Cloud Computing for Education Workshop

Introduction to Cloud Computing

© 2012 Copyright REZA CURTMOLA, NJIT
Parts of this presentation are courtesy of Prof. Cristian Borcea, NJIT

Organizers

- Reza Curtmola

- Center of Pre-College Programs (CPCP)
  - John Carpinelli
  - Howard Kimmel
  - Levelle Burr-Alexander

- Bo Chen (PhD Student)

- Kathy James
Workshop Overview

• July 2, 3, 5 (Monday, Tuesday, Thursday)
• One more day in beginning of December
• Each day:
  § Breakfast (8:30 – 9:00)
  § AM session
  § Lunch (12:00 – 1:00)
  § PM session
• WiFi connectivity:
  § NJIT network
  § User name: guest745
  § Password: stocky52

CCEW '12  Introduction to Cloud Computing  3

Workshop Overview

• Monday
  § AM: Welcome, Opening Remarks, Introductions, Pre-workshop Surveys, Overview of Cloud Services
  § PM: Storing and Sharing Data in the Cloud
• Tuesday
  § AM: Clouds in Education and Collaboration in and out of the Classroom
  § PM: Cloud-based Tools for Real-time Collaboration, Course Management using Piazza
• Thursday
  § AM: Standards-based Lesson Planning and Post-workshop Assignment, Creating a Lesson Plan
  § PM: Using Public Data Sets Available in Amazon's Cloud, Post-workshop Survey, Closing Remarks

CCEW '12  Introduction to Cloud Computing  4
Cloud computing at a glance

- Utility computing: our data and applications are hosted somewhere in the Internet ("in the cloud")
  - Most services we access over the Internet are in the cloud (e.g., Google, Amazon, Yahoo)

- Benefits:
  - Providers: economies of scale by having many users sharing the same infrastructure
  - Consumers: reduced cost and overhead

Cloud infrastructure = Data centers with 100,000’s servers

Tech predictions

- Gartner’s 2011 Strategic Technologies
  1. Cloud Computing
  2. Mobile Applications
  3. Social Computing
  4. …

![Worldwide Public IT Cloud Services Spending (SB) by Offering Category 2009, 2014](chart.png)
How it all started

• Very large data centers with low utilization on average
  ▪ How about renting computers when load is low?

• Google data center video:

What is cloud computing?

• Cloud = the name comes from the figures commonly used to represent the Internet as a cloud or as a series of connected clouds
  ▪ Idea: computing doesn’t have to happen on my desktop or in my building – it can be anywhere

• Technical view
  ▪ Use of computing resources available as a service (Software, Platforms, or Infrastructure)
  ▪ Resources can be accessed across the Internet

• Business view
  ▪ Reduced costs through sharing and centralization
  ▪ Flexibility and adaptability by rapid provisioning
What is cloud computing?

- The word “Cloud” has been used and abused
  - Different people have different definitions of what the cloud means
- “Cloud” is a buzz word nowadays: everyone is doing things related to the cloud, everything is being relabeled to include the word “cloud”, a lot of hype
- Examples:
  - Commercial platforms for cloud services (Amazon Web Services, Microsoft Azure, Google, IBM, etc.)
    - Financial companies (such as hedge funds) use cloud computing for financial data simulations that require intensive computation
    - Intensive scientific computations
    - The cloud is used for content distribution
  - Online storage providers (DropBox, SugarSync, …)
  - Online services for sharing photos, videos, etc.
  - Email (Gmail, Yahoo Mail, etc.)
  - The cloud is used as an extension for mobile devices
  - Cloud computing in governments more and more
  - Pretty much any web service runs in the cloud nowadays

Cloud attributes (IDC)

Key Cloud Services Attributes
(cloud offerings must meet all eight criteria)

- Off-Site, Third-party provider
- Accessed via Internet
- Minimal/no IT skills required to “implement”
- Provisioning = self-service requesting; near real-time deployment; dynamic & fine-grained scaling
- Pricing model = fine-grained, usage-based (at least available as an option)
- UI = browser and successors
- System Interface = web services APIs
- Shared resources/common versions (customization “around” the shared services)

Source: IDC, September 2008
Cloud requirements (Yahoo!)

