MIT Tier2 site report

Max Goncharov
March 3, 2009

USCMS Tier2 workshop

Hardware and infrastructure

- farm composition

- curent operations

- new location – Bates

- new infrastructure

- new network

- work in progress

CPU and batch slots

CPU power and batch slots:

Tier2 resources	Batch slots/CPU power	1116 / 1753 kSi
Other (CDF/HI)		530 / 432 kSi
Total		1646 / 2185 kSi

Storage figures:

Tier2 resources	309 TB
Other (CDF/HI)	34 TB
Total	343 TB

Join official dCache

Raids:

- continue using Raid5, with and without hot spare
- physical raw storage: 527 TB total

Farm hardware

Worker nodes:

2x4-core AMD 1.6-2.0 GHz 3.0 TB drive x70

2x4-core AMD 1.6 GHz 3.75 TB drive x55

2x2-core AMD 1.8 GHz 2.4 TB drive x20

Tier2 funded

2x2-core AMD 2.0 GHz 1.8 TB drive x30

2x1-core AMD 1.6 GHz x200

Other resources (CDF/CMS-HI)

Location: w91

- hosted by MIT IS&T, on campus
- network, racks, power, managed by MIT IS&T
- overseen 5x16 by operators
- generator/UPS backup for all servers (2 racks)
- 16 racks total size
- each rack – 38 U maximum
- running out of physical space

This year – moving to new location: Bates

Services Status

Software:

- all machines at SL4

OSG 1.0.0:

- reworking NFS mount scheme
- using NFS-lite

GUMS:

- currently running version 1.2.15 on a dedicated node

dCache:

- version 1.8, SRM version 2.2

Condor:

- version 7.2

PhEDEx:

- up2date, several new links established

New Location

Bates: former host to linear accelerator (40 miles from MIT)

- main building is completely refurbished

- 70 water cooled 47U racks, installed by May, CMS use 40 racks

- new networking (hardware, switches)

- new rack hardware

Resource Allocation

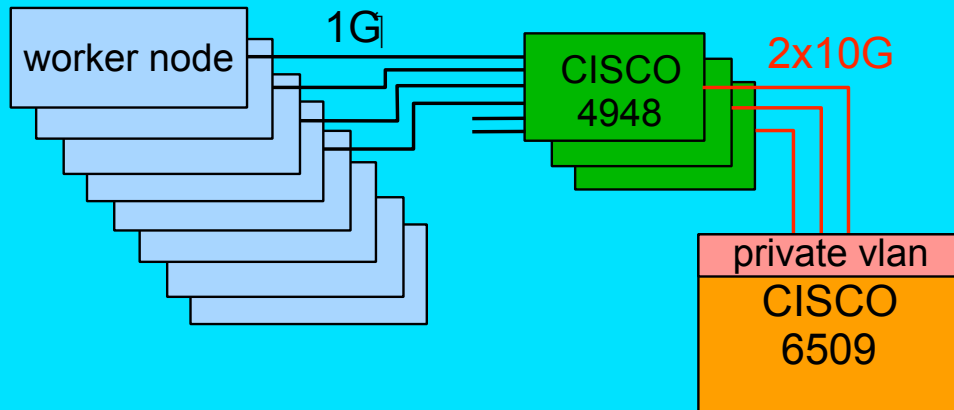| Total	250 K
network, power cables, …	50 K
networks switches	100 K
new servers	20 K
new worker nodes	80K (2U 8 core 5 TB)

In Bates we can easily triple in size.

