The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization

IAEA International Atomic Energy Agency

2494-23

Workshop on High Performance Computing (HPC) Architecture and Applications in the ICTP

14 - 25 October 2013

Sustainable Management of HPC Resources

Axel Kohlmeyer

ICTP, Trieste/Temple University
Philadelpia
USA

Sustainable Management of HPC Resources

Dr. Axel Kohlmeyer

Sr. Scientific Computing Expert
ICTP, Trieste

Associate Dean for Scientific Computing
Temple University, Philadelphia

http://sites.google.com/site/akohlmey/

The Abdus Salam
International Centre
for Theoretical Physics

Owls' Nest @ Temple University

185 Nodes, >2300 CPU cores, 8x15TB = 120 TB Raid-5, 4.5 TB RAM
32x Nvidia Tesla C2050, 4x AMD FireProV8800, 2x Nvidia GTX 580
~45 research groups, 150 total active users, ~30 active users each week

Section / Queue	Vendor / Model	# Nodes	Processors / Node	Cores / Node	Memory / Node	Interconnect	Local Disk	TFLOP/s
normal	Dell R610	72	2x Intel Xeon X5660, 2.8GHz	12	12GB	QDR Infiniband	120GB	9.7
manycore	Dell R815	10	4x AMD Opteron 6174, 2.2GHz	48	64GB	QDQ Infiniband	120GB	4.2
highmem	Dell R610	8	2x Intel Xeon X5677, 3.5GHz	8	96GB	QDR Infiniband	900GB	0.9
bigmem	Supermicro H8QG6	2	3x AMD Opteron 6238, 2.6GHz	36	384GB	3x GigE	660GB (SSD)	0.75
legacy	Dell PE1950	72	2x Intel Xeon E5430, 2.7GHz	8	8GB	DDR Infiniband	40GB	6.1
devel	Dell R610	4	2x Intel Xeon X5550, 2.7GHz	8	24GB	QDR Infiniband	120GB	0.34
gpu	Supermicro X8DTG-QF	8	2x Intel Xeon E5630, 2.5GHz, 4x Nvidia Tesla C2050	8, 1792	24GB, 4x 3GB	QDR Infiniband	1500GB	17.1
ati	Supermicro X8DTG-QF	1	2x Intel Xeon E5520, 2.3GHz, 4x AMD FirePro3D V8800	8, 6400	48GB, 4x 1GB	GigE	1500GB	10.4
Login Servers	**Vendor / Model**	**# Nodes**	**Processors / Node**	**Cores / Node**	**Memory / Node**	**Interconnect**	**Local Disk**	**TFLOP/s**
login3/4	Supermicro H8QG6	2	1/2x AMD Opteron 6238, 2.6GHz	6	64GB	3x GigE	660GB (SSD)	0.13
File Servers	**Vendor / Model**	**# Nodes**	**Processors / Node**	**Cores / Node**	**Memory / Node**	**Interconnect**	**Local Disk**	**TFLOP/s**
st01/02/03/04	Supermicro X8DTU	4	2x Intel Xeon E5620, 2.4GHz	16 (HT)	24GB	4x GigE	2x 15TB	n/a
Management Servers	**Vendor / Model**	**# Nodes**	**Processors / Node**	**Cores / Node**	**Memory / Node**	**Interconnect**	**Local Disk**	**TFLOP/s**
admin	Dell R610	4	2x Intel Xeon X5550, 2.8GHz	24 (HT)	12GB	GigE	45GB	n/a
login1/2	Dell R610	2	2x Intel Xeon X5660, 2.8GHz	24 (HT)	24GB	2x GigE	120GB	n/a

HPC Staff and IT-Support

- University operates the data center
  -> extremely reliable power & cooling
- University staff & student helpers at hand for replacing hardware, cabling and related tasks
- Pool of spare parts (HDs, DIMMs, IB-HCAs)
- Special help recruited for the initial installation
- Only 1 full-time HPC staff: initially A.K., since 07/2012 E.B. with research faculty appointment
  -> Cluster and account management, user support, HPC training, software optimization

The Abdus Salam International Centre for Theoretical Physics

General Concept

- “Cluster of Clusters” with multiple login nodes
- Identical software environment on all nodes
- Dedicated management node (software server)
- Home dirs distributed across 4 NFS servers
- Channel bonding for NFS & login nodes
- Infiniband (DDR or QDR) for MPI only
- Computing via batch only (Torque/Maui)
- 1 queue per hardware configuration

Hardware Selection

- Well balanced components, avoid extremes (e.g. not max. TDP, not max. CPU Clock, not max. RAM, not max. Number of CPU cores)
- Vendor choice not only based on price, but:
  => ease of remote management from linux
  => ease of replacing broken hardware
  => familiarity/satisfaction of IT staff with vendor
  => track record of vendor in specific hardware
- Bulk of cluster optimized for largest users
- Redundancy only for infrastructure hardware

