



北京航空航天大学
BEIHANG UNIVERSITY

云计算与大数据平台

王宝会 北航软件学院

2020年9月

我是谁？

- 北京航空航天大学软件学院教授级高工，主要从事互联网信息安全、互联网软件架构及互联网软件性能工程理论、方法与应用研究，对机器学习、大数据、无线传感网、云计算相关技术、应用架构及实施有深入的研究。先后参与国家重大项目《增强型搜索引擎研究及示范应用》、《食品安全大数据信息安全关键技术研究》、《食品安全突发事件及重大事件应急演练及应急保障决策系统研究》以及《中国移动POC项目设计与开发》、《华为公司网络高级应用协议测试》、《新疆自治区工商系统食品在线监管》。在国内外期刊多次发表学术论文及学术著作，获得10个国家发明专利。
- 手机：13601124825
- 邮箱：wangbh@buaa.edu.cn



王宝会

中国



扫一扫上面的二维码图案，加我微信

我是导游

- 第3-13周，每周三晚上7点到9点50
- YF504
- 共11次课，计划云计算5次，大数据6次
- 第15周考试
- 考试：考勤20%+作业20%+笔试30%+机试30%

课程安排

- 1.云计算与大数据平台概述：介绍云计算以及大数据的发展历程，并从一个总的图介绍课程的主要内容，包括云计算平台部分和大数据平台部分。
- 2.云计算基础：介绍虚拟机与容器，包括Xen、KVM、Docker、Kubernetes等。虚拟化管理工具、Libvirt、XenAPI、OpenStack通用组件、消息总线、RESTful API、OpenStack通用库Oslo等云计算的基础。
- 3.计算组件：介绍OpenStack中的Nova体系结构、调度器以及资源管理服务Placement。
- 4.存储组件：介绍对象存储Swift、块存储Cinder和分布式存储Ceph。
- 5.网络组件：介绍虚拟网络功能管理Neutron、Neutron插件以及Neutron Agent。
- 6.计量与监控：介绍采集计量数据及加工预处理Ceilometer、预警和计量通知服务Aodh、资源索引和存储时序计量数据Gnocchi以及事件存储服务Panko。
- 7.安全组件：介绍OpenStack安全概述、Keystone组件以及可信计算池。

课程安排

- 8.OpenStack其他组件：介绍编排Heat、物理机管理Ironic、控制面板Horizon、部署管理工具Bifrost、Kolla、以及加速设备管理Cyborg。
- 9.大数据平台：介绍大数据平台的概述以及大数据获取组件，包括Flume、Kafka、Camel、Sqoop等
- 10.大数据存储：介绍大数据存储部分，包括文件系统HDFS、Alluxio，以及数据库HBase、Redis等
- 11.调度与管理服务：介绍大数据平台的资源管理组件Yarn、Mesos、协调与管理Ambari、ZooKeeper以及调度器Oozie、Airflow等。
- 12.大数据处理：介绍大数据的处理组件，包括批处理组件MapReduce、Spark及流处理组件Flink、Storm、Kafka Stream、Spark Streaming等
- 13.大数据查询：主要介绍大数据的查询工具Hive、Drill、Impala和Solr。
- 14.大数据安全：介绍大数据的安全组件Kerberos、Sentry及Ranger。
- 15.大数据与机器学习：介绍大数据中一些机器学习相关的组件，包括MADLib、Mahout和Spark MLlib。

几点希望

- 按时出勤，我们有一个共同目标，你懂的。
- 带神来听课，不要心猿意马。
- 师傅领进门，修行在个人。
- 学好这门课，受益三生三世，收获十里桃花。



云计算

- 随着云计算中第三方功能服务的提出，越来越多的厂商提供功能服务，改变了原来依赖于应用系统情况，像地理信息系统厂商可通过地图服务接口为其他第三方应用系统提供位置信息服务与地理定位操作，这种基于功能服务通过Web Services的模式来实现，第三方厂商无须关心具体地理信息厂商实现的技术与方法，只需要把精力放在本身应用系统的设计开发上，只需要通过功能接口的调用来开发定制出符合应用系统的相应功能。

