

# Intel® Rack Scale Design

Architectural Requirements Specification

---

*August 2016*

*Revision 004*



No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and noninfringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

This document contains information on products, services, and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications, and roadmaps.

The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request.

Copies of documents that have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting <http://www.intel.com/design/literature.htm>.

Intel 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 © 2016 Intel Corporation. All rights reserved.



# Contents

<b>1</b>	<b>Introduction .....</b>	<b>6</b>
1.1	Disclaimer .....	6
1.2	Scope .....	6
1.3	Intended audience .....	6
1.4	Terminology .....	6
1.5	Conventions .....	7
1.6	Intel® Rack Scale Design platform overview .....	7
1.6.1	Intel® Rack Scale Design software Interfaces .....	8
1.6.2	Recommended Essential Intel® Rack Scale Design Hardware Elements .....	8
1.6.3	Platform hardware design guideline summary .....	9
<b>2</b>	<b>Intel® Rack Scale Design Platform Hardware Design General Specifications .....</b>	<b>11</b>
2.1.1	Rack must have one or more Pooled System Management Engine software (PSME) ...	11
2.1.2	Multiblade chassis capable of populating more than one blade inside the chassis .....	11
2.1.3	Shared or highly efficient power supply .....	11
2.1.4	Shared power bus bar across entire rack .....	11
2.1.5	Shared or highly efficient cooling .....	11
2.1.6	Centralized cooling across entire rack .....	12
2.1.7	JBOD support .....	12
2.1.8	Compute blade with one or more HDDs .....	12
2.1.9	Composed node with M.2 drive .....	12
2.1.10	At least one Intel® Rack Scale Design compute Module in POD .....	12
2.1.11	Compute module serviceability independence .....	12
2.1.12	Ethernet-based fabric for network connectivity .....	12
2.1.13	At least one Ethernet switch per rack .....	13
2.1.14	At least one Ethernet switch in the pod .....	13
2.1.15	Network switch support for network software agent .....	13
2.2	Components location identification support .....	13
2.2.1	Unique drawer ID number within the rack .....	14
2.2.2	Drawer ID numbering from bottom to top .....	14
2.2.3	Unique module ID number within the drawer .....	14
2.2.4	Module ID numbering consistency .....	14
2.2.5	Unique blade ID number within the module .....	14
2.2.6	Blade ID numbering consistency .....	15
2.2.7	Power supply position numbering consistency .....	15
2.2.8	Fan position numbering consistency .....	15
2.2.9	Unique ID for all FRUs .....	15
2.3	Intel® Rack Scale Design fabric and secure management network configuration .....	15
2.3.1	Management network and production network separation .....	15
<b>3</b>	<b>Composed Node Design Guidelines .....</b>	<b>16</b>
3.1	Node reset, power, and performance .....	16
3.1.1	Node reset support .....	16
3.1.2	Module power monitoring support .....	16
3.1.3	Module power budget support .....	16
<b>4</b>	<b>PSME Design Guidelines .....</b>	<b>17</b>
4.1	PSME configuration management .....	17
4.1.1	PODM-to-PSME communication via private network .....	17
4.2	PSME managed assets .....	17
4.2.1	Blade identification consistency .....	17



<b>5</b>	<b>RMM Design Guidelines.....</b>	<b>18</b>
5.1	RMM overview .....	18
5.1.1	PODM-to-RMM communication via private network .....	18
5.1.2	RMM-to-PSME communication via private network .....	18
5.1.3	Rack power monitoring support .....	18
5.2	RMM general support.....	18
5.2.1	High availability RMM support.....	18
<b>6</b>	<b>POD Manager Design Guidelines.....</b>	<b>19</b>
6.1	PODM configuration management.....	20
6.1.1	Secure communication channel for management network.....	20
<b>A</b>	<b>Intel® Rack Scale Design Platform Hardware Alignment Checklist.....</b>	<b>20</b>

## Figures

Figure 1	Pod logical hierarchy.....	7
Figure 2	API blocks.....	8
Figure 3	Intel® Rack Scale Design component location identification.....	14
Figure 4	Pod manager logical view .....	19

## Tables

Table 1	Terminology .....	6
Table 2	Intel® Rack Scale Design platform design guideline summary .....	9
Table 3	Design criteria .....	20

## §



## Revision History

---

Revision	Description	Date
0.3	Addressed additional user feedback	August 04, 2016
0.2	Addressed errors and user content feedback	August 21, 2015
0.1	Initial release.	August 17, 2015

§



# 1 Introduction

This document provides general hardware guidelines for Intel's development partners to align with Intel® Rack Scale Design specifications for hardware. The complete Intel® Rack Scale Design platform design specifications are available to Intel partners and customers under NDA from your Intel field representative.

## 1.1 Disclaimer

Technologies and concepts expressed in this document are in pathfinding and may go through substantial changes before being productized. No product commitment or even readiness should be inferred from this document.

## 1.2 Scope

The scope of this document is platform hardware design for Intel® Rack Scale Design hardware components. Time horizon: Guidance listed here covers for designs that target 2016 or later commercial platforms.

## 1.3 Intended audience

The intended audiences for this document are hardware vendors (for example, OEMs/ODMs) who build Intel® Rack Scale Design platforms or components that are integrated into the Intel® Rack Scale Design platform.

