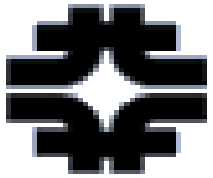
dCache.ORG

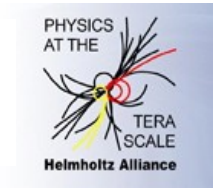
dCache.ORG

dCache, gridKa TABreport.

Patrick Fuhrmann et al.



additional funding, support or contributions by



d-grid
DGI II

Roadmap

dCache.ORG

Chimera

- Motivation
- Migration procedure
- NDGF experience

ACL's in dCache

Known issues

Plans : NFS 4.1

Preparation for analysis

Support

dCache.ORG

Chimera motivation

Slides stolen from

Tigran Mkrtchyan

What is wrong with PNFS?

dCache.ORG

Nothing is really wrong ... **however** :

- Only a single way of accessing pnfs : through the NFSv2 stack.
  - Has to be used by dCache plus all mounted clients.
  - Negative side effect : file size limit of 2GB
- Serialized access to individual database(s) through a global DB lock
- Multiple DB transactions per high level operation.
- All metadata stored as BLOB (no SQL queries)
- Internal structure is platform dependent
  - DB can't be moved to an other OS/Platform
- dCache designed to work with PNFS
- Design is based on late 90's technology

Slides stolen from
Tigran Mkrtchyan

Chimera pros (Design)

dCache.ORG

- Not a daemon – it's an API (and a Library)
- All 'clients' may work in parallel
- Single DB transaction per high level operation
  - Relies on DB transactional model (READ COMMITTED)
- File system view independent of metadata
  - Same objects may be represented by a different tree topology.
  - e.g. : would allow spaces to be represented as file system tree.
- Designed to benefit from the underlaying DB technology.
  - Easy to query.
  - Allows consistency check.
  - Some operations are delegated to *Stored Procedures* and *Triggers*.
- Designed to work with dCache

Slides stolen from
Tigran Mkrtchyan

Chimera pros (For you)

dCache.ORG

- Speed
  - Improves with good data base implementation (Oracle, postgres)
- Scalability
  - Speed improves with more cores, threads.(See next slide)
- Functionality
  - Professional backup (Depends on DB)
  - SQL queries (Examples later)
  - Vendor/platform independent.
- Maintenance
  - Support for PNFS will sooner or later be reduced and discontinued.

