



# 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

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.

UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

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.

All products, computer systems, dates, and figures specified are preliminary based on current expectations, and are subject to change without notice.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.  
Copyright © 2014, Intel Corporation. All rights reserved.

\*Other names and brands may be claimed as the property of others.

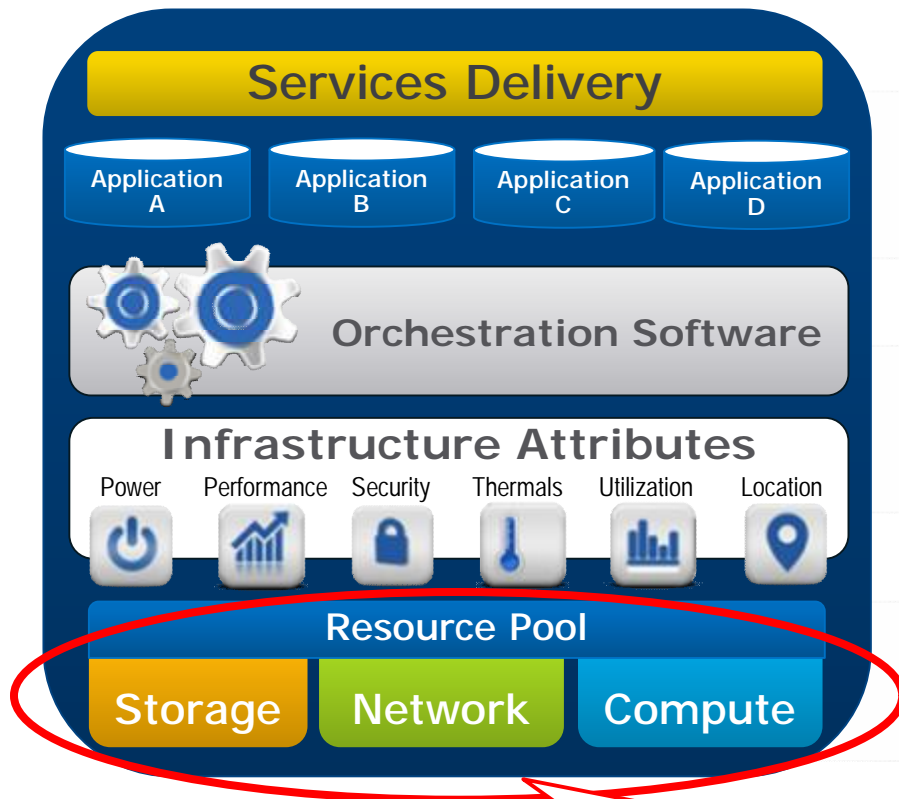


# Agenda

- **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

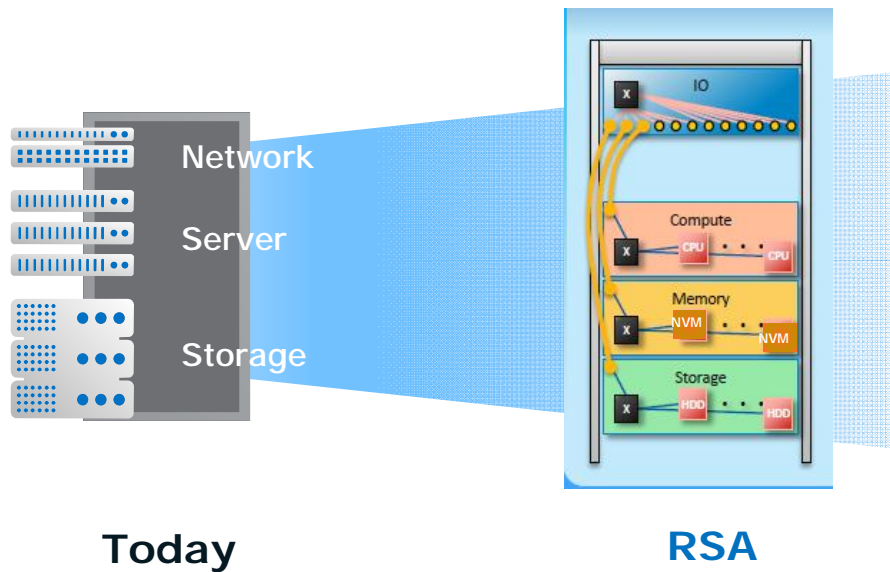
RSA optimizes systems to run your SDI solution



