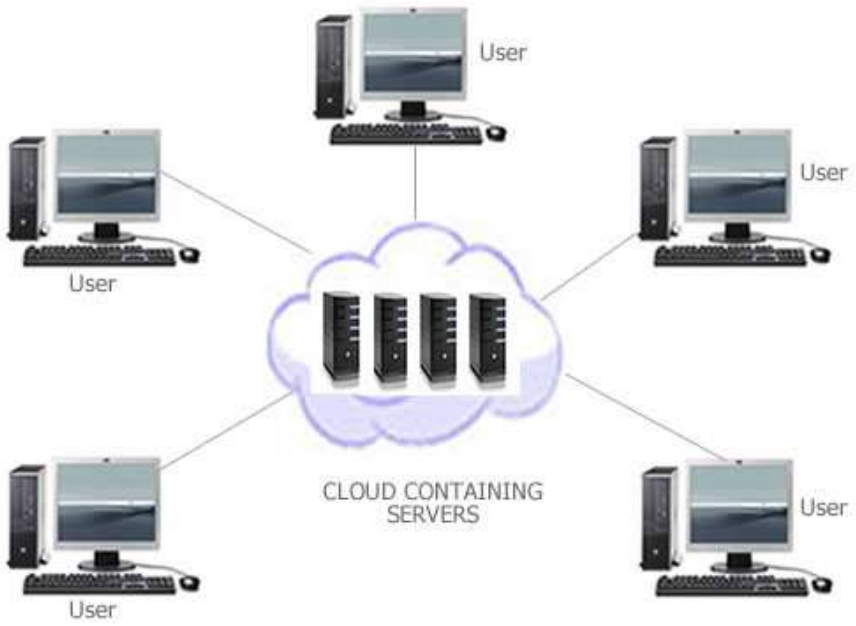
Distributed Systems (COMP9243)

Lecture 10a: Cloud Computing

① What is Cloud Computing?
② X as a Service
③ Key Challenges
④ Developing for the Cloud

What is Cloud Computing?

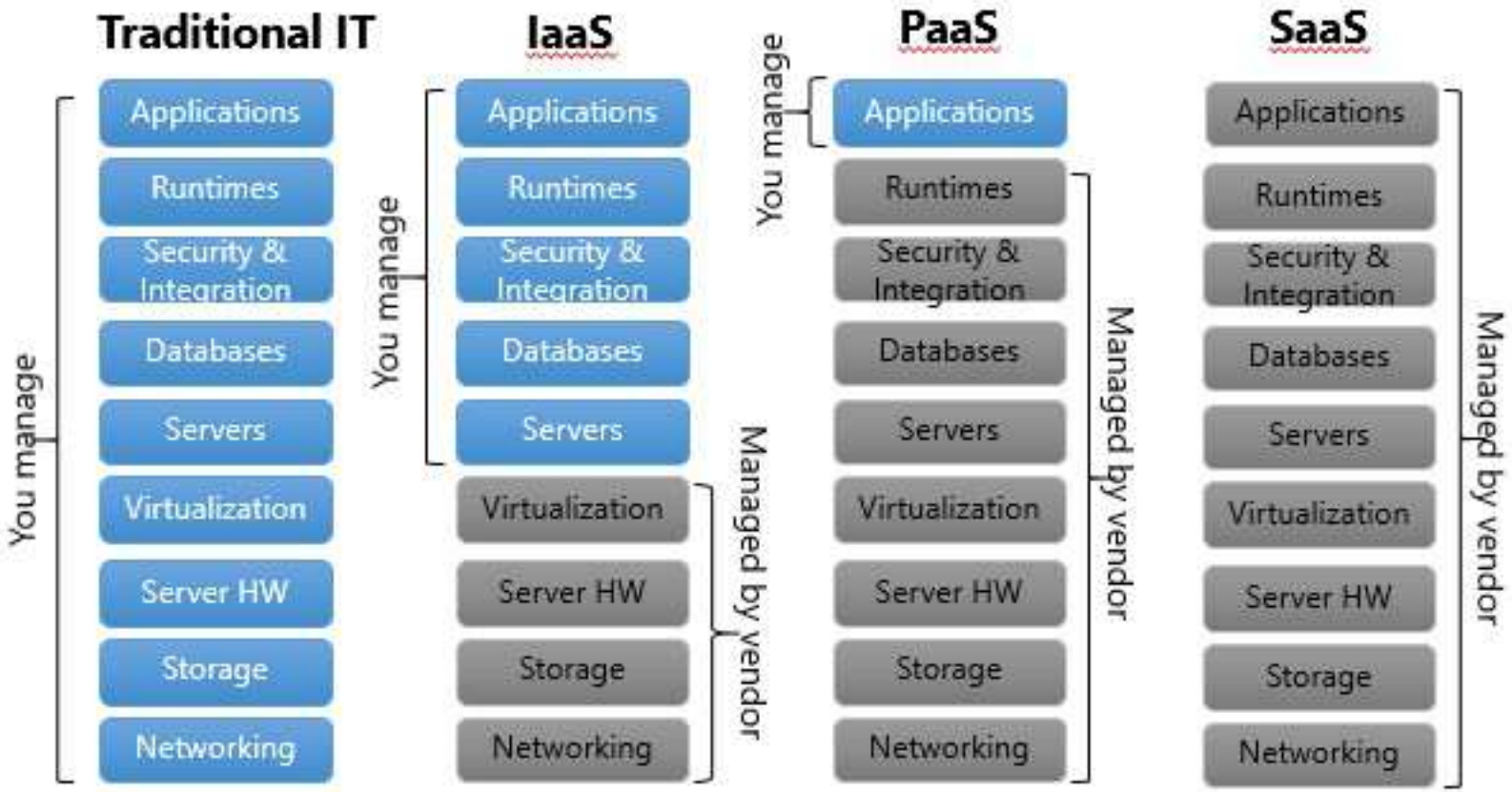
A style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet. (Wikipedia)



Why is it called *Cloud*?

- ➜ services provided on virtualised resources
- ➜ virtual machines spawned on demand
- ➜ location of services no longer certain
- ➜ similar to *network cloud*

Flavours of Cloud Computing:



http://www.mazikglobal.com/blog/cloud-computing-stack-saas-paas-iaas/

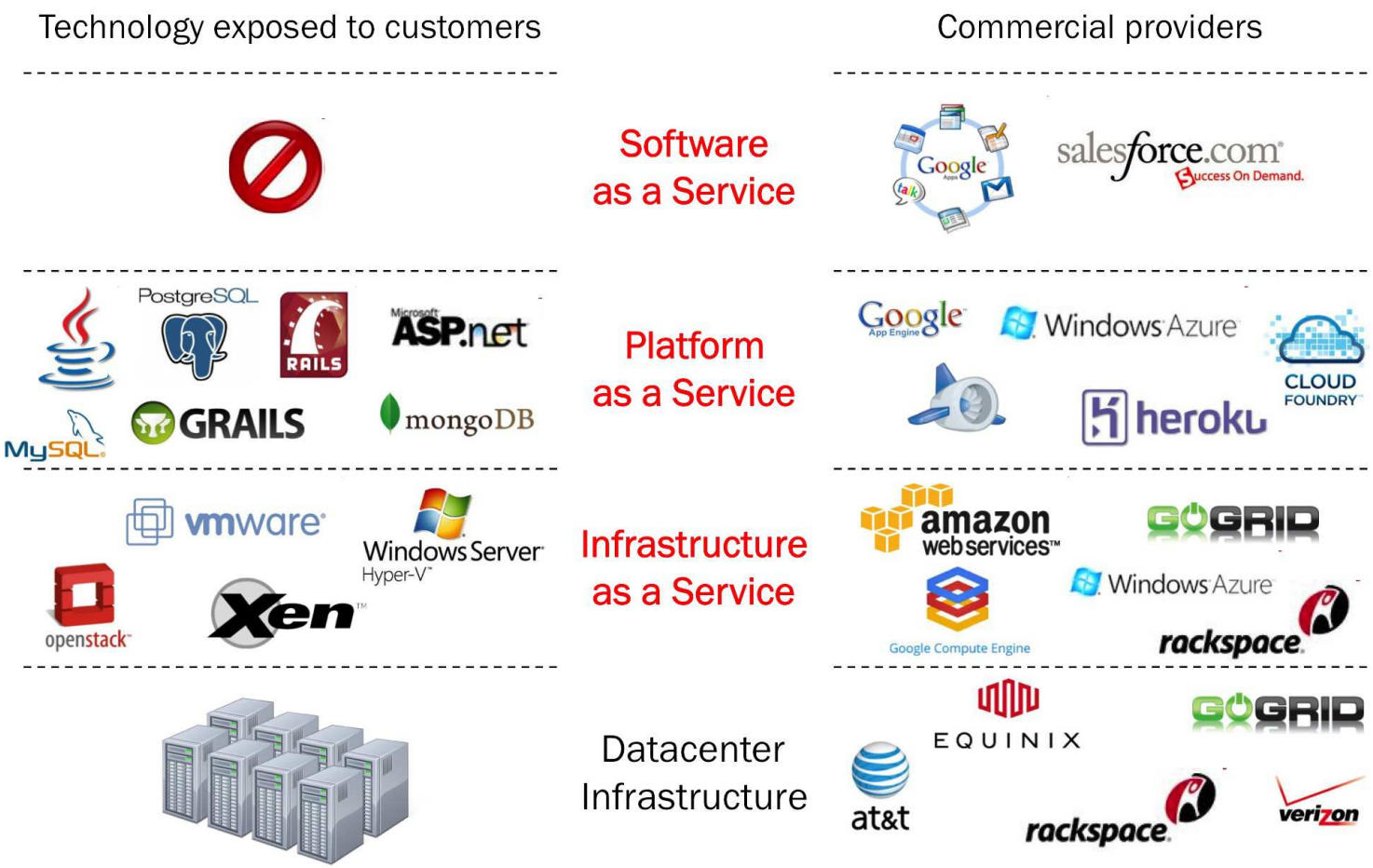


Figure from Hiroshi Wada

KEY CHARACTERISTICS OF CLOUD COMPUTING

SP 800-145. The NIST Definition of Cloud Computing:

① On-demand, self-service
- get resources (CPU, storage, bandwidth etc),
- automated: as needed, right now!