Slides stolen from
Tigran Mkrtchyan

dCache.ORG

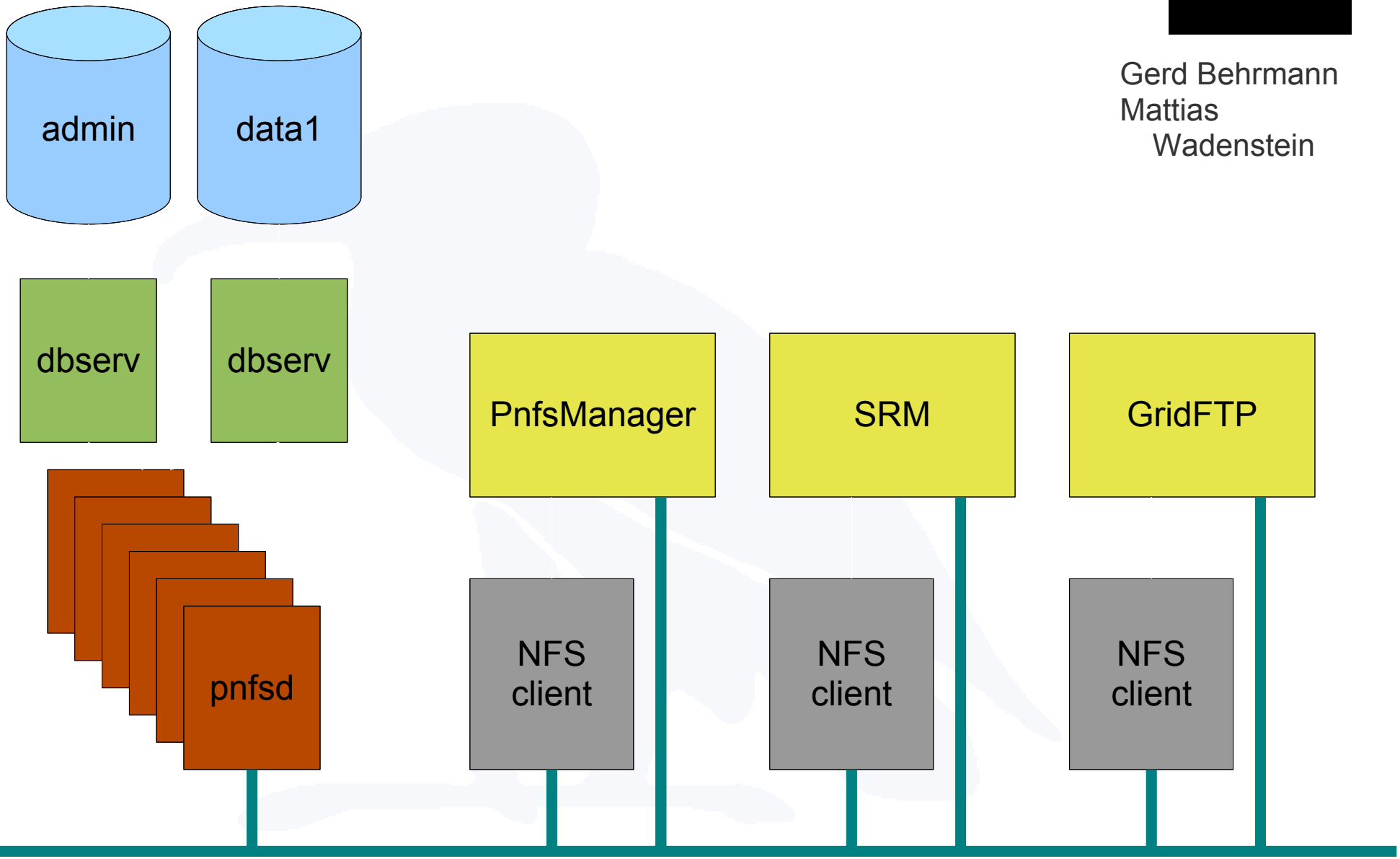
dCache.ORG

Pnfs to Chimera migration

Slides by
Gerd Behrmann
