Cloud Computing and Big Data

Understanding Cloud Services

Oxford University
Software Engineering
Programme
Nov 2015

Contents

- IaaS / PaaS / SaaS
- Using IaaS
- Tools
- Using PaaS

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!

AWS Reference Architectures

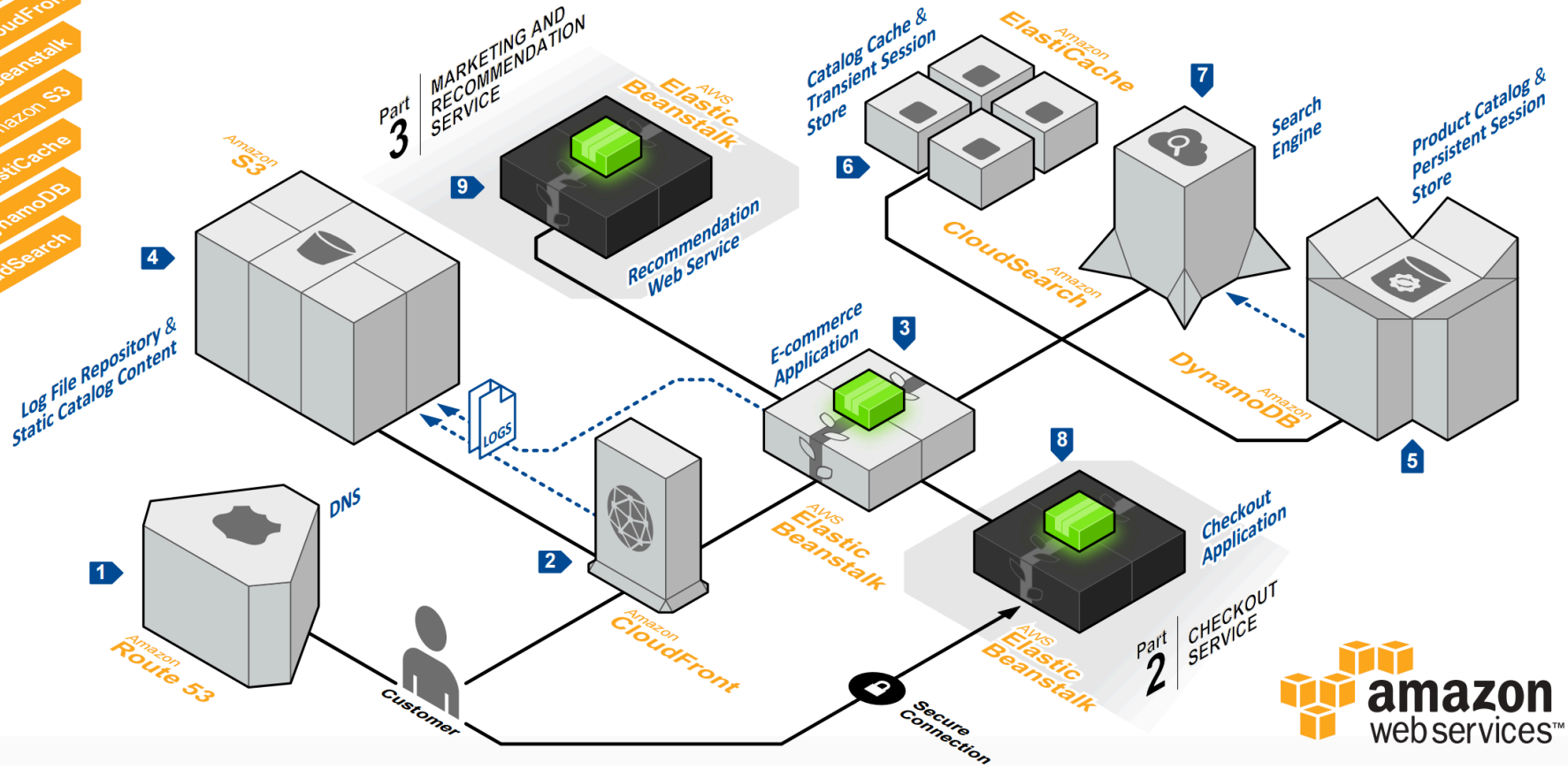
- Amazon Route 53
- Amazon CloudFront
- AWS Elastic Beanstalk
- Amazon S3
- Amazon ElastiCache
- Amazon DynamoDB
- Amazon CloudSearch

