LCG-LHCC Referees Meeting 24 Sep 2007

Status of Dress-Rehearsal Preparations

Material taken mostly from the WLCG workshop and Chep07 Victoria Sep 1-7 2007 (also recent CMS week). Additional information in (blue).

- Experiment plans/readiness (including cosmics runs)
 - •ALICE, ATLAS, CMS, LHCb
- Services readiness
- Site readiness
- Common Computing Readiness Challenge 2008

Dress Rehearsal Elements (I)

- General purpose of the Dress Rehearsal
 - Combined tests of all steps needed to produce the ESDs from RAW
- Data flow and systems concerned (I)
 - Generated and real data from detector commissioning: registration in CASTOR2 + Grid File Catalogue - DAQ/WLCG services/Offline
 - MC RAW for the detectors not yet being commissioned
 - Cosmics, pulser, other data for detectors already in the cavern
 - Registration in CASTOR2 and Grid FC already well tested and working
 - Replication of RAW to T1s Offline/WLCG services
 - Synchronous to RAW registration, the RAW is replicated to a T1
 - Replicate using FTD/FTS utilities
 - Replication shares are determined from the contribution factors of the T1s
 - The replication is random, depending on resources/channel availability
 - Replication with FTS is (has been) exercised, however not with FTS v.2.0 and SRM v.2.2

Dress Rehearsal Elements (III)

- Major steps and systems concerned (III)
 - Second pass reconstruction at T1 Offline/WLCG Services
 - After pass 0 is complete and new condition object are available in OCDB
 - Second iteration of conditions data, derived from pass 1, and eventually refinement of the reconstruction code
 - Triggered by a successful T0 processing
 - Produces final ESDs
 - As a part of the same job AOD production
 - Automatic validation procedures
 - A copy of the ESDs is stored at each T1
 - Asynchronous data flow to CAF, registration and analysis -Offline/WLCG services
 - Parts of RAW (on demand), calibration and alignment runs, parts of ESDs copied to CAF disk pool
 - Detector expert special calibration tasks
 - First and second pass ESDs analysis



FDR phase 1

Data flow

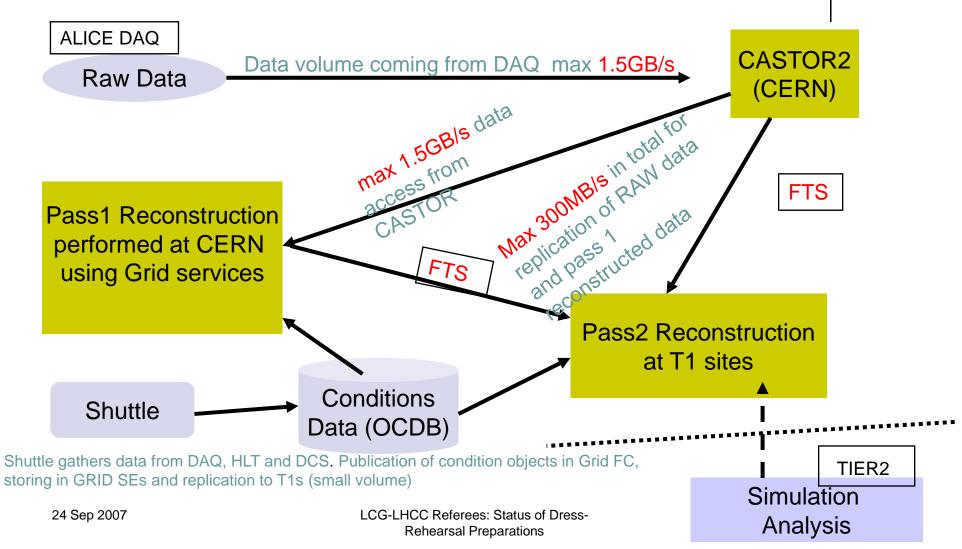
- Several detectors will begin commissioning in September
 - TPC, TRD, PHOS, HMPID
 - Cosmics, pulsers, laser tracks, black events, simulated RAW injection
 - An element of the TRD readout (Global Tracking Unit) will be used to inject simulated RAW see next slides (probably not ready for first phase)
 - Preparation of AliRoot for the reconstruction of these events well under way
 - The target transfer rates to CASTOR2/T1 depending on inclusion of detectors/type of data

