Open. Together.

OCP
SUMMIT

Power Capping with Cloud Super Apps Performance SLA

[H/W Management]

Speakers:
Justin Song @ Alibaba
Edmund Song, Nishi Ahuja @Intel

Contributors:
Hao Zhu, Guan Wang, Xiaolin Meng @ Alibaba
Feng Jiang, Mohan J. Kumar, Xiaoguo Liang, John Leung @ Intel

OPEN PLATINUM™

Agenda

MANAGEMENT

- Alibaba Power Capping with Performance SLA
- Fine Granularity Power Management Knobs
- Redfish Adoption and Practices
- Call to Action

Case Studies

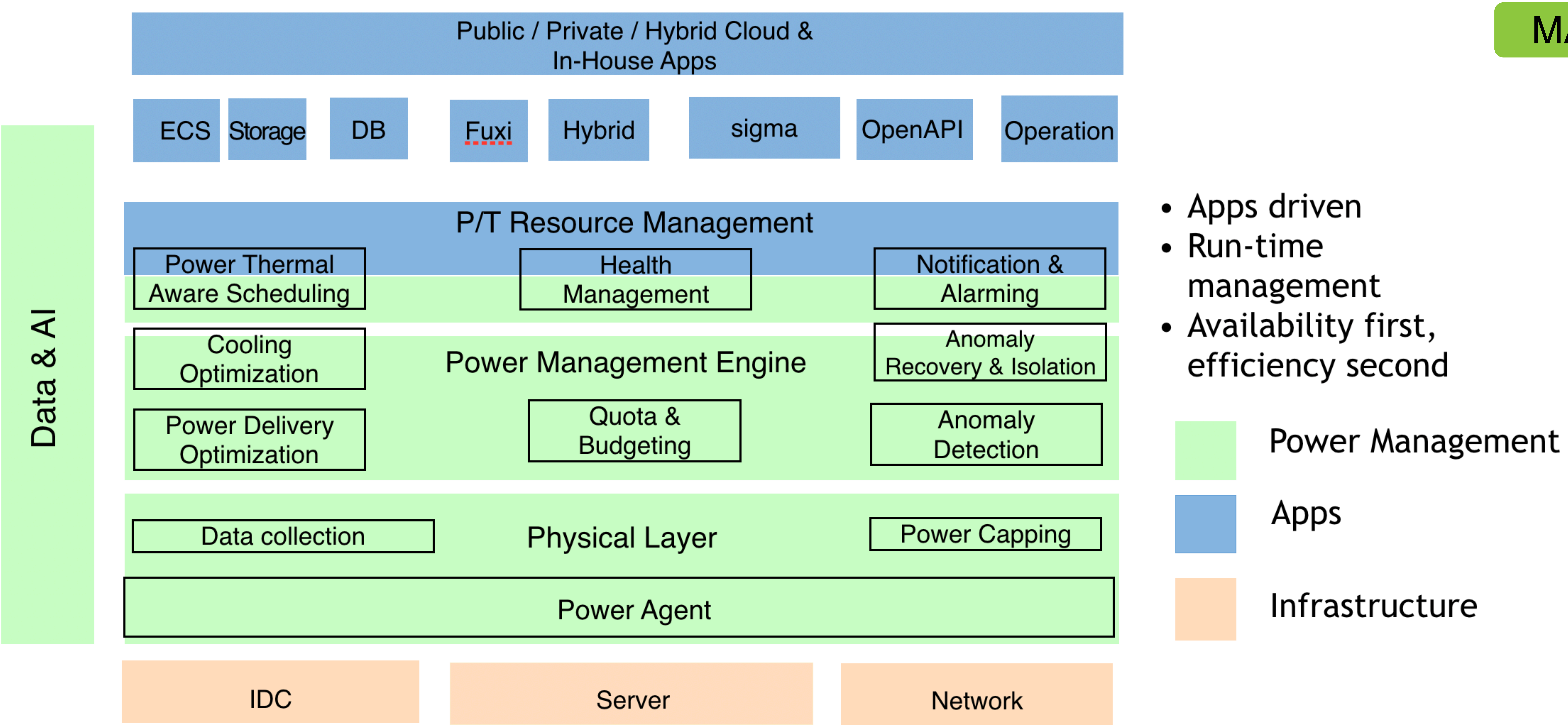
MANAGEMENT

Alibaba Power Capping with Performance SLA

Case Studies

Alibaba Power Management Architecture

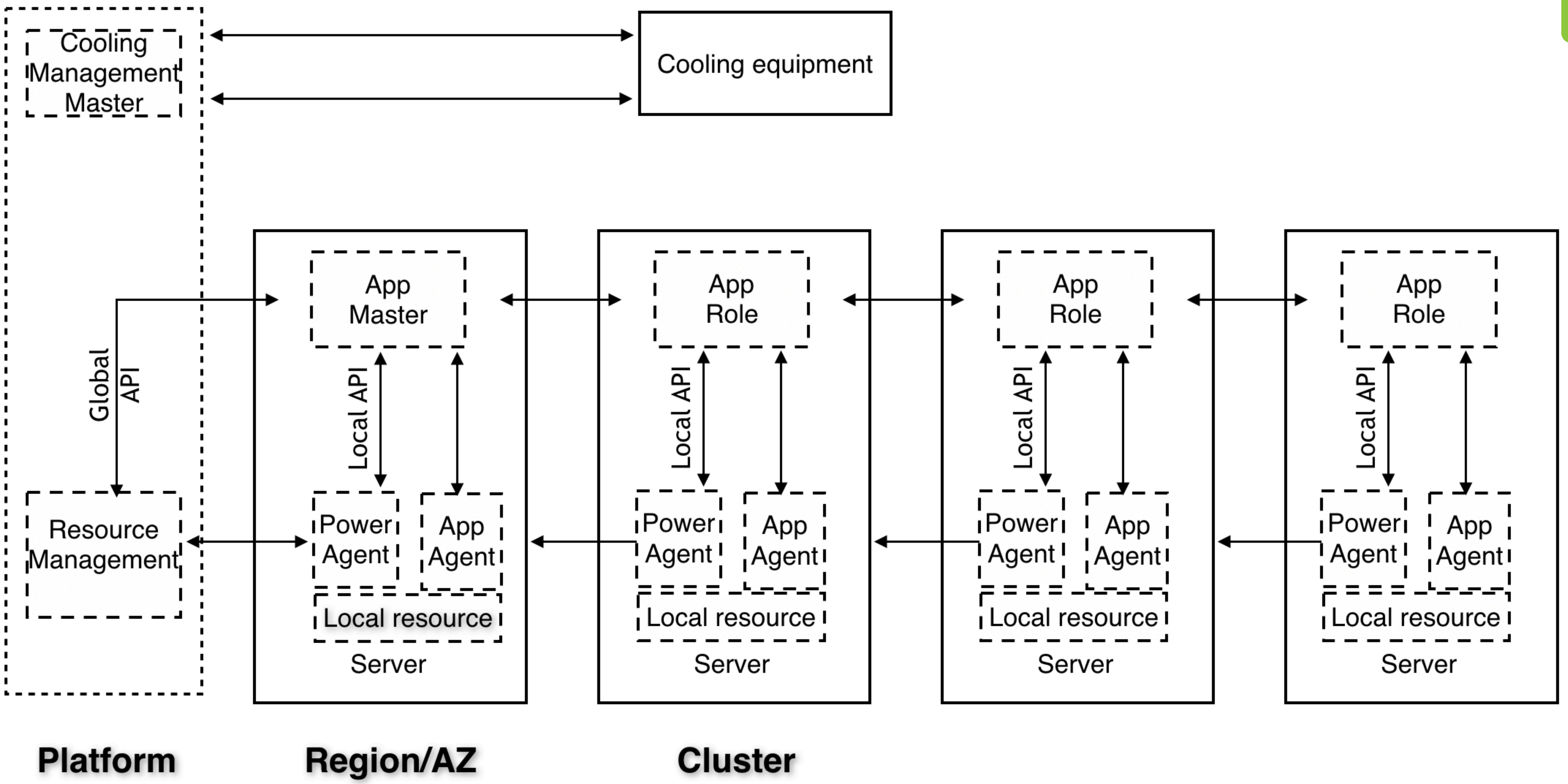
MANAGEMENT



- Apps driven
- Run-time management
- Availability first, efficiency second

Case Studies

Working with Apps

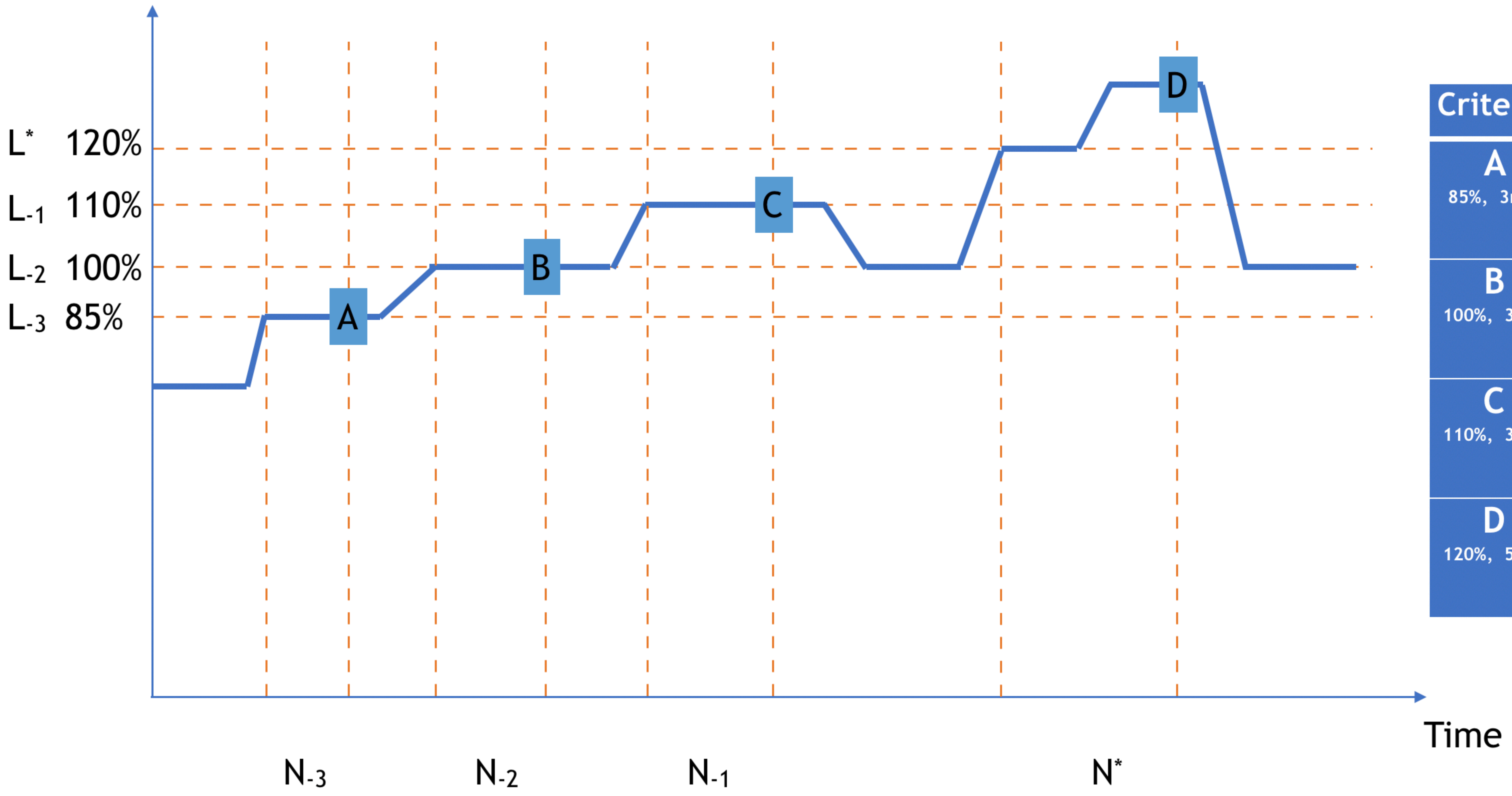


MANAGEMENT

Case Studies

Power Capping Triggering

MANAGEMENT



Criteria	Operation
A 85%, 3min	Rebalance
B 100%, 3min	Stop adding instances
C 110%, 3min	Super apps flow control
D 120%, 5min	Power capping with performance SLA

Case Studies

Performance SLA

MANAGEMENT

Type	Target	Note
Availability		Align w/ apps
Service delay	1s	Local
	30s	Global
Models coverage	Based on spec & test results	
Racks coverage	Based on spec & test results	
Power watermarks	Defined by apps & platform	
Capping accuracy	5%	
Priority	Defined by apps	Low priority nodes first capped
Fmin	Defined by apps Lifted by AI	Anytime higher than Fmin
Granularity	By core (CPU), rank (mem), link (IO) and device (storage)	
Capping - DVFS	Minimal performance impact	Defined by apps
Capping - CCx	**Minimal latency impact**	Defined by apps
In-Band	supported	
Out-of-Band	Partially supported	
Thermal watermarks	Defined by apps and platform	
Failover	Unconditional capping, autonomous capping, or S5	