Mattias Wadenstein

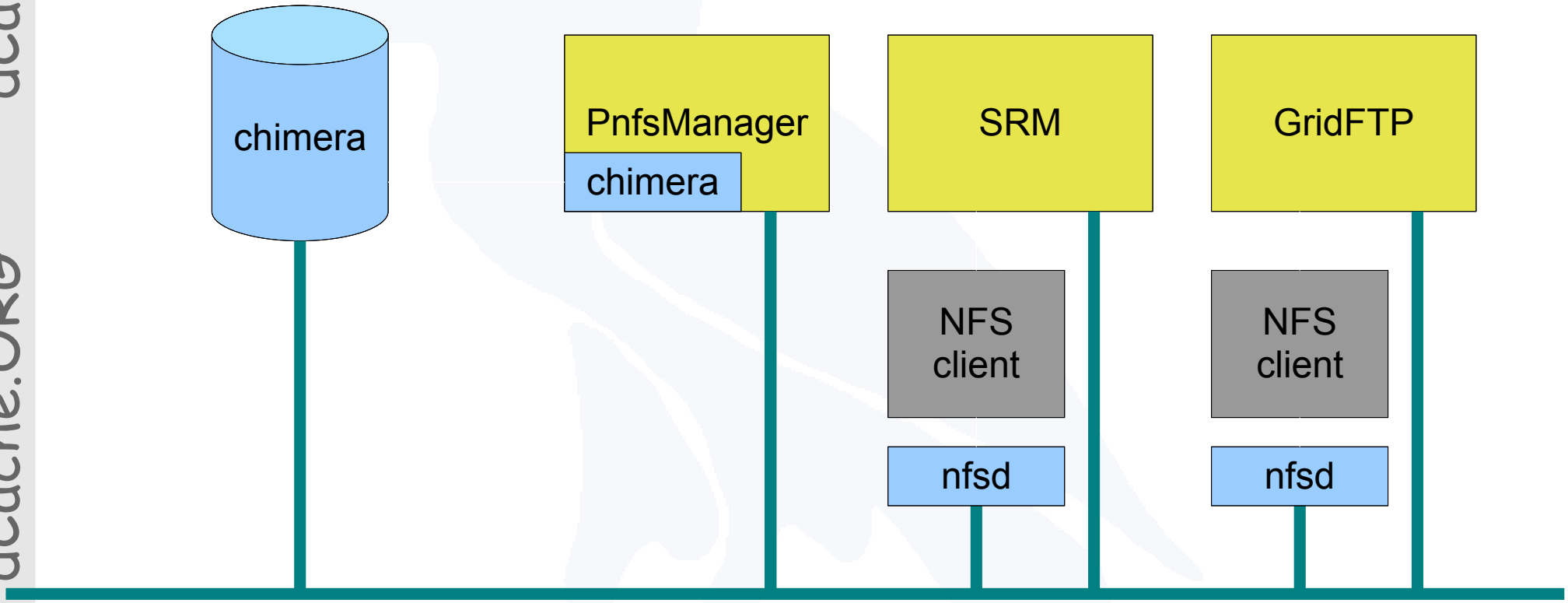
dCache – pnfs interactions

Gerd Behrmann
Mattias Wadenstein



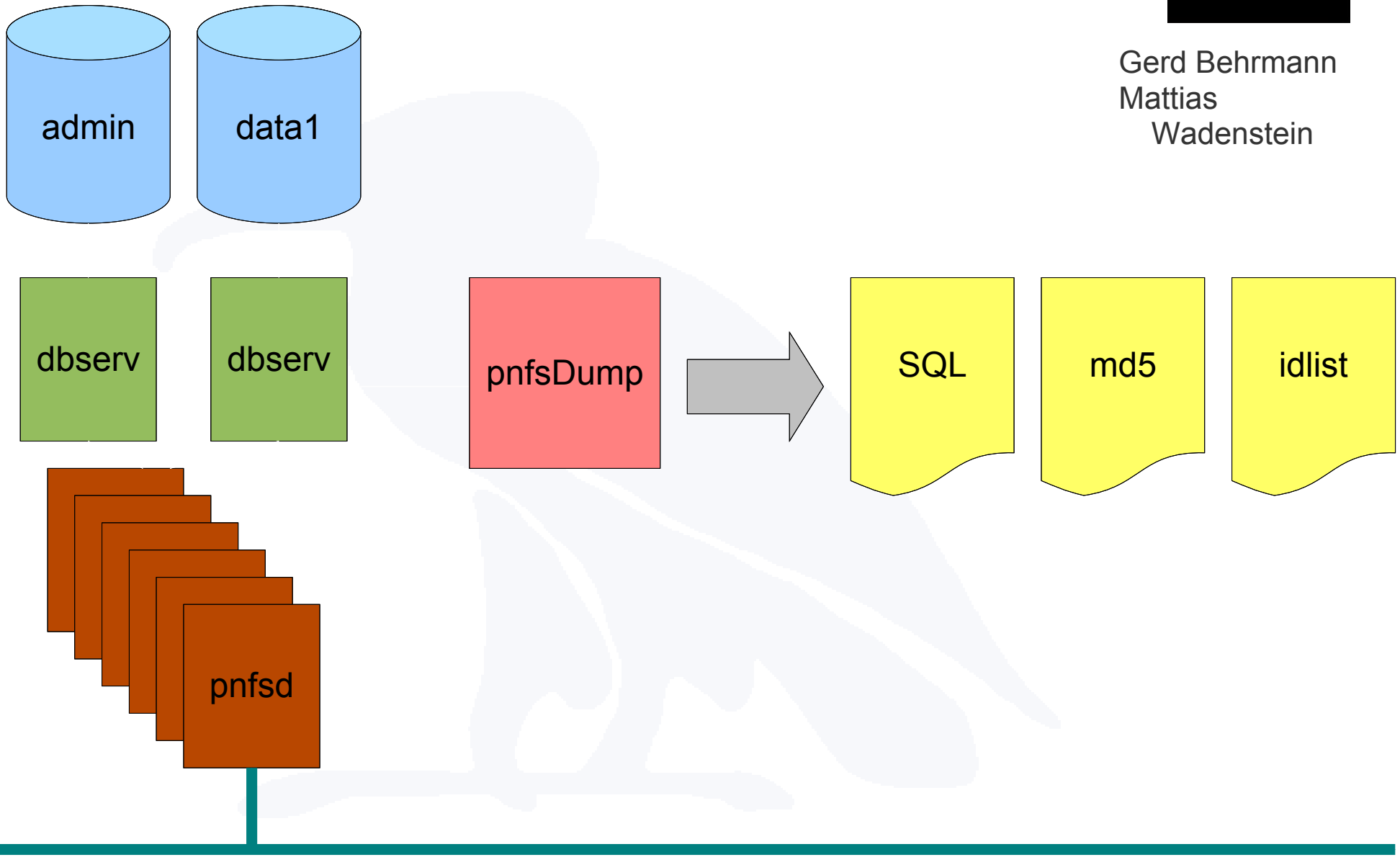
