BITS Pilani presentation

BITS Pilani
Hyderabad Campus

D. Powar
Lecturer,
BITS-Pilani, Hyderabad Campus

innovate achieve lead

BITS Pilani
Hyderabad Campus

SSZG527

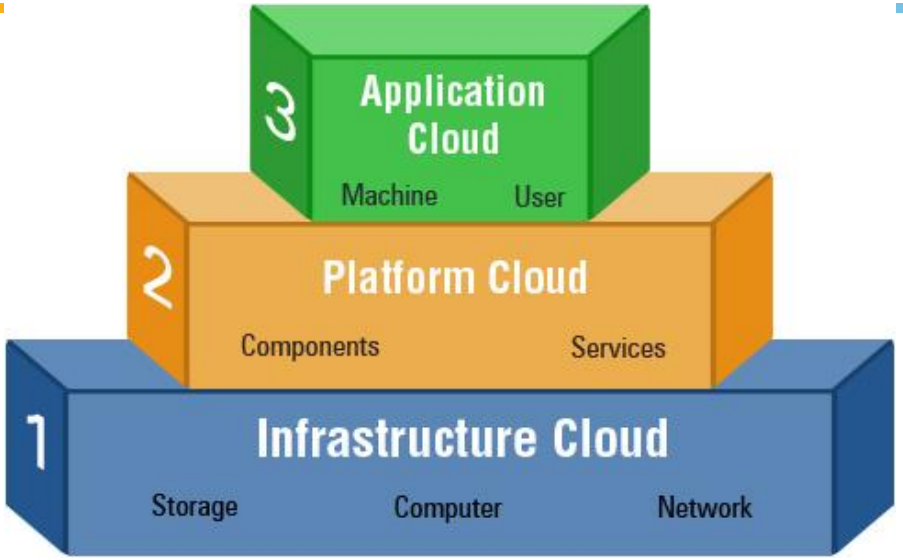
Lecture 5

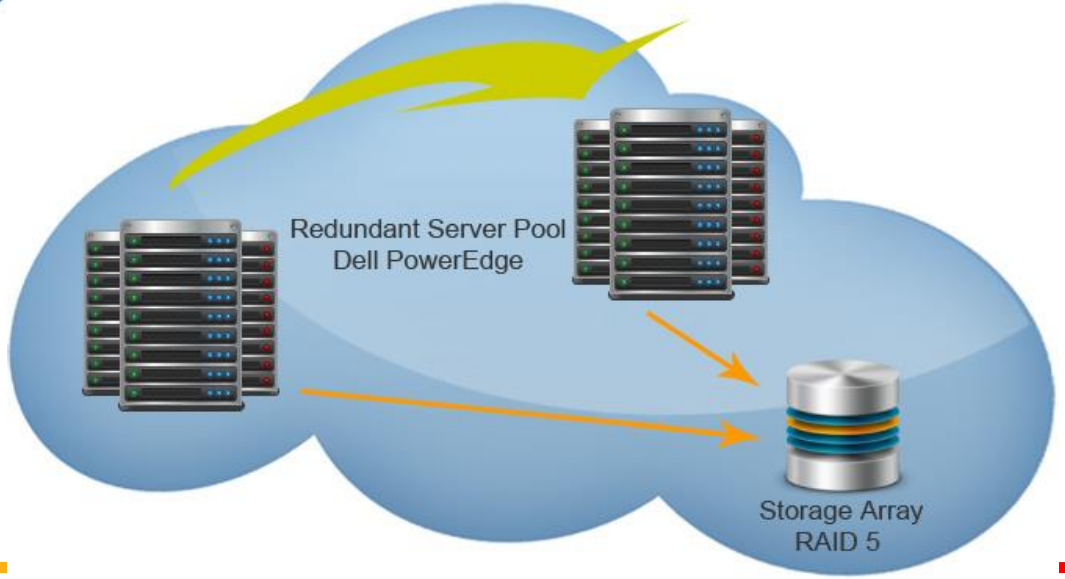
Cloud Computing

Agenda:

- Introduction to IaaS
- IaaS examples
- Reference Model of AWS
- Amazon compute cloud services
- Amazon Storage service

Introduction to IaaS





Example:


amazon
web services™


bitrefinery


rackspace®
the open cloud company


GOGRID


Go Daddy®.COM
The web
is your Domain!™


HOSTING
.com


NephoScale
CLOUD COMPUTING

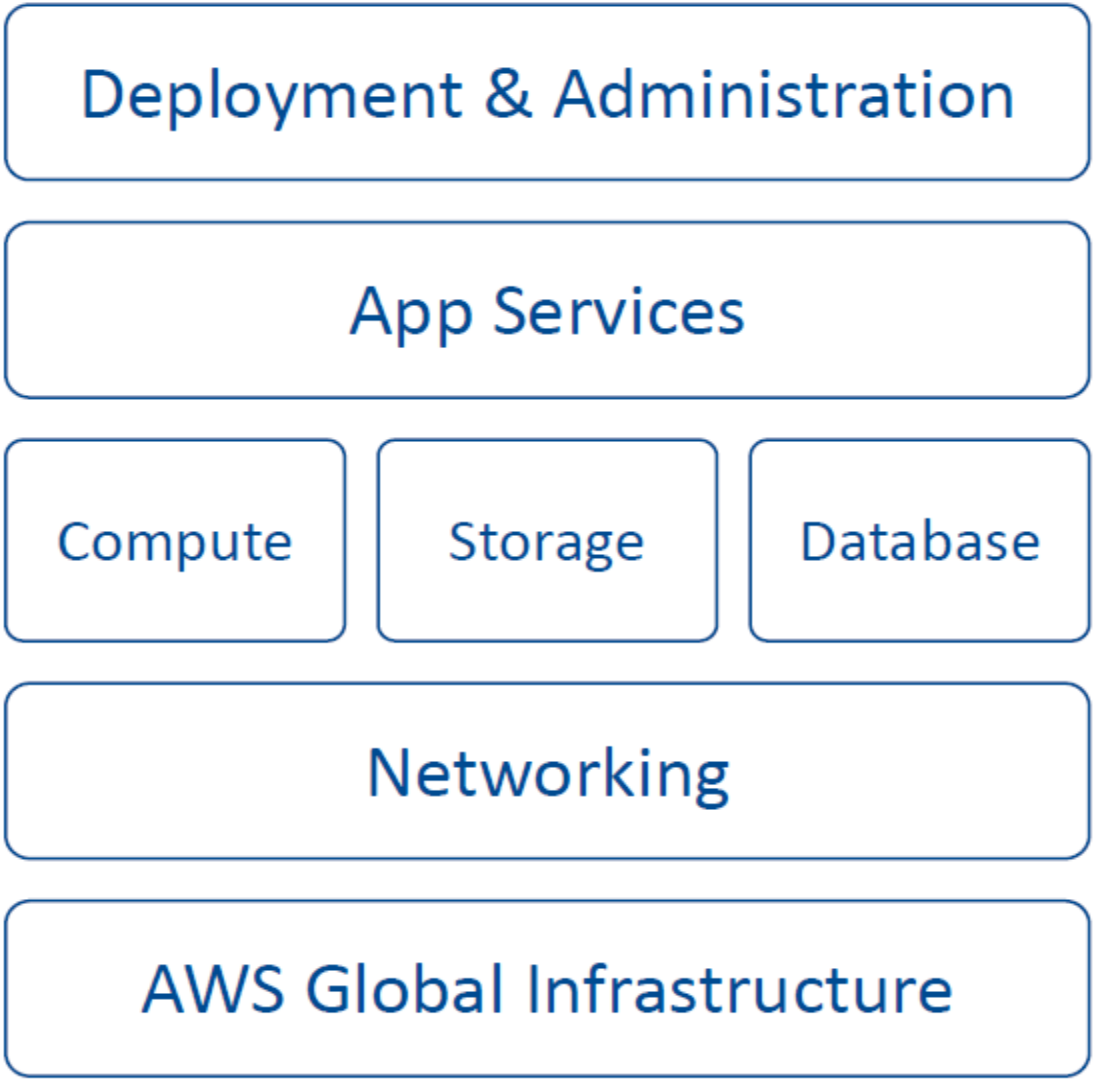

ReliaCloud™


OpSource™
The Business of Cloud Operations

SOFTLAYER®


terremark®

Reference Model



An overview of AWS

- AWS is Amazon's umbrella description of all of their web-based technology services.
- Main services include:
  - Compute
  - Storage
  - Database
  - Deployment & Management
  - Application Services
  - Networking
  - Content Delivery

Category	Service
Compute	❖ Amazon Elastic Compute Cloud (Amazon EC2) ❖ Amazon Elastic MapReduce (Amazon EMR) ❖ Auto scaling ❖ Elastic Load Balancing
Storage	▪ Amazon Simple Storage Service (Amazon S3) ▪ Amazon Glacier ▪ AWS Storage Gateway ▪ AWS Import/Export
Content Delivery	✓ Amazon CloudFront
Database	❑ Amazon Relational Database Service (Amazon RDS) ❑ Amazon DynamoDB ❑ Amazon ElastiCache
Deployment & Management	• AWS Identity and Access Management (IAM) • Amazon CloudWatch • AWS Elastic Beanstalk • AWS CloudFormation
Application Services	➢ Amazon Simple Queue Service (SQS) ➢ Amazon Simple Notification Service (Amazon SNS) ➢ Amazon Simple Email Service (Amazon SES) ➢ Amazon CloudSearch
Networking	○ Amazon Virtual Private Cloud (Amazon VPC) ○ Amazon Route 53 ○ AWS Direct Connect

Amazon Elastic Compute Cloud (Amazon EC2)

- Amazon EC2 provides resizable computing capacity in the AWS cloud. Using Amazon EC2 eliminates your need to invest in hardware up front, so you can develop and deploy applications faster. You can use Amazon EC2 to launch as many or as few virtual servers as you need, configure security and networking, and manage storage.
- Amazon EC2 enables you to scale up or down to handle changes in requirements or spikes in popularity, reducing your need to forecast traffic.

Features of Amazon EC2

- Virtual computing environments, known as *instances*
- Pre-configured templates for your instances, known as *Amazon Machine Images (AMIs)*, that package the bits you need for your server (including the operating system and additional software)
- Various configurations of CPU, memory, storage, and networking capacity for your instances, known as *instance types*
- Storage volumes for temporary data that's deleted when you stop or terminate your instance, known as *instance store volumes*
- Persistent storage volumes for your data using Amazon Elastic Block Store (Amazon EBS), known as *Amazon EBS volumes*
- Multiple physical locations for your resources, such as instances and Amazon EBS volumes, known as *regions* and *Availability Zones*
- Static IP addresses for dynamic cloud computing, known as *Elastic IP addresses*
- Virtual networks you can create that are logically isolated from the rest of the AWS cloud, and that you can optionally connect to your own network, known as *virtual private clouds* (VPCs)

