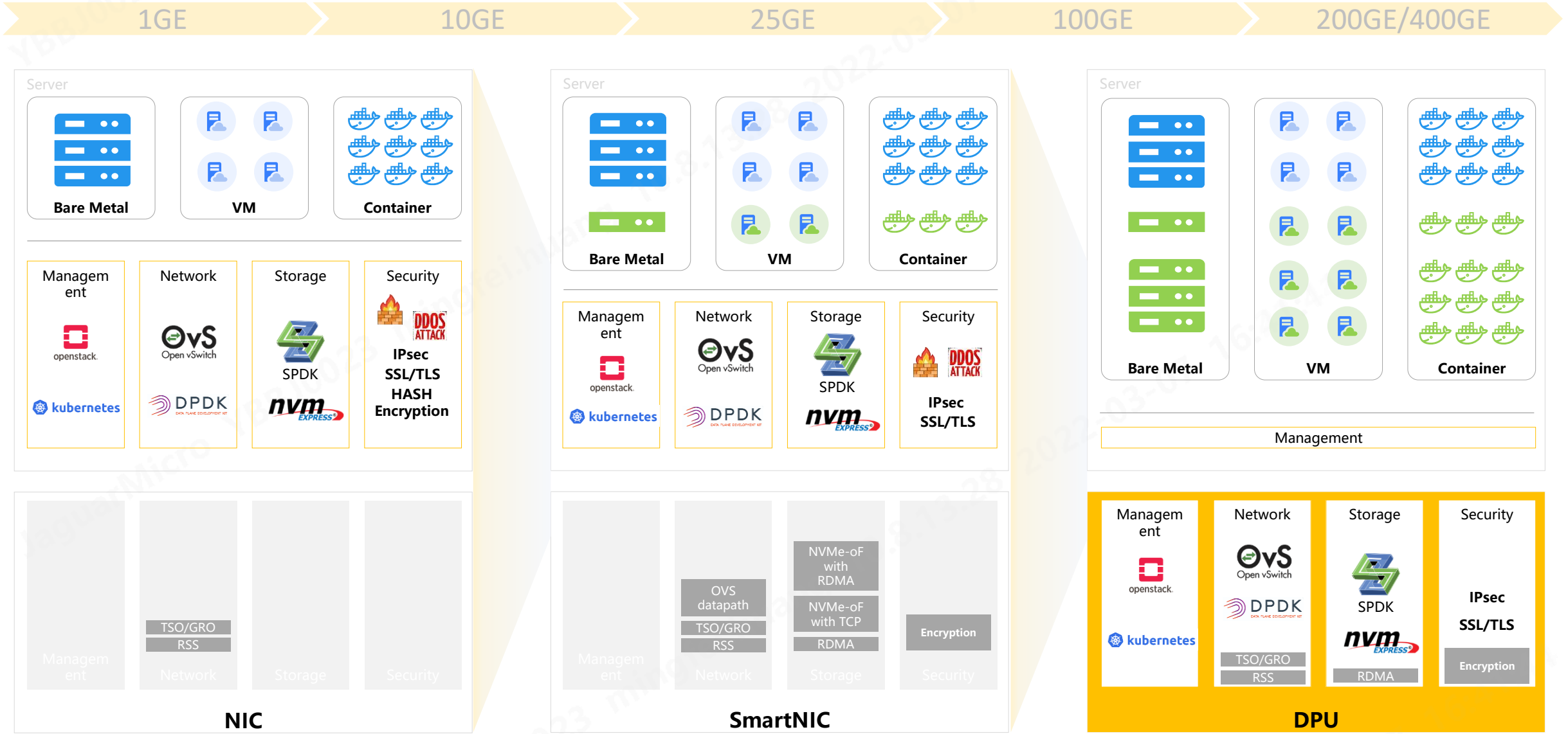
Improving Data Center Infrastructure by JaguarMicro DPU