dCache – Chimera interactions

Gerd Behrmann
Mattias Wadenstein



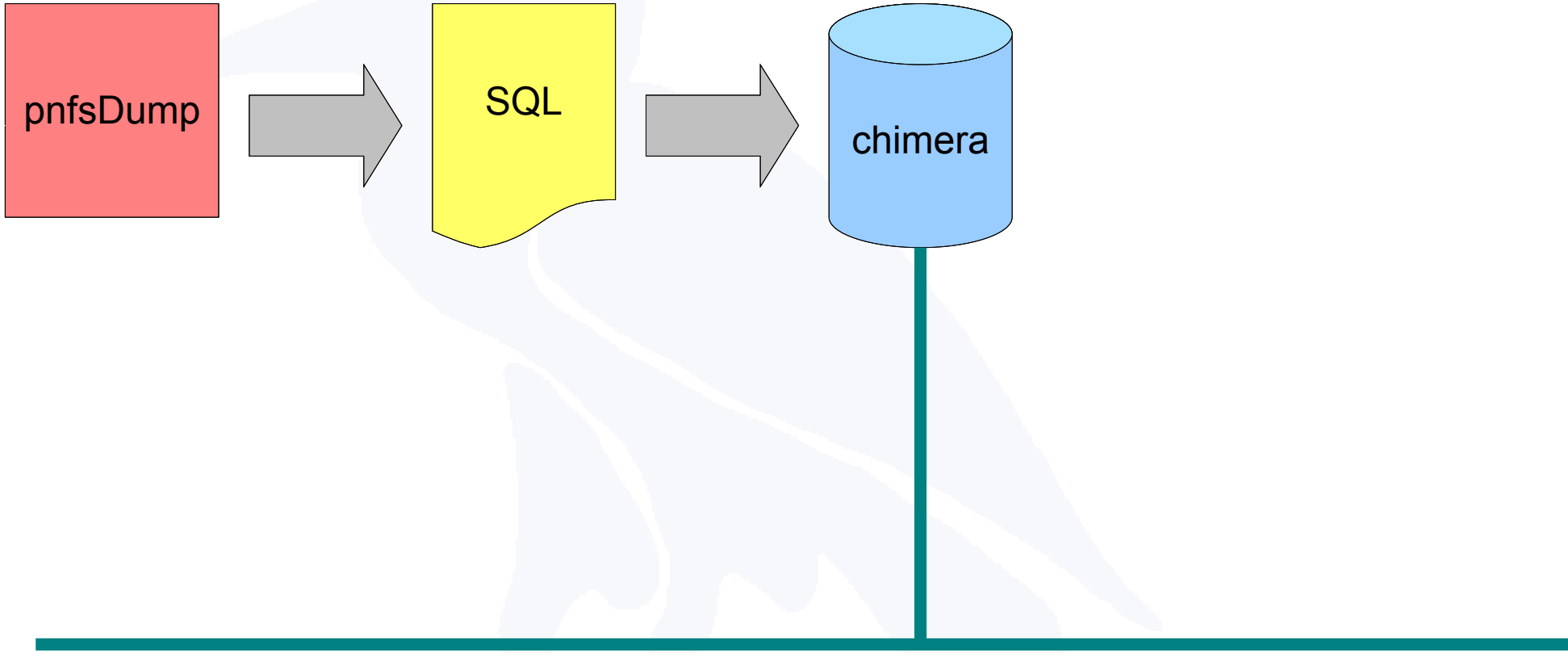
Dumping the Pnfs content.

