Rack Scale Architecture – Platform and Management

Mohan J. Kumar – Senior Principal Engineer, Intel Corporation
Agenda

• Rack Scale Architecture (RSA) Overview
• RSA Platform Overview
• Elements of RSA Platform
• RSA Management and API
• Summary and Call to Action
Rack Scale Architecture (RSA) Overview
Why RSA?

- Manual
- Static
- Hardware Defined
- Automated
- Dynamic
- Software Defined
Why RSA?

- Manual
- Static
- Hardware Defined
- Automated
- Dynamic
- Software Defined

Rack Scale Architecture
Why RSA?

Manual
Static
Hardware Defined

Rack Scale Architecture

Automated
Dynamic
Software Defined

Rack Scale Architecture is key to accelerating Software Defined Infrastructure adoption!
RSA – Enabling Software Defined Infrastructure

- Intel technologies optimized for flexibility, performance & cost
- Intel enabling broad range of OEM & end user implementation
RSA – Enabling Software Defined Infrastructure

- Intel technologies optimized for flexibility, performance & cost
- Intel enabling broad range of OEM & end user implementation

Exposed and Integrated Telemetry

Broadest Enabled Ecosystem

Platform and Architectural Leadership

Orchestration Partner Solutions

- OpenStack*
- VMware*
- Microsoft*

Hardware Attributes
Composability

Partner Hardware Solutions
Intel reference architectures

STORAGE
Building Blocks

NETWORK
Building Blocks

COMPUTE
Building Blocks

Intel technologies optimized for flexibility, performance & cost
Intel enabling broad range of OEM & end user implementation

RSA builds upon hyperscale best practices with improved capabilities for the infrastructure layer!
Intel® Technologies for Rack Scale Architecture

- Open network platform
- Storage – PCIE – SSD & Caching
- Photonics & Intel Ethernet Switch
- Silicon – Atom & Xeon (CPU / Mem Modules)

- Up to 1.5x Improvement
- Up to 6x Reduction

- Power provisioning Reduction
- Network uplink Improvement
- Network downlink Improvement
- Cable Reduction

- Intel technologies optimized for flexibility, performance & cost.
- Intel enabling broad range of OEM & end user implementation.
Intel® Technologies for Rack Scale Architecture

Open network platform

Storage – PCIE – SSD & Caching

Photonics & Intel Ethernet Switch

Silicon – Atom & Xeon

CPU / Mem Modules

Accelerating Rack Scale innovation by delivering suite of interoperable technologies

- Intel technologies optimized for flexibility, performance & cost.
- Intel enabling broad range of OEM & end user implementation.
Evolution of Rack Scale Infrastructure

Today
Physical Aggregation

- Shared power/cooling
- Modular platforms

Emerging
Composable Resource Pools

- RSA management framework
- Pooled system architecture
- Scalable fabric architecture

Future
Service Aware Orchestration

- Shared memory
- Service aware orchestration

Evolves from systems to composable resource pools
Rack Scale Architecture
Platform Overview
Rack Scale Architecture Platform

Rack Scale Architecture moves innovation focus to a POD Level
Physical and Logical Architecture
Enable rack scale innovation at solution level

1. Modular pod (multi-rack) manager
2. Composable pooled systems
3. Scalable pod wide storage.
4. Efficient configurable network fabric

Scalable architecture through hardware and software innovation
RSA Pooled System Type 1 Platform

- Supports range of server processors
- Shared Memory
- Pooled NVM
- Supports POD Wide Storage
- Supports RSA Ethernet fabric
- Service Model requires Full node replacement
RSA Pooled System Type 1 Platform

- Supports range of server processors
- Shared Memory
- Pooled NVM
- Supports POD Wide Storage
- Supports RSA Ethernet fabric
- Service Model requires Full node replacement
RSA Enterprise capable Pooled System Type 2 Platform

- Supports range of server processors
- Large local and Shared Memory
- Direct Attach Storage as well as pooled NVM
- Supports RSA ethernet fabric
- Redundant networks, scalable DAS storage, sub node FRUs

High IO Optimized Platform
Elements of RSA Platform
RSA Pooled System Elements

- Intel® Processors optimized for Cloud
- Disaggregated Networking using Ethernet compute fabric
- Pooled Memory Controller
- POD Wide Storage
- Pooled System Management Engine (PSME)
- POD Manager
RSA Pooled System Elements

- Intel® Processors optimized for Cloud
- Disaggregated Networking using Ethernet compute fabric
- Pooled Memory Controller
- POD Wide Storage
- Pooled System Management Engine (PSME)
- POD Manager

Enabling Rack Scale innovation at Solution Level
**Ethernet Compute Fabric**

- Low latency compute fabric
- Flexible ports
- Embedded Intel Ethernet controller technology
- Forth generation Ethernet switch architecture
- Synergistic with Intel® Silicon Photonics 100GbE interconnect technology
Ethernet Compute Fabric