Mingfei Huang

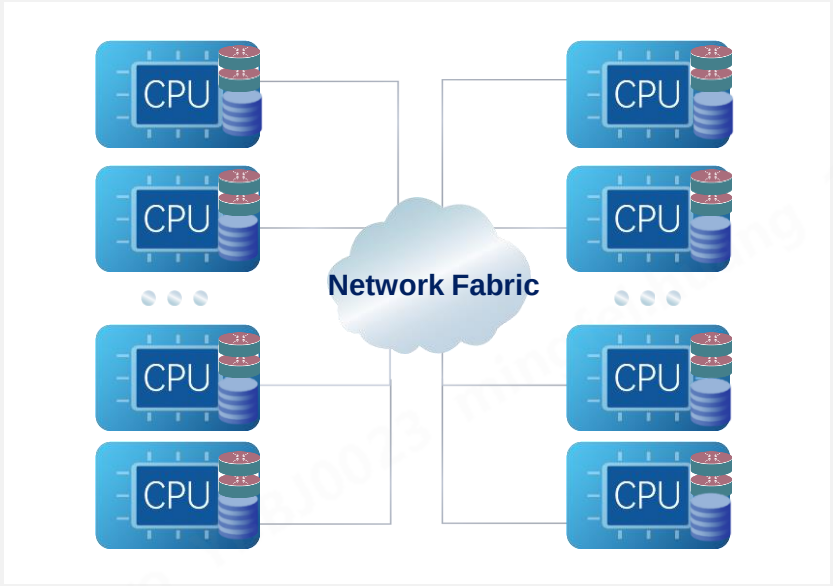
Mar. 2022

Why DPU

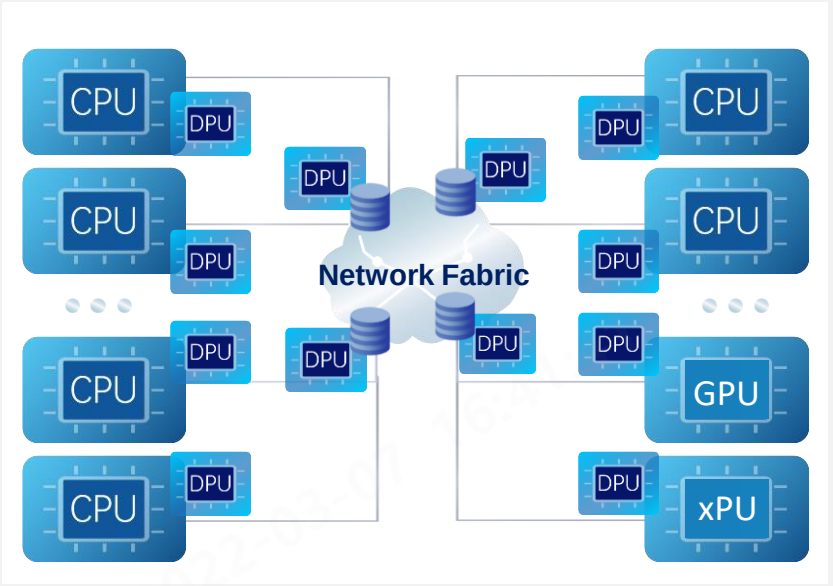


Data Center – shifting from CPU-centric to DPU-centric architecture

CPU-centric Data Center



DPU-centric Data Center



- As demand for big data and network bandwidth explodes, data center can't rely only on CPU to handle all workloads.
- All future data center infrastructure workloads (network, storage, virtualization, management) will be handled by data processing units (DPUs).

DPU becoming the "controller" of next-gen data center

Deliver China's first high-performance cloud native DPU SOC and solutions

Jaguar Microsystems

China's leader in Data Processing Unit (DPU) and Solutions

A leading technology company developing a new generation of DPUs and advanced silicon solutions for modern data centers. Founded by Dr. Sunny Siu and a core team each with over 20 years of experience in the semiconductor industry, the company has gathered a world-class silicon and software team formerly from Broadcom, Intel, Arm, HiSilicon (Huawei), and Alibaba.