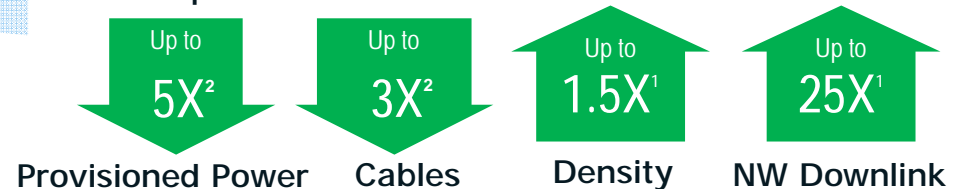
# Intel® Rack Scale Architecture – Optimized for SDI

Discrete Components,  
Self-Integration

Composable set of pooled  
and disaggregated resources



- 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

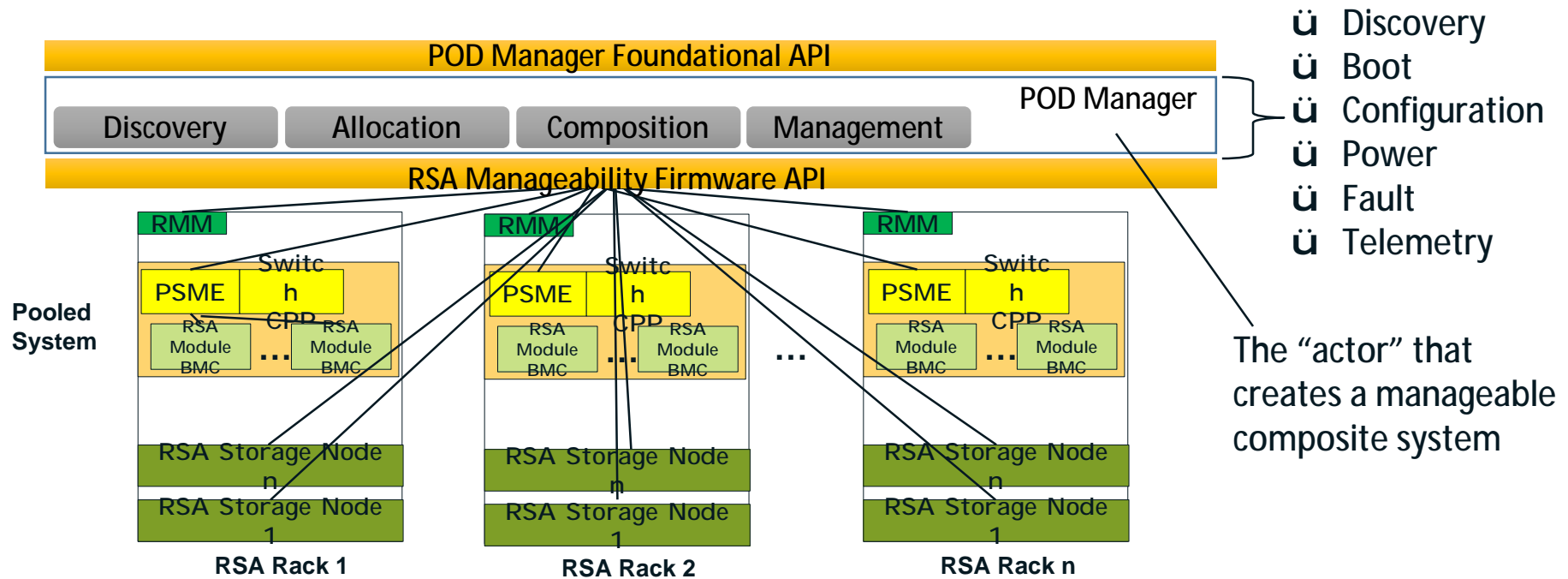


Flexibility, Capital Efficiency, Lower  
TCO



