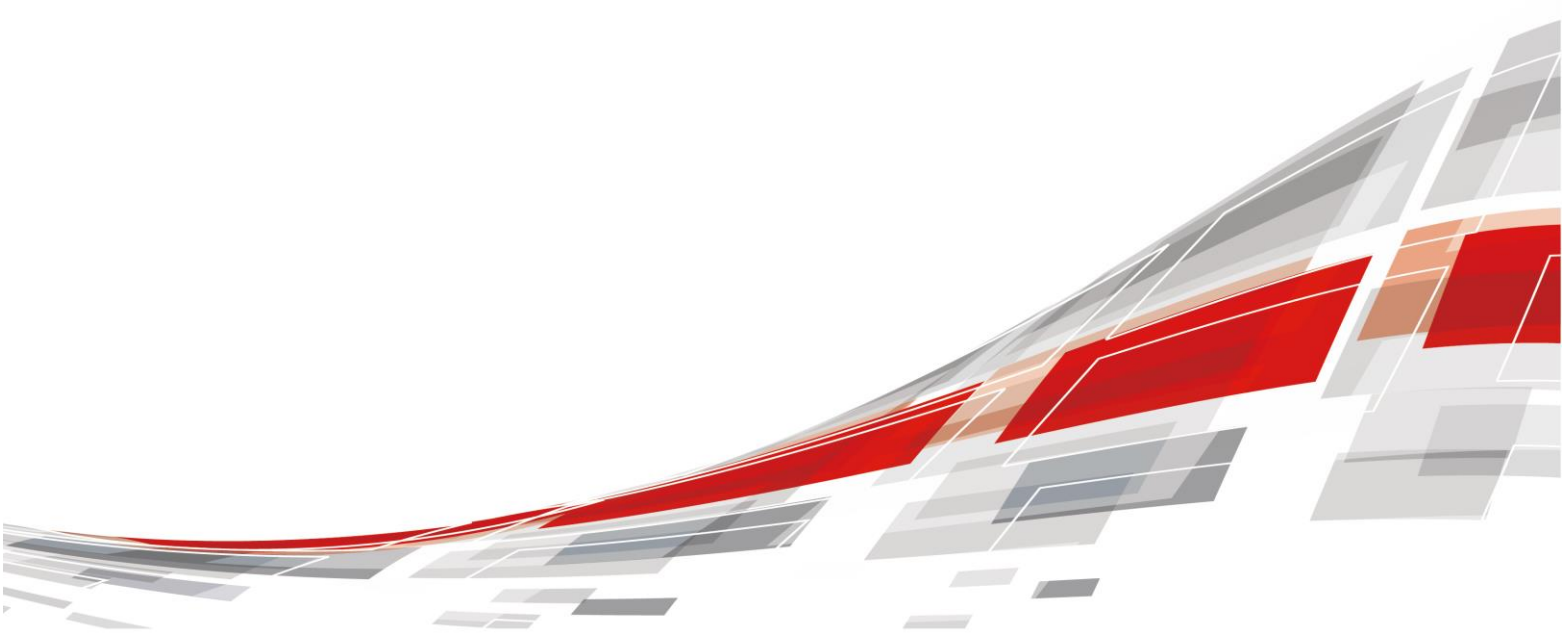
# FusionOne Compute
# 23.1.0
# 技术白皮书

**文档版本** 01

**发布日期** 2023-10-30



**超聚变数字技术有限公司**

**商标声明**



**注意**

本文中，只是为了描述的简洁和方便理解，用“xFusion”指代“xFusion Digital Technologies Co., Ltd.”，这并不代表“xFusion”还可以具备其它含义。基于本文中单独提及或描述的“xFusion”，不能用于“xFusion Digital Technologies Co., Ltd.”之外的理解或表达，超聚变数字技术有限公司也不承担因单独使用“xFusion”所带来的其它任何法律责任。

您购买的产品、服务或特性等应受超聚变数字技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，超聚变数字技术有限公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

超聚变数字技术有限公司

地址： 河南省郑州市郑东新区龙子湖智慧岛正商博雅广场 1 号楼 9 层 邮编：450046

网址： https://www.xfusion.com

# 前言

## 概述

本文档介绍了 FusionOne Compute 23 产品价值、产品架构、高性能、线性扩展、系统安全以及系统可靠性。

## 读者对象

本文档主要适用于以下工程师：

- 营销工程师
- 技术支持工程师
- 维护工程师

## 符号约定

在本文中可能出现下列标志，它们所代表的含义如下。

| 符号 | 说明 |
| --- | --- |
| ⚠ 危险 | 表示如不避免则将会导致死亡或严重伤害的具有高等级风险的危害。 |
| ⚠ 警告 | 表示如不避免则可能导致死亡或严重伤害的具有中等级风险的危害。 |
| ⚠ 注意 | 表示如不避免则可能导致轻微或中度伤害的具有低等级风险的危害。 |

| 符号 | 说明 |
| --- | --- |
| 须知 | 用于传递设备或环境安全警示信息。如不避免则可能会导致设备损坏、数据丢失、设备性能降低或其它不可预知的结果。<br>“须知”不涉及人身伤害。 |
| 说明 | 对正文中重点信息的补充说明。<br>“说明”不是安全警示信息，不涉及人身、设备及环境伤害信息。 |

# 修改记录

| 文档版本 | 发布日期 | 修改说明 |
| --- | --- | --- |
| 01 | 2023-10-30 | 第一次正式发布。 |

# 目 录



# 1 产品概述

随着数据不断增长以及互联网业务的兴起，新兴业务的激增、业务数据呈现几何倍数增加，传统服务器+存储的架构已经无法很好满足业务发展需求，分布式、云化技术应运而生。越来越多的企业采用虚拟化与云计算技术来构建 IT 系统，提升 IT 系统的资源利用率以及缩短业务上线周期。但在应用过程中，企业面临如下挑战：

- 虚拟平台部署和管理复杂，运维费用仍然维持增长趋势。
- 安装部署复杂，硬件来自多厂商，规划、部署、调优需要丰富的经验支撑。
- 多厂商设备，售后支持界面多，解决问题慢。
- 系统庞大（不同厂商硬件设备维护、虚拟平台管理），维护难度大。

企业越来越关注成本控制、业务敏捷、风险管控，希望能拥有总成本低、新业务的上线时间快、资源可弹性伸缩、安全可靠、高性能的 IT 系统。

FusionOne Compute 是一个开放的、可扩展的系统，具有计算/存储/网络融合、高性能、高可靠、高安全、业务自动化快捷部署、统一管理、资源智能弹性伸缩、运维简单的特点，可帮助客户业务快速上线，快速实现不同云应用的部署，同时降低维护管理的难度。

# 2 产品价值

FusionOne Compute 遵循开放架构标准，集成服务器硬件、计算虚拟化、网络虚拟化、存储虚拟化为一体，支持对接外置存储设备，资源可按需调配、线性扩展。

## 简单

FusionOne Compute 实现了工具安装、上电后的设备自动发现、统一的维护管理，端到端的简化了业务交付。

- 简化安装：硬件软件一键安装，设备进场后快速上线。
- 简捷交付：设备上电自动发现，参数自动配置，实现业务快速上线。
- 简单维护：统一界面管理，故障主动排查，简化日常运维。

## 优化

FusionOne Compute 通过采用业界领先硬件，为应用提供最优的业务体验。

- 计算虚拟化：集成并优化 KVM 虚拟化，结合 VRM 高可靠管理，提供简单易用的虚拟机生命周期管理。
- 存储虚拟化：通过 iSCSI 支持对接外部 SAN 存储，如 FC-SAN、IP-SAN，将对接的卷格式化为共享文件系统，创建 RAW 格式的磁盘文件，绑定给虚拟机提供存储服务。
- 网络优化：支持 GE、10GE、25GE 网络，提供高带宽的交换网络，并且支持普通模式、SR-IOV、DPDK 用户态模式、SDN 多种类型的网络虚拟化模式。

# 3 产品介绍

FusionOne Compute 虚拟化平台，主要负责硬件资源的虚拟化，以及对虚拟资源、业务资源、用户资源的集中管理。它采用虚拟计算、虚拟存储、虚拟网络等技术，完成计算资源、存储资源、网络资源的虚拟化。同时通过统一的接口，对这些虚拟资源进行集中调度和管理，从而降低业务的运行成本，保证系统的安全性和可靠性。

- 统一虚拟化平台

  将计算资源划分为多个虚拟机资源，为用户提供高性能、可运营、可管理的虚拟机。支持虚拟机资源按需分配，支持多操作系统，QoS 保证资源分配，隔离用户间影响。

- 大集群

  单个计算集群最大可支持 256 个主机，8000 台虚拟机（单站点支持多个计算集群）。

- 自动化调度

  支持自定义的资源管理 SLA（Service-Level Agreement）策略、故障判断标准及恢复策略。通过 IT 资源调度、热管理、能耗管理等一体化拉通，降低维护成本。自动检测服务器或业务的负载情况，对资源进行智能调度，均衡各服务器及业务系统负载，保证系统良好的用户体验和业务系统的最佳响应。

- 完善的权限管理

  可根据不同的角色、权限等，提供完善的权限管理功能，授权用户对系统内的资源进行管理。

- 丰富的运维管理

  提供多种运维工具，实现业务的可控、可管，提高整个系统运营的效率。

  - 支持“黑匣子”快速故障定位

系统通过获取异常日志和程序堆栈，缩短问题定位时间，快速解决异常问题。

- 支持全 Web 化的界面

  通过 Web 浏览器对所有硬件资源、虚拟资源、用户业务发放等进行监控管理。

- 云安全

  采用多种安全措施和策略，并遵从信息安全法律法规，对用户接入、管理维护、数据、网络、虚拟化等提供端到端的业务保护。



# 3.1 架构概述

FusionOne Compute 虚拟化平台将节点的计算、网络资源虚拟提供给节点上的业务虚拟机使用。根据节点提供的功能特性差异，又分为管理节点、计算节点，详细的节点架构如下图：

图3-1 集群架构

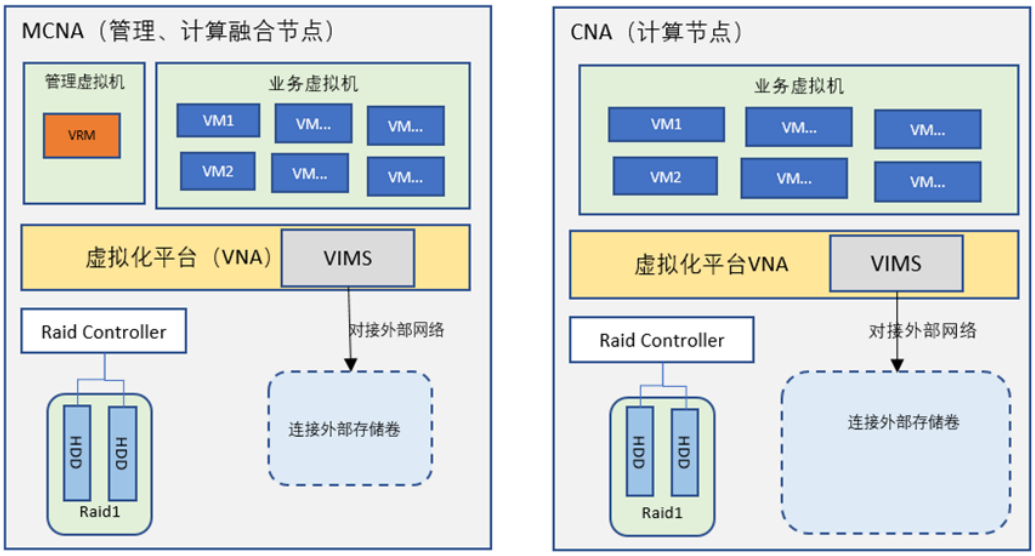



表3-1 节点说明

| 名称 | 说明 | 部署原则 |
| --- | --- | --- |
| MCNA（管理节点） | 具有管理功能的节点，其上部署了 VRM 管理虚拟机。同时也可提供计算功能。 | 根据需要部署 1 个～多个。 |
| CNA（计算节点） | 具有计算功能的节点，虚拟化计算资源。 | 根据需要部署 0 个～多个。 |

