dCache, releasing and testing

Patrick Fuhrmann

on behave of the team

additional funding, support or contributions by

new release coming soooon   1.9.x

1.9.1 and 1.9.2 : somewhere this week

  ➢ *new pool code*
  ➢ *information system resp. information provider*

1.9.3 : minor problems

  ➢ *eg : problems reported by Ron (gsidcap, int. interface)*
  ➢ *load sharing improved (gap ...)*

1.9.4 : Implementation of the short term MoU

1.9.5 : 4 – 6 weeks from now

  ➢ *new Authorization/Authentication (gPlazma)*
    ➢ *DN -> uid*
    ➢ *FQAN -> gid*
  ➢ *ACL's on Chimera or Pnfs (Chimera preferred)*

dCache internal/external testing

- *Internally : Automated regression testing (Hudson)*
  - *runs after release is cut*
  - *includes Flavias S2 tests*
- *We provide SRM/dCache endpoint : srm-devel.desy.de*
  - *Tested by Flavia*
- *We provide 1 – 2 dCache (v)machines to be tested against gLite middleware (theoretically by CERN)*

All this doesn't replace proper testing by (W)LCG.

The final testing for WLCG applications resp. middleware needs to be done by our customers (e.g. like OSG is doing)

We offer support@dCache.org   prioritizing Tier I's

We offer weekly Tier I phone conferences
(Atlas only joined when explicitly invited)

OSG storage people are present during dCache developers phone conference. Tickets are reviewed.

dCache.org organizes the German dCache support funded by HGF and German e-science projects. Organization of tutorials and workshops. e.g gridKa school.

dCache.org is hiring 2 more people Okt, 2008.

dCache.org is connected to the GGUS ticket system, though we have to prioritize requests.

dCache.ORG

dCache.ORG

➤ *The WLCG is a community approach.*

➤ *Storage issues are still challenging.*

➤ *dCache.org is entirely funded by FNAL, DESY and NDGF*

➤ *OSG is self organising storage and providing developers*

➤ *Germany is well organized.*

➤ *How about the rest. Who is responsible ?*

➤ *We need better collaboration if new requirements come up (xroot, ??gsi??xroot)*

- Providing software ... a reliable service ?
  - Commitments of the funding parties.
  - Stable Team.
  - Acquisition of new partners.
  - Transparent development process including
    - Detection of deficiencies and upcoming scalability issues.
    - Staying in touch with future technologies and industry solutions.
  - Customer interactions.
  - Transparent software design.
  - Software quality process and certification.
  - Package distribution
  - Support infrastructure
- dCache 1.8 and SRM 2.2
- dCache 1.8 deployment
- Did we learn something from the SRM 2.2 exercise ?

Providing software ... a reliable service ?

Commitments of the funding parties (Quotation from original sources)

FERMIlab

Fermilab intends to continue to support SRM 2.2 (on dCache) in the form of bug fixes as long as there is a reasonable HEP community using the software. Work on Future versions of the SRM depends on Fermilab's programmatic needs, and its ability to support the envisioned deployment community.

NDGF Nordic DataGrid Facility

- NDGF currently considers dCache the best solution to our storage needs. We will continue to contribute to dCache development while dCache is used at NDGF.
  - We provide first level support for sites deploying dCache in the Nordic countries.
  - NDGF will provide second level support for our contributions to dCache as long as there is a need and we are able to do so, which is at least for as long as we use dCache ourselves.
-

DESY

We will maintain the **dcache.org infrastructure** whilst there is a reasonably large community using the dCache technology.

Whilst **dCache software** is in use at DESY, support will be at least "best effort" for those components contributed by DESY. The availability of additional funding will allow DESY to provide an enhanced level of support. This additional support will be dictated by funding sources, appropriate to level of funding made available and guaranteed for the duration of this funding.(EGEE, d-grid, ...)

dCache.ORG

dCache.ORG

Head of dCache.ORG

Patrick Fuhrmann

Core Team (Desy, Fermi, NDGF)

Andrew Baranovski
Gerd Behrmann
Bjoern Boettscher
Ted Hesselroth
Alex Kulyavtsev
Iryna Koslova
Dmitri Litvintsev
David Melkumyan
Dirk Pleiter
Martin Radicke
Owen Synge
Neha Sharma
Vladimir Podstavkov

Head of Development FNAL :

Timur Perelmutov

Head of Development DESY :

Tigran Mkrtchyan

Head of Development NDGF:

Gerd Behrmann

External

Development

Abhishek Singh Rana, SDSC
Jonathan Schaeffer, IN2P3

Support and Help

Greig Cowan, gridPP
Stijn De Weirdt (Quattor)
Maarten Lithmaath, CERN
Flavia Donno, CERN

Acquisition of new partners.

