



# **USCMS Software and Computing Progress Report for the 3<sup>rd</sup> and 4<sup>th</sup> Quarter FY2003**

**For Period April 1 to September 3, 2003**

Technical, financial and management status is reported for the period April 1 through September 30, 2003.

## **1 Technical Status**

Two events, the upcoming Data Challenge DC04 and the start of the first deployment of an LCG service at the Fermilab Tier-1 center, drove the USCMSSC efforts through the end of the fiscal year. The production of a large, multi-million simulated and reconstructed even sample for DC04, the so-called Pre-Challenge Production (PCP), started this summer.

### ***1.1 User Facilities subproject***

#### ***1.1.1 Tier-1 Facilities***

Preparations for three major milestones are driving most of the Tier-1 activities now, and over the next months:

- The DC04 Data Challenge and the preparation for the computing TDR (C-TDR)
- The preparation for the Physics TDR (P-TDR)
- The roll out of the LCG Grid service and federating it with the U.S. facilities.

During the past two quarters the following efforts contributed to achieving these goals.

We report the major activities along the work breakdown structure.

##### **1.1.1.1 WBS 1.1.1 Building Components**

The primary areas of R&D efforts included

- Development of the Grid Storage Element and integration of the prototype into the USCMS production environment. The prototype is based on dCache and the Storage Resource Management (SRM) protocol. Substantial parts of the SRM protocol version 2.1 have been implemented as a JAVA plug-in to extend the management capabilities of dCache. This was built according to the definition of the Storage

Element and demonstrated to work between the CMS Tier-1 center at Fermilab and the CMS Tier-2 prototype center at UCSD. We are planning to deploy the described Storage Element at all USCMS Tier-2 centers and eventually at CERN. Areas of further activities include understanding how the Storage Element, in particular the catalogue service, communicates with POOL and the applications.

- Increase the efficiency of data transfer. Fermilab is in the process of obtaining an optical network connection to StarLight in Chicago, the premier optical network-switching center on the North American continent. This is of immediate benefit to production bulk data transfer and a test bed for high performance network investigations and scalability into the area of LHC operations. The Tier-1 facilities group in collaboration with the Fermilab Computing Division and the EU sponsored DataTAG project has launched a project to actively pursue the integration of network stack implementations supporting ultra scale networking for rapid data transactions and data-intensive dynamic workspaces.
- Improvement of data serving for data intensive applications. Started to exploit dCache as a way to utilize disk space on the worker nodes and efficiently supply data in intense applications like simulation with pile-up. The results of this evaluation look very promising and they were presented to international CMS. CMS currently sees no alternative to this method and asked the Tier-1 facilities group to help with a dCache installation at CERN.

#### **1.1.1.2 WBS 1.2.1 Preparations for Analysis**

After DC04 is completed, CMS will enter a year of intense analysis activity preparing for the Physics TDR. Therefore USCMS needs to increase the analysis capability of the Tier-1 center and the number of users working in the US. While many of the preparations overlap with DC04 preparations some extensions, especially in the area of user management in the environment of Virtual Organizations and extensions to the production tools to allow custom user simulation and distributed analysis applications became necessary. To appropriately support interactive users in a farm environment a load balancing mechanism based on the Fermilab batch system was adopted. Also, significant improvements were made to the software and developer environment, now matching what is available to CMS researchers working at CERN.

#### **1.1.1.3 WBS 1.1.3 Production Facilities and WBS 1.1.5 Procurements, Equipment Costs and Investments**

According to conservative estimates, the required resources at the US CMS Tier-1 for contributing a fair-share to the CMS DC04 would be 240 CPUs. This is approximately twice the currently available resources. USCMS has a small number of systems reserved for data analysis. As planned, in preparation for the Physics TDR analysis, these computing resources are being ramped up.

For production of simulated and reconstructed events, USCMS needs to deploy a system large enough to satisfy the commitments to CMS, while allowing high performance network connections between systems to enable efficient simulation processing and, at the next stage, improved analysis capability. In order to accomplish this, the large cluster broken up into cells. Each cell consists of 20

dual CPU compute nodes with one GigE copper connection and 1 to 4 RAID systems, all connected via a 24 port copper GigE switch. The switches have a 32-gigabit interconnect and multiple fiber uplinks can be aggregated to connect the cells to the FNAL site network

To provide disk space to physics users at the Tier-1 center, US CMS is procuring a large NAS device. The system is scoped to have approximately 3TB of usable storage space and will allow for high availability and redundancy. The 3TB system will allow to support a community of up to 30 active users, fulfilling the requirements for use space as estimated and experience in user requests for analysis. This size of user storage is also matching well with resources available on desktop systems, allowing selections and data summaries to be made and then transferred to local resources.

