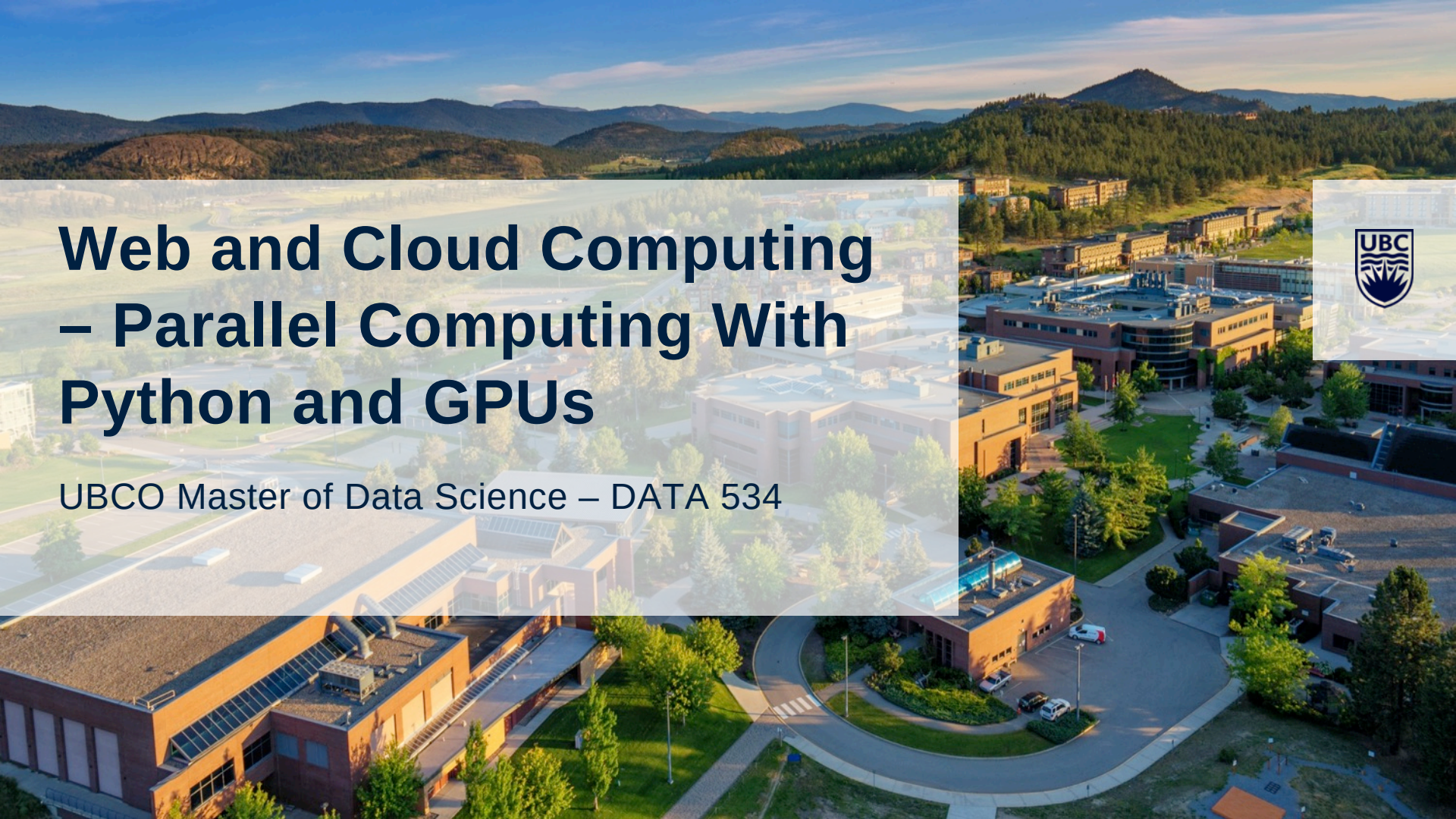


Web and Cloud Computing – Parallel Computing With Python and GPUs

UBCO Master of Data Science – DATA 534

What is Cloud Computing?