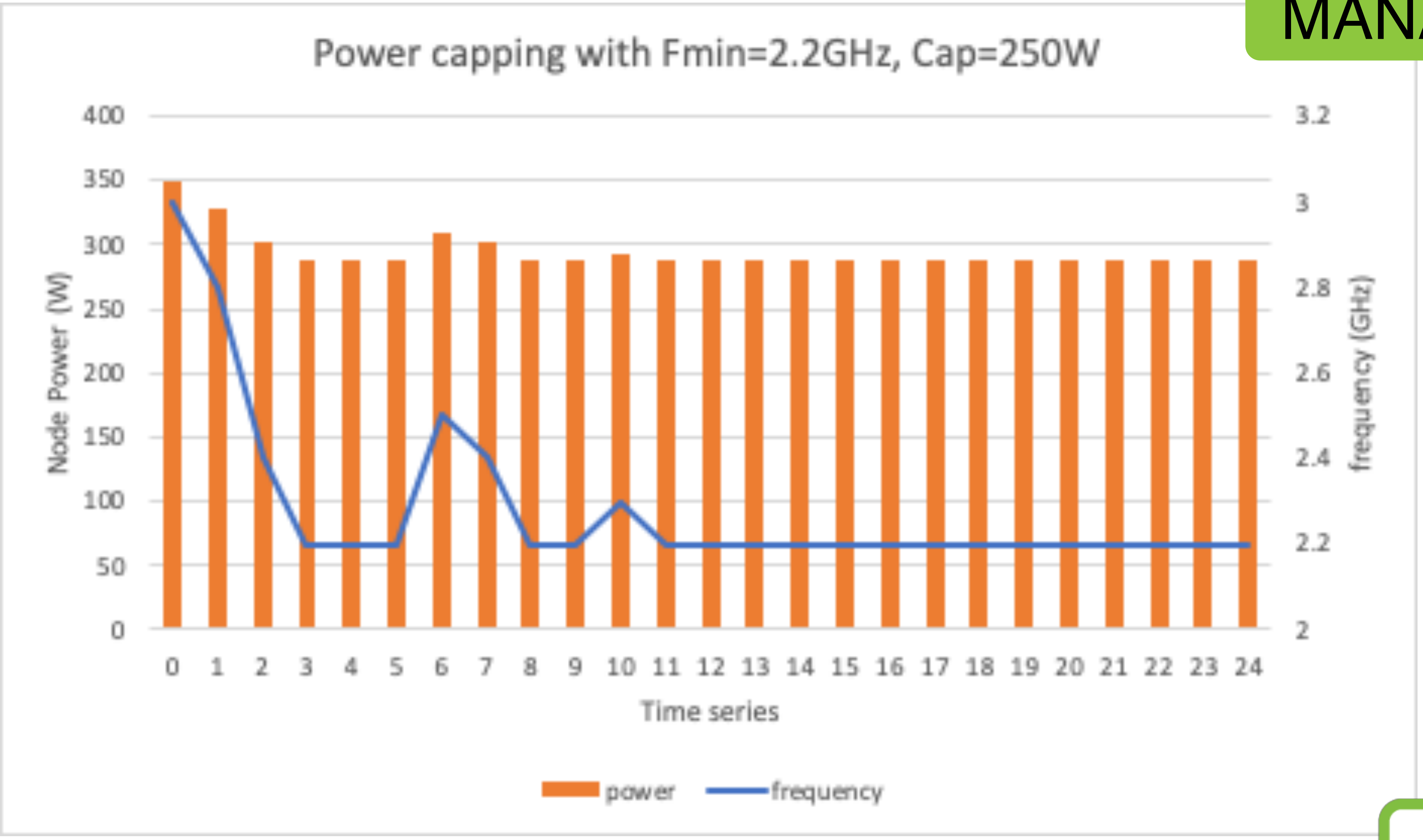
Preventing performance downgrade

DVFS: Dynamic Voltage Frequency Scaling
CCx: Core C-States

Case Studies

Results — Examples

MANAGEMENT



Scenario: high priority instances performance guaranteed with low priority instances capped

Scenario: capping with no Fmin violation

Case Studies

Fine Granularity Power Management Knobs

Case Studies

