Are you ready for Software Defined Data Center?

Easily manage both bare-metal and virtualized platforms in your data center

Wiwynn® Cluster Manager with Intel® Rack Scale Design Technology

White Paper

Contents

Why Virtualization Cannot Deliver the Software Defined Data Center on Its Own

Virtualization technologies, as the keystone of cloud computing, have been widely deployed in the past decade. With virtualization technologies (*Figure 1*), data center operators can reduce over-provisioning, increase server utilization, provide high availability for applications deployed within virtual machines (VMs), simplify application deployment and migrations, as well as respond faster to changing business demands.

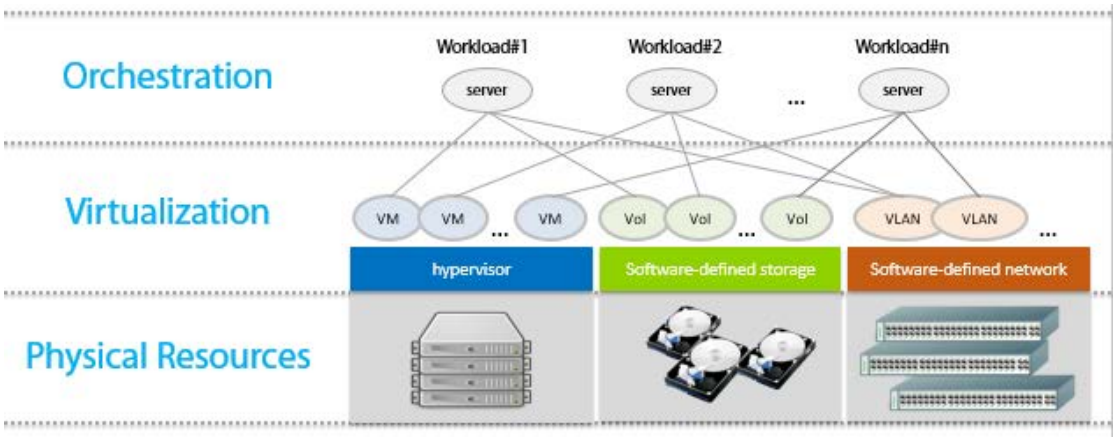

Figure 1. Virtualization deployed in datacenters today.

However, not all workloads are satisfied with VMs. Some require hardware that are not compatible with the virtualization software, or need to run on bare metal for performance reasons.

For example, FPGA cards may need extra space and power with dedicated PCIe switches to connect with high performance storage resources, such as NVMe SSDs. Therefore, the specialized FPGA servers cannot be easily assigned to a VM and applied to a workload without manual operations.

When workloads change, it is not easy to reassign these specialized servers to new workloads. As a result, new workloads may be under-provisioned, or the servers may remain idle.

Data center operators need to manually change configuration for better resource utilization. This is not only inefficient and costly, but may also damage hardware components.

New technologies are required to complement virtualization technologies and realize advanced software defined data center (SDDC).

Intel® Rack Scale Design Technology

Intel® created Intel® Rack Scale Design (Intel® RSD) to address problems that virtualization alone cannot solve while reducing Total Cost of Operation (TCO) for modern data centers as well.

Intel® RSD is a logical architecture. The key concept is to disaggregate hardware, such as compute, storage and network resources, from preconfigured servers and deploy them in sharable resource pools (*Figure 2*).