- The Nordic Data Grid Facility (NDGF) is now part of the dCache development team and provides input on issues arising from the fact that the NDGF production dCache spans 4 countries and has to drive different Tape Storage Systems at the same time in different locations.

  - ➢ FTP Version II implementation
  - ➢ Multi HSM implementation

dCache.ORG

dCache.ORG

Transparent development process including

- Weekly phone conferences with all partners

- Regular all hands meeting (BNL August, DESY November, ??? End of January)

- Meeting agenda and minutes at www.dCache.org

dCache.ORG

dCache.ORG

Software design

e.g. :

- ➤ **Modern Programming Language   (java ©)**
  - ➤ Attracts young developers
  - ➤ Permanent improvements by Sun MS © and Java community
  - ➤ Industry/Eduational   standard

- ➤ **Message Passing between components (Cells)**

  - ➤ Multi path message passing for scaling reasons.

  - ➤ Easy plug-in of new Components (no relink necessary)

- ➤ **Distributed Queues to   protect the system from meltdown**

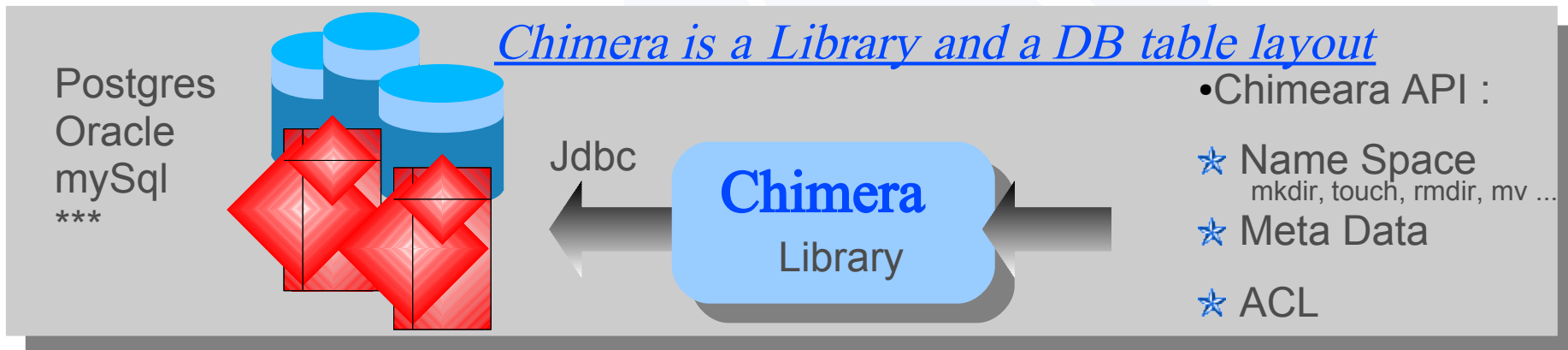- ➤ **Fast central queue (get, decide , forward and forget)**

dCache.ORG

Detection of deficiencies and upcoming scalability issues.

> ➤ Pnfs name space turned out to be the upcoming bottleneck.
> ➤ Will be replaced by Chimera.
> ➤ Chimera has been test at various independent places.
> ➤ Chimera is already optionally available in the 1.8.0-4 rpm.

Chimera ad

✻ *Chimera provides the same functionality to dCache as Pnfs does.*
✻ *Chimera is a Java API, a library and a database table layout.*
✻ *Chimera doesn't have any server by itself.*
✻ *Consequently it can make use of any DB performance improvements.*

Chimera is a Library and a DB table layout

Postgres
Oracle
mySql
***

Jdbc

Chimera
Library

• Chimeara API :

✻ Name Space
   mkdir, touch, rmdir, mv ...

✻ Meta Data

✻ ACL

Staying in touch with upcoming technologies and industry standards

NFS 4.1

Technical Advantages :

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

Deployment Advantages :

> Clients are coming for free (provided by all major OS vendors).

Staying in touch with requested technologies

ALC's

ALC's are ready for testing (in Chimera, maybe in pnfs as well)

xRoot

dCache supports the xRoot protocol as used by Alice and the

Alice security mechanism.

