intel.

Storage virtualization and HW-agnostic acceleration at ease with OPI and Kubernetes

EXECUTIVE SUMMARY

Problem statement

In modern data center environments, managing storage resources efficiently while ensuring performance and scalability is a complex challenge. Traditional storage solutions often lead to vendor lock-in, large integration/automation effort and struggle to keep up with evolving workload demands.

Solution

Intel proposes a local disk emulation approach using a new type of device for infrastructure offload called Infrastructure Processing Unit (IPU). Standard host-side drivers like NVMe or virtio-blk and legacy applications can be enabled to access disaggregated storage at scale using state-of-art protocols like NVMe over TCP while increasing performance through offload of storage services to the IPU. The solution is HW- and vendor-agnostic by leveraging the Open Programmable Infrastructure (OPI) API layer and provides integration with Kubernetes orchestration to facilitate ease of deployment and automation.

Benefits

By adopting this approach, customers can achieve enhanced storage performance, scalability, and flexibility without touching their existing storage applications.

Customers avoid vendor lock-in and build on top of best-of-breed open source and multi-vendor community standards for a holistic IPU/DPU management.

With Kubernetes integration they are given a full stack solution which is easy to deploy and automate based on their existing ecosystem and expertise.

The solution brief presents how a new class of data center devices for infrastructure offload (IPUs) can be used to virtualize and dynamically manage disaggregated storage resources at scale while reducing server-side computational load, improving performance, and avoiding vendor lock-in. The proposed approach is based on an open source, multi-vendor project hosted at the Linux Foundation called Open Programmable Infrastructure (OPI). OPI defines standard APIs for utilizing SmartNICs, DPUs and IPUs. Through the integration with Kubernetes and its Container Storage Interface (CSI) framework customers gain an end-to-end solution that seamlessly integrates storage management and automates volume provisioning for their containerized workloads.

Virtualized disaggregated storage with IPU

An Infrastructure Processing Unit (IPU) **[1]** is a specialized networking device designed to accelerate and manage critical infrastructure functions similarly to a DPU. It features dedicated programmable cores and Ethernet features, offering a host of benefits for both multi-tenant and bare metal cloud infrastructure, as well as for Enterprise and Telco networks.

Some key benefits of an IPU include:

1. **Enhanced Security**: IPUs isolate tenant applications from provider services, creating an extra layer of security. This separation ensures that sensitive infrastructure tasks are kept separate and secure, reducing the risk of security breaches.
2. **Infrastructure Offload**: IPUs offload resource-intensive infrastructure tasks from server CPUs. This optimization not only enhances performance but also frees up server resources and allows CPU cores to be utilized for revenue-earning applications instead of infrastructure management.
3. **Virtual Storage Enablement**: IPUs enable the virtualization of storage resources, remote storage connectivity and acceleration of storage-related computations. They provide data centers with greater flexibility in managing their storage infrastructure which is crucial in adapting to evolving data storage needs.

The Intel® IPU E2100 is a 200 Gbps IPU which enables local storage emulation of NVMe and virtio-blk devices. This means it provides a virtualization layer for remote storage access and makes it indistinguishable from local storage from the perspective of a host. In addition, it offers HW-accelerated data digest, crypto and compression capabilities. Figure 1 illustrates a remote storage access with IPU using NVMe/TCP. In this scenario, the IPU emulates a local NVMe drive connected over PCIe to the host so that standard SW stacks and OSs can

