CASTOR 2 experience

Vladimír Bahyl
CASTOR Operations

Vladimir.Bahyl@cern.ch

Outline

- SRM changes
- Upgrade guide for your GridFTP/SRM facility
- CASTOR2 status update
- CASTOR2 disk pool configuration for the Service phase
- Summary

SRM changes 1/2

- Since version 2.0.0, important features have been introduced and many bugs have been fixed
- 2.1.0
  - USE_DISK_SERVER_NAME_IN_TURL
    - Enable/disable the use of the disk server name in the TURLs
    - Usefull if you run GridFTP directly on the disk server
  - RESOLVE_CLUSTER_HOSTS
    - Use gethostbyname() to look up IP addresses of all hosts of a DNS alias and rotate returned TURLs through them
    - Usefull if your SRM server is in fact a DNS alias to a cluster of GridFTP doors

SRM changes 2/2

- 2.2.0
  - XML requests repository
    - Store requests in XML format to allow repository portability between IA32 and IA64 architectures
    - Old requests can still be read/written to ensure smooth migration
  - Added -nodcau option to GSIFTP CMD (globus-url-copy) in /etc/srm.conf
    - Necessary to ensure compatibility with dCache
- Both CASTOR 1 & 2 are supported

➔ All this is useful and should be deployed

Upgrade guide for your GridFTP/SRM facility

- Motivation
  - Benefit from all new features and bugfixes described earlier
  - Compatibility reasons
    - have same versions as at CERN
- Components needed for the upgrade:
  - SRM v2
  - GridFTP v2
  - CASTOR 2 client

Upgrade guide – RPMs

- SRM
  - SRM-2.2.2-1.slc3
- GridFTP
  - castor_v2_gridftp_server-VDT1.2.0rh9-1
- CASTOR 2 client
  - Packaging policy has changed
    - Before: single monolithic RPM
    - Now: each component packaged separately
  - CASTOR-client-2.0.0-59
  - + all 15 packages it requires
- This setup will work with CASTOR 1 storage system in background = no need to upgrade the rest to CASTOR 2
- Tested at CERN with SLC3 for IA32 and IA64
- Contact me for full list of RPMs

Upgrade guide – configuration files changes

- SRM
  - Options added to /etc/srm.conf; example is provided
  - Quattor managed nodes can benefit from NCM srm component
- GridFTP
  - Some config files are new; some have been modified/relocated:
    - /etc/castor/stagemap.conf
      - Mapping of group accounts to stagers
    - /etc/castor/stagetype.conf
      - Definition of stager versions
      - Default = CASTOR 1
  - Unmodified:
    - /opt/edg/etc/edg-mkgridmap.conf, /opt/edg/etc/grid-mapfile-local, /etc/grid-security/*, …
- CASTOR 2 client
  - New directory – /etc/castor:
    - /etc/castor/castor.conf → /etc/shift.conf
    - /etc/castor/castor.localhosts → /etc/shift.localhosts

CASTOR2 status

- Update since the last GDB

Issues solved 1/2

- LSF plug-in memory leak
  - Memory delivered by the LSF API has not been freed properly
  - ☺ No other memory leaks in the system observed
- DLF database space problem
  - We run out of free space in the data tablespace
    - Since this was only log data, we decided to drop it
    - All data were in logs as well
  - ☺ Partitioning by time is being considered
- Improved file open latency
  - LSF interface to resource monitoring master has been optimized
  - ☺ Gained around 1 second per file