Upgrade Path

- Start new cluster from scratch with updated software environment (e.g. RHEL 5 -> RHEL 6) and new home / data directory storage
- Move active users to new machine (in stages)
- When in full production, consolidate old cluster
  => integrate still useful components
  => select some unused components as spares
  => discard reset of use for training purposes
- Expand with additional hardware for ~2 years
  => then start next cluster

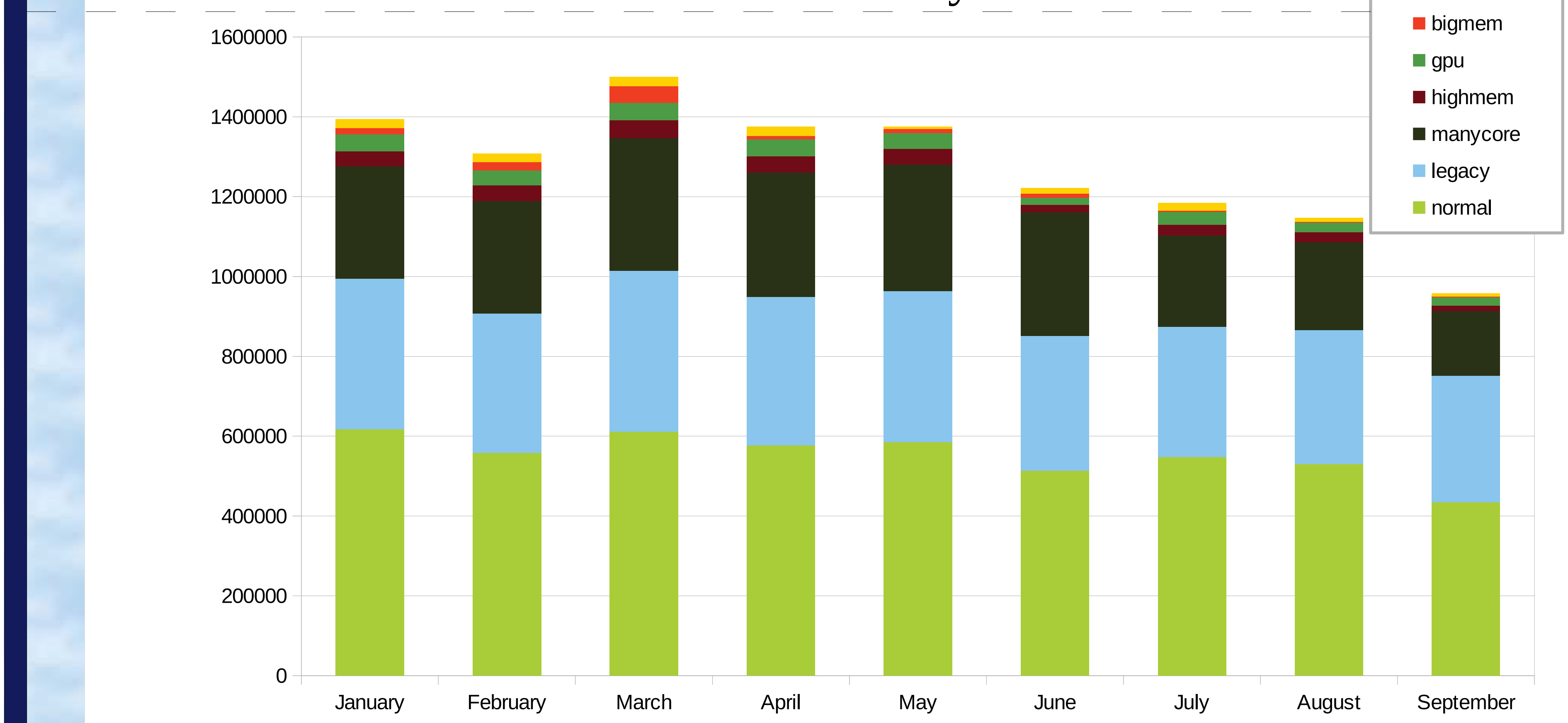
Monitoring / Health Check

- Use ganglia for overall cluster status monitoring
  => look for fingerprints of pending problems
  => investigate unusual use indicators
  => contact individual users about issues
- Use pre- and post-job health check script
  => search for indicators of problems by parsing log files and running test tools
  => in case of problem, node sets itself offline
  => investigate causes of node problems
- Keep detailed incident record (trouble log)

