

计算服务 最佳实践

产品版本: v7.0.1
发布日期: 2026-08-18

目录

1 最佳实践	1
1.1 创建启用SR-IOV网卡的云主机	1
1.2 创建GPU计算加速型云主机	4
1.3 在线扩容云主机	6
1.4 开启内置Agent拓展云主机监控能力边界	8
1.5 创建网络优化型和计算网络优化型云主机	13
1.6 云主机挂载RoCE无损网络	15

1 最佳实践

1.1 创建启用SR-IOV网卡的云主机

背景描述

SR-IOV（Single Root I/O Virtualization）定义一种用于虚拟化PCIe（Peripheral Component Interconnect Express，快速外设组件互连）设备的标准化机制，该机制可以将单个PCIe以太网控制器虚拟化为多个PCIe设备。然后，每个PCIe设备可以绕过系统管理程序和虚拟交换层直接分配给云主机。

本文将详细介绍如何在该云平台中快速创建启用SR-IOV网卡的云主机，以缩短其数据传输路径，获得接近物理设备的I/O性能。

说明：

- 在同一云主机中，支持挂载多个SR-IOV网卡。
- 暂不支持千兆网卡。
- 为了保证虚拟网卡被云主机成功识别并正常工作，请在云主机创建成功后，为其安装虚拟网卡驱动。在获取并安装虚拟网卡驱动的过程中，如有疑问请联系网卡提供厂商获取帮助。
- 当冷迁移启用SR-IOV网卡的云主机时，若目标节点上可用的SR-IOV虚拟网卡数量不足，仍可以完成迁移，但是云主机状态为错误。此时，请重置云主机状态后，将其迁移到其他可用SR-IOV虚拟网卡数量充足的计算节点。

在冷迁移成功后，该云主机SR-IOV网卡的名称会发生变化，更新为新挂载的SR-IOV网卡。

前提条件

- 已确认计算节点插有提供SR-IOV功能的物理网卡。
- 已完成 [制作并上传镜像](#)、[创建安全组（可选）](#) 和 [云主机规格](#) 操作。
- 已完成 [前置条件准备](#) 操作。
- （可选）已根据需要完成 [创建可用区](#)、[创建SSH密钥对](#) 和 [创建云主机组](#) 操作。

操作步骤

1. 绑定物理网络。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[计算节点]，进入“计算节点”页面。
2. 单击待操作节点名称，进入其详情页面后，选择[网卡]页签，并勾选待操作物理网卡，单击 **绑定物理网络**，弹出“绑定物理网络”对话框。
3. 配置参数后，单击 **确认**，完成操作。

2. 虚拟化切割物理网卡。

1. 在上述计算节点详情页面的[网卡]页签中，勾选待操作物理网卡后，单击 **虚拟化切割**，弹出“虚拟化切割”对话框。
2. 配置参数后，单击 **确认**，完成操作。

3. 创建支持SR-IOV的网络。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[网络]-[网络]，进入“网络”页面。
2. 单击 **创建网络**，进入“创建网络”页面。
3. “网络类型”选择“内部网络”，“网络模式”选择“VLAN”，“物理网络”选择上述已绑定SR-IOV网卡的网络，并输入其他参数信息后，单击 **创建网络**，完成操作。其中，该页面中各参数的具体说明，请参考“网络”帮助中“网络”的相关内容。

4. 创建启用SR-IOV网卡的云主机。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
2. 单击 **创建云主机**，进入“创建云主机”的“基础配置”页面。
3. 配置参数后，单击 **下一步：网络配置**，进入“网络配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#)。
4. “虚拟网卡”选择上述新建的SR-IOV网络，勾选“启用SR-IOV”，并输入其他参数信息后，单击 **下一步：系统配置**，进入“系统配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#)。

说明：

启用SR-IOV的虚拟网卡，不支持安全组功能。

5. 配置参数后，单击 **下一步：确认配置**，进入“确认配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#)。

说明：

为了确保云主机成功创建，建议“高级选项”中的“所属计算节点”参数设置为“手动指定”，以免计算节点上可用的SR-IOV虚拟网卡数量不足，导致云主机创建失败。

6. 在“确认配置”页面中，确认云主机的配置信息后，单击 **创建云主机**，完成操作。

结果验证

1. 确认云主机网络已启用SR-IOV。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
2. 单击上述云主机名称，进入其详情页面后，选择[网络]页签，确认“SR-IOV”状态为“已启用”。

