Bringing the Public Cloud to Your Data Center

Jim Pinkerton

Partner Architect Lead

1/20/2015

Microsoft Corporation

A Dream…

- Hyper-Scale Cloud efficiency is legendary
  - Reliable, available services using high volume, failure prone, hardware
- But… enterprises do not operate at the same scale

5.8+ billion
worldwide queries each month
bing

250+ million
active users
OneDrive

400+ million
active accounts
Outlook.com

2.4+ million
emails per day
Microsoft Exchange Hosted Services

8.6+ trillion
objects in Microsoft Azure storage
Microsoft Azure

48+ million
users in 41 markets
XBOX LIVE

50+ million
active users
Microsoft Office Web Apps

1 in 4
enterprise customers
Office 365

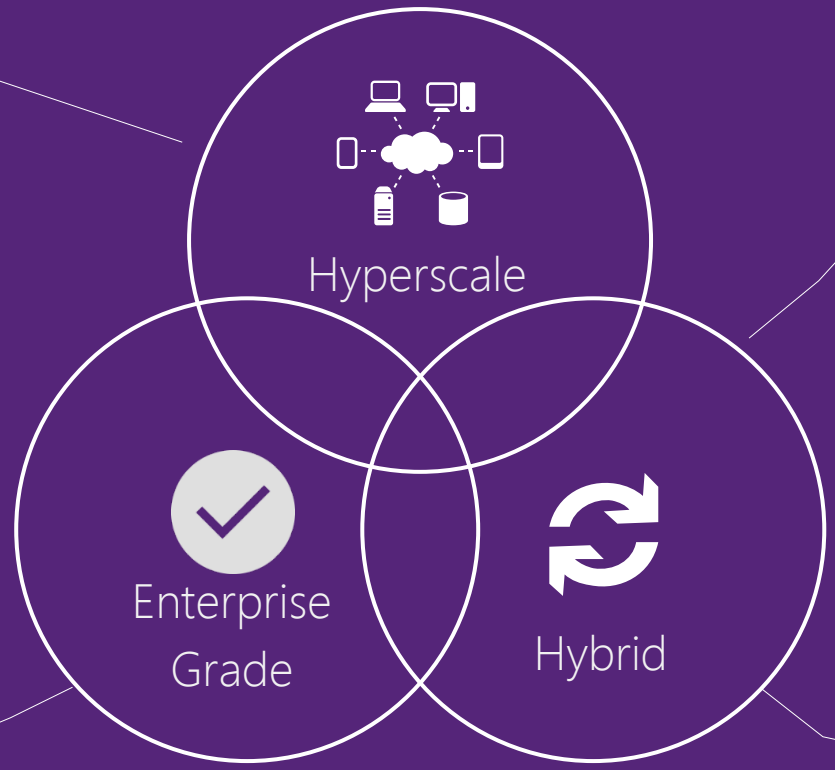
50+ billion
minutes of connections handled each month
skype

200+ Cloud Services

1+ billion customers · 20+ million businesses · 90+ markets worldwide

Microsoft Uniquely Positioned

Innovation in datacenter management (led by Azure, O365, Bing, Skype)



Microsoft in the leaders magic quadrant in four segments

- x86 Virtualization
- Cloud IaaS
- Enterprise Application PaaS
- Public Cloud Storage Services

Cloud OS

Common hypervisor and OS

Common tenant UX

But what is a “cloud”?
- A new paradigm for managing servers

- Provides tenant services
  - IaaS solution – host tenant VMs
  - PaaS solution – host tenants on “less than a VM”
  - SaaS solution – multi-tenant end-user services
- Extreme computing
  - Extreme virtualization – compute, networking, storage
  - Extreme automation
    - Tenant self-service
    - Self-healing fabric
  - Extremely efficient use of high-volume hardware
  - Extreme dynamic scaling of services

Requirement: Software Defined Everything

- Integrated multi-machine management
  - Tenant self service
  - “Fabric” management – networking, compute, storage
- Software Defined Everything
  - Virtualized
    - Networking – each tenant believes they are on a private network
    - Storage – each tenant believes they own all the storage resources
    - Compute – each tenant believes they own the server
  - Make tenant services reliable, scalable, available
    … on top of high volume, fault prone hardware