## 1.4 Terminology

**Table 1** Terminology

Term	Definition
API	Application program interface. A set of routines, protocols, and tools for building software applications. API defines operations, inputs, and outputs.
BIOS	Basic input/output system. To initialize and test compute/storage node hardware components, and to load a boot loader or an operating system from a mass memory device. The BIOS supports UEFI interface.
BMC	Baseboard management controller. A specialized service processor that monitors the physical state of a computer and provides services to monitor and control certain node operations. The BMC supports Intelligent Platform Management Interface (IPMI).
composed node	Node composed by PODM. PODM composes nodes within the rack by communicating with PSME for compute node and storage allocation based on the user input.
CPP	Control Plane Process or is the Switch Management CPU. This is the host that Intel uses to run the PSME on the Intel Software Delivery Vehicle reference platform.
DMC	Drawer management controller. Controller that manages the drawer, where the PSME functionality is normally implemented.
EORS	End-of-row switch. Switch that is connected to the end of the row, either through copper or fiber.
HA	High availability. Generally a redundancy component available. For example, if a rack supports HA RMM, then more than one RMM is present in the rack and if primary RMM fails, a secondary RMM provides the RMM functionality.
iPXE	An open-source implementation of the PXE client firmware and bootloader.
MMC	Module management controller. The controller that manages the blades in the module.
node	Any compute node, such as an Intel® Xeon® or Intel® Atom™ processor, in the module under a drawer.
pod	A collection of racks within a shared infrastructure management domain.
PODM	Pod manager. Logical management functionality across all infrastructure in a pod.
PSME	Pooled system management engine. A microcontroller responsible for configuration of shared and pooled memory controlled by the SMC, pooled storage by the PNC, the nodes and network (SDN) of the compute nodes and switch. Also known as DMC (drawer management controller).
PTAS	Power thermal aware scheduling.

Term	Definition
PXE	Preboot eXecution Environment. A specification that allows devices to boot over a network.
REST	Representational State Transfer
RMM	Rack management module. Module that is responsible for managing the rack, which normally assigns IDs for PSME and manages the rack power and cooling.
TORS	Top-of-rack switch. Switch that is connected to the top of the rack, either through copper or fiber (but mostly fiber).

