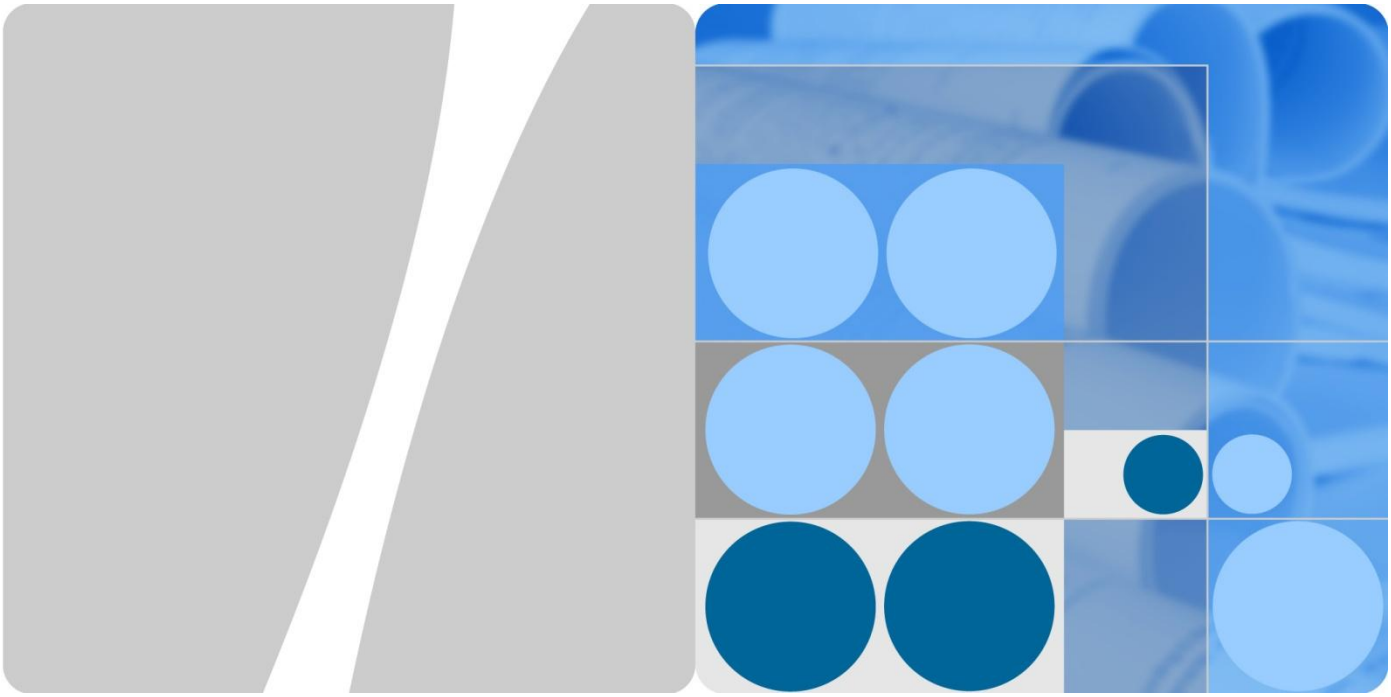


华为 FusionCloud 桌面云解决方案
5.3 系统高可用性白皮书

文档版本 01

发布日期 2015-06-23

华为技术有限公司

HUAWEI

商标声明


和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为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，华为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。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： http://www.huawei.com

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

目录

1 华为桌面云解决方案

桌面云解决方案的架构组件部署在云计算提供的虚拟机中，对外提供桌面服务，结构图如下图所示。

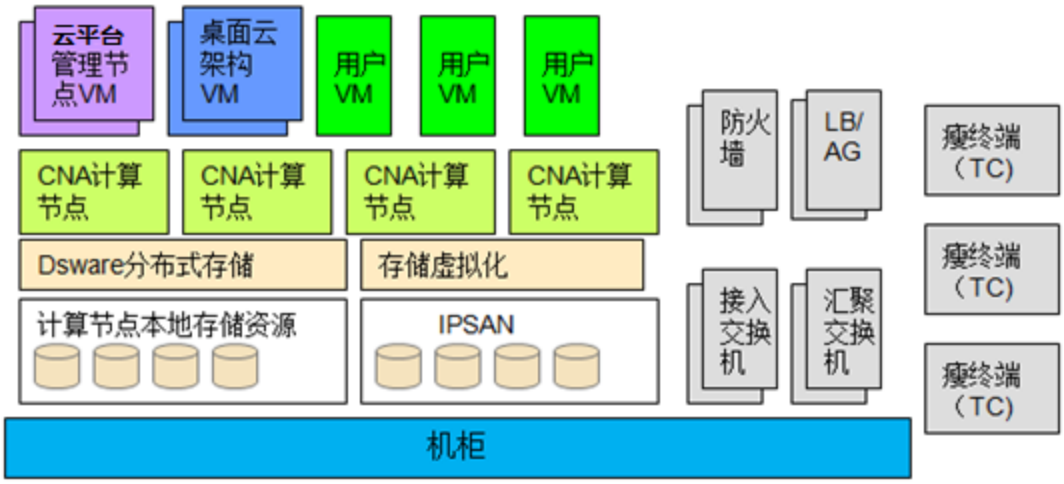


图1　桌面云平台系统组成图

2 系统可用性指标

- 全局虚拟机年度平均可用度99.9%

指标介绍：虚拟主机在任何随机时刻能够满意工作的概率。可用度是由产品可用性和维修性共同决定的：

$$A = \frac{MTBF}{MTBF + MTTR}$$

其中：

A——产品的可用度

MTBF——产品的平均故障间隔时间

MTTR——产品的平均修复时间

指标达成：参考本文第 2、3 章 。

- 虚拟机业务运行时间（duty time）： 24*7

  指标说明：可提供一天 24 小时，一周七天的不间断虚拟机服务。

- 电源恢复时间< 2 hour:

  指标说明：云平台系统从电源恢复到业务完全恢复的时间。

  指标达成：云平台系统软件可以不分先后进行加载，包括管理软件和计算服务器软件；单个服务器加载时间不超过 5 分钟，可并行进行 20 个服务器的加载。

- 虚拟机故障迁移的时间4分钟：

  指标说明：由于各类故障导致虚拟机掉电或死机，云计算将故障虚拟机重新启动完成的时间；系统检查到虚拟机故障，开始在其他正常服务器上启动虚拟机的时间。 最终 VM 启动完成依赖于客户机 OS 启动时间。

  指标达成：系统管理服务器通过心跳进行虚拟机的故障检测，在 40 秒内没有响应，就在其他服务器上进行此虚拟机的启动，此过程称为故障迁移(HA)。虚拟机本身启动的时间不包括在内。系统通过锁的方式，可防止虚拟机的脑裂行为。

- 虚拟机热迁移时间和虚拟机内存相关，每1G内存需要20秒：

  指标说明：计划内的定时或手工热迁移，虚拟机从一台服务器上平滑无损的迁移到另外一台物理机上运行起来的时间。

  指标达成：热迁移时，系统内的虚拟化软件进行虚拟机的内存拷贝到目的物理服务器，速度大致为每 20 秒 1G，拷贝完成后，再将拷贝这段时间内的内存变化数据同步到目的物理机，然后如此循环，最终进行新虚拟机的启动，关闭旧虚拟机，切换时间为毫秒级业务无感知。

- 瘦客户端(TC)的年失效率小于3%。

3 系统软硬件可靠性

3.1 机柜

云计算机房采用的机柜具有以下可靠性规格：

- 双 PDU，支持设备双路供电，所有输出具有过流保护功能；

- 最高规格的抗震设计：抗 9 级烈度地震。
- 满足北美 NEBS-L3 的要求

3.2 服务器

3.2.1 内存可靠性

内存错误主要包括硬件错误和软件错误，其中硬件错误是由硬件失效或者损坏造成的，器件会不断返回不正确的数据，硬件错误可以通过 E6000、E9000 和 RH 2288H 启动时 BIOS 的内存自检发现。

内存使用中经常碰到的为软件错误，软件错误不能通过内存自检发现，只有通过一些内存检错和纠错的算法来保护内存中的数据。服务器在内存软件错误纠正上采用内存 ECC（Error Checking and Correction）技术，采用工业标准的纠错算法，能够检测内存 2bit 错误，并修复内存单 bit 错误。

3.2.2 硬盘可靠性

硬盘热插拔：服务器支持系统运行时的硬盘（SATA/SAS）热插拔；

硬盘 RAID：服务器支持 RAID0、1、5 等多种 RAID 方式，支持 RAID 下另加热备盘的配置，保证了硬盘数据的高可靠性，在 RAID 组的某颗硬盘坏掉后，支持数据恢复、RAID 组恢复和在线更换硬盘。其中 RAID 卡支持电池，可以对 Cache 数据进行保护，既可以提高对硬盘的访问性能，又可以防止意外掉电时数据的丢失。

3.2.3 支持磁盘在线定时故障检测和预警

桌面云解决方案虚拟块存储模块采用了业界先进的 SMART 技术标准来实现对基于 ATA 和 SCSI 接口的硬盘进行监控和可靠性管理，检查其可靠性并预测磁盘错误。他的技术原理是主要通过侦测硬盘各属性，如数据吞吐性能、马达起动时间、寻道错误率等属性值和标准值进行比较分析，推断硬盘的故障情况并给出提示信息，帮助用户避免数据损失。

SMART 是 Self-Monitoring Analysis and Reporting 系统的英文系统缩写，中文就是自监测、分析和报告技术。这个技术是现在普遍应用于硬盘的数据可靠性技术， 一般情况下 SMART 的几个主要的关键检测属性包含如下：

- Read Error Rate 错误读取率

- Start/Stop Count 启动/停止次数(又称加电次数)
- Relocated Sector Count 重新分配扇区数
- Spin up Retry Count 旋转重试次数(即硬盘启动重试次数)
- Drive Calibration Retry Count 磁盘校准重试次数
- ULTRA DMA CRC Error Rate (ULTRA DMA 奇偶校验错误率)
- Multi-zone Error Rate 多区域错误率

3.2.4 电源可靠性

服务器（例如 RH2288H、E6000 和 E9000 等）配置多组冗余电源（PSU），提供电源故障告警，支持电源冗余和热插拔，可以在 1 组电源故障后，系统持续运行而不影响业务；并且可以在线更换故障电源。

