

keeper^{technology}

access | manage | protect

Presents...

Unlocking  ceph



*LOC Designing Storage Architecture
for Digital Collections*

Glenn Heinle
Chief Technology Officer
Keeper Technology
9/09/19

Ceph

What is it?

- Reliable
- Scalable
- SDS
- Clustered
- Open Source
- Object, Block, File

When? Where?

- Doctoral Dissertation
- 2004, UCSC
- Inktank, RedHat, IBM

What Impact?

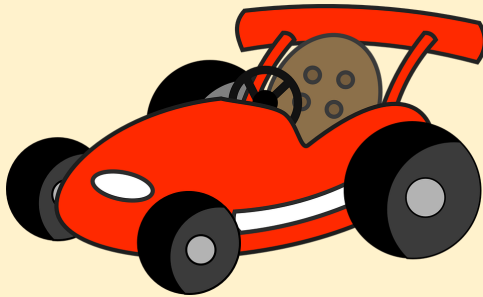
- DIY
- OpenStack
- On-Prem Cloud
- Multiple Entries in Top 50 of IO-500 list



Identify Your Use Case

Throughput-Optimized

- HDD/SSD
- 10/25 Gb
- Medium Density



Capacity-Optimized

- HDD
- 10 Gb
- High Density



IOPS-Optimized

- SSD/NVMe
- 25/40/100
- IB, RDMA
- Low Density



Hardware



- The hardware you use is very important
- “*Commodity Hardware*” does not mean old or low performance
- Trying to save money on hardware is likely to increase cost elsewhere
- Maintainable
 - Supported



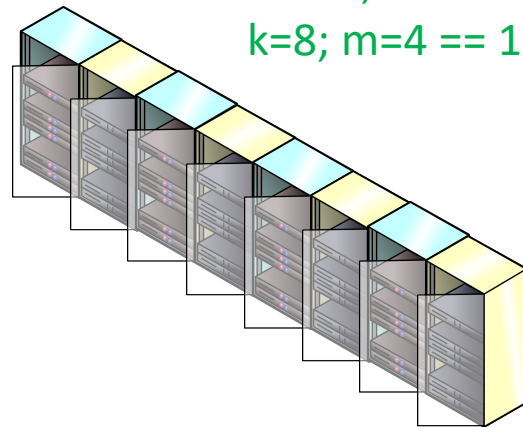
Storage Nodes

SSD/NVMe drives
play an important
performance role


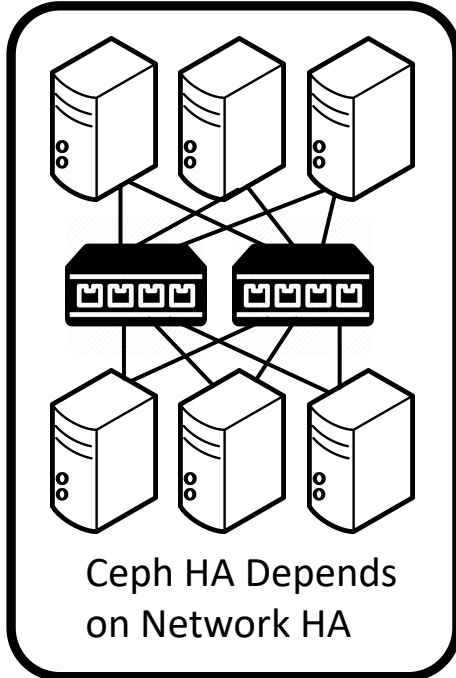


You can put 1.5
PB in a shelf.
But...should you?

Erasure-coding requires
more nodes, not less.
 $k=8; m=4 == 12$



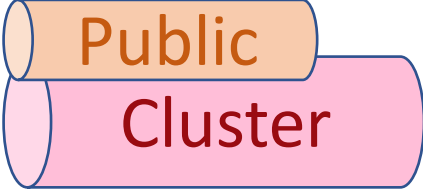
Network



MTU
9000

Ceph
Loves
Jumbo
Frames

This block features a blue scroll icon with the text 'MTU 9000' written on it. To the right of the scroll, the text reads 'Ceph Loves Jumbo Frames'.



Public
Cluster


Cluster Network will use more bandwidth than Public Network

This block contains an icon of two overlapping cylinders, one orange and one pink. Below the icon, the text reads 'Public Cluster'. Below that, the text states 'Cluster Network will use more bandwidth than Public Network'.