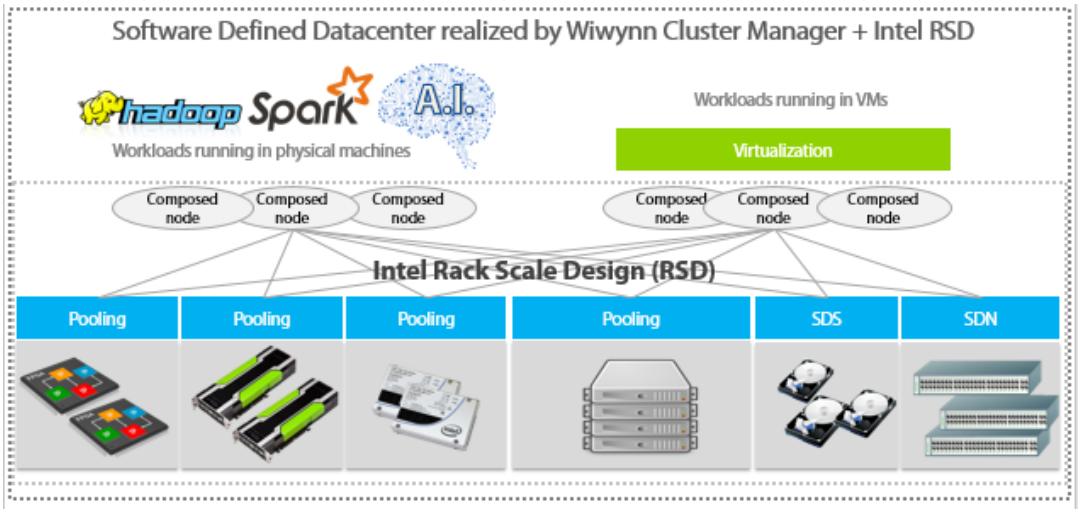

Figure 2. Intel® RSD, virtualization, and Cluster Manager realize complete SDDC

A logical node can be created based on the minimum workload requirements of the workload. Selected resources are physically connected through an interconnect fabric for the composed node and released to resource pools again once the workload completes. Configurations can also be adjusted in real time for the changing demand of application usage to maintain a target performance level. .

For example, there is a rack of preconfigured high performance servers equipped with two Intel® Xeon® processors, four accelerators and four NVMe SSDs while other racks are filled with 1U two-socket server and JBOD in the data center.

All of these high performance servers are running some workloads but not all of the equipped accelerators and NVMe SSDs are fully occupied. In the past, these unoccupied resources cannot be used by other servers and stay idle.

However, if the data center adopts Intel® RSD, theses accelerators and NVMe SSDs act as disaggregated resources from the preconfigured servers and can be composed with other compute nodes for specific workloads.

Therefore, for example, if there is a new deep learning demand requesting two Intel® Xeon® processors, eight accelerators and six NVMe SSDs, data center operators can easily get these resources from multiple preconfigured servers and compose a logical node for it. This process takes only minutes instead of hours or even days in the past since there is no hardware upgrade or manual reconfiguration required.

In summary, Intel® RSD provides more efficient infrastructure to keep up with increasing business demands, including

- Delivery of new services in minutes, not hours or days.
- Addressing application workload needs with agility
- Scaling capacity without interruption
- Reduction of operational and capital expenses

Wiwynn® Cluster Manager with Intel® RSD Integration

Wiwynn® Cluster Manager is a system software that makes data center easier to manage with features such as resource planning, massive firmware and OS deployment, real-time rack level visual monitoring.

For advanced SDDC capabilities, Wiwynn integrates Intel® RSD into Wiwynn® Cluster Manager and offers data center products that are compliant with Intel® RSD specifications. Wiwynn is one of the first OEMs in the industry to deliver Intel® RSD version 2.1 with high performance NVMe pooling.

Important Wiwynn® Cluster Manager's features includes

1. Integration with Intel® RSD for best-fit compute blades to meet given workload requirements with
2. NVMe pooling and dynamic composition
3. iSCSI remote boot and provisioning
4. Remote and local OS provisioning
5. Rack-level power monitoring
6. VLAN provisioning and switch port ACL assignment

Compared with widely used virtualization technologies, benefits of Wiwynn® Cluster Manager with Intel® RSD are shown in the following table.

Table 1. Benefits of Wiwynn® Cluster Manager with Intel® RSD

		Wiwynn® Cluster Manager with RSD	Virtualization platforms without RSD
H/W Utilization			
• CPU-shared by VMs		●	●
• HDD-shared by VMs		●	●
• CPU pooling on bare-metals		●	
• Provisioning	HDD[1]	●	3rd party option
	NVMe SSD[2]	●	
	Accelerators[2]	●	
H/W Reconfiguration (Bare Metal)			
• Operation		S/W commands	Manual
• Time required		Minutes	Hours or even days
• Hardware damage		No	Happens
H/W Upgrade			
• Cost		Cost effective	Costly
• Upgrade scale		Partial	Massive
• New NVMe, accelerator utilized by existing equipment		●	
S/W Deployment			
• OS and Firmware		Multiple deployment in the same time with one command	3rd party option