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™

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

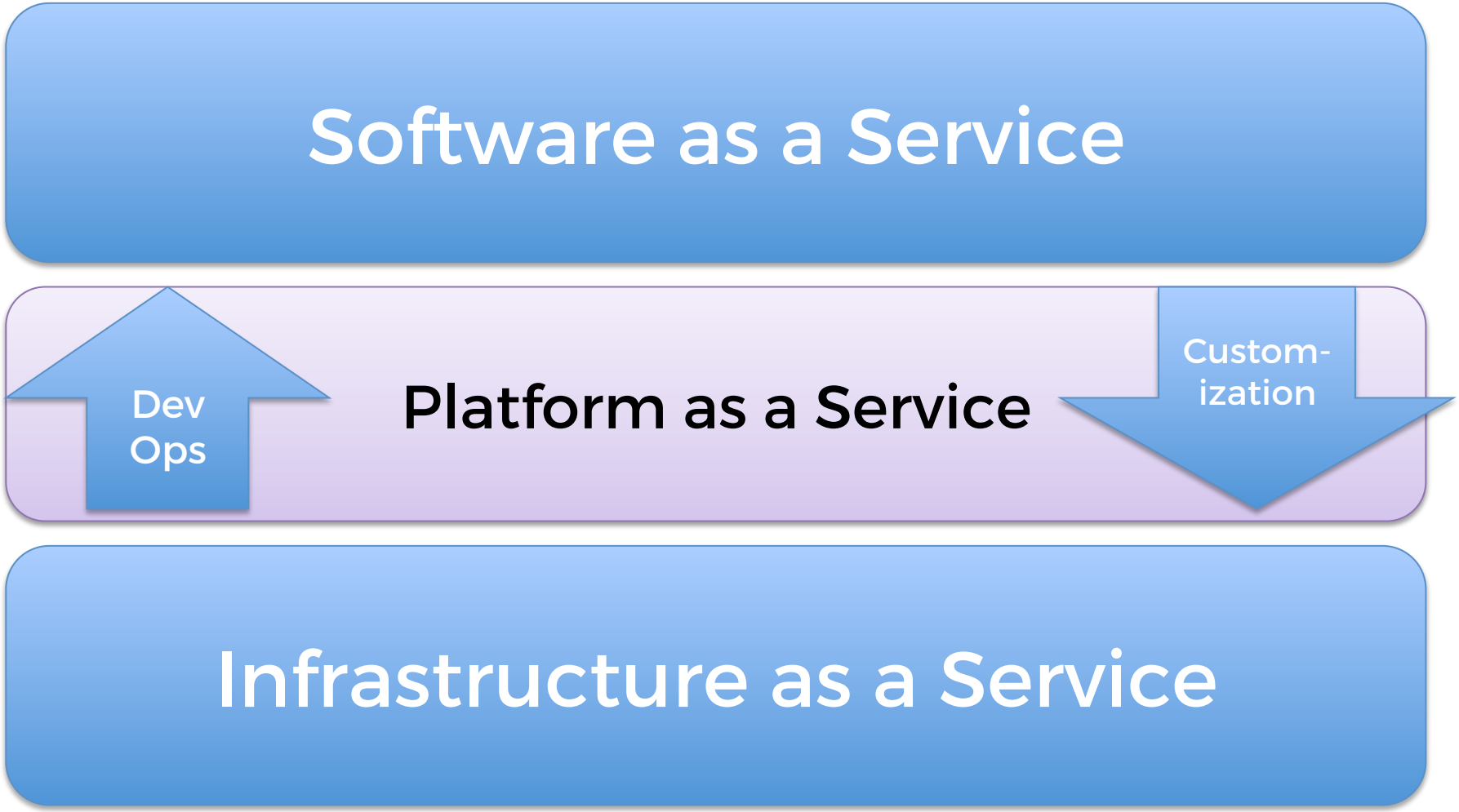
- **Mobile Hub** BETA — Build, Test, and Monitor Mobile apps
- **Cognito** — User Identity and App Data Synchronization
- **Device Farm** — Test Android, Fire OS, and iOS apps on real devices in the Cloud
- **Mobile Analytics** — Collect, View and Export App Analytics
- **SNS** — Push Notification Service

