Ceph Based Storage Systems at the RACF

Alexandr Zaytsev
alezayt@bnl.gov

BNL, USA
RHIC & ATLAS Computing Facility

Outline

- Ceph Evolution
  - New and upcoming features
- Ceph Installations in RACF
  - Evolution of our approach to Ceph
  - Current RACF Ceph cluster layouts
  - Production experience with RadosGW/S3
  - Production experience with CephFS/GridFTP
- Future Plans (2015-2016)
- Summary & Conclusion
- Q & A

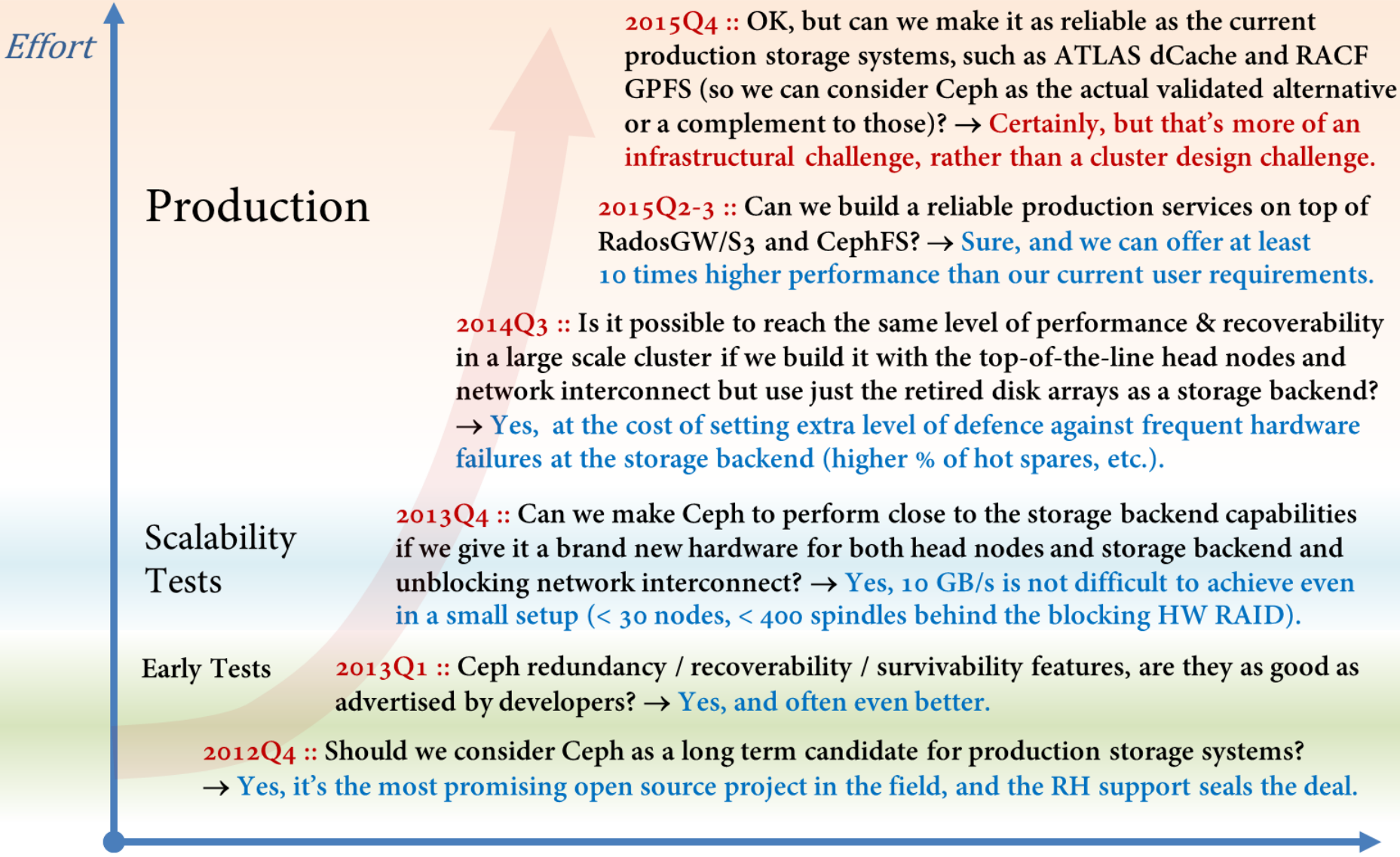
Evolution of Production Releases of Ceph

- *Hammer* production releases (v0.94.3)
  - Many critical CRUSH, OSD, RadosGW fixes
  - RadosGW object versioning and bucket "sharding" (splitting the index in order to improve performance on very large buckets)
  - CephFS recovery tools and more internal performance monitoring tools
  - Ceph over RDMA features (RDMA "xio" messenger support) support (seems not quite production ready yet)
    - Based on the 3rd party Accelio high-performance asynchronous reliable messaging and RPC library
    - Enables the full potential of the low latency interconnects such as Infiniband for Ceph components (particularly for the OSD cluster network interconnects) via eliminating additional IPoIB / EoIB layers and dropping the latency of the cluster interconnect down to the microsecond range
  - *We are currently targeting v0.94.3 for production deployment on both RACF Ceph clusters*

