整机柜架构中的电源与散热管理
Power and Thermal Management in Rack Scale Architecture

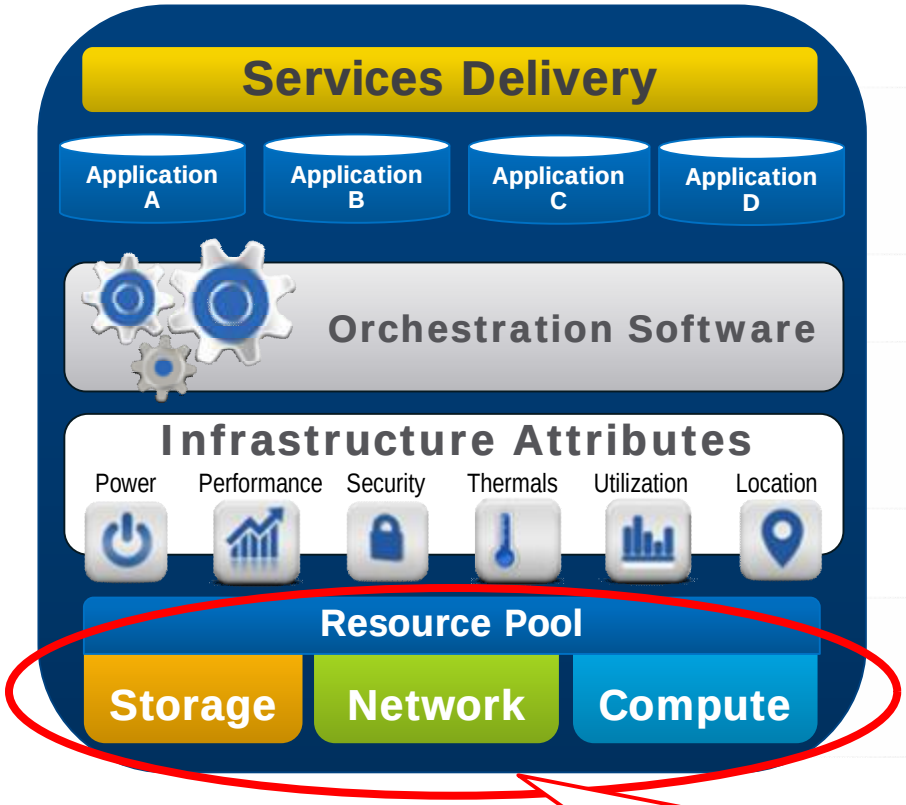
Aug 29, 2014

Legal Disclaimer

Agenda

- **Rack Scale Management Architecture Overview**
- **Power Events Management in Rack Scale Architecture**
- **Advanced Thermal Management in Rack Scale Architecture**
- **Summary**
- **Q&A**

Software Defined Infrastructure



SERVICE ASSURANCE

Policy based automation provides dynamic provisioning and service assurance as applications are deployed and maintained

PROVISIONING MANAGEMENT

Orchestration provisions and optimally allocates resources based on the unique requirements of an application

POOLED RESOURCES

Network, Storage and Compute elements are abstracted into resource pools

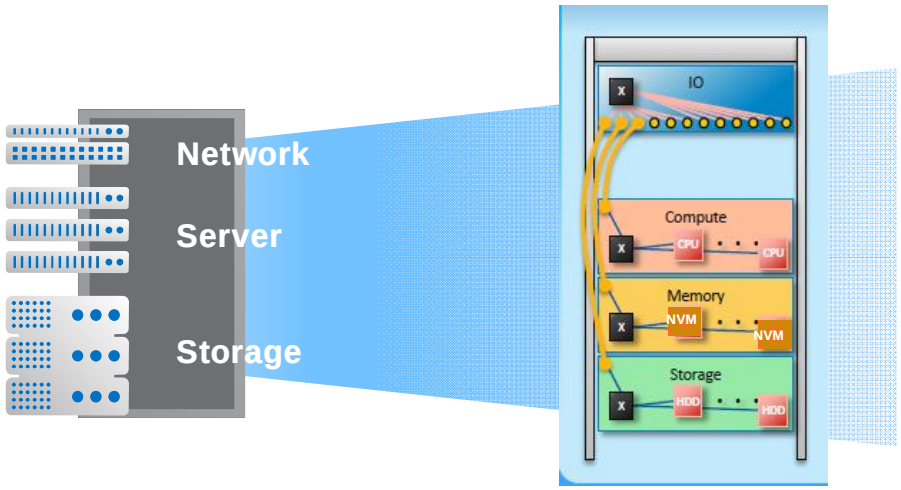
RSA optimizes systems to run your SDI solution