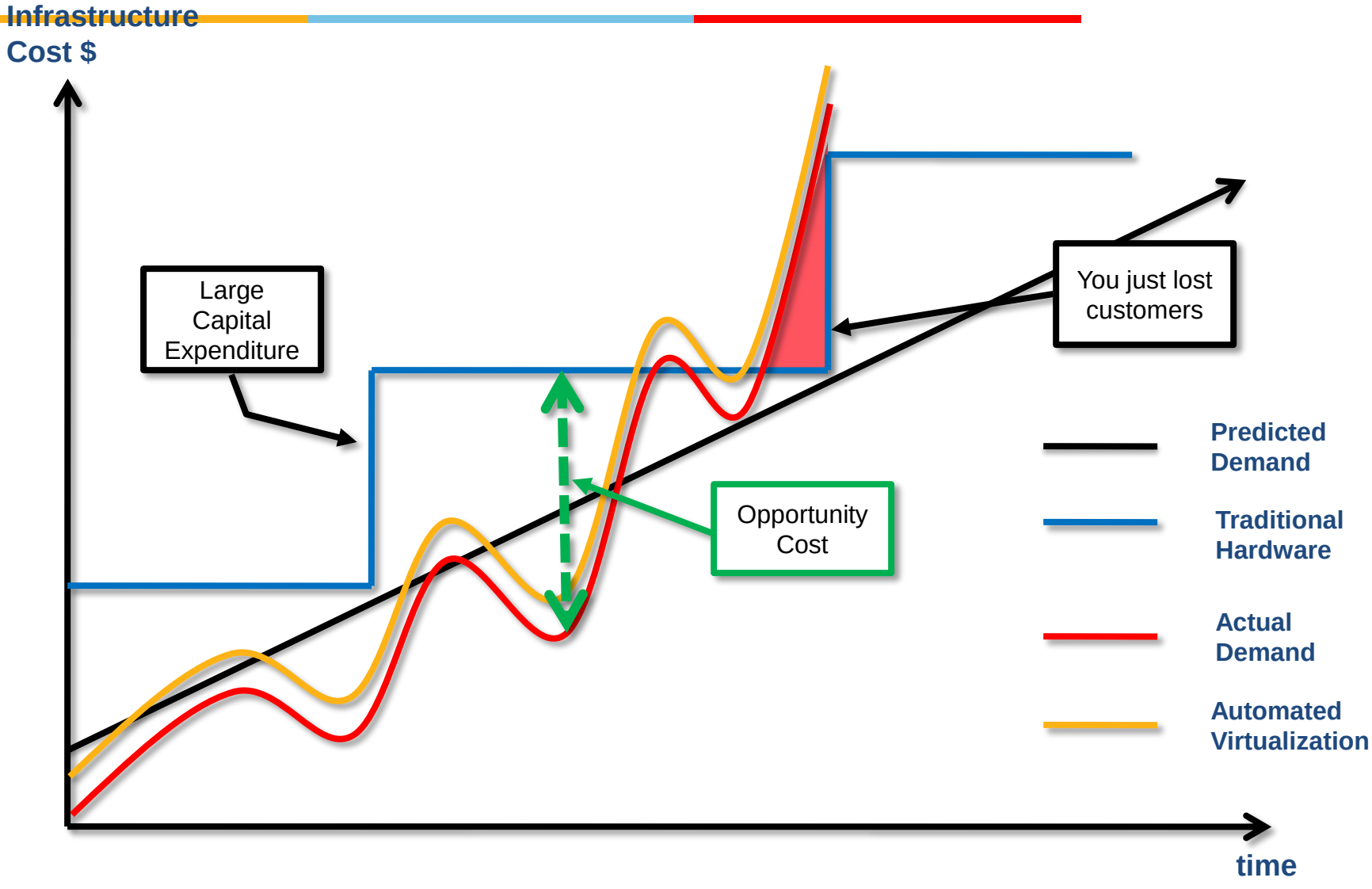
Amazon EC2 (contd..)

- EC2 provides web services API for provisioning, managing, and deprovisioning virtual servers inside amazon cloud.
- Applications anywhere on the Internet can launch a virtual server in the amazon cloud with a single web services call (either REST or SOAP WS call)

Amazon EC2 (Contd..)

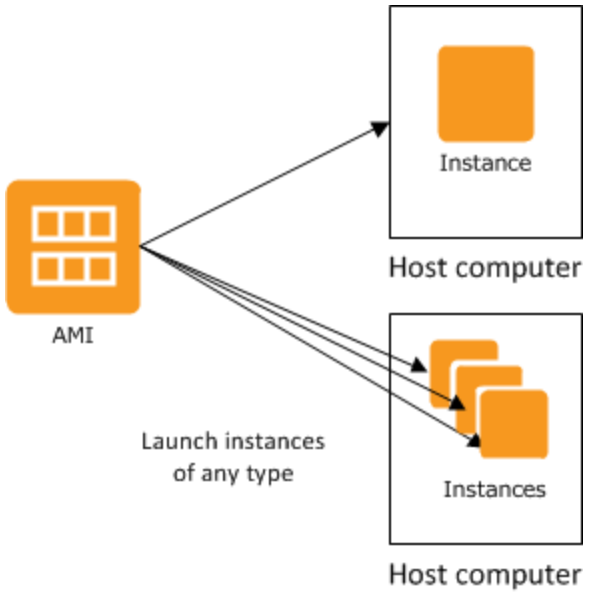
- Resizable compute capacity in the cloud
  - Obtain and boot new server instances in minutes
  - Quickly scale capacity, up or down, as your computing requirements change
- Full root access to a blank Linux machine
- Simple Web service management interface
- Changes the economics of computing

Predictions Cost Money



Amazon Machine Image (AMI)

- An AMI is a template that contains a software configuration (for example, an operating system, an application server, and applications). From an AMI, you launch an *instance*, which is a copy of the AMI running as a virtual server in the cloud. You can launch multiple instances of an AMI



Three Flavors of Amazon Machine Images

Public AMIs: Use pre-configured, template AMIs to get up and running immediately.

Private AMIs: Create an Amazon Machine Image (AMI) containing your applications, libraries, data and associated configuration settings

Paid AMIs: Set a price for your AMI and let others purchase and use it (Single payment and/or per hour)

Three Amazon EC2 Choices

	Small	Large	Extra Large
Bits	32	64	64
RAM	1.7 GB	7.5 GB	15 GB
Disk	160 GB	850 GB	1690 GB
EC2 Compute Units	1	4	8
I/O Performance	Medium	High	High
Firewall	Yes	Yes	Yes

Amazon EC2 Instance Types

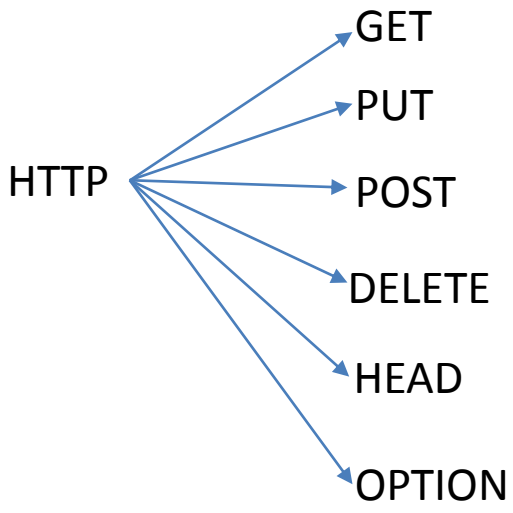
Instance Family	Instance Type	Processor Arch	vCPU	ECU	Memory (GiB)	Instance Storage (GB)	EBS-optimized Available	Network Performance
General purpose	m3.medium	64-bit	1	3	3.75	1 x 4 SSD*6	-	Moderate
General purpose	m3.large	64-bit	2	6.5	7.5	1 x 32 SSD*6	-	Moderate
General purpose	m3.xlarge	64-bit	4	13	15	2 x 40 SSD*6	Yes	Moderate
General purpose	m3.2xlarge	64-bit	8	26	30	2 x 80 SSD*6	Yes	High
General purpose	m1.small	32-bit or 64-bit	1*1	1	1.7	1 x 160	-	Low
General purpose	m1.medium	32-bit or 64-bit	1	2	3.75	1 x 410	-	Moderate
General purpose	m1.large	64-bit	2	4	7.5	2 x 420	Yes	Moderate
General purpose	m1.xlarge	64-bit	4	8	15	4 x 420	Yes	High
Compute optimized	c3.large	64-bit	2	7	3.75	2 x 16 SSD	-	Moderate
Compute	c3.xlarge	64-bit	4	14	7.5	2 x 40 SSD	Yes	Moderate

