



CERN data services vs. Service Challenge

Status & Plans

Vladimír Bahyl

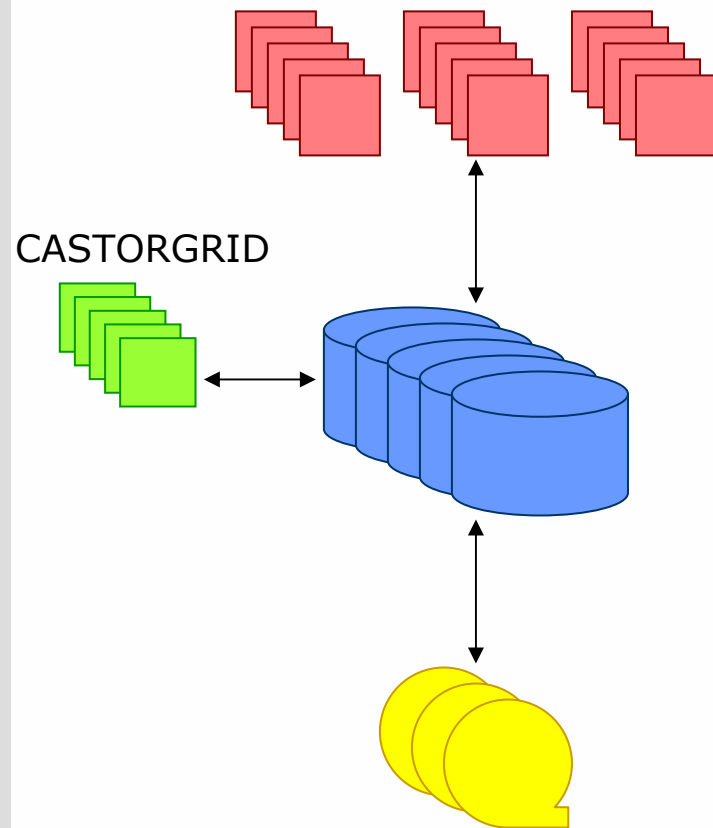
Fabric Infrastructure & Operations

Vladimir.Bahyl@cern.ch

Overview

- Introduction to CERN's data services
- CASTORGRID
 - The WAN data transfer service
- Network connectivity
- Tapes involvement

