MIT Tier2/Tier3 Clusters

Max Goncharov
Jan 26, 2022

CMS Computing

CMS collaboration:

- 7 Tier1 (T1) centers
- ~50 Tier2 (T2) centers

In US:

- 1 T1 center
  - Fermi National Lab
- 8 T2 centers
  - Caltech
  - MIT
  - Purdue University
  - University of California (San Diego)
  - University of Florida
  - University of Nebraska
  - University of Wisconsin
  - Vanderbilt

T1 and T2 Clusters

T2 cluster

- worker nodes to run user applications
- mass storage
- infrastructure for data transfers

T3 cluster

- small version of T2
- does not get tested as much at T2
- often customized for research groups

+

T1 : T2 cluster plus tape storage

Cluster Configuration

CPU/Storage Mix Model

Worker Nodes (WNs)
~22000 cores, ~750 servers

Storage
16.5 PB
(resilience through double replication)

Condor
batch scheduling framework
RAM - 2GB/job; Disk - 20GB/job

Hadoop (HDFS)
mass storage

Hardware Overlap – 99%

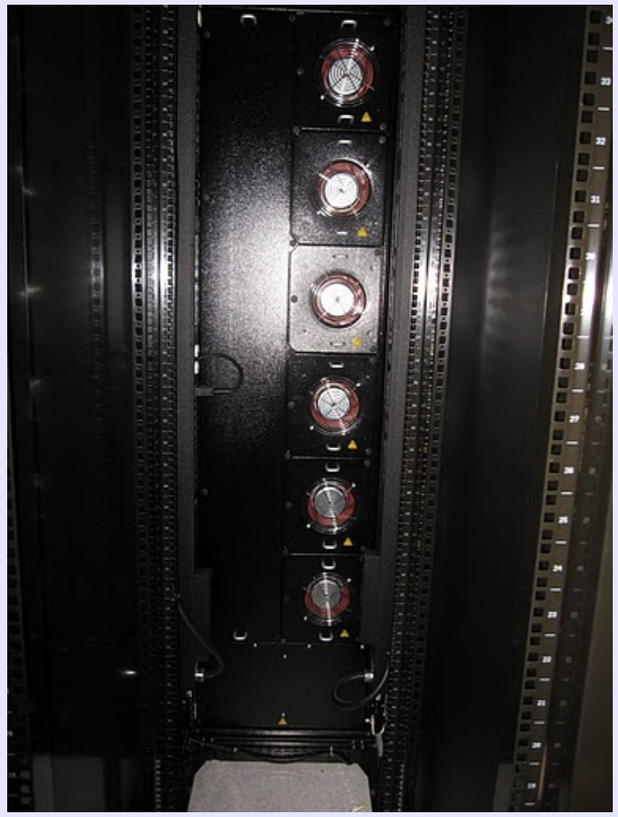
Infrastructure

Water Cooling

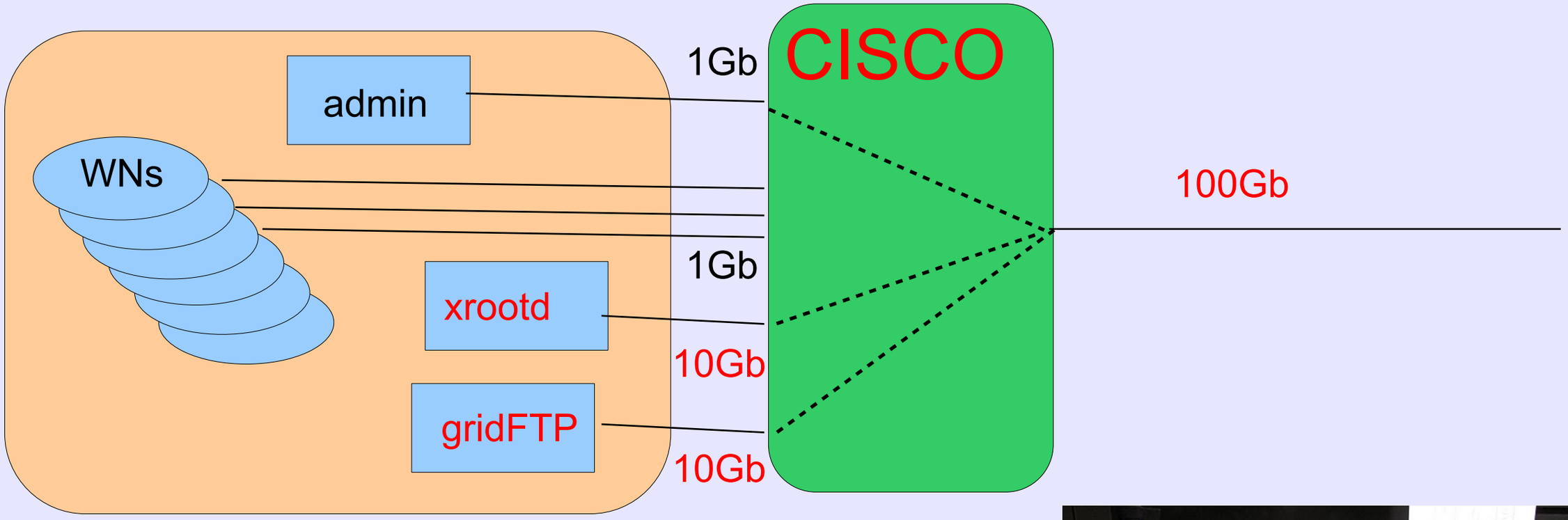
- infrastructure on the ground floor
- interlock in case of significant leak

Water Cooled Racks