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) [Mell_2009], [Berkely_2009].

It can be *rapidly provisioned* and *released* with minimal management effort.

It provides *high level abstraction* of computation and storage model.

It has some essential **characteristics, service models**, and **deployment models**.

Essential Characteristics

On-Demand Self Service:

- A consumer can unilaterally provision computing capabilities, automatically without requiring human interaction with each service's provider.

Heterogeneous Access:

- Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous ***thin*** or ***thick*** client platforms.

Essential Characteristics (cont.)

Resource Pooling:

- The provider's computing resources are pooled to serve multiple consumers using a *multi-tenant model.*
- Different physical and virtual resources dynamically assigned and reassigned according to consumer demand.

Measured Service:

- Cloud systems *automatically control* and *optimize* resources used by leveraging a metering capability at some level of abstraction appropriate to the type of service.
- ***It will provide analyzable and predictable computing platform.***

Service Models

Cloud 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 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,…
- ***Examples: Caspio, Google Apps, Salesforce, Nivio, Learn.com.***

Service Models (cont.)

Cloud Platform as a Service (PaaS):

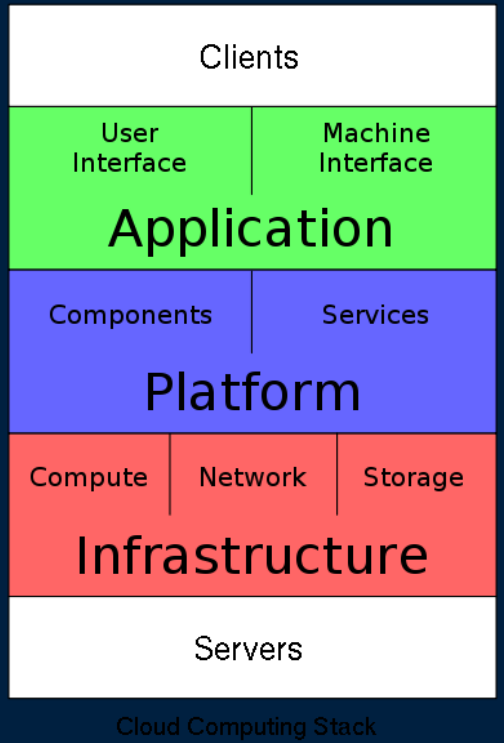
- The capability provided to the consumer is to deploy onto the cloud infrastructure *consumer-created* or *acquired applications* created using *programming languages and tools* supported by the provider.
- The consumer does not manage or control the underlying cloud infrastructure.
- Consumer has control over the deployed applications and possibly application hosting environment configurations.
- ***Examples: Windows Azure, Google App.***

Service Models (cont.)

Cloud Infrastructure as a Service (IaaS):

- The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources.
- The consumer is able to deploy and run arbitrary software, which can include operating systems and applications.
- The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).
- ***Examples: Amazon EC2, GoGrid, iland, Rackspace Cloud Servers, ReliaCloud.***

Service Models (cont.)



Cloud Computing Stack

Service Model at a glance: Picture From http://en.wikipedia.org/wiki/File:Cloud_Computing_Stack.svg

