第3章 云计算机制

本章主要介绍常见的云计算机制，包括云基础设施机制、云管理机制、云监控机制和特殊云机制。通过本章的学习，应能够对云计算的机制有所了解。

3.1 云基础设施机制

云基础设施机制是云环境的基础构建块，它是形成云技术架构基础的主要构件。云基础设施机制主要包括如下。

- 虚拟网络边界；
- 虚拟服务器；
- 云存储设备；
- 云使用监控；
- 资源备份；
- 就绪环境。

这些机制并非全都应用广泛，也不需要为其中的每一个机制都建立独立的架构层。相反，它们应被视为云平台中常见的核心组件。

3.1.1 虚拟网络边界

虚拟网络边界(virtual network perimeter)通常是由提供和控制数据中心连接的网络设备建立，一般是作为虚拟化环境部署的，如虚拟防火墙、虚拟网络(VLAN、VPN)。该机制被定义为将一个网络环境与通信网络的其他部分隔开，形成一个虚拟网络边界，包含并隔离了一组相关的基于云的IT资源，这些资源在物理上可能是分布式的。

该机制可被用于如下的几个方面。

- 将云中的IT资源与非授权用户隔离；
- 将云中的IT资源与非用户隔离；
- 将云中的IT资源与云用户隔离；
- 控制被隔离IT资源的可用带宽。

1. 虚拟防火墙

图3.1是虚拟防火墙的示意图，虚拟防火墙是一个逻辑概念，该技术可以在一个单一的硬件平台上提供多个防火墙实体，即把一台防火墙设备在逻辑上划分成多台虚拟防火墙，每台虚拟防火墙都可以被看成是一台完全独立的防火墙设备，可拥有独立的管理员、安全策

略、用户认证数据库等。

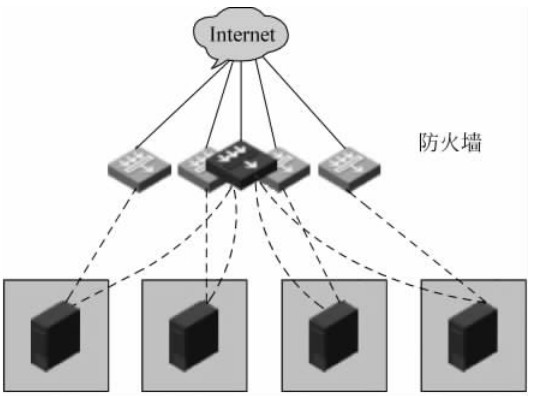

图 3.1　虚拟防火墙示意图

每个虚拟防火墙能够实现防火墙的大部分特性，并且虚拟防火墙之间相互独立，一般情况下不允许相互通信。

虚拟防火墙具有以下技术特点。

(1) 每个虚拟防火墙独立维护一组安全区域；

(2) 每个虚拟防火墙独立维护一组资源对象(地址/地址组、服务/服务组等)；

(3) 每个虚拟防火墙独立维护自己的包过滤策略；

(4) 每个虚拟防火墙独立维护自己的 ASPF 策略、NAT 策略、ALG 策略；

(5) 可限制每个虚拟防火墙占用资源数，如防火墙 Session 以及 ASPF Session 数目。

虚拟防火墙不仅解决了业务多实例的问题，更主要的是，通过它可将一个物理防火墙划分为多个逻辑防火墙来用。多个逻辑防火墙可以单独配置不同的安全策略，同时在默认情况下，不同的虚拟防火墙之间是默认隔离的。

2. 虚拟专用网络

虚拟专用网络(VPN)是一种通过公用网络(如 Internet)连接专用网络(如办公室网络)的方法。

它将拨号服务器的拨号连接的优点与 Internet 连接的方便与灵活性相结合。通过使用 Internet 连接，用户可以同时在大多数地方通过距离最近的 Internet 访问电话号码连接到自己的网络。

VPN 使用经过身份验证的链接来确保只有授权用户才能连接到自己的网络，而且这些用户使用加密来确保他们通过 Internet 传送的数据不会被其他人截取和利用。Windows 使用点对点隧道协议(PPTP)或第二层隧道协议(L2TP)实现此安全性。

通过图 3.2 所示，VPN 技术使得公司可以通过公用网络(如 Internet)连接到其分支办事处或其他公司，同时又可以保证通信安全。通过 Internet 的 VPN 连接从逻辑上讲相当于一个专用的广域网(WAN)链接。

VPN 系统的主要特点如下。

(1) 安全保障

虽然实现 VPN 的技术和方式很多，但所有的 VPN 均应保证通过公用网络平台传输数

图 3.2　虚拟专用网络(VPN)基本原理

据的专用性和安全性。在安全性方面,由于 VPN 直接构建在公用网上,实现简单、方便、灵活,但同时其安全问题也更为突出。企业必须确保其 VPN 上传送的数据不被攻击者窥视和篡改,并且要防止非法用户对网络资源或私有信息的访问。

(2) 服务质量保证

VPN 应当为企业数据提供不同等级的服务质量保证(QoS)。不同的用户和业务对服务质量保证的要求差别较大。在网络优化方面,构建 VPN 的另一重要需求是充分有效地利用有限的广域网资源,为重要数据提供可靠的带宽。广域网流量的不确定性使其带宽的利用率很低,在流量高峰时引起网络阻塞,使实时性要求高的数据得不到及时发送;而在流量低谷时又造成大量的网络带宽空闲。QoS 通过流量预测与流量控制策略,可以按照优先级实现带宽管理,使得各类数据能够被合理地先后发送,并预防阻塞的发生。

