Report from SARA/NIKHEF T1 and associated T2s

Ron Trompert

SARA

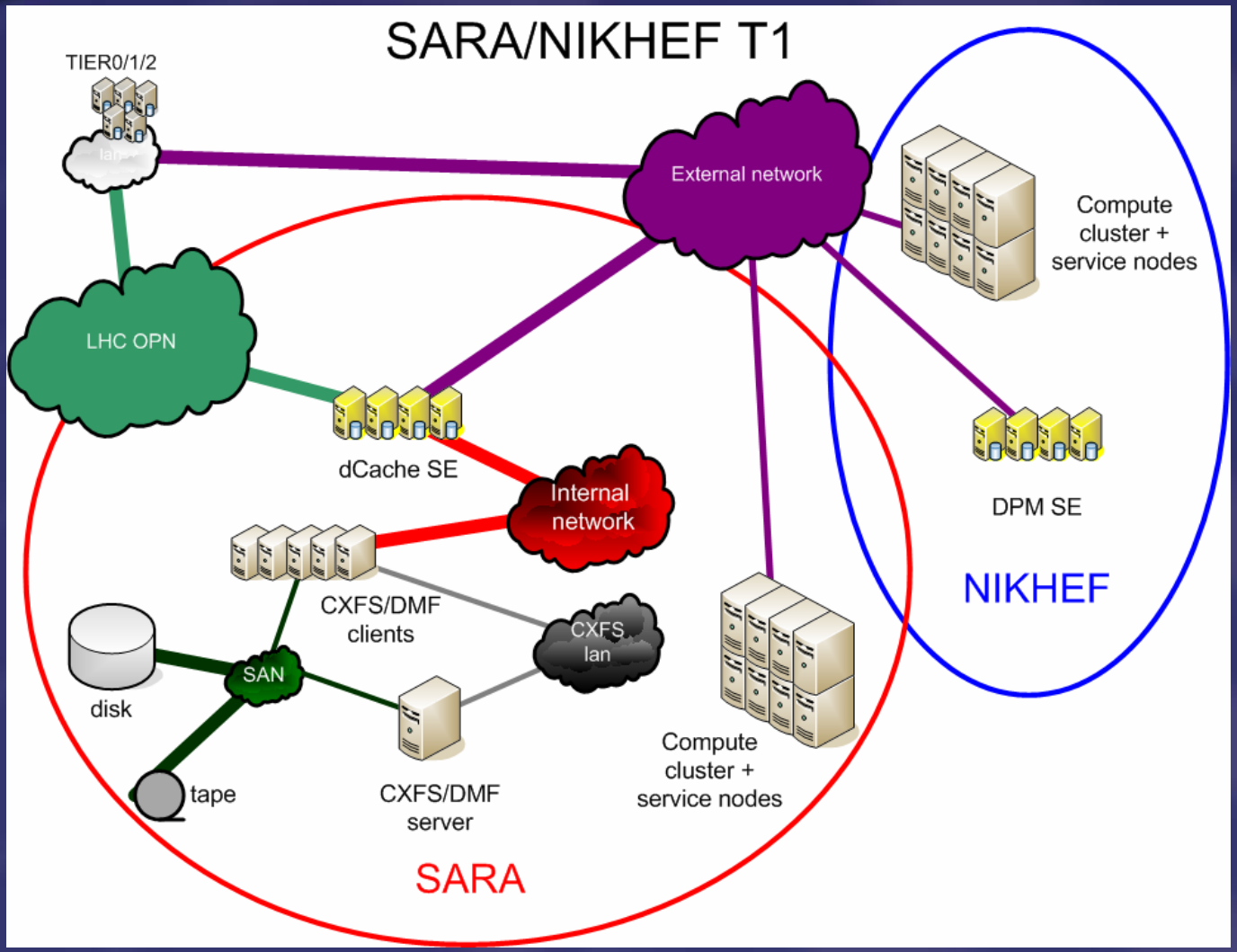
About SARA and NIKHEF

- NIKHEF
  - High Energy Physics Institute
- SARA
  - High performance computing centre
  - Manages the Surfnet 6 network for the Dutch NREN Surfnet.
  - Manages HSM environments
- SARA and NIKHEF are located in the same building but are separate organisations collaborating in WLCG, EGEE and a Dutch eScience project VLe

