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

Growth
Data growth doubles every 18 months

Agility
New services can take a week or more to provision

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.

---

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


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
Assign Drive2 to Node1

Assign Drive1, Drive 3 to Node 2

Logical effect of the assignment
Rack Scale Pooled NVMe Controller (PNC)

Assign ½ capacity of Drive1 to Node1

Assign ½ capacity of Drive1 to Node2

Logical effect of the assignment
Local vs. Pooled Storage

Consolidation of Storage

Consolidation of compute

Pooled Infrastructure cost

Op Ex Value of pooling

Diverse Workload deployment

Avg IOPS/SVR < Local Capacity

Avg IOPS/SVR = Local Capacity

Avg IOPS/SVR > Local Capacity

WL/SVR

Pooling Value

Local Deployment model?

less than capacity deployed

More than capacity deployed

Consolidation

IOPS/SVR > Local Capacity

IOPS/SVR < Local Capacity

Consolidation

Consolidation

Diverse Workload deployment

Op Ex Value of pooling

Local Deployment model?

Consolidation of Storage

Consolidation of compute

Diverse Workload deployment
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

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