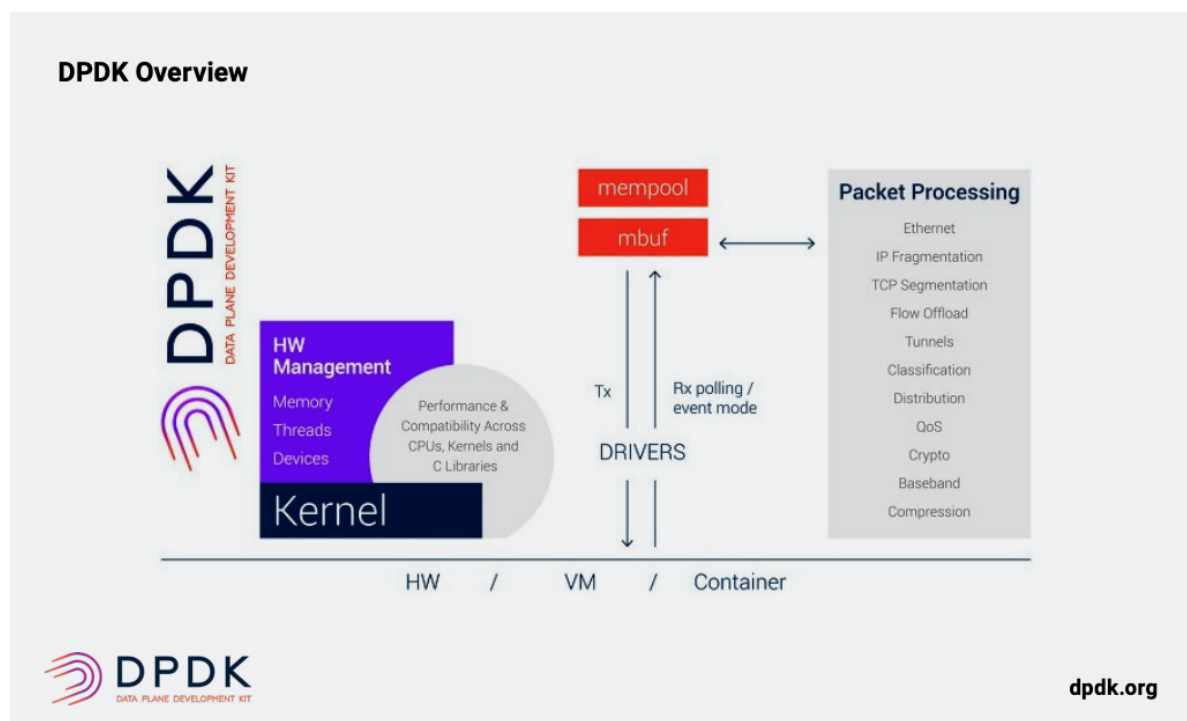


## 简介

数据平面开发套件（DPDK）软件是一组用户空间库和驱动程序，可加速在所有主要CPU架构上运行的数据包处理工作负载。由Intel大约10年前创建，现在是Linux基金会旗下的开源项目之一。从企业数据中心到公共云的环境，尤其是在电信网络中，DPDK在推动高性能通用CPU的使用方面都起了重大作用。

## DPDK是什么

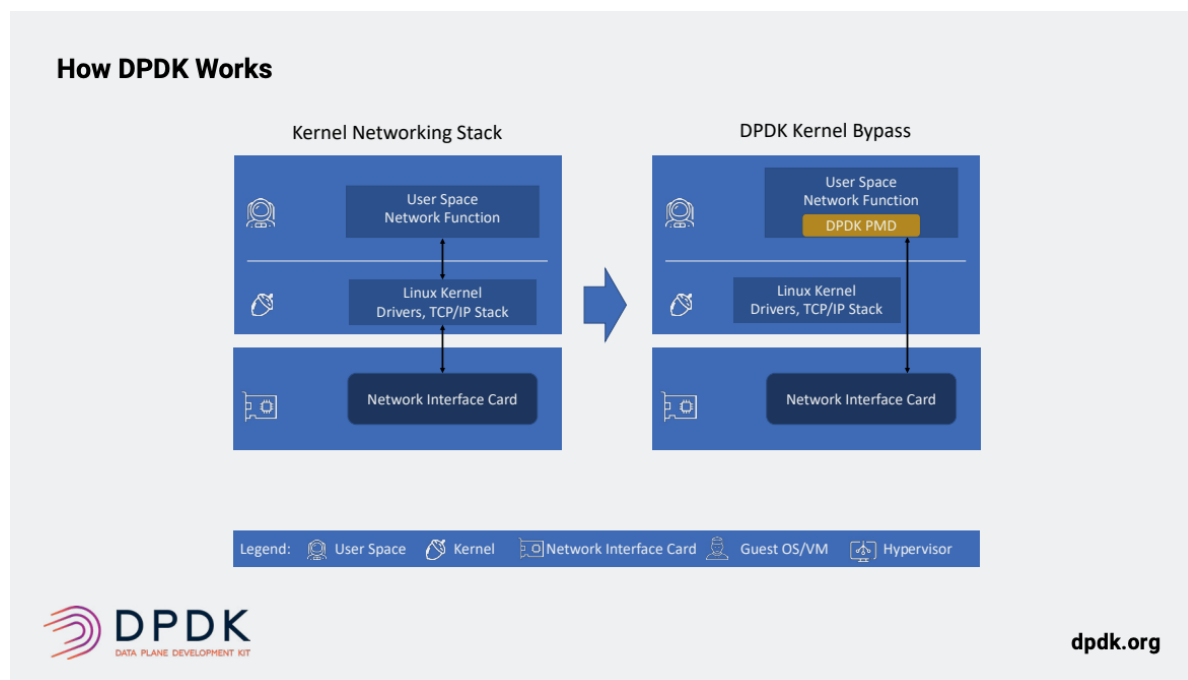
DPDK是Linux基金会托管的一个开源项目。为了加快网络I/O，DPDK允许传入网络数据包过渡到用户空间，进而消除了内存复制的开销，在不进行内存复制的情况下，这些数据包将被快速处理用户空间和内核空间之间上下文切换的开销。DPDK扩展了标准内核网络堆栈，在高吞吐量和/或低延迟至关重要，例如无线核心，无线访问，有线基础设施，路由器，负载均衡器，防火墙，视频流，IP语音等。许多流行的Linux发行版，包括由Red Hat和Canonical支持的发行版将DPDK支持作为其标准包装的一部分。



