

Overview of dCache Systems at BNL

Iris Wu

Hiro Ito

Jane Liu

dCache workshop 2018
DESY, Hamburg Germany

70 YEARS OF
DISCOVERY

A CENTURY OF SERVICE



Outline

- Overview
- Future operation
- Needs help from developers

ATLAS dCache

Largest ATLAS T1 site

Core server 3.0.11, NFS: 3.0.38,
pool:3.043

17.5 PB

TAPE backend

Belle II dCache

Belle II T1 site

Version 3.0.11

1.7 PB disk space

TAPE backend

Simon's dCache

Version 2.1.6

0.26PB disk space

PHENIX dCache

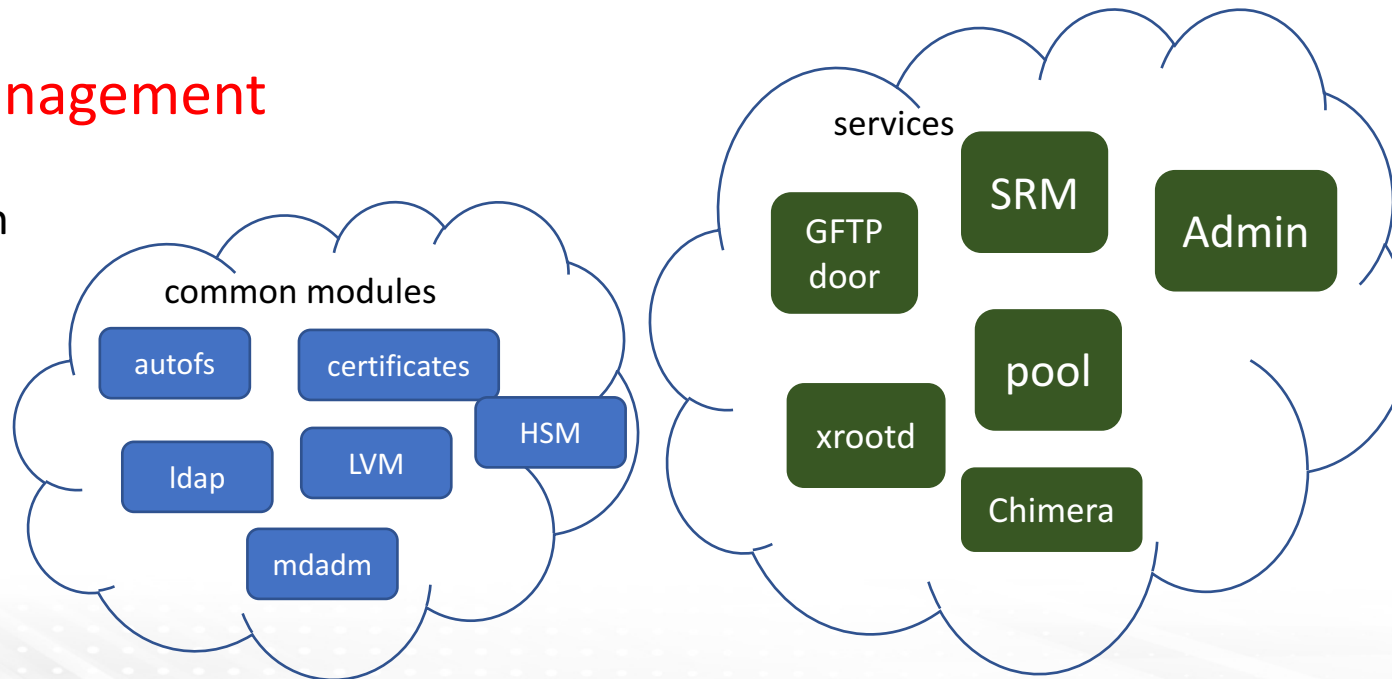
PHENIX experiment T0 site

In the progress of upgrade

TAPE backend

Puppet Management

- Efficient
- Automation



```
atlasdcache::pool::data:  
dc016:  
1:  
  diskspace: 51938544204217  
  lan: 1000  
  wan: 10  
  p2p: 20  
  pp: 16  
  rh: 1000  
  st: 2  
  checksum: ADLER32  
  type: disk  
  tags: CDCE  
  disk: data  
  device: /dev/md0
```