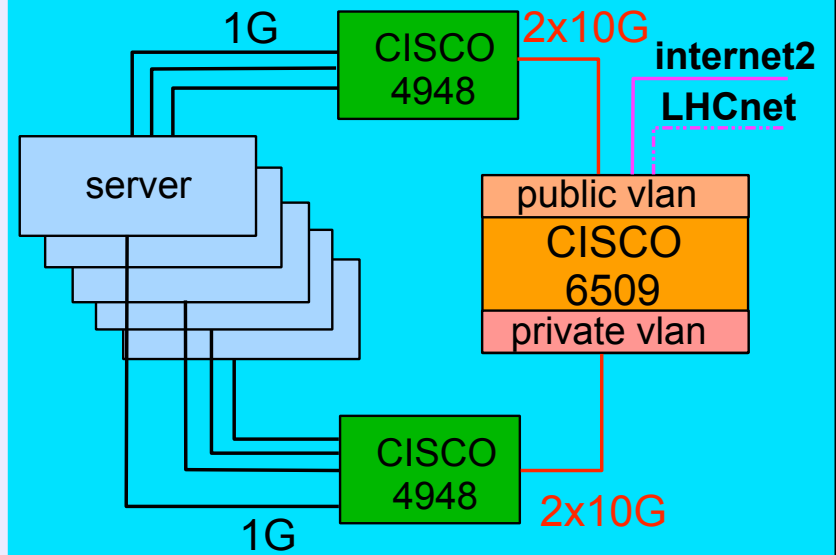Water cooled racks → increased reliability
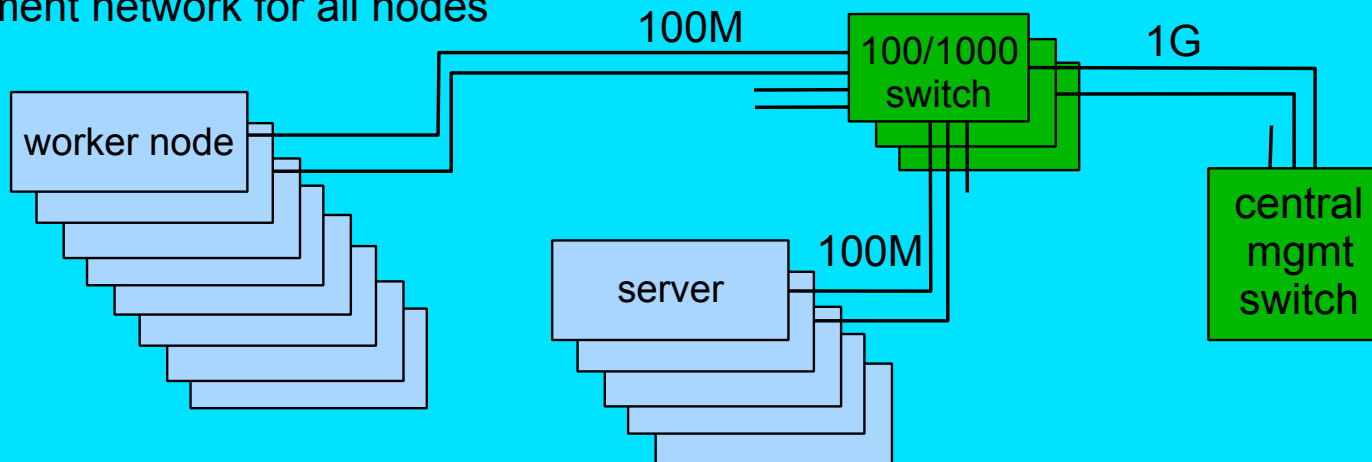
Network now

Network for worker nodes

worker node — 1G — CISCO 4948 — 2x10G — private vlan / CISCO 6509

Network provided and managed by MIT IS&T

Network for servers

server — 1G — CISCO 4948 — 2x10G — public vlan / CISCO 6509 / private vlan — internet2 / LHCnet — CISCO 4948 — 2x10G — 1G

Management network for all nodes

worker node — 100M — 100/1000 switch — 1G — central mgmt switch

server — 100M

Network to be

Network for worker nodes

worker node

1G(+1G)