## 为什么需要DPDK-DPDK的发展史

传统上，由思科，爱立信，华为，瞻博网络，中兴和诺基亚等供应商提供的联网设备使用ASIC来执行低级数据平面功能，例如数据包处理。在某些情况下，这些ASIC是专有的，而在其他产品则是诸如Broadcom或Marvell之类的芯片供应商提供的标准产品。反过来，利用了ASIC通过专有软件来实现防火墙，路由器，交换机，基站支持的各种网络协议站和其他网络设备。这些架构提供了高性能所需的吞吐量网络方面，新产品推出的时间表受到漫长的芯片开发/调试周期以及供应商之间没有软件可移植性的机会。到2007年，半导体公司如Intel，Cavium（现在是Marvell的一部分），Freescale（现在是NXP的一部分）和NetLogic（现在Broadcom的一部分）正在引入具有足够处理能力的标准多核处理器，以执行低级数据包处理功能需要以与基于ASIC的网络有效竞争所需的成本和性能产品。问题出在软件上：Linux是网络设备的默认操作系统，而Linux内核包含许多瓶颈，无法有效处理数据包。需要一种在保持与Linux兼容性的同时消除这些性能限制的解决方案应用程序。它适合作为Linux发行版中存在的库进行打包，以供在需要时使用。管理各种网络设备。

这些目标最初是在2010年实现的，当时英特尔推出了针对Xeon一代的DPDK初始版本基于Nehalem微体系结构的处理器。DPDK绕过Linux内核，在用户中执行数据包处理空间以最大化网络性能。DPDK通过使用用户中运行的轮询模式驱动程序（PMD）来实现此目的空间，不断检查传入的数据包队列以查看是否有新数据到达，从而实现了高吞吐量和低吞吐量延迟（下图）。目前，根据开放源BSD许可，通过.zip文件提供了DPDK



2013年，网络软件供应商6WIND建立了DPDK.org开源社区，以开发人员为中心资源，托管库，驱动程序，文档，开发人员的邮件列表和git存储库。DPDK.org的出现促进了DPDK的使用量的增加，以及来自不断扩大的IPDK池对DPDK的捐款的显着增长全球开发商。

## 误区 #1: DPDK是英特尔控制的

在DPDK的早期，英特尔确实对该项目产生了巨大的影响。但是，随着项目的发展社区的多样性也不断发展，例如Chelsio和Mellanox等设备供应商的贡献以及其他成员的贡献都占了重要比例。英特尔继续为改进了DPDK提供支持，但其他非x86体系结构的供应商（例如Arm，Power和Tilera）拥有将DPDK推到了纯粹的x86外都多种平台。如前所述，Linux基金会的治理结构会引导DPDK项目容纳更多都参与者，包括英特尔。

## DPDK在NFV和其发展中扮演的角色

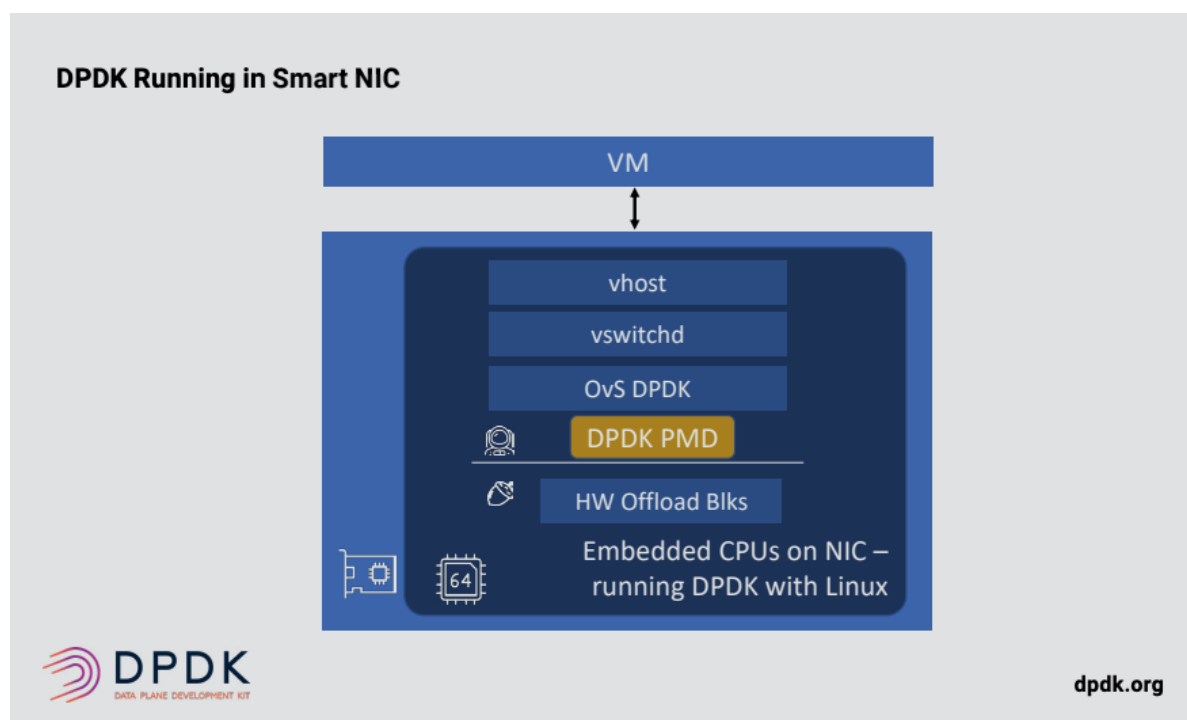
在电信行业中，随着公司开发解决方案来满足DPDK的要求，DPDK成为关键的使能技术。网络功能虚拟化（NFV）的目标。NFV计划始于2012年，当时有七个通讯服务提供商（CSP）同意在ETSI的主持下进行合作：AT & T，BT，德国电信，Orange，意大利电信，西班牙电信和Verizon。成员公司的列表增长迅速，现在包括800多个组织，其中包括 CSP，供应商，公共云提供商，研究人员，学术界和政府实体。

NFV的主要目标是实现网络功能作为标准托管的虚拟化软件的实现服务器平台，而不是传统的网络设备，后者包含在专有软件上运行的专有软件硬件。为了使这种方法具有成本效益，虚拟网络功能（VNF）至关重要。在以下情况下在标准CPU架构上运行时，能够获得足够的性能，吞吐量和延迟管理程序的控制。原始的ETSI NFV白皮书概述了该计划的目标和高层架构，将DPDK与多核处理器和电信级一起确定为NFV的关键启用技术虚拟化平台。大部分电信流量由小数据包组成；没有DPDK，瓶颈与在Linux内核中处理这些数据包相关联将阻止VNF达到所需的性能水平。但是，通过在VNF本身和虚拟交换机（vSwitches）中利用DPDK，开发人员能够实现高吞吐量和低延迟，从而使NFV概念成为可行且具有成本效益的固定功能设备的替代品。

尽管NFV的最初前提是在标准服务器上运行虚拟化工作负载，但现在重点是加速平台，包括智能网络接口卡（SmartNIC）和/或图形处理单元（GPU），这样可以大大减少主机CPU资源需求，从而节省大量的CAPEX和OPEX。工作是DPDK社区中正在进行的工作，以确保SmartNIC和GPU都可以作为DPDK的加速器进行接口，当系统中有SmartNIC和GPU资源时，使应用程序能够利用它们。