## 3.1.1 软件架构介绍

FusionOne Compute 虚拟化平台采用 KVM 虚拟化架构，将节点的计算、网络资源虚拟提供给节点上的业务虚拟机使用。通过集成共享文件系统，支持对接外部存储作为虚拟机的存储资源。根据节点提供的功能特性差异，又分为管理融合节点、存储融合节点、计算节点，详细的节点架构如下图：

图3-2 节点架构

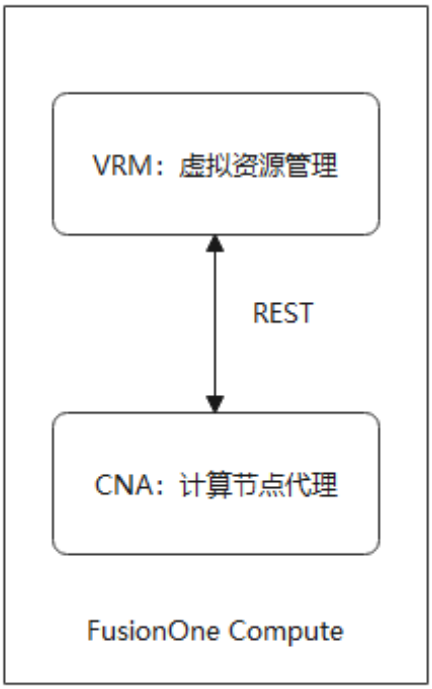



表3-2 FusionOne Compute 场景各类节点说明

| 名称 | 说明 |
| --- | --- |
| VRM | VRM 主要提供以下功能：<br>• 管理集群内的块存储资源。<br>• 管理集群内的网络资源(IP/VLAN)，为虚拟机分配 IP 地址。<br>• 管理集群内虚拟机的生命周期以及虚拟机在计算节点上的分布和迁移。<br>• 管理集群内资源的动态调整。<br>• 通过对虚拟资源、用户数据的统一管理，对外提供弹性计算、存储、IP 等服务。<br>• 通过提供统一的操作维护管理接口，操作维护人员通过 WebUI 远程访问 FusionOne Compute 对整个系统进行操作维护，包含资源管理、资源监控、资源报表等。 |
| CNA | 具有计算功能的节点，虚拟化计算资源。<br>• 提供虚拟计算功能。 |

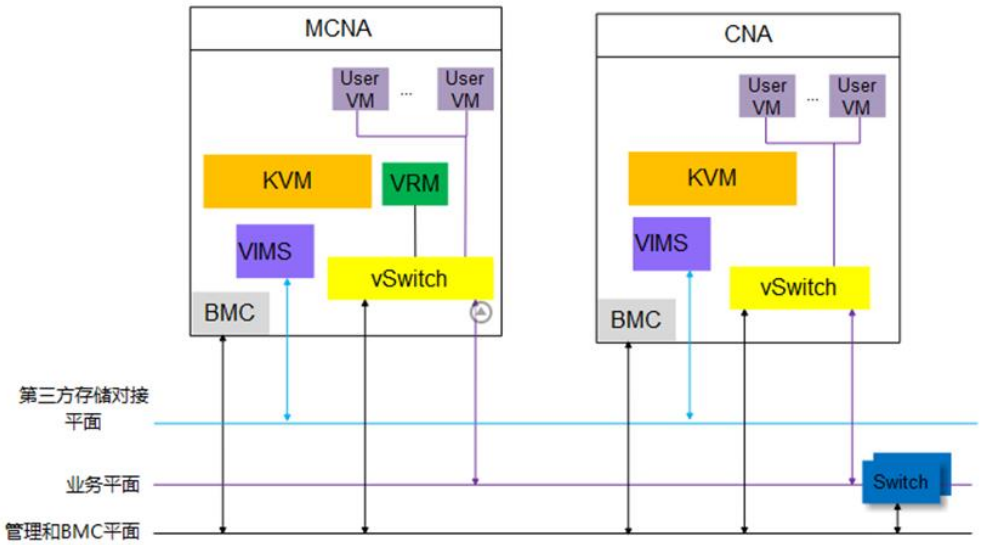
  

| 名称 | 说明 |
| --- | --- |
| | • 管理计算节点上的虚拟机。<br>• 管理计算节点上的计算、存储、网络资源。 |

## 3.1.2 组网介绍

FusionOne Compute 23 版本系统组网包含：管理平面、存储平面、业务平面、BMC 平面。详细的组网情况如下：

图3-3 系统组网图




通信平面类型说明介绍：

- 管理平面：用于系统的业务操作和运维管理，支持 TCP/IP 协议，支持 GE/10GE 组网，可以与业务平面共网卡，通过 VLAN 隔离。
- BMC 平面：服务器设备管理平面，访问服务器设备的运维管理平台。
- 第三方存储对接平面：存储业务数据读写的网络，对应存储网络的存储业务网络平面。
- 业务平面：客户业务通信网络平面，支持 TCP/IP 协议，支持 GE/10GE 组网，可以与管理平面共网卡，通过 VLAN 隔离。

## 3.1.3 典型配置介绍

FusionOne Compute 23 典型配置具体如下：

- **典型配置：**

| 配置项 | 典型配置 | 说明 |
| --- | --- | --- |
| 服务器类型 | 机架服务器 | 根据客户对机柜空间、磁盘大小、密度、PCIE 网卡数量等选择合适的服务器类型。<br>机架服务器：优势为灵活，支持多种硬盘类型，预留多个 PCIE 槽位，支持 GPU 卡。 |
| CPU/内存配置 | 支持 Intel® Xeon®系列 CPU 和 澜起®津逮®系列 CPU；支持鲲鹏 920 处理器<br>支持 DDR4 或 DDR5 内存 | CPU/内存配置根据客户的业务规格和配置可以动态调整配置，提供更多的计算资源。<br>建议 CPU 最低主频 2.0GHz，最少 8 个物理核心。 |
| 磁盘 | SAS HDD：600GB<br>SATA SSD：480GB、960GB | 操作系统盘通常使用 2pcs 相同容量的磁盘创建 RAID1。 |
| 网卡 | 2*10GE+2*10GE | 推荐管理 2*10GE 网口（业务平面可与管理网口复用），外部存储对接网络平面独占 2*10GE 网口。 |

# 3.2 主要功能介绍

## 3.2.1 虚拟化计算

### 服务器虚拟化

将服务器物理资源抽象成逻辑资源，让一台服务器变成几台甚至上百台相互隔离的虚拟服务器，不再受限于物理上的界限，而是让 CPU、内存、磁盘、I/O 等硬件变成可以动态管理的“资源池”，从而提高资源的利用率，简化系统管理。同时硬件辅助虚拟化技术提升虚拟化效率，增加虚拟机的安全性。

- CPU 虚拟化

  FusionOne Compute 将物理服务器的 CPU 虚拟成虚拟 CPU（vCPU），供虚拟机运行时使用。当多个 vCPU 运行时，FusionOne Compute 会在各 vCPU 间动态调度物理 CPU 的能力。

- GPU 直通

  FusionOne Compute 支持将物理服务器上的 GPU（Graphic Processing Unit）直接关联给特定的虚拟机，来提升虚拟机的图形视频处理能力，以满足客户对于图形视频等高性能图形处理能力的需求。GPU 虚拟化。

- GPU 虚拟化

  FusionOne Compute 支持将物理服务器上的 GPU（Graphic Processing Unit）根据资源比例划分，将 GPU 划分为 vGPU，绑定给业务升虚拟机的图形视频处理能力，以满足客户对于图形视频等高性能图形处理能力的需求。

- USB 设备直通

  FusionOne Compute 支持将物理服务器上的 USB 设备直接关联给特定的虚拟机，以满足用户在虚拟化场景下对 USB 设备的使用需求。

### 虚拟机资源管理

客户可以通过自定义方式或基于模板创建虚拟机，并对集群资源进行管理，包括资源自行动态调度（包含负载均衡和动态节能）、虚拟机管理（包含创建、删除、启动、关闭、重启、休眠（X86）、唤醒（X86）虚拟机等）、存储资源管理（包含普通磁盘和共享磁盘的管理）、虚拟机安全管理（包含自定义 VLAN 等），此外，还可以根据业务负载灵活调整虚拟机的 QoS（CPU QoS）。

- 虚拟机生命周期管理

虚拟机支持多种操作方式，用户可根据业务负载灵活调整虚拟机状态。虚拟机操作方式包括：

- 创建/删除/启动/关闭/重启/查询虚拟机

  FusionOne Compute 接受来自业务管理系统的创建虚拟机请求，依据请求中定义的虚拟机规格（vCPU、内存大小、盘大小）、镜像要求、网络要求等，选择合适的物理资源创建虚拟机，并在虚拟机创建完成之后，查询虚拟机运行状态和属性。在使用虚拟机的过程中，用户可以停止、重启、甚至删除自己的虚拟机。该功能为用户提供了基本的虚拟机操作和管理功能，方便用户对虚拟机的使用。

- 休眠/唤醒虚拟机（仅 X86 架构支持）

  当业务处于低负载量运行时，可以只保留部分虚拟机满足业务需求，将其他空闲虚拟机休眠，以降低物理服务器的能耗；当需要业务高负载运行时，再将虚拟机唤醒，以满足高负载业务量正常运行需求。该功能满足业务系统对资源需求的灵活性，提高系统的资源利用率。

- 虚拟机模板

  通过使用虚拟机模板功能，用户可对虚拟机定义规格化模板，并使用模板方式完成虚拟机创建。

- CPU QoS

  虚拟机的 CPU QoS 用于保证虚拟机的计算资源分配，隔离虚拟机间由于业务不同而导致的计算能力相互影响，满足不同业务对虚拟机计算性能的要求，最大程度复用资源，降低成本。

  创建虚拟机时，可根据虚拟机预期部署业务对 CPU 的性能要求而指定相应的 CPU QoS。不同的 CPU QoS 代表了虚拟机不同的计算能力。指定 CPU QoS 的虚拟机，系统对其 CPU 的 QoS 保障，主要体现在计算能力的最低保障和资源分配的优先级。

  CPU QoS 包含如下三个参数：

  - **CPU 资源份额**

    CPU 份额定义多个虚拟机在竞争物理 CPU 资源的时候按比例分配计算资源。

    以一个主频为 2.8GHz 的单核物理主机为例，如果上面运行有三台单 CPU 的虚拟机。三个虚拟机 A，B，C，份额分别为 1000，2000，4000。当三个虚拟机 CPU 满负载运行时，会根据三个虚拟机的份额按比例分配计算资源。份额为 1000 的虚拟机 A 的计算能力约为 400MHz 的，份额为 2000 的虚拟机 B 获得的计算能力约为 800MHz，份额为 4000 的虚拟机 C 获得的计算能力

约为 1600MHz。(以上举例仅为说明 CPU 份额的概念，实际应用过程中情况会更复杂)。

CPU 份额只在各虚拟机竞争计算资源时发挥作用，如果没有竞争情况发生，有需求的虚拟机可以独占物理 CPU 资源，例如，如果虚拟机 B 和 C 均处于空闲状态，虚拟机 A 可以获得整个物理核即 2.8GHz 的计算能力。

- **CPU 资源预留**

  CPU 预留定义了多个虚拟机竞争物理 CPU 资源的时候分配的最低计算资源。

  如果虚拟机根据份额值计算出来的计算能力小于虚拟机预留值，调度算法会优先按照虚拟机预留值的能力把计算资源分配给虚拟机，对于预留值超出按份额分配的计算资源的部分，调度算法会从主机上其他虚拟机的 CPU 上按各自的份额比例扣除，因此虚拟机的计算能力会以预留值为准。

  如果虚拟机根据份额值计算出来的计算能力大于虚拟机预留值，那么虚拟机的计算能力会以份额值计算为准。

  以一个主频为 2.8GHz 的单核物理机为例，如果运行有三台单 CPU 的虚拟机 A、B、C，份额分别为 1000、2000、4000，预留值分别为 700MHz、0MHz、0MHz。当三个虚拟机满 CPU 负载运行时：

  - 虚拟机 A 如果按照份额分配，本应得 400MHz，但由于其预留值大于 400MHz，因此最终计算能力按照预留值 700MHz 算。
  - 多出的(700-400)MHz 按照 B 和 C 各自的份额比例从 B 和 C 处扣除。
  - 虚拟机 B 获得的计算能力约为(800-100)MHz，虚拟机 C 获得的计算能力约为(1600-200)MHz。