3.2.5 系统监控

系统支持对 CPU，内存等热关键器件的温度实时监控，配合智能的风扇调速和监控，确保系统运行的可靠性。

系统支持对风扇，电源，硬盘等关键器件的运行状态监控，设备故障时会产生告警，可以灵活对支持热插拔设备进行在线更换，不支持热插拔设备提前安排好业务后进行下电更换。

3.2.6 板载软件可靠性

BMC 软件支持双 Image，当 Flash 中的某个 Image 遭到破坏时，支持从另一个 Image 启动 BMC 系统，而不会造成系统无法启动的情况。

BMC 软件支持进程监控，某个进程死掉后，支持重启恢复功能。

3.3 存储设备

FusionStorage 分布式存储系统用于一体机解决形态下。利用计算节点上的本地存储资源存储用户数据，采用冗余备份的强一致性技术及分布式 cache 技术，在保证数据一致性的前提下，提供高存储性能的解决方案。FusionStorage 分布式存储在桌面云场景下，采用三副本技术，可用性高达 99.9993%。

IPSAN 产品（华为的 S5500T 系统）每套配置 1 个主控框，可扩展三个级联框，共 96 块硬盘。可靠性措施如下：

- 配置 8 条物理链路做存储多路径（multi-path）
- 每套配置两个全局的热备硬盘
- 每套配置 7 个 9+1 的 raid5 硬盘组

可靠性指标如下表：

表1　IPSAN 的可靠性指标

场景	可用度	MTBF(年)	年中断时间(分钟/年)
10 块硬盘配置 RAID5	99.9991%	12.684	4.73

IPSAN 的整体可用度为 99.9991%，MTBF=12.684 年（111110.1h），年中断时间为 4.73min。

FusionStorage 系统及 IPSAN 设备提供了掉电保护、后台扫描、数据预重建等核心数据保护机制，保障用户数据不丢失，用户数据不丢失率（数据持久度）高达 99.99%以上，而普通 PC 机的数据持久度不到 95%。

3.4 网络设备

网络子系统主要采取以下五个措施来增强系统的可靠性：

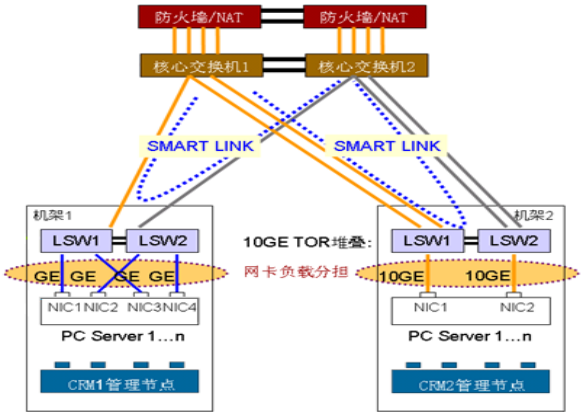


图2 网络子系统部署示意图

3.4.1 网卡负荷分担

如图 2 所示，对于物理服务器提供的多块网卡，出于可靠性以及流量负载均衡的考虑，系统采用了 Bonding 模式。使用绑定模式之后，网卡被绑定成逻辑上的“一块网卡”后，同步一起工作，对服务器的访问流量被均衡分担到多块网卡上，这样每块网卡的负载压力就小多了，抗并发访问的能力提高，保证了服务器访问的稳定和畅快，而且当其中一块发生故障的时候，另外的网

卡立刻接管全部负载，过程是无缝的，服务不会中断。避免单个网卡或者链路故障引发的业务中断。

服务器绑定多网卡的实际意义在于当系统绑定多网卡之后，不仅可以扩大服务器网络进出口带宽，而且可以实现有效负载均衡和提高容错能力，避免服务器出现传输瓶颈或者因某块网卡故障而停止服务。

3.4.2 交换机堆叠

堆叠是将同一物理位置上的交换机通过堆叠电缆或高速上行口组成一个高可靠的设备组，接入交换机设备是通过堆叠口实现堆叠的。通过堆叠，在提高可靠性的同时，可以实现对交换机的集中管理和维护，降低用户的维护成本。

通过堆叠技术，将两台物理交换机作为一台交换机进行处理，交换机之间无需配置 TRUNK，对于接入设备服务器而言，相当于只看到一台物理设备。处于堆叠组中的两台物理交换机处于主备状态，单台设备故障，由另外一台设备接管。

交换机通过堆叠线缆连接成环型或链型，运行堆叠管理协议，选举出主交换机，负责堆叠系统的管理，包括分配堆叠成员的 ID、收集堆叠的拓扑信息，并将拓扑信息通告给所有的堆叠成员；主交换机指定备用交换机，备交换机在主交换机出现故障的时候升级为主交换机来管理整个堆叠。

3.4.3 交换机互连冗余

Smart Link，中文译为灵活链路，又称为备份链路，是一种为链路双上行提供可靠高效的备份和切换机制的解决方案，常用于双上行组网。相比 STP（Spanning Tree Protocol，生成树协议），Smart Link 技术能够提供更高的收敛性能，相比 RRPP（Rapid RingProtection Protocol）和 SEP(Smart Ethernet Protection)，Smart Link 技术提供了更简洁的配置使用方式。

