Seagate Kinetic Open Storage Platform

James Hughes

…and many others

Seagate: Fibre Channel? RAID? SATA? File System? All RUBBISH

App to disk via ethernet, baby. The rest of you, clear out your desks

By Chris Mellor, 22nd October 2013

Seagate is building hard disk drives with a direct Ethernet interface and object-style API access for scalable object stores, a plan which - if it works - would destroy much of the existing, typical storage stack.

Drives would become native key/value stores that manage their own space mapping with accessing applications simply dealing at the object level with gets and puts instead of using file abstractions.

Seagate says it has developed its Kinetic technology because the existing app-to-drive storage stack is clumsy, inefficient and delays data access. Put an Ethernet interface

RELATED STORIES

What's the first Kinetic Ethernet hard drive? Psst, it's the 4TB Terascale

Seagate to

MOST READ MOST COMMENT

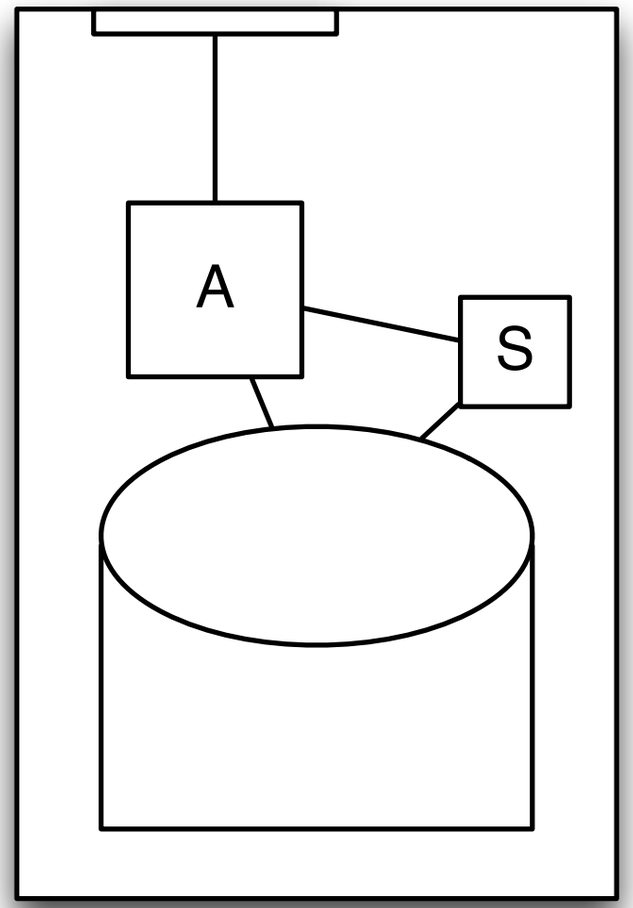
LG G Flex: A new cheeky curvy mobe wit 'SELF-HEALING' bottom

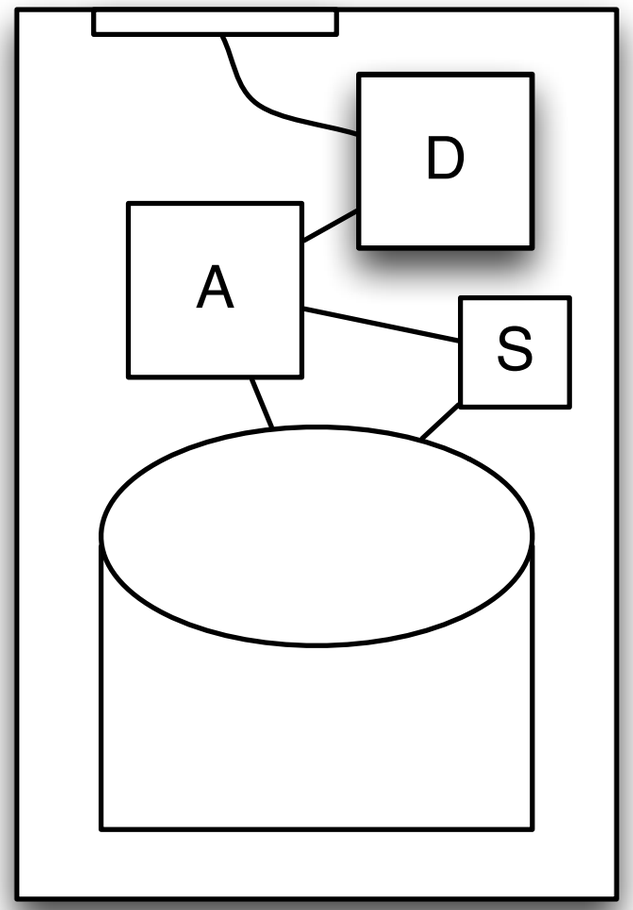
Everything's going to be all white: Google Nexus 5 mobe expected Friday

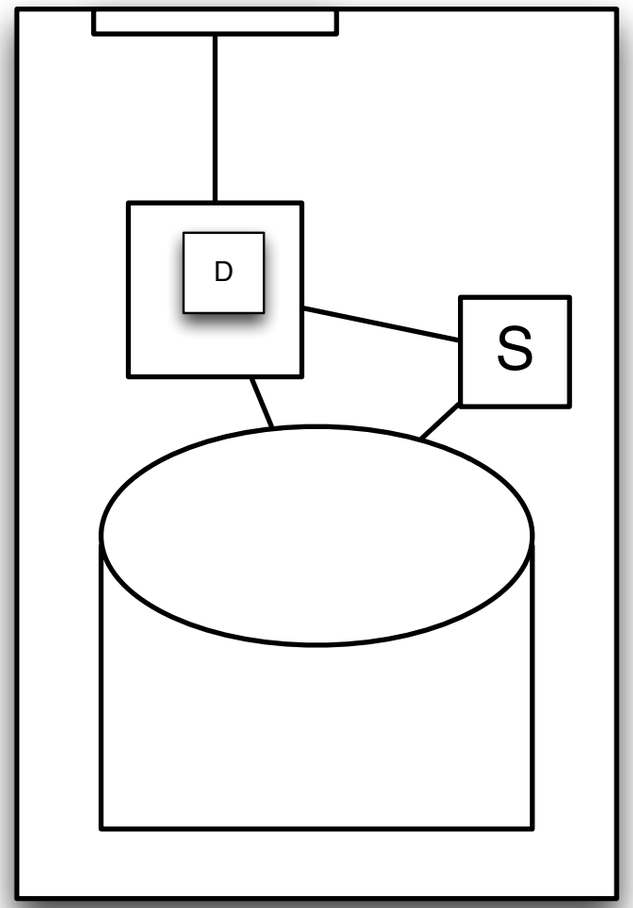
In a meeting with a woman? For pity's sak

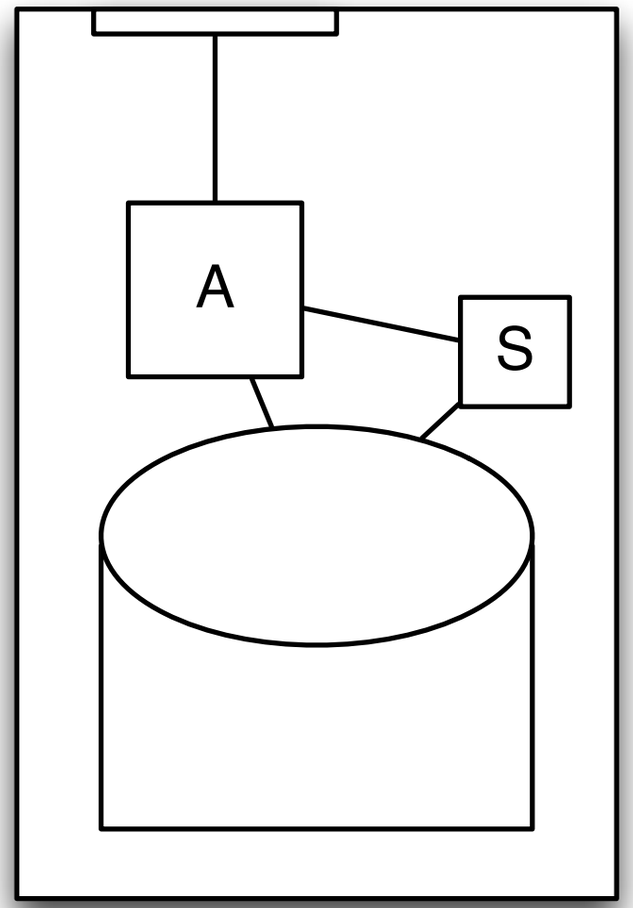
Seagate is building hard disk drives with a direct Ethernet interface and object-style API access for scalable object stores, a plan which - if it works - would destroy much of the existing, typical storage stack.

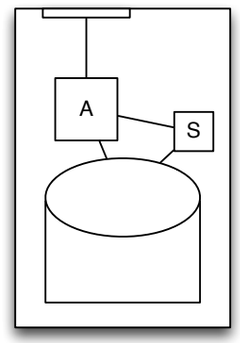
Drives would become native key/value stores that manage their own space mapping with accessing applications simply dealing at the object level with gets and puts instead of using file abstractions.

