

LHCb status and plans

Ph. Charpentier
CERN

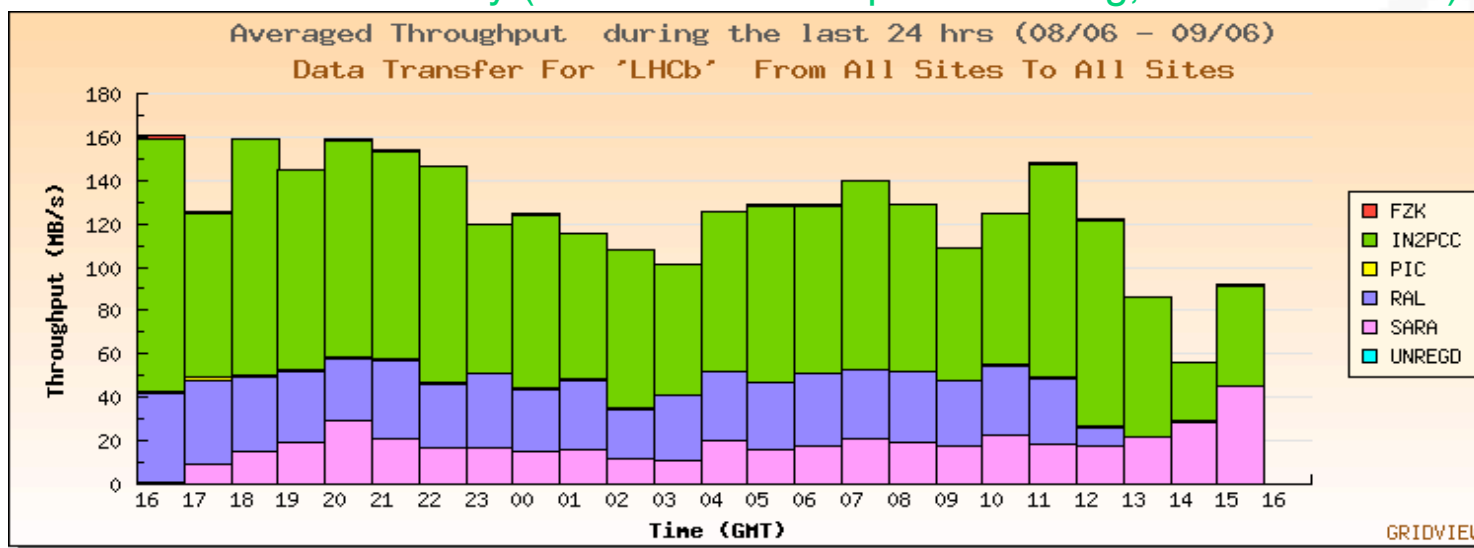


**International Conference on Computing
in High Energy and Nuclear Physics**
2-7 Sept 2007 Victoria BC Canada

- **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
 - ☆ Job submission and data upload (simulation: no input data)
 - ☆ Data export from CERN (FTS) using MC raw data (DC06-SC4)
 - ☆ Job submission with input data (reconstruction and re-reconstruction)
 - ✧ For staged and non-staged files
 - ☆ Data distribution (DSTs to Tier1s T0D1 storage)
 - ☆ 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)



- Summer 2006
 - Data production on all sites
 - ☆ Background events (~100 Mevts b-inclusive and 300 Mevts minimum bias), all MC raw files uploaded to CERN
- Autumn 2006
 - MC raw files transfers to Tier1s, registration in the DIRAC processing database
 - ☆ As part of SC4, using FTS
 - * Ran smoothly (when SEs were up and running, never 7 at once)



- Summer 2006
 - Data production on all sites
 - ☆ Background events (~100 Mevts b-inclusive and 300 Mevts minimum bias), all MC raw files uploaded to CERN
- Autumn 2006
 - MC raw files transfers to Tier1s, registration in the DIRAC processing database
 - ☆ As part of SC4, using FTS
 - ✱ Ran smoothly (when SEs were up and running, never 7 at once)
 - ☆ Fake reconstruction for some files (software not finally tuned)
- December 2006 onwards
 - Simulation, digitisation and reconstruction
 - ☆ Signal events (200 Mevts)
 - ☆ DSTs uploaded to Tier1 SEs
 - ✱ Originally to all 7 Tiers, then to CERN+2