intel

Intel® Rack Scale Architecture – Optimized for SDI

Discrete Components, Self-Integration

Composable set of pooled and disaggregated resources



Today **RSA**

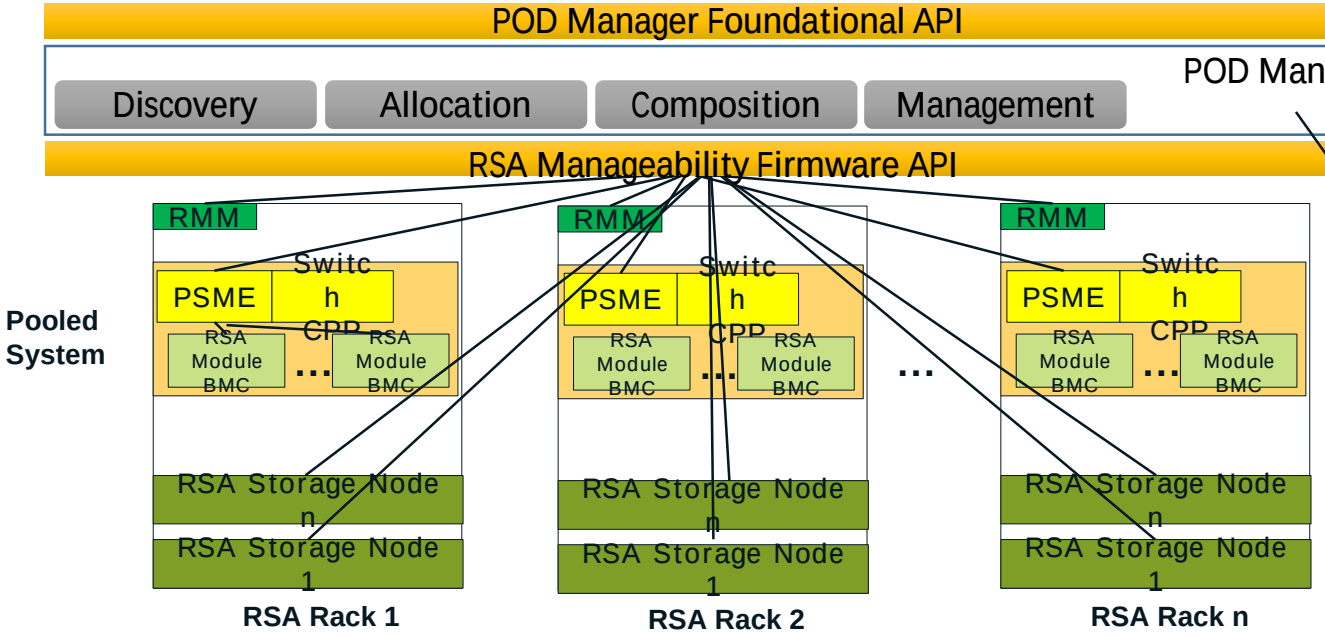
- Rack level pooled resources with discoverability & serviceability for uninterrupted maintenance
- Enables orchestration software to compose server, increase rack density & utilization
- Enables service assurance software for optimization & automation

Up to 5X[2]	Up to 3X[2]	Up to 1.5X[1]	Up to 25X[1]
Provisioned Power	**Cables**	**Density**	**NW Downlink**