云计算

- 云计算如同人类的大脑，是物联网的神经中枢。2019年云计算产业达到4300亿。阿里云、百度云、腾讯云、金山云、电信云...
- 云计算的快速发展离不开互联网的普及，互联网改变了用户对于使用权和所有权的认知，从而直接推动了云的落地，企业甚至不用购买硬件设备，直接基于互联网访问，按使用量购买。这并非简单的服务器硬件“资源池化”的技术变革，更推动了一场商业模式的变革---**当互联网已经成为社会的基础设施之后，云计算将是下一个基础设施。**

大数据

- 大数据提高数据的处理能力，通过处理实现数据的增值。
- 大数据与云计算就像硬币的两面一样密不可分，大数据不能由一台计算机处理，它必须由分布式体系结构处理，大数据本身就是新兴的分布式数据挖掘。
- **大数据的处理技术必须依赖于云计算的分布式数据处理以及云计算的虚拟化处理技术。**

人工智能

- 人工智能与大数据密不可分，也是基于云计算平台完成的深度学习进化。人工智能的训练过程，离不开大数据的高维度数据样本。
- 人工智能、大数据和云计算正在出现三位一体式深度融合，构成ABC金三角。三者独立、互补、相辅相成的。大数据的开发和应用离不开云计算的强大支持。云计算的发展和大数据的积累是人工智能快速发展的基础，是实现实质性突破的关键。大数据和人工智能的进步将拓展云计算的深度和广度。
- **现在是人工智能、大数据和云计算发展的黄金时期。**

物联网

- 物联网是大数据的网络载体，是互联网应用的拓展

任务四：研究全景异构感知数据融通关键技术

数据跨分区缓存与数据同步（含异步模式等）技术

设备全维度分析技术

研究快速接口集成开发技术

任务三：研究物联网平台边缘智能实现策略

应用（能力模块）协同技术

应用（能力模块）下发部署策略研究

人工智能协同技术

AI模型下发策略

任务一：研究多源多态海量异构终端泛接入互联协议体系

异构终端统一标准接入协议

异构终端物理描述模型

异构终端物理描述模型的统一解析技术

敏捷开发SDK工具

任务二：研究物联网平台大规模异构终端管理技术

超大规模管理、大并发、高可靠的平台实现

可信设备管理

OTA服务能力

云计算和大数据推动软件开发架构变化

- 架构开发方法ADM(Architecture Development Method)是以开发架构为目的，为其提供的一种可重复的测试过程。ADM包含了建立体系结构框架，开发架构内容，过渡及管理架构的实现。
- 面向服务架构（SOA）是一种架构模式，开放组织架构框架（TOGAF）是一种架构框架。许多开放组的成员一直在努力将两种概念结合。
- SOA/TOGAF实用指南项目意在TOGAF架构开发方法中开发具体的SOA扩展，该TOGAF架构开发方法是TOGAF组成的核心部分。该实用指南项目的原理是：如果SOA是一种架构模式，这样就有可能利用某种具体的SOA活动和可交付投递，扩展风格独立的TOGAF ADM方法以便生成一个服务定向ADM。

云计算和大数据推动软件开发架构变化



云计算和大数据推动软件开发模型变化

- 敏捷开发是以用户需求进化为核心，采用迭代、循序渐进的方法进行软件开发。在敏捷开发中，一个软件项目在构建初期被细分成多个子项目，各个子项目的成果都经过测试，是独立的，而又是可以合在一起变成整体的。

