

第一部分

云计算概念与基础

第 1 章

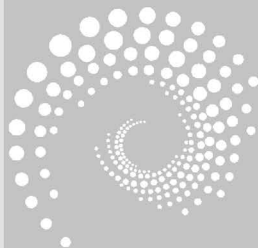
云计算概论

CHAPTER 1

本章首先介绍云计算的定义,旨在让读者对云计算有一个宏观的概念,然后介绍云计算的产生背景及其发展历史。通过本章的学习,读者可以对云计算产生一个初步的认识。

本章目标

- 了解云计算;
- 了解云计算的产生背景;
- 了解云计算的发展历史;
- 了解云计算与传统 IDC 的区别;
- 了解亚马逊云。



1.1 云计算的概念和由来

1.1.1 什么是云计算

云计算(cloud computing)是基于互联网相关服务的增加、使用和交付模式,通常涉及通过互联网提供动态易扩展且经常虚拟化的资源。云是网络、互联网的一种比喻说法。过去在图中往往用云表示电信网,后来也用于抽象表示互联网和底层基础设施。云计算甚至可以让用户体验每秒 10 万亿次的运算能力,拥有这么强大的计算能力可以模拟核爆炸、预测气候变化和市场发展趋势等。用户可以通过计算机、手机等方式接入数据中心,按自己的需求进行运算。

云计算的定义不一而论。对于到底什么是云计算,至少可以找到 100 种解释。现阶段广为接受的是美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)的定义:云计算是一种按使用量付费的模式,这种模式提供可用的、便捷的、按需的网络访问,进入可配置的计算资源共享池(资源包括网络、服务器、存储、应用、软件、服务),这些资源能够被快速提供,只需投入很少的管理工作或服务供应商进行很少的交互。

1.1.2 云计算的产生背景

云计算的产生是由很多推动因素构成的,其中包括以下几个方面。

1. 硬件成本下降和能效提升

硬件制造技术进步使服务器和网络设备性价比大幅提升,云计算可以利用大规模廉价的商用服务器组建数据中心。同时芯片技术也在持续进步,计算性能提高而能耗下降。这为云计算提供了基础设施支持。

2. 互联网高速发展

高速、无处不在的宽带互联网连接为访问云服务提供了可能。用户可以随时随地通过网络访问云服务提供的软件、计算能力、存储资源等。互联网的进步推动了分布式云架构的兴起。

3. 企业数字化转型需求

企业面临数字化和网络化转型的压力,需要更高弹性、敏捷性和速度。云计算具有弹性扩展、按需使用等优势,可以降低企业使用信息系统的成本,帮助企业实现数字化转型目标。

4. 大数据时代的数据分析需求

海量数据的生成需要强大的分布式数据处理能力。云计算基础架构正好提供了这个能力,可以按需调度大量服务器进行分析处理,使云平台成为大数据最合适的基础设施。

综上所述,云计算得以快速发展的背景主要在于技术条件的成熟、互联网进步的推动、

企业数字化转型的潮流和数据分析的需要。

1.1.3 云计算的发展历史

云计算的发展历史可以追溯到 20 世纪 90 年代初期,其后主要经历的阶段如下。

1. 雏形阶段(20 世纪 90 年代初至 2006 年)

这一时期出现了网格计算、应用服务提供商和部分公有云服务,诞生了云计算的雏形。例如,Salesforce 推出了在线客服和销售软件服务,亚马逊发布了 S3 云存储和 EC2 云服务器服务。在该阶段,云计算从理论走向实践。

2. 起步阶段(2006—2009 年)

2006 年,亚马逊正式推出 AWS,标志着公有云时代的到来。2007 年,IBM 和谷歌分别启动蓝云平台 and App Engine。2008 年,NASA 和 Rackspace 共同开发开放源代码 OpenStack 项目,并将其用于构建公有及私有云。在该阶段,云计算概念和技术不断成熟。

3. 高速发展阶段(2009 年至今)

云计算进入高速增长发展期,成为企业数字化转型的关键路径。越来越多的企业将核心业务移动到云平台,主要公有云服务商像 AWS、Azure、Google Cloud、AliCloud 平台等不断扩大市场规模。其中比较有代表性的是互联网企业 Netflix,其快速发展成为典型的云原生公司。

总体来说,经过多年的发展,云计算已经成为继个人计算机、互联网、移动互联网之后的新一代信息技术变革浪潮,并深刻地改变了人们的生产和生活方式。

1.2 云计算与传统 IDC 的区别

IDC(Internet Data Center,互联网数据中心)是一种集中提供计算、存储和网络等基础设施服务的数据中心,通常由专业的数据中心管理公司运营,为各类互联网企业和机构提供服务器托管、虚拟主机、数据存储、网络带宽等服务。IDC 的主要特点如下。

(1) 高可用性: IDC 通常具有高可用性和冗余设计,以确保在任何情况下都能够保持高水平的可用性和稳定性。

(2) 安全可靠: IDC 通常具有严格的安全措施和可靠的数据备份机制,以确保数据的安全性和可靠性。

(3) 快速响应: IDC 通常具有高效的运维团队和快速响应能力,能够及时处理各种问题和故障。

(4) 定制化服务: IDC 通常能够提供各种定制化的服务,以满足不同客户的需求。

随着云计算和大数据技术的不断发展,IDC 也在不断演进和变革,以满足新的业务需求和技术趋势。例如,IDC 正在越来越多地采用虚拟化技术、自动化运维技术、绿色节能技术