Gerd Behrmann
Mattias Wadenstein



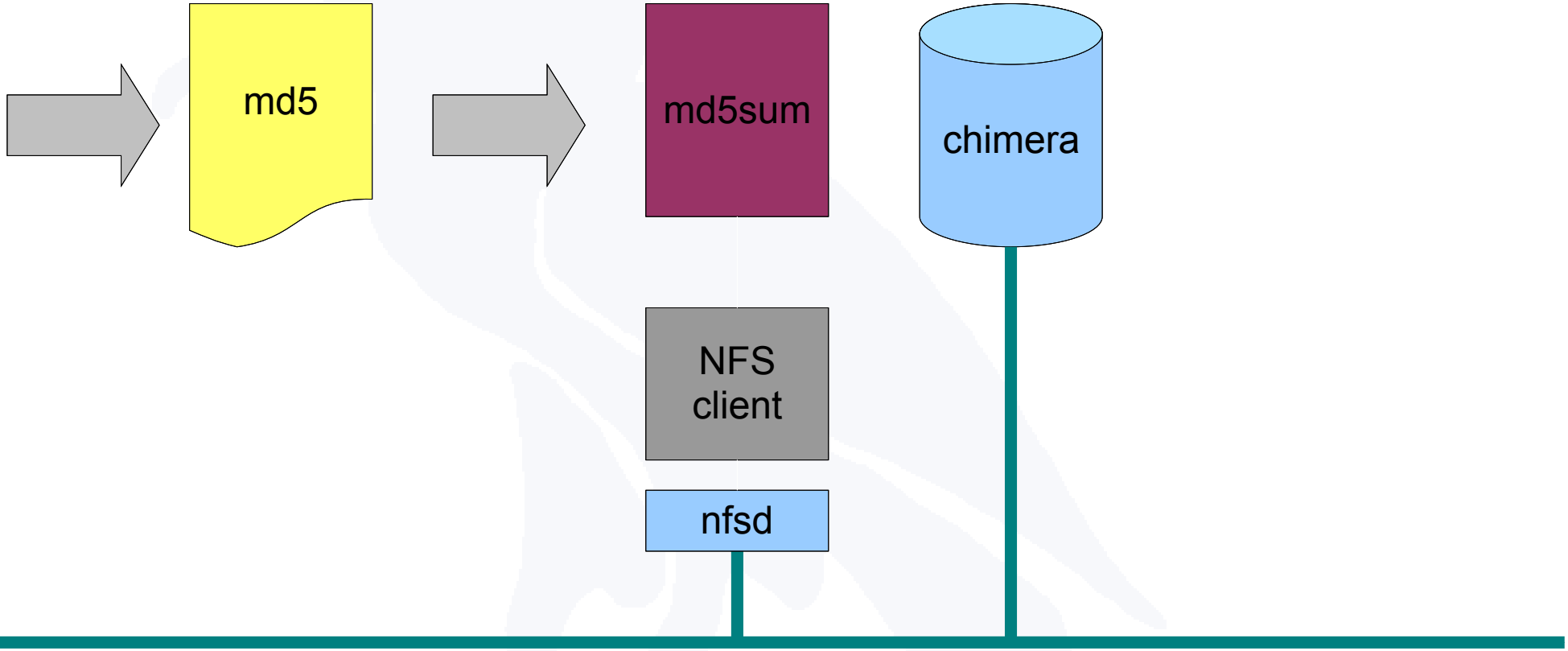
Filling the Chimera DB

Gerd Behrmann
Mattias Wadenstein



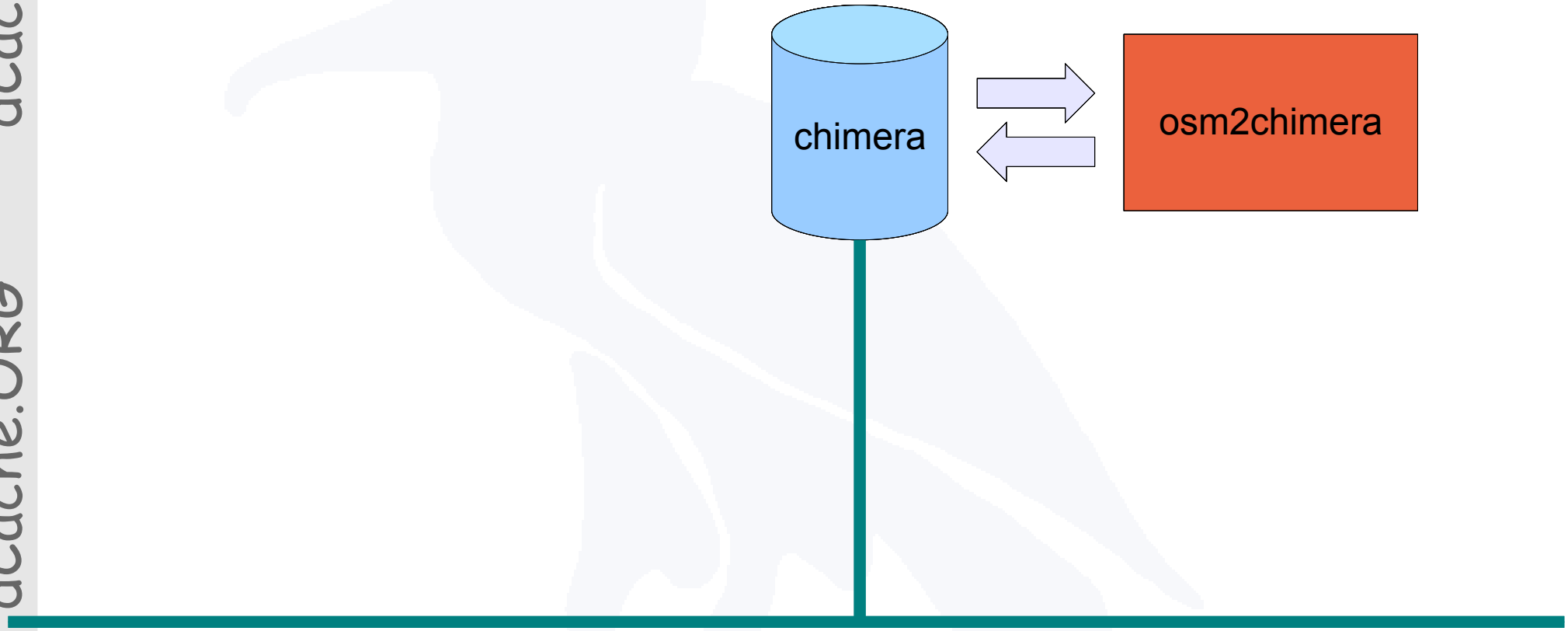
Checking Chimera content (md5)