② Network access
- services accessible over the network, standard protocols

③ Pooled resources
- provider: multi-tenant pool of resources
- dynamically assigned and reassigned per customer demand

④ Elasticity
- Scalability: rapidly adjust resource usage as needed

⑤ Measured service
- monitor resource usage
- billing for resources used

Benefits

Flexibility:

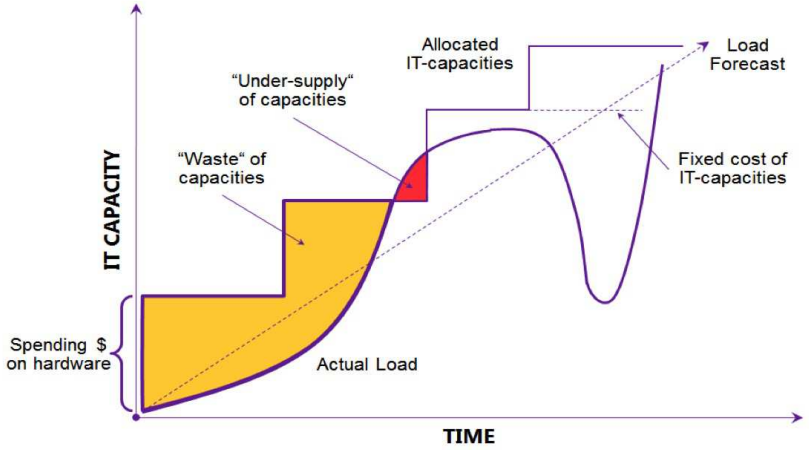
- ➜ Flexible provisioning
- ➜ Add machines on demand
- ➜ Add storage on demand

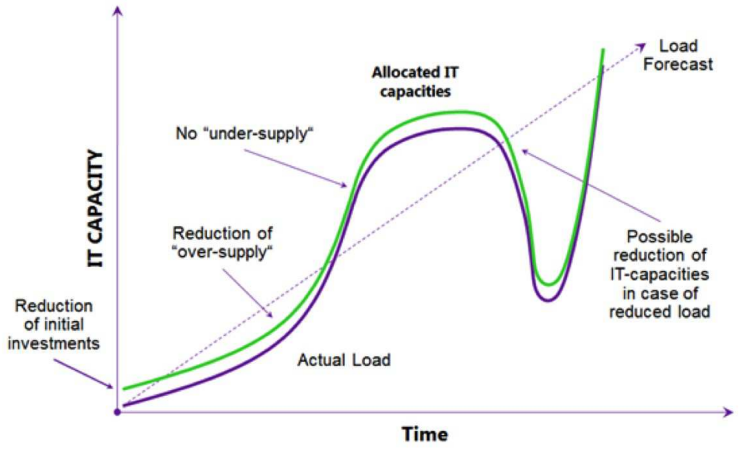
Effort:

- ➜ Low barrier to entry
- ➜ Initial effort: no need to spec and set up physical infrastructure
- ➜ Continuing effort: no need to maintain physical infrastructure

Cost:

- ➜ Low initial capital expenditure
- ➜ Avoid costs of over-provisioning for scalability
- ➜ Pay for what you use



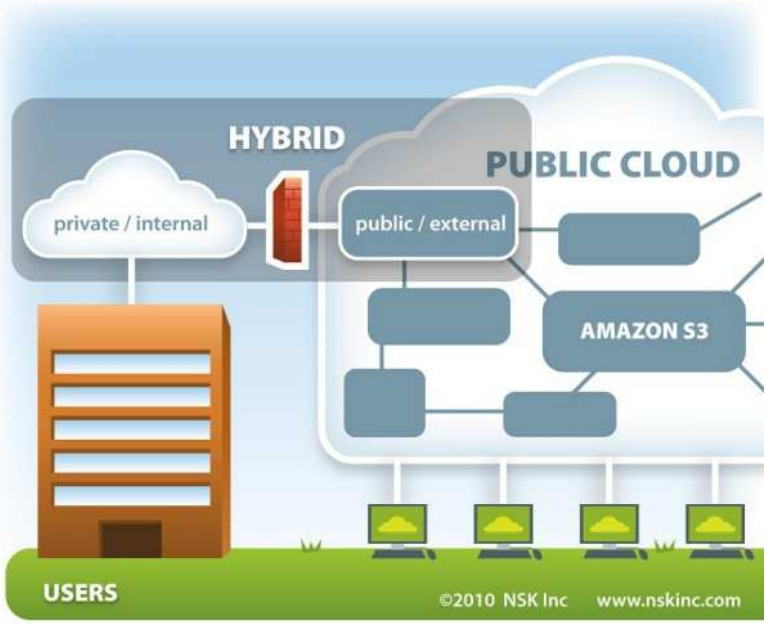


in "Developing and Extending Applications for Windows Azure with Visual Studio"

Reliability:

- Redundancy
- Trust reliability of provider
- Data backups
- *What happens when provider goes down?*
- *What about Security? Privacy?*

Public vs Private Clouds?



Public: open services available to everyone

Private: owned, operated, and available to specific organisation

Is this still cloud computing?

Hybrid: system uses some private cloud services and some public cloud services.

`http://blog.nskinc.com/IT-Services-Boston/bid/32590/Private-Cloud-or-Public-Cloud`

Infrastructure as a Service: IaaS

Service provider provides:

- Server and network hardware
- Virtual machines
- IP addresses
- Services to manage VMs (create, start, stop, migrate)
- Optional: storage, database, synchronisation, communication

Client provides:

- OS and OS environment
- Web server, DBMS, etc.
- Middleware
- Application software

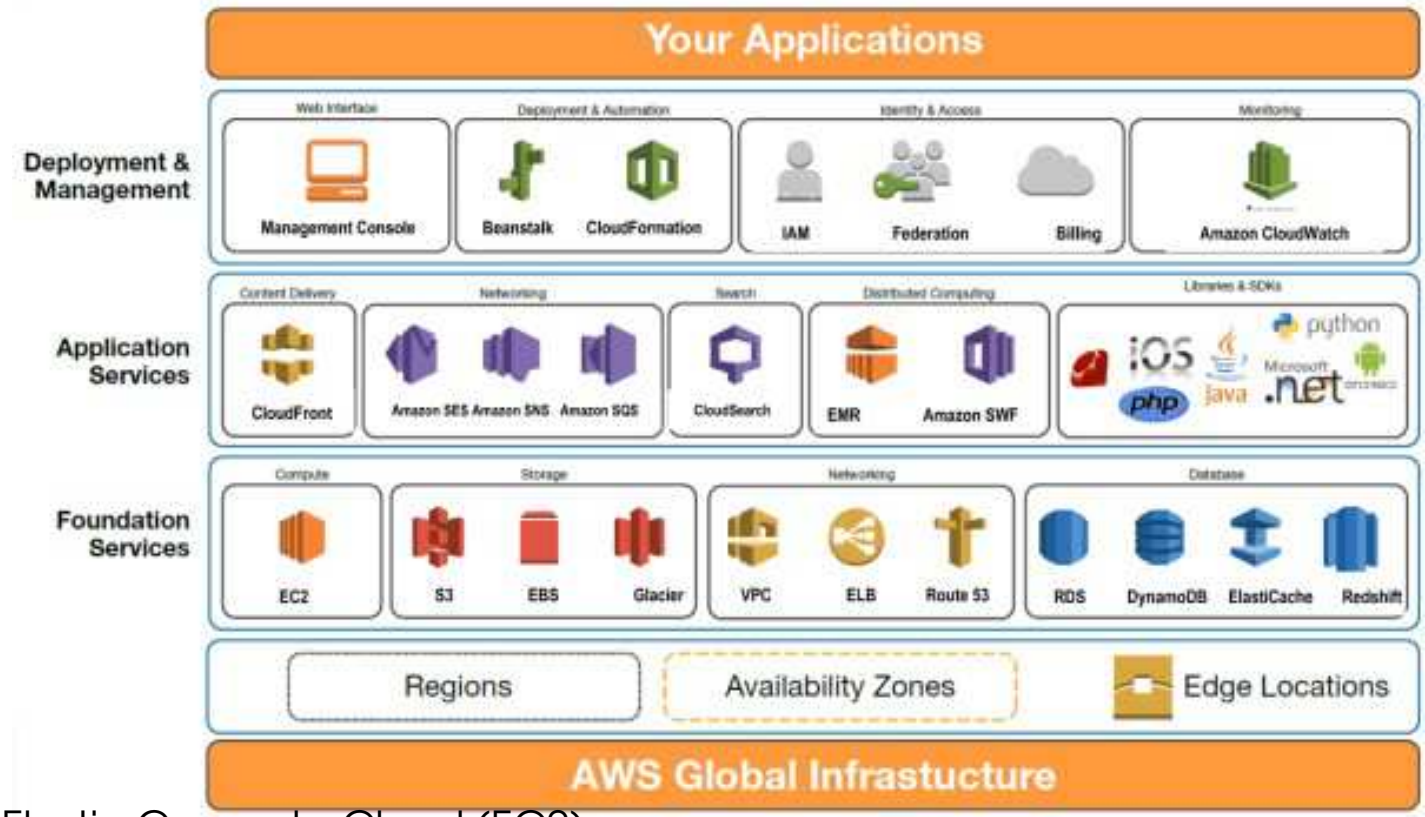
Challenges – Client:

- Transparency (naming, redirection)
- Scalability: replication and load balancing decisions
- Synchronisation and coordination
- Security
- Fault tolerance
- Software maintenance and sys admin

Challenges – Provider:

- Hardware provisioning and maintenance
- Load management
- IP address management, DNS management
- Infrastructure fault tolerance
- Monitoring, logging, billing
- Storage

EXAMPLE 1: AMAZON WEB SERVICES (AWS)



