Compact Muon Solenoid: Cyberinfrastructure Solutions

Ken Bloom

UNL Cyberinfrastructure Workshop -- August 15, 2005

Computing Demands

CMS must provide computing to handle huge data rates and sizes, and strenuous demands on processing. In 2007, expect

- 225 MB/s raw data rate to tape (>2 PB raw data/year)
- 40000 kSI2k CPU resources (Intel P4 = ~1 kSI2k)
- 14000 TB disk (Library of Congress = ~3000 TB)

Many different types of data processing required:

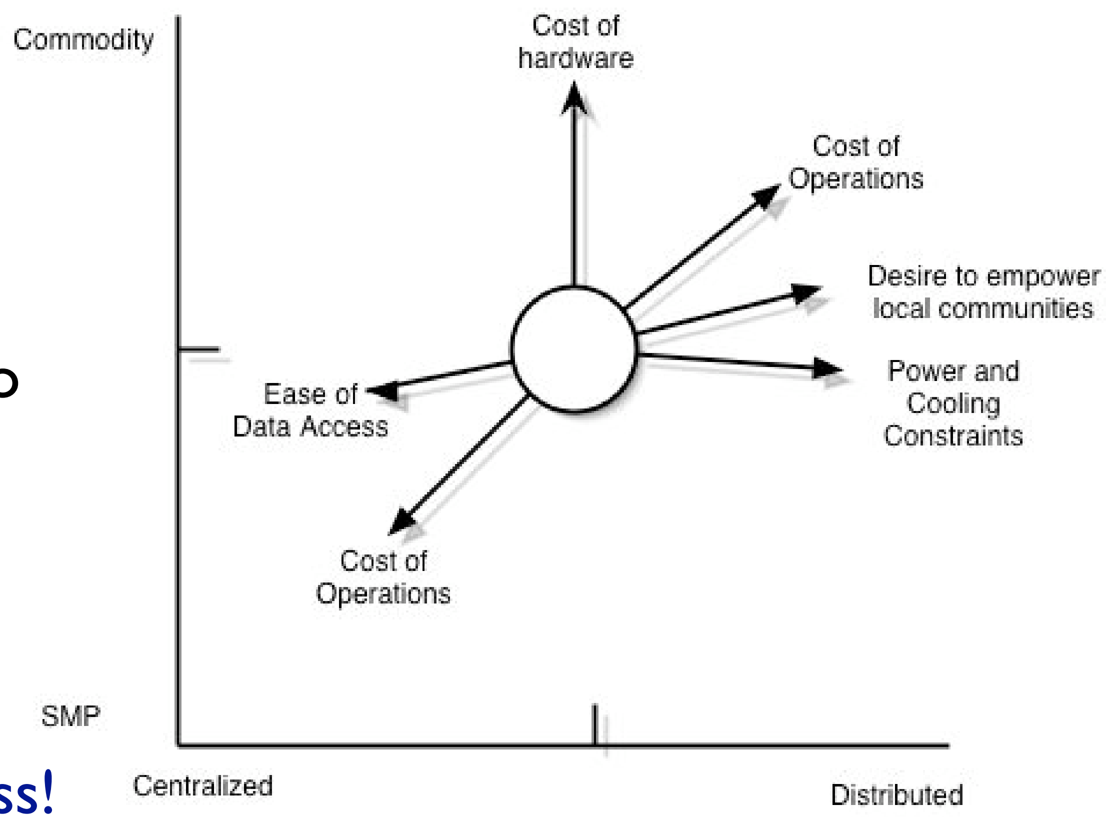
- Event reconstruction: CPU-intensive, large input, larger output
- First-pass event selection: Less CPU-intensive, large input, smaller output
- Data analysis: Moderately CPU-intensive, small input, small output
- Simulations: Extremely CPU-intensive, tiny input, large output

Historic computing solution: a large symmetric multiprocessor (SMP) computer, attached to large disk pool, located at a central facility

