

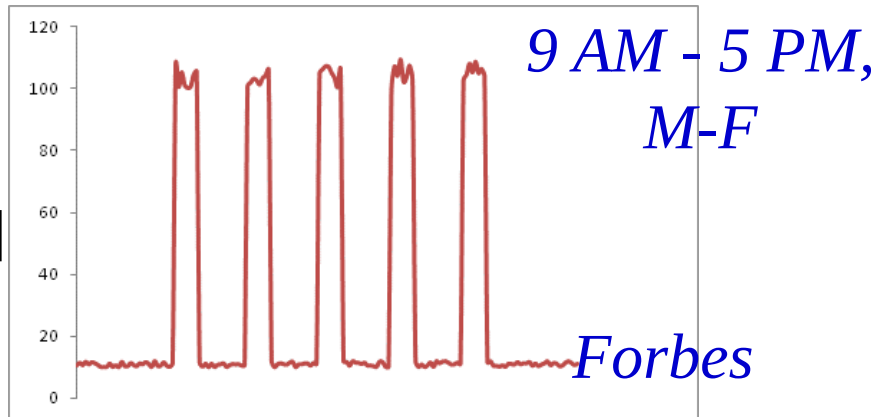
Cloud systems

dacosta@irit.fr



Why a new technology

- Large amount of data
 - Impossible to manage on a single computer
- Large amount of users
 - Workload varies a lot
- Grids are not always ad
 - Must know final size
 - Start big
 - Be well-known
 - Expertise in software, hardware, administration
- How to start the next best service / web-site ?



Cloud vision of the solution

- Host the web site in a *Cloud*
- Provision new servers every day
 - unprovision them every night
- Pay just \$0.10 per server per hour
- Let the cloud provider manage everything
 - Hardware
 - Administration!
- Just take care of the code and storage

Definition

- Cloud Computing is a general term used to describe a new class of network based computing that takes place over the Internet
- It is a collection/group of integrated and networked hardware, software and Internet infrastructure (called a platform)
- It uses the Internet for communication and 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).

Same as Grid initial Goal!