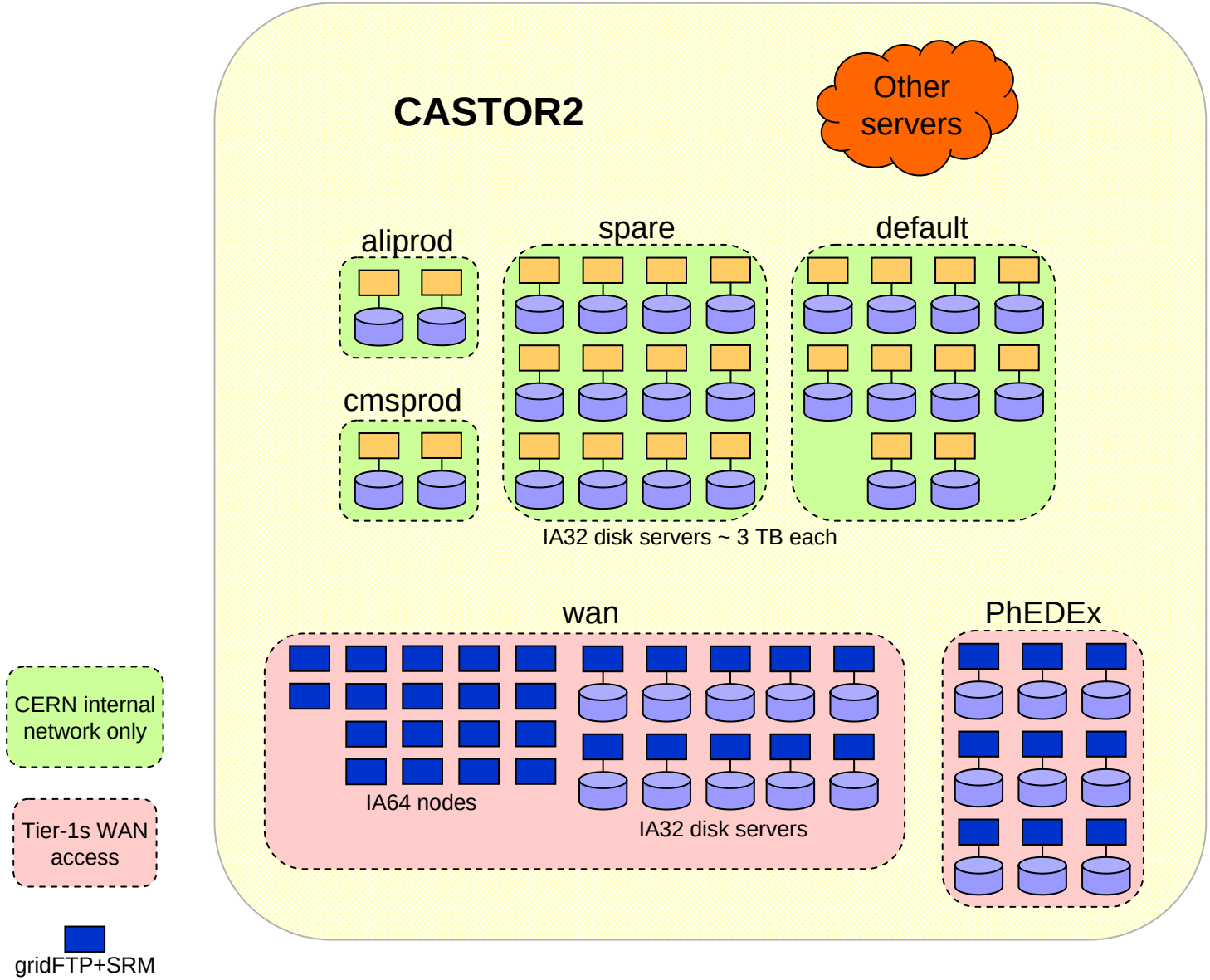
Issues solved 2/2

- ☺ Optimization of databases
  - Indexes on new columns were introduced and function based indexes for various queries have been created
    - This helped to lower the number of I/O operations
  - Partitioning by request type in the stager database is being considered
- 😐 What remains
  - Because gsiftp is not fully supported protocol in CASTOR2, requests are treated twice
    - SRM request for a file schedules an LSF job
    - Once the file is ready, gridFTP does rfio_open() which schedules another LSF job

