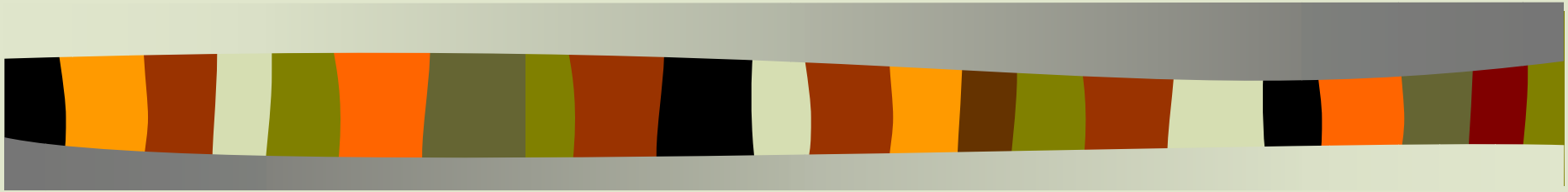


IN2P3 Site Report

John O'Neill, Philippe Gaillardon

Centre de Calcul de l'IN2P3

<http://doc.in2p3.fr/hpss/>



HPSS User Forum

4-6 May 2004, Santa Fe

Site report overview

- Who are we?
- Configuration and statistics
- Site-specific – interface, problems
- GRIDs
- HPSS 5.1
- Summary



Who are we?

- IN2P3 = IN²P³ = National Institute of Nuclear Physics and Particle Physics
- Institute of the CNRS (National Center for Scientific Research)
- 18 laboratories + 1 Computing Center (~50 people)
- Computing for *all* French particle physics, including for CEA



Hardware configuration

- HPSS 4.5
- Recently upgraded core server from an F50 to a p630
- 19 tape movers (AIX, Solaris) in “fast” STK silos
 - 28 Storagetek 9840s
 - 23 9940s
 - SCSI, FC
- 6 disk movers, 12 TB
- GigE data network