2. 确认虚拟网卡已绑定云主机。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[计算节点]，进入“计算节点”页面。
2. 单击上述计算节点名称，进入其详情页面后，选择[虚拟网卡VF]页签，确认上述物理网络的某一虚拟网卡已绑定上述新建的云主机。

1.2 创建GPU计算加速型云主机

背景描述

GPU计算加速型云主机不仅可以提供出色的浮点计算能力，还能够实现更灵活的资源部署，有效提高资源利用率，降低企业成本，以应对高并发、低时延的海量计算场景，如云游戏、VDI、VR/AR、AI推理以及机器学习等。

本文将详细介绍如何在该云平台中通过GPU虚拟化技术将物理GPU切割成更细粒度的vGPU后，使用已规划的vGPU云主机规格，快速创建GPU计算加速型的云主机。

前提条件

- 已成功购买并激活GPU。
- 已完成 [前置条件准备](#) 操作。
- （可选）已根据需要完成 [创建可用区](#)、[创建SSH密钥对](#) 和 [创建云主机组](#) 操作。

操作步骤

1. 创建vGPU云主机规格。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机规格]，进入“云主机规格”页面。
2. 选择[GPU计算加速型]页签后，单击 [创建云主机规格](#)，弹出“创建云主机规格”对话框。
3. 配置参数后，单击 [创建](#)，完成操作。其中，该页面中各参数的具体说明，请参考 [创建云主机规格](#)。

2. 开启计算节点的vGPU功能。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[计算节点]，进入“计算节点”页面。
2. 单击待操作节点名称，进入其详情页面后，单击 [更多操作](#)，在下拉列表中选择 [开启vGPU](#) 或 [关闭vGPU](#)，弹出“开启/关闭vGPU”提示框。
3. 单击 [开启](#) 或 [关闭](#)，完成操作。

3. 虚拟化切割物理GPU。

1. 在指定计算节点详情页面的[GPU]页签中，勾选待操作GPU后，单击 **虚拟化切割**，弹出“虚拟化切割”对话框。
 2. 配置参数后，单击 **确认**，完成操作。
4. 创建GPU计算加速型云主机。
1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
 2. 单击 **创建云主机**，进入“创建云主机”的“基础配置”页面。
 3. “规格”选择“GPU计算加速型”的云主机规格，并输入其他参数信息后，单击 **下一步：网络配置**，进入“网络配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#)。
 4. 配置参数后，单击 **下一步：系统配置**，进入“系统配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#)。
 5. 配置参数后，单击 **下一步：确认配置**，进入“确认配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#)。
 6. 在“确认配置”页面中，确认云主机的配置信息后，单击 **创建云主机**，完成操作。

1.3 在线扩容云主机

背景描述

当承载用户业务的云主机，已不能满足其上运行的应用对算力日益增长的需求时，通常需要通过横向扩容或纵向扩容的方式进行云上扩容。两种扩容方式的具体说明如下：

- 横向扩容：即通过创建更多的云主机，分担业务处理压力。例如：当一个Web应用服务器的CPU消耗和内存消耗都已达到85%以上，且客户侧的访问压力还在不断增长时，管理员可以通过在负载均衡下创建更多的Web应用服务器副本，支持不断增长的业务访问压力，使整个系统的处理能力得到横向扩展。
- 纵向扩容：即通过扩容原云主机的vCPU和内存，提高云主机自身算力。针对不支持横向扩容的应用系统，例如NFV负载（如虚拟路由器或网关等），只能通过对原云主机纵向扩容实现云上扩容。

针对云主机的纵向扩容，在V6.1.1之前版本的计算云产品中，只支持通过关机后再调整规格的方式实现，但是在V6.1.1及其之后版本的计算云产品中，支持利用云主机操作系统的热添加能力，在不停机的情况下调整云主机规格，实现在不影响业务连续性的前提下对云主机实施在线纵向扩容。本文将详细介绍如何利用云主机的在线调整规格功能，实现云主机的在线纵向扩容。

说明：

- 由于在线扩容场景有使用边界的限制，例如需要云主机操作系统支持热添加能力等，强烈建议用户在创建云主机时提前做好容量规划。如若后期确需扩容，请谨慎使用在线扩容云主机功能。
- 在线扩容场景的具体使用限制，请参考[在线调整云主机规格使用限制](#)。

