

dCache introduction

Paul Millar

On behalf of the dCache team.

EGI-EISCAT-3D ad-hoc meeting



High-level Overview



dCache is...

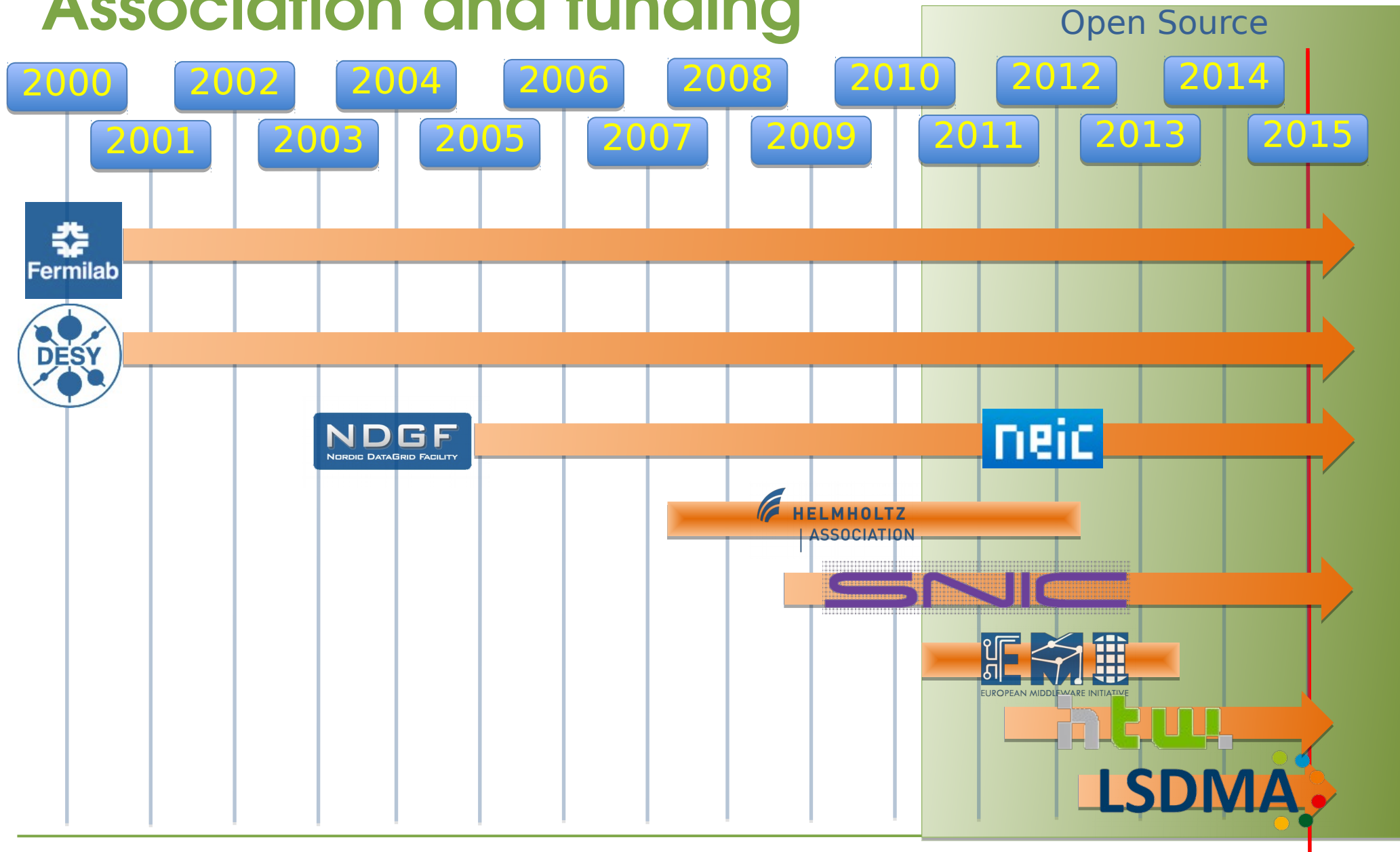
software for providing **scalable**, managed storage for huge amounts of data.

deployed at research institutes throughout the world and used by a diverse collection of user-communities.

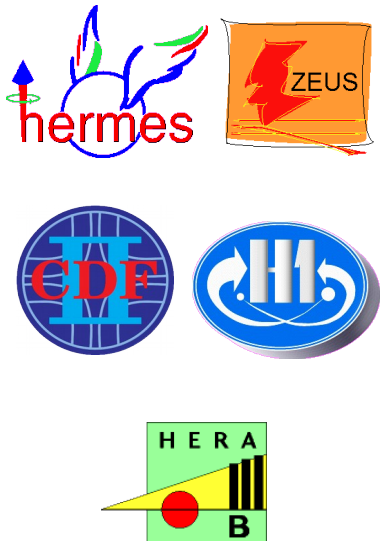



supported through the **dCache.org** collaboration, which provides:

- regular feature releases that are maintained with subsequent bug-fix releases.
 - Support and advice through a variety of channels.
-

Association and funding

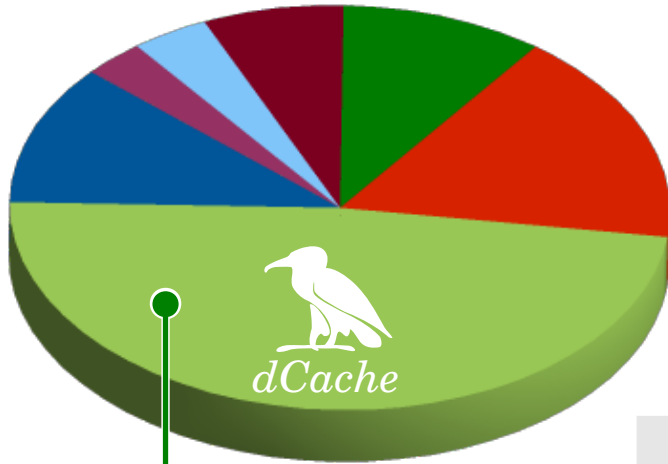


dCache history

Era	Disk cache	Grid Storage	Generic Storage	Cloud Storage
Additional Communities				
Additional Authentication	Trusted host	X.509, Kerberos	Username+PW	SAML, OpenID, OAuth, Token, ...

What is dCache today?


