CMS Service Challenge 3 Planning

Lassi A. Tuura
Northeastern University

Service Challenge Context

- CMS schedule at high level
  - LHC start in April 2007
  - Last major data challenge DC06 in summer/autumn 2006
  - Magnet test end of 2005, cosmic challenge beginning of 2006
  - *Service challenge 3: July - December 2005*
- Compact computing and software planning overview
  - Baseline system assumes thin grid middleware, experiments significant stake-holders with major input to operation and choices made
  - Full computing system essentially being put into operation now
  - New event data model / framework being deployed this year
- Currently identified top three issues and/or missing functionality
  - Priorities and share allocation in both workload and data management
    - Data management aiming at first solution this summer
  - Comprehensive monitoring to understand resource usage
    - To guide re-prioritisation and policy changes
  - Operational stability and babysitting required

Integration Program of Work

- Establish an environment where we can work and test components
  - Integration cycles of approximately three months
  - Reasonably complete testbed environment: smaller-scale replica of the entire system, complete set of services expected to be always present
  - Measure performance of important components: data transfers, data serving, workload management components (resource broker, computing elements)
  - Becomes next production service
- Success is mandatory
  - Main focus on alternatives with reasonable expectation of success
  - We cannot afford total failure on any significant component
  - Clarify long-term goals and temporary solutions
    - Example: Which data serving solutions are temporary solutions for current problems, which are the long-term strategy? Make clear to decision makers!

Service Challenge 3: Test Facility

- Exercise roughly realistic scenario, but small amount of data
  - Data produced centrally and distributed to Tier 1 centres (MSS)
  - Strip jobs at Tier 1 produce analysis datasets
    - Approximately 1/10th of original data, also stored in MSS
  - Analysis datasets shipped to Tier 2 sites for analysis, published locally
  - Tier 2 sites produce MC data, ship to Tier 1 MSS
    - May not be the local Tier 1
  - Transfers between Tier 1 sites
    - Analysis datasets, 2nd replica of raw
- Volumes + sites
  - 50 TB from CERN T1 to at least two T1s, plus smaller ones (~10 TB)
  - 5 TB to T2s, at least one per T1
  - 5-10 TB T1/T1 analysis dataset transfers
  - 50 TB T1/T1 2nd raw replica transfers

Service Challenge 3: Services I

- Expect to treat this as a continuous testbed environment
  - All services available at all times, we should be able to execute integration tests while throughput tests take place
  - Expect reasonably complete testbed early enough
  - With reasonable computing resources
- Main points of focus
  - Structured data flow using PhEDEx transfers
  - Transfer and data serving throughput
    - Simultaneous data import, export and analysis
  - Production job throughput at Tier 2 sites
  - Exercising job pull model
  - O(10M) events / month continued (to be verified!)
- Specific services: next slides
  - Expect sites to negotiate with us what is used where

Service Challenge 3: Services II

- Data transfers: PhEDEx; to be installed at each participating site
  - Underlying transfers expected to be either SRM or globus-url-copy
  - If sites feel strongly about EGEE FTS, should discuss use with us *now*
- File catalogue
  - POOL API, local, relational; MySQL / Oracle default; LFC and Globus RLS are an option, sites should contact CMS if they expect to use these
  - We do not expect to test EGEE Fireman
- General environment
  - Data serving infrastructure -- dCache, Castor or xrootd
  - Computing element, job submission (UI)
  - Output harvesting for transfers (CMS agents, can use UI-type machine)
  - CMS software installation and publishing into the information system
  - Bookkeeping / monitoring databases for production (CMS BOSS, MySQL)
  - Above-mentioned file catalogue
  - PubDB or successor

Service Challenge 3: Schedule

- July: SC3 throughput phase
  - T0/T1/T2 simultaneous import/export
  - To and from tape at T1s — *application level end-to-end transfers*
  - Real files, real storage
  - Expect to test other things on testbed at the same time
- September: service phase 1 — modest throughput
  - Demonstrate bulk data processing, simulation at T1, T2s
    - Requires software, job submission, output harvesting, monitoring, …
    - Not everything everywhere, something reasonable at each site
- November: service phase 2 — modest throughput
  - Phase 1 + continuous data movement
  - New task-queue based job system for production
    - Already in September if available then
  - Precursor for next production service

More Information

- CMS computing
  - David Stickland (david.stickland@cern.ch)
  - Lothar Bauerdick (bauerdick@fnal.gov)
  - Computing TDR: http://cmsdoc.cern.ch/cms/cpt/tdr
  - SC3 wiki: http://uimon.cern.ch/twiki/bin/CMS/SWIntegration
- Data management — Peter Elmer
  - cms-dm-developers@cern.ch
- Workload management — Stefano Lacaprara
  - cms-computing-wm@cern.ch
- Production — Tony Wildish
  - cms-production@cern.ch
- Data transfers — Lassi Tuura, Tim Barrass
  - http://cern.ch/cms-project-phedex
  - cms-phedex-developers@cern.ch