

Cloud Computing and Big Data

Understanding Cloud Services

Oxford University
Software Engineering
Programme
July 2020



Capabilities offered as-a-Service

- Software-as-a-Service
 - Salesforce, Quickbooks Online, Gmail, Gdrive, Office 365, etc
- Infrastructure-as-a-Service
 - CPUs, Memory, Disk, Networks, Firewalls, etc
 - Amazon AWS, Joyent, Microsoft Azure, IBM Softlayer, Rackspace, Google Compute Engine
- Platform-as-a-Service
 - Somewhere between!

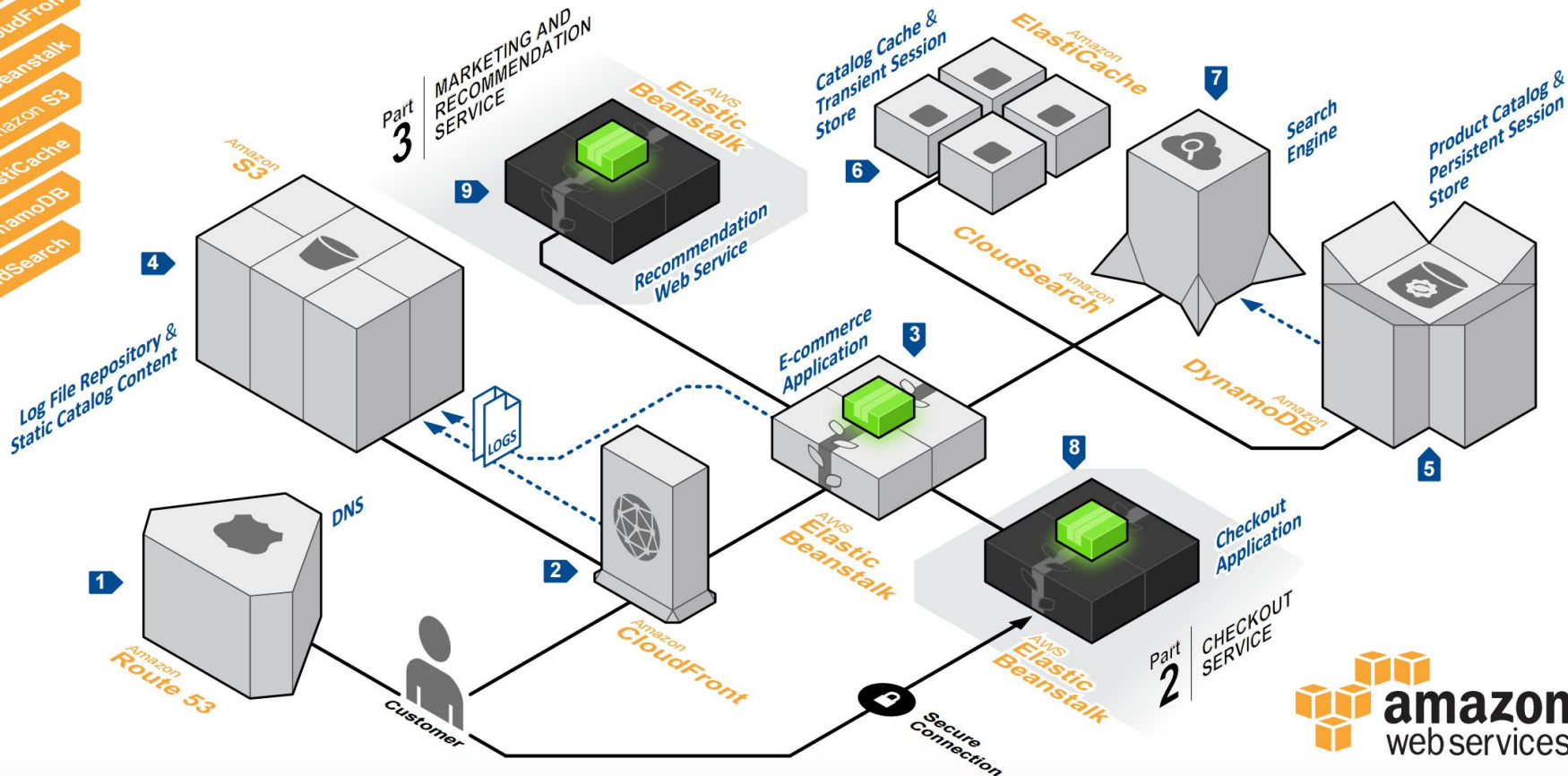


E-COMMERCE WEB SITE

PART 1: WEB FRONT-END





With Amazon Web Services, you can build a highly available e-commerce website with a flexible product catalog that scales with your business. Maintaining an e-commerce website with a large product catalog and global customer base can be challenging. The catalog should be searchable, and individual product pages should contain a rich information set that includes, for example, images, a PDF manual, and customer reviews.

Customers want to find the products they are interested in quickly, and they expect pages to load quickly. Worldwide customers want to be able to make purchases at any time, so the website should be highly available. Meeting these challenges becomes harder as your catalog and customer base grow. With the tools that AWS provides, you can build a compelling, scalable website with a searchable product catalog that is accessible with very low latency.









Amazon Web Services





Compute

-  **EC2**
Virtual Servers in the Cloud
-  **EC2 Container Service**
Run and Manage Docker Containers
-  **Elastic Beanstalk**
Run and Manage Web Apps
-  **Lambda**
Run Code in Response to Events

Storage & Content Delivery

-  **S3**
Scalable Storage in the Cloud
-  **CloudFront**