REST

- **Representational State Transfer** (**REST**) is a style of software architecture for distributed systems such as the World Wide Web.
- REST has emerged as a predominant web service design model
- REST-style architectures consist of clients and servers.
- REST facilitates the transaction between web servers by allowing loose coupling between different services.
- Facebook, twitter, google, flickr, etc

REST

- HTTP is the foundation for REST
- Correlated to CRUD



REST (best fit)

- Web Services
- Limited bandwidth (Smaller mesg size)
- Limited resources (No xml parsing required)
- Exposing data over the Internet
- Combining content from many different sources in a web browser
- Pros:
  - ✓ Uniform Interface
  - ✓ Interoperable
  - ✓ Stateless
  - ✓ Easy to Access
  - ✓ Scalable

SOAP

- **Simple Object Access Protocol**, is a protocol specification for exchanging structured information in the implementation of Web Services in computer networks.
- It relies on XML for its message format, and usually relies on other Application Layer protocols, most notably HTTP and SMTP, for message negotiation and transmission

SOAP (best fit)

- Enterprise services (reliable, transactional, secured)
- Asynchronous processing
- Stateful/conversational operations
- Standards support (WS - standard), interoperability with business support

EC2 concepts

- Elastic IP address: Not static not dynamic but elastic; an IP reserved for your use; disappears once the server is terminated.
  - Elastic IP Address belongs to the account and not to a virtual machine instance
  - It exists until it is explicitly removed, and remains associated with the account even while it is associated with no instance
  - Allow you to mask instance or availability zone failures by programmatically remapping your public IP addresses to any instance associated with your account.
  - Rather than waiting for a data technician to reconfigure or replace your host, or waiting for DNS to propagate to all of your customers, Amazon EC2 enables you to engineer around problems with your instance or software by programmatically remapping your Elastic IP address to a replacement instance

EC2 Private IP Address	EC2 Public IP Address	EC2 Elastic IP Address
The internal RFC 1918 address of an instance that is only routable within the EC2 Cloud. Network traffic originating outside the EC2 network cannot route to this IP, and must use the Public IP or Elastic IP Address mapped to the instance	Internet routable IP address assigned by the system for all instances. Traffic routed to the Public IP is translated via 1:1 Network Address Translation (NAT) and forwarded to the Private IP address of an instance. The mapping of a Public IP to Private IP of an instance is the default launch configuration for all instance types. Public IP Addresses are no longer usable upon instance termination	Internet routable IP address allocated to an AWS EC2 account. Similar to EC2 Public Address, 1:1 NAT is used to map Elastic IP Addresses with their associated Private IP addresses. Unlike a standard EC2 Public IP Address, Elastic IP Addresses are allocated to accounts and can be remapped to other instances when desired

Auto Scaling

- It allows you to scale your Amazon EC2 capacity up or down automatically according to conditions you define.
- With Auto Scaling, you can ensure that the number of Amazon EC2 instances you're using increases seamlessly during demand spikes to maintain performance, and decreases automatically during demand lulls to minimize costs.
- Auto Scaling is particularly well suited for applications that experience hourly, daily, or weekly variability in usage. Auto Scaling is enabled by Amazon CloudWatch and available at no additional charge beyond Amazon CloudWatch fees.

Elastic Load Balancing

- It automatically distributes incoming application traffic across multiple Amazon EC2 instances.
- It enables you to achieve even greater fault tolerance in your applications, seamlessly providing the amount of load balancing capacity needed in response to incoming application traffic.
- It detects unhealthy instances and automatically reroutes traffic to healthy instances until the unhealthy instances have been restored.
- Customers can enable Elastic Load Balancing within a single Availability Zone or across multiple zones for even more consistent application performance.

Amazon Virtual Private Cloud (Amazon VPC)

- With Amazon VPC, you can define a virtual network topology that closely resembles a traditional network that you might operate in your own data center.
- You have complete control over your virtual networking environment, including selection of your own IP address range, creation of subnets, and configuration of route tables and network gateways.
- You can easily customize the network configuration for your Amazon VPC.
- For example: you can create a public-facing subnet for your webservers that has access to the Internet, and place your backend systems such as databases or application servers in a private-facing subnet with no Internet access.

Amazon Route 53

- It is a highly available and scalable Domain Name System (DNS) web service.
- It is designed to give developers and businesses an extremely reliable and cost-effective way to route end users to Internet applications by translating human readable names, such as www.example.com, into the numeric IP addresses, such as 192.0.2.1
- Route 53 effectively connects user requests to infrastructure running in AWS, such as an EC2 instance, an elastic load balancer, or an Amazon S3 bucket

Amazon EC2 Key Pairs

- Amazon EC2 uses public–key cryptography to encrypt and decrypt login information
- EC2 Key Pair is used to access the instance
- EC2 user has to create the Key Pair (.pem file)
- Public key will available with amazon and private key is with client machine
- Linux/Unix instances have no password, and you use a key pair to log in using SSH. With Windows instances, you use a key pair to obtain the administrator password and then log in using RDP

