Using FPGAs to Accelerate Data Centers

achronix
SEMICONDUCTOR CORPORATION

With the technology industry at a crossroads — the effective repeal of Moore's Law and the stagnation and decline of PC, tablet, consumer electronics and smartphone markets — data centers have become the sweet spot of the technology sector, showing healthy revenue growth and attracting new system solutions in both hardware and software. Unlike the ethereal promise of upcoming wonders from AI, robotics and the IoT, data center growth and innovation is happening in the here and now, with an even brighter future ahead the moment other nascent markets emerge from their chrysalis with killer apps of their own.

Evolution of the Data Center

The dynamism of growing markets always braces its participants with a continual stream of daunting challenges. Led by the examples set by the gigantic server farms of Google, Amazon and Facebook, the arrival of the Big Data era and the oceans of information it has spawned - a full 80% of it unstructured — have forced data center designers to move away from highly customized hardware architectures towards commoditized hardware with high port counts, low power consumption and high performance. This approach supports scalability and redundancy while minimizing cost and the need to expand or re-architect the installed plant and its associated ventilation and power infrastructure. In effect, data center technicians want to be able to leave their installation unchanged while simply swapping out boards or blades with other products that offer higher performance, lower power, and feature upgrades.

Yet some levels of customization are still required to maintain flexibility for supporting new applications and user demands. This is precisely what gave Intel the opportunity to branch into a non-PC market segment with tremendous success. x86-based Xeon CPUs have been able to follow Intel's PC-centric semiconductor process technology roadmap into deep submicron, along with all the performance benefits this has entailed. Coupled with the broad software ecosystem support for Xeon, it appeared at first that Intel had provided the data center segment with just enough support for needed software-level personalization to meet individual customization needs while leaving installations safe and secure in their commodity hardware and software architectures that maximized performance and minimized cost.

A New Set of Problems

Yet today, data center architects are beginning to find the adoption of commodity CPU hardware is no longer sufficient for all their customization needs. Taking networking as an example, CPUs serve admirably in performing administrative and protocol tasks for networking control plane applications but lack the performance to support the packet based processing requirements of layers 4 and below.

This is understandable, in that various bit-intensive tasks of the lower OSI layers are far less efficient in execution on a CPU than the word/block-oriented data structures at the transport, network and higher layers. The problem is so severe that data center architects are finding it increasingly difficult to properly control power consumption, maintain performance and architectural scalability for their installed plant. Previously, a custom multi-core CPU array or a processor with associated hardware accelerators might have been called for, but this goes against all the cost and commodity themes of today's data center design precepts.

A New Approach

The new Achronix PCIe Accelerator-6D board is a programmable NIC developed to meet these exact needs at the networking layers L4 and below, supporting custom bit-intensive tasks through a configurable hardware acceleration engine. With four quad small form-factor pluggable (QSFP+) hot-swap transceivers that support 4 × 40G or 16 × 10G Ethernet communications, twelve DDR3 small-outline, dual in-line memory modules (SODIMM) comprising a total 192 GB of memory and 690 Gbps bandwidth, a PCIe Gen3 ×8 pluggable form factor with 64 Gbps throughput and an HD1000 FPGA, the Accelerator-6D board can be used to implement a broad variety of accelerators for data shaping, header analysis, encapsulation, security, network function virtualization (NFV) and test and measurement.

Refining the Data Center Architecture

The Accelerator-6D board offers immediate benefits to communications within a data center itself, improving system power, performance and efficiency by unburdening CPUs of laborious memory accesses and pipeline executions and directly addressing system memory for protocol stack processing and physical layer transactions.

This opens up the possibility for data centers to employ the Accelerator-6D for a broad variety of general DMA applications, via the RDMA over converged Ethernet (ROCE) or iWARP protocols. Currently, low-cost generic NIC cards transfer the processing load for ROCE directly to the system software stack, which negatively impacts system performance and power. A custom NIC solution is out of the question because of cost and scalability concerns.