# RSA Management Architecture

## System Level Strawman



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



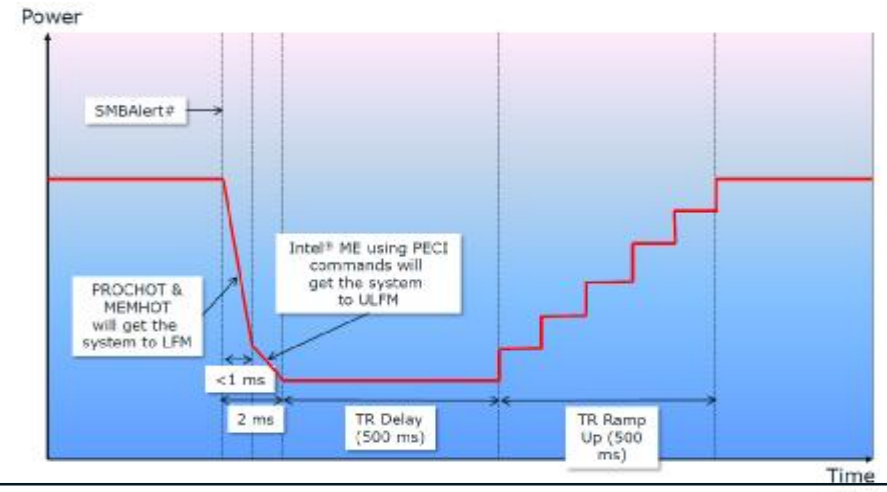
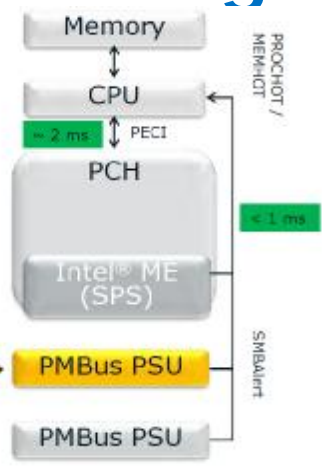


# Power Events Management: CLST and SMART

**CLST**  
 Closed Loop System Throttling  
 PSU Overcurrent protection



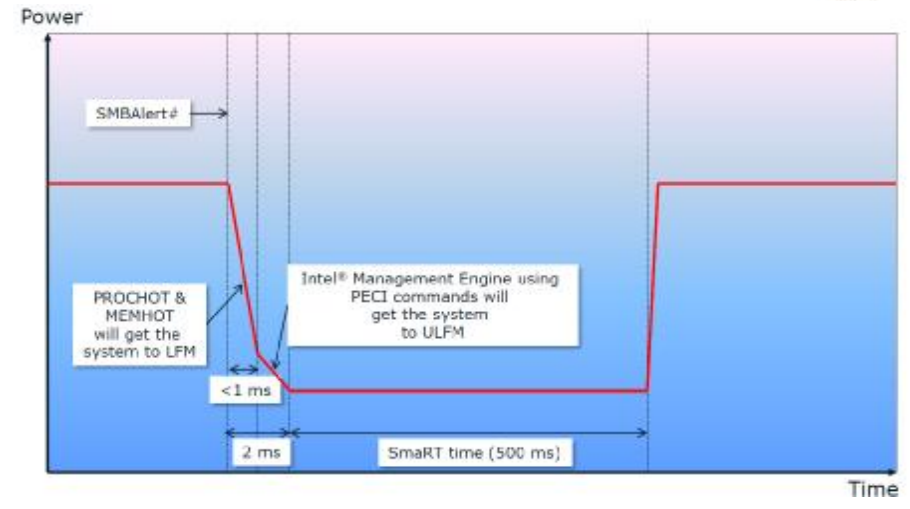
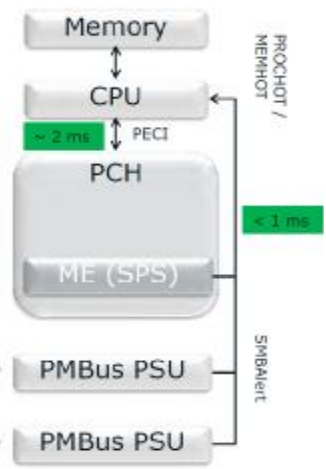
Over Current/  
Over Temp



