

An Introduction and Overview of Cloud Computing

Presented by: Prof Mark Baker

ACET, University of Reading

Tel: +44 118 378 8615

E-mail: Mark.Baker@computer.org

Web: <http://acet.rdg.ac.uk/~mab>



What is Cloud Computing?

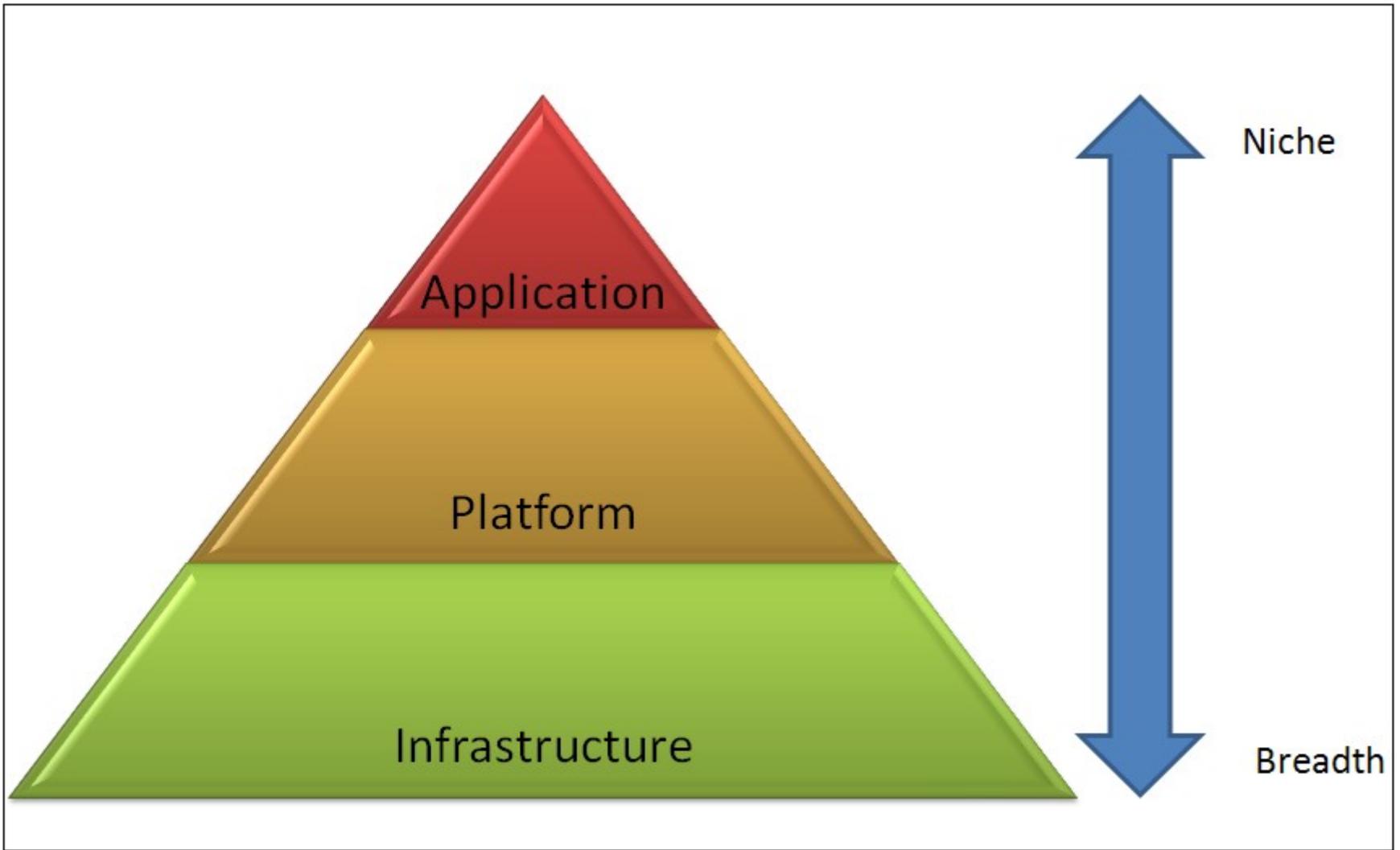
- **Cloud Computing** is a general term used to describe a new class of network based computing that takes place over the Internet, basically a step on from **Utility Computing**.
- In other words, this is a collection/group of integrated and networked hardware, software and Internet infrastructure (called a platform).
- Using the Internet for communication and transport provides hardware, software and networking services to clients.
- These platforms hide the complexity and details of the underlying infrastructure from users and applications by providing very simple graphical interface or API (Applications Programming Interface).

What is Cloud Computing?

