

Piston
PISTON{CLOUD}COMPUTING

OpenStack: Storage

Object, Block, and Beyond

Christopher MacGown

Co-Founder and Chief Scientist



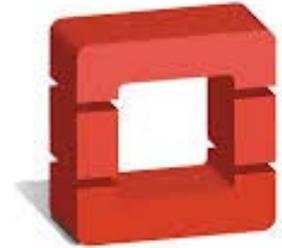
Storage in the cloud

- Virtually attached block devices
 - Roughly equivalent to traditional SAN/NAS
- Distributed object storage
 - Scalable, distributed file storage



Object Storage: Swift

- Horizontally scalable storage service
- Redundancy via full-replicas
 - Synchronous replication for uploads
 - Background replication for failures
 - 5GB file size limit unless file is broken up or uploaded via special command
- Placement handled via distributed hashing ring
- Supports both expiring and versioned objects
- Provides container synchronization across data centers



openstack™
CLOUD SOFTWARE



Object Storage: Ceph

- Horizontally scalable storage service
- Redundancy via full replicas
 - Erasure coding a proposed alternative
 - Synchronous replication during upload
 - Background replication upon failure
 - Can handle unlimited file sizes
- Policy based algorithmic placement based on CRUSH algorithm
- No container synchronization across data centers until next release
- Can expose virtual block devices on top of object store



Object Storage

- Use commodity hardware to provide effectively unlimited scaling
- Self-healing: assuming available capacity can absorb failures
- Already in use at large scale
 - SDSC Cloud Storage serves 5.5PB of swift and scaling to >100PB
 - Commercial vendors like Rackspace are MUCH larger but won't talk about their numbers

BUT...

- Assumed 3x replicas can be expensive
- De-duplication is a commercial feature
- Available bandwidth limits ability to scale



Object Storage: APIs

- Converging to a couple common APIs
 - Amazon's S3 API
 - Supported generally across the board at varying levels of fidelity
 - OpenStack Swift API
 - Supported by Swift, as well as Ceph, and Riak, but some features not implemented
- API fragmentation is slowly disappearing, but feature fragmentation is still an issue



Block Storage: Cinder

- Virtualized iSCSI attached block devices
- Multi-Backend
 - Likely supports your existing SAN and NAS infrastructure
 - EMC, NetApp, Nexenta
 - Supports open source alternatives as well
 - NFS, GlusterFS, Ceph RBD
- Attached directly to virtualized compute infrastructure
 - POSIX storage, but not POSIX shared storage



Storage in the cloud

- Not a matter of 1, 5, or 10 year horizon
 - The use-case drives storage decisions
 - Technology is limited by exponential laws of computing
 - So it isn't merely:
 - Block storage for performance sensitive computation
 - Object storage for performance insensitive computation
 - Maybe it's both...
- Understand the model of your storage growth
 - A more accurate forecast lets you plan your scalability around a shorter deployment horizon
- Linear scaling is the goal



Thank you.

