

VMware vSphere Bitfusion 安装指南

2021 年 5 月 11 日

更新以包含 VMware vSphere Bitfusion 3.5

VMware vSphere Bitfusion 3.0

您可以从 VMware 网站下载最新的技术文档:

<https://docs.vmware.com/cn/>。

VMware, Inc.
3401 Hillview Ave.
Palo Alto, CA 94304
www.vmware.com

**威睿信息技术（中国）有
限公司**
北京办公室
北京市
朝阳区新源南路 8 号
启皓北京东塔 8 层 801
www.vmware.com/cn

上海办公室
上海市
淮海中路 333 号
瑞安大厦 804-809 室
www.vmware.com/cn

广州办公室
广州市
天河路 385 号
太古汇一座 3502 室
www.vmware.com/cn

版权所有 © 2020-2021 VMware, Inc. 保留所有权利。 [版权和商标信息](#)

目录

关于 VMware vSphere Bitfusion 安装指南	4
1 了解 VMware vSphere Bitfusion	5
2 vSphere Bitfusion 安装过程概览	8
3 vSphere Bitfusion 服务器的系统要求	10
4 部署 vSphere Bitfusion 设备	12
准备部署 vSphere Bitfusion	12
查找 vCenter Server TLS 证书指纹	12
为 GPU 启用直通	13
启动 vSphere Bitfusion 设备部署	14
自定义 vSphere Bitfusion 设备 OVF 模板	15
将 GPU 直通到 vSphere Bitfusion 虚拟机	17
验证 vSphere Bitfusion 插件是否已注册到 vCenter Server	19
添加后续的 vSphere Bitfusion 服务器	20
5 安装 vSphere Bitfusion 客户端	23
在 CentOS 和 Red Hat 上安装 vSphere Bitfusion 客户端	25
在 Ubuntu 上安装 vSphere Bitfusion 客户端	25
6 启用 vSphere Bitfusion 客户端	27
启用 vSphere Bitfusion 客户端	28
生成客户端身份验证令牌	29
7 安装 NVIDIA 软件以与 vSphere Bitfusion 配合使用	31
从 Internet 安装适用于 vSphere Bitfusion 的 NVIDIA 软件	31
在气隙网络环境中安装 NVIDIA 软件	32
8 将准虚拟 RDMA 网络适配器与 vSphere Bitfusion 结合使用	36
9 升级 vSphere Bitfusion	38
将 vSphere Bitfusion 集群从 2.0 升级到 2.5	39
将 vSphere Bitfusion 集群从 2.5 升级到 3.0	42
将 vSphere Bitfusion 集群从 3.0 升级到 3.5	45

关于 VMware vSphere Bitfusion 安装指南

VMware vSphere Bitfusion 安装指南介绍如何在 VMware® vSphere 环境中安装和配置 VMware vSphere Bitfusion。

VMware 非常重视包容性。为了在客户、合作伙伴和内部社区中促进这一原则，我们采用包容性语言创建内容。

VMware vSphere Bitfusion 安装指南面向需要安装和配置 vSphere Bitfusion 且具有丰富经验的 vSphere 管理员。

目标读者

本信息面向需要安装 vSphere Bitfusion 的所有用户。本信息的目标读者为熟悉使用 VMware vSphere 和 vCenter Server 实施虚拟机技术和执行数据中心操作且具有丰富经验的 Linux 系统管理员。

了解 VMware vSphere Bitfusion

1

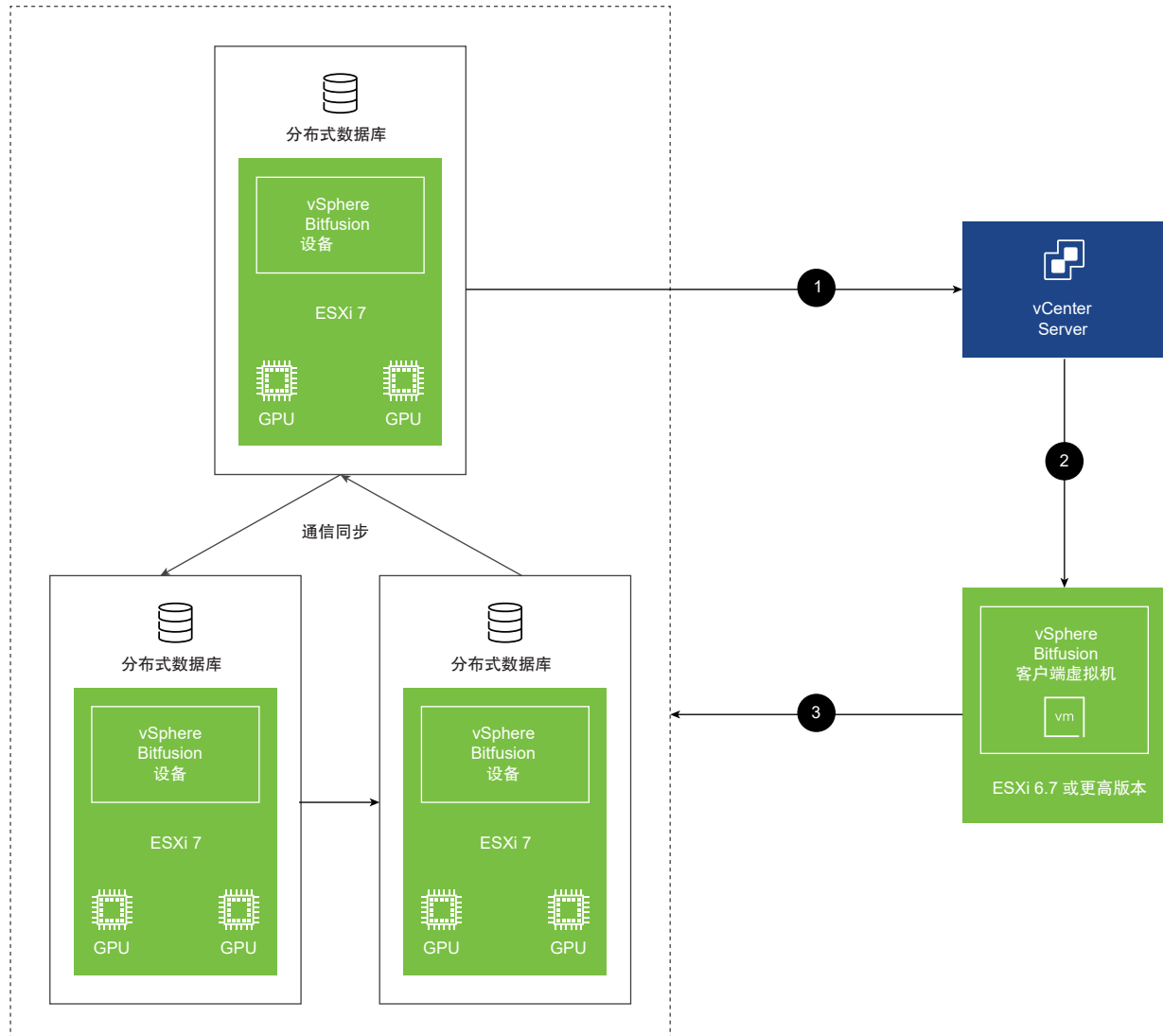
VMware vSphere Bitfusion 可虚拟化硬件加速器（例如图形处理单元 (GPU)），以提供可通过网络访问的共享资源池，从而支持人工智能 (AI) 和机器学习 (ML) 工作负载。

vSphere Bitfusion 架构

vSphere Bitfusion 采用客户端-服务器架构。该产品允许多个客户端虚拟机 (VM) 运行人工智能 (AI) 和机器学习 (ML) 应用程序，以便在运行 vSphere Bitfusion 服务器软件的虚拟机上共享对远程 GPU 的访问。您可以在 vSphere Bitfusion 客户端计算机上运行应用程序，而提供加速的 GPU 则安装在网络中的 vSphere Bitfusion 服务器计算机上。应用程序可以打开文件、分配内存并调用 CUDA，就像在具有本地 GPU 的计算机上运行一样。

下图是一个小型 vSphere Bitfusion 集群的示例，例如一组 vSphere Bitfusion 服务器-客户端计算机和交换网络中的 vCenter Server。最小 vSphere Bitfusion 集群配置为一个客户端、一个服务器和一个 vCenter Server。您可以创建具有多个客户端和多个服务器的大型集群。

图 1-1. 小型 vSphere Bitfusion 集群的示例



- 1 主 vSphere Bitfusion 服务器将向 vCenter Server 注册 vSphere Bitfusion 插件。
- 2 vSphere Bitfusion 插件可启用 vSphere Bitfusion 客户端虚拟机。
- 3 vSphere Bitfusion 客户端有权访问 vSphere Bitfusion 集群中的所有 vSphere Bitfusion 服务器。

注 在使用 VMware vSphere Bitfusion 之前，您必须部署 vSphere Bitfusion 服务器以及安装并启用 vSphere Bitfusion 客户端。有关详细信息，请参见《VMware vSphere Bitfusion 安装指南》。

vSphere Bitfusion 功能

在 vSphere Bitfusion 客户端上启动 AI 或 ML 应用程序时，vSphere Bitfusion 会拦截应用程序的 CUDA 调用并查看调用的数据和数据指针。vSphere Bitfusion 服务器不需要连接到数据，只需连接到 vSphere Bitfusion 客户端。客户端会将数据和其余的 CUDA 调用传输至服务器。vSphere Bitfusion 服务器将处理调用，并将结果返回给客户端。

运行 AI 和 ML 应用程序时，vSphere Bitfusion 可以执行以下任务。

- 从 vSphere Bitfusion 服务器动态分配和访问 GPU 资源。

应用程序可以共享并非专用于单个计算机的 GPU 资源，并且可以在配置的计算机、容器和环境中运行每个应用程序。应用程序将消耗网络上 vSphere Bitfusion 服务器池中的 GPU 加速服务，但仅在应用程序或会话运行期间消耗资源。在应用程序或会话完成后，GPU 将返回到池。

- 访问 GPU 资源的分区，以与其他应用程序进行并发共享。

共享 GPU 的另一个选项是对 GPU 进行分区。物理 GPU 的内存可以分为任意大小的部分，并同时分配给不同的应用程序。vSphere Bitfusion 使用介入技术执行共享。vSphere Bitfusion 会截获通常寻址本地加速器（在 PCIe 主机总线上）的 API 调用，并通过网络发送 API 调用及相关数据。vSphere Bitfusion 为 AI 和 ML 应用程序提供共享服务，并支持 CUDA API 将 NVIDIA GPU 作为目标。

vSphere Bitfusion 组件

vSphere Bitfusion 服务器

vSphere Bitfusion 服务器在具有本地安装 GPU 的 ESXi 主机上作为 VMware 设备运行，即具有预打包软件和服务的预配置虚拟机 (VM)。服务器需要访问本地 GPU，通常通过 VMware vSphere® DirectPath I/O™ 访问。

vSphere Bitfusion 客户端

vSphere Bitfusion 客户端在运行 AI 和 ML 应用程序的虚拟机上运行。

vSphere Bitfusion 插件

vSphere Bitfusion 服务器将向 VMware vCenter Server 注册 vSphere Bitfusion 插件。该插件可监控和管理 vSphere Bitfusion 客户端与服务器。

vSphere Bitfusion 个集群

vSphere Bitfusion 集群是 vCenter Server 实例中所有 vSphere Bitfusion 服务器和客户端的集合。

vSphere Bitfusion 组

vSphere Bitfusion 客户端会在安装过程中创建一个 vSphere Bitfusion 组。只有该组的成员才能使用 vSphere Bitfusion。某些配置文件设置有适当的权限，组成员将继承相应的限制，以高效使用 vSphere Bitfusion。

vSphere Client

通过 vSphere Client，可以使用 Web 浏览器连接到 vCenter Server 实例，以便管理 vSphere 基础架构。您可以通过 vSphere Client 访问 vSphere Bitfusion 插件。

命令行界面 (CLI)