LHC data stored on each storage system




- dCache (96 PB)
- DPM (34 PB)
- EOS (0 PB)
- StoRM (20 PB)
- CASTOR (14 PB)
- BeStMan (7.6 PB)
- Globus FTP (6.1 PB)
- ARC (0.01 PB)
- xrootd (22 PB)

Source: BDII (2014-11-14)







8 FTEs



2 FTEs



1.5 FTEs




Core team


Student mentor programme

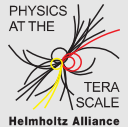



Hochschule für Technik und Wirtschaft Berlin
3 students


Collaborations












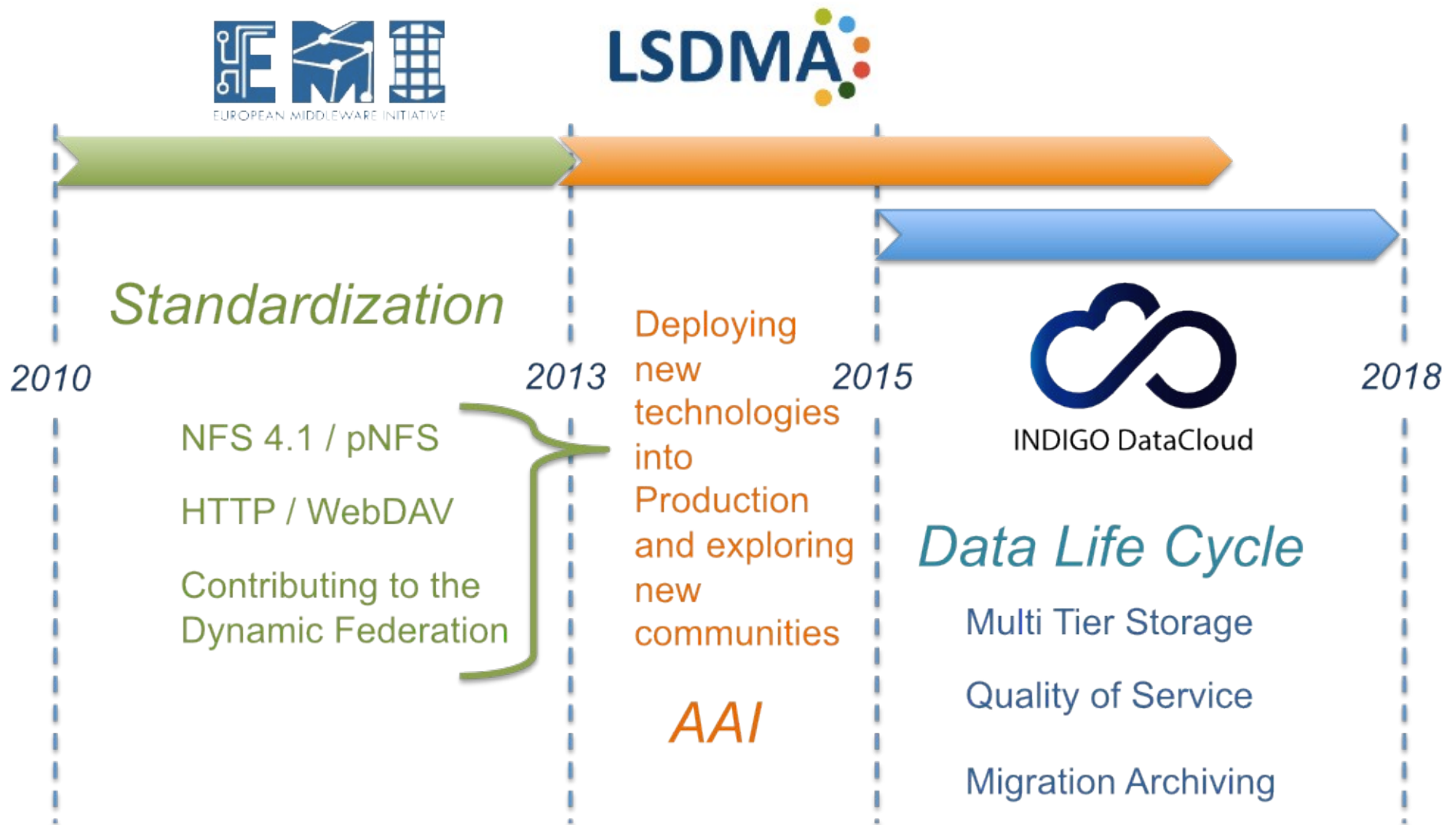








Current and future project funding

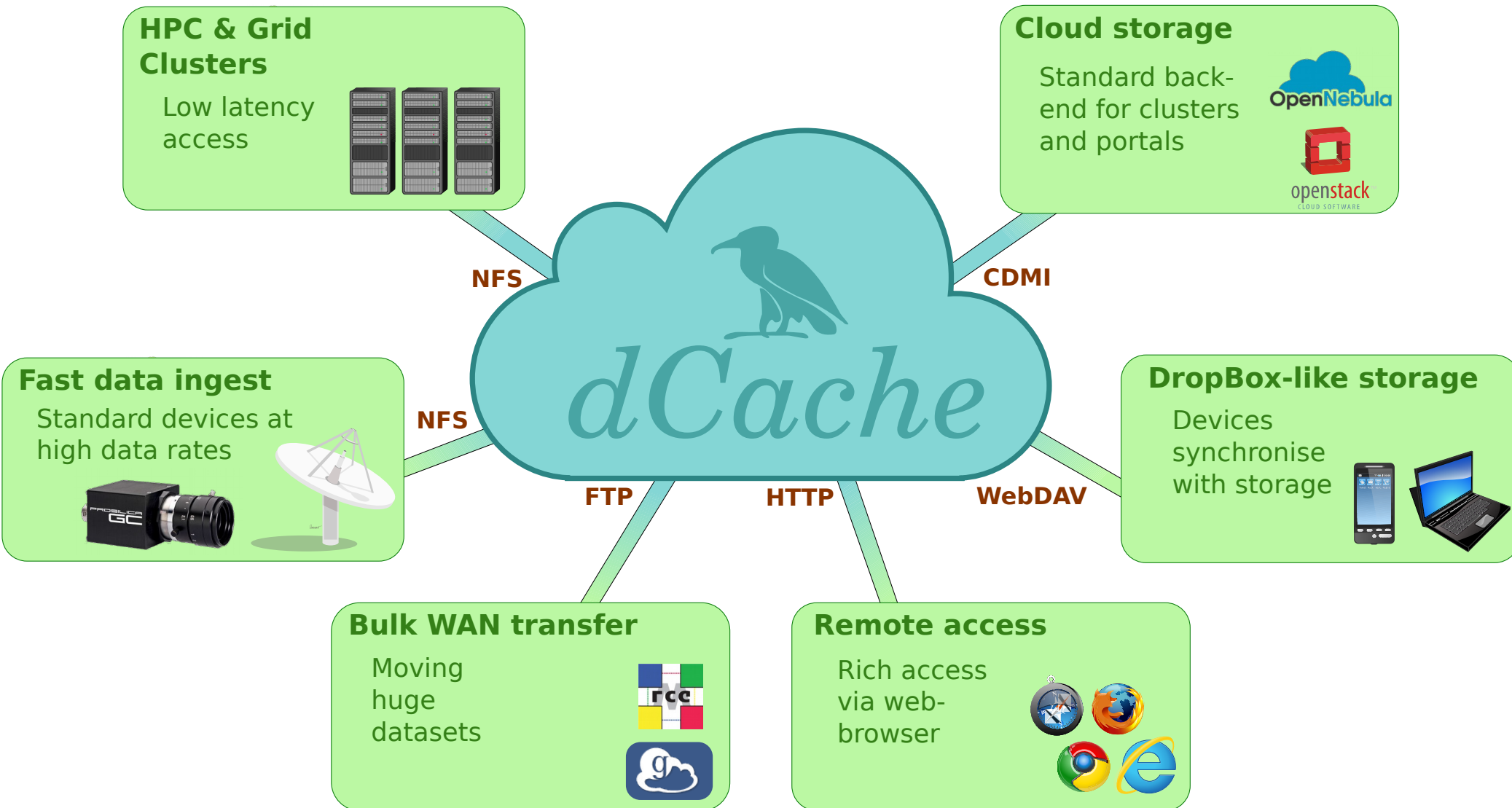


dCache key features include...

- Users see a single POSIX filesystem (hard- & soft-links, etc),
- Transparent support for tertiary (tape) storage,
- Scalable bandwidth,
- Steerable target when reading and writing,
- Space management,
- Resilience to storage node failure,
- Supports transparent storage device life-cycle,
- Hot-spot detection and mitigation,
- Differentiable quality of service,
- Pluggable authentication,

...

