

Cloud Computing

- On demand delivery of compute power, database storage, applications & other IT resources
- Pay-as-you-go pricing

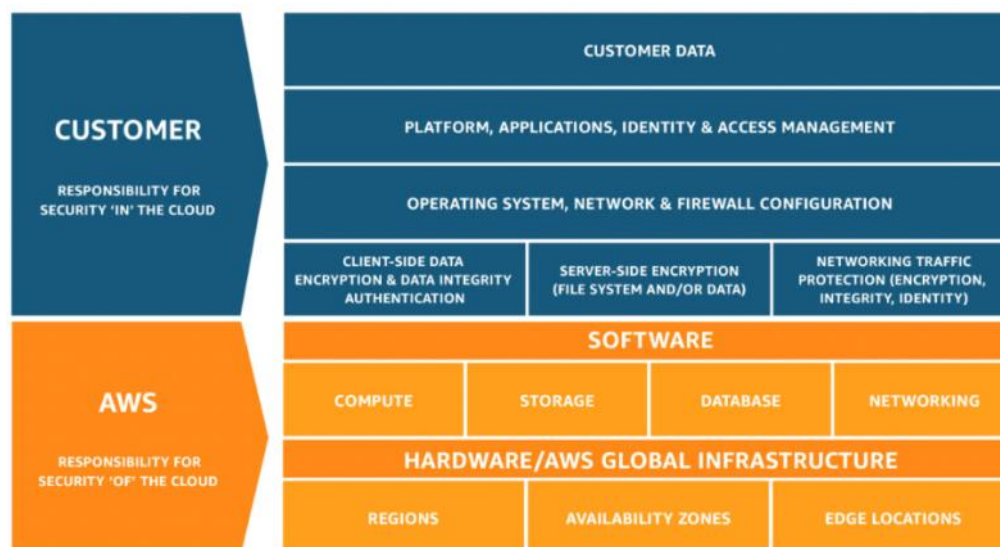
Shared Responsibility Model

All the services are mostly categorized in three types of models

- IaaS (Infrastructure as a Service)
- PaaS (Platform as a Service)
- SaaS (Software as a Service)
- And one regular model On-Premise (everything handled by consumer)

	Responsibility	SaaS	PaaS	IaaS	On-prem
Responsibility always retained by the customer	Information and data	Blue	Blue	Blue	Blue
	Devices (Mobile and PCs)	Blue	Blue	Blue	Blue
	Accounts and identities	Blue	Blue	Blue	Blue
Responsibility varies by type	Identity and directory infrastructure	Blue	Blue	Blue	Blue
	Applications	Blue	Blue	Blue	Blue
	Network controls	Blue	Blue	Blue	Blue
	Operating system	Blue	Blue	Blue	Blue
Responsibility transfers to cloud provider	Physical hosts	Blue	Blue	Blue	Blue
	Physical network	Blue	Blue	Blue	Blue
	Physical datacenter	Blue	Blue	Blue	Blue

S3
KMS
DynamoDB RDS EC2



Consumer is always responsible for,

- Information and data stored in cloud (**security in the cloud**)
- Devices that are allowed to connect to your cloud
- Accounts & identities of people, services, & devices

CP is always responsible for, (**security of the cloud**)

- Physical datacenter

- Physical network
- Physical hosts

Your service (service model) will determine responsibilities like,

- OS
- Network controls
- Applications
- Identity & infrastructure

Cloud Models

- It defines the deployment model of services

Private

- Cloud services used by a single organization not exposed to public
- Complete control
- Security for sensitive apps
- E.g. rackspace

Public

- Cloud resources owned & operated by third party CSP delivered over internet
- E.g. Azure, AWS, GCP

Hybrid

- Keep some servers on premise & extend some to cloud

Characteristics of Cloud Computing

- On-demand self service
- Broad network access
- Multi-tenancy & resource pooling
- Rapid elasticity & scalability
- Measured service
- Operational resilience

Consumption Based Model

- Cloud services work on consumption based model
- It falls under operational expenditure (OpEx), where-as on premise cloud model is capital expenditure (CapEx + OpEx)
- No upfront cost
- Pay only for what and how much you used
- Pay-as-you-go model
- **3 pricing fundamentals** of AWS cloud are **compute, storage, data transfer out** of AWS cloud

Metrics of Cloud

Availability

- Uptime
- High availability means running your application / system in at least 2 AZ
- It is achieved using Auto Scaling Group multi AZ & Load Balancer multi AZ

Scalability

- If suddenly traffic peaks up, you can add more resources to better handle the demand, vice-versa
- Vertical Scaling
 - o Increasing or decreasing capabilities of resources
 - o Like more/less CPUs or RAM to VM
 - o Scale up and down policy

From: t2.nano - 0.5G of RAM, 1 vCPU
To: u-12tb1.metal – 12.3 TB of RAM, 448 vCPUs

- Horizontal Scaling (**Elasticity**)
 - o Adding or subtracting number of resources
 - o Like adding/removing VMs
 - o Scale in and out policy
 - o Horizontal Scaling is achieved using Auto Scaling Group (ASG) & Load Balancer (LB)
- Scalability is like manual, Elasticity is like automatic

Agility

- Develop, test, deploy and configure cloud based resources quickly just on click by reducing time

Reliability (Durability)

- Ability of system to recover from failures and continue to function
- Decentralized design, enables you to have services in region around the world
- In some cases, your cloud env will automatically shift to different region for you, with no action needed from your side

Cost Predictability

- Forecasting cost using tools like Total Cost of Ownership (TCO) or Pricing Calculator

Fault Tolerance

- Ability to prevent a failure
- Failover to secondary system is done in case of failure in primary using Azure Traffic Manager

Sustainability

- Focuses on minimizing environmental impacts of running cloud workloads

Physical Infrastructure

Availability Zone

- Made up of **one or more data centers** equipped with independent power, cooling and networking
- AZ's are connected through high speed, private fiber optic networks, with low-latency network

Region

- It will contain **min 3 or max 6, AZ's**
- Few services or VM features are only available in specific region
- There are few services that don't require you to select particular region like Route 53, IAM

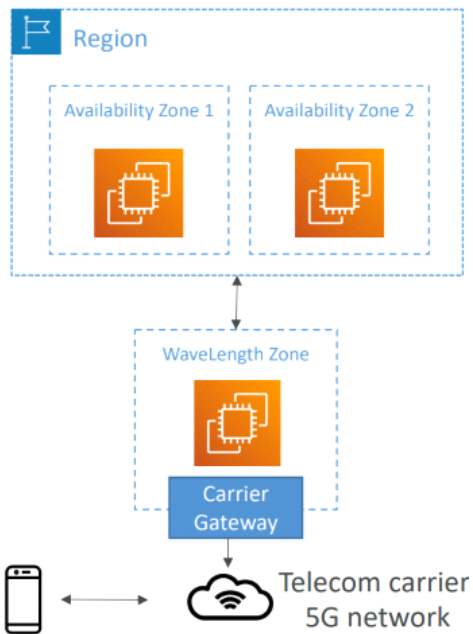


Edge Locations

- Nearest location, point of presence
- DC used to deliver content fast to users

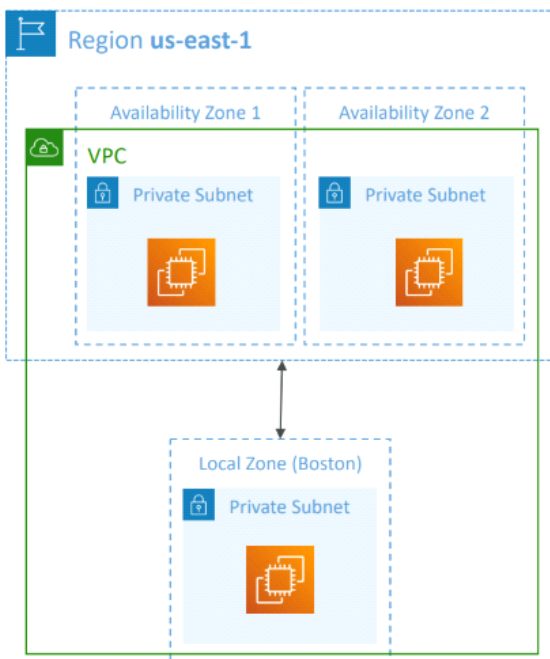
WaveLength Zone

- WaveLength Zones are infrastructure deployments embedded within the telecommunications providers datacenters at the edge of 5G networks
- Bring AWS services to the edge of the 5G networks
- Ultra-low latency applications



Local Zone

- Places AWS compute, storage, database and other selected services closer to end users to run latency sensitive applications
- Extend your VPC to more locations



Settings

Data protection and security

Zones

Default credit specification

EC2 Serial Console

EC2 console preferences

Zones (33)



Actions ▾

Switch regions to manage Zones for a different AWS region.

All Zones ▾

< 1 2 > ⚙

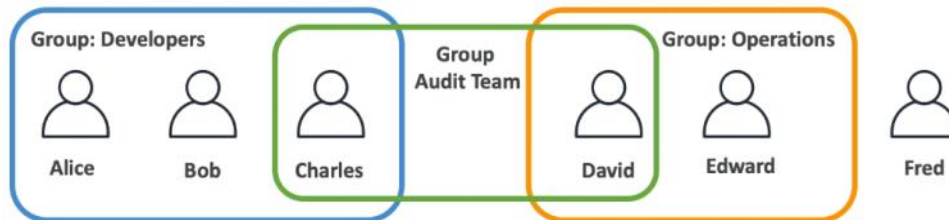
	Zone ID	Zone name ▾	Zone type ▾	Location ▲	State ▾	Opt-in status ▾	Network border
<input checked="" type="radio"/>	use1-az4	us-east-1a	Availability Zone	-	✓ Available	✓ Enabled by default	us-east-1
<input checked="" type="radio"/>	use1-az6	us-east-1b	Availability Zone	-	✓ Available	✓ Enabled by default	us-east-1
<input checked="" type="radio"/>	use1-az1	us-east-1c	Availability Zone	-	✓ Available	✓ Enabled by default	us-east-1
<input checked="" type="radio"/>	use1-az2	us-east-1d	Availability Zone	-	✓ Available	✓ Enabled by default	us-east-1
<input checked="" type="radio"/>	use1-az3	us-east-1e	Availability Zone	-	✓ Available	✓ Enabled by default	us-east-1
<input checked="" type="radio"/>	use1-az5	us-east-1f	Availability Zone	-	✓ Available	✓ Enabled by default	us-east-1
<input type="radio"/>	use1-bue1-az1	us-east-1-bue-1a	Local Zone	Argentina (Buenos Aires)	✓ Available	⊖ Disabled	us-east-1-bue-1
<input type="radio"/>	use1-wl1-atl-...	us-east-1-wl1-...	Wavelength Zone	Atlanta	✓ Available	⊖ Disabled	us-east-1-wl1-a

Identity Access & Management (IAM)

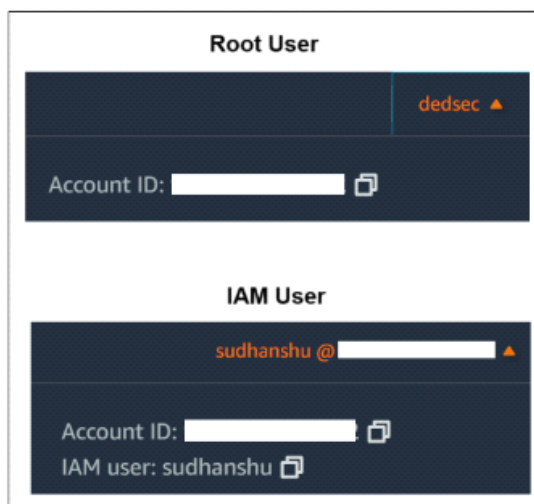
- Global service
- **Root account** created by default

IAM User & User Group

- **Users** are people within your organization, & **Group** can be created for them
- Similar to Linux users & groups



- Users & Groups are created to assign different kinds of permissions & access based on their role
- IAM User is always created in same account ID



- To sign in as IAM user, need three things
 - o Sign-in URL (optional)
 - o Account ID or Alias (can be created on IAM Dashboard)
 - o Username
 - o Password

IAM Policy

- Permissions are assigned in JSON document called policies, apply **least privilege principle**, don't give access more than needed
- Components in policy
 - o Version : policy language version
 - o Id : identifier for policy (optional)
 - o **Statement** : one or more individual statement (permissions)

- **Sid** : identifier for statement (optional)
- **Effect** : whether statement allows or denies access (Allow/Deny)
- **Principal** : account / user / role to which this statement is applied to
- **Action** : list of action this statement allows or denies
- **Resource** : list of resources to which actions are applied
- **Condition** : when this policy is in effect (optional)

```
{
  "Version": "2012-10-17",
  "Id": "S3-Account-Permissions",
  "Statement": [
    {
      "Sid": "1",
      "Effect": "Allow",
      "Principal": {
        "AWS": ["arn:aws:iam::123456789012:root"]
      },
      "Action": [
        "s3:GetObject",
        "s3:PutObject"
      ],
      "Resource": ["arn:aws:s3:::mybucket/*"]
    }
  ]
}
```

- When you create a user, it has to be assigned permissions via three types
 - Add user to existing group or create new one
 - Attach managed policies to user
 - Copy all group membership, attached policies
- You can create inline policy (separate policy) for the user

Permissions policies (4)

Permissions are defined by policies attached to the user directly or through groups.

