dCache, the Update

Patrick Fuhrmann

for the dCache people

The Team

Responsibility, dCache

Patrick Fuhrmann

Rob Kennedy

Responsibility, SRM

Timur Perelmutov

Core Team (Desy and Fermi)

Jon Bakken

Mathias de Riese

Micheal Ernst

Alex Kulyavtsev

Birgit Lewendel

Dmitri Litvintsev

Tigran Mrktchyan

Martin Radicke

Neha Sharma

Vladimir Podstavkov

External Development

Nicolo Fioretti, BARI

Abhishek Singh Rana, SDSC

Support and Help

Maarten Lithmaath, CERN

Owen Synge, RAL

dCache, the Idea

The story so far ...

what's new , what's coming next

Feature Track

SRM V2

Replica Manager

Information Provider

File Hopping

Smart Tape System flushing

Generic Tertiary Storage Interface

dCache partitioning

VOMS integration

Performance Track

medium term strategy

Multiple I/O queues

Multiple Pnfs (file systems)

Chimera

Distributed Peta Byte Disk Store with single rooted filesystem providing posix like and wide area access protocols.

Distributed cache system to optimize access to Tertiary Storage Systems

Grid Storage Element coming with standard data access protocols, Information Provider Protocols and Storage Resource Manager.

The story so far ...

What's new ?

What's coming next ?

Feature Track

Please refer to talk by Alexander Kulyavtsev

Resilient dCache

- Controls number of copies for each dataset in dCache
- Makes sure n < copies < m
- Adjusts replica count on pool failures
- Adjusts replica count on scheduled pool maintenance
- Makes use of local disk space when running on farm nodes
- doesn't work with HSM back-end yet

Please refer to talk by Timur Perelmutov

Storage Resource Manager V2

- Directory methods/function available

- Explicit Space Management is in developmenet phase

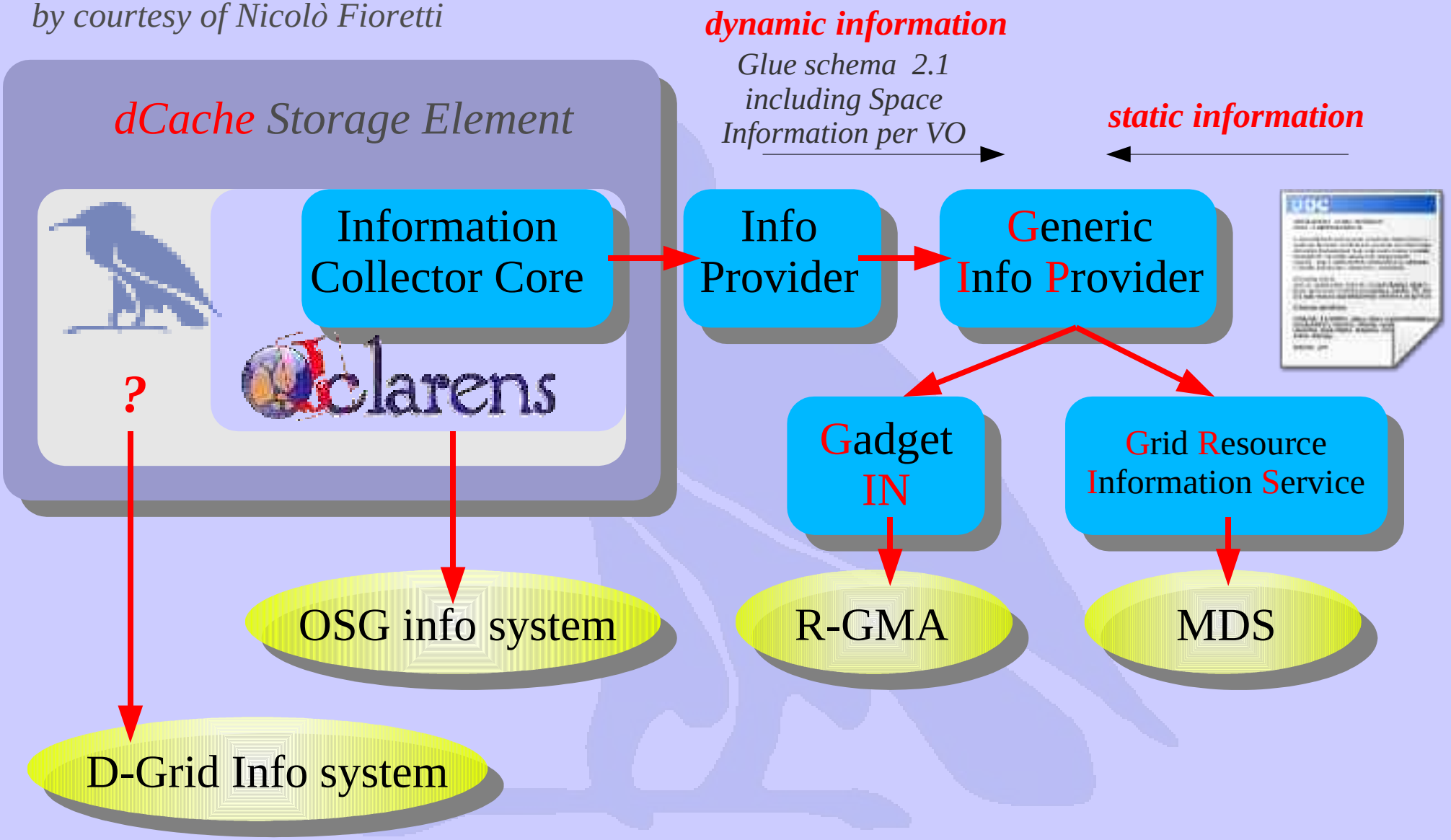
Please refer to talk by Abhishek Singh Rana