您可以使用命令行界面 (CLI) 命令管理 vSphere Bitfusion 服务器和客户端。

vCenter Server

vCenter Server 是服务器管理软件，提供了一个集中式平台来控制您的 vSphere 环境。

vSphere Bitfusion 安装过程概览

2

VMware vSphere Bitfusion 是一个复杂的产品，具有多个要安装和设置的组件。为确保成功部署 vSphere Bitfusion，需要了解所需的任务序列。

图 2-1. vSphere Bitfusion 安装 workflow



执行如下步骤，可以成功安装 vSphere Bitfusion。

- 1 阅读 vSphere Bitfusion 发行说明。
- 2 请确保您的环境满足最低系统要求以及运行您计划运行的人工智能和机器学习工作负载所需的任何其他资源。请参见第 3 章 [vSphere Bitfusion 服务器的系统要求](#)和第 5 章 [安装 vSphere Bitfusion 客户端](#)。
- 3 安装主 vSphere Bitfusion 服务器。请参见第 4 章 [部署 vSphere Bitfusion 设备](#)。
- 4 根据需要安装其他 vSphere Bitfusion 服务器。请参见[添加后续的 vSphere Bitfusion 服务器](#)。
- 5 安装 vSphere Bitfusion 客户端。请参见第 5 章 [安装 vSphere Bitfusion 客户端](#)。
- 6 启用 vSphere Bitfusion 客户端。请参见第 6 章 [启用 vSphere Bitfusion 客户端](#)。

vSphere Bitfusion 服务器的系统要求

3

vSphere Bitfusion 需要一个 ESXi 主机，以用于安装 vSphere Bitfusion 服务器。

vSphere Bitfusion 服务器的系统要求

vSphere Bitfusion 服务器必须在符合以下系统要求的 vSphere 部署上运行。

- vSphere Bitfusion 服务器设备的最低磁盘空间要求为 50 GB。
- 运行 vSphere Bitfusion 服务器的 ESXi 主机版本必须为 7.0 或更高版本。
- vSphere Bitfusion 服务器的最低内存要求为 32 GB 或服务器上安装的 GPU 内存总量的 150%（取较高者）。
- vSphere Bitfusion 服务器的最低虚拟 CPU (vCPU) 要求是 GPU 卡数乘以 4。
- 支持 TCP/IP 或 RoCE 的网络（PVRDMA 适配器）。
- 任何访问两个或更多 GPU 的计算机至少要有 10 Gbps 的带宽。
- 客户端计算机与服务器虚拟机之间的延迟不得超过 50 微秒。这并非严格要求，但延迟越低，vSphere Bitfusion 部署的性能越好。
- 所有 vSphere Bitfusion 服务器都必须连接到同一组有效的 NTP 服务器。

vSphere Bitfusion 服务器所需的端口

使用拒绝列表或防火墙规则确认这些端口未被阻止。通信需要这些端口。

自 vSphere Bitfusion 3.5 起，出于安全考虑，将默认禁用 TLSv1.0 和 TLSv1.1 协议。有关详细信息，请参见知识库文章 [2145796](#)。

端口	描述
443	此端口用于 vSphere Bitfusion 与 vSphere Client 之间的通信。
7000 和 7001	Apache Cassandra 使用此端口与 Cassandra 集群通信。
9042	Apache Cassandra 使用此端口与本机协议客户端进行通信。
9142	Apache Cassandra 将此端口用于 Cassandra Thrift API。
9160	此端口由 Apache Cassandra 使用。CQL 本机传输会在此端口上侦听 vSphere Bitfusion 客户端，并在需要加密连接和未加密连接时使用此端口。

端口	描述
45201 - 46225	vSphere Bitfusion 客户端使用这些端口与处理 CUDA 请求的 vSphere Bitfusion 服务器进程通信。
55001 - 55201	vSphere Bitfusion 客户端使用这些端口与在 vSphere Bitfusion 服务器上运行的任务特定分派程序进程通信。该进程将请求会话，并启动 vSphere Bitfusion 服务器上用于处理工作负载的服务工作程序。
56001	此端口用于 vSphere Bitfusion 内部通信。vSphere Bitfusion 服务器通过此端口相互通信，vSphere Bitfusion 客户端使用此端口在 vSphere Bitfusion 服务器上启动任务。

vCenter Server 的 Web 浏览器要求

要使用 vSphere Bitfusion，您需要一个受 vCenter Server 支持的 Web 浏览器版本。有关详细信息，请参见 [vSphere Client 软件要求](#)。

vSphere Bitfusion 兼容性和互操作性

有关与 vSphere Bitfusion 兼容的版本、型号和产品的列表，请参见 [VMware vSphere Bitfusion 兼容性和互操作性](#) 页面。

部署 vSphere Bitfusion 设备

4

vSphere Bitfusion OVA 文件包含构成 vSphere Bitfusion 服务器的压缩开放式虚拟化格式文件。设置 vSphere 环境后，请下载 vSphere Bitfusion OVA 文件，并将其作为 vSphere 设备进行部署。

可以采用开放式虚拟格式 (OVF) 和开放式虚拟设备 (OVA) 部署虚拟机、虚拟设备和 vApp。通过部署 OVF 或 OVA 模板，您可以将预配置的虚拟机或 vApp 添加到 vCenter Server 或 ESXi 清单中。部署 OVF 或 OVA 模板类似于从模板部署虚拟机。但是，可从 vSphere Client 可访问的任何本地文件系统或从远程 Web 服务器部署 OVF 或 OVA 模板。

vSphere Bitfusion 作为 OVA 文件进行分发。要了解有关 OVA 或 OVF 文件的更多信息，请参见《vSphere 虚拟机管理》文档。

本章讨论了以下主题：

- 准备部署 vSphere Bitfusion
- 启动 vSphere Bitfusion 设备部署
- 自定义 vSphere Bitfusion 设备 OVF 模板
- 将 GPU 直通到 vSphere Bitfusion 虚拟机
- 验证 vSphere Bitfusion 插件是否已注册到 vCenter Server
- 添加后续的 vSphere Bitfusion 服务器

准备部署 vSphere Bitfusion

开始 vSphere Bitfusion 部署之前，必须执行几项任务。完成这些任务是部署过程的必备条件。

查找 vCenter Server TLS 证书指纹

vCenter Server TLS 证书指纹是 vCenter Server TLS 证书的安全哈希算法 (SHA1) 签名。

执行以下过程，查找环境对应的 TLS 证书指纹。您必须先复制指纹，稍后再将其添加到 OVF 模板的部署属性中。

步骤

- 1 打开 Web 浏览器，然后输入 vSphere Client 的 URL：
`https://vcenter_server_ip_address_or_fqdn/ui`
- 2 输入具有 vCenter Server 权限的用户的凭据，然后单击**登录**。

3 查找 TLS 证书指纹。

- 在 Google Chrome 中查找 TLS 证书指纹。
 - a 单击 Web 地址左侧的**安全**图标，然后选择**证书**。
 - b 在**证书**对话框中，单击**详细信息**选项卡。
 - c 在**详细信息**选项卡上，向下滚动列表，在**字段**列中选择**指纹**。
 - d vCenter Server TLS 证书指纹显示在列表框下方的文本框中。
- 在 Mozilla Firefox 中查找 TLS 证书指纹。
 - a 单击 Web 地址左侧的**安全**图标，选择连接状态右侧的箭头，然后单击**详细信息**。
 - b 在**页面信息**对话框的**安全**选项卡上，选择**查看证书**。
 - c 在**证书**浏览器选项卡上，指纹显示在**指纹**部分中。

后续步骤

自定义 vSphere Bitfusion 设备 OVF 模板

为 GPU 启用直通

要在 vSphere Bitfusion 服务器中使用 GPU，必须在直通模式下启用该设备。该操作允许服务器绕过 ESXi 管理程序直接访问 GPU，从而获得与本机系统上 GPU 类似的性能级别。

使用直通模式时，每个 GPU 设备都专用于 vSphere Bitfusion 服务器的虚拟机 (VM)。您可以在直通模式下使用多个物理 GPU。必须为计划在 vSphere Bitfusion 服务器中使用的所有 GPU 设备执行以下过程。

必备条件

- 确认您的服务器供应商支持您的 GPU 设备。
- 确认您的 GPU 可在直通模式下使用。
- 确认您已为 vSphere Bitfusion 服务器创建了虚拟机。
- 验证 GPU 设备是否可以映射总大小达到或超过 16 GB 的内存区域。

注 通常，高端 GPU 卡需要大量的内存映射。这些内存映射是在设备的 PCI 基本地址寄存器 (BAR) 中指定的。您可以在 GPU 的供应商文档中找到相关信息。

过程

- 1 如果 GPU 需要 16 GB 或更多的内存映射，请在 ESXi 主机的 BIOS 设置中为 GPU 启用直通。
通常情况下，设置的名称为 Above 4G decoding、Memory mapped I/O above 4GB 或 PCI 64-bit resource handing above 4G。
- 2 在 ESXi 主机上为 GPU 启用直通。
 - a 在 vSphere Client 中，右键单击 ESXi 主机，然后选择**设置**。
 - b 在**配置**选项卡上，选择**硬件 > PCI 设备**，然后单击**配置直通**。

- c 在**编辑 PCI 设备可用性**对话框中，在“ID”列中选中 GPU 设备对应的复选框。
 - d 单击**确定**。
- GPU 将在**启用直通的设备**选项卡上显示。
- e 重新引导 ESXi 主机。

3 在虚拟机的引导选项中启用 UEFI 或 EFI。

vSphere Bitfusion 服务器的虚拟机必须处于 EFI 或 UEFI 模式才能正确使用 GPU。

- a 在 vSphere Client 中，右键单击虚拟机。
- b 选择**编辑设置 > 虚拟机选项 > 引导选项**。
- c 从**固件**下拉菜单中选择 UEFI 或 EFI。
- d 单击**确定**。

在 vSphere Bitfusion 设备的部署过程中，可以将 GPU 直通到 vSphere Bitfusion 服务器的虚拟机。

启动 vSphere Bitfusion 设备部署

要开始 vSphere Bitfusion 部署，请指定有关产品的信息，包括产品名称、位置和在 vSphere 环境中的存储。

安装其他 vSphere Bitfusion 服务器时，必须在安装的第一个或主 vSphere Bitfusion 服务器中注册后续服务器。在打开 vSphere Bitfusion 虚拟机的电源之前，必须执行此额外步骤。请参见[添加后续的 vSphere Bitfusion 服务器](#)。

前提条件

- 从 <https://my.vmware.com/downloads/> 下载 vSphere Bitfusion OVA 文件。
- 确认要在其上部署 vSphere Bitfusion 设备的 vSphere 环境满足最低系统要求。请参见[第 3 章 vSphere Bitfusion 服务器的系统要求](#)。
- 确认您能够以管理员身份登录到 vSphere Client。
- 确认要在其上部署 vSphere Bitfusion 设备的 ESXi 主机正在运行。

步骤

- 1 以管理员身份登录 vSphere Client。
- 2 在 vSphere Client 中，右键单击要在其上部署 vSphere Bitfusion 设备的 ESXi 主机，然后选择**部署 OVF 模板**。
- 3 在**选择 OVF 模板**页面上，输入 ova 文件的 URL 或浏览到该文件，然后单击**下一步**。
- 4 在**选择名称和文件夹**页面上，输入 vSphere Bitfusion 虚拟机的名称，并选择部署位置，然后单击**下一步**。
- 5 在**选择计算资源**页面上，选择要运行已部署的虚拟机模板的资源，然后单击**下一步**。

- 在**查看详细信息**页面上，验证 OVF 模板详细信息并单击**下一步**。

查看详细信息页面将显示一条警告，指明 vSphere Bitfusion OVF 使用高级配置值，可能会带来安全风险。触发警示的配置值是 `pciPassthru.use64bitMMIO = true` 和

`pciPassthru.64bitMMIOSizeGB = 256`。第一个参数将为 GPU 设备启用 PCI 直通，GPU 设备需要具有 16 GB 或更多的内存映射，第二个参数将配置内存映射 I/O (MMIO) 大小 256 GB。您可以稍后在 vSphere Bitfusion 虚拟机的设置中调整此值。

