

# Infrastructure as a Service (IaaS)

## Week 1

School of Computer Science

Faculty of Engineering and Information Technology

University of Technology Sydney

# Learning Objectives

- Understand Computing Infrastructure in Enterprises
- Understand and define Infrastructure as a Service (IaaS)
- Understand Generic Architecture of IaaS Service Provisioning
- Brief introduction to Amazon Web Services

# Computing Infrastructure in Enterprises

- Web Servers
  - Load balancers between Web Servers ... etc.
- Application servers
  - Load balancers between application servers .... etc.
- Database Servers
  - Load Balancers between Database Servers
  - Query Optimizers ...etc.
- Memory
  - RAM (SD-, RD-, DDR-, DDR2-)
  - Non-Volatile Memory
- Variations and enhancements in processing speed, capacity, latency, and cost over the years .... etc



# Computing Infrastructure in Enterprises

- Enterprise applications have specific hardware requirements
  - Different applications have different hardware requirements (one-size does not fit all)
    - Large scale data mining application vs. Microsoft office applications
- Implications of procuring, maintaining and de-registering of IT Infrastructure
  - Budgeting for equipment;
  - Personnel hiring and management;
  - Real-estate
- Would make (business) sense to rent these (depending on the specific application requirements), and use them.
- Cloud computing provides a solution to the above issues in the form of Infrastructure-as-a-Service (IaaS)

# Infrastructure-as-a-Service

- The NIST Definition of 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, which can include operating systems and applications.” (Mell and Grance 2011)



# Understanding the IaaS definition

- Key elements of the IaaS definition
  - Provision (processing, storage and other fundamental computing resources) on demand;
  - Cloud Consumer is able to deploy and run arbitrary software; (operating system or application)
  - The cloud consumer does not manage the underlying cloud infrastructure
- Implications of IaaS:
  - No up-front capital investment for the cloud consumer;
  - No administrative or maintenance expenses for the cloud consumer;
  - No real-estate expenses for the cloud consumer;
  - Cloud Consumer is charged for resources (Pay-as-you-go)
- Simplest (or basic) flavour of cloud offering

# Generic Architecture of IaaS Service Provisioning Process

- IaaS provider has to setup the IaaS stack
  - IaaS Service Provisioning Process is carried out by the IaaS Stack
- “Layers” of the IaaS Stack are
  - Co-location Layer
  - Hardware Layer
  - Virtualization Layer
  - Service Layer



