# Abstract


Tile: Intel Rack Scale Architecture

- This talk provides an overview of Intel Rack Scale Architecture and discusses how this architecture addresses underutilized and stranded resources in a Data center through resource pooling!
- We will also specifically discuss concept of a pooled system, storage node, pooling of PCIe as well as NVMe based storage.
- The impact of pooling on latency, radix and failure domains will also be discussed.
- Further pooling introduces a need for composition of the platform.
  We will also discuss the characteristics of such platform composition how software can emerge to take advantage of these capabilities!


# Rack Scale Architecture

**Mohan J. Kumar**

**Intel Fellow, Data  Center Group**

# Data Center Challenges

Infrastructure has not kept up with increasing business demands

## Inefficiency

Less than 50% server utilization[2]


## Growth

Data growth doubles every 18 months[1]

## Agility

New services can take a week or more to provision[1]

## Business Needs

- **Reduce** operational and capital expenses.
- **Deliver** new services in minutes, not months.
- **Optimize** data center based on real-time analytics.
- **Address** application workload needs with agility.
- **Scale** capacity without interruption.

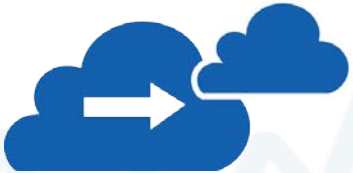
1 *Worldwide and Regional Public IT Cloud Services 2013–2017 Forecast.* IDC (August 2013)
2 IDC's Digital Universe Study, sponsored by EMC, December 2012

# Today's Architecture

- Proprietary and preconfigured
- Upgrade as a system
- Limited flexibility

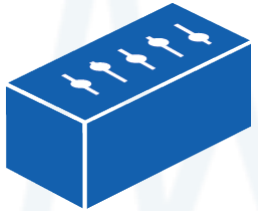
# What's Next?

