good, and very limited computing resources to run jobs.
2. Lack of up-to-date HPC makes my ability to attract funding (to be competitive with others in my field) impossible. an overall lack of infrastructure to support research at NJIT - of which HPC resources is one example.
3. No strategic plans at the Institute level to invest in needed HPC resources (hardware/software) for research so that faculty can be competitive in seeking funds. This is extremely short-sighted given the thrust by external agencies (NSF, for example) to support computational research ("cyberinfrastructure", petascale computing, materials genome, etc.)
4. I do not see any movement at NJIT to initiate a long-term strategy to address the current deficiences in its high performance computing systems.
5. Number of people that support HPC is even now clearly inadequate to handle to the needs of faculty who are trying to conduct this type of research. The fact that the current staff is able to do such a decent job is a credit to them.
6. There is an overall lack of infrastructure to support research at NJIT - of which HPC resources is one example.

continued…

Open comments *continued*

11 participants

- Training sessions will be good; Support for MapReduce will be good

- I am quite interested in big data mining in the cloud. It would be helpful if the university provides Hadoop so students can learn map/reduce programming.

- Technical support is very important, and there are not enough people to provide this support. Sometimes jobs sit in a queue for too long. This needs to be improved.

- Most of the responses are in anticipation of a new hire whose research work easily fits under the Big Data category.

- My AFS account had NOT been reset, even after multiple request, to enable me to use it. Right now I pretty much use my notebook PC and iMac to do research and teaching.

- Some of the questions were confusing because without the resources of for example updated documentation, you can not comment on libraries and applications that may exist but may or may not be currently available for your research.

- If we get fund to establish a software platform for tampered image detection, I hope to get help from computing Center of NJIT.