

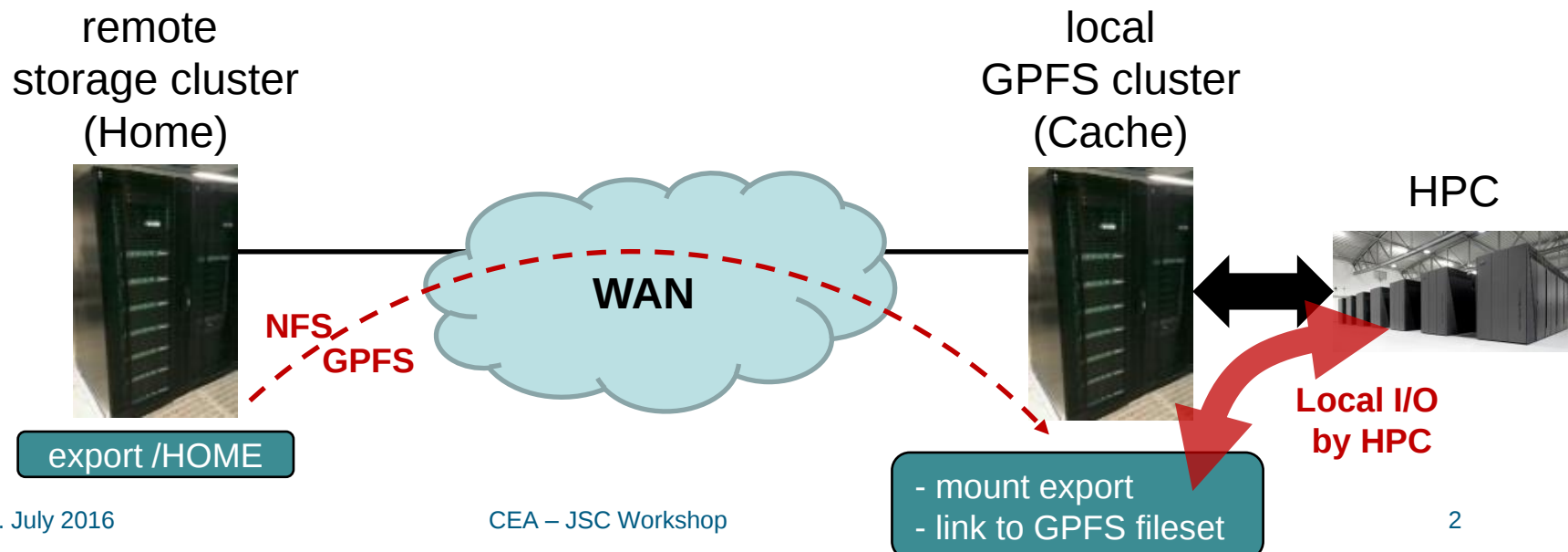


AFM Testbed First Experience

Stephan Graf (JSC)

What is AFM?

- AFM = Active File Management
- File system caching, part of *GPFS* software
- Purpose:
 Cache remote data (asynchrony) to local storage cluster to make it (high-performance) accessible by local systems
 → minimize latency for remote data access
 → data accessible though connection broken



Some AFM Details

- Several caching modes available

Mode	Purpose
Local-update (LU)	Read-only, local update allowed, no resync
Read-only (RO)	Read-only, no create or modify
Single-writer (SW)	One side is allowed to modify, all other read-only
Independent-writer (IW)	Multiple sides modify data, conflicts possible

- Asynchron Operations:
 - write, chmod, chown, create, mkdir, remove, rmdir, rename, link, symlink, and attribute updates.
- Synchron Operations:
 - Metadata queries (e.g. **ls**)
 - Read (direct access to first blocks, staged in completely in background)

Setting up HOME (GPFS based)

- Create local GPFS file system
- Enable AFM on this file system
`mmapfmconfig enable /gpfs/jsc_export/`
- Remote cluster mount
 - Setup remote cluster authentication
 - Grant access to remote cluster for the local GPFS file system

