SC4 Plans of the SARA/NIKHEF Tier1

Ron Trompert, Mark van de Sanden
SARA

Version 0.3
Date: October 4th 2005

1.   Introduction

In this document we will outline the plans for the SARA/NIKHEF tier1 site to meet the requirements to fulfill the SC4 goals. In Section 2, we will review the SC4 time line and milestones and derive from them the requirements imposed on the SARA/NIKHEF tier1 site. In Section 3, the infrastructure will be discussed and in Section 4, we will estimate what hardware needs to be in place to fulfill SC4's goals and provide a timeline when to purchase and install the various hardware items. Section 5 deals with the SC4's services and service levels and Section 6 contains some concluding remarks.

2.   Requirements

At the time of writing, LCG has the following time line, [1].

- Now – September 2005: clarification of SC4 Use Cases, components, requirements, services etc.
- October 2005: SRM 2.1 testing starts; FTS/MySQL; target for post-SC3 services.
- January 31st 2006: basic components delivered, in place and production ready.
- February / March: integration testing.
- February: SC4 planning workshop at CHEP (w/e before).
- March 31st 2006: integration testing successfully completed.
- April 2006: throughput tests.
- May 1st 2006: Service Phase starts.
- September 30th 2006: Initial LHC Service in stable operation.
- Summer 2007: first LHC event data.

In addition, the following milestones have been defined for SC4 in the TDR, [2], and in [1]:

Date	Description
31 Jan 06	All required software for baseline services deployed and operational at all Tier-1s and at least 20 Tier-2 sites.
30 Apr 06	Service Challenge 4 Set-up: Set-up complete and basic service demonstrated. Performance and throughput tests complete: Performance goal for each Tier-1 is the nominal data rate that the centre must sustain during LHC operation (see Table 7.2 below) CERN-disk ◊ network ◊ Tier-1-tape. Throughput test goal is to maintain for three weeks an average throughput of 1.6 GB/s from disk at CERN to tape at the Tier-1 sites. All Tier-1 sites must participate. The service must be able to support the full computing model of each experiment, including simulation and end-user batch analysis at Tier-2 centres.
31 May 06	Service Challenge 4: Start of stable service phase, including all Tier-1s and 40 Tier-2 centres.
30 Sept 06	1.6 GB/s data recording demonstration at CERN: Data generator◊ disk ◊ tape sustaining 1.6 GB/s for one week using the CASTOR mass storage system.

|	
30 Sept 06	Initial LHC Service in operation: Capable of handling the full target data rate between CERN and Tier-1s (see Table 7.2). The service will be used for extended testing of the computing systems of the four experiments, for simulation and for processing of cosmic-ray data. During the following six months each site will build up to the full throughput needed for LHC operation, which is twice the nominal data rate.
1 Apr 07	LHC Service Commissioned: A series of performance, throughput and reliability tests completed to show readiness to operate continuously at the target data rate and at twice this data rate for sustained periods.

Table 1: Milestones

The time line and milestones imply that all hardware needed to fulfill SC4's goals should be in place before the end of January 2006, before the start of the integration tests. At that time SARA/NIKHEF should be capable of achieving the target data rate of 150 MB/s[*], [3], disk to tape.

[*]The TDR [2] states that a rate of 175 MB/s should be achieved but since [3] is more recent than the TDR, we assume that 150 MB/s is the rate the SARA/NIKHEF tier1 should achieve.