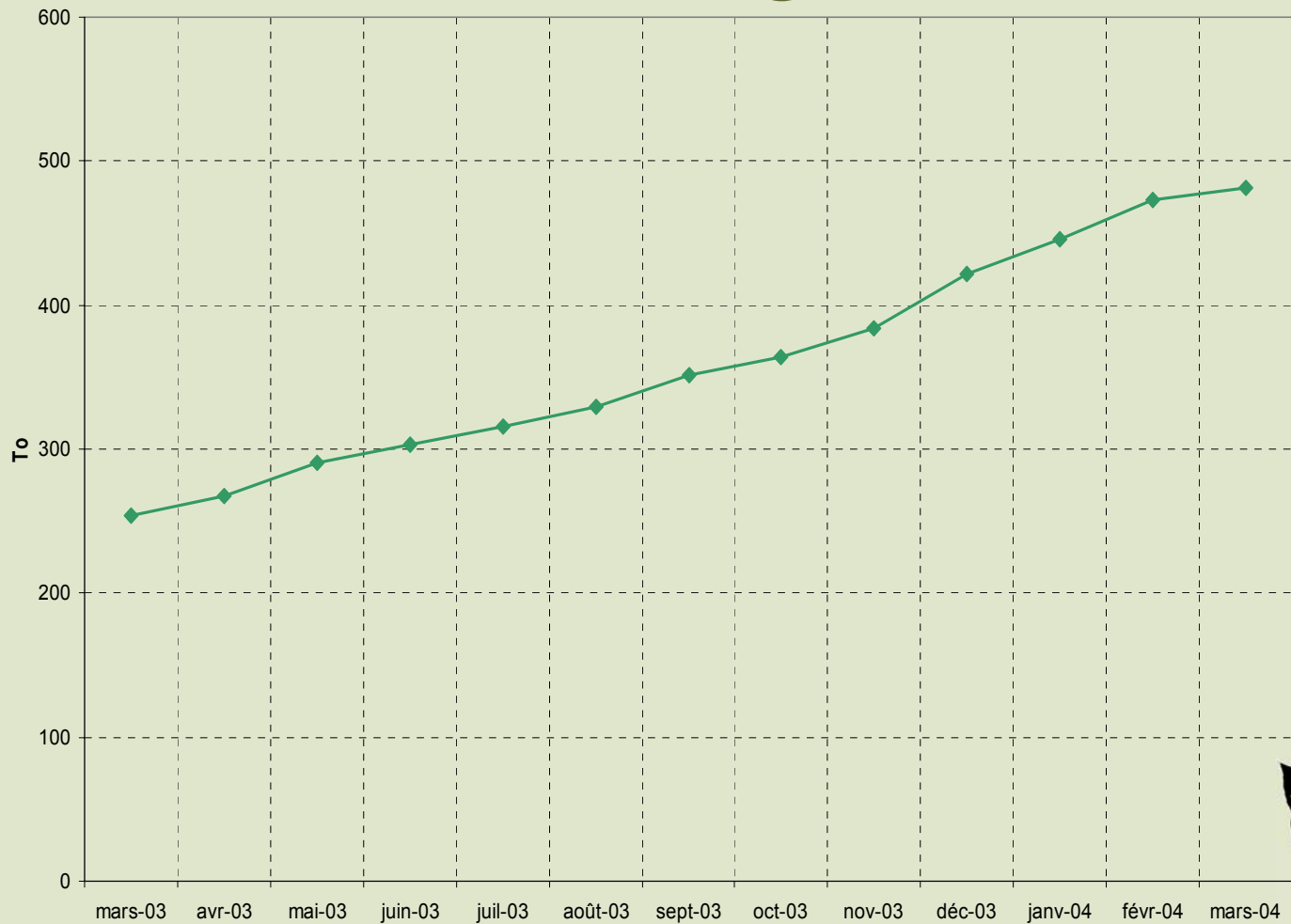


Statistics

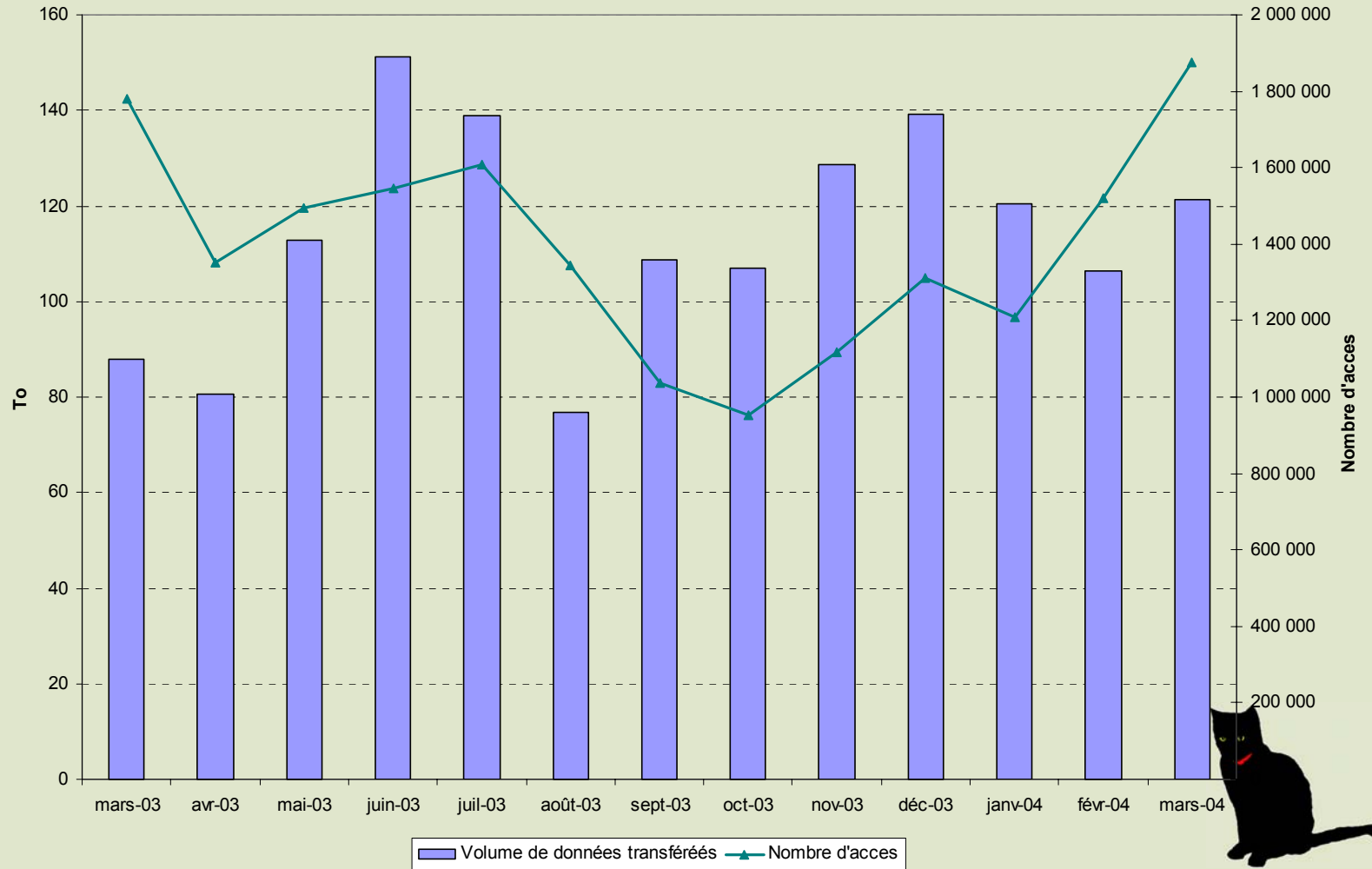
- Now have almost 500 TB stored for ~ 30 different experiments in particle and astro-particle physics
- 6.3×10^6 files, average <80 MB> per file
- Up to 7200 tape mounts/day
- Current usage statistics at <http://doc.in2p3.fr/hpss>