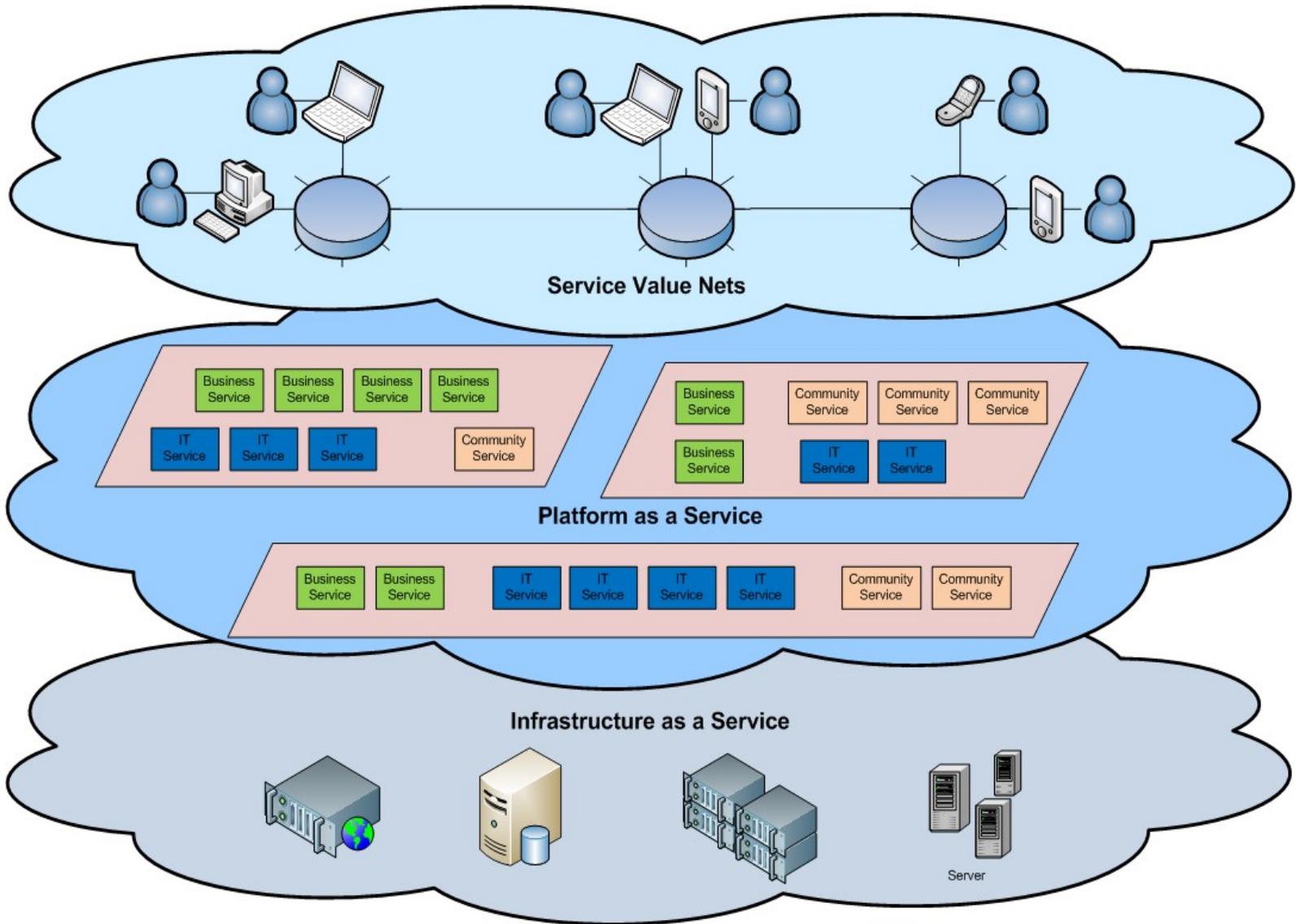
- In addition, the platform provides on demand services, that are always on, anywhere, anytime and any place.
- Pay for use and as needed, elastic (scale up and down in capacity and functionalities).
- The hardware and software services are available to the general public, enterprises, corporations and businesses markets.

Cloud Summary

- Cloud computing is an umbrella term used to refer to Internet based development and services.
- A number of characteristics define cloud data, applications services and infrastructure:
 - **Remotely hosted:** Services or data are hosted on remote infrastructure.
 - **Ubiquitous:** Services or data are available from anywhere.
 - **Commodified:** The result is a utility computing model similar to traditional that of traditional utilities, like gas and electricity - you pay for what you would want!



Cloud Architecture



Different Cloud Computing Layers

Application Service (SaaS)	MS Live/ExchangeLabs, IBM, Google Apps; Salesforce.com Quicken Online, Zoho, Cisco
Application Platform	Google App Engine, Mosso, Force.com, Engine Yard, Facebook, Heroku, AWS
Server Platform	3Tera, EC2, SliceHost, GoGrid, RightScale, Linode
Storage Platform	Amazon S3, Dell, Apple, ...

