



U.S. CMS Software and Computing Progress Report for the 2nd Quarter FY2006

Technical, financial and management status is reported for the period of January 1st to March 31st, 2006.

Technical Status

U.S. CMS Software and Computing efforts are being driven by the major milestones and activities required for FY06:

- Reliable and automated transfers of data between the CERN Tier-0 and USCMS Tier-1 centers
- Participation in the LCG service challenges to commission a worldwide Grid service for CMS
- Preparation, development and integration of CMS software and computing systems

WBS 1.1 USCMS Tier-1 Facilities

Worker nodes: The Fermilab Tier-1 team has completed a study of motherboards for the FY06 purchases - the ASUS model with dual Opteron 270 dual-core CPUs was selected. They have submitted a purchase order for 240 units, and it will be bid by the normal FNAL purchasing process. Delivery is expected in mid June. Otherwise, scheduled and user analysis processing on the installed worker nodes (1 million Si2K) is routine. Condor groups are being installed to help with priorities, investigating how to set limits on how much output a single user job can generate, and how to kick jobs out more quickly when they use too much memory.

Data disk: We have completed the study of available disk arrays and the newer model of last year's choice is the best choice for our FY06 order. An order for 300 TB of Nexsan's SataBeast was placed and it will be bid by the normal FNAL purchasing process. This is about half of the foreseen FY06 procurement planned for storage. We have also studied the new 3ware 9500 model of an integrated CPU/disk unit. This unit is more stable than previous 3ware models, but still has a higher measured error rate than the Nexsan model. Since the 3ware solution is integrated, the unit's disks are less flexible and do not allow to move disks from system to system as the Nexsan disk arrays do. The 3ware unit is under consideration for the LPCCAF dCache pools – where the disk reliability requirements are lower and should match Tier-2 disk reliability.

We have created a new type of dCache pool - a stage pool. The stage pools stage/read data from the Enstore tape system. These stage pools are similar to existing read pools,

but data is pool-to-pool copied to an Read/Write pool for subsequent user access. This eliminates user access to the stage pool disks, which reduces disk contention and makes transfers run quicker and more smoothly. A similar mechanism is under investigation for writing data from the dCache disks to the Enstore tape system.

User disk: IBRIX has been more stable and we are considering increasing the number of segment servers as the IO and attached disks increase. We are constructing a separate IBRIX test stand to perform initial deployments of new software versions before installing on the production system. We are also planning on having IBRIX come for detailed admin training when we deploy the test stand. The computing division has select BlueArc as its NAS file server. We performed tests with the unit and it performed well. We are deploying one of our Nexsans on the BlueArc unit and letting users test it with real (scratch) data before committing to proceeding with the BlueArc solution.

User Analysis Facility: The version of ssh deployed by Fermilab CD is getting so far out of date that it is starting to be incompatible with deployments of current ssh installs from the standard Linux distributions. We are investigating deploying the standard kerberized ssh server on our UAF nodes. This works fine except for cryptocard access. We did a poll, and to our surprise, many users routinely rely on their cryptocard even though other access methods are available to them. We are considering deploying some type of transparent gateway for the cryptocard users.

Data transfers: It is difficult to understand where problems really are when a SRM transfer stalls or fails. A program to provide more information on the web about SRM transfers has been started by the SRM developers to try and overcome this nagging issue. A change in the cost model for pool selection has allowed a better balance of traffic, and higher rates, for WAN gridftp transfers. A similar change in the model for selecting gridftp doors has also allowed better balancing of traffic for the USCMS Tier-2 sites behind firewalls. The initial version of gPlazma, an authorization and mapping service for the SE has been deployed and works for gridftp transfers. Dcap and SRM gPlazma modes are still being investigated. The initial version of SRM-2, running under tomcat, has been deployed. It works fine for everything except CERN's FTS, and even there files are transferred properly and we consider the transfer a success. Only FTS's final handshake has an error and fails the entire transfer. We are working with CERN to understand what is causing the problem. We expect to have this solved quickly, and then will make SRM-2 the default SRM server at FNAL.

Infrastructure: The accelerator is down at FNAL and there have been several power and cooling outages to improve service. These are a minor inconvenience and operations do seem to be improved as a result of the work.

