

版权申明

版权归Esri所有

由Esri中国（北京）有限公司翻译印制

在该资料中所包含的内容均属于Esri财产。此著作受到美国版权法和其他国际版权条约及惯例所保护。不得以任何方式、任何文字或图片作全部和局部变更、发行、复制、转载、引用除非注明由Esri编写。所有疑问都可反馈到Esri的合同及法律服务管理部，地址：380 New York Street, Redland, CA 92373—8100USA

该资料如有变更恕不另行通知

Esri、the Esri globe logo、ArcGIS、ArcInfo、ADF、ArcEditor、ArcView、ArcPad、ArcWeb、社区StreetMap、ArcIMS、ArcMap、ArcXML、ArcObjects、ArcSDE、JTX、ArcExplorer、ArcGlobe、ArcToolbox、ArcCatalog、ModelBuilder、SDE、Spatial Database Engine、www.esri.com和@esri.com是Esri在美国、欧盟和其他地区的商标、注册商标或服务标志。本资料提及的其他的公司和产品都已由各自的商标拥有者注册。

目 录

1.概述.....	1
2.什么是云计算.....	2
3.Esri的云计算支持	10
4. “云端部署” 案例.....	18
5.ArcGIS的云安全	26
附录1：云计算演变.....	29
附录2：剖析云计算技术.....	31

1. 概述

一说到云计算，可能大家都会一头雾水，都会存在一个疑问，什么是云计算？这里我们做个简单的解释。从事企业IT建设的人员都知道，我们花了大量时间做一些基础架构的维护、软件的维护、运维的日常的工作，这大概占了70%以上的时间，我们大概只有20%甚至更少的时间去想我们的IT结构、IT组织，怎么满足业务的需求。今天我们发觉面临的压力越来越多的来自业务部门，怎样快速满足业务部门的需求，怎么改变目前的窘迫的状况，快速适应IT快速变化的需求。云计算应该是一个很好的利器，解决目前的状况。我们说的云计算实际上说的简单一点，说成类似今天的电话，类似墙上的插座，或者类似自来水的水龙头，我们只需要拧开水龙头就可以有自来水。具体说水来自长江还是黄河，插座的电是核电厂还是火电厂的发电，都无所谓，知道不知道对我影响不大。我们的IT架构如何走到这样的一个程度？云计算就是很好的解决和处理的平台去达到这样的目标。

一般来说，企业或机构都会管理和维护自己的计算机系统。来自IBM的一项调查表明，每年企业花在硬件投资和维护上的费用越来越高，如下图所示：



图1-1 全球年度IT硬件费用投资

（图注：该图内容来自因特网）

随之而来，大家想着各种办法来解决如何把系统建设的更加容易、更加节省。因此各种新概念层出不穷，虚拟化技术、集群计算、网格计算、效用计算等。谷歌、亚马逊、微软、IBM、雅虎等多家厂商不断在云计算技术上

推陈出新，推进了云计算时代的来临。可以说，云计算是并行计算、分布式计算和网格计算的发展，或者说这些计算机科学概念的商业实现。也是虚拟化、效用计算、IaaS、SaaS、PaaS等概念混合演进并跃升的结果。

随着云计算时代的到来，目前已有多种软件部署方法的选择，可以大大的减少企业或政府机构的大量的时间，金钱，和IT资源的需求，并带来大量的效益，例如降低费用预算，提高客户满意度等。本篇文档提供了一个云计算的概览，并探讨一些ArcGIS Server在云计算中的应用。

2. 什么是云计算

(以下内容部分摘自百度百科; <http://baike.baidu.com/view/1316082.htm>)

2.1 基本概念

“云计算”的概念起源于大规模分布式计算技术。是并行计算 (Parallel Computing)、分布式计算 (Distributed Computing) 和网格计算 (Grid Computing) 的发展。云计算是虚拟化 (Virtualization)、效用计算 (Utility Computing)、IaaS(基础设施即服务)、PaaS(平台即服务)、SaaS (软件即服务) 等概念混合演进并跃升的结果。有人把云计算比喻为：这好比是从古老的单台发电机模式转向了电厂集中供电的模式。它意味着计算能力也可以作为一种商品流行流通，就像煤气、水电一样，取用方便，费用低廉。最大的不同在于，它是通过互联网进行传输的。

云计算是分布式计算技术的一种，是透过网络将庞大的计算处理程序自动分拆成无数个较小的子程序，再交由多部服务器所组成的庞大系统经搜索、计算分析之后将处理结果回传给用户。透过这项技术，网络服务提供者可以在数秒之内，达成处理数以千万计甚至亿计的信息，达到和“超级”计算机同样强大的网络服务。

2.2 云计算特点

(1) **超大规模。**“云”具有相当的规模，Google云计算已经拥有100多万台服务器，Amazon、IBM、微软、Yahoo等的“云”均拥有几十万台服务器。企业私有云一般拥有数百上千台服务器。“云”能赋予用户前所未有的计算能力。

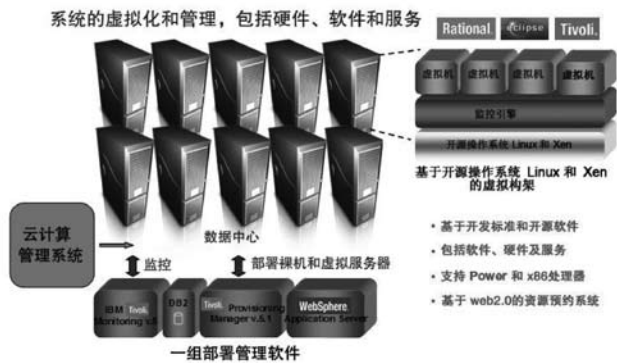


图2-1 云计算系统

(2) **虚拟化。**云计算支持用户在任意位置、使用各种终端获取应用服务。所请求的资源来自“云”，而不是固定的有形的实体。应用在“云”中某处运行，但实际上用户无需了解、也不用担心应用运行的具体位置。只需要一台笔记本或者一个手机，就可以通过网络服务来实现我们需要的一切，甚至包括超级计算这样的任务。

(3) **高可靠性。**“云”使用了数据多副本容错、计算节点同构可互换等措施来保障服务的高可靠性，使用云计算比使用本地计算机可靠。

(4) **通用性。**云计算不针对特定的应用，在“云”的支撑下可以构造出千变万化的应用，同一个“云”可以同时支撑不同的应用运行。

(5) **高可扩展性。**“云”的规模可以动态伸缩，满足应用和用户规模增长的需要。

(6) **按需服务。**“云”是一个庞大的资源池，你按需购买；云可以象自来水、电、煤气那样计费。

(7) **极其廉价。**由于“云”的特殊容错措施可以采用极其廉价的节点来构成云，“云”的自动化集中式管理使大量企业无需负担日益高昂的数据中心管理成本，“云”的通用性使资源的利用率较之传统系统大幅提升，因此用户可以充分享受“云”的低成本优势，经常只要花费几百美元、几天时间就能完成以前需要数万美元、数月时间才能完成的任务。

云计算可以彻底改变人们未来的生活，但同时也要重视环境问题，这样才能真正为人类进步做贡献，而不是简单的技术提升。

(8) **潜在的危险性：**云计算服务除了提供计算服务外，还必然提供了存储服务。但是云计算服务当前垄断在私人机构（企业）手中，而他们仅仅能够提供商业信用。对于政府机构、商业机构（特别象银行这样持有敏感数据的商业机构）对于选择云计算服务应保持足够的警惕。一旦商业用户大规模

使用私人机构提供的云计算服务，无论其技术优势有多强，都不可避免地让这些私人机构以“数据（信息）”的重要性挟制整个社会。对于信息社会而言，“信息”是至关重要的。另一方面，云计算中的数据对于数据所有者以外的其他用户云计算用户是保密的，但是对于提供云计算的商业机构而言确实毫无秘密可言。这就象常人不能监听别人的电话，但是在电讯公司内部，他们可以随时监听任何电话。所有这些潜在的危险，是商业机构和政府机构选择云计算服务、特别是国外机构提供的云计算服务时，不得不考虑的一个重要的前提。

2.3 判断云计算标准

到底什么是云计算？这是大家比较关注的一个问题。现在我们发现了有很多种不同的说法，到底什么是云，什么不是云，让人很费解。有人讲公有云是云，私有云不是云；还有人说支持虚拟化叫云，不支持虚拟化不叫云；还有人讲有1000台服务器是云，好像999台就不是云。现在有个别高性能计算中心，什么都没变，就是名字改成叫云计算中心啦。为什么有那么多人要打着云计算的旗号呢？要么赚眼球，要么圈钱，各有各的目的。

作为公众，需要鉴别哪些是真云，哪些是假云。为此，刘鹏（中国电子学会云计算专家委员会委员、解放军理工大学教授）提出判断是不是云计算的三条参考标准，供大家参考：

第一条：用户所需的资源不在客户端而来自网络。这是云计算的根本理念所在，即通过网络提供用户所需的计算力、存储空间、软件功能和信息服务等。

第二条：服务能力具有分钟级或秒级的伸缩能力。如果资源节点服务能力不够，但是网络流量上来，这时候需要平台在一分或几分钟之内，自动地动态增加服务节点的数量，从100个节点扩展到150个节点。能够称之为云计算，就需要足够的资源来应对网络的尖峰流量，哪怕是突然出现了“贾君鹏”这样的热点。过了一阵子，流量下来了，服务节点的数量在随着流量的减少而减少。现在有的传统IDC自称也能提供伸缩能力，但需要多个小时之后才能提供给用户。问题是网络流量是不可预期的，不可能等那么久。

第三条：具有较之传统模式5倍以上的性能价格比优势。看了上面一条，有些人在想，没关系，多配一些机器，流量再大也应付得了。但这不是云计算的理念。我们还有性能价格比指标。云计算之所以是一种划时代的技术，就是因为它将数量庞大的廉价计算机放进资源池中，用软件容错来降低硬件成本，通过将云计算设施部署在寒冷和电力资源丰富的地区来节省电力成本，通过规模化的共享使用来提高资源利用率。国外代表性云计算平台提供商达到了惊人的10-40倍的性能价格比提升。国内由于技术、规模和统

一电价等问题，暂时难以达到同等的性能价格比，我们暂时将这个指标定为5倍。拥有256个节点的中国移动研究院的云计算平台已经达到了5到7倍的性能价格比提升，其性能价格比随着规模和利用率的提升还有提升空间。

这三条标准相当于三张滤网，可以一层层地滤掉那些不属于云计算范畴的东西。让我们来试试灵不灵——来看看下面这些是不是云计算：

PC系统：第一关过不了，因用户所需的资源都在客户端，所以不属于云计算范畴；

iPhone应用软件：如果下载到iPhone上就能独立运行，与外界只是通信关系，则过不了第一关；如果是依托于网络平台运行，计算和数据资源来自网络，iPhone端只是个访问界面，则有可能进入云计算范畴；

Web网站：过得了第一关。但如果还过得了可伸缩性这一关，会很难过性能价格比这一关；

上网本：能过第一关，但它本身只是个终端，如果只是用于上网，则在传统范畴，如果用于与云计算设施交互，只可划入云计算范畴。

广告联盟：难过第二关。将不同网站的广告组织在一起，就构成了广告联盟。发布一个广告，可以出现在众多的网站上。不过，由于公众对广告的点击率不高，广告联盟不需要有太多的可伸缩性，也不容易做到——因为参与者都是众多的小网站。