Flexibility, Capital Efficiency, Lower TCO

intel

RSA Management Architecture
System Level Strawman



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

The "actor" that creates a manageable composite system

Unified Management APIs to support flexible and scalable usage models

Agenda

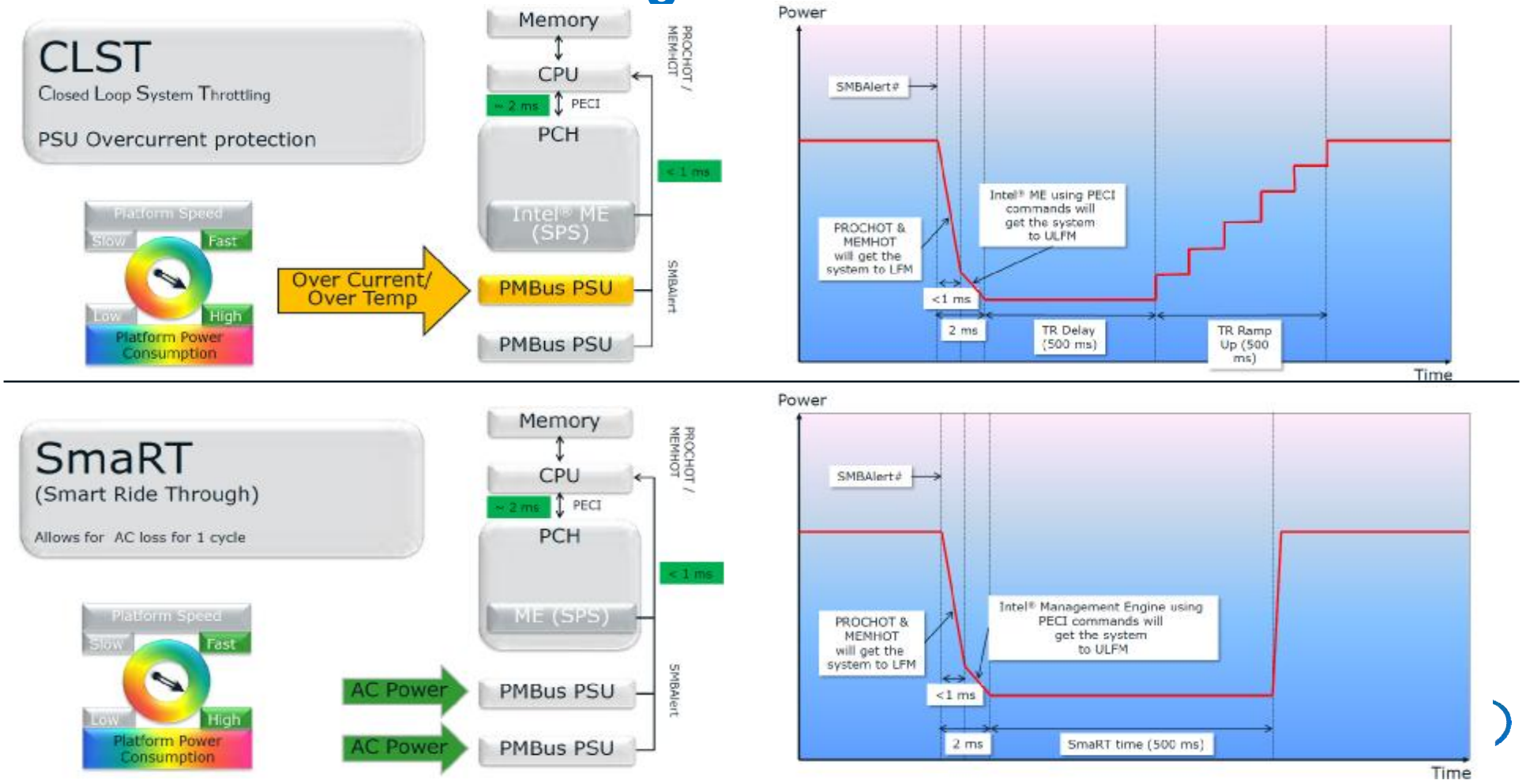
- Rack Scale Management Architecture Overview
- **Power Events Management in Rack Scale Architecture**
- Advanced Thermal Management in Rack Scale Architecture
- Summary
- Q&A

