Scale Test for the Experiment Program: STEP'09

Jamie.Shiers@cern.ch

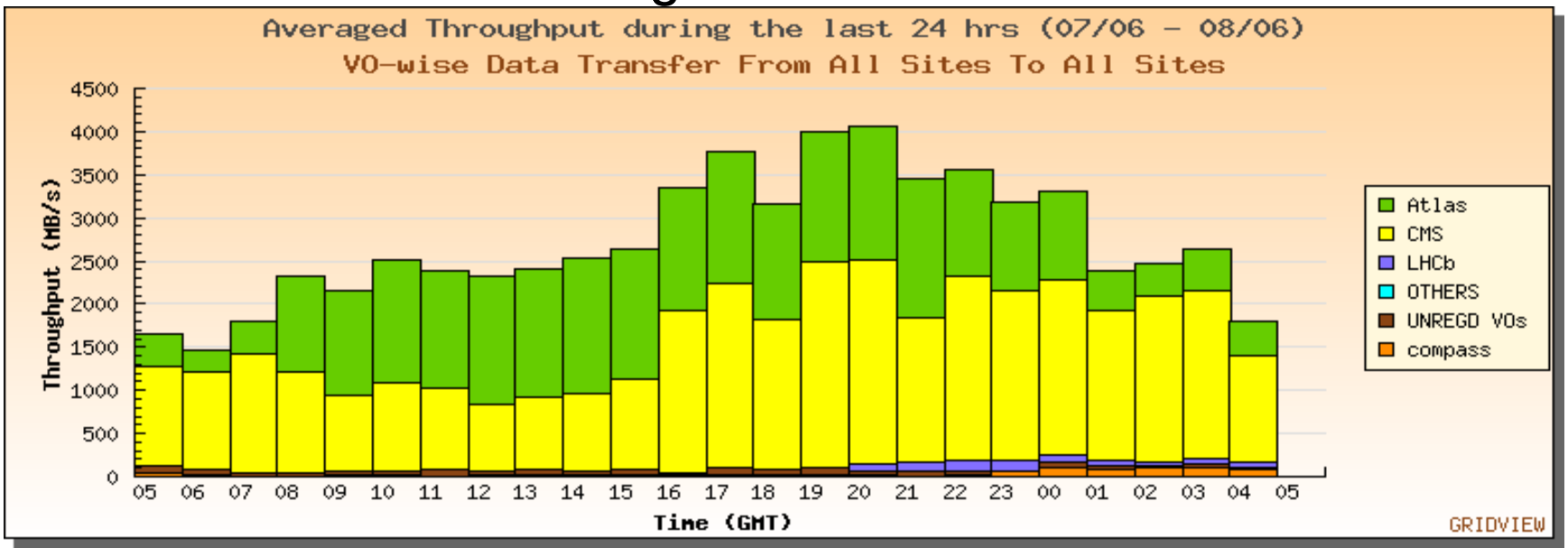
~ ~ ~

June GDB

Overview

- In this talk I intend to take a somewhat wider view of STEP'09 than simply "the first two weeks of June" – closer to "Q2 2009"
  - **See also Harry's talk to yesterday's MB F2F**

- I hope that this is consistent with Matthias' comment at the kick-off in Prague:



Site Readiness

- In the past, we attempted to achieve a period of stability prior to the start of a challenge (which this is not...) but never succeeded
- In other words, we ***failed*** to achieve a one-month period of stability (and we said 1 month wasn't long enough for all sites)
- This time we did not try explicitly, so you could say we **succeeded** in **not** having a period of **stability**,
- Or that we **succeeded** in having a period of **instability**
- Or that we were **unsuccessful** in having a period of **stability**
- This was – at least partially – as not everyone buys in to MB decisions (e.g. scheduling changes right up to the last minute and beyond, including Fridays, is contrary to our principles)
- Several **major** site issues in the weeks & months leading up to June: **we must learn from these!**

Major Service Incidents

- In February 2008 we agreed that sites should spontaneously produce a Service Incident Report whenever there was a major service degradation or downtime. We have re-discussed this innumerable times since, including in Prague (see Olof's presentation)

➢ **This is the one site metric that I set in Prague.**

- To rephrase it: when you have such an incident, do you:

Produce SIR?	**Usually**	**Sometimes**	**Rarely**
Use Template?	**Usually**	**Sometimes**	**Rarely**

☹ It is clear from the archive of EGEE broadcasts that there have been major incidents or service degradation / downtime for which no SIR has been produced...