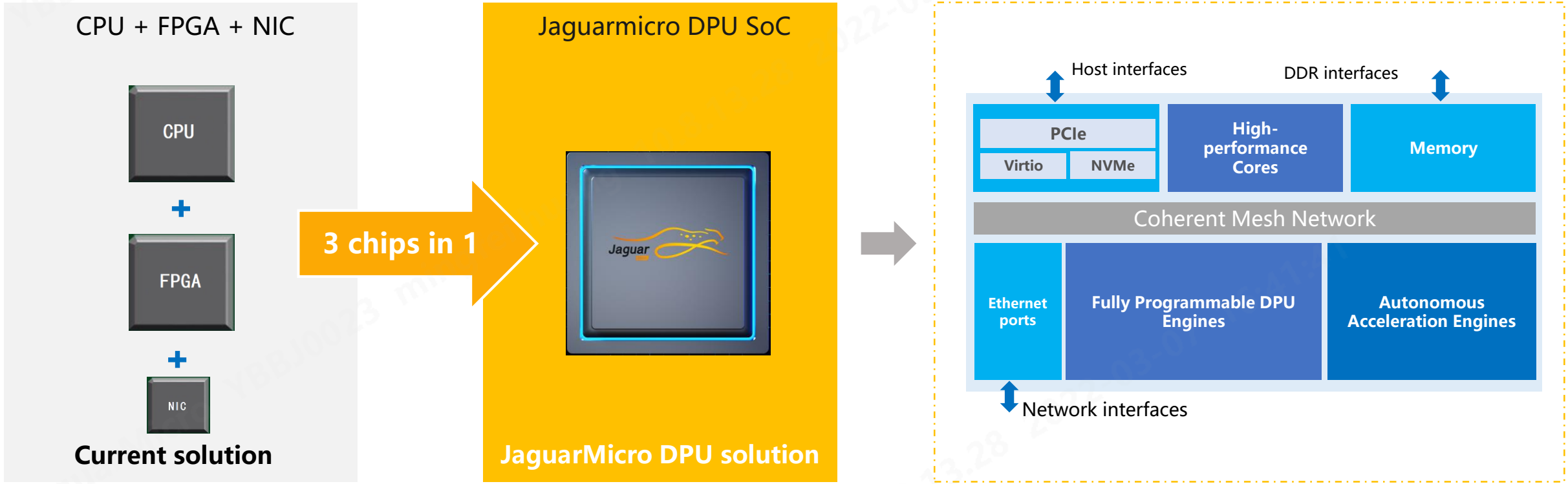
Major investors

Tencent 腾讯 · SEQUOIA CAPITAL 红杉资本 | CHINA · NIO Capital · 中芯聚源 China Fortune-Tech Capital · 五源资本 · SCGC 深创投 · 弘卓资本 HUNDREDS CAPITAL · 耀途资本 Glory Ventures · ZWC 众为资本 · 正心谷资本 Loyal Valley Capital

Challenges in Data Center

1	**Unified Management**	How to unify the management of VM, Container and Bare Metal services
2	**Hot Migration**	Offloading Network, Storage, Management workload from Host CPU to DPU makes VM Migration more complex
3	**Elastic Cluster**	How to support elastic cluster resource management for VM, Container and Bare Metal services
4	**Flexibility**	How to achieve programmability with performance acceleration
5	**Compatibility**	How to ensure minimal disruption to existing systems

Jaguarmicro DPU combines 3 chips 'functionalities into an SoC



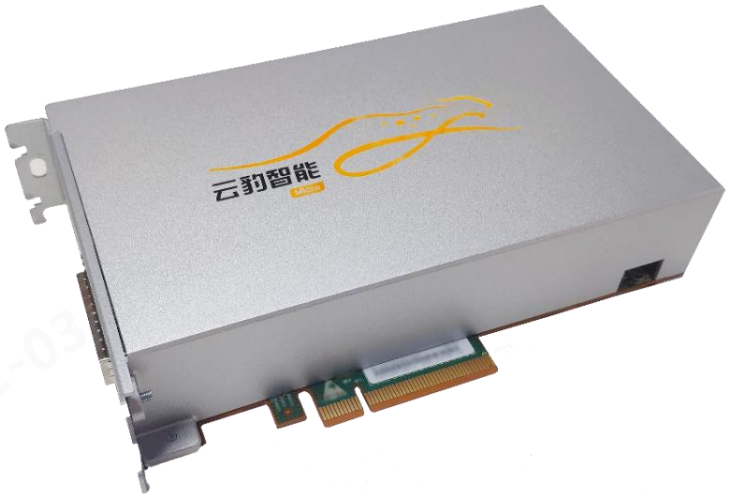
- 3 chips solution, very expensive
- Power exceeding 150W
- Cannot reach 2*100Gbps+.

- Single chip solution reducing cost & power each by over 50%
- Multi-fold increase in performance.

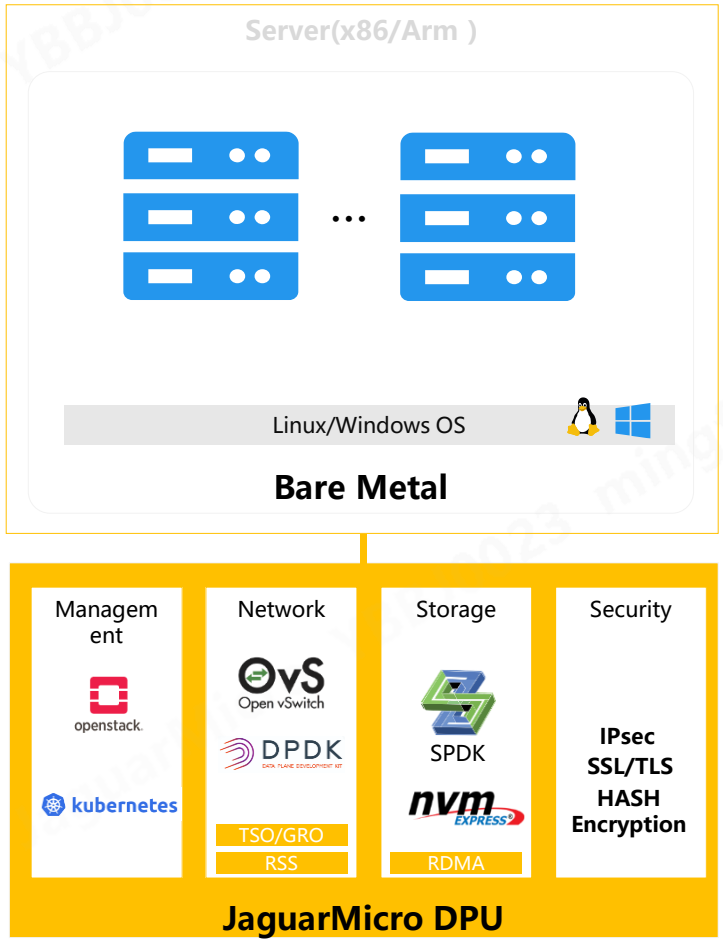
FPGA-based JaguarMicro DPU solution already in production