But SDI requires a lot of work to put the pieces together

Microsoft Cloud Platform System - powered by Dell

Azure-consistent Cloud in a Box

Microsoft

Windows Server 2012 R2,
System Center 2012 R2,
Windows Azure Pack

Microsoft-designed architecture based on Public Cloud learning

Microsoft-led support & orchestrated updates

Optimized run-books for Microsoft applications

DELL

Dell PowerEdge servers

Dell dense Storage enclosures

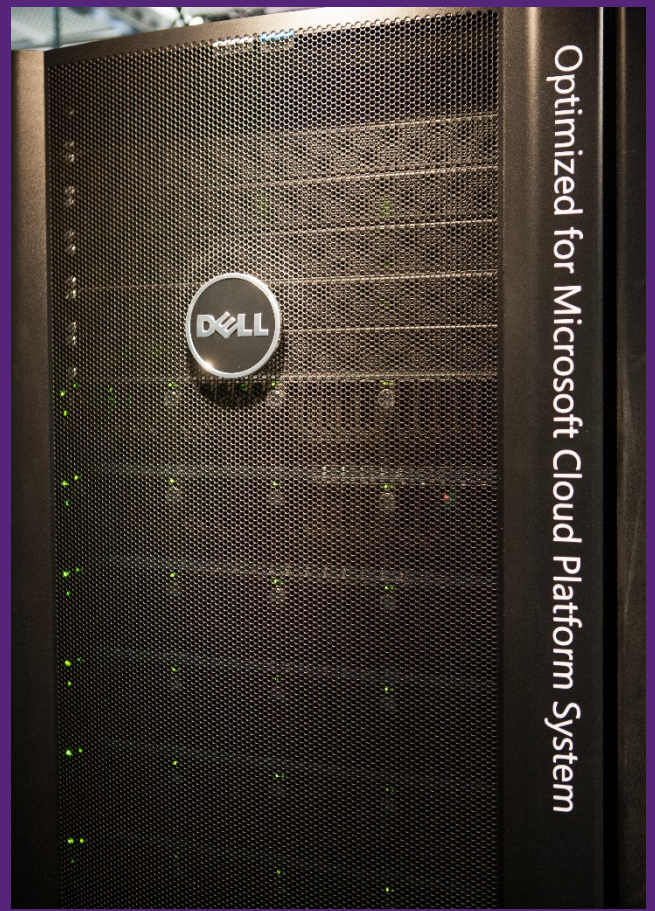
Dell Networking switches

Tightly integrated components

Microsoft-led support & orchestrated updates

Built on standard, high volume hardware

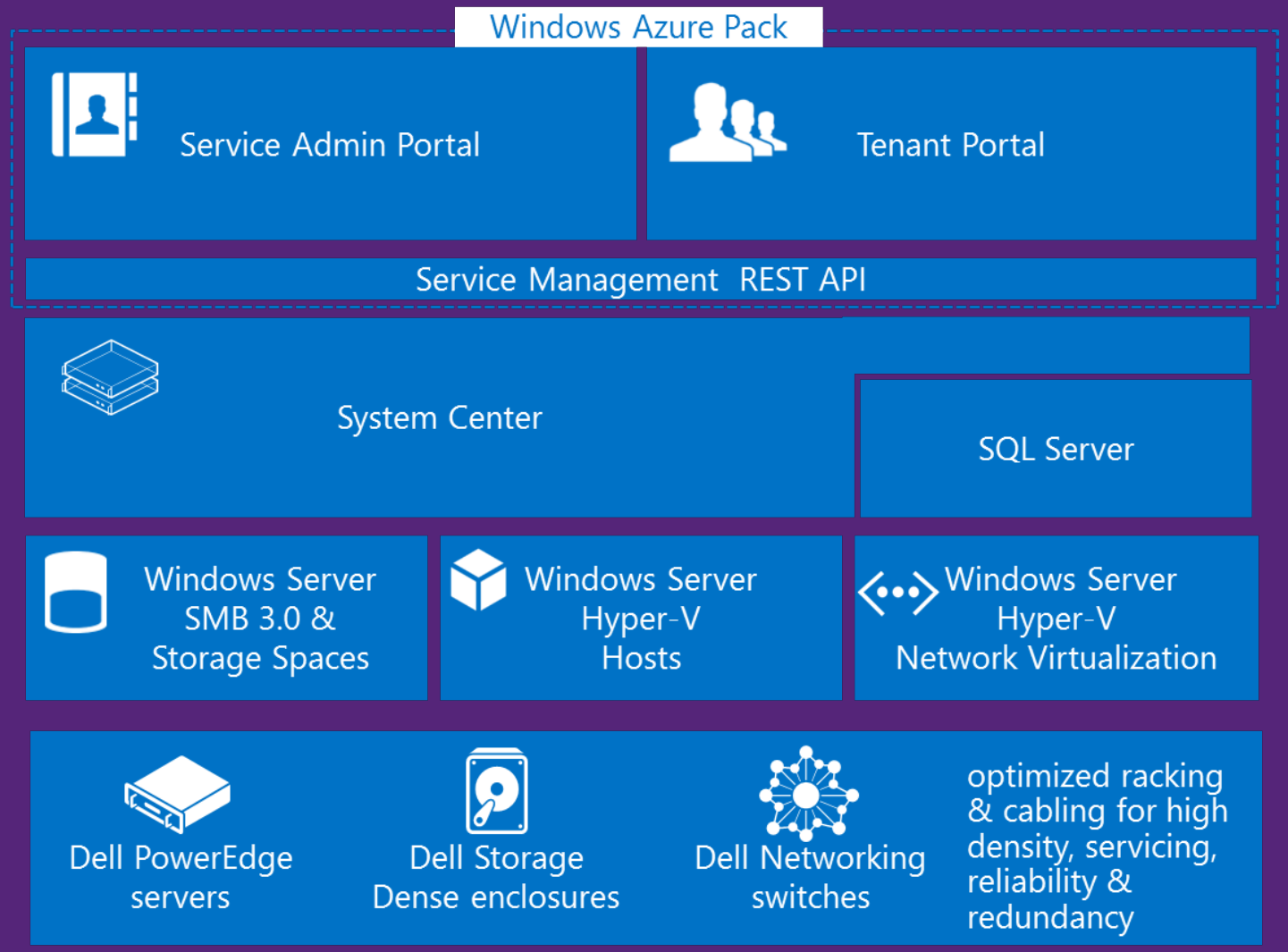
pre-packaged, pre-wired



standard servers, switches

JBODs

Cloud Platform System - Capabilities



- Pre-deployed infrastructure
  - Switches, load balancer, storage, compute, network edge
  - N+2 fault tolerant (N+1 networking)
- Pre-configured as per best practices
- Integrated Management
  - Configure, deploy, patching
  - Monitoring
  - Backup and DR
  - Automation
- Up to 8000 VM's* and 1.1 PB of total storage
- Optimized deployment and operations for Microsoft and other standard workloads

