大禹智芯

云计算第三引擎

1. 公司简介

北京大禹智芯科技有限公司（以下简称为大禹智芯）成立于 2020 年，是一家专注于提供 DPU （数据处理单元）产品设计、研发与服务的高新科技企业。

大禹智芯通过现有和自主研发的先进制成芯片，自研高性能 IaaS 组件和针对特定协议的加速能力，提供包括芯片、硬件产品、系统软件、应用集成等一整套围绕 DPU 的软硬件产品及服务，致力于打造新一代的云计算引擎，助力用户构建领先的 IT 基础设施 ，加速企业数字化转型步伐。

公司愿景是“以大禹之道释放高速网络的应用潜力，让企业有选择地享受技术红利，助力企业又好又快地搭建领先的 IT 基础设施”。

1.1. 团队介绍

1.1.1 创始人团队：

李爽 首席执行官

- 美团云总经理
- 阿里巴巴集团网络部总监
- 百度系统技术委员会主席

高亚滨 首席运营官

- 思科 XaaS 与云协作业务大中华区总经理
- 阿里巴巴全球技术战略合作总监

王昕溥 首席技术官

- 美团云技术总监
- 阿里云和蚂蚁金服网络产品研发负责人
- 百度 CDN 平台研发负责人

Patrick 首席科学家

- 10 年以上芯片研发经历，硅谷一线芯片公司担任芯片架构师
- 先后任全球两家头部云计算公司智能网卡软硬件团队负责人

1.1.2 团队核心人员：

大禹智芯拥有布局完整、分工明确、合作默契的软硬件技术精英团队，总部位于北京，在上海、西安、南京和杭州设有研发办公室。

核心团队成员来自于国内外互联网/云计算头部公司及传统网络/芯片/安全头部厂商，拥有十年以上云计算平台设计、研发和运营的经验，具有扎实的技术功底及丰富的行业经验，对于云计算平台的使用场景及基础设施的搭建有着清晰的理解和丰富的实践经验；同时，大禹智芯团队成员彼此合作十余年，具备深刻理解技术发展趋势、用户需求和业务场景的综合能力。

1.2. DPU 技术的发展历程

DPU 相关技术概念起源于国内外公有云头部厂商，随后逐渐被市场认可，如今被公认为继 CPU、GPU 后云计算的第三引擎。

2016 年起，随着基础设施规模不断扩大、网络带宽不断提高、数据迁移量和频度的变化，底层数据处理任务（包括网络，存储，安全）所消耗的算力逐渐增加，这导致有效算力在总体算力中的占比逐渐下降。在此背景下，业界开始探索一种高效的解决方案，以实现在满足云计算业务要求的基础上，降低基础数据处理任务消耗的算力，从而释放和提高可用算力。在经过约 3 年的探索后，DPU 概念在 2019 年孕育而生。目前，DPU 已在公有云头部厂商实现大规模部署，其成功应用充分证明 DPU 类产品可极大提高云计算场景下的基础设施使用效率，大幅降低虚拟化计算、存储资源的单位成本。

大禹智芯创始人团队希望通过自身多年在云计算行业的技术积累和对 DPU 类产品的充分理解，向广泛市场推广 DPU 技术，让这一目前仅为少数头部公有云厂商掌握的领先技术惠及各行各业。在未来，大禹智芯将以 DPU 相关产品和技术在云计算场景应用的丰富实践经验为基础，在其他领域发掘更多 DPU 应用场景。

1.3. 围绕 DPU 的新型云计算数据中心基础架构

在现有数据中心基础架构中，除以 GPU 代表的异构计算由 GPU 来处理外，其余基础数据处理任务（包括网络、存储、安全以及通用计算类任务等）均由 CPU 处理（如图 1 所示）。从“物尽其用”的角度出发，CPU 应专注于处理通用计算类任务，而不应该承载除此之外的

其他任务。

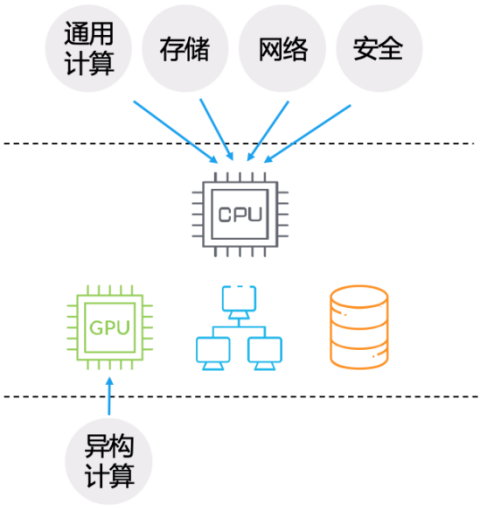

图 1

DPU 的出现分担 CPU 在传统基础架构中所承载的基础数据处理任务，专门负责网络、存储和安全相关的处理任务，进而大大解放 CPU 的计算能力，使 CPU 能够最大化承载其本身最擅长的通用计算类任务（如图 2 所示）。

