

Cloud Computing

# **Cloud Computing Technologies and Types**

Dell Zhang

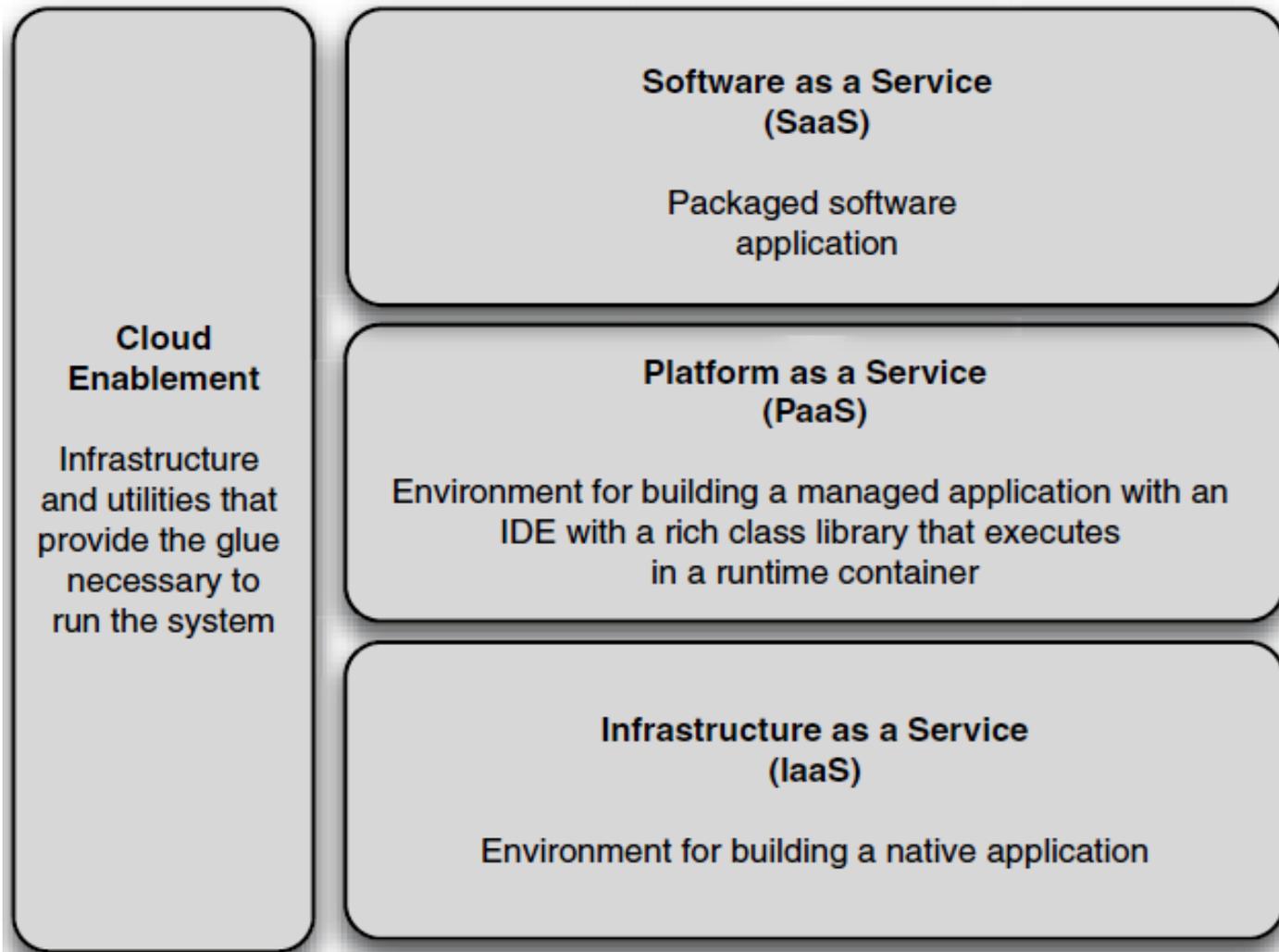
Birkbeck, University of London

2018/19

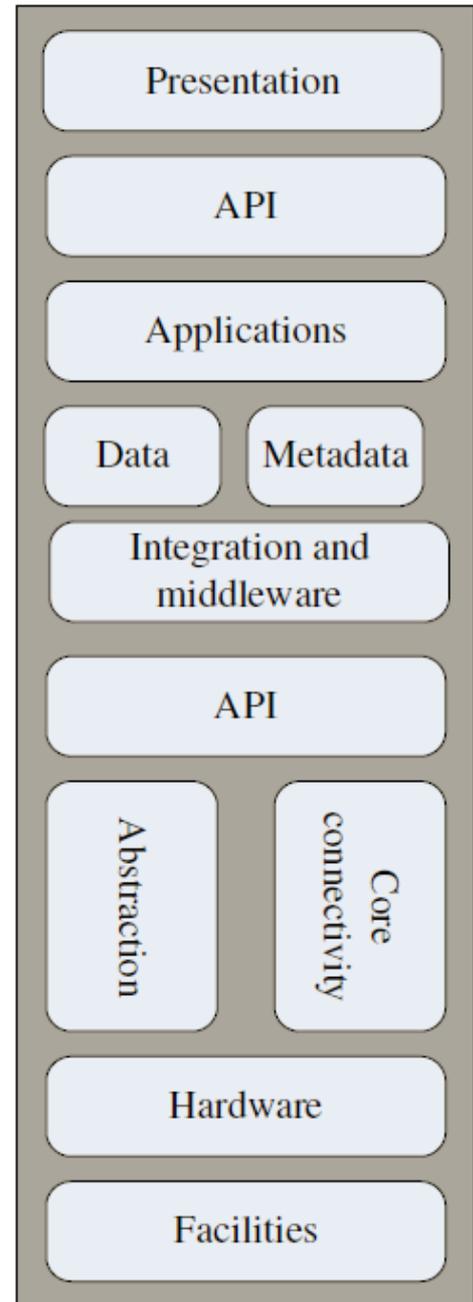
# The Technological Underpinnings of Cloud Computing

- Data centres (see former lectures)
- Virtualisation (see former lectures)
- RESTful APIs (see later lectures)
- Cloud storage (see former lectures)
- NoSQL databases (see later lectures)
- Elasticity (see former lectures)

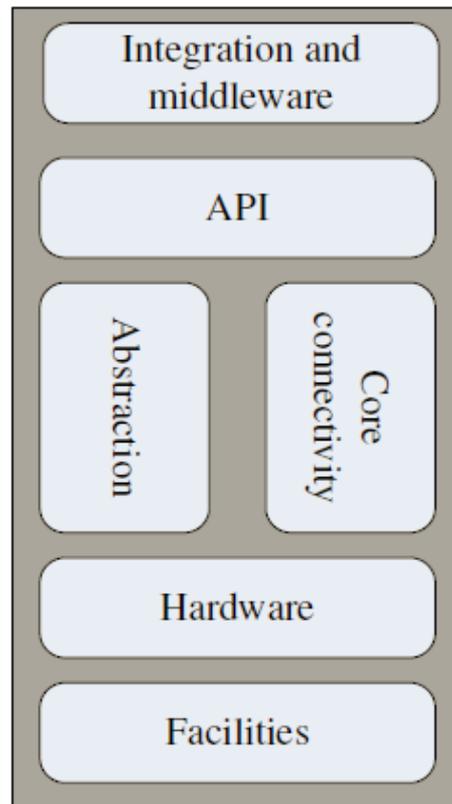
# Understanding the Different Classifications of Clouds



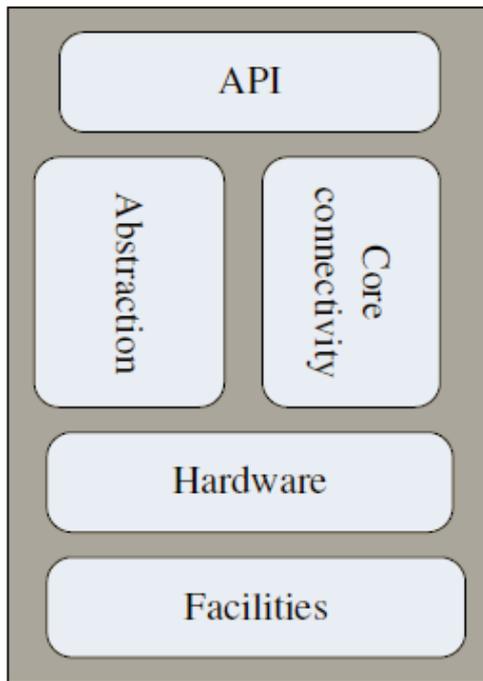
Software as a Service



Platform as a Service



Infrastructure as a Service



# X as a Service

- Infrastructure as a Service (IaaS)
  - Why buy machines when you can rent cycles?
  - Examples: Amazon EC2, Rackspace
- Platform as a Service (PaaS)
  - Give me nice API and take care of the maintenance, upgrades, ...
  - Example: Google App Engine (GAE)
- Software as a Service (SaaS)
  - Just run it for me!
  - Example: Gmail, Salesforce's Online CRM

# laaS

- An IaaS supplies virtual machine images of different operating system flavours.
  - These images can be tailored by the developer to run any custom or packaged application.
  - The user can bring online and use instances of these virtual machine images when needed.