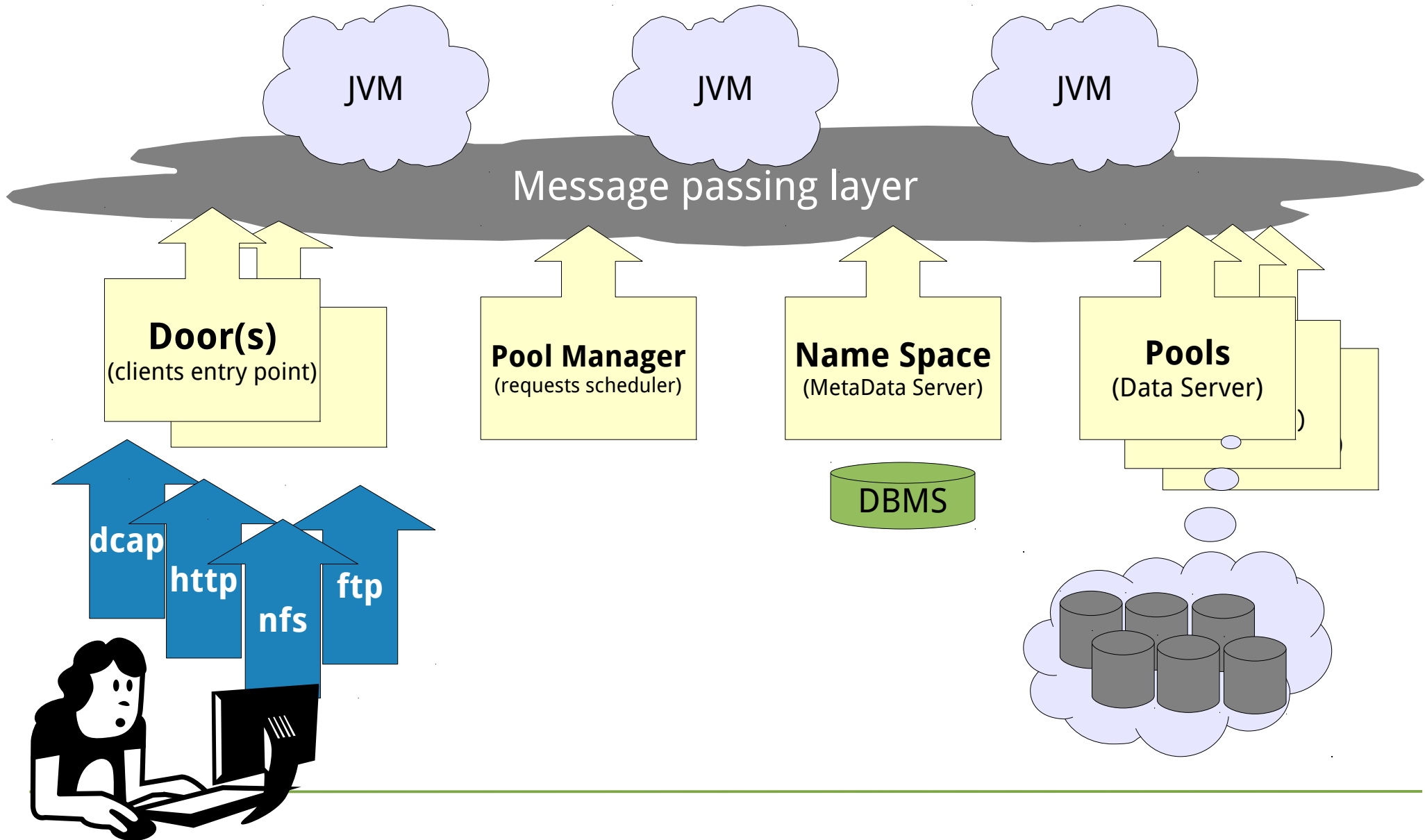
The scientific cloud vision



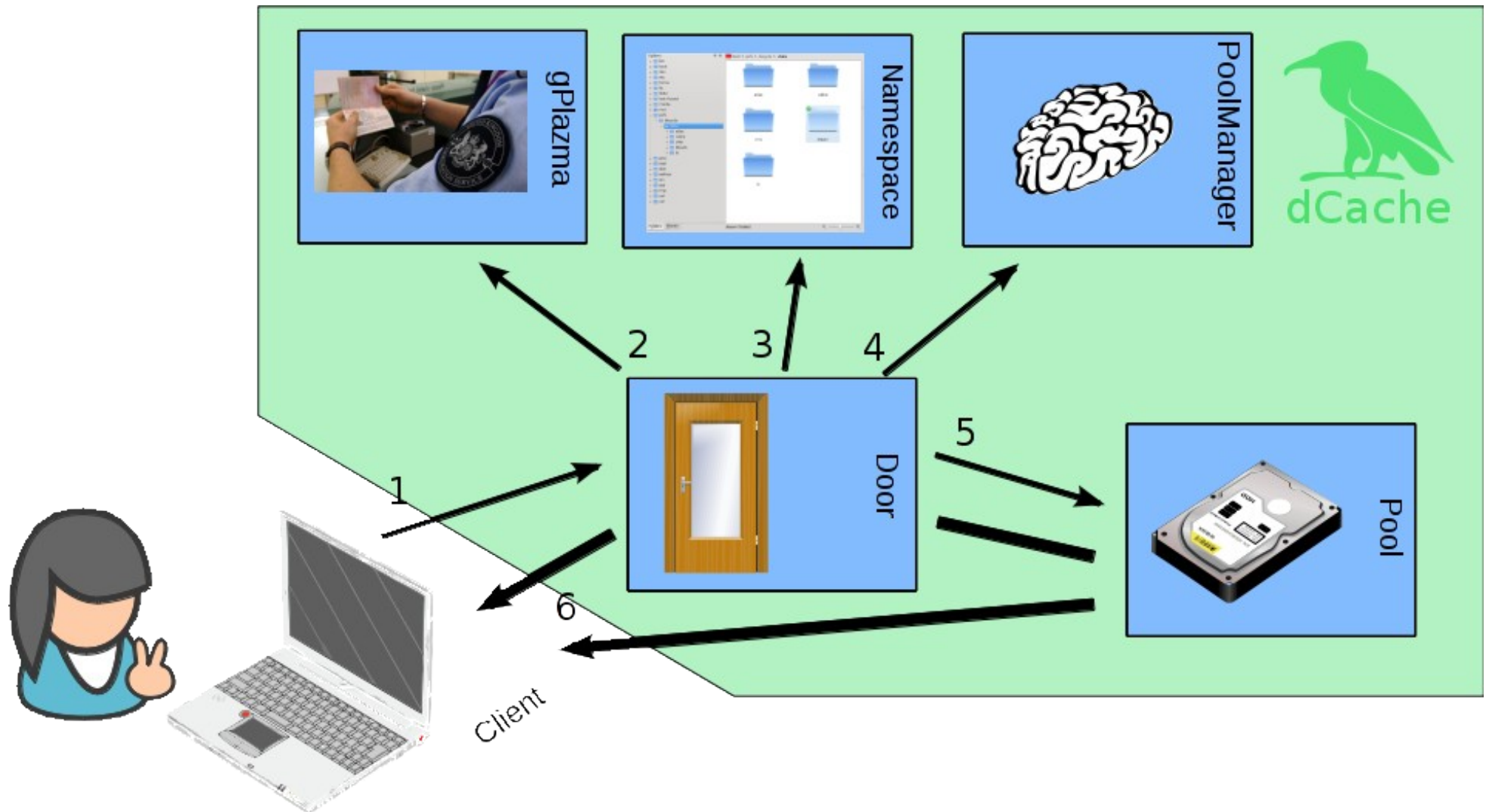
Some details* on how dCache operates...