# Pictorial Representation of IaaS Service Provisioning

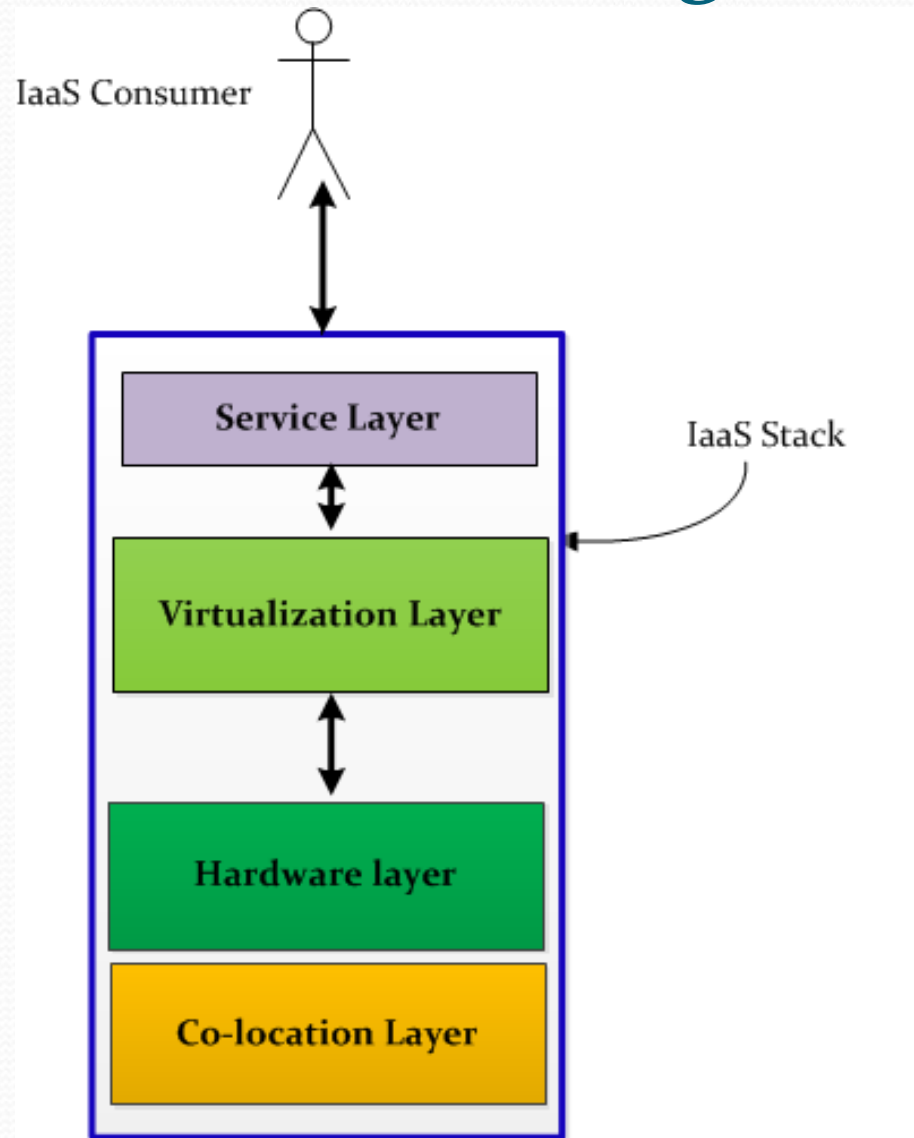


Figure: IaaS Stack



# Co-location Layer

- Provide the basic requirements needed to deliver IaaS services
  - Electrical Power
    - Redundant electric power supply lines
    - Power supply lines from multiple electricity suppliers
    - Backup power supply
  - Bandwidth
    - Redundant network connectivity
    - Fault-tolerant network connectivity (multiple “physical” links from different network providers)
  - Cooling (efficient design, major cost-constraint)
  - Real Estate (choice of location, cheap yet reliable, “safe” from natural disasters, political stability ...etc)

# Hardware Layer

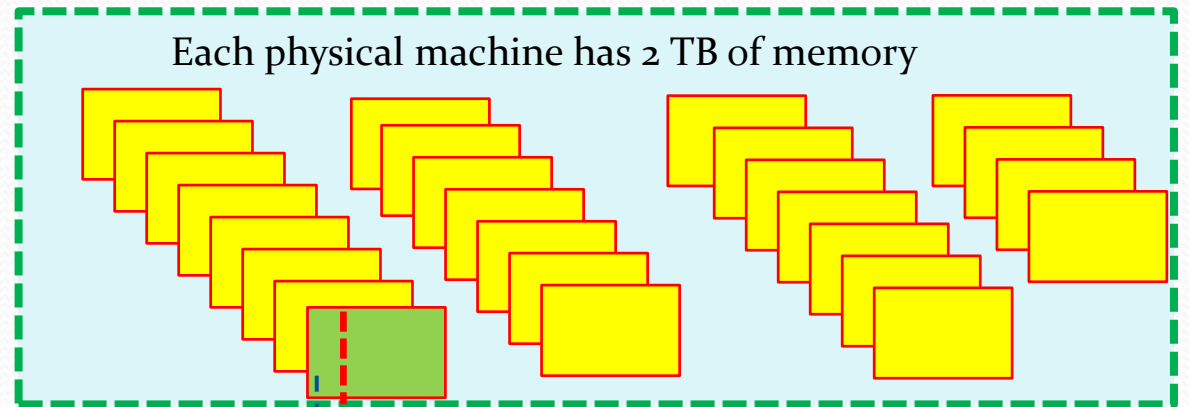
- Comprises of the physical hardware that will be provisioned “*virtually*” to the cloud consumers (data centres)
- Actual or physical resource present in the data centre
- Maximize the hardware capacity to give the cloud consumers the illusion of infinite capacity.



# Virtualization Layer

- Key layer to enable IaaS Service Provisioning

Data centre

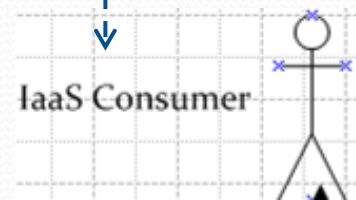


The virtualization layer is responsible for automatically:

- Finding an appropriate physical machine that can host the consumer's request;
- Slicing it to meet customer's requirements;
- Allocating the sliced part (virtual view) to the consumer automatically
- Maintaining a snapshot of resource usage (free/being used etc..)



50 GB of memory for  
2 hours



# Virtualization Layer

- The virtualization layer is responsible for automatically:
  - Finding an appropriate physical machine that can host the consumer's request;
  - Slicing it to meet customer's requirements;
  - Allocating the sliced part (virtual view) to the consumer automatically
  - Maintaining a snapshot of resource usage (free/being used etc..)



# Service Layer

Define the interface to the IaaS Consumer to interact with the provisioned resources (which have been virtualized)

The screenshot displays the AWS Management Console for the Amazon EC2 service. The browser address bar shows the URL `https://console.aws.amazon.com/ec2/home?region=us-east-1`. The navigation sidebar on the left lists various AWS services, with the 'EC2 Dashboard' selected. The main content area features a 'Getting Started' section with a 'Launch Instance' button and a note about the region. Below this is the 'Service Health' section, which includes a table showing the current status and details of the Amazon EC2 service and its availability zones.