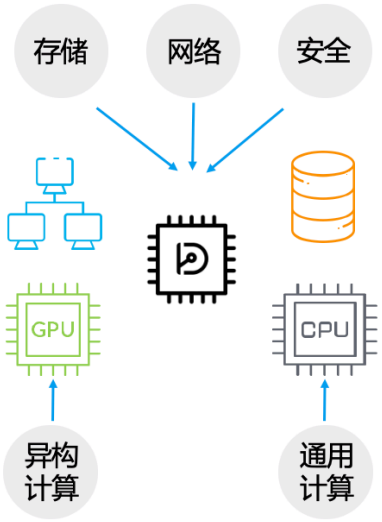

图 2

目前，整个云计算数据中心基础设施架构正从以 CPU 为中心向以 DPU 为中心逐步演

进。该演进过程将不断提升数据中心的整体处理效率。届时，CPU、GPU、DPU 将发挥“各司其职”、“物尽其用”的作用，均专注于承担自己擅长的处理任务。当中，DPU 作为 CPU、GPU，存储介质与数据中心网络间的桥梁，为各个组件实现数据的高速流转。因此，该架构变化也可解释为从以算力为中心的计算架构向以数据通信为中心的计算架构的演进。

DPU 为实现其功能，一端连接网络，另一端与各服务器组件互联，在数据中心基础设施中则以一张近似“网卡”的形态呈现。(如下图 3 所示)。

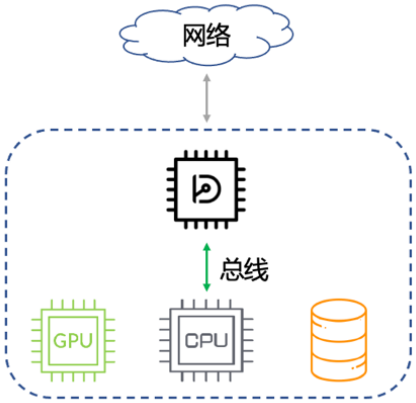

图 3

网卡在整个计算机系统里面存在了很久，其自身也经历了技术革新。从功能角度出发，网卡基本上可以分为三类，第一类为普通网卡，第二类为特殊功能网卡，第三类为 DPU 网卡，他们之间的功能对比（参考表格 1)。

	普通网卡	特殊功能网卡	DPU 网卡
处理单元	网卡芯片 ASIC，不可编程	有限编程能力的网卡芯片 ASIC，无额外处理器单元	完全可编程，集成有处理器单元，由 SoC/FPGA 构成
主要功能	网络转发，Checksum Offload，GRO/GSO 等	普通网卡功能，特殊功能的卸载，例如 RDMA/RoCEv2，特定网络卸载能力，如 OVS fast-path offload	不仅能够实现数据面的卸载，同时能够运行控制面程序，对转发面和控制面均有极强的可编程能力，能够实现定制化功能

CPU 消耗*	高	中	低或趋于 0

表格 1

*注释：指 CPU 处理网络，存储，安全流量所需要消耗的资源

由于 DPU 网卡具备普通网卡与特殊功能网卡没有的控制平面/转发平面的可编程能力，其核心价值体现在如下三方面：

1、实现原先无法实现的功能：典型案例就是云化裸金属

2、达到渴望的性能：利用 DPU 实现卸载和加速，用更高效的处理方式处理

3、综合性价比：同等云计算算力能力下，DPU 可大幅降低物理服务器数量，进而减小对机架空间，网络端口等需求，降低综合成本。

3. 大禹智芯 DPU Paratus 系列

3.1. Paratus 1.0

- ✓ 端口：4x10G/25G + 1x MGMT RJ45
- ✓ 处理器单元：16x ARM Core 2.0GHz
- ✓ 处理器内存：最大 64GB
- ✓ PCIe：3.0 x8
- ✓ 尺寸：FH¾L，单宽

3.2. Paratus 2.0

- ✓ 端口：4x/25G or 2x100G + 2x MGMT RJ45
- ✓ 处理器单元：16x ARM Core 2.0GHz + FPGA
- ✓ 处理器内存：ARM 最大 32GB，FPGA 最大 64GB
- ✓ PCIe：3.0 x16
- ✓ 尺寸：FHHL，双宽