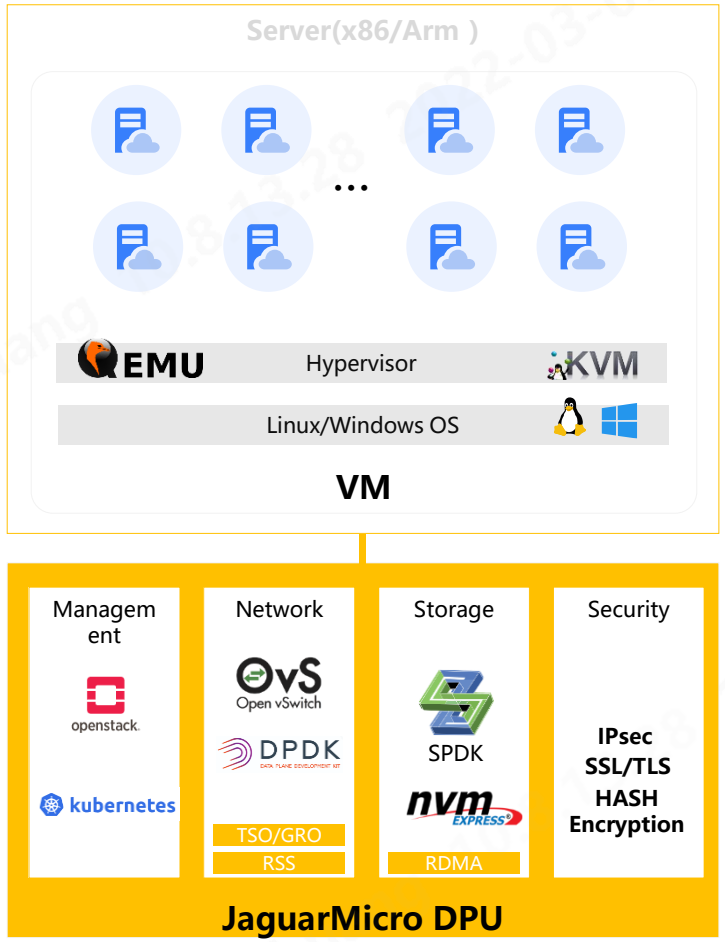
Features

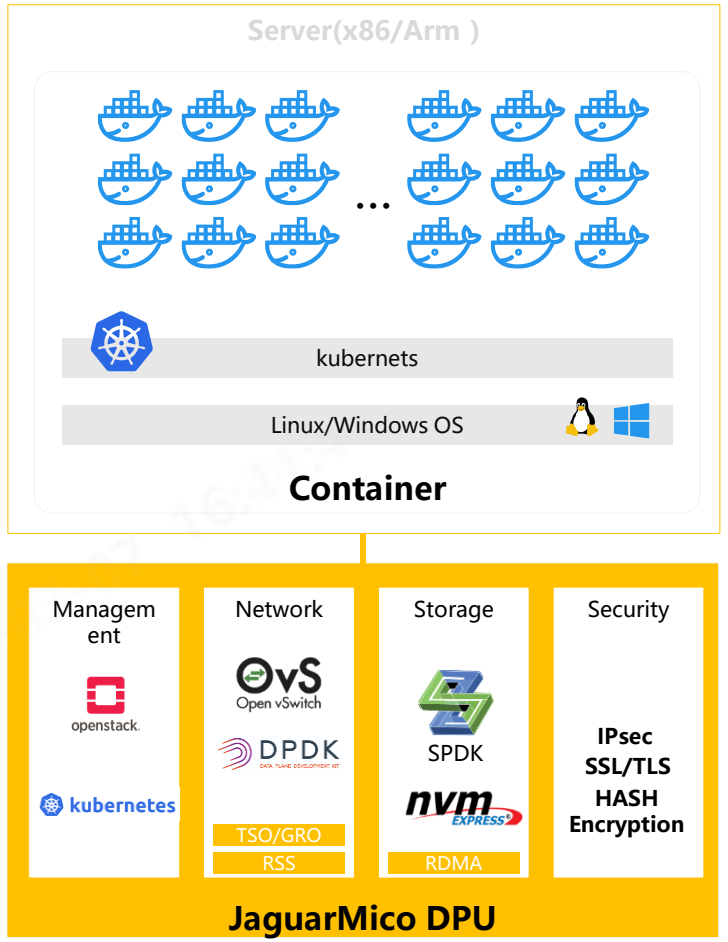
- ✓ Unified support of Bare-metal, VMs and Containers services
- ✓ 2x25G network support and high-performance storage
- ✓ Enable elastic and on-demand virtual network and storage interfaces
- ✓ Support hot migration, hot upgrade, and hot pluggable
- ✓ Allow smooth migration from standard NIC card to FPGA-based DPU



