Intel® Rack Scale Architecture Integration with Orchestration solutions
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.

---

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
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
Intel® Rack Scale Architecture Framework

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

Modular scalable management architecture
Pooled Systems

Intel Rack Scale Architecture transforms resources from physical servers into virtual pools
Evolution of Rack Scale Infrastructure

**Today**
- Physical Aggregation
  - Shared power/cooling
  - Modular platforms

**Emerging**
- Composable Resource Pools
  - Rack scale management framework
  - Scalable fabric architecture

**Future**
- Service Aware Orchestration
  - Composable resource pools
  - Service aware orchestration
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
Intel® Rack Scale Architecture Management Hierarchy and API

- Intel® Rack Scale Architecture API is REST based API interface with JSON data format and utilizes HTTPS web protocol
- Intel® Rack Scale Architecture API comprehends multi-node, pooled system, storage and network
- Intel® Rack Scale Architecture API is built using Redfish as base profile
Intel® Rack Scale Architecture and OpenStack

Intel® Rack Scale Architecture software stack provisions hardware

Detects and registers compute, network and storage to respective OpenStack controllers

Intel® Rack Scale Architecture software stack reprovisions hardware based on demand

Enables location aware resource allocation for OpenStack
RSA Futures
Intel® Rack Scale Architecture Futures

Composable High-performance Storage
Composable Accelerators
Composable Memory Tiers
Composable Security
Telemetry Support
...

...
Intel® Rack Scale Architecture Pooled NVMe Controller (PNC)

- Enable disaggregation of NVM Express devices
- Enable disaggregation of PCIe devices including NICs, accelerators
- Assign high performance storage to nodes based on workload demand
- Prevent SPOF by host failover
- Enables ease of workload migration in hyperscale cloud environment
- Enables better utilization of DC resources by allowing composable high performance IO capacity
PNC
Full Drive Assignment

Capacity: 1TB

Assign Drive2 to Node1

Logical effect of the assignment
**PNC**

**Partial Drive Assignment**

Assign ½ capacity of Drive1 to Node1 and Node2

Capacity: 1TB

Drive 1

Drive 2

Drive n

Pooled NVMe Controller

Port : N1

Port : N2

Port : Nn

Server Node 1

Server Node 2

Server Node n

PSME

Port : N1

Port : Nn

Assign ½ capacity of Drive1 to Node1 and Node2

Capacity: ½ TB

Drive 1’

Drive 1”

Server Node 1

Server Node 2

Logical effect of the assignment
PNC
Partial Drive Assignment

Assign ½ capacity of Drive 1 to Node 1 and Node 2

Logical effect of the assignment
Summary
Rack Scale Architecture

**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”