- Low latency compute fabric
- Flexible ports
- Embedded Intel Ethernet controller technology
- Fourth generation Ethernet switch architecture
- Synergistic with Intel® Silicon Photonics 100GbE interconnect technology

Ideal Networking solution for Rack Scale Architecture Data Center
Pooled NVM Express Solutions

- Enable disaggregation of NVM Express devices in cloud
- Assign high performance storage to nodes based on workload demand
- Enable host failover and NVM Express device failover
- Enables ease of workload migration in hyperscale cloud environment
Pooled NVM Express Solutions

- Enable disaggregation of NVM Express devices in cloud
- Assign high performance storage to nodes based on workload demand
- Enable host failover and NVM Express device failover
- Enables ease of workload migration in hyperscale cloud environment

NVM Express Pooling supports Software Defined Storage (SDS) through on demand tiered storage
Pooled Memory Solutions

- Large disaggregated memory pool using standard DIMMs (including NVDIMMs)
- Partitionable memory to nodes based on workload demand
- Support per node partition and sharable partition(s)
- Sharable partitions can be used for de-dup, VM Migration and other advanced functions
Pooled Memory Solutions

- Large disaggregated memory pool using standard DIMMs (including NVDIMMs)
- Partitionable memory to nodes based on workload demand
- Support per node partition and sharable partition(s)
- Sharable partitions can be used for de-dup, VM Migration and other advanced functions
Pooled Memory Solutions

- Large disaggregated memory pool using standard DIMMs (including NVDIMMs)
- Partitionable memory to nodes based on workload demand
- Support per node partition and sharable partition(s)
- Sharable partitions can be used for de-dup, VM Migration and other advanced functions

Memory Pooling drives Data Center efficiency as well as cloud innovation
RSA Storage Overview

POD Wide Storage Pool Based on configurable "RSA Storage Bricks"

Ethernet to DC Fabric
RSA Storage Overview

Modularize Storage hardware to economically support full range of Storage solutions
POD Wide Storage Architecture

- POD-wide supporting Services model
- Delivered on RSA Storage Bricks
- Dynamically Assigned Usage Specific Storage Service Stacks
- POD manager boots and “personalizes them” to meet RSA use cases
POD Wide Storage Architecture

- POD-wide supporting Services model
- Delivered on RSA Storage Bricks
- Dynamically Assigned Usage Specific Storage Service Stacks
- POD manager boots and “personalizes them” to meet RSA use cases

Enable Storage hardware with storage personalities for usage based architecture
RSA Management API
RSA Management Model

- RSA POD Manager
- Ethernet to DC Fabric
- TORS Manager
  - If TORS used
- Pooled System Manager (PSME)
- PDU and Rack Environment Manager (RMM)
- POD Storage
- POD
RSA Management Model

Modular, Expandable Management Architecture for Rack Scaling
RSA Management Hierarchy and API

- RSA API is REST based API interface with JSON encoding and utilizes HTTPS web protocol
  - Also exploring OData

- RSA API comprehends multi-node, pooled system

- RSA API is built using Red Fish as base profile (http://www.redfishspecification.org)

- RSA API requires consistency across vendor implementation

- For node and sub-node level, Intel recommends Red Fish as long term direction

Orchestration Partner Solutions

- OpenStack*
- VMware*
- Microsoft*

POD Manager

RSA Manageability Engine e.g. PSME

Node Node Node

Multi-Node
RSA Management Hierarchy and API

- **RSA API** is REST based API interface with JSON encoding and utilizes HTTPS web protocol
  - Also exploring OData

- **RSA API** comprehends multi-node, pooled system

- **RSA API** is built using Red Fish as base profile ([http://www.redfishspecification.org](http://www.redfishspecification.org))

- **RSA API** requires consistency across vendor implementation

- For node and sub-node level, Intel recommends Red Fish as long term direction

**Orchestration Partner Solutions**

- **OpenStack**
- **VMware**
- **Microsoft**

**POD Manager**

**RSA Manageability Engine** e.g. PSME

**Node**  **Node**  ...  **Node**

**uS**  **uS**

**Multi-Node**

**RSA API builds upon industry standards for ease of adoption**
RSA Management Architecture

- Topology discovery
- Disaggregated resource management
- Composable system support
- Support Compute, network, storage, NVM elements

Orchestration Partner Solutions

- OpenStack*
- VMware*
- Microsoft*

POD Manager

Discovery
Boot
Configuration
Power
Fault
Telemetry

RMM
Distributed Switch Manager

POD Fabric

RSA API

Rack Temp/Power

Storable Management

Operations/Node Management

Boot Service

Switch Configuration

NVM Mgmt

Operations Management

Rack Temp/Power

POD Fabric

IDF14
RSA Management Architecture

Orchestration Partner Solutions

- OpenStack*
- VMware*
- Microsoft*

POD Manager

- Discovery
- Boot
- Configuration
- Power
- Fault
- Telemetry

POD Fabric

Enables Cloud Integration through common API

- Topology discovery
- Disaggregated resource management
- Composable system support
- Support Compute, network, storage, NVM elements