** VM Topology - 2vCPU, 1.75 GB Ram, 50 GB Disk*

CPS – Cloud in a Box

Per Rack
(1-4 racks)

512 Cores
8TB RAM
262 TB usable storage

1360 Gb/s internal rack connectivity
560 Gb/s inter-rack connectivity

60 Gb/s external

2322 Lbs
42U
16.6 KW Maximum

Azure Consistent

- Same tenant self service portal – integrated with SDN/SDS/SDC
- Same management APIs
- Same extremely efficient packing of compute and storage

Software Defined Compute (SDC)

- Virtualized compute (virtual cores, virtual memory, etc)

Software Defined Networking (SDN)

- Hyper-V Network Virtualization overlay network for isolation
- Integrated load balancer for dynamic service scaling

Software Defined Storage (SDS)

- Reliable, Available, Servicable storage on high volume hardware
- JBOD based storage
- Advanced storage configurations (dedup, tiering, write-back-cache, SMB3)

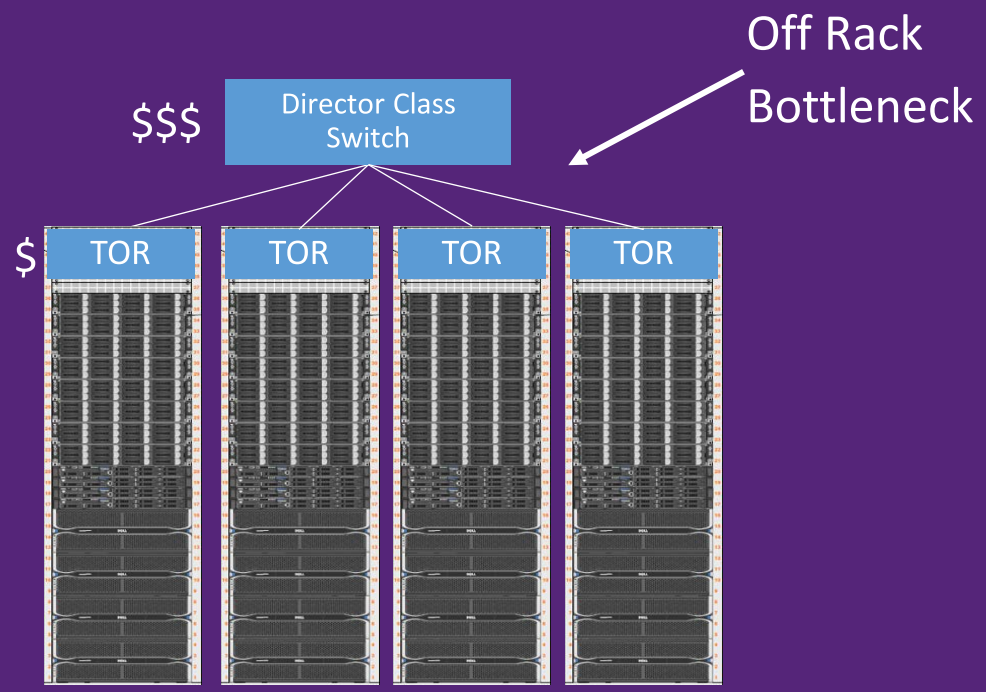
CPS – An Azure Consistent Innovative Architecture

- Same hypervisor
  - In CPS, all management functions have been virtualized
- IaaS and PaaS solution
  - Azure Portal
  - Azure Web Sites
- Efficient use of hardware
  - Efficient packing of VMs
  - Efficient packing of storage
  - State of the art network offloads
- Dynamic scaling
  - Dynamic VM scaling
  - Integrated load balancer
  - Incremental Hardware scaling

Flat Network – Efficient Resource Packing

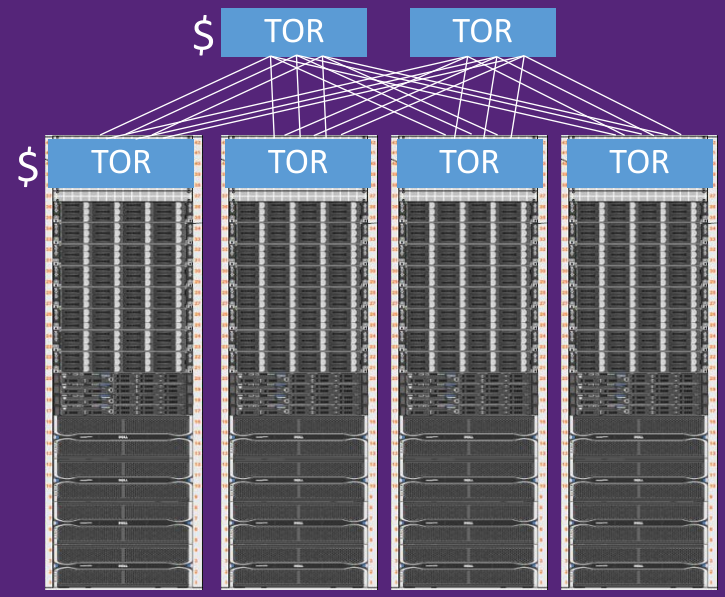
Traditional Deployment

- Within rack: 1360 gb/sec
- Off-rack typically 10-20 gb/sec
- **VM and storage placement is critical**
  - Orphaned resources are common