1. Bootable remote volumes for bare-metal with open APIs
2. Dynamic provisioning for compute nodes with open APIs

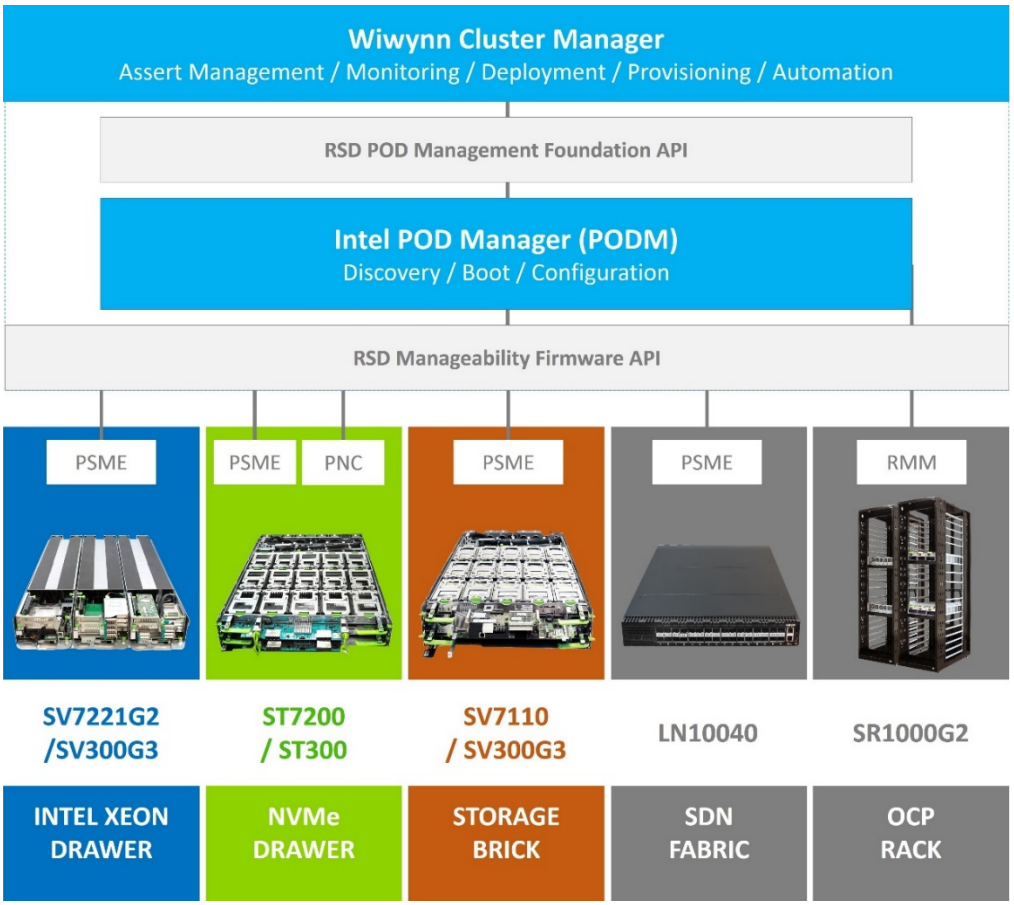

Figure 3. Architecture and building blocks of Wiwynn RSD solution

Figure 3 shows Wiwynn® Cluster Manager acting as the orchestration software above Intel® RSD technology. Wiwynn® RSD Enabled Cluster Manager includes the following software components:

1. **Pod Manager (PODM):**

   A pod is a collection of physical racks, drawers, modules, and blades. Pod Manager manages the physical infrastructure in a pod. It uses Redfish APIs to communicate with the lower-level software modules including PSMEs and RMMs that manage the individual disaggregated hardware resources within the racks.

2. **Rack Management Module (RMM):**

   RMM software handles infrastructure functions such as security certificates, IP address assignment, as well as power and thermal control and monitoring within a rack.

3. Pooled System Management Engine (PSME):

PSME software manages drawer and chassis. It communicates with the baseboard management controller (BMC) in a drawer or chassis to issue requests and report telemetry and asset information to PODM.

