

# A highly distributed, petascale migration from dCache to HDFS

*Thursday, 27 October 2011 11:00 (30 minutes)*

The University of Wisconsin CMS Tier-2 center serves nearly a petabyte of storage and tens of thousands of hours of computation each day to the global CMS community. After seven years, the storage cluster had grown to 250 commodity servers running both the dCache distributed filesystem and the Condor batch scheduler. This multipurpose, commodity approach had quickly and efficiently scaled to meet growing analysis and production demands. By 2010, when alternatives to dCache became available in the CMS community, the center was ready to test alternatives that might be a better fit for its hybrid model. HDFS had become widely accepted in the web world and was designed to run in a similarly mixed storage and execution environment. In early evaluations, it performed as well as dCache while also reducing the operational burden. So, in the spring of 2011, the center successfully migrated all of its production data to HDFS with only a few hours downtime.

This migration was one of the largest to date within the CMS community. A unique and highly distributed mechanism was developed to complete the migration while maximizing availability of data to the thousands of jobs that run at Wisconsin each day. This talk presents the migration technique and evaluates its strengths, weaknesses and wider applicability as peers within the CMS community embark on their own migrations.

## Summary

**Primary authors:** Dr MOHAPATRA, Ajit (University of Wisconsin (US)); Mr BRADLEY, Dan (University of Wisconsin (US)); Dr DASU, Sridhara (University of Wisconsin (US)); Mr MAIER, William (University of Wisconsin (US))

**Presenter:** Mr MAIER, William (University of Wisconsin (US))

**Session Classification:** Storage & File Systems

**Track Classification:** Storage & Filesystems