Separate network for IPMI & Management

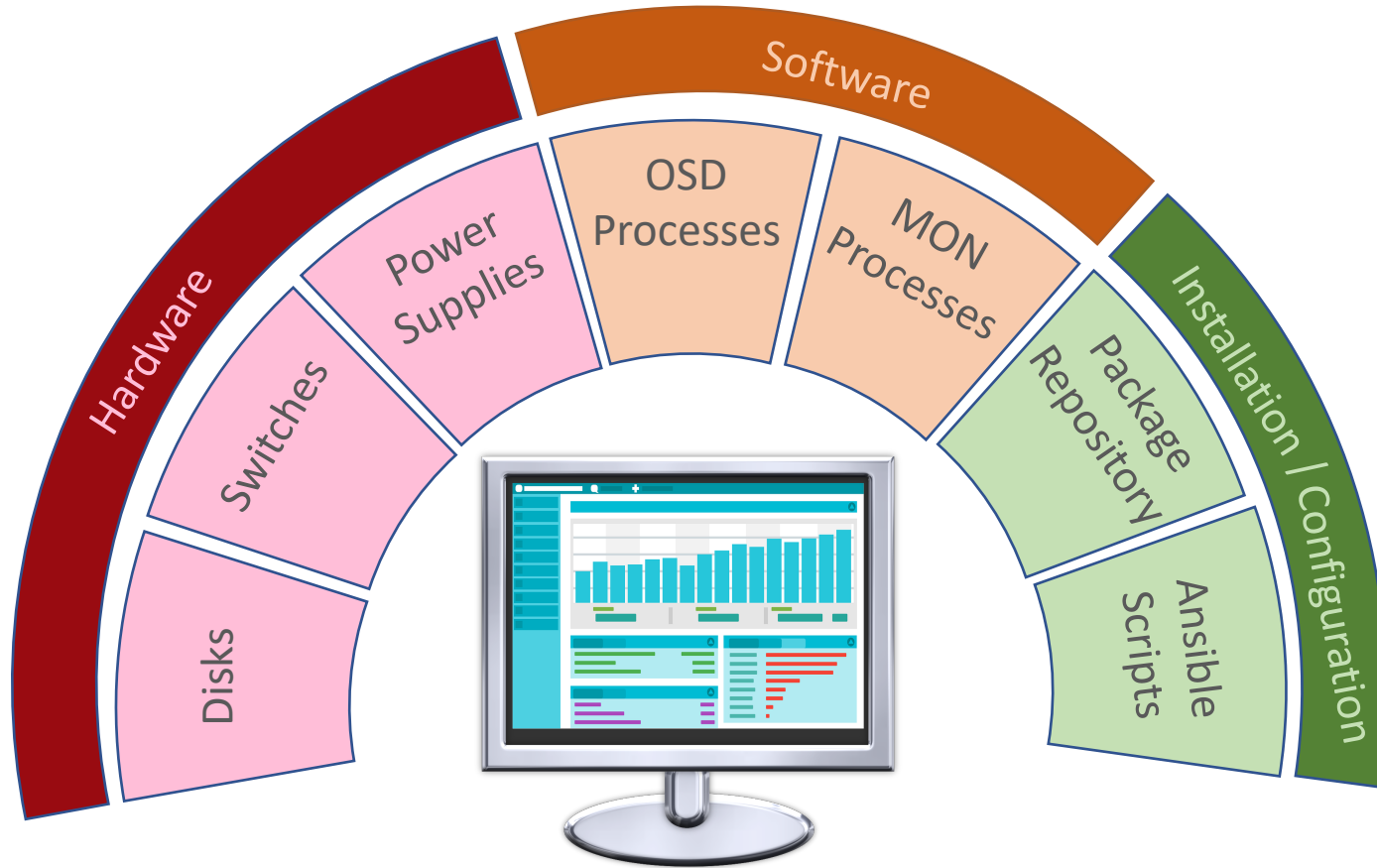
The icon shows a green cloud with a network diagram inside, consisting of a central square node connected to three other square nodes below it.