## 误区 #2 智能网卡会杀死 DPDK

在SmartNIC本身上，有一些使用DPDK的SmartNIC体系结构将数据包移入用户的用户空间。在NIC上运行的操作系统。DPDK还提供了与SmartNIC的接口，从而简化了相互之间的交互。智能网卡。无论是基于ASIC，CPU，FPGA还是NPU的DPDK都能为SmartNIC增加价值。



## Linux 基金会和 DPDK

在2017年，DPDK加入Linux基金会，一个以为促进开源技术广泛合作的中立家园。现在，DPDK项目包括一个核心软件项目，DPDK本身以及各种与DPDK密切相关的较小子项目。DPDK的核心项目由许多软件git组成存储库，包括主存储库，几个“下一个”存储库，这些存储库用于将更改应用到存储库的特定部分将DPDK放入主存储库和稳定发行版存储库之前。每个项目都有自己的维护者和维护过程。

DPDK的治理是由两个不同的委员会（管理董事会和技术委员会）推动的。管理董事会负责预算，市场营销，实验室资源，管理，法律和许可问题。技术委员会负责技术问题，包括批准新的子项目，弃用旧的子项目以及解决技术纠纷。这两个委员会是对等的，并且一起监督DPDK项目。BSD下提供了主要的DPDK代码许可，而与Linux内核相关的部分则根据GPL许可。

Linux基金会在世界各地组织了许多针对DPDK的活动。DPDK峰会涵盖了最新的DPDK框架的开发和相关项目已在中国，欧洲，印度和北美的多个城市举行。在欧洲则举行了规模较少但技术含金量更高的DPDK userspace峰会，并提供了包括路线图在内的更具技术性的议程。

## DPDK 社区的发展

到2019年末，DPDK的贡献者来自25个以上组织的160多位专家。独立于技术贡献者之外，DPDK项目有两种成员资格。有10个金牌会员：Arm，AT&T，爱立信，F5，英特尔，Marvell，Mellanox，恩智浦，红帽和中兴通讯。银牌会员三名：6WIND，华为和博通。（AMD将成为DPDK第四个银牌会员）金牌会员在董事会中具有更大的代表权。

越来越多的开源项目利用DPDK。其中包括ANS, BESS, Butterfly, DPVS, FD.io / VPP, FastClick, Lagopus, MoonGen, mTCP, OPNFV, OpenDataPlane, Open vSwitch, Packet-journey, Pktgen-dpdk, PcapPlusPlus, Ruru, Seastar, SPDK, TRex, WARP17, YANFF和 AStack。作为 Open vSwitch (OvS) 的一部分, DPDK被广泛部署全球云, 企业和电信数据中心。

## 误区 #3 DPDK是一个封闭且排他的社区

的确, DPDK项目是由一小组专家和非常熟悉Linux内核和设备驱动程序编程的Guru们组成的。但是, 伴随社区显着和持续的增长, 已经有数百人参与该项目, 包括硬件开发人员, 软件开发人员, 质量保证人员, 文档编写者和最终用户的开发人员。DPDK已经超越最初只有少数公司参与, 现在有数十家公司参与DPDK项目的发展。任何人都可以通过<https://www.dpdk.org/contribute/> 成为DPDK的成员。所有提交均经过项目维护专家审查吸收变更以确保符合质量标准并遵循最佳的实践。

## DPDK市场发展

DPDK的大多数应用最初是在电信领域。随着CSP采用网络虚拟化来降低运营成本并加快新服务的部署, 他们虚拟化了需要高吞吐量和/或低延迟的用例, 例如路由器, 防火墙, 无线电访问网络 (RAN) 和演进的分组核心 (EPC)。虚拟化平台的供应商, 在这些情况下, VNF和应用程序已在其产品中利用了DPDK, 以实现CSP的性能目标。随着CSP探索新的边缘托管应用, 例如视频缓存, 监控, 增强现实 (AR), 辅助驾驶, 零售和工业物联网, DPDK仍然是实现积极性能目标的关键技术。

类似DPDK最先在电信应用程序中, 对数据包处理功能的性能要求一样, DPDK越来越多地应用于企业和云当中。例如, 在2018年, VMware引入了他们的NSX-T数据中心软件定义基础结构的基于DPDK的边缘配置。此版本的NSX-T地址需要具有可变数据包大小的高数据包吞吐量的应用程序以及支持具有以下功能的高速NIC的服务器高达100Gbps的北/南流量。通常, 南北向流的包大小各不相同, 数据包处理要求, 即使它们只占总流量的不到20%。在此用例中, 通过使用带有小数据包 (64字节) 的DPDK, 英特尔和VMware的分析表明, 性能提高了五倍。

同时有几家公司已将DPDK用于金融应用, 其中低延迟带来了巨大的竞争优势。例如, 在高频交易 (HFT) 中, 延迟会直接影响交易者的交易效率

算法策略及其超越竞争对手的能力。信息周刊估计, 对于一家大型经纪公司来说, 一毫秒每年价值1亿美元。DPDK是这个市场解决方案供应商开发所依赖的关键技术。

## 误区 #4 DPDK只适用于电信行业

随着NFV的到来, DPDK对于虚拟化网络功能的成功变得尤为重要, DPDK可用于企业数据中心以及云服务提供商中部署的解决方案。在

除了前面提到的各种用例之外, DPDK的其他应用程序还包括虚拟交换, 软件定义的网络 (SDN), 数据分析, 人工智能 (AI) 推理, 视频转码和在线赌博。七大云提供商—阿里云, 亚马逊网络服务, 百度网盘, 谷歌云平台, Microsoft Azure, 腾讯云-所有产品都使用DPDK。

## DPDK 跨架构的快速增长

英特尔架构, AMD, ARM, Power和RISC-V

英特尔在2010年推出支持x86 Xeon平台的DPDK时, 其他几家供应商也提供了多核面向网络应用程序的处理器及其自己的SDK库, 以实现高性能数据包处理。例如, Cavium和NetLogic都推出了基于MIPS CPU架构的处理器, 而飞思卡尔则是基于PowerPC架构的。在接下来的几年中, 这三家公司都推出了基于Arm架构的新多核处理器系列。有意思的是, 随着半导体行业的持续发展, 这三家公司最终都被收购了。如今, 面向高端网络应用的处理器的供应商格局已经逐步发展到市场领先的产品基于Intel和AMD的x86架构或包括Broadcom, Marvell和NXP在内的多家供应商提供的Arm架构。IBM Power体系结构用于适用于企业和云应用程序的处理器。现在, DPDK支持所有这三种架构, 并在在添加开源RISC-V体系结构方面取得了进展。



作为DPDK的创始人，Linux Foundation项目的金牌会员和领先的贡献公司之一，Intel推动了DPDK中的许多创新，确保为所有相关的英特尔产品提供高度优化的支持，并且维护良好。除了适用的Xeon处理器外，这些产品还包括网络适配器，例如e1000，ixgbe，i40e，fm10k和ipn3ke。

DPDK在AMD的x86处理器（例如EPYC Embedded 3000系列）上运行。AXGBE轮询模式驱动程序（PMD）提供支持集成在EPYC SoC中的AMD 10Gbps网络适配器系列。