Customer interactions

User helping user : user-forum@dCache.org

SRM 2.2 deployment : srm-deployment@dCache.org

Ticket system : support@dCache.org

Regular phone conference with some big sites (on request)

Wiki/SVN area for customers feedback and contributions

During SRM 2.2 development and deployment   phase

Weekly phone conferences with GSSD (and dCache Tier I's)

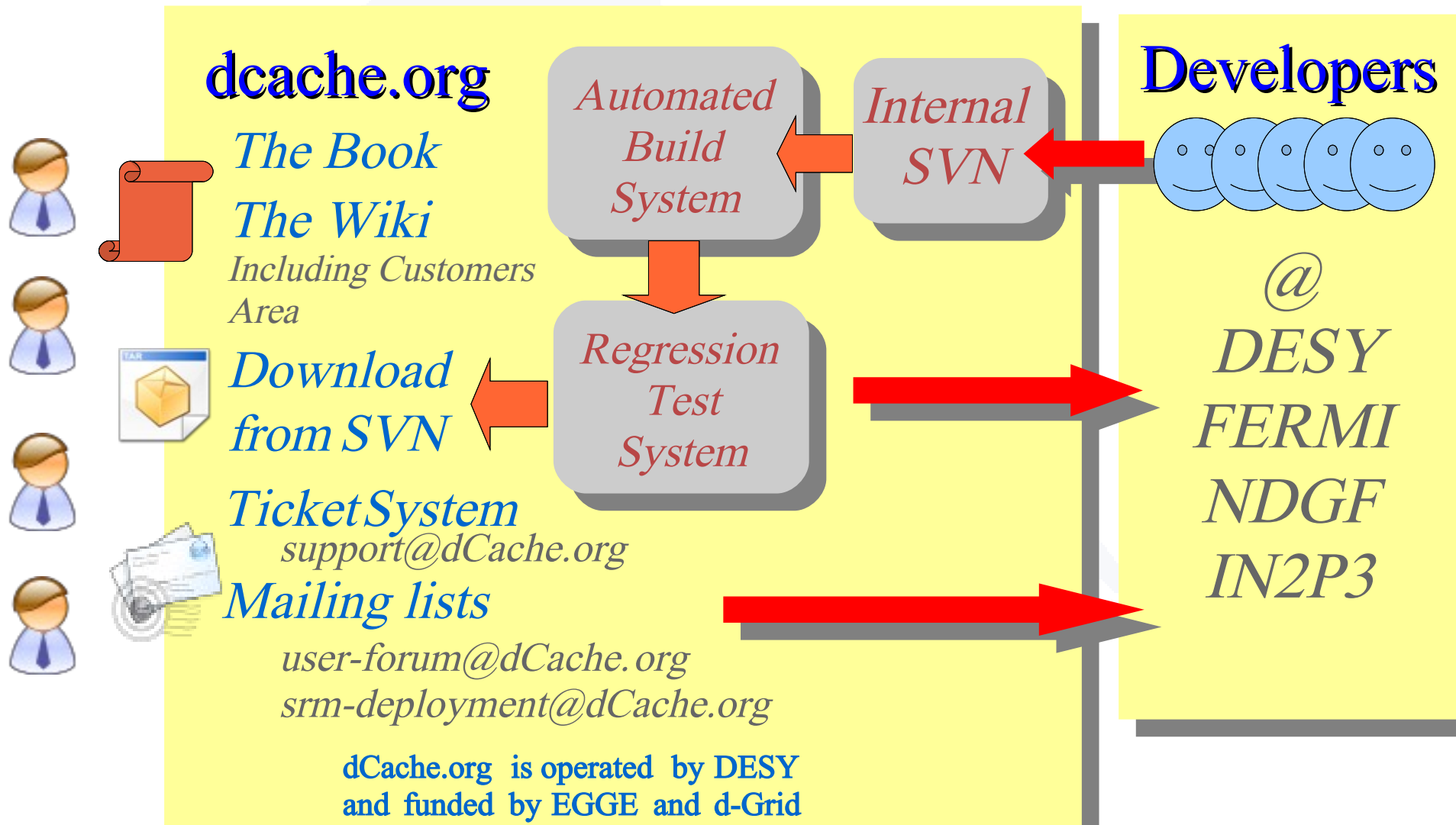Weekly MGMT phone conference of SRM 2.2 providers.

dCache.ORG

- *dCache.ORG is an infrastructure*
- *dCache.ORG is the door into the dCache team*

dcache.org

Developers

Automated Build System

Internal SVN

The Book

The Wiki

Including Customers Area

Download

from SVN

Regression Test System

@

DESY

FERMI

NDGF

IN2P3

Ticket System

support@dCache.org

Mailing lists

user-forum@dCache.org

srm-deployment@dCache.org

dCache.org is operated by DESY
and funded by EGGE and d-Grid

dCache.ORG

dCache.ORG

dCache distribution

dCache is distributed and configured through YAIM (CERN, DESY) for the Tier II's in Europa.