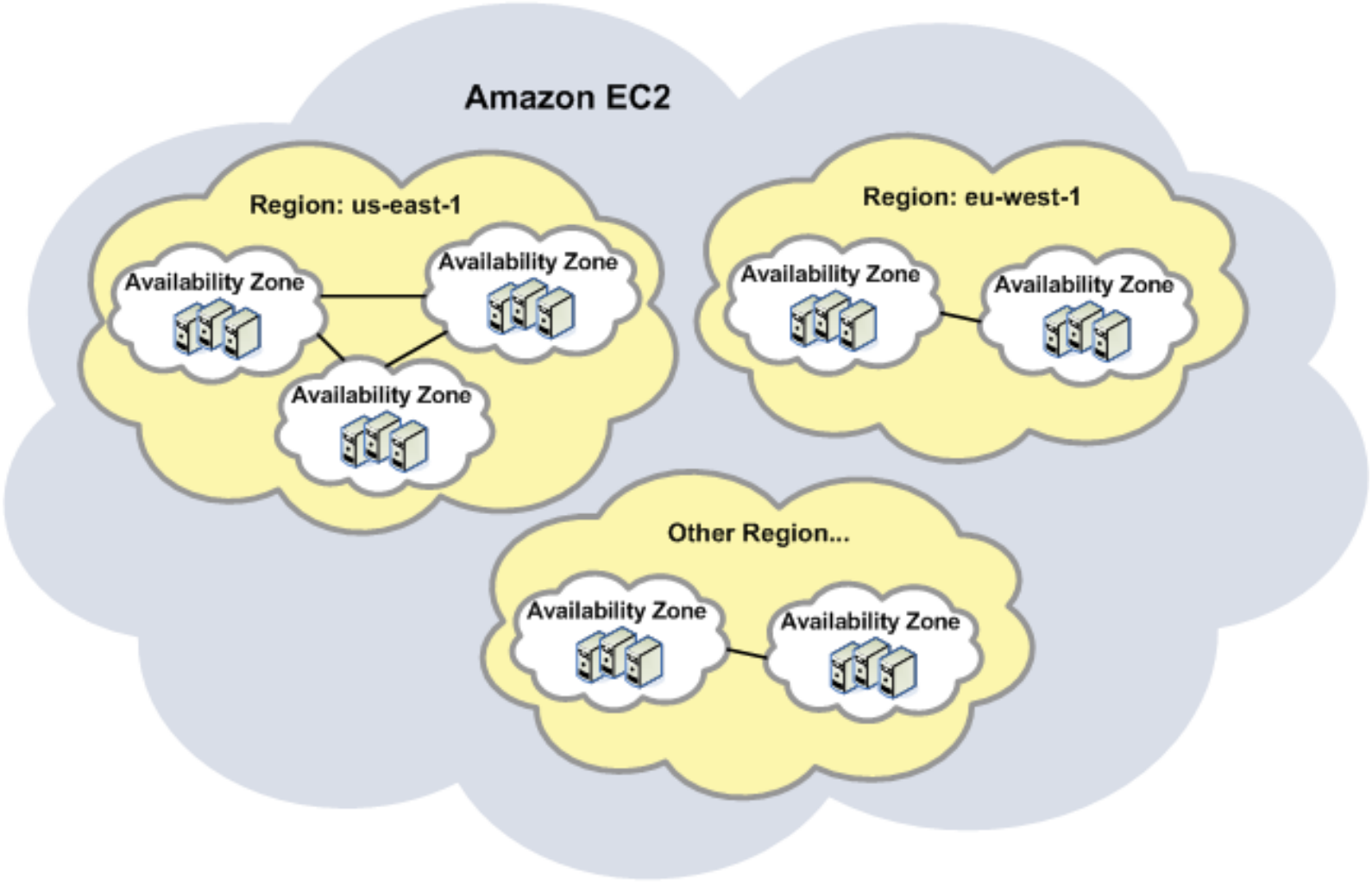
Security Groups

- Defines network security policies that restrict the ports through which any machine (instance) can be accessed
- The default security group when creating an EC2 instance allows the instance to connect to any outside IP address but disallows incoming connections
- Different security groups should be created for different server (instance) types;
  - for example, the web server security group could specify that port 80 may be opened for incoming connections.

Region Vs Availability zones

- An AWS region contains two or more availability zones.
- Each zone is basically a separate datacenter, and provides low latency connectivity to all other zones in the region.
- EC2 instances, reside in the region of your choice. The AWS regions are isolated from each other, but you can seamlessly manage resources in different availability zones within the same region.

Availability Zones (Amazon EC2)