Global Content Delivery Network
-  **Elastic File System** **PREVIEW**
Fully Managed File System for EC2
-  **Glacier**
Archive Storage in the Cloud
-  **Import/Export Snowball**
Large Scale Data Transport
-  **Storage Gateway**
Integrates On-Premises IT Environments with Cloud Storage








Database

-  **RDS**
Managed Relational Database Service
-  **DynamoDB**
Predictable and Scalable NoSQL Data Store
-  **ElastiCache**
In-Memory Cache
-  **Redshift**
Managed Petabyte-Scale Data Warehouse Service





Developer Tools

-  **CodeCommit**
Store Code in Private Git Repositories
-  **CodeDeploy**
Automate Code Deployments
-  **CodePipeline**
Release Software using Continuous Delivery

Management Tools

-  **CloudWatch**
Monitor Resources and Applications
-  **CloudFormation**
Create and Manage Resources with Templates
-  **CloudTrail**
Track User Activity and API Usage
-  **Config**
Track Resource Inventory and Changes
-  **OpsWorks**
Automate Operations with Chef
-  **Service Catalog**
Create and Use Standardized Products
-  **Trusted Advisor**
Optimize Performance and Security






Security & Identity

-  **Identity & Access Management**
Manage User Access and Encryption Keys
-  **Directory Service**
Host and Manage Active Directory
-  **Inspector** **PREVIEW**
Analyze Application Security
-  **WAF**
Filter Malicious Web Traffic








Internet of Things

-  **AWS IoT** **BETA**
Connect Devices to the cloud

Mobile Services

-  **Mobile Hub** **BETA**
Build, Test, and Monitor Mobile apps
-  **Cognito**
User Identity and App Data Synchronization
-  **Device Farm**
Test Android, Fire OS, and iOS apps on real devices in the Cloud
-  **Mobile Analytics**
Collect, View and Export App Analytics
-  **SNS**
Push Notification Service