CAF facility at CERN will receive portions of the data

- Automatically for special run types (f.e. calibration), small portions of the data will be copied on demand
- Access to data reserved for detector experts (10-20) at any given moment
- Events will be of various sizes, depending on the data type
- The data will be accessed in a random pattern
- The data will be copied from CASTOR2 to the CAF disk pool, access is only from CAF pool

Transfers for the phase 1





Grid services: general conditions



- ALICE requires the services as provided during the Data Challenges
 - VOBOXES deployed at all ALICE T0-T1-T2 sites
 - ALICE requires the latest version of the gLite3.1
 VOBOX (will use current LCG RB to start with)
 - Still under configuration, testing and deployment
 - This new configuration is mandatory to migrate to WMS3.1
 - All sites will have to be updated to the latest version
 - Pilot version in production deployed in voalice03@CERN
 - FTS service from T0-T1
 - This exercise tests also SRM2.2
 - FTS channel sharing and rates as during the T0-T1 exercise in 2006/2007 (p-p running at 100 MB/sec)



Grid services: CASTOR

Considerations

- The interface has been successfully tested so far, minor problems are being fixed
- New xrootd expert in ARDA from September 2007
- The xrootd-CASTOR interface will be part of the next major Castor release which will be deployed during the 1st half of September 2007 (now scheduled for 26 Sep)
- The xrootd-CASTOR setup needs to be defined in detail and the responsibilities between ALICE and IT clarified (ongoing – server hardware being setup now)



Plan of the FDR

- Mid September 2007
 - Strategy and setup fully defined (done)
- October 2007 FDR Phase 1
 - Cosmic Rays data taking, calibration runs, special runs from detector commissioning (already started including registration in CASTOR2)
 - Registration in CASTOR2/Replication T0-T1, Pass 1 reconstruction, expert analysis
- November-end 2007 FRD Phase 1+2
 - All elements of Phase 1
 - Pass 1 and Pass 2 reconstruction
 - Conditions data with Shuttle (online conditions data collector)
- February-May 2008 FDR Phase 1+2+3
 - All elements of Phase 1+2
 - Gradual inclusion of DA and QA (online detector algorithms and quality assurance)

ATLAS Cosmics Runs



M5 and M6 and beyond

- M4 Ran August 23 to Sep 3rd with good results
- M5: October 16 − 23
- M6: end December
 - With incremental goals
 - Reprocessing challenges in between runs
- In 2008: cosmic ray data taking is the default





- Simulated events injected in the tdaq
- Realistic physics mix
- Bytestream format including luminosity blocks
- File & dataset sizes as expected for real data
- Realistic trigger tables
- datastreaming
- Use of conditions database
- Data quality-, express line-, callibration- running
- T0 reconstruction: ESD, AOD, TAG, DPD
- Exports to T1&2's
- Remote analysis





- at the T1's
 - Reprocessing from RAW → ESD, AOD, DPD, TAG
 - Remake AOD from ESD
 - Group based analysis → DPD (Derived Physics Data)
- At the T2&3's
 - Root based analysis
 - Trigger aware analysis with Cond. and Trigger db
 - No MC truth, user analysis
 - MC/Reco production in parallel





Round 1

- Data streaming tests DONE
- 2. Sept/Oct 07 Data preparation STARTS SOON
- 3. End Oct07: Tier 0 operations tests
- 4. Nov07-Feb08. Reprocess at Tier1, make group DPD's

Round 2 ASSUMING NEW G4

- 1. Dec07-Jan08 New data production for final round
- Feb08 Data prep for final round using
- 3. Mar08. Reco final round ASSUMING SRMv2.2
- 4. Apr08. DPP prod at T1's
- 5. Apr08 More simulated data prod in preparation for first data.
- 6. May08 final FDR (proposed to be part of ccrc'08)

See also Dario's slides later on Combined Data Management

CMS CSA07

The preparation tests in CMS are call Computing Software and Analysis Challenges (CSA07)

