SuperMike-II Launch Workshop: System Overview and Allocations

Dr Jim Lupo
CCT Computational Enablement
jalupo@cct.lsu.edu

SuperMike-II: Serious Heterogeneous Computing Power

System Hardware

SuperMike provides 442 nodes, 221TB of local disk, and 400TB of storage.

It is considered heterogeneous because of the types of nodes provided. However, all share some common features:

- Dual 8-core Intel Sandy Bridge 2.6GHz Xeon CPUs
- Minimum of 2GB per core (32 – 256GB total per node)
- 500 GB local hard disk drive.
- 40 Gbps QDR InfiniBand communication fabric.
- 1 GbE Ethernet management network.
- Red Hat Enterprise Linux 6 operating system.

Heterogeneous System?

SuperMike-II supports 5 different computing models:

Distributed-memory: message passing processing using MPI across multiple mult-core nodes.

Shared-memory: multi-threading using OpenMP on a single multi-core node.

Accelerated: processing using GPU's, developed using CUDA or compiler extensions.

Multi-node SMP: (Symmetric Multi-Processor) ScaleMP 8-node environment.

Hybrid programming: mixing and matching models as needed in a single application.

Interactive Nodes

There are 2 interactive nodes (also called head nodes). These support user login for development work, job management, and general computation.

More memory: 4 GB per core (64 GB total).

Do Not run production work on these nodes! Just can't seem to say this enough times.

Compute Nodes

382 nodes are available for production work. They are accessed via the PBS job manager, just as on all the other Intel clusters available here.

Basic memory: 2 GB per core (32 GB total per node)

These can be used for serial jobs, parallel jobs with MPI, single node OpenMP jobs, or parallel hybrid MPI-OpenMP jobs on multiple nodes.

GPU Nodes

50 nodes are each equipped with 2 NVIDIA M2090 GPUs. Each M2090 has a peak double-precision processing rate of 665 TFlops.

Each has 4 GB per core, giving 64 GB per node.

Some applications, such as LAMMPS and NAMD, are already available to run on GPUs.

CUDA and PGI compilers with PGI Accelerator are available for GPU programming.

SMP SuperNode

8 nodes, each with 256GB memory, are joined into a virtual SMP processor. That means 128 cores have a single system image with 2TB of shared memory.

1 process using 2TB for a large memory application.

128-core shared-memory OpenMP process.

Noteworth Items

#PBS -l ppn=16

#PBS -l ppn=1

OpenMPI is default MPI library

Queues: workq, checkpt, single, gpu
mwfa, lasigma

SMP node not integrated yet.

Job Queues

Name	Max Wall Time (hrs)	Max Nodes
single	72	1
workq	72	128
checkpt	72	200
gpu	24	16
mwfa	24	8
lasigma	24	28

Performance Measures

- 250'th in the Top500 rankings
- 146 TFlops peak performance
- 10x faster than Tezpur

Real world comparison – more this afternoon – using FVCOM (a shallow water CFD code) on 64 cores of Eric, Queen Bee, Pandora, and SuperMike-II. The metric was *time per iteration*:

SuperMike-II:	**0.033 secs**
Queen Bee:	0.102 secs (3.1x slower)
Pandora:	0.130 secs (3.93x slower)
Eric:	0.135 secs (4.1x slower)

Storage Subsystem

/home – 2TB of NFS-mounted space.
/work – 200TB of high-performance Lustre FS ephemeral space.
/project – 200TB of high-performance Lustre FS persistent space.

/work is shared access space – no quota limits on size – but is subject to purging if consumption gets too high.

/project is quota-controlled space with at least 6-month life-time. It is allocation controlled, much as SUs are controlled.

Software Support

Basic Utilities: RHEL 6

Private Utilities: Installed in user directories for private use.

Group Utilities: Installed in user directories for group use.

Global Utilities: Installed for all, environment via Softenv.