Bare-metal, VM & Container service support

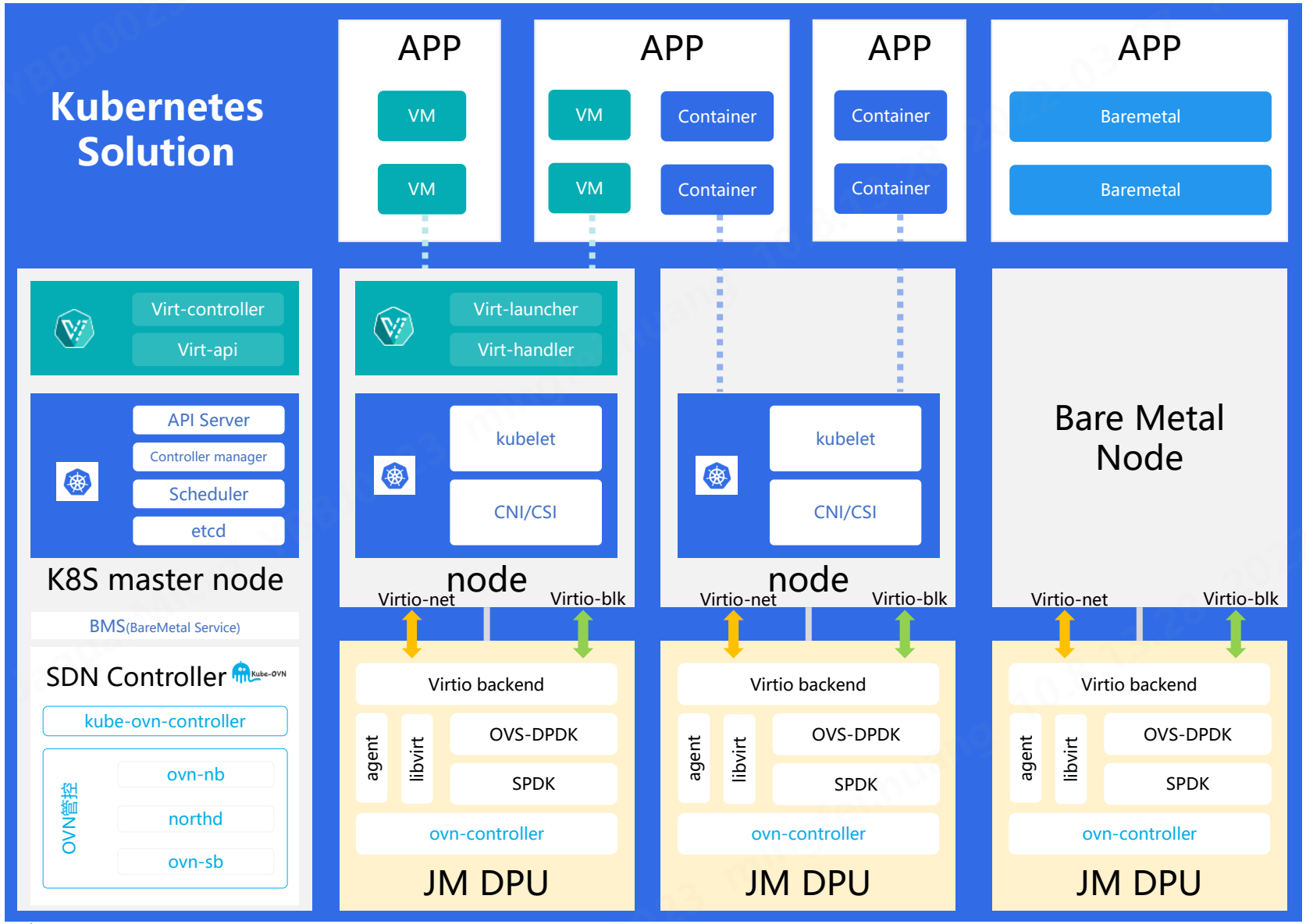






Unified Management and Single Pool for BM, VM and Container

Kebernetes Solution



Bare Metal

- Kube-apiserver add and configure virtio-net and virtio-blk devices to bare-metal host.