<b>L</b>	Linux	Operating system
<b>A</b>	Apache	Web server
<b>M</b>	MySQL	Relational database
<b>P</b>	PHP	Server side of website

# PaaS

- A PaaS abstracts away the interaction with the bare metal of the system, and thus lets software developers concentrate specifically on writing the application.
  - You don't need to directly administer the virtual operating systems.
  - This simplification generally comes at the cost of less flexibility and the requirement to code in the specific languages supported.

# SaaS

- An SaaS provides software applications that are available on an on-demand basis.



# Serverless Computing

- In the majority of [cloud computing] cases,
  - the user must *deploy* VMs (either directly or indirectly) to support these capabilities and *delete* the VMs when they are no longer needed.
- At times, however, this overhead is unacceptable.
  - When you want an action to take place in response to a relatively rare event, the cost of keeping a VM running continuously so that a program can wait for the event may be unacceptably high.

# Serverless Computing

- For example
  - the user may wish to perform some bookkeeping when a new file is created in a cloud repository
  - The user may wish to receive a notification when an important event occurs

# Serverless Computing

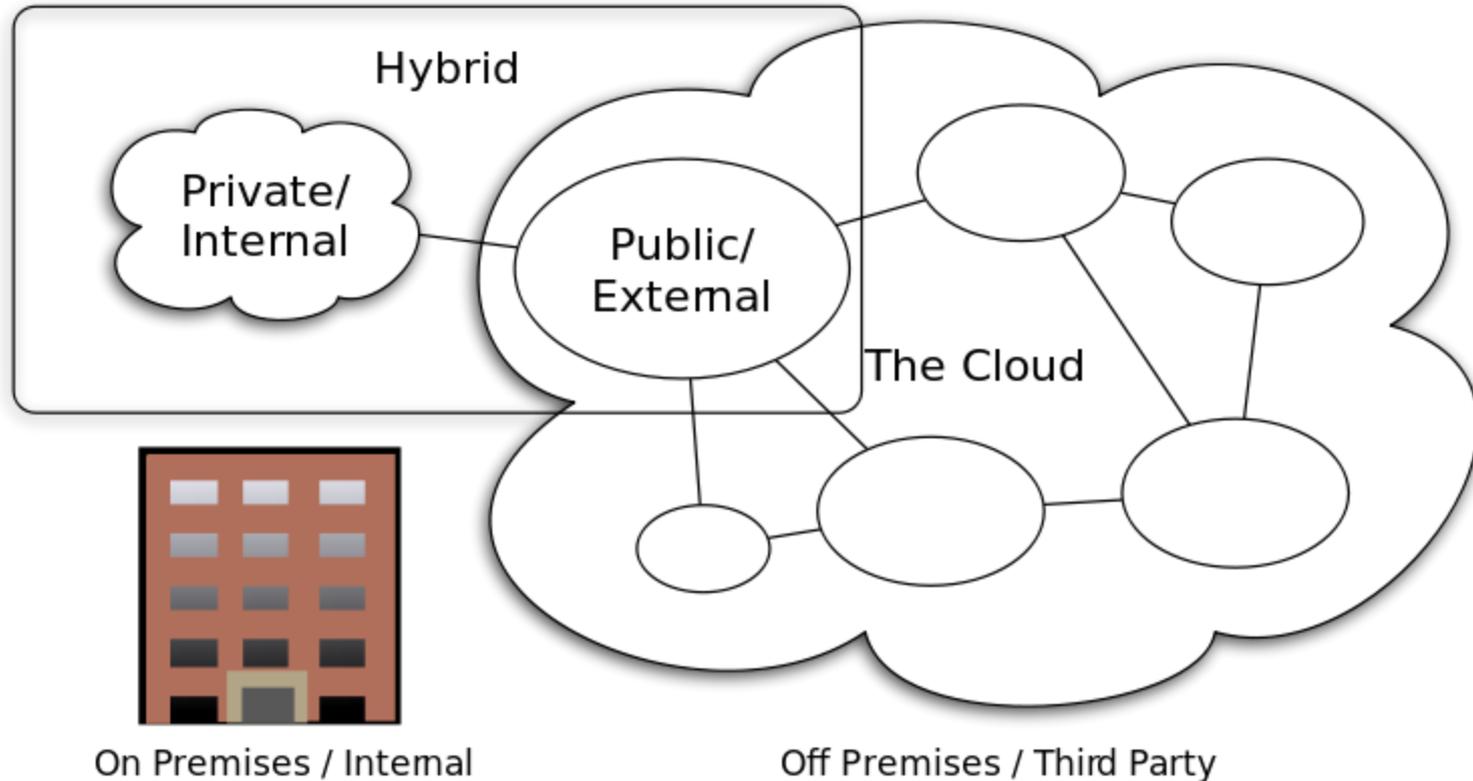
- In *serverless computing*,
  - the user provides a simple function to be executed under certain conditions;
  - the cloud provider keeps a set of machines running to execute these functions on the user's behalf;
  - the user is charged only for the execution of the task, not for maintaining the servers.

# FaaS

- Most serverless vendors offer compute runtimes aka Function as a Service (FaaS) platforms, which *execute application logic but do not store data*.
  - AWS Lambda, introduced by Amazon in 2014, was the first public cloud infrastructure vendor with an abstract serverless computing offering.
  - Google Cloud Functions
  - Microsoft Azure Functions



# Cloud Types



## Cloud Computing Types

# Cloud Types

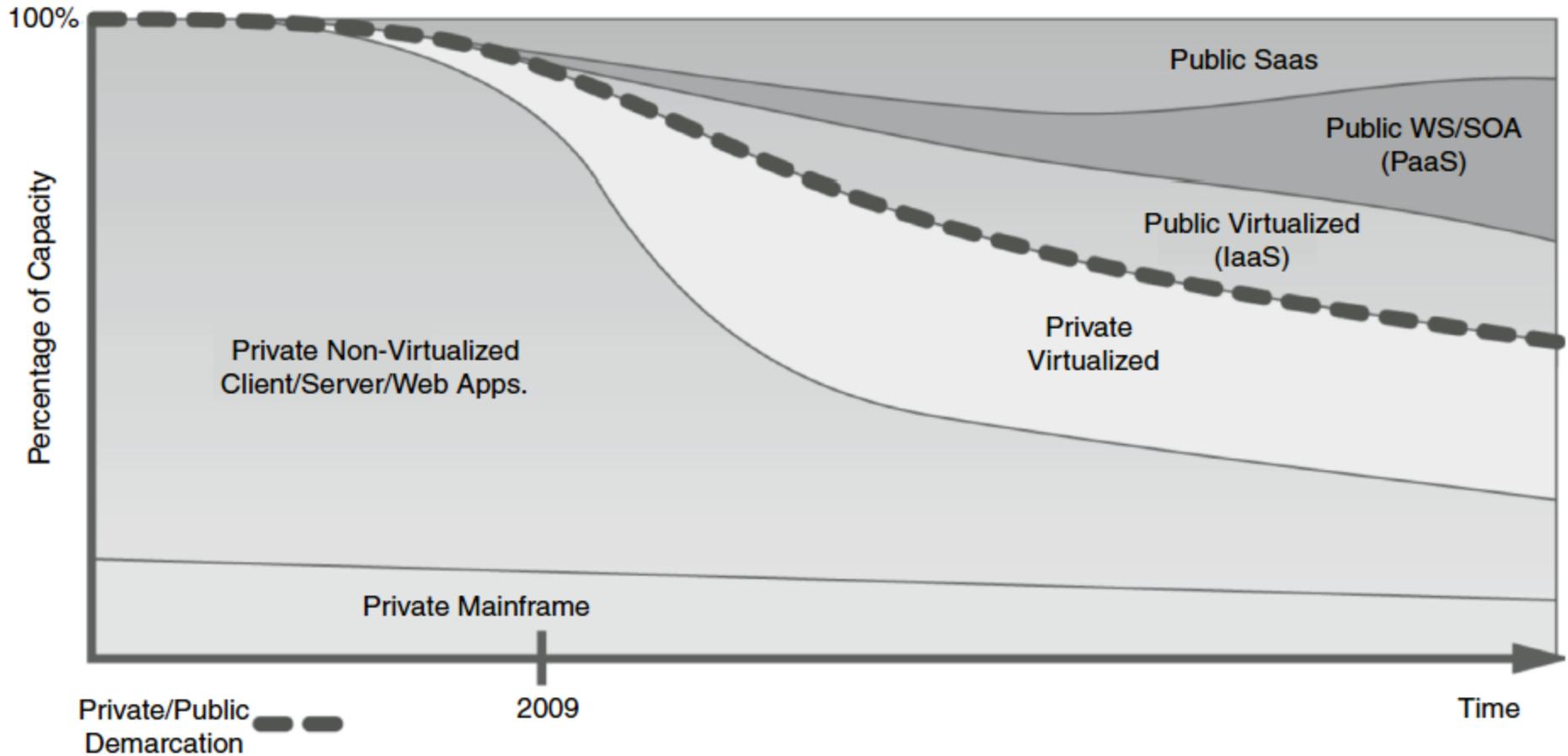
- **Public Cloud**
  - The cloud made available to the general public by a service provider, either free or offered on a pay-per-use model.
- **Private Cloud**
  - The cloud operated solely for a single organization and not shared with others.
- **Community Cloud**
  - The cloud shared between several organizations from a specific community with common concerns.
  - such as ...?