Tape Robotics: The CCF department has placed an order for the STK SLA8500 and it is planned for arrival in May. We are closely watching its deployment and hope to help start testing soon.

WBS 1.2 USCMS Tier-2 Facilities Program

After the major effort for Service Challenge 3 and many hardware purchases last quarter, the U.S. CMS Tier-2 facilities were largely focused on improving the performance and reliability of their systems during the first months of 2006. For instance, all sites

demonstrated that they could deliver data from their storage systems to computing elements at the rate of 1 MB/s/batch slot. To achieve this goal, several sites were inspired to restructure their storage systems, which made them more robust and more prepared for the greater demands to come. In addition, all sites had a period of at least three days in which their CPU's were at least 75% occupied by jobs that had been submitted through the Grid interfaces (as opposed to jobs submitted locally). This demonstrates the reliability of these interfaces over a long period of time. During the quarter, Tier-2 sites delivered about 800,000 hours of CPU time -- roughly 90 CPU years -- to various virtual organizations. About 320,000 of those CPU hours were used by CMS.

In addition, the sites continue the development and deployment of the necessary software and services. All sites moved to the newest version of the Open Science Grid software, 0.4.0, in March. All now host a version of CMSSW, the new CMS software library. All are ready to host data through the new DBS/DLS system.

On March 3 2006, Caltech hosted a workshop on networking issues for the US CMS Tier-2 sites. Participants included representatives from all of the sites, Fermilab networking experts, and representatives from ESNet. The sites are fully engaged in the issue of network capacity and reliability. The best news to come out of the workshop is that all sites are now on track to have WAN capacity of 10 Gb/s by the end of 2006 -- greater capacity than expected, and sooner than expected. During the workshop, the sites agreed to a set of milestones that will test out network capacity, reliability and topology during the coming year.

As the quarter ended, sites were preparing for the annual (for now) Tier-2 workshop, hosted by the University of Nebraska-Lincoln on April 6.

WBS 1.3 Grid Services and Interface with the Open Science Grid (OSG)

Now that OSG 0.4.0 is in use we moved on to version 0.4.1, which has additional features for making the Compute Element and Storage Elements performant and integrated. This adds support for the *Jobmon* Grid debugging tool to the software stack and extends the Information Services integration with the EGEE for the WLCG. There was much end-to-end testing of the software release to ensure smooth installation and configuration,

A test suite for the end-to-end security components is being developed, as well as a framework for performance testing of *Prima* and GUMS.

CRAB jobs were successfully submitted through the WLCG CMS Resource Broker to the Nebraska Tier-2 on the Open Science Grid. Several issues of detail through the OSG EGEE bridge to the WLCG were worked out, and the interoperability support people at CERN continue to be very helpful. A template for the OSG Site policies is being circulated round the consortium for comment and then release.

Support continued for the use of resources at the Tier-2s and on the OSG. An offer of employment was made to replace the person who left last October.

WBS 1.4 Application Services

U.S.CMS engineers are working with the international CMS efforts in the CPT-Computing project to provide the application-level computing services. The requirements, interfaces and functionalities of the application areas are now reasonably well described in the technical baseline documented in the CMS Computing TDR.

The initial design of the Dataset Bookkeeping System DBS proceeded in this quarter, with a design report being released within the CMS data management group. Developers at Fermilab worked on a prototype implementation of the database server, and on the mechanisms to fill the DBS schema from the pre-existing reldb (MC production database) information.

The Fermilab group had lost one of its lead developers in this area, and it took until September that he could be replaced by moving an experienced developer and designer with a lot of expertise on data management gained in the SAM project, S.Viseli, into the CMS project. However, specific deliverables in the data management area like the prototype of the DBS was slipping by several months. As a backup, one of the U.S. engineers based at CERN, L.Tuura, took up coordinating the DBS efforts in particular to make sure that the required deliverables for the Service Challenge 3 were delivered.

The PhEDEx data transfer system was deployed at all SC3 sites and monitoring and operations support tools were developed and installed. PhEDEx orchestrates the production dataflows of moving datasets to the hosting sites. Also PhEDEx is being successfully used to orchestrate the SC3 related data transfers.