4. Pooled Node Controller (PNC):

PNC manages pooling of NVMe devices by logically binding PCIe end point devices to upstream ports that are typically attached to compute modules in the system. Wiwynn designed the PNC software to run in a BMC chip in our NVMe storage system (*Figure 4*) currently but will support systems for FPGAs and other accelerators in the future.

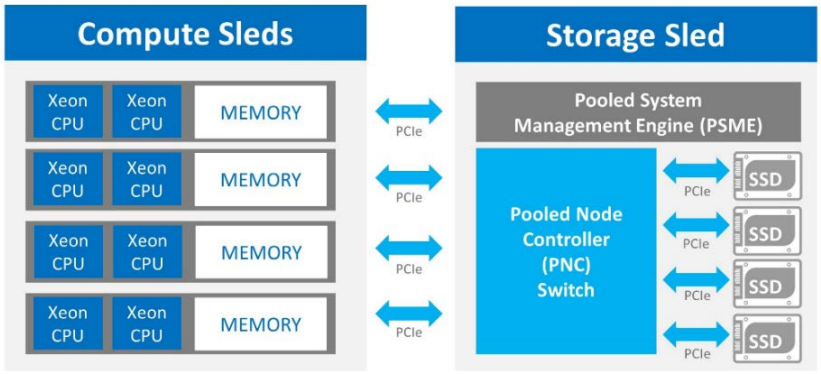

Figure 4. NVMe Pooling

5. Integrated User Interface with Cluster Manager:

Wiwynn® Cluster Manager provides a browser based user interface to simplify the management of an Intel® RSD system. It also provides additional functions that complement Intel® RSD to help end users manage data centers more effectively.

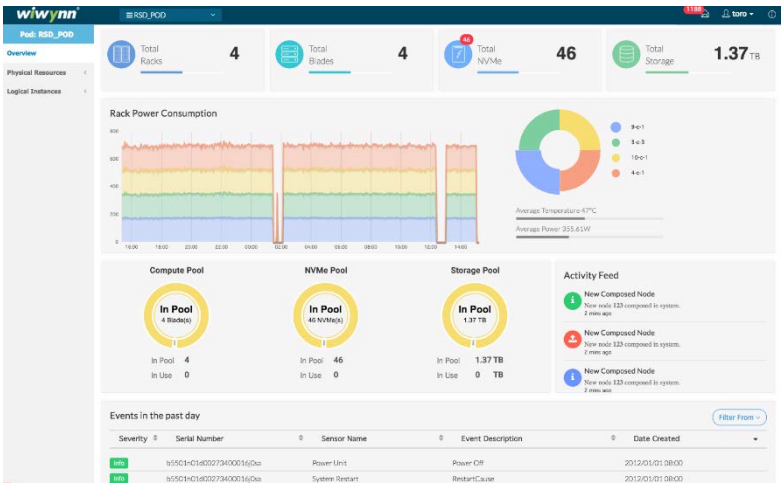

Figure 5. User Interface of Wiwynn Cluster Manager

Wiwynn Adds Innovations to Intel® RSD

Based on Intel® RSD, Wiwynn developed some enhanced functions and contributed to Intel® RSD spec to make the architecture more complete.

- **Ceph Integration with Enhanced Data Model**

  To make the provisioning and OS deployment on bare-metal servers as agile as virtual machines, bare-metal servers need to be diskless, i.e., to boot from a remote volume. Intel® RSD reference code provides Logical Volume Manager (LVM) services to create iSCSI volumes mapped to compute nodes. However, LVM is not a distributed storage system which may cause a system failure with a single disk failure.

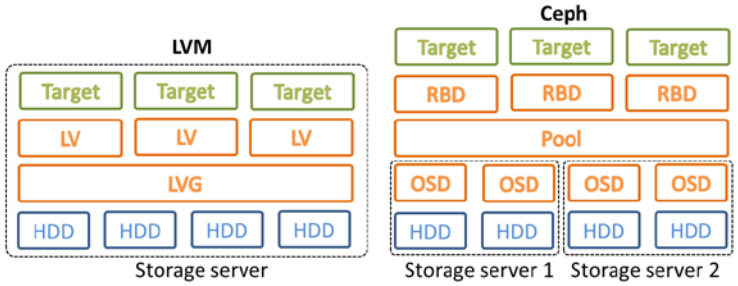

Figure 6. Ceph vs. LVM

To ensure highly reliable bootable iSCSI volumes for composed nodes, Wiwynn has integrated the Ceph storage service with Intel® RSD.