**Region:** US East (Virginia)

**EC2 Dashboard**

Scheduled Events

INSTANCES

- Instances
- Spot Requests
- Reserved Instances

IMAGES

- AMIs
- Bundle Tasks

ELASTIC BLOCK STORE

- Volumes
- Snapshots

NETWORK & SECURITY

- Security Groups
- Elastic IPs
- Placement Groups
- Load Balancers
- Key Pairs
- Network Interfaces

**Getting Started**

To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

[Launch Instance](#)

Note: Your instances will launch in the US East (Virginia) region.

**Service Health**

Service Status

Current Status	Details
Amazon EC2 (US East - N. Virginia)	Service is operating normally

[View complete service health details](#)

Availability Zone Status

Current Status	Details
us-east-1a	Availability zone is operating normally
us-east-1b	Availability zone is operating normally
us-east-1c	Availability zone is operating normally
us-east-1d	Availability zone is operating normally

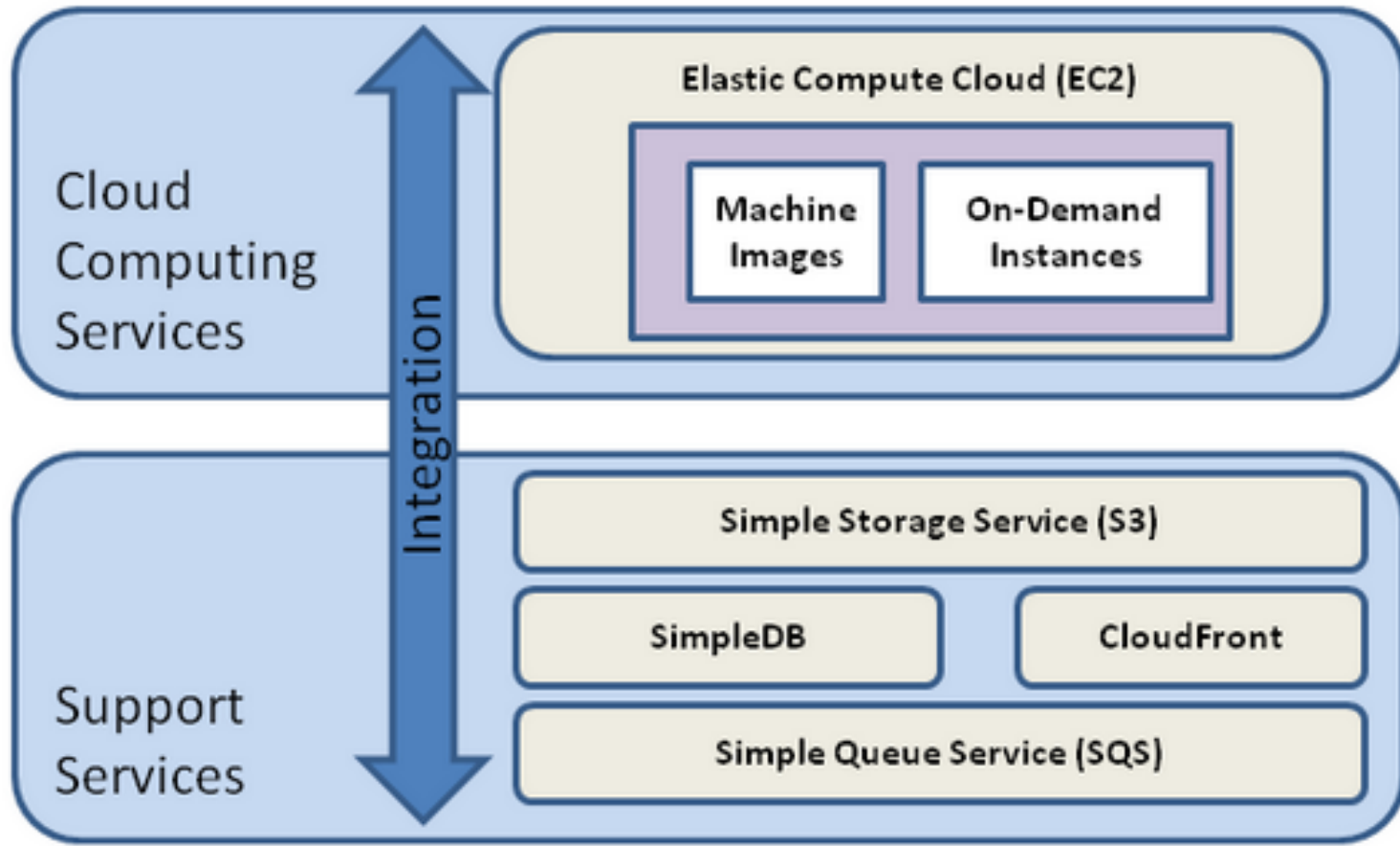
# Amazon Web Services (AWS)