Evolution of Production Releases of Ceph

- Future production *Infernalis* release (v9.x.x under the new versioning schema, development v9.0.3 is now available)
  - Many more improvements across the board
  - New OSD storage backend interfaces:
    - NewStore (replacement for the FileStore that implements the ObjectStore API)
      – expected to deliver performance improvements for both capacity-oriented and performance oriented Ceph installations
    - Lightning Memory-Mapped Database (LMDB) key/value backend for Ceph (currently available LSM-tree based key/value store backend turned out to be less performant than the current FileStore implementation)
      – expected to deliver performance improvements for performance-oriented Ceph installations, especially for handling small objects
  - New filesystem interfaces to Ceph:
    - Hadoop FileSystem Interface
    - Access to RadosGW objects through NFS
  - *Hopefully going to provide production ready RDMA support*

Evolution of the RACF Approach to Ceph



2015Q4 :: OK, but can we make it as reliable as the current production storage systems, such as ATLAS dCache and RACF GPFS (so we can consider Ceph as the actual validated alternative or a complement to those)? → Certainly, but that's more of an infrastructural challenge, rather than a cluster design challenge.

2015Q2-3 :: Can we build a reliable production services on top of RadosGW/S3 and CephFS? → Sure, and we can offer at least 10 times higher performance than our current user requirements.

2014Q3 :: Is it possible to reach the same level of performance & recoverability in a large scale cluster if we build it with the top-of-the-line head nodes and network interconnect but use just the retired disk arrays as a storage backend? → Yes, at the cost of setting extra level of defence against frequent hardware failures at the storage backend (higher % of hot spares, etc.).

2013Q4 :: Can we make Ceph to perform close to the storage backend capabilities if we give it a brand new hardware for both head nodes and storage backend and unblocking network interconnect? → Yes, 10 GB/s is not difficult to achieve even in a small setup (< 30 nodes, < 400 spindles behind the blocking HW RAID).

2013Q1 :: Ceph redundancy / recoverability / survivability features, are they as good as advertised by developers? → Yes, and often even better.

2012Q4 :: Should we consider Ceph as a long term candidate for production storage systems? → Yes, it's the most promising open source project in the field, and the RH support seals the deal.

The Infrastructural Challenges

Storage backend hardware and OS level recovery / rediscovery must be ensured outside Ceph as well.
Fully Puppetized installation for quick re-installation of the head nodes, periodic checks for LUNs health, automatic recovery of the LUNs becoming available again after disk arrays failure, etc.

Internal Ceph performance monitoring tools are great for early problems detection
but something outside Ceph must look at them

Automatic power up of the Ceph cluster components and disk arrays after the complete power loss; minimizing the downtime for the storage arrays with the old disks inside to prevent mass failures.
All our Ceph cluster components are tuned to do so after 30 sec timeout (since our clusters are still mainly on normal power – plenty of opportunity to battle test).

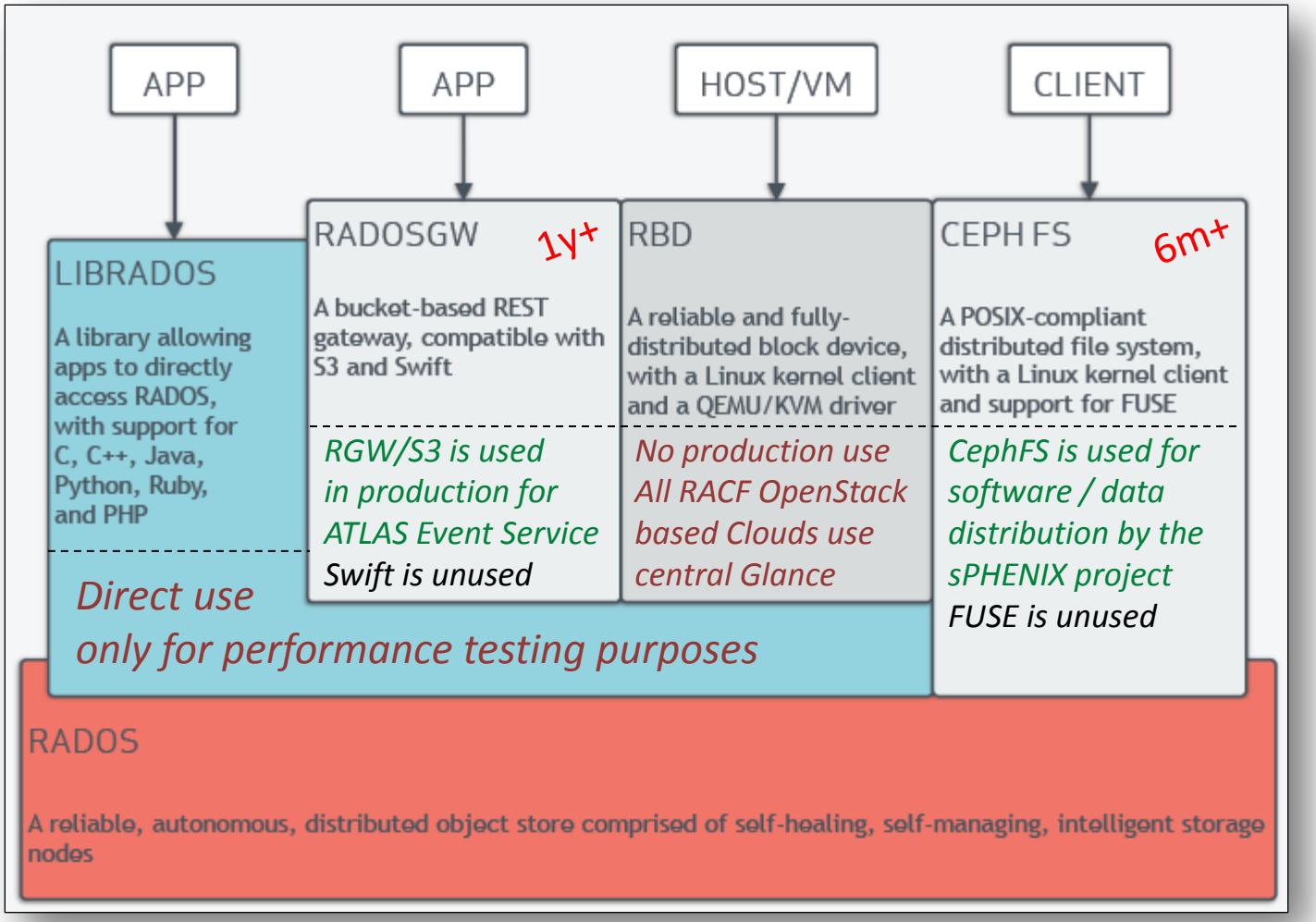
Keep all the internal low level storage traffic (such as iSCSI) on the links that a physically bound to the Ceph cluster if possible