A seismic shift in how data centers are built and managed—powered by Intel



All infrastructure
delivered as a service


Hyper-scalable to keep
up with business demands



Resources automatically
tuned to application workloads

# Software Defined Infrastructure

## Dedicated Appliances

SAN

NAS

Network Appliance

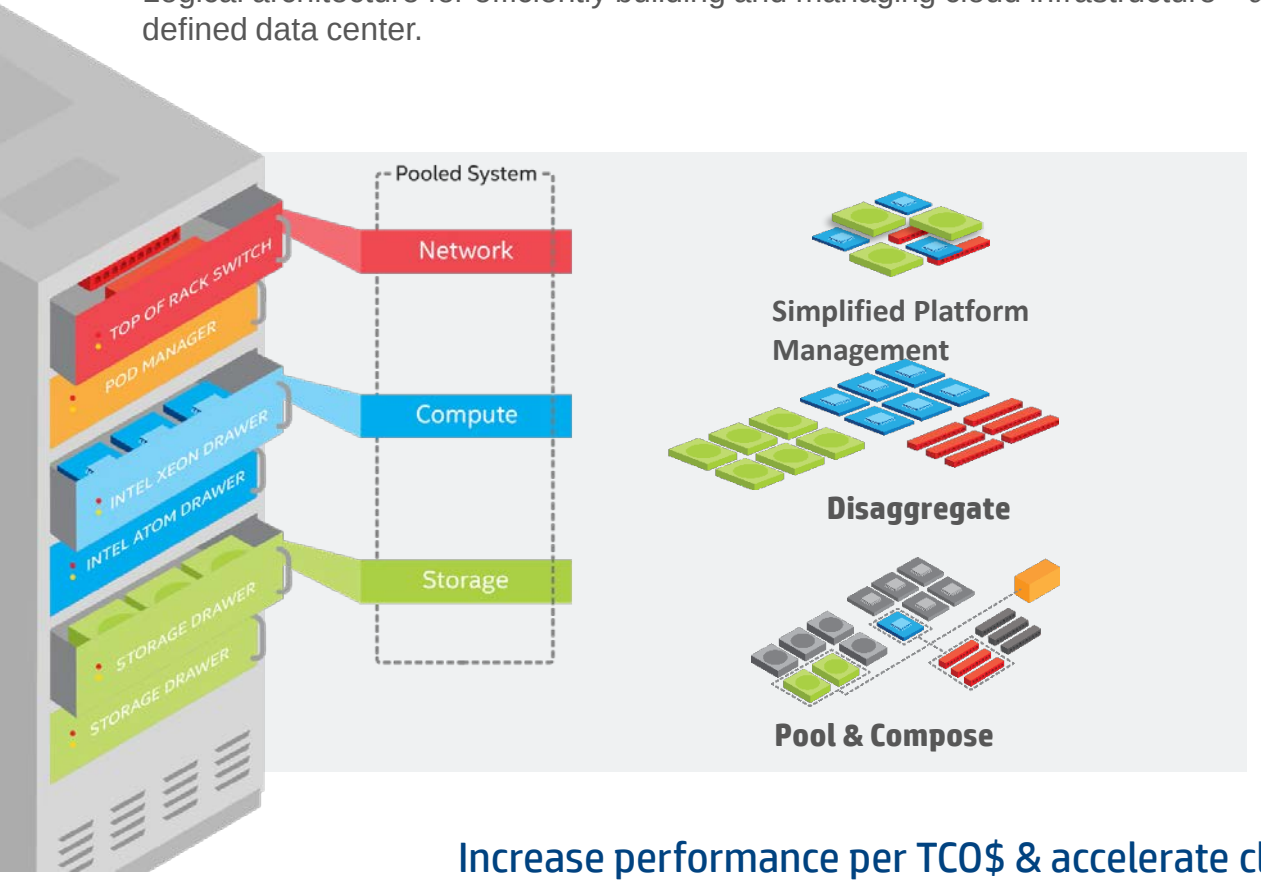
Telco Appliance

⋮

## Software Defined Infrastructure

# Intel® Rack Scale Architecture

Logical architecture for efficiently building and managing cloud infrastructure—and providing the simplest path to a software defined data center.

