



STORAGE FOR DATA PLATFORMS

Federico Lucifredi & Kyle Bader

OpenStack Summit, Vancouver 2018

COMMON ARCHITECTURAL MODEL - PUBLIC OR PRIVATE CLOUD

PUBLIC CLOUD (AWS)

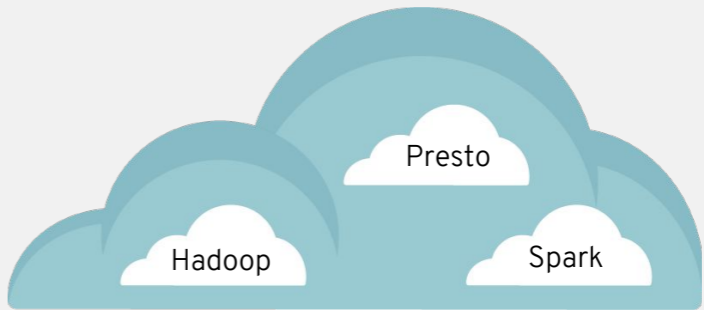


**AWS EC2
PROVISIONING**



**AWS S3
SHARED DATASETS**

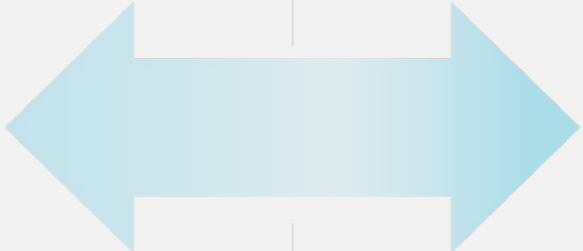
PRIVATE CLOUD



**OPENSTACK
PROVISIONING**



**CEPH S3/SWIFT
SHARED DATASETS**



THREE PILLARS FOR DATA PLATFORMS



Object



Persistent Block



Low latency
local block

OBJECT

- Shared persistence layer across compute clusters
- Extreme scalability
 - Capacity
 - Throughput
- Economical



OBJECT WORKLOADS



PERSISTENT BLOCK

- Arbitrary block capacity
- Capacity based IOPS
- Persistence helps with management of database lifecycles
 - Detach and reattach on larger instance
 - Snapshots for backups and copy-on-write for secondaries



PERSISTENT BLOCK WORKLOADS



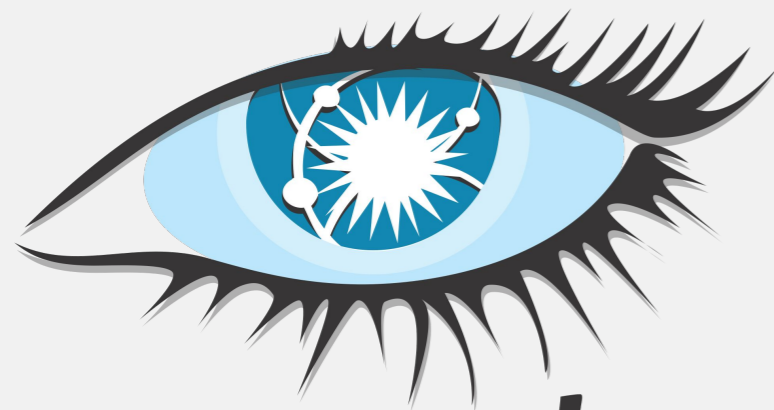
PostgreSQL



LOW LATENCY LOCAL BLOCK

- For distributed applications that tolerate instance failures
- Big data workloads
 - Scratch data
 - Intermediate spill data
 - Buffers

LOW LATENCY LOCAL BLOCK



cassandra



elasticsearch



etcd

COMMON ARCHITECTURAL MODEL - PUBLIC OR PRIVATE CLOUD

PUBLIC CLOUD (AWS)

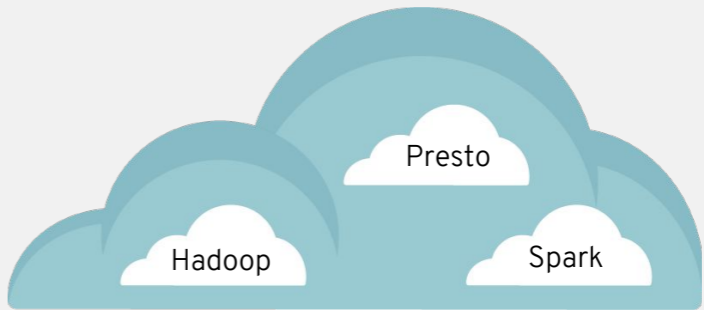


**AWS EC2
PROVISIONING**



**AWS S3
SHARED DATASETS**

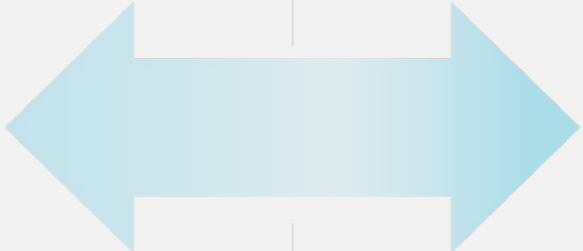
PRIVATE CLOUD



**OPENSTACK
PROVISIONING**



**CEPH S3/SWIFT
SHARED DATASETS**



THANK YOU!

COME TO SEE US AT THE RED HAT BOOTH