Alert about any changes in the RAID system to minimize chances of loosing the OSDs completely (or even groups of OSDs, as one physical array can carry several OSD volumes).
Over the last year we never went near to a complete raid set loss thanks to preventive maintenance.

Localize all internal OSD-to-OSD traffic on isolated & well protected switches

OSD mapping to the head nodes must ensure that a temporary or permanent loss of equipment in any single rack of storage equipment cant affect more than 30% of the head nodes in the cluster.
Our Ceph equipment is distributed across 3 rows in the CDCE area of RACF and the OSD mapping takes that into account.

RACF: Current Use of Ceph Components



RACF Ceph Clusters: Building Blocks

First gen. head nodes, first and second gen. gateways

x18

Dell PowerEdge R420 (1U)

2x 1 TB HDDs in RAID-1 + 1 hot spare
50 GB RAM + 1x 250 GB SSD (up to 10 OSDs)
1x 40 GbE + 1x IPoIB/4X FDR IB (56 Gbps) – *Head nodes*
2x 10 GbE – *Gateways*

Second gen. head nodes

x8

Dell PowerEdge R720XD (2U)

8x 4 TB HDDs in RAID-10 + 2 hot spares
128 GB RAM + 2x 250 GB SSDs (up to 24 OSDs)
1x 40 GbE + 1x IPoIB/4X FDR IB (56 Gbps) +
12x 4 Gbps FC ports

Storage backend (retired ATLAS dCache HW RAID disk arrays)

iSCSI export nodes

SUN Thor servers (Thors)

48x 1 TB HDDs under ZFS
8 GB RAM
1x 10 GbE
4x 4 Gbps FC (no longer used)