OpenDataPlane是第一个定义用于在基于Arm的SoC上进行网络数据平面处理的API的开源项目（ODP），由Linaro Networking Group在2013年发起。Linaro是由Arm，Freescale Semiconductor，IBM，三星，ST-Ericsson和Texas Instruments于2010年成立的工程组织，致力于Arm的开源软件的开发。体系结构平台。2018年，Linaro Networking Group解散，ODP的管理移至OpenFastPath（OFP）基金会，由Arm，Enea和Nokia于2015年成立。启用与DPDK，OFP通过ODP-DPDK层实现了DPDK支持。

## 误区 #5 DPDK是一个纯软件项目，是反硬件的

DPDK最初是一个旨在通用CPU上提高OS的数据包处理性能的项目。它已经发展到远远不止于此了，如今的DPDK更像是一个框架，在Linux和其他UNIX类型的OS和Windows上的加速（请参阅<http://doc.dpdk.org/guides/>上列出的OS支持）。DPDK支持从SmartNIC到加密加速器的各种硬件设备。它提供了一个统一的库可以在可用时利用硬件功能，但在硬件不可用时，使用全软件路径的实现。这样可以确保依赖DPDK的应用程序始终可以正确运行，尽管当没有独特的硬件加速挂钩时，性能会有所降低。

由于这两个ODP-OFP计划，Arm体系结构在DPDK中具有强大的支持。多个采用Arm执照的公司为DPDK做出了贡献，致力于确保DPDK利用该架构的所有功能。由某些CSP和设备提供商提倡，与具有相同功能的x86平台相比，其电源效率可能更高性能。DPDK中提供了对IBM Power架构的支持，并且最新的DPDK版本已合并Power9的库。IBM和Canonical合作开发了一个版本的Ubuntu Server，可以直接从IBM订购，其中包括DPDK性能优化。

为了完善DPDK中支持的处理器体系结构，目前DPDK社区正在努力添加RISC-V。RISC-V基于2010年在伯克利的加利福尼亚大学开始的一个项目，是一个免费的开源指令集，RISC-V基金会现在促进和维护体系结构（ISA）。通过2018年宣布的合作，Linux Foundation提供了RISC-V生态系统的资源，例如培训程序，基础架构工具和行销。SiFive是一家无晶圆厂的半导体公司，其基于RISC-V架构（例如Freedom）生产SoC，并在2016年推出的Everywhere 310 SoC。

鉴于其广泛的架构支持，DPDK越来越不被视为NFV的加速器，而被视为Linux用户空间框架，为对高性能I/O和操作系统至关重要的核心功能提供统一的抽象层包处理，包括加密功能。DPDK支持包括FPGA，ASIC和SmartNIC在内的各种硬件设备，从而充分利用了所有可用的，支持的资源，最大程度优化系统级别的整体网络性能。

## 误区 #6 DPDK 只与 Linux 相关

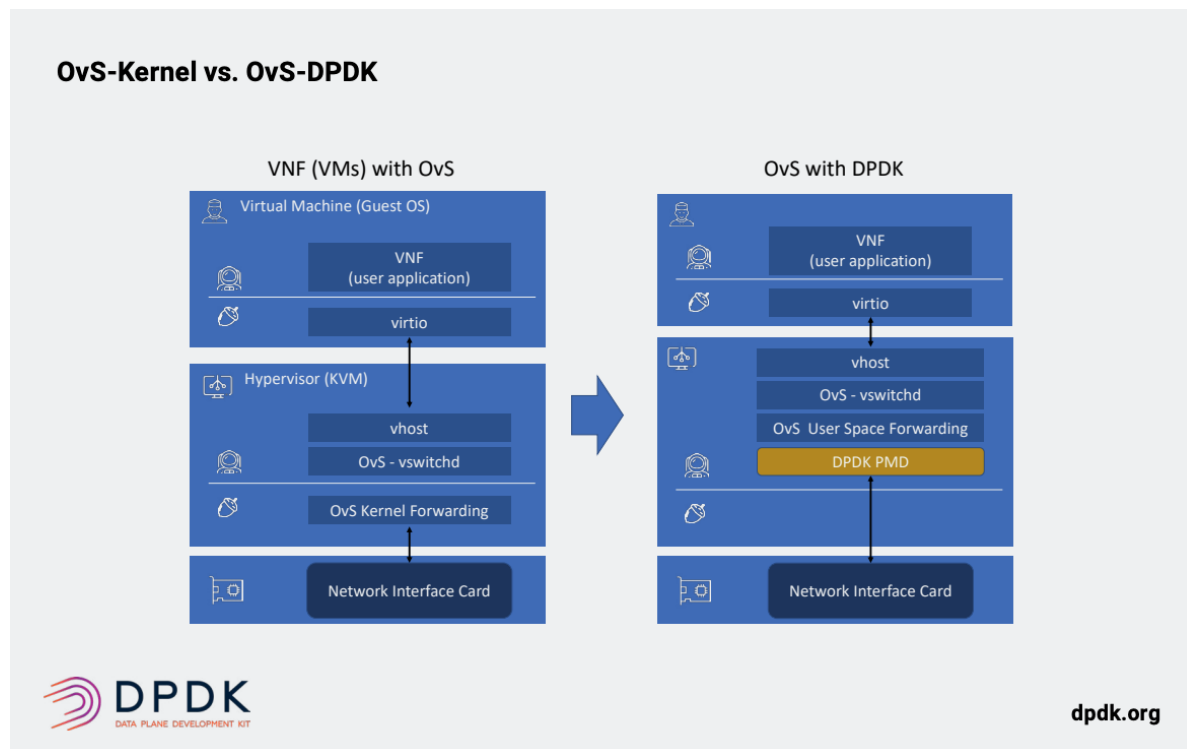
由于在数据中心内的广泛使用，DPDK最初是在Linux上实现的。但是，今天的DPDK支持更广泛的操作系统，例如FreeBSD。此外，DPDK正在开发移植到Windows操作系统。

## Open vSwitch 和 开放虚拟网络

Open vSwitch是Linux Foundation Networking项目，是分布式虚拟多层的开源实现。使用Apache 2.0许可证。它可以通过程序扩展实现大规模的网络自动化，同时仍支持标准管理接口和协议，例如NetFlow，sFlow，IPFIX，RSPAN，CLI，LACP和802.1ag。OvS的Linux内核实现已于2012年合并到内核主线中。

最初英特尔与OvS社区接触，共同探索一种架构增强功能，该功能将使DPDK成为OVS额外的数据平面，但由于当时社区致力关注企业应用程序，认为这个主张的价值尚不明确，该建议也失败了。然后，根据NFV用例的性能要求，英特尔与两个电信客户合作创建了DPDK。加速开放式vSwitch，简称ovs-dpdk，是OvS的分支，它利用DPDK优化了电信网络所需的流量以小数据包为主的OvS的性能。