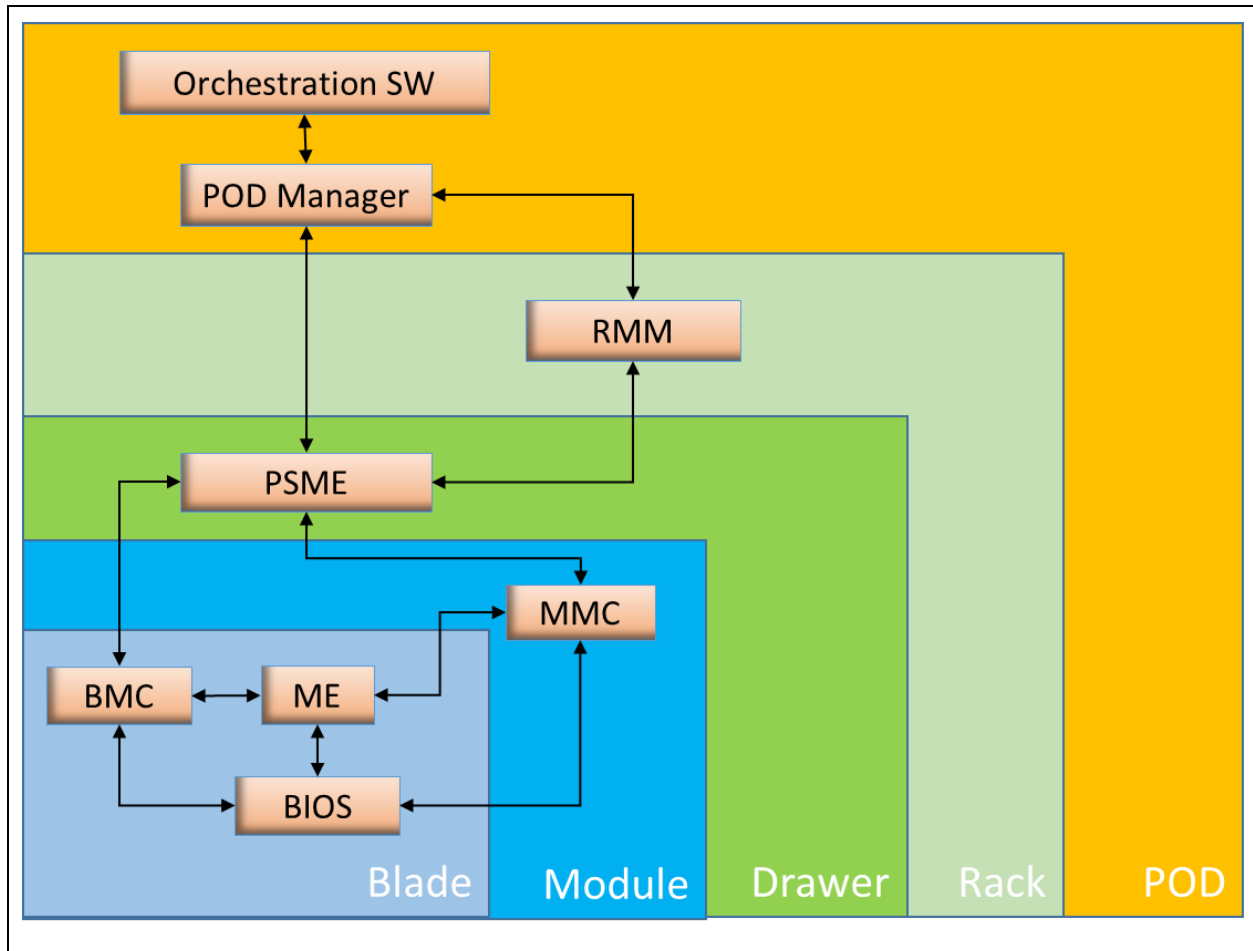
## 1.5 Conventions

The key words/phrases "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#).

## 1.6 Intel® Rack Scale Design platform overview

Figure 1 illustrates the various elements for the Intel® Rack Scale Design platform in a pod logical view.

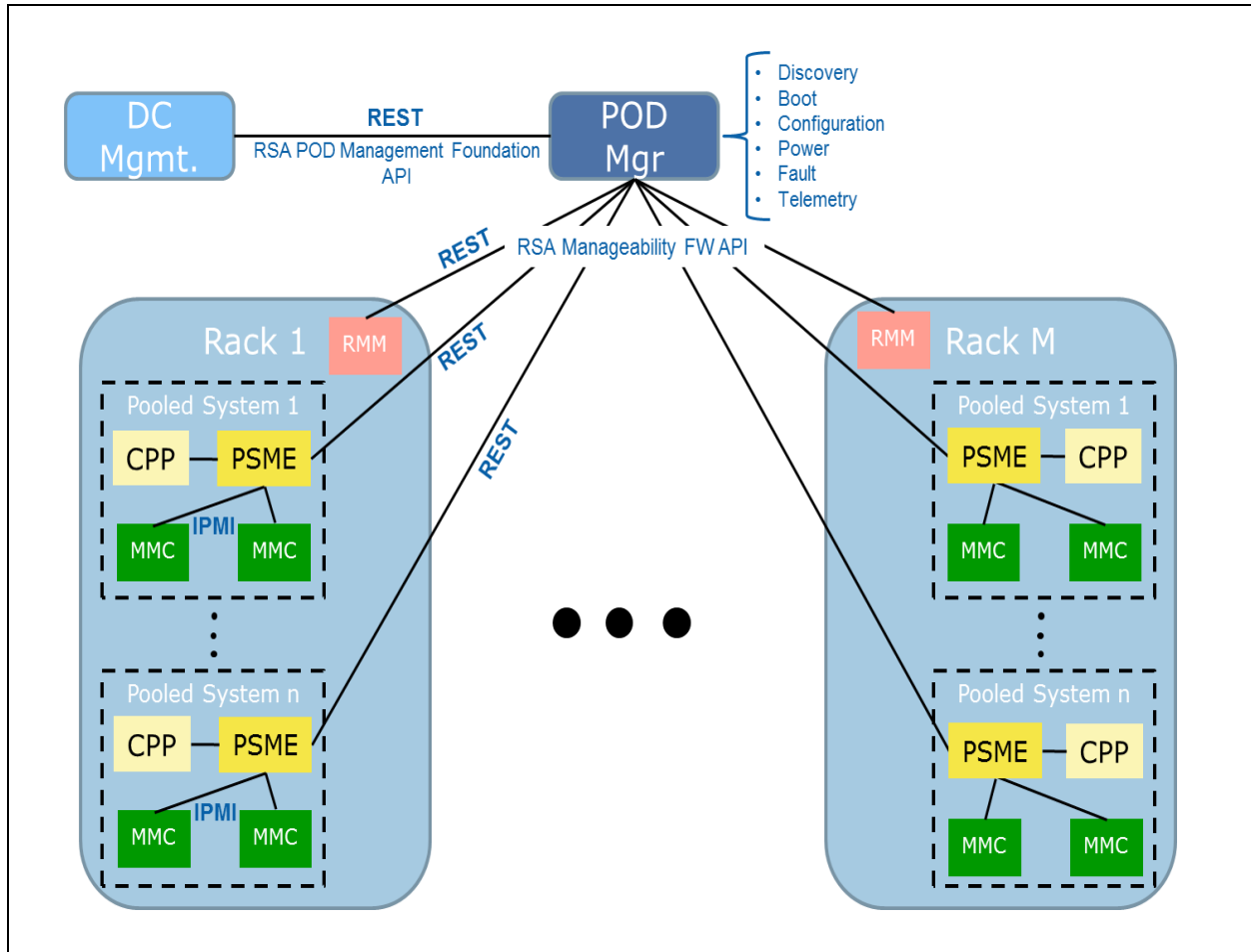
**Figure 1 Pod logical hierarchy**



## 1.6.1 Intel® Rack Scale Design software Interfaces

A complete hardware solution will require a management software interface with a set of APIs that facilitate use cases. An example of a REST-based API framework for the Intel® Rack Scale Design PODM, RMM, and PSME is shown in Figure 2.

**Figure 2** API blocks



Customers should check <https://github.com/01org/intelRSD> to download the latest available onboard device drivers, system firmware, and system software. For further assistance, please contact the Intel Field Representative.