About SARA and NIKHEF

- Separate computational domains
- Weekly meetings to discuss operational matters, middleware development (EGEE, VLe), etc
- Bi-weekly meetings to discuss planning and other T1 matters

Infrastructure



Deployed hardware

- SARA:
  - 59 TB disk storage for WLCG (dCache)
    - 1.8 TB disk cache for tape silos
    - Alice T0D1: 2TB T1D0: 3TB
    - Atlas T0D1: 12TB T1D0: 4TB
    - LHCb T0D1: 34TB T1D0: 4TB
  - 50 TB of tapes
    - 6 dedicated 9940B drives
    - SL8500 and Powderhorn
  - Computing
    - 37 kSI2k MATRIX 3 GHz Xeon 2 CPUs/node
      - WLCG shared with other VOs
    - 2231 kSI2k LISA (Atlas and LHCb)
      - Shared with other users (shared fairshare of 18%)

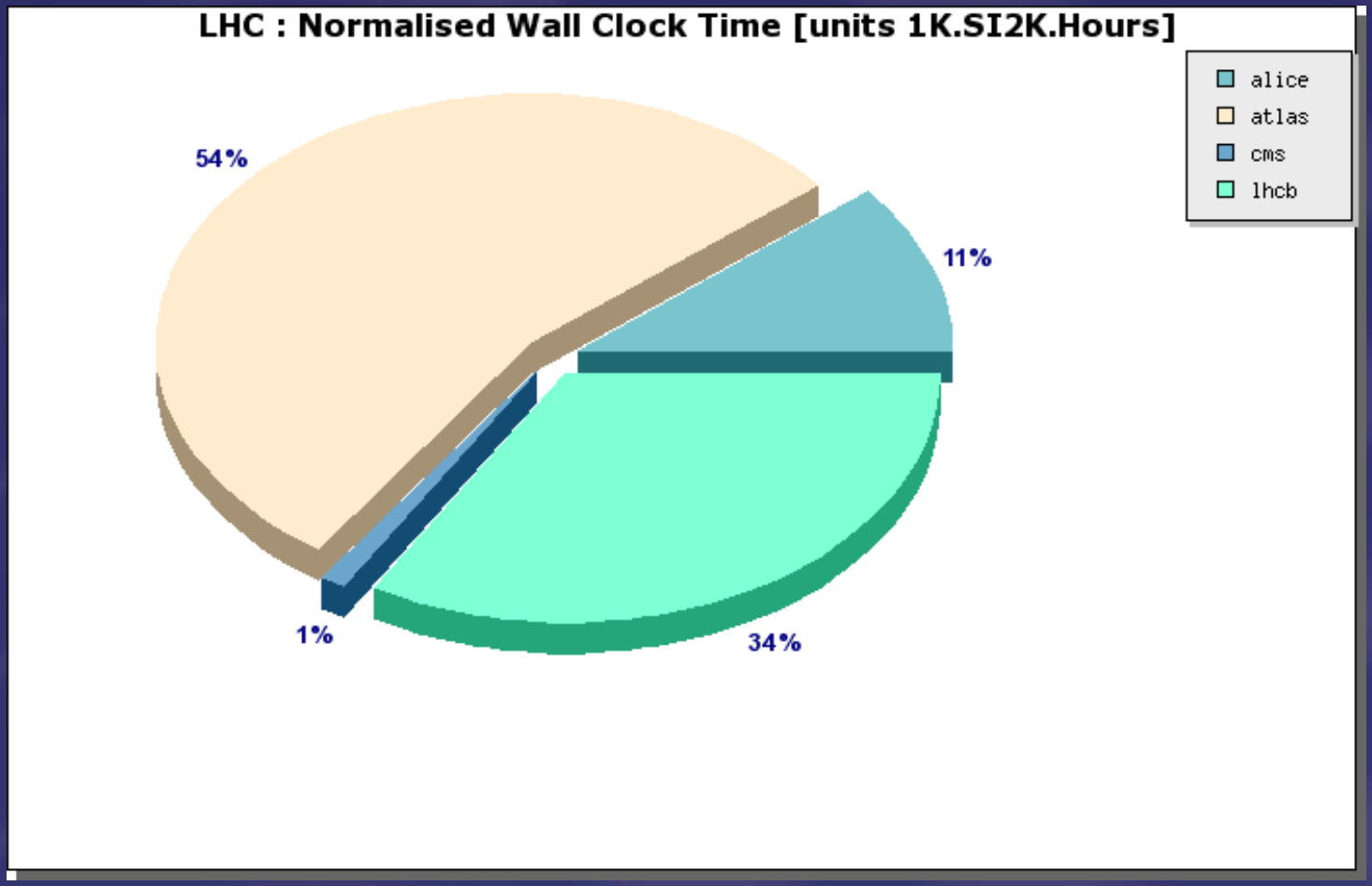
Deployed hardware

- NIKHEF:
  - 14 TB disk storage for WLCG (DPM)
    - Alice T0D1: 1TB
    - Atlas T0D1: 11TB
    - LHCb T0D1: 2TB
  - Computing
    - 305 kSI2k Luilak-1 2.66 GHz Xeon 5150 4 CPUs/node
      - Dedicated for WLCG
    - 305 kSI2k Luilak-2 2.66 GHz Xeon 5150 4 CPUs/node