Upcoming Events

HPC@LSU Announcements

Alphabetical List of Software

- aces
- amber
- ansys
- ansys ls-dyna
- apache_ant
- arpack
- atlas
- autodock
- autotools [automake autoconf m4]
- bbcp
- blacs
- blast
- boost
- boostjam
- cmake
- cmg-stars
- condor
- cp2k
- cpmd
- cuda
- ddd
- ddt
- delft3d
- dl_poly
- espresso md
- fftw
- fluent
- gamess
- ganga
- gaussian
- gaussian view
- gaussview
- gcc-4.3.2
- gcc-4.7.0
- gcc-4.7.2
- git
- globus
- globus toolkit
- gnuplot
- graphviz
- gromacs
- gsl
- hdf5
- hypre
- imagemagick
- intel
- intel-cc
- intel-fc
- intel compiler suite
- java 2 sdk
- jdk
- jmol
- lammps
- lapack
- lmto
- maple
- material studio
- mathematica
- matlab
- maya
- metis
- migrate
- mkl
- mpiblast
- mpich
- mpich2
- mpip
- mrbayes
- mvapich
- mvapich2
- namd
- ncbiblast
- nciplot
- netcdf
- netcdf_fortran
- nvidia driver
- nwchem
- octave
- openbabel
- openeye
- openmpi
- orca
- parallel_netcdf
- parmetis
- parpack
- paup
- pdtoolkit
- perl
- petsc
- piny-md
- piny_md
- postgresql
- python
- quantum espresso
- r
- ruby
- saga
- scalapack
- scons
- sprng
- sprng random number generator
- superlu_dist
- tau
- tcl
- the portland group compilers
- totalview
- uberftp
- valgrind
- visit
- visual molecular dynamics
- vmd
- wien2k
- xcrysden

The information in this database may not reflect all software available on HPC systems. If you do not see an application that you wish to use, or if you have questions about software that is currently available, please contact the HPC Help Desk.

Softenv Environment Manager

Version and Availability

Machine	Version	Softenv Key
eric	4.2	+autodock-4.2-gcc-4.3.2
eric	4.3	+autodock-4.3-gcc-4.3.2
qb	4.2	+autodock-4.2-gcc-4.3.2
qb	4.3	+autodock-4.3-gcc-4.3.2
oliver	4.2	+autodock-4.2-gcc-4.3.2
oliver	4.3	+autodock-4.3-gcc-4.3.2
louie	4.2	+autodock-4.2-gcc-4.3.2
louie	4.3	+autodock-4.3-gcc-4.3.2
poseidon	4.2	+autodock-4.2-gcc-4.3.2
poseidon	4.3	+autodock-4.3-gcc-4.3.2
painter	4.2	+autodock-4.2-gcc-4.3.2
painter	4.3	+autodock-4.3-gcc-4.3.2
tezpur	4.2	+autodock-4.2-gcc-4.3.2
tezpur	4.3	+autodock-4.3-gcc-4.3.2
supermike2	4.2.3	+autodock-4.2.3-Intel-13.0.0

Decoding The Key

+namd-2.8-intel-11.1-mpich2-1.3.2p1-cuda-4.0.17

- **namd-2.8** ..... application name and version
- **intel-11.1** ..... compiler (runtime libraries)
- **mpich2-1.3.2p1** .. MPI (launcher and runtime libraries)
- **cuda-4.0.17** ..... GPU support (runtime libraries)

The presence of GPUs on the system means some applications may have builds supporting them.

Take-aways

Support for software at multiple levels.

Environment controlled by Softenv

Applications must be GPU enabled to use them.

Allocations For LSU HPC

Dr Jim Lupo
CCT Computational Enablement
jalupo@cct.lsu.edu

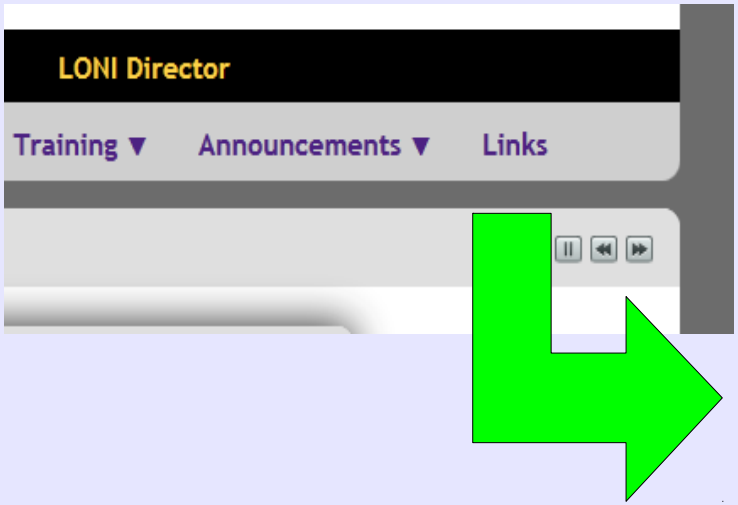
System Policies

Usage – Who is allowed on the systems, and what may be done.

Allocation – Control of computational resources.

Storage – Control of data storage resources.

Quick Links Page



http://www.hpc.lsu.edu/

Quick Links

Quick Links if you do not want to navigate through the website.