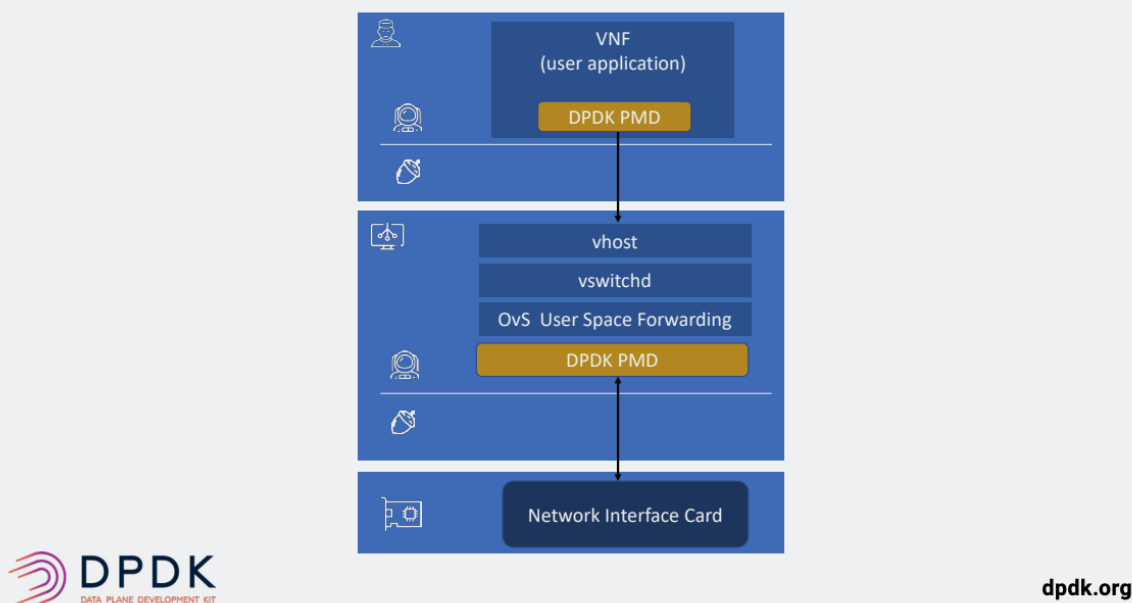
2013年，英特尔公开宣布OVDK为开放项目，并有外部贡献者的参与。他们在2014年表示OVDK项目已达到演示所需性能加速的目标，并且OVDK将停止支持将DPDK集成到主线OvS代码库中，并致力于在OvS中保持这种支持。



到2015年，OvS-DPDK已作为OvS的构建选项合并到主线中，从而提供了实施OvS中的用户空间数据平面（图4）。OvS-DPDK的小数据包吞吐量提高了10倍以上，并且低得多的延迟。使用DPDK数据包处理还优化了OvS内部的几个性能热点区域库。开放虚拟网络（OVN）是一个开源项目，最初由Nicira的OvS团队于2015年启动（现在作为支持虚拟网络抽象的系统。OVN通过添加对虚拟网络抽象（例如虚拟第2层和第3层）的本地支持，对OvS的功能（包括OvS-DPDK配置）进行了补充覆盖和安全组。OVN的开源绑定可用于许多平台，例如OpenStack和Kubernetes。OVN是许多商业产品中使用的SDN平台，包括Red Hat Virtualization，Red Hat OpenStack和Red Hat OpenShift。OvS-DPDK为基于这些的部署带来了高数据包处理性能产品。

如今，对于大多数NFV部署，我们看到OvS-DPDK与VNF本身中启用的DPDK的组合（下图）。这提供了从NIC到应用程序的全程强大网络性能，同时保留了将vSwitch提供给可能相互通信的多个VNF的灵活性（东西向服务链接 Service Chain）。正如我们将看到的，DPDK可以用于多种部署架构中，并可以与其他方法结合使用加速，包括基于硬件的解决方法。

## DPDK in Both OvS and the VNF



## 相关开源项目和标准

尽管DPDK是软件加速工具包中最著名的一个，但也有其他开源项目和与DPDK重叠或相近的框架。其中包括Linux Foundation FD.io的矢量处理平面（VPP）项目，以及Linux Foundation ioVisor项目中的Express Data Path（XDP），以及AF\_XDP和虚拟数据路径加速度（vDPA）。还有Snabb数据包工具包和netmap项目，它们都采用类似的方法，尽快将数据包移入用户空间。尽管Snabb仍是一个活跃的项目，但由于DPDK的卓越的性能，现在已基本取代了netmap。

在Linux和其他UNIX风格的操作系统中，有许多方法可以改善网络性能。从诸如将有效负载校验和添加到NIC的卸载功能到成熟的TCP卸载引擎（TOE）。随着最近一波NFV工作负载的出现，人们对各种方法重新产生了兴趣，DPDK只是其中一种。为了简洁起见，我们将简要介绍不同的技术，并为读者提供有关DPDK当今适合的背景以及有关其他流行且有希望的网络加速计划的信息。我们将从硬件辅助机制开始，其中许多机制是DPDK的补充：

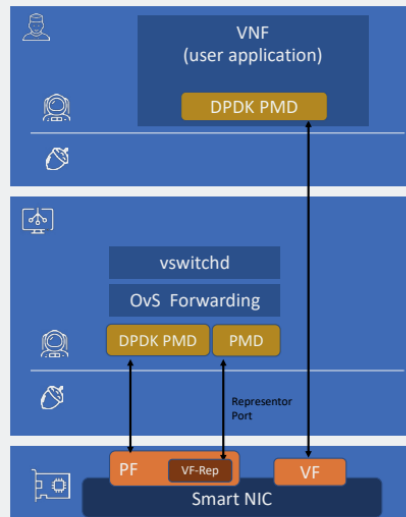
### PCI 直通

PCI直通涉及将VM guest虚拟机绑定到特定的PCI卡，就像VM是具有完全访问权限的裸机系统一样，包括NIC的所有权。可以使用类似的技术将特定的NIC绑定到容器。这意味着没有NIC和网络应用程序的其他应用程序争用具有独占访问权。数据包进入后VM或容器，它们仍必须在用户或内核空间中进行处理，但这是一个单独的考虑因素。

### 单根I/O虚拟化

单根I/O虚拟化（SR-IOV）是PCI标准的扩展，涉及创建虚拟功能（VF），该功能可以被视为独立的虚拟PCI设备。可以将每个VF分配给VM或容器，并且每个VF具有传入数据包的专用队列。同样，数据包到达后如何处理是一个单独的决定。SR-IOV是通常在实际部署中与DPDK结合使用，目标是绕过主机和来宾OS的内核。它可用于提供从VNF（VM）中的用户空间到虚拟接口上的传入数据包队列的直接访问在NIC上，减少了等待时间，并减少了复制过程（图6）。容器可以使用相同的技术。主要的SR-IOV的问题在于它需要支持该功能的NIC，并且每个VF都占用NIC上的物理资源，因此虽然理论上VF限制很高，但实际上可以限制多少个VF存在实际的内存限制。

