20180502

INTRODUCTION TO AWS CLOUD – Jody Soeiro de Faria

**Before cloud**

Pay as you go pricing

Guessing theoretical maximum peaks

Expensive resources that stay idle

Insufficient capacity to meet needs

**After AWS**

Initiated within seconds

Temporary and disposable

Free from inflexibility and efficient

“AGILITY”

Speed, experimentation, culture of innovation

*Increase speed and global reach within a moment’s notice*

New resources are a click away

*Increase experimentation*

Operations as code

Safely experiment

For example:

Spin up servers in minutes for experimenting

Return or re-purpose servers

Testing using different configurations

AWS CloudFormation enables templated environments

*Increase innovation*

Experiment quickly with low cost / risk

More experimentation more often

Discovery of new configurations

THE AWS INFRASTRUCTURE

Elasticity, scalability, and reliability of computing resources

Regions

Physical location in the world

Contains multiple Availability Zones

Availability Zones (AZs)

One or more discrete data centers

Redundant: power/networking/connectivity

*Highly Available*

Fault tolerance (ability to remain operational even if other systems fails):

Applications operational during component failure

Built-in redundancy of components

High availability:

Systems “always” functioning and accessible

Downtime is minimized as much as possible

Without human intervention

*Elasticity, Scalability and High Performance*

**Elasticity is the power to scale resources up or down easily** while only paying actual resources used

Quickly deploy new applications

Instantly scale up as the workload grows

Instantly shut down resources that are no longer required

Scale down and don’t pay for the infrastructure

AWS

Use services at your pace

Adapt their consumption

Launch new services or products

Accommodate new strategic directions

Tools to run a wide range of applications

Auto Scaling and Elastic Load Balancing

Scale up or down based on demand

Deploy your system in multiple regions

Lower latency and better experience

Innovative services and cutting-edge technology

Virtually any workload

**SECURITY AND COMPLIANCE**

Customer retains control over region where data located

Easy to make regional compliance

Security auditing often periodic and manual

AWS cloud provides governance capabilities

Industry-leading capabilities that meet the strictest security requirements

Data centers

State-of-the-art electronic surveillance

Multi-factor access control systems

Staffed 24x7 by trained security guards

Access is strictly least-privileged basis

Environmental systems minimize the impact of disruptions

Multiple regions and Availability Zones enable resiliency

Security policy can be formalized and embedded with the design of your infrastructure

Reliability

High performing and reliable solutions

Implement solutions quickly and with limited friction

Reliability is…

Ability to recover from failures

Dynamically acquire resources to meet demand and mitigate disruptions

Must have well-planned foundation

Handles changes in demand

Detect failure and automatically heal itself

AWS  
 Achieve greater flexibility and capacity

Reduce uncertainty of forecasting hardware needs

Capacity and reliability that is difficult to match by on-premise solutions

**PRICING: PAY AS YOU GO**

Avoid dedicating resources to infrastructure

*Benefits of pay-as-you-go*

Redirect focus to innovation and invention  
Reduce performance complexity

Reduce procurement complexity

Enabling elasticity

Adapt to changing business

Improve responsiveness to change

KNOWLEDGE CHECK

NOT an advantage of cloud computing over on-premise computing?

*(NOT SURE) Answer: Pay for racking, stacking, powering servers*

Following are NOT a benefit of AWS cloud computing:

Answer: Multiple procurement cycles and high latency

What is TRUE about Regions?

Answer: Each region is located in a separate geographic area and physical location with multiple Availability Zones

Advantages of AWS cloud security?

Answer: You retain complete control and ownership of your data region and AWS uses multi-factor access control systems

Which of the following AWS tools help your application scale up or down based on demand?

Answer: Auto Scaling and Elastic Load Balancing

What pricing model that allows AWS customers to pay for resources on an as needed basis?

