云计算：虚幻还是现实？

中国科学院计算技术研究所

先进计算机系统实验室

詹剑锋

2012年6月12日

云：诗意的存在

大地：坚实广阔

中科院计算所

云计算什么时候落地？

提纲

- 什么是云计算?
  - 为什么会有云计算?
- 现状
- 我对云计算的一些理解

众说纷纭Cloud

- **If you have a hammer, everything looks like a nail.**

- Service computing community
  - XX as a service
    - Software, platform, infrastructure
- Grid computing community
  - 新瓶装旧酒
- High performance computer community
  - HPC in Cloud
- Data management community
  - 数据为王!
- System architecture community
  - 新的负载特性! 新的体系结构!

Cloud定义就是一篇高引用论文！

Author/Reference	Year	Definition/Excerpt
M. Klems [11]	2008	*you can scale your infrastructure on demand within minutes or even seconds, instead of days or weeks, thereby avoiding under-utilization (idle servers) and over-utilization (blue screen) of in-house resources...*
P. Gaw [11]	2008	*using the internet to allow people to access technology-enabled services. Those services must be 'massively scalable...*
R. Buyya [6]	2008	*A Cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers*
R. Cohen [11]	2008	*Cloud computing is one of those catch all buzz words that tries to encompass a variety of aspects ranging from deployment, load balancing, provisioning, business model and architecture (like Web2.0). It's the next logical step in software (software 10.0). For me the simplest explanation for Cloud Computing is describing it as, "internet centric software...*
J. Kaplan [11]	2008	*a broad array of web-based services aimed at allowing users to obtain a wide range of functional capabilities on a 'pay-as-you-go' basis that previously required tremendous hardware/software investments and professional skills to acquire. Cloud computing is the realization of the earlier ideals of utility computing without the technical complexities or complicated deployment worries...*
D. Gourlay [11]	2008	*...the next hype-term...building off of the software models that virtualization enabled*
D. Edwards [11]	2008	*...what is possible when you leverage web-scale infrastructure (application and physical) in an on-demand way...*
B. de Haff [11]	2008	*...There really are only three types of services that are Cloud based: SaaS, PaaS, and Cloud Computing Platforms. I am not sure being massively scalable is a requirement to fit into any one category.*
B. Kepes [11]	2008	*...Put simply Cloud Computing is the infrastructural paradigm shift that enables the ascension of SaaS. ... It is a broad array of web-based services aimed at allowing users to obtain a wide range of functional capabilities on a pay-as-you-go basis that previously required tremendous hardware/software investments and professional skills to acquire*
K. Sheynkman [11]	2008	*Clouds focused on making the hardware layer consumable as on-demand compute and storage capacity. This is an important first step, but for companies to harness the power of the Cloud, complete application infrastructure needs to be easily configured, deployed, dynamically-scaled and managed in these virtualized hardware environments*
O. Sultan [11]	2008	*...In a fully implemented Data Center 3.0 environment, you can decide if an app is run locally (cook at home), in someone elses data center (take-out) and you can change your mind on the fly in case you are short on data center resources (pantry is empty) or you having environmental/facilities issues (too hot to cook). In fact, with automation, a lot of this can can be done with policy and real-time triggers...*
K. Hartig [11]	2008	*..really is accessing resources and services needed to perform functions with dynamically changing needs...is a virtualization of resources that maintains and manages itself.*
J. Pritzker [11]	2008	*Clouds are vast resource pools with on-demand resource allocation...virtualized ...and priced like utilities*
T. Doerksen [11]	2008	*Cloud computing is ... the user-friendly version of Grid computing*
T. von Eicken [11]	2008	*outsourced, pay-as-you-go, on-demand, somewhere in the Internet, etc*
M. Sheedan [11]	2008	*...'Cloud Pyramid' to help differentiate the various Cloud offerings out there...Top: SaaS; Middle: PaaS; Bottom: IaaS*
A. Ricadela [11]	2008	*...Cloud Computing projects are more powerful and crash-proof than Grid systems developed even in recent years*
I. Wladawsky Berger [11]	2008	*...the key thing we want to virtualize or hide from the user is complexity...all that software will be virtualized or hidden from us and taken care of by systems and/or professionals that are somewhere else - out there in The Cloud*
B. Martin [11]	2008	*Cloud computing encompasses any subscription-based or pay-per-use service that, in real time over the Internet, extends IT's existing capabilities*
R. Bragg [5]	2008	*The key concept behind the Cloud is Web application... a more developed and reliable Cloud. Many find it's now cheaper to migrate to the Web Cloud than invest in their own server farm ... it is a desktop for people without a computer*
G. Gruman and E. Knorr [14]	2008	*Cloud is all about: SaaS...utility computing...Web Services... PaaS...Internet integration...commerce platforms....*
P. McFedries [22, 15]	2008	*Cloud Computing, in which not just our data but even our software resides within the Cloud, and we access everything not only through our PCs but also Cloud-friendly devices, such as smart phones, PDAs... the megacomputer enabled by virtualization and software as a service...This is utility computing powered by massive utility data centers.*