Cloud Computing Service Layers

	Services	Description
Application Focused	Services	Services - Complete business services such as PayPal, OpenID, OAuth, Google Maps, Alexa
	Application	Application - Cloud based software that eliminates the need for local installation such as Google Apps, Microsoft Online
	Development	Development - Software development platforms used to build custom cloud based applications (PAAS & SAAS) such as SalesForce
Infrastructure Focused	Platform	Platform - Cloud based platforms, typically provided using virtualization, such as Amazon ECC, Sun Grid
	Storage	Storage - Data storage or cloud based NAS such as CTERA, iDisk, CloudNAS
	Hosting	Hosting - Physical data centers such as those run by IBM, HP, NaviSite, etc.

Basic Cloud Characteristics

- The “no-need-to-know” in terms of the underlying details of infrastructure, applications interface with the infrastructure via the APIs.
- The “flexibility and elasticity” allows these systems to scale up and down at will - utilising the resources of all kinds (CPU, storage, server capacity, load balancing, and databases).
- The “pay as much as used and needed” type of utility computing and the “always on!, anywhere and any place” type of network-based computing.

Basic Cloud Characteristics

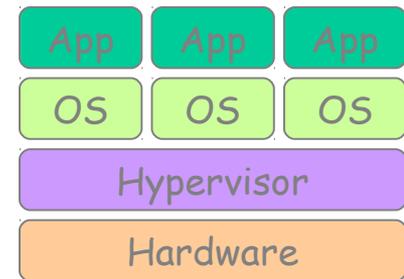
- Cloud are transparent to users and applications, they can be built in multiple ways - branded products, proprietary open source, hardware or software, or just off-the-shelf PCs.
- In general, they are built on clusters of PC servers and off-the-shelf components plus Open Source software combined with in-house applications and/or system software.

Software as a Service (SaaS)

- **SaaS** is a model of software deployment where an application is hosted as a service provided to customers across the Internet.
- **SaaS** is generally used to refer to business software rather than consumer software, falls under Web 2.0!
- By removing the need to install and run an application on a user's own computer it is seen as a way for businesses to get the same benefits as commercial software with smaller cost outlay.
- **SaaS** alleviates the burden of software maintenance/support, but users relinquish control over software versions and requirements.
- Terms that are used in this sphere include *Platform as a Service (PaaS)* and *Infrastructure as a Service (IaaS)*.

Virtualization

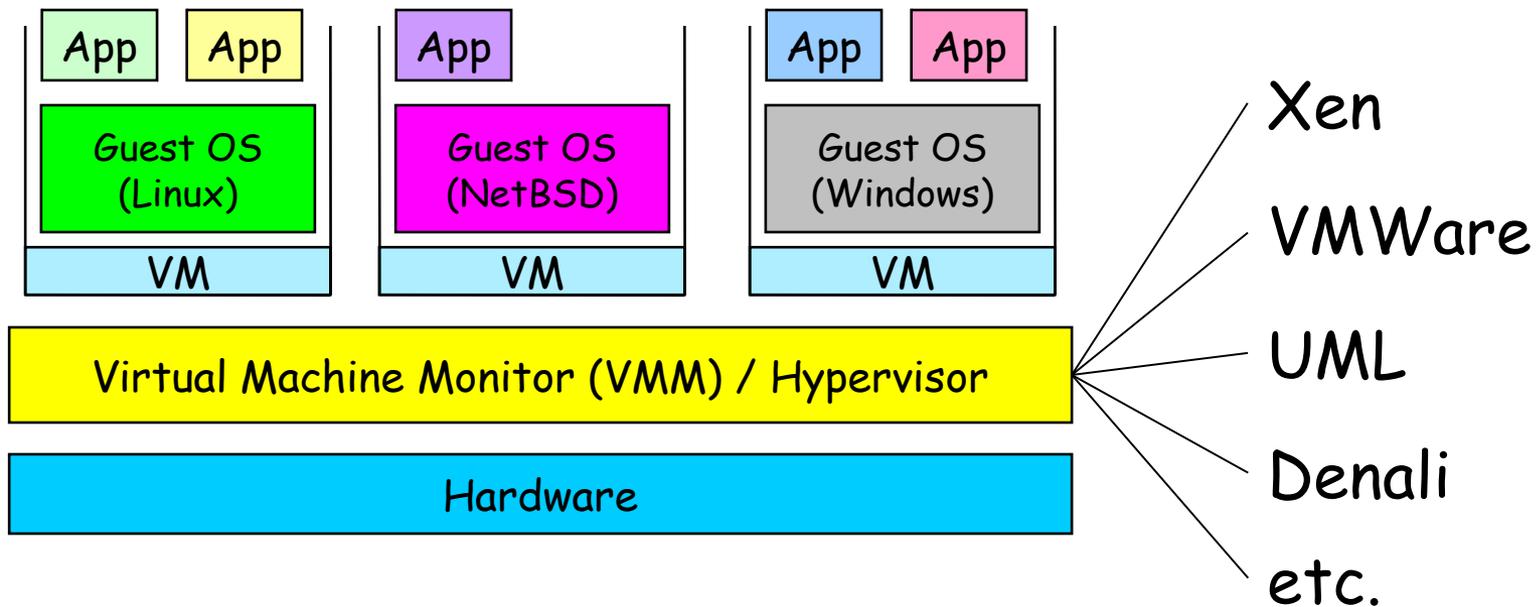
- Virtual workspaces:
 - An abstraction of an execution environment that can be made dynamically available to authorised clients by using well-defined protocols,
 - Resource quota (e.g. CPU, memory share),
 - Software configuration (e.g. O/S, provided services).
- Implement on Virtual Machines (VMs):
 - Abstraction of a physical host machine,
 - Hypervisor intercepts and emulates instructions from VMs, and allows management of VMs,
 - VMWare, Xen, etc.
- Provide infrastructure API:
 - Plug-ins to hardware/support structures