Gerd Behrmann
Mattias Wadenstein



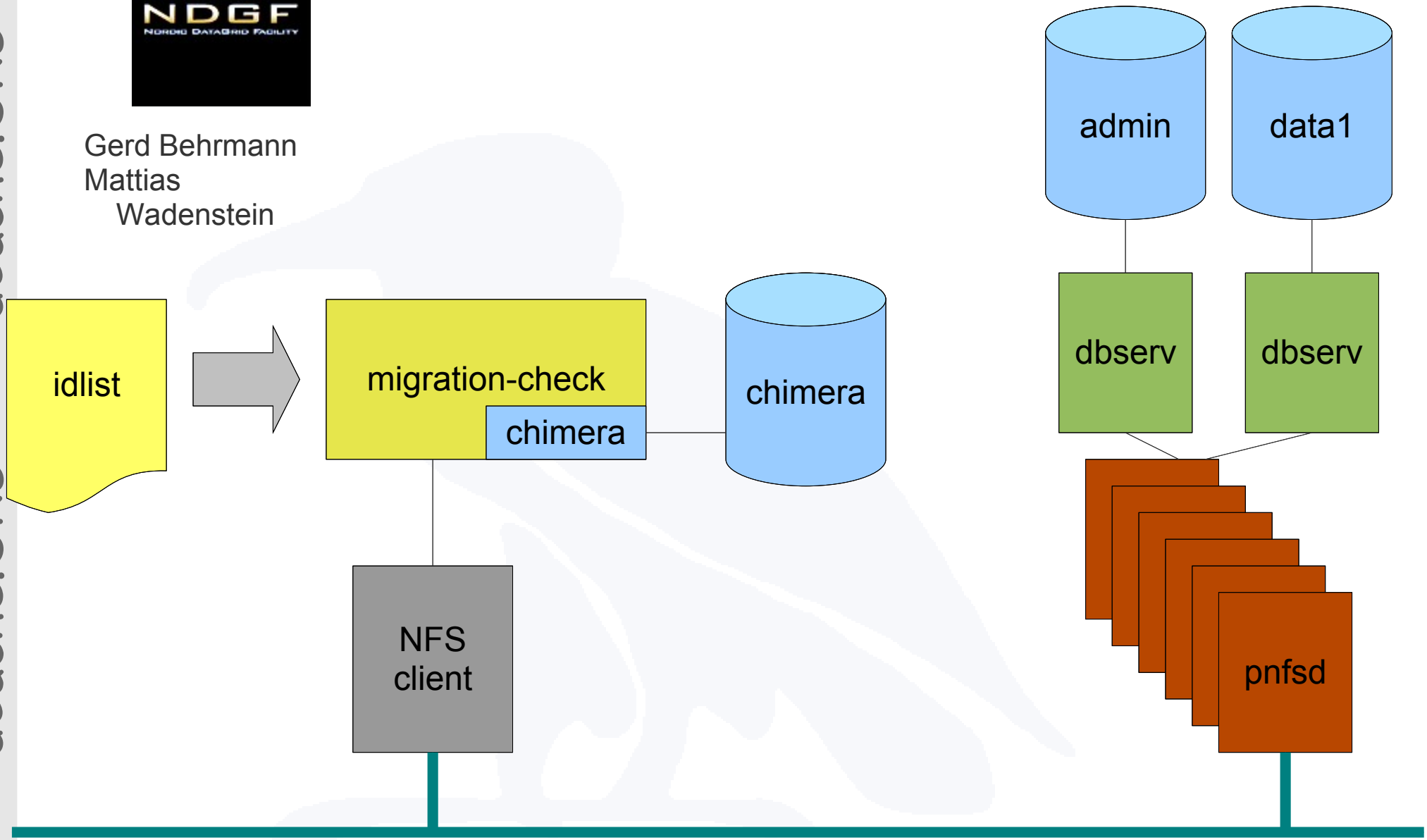
Inserting HSM info into Chimera

Gerd Behrmann
Mattias Wadenstein

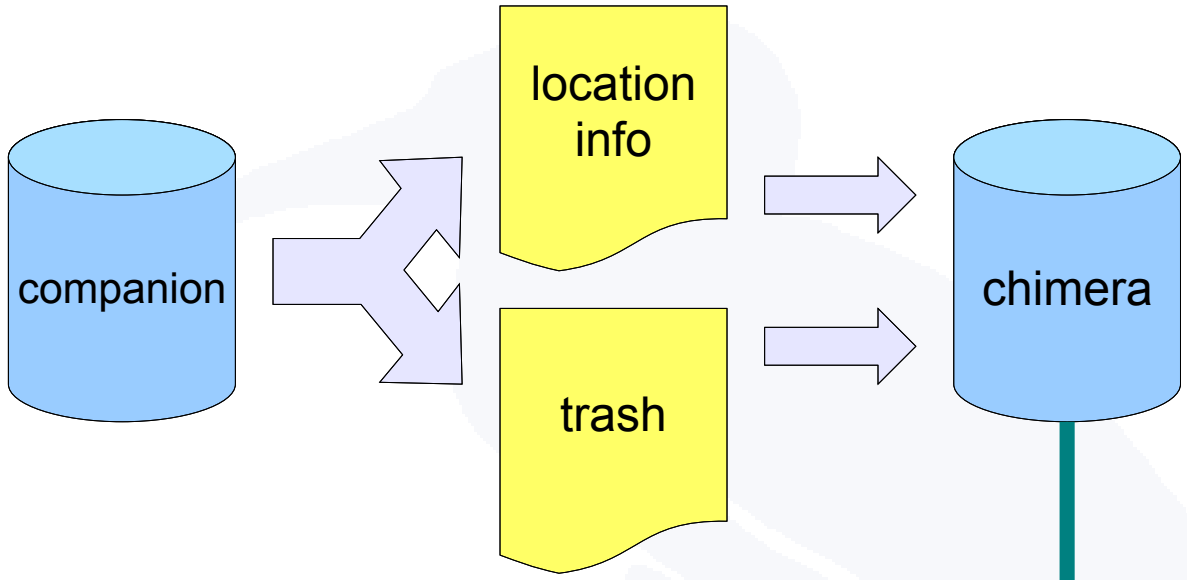


Inserting HSM info into Chimera

Gerd Behrmann
Mattias Wadenstein

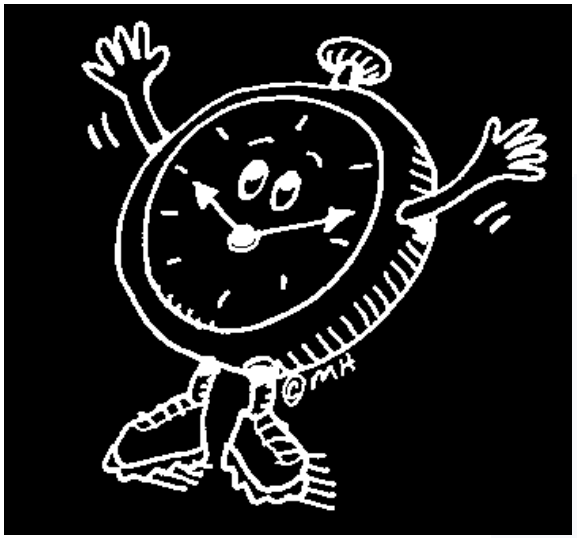


Inserting file 'location' info into Chimera.