Cloud Power-Performance Requirement

MANAGEMENT

- Rack density
- Utilization of provisioned resource

Capex Optimization

- Power-Performance proportionality(e.g. SLA matched energy efficiency)
- Performance-per-watt efficiency

Opex Optimization

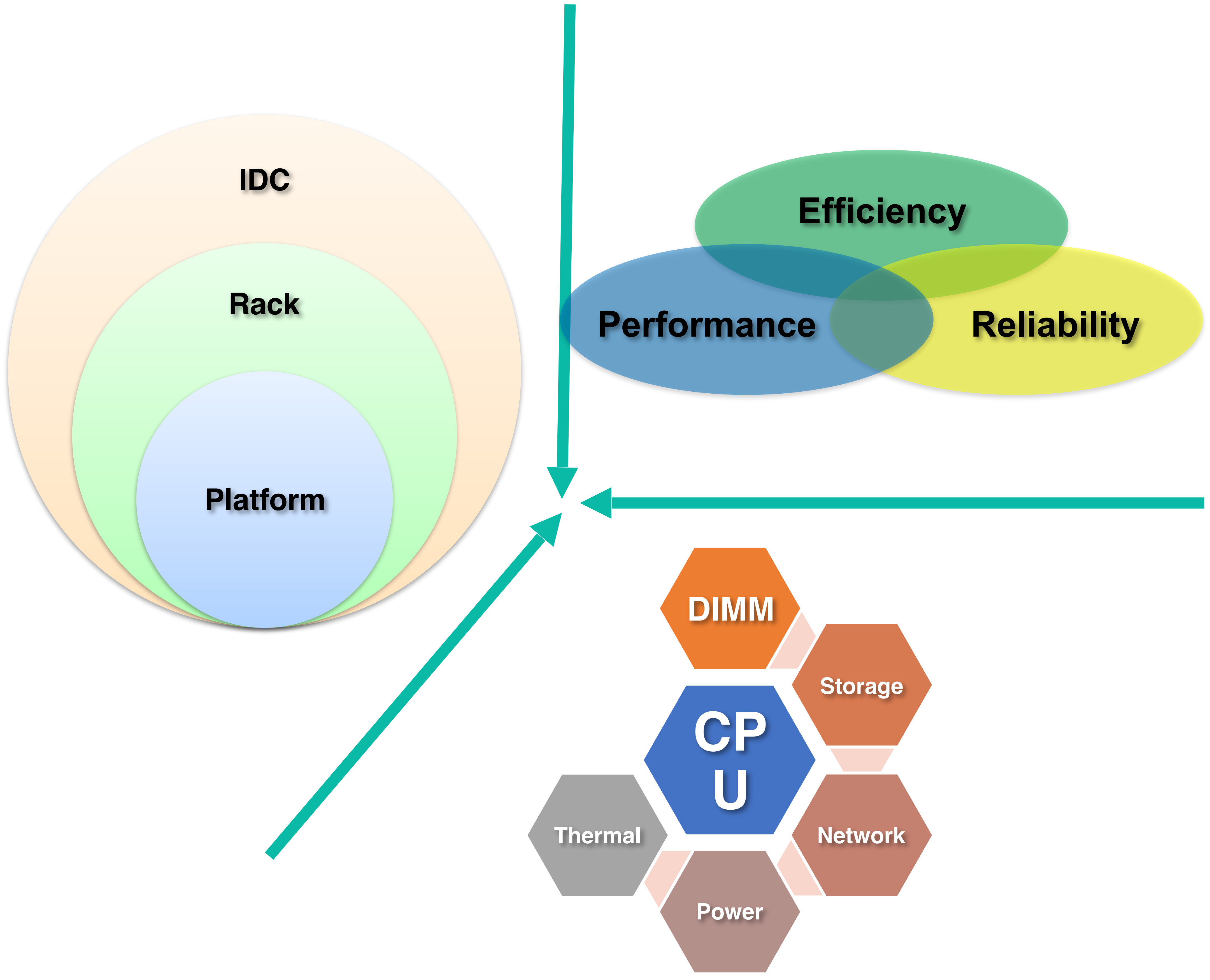
- Convergence between infrastructure plane and resource plane
- Hardware intelligences into resource scheduling and orchestration