Article	Organization	
User Account Request	LONI	LSU
User Allocation Request(*)($)	LONI	LSU
User Password Reset(*)	LONI	LSU
User Profile Settings(*)	LONI	LSU
Policy – Usage	LONI	LSU
Policy – Allocations	LONI	LSU
Policy – Storage	LONI	LSU
Available Systems	LONI	LSU
Available Software	LONI	LSU
System User Guides	LONI	LSU

User Accounts

LSU faculty, research staff, and students may apply for HPC accounts. Students need a faculty sponsor.

Outside collaborators working with LSU faculty and staff may be sponsored for an account.

User accounts only gain you access to the head nodes, allowing use of basic utilities and access to file systems.

Production computing requires an allocation of CPU time.

Controlled Access to Resources

NSF XSEDE	Peer-reviewed proposals, examining the science and technical components, awards arbitrated by committee.
LONI	Committee reviewed technical proposal.
LSU	Provided unrestricted access since inception of centralized HPC service. Changing to introduce a low barrier-to-entry committee reviewed system.

Rational for Implementation

Allocations will allow system resources to be dedicated (and tracked) by specific usage categories:

Category	Available Resources	Allocation Authority
Economic Development	10%	Vice-Chancelor RED
Discretionary	10%	CCT Director
Default Allocations (2,000 SU)	5%	HPC@LSU Staff
Startup Allocations (2,000 – 50,000 SU)	15%	HPCRAC Chair
Research Allocations (> 50,000 SU)	60%	HPCRAC

HPCRAC = HPC Resource Allocation Committee

SU = Service Unit = 1 core-hour

HPCRAC

Name	Department	Contact Email
Honggao Liu (Chair)	CCT	honggao@cct.lsu.edu
Juana Moreno	Physics	moreno@physics.lsu.edu
Jim Q. Chen	Civil & Environ. Engineering	qchen@lsu.edu
Shawn W. Walker	Mathematics	walker@math.lsu.edu
Krishnaswamy Nadakumar	Chemical Engineering	nandakumar@lsu.edu
Jeremy Brown	Biological Sciences	jembrown@lsu.edu

What is an Allocation?

Service Units (SU): 1 wall-clock hour on 1 processor core.

Allocation: Bank account of SUs

Job submission system requires an allocation account code with a positive balance before allowing execution.

"Free", but resource consumption is limited.

"Competitive", in that requests are reviewed for suitability and available resources.

How is it charged?

Service Units (SU): 1 wall-clock hour on 1 processor core.

Under most circumstances, only entire nodes are assigned to jobs. That means 16 cores per node.

Running for 1 wall-clock hour on 1 node? 16 SU.
Running for 10 wall-clock hours on 50 nodes? 8,000 SU.

Its all arithmetic from here....

Only time on compute nodes is charged: batch or interactive

Who May Request an Allocation

LSU faculty or permanent research staff may serve as a PI on an allocation request.

The PI controls authorization of other users who may then charge to allocation.

How Do I Get Me One O' Them Allocations?

Default Allocation – granted with user account.

Startup Allocation – lightweight request.

Research Allocation – request requires proposal.

Join an existing allocation?

default Allocation

- Awarded to every user upon account creation.
- 2,000 SU's.
- 1 year expiration.
- Non-renewable.
- May be awarded anytime during the year.

Supports access for basic processing needs to all eligible users. Meets most student needs, and provides basis for deciding if some other type of allocations is needed.

startup Allocation

- minimal proposal required (web form).
- 2,000 – 50,000 SUs.
- 1 year expiration, subject to renewal request.
- Max of 2 active per PI.
- May be awarded anytime during the year.

Supports development or proof-of-concepts work in preparation for applying for a research allocation, and other low-intensity efforts.