Power Events in Rack Scale Architecture

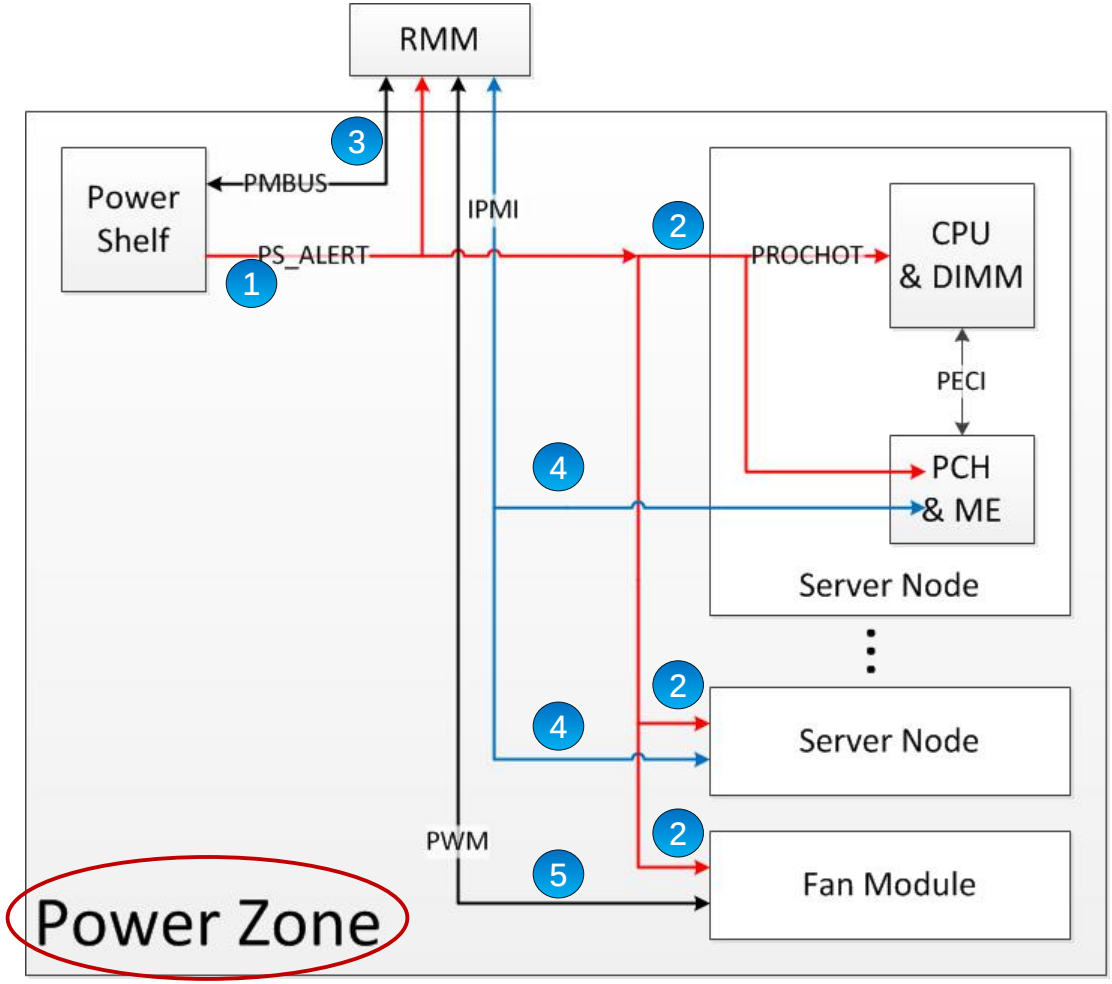
- ü Over current
- ü Power supply module overheat
- ü Power supply module failure
- ü Main power temporary loss / interrupt
- ü Main power out of range
- ü …

Shared Power Supply in RSA Requires More Sophisticated Power Events Management

Power Events Management: CLST and SMART



Rack Scale SMART/CLST Concept



(1) **Power Shelf asserts PS_ALERT upon the power event occurs;**

(2) **PS_ALERT triggers CPU/DIMM throttling, and set Fans to the low power mode;**

(3) **PS_ALERT also triggers RMM to poll Power Shelf for the event information;**

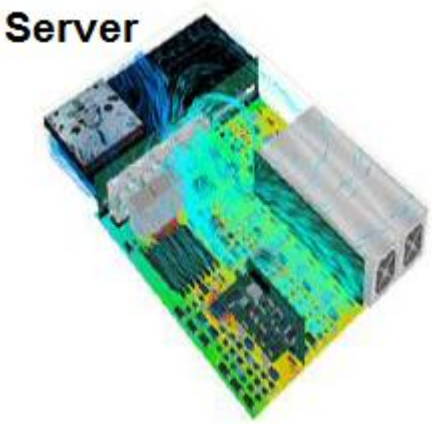
(4) **RMM informs the Server Nodes for the next actions;**