    CPU 预留只在各虚拟机竞争计算资源的时候才发挥作用，如果没有竞争情况发生，有需求的虚拟机可以独占物理 CPU 资源。例如，如果虚拟机 B 和 C 均处于空闲状态，虚拟机 A 可以获得整个物理核即 2.8GHz 的计算能力。

- **CPU 资源限额**

  控制虚拟机占用物理 CPU 资源的上限。以一个两 CPU 的虚拟机为例，如果设置该虚拟机 CPU 上限为 3GHz，则该虚拟机的两个虚拟 CPU 计算能力被限制为 1.5GHz。

- 虚拟资源动态复用

  虚拟机空闲时，可自动根据可设置的条件将其部分内存、CPU 等资源释放并归还到虚拟资源池，以供系统分配给其他虚拟机使用。用户可在 Web 界面上对动态资源进行监控。

- 虚拟机统计

  支持虚拟机、磁盘资源使用情况统计报表。

### 虚拟机资源动态调整

FusionOne Compute 支持虚拟机资源动态调整，用户可以根据业务负载动态调整资源的使用情况。虚拟机资源调整包括：

- 离线/在线调整 vCPU 数目

  无论虚拟机处于离线（关机）或在线状态，用户都可以根据需要增加虚拟机的 vCPU 数目。虚拟机处于离线状态时，用户可以根据需要减少虚拟机的 vCPU 数目。通过离线/在线调整虚拟机 vCPU 数目，可以满足虚拟机上业务负载发生变化时对计算能力灵活调整的需求。

- 离线/在线调整内存大小

  无论虚拟机处于离线或在线状态，用户都可以根据需要增加虚拟机的内存容量。虚拟机处于离线状态时，用户可以根据需要减少虚拟机的内存容量。通过离线/在线调整内存大小，可以满足虚拟机上业务负载发生变化时对内存灵活调整的需求。

- 离线/在线添加/删除网卡

  虚拟机在线/离线状态下，用户可以挂载或卸载虚拟网卡，以满足业务对网卡数量的需求。

- 离线/在线挂载虚拟磁盘

  无论虚拟机处于离线或在线状态，用户都可以挂载虚拟磁盘，在不中断用户业务的情况下，增加虚拟机的存储容量，实现存储资源的灵活使用。

**说明**

虚拟机处于离线或在线状态，且虚拟机使用的磁盘是虚拟化存储时，用户可通过增加已有磁盘容量的方式进行虚拟机存储容量的扩容。

## 3.2.2 虚拟化网络

### 虚拟网卡

虚拟网卡均有自己的 IP 地址、MAC 地址，从网络角度来看，虚拟网卡与物理网卡一致。FusionOne Compute 实现多队列、虚拟交换、QoS、上行链路聚合功能，提升虚拟网卡的 I/O 性能。

### 网络 I/O 控制

网络 QoS 策略提供带宽配置控制能力，包含如下方面：

- 基于端口组成员接口发送方向与接收方向的带宽控制
- 基于端口组的每个成员接口提供流量整形、带宽优先级的控制能力。

### 分布式虚拟交换机

分布式交换机的功能类似于普通的物理交换机，每台主机都连接到分布式交换机中。分布式交换机的一端是与虚拟机相连的虚拟端口，另一端是与虚拟机所在主机上的物理以太网适配器相连的上行链路。通过它可以连接主机和虚拟机，实现系统网络互通。另外，分布式交换机在所有关联主机之间作为单个虚拟交换机使用。此功能可使虚拟机在跨主机进行迁移时确保其网络配置保持一致。

## 3.2.3 虚拟化存储

### 虚拟存储管理

存储虚拟化是将存储设备抽象为数据存储，虚拟机在数据存储中作为一组文件存储在自己的目录中。数据存储是逻辑容器，类似于文件系统，它将各个存储设备的特性隐藏起来，并提供一个统一的模型来存储虚拟机文件。存储虚拟化技术可以更好的管理虚拟基础架构的存储资源，使系统大幅提升存储资源利用率和灵活性，提高应用的正常运行时间。

能够封装为数据存储的存储单元包括：

- SAN（Storage Area Network ）存储（包括 iSCSI 或光纤通道的 SAN 存储）上划分的 LUN（Logical Unit Number）。
- 超聚变分布式块存储上的存储池。
- 主机的本地硬盘。

数据存储可支持如下文件系统格式：

- 虚拟镜像管理系统（VIMS）

  为存储虚拟机而优化的高性能文件系统。主机可以将虚拟镜像管理系统数据存储部署在任何基于 iSCSI 的本地或联网存储设备上，包括光纤通道、以太网光纤通道的 iSCSI SAN 设备。
- 本地存储（EXT4）

FusionOne Compute 支持服务器的本地磁盘虚拟化。

- NAS

  FusionOne Compute 支持对接外部的 NAS 存储，用于存储虚拟机部署的镜像文件。

## 虚拟存储精简置备

虚拟存储精简置备是一种通过灵活的按需分配存储空间来优化存储利用率的方法。精简置备可以为用户虚拟出比实际物理存储更大的虚拟存储空间，只有写入数据的虚拟存储空间才会为之真正分配物理存储，未写入的虚拟存储空间不占用物理存储资源，从而提高存储利用率。

虚拟存储精简配置基于磁盘提供，管理员可以按“普通”格式或“精简”格式分配虚拟磁盘文件。

- 存储无关

  虚拟存储精简配置与操作系统、硬件完全无关，因此只要使用虚拟镜像管理系统，就能提供虚拟存储精简配置功能。

- 容量监控

  提供数据存储容量预警，可以设置阈值，当存储容量超过阈值时产生告警。

## 虚拟机快照

虚拟机快照功能是指把某一时刻的虚拟机状态像照片一样保存下来，在需要的时候用快照把虚拟机恢复到快照时的状态。虚拟机快照保存的内容包括虚拟机所有磁盘的信息。

## 存储热迁移

虚拟机正常运行时，管理员可通过手动操作，将虚拟机的磁盘迁移至其他存储单元。存储热迁移可在存储虚拟化管理下的同一个存储设备内、不同存储设备之间进行迁移。热迁移使客户在业务无损的情况下动态调整虚拟机存储资源，以实现设备维护等操作。

# 3.3 关键特性

## 3.3.1 跨主机热迁移

虚拟机热迁移是指在不中断业务的情况下，将同一个集群中虚拟机从一台物理服务器移动至另一台物理服务器。虚拟机管理器提供内存数据快速复制和共享存储技术，确保虚拟机迁移前后数据不变，主要应用在如下几个场景：

- 在进行服务器操作维护前，系统维护人员将该服务器上的虚拟机迁移到其他服务器，降低操作维护过程中业务中断的风险。
- 在进行服务器升级操作前，系统维护人员将该服务器上的虚拟机迁移到其他服务器，升级完成后将所有虚拟机迁回，降低服务器升级过程中业务中断的风险。
- 将空闲服务器上的虚拟机迁移到其他服务器，将没有负载的服务器关闭，降低业务运行成本。

FusionOne Compute 支持 IMC 特性，支持不同代次的 Intel CPU 跨代次热迁移。通过对计算集群配置 IMC 策略，可以确保集群内的主机向虚拟机提供相同的 CPU 功能集，即使这些主机的实际 CPU 不同，也不会因 CPU 不兼容而导致迁移虚拟机失败。

虚拟机热迁移类型如下表所示。

| 迁移类型 | 子类 | 说明 |
| --- | --- | --- |
| 手动迁移 | 按目的迁移 | 系统维护人员通过 FusionOne Compute 的虚拟机迁移功能，手动迁移一台虚拟机到另一台服务器上。 |
| 自动迁移 | 虚拟机资源调度 | 在同一个集群内，系统根据预先设定的虚拟机调度策略，对虚拟机进行自动迁移。 |

当服务器节点配置 Intel Sapphire Rapids CPU 时，FunsionOne Compute 可开启 QAT 加速来实现虚拟机迁移时内存数据的压缩和解压缩，减少数据的传输量，提高迁移效率。从而大大提升虚拟机的迁移速度和成功率，并且不会带来额外的 CPU 资源占用。

## 3.3.2 跨存储热迁移

虚拟机正常运行时，管理员可通过手动操作，将虚拟机的磁盘迁移至其他存储单元。存储热迁移可在存储虚拟化管理下的同一个存储设备内、不同存储设备之间进行迁

移。热迁移使客户在业务无损的情况下动态调整虚拟机存储资源，以实现设备维护等操作。

## 3.3.3 虚拟机高可用性（HA）

虚拟机高可用性是当计算节点上的虚拟机出现故障时，系统自动将故障的虚拟机在正常的计算节点上重新创建，使故障虚拟机快速恢复。

当系统检测到虚拟机故障时，系统将选择正常的计算节点，将故障虚拟机在正常的计算节点上重新创建。

- 计算节点掉电恢复或重启

  当计算节点掉电恢复或重启时，系统将计算节点上具有 HA 属性的虚拟机重新创建至其他计算节点。

- 虚拟机蓝屏

  当系统检测到虚拟机蓝屏故障且该虚拟机蓝屏处理策略配置为 HA 时，系统选择其他正常的计算节点重新创建虚拟机。

## 3.3.4 虚拟机回收站

FusionOne Compute 支持临时存放用户通过安全删除/普通删除执行的放入回收站中的虚拟机，存放在回收站的虚拟机可以恢复，便于虚拟机的日常维护和使用。

在删除虚拟机的时候，可以选择 xx 天之后删除，虚拟机不会立即删除，而是在用户选择的天数（1~60 天，默认为 30 天）之后进行自动回收删除。用户也可以进入回收站页面，对待删除的虚拟机进行恢复，或者在回收站彻底删除。

## 3.3.5 虚拟机安全删除

FusionOne Compute 支持两种虚拟机的删除方式，安全删除与普通删除，安全删除虚拟机可以针对安全隐私比较高的虚拟机的数据销毁，将对删除的磁盘内部的数据进行安全低格处理，确保关键信息资产不泄漏。

- 安全删除：通过覆盖性擦写对磁盘空间进行删除，避免数据被恢复，安全性高，但删除速度较慢，且删除时会占用系统资源。
- 普通删除：通过破坏磁盘文件系统对磁盘空间进行删除，删除速度快，但存在通过残余信息恢复数据的可能，安全性差。

## 3.3.6 动态资源调度（DRS&DPM）

动态资源调度 DRS（Dynamic Resource Scheduler），指采用智能负载均衡调度算法，并结合动态电源管理功能，通过周期性检查同一集群资源内各个主机的负载情况，在不同的主机间迁移虚拟机，从而实现同一集群内不同主机间的负载均衡，并最大程度降低系统的功耗。

- 系统轻载时，将迁移部分虚拟机，并将其集中在部分物理主机，随即将空闲主机下电。
- 系统重载时，将启动部分物理主机，并将虚拟机均衡分布在各主机中，以保证资源的供应和用户的体验。
- 通过计划任务，可根据系统运行的情况，分时段采取不同的资源调度策略，以满足不同场景的用户需求。

动态节能调度 DPM（Dynamic Power Management）

动态节能调度和负载均衡配合使用，仅在负载均衡调度打开之后才能使用动态节能调度功能。在一个集群内，对计算服务器和虚拟机运行状态进行监控的过程中，如果发现集群内业务量减少，系统将业务集中到少数计算服务器上，并自动将剩余的计算服务器关机；如果发现集群内业务量增加，系统将自动唤醒计算服务器并分担业务。

## 3.3.7 虚拟机资源 QoS

- CPU QoS

  虚拟机的 CPU QoS 用于保证虚拟机的计算资源分配，隔离虚拟机间由于业务不同而导致的计算能力相互影响，满足不同业务对虚拟机计算性能的要求，最大程度复用资源，降低成本。

  创建虚拟机时，可根据虚拟机预期部署业务对 CPU 的性能要求而指定相应的 CPU QoS。不同的 CPU QoS 代表了虚拟机不同的计算能力。指定 CPU QoS 的虚拟机，系统对其 CPU 的 QoS 保障，主要体现在计算能力的最低保障和资源分配的优先级。

