CS4650 Topic 18:
Cloud Computing Goals and Benefits

A Definition

- Cloud computing is distributed computing using remotely provisioned scalable and measured resources.
  - Resources: CPUs, Data Storage, Software, etc.
  - Distributed: Not isolated to a single facility
  - Remotely: Resources typically not owned or managed by the users
  - Provisioned: The resources are configured as needed by the customer on demand
  - Scalable: If the customer's demand grows, additional resources are added, and when the demand shrinks, the extra resources are released.
  - Measured: The use of resources is monitored, so that the user only pays for the resources used.

Why Cloud Computing?

- Cost Reduction
  - In-house IT resources can be difficult to acquire and maintain.
  - Staffing required to keep the environment operational.
  - Upgrades and patches require time and effort.
  - Capital investments costs to purchase, refurbish and replace.
  - Utility bills.
  - Security, both to keep data and equipment safe and private.
  - Tracking licenses and support arrangements.

Why Cloud Computing?

- Capacity Planning
  - If a company uses in-house IT resources, then the company must decide the amount of each type of resource to purchase and support.
  - If the company decides on a level that is too small, then it will be *under-provisioned*, and the system will not be able to supply the level of service the company requires or desires.
  - If the company decides on a level that is too large, then it will be *over-provisioned*, and the system will cost the company more than necessary, as the extra capacity will be unused.
  - The demand for services will fluctuate based on the time of day and the day of the week, etc. To prevent under-provisioning, the company has to plan for the largest expected load. During non-peak times, the extra capacity will be unused.
  - Sometimes the demand will vary depending upon special projects or needs of the company's customers. These expected loads will add to the company's plan for required level of service.

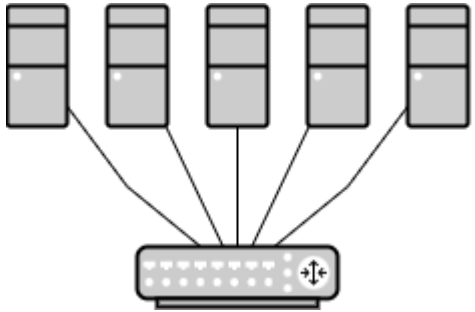
Why Cloud Computing?

- Capacity Planning
  - Companies use different capacity planning strategies:
  - *Lead Strategy* -- adding capacity to one or more of the IT resources in anticipation of an expected increase in demand. This strategy costs the company more money in the short-term, but typically keeps the company running "at full speed".
  - *Lag Strategy* -- adding capacity as an IT resource reaches its full capacity. This saves money, but this may be a false assumption since when a resource reaches capacity, work throughput is reduced until those new resources can be brought on-line.
  - *Match Strategy* -- Adding IT resources in small increments as the demand increases. However, for an in-house IT system, how are resources incrementally installed or deinstalled?

Why Cloud Computing?

- Organizational Agility
  - Businesses need to adapt and evolve to face changes in their market.
  - It may be difficult to accurately estimate the required level of IT services, especially if a sudden change brings a need for rapid IT scaling. Conversely, if there is a business downturn, there may be a need to scale back on the IT services.
  - If new technology becomes available, it may be prohibitively expensive in terms of both time and money to migrate to the new technology.
  - The cost to scale the IT services may make it prohibitive for a company to enable new or expanded business opportunities.
  - A company may be considering automating a portion of its operation. Again, the cost to provide these IT services in-house may be prohibitive.

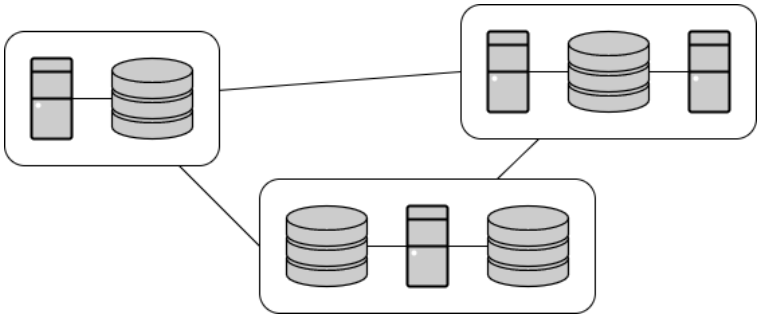
Technological Innovations



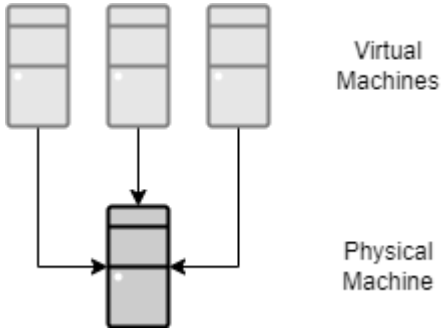
- *Clustering*: A group of roughly equivalent but independent IT resources are interconnected and operate as a single system.
- If any device fails, its workload is distributed to other devices in the system, reducing system failure rate and increasing availability and reliability.
- Components in the cluster are kept in synchronization through dedicated, high-speed communication links.
- The devices should have reasonably identical hardware and operating systems.