# Cloud Types

- Hybrid Cloud
  - A composition of two or more different types of clouds.
  - Why?
    - (1) The organisation may want to host some critical, secure applications in their private cloud, but use a public cloud for the not-so-critical applications.
    - (2) Cloud bursting: the organisation uses its own cloud for normal usage, but an external cloud is used for peak loads when excess capacity is needed.

# To Be Private, or Not?

- Private clouds are at small scale
  - Big cost savings are driven by huge volume
- Legacy applications don't cloudify easily
  - You can achieve only marginal improvements without re-architecting them
- On-premises doesn't mean more secure
  - Unless your company spends more money and energy thinking about security than Amazon etc.
- Do what you do best
  - Private clouds will always be many steps behind public clouds in the rate of innovation and optimisation



# Matching Cloud Providers to Your Needs

- Amazon Web Services (AWS)
- Google App Engine (GAE)
- Windows Azure

# AWS

- IaaS + PaaS



# Reference Model

Deployment & Administration

App Services

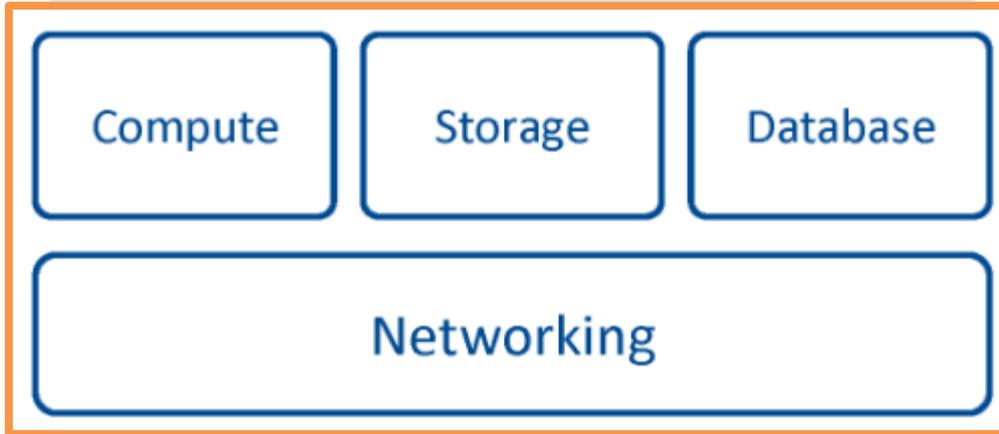
Compute

Storage

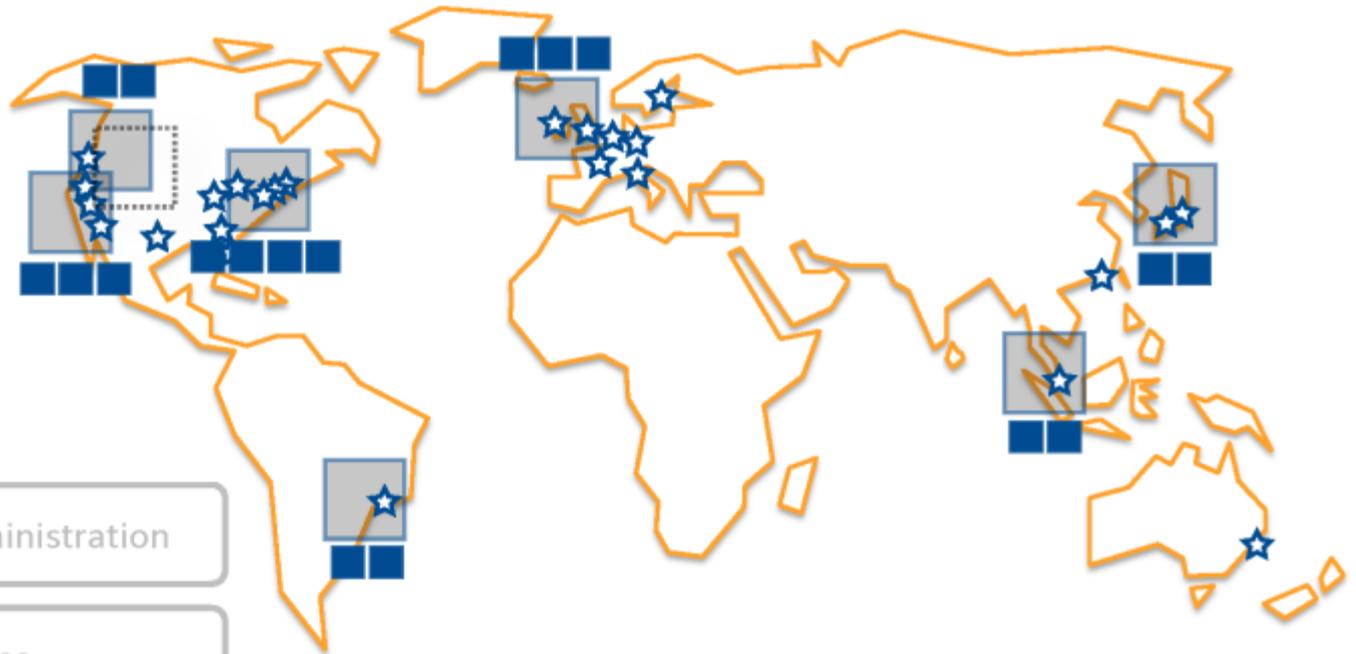
Database

Networking

AWS Global Infrastructure



# Global infrastructure



Deployment & Administration

App Services

Compute

Storage

Database

Networking

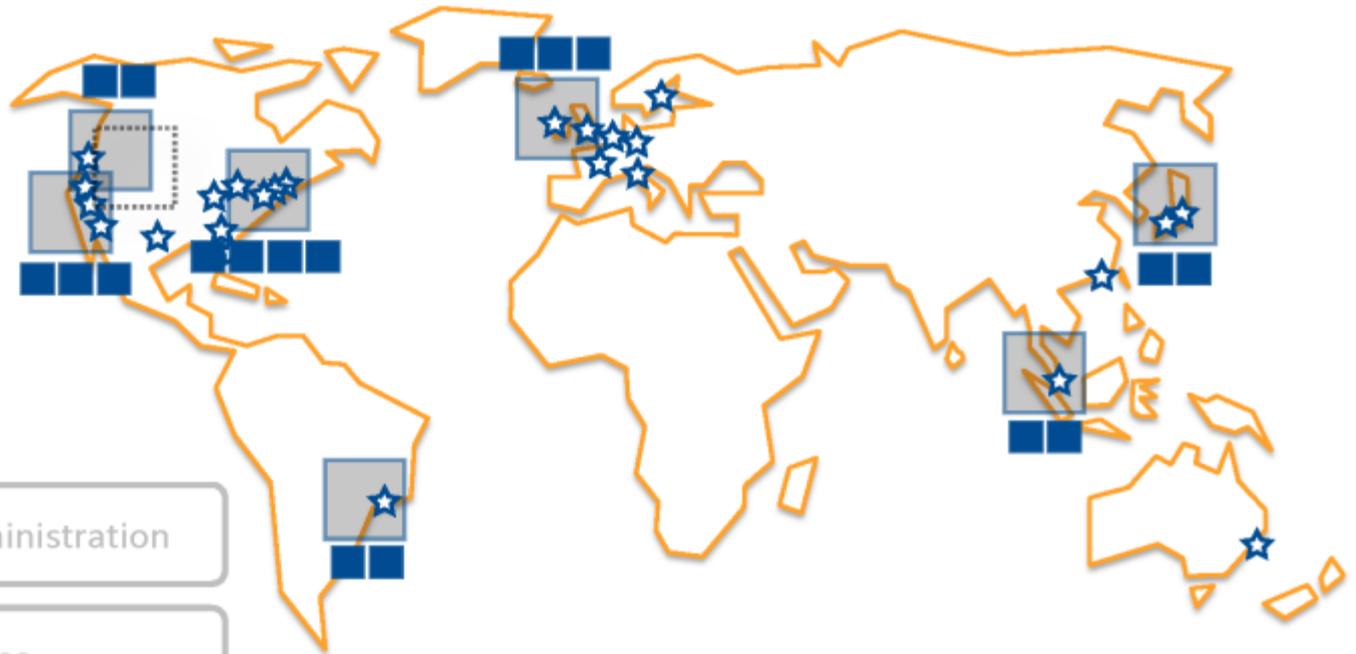
**AWS Global Infrastructure**

## Regions

*An independent collection of AWS resources in a defined geography*

*A solid foundation for meeting location-dependent privacy and compliance requirements*