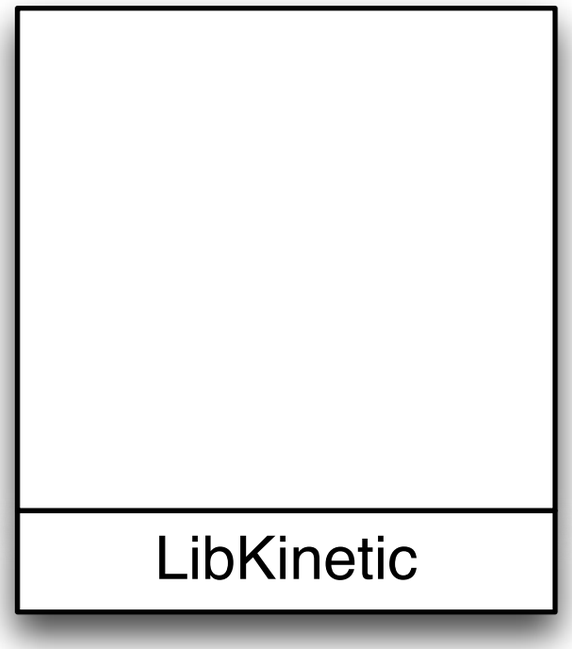

A
S

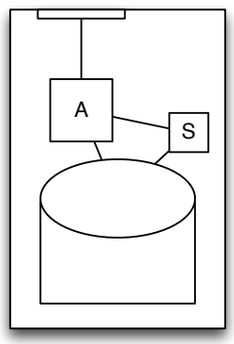

D
A
S

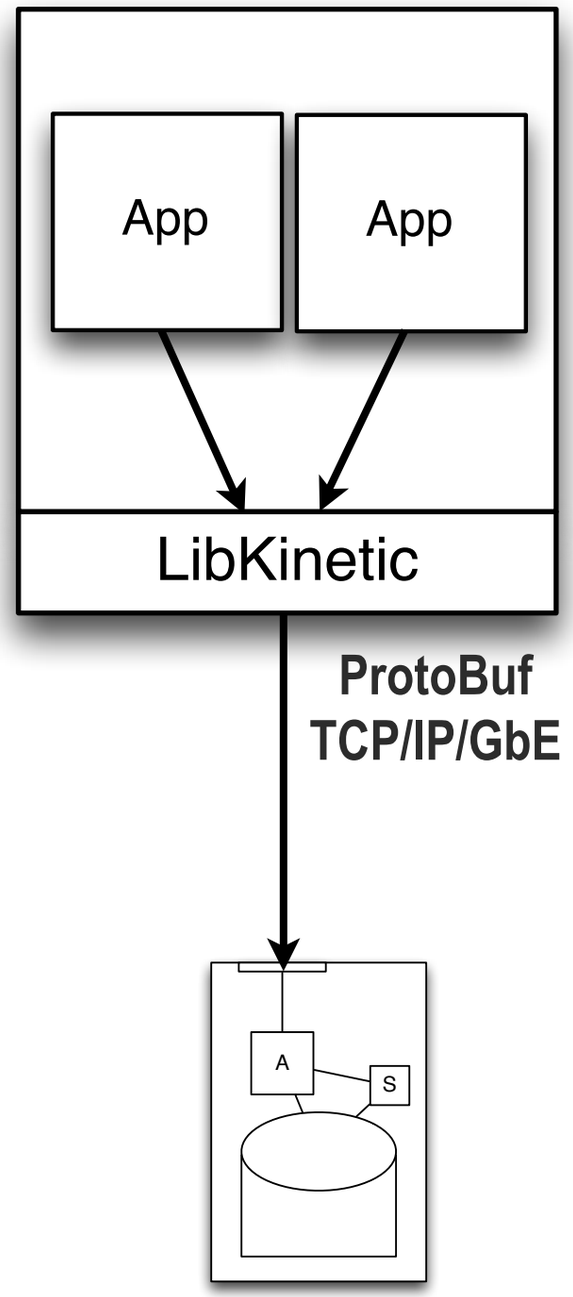

D
S


A
S

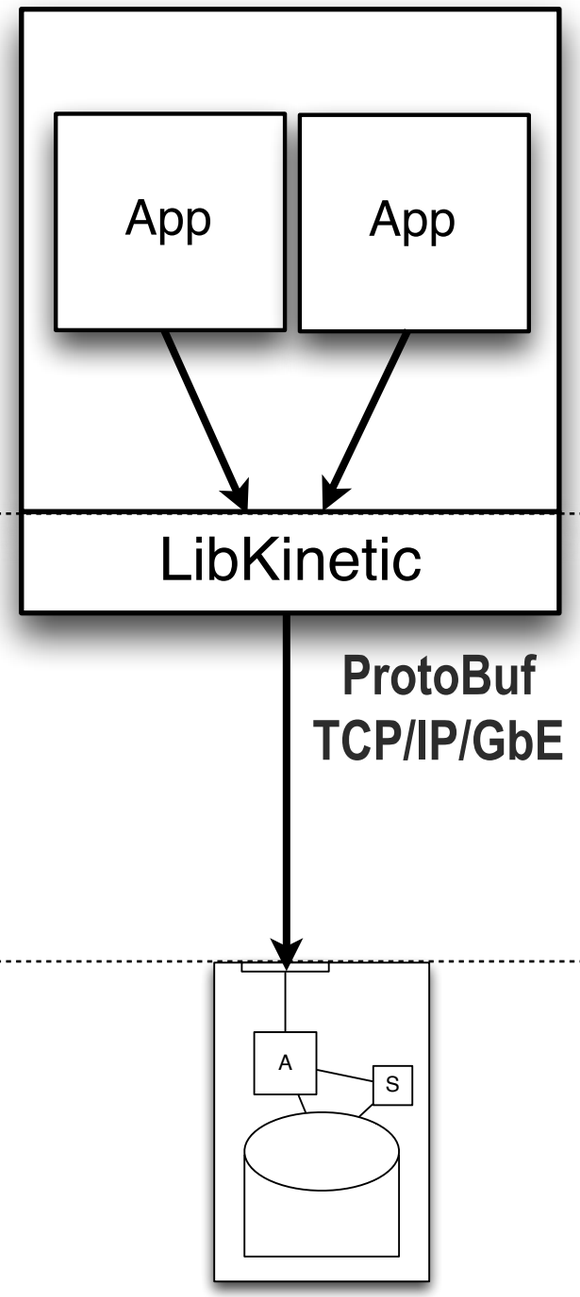

A
S


LibKinetic


A
S


App
App
LibKinetic
ProtoBuf
TCP/IP/GbE
A
S

- Application
- Clustering
- Management



- Interconnect

- Storage

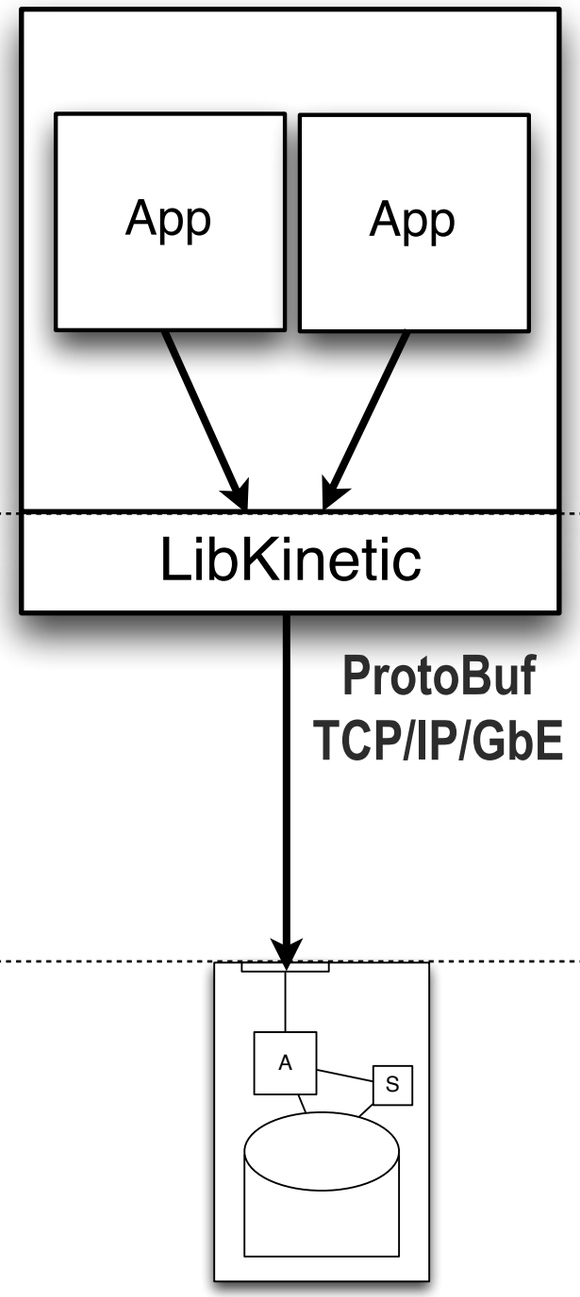
Proprietary
to System Vendor

GPL
Standard

Proprietary
to Seagate

- Application
- Clustering
- Management

Proprietary to System Vendor



C++, Java, Python, Erlang, DIY

- Interconnect

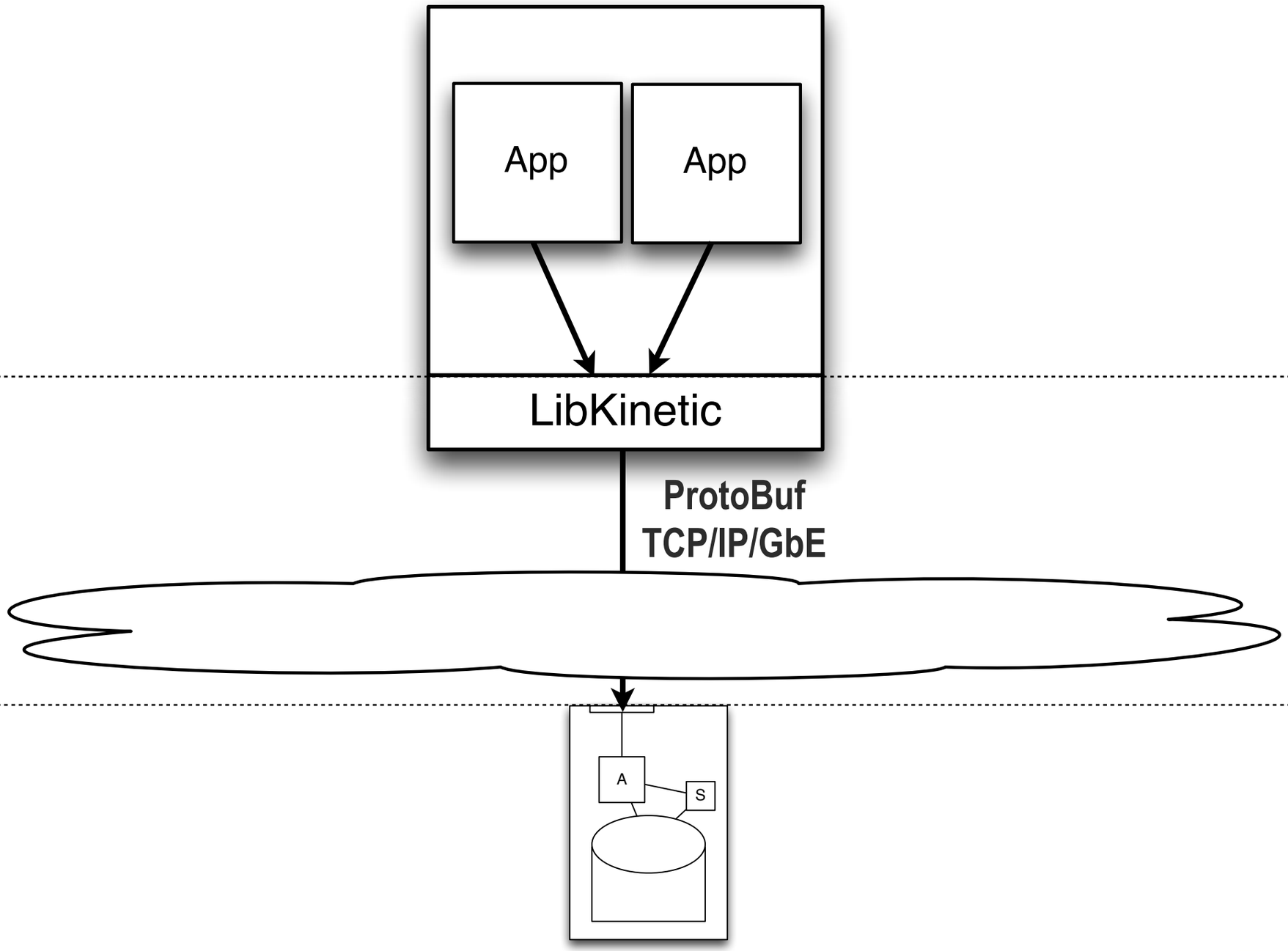
GPL Standard

- Storage

Proprietary to Seagate

- Application
- Clustering
- Management

Proprietary to System Vendor



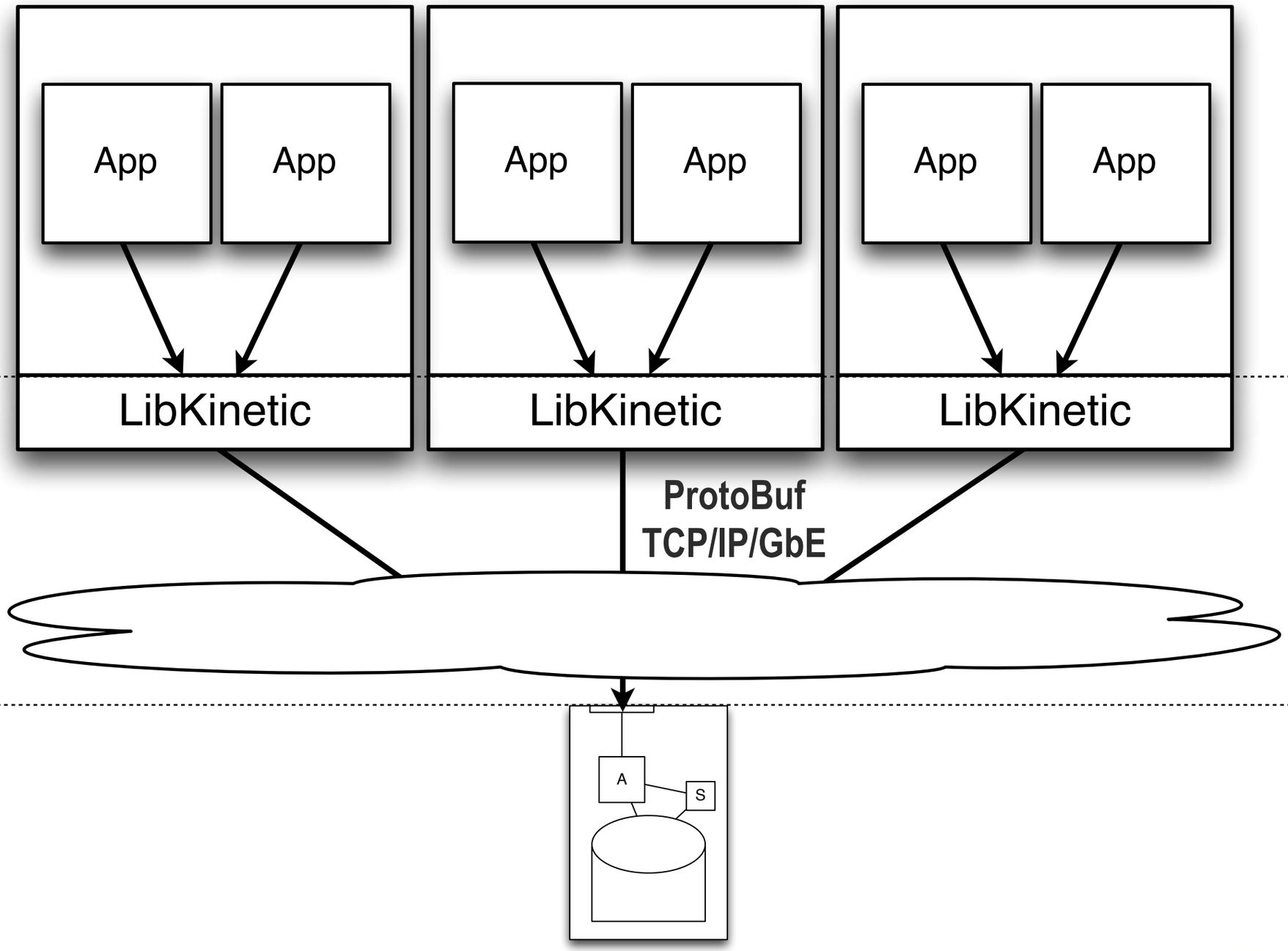
- Interconnect

GPL Standard

- Storage