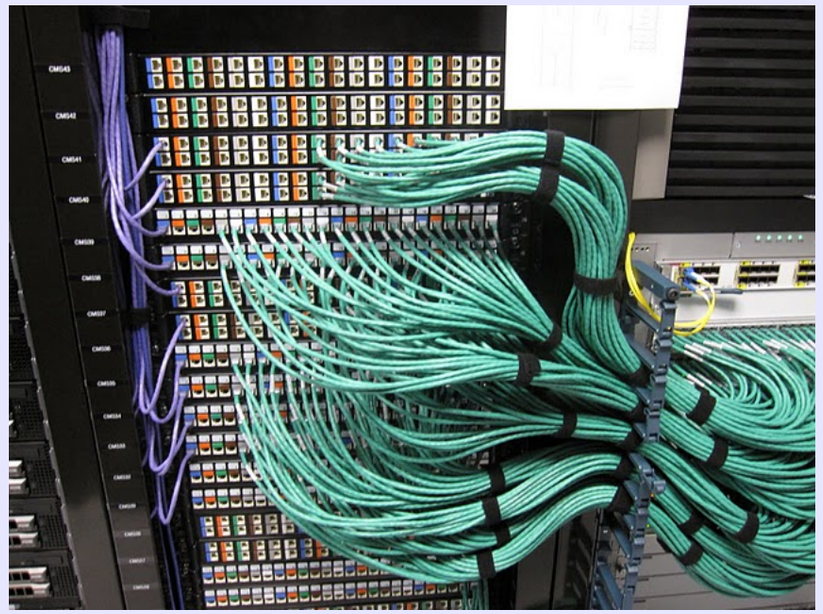
- 47 U, 10 kW of power
- 90% filled, room for expansion
- 4 UPS racks for central servers
- UPS support is on the ground floor



Network



- Network managed by MIT IS&T
- 100 Gb Fiber link to the outside world (no ipV6)
- CISCO handles all communications
- Machines can talk at 1Gb through copper links
- xrootd/gridFTPs are on 10 Gb links



Current Projects

Attach tape storage to T2

- 15 MW $90M single purpose data center
- Near zero Carbon footprint
- Space, power, and cooling for 780 racks
- More than 300,000 cores, thousands of GPUs
- 100 Gb multi-fiber ring to Boston, NYC and Albany
- Three new top500 in the past year
- Located in Holyoke, MA

MGHPCC

Boston University
Harvard University
MIT
Northeastern University
University of Massachusetts

Current Projects

Current mass storage (Hadoop) is becoming obsolete

- other technologies are available (CephFS)
- deploy erasure coding
  - now we use duplication for resiliency
  - with erasure coding we expand storage
  - … but because of our mixed hardware model not clear if it would work
  - spread storage at various physical locations (data lakes?)

-