gPLAZMA (grid-aware PLuggable AuthoriZation MAnagement): Introducing RBAC (Role Based Access Control) Security in dCache

Virtual Organization Membership Service Integration

- Pluggable authorization
- Backward compatible to current system.
- Storage Authorization Call-outs for SRM and GridFtp

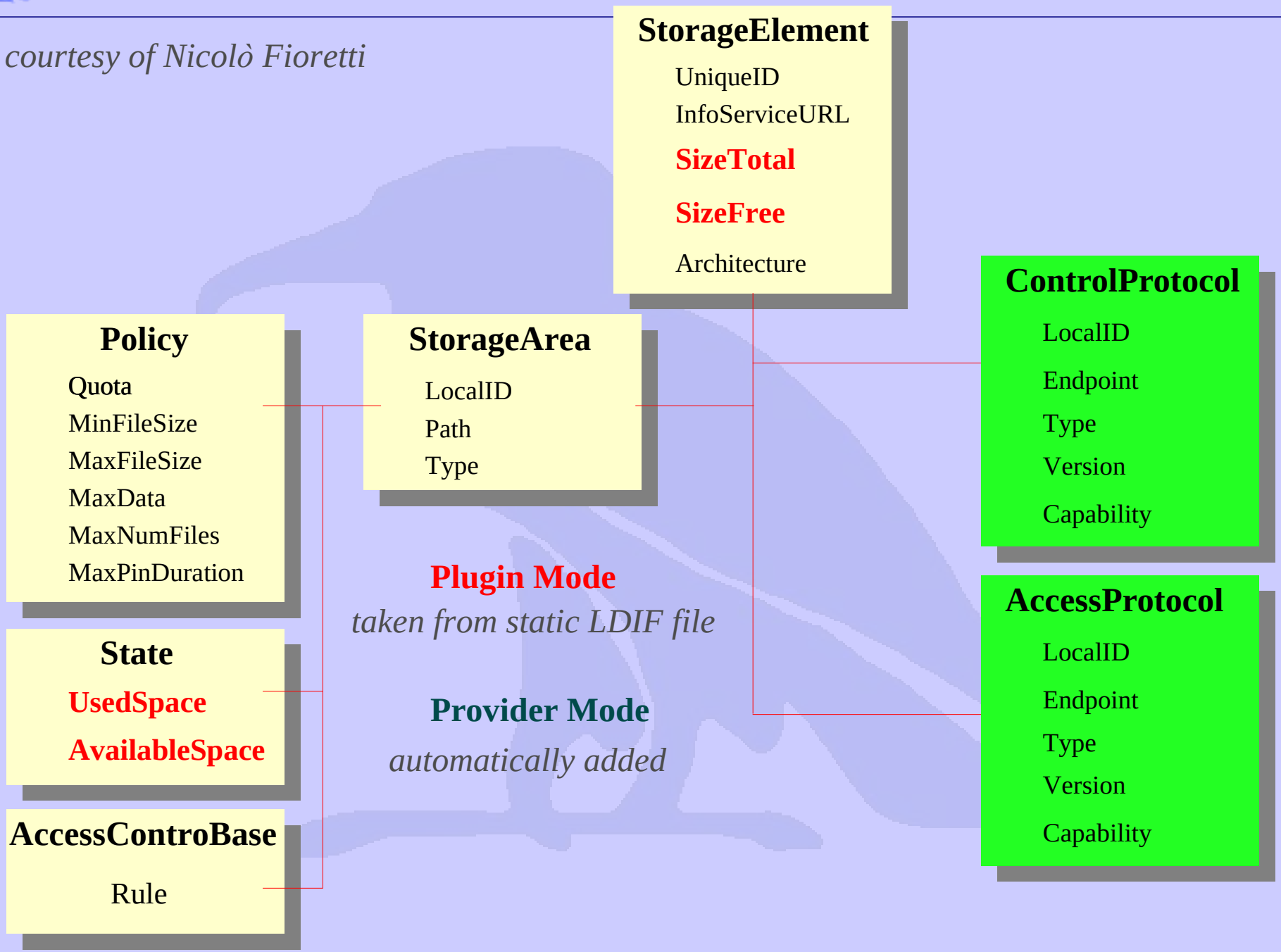
by courtesy of Nicolò Fioretti



with many thanks to Jean-Philippe Baud and Lawrence Fields