Filter by Type

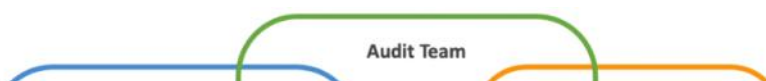
Search

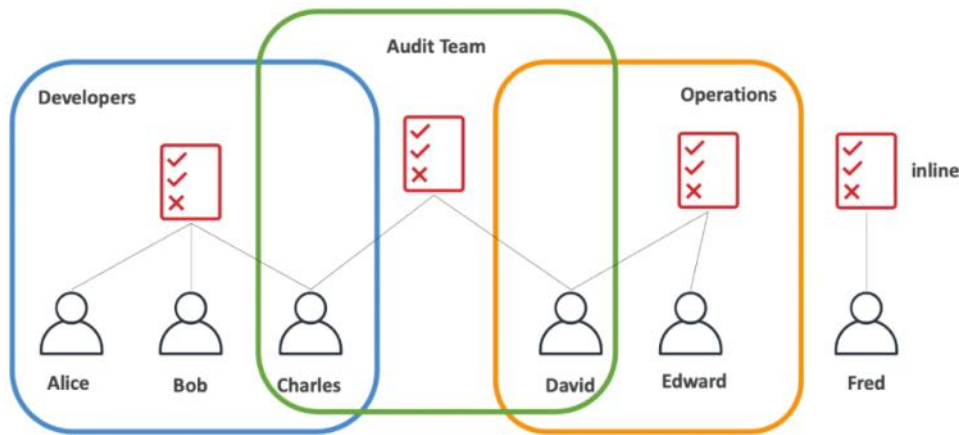
All types

< 1 > ⚙

<input type="checkbox"/>	Policy name	▲	Type ▼	Attached via
<input type="checkbox"/>	AdministratorAccess		AWS managed - job function	Group Admini
<input type="checkbox"/>	AlexaForBusinessDevice...		AWS managed	Group developers
<input type="checkbox"/>	IAMReadOnlyAccess		AWS managed	Directly
<input type="checkbox"/>	sudhanshu-inline-policy		Customer inline	Inline

- Policy Inheritance





- AWS provides many pre created policies

IAM Password Policy

- By default AWS applies IAM default policy
 - o Minimum length 8 characters
 - o Include min of three of mix of character types : Uppercase, Lowercase, Numbers, Non-alphanumeric characters
 - o Never expire password
 - o Must not be identical to your AWS account name or email address
- Can create custom password policy as well

Password policy

☐ IAM default
Apply default password requirements.

☒ Custom
Apply customized password requirements.

Password minimum length.
Enforce a minimum length of characters.

characters

Needs to be between 6 and 128.

Password strength

☐ Require at least one uppercase letter from the Latin alphabet (A-Z)

☐ Require at least one lowercase letter from the Latin alphabet (a-z)

☐ Require at least one number

☐ Require at least one non-alphanumeric character (! @ # \$ % ^ & * () _ + - = [] { } | ')

Other requirements

☐ Turn on password expiration

☐ Password expiration requires administrator reset

☐ Allow users to change their own password

☐ Prevent password reuse

IAM MFA

- AWS provides three ways to manage MFA
 - o Passkey or security key (fingerprint, face, screen lock or passkey)
 - o Authenticator app
 - o Hardware TOTP token

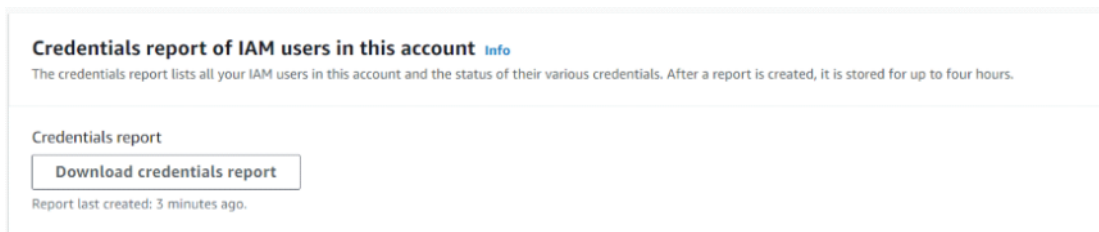
- All the configuration related to MFA can be managed in User > Security credentials

IAM Roles

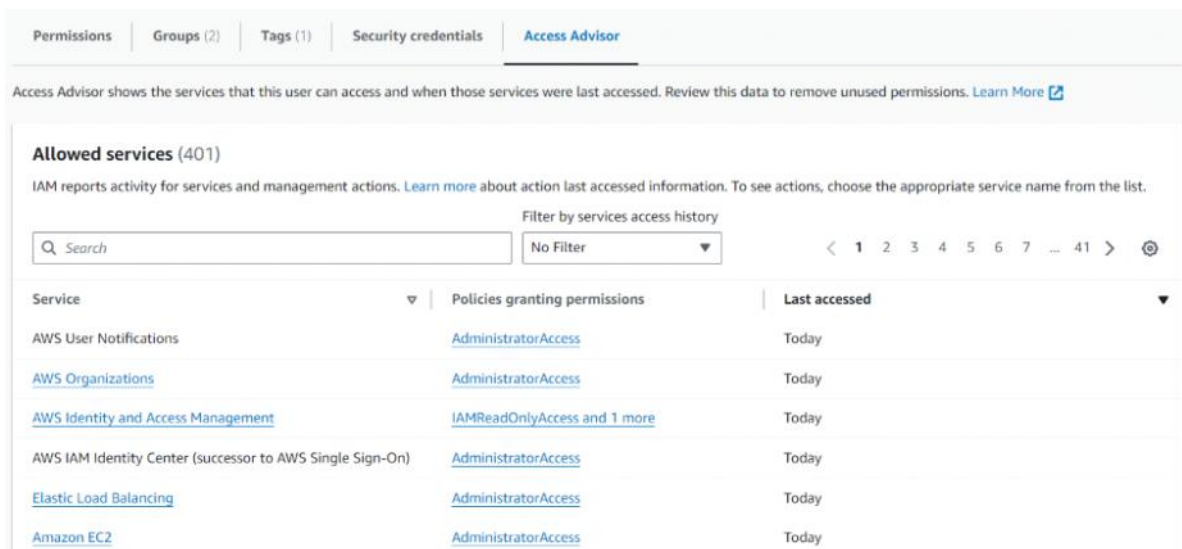
- Some AWS services will need to perform actions on your behalf
- To do so, we will assign permissions to AWS services with IAM Roles
- Common roles are
 - o EC2 Instance Roles
 - o Lambda Function Roles
 - o Roles for CloudFormation

IAM Security Tools

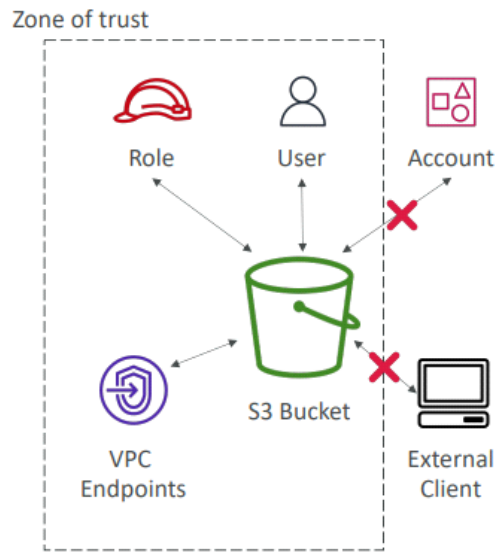
- **Credentials Report**
 - o Account level
 - o Report that list all your users & along with all the credentials details including MFA, access keys



- **Access Advisor**
 - o User level
 - o It shows the service permissions granted to a user & when those services were last accessed
 - o Review this data to revise permissions



- **Access Analyzer**
 - o Find out resources which are shared externally
 - o Define Zone of Trust
 - o Access outside ZoT are findings



AWS Access

AWS Management Console

- Protected by password + MFA

AWS CloudShell

- Terminal inside AWS, to access AWS CLI & all bash commands in CloudShell environment
- No need of access key explicitly
- Session of logged in user is used

AWS CLI

- CLI tool to manage all the AWS services through
- AWS CLI is built using AWS SDK of python which is boto3
- To use access key in AWS CLI need to run **aws configure**
- Protected by access keys

```
~$ aws configure
AWS Access Key ID [None]: 
AWS Secret Access Key [None]: 
Default region name [None]: eu-west-1
Default output format [None]:
```

AWS SDK

- Embed within your applications
- Supports all the languages
- Protected by access keys
- Access keys are generated through AWS console (User -> Security credentials)
- Users manage their own access keys

Elastic Cloud Compute (EC2)

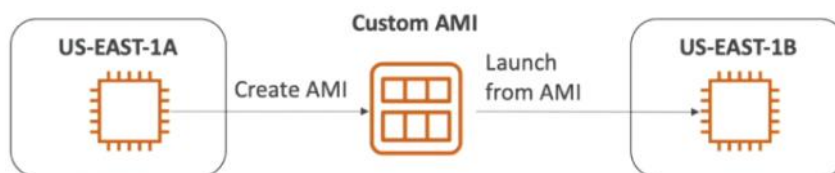
- IaaS
- Virtual server

Configuration options

- OS or Amazon Machine Image (AMI)
 - o Linux, Windows or MacOS
- CPU
- RAM
- Storage
 - o EBS & EFS
 - o EC2 Instance Store
- Network
 - o Speed of card
 - o Public IP
- Security Group
- EC2 User Data
 - o Bootstrapping (running commands) when a machine starts
 - o This script runs only once after creation of EC2
 - o Used to automate tasks like installing updates, software, downloading files
 - o Runs with sudo user

Amazon Machine Image (AMI)

- Template of customized EC2 instance
- Built for specific region, can be copied across regions
- While building AMI, EBS snapshots are also created
- EC2 instance can be launched from
 - o Public AMI : AWS provided
 - o Own AMI
 - o AWS Marketplace AMI : Vendors on AWS Marketplace



EC2 Instance Types

- There are many EC2 instance types categorized based on configurations
 - o General Purpose
 - o Compute Optimized
 - o Memory Optimized
 - o Accelerated Computing
 - o Storage Optimized
 - o HPC Optimized
 - o Instance Features

- Measuring Instance Performance
- General Purpose
 - M & T series
 - Web servers or code repositories
 - **Balance** between compute, network, memory
- Compute Optimized
 - C series
 - Compute intensive tasks, that require **high performance processors**
- Memory Optimized
 - R & X series
 - Fast performance for **processing large data sets** in memory
- Storage Optimized
 - I & D series
 - Require **high, sequential read & write** access to large data sets on local storage
 - Suitable for **data warehousing** applications
- Accelerated Computing
 - Best to perform calculations that include floating point numbers
 - Graphics processing
 - Data pattern matching
- HPC Optimized
- Naming convention

m5.2xlarge

- m: instance class
- 5: generation (AWS improves them over time)
- 2xlarge: size within the instance class

