LA-UR-21-27657
Approved for public release; distribution is unlimited.

Title: Exploring OpenSNAPI Use Cases and Evolving Requirements

Author(s): Williams, Brody Kyle
Poole, Stephen Wayne
Poole, Wendy Kinton

Intended for: HPC Intern Symposium 2021

Issued: 2021-08-02 (Draft)

Title: Exploring OpenSNAPI Use Cases and Evolving Requirements

Abstract:

Emerging system architectures are rapidly transforming in order to meet shifting requirements.
Motivated by expanding data volumes, energy efficiency concerns, and the omnipresent need to improve performance, architectures are increasingly adopting a data-centric approach.
At the core of this concept is the goal of minimizing data motion and instead processing data in-situ to the greatest degree possible.
Therefore, data-centric designs, in contrast to conventional CPU-centric models, typically distribute compute capabilities throughout the architecture.
As part of this paradigm shift, a novel class of devices known as data processing units (DPUs), alongside CPUs and GPUs, are quickly forming a third pillar of data-centric systems.
These devices, which include smart network adapters and switches, seek to offload computation on data at the network edge as well as in-flight within the network fabric.

The Open Smart Network API (OpenSNAPI) project seeks to develop a unified API for DPU devices.
In our previous talks, we introduced the OpenSNAPI project and detailed our investigations regarding the viability of offloading compute intensive kernels to BlueField DPUs.
In contrast, in this talk we detail our efforts to offload application level file I/O to the DPU.
We also discuss plans and early efforts to explore in-network compute capabilities.
Finally, we describe our observations with respect to the evolving design of OpenSNAPI.