* Some details are deliberately omitted to keep slides manageable.

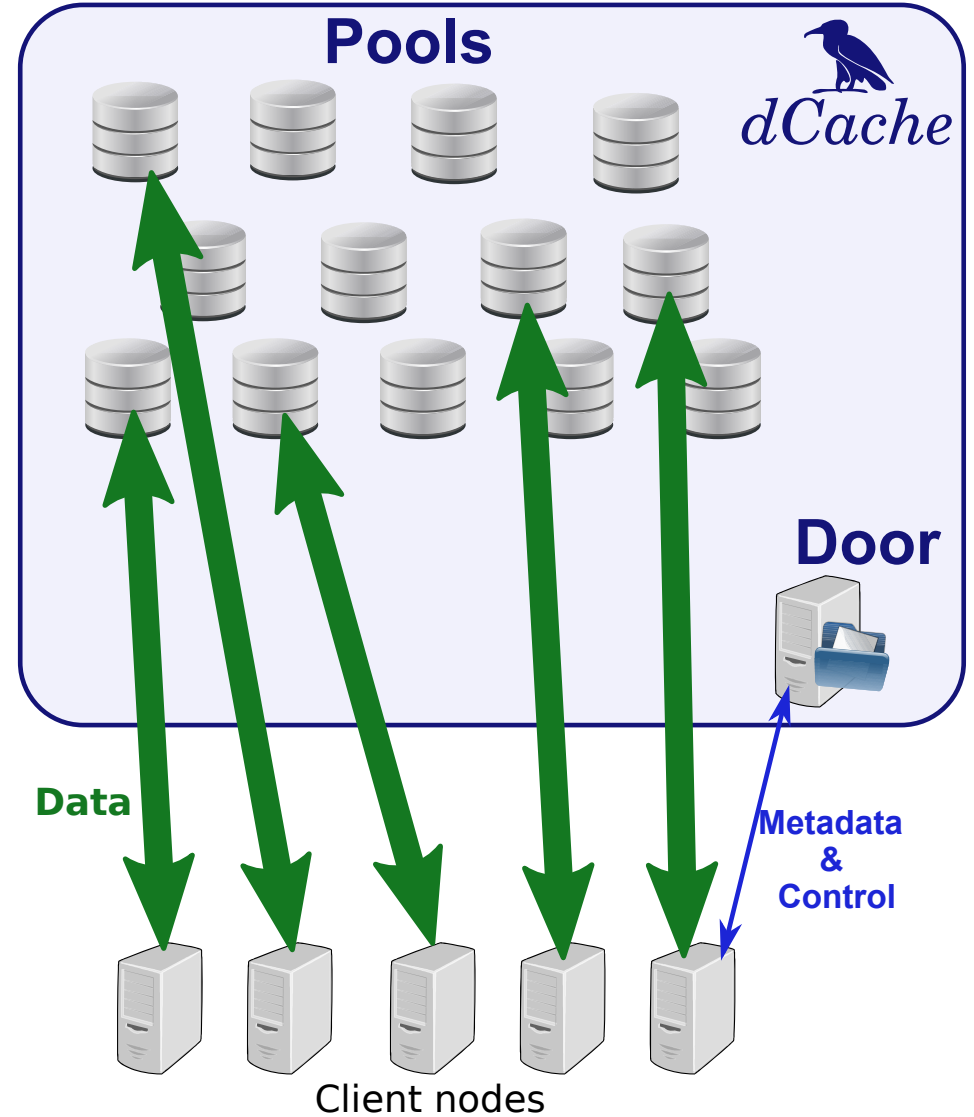
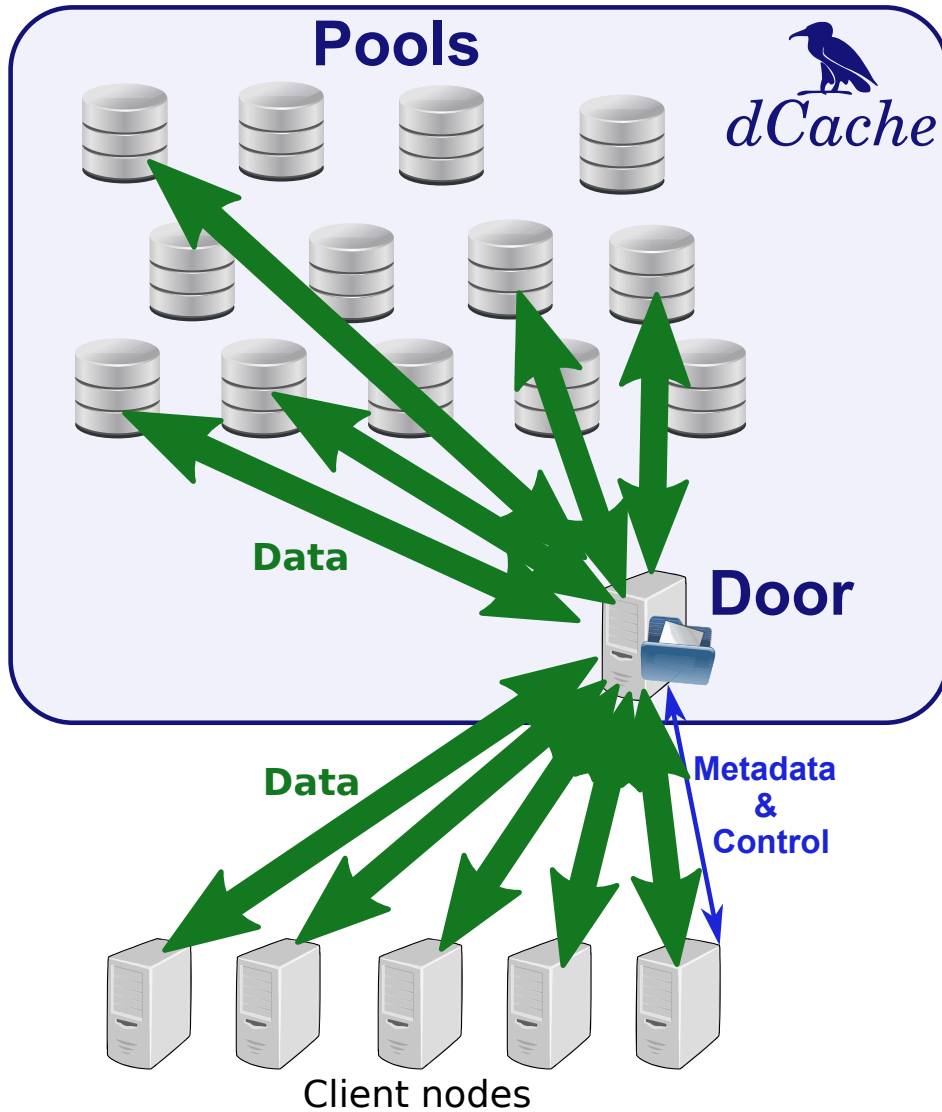
dCache – under the hood

