The Worldwide LHC Computing Grid

The WLCG Service Challenges

Closeout Review, September 2006

Jamie Shiers, CERN, Geneva, Switzerland

Worldwide LHC Computing Grid

Distributed Production Environment for Physics data Processing

Agenda

4 main components:

1. **Summary of Observations & Recommendations**
2. High-level experiment-by-experiment review
3. Service Issues: problem response / resolution
4. Workshops & Other Meetings

The State of the Grid

- The Services provided as part of SC4 are mainly an evolution of those provided in SC3
  - Essentially those established in the BSWG, although SC3 timelines required some 'second guessing'
    - SC3 service setup target: May 2005; BSWG report out: July 2005
- Despite the problems encountered – and those yet to be faced and resolved – I believe that it is correct to say we have a usable service (not a perfect one)
  - There **have** been serious problems, but **once** understood, these have (usually) been resolved rather rapidly
    - Some could have been avoided – with 20/20 hindsight…
  - **It is surely much more productive to focus on the former**

Acknowledgements

- The material in this talk came from many sources…
  - The review documents mentioned, but also presentations, e-mails, Wiki pages, blogs, …
- It is (IMHO) important to recognise the **very hard work** performed by people **too numerous** to name, but at all sites & in all experiments
  - (and of course to GridPP for this template…)

Service Preparation

- Along with other CERN IT groups and other sites, there are many many aspects of Service Preparation / Hardening in which the SC Team has been (pro-)actively involved.
- **I will skip this, simply due to lack of time**

➢ Except to say that the infrastructure is in much better shape than one year ago, but we still have a lot of additional work ahead.

WLCG - Major Challenges Ahead (Status at time of last comp. review)

1. Get data rates at all Tier1s up to MoU Values
   - Stable, reliable, rock-solid services
   - **We are currently about 1/2 the target level, without including tape**

2. (Re-)deploy Required Services at Sites to meet MoU Targets
   - Measured, delivered Availability, maximum intervention time etc.
   - Ensure that the services delivered match the experiments' requirements

➢ **T0 and T1 services are tightly coupled!**
   - Particularly during accelerator operation

- Need to build strong collaborative spirit to be able to deliver required level of services
   - And survive the inevitable 'crises'…

Observation #1

- We are still **not** able to demonstrate full nominal Tier0-Tier1 transfer rates (1.6GB/s) over extended periods, let alone **recovery** rates (targeted at twice nominal);
- **As the services involved are still far from stable, this recovery capability is needed**
  - Both **scheduled** & unscheduled interruptions
- Also a factor in extended coverage & MoU targets (see under Service Issues later)

Presented to MB, SC Tech Day & FZK-cloud w/s.

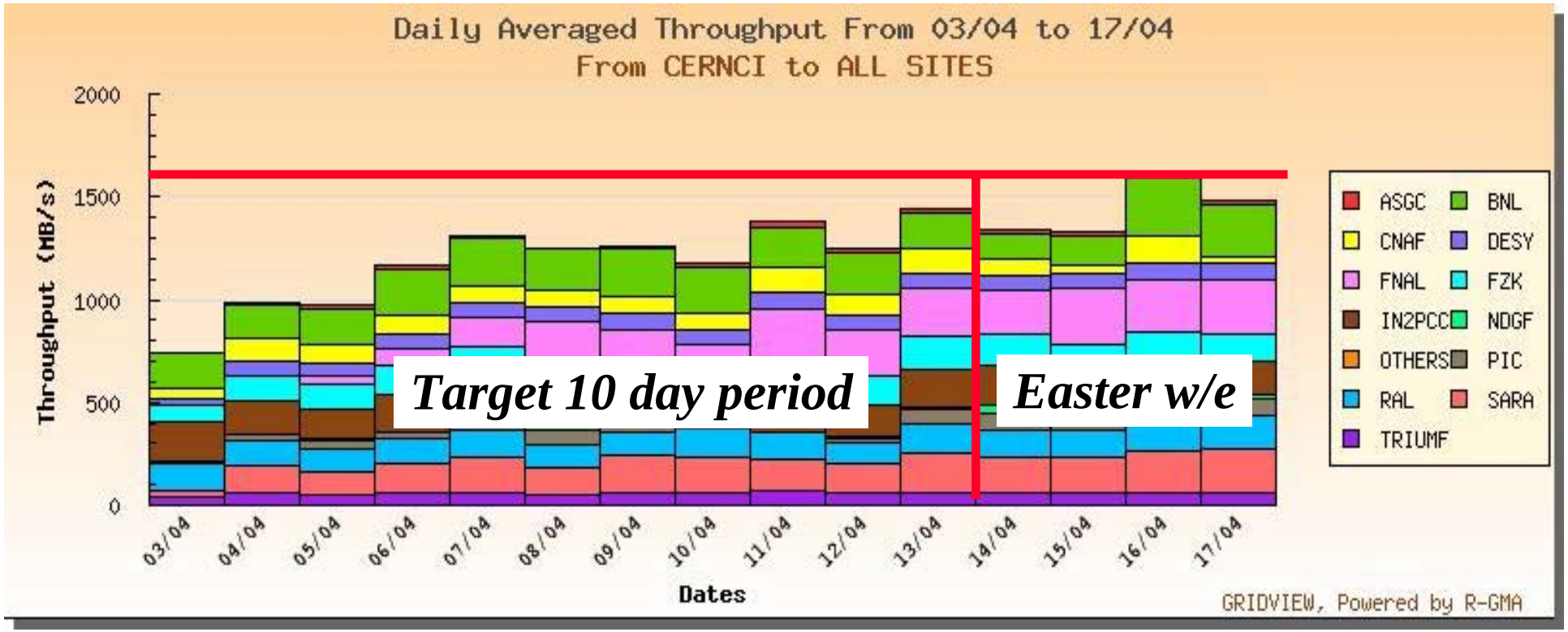
Prepared & reviewed through LCG ECM (RSM) & experiment meetings

Heat

SC4 Tier0 – Tier1 Data Rates (pp)

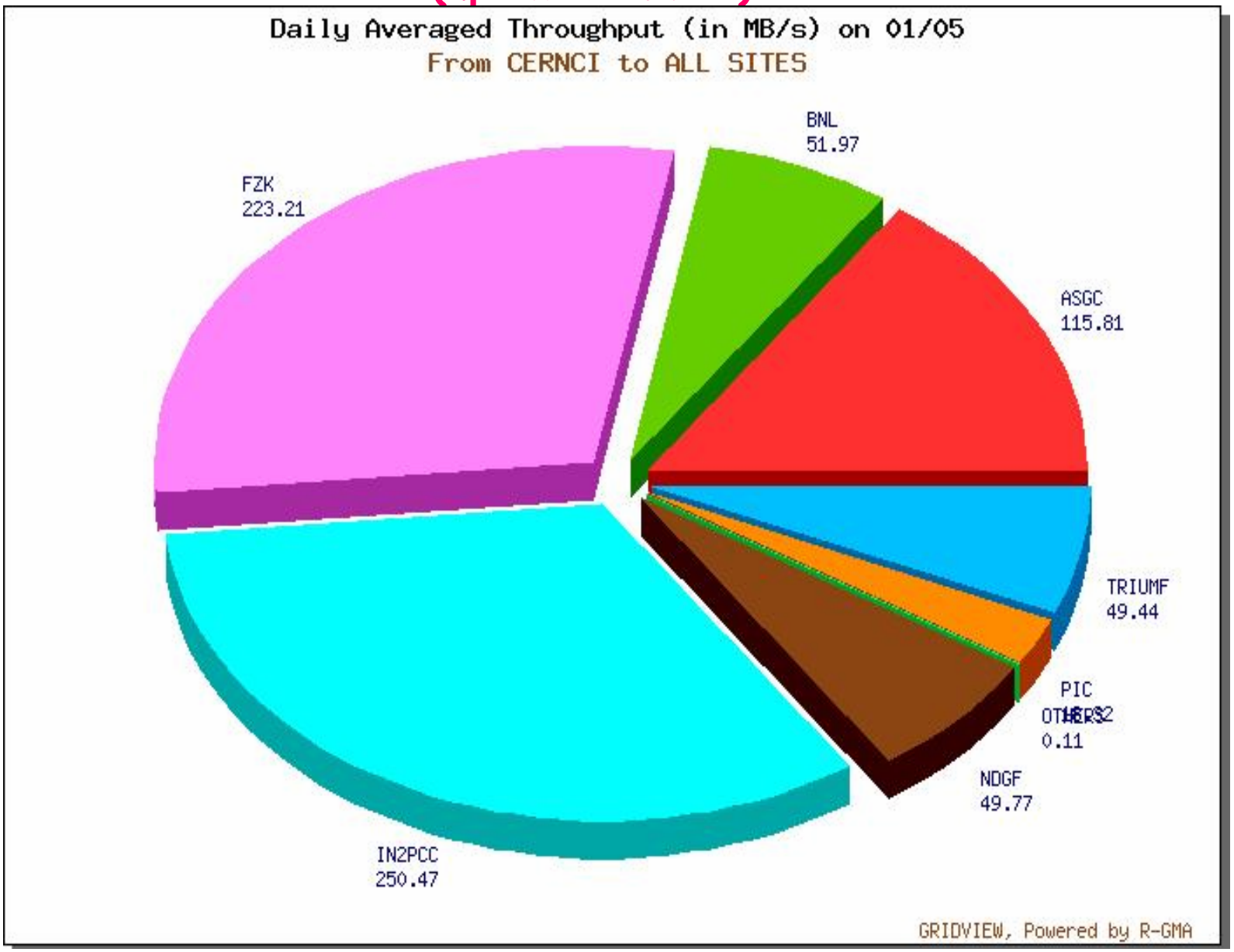
Tier1 Centre	*ALICE*	*ATLAS*	*CMS*	*LHCb*	*Target*
IN2P3, Lyon	9%	13%	10%	27%	**200**
GridKA, Germany	20%	10%	8%	10%	**200**
CNAF, Italy	7%	7%	13%	11%	**200**
BNL, USA	-	22%	-	-	**200**
FNAL, USA	-	-	28%	-	**200**
RAL, UK	-	7%	3%	15%	**150**
NIKHEF, NL	(3%)	13%	-	23%	**150**
ASGC, Taipei	-	8%	10%	-	**100**
PIC, Spain	-	4% (5)	6% (5)	6.5%	**100**
Nordic Data Grid Facility	-	6%	-	-	**50**
TRIUMF, Canada	-	4%	-	-	**50**
TOTAL					**1.6GB/s**