前提条件

- 该云主机在创建时所用的镜像已开启在线变更规格属性。具体开启在线变更规格的操作，请参考“镜像”帮助中“修改配置”的相关内容。
- 该云主机操作系统支持热添加能力。

操作步骤

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
2. 勾选待操作云主机后，单击 更多 - 调整规格 ，弹出“调整云主机规格”对话框。

3. “调整规格方式”选择“在线调整”，并选择调整规格后，单击 **确定**，完成操作。

调整规格方式

☐ 离线调整  ☒ 在线调整 

调整规格

通用计算型

计算优化型

GPU计算加速型

点击选择过滤条件



规格名称	vCPU	内存
☐ 121	12	12 GiB
☐ 16C-16G	16	16 GiB
☐ 16C-32G	16	32 GiB
☐ 16C-64G	16	64 GiB
☒ 16C-8G	16	8 GiB

< 1 2 3 4 5 ... 10 > 5条/页

取消

确认

说明：

- 由于只支持在线调大云主机的CPU和内存规格，所以在上述对话框的规格列表中，仅展示比原规格大的规格（即：vCPU或内存至少有一个大于原规格中的对应值）。
- 目前仅支持针对通用计算型的云主机在线调整规格，其规格调整范围也仅限于通用计算型范围内。

结果验证

1. 在云平台页面确认云主机规格已调整。

- 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
- 单击上述云主机名称，进入其详情页面。在该页面的“详细配置”区域框中，确认“规格名称”参数的值已变更为上述新调整规格的名称。

2. 在云主机操作系统确认云主机规格已调整。

- 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
- 单击上述云主机的VNC图标，进入VNC Console页面，执行相应命令或操作查看操作系统当前CPU和内存的值，确认已成功调整。

1.4 开启内置Agent拓展云主机监控能力边界

背景描述

云主机作为客户应用的运行载体是平台多元算力的提供者之一，其各项指标是否工作正常与客户应用能否持续稳定运行息息相关，所以云主机的可观测性向来是客户比较关注的云平台基础能力之一。

在计算云产品的V6.1.1之前版本中，仅为云主机提供基础监控能力（基础监控模式），即通过调用虚拟化管理程序的API接口，从外部获取云主机的基础监控数据。该监控模式的优点是可以以非侵入的方式，简单方便地获取到云主机的基础监控数据，但不足之处是一些特定的监控数据获取不到或是特定场景下获取的某些监控指标不是特别精准，例如云主机的系统盘使用量无法获取，Windows操作系统场景下获取的内存使用数据不够精准等。

为了解决上述问题，在计算云产品的V6.1.1版本中，为云主机引入内置Agent的监控模式（详细监控模式），该模式需要云主机内预先安装qemu-guest-agent（以下简称qga），qga的工作原理及安装步骤请参考[如何安装QEMU Guest Agent](#)。在详细监控模式下，云主机的各项监控数据均由运行在云主机操作系统里的后台守护进程qga获取并返回给云平台控制面，所以在该模式下能获取到任意维度的云主机监控数据，同时获取到的数据会更加精准。

本文将详细介绍如何使用云主机的详细监控模式。

说明：

云主机的两种监控模式互有利弊，请根据客户实际业务需求酌情选择适合自己的监控模式。

前提条件

- 已完成 [制作并上传镜像](#)、[创建安全组（可选）](#) 和 [云主机规格](#) 操作。其中，在执行“制作并上传镜像”操作时，请确保已配置“启用云主机内置Agent”为“是”。
- 已完成 [前置条件准备](#) 操作。
- （可选）已根据需要完成 [创建可用区](#)、[创建SSH密钥对](#) 和 [创建云主机组](#) 操作。

操作步骤

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。

- 单击 **创建云主机** ，进入“创建云主机”的“基础配置”页面。
- 配置参数后，单击 **下一步：网络配置** ，进入“网络配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#) 。
- 在“高级选项”中勾选“开启云主机详细监控”，并配置其他参数后，单击 **下一步：系统配置** ，进入“系统配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#) 。