      - Shared with other (non-LHC) VOs
    - 70 kSI2k Halloween 2.8 GHz Xeon 2 CPUs/node
      - Dedicated for WLCG
    - 106 kSI2k Bulldozer 3.2 GHz Xeon 2 CPUs/node
      - 1/3 for WLCG

3D project

- Two dual Xeon nodes (Oracle RAC) at SARA with SAN storage (300 GB for now)
- Infrastructure is being tested
- Participate in ATLAS replication tests
- DBA workshop at SARA on march 20-21 2007

Metrics



Normalised Wallclock over 2006

Metrics

Normalised Wall Clock Time [units 1K.SI2K.Hours]							
Tier-1	Site	alice	atlas	cms	lhcb	Summed Usage	As a Percentage
SARA/NIKHEF	NIKHEF.NL	215878	1113194	19257	703729	2052058	94.4%
SARA/NIKHEF	SARA-LISA	0	27089	0	4949	32038	1.5%
SARA/NIKHEF	SARA-MATRIX	22527	42376	3632	21214	89749	4.1%
Total per VO		238405	1182659	22889	729892		
As a Percentage		11%	54.4%	1.1%	33.6%		

Normalised Wallclock over 2006

Metrics

Datatransfers in 2006 for SARA SRM

VO	TRANSFERS IN	TRANSFERS OUT	STORE TO TAPE	RESTORE FROM TAPE
ALICE	158 TB	0 TB	0 TB	0 TB
ATLAS	12 TB	29 TB	11 TB	5 TB
LHCB	8 TB	9 TB	5 TB	0 TB