等,以降本增效并更好地满足客户的需求。

云计算与传统 IDC 的区别主要体现在以下几方面。

1. 资源供给方式

- IDC: 采用传统的自建模式,企业需要自行购买服务器、存储、网络等硬件设备,并建立机房环境。资源供给是有限且固定的。
- 云计算: 采用按需、弹性的资源供给模式。云服务提供商拥有大规模的数据中心资源池,用户可以根据实际需求按量获取所需的计算、存储、网络等资源,资源可以动态扩展和收缩。

2. 资源管理方式

- IDC: 企业需要自行管理和维护整个 IT 基础架构,包括硬件、软件、网络、安全等,需要大量的人力和技术投入。
- 云计算: 云服务提供商负责底层资源的管理和维护,用户只需要关注自身的应用程序和数据,大大降低了 IT 运维的复杂性。

3. 成本结构

- IDC: 采用高前期资本支出(CapEx)模式,需要购买昂贵的硬件设备和建设机房,并持续投入运维成本。这种成本结构的资源利用率相对较低。
- 云计算: 采用按需付费的运营支出(OpEx)模式,用户只需要为实际使用的资源付费。这样可以避免大量前期投资,降低了资源闲置的浪费,提高了资源利用效率。

4. 灵活性和扩展性

- IDC: 资源扩展受硬件容量限制,扩展往往需要购买新硬件、重新部署,响应速度较慢。
- 云计算: 资源可以根据需求实现快速、无缝的扩展和收缩,具有很高的灵活性和敏捷性。

5. 地理分布

- IDC: 服务设备通常集中在某些特定的物理位置,受地理位置的限制。
- 云计算: 云服务提供商通常在全球范围内部署多个可用区。用户可以根据需求选择不同区域的云资源,实现业务的分布式部署。

1.3 AWS

“亚马逊云”通常是指亚马逊的云计算服务平台,即 Amazon Web Services(AWS)。AWS 提供广泛的全球云基础设施服务,如计算、存储、数据库、分析、网络、移动、开发工具、管理工具、物联网、安全和企业应用等。这些服务帮助组织缩放和增长,被各种规模的公司用来运行应用程序和服务。

目前,AWS 是全球云计算市场的主导力量,在全球云基础设施服务市场的份额保持领先地位。在全球范围内,AWS 通过持续的创新和早期市场进入优势,建立了广泛的客户基础和服务生态系统,持续引领市场。

习题

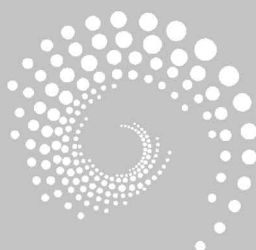
本书提供在线测试习题,扫描下面的二维码可以获取本章习题。



在线测试

云计算基础

CHAPTER 2



本章主要介绍关于云计算的各种基础知识,包括分布式计算、实现云计算的关键技术、云交付模型和云部署模式,同时介绍云计算的优势、面临的挑战和典型应用。通过本章的学习,读者能够对云计算有一个基本的认识。

本章目标

- 熟悉分布式计算;
- 熟悉云计算的关键技术;
- 掌握云交付模型;
- 掌握云部署模式;
- 了解云计算的优势与挑战;
- 了解典型云应用;
- 了解 AWS 的实验。

2.1 分布式计算

2.1.1 应用场景

分布式的概念十分广泛,凡是去中心的架构都可以理解为分布式。人们日常生活中最早接触到的分布式应该是 P2P。用户下载的文件不是集中存放到某个中心,而是分别存储在网络中的不同节点。当用户有下载需求时,可以从网络上的节点获取相应资源碎片,并形成下载文件。例如,用迅雷下载文件的方式就是 P2P。

除 P2P 外,还有很多分布式架构的应用场景。例如,CDN 技术将视频网站中的内容分布存储在就近的服务器上,从而形成分布式网络;大数据技术应用到分布式存储架构,将数据存储于不同的节点磁盘中,需要执行分析任务时切分为片段在分散的服务器节点中进行运算;区块链技术实现去中心化,将账目信息记录在不同的节点,交易时更新网络上的所有账目副本;应用架构中的分布式计算架构多应用于微服务。

分布式计算是一种计算方法,与集中式计算是相对的。随着计算技术的发展,一些应用需要巨大的计算能力才能完成,如果采用集中式计算,则需要耗费很长的时间才能完成。而分布式计算将应用分解成许多更小的部分,分配到多台计算机进行处理,这样可以节省整体计算时间,从而大大提高计算效率。云计算是分布式计算技术的一种,也是分布式计算这种科学概念的商业实现。

2.1.2 应用架构中的分布式

分布式计算的优点是发挥“集体的力量”,将大任务分解成小任务,分配给多个计算节点同时计算。分布式计算将计算扩展到多台计算机,甚至是多个网络,在网络有序地执行一个共同的任务。各节点间通信方式可以通过 RPC 调用、Q 消息队列或 Web Service 方式。在分布式计算发展起来之前的网络协议并不能满足分布式计算的要求,于是产生了 Web Service 技术。

分布式计算的实现需要借助 Web Service 接口。Web Service 是一个平台独立的、低耦合的、自包含的、基于可编程的 Web 应用程序接口,可使用开放的 XML(标准通用标记语言下的一个子集)标准来描述、发布、发现、协调和配置这些应用程序,用于开发分布式的、互操作的应用程序。