Vaquero, L. M., Rodero-Merino, L., Caceres, J., and Lindner, M. 2008. A break in the clouds: towards a cloud definition. *SIGCOMM Comput. Commun. Rev.* 39, 1 (Dec. 2008), 50-55.

Cloud重要术语和定义

- Data-intensive scalable computing (DISC)
  - CMU计算机学院院长. Randal E. Bryant
- Above the cloud
  - Berkeley. UC Berkeley RAD Lab.
- Warehouse-scale computer
  - Googler. The data center as a computer

Luiz André Barroso and Urs Hölzle

Data-intensive scalable computing (DISC)

- 在这个研究领域，大学正处于一个**相反的位置.**
  - Follow 工业界，而不是通常的研究领先于产业的顺序。
- **Google和它的竞争者证明了一种新的计算模式。**大学能够学习和拓展这些想法非常重要。
- 2007年5月，CMU计算机学院院长Randal E. Bryant

伯克利云计算视图

- **Utility Computing: pay-as-you-go computing**
  - Illusion of infinite resources
  - Fine-grained billing (e.g. hourly)

- 云计算是SaaS和Utility Computing的结合。
  - Public Cloud.
  - Private Cloud.
- Berkeley, UC Berkeley RAD Lab， 2009年2月

Google: warehouse-scale computer (WSC)

- The massive scale of their software infrastructure, data repositories, and hardware platform.
  - 通常属于一个组织。
  - 使用相对同构的硬件和软件系统平台。
  - 应用、中间件和系统软件通常自行开发，而不是从第三方购买。
  - 运行数目少，但规模非常大的互联网服务。
  - 公共的资源管理设施，支持快速灵活部署。

Luiz André Barroso and Urs Hölzle

Vaquero et al. 的Cloud 定义

- Cloud是大量容易访问和使用的虚拟资源池（如硬件、开发平台或者服务）.
- 资源可以根据变化的负载规模动态配置。
- 通常使用pay-per-use model.
- Infrastructure provider为Service provider提供定制的SLA保证。

A break in the clouds: towards a cloud definition. *SIGCOMM Comput. Commun. Rev.* 39, 1 (Dec. 2008), 50-55.

Cloud分类

- Public Cloud
- Private Cloud
- Hybrid Cloud

Cloud summary

资源提供者:

IaaS

用户:

SaaS

开发人员:

PaaS

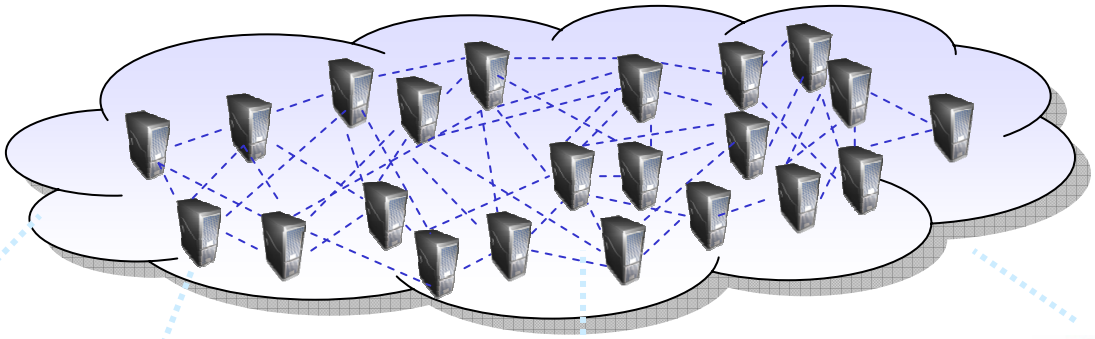
学术界:

DISC/Utility/Cloud

提纲

- 什么是云计算？
  - 为什么会有云计算？
- 现状
- 我对云计算的一些理解

Cloud的远景



TVs

PCs

PDAs

4+ billion phones *by 2010*
[Source: Nokia]

发展动力

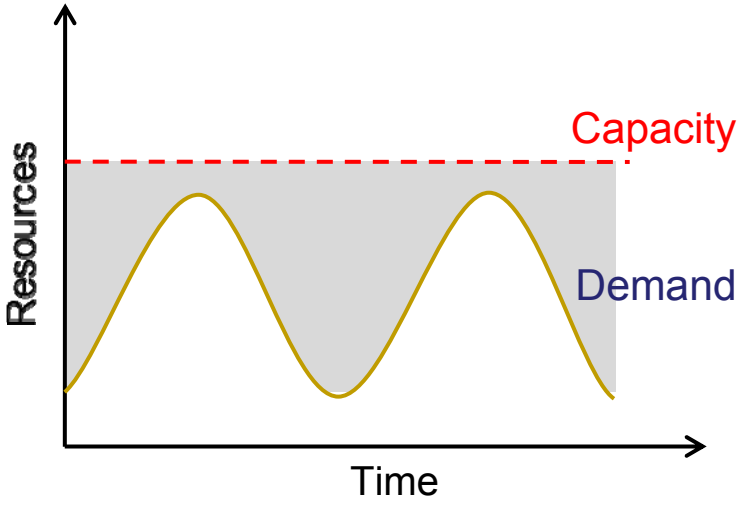
- 提高用户体验
  - 管理容易，无需配置、备份
  - Ubiquity of access (带浏览器的移动设备)
- 给厂商带来的好处
  - 软件更新简化
  - 需要适用的平台变少。
    - 猜猜**Microsoft**软件测试的机器总数。
- 规模化经济
- 导致新一轮计算范式改变
  - server-side computing

案例:搜索引擎的计算需求

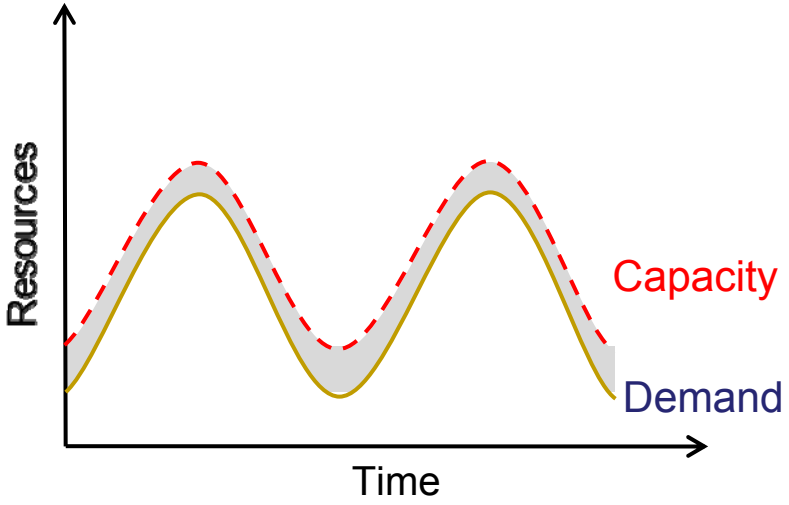
- 访问规模
  - 计算设备的普及，导致不断增长;
- 数据规模
  - the Web is growing by millions of pages per day
- 市场竞争
  - 不断提高检索结果的质量，更新索引的频率（实时性）
- 保证服务质量，控制成本!

