“Computing Heads for the Clouds”

BEBO WHITE
HKU ECOM-ICOM EXPERTS ADDRESS
HONG KONG
JANUARY 19, 2010

InformationWeek
BUSINESS INNOVATION POWERED BY TECHNOLOGY

IBM Launches Cloud Computing Lab

Hong Kong center will focus on support for Big Blue's LotusLive online service.

By Paul McDougall, InformationWeek
Dec. 10, 2009
URL: http://www.informationweek.com/story/showArticle.jhtml?articleID=222001544

IBM said Thursday that it opened a research lab in Hong Kong dedicated to developing products and services for the cloud computing market. At the outset, the lab will dedicate resources to supporting IBM's online LotusLive offering.

IBM currently counts more than 18 million users of LotusLive, which offers Web-based access to e-mail, instant messaging, calendaring, and other collaboration tools. The service starts at $3.00 per month, per user.

IBM said the Hong Kong center will research and develop best practices around various issues related to cloud computing, including security, privacy, and reliability. The lab is co-located with IBM's existing China Development Laboratory, which employs more than 5,000 developers.

"As the first cloud computing laboratory in Hong Kong to serve as a global resource for cloud-based collaboration services, the laboratory marks a milestone for IBM and for the information technology industry in Hong Kong," said Dominic Tong, IBM's general manager for Hong Kong/China, in a statement.

"The opening of the laboratory demonstrates Hong Kong's advantage as a global hub for world-class information technology and online services and we are delighted that it aligns with the government's agenda of developing Hong Kong into a center of excellence in innovation and technology," said Tong.

The Top 10 Strategic Technologies for 2010 – Gartner Report

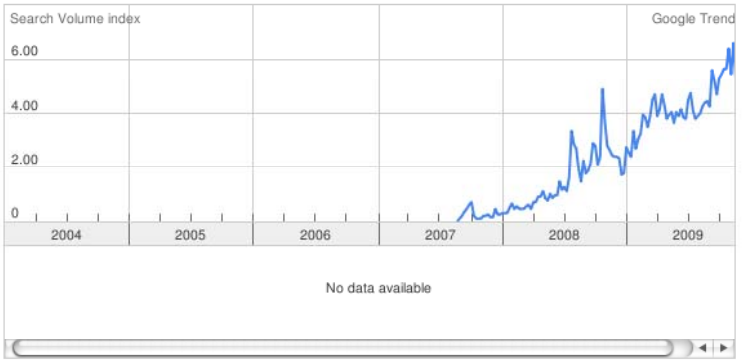
1. Cloud Computing ⟵
2. Advanced Analytics
3. Client Computing (Virtualization) ✔
4. IT for Green
5. Reshaping the Data Center ✔
6. Social Computing
7. Security – Activity Monitoring
8. Flash Memory
9. Virtualization for Availability ✔
10. Mobile Applications

Google trends "cloud computing" Search Tren
Tip: Use commas to compare multiple search terms.

Searches Websites All regions

Scale is based on the average worldwide traffic of "cloud computing" in all years. Learn more

"cloud computing" 1.00



Cloud Computing

- A style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet
- Users need not have knowledge of, expertise in, or control over the technology infrastructure in the "cloud" that supports them

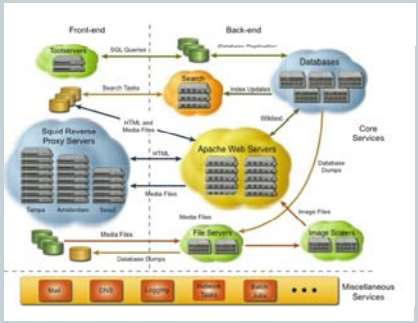
(Wikipedia)

Perfect Storm Fueling Cloud Computing

- Growth of Internet usage
  - Broadband networking
  - Mobile, location-aware services
  - Self-service
- Massive data – horizontal scale
  - User-generated content, digital media
  - Even more data ahead – environmental monitoring
- Moore's Law driving down cost of computing and storage
  - Low cost servers and disk drives
  - Consumer devices: smartphones, netbooks, gaming consoles
  - Enables new capabilities: speech, NLP, semantics, etc.

