Status and plans

Patrick Fuhrmann

On behave of the project team

INDIGO - DataCloud

neic

Fermilab

NDGF Nordic DataGrid Facility

DESY

HELMHOLTZ | ASSOCIATION

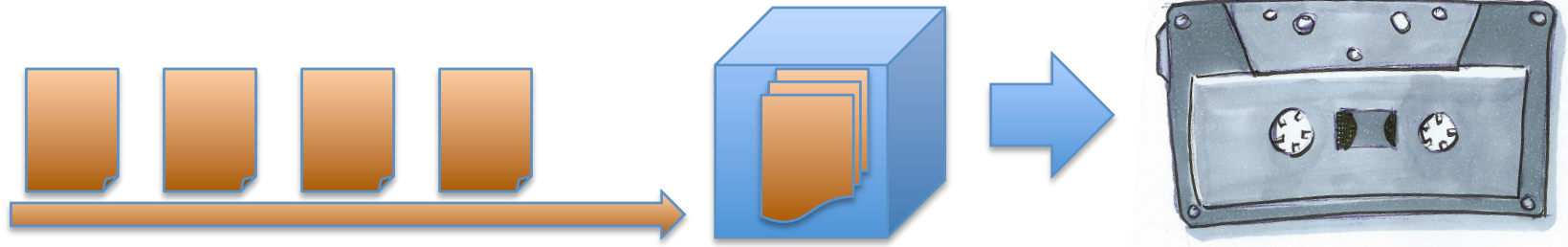Federal Ministry of Education and Research

LSDMA

What did we do since the last workshop ?

- Tidying up (See Gerds presentation)
- Small File Support for Tape (Karsten)
- Getting NFS into Production (Tigran)
  - CMS Worker nodes
  - Underneath OwnCloud
- Investigating in Cloud Semantics (Paul)
  - CDMI or S3 (With low priority)
- Completing new Resilient Manager (Al)
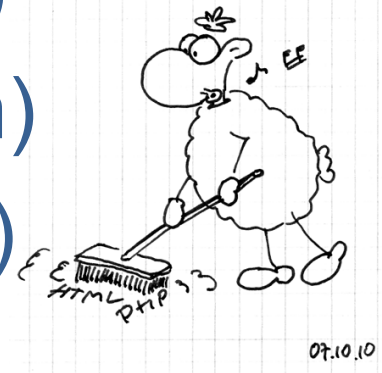
Small files for Tape

- Small files are collected in containers, which are again stored in dCache before written to tape.

- Transparent recall of the container if one file is recalled from tape archive.

- Container is kept on online storage for further usage (recalls) as long as cache space is available.

- First costumers
  - DESY Light source groups (small pictures)
  - DP-HEP (Long term data preservation in HEP)

- See CHEP'15 Poster or talk to Karsten

Deploying NFS into production

- The Desy-Cloud

- CMS Grid Infrastructure @ DESY

New Production Systems based dCache.org
on dCache NFS.

ownCloud

NFS 4.1 / pNFS
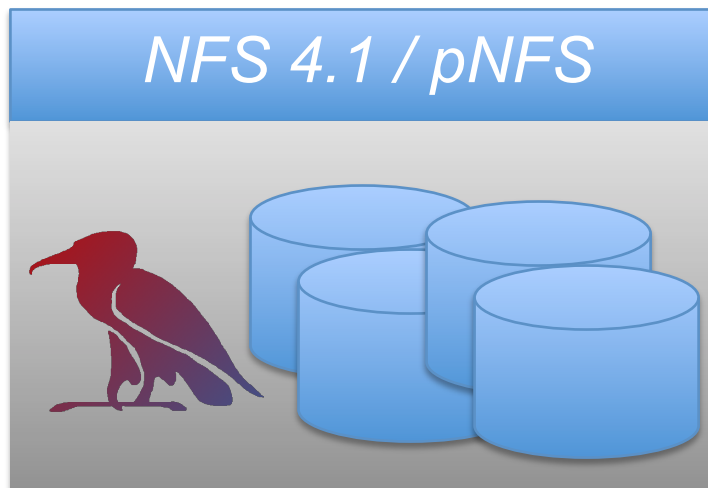
dCache Backend Storage Layer