Plugin and Provider mode

by courtesy of Nicolò Fioretti



Collect and flush

Write Only Cache

From Client

Forward on
read
or on
arrival

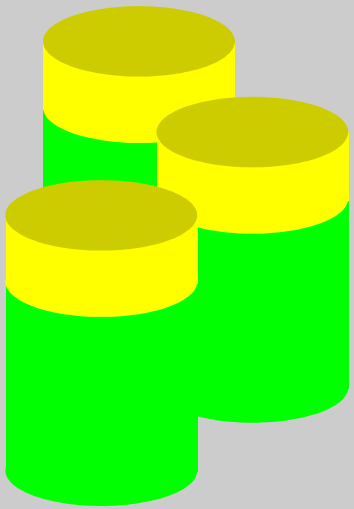
Read Only Cache

Replicate
on high load

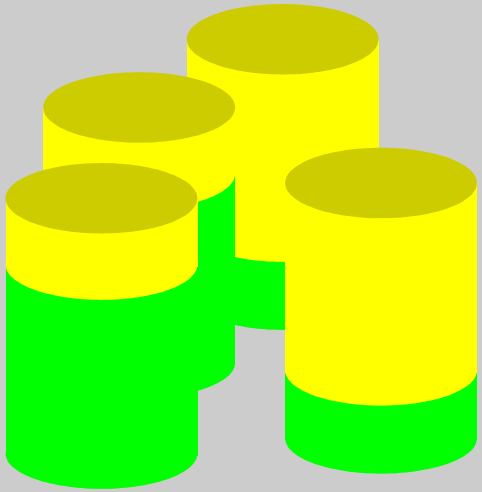
To Client

even more dynamic when using API

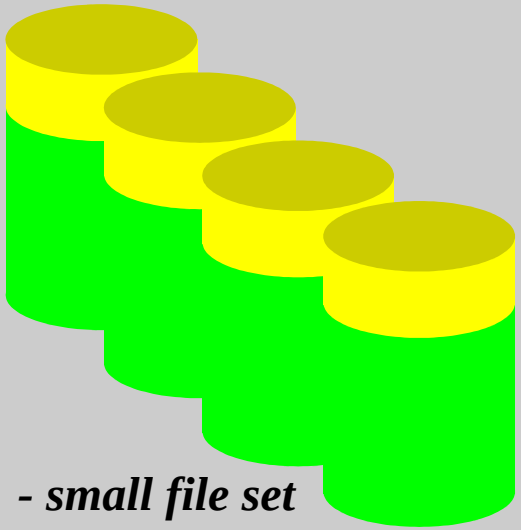
One dCache instance, but pool-groups with different characteristics



