THE STATE OF READINESS OF LHC COMPUTING

Jamie Shiers, CERN, Geneva, Switzerland

Abstract

This paper describes the state of readiness of LHC Computing at the time of the CHEP '06 conference in Mumbai. It first attempts to define the global requirements for LHC computing, what "readiness" means in this context and how it could be measured. Proceeding from such a definition, the various metrics by which it could be measured or monitored are reviewed – as well as the status of the monitoring itself – leading to an overall conclusion.

INTRODUCTION

As described in the LCG TDR [1], the LHC Computing Grid Project (LCG) was approved in September 2001 to develop, build and maintain a distributed computing infrastructure for LHC offline computing. Two distinct phases were foreseen:

1. from 2002 – 2005, during which time the necessary software and services would be prototyped and developed;
2. 2006 – 2008, covering the period when the services would be deployed and first data from collisions in the LHC machine delivered and analysed.

We are now well into phase 2 of the project, with first data foreseen for summer 2007 – little more than a year from the time of this conference.

Since the last CHEP conference, held in September 2004 in Interlaken, Switzerland, a number of important developments have taken place in terms of the delivery of key design documents. This includes the following documents:

- The LHC Computing Model documents and corresponding Technical Design Reports [2], [3], [4], [5];
- The associated LCG Technical Design Report [1];
- The LCG Memorandum of Understanding (MoU) [6], which is currently in the process of being signed by the various parties.

Together, these define not only the functionality required (the so-called Use Cases), but also the requirements in terms of Computing, Storage (disk & tape) and Network. Furthermore, a series of workshops, including one immediately preceding this conference, have helped to clarify the services required in terms that are more accessible to the sites.

In addition, we have close-to-agreement on the Services that must be run at each participating site, as well as the roll-out of Service upgrades to address critical missing functionality. Finally, we have an on-going programme to ensure that the service delivered meets the above requirements, including the essential validation by the experiments themselves.

MEASURING OUR READINESS

The LCG MoU [6] lists clear targets in terms of:

- Data transfer rates that must be maintained during data taking and other periods;
- Service availability and time to intervene / resolve problems;
- Resources provisioned across the sites as well as measured usage.

As such, it provides a clear yardstick by which the readiness of LHC computing can be measured.

Of perhaps more direct concern to a physicist would be the "challenge" established at CHEP 2004 [7]:

- [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"***

Whilst these metrics are clearly much more subjective, they are more likely to correspond to how a physicist will judge the service – rather than the mean time to intervene on a specific Grid middleware service at a Tier1 site. Thus, it is important that both the MoU targets and these more end-user oriented measures are kept in mind.

TIMELINE

The official startup date of the LHC machine is currently summer 2007 – a more concrete date, with a precision of some two weeks, is expected in June 2006. This must, therefore, remain our working hypothesis. From looking at previous accelerators, one might expect lower than design luminosity and / or efficiency. However, as was pointed out in [8], a number of physics papers – such as multiplicity and p_t distributions – are ready, simply waiting for data and can be prepared within weeks of first collisions (ALICE). Moreover, should the luminosity be lower than the target, the experiments will simply open the triggers so that the full nominal rate will be achieved – data is urgently required to align and

calibrate the detectors, to test the data acquisition systems and offline frameworks. In addition, cosmic data is already arriving and the need for fully functional production quality services is here today, albeit at a slightly lower scale than required in full data taking mode.

The WLCG service targets are thus:

- The SC4 (WLCG pilot) service phase from 1st June 2006 on;
- The official WLCG production service from 1st October 2006 on;
- Ramp-up to full scale services by April 2007, 6 months before scheduled pp data taking.

On the other hand, initial running is likely to be characterized by poorly understood and calibrated detectors, machine operational parameters and offline software. Whilst this may well mean that reduced datasets, such as AOD and TAG, will still be in the process of being defined, the ESD may well be correspondingly larger and the pressure to resolve these questions and provide first physics results enormous. These requirements cannot possibly be met by the resources that are foreseen to be available at CERN and hence a fully-functional distributed computing environment – aka the Worldwide LHC Computing Grid – will be required from day one.

WLCG SERVICE HIERARCHY

The main responsibilities of the different tiers of the WLCG computing model are as follows:

- Tier0 (CERN): safe keeping of RAW data (first copy); first pass reconstruction, **distribution of RAW data and reconstruction output to Tier1**; reprocessing of data during LHC down-times;

- Tier1: safe keeping of a proportional share of RAW and reconstructed data; large scale **reprocessing** and safe keeping of corresponding output; **distribution of data products to Tier2s** and **safe keeping** of a share of simulated data produced at these Tier2s;

- Tier2: Handling **analysis** requirements and proportional share of **simulated event** production and reconstruction.