Sync & Share
Laptops
Mobile Devices

Direct
Low-latency
access
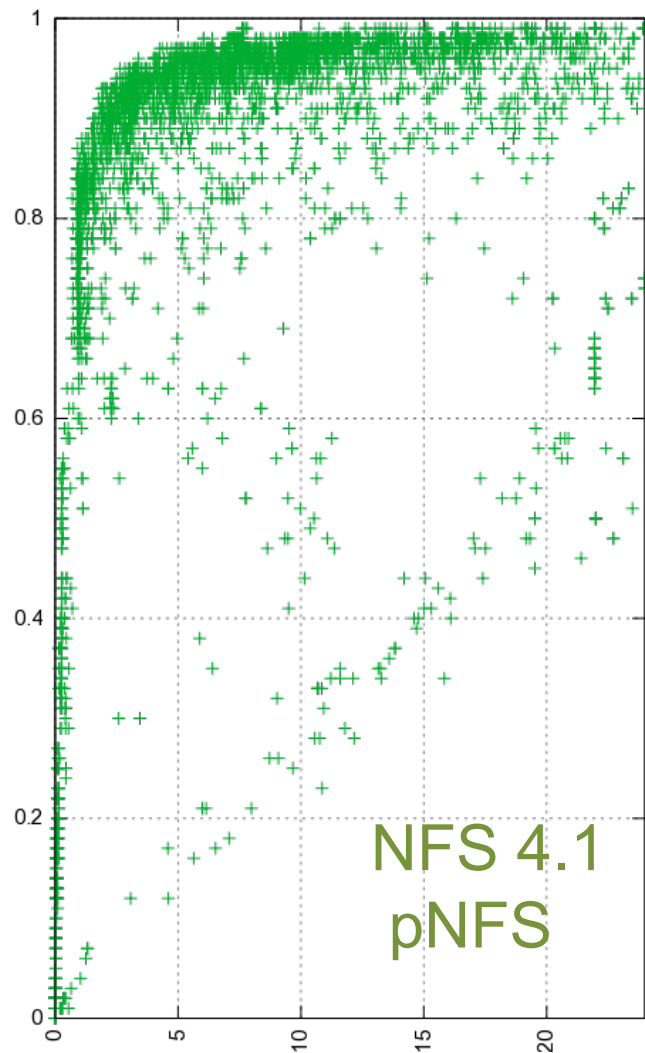Workernodes
HPC

Wide Area
FTS
GLOBUS (ONLINE)
HTTP (WebDAV)

CMS Tier II @ DESY

- Slowly migrating CMS Grid worker nodes to NFS4.1 data access.

- Good experience as long as the network is stable.

Job Efficiency (NFS – dCap)

NFS 4.1
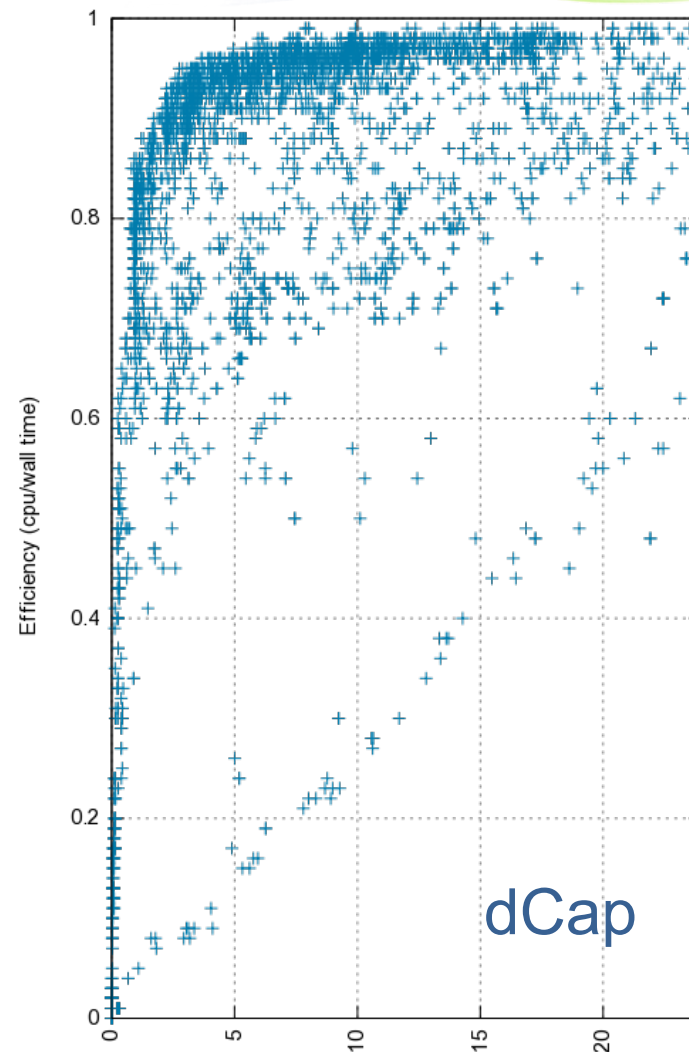pNFS

dCap

Job Efficiency (CPU / Wall Time)

Execution Time (hours)

As with all news spec's, there are issues

Stolen from Tigran

- Network problems cause the system to be behave unpredictable.

- Data Server behind firewalls

- Weak clients on VM's

- Specification Violation
  - infinite state recovery with Linux kernel

New Resilient Manager

- Keeping > 1 copies of each file
  - Redundancy
    - For pool maintenance
    - In case of pool failure
  - Higher overall performance
- First Steps towards : Software Defined Storage
- In development phase (AI)

Exploring new communities.