Proprietary to Seagate

- Application
- Clustering
- Management



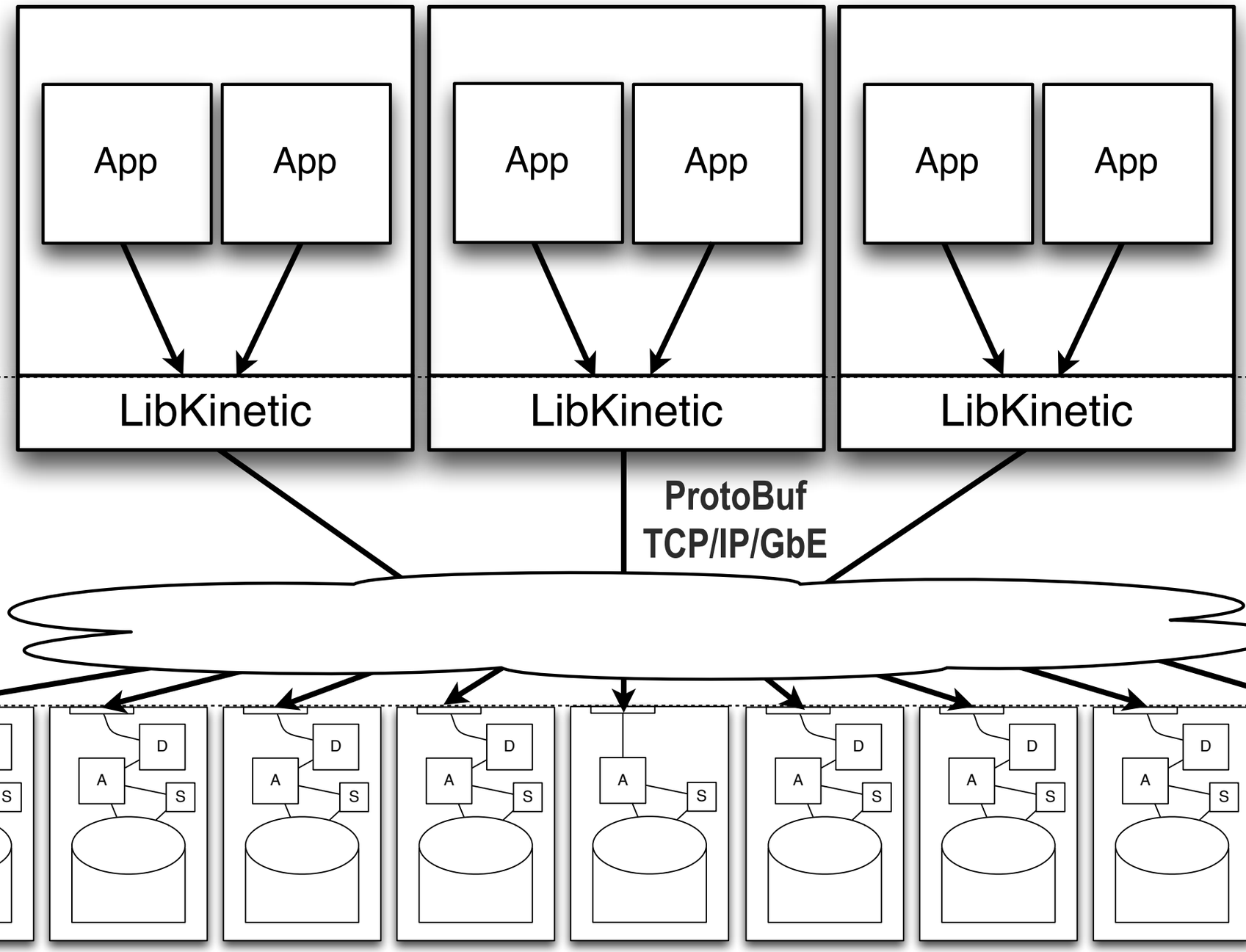
- Interconnect
- Storage

Proprietary to System Vendor

GPL Standard

Proprietary to Seagate

- Application
- Clustering
- Management



- Interconnect
- Storage

Proprietary to System Vendor

GPL Standard

Proprietary to Seagate

SAS versus Kinetic Open Storage

- Standard form factor
- 2 SAS ports
- SCSI command set
  - data = read (LBA, count)
  - write (LBA, count, data)
  - LBA :: [0, max]
  - data :: count * 512 bytes
  - CRC on cmd and PI on block

- Standard form factor
- 2 Ethernet ports (same connector)
- Kinetic key/value API
  - value = get (key)
  - put (key, value)
  - delete (key)
  - key :: 1 byte to 4 KiB
  - value :: 0 bytes to 1 MiB
  - HMAC on cmd and SHA on value