  CPU QoS 包含如下三个参数：

  - **CPU 资源份额**

    CPU 份额定义多个虚拟机在竞争物理 CPU 资源的时候按比例分配计算资源。

    以一个主频为 2.8GHz 的单核物理主机为例，如果上面运行有三台单 CPU 的虚拟机。三个虚拟机 A，B，C，份额分别为 1000，2000，4000。当三个虚

拟机 CPU 满负载运行时，会根据三个虚拟机的份额按比例分配计算资源。份额为 1000 的虚拟机 A 的计算能力约为 400MHz 的，份额为 2000 的虚拟机 B 获得的计算能力约为 800MHz，份额为 4000 的虚拟机 C 获得的计算能力约为 1600MHz。(以上举例仅为说明 CPU 份额的概念，实际应用过程中情况会更复杂)。

CPU 份额只在各虚拟机竞争计算资源时发挥作用，如果没有竞争情况发生，有需求的虚拟机可以独占物理 CPU 资源，例如，如果虚拟机 B 和 C 均处于空闲状态，虚拟机 A 可以获得整个物理核即 2.8GHz 的计算能力。

- **CPU 资源预留**

  CPU 预留定义了多个虚拟机竞争物理 CPU 资源的时候分配的最低计算资源。

  如果虚拟机根据份额值计算出来的计算能力小于虚拟机预留值，调度算法会优先按照虚拟机预留值的能力把计算资源分配给虚拟机，对于预留值超出按份额分配的计算资源的部分，调度算法会从主机上其他虚拟机的 CPU 上按各自的份额比例扣除，因此虚拟机的计算能力会以预留值为准。

  如果虚拟机根据份额值计算出来的计算能力大于虚拟机预留值，那么虚拟机的计算能力会以份额值计算为准。

  以一个主频为 2.8GHz 的单核物理机为例，如果运行有三台单 CPU 的虚拟机 A、B、C，份额分别为 1000、2000、4000，预留值分别为 700MHz、0MHz、0MHz。当三个虚拟机满 CPU 负载运行时：

  - 虚拟机 A 如果按照份额分配，本应得 400MHz，但由于其预留值大于 400MHz，因此最终计算能力按照预留值 700MHz 算。
  - 多出的(700-400)MHz 按照 B 和 C 各自的份额比例从 B 和 C 处扣除。
  - 虚拟机 B 获得的计算能力约为(800-100)MHz，虚拟机 C 获得的计算能力约为(1600-200)MHz。

  CPU 预留只在各虚拟机竞争计算资源的时候才发挥作用，如果没有竞争情况发生，有需求的虚拟机可以独占物理 CPU 资源。例如，如果虚拟机 B 和 C 均处于空闲状态，虚拟机 A 可以获得整个物理核即 2.8GHz 的计算能力。

- **CPU 资源限额**

  控制虚拟机占用物理 CPU 资源的上限。以一个两 CPU 的虚拟机为例，如果设置该虚拟机 CPU 上限为 3GHz，则该虚拟机的两个虚拟 CPU 计算能力被限制为 1.5GHz。

- 网络 QoS 策略提供带宽配置控制能力，QoS 功能不支持同一主机上虚拟机之间的流量限制。包含如下方面：

- 基于端口组成员接口发送方向与接收方向的带宽控制
- 基于端口组的每个成员接口提供流量整形、带宽优先级的控制能力。

## 3.3.8 自动精简配置

存储自动精简配置（Thin Provisioning），可以为客户虚拟出比实际物理存储更大的虚拟存储空间，为用户提供存储超分配的能力。只有写入数据的虚拟存储空间才能真正分配到物理存储，未写入的虚拟存储空间不占用物理存储资源。FusionOne Compute 虚拟化的存储自动精简配置不依赖于存储设备。

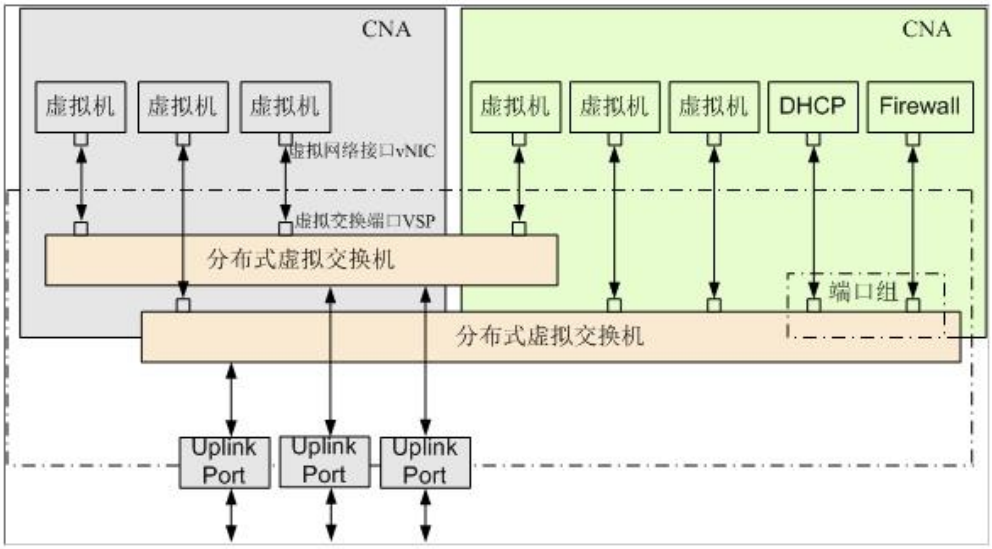
存储自动精简配置的应用场景主要针对虚拟机的用户数据卷。当企业或运营商宣称提供的容量较大，但用户实际却用不完时，可通过存储自动精简配置帮助企业或运营商大幅降低存储的初始投资成本。

## 3.3.9 分布式虚拟交换机

分布式虚拟交换管理，即实现系统管理员对一至多台 CNA 服务器上的虚拟交换机的物理端口和虚拟端口进行配置/维护。

分布式虚拟交换机的模型如下图所示。

图3-4 分布式交换机示意图



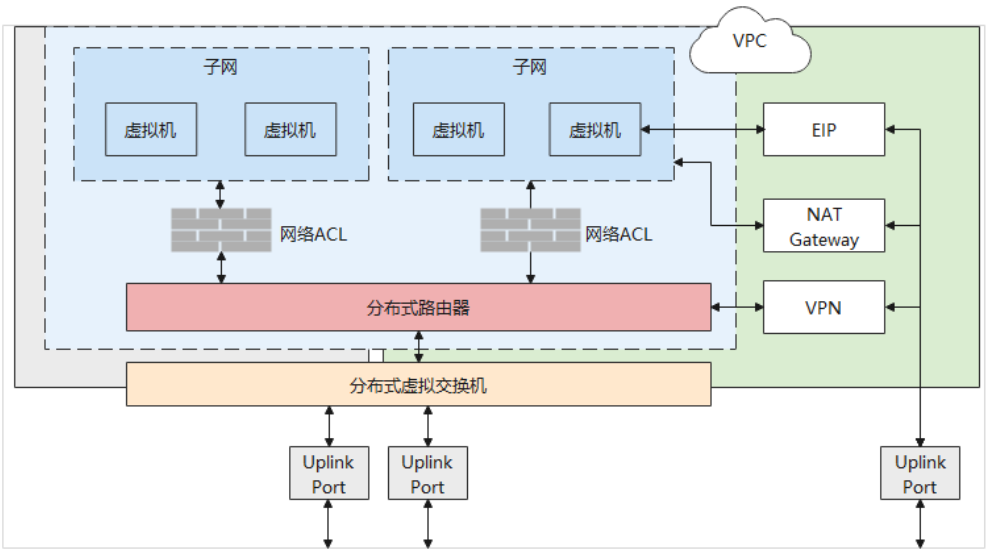

分布式虚拟交换机模型的基本特征：

- 用户可以配置多个分布式交换机，每个分布式交换机可以覆盖集群中的多个 CNA 节点。
- 每个分布式交换机具有多个分布式的虚拟端口 VSP（Virtual Switch Port），每个 VSP 具有各自的属性（速率、统计和安全组等），为了管理方便采用端口组管理相同属性的一组端口，相同端口组的 VLAN（Virtual Local Area Network）、QoS、安全属性相同。
- 每个分布式交换机可以配置一个 Uplink 上行链路组，用于虚拟机对外的通信，Uplink 上行链路组可以包含多个物理网卡，这些物理网卡可以配置网口聚合和负载均衡策略。
- 每个虚拟机可以具有多个 vNIC（Virtual Network Interface Card）接口，vNIC 可以和交换机的 VSP 一一对接。

  分布式虚拟交换默认支持的 OVS 网络模式，业务网络可以根据用户需要，开启 OVN 网络（SDN 控制器），开启后可支持虚拟私有云（Virtual Private Cloud，简称 VPC），支持用户自主配置和管理虚拟网络环境和虚拟化向云平台的无缝演进。用户可以在 VPC 中定义子网、路由表、网络 ACL、NAT 网关等网络特性，通过 VPC 方便地管理、配置内部网络，进行安全、快捷的网络变更。

图3-5 VPC 组网图




虚拟私有云的基本特征包括：

1. 用户可以配置多个 VPC，并管理 VPC 的路由表，每个 VPC 通过分布式路由器覆盖多个节点。

2. 每个 VPC 内部可以通过 IP 网段划分出多个子网，子网内的虚拟机具备相同的网络和安全属性（网络 ACL、VLAN、QoS 等），为方便 IP 地址管理，用户可以启用子网 DHCP，对子网内虚拟机进行 IP 地址的动态分配。

3. 子网内虚拟机内部交互和外部交互的基本方式由分布式路由器提供，同时可以根据业务需要为虚拟机分配 EIP 或通过 NAT 网关，VPN 等方式对外通信。

## 3.3.10 用户态交换模式（DPDK）

通过使用 DPDK（Data Plane Development Kit，DPDK 是一系列库和驱动的集合）技术，在数据平面开发套件用来在 x86 平台进行快速的数据包处理。它通过环境抽象层旁路内核协议栈、轮询模式的报文无中断收发、优化内存/缓冲区/队列管理、基于网卡多队列和流识别的负载均衡等多项技术，实现了在 x86 处理器架构下的高性能报文转发能力，提高虚拟机网络性能。

## 3.3.11 SR-IOV（x86）

支持 SR-IOV 特性，实现了将 PCI 功能分配到多个虚拟接口以在虚拟化环境中共享一个 PCI 设备的资源。SR-IOV 能够让网络传输绕过软件模拟层，直接分配到虚拟机。这样就降低了软件模拟层中的 I/O 开销。

## 3.3.12 网络安全组

针对部分安全要求进行控制的虚拟机，可以配置安全组，便于管理员对业务虚拟机的网络安全进行控制。用户可以根据实际需求，对出入口流量，根据配置规则，配置白名单、黑名单进行网络访问控制。

- 白名单：添加安全组后，默认禁止该安全组的虚拟机与外界互通，需要在安全组中添加规则后才可使网卡在该安全组的虚拟机与外界互通，默认为白名单。
- 黑名单：添加安全组后，默认该安全组的虚拟机可以与外界互通。需要在安全组中添加规则后才可拒绝网卡在该安全组的虚拟机与外界某些地址互通。

## 3.3.13 GPU 直通

FusionOne Compute 支持将物理服务器上的 GPU（Graphic Processing Unit）直接关联给特定的虚拟机，来提升虚拟机的图形视频处理能力，以满足客户对于图形视频等高性能图形处理能力的需求。