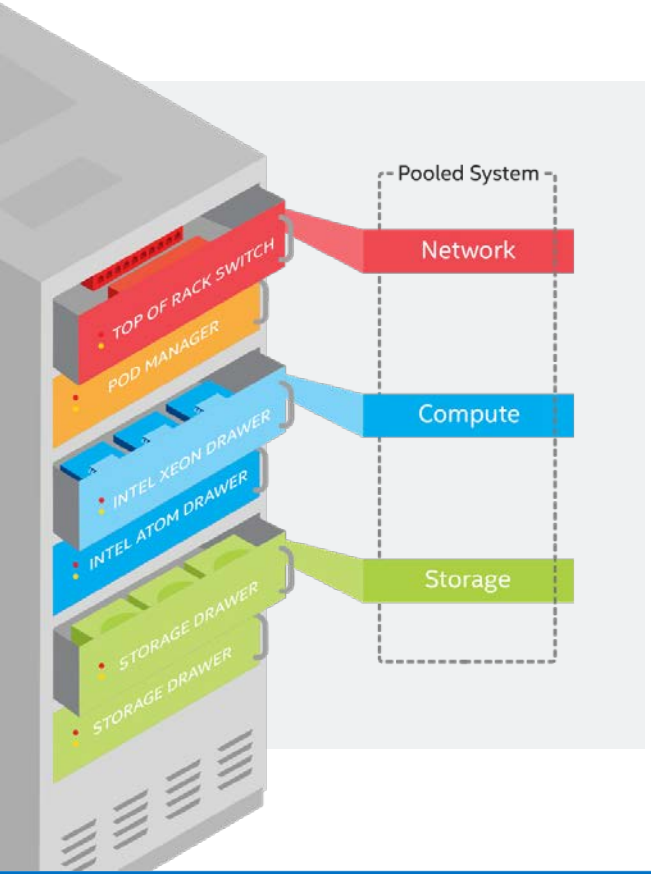


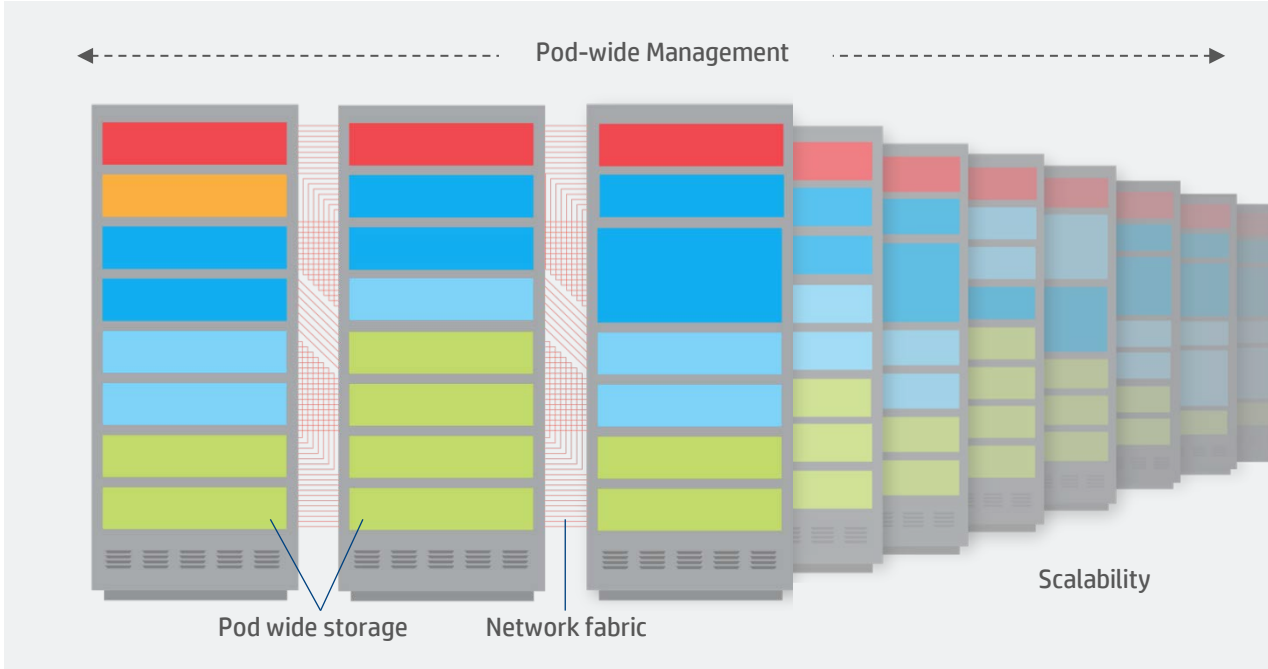

- ✓ **User-Defined Performance**
- ✓ **Maximum Utilization**
- ✓ **Interoperable Solutions**