Metrics

Data Stored

	ALICE	ATLAS	LHCB
Disk	2 TB	14 TB	7 TB
Tape	0 TB	18 TB	7 TB

Activities during the past year

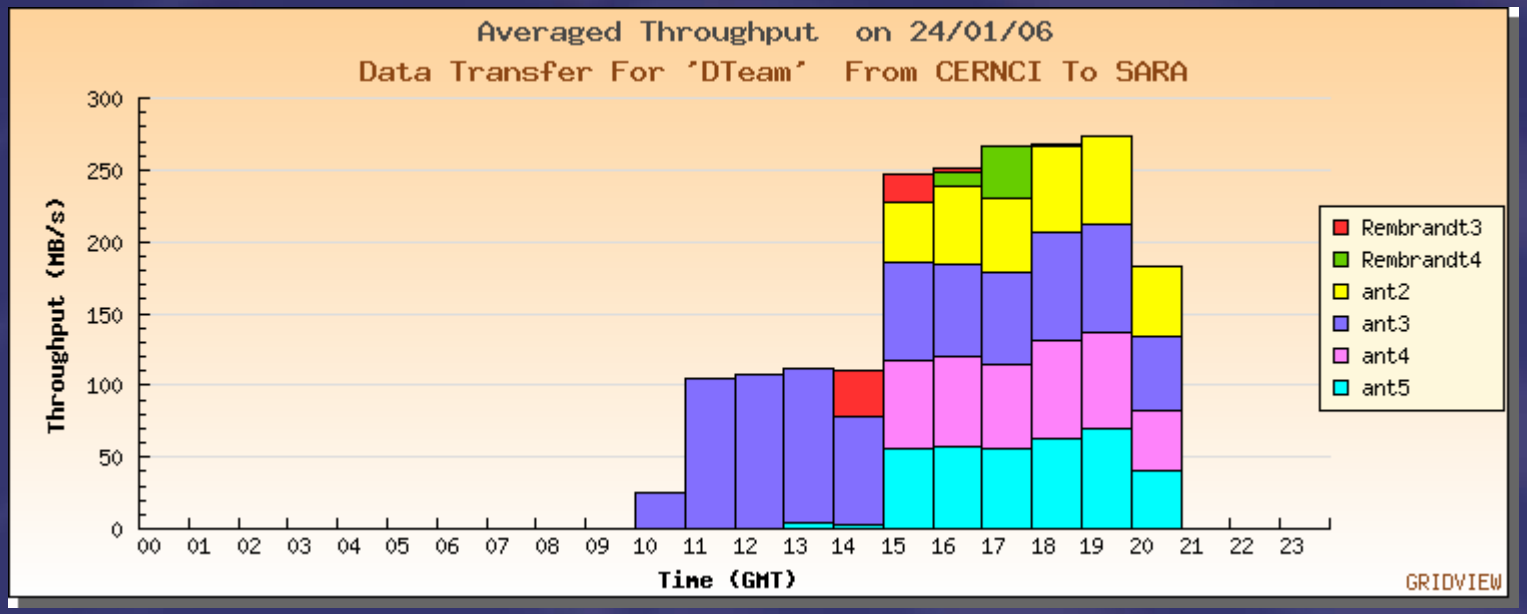
- NIKHEF
  - Ldap vo server grid-vo.nikhef.nl went into retirement
  - Installed latest version of torque and maui and did work on the scheduling (fairshares)
  - Installed new worker nodes
  - Installed DPM SE plus storage

Activities during the past year

- SARA
  - Installed 5 and later on 10 storage servers with a total of 78 TB
  - Installed 4 extra tape drives
  - Separated hsm environments
    - T1 now has it's own dedicated hsm environment
  - Network rearrangement
  - Installed 5 new service nodes for
    - Oracle RAC (2 nodes)
    - Monitoring s/w (Nagios)
    - Dcache admin
    - Install server

Activities during the past year

- SC3 throughput disk2disk test rerun
  - Met target rate of 150 MB/s
  - Peak up to 272 MB/s



Activities during the past year

- SC4 throughput test
  - Disk2disk
    - 154 MB/s for 15 days
    - Slow startup due to some problems
  - Disk2tape
    - Only about 50 MB/s
    - Suboptimal SAN configuration
      - Now write 28+ MB/s per 9940B drive

Funding

- WLCG resources funded by BIGGRID
  - Proposal for the funding of a Dutch e-Science infrastructure.
  - Started January 1st 2007
  - About 28M€ for 4 years.
  - To be shared by a number of participants in the Netherlands

Planning

- On the short term
  - Grid services nodes
    - CE,UI,FTS, …
    - Reliable hardware
      - Double system disks
      - Double power supplies
      - Mirrored memory
    - High availability linux