- Multi-tenancy: Many apps co-existing on the same infrastructure
- Elasticity: Fast and graceful response to changing resource requirements
- Scalability: Scaling to growing data and apps
- Load and tenant balancing: absorbing load spikes, not to overload the hardware
- Availability: The cloud must be always on
- Security: No security breach into the cloud
- Metering: Monitoring cloud usage for resource provisioning and costing
- Simple APIs: Simplify deploying and tuning applications in the cloud

Cloud Services

- Cloud services are available as a utility in a pay-per-use manner: only pay for what you use
  - no need to be concerned about over-provisioning or under-provisioning
  - Example: website for US Open tennis tournament
    - High traffic volume for two weeks
    - Reduced usage otherwise
  - This elasticity of resources, without paying a premium for large scale, is unprecedented in the history of IT
The NIST Definition of Cloud Computing

- NIST = National Institute of Standards and Technology
- Cloud computing is an evolving paradigm

Cloud computing is a model for enabling ubiquitous, 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.

The NIST Cloud Model – Five characteristics

1. **On-demand self-service**
   - A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

2. **Broad network access**
   - Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).
The NIST Cloud Model – Five characteristics

3. Resource pooling

- The provider’s computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.
- The customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter).
- Examples of resources include storage, processing, memory, and network bandwidth.

4. Rapid elasticity

- Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand.
- To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

5. Measured service

- Cloud systems automatically control and optimize resource use by leveraging a metering capability (done on a pay-per-use basis)
- Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.
Cloud entities

- **Cloud providers**: offer hardware & management tools
  - Many times provide system services (e.g., local OS + persistent storage + system software like compilers)
  - Can provide other (user-level) services
- **Cloud customers**
  - **Service providers**: use the cloud to provide user-level services (accessible across the Internet)
  - **Users**: use the services offered by service providers

Types of services

- **Infrastructure-as-a-Service (IaaS)**
  - Virtual servers with unique IP addresses and blocks of storage on demand (e.g., Amazon EC2)
- **Platform-as-a-Service (PaaS)**
  - Set of software and development tools (API) hosted on the provider’s servers (e.g., Windows Azure)
- **Software-as-a-Service (SaaS)**
  - The provider allows the customer only to use its applications (e.g., web-based email, web stores, etc)
The NIST Cloud Model – Three Service Models

1. Infrastructure as a Service (IaaS)
   • The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software.
   • The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications.
   • A cloud infrastructure is the collection of hardware and software that enables the five essential characteristics of cloud computing.
     § Physical layer (hardware: server, storage and network components)
     § Abstraction layer (software deployed across the physical layer)

2. Platform as a Service (PaaS)
   • The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider.
   • The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.
The NIST Cloud Model – Three Service Models

3. Software as a Service (SaaS)
   • The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure.
   • The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email).
   • The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.
What About Security?

- **Infrastructure-as-a-Service (IaaS)**
  - Security provisions beyond the basic infrastructure are carried out mainly by the cloud consumer.

- **Platform-as-a-Service (PaaS)**
  - Security provisions are split between the cloud provider and the cloud consumer.

- **Software-as-a-Service (SaaS)**
  - Security provisions are carried out mainly by the cloud provider.