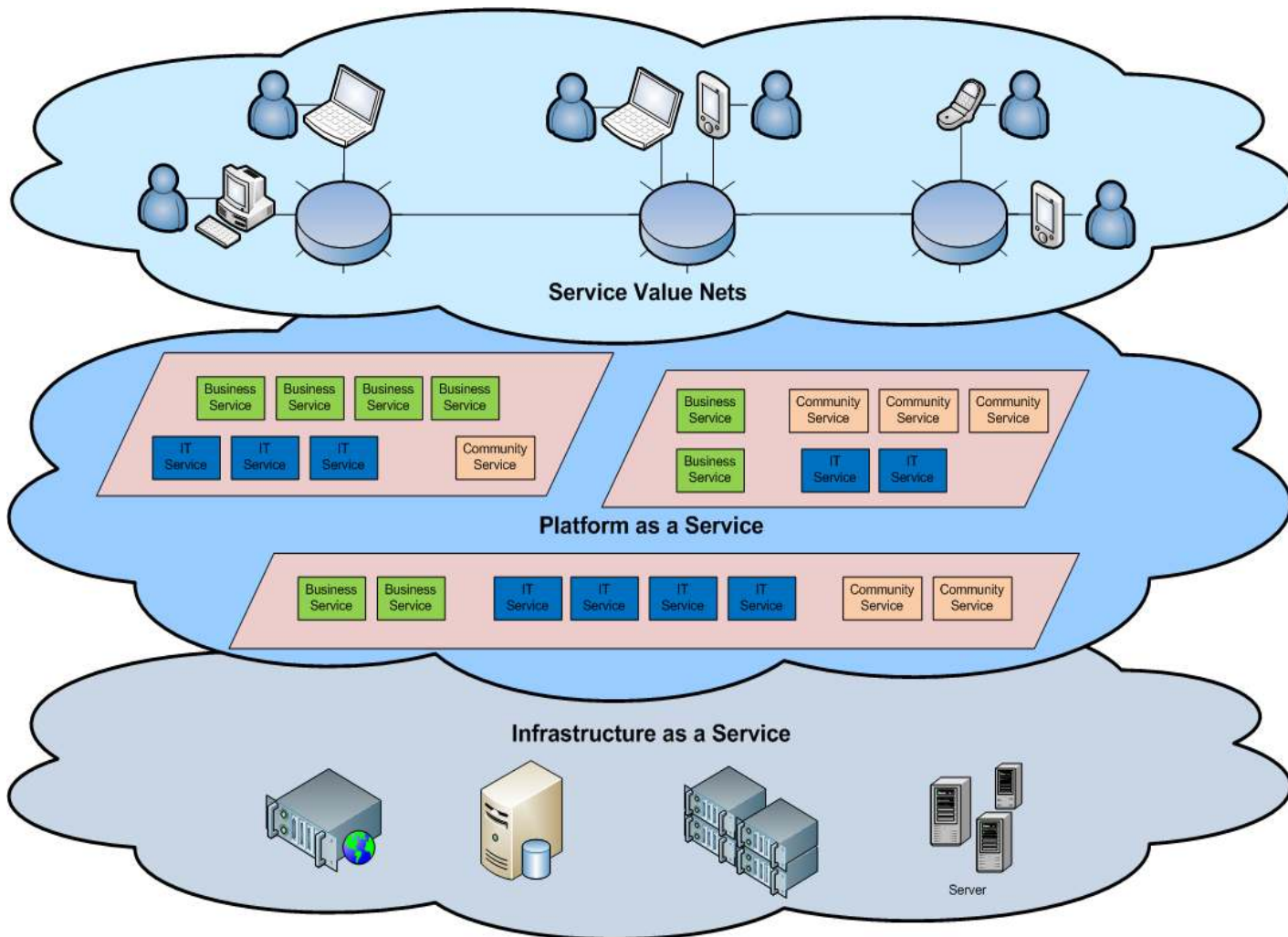
Some characteristics

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

Some characteristics

- **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 gas and electricity

Architecture



Cloud layers

- **Application as a Service**
 - 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, ...

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.
- They are “**transparent**” as they can run whatever the underlying layers
- They are composed of classical servers and open-source software

Exemple : 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).

Infrastructure

Virtualization

- Virtual workspaces:

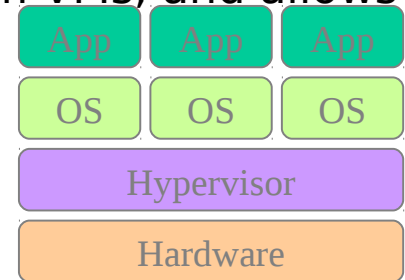
- An abstraction of an execution environment that can be made dynamically available to authorised clients by using well-defined protocols,
- Resource quota (e.g. CPU, memory share),
- Software configuration (e.g. O/S, provided services).

- 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:

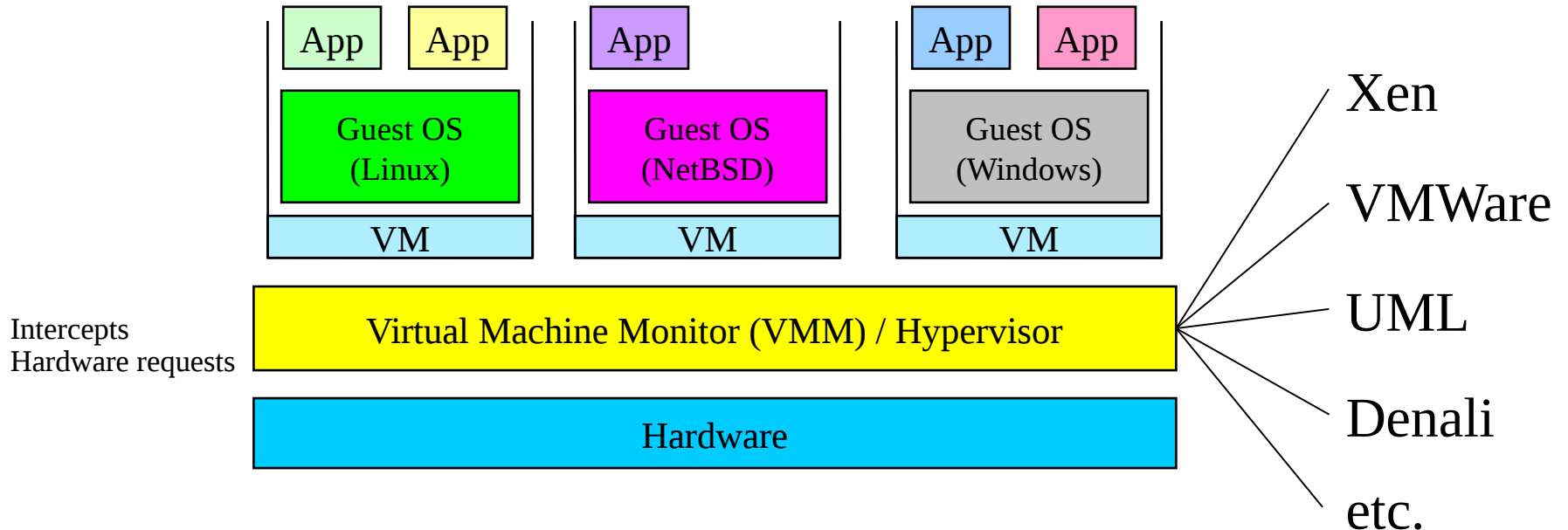
- Plug-ins to hardware/support structures



Virtualized Stack

Virtual Machines

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



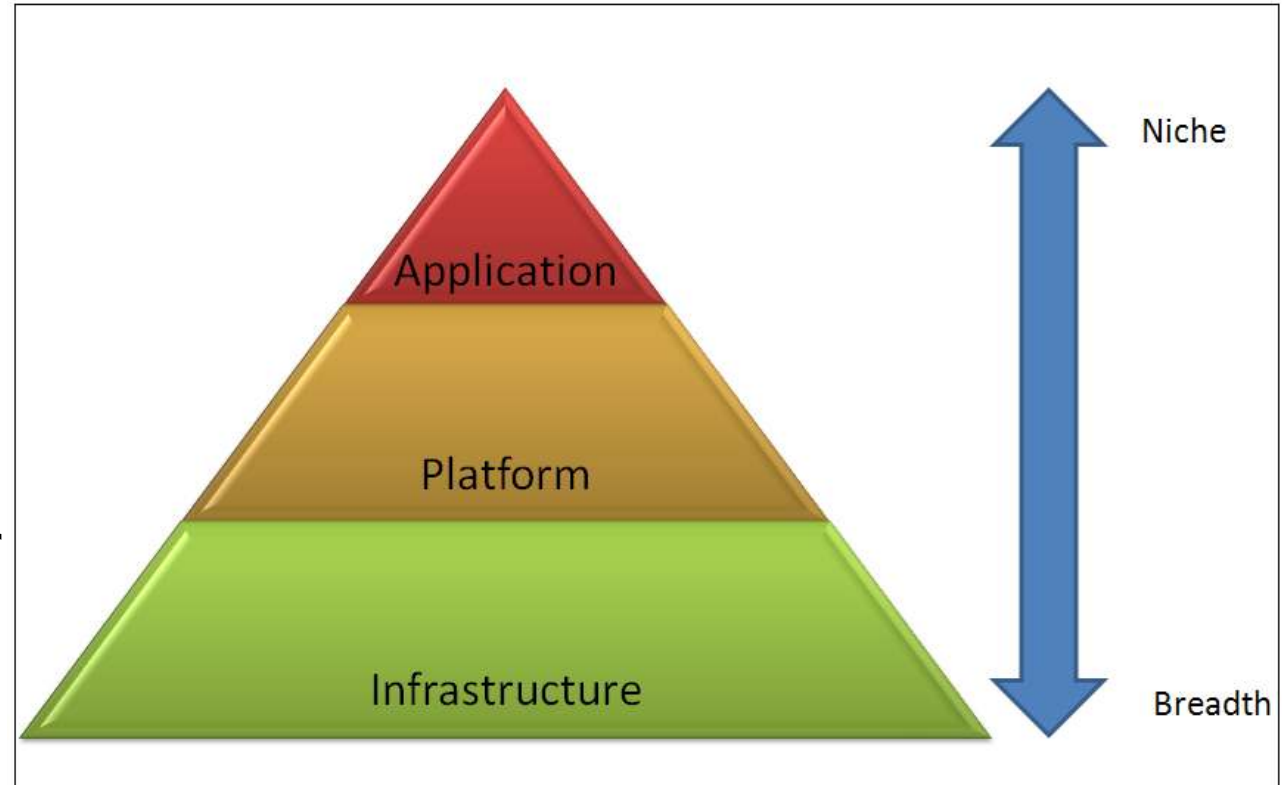
Performance: Para-virtualization (e.g. Xen, kvm) is very close to raw physical performance! Otherwise simulation-like which works but is slow