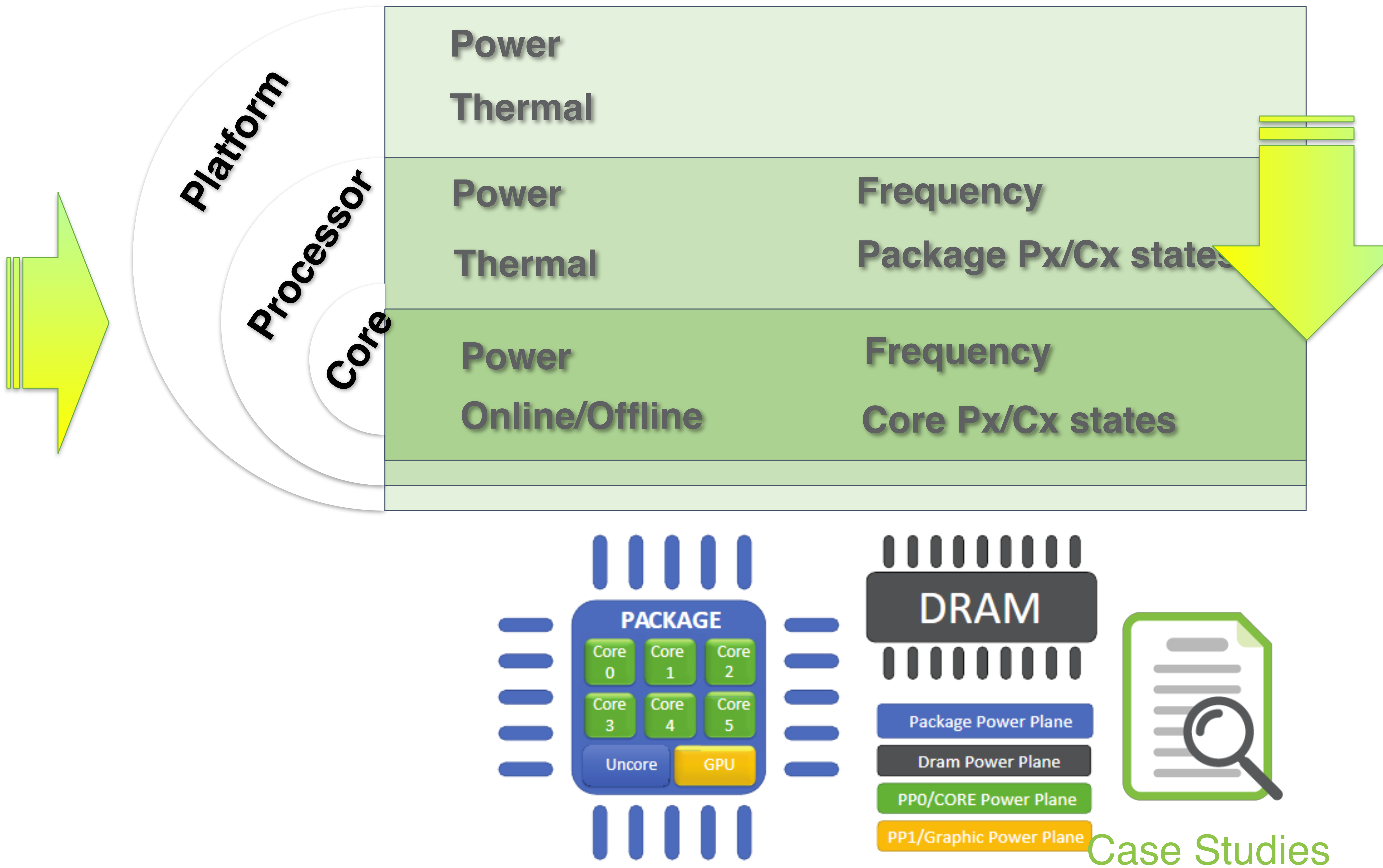
SLA and Reliability

Case Studies

Fine Grain Platform Power-Performance Knobs

MANAGEMENT





Case Studies

Intel Practices in Cloud Power-Performance Optimization

MANAGEMENT

Workload

- Group Power Capping
- Dynamic Core Mgmt.
- Peak Shaving via Smart BBU(Turbo Rack)
- HW Telemetry Awaress Scheduling
- Workload PnP Analytic and Simulation