SC4 Results (Disk – Disk)



Site by Site Debugging (April – May)

(updates in red)

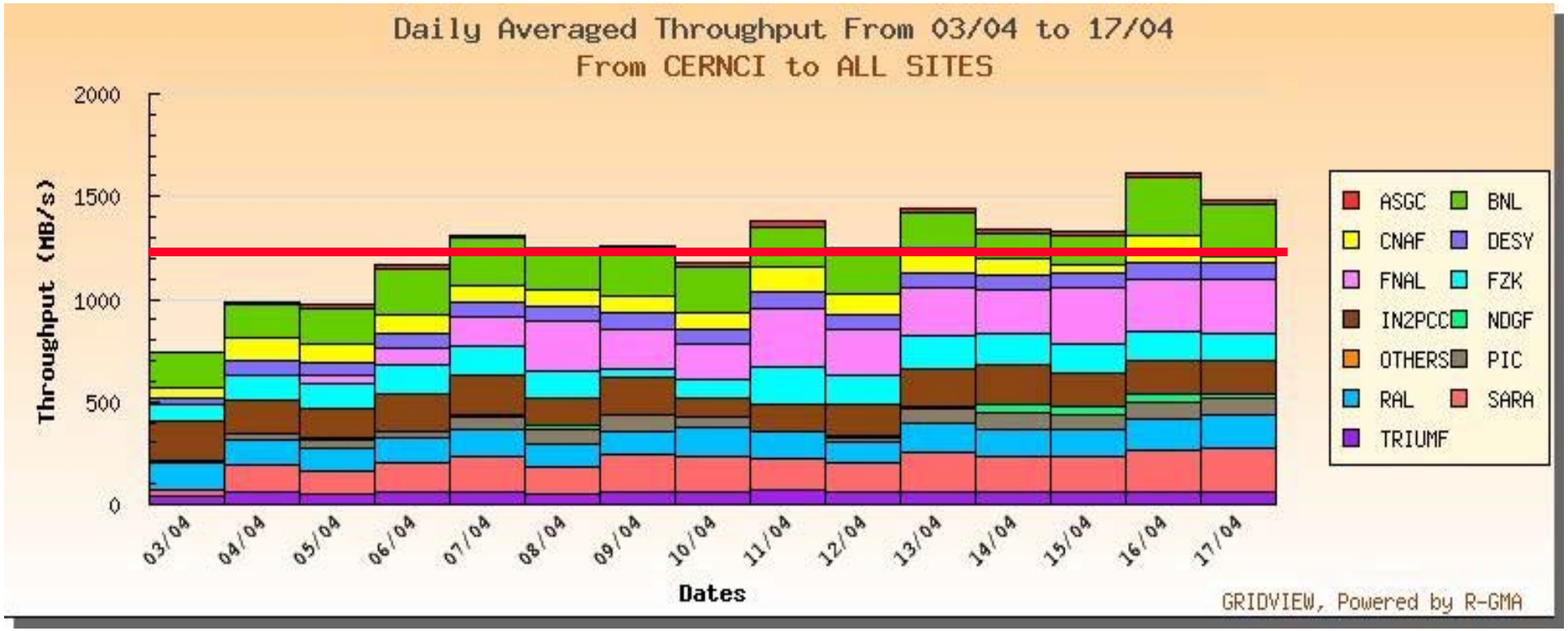


Heat

New Tier0 – Tier1 Rates (Megatable)

Tier1 Centre	ALICE	ATLAS	CMS	LHCb	Total
IN2P3, Lyon	27.9	75.4	43	22.6	**168.9**
GridKA, Germany	60.0	63.8	37	18.5	**179.3**
CNAF, Italy	34.6	107.0	55	18.1	**214.7**
BNL, USA	-	186.5	-	-	**186.5**
FNAL, USA	-	-	110	-	**110**
RAL, UK	8.8	76.8	7	18.5	**111.1**
NIKHEF, NL	13.8	72.0	-	21.2	**107.0**
ASGC, Taipei	-	35.7	37	-	**72.7**
PIC, Spain	-	20.6	19	15.7	**55.3**
Nordic Data Grid Facility	21.2	20.6	-	-	**41.8**
TRIUMF, Canada	-	19.2	-	-	**19.2**
US ALICE	46.4	-	-	-	46.4
TOTAL					**1312.9**

SC4 Disk – Disk Revisited



Observation #2

- However, experiment-driven data transfers (ATLAS and also CMS) achieved rates close to the target of **full nominal rates** for a single experiment (about half of the total rate for all experiments) under **much more realistic** conditions than for previous DTEAM transfers.
- For this reason, this is considered a (particularly) positive result;

ATLAS SC4 Targets

Centre	ATLAS SC4 (40% to tape)	Nominal (pp) MB/s (all experiments)
ASGC	60.0	100
CNAF	59.0	200
PIC	48.6	100
IN2P3	90.2	200
GridKA	74.6	200
RAL	59.0	150
BNL	196.8	200
TRIUMF	47.6	50
SARA	87.6	150
NDGF	48.6	50
FNAL	-	200

Global Inter-Site Rates - Megatable

Centre	T0->T1	T1->T2	T2->T1	T1<->T1
	Predictable – Data Taking	Bursty – User Needs	Predictable – Simulation	Scheduled Reprocessing
IN2P3, Lyon	168.9	286.2	85.5	498.0
GridKA, Germany	179.3	384.9	84.1	395.6
CNAF, Italy	214.7	321.3	58.4	583.8
FNAL, USA	110	415.0	52.6	417.0
BNL, USA	186.5	137.7	24.8	358.0
RAL, UK	111.1	108.3	36.0	479.4
NIKHEF, NL	107.0	34.1	6.1	310.4
ASGC, Taipei	72.7	126.5	19.3	241.2
PIC, Spain	55.3	167.1	23.3	294.5
Nordic Data Grid Facility	41.8	-	-	62.4
TRIUMF, Canada	19.2	-	-	59.0

ATLAS T1 – T1 Rates

- Take ATLAS as the example – highest inter-T1 rates due to multiple ESD copies
- Given spread of resources offered by T1s to ATLAS, requires "pairing of sites" to store ESD mirrors
- Reprocessing performed ~1 month after data taking with better calibrations & at end of year with better calibrations & algorithms

➢ **Continuous or continual? (i.e. is network load constant or peaks+troughs?)**

FZK (10%) + CCIN2P3 (13%)	BNL (22%)
CNAF (7%)	RAL (7%)
NIKHEF/SARA (13%)	TRIUMF (4%) + ASGC (8%)
PIC (4-6%)	NDGF (6%)

Heat

Nominal Tier0 – Tier1 Data Rates (pp)

Tier1 Centre	*Target*	*Date FTS channels setup and tested*
IN2P3, Lyon	**200**	**July 3**
GridKA, Germany	**200**	**May 17**
CNAF, Italy	**200**	**June 27**
BNL, USA	**200**	**August 4**
FNAL, USA	**200**	**-**
RAL, UK	**150**	**☺ March 17 ☺**
NIKHEF, NL	**150**	**~May 30**
ASGC, Taipei	**100**	**July 10**
PIC, Spain	**100**	**May 29**
Nordic Data Grid Facility	**50**	**-**
TRIUMF, Canada	**50**	**-**
TOTAL	**1.6GB/s**	**-**

Meeting the LCG challenge

Example: Tier-2 individual transfer tests

Initial focus was on getting SRMs understood and deployed…..