dCache is distributed and configured through VDT for OSG supported Tier II's

dCache is distributed through dCache.org for the Tier I's and configured
manually due to the complexity of the Tier I setups.

dCache.ORG

dCache.ORG
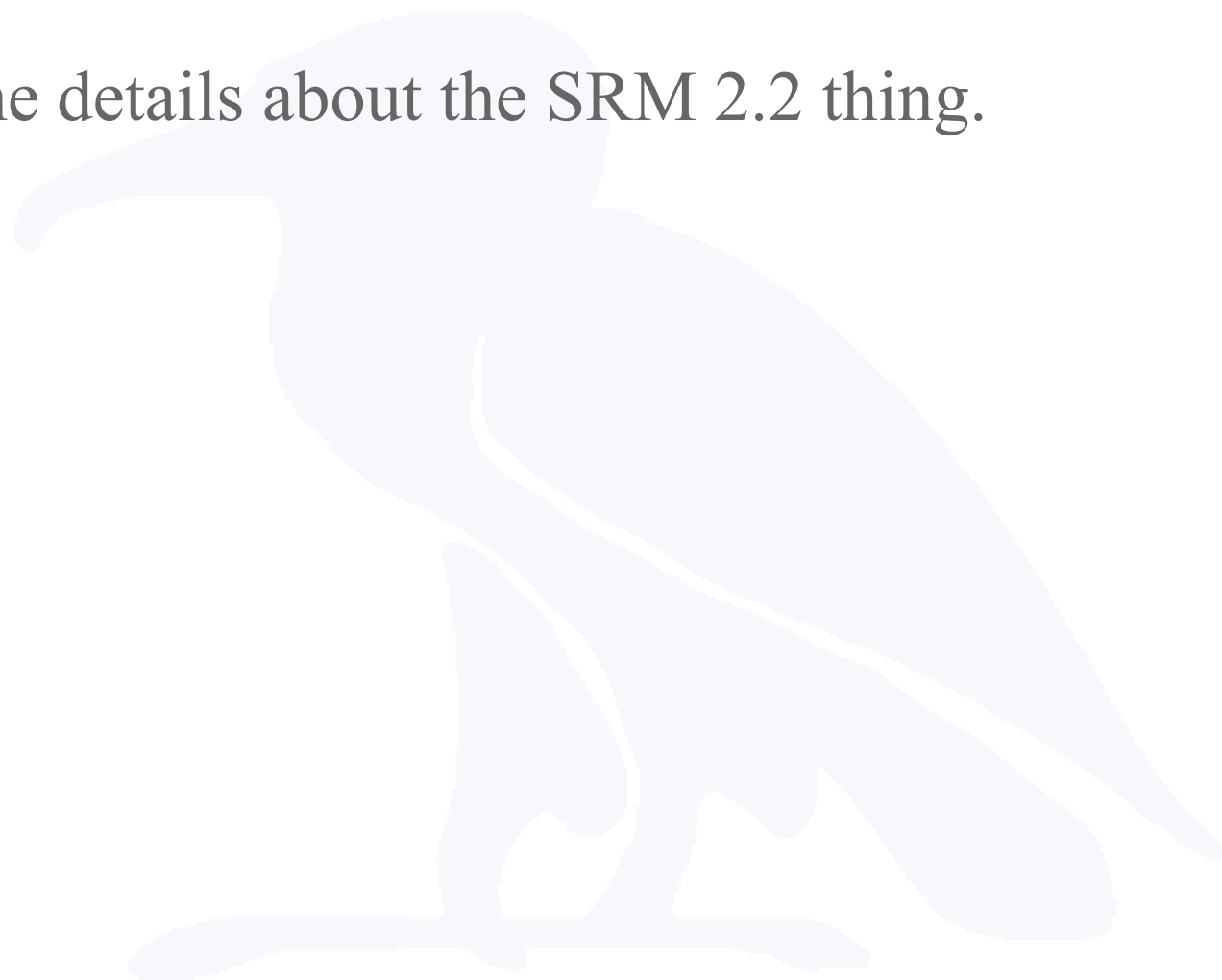
Support Infrastructure

Already established

- › US (Atlas, CMS) : Open Science Grid (OSG)
- › UK : gridPP

Starting

- › Germany (*Physics on the Tera-scale*)
  - › Funding starts : Jan 2008
  - › Aachen, Hamburg, Karlsruhe, Munich
- › Italy (BARI)
- › GGUS (Technically connected to support@dCache.org)
- › GGUS still needs skilled 'storage' team.

