CS5950 / CS6030
Cloud Computing

http://www.cs.wmich.edu/gupta/teaching/cs6030/6030cloudS17/cs6030cloud.php

Ajay Gupta
B239, CEAS
Computer Science Department
Western Michigan University

ajay.gupta@wmich.edu
276-3104

Acknowledgements

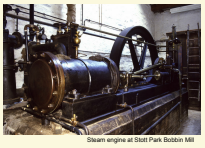
- I have liberally borrowed these slides and material from a number of sources including
  - Web, AWS Educate
  - MIT, Harvard, UMD, UPenn, UCSD, UW, Clarkson, . . .
  - Amazon, Google, IBM, Apache, ManjraSoft, CloudBook, . . .
- Thanks to original authors including Ives, Dyer, Lin, Dean, Buyya, Ghemawat, Fanelli, Bisciglia, Kimball, Michels-Slettvet,…
- If I have missed any, its purely unintentional. My sincere appreciation to those authors and their creative mind.

Recap: Computing at scale

- Modern applications require huge amounts of processing and data
  - Measured in petabytes, millions of users, billions of objects
  - Need special hardware, algorithms, tools to work at this scale
- Clusters and data centers can provide the resources we need
  - Main difference: Scale (room-sized vs. building-sized)
  - Special hardware; power and cooling are big concerns
- Clusters and data centers are not perfect
  - Difficult to dimension; expensive; difficult to scale

The power plant analogy


Waterwheel at the Neuhausen ob Eck Open-Air Museum


Steam engine at Stott Park Bobbin Mill

- It used to be that everyone had their own power source
  - Challenges are similar to the cluster: Needs large up-front investment, expertise to operate, difficult to scale up/down...

Scaling the power plant

- Then people started to build large, centralized power plants with very large capacity...

Metered usage model

- Power plants are connected to customers by a network
- Usage is metered, and everyone (basically) pays only for what they actually use

Why is this a good thing?

Electricity	Computing
• Economies of scale – Cheaper to run one big power plant than many small ones	Cheaper to run one big data center than many small ones
• Statistical multiplexing – High utilization!	High utilization!
• No up-front commitment – No investment in generator; pay-as-you-go model	No investment in data center; pay-as-you-go model
• Scalability – Thousands of kilowatts available on demand; add more within seconds	Thousands of computers available on demand; add more within seconds

What is cloud computing?

What is cloud computing?

The interesting thing about Cloud Computing is that we've redefined Cloud Computing to include everything that we already do.... I don't understand what we would do differently in the light of Cloud Computing other than change the wording of some of our ads.

Larry Ellison, quoted in the Wall Street Journal, September 26, 2008

A lot of people are jumping on the [cloud] bandwagon, but I have not heard two people say the same thing about it. There are multiple definitions out there of "the cloud".

Andy Isherwood, quoted in ZDnet News, December 11, 2008

So what is it, really?

- According to NIST:

  Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

- Essential characteristics:
  - On-demand self service
  - Broad network access
  - Resource pooling
  - Rapid elasticity
  - Measured service

Other terms you may have heard

- Utility computing
  - The service being sold by a cloud
  - Focuses on the business model (pay-as-you-go), similar to classical utility companies
- The Web
  - The Internet's information sharing model
  - Some web services run on clouds, but not all
- The Internet
  - A network of networks.
  - Used by the web; connects (most) clouds to their customers

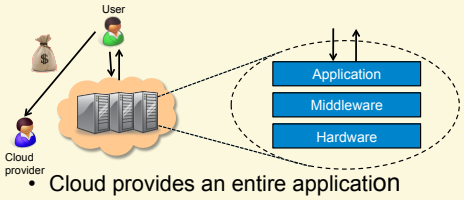
Plan for today

- AWS starter – EC2, VPC, SecurityGroups, Storage ✓
- Computing at scale
  - The need for scalability; scale of current services ✓
  - Scaling up: From PCs to data centers ✓
  - Problems with 'classical' scaling techniques ✓
- Utility computing and cloud computing
  - What are utility computing and cloud computing? ✓
  - What kinds of clouds exist today? NEXT
  - What kinds of applications run on the cloud?
  - Virtualization: How clouds work 'under the hood'
  - Some cloud computing challenges