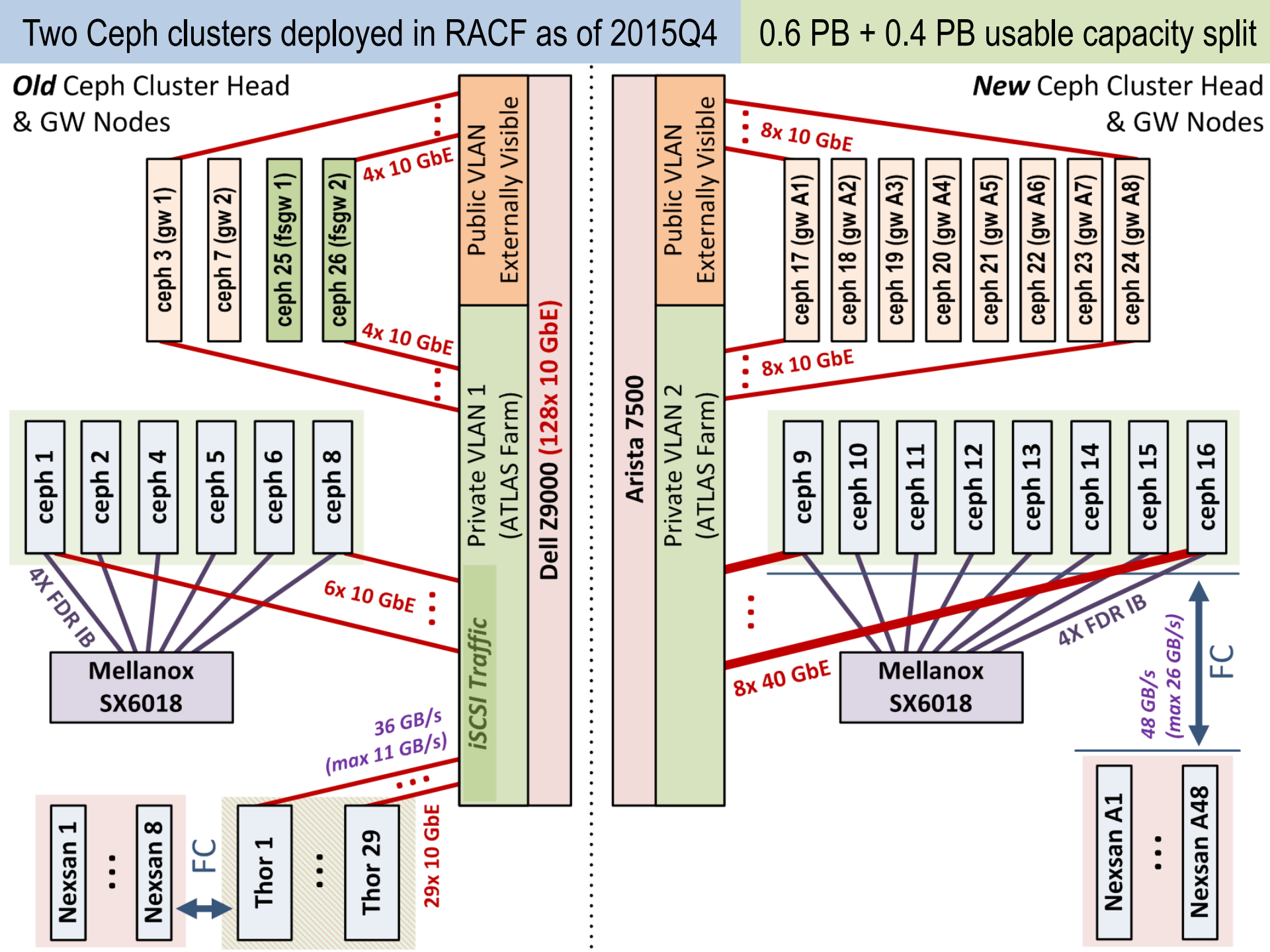
x29

FC attached storage arrays

Nexsan SATABeast arrays (Thors)

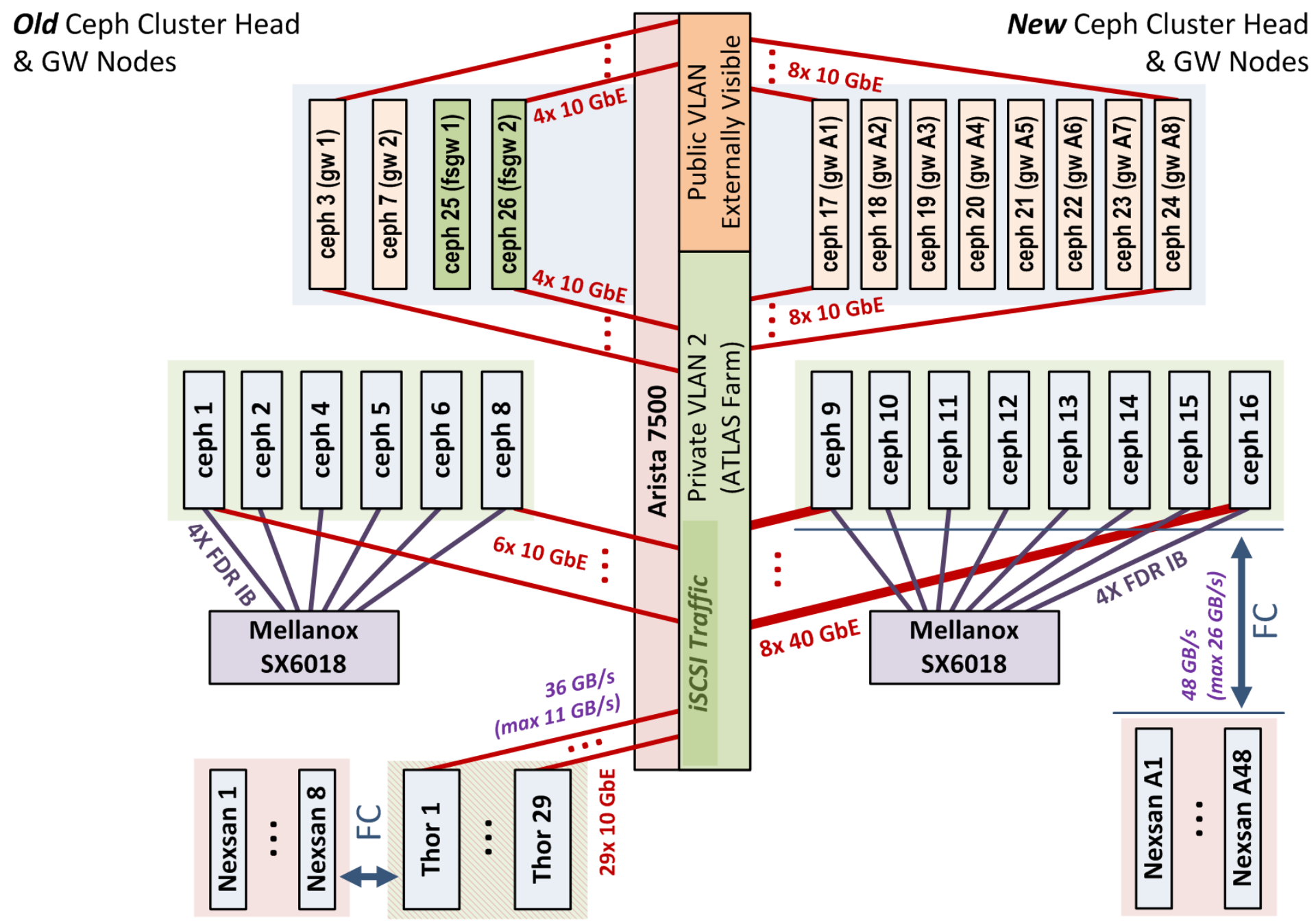
40x 1 TB HDDs in
HW RAID-6 + 2 hot spares
2x 4 Gbps FC (no longer used)