## 1.6.2 Recommended Essential Intel® Rack Scale Design Hardware Elements

Intel recommends that all of the following items be incorporated into 2016 Intel® Rack Scale Design hardware designs in order to be aligned with Intel vision and intent. The latter part of this document describes these criteria in more detail.

Intel® Rack Scale Design Systems should incorporate the following essential elements:

- At least one PSME in each rack
- Shared or highly efficient power supply
- Shared or highly efficient cooling
- At least one compute blade node





- Ethernet-based fabric for network connectivity
- At least one Ethernet switch in the pod
- Unique drawer ID number within the rack
- Unique module ID number within the drawer
- Unique blade ID number within the module
- Management network and production network separation
- Node reset support
- Power monitoring and power budget support
- Secure Private rackwide network
- Blade presence identification
- RMM – private rackwide management network
- Secure communication channel for management network

### 1.6.3 Platform hardware design guideline summary

This section lists the complete requirements for defining how the platform is designed to conform to the Intel® Rack Scale Design.

Intel considers a hardware system to be “Intel® Rack Scale Design aligned” if 75% of the items in this section are intended to be implemented by the partner by the end of 2016 as well as all of the items listed in section 1.6. Items listed as “Required” are more critical to use cases in the near term as compared to “Optional” or “Recommended” items.

The items listed in the previous section are essential for Intel® Rack Scale Design alignment as fundamental currently available building blocks. The remaining items in this section are highly recommended to be incorporated into the partner’s roadmap in the 2016 time horizon.

Note that each row of Table 2 contains a reference to a section in the appendix which contains a more complete description of the design criteria.

**Table 2 Intel® Rack Scale Design platform design guideline summary**

Section	Intel® Rack Scale Design criteria	Design criticality
2.1.1	Rack must have one or more Pooled System Management Engine software (PSME)	Required
2.1.2	Multiblade chassis capable of populating more than one blade inside the chassis	Recommended
2.1.3	Shared or highly efficient power supply	Required
2.1.4	Shared power bus bar across entire rack	Recommended
2.1.5	Shared or highly efficient cooling	Recommended
2.1.6	Centralized cooling across entire rack	Recommended
2.1.7	JBOD support	Optional
2.1.8	Compute blade with one or more HDDs	Optional
2.1.9	Composed node with M.2 drive	Optional
2.1.10	At least one Intel® Rack Scale Design compute module in POD	Required
2.1.11	Compute module serviceability independence	Required
2.1.12	Ethernet-based fabric for network connectivity	Required
2.1.13	At least one Ethernet switch per rack	Optional
2.1.14	At least one Ethernet switch in the pod	Required
2.1.15	Network switch support for network software agent	Required
2.2.1	Unique drawer ID number within the rack	Required
2.2.2	Drawer ID numbering from bottom to top	Recommended
2.2.3	Unique module ID number within the drawer	Required



Section	Intel® Rack Scale Design criteria	Design criticality
2.2.4	Module ID numbering consistency	Recommended
2.2.5	Unique blade ID number within the module	Required
2.2.6	Blade ID numbering consistency	Recommended
2.2.7	Power supply position numbering consistency	Required
2.2.8	Fan position numbering consistency	Required
2.2.9	Unique ID for all FRUs	Recommended
2.3.1	Management network and production network separation	Required
3.1.1	Node reset support	Required
3.1.2	Module power monitoring support	Recommended
3.1.3	Module power budget support	Recommended
4.1.1	PODM-to-PSME Communication Via Private Network	Recommended
4.2.1	Blade identification consistency	Required
5.1.1	PODM-to-RMM Communication Via Private Network	Required
5.1.2	RMM-to-PSME Communication Via Private Network	Required
5.1.3	Rack power monitoring support	Required
5.2.1	High availability RMM support	Optional
6.1.1	Secure communication channel for management network	Required





## 2 Intel® Rack Scale Design Platform Hardware Design General Specifications

---

This section describes Intel® Rack Scale Design platform-level feature design guidelines. Subsequent sections describe Intel® Rack Scale Design platform subcomponent level design guidelines that apply to platform hardware. The Intel® Rack Scale Design platform meets the following generic requirements as stipulated (optional, recommended, or required) in each section.

### 2.1.1 Rack must have one or more Pooled System Management Engine software (PSME)

Design criticality: Required

A rack is made-up of drawers. An Intel® Rack Scale Design system must provide a mechanism to manage rack level end point components down to the drawer level. The PSME provides management interface to manage the modules/blades at a drawer level. A rack must contain one or more PSME.

In some cases, a PSME may service multiple drawers, as long as the drawer is uniquely addressable and provides the necessary instrumentation. For example, if each drawer has a microcontroller to provide the necessary instrumentation for all drawer requirements (such as module presence detection) and is interfaced to the RMM, then the drawer PSME could physically run in the RMM and represent each drawer instance to meet this requirement.

### 2.1.2 Multiblade chassis capable of populating more than one blade inside the chassis

Design criticality: Recommended

It is recommended to have a drawer to support multiple blades and share a power supply across multiple blades in the platform. This feature allows modularity for upgrade, serviceability, and to effectively utilize the real estate footprint.

