Intel® Data Plane Development Kit (Intel® DPDK) with VMware vSphere®

August 2014

SOLUTION OVERVIEW

Table of Contents

Introduction

Businesses are moving away from their dependence on an ever-increasing variety of non-virtualized, dedicated, proprietary physical solutions. Instead, they are moving to agile, high performance, high throughput, and highly efficient virtualization solutions that are cloud ready.

In order to be cloud ready, virtualization solutions need to span across server, storage, and networking. For compute and I/O intensive applications and architectures that require high bandwidth and predictable latency, businesses typically discover that it is not enough to just deploy flexible, virtual server applications – the underlying I/O must also be flexible, virtual, and easy to configure and deploy.

Historically, traditional core networks used custom hardware based on the need for performance and capabilities that standard off-the-shelf hardware could not provide. This left the intelligence of the network embedded in custom hardware and chipsets that were expensive to buy, difficult to manage, and slow to change. These non-virtualized, legacy solutions also kept businesses locked into dedicated, proprietary hardware, and inflexible legacy applications and architectures.

Today, many Telecom service providers are using “Network Functions Virtualization” (NFV) to transform the economics of traditional proprietary hardware architectures around the core network. Instead of non-virtualized, proprietary hardware, service providers can now use standard Intel® architecture-based servers to execute most mission critical workloads. Moving beyond Software-Defined Data Center (SDDC) concepts, service providers are embracing NFV as the logical next step in the on-going effort to abstract hardware resources across all types of data center infrastructure.

This paper describes the Intel® Data Plane Development Kit (Intel® DPDK) 1.6 with VMware vSphere® (VMware® ESXi™) 5.5 solution. Intel DPDK packet processing software running on standard Intel® architecture-based servers can now perform tasks that were traditionally assigned to hardware, such as application-specific integrated circuits (ASICs) and field programmable gate arrays (FPGAs). The Intel DPDK with VMware vSphere solution can be used to rapidly migrate legacy applications from non-virtualized, dedicated proprietary hardware to high performance, high throughput, and highly efficient cloud-ready environments.

VMware and Intel Partnership

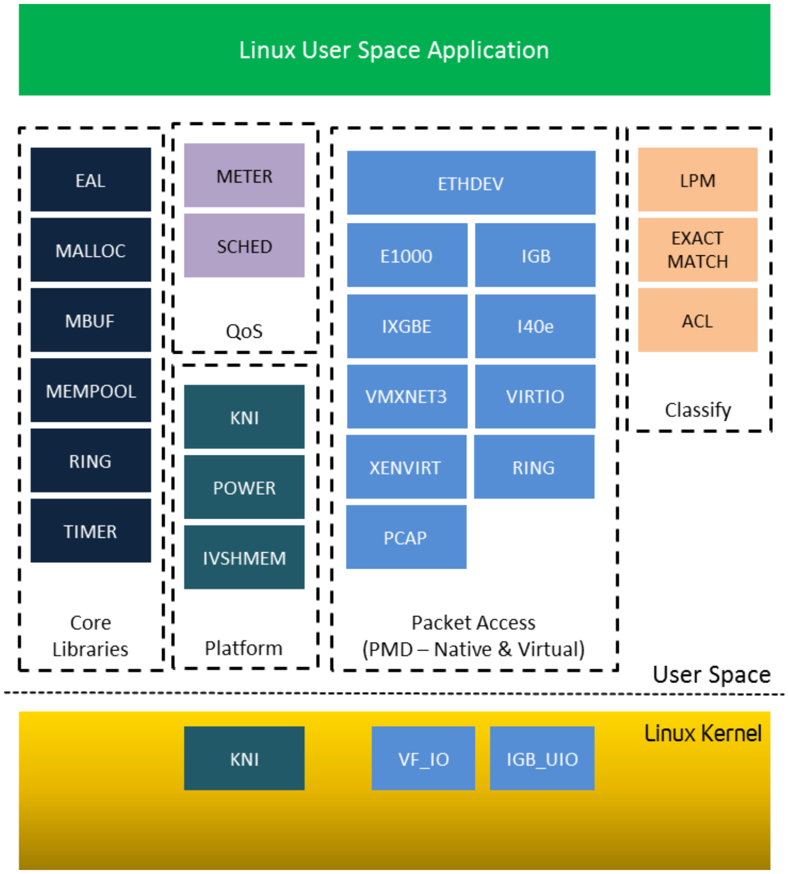
Intel® and VMware® are working together through collaboration and joint innovation with virtualized data centers to migrate legacy solutions, and then to “future proof” applications with next-generation capabilities. This proven partnership has resulted in a differentiated Intel DPDK solution that delivers agile and scalable virtual and physical architectures that can provide for high throughput and predictable latency.