x56


Two Ceph clusters deployed in RACF as of 2015Q4
0.6 PB + 0.4 PB usable capacity split
Old Ceph Cluster Head & GW Nodes
New Ceph Cluster Head & GW Nodes
ceph 3 (gw 1)
ceph 7 (gw 2)
ceph 25 (fsgw 1)
ceph 26 (fsgw 2)
4x 10 GbE
4x 10 GbE
Public VLAN Externally Visible
Private VLAN 1 (ATLAS Farm)
iSCSI Traffic
Dell Z9000 (128x 10 GbE)
Arista 7500
Public VLAN Externally Visible
Private VLAN 2 (ATLAS Farm)
8x 10 GbE
8x 10 GbE
ceph 17 (gw A1)
ceph 18 (gw A2)
ceph 19 (gw A3)
ceph 20 (gw A4)
ceph 21 (gw A5)
ceph 22 (gw A6)
ceph 23 (gw A7)
ceph 24 (gw A8)
ceph 1
ceph 2
ceph 4
ceph 5
ceph 6
ceph 8
4X FDR IB
6x 10 GbE
Mellanox SX6018
ceph 9
ceph 10
ceph 11
ceph 12
ceph 13
ceph 14
ceph 15
ceph 16
8x 40 GbE
4X FDR IB
Mellanox SX6018
48 GB/s (max 26 GB/s)
FC
36 GB/s (max 11 GB/s)
29x 10 GbE
Nexsan 1
Nexsan 8
FC
Thor 1
Thor 29
Nexsan A1
Nexsan A48

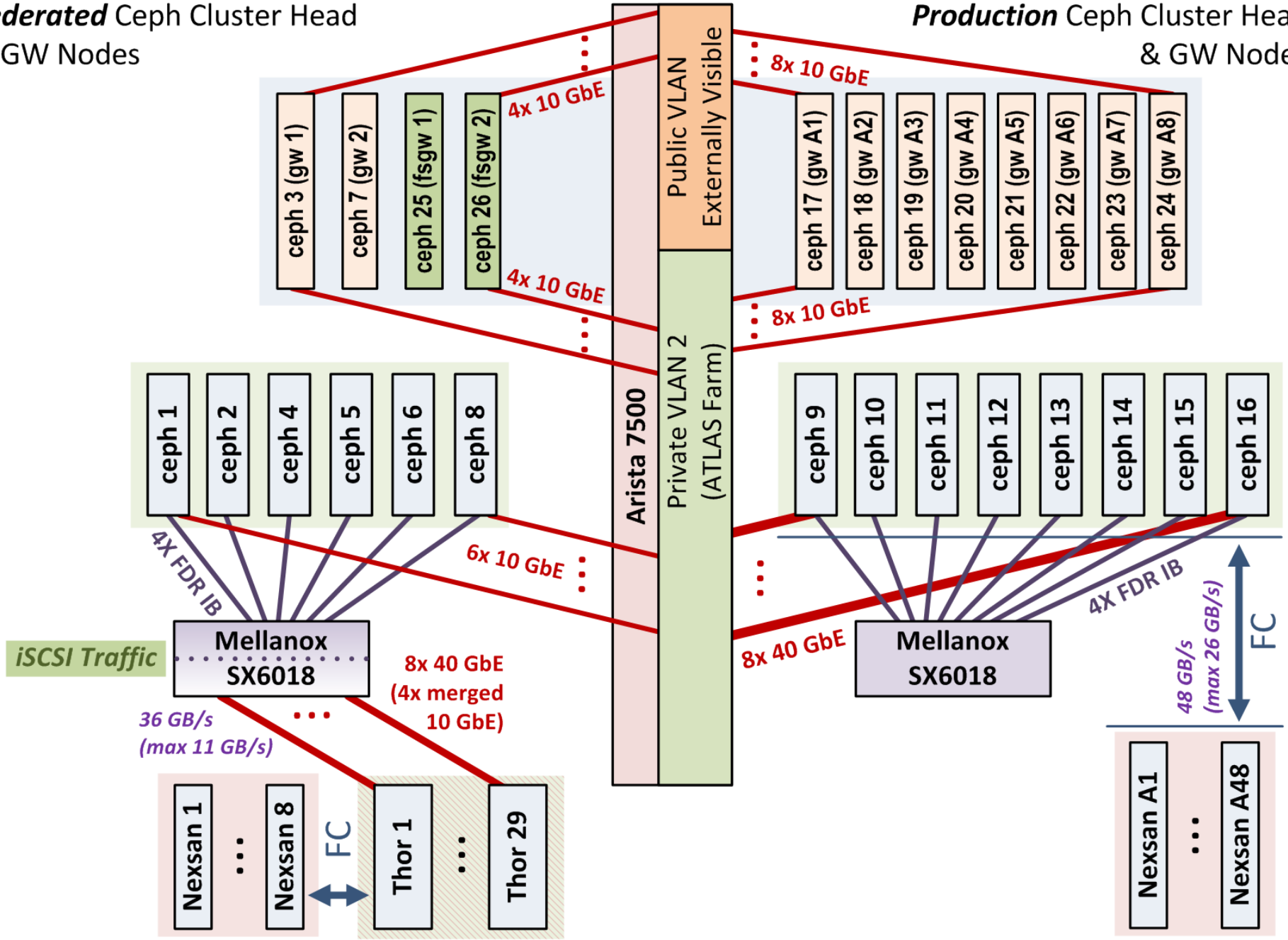
Finalize migration to the new Arista based ATLAS networking core