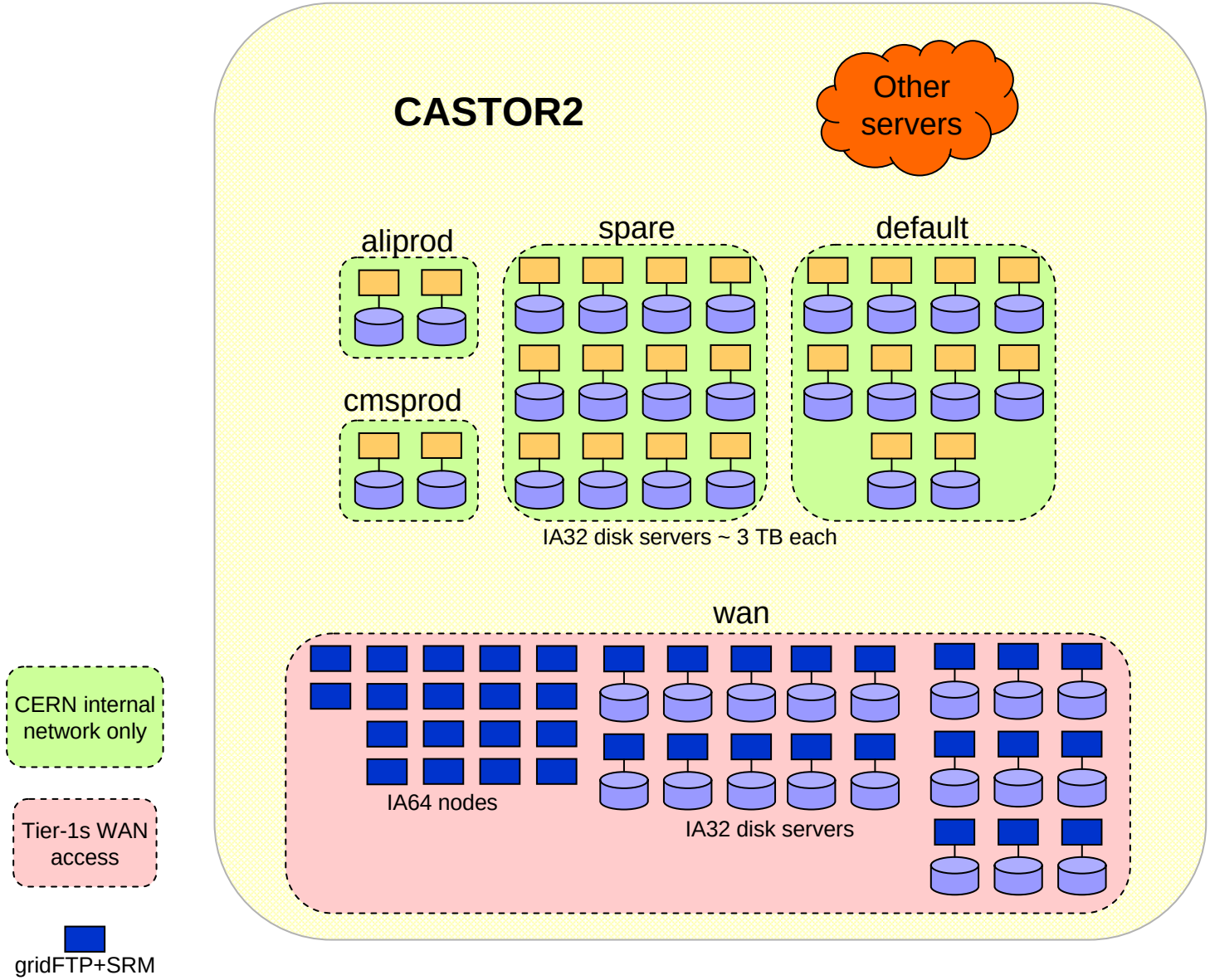
CASTOR disk pools

- Configuration for the Service phase

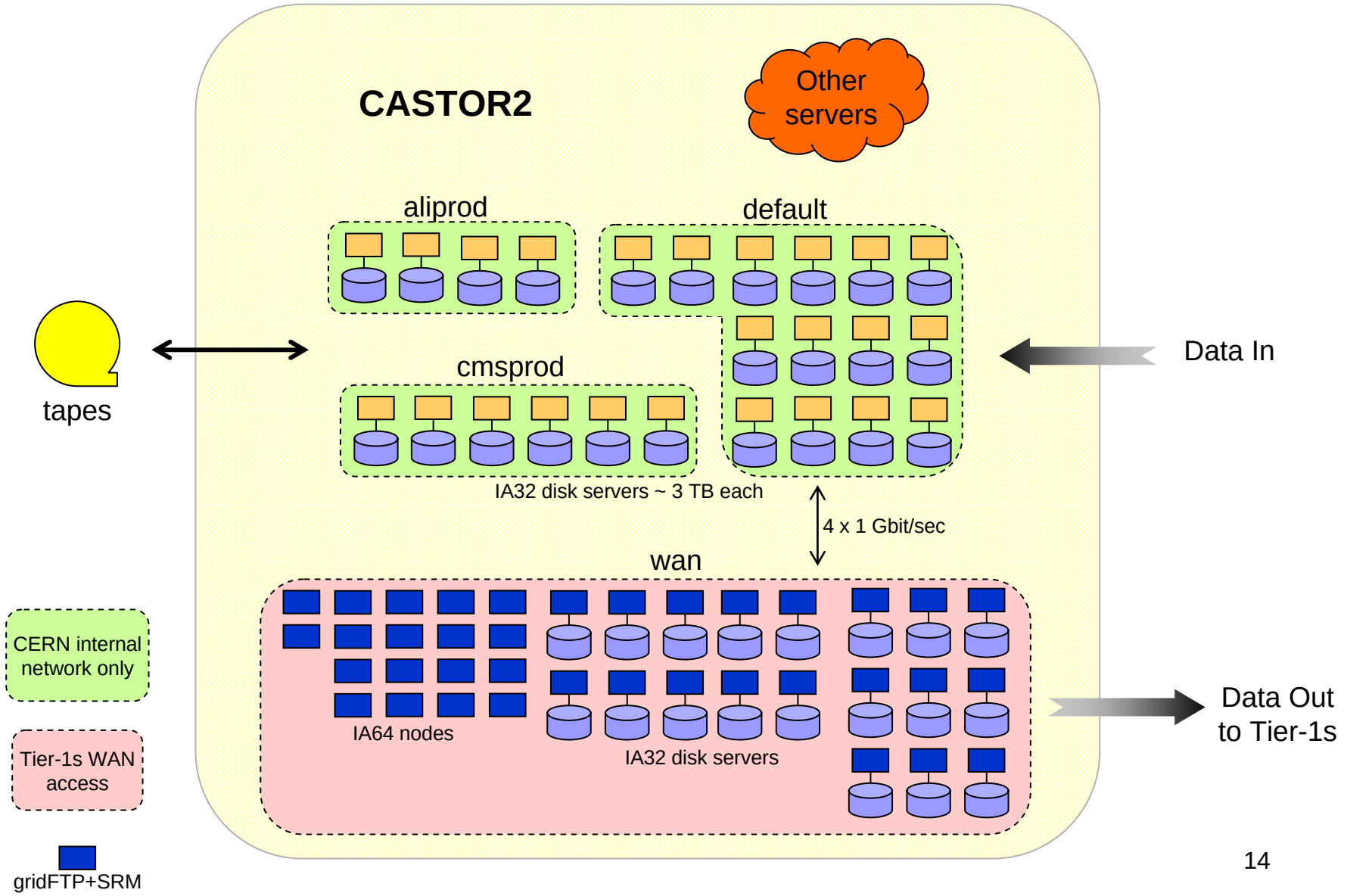
Previous state



Transition …


CASTOR2
Other servers
aliprod
spare
default
cmsprod
IA32 disk servers ~ 3 TB each
wan
IA64 nodes
IA32 disk servers
CERN internal network only
Tier-1s WAN access
gridFTP+SRM

Current state



Reasons behind the configuration

- Most realistic model
- Allows separation of data flows
  - Incoming data from the accelerator being written to tapes
  - Data being distributed to Tier-1 institutes
- ALICE
  - Plans to transfer files to CERN via the WAN pool
  - Will read and distribute data from WAN pool
  - If needed data can be replicated into the internal *aliprod* pool
- CMS
  - Will stage already existing data from tapes
  - Data will be stored in CERN internal *cmsprod* pool
  - Replicated to WAN pool and distributed to Tier-1s
- Other experiments will be added later

Summary

- Please upgrade to the latest version of SRM and GridFTP together with CASTOR 2 client to benefit from numerous features and bug fixes
  - Compatibility with CASTOR 1 assured

☺ Improvements to the CASTOR 2 system have been made even over the quieter summer period

☺ Flexibility of CASTOR 2 disk pool configuration allows modifications on short timescale if necessary

Thank you

- Vladimir.Bahyl@cern.ch
- http://cern.ch/vlado