Deployment Models

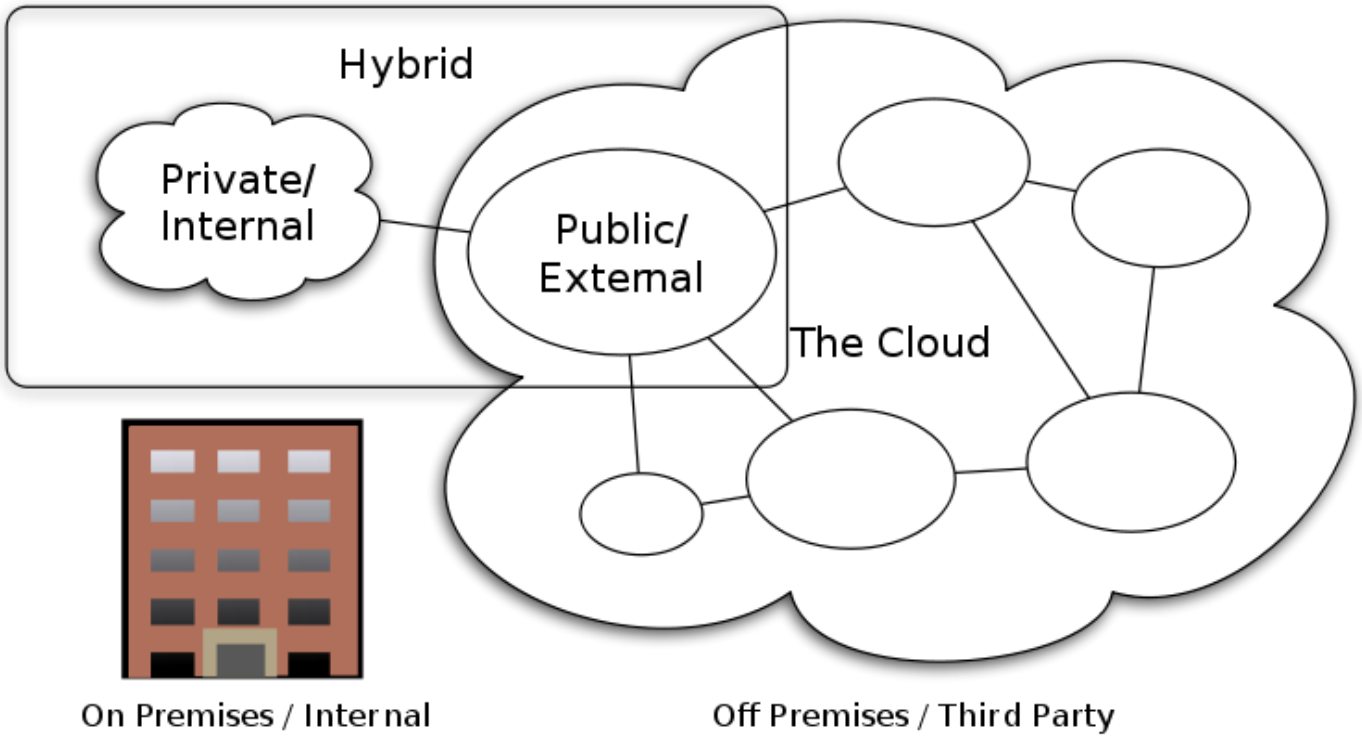
Deployment Models

- ***Private Cloud:***
  - The cloud is operated **solely** for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

- ***Community Cloud:***
  - The cloud infrastructure is shared by several organizations and supports a specific community that has **shared concerns**
  - It may be managed by the organizations or a third party and may exist on premise or off premise

- ***Public Cloud:***
  - The cloud infrastructure is made available to the general public or a large industry group and it is owned by an organization selling cloud services.

- ***Hybrid cloud:***
  - The cloud infrastructure is a composition of two or more clouds (private, community, or public).



Cloud Computing Types

CC-BY-SA 3.0 by Sam Johnston

Advantages of Cloud Computing

- Cloud computing do not need high quality equipment for user, and it is very easy to use.
- Provides dependable and secure data storage center.
- Reduce run time and response time.
- Cloud is a large resource pool that you can buy on-demand service.
- Scale of cloud can extend dynamically providing nearly infinite possibility for users to use internet.

amazon web services™

EC2

What is Infrastructure as a Service ?

A category of cloud services which provides capability to provision processing, storage, intra-cloud network connectivity services, and other fundamental computing resources of the cloud infrastructure.