Instance	vCPU	Mem (GiB)	Storage	Network Performance	EBS Bandwidth (Mbps)
t2.micro	1	1	EBS-Only	Low to Moderate	
t2.xlarge	4	16	EBS-Only	Moderate	
c5d.4xlarge	16	32	1 x 400 NVMe SSD	Up to 10 Gbps	4,750
r5.16xlarge	64	512	EBS Only	20 Gbps	13,600
m5.8xlarge	32	128	EBS Only	10 Gbps	6,800

- t2.micro is free, up to 750 hours per month

Security Groups

- Acts as firewall for EC2
- Controls which kind of traffic is allowed in & out of EC2 instance, by regulating
 - Ports
 - IP ranges
 - Inbound & Outbound rules
- SG only contain allow rules
- SG can be attached to multiple EC2 & EC2 can have multiple SG

Inbound rules

Outbound rules

Tags

Inbound rules (2)

↺

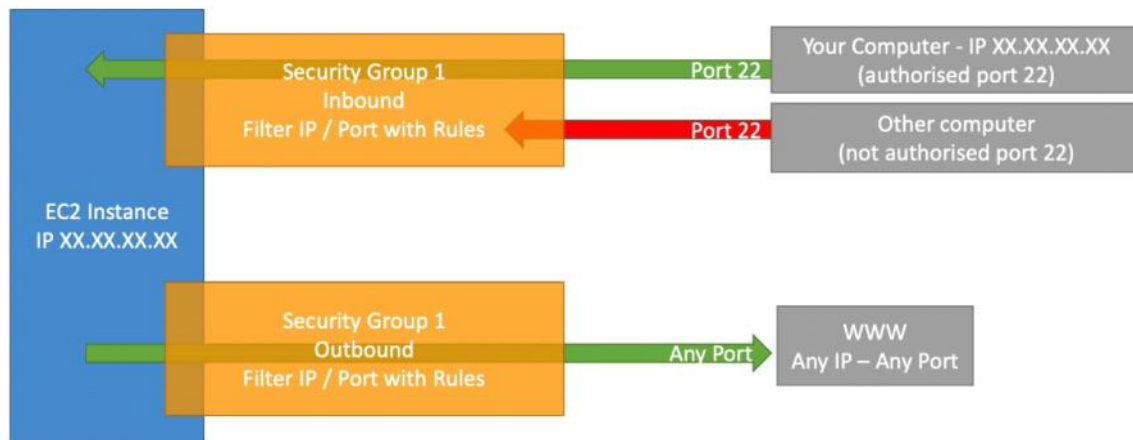
Manage tags

Edit inbound rules

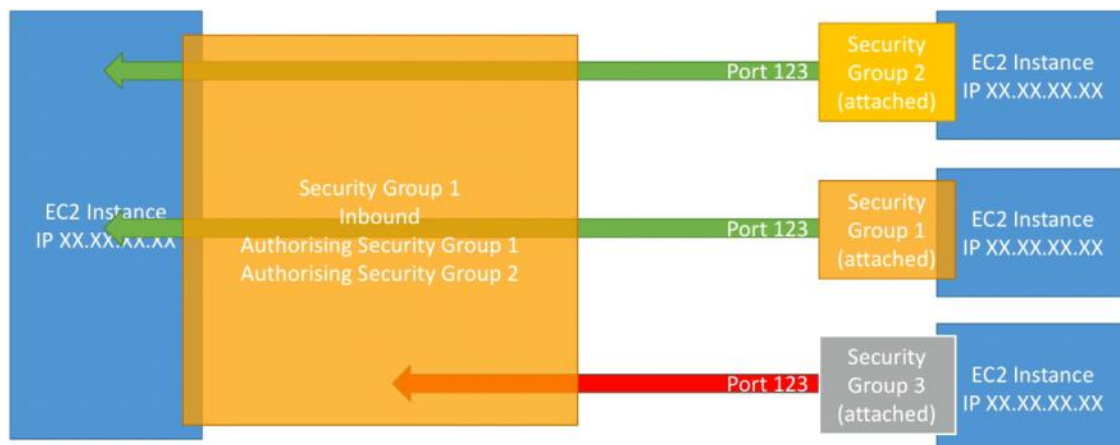
🔍 Search

< 1 > ⚙

<input type="checkbox"/>	Name	Security group rule...	IP version	Type	Protocol
<input type="checkbox"/>	-	sgr-02eb08c62834b12...	IPv4	HTTP	TCP
<input type="checkbox"/>	-	sgr-04268b038ea713...	IPv4	SSH	TCP



- SG can reference other SG as well



- Ports
 - o SSH : 22
 - o RDP : 3389
 - o FTP : 21
 - o SFTP : 22
 - o HTTP : 80
 - o HTTPS : 443

EC2 Access

	Mostly CLI SSH	Windows Application Putty	Browser Terminal EC2 Instance Connect
Mac	✓		✓
Linux	✓		✓
Windows < 10		✓	✓
Windows >= 10	✓	✓	✓

- Command to connect EC2 using SSH

```
ssh -i <pem file> <user>@<public IP>
```

- To use AWS CLI inside EC2, it is always recommended to use IAM Roles, and not aws configure directly

EC2 Instances Options

- On-Demand Instances
 - o Normal EC2
 - o Good for short-term & uninterrupted workload
 - o Pay by second for Linux & Windows, for others pay per hour
 - o Highest cost but no upfront payment
 - o No long term commitment
- Reserved Instances
 - o Long workloads
 - o It has reservation period of 1 year or 3 years
 - o Payment options : No Upfront, Partial Upfront, All Upfront
 - o Good discount compared to On-Demand based on reservation period & payment option
 - o Can buy or sell it in reserved instances marketplace
 - o Convertible Reserved Instances can change instance type, family, OS, scope & tenancy
- Savings Plans
 - o It has reservation period of 1 year or 3 years
 - o Commitment to an amount of usage
 - o Long workload
 - o Specific instance family & region
 - o Flexible in instance size, OS, tenancy (Host, Dedicated, Default)
- Spot Instances
 - o Maximum discount, 90% compared to On-demand
 - o Short workloads
 - o Cheap

- Can lose instances (less reliable)
- Dedicated Hosts
 - Book entire physical server
 - Most expensive
 - Two types : On-demand or Reserved
 - Control instance placement
- Dedicated Instances
 - No other customer will share your hardware
- Capacity Reservations
 - Reserve capacity in a specific AZ for any duration
 - No time commitment
 - No billing discount
 - Combine with regional reserved instances & saving plan to get discount



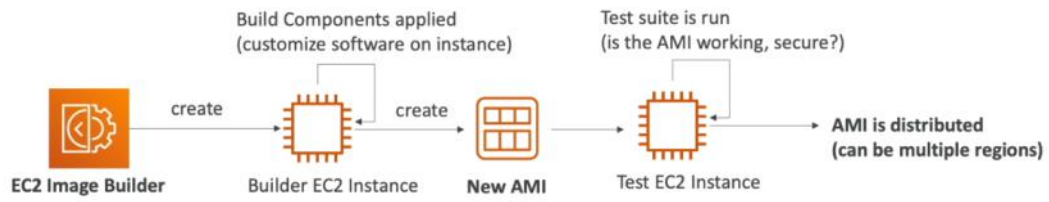
- On demand: coming and staying in resort whenever we like, we pay the full price
- Reserved: like planning ahead and if we plan to stay for a long time, we may get a good discount.
- Savings Plans: pay a certain amount per hour for certain period and stay in any room type (e.g., King, Suite, Sea View, ...)
- Spot instances: the hotel allows people to bid for the empty rooms and the highest bidder keeps the rooms. You can get kicked out at any time
- Dedicated Hosts: We book an entire building of the resort
- Capacity Reservations: you book a room for a period with full price even you don't stay in it

Example – m4.large – us-east-1

Price Type	Price (per hour)
On-Demand	\$0.10
Spot Instance (Spot Price)	\$0.038 - \$0.039 (up to 61% off)
Reserved Instance (1 year)	\$0.062 (No Upfront) - \$0.058 (All Upfront)
Reserved Instance (3 years)	\$0.043 (No Upfront) - \$0.037 (All Upfront)
EC2 Savings Plan (1 year)	\$0.062 (No Upfront) - \$0.058 (All Upfront)
Reserved Convertible Instance (1 year)	\$0.071 (No Upfront) - \$0.066 (All Upfront)
Dedicated Host	On-Demand Price
Dedicated Host Reservation	Up to 70% off
Capacity Reservations	On-Demand Price

EC2 Image Builder

- Used to automate creation of VM or container images
- Automate creation, maintain, validate & test EC2 AMIs
- Can be run on schedule (weekly, whenever package is updated, etc.)
- Free service, only pay for underlying service



Elastic Block Storage (EBS)

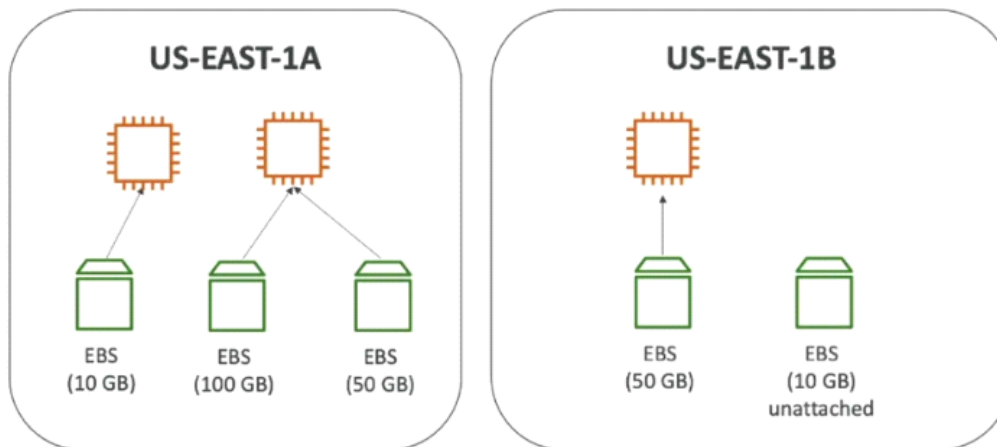
- EBS volume is network drive (think it as pen drive) you can attach it to your instances while running
- It can persist data, even after instance is terminated
- Volumes can be deleted on termination of instance by Delete on Termination attribute, this can be updated through AWS CLI only

```
aws ec2 modify-instance-attribute --instance-id i-08fc67f21b9314df4 --block-device-mappings "[{"DeviceName": "/dev/sdb", "Ebs": {"DeleteOnTermination": false}}]"
```

▼ Block devices

Volume ID	Device name	Volume siz...	Attachmen...	Attachmen...	Encrypted	KMS key ID	Delete on termination
vol-0eb9f8...	/dev/xvda	8	✓ Attached	2024/08/0...	No	-	Yes
vol-094a31...	/dev/sdb	5	✓ Attached	2024/08/0...	No	-	No

- Can be mounted to only one EC2 at a time (except io1 & io2 volumes, they support multi attach)
- AZ specific, can't be attached to EC2 in different AZ directly (can be done through snapshot)

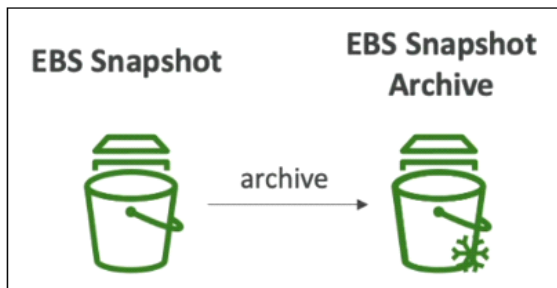


EBS Snapshot

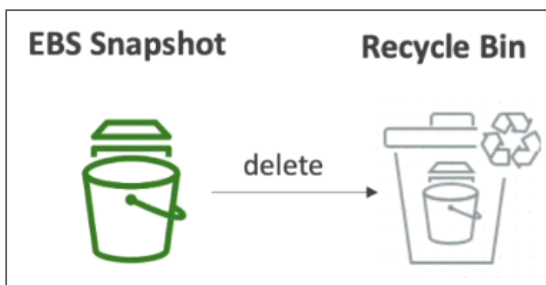
- Snapshot is same as backup at a point in time recovery
- To take snapshot it is recommended to detach volume, not necessary
- Snapshot can be copied across AZ or Regions
- Volume can be created using snapshot