Virtualized Stack

Virtual Machines

- VM technology allows multiple virtual machines to run on a single physical machine.



Performance: Para-virtualization (e.g. Xen) is very close to raw physical performance!

What is the purpose and benefits?

- **Cloud computing** enables companies and applications, which are system infrastructure dependent, to be infrastructure-less.
- By using the Cloud infrastructure on “**pay as used and on demand**”, all of us can save in capital and operational investment!
- Clients can:
 - Put their data on the platform instead of on their own desktop PCs and/or on their own servers.
 - They can put their applications on the cloud and use the servers within the cloud to do processing and data manipulations etc.

What is the purpose and benefits?

- The term **cloud** is used to describe and to reflect this class of Internet centric computing infrastructure being transparent (users do not need to know what is behind the scenes!), highly scalable (scale up and down as needed), on-demand, pay as needed and as used.
- After so many years, Cloud Computing is today the a **network based computing** over Internet.

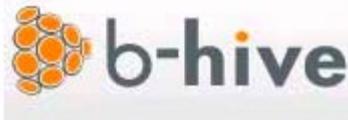
Cloud-Sourcing

- Why is it becoming a Big Deal:
 - Using high-scale/low-cost providers,
 - Any time/place access via web browser,
 - Rapid scalability; incremental cost and load sharing,
 - Can forget need to focus on local IT.
- Concerns:
 - Performance, reliability, and SLAs,
 - Control of data, and service parameters,
 - Application features and choices,
 - Interaction between Cloud providers,
 - No standard API - mix of SOAP and REST!
 - Privacy, security, compliance, trust...