NFS4.1/pNFS

dCache

DAV
HTTP/WebDav



Globus Online
GFTP

XRootD

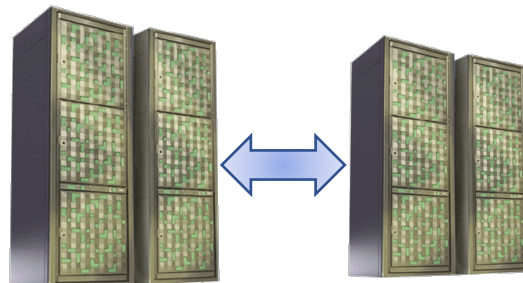


Admin nodes



Data transfer nodes

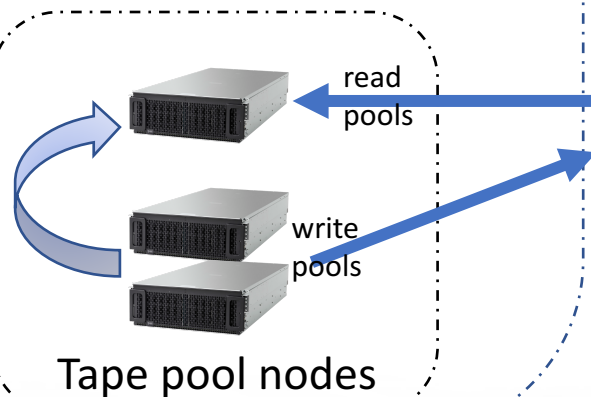
Disk pool nodes



primary

secondary

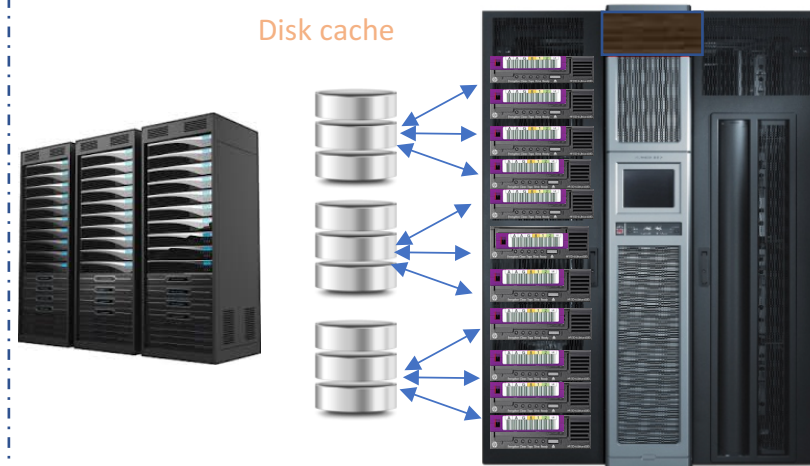
Resilience Management



US ATLAS Tier1 dCache @BNL

HPSS
High Performance Storage System

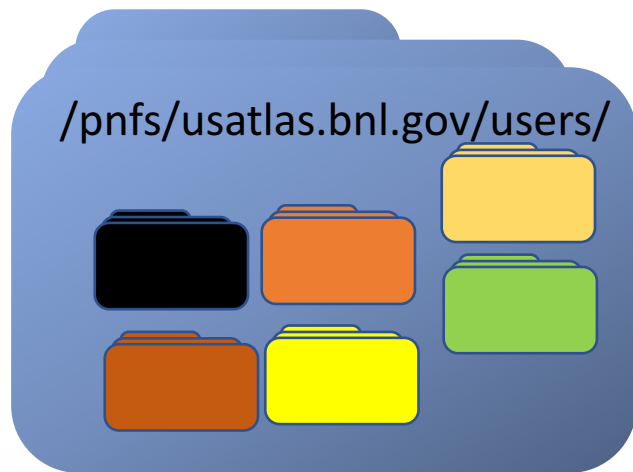
Disk cache



Protocols

- NFS 4.1

- Mounted in the Linux farm
- US ATLAS Tier3 facility



- XRootD

- Read is in production
- Write is tested through
 - GSI authentication

Support of XRootD third party transfer?