☹ **An e-mail is not good enough (traceability?)**

TWiki - Mozilla Firefox

arks Tools Help

ome https://twiki.cern.ch/twiki/bin/view/LCG/WLCGServiceIncidents Google

mie S... AudioPresenter Document Search Geneva Weather Indico [IT-GS] Indico [LCG] SIMBA2 WHO/UN Per-diem

WLCGServiceIncidents < LCG < T... PostMortem20090603 < FIOgroup < T... Panda monitor and browser

Jump

Edit WYSIWYG

You are here: TWiki > LCG Web > WLCGCommonComputingReadinessChallenges > WLCGOperationsWeb > WLCGServiceIncidents r22 - 08 Jun 2

WLCG Service Incident Reports

- N.B. Downtimes / degradations are always "user visible" (which is what counts...)

Site	Date	Duration	Service	Impact	Report
CERN	4 Jun 09	n/a	CASTOR LHCb	accidental garbage collection of tape0disk1 files	https://twiki.cern.ch/twiki/bin/view/FIOgroup/PostMortem2
CERN	3 Jun 09	n/a	CASTOR LHCb	accidental re-enabling of garbage collection in lhcbdata	https://twiki.cern.ch/twiki/bin/view/FIOgroup/PostMortem2
CERN	1 Jun 09	~4 hours	DB services	unavailable	https://twiki.cern.ch/twiki/bin/view/PSSGroup /StreamsPostMortem#Network_hardware_problem_affec
PIC	23 - 26 May 09	3 days	LFC	instability	https://twiki.cern.ch/twiki/pub/LCG/WLCGServiceIncident /Post_mortem_LFC_indicent_23-26_May_2009_-_WikiP
PIC	14 May 09	5 hours	cooling	down	SIR_PIC_COOLING_OUTAGE_2009_05_14.pdf
SARA	04 May 09	36 hours	MSS	down	SIR_SARA_TAPEBACKEND_OUTAGE_2009_05_04.pdf
IN2P3	3 May 09	44 hours	cooling	down	SIR_COOLING_OUTAGE_2009_05_03.pdf
IN2P3	25 Apr 09	7.5 hours	MSS	down	SIR_ROBOTIC_LIBRARY_OUTAGE_2009_04_26-3.pdf
IN2P3	20 Apr 09	12 hours	MSS	down	SIR_ROBOTIC_LIBRARY_OUTAGE_2009_04_22.pdf
CERN	12 Apr 09	VOMS: 2	VOMS, SRM	Degraded	VomsPostMortem2009x04x10

Production Services

- In November 2007 we had a workshop focussing on service reliability
- This followec deploy the se
- Much of wha "Common Se
  - Not deployir
  - Not deployir
  - Not countin production a
  - Etc.
- Can we affor providers anc

Pros & Cons – Managed Services

☺ Predictable service level and interventions; fewer interventions, lower stress level and more productivity, good match of expectations with reality, steady and measurable improvements in service quality, more time to work on the physics, more and better science, …

☹ Stress, anger, frustration, burn-out, numerous unpredictable interventions, including additional corrective interventions, unpredictable service level, loss of service, less time to work on physics, less and worse science, loss and / or corruption of data, …

This workshop is about the 1st column

Site Issues

- Pause for some comments from:

Site	Issue(s)
ASGC	Recovery from fire – many issues
FZK	Tape access issues
IN2P3	HPSS scheduled intervention
Others?	

- N.B. we agreed at the WLCG MB that sites should keep the experiments & WLCG service coordination informed in case of major problems – regular updates in case of prolonged issues.

💣 **When this is not done it leads to major time loss and considerable frustration!**

So What Next?