- 在**选择存储**页面上，定义在哪里以及如何存储已部署 OVF 模板的文件，然后单击**下一步**。
- 在**选择网络**页面上，选择要用于网络适配器 1 的网络，然后单击**下一步**。

与网络适配器 1 结合使用的网络传输管理流量。您可以同时将该网络用于传输数据流量，也可以在以后添加额外的网络适配器。

后续步骤

vSphere Bitfusion 设备需要多个自定义参数。完成**自定义模板**页面，以自定义 OVF 模板的部署属性。请参见[自定义 vSphere Bitfusion 设备 OVF 模板](#)。

自定义 vSphere Bitfusion 设备 OVF 模板

在 vSphere Bitfusion 设备部署过程中，必须在**部署 OVF 模板**对话框中指定多个自定义参数。

在**部署 OVF 模板**对话框的**自定义模板**页面上，可以自定义 OVF 模板的部署属性。

前提条件

- 确认您具有 vCenter Server TLS 证书指纹。请参见[查找 vCenter Server TLS 证书指纹](#)。
- 验证您使用的 DNS 和 NTP 服务是否已在您的环境中正确设置。
 - 如果使用 DHCP，请验证它是否提供 DNS 和 NTP 地址。
 - 如果不使用 DHCP，请在 OVF 模板中指定 DNS 和 NTP 服务器地址。

注 时钟同步对于 vSphere Bitfusion 功能非常重要。

步骤

- 在 **Bitfusion 服务器设置** 部分中，指定服务器的主机名。

主机名的有效字符为 ASCII 字符：A-Z（包括大小写）、数字 0-9 和连字符 (-)。主机名不能以连字符开头。主机名存储在 `/etc/hostname` 中。

- 在 **Bitfusion 服务器设置** 部分中，输入在 Web 浏览器的地址栏中显示的 vCenter Server GUID 和 URL。

例如，如果导航栏显示 URL `https://example.vslab.local/ui/app/vm;nav=h/urn:vmomi:VirtualMachine:vm-4450:612d27ff-d297-4573-bdc0-2c0dac8589a5/summary`，则 vCenter Server URL 为 `https://example.vslab.local`，GUID 为 `612d27ff-d297-4573-bdc0-2c0dac8589a5`。

- 3 在 **Bitfusion 服务器设置** 部分中，输入要在其上部署 vSphere Bitfusion OVF 模板的 vCenter Server 实例的用户名和密码。

- 4 在 **Bitfusion 服务器设置** 部分中，输入 vCenter Server TLS 证书指纹。

- 5 在 **凭据** 部分指定客户密码。

部署完成后，使用客户用户帐户通过控制台 shell 或 SSH 登录到 vCenter Server Appliance。

- 6 （可选）在 **NVIDIA 驱动程序** 部分中，选中 **下载并安装 NVIDIA 驱动程序** 复选框以接受 NVIDIA 许可证。

通过接受 NVIDIA 许可证，vSphere Bitfusion 会在首次引导虚拟机期间下载并安装 NVIDIA 驱动程序、CUDA 库和 NVIDIA Fabric Manager。

注 如果在无法访问 Internet 的环境（例如，使用气隙网络）中运行 vSphere Bitfusion，请不要选中该复选框。必须在部署 vSphere Bitfusion 设备后手动下载并安装 NVIDIA 软件。

- 7 在 **网络适配器** 部分中，指定环境的网络连接配置。

必须指定用于管理和数据流量的网络适配器 1 的配置。网络适配器 1 必须连接到与 vCenter Server 实例通信的网络。

网络适配器 2、3 和 4 是可选的，并且仅用于数据流量。每个网络适配器都必须连接到单独的网络。vSphere Bitfusion 会选择可将数据最高效地传输到 vSphere Bitfusion 服务器的网络。

选项	描述
IPv4 地址	输入网络适配器的 IPv4 地址。如果使用 DHCP，请将此文本框留空。 注 IPv6 不受支持。
网络 CIDR 前缀	输入网络无类别域际路由 (CIDR) 设置。 例如，如果您的网络使用网络掩码 /24，请从下拉菜单中选择 24 (255.255.255.0)。
MTU	输入 MTU 大小。默认值为 1500。为获得最佳性能，请将 MTU 大小指定为网络硬件支持的最大 MTU 大小。 注 如果将 MTU 大小设置为大于 1500，请确认数据中心的网络交换机是否已启用巨型帧。
网关	输入要用于设备的网络网关地址。如果使用 DHCP，请将此文本框留空。
DNS	输入要用于设备的 DNS 服务器地址。如果使用 DHCP，请将此文本框留空。
DNS 搜索域	输入要用于设备的 DNS 搜索域地址。如果使用 DHCP，请将此文本框留空。
NTP	输入要用于设备的 NTP 服务器地址。如果您使用的是 DHCP，并且 DHCP 服务器支持发送 NTP 服务器信息，请将此文本框留空。
配置网络适配器	如果要配置网络适配器 2，请选中该复选框。对每个后续网络适配器重复此操作。

- 8 单击 **下一步**。

- 9 在 **即将完成** 页面上，检查 vSphere Bitfusion 服务器配置并单击 **完成**。

结果

用于创建 vSphere Bitfusion 设备的新任务将显示在“近期任务”窗格中。该任务完成后，会在所选资源上创建新设备。

后续步骤

- 将 GPU 直通到 vSphere Bitfusion 虚拟机。请参见[将 GPU 直通到 vSphere Bitfusion 虚拟机](#)。
- 如果您在首次引导期间没有选择下载并安装 NVIDIA 驱动器、CUDA 库和 NVIDIA Fabric Manager，则必须手动安装此软件。请参见[第 7 章 安装 NVIDIA 软件以与 vSphere Bitfusion 配合使用](#)。
- 您可以添加更多用于数据流量的网络适配器。请参见 VMware vSphere Bitfusion 用户指南中的修改 vSphere Bitfusion 服务器的网络设置。

将 GPU 直通到 vSphere Bitfusion 虚拟机

VMDirectPath I/O 允许客户机操作系统直接访问 GPU，从而绕过 ESXi Hypervisor。通过使用直通设备，可以更高效地使用资源，并提高 vSphere Bitfusion 环境的性能。启用 GPU 直通可在 vSphere 上提供接近于其本机系统的性能级别。

前提条件

- 确认您具有计划执行的任务所需的特权。
 - 验证您是否具有**虚拟机.配置.添加或移除设备**特权。
 - 验证您是否具有**虚拟机.配置.高级配置**特权。
 - 如果您计划在编辑虚拟机时增加内存预留，请确认您具有 **虚拟机.配置.更改资源** 特权。
 - 验证您是否具有**虚拟机.配置.更改内存**特权。
- 确认 vSphere Bitfusion 服务器的虚拟机电源已关闭。
- 要使用 DirectPath，请确认已在 ESXi 主机的 BIOS 中启用 Intel Virtualization Technology for Directed I/O (VT-d) 或 AMD I/O Virtualization Technology (IOMMU)。
- 验证 GPU PCI 设备是否已连接到主机并标记为可用于直通。请参见[GPU 启用直通](#)。
- 如果 ESXi 主机配置为从 USB 设备引导，或者活动 coredump 分区配置为位于通过 USB 通道连接的 USB 设备或 SD 卡上，请禁用 USB 控制器直通。

注 对于从经由 USB 通道连接的 USB 设备或 SD 卡进行引导的 ESXi 主机，VMware 不支持 USB 控制器直通。活动 coredump 分区配置为位于通过 USB 通道连接的 USB 设备或 SD 卡上的配置也不受支持。有关信息，请参见 <http://kb.vmware.com/kb/1021345>。

步骤

1 添加 GPU 设备。

- a 在 vSphere Client 中，右键单击清单中的 vSphere Bitfusion 虚拟机，然后选择**编辑设置**。
- b 在**虚拟硬件**选项卡上，单击**添加新设备**按钮。
- c 从下拉菜单中的**其他设备**下，选择 **PCI 设备**。
- d 展开 **新 PCI 设备** 部分，然后选择访问类型。
- e 在**新 PCI 设备** 部分中，从 **PCI 设备** 下拉菜单中选择一个 GPU。

注 默认情况下，将为每个新 GPU 列出相同的 PCI 设备地址。添加多个设备时，必须分别为每个设备选择 PCI 地址。

- f 单击**确定**。

2 为 ESXi 主机配置 CPU 和内存资源。

如果 ESXi 主机专用于 vSphere Bitfusion 服务器，请将 CPU 和内存设置为其最大值。如果主机不是专用于 vSphere Bitfusion，请将最小 CPU 值指定为 GPU 数量乘以 4，将最小内存值指定为汇总 GPU 卡内存的 1.5 倍或 32 GB（取较大者）。

- a 在 vSphere Client 中，右键单击 vSphere Bitfusion 虚拟机，然后选择**编辑设置**。
- b 展开 **CPU** 部分，然后编辑资源。
- c 展开**内存**部分，然后编辑资源。
- d 在**内存**下，选中**预留所有客户机内存 (全部锁定)** 复选框。
- e 单击**确定**。

3 调整内存映射 I/O (MMIO) 大小。

默认情况下，vSphere Bitfusion 安装程序会配置 256 GB 的 MMIO 大小。要计算必须为 MMIO 预留的实际内存，对于两个和三个内存均为 16 GB 的卡，请考虑以下 MMIO 内存计算。

- $2 \times 16 \text{ GB} = 32$. 将 32 GB 向上舍入到下一个 2 次幂，所需的内存映射 I/O 大小为 64 GB。
- $3 \times 16 \text{ GB} = 48$. 将 48 GB 向上舍入到下一个 2 次幂，所需的内存映射 I/O 大小为 64 GB。

- a 在 vSphere Client 中，右键单击 vSphere Bitfusion 虚拟机，然后选择**编辑设置**。
- b 单击**虚拟机选项**，然后展开**高级**部分。
- c 在**配置参数**下，单击**编辑配置**。
- d 在**配置参数**对话框中，找到参数 `pciPassthru.64bitMMIOSizeGB`，然后输入 MMIO 大小（以千兆字节为单位）。
- e 单击**确定**。

4 （可选）生成虚拟机的快照。

快照在执行时会捕获虚拟机的状况。如果在启动虚拟机时出现错误，则可以从快照恢复 vSphere Bitfusion 安装。

- a 在 vSphere Client 中，右键单击 vSphere Bitfusion 虚拟机，然后选择**快照 > 生成快照**。
- b 输入快照的名称和描述。
- c 单击**创建**。

5 如果要部署后续 vSphere Bitfusion 服务器，请启用该服务器。

注 当您启用其他 vSphere Bitfusion 服务器时，该服务器可识别您之前部署的主 vSphere Bitfusion 服务器。如果在打开 vSphere Bitfusion 虚拟机电源之前未执行此步骤，则后续服务器将成为主服务器，并覆盖 vSphere Bitfusion 集群的配置。

- a 在 vSphere Client 中，右键单击清单中的虚拟机，然后选择 **Bitfusion > 启用 Bitfusion**。
- b 在**启用 Bitfusion** 对话框中，选择**对于服务器，这将允许将其用作 GPU 服务器**单选按钮，然后单击**启用**。

6 在 vSphere Client 中，右键单击 vSphere Bitfusion 虚拟机，然后选择**电源 > 打开电源**。

如果要打开多个 vSphere Bitfusion 服务器的电源，请一次打开一个。每打开一个服务器的电源后，请等待三分钟或更长时间再打开下一个服务器的电源。

结果

虚拟机打开电源后，先让它运行 10 分钟或更长时间，再执行任何其他配置任务或操作。在此期间，虚拟机将注册到 vCenter Server，如果您选择此安装选项，它还将下载和安装 NVIDIA 驱动程序。

后续步骤

- 如果首次引导过程中没有选择下载并安装 NVIDIA 驱动程序，现在可以手动安装该驱动程序。请参见[从 Internet 安装适用于 vSphere Bitfusion 的 NVIDIA 软件](#)。
- 验证 vSphere Bitfusion 插件是否已注册到 vCenter Server。请参见[验证 vSphere Bitfusion 插件是否已注册到 vCenter Server](#)。