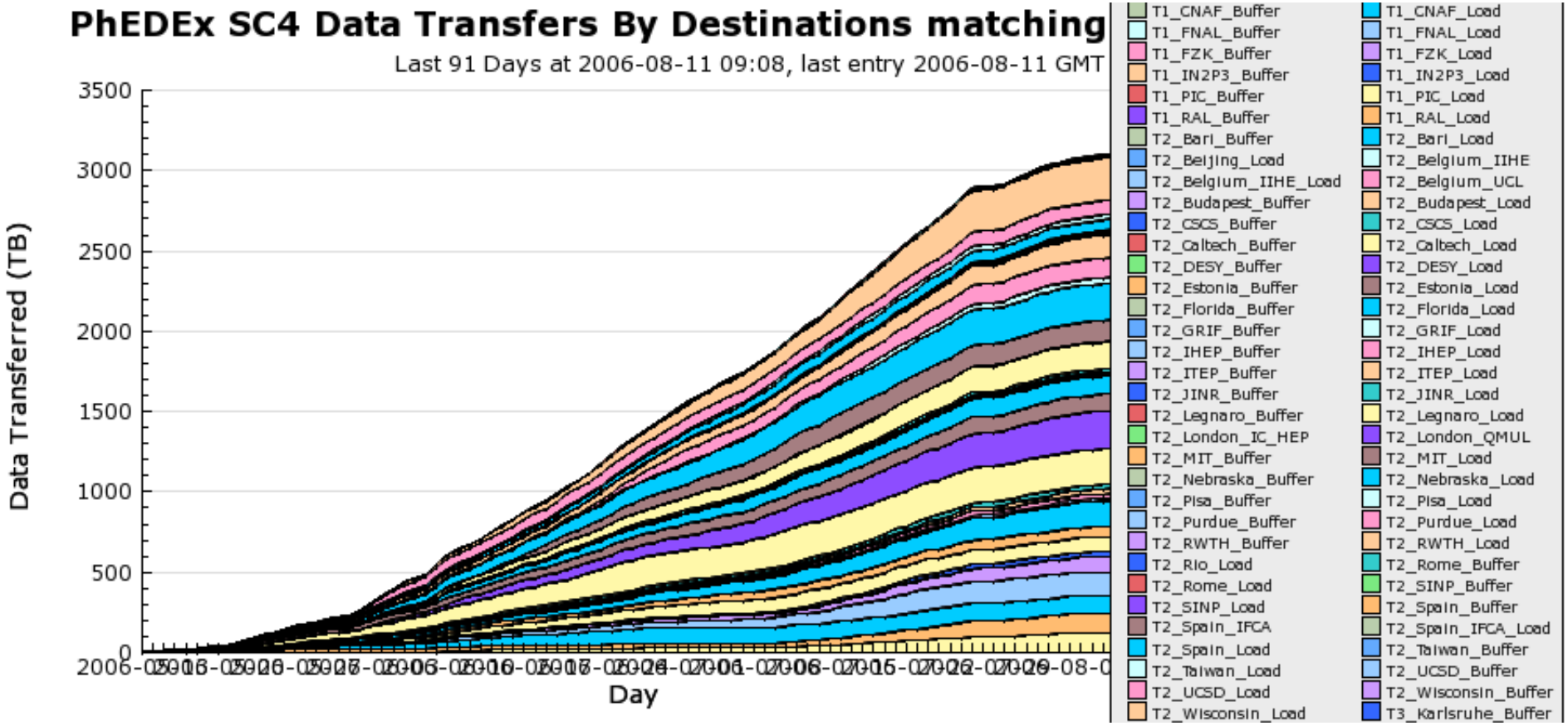
	Receiving											
	RAL Tier-1	Lancaster	Manchester	Edinburgh	Glasgow	Birmingham	Oxford	Cam	Durham	QMUL	IC-HEP	RAL-PPD
RAL Tier-1		~800Mb/s	350Mb/s	156Mb/s	166 Mb/s	289 Mb/s	252 Mb/s			118 Mb/s	84Mb/s	397 Mb/s
Lancaster												
Manchester	150 Mb/s											
Edinburgh	440Mb/s											
Glasgow	331Mb/s											
Birmingham	461 Mb/s											
IC-HEP												
Oxford	456 Mb/s											
Cambridge	74 Mb/s											
Durham	193 Mb/s											
QMUL	172 Mb/s											
IC-HEP												
RAL-PPD	388 Mb/s											

- Big variation in what sites could achieve
  - Internal networking configuration issues
  - Site connectivity (and contention)
  - SRM setup and level of optimisation
- Rates to RAL were generally better than from RAL
  - Availability and setup of gridFTP servers at Tier-2s
  - SRM setup and level of optimisation
- Scheduling tests was not straightforward
  - Availability of local site staff
  - Status of hardware deployment
  - Availability of Tier-1
  - Need to avoid first tests during certain periods (local impacts)

Example rates from throughput tests

http://wiki.gridpp.ac.uk/wiki/Service_Challenge_Transfer_Tests

Observation #3



Observation #4

- By definition, these activities tested site services, such as LFCs, VO boxes, and overall production readiness significantly more than the DTEAM-driven transfers. A number of issues have been found at a variety of sites and solutions have been found or are planned
  - see under the ATLAS section below
- However, they underline the fact that certain sites / regions still have to make **significant progress** to achieve the required service level;

➢ **Sites are addressing the issues seen with these and other services, but timescales typically weeks or (much) longer**

- **Debugging of complex problems, hardware acquisition cycles, preparation & scheduling of e.g. dCache / CASTOR upgrades etc.**
- **Communication issues again – Observation #10**

Observation #5

- A particular effective model, as demonstrated by Lyon for ATLAS, is to have a contact person for the experiment **both** at the Tier0 and the Tier1;
- US-ATLAS staffing / experience also suggests strong manpower shortage at non-US sites
- See also related observations (next)

Observation #6

- Sites appear to be able to focus their full attention on a specific experiment or challenge for a **few days only**
- This is clearly indicative of the high workload at the sites and should be built into the experiments' operational models
  - i.e. a few days at high priority per month per experiment already completely drains the sites involved;
- See also observations 5 & 9;
- This is true generally – see various attempts to start blogs / wiki pages (SC team, expts)
- See Recommendations & Actions later

Observation #9

- Several – if not many – sites appear to suffer from significant manpower shortages, which impacts both the service level that they are able to provide and the response time to requests;
  - Both “setup” and problem resolution
- **This was particularly evident both around Easter and Summer vacation periods**
- Sites affected include CNAF & FZK

Observation #7

- Upgrades to **CASTOR2** at a number of sites have led to further instabilities. Once all such migrations have been completed, a (further) test needs to be made to ensure that these sites can now meet both throughput and stability targets;
  - Note CASTOR **operations training** programme
  - Phone-in offered -- slides on Web
  - VRVS could be provided too
- Upgrades to dCache also pending which – once completed – are also expected to result in improvements
  - FZK scheduled for October – other sites?

Observation #8

- Several sites have experienced significant power and / or cooling problems, resulting in prolonged service downtime;

- [If we had to ‘design’ a year to stress test our basic infrastructure, would we have been so ‘dramatic’?]

Observation #10

- Reporting to and attendance at the weekly Joint Operations Meetings[1] has improved since the previous report in May 2006 but still leaves considerable room for further improvement
  - Reports are often written in a style that is clearly oriented at local consumption,
  - Some sites still do not provide reports on a regular basis, even though there is significant activity at that site;
  - **Major issues – such as problems & their resolution at sites including FZK & RAL (given as examples) – seem to have 'slipped through' unnoticed.**
    - **See RAL quarterly report, FZK status at FZK – T2 federation workshop**

[1] See http://agenda.cern.ch/displayLevel.php?fid=258 to access agendas, reports, action items and minutes.

Observation #11

- **Opportunistic** use of resources – **used** or expected to be used by **all** experiments – may result in the use of CPU resources at sites with insufficient local storage. As an interim solution, unrestricted WAN access to the CERN SE has been provided, but this can result in **poor** and/or **unpredictable** network performance and result in problems that are highly **complex** to debug.
- It is considered important to clearly separate this opportunistic use of resources from the standard production model, where data is typically written to the local storage element (and eventually archived to the associated Tier1 site in the case of Monte Carlo production at Tier2s.);

Observation #12

- A bug in Oracle 10.2.0.2 led to logical data corruption in the LFC and VOMRS instances at CERN. Once the problem had been sufficiently understood, it was successfully escalated to Oracle as a top priority issue. A work-around was put in place and the experiments and all outside sites were advised accordingly.
  - At the time of writing a patch that passes all test cases has still not been received, although the workaround – effectively to turn off the faulty code path – solves most of the problems and eliminates the risk of further data corruption (but introduces some side effects…)

➢ This can be viewed as an important test case both of our ability to escalate such problems within the Oracle support structure as well as to handle bugs that potentially affect a large number of sites.
  - **Important also to review strategy wrt adoption of new releases – and regular security patches / alerts (proposal to GDB/MB)**

Recommendations & Actions ✓

- Streamlining of reporting to the weekly combined operations meeting – now to held on **Thursdays at 16:00** Geneva time – and the various LCG coordination meetings (LCG **Experiment Coordination** Meeting Mondays at 15:00, LCG Service Coordination Meeting Wednesdays at 10:00) has been proposed to the WLCG Management Board and has been put in place;

- **Geneva time: UTC+1 in Winter, UTC+2 in Summer**

✓= already agreed at MB/GDB. (All have been presented).

Recommendations & Actions ✓

- The use of the EGEE **broadcast tool** for announcing both scheduled and unscheduled interruptions has greatly improved. Improvements in the tool to clarify broadcast targets are underway. Sites are requested to ensure the **nature** and **scope** of the event are **clear** both from the subject and text of the announcement (and are not, for example, deduced from the e-mail address of the sendee);
  - ☺ Tape robot maintenance at CERN 10.30-16.00 Thursday 13 July
  - ☹ Tape access interrupted
- All times should be given in **UTC**! (or local + UTC)

Recommendations & Actions

- Site monitoring of local services still needs **considerable further improvement** – many issues that could be spotted locally are still first found by the central Service Coordination Team or – worse still – by the users;
- Sites are encouraged to share their monitoring tools and experience. To this end, a focussed discussion on monitoring is foreseen at the Service Challenge Technical Day, September 15th at CERN.
- **Done – much more available than previously understood: will build on this! (See also next slide)**
- **More coordination of the different groups and approaches would be valuable…**
- ¿ **Work on this over coming months → WLCG Collaboration Workshop in January?**

Statistics concerning all the transfers performed during last week

Time frame concerned: Between **2006-09-13 00:00:00 +00:00** and **2006-09-20 00:00:00 +00:00**

