intel.

Optimizing Data Center Storage workloads with Napatech's IPU Solution

Napatech is the leading supplier of smart network interface card (SmartNIC) and IPU solutions in cloud, enterprise, and telecom data center applications. Through commercial-grade software suites integrated with high-performance hardware, Napatech accelerates network infrastructure, security, and storage workloads to deliver best-in-class system-level performance while maximizing the availability of server compute resources for applications and services.

Enterprise and cloud data centers are increasingly adopting the Non-Volatile Memory Express over Transmission Control Protocol (NVMe*/TCP) storage technology because of its performance, latency, scalability, management, and resource utilization advantages. However, implementing the required storage initiator workloads on the server's host CPU imposes significant computational overheads and limits the number of CPU cores available for running services and applications.

This solution brief explains how an integrated hardware-plus-software solution from Napatech addresses this problem by offloading the storage workloads from the host CPU to an infrastructure processing unit (IPU) while maintaining full software compatibility at the application level.

The solution frees up host CPU cores that would otherwise be consumed by storage functions and delivers significantly higher performance than a software-based implementation. This significantly reduces data center CAPEX, OPEX, and energy consumption. It also introduces security isolation into the system, increasing protection against cyber-attacks, which reduces the likelihood of the data center suffering security breaches and compromising high-value customer data.

NVMe/TCP: The Optimum Storage Technology for Modern Data Centers

NVMe/TCP is a storage technology that allows Non-Volatile Memory Express (NVMe) storage devices to be accessed over a network using standard data center fabrics as show in Figure 1.
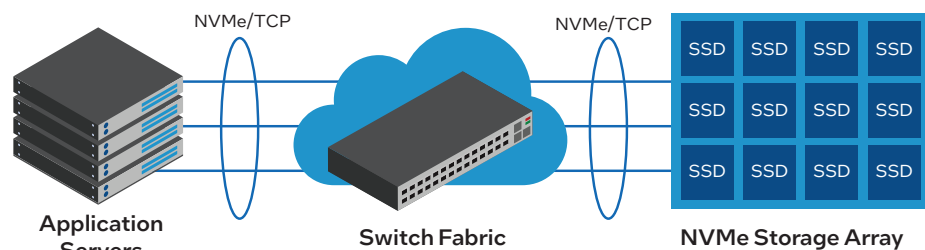
Authors

Charlie Ashton
Senior Director of Business Development
Napatech

Rich Howell
IPU Product Line Manager
Intel Programmable Solutions Group

Figure 1. Disaggregated Storage Architecture

Modern cloud and enterprise data centers are increasingly adopting NVME/TCP due to the compelling advantages that it offers over older storage protocols such as Internet Small Computer System Interface (iSCSI) Fibre Channel:

- Higher performance: NVMe is designed to take full advantage of modern high-speed NAND-based solid-state drives (SSDs) and offers significantly faster data transfer rates than traditional storage protocols. NVMe/TCP extends these benefits to a networked storage environment, allowing data centers to achieve high-performance storage access over the fabric.

- Reduced latency: The low-latency nature of NVMe/TCP is critical for data-intensive applications and real-time workloads. NVMe/TCP can help reduce storage access latencies and improve overall application performance by minimizing the communication overhead and eliminating the need for protocol conversions.

- Scalability: Data centers often deal with large-scale storage deployments, and NVMe/TCP allows seamless scalability by providing a flexible and efficient storage access solution over a network. As the number of NVMe devices grows, data centers can maintain high-performance levels without significant bottlenecks.

- Shared storage pool: NVMe/TCP enables the creation of shared storage pools accessible to multiple servers and applications simultaneously. This shared storage architecture improves resource utilization and simplifies storage management, significantly saving costs.

- Legacy infrastructure compatibility: Data centers often have existing infrastructure built on Ethernet, InfiniBand, or Fibre Channel networks. NVMe/TCP allows them to leverage their current fabric investments while integrating newer NVMe-based storage technology without overhauling the entire network infrastructure.

- Efficient resource utilization: NVMe/TCP enables better resource utilization by reducing the need for dedicated storage resources on each server. Multiple servers can access shared NVMe storage devices over the network, optimizing the use of expensive NVMe storage resources.

- Future-proofing: As data centers evolve and adopt faster storage technologies, NVMe/TCP provides a forward-looking approach to storage access, ensuring that storage networks can keep up with the growing demands of modern applications and workloads.