Technological Innovations

- *Grid Computing*: Computing resources are organized into logical pools. These pools are coordinated to provide a high performance distributed grid.
- Grid systems are loosely coupled and distributed.
- Grid systems can use heterogeneous devices that can be geographically dispersed.

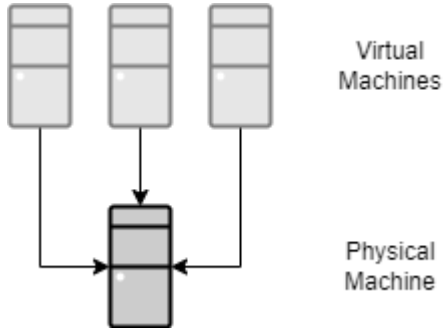


Technological Innovations



- *Virtualization*: A layer of virtualization software allows a *virtual machine* to run on a *physical machine*.
- The virtual machine acts like a standalone computer, and may have different software and a different operating system from the actual physical machine.
- Multiple virtual machines can be simultaneously hosted on a single physical machine.
- An error in one virtual machine will only affect that one virtual machine, the other virtual machines and the host machine are unaffected.

Technological Innovations



- With virtualization, the virtual machine can be configured to the exact requirements of a particular customer.
- The virtual machine can be ported to run on any physical machine, even on upgraded or enhanced processors.
- Virtual machines can also be replicated easily (it is just a software bundle), to allow for scaling as needed and on demand.
- In addition to processors, data storage and network interconnectivity hardware can also be virtualized.

Enabling Technologies

- In addition to the enabling technologies listed above, there are a number of other technologies that were being developed for other uses that are essential to Cloud Computing:
  - Internet Architecture
  - Broadband Networks
  - Web Technology
  - Service Technology
  - Data Center Technology

Key Points

- The primary business drivers that drove the need for cloud computing include cost reduction, capacity planning, and organizational agility.
- The primary technical innovations that enabled cloud computing include clustering, grid computing, and virtualization.

Basic Cloud Computing Concepts

- A *cloud* is a distinct IT environment that is designed for remotely provisioning scalable and measured IT resources.
- Originally, the internet itself was called 'the cloud', and it approximately satisfied the above definition. Various 'web requests' would be processed by servers located somewhere in the web, these would perform some IT services, and responses would be returned.
- The more modern description focuses on a *cloud* being IT resources that essentially can be *rented* from the provider to expand the IT capability of the client.

Basic Cloud Computing Concepts

- With the modern usage of the term *cloud*, a cloud has a finite boundary (whereas the Internet is not considered to have a finite boundary). Also, there are many individual clouds that are accessible via the Internet, although not all clouds are connected to the Internet.

Basic Cloud Computing Concepts

- *On-Premise*: The term *on-premise* refers to IT resources of a company that hosted within the boundary of the company, and which are not connected to a cloud.
  - A company may have a private cloud, or even host a public cloud.
  - Any of the resources that are in these clouds are considered not on-premise, even if they are located in the company's building.
  - An on-premise IT resource can access and interact with a cloud-based IT resource.
  - The distinction between cloud-based and on-premise is whether the resource is part of a cloud.

Basic Cloud Computing Concepts

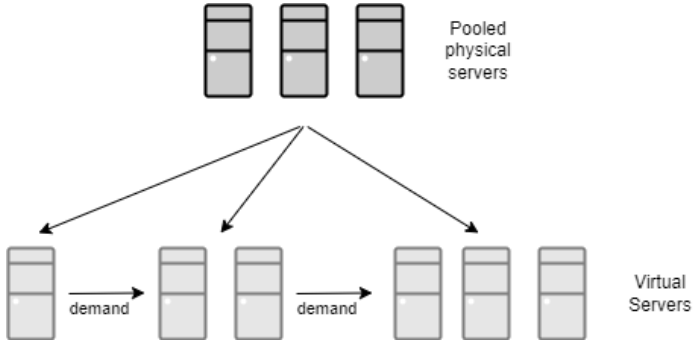
- *Cloud Consumers:* The party that uses cloud-based IT services is a *cloud consumer*.
- *Cloud Providers:* The party that provides cloud-based IT services is a *cloud provider.*
- The some situations a company may simultaneously be a cloud consumer and a cloud provider: Company *A* may provide cloud-based servers upon which company *B* implements cloud-based applications (or services or storage), which company *C* uses. With respect to company *A*, company *B* is a cloud consumer, but with respect to company *C*, company *B* is a cloud provider.