- Archive
 - o Move snapshot to archive tier which is 75% cheaper
 - o Takes 24 to 72 hours to restore the archive

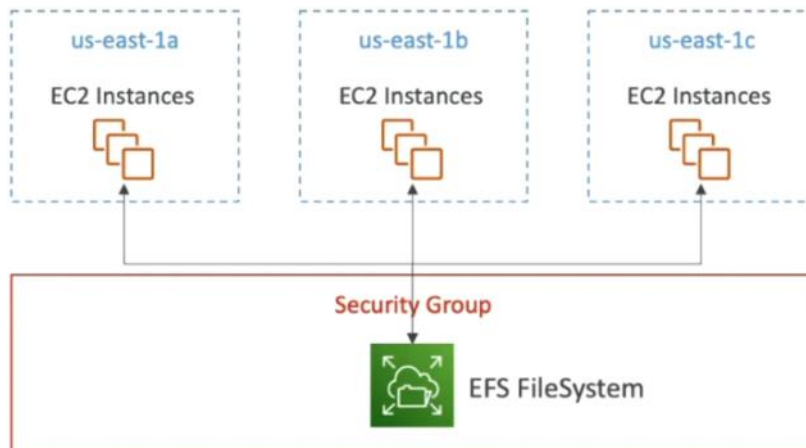


- Recycle bin
 - o Setup rules to retain deleted snapshots so you can recover them after an accidental deletion
 - o Specify retention from 1 day to 1 year



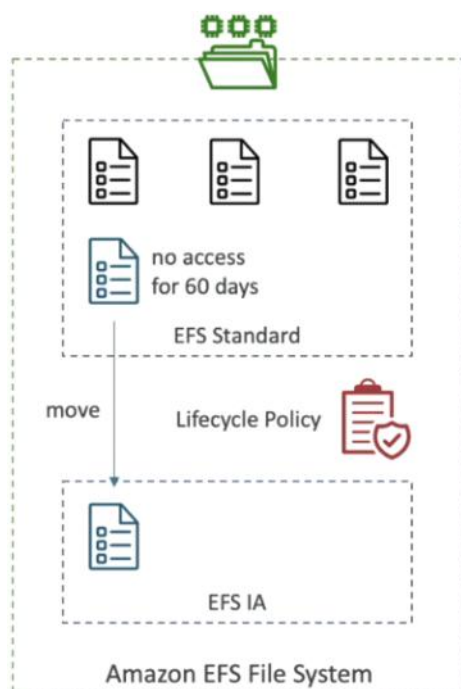
Elastic File Storage (EFS)

- Network file system
- Can be mounted to multiple EC2
- Works only with Linux EC2
- Multiple AZ
- Highly available, scalable
- Expensive
- Pay per use
- No capacity provisioning



EFS Infrequent Access (EFS-IA)

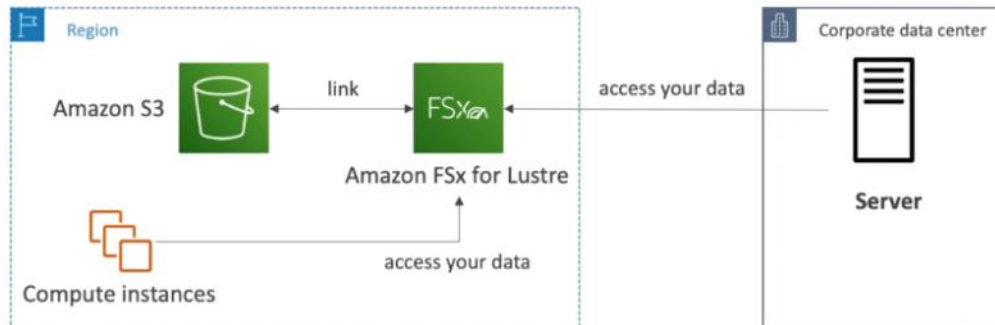
- Storage class that is cost optimized for files that are not accessed every day
- Up to 92% lower cost compared to EFS Standard
- EFS will automatically move files to EFS-IA based on last time it was accessed
- Enabled EFS-IA using lifecycle policy



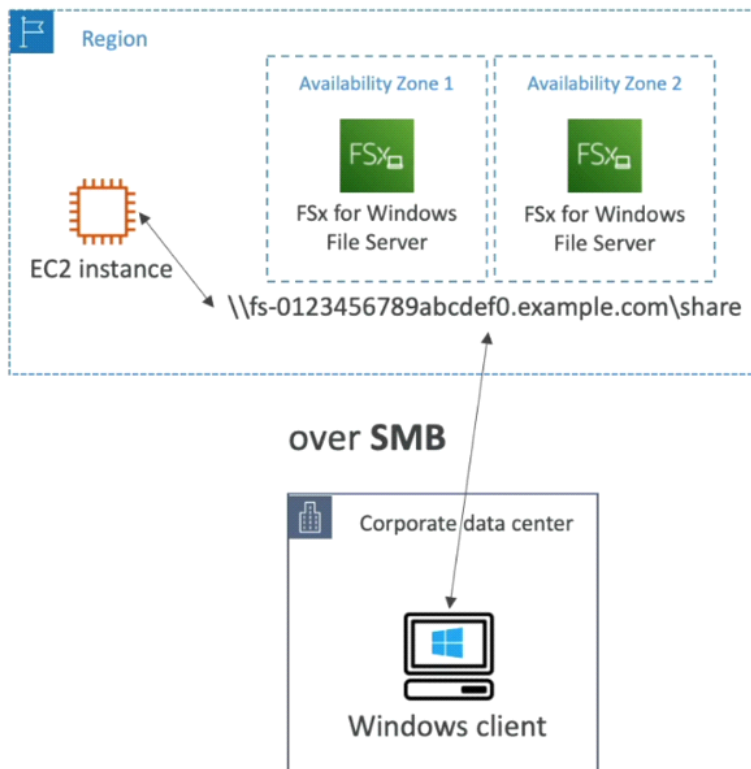
Amazon FSx

- Third party high performance file system on AWS
- Fully managed service

FSx for Lustre (High Performance Computing)



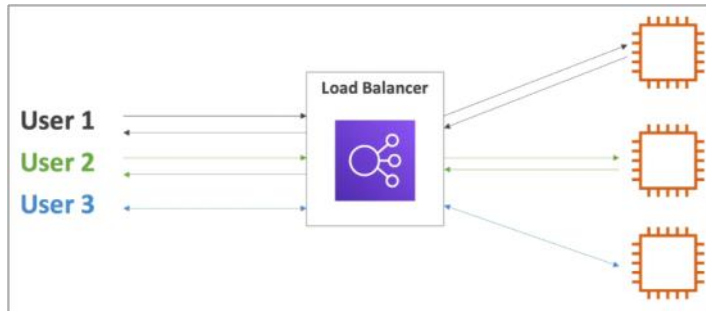
FSx for Windows File Server (over SMB, integrated with MS Active Directory)



FSx for NetApp ONTAP

Elastic Load Balancer (ELB)

- Load balancer is server that forward internet traffic to multiple servers (EC2) based on some condition
- Helps in spreading load across multiple servers
- Expose single point of access (DNS) of your application



ELB

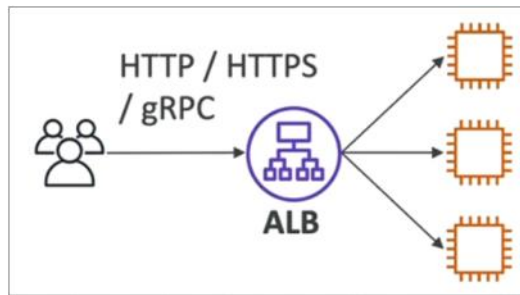
- AWS managed LB

Load balancers (1/1)						Actions	Create load balancer
Elastic Load Balancing scales your load balancer capacity automatically in response to changes in incoming traffic.							
Filter load balancers							
<input checked="" type="checkbox"/>	Name	DNS name	State	VPC ID	Availability Zones		
<input checked="" type="checkbox"/>	ALB	ALB-200877346.ap-south...	Active	vpc-0bc3e1fab83ac6e8b	3 Availability Zones		

- We need to link Target Group (TG) to ELB
- TG is a group of instances or lambda functions or ALB or IP address
- If any instance in TG is stopped or down, it is treated as unhealthy and ELB will not redirect traffic to that instance

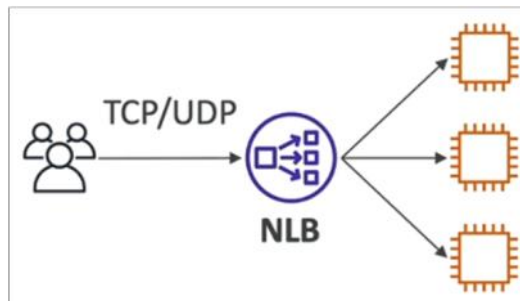
Instance ID	Name	Port	Zone	Health ...	Health status details
i-0c1cb160b3...	ELB-demo	80	ap-southe...	Healthy	-
i-0b50639bda...	ELB-demo	80	ap-southe...	Unused	Target is in the stopped state

- 3 types of ELB
 - o Application
 - o Network
 - o Gateway
- **Application Load Balancer (ALB)**
 - o HTTP / HTTPS / gRPC protocols
 - o Layer 7
 - o HTTP routing features
 - o Static DNS (URL)



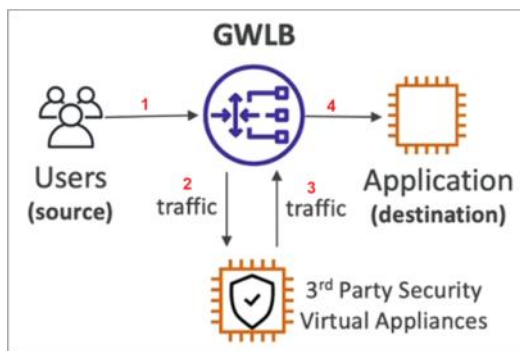
- **Network Load Balancer (NLB)**

- TCP / UDP protocols
- Layer 4
- Ultrahigh performance, millions of request per seconds
- Static IP through Elastic IP



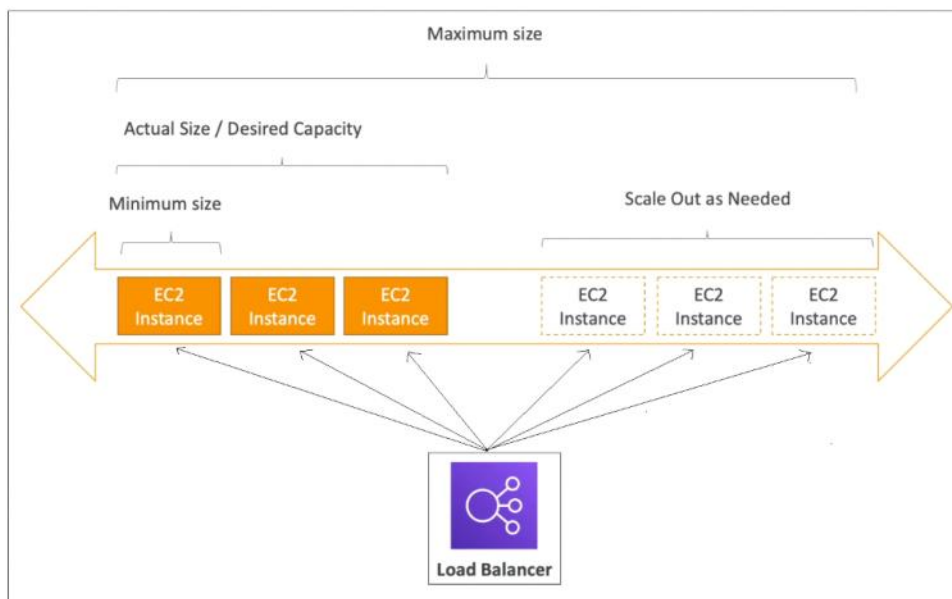
- **Gateway Load Balancer (GLB)**

- GENEVE protocol on IP packets
- Layer 3
- Route traffic to firewalls that you manage on EC2 instances
- Intrusion detection