Everything as a Service

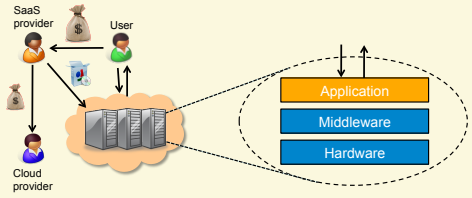
- What kind of service does the cloud provide?
  - Does it offer an entire application, or just resources?
  - If resources, what kind / level of abstraction?
- Three types commonly distinguished:
  - Software as a service (SaaS)
    - Analogy: Restaurant. Prepares&serves entire meal, does the dishes, ...
  - Platform as a service (PaaS)
    - Analogy: Take-out food. Prepares meal, but does not serve it.
  - Infrastructure as a service (IaaS)
    - Analogy: Grocery store. Provides raw ingredients.
  - Other xaaS types have been defined, but are less common
    - Desktop, Backend, Communication, Network, Monitoring, ...

Software as a Service (SaaS)



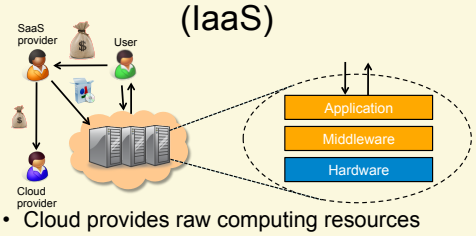
- Cloud provides an entire application
  - Word processor, spreadsheet, CRM software, calendar...
  - Customer pays cloud provider
  - Example: Google Apps, Salesforce.com

Platform as a Service (PaaS)



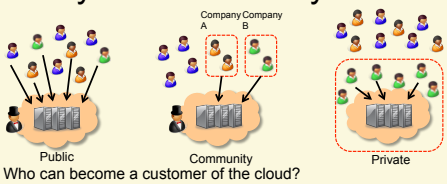
- Cloud provides middleware/infrastructure
  - For example, Microsoft Common Language Runtime (CLR)
  - Customer pays SaaS provider for the service; SaaS provider pays the cloud for the infrastructure
  - Example: Windows Azure, Google App Engine

Infrastructure as a Service (IaaS)



- Cloud provides raw computing resources
  - Virtual machine, blade server, hard disk, ...
  - Customer pays SaaS provider for the service; SaaS provider pays the cloud for the resources
  - Examples: Amazon Web Services, Rackspace Cloud, GoGrid

Private/hybrid/community clouds



- Who can become a customer of the cloud?
  - Public cloud: Commercial service; open to (almost) anyone. Example: Amazon AWS, Microsoft Azure, Google App Engine
  - Community cloud: Shared by several similar organizations. Example: Google's "Gov Cloud"
  - Private cloud: Shared within a single organization. Example: Internal datacenter of a large company.

Focus of this class →

Is this a 'real' cloud? →

Plan for today

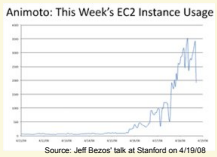
- AWS starter – EC2, VPC, SecurityGroups, Storage ✓
- Computing at scale
  - The need for scalability; scale of current services ✓
  - Scaling up: From PCs to data centers ✓
  - Problems with 'classical' scaling techniques ✓
- Utility computing and cloud computing
  - What are utility computing and cloud computing? ✓
  - What kinds of clouds exist today? ✓
  - What kinds of applications run on the cloud? NEXT
  - Virtualization: How clouds work 'under the hood'
  - Some cloud computing challenges

Examples of cloud applications

- Application hosting
- Backup and Storage
- Content delivery
- E-commerce
- High-performance computing
- Media hosting
- On-demand workforce
- Search engines
- Web hosting

Case study: animoto

- Animoto: Lets users create videos from their own photos/music
  - Auto-edits photos and aligns them with the music, so it "looks good"



Source: Jeff Bezos' talk at Stanford on 4/19/08

- Built using Amazon EC2+S3+SQS
- Released a Facebook app in mid-April 2008
  - More than 750,000 people signed up within 3 days
  - EC2 usage went from 50 machines to 3,500 (x70 scalability!)

http://animoto.com/blog/company/amazon-com-ceo-jeff-bezos-on-animoto/

Case study: The Washington Post

- March 19, 2008: Hillary Clinton's official White House schedule released to the public
  - 17,481 pages of non-searchable, low-quality PDF
  - Very interesting to journalists, but would have required hundreds of man-hours to evaluate
  - Peter Harkins, Senior Engineer at The Washington Post: Can we make that data available more quickly, ideally within the same news cycle?
  - Tested various Optical Character Recognition (OCR) programs; estimated required speed
  - Launched 200 EC2 instances; project was completed within nine hours (!) using 1,407 hours of VM time ($144.62)
  - Results available on the web only 26 hours after the release

http://aws.amazon.com/solutions/case-studies/washington-post/

Other examples

- DreamWorks is using the Cerelink cloud to render animation movies
  - Cloud was already used to render parts of *Shrek Forever After* and *How to Train your Dragon*

