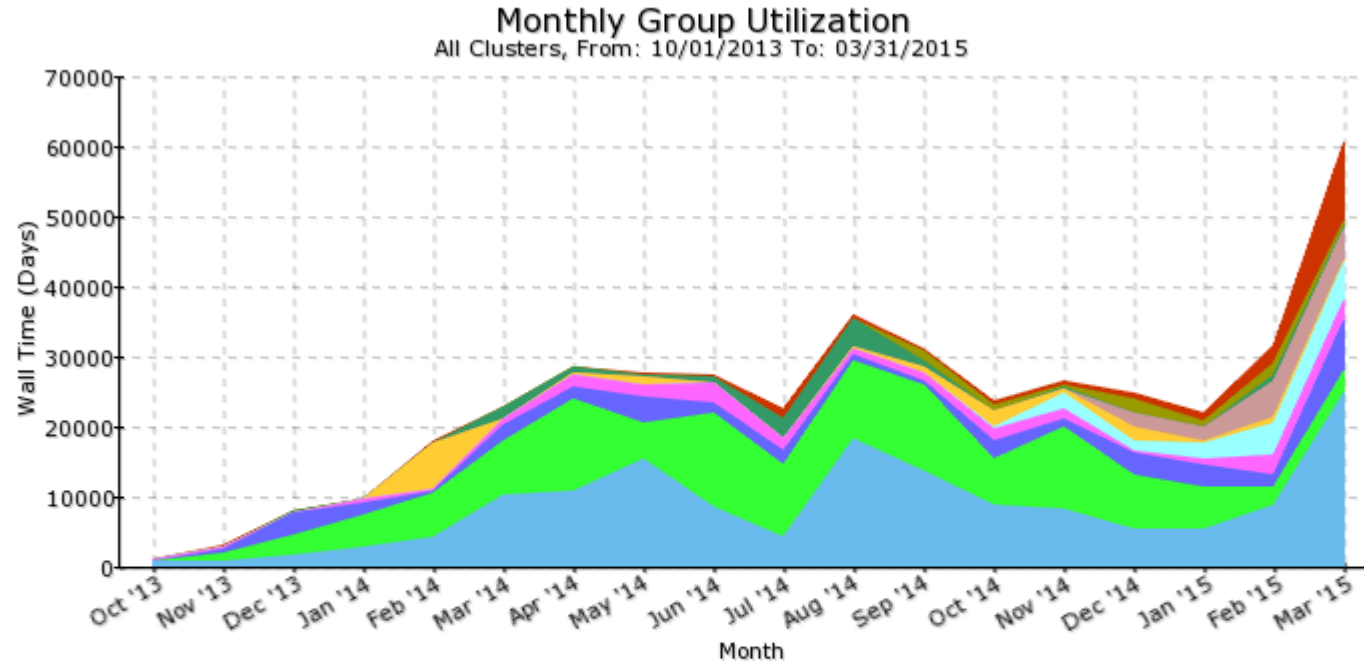# Yeti Operations Committee

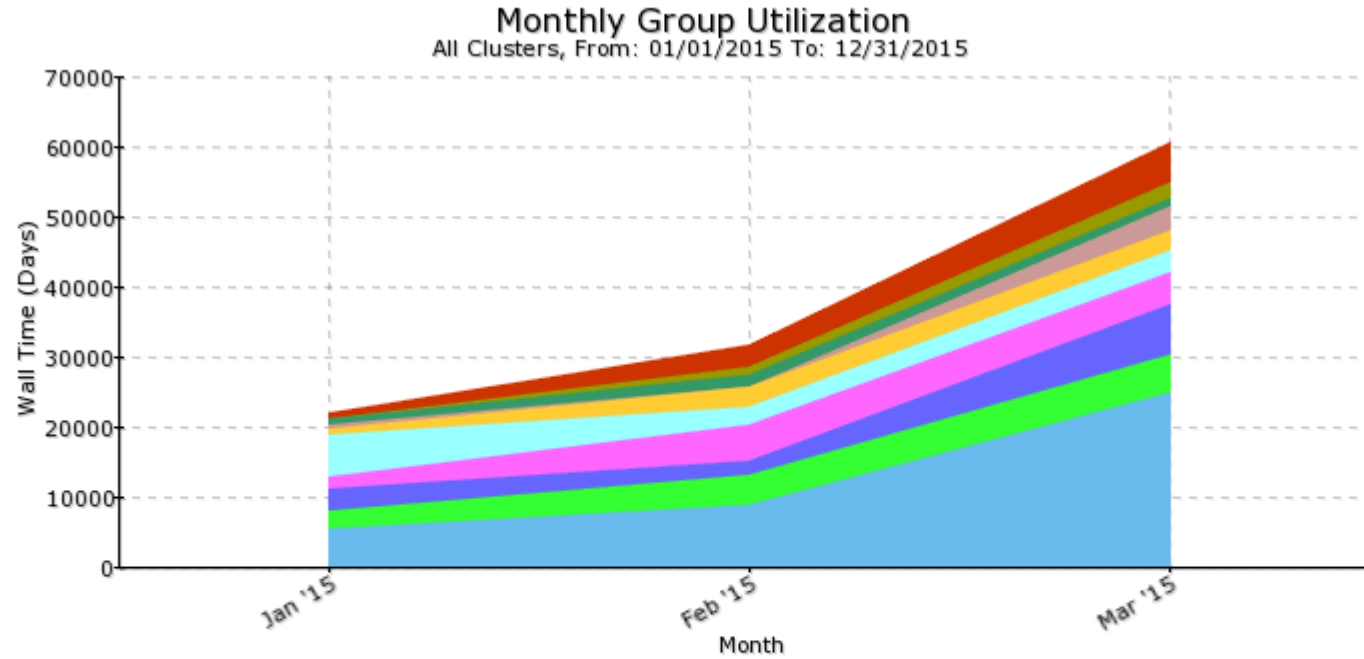APRIL 15 2015 MEETING

Topics

1. Usage Report
2. Free Tier, Education Tier, Publications
3. Operations Overview
4. Operations Discussion
5. Executive Session

Monthly Wall Time Since Launch

Monthly Walltime 2015

Monthly Group Utilization
All Clusters, From: 01/01/2015 To: 12/31/2015

Overall Statistics
2015

Users	164
Groups	24
Total Jobs	1,287,082
Average Job Size	1.5 cores
Average Wall Time	0.1 days
Average Job Size	1.2 nodes
Average Wait Time	2.2 hours
Average Execution Time	4.5 hours

Queue Statistics
2015

Queue	Jobs	Avg. Cores	Avg. Wait (h)	Wall Time (d)	Avg. Mem (M)
batch1	839,618	1.4	2.0	28,529	138
batch2	208,466	1.4	0.7	8,656	2,115
batch2a	3,751	15.4	6.4	4,797	7,468
batch3	220,299	1.2	4.3	38,581	944
batch4	12,460	8.0	7.3	31,068	8,346
interactive	2,411	4.2	0.1	692	4,268
interlong	77	10	0.8	352	1,561

Queue Statistics
March 18 - 31

Queue	Jobs	Avg. Cores	Avg. Wait (h)	Wall Time (d)	Avg. Mem (M)
batch1	195,872	1.9	3.4	9,339	205
batch2	16,153	1.8	0.9	1,461	3,060
batch2a	3,227	10.6	7.1	3,558	7,793
batch3	37,819	1.1	6.5	10,835	1,383
batch4	1,918	6.1	13.2	5,710	7,289
interactive	525	2.0	0.0	102	6,269
interlong	24	10.8	1.3	89	2,031

Free Tier

- Eligibility