The Abdus Salam
International Centre
for Theoretical Physics

Service Units Delivered in 2013

100% = 1.63 SU/month => 20M SU / year



The Abdus Salam
International Centre
for Theoretical Physics

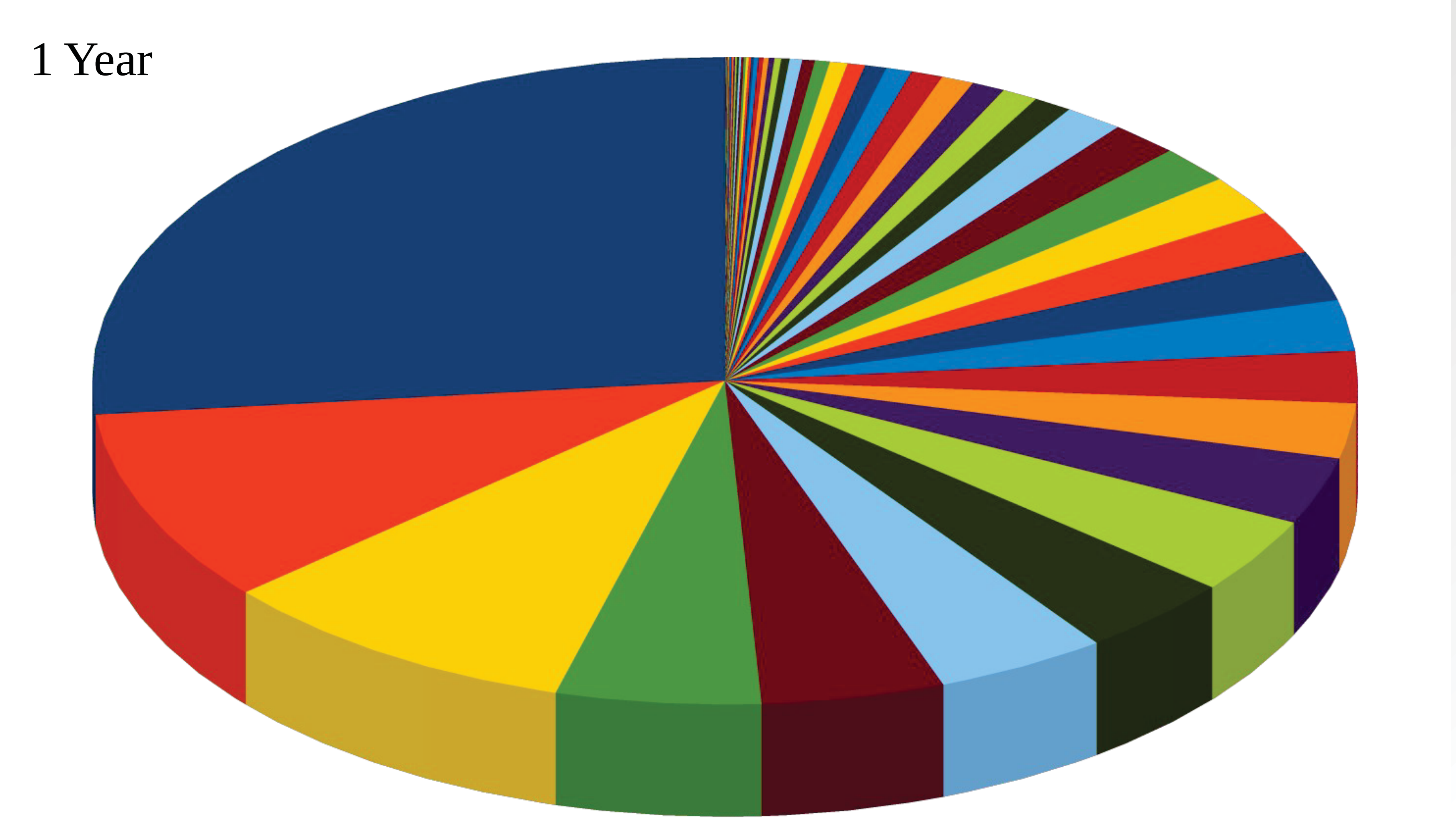
Consolidated Software Environment

- Use rpm and modules for software deployment
- Modules tree on read-only NFS export
- System software via local RPM repository
- Compile all applications to work on all nodes
- Reduce choice of alternatives where possible
- Reduce number of installed software packages
- Development and GUI tools only on login nodes
- Maintain internal wiki to document all installs

User Management

- Custom script to account generation
  => enforces consistency
- No local password management:
  - Access through university credentials via LDAP
  - Discretionary accounts via ssh keys only
- Monitor new users, advise on bad practices
- Disseminate user's guide via web page with examples for all important steps
- Update as needed based on usage patterns

