Product Brief

Hyperconverged Infrastructure (HCI)
Intel® Volume Management (Intel® VMD) Device Assign

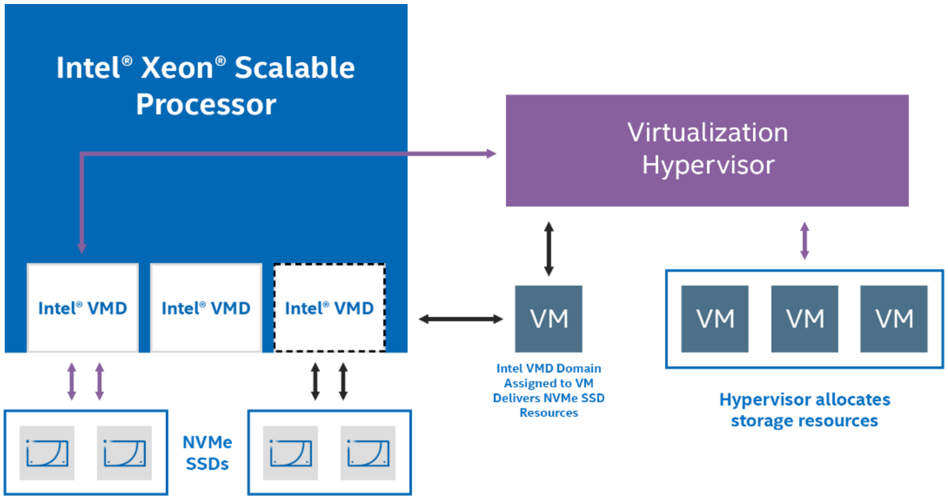
# Modernizing HCI Storage Allocation to Improve Performance with NVMe SSDs

**Intel® Volume Management (Intel® VMD) Direct Assign is a new HCI storage architecture to bypass hypervisor resources, delivering NVMe performance and Intel VMD functionality directly to the Virtual Machines.**

Hypervisor technology has evolved to functionally divide bare-metal server resources into virtual instances for increased IT flexibility and maximize hardware utilization. However, this resource management can cause increased latency and decreased performance for certain components. For example, hypervisor storage stacks can limit the throughput of next generation NVMe SSDs due to added software overhead. To take advantage of improving storage technologies and to modernize the hyperconverged infrastructure (HCI) architecture, a new storage allocation process is required.

Intel® Volume Management Device (Intel® VMD) Direct Assign is a specific application of Intel VMD technology that aims to improve the storage experience in HCI architectures. Intel VMD is a feature in next-generation Intel® Xeon® Scalable processors that acts like an integrated controller inside the processor PCIe* root complex. The NVMe SSDs are directly connected to Intel VMD domains and the Intel VMD driver stack owns those devices. This creates a robust and consistent storage environment to provide functionalities such as storage isolation, error handling, LED management, and hot-plug support. Intel VMD Direct Assign then takes this isolated storage subsystem and bypasses the hypervisor to give (assign) Intel VMD control directly to a virtual machine (VM) running on the host. This VM inherits all the functionality of Intel VMD and the performance of the NVMe SSDs, without the storage bottleneck of the hypervisor. This VM can then allocate storage resources out to other VMs, therefore acting like a virtual storage controller, or use the fast storage resources directly.

## Utilize a Storage Controller VM




Hyperconverged infrastructures are often specifically designed to maximize the utilization of underlying hardware, which is often measured in the total number of supportable VMs. This is the predominant architecture in cloud deployments and can mean that one platform needs to support the storage resources assigned to tens or even hundreds of virtual machines. Using NVMe SSDs can increase the available storage bandwidth, allowing more storage throughput allocated to each VM. HCI architectures, however, need to change to properly deliver that bandwidth to the virtual environments. To accomplish this, HCI vendors can implement a storage controller VM. This is a special VM that allocates storage resources at the virtualized level rather than the hypervisor level. Intel VMD drivers can be used to pass storage resources directly to the storage controller VM, essentially removing any hypervisor overhead. The storage controller VM can use an efficient storage stack to better allocate storage resources. A single NVMe SSD can be assigned to a storage controller VM, but this means the VM OS must provide all functionality for the PCIe subsystem (i.e. LED management, hot-plug, error handling). This can deliver excellent performance but could cause instability on the platform. Intel VMD Direct Assign is a more robust implementation of this architecture, where the entire Intel VMD domain and any NVMe SSDs connected to that domain are passed to the storage controller VM. All Intel VMD functionalities are also passed along, delivering storage reliability and consistency along with the performance boost of NVMe SSDs.

## Accelerate Performance Sensitive VM Instances

Hyperconverged infrastructure is not always used for high density utilization of hardware. Sometimes, it aims to support increased flexibility of application deployments instead. This is more common when a small number of VMs (<4) is supported on a platform. The specific instances may very over time, meaning new VMs can be launched or removed based on the current needs of the organization. Also, in case of hardware failure, VMs can easily be shifted to a new platform to prevent downtime. This is particularly important for mission critical tasks such as security or network processing. The usage of VMs instead of bare-metal servers gives users the flexibility to manage applications in a simpler way. As the broader ecosystem around these HCI deployments (processing power and networking) is accelerating, storage needs to keep up in order to maintain the virtualized usage model. Intel VMD Direct Assign can deliver the same benefit of storage reliability and consistency along with the performance boost of NVMe SSDs to this flexible IT implementation. The primary difference here is that the VM itself consumes the storage resources directly to accelerate the application running. However, this does mean that each VM would need its own Intel VMD domain and storge devices. If 4 VMs are running and need storage acceleration, then 4 Intel VMD domains would need to be populated and assigned accordingly. Alternatively, if only one of the VMs on a platform is storage dependent, then it can leverage Intel VMD Direct Assign, while remaining VM's use standard storage allocation from the hypervisor.

| Features at a Glance | Details |
|---|---|
| Supported Platforms | Platforms with Intel® Xeon® Processor Scalable processor family (Generation 1, 2, and 3).<br>Requires Intel® Volume Management Device (Intel® VMD) hardware and proper enabling at the platform level. |
| Supported Configurations | • NVMe SSDs must be connected to Intel VMD enabled PCIe lanes.<br>• Each Intel VMD Domain can be assigned to a maximum of 1 VM.<br>• All NVMe SSDs connected to an Intel VMD domain will be assigned to the guest OS. Partial assignment is not supported.<br>Supported hypervisor AND guest OS must be used for Direct Assign to be functional. |
| Supported Hypervisors | VMware ESXi, Linux KVM, Nutanix AHV |
| Supported Guest OS | RHEL, SUSE, Ubuntu |
| Key Features | • Storage performance from hypervisor bypass of storage resources<br>• Management Tools (UEFI CLI, UEFI HII, OS CLI, GUI)<br>• Surprise hot-plug<br>• Status LED indication<br>• Storage isolation of PCIe subsystem from VM guest OS<br>Error handling |