    - Any Columbia faculty or researcher

    - Anyone (with Columbia UNI) with faculty or researcher endorsement

    - Reinstate 16:32 core maximum?

Free Tier

Setting	Free Tier User	Renter
Fairshare (Servers)	1/64th	1/16th
Quota	16 GB	64 GB
Core Maximum (Soft)	4	n/a
Core Maximum (Hard)	8	512

Education Tier

- SRCPAC Education Subcommittee

- Support for classes and workshops on Yeti

- Will support at least one class this fall

- Contact hpc-support to discuss a class

Publications

- Starting a publications page

- Will be soliciting contributions at regular intervals

- Twice a year?

- Hotfoot's page: https://wikis.cuit.columbia.edu/confluence/display/rcs/Research+Products

Operations Overview

- Background for policy discussion

- Topics
  - Fairshare

  - Job Priority

  - Job Queues

  - User Limits

Operations Overview - Fairshare

- Every group has a target share

- Based on servers purchased

- Example: Astro share is 4.55%

Operations Overview - Fairshare

- Recent usage tracked

- Yeti configured to look back one week

- Actual usage is compared to fairshare target

- Usage measured in core hours

Operations Overview – Fairshare

- Astro target fairshare: 4.55 %

- Astro actual current share: 5.70 %

- Therefore Astro jobs waiting in queue have priority lowered.