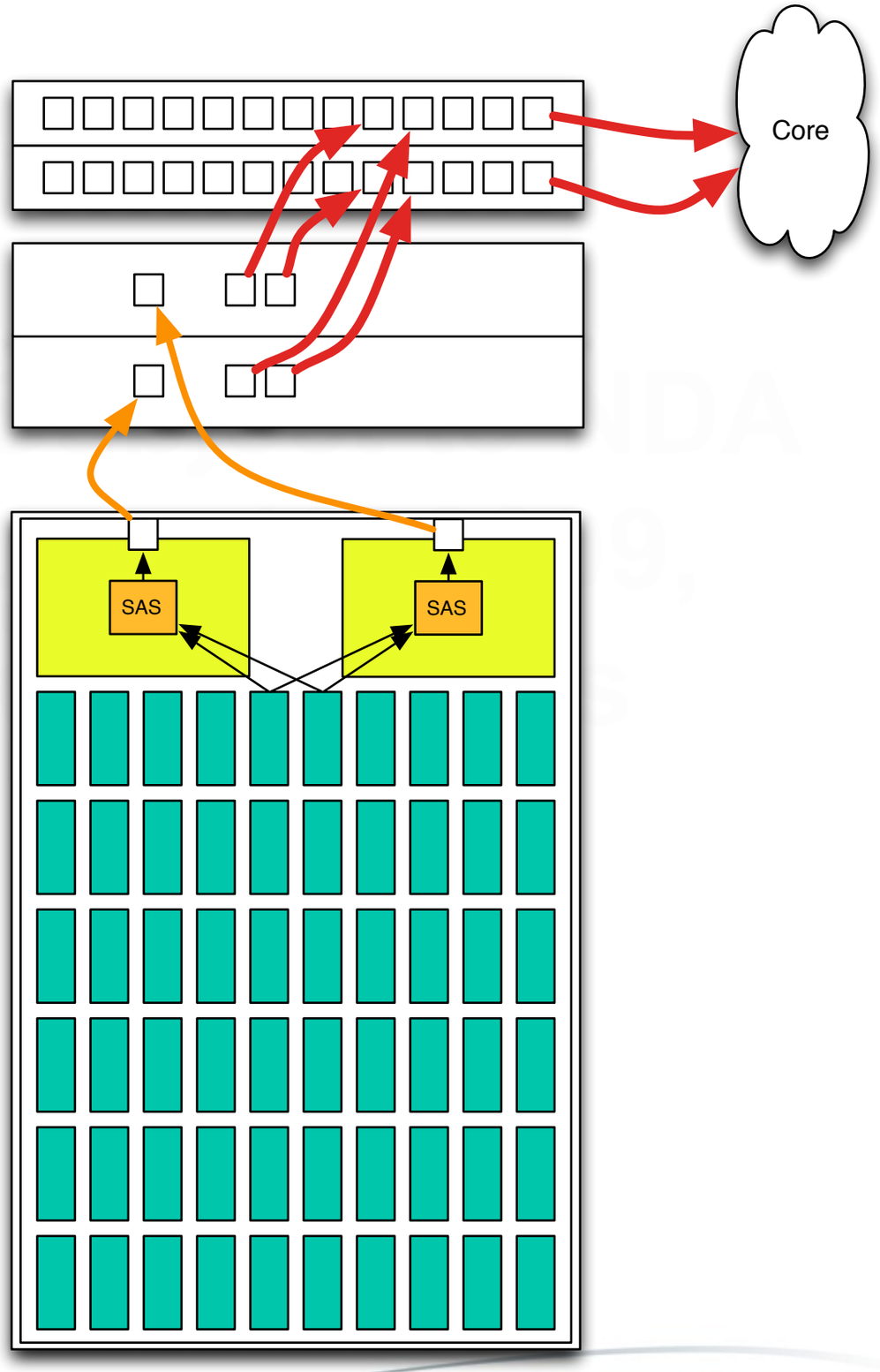
Typical HA High Density

Intel server

- Double Socket
- 48GB Ram
- 1000w

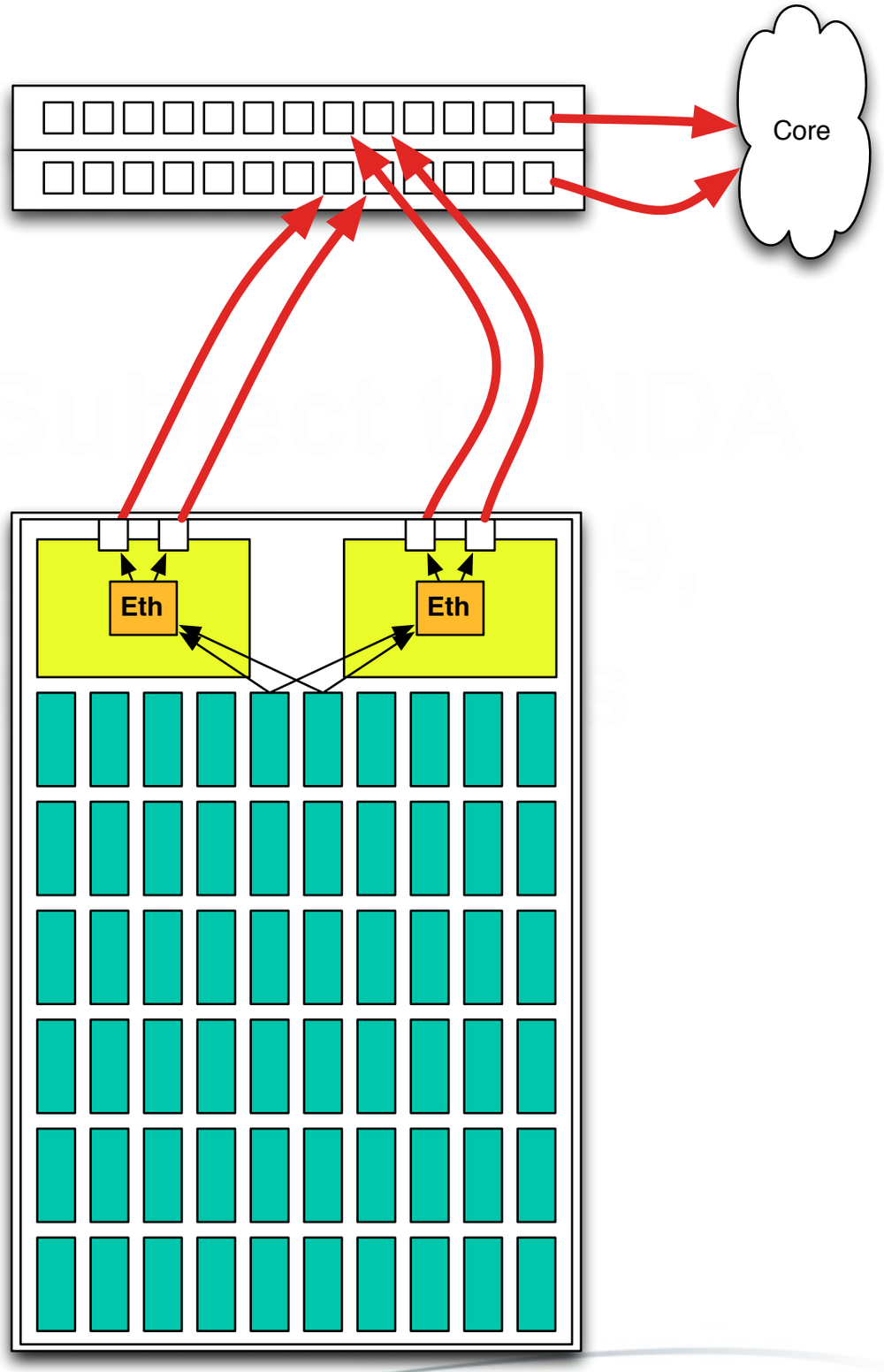
SAS tray

- Connected to the server



Low cost HA Configuration

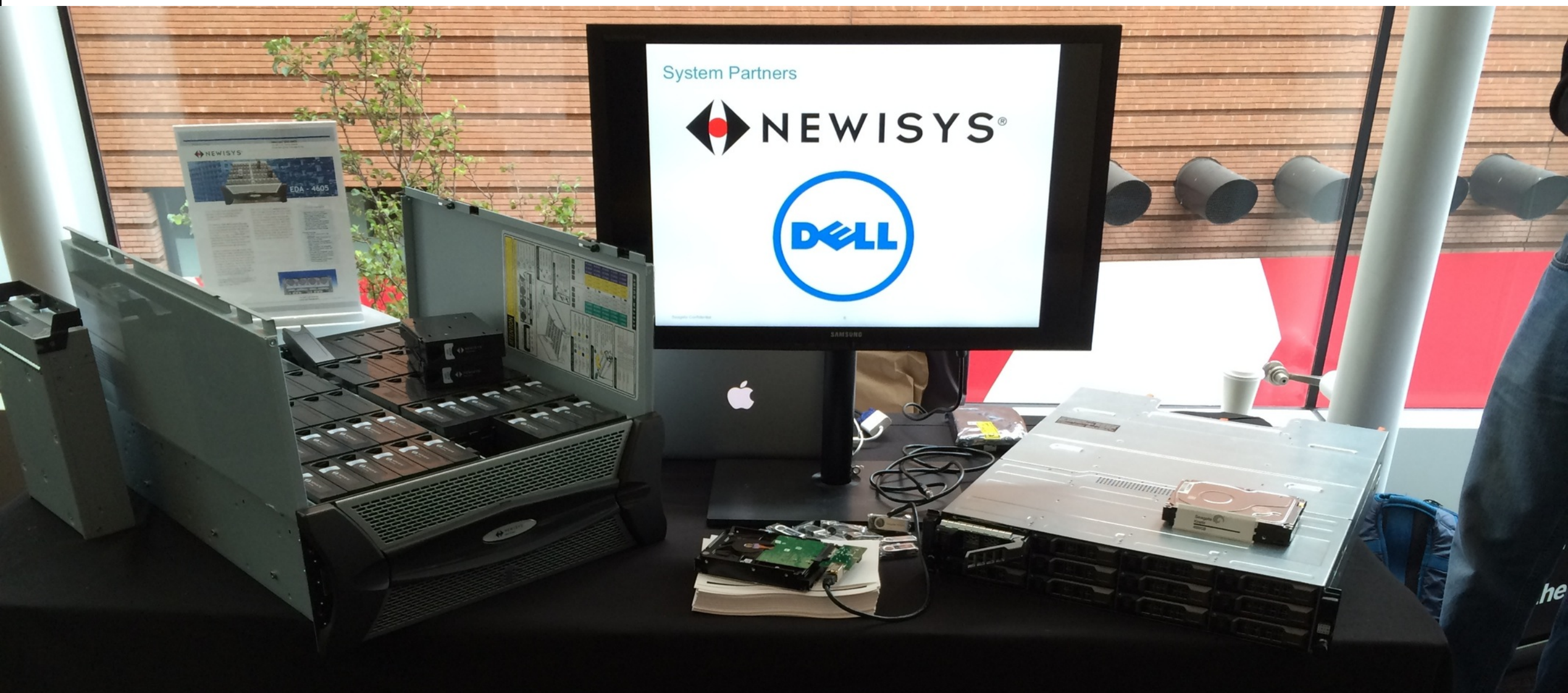
- Each drive talks to both switches
- Each switch has 2 by 10Gb/s Ethernet
- Kinetic Tray talks directly to ToR
- No servers