对于目前比较流行的微服务架构,主要采用 RPC 和 Web Service 方式提供服务间访问,基于 Web Service 的 API 访问获得了更多的应用和认可。如图 2.1 所示,微服务的体系结构是基于微服务提供者、微服务请求者、微服务注册中心三个角色和发布、发现、绑定三个动作构建的。简单地说,微服务提供者是微服务的拥有者,等待为其他服务和用户提供自己已有的功能;微服务请求者是微服务功能的使用者,利用 SOAP 或 RESTful 消息向微服务提供者发送请求以获得服务;微服务注册中心的作用是将服务请求者与合适的微服务提供者联系在一起,充当管理者的角色。这三个角色是根据逻辑关系划分的,在实际应用中,角色之间很可能有交叉:一个微服务既可以是微服务提供者,也可以是微服务请求者,还可以二者兼而有之。对于微服务角色之间的关系,“发布”是为了让用户或其他服务知道某个微

服务的存在和相关信息,“查找(发现)”是为了找到合适的微服务,“绑定”则是在提供者与请求者之间建立某种联系。在更为复杂的技术架构中,通常还会采用 Web Service 网管实现对服务请求的分发和处理,包括实现熔断、权限控制等高级功能。

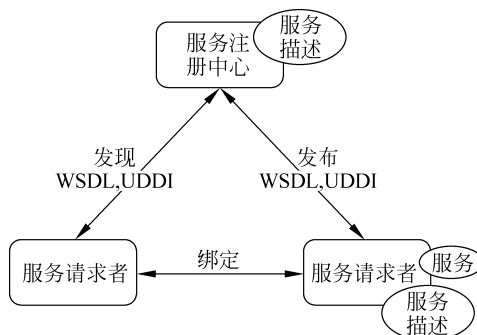


图 2.1 微服务的体系结构

这种技术的功能与中间件的功能有相似之处：微服务技术屏蔽不同开发平台功能模块相互调用的障碍,利用 HTTP 和 SOAP/RESTful 使商业数据在微服务上传输,从而调用这些开发平台不同的功能模块完成计算任务。由此可见,如果要在互联网上实施大规模的分布式计算,则需要微服务做支撑。

2.1.3 分布式计算和云计算的区别与联系

人们经常听到很多新的技术名词,如区块链、大数据、微服务、人工智能、容器等。这些概念(包括 2.1 节介绍的分布式计算)与云计算的关系是什么样的呢?应该说,云计算是更抽象也更广泛的一个概念。云计算可以简单理解为,用户的所有需求都可以通过服务的形式进行封装,当用户申请一个服务时,云平台可以自动地将服务请求转换为技术请求,自动在云平台的数据中心处理该服务器请求,并将处理完的结果返回给用户。用户可以更加专注在业务需求本身上,而不需要再关注和维护为了实现该业务需求而衍生的安装、调试和维护等工作。因此,无论是对企业内的私有云还是面对公众的公有云,云平台都将成为一个对外提供服务的统一窗口,同时借助自动化引擎和策略调度机制,将服务进行自动转换和处理。简而言之,云计算解决的是人和 IT 资源的关系。如同 QQ 解决的是人和人的关系,淘宝解决的是人和实体物品的关系,而云计算解决的是人和 IT 资源的关系。区别在于,淘宝并不生产物品,它只是实现了一个信息交易和共享平台,某种角度上是一个大的集成商的角色。而在云计算平台中,不仅要解决交易和信息,而且要实际地提供基础架构和应用服务的租赁,实现端到端的交付。这也就不难理解为什么 AWS 和阿里云在云计算领域做得最早、发展得最好,因为它们都是在解决人和物、人和资源的问题。

对区块链、大数据、微服务、人工智能、容器这些概念仔细分析,不难发现它们都不是解决人和服务或人和物品的。这些技术大多是对传统架构的升级和发展,对某个问题提供了更智能的算法模型,或者提供了更加高效、可靠、低成本的实现方式和技术变革。因此,这些技术都应涵盖在云计算概念中。这些技术既可以通过云计算实现,以服务的方式提供给用户进行使用,同时也可以不运用云计算技术进行实现。值得注意的是,通常这些新技术会和

云计算技术一并出现,因为这些新技术(包括所运用的分布式技术等)都需要创建多个计算或存储节点进行实现,为此需要大批量地创建和弹性伸缩这些节点,此时云计算的弹性服务可以提供便利的部署和使用。

2.1.1 节介绍了分布式存储的应用场景,而作为最典型的分布式场景,云存储与云计算有着天然的结合。就如本节所述,云计算解决的是人和 IT 资源之间的关系,而云存储则是作为基础架构中的重要部分对外提供服务。

2.2 云计算的关键技术

云计算是一种新型的超级计算方式,以数据为中心,是一种数据密集型的超级计算。云计算的目标是以低成本的方式提供高可靠、高可用、规模可伸缩的个性化服务,为此需要分布式海量数据存储、虚拟化技术、云平台技术、并行编程技术、数据管理技术等关键技术加以支持。

2.2.1 分布式海量数据存储

随着信息化建设的不断深入,信息管理平台已经完成了从信息化建设到数据积累的职能转变,在一些信息化起步较早、系统建设较规范的行业,如通信、金融、大型生产制造等领域,海量数据的存储、分析需求的迫切性日益明显。

以移动通信运营商为例,随着移动业务和用户规模的不断扩大,每天都会产生海量的业务、计费 and 网管数据,庞大的数据量使得传统的数据库存储已经无法满足存储和分析需求。主要面临的问题如下。

(1) 数据库容量有限。

关系数据库不是为海量数据而设计的,设计之初并没有考虑到数据量能够庞大到 PB 级。为了继续支撑系统,不得不进行服务器升级和扩容,成本高昂而难以接受。

(2) 并行取数困难。

除分区表可以并行取数外,其他情况都要对数据进行检索才能将数据分块,并行读数效果不明显,甚至增加了数据检索的消耗。虽然可以通过索引提升性能,但实际业务表明数据库索引作用有限。