5. 大禹智芯 DPU 的核心功能与场景

大禹智芯团队有多年云计算场景的实际开发部署经验，充分理解云计算场景下功能需求以及如何实现。利用我们的技术优势，大禹智芯 DPU 产品从功能上充分覆盖云计算场景中的网络，存储及安全各功能点，具体如下：

1、网络功能

- 云计算虚拟网络功能的全卸载
- 数据中心内高性能网络的实现

2、存储

- 云计算高性能解耦存储资源池
- 高性能网络存储目标端（Target）
- 高性能网络存储起始端（Initiator）

3、安全

- 虚拟化场景下网络安全
- 第三方安全能力部署平台
- 为第三方安全能力提供流量编排

大禹智芯完整 DPU 解决方案（即对芯片、硬件产品、系统软件、应用的集成）可通过架构变革实现云计算基础设施的“减耗增效”，帮助客户实现创新型云计算产品的落地，构建创新型网络/信息/数据安全相关的处理平台，为“国产可替代”做强有力的功能与能力补充。

7. 大禹智芯 DPU 核心功能实现

7.1. 基于大禹智芯 DPU 的云计算虚拟网络功能实现

虚拟化网络能力是云计算中的一个重要基础，不同租户之间的网络隔离，各租户不同的网络服务及服务质量等级，均依赖虚拟化网络来实现。对虚拟化网络的处理，往往会占用大量的 CPU 计算资源，对虚拟化网络处理的卸载能够得到非常明显的收益。大禹智芯 DPU 可完全代替 CPU 来处理虚拟化网络的全部功能。图 4 展示了普通网卡/DPDK 普通网卡，特殊功能网卡/智能网卡和大禹智芯 DPU 网卡在虚拟化网络中以 OVS（Open Virtual Switch）为例的不同处理方式。大禹智芯 DPU 网卡可以实现 OVS 的全卸载，把原先有宿主机侧 CPU 负责处理的 OVS 控制平面（CP）和数据平面（DP）转移到 DPU 上处理，从而解放 CPU 计算资源。当 OVS 下放到 DPU 后，由 DPU 负责虚拟化网络的控制及转发。宿主机侧的全部计算资源均能够以虚拟计算资源的方式对外输出。

