

Jet User Meeting

16 January 2013

Agenda

- New Allocation Strategy
- Post-Job Memory Statistics
- Upcoming Downtimes
- Changes in login timeouts on bastion hosts
- Open Forum

New Allocation Strategy

- On Jet, controlling access to resources has been by maximum cores in use at any one time
- To ensure that no single user or account could block real-time projects (before reservations)
- Now, all projects will get a set number of core-hours per month
 - This system matches what is done on Zeus and Gaea.
 - This system matches what most all other HPC center do.

How Allocations Work

- Each project is given a number of equivalent core hours (ECH) per year
- Allocations are divided by month
- Allocations are like a bank
 - Deposits are made at the beginning of the month
 - Each job is a withdraw from your balance
 - The amount deducted is $\$NUMBER_OF_CORES * \$WALLTIME_USED * \$CHARGE_RATE$
- Any unused allocations at the end of the month are lost
 - “Use it or lose it”
- Serial jobs are charged as a single core
- Parallel jobs are charged as if the whole node is used, regardless of what is requested

Allocation Assignment

- Allocations are determined by the allocation board, a group of executives within OAR
- Allocations are grouped and assigned by Modeling Center
- These allocations are then broken up by project by the Modeling Centers
- Allocations are by fiscal year
- If there are problems with the allocations, or you need to request a change email rdhpcs.jet.help@noaa.gov and let us know

Equivalent Core Hours

- Different systems have different performance characteristics.
- To account for this, we are using the concept of “Equivalent Core Hours”
- The rate at which you are charged will vary depending on which system your job runs.
- The charge rates may vary over time to help balance system usage.

Jet Charge Rates

Hjet	Njet	Tjet	Ujet	Sjet	Tgpu	service
0.0	1.2	1.0	1.0	2.0	0.0	10.0

Notes:

- Anyone can use hjet now, but it will be decommissioned soon with minimal warning.
- Anyone can access the GPU cluster, but please contact rdhpcs.jet.help@noaa.gov and discuss with us what you want to do with that system.
- The charge rate on the service partition is high to encourage users to use compute nodes for the majority of the work. The service PE should only be used for the parts of workflow that require external connectivity.

Getting information about your account

account_params

Account Params -- Information regarding project associations

User: jsmith

Project: sepp

Initial Allocation: 101917.81

Allocation:	Id	Name	Available	Allocated	PercentUsed
-------------	----	------	-----------	-----------	-------------

Allocation:	--	-----	-----	-----	-----
-------------	----	-------	-------	-------	-------

Allocation:	34	Project=sepp	101917.81	101917.81	0.00
-------------	----	--------------	-----------	-----------	------

Directory: /lfs1/projects/sepp DiskInUse=1702 GB, Quota=2000 GB

Project: jetmgmt

Initial Allocation: 101805.49

Allocation:	Id	Name	Available	Allocated	PercentUsed
-------------	----	------	-----------	-----------	-------------

Allocation:	--	-----	-----	-----	-----
-------------	----	-------	-------	-------	-------

Allocation:	35	Project=jetmgmt	101803.49	101917.81	0.11
-------------	----	-----------------	-----------	-----------	------

Directory: /lfs1/projects/jetmgmt DiskInUse=8111 GB, Quota=15000 GB

Directory: /lfs2/projects/jetmgmt DiskInUse=148894 GB, Quota=300000 GB

Directory: /pan2/projects/jetmgmt DiskInUse=250713 GB, Quota=668609 GB

Allocation Terms:

Name -- Name of allocated project

Amount -- Total Allocation

Reserved -- Reserved requests from current jobs in queue.

Balance -- Total allocation remaining for new jobs.

Available -- Amount currently available for new jobs.

Initial Allocation -- Current Month's Allocation.

All amounts are in core-hours.

Note: If the allocation of the disk space seem unusually large,
it is most likely because no allocation is defined for your project.

Getting information about your account

account_params

```
Account Params -- Information regarding project association

User: jsmith
  Project: sepp
  Initial Allocation: 101917.81
    Allocation: Id Name Available Allocated PercentUsed
    Allocation: -- -----
    Allocation: 34 Project=sepp 101917.81 101917.81 0.00

    Directory: /lfs1/projects/sepp DiskInUse=1702 GB, Quota=2000 GB

  Project: jetmgmt
  Initial Allocation: 101805.49
    Allocation: Id Name Available Allocated PercentUsed
    Allocation: -- -----
    Allocation: 35 Project=jetmgmt 101803.49 101917.81 0.11

    Directory: /lfs1/projects/jetmgmt DiskInUse=8111 GB, Quota=15000 GB
    Directory: /lfs2/projects/jetmgmt DiskInUse=148894 GB, Quota=300000 GB
    Directory: /pan2/projects/jetmgmt DiskInUse=250713 GB, Quota=668609 GB
```

Current Monthly
Deposit

Allocation Terms:

```
    Name -- Name of allocated project
    Amount -- Total Allocation
    Reserved -- Reserved requests from current jobs in queue.
    Balance -- Total allocation remaining for new jobs.
    Available -- Amount currently available for new jobs.
    Initial Allocation -- Current Month's Allocation.
```

All amounts are in core-hours.

