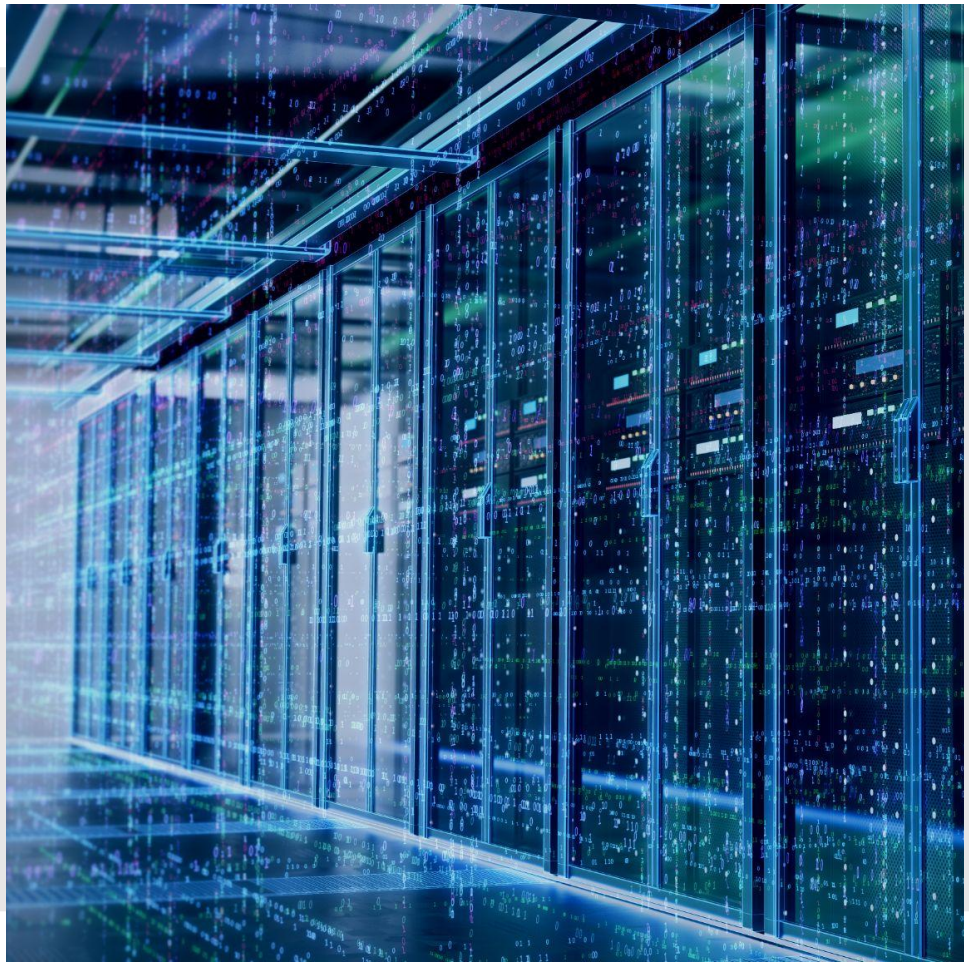
RISC-V Server SoC Specification

Ved Shanbhogue, Rivos Inc.