Channel Name	VO Name	Total	% Failures	# Succ.	# Fail.	1st Failure Reason	% 1st Failure Reason	2nd Failure Reason	% 2nd Failure Reason	Avg. Size (GB)	Avg. Duration (sec)	Avg. Tx Rate (MB/sec)	Eff. Tx Bytes (GB)	Tx Bytes (GB)
INFN-CERN	[All]	28309	94.52	1551	26758	Source SRM	88.26	Dest SRM	9.72	0.27	129.38	2.12	832.71	924.95
CERN-SARA	[All]	33820	92.22	2632	31188	Other	97.77	Source SRM	1.31	0.34	47.32	7.37	2681.42	2692.43
FNAL-CERN	[All]	4918	91.85	401	4517	Source SRM	90.95	Dest SRM	7.9	1.96	868.14	2.31	786.73	892.01
AACHEN-CERN	[All]	4820	90.93	437	4383	Source SRM	98.86	Dest SRM	0.64	1.78	289.62	6.31	779.29	784.49
SARA-CERN	[All]	27745	89.21	2993	24752	Other	59.62	Dest SRM	38.76	0.12	154.83	0.8	362.19	446.56
RAL-CERN	[All]	25846	88.99	2846	23000	Source SRM	60.36	Dest SRM	30.76	0.16	78.45	2.09	910.21	1045.27
PISA-CERN	[All]	418	78.95	88	330	Other	100			1.97	257.64	7.84	173.68	173.68
CERN-INFN	[All]	30901	77.41	6980	23921	Dest SRM	67.01	Source SRM	16.98	0.11	76.65	1.51	3139.71	4872.26
	alice	**7105**	**74.61**	**1804**	**5301**	**Source SRM**	**45.92**	**Dest SRM**	**44.95**	**0.23**	**89.13**	**2.64**	**412.71**	**418.09**
	atlas	**1023**	**94.82**	**53**	**970**	**Other**	**62.27**	**Source SRM**	**30.93**	**0.41**	**150.7**	**2.79**	**21.75**	**34.96**
	cms	**18410**	**74.33**	**4725**	**13685**	**Dest SRM**	**72.03**	**Other**	**12.42**	**0.57**	**409.3**	**1.42**	**2683.61**	**4393.81**
	lhcb	**4363**	**90.88**	**398**	**3965**	**Dest SRM**	**94.17**	**Transfer**	**3.08**	**0.05**	**94.2**	**0.59**	**21.65**	**25.39**
BNL-CERN	[All]	3400	75.62	829	2571	Dest SRM	79.19	Other	12.99	0.13	165.63	0.81	107.55	178.75
GRIDKA-CERN	[All]	13396	74.22	3454	9942	Dest SRM	68.53	Other	26.54	0.12	96	1.3	839.17	933.02
CERN-CERN	[All]	4251	73.91	1109	3142	Dest SRM	72.76	Source SRM	22.63	0.15	110.27	1.37	163.82	210.47
IN2P3-CERN	[All]	14830	69.12	4580	10250	Dest SRM	85.94	Other	12.38	0.06	87.99	0.68	534.31	656.54
STAR-CERN	[All]	29205	68.38	9236	19969	Dest SRM	73.87	Source SRM	17.04	0.19	134.55	1.45	3448.66	3895.67
TRIUMF-CERN	[All]	1531	67.28	501	1030	Dest SRM	59.13	Source SRM	35.83	0.15	178.23	0.84	73.57	113.85
UNL-CERN	[All]	1016	62.7	379	637	Transfer	96.08	Other	3.92	1.75	4698.35	0.38	0	0
CERN-TRIUMF	[All]	3395	58.56	1407	1988	Other	54.18	Dest SRM	44.87	0.49	181.29	2.75	1366.75	1567.93
CERN-IN2P3	[All]	35688	56.23	15621	20067	Dest SRM	47.91	Other	43.15	0.13	33.52	4.11	10450.14	10450.14
PIC-CERN	[All]	619	53.31	289	330	Dest SRM	70	Other	16.36	0.91	734.09	1.26	261.74	320.06
CERN-ASCC	[All]	25741	48.79	13182	12559	Dest SRM	63.06	Other	31.61	0.66	372.51	1.8	17307.79	19303.83
CERN-RAL	[All]	17841	44.39	9922	7919	Other	64.15	Source SRM	30.82	0.05	122.03	0.4	1412.93	1904.6
CERN-BNL	[All]	24136	42.69	13833	10303	Other	61.76	Source SRM	34.09	0.6	238.85	2.56	16531.26	16946.37
CERN-PIC	[All]	22091	38.24	13643	8448	Dest SRM	58.07	Other	30.02	0.15	134.54	1.15	8202.38	9126.42
KBFI-CERN	[All]	1493	36.64	946	547	Source SRM	55.94	Other	29.8	1.76	351.11	5.12	1660.55	1712.82
LNL-CERN	[All]	1249	32.11	848	401	Source SRM	86.53	Dest SRM	12.47	1.79	313.61	5.83	1514.93	1526.18
ULBVUB-CERN	[All]	1790	30.34	1247	543	Source SRM	54.33	Other	35.91	1.74	1001.78	1.78	2169.91	2276.63
DESY-CERN	[All]	330	28.79	235	95	Dest SRM	57.89	Source SRM	40	1.87	437.33	4.37	438.95	438.95

Recommendations & Actions

- Problem resolution – and reporting – needs to be improved, particularly in the case of complex problems which require a range of expertise and / or sites to resolve (see below);
  - Be precise; be concise; be nice;

- See also next slides…
- [and also Service Issues later…]

Recommendations & Actions

- Regular reviews of open tickets and identification of complex / unresolved problems are held with escalation (depending on exact problem) as required.

➢ This has proved successful in the resolution of chronic LHCb problems as well as the CMS CSA06 preparation.

💣 **But this is still a very difficult area with vast mismatches between expectations & resources**

Recommendations & Actions

- Phone and / or physical participation of the experiments in the CERN daily operations meeting[1] (~10-15' starting at 09:15) is encouraged to highlight new problems and ensure that there is adequate information flow. These meetings are also be open to **external sites** 'wishing' to participate;
  - Starts at 09:00 with a review of internal tickets

[1] Typically held in the "openspace" in B513

Dial-in +41 22 767 6000, code 0175012.

Proposal to include ATLAS sites based on summary mail of outstanding issues distributed < 17:00 previous evening

Recommendations & Actions

- A **regular** (**3**-4/year?) WLCG Service Coordination meeting, where the Tier0 and all Tier1+Tier2 "clouds" (federations) as well as the experiments are represented, should be established. This should review the services delivered by that federation, main issues encountered and plans to resolve them, possibly following the model used by GridPP for their collaboration meetings
  - See, for example Deployment Metrics and Planning, presented at GridPP16).
- It should also cover the experiments' plans for the coming quarter in more detail than can be achieved at the weekly joint operations meetings (which nevertheless could cover any updates). This meeting should not require physical presence, but would require the reports / presentations to be submitted in advance;

Recommendations & Actions

- A "Service Coordinator (On Duty – SCOD)" – a rotating, full-time activity for the length of an LHC run (but almost certainly required also outside data taking) should be established as soon as possible. The person assuming this activity would, for their period on duty:
  - Attend the daily and weekly operations meetings, relevant experiment planning and operations meetings, CASTOR deployment meetings;
  - Liaise with site and experiment contacts (**MOD, SMOD, GMOD, DBMOD**, …);
  - Maintain a daily log of on-going events, problems and their resolution;
  - Act as a single point of contact for all immediate WLCG service issues;
  - Escalate problems as appropriate to sites, experiments and / or management;
  - Write (**and present**) a detailed 'run report' at the end of the period on duty.
- It is proposed that this rota be staffed by the Tier0 and Tier1 sites, each site manning ~2 2-week periods per year (or 4 1-week periods);
- Could also be supplemented by "Yet-Another-Grid" projects (TT)

SC4 Review

Summary of Experiment Results

Worldwide LHC Computing Grid

Distributed Production Environment for Physics data Processing

ATLAS Summary (1/2)

- The overall plan for the ATLAS SC4 exercise was to send data out to all ATLAS Tier1 sites at the full nominal rate expected for that site during LHC pp running.
- Whilst these data rates were not achieved for the target of one week, this exercise uncovered a number of problems – many of which have since been resolved – and was clearly an important step towards reaching full nominal rates under realistic conditions.
- Key accomplishments were:
  - Ran a full-scale exercise, from EF, reconstruction farm, T1 export, T2 export with realistic data sizes, complete flow
  - Included all T1s sites in the exercise from first day
  - Included ~ 15 T2s sites on LCG by the end of the second week
  - Maximum export rate (per hour) ~ 700 MB/s (Nominal rate ~ 780 MB/s (with NGDF))
  - ATLAS regional contacts were actively participating in some of the T1/T2 clouds
  - Put in place monitoring system allowing sites to see their rates (disk/tape areas), data assignments, errors in the last hours, per file, dataset, …
  - FTS channels in place between T0 and T1 and now progressing between T1 and T2s
  - Exported a total of 1PB of data by Sunday August 6th
- Problems with VO box load have been identified and resolved, whereas adequate monitoring of LFC services at Tier1 sites remains an outstanding issue;
- Major concerns include communication issues with the sites and the serious lack of manpower globally;
- **A range of LFC issues (functionality; usage; deployment; operation) is being addressed**

ATLAS Summary (2/2)

ASGC	after VO BOX upgrade, went very well. 100 MB/s when ATLAS runs; 40~50 MB/s when CMS runs (should be 60 MB/s); communication problems during start-up of exercise
BNL	not using realistic tape area; suffering from read/write contention when using 'production' areas (as opposed to SC4 /dev/null area); very good support for ATLAS
CNAF	unstable Castor-1; now fighting Castor-2 installations. Needs re-evaluation during next phase
LYON	very good service T0->T1 and T1->T2! The only site that was constantly part of the exercise (except for scheduled downtimes).
FZK	after VO BOX upgrade, went better. Still very unstable service (in/out of the exercise all the time)
PIC	stable service; dCache disk area and Castor tape area occasionally suffering some timeouts/overload issues
RAL	not stable; difficult to understand status; could not sustain rate for a few hours. See the LCG Quarterly Report for Q2 2006 for further details of on-going storage issues at RAL.
SARA	very stable service overall
TRIUMF	remains stable; network distance leads to occasional LFC connection glitches

CMS Summary (1/2)

- The main activity during this period was preparation work for CMS CSA06. This involved debugging of data rates into and out of CERN (using PhEDEx over FTS), clarification of FTS channel setup, monitoring and operations and testing of the gLite RB;

- Problems resolved include poor transfers both into and out of CERN (related to the use of the loopback interface for SRM transfers and to incorrect handling at the SRM level of duplicate nameserver entries. Once these problems were resolved, and following tuning at the PhEDEx level, CMS were able to drive transfers at the target rate for CSA06 of 150MB/s (1/4 of the nominal rate);

- Following this successful debugging exercise, an attempt to run at 500MB/s out of CERN for at least 3 days was made. Whilst this target was not reached, the 'threshold' of 300MB/s was attained, with a daily average of 450MB/s on 8th August, with ATLAS and other transfers proceeding in parallel.