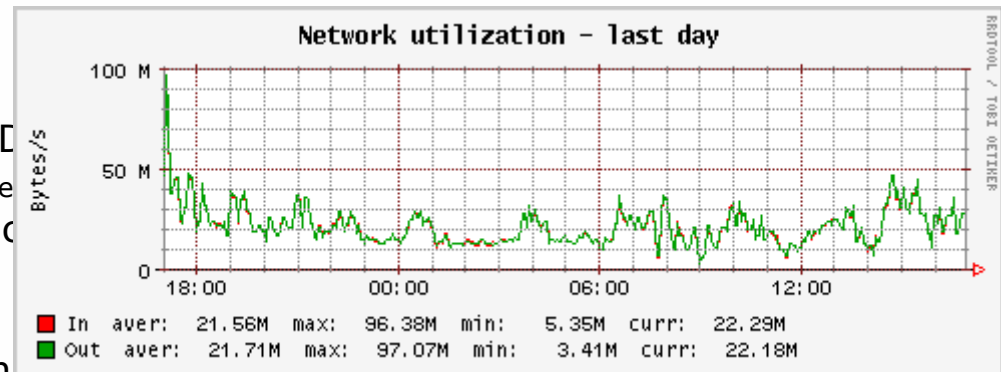
Data Services



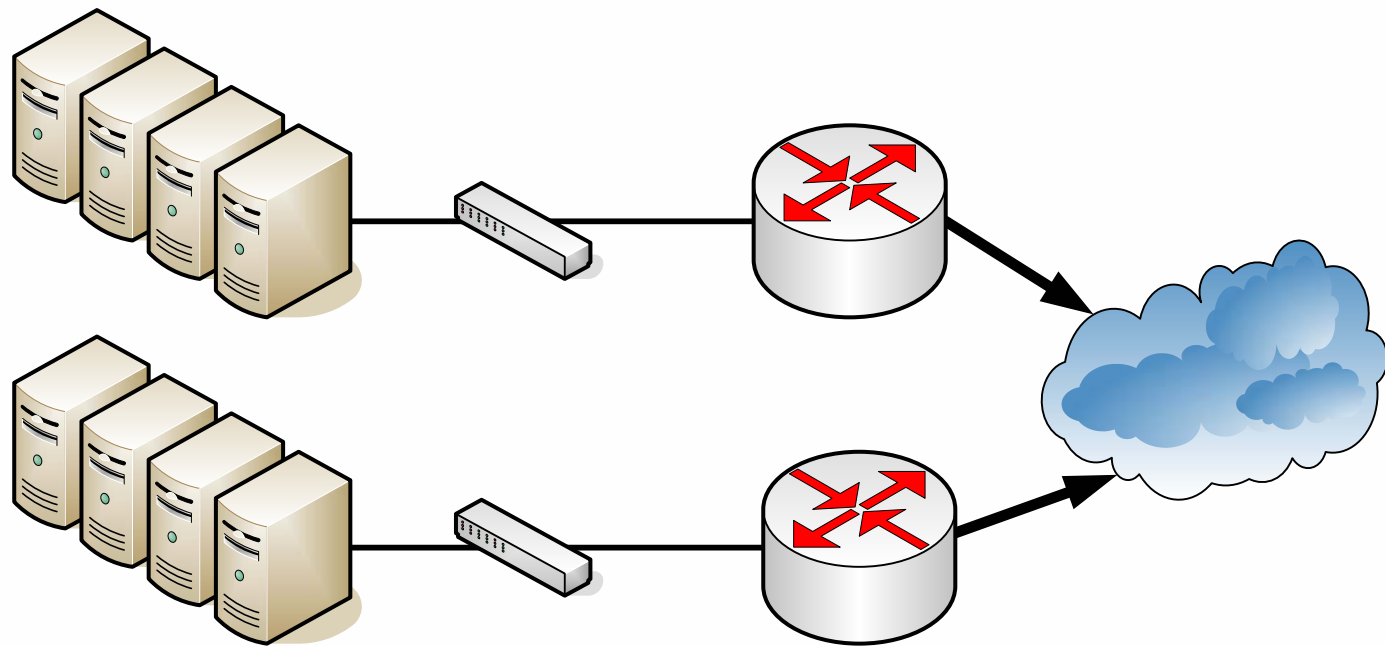
- Hierarchical data storage system
- Batch system
 - 1300 dual CPU nodes
 - Data producers & consumers
- Disk layer – intermediate
 - 750 TB in total
 - Distributed across 420 disk servers
- Tape layer – permanent
 - STK: 50 9940B, 20 9840 drives
 - IBM: 8 3592
- CASTOR used for data management

CASTORGRID – description

- Main WAN data transfer service
- Load balanced service (more later)
- Special high network throughput routing
 - Not to overload the firewall
- Runs both:
 - GridFTP
 - castor_gridftp_server-VD
 - per host certificates (more
 - gsiftp://castorgrid.cern.c
 - SRM
 - SRM-1.2.12-0
 - srm://castorsrm.cern.ch
- Recently extended 8 nodes (2 GB RAM)
- 2 x 1 Gbit/sec connectivity per 4 nodes
- Basic TCP/IP stack tuning
- Proper trouble tracking system
 - Wan-Data.Operations@cern.ch