Some Commercial Cloud Offerings



Amazon Elastic Compute Cloud (Amazon EC2) - Beta

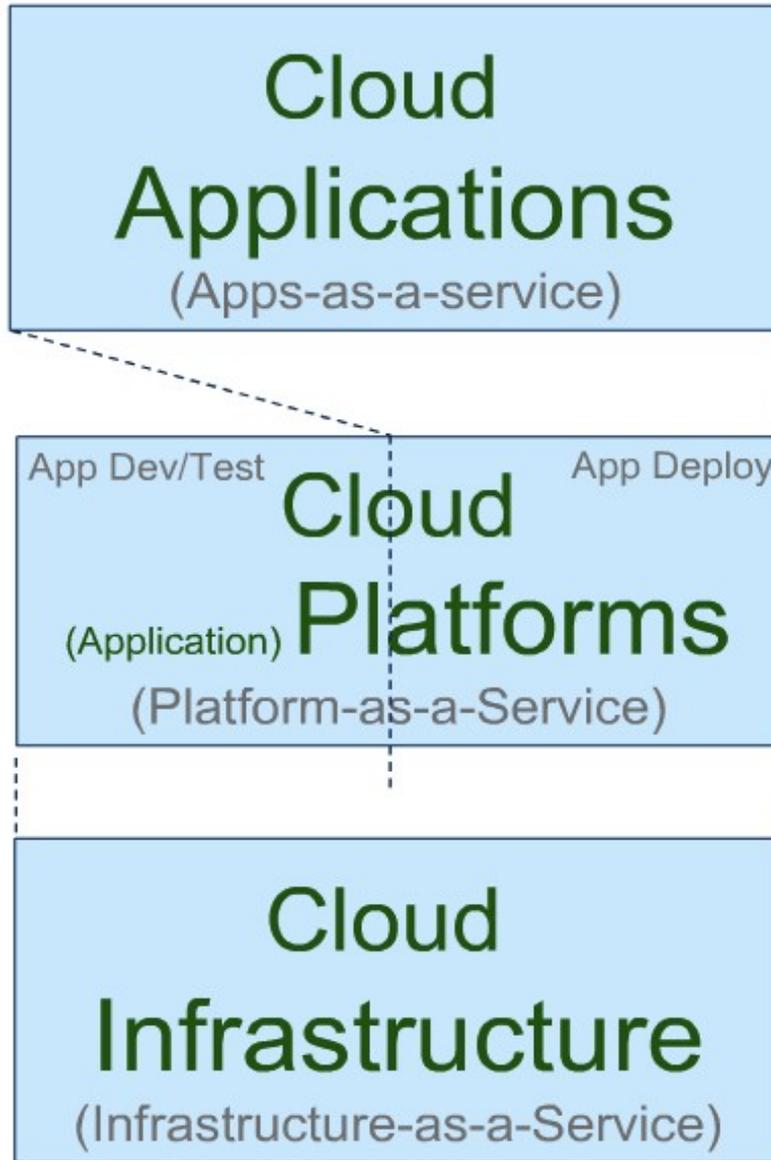


Google App Engine



- Problem: Commercial offerings are proprietary and usually not open for cloud systems research and development

IT Cloud Services



Infrastructure Services

Storage

- Amazon S3 & EBS
- Rackspace Cloud Files
- Nirvanix
- AT&T Synaptic
- Zetta

Cloud Broker

- RightScale
- enStratus
- Kaavo
- Elastra
- CloudKick
- CloudSwitch

Compute

- Amazon EC2
- Serve Path GoGrid
- Rackspace Cloud Servers
- Joyent Cloud
- Flexiant Flexiscale
- ElasticHosts
- Terremark
- ITRiCITY
- LayeredTech
- Savvis Cloud Compute
- Verizon CaaS
- AT&T Synaptic
- Sungard Enterprise Cloud
- Navisite

Services Management

- Scalr
- CohesiveFT
- Ylastic
- CloudFoundry
- NewRelic
- Cloud42
- Amazon CloudWatch
- Amazon VPC

Cloud Software

SaaS Data Security

- Navajo
- PerspecSys

Data

- 10Gen MongoDB
- Apache CouchDb
- Apache HBase
- Hypertable
- Tokyo Cabinet
- Cassandra
- memcached
- Clustrix
- FlockDB
- Gizzard
- Redis
- BerkeleyDB
- Voldemort
- Terrastore

Compute

- Globus Toolkit
- Xeround
- Sun Grid Engine
- Hadoop
- OpenCloud
- Gigaspace
- DataSynapse

File Storage

- EMC Atmos
- ParaScale
- Zmamda
- CTERA
- Appistry

Cloud Management

- CA Turn-key Cloud
- OpenNebula
- Open.ControlTier
- Enomaly Enomalism
- VMware vCloud
- CohesiveFT VPN Cubed
- Hyperic
- Eucalyptus
- Puppet Labs
- Appistry
- IBM CloudBurst
- Cisco UCS
- Zenoss
- Surgent

CLOUD TAXONOMY

Platform Services

General Purpose

- Force.com
- Etelos
- LongJump
- Rollbase
- Bungee Connect
- Google App Engine
- Engine Yard
- Caspio
- Qrmp
- MS Azure
- Mosso Cloud Sites
- VMforce
- Intuit Partner Platform
- Joyent Smart Platform

Business Intelligence

- Aster DB
- Quantivo
- Cloud9 Analytics
- K2 Analytics
- LogiXML
- Oco
- PivotLink
- Clario Analytics
- ColdLight Neuron
- Vertica

Integration

- Amazon SQS
- Amazon SNS
- Boomi
- SnapLogic
- IBM Cast Iron
- gnip
- Appian Anywhere
- HubSpan
- Informatica On-Demand

Development & Testing

- Keynote Systems
- SOASTA
- SkyTap
- Aptana
- LoadStorm
- Collabnet
- Rational Software Delivery Services

Database

- Amazon SimpleDB
- Mosso Drizzle
- Amazon RDS

Software Services

Financials

- Concur
- Xero
- Workday
- Expensify
- Intuit Quickbooks Online

Content Management

- Clickability
- SpringCM
- CrownPoint

Collaboration

- Box.net
- CubeTree
- SocialText
- Basecamp
- Assembla
- DropBox

Sales

- Xactly
- StreetSmarts
- Success Metrics

Desktop Productivity

- Zoho
- Google Apps
- HyperOffice
- MS Office Web Apps

Billing

- Aria Systems
- eVapt
- Redi2
- Zuora

Social Networks

- Ning
- Zemby
- Amitive
- Jive SBS

CRM

- NetSuite
- Parature
- Responsys
- Rightnow
- LiveOps
- MSDynamics
- Salesforce.com
- Oracle On Demand

Document Management

- NetDocuments
- DocLanding
- Knowledge TreeLive
- SpringCM

Cloud Storage

- Several large Web companies (such as Amazon and Google) are now exploiting the fact that they have data storage capacity that can be hired out to others.
- This approach, known as **cloud storage** allows data stored remotely to be temporarily cached on desktop computers, mobile phones or other Internet-linked devices.
- Amazon's Elastic Compute Cloud (EC²) and Simple Storage Solution (S3) are well known examples.