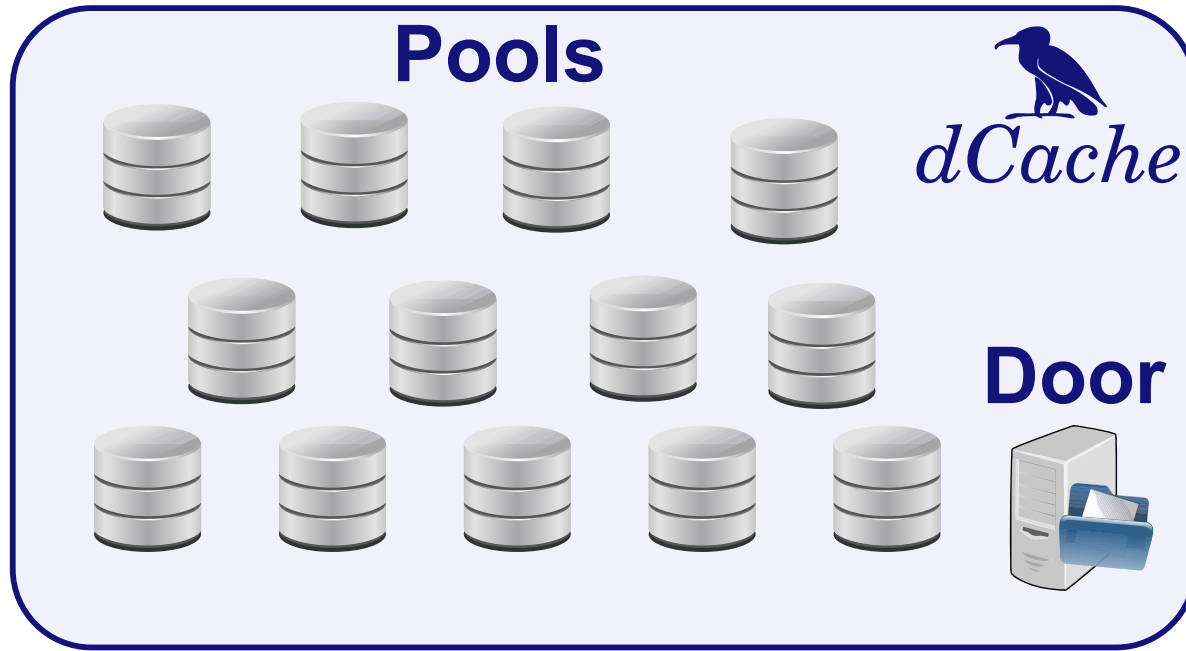


Core components when transferring

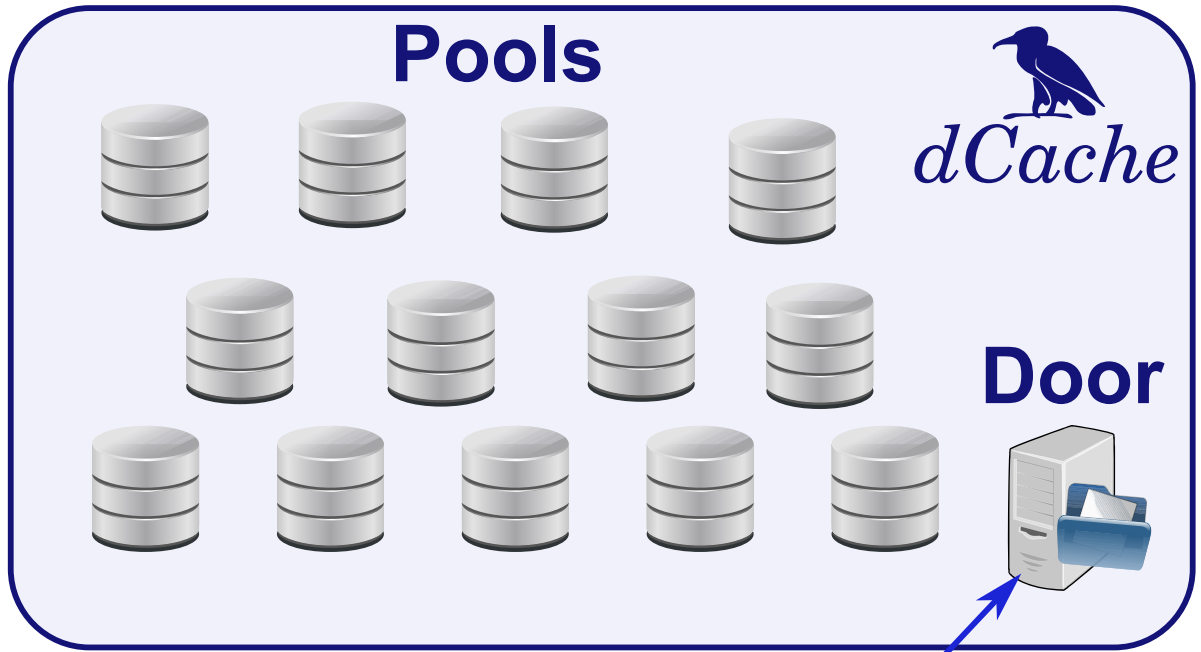