(3) 可扩充性和灵活性

VPN 必须能够支持通过 Intranet 和 Extranet 的任何类型的数据流,方便增加新的节点,支持多种类型的传输媒介,可以满足同时传输语音、图像和数据等应用对高质量传输以及带宽增加的需求。

(4) 可管理性

从用户角度和运营商角度应可方便地进行管理、维护。VPN 管理的目标为:减小网络风险、具有高扩展性、经济性、高可靠性等优点。事实上,VPN 管理主要包括安全管理、设备管理、配置管理、访问控制列表管理、QoS 管理等内容。

3.1.2　虚拟服务器

服务器通常通过虚拟机监视器(VMM)或虚拟化平台(Hypervisor)来实现硬件设备的抽象、资源的调度和虚拟机的管理。虚拟服务器(virtual server)是一种模拟物理服务器的虚拟化软件。虚拟服务器与虚拟机(VM)为同义词,虚拟基础设施管理器(VIM)用于协调与 VM 实例创建相关的物理服务器。虚拟服务器需要对服务器的 CPU、内存、设备及 IO 分别实现虚拟化。

通过向云用户提供独立的虚拟服务实例,云提供者使多个云用户共享同一个物理服务器。如图 3.3 所示。

每个虚拟服务器都可以存储大量的 IT 资源、基于云的解决方案和各种其他的云计算机制。从映像文件进行虚拟服务器的实例化是一个可以快速且按需完成的资源分配过程。通过安装和释放虚拟服务器,云用户可以定制自己的环境,这个环境独立于其他正在使用由同一底层物理服务器控制的虚拟服务器的云用户。虚拟服务器的具体内容将在 4.2 节详细

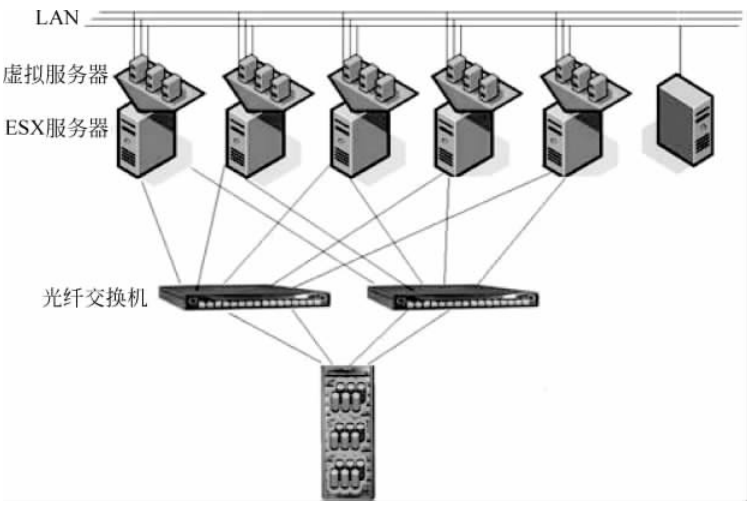

图 3.3　虚拟服务器基本架构

地介绍。虚拟服务器有以下几个特性。

(1) 多实例

通过服务器虚拟化,一台物理机上可以运行多个虚拟服务器,支持多个客户操作系统,并且物理系统的资源是以可控的方式分配给虚拟机。

(2) 隔离性

虚拟服务器可以将同一台物理服务器上的多个虚拟机完全隔离开来,多个虚拟机之间就像多个物理机器之间一样,每个虚拟机都有自己独立的内存空间,一个虚拟机的崩溃并不会影响到其他虚拟机。

(3) 封装性

一个完整的虚拟机环境对外表现为一个单一的实体,便于在不同的硬件设备之间备份、移动和复制。同时,虚拟服务器将物理机器的硬件封装为标准化的虚拟硬件设备提供给虚拟机内的操作系统和应用程序,提高了系统的兼容性。

基于以上这些特性,虚拟服务器带来了如下的优点。

(1) 实时迁移

实时迁移是指在虚拟机运行时,将虚拟机的运行状态完整、快速地从一个宿主平台迁移到另一个宿主平台,整个迁移过程是平滑且对用户透明的。由于虚拟服务器的封装性,实时迁移可以支持原宿主机和目标宿主机硬件平台之间的异构性。

当一台物理机器的硬件需要维护或更新时,实时迁移可以在不宕机的情况下将虚拟机迁移到另一台物理机器上,大大提高了系统的可用性。

(2) 快速部署

在传统的数据中心中,部署一个应用需要安装操作系统、安装中间件、安装应用、配置、测试、运行等多个步骤,通常需要耗费十几个小时甚至几天的时间,并且部署过程中容易产生错误。

而采用虚拟服务器之后,部署一个应用其实就是部署一个封装好操作系统和应用程序的虚拟机,部署过程只需要以下几个步骤:拷贝虚拟机、启动虚拟机和配置虚拟机,通常只

需要十几分钟，且部署过程自动化，不易出错。

(3) 高兼容性

虚拟服务器提供的封装性和隔离性使应用的运行平台与物理底层分离，提高了系统的兼容性。

(4) 提高资源利用率

在传统的数据中心，出于对管理性、安全性和性能的考虑，大部分服务器上都只运行一个应用，导致服务器的 CPU 使用率很低，平均只有 5%～20%。采用虚拟服务器之后，可以将原来多台服务器上的应用整合到一台服务器之中，提高了服务器资源的利用率，并且通过服务器虚拟化固有的多实例、隔离性和封装性保证了应用原有的性能和安全性。