- CERN is working on a "science cloud" to process experimental data
- Virgin Atlantic is hosting their new travel portal on Amazon AWS

Recap: Utility/cloud computing

- Why is cloud computing attractive?
  - Analogy to 'classical' utilities (electricity, water, ...)
  - No up-front investment (pay-as-you-go model)
  - Low price due to economies of scale
  - Elasticity - can quickly scale up/down as demand varies
- Different types of clouds
  - SaaS, PaaS, IaaS; public/private/community clouds
- What runs on the cloud?
  - Many potential applications: Application hosting, backup/ storage, scientific computing, content delivery, ...
  - Not yet suitable for certain applications (sensitive data, compliance requirements)

Is the cloud good for everything?

- No.
- Sometimes it is problematic, e.g., because of auditability requirements
- Example: Processing medical records
  - HIPAA (Health Insurance Portability and Accountability Act) privacy and security rule
- Example: Processing financial information
  - Sarbanes-Oxley act
- Would you put your medical data on the cloud?
  - Why / why not?

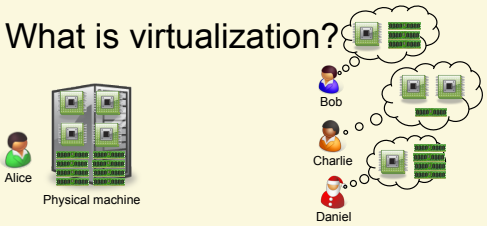
Recap: Cloud applications

- Clouds are good for many things...
  - Applications that involve large amounts of computation, storage, bandwidth
  - Especially when lots of resources are needed quickly (Washington Post example) or load varies rapidly (TicketLeap example)
- ... but not for all things

Plan for today

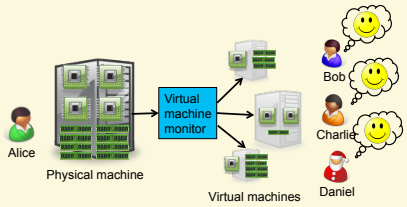
- AWS starter – EC2, VPC, SecurityGroups, Storage ✓
- Computing at scale
  - The need for scalability; scale of current services ✓
  - Scaling up: From PCs to data centers ✓
  - Problems with 'classical' scaling techniques ✓
- Utility computing and cloud computing
  - What are utility computing and cloud computing? ✓
  - What kinds of clouds exist today? ✓
  - What kinds of applications run on the cloud? ✓
  - Virtualization: How clouds work 'under the hood' NEXT
  - Some cloud computing challenges

What is virtualization?



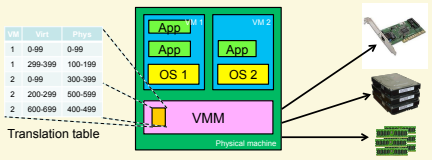
- Suppose Alice has a machine with 4 CPUs and 8 GB of memory, and three customers:
  - Bob wants a machine with 1 CPU and 3GB of memory
  - Charlie wants 2 CPUs and 1GB of memory
  - Daniel wants 1 CPU and 4GB of memory
- What should Alice do?

What is virtualization?



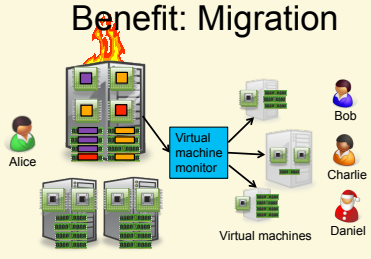
- Alice can sell each customer a virtual machine (VM) with the requested resources
  - From each customer's perspective, it appears as if they had a physical machine all by themselves (isolation)

How does it work?



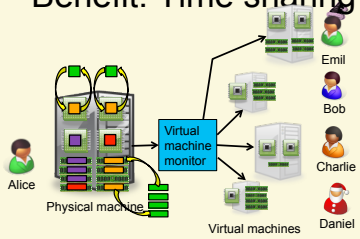
- Resources (CPU, memory, ...) are virtualized
  - VMM ("Hypervisor") has translation tables that map requests for virtual resources to physical resources
  - Example: VM 1 accesses memory cell #323; VMM maps this to memory cell 123.
  - For which resources does this (not) work?
  - How do VMMs differ from OS kernels?