Figure 6 shows a typical LVM storage service residing on a single storage server, and its Logical Volume Group (LVG) managing physical drives in this storage server only.

In contrast, a Ceph storage service usually runs multiple object storage daemons (OSDs) across storage servers, allowing it to pool storage resources across physical enclosures. The data model designed for LVM is shown in *Figure 7*.

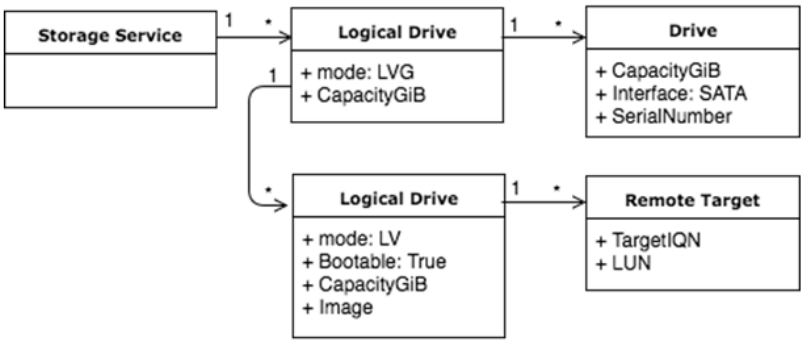

Figure 7. Data model designed for LVM

Therefore, we proposed the enhanced data model shown in *Figure 8* to enable Ceph in Intel® RSD. Wiwynn worked with Intel on design verification and validation before the software release, and it is now part of Intel® RSD 2.1.4 and its corresponding Compliance Test Suite (CTS).

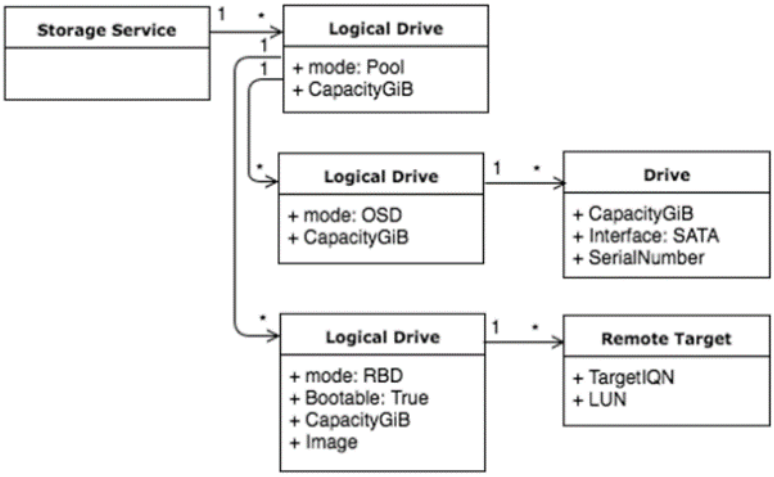

Figure 8. Data model for Ceph

- **Integration of PCIe Fabrics with new Redfish Schema**

NVMe pooling is a major new feature of Intel® RSD 2.1. It increases the utilization of high performance NVMe SSD with flexibility. However, the original design supports only one physical PCIe fabric and cannot satisfy various JBOF configurations. Take Wiwynn® RSD Enabled ST300, a 1U24B powerful NVMe JBOF, for example. There are two SKUs shown in *Figure* 9.