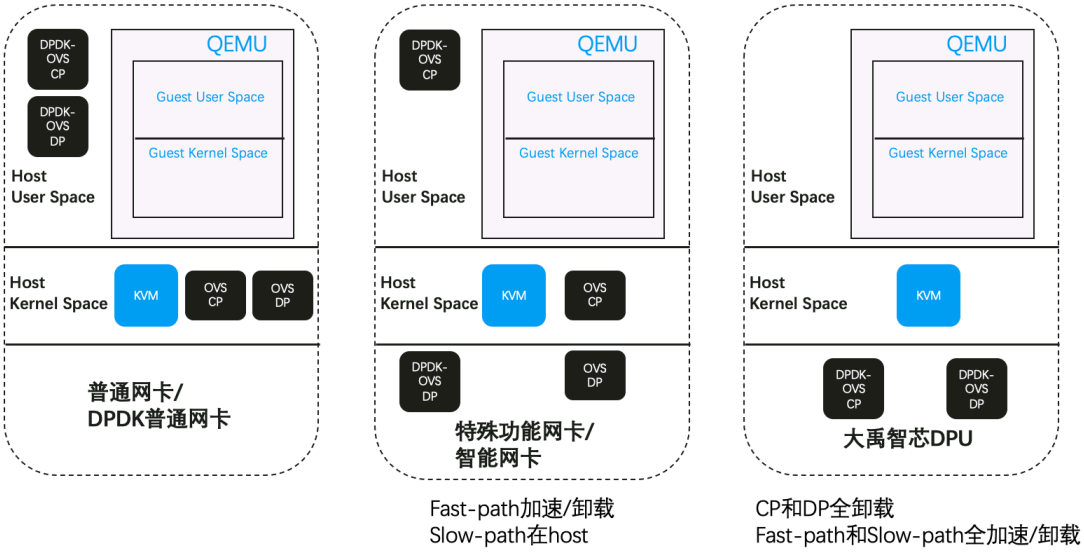

图 4

大禹智芯DPU与虚拟机之间的互动遵守主流开源虚拟化网络框架，使用VirtIO或DPDK-vDPA 的方式向虚拟机提供虚拟网络功能，实现方式如下图 5 所示。

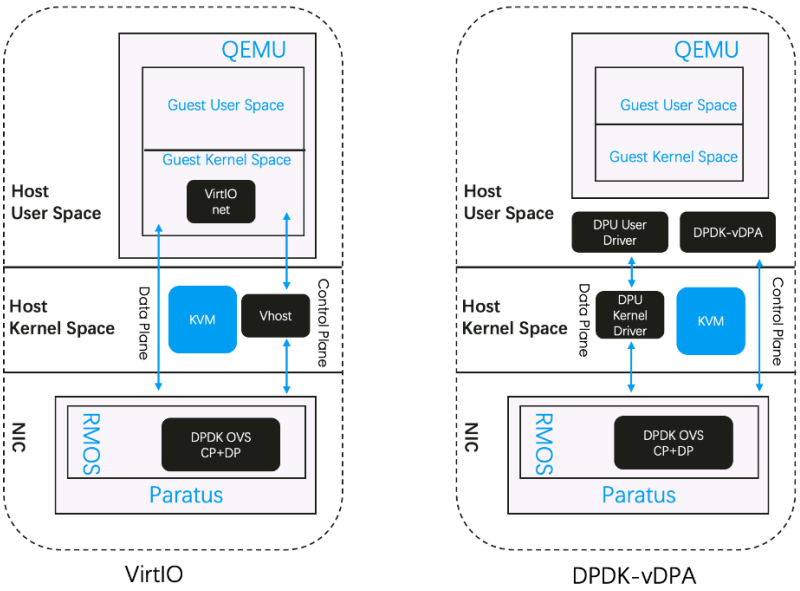

图 5

7.2. 基于大禹智芯 DPU 的云计算存储能力实现

在云计算场景和企业网场景中，存储应用所占比例逐渐提高。社会整体数据产生量呈现爆发性增长，基于海量数据的新型应用孕育而生。对数据的存储、移动、分析、应用，占据了大部分数据中心内部可用网络带宽。构建一个高效的存储系统对于云计算场景和企业网场景均十分重要。数据的流转是 DPU 专门负责且其擅长的工作，围绕 DPU 构建的新型存储服务基础架构正在逐渐被市场所接受。

这里谈到的存储，不是传统意义上的存储，而是基于数据中心网络互联，由软件定义的存储协议栈构建的分布式可扩展存储。此类存储提供了充足的可扩展空间，与云计算虚拟化有更好的耦合性，并且从架构上实现更高的稳定性和可靠性，这是未来存储领域的发展发向。此类存储从功能上可分为块存储、文件存储和对象存储等。块存储作为所有存储类型的基础，从使用场景上又为目标端（Target）和起始端（Initiator），Target 提供块存储资源池的功能并通过数据中心网络提供块存储服务，Initiator 为使用端，通过网络获取 Target 提供的块存储（如下图 6 所示）。

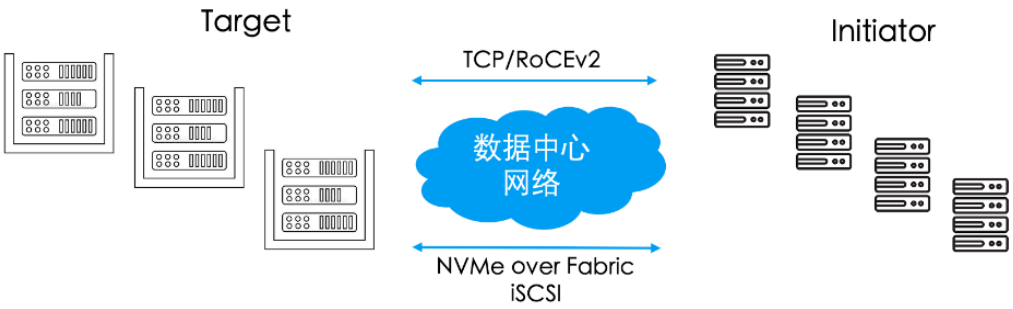

图 6

如果把文件存储和对象存储看作是构建在块存储上面的存储类应用，那么存储类应用不仅需要底层块存储支撑，同时还需要计算资源的参与以实现高级存储逻辑。基于大禹智芯 DPU，可以构建一个全新的存储架构（如下图 7 所示）。此架构充分利用 DPU 的能力实现“解耦的存算分离”。使用 DPU 实现独立于传统 CPU 计算资源的高性能网络块存储服务与资源池化，依赖 CPU 计算资源的文件存储和对象存储则可以专注在上层存储逻辑而不用负责底层块存储，也无需使用本地存储；计算节点按需使用块存储或文件/对象存储，而无需使用本地存储。