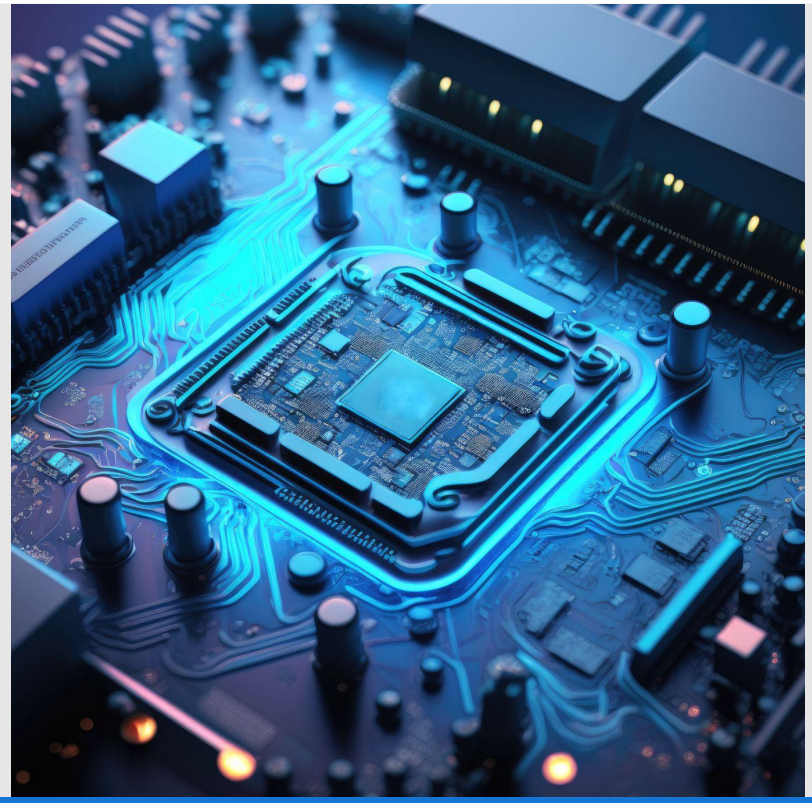
Server

- A computing system designed to manage and distribute resources, services, and data to other computers or devices on a network
- Serves or provides information and resources upon request
- Operate continually and have higher requirements for capabilities such as RAS, security, performance, and quality of service



RISC-V Server SoC Specification Goals

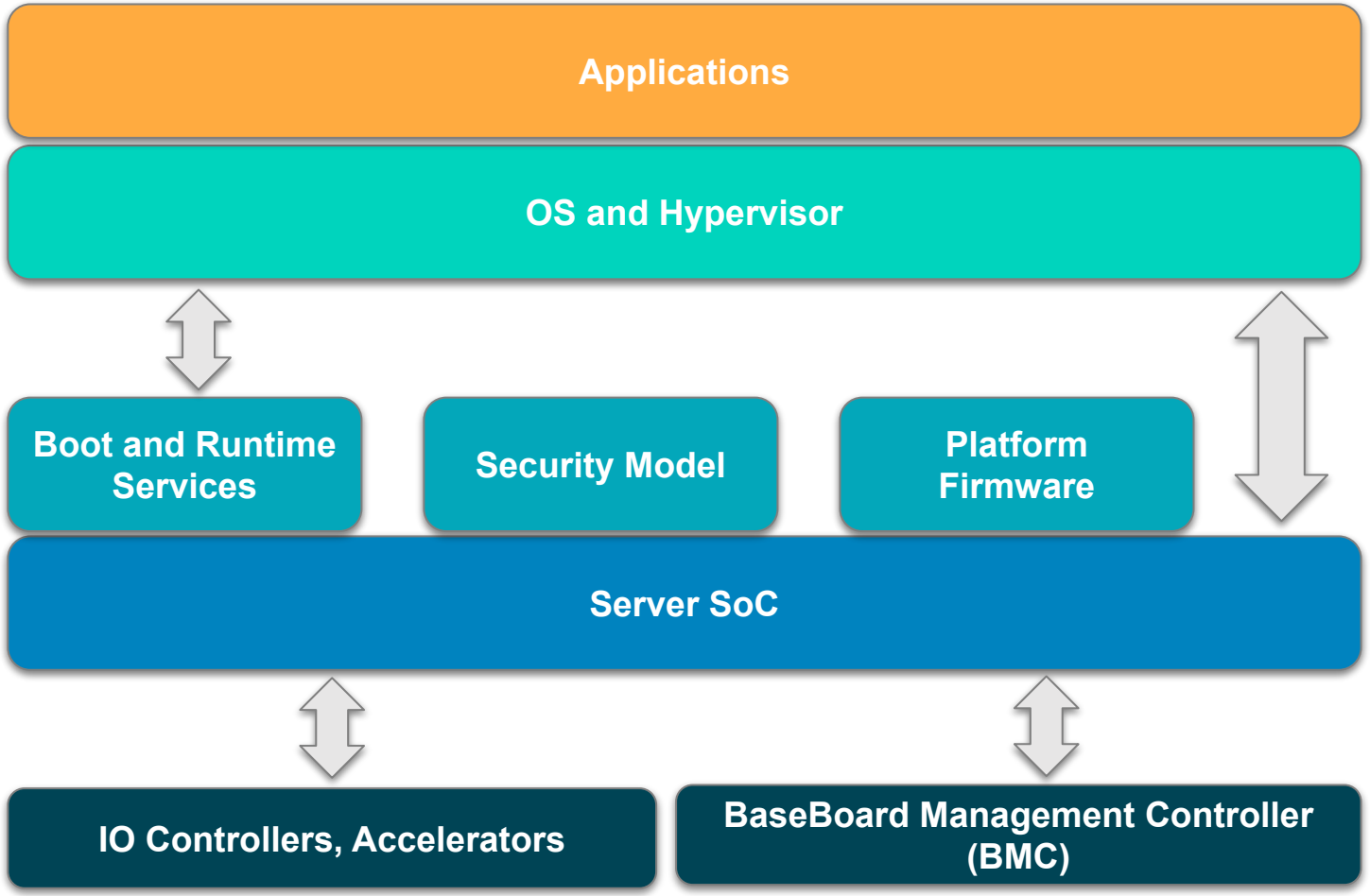
- Standardizes the requirements for the hardware interfaces and capabilities provided by the SoC to software executing on the application processor
- A standard set of capabilities, encompassing areas where divergence is typically unnecessary and where novelty is absent across implementations