超级计算中心：如果规模够大，勉强能过第二关，但性价比不好，过不了第三关。超级计算机都是很昂贵的，在性价比上没有优势。我们认为不是云。

在线Office：规模化运作后容易达到云计算的要求。例如，今天大家看到的百汇Office，还有Google的DOCS也是。

在线CRM：规模化运作后容易达到云计算的要求。提供给用户一个月使用成本是几十块钱，这个远远优于我们使用传统模式，是云。

IaaS：像亚马逊租用机器的服务，所以这种性价比也非常好，租用一个虚拟机1小时只要0.1美元，也是云。

PaaS：如微软的Azure，可免费提供25GB的存储空间，必定有非常好的性价比，也是云。

云安全：规模化运作后容易达到云计算的要求。比如360安全卫士，提供给用户完全免费的服务，之所以能够这样，是因为有很高的性价比。

2.4 什么是云服务

尽管在云计算环境中已经有多种解决方案的方式，云选择包含至少三层的IT技术架构：

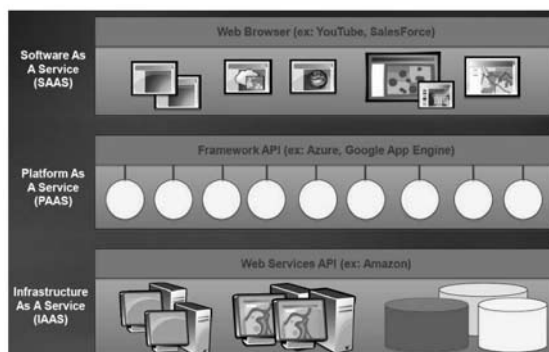


图2-2 云计算架构

软件即服务 (SaaS)： 是通过网页浏览器把程序和功能传给成千上万的用户。实例包括：Salesforce.com, Esri的Business Analyst Online (BAO), ArcGIS Online Sharing, 以及GIS Portal Toolkit。

平台即服务 (PaaS)： 即大家所熟知的“云件”或“Cloudware”，PaaS能够将私人电脑中的资源转移至网络云，是SaaS的延伸，这种形式的服务把开发环境作为一种服务来提供。允许开发者进行创建、测试和部署应用，即使用中间商的设备来开发自己的程序并通过互联网和其服务器传到用户手中。实例包括：ArcGIS Online共享的REST API和ArcGIS Web Mapping APIs。

基础设施即服务 (IaaS)： 由计算机架构如虚拟化组成，并作为服务实现为用户提供。基于Internet的服务（如存储和数据库）是IaaS的一部分。IaaS提供了动态和高效的部署架构，IaaS例子中大家所熟知的有Amazon Simple Storage Service (S3), Amazon Elastic Cloud Compute (EC3)即弹性云计算，Akamai，以及ArcGIS Online Data Centers等。

下面简要该素一下IaaS的实现机制，如下图所示：

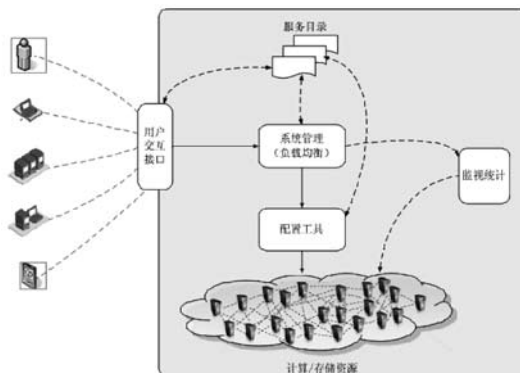


图2-3 IaaS机制

用户交互接口向应用以Web Services方式提供访问接口，获取用户需求。服务目录是用户可以访问的服务清单。系统管理模块负责管理和分配所有可用的资源，其核心是负载均衡。配置工具负责在分配的节点上准备任务运行环境。监视统计模块负责监视节点的运行状态，并完成用户使用节点情况的统计。执行过程并不复杂：用户交互接口允许用户从目录中选取并调用一个服务。该请求传递给系统管理模块后，它将为用户分配恰当的资源，然后调用配置工具来为用户准备运行环境。

2.5 云类型

云计算可以有三种部署模式，即公共云、私有云和混合云。

私有云：私有云是为一个客户单独使用而构建的，因而提供对数据、安全性和服务质量的最有效控制。那么虚拟私有云是什么概念？对于企业应用来说，在这中间可能跨内部云、外部云，也可能是自己建立的有几个数据中心。比如说你的企业上海、北京、广州都有数据中心，那去跨这些数据中心形成的虚拟私有云，这个云是一个逻辑上是一个整体，但物理上跨很多数据中心，这就类似于今天在网络里看到的VPN概念。

私有云可部署在企业数据中心的防火墙内，也可以将它们部署在一个安全的主机托管场所。私有云可由公司自己的IT机构，也可由云提供商进行构建。下图是一个典型私有云的框架图：



图2-4 私有云框架图

私有云的建设关键点在于它是一种动态、易扩展的虚拟化计算服务，从层次上来看，包括如下：

- 基础设施即服务 (Infrastructure as a Service, IaaS)

- 以服务的形式提供虚拟硬件资源, 如虚拟主机/存储/网络/安全等资源。

- 用户无需购买服务器、网络设备、存储设备, 只需通过互联网租赁即可搭建自己的应用系统

- 平台即服务 (Platform as a Service, PaaS)

- 提供应用服务引擎, 如互联网应用程序接口/运行平台等。

- 用户基于该应用服务引擎, 可以构建该类应用。

- 软件即服务 (Software as a Service, SaaS)

- 用户通过标准的Web浏览器来使用Internet上的软件。

- 客户端软件通过标准的Web服务来使用Internet上的服务

- 用户不必购买软件, 只需按需租用软件

公共云: 指为外部客户提供服务的云, 它所有的服务是供别人使用, 而不是自己用。云服务遍布整个因特网, 能够服务于几乎不限数量的拥有相同基本架构的客户。如亚马逊、Rackspace、Salesforce.com、微软、Google等推出的公共云产品。

混合云: 指供自己和客户共同使用的云, 它所提供的服务既可以供别人使用, 也可以供自己使用。混合云表现为多种云配置的组合, 数个云以某种方式整合在一起。例如有时用户可能需要用一套单独的证书访问多个云, 有时数据可能需要在多个云之间流动, 或者某个私有云的应用可能需要临时使用公共云的资源。



图2-5 云类型

2.6 云计算的效益

云计算部署模型为企业机构提供多种效益:

(1)、简化应用程序部署与管理。

对于企业或政府机构部门应用GIS来说，最头疼的不是怎么使用功能，而是在于怎么搭建和维护GIS的资源。我们必须关注的不仅仅是怎么采集数据、处理数据、构建数据库、制图、发布地图服务，更重要的是，要考虑怎么部署GIS软件和优化软件系统配置，才能达到最优的性能，这占到了一个GIS系统应用建设的80%的精力和资源，将来还得考虑系统升级和维护。而采用云计算，大大的简化GIS服务器的部署，减少复杂的服务器管理。我们可以通过增加或减少ArcGIS Server进程的数量来快速满足负载的需求，而不需要投资新的硬件。用户在几分钟之内可以直接访问ArcGIS Server，并可以开始发布服务，并直接支持Web制图应用。

(2)、降低投资和运营成本

我们知道，GIS系统建设的管家在于数据，而且系统建设的最大的成本也在于数据。采用云计算模式，可以集中统一的维护数据，通过共享的方式为所有的客户端提供数据服务和软件服务。

公共云服务：ArcGIS 的Online上提供了公共的云地图服务，使用者无需关注数据如何采集、更新或维护，也无需购买数据。如果需要采用公共云服务，只需要按照流量的方式支持一定的费用即可，大大节省了成本。

企业私有云：我们不再需要为每个项目、每个下属单位重复的购买数据，服务器，甚至人力成本。通过企业私有云，只需要构建一个数据中心，集中统一维护，企业集团或政府机构部门内部都可以共享该私有云服务。硬件资源、人力资源、维护资源都可以从成本链中减少。

下图为一个使用传统数据中心与采用Amazon云计算Web服务方式的费用与满意度的比较：

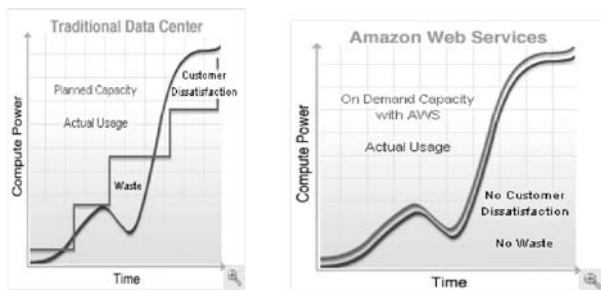


图2-6 云计算费用与满意度比较

根据不完全的统计，云计算能够带来更低的成本和管理费用，而大大提高利用率，因此综合来说可以比现有系统提高30倍的价值。



图2-7 云计算效率

(3)、缩短建设周期

对于一个传统的系统建设来说，需要一个基本的周期，如下：

- a. 硬件维护
- b. 硬件安装
- c. IT安装
- d. 软件安装
- e. 应用和数据装载
- f. 测试

这其中a—d的四个步骤过程都将运行在云端，即无需再重复考虑做这些工作之后，我们的系统建设周期将大大缩短。

(4)、提高应用程序和基础设施的灵活性。

我们在建设系统中，采用云计算模式，可以把精力更多的关注在新功能的开发和集成上，而不需要关注基础架构上。可以把应用程序功能扩展到互联网中实现。在需要日常访问和应答高峰期流量访问时候，只需要动态、弹性增加或移除硬件设备就可以实现，而无需增加重新部署或编码的工作量。达到程序与基础设施灵活应用的效果。

3. Esri的云计算支持

尽管ArcGIS Server尚未完成达到成熟的云环境，但是ArcGIS Server提供的是一个按需架构的可用组件。你可以将你的缓存地图切片上传到效用计算供应商那，例如亚马逊的S3，在云端创建数据中心。在Web ADF和至少小修改的ArcGIS API就可以直接使用地图服务。

●运行在 RHEL5上的ArcGIS Server的Java平台可以直接为亚马逊弹性云计算（EC2）来配置。但是无论如何这都不重要，只是推荐高级Server使用的个例。这个配置只能通过专业服务合同（Professional Services）来支持。

●由于是通过Internet连接来访问ArcGIS Server服务，因此Web ADF应用可以配置到云端。这个配置只能通过专业服务合同（professional Services）来支持。

●现今的部署还不能够完全利用上动态群集的进程，或亚马逊云平台所提供的所有优势。

●将来，ArcGIS Server的Java平台架构设计将遵循这样的方式：即有效的运行服务，并充分的利用亚马逊EC2的云架构所提供的优势。在不中断服务情况下，服务器进程也可以从EC2上动态、实时的添加或移除。

●ArcGIS Server的安装将为AMI/EC2进行优化

●专门设计云/可视化环境的管理

●为ArcGIS Server提供共享配置、数据等的机制，特别是为亚马逊EC2，S3设计。

●ArcGIS Server平台将更加适应EC2，从JAVA架构到Linux操作系统运行时的支持。将为亚马逊映像准备。

●ArcGIS Server产品历史和云历史现在将在群集进程内同步，并与云端系统一致，如Server objects Extensions的热点部署，“实用”软件对象等。

3.1 ArcGIS云计算部署模型

目前有多种云计算部署场景。美国国家标准技术研究院（NIST）目前正在采用事实上的云计算和分布式模型，如下Esri的例子。

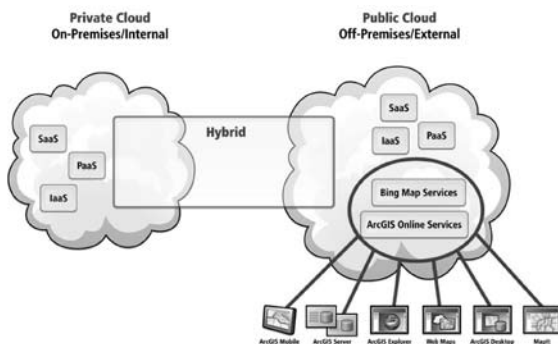


图3-1 ArcGIS 云计算模型

图：ArcGIS云计算部署模型：一些机构考虑到安全问题，选择私有云或“混合云”的部署模型

公共云是最通用的云计算，基础设施和应用都由提供云服务的机构维护。当然由于许多的传统提供商和用户目前还不想使用公共云计算，因此云服务层只能在私有云环境中复制使用。

3.2 ArcGIS云计算产品

ArcGIS产品提供了全系列的软件体系，对于云计算架构来说，我们考虑到基础设施、平台、应用程序，都可以在公共云上实现。用户端可以通过多种方式来访问云端的服务，如下图所示：