双上行组网是目前常用应用组网之一，该组网下通过生成树协议阻塞冗余链路，起备份作用。当主用链路故障时，将流量切换到备用链路。虽然这种方案从功能上可以实现客户冗余备份的需求，但是在性能上却不能达到很多用户的要求，因为即使采用快速生成树协议的快速迁移，也只能是秒级的收敛速度。这对于应用于电信级网络核心的高端以太网交换机，是非常不利的一个性能参数。

基于上述原因，桌面云解决方案引入了 Smart Link 解决方案，针对双上行组网，实现主备链路冗余备份及快速迁移。该方案为双上行组网量身定做，即保证了性能，又简化了配置，同时，作为对 Smart Link 的一个补充，还引入了端口联动的方案，也即是 Monitor Link，用于监控上行链路，使 Smart Link 备份作用更为完善。

3.4.4 虚拟路由冗余保护

VRRP（Virtual Router Redundancy Protocol）虚拟路由冗余协议，是一种容错协议。该协议通过把几台路由设备联合组成一台虚拟的路由设备，使用一定的机制保证当主机的下一跳交换机出现故障时，及时将业务切换到其它交换机，从而保持通讯的连续性和可靠性。

VRRP 将局域网的一组路由设备构成一个 VRRP 备份组，相当于一台虚拟路由器。局域网内的主机只需要知道这个虚拟路由器的 IP 地址，并不需知道具体某台设备的 IP 地址，将网络内主机的缺省网关设置为该虚拟路由器的 IP 地址，主机就可以利用该虚拟网关与外部网络进行通信。

VRRP 将该虚拟路由器动态关联到承担传输业务的物理设备上，当该设备出现故障时，再次选择新设备来接替业务传输工作，整个过程对用户完全透明，实现了内部网络和外部网络不间断通信。

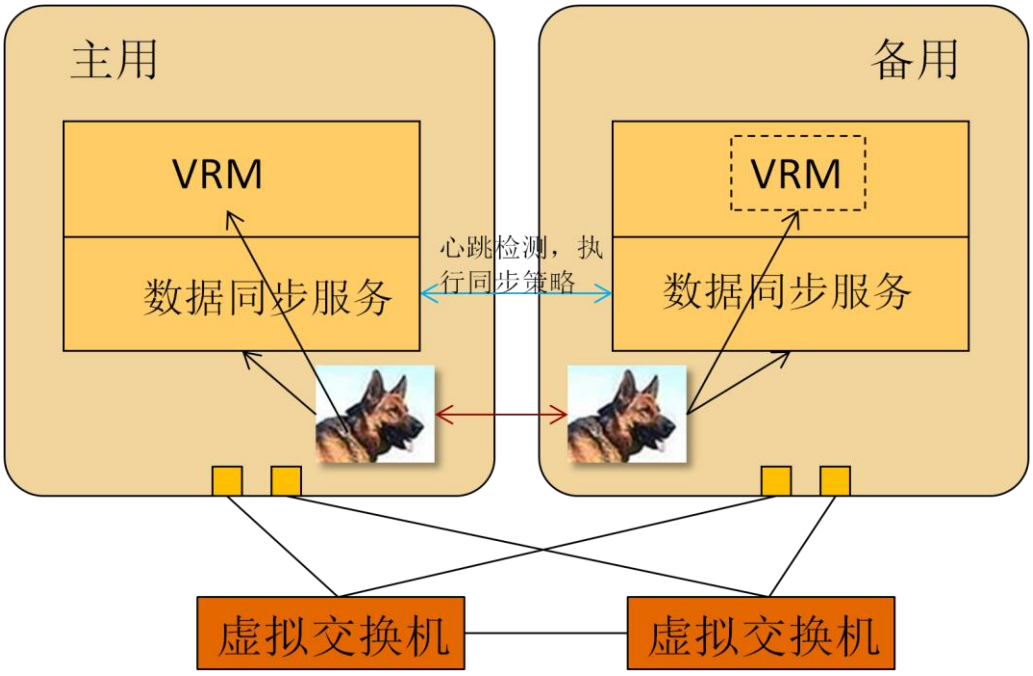
3.4.5 网络分平面通信

整个云计算系统逻辑上可以分为三个平面：管理平面、存储平面和业务平面。为了保证各种网络平面数据的可靠性，桌面云解决方案采用分网络平面的架构方案，不同平面间采用 VLAN 进行隔离，单个平面的故障不影响其余的两个平面继续工作，例如当管理平面暂时故障时，业务平面还能够继续为云终端用户提供服务。此外，系统还支持基于 VLAN 的优先级设定，使得内部的管理/控制报文具备最高的权限，从而使得在任何时候，管理员和用户均可以管控系统。

3.5 云平台软件

3.5.1 管理节点 HA

虚拟化平台的业务主备管理节点采用管理平面的心跳检测，备用节点实时检测主用节点的健康状态，一旦发现主用管理节点故障，备用管理节点将立刻接管主用节点业务，持续对外提供服务。针对管理节点上的应用进程，通过采用软件狗的方式对运行在管理节点上的进程进行实时监控，如发现进程吊死或进入死循环，软件狗将会检测到相关进程的异常状态，并触发相关进程的重启恢复；如果发现进程重启后仍不能恢复正常，则进行业务管理节点的主备倒换并出主备心跳异常告警以保证应用进程的可靠性。