(3) 对于 J2EE 应用而言,JDBC 的访问效率太低。

Java 的对象机制使得读取的数据都需要序列化,导致读数速度很慢。

(4) 数据库并发访问数太多。

由于数据库并发访问数太多,因此存在 I/O 瓶颈和数据库的计算负担太重两个问题,甚至出现内存溢出等现象,而数据库扩容成本太高。

为了解决以上问题,分布式存储技术得以发展。在技术架构上可以分为 3 类:解决企业数据存储和分析使用的大数据技术、解决用户数据云端存储的对象存储技术、满足云端操作系统实例需要用到的块存储技术。

(1) 大数据技术。

对于大数据技术,理想的解决方案是将大数据存储到分布式文件系统中。云计算系统

由大量服务器组成,为大量用户服务,因此云计算系统采用分布式存储的方式存储数据,用冗余存储的方式(集群计算、数据冗余和分布式存储)保证数据的可靠性。冗余的方式通过任务分解和集群,用低配机器替代超级计算机的性能来保证低成本,这种方式保证分布式数据的高可用、高可靠和经济性,即为同一份数据存储多个副本。云计算系统中广泛使用的数据存储系统是谷歌的 GFS 和 Hadoop 团队开发的 GFS 的开源实现 HDFS。需要注意的是,在处理交易系统时,与传统的数据库存储方式相比,目前大数据技术的 TPS(Transactions Per Second,每秒交易量)表现仍然存在差距。因此,大数据多用于分析系统,而在线实时交易还是采用数据库方式。

(2) 对象存储技术。

对于对象存储,人们非常熟悉的云盘就是基于该技术实现的。用户可以将照片、文本、视频直接通过图形界面进行云端上传、浏览和下载。其实,上传等操作界面最终都是通过 Web Service 与后台的对象存储系统进行交互,前端界面更多的是在用户、权限和管理层面上提供支持。对象存储的主要特点如下。

① 所有的存储对象都有自身的元数据和 URL,这些对象在尽可能唯一的区域复制 3 次,而这些区域可被定义为一组驱动器、一个节点、一个机架等。

② 开发者通过一个 RESTful HTTP API 与对象存储系统进行相互作用。

③ 对象数据可以放置在集群的任何地方。

④ 在不影响性能的情况下,集群通过增加外部节点进行扩展。这是相对全面升级、性价比更高的近线存储扩展方式。

⑤ 数据无须迁移到一个全新的存储系统。

⑥ 集群可无宕机增加新的节点。

⑦ 故障节点和磁盘可无宕机调换。

⑧ 在标准硬件上运行,普通的 x86 服务器即可接入。

(3) 块存储技术。

云平台的块存储技术是否有 S3 这一种就足够了? 答案是否定的。正所谓术业有专攻,S3 搭建的对象存储可以利用普通的 PC 服务器组建集群,从而便于实现对象的分布式存储。在商业应用中,类似数据库和操作系统,都是要在裸存储上进行安装才能发挥其最大的性能。在云平台上可以独立创建块存储并挂接到某个云实例。正如 2.2.3 节中提到的,云计算平台的优势在于提供简化的服务给用户使用。因此对于数据块的开通和挂载,云平台会完成相应的处理,用户只需要使用即可。如果按传统方式进行处理,则需要人工在存储上做大量操作和处理才能进行划分和挂载。

2.2.2 虚拟化技术

虚拟化技术是云计算系统的核心组成部分之一,是将各种计算及存储资源充分整合和高效利用的关键技术。云计算的虚拟化技术不同于传统的单一虚拟化,它是涵盖整个 IT 架构的,包括资源、网络、应用和桌面在内的全系统虚拟化。通过虚拟化技术可以隔离所有硬件设备、软件应用和数据,打破硬件配置、软件部署和数据分布的界限,实现 IT 架构动态化和资源集中管理,使应用能够动态地使用虚拟资源和物理资源,提高系统适应需求和环境的能力。

虚拟化技术的主要特点如下。

(1) 资源分享。通过虚拟机封装用户各自的运行环境,有效实现多用户分享数据中心资源。

(2) 资源定制。用户利用虚拟化技术配置私有服务器,指定所需的 CPU 数目、内存容量、磁盘空间等,实现资源的按需分配。

(3) 细粒度资源管理。将物理服务器拆分成若干虚拟机,可以提高服务器的资源利用率,且有助于服务器的负载均衡和节能。

基于以上特点,虚拟化技术已成为实现云计算资源池化和按需服务的基础。

2.2.3 云平台管理技术

云计算资源规模庞大,服务器数量众多且分布在不同的地点,同时运行着数百种应用。如何有效地管理这些服务器,保证整个系统提供不间断的服务是一项巨大的挑战。

云平台技术能够使大量的服务器协同工作,便于进行业务部署,快速发现和恢复系统故障,通过自动化、智能化的手段实现大规模系统的可靠运营。

云计算平台的主要特点是用户不必关心云平台底层的实现,只需要调用平台提供的接口就可以在云平台中完成自己的工作。利用虚拟化技术,云平台提供商可以实现按需提供服务,一方面降低了云的成本,另一方面保证了用户的需求得到满足。云平台基于大规模的数据中心或网络,因此云平台可以提供高性能的计算服务,且云资源对于云平台用户而言几乎是无限的。

云平台所服务的对象除个人外,大部分都是企业级用户。企业级用户无论是内部使用(私有云)还是外部租赁(公有云)都会涉及管理问题,不同部门使用资源的监控、预算、计量、自动化运维、审计、安全管控、流程控制、容量规划和管理等都是云平台管理所涉及的问题,在后续章节中会展开介绍。