企业智能化应用场景

- 智能化应用本质有两点，一是减少工作量（降本），二是反哺业务（增效），说得高端一点，叫赋能。常见的应用场景需求：
- **报表自动化和智能化。**减少大量的员工工作量，明显改善信息用户对智能化应用的接收程度。通过自动出具、智能定性、穿透溯因、分析建议的目标实现，形成业务报表“透明、共享、反馈、持续改善”的良好循环。
- **单据智能化和审批智能化。**在业务逻辑支持以及合规管控下，又一个能为业务部门大量减少工作量的场景需求，最典型的应用例子莫过于近几年很火的财务RPA（财务机器人）。
- **会议材料准备和任务跟进智能化。**经营会议这种工作分析会最麻烦的是会前的材料准备和会后的任务跟进完结。对智能化应用的需求，在于会前的材料数据是否能够按规则整理输出，会后任务是否能够形成对应会议议题的任务跟进和反馈闭环，甚至纳入绩效体系。这类协同闭环的反馈需求，是很典型的提高工作效率的应用场景。

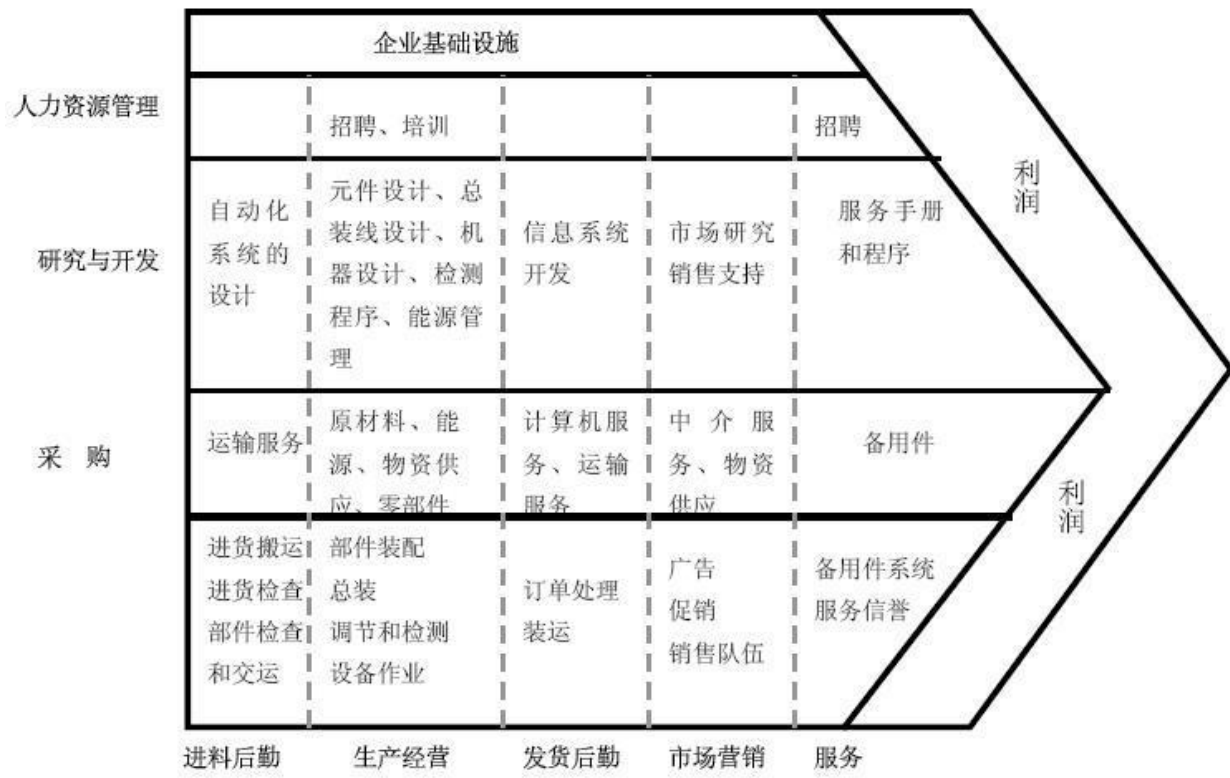
企业智能化应用场景

- **业务监控和预警体系智能化。**对于各类业务和运维问题而言，企业希望看到的是事前的状态、预警、管控处理建议，例如本周销量同比环比或趋势线进入预警区间，又例如库存库龄或周转率出现异常变化，再例如在超维护期引起的OEE及效率下降和加班生产之间如何选择等等，智能化应用提供的是对业务的精益管控和增效保障，而不是事后的各类原因分析报告和改善措施。
- **产销协同计划模式的智能化**
- **渠道管理与服务能力的智能化**
- **产品设计与工艺路线的智能化**