管理节点负责对全系统的业务进行管理，采用主备高可靠性的工作方式，如果主备管理节点同时故障，相关的新增业务会受影响，例如虚拟机的创建和删除等，但对于已经存在并运行中的虚拟机无影响，用户继续使用虚拟机上的应用程序，不会有任何感知。

3.5.2 管理节点数据备份

管理节点所有数据，包括配置文件、数据库记录等，均会定期自动备份，即使管理服务器主备都故障且数据丢失，也可以快速恢复。

备份方式：

- 每周全量备份；
- 每天增量备份；
- 高危操作前的即时备份

主备同时故障且数据全部丢失的恢复过程：

- 更换管理服务器；
- 重新加载管理节点；
- 拷入备份的数据，启动管理节点，即可恢复。恢复过程在 30 分钟内完成。

3.5.3 虚拟机备份

HyperDP 虚拟机备份方案，是使用华为 HyperDP 设备，配合 FusionCompute 的快照备份功能实现的虚拟机数据备份方案。HyperDP 通过与 FusionCompute 配合，实现指定虚拟机或虚拟机内指定卷对象按指定策略的备份。当虚拟机数据丢失或故障时，可通过备份的数据进行恢复。数据备份的目的端为 HyperDP 虚拟机挂载的虚拟磁盘或外接的 NFS/CIFS 共享文件系统存储设备。

HyperDP 虚拟机备份方案具有以下特点：

1、简单易用，无需安装备份代理，备份服务器通过虚拟机模板安装，通过浏览器即可进行管理。

2、灵活备份策略设置，支持设置周期性全量与增量备份策略，可灵活设置备份周期与备份时间窗口，支持设置备份数据过期策略以自动清理过期备份数据，并可针对不同类型 VM 设置不同备份策略。

3、高效备份与恢复，全量备份时只备份有效数据，增量备份时只备份非重复更改数据，减少无效数据备份，最大限度减少备份通信流量与备份存储空间需求。

4、并发备份与恢复，每个备份设备支持 200 个 VM，可最多并发备份与恢复 8 个 VM，每个备份域支持 10 个备份设备；由于 HyperDP 位于专用的虚拟设备上，备份处理对生产 VM 基本没有影响。

3.5.4 虚拟机 HA

当计算节点物理服务器宕机或者重启，系统可以将具有 HA 属性的虚拟机故障迁移到其他计算服务器，保证虚拟机能够快速恢复。

云计算解决方案提供多种迁移策略，当计算服务器宕机后，由于单个集群内可以运行上千个虚拟机（单个 CNA 节点可以运行 40 个 VM），为避免大量虚拟机迁移造成网络拥塞和目的服务器过载，系统会根据网络流量、目的服务器负荷选择将虚拟机迁移到不同的目的服务器。

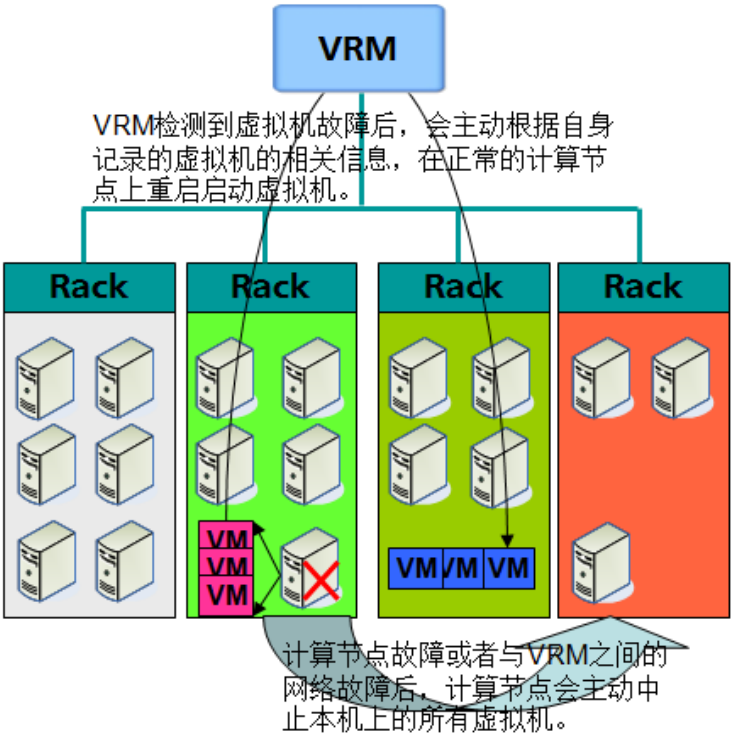


图3 虚拟机 HA 特性示意图

在不同虚拟机故障的场景，当虚拟机与 VRM 无法进行连接，则系统认为虚拟机故障，系统选择新的计算节点重新启动故障虚拟机。

虚拟机 HA 过程：当虚拟机与 VRM 无法进行连接时， VRM 会认为该 VM 已经“故障”，于是下命令在其他的计算节点上重新启动该虚拟机，达到自动快速恢复故障的目的。同时为了防止错误决策导致的虚拟机脑裂，系统中引入了防脑裂锁机制，杜绝了这种现象的发生。。