Importance of redirection





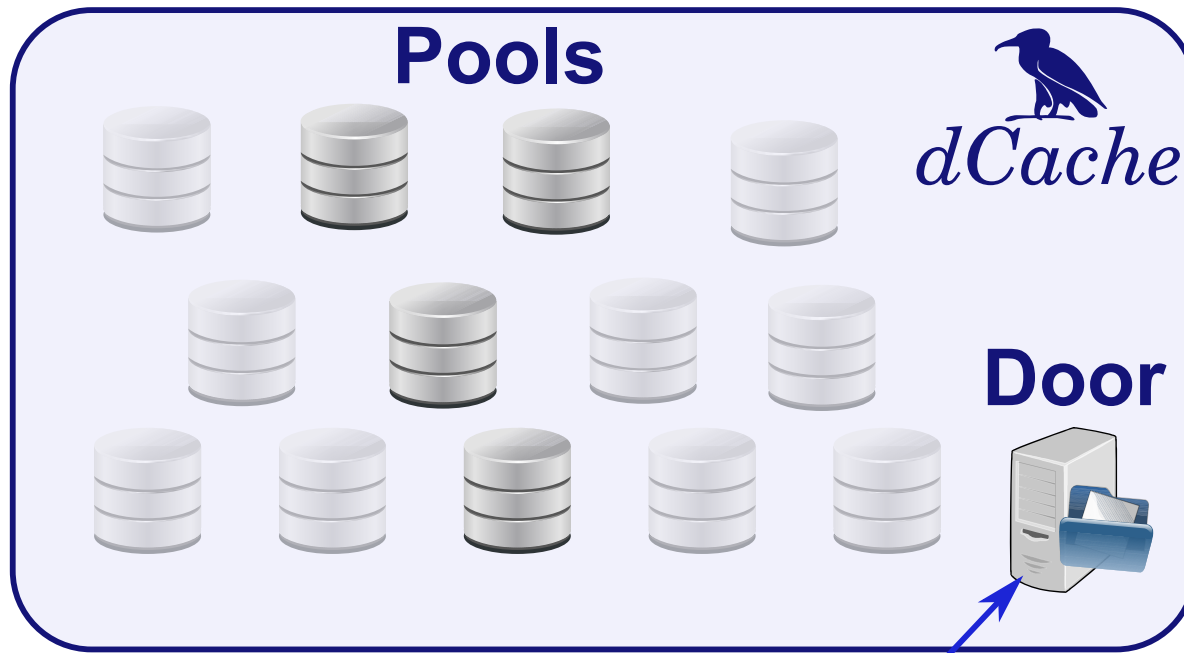
Client node



**Metadata
&
Control**



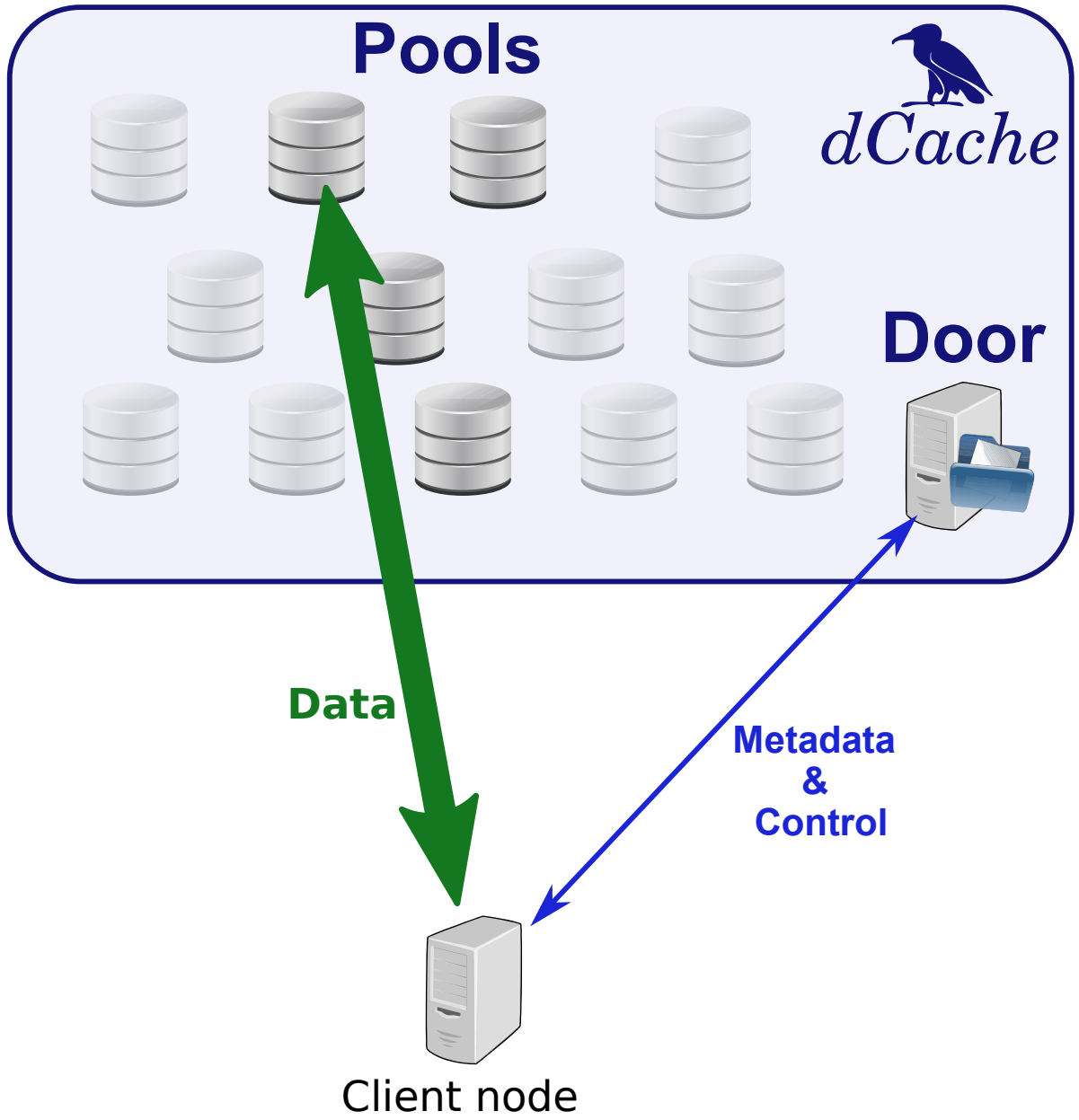