## DPDK with SR-IOV and a SmartNIC

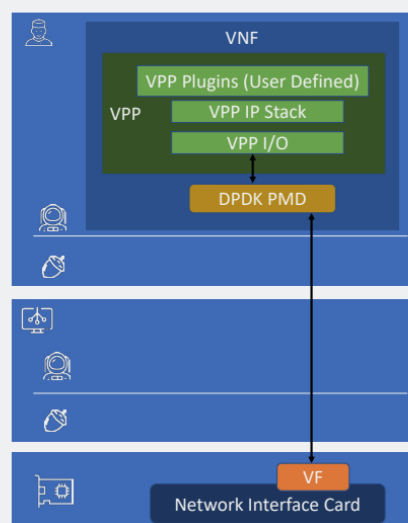


dpdk.org

## FD.io/VPP

矢量数据包处理器（VPP）是Linux Foundation下的Fast Data（FD.io）项目的一部分。VPP最初是由思科作为一个开源项目贡献。VPP的目标是提供一个快速的2-4层用户空间网络堆栈，在x86，Arm和Power等常见架构上运行。如今，大多数VPP实施都利用DPDK作为插件，加速通过DPDK PMD将数据包进入用户空间。VPP专注于上层网络协议（下图）。VPP通过在以下位置执行功能来获得其大部分性能分批或矢量分组，而不是单个分组。VPP为第2层和第2层提供了虚拟交换机和虚拟路由器第三层数据包处理。它还集成了优化的TCP / IP堆栈，该堆栈使用矢量化数据包处理来提高性能。性能。与使用DPDK相比，这使在VPP之上构建高层（L4-7）网络功能更加容易。VPP专注于招募虚拟网络网关和安全性等上层网络功能的开发人员。它还正在努力确保使用容器的云原生平台可以有效地使用VPP。

## VPP (FD.io) with DPDK and SR-IOV



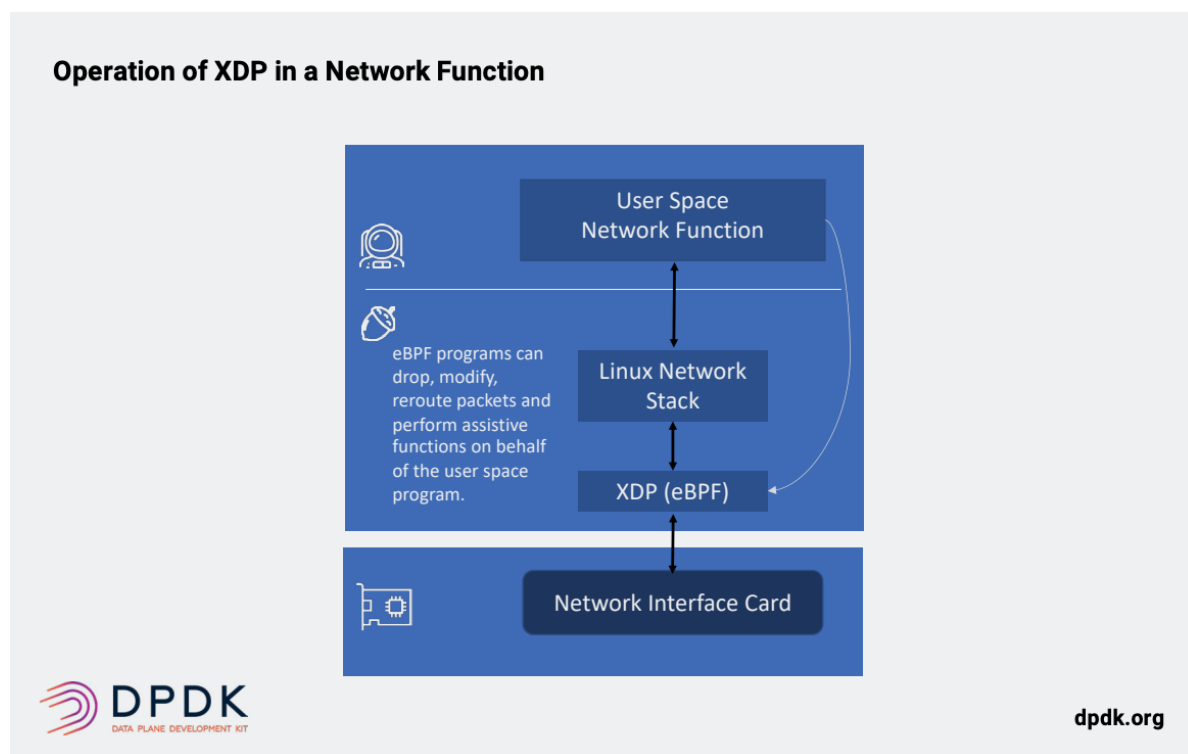
dpdk.org



## XDP 和 eBPF/BPF

eXpress数据路径是ioVisor项目下的内核中数据包处理框架，该框架也由Linux基金会托管。它旨在在内核空间中提供高性能的数据包处理。XDP通过挂钩实现了这一目标，进入内核以对传入的数据包运行优化的代码。这些程序在扩展的伯克利分组过滤器（eBPF）框架下执行。eBPF是1992年创建的原始BPF内核解决方案的改进版本，该解决方案利用了有限的机器指令集，以在内核空间中运行用户创建的程序，基本上在轻量级虚拟机中运行。在将这些用户程序加载到内核之前，已经过验证，以确保它们没有恶意，并且不会无休止运行。BPF的最初用途是通过过滤进行网络故障排除和分析，不久将扩展到安全用例。eBPF诞生于2013年，是对原始BPF的改进，并引入了更复杂的虚拟机。通过低级虚拟机(LLVM)将eBPF用户程序编译为eBPF指令，然后加载到内核中执行。今天，我们使用术语BPF来指代eBPF，将旧的BPF降级为经典的BPF。

因此，XDP方法通过运行自定义逻辑来处理内核中的数据包来实现性能（下图）。XDP BPF和BPF已用于创建复杂的网络解决方案，包括Cilium等安全项目。以内核内方法与内核旁路方法不同，XDP能够在处理数据包的同时利用内核网络堆栈。它正在积极开发中，在除了删除，切换或传递到内核堆栈之外，增加复杂动作的同时改善其性能。

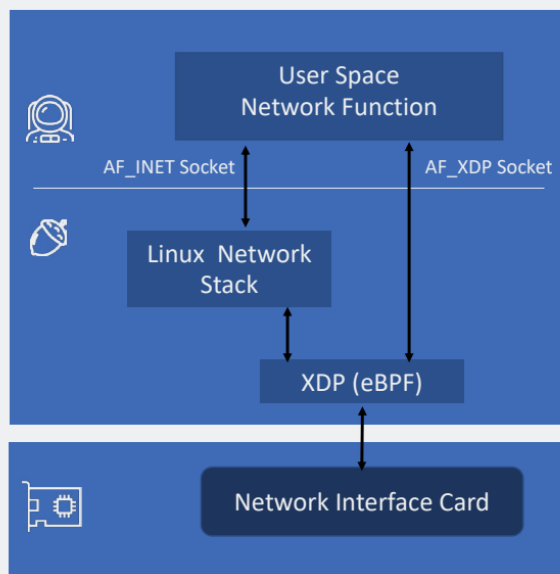


## AF\_XDP

AF\_XDP代表地址族XDP，这是Linux中新的套接字地址族类型。AF\_XDP与XDP有关，因为它使用eBPF机制以及VXP驱动程序层。它会将符合特定条件的数据包直接引导到用户空间中。类似于DPDK的内核绕过。当与XDP结合使用时，这种方法结合了最好的内核逻辑使用DPDK样式的内核绕过。AF\_XDP套接字允许内核XDP程序将帧重定向到用户空间中的缓冲区进行处理，或继续通过内核的现有网络堆栈，TCP / IP等分流一些流量（下图）。