挂载ISO **+ 添加**

挂载USB 您需要手动指定物理节点后才能挂载USB设备。

启动顺序 **1**

云主机监控 ☐ **开启云主机详细监控**

说明：

当所选云主机启动镜像的“启用云主机内置Agent”属性未开启时，此处的“开启云主机详细监控”复选框将被置灰不可选。

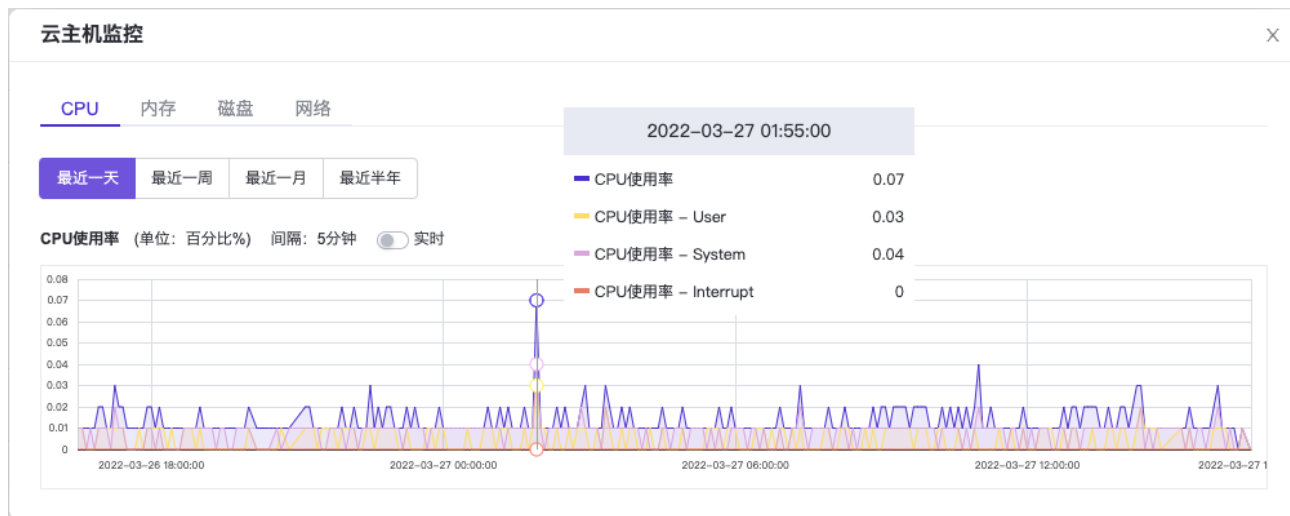
- 配置参数后，单击 **下一步：确认配置** ，进入“确认配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#) 。
- 在“确认配置”页面中，确认云主机的配置信息后，单击 **创建云主机** ，完成操作。
- (可选)如果云主机启动镜像中没有预先安装qga，可以在云主机启动后再安装，安装方法同样参考[如何安装QEMU Guest Agent](#)。

结果验证

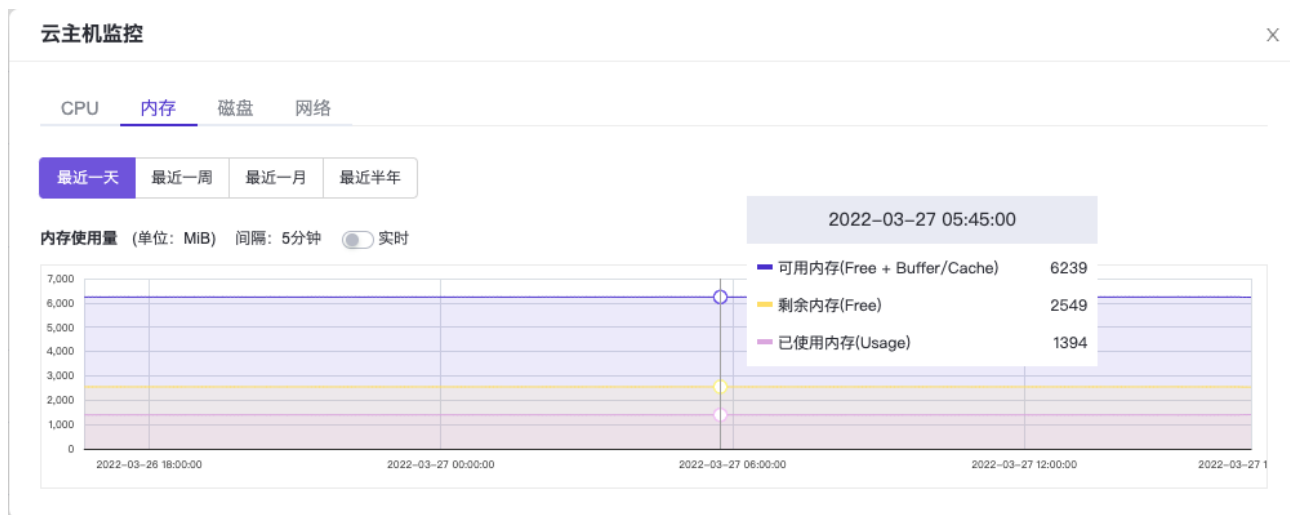
- 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
- 单击上述云主机的监控图标，查看云主机当前的监控数据。具体支持获取和展示的监控指标请参考[详细监控功能指标展示详情](#)。