Increase performance per TCO$ & accelerate cloud adoption

# Rack Scale – Architecture Framework

1. Pooled systems 2. Network fabric management 3. Pod-wide storage 4. Pod management



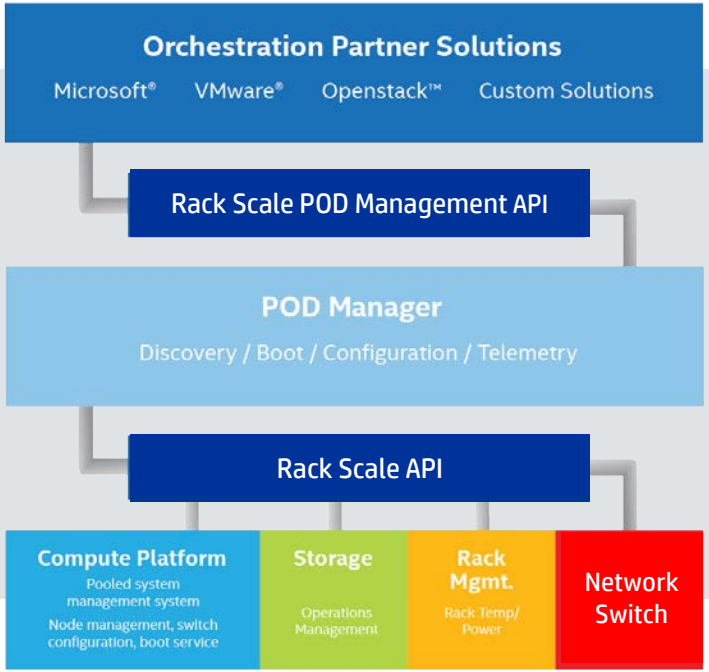




- Modular scalable management architecture

# Management Software Framework

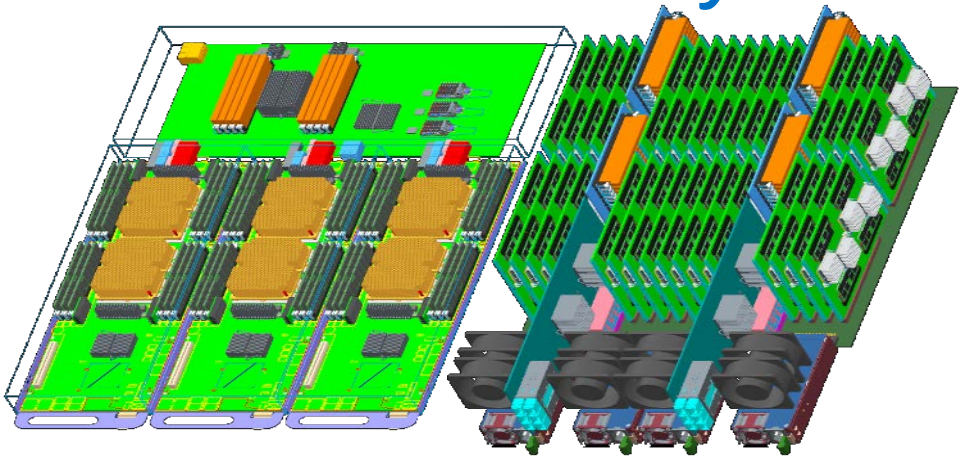
Flexible management architecture allowing for range of implementation options

- Asset & location discovery
- Disaggregated resource management
- Composable system support
- Support compute, network, and storage