Auto Scaling Group (ASG)

- Incoming traffic on your application can change in real time due to many factors
- You can create and get rid of servers very quickly using ASG
- Integrated with ELB
- Offers horizontal scaling
- It needs an EC2 instance launch template
- Scale out (add EC2) to match an increased load
- Scale in (remove EC2) to match a decreased load
- Ensure min and max number of instances are running
- Automatically register new instances to load balancer
- Replace unhealthy instances with new instances



Scaling Strategies

- **Manual**
 - o Update the size of an ASG manually
- **Dynamic**
 - o Respond to changing demand
 - o **Simple / Step** scaling : ex. based on CloudWatch alarm, CPU utilization
 - o **Target Tracking** scaling : ex. I want avg ASG CPU to be 40%
 - o **Scheduled** scaling : based on date and time
 - o **Predictive** Scaling : uses ML to predict future traffic AOT

Simple Storage Service (S3)

- Storage service for backup, disaster recovery, archive, application hosting, media hosting, data lakes, software delivery, static website
- It store objects (files) in buckets (directories)
- High durability (11 nines), same for all storage classes
- Availability (4 nines)

S3 Bucket

- Bucket should have globally unique name, across all regions all accounts
- Buckets are defined at region level
- Naming convention
 - o No uppercase
 - o No underscore
 - o 3-63 characters
 - o Must start with lowercase letter or number
 - o Must not start with prefix xn--
 - o Must not end with suffix -s3alias

S3 Objects

- Objects (files) have a key (path of that file)
- There is no concept of directories inside bucket, but UI talks different, hierarchy is completely based on keys
- Max object size is 5TB
- Version ID
- Metadata
- Tags

S3 Security

- User Based
 - o IAM Policies : which API calls should be allowed for specific user from IAM
- Resource Based
 - o Bucket Policies : bucket wide rules, allow cross account
 - o Object ACL - finer grain (can be disabled)
 - o Bucket ACL - less common (can be disabled)
- Encryption
 - o Encrypt objects in S3 using encryption keys

S3 Bucket Policy

- JSON based policies
- Similar to IAM policy
- Examples
 - o Public access : bucket policy
 - o IAM user access to S3 : IAM permissions in IAM policy
 - o EC2 instances access to S3 : IAM roles
 - o Cross account access to S3 : bucket policy
- If bucket settings for block public access are enabled, then no matter what bucket policy it has, public access will always be blocked

Block all public access

On

Block public access to buckets and objects granted through *new* access control lists (ACLs)

On

Block public access to buckets and objects granted through *any* access control lists (ACLs)

On

Block public access to buckets and objects granted through *new* public bucket or access point policies

On

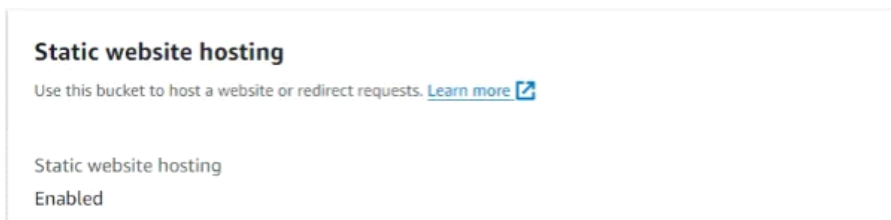
Block public and cross-account access to buckets and objects through *any* public bucket or access point policies

On

```
{
  "Version": "2012-10-17",
  "Id": "Policy1723620450464",
  "Statement": [
    {
      "Sid": "Stmt1723620449049",
      "Effect": "Allow",
      "Principal": "*",
      "Action": "s3:GetObject",
      "Resource": "arn:aws:s3:::demo-sudhanshu-s3/*"
    }
  ]
}
```

S3 Static Website Hosting

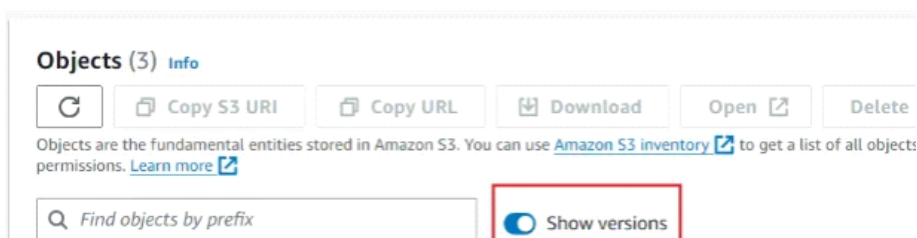
- To use S3 for hosting the static website, you need to enable it first in properties



- And upload the index.html file

S3 Versioning

- Versioning a file is possible in S3
- Enabled at bucket level
- Same key overwrite will change the version
- Easy roll back to previous version



Objects (3) [Info](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects permissions. [Learn more](#)

- Any object deleted, will not be deleted permanently (delete marker), possible to restore due to versioning, but version can be deleted permanently
- If version is deleted, then object will roll back to previous version
- Upload

<input type="checkbox"/>	index.html	html	nXhrE5mITORceK5vEZDu KvB1kTs21EFn	August 14, 2024, 13:18:41 (UTC+05:30)	438.0 B	Standard
<input type="checkbox"/>	index.html	html	dRDN_4_t1NHL31ML_RN CR0xUBRj5Dp5z	August 14, 2024, 13:18:29 (UTC+05:30)	438.0 B	Standard

- Delete

<input type="checkbox"/>	index.html	Delete marker	OPZ1UptsKHaJpq_zr6Nu4oK G2sYY8Nlz	August 14, 2024, 13:16:42 (UTC+05:30)	0 B	-
<input type="checkbox"/>	index.html	html	S5k6s2.yWFdaQKMyF7zNLg5 dvn8RmCTH	August 14, 2024, 13:16:33 (UTC+05:30)	438.0 B	Standard

S3 Replication

- Must have versioning enabled in source and destination buckets
- Buckets can be in different AWS accounts
- Copying is asynchronous
- Must give proper IAM permission to buckets
- Need to create replication rule in source bucket under management

Replication rules (1)

Use replication rules to define options you want Amazon S3 to apply during replication such as server-side encryption, replica ownership, transitioning replicas to another storage class, and more. [Learn more](#)

Replication rule name	Status	Destination bucket	Destination Region	Priority	Scope	Storage class	Replica owner	Replication Time Control	KMS-encryption object (SSE-I or DS-KMS)
<input type="radio"/> demo-replication	Enabled	s3://demo-replication-sudhanshu-2	Asia Pacific (Sydney) ap-southeast-2	0	Entire bucket	Same as source	Same as source	Disabled	Do no replic

- **Cross Region Replication (CRR)**
 - o Compliance
 - o Low latency access
 - o Replication across accounts
- **Same Region Replication (SRR)**

- Log aggregation
- Live replication between prod and test accounts

S3 Storage Classes

- Applied on object level
- Can be changed
- lifecycle rules are used to define actions for objects such as changing storage class, archiving, deleting after a specified period of time
- **Standard - General Purpose**
 - 4 nines of availability
 - Used for frequently accessed data
 - Low latency
 - High throughput
 - Big data analytics, mobile & gaming application, content distribution
- Infrequent Access (IA)
 - **Standard IA**
 - 3 nines of availability
 - Less frequently accessed
 - Require rapid access when needed
 - Lower cost than S3 Standard
 - Disaster recovery, backups
 - **One Zone IA**
 - 11 nines of durability
 - 99.5% of availability
 - Data lost when AZ is destroyed
 - Storage for type of data which can be created in case of loss
- **Glacier**
 - Low cost object storage meant for archiving / backup
 - Pricing : price for storage + object retrieval cost
 - **Glacier Instant Retrieval**
 - Millisecond retrieval, great for data accessed once a quarter
 - Minimum storage duration of 90 days
 - **Glacier Flexible Retrieval**
 - Expedited (1 to 5 min), Standard (3 to 5 hours), Bulk (5 to 12 hours) - free
 - Minimum storage duration of 90 days
 - **Glacier Deep Archive - long term storage**
 - Standard (12 hours), Bulk (48 hours)
 - Minimum storage duration of 180 days
- **Intelligent Tiering**
 - Small monthly monitoring & auto tiering fee
 - Moves object automatically between access tiers based on usage
 - No retrieval charges in S3 IT

- Frequent AT : default tier
- IA AT : objects not accessed for 30 days
- Archive Instant AT : objects not accessed for 90 days
- Archive AT : configurable from 90 days to 700+ days
- Deep Archive AT : configurable form 180 days to 700+ days

	Standard	Intelligent-Tiering	Standard-IA	One Zone-IA	Glacier Instant Retrieval	Glacier Flexible Retrieval	Glacier Deep Archive
Durability	99.999999999% == (11 9's)						
Availability	99.99%	99.9%	99.9%	99.5%	99.9%	99.99%	99.99%
Availability SLA	99.9%	99%	99%	99%	99%	99.9%	99.9%
Availability Zones	>= 3	>= 3	>= 3	1	>= 3	>= 3	>= 3
Min. Storage Duration Charge	None	None	30 Days	30 Days	90 Days	90 Days	180 Days
Min. Billable Object Size	None	None	128 KB	128 KB	128 KB	40 KB	40 KB
Retrieval Fee	None	None	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved


	Standard	Intelligent-Tiering	Standard-IA	One Zone-IA	Glacier Instant Retrieval	Glacier Flexible Retrieval	Glacier Deep Archive
Storage Cost (per GB per month)	\$0.023	\$0.0025 - \$0.023	\$0.0125	\$0.01	\$0.004	\$0.0036	\$0.00099
Retrieval Cost (per 1000 request)	GET: \$0.0004 POST: \$0.005	GET: \$0.0004 POST: \$0.005	GET: \$0.001 POST: \$0.01	GET: \$0.001 POST: \$0.01	GET: \$0.01 POST: \$0.02	GET: \$0.0004 POST: \$0.03 Expedited: \$10 Standard: \$0.05 Bulk: free	GET: \$0.0004 POST: \$0.05 Standard: \$0.10 Bulk: \$0.025
Retrieval Time	Instantaneous					Expedited (1 – 5 mins) Standard (3 – 5 hours) Bulk (5 – 12 hours)	Standard (12 hours) Bulk (48 hours)
Monitoring Cost (pet 1000 objects)		\$0.0025					


S3 Encryption



- Server Side Encryption (default)
 - AWS encrypts file after receiving it
- Client Side Encryption
 - Client will encrypt the file and upload it to AWS





IAM Access Analyzer for S3

- Ensures that only intended people have access to your S3 buckets
- Evaluates S3 bucket policies, S3 ACLs, S3 Access Point Policies
- Powered by IAM Access Analyzer

 **Buckets with public access (4)**
 These buckets can be accessed by anyone on the internet. Unless you require a public configuration for a specific and verified use case, AWS recommends that you block all public access to your buckets. [Learn more](#)

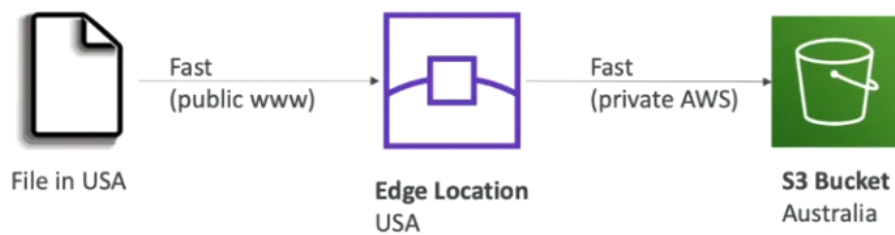
View findings  Mark as active Archive Block all public access

