

GDM drivers for CDF

R. Snider
for CDF offline

GDM meeting
April 12, 2005

Listed CDF drivers

- Production use of SAM for MC, reconstruction
- Production use of SAM on CAF, dCAFs
- Optimize use of tape resources
- Support for 2x, 3x data taking rate
- Tape migration support
- Use of OSG/LCG/Fermigrid resources
- Load balancing on analysis cluster
- Accounting for Finance Committee

Production use of SAM for MC

- Use SAM to import, catalog MC files produced off-site
- Priority: high
- Status/timeline
 - Tools tested, ready
 - Replace DFC store with sam store: 1 day
 - Auto-destination config, integration testing: 2 weeks
- GDM dependencies:
 - SAM (consultation)
 - Grid services (1 year time scale)

Production use of SAM for MC

- Non-GDM dependencies:
 - Experiment (user training to access data via SAM)
- Risks
 - Effort, expertise limited: schedule risk

Production use of SAM for reconstruction

- Use SAM to automate farm processing
- Priority: high
- Status/timeline
 - Validate single stream processing: **completed**
 - Prototype farm shadow processing “Stream A” now
 - Begin production processing of Stream A: **April 18**
 - Requires 6.1.0 production executable
 - Validate split stream processing: **April 25**
 - Begin production processing of split stream: **about May 2**
 - Requires final calibrations
 - Deploy production farm upgrade
 - **Defer details to Jerry Guglielmo's talk next week**

Production use of SAM for reconstruction

- GDM dependencies
 - SAM (consultation)
 - Grid services (1 year time scale)
- Non-GDM dependencies
 - Experiment (production exe, calibrations, user training)
- Risks: mainly schedule
 - Validation: low risk
 - Start of production processing: may slip by few weeks
 - Production upgrade deployment
 - Critical path: head node procurement

Production use of SAM on CAF, dCAFs

- Use SAM for general user analysis
- Priority: high
- Status/timeline (detailed report next week)
 - Production deployment: **May 13**
 - Detailed effort loaded schedule and milestones
- GDM dependencies
 - Grid services (1 year time scale)
- Risks
 - Schedule risk in production deployment
 - Grid solution not yet specified

Optimize use of tape resources

- Efficient utilization of existing tape slots
- Priority: high
- Status/timeline
 - Tape re-cycling: in progress
 - Tape consolidation
 - Aggregation for user tapes, MC import: **May**
 - Completed as part of SAM for MC project
 - Consolidation of existing tapes: **Summer (?)**
 - One-pass production processing
 - Requires production executable, final calibrations
 - Data compression

Optimize use of tape resources

- GDM dependencies
 - Lower storage (re-cycling, consolidation plan)
 - SAM (consultation for auto-destination)
- Non-GDM dependencies
 - Experiment
 - Development of event compression, data reduction scheme
- Risks
 - Tape consolidation plan unspecified
 - Still evaluating potential yield, effort required to recover tapes
 - Compression scheme not yet defined.
 - Yield, effort required unknown

Support for 2x data taking rate

- Reconstruction, storage, analysis capacity to support data logged at rates up to 40 MB/sec
- Priority: high
- Timeline
 - Experiment can currently log data at 30 MB/sec to 40 MB/sec
 - Has not yet chosen to do so
 - Offline is ready

Support for 3x data taking rate

- Reconstruction, storage, analysis capacity to support data logged at rates up to 60 MB/sec
- Priority: medium
- Status/timeline
 - Logging rate increase expected in FY2006
 - Re-evaluate plan: **summer 2005**
 - Planned farm upgrades will cover processing requirements
 - Data storage is main issue

Support for 3x data taking rate

- GDM dependencies
 - Lower storage (tape storage plan)
- Risks
 - Run out of space in tape archives

Tape migration support

- Need a new tape technology
 - Evaluate cost / benefit of density migration
 - Recover slots at the cost of tapes, operations
- GDM dependencies
 - Tape migration support
- Risks
 - Schedule tied to availability of new technology

Use of OSG/LCG/Fermigrid resources

- Needed to maintain access to existing remote resources
 - Also opportunistic access as available
- Priority: medium
 - Need access to LCG/OSG resources by 2006

Use of OSG/LCG/Fermigrid resources

- Status/timeline (details presented at CDF status report on March 8)
 - Formalize a plan, schedule: draft exists
 - Review the plan: May, June (?)
 - Demonstrate Condor glide-in for CDF jobs: done
 - Establish glide-in CAF at Tier 1 center (CNAF): in beta test now
 - Test scalability, stability on large LCG/OSG clusters
 - May provide low-cost solution for portion of requirements
 - Requires access to CDF software, dedicated data storage, head node
 - Enable CAF user services directly on basic grid services
 - Centralized queue for MC production
 - Support interactive services for analysis jobs with generic grid submission

Use of OSG/LCG/Fermigrid resources

- GDM dependencies
 - Grid services
 - Fermigrid (authentication)
 - SAMgrid
- Non-GDM dependencies
 - Experiment
- Risks
 - Technical risks in various areas
 - Many issues to be resolved
 - Need to define schedule

Load balancing on analysis cluster

- Priority: low
- Should be considered as part of overall grid strategy
 - Will adopt suitable brokering tools as they become available
- Current load balancing scheme should be adequate for next year
 - User self-balancing
 - Administrative control via dataset location

Accounting for Finance Committee

- Is this a GDM driver?