By implementing ROCE/iWARP on the Accelerator-6D, system CPU overhead can be bypassed for East-West transactions between servers in a data center, while supporting standard networking and tunneling protocols for more conventional North-South communications enabling server applications to access the Internet or other remote resources. Using the Accelerator-6D for such local communications and processing eliminates the need to burden the network software stack and system memory resources.

Improving Network Security

Cloud services providers must provide their users not only with server and network resources capable of high performance, but also with exemplary security. The Accelerator-6D offers system administrators a resource that is decoupled from system CPUs to implement encryption schemes in whatever configurations are desired across the data center.

Further security enhancements are now possible without burdening system CPUs by employing the Accelerator-6D for NFV applications, particularly in network visibility and monitoring. Both physical and virtual taps can be implemented, with applications including traffic management, performance monitoring, load balancing, filtering, firewalls, intrusion detection and data loss prevention. All these applications and more are possible using the Accelerator-6D as a configurable hardware accelerator, reducing CPU loads to improve data center power consumption, performance and efficiency while helping service providers to ensure service-level agreement (SLA) compliance.

Tackling Other Sources of Cost

The network test and measurement market is a specialized niche that directly affects the data center. Their highly customized and expensive equipment is required to test new networking installations. The Accelerator-6D helps test and measurement vendors take a much needed

step towards standardization by allowing test functions to be instantiated in a reconfigurable hardware accelerator. Frame generation, packet and protocol checking and other analysis functions can be implemented using the Accelerator-6D as an agnostic NIC.