### 2.1.3 Shared or highly efficient power supply

Design criticality: Required

Compute/Storage Modules must support cost effective, efficient, and manageable shared power. The solution is achieved by either, 1) sharing power across two or more Modules, or 2) having a non-redundant > 90% efficient (delivered power to Module/input AC power) power configuration. Efficient data center power delivery is necessary to reduce operating cost which is a core value proposition for Intel® Rack Scale Design.

If shared power is used, then the shared cooling solution is recommended.

### 2.1.4 Shared power bus bar across entire rack

Design criticality: Recommended

Use of shared power bus bar across the entire rack would allow consolidation of power supplies, reduce real estate usage, and reduce cost.

### 2.1.5 Shared or highly efficient cooling

Design criticality: Recommended

Compute/ Storage Modules recommended to support cost-effective and manageable shared cooling to positively affect TCO. The platform recommended to support shared cooling across two or more modules. One option to



achieve efficient cooling is to use a shared cooling fan larger than 2U (3.5 inches) in diameter that can cool multiple modules simultaneously. If the system does not utilize fans, the system could implement a more efficient system than shared cooling, such as a liquid cooled system.

If shared power is used, then the shared cooling solution is recommended.

## 2.1.6 Centralized cooling across entire rack

Design criticality: Recommended

Centralized cooling generally increases efficiency and reduces cost when the racks are fully populated. It is recommended that the platforms be designed with centralized cooling across the entire rack. In an Intel® Rack Scale Design, centralized cooling increases the platform cooling efficiency and operating cost.

## 2.1.7 JBOD support

Design criticality: Optional

JBOD allows for efficient pooling of storage resources. Composed nodes could have local storage, JBOD, network-based storage, or a combination of these. This feature allows efficient pooling of storage resources as well as more efficient management of storage resources.

## 2.1.8 Compute blade with one or more HDDs

Design criticality: Optional

An Intel® Rack Scale Design compute blade has iSCSI or iPXE support, therefore, the boot or storage could be accessed through the network. Compute blades sometimes use local storage such as HDD for boot or delta-file storage.

The benefit of local compute HDD is that, in the event of network errors, the system may be able to store the errors as well as the current data in the local storage if local storage is present.

## 2.1.9 Composed node with M.2 drive

Design criticality: Optional

Use of M.2 as local storage would provide better performance, compared to an HSD.

In an Intel® Rack Scale Design the M.2 drive will require less rack real estate as compared to drive bay-based local storage.

## 2.1.10 At least one Intel® Rack Scale Design compute Module in POD

Design criticality: Required

The pod must have at least one compute drawer with at least one compute module. It is possible that some racks support storage and some support compute only, but a pod must have at least one compute module. A compute module is essential as it is the most basic building block for the rack.

## 2.1.11 Compute module serviceability independence

Design criticality: Required

Compute modular must support modular CPU and memory that can be serviced or upgraded independent of shared resources (shared power, shared cooling, shared network, shared storage).

## 2.1.12 Ethernet-based fabric for network connectivity

Design criticality: Required



The platform must support Ethernet-based fabric for network connectivity.

The use of a common ubiquitous transport is important for standard Intel® Rack Scale Design and interoperability. Ethernet is a network used widely today and Intel® Rack Scale Design devices need to support widely used network technologies.

### 2.1.13 At least one Ethernet switch per rack

Design criticality: Optional

The network in the rack is generally connected via a disaggregated switch or TORS, but it is possible to build an Intel® Rack Scale Design system with the network connected to an EORS with a patch panel in the compute rack. The Ethernet switch is necessary to provide connectivity from the rack to the rest of the datacenter and the user workloads.

### 2.1.14 At least one Ethernet switch in the pod

Design criticality: Required

The pod must have at least one Ethernet switch component to connect from the pod to the external network. The Ethernet switch is necessary to provide PODM connectivity from the rack to the user workloads.

### 2.1.15 Network switch support for network software agent

Design criticality: Required

Network switch components (such as TORS and disaggregated switches) must support running a network software agent to monitor and configure the switches. The combined network switch and network software agent provide a mechanism to deliver security fixes, performance fixes, and bug fixes, which will keep the system up to date.

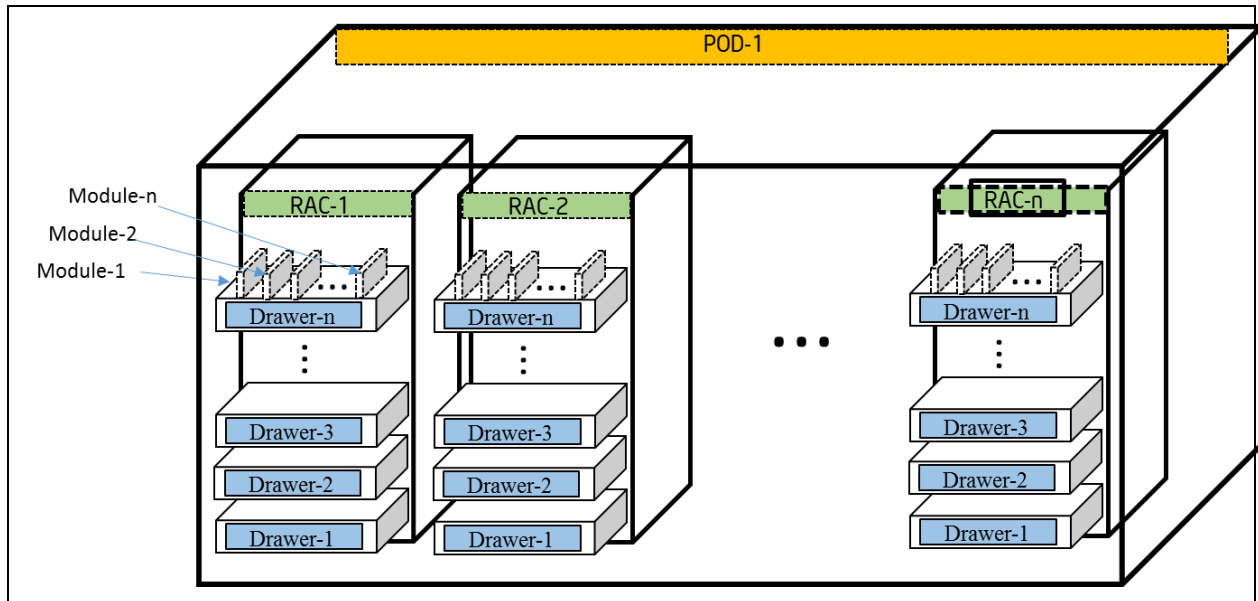
## 2.2 Components location identification support