CASTORGRID – topology



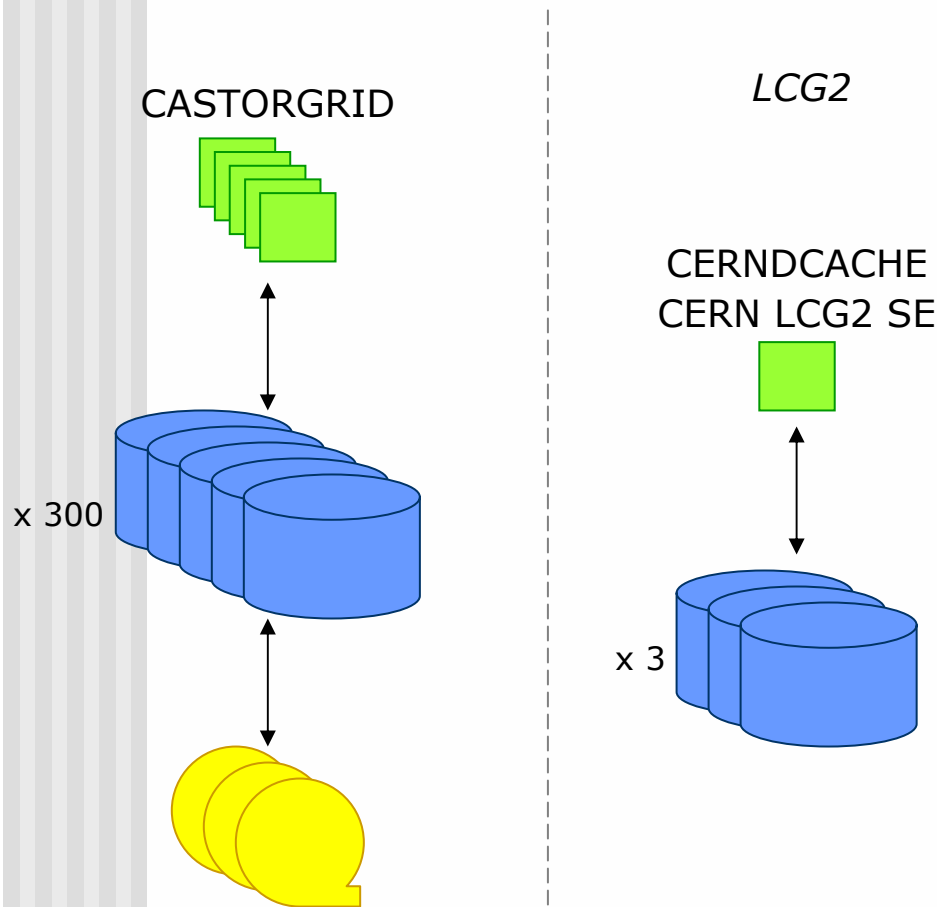
CASTORGRID – lesson 2

- Host vs. service LCG certificates
- globus-ftp-control
 - Prior to version 1.11
 - Clients did twice IP DNS lookup
 - 1st time once they received the certificate
 - 2nd time when initiating data transfer
 - 1st IP \neq 2nd IP \rightarrow service certificate
 - Version 1.11 onwards
 - Single IP lookup only \rightarrow host certificates in use

CASTORGRID – lesson 3

- User mapping
 - Users' identification is collected from virtual organizations' LDAP servers
 - All users from one experiment are mapped onto a group account
 - Confusion when accessing user's own directories
 - For exceptions, local map file is used
 - LCG certificate is not enough
 - Non LHC experiments / without virtual organization