Client node



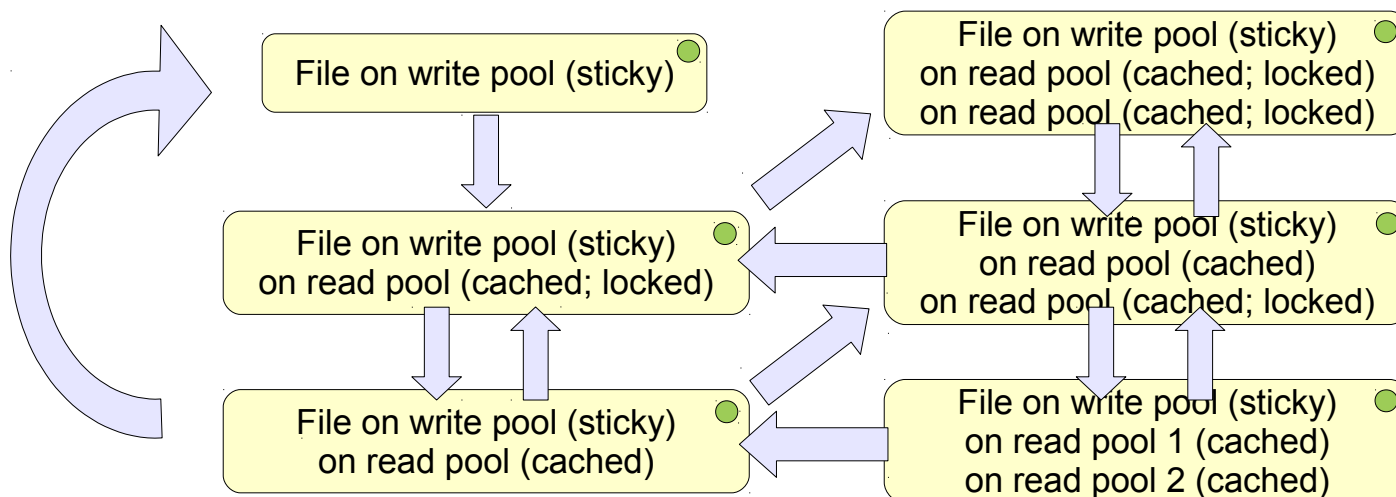
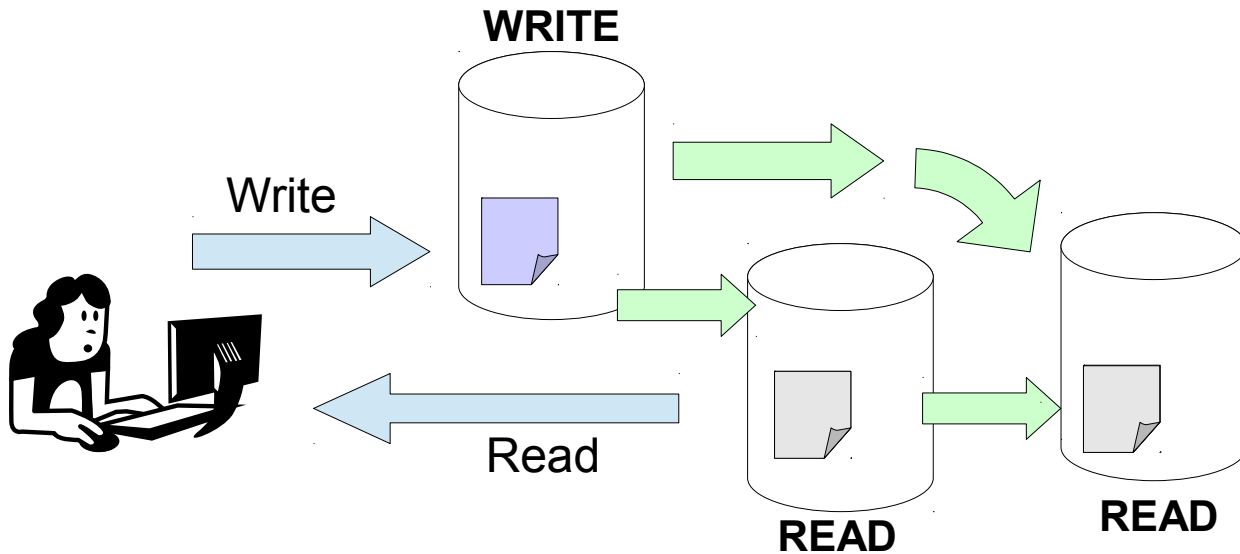
**Metadata
&
Control**