- **→** The goal is to exercise aspects of the computing model and the software development program with analysis activities and users
- Dedicated tests of components do not show interference problems
- **→** CSA07 is intended to exercise the computing model at greater than 50% of the target for 2008
- The CSA06 challenge was an exercise at 25% of scale
 We have a number of elements that have not been exercised previously
- ➡ Integration of the computing components up to storage manager
- → Some data transfer channels: Tier-1 to Tier-1 Transfers, Tier-2 to Tier-1
- **→** Balancing of simulation and analysis

Desire to demonstrate computing and offline tools with a diverse and active user community

Previous exercises have relied heavily on load generators

CMS

Basic Scaling Items Checked in CSA07

Service	2008 Goal	CSA07 Goal	CSA06 Goal	Status 2006
Tier-0 Reco Rate	150Hz - 300Hz	100Hz	50Hz	Achieved
Network Transfers between T0-T1	600MB/s	300MB/s	150MB/s	Achieved (6/7 cont.)
Network Transfers between T1-T2	50-500 MB/s	20-200 MB/s	10-100 MB/s	Achieved (15 sites)
Network Transfers T1-T1	I00MB/s	50MB/s	NA	Not Attempted
Job Submission to Tier-1s	50k jobs/d	25k jobs/d	12k jobs/d	3k jobs/d
Job Submissions to Tier-2s	150k jobs/d	75k jobs/d	48k jobs/d	Achieved
MC Simulation	1.5 10^9 events/year	50M per month	NA	Not Attempted

lan FiskFermilab CHEP Conference September 4, 2007 3



Data From CERN

Once events are reconstructed they head for Tier-I centers

- Total rate of reconstruction is expected to be 100Hz
 - Application is more complete than CSA06.
 - Large variations in the time of reconstruction. Memory requirements are higher than expected
- Expected aggregate rate is 300MB/s

Once data gets to Tier-I centers it is the hot copy

- There are very few resources at CERN except for the Tier-0 and the small queue at the CAF
- Skimming and Re-Reconstruction will happen at Tier-1 centers
 - Reprocessing will involve creating the samples from the individual simulation constituents
 - TTbar, Jets, DrellYan, etc.

lan FiskFermilab WLCG Workshop Victoria September 2, 2007 8



Status

We believe all the hardware at the sites is available to complete the challenge

- Need roughly 2500 batch slots at CERN (3000 are available)
- Need functional tape resources at a nominal Tier-1s (600slots and 150TB storage)
 - Large variation in Tier-1 capacity
- Need credible disk mass storage at Tier-2s (150slots and 20TB)
 - Even larger variation here than in the Tier-2s
- Storage scales as time not as percentage of the final system

We expect a software release to start the pre-challenge steps on Monday

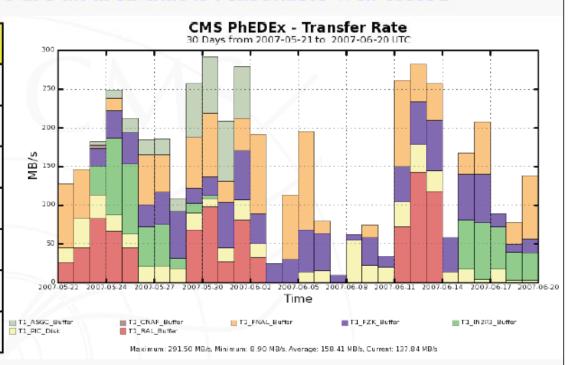
- This is roughly three weeks later than we expected
- Application memory targets are higher than expected
 - After some effort from the developers should still work



Status

CERN to Tier-I transfers are an area that is reasonable well tested

Site	Rate (MB/s)	
ASGC	26	
CNAF	37	
FNAL	105	
FZK	26	
IN2P3	32	
PIC	13	
RAL	26	



An open item is the ability to reliably write data to tape at all centers

 CSA07 also includes some elements specifically designed to exercise the ability to retrieve data from tape

Ian FiskFermilab WLCG Workshop Victoria September 2, 2007

CMS CSA07 - Schedule

We need to convert the simulated events to looking like events that came from the HLT farm

- → This is divided into 3 steps and we expect this will take about three weeks
- → Start hopefully Monday (started 10 September)

Begin Tier-0 reconstruction activities on September 24

Simulation at the Tier-2s will continue from the beginning

About a week after the beginning we expect to start the skimming at the Tier-1 sites