- Amazon Web Services (AWS) is a collection of computing services offered over the Internet by Amazon.com
- Amazon.com is one of the heavily trafficked websites globally
- Houses and provides a wide range of computing infrastructure (compute, storage etc...); and provisions them “as-a-service” to the end-users on-demand
- AWS value proposition - *“You pay for what you use”*
- Amazon Infrastructure can be provisioned for any type of application usage
- Increasingly popular platform for website hosting and deployment



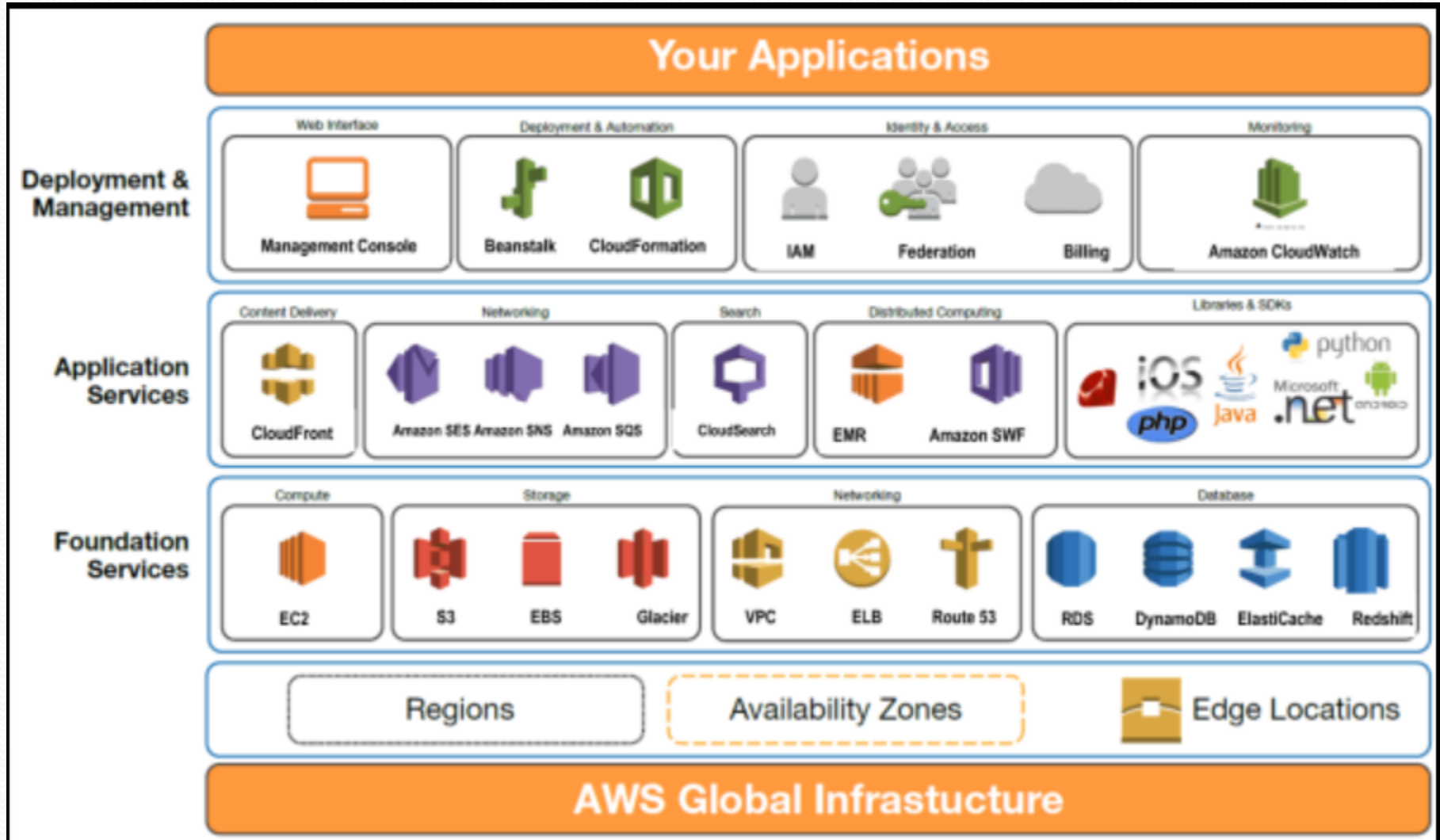


# Architecture of Amazon Web Services



**Figure:** Architecture of Amazon Web Service (AWS)

# AWS service offerings

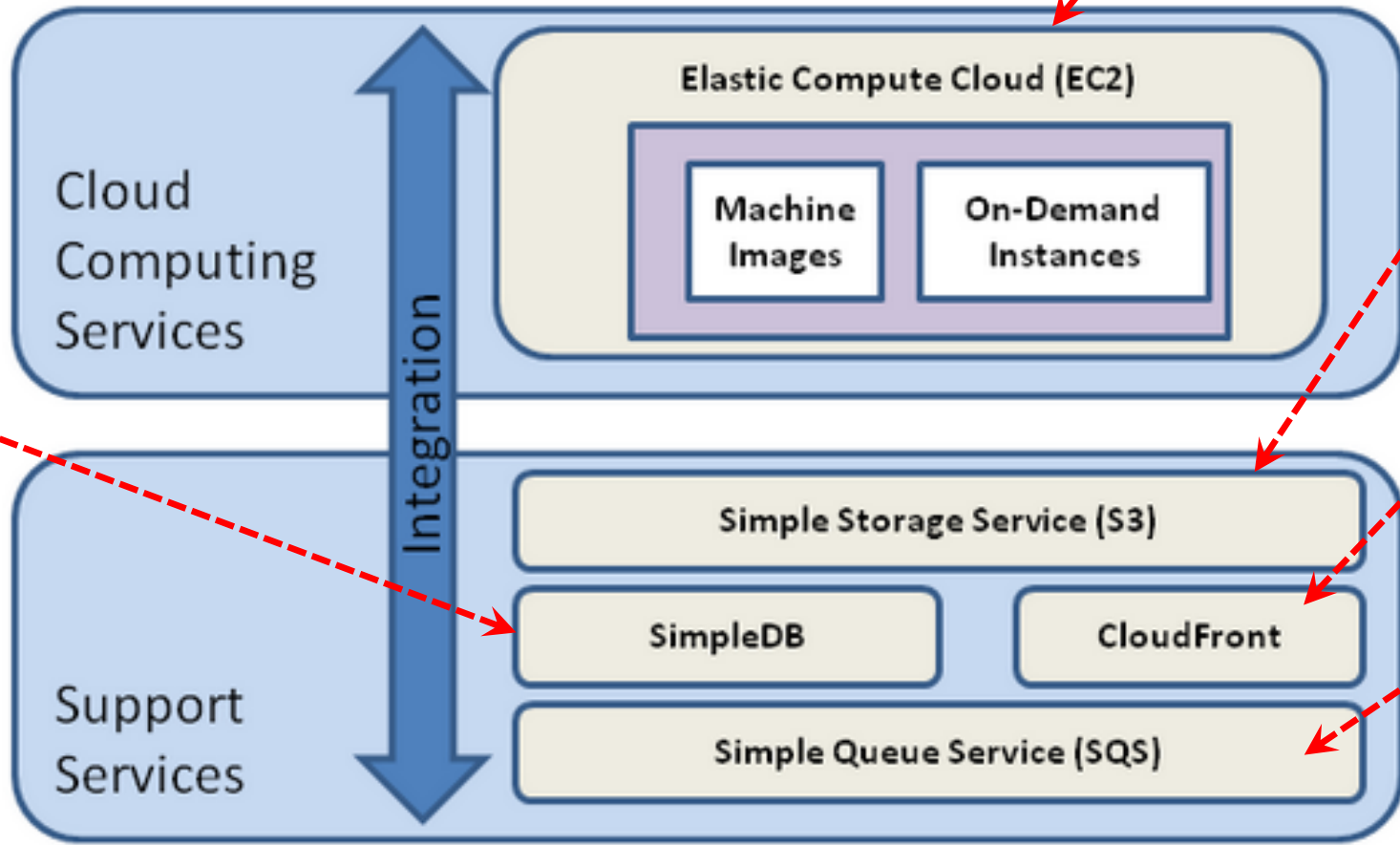




# Components of AWS

- Amazon Elastic Compute Cloud (EC2) Instances
- Amazon Simple Storage Service (S3)
- Amazon Relational Database Service
- Amazon Cloud Front Service
- Amazon Simple Queue Service (SQS)
- Amazon Glacier

# Architecture of Amazon Web Service



**Figure:** Architecture of Amazon Web Service (AWS)



# AWS EC2 instances (on demand)



- Provide virtual machines to cloud consumers
- (As of 2020) EC2 instances are available in six different types
  - General;
  - Compute-optimized (lowest price per compute);
  - Memory-optimized (lowest price per GiB of RAM);
  - GPU-optimized (optimized for graphical and GPU-applications)
  - Storage-optimized (optimized for high I/O)
- Choose as per your application requirements
- Combine instances from different types



# Different types of EC2 pricing models



- Comes with three pricing models
  - On-Demand Instance - Pay by the hour with no long term commitments;
  - Reserved Instance - Very cheap compute power (compared to the on-demand instances);
    - Up-front payment (All up-front, no up-front, partial up-front);
    - Fixed Date and time for resource usage
  - Spot Instances – Bid and use spare EC2 instances.