Internet Scale Demands Have Driven New Application Design Patterns

- Very large, horizontal scale systems built from commodity components
- Component failure must be handled gracefully
- Non-traditional, highly replicated data storage and caching solutions
- Applications built from cooperating set of services



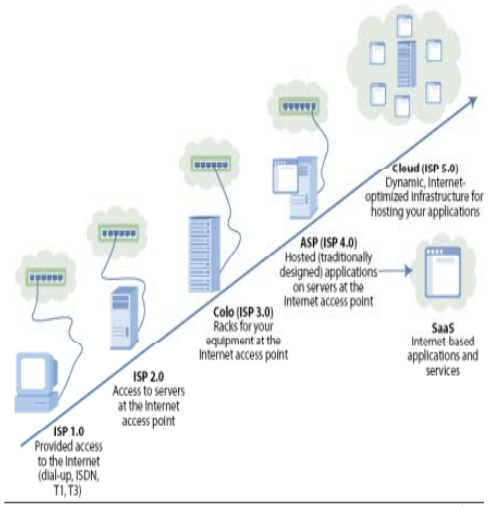
Age of "Warehouse Scale" Machines


Google's data center on the Columbia river, Oregon

•Thousands and thousands of commodity parts built into a system to essentially serve a single application
•Power and cooling are the major drivers of cost

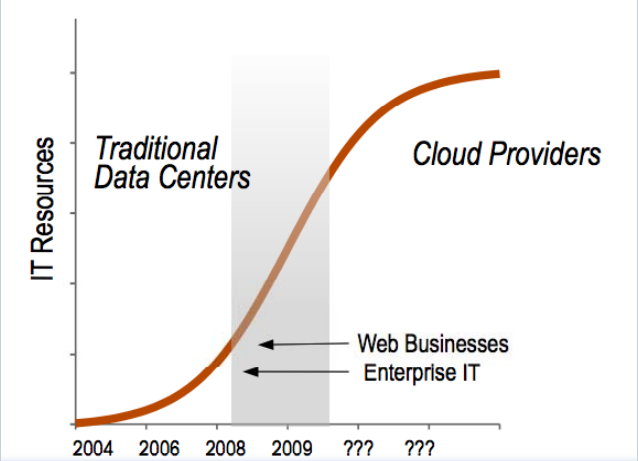
Age of "Containerized" Datacenters

Figure 3 Cloud Computing: The Latest Evolution Of Hosting



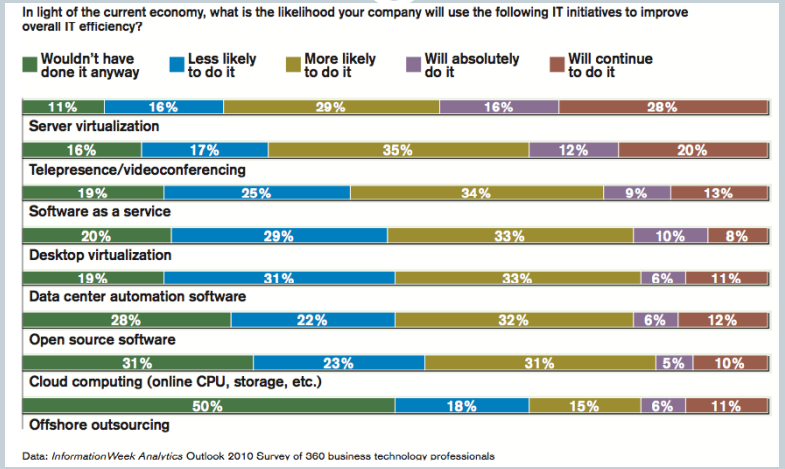
44229 Source: Forrester Research, Inc.

Adoption Curve of Cloud Computing?