CPS Deployment

- Flat network between racks
  - 560 gb/sec off-rack connectivity
- Equal Cost Multi-Path Routing (ECMP)
  - Network fault tolerance
- **VMs and storage can be placed anywhere resources are available**

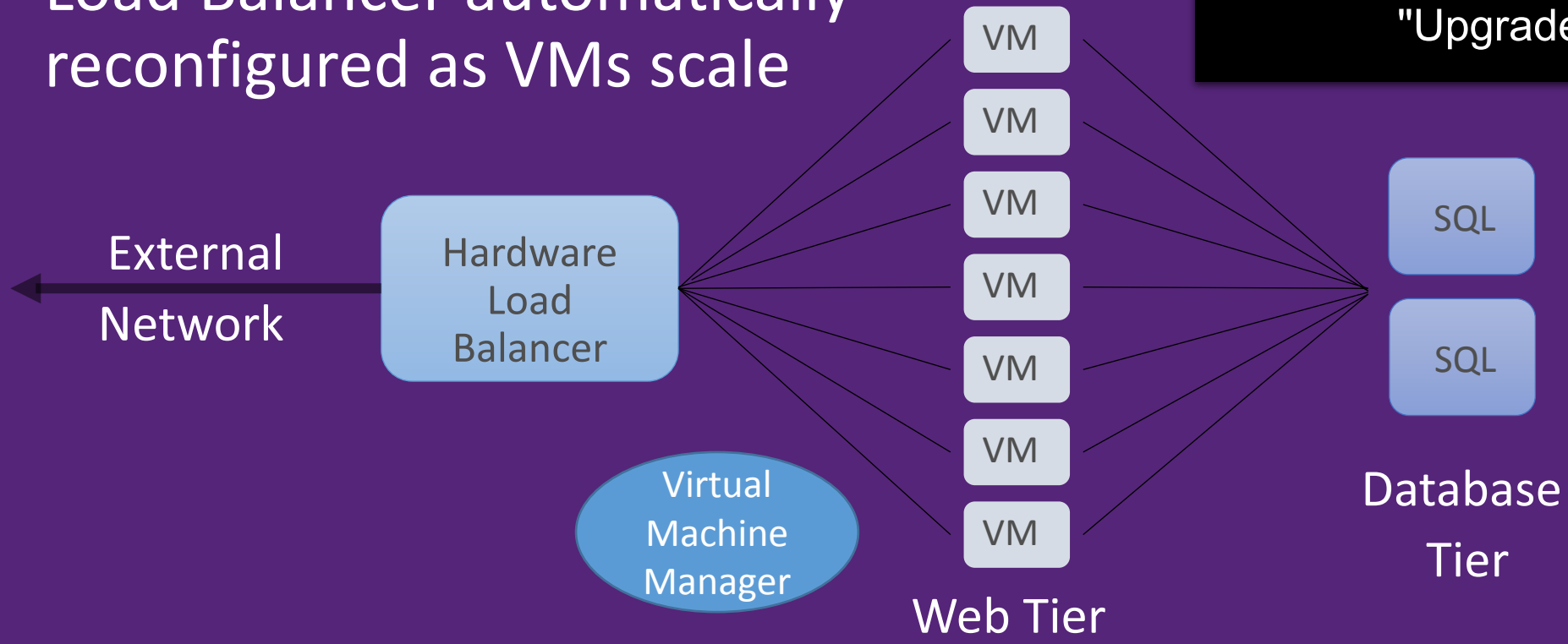


Dynamic Scaling

- VM Role enables dynamic VM scaling
- Load Balancer automatically reconfigured as VMs scale

VM Role Instance

```
"ResourceDefinition": {
    "ScaleOutSettings": {
        "InitialInstanceCount": "1",
        "MaximumInstanceCount": "5",
        "MinimumInstanceCount": "1",
        "UpgradeDomainCount": "1" },
```



Optimized for Enterprise and Hosters

- Low operational overhead
  - Transparent patching
  - Integrated management
  - Integrated and tested at full scale

- Converged Infrastructure
  - Storage, Compute, Networking, Management

- Highly fault tolerant
  - N+2 for compute and storage
  - N+1 for networking

Cloud Reliability: Cloud Operations Simulation: **Storage**

1 year in 7 days

Cloud Learnings
- Design for loosely coupled system > Validate E2E
- Fault Tolerance → MTTR
- Continuous monitoring and measurements