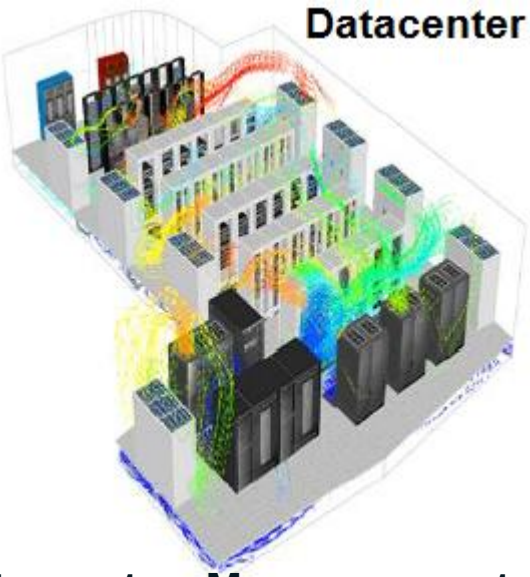
RMM to Server Nodes and Power Shelf communications are based on RSA manageability API

Agenda

- Rack Scale Management Architecture Overview
- Power Events Management in Rack Scale Architecture
- **Advanced Thermal Management in Rack Scale Architecture**
- Summary
- Q&A

PTAS (Power Thermal Aware Solution)





Server Platform Additional Telemetry

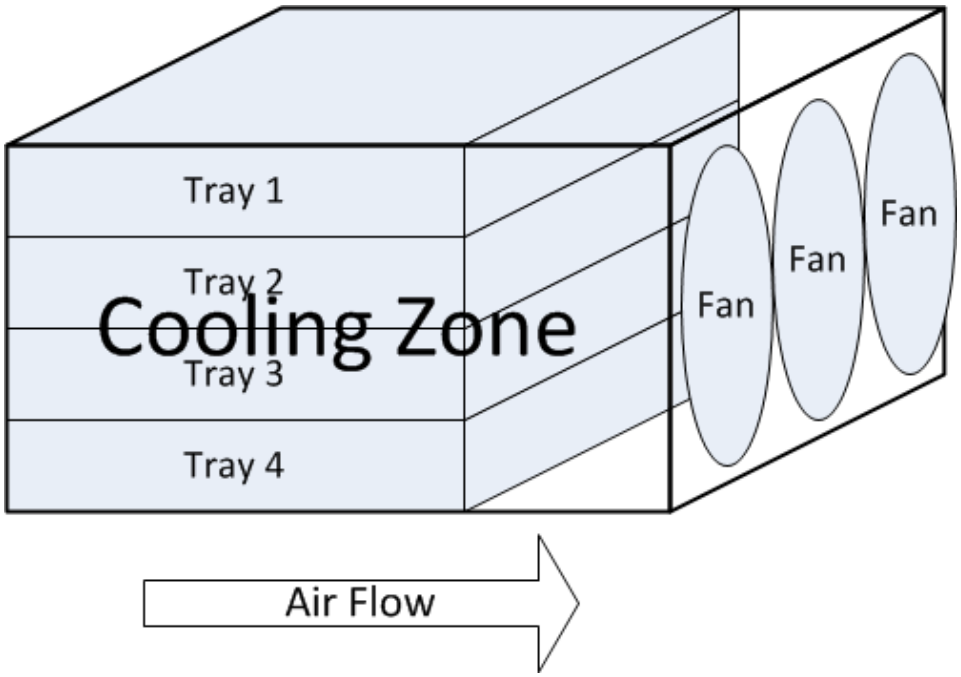
- **Volumetric Airflow**
- **Outlet Air Temperature**
- **CUPS (Compute Usage Per Second)**

Datacenter Management

- **Temperature, Airflow**
- **Power management policy**
- **Workload orchestration**

PTAS connects the server platform and datacenter for optimized thermal control

Rack Scale Level PTAS Concept



Basic Theory:

$Q = f(RPM)$ [1]

Where:

Q is the Volumetric Airflow of the cooling zone;
RPM is the speed (rotate per minute) of the fans.