Source: Sun Microsystems)

Current Strategies for IT Efficiency



So Back to Clouds

- What is cloud computing?
  - When computing services are provided over the Internet rather then locally on a user's own machine
  - Computation is run on an supporting infrastructure which is independent of the applications themselves
  - The infrastructure can take on many forms, but to the end user, the implementation is irrelevant, hence the "cloud" abstraction
- What *isn't* cloud computing?
  - Cloud computing does not *necessarily* include grid computing, resources as a utility, or self managing computing
  - Each of these can however be used in some cloud computing systems, but cloud computing can also be done with free and decentralized architectures

What is New in Cloud Computing

- The illusion of infinite computing resources
- The elimination of an up-front commitment by users
- The ability to use and pay on demand

- Cloud Computing vs. P2P?
  - Both take advantage of remote resources
  - P2P: does not use clouds (datacenters), peers do not get paid, lower reliability
- Cloud Computing vs. Grid Computing?
  - Both use clouds
  - Grid Computing requires commitment, share based on common interests. Not public cloud

Is Cloud Computing Real? (1/2)

- *I don't understand what we would do differently in the light of Cloud Computing other than change the wordings of some of our ads*

  Larry Ellison, Oracle's CEO
- *I have not heard two people say the same thing about it [cloud]. There are multiple definitions out there of "the cloud"*

  Andy Isherwood, HP's Vice President of European Software Sales
- *It's stupidity. It's worse than stupidity: it's a marketing hype campaign.*

  Richard Stallman, Free Software Foundation founder

Is Cloud Computing Real? (2/2)

- *On-demand environments like clouds are poised to become the default deployment environment for distributed and scale-out compute infrastructures*

  Erik Troan, CTO rPath
- *69% of all Internet users have either stored data online or used a Web-based software application*

  Pew Research Center, September 2008

Cloud Service Providers "Level the Field"

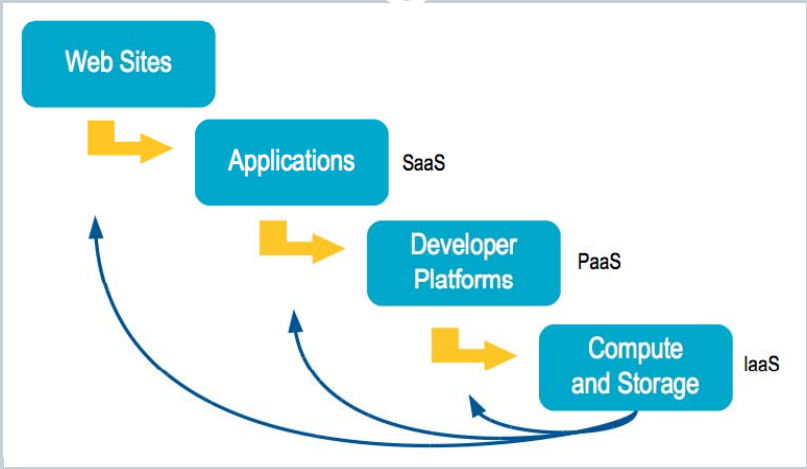
- Anyone can gain efficiencies of large scale
- Pay-as-you go, pay only for what you need
- Automation and programmatic API control
- Scale up, scale down
- Better agility, faster response, more innovation
- "Let me be very clear here: I don't want to operate datacenters anymore…We'd rather spend our time giving our customers great service and writing great software rather than managing physical hardware" --- Don MacAskill, CEO, Smugmug

What Makes a Cloud?

- Virtual Machines
- VM Manager
  - Scalability
- File system Infrastructure
- Remote access (portal)
- Cost?
- Security?