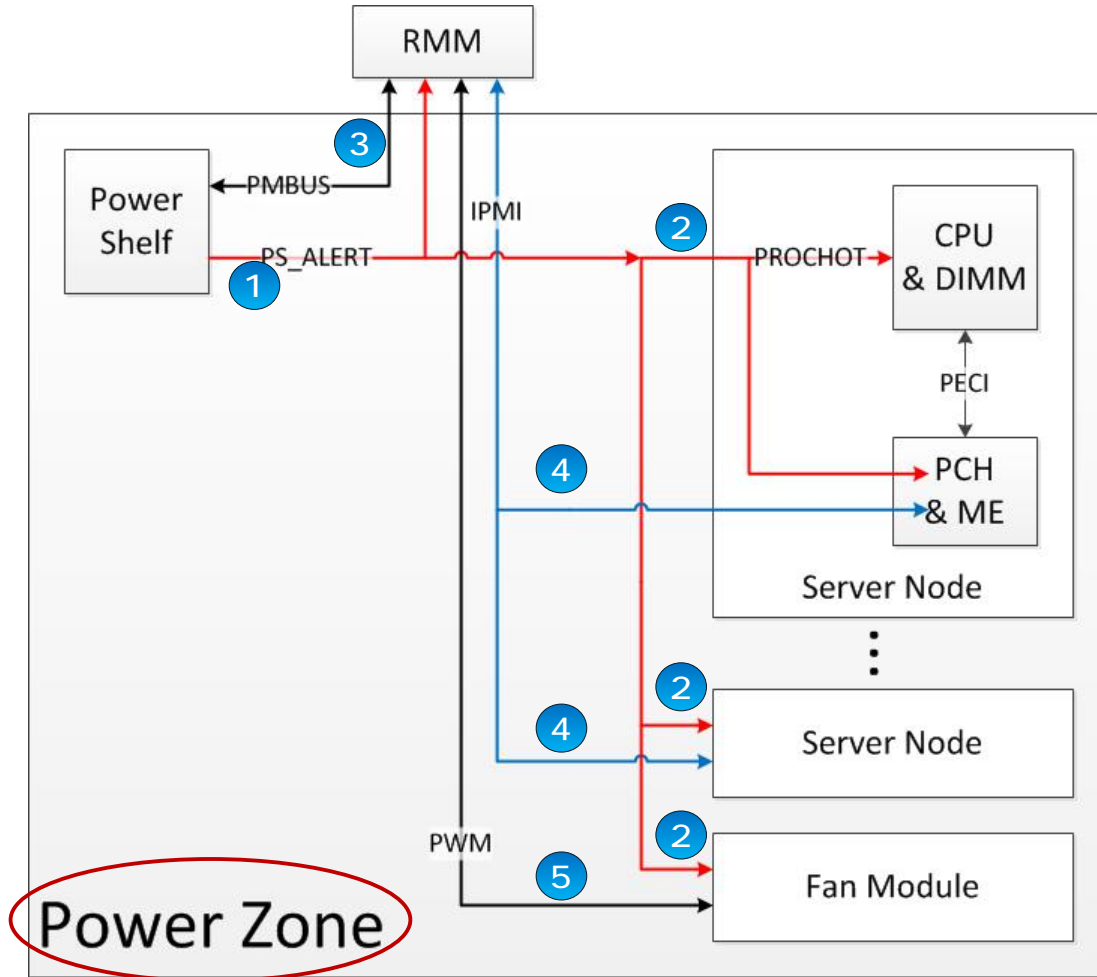
**SmaRT**  
 (Smart Ride Through)  
 Allows for AC loss for 1 cycle