New Solutions for New Times

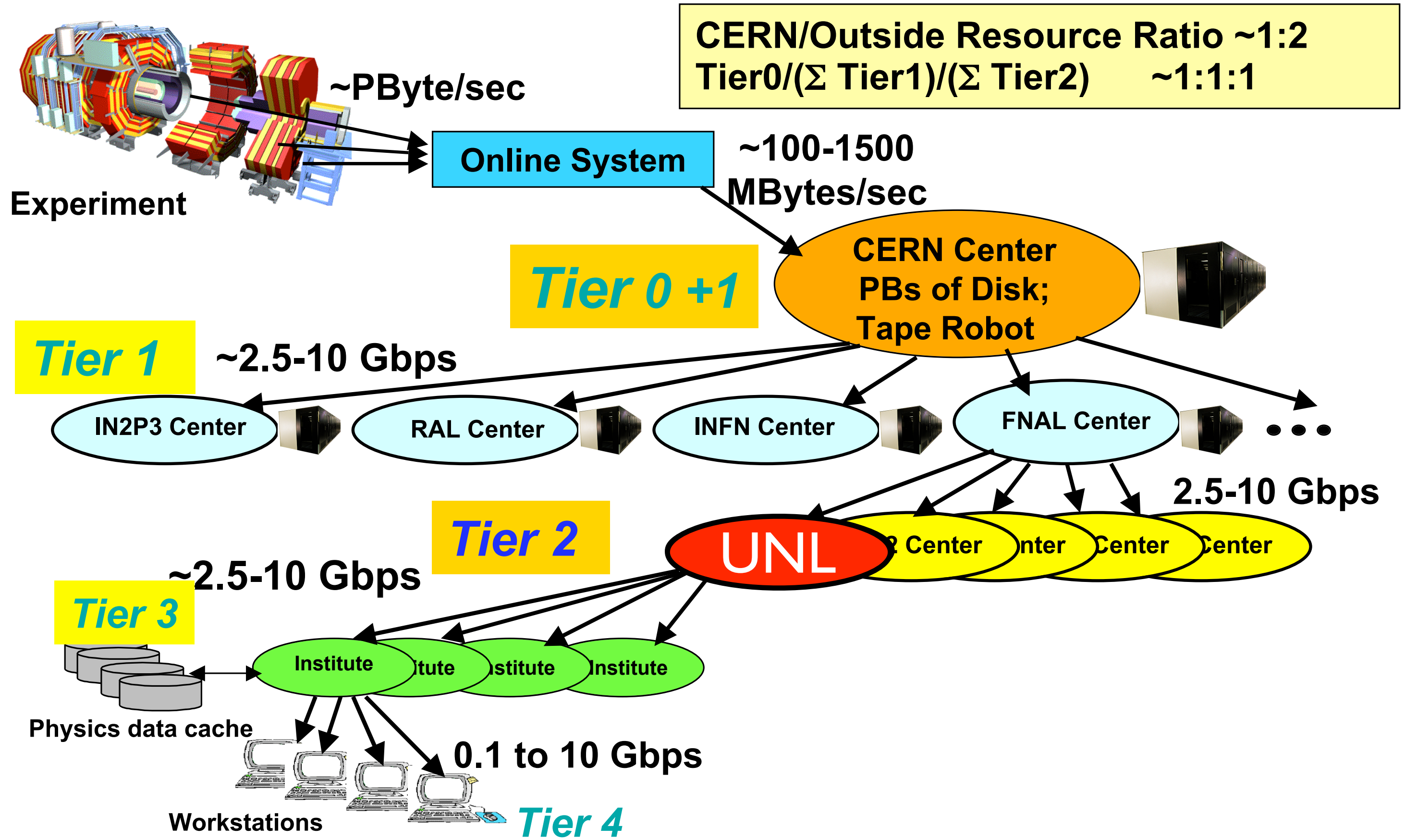
Centralized SMP good for ease of data access and operational economics. But:

- Physical infrastructure a precious resource -- power, cooling expensive
- Commodity machines with fewer CPU's a lot cheaper
- Operational expertise might be distributed
- Institutions and countries more willing to pay for local computing than central computing elsewhere
- Tendency to buy for peak load, but not always at peak -- good to share computing with scientists on different "schedules"
- Fast networks allow data and jobs to be shipped anywhere!