- February 2007 onwards
 - Background events reconstruction at Tier1s
 - ☆ Uses 20 MC raw files as input
 - ✧ were no longer on cache, hence had to be recalled from tape
 - ☆ output rDST uploaded locally to Tier1
- June 2007 onwards
 - Background events stripping at Tier1s
 - ☆ Uses 2 rDST as input
 - ☆ 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

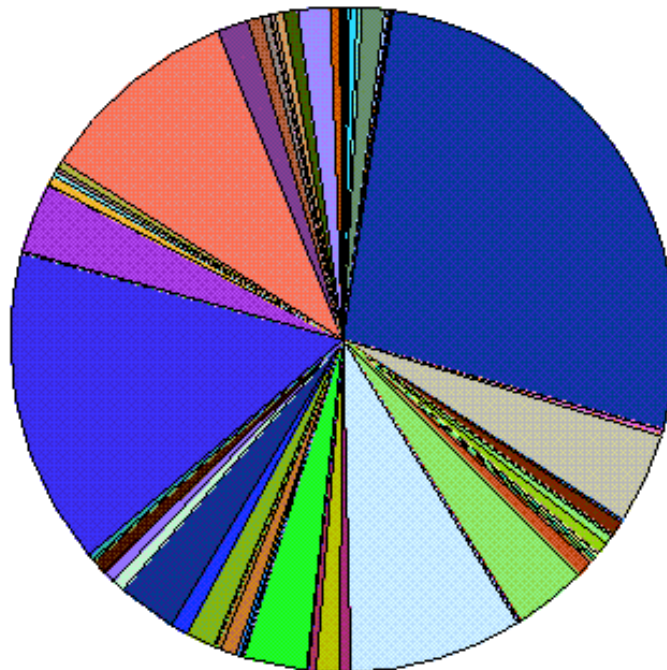


- ❑ Performed by LHCb SAM jobs
 - ☆ See Joël Closier's poster at CHEP
 - ☆ Install software, checks site capabilities (platform), runs a test job using all applications
- ❑ Problems encountered
 - ☆ Reliability of shared area: scalability of NFS?
 - ☆ Access permissions (lhcbasm)
 - ☆ Move to pool accounts...
- ☆ Important: beware of access permissions when changing accounts' mapping at sites!!!
 - ❄ moving to pool accounts was a nightmare



- Up to 10,000 jobs running simultaneously
 - ▣ Continuous requests from physics teams

Total Running Jobs: 9654
DIRAC: 0.32% LCG: 99.68%



Feb 28 2007, 06:40

DIRAC.Zurich-MH.ch	0.18%	LCG.IPSL-IPGP.fr	0.06%
DIRAC.Zurich.ch	0.15%	LCG.IRB.hr	0.11%
LCG.ACAD.bg	0.36%	LCG.ITEP.ru	0.21%
LCG.AUVER.fr	0.24%	LCG.KFKI.hu	0.77%
LCG.Barcelona.es	1.08%	LCG.KIAC.ru	0.28%
LCG.Bari.it	0.20%	LCG.KIAM.ru	0.09%
LCG.BHAM-HEP.uk	0.03%	LCG.Krakow.pl	1.42%
LCG.Bologna.it	0.01%	LCG.LAL.fr	0.85%
LCG.Cagliari.it	0.06%	LCG.Lancashire.uk	2.88%
LCG.Catania.it	0.11%	LCG.LISA.nl	0.63%
LCG.CERN.ch	26.94%	LCG.Liverpool.uk	0.58%
LCG.CESGA.es	0.04%	LCG.LPC.fr	0.71%
LCG.CGG.fr	0.26%	LCG.LPH-fails.fr	0.10%
LCG.CNAF-GRIDIT.it	0.02%	LCG.LPH.fr	0.29%
LCG.CNAF.it	4.42%	LCG.Manchester.uk	15.38%
LCG.CNB.es	0.16%	LCG.Napoli-Atlas.it	0.03%
LCG.CPPM.fr	0.70%	LCG.Napoli.it	0.05%
LCG.CSCS.ch	0.12%	LCG.NIKHEF.nl	3.39%
LCG.Dortmund.de	0.40%	LCG.OU.il	0.04%
LCG.Durham.uk	0.60%	LCG.Oxford.uk	0.50%
LCG.Edinburgh.uk	0.05%	LCG.Padova.it	0.24%
LCG.EELA-CIEMAT.es	0.19%	LCG.Pisa.it	0.20%
LCG.ETF-RIT.lv	0.15%	LCG.PUFI.ru	0.37%
LCG.Ferrara.it	0.15%	LCG.QMUL.uk	9.85%
LCG.FESB.hr	0.05%	LCG.RAL-HEP.uk	1.51%
LCG.FORTH.gr	0.83%	LCG.RHUL.uk	0.70%
LCG.Glasgow.uk	3.54%	LCG.SINP.ru	0.36%
LCG.GR-01.gr	0.10%	LCG.Sofia.bg	0.06%
LCG.GR-05.gr	0.06%	LCG.TCD.ie	0.21%
LCG.GRIDKA.de	8.41%	LCG.Torino.it	0.42%
LCG.HG-02.gr	0.55%	LCG.ULANBIM.tr	0.68%
LCG.HG-06.gr	1.12%	LCG.USC.es	1.55%
LCG.HPC2H.se	0.38%	LCG.WARSAW.pl	0.49%
LCG.Imperial.uk	0.08%	LCG.WCSS.pl	0.11%
LCG.IH2P3.fr	3.08%	LCG.WEIZMANN.il	0.02%