Source- [ITU –Cloud Focus Group]

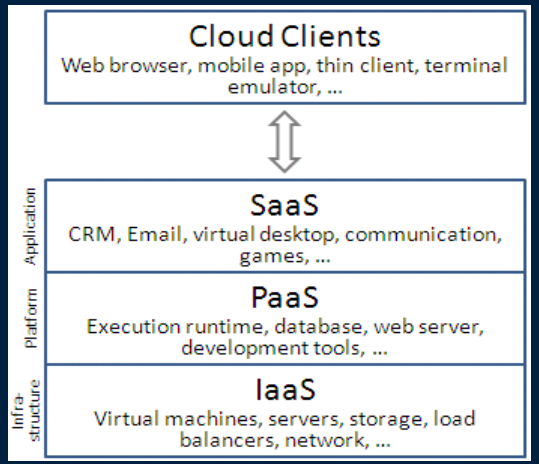


Diagram Source: Wikipedia

Highlights of IaaS

On demand computing resources

- Eliminate the need of far ahead planning

No up-front commitment

- Start small and grow as required
- No contract, Only credit card!

Pay for what you use

No maintenance

Measured service

Scalability

Reliability

What is EC2 ?

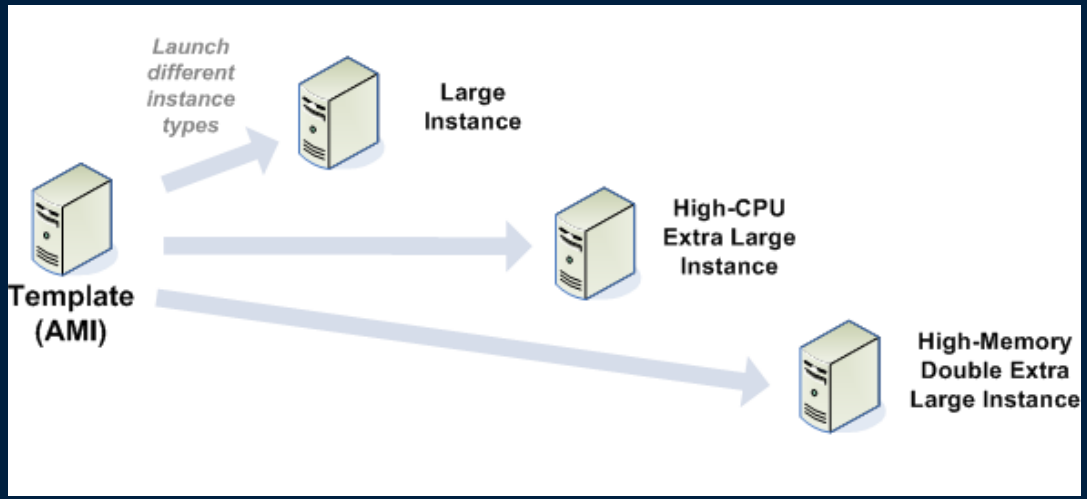
- Amazon Elastic Compute Cloud (EC2) is a web service that provides resizeable computing capacity that one uses to build and host different software systems.
- Designed to make web-scale computing easier for developers.
- A user can create, launch, and terminate server instances as needed, paying by the hour for active servers, hence the term "elastic".
  - Provides scalable, pay as-you-go compute capacity
  - Elastic - scales in both direction

EC2 Concepts

- AMI & Instance
- Region & Zones
- Storage
- Networking and Security
- Monitoring
- Auto Scaling
- Load Balancer

Amazon Machine Images (AMI)

- Is an immutable representation of a set of disks that contain an operating system, user applications and/or data.
- From an AMI, one can launch multiple instances, which are running copies of the AMI.