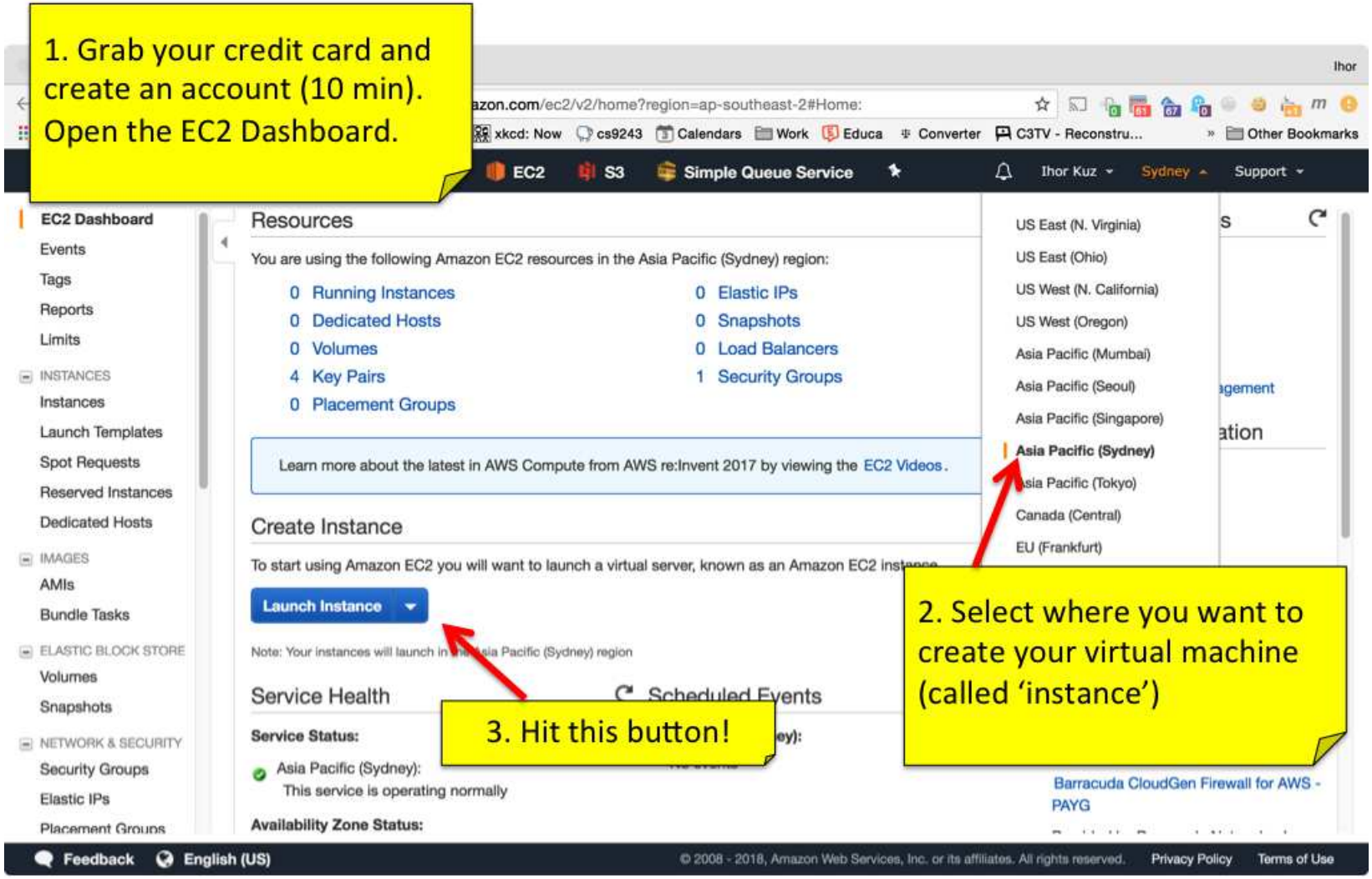
- ➜ Elastic Compute Cloud (EC2)
- ➜ Simple Storage Solution (S3)
- ➜ Simple DB
- ➜ Simple Queue Service

http://vmtoday.com/2013/07/introduction-to-amazon-web-services-aws/

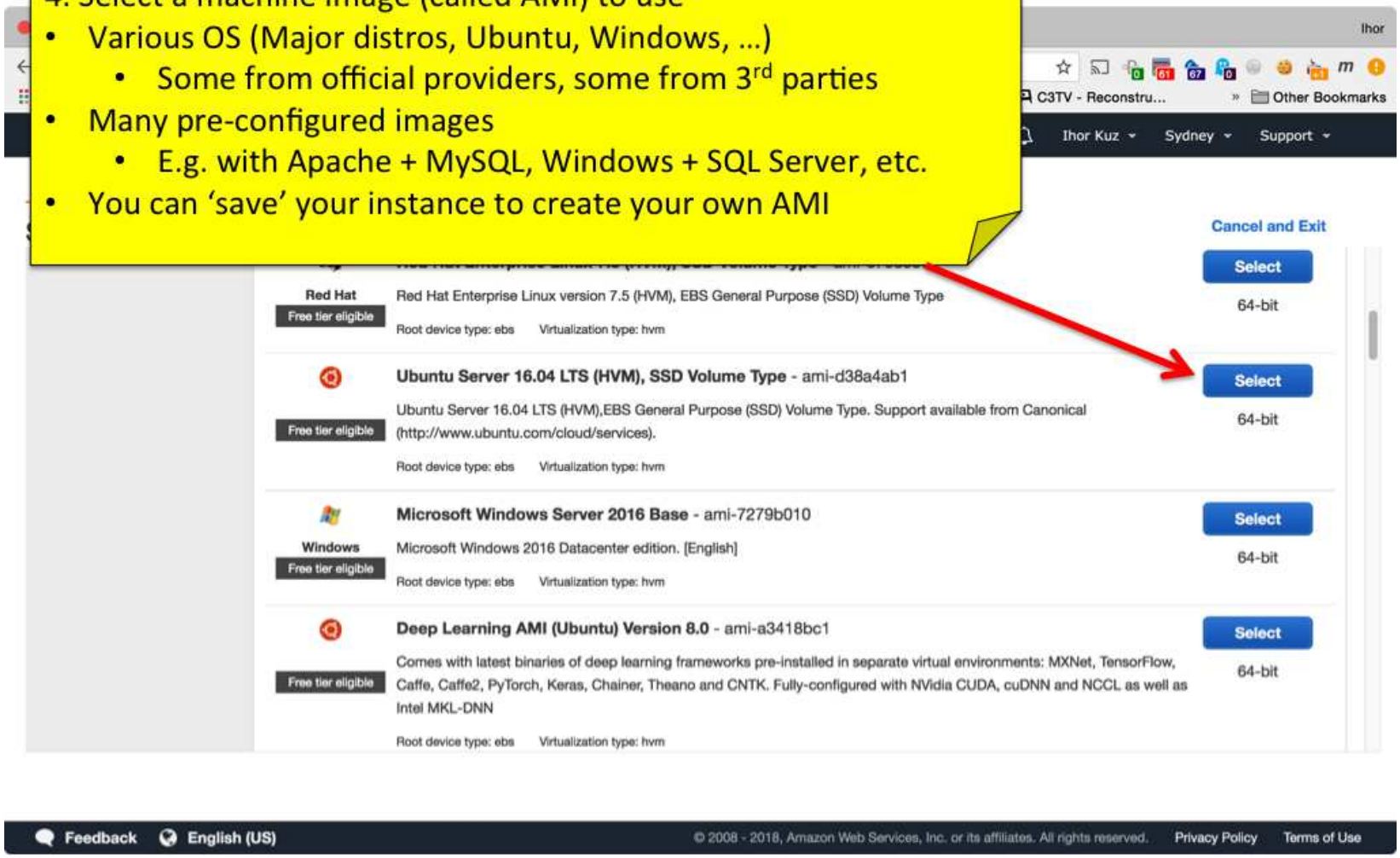
Elastic Compute Cloud (EC2):

- ➜ Instances: virtual cores, memory, storage
  - instance types (cpu,memory,net, storage options):
  - t, m, c, p, g, x, r, i, d
  - micro, small, medium, large, xlarge, ...
- ➜ Cost:
  - free tier: limited instances, free CPU hours
  - on-demand: $0.007 - $39 per hour
  - reserved: 1-3 years, discounted, fixed cost
- ➜ Launch Amazon Machine Image (AMI) on instances
- ➜ Preconfigured or custom images

USING EC2


1. Grab your credit card and create an account (10 min). Open the EC2 Dashboard.
EC2
S3
Simple Queue Service
Ihor Kuz
Sydney
Support
EC2 Dashboard
Events
Tags
Reports
Limits
INSTANCES
Instances
Launch Templates
Spot Requests
Reserved Instances
Dedicated Hosts
IMAGES
AMIs
Bundle Tasks
ELASTIC BLOCK STORE
Volumes
Snapshots
NETWORK & SECURITY
Security Groups
Elastic IPs
Placement Groups
Resources
You are using the following Amazon EC2 resources in the Asia Pacific (Sydney) region:
0 Running Instances
0 Dedicated Hosts
0 Volumes
4 Key Pairs
0 Placement Groups
0 Elastic IPs
0 Snapshots
0 Load Balancers
1 Security Groups
Learn more about the latest in AWS Compute from AWS re:Invent 2017 by viewing the EC2 Videos.
Create Instance
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 (Sydney) region
Service Health
Service Status:
Asia Pacific (Sydney):
This service is operating normally
Availability Zone Status:
Scheduled Events
US East (N. Virginia)
US East (Ohio)
US West (N. California)
US West (Oregon)
Asia Pacific (Mumbai)
Asia Pacific (Seoul)
Asia Pacific (Singapore)
Asia Pacific (Sydney)
Asia Pacific (Tokyo)
Canada (Central)
EU (Frankfurt)
2. Select where you want to create your virtual machine (called 'instance')
3. Hit this button!
Barracuda CloudGen Firewall for AWS - PAYG
Feedback
English (US)
© 2008 - 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use

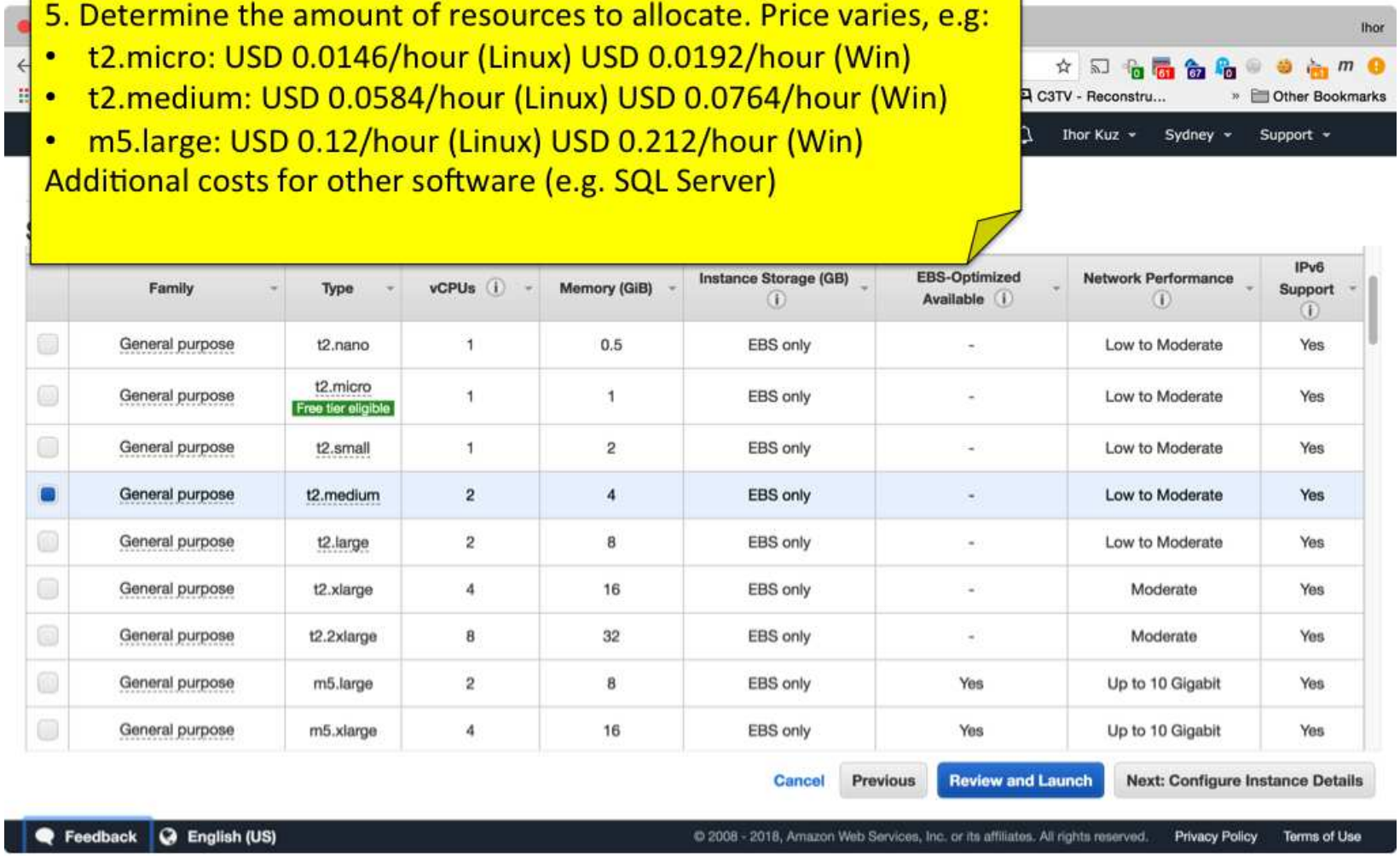
4. Select a machine image (called AMI) to use
- Various OS (Major distros, Ubuntu, Windows, …)
  - Some from official providers, some from 3rd parties
- Many pre-configured images
  - E.g. with Apache + MySQL, Windows + SQL Server, etc.
- You can 'save' your instance to create your own AMI



