The CMS Data Management System

M Giffels[1], Y Guo[2], V Kuznetsov[3], N Magini[1] and T Wildish[4]

[1] CERN, CH-1211 Genève 23, Switzerland
[2] Fermi National Accelerator Laboratory, Batavia, Il, USA
[3] Cornell University, Ithaca, NY, USA
[4] Princeton University, Princeton, NJ, USA

E-mail: Nicolo.Magini@cern.ch

Abstract. The data management elements in CMS are scalable, modular, and designed to work together. The main components are PhEDEx, the data transfer and location system; the Dataset Booking System (DBS), a metadata catalogue; and the Data Aggregation Service (DAS), designed to aggregate views and provide them to users and services. Tens of thousands of samples have been cataloged and petabytes of data have been moved since the run began. The modular system has allowed the optimal use of appropriate underlying technologies. In this presentation we will discuss the use of both Oracle and nonSQL databases to implement the data management elements as well as the individual architectures chosen. We will discuss how the data management system functioned during the first run, and what improvements are planned in preparation for 2015.

1. Introduction
2. PhEDEx
2.1. PhEDEx
blah blah

3. DBS
3.1. DBS
blah blah

4. DAS
4.1. Data Aggregation Service
blah blah

5. Conclusions

References
[1] *"Egeland R, Wildish T and Metson S 2008 Data transfer infrastructure for CMS data taking"*, XII Advanced Computing and Analysis Techniques in Physics Research (Erice, Italy: Proceedings of Science)
[2] The CMS Data Aggregation System, V. Kuznetsov, D. Evans, S. Metson, doi:10.1016/j.procs.2010.04.172
[3] Keyword Search over Data Service Integration for Accurate Results, V. Zemleris, V. Kuznetsov, to be published in CHEP 2013 proceedings.