public vlan
New T2 switch

10G

MIT switch

10G

MIT campus

LHCnet

internet2

server

Bates

10G

Management network for all nodes

worker node

100M

100/1000 switch

1G

100M

server

central mgmt switch
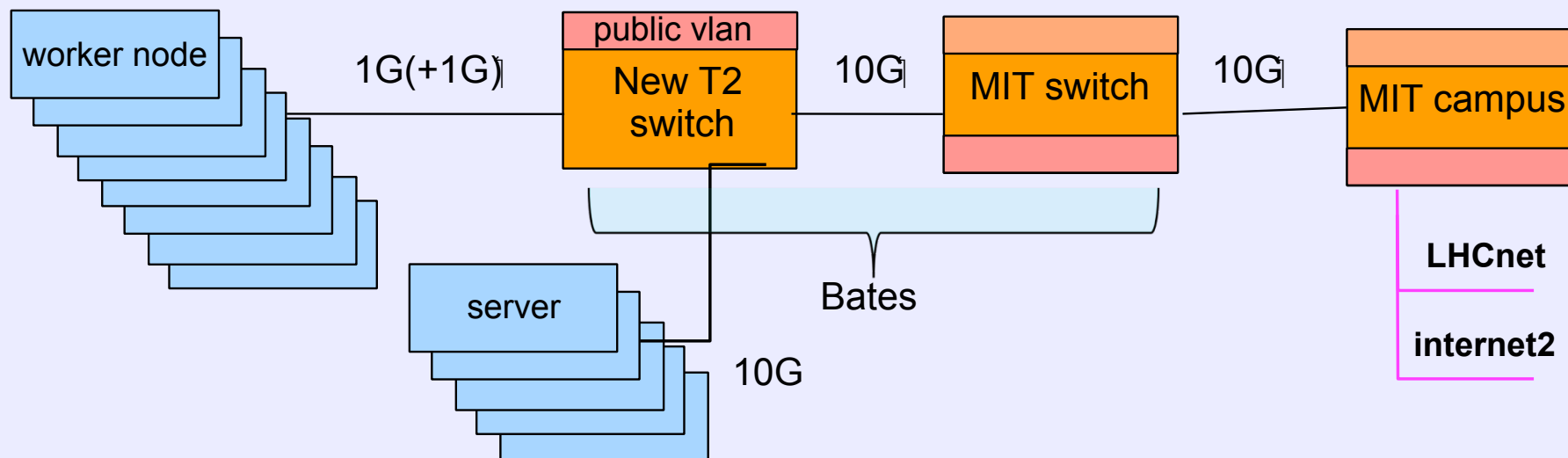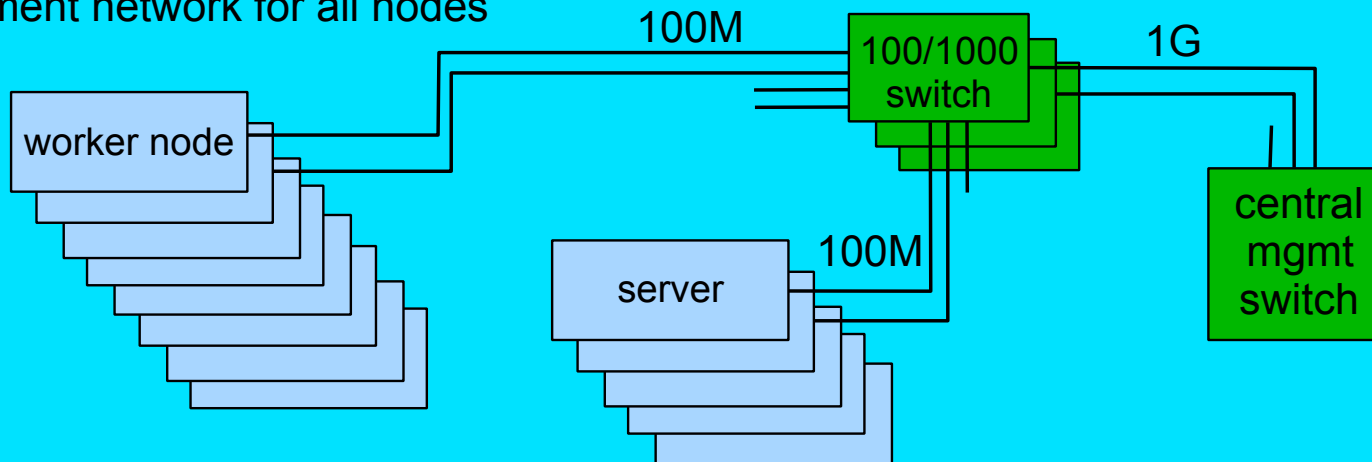
gridftp doors (theoretically, 10G possible)
real life transfers: ~2 Gbps observed with PhEDEx