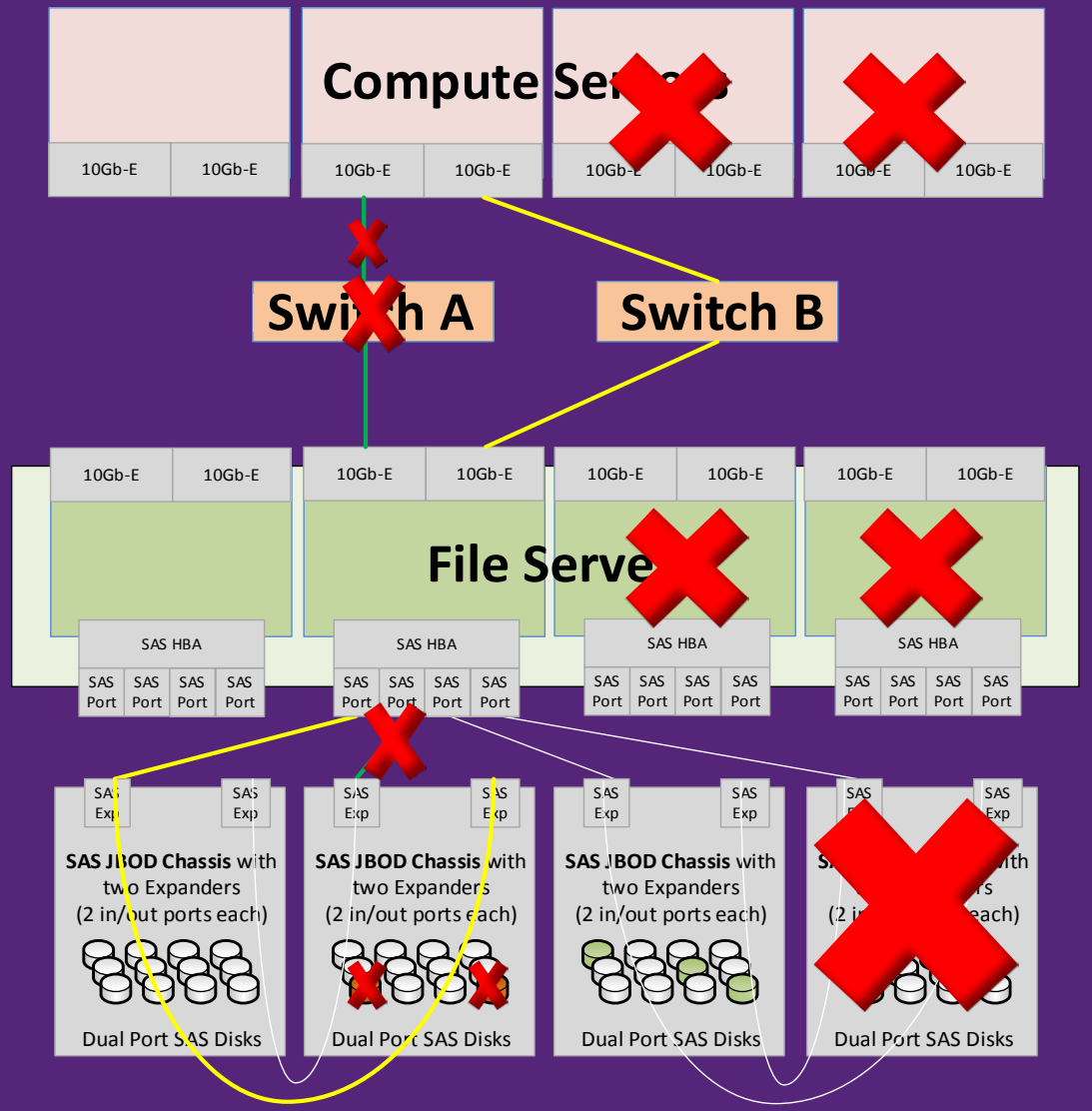
TOPOLOGY & TENANT WORKLOADS
- CPS Full Rack or Stamp Configuration
- IaaS VM Roles & variety workloads

TRIGGER FAULTS & ADMIN ACTIONS
- Actions: Deploy, Live Migrate, VM Meta Ops
- CSU Faults: Patching, Node drain/crash
- SSU Actions: Rebuilds, Backup, Tiering
- SSU Faults: Power, NIC, HBA, JBOD

MEASURE SUCCESS CRITERIA / SLA
- Zero downtime for tenant workloads
- Failover within minutes; zero app errors
- Meet data consistency & resiliency SLA

Profile	Metric
Cloud Topology	Tenant VMs: 2000 VM Profile: 1 vCPU, 1.75GB Compute Nodes: 30 Storage Nodes: 4 (SOFS+SPACES)
Failures: Compute Node Storage Node JBOD Power Shared Disk NIC, Cable, Switch SAS HBA & Cable	376 node drain & failovers 244 unplanned failovers 1 JBOD failure per day 2 drive failures per pool per day 28 NIC/Cable & 2 Switch failures 8 SAS Cable Pulls & 2 HBA failures
Tenant Workloads & SLA VM Live Migrations Storage Migrations	Variety workloads always running SLA: Zero impact on workloads SLA: Zero IO errors or timeouts SLA: Failovers within a minute 10,152 VMs live migrated 5,734 VMs storage migrated

Highly Fault Tolerant



- All systems are active-active
  - N+2 Fault Tolerant for
    - HDD and SSD
    - Storage Servers
    - Compute Servers
  - N+1 Fault Tolerant for:
    - SAS network (cable, HBA)
    - JBOD
    - Ethernet network (cable, switch)