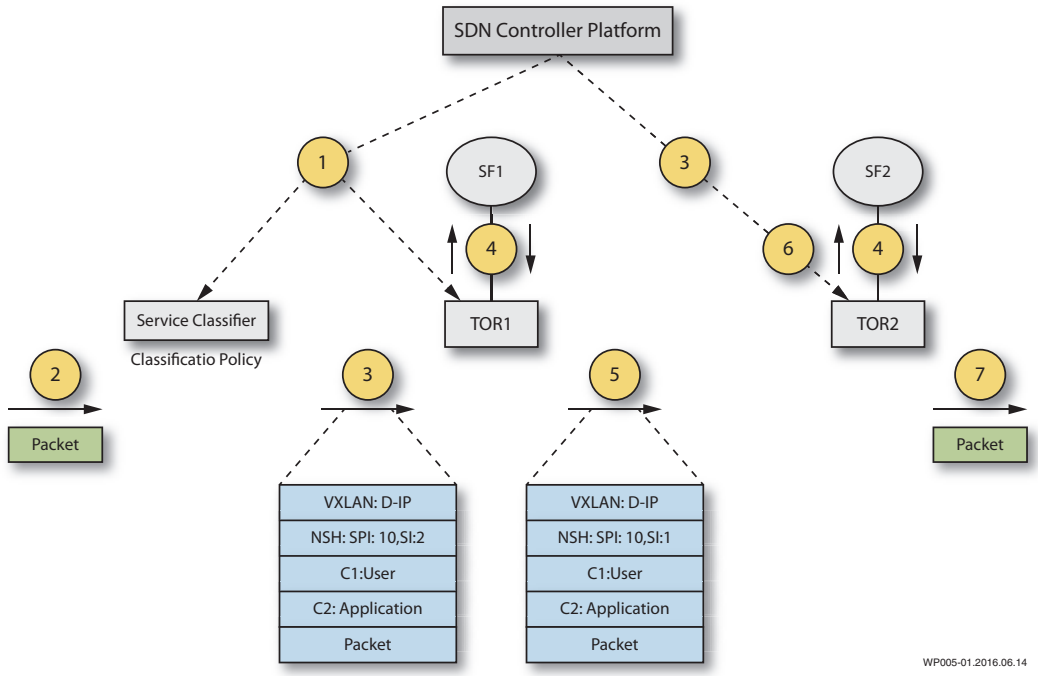

Figure 1: *Open vSwitch (OVS) NSH Service Path Forwarding*

Source: Cisco Systems (http://blogs.cisco.com/data center/nsh

1) A software-defined networking (SDN) controller platform configures the classification policies associated with a given service chain.
2) NSH is added to the packets at the ingress service classifier.
3) Based on the service path identifier (SPI)/service index (SI) combination, the appropriate transport encapsulation delivers the packet.
4) Top-of-rack switch (TOR1) removes the transport encapsulation, and the packet and NSH are delivered to the initial service function (SF1).
5) The packets and NSH return to TOR1, which determines the next service function to apply.
6) TOR2 uses the SPI/SI to determine that the packet should be delivered to SF2.
7) When the packet exits the last service function of the chain, it is forwarded to its initial destination. (Source: IEEE Computer, November, 2014, p. 42).

The Apache OVS is an open source network switching hardware virtualization engine that supports L2, L3 and L4 protocols for multiple networking standards (**Figure 1**). Acting either as a purely virtual switch or as the controlling software for underlying hardware, OVS provides traffic management, QoS and security functions across multiple servers.

Among the many offload operations the Accelerator-6D can perform for the OVS standard is the interface between an OVS-virtualized server environment and the real world. For such an interface to work properly across multiple layers and functions, a data path must be created that integrates a very particular set of services for that path. The network services header (NSH), originally developed by Cisco Systems, offers the necessary protocol for creating this virtual chain of network services. The NSH instantiates a unique network connection that supports multiple, disparate services along its path through the insertion of specific metadata in the packet header. This renders the services for a given packet functionally independent of traffic protocols.

Since almost every such service chain is likely to be exceptional, or *sui generis*, supporting the NHS protocol for SDN-enabled data centers would constitute a significant source of inefficiency for system CPUs and could easily prove to be outside the scope of a conventional ASIC offload engine. Either approach would also break the current data center design paradigm for scalability, generic hardware architecture, cost minimization, power and performance. The Accelerator-6D, however, provides the needed programmability, hardware-based acceleration, system CPU offloading and commodity-like modularity to make OVS and NSH implementations successful.