Other efforts included making progress on the definition of the data flow from the HLT into the Tier-0 center and the related data formatting and data bookkeeping issues.

In the September PMG meeting project oversight agreed that P.Elmer of Princeton University would take up coordination of this area. Since then he has become part of the Project Execution Team and is closely coordinating these efforts.

The DISUN project also contributed to the Grid Services and Interfaces WBS deliverables in a variety of ways, as follows:

- Testing of the OSG itb release for CMS applications.
- Testing the srmcp client installation on itb.
- Provisioning OSG 0.4.1. This is ongoing still.
- Developing a CE configuration that does not require NFS exports to the worker nodes.

WBS 1.5 Distributed Computing Tools

DISUN is contributing effort to the Distributed Computing Tools WBS deliverables, in particular to the ProdAgent development, integration, and beta testing. This work is ongoing.

Work on CRAB is also ongoing. DISUN continues to be responsible for the vast majority of CMS job submissions on OSG. In addition to the standard production MC, we have started large scale generation of Alpgen MonteCarlo events.

DISUN is working on establishing high bandwidth SRM transfers between FNAL and two of the DISUN sites. A bandwidth limitation due to scheduling in SRM was identified, and reported to the FNAL SRM team. Further work is now stalled while we wait for an SRM update that is being worked on.

DISUN has started deploying CRAB UI at two of the DISUN sites in order to transition local users to using CRAB for their work at the tier-2. This work is ongoing. DISUN has also worked with Fermilab on integration of the GUMS callout from gPLAZMA. This work is ongoing. In addition, DISUN has started deployment of releases of the new software, CMSSW, at the Tier-2 centers.

WBS 1.6 Software and Support

During this quarter, for the new CMS framework/EDM there has been a lot of work on usability issues, like finer tuned output branch selection capabilities, and configuration support. We have also developed software quality assurance tools like timing and memory tracking utilities.

The main focus of attention has been on the needs of our collaborating projects including:

- Simulation: for whom we developed a notification mechanism, FileInPath, random number and geometry services
- Reconstruction: which needed special EDM container templates for calorimetry and tracking
- Calibration: which had additional feature requests from the event setup system
- Event Filter Farm: for which we developed event streaming code and a first prototype of the storage manager
- Data Management: which has a long list of joint software and computing features

In the software support area, we have taken a lead role in developing a new distribution system in which rpms are generated from source even for external projects like the LCG. This will give us a greater flexibility to determine when we are ready to move to new platforms and will allow for more automation since we will have full control of the release procedure and distribution building system. Software and Support has succeeded in copying the nightly build system developed at CERN to FNAL. Users at FNAL can now use the results from nightly builds for rapid software development just as if they were at CERN. This has made a big difference to our users. Work to support and install the many frozen releases needed for computing and physics challenges continue. User desktop and software consulting continues.

WBS 2.0 Core Applications and Support, CAS

M.Case (UC Davis) is now working on and coordinating the Geometry project. He continues to work on developments of the Geometry as CMSSW continues to evolve.

R.Wilkinson (Caltech) has started working on the framework. His first tasks have been working on ways to include bits of parameter files into larger sets, to allow easier parameter changes by the users. He has also started looking at ways of filtering events based on trigger paths.

I.Osborne (Northeastern U) continues to work and debug the event visualization and event display package IGUANA in the new framework. More users are starting to request features in IGUANA. IGUANA is heavily dependent on the framework, and the framework is still evolving rapidly. However, the work is on track to provide an important tool for the upcoming Magnet Test/Cosmic Challenge.

Z.Xie (Princeton) and M.Case are continuing their database efforts. All of the effort now is in the new CMSSW framework. Z.Xie has been involved in helping the various detector groups define their database objects within POOL to be able to make the access as efficient as possible. She is one of the main developers for the online to offline db transfer. She is also continuing her development work on POOL. She is responsible for the SQLiteAccess plugin for POOL.

S.Muzaffar (Northeastern U) is continuing the maintenance of the *Ignominy* tools. This includes the migration to the new framework. He has also helped develop the tools for the nightly build system. He has also contributed to IGUANA bug fixing.

