

# HPSS at the European Centre for Medium-range Weather Forecasts

Francis Dequenne

francis.dequenne@ecmwf.int

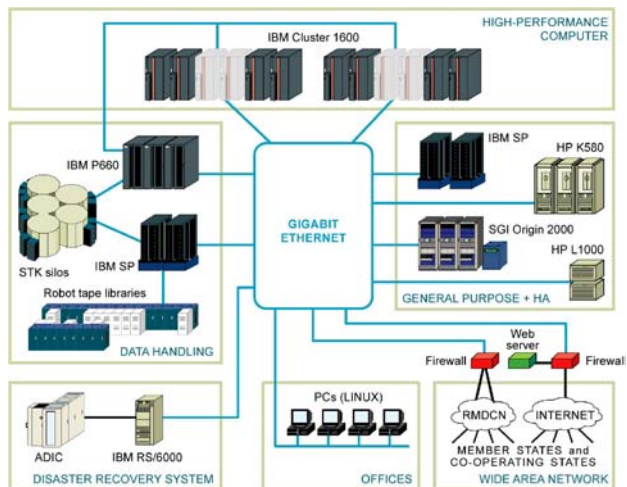
June 2003

HUF 2003



## Who are we?

- European based international scientific organisation, specialising in weather modelling through supercomputers.
- We predict the weather!

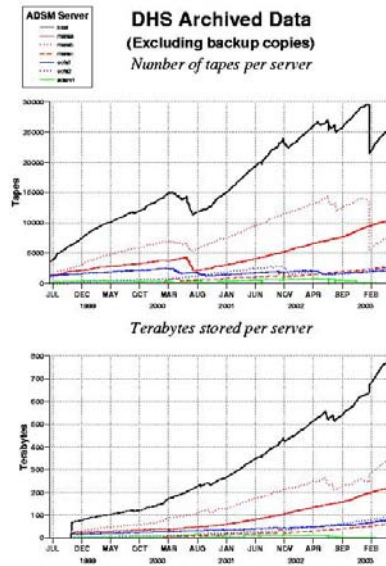


HUF 2003



## Basic facts:

- 800 TB of data stored.
- 1 TB saved daily.
- Several hundred GB retrieved daily.
- Two main applications:
  - MARS (storage and retrieval of organised meteorological objects)
  - ECFS (general purpose file archiving system).



HUF 2003

ECMWF 

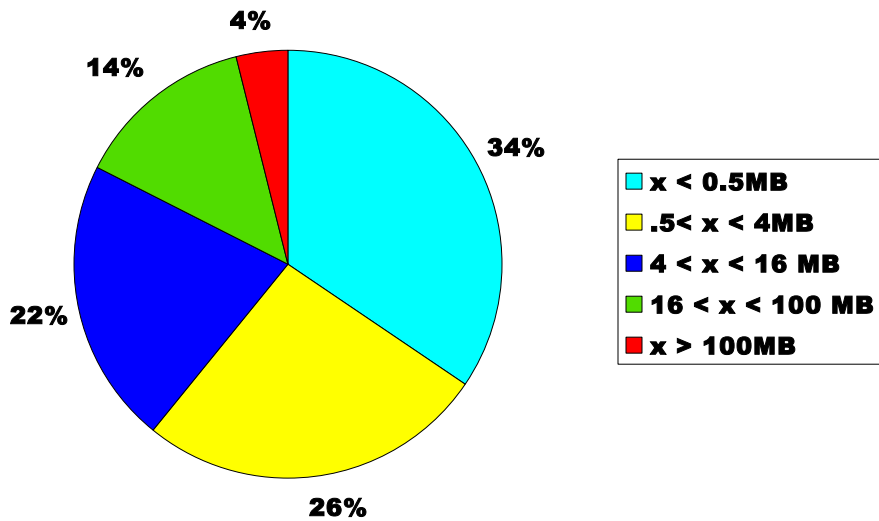
## Main application characteristics.

- MARS
  - Around 630 TB of data, few files.
  - Manages its own disks, HPSS only used for tape storage.
  - Requires very efficient management of retrievals of parts of files from tape (partial gets).
  - Access through the HPSS API.
- ECFS
  - Around 170 TB of data, 8.3 million files.
  - Will start using HPSS by the end of the year.
  - Clients will use pftp to read and write data in HPSS disk-tapes hierarchies.

