



The ALICE Data Challenges

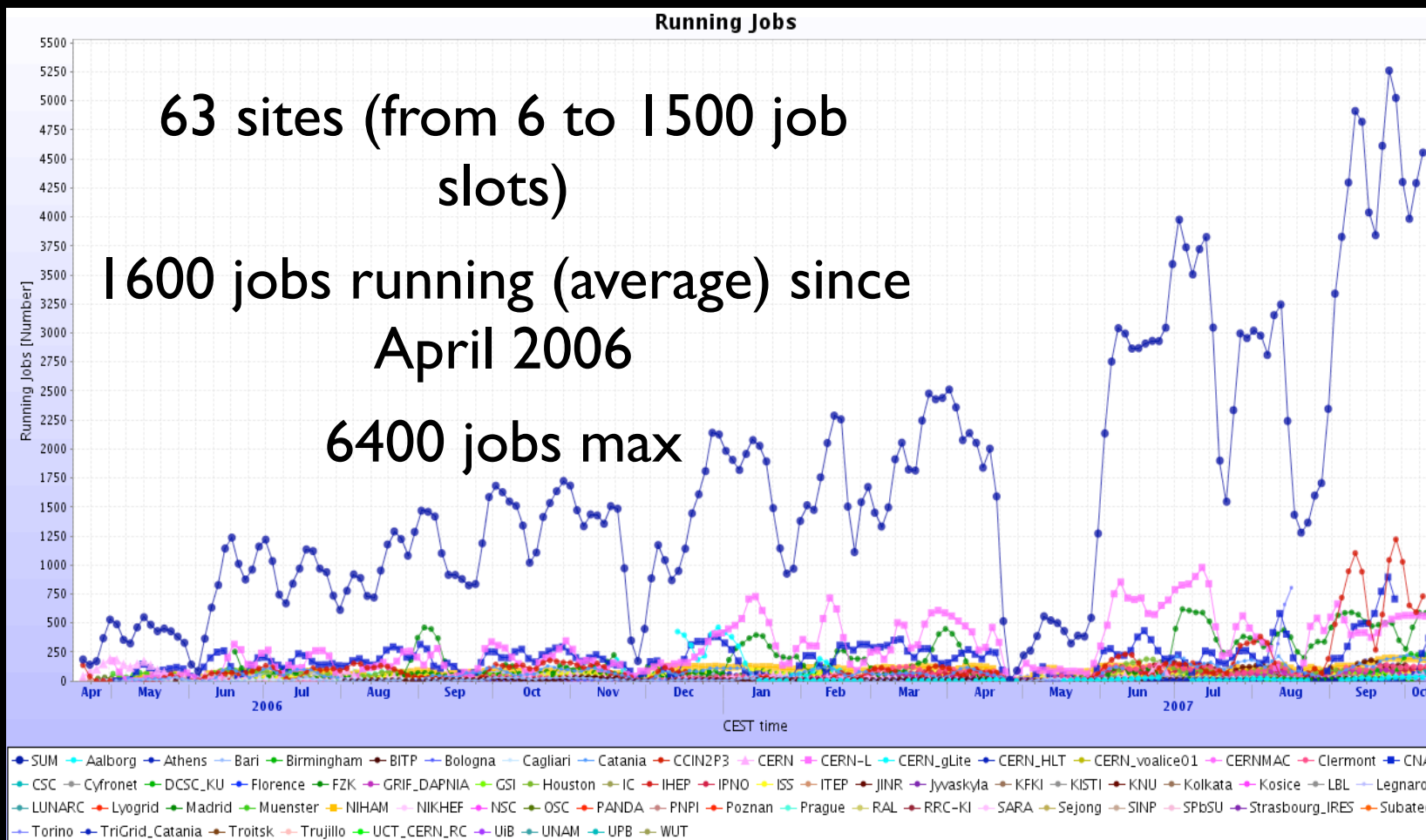
Federico Carminati
GDB October 10, 2007

Some statistics

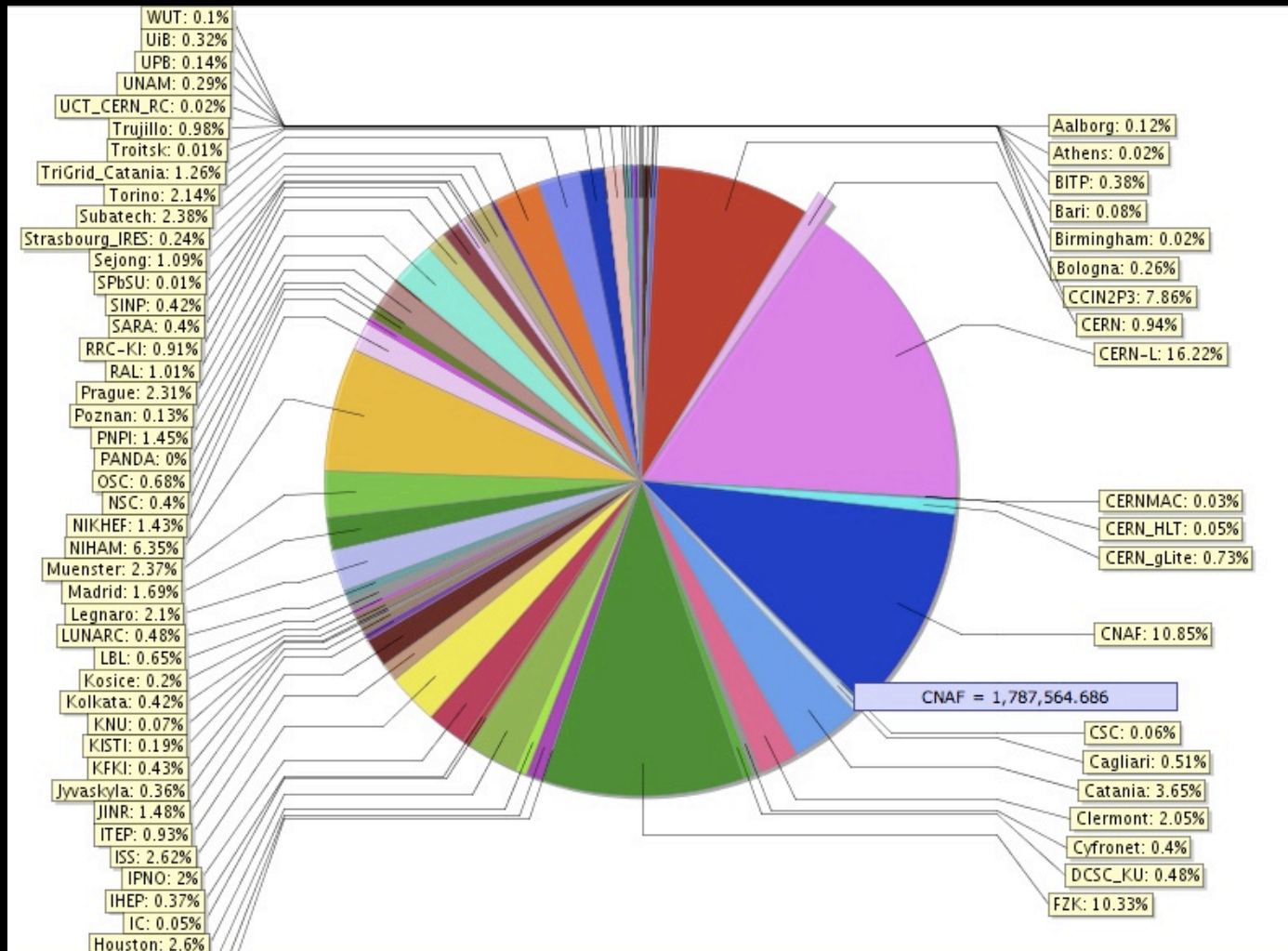
- Production statistics
 - MC events for detector and physics performance studies, various configurations requested by ALICE PWGs
 - 90 Million p+p events (2006+2007)
 - 1 Million Pb+Pb events
 - Comprehensive test of the reconstruction and analysis framework AliRoot
- Storage statistics
 - Since 01/01/2007 - 600TB of ESDs
 - Last year (with data replication during FTS exercise) - 1 PB
 - Most of the data is stored at CERN (CASTOR2), pending wider deployment of xrootd-enabled storage (see later)



Job profiles



Site relative contribution



50% T1s, 50% T2s



ALICE operation structure

- AliEn central services
 - Catalogue, API service (user Grid interaction), Task Queue, Authorization, CPU and storage quotas, Software builds&checks and repository
 - Close to production quality (>95% availability over the last 4 months)
 - Services have undergone substantial improvement
 - Ongoing effort to build redundant hardware setup and failover capabilities (many thanks to CERN IT!)