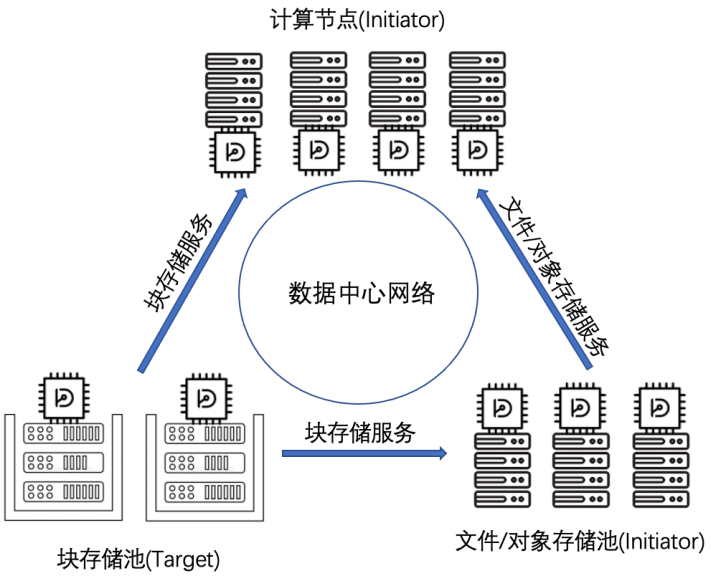

图 7

大禹智芯 DPU 在各个场景下的使用方式和功能具体如下：

7.2.1. 块存储节点（Target）

如下图 8 所示，与现有使用 CPU 来实现网络块存储不同，DPU 作为 PCIe RC 直接管理 NVMe ，替代了 CPU+网卡+CPU 内存的工作，让块存储软件栈直接运行在 DPU 上。大禹智芯 DPU 支持 DPDK 及 SPDK，并且提供底层驱动，方便用户把自己基于 SPDK 开发的块存储软件栈迁移到 DPU 上运行。DPU 同时提供压缩/解压缩，加密/解加密的加速能力，可以实现比 CPU 更高效的存储数据服务。典型的物理形态是 DPU+JBOF。

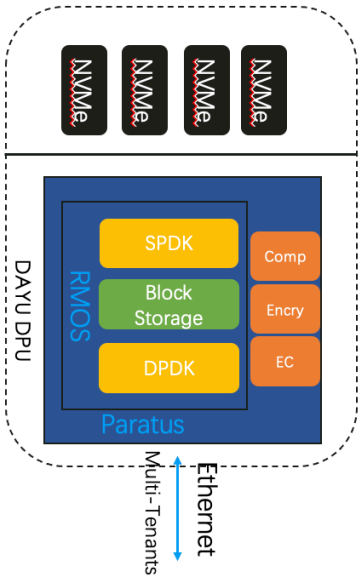

图 8

7.2.2. 计算节点（Initiator）

如下图 9 所示，计算节点上大禹智芯 DPU 提供了 Initiator 的功能，可以把远端块存储池的存储资源拉到本地，并且通过 DPU 的处理，暴露给主机侧 CPU 为“本地”存储。本地存储可以为 VirtIO-blk 或 NVMe 类型的块设备。从”网络“到”本地“的转化过程，DPU 起到了网络协议卸载的功能，使原先需要由 CPU 处理的网络协议栈解析/封装工作转移到 DPU 处理，释放了 CPU 的计算资源。同时，如果用户有自己的 Initiator 程序，大禹智芯 DPU 也提供运行环境，用户可以使用自己定制化的 Initiator 程序运行在 DPU 上，同样可以实现存储网络协议的卸载。主机 CPU 上可以有更多的计算资源用来运行应用，提供虚拟化计算能力，或者运行文件/对象存储的客户端。

除了网络卸载，DPU 还可以为存储提供压缩/解压缩，加密/解加密的加速能力，配合远端存储资源池共同使用，把原先需要在存储后端处理的数据服务，前移到使用端处理，可以大幅提高存储数据服务处理能力和性能。