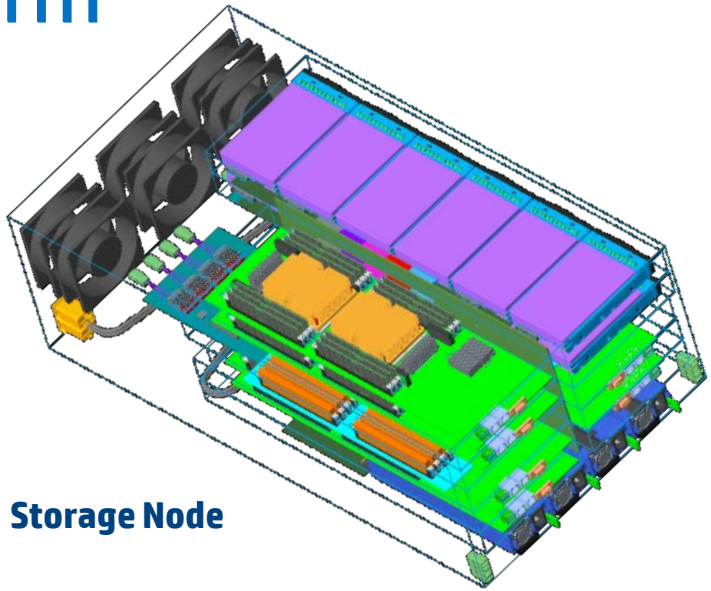

- Comprehensive management architecture

# Rack Scale Pooled System Platform



**Compute Node**

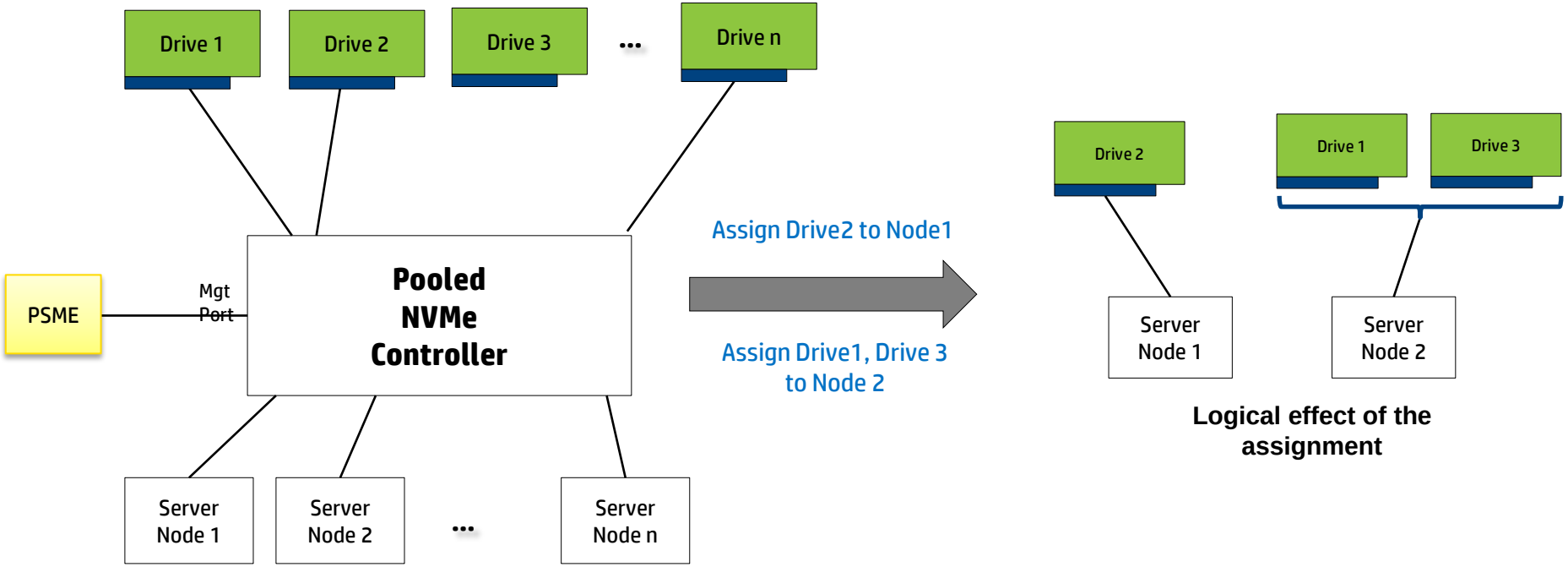
- Supports range of server processors
- Pooled NVM
- Supports Ethernet fabric
- Service Model requires Full node replacement