Availability zones

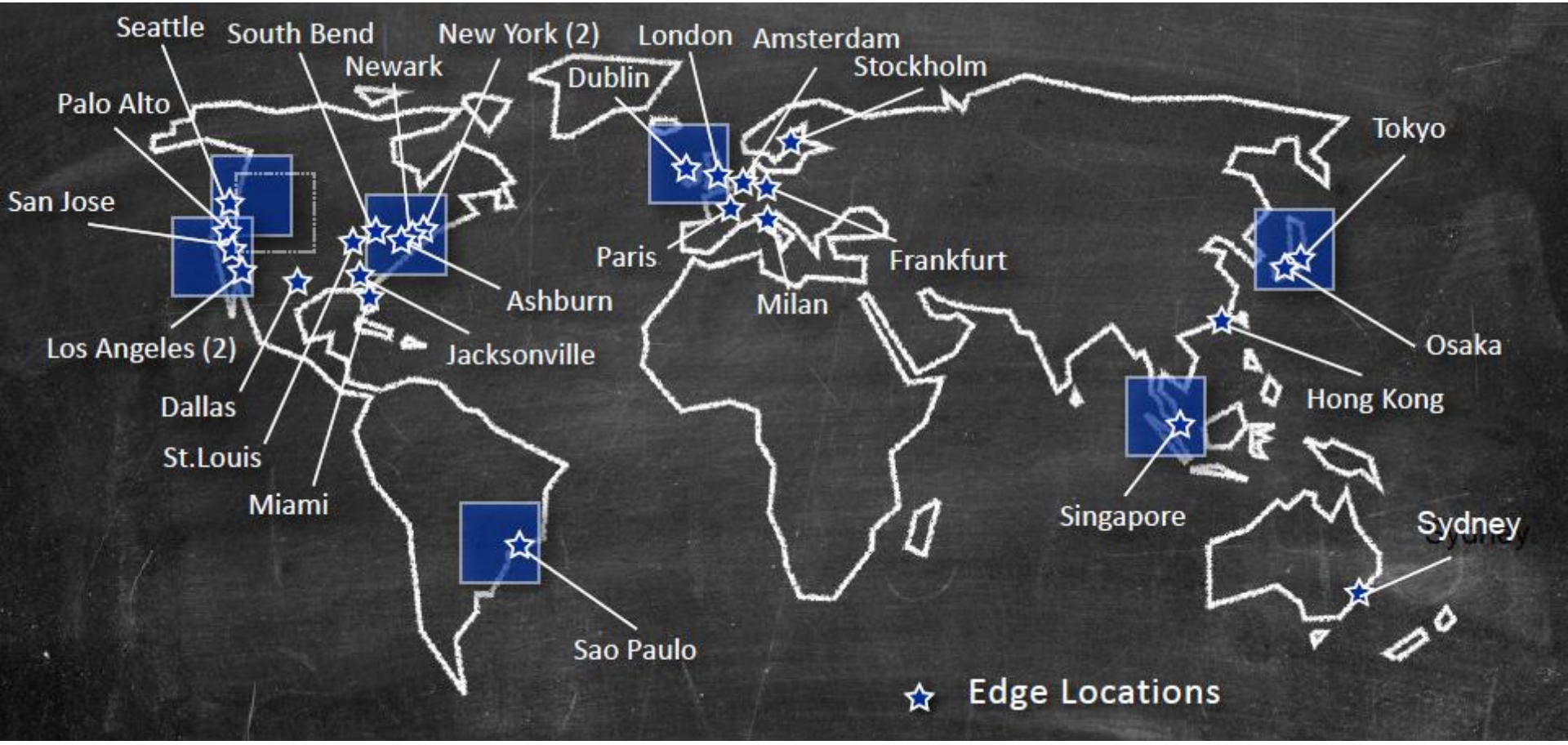
Regions

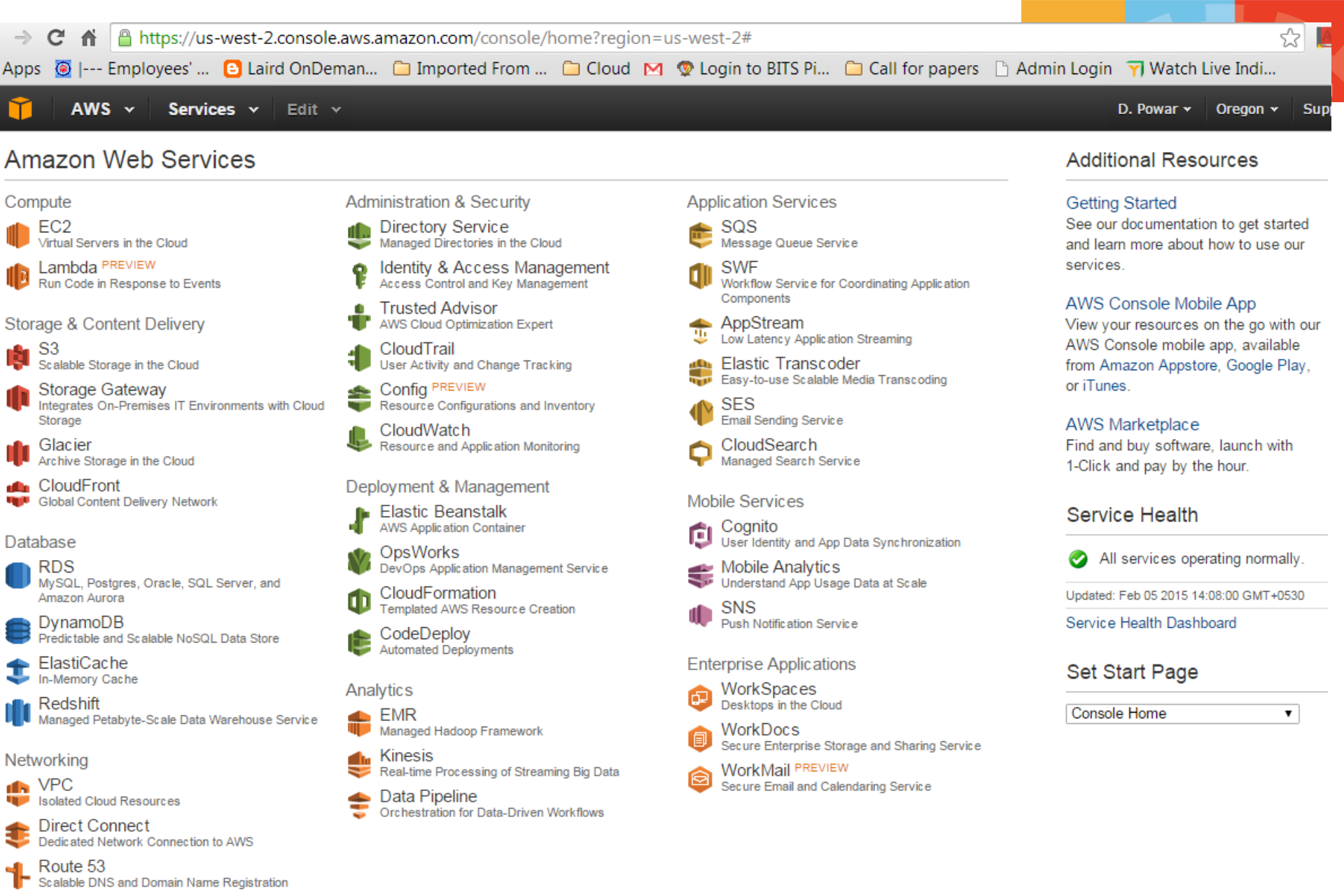
Code	Name
ap-northeast-1	Asia Pacific (Tokyo) Region
ap-southeast-1	Asia Pacific (Singapore) Region
ap-southeast-2	Asia Pacific (Sydney) Region
eu-west-1	EU (Ireland) Region
sa-east-1	South America (Sao Paulo) Region
us-east-1	US East (Northern Virginia) Region
us-west-1	US West (Northern California) Region
us-west-2	US West (Oregon) Region