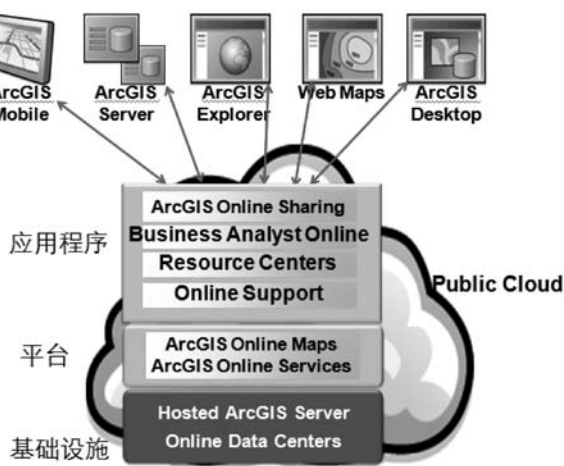


图3-2 ArcGIS公共云架构

Esri提供了云计算产品的解决方案，主要包括以下：

(1)、IaaS：云端地图切片，缓存的地图切片可以上传到云端，并在云端建立数据中心。具体参见下一节的内容。

(2)、SaaS：Esri针对SaaS目前提供了Esri Business Analyst Online，允许用户将GIS技术结合整个美国的大量的统计专题、消费者数据，以及商业数据。这可以将按需分析、报表和地图通过Web进行传递。因为Esri维护Business Analyst Online，用户不需要担心数据管理和技术更新。

(3)、PaaS：将来，Esri开发人员将此内容和功能扩展至Esri的 PaaS上，并通过ArcGIS Web Mapping APIS，例如JavaScript，Flex，Silverlight/WPF等来提供，并在ArcGIS Online中管理。

(4)、软件加服务（S+S）：ESRI已经提供了软件加服务的模式，可以让用户按需配置所需要的服务。Esri的ArcGIS Online Map和GIS Services提供S+S的用户可以快速访问制图设计，无缝的基础底图，并可以添加用户自己的数据到Esri的按需配置产品上。MapIt是另一个软件加服务的应用，可以让业务信息通过访问Esri和Bing Maps的在线数据、基础底图、和任务服务，来进行显示和更加精确的分析，并支持Windows Azure平台和 Microsoft的SQL Azure。作为一个社区云，ArcGIS的在线内容共享项目可以让用户或组织享受公共云的地理数据内容。亚马逊的EC2和S3计算和存储服务，可以让ESRI进行7*24小时的访问和维护内容。

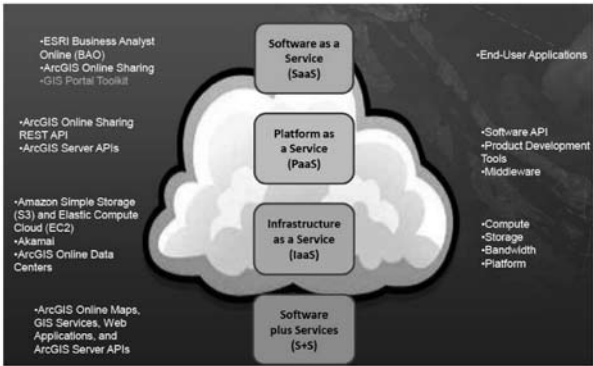


图3-3 Esri云产品解决方案

3.2.1 云端缓存

对于构建GIS系统来说，众所周知最重要的是数据。随着网络技术的发展，地图缓存已经成为提高地图服务访问性能的一个重要技术手段。那么缓存地图是否可以部署到云端或是否支持通过云端访问缓存地图，缓存地图与动态地图是否可以无缝结合应用，则是目前GIS云计算建设的重要因素。

Esri的合作伙伴ROK技术公司（Roktech）建立FltPlan.com网站来服务与飞机的跟踪地图服务，Roktech使用ArcGIS Server缓存切片和相对新的云计算类型服务HaaS（硬件即服务）。

该应用结合动态与缓存切片地图的结合，大大的提高了地图显示的性能，并且切片可以来自多个云端的地图服务，如下图：



图3-4 云端地图服务来源

可以看到不同的地图切片来自不同的切片服务器云上，并且结合了动态的地图服务内容。该应用的架构图如下：

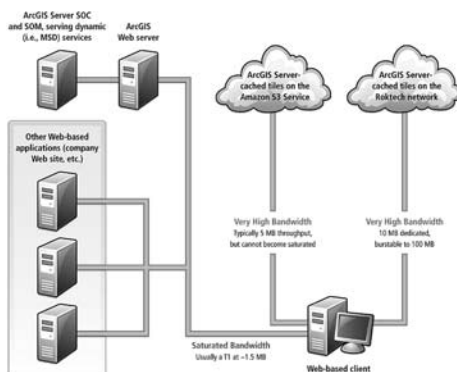


图3-5 云端缓存架构

(关于此应用更多内容请访问网址：<http://www.esri.com/news/arcuser/1009/cloudcomputing.html>)

Esri的ArcGIS Server的地图缓存和Globe三维地图服务缓存，目前都已经可以实现在Amazon上进行创建和存储。对于小的缓存来说，你不需要注意缓存创建的步骤，可以直接相对简单的方式传到云上。对于大的缓存来说，你就需要确定是否需要转移预先创建好的缓存到云端，还是在云端来临时创建。我们以Amazon为例子说明，因为移动文件到Amazon上不会和通过网络从一台机器拷贝到另外一台机器上那样，那么容易实现。因为你不仅需要确定文件传输的方法，你还需要保证敏感数据传送过程中的安全性。而且传输的速度受Internet连接的限制，需要花费更加大量的时间。

所以当你有特别大量的缓存地图需要传送到Amazon上，可以采用AWS Import/Export工具来实现。

3.2.2 ArcGIS Online

ArcGIS Online是Esri公司的在线的公共资源网站，提供了一站式的地图云服务。通过ArcGIS Online，你可以查找，共享和组织地理内容来建立GIS应用。我们可以通过ArcGIS.com，访问Esri和其它GIS用户所发布的地图，应用和工具，并可共享地图内容。通过ArcGIS Online你可以：

- 访问免费，高质量的基础地图服务，来应用到GIS项目中
- 创建和加入工作组，可以将地图内容进行共享和协作
- 快速开发Web应用。通过使用ArcGIS Web Mapping API和已有的地图内容，可以迅速搭建地图应用。

3.2.3 ArcGIS Sharing

ArcGIS Sharing是ArcGIS的社区协作模式，可以实现：

- 浏览和使用Esri和ArcGIS用户发布的地图
- 上传地图和注册ArcGIS Server服务
- 共享地图和数据
- 创建在线地图

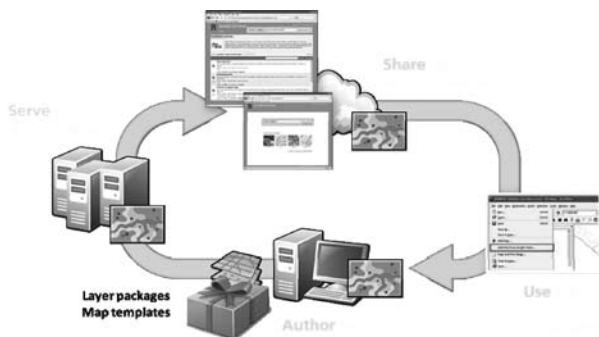


图3-6 ArcGIS sharing 社区

3.2.4 ArcGIS Web Mapping APIs

ArcGIS Web Mapping API包括Flex JavaScript和Silverlight，可以用来创建富互联网应用。开发人员可以免费的访问ArcGIS Web制图内容，包括在线帮助系统、API内容、代码库、示例代码和配置应用模板。只用制图应用开发API，你可以访问标准的ArcGIS Online地图和任务服务。这些任务服务包括地理编码服务、路径分析服务、专题统计数据分析和报表服务等。

3.2.5 ArcGIS Business Analyst Online

ArcGIS Business Analyst Online是Esri所提供一个基于SaaS的解决方案。是基于Flex API之上开发的云中应用程序，提供了软件即服务的内容。Esri的商业云架构如下图所示：

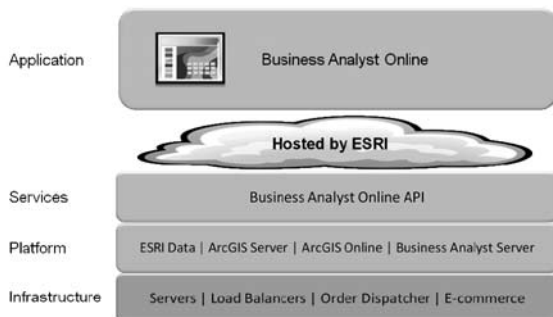


图3-7 ArcGIS Business Analyst 的 SaaS 方案

Business Analyst Online提供的API可以用来定制包含统计数据报表能力的商业Web应用。可以用于统计人口和位置信息，按照邮政编码、指定区域、驾车时间、或距离等来进行统计。报表的内容可以通过PDF，Excel表格或XML流来提供。结合数据与分析能力，可以实现新的商业点的选址分析。

3.2.6 ArcLogistics Online

ArcLogistics是Esri提供的解决物流配送车辆线路优化的解决方案软件。ArcLogistics可以实现多点、多线路的线路配送，并充分考虑车辆的使用情况，如车辆容积率、车辆成本、配送时间、动态分配调节配送、配送区域规划问题等，并解决从配送规划线路到车辆配送导航问题的整体解决方案。如下图所示：

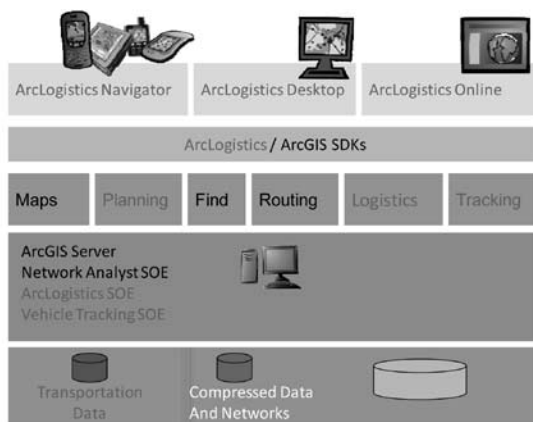


图 3-8 ArcLogistics Online 架构

ArcLogistics Online是Esri除了部署桌面产品之外所提供的的一个富客户端软件解决方案，直接使用ArcGIS Online所提供的地图、地理编码和路径服务。这些服务通过云端服务提供，并作为SaaS模型的支持客户端软件。Esri还提供另外一种部署模式，与上面的部署模式类似，也是一个富客户端软件解决方案，但是是由用户采用ArcGIS Server的服务所发布出来的自由的地图服务、地理编码服务和路径服务（需要有ArcGIS Server的网络扩展模块）。

ArcLogistics Online将作为一种SaaS的模式提供的服务，由以下优势：

- 可伸缩，基于用户界面，比较符合用户现有业务流程。容易学习和使用。对于使用客户端软件无需GIS经验和知识就可以共享ArcLogistics技术。重点是可以快速优化路径，并对现有业务处理流程进行快速建模；