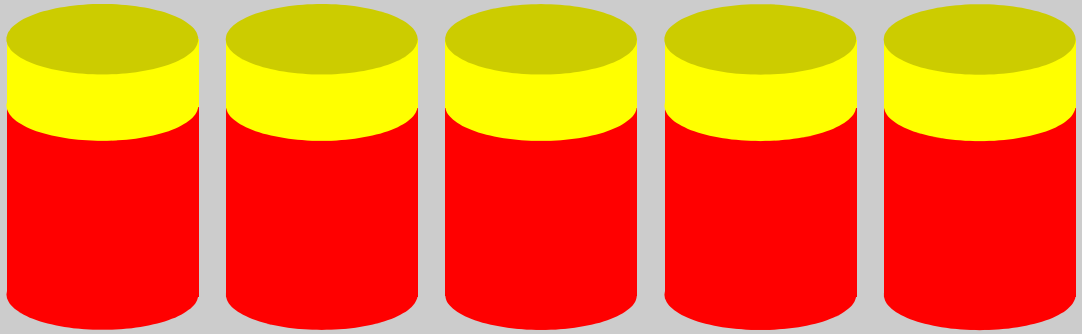
prefer balanced movers



prefer filling empty pools first



- small file set
- high throughput
- all files on all pools



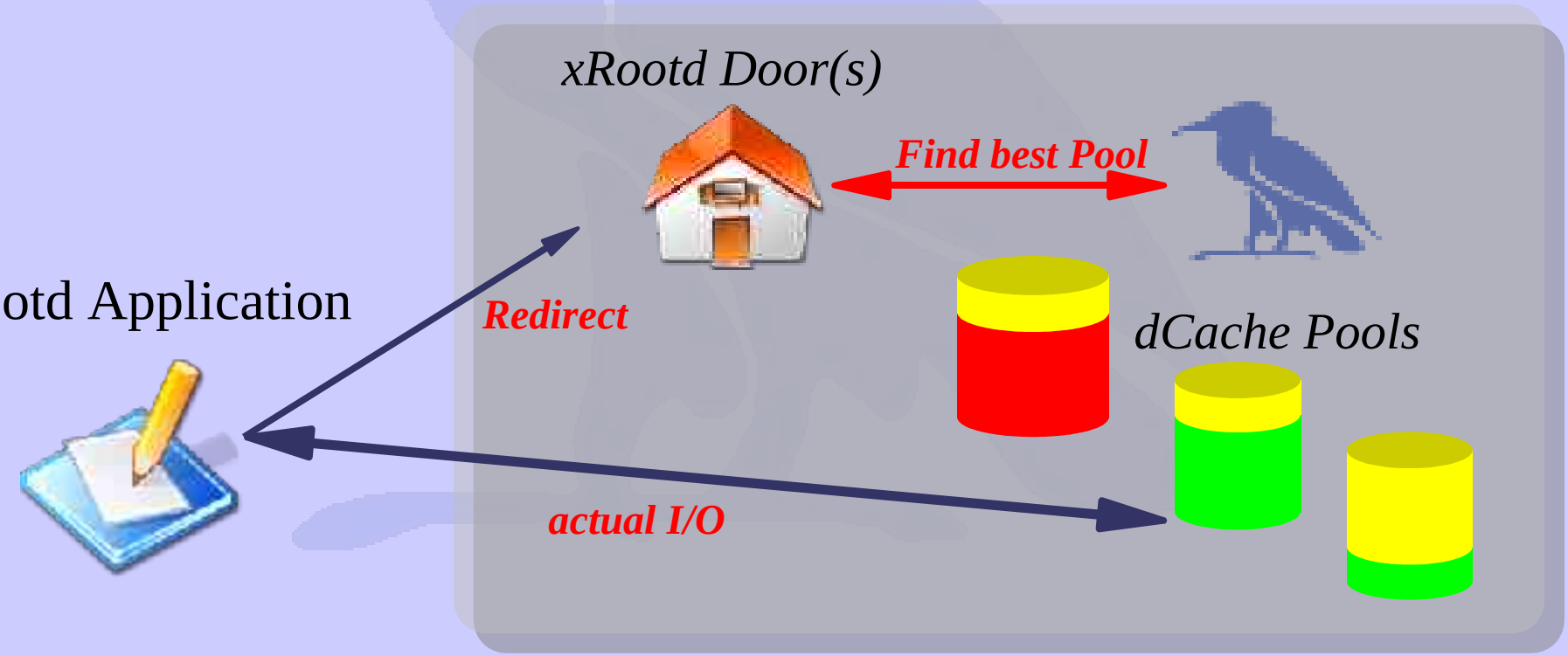
Resilient Pool Group,
file multiplicity centrally managed,
no HSM back end