Protocols



dCache Native Gridftp Server

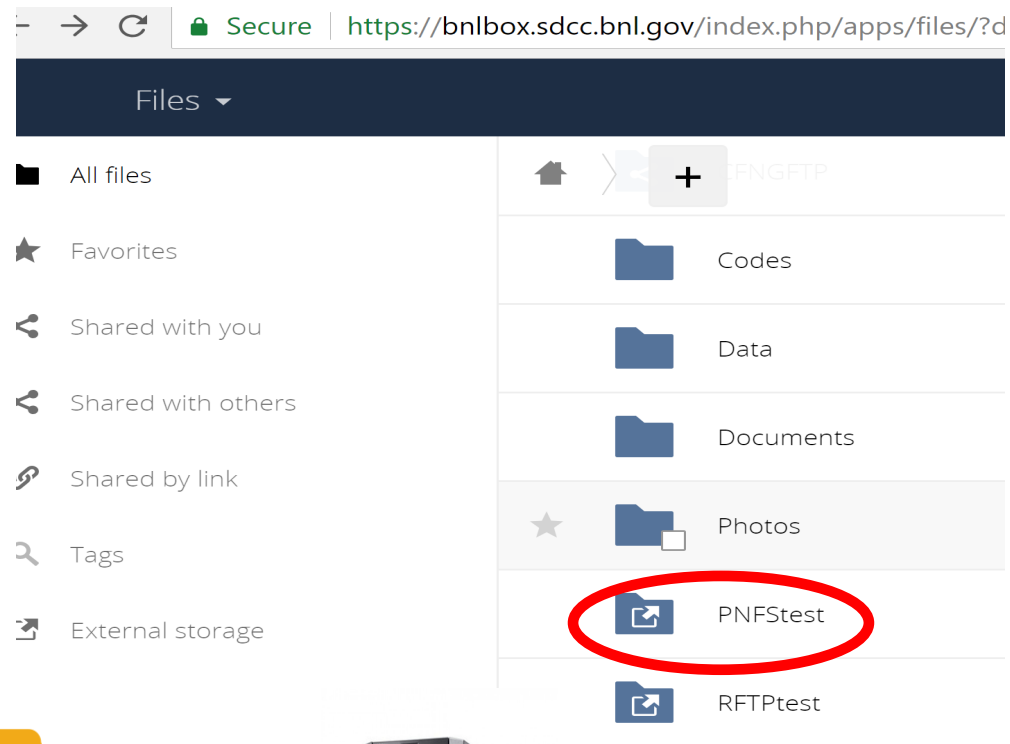
- Pros
 - It is a native dCache!!!
 - Performance and stability is great
- Cons
 - Can dCache developer keep supporting?
 - Can it be officially supported by Globus



