



CMS Activities at Tier-2 Centers

Ian Fisk
Tier-2 Workshop
June 12, 2006



Introduction



In CMS the Tier-2 centers are the primary resources for user analysis in the experiment

- ➔ They have the highest percentage of computing resources (disk and CPU) devoted to user specified tasks

Tier-2s are also the only experiment controlled resource for event simulation.

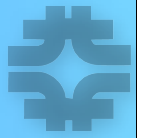
- ➔ The total tier-2s are expected to produce a yearly simulated data sample approximately equal to the real data sample

The Tier-2 centers are a mix of experiment controlled simulation resources and user analysis resources.

- ➔ The prioritization of user analysis resources will be done with input from the local community, regions of interest that collaborate at Tier-2 centers, and the international experiment.



Tier-2 Capacity



The Tier-2s are designed to support 40 active physicists performing analysis.

- ➔ Added to this in the simulation resources for the international experiment

Current plans for capacity by 2008

- ➔ 900kSi2k of computing
 - We estimate around 100-125 dual CPU nodes by the start of CMS
 - Split between analysis and simulation
- ➔ 200TB of disk
 - Mostly used for staging space for local analysis with small quantities for MC staging for transfer to archiving centers
 - Approximately 200MB/s of IO serving capability for applications from mass storage
- ➔ 1-10Gb/s of networking
 - Wide variation in centers



Tier-2 Functionality (1/2)



CMS Event Simulation

- ➔ The CMS estimate of a simulation sample that is roughly equal to the raw sample which corresponds to 100M events per month
- This assumes Tier-2 centers run efficiently throughout the year
 - Roughly 4 times what we hope to do in preparation for CSA06 and roughly ten times what we have currently demonstrated
 - After the experiment starts there is no spare capacity at CERN or the Tier-1 centers for simulation
- Event simulation will be managed centrally
 - Submitted and retrieved through grid interfaces. There is a facility operations load on local support for hardware, interfaces, etc.
- Simulation is sent from Tier-2 centers to the Tier-1s for archiving and serving.
 - In the CMS model Tier-2 sites are not required to have tape based mass storage, so all simulation data is sent to Tier-1 centers.



Tier-2 Functionality (2/2)



Analysis

- ➔ Analysis requires the grid interface for submission, but also requires data be hosted at the Tier-2 center on disk.
- ➔ The disk for analysis for the Tier-2s is large.
 - After the summarized data formats stabilize and the detector is understood, the disk represents a reasonable fraction of the total AOD.
 - During the early days of the experiment, it is anticipated that the analysis will frequently have to use raw or reconstructed data and not small AOD samples
 - Reconstructed data will be updated frequently
 - new calibrations, code and format
 - Leads to larger requirements on size of datasets being imported for analysis
 - Tier-2 disk will act as a cache
 - Increases the importance of networking



Tier-2 Services (1/6)



The Tier-2 centers require a functional grid interface to the processing resources

- ➔ Currently this is either an EGEE or OSG software distribution
- ➔ CMS Simulated Event production is expected to be centrally operated and submitted to sites
- ➔ CMS analysis jobs are expected to enter the site through the grid door
 - CMS analysis workflow tools are designed to provide transparent access to remote resources

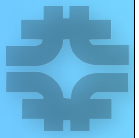
Currently CMS installs the software environment and local configuration in an organized way, though local installation is supported

Additionally some Tier-2 centers are expected to serve as the User Interface for local analysis communities

- ➔ Tier-2 centers may support a local or geographical community



Tier-2 Services (2/6)



CMS Expects to support finer grained authorization at Tier-2 centers for roles and groups through VOMS

- ➔ Enable the experiment to control the prioritization on the fraction of resources pledged to CMS

For Analysis

- ➔ We anticipate 20 analysis groups
 - Each with one analysis coordinator role for common tasks
 - Pool of group members for analysis activities
- ➔ All analysis groups will probably not be supported at all Tier-2 centers, but CMS wishes to be able to assign priority for common resources

For Production

- ➔ We anticipate 3 priority tiers
 - Normal priority, high priority for time critical items and backfill priority for opportunistic resource use.