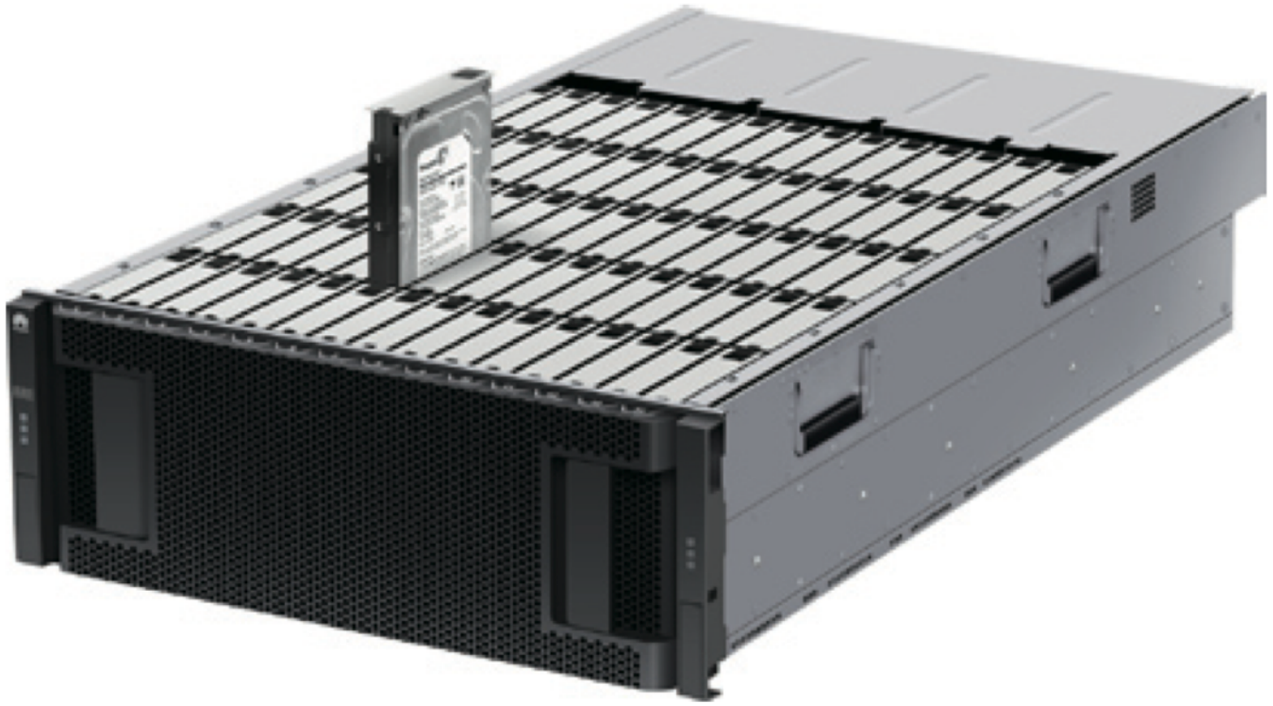
System Hardware

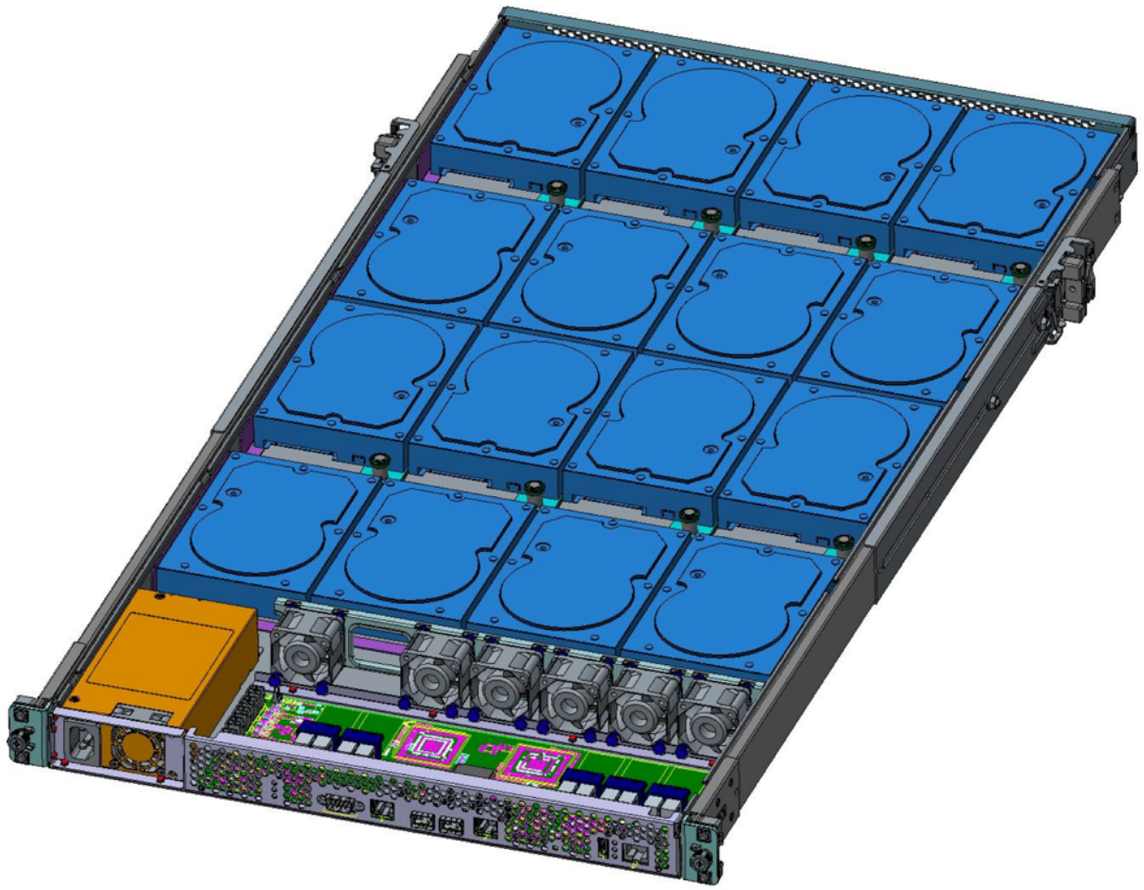
Typical JBOD architecture

- Does not require a server, just JBODs to the ToR Switch
- 10 JBODS × 60 drives × 4TB = 2.4PB/Rack



System Hardware





Kinetic *Drive*

Provides RPC to Key/Value database

- Data is pre-indexed
- Compression and other value is easy and transparent

P2P (Drive to Drive) copy of key ranges

Communicate using existing Data Center Plumbing (TCP/IP)

Multiple masters - Data sharing between machines

Configurable caching per command

- WriteThrough, WriteBack, Flush

Local space management

Kinetic *Systems*

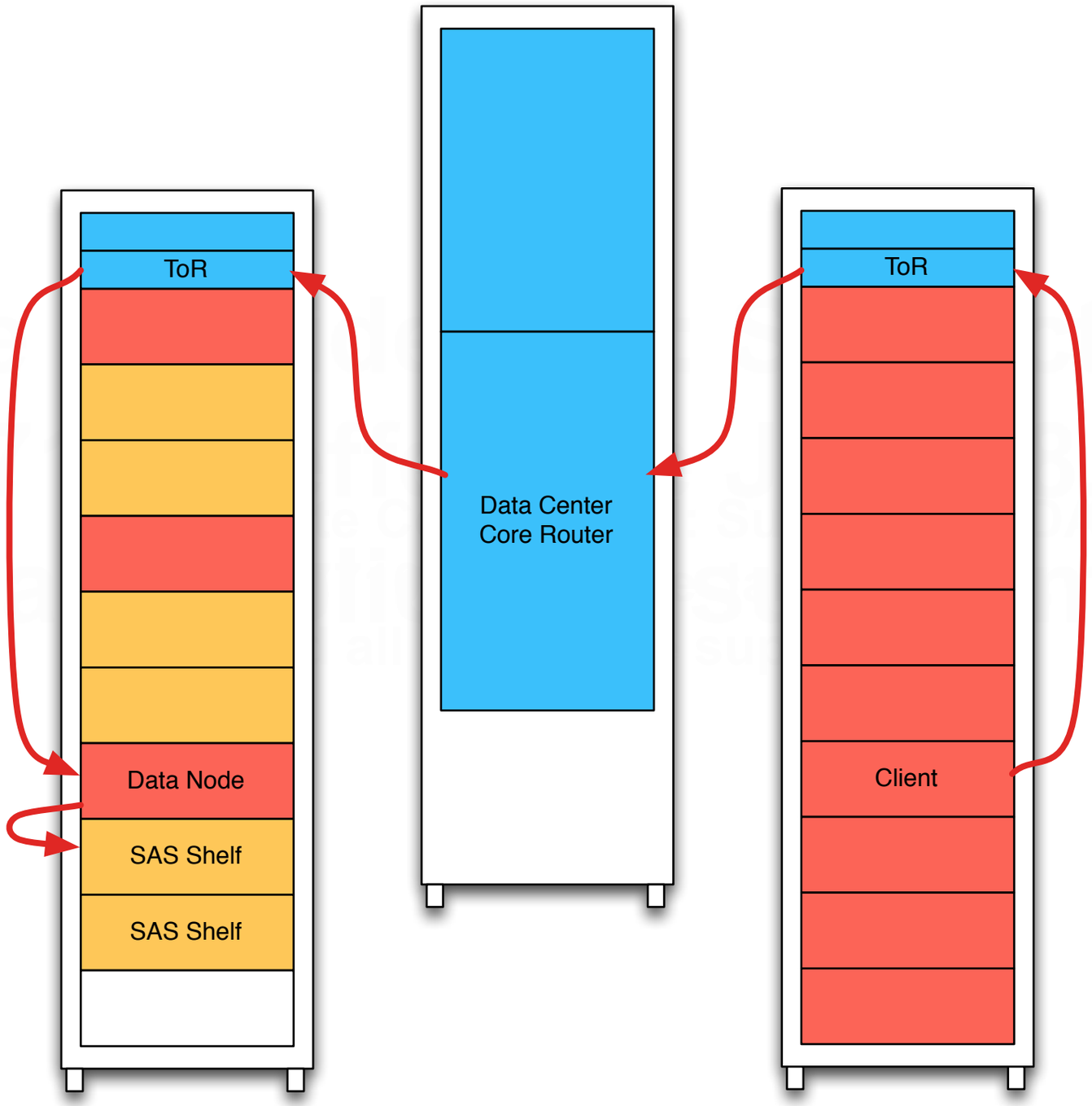
Clustering (performance, reliability, management)

Compatibility with large scale applications (S3, etc.)