→ Data movement and analysis access

By two weeks we expect to begin reprocessing at Tier-1 sites



Outlook

Event production for the challenge is finished

This went nicely. Largest number of events produced per month

Site commissioning and data channel commissioning is in progress

- The SAM tests are a good indicator for the eventual success of user jobs
- CMS computing model call for the mesh of Tier-1 to Tier-2 transfers to work efficiently. Testing in progress

Still testing a number of workflows

Software releases and techniques are new

Previous challenges have validated elements of the CMS computing model

- Also gave directions for development and refinement
- It will be a busy fall.

lan FiskFermilab WLCG Workshop Victoria September 2, 2007 20



Connectivity

DDT team has done a nice job

- Sites have also done a nice job following up on problems
- Things are improving

Even so, on any reasonable challenge start schedule we are going to go into the challenge without even close to a full set of links

- The mesh of Tier-I to Tier-2 connectivity will be restricted
- Tier-I to Tier-I mesh is really limited

I expect the CERN to Tier-I transfers will work

Tier-I to Tier-I is not in good shape

Currently 3 l Tier-2s have at least one download link

- One has 4
- Several have 3 and a number have 2

CSA07 Meeting

September 19, 2007



Outlook

We are going into the challenge with roughly the number of events we wanted

- Tier-I and Tier-2 resources are sufficient to complete the activity, I believe
- We have some extra flexibility because we have some reasonable size Tier-2s
 - The challenge storage requirements scale as the time, so it's not 50% of the data for a year (schedule is for 30 days)

We are not going in with as complete a set of commissioned network links as we would have liked

- Things may improve during the challenge
- May have to be more creative in how we move data around

CSA07 Meeting

September 19, 2007





Status of DC06

Reminder:

- DC06 is a generic name for activities that will last until end 2007 (physics book simulation, reconstruction, analysis)
- Two-fold goal: produce and reconstruct useful data, exercise the LHCb Computing model, DIRAC and ganga
- To be tested:
 - ☆ Software distribution

 - □ Data export from CERN (FTS) using MC raw data (DC06-SC4)
 - - For staged and non-staged files

 - Batch analysis on the Grid (data analysis and standalone SW)
 - → Datasets deletion
- LHCb Grid community solution
 - ☆ DIRAC (WMS, DMS, production system)
 - ☆ ganga (for analysis jobs)



WLCG Workshop 1-2 Sept 2007, Victoria, BC

.





DC06 phases (cont'd)

- February 2007 onwards
 - Background events reconstruction at Tier1s
 - - were no longer on cache, hence had to be recalled from tape
- June 2007 onwards
 - Background events stripping at Tier1s

 - → Accesses the 40 corresponding MC raw files for full reconstruction of selected events
 - □ DST distributed to Tier1s
 - Originally 7 Tier1s, then CERN+2
 - need to clean up datasets from sites to free space

Ph.C.

WLCG Workshop 1-2 Sept 2007, Victoria, BC





What is still missing?

- o gLite WMS
 - Several attempts at using it, not very successful
 - ★ Still not used in production (not released as such...)(Released 18 Sep)
- Full VOMS support by middleware
 - Many problems of mapping when using VOMS
 - Was working, had to move back to plain proxies due to dCache problems
 - → Problems of LFC registration in existing directories
 - e.g. when moving to pool accounts for production group
 - ☆ No castor proper authentication (i.e. no security for files)
- o SRM v2.2
 - Tests ongoing
- Agreement and support for generic pilot agents
 - Essential for good optimisation at Tier1s
 - ♣ Prioritisation of activities (simulation, reconstruction, analysis)



WLCG Workshop 1-2 Sept 2007, Victoria, BC





Plans and outlook

- Re-processing of background
 - Just restarted (software fault found): 6,000 jobs
 - Stripping will follow: 3,000 jobs
- SRM v2.2 tests
 - Ongoing, many issues found and fixed

 - → Difficult to get space tokens and corresponding pools properly configured
- Analysis
 - Rapidly growing (batch data analysis, ROOT scripts for fits, toy MC)



WLCG Workshop 1-2 Sept 2007, Victoria, BC





Plans (cont'd)