2015Q4



Remove all remaining iSCSI traffic from central ATLAS networking core

2015Q4- 2016Q1



Short Term Plans for Ceph Installations in RACF

- Migrate all production services to the new Ceph v0.94.3 based cluster (1.8 PB raw capacity, 1 TB of RAM on the head nodes) provided with mixed simple replication factor 3x pools and erasure coded pools (Oct 2015)
  - **192 OSDs**, 0.6-1.0 PB of usable capacity (depending on erasure code pool profile, which is not finalized yet)
  - MON and MDS components on separate hardware from the OSDs
  - Try the full Amazon S3 compliant RadosGW/S3 interfaces (using DNS name based bucket identification in the URLs)
  - Perform the new performance/stability tests for CephFS with 8x 10 GbE attached client nodes, targeting network limited maximum aggregated I/O throughput of **10 GB/s**
- Rebuild the "old" Ceph v0.80.1 based cluster (1.2 PB raw capacity, 0.3 PB of RAM on OSD nodes) under Ceph v0.94.3 with the isolated low latency networking solution for the iSCSI transport (2015Q4-2016Q1)
  - Finalize the public Ethernet network reorganization on both Ceph clusters in the process
  - Make it a testbed for the Federated Ceph cluster setup
- Ensure the full 1+1 input power redundancy for critical components of both clusters (normal power plus central UPS) by the end of 2015