---

**Spectrum of clouds (1)**

- Lower-level, Less management
- Higher-level, More management

<table>
<thead>
<tr>
<th>Amazon EC2</th>
<th>Microsoft Azure</th>
<th>Google AppEngine</th>
</tr>
</thead>
</table>

---

CCEW '12 Introduction to Cloud Computing
Spectrum of clouds (2)

- Amazon EC2: an instance looks more like the physical hardware & users control entire software stack
  - Difficult to offer automatic scalability and failover because they are application dependent
- Microsoft Azure: applications written in .NET and compiled to CLR (language independent environment)
  - Recently included inter-operability support for other languages
  - Libraries offer some support for failover/scalability, but the developer must declare some application properties
- Google AppEngine: targeted exclusively at web applications (not general computing)
  - Clean separation between stateless computation tier and stateful storage tier allows for automatic scaling & high availability mechanisms

The NIST Cloud Model – Four Deployment Models

1. Public Cloud
   - The cloud infrastructure is provisioned for open use by the general public.
   - It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them.
   - It exists on the premises of the cloud provider.

2. Private Cloud
   - The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units).
   - It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.
The NIST Cloud Model – Four Deployment Models

3. Community Cloud
   - The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations).
   - It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

4. Hybrid Cloud
   - The cloud infrastructure is a composition of two or more distinct cloud infrastructures (public, private, or community) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

New application opportunities

- Parallel batch processing
  - Get fast answers when processing large amounts of data
  - “Using hundreds of computers for a short time costs the same as using a few computers for a long time”

- Mobile applications & services
  - Store large data sets & perform complex computations in the cloud (e.g., mashups, augmented reality)
  - Could maintain a virtual copy of your device in the cloud & potentially offload computations there

- Extensions of desktop software
  - Matlab, Mathematica
Business view

• Shift in economic model: the relevance and weight of technology diminishes as it becomes commodity and standardized (electricity grid, phone, water supply)

• Providers: economies of scale by having many users sharing the same infrastructure
  ▪ Main reason why cloud computing happened when very large data centers have started to appear

• Consumers: reduced cost and overhead

Provider’s benefits

• Renting an IT infrastructure to many users
• Sharing of resources
• Reduction of costs through scale
• Centralized monitoring and maintenance
• Control over software evolution
• Control over service level agreements
Economics of cloud providers

• 5-7x economies of scale [Hamilton 2008]

<table>
<thead>
<tr>
<th>Resource</th>
<th>Cost in Medium DC</th>
<th>Cost in Very Large DC</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network</td>
<td>$95 / Mbps / month</td>
<td>$13 / Mbps / month</td>
<td>7.1x</td>
</tr>
<tr>
<td>Storage</td>
<td>$2.20 / GB / month</td>
<td>$0.40 / GB / month</td>
<td>5.7x</td>
</tr>
<tr>
<td>Administration</td>
<td>≈140 servers/admin</td>
<td>&gt;1000 servers/admin</td>
<td>7.1x</td>
</tr>
</tbody>
</table>

• Extra benefits
  ▪ Amazon: utilize off-peak capacity
  ▪ Microsoft: sell .NET tools
  ▪ Google: reuse existing infrastructure

Consumer’s benefits

• Reduced capital expenditure for hardware, software, services
• Reduced operational expenses - Pay as you go (dynamic provisioning)
• Simple to achieve scalability and flexibility
• More predictable costs
• Note: there are scenarios where cloud is more expensive than owning an infrastructure
Consumer’s benefits

- Applications run in a scalable, reliable, available, and fault tolerant manner
- Customers (application owners) are shielded from the complexities of what’s needed to run a service
  - Purchase hardware
  - Setup infrastructure for a production environment
- Customers are shielded from maintaining the service running
  - Installing and maintaining the OS
  - Managing the networking issues
  - Load balancers
  - Troubleshooting hardware when it fails or when it needs more capacity
- Money is saved and complexity is reduced
- App-centric environment: app owners focus on the application and the cloud computing infrastructure executes it

Cloud benefits

- Watch video:

Consumer’s benefits: example

Using cloud services for governments

- Usage of cloud services for governance has been exploding
  - Both federal and state governments
- No more bureaucratic struggling with technology: procurement of hardware, vendor selection or storage capacity
  - All the technological firepower without any of the headaches of ownership and maintenance
- No longer have to predict demand: the cloud is infinitely scalable
- Using a cloud service like Amazon Web Services, anyone can access nearly unlimited computing power in a matter of minutes
  - With a few clicks, anyone can rent virtually unlimited computing capacity and storage
- Cloud computing has the potential to overhaul IT in the public sector
- Huge cost savings potential: the feds expect to close 800 of their roughly 2,100 data centers by 2015.