(5) 动态调度资源

虚拟服务器可以使用户根据虚拟机内部资源的使用情况即时地灵活调整虚拟机的资源，如 CPU、内存等，而不需要像物理服务器一样需要打开机箱变更硬件。

3.1.3 云存储设备

云存储设备(cloud storage device)机制是指专门为基于云配置所设计的存储设备。这些设备的实例可以被虚拟化。其单位如下。

- 文件(file)。数据集合分组存放于文件夹中的文件里。
- 块(block)。存储的最低等级，最接近硬件，数据块是可以被独立访问的最小数据单位。
- 数据集(dataset)。基于表格的、以分隔符分隔的或以记录形式组织的数据集合。
- 对象(object)。将数据及其相关的元数据组织为基于 Web 的资源，各种类型的数据都可以作为 Web 资源被引用和存储，如利用 HTTP 的 CRUD(create、retrieve、update、delete)操作(如 CDMI、Cloud Data Management Interface)。

随着图 3.4 所示的云存储的广泛应用，一个与云存储相关的主要问题出现了，就是数据的安全性、完整性和保密性，当数据被委托给外部云提供者和其他第三方时，就更容易出现危险。此外，数据出现跨地域或国界的迁移时，也会导致法律和监管问题。

图 3.4 云存储广泛应用

(1) 用户的操作安全

当一个用户在公司编辑某个文件后,回到家中再次编辑,那么他再次回到公司时文件已是昨晚更新过的,这是理想状态下的情况。在很多时候用户编辑一个文件后,会发现编辑有误,想取回存在公司的文件版本时,可能在没有支持版本管理的云存储中用户的副本也已经被错误地更新了;同样的道理,当删除一个文件的时候,如果没有额外的备份,也许再到网盘回收站中也找不到了。版本管理在技术上不存在问题,但是会加大用户的操作难度。目前的云存储服务商只有少数的私有云提供商提供有限的支持,多数情况下这种覆盖是时常发生的。

(2) 服务端的安全操作

云存储设备早已成为黑客入侵的目标,因为设备上不仅有无穷的用户数据,对此类大用户群服务的劫持更加是黑色收入的重要来源。也就是说云存储设备的安全性直接影响着用户上传数据的安全。在虚拟服务器技术的支撑下,V2V(virtual to virtual)迁移的可靠性相当高,多数的云存储厂商都预备安全防护方案。

3.1.4 资源备份

如图 3.5 和图 3.6 不同视角的展示,与传统机构视角不同的是,云计算集中部署计算和存储资源,提供给各个用户。这既避免了用户重复建设信息系统的低效率,又能赋予用户价格低廉且近乎无限的计算能力。云计算提供的资源是弹性可扩展的,可以动态部署、动态调度、动态回收,以高效的方式满足业务发展和平时运行峰值的资源需求。云计算使用了资源备份容错、计算节点同构可互换等各种措施来保障服务的高可靠性,并拥有专业的维护队伍。

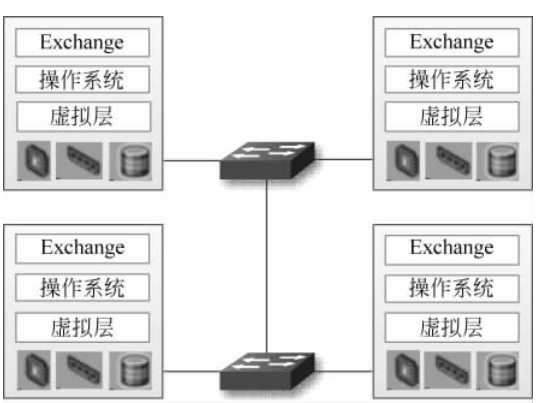

图 3.5 传统机构视角

图 3.6 云计算架构视角

资源备份(resource backup)可对同一个 IT 资源创建多个实例。资源备份用于加强 IT 资源的可用性和性能。使用虚拟化技术来实现资源备份机制,可以复制基于云的 IT 资源(如整个数据中心中的应用、数据)实现集中的备份和恢复,确保当出现系统故障、误操作等情况下应用系统仍然可用和可恢复。

3.1.5 就绪环境

PaaS平台就是指云环境中的应用即服务(包括应用平台、集成、业务流程管理和数据库服务),也可以说是中间件即服务。PaaS平台在云架构中位于中间层,其上层是SaaS,其下层是IaaS,基于IaaS之上的是为应用开发(可以是SaaS应用,也可以不是)提供接口和软件运行环境的平台层服务。

就绪环境机制是PaaS云交付模型的定义组件,基于云平台,已有一组安装好的IT资源,可以被云用户使用和定制。云用户利用就绪环境机制进行远程开发和配置自身的服务和应用程序。如数据库、中间件、开发工具和管理工具以及进行开发和部署Web应用程序。

Oracle的共享、高效的PaaS框架如图3.7所示,其中解释了就绪环境机制的实现位于应用运行环境层(aPaaS),为用户提供一套完整的运行环境。

(1) iPaaS:基于SOA、ESB、BPM等架构,是云内/云与企业间的集成平台;