Application Services

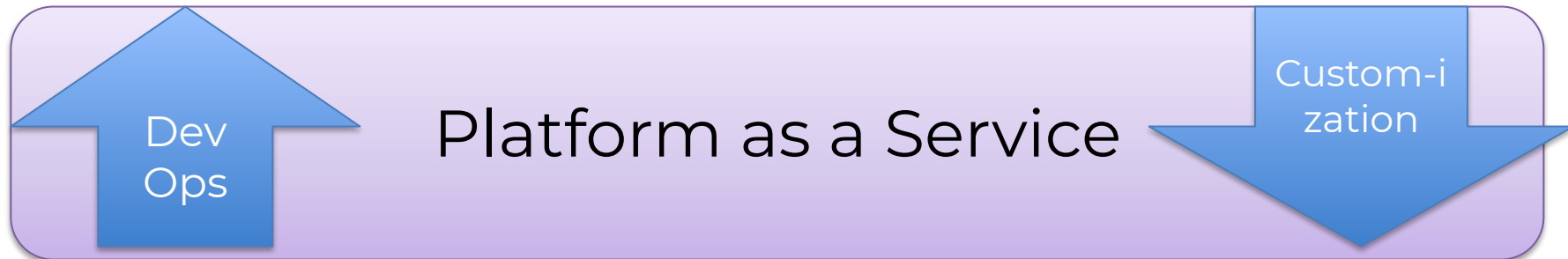
-  **API Gateway**
Build, Deploy and Manage APIs
-  **AppStream**
Low Latency Application Streaming
-  **CloudSearch**
Managed Search Service
-  **Elastic Transcoder**
Easy-to-use Scalable Media Transcoding
-  **SES**
Email Sending Service
-  **SQS**
Message Queue Service
-  **SWF**
Workflow Service for Coordinating Application Components

Enterprise Applications



IaaS, PaaS, SaaS

Software as a Service



Infrastructure as a Service

Public IaaS

- Main Infrastructure-as-a-Service options:
 - Amazon AWS (largest market share)
 - Microsoft Azure
 - Google GCE
 - Alibaba
 - DigitalOcean, Linode, etc



Gartner's view

2019 Magic Quadrant



Market Overview

The market for cloud IaaS is maturing, but revenue is growing unabated. Gartner projects revenue in the cloud IaaS market to increase to \$81.5 billion by 2022, up from \$41.4 billion in 2019. But most of the enterprise interest and revenue are currently directed toward two providers: AWS and Microsoft. The market views both AWS and Microsoft as being general-purpose providers capable of supporting a broad range of workloads. Google is making steady progress in terms of enterprise adoption, but it remains in a distant third place in terms of overall annual revenue and interest among Gartner's enterprise clients. All other vendors in this market are forced to focus on regional dominance or niche workloads given the momentum of AWS and Microsoft, and the scale at which they operate. Examples of regional and niche-focused vendors are Alibaba and Oracle. Alibaba dominates the market for cloud IaaS in China, and Oracle is, naturally, mostly focused on Oracle workloads as it attempts to scale in the process of rebooting its cloud endeavors. Lastly, IBM remains in a precarious position due to being slow to improve its cloud IaaS offerings, which are ultimately not competitive with the market leaders.



MACRO PERSPECTIVE

MEGACLOUDS ARE FIGHTING TO BE #1 PLUMBING FOR DIGITAL BUSINESS

Besides a few serious regional players like Alibaba, global enterprises have 3 main marketplace bazaars to choose from to power their digital transformation



PLAYER #1 (CATEGORY LEADER): MOMENTUM AND BRAND NAME

\$17.1 Billion (2017 Revenue Est.)
40% YoY Growth

PRODUCT STRATEGY

The monocloud that's good enough for most things, not amazing for anything. Heading down proprietary path as most services are integrally tied to their public cloud architecture.

GTM STRATEGY

Aggressive enterprise sales: lock-in, land-and-expand.

BIG EXISTENTIAL QUESTION

Amazon can't allocate 30 top PhDs to solve a single problem. Who will hit Amazon in the achilles heel?



PLAYER #2 (FOR NOW): ENTERPRISE HERITAGE

\$6.1 Billion (2017 Revenue Est.)
81% YoY Growth

PRODUCT STRATEGY

Play to internal strengths: Underserved enterprise workloads like legacy Microsoft products, platform and application services for modern enterprise apps.