- 简化软件部署。客户端软件可以快速下载和安装。所支持的地图服

务、地理编码服务和路径服务可以由ArcGIS Online或客户自由发布的服务。如果客户选择使用ArcGIS Online的云服务，则无需管理空间数据库。并且可以自动更新数据和软件，减少软件维护和数据维护费用；

- License费用根据车队的数量确定。License由参与计算线路的车辆的数量确定。

- 和桌面部署类似，客户端软件也可以定制和扩展来满足特定的需求，例如可以和企业应用和数据源集成。应用框架可以根据不同工作流程，语言或纵向市场情况来配置。

3.3 “云计算” 资源需求

既然云计算是将多资源的硬件串起利用起来，那么许多机构都有一个疑虑，云计算到底需要多少服务器资源？以下是一个可供参考的实例，澳大利亚Esri公司实施了布里斯班的格拉德斯顿游艇赛跑跟踪网站。根据统计，网站接收的点击率超过了480万，在2010年，2天半的时间，网站点击率超过了1450万点击率。而应对目前这种大量的访问方式，采用多机器负载均衡的环境模式：

- 1x3core Web Server—这台机器用来响应用户所访问的Flex API的应用
- 3x8core ArcGIS Server(负载均衡)—这些机器用来响应Flex应用所访问的所有的ArcGIS Server的地图服务
- 1x8core数据库服务器—这台机器用来支持ArcGIS Server应用的Geodatabase数据库。

经过实际运行，以上机器没有超过40%的负荷，在高峰期，超过1000个的独立进程运行。

3.4 Esri云计算 计划

Esri将提供ArcGIS Server企业级高级版V10，并提供一些扩展，可运行在AWS中。当然，对于ArcGIS Server9.3.1的用户也可以运行在云中，但需要Esri的专业服务部门（PS）来帮助运行和部署。

Esri已经通过ArcGIS Online服务提供了在线的地图和数据，包括去年与微软协作所提供商业分许可软件的MapIt服务。Esri同时提供自由的“Business Analyst Online”来共享软件本身的统计报表和制图服务。

Esri的总裁Jack Dangermond在接受计算机世界采访中，下一步我们计划将不仅仅把数据存储到云端，同样可以让用户把ArcGIS Server产品运行在云端。

Esri目前正接受计划在云计算平台上实现产品的应用，包括将来使用Google App Engine来实现一些应用。将来也将支持ArcGIS Server的一些组件在Windows Azure上实现。

4. “云端部署” 案例

目前市场上多家公司都提供了对于云计算部署应用程序的解决方案，而ArcGIS软件也可以部署在多种云平台之上，下面举几个例子来说明：

4.1 Applogic部署

4.1.1 3Tera公司简介

3Tera公司是全球最大的IT管理软件供应商之一。作为云计算领域的领导企业，3Tera的AppLogic采用直观的图形用户界面为共有云及私有云环境提供创新的解决方案，以帮助客户提供云服务，并部署复杂的企业级应用软件。AppLogic可将符合应用程序的设计和部署简化为云中单一逻辑实体。通过统一应用程序的设置，应用程序部署，以及利用虚拟服务器构造功能，AppLogic能帮助降低成本，提高生产率并增加服务质量。

AppLogic是一个云计算平台，可以运行分布式应用。AppLogic可以打包整个N层应用或服务成一个逻辑上的整体，并作为单一系统管理。AppLogic可以在可视化浏览器中方便的组织，部署，监控，控制系统，如下图所示：

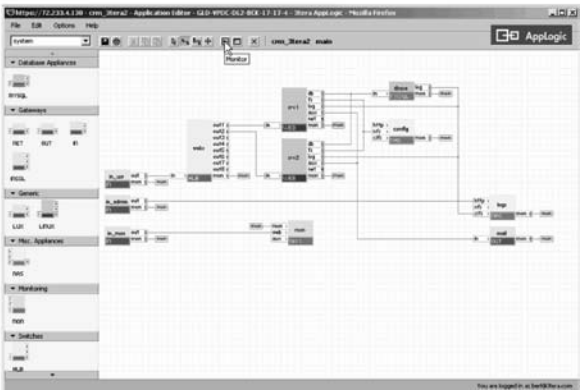


图4-1 AppLogic云计算的可视化系统

(更多关于AppLogic信息请参考：<http://www.3tera.com/AppLogic/index.php>)

4.1.2 ArcGIS Server在AppLogic上部署

(以下部分内容来自：<http://www.skygoneinc.com>)

AppLogic允许复杂的应用部署，例如ArcGIS Server的部署。AppLogic可以实现将模板中的一个多层配置和管理作为一个单独的对象来处理，如下图所示

示：

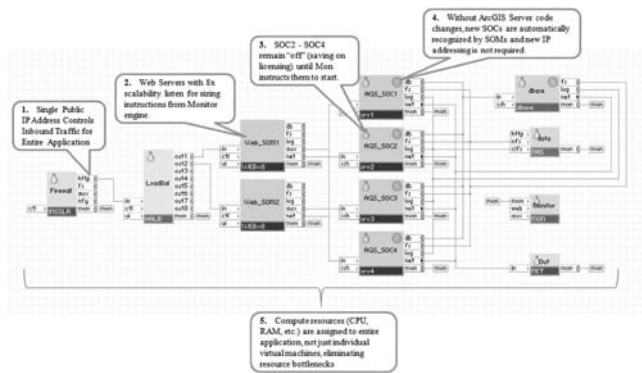


图4-2 ArcGIS Server在AppLogic的部署

从图上可以看出来，ArcGIS Server的SOM分布在两个服务器上，前端负载均衡自动进行消息的分配。SOC分布在其它几台服务器上，其中SOC2—SOC4保留停止状态，当监视器请求启动时候再工作。当然在不改变ArcGIS Server代码情况下，新的SOC机器可以自动的被SOM识别。

4.2 ArcGIS Server与Amazon的云计算

4.2.1 Amazon公司简介

亚马逊是互联网上最大的在线零售商，但是同时也为独立开发人员以及开发商提供云计算服务平台。亚马逊将他们的云计算平台称为弹性计算云 (Elastic Compute Cloud, EC2)，它是最早提供远程云计算平台服务的公司。弹性计算云从沿革上来看，并不是亚马逊公司推出的第一项这种服务，它由名为亚马逊网络服务的现有平台发展而来。早在2006年3月，亚马逊就发布了简单存储服务 (Simple Storage Service, S3)，这种存储服务按照每个月类似租金的形式进行服务付费，同时用户还需要为相应的网络流量进行付费。亚马逊网络服务平台使用REST (Representational State Transfer)和简单对象访问协议 (SOAP)等标准接口，用户可以通过这些接口访问到相应的存储服务。2007年7月，亚马逊公司推出了简单队列服务 (Simple Queue Service, SQS)，这项服务使托管主机可以存储计算机之间发送的消息。通过这一项服务，应用程序编写人员可以在分布式程序之间进行数据传递，而无需考虑消息丢失的问题。通过这种服务方式，即使消息的接收方还没有模块启动也没有关系。服务内部会缓存相应的消息，而一旦有消息接收组件被启动运行，则队列服务将消息提交给相应的运行模块进行处理。同样的，用户必须为这种消息传递服务进行付费使用，计费的规则与存储计费规则类似，依据消息的个数以及

消息传递的大小进行收费。

在用户使用模式上，亚马逊的弹性计算云要求用户要创建基于亚马逊规格的服务器映像(名为亚马逊机器映像即亚马逊 Machine Image, AMI)。弹性计算云的目标是服务器映像能够拥有用户想要的任何一种操作系统、应用程序、配置、登录和安全机制，但是当前情况下，它只支持Linux内核。通过创建自己的AMI，或者使用亚马逊预先为用户提供的AMI，用户在完成这一步骤后将AMI上传到弹性计算云平台，然后调用亚马逊的应用编程接口(API)，对AMI进行使用与管理。AMI实际上就是虚拟机的映像，用户可以使用它们来完成任何工作，例如运行数据库服务器，构建快速网络下载的平台，提供外部搜索服务甚至可以出租自己具有特色的AMI而获得收益。用户所拥有的多个AMI可以通过通信而彼此合作，就像当前的集群计算服务平台一样。

在亚马逊提供上述服务的时候，并没有从头开始开发相应的网络服务组件，而是对公司已有的平台进行优化和改造，一方面满足了本身网络零售购物应用程序的需求，另一方面也供外部开发人员使用。

在开放了上述的服务接口之后，亚马逊公司进一步在此基础上开发了EC2系统，并且开放给外部开发人员使用。

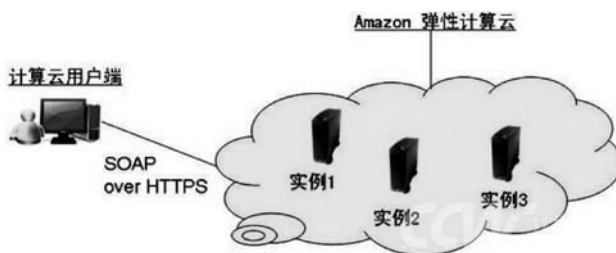


图4-3 亚马逊弹性计算云使用模式

4.2.2 ArcGIS Server在Amazon上部署流程

在Amazon上部署企业级ArcGIS Server需要一些架构设计和了解Amazon Web Services。以下是一个基本的部署过程，当然也会根据你的部署需求的不同而有所不同。

1、准备部署ArcGIS Server到Amazon的云上

- 创建一个Amazon帐户，并确认包含了Elastic Compute Cloud(EC2)的访问。
- 访问Esri-created Amazon Machine Images (AMIs)。
- 创建“Amazon为ArcGIS Server的安全组”，配置至少一个安全组，来允许远程桌面的访问。

- 2、 激活ArcGIS Server
 - 使用ArcGIS Server AMI来登陆EC2进程
 - 使用Windows远程桌面来连接你的新的进程
- 3、 设置存储数据的位置
 - 选择在云端存储数据的位置
 - 可选的，配置更多的EBS来存储本地数据
 - 可选的，使用企业级Geodatabase AMI来配置ArcSDE
- 4、 将数据迁移到Amazon
 - 选择数据迁移方法，传输数据
 - 给定SOC帐户权限，访问数据
- 5、 创建GIS服务和应用
 - 使用ArcGIS Server创建服务，来进行制图，地理编码等
 - 迁移应用到EC2进程，并重新指定他们的服务到运行在EC2进程上的服务
- 6、 定制EC2进程的安全
 - 改变你的EC2进程的管理员密码
 - 可选的，配置ArcGIS Server安全
 - 可选，调整EC2进程的安全，例如防火墙设置，SSL配置等
- 7、 创建自定义的AMI
 - 删除一些敏感或特殊的用户信息
 - 保存你的配置为自定义AMI。
- 8、 访问你自由的AMI的一个新的“production”进程
 - 使用你定制的AMI登陆EC2进程，对新的“production”部署
 - 设置远程桌面连接到生产的EC2进程为不可用
 - 为生产EC2进程配置一个Amazon Elastic IP地址
 - 可以选择的是，可以按需创建或删除进程，可以使用Amazon Auto Scaling API进行人工或程序化自动实现。多个机器都可以通过 Amazon Elastic Load Balancer (ELB)连接在一起。

下图是基本的流程图：



图4—4 ArcGIS亚马逊云服务流程图

4.2.3 ArcGIS Server在Amazon上部署架构

ArcGIS Server架构在Amazon上的部署可以从简单到复杂。如果仅仅是为了开发或测试ArcGIS Server，可以使用一个 EC2进程，并把数据放到EBS上。

当我们的应用部署到生产系统中，或用户访问量比较大了以后，可以添加更多的进程来进行扩容。当然也可以把企业级Geodatabase部署到云端，并保持与已有的数据进行同步。以下是Amazon上部署ArcGIS Server的集中基本架构：