(2) aPaaS共享:基于Java等应用技术架构,是应用的部署与运行环境平台;

(3) dPaaS可灵活伸缩:是数据存储与共享平台,提供多租户环境下高效与安全的数据访问;

(4) 硬件资源池:为PaaS平台提供所需要的高性能硬件资源系统。

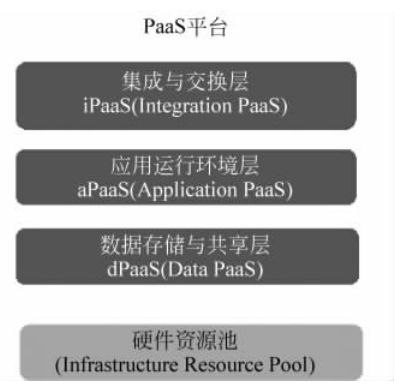

图3.7 Oracle的PaaS框架

3.2 云管理机制

云管理与传统管理的比较如表3.1所示。

表3.1 云管理与传统管理的比较

比较内容	传统管理	云管理
管理对象	网络、存储、服务器、OS、数据库、中间件、应用	IaaS、PaaS、SaaS等各种云服务
管理目标	实现IT系统的正常运作	实现云服务的端对端交付
管理特色	需要专业的管理技能 手动管理 竖井式管理	通过封装屏蔽底层细节 自服务 多租户,共享管理平台
管理平台易用性	安装配置复杂	自配置、自修复、自优化
管理规模	100节点	10 000节点+
用户	管理员	分层管理,多租户
整合	基于事件、数据库、私有接口的整合	面向服务的整合
管理手段	离散的工具	充分自动化

经过表 3.1 所示的云管理与传统管理的比较，不难发现基于云的 IT 资源需要被建立、配置、维护和监控。远程管理系统是必不可少的，它们促进了形成云平台与解决方案的 IT 资源的控制和演化，从而形成了云技术架构的关键部分，与管理相关的机制如下。

- 远程管理系统；
- 资源管理系统；
- SLA 管理系统；
- 计费管理系统。

这些系统通常提供整合的 API，并能够以个别产品、定制应用或者各种组合产品套装和多功能应用的形式提供给用户。

3.2.1 远程管理系统

远程管理系统(remote administration system)向外部的云资源管理者提供工具和用户界面来配置并管理基于云的 IT 资源。

如图 3.8 所示，远程管理系统能建立一个入口以便访问各种底层系统的控制和管理功能，这些功能包括了资源管理、SLA(服务等级协议)管理和计费管理。

图 3.8 远程管理系统的主要功能

远程管理系统主要创建如下两种类型的入口。

(1) 使用与管理入口。一种通用入口，集中管理不同的基于云的 IT 资源，并提供资源使用报告。

(2) 自助服务入口。该入口允许云用户搜索云提供者提供的最新云服务和 IT 资源列表。然后，云用户向云提供者提交其选项进行资源分配。

这个系统也包括 API，云用户可以通过这些标准 API 来构建自己的控制台。云用户可能使用多个云提供者的服务，也可能更换提供者，云用户通常能执行的任务包括如下。

- 配置与建立云服务；
- 为按需云服务提供和释放 IT 资源；
- 监控云服务的状态、使用和性能；
- 监控 QoS 和 SLA 的实行；
- 管理租赁成本和使用费用；
- 管理用户账户、安全凭证、授权和访问控制；
- 追踪对租赁服务内部与外部的访问；
- 规划和评估 IT 资源供给；
- 容量规划。

3.2.2 资源管理系统

资源管理系统(resource management system)可帮助协调 IT 资源，以便响应云用户和云提供者执行的管理操作。如图 3.9 所示，资源管理系统包含一个 VIM 平台和一个虚拟机映像库。VIM 也可能有额外的库，包括专门用来存放操作数据的库。

通常通过资源管理系统自动化实现的任务包括如下。

(1) 管理用来创建预构建实例的虚拟 IT 资源模板,如虚拟服务器映像;

(2) 在可用的物理基础设施中分配和释放虚拟 IT 资源模板,以响应虚拟 IT 资源实例的开始、暂停、继续和终止;

(3) 在有其他机制参与的条件下,协调 IT 资源,如资源复制、负载均衡和故障转移系统;

(4) 在云服务实例的生命周期中,强制执行使用策略与安全规格;

(5) 监控 IT 资源的操作条件。

图 3.9 资源管理系统组成

3.2.3 SLA 管理系统

SLA 管理系统(Service Level Agreement Management,SLA)代表的是一系列商品化的可用云管理产品,这些产品提供的功能包括:SLA 数据的管理、收集、存储、报告以及运行时通知。对相关数据的管理、收集、存储、报告以及运行时通知。通常会有一个服务资料测量库。对相关数据的管理、收集、存储、报告以及运行时通知,通常会有一个服务资料测量库。图 3.10 是一个包含一个 SLA 管理器和 QoS 测量库的 SLA 管理系统。

3.2.4 计费管理系统

计费管理系统(billing management system)专门用于收集和处理使用数据,它涉及云提供者的结算和云用户的计费。计费管理系统依靠按使用付费监控器来收集运行时的使用数据。这些数据存储在系统组件的一个库中,为了计费、报告和开发票等目的,可以从库中提取数据,图 3.11 是一个由定价与合同管理器和一个按使用付费测量库构成的计费管理系统。