2.2.4 并行编程技术

目前两种最重要的并行编程模型是数据并行和消息传递。数据并行编程模型的编程级别比较高,编程相对简单,但它仅适用于数据并行问题。消息传递编程模型的编程级别相对较低,但消息传递编程模型具有更广泛的应用范围。

1. 数据并行

数据并行是一种较高层次上的模型,它提供给编程者一个全局的地址空间,一般这种形式的语言本身就提供并行执行的语义。因此,编程者只需要简单地指明待执行的并行操作及对象,即可实现数据并行的编程。

例如,对于数组运算,数组 B 和 C 的对应元素相加后送给 A,则通过语句 $A=B+C$ 或其他表达方式能够实现上述功能,使并行机对 B、C 的对应元素并行相加,并将结果并行赋给 A。因此,数据并行的表达是相对简单和简洁的,它不需要编程者关心并行机是如何对该操作进行并行执行的。

数据并行编程模型虽然可以解决一大类科学与工程计算问题,但是对于非数据并行类

的问题,通过数据并行的方式来解决则难以取得较高的效率。

2. 消息传递

消息传递是指各个并行执行的部分之间通过传递消息实现交换信息、协调步伐、控制执行,一般是面向分布式内存的,但也可适用于共享内存的并行机。消息传递为编程者提供了更灵活的控制手段和表达并行的方法,是消息传递并行程序能提供较高执行效率的重要原因。

消息传递模型一方面为编程者提供了灵活性,另一方面也将各个并行执行部分之间复杂的信息交换和协调、控制的任务交给了编程者,在一定程度上增加了编程者的负担。同时,这也是消息传递编程模型编程级别低的主要原因。消息传递的基本通信模式是简单、清楚的,学习和掌握这些部分其实并不困难。因此,目前大量的并行程序设计仍然是消息传递并行编程模式。

云计算通常采用并行编程模式。在并行编程模式下,并发处理、容错、数据分布、负载均衡等细节都被抽象到一个函数库中,通过统一接口,用户大型的计算任务被自动并发和分布执行,即将一个任务自动分成多个子任务后并行地处理海量数据。

2.2.5 数据管理技术

云计算系统对大数据集进行处理、分析,向用户提供高效的服务。为此,数据管理技术必须能够高效地管理大数据集。其次,如何在规模巨大的数据中找到特定的数据,也是云计算数据管理技术所必须解决的问题。

在云计算应用中,最为常见的是谷歌的 BigTable 数据管理技术。由于采用列存储的方式管理数据,如何提高数据的更新速率和进一步提高随机读速率是未来的数据管理技术必须解决的问题。

谷歌提出的 BigTable 技术是建立在 GFS 和 MapReduce 上的一个大型分布式数据库。BigTable 实际上是一个很庞大的表,它的规模可以超过 1PB(1024TB),它将所有数据都作为对象进行处理。谷歌对 BigTable 给出了如下定义: BigTable 是一种为了管理结构化数据而设计的分布式存储系统,这些数据可以扩展到非常大的规模,如在数千台商用服务器上达到 PB 规模的数据。现在有很多的谷歌应用程序建立在 BigTable 上,如 Google Earth 等,而基于 BigTable 模型实现的 Hadoop HBase 也在越来越多的应用中发挥作用。

2.3 云交付模型

根据现在最常用且比较权威的 NIST 的定义,云计算主要分为三种交付模型,且这三种交付模型主要从用户体验的角度出发。

如图 2.2 所示,这三种交付模型分别是软件即服务(Software as a Service,SaaS)、平台即服务(Platform as a Service,PaaS)和基础设施即服务(Infrastructure as a Service,IaaS)。对普通用户而言,他们主要面对的是 SaaS 这种服务模式,而且几乎所有的云计算服务最终的呈现形式都是 SaaS。