Answer: Pay as you go

Number one reason customers are switching to cloud computing

Answer: agility

SECTION TWO: AWS TECHNOLOGY OVERVIEW – MIKE BLACKMER

**AWS Services and Categories**

**Intro to Documentation**

Discussion for now:

Compute

Storage

Database

Networking & Content Delivery

Security, Identity & Compliance

In AWS browser,

“Explore Our Products”

Compute -> Amazon EC2

Amazon EC2 page

Product details

Instance types

Pricing

Getting started

FAQs

Resources

Compute -> Amazon VPC (isolated cloud resources)

Storage ->

Database ->

Network & Content Delivery - > Amazon VPC

Security, Identity & Compliance - > AWS Identity & Access Management

**Documentation**

AWS (Navigation Bar) - > More - > Documentation

Amazon EC2 - > EC2 API Reference / EC2 CLI Reference

Categories and Services

http://aws.amazon.com

AWS GLOBAL INFRASTRUCTURE – ANNA FOX

3 topics:

Regions

Availability zones

Edge locations

AWS Home page - > scroll down - > “Global Network of Regions and Edge Locations

Regions are areas that have 2 or more availability zones

Choose which region will let you optimize latency, cost, regulatory requirements

If leveraging cloud computing services,

You can have an application near HQ, San Diego and deploy another application in the East Coast.

Let’s say customer base is in Virginia, choose the nearest region to optimize latency, agility and cost

Click “learn more” - > AWS Global Infrastructure -> “See detailed list of offerings at all AWS locations - > Region Table

AVAILABILITY ZONES

Collection of data centers within a region

Each are isolated to one another yet connected by a vast low latency link)

What is the benefit then?

When common points of failure occur, it does not affect all of your availability zones cause they are isolated.

How are they isolated?

Each of the availability zone are physically distinct independent infrastructure

Logically and physically separated

Each zone has their own discrete power supply, on-site backup generators, cooling equipment, and networking and connectivity.

They are also planned via different grids in independent utility companies

Connected to multiple Tier 1 transit providers

Isolation protects them from failures of other AZs

Best practice: multiple availability zones

EDGE LOCATIONS

Content delivery network – Amazon CloudFront

Deliver content to customers

Request for content is automatically routed to the nearest Edge Location

Content is delivered faster

Utilizing the global network locations and regions, you have access to quicker content delivery.

Edge locations are typically located in highly populated areas

Full list of edge locations: aws.amazon.com/cloudfront/details

AMAZON VIRTUAL PRIVATE CLOUD (VPC) – Kent Rademacher

Offers pay as you go, on demand compute as well as managed services

These compute resources must be available under normal IP protocols.

Implemented with familiar network structures.

Customers need to adhere networking best practices as well as meet regulatory and organizational requirements.

VPC - is the networking AWS service that will meet your networking requirements

A private, virtual network in the AWS Cloud

Uses same concepts as on premise networking

Allows complete control of network configuration

Ability to isolate and expose resources inside VPC

Offers several layers of security controls

Ability to allow and deny specific internet and internal traffic

Other AWS services deploy into VPC

Services inherent security built into network

VPC is a AWS foundational service. Integrates with numerous AWS service

For example, Amazon EC2 instances are deployed on Amazon VPC. Amazon Relational Database Services (RDS), database instance is deployed on VPC with database protected with network security just like on-premise network.

VPC Features

Builds upon high availability of AWS Regions and Availability Zones (AZ)

Amazon VPC lives within a Region

Multiple VPCs per account

Subnets (deployed in AZs) (able to create multiple subnets but will be complex)

Used to divide Amazon VPC

Allows Amazon VPC to span multiple AZs

Route tables (configurable)

Control traffic going out of the subnets

Internet Gateway (IGW) – for a subway to be public, attach IGW.