Enables OS and hypervisor vendors to support such SoCs with a single binary OS image distribution model

RISC-V Server Platform

- Standardized SoC hardware interfaces such as PCIe root ports, IOMMU, and Interrupt Controllers
- Boot and Runtime services using UEFI and ACPI
- BMC for provisioning and management using standards such as MCTP, PLDM, IPMI, and Redfish
- Security Model guides debug authorization, secure boot, firmware updates, firmware resilience, and other use cases

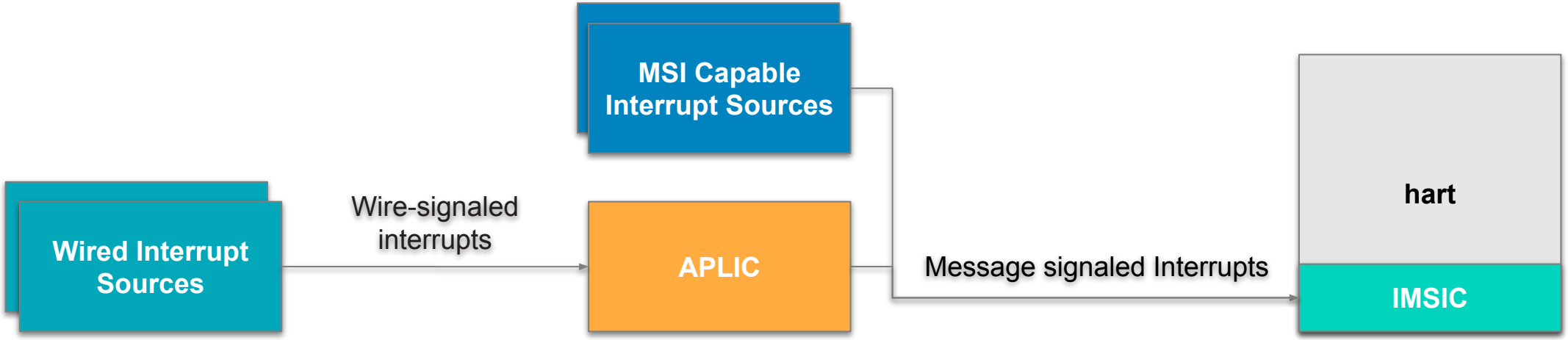


Outline of RISC-V Server SoC specification

- Clocks, Timers, and Interrupt Controllers
- IOMMU
- PCIe Subsystem
  - ECAM and PCIe memory space
  - Access Control Services
  - Handling of ID and address routed transactions
  - Message Signaled Interrupts
  - Precision Time Management
- Reliability, Availability, and Serviceability
- Quality of Service
- Manageability
- Performance Monitoring
- Security

Clocks, Timers, and Interrupt Controllers

Standardize the unit of time as 1ns
Requires RISC-V Advanced Interrupt Architecture (AIA)
Message signaled interrupt to the hart
Specifies the minimum number of VS-mode interrupt files that must be supported
Specifies the minimum number of interrupt identities that must be supported