New communities

- Through German national projects we became involved in HPC (Dresden).
  - VAVID (Wind Power Generators simulation
  - MOSGRID (Molecular Structure Simulation)
- envision to access mass data PLUS meta data from Super Computers (CDMI)

Exploring more …

dCache.org

- ## German support for the Human Brain Project (SMHB)

  - Jülich – Aachen Research Alliance
  - Distributed dCache between Aachen and Jülich
  - dCache's ability to select pools close to the client or to move data closer to the client made it a perfect match for their requirements.
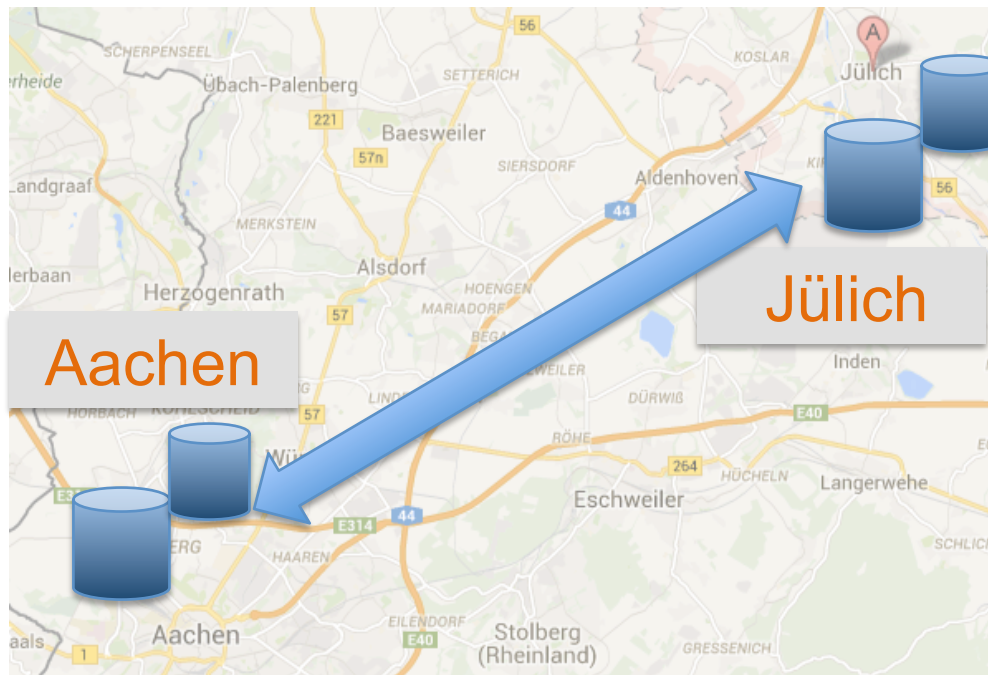
- Two cities, one system.
  - Similar to NDGF (4 Countries one system)
  - Second copy automatically generated at the other location.
  - Or second location just used as a cache.

And more

- Intensity Frontier (IF) at Fermilab.
  - Quote "Craig Group" (plenary talk at CHEP'15)

Nice

- dCache
  - Highly distributed storage with central name space
  - Much lower cost (~$100/TB), ~4PB shared by IF experiments
  - Read / Write interfaces, but does not look like usual file systems
  - Accessible from off-site
  - A cache (optionally front-end to tape system) -- old files are flushed

Hm, actually it does … That's why we have NFS4.1

dCache: responding to new technologies

Responding to new technologies

- New Disk Technologies (dCache in a BOX)
  - Open Ethernet Disks (HGST, Manfred's Talk)
  - dCache in DDN (Triumf, Simon's Talk)
- New Object-Store Back-ends
  - CEPH
- New European Projects (INDIGO DC)
  - Focusing on Data Quality of Service and
  - Data Lifecycle Management