Gerd Behrmann
Mattias
Wadenstein

Converting the entire NDGF Namespace



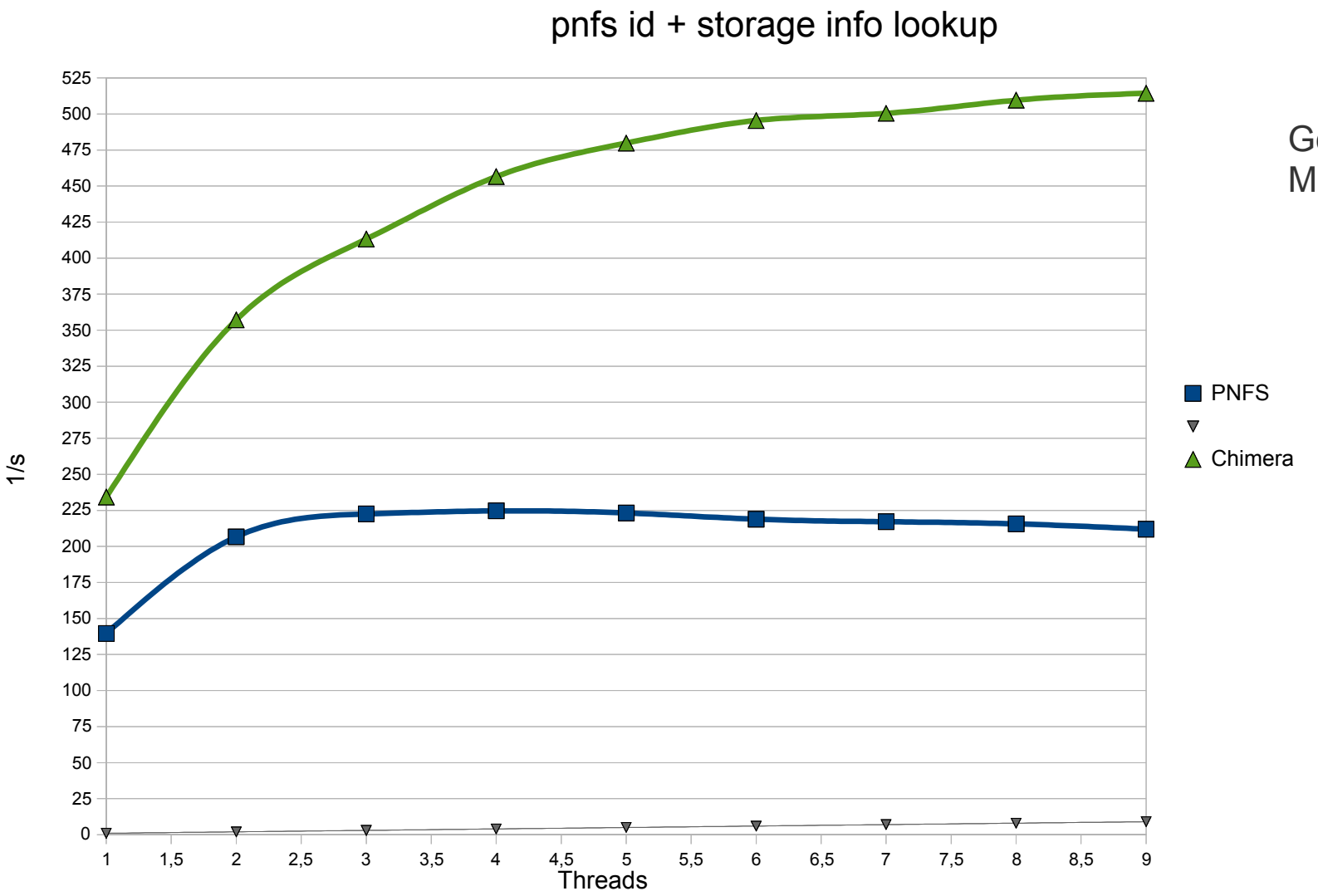
Gerd Behrmann
Mattias
Wadenstein

8 mill. files

pnfsdump:	11h
importing the SQL:	3.5h
md5sum verification:	11h
companion import:	4h
total:	30h

Chimera Performance

dCache.ORG



NDGF
Nordic DataGrid Facility

Gerd Behrmann
Mattias
Wadenstein

dCache.ORG

dCache.ORG

Access Control in dCache

ACLs in dCache

dCache.ORG

- Is required by Atlas. Not CMS at Tier I.
- Code development finished.
- Available in 1.9.3
- Introduced at the Aachen workshop 7. April
- As a result of the workshop, very good documentation is available.
- If Acl's are switched off, regular Unix permissions apply.
- Educational 1.9.3 Virtual Box is available.

dCache.ORG

Known Issues

Short term (pre D-day) improvements

dCache.ORG

- SRM has at least two duties :
  - Serve user requests as fast as possible.
  - Protect back-end storage system from overload.
- And two problems :
  - It doesn't do either.
    - Implementation problem
    - Protocol interaction problem
  - To much of an abstraction (Graeme S.).

dCache.ORG

How do we improve ?

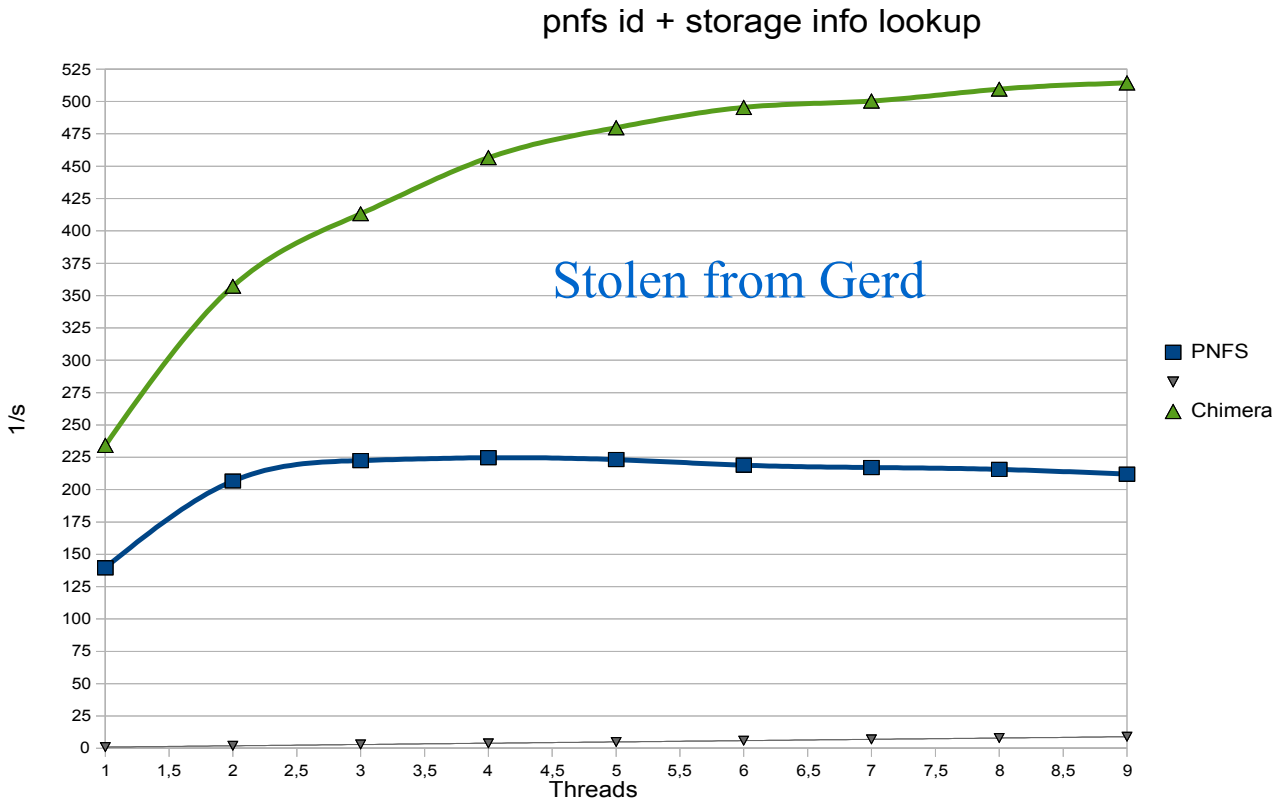
Gradually

Improvements NOW : Chimera

dCache.ORG

Obvious improvements : Make the back-end faster.