Overall, NVMe/TCP offers a powerful and flexible storage solution for data centers, enabling high performance, low latency, and efficient resource utilization in a shared and scalable storage environment.

Limitations of Software-only Storage Architectures

Despite the compelling benefits of NVMe/TCP for storage, data center operators must be aware of significant limitations associated with an implementation in which all the required storage initiator services run in software on the host server CPU, see Figure 2.
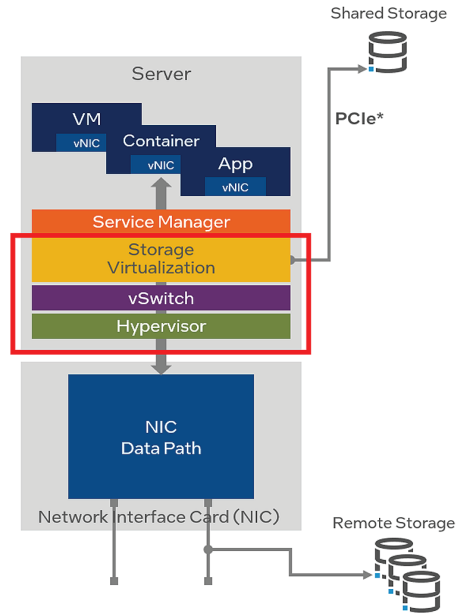

Figure 2. Software based Networked Storage

First, a system-level security risk is presented if the storage virtualization software, the hypervisor, or the virtual switch (vSwitch) is compromised in a cyber-attack.

Second, there is no way to ensure full isolation between tenant workloads. A single architecture hosts multiple customers' applications and data in a multi-tenant environment. The "noisy neighbor" effect occurs when an application or virtual machine (VM) uses the most available resources and degrades system performance for other tenants on the shared infrastructure.

Finally, a significant portion of the host CPU cores is required for running infrastructure services such as the storage virtualization software, the hypervisor, and the vSwitch. This reduces the number of CPU cores that can be monetized for VMs, containers, and applications. Reports indicate that between 30% and 50% of data center CPU resources are typically consumed by infrastructure services.

In a high-performance storage subsystem, the host CPU might be required to run several protocols such as Transmission Control Protocol (TCP), Remote Direct Memory Access over Converged Ethernet (RoCEv2), InfiniBand, and Fibre Channel. When the host CPU is heavily utilized to run these storage protocols and other infrastructure services, the number of CPU cores available for tenant applications is significantly reduced. For example, a 16-core CPU might only deliver the performance of a 10-core CPU.

For these reasons and more, a software-only architecture presents significant business and technical challenges for data center storage.

2

IPU-based Storage Offload

Offloading the NVMe/TCP workload to an IPU, in addition to other infrastructure services such as the hypervisor and vSwitch, addresses the limitations of a software-only implementation and delivers significant benefits to data center operators:

- CPU utilization: NVMe/TCP communication involves encapsulating NVMe commands and data within the TCP transport protocol. The host CPU processes these encapsulation and de-encapsulation tasks without offloading. Offloading these operations to dedicated hardware allows the CPU to focus on other critical tasks, improving overall system performance and CPU utilization.

- Lower latency: Offloading the NVMe/TCP communication tasks to specialized hardware can significantly reduce the latency of processing storage commands. As a result, applications can experience faster response times and better performance when accessing remote NVMe storage devices.

- Efficient data movement: Offloading non-CPU application tasks to discrete hardware accelerators enables data movement operations to be performed more efficiently than using a general-purpose CPU. It can handle large data transfers and buffer management effectively, reducing latencies and improving overall throughput.

- Improved scalability: Offloading NVMe/TCP tasks improves scalability in large-scale storage deployments. By relieving the CPU from handling the network communication, the system can support more concurrent connections and storage devices without becoming CPU-bound.

- Energy efficiency: By offloading certain tasks to dedicated hardware, power consumption on the host CPU can be reduced. This energy efficiency can be especially important in large data center environments where power consumption is a significant consideration.