云平台的管理节点部署在虚拟机上，因此也具备虚拟机 HA 功能。主用管理节点虚拟机和备用管理节点虚拟机部署在不同的计算节点中，如果主用的管理节点虚拟机所在计算节点故障，备用的管理节点虚拟机会升为主用，同时，原主用管理节点虚拟机会在除了备用管理节点虚拟机所在计算节点之外的其他计算节点重启，重启后作为备用节点使用，这个功能叫做主备互斥功能，已保证主备管理节点始终不在同一个计算节点中的，因此任何一个计算节点故障都不会导致主备管理节点同时故障。

3.5.5 虚拟机故障检测和处理

桌面应用中，很多虚拟机采用了 windows 操作系统，windows 操作系统往往运行不稳定，会出现蓝屏等故障现象。云平台能够捕获到虚拟机蓝屏信息，

在虚拟机出现蓝屏时自动对虚拟机进行自动重启来恢复，不需人为干预和处理，用户只需在虚拟机启动后重新连接虚拟机即可。

如果用户虚拟机的操作系统崩溃后，往往只能靠重装系统来恢复，用户会担心自己的数据卷中的数据是否丢失，华为云平台提供了一键式“故障虚拟机磁盘迁移功能”，快速创建和用户虚拟机相同规矩的系统卷，并将用户原故障虚拟机的数据卷自动挂载，用户不需另外做其他的操作，就可以直接登录新虚拟机，找回数据卷中的数据。

3.5.6 虚拟机热迁移

虚拟机是云平台提供弹性计算服务的资源实体，为保证虚拟机的可用性，规避业务中断的风险，系统提供虚拟机热迁移能力，即虚拟机在不中断业务的情况下实现迁移。在迁移过程中，为保证内存的同步，虚拟机管理器（Hypervisor）提供了内存数据的快速复制技术，从而保证了在不中断业务的情况下将虚拟机迁移到目标主机（如图 4）。同时，通过共享存储保证了虚拟机迁移前后持久化数据不变。

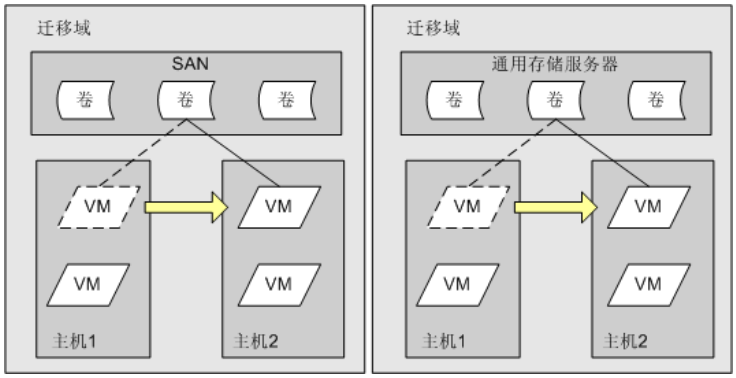


图4 虚拟机热迁移特性示意图

降低客户的业务运行成本：根据时间段的不同，客户的服务器会在一定时间内处于相对空闲状态，此时若将多台物理机上的业务迁移到少量或者一台物理机上运行，而将没有运行业务的物理机关闭，就可以降低客户的业务运行成本，同时达到了节能减排的作用。

保证客户系统的高可靠性：如果某台物理机运行状态出现异常，在进一步恶化之前将该物理机上运行的业务迁移到正常运行的物理机上，就可以为客户提供高可用性的系统。

硬件在线升级：当客户需要对物理机硬件进行升级时，可先将该物理机上的所有虚拟机迁移出去，之后对物理机进行升级，升级完成再将所有虚拟机迁移回来，从而实现在不中断业务运行的情况下对硬件进行升级，保证服务的持续可用性。

目前系统支持的虚拟机热迁移应用场景：

- 根据需要按照迁移目的手动把虚拟机迁移到空闲的物理服务器
- 根据资源利用情况将虚拟机批量迁移到空闲的物理服务器

3.5.7 存储迁移

云平台的存储虚拟化模块提供了存储迁移的能力。在对存储设备进行维护的场景中，或用户使用存储的能力要求提供，旧的存储设备能力以及不能满足要求等场景中，需要将用户的数据从一个存储设备迁移到另外一个存储设备中，云平台提供了存储热迁移的能力，可以保证 VM 无感知的情况下，把用户 VM 的存储数据从一个存储设备迁移到另外一个存储设备上。

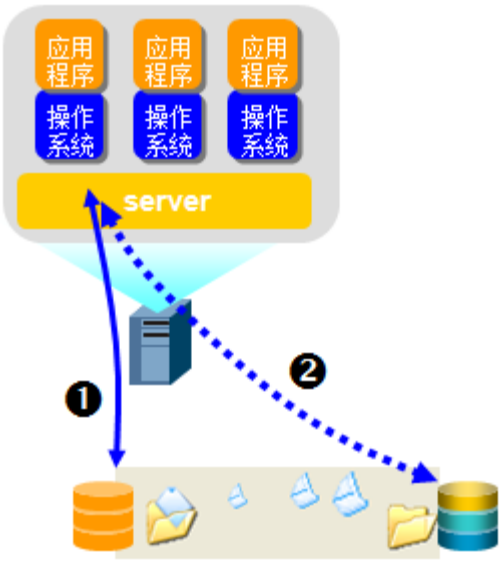


图5 存储迁移特性示意图

3.5.8 虚拟机负载均衡

系统节点在新开一个虚拟机、虚拟机热迁移或者计算节点故障异地重启恢复的时候，在系统配置成负载均衡模式的情况下，根据各个物理计算服务器节点现有的负载状况进行动态的调配，使得集群中各个物理计算服务器的负载处于一个动态的均衡状态。