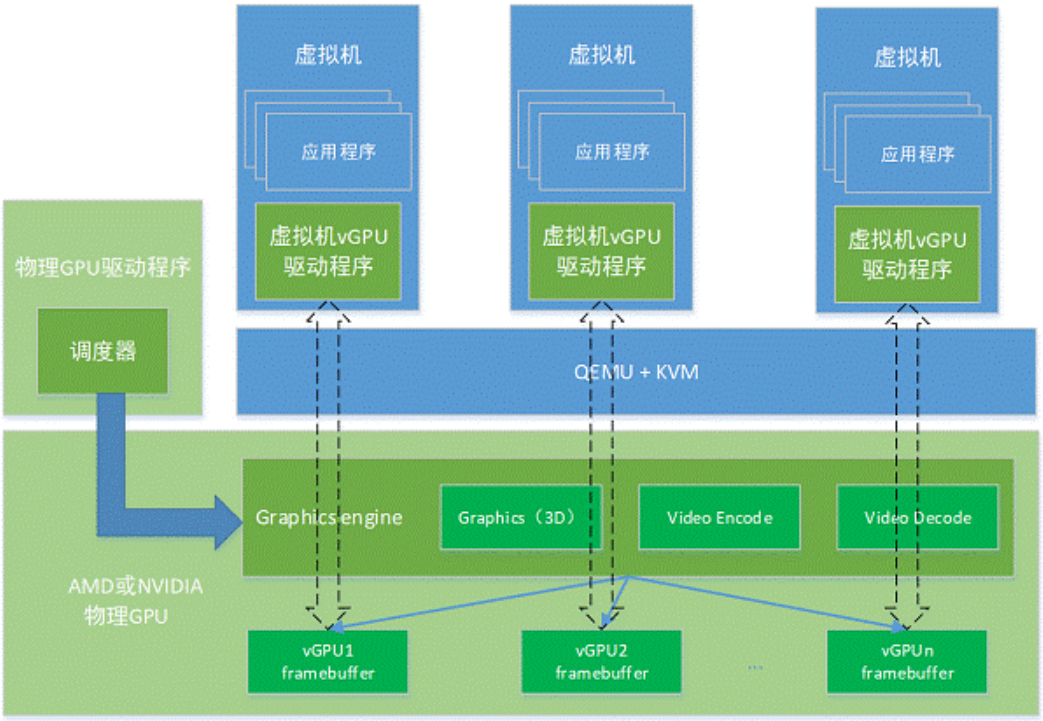
## 3.3.14 GPU 虚拟化

vGPU（Virtual GPU）是一种在虚拟化环境中提供虚拟显卡的技术，通过在宿主机上运行虚拟化软件，在虚拟机中模拟出一个独立的显卡设备，让虚拟机可以直接访问显卡资源，从而实现高性能图形加速。

每个 vGPU 都类似于物理 GPU，有固定的显存大小，一个或者更多的虚拟显示输出。vGPU 的存在创建时就从物理 GPU 中分配出来，并且是独占的。

vGPU 跟传统的 GPU 类似，具有固定数量的 GPU 帧缓冲器和一个或多个虚拟显示输出。在 vGPU 虚拟机被创建的时候，帧缓冲区就被分配到物理 GPU 的帧缓冲区中，vGPU 保留对该缓冲区的独占使用，直到它被销毁为止。留在物理 GPU 上的所有 vGPU 虚拟机共享对 GPU 引擎的访问，包括图形（3D）、视频编解码引擎等。

vGPU 技术解决了在虚拟化环境中无法满足图形处理需求的问题。在传统虚拟化环境中，虚拟机使用的是宿主机的物理显卡，虚拟机的图形性能和宿主机的性能有很大关系。如果宿主机的显卡性能较低，那么虚拟机的图形性能也会很差。而 vGPU 技术则让虚拟机可以独立使用虚拟显卡，不受宿主机的限制，从而提高了虚拟机的图形性能。




FusionOne HCI 提供对多种 Gpu 卡的虚拟化支持，通过将一块 GPU 虚拟化为多块虚拟 GPU，在虚拟机中可以使用和直接硬件访问 GPU 一样的方式进行计算，并提供较

高的隔离性、安全性和通用性，能够满足虚拟化环境下不同用户的需求，并充分利用宿主机上的 GPU 资源，提高整体的计算能力和性能。

## 3.3.15 无代理防病毒

为了对主机中所有虚拟机进行病毒防护，若采用传统防病毒产品，则需要在每台虚拟机本地安装防病毒产品，这样不仅会占用较多资源，而且在全盘扫描、病毒更新等场景下会造成病毒风暴。

为了解决该问题， FusionOne Compute 提供了防病毒所需 API，防病毒厂家可基于 API 进行二次开发，形成虚拟化防病毒解决方案，做到仅需在一台特殊的安全虚拟机中部署防病毒引擎，其他用户虚拟机在本地安装轻量级驱动，通过安全服务虚拟机提供的服务即可完成杀毒。

## 3.3.16 VMware 虚拟机模板导入

FusionOne Compute 支持导入 Vmware 的 OVF 和 OVA 格式的虚拟机模板，其原理如下：

- 按要求从 Vmware 中把虚拟机导出为 OVF 或者 OVA
- 在 FusionOne Compute 中选择此 OVF 或者 OVA 进行导入
- FusionOne Compute 对配置自动识别和转换
- 对于无法自动转换的配置，提示用户进行选择
- FusionOne Compute 执行磁盘格式转换、驱动替换和注入
- 在 FusionOne Compute 中启动此虚拟机进行验证

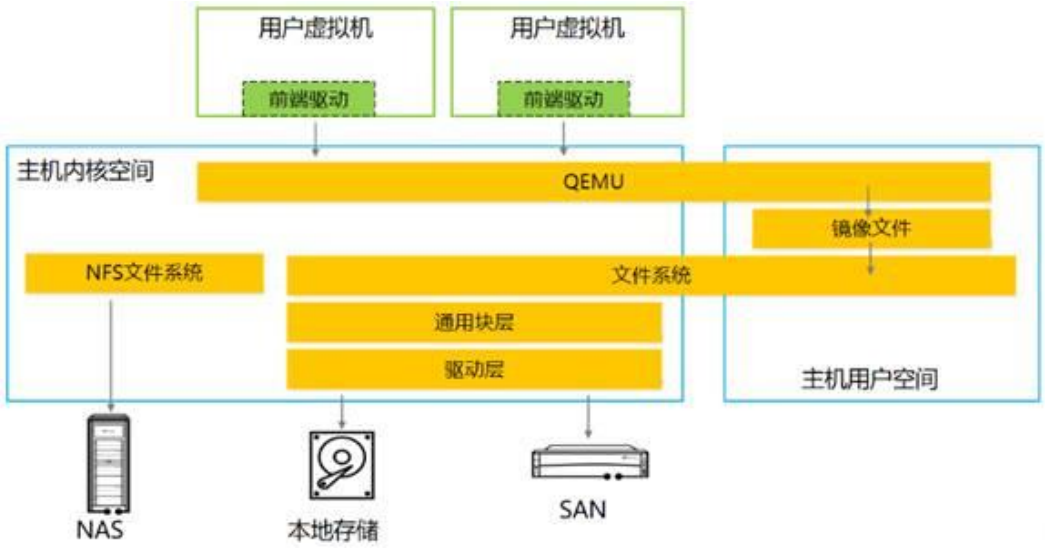
## 3.3.17 虚拟镜像管理系统（VIMS）

Vims 是一种高性能的集群文件系统，使虚拟化技术的应用超出了单个存储系统的限制可让多个虚拟机共同访问一个整合的集群式存储池，从而显著提高了资源利用概。

存储区域网络 SAN，是一种高速的、专门用于存储操作的网络，通常独立于计算机局域网（LAN）。提供主机和存储系统之间的数据传输，网络内部数据传输的效率快。常见的架构有 FC SAN、IP SAN，具有存储容量利用率高、兼容性高、传输距离远、高带宽、主机、存储设备可以独立扩展等优点。

FusionOne HCI 通过在 IPSan 和 FCSan 的卷上创建一个 VIMS 共享文件系统，并在多个节点上挂载，实现了多个节点共享一个存储卷。可以在此共享文件系统上创建虚拟机，由于数据在多个节点间共享，可以实现虚拟机迁移时免迁移存储数据。



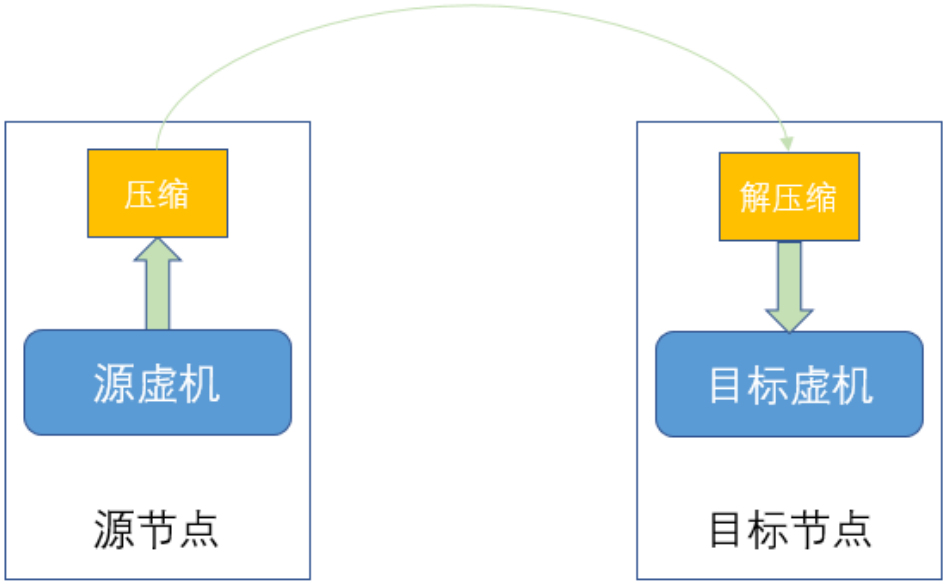

## 3.3.18 虚拟机 QAT 迁移加速

Intel 最新一代 CPU 集成了 QAT 设备，此设备可用于数据的压缩和解压缩，并且处理过程不会占用 CPU 的资源。对于 HCI 虚拟化平台而言，可以利用此“免费”的设备来对我们一些功能进行加速，充分释放硬件能力。典型的场景是虚拟机的热迁移。

虚拟机在热迁移过程中需要传输大量的内存数据以及迁移过程中产生的“脏页”，由于传输数据量大，并且要求传输速度比脏页产生速度快（否则会出现热迁移失败），这个过程对网络有较高要求，并且会与其他业务竞争带宽资源。

传统的方案是在源端利用 CPU 对数据进行压缩，在目的端利用 CPU 对数据进行解压缩，从而大大减小需要传输的数据。但此方案需要使用较多的 CPU 资源。利用 CPU 集成的 QAT 设备来做迁移数据的压缩解压缩很好解决了对网络带宽占用，提高传输效率。同时又不占用 CPU 资源，解决了 HCI 平台中虚拟机热迁移占用资源多，迁移慢的问题，同时降低了虚拟机热迁移失败的概率。




## 3.4 不同架构关键特性对比

| 特性 | x86(Intel) | ARM（鲲鹏） | 备注 |
| --- | --- | --- | --- |
| 跨主机热迁移 | 支持 | 支持 | 二者特性相同 |
| 跨存储热迁移 | 支持 | 支持 | 二者特性相同 |
| 虚拟机高可用 | 支持 | 支持 | ARM 不支持虚拟机故障 HA 处理策略 |
| 虚拟机回收站 | 支持 | 支持 | 二者特性相同 |
| 虚拟机安全删除 | 支持 | 支持 | 二者特性相同 |
| 动态资源调度 | 支持 | 支持 | 二者特性相同 |
| 虚拟机资源 QoS | 支持 | 支持 | 二者特性相同 |
| 自动精简配置 | 支持 | 支持 | 二者特性相同 |
| 分布式虚拟交换机 | 支持 | 支持 | 二者特性相同 |
| 用户态交换模式 | 支持 | 支持 | 二者特性相同 |

| 特性 | x86(Intel) | ARM（鲲鹏） | 备注 |
| --- | --- | --- | --- |
| SR-IOV | 支持 | 支持 | 二者特性相同 |
| 网络安全组 | 支持 | 支持 | 二者特性相同 |
| GPU 直通 | 支持 | 支持（仅支持 T4） | - |
| GPU 虚拟化 | 支持 | 不支持 | 二者特性相同 |
| 无代理防病毒 | 支持 | 不支持 | - |
| Vmware 虚拟机模板导入 | 支持 | 不支持（指令集差异无法兼容） | - |
| 虚拟镜像管理系统 | 支持 | 支持 | 二者特性相同 |
| 虚拟机 QAT 迁移加速 | 支持 | 不支持 | 仅 Intel V7 CPU 支持。 |