Tier-2 Services (3/6)



CMS makes only interface requirements on the storage technology used

- ➔ The local area access needs to be readable by Pool and ROOT
 - Files are only expected to be read from mass storage with ROOT and will be written into mass storage with SRM
 - Files once closed in mass storage and registered with the experiment are assumed to be immutable
- ➔ Wide area access and write access are expected to be performed with an SRM interface to the site

At 200TB the expected storage capacity of a Tier-2 center in 2008 is large

- ➔ This has pushed some Tier-2 centers to deploy dCache, though any solution that meets the interface requirements and scales is amenable to CMS

CMS anticipates roughly 1MB/s per batch slot on average for analysis applications this year

- ➔ There may be specific user applications that require higher rate access



Tier-2 Services (4/6)

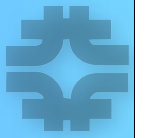


CMS expects to make use of a “trivial file catalog” (TFC) for data resolution on the site.

- ➔ The idea of the TFC is to use a consistent name space on all sites hosting CMS data to eliminate the need for more complicated services to resolve the mapping between logical and physical file names
- ➔ At each site there is a local site configuration file that indicates to CMS applications how to resolve a logical file name
 - For example:
 - if the local file name was
 - /store/preprod/2006/05/05/PreProdR3Minbias/0000/GUIDNAME
 - Then the local configuration file would indicate to append a head to the logical file to resolve the physical file
 - dcap://pnfs/cms/WAX/11/store/preprod/2006/05/05/PreProdR3Minbias/0000/GUI.
 - The name space is designed to scale by providing sufficient granularity that there are not too many entries in any directory



Tier-2 Services (5/6)



The networking requirements for CMS are designed not to exclude sites from participating, but also to encourage sites to have sufficiently strong networking to make reasonable use of the available storage

- ➔ At 200TB
 - With a 10Gb/s network it requires roughly 2 days to flush the entire cache
 - A 2.5 Gb/s network requires a little over a week to flush the disk cache
 - 1 Gb/s network requires over 3 weeks to flush the cache
- ➔ Any less than 1 Gb/s and the disk is essentially static,



Tier-2 Services (6/6)



CMS has a system for managing datasets called PhEDEx

- ➔ Installed as a VOBOX service with access to the mass storage namespace
- ➔ All sites that support CMS data hosting have installed it
 - Lots of successful sites
 - Relies on underlying transfer tools (srmcp, FTS, gridFTP)
- ➔ Manages the replication of datasets, understands windows of data on disk, deals with layout of data into trivial file catalog space

CMS is using the Frontier infrastructure from LCG-3D project to deploy a distributed system for providing calibration data to remote applications

- ➔ Relies on web server technology to provide caches of database queries in squids
- ➔ Easy to deploy and support, 10 instances we installed for testing

We will discuss more about both tomorrow when we talk about analysis



Tier-2 Operations



Tier-2 centers have a significant central contribution to CMS computing

- ➔ They are not considered to be 24/7 operations
 - Though specific sites may have the capability of supporting them that way

The amount of effort required to support the user community and the hardware will probably vary with

- ➔ Size of facility
- ➔ Technology deployed
- ➔ Level of experience

The current estimates for CMS in terms of people needed to support the facility and the experiment services is 2FTE



Tier-2 Status



We had 25 Tier-2 sites come forward for SC4

- ➔ These all will hopefully participate in the CMS data challenge in the fall (CSA06)

Sites have worked through the steps to support

- ➔ Data Transfer
- ➔ Analysis Workflow
- ➔ Production Workflow

So far the interactions have been good

- ➔ the majority of the sites have completed all the steps
- ➔ Still working with many sites to help them succeed.



Outlook



The Tier-2 program is a critical piece in the CMS computing model

- ➔ It represents all the simulation
- ➔ The bulk of the user analysis resources

The services required at the Tier-2 centers should be supportable with a reasonable amount of effort

- ➔ There should be significant overlap with the requirements of other experiments
- ➔ The scale of the hardware is large enough that sites should not underestimate the deployment and commissioning time
 - Storage especially is going require growing some local expertise

The program is ramping rapidly and there is a lot of activity

- ➔ It is not too late to join the deployment activities, but there is not unlimited time especially to contribute to CSA06.