GTM STRATEGY

Strong enterprise support model.

BIG EXISTENTIAL QUESTION

Will enterprise chops trump Amazon's scale and scope?



Google Cloud Platform

PLAYER #3 (KILLER PRODUCTS): BUT WHERE'S THE ENTERPRISE LOVE?

\$950 Million (2017 Revenue Est.)
75% YoY Growth

PRODUCT STRATEGY

Google shines strength in machine learning, developer tools, and container orchestration (Kubernetes).

GTM STRATEGY

Historically Google hasn't catered to the enterprise with sales & support. They're apparently trying to change this though.

BIG EXISTENTIAL QUESTION

Can Diane Green, Sam Ramji, and the first-class GTM team from Apigee bring Google from enterprise 0 to hero?

Source: Estimates from Bank of America Merrill Lynch's "Server & Enterprise Software: Cloud Wars 9: AI : From faster to smarter powered by ABC." May 8, 2017. Revenue includes PaaS & IaaS.

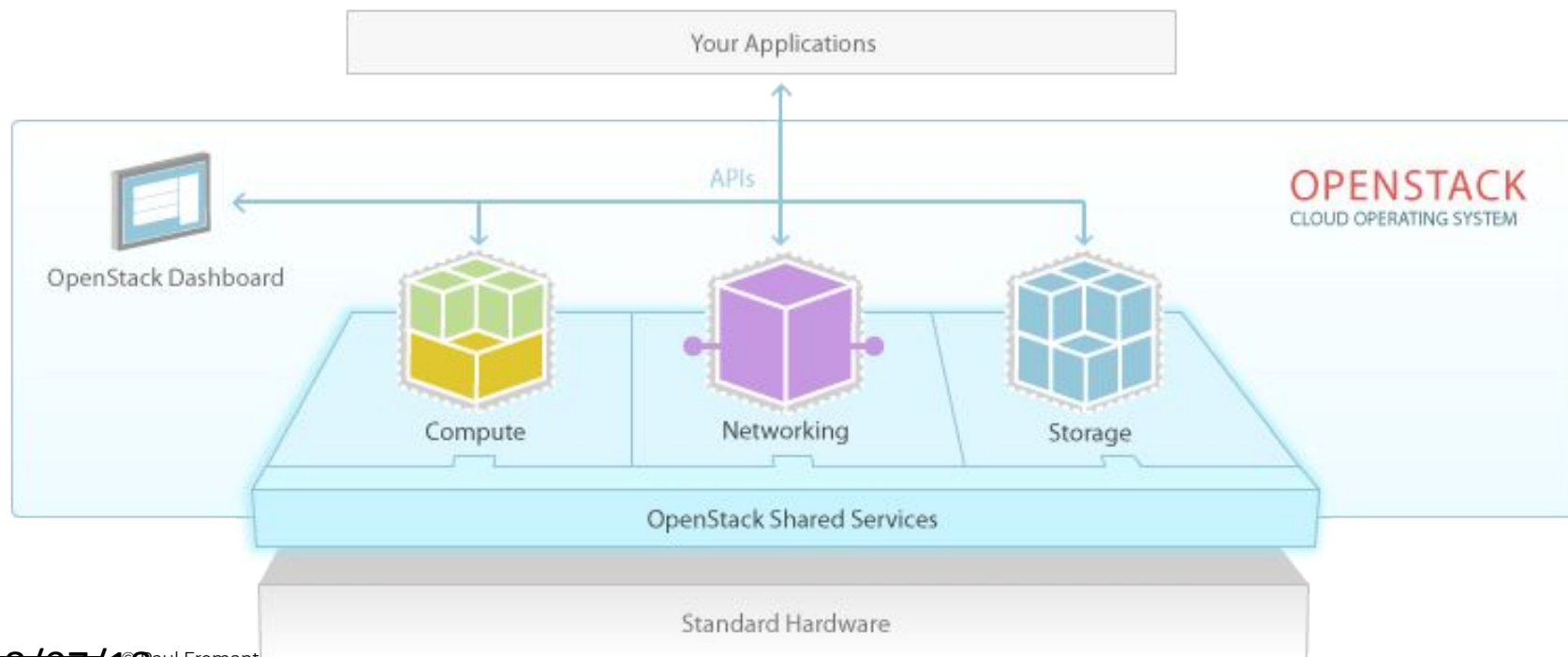
Private IaaS

- OpenStack
- HP Enterprise Eucalyptus
- VMware vSphere / vCloud
- Apache CloudStack



OpenStack

- Compute - *Nova*
- Networking - *Neutron* (formerly Quantum)
- Block Storage – *Cinder*
- Object Storage – *Swift*
- Image Service – *Glance*
- Identity Service - *Keystone*



EC2 / AWS main functions

- EC2 (Elastic Compute Cloud)
 - Instances
 - Servers of various sizes
 - AMIs (Amazon Machine Images)
 - Server images
 - Elastic Block Storage (EBS)
 - Virtualized Hard drives
 - VPC (Virtual Private Cloud)
 - Secure network space
- S3 (Simple Storage Solution)
 - “Buckets” of data
 - Longer term storage of data



Platform-as-a-Service

PaaS

- IaaS is about provisioning
 - machines, disk, network
- PaaS is about provisioning services for developers
 - Hadoop, Spark, JEE containers
 - Databases, Queues, Pub/Sub
 - Cache, Email services, Notifications
- Sort of SaaS for developers



Public PaaS options

- Amazon AWS is the most successful PaaS
 - RDS (Database), DynamoDB
 - ElastiCache (memcache as a service)
 - Elastic Beanstalk (deployment as a service)
 - Simple Notification Service
 - API Gateway
 - CloudSearch
 - Etc, etc, etc!