目前AF\_XDP还无法达到DPDK所能提供的高性能水平，该项目得到了英特尔，红帽和Mellanox等供应商的支持。其中一个AF\_XDP方法的潜在主要优点是减少了对特定于供应商的PMD的需求，而是允许构建与基础网卡无关的便携式网络功能应用程序。实际上，DPDK该项目使用AF\_XDP驱动程序框架来简化大量特定于供应商的PMD，从而提供了没有本地分叉的设备的模型。

## Using AF\_XDP in a Network Function



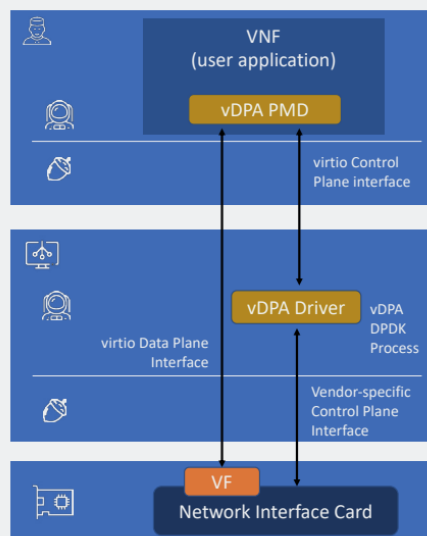
dppk.org

## vDPA

virtio数据路径加速 (vDPA) 项目 (有时称为“虚拟”或“vhost”数据路径加速) 目的是标准化NIC SR-IOV数据平面。它使用了众所周知的virtio框架, 利用了来宾VM中的virtio驱动程序 (图10) 或容器中DPDK的virtio用户PMD, 两者均与供应商无关。在下面vDPA框架中, 每个NIC供应商都将遵循virtio环布局进行数据通信。他们还将提供供应商vDPA驱动程序, 可帮助DPDK将控制平面指令转换为供应商特定的控制平面命令。目标是以减少DPDK下必须维护的大量特定于供应商的驱动程序 (PMD)。像AF\_XDP一样, 它通过减少供应商驱动程序的依赖性, 提高网络功能的可移植性。

vDPA仍处于早期阶段, 并且正在积极开发中。它得到了主要供应商的支持, 例如英特尔, Mellanox和红帽

## DPDK vDPA Framework



dppk.org

## Switchdev和Linux内核

以太网交换机设备驱动程序模型（switchdev）是用于减轻数据平面负担的内核设备模型的框架从内核到交换ASIC的流量。switchdev的目标是允许用户使用Linux内核之上的著名API访问现有的ASIC功能。然后，这些用户可以使用相同类型的命令和工具来配置服务器或交换机。在switchdev下模型，而不是将数据包移至用户空间或在内核空间中对其进行处理，可以将数据包处理分流到NIC卡上的ASIC或机架顶部的外部交换机，它们在硬件中执行高速交换功能。

尽管有新方法可以改善数据包处理性能，但DPDK仍然是最受欢迎的方法并被广泛采用。DPDK社区还积极与许多其他举措进行合作，包括VPP，AF\_XDP和vDPA将继续发展DPDK，并提高其性能和易用性。

### 误区 #7 DPDK仅对硬件供应商感兴趣

是的，DPDK中涉及许多硬件供应商。但是，DPDK有很多甚至更多的服务供应商，系统集成商，最终用户和软件开发人员。

## 基础设施的加速趋势促进 DPDK 的发展

随着媒体器件更高分辨率的需求，物联网设备的数据收集量的增加，加上对业务的大数据和AI分析的日益增长，使得数据中心在数量和规模上都在不断增长。同时围绕5G的市场计划，云和电信市场都看到了关于物联网和边缘计算的新的需求：虚拟化无线接入网络和电信中心局（CO）将商品服务器推向了新的边缘位置。在x86或Arm系统上运行的 Disaggregated RAN，处理视频分析的边缘工作负载以及运行的边缘系统，内容交付网络等都要求更高的网络速度和I/O处理能力。同时随着对隐私的不断关注以及需要防止对通信流进行的恶意攻击，加密已经变得至关重要。当把这些需求添加到对电源，冷却和空间的日益严格的限制中时，对于DPDK或其他提高数据包处理效率的方法的需求就变得越来越强烈。

### 误区 #8： DPDK 不够绿色，它一直占用 CPU的资源

DPDK现在主要依赖于PMD来提高性能，并且因为驱动程序不断以polling的模式查看数据包是否已经到达，这样即使在没有数据包的时候，CPU也会100%的时间处于繁忙状态。这个现象使得有些人相信DPDK不是绿色的。但是从另一个角度看，DPDK允许电信基础设施提高其现有平台的效率，减少了对额外CPU的需求，同时也提高了数据包处理能力。目前DPDK团队正在研究通过扩展核心频率以减少空闲流量时段的功耗的方法，例如允许CPU通过Rx中断唤醒的API。

## FPGA，ASIC，GPU 与DPDK的关系

在运行移动工作负载的边缘无线网关和安全设备中，例如虚拟演进分组核心（vEPC），大多数网关功能都涉及摄取数据包，检查报文头和有效负载以及执行某些操作。DPDK非常适用于这些用例，特别是考虑到它不同CPU类型（x86，Arm）上运行的灵活性，尤其如此。当我们更详细地检查这些边缘位置时，我们发现设备类型和新硬件的多样性在增加，功能包括FPGA，GPU，NPU和自定义ASIC。DPDK必须不断地发展以适应这些特殊的NIC以及融合那些看起来可能与当今数据中心标准服务器不同的基础架构系统。在当今的云数据中心内，SmartNIC被越来越多地采用来分担I/O的处理。

在节约成本和功耗的驱动下，我们希望数据包的低层处理可以转移到硬件上，诸如FPGA，ASIC或NPU的同类产品。简单的交换机，防火墙，基本的负载均衡或路由功能可能不需要通用CPU编程和灵活性。但是，对于更复杂的L4-7功能，DPDK在有效地将数据包发送到用户空间以使应用程序执行更深入的分析中发挥关键的作用。例如，任何类型涉及L7代理中的HTTP标头检查和注入或有效负载处理时，都会更加倾向于使用DPDK。