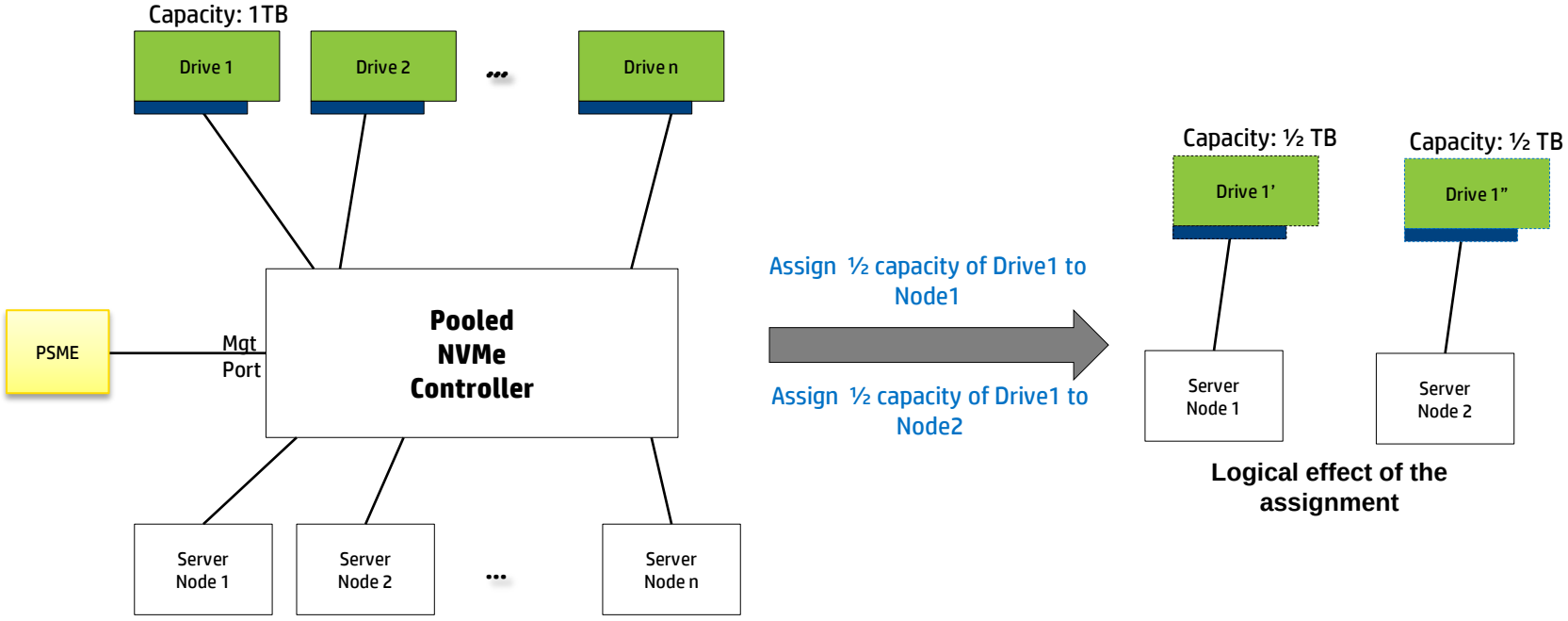
**Storage Node**

- Supports range of server processors
- Direct Attach Storage
- Pooled NVM
- Supports Ethernet fabric
- Redundant networks, scalable DAS storage, sub node FRUs

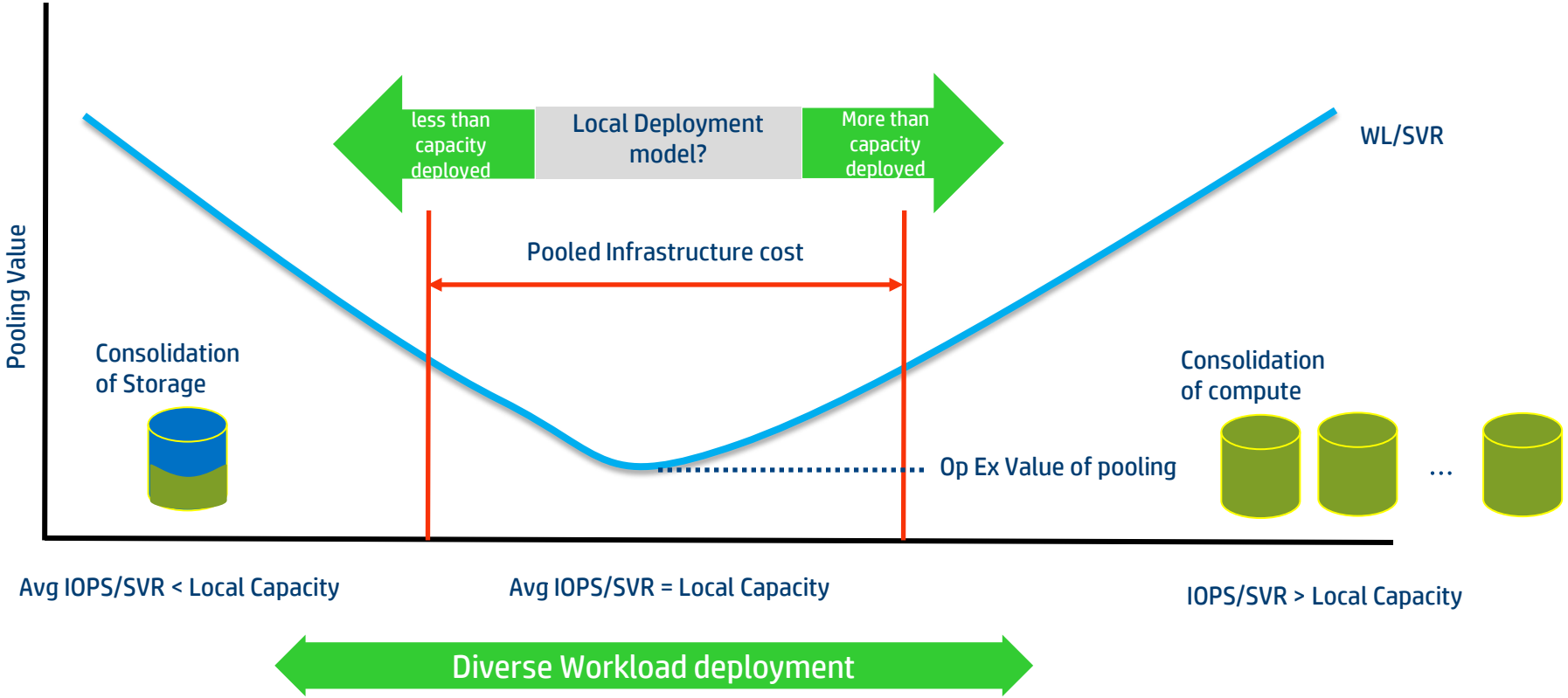
# Rack Scale Pooled NVMe Controller (PNC)