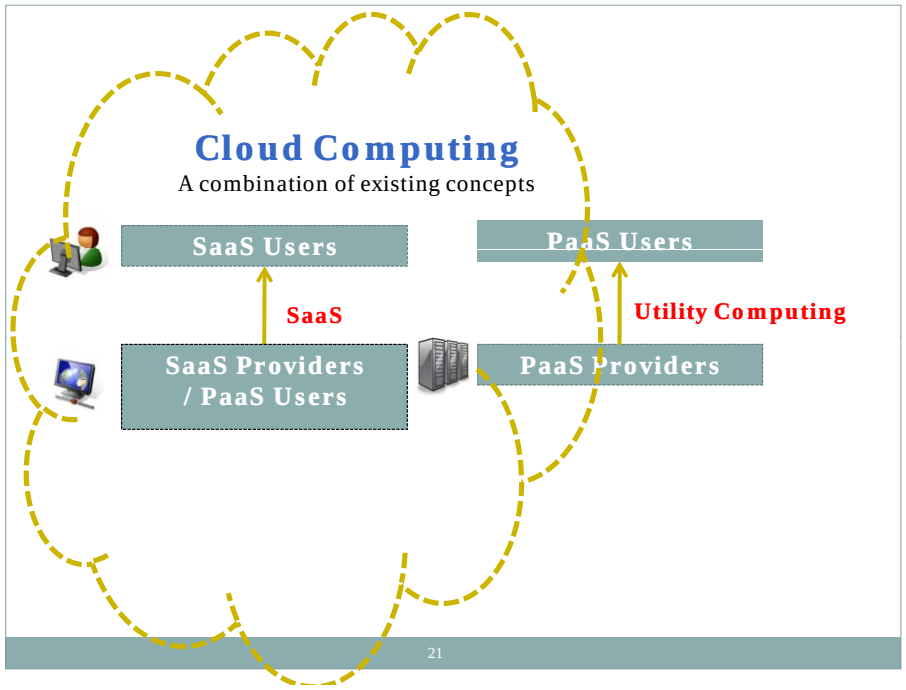
Key Parts of Cloud Definition

Natural Evolution of the Web

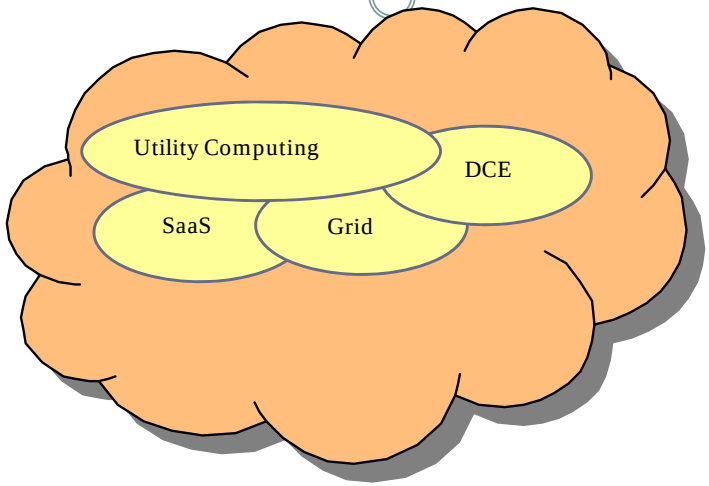


Types of Cloud Computing Services

- Software as a service (SaaS)
  - Software is provided to end users in an "On-demand" fashion.
  - Reduces upfront costs, i.e. buying multiple licenses
  - "Utility-based" computing
- Platform as a service (PaaS)
  - When the software needed to develop cloud applications are *themselves* provided in a "software as a service" fashion
- Infrastructure as a service (IaaS)
  - An "infrastructure" referring to much of the background *hardware* (contrast to software) needs of an organization



Clouds Can Have Any/All of These



And the descriptions often overlap !

Software as a Service (1/2)

- A common problem:
  - Business software installed on a work computer cannot be used from other computers (home, mobile computers) because the EULA doesn't allow it (licensing costs)
  - Remote access a possibility, but issues with uptime, security...
  - Instead the software is abstracted to a *cloud application* which can be accessed anywhere via *cloud client* (e.g. Web browser)
- Features of Software as a Service:
  - Commercial software that can be used and configured over a network, file servers and remote configuration storage, one application used by many clients, application updateability.