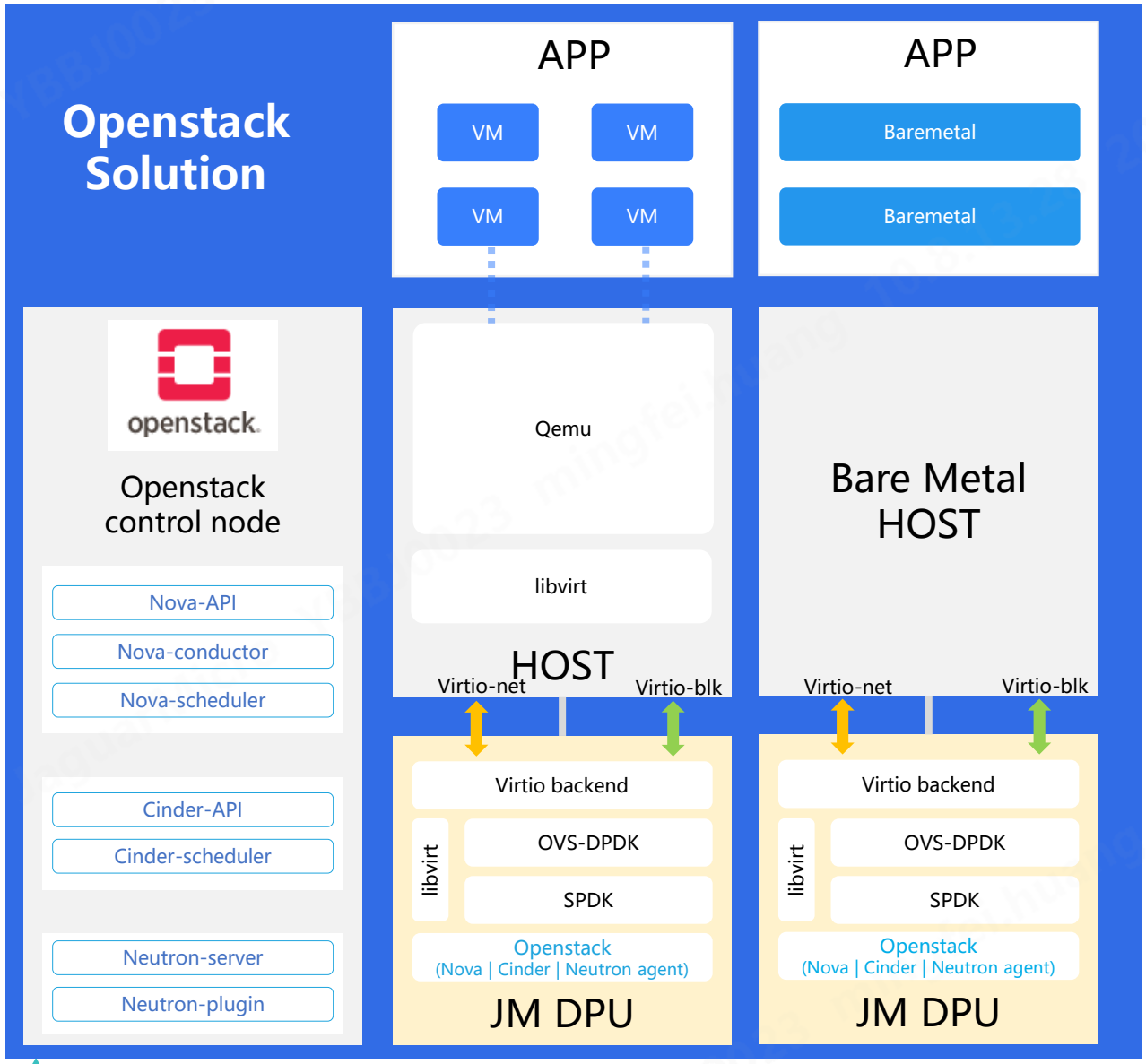
Container

- Kube-apiserver create | delete | manage pods through kubelet on work node.
- Kubelet talks with DPU through CNI/CSI plugin to add and configure virtio-net and virtio-blk devices.

Virtual Machine

- Kube-apiserver create | delete | manage VMs through kubelet & Virt-handler & Virt-lancher on work node.
- Kubelet talks with DPU through CNI/CSI plugin to add and configure virtio-net and virtio-blk devices.