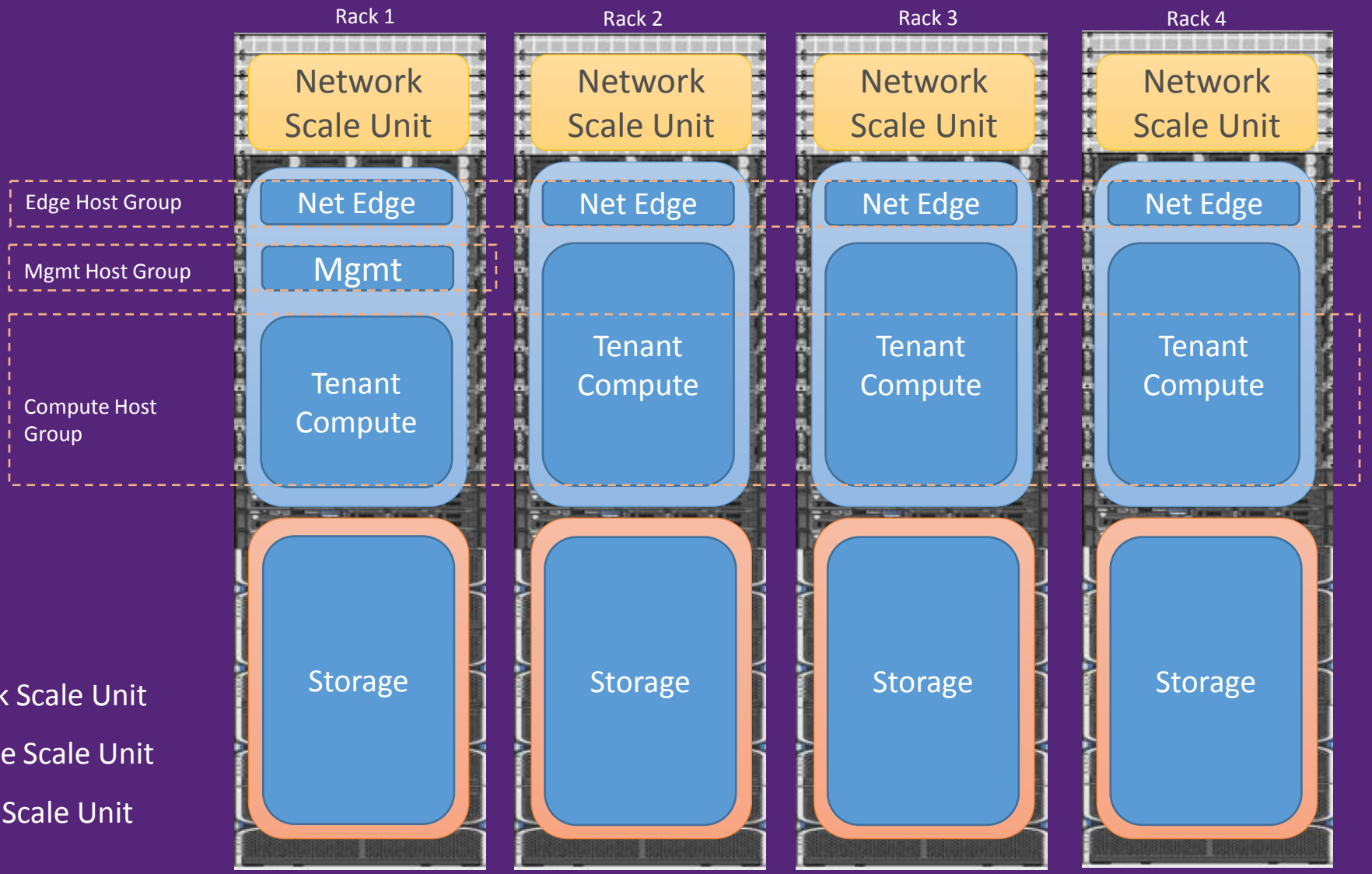
Illustrative of the CPS Architecture

All links not shown for clarity

Converged Infrastructure

Converged Solution Combines:

- Software Defined Storage
- Software Defined Network
- Software Defined Compute
- Integrated Management



Storage Scale Unit (SSU) Performance - Reads

Test Configuration

Load: 112 VMs, on 14 servers generating load
- Diskspd Load generator in each VM, single threaded
- Random 4 KB IO, using RDMA offload NICs

Single SSU (4 servers),
- 14* tenant volumes, 3-way mirror
- 2 pools, 20 SSD per pool (capacity of 2 in reserve)