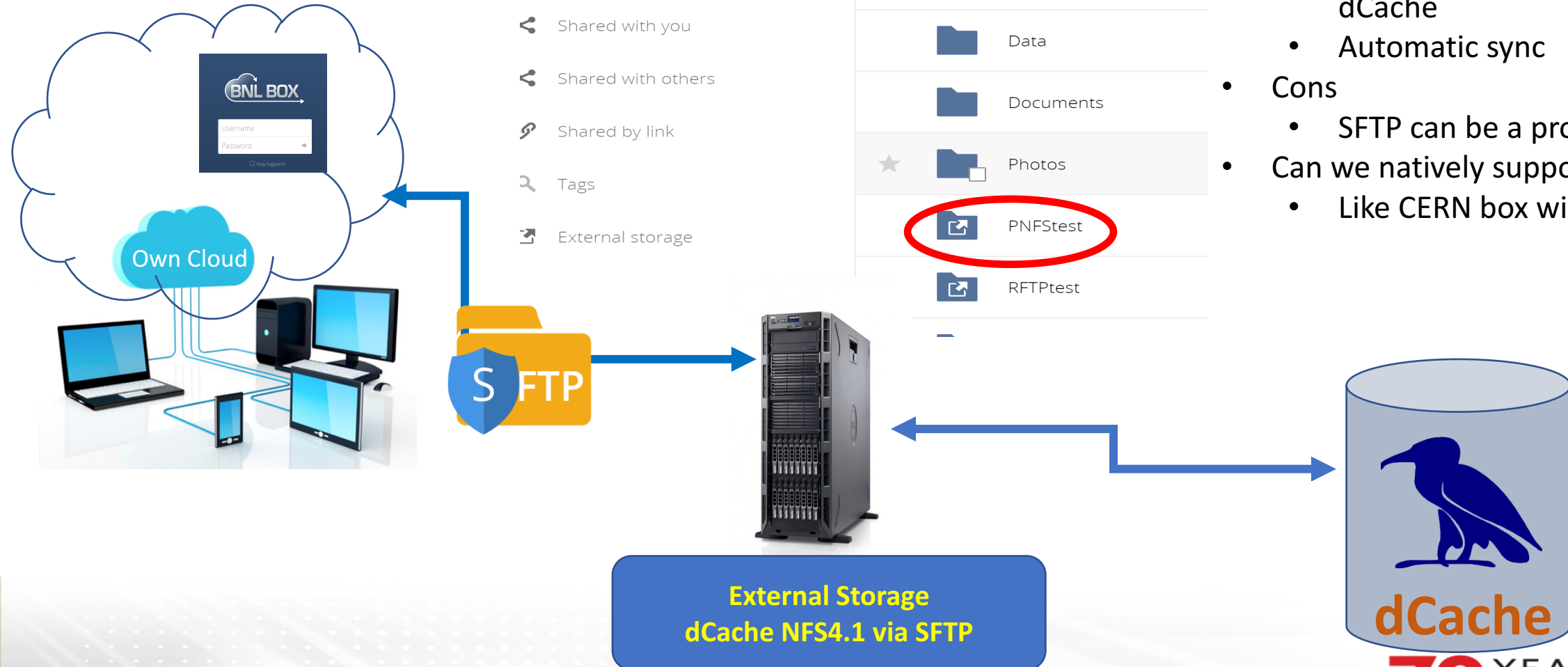
Globus Gridftp Server + dCache NFS4.1

- Pros
 - It is a native Globus. The server is always supported by Globus.
 - NFS4.1 is a standard.
- Cons
 - The performance is not as good as the dCache gridftp option.
 - The stability issue has been reported.

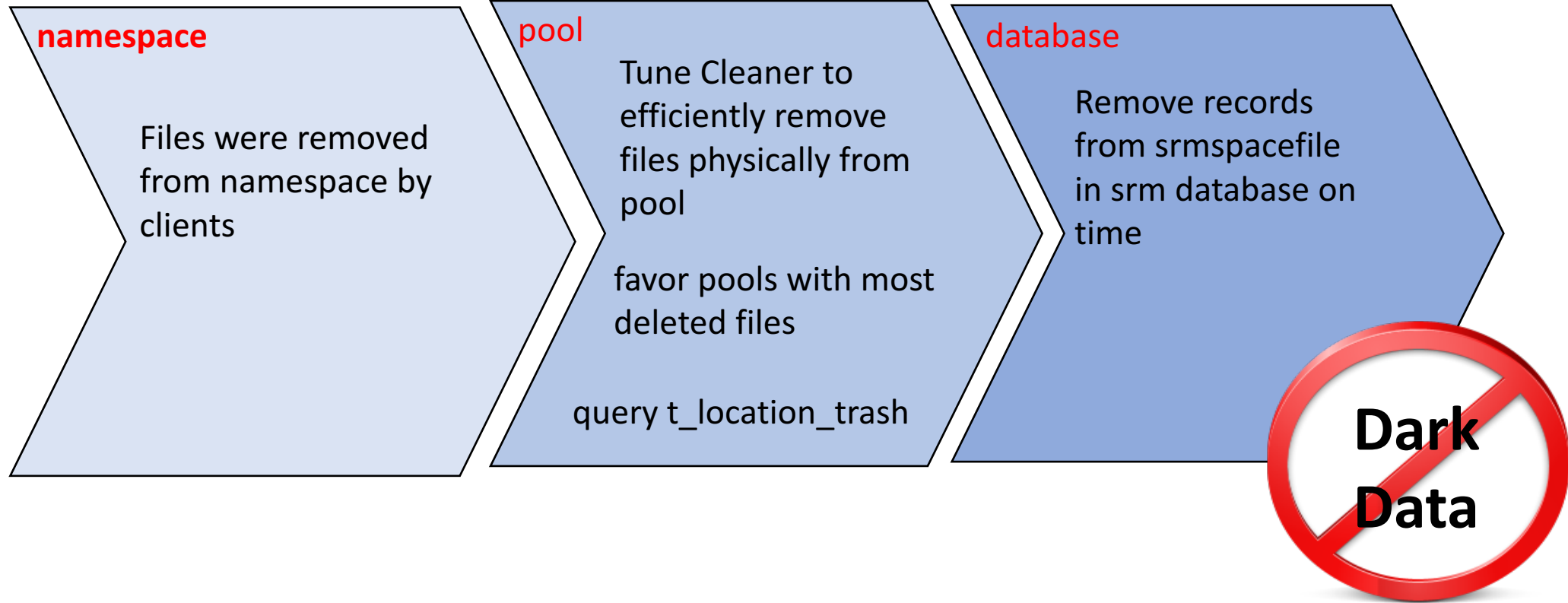
BNLBox (Cloud storage/Owncloud)



- Pros
 - Easiest to transfer data in/out dCache
 - Automatic sync
- Cons
 - SFTP can be a problematic.
 - Can we natively support?
 - Like CERN box with EOS?