图 3.10 SLA 管理系统组成

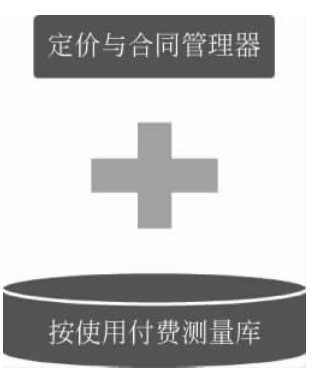

图 3.11 计费管理系统组成

3.3 云监控机制

云监控机制是一种轻量级的自治软件程序,用于收集和处理 IT 资源的使用数据,其主要内容包括如下。

- 资源监控；
- 能量监控；
- SLA 监控；
- 安全监控。

3.3.1 资源监控

资源监控是为了保证应用和服务的性能，开发者必须依据应用程序、服务的设计和实现机制估算工作负载，确定所需资源和容量的数量，避免资源供应不足或供应过量。

虽然负载估计值可通过静态分析、测试和监控得到，但实际上系统负载变化迅速，难以预测。云提供商通常负责资源管理和容量规划，提供 QoS 保证。因此，监控对于云提供商是至关重要的。提供商根据监控信息追踪各种 QoS 参数的变化，观察系统资源利用情况，从而准确规划基础设施和资源，遵守 SLA。

3.3.2 能量监控

云计算是对既有的计算资源在一种全新模式下的重组。在云端，数以万计的服务器提供近乎无穷的计算能力，而云用户根据自己的需求获取相应的计算能力。集中的存储和计算，形成了云能耗黑洞。云计算系统作为未来信息通信系统中内容与服务的源头与处理核心，也已成为信息通信系统的能耗大户。能量支出已经成为云计算系统运营不断增加的成本，有可能超过购买硬件资源的成本。为充分利用能量，提供系统能效，降低能量成本，需要从监控能耗入手，利用采集来的系统运营状态参数对服务器中的主要耗能部件进行建模分析，为节能策略的构建提供依据。

云计算是一种按使用量付费的模式，使用付费监控器(pay-per-use monitor)按照预先定义好的定价参数测量基于云的 IT 资源使用，使用期间生成的使用日志可以用于计算费用，日志主要包括如下。

- 请求/响应消息数量；
- 传送的数据量；
- 带宽消耗量。

3.3.3 SLA 监控

云计算市场在持续增长，用户如今关注的不仅仅是云服务的可用性，他们想要知道厂商能否为终端用户提供更好的服务。因此，用户更关注服务级协议，也就是 SLA(Service Level Agreement)，并需要监控 SLA 的执行情况。

服务等级协议(SLA)是服务提供者和客户之间的一个正式合同，用来保证可计量的网络性能达到所定义的品质。

SLA 监控器(SLA monitor)被用来专门观察云服务运行时性能，确保它们履行了 SLA 公布的约定 QoS 需求。例如轮询检测是否在线，检测 QoS 是否达到 SLA 的要求。

SLA 监控器保证的服务体系架构如图 3.12 所示，需要三个服务角色：一个服务提供者、一个服务客户和一个服务代理。

首先，通过在适当的平台上创建一个 Web 服务并生成 WSDL 文档和服务的基本 SLA，服务提供者发布一个由 SLA 保证的 Web 服务。

然后，它把服务细节发送到服务代理以存储在资源库中。服务客户向代理注册，然后在代理的资源库中搜索并发现适当的 Web 服务，检索服务的 WSDL 和 SLA。

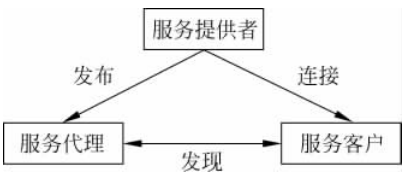

图 3.12　SLA 监控器保证的服务体系架构

最后它再与提供者协商把 SLA 正规化、确定下来并绑定到它的 Web 服务。

使用 SLA 监控器需要注意这些问题：第三方监控、告警装置、转换 SLA 以及有效的后备设施。

(1) 第三方监控

审计是很重要的一步，能够确保安全，保证 SLA 的承诺和责任归属，保持需求合规。用户可以用第三方监控。如果用户在云中运行业务关键的应用，这项服务应该保持定期审查，确保合格，敦促厂商与 SLA 步调一致。

(2) 转换 SLA，帮助整个业务成果

尽管云计算市场正在迅猛增长，中小企业的 IT 大多数都不够成熟，不足以支撑基于基础设施的 SLA 来帮助业务发展。企业应该选择最适合业务需求的 SLA，而不是急急忙忙签署协议。

如果企业操之过急，直接选择基础设施级别的 SLA，可能会由公司内部产生花费。例如，某企业想要 99.999%的高可用性，服务商就会提供更多冗余和灾难恢复，结果花费大幅提高。

当聚焦于节俭型业务级别 SLA 时，云计算 SLA 监控应该具有逻辑性和可行性，而不仅仅是基础设施级别的 SLA。

(3) 确保告警装置

为了让 SLA 监控更高效，用户要确保可以通过 Web 门户定期报告可用性和责任时间。用户应该保证及时的 Email 告警。

(4) 确保厂商有高效的后备设施

不同的厂商对于数据保护的系统也不同。但是有的厂商会把该职责推给客户，这样的话客户只好自己保护数据。因此用户应该确定服务商在签署 SLA 时，是否对此负有责任。

3.3.4　安全监控

由于服务模式的差异和新技术的应用，云计算还面临着新的安全风险挑战，如密钥和数据的防泄漏、动态数据隔离、虚拟化计算安全等。

云计算的安全风险是由其服务模式和自身的特性决定的。云端使用浏览器来接入云计算中心，以访问云中的 IaaS、PaaS 或 SaaS 服务，接入端的安全性直接影响到云计算的服务安全。为此，需要对云系统进行有效的监控，获取云系统内部的全局状态信息，向用户报告其服务的运行状态，通过对全局的监控信息进行关联分析，从而及时发现云平台内部可能的攻击，提高云用户的服务质量。