Software as a Service (2/2)

- Application is used as an on demand service. Often provided via the Internet
- Example: Google Apps (online office)
- Benefits to users
  - Reduce expenses: multiple computers, multiple users
  - Ease of usage: easy installation, access everywhere
- Benefits to providers
  - Easier to maintain
  - Control usage (no illegal copies)

Software as a Service: Types

- Multi-level architectures:
  - Level 1: Customized, ad-hoc basis
  - Level 2: Configurability
  - Level 3: Configurability + multi-tenancy (a single instance of software)
  - Level 4: Configurability + multi-tenancy + Scalability
- Virtualization in contrast with multi-tenancy

Infrastructure as a Service

- Components of Infrastructure as a service:
  - Computing hardware
    - E.g. Client Storage Space, Processing power needs
  - Virtualization
    - E.g. VM Ware, VirtualBox
  - Networking
    - Security, communication speeds, servers
  - Internet Access
    - Connecting clients applications to servers anywhere
  - Utility Computing
    - E.g. Charging by hour, gigabyte, process load
  - Contracts
    - Specific service levels which must be adhered too.

Platform as a Service

- Implements the "Software Lifecycle" on the cloud.
- Common software engineering practices: client elicitation, software design, component design, implementation, testing, maintenance, data handling and documentation
- How does it work?
  - Cloud servers allow not just hosting of completed application, but applications *in progress*
  - U.I. design tools right inside the web browser to use them
  - Provides testing in real situations: heavy client loads
  - Automatic integration with other data sources on the cloud
  - Tools for the developers to work efficiently with one another
  - Profit simulations

Cloud "Killer Apps"

- Mobile and Web applications
  - Mobile devices: low memory & computation power
- Extensions of desktop software
  - Matlab, Mathematica

Clouds: Heavier Than You Think

- A lot of constituent elements are necessary to make Cloud Computing work:
  - Cloud applications
  - Cloud clients
  - Cloud hardware
  - Cloud platform
  - Cloud services
  - Cloud storage

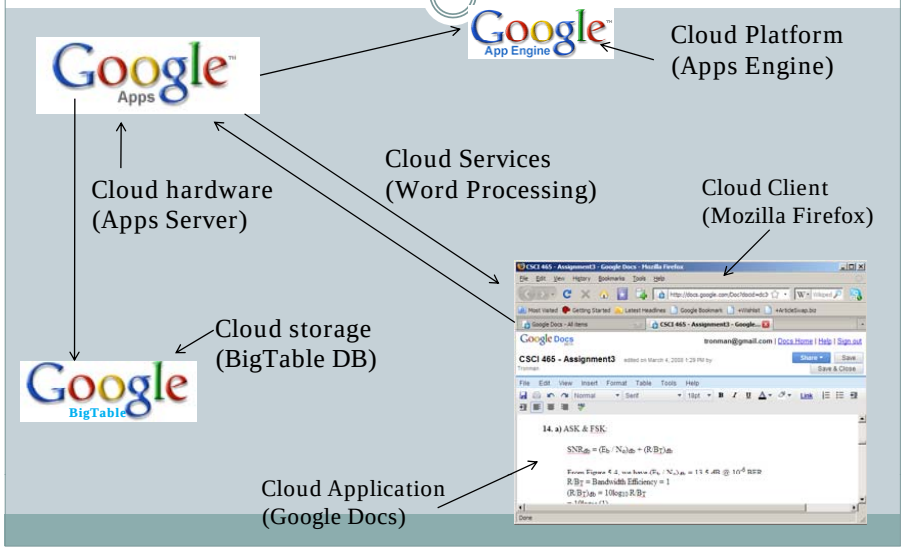
Cloud Architecture

- Heavily relies on communication and communication protocols
- Standards are implemented openly (publicly) to aid collaboration between services
- Software designers are “cloud architects”, software is integrated into the cloud system by a “cloud integrator”
- API’s extremely important
  - “Without APIs, there is no cloud computing.” – Dave Rosenberg