Status: All  1 

	Bucket name	Discovered by Access Analyzer	Shared throu...	Status	Access le...
<input type="radio"/>		2 minutes ago	Bucket policy	Active	Read
<input type="radio"/>		2 minutes ago	Bucket policy	Active	Read
<input type="radio"/>		2 minutes ago	Bucket policy	Active	Read
<input type="radio"/>		2 minutes ago	Bucket policy	Active	Read

S3 Transfer Acceleration

- Increase transfer speed by transferring file to an AWS nearest edge location which will forward data to S3 bucket in the target location



Snow Family

- Highly secure, portable devices to collection and process data at the edge and migrate data into and out of AWS
- Job types
 - o Import into S3
 - o Export from S3
 - o Local compute and storage only
 - o Import virtual tapes into AWS Storage Gateway




Snowcone



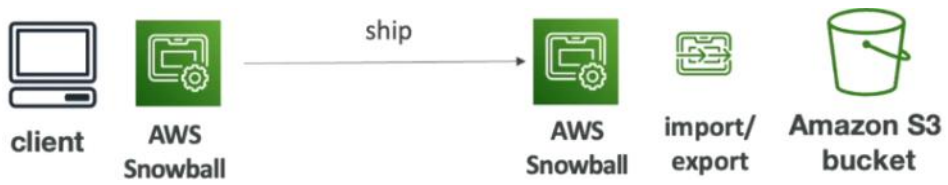

Snowball Edge

	Snowcone	Snowball Edge
Storage Capacity	8 TB HDD - 14 TB SSD	80 TB - 210 TB
Migration Size	Up to terabytes	Up to petabytes

1 Exabyte = 1000 Petabytes = 1000000 Terabytes

- Snow devices
 - o Snowcone
 - o Snowcone SSD
 - o Snowball Edge storage optimized
 - o Snowball Edge compute optimized
 - o Snowball Edge compute optimized with GPU
 - o Snowmobile
- Snowball Edge Pricing
 - o Pay for device usage and data transfer out of AWS
 - o Data transfer in S3 is free
 - o On Demand
 - One time service fee per job includes 10 days (80TB), 15 days (210 TB)
 - Shipping days not counted
 - Pay per day for additional days
 - o Committed Upfront
 - Pay in advance for monthly, 1 year, 3 years (Edge computing)
 - 62% discounted pricing

	Time to Transfer		
	100 Mbps	1Gbps	10Gbps
10 TB	12 days	30 hours	3 hours
100 TB	124 days	12 days	30 hours
1 PB	3 years	124 days	12 days



Storage Gateway

- Bridge support for on premise storage and cloud data in S3
- Hybrid storage service to allow on premises to seamlessly use the AWS cloud
- Types
 - o File Gateway
 - o Volume Gateway
 - o Tape Gateway



AWS Database

- Structure the data
- Build indexes to efficiently query / search through the data
- Define relationships between your datasets

Types of DB

- Relational DB
 - Table format
- Non-relational DB
 - JSON format
 - Flexible
 - Scalable
 - High performance
 - Highly functional

Relational Database Service (RDS)

- Managed DB service for relational database like SQL as query language
- Postgres, MySQL, MariaDB, Oracle, Microsoft SQL Server, IBM DB2, Aurora

Advantages of RDS over own DB in EC2

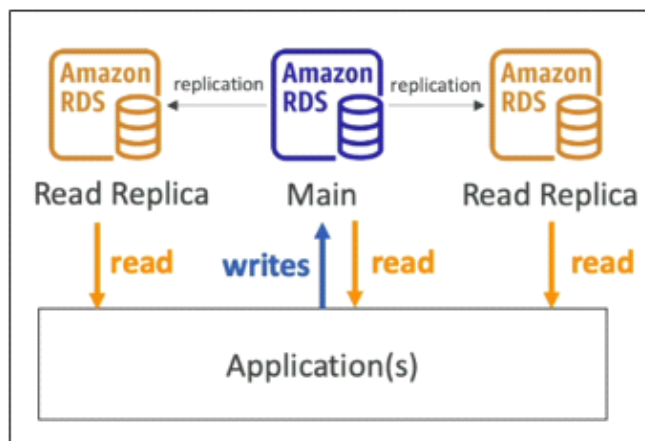
- Continuous backups and restore to specific timestamp (point in time restore)
- Monitoring dashboards
- Multi AZ setup for disaster recovery
- Horizontal and Vertical scaling
- Storage backed by EBS

Snapshots

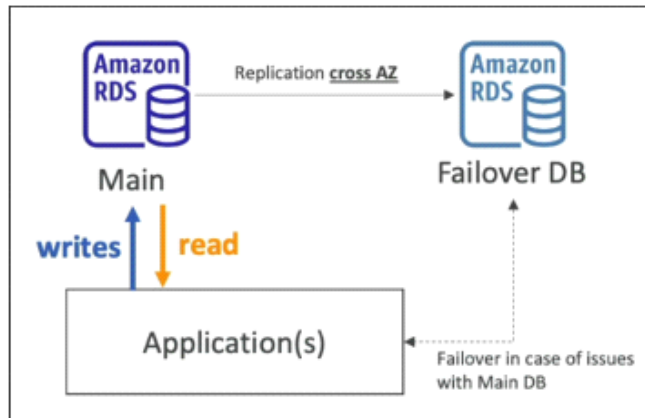
- Restore snapshot to create new DB from it
- Copy snapshot to different region
- Share snapshot with other accounts

Deployment Options

- **Read Replicas**
 - o Scale the read workload of your DB
 - o Can create up to 15 Read Replicas
 - o Data is written to main DB only

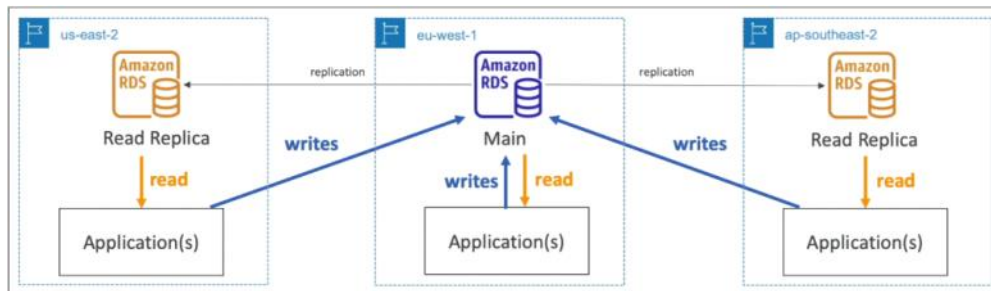


- **Multi-AZ**
 - o Failover in case of AZ outage
 - o High availability
 - o Data is only read/written to main DB until there is some issue with main DB
 - o Can only have 1 other AZ as failover



- Multi-Region (Read Replicas)

- Same as read replicas but instead of same region, different regions for read replicas
- Disaster recovery in case of region issue
- Local performance for global reads
- Replication cost

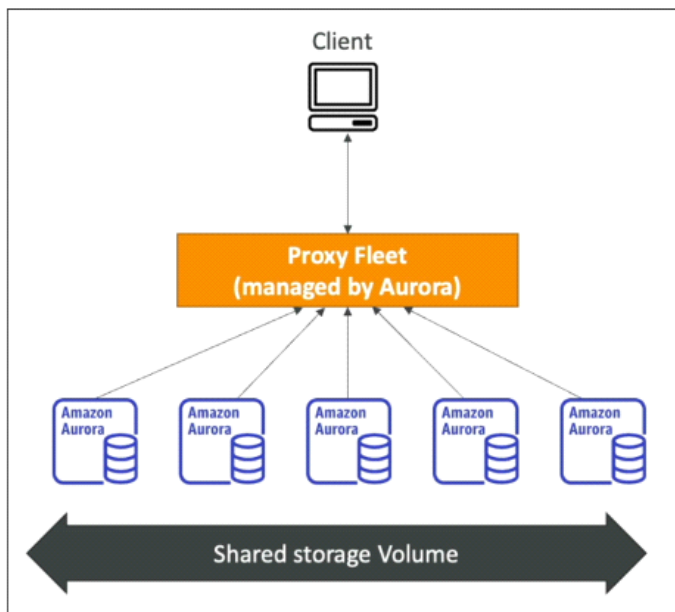


Amazon Aurora

- Proprietary technology from AWS
- Not open sourced
- Supports PostgreSQL & MySQL
- Claims 5x performance improvement over MySQL on RDS & 3x performance improvement over PostgreSQL on RDS
- Automatically grows in increments of 10GB, up to 128TB
- Cost 20% more than RDS, but more efficient

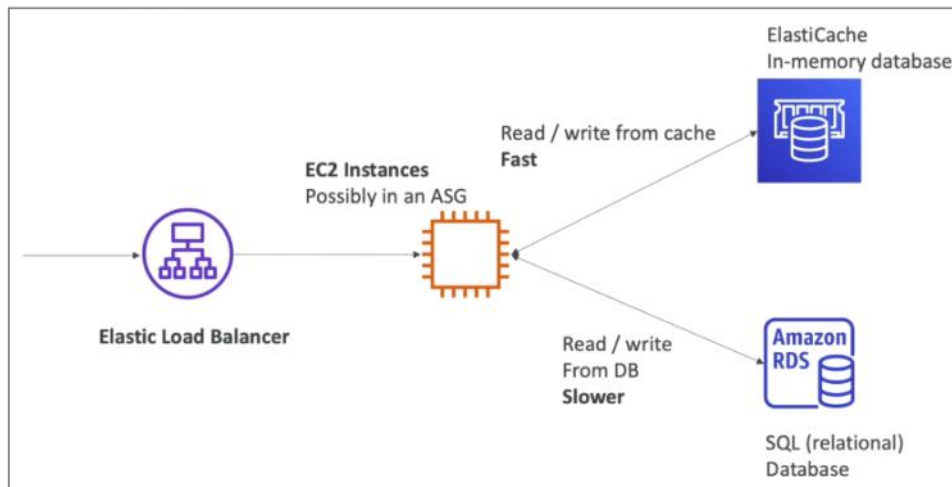
Amazon Aurora Serverless

- Automated DB instantiation
- Auto scaling
- No capacity planning needed
- Pay per second
- No management overhead
- More cost effective



Amazon ElastiCache

- AWS managed Redis or Memcached DB
- Caches are in-memory database with high performance
- Low latency
- Helps reduce load off databases for read intensive workloads

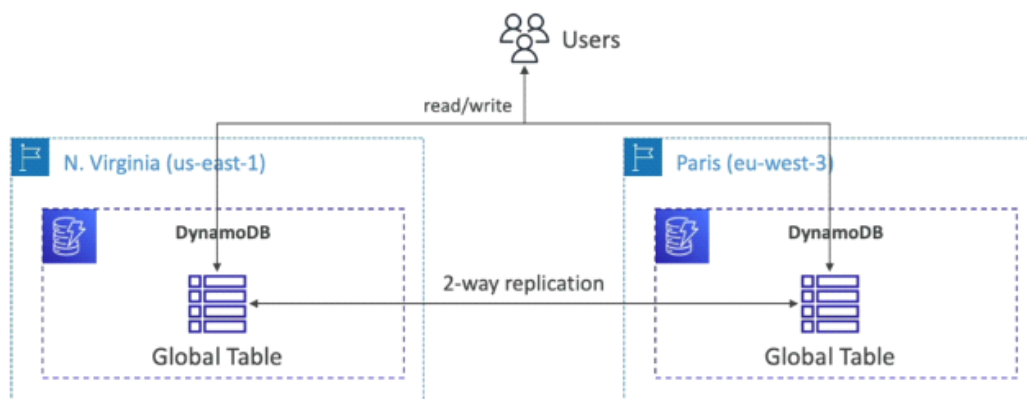


Dynamo DB

- AWS managed NoSQL (non-relational) DB
- Highly available
- Replication across 3 AZ
- Scales to massive workloads, distributed serverless DB
- No need of creating DB, directly create tables (serverless)
- Millions of request per second, trillions of row, 100s of TB
- Fast & consistent in performance
- Single digit millisecond latency
- Low latency retrieval
- Low cost & auto scaling
- Standard & Infrequent Access (IA) classes