CMS Summary (2/2)

- In the 3 month period ending mid-August CMS transferred over 3.3 PB in wide-area transfers between storage systems. Of this, disk-to-disk SC4 transfers account for just over 3 PB and our recent two high-throughput Tier-0/Tier-1 disk-to-disk tests for most of the rest. This translates to an achieved rate of ~1 PB/month in CMS world-wide.

- Specific problems encountered during these tests include various CASTOR2 bugs, such as the fact that CASTOR's reply to the stager_qry command was an arbitrary string that the PhEDEx stager agent had no chance to interpret in a sense that it could determine whether the requested file was on disk or on tape. Therefore it did what it was supposed to do, it submitted a stager_get request for that file. This resulted in a very large number (40K) of stager requests which rapidly overloaded the system. Thanks to Sebastien Ponce and his team the problem was quickly analyzed and a temporary fix was made available to CMS yesterday noon. The permanent fix is expected to be rolled-out by mid September;

- Both CMS and LHCb experienced poor transfer rates into CERN (LHCb from worker nodes used opportunistically, CMS during the centralization of MC data as preparation for CSA06). These problems were eventually traced to the HTAR and have now been resolved. However, the intervention on the HTAR that led to these problems did not follow the agreed procedure for scheduling and announcing such changes and it is imperative that these procedures are rigorously followed in the future;

- Work on patching and tuning the gLite RB as preparation for CSA06 (in collaboration with ATLAS) has been successful. Thus the CMS requirement to handle 50K jobs / day on less than 10 RBs can be met.

CMS Site Summary

ASGC	Initial problems with transfers, mass storage and hardware reliability. Good progress on solving the known problems. Good response in implementing the workflow validation steps. Good support for CMS
CNAF	Initial problems with transfers due to CASTOR2 migration. Good local support for CMS. Good response implementing the workflow and validation steps. Excellent participation in CMS activities. Good support for CMS.
LYON	Some problems with the mass storage configuration and maintaining reasonable transfers under load. Good response implementing the workflow validation steps. Good support for CMS.
FZK	Initial slow completion of workflow validation steps and slow progress on transfer activities. Since the early part of the summer, progress has quicken substantially. Good participation in Tier-1 to Tier-2 transfers and good support for CMS.
PIC	Good response implementing the workflow validation steps. Good support for CMS.
RAL	Continued problems with transfer and the mass storage transition. Completion of workflow validation steps was slow due to hardware availability. Responsive local support.
FNAL	Good response in completing the workflow validation steps. Excellent participation in Tier-1 to Tier-2 transfers during the summer, though slowed by stability limitations toward the end. Good support for CMS.

ALICE Summary (1/2)

- PDC'06 includes the integration of the FTS service into the ALICE File Transfer Daemon (FTD) and to test the operation and stability of the combined system.
  - T0-T1: migration of raw data produced at T0 and 1st pass ESDs also produced at T0
  - T1-T2, T2-T1: transfers of ESD and MonteCarlo data and AODs for custodial storage respectively
  - T1-T1: replication of ESDs and AODs
- Multiple successful transfers have been performed to all T1 sites involved in the exercise, however the target rate of 300MB/sec sustained for a week has not been met. The exercise allowed to expose a number of critical areas and as such was a very important step toward achieving full nominal rates under realistic conditions.
- Some concerns about #files and filesizes used for transfers – this is being addressed

ALICE Summary (2/2)

CNAF	Unstable overall with different sources of errors, most frequent are related to inaccessible storage (CASTOR2). Max rate achieved: 28.4 MB/s.
RAL	Joined the exercise late (site-wide issues with disk storage). Difficult to debug problems, transfers stay in waiting status without clear reason or fail. Resources for ALICE are not sufficient for the duration of the exercise – 1.8 TB of disk without garbage collector. Still in a setup phase.
CCIN2P3	Generally very stable. Problems with srm_get. Max rate achieved: 121.4 MB/s.
SARA	Problems with the LFC catalogue (backend ORACLE), associated to null comments inserted by ALICE. VO-box instabilities have adverse effect on transfers. Max rate achieved 47.6 MB/s.
GridKA	Unstable overall with a variety of error: SRM connection refused and transfer timeouts. Max rate achieved 164.3 MB/s.
CERN	Hardware problems with the VO-box, affecting the transfers to all centres. VO-box replaced.

LHCb Summary (1/2)

- The goals of the LHCb DC06 activity are as follows:
  - Distribution of RAW data from CERN to Tier-1's
  - Reconstruction/stripping at Tier-1's including CERN
  - DST distribution to CERN & other Tier-1's
- Simulated data are shipped to the 6 T1s + CERN with a share that depends on the computing power and status of the site. The amount of data processed is correlated to the amount of integrated data transferred out of CERN to various T1. So far the integrated rate is small (but close to a final draft of the computing model : ~3MB/s to each T1).
- Problems at NIKHEF/SARA (dcap callback mechanism incompatible with network setup – resolved in a beta version of dCache) and at Lyon (use of gsidcap not yet supported by a production version of ROOT) impacted production, although temporary workarounds were found in both cases. For the above reason, the NIKHEF/SARA share is set to 0;

LHCb Summary (2/2)

CERN	ran smoothly its share of jobs during the first month. Some issues with the AFS area serving the Software Installation Area that currently prevents to install jobs through a normal grid job. Problems with the Castor storage in uploading files from simulation jobs running on the small centers (due to the HTAR configuration) and also in the grid mapfile creation that seems to be uncorrelated to VOMS/LDAP mechanism as it happens somewhere else. Flickering behavior of the Information System.
CNAF	potentially CNAF is the largest center and could process the largest share of data. However it suffered a long standing problem with Castor2 stager. Basically at CNAF are using a different configuration to at CERN where for each VO there is a dedicated instance of the DB and LSF. There are several reasons behind: 1. The single disk server serving the LHCb requests from LSF was not enough. There was also a limit on the max number of jobs per disk server increased to 300. (Fixed) 2. The DB is overloaded (deadlocks) and all the requests to the stager are stuck (fixed) 3. The pure disk pool (no Garbage Collector) seems to have problem in accessing files in case it becomes full (with consequent pending jobs overloading the LSF queue) Now CNAF should be OK.
LYON	ran smoothly DC06. Some minor issues due to the storage. They are using at Lyon the disk only storage instead of the tape endpoint (this last supporting only gsidcap protocol). Length of the largest queue doesn't fit with the LHCb Simulation jobs. Flickering Information System also experienced there.
FZK	Poor the usage of GridKA for reconstruction jobs of this DC06 (because it prevents to access data directly from the application), it has been rather used for production. The main problem (under investigation) seems related to their gridftp daemons that decide to close their sockets from time to time.
PIC	some issues with the storage; recent issue with pilot jobs that were not picking up any production (either reconstruction or simulation) job. PIC ran its share without any other major problem.
RAL	also ran smoothly DC06 reco jobs without major issues. Experienced a slowness accessing data at some point and problem fixed by adding another disk server.
SARA	NIKHEF/SARA never used for reconstruction: it is currently impossible accessing (through Root) data stored in the WAN connected Storage at SARA from WN via dcache.A patched version of the dCache client has been released for test. This version doesn't require Inbound connectivity on the WN because it wouldn't require calls client back. Site admins at NIKHEF are very collaborative and are pushing for testing/certifying new dcache libraries needed by LHCb. Once there will be prove that new clients are working fine they will install in their nodes without waiting official release of LCG. Experiment side also tests with the application are ongoing. They didn't yet manage to access file with gsidcap and these new dCache clients. Until further news, NIKHEF sits out DC06 activity.

SC4 Review

Discussion of Service Levels, Intervention Times & Availability Targets

Worldwide LHC Computing Grid

Distributed Production Environment for Physics data Processing

WLCG Service Availability Targets - CERN

- Based on experience of Service Phases of SC3 & SC4, where do we stand with respect to the Service Availability targets in the MoU?

➢ Take 2 concrete examples:

1. Event reconstruction;
2. Distribution of data to Tier1s during run.

– What are the main WLCG & VO-specific services involved?

– How can targets be met? Implications?

Services Examples

1st pass processing, data export

- These two services are characterised by strong dependence on both VO and IT provided services
- Data export introduces a further coupling to storage services at Tier1 sites
- Cannot meet targets without on-call services!
- Typical interruptions:
  - 02:00 weekdays until 10:00
  - 14:00 Saturday until Monday 10:00

WLCG MoU Targets

Service	*Maximum delay in responding to operational problems*			*Average availability[1] measured on an annual basis*	
	Service interruption	Degradation of the capacity of the service by more than 50%	Degradation of the capacity of the service by more than 20%	During accelerator operation	At all other times
Raw data recording	4 hours	6 hours	6 hours	99%	n/a
Event reconstruction or distribution of data to Tier-1 Centres during accelerator operation	6 hours	6 hours	12 hours	99%	n/a
Networking service to Tier-1 Centres during accelerator operation	6 hours	6 hours	12 hours	99%	n/a
All other Tier-0 services	12 hours	24 hours	48 hours	98%	98%
All other services[2] – prime service hours[3]	1 hour	1 hour	4 hours	98%	98%
All other services[2] – outside prime service hours[3]	12 hours	24 hours	48 hours	97%	97%

[1] (time running)/(scheduled up-time)
[2] Services essential to the running of the Centre and to those who are using it.
[3] Prime service hours for the Host Laboratory: 08:00-18:00 in the time zone of the Host Laboratory, Monday-Friday, except public holidays and scheduled laboratory closures.