波特价值链分析模型

- 美国哈佛商学院著名战略学家迈克尔·波特提出的“价值链分析法”，把企业内外价值增加的活动分为基本活动和支持性活动，基本活动涉及企业生产、销售、进料后勤、发货后勤、售后服务。支持性活动涉及人事、财务、计划、研究与开发、采购等，基本活动和支持性活动构成了企业的价值链。不同的企业参与的价值活动中，并不是每个环节都创造价值，实际上只有某些特定的价值活动才真正创造价值，这些真正创造价值的经营活动，就是价值链上的“战略环节”。企业要保持的竞争优势，实际上就是企业在价值链某些特定的战略环节上的优势。运用价值链的分析方法来确定核心竞争力，就是要求企业密切关注组织的资源状态，要求企业特别关注和培养在价值链的关键环节上获得重要的核心竞争力，以形成和巩固企业在行业内的竞争优势。企业的优势既可以来源于价值活动所涉及的市场范围的调整，也可来源于企业间协调或合用价值链所带来的最优化效益。

波特价值链分析模型



复印机生产企业的价值链

云计算的前世今生

The Next Frontier - On Demand Solutions For Your Business

Cloud

Web

Client-Server Computing

Personal Computers

Mainframe

云计算是根据面向需求的服务发展起来的，2007年10月IBM和Google公布了它们在云计算领域的合作计划后，云计算的概念才为人们所熟知。特别是IBM的“蓝色的云”计划推出后，云计算成为业界无人不谈的热门话题，维基百科也顺天应人，在网站中加入了该词条。