Virtualization in General

- Advantages of virtual machines:
 - Run operating systems where the physical hardware is unavailable,
 - Easier to create new machines, backup machines, etc.,
 - Software testing using “clean” installs of operating systems and software,
 - Emulate more machines than are physically available,
 - Timeshare lightly loaded systems on one host,
 - Debug problems (suspend and resume the problem machine),
 - Easy migration of virtual machines (shutdown needed or not).
 - Run legacy systems!
 - Not necessary to have one server for each service
 - Easier to manage increase of a service workload

Key concept in Cloud : Multi-layers

- Application
 - build over
- Platform
 - build over
- Infrastructure



- Each level managed by different people

Three levels / Three point of view

SaaS
Software as a Service

PaaS
Platform as a Service

IaaS
Infrastructure as a Service

Three levels / Three point of view

SaaS Software as a Service

- Software delivery model
 - No hardware or software to manage
 - Service delivered through a browser
- Advantage for users
 - Pay per use, updates free, trouble free, homogeneity
- Advantage for service provider
 - User lock-in, less fragmentation, recurrent fees

Three levels / Three point of view

- Consumer creates the software using tools and/or libraries from the provider
- The provider provides the networks, servers, storage and other services

PaaS Platform as a Service

- Same benefits as SaaS
- Some vendors
 - Google App Engine, Mosso, AWS: S3, Windows Azure
- Some classical services : **Hadoop**, Data Base, Web Server,...

Three levels / Three point of view

- Offers resources : Virtual machines, routers, servers, disks,...
- Same benefits as SaaS and PaaS
- Amazon EC2, Azure Services Platform, DynDNS, Google Compute Engine, HP Cloud, Rackspace Cloud, ...
- Main open-source middlewares
 - OpenNebula, OpenStack

IaaS
Infrastructure as a Service

PaaS & IaaS as tools

Cloud computing takes virtualization to the next step

- You don't have to own the hardware
- You “rent” it as needed from a cloud
- There are public clouds
 - e.g. Amazon EC2, and now many others (Microsoft, IBM, Sun, and others ...)
- A company can create a private one
 - With more control over security, etc.
- Classical method : Hybrid cloud
 - Internal cloud for core functions
 - External cloud for absorbing peak demand

Goal 1 - Cost Control

- **Cost**
 - Many systems have variable demands
 - Batch processing (e.g. New York Times)
 - Web sites with peaks (e.g. Forbes)
 - Startups with unknown demand (e.g. Instagram)
 - Reduce risk
 - Don't need to buy hardware until you need it

Goal 2 - Business Agility

- More than scalability - elasticity!
 - In big health care business
 - Used to take 3 - 4 months to give a department a server cluster, then they would keep it for themselves!
 - Using EC2, about 5 minutes!
 - And they give it back when they are done!
- Scaling back is as important as scaling up

Goal 3 - Stick to Our Business

- Most companies don't WANT to do system administration
 - Forbes says:
 - We are is a publishing company, not a software company
- But beware:
 - Do you really save much on sys admin?
 - You don't have the hardware, but you still need to manage the OS!