CCEW '12

Consumer’s benefits: example

Using cloud services for governments

- The state of Utah is looking to save $4 million a year in hosting services by consolidating data centers, virtualizing servers and moving to a private cloud platform which will eventually deliver hosted e-mail and Web applications to cities and counties throughout the state.
- Oregon will save an estimated $1.5 million annually by partnering with Google to offer the state’s schools cloud-based computing. The initiative provides Oregon’s public schools the ability to transition e-mail, calendars, online documents, video conferencing and website creation to Google’s Apps for Education services.
- NJ Transit uses a cloud-hosted Customer Relationships Management (CRM) software from Salesforce.com
  - NJ Transit is the nation’s largest statewide public transportation system providing bus, rail, and light rail services of over 900,000 daily trips on 247 bus routes, 26 bus stations, 11 commuter rail lines, and three light rail lines.
  - Serves 164 rail stations, 60 light rail stations and 19,800 bus stops.
- More examples:
  http://www.info.apps.gov/content/state-and-local-cloud-computing-case-studies
Traditional data center: heavy penalty for under-provisioning

Resources

Capacity

Demand

Time (days)

1

2

3

Lost revenue

Lost users

Economics of cloud model

Resources

Capacity

Demand

Time

Static data center

Data center in the cloud

Unused resources
Drawbacks of the cloud

- Too much control on the provider side
  - Data owners lose control over their data
  - Can outsource data storage and computation, but cannot outsource liability for losing data and service

- Privacy and security
  - Biggest hurdle to widespread cloud use (especially when using the cloud for governance, or when handling sensitive data)

- Legislation related to data management
- Data movement costs (data lock-in)
- Data transfer bottlenecks (hard to upload large amounts of data)
IDC cloud survey (2)

Q: Rate the benefits commonly ascribed to the 'cloud'/on-demand model

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy/fast to deploy</td>
<td>63.6%</td>
</tr>
<tr>
<td>Pay only for what you use</td>
<td>61.5%</td>
</tr>
<tr>
<td>Less in-house IT staff, costs</td>
<td>57.0%</td>
</tr>
<tr>
<td>Low monthly payments</td>
<td>53.3%</td>
</tr>
<tr>
<td>Offers the latest functionality</td>
<td>50.0%</td>
</tr>
<tr>
<td>Encourages more standard IT</td>
<td>46.3%</td>
</tr>
<tr>
<td>Sharing systems/information simpler</td>
<td>43.4%</td>
</tr>
<tr>
<td>It's the way of the future</td>
<td>29.1%</td>
</tr>
</tbody>
</table>

Source: IDC Enterprise Panel, August 2008 (n=244)

CCEW '12 Introduction to Cloud Computing 41

IDC cloud survey (3)

Q: Rate the challenges/issues ascribed to the 'cloud'/on-demand model

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security</td>
<td>74.6%</td>
</tr>
<tr>
<td>Performance</td>
<td>63.1%</td>
</tr>
<tr>
<td>Availability</td>
<td>63.1%</td>
</tr>
<tr>
<td>Hard to integrate with in-house IT</td>
<td>51.1%</td>
</tr>
<tr>
<td>Not enough ability to customize</td>
<td>55.0%</td>
</tr>
<tr>
<td>Worried on-demand will cost more</td>
<td>50.4%</td>
</tr>
<tr>
<td>Bringing back in-house may be difficult</td>
<td>50.0%</td>
</tr>
<tr>
<td>Regulatory requirements prohibit cloud</td>
<td>49.2%</td>
</tr>
<tr>
<td>Not enough major suppliers yet</td>
<td>44.3%</td>
</tr>
</tbody>
</table>