云计算改变了IT基础架构的部署模式

- 传统的烟囱式模式，相互孤立，利用率低
- 云计算基础架构将服务器、存储、网络形成资源池，扩展灵活，部署方便

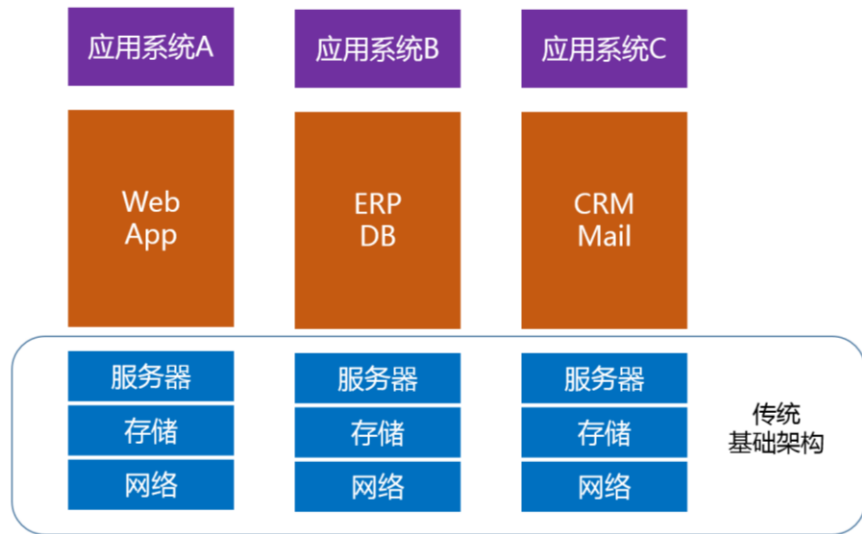


图1 传统的烟囱式信息系统

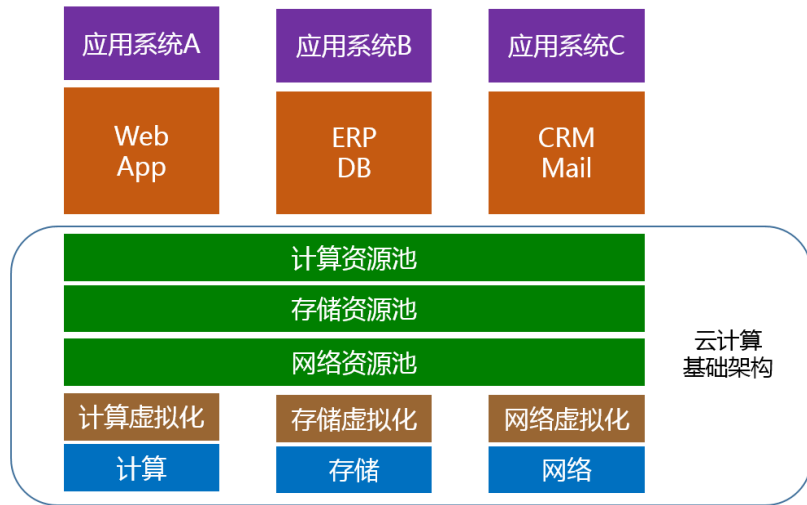
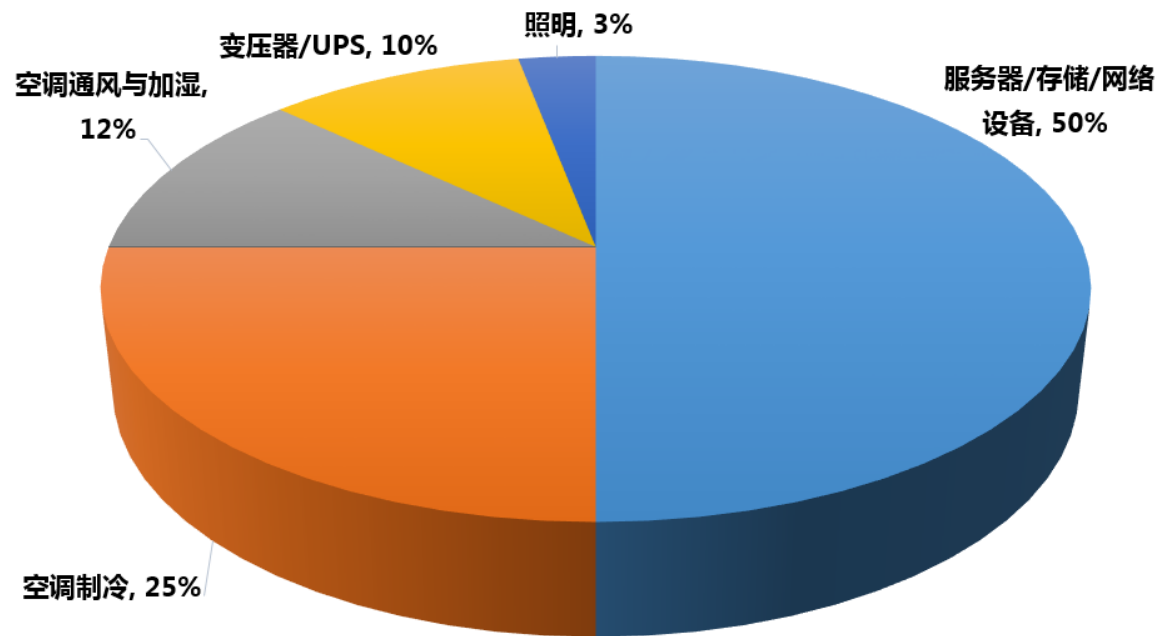