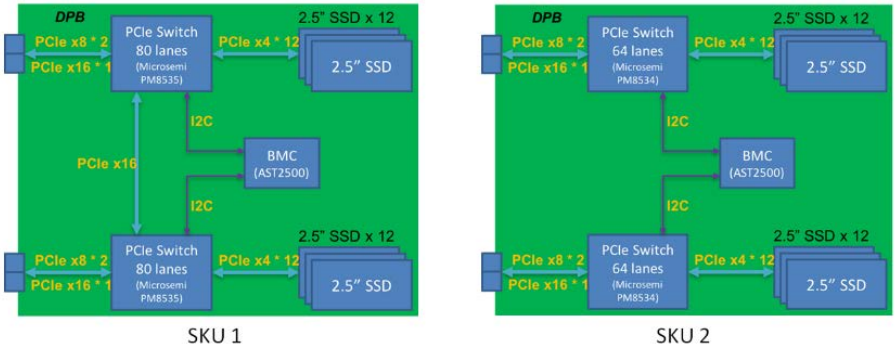

Figure 9. Architecture of ST300 NVMe JBOF

SKU1 has two 80-lane PCIe switches and a 16-lane interconnection between the two switches. This implementation enables all the PCIe endpoints inside the chassis to act as a single NVMe fabric and it works fine with the Intel® RSD reference code. SKU2 has two separate 64-lane PCIe switches, allowing them to be used as two independent NVMe fabrics inside a single chassis. However, the reference code cannot support this configuration.

More JBOF configurations benefit from Intel® RSD after Wiwynn worked with Intel to modify the reference code and proposed a new Redfish schema shown in *Figure 10*. Fabric 1 is independent of fabric 2, and the endpoints of fabrics 1 and 2 are associated with physical drives inside the same chassis. Both fabrics 1 and 2 are controlled by the same PNC agent residing in the shared BMC to ensure that control and data paths are properly coordinated to maintain data integrity and security.

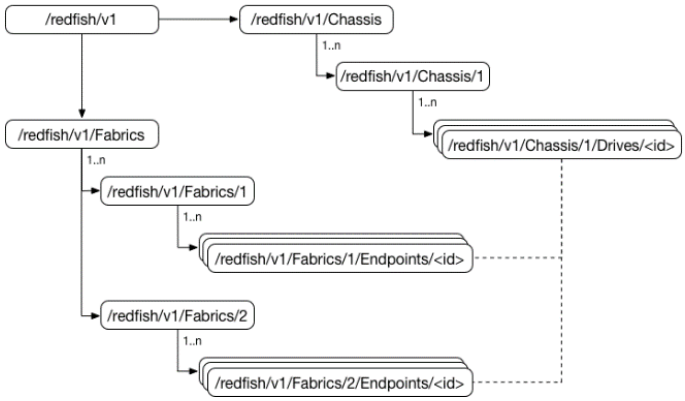

Figure 10. Redfish Schema to support the two-fabric ST300

- **Implementation of Flexible Rack-level Power Monitoring**

  Power is always one of the top priorities for datacenter management, as rack power consumption must be maintained within prescribed limits. Different data centers have unique preferences and considerations with regard to power monitoring, so Wiwynn implemented three options for end customers as shown in *Figure 11.*

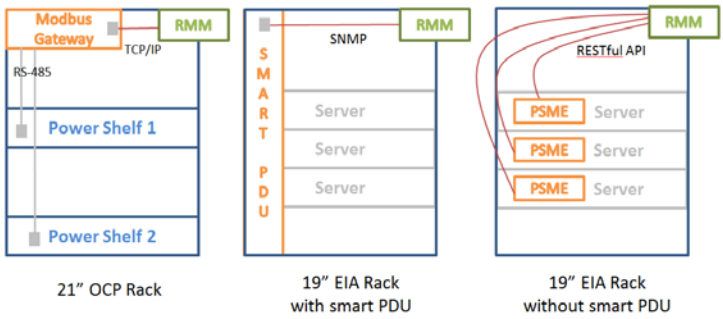

Figure 11. Flexible options of rack-level power consumption monitoring

For 21" OCP rack, Modbus gateway is adopted to monitor and control PSUs in power shelves. As for traditional 19" racks, RMM can read the power data via a smart PDU or calculate power consumption with PSME reported data while the smart PDU is not available.

Future Plans for Wiwynn® Cluster Manager

Wiwynn and Intel are working closely to make Intel® RSD solutions more complete and compelling. Here are some new features targeted for future releases by 18Q2.

- **Out-of-band Telemetry**

  Wiwynn will deliver Intel® RSD 2.2 compliant compute products with new telemetry capabilities based on hardware features in Intel® Xeon® Scalable processors and chipsets.