There are variations between the experiments – for example LHCb does not plan analysis at Tier2 sites, whereas ATLAS foresee storing two copies of the ESD at Tier1 sites, with a further full copy at BNL. Given the spread in resources that the various ATLAS Tier1 sites will offer, this requires some pairing of sites with approximately balanced resources.

THE WLCG DASHBOARD

As described above, this sounds like a conventional problem for a 'dashboard'. Nevertheless, it is important to stress that there is no single viewpoint. For example, a funding agency may be concerned with how well the resources provided being used. A VO manager may wish to see how well their production is proceeding. A site administrator on the other hand may simply want to see if his or her services up and running and meeting the agreed MoU targets. The on-duty operations team will typically want to know if there any outstanding alarms. Finally, an LHCC referee may want to see how the overall preparation progressing with any areas of concern highlighted. Nevertheless, much of the information that would need to be collected is common and so it is important to separate the collection from presentation (views…), as well as the discussion on metrics.

THE REQUIREMENTS

The overall requirements can be found from a number of sources. For example, information on the **resource requirements**, including the ramp-up in Tier*N* CPU, disk, tape and network, can be found in:

- The Computing TDRs;
- From the resources pledged by the sites (in the appendices of MoU and as reviewed by the C-RRB [11]);
- From the plans submitted by the sites to the LCG Management Board regarding acquisition, installation and commissioning [12].

Most importantly, one needs to measure what is currently and historically available and to signal anomalies.

Similar, the **functional requirements** – in terms of services and service levels, including operations, problem resolution and support, can be found:

- Implicit / explicit requirements in Computing Models;
- Agreements from Baseline Services Working Group and Task Forces;
- Service Level definitions in MoU.

Once again, it is essential to measure what is currently and historically available and to signal anomalies.

Finally, in terms of **data transfer rates**, as well as the overall the Tier*X* →← Tier*Y* matrix, it is necessary to understand the key Use Cases, define realistic tests and once again measure and signal anomalies.

We cover each of these areas in more detail below.

RESOURCE REQUIREMENTS

The computing resource requirements of the LHC experiments are reviewed regularly by the C-RRB. In addition to this somewhat static view, the LCG project requires the Tier0 and Tier1 sites to complete regular planning reports, indicating their acquisition and deployment schedule against the overall resource requirements of the support experiments. Ideally, these would be compared automatically and regularly with the measured, delivered capacity – in terms of CPU, disk and tape storage. This, however, would require sites to systematically publish accounting information in an agreed format – an area where there is still work to be done before full agreement is reached.

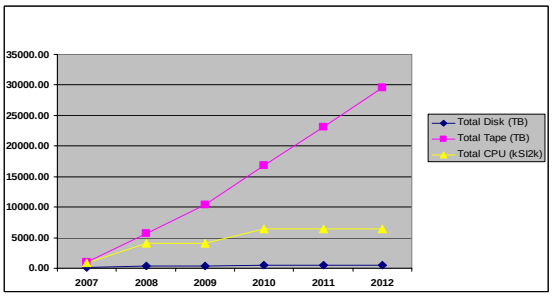

Figure 1 - ATLAS Tier0 Resource Requirements

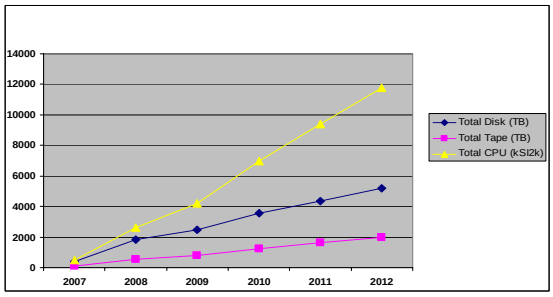

Figure 2 - ATLAS CAF Resource Requirements

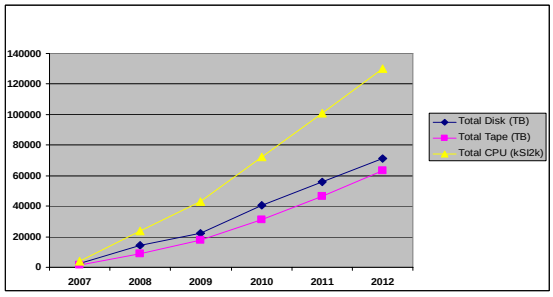

Figure 3 - ATLAS Tier1 Resource Requirements

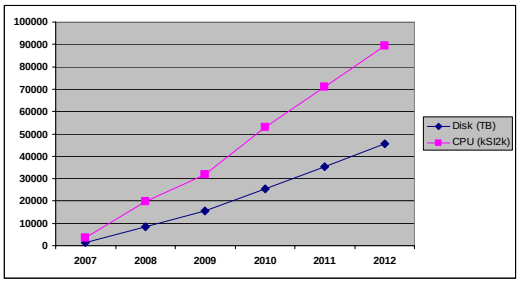

Figure 4 - ATLAS Tier2 Resource Requirements

FUNCTIONAL REQUIREMENTS

The WLCG MoU lists services and target services levels that each site (tier) must provide. Some examples are listed below.

