OPEN SOURCE STORAGE FOR ARCHIVAL: FILE → OBJECT

SAGE WEIL – RED HAT

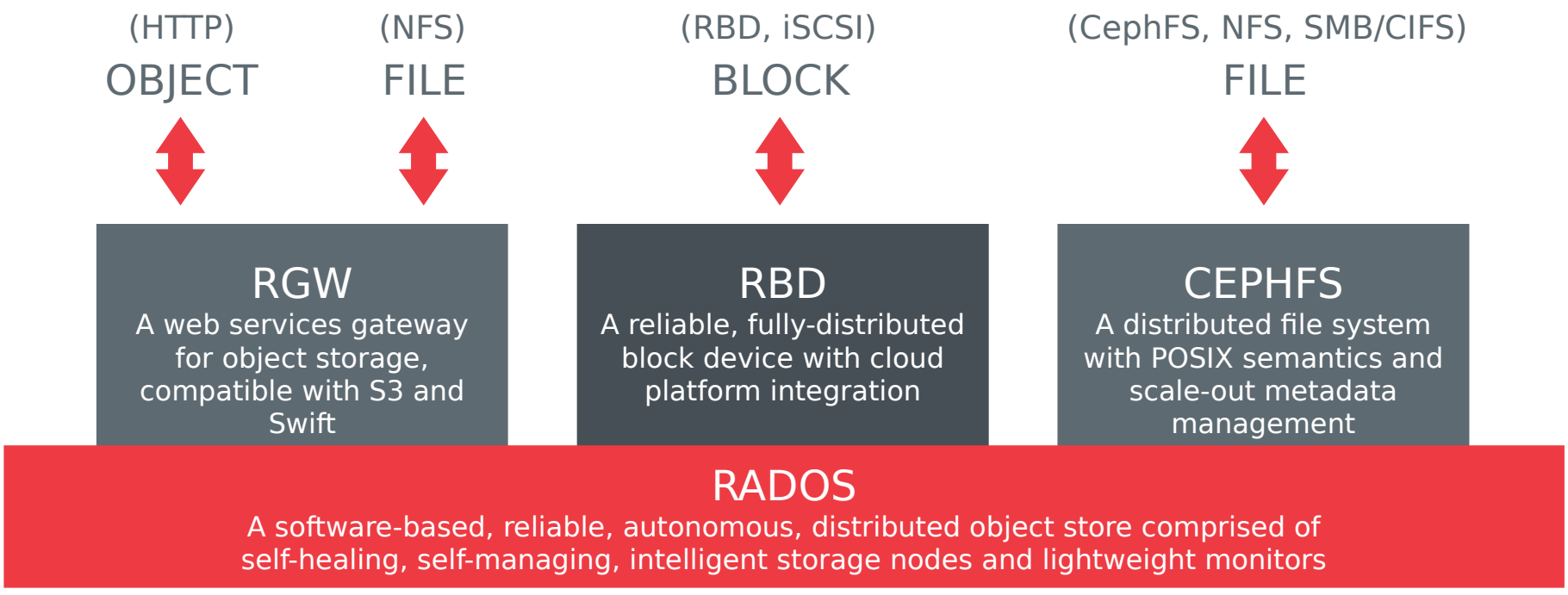
2017.09.18

CEPH

- Object, block, and file storage in a single cluster
- All components scale horizontally, no single point of failure
- Hardware agnostic, commodity hardware
- Open source (LGPL)

- Hardware and software vendor independence
  - cost
  - data lifetime > vendor lifetime
- Transparency
  - how to read your data in 10, 20, 50 years?
  - data is never hostage to proprietary platform or org
- Efficient investment of engineering effort

CEPH COMPONENTS AND PROTOCOLS



TRANSITIONING TO OBJECT APIS

- Objects are
  - big, (usually) immutable, efficient
  - easy to replicate, mirror, proxy, cache
- Ceph RGW object gateway
  - compression, encryption, quota, multi-site federation, erasure coding, tiering, ...
  - indexed and searchable (new!)
- Primary API is S3 (or Swift)
  - NFS as secondary API: ingest and export, broad compatibility

- Choose your use-case
  - file storage workloads that read/write entire files/objects

    (object stores are not full-blown file systems with small, in-place files updates)
  - aligns well with archives!

FILE → OBJECT

- “A big financial institution”
  - Ceph RGW for log storage and later analysis
  - NFS perfect for log ingest from broad range of hosts (using existing tools without modification)

- “A large insurance company”
  - custom application using lots of files and NFS
  - new generation of app will be object-based
  - long transition to convert and phase out old NFS users

MORE FILE → OBJECT

- “An oil and gas company”
  - import data over NFS from processing cluster
  - use Ceph RGW multi-site replication to distribute data globally to other data centers
  - read access via both object and NFS

- CERN
  - self serve Ceph RGW object storage service
  - range of physics apps with custom, ad hoc storage backends
  - CephFS and XRootD: global hierarchical scientific data archive

CEPH VS ARCHIVAL REQUIREMENTS

- Integrity
  - full data + metadata checksums
    - (but no efficient fixity audit)
  - background scrubbing
- Cost-efficient
  - open source, vendor choice
- High availability
  - replication or erasure coding
  - no single points of failure
- Multi-protocol
  - S3, Swift, NFS
  - CIFS, block, ...
- Access control
  - S3 or Swift auth model

    (bucket/container based)
  - STS: federation with kerberos (under development)
- Audit and logging
  - all administrator actions
  - internal consistency checks, inconsistencies and repair actions, ...
  - data access logs are optional
  - log preservation (e.g., immutable log storage) out of scope for Ceph itself

THANK YOU

- Minimal IT staff training?

Sage Weil

Ceph Project Lead / Red Hat

sage@redhat.com

@liewegas