3.5.9 黑匣子

管理节点和计算节点引入电信领域“黑匣子”技术，在系统出现异常或宕机时自动存储 VMM 内核日志、系统快照、内核诊断信息及异常退出之前屏幕输出信息，并保存至非易失性存储设备（计算节点）或通过 netpoll 方式实时发送至远端服务器（TFTP 服务器），以便系统死机后，导出分析定位。

3.5.10 数据一致性保证

整个云系统的实现，采用了电信领域的高可靠性要求，系统开发的代码中 80%以上用于考虑对各种各样的异常情况进行处理，以及为了保证数据一致性而进行的 checkpoint 机制及回滚机制，从而保证了系统高质量稳定运行。同时内部集成了数据审计机制，对由于各种各样的故障原因导致的潜在数据垃圾进行审计清理，并进行垃圾回收，保障各节点不会因为垃圾数据导致数据不一致影响业务。

3.5.11 健康检查工具及故障信息收集工具

系统还提供了健康检查工具及故障信息收集工具，同样源于电信领域的高可靠性理念。

健康检查工具可以例行对系统进行体检，观察系统运行情况、告警及日志信息、进程运行情况、关键配置信息是否被破坏、系统关键资源的占用情况及变化趋势等等诸多信息来发现系统潜在的性能、安全和可靠性风险，并一一加以弥补，同时，该工具还可以用于高危操作及升级前后的系统检查，以验证系统的健康状况。

故障信息收集工具可以根据典型的故障种类，在系统发生故障时，方便并准确地收集和该故障相关的日志及告警信息，以便快速提供给技术支持人员进行故障定位分析，简化故障信息收集动作，缩短了这部分时间，从而缩短了因故障导致业务停机的整体时间。

3.6 FusionAccess 桌面接入系统可用性

3.6.1 FusionAccess 服务的高可用性

下表为 FusionAccess 服务软件部署的方式，除了 License 服务，其它所有的 FusionAccess 服务采用冗余部署。由于 HDC 能够缓存 Licence，即使 License 服务宕机，在一个月内都不影响业务。任何服务出现故障，系统会及时发现并且进行故障隔离。综上所述，任何单点故障，不影响业务。除了通过软件

的冗余部署提高 FusionAccess 系统的可用性，所有的 FusionAccess 服务都有本地业务监控，当服务不能提供业务时，服务会被自动重启继续提供业务。

特别的，在需要 AD 提供用户鉴权的场景，即使 AD 系统出现异常，不能正常的进行鉴权，这时，系统仍然可以利用虚拟机的本地鉴权，不妨碍系统使用桌面云服务。

服务名	功能	部署方式	单点故障对业务影响
WI	提供用户的登陆	负载均衡	无
HDC	桌面接入控制	负载均衡	无
License	License 控制	单点部署	在设计上，License 宕机后，一个月之内不影响业务
ITA	业务发放	主备	无
vLB	WI 的负载均衡器	主备	无
vAG	自助维护登陆网关和桌面接入网关	负载均衡	无
GaussDB	保存桌面业务数据	主备	无
AD/DNS/DHCP	IT 基础设施	主备	无

3.6.2 FusionAccess 服务的监控

FusionAccess 的 VDI 基础架构服务器进行实时监控，对于服务(器)的异常，将会在 ITA 的 Portal 提供统一的告警呈现给用户。每种告警的处理都有对应的指导书。不同的服务(器)采用不同方式进行。一般来说，Linux 服务器由服务主动向 ITA 上告心跳，心跳信息中包括 CPU 和内存占用率。ITA 如果连续三个周期没有收到服务器的心跳，则会产生服务异常的告警。如果心跳信息中的 CPU 和内存占用超过 80%，则也会产生告警。部署在 windows 的服务器，则通过检查服务状态方式进行监控。下表为 FusionAccess 的主要监控告警。

服务名	功能	监控方式	备注

WI	提供用户的登陆	心跳	
HDC	桌面接入控制	心跳	
License	License 控制	心跳	
ITA	业务发放	服务状态检查	两台 ITA 相互检查
vLB	WI 的负载均衡器	心跳	
vAG	自助维护登陆网关和桌面接入网关	心跳	
GaussDB	保存桌面业务数据	HDC 和 ITA 业务触发	
AD/DNS/DHCP	IT 基础设施	ITA 主动进行进程检查	
LogGetter	用于管理配置数据备份	通过检查备份结果进行监控	
磁盘	对于 VDI 基础架构的所有磁盘进行监控	服务器周期上告磁盘状态	确保磁盘没有被异常写满，Linux 系统中，还需要检查 inode 个数没有超过磁盘分区的 80%
时钟同步	全系统时钟同步	服务器周期上告时钟同步状态	