The Host Laboratory shall supply the following services in support of the offline computing systems of all of the LHC Experiments according to their computing models.

i. Operation of the Tier0 facility providing:

1. *high bandwidth network connectivity from the experimental area to the offline computing facility (the networking within the experimental area shall be the responsibility of each Experiment);*
2. *recording and permanent storage in a mass storage system of one copy of the raw data maintained throughout the lifetime of the Experiment;*
3. *distribution of an agreed share of the raw data to each Tier1 Centre, in-line with data acquisition;*
4. *first pass calibration and alignment processing, including sufficient buffer storage of the associated calibration samples for up to 24 hours;*
5. *event reconstruction according to policies agreed with the Experiments and approved by the C-RRB (in the case of pp data, in-line with the data acquisition);*
6. *storage of the reconstructed data on disk and in a mass storage system;*
7. *distribution of an agreed share of the reconstructed data to each Tier1 Centre;*
8. *services for the storage and distribution of current versions of data that are central to the offline operation of the Experiments, according to policies to be agreed with the Experiments.*

…

The following services shall be provided by each of the Tier1 Centres in respect of the LHC Experiments that they serve, according to policies agreed with these

Experiments. With the exception of items Annex 1.1.i, ii, iv, these services also apply to the CERN analysis facility:

i. acceptance of an agreed share of raw data from the Tier0 Centre, keeping up with data acquisition;

ii. acceptance of an agreed share of first-pass reconstructed data from the Tier0 Centre;

iii. acceptance of processed and simulated data from other centres of the WLCG;

iv. recording and archival storage of the accepted share of raw data (distributed back-up);

v. recording and maintenance of processed and simulated data on permanent mass storage;

vi. provision of managed disk storage providing permanent and temporary data storage for files and databases;

…

Similarly the MoU defines target availability and maximum delay before an intervention to resolve a problem occurs, as defined in the following (simplified) table for Tier1 sites.

	Maximum delay in responding to operational problems			***Uptime***	
	Service down	Degradation > 50%	Degradation > 20%	Run	No run
Data from T0	12 hours	12 hours	24 hours	99%	n/a
Network to T0	12 hours	24 hours	48 hours	98%	n/a
Analysis services	24 hours	48 hours	48 hours	n/a	98%
Other – prime shift	2 hour	2 hour	4 hours	98%	98%
Other – outside prime shift	24 hours	48 hours	48 hours	97%	97%

Table 1 - simplified MoU Service Targets for a Tier1

These targets can only be met by design, using a combination of appropriate hardware, with the necessary robustness and failover capabilities in the middleware, as well as appropriate adapted, tested and documented procedures. Over the past months, CERN has redeployed those services with "critical" or "high" requirements in terms of service level availability [13], using a range of standard techniques that would be appropriate to other sites wishing to implement similar services.

The agreement for all sites is that the services will be monitored using the Service Availability Monitoring Environment (SAME) – basically an extension of the existing Site Functional Test framework that has been used successfully for some time. This framework would monitor individual services, which would then be aggregated into higher level services – such as those described in the MoU – and service level availability published and reviewed on a regular basis.

DATA TRANSFER RATES

The data rates that each site must sustain over prolonged periods – basically each LHC running period – are shown in the table below. Furthermore, sites must be capable of running for extended periods – 4 / 8 / 12 hours or more – at up to twice these rates, in order to recover from backlogs in a timely manner. Moreover, as we have recently seen, an outage of the Tier0 centre is by no means implausible, and hence we need to have the capacity to recover from at least 4 hour interruptions in Tier0 operations. The results that have been achieved so far are described further in [9], focusing on the Service Challenge 3 Tier0 to Tier1 disk – disk and disk – tape tests performed in January and February of this year.