- Conditions DB test
 - Deployed and 3D streaming working at all Tier1s
 - Stress tests starting (Bologna)
 - Usage in production during Autumn
- LFC replication
 - Requested at all Tier1s
 - In production for over 6 months at CNAF
- Dress rehearsal(s)
 - Assuming it means producing data at Tier0, shipping to Tier1s and processing there...
 - Pit Tier0: ongoing
 - Autumn: include Tier1 distribution and reconstruction
 - LHCb welcomes a concurrent DR in Spring 08
 - ☆ Will use 2 GB "raw" (simulated) files

WLCG Workshop 1-2 Sept 2007, Victoria, BC



Conclusions

- LHCb using WLCG/EGEE infrastructure successfully
 - Eagerly waiting for generic pilots general scheme
- Still many issues to iron out (mainly DM)
 - SE reliability, scalability and availability
 - Data access
 - SRM v2.2
 - SE migration at many sites
- Trying to improve certification and usage of middleware
 - LCG-AA deployment, production preview instances
- Plans to mainly continue regular activities
 - Move from "challenge mode" to "steady mode"



WLCG Workshop 1-2 Sept 2007, Victoria, BC

WLCG Workshop Plenary by John Gordon (RAL) The State of the services: Critical issues (1/2)

dCache

- SRM calls must comply with the specification
- Clear configuration and management documentation needed
 - Multi-VO configuration
 - Default space
 - Separate pools for different space tokens
- TxD1 space tokens do not recover space from deleted files
- More management tools needed
 - List associations between files, tokens, disk pools and tapes
 - Space management, reconfiguration

CASTOR

- SRM calls must comply with the specification
- Thread bookkeeping issue ("too many threads")
- Race condition ("current user does not own this request")
- Get requests slow down with time

The State of the services: Critical issues (2/2)

- DPM
 - Occasional srmv2.2 daemon crashes
- StoRM
 - Space tokens not independent of paths
 - SRM calls must comply with specification
- GFAL/lcg-utils
 - Light version not yet fully independent of BDII
 - More robust directory creation algorithm needed
 - Incorrect handling of TURLs returned by CASTOR
- FTS
 - None

The State of the services: SRMv22

- All implementations delivered for testing
- Bugs, inconsistencies, configuration issues
- Critical issues identified
- Experiment testing one month late
- Good progress but not there yet.

Non-critical for 2007 but need addressing:

- Job Priorities
- User DN Accounting
- Fully Qualified Attribute Name Accounting
- Glexec on WN

The State of the services: Conclusions

- Are we ready?
 - Yes, but
- We are ready for the dress rehearsals
 - But partly because the experiments are not relying on anything that isn't there now.
- Experiments are running large numbers of jobs
 - Despite complaints about RBs etc
- Data Transfer is still the weak link.
 - Components work but not successfully stress tested
 - Despite many years of planning
- Much work to be done for next year

Site Readiness

LCG Management Board Tier0/1 procurement milestones

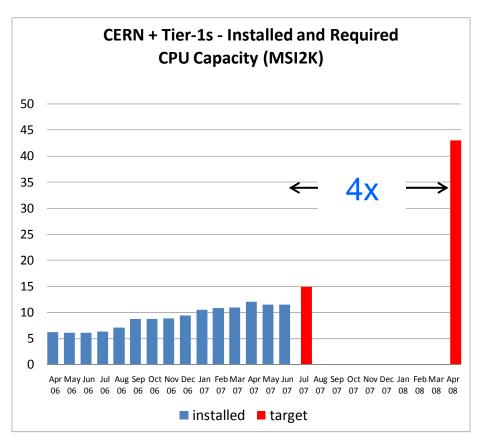
See https://twiki.cern.ch/twiki/pub/LCG/QuarterlyReports/QR_2007Q2.pdf