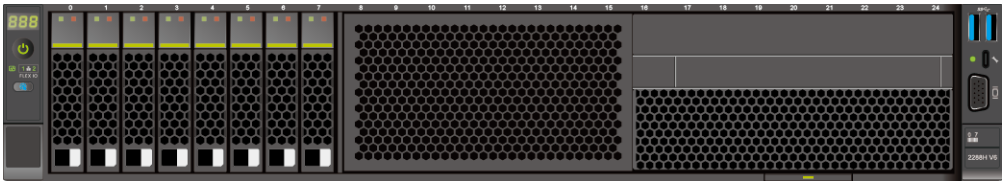
# 4 硬件配置介绍

FusionOne Compute 典型配置机型支持包括 2288H V6、2488H V6、2288H V7 的 x86 服务器，KunLun 2280、KunLun G2280 的 ARM 服务器。可根据客户要求，灵活的配置客户需要的硬件设备。详细配置介绍如下：



## 4.1 x86 节点

图4-1 2288H V6 计算节点服务器

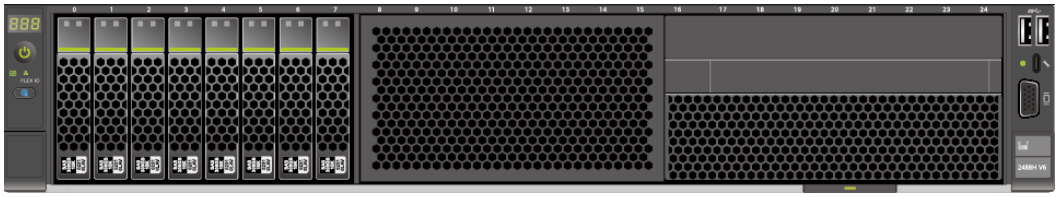


| | |
|---|---|
| 形态 | 2U2P 机架服务器 |
| 处理器 | 2 个 Intel Xeon Icelake Processors (43、53、63、83 系列) |
| 内存插槽 | 32 个 DDR4 DIMM 插槽 |
| 硬盘数量 | 8 个 2.5 英寸 SAS/SATA 硬盘 |

| | |
|---|---|
| RAID 支持 | 支持 RAID1 |
| OCP 网卡 | 支持 GE\10GE\25GE |
| PCIe 扩展 | • 服务器后面板配置硬盘模组/PCIe Riser 模组机型：支持 1 个 RAID 控制扣卡专用的 PCIe 扩展槽位，2 个 OCP 3.0 网卡专用槽位，8 个标准的 PCIe 扩展槽位。<br>• 服务器后面板配置 4 张 GPU 卡机型：支持 1 个 RAID 控制扣卡专用的 PCIe 扩展槽位，2 个 OCP 3.0 网卡专用槽位，1 个标准的 PCIe 扩展槽位，4 个 GPU 专用的 PCIe 扩展槽位。<br>说明<br>具体可用的插槽以实际配置的为准。 |

图4-2 2488H V6 计算节点服务器



| | |
|---|---|
| 形态 | 2U4P 机架服务器 |
| 处理器 | 4 个 Intel Xeon CooperLake Processors (53、63、83 系列) |
| 内存插槽 | 48 个 DDR4 DIMM 插槽 |
| 硬盘数量 | 8 个 2.5 英寸 SAS/SATA 硬盘 |
| RAID 支持 | 支持 RAID1 |
| OCP 网卡 | 支持 GE\10GE\25GE |
| PCIe 扩展 | 支持 11 个 PCIe 3.0 扩展槽位。 |

| | |
|---|---|
| | 支持 1 个 OCP 3.0 网卡专用的 FLEX IO 插卡扩展槽位，6 个 Riser 卡转出的 PCIe 扩展槽位，4 个板载 PCIe 扩展槽位。<br>说明<br>具体可用的插槽以实际配置的为准。 |

图4-3 2288H V7 计算节点服务器


| | |
|---|---|
| 形态 | 2U2P 机架服务器 |
| 处理器 | 2 个 Intel Xeon Sapphire Rapids Processors<br>(44、54、64、84 系列) |
| 内存插槽 | 32 个 DDR5 DIMM 插槽 |
| 硬盘数量 | 8 个 2.5 英寸 SAS/SATA 硬盘 |
| RAID 支持 | 支持 RAID1 |
| OCP 网卡 | 支持 GE\10GE\25GE |
| PCIe 扩展 | • 服务器后面板配置硬盘模组/PCIe Riser 模组机型：支持 2 个 OCP 3.0 网卡专用槽位，8 个标准的 PCIe 扩展槽位。<br>• 服务器后面板配置 4 张 GPU 卡机型：支持 2 个 OCP 3.0 网卡专用槽位，2 个标准的 PCIe 扩展槽位，4 个 GPU 专用的 PCIe 扩展槽位。<br>说明<br>具体可用的插槽以实际配置的为准。 |

# 4.2 ARM 节点

图4-4 KunLun G2280 计算节点服务器

| | |
|---|---|
| 形态 | 2U2P 机架服务器 |
| 处理器 | 2 个鲲鹏 920 处理器 |
| 内存插槽 | 配置鲲鹏 920 5220、3210、5225F 和 5221K 处理器时 最多支持 16 个 DDR4 DIMM 插槽<br>配置鲲鹏 920 5250、7260、7265F 和 5255F 处理器时 最多支持 32 个 DDR4 DIMM 插槽 |
| 硬盘数量 | 8 个 2.5 英寸 SAS/SATA 硬盘 |
| RAID 支持 | 支持 RAID1 |
| OCP 网卡 | 支持 GE\10GE\25GE |
| PCIe 扩展 | 服务器后面板配置硬盘模组/PCIe Riser 模组机型：支持 2 个 OCP 3.0 网卡专用槽位，8 个标准的 PCIe 扩展槽位（最多可支持安装 7pcs Atlas 300I）。<br>说明<br>具体可用的插槽以实际配置的为准。 |

图4-5 KunLun 2280 计算节点服务器

| | |
|---|---|
| 形态 | 2U2P 机架服务器 |
| 处理器 | 2 个鲲鹏 920 处理器 |
| 内存插槽 | 配置鲲鹏 920 5220、3210、5225F 和 5221K 处理器时 最多支持 16 个 DDR4 DIMM 插槽<br>配置鲲鹏 920 5250、7260、7265F 和 5255F 处理器时 最多支持 32 个 DDR4 DIMM 插槽 |
| 硬盘数量 | 8 个 2.5 英寸 SAS/SATA 硬盘 |
| RAID 支持 | 支持 RAID1 |
| OCP 网卡 | 支持 GE\10GE\25GE |
| PCIe 扩展 | 服务器后面板配置硬盘模组/PCIe Riser 模组机型：支持 2 个 OCP 3.0 网卡专用槽位，8 个标准的 PCIe 扩展槽位。<br>说明<br>具体可用的插槽以实际配置的为准。 |

# 5 系统可靠性

分布式存储系统提供了数据跨节点的保护能力在多个硬盘或者节点故障时也能够继续提供服务，将数据放置到同一个节点池内不同节点的不同硬盘上，数据获得了跨节点的可靠性和故障快速恢复的能力。同时通过硬件的冗余配置提供系统的可用性。



## 5.1 硬件可靠性

FusionOne Compute 选用高可靠的自研硬件，通过系统冗余设计保证系统可靠性，具有如下特点：

- 单板硬件采用电信级器件及加工工艺流程，可显著提高系统可靠性。
- 支持热插拔的 SAS/SATA/NVMe 硬盘，支持 RAID1 保护。使用 SSD 后的可靠性远远高于传统机械硬盘，从而能够延长系统运行时间。
- 整机提供 2 个热插拔电源模块，支持 1+1 冗余；提供 4 个热插拔风扇模块，支持 N+1。
- 提供 iBMC 直连管理接口，支持 iBMC 近端运维，提升运维效率。
- 提供 iBMC 直连管理接口，支持 iBMC 近端运维，提升运维效率。
- 支持 FPC Failure Prediction and Correction ）功能，对内存故障进行预测，并利用多种自愈技术，做出自愈隔离，以免影响业务正常运行。
- 网络双平面设计。

- 支持节点内磁盘漫游，磁盘节点换位插拔可自动识别，5 分钟内互换槽位插入，系统不发生数据重构。

## 5.2 软件可靠性

### 管理节点冗余

管理节点 VRM 支持主备冗余部署，当其中一个 VRM 节点故障后，备节点可自动升为主节点。

### 虚拟机 HA

虚拟机 HA 是当计算节点上的虚拟机出现故障时，系统自动将故障的虚拟机在正常的计算节点上重新创建，使故障虚拟机快速恢复。

当系统检测到虚拟机故障时，系统将选择正常的计算节点，将故障虚拟机在正常的计算节点上重新创建。

- 计算节点掉电恢复或重启

  当计算节点掉电恢复或重启时，系统将计算节点上具有 HA 属性的虚拟机重新创建至其他计算节点。

- 虚拟机蓝屏（Intel）

  当系统检测到虚拟机蓝屏故障且该虚拟机蓝屏处理策略配置为 HA 时，系统选择其他正常的计算节点重新创建虚拟机。

### 管理数据备份和恢复

系统提供配置数据和业务数据定期在本地和异地备份（将数据备份到第三方服务器）的能力。当管理节点服务异常无法自动修复时，通过本地备份数据立即恢复；当由于灾难性的故障导致管理节点双节点同时故障且不能通过重启等操作进行恢复时，可使用异地备份数据立即恢复（1 个小时之内完成），减少故障恢复时间。

### 黑匣子

内置黑匣子，黑匣子用于收集当前系统的信息，当系统出现故障的时候，黑匣子会保存系统的故障现场信息。借助系统的故障现场信息，可以方便地进行故障定位。

黑匣子保存如下信息：

- 存储内核日志
- 系统快照
- 异常退出前屏幕输出信息
- 诊断工具的诊断信息

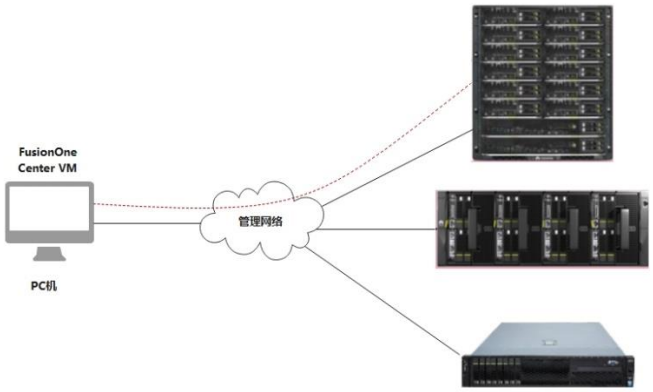
# 6 运维管理



## 6.1 一键安装

FusionOne Installer 快速安装部署工具完成系统软件的安装，大大简化虚拟化的安装效率。

图6-1 部署系统示意图




- FusionOne Installer 安装工具可部署在 PC 机或者虚拟机中。

- FusionOne Installer 通过 Simple Service Discovery Protocol（SSDP）简单服务发现协议或者扫描 IP 方式发现服务器，读取服务器信息。
- FusionOne Installer 连接服务器 BMC，使用 KVM 挂载光盘功能进行引导，启动安装。最大支持 8 个节点并行安装。
- 安装过程中使用 NFS 共享传送安装软件包和配置。

FusionOne Installer 提供安装向导和统一的安装配置界面，可协助用户快速完成参数设置，FusionOne Installer 根据相关的安装配置快速自动完成系统软件的安装。

# 6.2 统一管理

## 6.2.1 集群可靠性管理

### 虚拟机快照策略配置

FusionOne HCI 支持对虚拟机配置快照的策略，可以支持定制快照、周期性快照，并可以对快照的类型配置（普通快照、内存快照、一致性快照)。快照数量可以配置最大保留个数，可以帮助用户自动化运维，无需用户清理快照。

### 虚拟机 QoS 保护组

FusionOne HCI 支持对虚拟机的 IOPS 进行上限设置，防止个别虚拟机占用了存储的 IOPS 资源，影响其他虚拟机的运行。另外提供 Burst IOPS 策略，可以允许虚拟机在短时间突破 IOPS 的 QoS 限制。