Source: IDC Enterprise Panel, August 2008 (n=244)

CCEW '12 Introduction to Cloud Computing 42
IDC cloud survey (4)

Q: Importance of IT cloud services supplier attributes
(1=not important, 5=very important)

- Offer competitive pricing: 83.2%
- Offer performance-level assurances/SLAs: 81.1%
- Understand my business and industry: 68.0%
- Can move cloud offerings back on-premise: 67.2%
- Can provide a complete solution: 61.9%
- Are future-oriented, an innovator: 58.2%
- Can support many of my IT needs: 55.4%
- Have a large network of partners: 54.1%
- Have local presence: 46.3%
- Are a large, established company: 40.6%
- Have done business with my organization: 30.4%

Source: IDC Enterprise Panel, August 2009 n=244

CCEW '12
Introduction to Cloud Computing

Amazon Web Services
(Amazon’s cloud computing platform)
Amazon Web Services (AWS) - Outline

- AWS Overview
- Application example
- Using Amazon EC2 instances (Thursday!)

Amazon Teaching Grants

- With AWS in Education, educators, academic researchers, and students can apply to obtain free usage credits to tap into the on-demand infrastructure of the Amazon Web Services cloud to teach advanced courses, tackle research endeavors, and explore new projects – tasks that previously would have required expensive up-front and ongoing investments in infrastructure

http://aws.amazon.com/education/
Amazon Web Services (1)

• Cloud services that can be managed (e.g., created, terminated), monitored, and accessed over the web
  ▪ The name is a bit misleading: they are not your typical multi-tier web services

• Compute
  ▪ Elastic Compute Cloud (EC2): launch on-demand virtual machines (instances)
  ▪ Elastic MapReduce: automatically starts Hadoop implementation of MapReduce for parallel applications
    • Amazon handles cluster management
  ▪ Auto Scaling: seamlessly increase/decrease number of EC2 instances function of load
    • Done based on metrics reported by CloudWatch

• Monitoring
  ▪ CloudWatch: monitor cloud resources such as CPU cycles, disk access, network traffic

Virtual Machines

Applications
Operating System
Virtual Machine

Applications
Operating System
Virtual Machine

Applications
Operating System
Virtual Machine

Hypervisor

HARDWARE
(CPU, Memory, Network card, HDD)

A machine in the cloud data center
Hardware Virtualization

- On a physical machine (which usually has a CPU with many cores), multiple virtual machines can be run
- A Virtual Machine Manager (VMM), also known as Hypervisor, sits on top of the machine hardware, and allows multiple operating systems (guests) to run concurrently on the host machine
  - Each operating system runs inside of a Virtual Machine
  - The Hypervisor connects the Virtual Machines with the physical hardware
  - The Hypervisor presents to guest operating systems a virtual operating platform; the guest OS thinks it is using the hardware directly; the guest OS is not aware that other OSs are using the same hardware.
  - Multiple virtual machines share the virtualized hardware resources

Virtual Machines

![Diagram of virtual machines and hypervisor](image)
Amazon Web Services (2)

- **Storage**
  - *Simple Storage Service (S3)*: provide persistent storage
    - Independent of EC2 instances
    - EC2 instances need to “download” data from S3 in order to access it (cannot issue read/write to S3)
  - *Elastic Block Store (EBS)*: provide block level storage volumes (i.e., disk-like) to EC2 instances
    - Persistent even after instances are terminated
    - Instances have to mount EBSs
  - *Import/Export*: move large amounts of data into and out of AWS using portable storage devices mailed to Amazon
    - Faster & cheaper than using the Internet to transfer the data

Amazon Web Services (3)