Operations Overview - Job Queues

Original Configuration

Queue	Time Limit	Memory Limit	Max. User Run
Batch 1	12 hours	4 GB	512
Batch 2	12 hours	16 GB	128
Batch 3	5 days	16 GB	64
Batch 4	3 days	None	8
Interactive	4 hours	None	4
Special Request	Varies	Varies	Varies

Operations Overview - Job Queues

Batch 2a Added

Queue	Time Limit	Memory Limit	Max. User Run
Batch 1	12 hours	4 GB	512
Batch 2	12 hours	16 GB	128
Batch 2a	12 hours	None	8
Batch 3	5 days	16 GB	64
Batch 4	3 days	None	8
Interactive	4 hours	None	4
Special Request	Varies	Varies	Varies

Operations Overview – User Limits

- Job Limit

  - 256 soft
  - 512 hard

- Core Limit

  - 512 hard

Operations Overview – Guaranteed Access

- Groups have priority access to the systems they purchased

- Job walltime must be < 12 hours to run on somebody else's system

Operations Discussion – Infiniband Job Queue

- Job queue for MPI jobs

- Will only use Infiniband nodes

- Other queues lose access to Infiniband nodes

- Possible exceptions
  - Very short jobs (1-2 hours?)
  - Preemptable jobs

Operations Discussion – Infiniband Job Queue

Proposed Settings

- Maximum wall time 36 hours

- Full nodes only

- No memory limit

- Max user run: 8?

Operations Discussion – Possible Exceptions

- Very Short Queue

  - Maximum walltime of 2 hours

- Preemptable Queue

  - Jobs can be interrupted at any time

Operations Discussion – Other Job Queues

- GPU

  - Similar to Infiniband Queue

- Admin

  - Administrator use only

Operations Overview - Job Queues

Possible Configuration

Queue	Time Limit	Memory Limit	Max. User Run
Batch 1	12 hours	4 GB	512
Batch 2	12 hours	16 GB	128
Batch 3	5 days	16 GB	64
Batch 4	3 days	None	8
Interactive	4 hours	None	4
Infiniband	36 hours	None	8
GPU	3 days	None	4
Special Request	Varies	Varies	Varies

Operations Overview - Job Queues

Possible Configuration

Queue	Time Limit	Memory Limit	Max. User Run
Batch 0	2 hours	4 GB	512
Batch 1	12 hours	4 GB	512
Batch 2	12 hours	16 GB	128
Batch 3	5 days	16 GB	64
Batch 4	3 days	None	8
Interactive	4 hours	None	4
Infiniband	36 hours	None	8
GPU	3 days	None	4
Special Request	Varies	Varies	Varies

Operations Overview - Job Queues

Possible Configuration

Queue	Time Limit	Memory Limit	Max. User Run
Preemptable	5 days	None	512
Batch 1	12 hours	4 GB	512
Batch 2	12 hours	16 GB	128
Batch 3	5 days	16 GB	64
Batch 4	3 days	None	8
Interactive	4 hours	None	4
Infiniband	36 hours	None	8
GPU	3 days	None	4
Special Request	Varies	Varies	Varies

Operations Discussion – Special Request Queues

- Users occasionally request special treatment

- Typical example: longer wall times

- Requests should have a time limit

- Support will forward requests to Yeti OC Chair for approval

Operations Discussion – Special Request Queues

- We have one now: Interlong

- Interactive queue with 96 hour walltime

- Needs to be formally approved

Operations Discussion – Requested Topics

1. Further explanation of fairshare
2. Remove stack limit
3. Removal of memory specification
4. Job submission from execute nodes
5. Interactive visualization on GPU nodes
6. Improved package maintenance (ROOT, Geant4, COMSOL)
7. Install GRID tools
8. Allow > 512 cores for jobs < 6 hours

Operations Discussion – Requested Topics

1. Update Yeti documentation
2. Matlab 2014+ parallelization changes
3. More environment modules
4. Retention of interlong queue

Executive Session

(Time Permitting)

Addendum - Yeti Execute Nodes

Round	Node Type	Number of Nodes	Percent of Cluster
1	Standard (64 GB)	38	22.8
	Intermediate (128 GB)	8	4.8
	High Memory (256 GB)	35	21.0
	Infiniband	16	9.6
	GPU	4	2.4
2	Standard (64 GB)	10	6.0
	High Memory (256 GB)	4	2.4
	Infiniband	48	28.7
	GPU	5	3.0
	Total	**167**	

Addendum – Group Shares (Includes Grant)

Group	Systems	Share
afsis	1.2	0.72%
astro	3.6	2.16%
ccls	11	6.59%
eeeng	1.2	0.72%
jrn	1.2	0.72%
ocp	6	3.59%
psych	1.2	0.72%
sscc	10.8	6.47%
stats	19.2	11.50%
xenon	1.2	0.72%
apam	10	5.99%
astro	2	1.20%
brain	1	0.60%
cheme	8	4.79%
cmt	2	1.20%
comb	1	0.60%
dsi	8	4.79%
dvmm	2	1.20%
hblab	1	0.60%
heat	1	0.60%
physo	17	10.18%
qmech	10	5.99%
sfil	1	0.60%
stock	1	0.60%
zmbbi	1	0.60%
grant	44.4	26.59%
total	167	100.00%