Defining Software Defined Storage

Laz Vekiarides
Entrepreneur and Technology Executive

Agenda

- Storage economics - misconceptions
- A working definition
- Use cases

- Disclaimer: Opinions are my own!

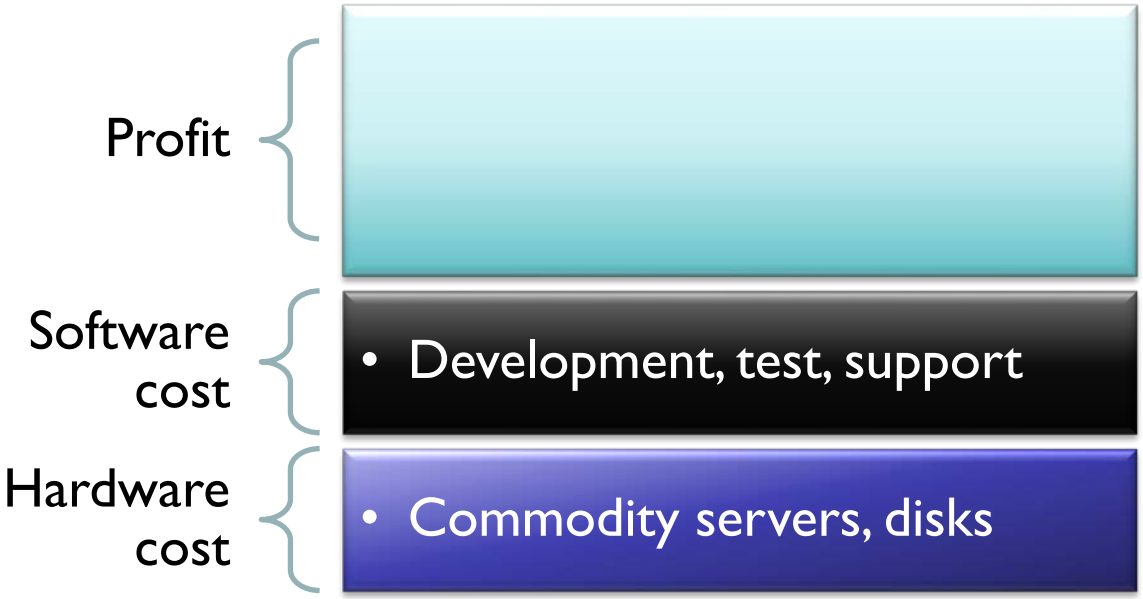
The New New Idea

- Software and hardware should be de-coupled
- Hardware is commodity servers and disks
- Software should run data management on commodity hardware

Is this really new thinking?

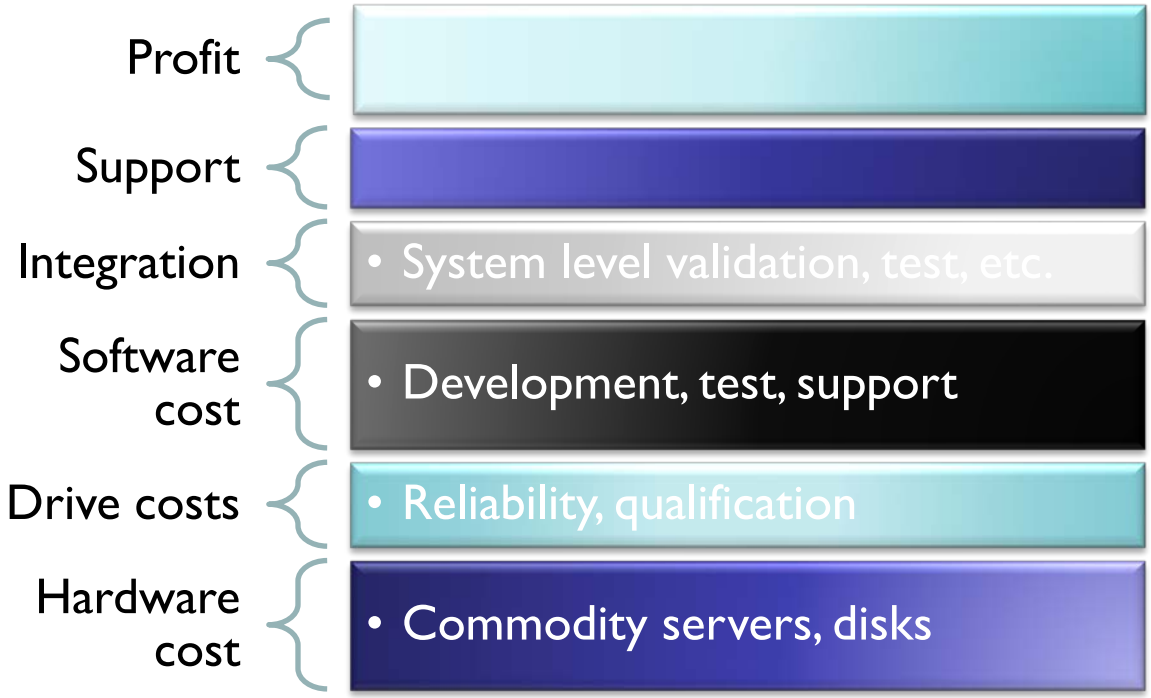
The Dissection Fallacy

Typical mistaken view of Storage Array Economics:



Reality is complicated

A more accurate breakdown:



The “Commodity” Fallacy

≠

SDS: An Aspirational Definition

Current working definition

- Software Defined Storage
  - Software-based storage solution comprised of:
    - Software
    - Commodity hardware (i.e. rack servers, DAS)
    - Standards-based non-commodity hardware
  - Offering:
    - Storage service not tied to the physical platform SDS is running
    - Virtualization and migration amongst physical platforms
  - And is generally tied to compute virtualization

Use cases: Cloud archive

- Unabated data growth in numerous industries
  - Data needs to be stored, organized and archived for long periods
  - Pressure to lower OPEX - floor space, cooling costs, FTE etc
  - Unstructured data (MRI results, Video Surveillance Records)
  - Regulatory Compliance
- Systems optimized for cost and ingest rate
  - Write-only storage
- Probably NOT SDS

Use case: Hosted virtual app

- Small and Mid sized companies moving to Public Cloud Storage
  - Small companies willing to utilize Cloud hosted SAN's for applications
- Generally for secondary storage, e.g. replication
  - Performance requirements moderate
  - Possibly DR in Cloud
- Driven by hosters – implemented as a VM on commodity servers
- Possible to achieve multi-tenant architectures via network virtualization

Use case: Private cloud app

- Proposed by large scale enterprise datacenters as a replacement for conventional SAN
  - High density compute environment
  - >200TB of server attached storage
  - On demand creation
- More rigorous performance and availability requirements
- Per-app, per-user multi-tenancy usually for chargeback
- Requires a deep orchestration and management framework.

Use case: Converged workloads

- Small business seeking a simple way to get started with virtualization
- Small/remote offices w/ footprint constraints
- Single integrated appliance
  - Focus on ease of use for IT generalist
  - Integrated HA/clustering with virtualization stack
  - Ability to scale out
  - Storage stack integrated with hypervisor
- Spoiler Alert: Some have proprietary HW
  - Can be hard to differentiate from conventional arrays.

Use case: Unstructured data

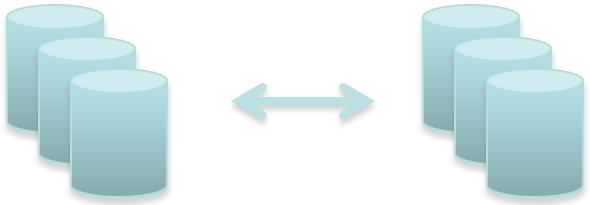
- Mid-size and larger enterprises that have massive unstructured data needs
  - Video processing, CAD, logs, other machine generated content
  - Performance requirement high enough to rule out Cloud Archive solution
  - Data to voluminous to store cost effectively in premium solutions (scale out required)
- Generally based on open source solutions
  - ZFS, object storage
- Is converting a server into an array permanently SDS?

The Big Performance Challenges

- Performance of virtual appliances is suspect:
  - Heavy compute requirements for data services
    - RAID
    - Erasure Coding
    - Encryption
    - Compression
    - Deduplication
  - Hypervisor economics can make good old fashioned arrays attractive