Amazon Web Services

Products ▾

Solutions ▾

Resources ▾

Infrastructure Services

- » Amazon Elastic Compute Cloud (Amazon EC2)
- » Amazon SimpleDB
- » Amazon Simple Storage Service (Amazon S3)
- » Amazon CloudFront
- » Amazon Simple Queue Service (Amazon SQS)
- » AWS Premium Support

Payments & Billing Services

- » Amazon Flexible Payments Service (Amazon FPS)
- » Amazon DevPay

On-Demand Workforce

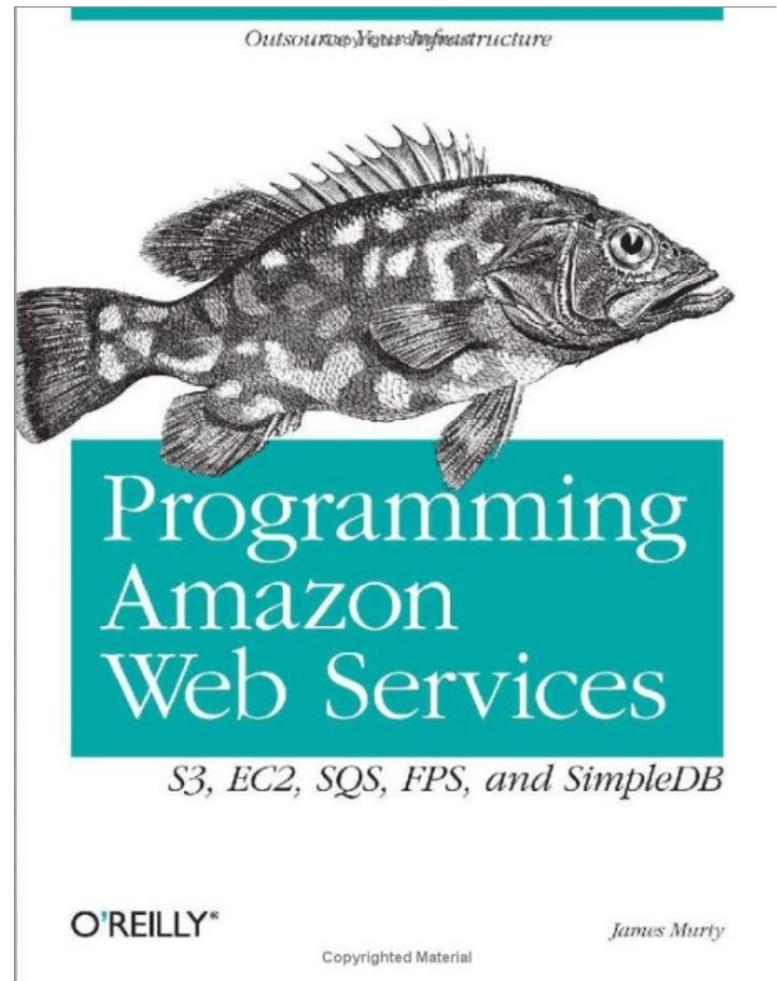
- » Amazon Mechanical Turk

Alexa Web Services

- » Alexa Web Information Service
- » Alexa Top Sites
- » Alexa Site Thumbnail

Amazon Fulfillment & Associates

- » Amazon Fulfillment Web Service (Amazon FWS)
- » Amazon Associates Web Service



19th May, 09

mark.baker@computer.org

Amazon Web Services



S3 - Simple Storage Service
CloudFront
EC2 - Elastic Compute Cloud
SQS - Simple Queue Service

Premium Support:

- SimpleDB,
- Elastic MapReduce,
- Mechanical Turk,
- Workflow System.

Amazon Simple Storage Service (S3)

- Unlimited Storage.
- Using the same scalable, reliable, fast infrastructure that Amazon uses.
- Pay for what you use:
 - \$0.20 per GByte of data transferred,
 - \$0.15 per GByte-Month for storage used,
 - Second Life Update:
 - 1TBytes, 40,000 downloads in 24 hours - \$200,



19th May, 09

Amazon Simple Storage Service (S3)

- REST/SOAP based!
- Flexible Download Protocol:
 - HTTP,
 - BitTorrent.
- Implementations of the WS already made:
 - NS3 Manager:
 - .NET Based, Core Lib available for MS Build Tasks.
 - S3sync
 - Ruby Based,
 - Works similar to rsync (for those familiar).
 - S3Fox
 - Firefox extension (if you use Firefox!).
 - Jungle Disk (and other 3rd party ones):
 - Work as WEBDAV, but interoperability issues!!