- Whilst there is no doubt that the service has “stepped up” considerably since e.g. one year ago, can:

1. We (providers) live with this level of service and the operations load that it generates?
2. The experiments live with this level of service and the problems that it causes? (Loss of useful work, significant recovery additional work, ...)

- Where are we wrt “the challenge” of CHEP 2004?
  - [The service] *"should not limit ability of physicist to exploit performance of detectors nor LHC's physics potential"*
    - *"…whilst being stable, reliable and easy to use”*


Fragile!
Please do not touch
Ne touchez pas s.v.p.

Caveat

- **N.B. I am focussing on major service incidents / degradations**

☺ **Obviously, its not all like this – some aspects of the service, as I've reported before, are running extremely well**

- It is also true of the experiments & sites and the people involved – the level of commitment and achievement is largely in the range **remarkable** to **amazing**

- But there are clearly areas in which we need to do better – possibly a lot better – hence the comments above

Stepping Up WLCG Operations

- Realistically, major changes are not practical, but operational improvements, such as those introduced in 2008, are!
  - Daily meeting, GGUS enhancements, KPIs, monitoring, etc.
- It has been suggested that something like a monthly **F2F WLCG Operations Meeting** (fitting somehow into the pre-GDB/MB/GDB timetable) is needed
- This would be complementary to the current week-daily operations call and the ~bi- / ~tri-annual WLCG workshops and other events
- Beyond that, I believe that the fairly loose procedures that we have developed together should be adhered to – also by people not in this room!

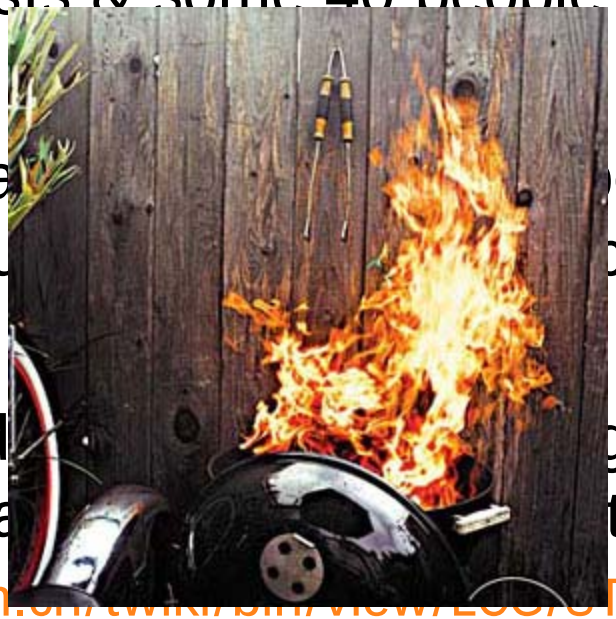
Daily WLCG Operations Meeting

- Has been particularly well attended during the last 10 days or so...
- And has been more useful as a result...
- LHCb has followed the example of CMS in providing a written report through a Wiki – very useful for finding further details and avoiding transcription errors!
- ATLAS, CMS & RAL have STEP'09 wikis – also very useful!
- More details through WLCG Operations wiki
- Please try and join so that we can **start** at 15:00 UTC+2!

- If you send a report or information by e-mail, please send to wlcg-scod@cern.ch (and not to me/Harry!)

STEP'09 Post-Mortem

- Currently scheduled for July 9-10 at CERN (dates imposed by room availability as much as anything else)
- Draft agenda exists & some 40 people have registered
- We will clearly ha[illegible]bout not only the experiments' resu[illegible]ce!
- A BBQ is planned[illegible]g at the Prevessin site (where we ca[illegible]ther things...)
  - https://twiki.cern.ch/twiki/bin/view/LCG/STEP09Barbecue
- Not withstanding the significant site & service problems, the numbers – still to be analyzed – look promising!



