Metadaten Workshop
26./27. März 2007
Göttingen

Chimera

a new grid enabled namespace service

Martin Radicke

Tigran Mkrtchyan

What is Chimera?

- a new namespace provider
- provides a simulated filesystem with additional metadata
- fast, scalable and based on RDBMS
- developed from scratch at DESY
    - lead development: Tigran Mkrtchyan

Motivation

- the dCache Storage Element as part of the LHC Data Grid
  - manages storage and exchange of data up to the petabyte-range
  - combines up to thousands of disk servers connected to tertiary storage providing a giant data repository
- all files presented under a central, single-rooted namespace
  - completely separated from the data servers

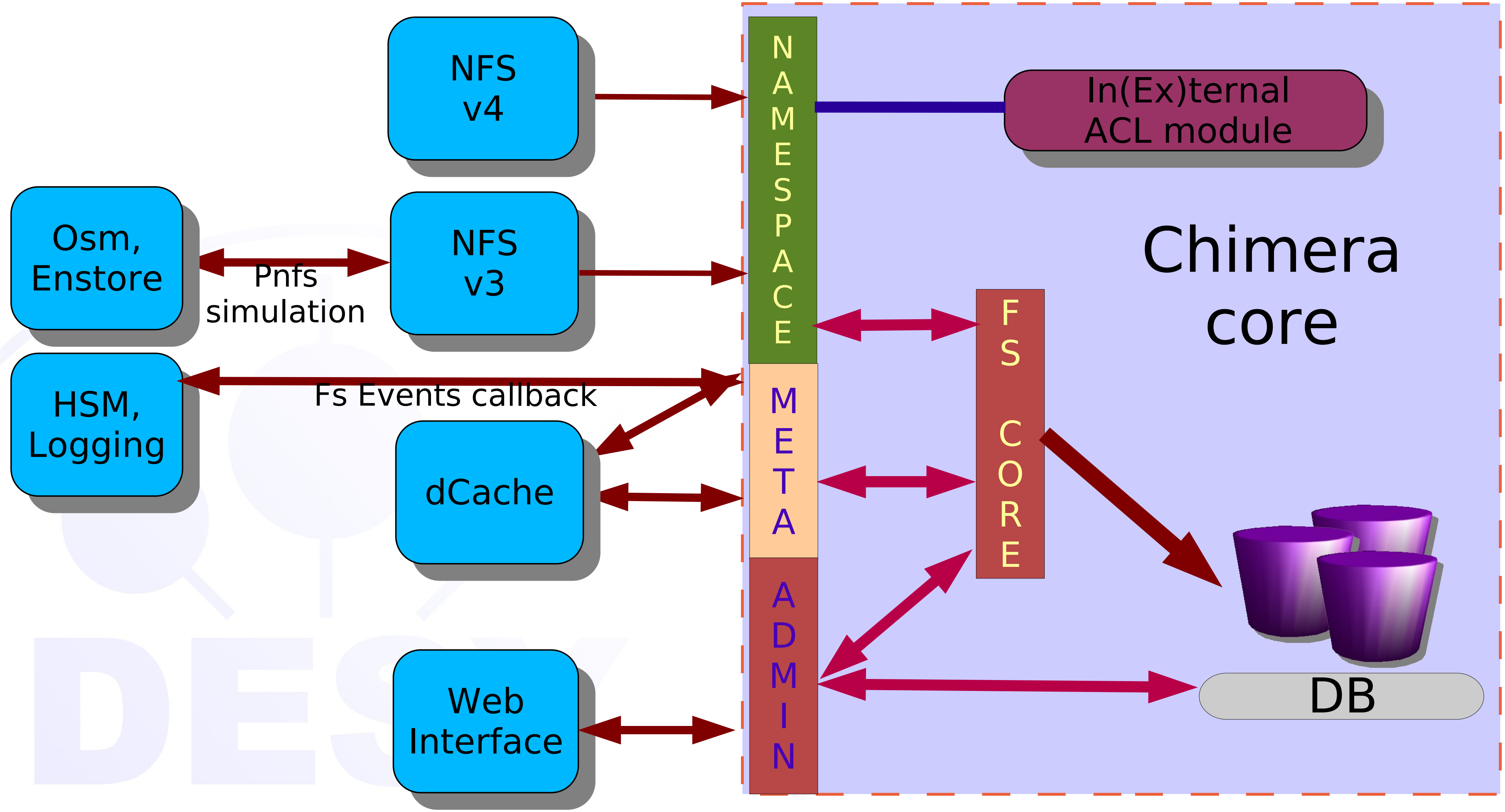
dCache Metadataprovider

- namespace provider currently in use: PNFS
- features:
  - filesytem emulation, extended by dCache-specific metadata regarding file location and tape backend
  - external access: mountable via NFSv2, gridFTP ('ls')
- expected performance bottleneck when dCache scales into the Petabyte-range with millions of file entries
  - heavy access to NFS operations has performance impacts on regular metadata queries done by dCache
  - no security (NFS v2 limitations), no ACLs