In summary the following Tier-1 facility hardware and software was either procured and installed or ordered during FY03:

- Upgrade of Production Facility
  - 60 dual Intel 2.4GHz XEON nodes
  - 17 Server nodes
  - 4 Storage systems (20TB usable capacity)
  - 3 Copper GigE network switches (84 ports)
  - 1 Fiber Channel switch (24 ports)
- Upgrade of the Analysis Farm
  - 16 dual Intel 2.4GHz XEON nodes
  - 5 server nodes
  - 3 Storage systems (8TB usable capacity)
  - 1 Copper GigE network switch
  - 4 Licenses for the IBRIX Cluster File System
- Miscellaneous
  - 25TB of tape media for the STK 9940B tape drives (200GB/cartridge)

#### **1.1.1.4 Rollout of the LCG-1 service**

The deployment of the LCG prototypes is not by itself a CMS milestone, but the functionality expected from the releases is coupled to the major CMS milestones. It is CMS expectation that the LCG-1 service can provide a consistent set of grid middleware components and provide required functionality for DC04 data and process management.

Installing the LCG pre-release LCG-0 and the first production release LCG-1, in particular at the Fermilab (and for US Atlas the Brookhaven) Tier-1 centers, has shown that the technology and the process the LCG project has chosen is more intrusive than is acceptable to the computing organizations at the lab.

Considerable effort had to be spent to test, verify and to assure that the installation of the LCG-1 components is in compliance with the policies and standards at Fermilab. Some of the assumptions

the LCG project at CERN made for the initial LCG-1 release, about the architecture and security model etc, do not take into account issues of on how to integrates with existing site policies.

We found that the LCG ramp up of functionality could not be achieved as rapidly as the LCG project had initially expected. We are evaluating ways of interfacing more efficiently with the LCG project.

### ***1.1.2 WBS 1.3 Distributed Production Environement***

The Distributed Production Environment made significant progress, despite being currently understaffed. US CMS could however get significant additional manpower contributions from Universities, funded through Grid-related activities.

- Continued support for CMS Monte Carlo production operations on the USCMS grids
- Release of DPE 2.0, which included upgraded versions of MOP, including maintenance contributions from the PPDG-supported developer, Anzar Afaq, and a new Condor matchmaking module for MOP contributed by an UW Madison supported person, Dan Bradley.
- An iVDGL-supported person, Bockjoo Kim of University Florida, was integrated into the release project to help the US CMS librarian to manage DPE releases.
- DPE versions 2.1 and 2.2 were released, where DPE middleware components were successfully migrated into the VDT and Grid3 pacman caches were employed.
- A new MDS-based configuration monitoring tool that we currently use to distribute MOP variables was release by Yujun Wu and a summer student, Xin Li.

The DPE was successfully migrated into the new Grid2003 environment (see below), and CMS production could be run on Grid2003 using the DPE components.

### ***1.1.3 Distributed production and Tier-2 Facilities***

The primary activities of the U.S. CMS prototype Tier-2 facilities during these two quarters were preparing for Pre-Challenge Production (PCP) for DC04 and participating in Grid2003. In the past, large-scale event production has required dedicated personnel at each facility to shepherd the production jobs. The goal this year was to have all Monte Carlo event production run centrally using prototype grid services. We had succeeded at this goal on smaller scale production, but substantial effort was spent in diagnosing and debugging the system. During the most recent production, the distributed production system was relatively stable, but the efficiency has not been as high as expected. The User Facilities project is working to raise the overall computing efficiency. All prototype Tier2 facilities are participating in this production effort with production requests launched centrally.

The other major activity of the prototype Tier-2 facilities during the last two quarters has been the development and deployment of Grid2003. Grid2003 is a multi-organization effort to form a large-scale grid infrastructure to meet the distributed computing needs of the stakeholders. It is a joint project between U.S. CMS, U.S. ATLAS, and the Trillium Grid Projects. The effort started with the deployment of the Grid2003 software onto the US CMS development grid testbed, that includes the US CMS Tier-1 and prototype Tier-2 sites and additional university installations. After an initial successful commissioning Grid2003 has been moved now to the Tier-2 production resources. These

join computing resources at over 20 other sites to form a grid enabled computing facility with over 2000 CPUs. Personnel at the USCMS prototype Tier-2 centers have been responsible for installing and maintaining software at their sites, monitoring performance, and diagnosing problems.

## ***1.2 Core Application Software***

### ***1.2.1 Architecture Framework Toolkits and Reconstruction Software (WBS 2.1)***

Work continued on implementing POOL into the CMS software framework. Bill Tanenbaum at Fermilab worked closely with the POOL software engineers at CERN to get a version of POOL ready that was usable by CMS. Version 7\_4\_0 of the CMS application framework COBRA was released during this time that was able to use POOL. This important milestone enabled the production of events for DC04 to proceed. With the POOL and COBRA releases, OSCAR was able to store events. Thus, for DC04 there will be a significant number of events simulated with OSCAR/GEANT4. ORCA is also able to read and write POOL files. Zhen Xie continued to work on the file catalog of POOL, developing command line tools as well as python bindings to the file catalog. Work started on developing a user's guide.