Each of the enabling technologies of the Intel DPDK with VMware vSphere solution is described in the sections below.

Intel Data Plane Development Kit

The Intel® Data Plane Development Kit (Intel® DPDK) is an open source, BSD-licensed optimized software library for Linux User Space applications. Intel DPDK enables higher levels of packet processing throughput than what is achievable using the standard Linux kernel network stack. This optimized library gives application developers the ability to address challenging data plane processing needs, typically found in Telecom and networking workloads, all in software and on general purpose, Intel architecture-based processors.

Figure 1. Intel DPDK high level architecture



The key software components for Intel DPDK include the Environment Abstraction Layer (EAL), Memory Pool Manager, Buffer Manager, Queue Manager, Ring Manager, Flow Classification, and Poll Mode Drivers for 1 Gigabit Ethernet (GbE) and 10 GbE controllers.

The Intel DPDK fundamentals include:

- Implements a “run-to-completion” model.
- Accesses all devices by polling without a scheduler.
- Accesses all devices directly from Linux User Space.

- Runs in 32-bit and 64-bit mode, with or without non-uniform memory access (NUMA).
- Scales from Intel® Atom™ processors to Intel® Xeon® processors.
- Supports an unlimited number of processors and processor cores.
- Optimizes packet allocation across dynamic random access memory (DRAM) channels.
- Allocates memory from the local node when possible.
- Ensures that all data structures and objects are cache-aligned for superior performance.

Designed to accelerate packet processing performance, Intel DPDK provides Intel architecture-optimized libraries that allow developers to focus on their applications. Intel DPDK provides non-GPL source code libraries that can enable exceptional data plane performance, ease software development, and minimize development time. Developers can make additions and modifications to the Intel DPDK source code in a production network element to meet individualized system needs.

Intel DPDK provides software pre-fetching, which increases performance by bringing data from memory into cache before it is needed, thereby significantly reducing memory latency. Developers can build applications with the libraries using “run-to-completion” and/or “pipeline” models that enable the equipment manufacturer’s application to maintain complete control.

VMware vSphere

VMware vSphere® is an optimal, industry leading virtualization platform that can be used as an enabling technology for building cloud computing architectures. vSphere delivers control over IT resources with the highest efficiency and choice in the industry, and it permits IT to meet service level agreements (SLAs) for the most demanding business-critical applications. This helps to lower the total cost of ownership (TCO).

vSphere virtualization creates a layer of abstraction between the resources that are required by an application and operating system, and the underlying hardware that provides those resources. It enables multiple, isolated execution environments to share a single hardware platform.

vSphere accelerates the shift to cloud computing for existing data centers. It also supports compatible public cloud offerings, forming the foundation for the industry’s only hybrid cloud model. This makes VMware vSphere a trusted platform for any application.

Figure 2. VMware vSphere virtual infrastructure

VMware vSphere virtualization provides for:

- **Consolidation**. VMware virtualization allows multiple application servers to be consolidated to one physical server, with little or no decrease in overall performance. This helps to minimize or eliminate underutilized server hardware, software, and infrastructure.
- **Manageability**. The live migration of virtual machines from server to server, and the migration of the associated storage, is performed with no downtime using VMware vSphere® vMotion® and VMware vSphere® Storage vMotion®.
- **Availability**. High availability can be enabled to reduce unplanned downtime and enable higher service levels for applications. VMware vSphere® High Availability (HA) ensures that, in the event of an unplanned hardware failure, the affected virtual machines are automatically restarted on another host in the vSphere cluster.
- **Automation**. VMware's automated load balancing takes advantage of vMotion and Storage vMotion to migrate virtual machines among a set of VMware® ESXi™ hosts. VMware vSphere® Distributed Resource Scheduler™ (DRS) and VMware vSphere® Storage DRS™ enable automatic resource relocation and optimization for virtual machines and storage.