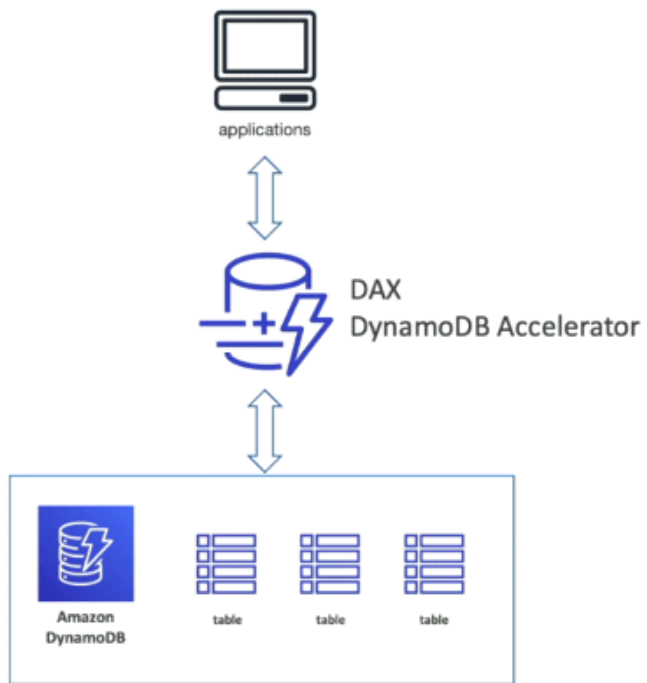
Global Tables

- Make table accessible with low latency in multiple regions
- Closest DB is accessed by user for read/write
- 2 way replication, Active-Active replication, read/write to any AWS region



DynamoDB Accelerator - DAX

- Fully managed in-memory cache for DynamoDB
- 10x performance improvement, single digit millisecond latency
- Secure
- Highly available
- Highly scalable



Database & Analytics Services

Redshift

- Based on PostgreSQL
- But it's not used for OLTP (online transactional processing)
- It is OLAP (online analytical processing - analytics and data warehousing)
- Load data once every hour, not every second
- 10x better performance than other data warehouse, scale to PBs of data
- Columnar storage of data (instead of row)
- Massively parallel query execution (MPP), highly available
- BI tools such as AWS quicksight or tableau integrate with it
- Use case: data analytics, data warehouse

Redshift Serverless

- Automatically provisions and scales data warehouse underlying capacity
- Run analytics without managing infrastructure

Elastic MapReduce (EMR)

- Helps in creating Hadoop clusters (Big data) to analyze and process vast amount of data
- Clusters can be made of hundreds of EC2 instances
- EMR takes care of all provisioning and configuration
- Auto scaling & integrated with spot instances
- Use case: data processing, machine learning, web indexing, big data

Athena

- Serverless **query service to perform analytics against S3** objects
- Uses standard SQL language to query files
- Supports csv, json, orc, avro & parquet
- Pricing \$5 per TB of data scanned
- Use compressed or columnar data for cost saving
- Use case: business intelligence, analytics, query logs, cloud trails



QuickSight

- Serverless machine learning powered business intelligence service to create **interactive dashboard**
- Fast, automatically scalable, embeddable, with per-session pricing
- Integrated with RDS, Aurora, Athena, Redshift, S3
- Use case: visualizations, perform ad-hoc analysis, insights

DocumentDB

- AWS implementation of **MongoDB**
- Non-relational database
- Fully managed, highly available with replication across 3 AZ
- Storage automatically grows in increments of 10GB
- Automatically scales to workloads with millions of requests per seconds

Neptune

- Fully managed **graph database**
- Highly available across 3 AZ, with up to 15 read replicas
- Can store up to billions of relations and query the graph with milliseconds latency
- Highly available with replication across multiple Azs

Timestream

- Fully managed, fast, scalable, serverless **time series database**
- Time series data is a data which evolves with the time
- Automatically scales up/down to adjust capacity
- Store & analyze trillions of events per day
- 1000s times faster & 1/10th the cost of relational DB
- Built in time series analytics function helps you identify patterns in your data in near real time

Quantum Ledger Database (QLDB)

- Ledger is book recording **financial transactions**

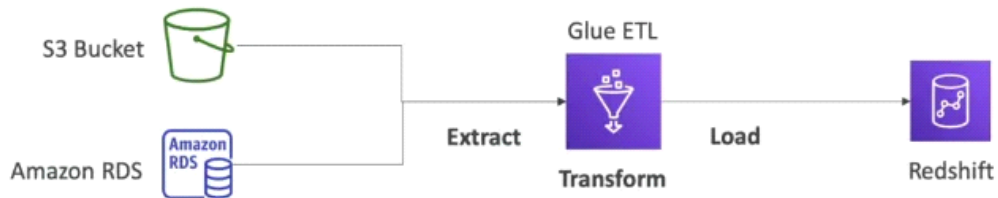
- Fully managed, serverless, high available, replication across 3 AZ
- Used to review history of all the changes made to your application data over time
- Immutable system, cryptographically verifiable
- 2-3x better performance than common ledger blockchain frameworks, manipulate using SQL

Amazon Managed Blockchain

- **Blockchain** makes it possible to build applications where multiple parties can execute transactions without the need for a trusted, central authority
- Amazon managed blockchain is a managed service to
 - o Join public blockchain networks
 - o Create your own scalable private network
- Compatible with frameworks Hyperledger Fabric & Ethereum

Glue

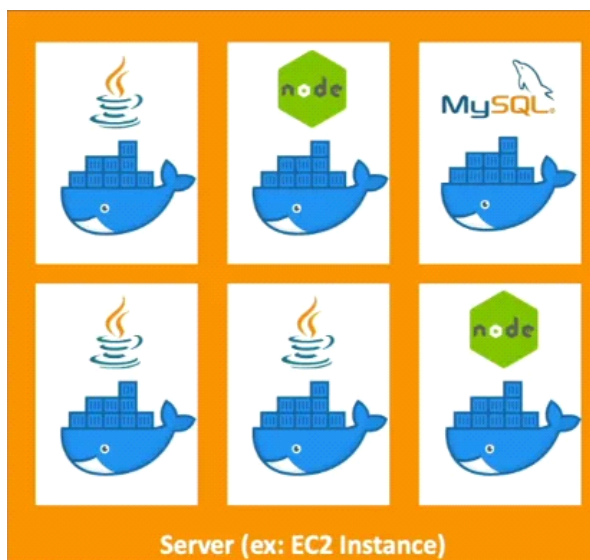
- Managed extract, transform & load (**ETL**) service
- Useful to prepare & transform data for analytics
- Fully serverless service



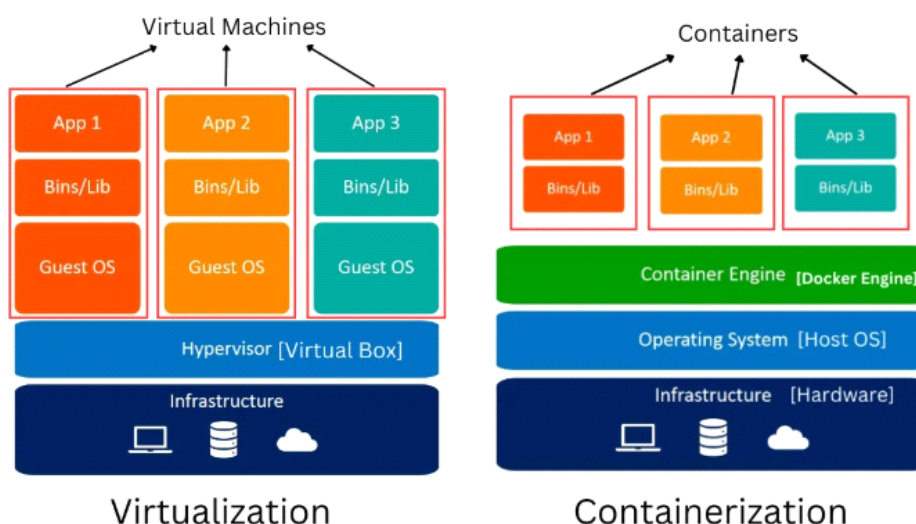
- Glue Data Catalog: catalog of datasets, can be used by Athena, Redshift, EMR

Docker

- Software development platform to deploy apps
- Apps are packaged in container that can be run on any OS
- Apps run the same, regardless of where they are run
 - o Any machine
 - o No compatibility issues
 - o Predictable behavior
 - o Less work
 - o Easier to maintain and deploy
 - o Works with any language, any OS, any technology
- Scale containers up and down, very quickly (seconds)



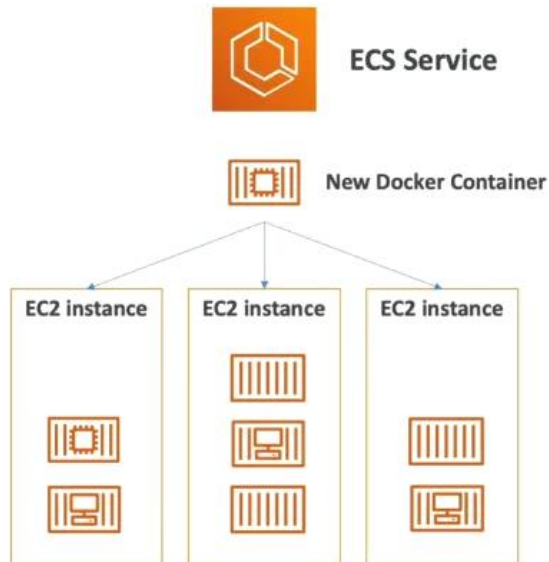
- Docker images are stored in Docker repositories
 - o Public: Docker Hub
 - o Private: Amazon ECR (Elastic Container Registry)
- Docker is like virtualization but not exactly



Container Services

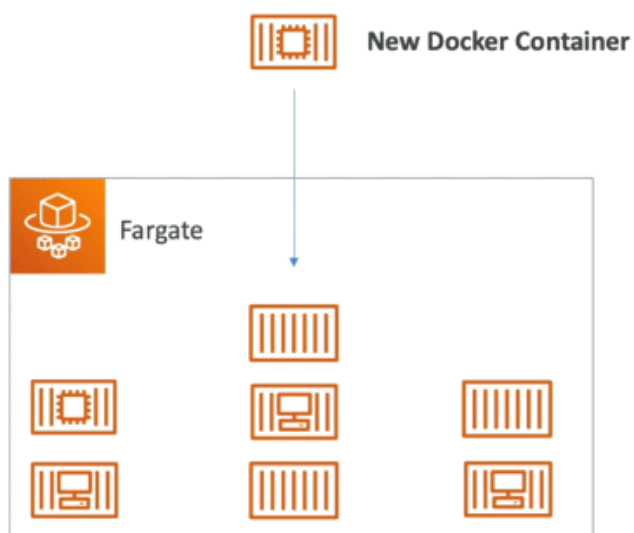
Elastic Container Service (ECS)

- Launch docker containers on AWS
- You must provision & maintain the infrastructure (EC2 instances)
- AWS takes care of starting / stopping containers
- Has integration with Application Load Balancer



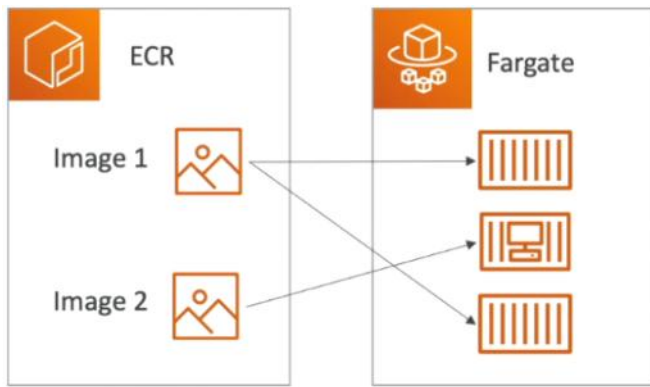
Fargate

- Launch docker containers on AWS
- You do not provision the infrastructure (no EC2 instance to manage)
- Serverless offering
- AWS just runs containers for you, based on CPU/RAM you need