Intel Cloud Power and Performance Solution

Cloud OS

- K8S
- Open Stack
- OpenDCM
- Proprietary Cloud OS

Cloud Power-Performance Telemetry and Mgmt. API (Redfish Based)

CPU
- RAPL/DVFS
- Px/Cx
- Core Freq.
- Core Off/On
- Un-core Freq.

DIMM
- RAPL

NIC, Storage…

Platform
- NM
- Open BMC

Rack & DC Facility
- Open RMC
- DC Power
- Smart BBU
- DC Cooling

Case Studies

MANAGEMENT

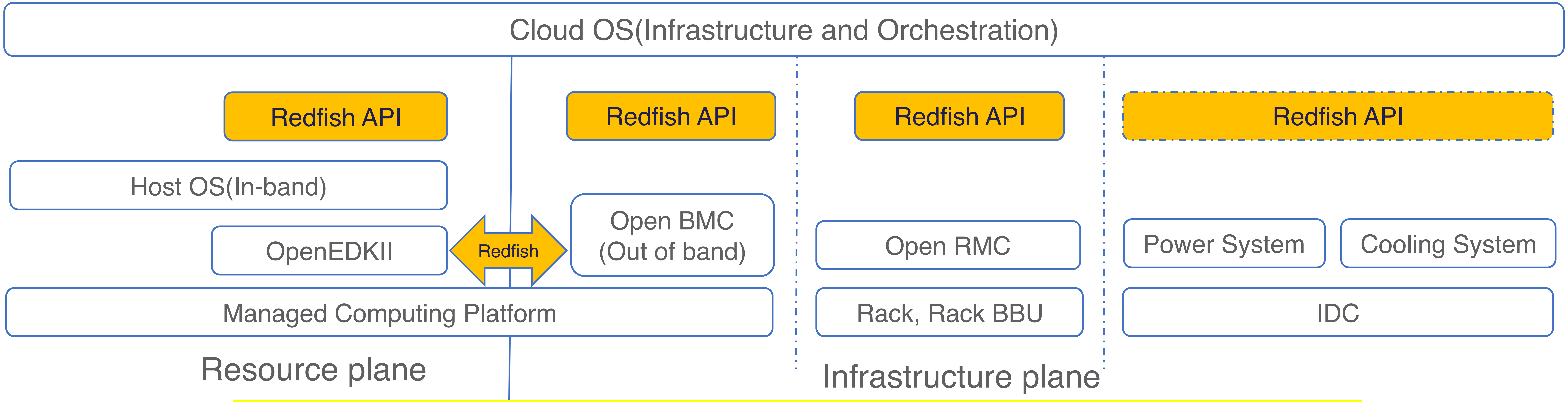
Redfish Adoption and Practices

Case Studies

API Requirement in Cloud Power-Performance Optimization

MANAGEMENT

- Interoperability - Infrastructure plane vs. Resource plane, server vs. facility (rack, IDC) etc.
- Consistent API model for In band interface and Out of band interface
- Support runtime configuration and cloud scale deployment



