

# CMS Requirements for the Grid

K. Holtman<sup>1</sup>, J. Amundson<sup>2</sup>, P. Avery<sup>3</sup>, S. Aziz<sup>2</sup>, L.A.T. Bauerdick<sup>2</sup>, J. Branson<sup>4</sup>, J.J. Bunn<sup>1</sup>, P. Capiluppi<sup>5</sup>, R. Clare<sup>6</sup>, A. Dominici<sup>7</sup>, F. Donno<sup>7</sup>, I. Fisk<sup>4</sup>, I. Gaines<sup>2</sup>, G. Graham<sup>2</sup>, C. Grandi<sup>8</sup>, T.M. Hickey<sup>1</sup>, V. Innocente<sup>9</sup>, W. Jank<sup>9</sup>, V. Kolosov<sup>10</sup>, N. Kruglov<sup>11</sup>, A. Kryukov<sup>11</sup>, I. LeGrand<sup>1</sup>, V. Litvin<sup>1</sup>, M. Livny<sup>12</sup>, M. Mennea<sup>13</sup>, V. Mitcin<sup>14</sup>, S. Muzaffar<sup>2</sup>, H.B. Newman<sup>1</sup>, V. O'Dell<sup>2</sup>, P. Ronchese<sup>15</sup>, A. Samar<sup>1</sup>, A. Sciaba<sup>7</sup>, L. Silvestris<sup>13</sup>, S. Singh<sup>1</sup>, C. Steenberg<sup>1</sup>, D. Stickland<sup>16</sup>, H. Stockinger<sup>9</sup>, K. Stockinger<sup>17</sup>, L. Taylor<sup>18</sup>, H. Wenzel<sup>2</sup>, T. Wildish<sup>16</sup>, R. Wilkinson<sup>1</sup>, I. Willers<sup>9</sup>, S. Wynhoff<sup>16</sup>, for the CMS collaboration.

<sup>1</sup>(Caltech)

<sup>7</sup>(INFN-Pisa)

<sup>13</sup>(INFN-Bari)

<sup>2</sup>(Fermilab)

<sup>8</sup>(INFN-Bologna)

<sup>14</sup>(JINR Dubna)

<sup>3</sup>(Univ. of Florida)

<sup>9</sup>(CERN)

<sup>15</sup>(INFN-Padova)

<sup>4</sup>(UCSD)

<sup>10</sup>(ITEP)

<sup>16</sup>(Princeton)

<sup>5</sup>(Bologna Univ.)

<sup>11</sup>(SINP MSU)

<sup>17</sup>(Caltech/CERN)

<sup>6</sup>(UC Riverside)

<sup>12</sup>(Univ. of Wisconsin Madison)

<sup>18</sup>(Northeastern Univ.)

## Abstract

CMS physicists need to seamlessly access their experimental data and results, independent of location and storage medium, in order to focus on the exploration for the new physics signals rather than the complexities of worldwide data management. In order to achieve this goal, CMS has adopted a tiered worldwide computing model which will incorporate emerging Grid technology.

CMS has started to use Grid tools for data processing, replication and migration. Important Grid components are expected to be delivered by the Data Grid projects, like EU DataGrid, PPDG and GriPhyN. As part of the activity of interfacing with these projects, CMS has created a set of long-term requirements to the Grid projects. These requirements are presented and discussed.

Keywords: CMS, Grid, Data Grid, Requirements, GriPhyN, PPDG, EU DataGrid

## 1 Introduction

In the period December 2000 – July 2001, CMS [1] conducted a major requirements and consensus building effort that resulted in a series of documents with concrete requirements for the Grid projects (GriPhyN [2], PPDG [3], and the EU DataGrid [4]) that CMS is involved in as a 'customer'. At the highest level, the requirement is simply that the Grid projects should deliver software components to CMS which can be used by CMS in the construction of the CMS Data Grid system. At a more detailed level, the requirements give a comprehensive overview of the Data Grid system that CMS intends to operate around December 2003. This CMS Data Grid system contains a large number of software components from different sources, including the Grid projects. Final selection, integration, and operation of these components will remain the responsibility of CMS.

The requirements effort focused in particular on the *architectural constraints*, which the Grid components to be delivered to CMS need to take into account, on the *scalability requirements* for these components, and on the *level of complexity of the workload* that needs to be supported.