Event Reconstruction

- It is assumed that event reconstruction is performed using the local batch system, i.e. LSF
- Other services involved include the conditions database service used by the experiment in question (an Oracle-based application for all except ALICE), the experiment-specific book-keeping system(s) (typically based on Oracle and/or MySQL), the LFC (either as a file catalog or as the basis of the CMS DLS), as well as CASTOR2;
  - In the recent ATLAS Tier0 exercise, DDM/LFC operations were decoupled leaving dependencies only on CASTOR, LSF and AFS;
  - In this exercise, AFS was the primary bottleneck and cause of job failures. This is being followed up (e.g. by the use of volume replication);
  - Overall LSF performed worse than in the previous test – leading to the suggestion that a dedicated instance for first pass processing might be needed;
  - CASTOR exceeded the goal of 1 week of stable operation but with a pool 2-times over-dimensioned and Atlas wasted time trying to understand its performance;
- In summary, steps are being taken to ensure reliable services, although coupling to CASTOR, LSF and AFS (and presumably experiment-specific services) remains. All of these services are complex and problems typically require 'the expert' to be solved;

Distribution of Data (1/2)

- This activity is loosely coupled to the former, in that it requires the output of the reconstruction phase. It is, by definition, tightly coupled to the storage management services of the host laboratory (CASTOR + SRM, hence also Oracle and LSF), as well as the FTS (which also depends on Oracle), the experiment-specific framework that drives the FTS, as well as the corresponding storage management services at all of the Tier1 sites supporting a given VO;
- Except in the case of failure or severe degradation of host laboratory services, problems with a single site can, in principle, be tolerated (provided that the site in question has the proven ability to rapidly catch up with a backlog, however caused (e.g. source/sink error, or both));
- On the assumption that recovery from backlogs is demonstrated, expert coverage can probably be limited to ~12-16 hours per day. Although inter-site problems typically require dialog between experts on both sides, more than 2/3 of the data is sent to European sites, where the maximum time difference is 1 hour;
- (Sites must still respond to site-local problems as per MoU)

Distribution of Data (2/2)

- In the case of data export to the Tier1 sites, corresponding on-call services are required at the Tier1s as well, together with inter-site contacts and escalation procedures;
- We note that GGUS and COD currently provide a service during office hours (of the site in question) only, but should provide the primary problem reporting route during such periods. This requires that realistic VO-specific transfer tests are provide in the SAM (or equivalent) framework, together with the appropriate documentation and procedures;
- The list of contacts and the procedures for handling out-of-hours problems will be elaborated by the WLCG Service Coordination team and presented to the Management Board for approval. These procedures will be constructed to facilitate their eventual adoption by standard operations teams, should extended cover ever be provided. We note that such a service may address problem determination, but will not, with the current structures, provide problem resolution.
  - Detailed proposal for service monitoring enhancements also available

The Worldwide LHC Computing Grid

Summary of Tier2 Workshop / Tutorials Questionnaire

Workshop Agenda

Worldwide LHC Computing Grid

Distributed Production Environment for Physics data Processing

WLCG Tier2 Workshop

- 2nd SC4 Workshop, with primary focus on "new Tier2s"
  - i.e. those not (fully) involved in SC activities so far
  - 1-2 people obviously didn't know this from responses ☺
  - Complementary to Mumbai "Tier1" workshop
- Attempted to get Tier2s heavily involved in:
  - Planning the workshop (content)
  - The event itself
    - Chairing sessions, giving presentations and tutorials, …
- Less successful in this than hoped – room for improvement!
  - Some 'volunteers' from FZK-cloud w/s – more from Asia-Pacific?

- Questionnaire from Jeremy Coles, with input from Michel Jouvin, Graeme Stewart, Kilian Schwarz, Michael Ernst, JDS

Workshop Feedback

- >160 people registered and (a few more) participated!
  - This is very large for a workshop – about same as Mumbai
    - Some comments related directly to this (~40 replies received so far)
- Requests for more:
  - Tutorials, particularly "hands-on"
  - Direct Tier2 involvement
  - Feedback sessions, planning concrete actions etc.
- **Active help from Tier2s in preparing / defining future events would be much appreciated**
  - Please not just the usual suspects…
- See also Duncan Rand's talk to GridPP16
  - Some slides included below (hidden)

What did I expect?

- An overview of the future
  - the big picture
  - more details about the experiments
  - data flows and rates
  - how were they going to use the Tier-2 sites?
  - what did they expect from us?

- Perhaps, a tour of the LHC or an experiment

Questionnaire (1/2)

1. What aspect(s) of the workshop/tutorial did you find most useful?
2. What aspect(s) of the workshop/tutorial did you find least useful?
3. If you were organising the next event what would you do differently?
4. What is the single most important thing you learnt this week that you would like to have known before?
5. On a scale of 1 (no use) to 10 (couldn't live without it) how would you rate the workshop?
6. On a scale of 1 (no use) to 10 (couldn't live without it) how would you rate the tutorials?
7. Did you ask any questions or enter the discussions during the workshop?
8. Did you ask any questions or enter the discussions during the tutorials?
9. If you answered "no" to either 7 or 8 but had questions or points you think should have been considered, what stopped you from making more of a contribution? What were those questions or comments?

Questionnaire (2/2)

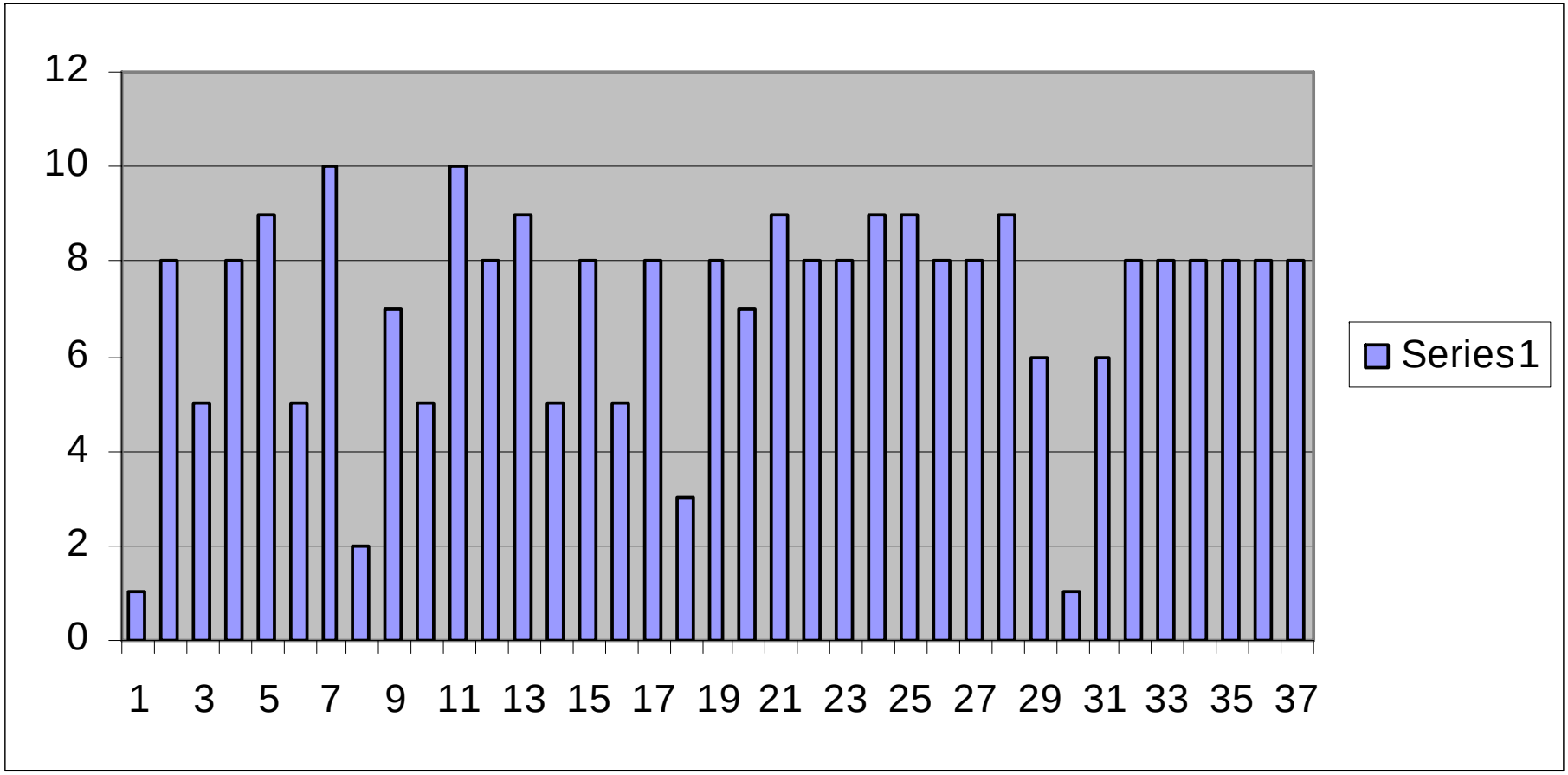
10. Are you confident that you are well positioned to contribute committed resources to WLCG at required service levels on behalf of your institute and country? If not what do you think is missing?
11. What sources of operational information do you find most useful in day-to-day running of your site?
12. What information do you feel you currently lack which if available would greatly increase your ability to contribute to WLCG and meeting the experiment needs?
13. What is currently your single biggest concern in respect of the WLCG project?
14. What would you like covered at the next workshop or tutorials?
15. Was this the first service challenge related meeting / workshop that you attended?
16. How did you find the sessions on experiment different use cases? (Too long, Perfect, Too short)
17. What is your background? (Computer Scientist, Physicist, Both, Other)
18. Are there any other comments or suggestions you would like to make?