Centrally controlled flushing of precious datasets to back-end tape system(s)

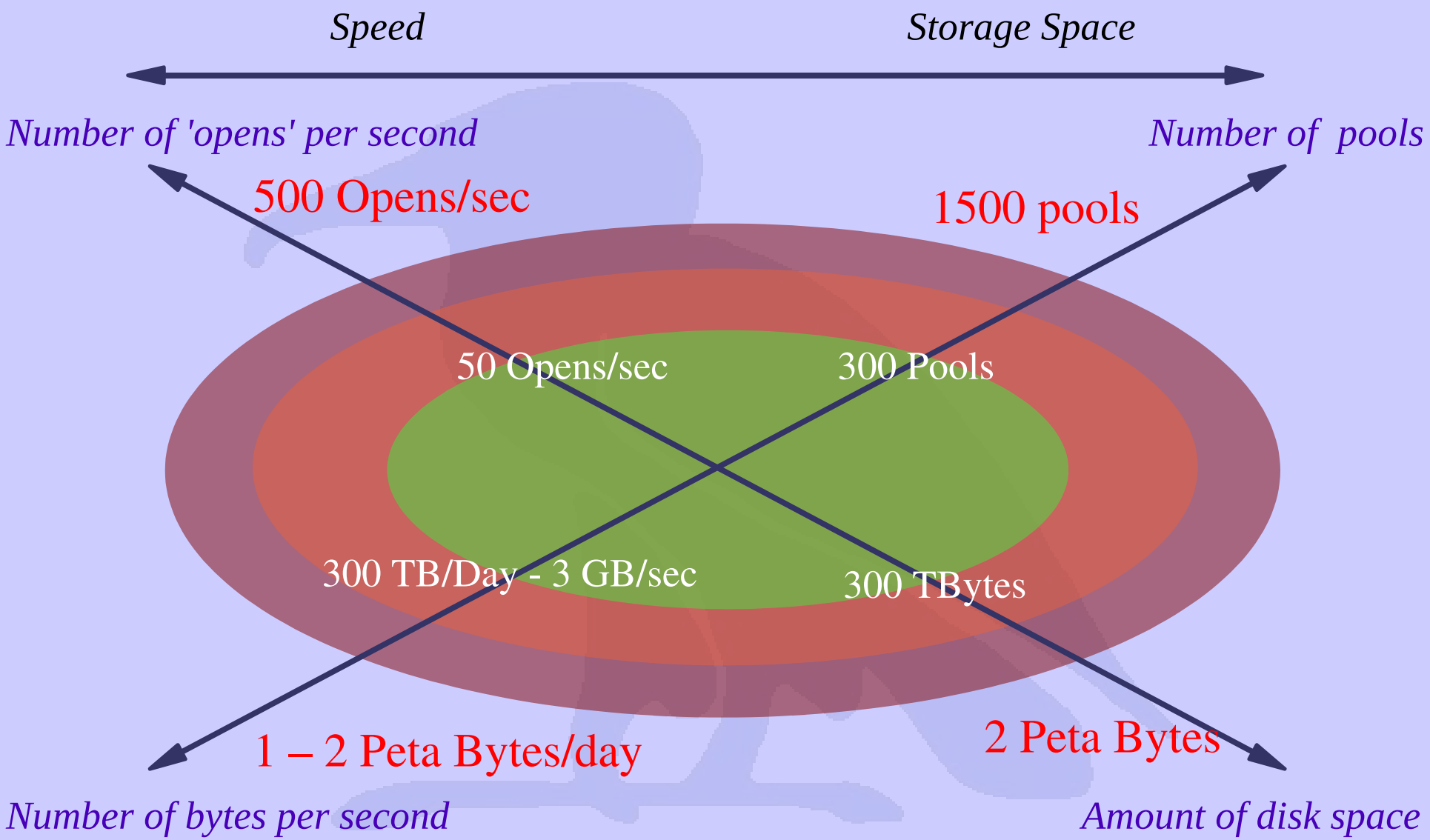
- determines number of concurrent streams to HSM system.
- never allow client data flowing onto pools, while same pool is flushing.
- never allow pool flushing, while data coming in from clients.
- event driven API for fine grained customization of HSM flush.