Vladimir Litvin of Caltech continued coordinating calorimetry reconstruction framework.

Michael Case of UC Davis continued with the DDD deployment in CMS software clients. ORCA now uses fully XML based DDD geometry. He started working together with HCAL physicists to develop an "ideal" geometry, the first step toward implementing a "human-maintainable" geometry. Up until now, the DDD has been machine generated from GEANT3 files. In the longer term this will be untenable, and thus a better way to generate the DDD must be devised. Michael is working on this.

Preliminary work started on the conditions database.

### ***1.2.2 Visualization and Analysis (WBS 2.2)***

No major developments in visualization occurred during this period. Instead, a number of steps were taken to consolidate and improve the code. All CMS specific parts were moved to the IGUANACMS package. This helps separate the generic visualization framework from the experiment-specific one, and should allow easier maintenance.

### ***1.2.3 Distributed Computing Software (WBS 2.3)***

A significant effort during this period went into the "PCP," with the coordination of the production of about 30 million events worldwide, on the way to producing at least 50 million events for DC04. Several new versions of McRunJob were released to CMS by Greg Graham. The new releases included:

- Functionality from a partly UC Riverside-support developer at CERN, Julia Andreeva, for new production applications in OSCAR and updated applications in ORCA 7.

- Functionality from an INFN-supported developer, Alessandra Fanfani, for LCG integration, and maintenance.
- A Fermilab-based effort to provide a common basic code distribution for the CMS and Dzero versions of Mcrunjob, the ShahKar project, produced its first RPM for distribution to the DZero and CMS experiments.

### 1.2.4 Support

The CAS subproject continued with various support activities related to production and data challenge support. Also work continued on various software support tools. DAR and REFDBDAR continued to be developed in light of the PCP requests.

As indicated above, some support of DDD for HCAL physicists was started.

## 2 Financial Status

### 2.1 DOE/NSF Funding

A first allocation of \$1,185K in DOE operations funds for FY03 arrived in April 2003 and \$1,640K arrived in July for a total of \$2,825K.

NEU administered \$750K in NFS fund for USCMS in FY03. Through the iVDGL project, an additional \$466k of NSF funds were made available to Prototype Tier 2 centers.

The following table shows the total FY03 funding allocation, and the budgeted funding and effort.

<b>FY03 Funds and Costs</b>	<b>FY02 total</b>	<b>FY03 total</b>	<b>FNAL</b>	<b>Caltech</b>	<b>UCSD</b>	<b>UFlorida</b>	<b>NEU</b>	<b>Princetor</b>	<b>UCDavis</b>	<b>UWMadik</b>	<b>U.Iowa</b>	<b>UofM</b>
<b>Funding Allocated</b>												
CAS Personnel	\$1,245k	<b>\$1,370k</b>	\$420k	\$175k				\$380k	\$245k	\$150k		
UF Personnel	\$1,239k	<b>\$2,102k</b>	\$1,385k	\$341k	\$124k	\$166k					\$9k	\$25k \$52k
Tier-1 Equipment	\$700k	<b>\$550k</b>	\$550k									
Tier-2 Equipment	\$232k	<b>\$192k</b>		\$48k	\$48k	\$96k						
Project Office, Reser	\$180k	<b>\$179k</b>	\$179k									
<b>Total Allocated</b>	<b>\$3,596k</b>	<b>\$4,393k</b>	<b>\$2,534k</b>	<b>\$564k</b>	<b>\$172k</b>	<b>\$262k</b>	<b>\$380k</b>	<b>\$245k</b>	<b>\$150k</b>	<b>\$9k</b>	<b>\$25k</b>	<b>\$52k</b>
<b>USCMSSC DOE Fun</b>												
USCMSSC DOE Fun	\$2,440k	<b>\$3,115k</b>	\$2,534k	\$100k					\$245k	\$150k	\$9k	\$25k \$52k
USCMSSC NSF Fun	\$690k	<b>\$750k</b>		\$290k	\$40k	\$40k	\$380k					
iVDGL Funds	\$466k	<b>\$528k</b>		\$174k	\$132k	\$222k						
<b>Effort Allocated</b>												
CAS FTE years	8.33	<b>11.0</b>	3	1				3	2	1		1
UF FTE years	9.42	<b>18.7</b>	11	3	2	2					0.2	0.5

**Table 0** Budgeted Costs of Work Scheduled for FY03

The next table shows the actual effort spent and the actual costs of work performed, as reported, during the first two quarters of FY03.