WAN

admin

LAN

pool

SRM

1G — door 1 — 1G

3G — door 2 — 3G

GUMS

10G — door 3 — 10G
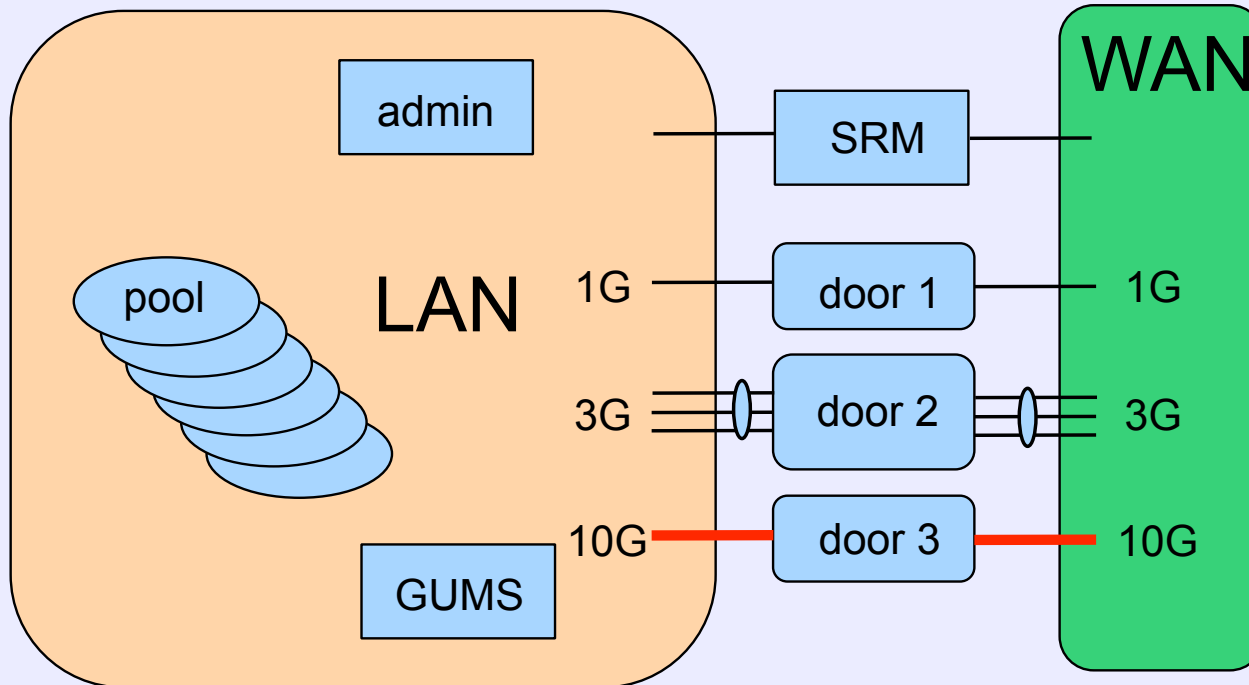
Network is a big challenge

- need to buy switches, set up cabling

- set up network monitoring and debugging

- switch from private to public network

would like to thank experts at Fermilab for providing their expertise

Work in progress

Storage:
- replica manager
- more pools

Network:
- set it up in Bates

OSG/dCache/Condor/PhEDEx:
- reworking the setup to become more reliable and not depend on NFS

Administration:
- OS upgrades
- many improvements scheduled in reliability and manageability

Personnel:
- we have completely new crew: Max Goncharov, Wei Li
- several students: Burak Alver, Michael Betancourt, Josh Bendavid, Steve Jaditz
- Started hiring process for mid-level system administartor

Would like to thank Fermilab, especially Burt Holzman and Paul Rossman for helping with the transition

Conclusions

- steady development, up to speed with milestones

- successful operation
  - good availability of storage and CPU to the grid

- at full capacity by LHC startup