Customization - Deletion



Customization— schema change

dcache. srmspacefile

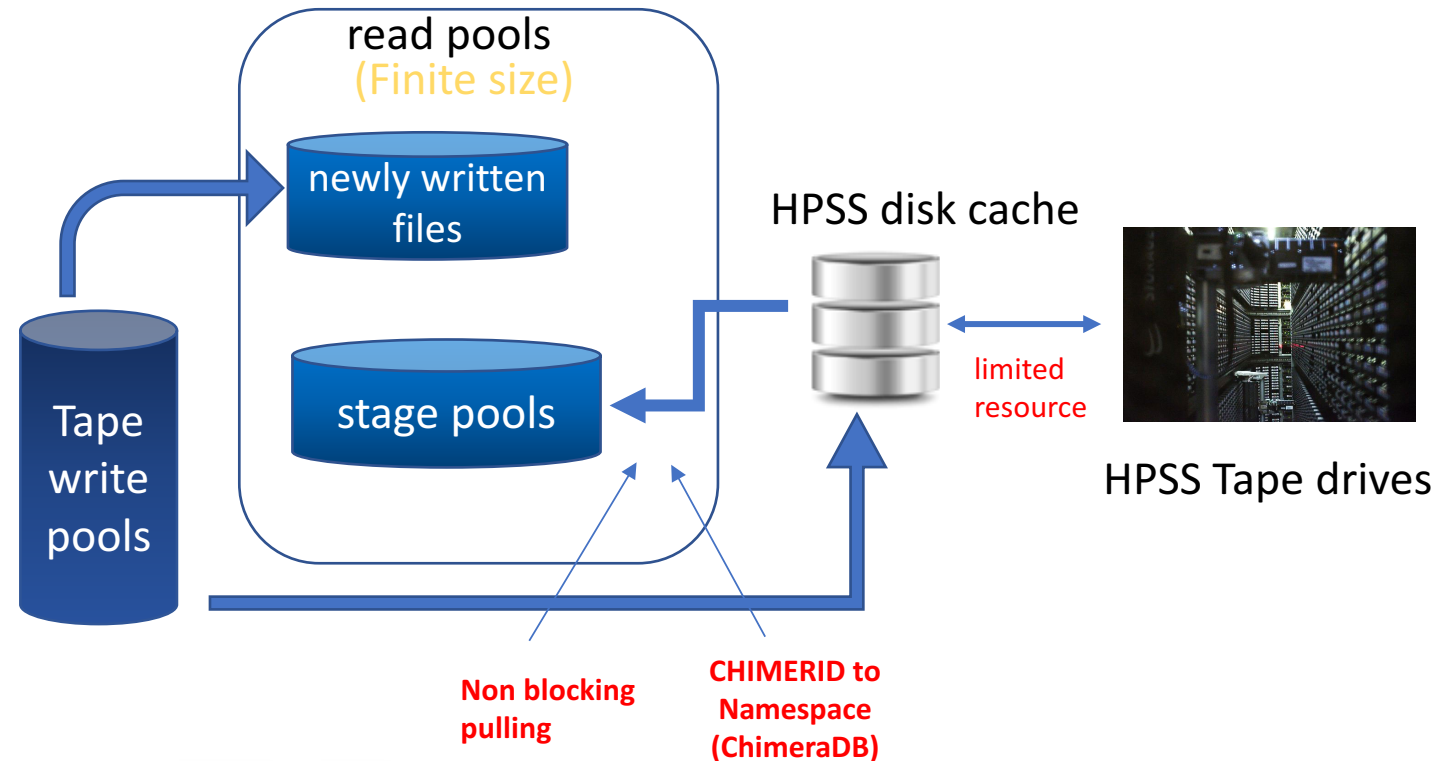
id	vogroup	vorole	spacereservation	sizeinbytes	creationtime	pnfsid	state



id	vogroup	vorole	spacereservation	sizeinbytes	creationtime	pnfsid	state	chimerapath