验证 vSphere Bitfusion 插件是否已注册到 vCenter Server

部署 vSphere Bitfusion 服务器并安装 NVIDIA 驱动程序后，验证 vSphere Bitfusion 插件是否在 vSphere Client 中显示。

前提条件

- 在验证 vSphere Bitfusion 插件是否在 vSphere Client 中显示之前，请先重新启动 vSphere Bitfusion 虚拟机。

注 如果您选择在 vSphere Bitfusion 部署过程中安装 NVIDIA 驱动程序，则服务器将重新启动。

- 允许虚拟机运行 10 分钟或更长时间后，再执行任何其他配置任务或操作。在此期间，将向 vCenter Server 注册虚拟机。

步骤

- 1 打开 Web 浏览器，然后输入 vCenter Server 实例的 URL: `https://vcenter_server_ip_address_or_fqdn`。
- 2 选择**启动 vSphere Client (HTML5)**。
- 3 输入具有 vCenter Server 权限的用户的凭据，然后单击**登录**。
- 4 （可选）要更新当前 vSphere Client 视图中的所有数据，请单击刷新图标 ()。
- 5 在 vSphere Client 中，选择**菜单 > Bitfusion**。
将加载 vSphere Bitfusion 插件。

后续步骤

如果 vSphere Bitfusion 插件已正确注册，您可以部署其他 vSphere Bitfusion 服务器和客户端。

如果 vSphere Bitfusion 插件无法正常工作或不可用，请验证是否在 OVF 模板中正确配置了 NTP、DNS、GUID 和 SHA1 设置。请参见[自定义 vSphere Bitfusion 设备 OVF 模板](#)。

添加后续的 vSphere Bitfusion 服务器

需要更多 GPU 资源时，可以向 vSphere Bitfusion 集群中添加更多服务器。

主 vSphere Bitfusion 服务器启动后，vSphere Bitfusion 将在 vCenter Server 中注册 vSphere Bitfusion 插件，从而形成一个 vSphere Bitfusion 集群，其中包含一个 vSphere Bitfusion 服务器。注册 vSphere Bitfusion 插件后，可以按照此过程中的步骤添加后续服务器。该 vSphere Bitfusion 插件使用主服务器的配置数据，因此可以加快后续服务器的部署。

或者，您也可以按照主服务器的部署过程在 vSphere Bitfusion 集群中添加新服务器。您可以在虚拟机 (VM) 上部署 vSphere Bitfusion 设备，自定义 vSphere Bitfusion OVF 模板，将 GPU 传递到 vSphere Bitfusion 服务器虚拟机，以及将虚拟机启用为 vSphere Bitfusion 服务器。

添加的 vSphere Bitfusion 服务器必须与第一个 vSphere Bitfusion 服务器属于同一个 vCenter Server 实例。

前提条件

- 确认已安装主 vSphere Bitfusion 服务器。
- 确认 vSphere Bitfusion 已注册到 vCenter Server 服务器。

步骤

- 1 在 vCenter Server 的**主机和集群**视图中，右键单击 ESXi 主机，然后选择 **Bitfusion > 安装 Bitfusion 服务器**。
此时将出现**安装 Bitfusion 服务器**对话框。

- 2 在**选择 OVA 映像**页面上，输入 vSphere Bitfusion OVA 文件的 URL 或浏览到该文件，然后单击**下一步**。
- 3 在**验证模板详细信息**页面上，查看 OVA 模板详细信息，然后单击**下一步**。
- 4 在**选择名称和主机名**页面上，输入虚拟机名称和 vSphere Bitfusion 服务器的主机名，然后单击**下一步**。

（可选）您可以指定 vSphere Bitfusion 服务器的主机 ID，例如，在升级 vSphere Bitfusion 服务器时。如果跳过此步骤，将自动生成并分配主机 ID。

- 5 在**选择存储**页面上，定义在哪里以及如何存储已部署虚拟机的文件，然后单击**下一步**。
- 6 在**选择网络**页面上，指定网络适配器 1 的网络配置，然后单击**下一步**。

必须指定用于管理和数据流量的网络适配器 1 的配置。网络适配器 1 必须连接到与 vCenter Server 实例通信的网络。

如果您的 vSphere Bitfusion 服务器需要其他网络适配器处理数据流量，可以单击**添加网络适配器**并指定其他适配器的网络配置。

选项	描述
网络适配器	从下拉菜单中选择一个网络。
适配器类型	选择要分配给虚拟机的网络适配器。 注 vSphere Bitfusion 支持 VMXNET3 和 PVRDMA 适配器。
DHCP/固定 IP	指定由 DHCP 服务器分配网络适配器的地址，还是使用固定 IPv4 地址。
IPv4 地址	输入网络适配器的 IPv4 地址。如果使用 DHCP，请将此文本框留空。 注 IPv6 不受支持。
网络掩码	从下拉菜单中选择网络掩码。 例如，如果您的网络使用网络掩码 /24，请选择 24 (255.255.255.0)。
网关	输入要用于设备的网络网关地址。如果使用 DHCP，请将此文本框留空。
MTU	输入 MTU 大小。默认值为 1500。为获得最佳性能，请将 MTU 大小指定为网络硬件支持的最大 MTU 大小。 注 如果将 MTU 大小设置为大于 1500，请确认数据中心的网络交换机是否已启用巨型帧。
DNS 服务器	输入要用于设备的 DNS 服务器地址。如果使用 DHCP，请将此文本框留空。
DNS 搜索域	输入要用于设备的 DNS 搜索域地址。如果使用 DHCP，请将此文本框留空。
NTP	输入要用于设备的 NTP 服务器地址。如果您使用的是 DHCP，并且 DHCP 服务器支持发送 NTP 服务器信息，请将此文本框留空。

- 7 在**选择 GPU**页面上，将 GPU 添加到后续服务器，然后单击**下一步**。
 - a 单击**添加 GPU**。
 - b 从 **GPU 设备**下拉菜单中选择一个 GPU。

- c （可选）指定 GPU 的总内存。

vSphere Bitfusion 插件使用在**选择 GPU**页面上添加的所有 GPU 的总 GPU 内存计算 vSphere Bitfusion 服务器虚拟机的最小内存值和建议的内存映射 I/O 大小值。

- d （可选）要接受 NVIDIA 许可证，请选中**下载并安装 NVIDIA 驱动程序**复选框。

通过接受 NVIDIA 许可证，vSphere Bitfusion 会在首次引导虚拟机期间下载并安装 NVIDIA 驱动程序、CUDA 库和 NVIDIA Fabric Manager。

注 如果在无法访问 Internet 的环境（例如，使用气隙网络）中运行 vSphere Bitfusion，请不要选中该复选框。必须在部署 vSphere Bitfusion 设备后手动下载并安装 NVIDIA 软件。

如果您的 vSphere Bitfusion 服务器需要额外的 GPU，您可以再次单击**添加 GPU 设备**并指定 GPU 的设置。

- 8 在**自定义服务器**页面上，指定 vSphere Bitfusion 服务器详细信息，然后单击**下一步**。

- a 指定虚拟机的 CPU 数量。
- b 指定虚拟机的内存映射 I/O (MMIO) 大小（以 GB 为单位）。
- c （可选）输入客户帐户的密码。

部署完成后，使用客户用户帐户通过控制台 shell 或 SSH 登录到 vSphere Bitfusion 服务器。如果跳过此步骤，您无法登录到后续服务器。

- d （可选）选中**创建后打开虚拟机电源**复选框。

如果在打开虚拟机电源之前对其进行更改，可以取消选中该复选框。

- 9 在**摘要**页面上，查看部署详细信息，然后单击**完成**。

结果

用于安装 vSphere Bitfusion 服务器的新任务将显示在“近期任务”窗格中。该任务完成后，会在所选资源上创建新设备。

当新的 vSphere Bitfusion 服务器加入集群时，vCenter Server 将提供令牌、证书和配置以访问 vSphere Bitfusion 集群。

安装 vSphere Bitfusion 客户端

5

可以在 vSphere Bitfusion 客户端上运行 AI 和 ML 应用程序。从 vSphere Bitfusion 2.5 开始，可以在任意计算机上安装并启用 vSphere Bitfusion 客户端。

vSphere Bitfusion 客户端的系统要求

- vSphere Bitfusion 客户端的最低磁盘空间要求为 2 GB。
- vSphere Bitfusion 客户端的最低内存要求是至少为应用程序请求使用的 GPU 内存的 150%。
- vSphere Bitfusion 客户端的最低虚拟 CPU (vCPU) 要求与使用本地专用 GPU 运行应用程序的要求相同。
- vSphere Bitfusion 客户端必须安装在使用以下操作系统之一的计算机上。
 - CentOS 7
 - CentOS 8
 - Red Hat Linux 7.4 或更高版本
 - Ubuntu 16.04
 - Ubuntu 18.04
 - Ubuntu 20.04

虚拟机的必备条件

如果 vSphere Bitfusion 客户端在虚拟机 (VM) 上运行，则必须启用所有 VMware Tools 脚本。在创建新的虚拟机时，将在默认配置中启用脚本。

虚拟机的其他必备条件

如果 vSphere Bitfusion 客户端在与 vSphere Bitfusion 服务器属于同一 vCenter Server 实例的虚拟机上运行，则其他系统要求也适用。

- vSphere Bitfusion 客户端虚拟机必须在由 vCenter Server 7.0 管理的 vSphere 部署上运行。
- vSphere Bitfusion 客户端必须安装在 ESXi 主机 6.7 或更高版本上。
- 必须启用所有 VMware Tools 脚本。在创建新的虚拟机时，将在默认配置中启用脚本。

vSphere Bitfusion 客户端启用

您可以通过以下方式之一启用客户端。

- 如果客户端不是在与 vSphere Bitfusion 服务器属于同一 vCenter Server 实例的虚拟机上运行，请参见[生成客户端身份验证令牌](#)。
- 如果客户端在与 vSphere Bitfusion 服务器属于同一 vCenter Server 实例的虚拟机上运行，请参见[启用 vSphere Bitfusion 客户端](#)。

vSphere Bitfusion 插件必须注册到 vCenter Server。请参见[验证 vSphere Bitfusion 插件是否已注册到 vCenter Server](#)。

vSphere Bitfusion 客户端所需的端口

使用拒绝列表或防火墙规则确认以下端口未被阻止。vSphere Bitfusion 客户端通过以下端口与 vSphere Bitfusion 服务器进行通信。

自 vSphere Bitfusion 3.5 起，出于安全考虑，将默认禁用 TLSv1.0 和 TLSv1.1 协议。有关详细信息，请参见[知识库文章 2145796](#)。

端口	描述
45201 - 46225	vSphere Bitfusion 客户端使用这些端口与处理 CUDA 请求的 vSphere Bitfusion 进程通信。
55001 - 55201	vSphere Bitfusion 客户端使用这些端口与在 vSphere Bitfusion 服务器上运行的任务特定分派程序进程通信。该进程将请求会话，并启动 vSphere Bitfusion 服务器上用于处理工作负载的服务工作程序。
56001	此端口用于 vSphere Bitfusion 内部通信。vSphere Bitfusion 服务器通过此端口相互通信，vSphere Bitfusion 客户端使用此端口在 vSphere Bitfusion 服务器上启动任务。

vCenter Server 的 Web 浏览器要求

要使用 vSphere Bitfusion，您需要一个受 vCenter Server 支持的 Web 浏览器版本。有关详细信息，请参见[vSphere Client 软件要求](#)。

vSphere Bitfusion 兼容性和互操作性

有关与 vSphere Bitfusion 兼容的版本、型号和产品的列表，请参见[VMware vSphere Bitfusion 兼容性和互操作性](#)页面。

本章讨论了以下主题：

- [在 CentOS 和 Red Hat 上安装 vSphere Bitfusion 客户端](#)
- [在 Ubuntu 上安装 vSphere Bitfusion 客户端](#)