- Up to 10,000 jobs running simultaneously
 - Continuous requests from physics teams
- Problems encountered
 - SE unavailability for output data upload
 - ☆ Implemented a fail-over mechanism in the DIRAC DMS
 - ☆ Final data transfer filed in one of the VOBOXes
 - ✧ Had to develop multithreaded transfer agent
 - too large backlog of transfers
 - ☆ Had to develop an lcg-cp able to transfer to SURL
 - ✧ Request to support SURL in lcg-cp
 - ✧ Took 10 months to be in production (2 weeks to implement)
 - Handling of full disk SEs
 - ☆ Handled by VOBOXes
 - ☆ Cleaning SEs: painful as no SRM tool (mail to SE admin)



- Needs files to be on disk cache
 - Easy for first prompt processing, painful for reprocessing
 - ☆ For first pass of real data: “pinning” vs processing latency
 - Developed a DIRAC stager agent
 - ☆ Jobs are put in the central queue only when files are staged
- File access problems
 - Inconsistencies between SRM tURLs and root access
 - ☆ LHCb paradigm is to access files from the servers
 - ✧ Favor long connections with small throughput
 - ✧ Caveat: sites have to provide sufficient concurrent access connections
 - unreliability of rfio, problems with rootd protocol authentication on the Grid (now fixed by ROOT)
 - lcg-gt returning a tURL on dCache but not staging files
 - ☆ Workaround with dccp, then fixed by dCache



- gLite WMS
 - Several attempts at using it, not very successful
 - ☆ Still not used in production (not released as such...)
- 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)
- SRM v2.2
 - Tests ongoing
- Agreement and support for generic pilot agents
 - Essential for good optimisation at Tier1s
 - ☆ Prioritisation of activities (simulation, reconstruction, analysis)

- Re-processing of background
 - Just restarted (software fault found): 6,000 jobs
 - ☆ 20 files as input per job
 - Stripping will follow: 3,000 jobs
 - ☆ 42 files as input per job
- SRM v2.2 tests
 - Ongoing, many issues found and fixed
 - ☆ Very collaborative work with GD
 - ☆ Difficult to get space tokens and corresponding pools properly configured
- Analysis
 - Rapidly growing (batch data analysis, ROOT scripts for fits, toy MC)

- 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
 - ☆ Oracle backend, 3D streaming
 - 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
 - ☆ Using existing simulation files (200 MBytes)
 - ☆ Useless without SRM v2.2
 - LHCb welcomes a concurrent DR in Spring 08
 - ☆ Will use 2 GB “raw” (simulated) files



- Main problem encountered is with Disk1TapeX storage
 - 3 out of 7 Tier1s didn't provide what had been requested
 - ☆ Continuously change distribution plans for LHCb
 - ☆ Need to clean up datasets to get space (painful with SRM v1)
 - Not efficient to add servers one by one
 - ☆ When all servers are full, puts a very large load on the new server
 - Not easy to monitor the storage usage
- Too many instabilities in SEs
 - Full time job checking availability
 - ☆ Enabling/disabling SEs in the DMS
 - ☆ VOBOX helps but needs guidance to avoid DoS
- Several plans for SE migration
 - RAL, PIC, CNAF, SARA (to NIKHEF): to be clarified
 - ☆ Easier sooner than later....