$T_{outlet} = T_{inlet} + 1.76 * P * k_{alt} / Q$ [2]

Where:

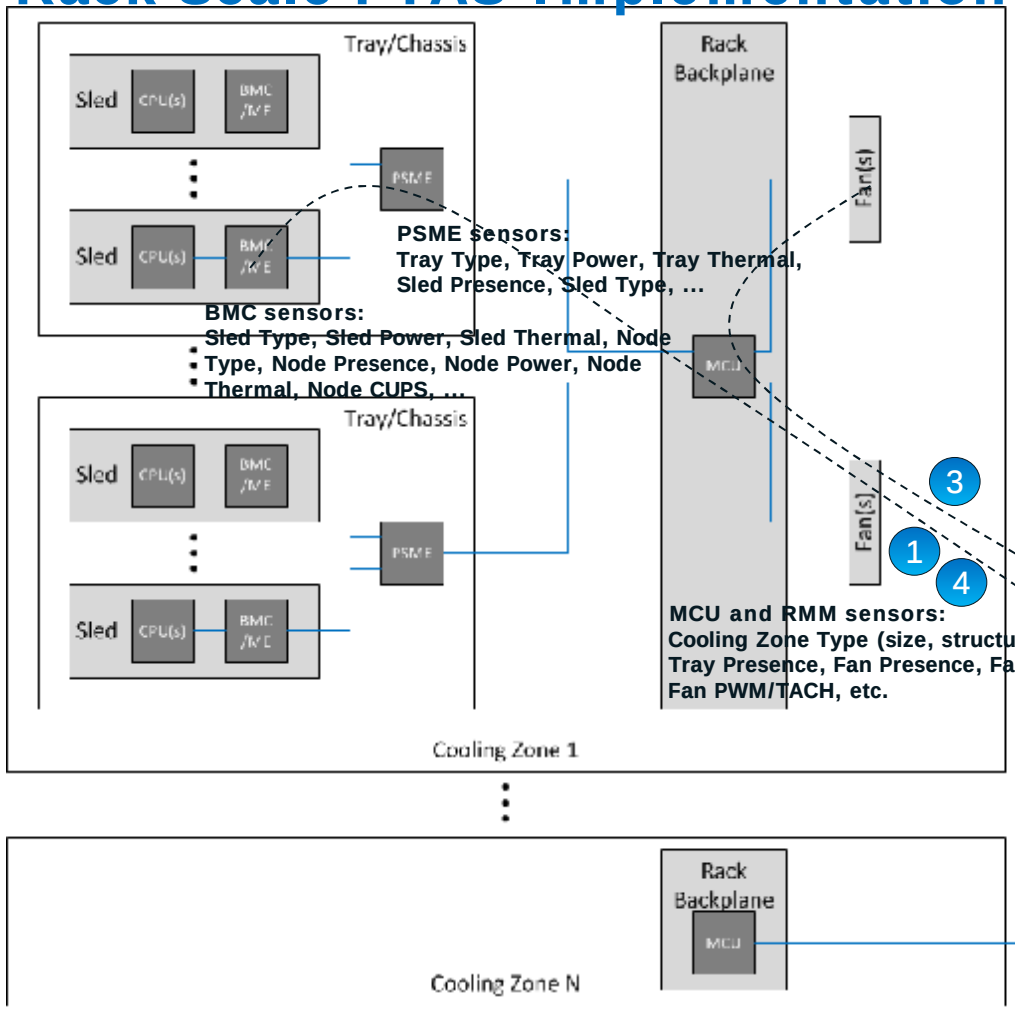
T_{outlet} is the outlet temperature of the cooling zone.
T_{inlet} is the inlet temperature of the cooling zone.
P is the total power dissipation of the cooling zone;
k_{alt} is the altitude correction factor.
Q is the Volumetric Airflow of the cooling zone;

Challenges at Rack Scale:

- **Function 1 is highly correlated to the cooling zone configuration, e.g. number and types of the trays.**
- **Equation 2 needs real time power data from all components within the cooling zone.**

Need to support the dynamical configuration of the cooling zone

Rack Scale PTAS Implementation using RSA Manageability API



(1) RMM gather the cooling zone configuration information from the BMC, PSME and Rack Backplane through the RSA manageability API. (example: presence and type of trays, presence and type of sleds, presence and type of fans, etc.)