Other Public PaaS options

- Google App Engine
- Force.com App Cloud
- Heroku
- RedHat OpenShift

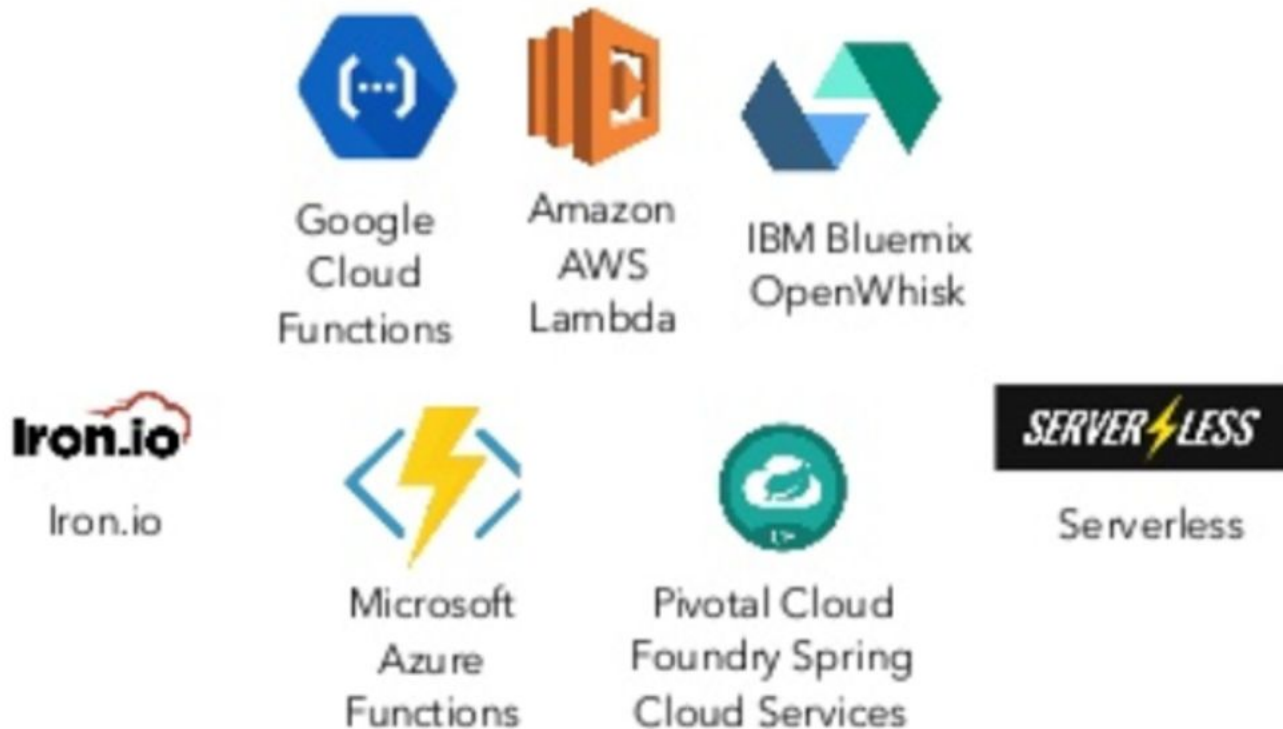


Serverless

- Invisible infrastructure
- Not yet clearly defined, but centering around the concept of Function-as-a-Service
 - E.g. AWS Lambda
 - Deploy a function that is fired when an event happens



Serverless = Functional pipelines



Amazon Lambda

- Serverless computing
 - Runs a small amount of code in response to an event
 - Billed in increments of 100 ms
 - Really pay as you go
 - Scalable (you don't need to define servers)




Back to Amazon AWS

- Three ways to interact
 - Amazon Dashboard (web)
 - APIs and Command-Line
 - Third-party tools
 - ElasticFox, HybridFox
 - Scalr



Amazon EC2 Dashboard

 **AWS** ▾ **Services** ▾ **Edit** ▾

Paul Fremantle ▾ Ireland ▾ Support ▾

EC2 Dashboard

Events

Tags

Reports

Limits

INSTANCES

Instances

Spot Requests

Reserved Instances

Commands

IMAGES

AMIs

Bundle Tasks

ELASTIC BLOCK STORE

Volumes

Snapshots

NETWORK & SECURITY

Security Groups

Elastic IPs

Placement Groups

Key Pairs

Network Interfaces


LOAD BALANCING

Load Balancers

Resources