- **Provisioning**. VMware virtualization encapsulates an application into an image that can be duplicated or moved, greatly reducing the cost of application provisioning and deployment.

In addition, VMware vSphere® 5.5 extends the core capabilities of the vSphere platform with the introduction of new networking features and enhancements. These include Lightweight Access Control Protocol (LACP) enhancements, traffic filtering, quality-of-service (QoS) tagging, single root I/O virtualization (SR-IOV) enhancements, enhanced host-level packet capture, and 40 GbE network interface card (NIC) support.

VMware Infrastructure Components

The Intel DPDK virtualization solution employs VMware vSphere virtual networking elements that are similar to those used in the physical environment, but with some advanced capabilities.

The VMware infrastructure used in the Intel DPDK virtualization solution includes:

- VMXNET3
- Virtual Distributed Switch
- Direct assignment using either VMware vSphere® DirectPath I/O™ or SR-IOV

Each of these elements is described in the sections below.

VMXNET3

Migrating legacy applications to a virtual environment can incur unwanted overhead. VMware VMXNET3 is a para-virtual network interface card (vNIC) that is optimized to provide high performance, high throughput, and predictable latency to ESXi virtual machines, while enabling advanced features such as vMotion and High Availability. Each virtual machine on an ESXi host can have up to 10 VMXNET3 virtual NICs.

Virtual Distributed Switch

VMware Virtual Distributed Switch (vDS) acts as a single, built-in Layer 2 switch across all associated hosts that are managed by administrators. The vDS infrastructure enables virtual machines to maintain a consistent network configuration. vDS is used to abstract physical network interfaces and provide access-level switching in the ESXi hypervisor.

Direct Assignment

Direct assignment using either VMware vSphere DirectPath I/O or SR-IOV allows guest operating systems to directly access an I/O device, bypassing the virtualization layer. This direct path, also called “passthrough”, can deliver higher performance for ESXi systems that utilize high speed I/O devices, such as 10 GbE.

- **vSphere DirectPath I/O**. This option enables virtual machines to directly access physical PCI express adapters on platforms with an I/O memory management unit. This includes Intel® Xeon® processor 5500 series systems that feature an implementation of the I/O memory management unit (IOMMU), also called Intel® Virtual Technology for Directed I/O (Intel® VT-d).

For more information about vSphere DirectPath I/O, go to "Configuring VMDirectPath I/O pass-through devices on a VMware ESX or VMware ESXi host (1010789)" at: http://kb.vmware.com/kb/1010789.

- **Single Root I/O Virtualization (SR-IOV)**. This option allows the physical PCI Express adapter to be partitioned into multiple virtual functions using an industry-standard mechanism. A single physical NIC is enabled using an ESXi physical function driver or PF, and each partitioned virtual function is enabled using a virtual function driver or VF, for the specific guest operating system used in each virtual machine. For example, the Intel® 82599 10 Gigabit Ethernet Controller (Niantic) can be partitioned into 63 VFs to assign to any virtual machines.

  In addition, Intel® DPDK 1.7 provides support for the Intel 40 GbE SR-IOV VF driver.

  VMware fully supports SR-IOV with an ecosystem and partners that can assist businesses to combine SR-IOV with best-of-breed virtualization.