RISC-V IOMMU

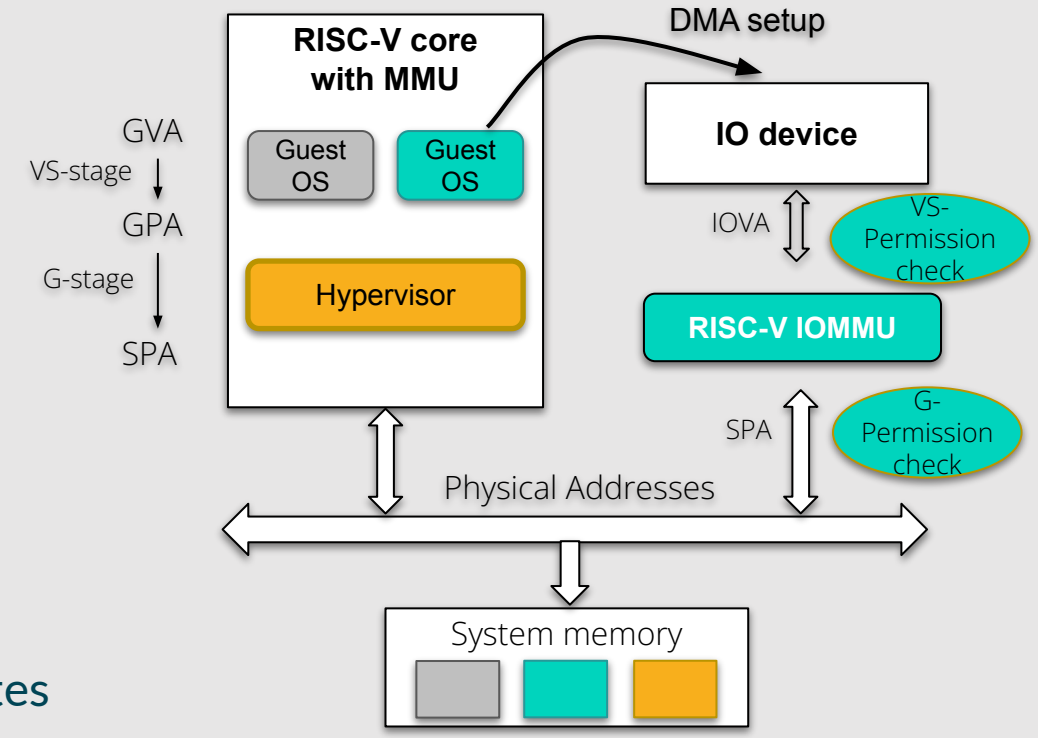
IOMMU needed to provide

- Memory protection
- Virtual address translation
- Virtual address space sharing between devices and CPU
- Interrupt remapping and virtualization