ALICE operation structure (2)

- Site AliEn services
 - Running on the VO-boxes
 - Interface to the site Grid services - stable for gLite and ARC middleware, OSC interface being build
 - Substantial improvement in the services over the past year has made them very stable, minimizing the number of interventions



LCG/gLite services

- LCG RB is now 'legacy software'
 - Still using it in production
- gLite WMS
 - Intensive test under way - initially at CNAF, now at CERN
 - ALICE dedicated WMS installed
 - Using same production queues, thus assuring adequate scale test
- FTS for T0->T1 transfers
 - V.1 fully tested and integrated with AliEn FTD
 - Three common challenges SC(LCG)/ALICE
 - Target rates achieved, but not the time stability
 - V.2 ALICE is entering testing phase in the framework of the FDR
 - This is also an indirect test of SRV v.2 (wherever available)



xrootd storage solutions

- dCache (xrootd emulation) – most advanced
 - In production GSI, CCIN2P3, NDGF, SARA
 - Being deployed at GridKA
- CASTOR2 (integrated xrootd) - first deployment
 - Common effort of CASTOR2 and ARDA teams at CERN, xrootd at SLAC
 - First production version installed by IT/FIO on the ALICE CASTOR2 instance at CERN since 10 days
 - Currently in 'tuning' phase of xrootd redirectors
 - Building operational experience
 - Other CASTOR2 sites for ALICE: CNAF, RAL
- DPM (integrated xrootd) – first choice for T2 sites
 - Preliminary version being tested at Torino and CERN



xrootd storage solutions - usage

- Unlike job management, storage not yet at “production quality”
 - Installation, operation and stability is still an issue
- Limited capacity (few T1 sites are offering large storage capacities for ALICE) is limiting data analysis efficiency
 - Currently ~20Mio (of 90Mio) events can be made available for user analysis on spinning media at any given time
 - To avoid ‘trashing’ of disk servers, users can work only with restricted datasets
 - As a consequence, a chaotic analysis of the entire ESD volume is not possible
- The storage capacity will remain a problem for ALICE
 - Simply missing from pledges



Grid operational principles

- Central services
 - Supported by Core Offline
 - Emphasis on high-quality monitoring tools, services failover capabilities and redundant hardware
 - Minimal human intervention
- Site services
 - Supported by Regional Experts - typically one per country/region
 - One expert supports the operation at ~8 centres
 - Overlapping responsibilities
 - Some large T2s have local experts, also acting as a first-line user support for the Grid analysis
 - The regional experts work with the site fabric/Grid experts to resolve any outstanding operational issues



Grid operational principles - communications and tests

- Weekly ALICE-LCG Task Force meetings
 - Escalation of unresolved issues to the OPS meeting
- Heavy use of GGUS ticketing system
- Specific problems are also reported to site expert lists (support@)
- Recently introduced - ALICE test suite in SAM for gLite site services
 - 5 proxy tests, accessibility of the common software area, test of the UI and ALICE queue(s) in the CE
 - In development - documentation and notification mechanisms (according to what the site prefers - GGUS or local contacts)
 - Expected that these tests will improve substantially the site services availability



User analysis - interactive

- Powered by PROOF
- Clusters operational at CERN (CAF) and GSI (GSIAF)
 - Other sites are planning to provide similar facilities
- Scope - prompt data analysis
 - Portions of ESDs - quick feedback on reconstruction quality
 - Detector calibration by experts
 - Time critical tasks
- Currently used by ~10-20 users in parallel
- Data management is integrated with the Grid
- Development efforts
 - Code stability
 - Disk and CPU quotas