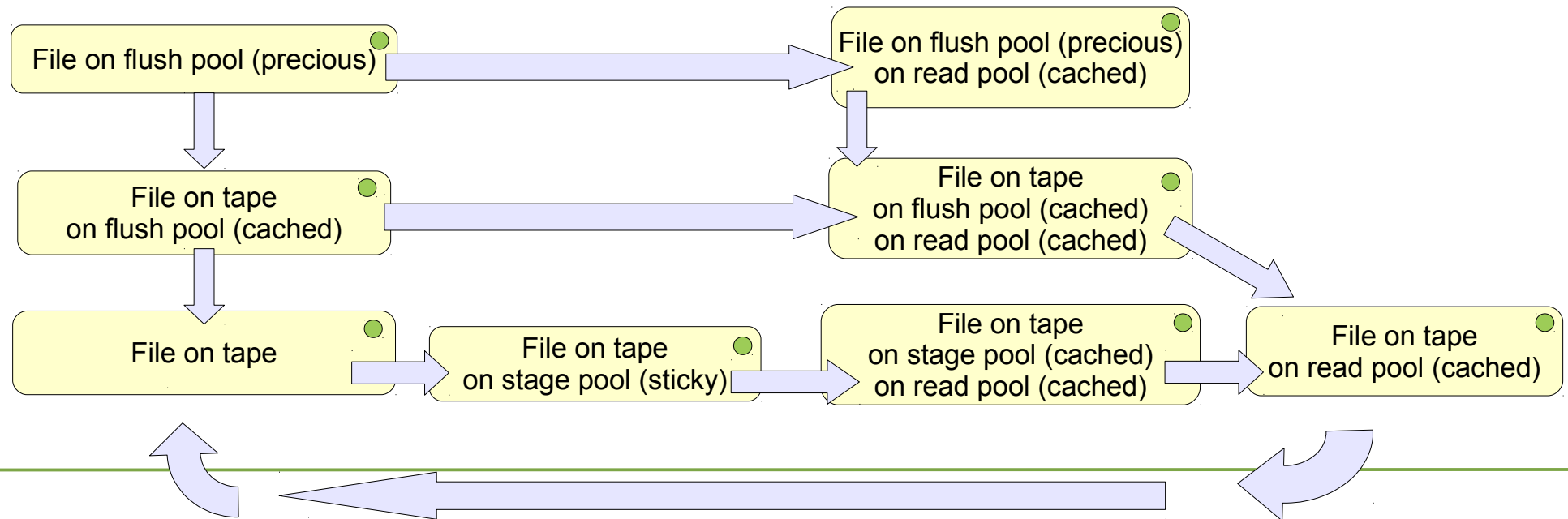
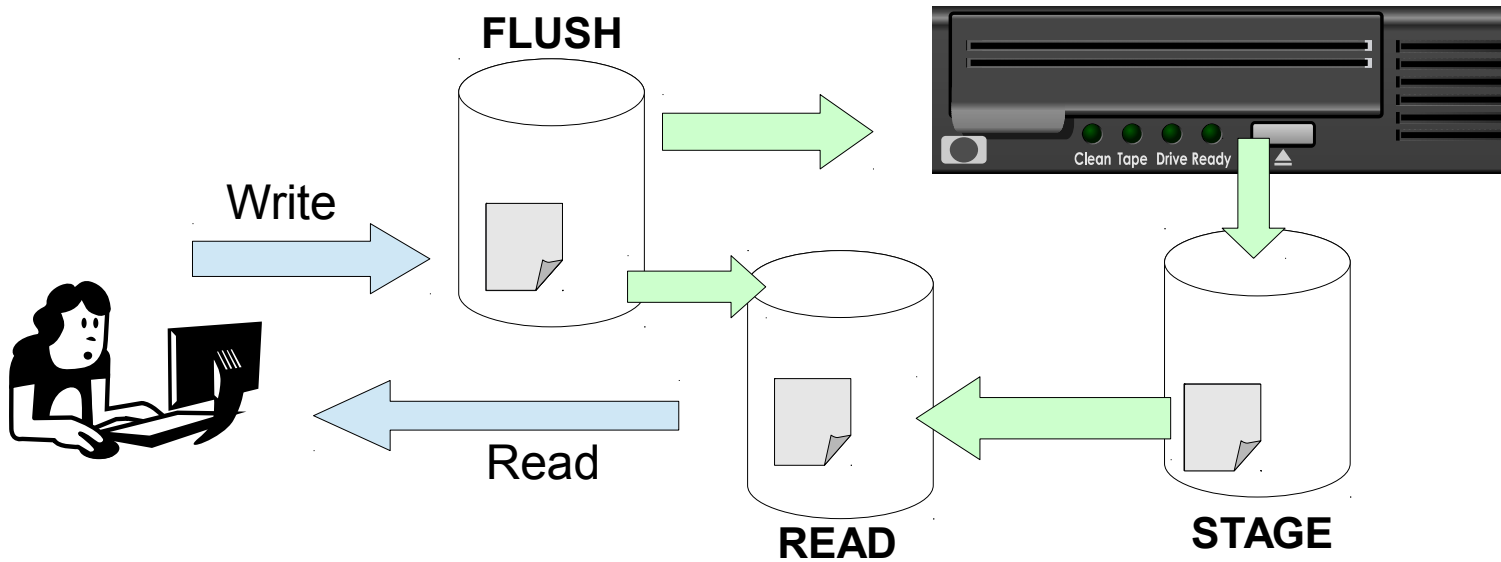
Client node



Guaranteeing QoS for write



Guaranteeing QoS for tape activity



Operational experience



Storage at DESY

- 6 dCache instances: Hera, CMS, ATLAS, Photon, “DESY” and Cloud:

Hera is officially switched off,

CMS & ATLAS for WLCG experiments,

Photon is for various photon user-communities,

Cloud is for sync-and-share service,

DESY is for the rest.

Comparative numbers

CMS	ATLAS	Photon	DESY	Cloud
$\sim 5 \times 10^6$ files	$\sim 1 \times 10^7$ files	$\sim 8 \times 10^7$ files *	$\sim 1 \times 10^7$ files	$\sim 2 \times 10^6$ files
~ 3 PiB	~ 3 PiB	~ 2.5 PiB *	~ 3 PiB	~ 10 TiB
~ 300 pool-nodes	~ 300 pool-nodes	~ 30 pool-nodes	~ 30 pool-nodes	~ 6 pool-nodes
~ 580 GiB/s ‡		~ 200 GiB/s ‡	~ 12 GiB/s ‡	~ 3 GiB/s ‡
~ 400 Hz (read)†		~ 180 Hz (write)†	~ 200 Hz (read)†	

* Photon instance accepts ~ 1 TiB per month as $\sim 1 \times 10^7$ files.

‡ Value is peak observed bandwidth aggregate over all clients within last 7 days.

† Value is peak observed open rate (either read rate or write rate) observed within last 7 days.

Other dCache instances

NT1	US-CMS T1	BNL	SARA
~5x10 ⁷ files			
~6.3 PiB (2.1 PiB tape; 4.2 PiB disk)	~20 PiB (disk)	~15 PiB (disk)	~6.2 PiB (disk)

Backup slides