3.6.3 桌面接入的高可用性

主要通过下面三种方式提高桌面接入的高可用性。

✧ 桌面接入自动重连

由于网络闪断或者其它原因导致桌面接入断连时，客户端会自动发起重新连接，不需要用户重新登录。

✧ 桌面服务端口自动切换

如果桌面服务的端口是固定的，用户安装的应用程序可能会占用用户虚拟机上桌面服务的端口，从而导致客户端无法连接。为了避免这类软件兼容性问题，华为的桌面服务采用了端口自动切换技术，当某个端口被占用时，能够自动切换到一个未占用端口，保证客户端能够正常连接。

✧ 桌面服务程序的高可用性

运行在用户虚拟机上的桌面服务程序，异常终止（无论是用户错误终止进程或者被其它程序终止）时，能够自动恢复进程，保证桌面服务的高可用性。

3.6.4 FusionAccess 管理数据备份

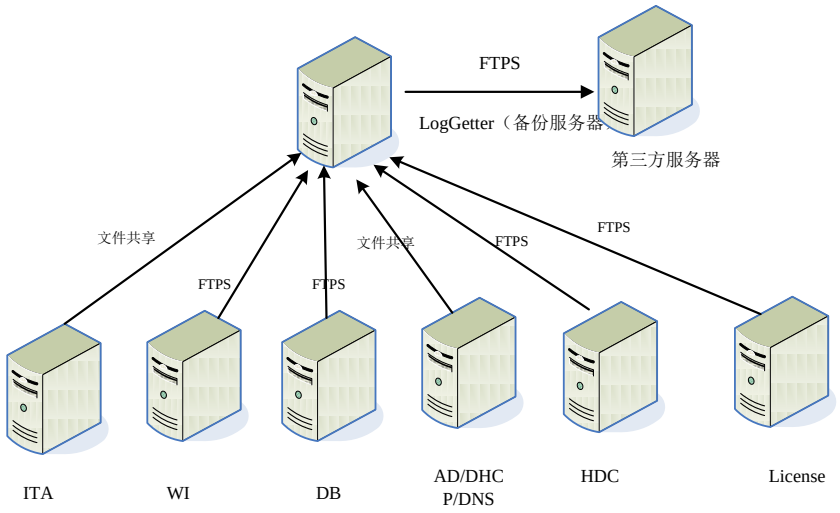


图6 管理数据备份示意图

上图是管理数据备份的示意图，凌晨一点，各个节点进行数据备份，备份完成后，打包成压缩文件。WI/HDC/License/DB 的备份文件通过 FTPS 把备份文件传输给 LogGetter 服务器，ITA/AD/DHCP/DNS 服务器则通过文件共享方式，共享给 LogGetter。如果配置了第三方的 FTP(s)备份服务器，当所有备份完成后，LogGetter 会把备份文件传输给第三方的 FTP(s)备份服务器。当没有配置三方的 FTP(s)备份服务器时，备份数据保存在 LogGetter 服务器上。

无论备份数据保存在 LogGetter 服务器还是客户提供的第三方 FTP(s)备份服务器，都能保存最近 10 次备份数据。当数据损坏时，能够根据 GPI 资料迅速恢复。

3.6.5 上电恢复可靠性设计

当数据中心意外停电并且恢复供电时，系统会自动拉起所有服务器。由于服务器节点启动没有先后顺序的依赖关系，所以在意外断电然后恢复供电的场景中，系统仍然会正常工作。

4 虚拟机桌面业务可用性

用户使用 VM 业务的 RBD 图如图 7 所示：

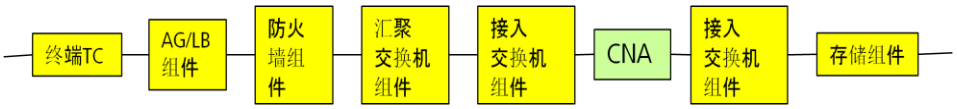


图7 用户使用 VM 的业务 RBD 图

所有桌面云组件内部均采用主备方式或集群方式，无单点。WI、HDC、ITA、AD/DNS/DHCP、DB 的数据都进行了备份，当出现双点故障时，能够在两个小时内恢复业务。根据上图的串并联关系，计算获得用户使用 VM 的业务可用性指标，详请请参考下表。注意该数据用于和客户交流，不能作为 SLA 写入合同中。

表2 虚拟桌面业务可用性指标

	MTBF（y）	MTBF（h）	可用度	年中断时间（min）
虚拟桌面 VM 的业务可用度	6.1859354	54188.7938	0.99998154	9.702576

5 术语表

AD	Active Directory	活动目录
ATA	Advanced Technology Attachment	高级技术附加装置，一种磁盘接口
BIOS	Basic Input/Output System	基本输入输出系统
CNA	Computing Node Agent	计算节点代理
DB	DataBase	数据库

DHCP	Dynamic Host Configuration Protocol	动态主机配置协议
DNS	Domain Name Server	域名服务器
ITA	IT Adaptor	IT 适配器
MTBF	Mean Time Between Failure	平均无故障时间
MTTR	Mean Time To Repair	平均维护时间
NC	Network Computer	网络计算机
NEBS	Network Equipment Building System	网络设备构建系统
PDU	Power Distribution Unit	配电单元
RAID	Redundant Array of Independent Disks	独立磁盘冗余阵列
RBD	Reliability Block Diagram	可靠性逻辑块图
SCSI	Small Computer System Interface	小型计算机系统接口
TFTP	Trivial File Transfer Protocol	简单文件传输协议