Basic Cloud Computing Concepts

- *Scaling:* From an IT resource perspective, scaling represents the ability of the resource to handle increased or decreased usage demands.
- There are two forms of scaling:
  - Horizontal scaling -- scaling out and scaling in
  - Vertical scaling -- scaling up and down

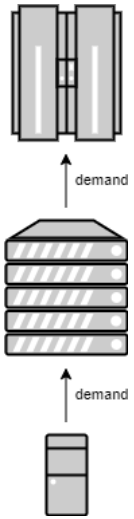
Basic Cloud Computing Concepts

- *Horizontal Scaling:* The allocating and releasing of IT resources that are of the same type and general capability, essentially the replicating of the resource, is called horizontal scaling.
- Allocating additional resources is considered *scaling out*.
- Releasing resources is considered *scaling in*.
- Sometimes this scaling is performed by creating copies of the virtual server. In other cases, this means utilizing additional physical servers.



Basic Cloud Computing Concepts

- *Vertical Scaling:* When an existing IT resource is replaced by another with either higher or lower capacity, this is considered vertical scaling.
- Replacing an IT resource with another more capable resource is considered *scaling up*.
- Replacing an IT resource with another less capable resource is considered *scaling down*.



Basic Cloud Computing Concepts

- Vertical scaling is limited, as physical devices can only expand so far.
- Horizontal scaling, while technically limited, is essentially unlimited, all that is required is to utilize additional resources. There is no inherent limit to the number of devices.
- Vertical scaling is more expensive: Physical devices with more capability are significantly more expensive than devices with more typical specifications.
- Vertical scaling takes time, as the new resource have to be purchased, installed, configured, and connected; Horizontal scaling is essentially instantaneous.
- Cloud computing most commonly uses horizontal scaling.

Risks and Challenges

- *Increased Security Vulnerabilities*
  - If a company keeps its business data on-premise, the company has the responsibility over data security. A company that is concerned about security can adopt strict security protocols and invest in software and hardware, as well as physical procedures and systems, to the degree the company determines is satisfactory.
  - The if the business data is moved to the cloud, data security becomes a shared responsibility with the cloud provider. The cloud provider security protocols adopted by the cloud provider may not be as strict. It would be the cloud provider that invests in the security software and hardware, along with the physical procedures and systems. For public clouds, it is unlikely that all desired security protocols would be in place.
  - Another source of vulnerability is in the Internet itself, through which the business connects to the cloud. With an on-premise network, external third parties would have no access to any sensitive data, while through the Internet vulnerabilities are present.

Risks and Challenges

- *Increased Security Vulnerabilities*
  - Another issue is that the cloud provider has privileged access to cloud consumer data. The company's trust boundary has to extend to include the cloud provider, yet the company has no access to the cloud provider's personnel, facilities, or policies.
  - Moreover, with the shared IT resources of the cloud provider, the trust boundary of one company will overlap the trust boundary of other consumers of that cloud. This means there is increased exposure of the business data.

Risks and Challenges

- *Reduced Operational Governance Control*
  - Cloud consumers are usually allotted a level of governance control that is lower than that over on-premise IT resources.
  - By keeping all resources on-premise, the IT resources are more under the control of the company.
  - When moving to the cloud, a company depends upon the quality of service of the cloud provider. The company must now depend upon an outside entity, and if the cloud provider does a poor job, the company may be jeopardized.
  - If the cloud provider goes out of business, the company may be seriously at risk!

Risks and Challenges

- *Limited Portability Between Cloud Providers*
  - There is no standardization between cloud providers. Each provider has its own set of capabilities and has its own set of IT resources being provided.
  - The protocols and software packages will be different.
  - The security protocols and software will be different.
  - Consequently, it can be very difficult to switch to a different cloud provider.

Risks and Challenges

- *Multi-Regional Compliance and Legal Issues*
  - Cloud providers establish data centers in affordable or convenient geographical locations.
  - Cloud consumers are often not aware of the physical location of their IT resources and data.
  - For some organizations, this can pose serious legal concerns pertaining to industry or governmental regulations.
  - For example, governments may have regulations that specify data privacy and storage policies.
    - For example, some UK laws require that personal data belonging to UK citizens be kept within the UK.
    - Some countries have laws that require some types of data to be disclosed to certain governmental agencies.