# AWS EC2 instances



AWS Management Console > Amazon EC2

Navigation

Region: Asia Pacific (Singapore)

EC2 Dashboard

Scheduled Events

INSTANCES

- Instances
- Spot Requests
- Reserved Instances

IMAGES

- AMIs
- Bundle Tasks

ELASTIC BLOCK STORE

- Volumes
- Snapshots

NETWORK & SECURITY

- Security Groups
- Elastic IPs
- Placement Groups
- Load Balancers
- Key Pairs

Amazon EC2 Console Dashboard

Getting Started

To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

[Launch Instance](#)

Note: Your instances will launch in the Asia Pacific (Singapore) region.

Service Health

Service Status

Current Status	Details
Amazon EC2 (APAC - Singapore)	Service is operating normally

[View complete service health details](#)

Availability Zone Status

Current Status	Details
ap-southeast-1a	Availability zone is operating normally

Figure: Launching an Amazon EC2 instance

Request Instances Wizard

CHOOSE AN AMI INSTANCE DETAILS CREATE KEY PAIR CONFIGURE FIREWALL REVIEW

Choose an Amazon Machine Image (AMI) from one of the tabbed lists below by clicking its **Select** button.

Quick Start My AMIs Community AMIs

**Basic Amazon Linux AMI 2011.09**  
Amazon Linux AMI 2011.09, EBS boot with Amazon EC2 AMI Tools.  
Root Device Size: 8 GB ☒ 64 bit ☐ 32 bit [Select](#)

**Red Hat Enterprise Linux 6.2**  
Red Hat Enterprise Linux version 6.2, EBS-boot.  
Root Device Size: 6 GB ☒ 64 bit ☐ 32 bit [Select](#)

**SUSE Linux Enterprise Server 11**  
SUSE Linux Enterprise Server 11 Service Pack 1 basic install, EBS boot with Amazon EC2 AMI Tools preinstalled; Apache 2.2, MySQL 5.0, PHP 5.3, Ruby 1.8.7, and Rails 2.3.  
Root Device Size: 10 GB ☒ 64 bit ☐ 32 bit [Select](#)

**Ubuntu Server Cloud Guest 11.10 (Oneiric Ocelot)**  
Ubuntu Server version 11.10 (Oneiric Ocelot) optimized for use on AWS. Commercial support available at <http://www.canonical.com/enterprise-services/ubuntu-advantage/cloud>  
Root Device Size: 8 GB ☒ 64 bit ☐ 32 bit [Select](#)

**Ubuntu Server Cloud Guest 10.04 LTS (Lucid Lynx)**  
Ubuntu Server version 10.04 (Lucid Lynx) optimized for use on AWS. Commercial support available at <http://www.canonical.com/enterprise-services/ubuntu-advantage/cloud>  
Root Device Size: 8 GB ☒ 64 bit ☐ 32 bit [Select](#)

Figure: Select an appropriate EC2 instance





EC2

# Importing VM images to Amazon EC2

- VM Import
  - Provides the ability to import virtual machine images to Amazon EC2 instances
  - This feature comes pre-packaged with Amazon EC2 instances

# Persistence mechanisms in AWS

- AWS provides the following (web) services for data storage:
  - Amazon Relational Database Service (Amazon RDS)
  - Amazon Simple Storage Service (Amazon S3)
  - Amazon Glacier
- For different purposes and audience



# Amazon Relational Database Service (Amazon RDS)



- Amazon RDS provides “relational database-as-a-service” to consumers
  - Set up and access your relational database in the cloud
- Users can select from MySQL, Oracle, or Microsoft SQLServer, Postgre SQL, Amazon Aurora, and Maria DB
  - Database administration and maintenance activities performed by Amazon
- Carries out regular backup and “point-in-time” (time stamped) recovery
- Provides the added advantage of scale-up, scale down of RDS database instances on demand