Openstack Solution



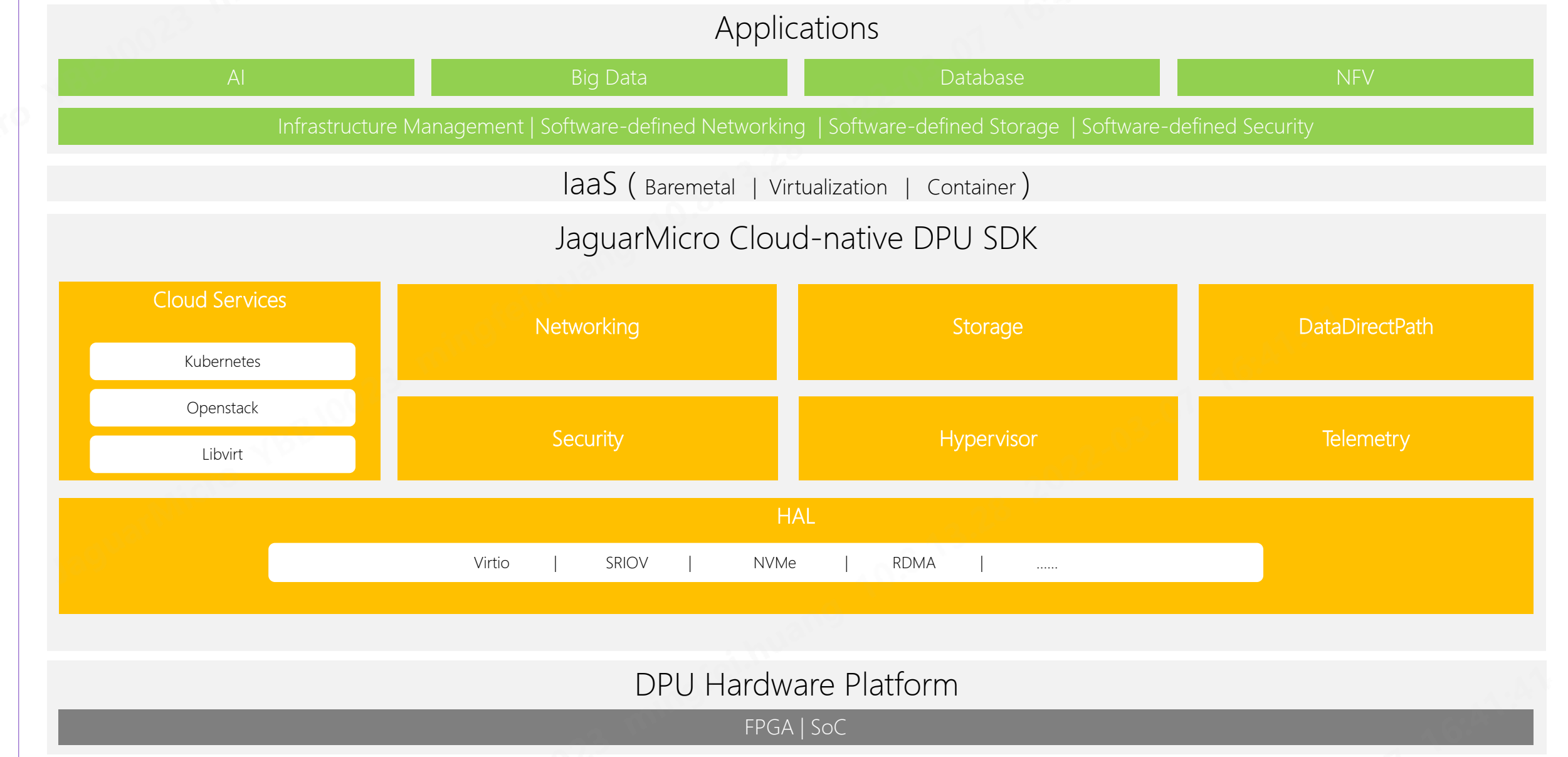
Bare Metal

- By libvirt on DPU, Nova, Cinder and Neutron agents add or configure virtio-net or virtio-blk devices.