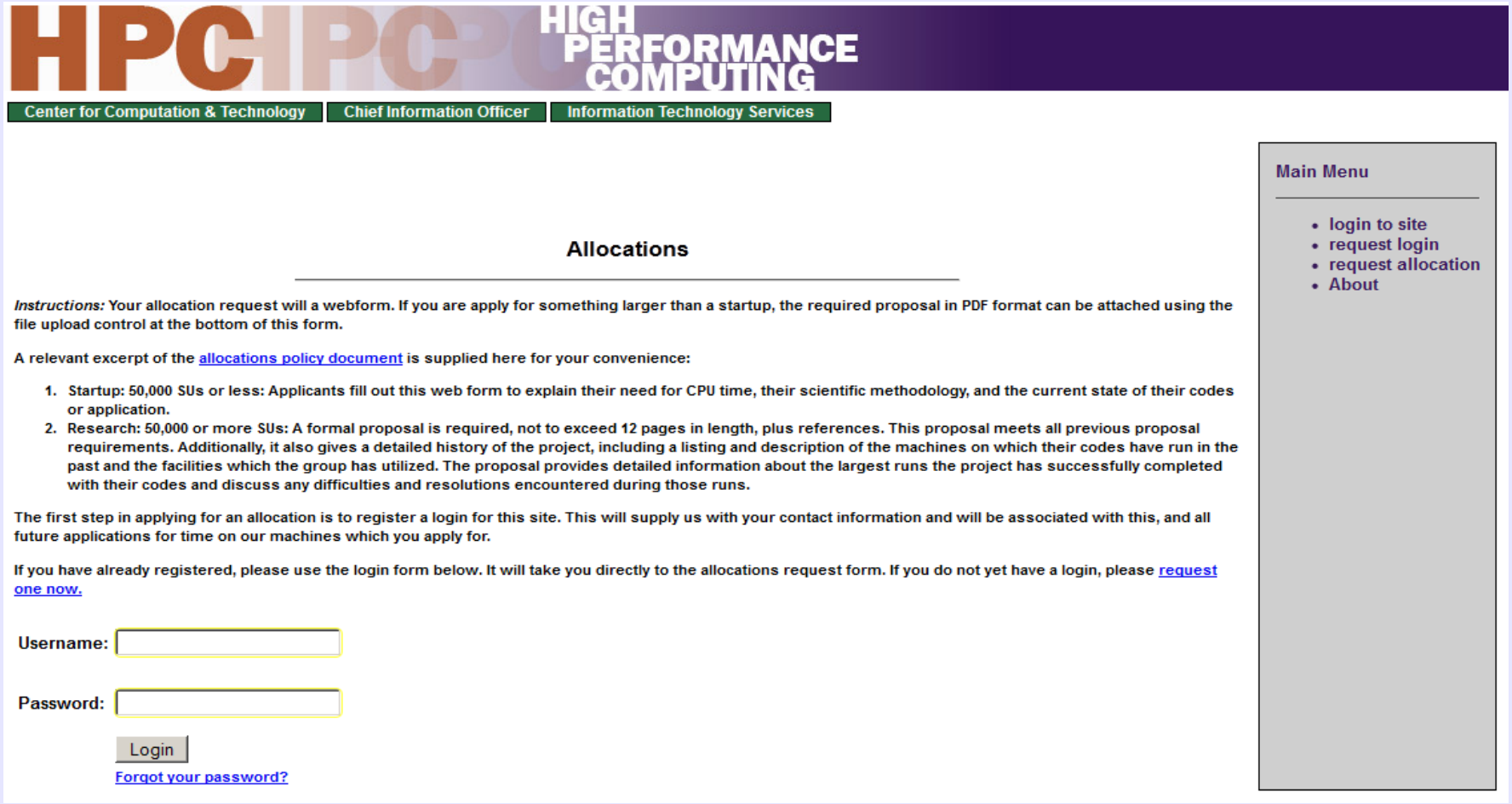
research Allocation

- Proposal required (1 – 5 pages)
- Limited to 3M SU.
- 1 year expiration.
- May be renewed via follow-on proposal.
- PI limited to 5M SU total active allocations.
- Awarded quarterly, proposals due 1 month prior.