在 CentOS 和 Red Hat 上安装 vSphere Bitfusion 客户端

您可以在 CentOS 和 Red Hat 上安装 vSphere Bitfusion 客户端。

前提条件

- 验证 CentOS 或 Red Hat 操作系统的版本是否受支持。请参见第 5 章 [安装 vSphere Bitfusion 客户端](#)。
- 确认您安装的 vSphere Bitfusion 客户端的版本不高于 vSphere Bitfusion 服务器的版本。请参见第 9 章 [升级 vSphere Bitfusion](#)。

步骤

- 1 对于 CentOS，请安装 Extra Packages for Enterprise Linux (EPEL)，这个额外的软件包存储库可提供访问权限来安装常用软件包的软件包。

```
sudo yum install -y epel-release
```

- 2 将 VMware Bit 的公钥添加到 GNU Privacy Guard (GPG)。

```
sudo rpm --import https://packages.vmware.com/bitfusion/vmware.bitfusion.key
```

- 3 从 VMware 网站下载 vSphere Bitfusion 客户端版本，网址为 <https://packages.vmware.com/bitfusion/centos/>。

例如，运行 `wget https://packages.vmware.com/bitfusion/centos/8/bitfusion-client-centos8-3.5.0-5.x86_64.rpm`

- 4 通过运行 `sudo yum install -y ./bitfusion_client_version` 命令来安装客户端软件包，其中 **bitfusion_client_version** 是 vSphere Bitfusion 客户端的文件名。

例如，`sudo yum install -y ./bitfusion-client-centos8-3.5.0-5.x86_64.rpm`。

- 5 （可选）验证 vSphere Bitfusion 客户端的版本。

```
bitfusion version
Bitfusion version 3.5.0
```

后续步骤

在客户端虚拟机 (VM) 上启用 vSphere Bitfusion 客户端。请参见 [启用 vSphere Bitfusion 客户端](#)。如果客户端未安装在与服务器属于同一 vCenter Server 实例的虚拟机上，请参见 [生成客户端身份验证令牌](#)。

在 Ubuntu 上安装 vSphere Bitfusion 客户端

您可以在 Ubuntu 上安装 vSphere Bitfusion 客户端。

前提条件

- 验证您的 Ubuntu 操作系统的版本是否受支持。请参见第 5 章 [安装 vSphere Bitfusion 客户端](#)。

- 确认您安装的 vSphere Bitfusion 客户端的版本不高于 vSphere Bitfusion 服务器的版本。请参见第 9 章 升级 vSphere Bitfusion。

步骤

- 1 从 VMware 网站 <https://packages.vmware.com/bitfusion/ubuntu/> 下载适用于您的 Linux 分发包的 vSphere Bitfusion 客户端。

例如，运行 `wget https://packages.vmware.com/bitfusion/ubuntu/20.04/bitfusion-client-ubuntu2004_3.5.0-5_amd64.deb`。

- 2 通过运行 `apt-get update` 命令更新软件包。

```
sudo apt-get update
```

- 3 通过运行 `sudo apt-get install -y ./bitfusion_client_version` 命令来安装软件包，其中 **bitfusion_client_version** 是 vSphere Bitfusion 客户端的文件名。

例如，`sudo apt-get install -y ./bitfusion-client-ubuntu2004_3.5.0-5_amd64.deb`

- 4 验证 vSphere Bitfusion 客户端的版本。

```
bitfusion version
Bitfusion version 3.5.0
```

后续步骤

在客户端虚拟机 (VM) 上启用 vSphere Bitfusion 客户端。请参见 [启用 vSphere Bitfusion 客户端](#)。如果客户端未安装在与服务器属于同一 vCenter Server 实例的虚拟机上，请参见 [生成客户端身份验证令牌](#)。

启用 vSphere Bitfusion 客户端

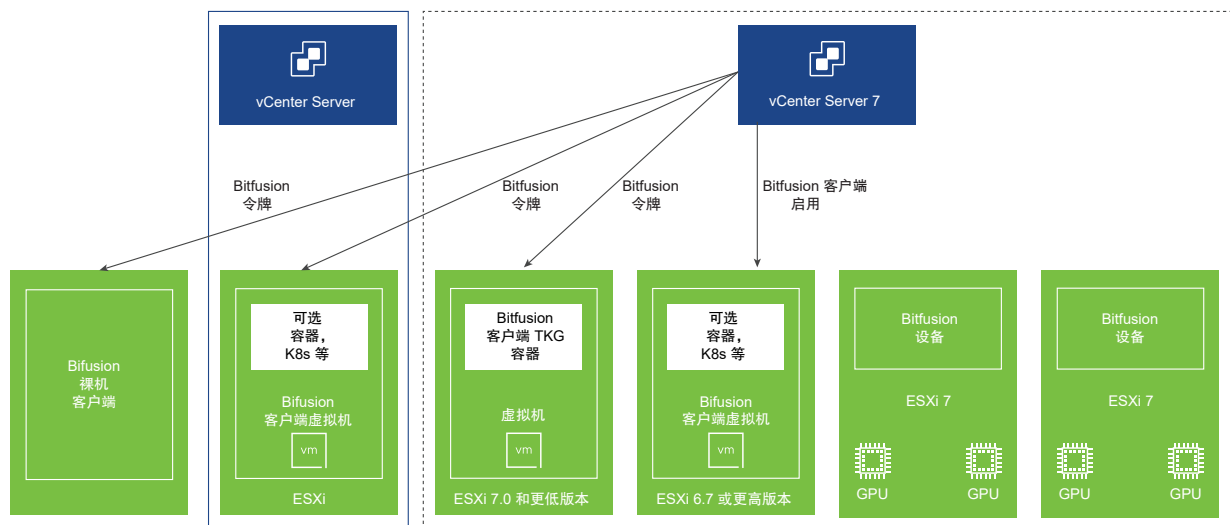
6

从 vSphere Bitfusion 2.5 开始，可以在多个平台上安装并启用 vSphere Bitfusion 客户端。

可以通过两种方式启用 vSphere Bitfusion 客户端。

- 对于与服务器在同一 vCenter Server 实例中的客户端，可从 vSphere Bitfusion 插件启用客户端。
- 在 vSphere Bitfusion 2.5 中，可在 Tanzu Kubernetes Grid (TKG) 容器、不同 vCenter Server 实例和裸机计算机上启用客户端。通过使用 vSphere Bitfusion 插件，可以生成授权令牌，并使用该令牌启用单个或多个客户端。您可以创建多个令牌来启用客户端组。要管理客户端或客户端组，可以启用或禁用特定令牌。

下图显示了多个平台上 vSphere Bitfusion 客户端的可用启用选项。



本章讨论了以下主题：

- 启用 vSphere Bitfusion 客户端
- 生成客户端身份验证令牌

启用 vSphere Bitfusion 客户端

可以在客户端虚拟机 (VM) 上启用 vSphere Bitfusion 客户端。

注 Bitfusion 客户端必须与 Bitfusion 服务器属于同一个 vCenter Server 实例。要添加在 vCenter Server 实例外部安装的客户端，请参见[生成客户端身份验证令牌](#)。

前提条件

- 安装适用于 Linux 发行版的 vSphere Bitfusion 客户端。请参见在 [CentOS 和 Red Hat 上安装 vSphere Bitfusion 客户端](#)和在 [Ubuntu 上安装 vSphere Bitfusion 客户端](#)。
- 关闭 vSphere Bitfusion 客户端虚拟机的电源。

步骤

- 1 在 vCenter Server 清单中，右键单击 vSphere Bitfusion 客户端虚拟机，然后选择 **Bitfusion > 启用 Bitfusion**。
- 2 在 **Bitfusion 启用**对话框中，选择**对于客户端**，这将允许用户运行 **Bitfusion 工作负载**单选按钮，然后单击**启用**。
- 3 打开客户端虚拟机的电源。
- 4 在虚拟机终端中，使用 `sudo usermod -aG bitfusion username` 命令将用户添加到 vSphere Bitfusion 组。
- 5 （可选）确认用户已成功添加到 vSphere Bitfusion 组。
 - a 注销并登录 vSphere Bitfusion 虚拟机终端。

注 如果未注销就返回到虚拟机终端，则新用户及其组分配将不会注册。

- b 在虚拟机终端中，运行 `groups` 命令列出用户及其关联的组。

```
groups
testuser bitfusion
```

- 6 （可选）通过运行 `bitfusion list_gpus` 命令列出 vSphere Bitfusion 部署中的可用 GPU，验证 vSphere Bitfusion 客户端是否正常工作。

```
/home/bitfusion$ bitfusion list_gpus
- server 0 [10.202.8.185:56001]: running 0 tasks
|- GPU 0: free memory 16160 MiB / 16160 MiB
|- GPU 1: free memory 16160 MiB / 16160 MiB
|- GPU 2: free memory 16160 MiB / 16160 MiB
|- GPU 3: free memory 16160 MiB / 16160 MiB
```

结果

您已成功启用 vSphere Bitfusion 客户端。

后续步骤

在 vSphere Bitfusion 客户端中启动应用程序。首次运行后，vSphere Bitfusion 客户端将加入集群。

生成客户端身份验证令牌

启用安装在 Tanzu Kubernetes Grid (TKG) 容器中、不同的 vCenter Server 实例中或裸机计算机上的 vSphere Bitfusion 客户端。

要启用与服务器不在同一 vCenter Server 实例中的 vSphere Bitfusion 客户端，请执行以下过程。必须生成授权令牌，下载相关的 tar 文件，然后在客户端的文件系统中提取该文件的内容。

要启用与服务器位于同一 vCenter Server 实例中的客户端，请参见[启用 vSphere Bitfusion 客户端](#)。

前提条件

- 确认已安装 vSphere Bitfusion 2.5 服务器或更高版本。
- 确认 vSphere Bitfusion 客户端的版本不高于 vSphere Bitfusion 服务器的版本。请参见[第 9 章 升级 vSphere Bitfusion](#)。
- 验证 vSphere Bitfusion 客户端是否可以通过网络访问集群中的服务器。

步骤

- 1 在 vSphere Client 中，选择**菜单 > Bitfusion**。
- 2 在**令牌**选项卡上，选择**新建令牌**。
- 3 在**创建令牌**对话框中，输入描述，然后单击**创建**。
- 4 从列表中选择令牌，单击**下载**，然后将 tar 文件保存到本地计算机。
- 5 将 tar 文件复制到客户端计算机或计算机的文件系统中。
- 6 提取 tar 文件的内容，然后将其复制到以下文件夹。
 - a 将 ca.crt 复制到 /etc/bitfusion/tls。
 - b 将 client.yaml 复制到 ~/.bitfusion。
 - c 将 servers.conf 复制到 ~/.bitfusion。
- 7 在计算机的终端中，通过运行 `sudo usermod -aG bitfusion username` 命令将用户添加到 Bitfusion 组。
- 8 （可选）确认用户已成功添加到 vSphere Bitfusion 组。
 - a 注销并登录到计算机的终端。
 - b 在终端中，运行 `groups` 命令。
 将列出用户及其关联的组。
- 9 （可选）通过运行 `bitfusion list_gpus` 命令列出 vSphere Bitfusion 部署中的可用 GPU，验证 vSphere Bitfusion 客户端是否正常工作。

结果

您已成功启用 vSphere Bitfusion 客户端。

后续步骤

在 vSphere Bitfusion 客户端中启动应用程序。首次运行后，vSphere Bitfusion 客户端将加入集群。

安装 NVIDIA 软件以与 vSphere Bitfusion 配合使用

7

如果您在 vSphere Bitfusion 服务器虚拟机的初始引导期间没有选择下载并安装 NVIDIA 驱动器、CUDA 库和 NVIDIA Fabric Manager，则必须手动安装此软件。

安装 NVIDIA 软件的方法有三种，具体取决于您的 vSphere Bitfusion 集群环境。

- 直接从 Internet 安装。
- 在具有本地 Web 服务器的气隙网络环境中安装。
- 在没有本地 Web 服务器的气隙网络环境中安装。

本章讨论了以下主题：

- [从 Internet 安装适用于 vSphere Bitfusion 的 NVIDIA 软件](#)
- [在气隙网络环境中安装 NVIDIA 软件](#)