Additional WAN data transfer services

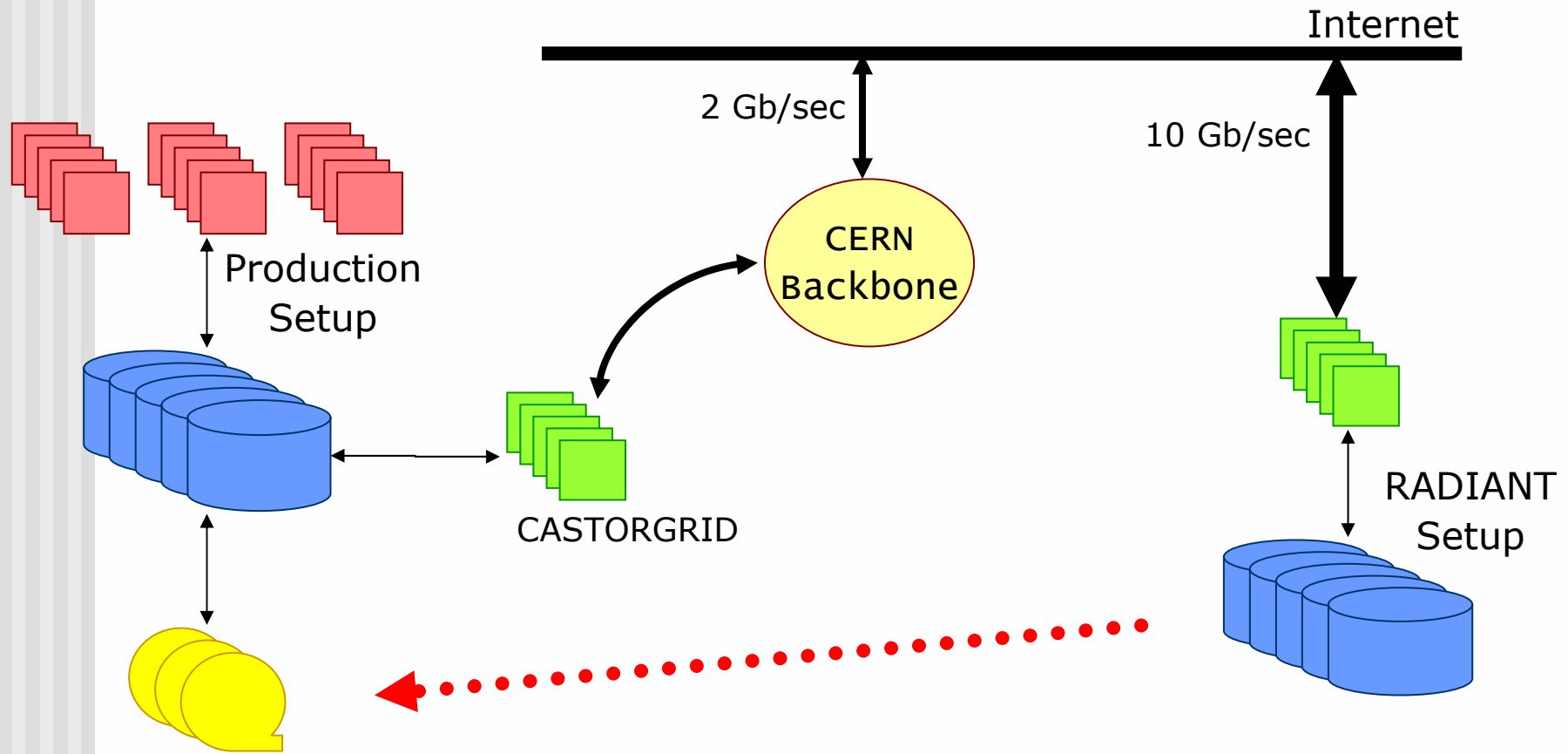


27 January 2005

RADIANT Meeting - Oxford

- CERN dCache storage element
 - lxn1180.cern.ch
 - Protocols : gsidcap, gsiftp
 - Small disk pool for experiments' small files
- CERN LCG2 storage element
 - lxn1183.cern.ch
 - Protocols : gsiftp, rfio
 - Development system

Network connectivity



27 January 2005

RADIANT Meeting - Oxford

10

Tapes involvement

- *July 2005: ... achieve 300 MB/sec from tapes ...*
- Fastest production ones
 - 30 MB/sec in streaming mode
- High in numbers
 - 50 STK 9940B
- Achieve 300 MB/sec
 - At least 10 drives needed
 - 20 % of total !
 - 2 STK Powderhorn installations
 - Drives would probably located in 1 of them
- IBM 3592s ?



IBM 3584



27 January 2005

RADIANT Meeting - Oxford

12

IBM 3584 involvement

- 8 IBM 3592
 - Detailed testing done
 - Up to 40 MB/sec
 - In streaming mode
 - $8 * 40 = 320$ MB/sec
- Room for up to 100 tapes cartridges

= All in all – more acceptable solution

Conclusion

- Network throughput capacity of production services is lacking behind
- Regarding non-network hardware
 - We are capable to do better
 - Awaiting CERN's internal network upgrade
- Tape drive resources scarce, but sufficient for now
 - If IBM 3592

Thank you

- Questions ?
- E-mail: Vladimir.Bahyl@cern.ch