Starting with extensible and scalable Redfish model

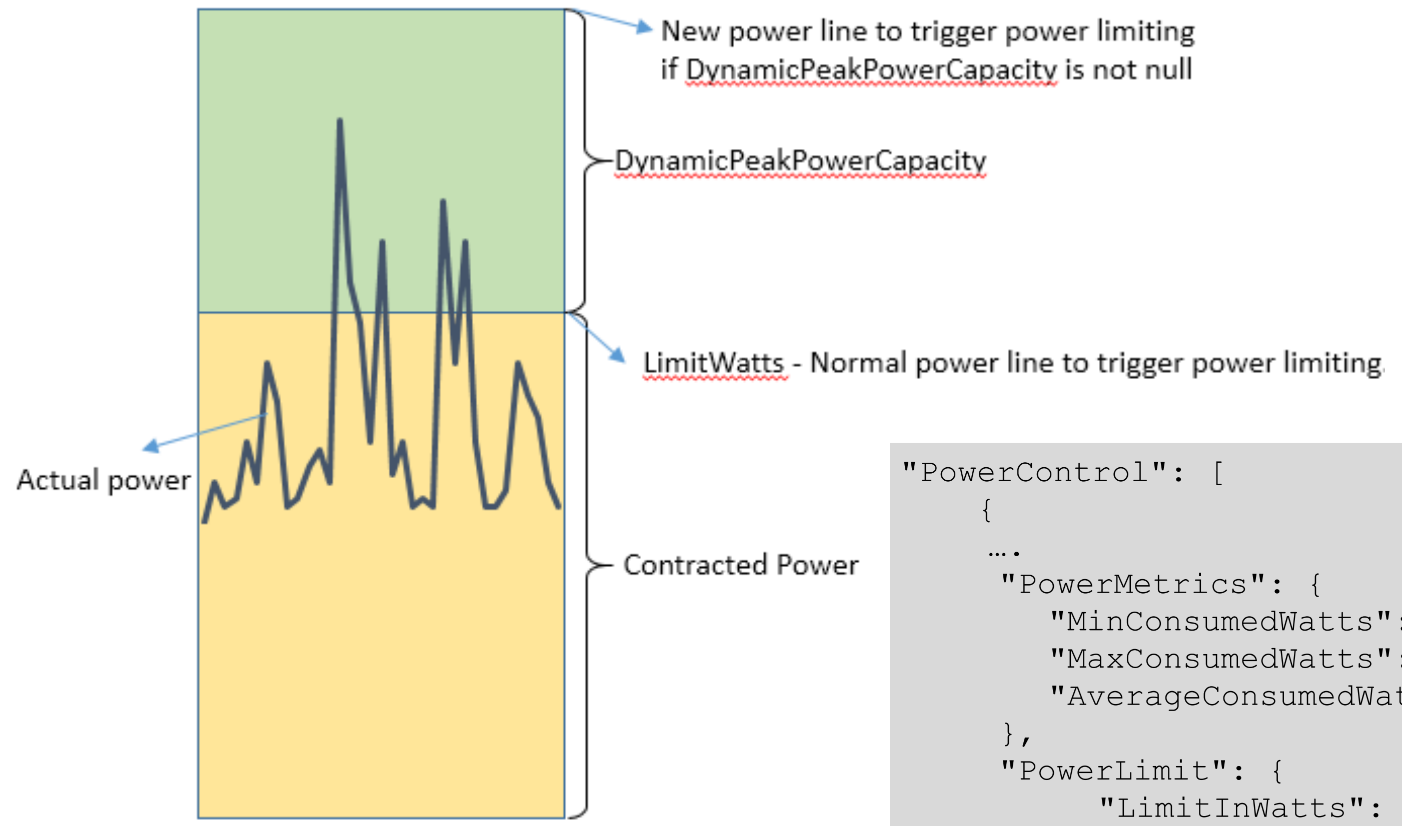
Case Studies

Redfish Practices

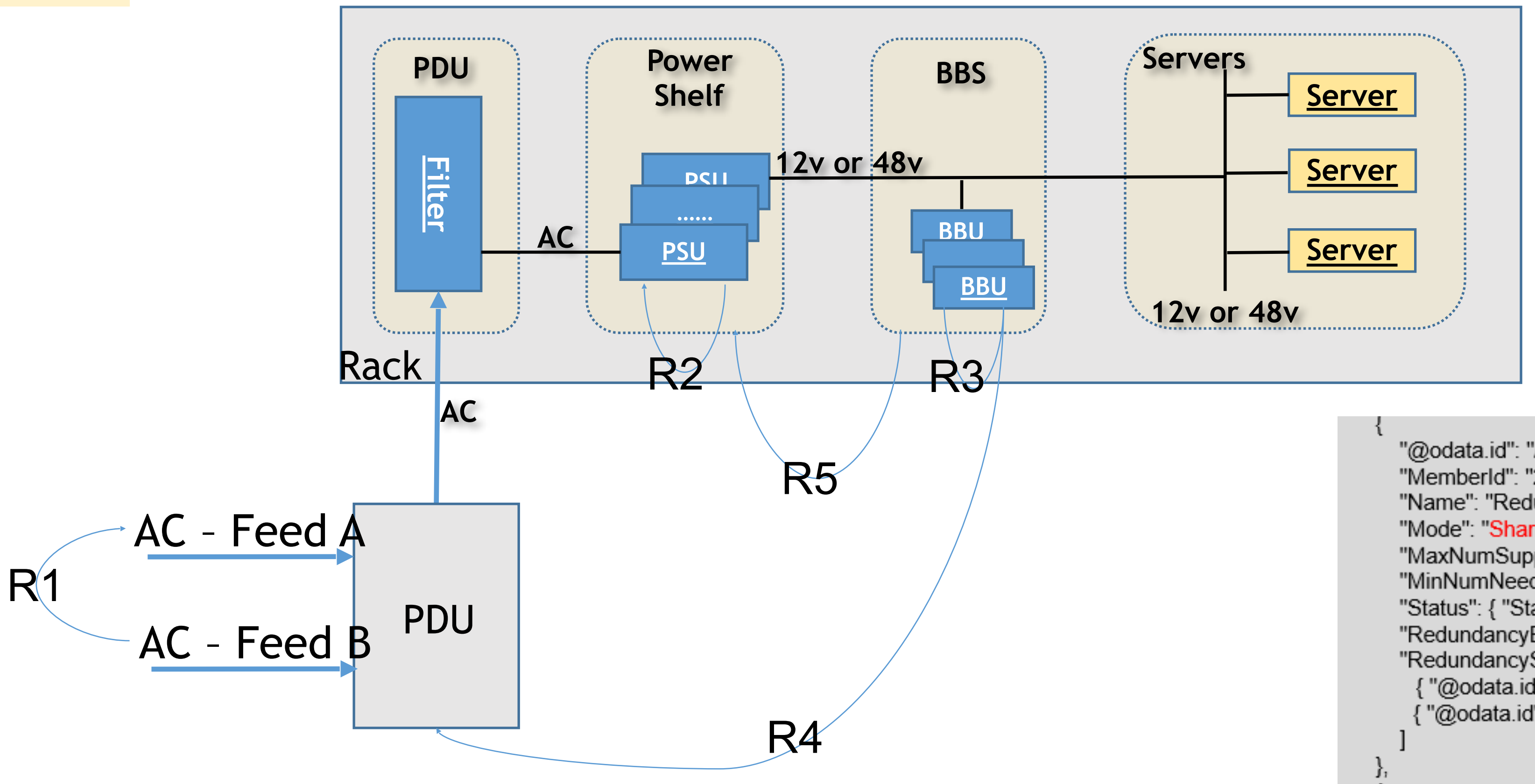
- Unified power control API model to support hierarchical power capping (platform, rack, cluster)
- Consistent model to support RAPL and Smart BBS based peak shaving.