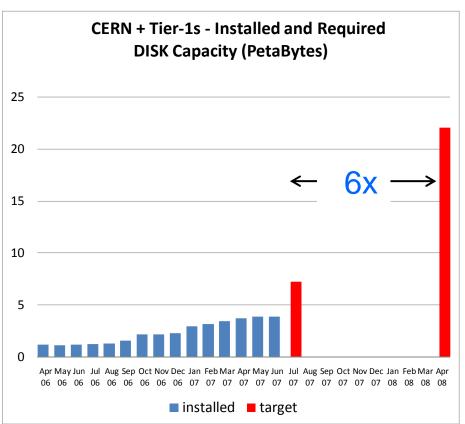


Sites well aware of 1 April 2008 milestone but it is a big increase from now. CERN has major acquisitions running and plans to be ready on time. Detailed ramp-up plans not yet known from all Tier1 sites. (Will they have enough for the dress rehearsals happening before the 1 April 2008? First feedback is from IN2P3 that they have 40% of the pledged increase already installed).

Site Readiness

Steep ramp-up still needed before first physics run

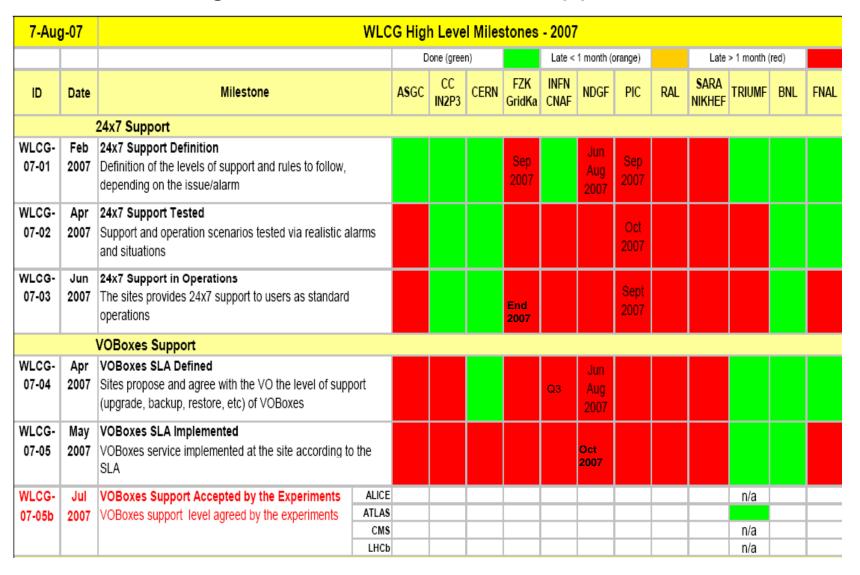




Evolution of installed capacity from April 06 to June 07
Target capacity from MoU pledges for 2007 (due July07)
and 2008 (due April 08)

Site Readiness

LCG Management Board Tier0/1 support milestones



Site Readiness Issues from WLCG workshop

- State of Tier 1/2 hardware resources for 2008 startup compared with pledges. Longer term planning readiness for future major expansion (2010 and beyond).
- The long time it will take for sites and experiments to validate the massive site capacity increases scheduled to be in place and available to the experiments by April 2008.
- Tested readiness of magnetic tape infrastructure.
- Tested readiness of site data management layer to disk and tape including for full rate multi-VO operation.
- Formal planning for 24 by 7 operations ready (including for staffing) and documented to the experiments and other sites. (PIC, NDGF planning to deploy by November)
- Experiment VO boxes in place with Service Level Agreements implemented. (ATLAS have accepted FZK SLA. LHCB working on an SLA for all Tier1. NDGF planning for October)
- Experiment local site support in place with clearly defined roles and interfaces to site operations.
- These issues support the need for multi-experiment dress rehearsals in the first half of next year: the Common Computing Readiness Challenge 2008 as proposed during the workshop by CMS and ATLAS.

Motivation and Goals (CMS CCRC'08)

What if:

- LHC is operating and experiments take data?
- All experiments want to use the computing infrastructure simultaneously?
- The data rates and volumes to be handled at the Tier0, the Tier1 and Tier2 centers are the sum of ALICE, ATLAS, CMS and LHCb as specified in the experiments computing model
- Each experiment has done data challenges, computing challenges, tests, dress rehearsals, at a schedule defined by the experiment
- This will stop: we will no longer be the master of our schedule...
 Once LHC starts to operate.
- We need to prepare for this ... together

A combined challenge by all Experiments should be used to demonstrate the readiness of the WLCG Computing infrastructure before start of data taking at a scale comparable to the data taking in 2008.