页面的可视化展现效果如下。在该页面中，单击 **实时** 开关，可以在历史数据和实时数据视图之间切换展示。

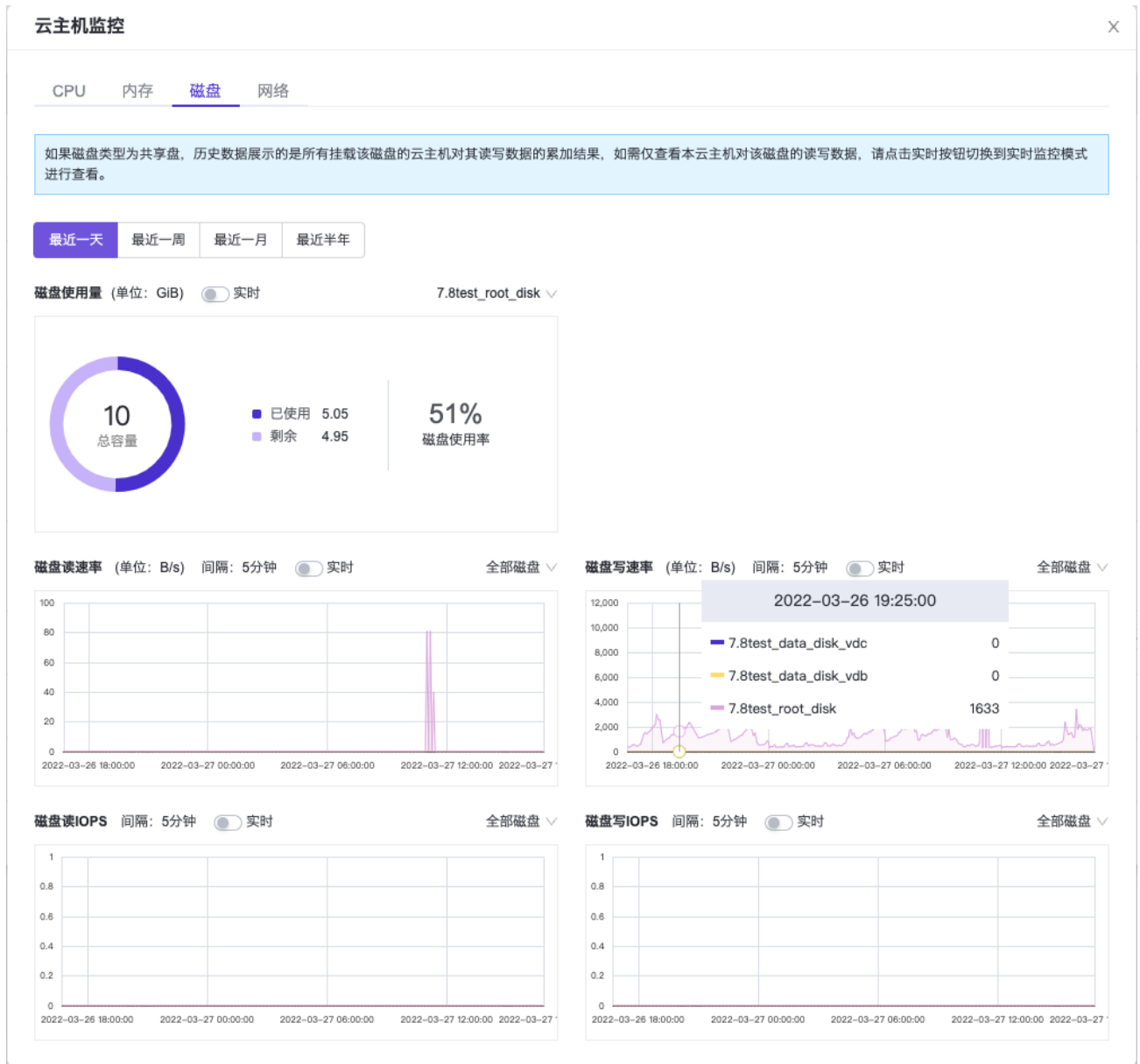
。 CPU监控：



。 内存监控：



。 磁盘监控：



网络监控：



1.5 创建网络优化型和计算网络优化型云主机

背景描述

DPDK (Data Plane Development Kit) 是一套工具集，可用于构建高性能数据平面应用程序。它可以帮助用户创建低延迟、高吞吐量的应用程序，以优化网络效率。云主机结合DPDK功能，可以为用户云主机提供更强的网络吞吐能力和网络包转发能力，以满足网络密集型应用的上云需求。

本文将详细介绍如何在云平台中快速创建DPDK类型的云主机。

前提条件

- 已成功激活计算节点DPDK功能。
- 已完成 [前置条件准备](#)、[创建云主机规格](#) 操作。
- (可选) 已根据需要完成 [创建可用区](#)、[创建SSH密钥对](#) 和 [创建云主机组](#) 操作。

操作步骤

1. 激活计算节点DPDK功能。

1. 请参考激活DPDK对接包制作及激活指南。
2. 计算节点在激活DPDK功能时，需确保节点上没有云主机。
3. 计算节点DPDK激活成功后，在节点的名称列上会添加DPDK标签。

2. 创建网络优化型云主机规格或者计算网络优化型云主机规格。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机规格]，进入“云主机规格”页面。
2. 选择[网络优化型]或[计算网络优化型]页签后，单击 **创建云主机规格**，弹出“创建云主机规格”对话框。
3. 配置参数后，单击 **创建**，完成操作。其中，该页面中各参数的具体说明，请参考 [创建云主机规格](#)。

3. 创建网络优化型云主机或计算网络优化型云主机。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
2. 单击 **创建云主机**，进入“创建云主机”的“基础配置”页面。

3. “规格”选择“网络优化型或者计算网络优化型”的云主机规格，并输入其他参数信息后，单击 **下一步：网络配置**，进入“网络配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#)。