To support its current wide-area distributed production effort, CMS has developed, and is already operating, a 'proto-Grid' system at a scale of about 10 sites, 500 CPUs, and a few TB of storage space. This proto-Grid system already satisfies some of the requirements outlined below, and uses several components created by the Grid community. The goal is to incorporate additional Grid components when they become available, evolving the system towards greater capabilities and greater scalability.

## 2 Requirements documents

The central requirements document is [5]. This document contains a *snapshot*, taken in 2001, of the vision that CMS has of the intended software capabilities of its production Data Grid system in 2003, and the expected scaling towards 2007. To capture the expected level of complexity, the vision is sometimes worked out to considerable detail, even though some of these details are likely to be adjusted in future. The document focuses on the relation between the CMS software components and the Grid software components operating inside the 2003 CMS Data Grid system, and contains the architectural constraints for the Grid components.

With respect to scalability, [6] provides comprehensive estimates of the hardware capacity needs. Grid component scalability requirements for the 2007 timeframe can be directly derived from [6]. Table 1 gives a compact indication. Intermediate scalability requirements for 2003 are driven by the CMS ‘20% data challenge’ milestone, for which the work starts in January 2004 with milestone completion in December 2004. In connection to this challenge, the Grid components delivered to CMS by the end of 2003 need to scale to supporting a hardware configuration that has 20% of the projected 2007 capacity.

Tier	CPU capacity	Nr of CPUs	Active tape	Archival tape	Disk
0 (CERN)	455,000 SI95	3000	1540 TB	2632 TB	796 TB
1 (5 all over the world)	105,000 SI95	750	590 TB	433 TB	313 TB
2 (25 all over the world)	26,000 SI95	180	none	50 TB	70 TB

Table 1: Estimates for 2007 CMS hardware capacity needs.

With respect to the complexity of the workload to be supported, [5] is an important source, but the most detailed source to date, especially for the ‘chaotic’ user analysis workload, is the HEPGRID2001 model [7], which gives a baseline for the workload that needs to be handled efficiently by the CMS Data Grid system around 2006. The HEPGRID2001 workload model is a model in terms of ‘virtual’ data products, that is in terms of objects which can be created when requested. A workload model at the level of files can be generated by extending the HEPGRID2001 model with a simulation of the CMS mapping of data products to files sets. The creation of such an extension is a planned activity.

## 3 Division of labor

The CMS requirements for the Grid projects [5] define the following high-level division of labor. Tasks in the 2003 CMS Data Grid system for components provided by the Grid community and Grid projects are as follows:

- Basic management and access interfaces for Grid resources such as storage systems, CPUs, and the network.
- Queuing of Grid jobs, Grid job execution management, integration with local site job submission systems.
- Distributed job scheduling: optimize job execution by efficiently allocating subjobs to sites, moving code to data if possible, by taking into account factors like data location and site loads, and by generating efficient data replication actions to pre-stage data when necessary.
- Error recovery services during job execution, which are configured with CMS-provided error recovery rules and scripts.
- Resource management, monitoring, accounting tools and services.

- Query estimation based on a decomposed job description and the current state of the Grid.
- Efficient wide-area data transfer in terms of files.
- Access by CMS executables to physical Grid files on site-local disk systems via POSIX calls on the regular UNIX filesystem interface.
- File catalog services mapping logical to physical files.
- File set catalog services.
- File replication services in terms of file sets with the ability to implement CMS-configured consistency management policies.
- Data management services: services to configure and maintain backups or mirrors to ensure the long-term availability and integrity of precious data.
- Resource optimization: longer term data migration (often based on user hints or initiated by system operator commands) to balance the use of resources in different Grid sites.
- Grid wide authentication and authorization services, security infrastructure.

Tasks for CMS components and commercial components selected by CMS are as follows:

- Physics analysis tools that provide user interfaces to the Grid services for end-user physicists, interfaces in terms of the high-level physics application semantics.
- Persistency layer that maintains data product values in files.
- Optimizing the strategy for mapping data product values into files.
- Local and remote extraction and packaging of data products to/from files.
- Configuration management for CMS data products and metadata.
- Generation and maintenance of configuration metadata for each file set, creation of services which allow CMS physicists to find the data they need using application-level queries.
- Efficient job decomposition into subjobs.
- Mapping of the application-level job description to a Grid-level description containing the names of input, work, and output file sets.
- Configuration of resource usage and access policies.
- Creation of error recovery rules and scripts.
- Making the tradeoff between pre-staging and dynamic staging of files, initiation of dynamic file staging operations in CMS executables.