Benefit: Migration



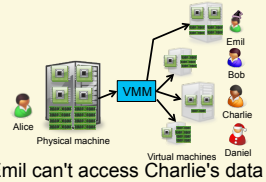
- What if the machine needs to be shut down?
  - e.g., for maintenance, consolidation, ...
  - Alice can migrate the VMs to different physical machines without any customers noticing

Benefit: Time sharing



- What if Alice gets another customer?
  - Multiple VMs can time-share the existing resources
  - Result: Alice has more virtual CPUs and virtual memory than physical resources (but not all can be active at the same time)

Benefit and challenge: Isolation



- Good: Emil can't access Charlie's data
- Bad: What if the load suddenly increases?
  - Example: Emil's VM shares CPUs with Charlie's VM, and Charlie suddenly starts a large compute job
  - Emil's performance may decrease as a result
  - VMM can move Emil's software to a different CPU, or migrate it to a different machine

Recap: Virtualization in the cloud

- Gives cloud provider a lot of flexibility
  - Can produce VMs with different capabilities
  - Can migrate VMs if necessary (e.g., for maintenance)
  - Can increase load by overcommitting resources
- Provides security and isolation
  - Programs in one VM cannot influence programs in another
- Convenient for users
  - Complete control over the virtual 'hardware' (can install own operating system, own applications, ...)
- But: Performance may be hard to predict
  - Load changes in other VMs on the same physical machine may affect the performance seen by the customer

Plan for today

- AWS starter – EC2, VPC, SecurityGroups, Storage ✓
- Computing at scale
  - The need for scalability; scale of current services ✓
  - Scaling up: From PCs to data centers ✓
  - Problems with 'classical' scaling techniques ✓
- Utility computing and cloud computing
  - What are utility computing and cloud computing? ✓
  - What kinds of clouds exist today? ✓
  - What kinds of applications run on the cloud? ✓
  - Virtualization: How clouds work 'under the hood' ✓
  - Some cloud computing challenges NEXT

10 obstacles and opportunities

1. Availability
   - What happens to my business if there is an outage in the cloud?
2. Data lock-in
   - How do I move my data from one cloud to another?
3. Data confidentiality and auditability
   - How do I make sure that the cloud doesn't leak my confidential data?
   - Can I comply with regulations like HIPAA and Sarbanes/Oxley?

Service	Duration	Date
S3	6-8 hrs	7/20/08
AppEngine	5 hrs	6/17/08
Gmail	1.5 hrs	8/11/08
Azure	22 hrs	3/13/09
Intuit	36 hrs	6/16/10
EBS	>3 days	4/21/11
ECC	~2 hrs	6/30/12

Some cloud outages

10 obstacles and opportunities

4. Data transfer bottlenecks
   - How do I copy large amounts of data from/to the cloud?
   - Example: 10 TB from UC Berkeley to Amazon in Seattle, WA
   - Motivated Import/Export feature on AWS
5. Performance unpredictability
   - Example: VMs sharing the same disk → I/O interference
   - Example: HPC tasks that require coordinated scheduling

Method	Time
Internet (20Mbps)	45 days
FedEx	1 day

Time to transfer 10TB [AF10]

Primitive	Mean perf.	Std dev
Memory bandwidth	1.3GB/s	0.05GB/s (4%)
Disk bandwidth	55MB/s	9MB/s (16%)

Performance of 75 EC2 instances in benchmarks

10 obstacles and opportunities

6. Scalable storage
   - Cloud model (short-term usage, no up-front cost, infinite capacity on demand) does not fit persistent storage well
7. Bugs in large distributed systems
   - Many errors cannot be reproduced in smaller configs
8. Scaling quickly
   - Problem: Boot time; idle power
   - Fine-grain accounting?

10 obstacles and opportunities

9. Reputation fate sharing
   - One customer's bad behavior can affect the reputation of others using the same cloud
   - Example: Spam blacklisting, FBI raid after criminal activity
10. Software licensing
   - What if licenses are for specific computers?
     - Example: Microsoft Windows
   - How to scale number of licenses up/down?
     - Need pay-as-you-go model as well

Stay tuned

Next you will learn about:
Programming at scale; Parallelism, Concurrency and Consistency