# Global infrastructure



Deployment & Administration

App Services

Compute

Storage

Database

Networking

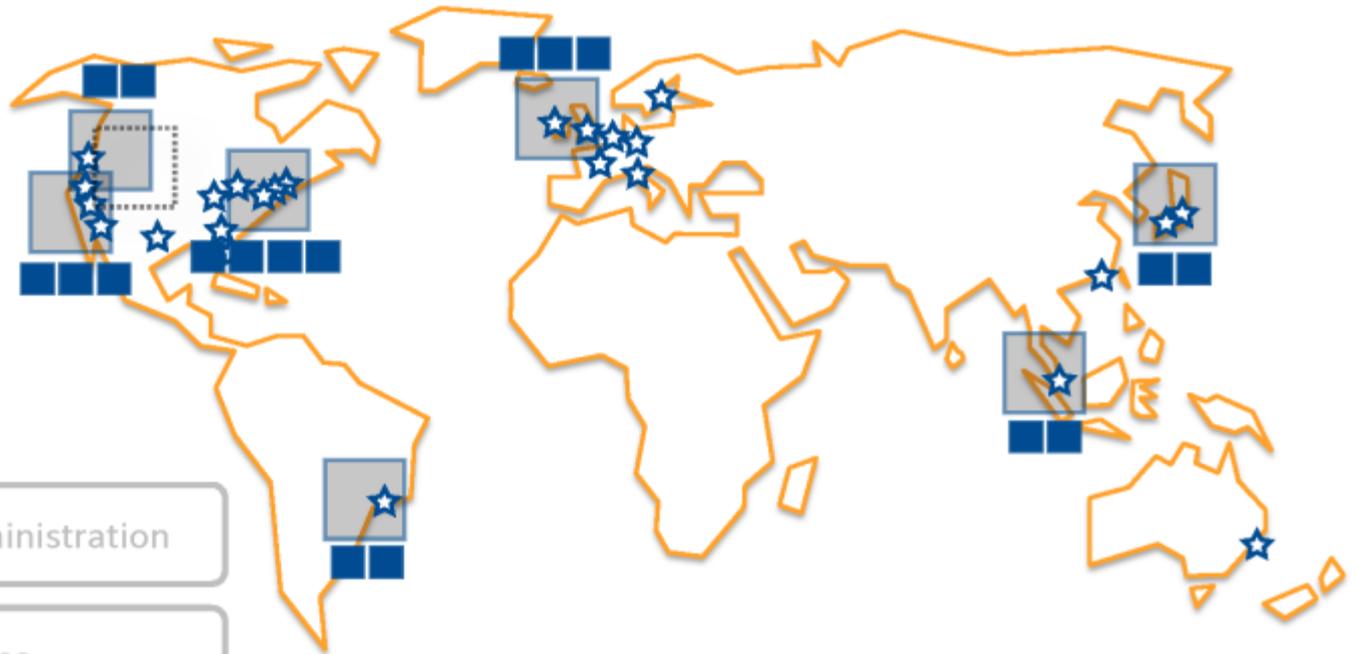
AWS Global Infrastructure

## ■ Availability Zones

*Designed as independent failure zones*

*Physically separated within a typical metropolitan region*

# Global infrastructure



Deployment & Administration

App Services

Compute

Storage

Database

Networking

AWS Global Infrastructure

## ★ Edge Locations

*To deliver content to end users with lower latency*

*A global network of edge locations*

*Supports global DNS infrastructure (Route53) and*

*Cloud Front CDN*

# Networking



Deployment & Administration

App Services

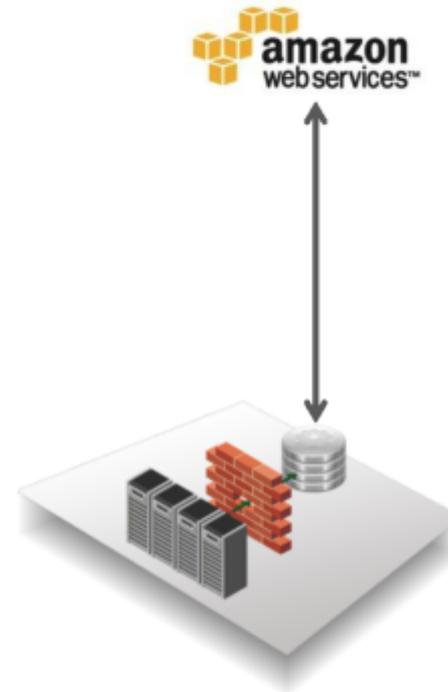
Compute

Storage

Database

**Networking**

AWS Global Infrastructure



## Direct Connect

*Dedicated connection between your IT infrastructure and the AWS datacenters  
Extend your network infrastructure and VLANs into AWS*

# Networking



Deployment & Administration

App Services

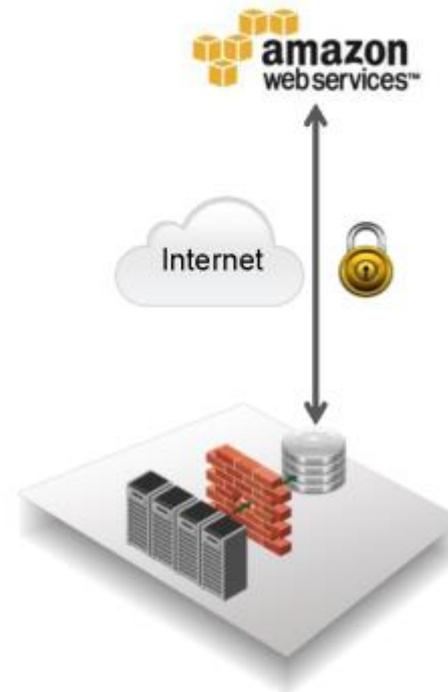
Compute

Storage

Database

Networking

AWS Global Infrastructure



## VPN Connection

*A Hardware VPN connection connects amazon environment to your datacenter*

*Internet Protocol security (IPsec) VPN connection*

*Commonly used hardware supported*

# Networking



Deployment & Administration

App Services

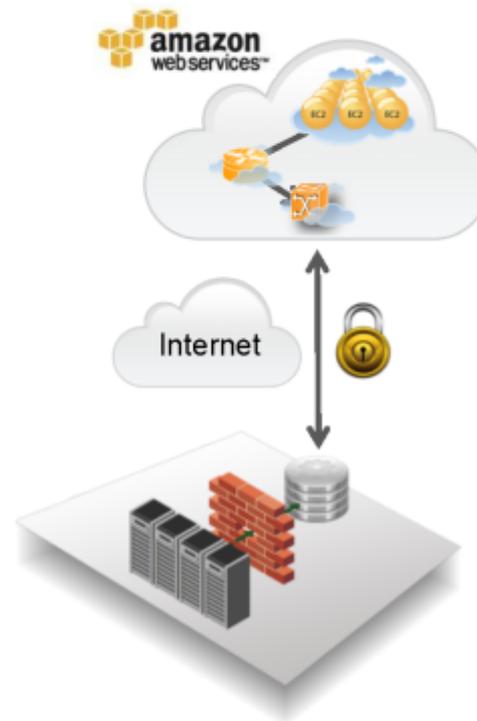
Compute

Storage

Database

**Networking**

AWS Global Infrastructure



## Virtual Private Cloud

*Private, isolated section of the AWS Cloud*

*Launch resources in a virtual network that you define  
complete control over your virtual networking  
environment*