HUF 2003

ECMWF 

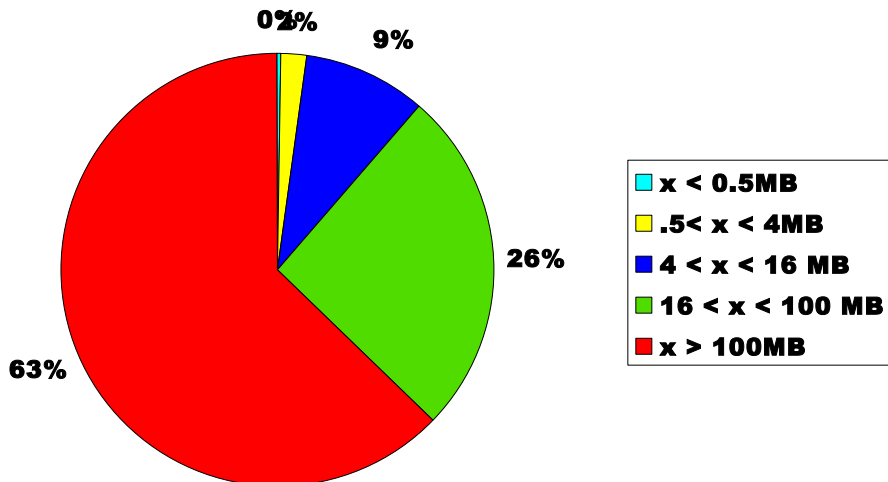
## Files in ECFS: 8.3 Million files



HUF 2003

ECMWF 

## ECFS files: 170 TB of data

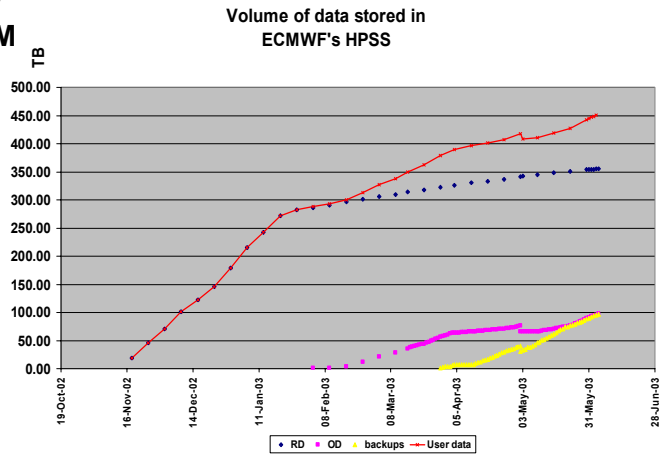


HUF 2003

ECMWF 

## HPSS: Our story so far.

- HPSS is progressively taking over from a TSM based data handling system.
- Tests started in 1Q02. V4.5 was installed in the summer 2002.
- System was put in production by end of October 2002.
- So far, we have transferred 2/3 of the MARS data to HPSS.



HUF 2003



## HPSS: First experiences.

- Pretty good!
- Applications ported without major problems.
- Performance more than adequate.
- Excellent support.