Total data storage over time



Accesses and data transfer



Tapes

- Tape exception handling could be improved: e.g., error during write → EOM, no log sense, some messages without vid, *etc.*
- Worse for non-AIX machines (error reporting, xmperv)
- No useful drive status window (with mounted cartridges)
- New tape technologies (20MB/s or more) require hi disk-to-tape thruput. Double GigE network, SANs?
- High drive demand → “thrashing” of tapes.



Control

- hpssadm
 - Provides longed-for bulk configuration
 - Initialization much faster with faster machine
 - Various needs: authorization domains, macro facility, more function, etc.
- Need dynamic configuration (disks, drives, COS, SCs, etc.) without stopping production
- SSM needs to be restarted regularly (memory leak)



Interface RFIO

- Developed at CERN in early '90s
- 64-bit, HPSS-knowledgeable (setcos, readlist/writelist)
- Available commands: rfcop, rfdir, rfstat, rfmkdir, rfrename, rfrm, rfcap
- Standard C API + readlist/writelist, setcos
- C++ streaming interface



RFIO (2)

- Daemon runs on each AIX disk mover and on 1 tape mover (tape-only COS)
- Permissions: correspondence between Unix perm bits and ACLs
- Mostly rfcop (copy) to/from local disks
- Many small files copied to NFS-mounted disk for applications; HPSS is permanent store
- Interfaced to Objectivity DB and, now, xrootd



bbftp

- “Babar” ftp (developed at IN2P3)
- Uses rfio API to HPSS
- Authentication by ssh, Grid certificate, other
- Creates its own parallel streams over line
- Can saturate any line
- Will eventually be replaced by tools from...



The G-word (Grid)

- Everything's comin' up GRIDs!
- 7 projects under way at CC-IN2P3
- Have GridFTP with RFIO I/f (API)
- Waiting for GridFTP interface for HPSS
- Using SRB (San Diego) with HPSS
 - Uses non-DCE HPSS API
 - Used by several experiments (Babar, LHC, astro)



More Grid -- SRM

- SRM needs local disk cache
 - Currently trying dCache (DESY, Fermilab)
 - dCache modified to use rfcpl I/f
- Interested in trying Berkeley HRM, but may be copyright problems (waiting)



HPSS 5.1 preparation

- Want to dump Sammi and, especially, Encina
- Installed and tested HPSS 5.1 and metadata conversion
- We like the new GUI
 - Really resizable windows
 - Fast, easy access to information
 - (How to change color?)
- We like DB2
 - Less space (adieu, LA files!)
 - Backs up DB2 directly to TSM (look out for exclusion file); much faster



Metadata conversion to 5.1

- Ran many tests of production metadata conversion
- Minor problems (vpath, pvlactivity, DB2 commands)
- Recommendation: start with a really clean SFS metadata base
- Finally got it down to 1.5 hours, 4 loads in parallel, sans checks (which take many hours)



Summary: wish list

- Better DB performance for small, frequently-accessed files; high hopes for DB2
- Kerberos, aka end of DCE
- GridFTP
- Dynamic configuration
- Per-file-family mount limit
- SAN and Fibre-channel for performance/recovery



More wishes

- Improved handling of tape and drive exceptions
- Scratch tapes
- Messages: more precise, less repetition, tracking problems
- Improved documentation (index, change bars)
- Btw: what happened to stand-alone facilities doc?



Some links

- HPSS at IN2P3 – statistics and control
 - <http://doc.in2p3.fr/hpss/>
- RFIO documentation (in French)
 - <http://doc.in2p3.fr/doc/public/products/rfio/rfio.html>
- Bbftp documentation
 - <http://doc.in2p3.fr/bbftp/>



Au revoir



<http://doc.in2p3.fr/hpss/>