Prodapt

Smart Server Network Solution

for GenAI workloads

SmartNIC & SmartSwitch-powered
Network accelerator platform to
reliably scale GenAI workloads

RDMA | Lossless Networking algorithms | Deterministic Latency

Author: Madhavendra (Shaan) Bhatnagar

Prodapt GenAI Network Accelerator Paper No.1

Contents

Introduction

Graphics Processor Units (GPUs) have become central to parallel computing, positioning them as the core processors for GenAI applications. These processors are purpose-built compute nodes driving image processing with AI inference, making them best suited for handling large volumes of simple computations.

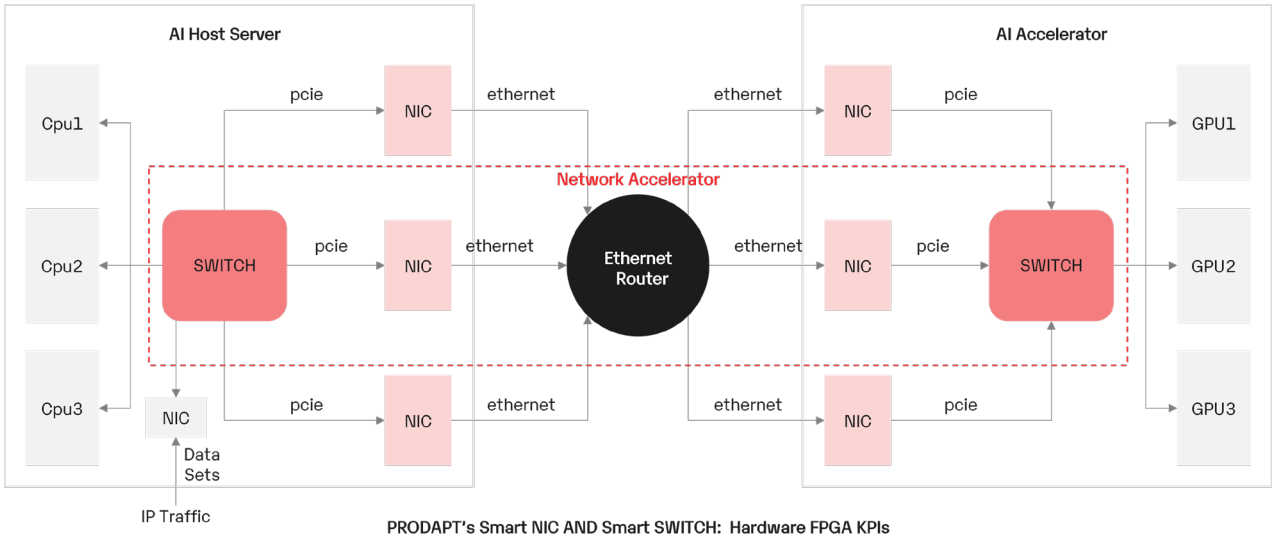
They drive the AI Acceleration fabric and network architecture using three industry standard interconnects: PCIexpress, Ethernet, InfiniBand, and two critical network fabric components: SmartSwitch for inter GPU collective communication and SmartNIC for routing AI Datasets to GPU nodes.

However, challenges remain in realizing lossless networking algorithms to power GenAI workloads at scale!

Prodapt's differentiator capabilities

Prodapt delivers the acceleration through RDMA, deterministic latency, and lossless networking algorithms driven thorough SmartSwitch and SmartNIC. The GenAI Hardware Fabric diagram below illustrates the NIC and Switch in fabric. High performing AI Clouds have 3 main subsystems: Host Servers, Routers (bridge between Host servers and AI Accelerators) and the AI Accelerators.

Gen AI Hardware Fabric



PRODAPT's Smart NIC AND Smart SWITCH: Hardware FPGA KPIs

Driving Factor for Network Accelerators

GPU memory subsystems, with their limited capacity, are struggling to manage the complexity of large models. This renders GPUs incapable of meeting training requirements.

For example, models with 100 billion parameters or more are spread across multiple GPUs to meet the current training requirements. An A800, 40 GB GPU from NVIDIA can accommodate 1-2 billion parameters based on computation requirements for training. A model with 100 billion parameters, as mentioned above, would need 50 GPUs for training and storage – and this does not include AI inference computational transitions.

Collective Communication and Computation across multiple GPUs or nodes is the only solution for distributed training. Prodapt's network platform is the backbone of CCC (Collective Communication and Computation), deploying state-of-the-art, industry-standard, lossless networking algorithms driven by Remote Direct Memory access (RDMA), Explicit Congestion Notification (ECN), Priority flow Control (PFC) and Approximate Fair drop (AFD). All these algorithms built into programmable state machines of NIC and Switch drive the Network Acceleration, meeting the latency requirements for collective communication of GPU Clusters.

Network acceleration plays an instrumental role in distributed computation nodes. In a distributed system, the network is categorized into single card, multi-card and multi-machine interconnections. GPU interconnections fall under the multiple cards category and PCIexpress is the most optimal interconnect technology for high bandwidth communication networks such as inter GPU clusters.

New classes of Data Center: AI Clouds and AI Factories depend upon accelerated computing and accelerated networking. Performance of Accelerated computing nodes such as GPUs is gated by huge nondeterministic network pipeline latency and lossy UDP traffic.

Bottomline: It's the network that is rearchitecting the data center platform towards ushering in the era of GenAI revolution and is the backbone of AI Clouds for distributed AI model training.

Prodapt's SmartNIC FPGA Solution

Prodapt's solution drives resilience and lossless networking capabilities across the NIC (Network Interface controller), besides promoting smart actions to de-congest layers and maximizing compute power.

Unique value proposition

- Remote direct memory access over Ethernet and PCIe interconnect between GPU clusters.
- Secured multi-tenancy guaranteeing no noisy neighbor problem between adjacent tenant streams.
- Programmable Congestion Control
- Lossless networking and connection tracking.
- Single root I/O Virtualization (SRIOV)

Prodapt's SmartSwitch FPGA Solution

The Prodapt Smart Switch enables load/store/forward of data over high speed PCIe or CXL (Compute express link), supports multiple interconnects to perform traffic routing – but with a high degree of intelligence and resilience.

Unique Value Proposition

- Real time Link failover for improved network resiliency
- Configurable fabric manager to support varying topology
- Low latency RDM protocols.
- Lossless networking and connection tracking.
- Single root I/O Virtualization (SRIOV)

POC (Proof of Concept)

Prodapt proposes to use the Synopsys HAPS platform (4 FPGA board) to develop and validate Network Acceleration algorithms. The platform offers extreme flexibility for developers to build a wide range of applications, from multi-user small designs to larger designs ones cobbling together multiple systems.

The HAPS architecture is common across all platforms, enabling users to mix the platforms to assemble any sized prototype.

HAPS-100 1 FPGA

Desktop form factor in a portable platform. Best debug performance in the desktop category for the industry.

Desktop form factor

HAPS-100 4 FPGA

Offers full flexibility for a range of applications, from small designs with multiple users sharing the system to larger designs built up by cabling together multiple systems. Our most popular model is the software workhouse in the industry.

Desktop or enterprise form factor

HAPS-100 12 FPGA

High capacity enterprise solution. Optimized for 1 BG+ designs and suitable for data-center installations.

Enterprise form factor