# Networking



Deployment & Administration

App Services

Compute

Storage

Database

**Networking**

AWS Global Infrastructure

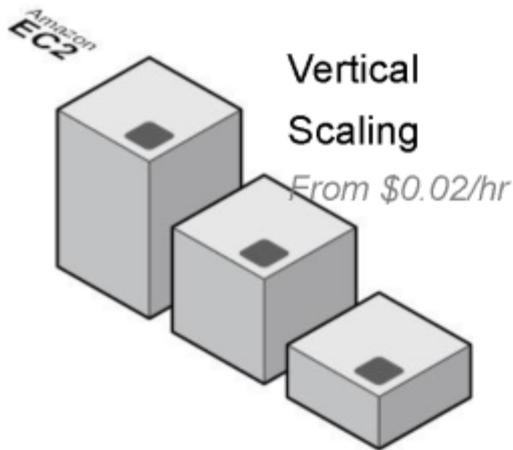
## Route 53

*Highly available and scalable Domain Name System*

*Extremely reliable and cost effective*

Feature	Details
<b>Global</b>	Supported from AWS global edge locations for fast and reliable domain name resolution
<b>Scalable</b>	Automatically scales based upon query volumes
<b>Latency based routing</b>	Supports resolution of endpoints based upon latency, enabling multi-region application delivery
<b>Integrated</b>	Integrates with other AWS services allowing Route 53 to front load balancers, S3 and EC2
<b>Secure</b>	Integrates with IAM giving fine grained control over DNS record access

# Compute



## Elastic Compute Cloud (EC2)

*Basic unit of compute capacity*

*Range of CPU, memory & local disk options*

*13 Instance types available, from micro to cluster compute*



Feature	Details
<b>Flexible</b>	Run windows or linux distributions
<b>Scalable</b>	Wide range of instance types from micro to cluster compute
<b>Machine Images</b>	Configurations can be saved as machine images (AMIs) from which new instances can be created
<b>Full control</b>	Full root or administrator rights
<b>Secure</b>	Full firewall control via Security Groups
<b>Monitoring</b>	Publishes metrics to Cloud Watch
<b>Inexpensive</b>	On-demand, Reserved and Spot instance types
<b>VM Import/Export</b>	Import and export VM images to transfer configurations in and out of EC2

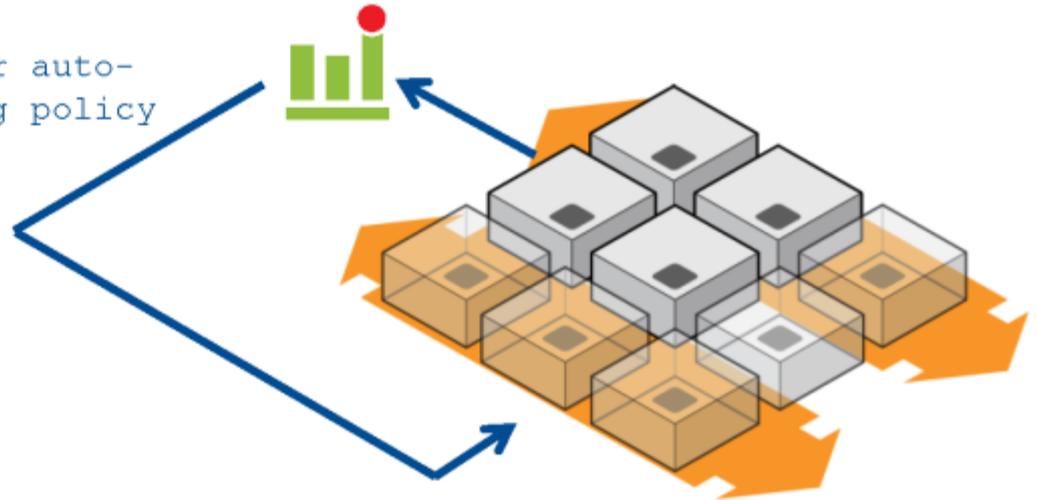
# Compute



Trigger auto-scaling policy



```
as-create-auto-scaling-group MyGroup
  --launch-configuration MyConfig
  --availability-zones eu-west-1a
  --min-size 4
  --max-size 200
```



Deployment & Administration

App Services

Compute

Storage

Database

Networking

AWS Global Infrastructure

## Auto-scaling

*Automatic re-sizing of compute clusters based upon demand*

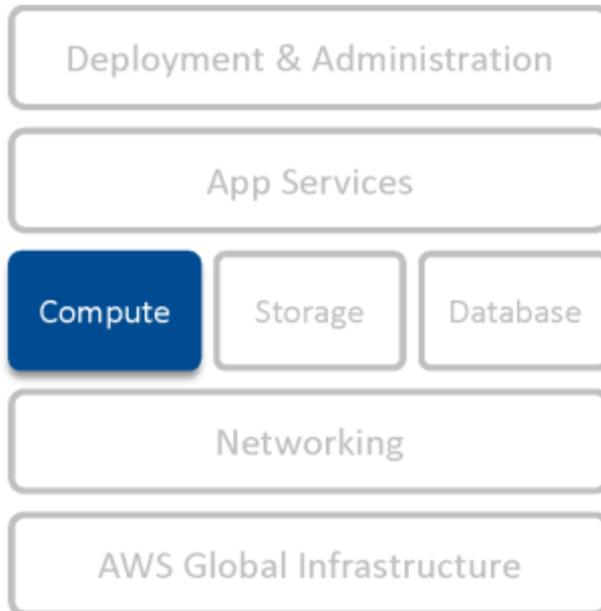
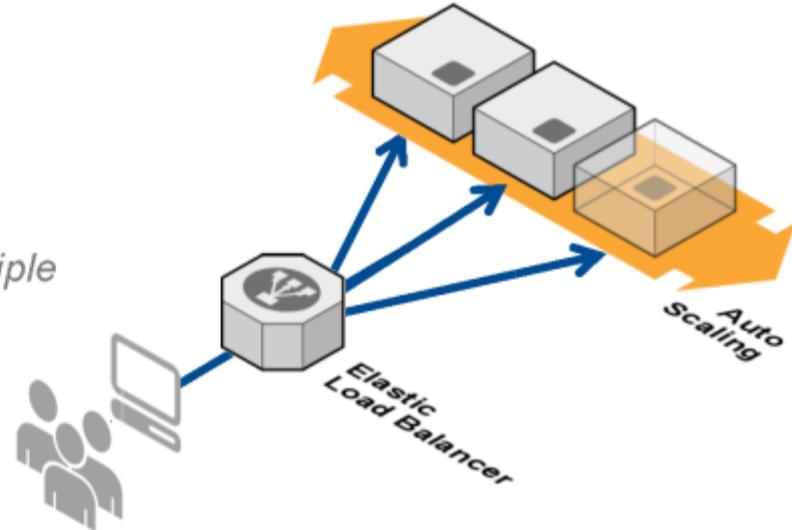
Feature	Details
Control	Define minimum and maximum instance pool sizes and when scaling and cool down occurs
Integrated to CloudWatch	Use metrics gathered by CloudWatch to drive scaling
Instance types	Run auto scaling for on-demand instances and spot. Compatible with VPC

