THE LCG SERVICE CHALLENGES – FOCUS ON SC3 RERUN

Jamie Shiers, CERN, Geneva, Switzerland

Abstract

The LCG Service Challenges are aimed at achieving the goal of a production quality world-wide Grid that meets the requirements of the LHC experiments in terms of functionality and scale. This talk highlights the main goals of the Service Challenge programme, significant milestones as well as the key services that have been validated in production by the LHC experiments. The LCG Service Challenge programme currently involves both the 4 LHC experiments as well as many sites, including the Tier0, all Tier1s as well as a number of key Tier2s, allowing all primary data flows to be demonstrated. The functionality so far addresses all primary offline Use Cases of the experiments except for analysis, the latter being addressed in the final challenge - scheduled to run from April until September 2006 - prior to delivery of the full production Worldwide LHC Computing Service.

INTRODUCTION

CERN is in the final stages of preparing a new flagship particle collider – the Large Hadron Collider (LHC). This facility is scheduled to enter operation in April 2007 and will generate unprecedented volumes of data that will require enormous amounts of computing power to process. Together with many collaborating institutes and national and regional projects including the EU-funded EGEE and the Open Science Grid (OSG) in the US, CERN is deploying a world-wide Grid – at a scale beyond what has previously been achieved in production – for this purpose. Whilst many of the individual components that make up the overall system are understood or even deployed and tested, much work remains to be done to reach the required level of capacity, reliability and ease-of-use. These problems are compounded not only by the inherently distributed nature of the Grid, but also by the need to get large numbers of institutes and individuals, all with existing, concurrent and sometimes conflicting commitments, to work together on an incredibly aggressive timescale.

In a nutshell, the LCG must be ready at full production capability, ready to run reliably for many months at a time, in less than one year from now. Problems related to hardware acquisition, personnel hiring and training and indeed product rollout from various industrial suppliers, must all be taken into account.

KEY PRINCIPLES

The most fundamental underlying principle is that these challenges result in a series of services that exist in parallel with the baseline production service and that rapidly approach the production needs of the LHC. The initial challenges will focus on the core services and will swiftly expand outwards to cover the full spectrum of the production and analysis chain. The service challenges must be run in an environment that is as realistic as possible, which includes end-to-end testing of all key experiment use- cases over an extended period, demonstrating that the inevitable glitches and longer-term failures can be handled gracefully and recovered from automatically. In addition, as the service level is built up by subsequent challenges, they must be maintained as stable production services on which the experiments test their computing models. As a result, the necessary resources and commitment from all partners are a pre-requisite to success. The overall effort and commitment should not be underestimated – we are effectively in (pre-)production mode already: next stop – 2020.

SERVICE CHALLENGE 3

Service Challenge 3 consisted of two distinct phases: a *setup phase*, starting in July 2005, during which disk-disk and disk-tape throughput tests between the T0 and T1s, as well as from a number of pilot T2 sites to participating T1s, followed by a *service phase*, from September until the end of the year, during which production activities from the LHC experiments were scheduled. All participating sites were required to install an SRM, with the transfers scheduled between SRMs. The target data rates from T2 sites to T1s were modest – simply to upload of the order of 3 1GB files per hour – corresponding to the order of magnitude data rate expected from Monte Carlo production at T2 sites. The main goal in this respect was to test the functionality of the underlying services, rather than a major throughput challenge. Between the T0 and T1s, on the other hand, the goals were more ambitious. It was foreseen that each participating T1 receive data at 150MB/s (disk – disk) over extended periods, followed by disk – tape at participating sites at 60MB/s. For the disk – disk tests, CERN was supposed to deliver an aggregate data rate of 1GB/s.

For a variety of technical reasons, these rates were not achieved – some 500MB/s was delivered out of CERN, but with widely varying rates site by site. This resulted in

an extensive period of debugging, leading to upgrading software and hardware configurations at all sites.

In order to test whether these changes had indeed solved the underlying issues, a re-run of the disk – disk and disk – tape tests was scheduled for early in 2006. The remainder of this paper focuses on these re-runs – other significant aspects of the Service Challenge programme and the state of readiness of LHC computing can be found in [1]. Further details of experiment activities as well as infrastructure preparations can be found in papers from the Distributed Event Processing and Production, the Grid middleware and e-Infrastructure operation and other tracks from this conference.