(1)、默认部署

这是最简单的部署方式。ArcGIS Server部署在一个EC2进程上，并提供一个EBS (Elastic Block Store) 来存储数据。你可以使用Esri所提供的ArcGIS Server Amazon Machine Image(AMI)来访问EC2进程。

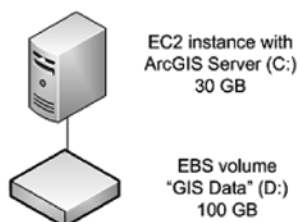


图4-5 ArcGIS 在 EC2 的默认部署

(2)、默认部署更多的EBS

你可以添加更多的EBS来存储更大量的数据。一般来说当EBS存储空间不足，可以增加EBS来存储更多的数据。增加的EBS也是一种你可以在不同驱动下，组织和备份数据的一种方法。

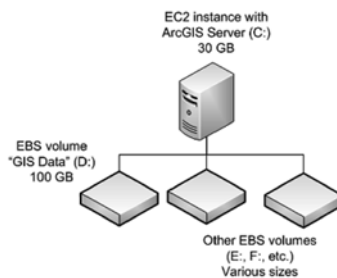


图4-6 ArcGIS 在 EC2 的扩展部署

(3)、默认部署的企业级Geodatabase

这种部署模式适用于，一些需要企业级Geodatabase作为数据源的地图服务和应用。Geodatabase保存在PostgreSQL中，并通过ArcSDE访问。这通过使用企业级 Geodatabase AMI来创建。

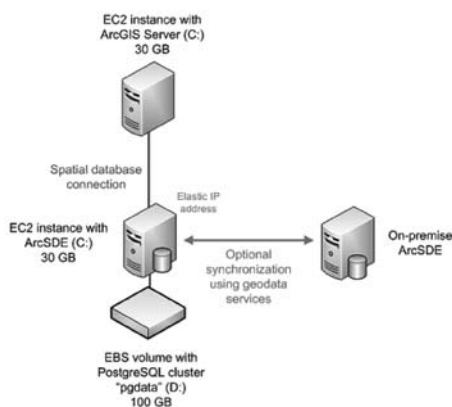


图4-7 企业级GDB在EC2上部署

这种部署需要你从ArcGIS Server AMI 和企业级Geodatabase AMI中登陆进程。你也可以通过使用ArcCatalog来连接空间数据库方式连接进程。你需要用Amazon Elastic IP地址来配置企业级Geodatabase EC2进程，这样连接才能够允许。

(4)、多EC2进程的ArcGIS Server访问本地数据

如果一台机器不足以支持你的部署，你可以选择添加包含ArcGIS Server的更多EC2进程，并使用Amazon Elastic Load Balancer (ELB)来连接他们。ELB提供了通用的Internet连接请求IP地址，并可以实现不同ArcGIS Server EC2进程的分布式工作。

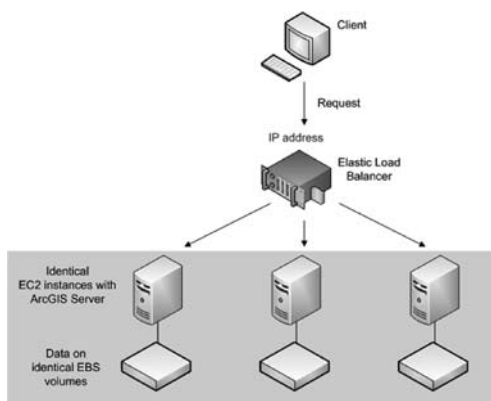


图4-8 ArcGIS在多EC2进程部署

(5) 多EC2进程的ArcGIS Server访问企业级Geodatabase数据

这种部署方式与上面的方式类似，只是数据部署在Amazon的企业级Geodatabase中。ArcGIS Server包含的EC2进程是统一的，并且使用定制的AMI

来登陆。所有都可以连接到同一个空间数据库中。

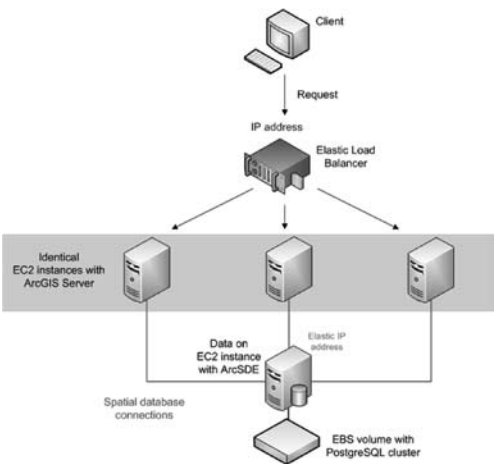


图4-9 多进程EC2的企业级GDB部署

这种部署模式的一个优点是，企业级Geodatabase可以通过Geodatabase服务定期的与已有的geodatabase数据进行同步更新。你也可以访问ArcSDE所提供的所有的时态、版本和复制功能。

（更多的关于ArcGIS Server与Amazon上的帮助，请参考网站：http://help.arcgis.com/en/arcgisserver/9.3/help/arcgis_server_for_amazon_help/index.html）

4.2.4 ArcGIS Server与Amazon云计算实例

Esri的专业服务部门已经将ArcGIS Server在Amazon上的云计算部署到一些应用中。下面是ArcGIS Server运行在Amazon云端的地址站点：

- Southern California Edison: <http://sce-lb-335936873.us-east-1.elb.amazonaws.com:8080/sce/index.html>
- Nature Conservancy: <http://climatewizard.org/>
- USDA Food Atlas: <http://snap-load-balancer-244858692.us-east-1.elb.amazonaws.com/index.html>

4.3 ArcGIS与虚拟化部署

什么是虚拟化（网络，磁盘，内存等）。对于ArcGIS Server来说，虚拟化是用于平台的虚拟化，可以允许多平台系统和应用运行在一个物理机器上。

目前主流的虚拟技术包括：

- ✓ VMware

- ✓ Microsoft Hyper-V
- ✓ Sun Microsystems Solaris Containers
- ✓ Citrix XenServer and Citix Essentials

ArcGIS Server可以完全支持虚拟环境，包括平台的支持，这些平台包括Windows, Red Hat Linux, SUSE Linux, Sun Solaris等。

Esri的ArcGIS Server可以完全的使用虚拟化的技术，例如VMware的架构。ArcGIS Server多层服务器部署模式，可以充分利用虚拟化架构的优势，例如提高管理，可用性和扩展性，以减少TCO。下图为ArcGIS Server运行在VMware ESXi上的虚拟机。

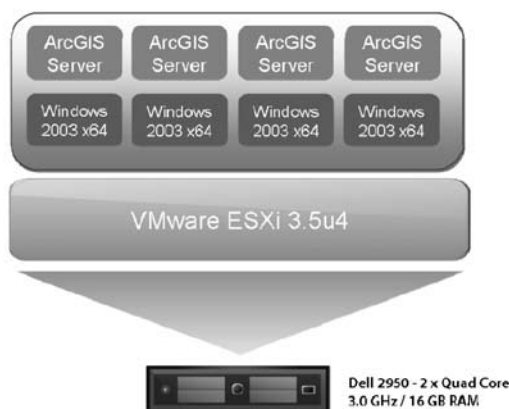


图4-10 ArcGIS与VMware

下图为ArcGIS Online中国地图服务的硬件虚拟服务器部署图，采用Windows2008服务器，虚拟平台采用Hyper-V2.0；R710配置为2*2.26 GHz E5520；其中切片地图都放在ArcSOM所在的服务器机器1上。

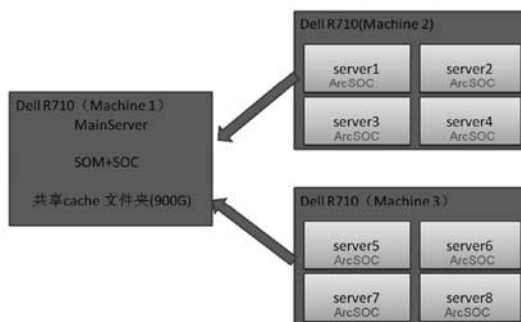


图4-11 ArcGIS虚拟化部署实例

ArcGIS虚拟化部署的优势：

- ✓ 可以将不同版本的ArcGIS Server部署在一台物理机器上。特别适用于开发环境和版本迁移情况
 - ✓ 通过分离ArcGIS Server应用可以减少整个架构的安全风险
 - ✓ 在发生灾害的时候，ArcGIS Server应用和服务可以快速恢复
 - ✓ 加速开发和部署ArcGIS Server自由应用
 - ✓ 最大化的利用硬件和计算机基础设施，简化服务器设备和系统规划。
- 例如新增的机器可以快速的配置为SOC机器，以满足高峰期访问的需要。

(更多资源请参考: <http://www.esri.com/library/whitepapers/pdfs/arcgis-server-virtualization.pdf>)

5. ArcGIS的云安全

云平台建设的一个重要建设考虑内容，其中云平台ArcGIS为云安全提供了多种级别的部署方式。

企业级安全:

ArcGIS安全可以在应用，网络，操作系统和DBMS层面控制。

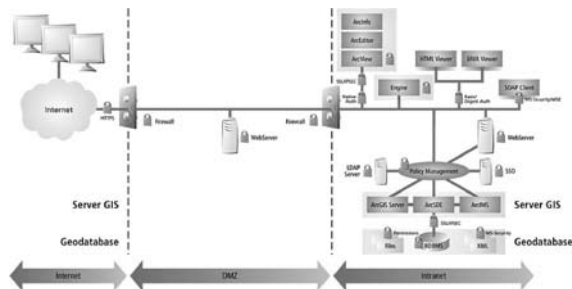


图5-1 ArcGIS安全

应用层提供安全控制的最大可伸缩度。通过使用ArcObjects，桌面应用，Web应用和Web Services可以和标准技术结合起来提供验证、授权，提供访问控制。ArcGIS可以为不同的用户访问权限提供功能限制和地理内容限制。ArcGIS Web应用和服务业可以通过安全通道（https）定制使用标准的授权方法（基础的，分类，图表，客户端授权）。而且，ArcGIS应用可以和已有轻量的目录访问协议，单点登录，以及策略管理系统结合起来为不同的角色用户指定特定的内容权限。

ArcGIS Server通过基于角色访问控制（RBAC）来保证安全。ArcGIS Server的JAVA应用提供了定制安全系统来管理存在于其它的关系数据库或目录服务（LDAP或Active Directory）的用户和角色，并通过API可以扩展安全系统。

应用安全：

(1)、富客户端

富客户端的安全主要可以通过三种方式来实现：

> 权限登陆

将Web Service与令牌（Token）服务集成

将SSO与 Windows授权集成

> 加密通信

直连：使用数据库客户端本身所提供的SSL或IPSec

应用连接：SDE所提供的网络协议安全通道端口5151

Web服务：SSL和HTTPS安全协议

> 定制开发

细粒度GUI访问控制，包括编辑、拷贝、剪切、粘贴、打印等

LDAP集成

(2)、移动安全

> ArcPAD

ArcPAD的AXF数据格式，提供了密码保护和加密机制。也可以使用内存卡加密

ArcGIS Server用户和组限制了可以发布ArcPAD数据的权限

Internet连接保证ArcPAD数据同步的安全

> ArcGIS Mobile

GeoData服务：通过HTTPS（SSL）或VPN通道

使用Token服务

Web服务认证

使用Windows Mobile加密API

(3)、Web应用安全

> ArcGIS Server Manager

自动应用ASP.NET和 Java EE的Web应用安全。例如修改ASP.NET中的web.config文件。

> 应用接口

.NET和 Java ADF：集成令牌安全服务的功能

REST API（JavaScript，Flex，Silverlight）：可以嵌入到URL中；动态产生令牌

(4)、Web服务安全

> ArcGIS Server Manager

权限继承：包括文件夹级别和个别服务级别

服务层安全限制：只允许Internet和Web连接

所有Web服务接口安全：如SOAP服务的WS-Security地址可通过第三方XML/SOAP网关实现。

➤ 移除本地连接访问

清空AGSUsers的用户组

云平台安全策略：

企业级云平台提供了本身的安全策略，下面以Amazon为例说明。

(1)、EC2安全

EC2通过多种模式保证安全，包括托管运营系统的SSH权限控制，客户运营系统的配对密钥，无状态防火墙方法，以及X.509认证的API请求等，来保证访问的安全，如下图为EC2隔离进程的安全模式：