You are using the following Amazon EC2 resources in the EU West (Ireland) region:

0 Running Instances	0 Elastic IPs
0 Volumes	0 Snapshots
0 Key Pairs	0 Load Balancers
0 Placement Groups	1 Security Groups

 Easily deploy and operate applications - use Chef recipes, manage SSH users, and more. [Try OpsWorks now.](#) [Hide](#)

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 EU West (Ireland) region


Service Health

Service Status:

 EU West (Ireland):
This service is operating normally

Availability Zone Status:

 eu-west-1a:
Availability zone is operating normally

 eu-west-1b:
Availability zone is operating normally

Scheduled Events

EU West (Ireland):

No events

Account Attributes

[Supported Platforms](#)
VPC

[Default VPC](#)
vpc-42fb9527

Additional Information

[Getting Started Guide](#)

[Documentation](#)

[All EC2 Resources](#)

[Forums](#)

[Pricing](#)

[Contact Us](#)

AWS Marketplace

Find **free software trial** products in the AWS Marketplace from the [EC2 Launch Wizard](#). Or try these popular AMIs:

[Tableau Server \(10 users\)](#)

Provided by Tableau

Rating ★★★★★

Pay by the hour for Tableau software and AWS usage

[View all Business Intelligence](#)

[SAP HANA One 244GiB](#)