This should be done well in advance of the start of data taking on order to identify flaws, bottlenecks and allow to fix those.

We must do this challenge as WLCG collaboration: Centers and Experiments

CCRC'08 - Proposed Scope (CMS)

- Test data transfers at 2008 scale:
 - Experiment site to CERN mass storage
 - CERN to Tier1 centers
 - Tier1 to Tier1 centers
 - Tier1 to Tier2 centers
 - Tier2 to Tier2 centers
- Test Storage to Storage transfers at 2008 scale:
 - Required functionality
 - Required performance
- Test data access at Tier0, Tier1 at 2008 scale:
 - CPU loads should be simulated in case this impacts data distribution and access
- Tests should be run concurrently
- CMS proposes to use artificial data
 - Can be deleted after the Challenge

CCRC'08: Constraints, Preconditions and Schedule (CMS)

- Mass storage systems are prepared
 - SRM2.2 deployed at all participating sites
 - CASTOR, dCache and other data management systems installed with appropriate version
- Data transfers are commissioned for CMS
 - Only commissioned links can be used
- Participating centers have 2008 capacity (this will be unlikely for the February pre-challenge – MoU commitment is April 1)
- Duration of challenge: 4 weeks
- Based on the current CMS schedule:
 - Window of opportunity during February 2008
 - In March a full detector COSMICS Run is scheduled
 - With all components and magnetic field
 This is the the first time with the final detector geometry
- Document performance and lessons learned within 4 weeks.

CCRC'08 - Proposed Organization (CMS)

Coordination: (1+4+nT1)

- WLCG overall coordination (1)
 - Maintains overall schedule
 - Coordinate the definition of goals and metrics
 - Coordinates regular preparation meetings
 - During the CCRC'08 coordinates operations meetings with experiments and sites
 - Coordinates the overall success evaluation
- Each Experiment: (4)
 - Coordinates the definition of the experiments goals and metrics
 - Coordinates experiments preparations
 - Applications for load driving (Certified and tested before the challenge)
 - During the CCRC'08 coordinates the experiments operations
 - Coordinates the experiments success evaluation
- Each Tier1 (nT1)
 - Coordinates the Tier1 preparation and the participation
 - Ensures the readiness of the center at the defined schedule
 - Contributes to summary document

CCRC'08 Proposal (WLCG)

- The need for a **Common Computing Readiness Challenge** has been clearly stated by ATLAS, CMS and WLCG Service Coordination.
 - Ideally, ALICE & LHCb should also participate at full nominal 2008 pp rates
- The goals & requirements such as production SRM v2.2 are common
- Two slots have been proposed: Feb (by CMS) & May '08 (by ATLAS)
- Given the goals & importance of this challenge, foresee to use <u>both</u> slots
- 1. Feb: pre-challenge; ensure pre-conditions are met; identify potentially problematic areas
 - Can be <100% successful
- 2. May: THE challenge;
 - Must succeed!

Must be pragmatic – focus on what can (realistically) be expected to work!

- Need to carefully prepare which means thorough testing of all components and successive integration prior to the full challenge
- In addition to the technical requirements, must ensure adequate computing and people resources are available throughout these periods
 - Neither of these slots is optimal in this respect, but when is?
 - Need to understand how to provide <u>production coverage</u> at all times!



In Summary: Dress Rehearsals Draft Timeline

Month	ATLAS	CMS	ALICE	LHCb
Sep'07	FDR 1	CSA07		MC->T1s
Oct'07	FDR 1	CSA07	FDR I	MC->T1s
Nov'07	FDR 1; Cosmics	Cosmics	FDR II	MC->T1s
Dec'07	FDR 1; FDR 2		FDR II	PROD
Jan'08	FDR 1; FDR 2			PROD
Feb'08	CCRC; FDR 1; FDR 2	CCRC	CCRC; FDR	CCRC; PROD
Mar'08	FDR 2; Cosmics	Cosmics	FDR III	
Apr'08	FDR 2		FDR III	
May'08	CCRC; FDR 2	CCRC	CCRC; FDR	CCRC
Jun'08				
Jul'08				
Aug'08				
Sep'08				