Cloud 用户的经济学

- 按需付费，而不是provisioning for peak
- statistical multiplexing



Static data center

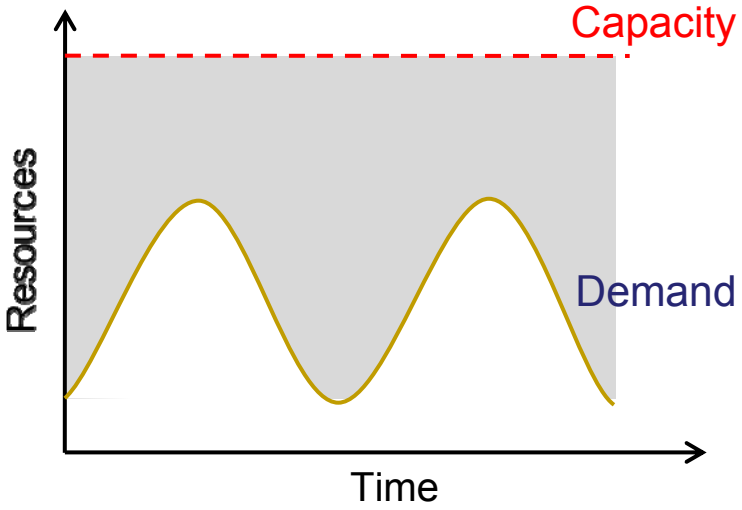


Data center in the cloud

Unused resources

Cloud 用户的经济学

- 过高配置的风险： underutilization

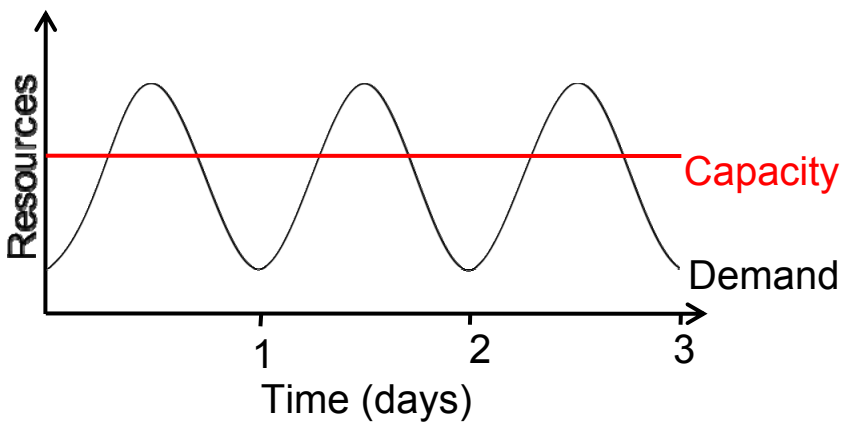


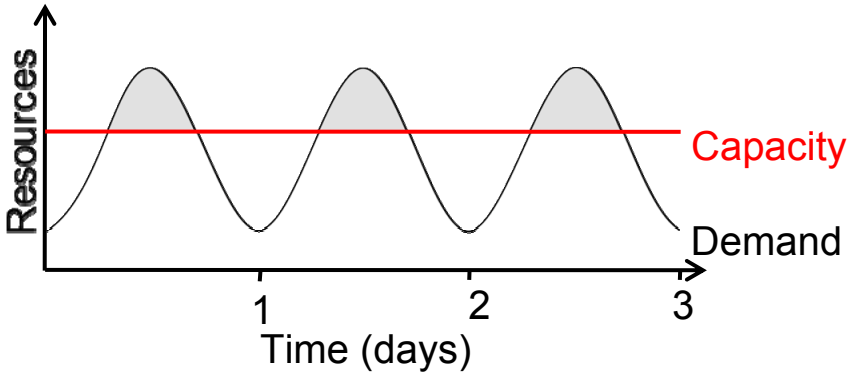
Unused resources

Static data center

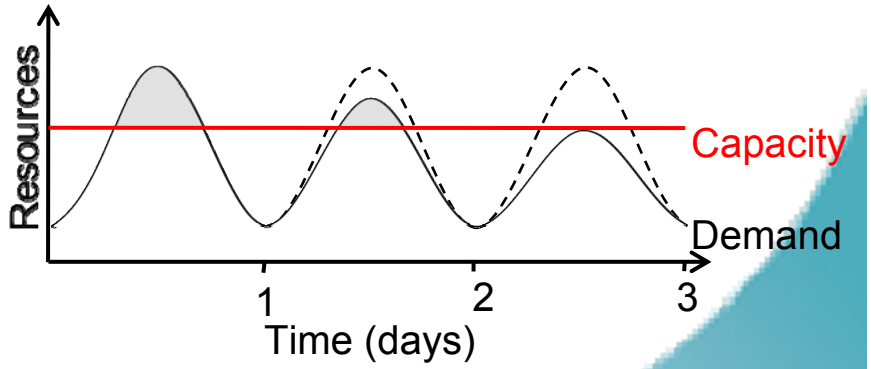
Cloud 用户的经济学

- 过低配置的带来的损失





Lost revenue



Lost users

提纲

- 什么是云计算？
  - 为什么会有云计算？
- 现状
- 我对云计算的一些理解

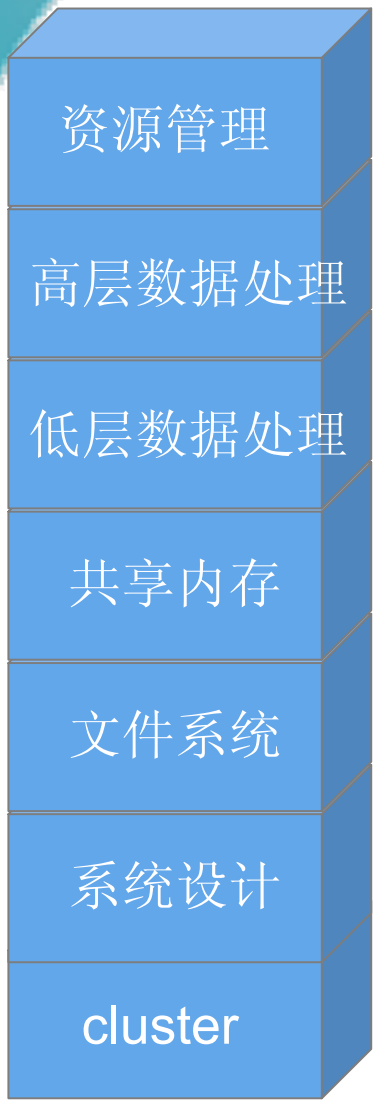
现状

- **Major players**
  - **Google/Yahoo/Microsoft**
  - Amazon
  - IBM
- 学术界的那些事

Google 为什么反对hardware specialization?

- 不同服务需求多样化!
  - Google 不愿意公开Benchmark和Workload trace
- workload 变化快!
  - 产品需求快速演变!
  - smart programmers
    - rewrite the baseline algorithms and data structures much more rapidly than hardware itself can evolve.
- Server 更新快! 3-4年
  - 让软件适应变化的硬件。

软件基础设施



Google	YAHOO!	Microsoft
GWQ	HOD	
Sawzall	Pig	Scope
MapReduce	Hadoop	Dryad
BigTable	HBase	CloudDB, SSDS
GFS	HDFS	Cosmos