# Compute

## Elastic Load Balancing

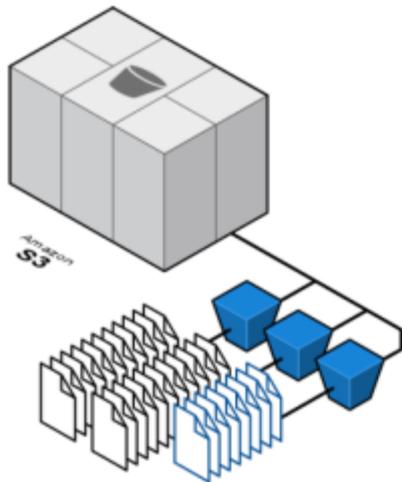
*Create highly scalable applications*

*Distribute load across EC2 instances in multiple availability zones*



Feature	Details
Auto-scaling	Automatically scales to handle request volume
Available	Load balance across instances in multiple availability zones
Health checks	Automatically checks health of instances and takes them in or out of service
Session stickiness	Route requests to the same instance
Secure sockets layer	Supports SSL offload from web and application servers with flexible cipher support
Monitoring	Publishes metrics to Cloud Watch

# Storage



Deployment & Administration

App Services

Compute

Storage

Database

Networking

AWS Global Infrastructure

## S3 - Durable storage, any object

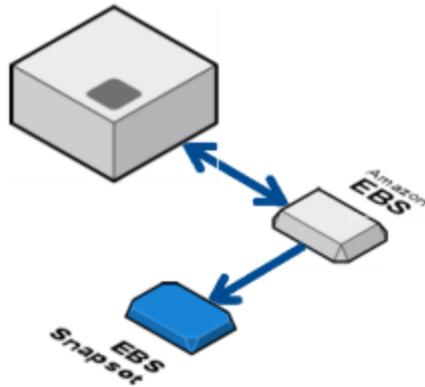
*99.999999999% durability of objects*

*Unlimited storage of objects of any type*

*Up to 5TB size per object*

Feature	Details
Flexible object store	Buckets act like drives, folder structures within
Access control	Granular control over object permissions
Server-side encryption	256bit AES encryption of objects
Multi-part uploads	Improved throughput & control
Object versioning	Archive old objects and version new ones
Object expiry	Automatically remove old objects
Access logging	Full audit log of bucket/object actions
Web content hosting	Serve content as web site with built in page handling
Notifications	Receive notifications on key events
Import/Export	Physical device import/export service

# Storage



## Elastic Block Store

*High performance block storage device*

*1GB to 1TB in size*

*Mount as drives to instances*

Deployment & Administration

App Services

Compute

Storage

Database

Networking

AWS Global Infrastructure

### Feature

### Details

**High performance file system**

Mount EBS as drives and format as required

**Flexible size**

Volumes from 1GB to 1TB in size

**Secure**

Private to your instances

**Available**

Replicated within an Availability Zone

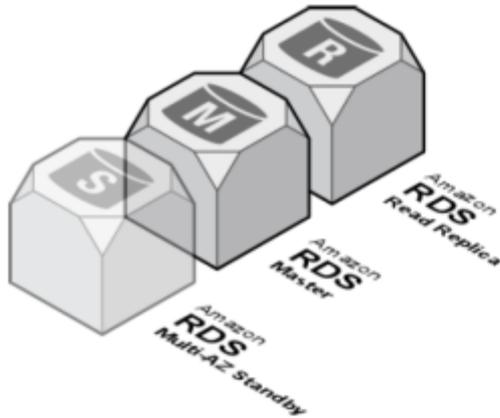
**Backups**

Volumes can be snapshotted for point in time restore

**Monitoring**

Detailed metrics captured via Cloud Watch

# Database



## Relational Database Service

*Database-as-a-Service*

*No need to install or manage database instances*

*Scalable and fault tolerant configurations*



Deployment & Administration

App Services

Compute

Storage

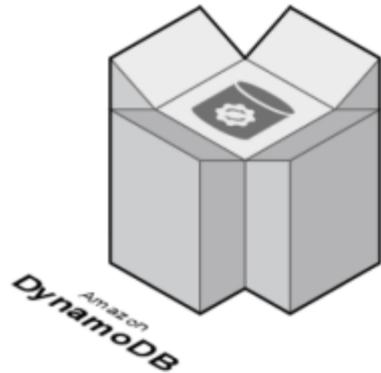
Database

Networking

AWS Global Infrastructure

Feature	Details
Platform support	Create MySQL, SQL Server and Oracle RDBMS
Preconfigured	Get started instantly with sensible default settings
Automated patching	Keep your database platform up to date automatically
Backups	Automatic backups and point in time recovery and full DB backups
Backups	Volumes can be snapshotted for point in time restore
Failover	Automated failover to slave hosts in event of a failure
Replication	Easily create read-replicas of your data and seamlessly replicate data across availability zones

# Database



## DynamoDB

*Provisioned throughput NoSQL database*

*Fast, predictable performance*

*Fully distributed, fault tolerant architecture*

Deployment & Administration

App Services

Compute

Storage

Database

Networking

AWS Global Infrastructure

Feature	Details
<b>Provisioned throughput</b>	Dial up or down provisioned read/write capacity
<b>Predictable performance</b>	Average single digit millisecond latencies from SSD backed infrastructure
<b>Strong consistency</b>	Be sure you are reading the most up to date values
<b>Fault tolerant</b>	Data replicated across availability zones
<b>Monitoring</b>	Integrated to Cloud Watch
<b>Secure</b>	Integrates with AWS Identity and Access Management (IAM)
<b>Elastic MapReduce</b>	Integrates with Elastic MapReduce for complex analytics on large datasets

# Application Services

## Amazon SQS

*Reliable, highly scalable, queue service for storing messages as they travel between instances*

Deployment & Administration

App Services

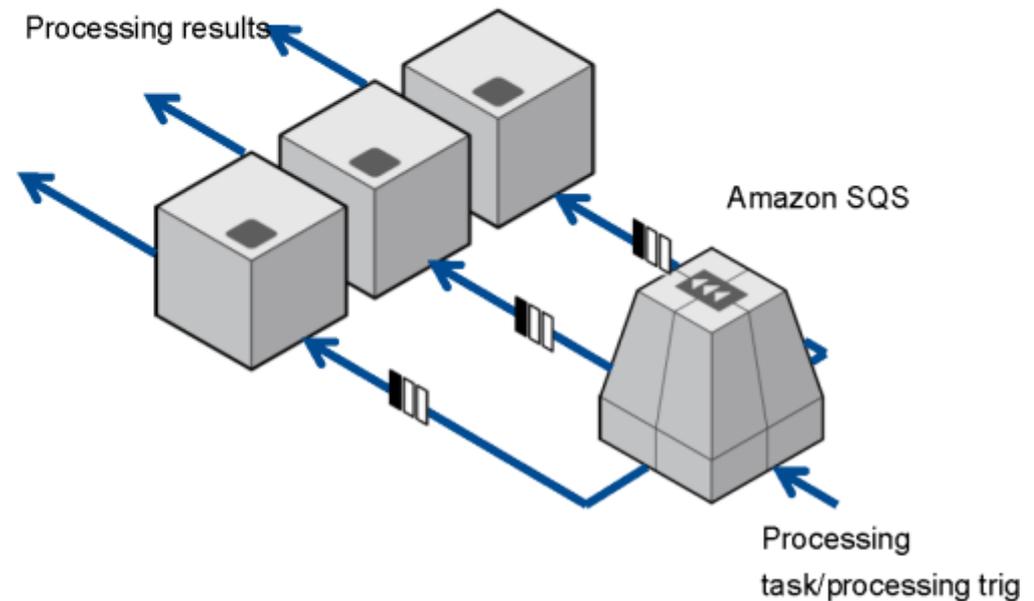
Compute

Storage

Database

Networking

AWS Global Infrastructure



Feature	Details
Reliable	Messages stored redundantly across multiple availability zones
Simple	Simple APIs to send and receive messages
Scalable	Unlimited number of messages
Secure	Authentication of queues to ensure controlled access