Real-Life Example #1

- Years ago, for my company, I bought a copy of Photoshop. It cost a lot of money, and it came on a disk.
- For a while, I could download any upgrades to the software.
- Adobe decided to switch to a cloud-based implementation of Photoshop.
- What were some of their motivations for switching?

Real-Life Example #1: Their Motivations

- By moving the application to the cloud, they could upgrade everyone's version of the software instantly: the software is on the cloud.
  - They did not have to distribute the updates.
  - They did not have to continue to support older versions of the software, making bug fixes a lot simpler.
- They had a significant issue with software piracy: People were copying and distributing copies of the disks.
  - One time I bought a new computer from a mom-and-pop computer shop. When I fired up the computer, I saw that they had pre-installed a version of Photoshop on the computer! Luckily I had purchased a copy of the program, so I was authorized to use it!
  - With the software being on the cloud, piracy ceased to be an issue.
  - This also means they saved a lot of time and money building or buying copy protection technology.

Real-Life Example #1: Their Motivations

- They also switched to a new pricing model, where you paid a monthly subscription to access the software.
  - The price was low enough that it was easy for new people to start using the program.
  - However, after about three years, the subscription was a lot more expensive than an outright purchase of the software.

Real-Life Example #1: My Motivations

- Because of the on-going monthly cost, Photoshop would have cost me a lot more money than the single purchase price for the software.
- I still have Photoshop running on my computer. Yes, it is the old version, but it does what I need. I did not need the occasional updates and new features that they provided.
- One advantage of the cloud-based version is that the software would be available to me on whatever computer I am using, as the software is on the cloud.

Real-Life Example #2

- I am currently working with a company (the *friends*) that is a seller of software from a third-party provider (the *authors*). The software has been sold to a few companies, and is in the process of being sold to a couple of new companies (the *customers*).
- The software is a very big ticket item.
- The software is mission critical to the customers, and contains sensitive company information.
- The authors are currently moving the software to the cloud (they have been talking about this for about 12 years).
- The customers *do not* want a cloud based solution!

Real-Life Example #2: Authors' Motivations

- The authors want a cloud based solution because it would be easier for them to administer, "install", and upgrade.
- This eliminates any copy protection issues.
- They are planning to increase the monthly service fees, since they are now hosting the software on their servers.

Real-Life Example #2: Customers' Motivations

- The customers do not want the new version of the software.
- The primary concern is that their sensitive company information would be on the cloud (see the security issues), and would be visible to the authors.
- The customers are in another country, and there are legal ramifications and governmental regulations that will be an issue.
- With the current version of the product, the software runs on the machines at the company, and are under their control. They make sure that everything is up-to-date and secure.
- The customer's site is on the opposite side of the world, and their operations would be jeopardized if there are any Internet disruptions. In addition, latency issues are also a large concern.
- They are not too happy about the increased fees.

A New Topic

- In a few weeks we will be talking more about the nuts and bolts of cloud computing.
- You will be doing some cloud computing using AWS (Amazon Web Services).
- You can get a jump on things by signing up now.

Sign Up for Free AWS Account

- You will be running one or more Cloud Computing demo on AWS.
- AWS has a free tier, which allows you to run these demos and use the AWS system for free, for a limited time (several months, if you place only a reasonable load on their system).
- Create your account at:

aws.amazon.com/free

AWS Tutorials

- After you have your AWS account, you can run some of the tutorials. Click on the 'Getting Started' tab, then browse for one or more tutorials that interest you.
- Some I found are:
  - Build, Train, and Deploy Machine Learning Model (10 min)
  - Build a basic web application (30 min)
  - Build a serverless web application (2 hrs) ← Unicorn rides!
  - Run a serverless "Hello World" with AWS Lambda (10 min)
  - Detect, Analyze and Compare Faces with Amazon Rekognition (10 min)
  - Deploy a LAMP Stack Application to Amazon Lightsail (10 min)
  - Getting Started with AWS CDK (35 min)

AWS Tutorials

- For one of your assignments, you will need to run an hour's worth of tutorials (using the time estimates they provide).
- Be sure to check the requirements of the tutorials, many are available in the free tier.
- Also check that you clean up the tutorial when done, so charges don't start accruing on your account.
- Pick tutorials that are of interest to you.
- When you are done with the tutorials, you will submit a quick report of which tutorial(s) you did, include a screenshot or two of each, and describe what you learned. There will be more details in the homework assignment.