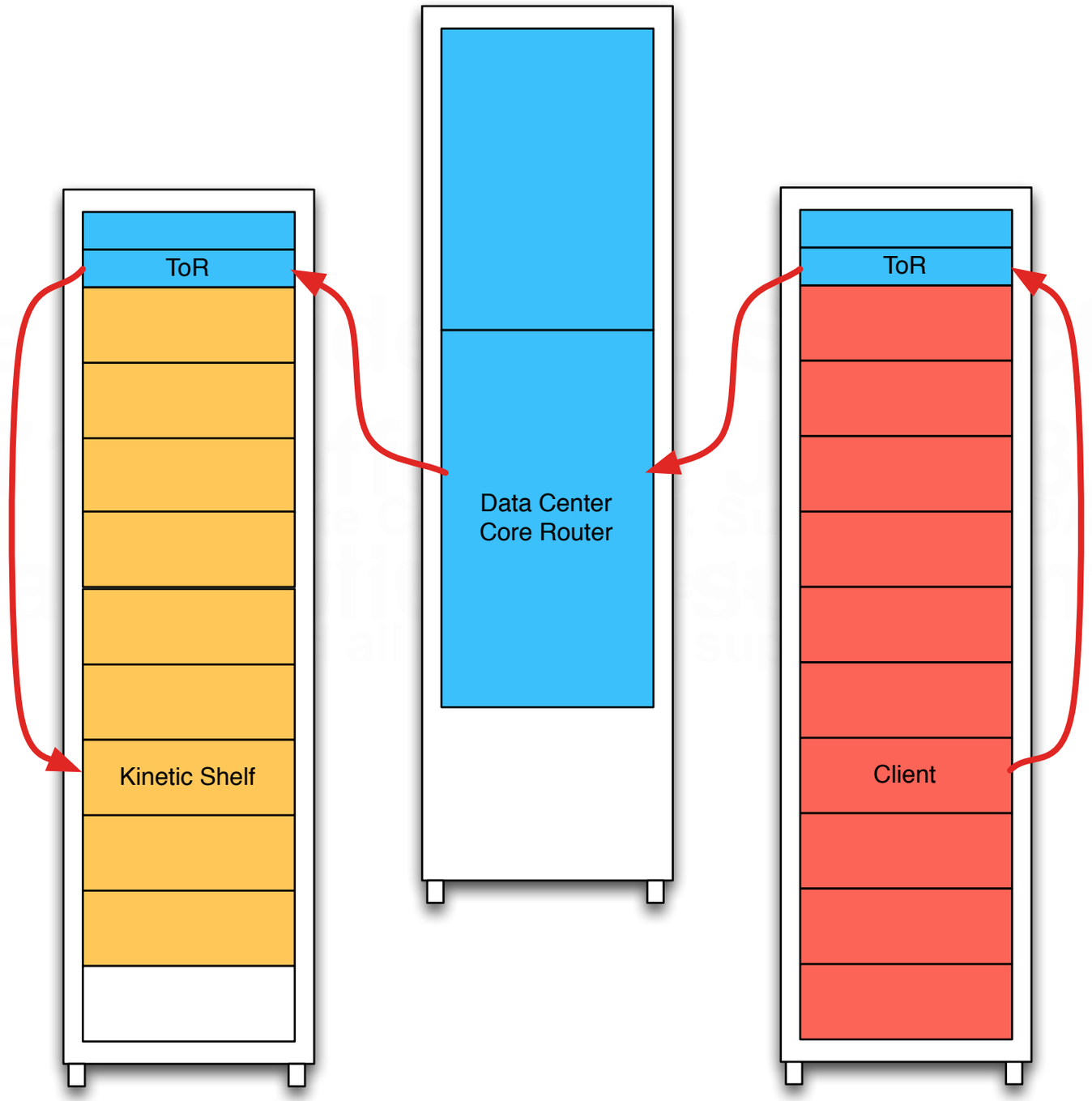
Centralized Management

- Reliability, availability, durability

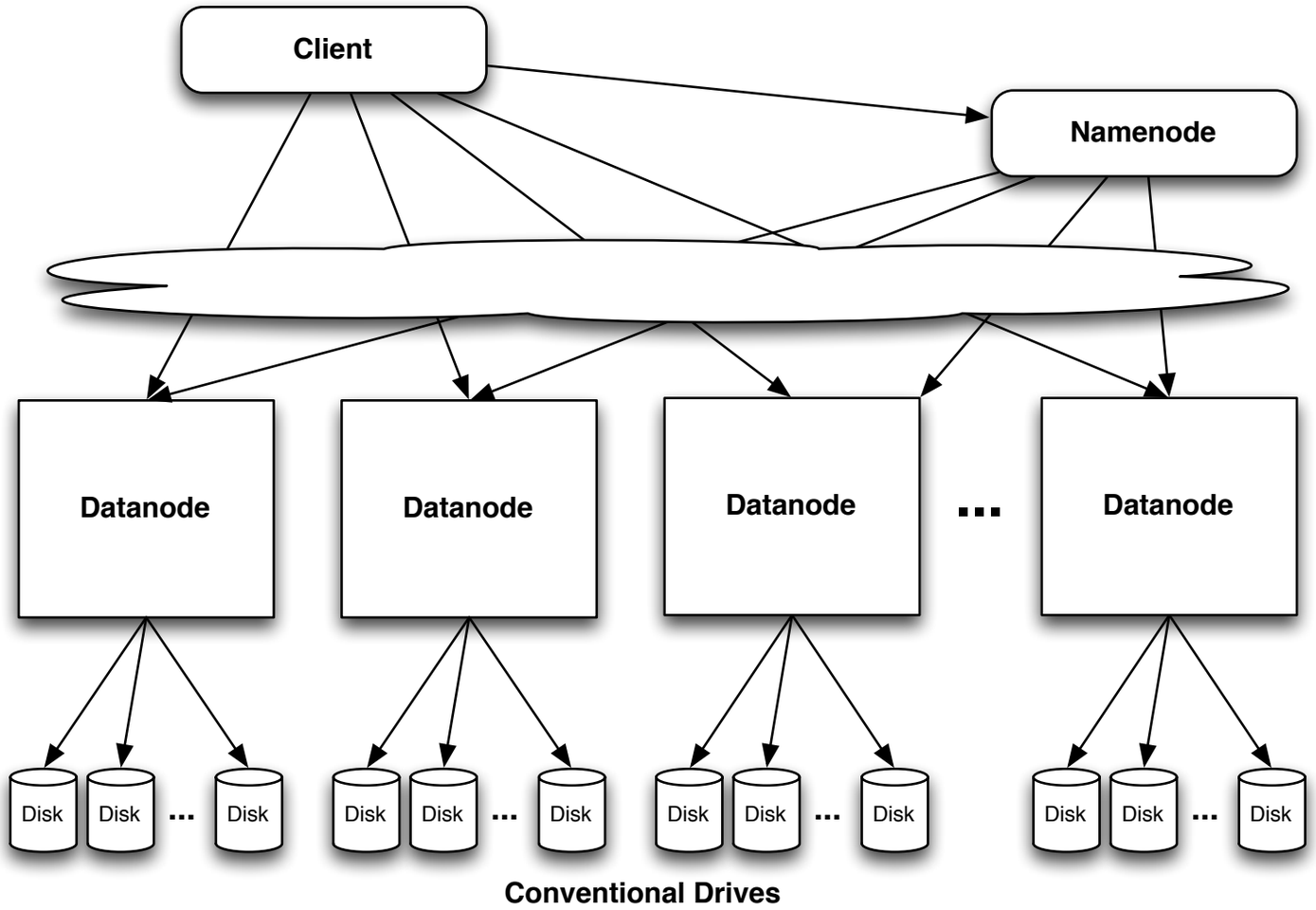
Existing Traffic Flow



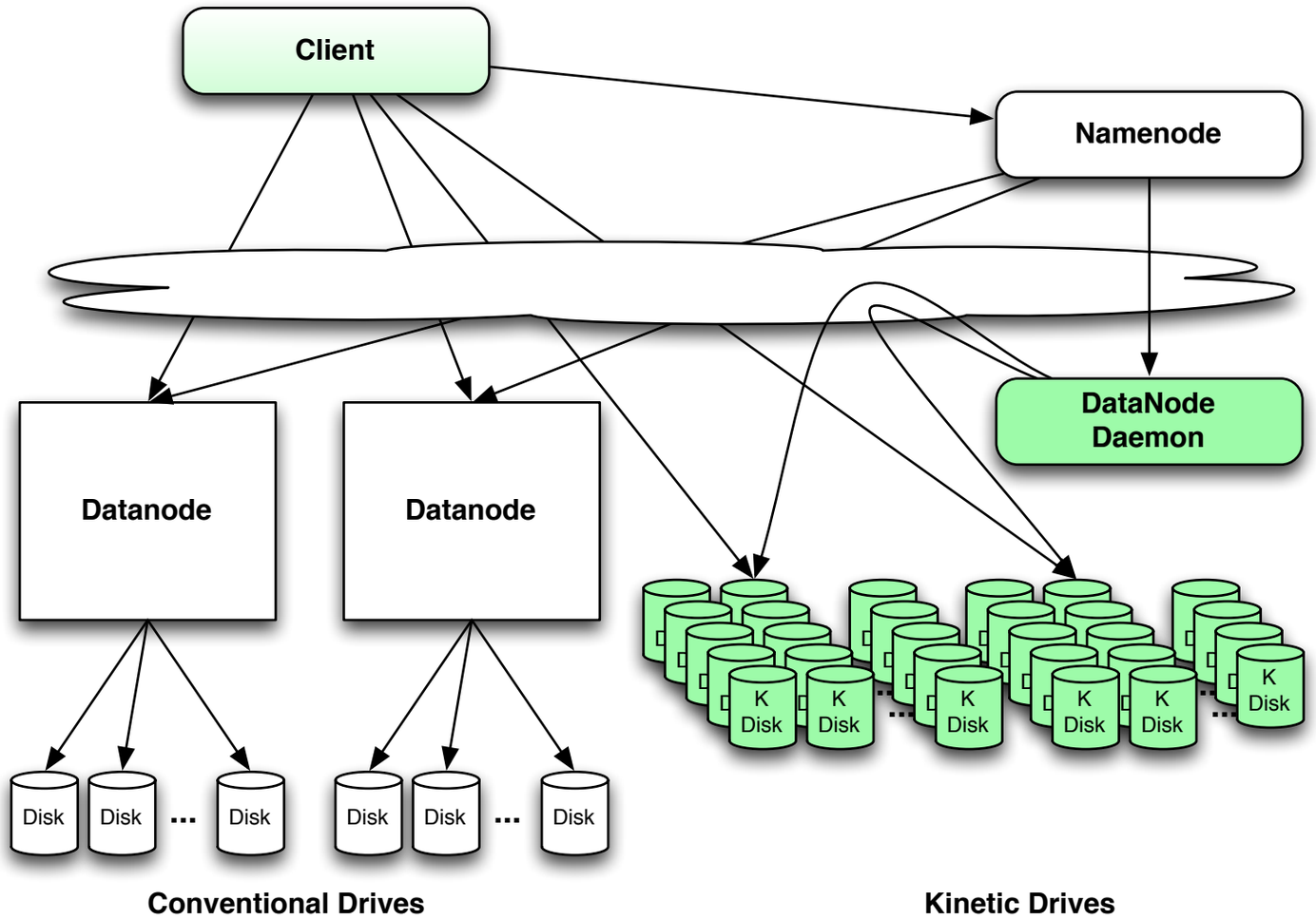
Kinetic Traffic Flow



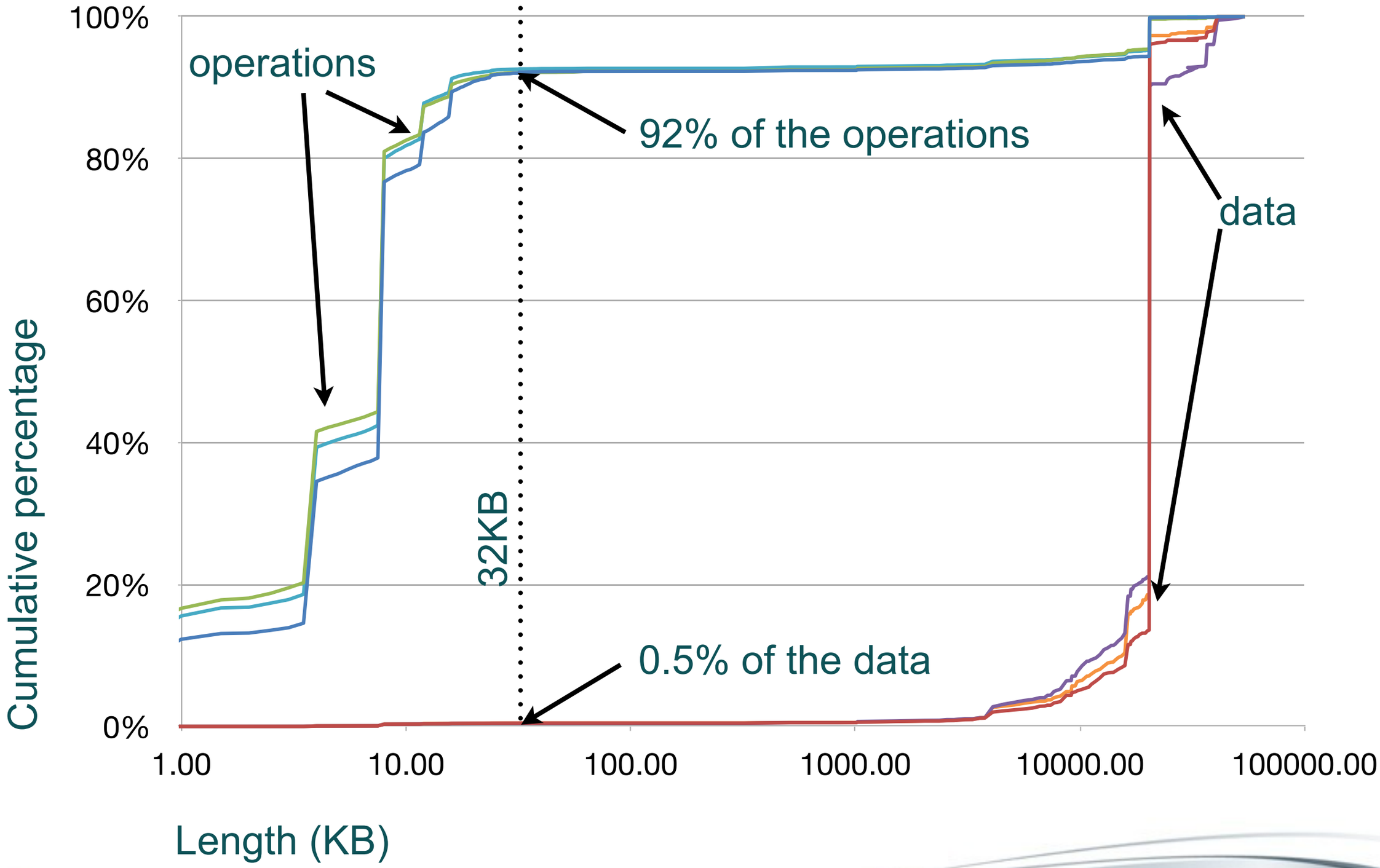
Conventional HDFS System



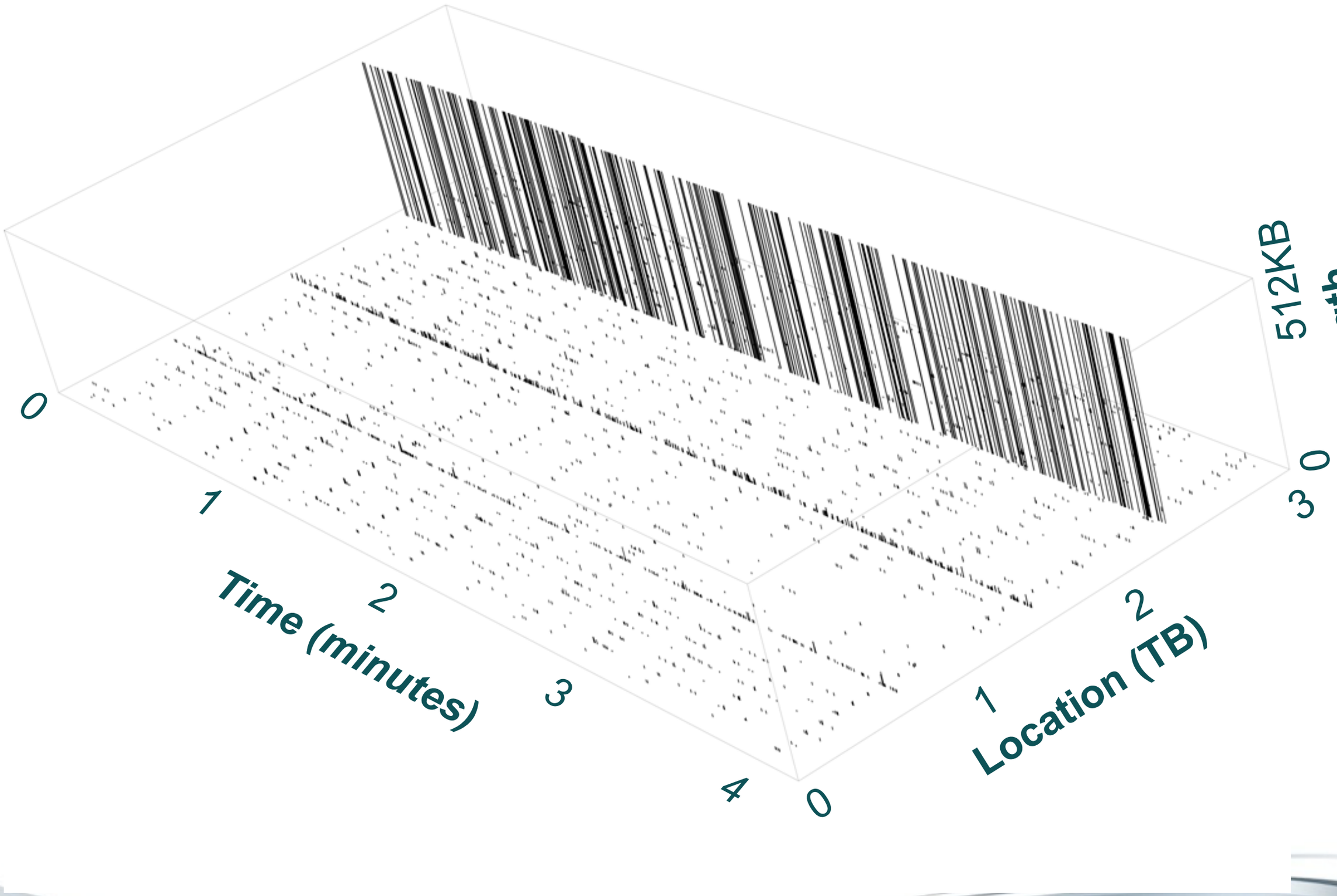
HDFS on Kinetic