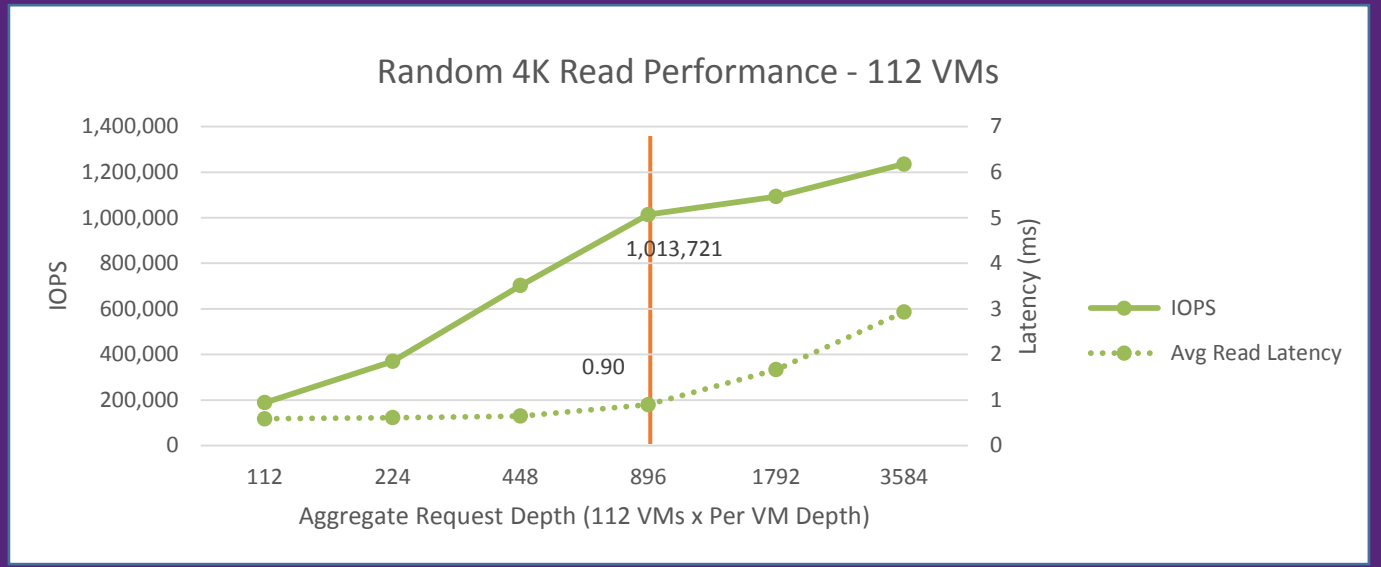
Spaces tiering placed all IOs in SSD Tier

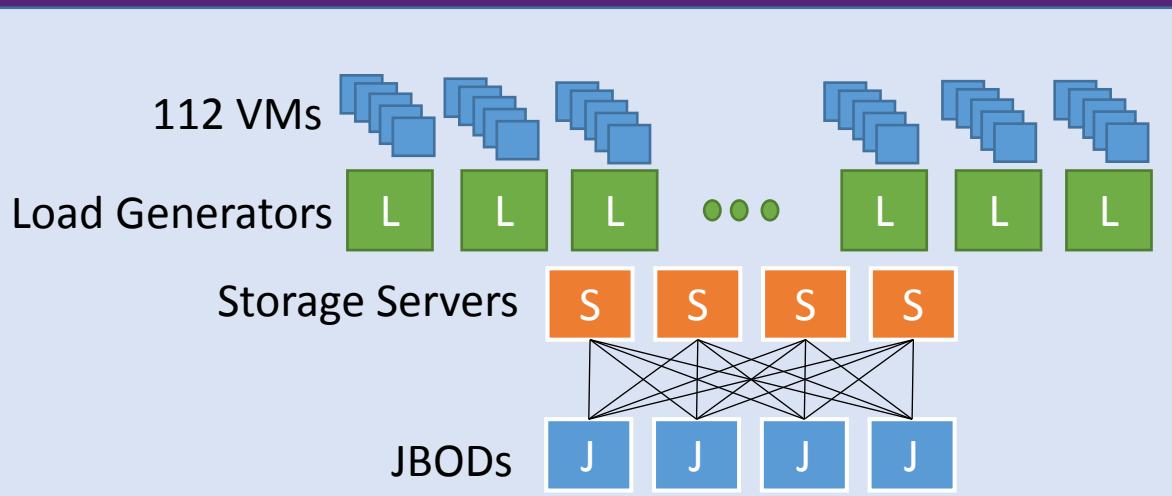
Variables

Queue depth (number of outstanding IOs per VM)
- Total IO queued to SSU is 112*QD

Single SSU Results (IOPs, Latency)

100% Read (non-saturated)
- 1M IOPs, queue depth of 8 per VM,
- 0.86 ms avg latency, 4 ms 95th percentile





*All of the data volumes available for tenant workloads

Storage Scale Unit (SSU) Performance – Read/Write

Test Configuration

Load: 112 VMs, on 14 servers generating load
- Diskspd Load generator in each VM, single threaded
- Random 4 KB IO, using RDMA offload NICs

Single SSU (4 servers),
- 14* tenant volumes, 3-way mirror
- 2 pools, 20 SSD per pool (capacity of 2 in reserve)