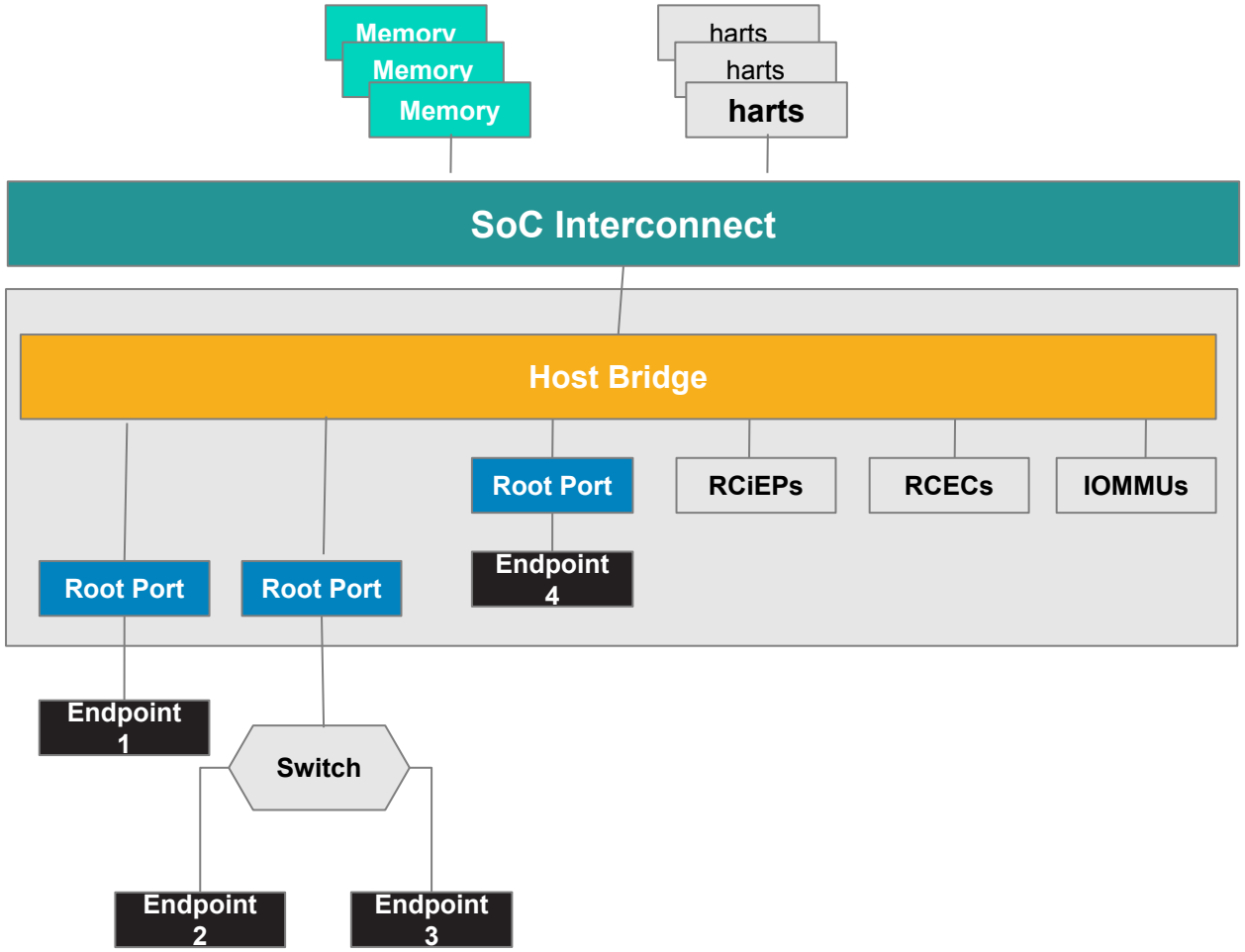
Require all DMA capable peripherals (RCIeP, non-PCIe, PCIe Root Ports) to be governed by a RISC-V IOMMU

Specifies:

- Minimum Device ID and PASID widths
- Support for PCIe Address Translation Services (ATS)
- Support for hardware performance monitor (HPM)
- Rules for integrating IOMMU with the host bridge such as for physical memory protection checks, physical memory attributes checks, and interpretation of IDs/Transactions



PCIe Subsystem



Root complex is a collection of root ports, root complex event collectors (RCEC), and root complex integrated endpoints (RCiEP)

Root complex uses a host bridge to connect to the CPU and system memory

Devices may be integrated into the SoC as either RCiEP or as an EP connected to a PCIe root port (endpoint 4 in this example)

One or more IOMMUs used to provide address translation and protection

Rules for Integrating the PCIe subsystem

Rules for routing of accesses to PCIe configuration space registers and handling of the completions

- Request Retry Status handling, Classification to Type 0/1 transactions, treatment of errors, unaligned accesses

Rules for mapping and routing memory mapped I/O (MMIO) registers

- Requirements on access width and alignment of accesses, treatment of errors, unaligned accesses
- Peer to peer request and completion routing

Rules for handling PCIe memory read/writes/atomics

- Ordering rules
- Rules for enforcement of coherence and cacheability

Rules for Error/Event reporting

- Advanced error reporting, error containment mechanism such as downstream port containment (DPC), associating RCiEP with RCEC