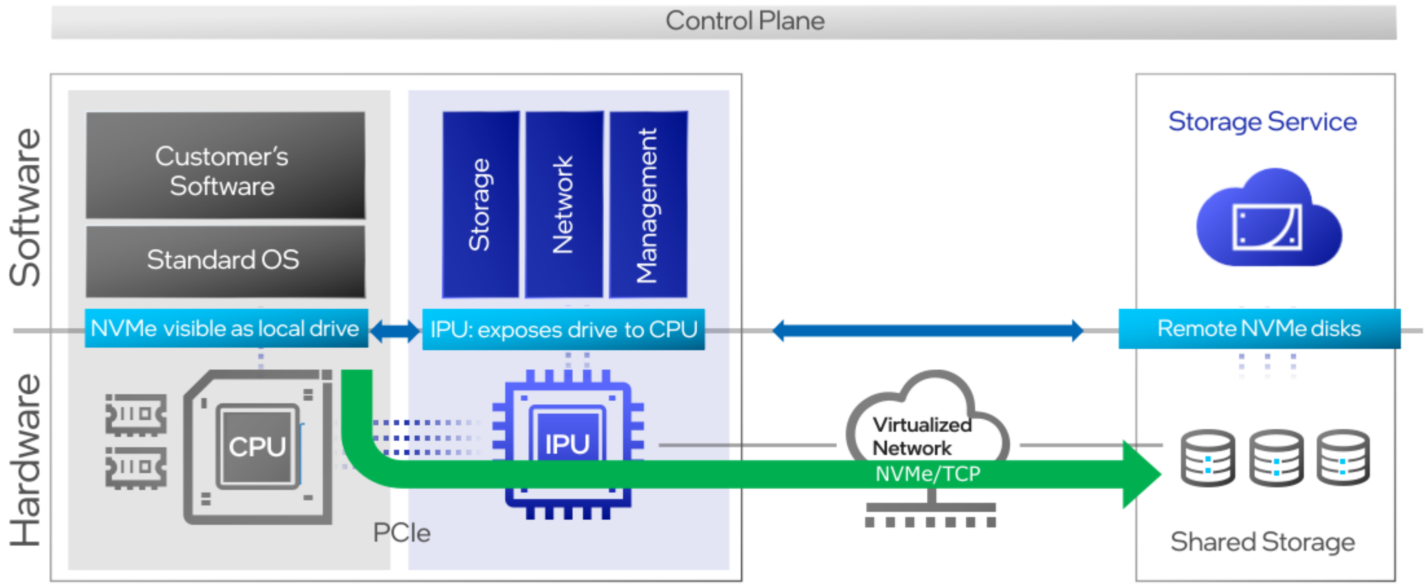

Figure 1 - Local NVMe disk emulation using IPU. Green path indicates the actual data flow to a remote storage appliance.

be used. Host compute resources are then free to run the customer's software since all infrastructure services like Storage, Networking and Management are offloaded to the IPU. IPU-side SW manages the actual over-the-network connection to a remote Storage Service using the NVMe/TCP protocol. A Control Plane needed to set up the scenario is indicated on the top.

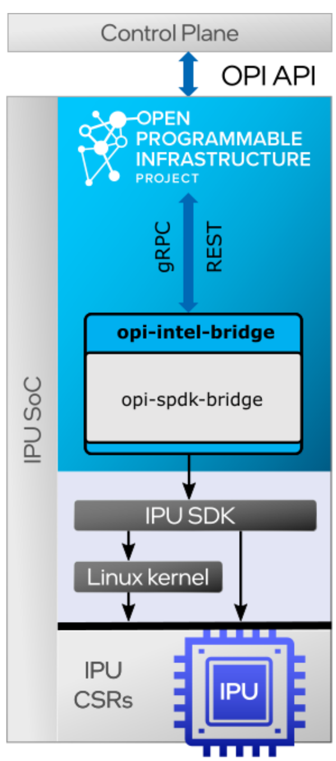

Figure 2 – OPI API implementation with OPI gRPC to IPU SDK bridge (opi-intel-bridge).

Open Programmable Infrastructure

As an interface to the Control Plane, a vendor-agnostic Open Programmable Infrastructure (OPI) API is used. OPI is a Linux Foundation open-source project developing standard APIs for utilizing SmartNICs, DPUs and IPUs, and other coprocessors or processing elements. It allows users to provision and orchestrate any vendor's device in a uniform manner and allows manufacturers to create standard APIs faster as well as benefit from a larger ecosystem [2]. The OPI Storage API is a modern gRPC-based API with simultaneous support for RESTful requests through grpc-gateway [3]. OPI relies on the concept of vendor-opi-bridges to provide OPI API integration with a vendor's SDK. Figure 2 visualizes this idea for Intel OPI integration, i.e. the opi-intel-bridge project. The implementation runs as a service on IPU SoC and enables IPU HW configuration and usage under OPI API control. At the same time, it makes a substantial reuse of another OPI project – opi-spdk-bridge which bridges the OPI API to an SPDK library [4] and thus provides a standalone SW target that can be run on a general-purpose CPU for experimentation and rapid prototyping purposes at no HW cost.

Kubernetes CSI integration

When used as the Control Plane, Kubernetes serves as a powerful orchestrator for this solution. With Kubernetes Container Storage Interface (CSI) [5] customers gain an end-to-end solution that seamlessly integrates and automates storage management with an IPU. Kubernetes automates volume provisioning, ensuring that resources are allocated efficiently, and eliminates the need for customers to develop their own automation scripts. This simplifies the deployment and scaling of storage resources and enhances overall