# Launching a RDS instance

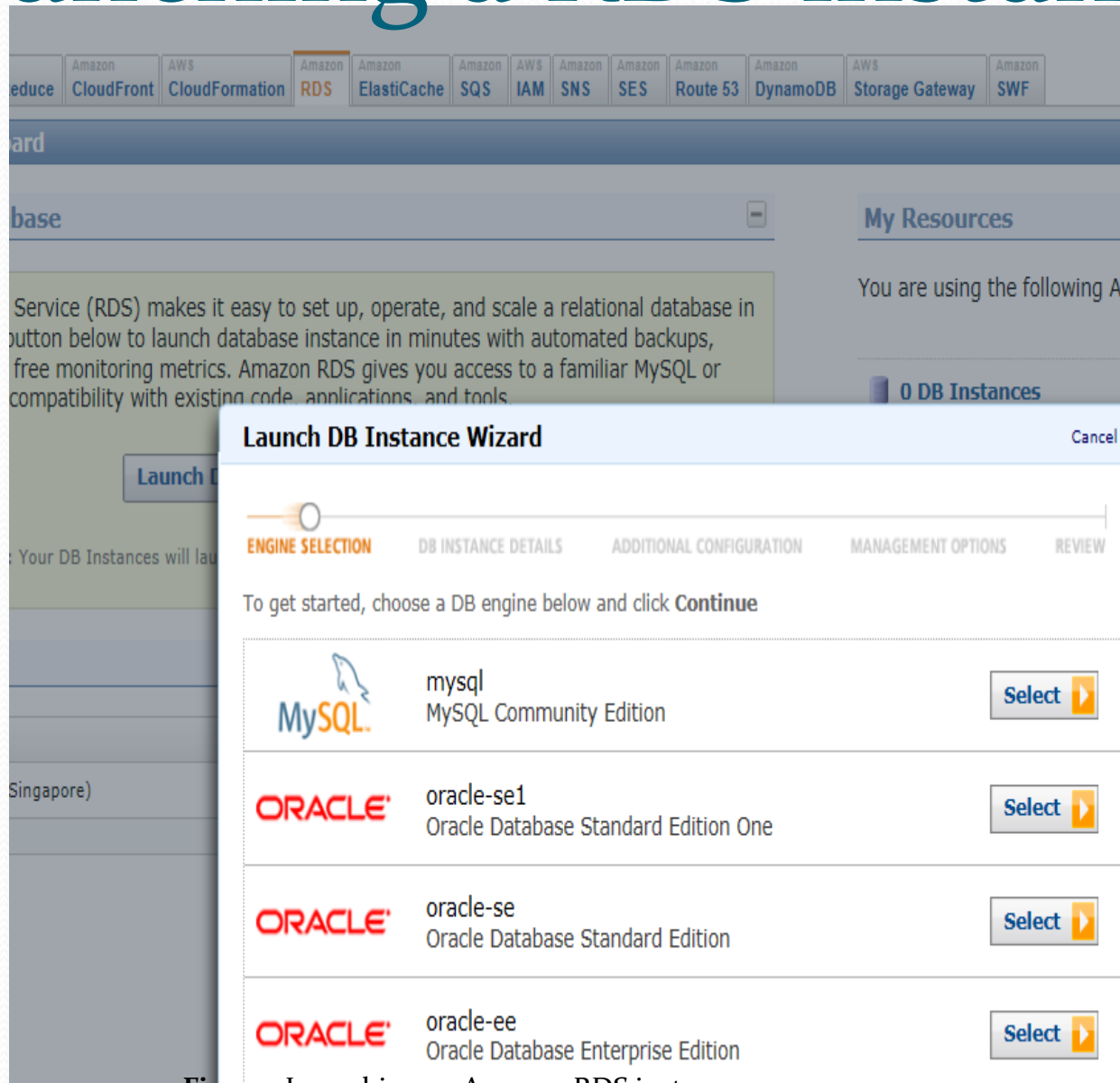


Figure: Launching an Amazon RDS instance



# Amazon Simple Storage Service (Amazon S3)



- Amazon S3 is an Online storage service from Amazon
- Very useful for storage of non-structured information (information without an underlying data model) such as images, videos..etc.
- Objects are stored in “buckets”, associated with an Amazon User.
  - Data is stored as “objects”, each of which is assigned unique key for referencing it
  - Size of objects (1 byte to 5 terabytes)

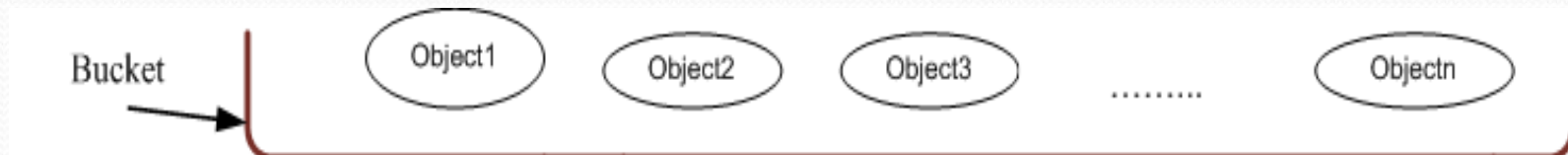




# Amazon Simple Storage Service (Amazon S3)



- Very popular web service from Amazon
- Has been used by many providers to provision innovative services (Ex: Dropbox)



**Figure:** Bucket and objects in Amazon S3

# Amazon Glacier

- Relatively new cloud-based storage service offered by Amazon
  - Useful for data archiving and backup
- Should be used for that is infrequently accessed and for which retrieval times of several hours are suitable.
- Very cheap to store data with Amazon Glacier (\$0.005 per gigabyte per month for Asia-Pacific (Sydney))
- Typical use-cases where it may be used:
  - Archiving media records or assets;
  - Very old transaction records



# Amazon CloudWatch

- Monitor the performance of AWS resources
- Available in two flavours (basic and advanced) – vary in terms of frequency and “type” of monitoring provided to the cloud consumer
- Define your own monitoring metrics and define alarm for these metrics
- Present the “health” of AWS resource to the consumer (response time, latency...etc)