(2) RMM choose the appropriate algorithm and coefficient sets based on the cooling zone configuration information.

(3) RMM polls the Fan RPM from the Rack Backplane through the RSA manageability API.

(4) RMM polls the Node/Sled/Tray Power from the BMC, PSME and Rack Backplane through the RSA manageability API.

(5) RMM calculate the Cooling Zone Level Volumetric Airflow and Cooling Zone Level Outlet Temperature, and expose these sensors through RSA manageability API.

RMM to Server Nodes and Tray communications are based on RSA manageability API.

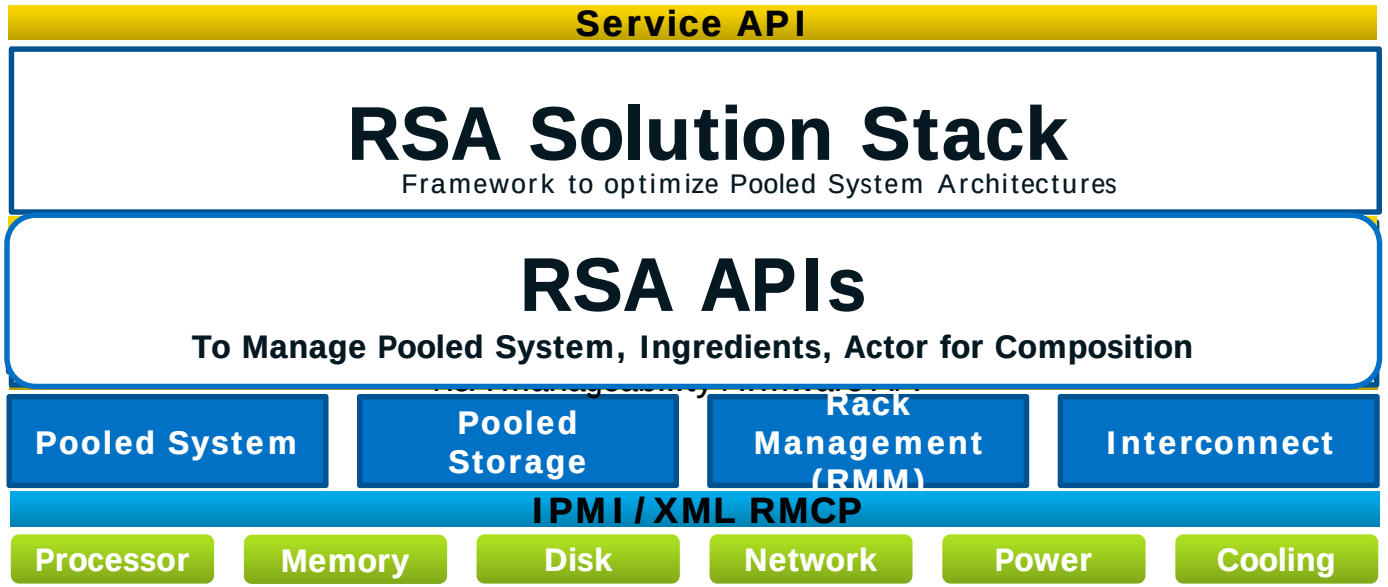
intel

Agenda

- Rack Scale Management Architecture Overview
- Power Events Management in Rack Scale Architecture
- Advanced Thermal Management in Rack Scale Architecture
- **Summary**
- Q&A

Rack Scale Software Architecture

Strawman with APIs

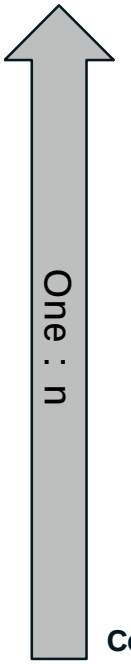


Agenda

- Rack Scale Management Architecture Overview
- Power Events Management in Rack Scale Architecture
- Advanced Thermal Management in Rack Scale Architecture
- Summary
- **Q&A**