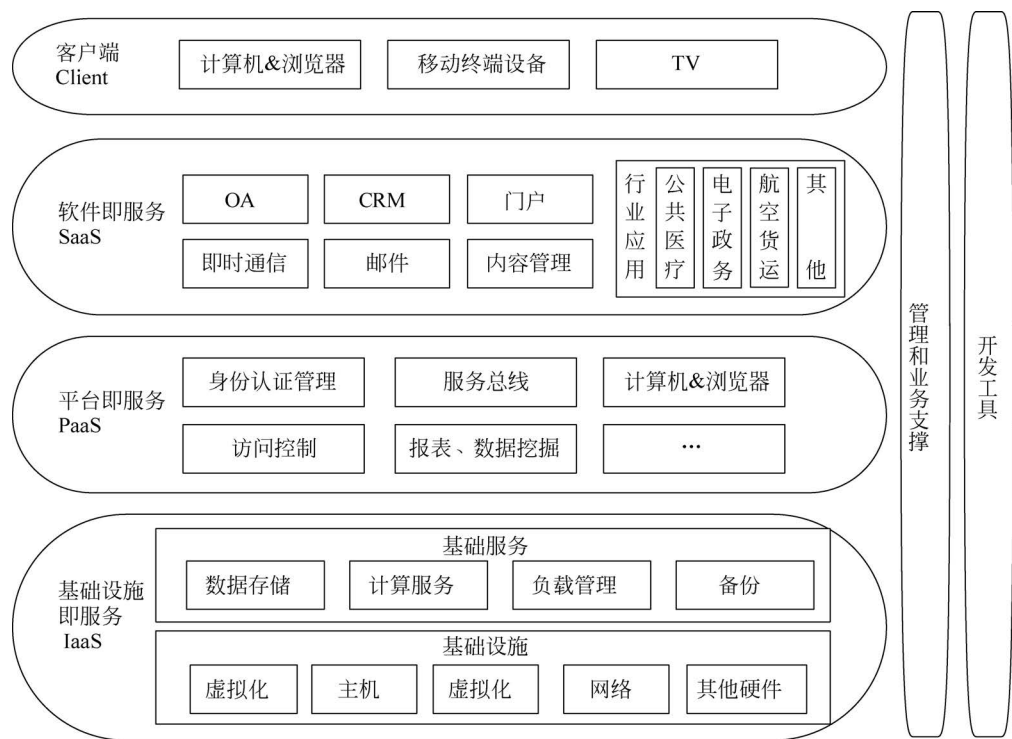


图 2.2 云计算三大交付模型

除此之外,人们还可能听到 DaaS、DBaaS、CaaS 等概念,这里认为所有 XaaS 都可以归为 IaaS、PaaS 和 SaaS 中的一种。例如,2.3.2 节将一种新型的交付模型 CaaS(容器即服务)作为开发平台的一部分。它是以容器为核心的公有云平台,被认为是具有革命性的云服务突破,可以看作 PaaS。

2.3.1 软件即服务

SaaS(软件即服务)是一种通过 Internet 提供软件的模式。用户无须购买软件,而是向提供商租用基于 Web 的软件,以此管理企业经营活动。相对于传统的软件,SaaS 解决方案具有明显的优势,包括前期成本较低、便于维护、展开使用迅速等。服务提供商维护和管理软件,并且提供软件运行的硬件设施,用户只需要拥有接入互联网的终端即可随时随地使用软件。因此,SaaS 软件被认为是云计算的典型应用之一。

SaaS 的主要功能如下。

- (1) 随时随地访问：在任何时间、任何地点,只要连接网络,用户就能访问 SaaS 服务。
- (2) 支持公开协议：SaaS 支持公开协议(如 HTML4/5),能够方便用户使用。
- (3) 安全保障：SaaS 供应商需要提供一定的安全机制,不仅应使存储在云端的用户数据处于绝对安全的环境,而且应在客户端实施一定的安全机制(如 HTTPS)来保护用户。
- (4) 多用户：通过多用户机制,不仅能更经济地支持庞大的用户规模,而且能提供一定的可指定性以满足用户的特殊需求。

用户消费的服务几乎完全是从网页(如 Netflix、MOG、Google Apps、Box. net、Dropbox)或苹果公司的 iCloud 进入的。尽管这些网页服务多用于商务和娱乐,但这也算是云技术的一部分。

常见的用于商务的 SaaS 应用包括 Citrix 公司的 GoToMeeting、Cisco 公司的 WebEx、Salesforce 公司的 CRM、ADP、Workday 和 SuccessFactors 等。

2.3.2 平台即服务

通过网络进行程序提供的服务称为 SaaS,而相应的,将服务器平台或开发环境作为服务进行提供就是 PaaS。所谓 PaaS 实际上是指将软件研发的平台作为一种服务,以 SaaS 的模式提交给用户。因此,PaaS 也是 SaaS 模式的一种应用。PaaS 的出现可以加快 SaaS 的发展,尤其是加快 SaaS 应用的开发速度。

在云计算应用的大环境下,PaaS 的优势显而易见,具体如下。

(1) 开发简单。开发人员能限定应用自带的操作系统、中间件和数据库等软件的版本(如 SLES 11、WAS 7 和 DB2 9.7 等),这样将非常有效地缩小开发和测试的范围,从而极大地降低开发测试的难度和复杂度。

(2) 部署简单。如果使用虚拟器件方式部署,则能将本来需要几天的工作缩短到几分钟,将操作精简到轻轻一击。此外,这样可以非常简单地应用部署或迁移到公有云上,以应对突发情况。

(3) 维护简单。因为整个虚拟器件都是来自同一个独立软件商(Independent Software Vendor,ISV),所以任何软件的升级和技术支持都只与一个 ISV 联系即可,不仅避免了常见的沟通不当现象,而且简化了相关流程。

PaaS 的主要功能如下。

(1) 良好的开发环境:通过 SDK 和 IDE 等工具使用户能在本地方便地进行应用的开发和测试。

(2) 丰富的服务:PaaS 平台会以 API 的形式将各种各样的服务提供给上层应用。

(3) 自动的资源调度:PaaS 具有可伸缩特性,不仅能优化系统资源,而且能自动调整资源,以帮助运行于其上的应用更好地应对突发流量。

(4) 精细的管理和监控:PaaS 能够提供应用层的管理和监控。例如,观察应用运行的情况和具体数值(如吞吐量和反应时间)以更好地衡量应用的运行状态,通过精确计量应用使用所消耗的资源以更好地计费。

PaaS 相关公司在网上提供各种开发和分发应用的解决方案,如虚拟服务器和操作系统等。这既节省了用户在硬件上的费用,也让分散的工作室之间的合作变得更加容易。这些解决方案包括网页应用管理、应用设计、应用虚拟主机、存储、安全和应用开发协作工具等。

常见的 PaaS 提供者包括 Google App Engine、Microsoft Azure、Force. com、Heroku、Engine Yard、AppFog、Mendix 和 Standing Cloud 等。

1. 服务平台交付

严格意义上,标准的 IaaS 提供的是虚拟实例,即虚拟机。用户申请的是一个干净的实

例或安装了某个软件的实例。实例开通后,用户需要在实例中安装软件或进行相应的配置,然后再将多个实例之间进行对接。这显然没有实现云计算开箱即用的服务理念。因此,服务平台交付(IaaS+)应运而生。