Utility Computing - EC2

- Started in 2006:
 - Amazon Elastic Compute Cloud (EC2):
 - Elastic, marshal 1 to 100+ PCs via WS,
 - Machine Specs...,
 - Fairly cheap!
- Powered by Xen - a Virtual Machine:
 - Different from Vmware and VPC as uses "para-virtualization" where the guest OS is modified to use special hyper-calls:
 - High performance!
 - Hardware contributions by Intel (VT-x/Vanderpool) and AMD (AMD-V).
 - Supports "Live Migration" of a virtual machine between hosts.
- Pay by the hour from S(small), 0.10 \$/h, to XL 0.80 \$/h.
- Linux, Windows, OpenSolaris.
- Management Console/AP.

EC2 - The Basics

- Load your image onto S3 and register it.
- Boot your image from the Web Service.
- Open up required ports for your image.
- Connect to your image through SSH.
- Execute you application...

Amazon Elastic MapReduce

Hosted Hadoop Framework

- Hadoop is a software framework that enables distributed manipulation of large amounts of data.
- But Hadoop does this in a way that makes it reliable, efficient, and scalable.
- Hadoop is efficient because it works on the principle of parallelisation, allowing data to process in parallel to increase the processing speed.
- Hadoop is also scalable, permitting operations on petabytes of data.

- Hadoop: Java Framework.
- Data Intensive distributed applications.
 - Amazon EC2 + Amazon S3.
 - Use cases:
 - Web Indexing.
 - Data Mining.
 - Machine Learning.
 - Financial Analysis.
 - Scientific Simulation.



19th May, 09

mark.baker@computer.org

Opportunities and Challenges

- The use of the cloud provides a number of opportunities:
 - It enables services to be used without any understanding of their infrastructure.
 - Cloud computing works using economies of scale:
 - It potentially lowers the outlay expense for start up companies, as they would no longer need to buy their own software or servers.
 - Cost would be by on-demand pricing.
 - Vendors and Service providers claim costs by establishing an ongoing revenue stream.
 - Data and services are stored remotely but accessible from "anywhere".

Opportunities and Challenges

- In parallel there has been backlash against cloud computing:
 - Use of cloud computing means dependence on others and that could possibly limit flexibility and innovation:
 - The **others** are likely become the bigger Internet companies like Google and IBM, who may monopolise the market.
 - Some argue that this use of supercomputers is a return to the time of mainframe computing that the PC was a reaction against.
 - Security could prove to be a big issue:
 - It is still unclear how safe out-sourced data is and when using these services ownership of data is not always clear.
 - There are also issues relating to policy and access:
 - If your data is stored abroad whose Freedom Of Information (FOI) policy do you adhere to?
 - What happens if the remote server goes down?
 - How will you then access files?
 - There have been cases of users being locked out of accounts and losing access to data.

Advantages of Cloud Computing

- Lower computer costs:
 - You do not need a high-powered and high-priced computer to run cloud computing's web-based applications.
 - Since applications run in the cloud, not on the desktop PC, your desktop PC does not need the processing power or hard disk space demanded by traditional desktop software.
 - When you are using web-based applications, your PC can be less expensive, with a smaller hard disk, less memory, more efficient processor...
 - In fact, your PC in this scenario does not even need a CD or DVD drive, as no software programs have to be loaded and no document files need to be saved.

Advantages of Cloud Computing

- Improved performance:
 - With few large programs hogging your computer's memory, you will see better performance from your PC.
 - Computers in a cloud computing system boot and run faster because they have fewer programs and processes loaded into memory...
- Reduced software costs:
 - Instead of purchasing expensive software applications, you can get most of what you need for free-ish!
 - That is right - most cloud computing applications today, such as the Google Docs suite, are totally free.
 - That is a lot better than paying \$200+ for similar Microsoft Office software - which alone may be justification for switching to cloud applications.

Advantages of Cloud Computing

- Instant software updates:
 - Another advantage to cloud computing is that you are no longer faced with choosing between obsolete software and high upgrade costs.
 - When the application is web-based, updates happen automatically - available the next time you log into the cloud.
 - When you access a web-based application, you get the latest version - without needing to pay for or download an upgrade.
- Improved document format compatibility.
 - You do not have to worry about the documents you create on your machine being compatible with other users' applications or operating systems.
 - Where Word 2007 documents cannot be opened on a computer running Word 2003, all documents can be read!
 - There are potentially no format incompatibilities when everyone is sharing documents and applications in the cloud.