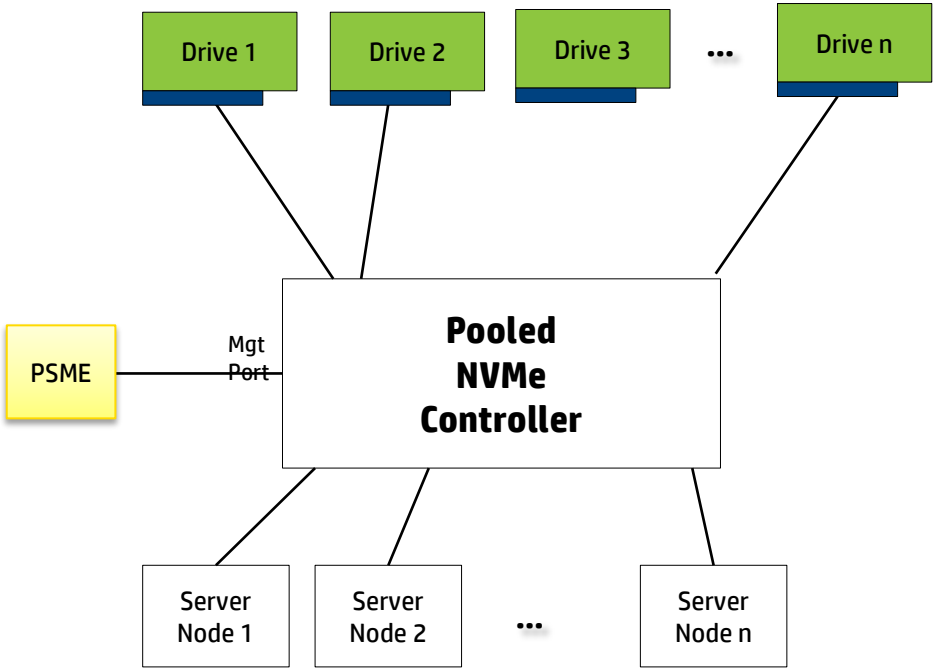
# Rack Scale Pooled NVMe Controller (PNC)



# Local vs. Pooled Storage

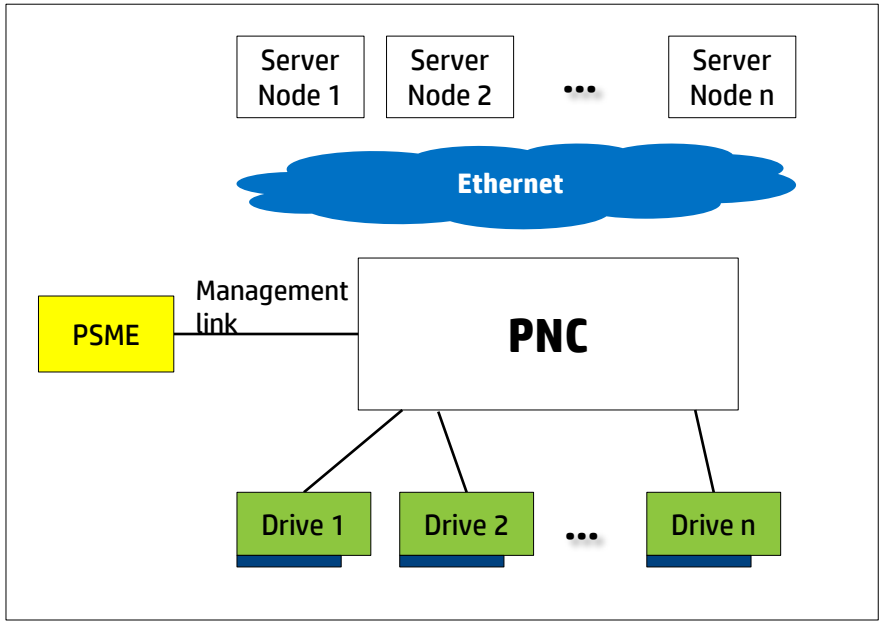


# Rack Scale Pooled NVMe Controller (PNC)




- Enable pooling and disaggregation of PCIe devices away from compute / storage nodes
- Enable disaggregation of PCIe devices including Storage, FPGA
- Assign high performance storage to nodes based on workload demand
- Allow full and partial drive assignment
- Prevent SPOF through host failover
- Enables ease of workload migration in hyperscale environment
- Enables better utilization of Data Center resources by allowing composable high performance IO capacity