How Cloud Computing Works

- Various providers let you create virtual servers
 - Set up an account, perhaps just with a credit card
- You create virtual servers ("virtualization")
 - Choose the OS and software each "instance" will have
 - It will run on a large server farm located somewhere
 - You can instantiate more on a few minutes' notice
 - You can shut down instances in a minute or so
- They send you a bill for what you use

Any Nasty Details? Yes !

- How do I pick a provider?
- Am I locked in to a provider?
- Where do I put my data?
- What happens to my data when I shut down?
- How do I log in to my server?
- How do I keep others from logging in (security)?
- How do I get an IP address?
- Etc.

What is the purpose and benefits?

- Classical companies and applications, are infrastructure dependent
 - Cloud computing is infrastructure-less.
- Cloud infrastructure is “pay as used and on demand”
 - Savings 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.
- It is :
 - transparent (users do not know what is behind the scenes!)
 - highly scalable (scale up and down as needed)
 - on-demand, pay as needed and as used.

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



- Problem of lock-in

Infrastructure Services

Storage

- Amazon S3
- Amazon EBS
- CTERA Portal
- Mosso Cloud Files
- Nirvanix

Compute

- Amazon EC2
- Serve Path GoGrid
- Elastra
- Mosso Cloud Servers
- Joyent Accelerators
- AppNexus
- Flexiscale
- ElasticHosts
- Hosting.com CloudNine
- Terramark
- GridLayer
- ITRICITY
- LayeredTech

Services Management

- RightScale
- enStratus
- Scalr
- CohesiveFT
- Kaavo
- CloudStatus
- Ylastic
- Dynect
- CloudFoundry
- NewRelic
- Cloud42

Data

- 10Gen MongoDB
- Oracle Coherence
- Gemstone Gemfire
- Apache CouchDb
- Apache HBase
- Hypertable
- TerraCotta
- Tokyo Cabinet
- Cassandra
- memcached

Compute

- Globus Toolkit
- Xeround
- Beowulf
- Sun Grid Engine
- Hadoop
- OpenCloud
- Gigaspaces
- DataSynapse
- Xeround

Cloud Management

- 3Tera App Logic
- OpenNebula
- Open.ControlTier
- Enomaly Enomalism
- Altor Networks
- VMware vSphere
- OnPathTech
- CohesiveFT VPN Cubed
- Hyperic
- Eucalyptus
- Reductive Lbs Puppet
- OpenQRM
- Appistry

File Storage

- EMC Atmos
- ParaScale
- Zmanda
- CTERA

Appliances

- PingIdentity
- Simplified
- rPath
- Vordel

CLOUD TAXONOMY

Platform Services

General Purpose

- Force.com
- Etelos
- LongJump
- AppJet
- Rollbase
- Bungee Labs Connect
- Google App Engine
- Engine Yard
- Caspio
- Qrimp
- MS Azure Services Platform
- Mosso Cloud Sites

Business Intelligence

- Aster DB
- Quantivo
- Cloud9 Analytics
- Blink Logic
- K2 Analytics
- LogiXML
- Oco
- Panorama
- PivotLink
- Sterna
- ColdLight Neuron
- Infobright
- Vertica

Integration

- Amazon SQS
- MuleSource Mule OnDemand
- Boomi
- SnapLogic
- OpSource Connect
- Cast Iron
- Microsoft BizTalk Services
- gnip
- SnapLogic SaaS Solution Packs
- Appian Anywhere
- HubSpan
- Informatica On-Demand

Development & Testing

- Keynote Systems
- Mercury
- SOASTA
- SkyTap
- Aptana
- LoadStorm
- Collabnet
- Dynamsoft