从 Internet 安装适用于 vSphere Bitfusion 的 NVIDIA 软件

您可以手动安装适用于 vSphere Bitfusion 部署的 NVIDIA 软件。如果在 vSphere Bitfusion 服务器虚拟机 (VM) 的初始引导期间没有选择下载并安装 NVIDIA 驱动程序、CUDA 库和 NVIDIA Fabric Manager，并且您的 vSphere Bitfusion 可以访问 Internet，请执行此过程。

如果在 vSphere Bitfusion 服务器虚拟机的初始引导过程中选择了下载并安装 NVIDIA 软件，则可以跳过此过程。

前提条件

- 使用 NVIDIA 驱动程序即表示接受 NVIDIA 软件许可协议。请参见[允许客户使用 NVIDIA 软件的许可证](#)。
- 经认证可与 vSphere Bitfusion 配合使用的 NVIDIA 驱动程序是 `NVIDIA-Linux-x86_64-460.32.03.run`。
- NCCL 操作所必需的且经认证可与 vSphere Bitfusion 配合使用的 CUDA 库为 `cuda_11.2.0_460.27.04_linux.run`。
- 经认证可与 vSphere Bitfusion 配合使用的 NVIDIA Fabric Manager 软件包为 `nvidia-fabricmanager-460-460.32.03-1.x86_64.rpm`。

步骤

- 1 登录到 vSphere Bitfusion 服务器虚拟机的设备 shell。

```
ssh customer@bitfusion_server_IP_address
```

- 2 要安装 NVIDIA 驱动程序、CUDA 库和 NVIDIA Fabric Manager，请运行 `sudo install-nvidia-packages --defaults --yes` 命令。
- 3 重新启动虚拟机。

结果

vSphere Bitfusion 虚拟服务器打开电源后，先运行 10 分钟或更长时间，再执行任何其他配置任务或操作。在此期间，vSphere Bitfusion 服务器将注册到 vCenter Server。

后续步骤

验证 vSphere Bitfusion 插件是否已注册到 vCenter Server

在气隙网络环境中安装 NVIDIA 软件

您可以在气隙网络环境中手动安装 NVIDIA 软件。如果在 vSphere Bitfusion 服务器虚拟机 (VM) 的初始引导期间没有选择下载并安装 NVIDIA 驱动程序、CUDA 库和 NVIDIA Fabric Manager，并且您的 vSphere Bitfusion 无法访问 Internet，请执行此过程。

如果在 vSphere Bitfusion 服务器虚拟机的初始引导过程中选择了下载并安装 NVIDIA 软件，则可以跳过此过程。

前提条件

- 使用 NVIDIA 驱动程序即表示接受 NVIDIA 软件许可协议。请参见 [允许客户使用 NVIDIA 软件的许可证](#)。
- 经认证可与 vSphere Bitfusion 配合使用的 NVIDIA 驱动程序是 `NVIDIA-Linux-x86_64-460.32.03.run`。您可以从 NVIDIA 网站下载驱动程序软件：http://us.download.nvidia.com/tesla/460.32.03/NVIDIA-Linux-x86_64-460.32.03.run
- NCCL 操作所必需的且经认证可与 vSphere Bitfusion 配合使用的 CUDA 库为 `cuda_11.2.0_460.27.04_linux.run`。您可以从 NVIDIA 网站下载库：https://developer.download.nvidia.com/compute/cuda/11.2.0/local_installers/cuda_11.2.0_460.27.04_linux.run
- 经认证可与 vSphere Bitfusion 配合使用的 NVIDIA Fabric Manager 软件包为 `nvidia-fabricmanager-460-460.32.03-1.x86_64.rpm`。您可以从 NVIDIA 网站下载库：http://developer.download.nvidia.com/compute/cuda/repos/rhel7/x86_64/nvidia-fabricmanager-460-460.32.03-1.x86_64.rpm

步骤

- 1 在能够访问 Internet 的计算机上，创建并导航到 nvidia-packages 文件夹。

```
mkdir ~/nvidia-packages  
cd ~/nvidia-packages
```

- 2 下载 NVIDIA 驱动程序、CUDA 库和 NVIDIA Fabric Manager。

```
wget http://us.download.nvidia.com/tesla/460.32.03/NVIDIA-Linux-x86_64-460.32.03.run  
wget https://developer.download.nvidia.com/compute/cuda/11.2.0/local_installers/  
cuda_11.2.0_460.27.04_linux.run  
wget http://developer.download.nvidia.com/compute/cuda/repos/rhel7/x86_64/nvidia-  
fabricmanager-460-460.32.03-1.x86_64.rpm
```

3 移动并安装 NVIDIA 软件。

根据您的 vSphere Bitfusion 网络环境是具有本地 Web 服务器还是没有 Web 服务器，按照以下过程安装 NVIDIA 驱动程序、CUDA 库和 NVIDIA Fabric Manager。

选项	描述
选项	操作
具有本地 Web 服务器	<p>a 要将 NVIDIA 软件文件夹复制到根目录或本地 Web 服务器上的类似目录中，请运行以下 scp 命令。</p> <pre>scp ~/nvidia-packages/* mylogin@mylocalwebserver:/var/www/html/</pre> <p>b 要登录到本地 Web 服务器，请运行 mylogin 命令。</p> <pre>ssh mylogin@mylocalwebserver@mylocalwebserver</pre> <p>c 要向 NVIDIA 驱动程序授予读取权限，请运行 chmod 命令。</p> <pre>chmod +r /var/www/html/*</pre> <p>d 要登录到 vSphere Bitfusion 服务器，请运行 ssh</p> <pre>customer@bitfusion_server_ip_address。</pre> <p>e 要从本地 Web 服务器安装 NVIDIA 软件，请运行 install-nvidia-packages 命令。</p> <pre>sudo install-nvidia-packages --yes --driver http:// mylocalwebserver/NVIDIA-Linux-x86_64-460.32.03.run \ --cuda http://mylocalwebserver/ cuda_11.2.0_460.27.04_linux.run \ --fm http://mylocalwebserver/nvidia- fabricmanager-460-460.32.03-1.x86_64.rpm</pre>
没有 Web 服务器	<p>a 要将 NVIDIA 软件复制到 vSphere Bitfusion 服务器，请运行 scp 命令。</p> <pre>scp NVIDIA-Linux-x86_64-460.32.03.run customer@bitfusion_server_ip_address:~/ scp cuda_11.2.0_460.27.04_linux.run customer@bitfusion_server_ip_address:~/ scp nvidia-fabricmanager-460-460.32.03-1.x86_64.rpm customer@bitfusion_server_ip_address:~/</pre> <p>b 要登录到 vSphere Bitfusion 服务器，请运行 ssh</p> <pre>customer@bitfusion_server_ip_address。</pre> <p>c 要从本地文件安装 NVIDIA 软件，请运行 install-nvidia-packages 命令。</p> <pre>sudo install-nvidia-packages --yes --driver NVIDIA- Linux-x86_64-460.32.03.run \ --cuda cuda_11.2.0_460.27.04_linux.run \ --fm nvidia- fabricmanager-460-460.32.03-1.x86_64.rpm</pre>

4 重新启动虚拟机。

结果

vSphere Bitfusion 虚拟服务器打开电源后，先运行 10 分钟或更长时间，再执行任何其他配置任务或操作。在此期间，vSphere Bitfusion 服务器将注册到 vCenter Server。

后续步骤

[验证 vSphere Bitfusion 插件是否已注册到 vCenter Server](#)

将准虚拟 RDMA 网络适配器与 vSphere Bitfusion 结合使用

8

您可以使用准虚拟 RDMA (PVRDMA) 适配器来提高 vSphere Bitfusion 部署的性能。

RDMA 提供从一台计算机内存到另一台计算机内存的直接内存访问，不会涉及操作系统或 CPU。内存的传输卸载至支持 RDMA 的主机通道适配器 (Host Channel Adapter, HCA)。PVRDMA 网络适配器在虚拟环境中提供远程直接内存访问。

前提条件

- 您的 vSphere 环境必须先设置 PVRDMA，然后才能将 vSphere Bitfusion 配置为使用 PVRDMA。要了解更多信息，请参见《vSphere 网络连接》文档。
- vSphere Bitfusion 服务器和客户端必须配置有两个网络适配器。第一个网络适配器用于使用默认适配器类型（如 VMXNET3）的管理流量。第二个网络适配器用于 PVRDMA 流量。
- 您必须先关闭 vSphere Bitfusion 服务器和客户端虚拟机的电源，然后才能将它们配置为使用 PVRDMA 适配器。

步骤

- 1 在 vSphere Client 中找到托管 vSphere Bitfusion 服务器和客户端的虚拟机。
- 2 右键单击清单中的虚拟机，然后选择**编辑设置**。
- 3 从**添加新设备**下拉菜单中，选择**网络适配器 2**。
“新网络”部分将添加到**虚拟硬件**选项卡下的列表中。
- 4 选择 PVRDMA 网络。
- 5 展开“新建网络”部分并将虚拟机连接到分布式端口组。
- 6 将**状态**设置更改为**打开电源时连接**。
- 7 在**适配器类型**下拉菜单中，选择 PVRDMA。
- 8 打开虚拟机电源。
- 9 如果您已打开托管 vSphere Bitfusion 客户端的虚拟机的电源，请安装 RDMA 驱动程序。

除了 RDMA 驱动程序之外，还应安装诊断工具。

- 对于 CentOS 和 Red Hat Linux，请运行以下命令。

```
yum install -y open-vm-tools rdma-core libibverbs libibverbs-utils infiniband-diags
```

- 对于 Ubuntu Linux，请运行以下命令。

```
sudo apt-get install -y rdma-core libmlx4-1 infiniband-diags ibutils ibverbs-utils  
rdmacm-utils perftest
```

结果

您已成功启用 vSphere Bitfusion 以使用 PVRDMA 网络适配器。

您可以使用 `ib_send_bw`（InfinBand 发送带宽）命令测试 vSphere Bitfusion 服务器和客户端之间的连接。例如，如果 vSphere Bitfusion 服务器和客户端的 IP 地址为 192.168.10.10 和 192.16.10.11，请运行以下命令。

```
#From the server 192.16.10.10  
ib_send_bw  
  
#From the client 192.16.10.11 - connects to the server  
ib_send_bw 192.168.10.10
```

vSphere Bitfusion 客户端会将带宽报告写入标准输出 (stdout)。

升级 vSphere Bitfusion

9

从 vSphere Bitfusion 2.5 开始，可以对 vSphere Bitfusion 环境执行升级。通过升级集群，可以保留当前配置数据和监控统计信息。

vSphere Bitfusion 支持多版本服务器和客户端环境。所有服务器都必须运行相同的 vSphere Bitfusion OVA 版本，而客户端版本可以混合使用。vSphere Bitfusion 客户端版本必须与 vSphere Bitfusion 服务器版本相同或更低。升级 vSphere Bitfusion 环境时，请先升级服务器，然后再升级客户端。

vSphere Bitfusion 版本	服务器版本	客户端版本
2.0.0	2.0.0	2.0.0
2.0.1	2.0.1	2.0.1
2.0.2	2.0.2	2.0.2
2.5.0	2.5.0	2.0.0 2.0.1 2.0.2 2.5.0
2.5.1	2.5.1	2.0.0 2.0.1 2.0.2 2.5.0 2.5.1
3.0.0	3.0.0	2.0.0 2.0.1 2.0.2 2.5.0 2.5.1 3.0.0

vSphere Bitfusion 版本	服务器版本	客户端版本
3.0.1	3.0.1	2.0.0
		2.0.1
		2.0.2
		2.5.0
		2.5.1
		3.0.0
		3.0.1
3.5.0	3.5.0	2.0.0
		2.0.1
		2.0.2
		2.5.0
		2.5.1
		3.0.0
		3.0.1
		3.5.0

本章讨论了以下主题：

- 将 vSphere Bitfusion 集群从 2.0 升级到 2.5
- 将 vSphere Bitfusion 集群从 2.5 升级到 3.0
- 将 vSphere Bitfusion 集群从 3.0 升级到 3.5