4. 配置参数后，单击 **下一步：系统配置**，进入“系统配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#)。
5. 配置参数后，单击 **下一步：确认配置**，进入“确认配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#)。
6. 在“确认配置”页面中，确认云主机的配置信息后，单击 **创建云主机**，完成操作。

1.6 云主机挂载RoCE无损网络

背景描述

RoCE技术凭借其低延迟、高吞吐量和高效的CPU利用率，成为现代数据中心、高性能计算、存储网络和企业应用中的关键技术。通过合理的硬件选择和软件配置，以及有效的网络规划和管理，企业和研究机构可以充分发挥RoCE的优势，提高其计算和存储系统的整体性能。

RoCE技术优势

- 低延迟和高吞吐量
 - RoCE通过绕过内核协议栈，直接在用户态进行数据传输，大大减少了数据传输的延迟。
 - 提供高吞吐量的数据传输能力，适合需要高速数据交换的应用场景。
- 高效的CPU利用率
 - RoCE减少了CPU在数据传输过程中的负担，使得CPU资源可以更多地用于实际的计算任务，提升系统整体性能。
- 网络协议兼容性
 - RoCE在标准以太网上运行，无需专用的网络基础设施，兼容现有的网络设备和协议，降低部署成本和复杂性。
- 数据中心网络优化
 - 支持数据中心网络中的优先级流量控制（PFC）和数据中心桥接（DCB），进一步优化网络性能和可靠性。

本文将详细介绍如何在该云平台中快速创建启用RoCE网卡的云主机以及给云主机挂载RoCE网卡

说明：

- RoCE网卡MTU默认是5500，不建议使用Bond。
- 只支持SR-IOV切割RoCE网卡挂载给云主机使用，云主机中的虚拟网卡可以共享计算节点物理网卡上基于PFC/ECN实现的可信的拥塞控制。
- 在同一云主机中，支持挂载多个RoCE网卡。
- 只支持Mellanox CX4/5系列网卡。
- 为了保证虚拟网卡被云主机成功识别并正常工作，请在云主机创建成功后，为其安装虚拟网卡驱动。在获取并安装虚拟网卡驱动的过程中，如有疑问请联系网卡提供厂商获取帮助。