Database

- Google BigTable
- Amazon SimpleDB
- FathomDB
- Microsoft SDS

Software Services

Billing

- Aria Systems
- eVapt
- OpSource
- Redi2
- Zuora

Financials

- Concur
- Xero
- Workday
- Beam4d

Legal

- DirectLaw
- Advologix
- Fios
- Sertifi

Sales

- Xactly
- LucidEra
- StreetSmarts
- Success Metrics

Productivity

- Zoho
- IBM Lotus Live
- Google Apps
- Desktoptwo
- Parallels
- ClusterSeven

Human Resources

- Taleo
- Workday
- iCIMS

Content Management

- Clickability
- SpringCM
- CrownPoint

Backup & Recovery

- JungleDisk
- Mozy
- Zmanda Cloud Backup
- OpenRSM
- Syncplicity

CRM

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

Document Management

- NetDocuments
- Questys
- DocLanding
- Aconex
- Xyθος
- Knowledge TreeLive
- SpringCM

Collaboration

- Box.net
- DropBox

Social Networks

- Ning
- Zemby
- Amitive

Cloud Storage

- Several large Web companies (such as Amazon and Google) have plenty of unused storage
 - Rent it to others
- Cloud storage
 - data are stored remotely
 - data are temporarily cached on desktop computers, mobile phones or other Internet-linked devices.
- Classical exemple :
 - Amazon's Elastic Compute Cloud (EC2)
 - Simple Storage Solution (S3)

Amazon Simple Storage Service (S3)

- Unlimited Storage.
- Pay for what you use:
 - \$0.20 per GByte of data transferred,
 - \$0.15 per GByte-Mon
 - Second Life Update:
 - 1TBytes, 40,000 downloads in 24 hours - \$200,



Utility Computing - EC2

- Amazon Elastic Compute Cloud (EC2):
 - Elastic, marshal 1 to 100+ PCs,
 - 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:
 - Hardware contributions by Intel (VT-x/Vanderpool) and AMD (AMD-V).
 - Supports “Live Migration” of a virtual machine between hosts.
- 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

- Hadoop: java framework
- Distributed manipulation of large amounts of data
- Data Intensive distributed applications.
- Amazon EC2 + Amazon S3.
- Use cases:
 - Web Indexing.
 - Data Mining.
 - Machine Learning.
 - Financial Analysis.
 - Scientific Simulation.

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**

- Dependence on others 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 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 energy 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:
 - Fewer large programs in local means better performance
 - Less memory used, same for CPU,...
 - 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!
 - most cloud computing applications are free today, such as the Google Docs suite.
 - better than paying for similar commercial software
 - which alone may be justification for switching to cloud applications.

Advantages of Cloud Computing

- Instant software updates:
 - No more choice between obsolete software and high upgrade costs.
 - For web-based applications, updates are automatic
 - available the next time you log into the cloud.
 - For web-based application, you log-in to the latest version
 - without needing to pay for or *download* an upgrade.
- Improved document format compatibility.
 - No more worry about the compatibility of your documents with other users' applications or OSes
 - No more format incompatibilities when is sharing documents and applications in the cloud.

Advantages of Cloud Computing

- **Unlimited storage capacity:**
 - Cloud computing offers virtually limitless storage.
 - 1 Tbyte hard drive compared to Pbytes in the cloud.
- **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.
 - 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:
 - Documents are not linked to a physical object
 - All you need is a computer and an Internet connection
 - Documents are instantly available from wherever you are
- Latest version availability:
 - 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
- 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.

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. ADSL/3G is the minimum.
 - Web-based applications require a lot of bandwidth to download, as do large documents.
- 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

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.
 - Changing fast due to improved javascript runtime and new paradigms (Hadoop)

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?
- 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. Also you must trust the Cloud.

Disadvantages of Cloud Computing

- HPC Systems:
 - Difficult to run HPC applications on MPI/OpenMP!
 - Scheduling is important with this type of application
 - as you want all the VM to be co-located to minimize communication latency!
- General Concerns:
 - Each cloud systems uses different protocols and different APIs
 - may not be possible to run applications between cloud based systems
 - Amazon has created its own DB system (not SQL), and workflow system (many popular workflow systems out there)
 - Normal applications need to be adapted

Conclusion

- No real novelty in Cloud computing ?
 - More merging of centralised and distributed ideas
 - The Grid idea made Reality !
- However some big concerns for users !
- Many new open source systems
 - Can run on local cluster
 - Can merge local cluster and outside clouds
- Next step : make it easy to manage
 - Currently most adaptation is done by hand !