Faster name space : **Chimera** instead of pnfs.



dCache system can now be modified to support bulk operations on the name space level, which would make better use of SRM bulk requests.

Implementation **independent** improvements.

dCache.ORG

(This is a collaborative effort)

SRM_INTERNAL_ERROR

Inform the client that we are currently really busy and that we would appreciate if it would back off for a moment.

Request Lifetime

If client and server would agree (in advance) on the maximum time before both time-out a request, unnecessary requests wouldn't have to be processed.

Asynchronous SRM ls

The server may queue the request and proceed with light weight requests (e.g. get status)

Short term (pre D-day) improvements

dCache.ORG

Implementation **dependent** modifications :

Faster name space (pnfs to Chimera)

Stolen from Timur

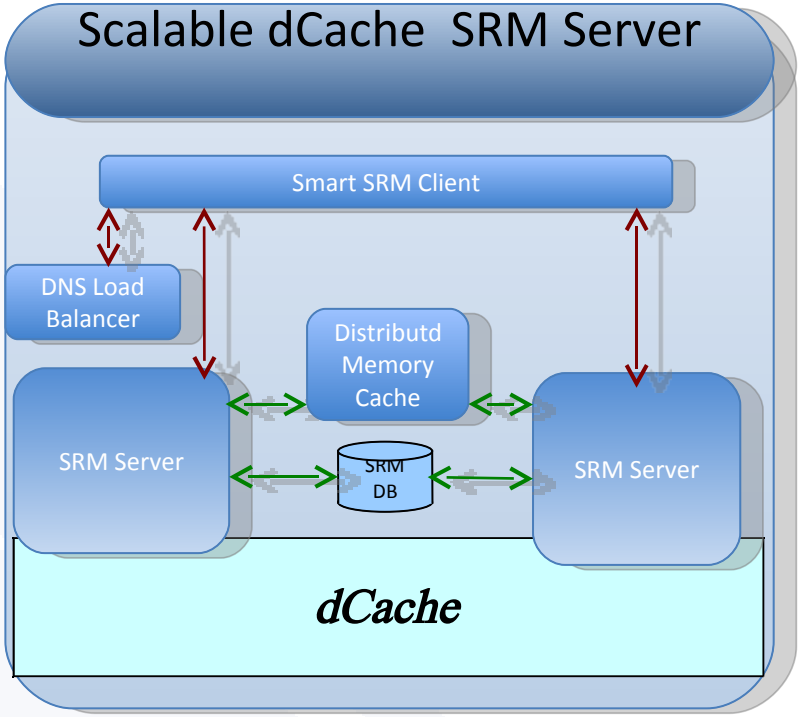
High CPU load due to GSI Authentication and Credential Delegation.

- Cache public and private key pairs used in GSI authentication and handshake.
- Work with Globus on improvements.
- Consider https as a long term solution.

Stolen from Timur

Scalability

- SRM is a single point of entry into a storage
  - Natural bottleneck
  - Single point of failure
- Distributed SRM
  - Scalable
  - More reliable



Further activities

dCache.ORG

Last month : External review of SRM in dCache.

- Good suggestions on SRM design improvements.

Next week : dCache developers meeting @ FNAL

Mid of May : SRM provider meeting @ DESY

- Improving interoperability of current implementation.
- Discussion of further SRM protocol simplification.

dCache.ORG

dCache.ORG

Further plans : NFS 4.1

Standardisation efforts : NFS 4.1 (pNFS)

dCache.ORG

- NFS 4.1(pNFS) is aware of distributed data.
- NFS 4.1 (pNFS) is an IETF standard.
- POSIX Clients are coming for free. No preload, no relinking. (provided by all major OS vendors).
- Widely adopted by major storage hardware vendors.
- Will make dCache useful to other (non-LHC) applications and communities.

NFS 4.1 : technical perspective

dCache.ORG

- NFS 4.1 is aware of distributed data
- Faster (optimized) e.g.:
  - Compound RPC calls
  - e.g. : 'Stat' produces 3 RPC calls in v3 but only one in v4
- GSS authentication
  - Built-in mandatory security on file system level
- ACL's
- dCache can keep track on client operations
  - OPEN / CLOSE semantic (so system can keep track on open files)
  - 'DEAD' client discovery (by client to server pings)
- smart client caching.

NFS 4.1 : status in dCache

dCache.ORG

- dCode is in trunk and can be released at any time.
- Waiting for standard kernel to have 4.1 intergrated.
- We expect this to happen end of the year. Up to then, kernel needs to be patched.
- Full fledged test instance available at DESY.
- Security not yet clear. Kerberos is standard. Certificates needs further discussions with CITI group.

dCache.ORG

Support

Support

dCache.ORG

Support for any dCache Tier I

Ticketsystem : Traffic is rather moderate.

Weeklyphone conference : not many problems rep..

Support for gridKa Tier I

Doris calls-in if necessary.

Support for german Tier II's

Should be done through HGF&DGI II storage support.
(Not well established yet. Mostly directly
by support@dcache.org , as well very moderate)

Support (Example)

dCache.ORG

Example for good collaboration : Atlas SE split-off at gridKa

- *Altas split off has been non trivial.*
- *dCache.org helped in preparing :*
  - *About 3 phone conferences*
  - *Some wikipages*
- *Doris and Silke did the split off all by themselves (excellent job).*
- *Finally an unfortunate issue has been encountered which required help by dCache.org.*

dCache.ORG

Further reading

www.dCache.ORG