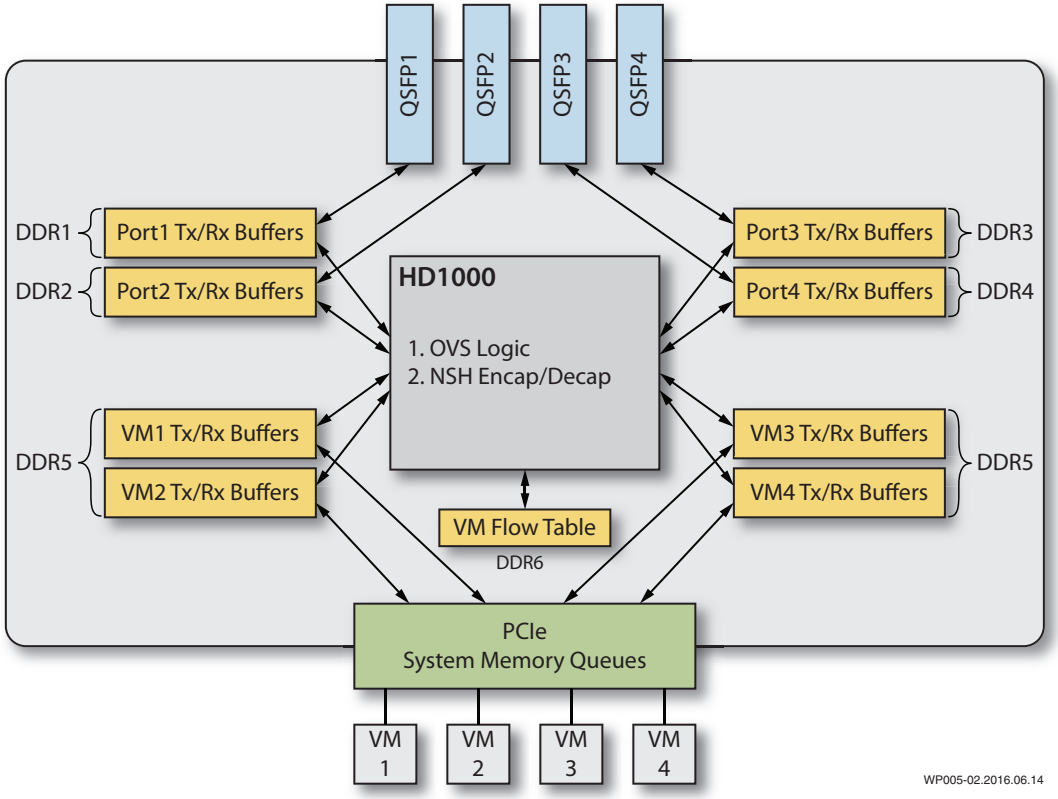

Figure 2: *Example: NFV*

To understand the differentiating value of the Accelerator-6D platform requires a deeper dive into the reference application example (**Figure 2**).

Supporting deployment and architecture flexibility, the Accelerator-6D platform with its 4 QSFP+ transceiver interfaces supports native NIC functionality while in addition supporting advanced capabilities such as redundant port interfaces, intelligent traffic redirect as well as per port dedicated memory bandwidth necessary to support higher levels of data loss-tolerance, leading to improved network utilization. Each 40G QSFP+ port can be allocated its own deep buffer (32 GB) along with over 100 Gbps of DDR bandwidth to enable guaranteed 100% receive packet buffering, eliminating the microburst absorption problem common with other solutions in the market today. With a combination of the abundant internal block memory and an additional two DDR interfaces provides the user with significant flexibility to map the other critical table functions required in the architecture, namely ingress and egress flow lookup tables, the associated information that details what actions to perform on the packets (e.g., destination port, NSH header encapsulation, etc.) as well as the DMA buffers, necessary to support the data transfer to and from the host resident VM queues.

Supporting the real world performance of today's data-center server platforms, the Accelerator-6D provides 64 Gbps of PCIe bandwidth, right sized to support 40 GE NFV Networking applications and high-performance OVS offload. The PCIe interface caps the bandwidth limit for both receive and transmit communications and as such reduces the DDR bandwidth over-subscription challenges when managing full-duplex network interfaces. With a combination of class-leading memory bandwidth and a rich programmable fabric, the Accelerator-6D platform accelerates the delivery of differentiating NFV and OVS solutions.

Leadership in the Data center Market

To be sure, there are other board solutions containing embedded programmable logic for the NIC market. However, none have the memory bandwidth and memory density linked to a large programmable FPGA of the Accelerator-6D.

Further value stems from the HD1000 FPGA at the heart of the board. The HD1000 offers a multitude of hardened cores for memory management and L1/L2 Ethernet functions — six DDR3 controllers, two 10/40/100G Ethernet MACs and two PCI3 Gen 3 controllers. The performance and low-power profile of the embedded hardened IP of the HD1000 free up programmable fabric for customization — an advantage to users that no other FPGA can hope to match. Together with the 12× SODIMM modules, the Accelerator-6D offers high performance, low power, cost and scalability perfectly in line with the design paradigms of today's data centers.

Most significantly, the Accelerator-6D can serve as a proof-of-concept platform for data center administrators to safely explore further optimization potential for installed plant. This new path for innovation will facilitate data center performance and efficiency improvements without creating the problems previously associated with customization of data center architectures.

Revision History

The following table lists the revision history of this document.

Version	Revision
1.0	Initial released version.

Achronix Semiconductor Corporation

2953 Bunker Hill Lane, Suite 101
Santa Clara, CA 95054
USA

Phone : 877.GHZ.FPGA (877.449.3742)
Fax : 408.286.3645
E-mail : info@achronix.com