- LHCb happy with the proposed agreement from JSPG (EDMS 855383)
 - Eager to see it endorsed by all Tier1s
 - ☆ Essential as LHCb run concurrent activities at Tier1's
 - ☆ Seems the most (only?) promising way to implement priorities
 - DIRAC prepared for running its payload through a glexec-compatible mechanism
 - ☆ Wait for sites to deploy the one they prefer
 - ☆ Already tested the IN2P3 minimal implementation (logging user's identity)
 - ✧ User's credentials downloaded for DM operations
 - ☆ LHCb doesn't require user level accounting by sites
 - ✧ DIRAC has its own accounting
 - ☆ LHCb is ready to make the police if needed
 - ✧ DIRAC has an authorisation system: users can be banned
 - ✧ Full traceability is ensured



- Problem of knowing “what runs where”
 - Reporting problems that was fixed long ago
 - ☆ but either were not released or not deployed
- Environment found on the WN cannot be trusted
 - DIRAC will import its own, fully controled
 - ☆ unsetenv PATH / LD_LIBRARY_PATH / PYTHONPATH !!
 - need to use the same version everywhere
 - ☆ even on different platforms
- Client MW from LCG-AA
 - very promising solution (using tarballs for deployment)
 - ☆ Installed in the shared area
 - very collaborative attitude from GD
 - ☆ versions for all available platforms installed as soon as ready
 - ☆ allows testing on LXPLUS and on production WNs
 - ✧ tarball shipped with DIRAC and environment set using CMT
 - ✧ not yet in full production mode, but very promising
 - ☆ allows full control of versions
 - ✧ possible to report precisely to developers
 - ✧ no way to know which version runs by default on a WN



- **Straightforward for LHCb applications**
 - problem was middleware clients used by them
 - ☆ dCache, gfal, lfc...
- **Usage by DIRAC**
 - binaries are OK
 - ☆ except lcg-cp that had a regression (2 weeks to find out)
 - python binding is not OK at some sites because...
- **Inconsistencies between MW and OS**
 - middleware is 32-bit only
 - hence WNs should by default expose a 32-bit architecture when being accessed from grid queues
 - ☆ at CERN, python is 64-bit
 - ☆ in addition unnecessary environment variables are making the case even more complicated
- **Message**
 - Developers: foresee concurrent multiple platform proper support
 - ☆ including compilers (gcc 4.1 is there...)



- Very impractical to test client MW on PPS
 - completely different setup for DIRAC
 - hard to verify all use cases (e.g. file access)
- Was used for testing some services
 - ☆ e.g. gLite WMS
- PPS uses a lot of resources in GD
 - worth discussing with experiments if needed...
 - ☆ no definite answer to the question from LHCb...
 - but easier to get a PS instance of the service
 - ☆ known to the production BDII
 - ☆ CEs available using a special tag: VOs can target those CEs
 - ☆ possibility to use or not depending on reliability
 - ✧ example: slc4 CEs were needed in order to find out all pbs
 - ☆ sees all production resources
 - ✧ caveat: should not break e.g. production CEs
 - but expected to be beyond that level of testing...
 - ☆ Already the case for SRM v2.2 tests.... and it seems not to pose problems

- Essential to test sites permanently
 - See J.Closier's poster at CHEP
 - Use the SAM framework
 - ☆ check availability of CEs open to LHCb
 - ☆ install LHCb and LCG-AA software
 - ✱ platform dependent
 - ☆ reports to the SAM database
 - ☆ LHCb would like to report the availability as they see it
 - ✱ no point claiming a site is available just for the ops VO
 - Faulty sites are “banned” from the DIRAC submission
 - ☆ LHCb can provide its view of site availability
 - ☆ Suggest all VOs do the same and report to the MB
 - Faulty SEs or full disk-SEs can also be “banned” from the DMS (as source and/or destination)
 - ☆ takes information from actual transfers (and announcements)



- 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”

- LHCb session this afternoon
 - Room : Sydney
 - Topics for informal discussion with sites and developers
 - ☆ SRM v2.2 interpretation / site deployment
 - ☆ Sites plans for storage
 - ✧ disk capacity
 - ✧ technology migration
 - ☆ Generic agents (a.k.a. glexec on WN)
 - ☆ LFC service using the read-only instances replicated by 3D



EDMS 853968

6. Provisioning of services to and use of the Grid is at your own risk. Any software provided by the Grid is provided on an as-is basis only, and subject to its own license conditions. There is no guarantee that any service operated by the Grid is correct or sufficient for any particular purpose. The Grid, the Sites and other VOs are not liable for any loss or damage in connection with your participation in the Grid

Would you buy a car to a vendor proposing you such conditions?.....