5. Determine the amount of resources to allocate. Price varies, e.g:

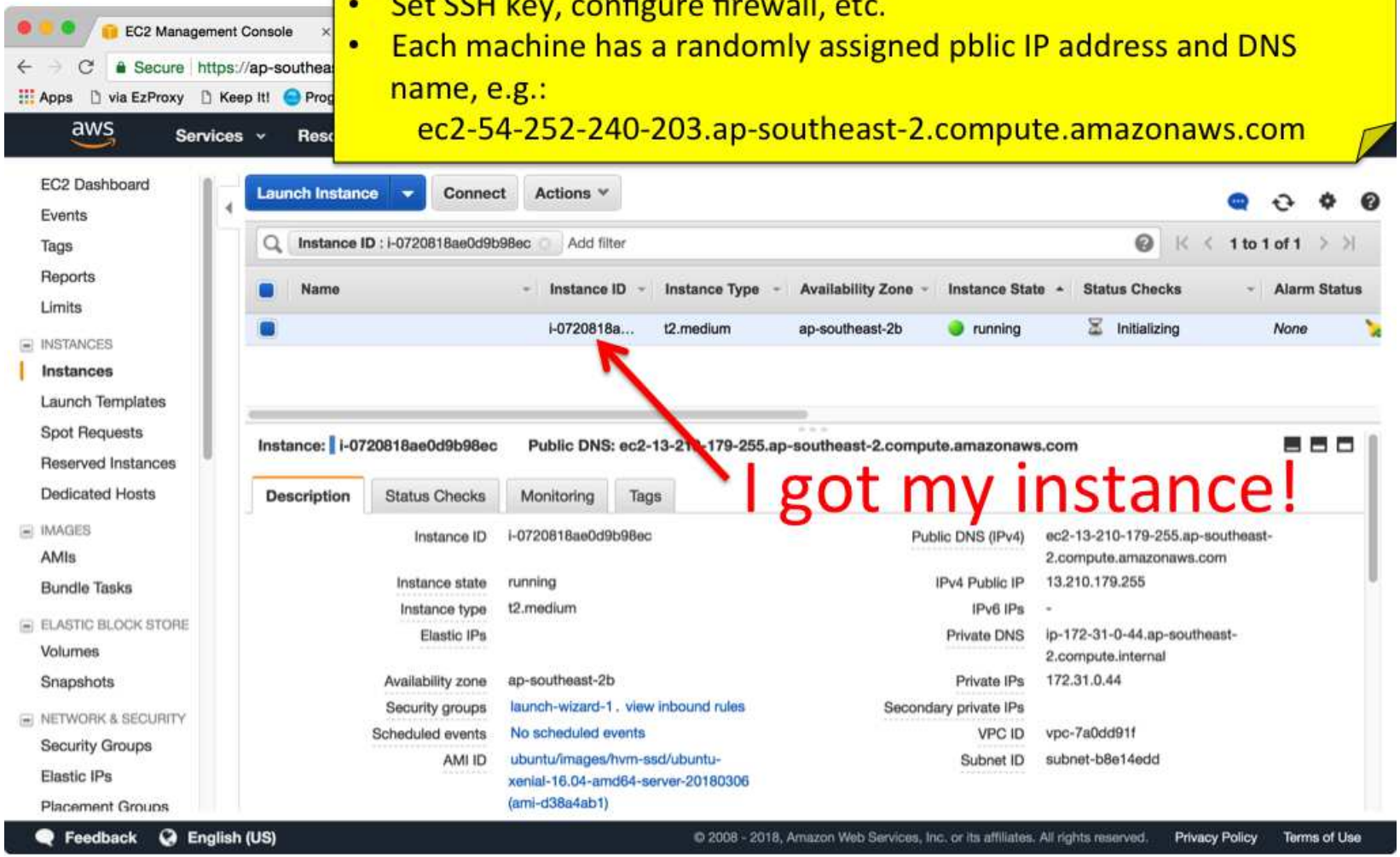
- t2.micro: USD 0.0146/hour (Linux) USD 0.0192/hour (Win)
- t2.medium: USD 0.0584/hour (Linux) USD 0.0764/hour (Win)
- m5.large: USD 0.12/hour (Linux) USD 0.212/hour (Win)

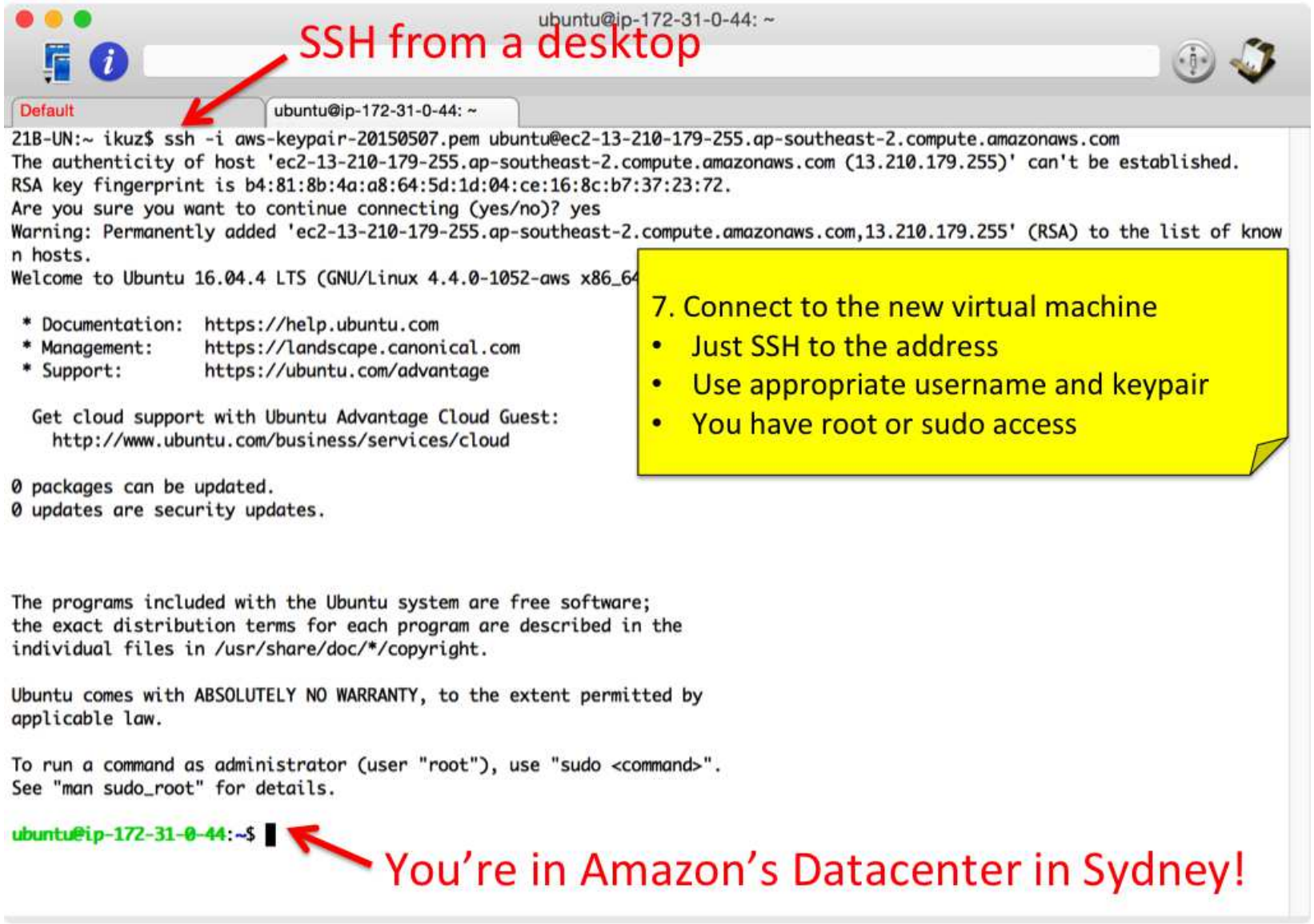
Additional costs for other software (e.g. SQL Server)