- Orchestration need insight of power redundancy of managed system for intelligent scheduling policy
- Describe complex power redundancy model which is across platform, chassis and rack

MANAGEMENT




```
"PowerControl": [
    {
    ….
     "PowerMetrics": {
        "MinConsumedWatts": 7200,
        "MaxConsumedWatts": 7900,
        "AverageConsumedWatts": 7600
     },
     "PowerLimit": {
           "LimitInWatts": 7900,
           "DynamicPeakCapacityWatts": 1000,
           "DynamicPeakCapacityDuration": "P3S",
           "LimitException": "LogEventOnly",
           "CorrectionInMs": 50
     },
     "Status": {
           "State": "Enabled",
           "Health": "OK"
     },
   }
],
```




```
"Redundancy": [
   {
     "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/Redundancy/0",
     "MemberId": "0",
     "Name": "Rack Power Sources Redundancy Policy",
     "Mode": "Failover",
     "MaxNumSupported": 2,
     "MinNumNeeded": 1,
     "Status": { "State": "Enabled", "Health": "OK" }
     "RedundancyEnabled": true,
     "RedundancySet": [
       { "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/PowerSources/0" },
       { "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/PowerSources/1" }
     ]
   },
   {
     "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/Redundancy/1",
     "MemberId": "1",
     "Name": "Rack Power Supply Units Redundancy Policy",
     "Mode": "Sharing",
     "MaxNumSupported": 2,
     "MinNumNeeded": 1,
     "Status": { "State": "Enabled", "Health": "OK" }
     "RedundancyEnabled": true,
     "RedundancySet": [
       { "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/PowerSupplies/0" }
```

R1: PSrc to PSrc

R2: PSU to PSU

```
   {
     "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/Redundancy/2",
     "MemberId": "2",
     "Name": "Redundancy policy for BBUs",
     "Mode": "Sharing",
     "MaxNumSupported": 3,
     "MinNumNeeded": 1,
     "Status": { "State": "Enabled", "Health": "OK" }
     "RedundancyEnabled": true,
     "RedundancySet": [
       { "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/PowerSupplies/2" },
       { "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/PowerSupplies/3" },
     ]
   },
   {
     "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/Redundancy/3",
     "MemberId": "3",
     "Name": "Redundancy policy for BBU and Power Source",
     "Mode": "Failover",
     "MaxNumSupported": 2,
     "MinNumNeeded": 1,
     "Status": { "State": "Enabled", "Health": "OK" }
     "RedundancyEnabled": true,
     "RedundancySet": [
       { "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/Redundancy/0" },
       { "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/Redundancy/2" }
     ]
   },
   {
     "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/Redundancy/4",
     "MemberId": "4",
     "Name": "Redundancy policy for BBUs and PSUs",
     "Mode": "Sharing",
     "MaxNumSupported": 2,
     "MinNumNeeded": 1,
     "Status": { "State": "Enabled", "Health": "OK" }
     "RedundancyEnabled": true,
     "RedundancySet": [
       { "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/Redundancy/1" },
       { "@odata.id": "/redfish/v1/Chassis/Rack1/Power#/Redundancy/2" }
```

R3: BBU to BBU

R4: BBUs to PSrcs

R5: BBUs to PSUs

Case Studies

Redfish based API reduce deployment complexity of data center power management

MANAGEMENT

Call to Action

Case Studies

Call to Action

MANAGEMENT

- Performance SLA driven power optimization is critical for TCO optimization and PUE efficiency. Need platform and solution co-innovation to support dynamic, flexible and workload aware optimization policies.

- Open and standard API able to reduce deployment cost in large scale cloud environment. Need collaboration to define common telemetries and control interface for OCP platform, e.g. via baseline OCP HW Mgmt. profile.

- Cloud developers and users need to understand and define their performance requirements for their cloud apps

- Get involved: https://www.opencompute.org/projects/hardware-management

Case Studies

Open. Together.

OCP Global Summit | March 14–15, 2019

OCP SUMMIT