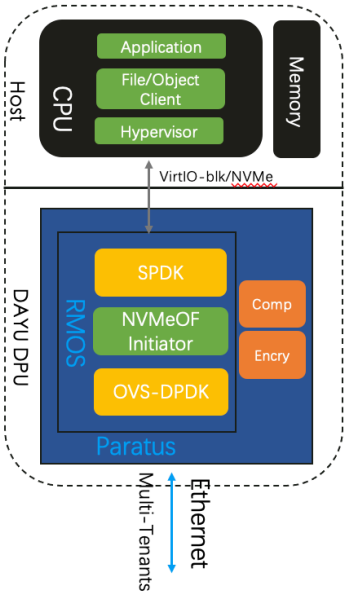

图 9

7.2.3. 文件/对象存储

如图 10 所示，文件/对象存储节点与上述的计算节点非常类似，利用 DPU 的 Initiator 能力，把远端块存储转换为“本地”块存储，暴露给主机侧 CPU VirtIO-blk 或 NVMe 类型。在此基础上，主机侧 CPU 上可以运行文件/对象存储软件栈，并同时对外提供文件/对象存储服务。

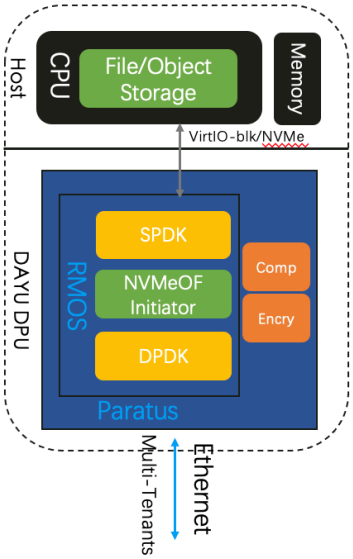

图 10

7.3. 基于大禹智芯 DPU 的下一代 HPC/AI Fabric

上述中提到了 DPU 的由来，DPU 在云计算场景中对网络，存储的处理，DPU 是如何帮助构建 “存算分离” 的基础设施架构的。我们再进一步理解 DPU 在 HPC/AI Fabric 中的作用。在现有架构中，CPU、GPU、NVMe 是紧耦合的部署方式，均在同一个物理机箱内，这在两个方面限制了现有架构的能力及扩展：一是 CPU 与 GPU 的紧耦合无法实现资源的灵活调配，不同的计算任务中 CPU 与 GPU 的比例往往不同，经常会出现 CPU 算力不足或 GPU 利用率不高的问题；二是随着计算模型的规模增大，完整的计算模型无法存放在本地，需要更多的借助远端存储能力，同时现有体系内对远端存储数据的利用必须经过 CPU，因此造成架构上的瓶颈。基于上述两点，下一代 HPC/AI Fabric 中的两个关键功能点将定义为：一是能够实现资源的池化及灵活分配；二是资源池之间有高性能网络支撑。针对这两个关键功能点，大禹智芯 DPU 均能提供相对应的解决方案。

如图 11 所示，CPU、GPU、NVMe 均在本地与 DPU 互联，DPU 与各个组件形成资源池。使 CPU、GPU、NVMe 打破现有紧耦合的部署方式。在 CPU+DPU 的组合中，DPU 负责把由 CPU 进行预处理后的数据分发给 GPU 和 NVMe，同时把 CPU 对 GPU 的管控从现有的总线层面延伸到以太网层面。在 GPU+DPU 的组合中，DPU 作为 GPU 的本地控制单元，除了与 CPU 端的 DPU 配合实现 CPU 与 GPU 的管控和数据分发外，还负责 GPU 到 GPU 或 GPU 到 NVMe 的数据流转，无需借助 CPU 的过多参与甚至是 bypass CPU。在 DPU+NVMe 的组合中，与上文阐述的基于大禹智芯 DPU 的存储架构保持一致，实现解耦的块存储+文件/对象

存储架构。

高性能网络在此架构中起到关键作用，资源池化后，资源池间的数据流转就变得越发重要，这就需要在各资源池间通过高性能网络保障计算任务的顺利进行，达到近似“本地”运行的效果。大禹智芯 DPU 通过提供端到端的先进流控技术，实现高吞吐，低延时，窄尾延时的可靠高性能网络。通过 DPU 提供的驱动接口，系统和应用调用网络能力变得便捷灵活，实现与应用的结合。