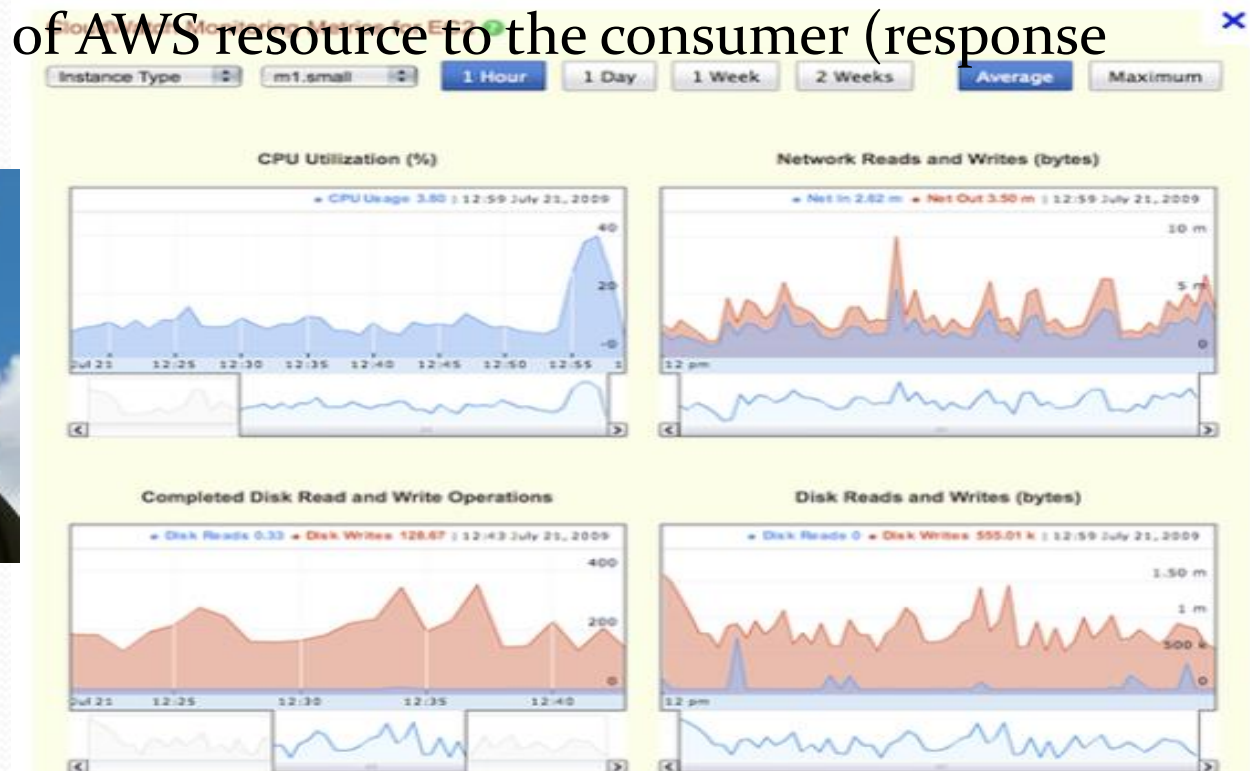


Figure: Amazon CloudWatch Monitoring

# Example of AWS metering and billing

- Fine-grained resource usage
- Pay only for what you use!!

**This Month's Activity as of December 25, 2010** ✕

The billing cycle for this report is December 1 - December 31, 2010. The AWS service usage charges on this page currently show activity through approximately 12/25/2010 16:59 GMT.

<a href="#">Expand All Services</a>   <a href="#">Collapse All Services</a> <a href="#">Printer Friendly Version</a>		<a href="#">Printer Friendly Version</a>
		Totals
<b>Amazon CloudFront</b> <a href="#">View/Edit Service</a>	<a href="#">Download Usage Report</a>	0.68
<b>Amazon Simple Storage Service</b> <a href="#">View/Edit Service</a>	<a href="#">Download Usage Report</a>	0.17
<b>AWS Data Transfer (excluding Amazon CloudFront)</b> <a href="#">View/Edit Service</a>		0.01
<b>Taxes</b> <a href="#">Estimated Taxes</a> (Due January 1, 2011)		0.00
<b>Total Charges due on January 1, 2011†</b>		<b>\$0.86</b>

Figure: Monthly activity statement from Amazon Web Services



# Additional References

1. M. Behrendt, B. Glasner, P. Kopp, R. Dieckmann, G., Breiter, S. Pappe, H. Kreger and A. Arsanjani. (2011). IBM Cloud Computing Reference Architecture 2.0.
2. P. Mell and T. Grance. (2011), The NIST Definition of Cloud Computing.
3. Tavis, M. (2010). Web Application Hosting in the AWS Cloud: Best Practices. Amazon Web Services
4. <http://aws.amazon.com/>