  A telemetry hub that supports health sensor reading and a CPU/DIMM utilization query will be integrated as the default monitoring mechanism. This allows customers to perform power and thermal optimization and workload balancing without installing in-band agents in the operating system.

- **Support of Compute Accelerator Pooling**

  Wiwynn products based on Intel® RSD 2.2 will also include data models for describing compute accelerators (ex. FPGA), allowing customers to specify accelerator requirements when composing a node.

- **NVMe Over Fabric**

  NVMe over Fabric (NVMeF) will make NVMe pooling even more flexible and scalable. As a leading supplier of NVMe JBOF solutions, Wiwynn has committed to deliver an NVMeF product compliant with Intel® RSD 2.3 in the second quarter of 2018.

Wiwynn® Cluster Manager provides advanced SDDC capabilities with Intel® RSD Through the standard API, Wiwynn® Cluster Manager can aggregate resource information and execute node composition without manual operation. Data center operators can utilize IT resources in more effective ways. With further development, Wiwynn® RSD Enabled solution will provide enhanced functions and realize SDDC further.

How to Deploy Wiwynn® RSD Enabled Solution in Data Center?

Wiwynn® Cluster Manager integrated with Intel® RSD can be deployed in a VM or container on the existing management server in the data center or on the storage server in the RSD Enabled rack.

For proof of concept (POC) purpose, the basic rack configuration (*Figure 12*) includes four compute blades, one NVMe JBOF, one storage server, one 1G management LAN switch and one 10G data LAN switch. Wiwynn® RSD Enabled firmware and configuration are required.

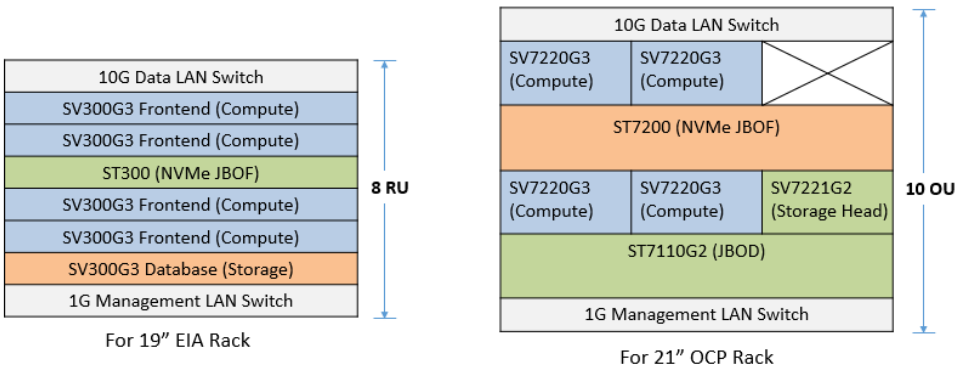

Figure12. Basic rack configuration for Wiwynn RSD POC

For mass deployment, rack configurations can be varied and dependent on the types of required workloads in the data centers as shown in *Figure 13*.

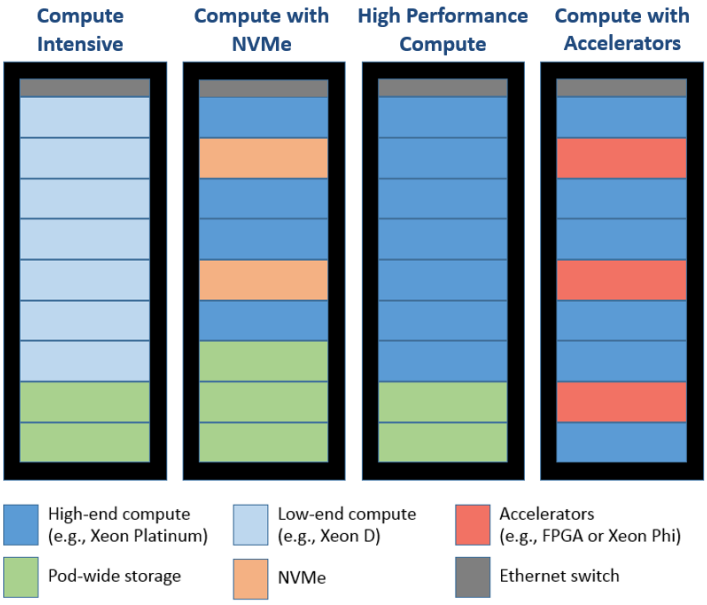

Figure 13. Example of Wiwynn® RSD Enabled pod

The Wiwynn® RSD Enabled pod could be a collection of racks mixed with compute blades based on different models of Intel® Xeon® processors, HDDs, NVMe drives, and accelerators. All of these compute and storage resources could be composed dynamically for specific workloads.