现状

- **Major players**
  - Google/Microsoft
  - **Amazon**
  - IBM
- 学术界的那些事

Amazon代表性系统

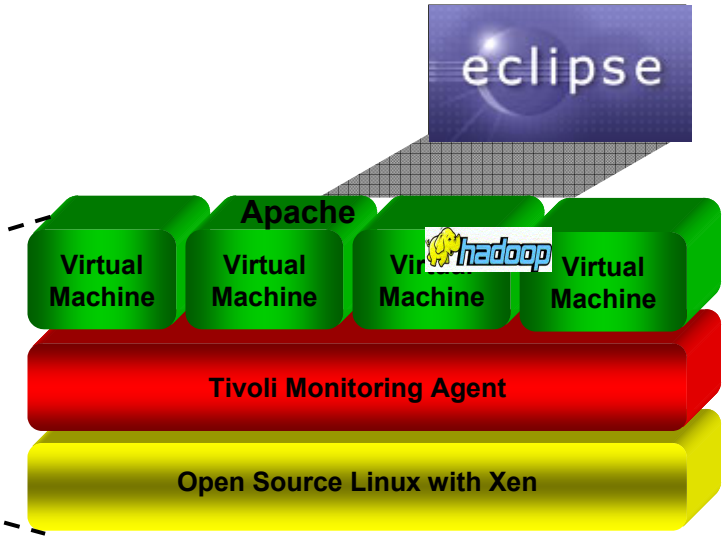
- 弹性计算云EC2(Elastic Computing Cloud)， $0.10 VM instance/hour
- 简单存储服务，S3(Simple Storage Service)，$0.15c/month/GB
- 简单数据库，SimpleDB
- 简单排队服务，SQS (Simple Queue Service)
- **DropBox**
  - 基于S3
- 托管环境，平台服务化，效用计算

现状

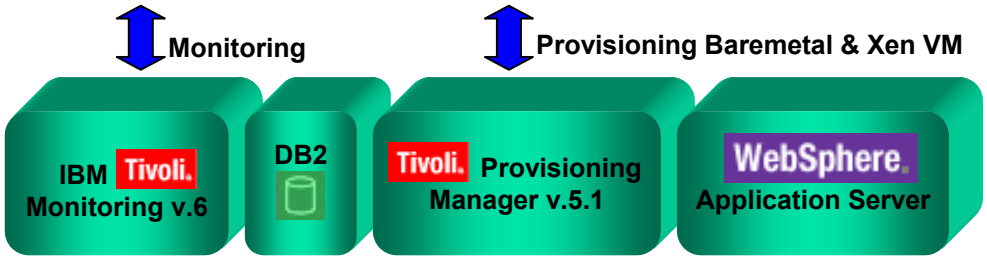
- **Major players**
  - Google
  - Amazon/Yahoo
  - **IBM**
- 学术界的那些事

IBM战略：向企业销售云计算软硬件设备

Data Center – System x



Virtualized Infrastructure based on Open Source Linux & Xen



Provisioning Management Stack

现状

- Google
- Amazon
- IBM
- 学术界的那些事

代表性事件(1)

- Google, Map Reduce, GFS, Big Table
  - SOSP, OSDI
- Hadoop
  - Yahoo's 开源软件
- CMU
  - DISC, data-intensive super/ scalable computing
- Amazon
  - EC2, S3
- DropBox
  - S3+add-on services

代表性事件(2)

- SUN/Oracle
  - (T1, T2, T3)/Rock, T4
- Conference:
  - The first workshop of Cloud computing and its application: http://*www.***cca08***.org* , Ian Foster etc.
  - USENIX Hot topic on Cloud Computing
  - **ACM Symposium on *Cloud Computing* (SoCC)**
  - Two Cloud conference , 争夺话语权!
- Berkley ‘s Technical Report: above the Cloud
- Google: The data center as a computer.

提纲

- 什么是云计算？
  - 为什么会有云计算？
- 现状
- 我对云计算的一些理解

云计算的形与神

- 云计算从形态来说是传统的高端计算的普及化。

- 本质是资源的统计复用(Statistical Multiplexing)，利用规模经济特性降低成本。

云计算是高端计算的普及化

- 成本驱动
- 云计算的几个核心技术都可以从高端计算中发现影子，
  - 虚拟化技术，早在上世纪60年代IBM的上就已成熟。
  - Cluster技术本身就是高端计算普及化的产物。
    - 在云计算浪潮中，集群的应用跳出了科学计算的圈子，进一步应用于数据处理和分析。
  - MPI→MapReduce
  - 资源管理

云计算本质是资源的统计复用

- 水、电、交通设施称为Utility，云计算同样具有“Utility”特性。
- 面对海量用户同时计算申请资源，机器负载在不同的时段波动性很大
- 对资源的需求、规模、计算时间的长短进行统一调配
- 将波动较大的资源曲线变平
  - 降低尖峰负载->降低成本!