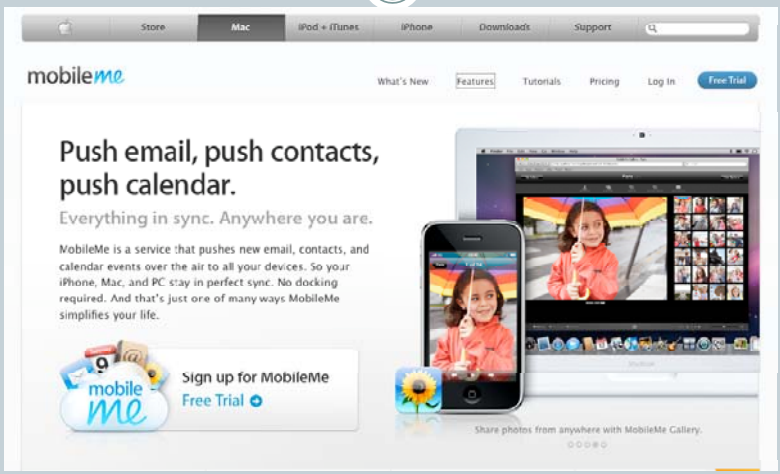
Role-based Cloud Computing

- Providers
  - Responsible for creation and maintenance of the hardware infrastructure which supports cloud computing applications.
- Developers
  - Responsible for the creation of software to be ran on a cloud computing system. Often works with provides for integration.
- Users
  - Responsible for the consumption of cloud-computing services.
- Vendors
  - Responsible for creation and distribution of lower-end platforms for cloud computing.