AMI and Instance

Amazon Machine Image (AMI) is a template for software configuration (Operating System, Application Server, and Applications)

Instance is a AMI running on virtual servers in the cloud

Each *instance type* offers different compute and memory facilities

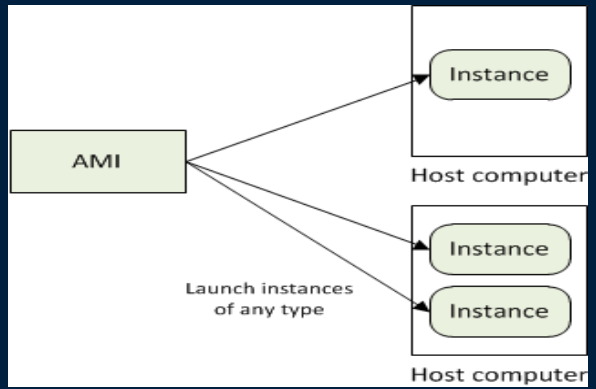


Diagram Source: http://docs.aws.amazon.com

Type	CPU	Memory	Local Storage	Platform	I/O	Name
Small	1 EC2 Compute Unit (1 virtual core with 1 EC2 Compute Unit)	1.7 GB	160 GB instance storage (150 GB plus 10 GB root partition)	32-bit	Moderate	m1.small
Large	4 EC2 Compute Units (2 virtual cores with 2 EC2 Compute Units each)	7.5 GB	850 GB instance storage (2 x 420 GB plus 10 GB root partition)	64-bit	High	m1.large
Extra Large	8 EC2 Compute Units (4 virtual cores with 2 EC2 Compute Units each)	15 GB	1690 GB instance storage (4 x 420 GB plus 10 GB root partition)	64-bit	High	m1.xlarge
Micro	Up to 2 EC2 Compute Units (for short periodic bursts)	613 MB	None (use Amazon EBS volumes for storage)	32-bit or 64-bit	Low	t1.micro
High-CPU Medium	5 EC2 Compute Units (2 virtual cores with 2.5 EC2 Compute Units each)	1.7 GB	350 GB instance storage (340 GB plus 10 GB root partition)	32-bit	Moderate	c1.medium

Region and Zones

Amazon have data centers in different region across the globe

An instance can be launched in different regions depending on the need.

- Closer to specific customer
- To meet legal or other requirements

Each region has set of zones

- Zones are isolated from failure in other zones
- Inexpensive, low latency connectivity between zones in same region

Storage

Amazon EC2 provides three type of storage option

- Amazon EBS
- Amazon S3
- Instance Storage

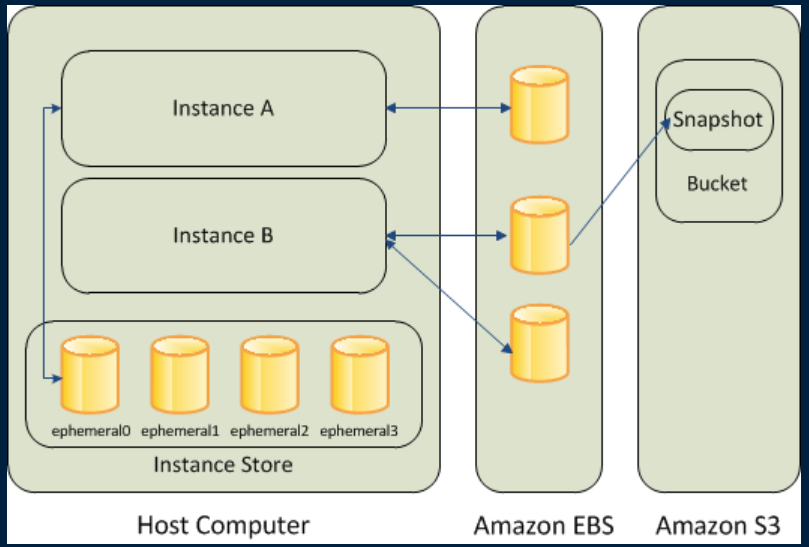
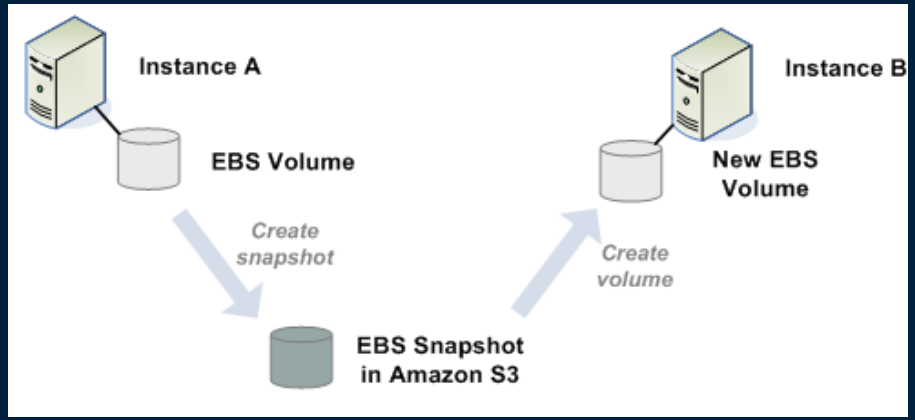