Future operation-Splitting read pools for Tape area

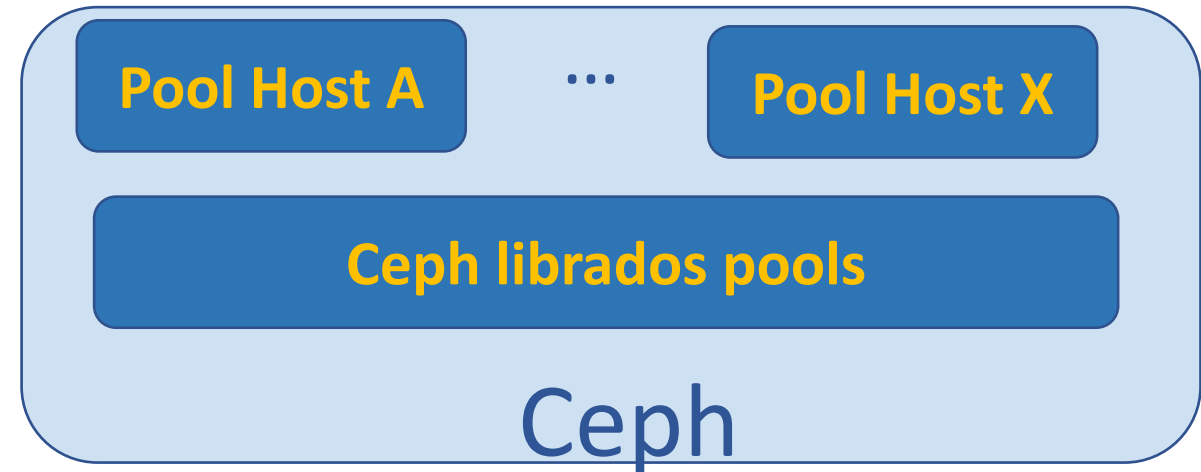
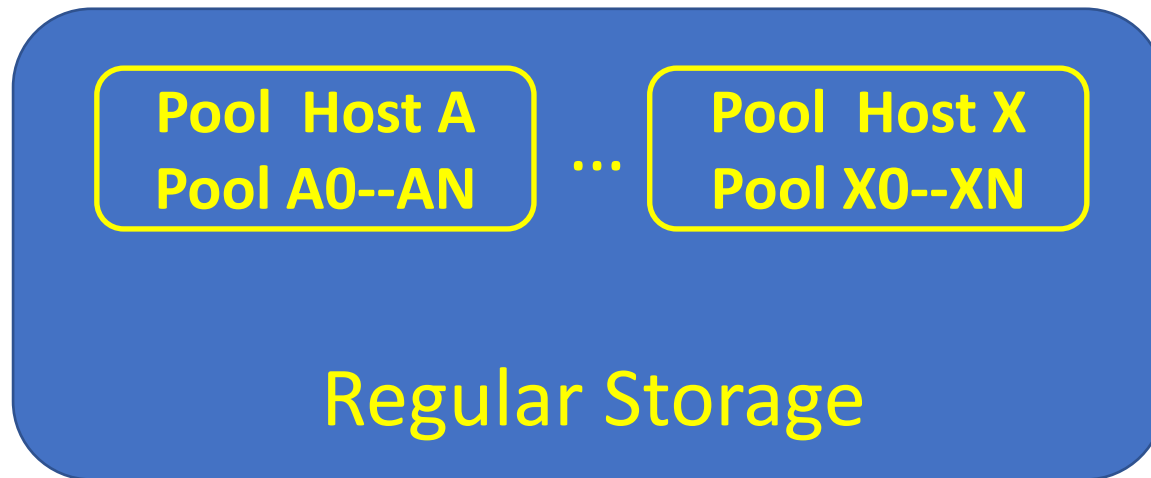
- Disk space crunch!
 - Various experiments and users have more data than the size of spinning disks available to them. And, it is expected to grow further.
- The tape storage is still cheaper than the spinning disks (or SSDs).
- Would like to use the archive storage (aka TAPE) as effective as reasonably possible.
 - TAPE requires a particular access to get the optimum performance.



The callback option for the future?

Future operation — Ceph Pools

- Pros:
 - Separate pool service hosts from actual storages.
 - remove the association of files in Ceph pools with t_location
 - Erasure codes allows the resiliency without duplicate copy



- Questions
 - Performance of partial read???
 - Scalability of Erasure code???

Needs help from developers

- continue the support for Ceph pools
- continue the support for Globus online
- support XRootD 3rd party transfer
- the issue with Resilience Management

Thank You!