AC Power  
AC Power



# 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 PSA manageability API

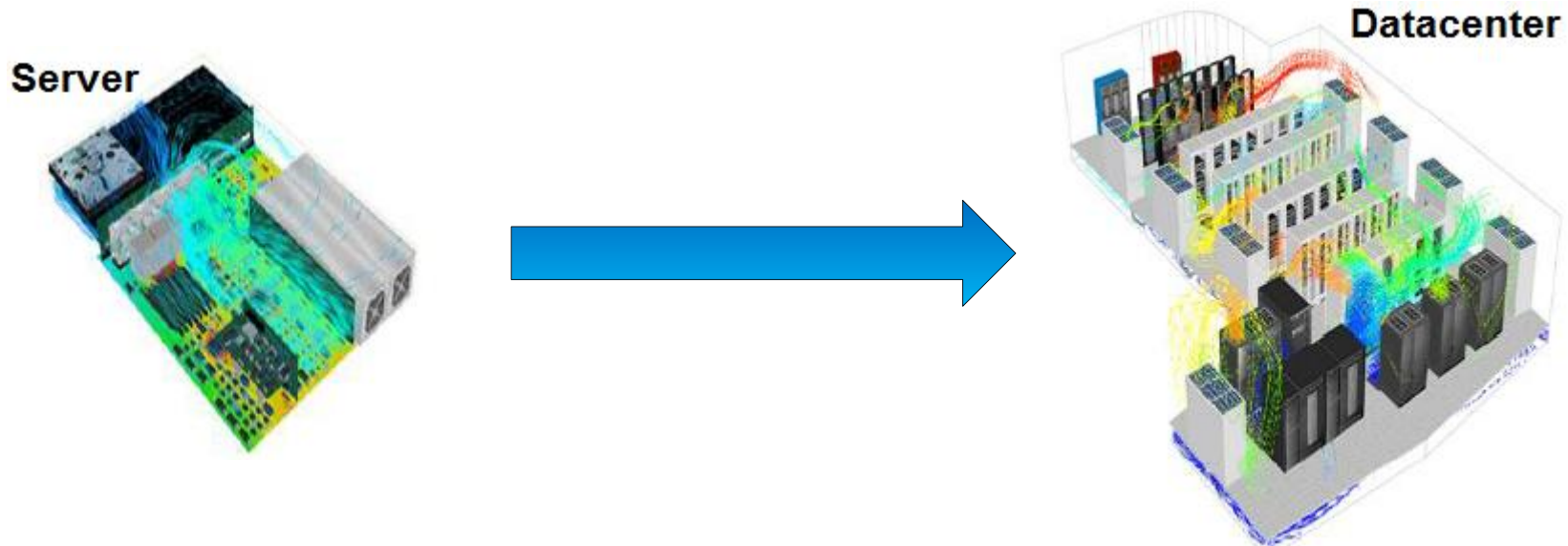


# Agenda

- 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)