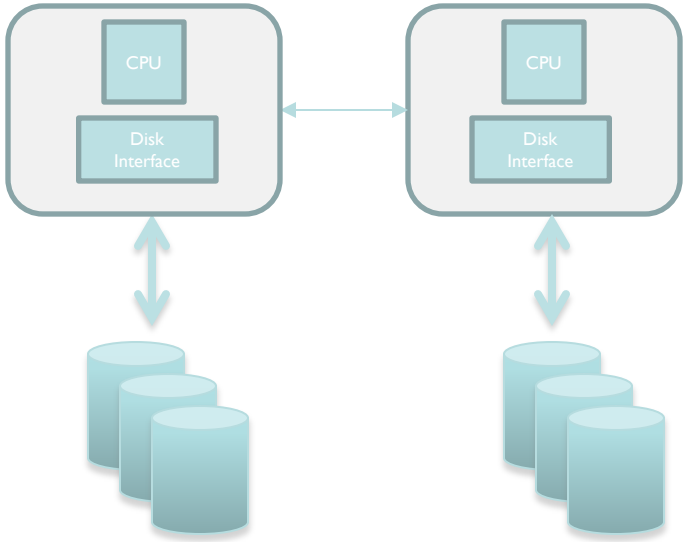
Reliability and Availability

- Numerous challenges exist:
  - Economics of mirroring vs. RAID



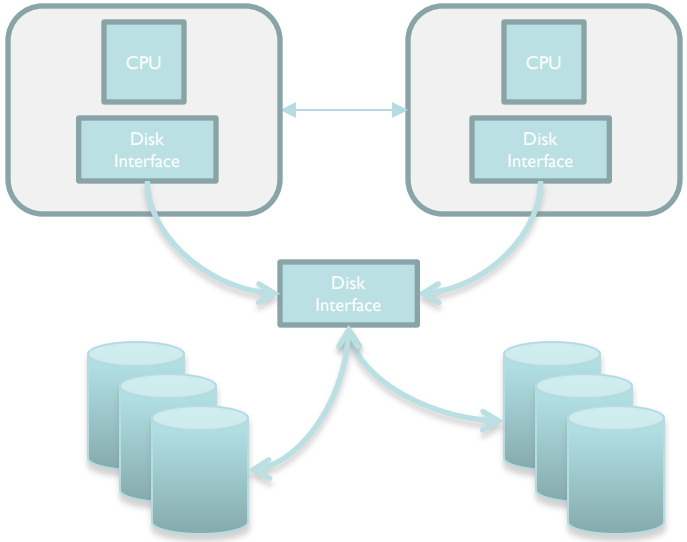
Reliability and Availability

- Numerous challenges exist:
  - Economics of mirroring vs. RAID
  - Economics of shared vs. mirrored media



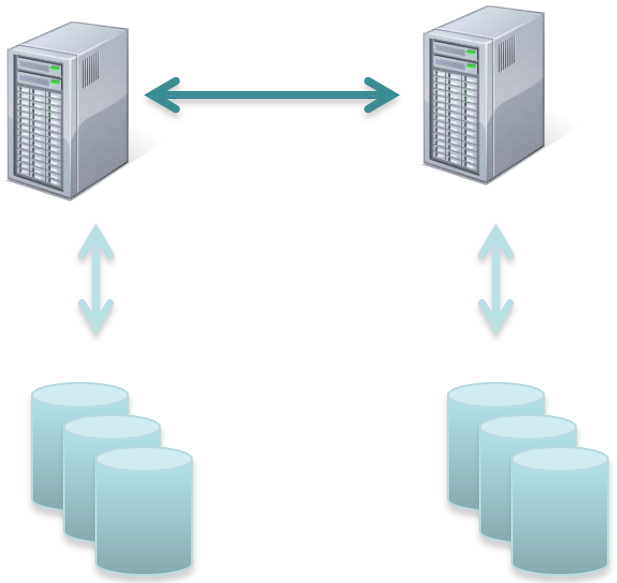
Reliability and Availability

- Numerous challenges exist:
  - Economics of mirroring vs. RAID
  - Economics of shared vs. mirrored media



Reliability and Availability

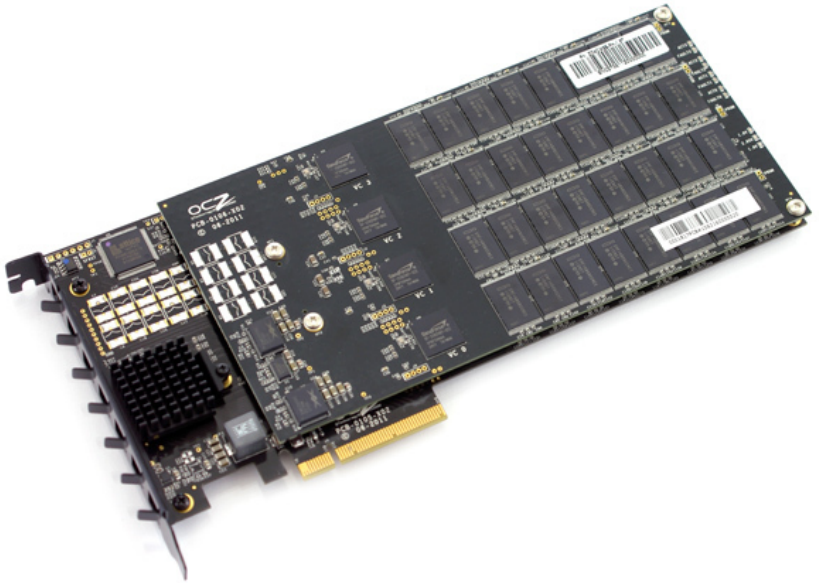
- Numerous challenges exist:
  - Economics of mirroring vs. RAID
  - Economics of shared vs. mirrored media
  - Challenges for the interconnect



What is "Standards-based Hardware"?

Some tests:

- Is it virtualized?
  - Can I move a VM's Storage?
- Is there a commonly agreed interface to access the functionality?



Hardware Reprised

- Hardware offloads are needed today
  - Storage workloads co-mingle with application workloads
- Numerous opportunities for discussion exist
  - Erasure coding
  - Real-time VM prioritization
  - I/O path optimization
  - Virtualized low latency interconnects
  - Network configuration/management

Questions?