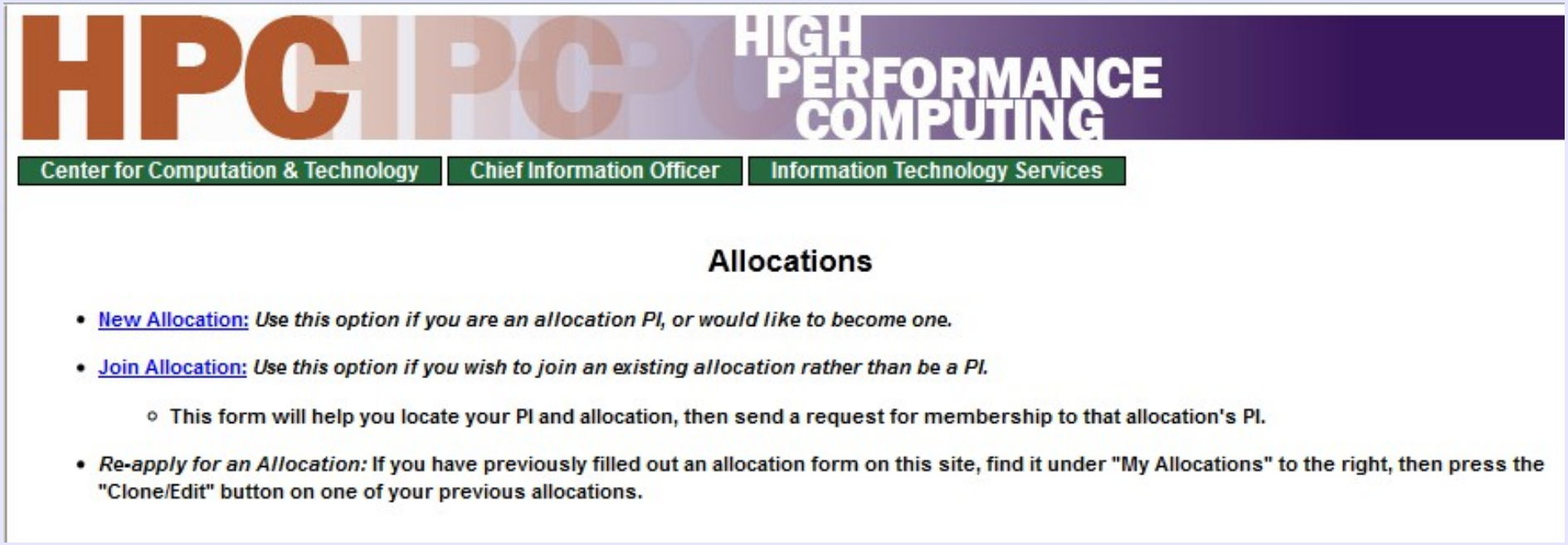
Supports major efforts, preparation for outside proposals.

Submit an Allocation Request

It all starts at: https://accounts.hpc.lsu.edu/allocations.php



Allocation Start Page



Note that you come here to ***request*** a new allocation or to ***join*** an existing allocation.

Join An Allocation

Join an Existing Allocation 1

To find the allocation PI, please enter one of these items:

1. The PI's email address,
2. The PI's full name, or
3. The PI's login ID.

search

Uniquely identify the PI.

Join an Existing Allocation 2

Search Term: jalupo				
Full Name	UID	E-mail address(es)	Phone Number(s)	
James Lupo	jalupo	jalupo@cct.lsu.edu	+1 225 578 8899 303-513-3376	Join Projects

To find the allocation PI, please enter one of these items:

1. The PI's email address,
2. The PI's full name, or
3. The PI's login ID.

jalupo search

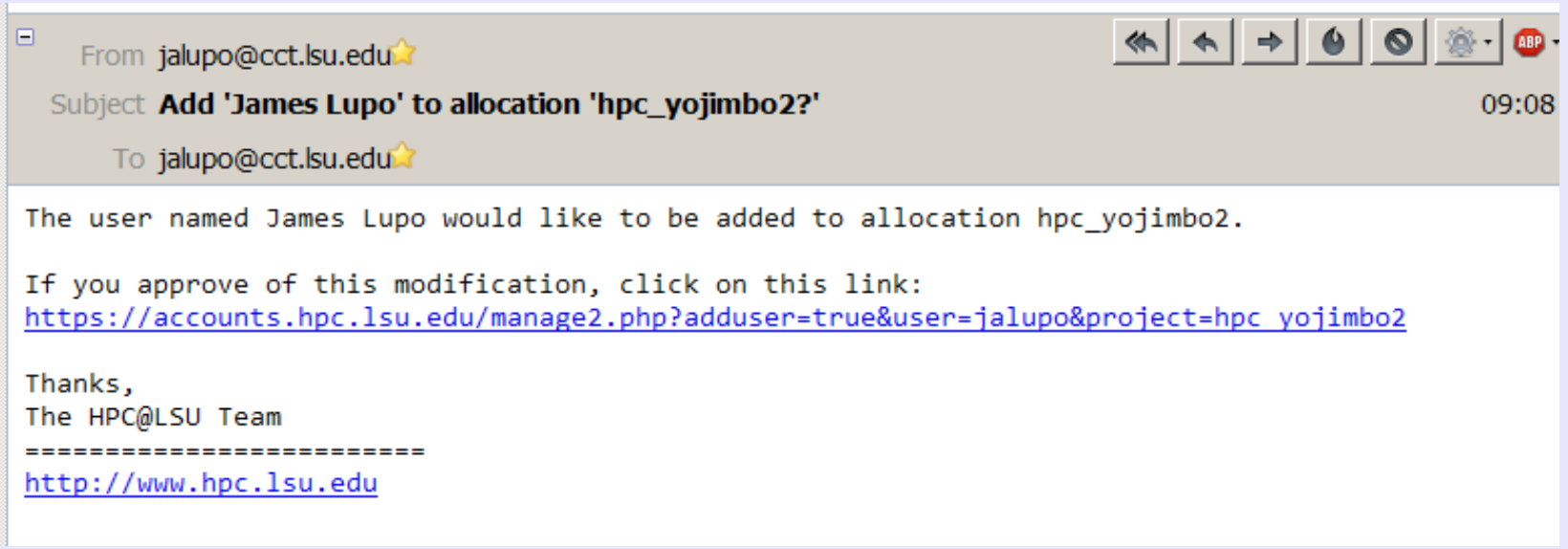
Send join request

Join an Existing Allocation 3

Projects		
allocation	description	
hpc_yojimbo2	Nibble, data driven simulation.	Join
hpc_enable01	Various	Join

PI Response

Selecting a Join button sends an email to the PI, who then just selects the link to approve the request.



From jalupo@cct.lsu.edu

Subject **Add 'James Lupo' to allocation 'hpc_yojimbo2?'** 09:08

To jalupo@cct.lsu.edu

The user named James Lupo would like to be added to allocation hpc_yojimbo2.

If you approve of this modification, click on this link:
https://accounts.hpc.lsu.edu/manage2.php?adduser=true&user=jalupo&project=hpc_yojimbo2

Thanks,
The HPC@LSU Team
==========
http://www.hpc.lsu.edu

Security Verification? Know the user; know the allocation code; hover over link to verify it matches the text.

Request an Allocation

Title of This Allocation:

This is a more verbose identifier than the name. It can contain spaces, and need not be unique. Typically it is the title you have given to your request document.

PI:

This is a list of the Principal Investigators for the project. Each P.I. should be identified by full name (first and last), login name, or email address. Each P.I. identifier should be separated by a comma. For example: John Doe,Jane Smith,Joe Brown.

Research Categories:

You need to provide a primary area of research. If applicable, you can also provide a secondary and tertiary area.

Primary

Secondary

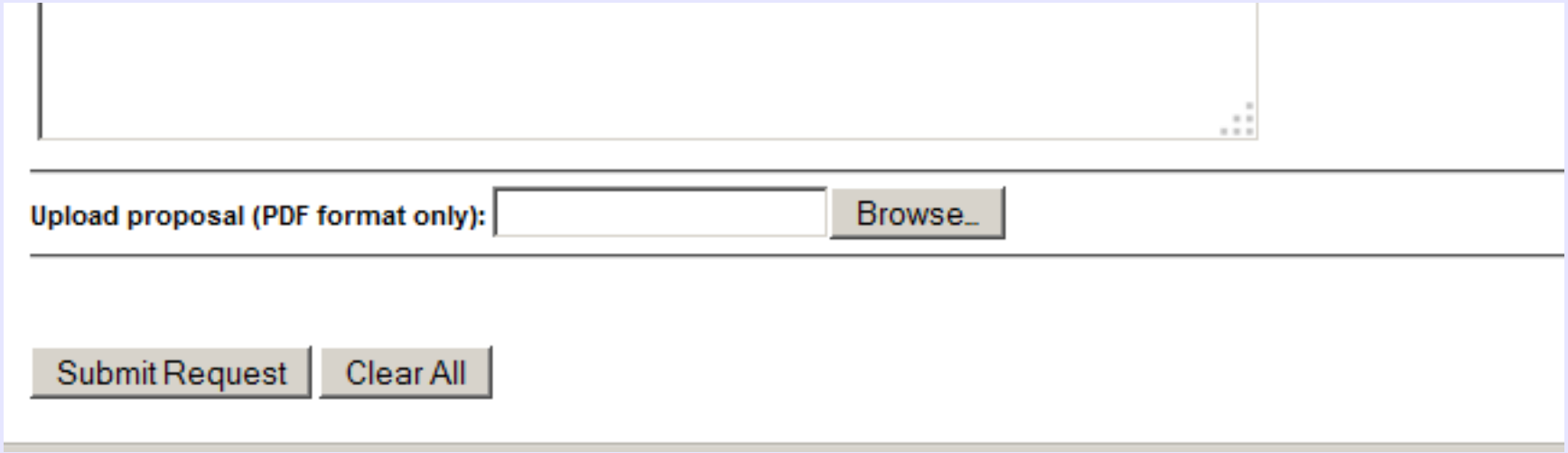
Tertiary

Description of Research:

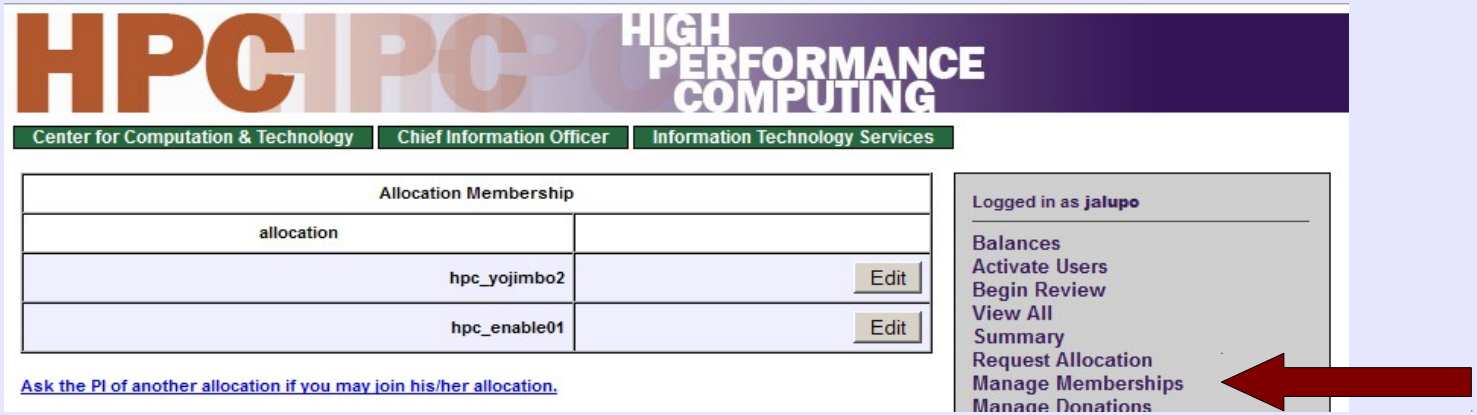
Startup Versus Research

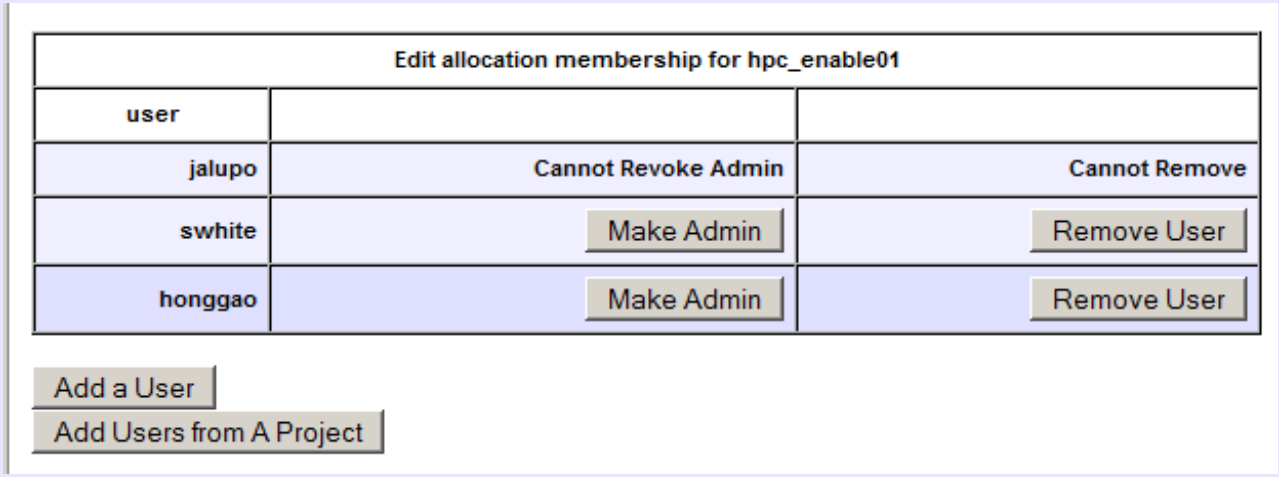
A Startup request just involves completing the form.