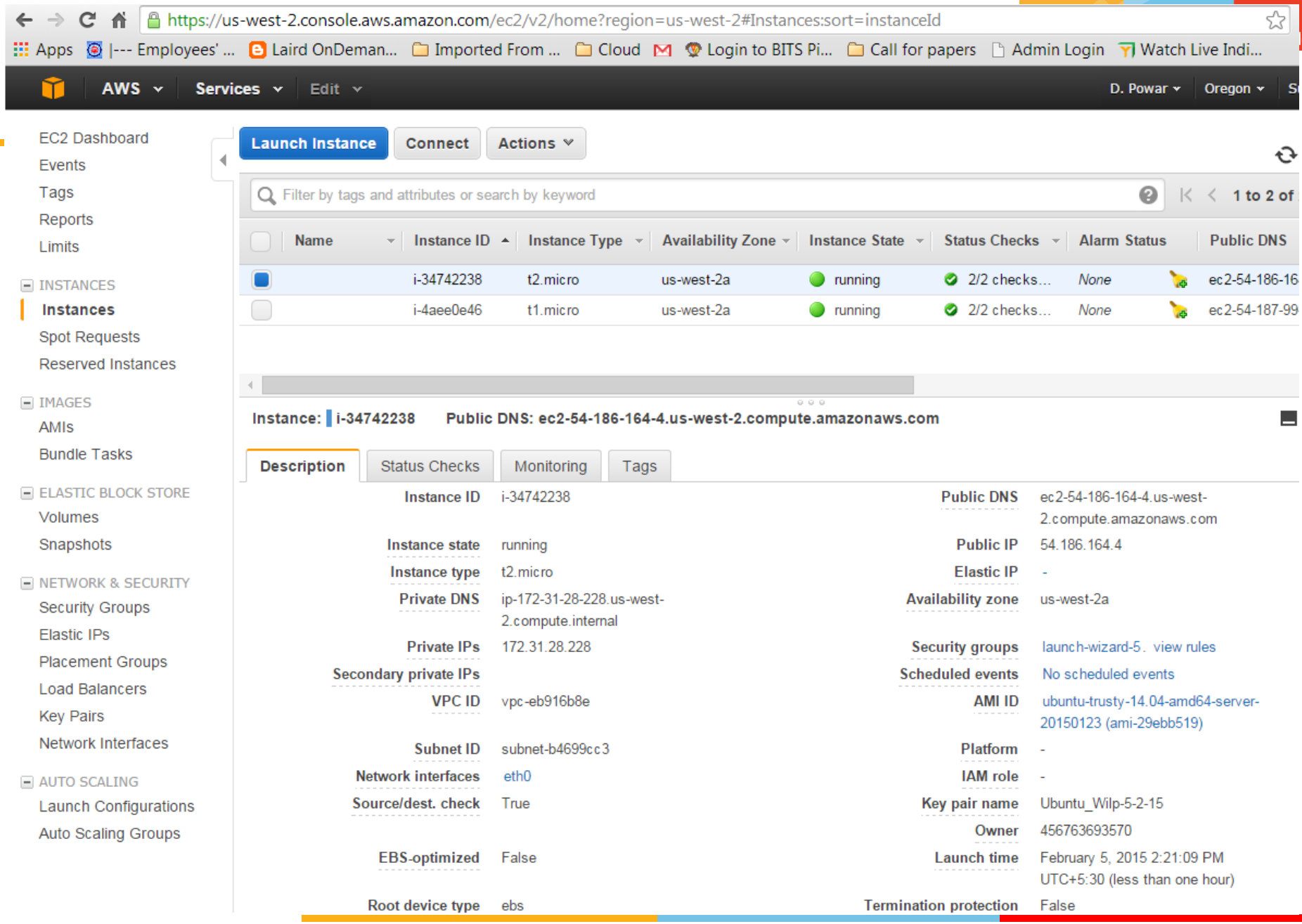
Edge Location

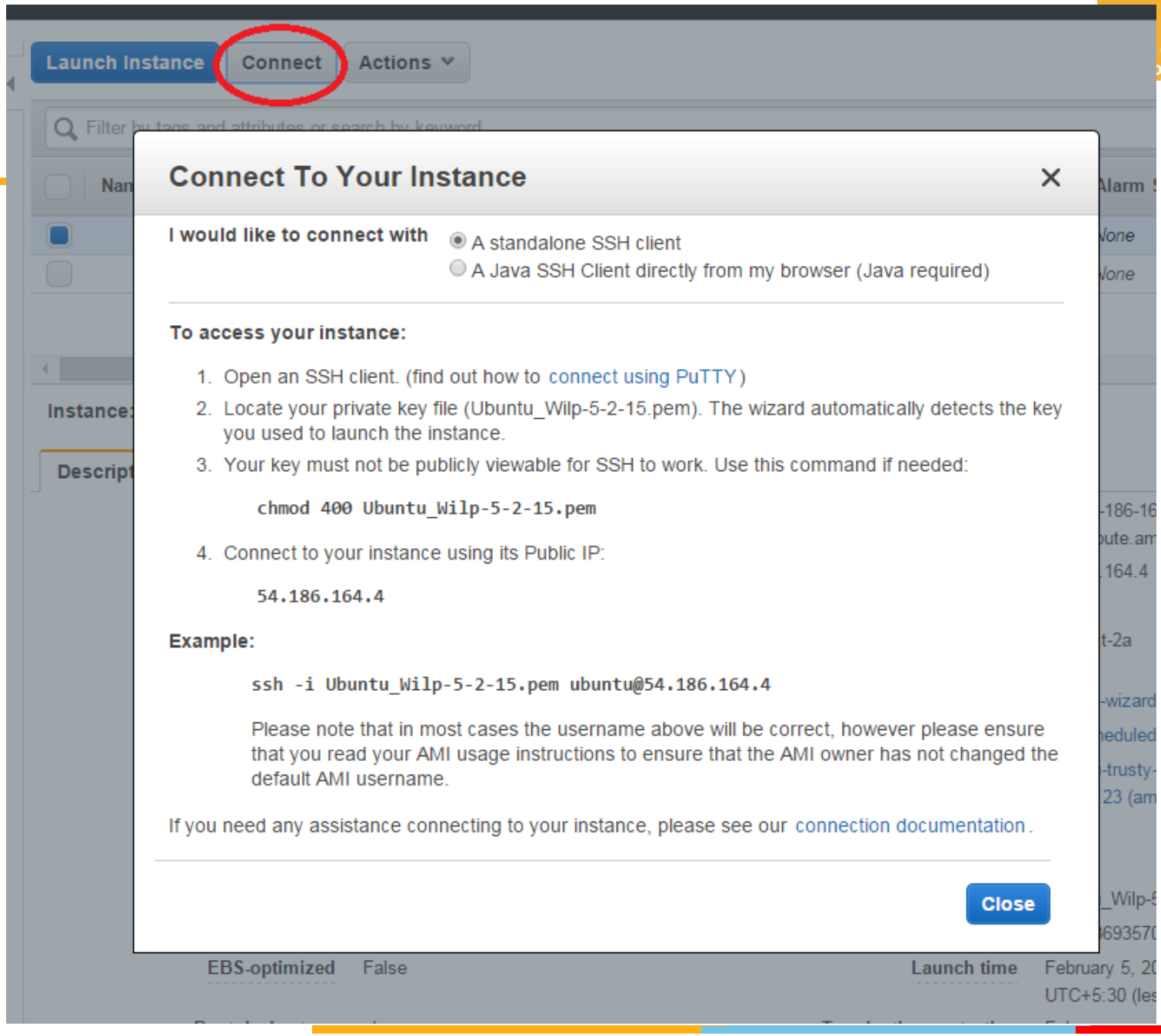
- Edge locations serve requests for CloudFront and Route 53.
- CloudFront is a content delivery network, while Route 53 is a DNS service.
- Requests going to either one of these services will be routed to the nearest edge location automatically.
- If edge server does not have the copy of the data, CloudFront will pull a copy from the origin server
- This allows for low latency no matter where the end user is located.