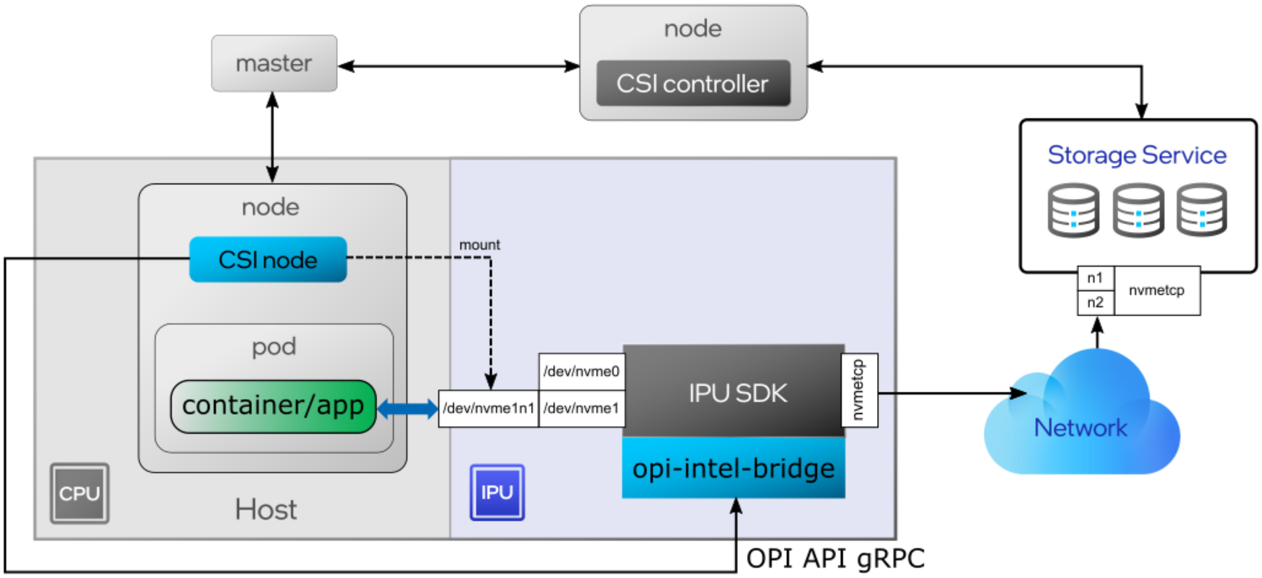

Figure 3 – Kubernetes-managed containerized app accessing locally emulated NVMe drive using IPU and NVMe/TCP connection to a remote storage service. A vendor-specific CSI controller plugin exposes the actual storage over network, while a vendor-agnostic CSI node plugin uses OPI API to configure the IPU for an NVMe drive exposition to the host and making it accessible by the workload.

integration and automation capabilities. Moreover, since Kubernetes is one of the leading container orchestration (CO) systems, there is a lot of existing in-house expertise and production-ready deployments available to make the transition to a HW-accelerated infrastructure as easy as possible. Figure 3 illustrates how a Kubernetes cluster is run together with OPI and IPU for making remote storage available to a containerized workload. A CSI controller plugin specific to a particular storage appliance is used to make storage available over the network. In the diagram two NVMe namespaces are exposed over TCP. On the host side, a CSI node plugin based on spdk-csi [6] is used to configure the IPU using the vendor-agnostic OPI API to connect to the remote storage and expose it to the host over PCIe as a standard NVMe drive, where it can be mounted to the pod and consumed by the workload which is run on it.

Contributors: Antti Kervinen, Artsiom Koltun, Yadong Li, Michael Lynch, Filip Szufnarowski, Jing Yan, Intel Corporation

SEO keywords: Kubernetes CSI, Disaggregated Storage, Storage Acceleration, Storage Virtualization. NVMe, virtio-blk, IPU, DPU, SmartNIC, Cloud Orchestration, OPI, Open Programmable Infrastructure

References:

[1] Intel, "IPU Based Cloud Infrastructure White Paper," [Online]. Available: https://www.intel.com/content/dam/www/central-libraries/us/en/documents/ipu-based-cloud-infrastructure-white-paper.pdf.

[2] Linux Foundation, "Open ecosystems for next generation architectures and frameworks," [Online]. Available: https://opiproject.org/.

[3] grpc-ecosystem, "gRPC to JSON proxy generator following the gRPC HTTP spec," [Online]. Available: https://github.com/grpc-ecosystem/grpc-gateway.

[4] spdk, "Build Ultra High-Performance Storage Applications with the Storage Performance Development Kit," [Online]. Available: https://spdk.io/.

[5] kubernetes-csi, "Kubernetes Container Storage Interface (CSI) Documentation," [Online]. Available: https://kubernetes-csi.github.io/docs/.

[6] spdk-csi, "CSI driver to bring SPDK to Kubernetes storage through NVMe-oF or iSCSI," [Online]. Available: https://github.com/spdk/spdk-csi.