Performance problems are often Network problems

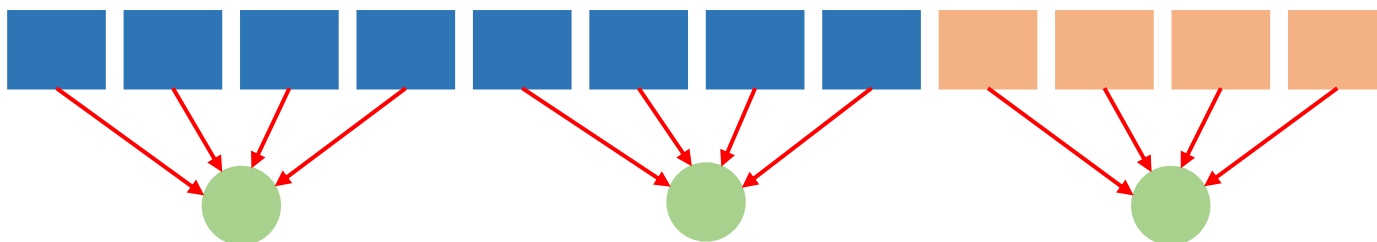
The icon depicts a yellow monitor with a clock face on the screen and a warning sign (a yellow triangle with a black exclamation mark) at the bottom.

Unified Management

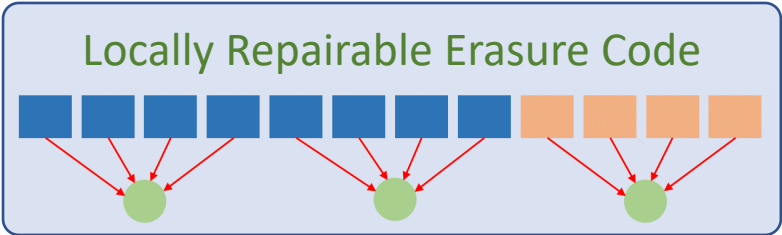
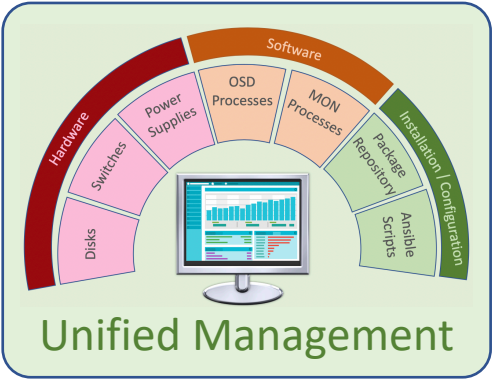
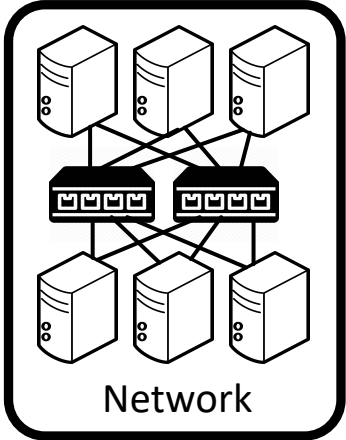
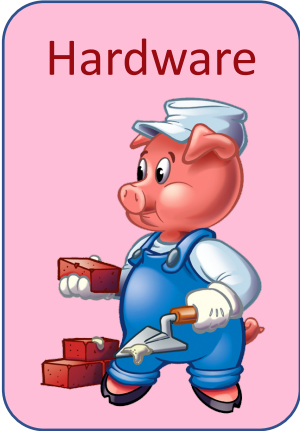
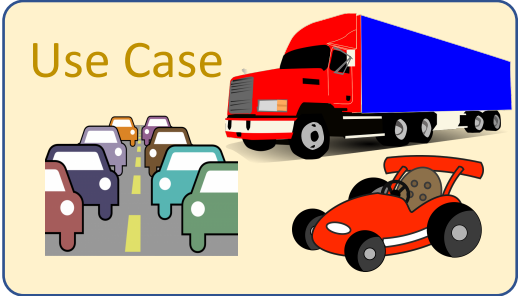


Locally Repairable Erasure Code (LRC)

- Create Local Parity Chunks
- Recover Using Fewer OSDs
- Faster Recover, More Overhead
- (*<data chunks>*, *<coding chunks>*, *<locality>*) – (k, m, l)
- Example: (8, 4, 4)



Summary



keepertechnology
access | manage | protect



Thank You