- 当冷迁移启用RoCE网卡的云主机时，若目标节点上可用的RoCE虚拟网卡数量不足，仍可以完成迁移，但是云主机状态为错误。此时，请重置云主机状态后，将其迁移到其他可用RoCE虚拟网卡数量充足的计算节点。

在冷迁移成功后，该云主机RoCE网卡的名称会发生变化，更新为新挂载的RoCE网卡。

平台预配置

- 加载RoCE网卡激活对接包。
- 计算节点BIOS中打开SMMU/IOMMU特性。
- 交换机侧配置PFC和ECN的拥塞控制。

前提条件

- 已确认计算节点插有提供RoCE功能的物理网卡。
- 已完成 [制作并上传镜像](#)、[创建安全组（可选）](#) 和 [云主机规格](#) 操作。
- 已完成 [前置条件准备](#) 操作。
- （可选）已根据需要完成 [创建可用区](#)、[创建SSH密钥对](#) 和 [创建云主机组](#) 操作。

操作步骤

1. 绑定物理网络。

- 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[计算节点]，进入“计算节点”页面。
- 单击待操作节点名称，进入其详情页面后，选择[网卡]页签，并勾选待操作物理网卡，单击 **绑定物理网络**，弹出“绑定物理网络”对话框。
- 配置参数后，单击 **确认**，完成操作。

2. 虚拟化切割物理网卡。

- 在上述计算节点详情页面的[网卡]页签中，勾选待操作物理网卡后，单击 **虚拟化切割**，弹出“虚拟化切割”对话框。
- 配置参数后，单击 **确认**，完成操作。

3. 创建支持RoCE的网络。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[网络]-[网络]，进入“网络”页面。
2. 单击 **创建网络** ，进入“创建网络”页面。
3. “网络类型”选择“内部网络”，“网络模式”选择“VLAN”，“物理网络”选择上述已绑定RoCE网卡的网络，并输入其他参数信息后，单击 **创建网络** ，完成操作。其中，该页面中各参数的具体说明，请参考“网络”帮助中“网络”的相关内容。

4. 创建启用RoCE网卡的云主机。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
2. 单击 **创建云主机** ，进入“创建云主机”的“基础配置”页面。
3. 配置参数后，单击 **下一步：网络配置** ，进入“网络配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#) 。
4. “虚拟网卡”选择上述新建的RoCE网络，勾选“启用RoCE”，并输入其他参数信息后，单击 **下一步：系统配置** ，进入“系统配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#) 。

说明：

启用RoCE的虚拟网卡，不支持安全组功能。

5. 配置参数后，单击 **下一步：确认配置** ，进入“确认配置”页面。其中，该页面中各参数的具体说明，请参考 [创建云主机](#) 。
6. 在“确认配置”页面中，确认云主机的配置信息后，单击 **创建云主机** ，完成操作。

结果验证

1. 确认云主机网络已启用RoCE。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[云主机]，进入“云主机”页面。
2. 单击上述云主机名称，进入其详情页面后，选择[网络]页签，确认“RoCE”状态为“已启用”。

2. 确认虚拟网卡已绑定云主机。

1. 在云平台的顶部导航栏中，依次选择[产品与服务]-[计算]-[计算节点]，进入“计算节点”页面。
2. 单击上述计算节点名称，进入其详情页面后，选择[虚拟网卡VF]页签，确认上述物理网络的某一虚拟网卡已绑定上述新建的云主机。

咨询热线：400-100-3070

北京易捷思达科技发展有限公司：

北京市海淀区西北旺东路10号院东区23号楼华胜天成科研大楼一层东侧120-123

南京分公司：

江苏省南京市雨花台区软件大道168号润和创智中心B栋一楼西101

上海office：

上海黄浦区西藏中路336号华旭大厦22楼2204

成都分公司：

成都市高新区天府五街168号德必天府五街WE602

邮箱：

contact@easystack.cn (业务咨询)

partners@easystack.cn(合作伙伴咨询)

marketing@easystack.cn (市场合作)

training@easystack.cn (培训咨询)

hr@easystack.cn (招聘咨询)