→ HOME is ready

Setting up AFM Cache

- Setup AFM Gateway node
 - GPFS server node which will mount remote file system
`mmchnode -gateway -N juniors1w1`
 - Remote cluster mount
 - *Get access to exported file system from HOME cluster*
 - *Mount file system on gateway node*
- Setup the AFM cache
 - Create local GPFS file system (here: “afm_caches”)
 - Create independent fileset in this file system
`mmcrfileset afm_caches bsc_export -p afmTarget=gpfs:///gpfs/bsc_export
--inode-space=new -p afmMode=read-only`
 - Link fileset into file system
- Mount afm_caches on all nodes → AFM_Cache is active

AFM Operator Command **mmafmctl**

- View cache state: **mmafmctl getstate**

active	Cache is active (NFS mounted)
inactive	Cache not connected to HOME yet (unused?)
disconnected	The Cache can not connect to the NFS HOME
unmounted	The NFS HOME is not mounted on gateway
Expired	Exceed disconnect timeout
Dirty	HOME data are not up-to-date
...	...

- Prefetch data:
mmafmctl <device> prefetch -j <FilesetName> ...
- Evict files
- ...

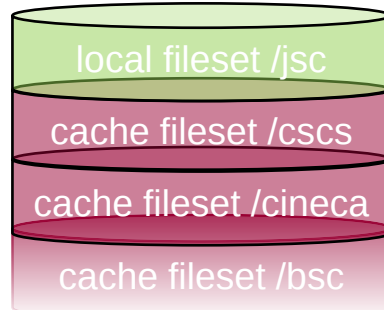
AFM & FeDaPP

- Proposal to use AFM for “active data repository”
- Testbed setup: (For each side)
 - GPFS cluster (separate from production?)
 - Get connected to the PRACE network
 - create export file system to all other sides (RO)
 - Mount remote file systems and create AFM caches
- Status:
 - JSC, BSC: OK
 - CSCS: test cluster connected to PRACE network
GPFS version mismatch
 - CINECA: use GPFS 3.5 + NFS → **Crashed JSC gateway**
- Ongoing discussion

FeDaPP: one global namespace



GPFS file system



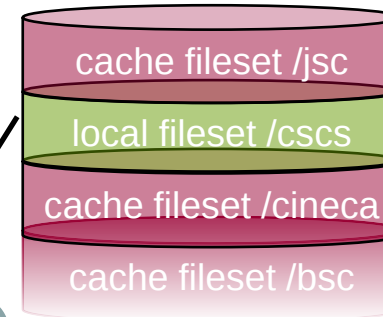
/fedapp/jsc
/fedapp/cscs
/fedapp/cineca
/fedapp/bsc
...



CSCS

Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

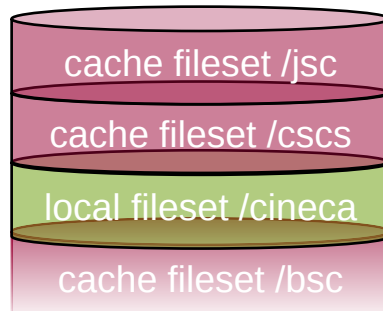
GPFS file system



/fedapp/jsc
/fedapp/cscs
/fedapp/cineca
/fedapp/bsc
...



GPFS file system



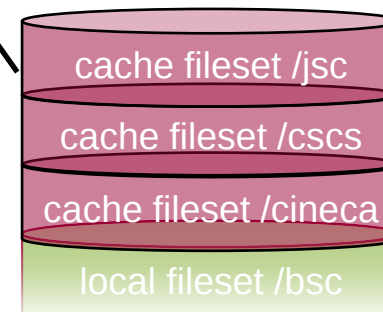
/fedapp/jsc
/fedapp/cscs
/fedapp/cineca
/fedapp/bsc
...



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

GPFS file system



/fedapp/jsc
/fedapp/cscs
/fedapp/cineca
/fedapp/bsc
...

First Results

- Good decision not to use production system
 - First tests with CSCS → JSC GPFS crashed
Reason: HOME was not AFM enabled
 - First tests with CINECA → JSC gateway node hanged
Reason: PMR in progress
- Bandwidth performance as expected (wire speed)
- First access with delays (staging in, expected), but then the data are cached (local file system access properties)