图2 云计算基础架构的部署模式

云计算数据中心的能耗分析



- 电费支出占总成本的60%-80%
- 我国的数据中心的PUE在2.5左右
- 节能降耗是迫切需求

什么是云计算

- 狭义云计算：提供资源的网络被称为“云”。“云”中的资源在使用者看来是可以无限扩展的，并且可以随时获取，按需使用，随时扩展，按使用付费。这种特性经常被称为像水电一样使用IT基础设施。
- 广义云计算：这种服务可以是IT和软件、互联网相关的，也可以是任意其他的服务。这种资源池称为“云”。“云”是一些可以自我维护 and 管理的虚拟计算资源，通常为一些大型服务器集群，包括计算服务器、存储服务器、宽带资源等等。云计算将所有的计算资源集中起来，并由软件实现自动管理，无需人为参与。这使得应用提供者无需为繁琐的细节而烦恼，能够更加专注于自己的业务，有利于创新和降低成本。

什么是云计算

- 云计算是并行计算、分布式计算和网格计算的发展，是这些计算机科学概念的商业实现
- 云计算基本思想就是通过网络将分散的、成本较低的计算资源如存储器、CPU、磁盘等进行整合，为最终用户提供功能强大的计算服务。
- 云计算是一种模式，该模式允许用户通过无所不在的、便捷的、按需获得的网络，接入到一个可动态配置的共享计算资源池（其中包括网络设备、服务器、存储、应用以及业务），并且以最小的管理代价或业务交互复杂度即可实现这些可配置计算资源的快速发放与发布。
- 云计算提供了计算和存储模型的高级抽象。它具有有一些**基本特征、服务模型和部署模型**
- 云计算中，云指经过整合的能够自我维护 and 管理的虚拟计算资源。在目前的实现中，通常为一些大型服务器集群，包括计算服务器、存储服务器、宽带资源等等。
- 云计算开启了一种基于Internet的新的应用模式，人们购买的可能不是性能各异的计算机，而是网络上经过整合的各种不同的计算资源。在新的应用模式下，一方面用户并不真正拥有这些资源，而是通过第三方服务供应商租用这些资源，从而降低了购买硬件的费用和配置软件的时间成本。另一方面，云计算服务供应商利用虚拟化、负载均衡等技术最大化的发挥各计算资源的效能，从而提高了资源利用率。

什么是云计算

- 资源与服务的动态可伸缩以及资源的虚拟化
- 对用户透明
- 虚拟化: software, databases, Web servers, operating systems, storage and networking as virtual servers
- 按需: add and subtract processors, memory, network bandwidth, storage
- 云计算底层实现的两个关键技术路线: 虚拟化和分布式

云计算系统五个层次

- **云客户端**可以是计算机、移动电话等硬件，也可以是操作系统、浏览器等软件。它是云计算系统的服务对象。最终用户通过客户端向云端发送请求并接受云计算服务。
- **云应用程序**使运行软件成为一种服务。这些应用程序在云计算系统而不是在用户本地运行，由云计算系统为其分配资源并维护其运行状态。用户通过网络获取运行结果。
- **云计算平台**是云应用程序运行的平台。它为云应用程序分配和调度计算资源。
- **云基础设施**实现了对软硬件的虚拟化，使云计算平台不用关心处理器、存储器、数据库、网络设备、软件等应如何组织。
- **云服务器**指实际的计算机软硬件，包括多核处理器、面向云计算的操作系统等等。这些软硬件能够提供云计算服务。