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



Public IaaS

- Main Infrastructure-as-a-Service options:
  - Amazon AWS (largest market share)
  - Google GCE
  - Microsoft Azure
  - IBM Softlayer
  - Rackspace
  - Joyent
  - DigitalOcean

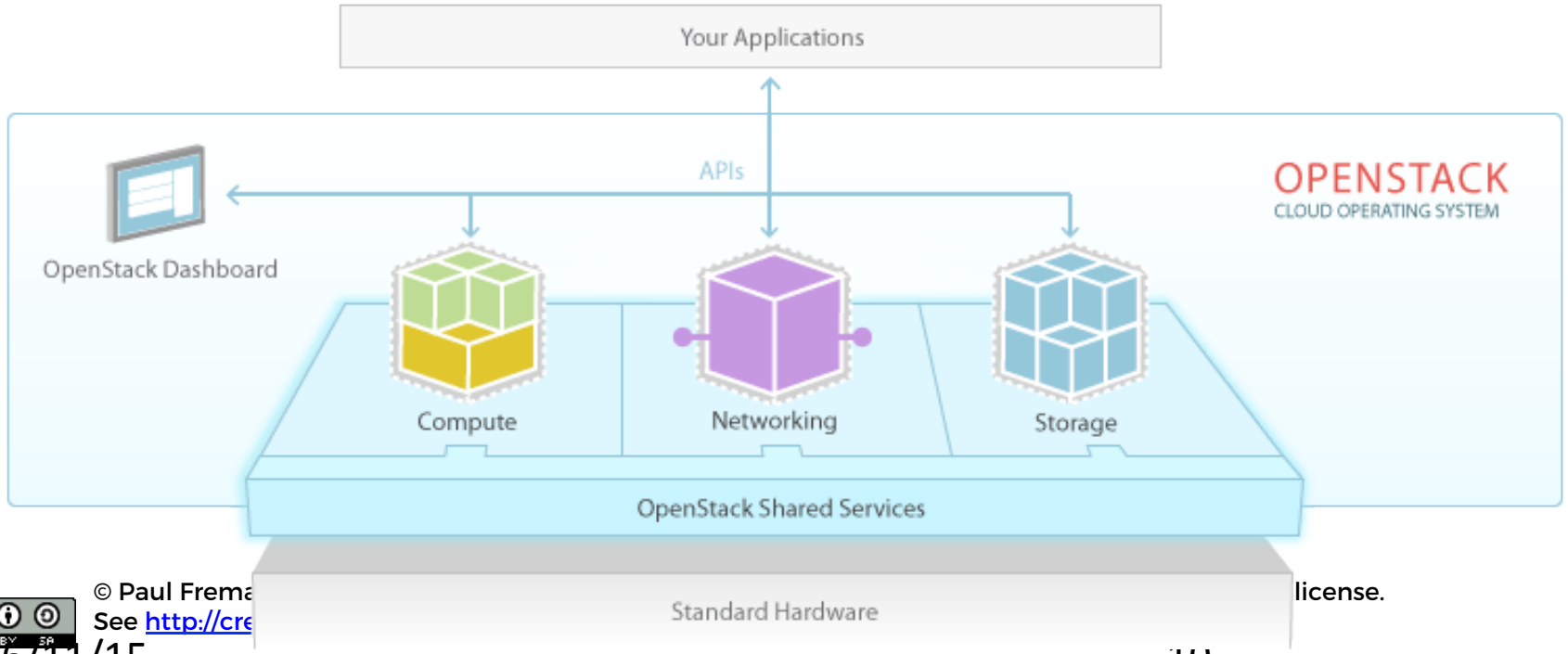
Gartner's view

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 becoming a PaaS
  - RDS (Database), DynamoDB
  - ElastiCache (memcache as a service)
  - Elastic Beanstalk (deployment as a service)
  - Simple Notification Service
  - API Gateway
  - CloudSearch
  - Etc, etc, etc!