# Rack Scale Pooled NVMe Controller (PNC)

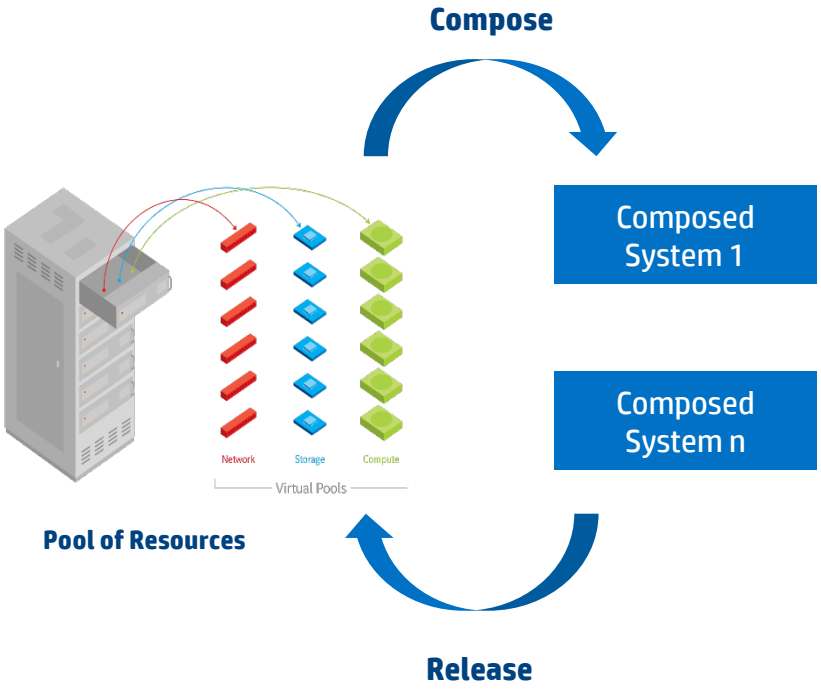



- Enable disaggregation of NVM Express devices
- Utilizes NVMeOF to expand radix of pooling
- Assign storage to Compute or Storage nodes based on workload demand
- Prevent SPOF through host failover
- Enables ease of workload migration in hyperscale environment
- Enables better utilization of DC resources through composition

# Local vs. Pooled Resources

| | Resource Local to a Node | Pooled Resource |
|---|---|---|
| Device Attach and Capacity | Limited by Physical constraints | Not constrained by node volumetrics |
| Device Availability (Failure Domain) | Node is SPOF | Pooled Fabric |
| Utilization | Limited by Local use | No stranded capacity/capability |
| Latency | Local Access | Incurs Additional Pooling Latency |
| Radix | Local | Limited by Pooling Fabric (one rack to multiple racks) |
| Refresh and Life Cycle Management | Node based | Component based |

# Composable Infrastructure – Software Implications




- Orchestration that comprehends composition capabilities
- Location aware placement of workloads
- Location aware placement in hyperscale storage
- Monitoring software that knows the physical bounds of the hardware
- Software (OS, VMM, App) capability to take advantage of dynamically added resources

# Summary

***USER-DEFINED PERFORMANCE***

- Tailor performance to meet application SLAs by selecting from pooled compute, storage & network resources
- Easily scale capacity with modular, buy-as-you-go architecture

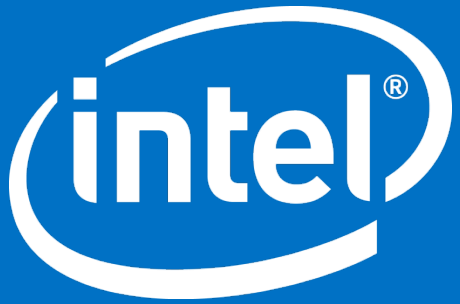
***MAXIMUM UTILIZATION***

- Autonomously manage compute, network & storage pools to virtually eliminate stranded resources

***INTEROPERABLE SOLUTIONS***

- Interoperable system architecture simplifies data center operations and integration of multi-vendor solutions

"Buy What You Need. Use What You Buy"


intel®