Diagram Source: http://docs.aws.amazon.com

Elastic Block Store(EBS) volume

- An EBS volume is a read/write disk that can be created by an AMI and mounted by an instance.
- Volumes are suited for applications that require a database, a file system, or access to raw block-level storage.



Amazon S3

- S3 = Simple storage Service
- A SOA – Service Oriented Architecture which provides online storage using web services.
- Allows read, write and delete permissions on objects.
- Uses REST and SOAP protocols for messaging.

Amazon SimpleDB

- Amazon SimpleDB is a highly available, flexible, and scalable non-relational data store that offloads the work of database administration.
- Creates and manages multiple geographically distributed replicas of your data automatically to enable high availability and data durability.
- The service charges you only for the resources actually consumed in storing your data and serving your requests.

https://aws.amazon.com/simpledb/

Networking and Security

Instances can be launched on one of the two platforms

- EC2-Classic
- EC2-VPC

- Each instance launched is assigned two addresses a private address and a public IP address.
  - A replacement instance has a different public IP address.

Instance IP address is dynamic.

- new IP address is assigned every time instance is launched

- Amazon EC2 offers Elastic IP addresses (static IP addresses) for dynamic cloud computing.
  - Remap the Elastic IP to new instance to mask failure
  - Separate pool for EC2-Classic and VPC

Security Groups to access control to instance

Monitoring, Auto Scaling, and Load Balancing

Monitor statistics of instances and EBS

- CloudWatch

Automatically scales amazon EC2 capacity up and down based on rules

- Add and remove compute resource based on demand
- Suitable for businesses experiencing variability in usage