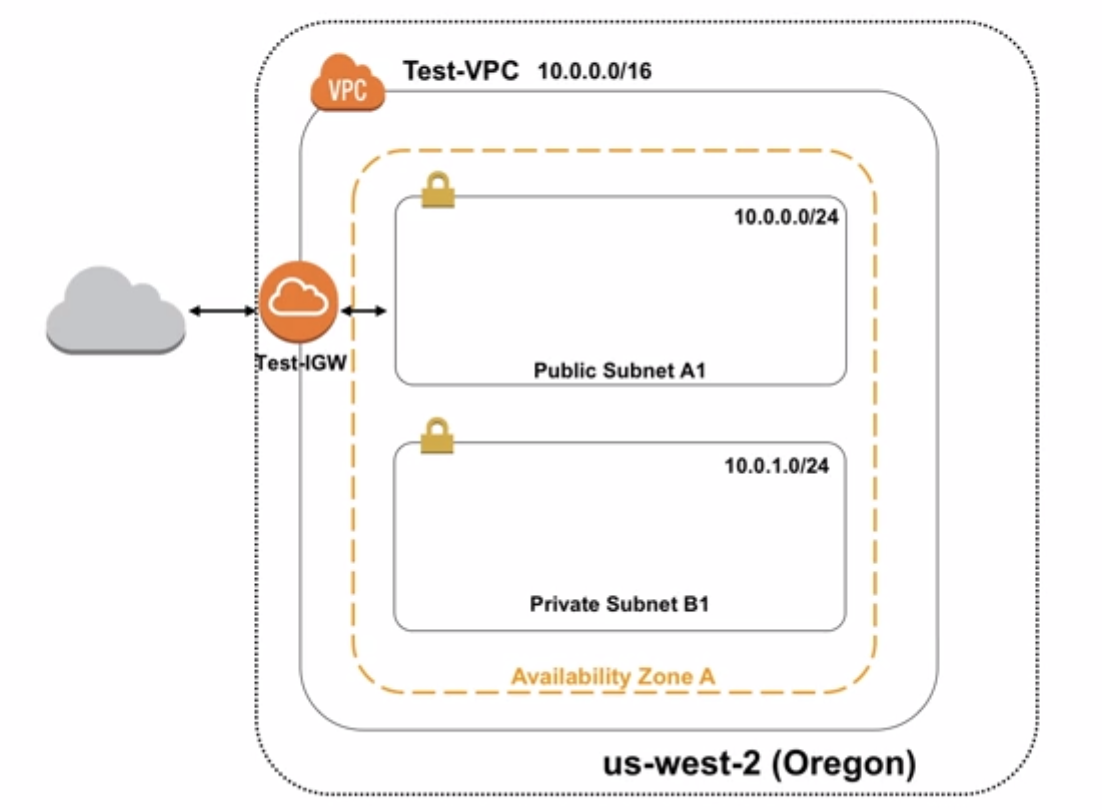
Allows access to the Internet from Amazon APC  
NAT Gateway

Allows private subnet resources to access Internet

Network Access Control Lists (NACL)

Control access to subnets; stateless

**Example VPC**



Select region “Oregon”

Name the VPC “Test-VPC”

VPC 10.0.0.0/16 = 65,000 IP address that can be used

Create new subnet

Subnet A1 10.0.0.0/24 256 IP addresses, Availability Zone A

Create another subnet

Subnet B1 10.0.2.0/23 512 IP address, Availability Zone A

Create Internet Gateway named “Test-IGW”

Subnet A1 – will become a public subnet where non-local traffic is routed through the IGW

Subnet B1 – will become a private subnet that is isolated from the Internet

**Summary**

What we accomplished

Created a VPC in the Oregon region

Created an Internet Gateway

Created one public subnets

Created one private subnets

Next steps

Learn about Route Tables and isolation methods

Learn about other Amazon VPC features like VPC Endpoints and Peering Connections

Learn about Security Groups

Learn about Amazon Elastic Cloud Compute (EC2)

Learn about Amazon Relational Database Service (RDS)