Note: If the allocation of the disk space seem unusually large,
it is most likely because no allocation is defined for your project.

Getting information about your account

account_params

Account Params -- Information regarding project associations

All Deposits

User: jsmith

Project: sepp

Initial Allocation: 101917.81

Allocation:	Id	Name	Available	Allocated	PercentUsed
-------------	----	------	-----------	-----------	-------------

Allocation:	--	-----	-----	-----	-----
-------------	----	-------	-------	-------	-------

Allocation:	34	Project=sepp	101917.81	101917.81	0.00
-------------	----	--------------	-----------	-----------	------

Directory: /lfs1/projects/sepp DiskInUse=1702 GB, Quota=2000 GB

Project: jetmgmt

Initial Allocation: 101805.49

Allocation:	Id	Name	Available	Allocated	PercentUsed
-------------	----	------	-----------	-----------	-------------

Allocation:	--	-----	-----	-----	-----
-------------	----	-------	-------	-------	-------

Allocation:	35	Project=jetmgmt	101803.49	101917.81	0.11
-------------	----	-----------------	-----------	-----------	------

Directory: /lfs1/projects/jetmgmt DiskInUse=8111 GB, Quota=15000 GB

Directory: /lfs2/projects/jetmgmt DiskInUse=148894 GB, Quota=300000 GB

Directory: /pan2/projects/jetmgmt DiskInUse=250713 GB, Quota=668609 GB

Allocation Terms:

Name -- Name of allocated project

Amount -- Total Allocation

Reserved -- Reserved requests from current jobs in queue.

Balance -- Total allocation remaining for new jobs.

Available -- Amount currently available for new jobs.

Initial Allocation -- Current Month's Allocation.

All amounts are in core-hours.

Note: If the allocation of the disk space seem unusually large,
it is most likely because no allocation is defined for your project.

Getting information about your account

account_params

Account Params -- Information regarding project associations

User: jsmith

Project: sepp

Initial Allocation: 101917.81

Allocation:	Id	Name	Available	Allocated	PercentUsed
-------------	----	------	-----------	-----------	-------------

Allocation:	--	-----	-----	-----	-----
-------------	----	-------	-------	-------	-------

Allocation:	34	Project=sepp	101917.81	101917.81	0.00
-------------	----	--------------	-----------	-----------	------

Directory: /lfs1/projects/sepp DiskInUse=1702 GB, Quota=2000 GB

Project: jetmgmt

Initial Allocation: 101805.49

Allocation:	Id	Name	Available	Allocated	PercentUsed
-------------	----	------	-----------	-----------	-------------

Allocation:	--	-----	-----	-----	-----
-------------	----	-------	-------	-------	-------

Allocation:	35	Project=jetmgmt	101803.49	101917.81	0.11
-------------	----	-----------------	-----------	-----------	------

Directory: /lfs1/projects/jetmgmt DiskInUse=8111 GB, Quota=15000 GB

Directory: /lfs2/projects/jetmgmt DiskInUse=148894 GB, Quota=300000 GB

Directory: /pan2/projects/jetmgmt DiskInUse=250713 GB, Quota=668609 GB

Allocation Terms:

Name -- Name of allocated project

Amount -- Total Allocation

Reserved -- Reserved requests from current jobs in queue.

Balance -- Total allocation remaining for new jobs.

Available -- Amount currently available for new jobs.

Initial Allocation -- Current Month's Allocation.

All amounts are in core-hours.

Note: If the allocation of the disk space seem unusually large,
it is most likely because no allocation is defined for your project.

Balance
Remaining

Getting information about your account

account_params

```
Account Params -- Information regarding
User: jsmith
Project: sepp
Initial Allocation: 101917.81
Allocation: Id Name Available Allocated PercentUsed
Allocation: -- -----
Allocation: 34 Project=sepp 101917.81 101917.81 0.00

Directory: /lfs1/projects/sepp DiskInUse=1702 GB, Quota=2000 GB

Project: jetmgmt
Initial Allocation: 101805.49
Allocation: Id Name Available Allocated PercentUsed
Allocation: -- -----
Allocation: 35 Project=jetmgmt 101803.49 101917.81 0.11

Directory: /lfs1/projects/jetmgmt DiskInUse=8111 GB, Quota=15000 GB
Directory: /lfs2/projects/jetmgmt DiskInUse=148894 GB, Quota=300000 GB
Directory: /pan2/projects/jetmgmt DiskInUse=250713 GB, Quota=668609 GB

Allocation Terms:
Name -- Name of allocated project
Amount -- Total Allocation
Reserved -- Reserved requests from current jobs in queue.
Balance -- Total allocation remaining for new jobs.
Available -- Amount currently available for new jobs.
Initial Allocation -- Current Month's Allocation.

All amounts are in core-hours.

Note: If the allocation of the disk space seem unusually large,
it is most likely because no allocation is defined for your project.
```