云计算安全涉及多个服务层次，需要构建跨层的、全面的监控机制，才能保证云的安全。

另外，各个国家不同的法律和隐私保护制度，以及监控标准的缺乏，使得云计算的安全监控变得更为复杂。

3.4 特殊云机制

典型的云技术架构包括大量灵活的部分，这些部分应对 IT 资源和解决方案有不同的使用要求，有如下特殊云机制。

- 自动伸缩监听器；
- 负载均衡器；
- 故障转移系统；
- 虚拟机监控器；
- 资源集群；
- 多设备代理；
- 状态管理数据库。

可以把所有这些机制看成对云基础设施的扩展。

3.4.1 自动伸缩监听器

自动伸缩监听器(automated scaling listener)是一个服务代理，它监听和追踪用户和云服务之间的通信或 IT 资源的使用情况。实际就是监听，如果发现超过阈值(大或者小，例如 CPU＞70％，用户请求每秒大于 10 个，并持续 10min)，通知云用户(VIM 平台)，云用户(VIM 平台)可以进行调整。注意，这只是监听器，监听自动伸缩的需求，不是处理自动伸缩。如果扩展需求在同一物理服务器上无法实现，则需要 VIM 执行虚拟机在线迁移，迁移到满足条件的另一台物理服务器上。

对于不同负载波动的条件，自动伸缩监控器可以提供不同类型的响应，例如：

(1) 根据云用户实现定义的参数，自动伸缩 IT 资源；

(2) 当负载超过当前阈值或低于已分配资源时，自动通知云用户。

3.4.2 负载均衡器

负载均衡器(load balancer)是一个运行时代理，有下面三种方式，它们都是分布式的，而不是主/备(备份)的方式，因为这是为了均衡。该机制可以通过交换机，专门的硬件/软件设备，以及服务代理来实现。

(1) 非对称分配(asymmetric distribution)：较大的工作负载被送到具有较强处理能力的 IT 资源；

(2) 负载优先级(workload prioritization)：负载根据其优先级别进行调度、排队、丢弃和分配；

(3) 上下文感知的分配(content-aware distribution)：根据请求内容分配到不同的 IT 资源。

负载均衡器被程序编码或者被配置成含有一组性能和 QoS 规则的参数，一般目标是优

化 IT 资源使用，避免过载并最大化吞吐量。负载均衡器机制可以是：多层网络交换机、专门的硬件设备、专门的基于软件的系统或服务代理。

负载均衡实现方式有以下几类。

1. 软件负载均衡技术

该技术适用于一些中小型网站系统，可以满足一般的均衡负载需求。软件负载均衡技术是在一个或多个交互的网络系统中的多台服务器上安装一个或多个相应的负载均衡软件来实现的一种均衡负载技术。

软件可以很方便地安装在服务器上，并且实现一定的均衡负载功能。软件负载均衡技术配置简单，操作也方便，最重要的是成本很低。

2. 硬件负载均衡技术

由于硬件负载均衡技术需要额外地增加负载均衡器，成本比较高，所以适用于流量高的大型网站系统。不过对于目前较有规模的企业网、政府网站来说，他们都会部署硬件负载均衡设备，原因是一方面硬件设备更稳定，另一方面也是为了合规性达标的目的。

硬件负载均衡技术是在多台服务器间安装相应的负载均衡设备，也就是负载均衡器来完成均衡负载的技术，与软件负载均衡技术相比，能达到更好的负载均衡效果。

3. 本地负载均衡技术

本地负载均衡技术是对本地服务器群进行负载均衡处理。该技术通过对服务器进行性能优化，使流量能够平均分配在服务器群中的各个服务器上，本地负载均衡技术不需要购买昂贵的服务器或优化现有的网络结构。

4. 全局负载均衡技术

全局负载均衡技术（也称为广域网负载均衡）适用于拥有多个服务器集群的大型网站系统。全局负载均衡技术是对分布在全国各个地区的多个服务器进行负载均衡处理，该技术可以通过对访问用户的 IP 地理位置判定，自动转向地域最近点。很多大型网站都使用这种技术。

5. 链路集合负载均衡技术

链路集合负载均衡技术是将网络系统中的多条物理链路，当作单一的聚合逻辑链路来使用，使网站系统中的数据流量由聚合逻辑链路中所有的物理链路共同承担。这种技术可以在不改变现有的线路结构，不增加现有带宽的基础上大大提高网络数据吞吐量，节约成本。

3.4.3 故障转移系统

故障转移系统（failover system）通过集群技术提供冗余实现 IT 资源的可靠性和可用性。故障转移集群是一种高可用的基础结构层，由多台计算机组成，每台计算机相当于一个冗余节点，整个集群系统允许某部分节点掉线、故障或损坏而不影响整个系统的正常运作。