Distribute incoming traffic across multiple instances

- Elastic Load Balancing

How to access EC2

AWS Console

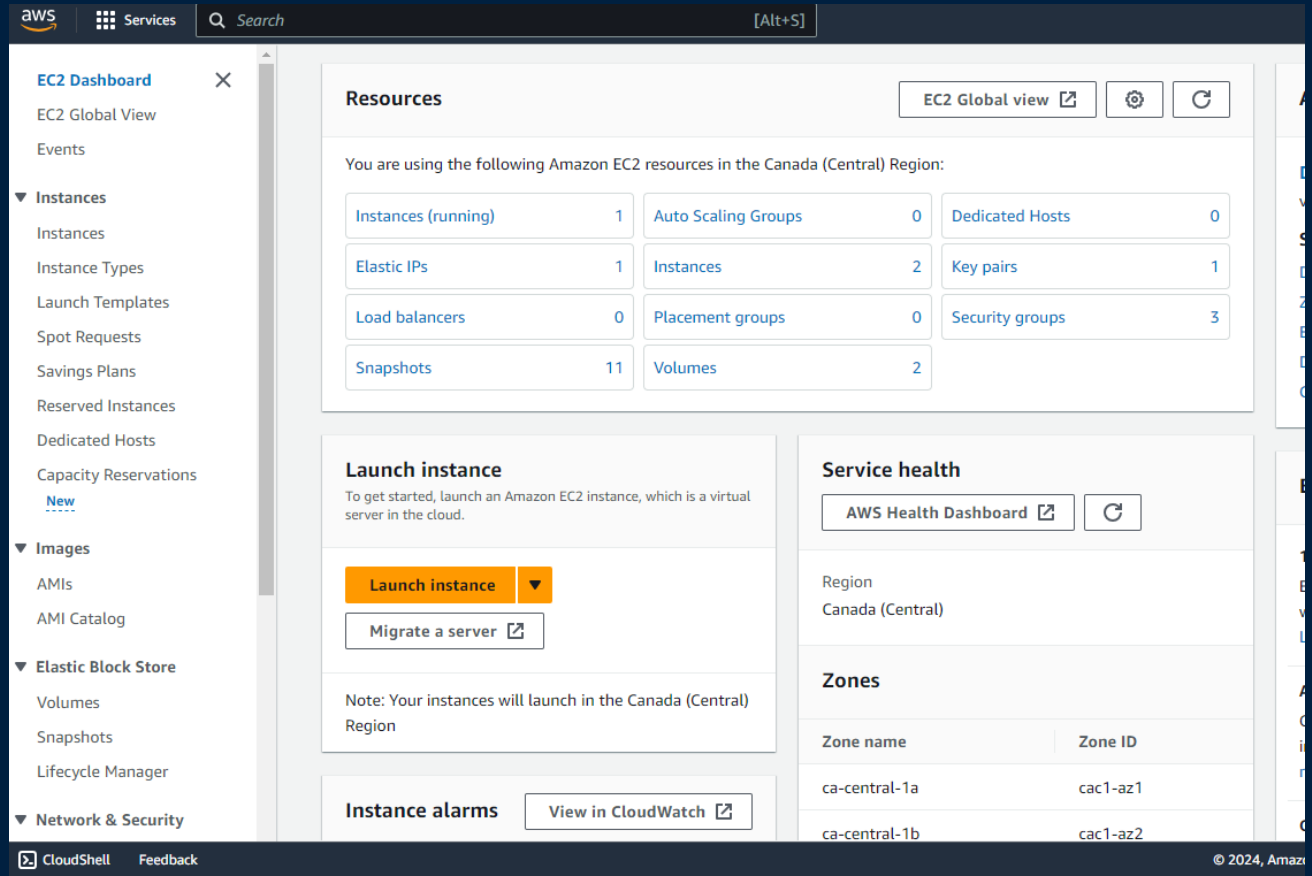
- http://console.aws.amazon.com

Command Line Tools

Programmatic Interface

- EC2 APIs
- AWS SDK

AWS Management Console



Region: US East (Virginia)

	Linux/UNIX Usage	Windows Usage
Standard On-Demand Instances		
Small (Default)	$0.085 per hour	$0.12 per hour
Large	$0.34 per hour	$0.48 per hour
Extra Large	$0.68 per hour	$0.96 per hour
Micro On-Demand Instances		
Micro	$0.02 per hour	$0.03 per hour
Hi-Memory On-Demand Instances		
Extra Large	$0.50 per hour	$0.62 per hour
Double Extra Large	$1.00 per hour	$1.24 per hour
Quadruple Extra Large	$2.00 per hour	$2.48 per hour
Hi-CPU On-Demand Instances		
Medium	$0.17 per hour	$0.29 per hour
Extra Large	$0.68 per hour	$1.16 per hour
Cluster Compute Instances		
Quadruple Extra Large	$1.60 per hour	N/A*
Cluster GPU Instances		
Quadruple Extra Large	$2.10 per hour	N/A*
* Windows® is not currently available for Cluster Compute or Cluster GPU Instances		

Pricing is per instance-hour consumed for each instance, from the time an instance is launched until it is terminated. Each partial instance-hour consumed will be billed as a full hour.

References

- Mobile cloud computing: Big Picture by M. Reza Rahimi
- http://aws.amazon.com/ec2, http://docs.aws.amazon.com
- Amazon Elastic Compute Cloud – User Guide, API Version 2011-02-28.

Above the Clouds: A Berkeley View of Cloud Computing - Michael Armbrust et.al 2009

International telecommunication union – Focus Group Cloud Technical Report


THE UNIVERSITY OF BRITISH COLUMBIA