Questionnaire (2/2)

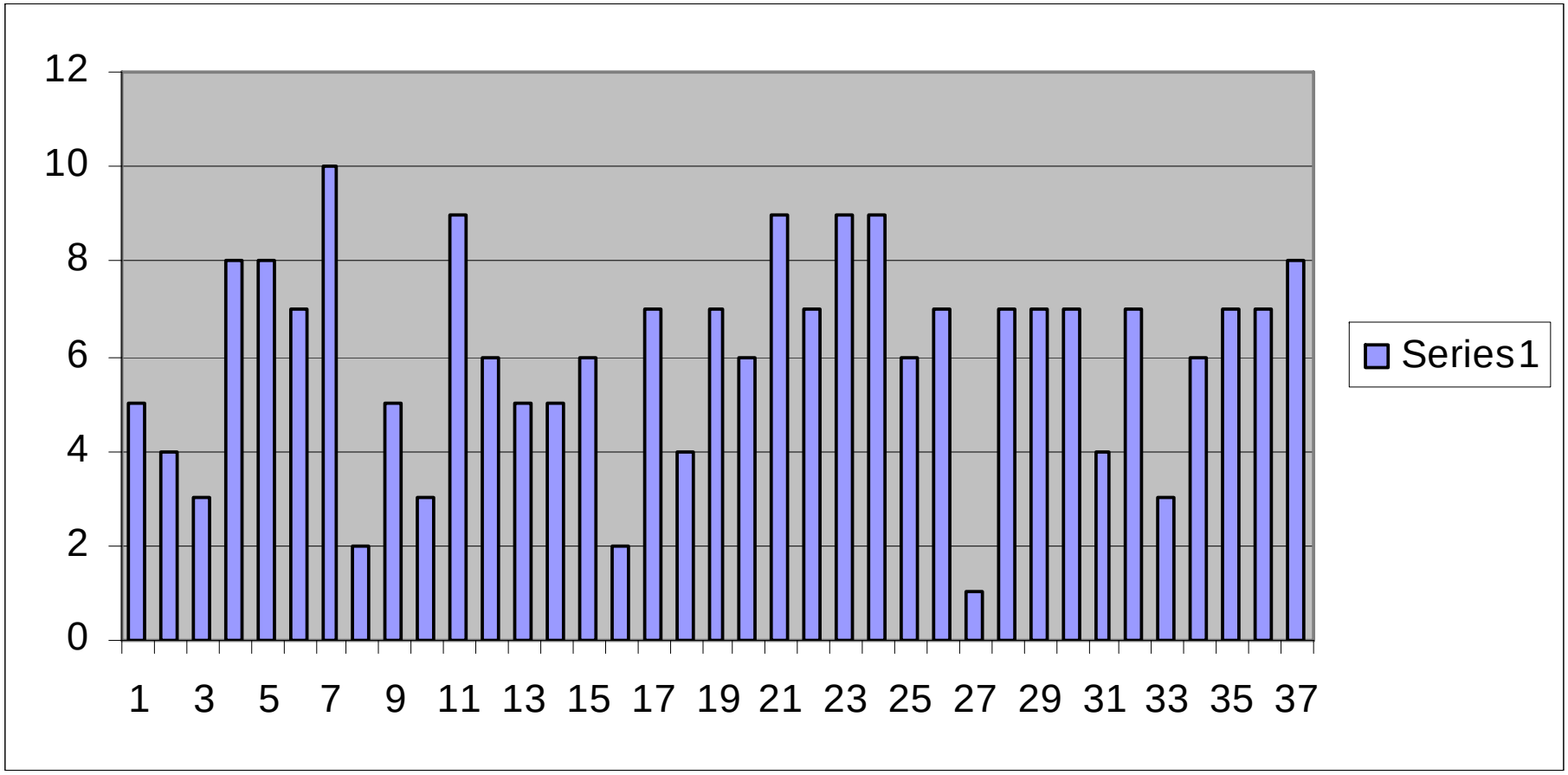
10. Are you confident that you are well positioned to contribute committed resources to WLCG at required ser[illegible] institute and country? If not what d[illegible]
11. What sources of o[illegible] day-to-day running of your si[illegible]
12. What inform[illegible] uld greatly increase you[illegible] nt needs?
13. What is curren[illegible] project?
14. What would y[illegible]?
15. Was this the first se[illegible] hop that you attended?
16. How did you find [illegible] ns on experiment different use cases? (Too long, Perfect, Too sh[illegible]
17. What is your background? (Computer Scientist, Physicist, Both, Other)
18. Are there any other comments or suggestions you would like to make?

Why are there so many questions in this questionnaire?

Tutorial Rating - 10=best



Workshop Rating



Question 10

10. Are you confident that you are well positioned to contribute committed resources to WLCG at required service levels on behalf of your institute and country? If not what do you think is missing?

- 1 NO
- 1 “not entirely”
- 1 “I hope so”
- The rest said YES!
  - Quite a few “Yes, but”s or “Yes, if”s

Middleware tutorials

- Popular – lots of discussion
- Understandable given fact that Tier-2 system admins more interested in middleware than experimental computing models
- Good to be able to hear roadmap for LFC, DPM, FTS, SFT's etc. from middleware developers and ask questions

Tier-2 interaction

- Didn't appear to be much interaction between Tier-2's
- Lack of name badges?
- Missed chance to find out how others do things
- Michel Jouvin from GRIF (Paris) gave a summary of his survey on Tier-2's
  - large variation between resources at Tier-2's
  - 1 to 8 sites per Tier-2; 1 to 13 FTE!
- Difference between distributed vs. federated Tier-2's?
- Post-workshop survey excellent idea

Providing a Service

- We are the users and customers of the middleware
- Tier-2 providing a *service* for experiments
  - CMS: ‘Your *customers* will be remote users’
- Tier-2's need to generate a customer service mentality
- Need good communication paths to ensure this works well
- CMS have VRVS integration meetings and email list – sounds promising
- Not very clear how other experiments will communicate pro-actively

Summary

- Learnt a lot about how the experiments intend to use Tier-2's
- Pretty clear about what they need from Tier-2 sites
- Could have been more feedback from Tier-2's
- Could have been more interaction between Tier-2's
- Tier-2's are critical to success of LHC: service mentality
- Communication between experiments and Tier-2's unclear

The LHC juggernaut is changing up a gear !

Workshop Gripes

- Why no visit to e.g. ATLAS? (Visit during SC Tech day)
- Why no introduction to particle physics?
- ➢ These things could clearly have been arranged
- ***Why no suggestion in the meeting Wiki?***
- Why no name badges? (We had CERN access cards, but not for 'locals'…)
- Start time (11:00 on Mon&Tue) (dictated by room availability)
- Better coffee, air conditioning, …
- ➢ More involvement of Tier1s and Tier2s

Workshop Comments

"Very very inspiring" "Hope to do it again soon"

"Tutorials were very useful"

"The organisation was excellent"

"Discussions were very enlightening"

"Information collected together in one place"

Future Workshops

- Suggest 'regional' workshops to analyse results of experiment activities in SC4 during Q3/Q4 this year – important to drill down to details / problems / solutions
- A 'global' workshop early 2007 focussing on experiment plans for 2007
- Another just prior to CHEP
- Given the size of the WLCG collaboration, these events are likely to be **BIG!**
- Few suitable meeting rooms at CERN – need to plan well in advance
- Something like 2 per year? Co-locate with CHEP / other events where possible?
- Quite a few comments suggesting HEPiX-like issues. Co-locate with HEPiX?
- A one-size-fits-all event is probably not going to succeed…

Communication

- U.S. Resource Allocation Committee (RAC)
- Tier1-Tier2 meetings every 2 months (1-2 days)
  - Rotate among sites
  - Attended by management, site coordinators, and most importantly technical staff from all sites
  - Site requirements are discussed and decided jointly
  - At UC - http://indico.cern.ch/conferenceDisplay.py?confId=a062200
  - At HU- http://indico.cern.ch/conferenceDisplay.py?confId=4897
  - Dec 8 – next meeting at UTA
- Site trouble ticket system
  - http://www.usatlas.bnl.gov/twiki/bin/view/Support/RTUsersGuide
- Panda mailing list, various Savannah, various hypernews
- Regular tutorials (examples: production, DDM, user…)

WLCG Collaboration Workshop

Jan 23-25 2007, CERN

- This workshop will cover: For each LHC experiment, detailed plans / requirements / timescales for 2007 activities.

- Exactly what (technical detail) is required where (sites by name), by which date, coordination & follow-up, responsibles, contacts, etc etc. There will also be an initial session covering the status of the various software / middleware and outlook. Do we also cover operations / support?