More info: aws.amazon.com/vpc/

AWS SECURITY GROUPS – Anna Fox

One of the highest priorities of AWS

Acts as built-in firewalls

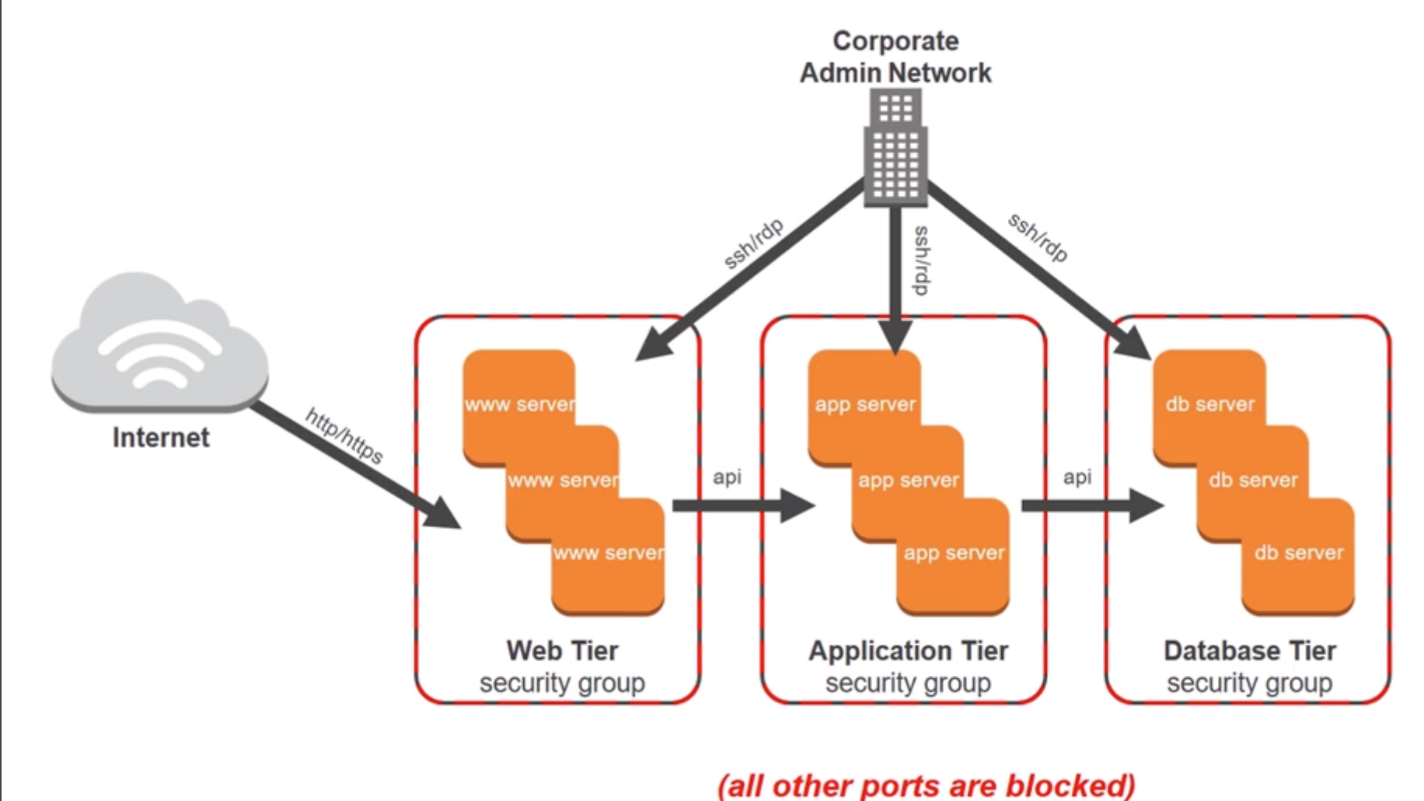
Accessibility

Provides control on what traffic you want to allow or deny

To determine who has access to your instances, you would configure a security group rule.

Rules can vary by giving instances private to totally public or somewhere in the middle.

Sample of AWS multiple Tier Security Group:



Multiple security group rule are created to accommodate multiple architecture.

If you started in the Web Tier you see that you have accepted traffic from anywhere in the Internet on port 80/443 by selecting the source of 0.0.0.0/0

If you move to the App Tier, it only accepts traffic from the Web Tier.

Similarly, the database Tier can only accept traffic from the Application Tier.

There is also a rule created to allow administration remotely from the corporate network over SSH port 22.

Creating a Security Group

AWS management console - > EC2 -> under navigation pane “Security Groups”.

“Create Security Group” - > Server group name “Web Server”, Description “Web Server Security Group”, VPC (choose a source).

Adjust rule by the drop down box under type. i.e., “HTTP”

Click “Add Rule”

AWS provides virtual firewalls that can control traffic for one or more instances in the area called Security Groups

You can control accessibility to your instances by creating Security Group Rules.

These Security Groups can be managed in the AWS Management Console.

as

as

as

as

as

as