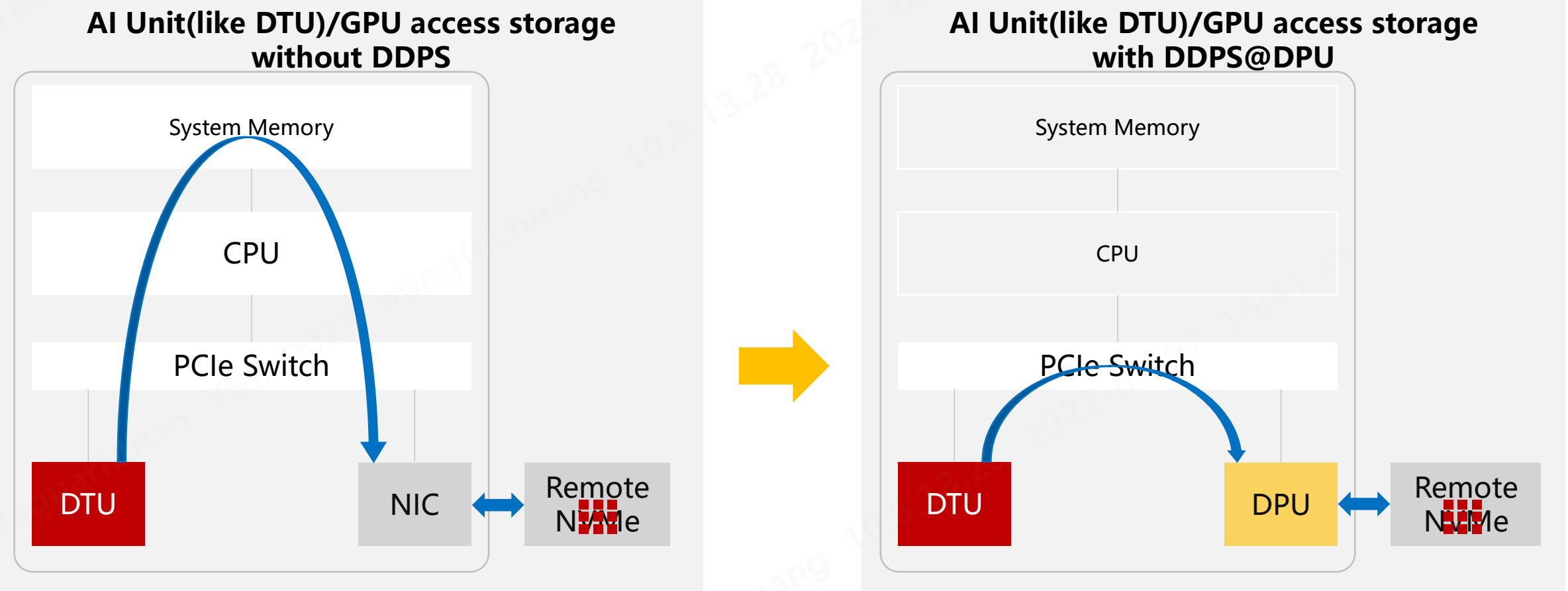
VM

- By libvirt on DPU, Nova, Cinder and Neutron agents add or configure virtio-net or virtio-blk devices.

Software Architecture



JaguarMicro DataDirectPath Storage



On DPU with DDPS enabled, AI Unit (like DTU) or GPU can access remote storage bypass Host, getting lower latency and higher performance.

Highlights

Network

- Hardware Accelerator
- Compatibility
- C programming

Storage

- Hardware offload
- Compatibility
- DataDirectPath Storage|RDMA

Security

- Zero Trust
- Hardware Accelerator

Management

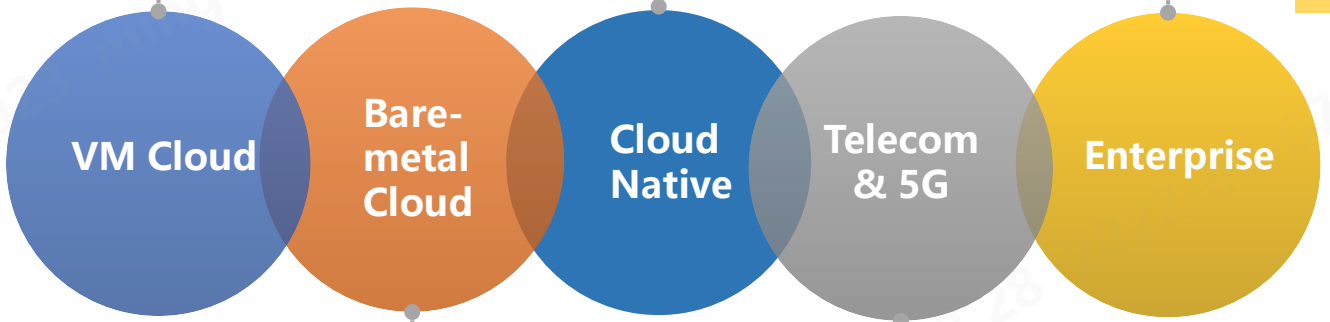
- Unified Infrastructure Mgmt
- VM Migration
- Hot Upgrade
- Diskless Boot
- Hot Plugging

Addressable Markets

Offloading infrastructure workloads from server CPUs

Support multi-level SLA for cloud-native applications

Private cloud and Enterprise applications (edge computing, control plane etc)



Support and unify management of Bare-metal cloud service

Support NFV for telecom and 5G operators

THANKS!

Contact us: info@jaguarmicro.com