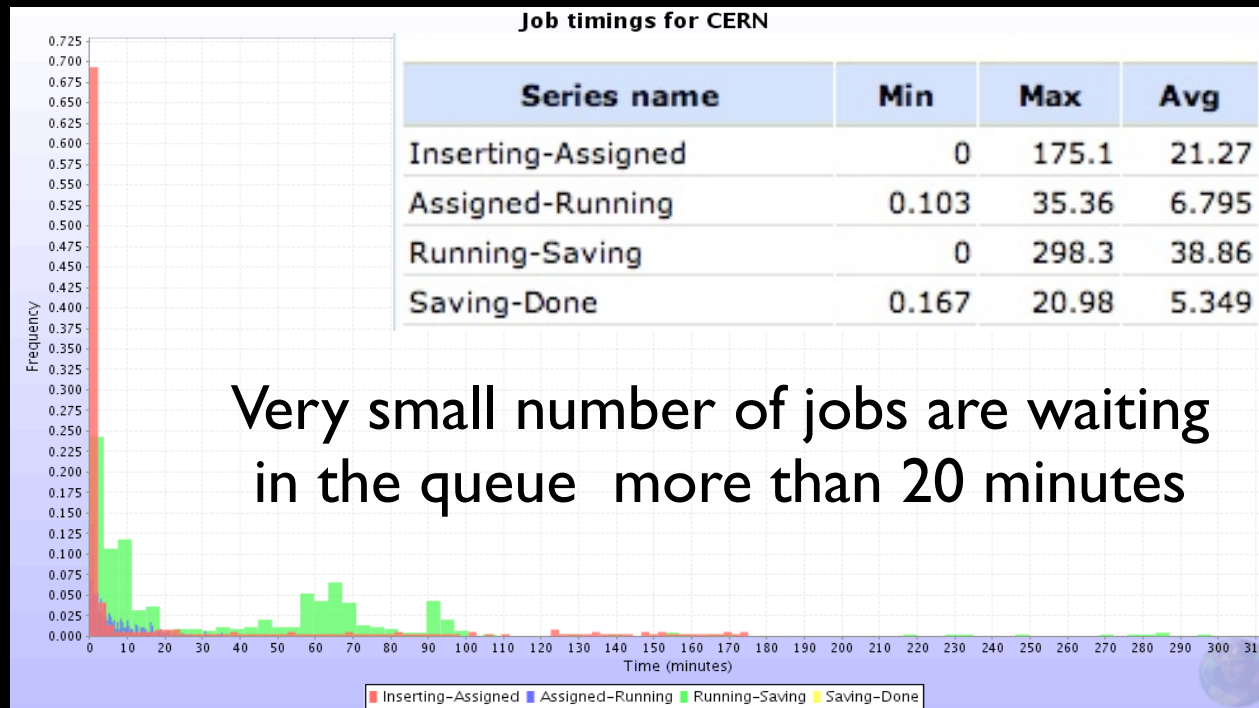
User analysis - Grid (==batch)

- Mainly at the T2s and if resources present - at the T1s
- After initial teething period, stable since 4 months and concentrated at two places
 - CERN (most of the ESDs are located here)
 - GSI - testing and tuning of analysis at a T2
 - Limitation - availability of data on spinning media
- Scope
 - Large volume data analysis, production of AODs
 - Production of specific portions of RAW data (calibration and special runs)
- Development efforts
 - Analysis Train: streamlining of user tasks and minimization of chaotic analysis
 - New functionality: merging at end of job is released and being tested



User analysis on the Grid

- Very important to get more users on the Grid
 - There is an upward trend, but it is very slow and sporadic
 - The analysis effort is taken very seriously by the core Offline team and the responses to all user queries and requests is fast