在数据传输过程中，DPU 的压缩/解压缩与加密/解密能力可实现端到端的数据安全。

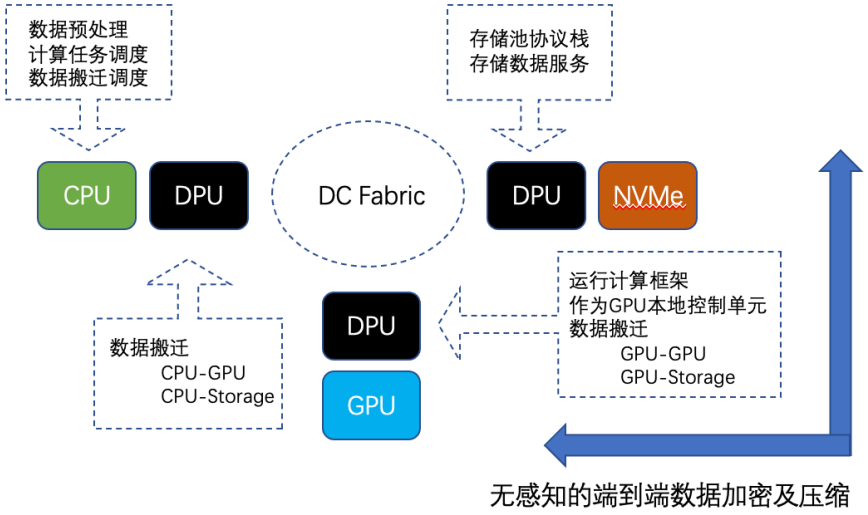

图 11

7.4. 大禹智芯 DPU 与管控平台的对接

云环境中的自动化部署是非常重要的一环，这一环节对云服务的用户体验具有直接影响。鉴于此，大禹智芯 DPU 充分考虑与主流云管平台的对接。从具体技术实现上，与现有部署方式和使用方式高度保持一致，保证在用户使用过程中与现有管控平台的对接更便捷。

对 DPU 自身的管理和控制具体如下：

7.4.1. 管理路径

大禹智芯 DPU 拥有独立的管理端口，可以与客户代外管理网络互联，提供独立的管理路径。除了独立管理路径外，DPU 自身也支持带内管理和主机内部管理两种方式。

7.4.2. Openstack

大禹智芯 DPU 通过提供 Neutron 和 Nova 的插件，实现与 Openstack 社区版本的集成。

7.4.3. Kubernetes（K8S）

大禹智芯 DPU 通过提供 CNI 和 CSI 的插件，实现与 Kubernetes 社区版本的集成。

9. 大禹智芯 DPU 的应用场景示例

9.1. 云化裸金属

大禹智芯 DPU 实现云化裸金属方案，提供 OVS 的全卸载，实现网络多租户，利用存储 Initiator 能力实现远程启动和远程挂载数据盘。裸金属的网络与存储全部可以由云管平台统一管理和部署，实现与虚拟化场景下的虚拟机/容器的互通。

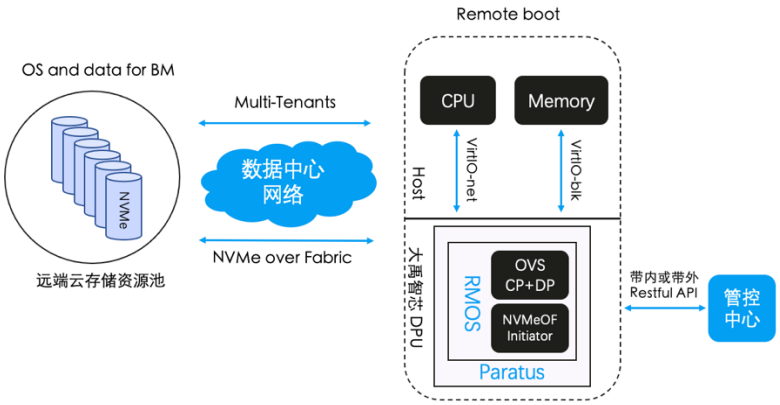

图 12

9.2. 虚拟机云

大禹智芯 DPU 为虚拟机所需的虚拟网络及存储实现全卸载，实现虚拟机镜像及数据使用远端存储，无需依赖本地存储。完全释放宿主机计算资源，大幅提高虚拟机部署密度，与用户云管平台结合可实现完全自动化部署和管控。