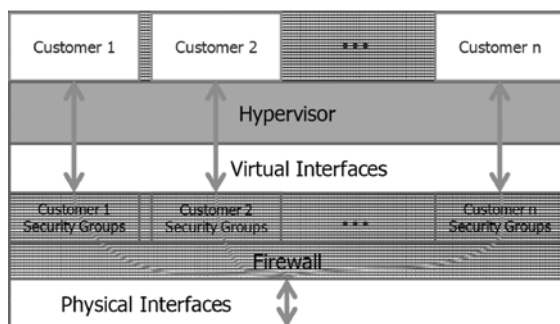


图5-2 EC2安全

(2)、虚拟专用云（VPC）安全

亚马逊的虚拟专用云（VPC）提供了多级的安全访问策略，如下图所示：

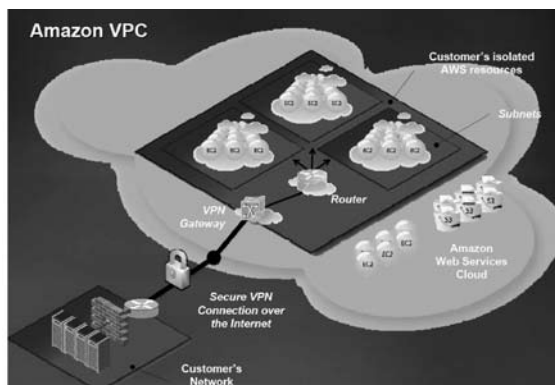


图5-3 VPC安全

(关于更多关于ArcGIS的亚马逊云服务安全, 请参考: http://help.arcgis.com/en/arcgisserver/9.3/help/arcgis_server_for_amazon_help/index.html#/ArcGIS_Server_security_in_Amazon/00rq0000000n000000/)

附录1: 云计算演变

(信息来源: C114中国通信网; <http://market.c114.net/187/a377086.html>)

“云计算”在2008年突然炙手可热, 大红大紫, 各大IT巨头非“云”不谈, 掀起阵阵“云团”: Amazon的云存储、Google的云开发、Trendmicro的云安全等等, 一时“风起云涌”, 热闹非凡。2008年甚至被称为“云计算”年, “云计算”也荣登2008年十大IT技术之一。

本文简述了“云计算”概念的由来, 并对“云计算”的“前世”——“云计算”是由哪些概念演变而来以及相互关系进行了探讨, 最后对“云计算”的发展趋势发表了自己的看法。

“云”从何来

云计算(Cloud Computing)源于Google工程师克里斯托弗于2006年秋季开始的“Google 101”项目。该项目作为华盛顿大学一门课程, 指导学生进行“云”系统的开发和编程。该门课程后来受到众多院校的欢迎, 随着IBM的加入, 变为Google-IBM的联合大学“云”。随后各个公司相继推出“云计算”相关的计划和应用, “云计算”如雨后春笋破土而出, 成为下一代互联网革命的代名词。

“云”之前世

“云计算”的概念虽然不长, 但是有人甚至将“云计算”的思想追溯到20世纪60年代的人工智能。从本人看来, “云计算”至少揉合了网格计算(Grid Computing)、软件即服务(SaaS)、平台即服务(PaaS)和硬件即服务(HaaS)四个方面内容, 可是说“云计算”是它们量变积累的质变。下面来看看各个技术的发展和运用, 以及它们与“云计算”的关系。

1. 网格计算(Grid Computing)

网格计算的概念是上个世纪90年代被提出, 目的是利用并行和分布式技术将众多单个的、零散的计算机形成一个超级“虚拟计算机”, 从而使人们可以像使用电力一样来使用该“虚拟计算机”的计算资源。

在科研领域中, 网格技术受到了极大的推崇, 同时也得到国家广泛的重视和支持, 像美国的NEES Grid, TeraGrid, 欧盟网格以及中国国家网格在本

地区甚至世界范围内都具有广泛影响，极大的加速了海量科研数据的处理和共享，一个曾经受到广泛关注的例子是1999年由加州大学伯克利分校发起的“搜寻外星文明(SETI)”计划，吸引了超过3,000,000台计算机参与。在商业领域，IBM和SUN走在前例：IBM于2003年发布了针对航天、汽车、政府和生命科学等领域的多种不同的网格解决方案，使网格计算从学术和研究领域推向商业市场；SUN则直言“网格就是计算机”，推出了SUN GRID，向科研机构或者企业提供网格服务。

网格计算可以说是“云”计算的萌芽，是“云计算”的能够成为可能的助推器。网格技术中的分布式和并行技术也正是“云”计算的核心技术之一。但是网格技术强调的是利用闲散众多的CPU资源来解决科研或者大型企业领域中日益增长的密集型计算需求，而这不一定是“云计算”所必须具有的特征，“云计算”强调的是“云”就是一切，理想状态下，人们在“云”上得到一切需求，至于“云”是怎样构建并不是用户关心的，也不需要用户参与。

2. 软件即服务(SaaS)

软件即服务的概念从提出到现在，有近10年时间，其核心内容是通过Internet将软件作为一个服务提供给使用者，而不像传统的软件商业模式——人们购买并维护软件。SaaS颠覆了软件发布和使用模式，也同时改变了软件厂商的盈利模式，与传统的软件发布方式相比，具有前期投入少、灵活、“按需服务(On-demand)”的优势。

现在SaaS应用和产品已经是国际上中小企业信息化的常见模式，无论是传统的IT巨头如Google、Oracle，还是新兴的SaaS厂商如Salesforce、NetSuite都吸引了众多客户，取得了很大成功。在国内，SaaS从概念萌芽到实际应用阶段过渡，用友、金蝶、阿里巴巴、800CRM等都推出了自己的SaaS产品，并吸引了相当一部分中小企业客户。从目前来看，SaaS的使用者大都是中小企业，SaaS应用的领域也比较集中于HR、CRM，大规模应用SaaS还有很远的路要走。

从使用者的角度来看，SaaS应用在很大程度上等同于“云计算”应用，SaaS应用的“在线服务(On-line service)”和“按需服务(On-demand Service)”的特点即是“云计算”的精髓。但是，“云计算”却是从更广、更深的角度来阐释、丰富、清晰这些特征，其旗帜鲜明的表示这是一个由内向外、从上向下的整个IT行业的变革，下一节提到的“平台即服务(PaaS)”和“硬件即服务(HaaS)”即是SaaS概念在IT行业其他范围的延伸和拓展。

3. 平台即服务(PaaS)和硬件即服务(HaaS)

平台即服务(PaaS)和硬件即服务(HaaS)是开发环境、服务器平台、以及硬件资源作为服务提供的商业模式，其同样具有SaaS服务的精髓——“在线服

务”和“按需服务”。平台即服务(HaaS)强调开发环境和服务器平台，如PC时代微软的Windows和Visual Studio，而硬件即服务(HaaS)则强调计算能力和存储资源等。两者虽有所侧重，但密不可分，例如使用一个在线平台开发一个应用，必然会用到数据存储空间。

在这两个方面，Amazon和Google是领头羊：Amazon在2006年开始推出存储服务，到现在已经拥EC2、S3、SimpleDB等一个完整的在线开发平台和资源；Google则在2008年4月推出了Google App Engine在线开发平台，提供类似Amazon的功能。此外，还有一些专业厂商如Salesforce、heroku、Bungee等提供了某些方面的应用程序在线开发平台。在国内，八百客的800APP是PaaS的代表产品，与Salesforce相似，其提供了一套比较完整的企业级API，可以供企业定制和开发自己的CRM、HR等管理程序的应用流程。

平台即服务(PaaS)和硬件即服务(HaaS)是“在线服务”和“按需服务”在更高层次应用的体现，其的出现加速了用户向“瘦客户端”的转变，而2008年“上网本”的热销恰恰是这种趋势的注脚。对于“云计算”而言，软件+平台+硬件才是其完整的生态链，三者相互影响、相互促进。只有这样，“云计算”才能引领下一代互联网革命。

“云”之今生

可见，“云计算”绝不是凭空冒出来的，正是因为有了前面谈到的各项应用基石，“云计算”才会一呼百应，炙手可热。在“云计算”时代，一切都是服务(EaaS, Everything as a Service)：存储资源、计算资源、开发环境、软件的使用和维护等等，一切服务都在“云”上。

从IT行业的横向发展来看，这种以“服务”为中心的趋势是让专业的人们更加专注于自己擅长的事情，从而大大提高整个IT行业的效率，促使IT行业在其他行业中发挥更大效益和影响。从IT行业的纵向发展来看，这种趋势更为明显，所谓“IT大势，分久必合”：计算机一出生就是庞然大物，成百上千人不得不通过终端来分享这个科技成果，这是“合”的时期；随后个人PC的兴起，成就了微软和比尔盖茨的梦想和预言，这是“分”的时期；毋庸置疑，下一个又将是“合”的时期，而“云计算”也许就是它的代言人。历史虽然不是简单的重复，但是历史前进发展的规律是无法阻挡，我们欢迎“云计算”时代的来临，同时也欢迎巨风网络(www.cn9g.com)时代的来临。

附录2：剖析云计算技术

(摘自：计算机世界报 2008年05月12日第17期 38—40 作者：清华大学 陈康 郑纬民)

实例1：Google的云计算平台与应用

Google的云计算技术实际上是针对Google特定的网络应用程序而定制的。针对内部网络数据规模超大的特点，Google提出了一整套基于分布式并行集群方式的基础架构，利用软件的能力来处理集群中经常发生的节点失效问题。

从2003年开始，Google连续几年在计算机系统研究领域的最顶级会议与杂志上发表论文，揭示其内部的分布式数据处理方法，向外界展示其使用的云计算核心技术。从其近几年发表的论文来看，Google使用的云计算基础架构模式包括四个相互独立又紧密结合在一起的系统。包括Google建立在集群之上的文件系统Google File System，针对Google应用程序的特点提出的Map/Reduce编程模式，分布式的锁机制Chubby以及Google开发的模型简化的大规模分布式数据库BigTable。

Google File System 文件系统

为了满足Google迅速增长的数据处理需求，Google设计并实现了Google文件系统（GFS，Google File System）。GFS与过去的分布式文件系统拥有许多相同的目标，例如性能、可伸缩性、可靠性以及可用性。然而，它的设计还受到Google应用负载和技术环境的影响。主要体现在以下四个方面：

1. 集群中的节点失效是一种常态，而不是一种异常。由于参与运算与处理的节点数目非常庞大，通常会使用上千个节点进行共同计算，因此，每时每刻总会有节点处在失效状态。需要通过软件程序模块，监视系统的动态运行状况，侦测错误，并且将容错以及自动恢复系统集成在系统中。

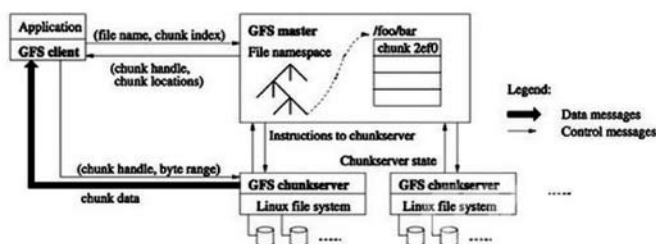
2. Google系统中的文件大小与通常文件系统中的文件大小概念不一样，文件大小通常以G字节计。另外文件系统中的文件含义与通常文件不同，一个大文件可能包含大量数目的通常意义上的小文件。所以，设计预期和参数，例如I/O操作和块尺寸都要重新考虑。