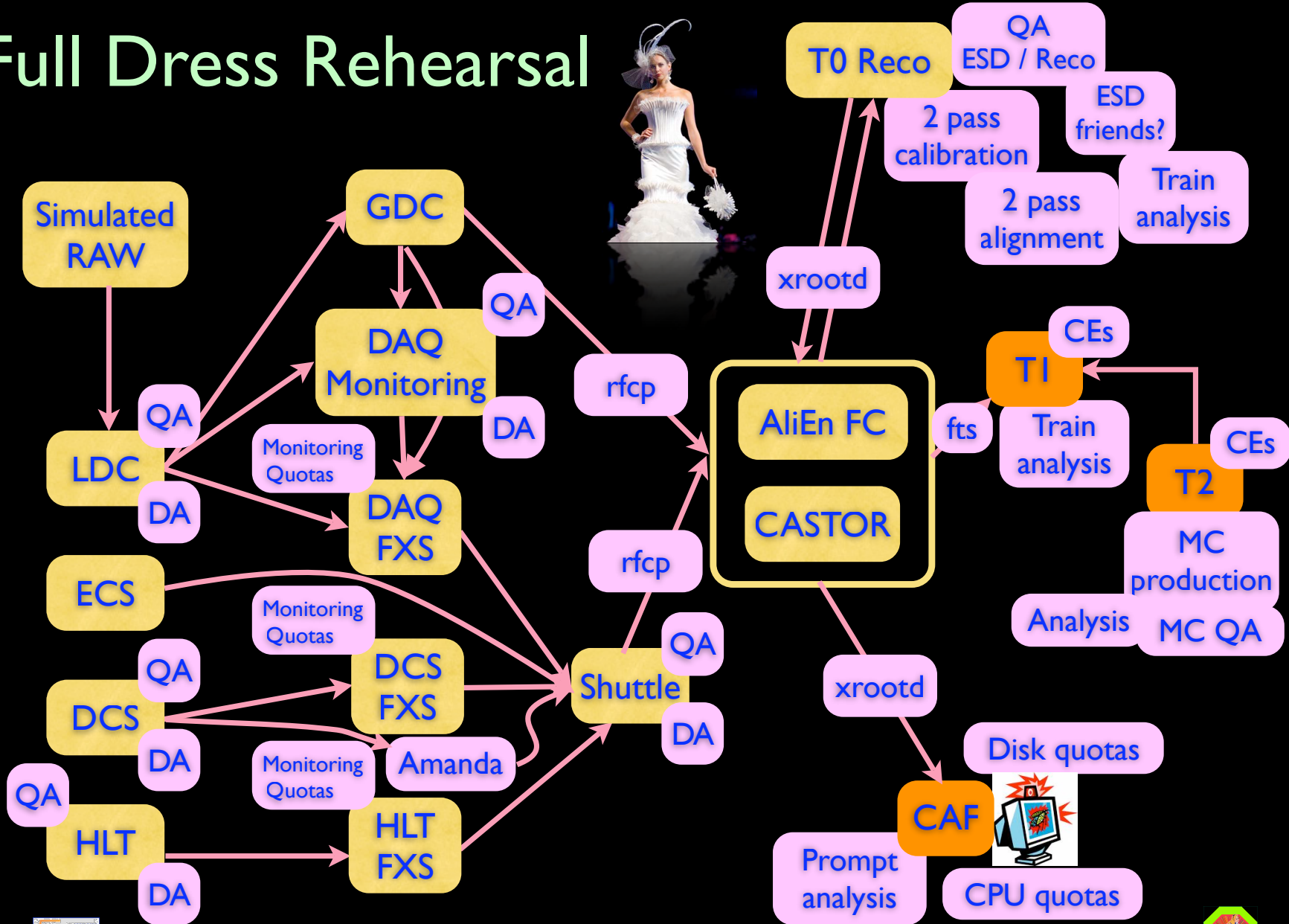


Data challenge plans

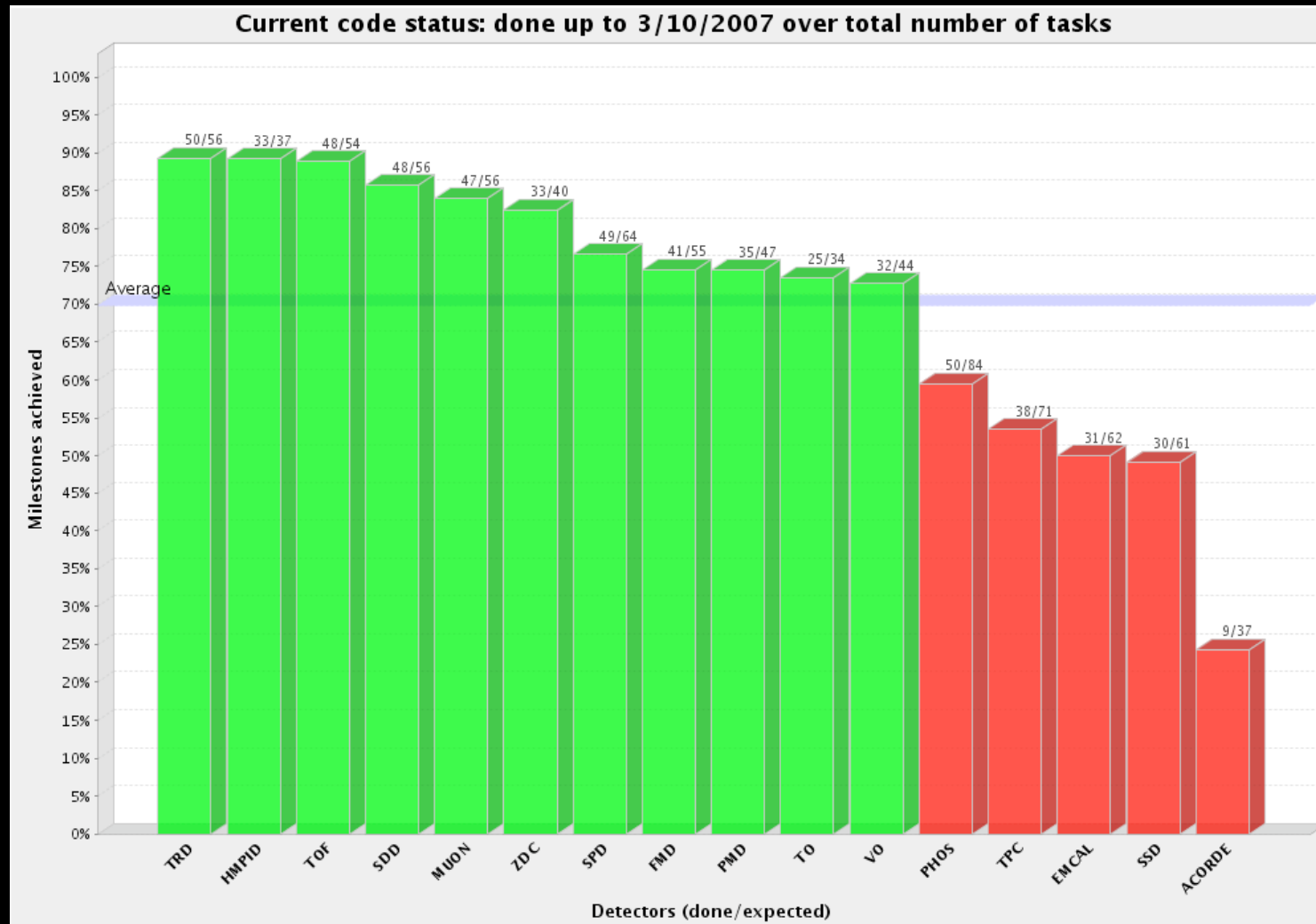
- Continue with the production and analysis of MC events
- Integration of the gLite WMS
- Main effort - deployment of storage
- In addition - execute the Full Dress Rehearsal (see ALICE presentation at yesterday pre-GDB)



Full Dress Rehearsal



Task summary



Present situation at external sites

Pledged by external sites versus required (new LHC schedule) all													
		2007		2008		2009		2010		2011		2012	
		T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
<u>CPU</u>	Requirement (MSI2K)	3.2	4.6	10.1	12.5	19.9	14.3	23.5	25.0	30.5	32.5	39.7	42.2
	Missing %	-45%	-5%	-31%	-37%	-47%	-20%	-39%	-40%	-53%	-54%	-64%	-65%
<u>Disk</u>	Requirement (PB)	1.3	0.7	4.0	1.7	6.8	4.0	12.0	4.3	16.6	5.6	22.4	7.3
	Missing %	-43%	4%	-32%	5%	-37%	-22%	-50%	-3%	-64%	-25%	-73%	-43%
<u>MS</u>	Requirement (PB)	1.4	-	5.8	-	12.4	-	19.7	-	27.0	-	33.7	-
	Missing %	-23%	-	-32%	-	-36%	-	-40%	-	-57%	-	-65%	-



Full Dress Rehearsal - phases

- Phase I - registration of RAW in CASTOR2 and Grid catalogue, replication to TIs, automatic reconstruction to ESDs (September - November)
 - Replication and automatic reconstruction postponed
- Phase II - all of the above + conditions data with Shuttle (December - January)
 - Interaction with DCS archive DB, DAQ and HLT is tested
 - Incorporation in the test beams / commissioning in P2 depends on detector plans
- Phase III - all of the above + Online Detector Algorithms + Quality Assurance (February - data taking)
 - Detailed plan existing, FDR contacts identified
 - We need synthetic planning and more frequent dedicated meetings
 - Progress OK, but time contingency small