DISK – DISK TRANSFER RATES

In contrast to SC3, the target data rates per T1 site were updated to reflect the nominal data rates foreseen for pp operation of the LHC accelerator, but limited to a maximum of 150MB/s – the global site target for the initial SC3 throughput phase. These data rates were calculated based on the percentage of requested resources (CPU, disk and tape) that a given site has pledged to the VOs that it will support. These rates – MB/s – are given in the table below.

Centre	ALICE	ATLAS	CMS	LHCb	Rate
ASGC		x	x		100
TRIUMF		x			50
BNL		x			150
FNAL			x		150
NDGF		x			50
PIC		x	x	x	30*(100)
RAL		x	x	x	
SARA	x	x		x	150
IN2P3	x	x	x	x	150
FZK	x	x	x	x	150
CNAF	x	x	x	x	150
DESY†		x	x		75

A two week transfer period in January 2006 was able to confirm that the improvements put in place since July 2005 resulted in both higher average throughput – as shown in the figure below – as well as more stable data rates to the individual sites.

Figure 1 - Average Disk - Disk Data Rates in Re-run

For comparison, the equivalent rates achieved in July are shown below – roughly half the overall rate.

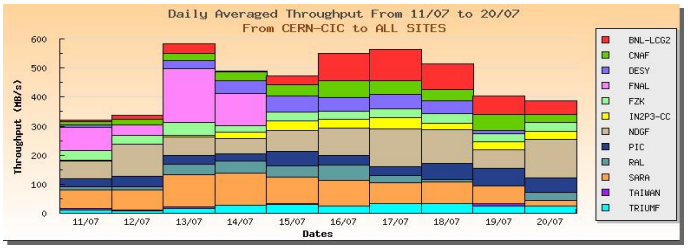


Figure 2 - Average Disk - Disk Rates in July

A snap-shot of a single day shows fairly constant data rates to the various sites, as shown in figure 2 below.

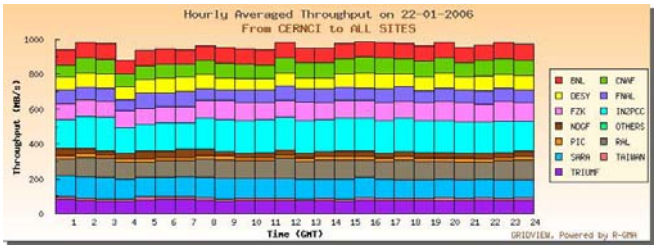


Figure 3 - Data Rates for a Single Day

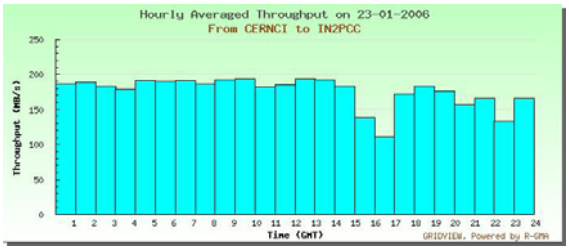


Figure 4 - Data Rates to IN2P3

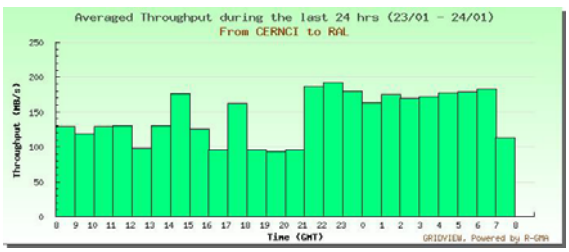


Figure 5 - Data Rates to RAL

* Due to network limitations, the target data rate was reduced wrt the nominal value.

† Although not a Tier1 site, DESY also participates to the major throughput exercises as a centre-of-excellence for Mass Storage and dCache in particular.

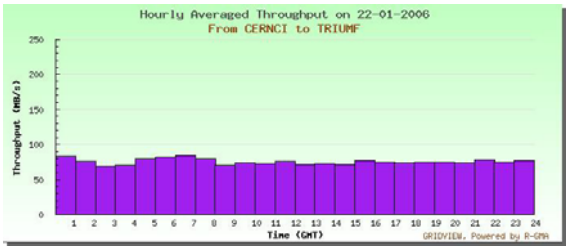


Figure 6 - Data Rates to TRIUMF

In summary, the data rates per site with respect to the targets are shown in the following table. It should be noted that these rates were in general not sustained over long periods. Since the time of CHEP several sites have managed rates significantly higher, as well as for sustained periods as part of the SC4 Throughput tests.