In addition to the above benefits that apply to the NVMe/TCP storage workload, the IPU-based system architecture provides incremental security isolation options, whereby the infrastructure services are isolated from tenant applications. This ensures that the storage, hypervisor, and vSwitch services are better protected from a cyber-attack launched by a tenant application. The infrastructure services themselves are secured since the boot process of the IPU itself is secure, while the IPU then acts as the root of trust for the host server.
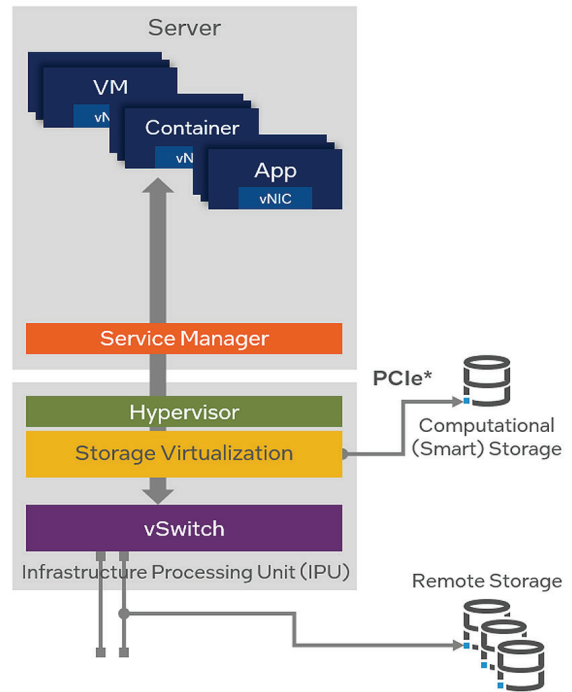

Figure 3. IPU enabled Disaggregated

Napatech Storage Offload Solution

Napatech provides an integrated, system-level solution for data center storage offload, comprising the high-performance Link-Storage* software stack running on the F2070X IPU see Figure 4.
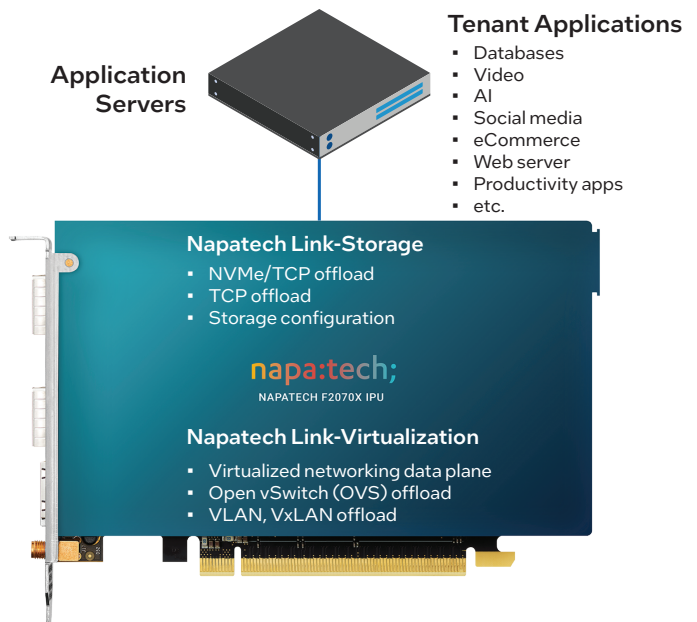

Figure 4. Napatech Storage Offload

3

The Link-Storage software incorporates a rich set of functions, including:

- Full offload of NVMe/TCP workloads from the host to the IPU

- Full offload of TCP workloads from the host to the IPU

- NVMeoF/TCP initiator

- Storage configuration over the Storage Performance Development Kit Remote Procedure Call (SPDK RPC) interface

- Multipath NVMe support

- Presentation of 16 block devices to the host via the virtio-blk interface

- Compatibility with standard virtio-blk drivers in common Linux* distributions

- Security isolation between the host CPU and the IPU, with no network interfaces exposed to the host.

In addition to Link-Storage, the F2070X also supports the Link-Virtualization* software, which provides an offloaded and accelerated virtualized data plane, including functions such as Open vSwitch (OVS), live migration, VM-to-VM mirroring, VLAN/VxLAN encapsulation/decapsulation, Q-in-Q, receive side scaling (RSS) load balancing, link aggregation, and Quality of Service (QoS).

Since the F2070X is based on an FPGA and CPU rather than ASICs, the complete functionality of the platform can be updated after deployment – whether to modify an existing service, add new functions, or fine-tune specific performance parameters. This reprogramming can be performed purely as a software upgrade within the existing server environment without disconnecting, removing, or replacing any hardware.