The new approach: Chimera

- many different users must be served concurrently without interfering each other and causing significant performance drop
- main features
  - dCache can access namespace directly, bypassing NFS for higher troughput
  - NFS v2 and v3 still supported for legacy clients (mount)
  - new client operations very lightweight (e.g. retrieving space used by a user/VO)
  - UNIX file permissions per default, Grid-enabled ACLs as plugin
  - callbacks on filesystem events (e.g. 'rm' oder 'move')

Architecture



Technical Overview

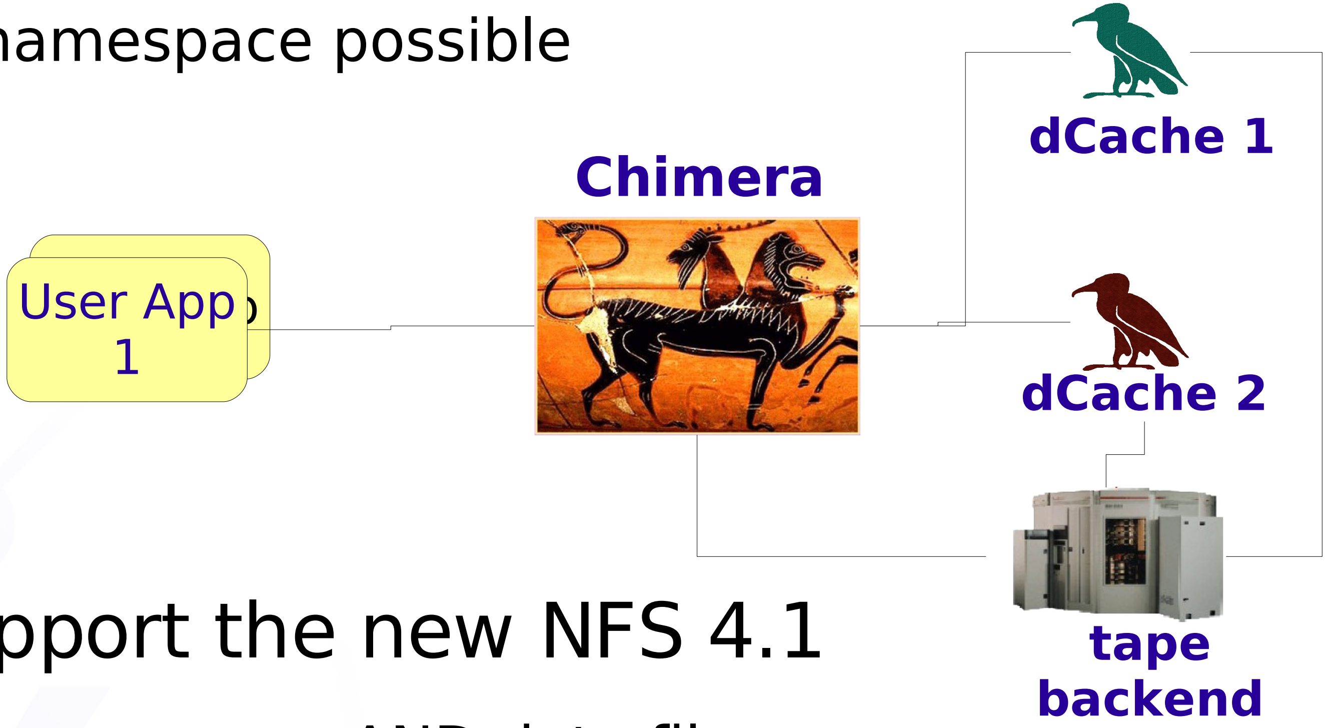
- complete redesign on top of relational databases
- smart DB schema allows isolation of queries of different types of metadata for better throughput
- well-defined API for namespace operations, metadata manipulations and admin interface
  - easy frontend creation (e.g. file browser, quota check)
  - can be extended by new metadata (7 levels per file prepared)

Technical Overview (II)

- plugin-interface for permission handler
- all major NFS versions supported
  - NFS v2: legacy
  - NFS v3: legacy, overcame the 2GB file size limit
  - NFS v4: more efficient communication, GSS authentification
  - NFS v4.1: client redirect allowed
- platform independent
  - pure Java
  - strict JDBC (no database-specific bindings, but possible)