Backup

RSA: Physical Manifestation



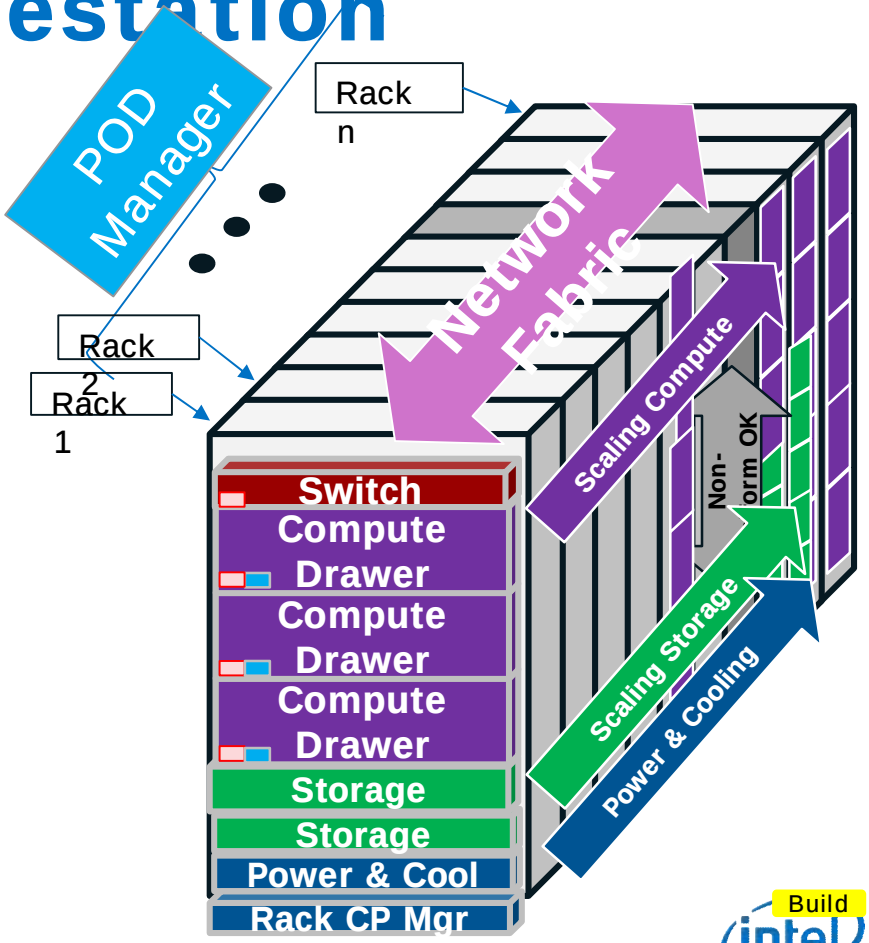
POD
- Collection of racks to be managed by a single POD Manager

Rack
- Standardized frame for deploying RSA drawers, and infrastructure (switch, cables, Power etc)

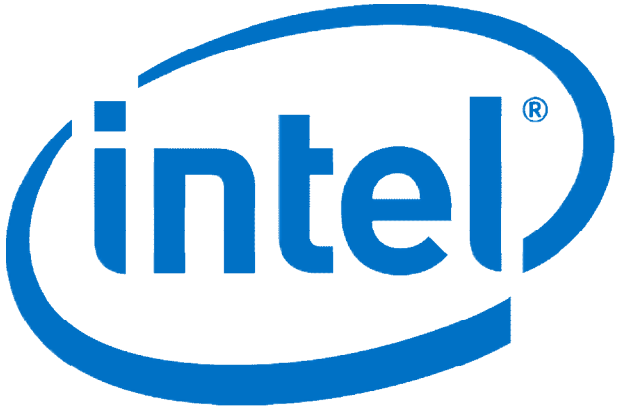
Drawer
- Sliding shelf that holds multiple modules
- Manageable Domain - a thermal or power zone
- Many drawers can fit into a rack

Module
- Field replaceable unit
- Can be pooled or shared
- Includes single or multiple elements of Compute, network, storage, memory, mgmt., power, cooling

Component
- Fundamental ingredients (CPU, DIMMs, cables, fans, boards etc)



Build intel


intel