A Research request must include a 5 page proposal, which shows up at the very end of the form:

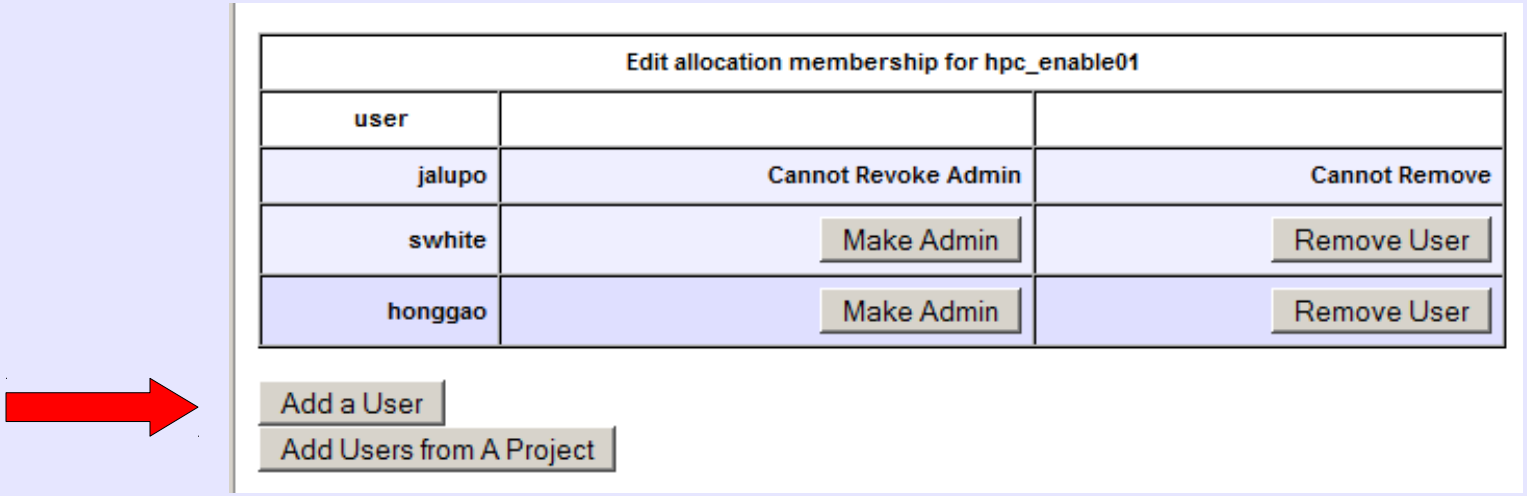


Manage Memberships





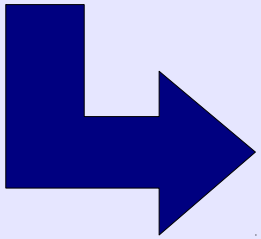
Allocation Admins May Add Users

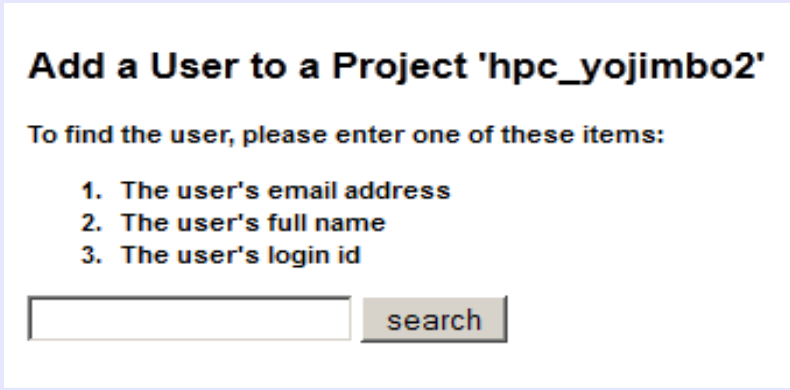


Edit allocation membership for hpc_enable01

user		
jalupo	Cannot Revoke Admin	Cannot Remove
swhite	Make Admin	Remove User
honggao	Make Admin	Remove User

Add a User
Add Users from A Project





Add a User to a Project 'hpc_yojimbo2'

To find the user, please enter one of these items:

1. The user's email address
2. The user's full name
3. The user's login id

search

Storage Allocation

Process is essentially identical to SU allocations.

Awarded in units of 100GB for 6 months at a time, subject to competitive renewal.

Small: 1 unit – basic information on what and why.

Medium: 2 – 10 units – 2 page proposal.

Large: > 10 units – formal proposal, not over 12 pages. History, science details, curation plan. Data must be used, not just stored (e.g. ***not for data archive purposes***)

More Information

The full policy can be found on the HPC web site at:

`www.hpc.lsu.edu`

Please feel free to send any questions to me at:

`jalupo@cct.lsu.edu`

or the HPC help desk at:

`sys-help@loni.org`

Questions?