云计算的数据价值

- 云计算最大的贡献是在最低端的PC服务器借助容错的系统软件进行数据分析
  - 一些大的互联网公司（数据公司）的趋势是把所有的软件开源。
  - 单个数据中心处理能力将从PB到EB

- 数字化世界的数据规模
  - 2011，总数据 1.8 ZB

统一的云计算标准

- 如果计算、存储都向服务化方向发展，统一的标准是必要的。
- 存储云
  - 如果存储接口不一致，无法在不同的服务提供商之间迁移数据。
  - 用户利益受损。
- 政府主导不如市场自然选择。

云计算安全问题

- 公共云发展的主要障碍。
- 并不是所有的计算对安全性要求都很高。
  - 只要确实能降低成本，仍然会有用户尝试。
- 安全问题并不是云计算系统的独有产物。

国内云计算发展状况

- Hadoop应用： 互联网公司应用成功
  - 部分公司也在此基础上推出了加强版本，
  - 如用C++重写部分Java代码，
  - 针对特定应用开发文件系统甚至定制服务器。
- 公共优势在于计算云： EC2-like system
  - 不等同于云计算
- DropBox-like systems
    - 百花齐放，但还未产生市场支配性企业。
- 传统高性能计算中心在**鼓吹发展**HPC in cloud，对异构负载可能有效。

优势

- 海量的中文用户（独特环境）和不断培育发展的用户需求。
  - 深刻理解用户需求的公司可能产生创新应用，在应用的驱动下发展创新技术。

云化策略

- 设备先行不可取。
- 新的中小型企业
  - 技术功底好的企业利用Hadoop技术自己建系统
  - 寻求托管外包
- 传统企业
  - 如果没有足够的IT基础设施，并且应用不具有很高的可用性要求，可以借助公有云来节省计算资源，
    - 《纽约时报》就租用亚马逊的机器进行大量报纸的数字化。
  - IT比较成熟，增量部分可以采用云计算通过Hadoop、虚拟机、服务器聚集、NoSQL等技术进行数据分析，来产生倍乘效应。
  - 混合云是值得尝试。
- 政府部门，IT服务可以通过托管和外包来实现;

CIO的重要能力

- 最重要的是要具备风险控制能力，亦即新技术的评估和消化能力
  - 不能没有人消化技术就上技术
  - 要组建一个强有力的成熟的技术团队;
  - 要以应用为驱动，不能没有应用就上技术。
- CIO要具备成本意识，
  - 主要是人的成本和基础设施成本，如租用机房、空调、供电，服务器本身的成本及功耗的成本、耗电，都要在很具体的场景下做成本分析。
- CIO还应具备全局的性能评估能力，
  - 要知道自己的应用系统瓶颈在哪里，比如说是存储的问题还是计算的问题等等。

Reading lists

Vaquero, L. M., Rodero-Merino, L., Caceres, J., and Lindner, M. 2008. A break in the clouds: towards a cloud definition. SIGCOMM Comput. Commun. Rev. 39, 1 (Dec. 2008), 50-55.

Jianfeng Zhan, Lei Wang, Xianna Li, Weisong Shi, Chuliang Weng, Wenyao Zhang, and Xiutao Zang. Cost-aware Cooperative Resource Provisioning for Heterogeneous Workloads in Data Centers. Accepted by IEEE Transaction on Computers (TC).

U. Hoelzle et al. 2009. The Datacenter as a Computer: an introduction to the Design of Warehouse-Scale Machines. 1st. Morgan and Claypool Publishers.

Armbrust, M., et al. Above the clouds: A Berkeley view of cloud computing. Tech. Rep. UCB/EECS-2009-28, EECS Department, UC Berkeley, Feb 2009

谢谢

个人联系方式

- 微博 http://weibo.com/jfzhan
- 主页 http://prof.ncic.ac.cn/jfzhan
- 邮件 zhanjianfeng@ict.ac.cn