用户可以结合上限 IOPS、突发缓冲区设置，灵活的为负载较高的虚拟机进行 QoS 配置。

### 虚拟机 HA 配置

FusionOne HCI 支持虚拟机 HA 故障处理策略配置，如基于优先级的故障恢复策略、蓝屏虚拟机等故障恢复策略等，用户可以根据集群内的业务虚拟机的实际需求定制虚拟机的故障处理策略。

### 虚拟机 DRS、DPM 配置

FusionOne HCI 支持计算集群的动态调度策略配置，可以基于用户配置的调度规则，对集群内的虚拟机调度策略机型配置，如虚拟机互斥、 主机负载均衡策略等。

### 虚拟机安全组

安全组配置可以满足用户不需要额外的防火墙软件，可以对虚拟机的网络访问进行控制，支持出入方向的流量控制、端口配置、基于 IP 子网等策略配置。以上策略配置均支持白名单、黑名单的方式使得规则生效。

### 管理软件数据备份和恢复

FusionOne HCI 支持对平台管理面数据进行备份，在系统异常、配置丢失或发生灾难性故障时，可使用已下载的备份数据进行恢复，快速把管理面业务恢复到备份时间点的状态。

- 本地备份：手动执行一次创建备份，备份数据保存在系统后台，但系统只能保存最近一次本地备份，请在创建备份后及时将其下载并存放在安全的位置以供将来使用。
- 远程备份：通过设置定时备份策略，自动备份到备份策略中的备份路径下，以便系统异常或配置丢失时，可以通过手工恢复已备份的数据将业务恢复到备份前的状态。

## 6.2.2 虚拟机生命周期管理

FusionOne HCI 23.0 版本支持虚拟机相关业务的发放管理特性，包括：虚拟机声明周期管理、磁盘创建管理以及网络 DVS、端口组、VLAN 管理。

### 虚拟机生命周期管理

FusionOne Center 管理平台提供了虚拟机发放管理特性，提供了虚拟机的创建以及常用的日常操作特性，包括：虚拟机上下电、重启关闭，虚拟机迁移，虚拟机导出导入，虚拟机规格调整，性能监控，快照等管理以及虚拟机模板管理等特性。

### 磁盘管理

虚拟机磁盘管理，提供了虚拟机卷设备创建、绑定虚拟机等操作，系统可提供：普通、共享卷设备，支持 IDE、VIRTIO、SCSI 的接口类型，系统默认提供的卷设备为瘦分配卷，可以有效的提升系统的磁盘利用率。

## 网络管理

网络管理主要为提供虚拟机发放中需要的网络资源，主要为分布式交换机、vlan 池、端口组以及 MAC 地址。FusionOne Center 提供了 VLAN 池、端口组、MAC 池的创建配置等功能。

## 虚拟机回收站管理

为了防止误删除虚拟机，还提供了虚拟机回收站的功能，可以快速用户恢复误删除的虚拟机。

## 平台直接修改虚拟机密码

为了方便管理员管理业务虚拟机，FusionOne Compute 提供平台修改虚拟机密码的能力，可以在客户忘记密码等场景下，可以快速重置掉密码。

## 快照管理

FusionOne Center 支持创建快照策略，实现对业务虚拟机的定期的保护，在业务虚拟机出现逻辑性错误时，可以使用快照恢复虚拟机，降低业务损失。

- 快照类型：支持创建普通快照或者内存快照，默认是普通快照。
- 快照策略功能：
  - 快照保留策略：可设置快照的最大保留个数，默认为 7。支持 1-32 的按需配置，超过保留个数的快照将自动删除(用户手动创建的快照不受快照保留策略控制)。
  - 定时创建快照：可设置定时自动创建快照的频率，支持每月、每周、每天的按需配置。
  - 立即执行：创建快照策略后，可支持用户按需立即执行创建快照。立即执行创建的快照，受快照保留策略控制。
  - 关联虚拟机：可为快照策略关联要作用的业务虚拟机。同一个策略支持关联多个业务虚拟机。
- 快照策略管理：支持修改、删除已创建的快照策略。。

### 镜像与模板管理

支持对接 NAS 存储，将客户的操作系统镜像、模板等公共文件存储在客户提供的 NAS 存储。在虚拟机发放、日常运维过程中，可以直接使用对接的 NAS 存储中的镜像文件，可以帮助客户快速部署业务。

## 6.2.3 扩容与减容

FusionOne Comoute 支持节点的平滑扩容、减容。

扩容：

使用 FusionOne Installer 扩容安装方式批量安装待扩容服务器节点，然后在 FusionOne Compute 页面添加主机的方式扩容主机即可。

- 手动添加：每次只能添加一个主机，适用于扩容主机数量较少的场景。
- 批量导入：用户模板填写主机信息，一次性导入，适用于主机数量较多的场景。

减容：

FusionOne Compute 减容操作通过移除主机操作完成。如果主机上存在运行的虚拟机、或者关联了共享存储且该存储仅关联到该主机，则需要先销毁数据存储。减容主机可以使用两个方式：移除（正常减容）、强制移除（故障减容）。

- 移除主机：移除主机之前需要取消关联的资源，如分布式存储、运行的虚拟机、绑定的虚拟机等。
- 强制移除：强制移除可以移除任意状态的主机，但是移除可能导致某些公共关联的资源异常，一般用于故障处理场景。

# 6.3 一键运维

虚拟化管理平台 VRM 可以支持对虚拟化平台的状态监控，包括 CPU、内存、IO、网络均可以进行可视化监控，可以方便运维人员监控平台的负载情况，健康状态。同时提供一键日志收集、一键下电等操作，替代传统日志收集、下电集群的方式。

## 6.3.1 一键日志收集

在 FusionOne Center 管理界面上集成日志收集功能，一键式收集系统故障时各个组件的相关日志，支持一键式收集系统所有日志，也支持针对性收集部件日志，为客户日常运维，问题定位，提供有效帮助。

- 可支持收集日志项包括：FusionOS、VNA、VRM 等系统组件的相关日志项。
- 日志收集一次收集的时间段暂只支持 2 天时间的日志文件，支持并发收集节点日志。

在 FusionOne Compute 日志收集页面，选择待收集的日志时间段，节点类型，日志类型以及需要收集的节点，即可进行收集日志。

日志收集完成后，可将相应的收集日志下载分析。

## 6.3.2 一键下电

在 FusionOne Compute 管理界面上，能一键式对 FusionOne Compute 所有节点执行安全下电操作，满足用户计划性下电的诉求。提供向导式配置引导用户一步步执行一键安全下电的操作。下电前检查站点中存在 vmtools 未正常运行的虚拟机时，提供安全关闭、强制关闭的选项，如果需要安全关闭，则需要用户手动关闭掉这些虚拟机，否则只能选择强制关闭的方式。

- 安全关闭：通过 vmtools 到虚拟机内自动关闭虚拟机。
- 强制关闭：虚拟机强制关闭，可能导致虚拟机移除，存在安全风险。

图6-2 下电操作示例

# 6.4 国产化平台支持

FusionOne Compute 支持用户信创场景使用需求，从安装部署到系统功能，支持 ARM 架构（鲲鹏）超融合集群。支持 x86 共集群，x86 的业务与 arm 的业务共管理平台，数据存储共享，给客户提供业务平滑迁移能力。

- ARM 架构部署：FusionOne Compute 支持部署在 ARM 架构（鲲鹏 CPU）的物理机上，整体功能与 X86 一致。
- X86 与 ARM 混部：管理节点要求同一个 CPU 架构，分布式存储节点在同一个 CPU 架构，例如：ARM 融合节点 3 个+x86 计算节点一个。
- 异构集群管理：支持在一个超融合站点内创建 ARM、X86 两种不同架构的计算集群，实现硬件资源高度集约的同时，保证集群操作的一致性。

# 7 系统安全



## 7.1 系统安全威胁

### 来自外部网络的安全威胁

- 传统的网络 IP 攻击

  如端口扫描、IP 地址欺骗、Land 攻击、IP 选项攻击、IP 路由攻击、IP 分片报文攻击等。

- 操作系统与软件的漏洞

  在计算机软件（包括来自第三方的软件，商业的和免费的软件）中已经发现了不计其数能够削弱安全性的缺陷。黑客利用编程中的细微错误或者上下文依赖关系，已经能够控制操作系统。常见的操作系统与软件的漏洞有：缓冲区溢出、滥用特权操作、下载未经完整性检查的代码等。

- 病毒、木马、蠕虫等。

- SQL 注入攻击

  攻击者把 SQL 命令插入 Web 表单的输入域或者页面请求的查询字符串中，欺骗节点执行恶意的 SQL 命令，在某些表单中，用户输入的内容直接用来构造（或者影响）动态 SQL 命令，或作为存储过程的输入参数，这类表单特别容易受到 SQL 注入攻击。

- 钓鱼攻击

钓鱼攻击是一种企图从电子通讯中，通过伪装成信誉卓著的法人媒体以获取如用户名、密码和信用卡明细等个人敏感信息的犯罪诈骗过程。这些通信都声称来自于著名的社交网站，拍卖网站，网络银行，电子支付网站或网络管理者，以此来诱骗受害者的轻信。钓鱼攻击通常是通过 email 或者即时通讯进行。

- 零日攻击

  “零日漏洞”通常指还没有打补丁的安全漏洞，而“零日攻击”则是指利用这种漏洞进行的攻击。由于安全漏洞出现后，厂商需要时间确认、验证、评估、修补漏洞，很难当日拿出补丁。因此，零日漏洞的利用程序对网络安全具有巨大威胁。

## 来自内部网络的安全威胁

- 攻击方法日新月异，内部安全难以防范

  内网 ARP 欺骗与恶意插件滥用问题等将产生新的安全威胁。被攻破的内网主机，容易被攻击者做为“肉鸡”进行内网的渗透攻击，导致重要数据泄露，或者将其作为 DDOS 工具向外发送大量的攻击包，占用网络带宽。员工滥用恶意插件或浏览被植入病毒或木马的网页，也易受到攻击。

- 补丁升级与病毒库更新不及时、蠕虫病毒利用漏洞传播危害大

  由于网络内主机和设备的操作系统、数据库、应用软件存在安全漏洞，没有及时安装最新的安全补丁，主机杀毒软件病毒库没有及时更新，给恶意的入侵者提供了可乘之机，使病毒和蠕虫的泛滥成为可能。大规模的蠕虫爆发可能导致企业内网全部陷于瘫痪，业务无法正常进行。

- 非法外联难以控制、内部重要机密信息泄露频繁发生

  企业员工通过电话、VPN、GPRS 无线等拨号方式绕过防火墙的监控直接连接外网，使得企业内网 IT 资料暴露在外，易导致重要机密信息泄露。

- 移动设备随意接入、网络边界安全形同虚设

  员工或临时外来人员的笔记本电脑、掌上电脑等移动设备，由于经常接入各种网络环境，很可能携带有病毒或木马等恶意软件，一旦未经审查就接入企业内网，将对内网安全构成巨大的威胁。

- 软硬件设备滥用、资产安全无法保障

  内网资产（CPU、内存、硬盘等）被随意更换与修改，缺乏有效的技术跟踪手段和统一管理，一旦出现攻击行为或者安全事故，责任定位非常困难。

- 应用软件缺乏监控，产生新的安全隐患

随着 QQ、MSN、微博等社交应用的普及，通过这些工具传播病毒、蠕虫、木马已成为新威胁的流行趋势；使用 BitTorrent、电驴等网络工具下载电影、游戏、和软件，可导致关键业务应用系统带宽无法保证。