云平台提供一个典型的开发平台服务,该开发平台基于传统的应用架构,用户申请时可以直接生成相应的开发平台实例。用户不需要进行相关配置(如修改配置文件,让 Web 服务器指向另一个 DB 实例),而只需要关注在部署业务代码上。从某种角度讲,通过 IaaS+ 服务开通的也是开发平台,用户不需要关注平台本身,而只需要在开通出的机器上部署业务代码即可。因为是 IaaS 平台的延伸,因此可以算是 PaaS 的一种,属于 IaaS+ 方式实现。典型的过程是,云平台提供一个开发平台开通服务,该服务生成实例时可以自动部署 Web 节点服务器、中间件节点服务器和数据库节点服务器;当用户申请该服务时,服务可以自动生成,根据该业务开发平台特点按顺序安装软件并互相对接访问关系,同时对于该平台安装的软件做好相应的配置。

2. 无服务器架构

顾名思义,无服务器架构(Serverless)是指无服务器的架构。用户不需要了解底层的部署和配置,开发人员直接编写运行在云上的函数、功能、服务,由云平台提供操作系统、运行环境、网关等一系列的基础环境,这里只需要关注编写业务代码即可。过去在实现一段业务逻辑时,需要调用很多方法或函数后进行程序编写,为此需要安装操作系统、JDK 和 Tomcat 并进行大量配置和调试,而无服务器架构的目的只有一个——基于现有函数进行扩展实现业务。无服务器架构的理念就在于此,用户可以直接访问云平台上的服务,实现函数调用和程序编写。

到底什么是 Serverless 呢?无服务器架构是基于互联网的系统,其应用开发不使用常规的服务进程。相反地,它仅依赖第三方服务(如 AWS Lambda 服务)、客户端逻辑和服务托管远程过程调用的组合。Serverless 的主要特点如下。

(1) 在 Serverless 应用中,开发者只需要专注于业务,剩下的运维等工作都不需要关心。

(2) Serverless 是真正的按需使用,请求到来时才开始运行。

(3) Serverless 是按运行时间和内存来计算成本的。

(4) Serverless 应用严重依赖特定的云平台和第三方服务。

Serverless 中的服务或功能代表的只是微功能或微服务。Serverless 是一种思维方式的转变,从过去“构建一个框架运行在一台服务器上,对多个事件进行响应”变为“构建或使用一个微服务或微功能来响应一个事件”。用户可以使用 Django or Node.js 和 Express 等实现服务,但是 Serverless 本身超越这些框架概念,因此框架也变得不那么重要了。

3. 容器即服务

容器即服务(CaaS, Container as a Service)也称容器云,是以容器为资源分割和调度的基本单位,封装整个软件运行时环境,为开发者和系统管理员提供用于构建、发布和运行分布式应用的平台。CaaS 具备一套标准的镜像格式,可以将各种应用打包成统一的格式,并在任意平台之间部署迁移。容器服务之间可以通过地址、端口服务互相通信,有序且灵活,

既支持对应用的无限定制,又可以规范服务的交互和编排。

容器云的 Docker 容器几乎可以在任意的平台上运行,包括物理机、虚拟机、公有云、私有云、个人计算机、服务器等。容器云的兼容性类似于 Java 的 JVM,Java 程序可以运行在任意已安装 JVM 的设备上,便于迁移和扩展。

4. CaaS 与 IaaS 和 PaaS 的关系

作为后起之秀的 CaaS 介于 IaaS 和 PaaS 之间,起到屏蔽底层系统 IaaS、支撑并丰富上层应用平台 PaaS 的作用。

CaaS 解决了 IaaS 和 PaaS 的一些核心问题。例如,IaaS 很大程度上只提供机器和系统,需要自己把控资源的管理、分配和监控,没有减少使用成本,对各种业务应用的支持也非常有限;而 PaaS 的侧重点是提供对主流应用平台的支持,其没有统一的服务接口标准,不能满足个性化的需求。CaaS 的提出可谓是应运而生,以容器为中心的 CaaS 很好地将底层的 IaaS 封装成一个大的资源池。用户只需要将自己的应用部署到这个资源池中,不再需要关心资源的申请、管理,以及与业务开发无关的其他事情。

2.3.3 基础设施即服务

基础设施即服务(IaaS)使消费者可以通过 Internet 从完善的计算机基础设施获得服务。基于 Internet 的服务(如存储和数据库)是 IaaS 的一部分。在 IaaS 模式下,服务提供商将多台服务器组成的“云端”服务(包括内存、I/O 设备、存储和计算能力等)作为计量服务提供给用户。IaaS 的优点是用户只需要提供低成本硬件,按需租用相应的计算能力和存储能力。

IaaS 的主要功能如下。

- (1) 资源抽象:使用资源抽象的方法,更好地调度和管理物理资源。
- (2) 负载管理:通过负载管理,不仅使部署在基础设施上的应用能更好地应对突发情况,而且还能更好地利用系统资源。
- (3) 数据管理:对云计算而言,数据的完整性、可靠性和可管理性是对 IaaS 的基本要求。
- (4) 资源部署:将整个资源从创建到使用的流程自动化。
- (5) 安全管理:IaaS 安全管理的主要目标是保证基础设施和所提供资源被合法地访问和使用。
- (6) 计费管理:通过细致的计费管理使用户更灵活地使用资源。

几年前如果用户想在办公室或公司的网站上运行一些企业应用,则通常需要购买昂贵的服务器或其他硬件来控制本地应用。如果使用 IaaS,则用户可以将硬件外包到其他地方。IaaS 公司会提供场外服务器、存储和网络硬件,用户可以租用。这样就节省了维护成本和办公场地成本,并且可以随时利用这些硬件运行其应用。

常见的 IaaS 公司包括亚马逊、微软、VMware、Rackspace 和 Red Hat。这些公司各自都有自己的专长。例如,亚马逊和微软提供的不只是 IaaS,它们还会将计算能力出租给用户来管理其网站。