Example: Google Apps

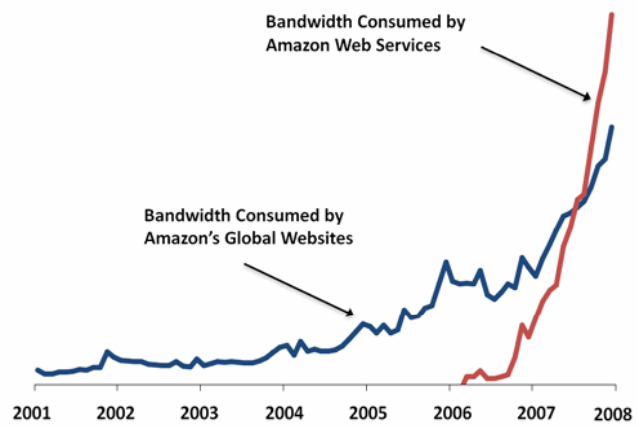


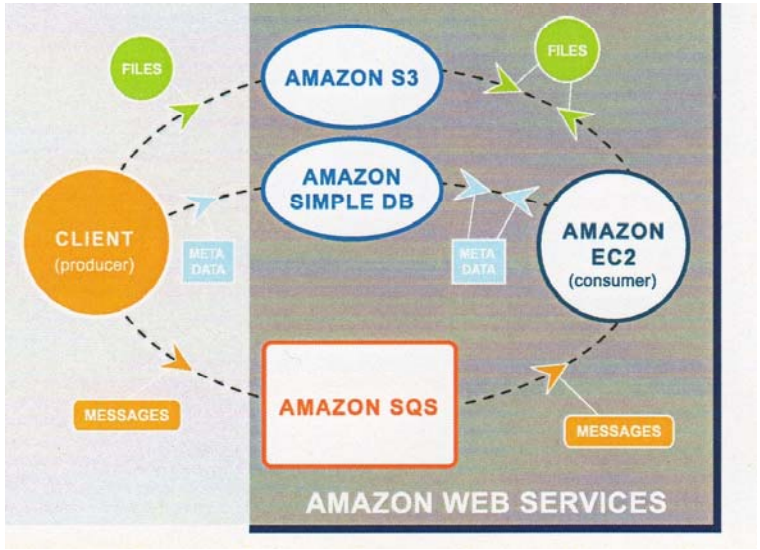
Example: MobileMe



Amazon EC2

- Characteristics:
  - Elastic: increase or decrease capacity ***within minutes***
    - Monitor and control via EC2 APIs
  - Completely controlled: root access to each instances
  - Flexible: choose your OS, software packages...
    - Redhat, Ubuntu, Windows Server 2003,...
    - Small, large, extra large instances
  - Reliable: Amazon datacenters, high availability and redundancies
  - Secure: Web interface to configure firewall settings
- Cost:
  - CPU: small instance, $0.10 per hour for Linux, $0.125 per hour for Windows (1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor)
  - Bandwidth: in $0.10, out $0.17 per GB
  - Storage: $0.10 per GB-month, $0.10 per 1 million I/O requests





Should You Move Into a Cloud?

- Does it really save money?

$$\text{UserHours}_{\text{cloud}} \times (\text{revenue} - \text{Cost}_{\text{cloud}}) \geq \text{UserHours}_{\text{datacenter}} \times \left(\text{revenue} - \frac{\text{Cost}_{\text{datacenter}}}{\text{Utilization}}\right)$$

  - $Cost_{cloud} > Cost_{datacenter}$, balance by *Utilization*
  - $UserHours_{cloud} > UserHours_{datacenter}$ (under-provisioning)
- Other factors
  - Re-implement programs
  - Move data into cloud
  - What else?
- Example:
  - Upload rate 20Mbits / s. 500GB takes 55 hours
  - If can process locally in less than 55 hours → moving into a cloud would not save time

EC2 Examples

- Batch processing / MapReduce
  - Peter Harkins at The Washington Post: 200 EC2 instances (1,407 server hours), convert 17,481 pages of Hillary Clinton's travel documents within 9 hours
  - The New York Times used 100 Amazon EC2 instances + Hadoop application to recognize 4TB of raw TIFF image into 1.1 million PDFs in 24 hours ($240)

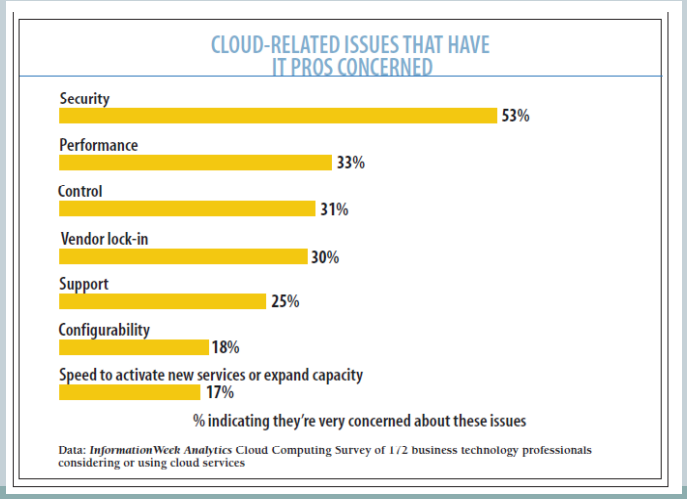
Ideal Cloud Computing

- Low cost for users and providers
- Independence of device and software
- Efficient utilization of all resources
- Constant, or near-constant uptime through the use of resource distribution
- Ability to continually add new users and applications by easily increasing resources when necessary
- Data is secure with policies to protect its security

How to keep the Clouds Floating

- Cloud Computing issues
  - Security of providers access to user data
  - Trustworthiness of providers
  - Security of data placement
  - Security of data representation
  - Recoverability of data
  - Tracking of illicit activities on the cloud
  - Long-term costs of cloud development
  - "Vendor" lock-in

Cloud Adoption Challenges



Cloud Control, InformationWeek Reports, 2009

Where is the Value?