- 缺乏外设管理手段，数据泄密、病毒传播无法控制

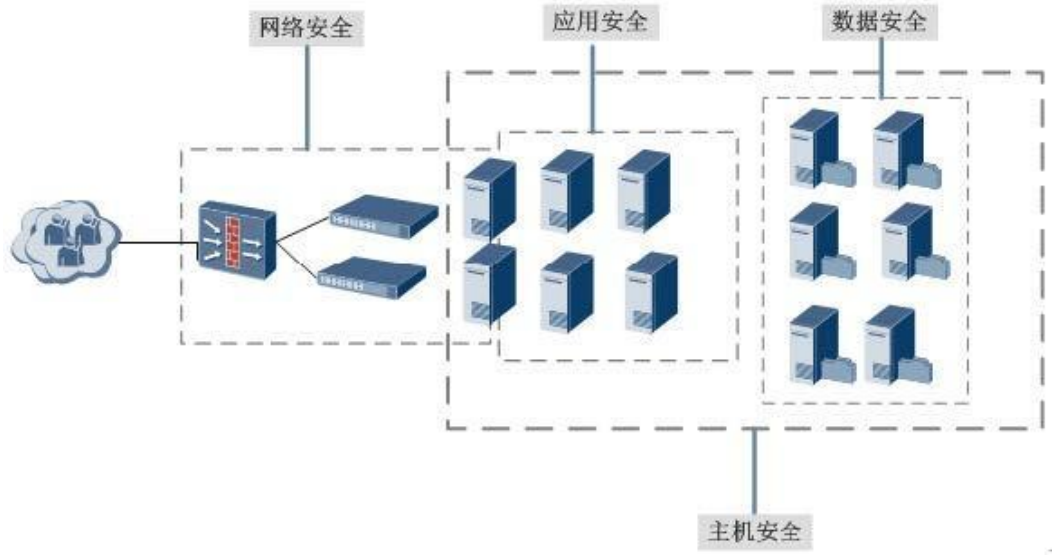
  U 盘、光驱、打印、红外、串口、并口等外设，由于使用方便，已成为数据泄密、病毒感染的出入口。通过封贴端口、制度要求等方式无法灵活对外设进行管理，特别是对 USB 接口的管理，因此，需通过其它技术手段解决存在的问题。

- 管理制度缺乏技术依据，安全策略无法有效落实

# 7.2 总体安全框架

依据系统面临的安全威胁和风险，FusionOne Compute 产品提供安全解决方案，如图 7-1 所示。FusionOne Compute 安全框架通过网络、主机、应用以及数据四个维度上来保证系统的安全性。

图7-1 安全架构图




简要介绍如下：

- 网络安全

  通过网络隔离，保证数据处理、存储安全和维护正常运行。

- 应用安全

从身份认证、权限控制、审计控制等方面介绍 FusionOne Compute 目前已经具备的安全措施。

- 主机安全

  通过对系统内节点的操作系统安全加固等手段保证节点正常运行。

## 7.2.1 网络安全

FusionOne Compute 的网络通信平面划分为业务平面、存储前端/后端平面和管理平面：

- 业务平面

  为用户提供业务通道，为虚拟机虚拟网卡的通信平面，对外提供业务应用。

- 外部存储对接平面

  为虚拟机提供访问存储资源的通信平面，但不直接与虚拟机通信，而通过虚拟化平台转化。

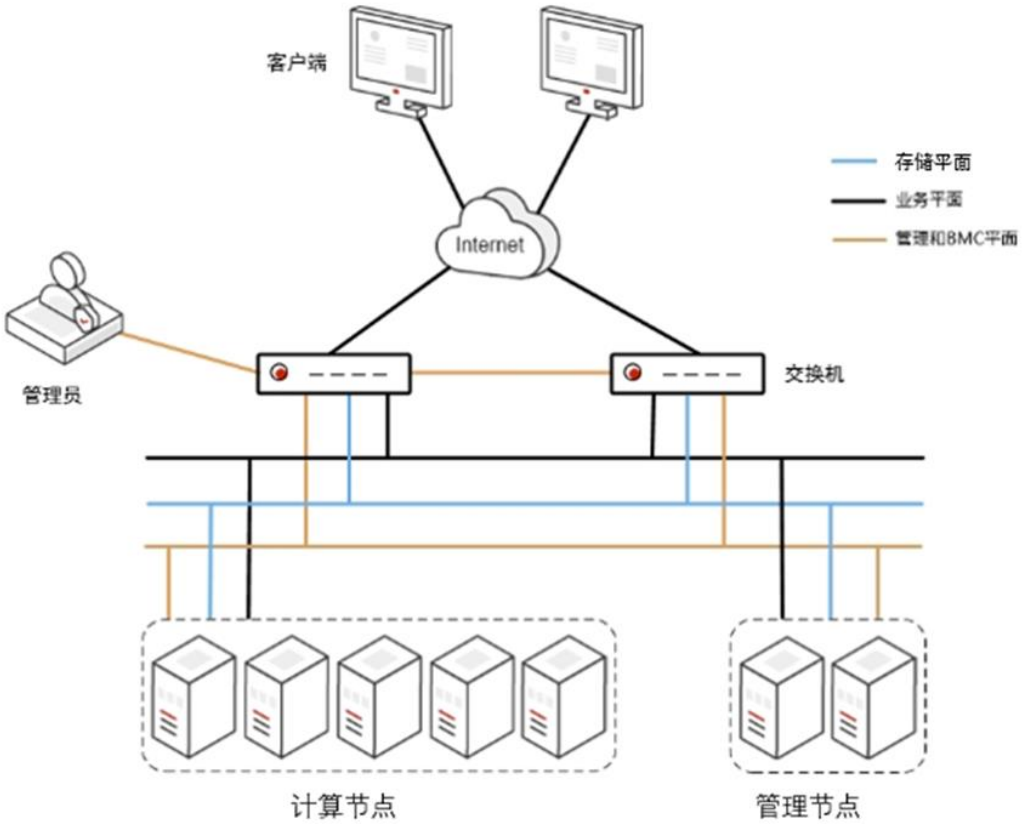
- 管理平面

  负责整个系统的管理、日常维护、业务配置、系统加载等功能的通信。

强烈建议三个平面之间相互隔离，平面隔离原理如图 7-2 所示。

图7-2 平面隔离示意图




## 7.2.2 应用安全

### 7.2.2.1 权限管理

FusionOne Compute 支持“三员分立”特性，通过角色对帐户进行权限控制，系统安装成功后会预置角色，分别是系统管理员、安全管理员（账户激活、安全策略管理）、安全审计员（安全审计）、使得不同用户具有不同的权限，从而保证系统的安全。单个帐户只能拥有系统管理员、安全管理员和安全审计员三者中的一种身份，便于管理员间的权限隔离和相互监督。安装完成之后默认帐户如下：

系统管理员（sysadmin）：仅具有系统业务的操作维护权限，以及创建、删除用户的权限。创建的用户不属于任何角色，处于锁定状态。

安全管理员（secadmin）：仅具有用户、角色的权限管理权，但不能创建用户。系统管理员创建的用户需要安全管理员赋予角色并解除锁定。

安全审计员（secauditor）：仅有日志查看和日志导出权限，用于对其他用户的操作进行审查。

### 7.2.2.2 Web 安全

FusionOne Compute 各 Web 服务具有的安全功能如下：

- 自动将客户请求转换成 HTTPS

  Web 服务平台能够自动把客户的请求转向到 HTTPS 连接。当用户使用 HTTP 访问 Web 服务平台时，Web 服务平台能自动将用户的访问方式转向为 HTTPS，以增强 Web 服务平台访问安全性。

- 防止跨站脚本攻击

  跨站点脚本攻击是指攻击者利用不安全的网站作为平台，对访问本网站的用户进行攻击。

- 防止 SQL 注入式攻击

  SQL 注入式攻击是指，攻击者把 SQL 命令插入到 Web 表单的输入域或页面请求的查询字符串，欺骗服务器执行恶意的 SQL 命令。

- 防止跨站请求伪造

  跨站请求伪造是指欺骗一个已登录的被攻击者装载一个包含恶意请求的页面，该请求利用浏览器自动发送鉴别凭证的功能，继承了被攻击者的身份和特权，执行一个对攻击者有益的恶意操作，如更改被攻击者的口令、地址等个人信息。

- 隐藏敏感信息

  隐藏敏感信息防止攻击者获取此类信息攻击系统。

- 限制上传和下载文件

  限制用户随意上传和下载文件，防止高安全文件泄漏，以及非安全文件被上传。

- 防止 URL 越权

  每类用户都会有特定的权限，越权指用户对系统执行超越自己权限的操作。

- 登录页面支持图片验证码

  在 Web 系统的登录页面，系统随机生成验证码；只有当用户名、密码和随机验证码全部验证通过时，用户才能登录。

### 7.2.2.3 数据库加固

FusionOne Compute 管理节点的数据库类型为 GaussDB 数据库。

数据库必须进行基础的安全的配置，保证数据库运行安全，GaussDB 数据库的主要安全配置如下：

- 访问源控制

  基于访问的实际业务需求与安全标准，只对本地开放访问。所有跨机访问数据库的连接请求都被拒绝，避免受到系统外部的攻击。

- 最小授权原则

  数据库超级管理员之外的其他用户，均按照最小权限的需求设定角色。

- 目录保护

  数据安装目录与其数据区目录属主为安装用户，且其以及其子目录权限控制为读写执行。

- 敏感文件保护

  对于数据库的核心配置文件，属主为安装用户，权限控制为读写。

- 连接数限制

  系统默认的最大连接数是 300，用户可根据实际需要修改配置文件中的最大连接数来防止超大连接数的恶意尝试攻击。

为保证数据安全，必须对数据库进行定期的备份，防止重要数据丢失。数据库支持本地在线备份方式和异地备份方式：

- 本地备份：数据库定时执行备份脚本进行备份。
- 异地备份：数据异地备份到第三方备份服务器。

#### 7.2.2.4 日志管理

日志查看时采取的安全措施如下：

- 任何人员不能在界面上修改或删除日志。
- 有查询权限的人才能导出日志。

### 7.2.3 主机安全

FusionOne Compute 中计算主机、管理节点均使用 Linux 操作系统，为保证此类设备的安全，必须对 Linux 操作系统进行基础的安全配置，基础安全配置的主要内容如下：

- 关闭不必要的服务，如屏蔽 Telnet 服务和 FTP 服务。
- 加固 SSH 的服务。

- 控制文件和目录的访问权限。
- 限制系统访问权限。
- 管理用户密码。
- 记录操作日志。
- 检测系统异常。
- 防火墙启用，只开启与业务相关的通信端口。

# 8 产品规格

## 管理容量

表8-1 管理容量

| 指标名称 | 指标值 |
| --- | --- |
| 单集群支持的最大主机数量 | 256 |
| 单集群支持的计算集群数量 | 64 |
| 单集群支持的最大虚拟机数量 | 5000 |
| 单计算集群支持的主机数量 | x86: 64<br>arm: 32 |
| VIMS 集群支持的最大主机数 | 32 |

## 主机规格

表8-2 主机规格

| 指标名称 | 指标值 |
| --- | --- |
| 每物理主机支持的最大逻辑 CPU | x86: 768<br>arm: 128 |
| 每物理主机支持的最大物理内存 | x86: 16T |

| 指标名称 | 指标值 |
| --- | --- |
| | arm: 4T |
| 每物理主机支持的最大虚拟机数量 | 1024 |
| 每物理主机最大挂载 LUN 数量 | 1024 |
| 每物理主机最大 vCPU 数 | x86: 4096<br>arm: 384 |
| 每物理主机最大虚拟网卡数 | 2048 |
| 每物理主机支持的最大虚拟磁盘数 | 2048 |
| 每物理主机的 NUMA 节点数 | x86: 16<br>arm: 4 |
| 每物理主机支持的在线迁移虚拟机并发数量 | 8 |

## 网络容量

表8-3 网络容量

| 指标名称 | 指标值 |
| --- | --- |
| 系统支持的最大分布式交换机数量 | 50 |
| 单个分布式交换机管理的最大主机数量 | 256 |
| 单个分布式交换机支持的最大虚拟交换端口数量 | 10000 |

## 虚拟机规格

表8-4 虚拟机规格

| 指标名称 | 指标值 |
| --- | --- |

| 指标名称 | 指标值 |
| --- | --- |
| 单个虚拟机支持的 vCPU 数量 | x86: 255<br>arm: 128 |
| 单个虚拟机支持的网卡数量 | 16 |
| 单个虚拟机支持的磁盘数量 | 60 |
| 单个虚拟机支持的内存容量 | x86: 6TB<br>arm: 256GB |
| 虚拟机支持的单磁盘规格 | 本地磁盘：16TB<br>VIMS 共享存储：64TB<br>分布式块存储：32TB |
| 虚拟机支持的 GPU 数量 | 1 |
| 单个虚拟机支持的快照数量 | 32 |