Planning

- Keeping up with Harry tables

	CPU	Disk	Tape
Q1 2007	742 kSI2k	129 TB	187 TB
Q2 2007	958 kSI2k	156 TB	214 TB

Planning

- 24x7 support
  - Initially not availably
  - Try to achieve the required MoU service level through redundancy.
  - See how it goes
  - If we do not achieve the required service level, we will reconsider this

T2s

- Israel
  - Weizmann
- Russia:
  - IHEP, ITEP, SINP ,JINR, RRC
- Czech republic
  - Prague, alternate data path for FZK
- UK
  - Northgrid sites: Lancaster, Liverpool, Manchester, Sheffield, alternate data path for RAL

ITEP

- 2.3 TB disk (dCache), 1 TB classic SE with xrootd for Alice, 1TB DPM SE.
  Plan to unite all SEs into the dCache SE.
- 32 WNs P4 2.4 GHz dual CPU nodes, lcg-CE and glite-CE
- VOBOX, UI, MON, WMSLB, BDII
- External 1G network

RRC

- 2 disk servers (3x2TB DPM storage + 2 TB shared fs for exp. soft, etc.)
- 50 computing nodes (dual P4 3.0 and 3.4 GHz) available soon
- External network connection is 100Mbps at the moment. Hope to get 1 G this year

JINR

- 30TB dCache SE. Plans for extra 30 TB. 1.8 TB DPM SE
- One farm for all VOs, 15 dual P3 1 GHz nodes. Will be upgraded to dual xeon dual core in Q2
- Second farm 4 dual P4.
- 2 lcg-CE, 2-3 additional CE are planned+torque server
- LFC, BDII, PROXY, WMSLB,RB

SINP

- Classic SE and dCache SE 5.4 TB. Plan to install extra 14 TB
- Lcg-CE with 29 WNs P4 2.8 GHz and dual Xeon 2.8GHz
  glite-CE with 12 WNs dual Opteron dual core 2.4 GHz
- 1 G network connectivity

Trouble tickets systems

- Current situation both NIKHEF and SARA have their own system with a person on duty watching over it.
- Considering single trouble ticket system for SARA/NIKHEF T1.
- EGEE Northern Federation has RT with a site on duty watching over it every other week

Networking

- 10G to CERN
  - Up and running (level 2) but not in production right now because of routing issues
- 10G to FZK
  - Aim to use SARA-FZK-CERN as backup when SARA-CERN fails
- 1G to Lancaster
  - In progress
- 2G connection SARA-NIKHEF
  - Plans to upgrade this to 10G

Issues

- Network
  - Security policy dictates that only traffic related to T0-T1 and T1-T1 data transfer is allowed to go through LHC OPN. Part of the SARA-CERN traffic should go through LHC OPN while the rest goes through Géant->source-based routing
  - We would like to use BGP for redundancy (use the SARA-FZK-CERN link as backup)
  - SBR excludes redundancy and BGP excludes the possibility to discriminate against source ip addresses
  - Sites have solved this issue using dedicated hardware
  - Will be discussed further

Issues

- Storage
  - dCache
    - Gridftp doors stopping due to hanging transfers
      - Watchdog script to restart the gridftpdoor
    - Srm timeouts due to growth of the postgres db which underlies the srm i/f
  - ROOT I/O for LHCb
    - LHCb jobs running at NIKHEF tried to open files at SARA using gsidcap-based ROOT I/O.
    - (gsi)dcap connects back to the WN which is on private NIKHEF IP space which failed
    - Installed dCache 1.7.0 as soon as it was released which has a passive version of dcap ☺

Issues

- LFC
  - Sometimes the LFC loses connection to the oracle db. It tries to reconnect which fails for an unknown reason.
    - Watchdog script checks this every 2 minutes and restarts lfcdaemon
- Randomly failing SAM tests
  - Usual procedure is to look if the service is still running and if so, repeat the failed test by hand. If the test succeeds, invoke the SAM testsuite manually to clear the error status