云计算的基本特征

- **灵活性**: 在花费代价很小的情况下, 为用户快速的重新配置资源。
- **应用程序兼容性**: 用户可以和调用本地API一样调用云端软件提供的API。
- **低花费**: 采用“租用”的方式使用第三方服务供应商提供的硬件资源, 用户既节省了购买硬件的费用, 又节约了配置软件的时间。
- **硬件设备位置无关性**: 通过网络组织各种硬件资源, 用户只需要通过Web访问这些资源, 不必关心到底使用了哪些硬件设备, 也不必关心这些硬件设备在何处。
- **资源共享**: 区别于以前资源和用户一对一的服务方式, 以一对多的方式为用户提供服务, 提高了资源利用率, 实现了资源共享。
- **可用性**
- **可靠性**: 为用户分配冗余的资源, 能够进行灾难恢复。
- **可扩展性**: 云端资源可以动态地扩展, 省却了用户的配置工作。
- **易维护性**: 对于最终用户来说, 只进行Web访问, 不用安装客户端, 易于维护。
- **可测量性**: 云系统通过在与服务类型相适应的抽象级别上利用度量功能, 自动控制和优化使用的资源。它将提供可分析和可预测的计算平台。为用户提供的各种资源量可以按照每个用户或每个应用的使用天数或月数、年数来计算。
- **按需应变自助服务**: 消费者可以单方面提供计算能力, 而不需要与每个服务提供者进行人工交互。
- **异构访问**: 通过网络获得功能, 并通过标准机制进行访问, 这些机制促进异构瘦客户机或胖客户机平台的使用。
- **资源池**: 提供者使用的的计算资源池多租户模型为多个使用者提供服务; 根据用户需求动态分配和重新分配不同的物理和虚拟资源。

云计算的基本特征



Create an AWS account for free.
Only pay for what you use.

[Sign Up Now](#)

Products & Services

[View all products & services »](#)

Compute

Scale to meet your application demands, whether one server or a large cluster. Choose from 10+ instance sizes and a variety of operating systems.

- ✦ [Amazon Elastic Compute Cloud \(EC2\)](#)
- ✦ [Amazon Elastic MapReduce \(EMR\)](#)

Storage

Utilize cost-effective solutions for storing and retrieving any amount of data, any time, anywhere.

- ✦ [Amazon Simple Storage Service \(S3\)](#)
- ✦ [Amazon Elastic Block Store \(EBS\)](#)

Database

Leverage scalable database solutions, from managed MySQL or Oracle, hosted enterprise database software, or non-relational database solutions.


- ✦ [Amazon DynamoDB](#)
- ✦ [Amazon Relational Database Service \(RDS\)](#)

Networking

Customize and control your network resources, both inside and outside the cloud.

- ✦ [Amazon Virtual Private Cloud \(VPC\)](#)
- ✦ [Amazon Route53](#)

云计算的基本特征



Region:	<input type="text" value="US East (Virginia)"/>	
	Linux/UNIX Usage	Windows Usage
Standard On-Demand Instances		
Small (Default)	\$0.080 per Hour	\$0.115 per Hour
Medium	\$0.160 per Hour	\$0.230 per Hour
Large	\$0.320 per Hour	\$0.460 per Hour
Extra Large	\$0.640 per Hour	\$0.920 per Hour

云计算的基本特征



云计算的基本特征



Microsoft Office 365



Email and calendars

Access your email and calendar from your computer, the web, and phone.*
[Watch the video >](#)

- Professional cloud-based email
- 25 GB of mailbox storage for each user
- Share calendars and schedule meetings



Office Web Apps

Create, store, and edit Office docs online.
[Watch the video >](#)

- Includes Word, PowerPoint, Excel, and OneNote Web Apps
- Edit and create files using a browser
- Open and edit files created in Office**



Web Conferencing

Schedule online meetings in seconds.
[Watch the video >](#)

- Connect with video conferencing
- Give live online presentations
- Share desktops and co-present



File sharing

Manage content and collaborate online.
[Watch the video >](#)

- Share files with customers and your team
- 10 GB storage for your company plus 300 MB per user



Website

Launch a professional website for your business.
[Watch the video >](#)

- No code required to easily create and publish your website



Mobility

Manage content and collaborate online.
[Watch the video >](#)

- Retrieve and send email, contact info, and appointments with Windows Phone, iPhone, iPad, Android, Symbian, and BlackBerry devices



谢谢!