HGST Open Ethernet Disks
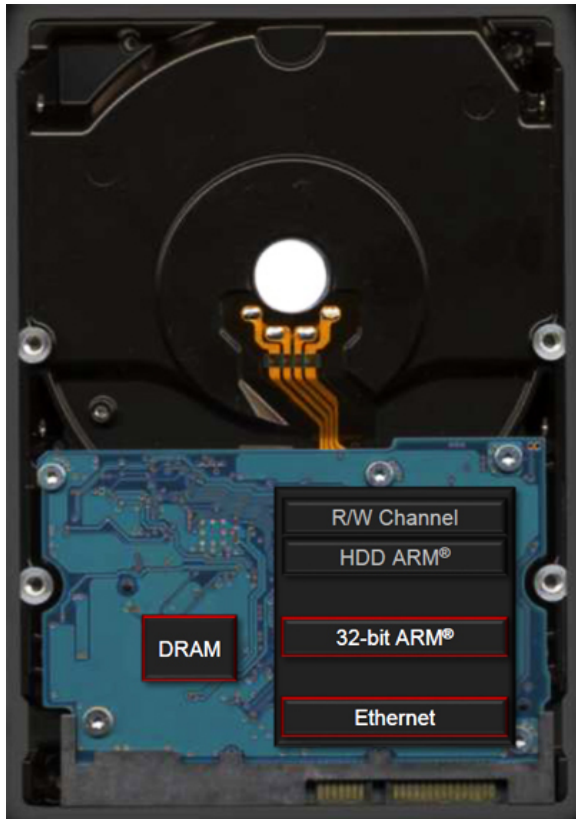
- Small ARM CPU with Ethernet piggybacked on regular Disk.
- Spec:
  - Any Linux (Debian on demo)
  - CPU 32-bit ARM, 512 Level 2
  - 2 GB DRAM DDR-3 Memory
    - 1792 MB available
  - Block storage driver as SCSI *sda*
 - Ethernet network driver as *eth0*
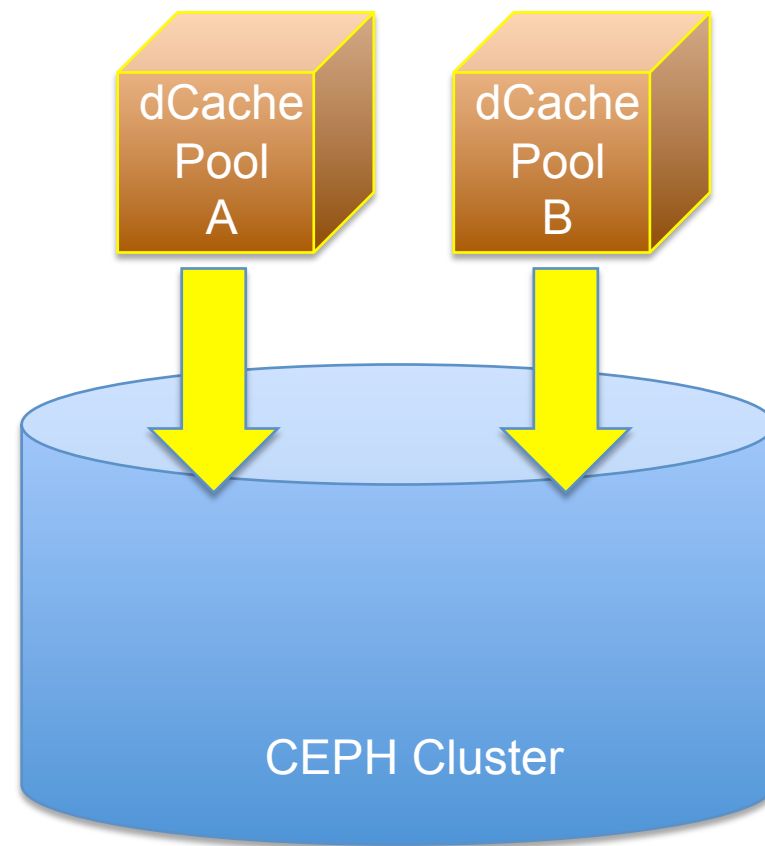
Response to ceph

- CEPH complements dCache perfectly. (Seems to be better than HADOOP)
  - Simplifies operating dCache disks.
  - dCache accesses data as object-store anyway already.
- dCache is evaluating a 'two step approach'.
  - Each pools sees it own object space in CEPH
  - All pools have access to the entire space, which is a slight change of dCache pool semantics.
- Would merge CEPH and dCache advantages
  - Multi Tier (Tape, Disk, SSD)
  - Multi protocol support for a common namespace.
    - All protocols see the same namespace
  - All the dCache AAI features
    - Support for X509, Kerberos, username/password

Possible CEPH deployments
(only Object Store variants)