3. Google文件系统上的文件读写模式和传统的文件系统不同。在Google应用（如搜索）中对大部分文件的修改，不是覆盖原有数据，而是在文件尾追加新数据。对文件的随机写是几乎不存在的。对于这类巨大文件的访问模式，客户端对数据块缓存失去了意义，追加操作成为性能优化和原子性（把一个事务看做是一个程序。它要么被完整地执行，要么完全不执行）保证的焦点。

4. 文件系统的某些具体操作不再透明，而且需要应用程序的协助完成，应用程序和文件系统API的协同设计提高了整个系统的灵活性。例如，放松了对GFS一致性模型的要求，这样不用加重应用程序的负担，就大大简化了文件系统的设计。还引入了原子性的追加操作，这样多个客户端同时进行追加的时候，就不需要额外的同步操作了。

总之，GFS是为Google应用程序本身而设计的。据称，Google已经部署了许多GFS集群。有的集群拥有超过1000个存储节点，超过300T的硬盘空间，被不同机器上的数百个客户端连续不断地频繁访问着。

下图1给出了Google File System的系统架构，一个GFS集群包含一个主服务器和多个块服务器，被多个客户端访问。文件被分割成固定尺寸的块。在每个块创建的时候，服务器分配给它一个不变的、全球惟一的64位块句柄对它进行标识。块服务器把块作为linux文件保存在本地硬盘上，并根据指定的块句柄和字节范围来读写块数据。为了保证可靠性，每个块都会复制到多个块服务器上，缺省保存三个备份。主服务器管理文件系统所有的元数据，包括名字空间、访问控制信息和文件到块的映射信息，以及块当前所在的位置。GFS客户端代码被嵌入到每个程序里，它实现了Google文件系统API，帮助应用程序与主服务器和块服务器通信，对数据进行读写。客户端跟主服务器交互进行元数据操作，但是所有的数据操作的通信都是直接和块服务器进行的。客户端提供的访问接口类似于POSIX接口，但有一定的修改，并不完全兼容POSIX标准。通过服务器端和客户端的联合设计，Google File System能够针对它本身的应用获得最大的性能以及可用性效果。



图附-1 Google File System的系统架构

MapReduce分布式编程环境

为了让内部非分布式系统方向背景的员工能够有机会将应用程序建立在大规模的集群基础之上，Google还设计并实现了一套大规模数据处理的编程规范Map/Reduce系统。这样，非分布式专业的程序编写人员也能够为大规模的集群编写应用程序而不用去顾虑集群的可靠性、可扩展性等问题。应用程序编写人员只需要将精力放在应用程序本身，而关于集群的处理问题则交由平台来处理。

Map/Reduce通过“Map（映射）”和“Reduce（化简）”这样两个简单的概念来参加运算，用户只需要提供自己的Map函数以及Reduce函数就可以在集群上进行大规模的分布式数据处理。

据称，Google的文本索引方法，即搜索引擎的核心部分，已经通过Map

Reduce的方法进行了改写，获得了更加清晰的程序架构。在Google内部，每天有上千个Map Reduce的应用程序在运行。

分布式大规模数据库管理系统BigTable

构建于上述两项基础之上的第三个云计算平台就是Google关于将数据库系统扩展到分布式平台上的BigTable系统。很多应用程序对于数据的组织还是非常有规则的。一般来说，数据库对于处理格式化的数据还是非常方便的，但是由于关系数据库很强的一致性要求，很难将其扩展到很大的规模。为了处理 Google内部大量的格式化以及半格式化数据，Google构建了弱一致性要求的大规模数据库系统BigTable。据称，现在有很多Google的应用程序建立在BigTable之上，例如Search History、Maps、Orkut和RSS阅读器等。

下图2给出了在BigTable模型中的数据模型。数据模型包括行列以及相应的时间戳，所有的数据都存放在表格中的单元里。BigTable的内容按照行来划分，将多个行组成一个小表，保存到某一个服务器节点中。这一个小表就被称为Tablet。

以上是Google内部云计算基础平台的三个主要部分，除了这三个部分之外，Google还建立了分布式程序的调度器，分布式的锁服务等一系列相关的云计算服务平台。

Google的云应用

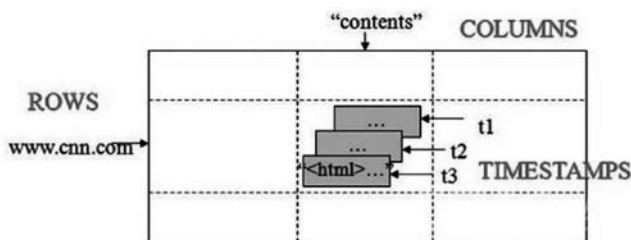
除了上述的云计算基础设施之外，Google还在其云计算基础设施之上建立了一系列新型网络应用程序。由于借鉴了异步网络数据传输的Web 2.0技术，这些应用程序给予用户全新的界面感受以及更加强大的多用户交互能力。其中典型的Google云计算应用程序就是Google推出的与 Microsoft Office软件进行竞争的Docs网络服务程序。Google Docs是一个基于Web的工具，它有跟Microsoft Office相近的编辑界面，有一套简单易用的文档权限管理，而且它还记录下所有用户对文档所做的修改。Google Docs的这些功能令它非常适用于网上共享与协作编辑文档。Google Docs甚至可以用于监控责任清晰、目标明确的项目进度。当前，Google Docs已经推出了文档编辑、电子表格、幻灯片演示、日程管理等多个功能的编辑模块，能够替代Microsoft Office相应的一部分功能。值得注意的是，通过这种云计算方式形成的应用程序非常适合于多个用户进行共享以及协同编辑，为一个小组的人员进行共同创作带来很大的方便性。

Google Docs是云计算的一种重要应用，即可以通过浏览器的方式访问远端大规模的存储与计算服务。云计算能够为大规模的新一代网络应用打下良好的基础。

虽然Google可以说是云计算的最大实践者，但是，Google的云计算平台是私有的环境，特别是Google的云计算基础设施还没有开放出来。除了开放

有限的应用程序接口，例如GWT（Google Web Toolkit）以及Google Map API等，Google并没有将云计算的内部基础设施共享给外部的用户使用，上述的所有基础设施都是私有的。

幸运的是，Google公开了其内部集群计算环境的一部分技术，使得全球的技术开发人员能够根据这一部分文档构建开源的大规模数据处理云计算基础设施，其中最著名的项目即Apache旗下的Hadoop项目。而下面的两个云计算的实现则为外部的开发人员以及中小公司提供了云计算的平台环境，使得开发者能够在云计算的基础设施之上构建自己的新型网络应用。其中IBM的蓝云计算平台是可供销售的计算平台，用户可以基于这些软硬件产品自己构建云计算平台。亚马逊的弹性计算云则是托管式的云计算平台，用户可以通过远端的操作界面直接使用。



图附-2 Google BigTable的数据模型

实例2：亚马逊的弹性计算云

亚马逊是互联网上最大的在线零售商，但是同时也为独立开发人员以及开发商提供云计算服务平台。亚马逊将他们的云计算平台称为弹性计算云（Elastic Compute Cloud, EC2），它是最早提供远程云计算平台服务的公司。

开放的服务

与 Google提供的云计算服务不同，Google仅为自己在互联网上的应用提供云计算平台，独立开发商或者开发人员无法在这个平台上工作，因此只能转而通过开源的Hadoop软件支持来开发云计算应用。亚马逊的弹性计算云服务也和IBM的云计算服务平台不一样，亚马逊不销售物理的云计算服务平台，没有类似于“蓝云”一样的计算平台。亚马逊将自己的弹性计算云建立在公司内部的大规模集群计算的平台之上，而用户可以通过弹性计算云的网络界面去操作在云计算平台上运行的各个实例（Instance），而付费方式则由用户的使用状况决定，即用户仅需要为自己所使用的计算平台实例付费，运行结束后计费也随之结束。

弹性计算云从沿革上来看，并不是亚马逊公司推出的第一项这种服务，它由名为亚马逊网络服务的现有平台发展而来。早在2006年3月，亚马逊就发布了简单存储服务（Simple Storage Service, S3），这种存储服务按照每个月类似租金的形式进行服务付费，同时用户还需要为相应的网络流量进行付费。亚马逊网络服务平台使用REST（Representational State Transfer）和简单对象访问协议（SOAP）等标准接口，用户可以通过这些接口访问到相应的存储服务。

2007年7月，亚马逊公司推出了简单队列服务（Simple Queue Service, SQS），这项服务使托管主机可以存储计算机之间发送的消息。通过这一项服务，应用程序编写人员可以在分布式程序之间进行数据传递，而无须考虑消息丢失的问题。通过这种服务方式，即使消息的接收方还没有模块启动也没有关系。服务内部会缓存相应的消息，而一旦有消息接收组件被启动运行，则队列服务将消息提交给相应的运行模块进行处理。同样的，用户必须为这种消息传递服务进行付费使用，计费的规则与存储计费规则类似，依据消息的个数以及消息传递的大小进行收费。

在亚马逊提供上述服务的时候，并没有从头开始开发相应的网络服务组件，而是对公司已有的平台进行优化和改造，一方面满足了本身网络零售购物应用程序的需求，另一方面也供外部开发人员使用。

在开放了上述的服务接口之后，亚马逊公司进一步在此基础上开发了EC2系统，并且开放给外部开发人员使用。

灵活的工作模式

亚马逊的云计算模式沿袭了简单易用的传统，并且建立在亚马逊公司现有的云计算基础平台之上。弹性计算云用户使用客户端通过SOAP over HTTPS协议来实现与亚马逊弹性计算云内部的实例进行交互。使用HTTPS协议的原因是为了保证远端连接的安全性，避免用户数据在传输的过程中造成泄露。因此，从使用模式上来说，弹性计算云平台为用户或者开发人员提供了一个虚拟的集群环境，使得用户的应用具有充分的灵活性，同时也减轻了云计算平台拥有者（亚马逊公司）的管理负担。

而弹性计算云中的实例是一些真正在运行中的虚拟机服务器，每一个实例代表一个运行中的虚拟机。对于提供给某一个用户的虚拟机，该用户具有完整的访问权限，包括针对此虚拟机的管理员用户权限。虚拟服务器的收费也是根据虚拟机的能力进行计算的，因此，实际上用户租用的是虚拟的计算能力，简化了计费方式。在弹性计算云中，提供了三种不同能力的虚拟机实例，具有不同的收费价格。例如，其中默认的也是最小的运行实例是1.7GB的内存，1个EC2的计算单元（1虚拟的计算核以相关的计算单元），160GB的虚拟机内部存储容量，是一个32位的计算平台，收费标准为每个小时10美

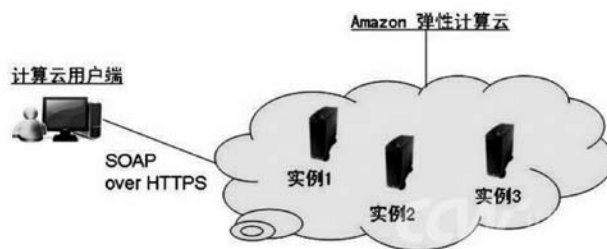
分。在当前的云计算平台中，还有两种性能更加强劲的虚拟机实例可供使用，当然价格也更加昂贵一点。

由于用户在部署网络程序的时候，一般会使用超过一个运行实例，需要很多个实例共同工作。弹性计算云的内部也架设了实例之间的内部网络，使得用户的应用程序在不同的实例之间可以通信。在弹性计算云中的每一个计算实例都具有一个内部的IP地址，用户程序可以使用内部IP地址进行数据通信，以获得数据通信的最好性能。每一个实例也具有外部的地址，用户可以将分配给自己的弹性IP地址分配给自己的运行实例，使得建立在弹性计算云上的服务系统能够为外部提供服务。当然，亚马逊公司也对网络上的服务流量计费，计费规则也按照内部传输以及外部传输进行分开。