➢ From feedback received so far, looks like an explicit interactive planning session would be a good idea

  - **Dates:** 22 January 2007 09:00 to 26 January 2007 18:00
  - **Location:** *CERN*, **Room**: Main auditorium
  - WLCG CB meeting Thursday afternoon in 6th floor conference room

➢ **Do we need tutorials? (YES!) If so, what topics? Who can help?**

➢ **Other ideas? Expert panel Q&A? International advisory panel?**

WLCG Collaboration Workshop

Sep 1-2 2007, Victoria, BC

- Workshop focussing on service needs for initial data taking: commissioning, calibration and alignment, early physics. Target audience: all active sites plus experiments

- *We start with a detailed update on the schedule and operation of the accelerator for 2007/2008, followed by similar sessions from each experiment.*

- *We wrap-up with a session on operations and support, leaving a slot for parallel sessions (e.g. 'regional' meetings, such as GridPP etc.) before the foreseen social event on Sunday evening.*

- **Dates:** 1-2 September 2007
- **Location:** *Victoria, BC, Canada*, co-located with CHEP 2007

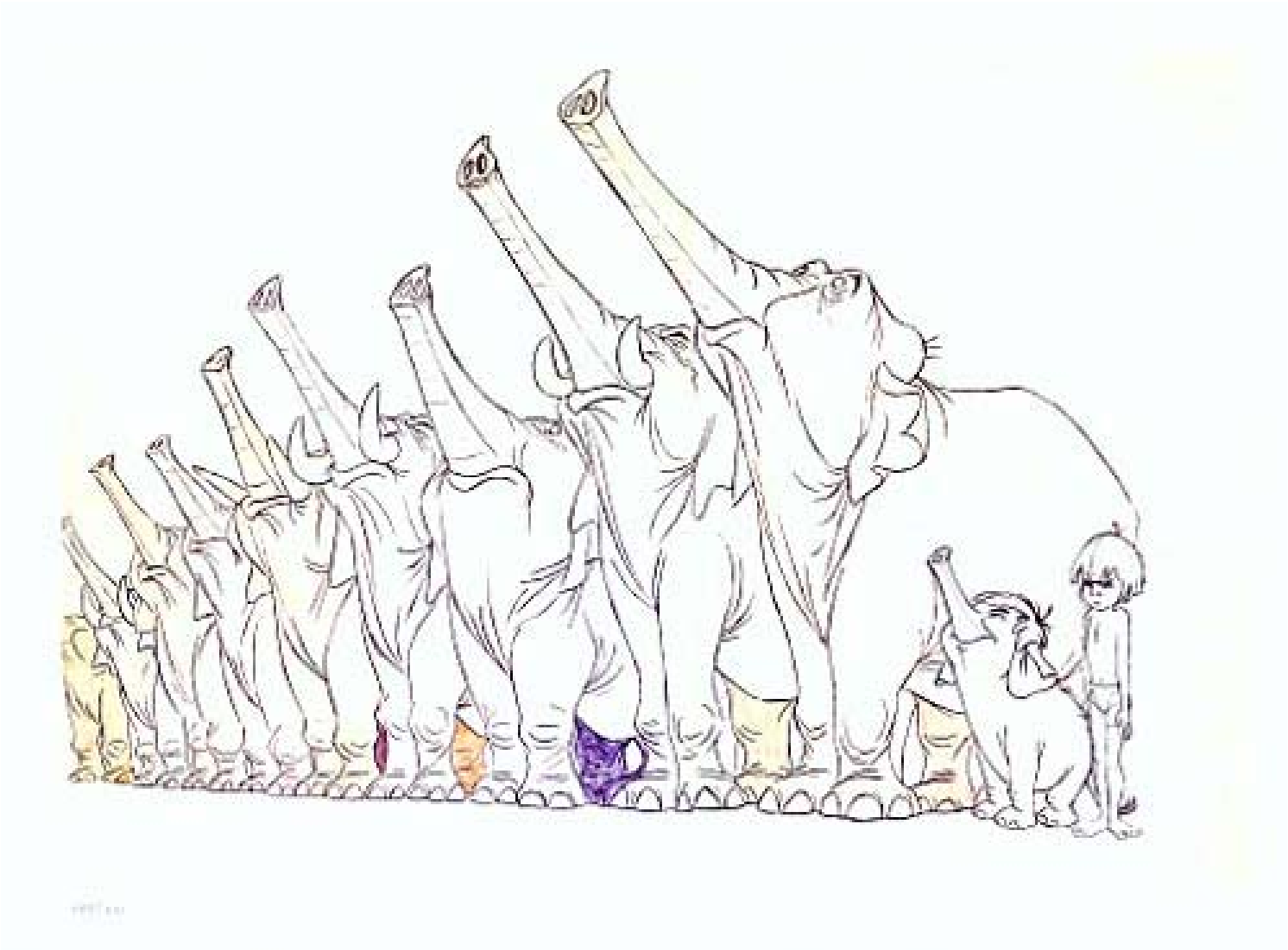
Summary

- Workshops have been well attended and received
  - Feedback will help guide future events

➢ Need to improve on Tier1+Tier2 involvement
  - Preparing agenda / chairing sessions / giving talks etc.

- Strong demand for more tutorials
  - Hands-on where possible / appropriate

- Hopefully lots of volunteers to help with future events...



SC4 Review

Outlook & Conclusions

Worldwide LHC Computing Grid

Distributed Production Environment for Physics data Processing

Service Outlook

- Good progress on improving services has been achieved over the past year
  - But there is clearly still a long way to go…

- **Quite a few additional / or upgraded components are expected prior to first data**
  - See summary of requirements from Dario

- We still have a lot to accomplish!

Outlook

- Service Challenge 3 to Service Challenge 4 involved only 'minor' changes
- From Service Challenge 4 to << LHC startup, we need to understand:
  - Migration to gLite 3.x;
  - Implications of SL(C)4;
  - Deployment of SRM 2.2(-compliant) solutions;
  - Production 3D-services as part of WLCG;
  - Other new services ???
- **We also need a coordinated exercise to prepare for Tier1<->Tier1 and Tier1<->Tier2 transfers**
- Continue to improve Service Level & Response times!

SC3 / SC4 – Conclusions

- For all its problems, SC3 and more completely SC4 have resulted in **production services** across **many sites**
- A great deal of work has been done in setting up the necessary infrastructures
  - Much work still remains to be done
  - New problems need to be uncovered and solved!
- We need to continue to work closely together on **concrete** and **realistic** targets!

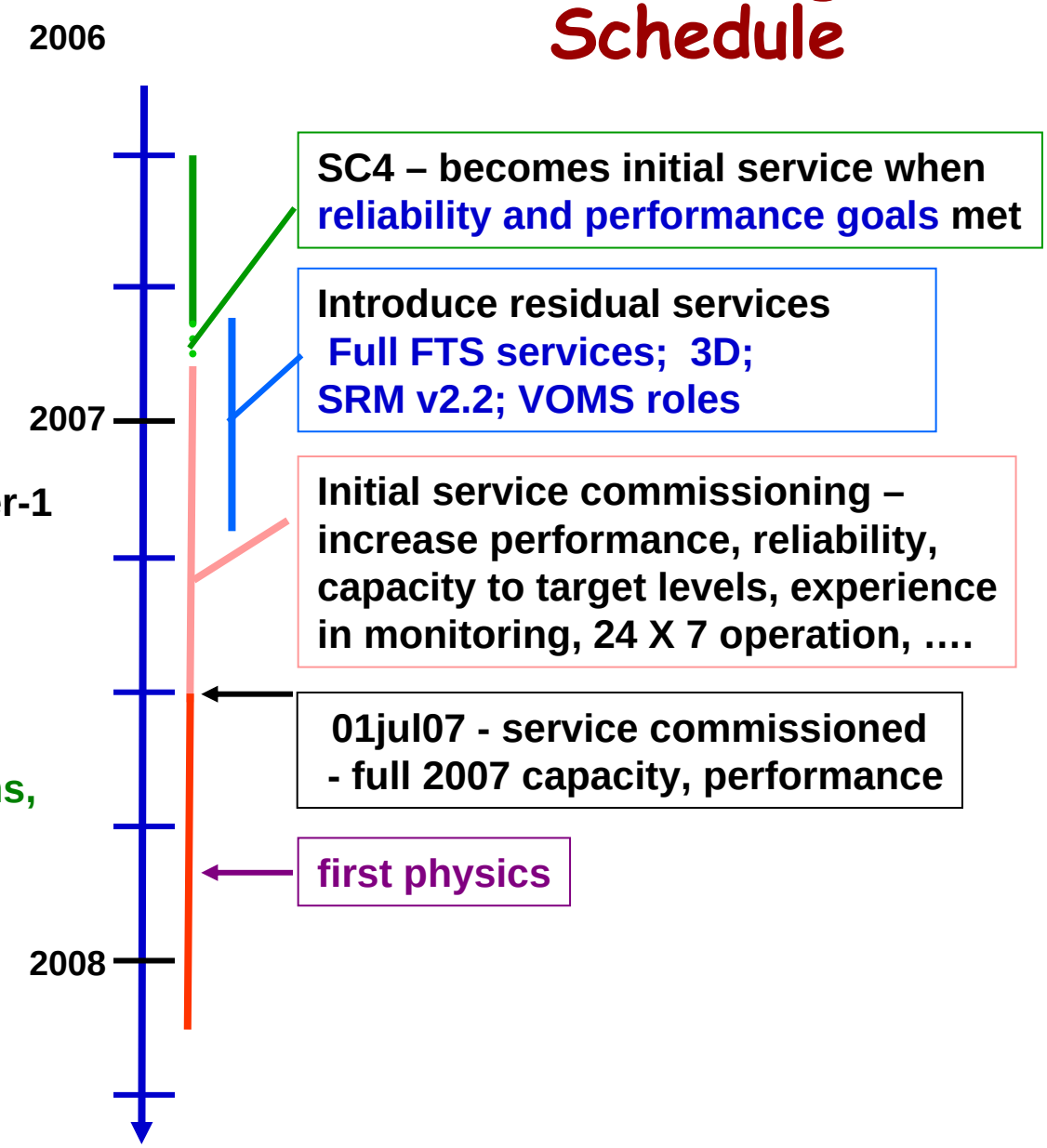
Commissioning Schedule

Continued testing of computing models, basic services

Testing DAQ→Tier-0 (??) & integrating into DAQ→Tier-0→Tier-1 data flow

Building up end-user analysis support

Exercising the computing systems, ramping up job rates, data management performance, ….



Summary & Conclusions

- Deploying a Worldwide Production Grid is not without its challenges
- Much has been accomplished; much still outstanding
- My two top issues?
  - Collaboration & communication at such a scale requires significant and constant effort
    - We are not yet at the level that this is just basic infrastructure
  - "Design for failure" – i.e. assume that things don't work, rather than hope that they always do!
    - A lesson from our "founding fathers" – the creators of the Internet?