Advantages of Cloud Computing

- Unlimited storage capacity:
 - Cloud computing offers virtually limitless storage.
 - Your computer's current 200 Gbyte hard drive is small compared to the hundreds of Pbytes available in the cloud.
 - Whatever you need to store, you can.
- Increased data reliability:
 - Unlike desktop computing, in which if a hard disk crashes and destroy all your valuable data, a computer crashing in the cloud should not affect the storage of your data.
 - That also means that if your personal computer crashes, all your data is still out there in the cloud, still accessible.
 - In a world where few individual desktop PC users back up their data on a regular basis, cloud computing is a data-safe computing platform!

Advantages of Cloud Computing

- Universal document access:
 - That is not a problem with cloud computing, because you do not take your documents with you.
 - Instead, they stay in the cloud, and you can access them whenever you have a computer and an Internet connection.
 - All your documents are instantly available from wherever you are.
- Latest version availability:
 - Another document-related advantage of cloud computing is that when you edit a document at home, that edited version is what you see when you access the document at work.
 - The cloud always hosts the latest version of your documents; as long as you are connected, you are not in danger of having an outdated version.

Advantages of Cloud Computing

- Easier group collaboration:
 - Sharing documents leads directly to better collaboration.
 - Many users do this as it is an important advantages of cloud computing - multiple users can collaborate easily on documents and projects.
 - Because the documents are hosted in the cloud, not on individual computers, all you need is an Internet connection, and you are collaborating.
- Device independence.
 - You are no longer tethered to a single computer or network.
 - Changes to computers, applications and documents follow you through the cloud.
 - Move to a portable device, and your applications and documents are still available.

Disadvantages of Cloud Computing

- Requires a constant Internet connection:
 - Cloud computing is impossible if you cannot connect to the Internet.
 - Since you use the Internet to connect to both your applications and documents, if you do not have an Internet connection you cannot access anything, even your own documents.
 - A dead Internet connection means no work and in areas where Internet connections are few or inherently unreliable, this could be a deal-breaker.
 - When you are offline, cloud computing simply does not work.

Disadvantages of Cloud Computing

- Does not work well with low-speed connections:
 - Similarly, a low-speed Internet connection, such as that found with dial-up services, makes cloud computing painful at best and often impossible.
 - Web-based applications require a lot of bandwidth to download, as do large documents.
 - If you are labouring with a low-speed dial-up connection, it might take seemingly forever just to change from page to page in a document, let alone to launch a feature-rich cloud service.
 - In other words, cloud computing is not for the broadband-impaired!

Disadvantages of Cloud Computing

- Can be slow:
 - Even with a fast connection, web-based applications can sometimes be slower than accessing a similar software program on your desktop PC.
 - Everything about the program, from the interface to the current document, has to be sent back and forth from your computer to the computers in the cloud.
 - If the cloud servers happen to be backed up at that moment, or if the Internet is having a slow day, you would not get the instantaneous access you might expect from desktop applications.

Disadvantages of Cloud Computing

- Features might be limited:
 - This situation is bound to change, but today many web-based applications simply are not as full-featured as their desktop-based applications.
 - For example, you can do a lot more with Microsoft PowerPoint than with Google Presentation's web-based offering.
 - The basics are similar, but the cloud application lacks many of PowerPoint's advanced features.
 - If you are a power user, you might not want to leap into cloud computing just yet.

Disadvantages of Cloud Computing

- Stored data might not be secure:
 - With cloud computing, all your data is stored on the cloud.
 - The questions is [How secure is the cloud?](#)
 - Can unauthorised users gain access to your confidential data?
 - Cloud computing companies say that data is secure, but it is too early to be completely sure of that.
 - Only time will tell if your data is secure in the cloud.
- Stored data can be lost:
 - Theoretically, data stored in the cloud is safe, replicated across multiple machines.
 - But on the off chance that your data goes missing, you have no physical or local backup.
 - Put simply, relying on the cloud puts you at risk if the cloud lets you down.

Disadvantages of Cloud Computing

- **HPC Systems:**
 - Not clear that you can run compute-intensive HPC applications that use MPI/OpenMP!
 - Scheduling is important with this type of application - as you want all the VM to be co-located to minimise communication latency!
- **General Concerns:**
 - Each cloud systems uses different protocols and different APIs... so it may not be possible to run applications between cloud based systems.
 - Amazon has created its own DB system (not SQL 92), and workflow system (many popular workflow systems out there) - so your normal applications will have to be adapted to execute on these platforms.

The Future

- Many of the activities loosely grouped together under cloud computing have already been happening and centralised computing activity is not a new phenomena:
- Grid Computing was the last research-led centralised approach.
- However there are concerns that the mainstream adoption of cloud computing could cause many problems for users.
- Whether these worries are grounded or not has yet to be seen.
- Many new open source systems appearing that you can install and run on your local cluster - should be able to run a variety of applications on these systems.