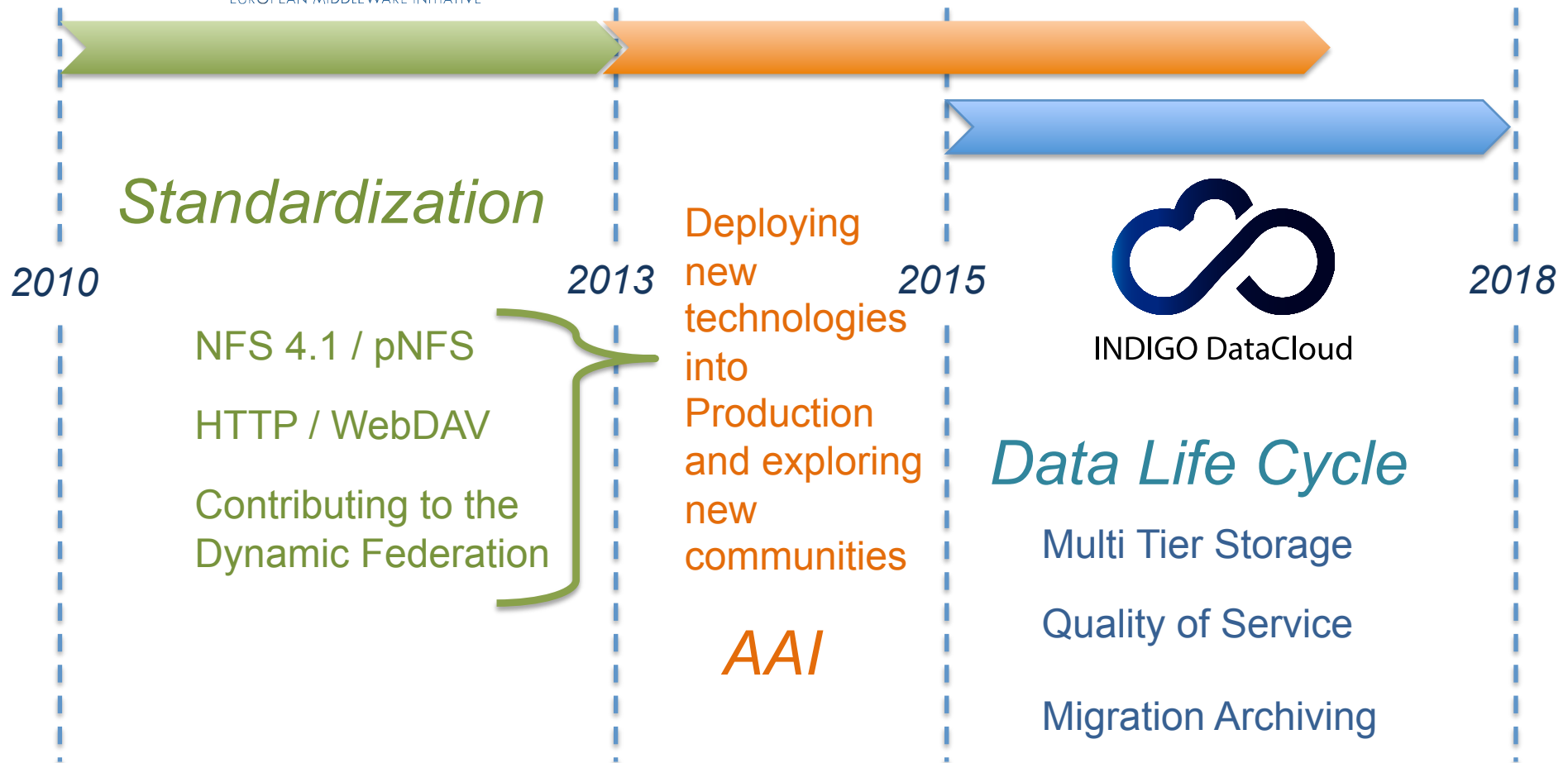Each dCache pool still only 'sees' his own private repository.
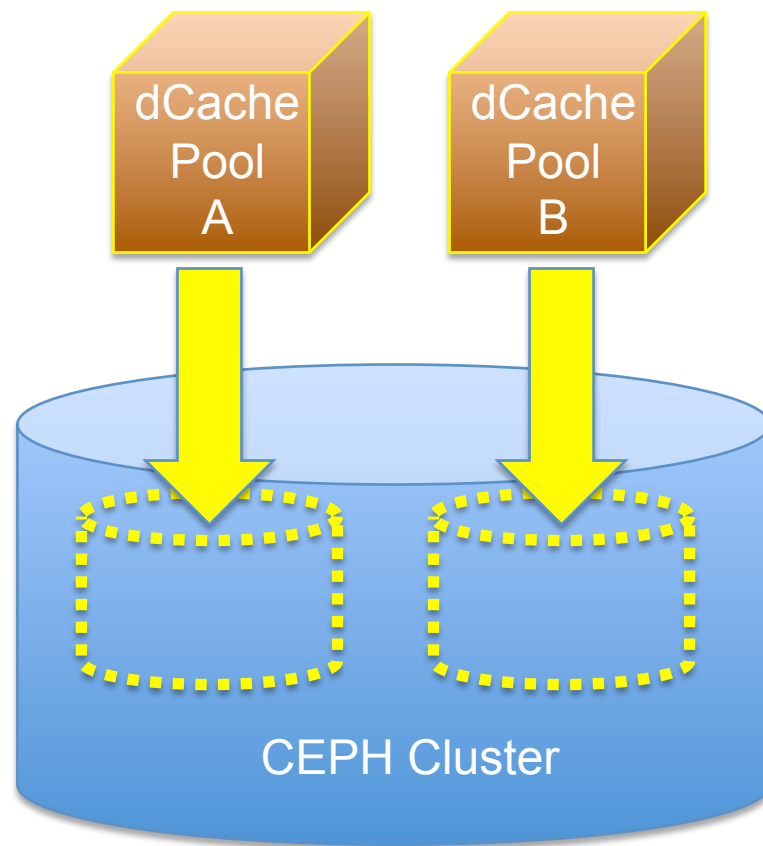
dCache pools can use shared repositories. Requires new pool semantics. (Focus on protocol engine)