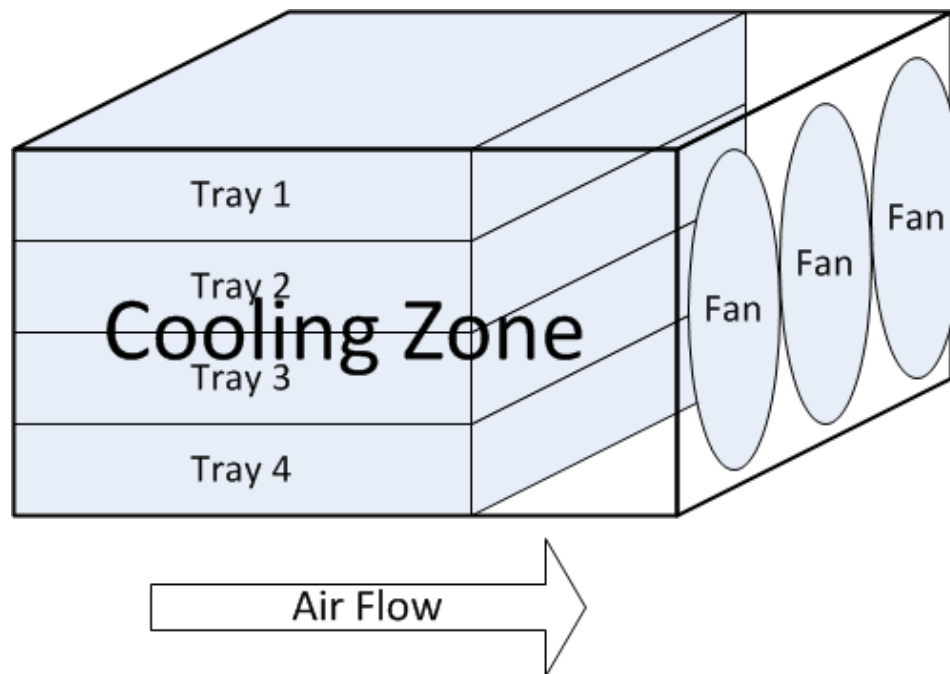
## Datacenter Management

- 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:

$$Q=f(\text{RPM}) \quad [1]$$

Where:

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

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

Where:

$T_{\text{outlet}}$  is the outlet temperature of the cooling zone.

$T_{\text{inlet}}$  is the inlet temperature of the cooling zone.

P is the total power dissipation of the cooling zone;

$k_{\text{alt}}$  is the altitude correction factor.

Q is the Volumetric Airflow of the cooling zone;

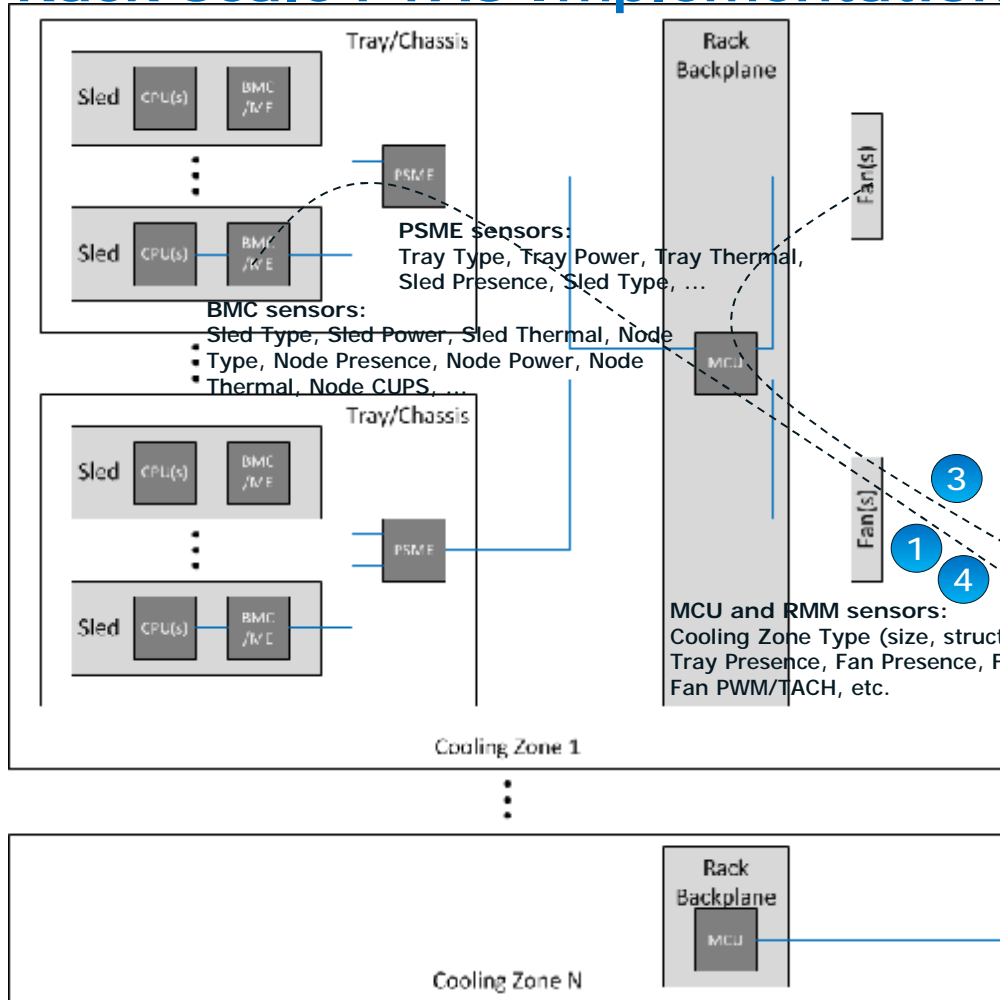
## Challenges at Rack Scale:

- 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.



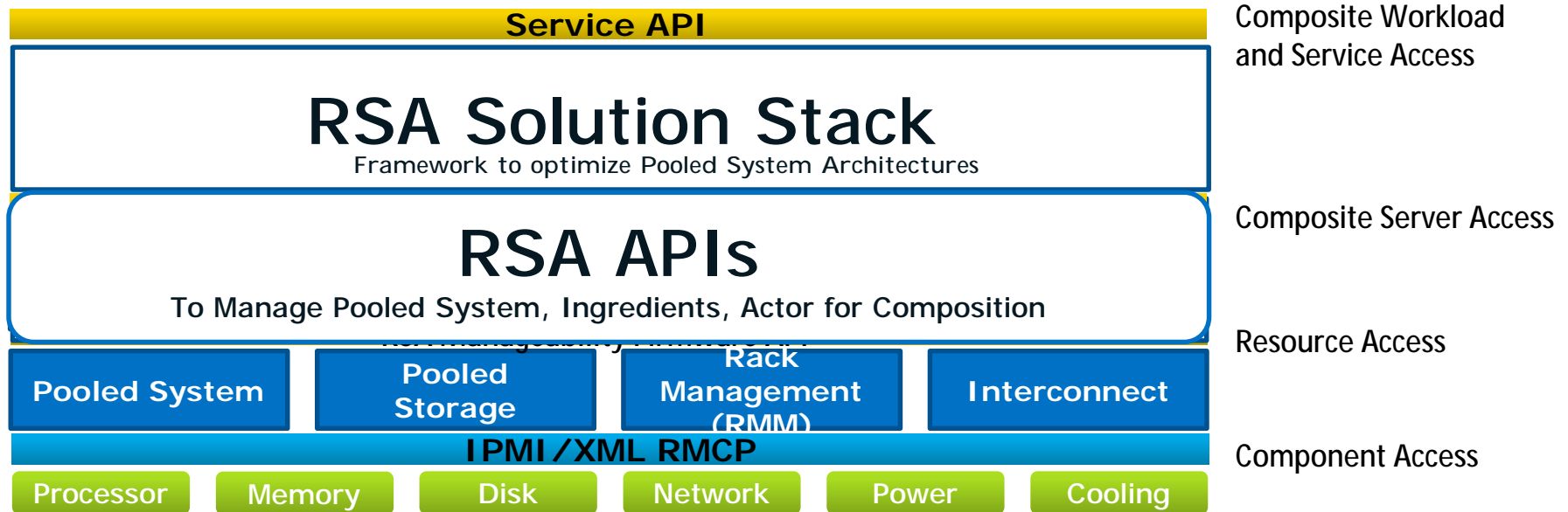
# Agenda

- 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