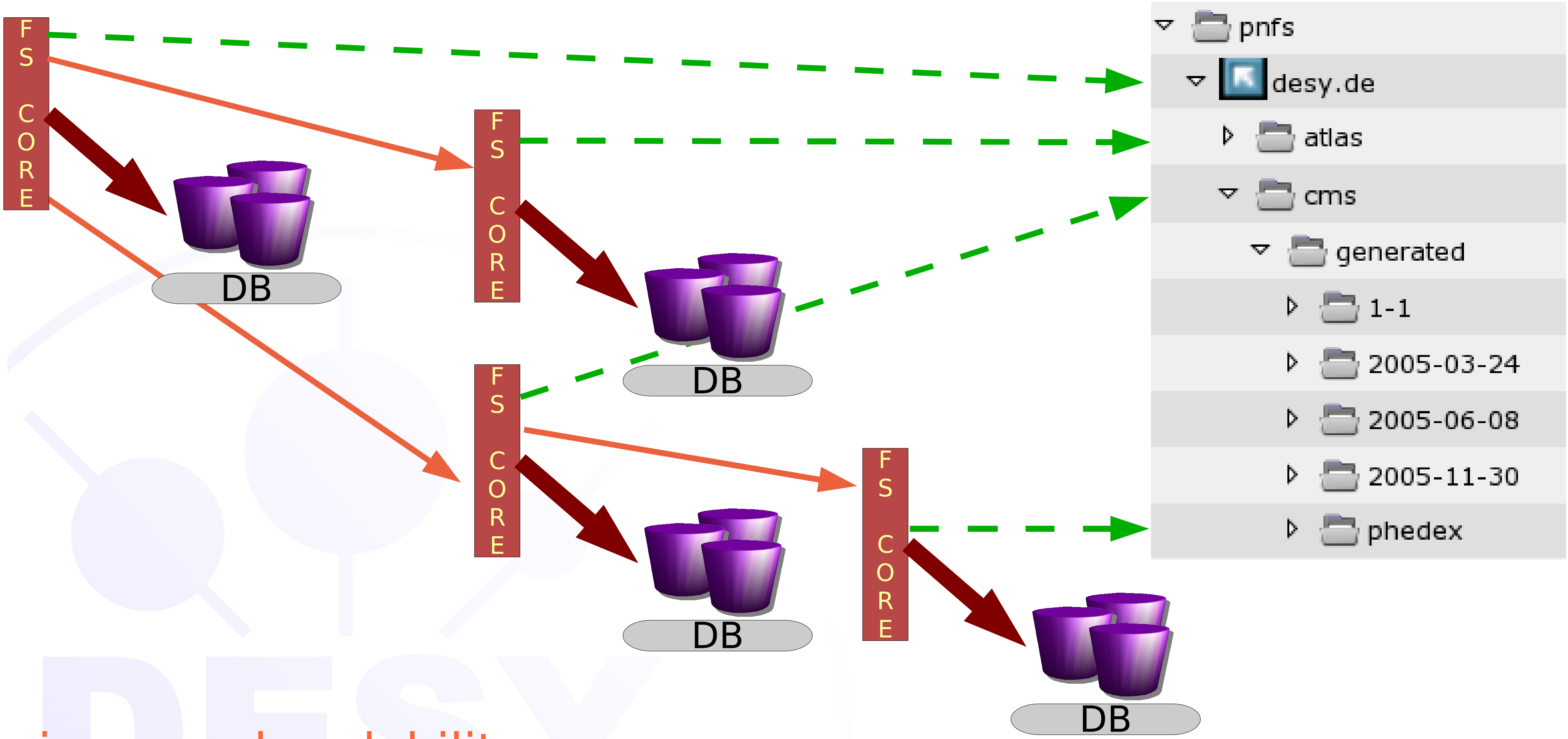
Benefits for dCache SE

- Chimera will overcome the limitations expected with PNFS handling several million file entries
  - still one shared namespace possible



- Chimera will support the new NFS 4.1
  - one protocol for namespace AND data file access
  - NFS 4.1 client will be part of Linux kernel

Filesystem chaining



- pnfs
  - desy.de
    - atlas
    - cms
      - generated
        - 1-1
        - 2005-03-24
        - 2005-06-08
        - 2005-11-30
        - phedex

- improved scalability
- each sub-chain can be root in other view

Status

- 200 file creates per second per thread
- tested against ORACLE, PostgreSQL, MySQL
- full working beta in test evaluation by FNAL
- to do
  - full scalable tests (any volunteers?)
- beta package will be public available soon (April)

The future

- Chimera, the next-gen Grid file catalogue ?
  - suggested by SARA, currently under discussion
- motivation: frequent inconsistencies between global LFC and local namespaces of the Site's SEs
- solution: distributed Chimera to provide a global, single-rooted and hierarchical namespace
  - possible through the previous mentioned chaining feature
  - additional replica metadata, file catalogue interface not yet there

Summary

- Chimera is a standalone and scalable namespace provider
- designed to serve many different users on top of it
- will be the main metadata provider for dCache to help it scale to the petabyte range
- extendable by new types of metadata and frontends using the API
- allows chaining for distributed and hierarchical namespaces

Thank you!

Contact: support@dcache.org