Future : Funding and Work Program

Funding influences dCache development topics

dCache.org

EMI — EUROPEAN MIDDLEWARE INITIATIVE

LSDMA

Standardization

2010

2013

2015

2018

NFS 4.1 / pNFS

HTTP / WebDAV

Contributing to the Dynamic Federation

Deploying new technologies into Production and exploring new communities

AAI

INDIGO DataCloud

Data Life Cycle

Multi Tier Storage

Quality of Service

Migration Archiving

Some words on INDIGO-DataCloud

INDIGO-DataCloud Cheat-Sheet

- Horizon 2020 project started April 2015

- Project led by INFN Italy

- Budget 11.1 Million Euros (800.000 for dCache)

- 26 Partners = 11 European Countries

- Duration 30 months

- The project aims for an Open Source Data and Computing platform targeted at scientific communities, deployable on multiple hardware, and provisioned over private and public e-infrastructures.

INDIGO in a nutshell

1. **Self-service, on-demand**
2. **Access through the network**
3. **Resource pooling**
4. **Elasticity (with *infinite resources*)**
5. **Pay as you go**

In the end,
Applications Rule.
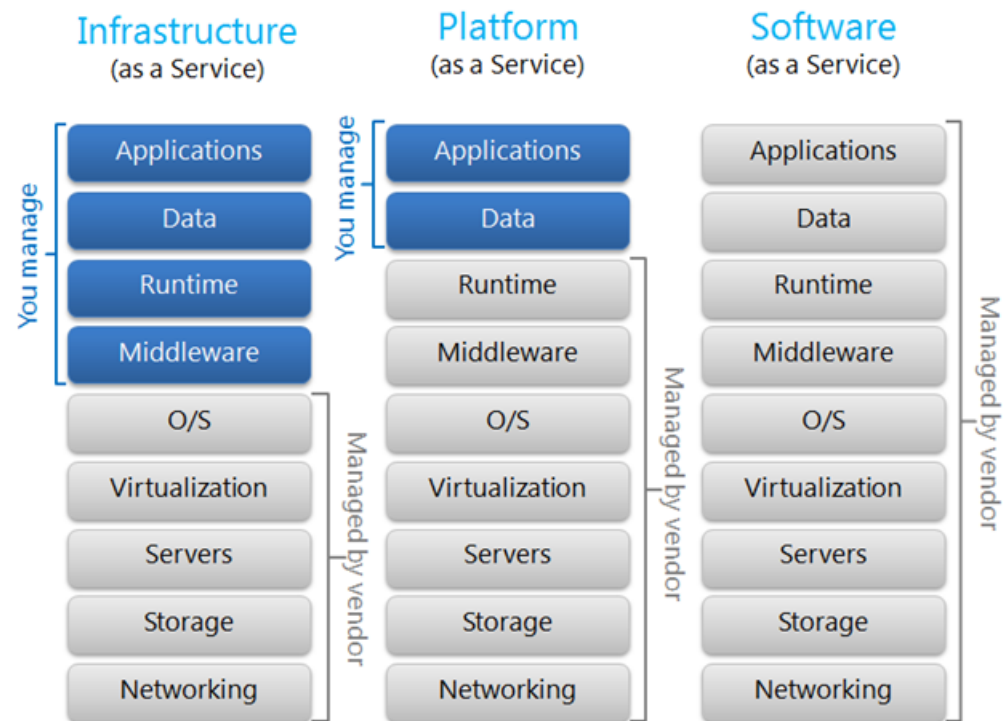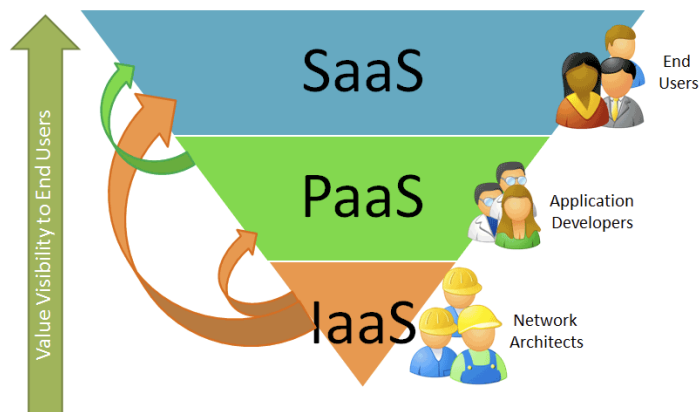
Stolen from Davide Salomoni (Project Director)

dCache involvement in INDIGO

dCache.org

- dCache is mostly involved in WP4, which is about Virtual Infrastructures. (IaaS)

- For storage systems, like dCache, this essentially means SDS (Software Defined Storage), which according to Wikipedia is:

  - **Software-defined storage (SDS)** is an evolving concept for computer data storage software to manage policy-based provisioning and management of data storage independent of hardware.
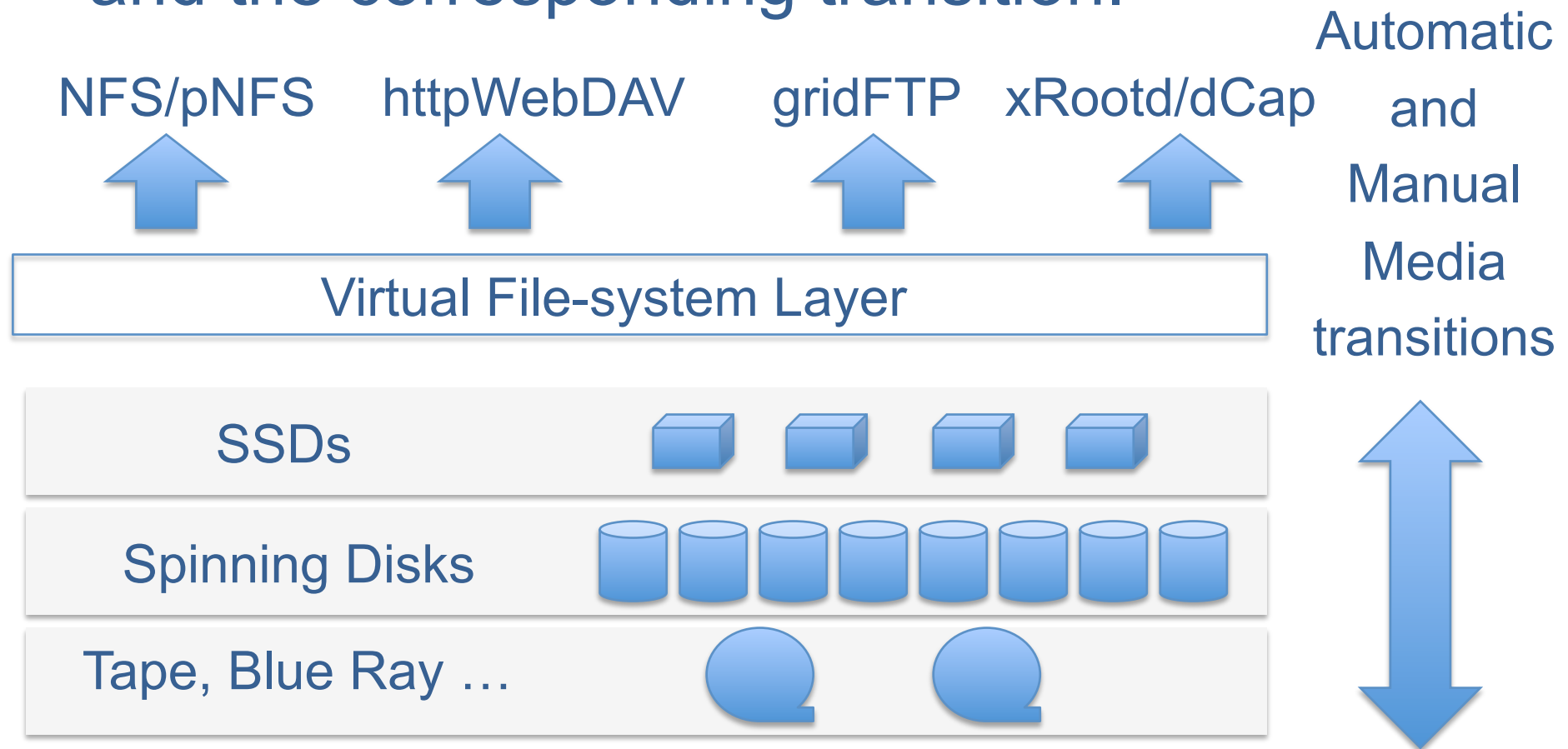
Software Defined Storage according to dCache

- User/PaaS defined "Quality of Service" management
  - User/PaaS defined "Access Latency"
    - SSD or Tape depending from application requirements.
  - User/PaaS Defined "Data Protection"
    - On one disk, two disks or tree tapes depending on how precious your data is.
  - User/PaaS Defined "Data Migration Policies"
    - Like Amazon Glacier vers. S3
- Automatic Storage-Tier migration
  - Based on access profile
- All this wouldn't be needed if SSD's would be cheap and 100 % reliable.

dCache is well prepared

dCache.org

Historically dCache supports multi-tier storage and the corresponding transition.

NFS/pNFS   httpWebDAV   gridFTP   xRootd/dCap

Automatic and Manual Media transitions

Virtual File-system Layer

SSDs

Spinning Disks

Tape, Blue Ray …

Work done in preparation for "Software defined Storage"

dCache.org

- Small Files for HSM

- Resilient Manager

- dCache CLOUD integration

- Work on CDMI

  - Not so much for data transfer but to manage data.