- **Databases**
  - *SimpleDB*: non-relational database (designed for high scalability & availability)
  - *Relational Database Service*: full-featured MySQL database
- **Messaging**
  - *Simple Queue Service (SQS)*: reliable & scalable mechanism for communication between EC2 instances or between machines outside the cloud and EC2 instances
Amazon Web Services (4)

- Networking
  - Route 53: DNS service that provides automatic availability & scalability
    - DNS (Domain Name System) translates domain to IP addresses
  - Virtual Private Cloud: allows integration of company machines with EC2 instances in one network (using IPSec VPN)
  - Elastic Load Balancing: automatically distributes incoming traffic to multiple EC2 instances

- Web
  - CloudFront: CDN (content distribution network) service
    - Like all CDNs uses caches at the edge of the network
    - Works for both static content and streaming
  - Alexa Web Information Service (AWIS): provides web analytics

Amazon Web Services (5)

- Deployment and management
  - Elastic Beanstalk: platform-as-a-service that allows simple deployment of Java + Apache Tomcat applications
    - Similar in nature with Windows Azure
    - Users can still access the infrastructure if needed

- Other Services
  - Fulfillment Web Service (FWS): process customer orders on behalf of third-party merchants
  - DevPay: billing system for customers to create billing apps
  - Flexible Payment Service (FPS): process Amazon user payments for third-party vendors (integrated in third-party apps)
  - Mechanical Turk: crowdsourcing (users are paid to do some simple tasks on the web)
EC2 instances

- **Instances**: Virtual machines that run in the EC2 environment
  - Each instance is like a physical machine that has its own CPU, memory, network interface, and disk space (volatile – data is lost when the instance is terminated)
  - *Xen* used as hypervisor for virtualization
- **AMI (Amazon Machine Image)**: Encrypted file that captures a complete snapshot of an EC2 instance at a point in time, including its software, configuration, and data
  - Images are stored in S3 and serve as boot disks for instances
  - Linux and Windows images publicly available
  - Users can create AMI from scratch: start from any public AMI, install & customize the software needed, and then store it as private AMI to use later on

EC2 environment

- Provides instance management and configuration services
  - Launch and terminate instances
  - Control the instance properties (e.g., type of instance, AMI) through a simple web interface
  - Set network access permissions for instances
  - User has Root (Administrator) access to instances

Instance creation:
Amazon EC2 Overview

- Enable users to launch and manage server instances
  - Users can control the instance properties through a simple web interface
- Each EC2 instance is like a physical machine:
  - CPU
  - Memory
  - Disk space
  - Network interface

Instance Types

- Computation baseline performance metric:
  - 1 EC2 Compute Unit (ECU) provides the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor
- Small Instance – default
  - 1.7 GB memory
  - 1 EC2 Compute Unit (1 virtual core with 1 ECU)
  - 160 GB instance storage
- Medium Instance
  - 3.75 GB memory
  - 2 EC2 Compute Unit (1 virtual core with 2 ECU)
  - 410 GB instance storage
Instance Types

- Large Instance
  - 7.5 GB memory
  - 4 EC2 Compute Units (2 virtual cores with 2 ECU each)
  - 850 GB instance storage

- Extra Large Instance
  - 15 GB memory
  - 8 EC2 Compute Units (4 virtual cores with 2 ECU each)
  - 1,690 GB instance storage

- Micro Instance
  - 613 MB memory
  - Up to 2 EC2 Compute Units (for short periodic bursts)

Simple Storage Service (S3)

- Store and retrieve any amount of data, at any time, from anywhere on the web
- Objects are stored in buckets (object containers)
- Resource are identified by URIs
  - [http://s3.amazonaws.com/CCEW/Presentation1.ppt](http://s3.amazonaws.com/CCEW/Presentation1.ppt)
  - CCEW is the bucket
  - Presentation1.ppt is the object
- Global naming scheme for buckets (names must be unique)
- Functionality: read, write, and delete data objects
- Basically, provides persistent file-level storage (S3 is accessed as a web service)
- Objects are replicated 3 times within different availability zones in the region where they are stored
- Has an access control mechanism to allow/deny access
Elastic Block Storage (EBS)