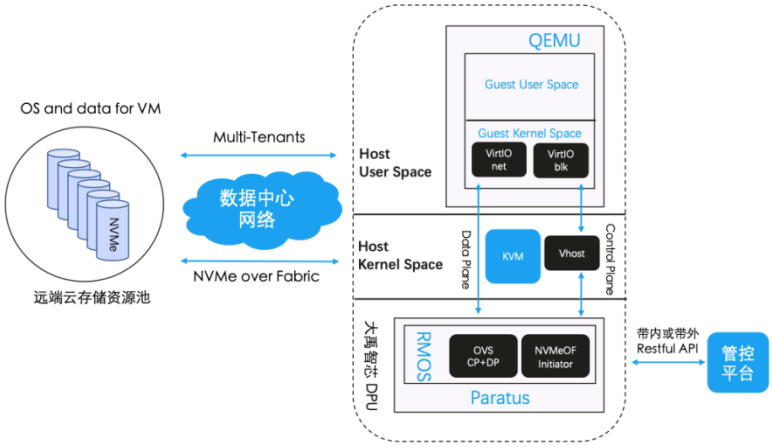

图 13

9.3. 容器云

在云化裸金属方案基础上构建容器云。容器 POD 内与 POD 间的虚拟网络流量均由大禹智芯 DPU 处理，所有容器 POD 间的网络访问策略也全部由 DPU 来承载，充分释放主机侧 CPU 资源，大幅提高单宿主机容器部署密度以及容器集群的容器/POD 数量。容器镜像及持久化存储均使用远端存储，无需本地存储。

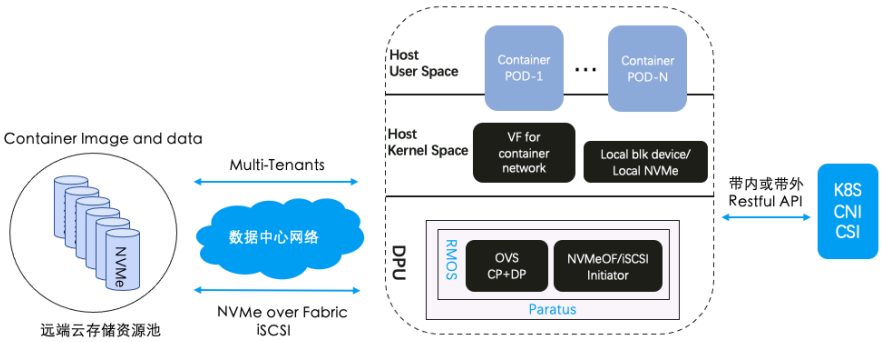

图 14

9.4. 流量编排

在云安全场景中，安全原子能力以虚拟机或容器形式存在，订阅安全服务的流量在各原子能力间需要实现服务链功能。大禹智芯 DPU 可实现原子能力间的流量编排，在保证功能和性能的前提下大幅降低 CPU 计算资源的损耗，使原子能力部署密度提升。

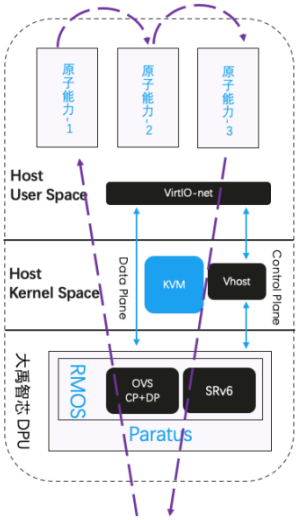

图 15

9.5. 分布式边缘安全平台

传统安全防护的部署往往是在数据中心边缘，此部署方式可以有效地对数据中心南北向流量进行安全防护。但随着应用部署方式的变化，分布式、集群化和虚拟化的广泛应用，数据中心流量模型发生了根本变化，数据中心流量为东西向流量和南北向流量的比例约为80:20。传统的安全能力部署方式已经不能应对数据中心流量模型的变化，对数据中心东西向流量的安全防护能力逐渐被广泛关注。基于大禹智芯 DPU 的分布式边缘安全平台，安全防护边界从数据中心边界下沉到应用与网络的边界，可对东西向和南北向流量均实现安全防护功能，同时由于 DPU 自身提供虚拟化网络能力，对虚拟网络下网络流量有天生感知能力。大禹智芯 DPU 不仅可以提供应用网络流量视图，还可以承载第三方安全应用，丰富安全能力维度。

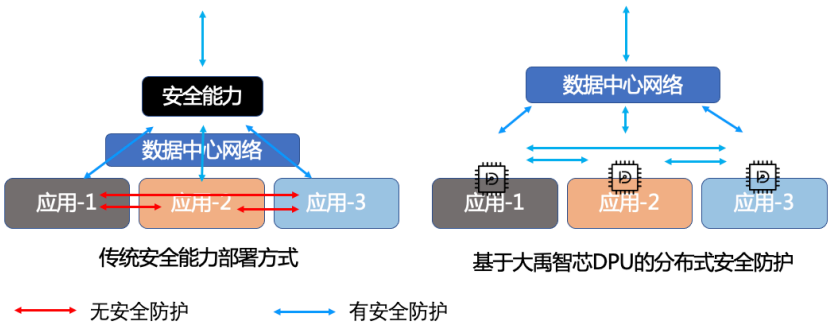

图 16