3.   Infrastructure

The infrastructure we intend to use during SC4 is outlined below in Figure 1. Here SARA will be a tier1 centre while NIKHEF will be a tier2 centre.
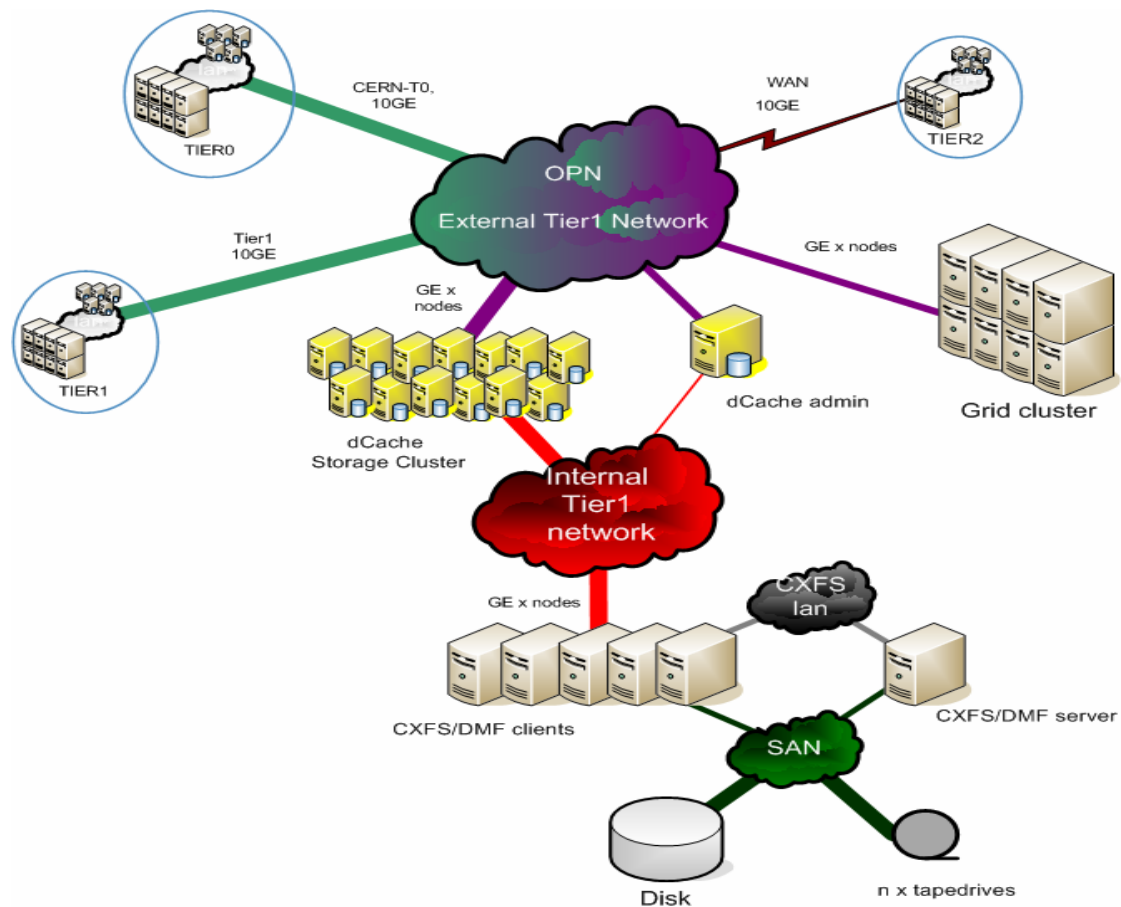
Figure 1: SARA/NIKHEF Tier1 Infrastructure

In Figure 1, the Grid cluster will comprise grid services like VOMS, BDII, RB, FTS, LFC, PX, SE classic, VO boxes, CE, WNs and so on, but also basic services like NTP, DHCP and an installation server to install all nodes of the infrastructure. The dCache storage cluster contains pool as well as gridftp and srm door nodes.

The mode of operation is typically that data is written to the dCache storage cluster. This data can come from CERN or other T1 and T2 sites, either through the OPN or the WAN. From the storage cluster, files are copied to the CXFS/DMF client nodes which are connected to a SAN with disk and tape storage. DMF will eventually copy the files from disk to tape.

4.   Hardware

In order to meet the requirements of SC4 the following hardware and software will be required..