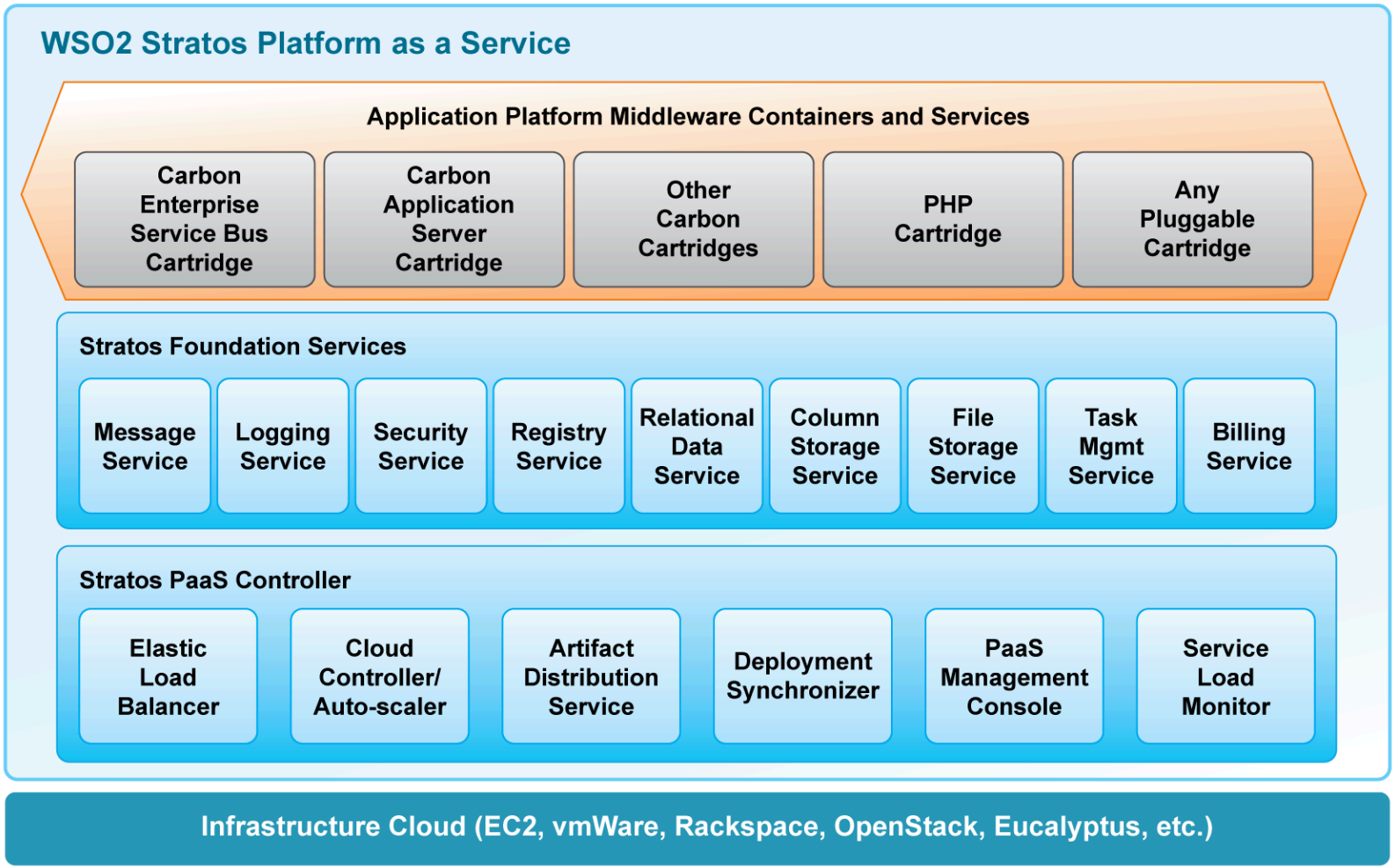
Other Public PaaS options

- Google App Engine
- Force.com App Cloud
  - Heroku
- IBM Bluemix
- RedHat OpenShift
- WSO2 Cloud
- EngineYard