POD Manager

RMM

Distributed Switch Manager

Rack Temp/Power

Storage Brick

Operations Management

NVM Mgmt

Switch Configuration

Boot Service

Operations/Node Management

Pooled System

RSA API

Discovery

Boot

Configuration

Power

Fault

Telemetry

Enables Cloud Integration through common API
RSA Reference Platform

- Half rack with shared cooling
- 2U – Xeon and Atom, 1/3 width boards
- Rack scale discovery enabled by Management back plane with rack management module
- Composability using POD Manager software
- Storage Node using Direct attached JBOD tray for Remote boot and CEPH Storage
- Configurable network architecture using Intel Ethernet Switch with Silicon photonics

Developed by Intel and Quanta
RSA Reference Platform

• Half rack with shared cooling
• 2U – Xeon and Atom, 1/3 width boards
• Rack scale discovery enabled by Management back plane with rack management module
• Composability using POD Manager software
• Storage Node using Direct attached JBOD tray for Remote boot and CEPH Storage
• Configurable network architecture using Intel Ethernet Switch with Silicon photonics

Reference platform demonstrates RSA value proposition and use cases
Summary & Call to Action
Summary

• RSA enables SDI – efficiency, agility, service compliance focus of SDI
• RSA arch enables solution scaling from half rack to large scale DC
• Consistent multi-vendor RSA API enables rapid solution integration and management stack innovation
Call to Action

• **Get educated** – Participate in detailed technical discussion, request overview white paper

• **Request** architecture, system management and manageability reference documents under NDA – Draft in Q4

• **Participate** in a hands-on hackathon in 1H 2015

• **Software Development platform** - Limited availability of development platform and API’s/Beta software for pilot engagements
INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL’S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL’S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS’ FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request. Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm

Intel, Xeon, Atom, Look Inside and the Intel logo are trademarks of Intel Corporation in the United States and other countries.

*Other names and brands may be claimed as the property of others. Copyright ©2014 Intel Corporation.
Risk Factors

The above statements and any others in this document that refer to plans and expectations for the second quarter, the year and the future are forward-looking statements that involve a number of risks and uncertainties. Words such as “anticipates,” “expects,” “intends,” “plans,” “believes,” “seeks,” “estimates,” “may,” “will,” “should” and their variations identify forward-looking statements. Statements that refer to or are based on projections, uncertain events or assumptions also identify forward-looking statements. Many factors could affect Intel's actual results, and variances from Intel's current expectations regarding such factors could cause actual results to differ materially from those expressed in these forward-looking statements. Intel presently considers the following to be important factors that could cause actual results to differ materially from the company's expectations. Demand for Intel's products is highly variable and, in recent years, Intel has experienced declining orders in the traditional PC market segment. Demand could be different from Intel's expectations due to factors including changes in business and economic conditions; consumer confidence or income levels; customer acceptance of Intel's and competitors' products; competitive and pricing pressures, including actions taken by competitors; supply constraints and other disruptions affecting customers; changes in customer order patterns including order cancellations; and changes in the level of inventory at customers. Intel operates in highly competitive industries and its operations have high costs that are either fixed or difficult to reduce in the short term. Intel's gross margin percentage could vary significantly from expectations based on capacity utilization; variations in inventory valuation, including variations related to the timing of qualifying products for sale; changes in revenue levels; segment product mix; the timing and execution of the manufacturing ramp and associated costs; excess or obsolete inventory; changes in unit costs; defects or disruptions in the supply of materials or resources; and product manufacturing quality/yields. Variations in gross margin may also be caused by the timing of Intel product introductions and related expenses, including marketing expenses, and Intel's ability to respond quickly to technological developments and to introduce new products or incorporate new features into existing products, which may result in restructuring and asset impairment charges. Intel's results could be affected by adverse economic, social, political and physical/infrastructure conditions in countries where Intel, its customers or its suppliers operate, including military conflict and other security risks, natural disasters, infrastructure disruptions, health concerns and fluctuations in currency exchange rates. Intel's results could be affected by the timing of closing of acquisitions, divestitures and other significant transactions. Intel's results could be affected by adverse effects associated with product defects and errata (deviations from published specifications), and by litigation or regulatory matters involving intellectual property, stockholder, consumer, antitrust, disclosure and other issues, such as the litigation and regulatory matters described in Intel's SEC filings. An unfavorable ruling could include monetary damages or an injunction prohibiting Intel from manufacturing or selling one or more products, precluding particular business practices, impacting Intel's ability to design its products, or requiring other remedies such as compulsory licensing of intellectual property. A detailed discussion of these and other factors that could affect Intel's results is included in Intel's SEC filings, including the company's most recent reports on Form 10-Q, Form 10-K and earnings release.

Rev. 4/15/14