Amazon Edge location




https://us-west-2.console.aws.amazon.com/console/home?region=us-west-2#
Apps
|--- Employees' ...
Laird OnDeman...
Imported From ...
Cloud
Login to BITS Pi...
Call for papers
Admin Login
Watch Live Indi...
AWS
Services
Edit
D. Powar
Oregon
Amazon Web Services
Compute
EC2
Virtual Servers in the Cloud
Lambda PREVIEW
Run Code in Response to Events
Storage & Content Delivery
S3
Scalable Storage in the Cloud
Storage Gateway
Integrates On-Premises IT Environments with Cloud Storage
Glacier
Archive Storage in the Cloud
CloudFront
Global Content Delivery Network
Database
RDS
MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
DynamoDB
Predictable and Scalable NoSQL Data Store
ElastiCache
In-Memory Cache
Redshift
Managed Petabyte-Scale Data Warehouse Service
Networking
VPC
Isolated Cloud Resources
Direct Connect
Dedicated Network Connection to AWS
Route 53
Scalable DNS and Domain Name Registration
Administration & Security
Directory Service
Managed Directories in the Cloud
Identity & Access Management
Access Control and Key Management
Trusted Advisor
AWS Cloud Optimization Expert
CloudTrail
User Activity and Change Tracking
Config PREVIEW
Resource Configurations and Inventory
CloudWatch
Resource and Application Monitoring
Deployment & Management
Elastic Beanstalk
AWS Application Container
OpsWorks
DevOps Application Management Service
CloudFormation
Templated AWS Resource Creation
CodeDeploy
Automated Deployments
Analytics
EMR
Managed Hadoop Framework
Kinesis
Real-time Processing of Streaming Big Data
Data Pipeline
Orchestration for Data-Driven Workflows
Application Services
SQS
Message Queue Service
SWF
Workflow Service for Coordinating Application Components
AppStream
Low Latency Application Streaming
Elastic Transcoder
Easy-to-use Scalable Media Transcoding
SES
Email Sending Service
CloudSearch
Managed Search Service
Mobile Services
Cognito
User Identity and App Data Synchronization
Mobile Analytics
Understand App Usage Data at Scale
SNS
Push Notification Service
Enterprise Applications
WorkSpaces
Desktops in the Cloud
WorkDocs
Secure Enterprise Storage and Sharing Service
WorkMail PREVIEW
Secure Email and Calendaring Service
Additional Resources
Getting Started
See our documentation to get started and learn more about how to use our services.
AWS Console Mobile App
View your resources on the go with our AWS Console mobile app, available from Amazon Appstore, Google Play, or iTunes.
AWS Marketplace
Find and buy software, launch with 1-Click and pay by the hour.
Service Health
All services operating normally.
Updated: Feb 05 2015 14:08:00 GMT+0530
Service Health Dashboard
Set Start Page
Console Home


https://us-west-2.console.aws.amazon.com/ec2/v2/home?region=us-west-2#Instances:sort=instanceId
Apps
|--- Employees' ...
Laird OnDeman...
Imported From ...
Cloud
Login to BITS Pi...
Call for papers
Admin Login
Watch Live Indi...
AWS
Services
Edit
D. Powar
Oregon
EC2 Dashboard
Events
Tags
Reports
Limits
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
AUTO SCALING
Launch Configurations
Auto Scaling Groups
Launch Instance
Connect
Actions
Filter by tags and attributes or search by keyword
1 to 2 of
Name
Instance ID
Instance Type
Availability Zone
Instance State
Status Checks
Alarm Status
Public DNS
i-34742238 t2.micro us-west-2a running 2/2 checks... None ec2-54-186-16
i-4aee0e46 t1.micro us-west-2a running 2/2 checks... None ec2-54-187-99
Instance: i-34742238 Public DNS: ec2-54-186-164-4.us-west-2.compute.amazonaws.com
Description
Status Checks
Monitoring
Tags
Instance ID i-34742238
Instance state running
Instance type t2.micro
Private DNS ip-172-31-28-228.us-west-2.compute.internal
Private IPs 172.31.28.228
Secondary private IPs
VPC ID vpc-eb916b8e
Subnet ID subnet-b4699cc3
Network interfaces eth0
Source/dest. check True
EBS-optimized False
Root device type ebs
Public DNS ec2-54-186-164-4.us-west-2.compute.amazonaws.com
Public IP 54.186.164.4
Elastic IP -
Availability zone us-west-2a
Security groups launch-wizard-5. view rules
Scheduled events No scheduled events
AMI ID ubuntu-trusty-14.04-amd64-server-20150123 (ami-29ebb519)
Platform -
IAM role -
Key pair name Ubuntu_Wilp-5-2-15
Owner 456763693570
Launch time February 5, 2015 2:21:09 PM UTC+5:30 (less than one hour)
Termination protection False


Launch Instance
Connect
Actions
Connect To Your Instance
I would like to connect with
A standalone SSH client
A Java SSH Client directly from my browser (Java required)
To access your instance:
1. Open an SSH client. (find out how to connect using PuTTY)
2. Locate your private key file (Ubuntu_Wilp-5-2-15.pem). The wizard automatically detects the key you used to launch the instance.
3. Your key must not be publicly viewable for SSH to work. Use this command if needed:
chmod 400 Ubuntu_Wilp-5-2-15.pem
4. Connect to your instance using its Public IP:
54.186.164.4
Example:
ssh -i Ubuntu_Wilp-5-2-15.pem ubuntu@54.186.164.4
Please note that in most cases the username above will be correct, however please ensure that you read your AMI usage instructions to ensure that the AMI owner has not changed the default AMI username.
If you need any assistance connecting to your instance, please see our connection documentation.
Close
EBS-optimized
False
Launch time

Summary

- Introduction to IaaS
- IaaS examples
- Reference Model of AWS
- Amazon compute cloud services
  - EC2
  - Regions & Availability zones
  - Elastic IP address
- Compute and network
  - Amazon Virtual Private Cloud
  - Amazon route 53
  - EC2 Key pair, Security Groups