Cumulative operations ordered by length



Map of Operations

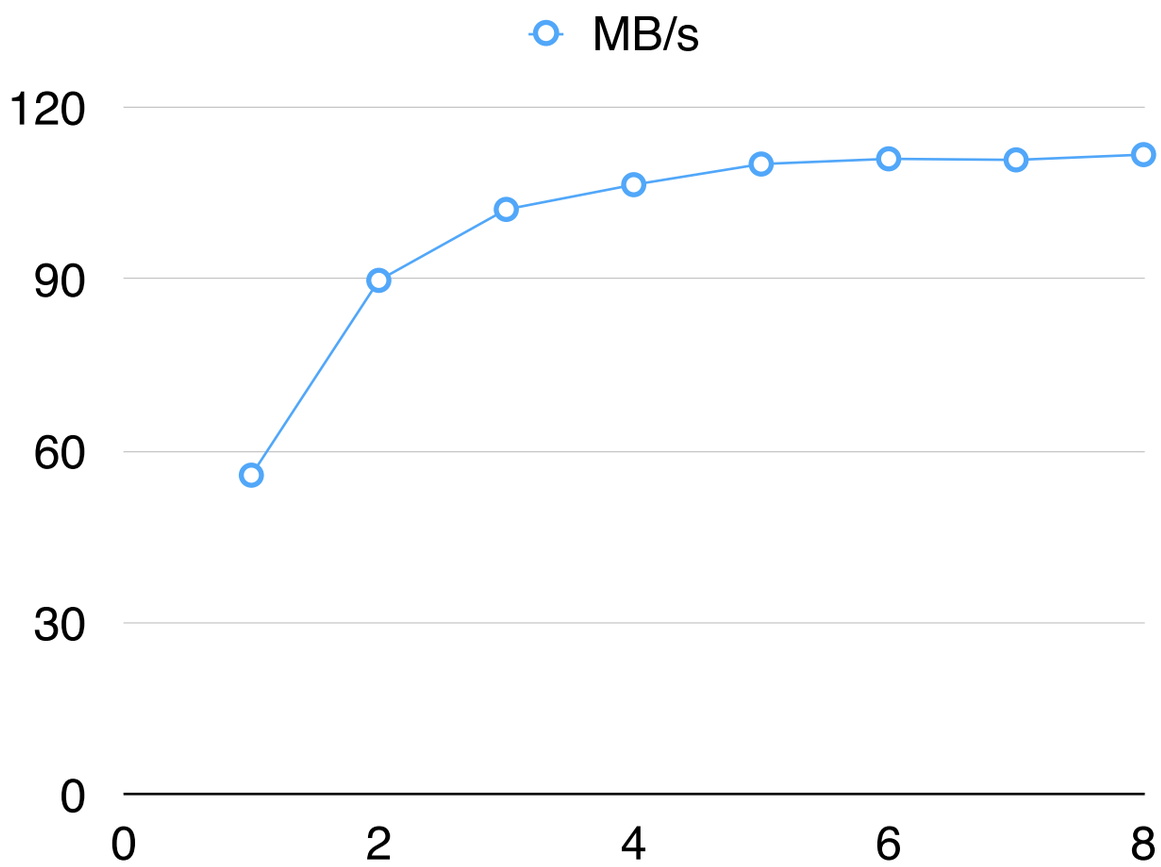


Performance Metrics

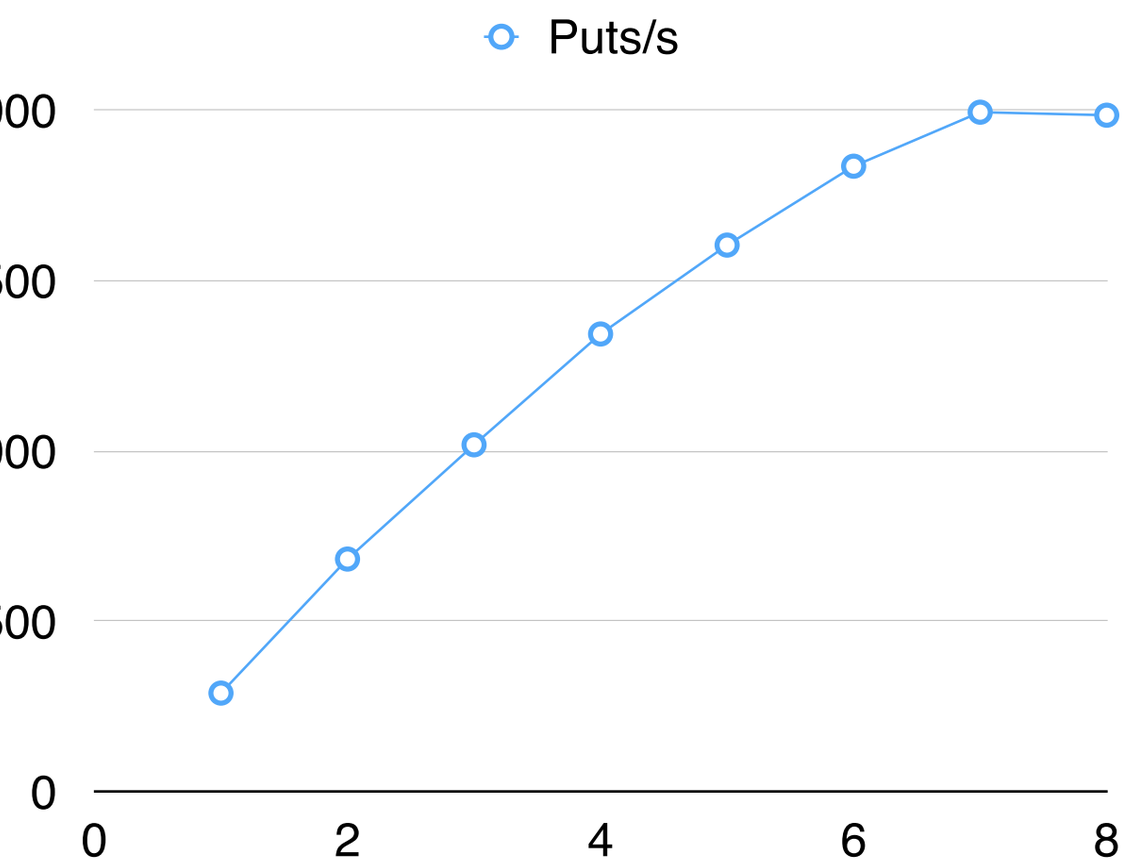
Same normal performance expectations

- Sequential Write: 50MB/s
- Random Write: 50MB/s
- Sequential Read: 50MB/s
- Random Read: 20% slower than traditional drives

Write Performance Results



1MB values put rate (MB/s)