Suggests a different way to do business!

A "Tiered" Solution



Each tier has a different responsibility:

- Tier 0: first-pass reconstruction, write data to tape, send copies to Tier 1
- Tier 1: data reduction, skimming, bulk re-processing, archiving
- Tier 2: hosting of smaller samples, data analysis, simulations

T2 @ UNL

UNL bid to host T2 center built on local strengths:

- Strong Research Computing Facility
- Experience in cluster computing
- Excellent support from Office of Research
- Robust, growing particle physics research group; large CMS involvement

UNL selected as T2 site in January 2005, joining Caltech, MIT, UCSD, Wisconsin, Purdue, Florida; development and operations begin May 2005

- Completed first round of hardware purchases
  - CPUs are dual-core Opteron chips, leading-edge technology
- Installed and commissioned major computing services, including CMS applications, data-transfer utilities, Grid interfaces
- Acknowledged leader among US CMS T2 sites in integration into tiered computing system
  - First to perform T1→T2 and T2→T1 data transfers

First Production UNL T2 Facility

Hardware purchases for 2005: 64 dual-core dual-processor Opteron nodes (~315 kSI2k), 20 TB disk server

- About 20% capacity/complexity of 2007 system

Tying Tiers Together I

Such a distributed system won't function unless there are technologies to connect all of the computers. Hence, "the Grid" and "middleware."

US CMS systems will be part of the Open Science Grid, consortium to develop infrastructure and shared resources for diverse communities of scientists.

- UNL a member -- researchers of all stripes can share/borrow computing

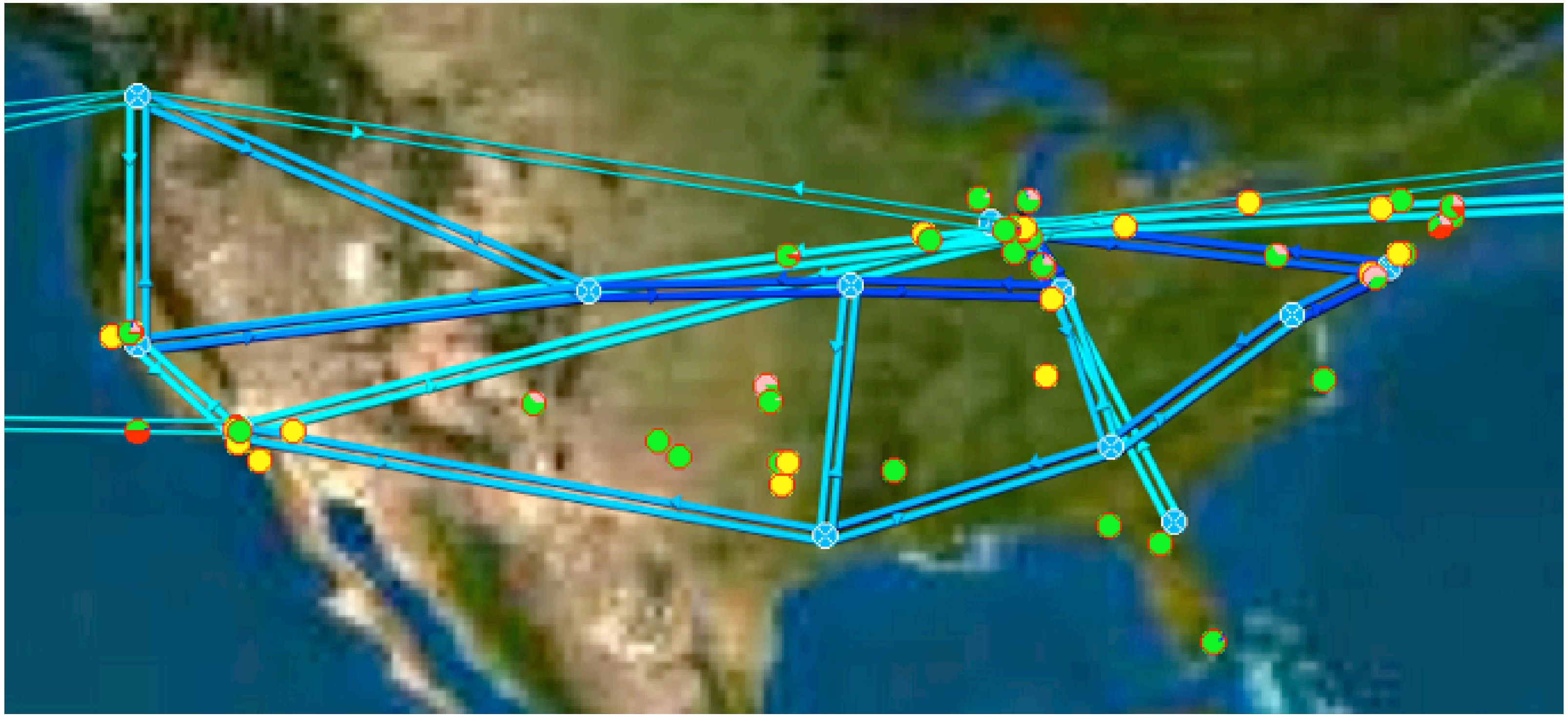
Some needs, and the corresponding middleware tools:

- Authorization and authentication tools
  - Who is allowed to borrow your CPU? Establish a "virtual organization"
  - VOMS breaks users into groups, defines roles and priorities; generates credentials "signed" by trusted authority
- Generic interfaces to diverse resources
  - GRAM and SRM -- generic interfaces to local batch queue, storage system
  - Can handle "load balancing" and "traffic shaping"

Tying Tiers Together II

Everyone wants to know what's happening on everyone else's cluster:

- Describing clusters to outside world
  - GLUE collects info about architecture and state, publishes out
- Monitoring and accounting
  - MonALISA gives real-time info about state of individual clusters and network traffic



Moving and Managing Data

CMS works with datasets that are much larger than those of typical Grid applications. Need to pay special attention to moving and tracking them.

- Obviously, BIG network pipes are needed!
  - CMS expects 2.5-10 Gb/s network connections T1→T2
    - Allows download of a new 200 TB dataset in a few days
  - Also need to use those pipes effectively -- "LambdaStation" project coordinates networking and storage facilities
    - Dynamically allocate network pipes for specific periods of time, optimize data "blasts"
- How do you know where the data are?
  - Can simplify by assigning particular datasets to particular centers; not so helpful for user data that may be in several places.
  - Research underway into efficient cataloguing. Keep track of immutable collections of files instead of files, don't move anything less than ~TB.

T2 Ramp-up and Challenges

CMS T2 project plan shows steep ramp-up of computing resources:

- 2005: ~200 kSI2k CPU, ~20 TB disk
- 2006: ~400 kSI2k CPU, ~70 TB disk
- 2007: 800 kSI2k CPU, 200 TB disk, 2.5 Gb network

Where are the challenges?

- Not the CPU -- RCF knows how to handle that
- Disk will be a challenge -- 200 TB can be operated, but needs lots of care and feeding by humans.
- Don't have the regional network capabilities yet, but expect (hope?) that it will come due to increased demand over time.
- Still much work to do in developing Grid middleware and making it robust -- not as easy as "telnet" yet, can't install and walk away.
- Need skilled labor to make it all go -- CMS has provided funds for 2 FTE system administrators for each site. Labor is biggest cost!

What is Cyberinfrastructure?

It's physical infrastructure:

- Power and cooling; Avery, M&P, WSEC, South Stadium

It's computing power:

- Faster processors, dual-core systems, 64-bit architecture

It's storage capacity:

- 200 TB for T2 possibly the largest disk server in Nebraska

It's network infrastructure:

- Huge datasets must be shipped around the world

It's software and middleware:

- CMS-specific applications, plus tools for functional, robust Grid

It's human resources:

- Still in intense R&D period; skilled labor is always in demand

The success of the CMS physics program depends on the successful development, integration and interaction of our cyberinfrastructure!