<b>FY03 Funds and Costs</b>	<b>FY02 total</b>	<b>FY03 total</b>	<b>FNAL</b>	<b>Caltech</b>	<b>UCSD</b>	<b>UFlorida</b>	<b>NEU</b>	<b>Princeton</b>	<b>UCDavis</b>	<b>UWMadison</b>	<b>U.Iowa</b>	<b>UofM</b>
<b>Total Allocated</b>	\$3,596k	<b>\$4,393k</b>	<b>\$2,534k</b>	<b>\$564k</b>	<b>\$172k</b>	<b>\$262k</b>	<b>\$380k</b>	<b>\$245k</b>	<b>\$150k</b>	<b>\$9k</b>	<b>\$25k</b>	<b>\$52k</b>
<b>Effort Allocated</b>												
CAS FTE years	8.33	<b>10.0</b>	3	1			3	2	1			
UF FTE years	9.42	<b>19.7</b>	11	3	2	2				0.2	0.5	1.0
<b>Effort Spent</b>												
CAS FTE years		<b>8.9</b>	1.9	1.0			3.0	2.0	1.0			
UF FTE years		<b>12.6</b>	7.9	3.0	0.5	1.0				0.2	0.25	0.1
<b>ACWP in k\$</b>												
CAS Labor		<b>\$1,209k</b>	\$269k	\$175k			\$380k	\$245k	\$140k			
UF Labor		<b>\$1,403k</b>	\$1,036k	<i>\$341k</i>	<i>no data</i>	<i>no data</i>				\$9k	\$13k	\$5k
T1 Equipment		<b>\$406k</b>	\$406k									
T2 Equipment		<b>\$144k</b>		<i>\$48k</i>	<i>\$48k</i>	<i>\$48k</i>						
Project Office		<b>\$155k</b>	\$155k									
<b>Total</b>		<b>\$3,318k</b>	<b>\$1,866k</b>	<b>\$564k</b>	<i>no data</i>	<i>no data</i>	<b>\$380k</b>	<b>\$245k</b>	<b>\$140k</b>	<b>\$9k</b>	<b>\$13k</b>	<b>\$5k</b>

**Table 0** Actual Costs and FTE effort of Work Performed in FY2003

Reporting and invoicing of NSF-funded effort is still sparse, and where insufficient information was available numbers are shown in italics. The US CMS S&C project office will work towards receiving a more complete picture of efforts spent at NSF-funded universities, and give that information in a later report.

## 2.2 Proposals submitted to the NSF

US CMS was advised to look for additional funding by submitting proposals for projects to the NSF. Project Management was directly involved in writing a number of proposals. In addition to the Research Program proposal, US CMS is involved in a “large ITR” proposal, together with U.S. Atlas on “Dynamic Workspaces Enabling Scientific Discovery” (PI J.Huth/Harvard, co-PIs L.A.T.Bauerdick/Fermilab, M.Livny/U.Wisconsin at Madison, M.Selzer/Harvard, R.Stevens/U.Chicago), and a “medium ITR” proposal on “A Global Grid-Enabled Collaboratory for Scientific Research” (PI H.Newman/Caltech).

## 3 Management Status

There was a personnel change in the Project Office: Bakul Banerjee took over new responsibilities in the Computing Division. Mark Leininger assumed her position as Project Engineer in July.

Ian Fisk took a position as an Associate Scientist at Fermilab and started as the Level2 Manager for User Facilities in April 2003. Bob Clare of UC Riverside started as the Level2 Manger for Core Application Software in August 2003. With these changes, the management team is again complete after more than a year of vacancy of one of the Level-2 manager positions.

There is an unresolved issue regarding computing equipment purchased through the CERN team accounts. Unlike detector components, we cannot transfer ownership of computing equipment to CERN. Consequently, someone at Fermilab has to retain custodial responsibility and the property office has to mail a property tag to CERN. This will become unwieldy and needs to be resolved. The business office is pursuing a solution with DOE.

### ***3.1 Meetings and Reviews***

The project reported to the DOE and NSF in a status meeting at the NSF on July 8, 2003 (see <http://uscms.org/s&c/reviews/doe-nsf/2003-07/>)

The following USCMSSC Program Management Group (PMG) meetings were held during this period:

June 13, 2003  
August 8, 2003

### ***3.2 Documents***

- MOU and FY04 SOW with Iowa on NSF-funded effort were approved.
- MOU and FY04 SOW with Minnesota on DOE funded effort is complete; purchase requisition is in the Fermilab system.
- MOU between iVDGL and USCMSSC is being signed.
- FY04 SOW for UCSD for iVDGL funds is ready but not signed by either party.
- FY04 SOW for Florida and Caltech for iVDGL funds waiting statement of deliverables and dates.
- FY04 SOW for UCD (CAS) prepared but funding decision not complete.
- unpaid invoices from CERN, Princeton, Caltech and UCSD were processed and paid.