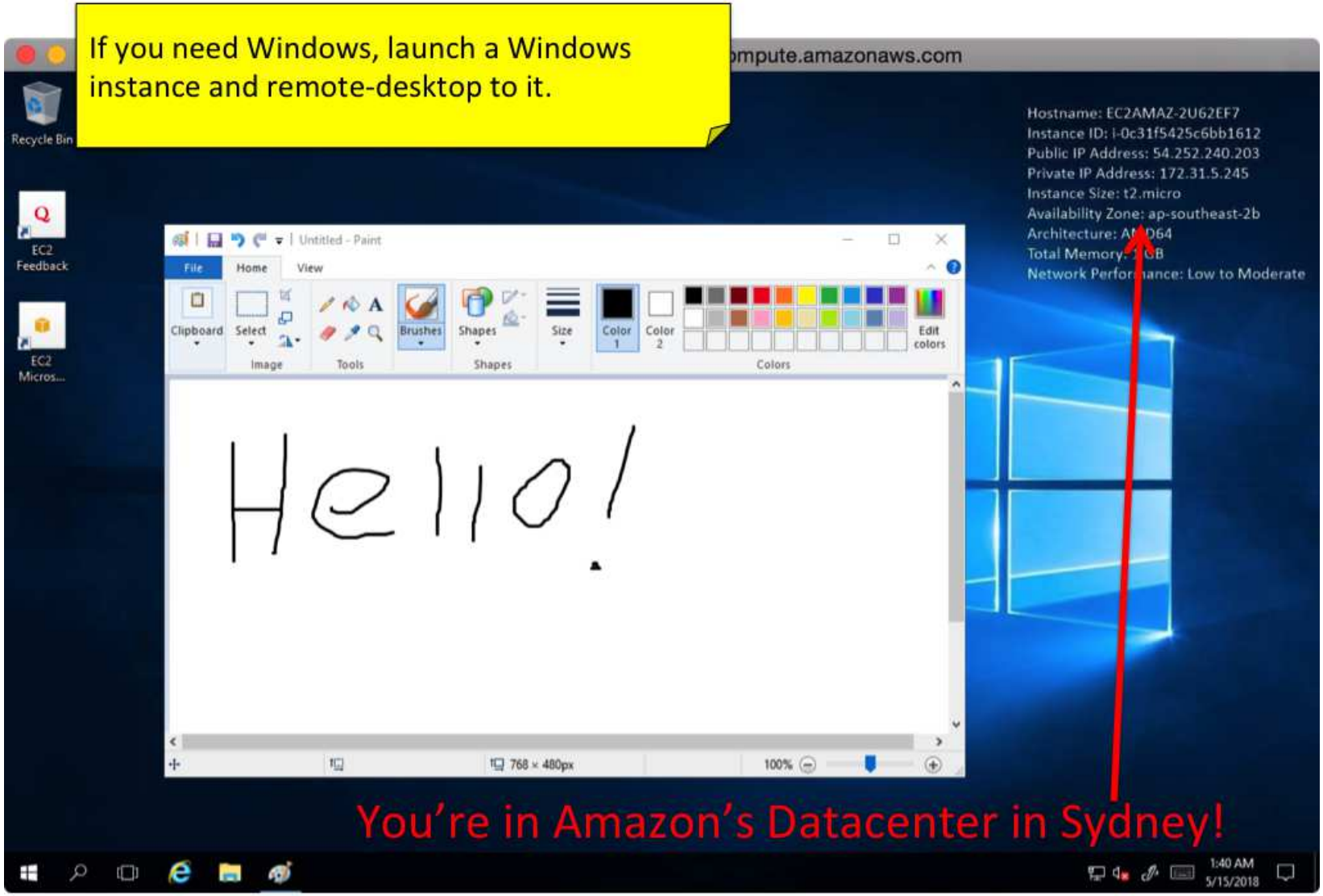


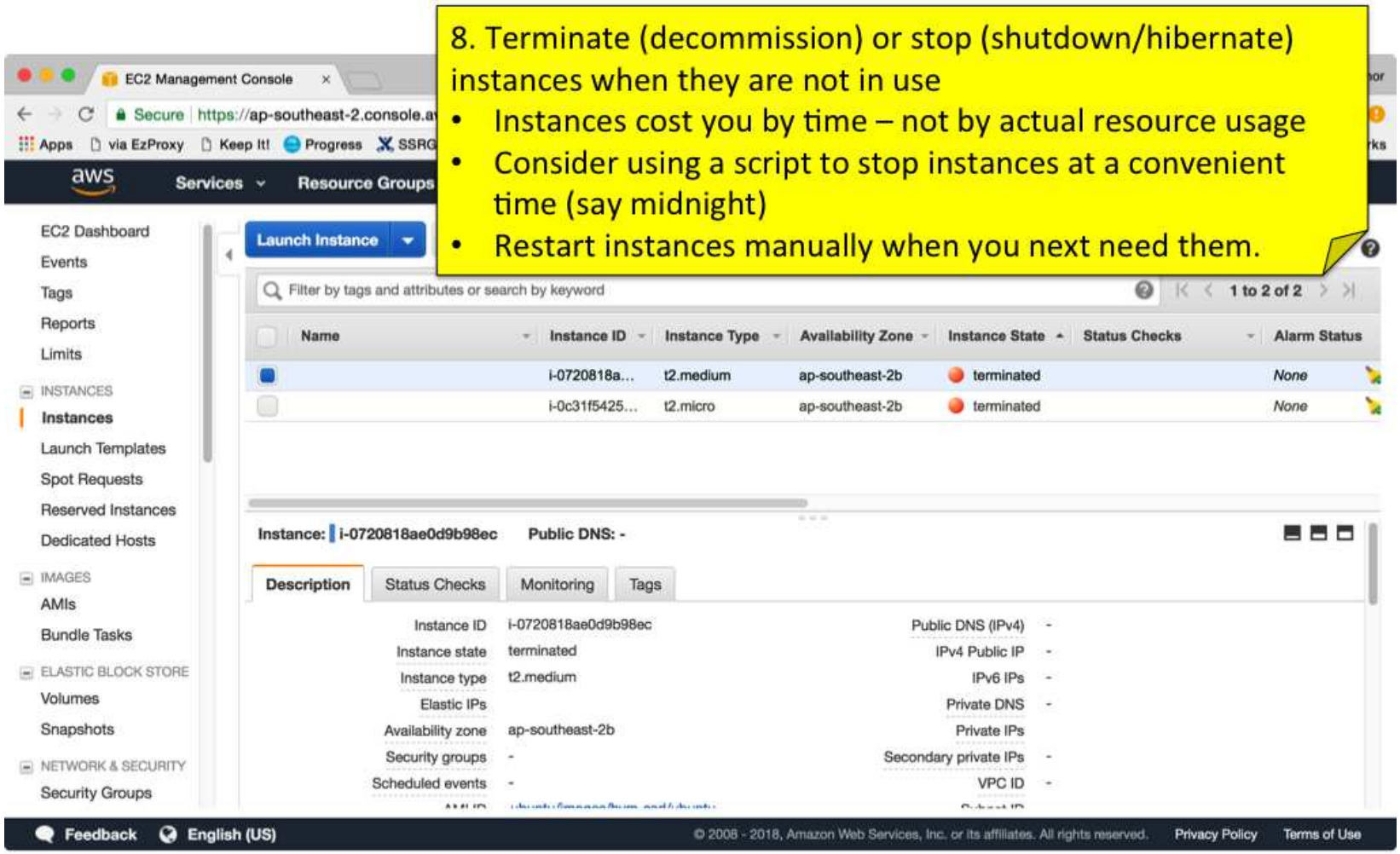
Family	Type	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance	IPv6 Support
General purpose	t2.nano	1	0.5	EBS only	-	Low to Moderate	Yes
General purpose	t2.micro Free tier eligible	1	1	EBS only	-	Low to Moderate	Yes
General purpose	t2.small	1	2	EBS only	-	Low to Moderate	Yes
General purpose	t2.medium	2	4	EBS only	-	Low to Moderate	Yes
General purpose	t2.large	2	8	EBS only	-	Low to Moderate	Yes
General purpose	t2.xlarge	4	16	EBS only	-	Moderate	Yes
General purpose	t2.2xlarge	8	32	EBS only	-	Moderate	Yes
General purpose	m5.large	2	8	EBS only	Yes	Up to 10 Gigabit	Yes
General purpose	m5.xlarge	4	16	EBS only	Yes	Up to 10 Gigabit	Yes

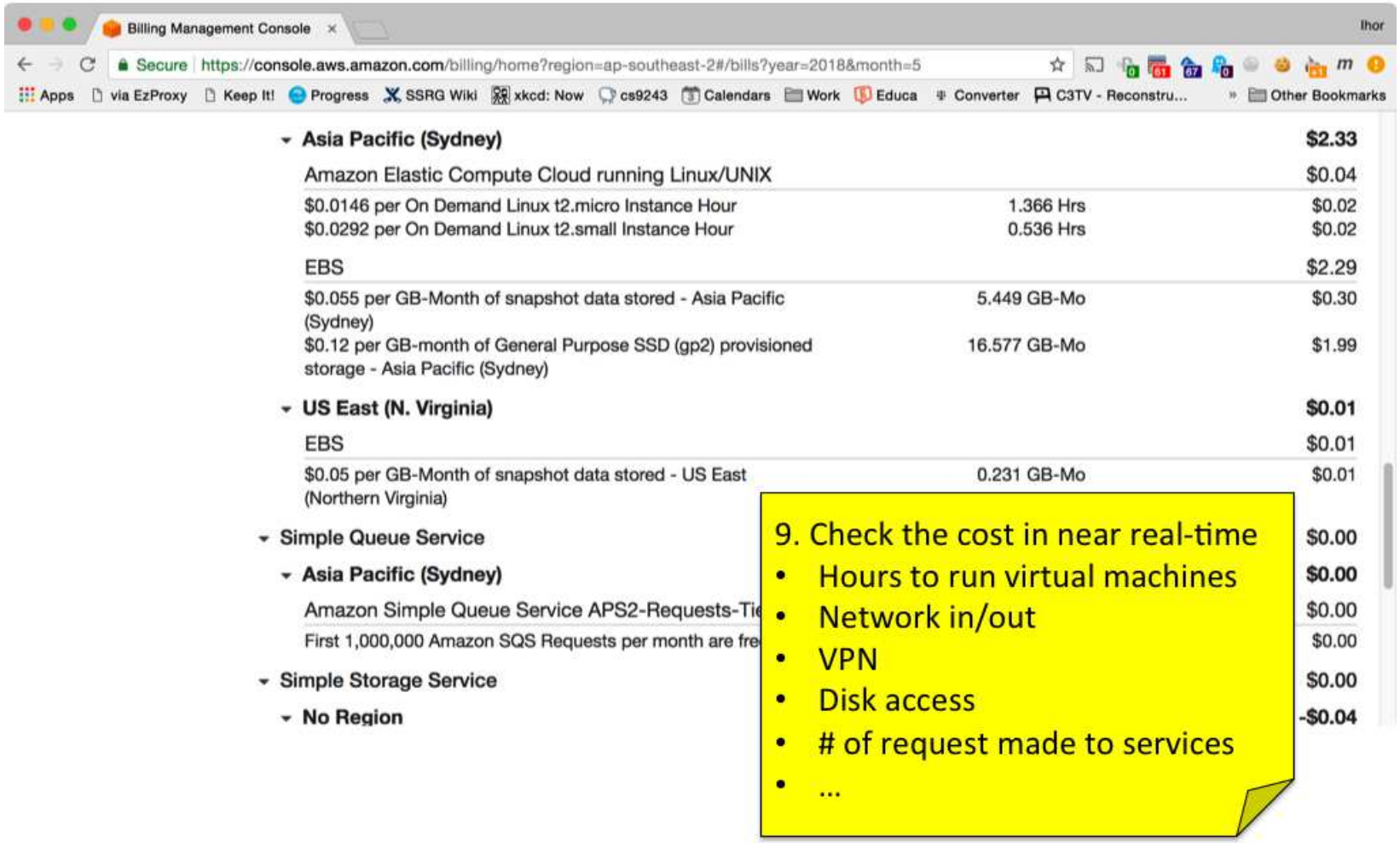
6. Done! (< 5 minutes in total)
- Set SSH key, configure firewall, etc.
- Each machine has a randomly assigned pblic IP address and DNS name, e.g.:
  ec2-54-252-240-203.ap-southeast-2.compute.amazonaws.com