Main EC2 components

- Instances
 - Your virtual computers
- Volumes (EBS)
 - Disk drives
- Elastic IPs
 - Specific IP address that can be assigned to systems
- Security Groups
 - Sets of firewall rules



More components

- Virtual Private Cloud (VPC)
 - A secure subnet for your instances which can be VPNed to/from your own datacentre
 - Includes/requires an Internet Gateway for creating public services
- Load Balancers
 - Network load-balancing system
- Key pairs
 - Security tokens for managing access
- Route 53
 - Amazon's DNS system



Other cool AWS stuff

- AWS IoT management
- Code tools
- Management
- Analytics
- API Gateway
- Machine Learning



Important

- There is a difference between **stopping** an instance and **terminating**
 - **Stopping**
 - Your instance is stopped, but the disk is still allocated
 - You will be charged for EBS disk
 - **Terminating**
 - Disk will also be removed and you will not be charged



EC2 machine sizes

- Families
 - T2, M4, M3, C4, C3, R3, G2, I2, D2
 - General purpose – T, M
 - Compute – C
 - Memory – R
 - GPU – G
 - IO – I
 - Data – D
- Numbers indicate the “family version”
 - E.g M4 supercedes M3



Amazon instance types (subset)

Instance Type	vCPU	Memory	Storage	Networking Performance	Physical Processor	Clock Speed	EBS OPT	Enhance Networking
t2.micro	1	1	EBS Only	Low to Moderate	Xeon family	Up to 3.3	-	-
t2.large	2	8	EBS Only	Low to Moderate	Xeon family	Up to 3.0	-	-
m4.large	2	8	EBS Only	Moderate	Xeon E5-2676 v3	2.4	Yes	Yes
m4.10xlarge	40	160	EBS Only	10 Gigabit	Xeon E5-2676 v3	2.4	Yes	Yes
c4.large	2	3.75	EBS Only	Moderate	Xeon E5-2666 v3	2.9	Yes	Yes
c4.8xlarge	36	60	EBS Only	10 Gigabit	Xeon E5-2666 v3	2.9	Yes	Yes
g2.2xlarge	8	15	1 x 60 SSD	High	Xeon E5-2670	2.6	Yes	-
g2.8xlarge	32	60	2 x 120 SSD	10 Gigabit	Xeon E5-2670	2.6	-	-
r3.large	2	15.25	1 x 32 SSD	Moderate	Xeon E5-2670 v2	2.5	-	Yes
r3.8xlarge	32	244	2 x 320 SSD	10 Gigabit	Xeon E5-2670 v2	2.5	-	Yes
i2.xlarge	4	30.5	1 x 800 SSD	Moderate	Xeon E5-2670 v2	2.5	Yes	Yes
i2.4xlarge	16	122	4 x 800 SSD	High	Xeon E5-2670 v2	2.5	Yes	Yes
i2.8xlarge	32	244	8 x 800 SSD	10 Gigabit	Xeon E5-2670 v2	2.5	-	Yes
d2.xlarge	4	30.5	3 x 2000	Moderate	Xeon E5-2676 v3	2.4	Yes	Yes
d2.8xlarge	36	244	24 x 2000	10 Gigabit	Xeon E5-2676 v3	2.4	Yes	Yes



An alternative – DigitalOcean

[Droplets](#)[Images](#)[Networking](#)[API](#)[Support](#)[Create Droplet](#)

Droplets

[Droplets](#)[Volumes](#)

Looks like you don't have any Droplets.

Fortunately, it's very easy to create one.

[Create Droplet](#)

© Paul Fremantle 2015. This work is licensed under a Creative Commons
Attribution-NonCommercial-ShareAlike 4.0 International License
See <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Other alternatives

- Microsoft Azure
 - Very strong brand around Microsoft toolkit
- Google Cloud
- Many others



Summary

- SaaS/PaaS/IaaS
- IaaS providers
- PaaS providers
- Working with AWS / EC2
- Now to do a lab!