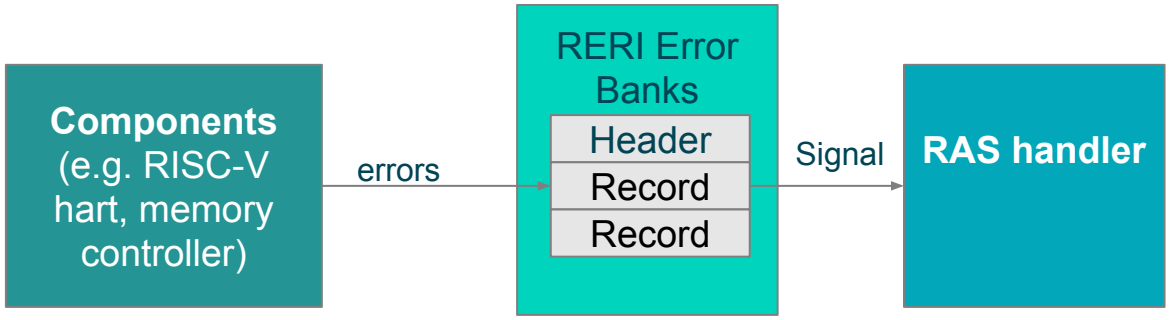
Rules for security capabilities

- Access control services such as source validation and translation blocking

Rules to implement precision time measurement

- Resolution of the PTM master time and software access to the master time

Reliability, Availability, and Serviceability



Reliability - probability that a system provides correct service

Availability - measure of tolerance of errors

Serviceability - time to restore the service to correct operation

- RISC-V RAS Error Record Register Interface (RERI) specification provides a standard interface for reporting errors including their severity, nature, and location

Server SoC expected to provide high levels of RAS

Guidelines on RAS implementation in the Server SoC

- Level of RAS depends on the reliability goals of the product - typically measured using metrics such as failure-in-time (FIT) and defects-per-million (DPM)
- Error detection and correction mechanisms for caches, system memory, and interconnects
- Periodic scrubbing of system memory for errors
- Error containment techniques such as data poisoning and rules for handling poisoned data

Rules for integrating RISC-V RERI for error recording and signaling

Quality Of Service

Modern SoCs feature tens of CPU cores, multiple levels of shared caches, shared interconnects, shared memory controllers

- Noisy neighbor problem - leads to non-deterministic workload performance
  - Interference among co-located tasks due to shared resource contention
- Underutilization problem
  - Difficulty in consolidating latency critical applications without compromising service level objectives (SLO)

Contention for shared cache capacity (e.g., LLC) or memory bandwidth significant source of interference

- Allocating dedicated capacity and bandwidth to workloads based on their SLO helps address these problems

RISC-V Ssqosid and RISC-V Capacity and Bandwidth QoS Register interface (CBQRI) specifications provide

- Methods to associate IDs for capacity/bandwidth allocation with workloads
- Configure capacity and bandwidth allocations in resource controllers

Server SoC specification provides guidelines on

- Mitigating unwarranted perf. interference by resource contention through capacity/bandwidth allocation mechanisms
- Integrating Ssqosid and RISC-V CBQRI extensions in the SoC
  - Significant caches and memory controllers, IOMMU
- Minimum number of resource control IDs and monitoring counter IDs

Manageability

Guidelines for the RISC-V server SoC to incorporate a standardized set of protocols and standards for server management

- Monitoring of sensors (temperature, power, etc.)
- Parameter control (power limits, etc.)
- Logging (RAS error records, etc.)

Using standards such as

- DMTF Redfish
- Platform level data model (PLDM)
- Management Component Transport Protocol (MCTP)

Guidelines on securing the management interface through use of standard protocols such as DMTF Security Protocol and Data Model (SPDM) for attestation and message encryption.

Guidelines on hardware interfaces for in-band and out-of-band management

- PCIe to BMC to facilitating uses such as remote KVM and management network
- I2C IPMI SSIF
- UART

Performance Monitoring

Guidelines for incorporating hardware performance monitors

- Significant caches
- Interconnects
- PCIe root ports

Guidelines on supporting collection of commonly used performance metrics such as bandwidth and latency to help guide workload placement and tuning

Guidelines for performance monitoring in NUMA configurations

Summary

RISC-V Server SoC specification provides rules and guidelines to control optionality in implementing RISC-V and Industry standards

To enables OS and hypervisor vendors to support such SoCs with a single binary OS image distribution model.

Calling on RISC-V member companies and ecosystem to implement and support the RISC-V Server SoC specifications and help make RISC-V servers predictable!