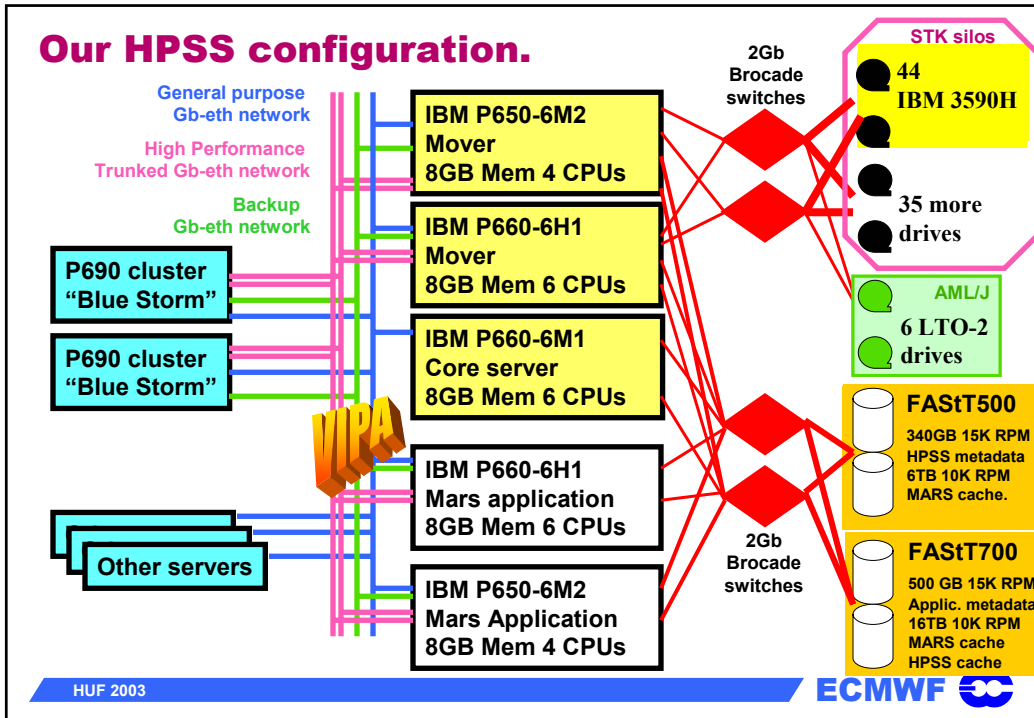


### • The WMDs that we found

- Old medias: Re-use of old SL tapes on 3590H drives resulted in some loss of data.
- HPSSADM: Need to restart SSM once/day.
- DCE issues. Fixed by going to ptf 3?
- Mover crashes.

HUF 2003





## Particularities of ECMWF set-up

- Abundant use of tape-only hierarchies.
- Hundreds of families.
- Intensive use of the API.
- Use of three subsystems.
- Partial file retrieval from tape.
- IBM Magstar drives in STK silos.
- LTO drives in AML/J libraries.
- VIPA.
- 24x7 support cover.

HUF 2003

ECMWF



## Next steps.

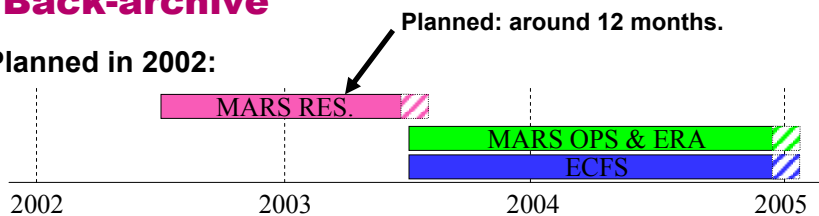
- Conclude MARS back-archive (More than 200 TB left)
- HPSS 5.1
  - Upgrade before we store millions of ECFS files in HPSS.
  - Support for new devices.
  - Go in production in early 4Q2003.
  - Test of HPSS 5.1 will start in the summer.
- ECFS installation.
  - Millions of files to be back-archived in HPSS.
  - Usage of large disk caches in HPSS.
  - Issues related to the management of small files.
  - Due 4Q2003. By then: around 220 TB to back-archive.

HUF 2003

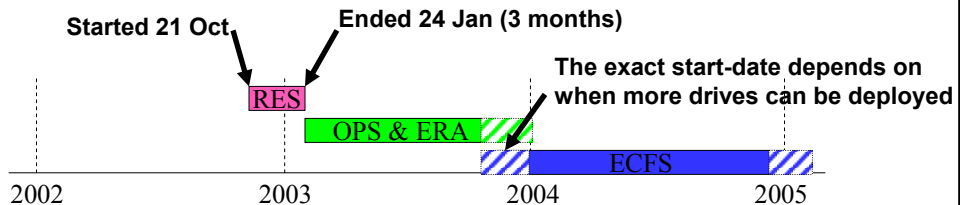


## Back-archive

- Planned in 2002:



- Current Plan:



HUF 2003



## What we would like to see...

- Handling of small files (esp. migration and repacking)
  - E.g by embedding very small files as fields in a DB2 table.
  - Limit the number of “small” files stored on a single media.
- Full support for HTAR and HSI.
- Authentication:
  - Hooks allowing use of site-written authentication-routines!
- Change/add drives/devices without stopping the PVL/PVR.
- Monitoring of system performance:
  - A single command-line query to obtain status of all drives, mounted volumes, recent I/O performances. Preferably through SQL instead of hpssadm.