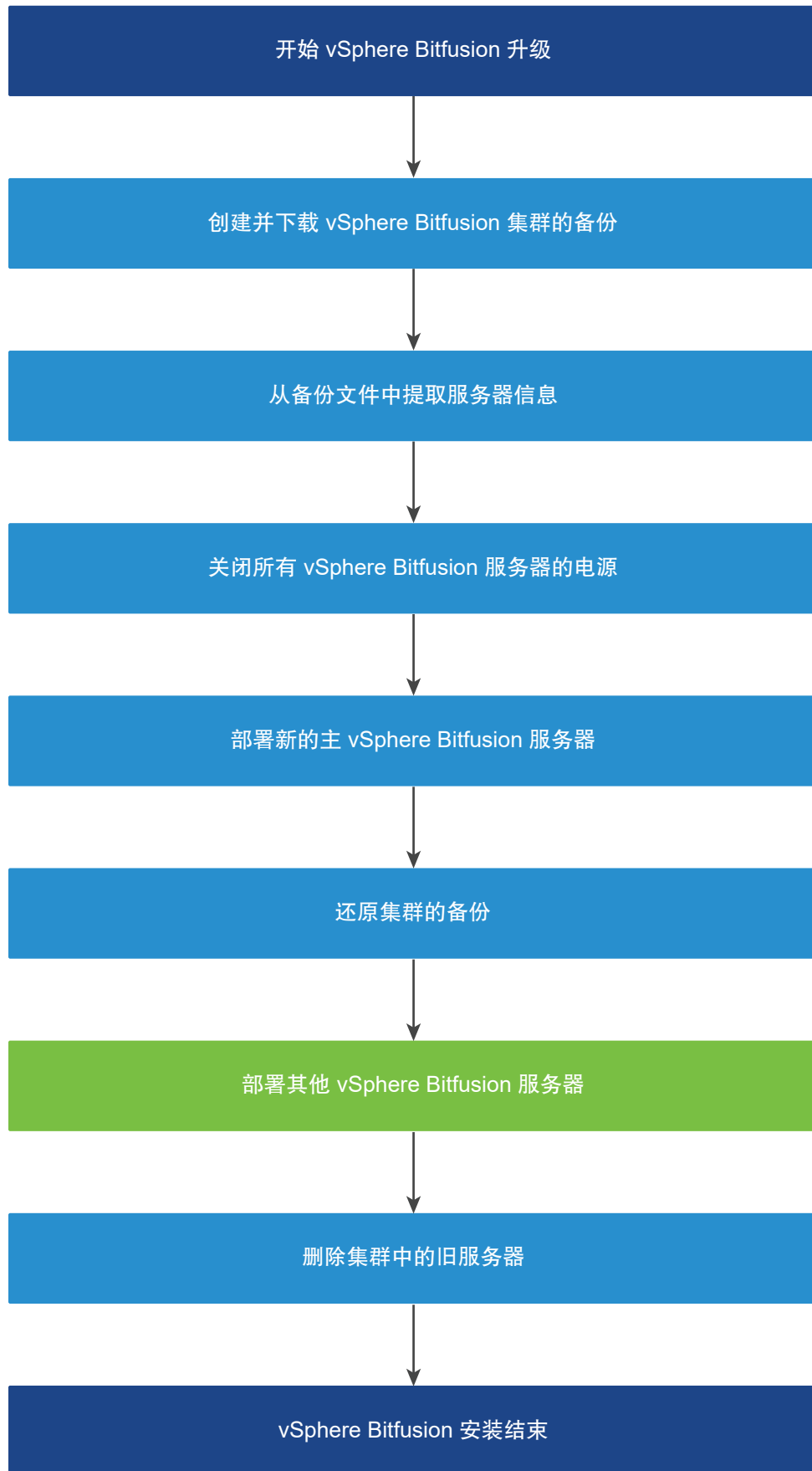
将 vSphere Bitfusion 集群从 2.0 升级到 2.5

您可以在 vSphere Bitfusion 2.5 上运行人工智能 (AI) 和机器学习 (ML) 工作负载，而且不会丢失当前的集群配置和监控数据。

要升级集群，必须备份环境，使用 vSphere Bitfusion 设备版本 2.5 部署新的服务器虚拟机 (VM)，然后还原备份。

您可以使用当前的 vSphere Bitfusion 2.0 客户端，也可以将客户端升级到版本 2.5。要升级客户端，必须安装最新的 CentOS、Red Hat 或 Ubuntu 软件包。有关详细信息，请参见第 5 章 [安装 vSphere Bitfusion 客户端](#)。

图 9-1. vSphere Bitfusion 升级 workflow



有关详细信息，请参见 VMware vSphere Bitfusion 用户指南中的备份 Bitfusion 集群。

- 2 从下载的 bitfusionbackup.tar.gz 存档文件中打开 manifest.json 文件，然后找到服务器部分。此部分包括有关备份时 vSphere Bitfusion 集群中服务器的信息。

记录每个服务器的主机 ID、主机名和 GPU 数量。

例如，以下服务器的主机 ID 为 6a2f4e80-70d8-4c51-bf10-00284f3ed2c6，主机名为 bitfusion-server-2.0.1-3-1，并且该服务器安装了一个 GPU。

```
"servers": [
  {
    "id": "6a2f4e80-70d8-4c51-bf10-00284f3ed2c6",
    "hostname": "bitfusion-server-2.0.1-3-1",
    "ip": "10.202.8.209",
    "port": "56001",
    "address": "10.202.8.209:56001",
    "mode": "manager",
    "health": "PASS",
    "num_devices": 1,
    "lastseen": "2020-10-14T21:29:38Z",
    "license": {
      "type": "vcenter-license",
      "name": "vSphere 7 Enterprise Plus",
      "license-id": "example",
      "key": "example",
      "expiry": "2025-09-30T00:00:00Z"
    }
  }
]
```

- 3 关闭集群中 vSphere Bitfusion 服务器的电源。
- 4 安装新的主 vSphere Bitfusion 服务器。
 - a 使用 vSphere Bitfusion 2.5 设备 OVF 模板部署新的主 vSphere Bitfusion 服务器虚拟机。
有关详细信息，请参见第 4 章 部署 vSphere Bitfusion 设备。
在部署过程中，输入主 vSphere Bitfusion 2.0 服务器所用的同一主机名。
 - b 在新虚拟机的设置中，添加与主 vSphere Bitfusion 2.0 服务器使用的相同数量的 GPU。
 - c 在新虚拟机的设置中，添加 guestinfo.bitfusion.server.host-id 配置参数。此参数值必须与主服务器版本 2.0 的主机 ID（列在 manifest.json 文件中）相匹配。
有关详细信息，请参见《vSphere 虚拟机管理》文档中的编辑配置文件参数。
 - d 打开服务器电源，并等待 vSphere Bitfusion 插件注册到 vCenter Server。
- 5 通过使用 vSphere Bitfusion 插件，还原 vSphere Bitfusion 2.0 集群的备份。
有关详细信息，请参见 VMware vSphere Bitfusion 用户指南中的还原 Bitfusion 集群。
- 6 对于集群中的每个后续 vSphere Bitfusion 服务器，请执行以下步骤。
 - a 使用 vSphere Bitfusion 2.5 设备 OVF 模板部署新的服务器虚拟机。
在部署过程中，输入相应的 vSphere Bitfusion 2.0 服务器所用的相同主机名。

- b 在新虚拟机的设置中，添加与相应的 vSphere Bitfusion 2.0 服务器使用的相同数量的 GPU。
- c 在新虚拟机的设置中，添加 `guestinfo.bitfusion.server.host-id` 配置参数。此参数值必须与对应的 2.0 版服务器的主机 ID（列在 `manifest.json` 文件中）相匹配。
- d 将虚拟机作为 vSphere Bitfusion 服务器。

有关详细信息，请参见[添加后续的 vSphere Bitfusion 服务器](#)。

- e 打开虚拟机电源。必须按顺序打开多个虚拟机的电源。

7 删除 vSphere Bitfusion 2.0 服务器虚拟机。

集群中的服务器将升级到版本 2.5。

将 vSphere Bitfusion 集群从 2.5 升级到 3.0

您可以将 vSphere Bitfusion 集群升级到版本 3.0，并保留当前的集群配置和监控数据。

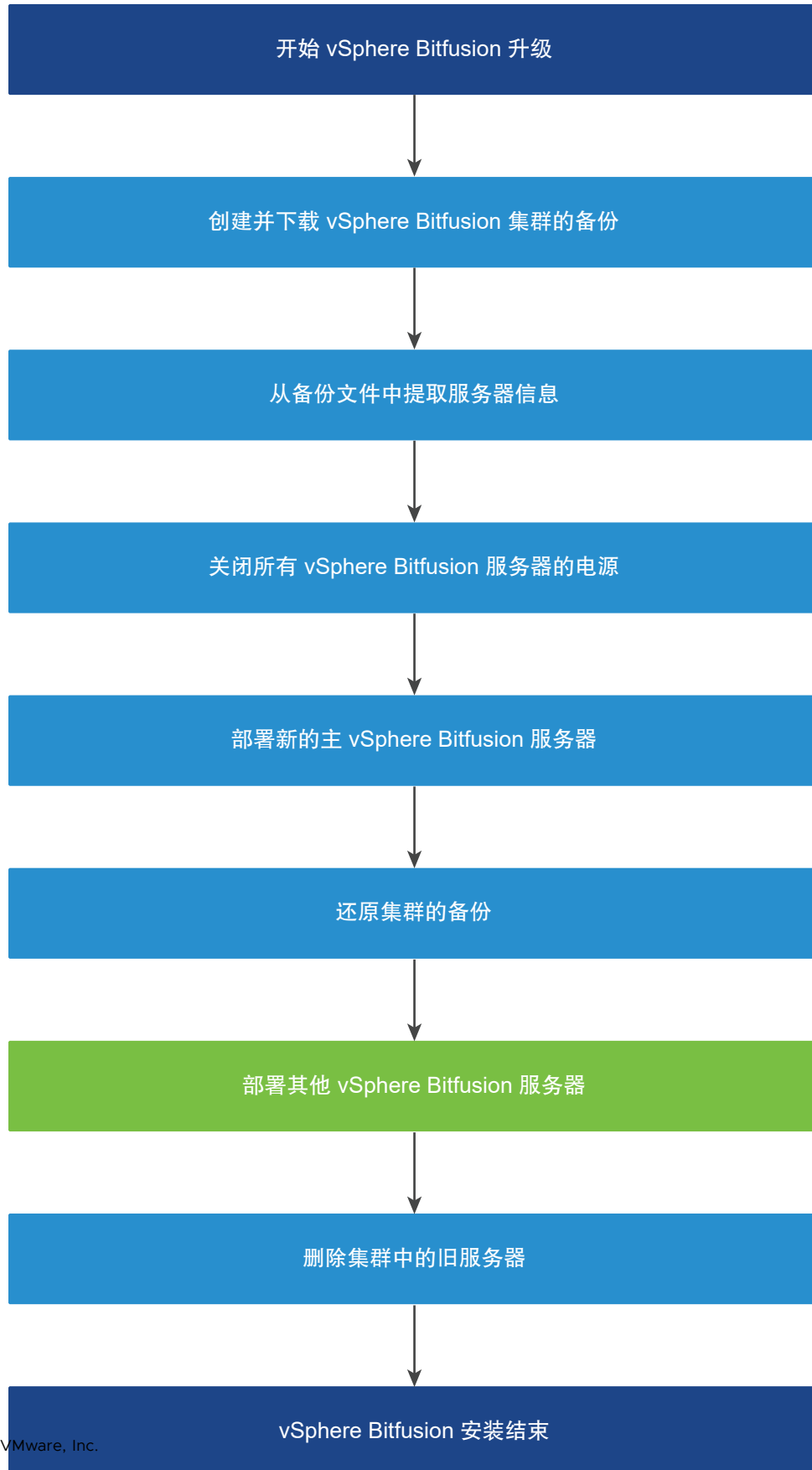
要升级集群，必须升级 vSphere Bitfusion 环境中的服务器。必须备份环境，使用最新版本的 vSphere Bitfusion 设备部署新的服务器虚拟机 (VM)，然后还原备份。

vSphere Bitfusion 客户端升级

您可以使用当前的 vSphere Bitfusion 2.x 客户端，也可以将客户端升级到版本 3.0。要升级客户端，必须在 Ubuntu、CentOS 或 Red Hat Linux 操作系统上安装最新软件包。客户端版本可以与 vSphere Bitfusion 服务器版本相同或更低。有关详细信息，请参见[第 5 章 安装 vSphere Bitfusion 客户端](#)。

vSphere Bitfusion 服务器升级

图 9-2. vSphere Bitfusion 升级 workflow



有关详细信息，请参见 VMware vSphere Bitfusion 用户指南中的备份 Bitfusion 集群。

- 2 从下载的 bitfusionbackup.tar.gz 存档文件中打开 manifest.json 文件，然后找到服务器部分。此部分包括有关备份时 vSphere Bitfusion 集群中服务器的信息。