10B value put rate

Goals of API

Data movement

- Get/put/delete/getnext/getprevious
- Versioned (== for success), options

Range operations

Multiple masters

- Authentication/Integrity/Authorization

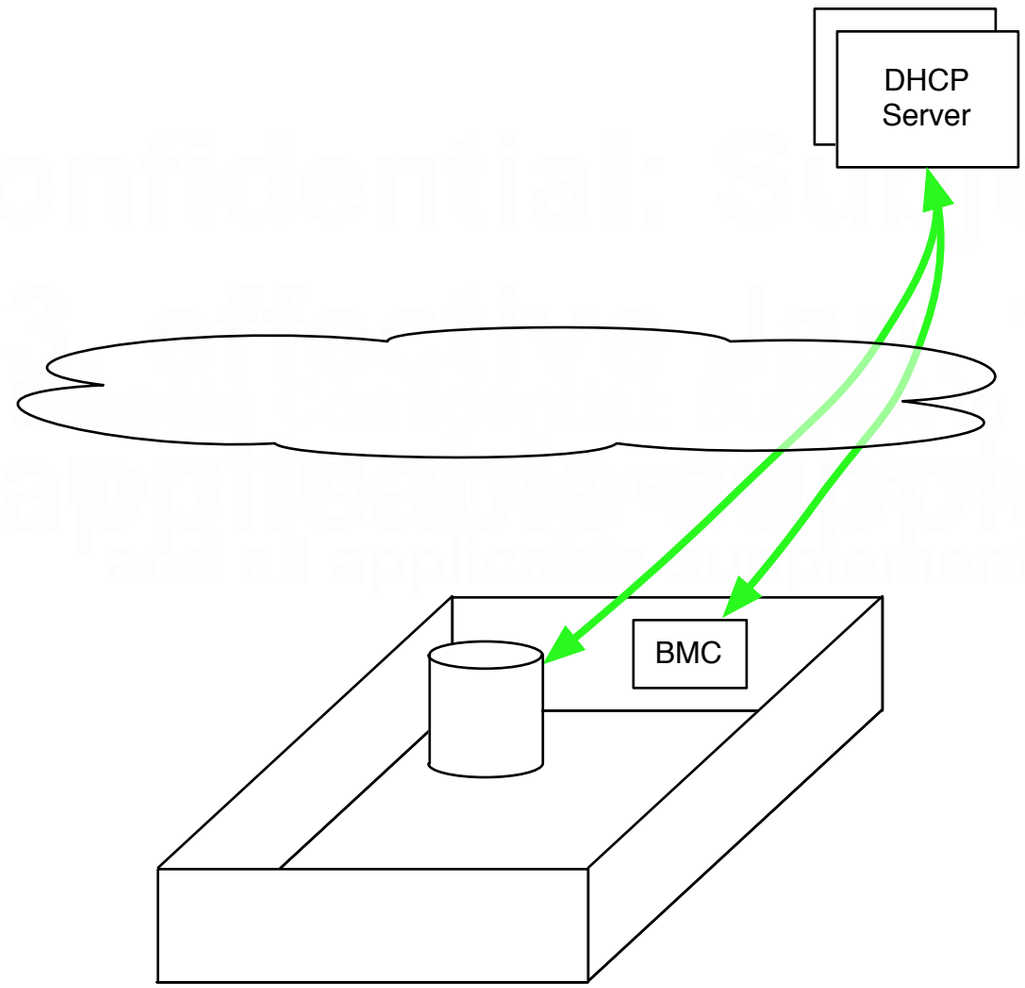
Cluster-able

- Simple cluster configuration version enforcement

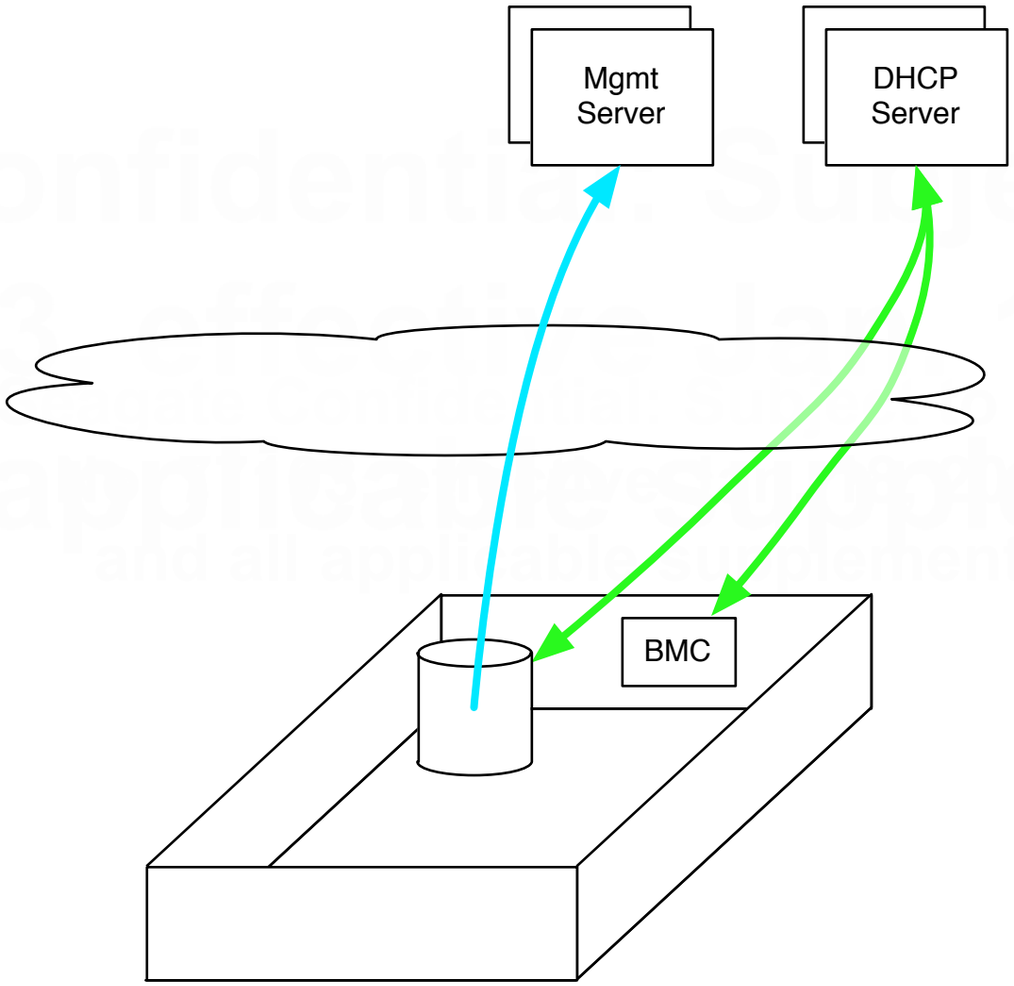
P2P copy

Management

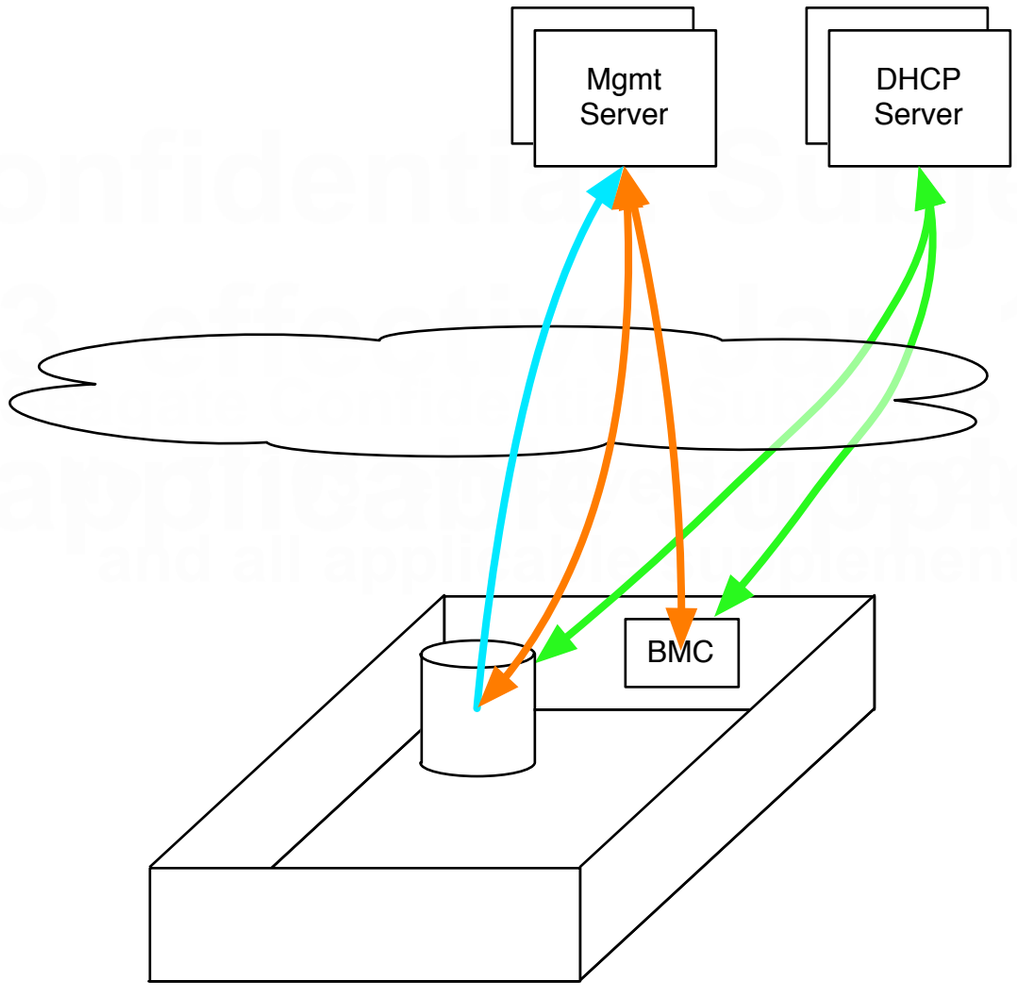
Bootstrapping devices



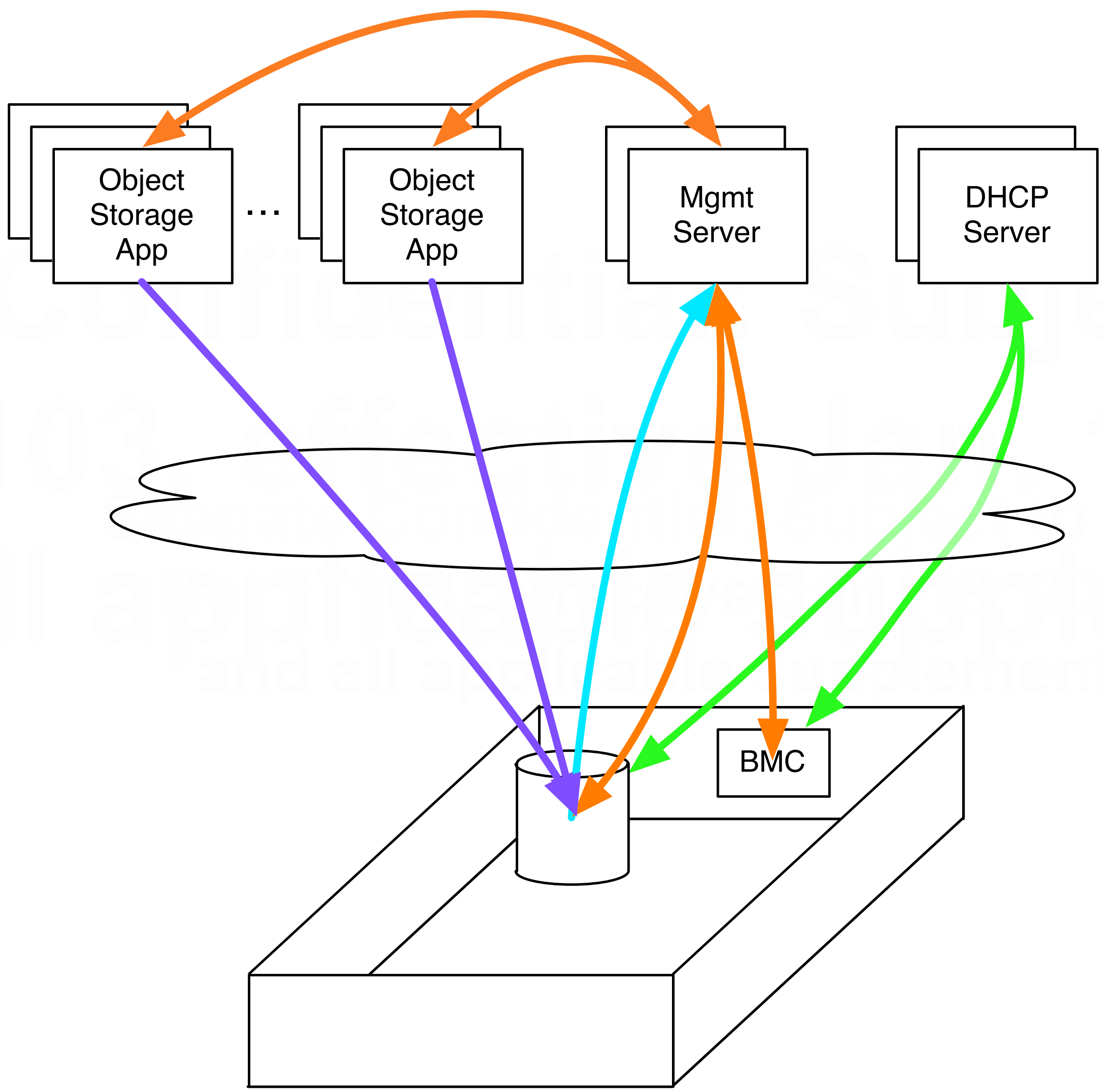
Bootstrapping devices



Bootstrapping devices



Bootstrapping devices



Conclusion

Next Generation Storage Devices

- Dis-intermediates cloud applications to drive
- Enable innovation in hardware and software ecosystem
- Lower TCO

Integration with:

- Swift
- HDFS
- Basho Riak
- Ceph
- Scality

More information

- http://seagate.com/www/kinetic
- https://developers.seagate.com
- http://github.com/Seagate

Seagate Confidential: Subject to NDA
and all applicable supplements