Furthermore, Wiwynn® RSD Enabled pods in different versions can also coexist with legacy pods in a data center (*Figure 14*). Wiwynn® Cluster Manager will act as the orchestration software and data center operators can gradually deploy Wiwynn® RSD Enabled racks for new demands or equipment upgrade.

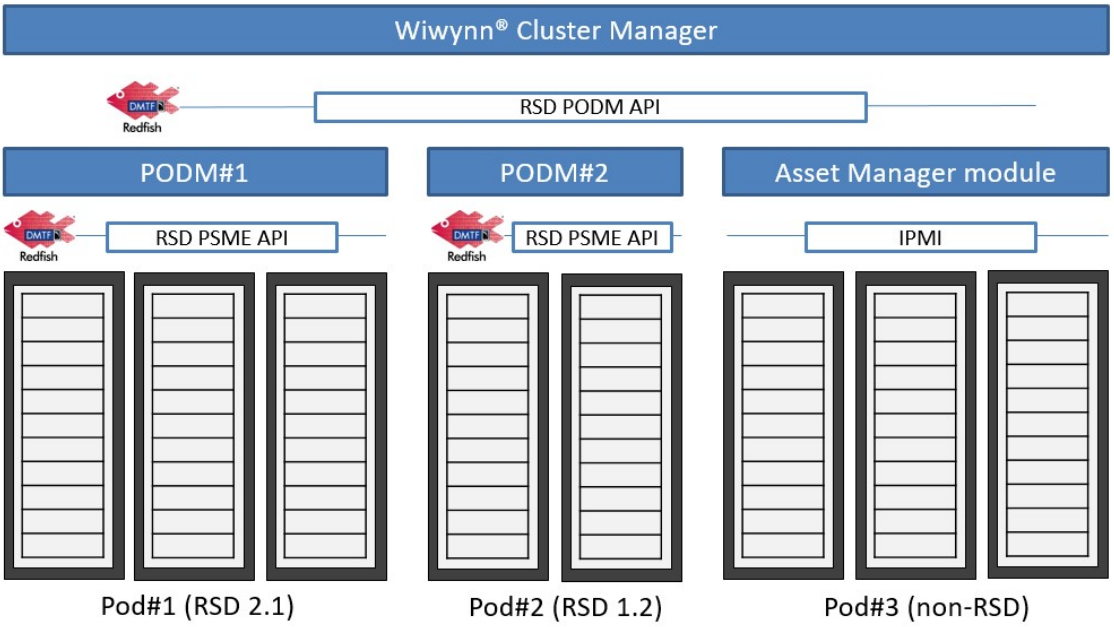

Figure 14. RSD and non-RSD pods coexist in a data center

To Learn More

For further information, please refer to the following sites:

- Wiwynn® Cluster Manager:

  http://www.wiwynn.com/english/product/type/details/34?ptype=40

- Wiwynn® RSD Enabled Products:

  http://www.wiwynn.com/english/product/type/44

- Intel® Rack Scale Design:
  https://www.intel.com/content/www/us/en/architecture-and-technology/rack-scale-design-overview.html

For requirements, please contact Wiwynn sales team at sales@wiwynn.com

About Wiwynn

Wiwynn® is an innovative cloud IT infrastructure provider of high quality computing and storage products, plus rack solutions for leading data centers. We aggressively invest in next generation technologies for workload optimization and best TCO (Total Cost of Ownership). As an OCP (Open Compute Project) solution provider and platinum member, Wiwynn® actively participates in advanced computing and storage system designs while constantly implementing the benefits of OCP into traditional data centers.

For more information, please visit http://www.wiwynn.com/english or contact sales@wiwynn.com
Follow Wiwynn® on Facebook and Linkedin for the latest news and market trends.