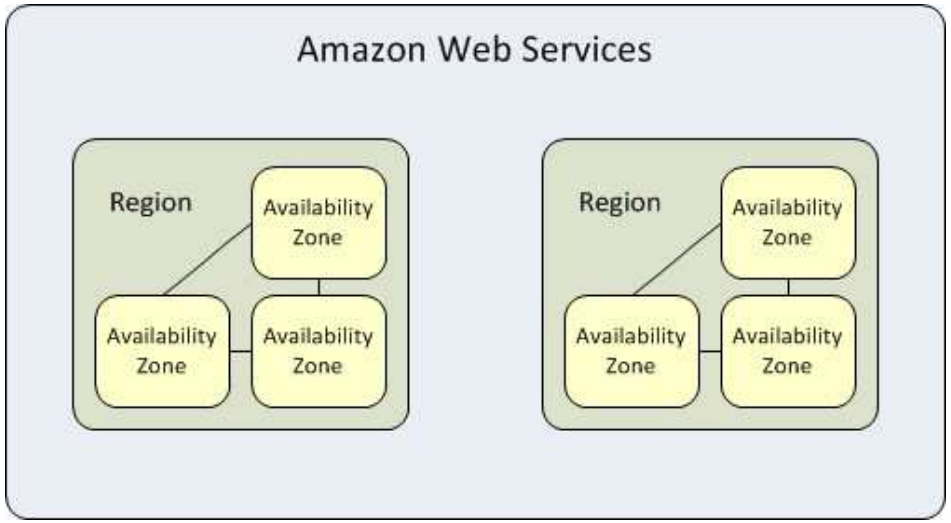

ubuntu@ip-172-31-0-44: ~
SSH from a desktop
Default
ubuntu@ip-172-31-0-44: ~
21B-UN:~ ikuz$ ssh -i aws-keypair-20150507.pem ubuntu@ec2-13-210-179-255.ap-southeast-2.compute.amazonaws.com
The authenticity of host 'ec2-13-210-179-255.ap-southeast-2.compute.amazonaws.com (13.210.179.255)' can't be established.
RSA key fingerprint is b4:81:8b:4a:a8:64:5d:1d:04:ce:16:8c:b7:37:23:72.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-13-210-179-255.ap-southeast-2.compute.amazonaws.com,13.210.179.255' (RSA) to the list of know
n hosts.
Welcome to Ubuntu 16.04.4 LTS (GNU/Linux 4.4.0-1052-aws x86_64
* Documentation: https://help.ubuntu.com
* Management: https://landscape.canonical.com
* Support: https://ubuntu.com/advantage
Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud
0 packages can be updated.
0 updates are security updates.
The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.
ubuntu@ip-172-31-0-44:~$
7. Connect to the new virtual machine
• Just SSH to the address
• Use appropriate username and keypair
• You have root or sudo access
You're in Amazon's Datacenter in Sydney!


If you need Windows, launch a Windows instance and remote-desktop to it.
Recycle Bin
EC2 Feedback
EC2 Micros...
Hostname: EC2AMAZ-2U62EF7
Instance ID: i-0c31f5425c6bb1612
Public IP Address: 54.252.240.203
Private IP Address: 172.31.5.245
Instance Size: t2.micro
Availability Zone: ap-southeast-2b
Network Performance: Low to Moderate
Untitled - Paint
File
Home
View
Clipboard
Select
Image
Tools
Brushes
Shapes
Size
Color 1
Color 2
Edit colors
Colors
Hello!
768 × 480px
100%
You're in Amazon's Datacenter in Sydney!
1:40 AM
5/15/2018


8. Terminate (decommission) or stop (shutdown/hibernate) instances when they are not in use
• Instances cost you by time – not by actual resource usage
• Consider using a script to stop instances at a convenient time (say midnight)
• Restart instances manually when you next need them.
EC2 Management Console
Secure https://ap-southeast-2.console.a
Apps via EzProxy Keep It! Progress SSRG
aws Services Resource Groups
EC2 Dashboard
Events
Tags
Reports
Limits
INSTANCES
Instances
Launch Templates
Spot Requests
Reserved Instances
Dedicated Hosts
IMAGES
AMIs
Bundle Tasks
ELASTIC BLOCK STORE
Volumes
Snapshots
NETWORK & SECURITY
Security Groups
Launch Instance
Filter by tags and attributes or search by keyword
1 to 2 of 2
Name Instance ID Instance Type Availability Zone Instance State Status Checks Alarm Status
i-0720818a... t2.medium ap-southeast-2b terminated None
i-0c31f5425... t2.micro ap-southeast-2b terminated None
Instance: i-0720818ae0d9b98ec Public DNS: -
Description Status Checks Monitoring Tags
Instance ID i-0720818ae0d9b98ec
Instance state terminated
Instance type t2.medium
Elastic IPs
Availability zone ap-southeast-2b
Security groups -
Scheduled events -
Public DNS (IPv4) -
IPv4 Public IP -
IPv6 IPs -
Private DNS -
Private IPs
Secondary private IPs -
VPC ID -
Feedback English (US)
© 2008 - 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy Terms of Use


Billing Management Console
Secure https://console.aws.amazon.com/billing/home?region=ap-southeast-2#/bills?year=2018&month=5
Apps via EzProxy Keep It! Progress SSRG Wiki xkcd: Now cs9243 Calendars Work Educa Converter C3TV - Reconstru... Other Bookmarks
Asia Pacific (Sydney) $2.33
Amazon Elastic Compute Cloud running Linux/UNIX $0.04
$0.0146 per On Demand Linux t2.micro Instance Hour 1.366 Hrs $0.02
$0.0292 per On Demand Linux t2.small Instance Hour 0.536 Hrs $0.02
EBS $2.29
$0.055 per GB-Month of snapshot data stored - Asia Pacific (Sydney) 5.449 GB-Mo $0.30
$0.12 per GB-month of General Purpose SSD (gp2) provisioned storage - Asia Pacific (Sydney) 16.577 GB-Mo $1.99
US East (N. Virginia) $0.01
EBS $0.01
$0.05 per GB-Month of snapshot data stored - US East (Northern Virginia) 0.231 GB-Mo $0.01
Simple Queue Service $0.00
Asia Pacific (Sydney) $0.00
Amazon Simple Queue Service APS2-Requests-Ti $0.00
First 1,000,000 Amazon SQS Requests per month are fre $0.00
Simple Storage Service $0.00
No Region -$0.04
9. Check the cost in near real-time
• Hours to run virtual machines
• Network in/out
• VPN
• Disk access
• # of request made to services
• ...

Reliability



http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-regions-availability-zones.html

Regions and Availability Zones:

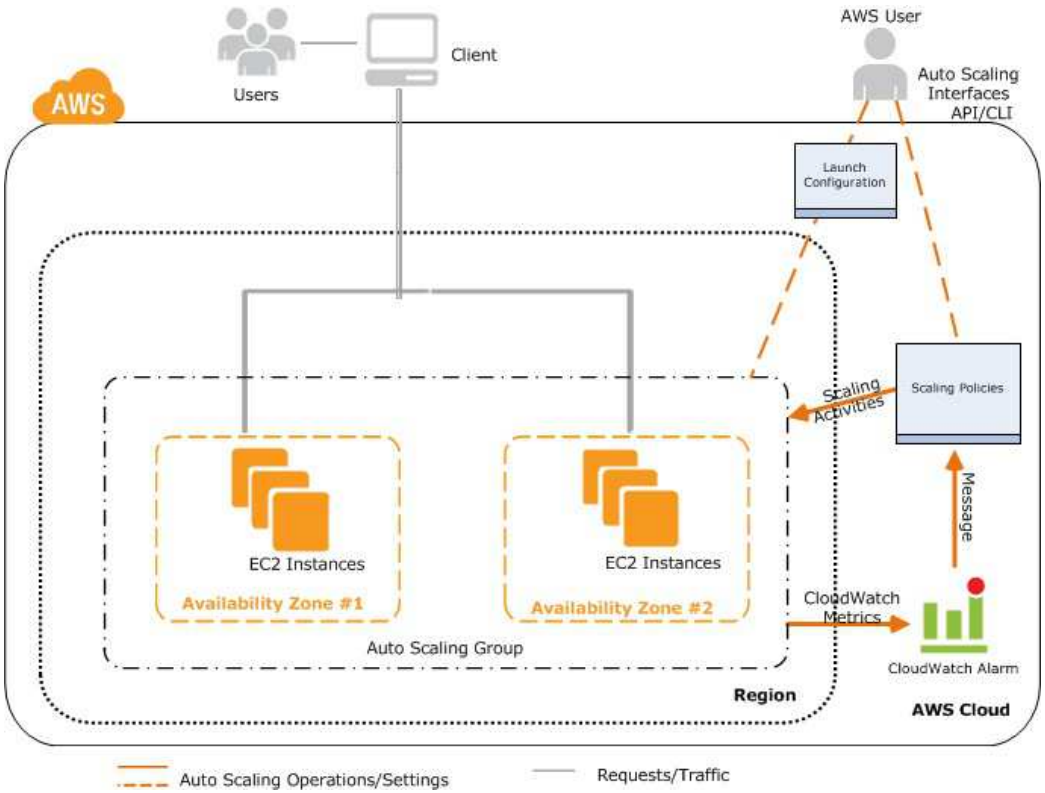
- 99.95% availability per service region
- Regions: geographically dispersed, independent
- Availability zones: contained in Regions
- Availability zones: isolated from failures in other zones, but connected

Elastic IP addresses:

- ➜ IP address associated with account
- ➜ Dynamic remapping to specific instances
  - instance has *private IP address* and *public IP address*
  - *Elastic IP* can be mapped (and re-mapped) to private IP

Elastic Load Balancing:

- ➜ Distributes traffic across instances
- ➜ Monitors 'health' of instances: customisable
- ➜ Routes to healthy instances



Auto Scaling:

- ➜ Automatically start or stop new instances
- ➜ User-defined conditions
  - manual (minimum group size), schedule
  - instance health, CloudWatch input

`https://docs.aws.amazon.com/autoscaling/ec2/userguide/what-is-amazon-ec2-auto-scaling.html`

Security:

- ➜ Infrastructure Security
  - Data centre physical security
  - Software and hardware maintenance
  - Monitoring and Testing (automatic and manual)
- ➜ Application Security
  - API access control (access keys)
  - Firewall settings for instances (security groups)
  - Virtual Private Cloud (VPC): private or public subnetworks
  - Encrypted storage support
  - Logging

STORAGE

Elastic Block Store:

- ➜ Network Attached Storage (NAS) (servers with disks)
- ➜ Block level storage volumes
- ➜ Mounted as block device (e.g. disk) on an instance
- ➜ Physical Servers and Disks shared by customers (no caching, competing for disk and net IO)
- ➜ Replicated in Availability zone
- ➜ Cost: per GB/per month

Simple Storage Service (S3):

- Buckets: store objects
  - Can be placed in specific regions
- Objects: data and metadata
  - metadata: key-value pairs describing the object
  - identified by key (unique within a bucket)
  - versioned
- Consistency:
  - highly replicated
  - eventual consistency, no locking
  - atomic object update
- Access control

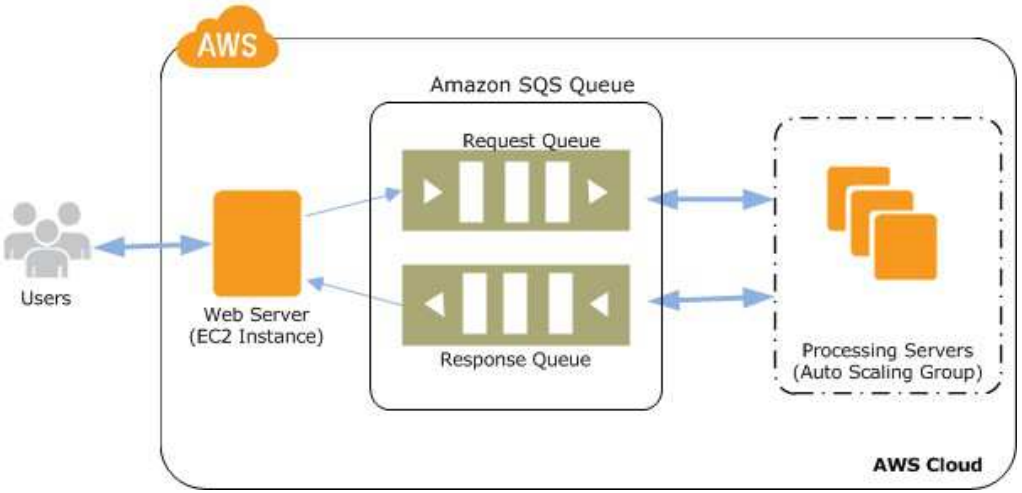
Snapshots:

- ➜ Point in time copy of EBS volume
- ➜ Stored in S3
- ➜ Differential
- ➜ Can be used to bootstrap image

Simple Database Service (SimpleDB):

- ➜ Non-relational database: key-value
- ➜ Partitioned into *domains*
- ➜ Consistency
  - highly replicated
  - eventual consistency
- ➜ Typical uses: logging, indexing S3 data
- ➜ Erlang!
- ➜ Replaced by DynamoDB

COMMUNICATION



Simple Queue Service (SQS):

- ➜ Message-queue oriented communication service
- ➜ Persistent, asynchronous messaging
- ➜ At-least once delivery guarantee
- ➜ No ordering guarantee
- ➜ Access control

`https://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/`

Platform as a Service

Service provider provides:

- Hardware infrastructure
- OS and platform software (middleware)
- Distributed storage management
- Load balancing, replication, migration
- Management and Monitoring services

Client provides:

- Application

Challenges – Client:

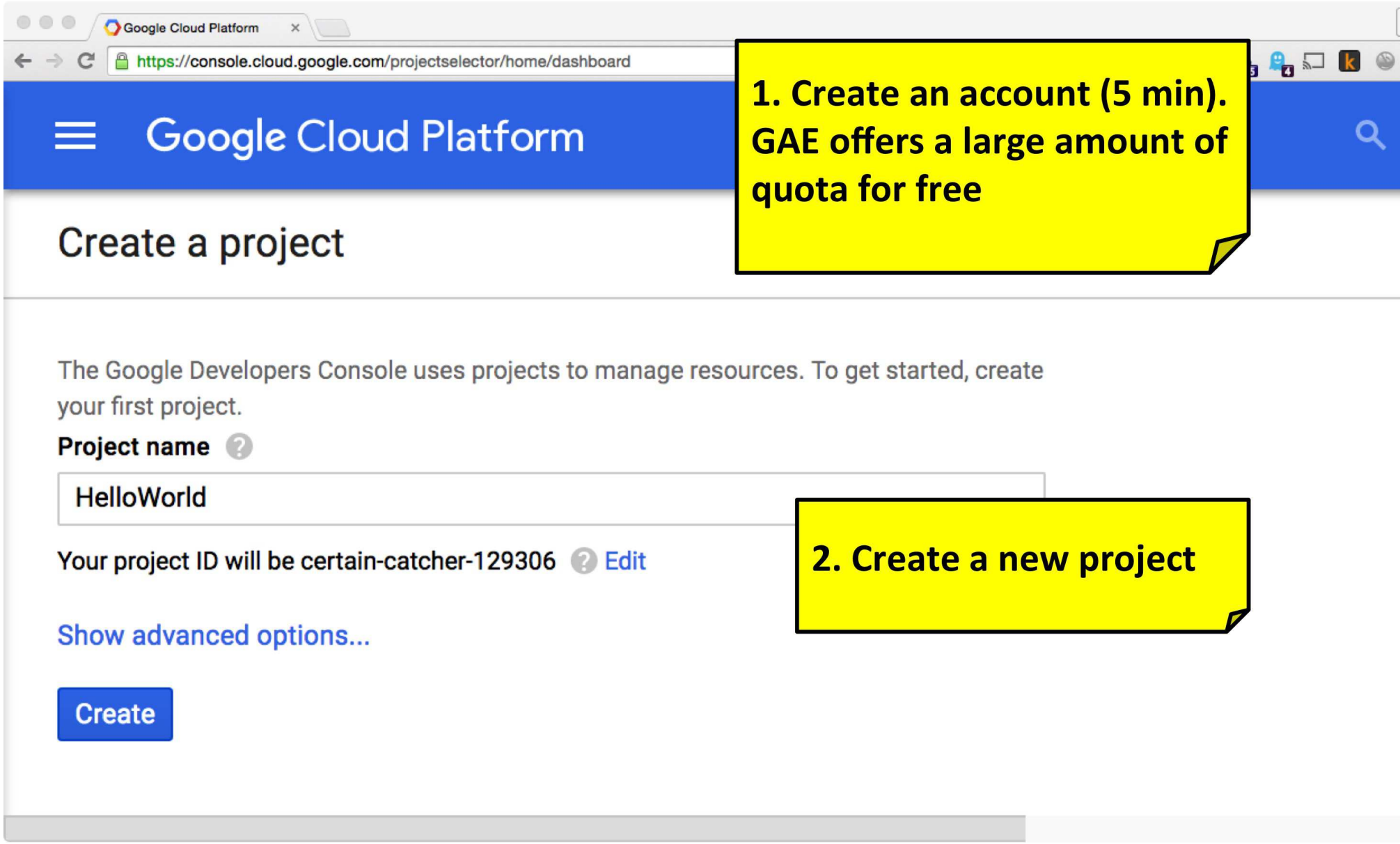
- ➜ Learn new API and environment
- ➜ Follow API
- ➜ Optimise to limits of API and platform
- ➜ Security for own app

Challenges – Provider:

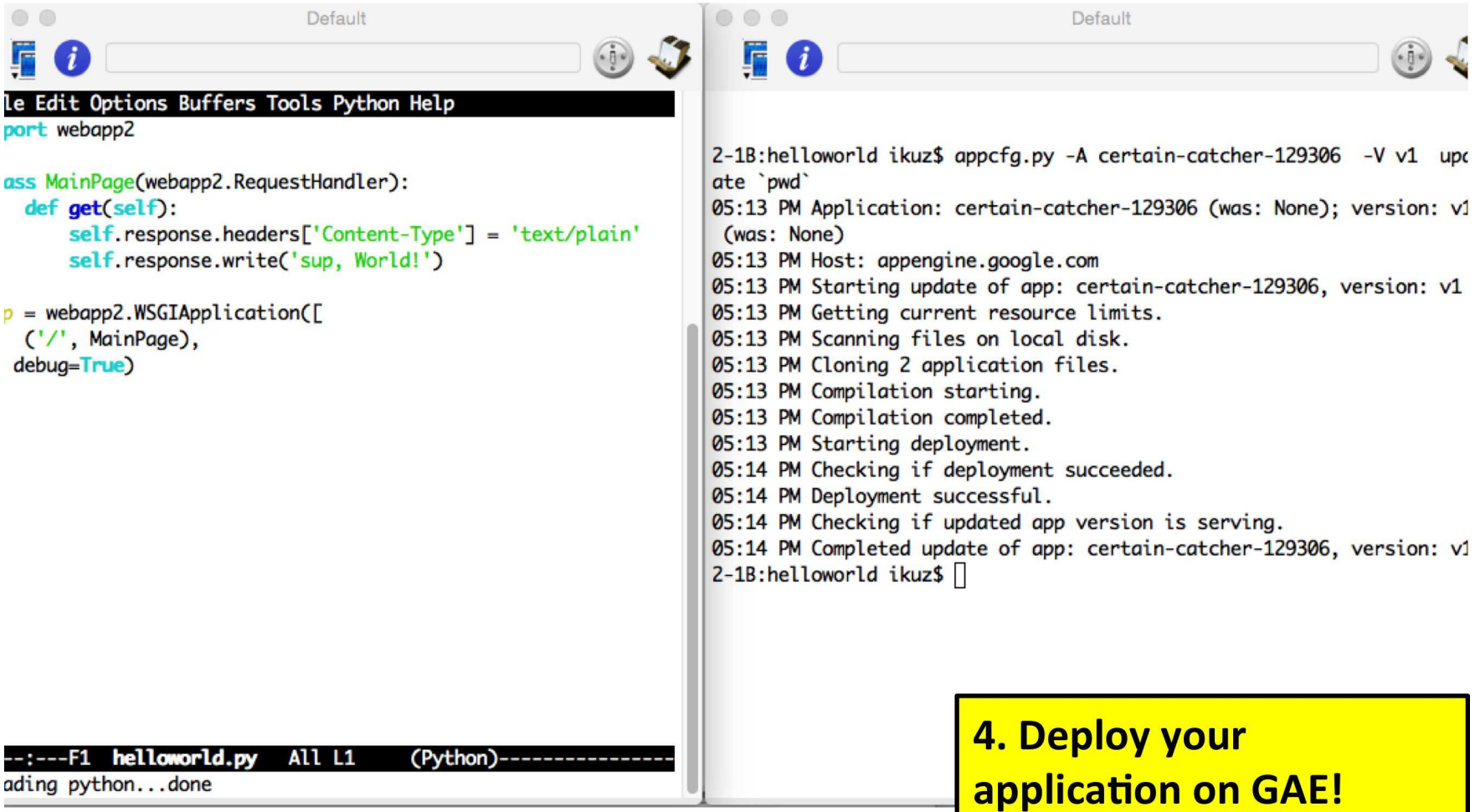
- ➜ Transparency (naming, redirection)
- ➜ Scalability: replication and load balancing decisions
- ➜ Synchronisation and coordination
- ➜ Security
- ➜ Fault tolerance
- ➜ Monitoring
- ➜ Software maintenance and sys admin