  - Could be a replacement for SRM

- Involvement in AAI (through LSDMA)

- dCache in a box

What's missing

- Mainly a common agreement (standard) on how to trigger transitions. (Protocol, API ??)

- We have some experience with SRM, however it seems not to be suitable for this purpose.

- Another candidate is CMDI (SNIA), which is an industry standard.

- Migration Policies are already discussed, documented and implemented within RDA (Practical Policy Working Group).
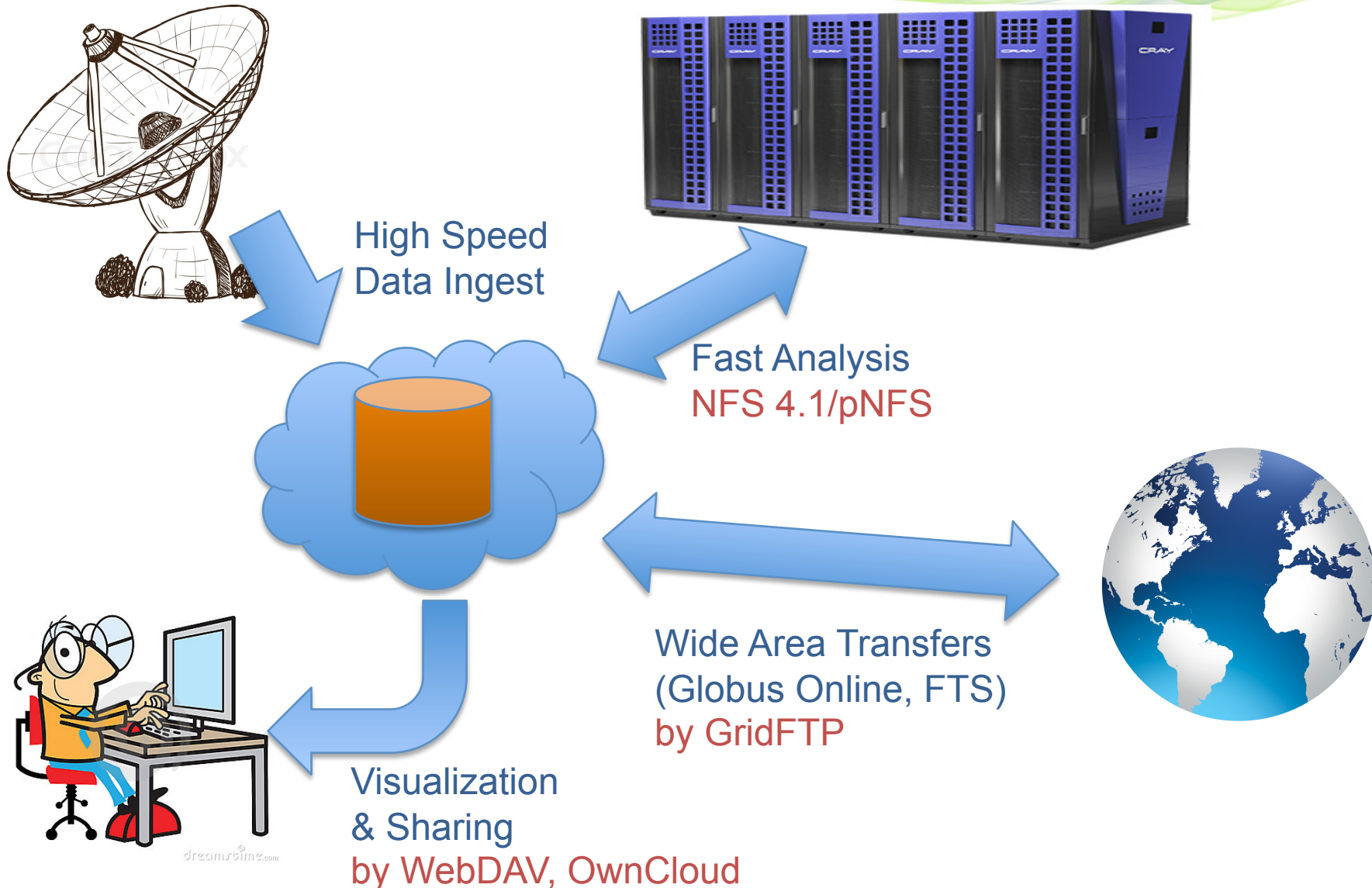
Summary

Magically, up to now, at the right moment, there was always an EU or National Project, funding dCache exactly for those features or activities, dCache was planning to do anyway and with that, they helped us following our master plan :

The support of the

Complete Scientific Big Data Life Cycle

Management.

Scientific Data Lifecycle

dCache.org

High Speed
Data Ingest

Fast Analysis
NFS 4.1/pNFS

Wide Area Transfers
(Globus Online, FTS)
by GridFTP

Visualization
& Sharing
by WebDAV, OwnCloud

The END

further reading
www.dCache.org