总而言之，亚马逊通过提供弹性计算云，减少了小规模软件开发人员对于集群系统的维护，并且收费方式相对简单明了，用户使用多少资源，只需要为这一部分资源付费即可。这种付费方式与传统的主机托管模式不同。传统的主机托管模式让用户将主机放入到托管公司，用户一般需要根据最大或者计划的容量进行付费，而不是根据使用情况进行付费，而且，可能还需要保证服务的可靠性、可用性等，付出的费用更多，而很多时候，服务并没有进行满额资源使用。而根据亚马逊的模式，用户只需要为实际使用情况付费即可。

在用户使用模式上，亚马逊的弹性计算云要求用户要创建基于亚马逊规格的服务器映像（名为亚马逊机器映像即亚马逊 Machine Image, AMI）。弹性计算云的目标是服务器映像能够拥有用户想要的任何一种操作系统、应用程序、配置、登录和安全机制，但是当前情况下，它只支持Linux内核。通过创建自己的AMI，或者使用亚马逊预先为用户提供的AMI，用户在完成这一步骤后将AMI上传到弹性计算云平台，然后调用亚马逊的应用编程接口（API），对AMI进行使用与管理。AMI实际上就是虚拟机的映像，用户可以使用它们来完成任何工作，例如运行数据库服务器，构建快速网络下载的平台，提供外部搜索服务甚至可以出租自己具有特色的AMI而获得收益。用户所拥有的多个AMI可以通过通信而彼此合作，就像当前的集群计算服务平台一样。

在弹性计算云的将来发展过程中，亚马逊也规划了如何在云计算平台之上帮助用户开发Web 2.0的应用程序。亚马逊认为除了它所依赖的网络零售业务之外，云计算也是亚马逊公司的核心价值所在。可以预见，在将来的发展过程中，亚马逊必然会在弹性计算云的平台上添加更多的网络服务组件模块，为用户构建云计算应用提供方便。



图附-3 亚马逊弹性计算云使用模式

实例3: IBM蓝云计算平台

IBM在2007年11月15日推出了蓝云计算平台，为客户带来即买即用的云计算平台。它包括一系列的云计算产品，使得计算不仅仅局限在本地机器或远程服务器农场（即服务器集群），通过架构一个分布式、可全球访问的资源结构，使得数据中心在类似于互联网的环境下运行计算。

通过IBM的技术白皮书，我们可以一窥蓝云计算平台的内部构造。“蓝云”建立在IBM大规模计算领域的专业技术基础上，基于由IBM软件、系统技术和服务支持的开放标准和开源软件。简单地说，“蓝云”基于IBM Almaden研究中心（Almaden Research Center）的云基础架构，包括Xen和PowerVM虚拟化、Linux操作系统映像以及Hadoop文件系统与并行构建。“蓝云”由IBM Tivoli软件支持，通过管理服务器来确保基于需求的最佳性能。这包括通过能够跨越多服务器实时分配资源的软件，为客户带来一种无缝体验，加速性能并确保在最苛刻环境下的稳定性。IBM新近发布的“蓝云（Blue Cloud）”计划，能够帮助用户进行云计算环境的搭建。它通过将Tivoli、DB2、WebSphere与硬件产品（目前是x86刀片服务器）集成，能够为企业架设一个分布式、可全球访问的资源结构。根据IBM的计划，首款支持Power和x86处理器刀片服务器系统的“蓝云”产品于2008年正式推出，并且计划随后推出基于System z“大型主机”的云环境，以及基于高密度机架集群的云环境。

在IBM的云计算白皮书上，我们可以看到如下的蓝云计算平台配置情况。

下图4演示了蓝云计算的高层架构。可以看到，蓝云计算平台由一个数据中心：IBM Tivoli部署管理软件（Tivoli Provisioning Manager）、IBM Tivoli监控软件（IBM Tivoli Monitoring）、IBM WebSphere应用服务器、IBM DB2数据库以及一些虚拟化的组件共同组成。图中的架构主要描述了云计算的后台架构，并没有涉及到前台的用户界面。

蓝云的硬件平台并没有什么特殊的地方，但是蓝云使用的软件平台相

较于以前的分布式平台具有不同的地方，主要体现在对与虚拟机的使用以及对于大规模数据处理软件Apache Hadoop的部署。Hadoop是网络开发人员根据Google公司公开的资料开发出来的类似于Google File System的Hadoop File System以及相应的Map/Reduce编程规范。现在也正在进一步开发类似于Google的Chubby系统以及相应的分布式数据库管理系统BigTable。由于Hadoop是开源的，因此可以被用户单位直接修改，以适合应用的特殊需求。IBM的蓝云产品则直接将Hadoop软件集成到自己本身的云计算平台之上。

“蓝云”中的虚拟化

从蓝云的结构上我们还可以看出，在每一个节点上运行的软件栈与传统的软件栈一个很大的不同在于蓝云内部使用了虚拟化技术。虚拟化的方式在云计算中可以在两个级别上实现。一个级别是在硬件级别上实现虚拟化。硬件级别的虚拟化可以使用IBM p系列的服务器，获得硬件的逻辑分区LPAR。逻辑分区的CPU资源能够通过IBM Enterprise Workload Manager来管理。通过这样的方式加上在实际使用过程中的资源分配策略，能够使得相应的资源合理地分配到各个逻辑分区。P系列系统的逻辑分区最小粒度是1/10颗中央处理器（CPU）。

虚拟化的另外一个级别可以通过软件来获得，在蓝云计算平台中使用了Xen虚拟化软件。Xen也是一个开源的虚拟化软件，能够在现有的Linux基础之上运行另外一个操作系统，并通过虚拟机的方式灵活地进行软件部署和操作。

通过虚拟机的方式进行云计算资源的管理具有特殊的好处。由于虚拟机是一类特殊的软件，能够完全模拟硬件的执行，因此能够在上面运行操作系统，进而能够保留一整套运行环境语义。这样，可以将整个执行环境通过打包的方式传输到其他物理节点上，这样就能够使得执行环境与物理环境隔离，方便整个应用程序模块的部署。总体上来说，通过将虚拟化的技术应用到云计算的平台，可以获得一些良好的特性。

1. 云计算的管理平台能够动态地将计算平台定位到所需要的物理平台上，而无需停止运行在虚拟机平台上的应用程序，这比采用虚拟化技术之前的进程迁移方法更加灵活。

2. 能够更加有效率地使用主机资源，将多个负载不是很重的虚拟机计算节点合并到同一个物理节点上，从而能够关闭空闲的物理节点，达到节约电能的目的。

3. 通过虚拟机在不同物理节点上的动态迁移，能够获得与应用无关的负载平衡性能。由于虚拟机包含了整个虚拟化的操作系统以及应用程序环境，因此在进行迁移的时候带着整个运行环境，达到了与应用无关的目的。

4. 在部署上也更加灵活，即将虚拟机直接部署到物理计算平台当

中。

总而言之，通过虚拟化的方式，云计算平台能够达到极其灵活的特性，而如果不使用虚拟化的方式则会有很多的局限。



图附-4 IBM蓝云产品架构

“蓝云”中的存储结构

蓝云计算平台中的存储体系结构对于云计算来说也是非常重要的，无论是操作系统，服务程序还是用户应用程序的数据都保存在存储体系中。云计算并不排斥任何一种有用的存储体系结构，而是需要跟应用程序的需求结合起来获得最好的性能提升。总体上来说，云计算的存储体系结构包含类似于Google File System的集群文件系统以及基于块设备方式的存储区域网络SAN两种方式。

在设计云计算平台的存储体系结构的时候，不仅仅是需要考虑存储的容量。实际上随着硬盘容量的不断扩充以及硬盘价格的不断下降，使用当前的磁盘技术，可以很容易通过使用多个磁盘的方式获得很大的磁盘容量。相较于磁盘的容量，在云计算平台的存储中，磁盘数据的读写速度是一个更重要的问题。单个磁盘的速度很有可能限制应用程序对于数据的访问，因此在实际使用的过程中，需要将数据分布到多个磁盘之上，并且通过对于多个磁盘的同时读写以达到提高速度的目的。在云计算平台中，数据如何放置是一个非常重要的问题，在实际使用的过程中，需要将数据分配到多个节点的多个磁盘当中。而能够达到这一目的的存储技术趋势当前有两种方式，一种是使用类似于Google File System的集群文件系统，另外一种是基于块设备的存储区域网络SAN系统。

Google文件系统我们在前面已经做过一定的描述。在IBM的蓝云计算平台

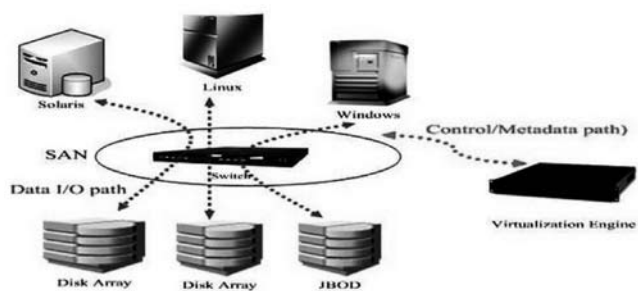
中使用的是它的开源实现Hadoop HDFS (Hadoop Distributed File System)。这种使用方式将磁盘附着于节点的内部，并且为外部提供一个共享的分布式文件系统空间，并且在文件系统级别做冗余以提高可靠性。在合适的分布式数据处理模式下，这种方式能够提高总体的数据处理效率。Google文件系统的这种架构与SAN系统有很大的不同。

SAN系统也是云计算平台的另外一种存储体系结构选择，在蓝云平台上也有一定的体现，IBM也提供SAN的平台能够接入到蓝云计算平台中。图5就是一个SAN系统的结构示意图。

从图5中可以看到，SAN系统是在存储端构建存储的网络，将多个存储设备构成一个存储区域网络。前端的主机可以通过网络的方式访问后端的存储设备。而且，由于提供了块设备的访问方式，与前端操作系统无关。在SAN连接方式上，可以有多种选择。一种选择是使用光纤网络，能够操作快速的光纤磁盘，适合于对性能与可靠性要求比较高的场所。另外一种选择是使用以太网，采取iSCSI协议，能够运行在普通的局域网环境下，从而降低了成本。由于存储区域网络中的磁盘设备并没有与某一台主机绑定在一起，而是采用了非常灵活的结构，因此对于主机来说可以访问多个磁盘设备，从而能够获得性能的提升。在存储区域网络中，使用虚拟化的引擎来进行逻辑设备到物理设备的映射，管理前端主机到后端数据的读写。因此虚拟化引擎是存储区域网络中非常重要的管理模块。

SAN系统与分布式文件系统例如Google File System并不是相互对立的系统，而是在构建集群系统的时候可供选择的两种方案。其中，在选择SAN系统的时候，为了应用程序的读写，还需要为应用程序提供上层的语义接口，此时就需要在SAN之上构建文件系统。而Google File System正好是一个分布式的文件系统，因此能够建立在SAN系统之上。总体来说，SAN与分布式文件系统都可以提供类似的功能，例如对于出错的处理等。至于如何使用还是需要由建立在云计算平台之上的应用程序来决定。

与Google不同的是，IBM并没有基于云计算提供外部可访问的网络应用程序。这主要是由于IBM并不是一个网络公司，而是一个IT的服务公司。当然，IBM内部以及IBM未来为客户提供的软件服务会基于云计算的架构。（注：本文受国家973计划资助，资助号：2007CB310900以及国家自然科学基金资助，资助号90718040）



图附-5 SAN系统的一般体系结构