A key attribute of Intel® Rack Scale Design management is location-aware discovery. A datacenter manager should be able to identify the physical location of hardware components to service them. The requirements below suggest a specific identification method (for example, left to right). An alternative identification method can be used as long as it is unique, communicates the location, is consistently used across the entire implementation, and is supported in the management interfaces (for example, right to left).

Identification support is essential to Intel® Rack Scale Design use cases because they enable the consistent disaggregation, assembly, and disassembly of assets in pooled system management. The Intel® Rack Scale Design system requires a mechanism to identify the components. The specifications in this section propose a mechanism for consistent tagging that will enable efficient management.

To support location identification, each Intel® Rack Scale Design component must meet the following requirements, as shown in Figure 3.

**Figure 3 Intel® Rack Scale Design component location identification**



### 2.2.1 Unique drawer ID number within the rack

Design criticality: Required

The drawers within the rack must have a unique number based on where the drawer is populated in the rack. The scope of the drawer ID is the rack.

### 2.2.2 Drawer ID numbering from bottom to top

Design criticality: Recommended

It is recommended that drawer numbers use the base as 1, and be numbered from bottom to top within the rack based on the physical location. If any location is not populated, then that location is skipped.

### 2.2.3 Unique module ID number within the drawer

Design criticality: Required

The modules within the rack must have a presence identification and unique number based on where the module is populated in the rack. The scope of the module ID is the drawer.

### 2.2.4 Module ID numbering consistency

Design criticality: Recommended

It is recommended that module numbers use the base as 1 and be numbered from left to right or right to left or front to back or back to front or top to bottom or bottom to top within the drawer, based on the physical location. If any module is not populated, then that location indicates module not present. If the system modules are aligned vertically, then they should be numbered as described, bottom to top.

### 2.2.5 Unique blade ID number within the module

Design criticality: Required

The blades within the module must have a presence identification and unique number based on where the blade is populated in the drawer. The scope of the blade ID is the module.



### 2.2.6 Blade ID numbering consistency

Design criticality: Recommended

It is recommended that blade numbers use the base as 1 and be numbered from left to right or right to left or front to back or back to front or top to bottom or bottom to top within the module, based on physical location. If a blade is not populated, then that location shall indicate blade not present.

### 2.2.7 Power supply position numbering consistency

Design criticality: Required

Service personnel should be able to easily identify the location of a power supply failure for replacement. It is recommended that the power supply position location label use the base as 1 and be numbered from left to right or right to left or front to back or back to front or top to bottom or bottom to top, within each subsystem (rack, drawer, or module). The location information could be implemented using straps or switches.

### 2.2.8 Fan position numbering consistency

Design criticality: Required

Service personnel should be able to easily identify the location of a failed fan for replacement. It is recommended that the fan position location label use the base as 1 and be numbered from left to right or right to left or front to back or back to front or top to bottom or bottom to top, within each subsystem (rack, drawer, or module). The location information could be implemented using straps or switches.

### 2.2.9 Unique ID for all FRUs

Design criticality: Recommended

It is recommended that each major field replaceable unit (FRU) in the rack (such as TORS, EORS, modules, and blades) have a GUID (such as manufacture ID, model number, and serial numbers) to help identify components for service. Specific FRU ID facilitates finding and fixing the right component when that component fails or requires service.

## 2.3 Intel® Rack Scale Design fabric and secure management network configuration

The RMM, PSME, CPP, TORS, and EORS should be able to communicate to their upstream ports, come up with IP address/port mappings, and respond to Intel® Rack Scale Design API requests.

### 2.3.1 Management network and production network separation

Design criticality: Required

The platform must support a separate IP address for Intel® Rack Scale Design management-related communication (control plane) such as PODM, RMM, and PSME configuration, and for production network (data plane).

A separate management network is essential to preventing unauthorized access from general access. Normally during the power up sequence, components go through an initialization and authentication process in which the network component may be vulnerable to an attack. This requirement attempts to prevent such attacks.