Typical Active User Distribution

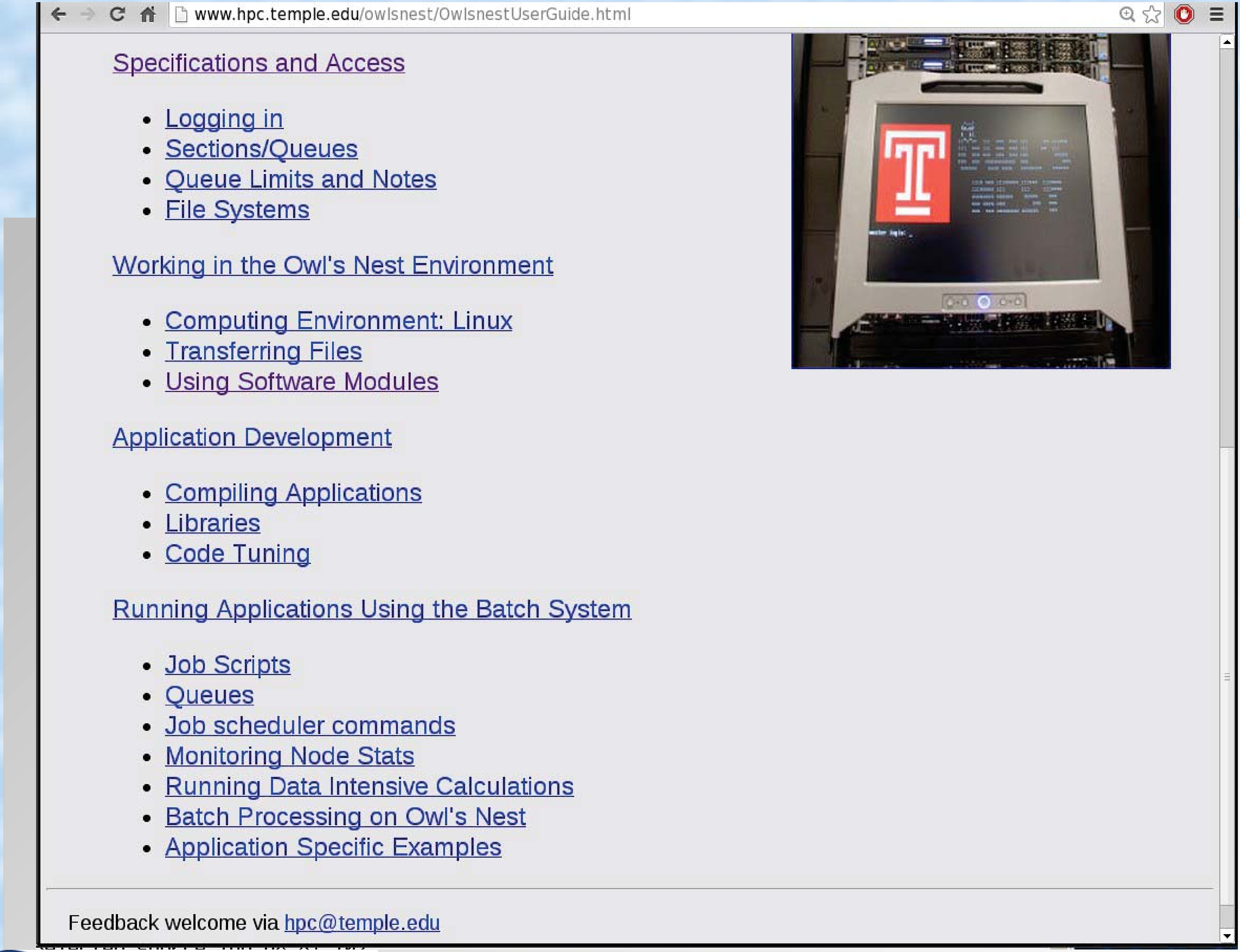


Batch Configuration

- One queue per hardware type
- Exclusive node access via Maui QoS policy
  => login to nodes via pbs_simpleauth.so only
- Wall time limits (12h/24h/48h/96h)
  shorter limits for queues with more nodes
- Max. # of jobs limit, max. # of nodes/job limit
  => adjusted based on usage patterns
- Fair share job priorities, expansion factor, no job age priority, aggressive backfilling

General Management Strategy

- Configure for automatic graceful degradation
- Use remote management wherever possible
- Use low-tech text mode tools, custom scripts
- Document all configuration choices in wiki
- Keep all machines as consistent as possible => c3 tools, rpm repository, read-only NFS
- Keep configuration simple, handle exceptions (e.g. walltime extension) on case-by-case basis
- Communication via hpc@temple.edu alias



Specifications and Access

- Logging in
- Sections/Queues
- Queue Limits and Notes
- File Systems

Working in the Owl's Nest Environment

- Computing Environment: Linux
- Transferring Files
- Using Software Modules

Application Development

- Compiling Applications
- Libraries
- Code Tuning

Running Applications Using the Batch System

- Job Scripts
- Queues
- Job scheduler commands
- Monitoring Node Stats
- Running Data Intensive Calculations
- Batch Processing on Owl's Nest
- Application Specific Examples

Feedback welcome via hpc@temple.edu