M.Stavrianakou (Fermilab) is working on and coordinating the efforts of the simulation part of CMSSW. Work continues on porting the elements of OSCAR into CMSSW. With the PTDR on schedule for release in May, the maintenance work on OSCAR has been minimal.

WBS 3.0 Project Office

Statements of Work are been updated for the new funding period, both for the DOE funded efforts with FY06, and for the need NSF funding period that started in February 2006. Funds have been allocated for DOE during FY06 and for NSF for the next 9-month funding period from Feb to Oct 2006.

We still see a large lag time between the effort being spent and the actual accounting. While the reporting of efforts is prompt and a system is in place for individual effort reports for each person on project, we are working with Universities to improve the invoicing.

For the DOE fraction of the S&C funds, in the following table we show the funding allocations (BCWS), the allocated effort, the reported effort and the actual invoiced costs (ACWP).

FY06 Funds and Costs	FY05 total	FY06 total	FNAL	Universities	Caltech	Princeton	UCD	Wisc	Cornell	UofC	Iowa
Funding Allocated											
Systems & Operations Personnel	\$5,492k	\$4,054k	\$2,724k	\$1,330k	\$400k	\$500k	\$160k	\$100k		\$50k	\$120k
USCMS Software		\$1,439k	\$1,439k								
Tier-1 Equipment	\$1,163k	\$3,495k	\$3,495k	\$0k							
Project Office, Reserve	\$982k	\$2,873k	\$2,873k	\$0k							
Total Allocated	\$7,637k	\$11,861k	\$10,531k	\$1,330k	\$400k	\$500k	\$160k	\$100k	\$0k	\$50k	\$120k
USCMSSC DOE Funds	\$7,637k	\$11,861k	\$10,531k	\$1,330k	\$400k	\$500k	\$160k	\$100k		\$50k	\$120k
Effort Allocated											
Systems & Operations FTE years	19.7			4					2.0	0.5	1.25
Effort Spent (6 months)											
Systems & Operations FTE years	38.5	9.5	7.8	1.8					1.0	0.25	0.5
USCMS Software FTE years		4.0	4.0								
ACWP in k\$, as invoiced											
Systems & Operations Personnel		\$1,292k	\$1,292k	\$0k							
USCMS Software		\$627k	\$627k	\$0k							
T1 Equipment	\$1,144k	\$186k	\$186k	\$0k							
Project Office	\$192k	\$117k	\$117k	\$0k							
Total		\$2,223k	\$2,223k	\$0k	\$0k	\$0k	\$0k	\$0k	\$0k	\$0k	\$0k

The NSF funding periods cross fiscal year boundaries, while in the past we have reported within fiscal year periods. During this NSF funding period and the addition of the DISUN project, the new Tier-2 centers started to receive funding according to a startup plan. The funding allocation of the NSF program is shown in the following table, as approved by project oversight in the PMG meeting. The Tier-2 centers receive funding through the iVDGL project, the DISUN project and the USCMS project (Research Program, RP). The table shows the allocated funds from these sources for the RP period nominally starting in May 2005 and running through April 2006.

NSF Funds and Costs	Total	Caltech	UFL	UCSD	Wisc	MIT	Purdue	UNL	NEU	UCLA
iVDGL Funding for CMS -- Funding through 7/2006										
Tier-2 Equipment	\$60k	\$20k	\$20k	\$20k						
Tier-2 Labor	\$390k	\$130k	\$130k	\$130k						
Total	\$450k	\$150k	\$150k	\$150k						
Research Program -- Funding through 4/2006										
Tier-2 Equipment	\$681k				\$180k	\$125k	\$188k	\$188k		
Tier-2 Labor	\$501k					\$125k	\$188k	\$188k		
Project Office	\$39k									\$39k
Software Labor	\$370k									
Management Reserve	\$461k								\$370k	\$461k
Total	\$2,052k				\$180k	\$250k	\$376k	\$376k	\$370k	\$500k
DISUN Project										
Tier-2C Equipment	\$750k	\$230k	\$230k	\$230k	\$60k					
Tier-2C Labor	\$1,173k	\$270k	\$270k	\$270k	\$363k					
Project Office	\$77k									\$77k
Total	\$2,000k	\$500k	\$500k	\$500k	\$423k					\$77k