Example 2: App Engine

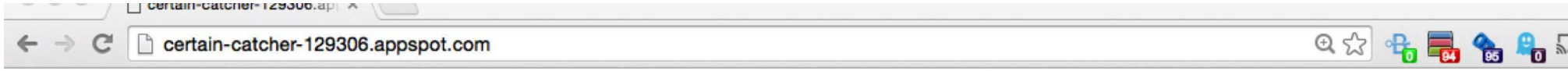
- Various development languages (Python, Java, PHP, Go)
- ... and runtime environments
- Storage based on Big Table
- Optimisation via Memcache
- Lots of APIs
- Per use billing
- Transparent scaling


Google Cloud Platform
https://console.cloud.google.com/projectselector/home/dashboard
Google Cloud Platform
1. Create an account (5 min). GAE offers a large amount of quota for free
Create a project
The Google Developers Console uses projects to manage resources. To get started, create your first project.
Project name
HelloWorld
Your project ID will be certain-catcher-129306 Edit
2. Create a new project
Show advanced options...
Create

3. Write an application using GAE's framework



4. Deploy your application on GAE!

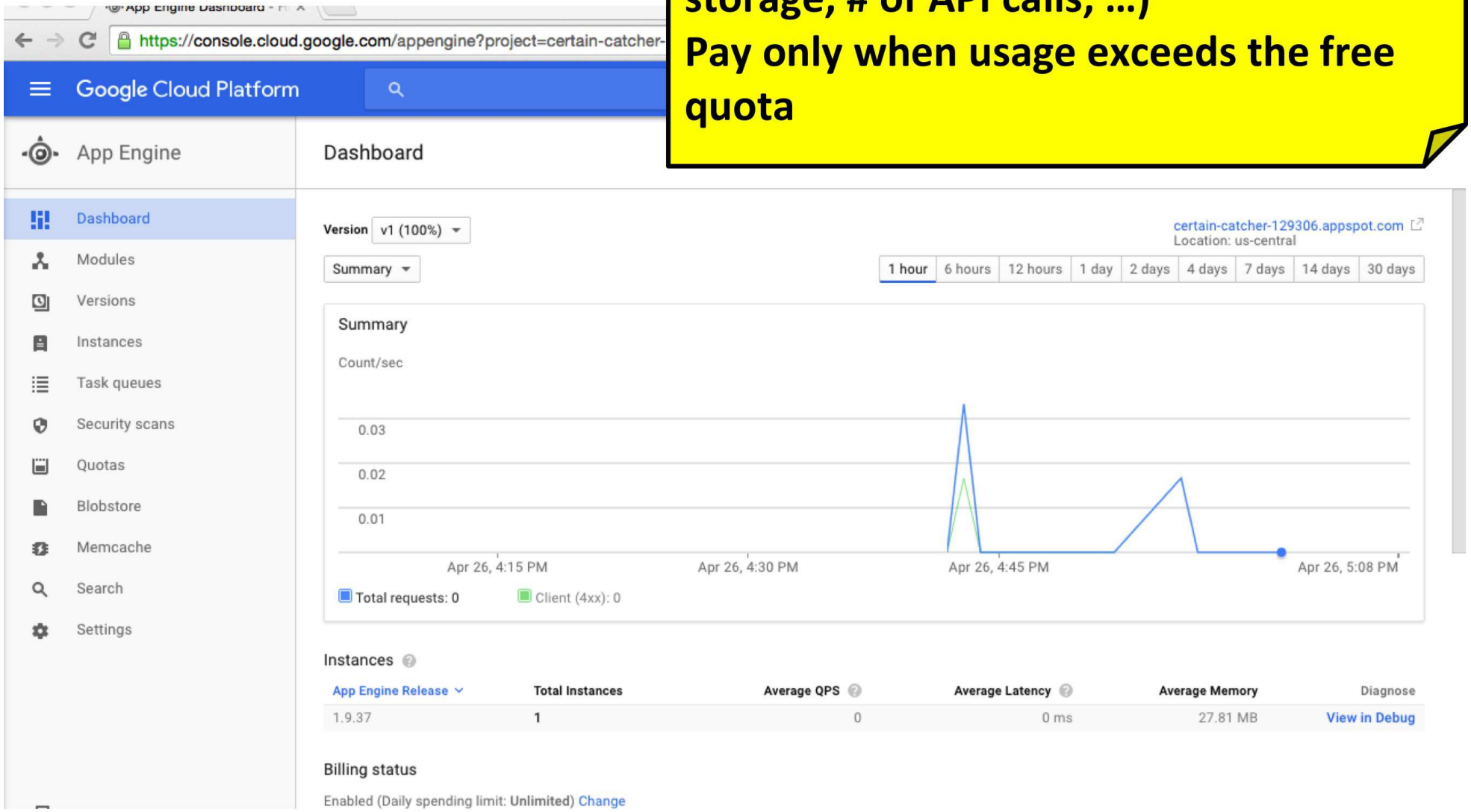


sup, World!

5. Running application.

Scale up/down, load balancing, replication, database management, … many services are provided by GAE.

6. Check your resource usage (CPU, storage, # of API calls, …)
Pay only when usage exceeds the free quota



Software as a Service

Service provider provides:

- Hardware infrastructure
- OS and platform software (middleware)
- Distributed storage management
- Load balancing, replication, migration
- Management and Monitoring services
- Application

Client provides:

- Data

Challenges – Client:

- ➜ Learn new application
- ➜ Deal with potential restrictions
  - Web interface, restricted functionality
  - No offline access, no local storage

Challenges – Provider:

- ➜ Transparency (naming, redirection)
- ➜ Scalability: replication and load balancing decisions
- ➜ Synchronisation and coordination
- ➜ Security
- ➜ Fault tolerance
- ➜ Monitoring
- ➜ Software maintenance and sys admin
- ➜ Application development and maintenance

Key Challenges of Cloud Computing

Scalability:

- ➜ Datacentre vs Global
- ➜ Partitioning
  - Services and Data
- ➜ Replication

Consistency:

- ➜ Dealing with consequences of CAP Theorem
- ➜ Dealing with un-usability of eventual consistency

Reliability:

- ➜ SLA (Service Level Agreement): guarantees given by provider
  - How reliable are the guarantees?
  - What is the consequence if they aren't met?
- ➜ Redundancy and Replication
  - within same provider (e.g. Availability Zones, Regions, etc.)
  - migration across providers
- ➜ Geographically distributed architecture

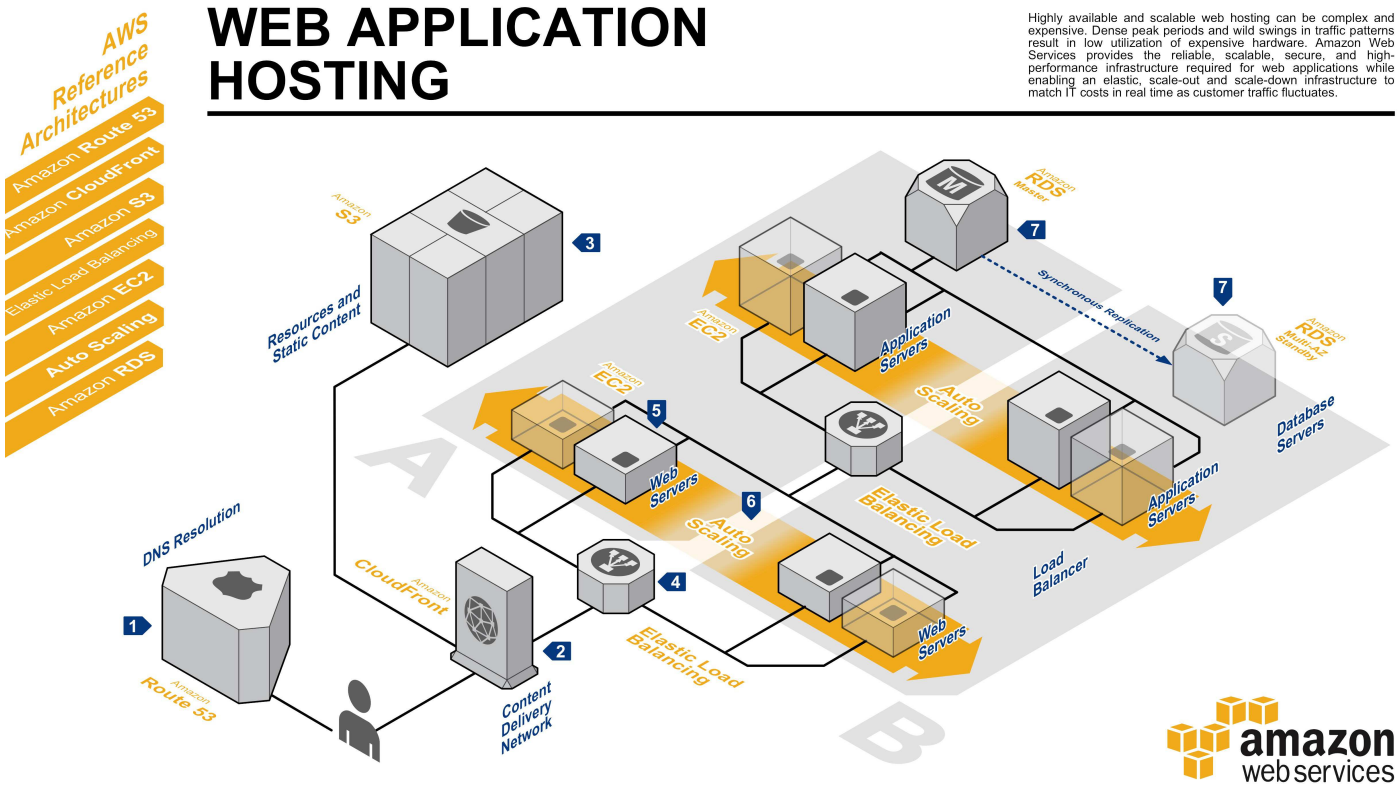
➜ Design for failure: Chaos Monkey

- test how well system deals with failure
- regularly and randomly kill system services

Security and Privacy:

- External threats
  - Denial of Service
  - Infrastructure or platform service compromise
  - SaaS compromise: data theft
- Co-located threats: other customers
  - Isolation: but, covert channels, bugs in isolation
- Privacy: data collected by providers
  - IaaS and PaaS providers: encryption only helps a bit
  - SaaS providers: at mercy of service provider
  - Governments and others: where is your data stored or processed? Which laws apply?

Examples from Amazon:



http://aws.amazon.com/architecture/