Some details about the SRM 2.2 thing.

dCache 1.8 has been "Certified" by Flavia's S2 test system to be fully compliant with the WLCG SRM 2.2 specification.

During the PPS test phase some misinterpretations of the agreements on how to handle spaces have been detected and fixed.

The evaluation of the PPS systems by the experiments have been "amazingly" successful. Results have been presented during the pre GDB and the GSSD meetings.

Weekly phone conferences with the dCache Tier I representatives as well as phone conferences with the WLCG, Castor/DPM and dCache management have been extremely helpful. (Many thanks to Jamie and Flavia).

dCache.ORG *(left margin, vertical)*

Stolen from a talk to the GSSD/pre-GDB by

Birger Koblitz, CERN-IT, for the ATLAS DDM team

- SRM2 works very well as a drop-in replacement for SRM1, at least for the dCache and StoRM implementations

- Pre-staging should be able to help a lot in efficiency

- Unfortunately: Files in SRM-Test instances also picked up for transfers to other sites

  – Reading tested in uncontrolled way, too!

- Next phase: Need to integrate better with ongoing tests...

Details on dCache 1.8 and the SRM protocol versions :

- To use SRM 2.2 in dCache, the upgrade to dCache 1.8 is mandatory.
- dCache 1.8 still supports the SRM 1.1 protocol.
- The use of SRM 2.2 and space management are optional in dCache 1.8
- The decision on *if and when* to enable the SRM 2.2 protocol and space management is up to the sites and the experiments they host.

Details on the dCache 1.8 upgrade procedure :

- The upgrade to dCache 1.8 is a normal rpm software upgrade.
- No data or meta data conversion needs to be done.
- The upgrade can be performed in parallel on all dCache components.

dCache 1.8 (including SRM 2.2) deployment

The Tier I production system deployment started Oct 29 with NDFG and is ongoing. dCache is currently installed at 9 Tier I's.

NDGF	Oct 29	OK
GridKa	Nov 6	OK
SARA	Nov 19	OK
IN2P3	Nov 26	Today
FermiLab	Nov 29	
RAL	Dec 3	
Triumf	Dec 3	
BNL	Dec 10	
Pic	Dec 17	

dCache 1.8 will replace dCache 1.7 in the VDT package presumably mid of February

The about 40 dCache Tier II's (LCG & OSG) may join at any time.

Issues detected during the NDGF upgrade :

➢ NDGF has not been enabled the space management subsystem yet. (Site decision)

➢ This will happen very soon.

➢ All issues detected during the upgrade phase have been fixed.

Issues detected during the gridKa upgrade :

➢ GridKa has been enabled the Space Management subsystem.

➢ There is still one remaining software issue in the space management subsystem, which is currently being worked on. (Problem has been fixed Saturday and will be made available this week.)

➢ Due to the fact that based on the WLCG SRM 2.2 agreement, only incoming data is supposed to be managed by the space subsystem, it is not yet clear how to treat data being restored from tape. This is primarily a policy decision. (Should those datasets go to managed space or not?)

➢ T0D1 switches to T0D0 in under some circumstances (Fixed with 1.8.0-4)

No issues at SARA

Up to now, all upgrades have been done within the scheduled time slot.

There have been no problems detected yet, which prevents us from suggesting to proceed with the scheduled deployment procedure.

Summary

- Sufficient number of funding sites (groups) committed to support dCache in the future.
- dCache.org nearly professional infrastructure. (download, source code, wiki, ticket system, mailing lists)
- 3 independent development sites (which all reply on the product) lower the risk of doing something really stupid.
- More sites asking to contribute (support and development)
- System design scales into the xx PetaByte range.
- New features in the queue : chimera , acls, nfs4.1
- Support infrastructure in the process of being extended.
- SRM 2.2 in functionality certified.
- SRM 2.2 deployment progressing as scheduled.

What did we learn from the SRM 2.2 lesson ?

- Never allow *specification. -> implementation -> customer.*
- Insist in *customer -> specification -> implementation.*
- Make sure your customers/users understand and need the specification before implementing it
- Never allow pressure when committing to a specification.
- Never agree on a simplified or partial specification, assuming this would speed up the implementation process.
- Never agree on a specification before knowing exactly how to implement it.
- Never allow a spec to reduce your system functionality.
- Never let a group (with their own interests) between you and your customers.

Further reading

www.dCache.ORG