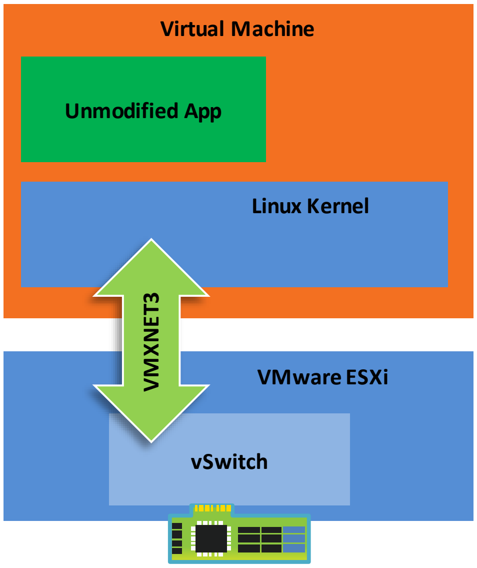
  For more information about SR-IOV, go to "SR-IOV support status FAQ (2038739)" at: http://kb.vmware.com/kb/2038739

The usage models for directly assigning PCI express devices to virtual machines are different between vSphere DirectPath I/O and SR-IOV, when used with Intel DPDK virtualization. These differences do not directly affect the overall design considerations for this solution.

Solution Architecture

As an introduction, typical workloads that do not require Intel DPDK use standard para-virtual device connectivity, which exposes the physical NIC using the VMXNET3 para-virtual network device as a standard Ethernet NIC. In this case, vSphere provides the connectivity between the physical NIC and the vNIC. The vNIC is used by the guest operating system TCP/IP networking stack and it offers the standard socket-based programming interface to applications in the virtual machine.

Figure 3. Standard para-virtual device connectivity



The ESXi hypervisor also supports the E1000 and E1000E virtual devices, which are emulations of Intel 1 GbE devices. Although each of these vNIC devices uses different implementation technologies, they provide a level of abstraction similar to that of the VMXNET3 para-virtual network device.

However, for I/O intensive workloads that demand significantly higher packet throughput performance, the Intel DPDK virtualization solution can use either of the following connectivity types:

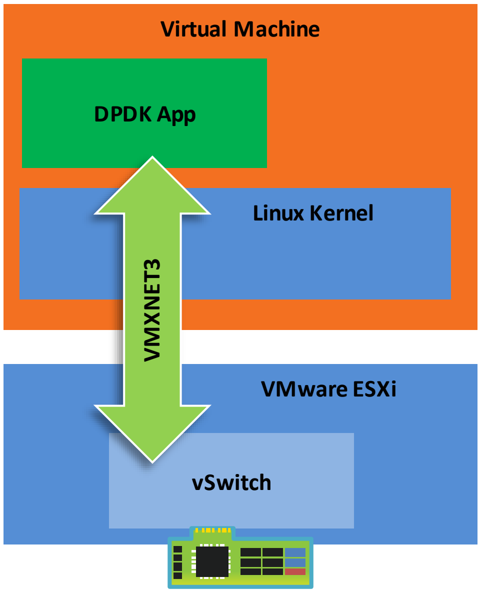
- Intel DPDK para-virtual device connectivity
- Direct assignment connectivity

Each of these solution architectures is described in the sections below.

Intel DPDK Para-virtual Device Connectivity

Intel DPDK para-virtual device connectivity provides an alternate, higher performance I/O data path to applications running in an ESXi virtual machine, as compared to the guest operating system TCP/IP stack with its socket-based application programming interface (API).

Figure 4. Intel DPDK para-virtual device connectivity



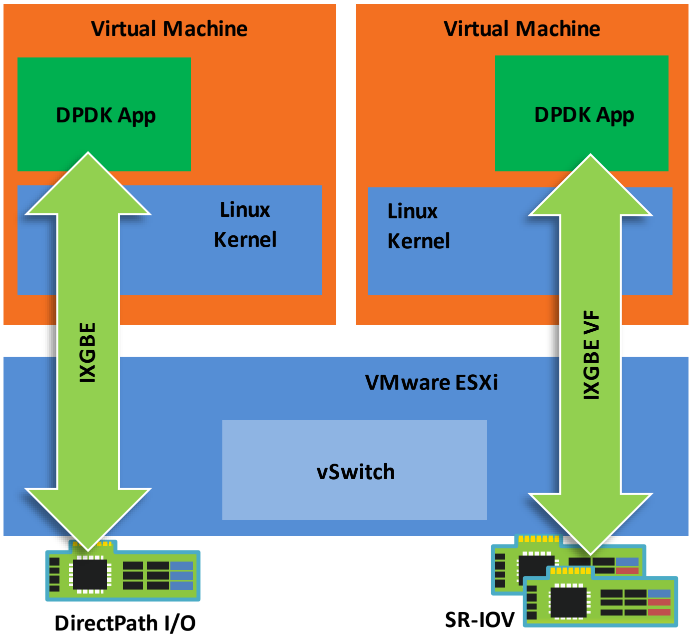
With Intel DPDK 1.6, Intel introduces built-in support for the VMXNET3 para-virtual NIC. Note that Intel DPDK on the VMware platform already provides support for the E1000 and E1000E virtual devices (introduced in DPDK 1.3.0).