• EBS are raw unformatted persistent virtual disks for EC2 instances
  • Size: 1 GB to 1 TB
  • Each user can use at most 20 EBS
• An EBS virtual disk (volume) can be formatted to hold a specific file system
• An EBS can be attached to only one instance at a time
  • EBS can be used as boot partition for instances: fast startup time
• An instance can mount multiple EBSs
• Volumes are replicated for availability
• Snapshot feature
  • Users can create incremental snapshots to S3
    • Good for sharing & instantiating new volumes
    • Extra availability

Security (1)

• Physical security
  • Data centers are in undisclosed-location facilities
  • Critical facilities have military grade perimeter control
  • Video surveillance and state of the art intrusion detection
  • Authorized staff must pass two factor authentication to access
  • Every operation is recorded for later examination
• OS security
  • Host OS
    • Amazon admins have access for pre-defined time
    • They use strong SSH keys to gain non privileged access and then use privilege escalation
    • Everything is logged for later examination
  • Guest OS: Amazon has no access to guest OS
  • Users use firewall to restrict access from outside
Security (2)

- Virtual Machine
  - Xen Hypervisor improves security for guest OSs
  - User cannot see other local instances (not owned)
- Network Security
  - Amazon use SYN cookies and connection limiting to fight DDOS
  - Users can use SSL to prevent man in the middle attack
  - Amazon never allows IP spoofing
  - Packets inside Amazon network cannot be sniffed
  - No ARP poisoning is possible
- S3 and SimpleDB
  - There is access control for these services
  - Users should use encryption to store data
  - Users can use SSL connections to upload/download data

Amazon Web Services (AWS) - Overview

- AWS Overview
- Application example
- Using Amazon EC2 instances
Online photo processing service (1)

- Users submit Photos and specify operations they want over their photos:
  - Red eye detection, Cropping, Re-coloring, etc

- Requests are put on Request Queue and photos are saved in S3 storage area
- Photo Processing Server gets the request, retrieves photos from S3 storage, performs processing, and sends result using Response Queue

Online photo processing service (2)

- If instance fails, the photos and messages are not deleted
- New instance can be created to do same task
- In the mean time users can submit tasks
- New instance gets the unfinished tasks
Online photo processing service (3)

• If more than one instance is needed, user can initiate more instances
• Message queue can be used by multiple instances
• Message retrieved by one instance is locked, such that other instances cannot see it

Online photo processing service (4)

• Pipeline processing: if an operation takes more time, user can create another instance to handle the time consuming jobs
  ▪ Suppose server B performs the time consuming jobs
  ▪ Server A and B can communicate using SQS
Case study

• The New York Times used AWS to create PDF files of its whole archive
  ▪ 100 Amazon EC2 instances running Hadoop application
  ▪ Processed 4TB of raw TIFF image data (stored in S3) into 11 million finished PDFs
  ▪ Running time: 24 hours
  ▪ Cost: $240 (not including bandwidth)


Case study

• The company CycleComputing build a 50,000 core supercomputer called Naga, using the Amazon cloud infrastructure
  ▪ 6,742 instances
  ▪ 51,132 cores
  ▪ 58.78 TB memory
• Ran a chemistry computational intensive job for 3 hours at a cost of about $5,000 per hour
• They estimate this job used over 20 million dollars in infrastructure

Amazon Web Services (AWS) - Overview

• Overview
• Application example
• Using Amazon EC2 instances

We will see this on Thursday:
  ▪ How to create EC2 instances and how to manage them
  ▪ How to access data, security issues
  ▪ How to connect to EC2 instances
  ▪ How to access public data sets available on Amazon (e.g., US census data, Human Genome data, etc.)
  ▪ How much all of this costs