Napatech F2070X IPU

The Napatech F2070X IPU is a 2x100G PCIe card with an Intel Agilex® 7 FPGA F-Series and an Intel® Xeon® D processor in a full height, half length (FHHL), dual-slot form factor, and based upon the Intel F2000X-PL see Figure 5 below for additional details..

The standard configuration of the F2070X IPU comprises an Intel Agilex® FPGA AGF023 with four banks of 4 GB DDR4 memory and a 2.3 GHz Intel Xeon D-1736 processor with two banks of 8 GB DDR4 memory. In contrast, other configuration options can be delivered to support specific workloads.

The F2070X IPU connects to the host via a PCIe 4.0 x16 (16 GTps) interface, with an additional PCIe 4.0 x16 (16 GTps) interface between the FPGA and the processor.

Two front-panel QSFP28/56 network interfaces support network configurations of

- 2x 100GbE

- 8x 10G or 8x 25G (using breakout cables)

Figure 5. F2000X-PL Block Diagram

A dedicated PTP RJ45 port provides optional time synchronization with an external SMA-F and internal MCX-F connector. IEEE 1588v2 time-stamping is supported.

A dedicated RJ45 Ethernet connector provides board management. Secure FPGA image updates enable new functions to be added or existing features to be updated after the IPU has been deployed.

The processor runs Fedora Linux, with a UEFI BIOS, PXE boot support, full shell access via SSH, and a UART.
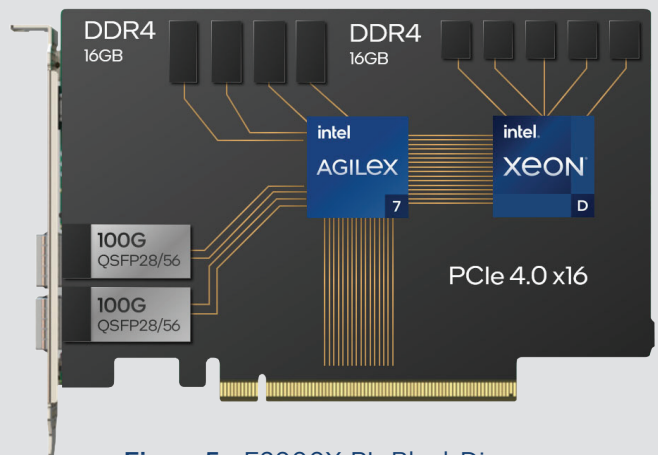
Summary

Enterprise and cloud data centers are increasingly adopting the NVMe/TCP storage technology because of its performance, latency, scalability, management, and resource utilization advantages. However, implementing the required storage initiator workloads on the server's host CPU imposes significant compute overheads and limits the number of CPU cores available for running services and applications.

Napatech's integrated hardware-plus-software solution, comprising the Link-Storage software stack running on the F2070X IPU (Figure 6), addresses this problem by offloading the storage workloads from the host CPU to the IPU while maintaining full software compatibility at the application level.

Napatech's storage offload solution frees up host CPU cores that would otherwise be consumed by storage functions and delivers significantly higher performance than a software-based implementation. This significantly reduces data center CAPEX, OPEX, and energy consumption.

The Napatech solution also introduces security isolation into the system, increasing protection against cyber-attacks, which reduces the likelihood of the data center suffering security breaches and compromising high-value customer data.
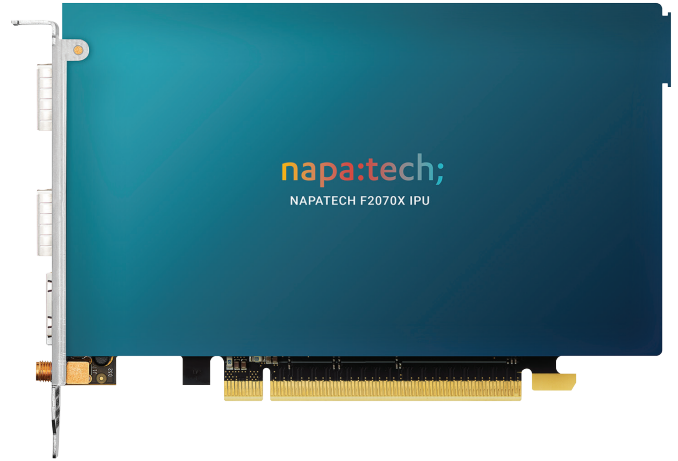
Figure 6. Napatech F2070X IPU

Learn More

- [Learn more about Napatech's storage offload solutions](#)