Intel DPDK para-virtual device connectivity fully enables all of the vSphere advanced virtualization features including vMotion, DRS, Fault Tolerance, High Availability, Snapshots, and others, which allows applications to be migrated, optimized, and then made cloud ready.

Direct Assignment Connectivity

Direct assignment using vSphere DirectPath I/O or SR-IOV with Intel DPDK can be used to migrate legacy applications to a virtualization environment. Both types of direct assignment connectivity can adjust the application environment to provide for near native packet performance, high throughput, and predictable latency, as compared to what is available through para-virtual NICs such as VMXNET3.

Figure 5. Direct assignment connectivity



However, when either vSphere DirectPath I/O or SR-IOV is used to migrate a legacy application, it bypasses the hypervisor layer for I/O, which disables the VMware advanced features that are needed to create a cloud-ready environment, including vMotion, DRS, Fault Tolerance, High Availability, Snapshots, and others. For example, because SR-IOV ties a virtual machine to the host, rapid service delivery cannot be provided because manual steps are required to provision each host.

Intel DPDK in the Cloud

With Intel DPDK and VMware vSphere, businesses can get the best of both the physical environment and virtual environment – all in one solution.

Intel is committed to delivering cloud-ready support for Intel DPDK. With Intel DPDK 1.6, Intel introduces built-in support for the VMXNET3 para-virtual NIC, which makes all vSphere advanced features fully available. Note that Intel DPDK already provides support for the E1000 and E1000E virtual devices on the VMware platform.

The Intel DPDK with VMware vSphere solution enables applications to be built with virtualization in mind. When used with Intel DPDK para-virtual device connectivity, the solution provides these key advantages:

- **Portability**. When developed with Intel DPDK, an application can maintain a single code base as it is migrated from the physical environment to the virtualization environment. This provides for easier portability between physical and virtual environments, lower maintenance costs, and faster time-to-market.

In addition, the application can run on a broader set of hardware platforms with different physical NICs, as long as the physical NIC has a supported ESXi driver associated with it.

- **Performance**. With Intel DPDK, a vNIC in a vSphere virtual machine can bypass the Linux kernel in the guest operating system, which gives the application direct access to the vNIC. Applications can be configured and tested to achieve significantly higher performance. In addition, Intel DPDK helps to minimize any application performance degradations by limiting the hypervisor overhead.
- **Agility**. By using Intel DPDK para-virtual device connectivity, the solution can enable an agile, cloud-ready environment using vSphere advanced features, including vMotion, DRS, Fault Tolerance, High Availability, Snapshots, and others. The solution provides for business agility, automated provisioning, scalability, and lower CAPEX and OPEX.

Certain Telecom applications will continue to have stringent network I/O performance requirements, after they are deployed to the virtualization environment. These key performance factors are packet throughput (due to the relatively small packet size), latency, and jitter.

For this type of workload, it is important to consider the factors and trade-offs involved in providing for flexibility, innovation, and rapid deployment. Horizontal scalability is generally preferred over vertical density in order to provide better overall performance. This type of architecture should be the cornerstone for future-proofing next generation Telecom applications.

Alternatively, Intel DPDK, in conjunction with either vSphere DirectPath I/O or SR-IOV connectivity, can be a viable option for migrating these types of applications to the virtualization environment, especially when redesigning for horizontal architecture is either not possible or cost prohibitive.