一台服务器接管发生故障的服务器的过程通常称为“故障转移”。如果一台服务器变为不可用，则另一台服务器自动接管发生故障的服务器并继续处理任务。集群中的每台服务

器在集群中至少有一台其他服务器确定为其备用服务器。故障转移系统有两种基本配置。

(1) 主动-主动。IT 资源的冗余实现会主动地同步服务工作负载,失效的实例从负荷均衡调度器中删除(或置为失效)。

(2) 主动-被动。有活跃实例和待机实例(无负荷,可最小配置),如果检测到活跃实例失效时,将被重定向到待机实例,该待机实例就成为了活跃实例。原来的活跃实例如果恢复或者重新建立,可成为新的待机实例。这是冗余机制。

负载均衡是对新请求进行保护,对于正在处理的请求(或者请求组)是会丢失的。采用哪种方式,具体由业务特性决定。

如图 3.13 所示,第一台服务器(Database01)是处理所有事务的活动服务器。仅当 Database01 发生故障时,处于空闲状态的第二台服务器(Database02)才会处理事务。故障转移集群将一个虚拟 IP 地址和主机名(Database10)在客户端和应用程序所使用的网络上公开。

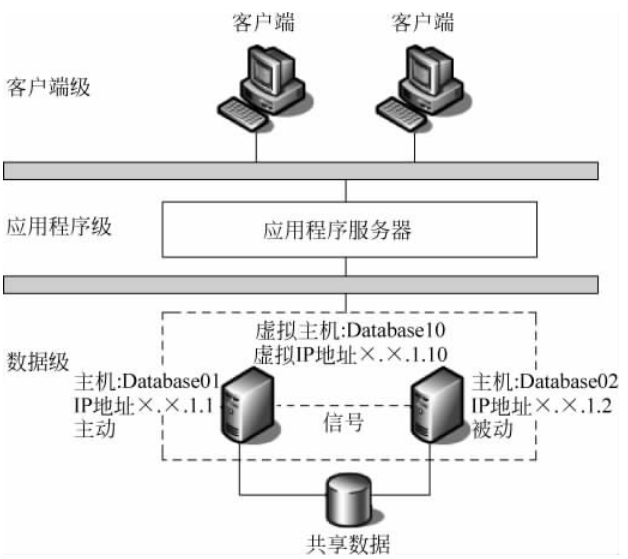

图 3.13　故障转移系统工作原理

3.4.4　虚拟机监控器

虚拟机监控器(Hypervisor)是虚拟化基础设施的最基础部分。Hypervisor 是一种运行在物理服务器和操作系统之间的中间软件层,可允许多个操作系统和应用共享一套基础物理硬件,因此也可以看作是虚拟环境中的"元"操作系统,它可以协调访问服务器上的所有物理设备和虚拟机,也叫虚拟机监视器(Virtual Machine Monitor),是 Hypervisor 的一部分。

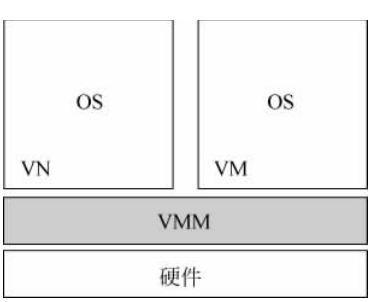

图 3.14　虚拟机监视器架构

如图 3.14 所示,VMM 提供了一组特性来管理跨物理服务器的多 Hypervisor,用来进行数据中心的硬件资源调度,例如分配合适 Hypervision,进行在线迁移到空

闲的物理机等。Hypervisor是所有虚拟化技术的核心。非中断地支持多工作负载迁移的能力是Hypervisor的基本功能。当服务器启动并执行Hypervisor时，它会给每一台虚拟机分配适量的内存、CPU、网络和磁盘，并加载所有虚拟机的客户操作系统。

常见的Hypervisor分两类：裸金属型和宿主型。

1. 裸金属型

裸金属型指VMM直接运作在裸机上，使用和管理底层的硬件资源，GuestOS对真实硬件资源的访问都要通过VMM来完成，作为底层硬件的直接操作者，VMM拥有硬件的驱动程序。裸金属虚拟化中Hypervisor直接管理调用硬件资源，不需要底层操作系统，也可以理解为Hypervisor被做成了一个很薄的操作系统。这种方案的性能处于主机虚拟化与操作系统虚拟化之间。代表是VMware ESX Server、Citrix Xen Server、Microsoft Hyper-V和Linux KVM。

2. 宿主型

宿主型指VMM之下还有一层宿主操作系统，由于Guest OS对硬件的访问必须经过宿主操作系统，因而带来了额外的性能开销，但可充分利用宿主操作系统提供的设备驱动和底层服务来进行内存管理、进程调度和资源管理等。主机虚拟化中VM的应用程序调用硬件资源时需要经过：VM内核→Hypervisor→主机内核，导致性能是三种虚拟化技术中最差的。主机虚拟化技术代表是VMware Server(GSX)、Workstation和Microsoft Virtual PC、Virtual Server等。由于主机型Hypervisor的效率问题，多数厂商采用了裸机型Hypervisor中的Linux KVM虚拟化，即为裸金属型。

3.4.5 资源集群

资源集群(resource cluster)将多个IT资源实例合并成组，使之能像一个IT资源那样进行操作。也就是“N in 1”。在实例间通过任务调度、数据共享和系统同步等进行通信。集群管理平台作为分布式中间件运行在所有的集群节点上。其类型包括如下。