Centre	ALICE	ATLAS	CMS	LHCb	Rate
ASGC		x	x		100
TRIUMF		x			50
BNL		x			200
FNAL			x		200
NDGF		x			50
PIC		x	x	x	100
RAL		x	x	x	
SARA	x	x		x	150
IN2P3	x	x	x	x	200
FZK	x	x	x	x	200
CNAF	x	x	x	x	200

Table 2 - nominal rates (to tape) during pp running

SUMMARY OF KEY ISSUES

There are clearly many areas where a great deal of work still remains to be done, including:

- Getting stable, reliable, data transfers up to full rates
- Identifying and testing all other data transfer needs
- Understanding experiments' data placement policy
- Bringing services up to required level – functionality, availability, (operations, support, upgrade schedule, …)
- Delivery and commissioning of needed resources
- Enabling remaining sites to rapidly and effectively participate

- Accurate and concise monitoring, reporting and accounting
- Documentation, training, information dissemination…

Nevertheless, a clear programme and timeline for addressing these issues exists, and is closely monitored by the LCG Grid Deployment and Management Board, as well as the LHCC.

THE DASHBOARD REVISITED

Whilst there are many possible visualisation techniques for our dashboard, one favoured by this author is the Spider or Kiviat plot. Not only does this allow multiple axes to be conveniently displayed – with possible drill-down or even selection of axes depending on the role or primary concerns of the viewer – but it also allows history or time progression to be monitored. In addition, although often used to display subjective information – such as the example below on web content – in the case of the key information that we need to monitor for the WLCG, the scale of the axes is well determined, such as the available CPU, disk and tape resources as compared against the Computing TDR requests.

Ideally, the dashboard should show the various axes gradually saturating, and not pulsating ominously like a tank of baby jellyfish in the cold blue electric light of the Melbourne aquarium.

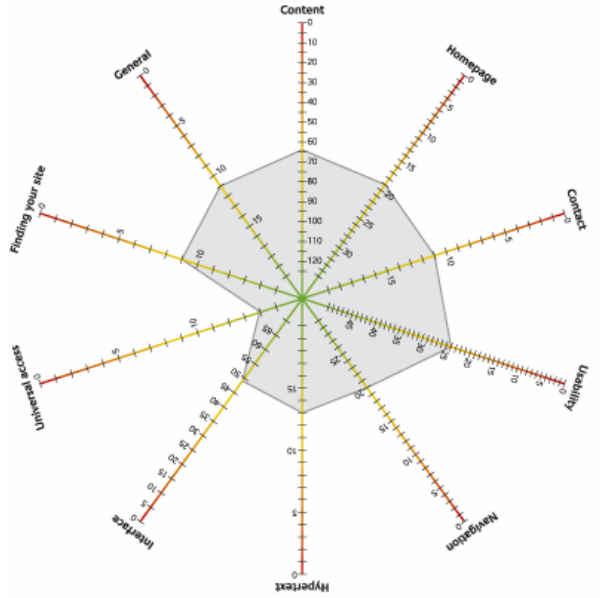

Figure 5 - a possible WLCG Dashboard (Kiviat Plot)

CONCLUSIONS

In the 3 key areas addressed by the WLCG MoU, namely:

- Data transfer rates;
- Service availability and time to resolve problems;
- Resources provisioned

we have made good – sometimes excellent – progress over the last year. There still remains a huge amount to do, but we have a clear plan of how to address these issues. We clearly need to be pragmatic, focussed and work together on these common goals in order to achieve our target by the time of the next CHEP conference in September 2007, to be held in Victoria, British Columbia, Canada.

REFERENCES

[1] The LHC Computing Grid Technical Design Report, LCG-TDR-001 / CERN-LHCC-2005-024.

[2] The ALICE Computing Technical Design Report, CERN-LHCC-2005-018.

[3] The ATLAS Computing Technical Design Report, CERN/LHCC/2005-022 / ATLAS-TDR-017.

[4] The CMS Computing Technical Design Report, CERN/LHCC 2005-023 / CMS TDR 7.

[5] The LHCb Computing Technical Design Report, CERN/LHCC 2005-019 / LHCb TDR 11.

[6] The Worldwide LHC Computing Grid Memorandum of Understanding, CERN-C-RRB-2005-01.

[7] Physics Validation of the LHC Software, Fabiola GIANOTTI (CERN), proceedings of CHEP 2004.

[8] State of readiness of LHC Experiment Software, Paris SPHICAS (CERN), proceedings of this conference.

[9] Service Challenge 3 Re-run, J. Shiers, in the proceedings of this conference.

[10] http://www.chep2007.com/ - Computing in High Energy and Nuclear Physics.

[11] The LHC Computing Resources Review board - http://lcg.web.cern.ch/lcg/Boards/crrb.html

[12] The LCG Management Board planning Wiki: https://uimon.cern.ch/twiki/bin/view/LCG/Planning

[13] The LCG Fabric Tasks Dash: https://twiki.cern.ch/twiki/bin/view/LCG/WlcgScDash