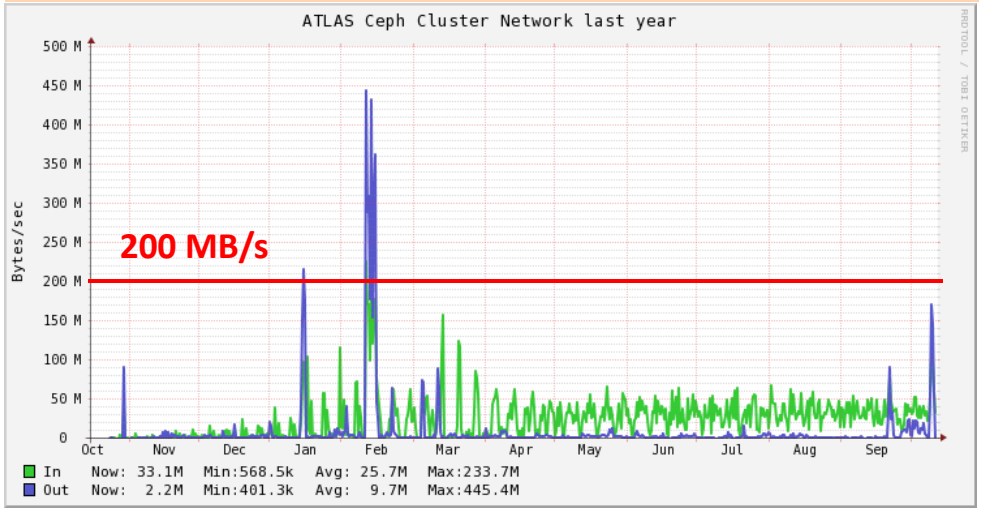
Production Experience with Ceph

https://www.flickr.com/photos/matthew-watkins/8631423045

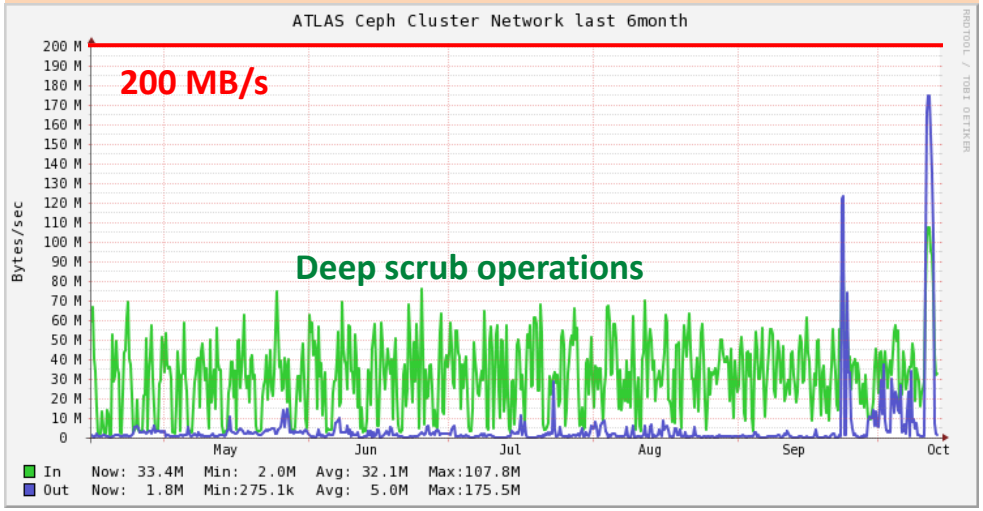
Production Experience with ATLAS Event Service

- 14.9M objects in PGMAP of the production cluster after 11 months of use by ATLAS ES
- 4.5M objects in a single atlas_pilot_bucket (no hard limit set)
- Only 4% of the available capacity of the cluster is used so far for ATLAS ES related data so far
- Only 7% of the maximum I/O capacity to the RadosGW/S3 clients (≈900 MB/s) of the current production cluster is used by ATLAS ES so far
- ATLAS ES load can be increased by factor of 10x while still staying well within the capabilities even of the *old* RACF Ceph cluster

Ceph internal cluster network activity over the last 12 months



Ceph internal cluster network activity over the last 6 months



Internal traffic within the cluster caused by ATLAS ES accessing it though the RadosGW/S3 (localized spikes up to 65 MB/s)

