



SMARTNICS TERMS AND DEFINITIONS

A brief, authoritative guide to the acronyms and terms
used in the rapidly emerging field of SmartNICs

Available online at
<https://smartnicssummit.com/terminology/>

SmartNICs Summit Technical Advisory Board
info@SmartNICsSummit.com

SmartNIC Definitions

A **SmartNIC** is a generic term for a programmable accelerator that makes data center networking, security and storage efficient and flexible. It is a NIC that has its own local processing power enabling it to perform tasks independently of the Host CPU. Typical tasks include packet processing, cybersecurity, video processing, and other data processing and analytics. They often consist of a general purpose processor cores, possibly in combination with an FPGA, application-specific or programmable processors, or a special-purpose logic.

Hardware Terminology - Processors

APU (Accelerated Processing Unit) is a single device designed to have both integrated CPU and GPU functionalities within one package. APUs combine the CPU and GPU onto a single chip to form a combined processing unit that reduces space and cost.

Bare-metal – refers to a form of cloud services in which the user rents a physical machine from a provider that is not shared with any other tenants. Traditional cloud computing models are based on virtual machines, bare metal servers do not come with a hypervisor preinstalled. This environment gives the tenant complete control over their server infrastructure.

DPU (Data Processing Unit) is a single ASIC containing an industry-standard CPU with a network interface capable of parsing, processing and transferring data at line rate. It usually also has a set of programmable acceleration engines that offload and improve performance for applications such as AI/ML, security, telecommunications and storage.

FPGA (Field-Programmable Gate Array) is an integrated circuit designed to be configured by a designer or user after manufacturing – hence the term field-programmable. The FPGA configuration is generally specified using a hardware description language (HDL), similar to that used when designing an application-specific integrated circuit (ASIC). Some SmartNIC designs use them for configurable special-purpose accelerators.

GPU (graphics processing unit) – a programmable device specially designed to implement graphical functions such as high-speed rendering, also used in other mathematically intensive applications such as simulation, high-performance computing, image processing, military/defense systems, computational physics, and financial analytics.

IPU (Infrastructure Processing Unit) is an advanced networking device with hardened accelerators and Ethernet connectivity that accelerates and manages infrastructure functions using tightly coupled, dedicated, programmable cores. An IPU offers full infrastructure offload and provides an extra layer of security by serving as a control point of the host for running infrastructure applications.

xPU (*generic* Processing Unit) is a generic term that describes a class of processing for SmartNICs that may include (but not limited to), CPU, GPU, APU, IPU, DPU, etc. It encompasses both general-

and specific-purpose processors that are separate from host processors and/or switching/storage processors, regardless of the applied function or use case. It is also the SNIA-approved term for processors of this type.

Software/Network Frameworks Terminology

Container – a combination of program code with everything needed to run it, typically including a runtime package, system tools, system libraries, and settings. Containers decouple applications from the underlying host infrastructure, allowing for easy deployment on different clouds and operating system (OS) environments. However, containers do not contain an OS, yet require an OS to run.

DOCA (Data Center-on-a-Chip Architecture) is an SDK and a runtime environment for programming NVIDIA DPUs. The SDK provides industry-standard open APIs and frameworks, including Data Plane Development Kit (DPDK) for networking and security and the Storage Performance Development Kit (SPDK) for storage. The frameworks simplify application offload with integrated acceleration packages.

eBPF (extended Berkeley Packet Filter) is a Linux kernel technology that allows programs to run within the kernel without requiring changes in the kernel source code or adding modules. It provides direct access to kernel facilities for new capabilities or programs that must run at higher speed than is available through standard kernel calls from the user space.

Fabric – A method of network and transport inter-connectivity that aggregates the control plane for each node. In this manner, each node participates and contributes to the holistic operation and establishes inter-dependencies for operation. A fabric may be built from several different transport networks.

Kubernetes (pronounced “koo-br·neh·teez”) - an open-source container-management system for automating application deployment. It was originally designed by Google and is now maintained by the Cloud Native Computing Foundation (CNCF).

Micro-services – a set of programs, or an architectural model that uses a set of programs, that perform basic system tasks that are usually very low-level, loosely coupled, and independently deployable. Micro-services are designed to be highly maintainable and testable, and organized around business capabilities.

Network Function Virtualization (NFV) – a method for implementing network services, such as routers, firewalls, and load balancers, as tasks that can be run on any compute platform in a network. Operations can thus be performed on Commercial-off-the-shelf (COTS) servers, rather than proprietary hardware. NFV is the platform virtualization, while SDN is the network virtualization.

Network Operating System (NOS) – an operating system designed specifically for networked computers that contain a variety of elements and allow access from multiple points, sharing of

resources, and cooperation in performing applications. In the general context, NOS is the OS of a networking equipment, which coordinates packet processing according to the protocols running on it. In the SDN context, a NOS is typically referred to as an SDN controller.

P4 is a language for describing packet processing by general-purpose CPUs, GPUs, FPGAs, and programmable ASICs. It is sometimes used for programming SmartNICs. P4 serves such applications as adaptive routing, software-defined networking, deep packet inspection, media processing, and network monitoring. The P4 community offers a language specification, a set of open-source tools, and sample P4 programs.

ROCm (Radeon Open Compute platform) is an open software platform developed by AMD for HPC and AI/ML. It provides access to deployment tools, libraries, compilers, programming models (such as the Heterogeneous Interface for Portability – HIP) and drivers/runtimes for AMD GPUs.

Software-Defined Networking (SDN) – a network architecture where the control plane is separated from the forwarding data plane. The control plane can be centralized, controlling multiple devices' data planes in the network.