	Required	Currently Available
Storage nodes with 2 TB disk space/node	7	2
Cxfs/dmf client nodes	4	2
Tape drives, 9940B or	10	2
Tape drives, Titanium 500	3	0
Service nodes (worker nodes class system)	5	1
DMF/CXFS License	1	1
DMF/CXFS Server	1	1
disk for SAN	10 TB	10
Tape cartridges	400	100

For the service phase of SC4 it is not yet known what the requirements of the experiments are. Therefore, we have assumed that these requirements will be similar to those of the service phase of SC3.

Our past experience indicates that it is reasonable to assume a throughput rate of about 30 MB/s per storage node and about 50 MB/s per CXFS/DMF client node so 5 storage nodes and 3 CXFS/DMF client nodes will be sufficient to meet the throughput goals of the throughput phase of SC4. Two additional storage nodes and one extra CXFS.DMF client node are necessary to maintain the production service during this period.

In addition to the hardware listed in the table above, we still need to resolve the issue of network equipment. For SC1-3 we have borrowed a Force10 switch from the University of Amsterdam and got the 10G link to CERN through Netherlight. For SC4 we plan to have a dedicated 10G link to CERN solely for the SARA/NIKHEF tier1 and a dedicated tier1 10G infrastructure.

We plan to be in operation in January 2006.

5.   Services and Service Levels

The MoU, [4], lists the following services that should be provided by Tier1 sites:

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;
vii.   provision of access to the stored data by other centres of the WLCG and by named AF's as defined in paragraph 1.4 of the MoU;
viii.   operation of a data-intensive analysis facility;
ix.   provision of other services according to agreed Experiment requirements;
x.   ensure high-capacity network bandwidth and services for data exchange with the Tier0 Centre, as part of an overall plan agreed amongst the Experiments, Tier1 and Tier0 Centres;
xi.   ensure network bandwidth and services for data exchange with Tier1 and Tier2 Centres, as part of an overall plan agreed amongst the Experiments, Tier1 and Tier2 Centres;
xii.   administration of databases required by Experiments at Tier1 Centres.

The SARA/NIKHEF tier1 site plans to provide these services for SC4.

Regarding service levels, the MoU, [4], states the following service levels for the period from the signing of the MoU to December 31st 2012.

Service	Maximum delay in responding to operational problems			Average availability 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
Acceptance of data from the Tier-0 Centre during accelerator operation	12 hours	12 hours	24 hours	99%	n/a
Networking service to the Tier-0 Centre during accelerator operation	12 hours	24 hours	48 hours	98%	n/a
Data-intensive analysis services, including networking to Tier-0, Tier-1 Centres outwith accelerator operation	24 hours	48 hours	48 hours	n/a	98%
All other service – prime service hours[1]	2 hours	2 hours	4 hours	98%	98%
All other services – outwith prime service hours[1]	24 hours	48 hours	48 hours	97%	97%

Table 2: Service levels

As far as SC4 is concerned, the accelerator will not be in operation so the first two rows of this table can be ignored. The maximum delay in responding to operational problems for services during prime hours will be observed for SC4 as these response times in the table are already within the limits of normal operation of the SARA/NIKHEF tier1. The response time outside prime hours is currently based on next business day service level.

6.   References

[1] SC4 Plans, Jamie Shiers, GDB meeting July 20st 2005, http://agenda.cern.ch/askArchive.php?base=agenda&categ=a045324&id=a045324s0t12%2Fmoreinfo%2Fsc4-gdb-july.pdf

[2] LHC Computing Grid Technical Design report, Jürgen Knobloch, LCG-TDR-001, CERN-LHCC-2005-024.

[3] SC4 Plans, Jamie Shiers, pre-GDB meeting September 6th 2005, http://agenda.cern.ch/askArchive.php?base=agenda&categ=a054962&id=a054962s0t13%2Fmoreinfo%2Fsc4-planning-Sep-gdb.pdf

[4] Memorandum of Understanding, CERN-C-RRB-2005-01/Rev. 30 May 2005, http://lcg.web.cern.ch/LCG/C-RRB/MoU/MoU.pdf

[1] Prime service hours for Tier1 Centres: 08:00-18:00 in the time zone of the Tier1 Centre, during the working week of the centre, except public holidays and other scheduled centre closures.