However, it is important to note that vSphere DirectPath I/O and SR-IOV cannot be used to enable a cloud-ready environment. When either one of these connectivity options is used to migrate a legacy application, it bypasses the hypervisor layer for I/O, which disables the VMware advanced features that are needed to build the cloud environment, thereby removing the key benefits that can be achieved through virtualization.

For example, with vSphere DirectPath I/O, DRS can be deployed, but with limited availability only – the virtual machine can be part of a cluster, but it cannot migrate across hosts.

In addition, because SR-IOV ties a virtual machine to the host, rapid service delivery cannot be provided since manual steps are required to provision each host. SR-IOV deployments can also be error prone when there are mismatches and incompatibilities between the VF driver in a virtual machine and the PF driver in the ESXi hypervisor, and this can potentially lead to system instability.

Conclusion

Businesses are caught between the increasing costs of infrastructure and decreasing margins. The Intel DPDK with VMware vSphere solution enables businesses to future proof applications to achieve the best of both the physical environment and virtual environment.

Intel DPDK with VMware vSphere allows businesses to rapidly migrate applications requiring high packet performance to a virtualization environment. Working together with VMware, Intel is committed to delivering cloud-ready support using Intel DPDK para-virtual device connectivity. In fact, Intel DPDK 1.6 provides built-in support for the VMXNET3 para-virtual NIC. This solution provides the advantages of portability, performance, and agility.

When the consideration for high packet performance is crucial, Intel DPDK can also be used with vSphere DirectPath I/O or SR-IOV to migrate applications to a virtualization environment. Although certain key benefits that provide for agility are lost, these connectivity options can be used as the “first step” in optimizing those applications.

As your partner for SDDC and Network Functions Virtualization, VMware can help you assemble, test, and deploy the right virtualization solution to fit your core network needs and business strategy.

Resources

For more information about the Intel DPDK and VMware products discussed in this paper, view the links and references below.

Intel References

Intel DPDK

- Packet Processing on Intel® Architecture:
  http://www.intel.com/go/dpdk
- DPDK Community Project (Intel® DPDK: Data Plane Development Kit):
  http://dpdk.org/

Intel Xeon Processor E7 / VMware vSphere

- VMware vSphere™ 5.0 and the Intel® Xeon® Processor E7 Family:
  http://www.vmware.com/files/pdf/partners/intel/VMware-Intel-Datacenter-VTSB-Solution-Brief.pdf
- VMware vSphere™ 5.1 and the Intel® Xeon® Processor E7 Product Family:
  http://www.vmware.com/files/pdf/partners/intel/Intel-E7-VMware-vSphere-SB.pdf

VMware References

VMware vSphere

- VMware vSphere Documentation:
  http://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-pubs.html
- Performance Best Practices for VMware vSphere 5.5:
  http://www.vmware.com/pdf/Perf_Best_Practices_vSphere5.5.pdf
- Featured VMware Documentation Sets:
  http://www.vmware.com/support/pubs/
- VMware Licensing Help Center:
  http://www.vmware.com/support/licensing/

VMware Community and TV

- VMware Community, VMware Technology Network (VMTN):
  https://communities.vmware.com/community/vmtn
- VMware Best Practices:
  https://communities.vmware.com/community/vmtn/bestpractices
- VMware Community, VMware Knowledge Base:
  http://communities.vmware.com/community/vmtn/resources/knowledgebase
- VMware Support Insider:
  http://blogs.vmware.com/kbtv/

- VMware TV:
  http://www.youtube.com/user/vmwaretv
- VMworld TV:
  http://www.youtube.com/user/VMworldTV
- VMware KBTV (external):
  http://www.youtube.com/user/VMwareKB

Acknowledgements

The following individuals contributed to the creation of this white paper:

- **John DiGiglio**, Software Product Marketing, Intel Corporation
- **Danny Lin**, Director of Systems Engineering – GSAO (Global Strategic Alliances Organization), VMware
- **Bhavesh Davda**, Senior Staff Engineer – Office of the CTO, VMware
- **Mike Richer**, Telecom Systems Engineer – GSAO (Global Strategic Alliances Organization), VMware
- **Donna Reineck**, Telecom PLM – vSphere Business Unit, VMware
- **Jon Catanzano**, Senior Technical Writer, Consultant