2.3.4 基本云交付模型的比较

三个交付模型之间没有必然的联系,只是三种不同的服务模式,都是基于互联网且按需按时付费。但是在实际的商业模式中,PaaS 的发展确实促进了 SaaS 的发展,因为 SaaS 的开发难度在获得开发平台的支持后就降低了。

从用户体验角度而言,它们之间的关系是独立的,因为它们面对的是不同的用户。从技术角度而言,它们并不是简单的继承关系,因为 SaaS 可以基于 PaaS 或直接部署在 IaaS 上,而 PaaS 可以构建在 IaaS 上或直接构建在物理资源上。

通过以上分析,表 2.1 对三种基本交付模型进行了比较。

表 2.1 三种交付模型的比较

云交付模型	服务对象	使用方式	关键技术	用户的控制等级	系统实例
IaaS	需要硬件资源 的用户	使用者上传数 据、程序代码、 环境配置	虚拟化技术、分 布式海量数据 存储等	使用和配置	Amazon EC2,Eucalyptus 等
PaaS	程 序 开 发者	使用者上传数 据、程序代码	云平台技术、数 据管理技术等	有限的管理	Google App Engine、 Microsoft Azure 等
SaaS	企业和需 要软件应 用的用户	使用 者 上 传 数据	Web 服务技术、 互联网应用开 发技术等	完全的管理	Google Apps,Salesforce CRM 等

这三种交付模式都是采用外包的方式,以此减轻云用户的负担,降低管理、维护服务器硬件、网络硬件、基础架构软件和应用软件的人力成本。从更高的层次上看,它们都试图解决同一个问题——用尽可能少(甚至为 0)的资本支出,尽可能多的获得功能、扩展能力、服务和商业价值。成功的 SaaS 和 IaaS 可以很容易地延伸到平台领域。

2.4 云部署模式

部署云计算服务的模式有三大类:公有云、私有云和混合云。如图 2.3 所示,公有云是云计算服务提供商为公众提供服务的云计算平台,理论上任何人都可以通过授权接入该平

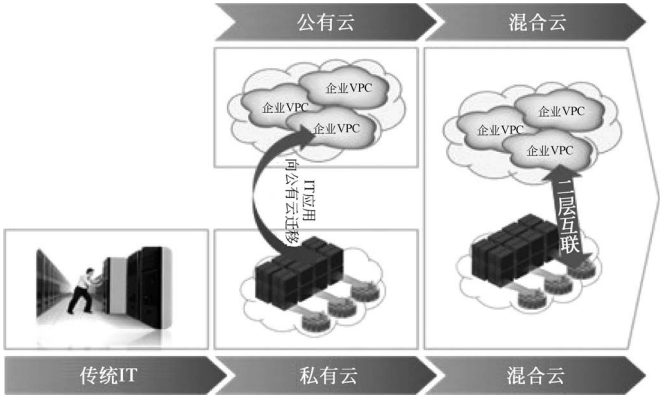


图 2.3 云部署模式

台。公有云可以充分发挥云计算系统的规模经济效益,但同时也增加了安全风险。私有云是云计算服务提供商为企业在其内部建设的专有云计算系统,存在于企业防火墙内,只为企业内部服务。与公有云相比,私有云的安全性更好,但管理复杂度也更高,其规模经济效益会受到限制,整个基础设施的利用率要远低于公有云。混合云是同时提供公有和私有服务的云计算系统,它是介于公有云和私有云之间的一种折中方案。

第三方评测机构曾经做过市场调查,发现公有云的使用成本在某些客户中会高于私有云,这往往与客户建设私有云所采用的厂商品牌有关。此外,公有云对外提供的服务是按月收取,大型部门有时可能无法及时准确洞察下属部门对公有云的使用量,因此会造成严重浪费。

2.4.1 公有云

公有云是指为外部客户提供服务的云,其服务供别人使用,而不供自己使用。

在该模式下,应用程序、资源、存储和其他服务都由云服务供应商提供给用户,这些服务多半都是付费的,也有部分出于推广和市场占有需要而提供免费服务。这种模式只能使用互联网进行访问和使用,且在私人信息和数据保护方面比较有保证。这种部署模型通常都可以提供可扩展的云服务并能高效设置。

目前,典型的公有云有微软的 Windows Azure Platform、亚马逊的 AWS、阿里巴巴的阿里云等。对于用户而言,公有云的最大优点是其所应用的程序、服务及相关数据都存放在公有云的提供者处,自己无须进行相应的投资和建设。目前最大的问题是数据不存储在用户自己的数据中心,其安全性存在一定风险。同时,公有云的可用性不受使用者控制,这方面也存在不确定性。

2.4.2 私有云

私有云是指企业自己使用的云,其服务不供别人使用,而供自己内部人员或分支机构使用。这种云基础设施专门为某一个企业服务,与自己管理或第三方管理、自己负责或第三方托管等都没有关系。

私有云的部署比较适合有众多分支机构的大型企业或政府部门。随着这些大型企业数据中心的集中化,私有云将会成为他们部署 IT 系统的主流模式。相对于公共云,私有云部署在企业自身内部,因此其数据安全性、系统可用性都可由自己控制。但其缺点是投资较大,尤其是一次性的建设投资较大。

2.4.3 混合云

混合云是指供自己和客户共同使用的云,其服务既可以供别人使用,也可以供自己使用。

混合云是两种或两种以上的云计算模式的混合体,如公有云和私有云混合。它们相互独立但在云的内部又相互结合,可以发挥出所混合的多种云计算模型的各自优势。相对而言,混合云的部署方式对提供者的要求较高。