

US-CMS Milestone Status Report

Title: Deployment of the Data Management Prototype

Year: 2005

Original Expected Completion Date: July 2005

Revised Expected Completion Date: February 2006

Description of Milestone: This milestone describes the US-CMS effort to deploy a data management prototype. The prototype is supposed to allow the definition and selection of CMS datasets in the initial release at a scale large enough to support CMS monte carlo production efforts and the current analysis efforts. The product is expected to grow both in functionality and in scale. The three primary components of data management in the CMS computing model are the dataset bookkeeping service (DBS), which is responsible for identifying CMS datasets; the dataset location service (DLS), which identifies at which center the pieces of a dataset are available; and the data transfer components, which are responsible for moving data between centers.

Conclusions:

This milestone has slipped from slipped from July 2005 to an expected release in February 2006. The primary reason for the shift has been a lack of available effort. The data management sub-project in US-CMS has lost three of the primary developers to industry positions over a 12 month period.

In spite of the limited effort and delay in deployment, there is progress to report in all three component development areas. The DBS has a detailed design document and a well-considered schema for the database. Additionally there is a prototype DBS server that is ready for initial deployment. Technical evaluations of grid components have been conducted to assess the usability for the DLS. The data transfer components are the most advanced of the three. CMS, with contributions from US-CMS, has developed a tool called PhEDEx for transfer management. PhEDEx was initially developed to support the 2004 data challenge, but has evolved and improved. PhEDEx was deployed to support CMS simulated event production as well as Service Challenge 3.

The strong coupling of computing activity and data placement in CMS makes data management a critical component. The slip of the prototype deployment is serious and US-CMS will attempt to recover some time during 2006 by increasing the total effort available.