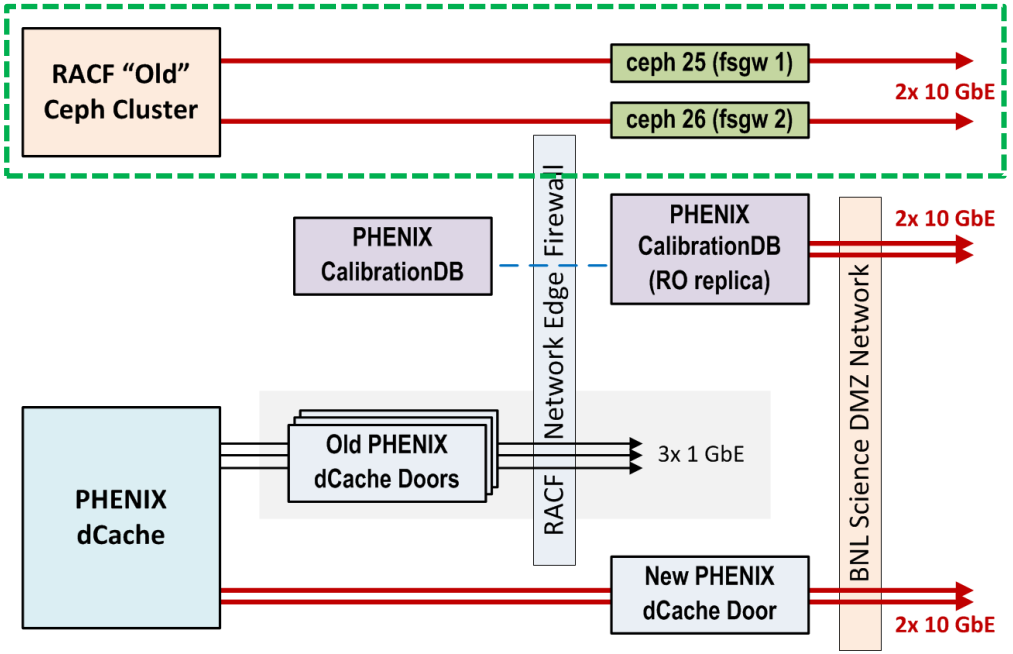
Ceph @ RACF: RadosGW/S3 Interfaces

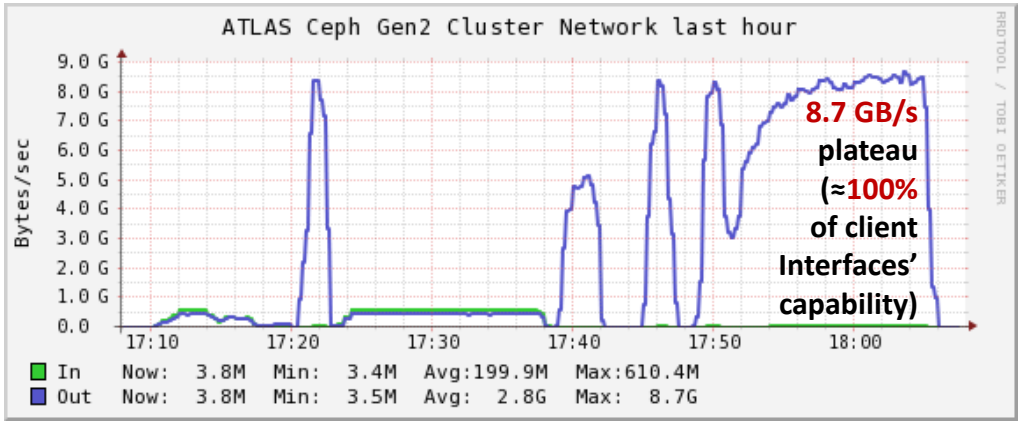
- First observed back to 2014Q4, still makes a valid point for the clients using the object store layer of Ceph:

- Comparing the performance of Ceph APIs on different levels for the large bulks of small objects (less than 1 kB in size)
  - Using the low level librados Python API to write a bulk of up to 20k objects of identical size per thread to a newly created pool in a synchronous mode (waiting for an object to fully propagate through the cluster before handling the next one)
    - **1.2k** objects/s (maximum achieved with 32 threads / 4 client hosts running in parallel)
  - Using the same low level librados Python API to write the same set of objects in an asynchronous mode (no waiting for object propagation through the cluster)
    - **3.5k** objects/s (maximum achieved in the same conditions)
  - Using the AWS/S3 Ruby API to access two RadosGW instances simultaneously and perform object creation in a synchronous mode
    - **0.43k** objects/s (maximum achieved with 128 threads / 4 hosts accessing both Rados GW instances in parallel)

Production Experience with CephFS/GridFTP

- New data access infrastructure enabling PHENIX production on the opportunistic OSG resources deployed in RACF in 2015Q3 includes a group of two CephFS / GridFTP gateways
  - Globus GridFTP server version 7.x
  - OSD striping is custom tuned for user directories in CephFS for maximum performance by using the *xattr* mechanism
- Available for production use since Aug 2015
  - Up to 300 TB of usable space is available (with factor of 3x replication protection)
  - Still deployed on top of the "old" Ceph cluster, yes capable of 1 GB/s data throughput which exceeds requirements at the moment
  - As it was already reported before, we are capable of serving data through CephFS at the level of 8.7 GB/s with the "new" Ceph cluster, but the demand is not there yet





Longer Term Plans for Ceph Deployments at RACF

- Eventually, the underlying 1 TB HDD based hardware RAID arrays are to be replaced with 2 TB HDD based arrays of similar architecture (2016)
  - Newer Nexsan SATABeast arrays retired by the BNL ATLAS dCache storage system
  - Fewer number of disk arrays for the same aggregated usable capacity of about 1 PB
- No radical storage architecture change is foreseen at the moment (FC attached disk arrays that are connected to the OSD nodes plus a mix of 40 GbE/10 GbE/4X FDR IB network interconnects are going to be used)
  - Yet using larger number of OSD nodes filled with local HDDs/SSDs and thus much larger number of OSDs per cluster looks like the best underlying hardware architecture for a Ceph cluster right now (no hardware RAID components are really needed for it)
  - Hitachi Open Ethernet Drive and Seagate Kinetic Open Storage architectures are very promising too, but may not be most price/performance efficient as of 2015Q4
- Possible changes in the objects to buckets mapping for the data written to BNL Ceph clusters through RadosGW/S3 interface:
  - Current approach with putting all objects into a single generic S3 bucket with millions (potentially – tens of millions objects in it) creates certain performance problems for pre-Hammer releases of Ceph (not affecting individual object access latency though)
  - Using multiple buckets can be an option (might also simplify deletion of the old objects)
  - Bucket “sharding” mechanism now available in Hammer release might alleviate the problem while still using a single bucket

Summary & Conclusion

- In 2014-2015 Ceph has become a well established storage platform for RACF after one year of gathering production experience
- RadosGW/S3 and CephFS/GridFTP are the most popular storage layers of Ceph for our clients
  - The scale and performance reached by RACF Ceph installations are staying well above our current user requirements
  - We are hoping that it is going to change soon as result of the evolution of the ATLAS Event Service project in particular
- Bringing Ceph in line with other storage system operated by RACF such as ATLAS dCache and GPFS in terms of quality of service and reliability is a challenging task, mostly from the facility infrastructure point of view
  - We are now leaving “testing and construction” period and arriving to the steady maintenance period
  - Large scale backend storage replacement operation is foreseen in 2016, but it is not going to change the overall architecture of our Ceph clusters
- We are open for cooperation with other Ceph deployments in the community, such as the recently funded “Multi-Institutional Open Storage Research InfraStructure (MI-OSiRIS)” project (NSF)


© Jose Antonio Hervás
Q & A
http://apod.nasa.gov/apod/ap150929.html