STEP'09 – Summary of Goals

Experiment	Summary
ALICE	T0-T1 data replication (100MB/s) Reprocessing with data recalled from tape at T1
ATLAS	Parallel test of all main ATLAS computing activities at the nominal data taking rate -Export data from T0 -Reprocessing and reconstruction at T1 ,tape reading and writing, post-reprocessing data export to T1 and further toT2 -Simulation at T2 (real MC) -Analysis at T2 using 50% of T2 CPU, 25% pilot submission, 25% submitted via WMS
CMS	-T0 multi-VO tape recording -T1, special focus on tape operations data archiving & pre-staging. -Data transfer -Analysis at T2
LHCb	-Data injection into HLT -Data distribution to Tier1 -Reconstruction at Tier1

STEP'09 – Activities During 1st Week

- ***"Though there are issues discovered on daily basis, so far STEP09 activities look good"***
  - IT-GS "C5" report

- More details:
  - https://twiki.cern.ch/twiki/bin/view/LCG/WLCGDailyMeetingsWeek090601

- See also Harry's WLCG Service Summary to yesterday's MB F2F

STEP'09: ATLAS

- Goals:
  - Parallel test of all main tasks at nominal data taking rate
    - Export from Tier 0
    - Reprocessing + reconstruction at Tier 1s; tape reading/writing
    - Export of processed data to other Tier 1s and Tier 2s
    - Simulation at Tier 2
    - Analysis at Tier 2 using 50% of T2 CPU, 25% pilot submission, 25% via WMS
- Progress
  - Started June 1
  - Simulation running at full rate
  - Load generator for data transfers reached 100% on 2nd June
  - Reprocessing running in 7 ATLAS clouds
  - Analysis in progress – using Hammercloud
    - 10-20k jobs concurrently between WMS, Panda, ARC; 130k jobs submitted so far (June 4 – less than 1 day of activity)
    - All clouds receive jobs from both WMS and Panda
    - ATLAS measures efficiency and read performance at each site

STEP'09: CMS

- **Goals:**
  - T0 data recording in parallel with other experiments
    - Plan to run high scale test around 48 hours cosmics data taking runs on 10-11 June & 17-18 June
      - 1st window: 6-8 June (ramping down June 9 to prepare for cosmics data taking)
      - 2nd window: 12-15 June (ramping down June 16 to prepare for cosmics data taking)
    - Ideally longer overlap runs would be desired: but in the time slot where CMS already shifted to meet ATLAS schedule, CMS can afford only two periods up to 4-days long
  - T1 focus on tape access, testing simultaneously prestaging, processing, transfers (2-14 June)
  - Data transfer goals:
    - T1 →T1: replicate 50 TB between all T1s in two phases, first 3-7 June, second 8-12 June
    - T1→T2 (2-14 June): stress T1 tapes and measure latency from T1 MSS to T2, burst tests rather than continuous operation
  - Analysis at T2
    - Demonstrate able to use 50% of pledged resources with analysis jobs, overlaps with MC work. Throughout June.
- **Progress:**
  - STEP'09 started June 2 with pre-staging at the T1 sites
  - June 2-5: CRUZET (Cosmics Running at 0 T) at CERN with export to T1 (so no major T0 activities)
  - STEP'09 activities on T0 ramped up on June 6
  - Reprocessing at T1s started June 3
  - T1→T1 transfers started June 4
  - Analysis: job preparation and ramp up June 2-7, reaching necessary scale and monitoring starting June 8

STEP'09: ALICE+LHCb

- ALICE goals:
  - Tier 0 – Tier 1 data replication at 100 MB/s
  - Reprocessing with data recall from tape at Tier 1s
- ALICE status:
  - Started June 1
  - 15k concurrent jobs running
  - FTS transfers to start this week

- LHCb goals:
  - Data injection into HLT
  - Data distribution to Tier 1s
  - Reconstruction at Tier 1s
- LHCb status:
  - Will join STEP'09 this week

Summary

1. There are still too many significant service degradations or outages – a reasonable fraction (1/3?) of which could be avoided or minimized (see below);

2. Significant improvements in WLCG operations and service delivery have been made over the past years: this includes orderly recovery from "unavoidable" problems, such as power & cooling;

3. A similar "step-up" in service level is required if we are to be able to offer an acceptable level of service at affordable cost for pp data taking, (re-)processing and analysis – all of which is coming "real soon";

4. Some sort of "re-run", following successful testing of sites / clouds that gave problems during STEP'09, as well as greater VO overlap, should be considered. "SEPT'09"?