Private PaaS options

- Pivotal CloudFoundry
  - The market leader
- Redhat OpenShift
- Apache Stratos

Platform as a Service model



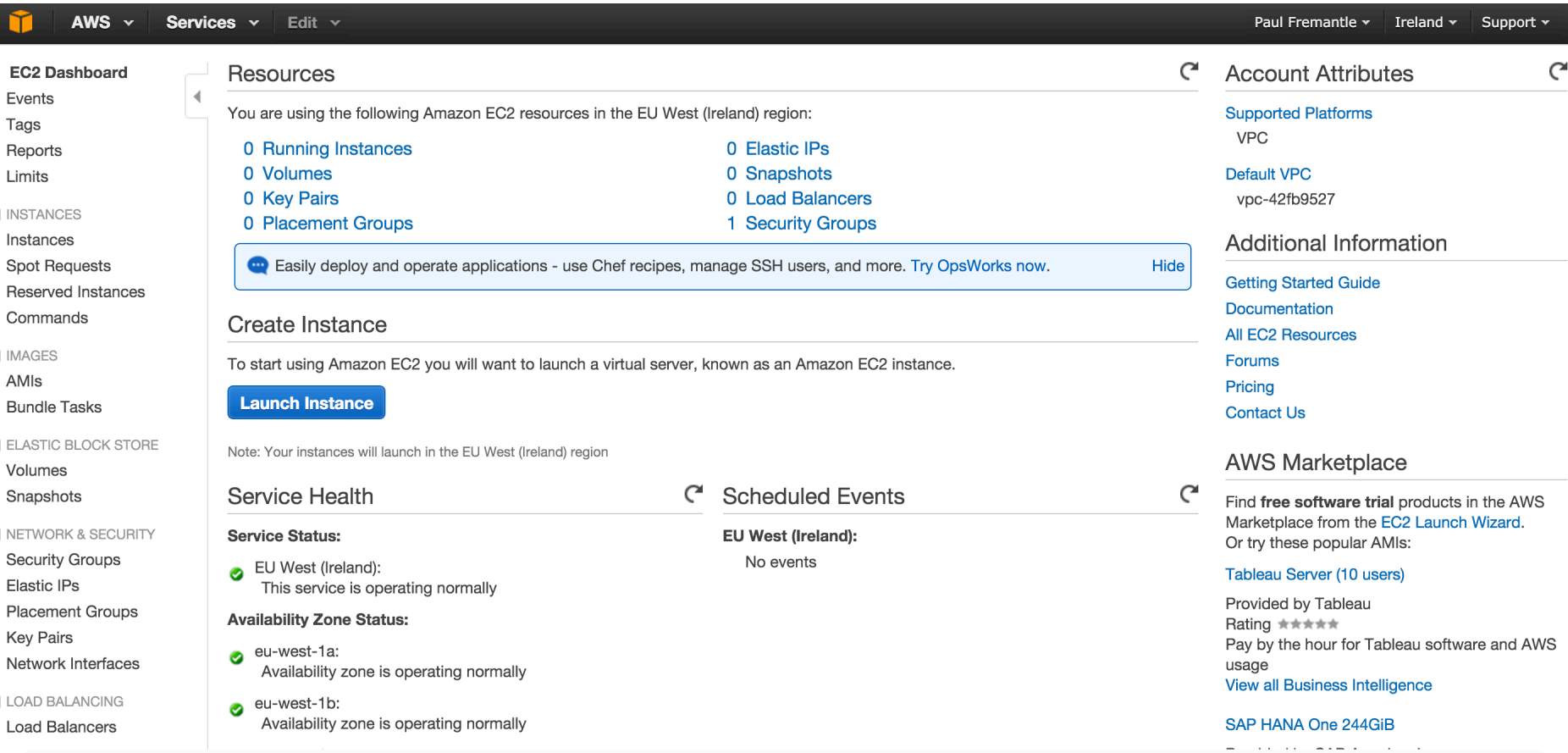
PaaS and Containers

To be covered more later!

Back to Amazon AWS

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

Amazon EC2 Dashboard



Amazon EC2 demo

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

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

Summary

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