HUF 2003



## Tools: Query\_hpss\_status

- Get status of all drives, and I/O rates recently observed.
- Problem: will hang the ssm daemons after a while, due to a memory leak.

HPSS Tape Drives report										Wed Jun 04 12:09:39 GMT+00:00 2003	
Drive Aix	Read	Write	Mover (Device)			Drive	Robot				
Id	device name	Volume	MB/s	MB/s	Admin	St Oper.	St	State	State	Status	Volume
2100	/dev/rmt2100	F0972900	0.000	0.000	Unlocked	Enabled	Enabled	Enabled	online	in use	F09729
2101	/dev/rmt2101	F1109100	0.000	0.000	Unlocked	Enabled	Enabled	Enabled	online	in use	F11091
2102	/dev/rmt2102		0.000	0.000	Unlocked	Enabled	Disabled				
2103	/dev/rmt2103		0.000	0.000	Unlocked	Enabled	Disabled				
2200	/dev/rmt2200	F1115300	0.000	0.000	Unlocked	Enabled	Enabled	Enabled	online	in use	F11153
2201	/dev/rmt2201	F1503100	0.000	10.917	Unlocked	Enabled	Enabled	Enabled	online	in use	F15031
2202	/dev/rmt2202	F1020900	0.000	0.000	Unlocked	Enabled	Enabled	Enabled	online	in use	F10209
2203	/dev/rmt2203	F1086300	17.917	0.000	Unlocked	Enabled	Enabled	Enabled	online	in use	F10863
3001	/dev/rmt3001	M0005500	0.000	17.401	Unlocked	Enabled	Enabled	UP			M00055
3002	/dev/rmt3002	M0076300	0.000	2.339	Unlocked	Enabled	Enabled	UP			M00763

HUF 2003



## Tools: Files\_in\_use

- Provides a list of all files currently opened, by whom, from which machine.
- Makes use of the gate-keeper site library.

Bitfiles currently in use.

=====

Monitoring started on 20030604:115240

Number of files opened: 8

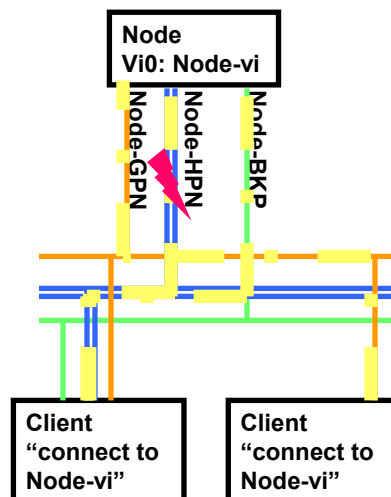
File Name	UID	Host	Time Opened
./marse4mnth/1/fc/19730100/sfc/37295.20030524.123706	marser	hdrv06	20030604:115312
./marse4mnth/1/an/19570900/sfc/37052.20030522.005317	marser	hdrv06	20030604:115242
/marsrdenfo/ec74/pf/20030213/sfc/645367.20030604.115444.tmp	marsrd	hdrg04	20030604:115445
marsrdseas/ed9t/fc/20011001/pt/1/644355.20030604.115455.tmp	marsrd	hdrg04	20030604:115456
./marse4wave/1/an/19640201/sfc/37415.20030522.212918	marser	hdrv06	20030604:115248
./marsrdenfo/edgh/cf/19530127/pt/644341.20030604.115447.tmp	marsrd	hdrg04	20030604:115453
./marse4oper/1/an/19641001/pl/33455.20030510.082353	marser	hdrv06	20030604:115359
./marsodenfo/12/cf/20020414/sfc/124099.20030604.115316.tmp	marsod	athos5-ge	20030604:115320

HUF 2003



## VIPA

- Feature of AIX V5.
- The system shows to the world a virtual address.
- Connections link to that virtual address instead of to an interface name.
- In conjunction with gated, allows automatic selection of best route.
- Allow automatic interface takeover without loss of connections.



HUF 2003