## 3 Composed Node Design Guidelines

---

This section describes Intel® Rack Scale Design platform compute/storage node design guidelines. A composed node is generally a combination of a compute blade and a module with network and/or storage connectivity. In some cases, the module and blade may be a single hardware element.

### 3.1 Node reset, power, and performance

#### 3.1.1 Node reset support

Design criticality: Required

The PSME should provide individual node reset support including power on and power off of the drawer and modules that are managed by PSME.

#### 3.1.2 Module power monitoring support

Design criticality: Recommended

Intel® Rack Scale Design recommends power monitoring at drawer, module and blade level. This feature helps PODM to take action on specific module/blade to keep rack and drawer within its power budget. It is recommended to follow specification [Intel® Intelligent Power Node Manager 3.0 External Interface Specification Using IPMI \(332200\)](#) or a later version or functional equivalent.

#### 3.1.3 Module power budget support

Design criticality: Recommended

Intel® Rack Scale Design recommends power control logic for the BMC/PSME to limit the power to the module/blades. It is recommended to follow specification [Intel® Intelligent Power Node Manager 3.0 External Interface Specification Using IPMI \(332200\)](#) or a later version or functional equivalent.





## 4 PSME Design Guidelines

---

The PSME is responsible for drawer identification management, as well as supporting the PSME Intel® Rack Scale Design API, and communicating with the BMC to perform node-level management. If the RMM is not present in the rack, one of the PSMEs in the rack would provide the RMM functionality. On-storage bricks may not have a PSME, in which case the BMC would perform PSME functionality and provide PSME Intel® Rack Scale Design API support.

### 4.1 PSME configuration management

#### 4.1.1 PODM-to-PSME communication via private network

Design criticality: Required

PSME to PODM and PSME to RMM must be connected through a private rackwide management network for the reasons outlined in section 2.3.1. The PSME must get its IP address from the datacenter admin before it communicates any management information.

### 4.2 PSME managed assets

#### 4.2.1 Blade identification consistency

Design criticality: Required.

The PSME should provide blade presence and location within the module identification.





## 5 RMM Design Guidelines

---

### 5.1 RMM overview

The rack manager module (RMM) is responsible for handling infrastructure functions such as power, cooling, and assigning PSME IDs.

#### 5.1.1 PODM-to-RMM communication via private network

Design criticality: Required

The RMM must be connected to a rackwide private network for the reasons outlined in section 2.3.1. The RMM must be able to get its IP address from the datacenter administrator before it communicates to the PODM.

#### 5.1.2 RMM-to-PSME communication via private network

Design criticality: Required.

The RMM and PSME must communicate via private network for the reasons outlined in section 2.3.1.

#### 5.1.3 Rack power monitoring support

Design criticality: Required

Intel® Rack Scale Design must support power monitoring at rack level. This feature helps PODM to take actions to keep the rack within its power budget.

### 5.2 RMM general support

#### 5.2.1 High availability RMM support

Design criticality: Optional

For racks that require high availability (HA), multiple RMMs could be incorporated into the rack. The following are recommended for HA RMM implementation:

- When multiple RMMs are present, there must be only one active RMM; others are passive RMMs.
- The active RMM and passive RMMs must coordinate such that a copy of the active RMM's data will be in sync with the passive RMMs.
- All RMMs must be on the rackwide private network.
- A passive RMM could implement a heartbeat mechanism to detect a failure of the active RMM. When multiple RMMs are present, generally a priority order will be assigned to determine who will take over upon failure of the active RMM.
- If an RMM is hot-added, the active RMM must provide the current configuration information, such as PSME ID, to the newly added RMM.
- The use of multiple RMM's enhances system availability through redundancy.