FPGA / ASIC之间的协作安全解决方案也在不断地发展，其中数据包通过DPDK进入用户空间进行分析。一旦确定流是安全的，FPGA和ASIC便能够直接引导流而无需涉及CPU周期。在GPU方面，我们看到了一些方法，例如Nvidia的并行计算平台CUDA与虚拟技术的结合。网络功能（cuVNF），可直接将有效载荷直接复制到GPU图形内存中，以进行信号处理vRAN解决方案。不过这些还在早期发展阶段，与DPDK的关系会在未来几年有所发展。

## 误区 #9: DPDK 的代码已经完成

DPDK 是一个正在进行和发展的项目，每年都要发布四个版本。DPDK 将继续根据用户的新需求，引入新的加速器设备，增加支持容器部署以及对其他设备的扩展，还有安全性和存储功能的扩展。

DPDK加速器功能： RegEx, Compress Dev, SPDK等

行业中的许多人将DPDK视为内核绕过机制，因其将数据包快速引进用户空间。DPDK框架其实要比这丰富得多。DPDK已经发展成为本质上可以加速使用的通用API基础平台的功能，包括通用CPU和其他功能（例如密码功能）。DPDK包括其他用于压缩和存储的库，以及正在开发中的正则表达式匹配（RegEx）的库。我们把这些DPDK新兴库做一个简要说明：

- **RegEx**：一个DPDK子系统，用于提供对正则表达式的加速匹配，常用于报文头部和用于安全性和L7负载均衡功能的有效负载匹配。
- **CryptoDev**：加速安全功能的库，提供诸如Hashing和公钥/私钥操作。
- **CompressDev**：一个可以方便地访问常见的压缩算法，同时支持两种无状态压缩和有状态压缩有助于减少有效负载大小，同时减轻CPU负担的库。
- **bbdev**：无线基带库提供了抽象硬件的通用编程框架，基于FPGA的加速器和/或辅助3GPP物理层处理的固定功能加速器，同时通过抽象化应用程序的优化功能将其与计算密集型无线功能分离库显示为虚拟bbdev设备。

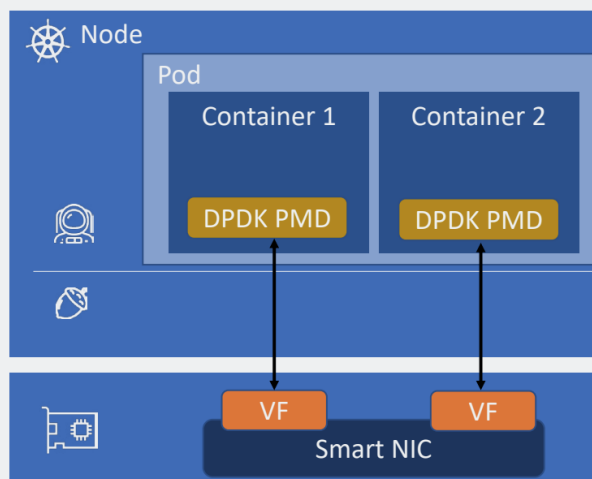
除了DPDK系列之外，还有存储性能开发套件（SPDK），该套件使用一些DPDK代码。尽管本身不是DPDK项目，但他们正在采用类似于DPDK的存储协议方法，试图加快存储速度，通过使用类似的PMD来实现 I/O功能，通过将存储协议直接带入用户空间，类似DPDK在网络方面的提升，来实现存储方面的性能提升。SPDK使用非易失性快速存储器（NVMe），用户空间中具有光纤NVMe（NVMe-oF）和存储协议iSCSI堆栈的标准PMD来提高存储系统的性能。

## DPDK 和 Linux 容器

DPDK最初设计是针对具有虚拟机的环境，在裸机部署和虚拟机中均能很好地工作。DPDK如今可以用于容器，对容器的支持也在不断地改善。当前容器环境中存在一些限制，例如需要为容器分配大量静态内存，数据包缓冲区，这使其不适用于临时的小规模微服务组件。最新的DPDK版本解决了此问题，实现了增强的动态内存分配的功能。

对于DPDK在容器启用特权的上下文的安全性一些开发人员表达了担忧，这可能会增加底层操作系统的受攻击可能性，并且DPDK可能仅仅只能在使用受信任的工作负载的平台中方可运行。不过，DPDK 开发人员正在解决这些问题，我们期望DPDK的容器支持随着容器在现代云原生架构中的兴起，并随着时间的推移而不断地改善。

## Example of Containers with DPDK using SmartNICs



dppk.org

## DPDK的优势与提高

随着SmartNIC的出现，曾经有一些业内人士预测 DPDK 会逐渐消失。然而事实是，DPDK依然在不断地发展，并且工作也在加速。展望未来，我们看到DPDK具有以下优势，以及需要提高的地方：

- **成熟性和稳定性：** API的稳定性是当今DPDK的独特属性之一。不像其他加速方案，DPDK相对成熟，项目维护人员了解保持需求的必要性，可靠的API，以防止代码损坏或持续的补丁程序和重新编译。
- **持续的性能改进：** DPDK将继续寻找方法和智能网卡SmartNIC和其他硬件加速的整合与集成。同样，我们可以期待DPDK 继续在软件性能上的改进。
- **降低复杂性：** DPDK并不是最容易部署和使用的，而且还必须包含多个组件的正确配置和部署，例如带有VNF并支持DPDK的OVS-DPDK。此外，还有无数必须维护的供应商特定的PMD。供应商并不想维护多个驱动程序，无论AF\_XDP统一PMD是否成功，DPDK 都需要更标准化的方式与硬件集成。
- **更好的云和容器支持：** 今天DPDK 虽然支持容器，但还并不理想。预期在用于容器部署的DPDK内存管理中 DPDK 会进行改进，来更好地与容器和微服务架构对齐。同样，安全模型也需要加强，以避免使用DPDK的容器需要特权升级。
- **效率更高：** 在Arm架构上，支持事件触发模式以提取数据包（相对于常量在x86上轮询）。这样的方法以及时钟CPU缩放将有助于DPDK改善其非绿色声誉。
- **多供应商：** 期望在x86以外的各种指令体系中继续提供支持（英特尔和AMD），Arm和PowerPC。同样，英特尔等供应商提供的加密和其他加速功能，也在Mellanox和Marvell以及其它新兴硬件供应商中得到进一步支持和发展。
- **最佳实践的合理化和规范化：** 虽然2020年将出现不同的用例和体系结构，但无论是使用在SmartNIC本身上使用DPDK或由API驱动模型中DPDK调用SmartNIC加速的技术，DPDK与SmartNIC（FPGA，ASIC，NPU）相结合的最佳实践方式的发展都将会继续下去。

无论如何，DPDK会继续成为基础设施加速中的主导框架，而DPDK项目也在寻找更多的贡献者和支持者。和所有开源项目一样，只有通过各种社区人员的参与才能成功。无论您是软件开发人员还是硬件开发人员，质量检查人员，开发测试工程师，文档专家或市场营销专家，DPDK该项目正在寻找新的贡献者和成员。如果需要了解更多信息，请访问[Linux Foundation DPDK](https://www.linuxfoundation.org/projects/dpdk/)网站查看详情。