# Application Services

## Simple Workflow

*Reliably coordinate processing steps across applications*

*Integrate AWS and non-AWS resources*

*Manage distributed state in complex systems*

Deployment & Administration

App Services

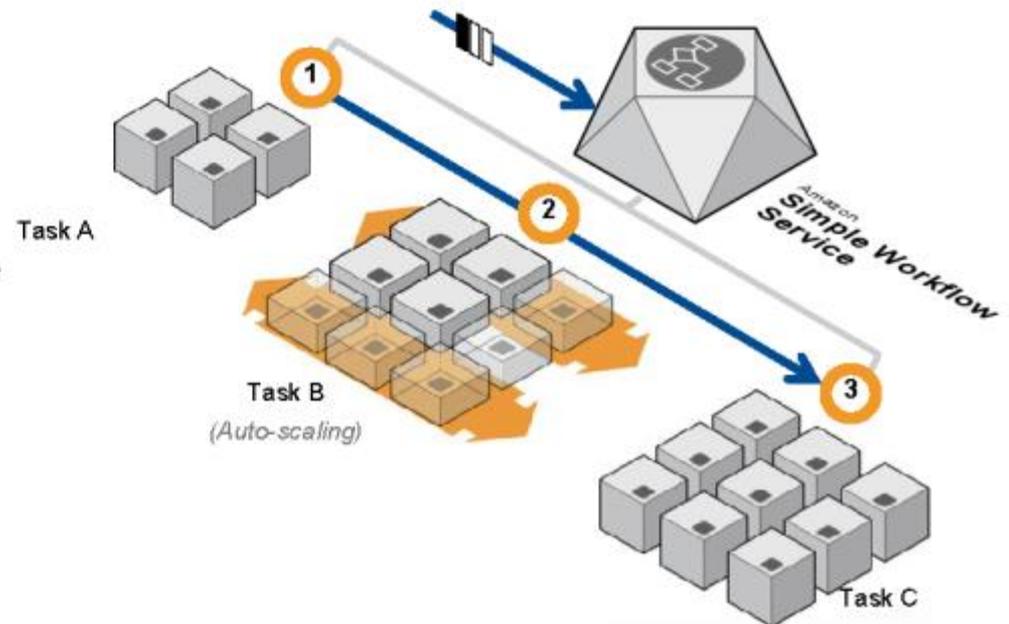
Compute

Storage

Database

Networking

AWS Global Infrastructure



Feature	Details
<b>Process state</b>	Maintain application state across complex workflows in a reliable and available manner
<b>Tracking</b>	Tracks executions and log process for audit purposes
<b>Consistency</b>	Ensures processing tasks are executed and duplicity of events does not occur
<b>Simple</b>	Simple Decider and Task programming model for rapid integration

# Application Services

## Cloud Search

*Elastic search engine based upon Amazon A9 search engine*  
*Fully managed service with sophisticated feature set*  
*Scales automatically*

Deployment & Administration

App Services

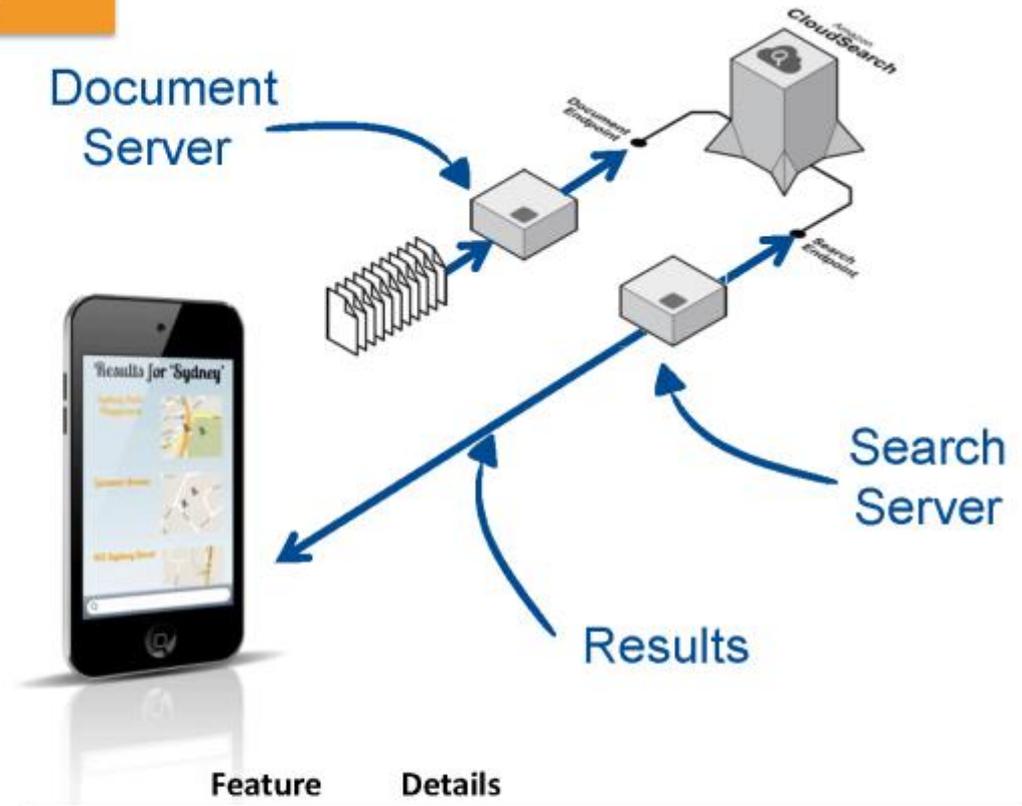
Compute

Storage

Database

Networking

AWS Global Infrastructure



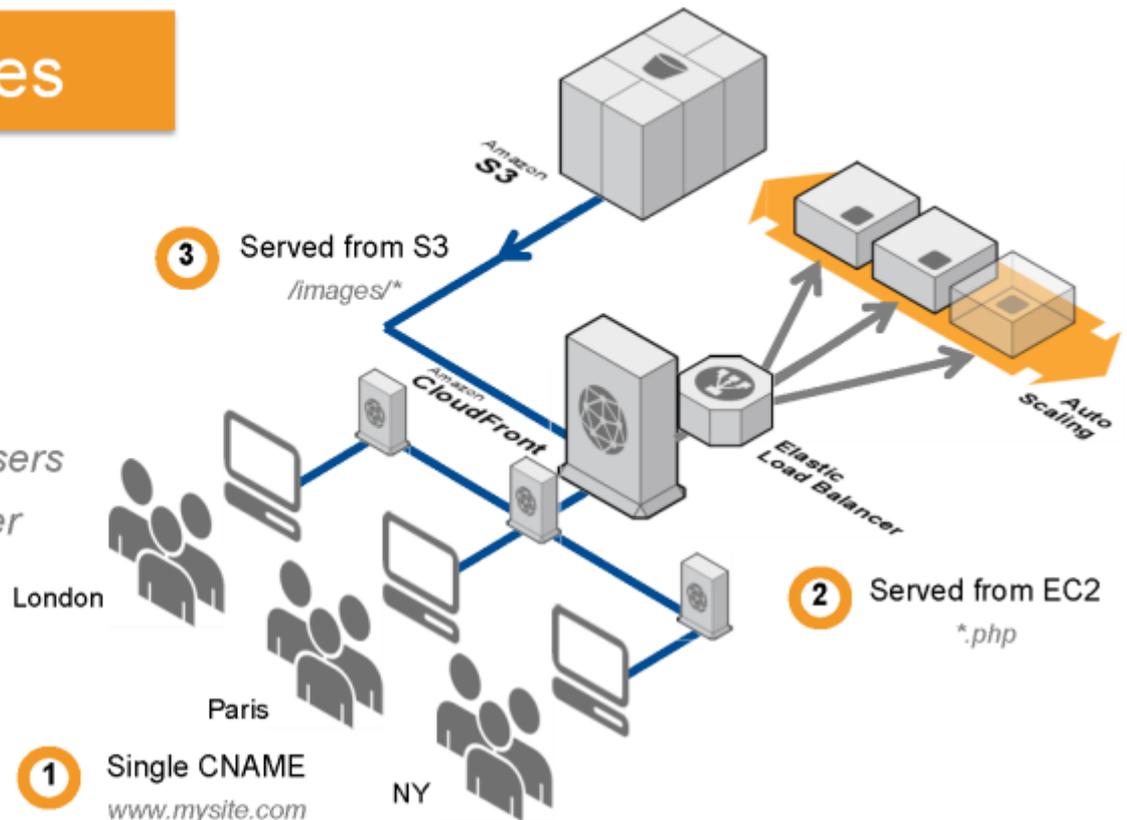
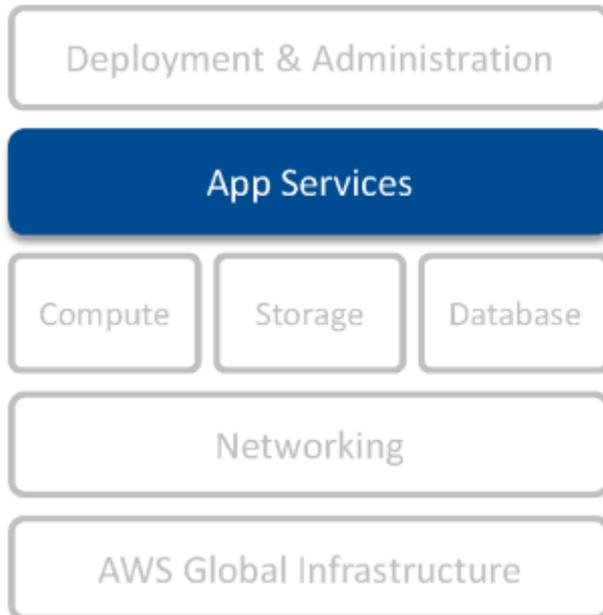
Feature	Details
<b>Auto-scaling</b>	Automatically scales based upon request volumes and data volumes
<b>High performance</b>	In memory operation means consistently low latency for search results
<b>Sophisticated features</b>	Support for faceting, stemming, synonyms, stop words and custom rank expressions
<b>Low cost</b>	Elastic service, pay for what you use