Partition Access

EMP	Partition List
HFIP	njet, tjet, ujet, sjet
RDHPCS	njet
FAA	nfaa
-- All of the Above --	service,tgpu

Notes:

- The ncomp partition is going away on the downtime on January 29th. RDHPCS users, please change your requests to njet at that time.
- All projects have access to the partitions listed in “All of the Above”.

Partition Access Best Practice

- For RDHPCS, just use njet
- For HFIP, it is more complicated
 - Three different system performance characteristics
 - The sjet system is drastically different in performance
 - The sjet system also requires different compiler options for the best performance
- It is possible to select all systems for the most flexibility
 - `-lpartition=njet:tjet:ujet:sjet`
- There are several problems with this
 - Code will not have been compiled as efficiently as it could for sjet, wasting resources
 - The walltime of your job will drastically change, you will have to pick the longest time possible which will cause scheduling difficulties on the faster system
 - Not all nodes have the same number of cores, and cores may be wasted based on your procs request.

Partition Access Best Practice

- We recommend that you use either:
 - `-lpartition=njet:tjet:ujet`
- OR:
 - `-lpartition=sjet`
- The performance difference between njet and tjet/ujet is small so there is not as much difficulty with specifying accurate wallclock times.

Windfall

- Windfall queue is a zero-charge queue.
- Jobs in the windfall queue run at the lowest priority.
- Jobs are automatically assigned to the windfall queue if their project does not have enough allocation left to run.
- It is NOT possible to submit to the windfall queue directly.
 - If you have allocation, you are required to use it first.

Queues and Priorities

Queue Name	Priority	Description
batch	10000	Default queue for all jobs
debug	1000000	High priority queue for debugging, maximum walltime is 30 minutes
urgent	100000	High priority queue for quick turn around time, maximum of one running job per project
novel	10000	Queue for jobs that require entire partition
service	10000	Access to service nodes
windfall	1	Zero charge queue to allow jobs to run after project allocations are exhausted. This queue is not directly accessible.

Scheduling Fairness

- Job scheduling is based on job priority
- Job priority is function of several different parameters
 - Base priority from queue
 - Time in queue (increases by one per minute)
 - Weighting based on past system usage
- Function will be tweaked over time to improve fairness
- When reservations are needed, new project will be defined (rt-) and allocations will be split
 - Accounts are debited for reservations even if they are not used
- We will NOT initially implement preemption for jobs in windfall. But we may in the future with little warning to improve fairness.

Changes in Session Timeouts

- NOAA policy requires that all ssh connections timeout on some intervals
- We have chosen to implement this by dropping connections Sunday morning at 2am local mountain time.
- We chose this as a straight-forward way to implement this function that would minimize impacts to working users.

Per-Job Memory Statistics

- We have installed developed new tool called cmem, which is run when your jobs exit
 - The tool can also be run during a job. See “cmem -h” for more information
- Per node memory usage is provided
- When determining how much memory a serial job uses, this is much better than the old method. Documentation has been updated.

https://jetdocs.rdhpcs.noaa.gov/wiki/index.php/Running_and_monitoring_Jobs#Submitting_a_serial_job

Per-Job Memory Statistics

Peak memory usage summary:

```
min = 1994848 KB
ave = 2024350 KB
max = 2063124 KB
```

All nodes sorted by nodelist order: (memory usage in KB)

Node	% of limit	user max	user limit	user current	total current	total phys
u495	8.7	2063124	23705600	1240	710176	24730116
u484	8.6	2046936	23705600	1304	685776	24730116
u485	8.7	2051152	23705600	1584	666108	24730116
u487	8.6	2038496	23705600	1416	683664	24730116
u488	8.4	1994848	23705600	1784	684356	24730116
u489	8.4	2000556	23705600	1468	685700	24730116
u493	8.5	2004288	23705600	1524	679168	24730116
u476	8.4	1995400	23705600	1524	674852	24730116

Upcoming Downtimes

- Jan 29th
 - Turn on allocation process
 - Remove ncomp partition
 - Implement maximum ssh session
- Mid to Late March
 - The IP addresses for all of the external Jet machines will be changing.
 - Any users who have firewalls on their local systems to allow for access MUST have them updated.
 - We will be sending emails in the next couple of weeks as to what changes local admins must make.

Open Forum

Any Questions?