- Much of the value is in the Virtual Machines
- What are VMs used for?
  - Server Consolidation (Fermilab)
  - Disaster recovery (commercial)
  - Component Isolation (sandboxing)
  - Hardware Independence (any OS on any Box)
  - Cluster Computing
    - E.g. Deploy a classroom environment
- Adding VM Management takes this to the "clouds"
  - Access resources on-demand
  - Isolate Users from each other
  - Schedule VM usage

Principles of an Open Cloud
The Open Cloud Manifesto (1/2)

- Cloud providers must work together to ensure that the challenges to cloud adoption (security, integration, portability, interoperability, governance/management, metering/monitoring) are addressed through open collaboration and the appropriate use of standards
- Cloud providers must not use their market position to lock customers into their particular platforms and limiting their choice of providers
- Cloud providers must use and adopt exiting standards wherever appropriate. The IT industry has invest heavily in existing standards and standards organizations; there is no need to duplicate or reinvent them

Principles of an Open Cloud
The Open Cloud Manifesto (2/2)

- When new standards (or adjustments to existing standards) are needed, we must be judicious and pragmatic to avoid creating too many standards. We must ensure that standards promote innovation and do not inhibit it
- Any community effort around the open cloud should be driven by customer needs, not merely the technical needs of cloud providers, and should be tested or verified against real customer requirements
- Cloud computing standards organizations, advocacy groups, and communities should work together and stay coordinated, making sure that efforts do not conflict or overlap

The Future?

- Application software:
  - Cloud and client parts, disconnection tolerance
- Infrastructure software:
  - Resource accounting, VM awareness
- Hardware systems:
  - Containers, energy proportionality

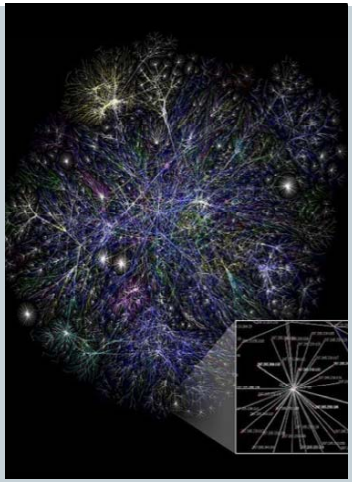
Project to Study Global Impact of Cloud Computing

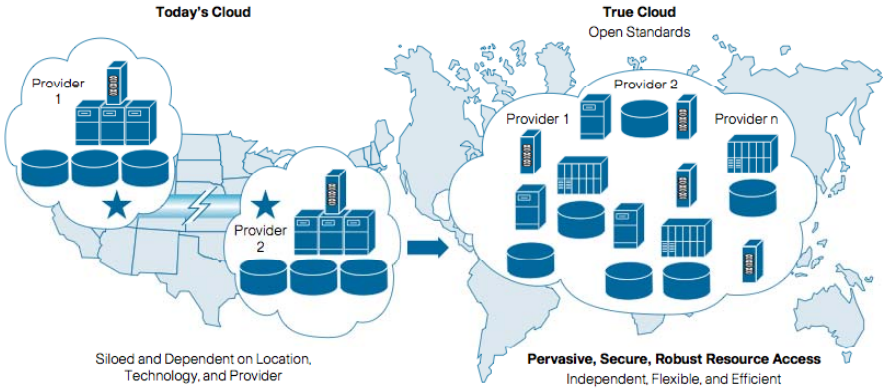
- For Individuals:
  - Greater access to information and collaboration
- For Society:
  - Developing nations may "leap frog" traditional datacenters
  - Revolutionize education
- For Business and Economic Growth:
  - Better decisions, better way to conduct business
  - Lower barriers of entry
  - Possibility of spurring job growth and new applications
- What will Cloud Computing look like in 2015?
- Who will it benefit?

WORLD ECONOMIC FORUM

Global Cloud of Clouds ("InterCloud")

- Interconnected network of servers, storage, and applications
- Segmented into public and private clouds
  - For security
  - For predictability
  - For regulatory compliance
- Unified and driven by a set of protocols, software API, and services
- Open to all





(Source: Cisco Systems)

Thank You!

BEBO@SLAC.STANFORD.EDU