# Application Services

## CloudFront

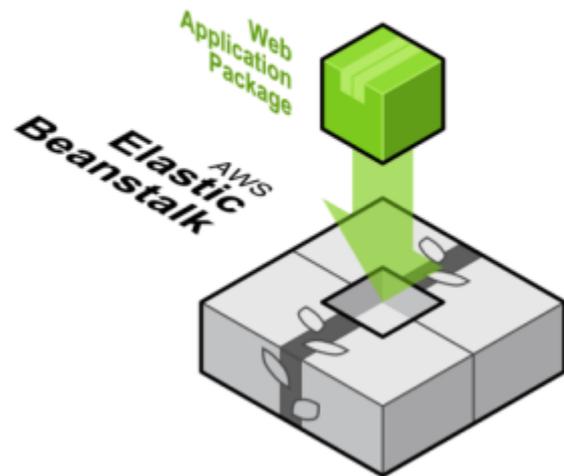
World-wide content distribution network

Easily distribute content to end users with low latency, high data transfer speeds, and no commitments.



Feature	Details
<b>Fast</b>	Multiple world-wide edge locations to serve content as close to your users as possible
<b>Integrated with other services</b>	Works seamlessly with S3 and EC2 origin servers
<b>Dynamic content</b>	Supports static and dynamic content from origin servers
<b>Streaming</b>	Supports rtmp from S3 and includes support for live streaming from Adobe FMS and Microsoft Media Server

# Deployment & Admin



## Elastic Beanstalk

*One-click deployment from Eclipse, Visual Studio and Git*

*Rapid deployment of applications*

*All AWS resources automatically created*



### Deployment & Administration

App Services

Compute

Storage

Database

Networking

AWS Global Infrastructure

Feature	Details
Platform support	Containers for Java, .net and PHP
Resource creation	Creates load balancer, instances, autoscaling and monitoring automatically
Monitoring & Logs	Integrated with Cloud Watch and consolidates server logs
Versioning	Manage versions of applications and easily rollback deployments
Notifications	Receive alerts on key events
Full resource access	Access all underlying AWS resources as necessary

# Deployment & Admin

```
{
  "AWSTemplateFormatVersion" : "2010-09-09",
  "Description" : "This template creates a CloudFormation stack that uses Amazon CloudFront and an Amazon EC2 AMI for Adobe Flash Media Server 4.5 to enable HTTP streaming of
your live event.",
  "Parameters" : {
    "InstanceType" : {
      "Type" : "String",
      "Description" : "The type of Amazon EC2 instance to launch. Valid values are: m1.large, m1.xlarge, m2.large, m2.xlarge, m2.2xlarge, m2.4xlarge, c1.xlarge.",
      "Default" : "m1.large",
      "AllowedValues" : [ "m1.large", "m1.xlarge", "m2.large", "m2.xlarge", "m2.2xlarge", "m2.4xlarge", "c1.xlarge" ],
      "ConstraintDescription" : ""
    }
  }
}
```

## Cloud Formation

*Automate creation of 'stacks' in a repeatable way*

*Scripting framework for AWS resource creation*

### Deployment & Administration

App Services

Compute

Storage

Database

Networking

AWS Global Infrastructure

Feature	Details
Platform support	Support for AWS resources from EC2 to IAM
Resource creation	Creates AWS resources behind the scenes and reports on progress
Declarative	Specify stacks in JSON format and source control your environments
Customizable	Drive stack creation with parameters

# Deployment & Admin

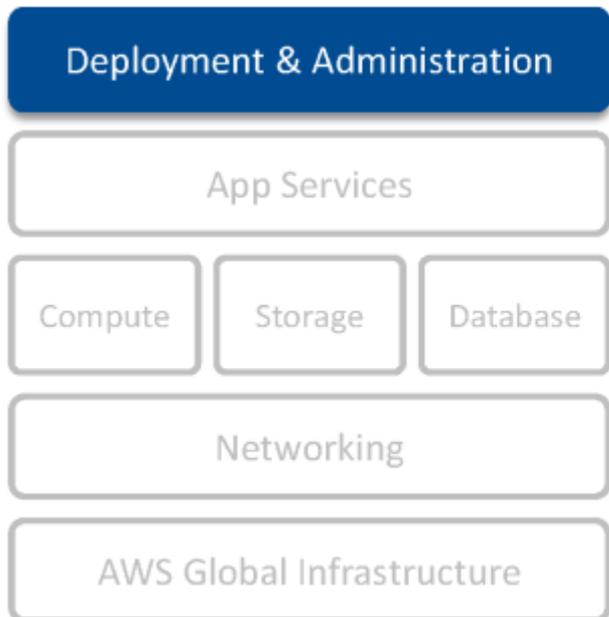
## Identity & Access Management

*Granular control of user rights with AWS*

*Automated granting of EC2 service rights*

## Software Developer Kits

*Comprehensive support of programming models for using AWS services*



# + others

Simple Email Service

Simple Notification Service

ElastiCache

Elastic MapReduce

CloudWatch

...

# Choose AWS

- If you
  - want to use third-party open-source software
  - have existing code
  - want to transfer a web app to your own machine/servers later
  - port code to another language
  - want complete control
  - need to stress/load test an app
    - for example, load up 1,000 instances

# GAE

- PaaS
  - Languages and runtime environments
    - Python
    - Java
    - Go
  - Sandbox
  - Datastore



# GAE vs AWS



Python/Java  
BigTable  
Other API's

VMs  
Flat File Storage



## **GAE:**

- **Higher-level functionality (e.g., automatic scaling)**
- **More restrictive (e.g., respond to URL only)**
- **Proprietary lock-in**

## **EC2/S3:**

- **Lower-level functionality**
- **More flexible**
- **Coarser billing model**

# Choose GAE

- If you
  - have no pre-existing code
  - are building request-response web apps or mashups
  - consider time-to-market the most important thing
  - aren't doing anything fancy (installing software)
  - aren't worried about lock-in to Google

# Windows Azure

- IaaS + PaaS

The image shows a screenshot of the Windows Azure Services Platform interface. At the top, there is a navigation bar with logos for Windows Live, Office Live, Exchange Online, SharePoint Online, and Microsoft Dynamics CRM Online. Below this, the main heading reads "Azure™ Services Platform". Underneath the heading, there are five white rounded rectangular buttons, each containing a logo and text: "Live Services", ".NET Services", "SQL Services", "Microsoft SharePoint Services", and "Microsoft Dynamics CRM Services". At the bottom of the interface, there is a large white rounded rectangular box containing the Windows logo and the text "Windows Azure™".

# Choose Azure

- If you
  - already use the .NET and SQL Server portions of the Microsoft stack
  - have existing code developed to those Microsoft APIs
  - have teams that normally develop in Visual Studio using C#
  - want to blend development from desktop to cloud
  - have no issue with lock-in to Microsoft

# Take Home Messages

- Understanding the Different Classifications of Clouds
  - IaaS, PaaS, SaaS
  - Public, Private, Community, Hybrid
- Matching Cloud Providers to Your Needs
  - Amazon, Google, Microsoft