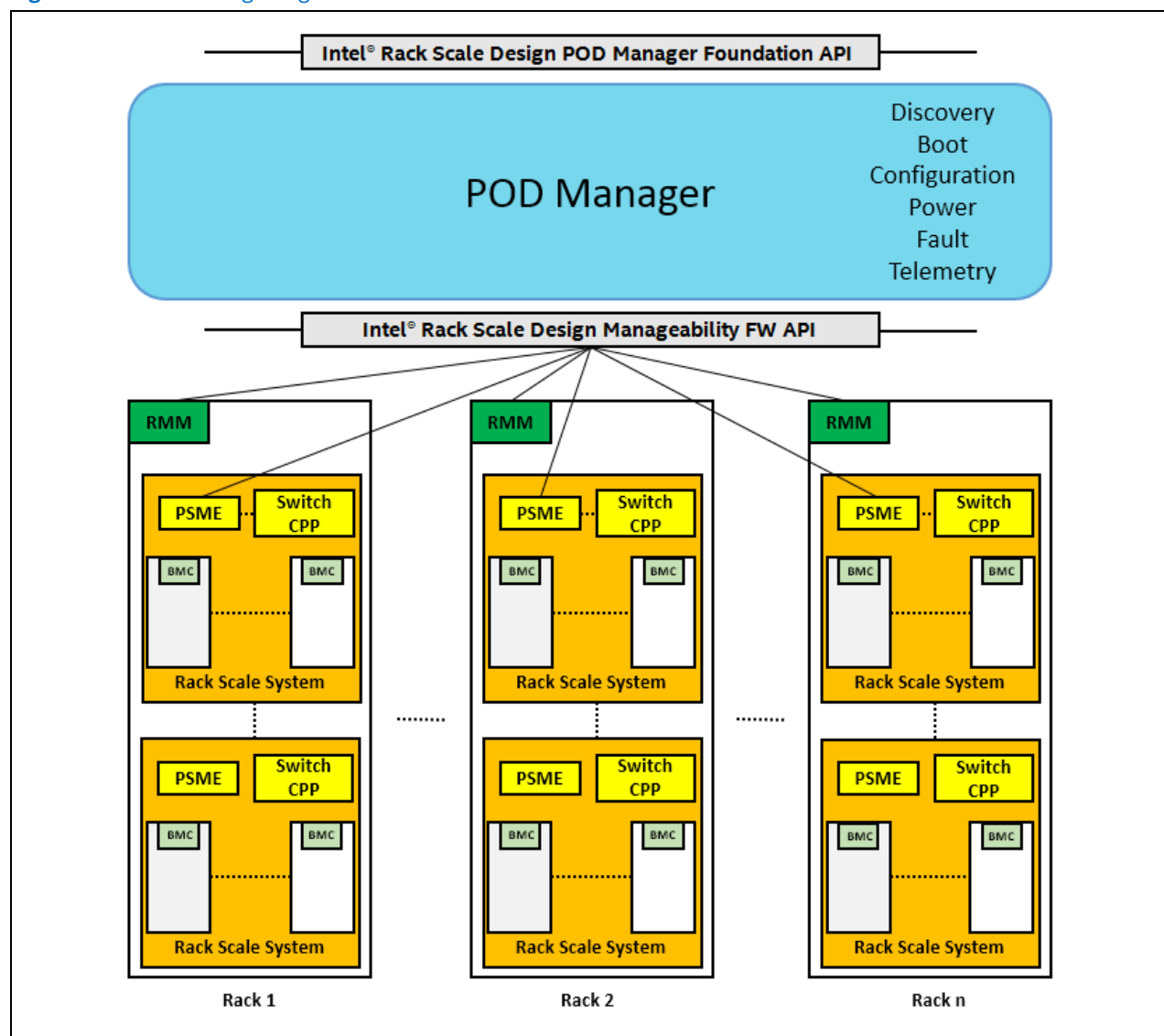
## 6 POD Manager Design Guidelines

The pod manager, as shown in Figure 4, is responsible for discovery of resources in the pod, configuring the resources, power and reset control, power management, fault management, monitoring the resources usage, and composing a server based on Intel® Rack Scale Design specifications.

The pod manager interacts with RMMs in the rack, PSMEs, and CPPs to create representation of the Intel® Rack Scale Design pod.

The pod manager allows composing a physical node to match the logical node requirements specified by the solution stack. Such composition is able to specify a system at a sub-composed node granularity, as shown in Figure 4.

**Figure 4** Pod manager logical view





## 6.1 PODM configuration management

### 6.1.1 Secure communication channel for management network

Design criticality: Required

The PODM must be connected to the RMM and PSME through a private network for the reasons outline in section 2.3.1. Any management related activity such as reconfiguration must be performed only after establishing secure a communication channel between the PODM to the PSME and the PODM to the RMM.

# A Intel® Rack Scale Design Platform Hardware Alignment Checklist

Table 3 lists the required criteria for incorporating into a 2016 Intel® Rack Scale Design platform design. Hardware manufacturers and designers can use this checklist as a guide to test how closely the design conforms to the most critical requirements.

**Table 3** Design criteria

Section	Intel® Rack Scale Design criteria	Design checklist	Comments
2.1.1	Rack must have one or more Pooled System Management Engine software (PSME)		
2.1.2	Multiblade chassis capable of populating more than one blade inside the chassis		
2.1.3	Shared or highly efficient power supply		
2.1.4	Shared power bus bar across entire rack		
2.1.5	Shared or highly efficient cooling		
2.1.6	Centralized cooling across entire rack		
2.1.7	JBOD support		
2.1.8	Compute blade with one or more HDDs		
2.1.9	Composed node with M.2 drive		
2.1.10	At least one Intel® Rack Scale Design compute module in POD		
2.1.11	Compute module serviceability independence		
2.1.12	Ethernet-based fabric for network connectivity		
2.1.13	At least one Ethernet switch per rack		
2.1.14	At least one Ethernet switch in the pod		
2.1.15	Network switch support for network software agent		
2.2.1	Unique drawer ID number within the rack		
2.2.2	Drawer ID numbering from bottom to top		
2.2.3	Unique module ID number within the drawer		
2.2.4	Module ID numbering consistency		
2.2.5	Unique blade ID number within the module		
2.2.6	Blade ID numbering consistency		
2.2.7	Power supply position numbering consistency		
2.2.8	Fan position numbering consistency		
2.2.9	Unique ID for all FRUs		
2.3.1	Management network and production network separation		
3.1.1	Node reset support		
3.1.2	Module power monitoring support		



Section	Intel® Rack Scale Design criteria	Design checklist	Comments
3.1.3	Module power budget support		
4.1.1	PODM-to-PSME Communication Via Private Network		
4.2.1	Blade identification consistency		
5.1.1	PODM-to-RMM Communication Via Private Network		
5.1.2	RMM-to-PSME Communication Via Private Network		
5.1.3	Rack power monitoring support		
5.2.1	High availability RMM support		
6.1.1	Secure communication channel for management network		