(1) 服务器集群：运行在不同物理服务器上的虚拟机监控器可以被配置为共享虚拟服务器执行状态(例如内存页和处理器寄存器状态)，以此建立起集群化的虚拟服务器，通常需要物理服务器共享存储，这样虚拟服务器可以从一个物理服务器在线迁移到另一个。

(2) 数据库集群：具有同步的特性，集群中使用的各个存储设备上的存储数据具有一致性，提供冗余能力。

(3) 大数据集集群(large dataset cluster)：实现数据的分区和分布，目标数据集可以有效地划分区域，而不需要破坏数据的完整性或计算的准确性。每个节点都可以处理负载，而不需要向其他类型那样，与其他节点进行很多的通信。

其中HA集群是资源集群的一种，Linux-HA的全称是High-Availability Linux，它是一个开源项目。这个开源项目的目标是：通过社区开发者的共同努力，提供一个增强Linux可靠性(reliability)、可用性(availability)和可服务性(serviceability)(RAS)的集群解决方案。

其中Heartbeat就是Linux-HA项目中的一个组件，也是目前开源HA项目中最成功的一个例子。它提供了所有HA软件所需要的基本功能，如心跳检测和资源接管、监测群

集中的系统服务、在群集中的节点间转移共享 IP 地址的所有者等。其中包括节点、资源、事件和动作四个相关术语。

1. 节点(node)

运行 Heartbeat 进程的一个独立主机,称为节点。节点是 HA 的核心组成部分,每个节点上运行着操作系统和 Heartbeat 软件服务。在 Heartbeat 集群中,节点有主次之分,分别称为主节点和备用/备份节点,每个节点拥有唯一的主机名,并且拥有属于自己的一组资源,如磁盘、文件系统、网络地址和应用服务等。主节点上一般运行着一个或多个应用服务,而备用节点一般处于监控状态。

2. 资源(resource)

资源是一个节点可以控制的实体,并且当节点发生故障时,这些资源能够被其他节点接管。在 Heartbeat 中,可以当作资源的实体如下。

(1) 磁盘分区、文件系统;

(2) IP 地址;

(3) 应用程序服务;

(4) NFS 文件系统。

3. 事件(event)

事件就是集群中可能发生的事情,例如节点系统故障、网络连通故障、网卡故障、应用程序故障等。这些事件都会导致节点的资源发生转移,HA 的测试也是基于这些事件来进行的。

4. 动作(action)

事件发生时 HA 的响应方式,动作是由 shell 脚步控制的。例如,当某个节点发生故障后,备份节点将通过事先设定好的执行脚本进行服务的关闭或启动,进而接管故障节点的资源。

图 3.15 是一个 Heartbeat 集群的一般拓扑图,在实际应用中,由于节点的数目、网络结构、磁盘类型配置的不同,拓扑结构可能会有不同。在 Heartbeat 集群中,最核心的是 Heartbeat 模块的心跳监测部分和集群资源管理模块的资源接管部分,心跳监测一般由串行接口通过串口线来实现,两个节点之间通过串口线相互发送报文来告诉对方自己当前的

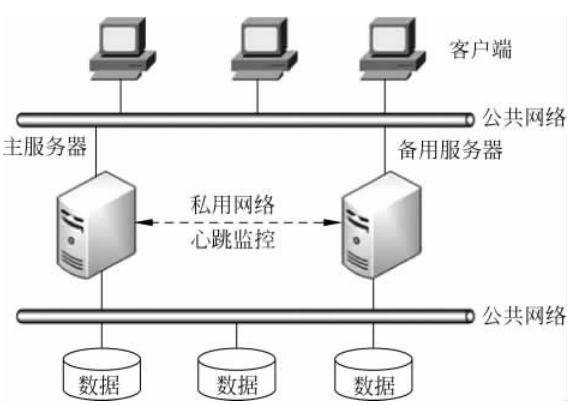

图 3.15 Heartbeat 集群的一般拓扑图

状态，如果在指定的时间内未收到对方发送的报文，那么就认为对方失效，这时资源接管模块将启动，用来接管运行在对方主机上的资源或者服务。

3.4.6 多设备代理

多设备代理(multi-device broker)机制用来帮助运行时的数据转换，使得云服务被更广泛的用户程序和设备所用。

多设备代理通常是作为网关存在的，或者包含网关的组件，例如：XML 网关、云存储网关和移动设备网关。

可以创建的转换逻辑层次包括如下：

- 传输协议；
- 消息协议；
- 存储设备协议；
- 数据模型/数据模式。

3.4.7 状态管理数据库

状态管理数据库(state management database)是一种存储设备，用来暂时地存储软件的状态数据。作为把状态数据缓存在内存中的一种替代方法，软件程序可以把状态数据卸载到数据库中，用以降低程序占用的运行时内存量。因此，软件程序和周边的基础设施都可以具有更大的可扩展性。

3.5 小　　结

基础机制是指在 IT 行业内确立的具有明确定义的 IT 构件，它通常区别于具体的计算模型和平台。云计算具有以技术为中心的特点，这就需要建立一套正式机制作为探索云技术架构的基础。本章介绍了云计算里常用的云计算机制，在实现过程中可以将它们组成不同的组合形式来具体应用。

3.6 习　　题

1. 云基础设施机制包括哪些？
2. 云管理机制包括哪些？
3. 特殊云机制包括哪些？