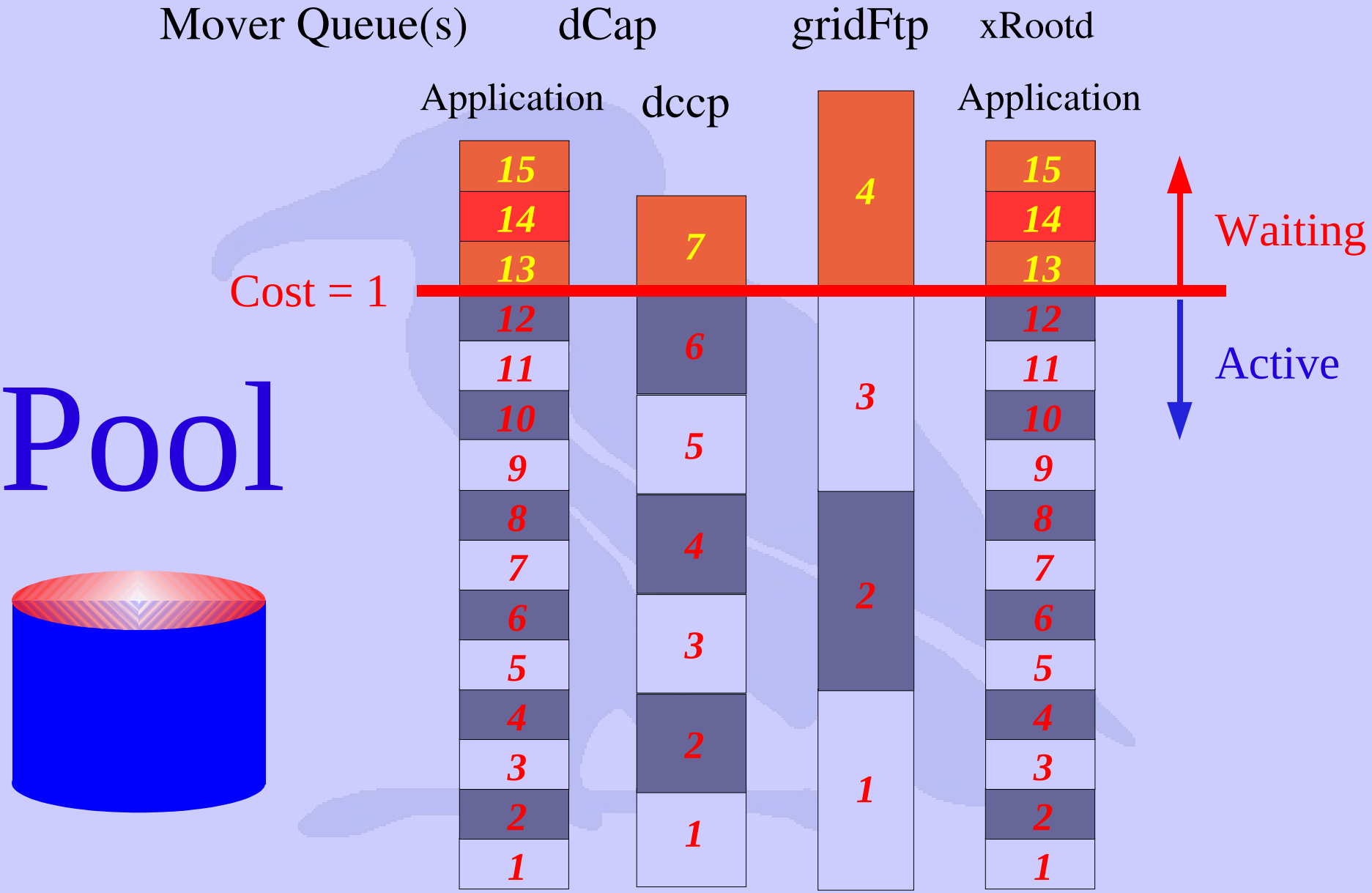
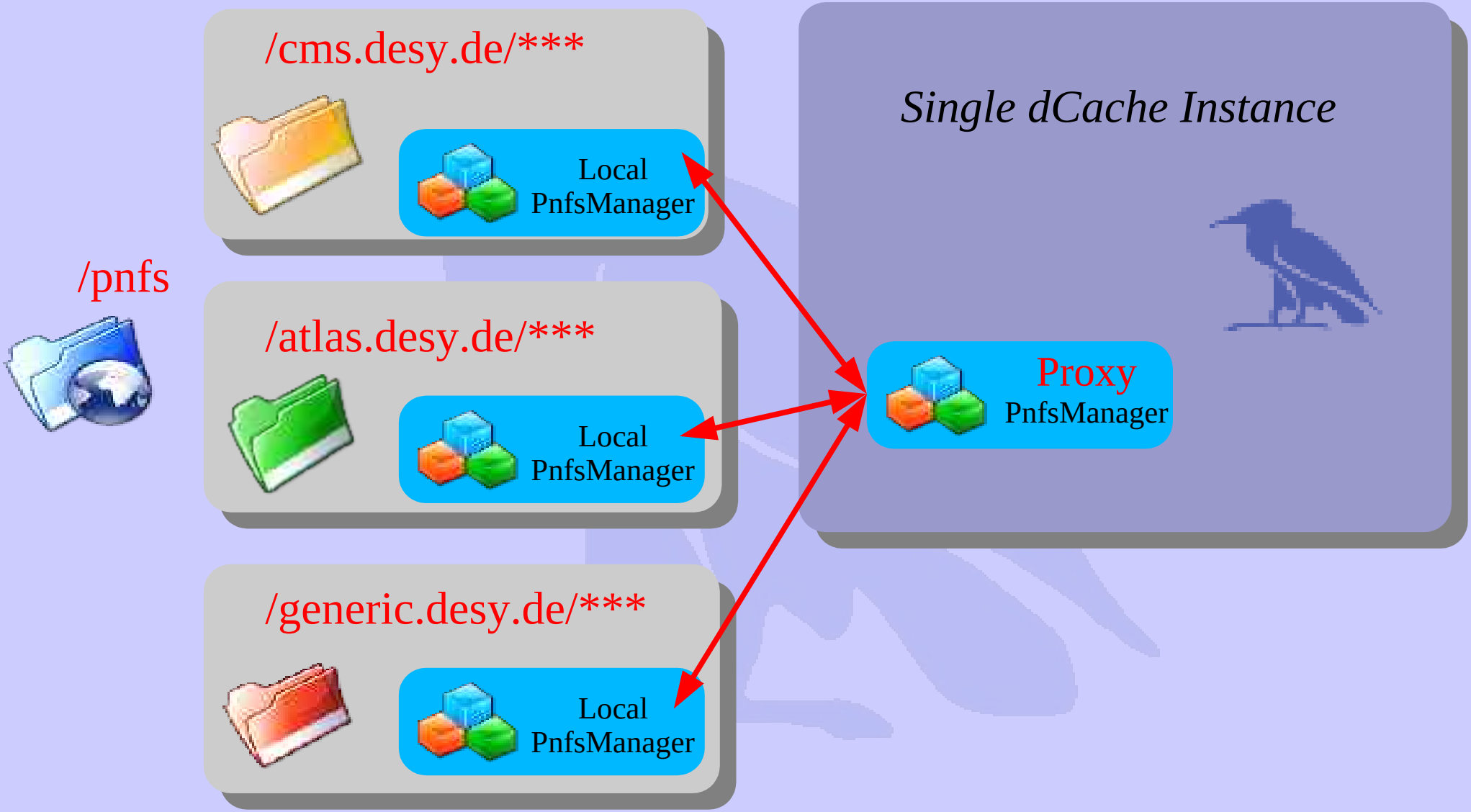
xrootd implemented :

- *regular I/O ok, including forwarding*
- *no asynchronous mode yet*
- *no security yet*
- *we need help from experts, specs sometimes not clear enough*



Performance Track


Speed
Storage Space
Number of 'opens' per second
Number of pools
500 Opens/sec
1500 pools
50 Opens/sec
300 Pools
300 TB/Day - 3 GB/sec
300 TBytes
1 – 2 Peta Bytes/day
2 Peta Bytes
Number of bytes per second
Amount of disk space


Mover Queue(s)
dCap
gridFtp
xRootd
Application
dccp
Application
Cost = 1
Waiting
Active
Pool
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
7
6
5
4
3
2
1
4
3
2
1

Separate Pnfs Server