Centre	Target	Achieved
ASGC	100	80 (peaks of 140)
TRIUMF	50	140
BNL	150	150
FNAL	150	>200
NDGF	50	150
PIC	30	>30
RAL	150	200
SARA	150	250
IN2P3	150	200
FZK	150	200
CNAF	150	200
DESY	75	100

DISK – TAPE TRANSFERS

Disk – tape transfers were performed to participating sites immediately following the disk – disk phase. Whereas the goal of the July tests was a flat 60MB/s per site, these rates were slightly modified for the rerun, to reflect what a site might achieve with today's technology tape drives – again capped at the nominal rate for that site. This led to targets of 75MB/s for most sites and 50MB/s for TRIUMF and others.

The transfers highlighted a hole in our ability to monitor the rates – whilst the existing GridView tool shows the rate at which data is transferred out of CERN, it is not able to monitor the rate at which data is written to tape at the far end. Perhaps unsurprisingly for an early attempt, these transfers showed more structure than the disk – disk ones, presumably as disk pools filled up and were drained at the remote sites.

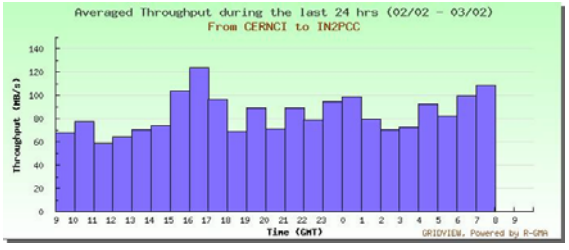
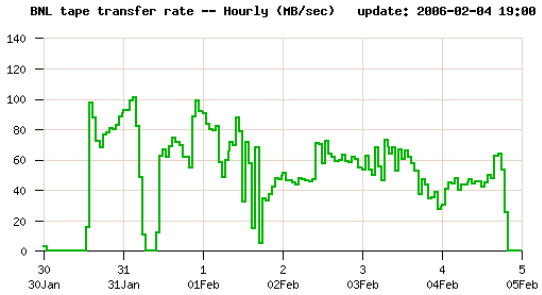


Figure 7 - Transfers to IN2P3 During Disk - Tape Phase



CONCLUSIONS

The Service Challenge 3 throughput re-runs – both disk-disk and disk-tape – where successful in their objectives of demonstrating improved rate and stability. In individual tests, many of the sites demonstrated rates at or even above their nominal rates. Whilst the steps required to ramp-up these data rates to the full nominal values at all sites – for disk-tape transfers – whilst adding the additional complexity of full T0 and T1 activities in parallel – should not be underestimated, this work nevertheless represents an important milestone in the preparations for ful-scalel LHC Computing services.

ACKNOWLEDGEMENTS

Many people have contributed to and continue to contribute to the overall LCG Service Challenge Programme. The focus of this paper has been on the Service Challenge 3 (SC3) re-run performed in January – February of this year (2006). As such, the teams at the participating sites – the CERN T0, all T1s, together with DESY – are warmly thanked for their commitment and hard work. The Service Challenges go far beyond simple throughput tests, and members of the LHC collaborations, the various Grid projects and all participating sites are thanked for their continuing contributions.

It is impossible – for simple space reasons – to mention every individual involved in this work. With apologies to those not explicitly mentioned – whilst acknowledging their work – the following warrant an explicit mention. The CERN-end of the transfers were driven by Maarten Litmaath of the Grid Deployment group. He was given significant support from the FTS developers and supporters, including Gavin McCance and Paolo Badino. The CERN external networking team and FIO group – in particular the CASTOR deployment team – are also warmly thanked for their work. James Casey, who drove the transfers for all preceding Service Challenges and developed much of the infrastructure, including the load generator scripts, deserves a special mention.

REFERENCES

[1] The State of Readiness of LHC Computing, J. Shiers, proceedings of CHEP ’06.

[2] LCG Service Challenge Wiki - https://twiki.cern.ch/twiki/bin/view/LCG/LCGServiceChallenges