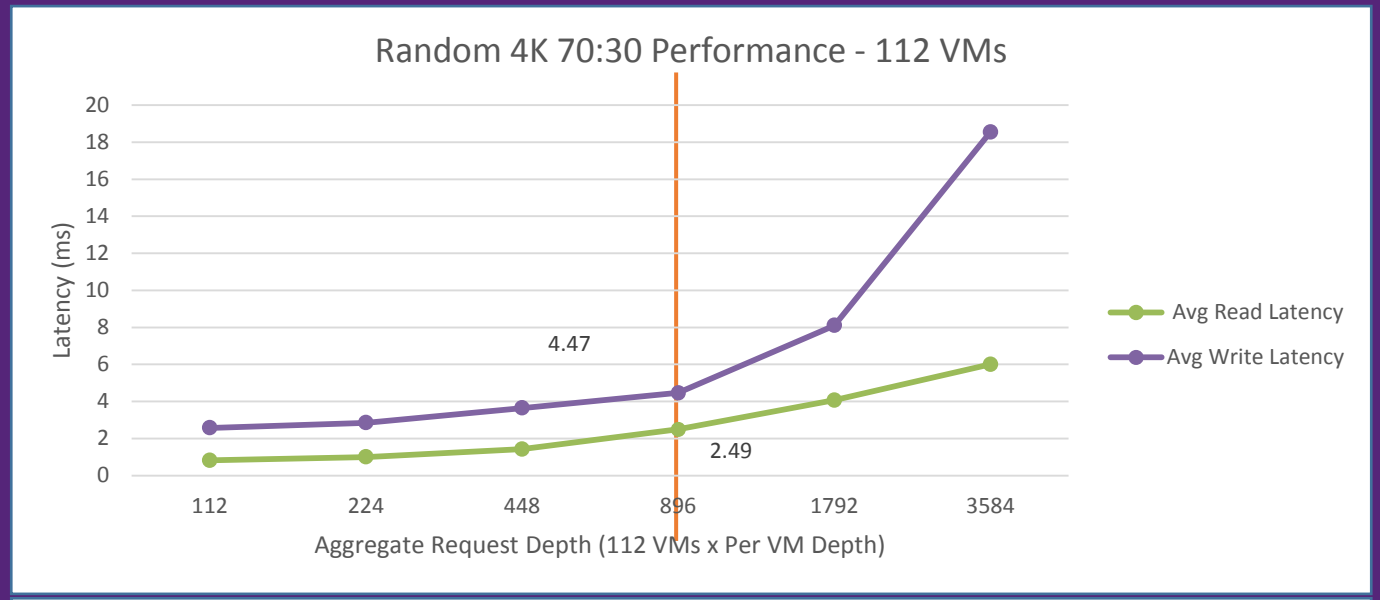
Spaces tiering placed all IOs in SSD Tier

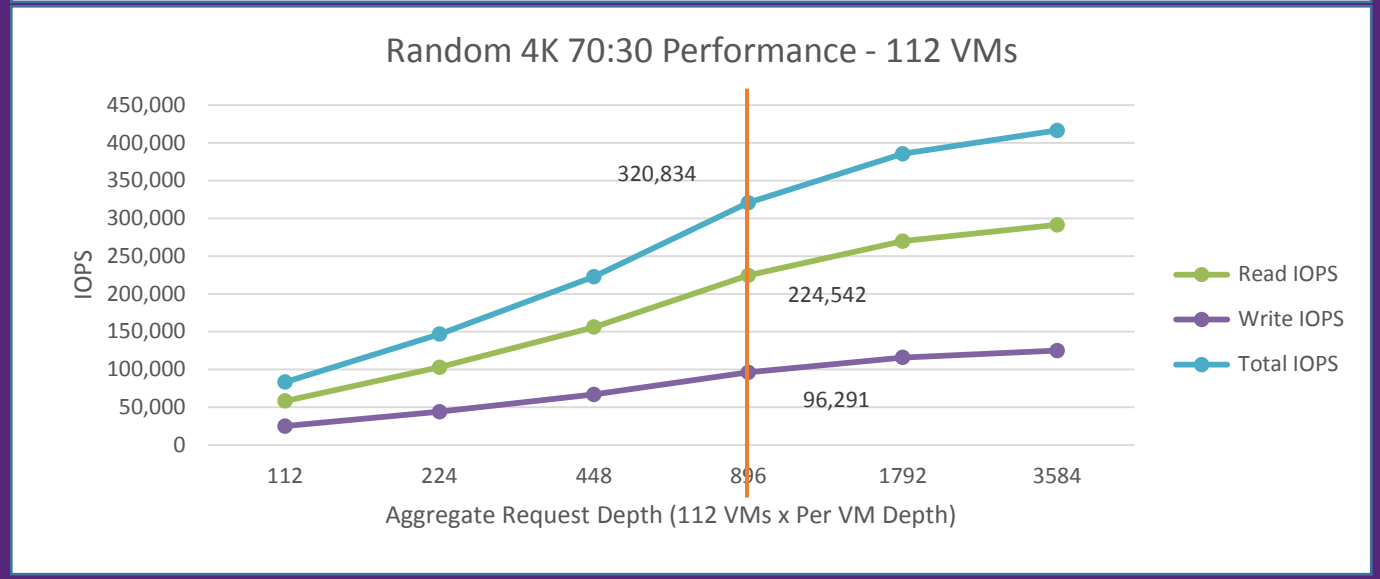
Variables

Queue depth (number of outstanding IOs per VM)
- Total IO queued to SSU is 112*QD

Single SSU Results (IOPs, Latency)

70%/30% Read/Write (non-saturated)
- 321K IOPs, queue depth of 8 per VM
- Reads: 2.5 ms avg, 5.0 ms 95th percentile
- Writes: 4.5 ms avg, 37.4 ms 95th percentile





*All of the data volumes available for tenant workloads

What's Next? The Cloud Marches on

- Continue to make Services reliable, scalable, available
  … on top of high volume, fault prone hardware

- Some cutting edge component technologies of interest
  - Storage - NVME, NVDIMM, ultra-high capacity hard drives
  - Networking - 40 gigabit Ethernet, new offloads, advanced network topologies
  - Next generation isolation - compute, network, storage
  - Next generation RESTful management

- See Microsoft's Open Compute contribution "Open Cloud Server" (OCS)