记录每个服务器的主机 ID、主机名和 GPU 数量。

例如，以下服务器的主机 ID 为 6a2f4e80-70d8-4c51-bf10-00284f3ed2c6，主机名为 bitfusion-server-2.5.1-3-1，并且该服务器安装了一个 GPU。

```
"servers": [
  {
    "id": "6a2f4e80-70d8-4c51-bf10-00284f3ed2c6",
    "hostname": "bitfusion-server-2.5.1-3-1",
    "ip": "10.202.8.209",
    "port": "56001",
    "address": "10.202.8.209:56001",
    "mode": "manager",
    "health": "PASS",
    "num_devices": 1,
    "lastseen": "2020-10-14T21:29:38Z",
    "license": {
      "type": "vcenter-license",
      "name": "vSphere 7 Enterprise Plus",
      "license-id": "example",
      "key": "example",
      "expiry": "2025-09-30T00:00:00Z"
    }
  }
]
```

- 3 关闭集群中 vSphere Bitfusion 服务器的电源。
- 4 安装新的主 vSphere Bitfusion 服务器。
 - a 使用 vSphere Bitfusion 3.0 设备 OVF 模板部署新的主 vSphere Bitfusion 服务器虚拟机。
有关详细信息，请参见第 4 章 部署 vSphere Bitfusion 设备。
在部署过程中，输入主 vSphere Bitfusion 2.5 服务器所用的相同主机名。
 - b 在新虚拟机的设置中，添加与主 vSphere Bitfusion 2.5 服务器使用的相同数量的 GPU。
 - c 在新虚拟机的设置中，添加 guestinfo.bitfusion.server.host-id 配置参数。此参数值必须与主服务器版本 2.5 的主机 ID（列在 manifest.json 文件中）相匹配。
有关详细信息，请参见《vSphere 虚拟机管理》文档中的编辑配置文件参数。
 - d 打开服务器电源，并等待 vSphere Bitfusion 插件注册到 vCenter Server。
- 5 通过使用 vSphere Bitfusion 插件，还原 vSphere Bitfusion 2.5 集群的备份。
有关详细信息，请参见 VMware vSphere Bitfusion 用户指南中的还原 Bitfusion 集群。
- 6 对于集群中的每个后续 vSphere Bitfusion 服务器，请执行以下步骤。
 - a 使用 vSphere Bitfusion 插件部署新的服务器虚拟机。

在部署过程中，输入相应 vSphere Bitfusion 2.5 服务器的 `manifest.json` 中列出的主机名和主机 ID。有关详细信息，请参见[添加后续的 vSphere Bitfusion 服务器](#)。

- b 在新虚拟机的设置中，添加与相应的 vSphere Bitfusion 2.5 服务器使用的相同数量的 GPU。
- c 在新虚拟机的设置中，添加 `guestinfo.bitfusion.server.host-id` 配置参数。此参数值必须与具有 2.5 版文件的相应服务器的主机 ID 匹配。
- d 打开虚拟机电源。必须按顺序打开多个虚拟机的电源。

7 删除 vSphere Bitfusion 2.5 服务器虚拟机。

将 vSphere Bitfusion 集群从 3.0 升级到 3.5

要使用最新的 vSphere Bitfusion 版本并保留当前的集群配置和监控数据，可以升级 vSphere Bitfusion 集群。

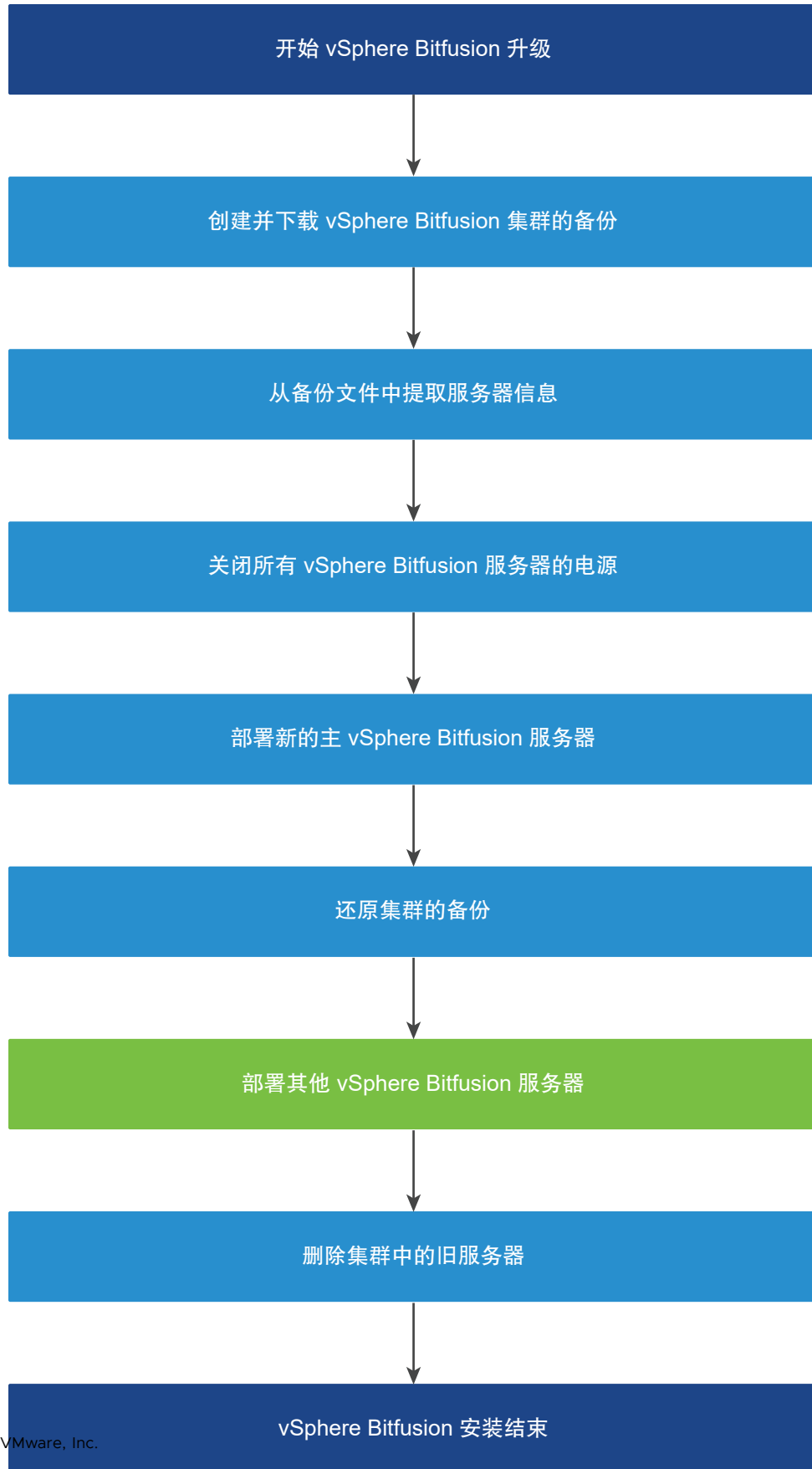
要升级集群，必须升级 vSphere Bitfusion 环境中的服务器。必须备份环境，使用最新版本的 vSphere Bitfusion 设备部署新的服务器虚拟机 (VM)，然后还原备份。

vSphere Bitfusion 客户端升级

您可以使用当前的 vSphere Bitfusion 2.x 客户端，也可以将客户端升级到版本 3.5。要升级客户端，必须在 Ubuntu、CentOS 或 Red Hat Linux 操作系统上安装最新软件包。客户端版本可以与 vSphere Bitfusion 服务器版本相同或更低。有关详细信息，请参见[第 5 章 安装 vSphere Bitfusion 客户端](#)。

vSphere Bitfusion 服务器升级

图 9-3. vSphere Bitfusion 升级 workflow



有关详细信息，请参见 VMware vSphere Bitfusion 用户指南中的备份 Bitfusion 集群。

- 2 从下载的 bitfusionbackup.tar.gz 存档文件中打开 manifest.json 文件，然后找到服务器部分。此部分包括有关备份时 vSphere Bitfusion 集群中服务器的信息。

记录每个服务器的主机 ID、主机名和 GPU 数量。

例如，以下服务器的主机 ID 为 6a2f4e80-70d8-4c51-bf10-00284f3ed2c6，主机名为 bitfusion-server-3.0.1-4，并且该服务器安装了一个 GPU。

```
"servers": [
  {
    "id": "6a2f4e80-70d8-4c51-bf10-00284f3ed2c6",
    "hostname": "bitfusion-server-3.0.1-4",
    "ip": "10.202.8.209",
    "port": "56001",
    "address": "10.202.8.209:56001",
    "mode": "manager",
    "health": "PASS",
    "num_devices": 1,
    "lastseen": "2020-10-14T21:29:38Z",
    "license": {
      "type": "vcenter-license",
      "name": "vSphere 7 Enterprise Plus",
      "license-id": "example",
      "key": "example",
      "expiry": "2025-09-30T00:00:00Z"
    }
  }
]
```

- 3 关闭集群中 vSphere Bitfusion 服务器的电源。
- 4 安装新的主 vSphere Bitfusion 服务器。
 - a 使用 vSphere Bitfusion 3.5 设备 OVA 模板部署新的主 vSphere Bitfusion 服务器虚拟机。
有关详细信息，请参见第 4 章 部署 vSphere Bitfusion 设备。
在部署过程中，输入主 vSphere Bitfusion 3.0 服务器所用的相同主机名。
 - b 在新虚拟机的设置中，添加与主 vSphere Bitfusion 3.0 服务器使用的相同数量的 GPU。
 - c 在新虚拟机的设置中，添加 guestinfo.bitfusion.server.host-id 配置参数。此参数值必须与主服务器版本 3.0 的主机 ID（列在 manifest.json 文件中）相匹配。
有关详细信息，请参见《vSphere 虚拟机管理》文档中的编辑配置文件参数。
 - d 打开服务器电源，并等待 vSphere Bitfusion 插件注册到 vCenter Server。
- 5 通过使用 vSphere Bitfusion 插件，将 vSphere Bitfusion 3.0 集群的备份还原到信息集群。
有关详细信息，请参见 VMware vSphere Bitfusion 用户指南中的还原 Bitfusion 集群。
- 6 对于集群中的每个后续 vSphere Bitfusion 服务器，请执行以下步骤。
 - a 使用 vSphere Bitfusion 插件部署新的服务器虚拟机。

在部署过程中，输入相应 vSphere Bitfusion 3.0 服务器的 `manifest.json` 中列出的主机名和主机 ID。有关详细信息，请参见[添加后续的 vSphere Bitfusion 服务器](#)。

- b 在新虚拟机的设置中，添加与相应的 vSphere Bitfusion 3.0 服务器使用的相同数量的 GPU。
- c 在新虚拟机的设置中，添加 `guestinfo.bitfusion.server.host-id` 配置参数。此参数值必须与具有 3.0 版文件的相应服务器的主机 ID 匹配。
- d 打开虚拟机电源。必须按顺序打开多个虚拟机的电源。

7 删除 vSphere Bitfusion 3.0 服务器虚拟机。