## 4 Job model

Physicists get work done on the CMS Data Grid by submitting jobs to it. A CMS Grid job generally consists of many subjobs. Each subjob involves the running of a single CMS executable, with a run time of seconds up to hours. The job model has been worked out to considerable detail (Figure 1), in particular where the use of Grid components is concerned. The emphasis in this job model is on efficient job scheduling by moving code to data, replication and consistency issues, and support for automatic error recovery.

## 5 Virtual data

The GriPhyN project [2] places particular emphasis on virtual data. This is related to the CMS concept of ‘on-demand reconstruction’. ‘Virtual’ refers to the many required data products that may not be physically stored, but exist only as specifications for how they may be computed from other data. Virtuality also means that data can be referred to in a location-independent way. In the context of the GriPhyN project, a detailed vision of 2006 CMS virtual Data Grid, with data handling in terms of ‘virtual data products’, was developed. However it is as yet

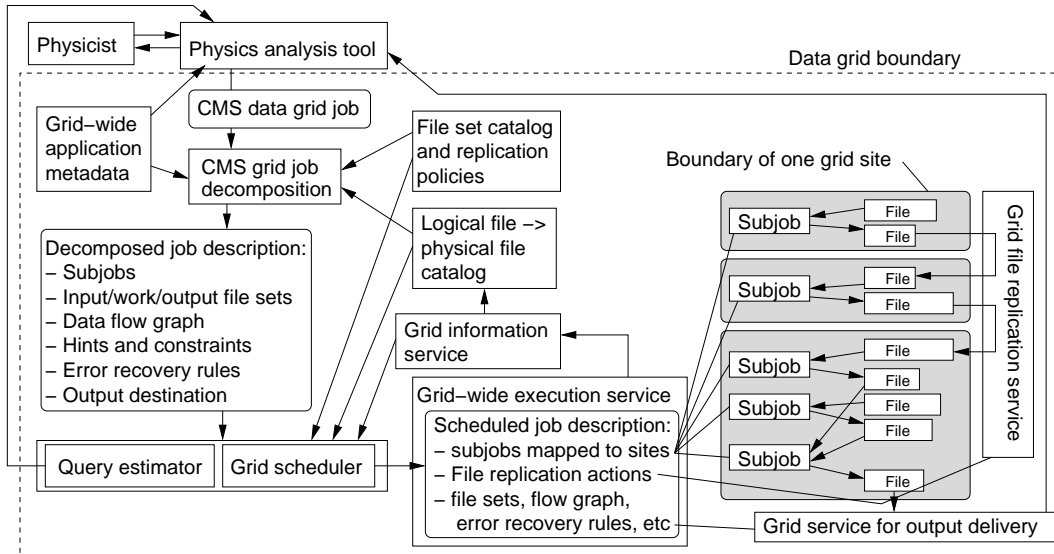


Figure 1: Creation and processing of a single CMS Data Grid job

unclear to what extent the Grid projects should get involved, in the long term, in realizing the different parts of this vision. It was decided [5] that in the time frame now–2003, the Grid projects are not required to deliver components that deal directly with CMS object level data handling, object-granularity data subsetting, or ‘on-demand reconstruction’.

## 6 Conclusions

Concerning the Grid components which are to be created by the Grid projects, and delivered to CMS from now until the end of 2003, [5] defines many requirements, and, even more importantly, many architectural constraints which these components need to take into account. An exact specification of the components to be delivered was however beyond the scope of the requirements activity that CMS conducted. The creation of such exact Grid component specifications is considered to be joint future work between the Grid projects and their customers, with the Grid projects taking the lead in this effort.

## References

- [1] <http://cmsdoc.cern.ch/>
- [2] <http://www.griphyn.org/>
- [3] <http://www.ppdg.net/>
- [4] <http://www.eu-datagrid.org/>
- [5] Koen Holtman, on behalf of the CMS collaboration. CMS Data Grid System Overview and Requirements. CMS Note 2001/037. <http://kholtman.home.cern.ch/kholtman/cmsreqs.ps>, .pdf
- [6] Ian Willers. CMS Interim Memorandum of Understanding: The Costs and How They are Calculated. CMS Note 2001/035.
- [7] Koen Holtman. HEPGRID2001: A Model of a Virtual Data Grid Application. Proc. of HPCN Europe 2001, Amsterdam, p. 711-720, Springer LNCS 2110. CMS Conference Report 2001/006. Web site: <http://kholtman.home.cern.ch/kholtman/hepgrid2001/>