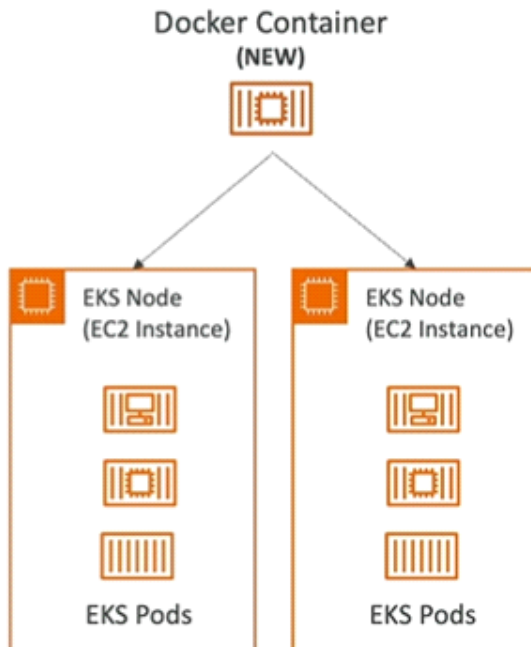
Elastic Container Registry (ECR)

- Private docker registry on AWS
- This is where you store your docker images so they can be run by ECS or Fargate



Elastic Kubernetes Service (EKS)

- Launch managed Kubernetes cluster on AWS
- Kubernetes is open source system for management, deployment and scaling of containerized apps

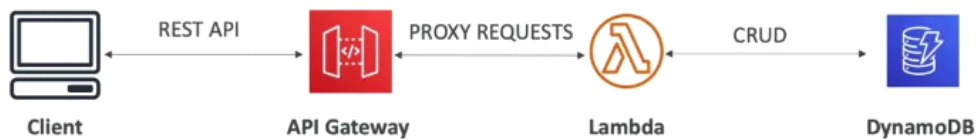


Lambda

- Virtual functions, no servers to manage
- Run on demand
- Scaling is automated
- Serverless

Benefits

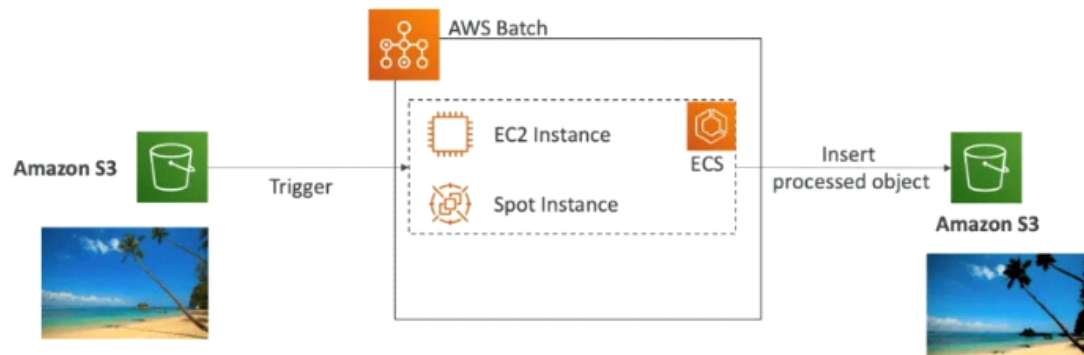
- Pay per request & compute time
- Integrated with many AWS services
- Event driven
- Supports many programming languages
- Supports Lambda Container Image, container image must implement Lambda Runtime API



Batch

- Fully managed batch processing at any scale
- Efficiently run 100000s of computing batch jobs
- Batch job is a job with start and end, it will dynamically launch EC2 instance or spot instance
- AWS batch provisions right amount of compute/memory
- You submit or schedule batch job
- Batch jobs are defined as Docker images & run on ECS

Example



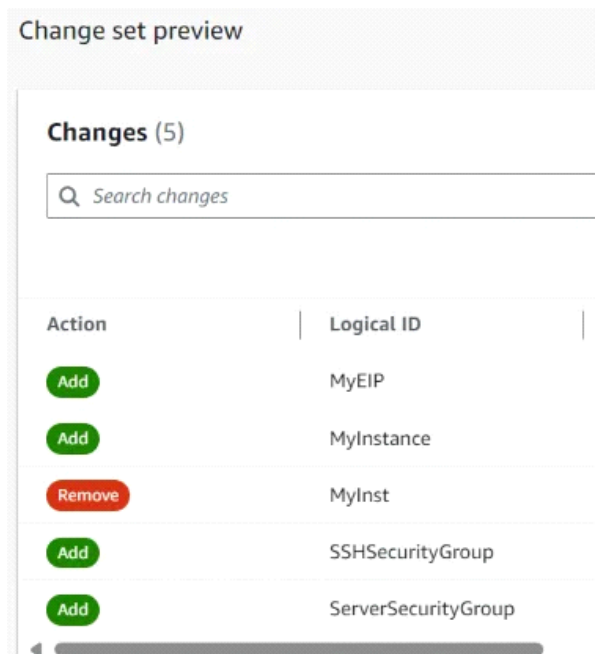
Lightsail

- Easy to use for beginners
- Includes everything you need to build websites or web applications
- Provides virtual servers, databases, storage, and more
- Low and predictable monthly cost
- Great for people with little cloud experience
- Has high availability, but no auto scaling
- Can setup notifications and monitoring
- Use cases
 - o Simple web applications
 - o Websites (templates for WordPress)
 - o Dev / Test environment
- Good for hosting WordPress instantly

Infrastructure Management

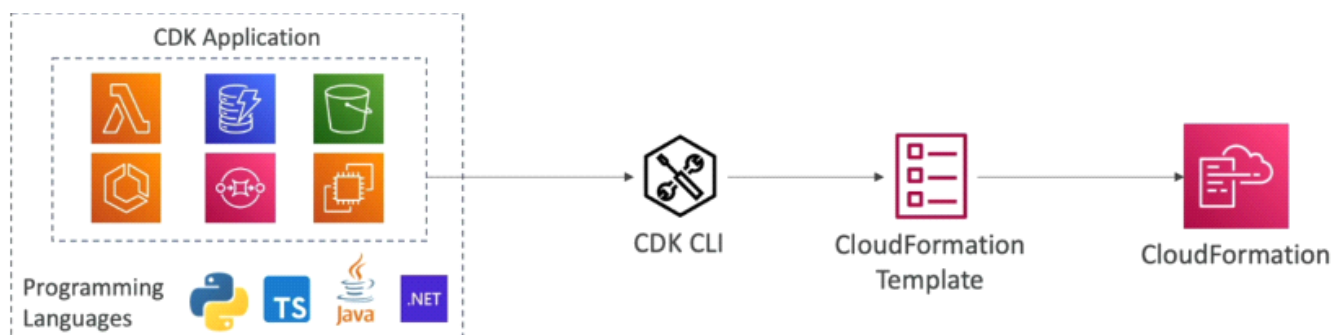
CloudFormation

- Declarative way of outlining your AWS infrastructure for any resources in a template
- IaC (YAML based)
- Each resource is tagged with an identifier for cost management
- Saving strategy : custom automation for deletion and creation
- It will generate diagrams automatically for your template
- Application Composer helps in giving visualization of template
- When you create CF stack, it will show you change set preview



Cloud Development Kit (CDK)

- Define cloud infrastructure using your familiar language
- Code is compiled into CloudFormation template (YAML/JSON)
- Can be used in lambda, or docker container in ECS/EKS




```

export class MyEcsConstructStack extends core.Stack {
  constructor(scope: core.App, id: string, props?: core.StackProps) {
    super(scope, id, props);

    const vpc = new ec2.Vpc(this, "MyVpc", {
      maxAzs: 3 // Default is all AZs in region
    });

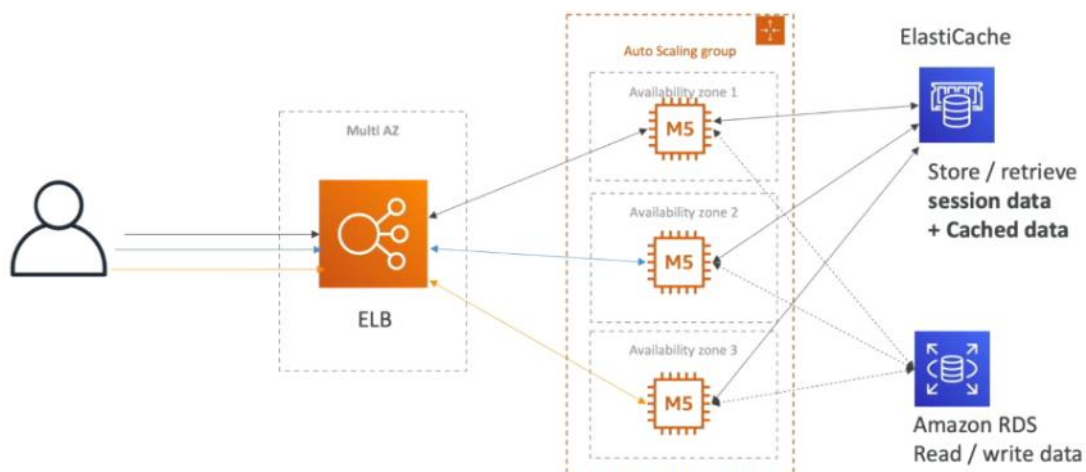
    const cluster = new ecs.Cluster(this, "MyCluster", {
      vpc: vpc
    });

    // Create a load-balanced Fargate service and make it public
    new ecs_patterns.ApplicationLoadBalancedFargateService(this, "My
      cluster: cluster, // Required
      cpu: 512, // Default is 256
      desiredCount: 6, // Default is 1
      taskImageOptions: { image: ecs.ContainerImage.fromRegistry("an
      memoryLimitMiB: 2048, // Default is 512
      publicLoadBalancer: true // Default is false
    });
  }
}

```

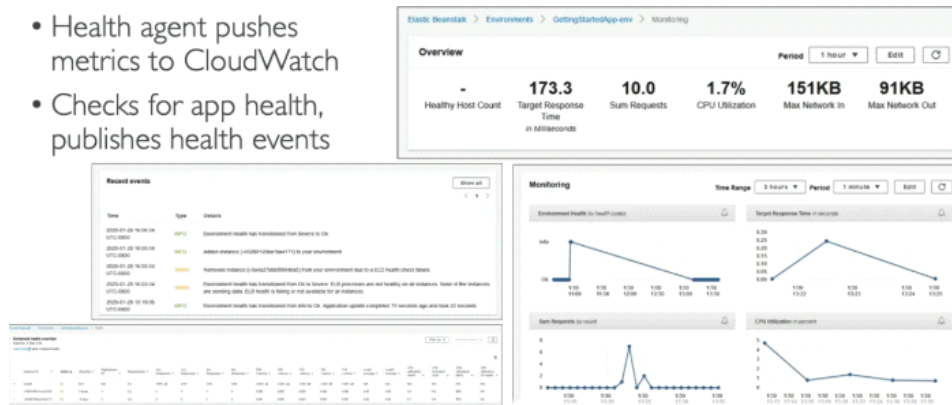
Elastic Beanstalk

- Webapp 3 tier



- EB is developer centric view of deploying an application on AWS
- It uses all the components like EC2, ASG, ELB, RDS, etc.
- We still have full control over the configuration
- PaaS
- Managed Service
 - o Deployment strategy
 - o Capacity provisioning
 - o Load balancing & auto scaling
 - o Instance configuration / OS
 - o Health monitoring

- Health agent pushes metrics to CloudWatch
- Checks for app health, publishes health events



- Three architecture models
 - Single Instance deployment : good for dev
 - LB + ASG : great for prod & pre-prod
 - ASG only : great for non-web apps in prod
- EB environment creates a CloudFormation stack

DevOps

CodeDeploy

- We want to deploy our application automatically
- Works with EC2 instances
- Works with on premise servers
- Hybrid service
- Instances must be provisioned and configured AOT with CodeDeploy Agent

CodeCommit

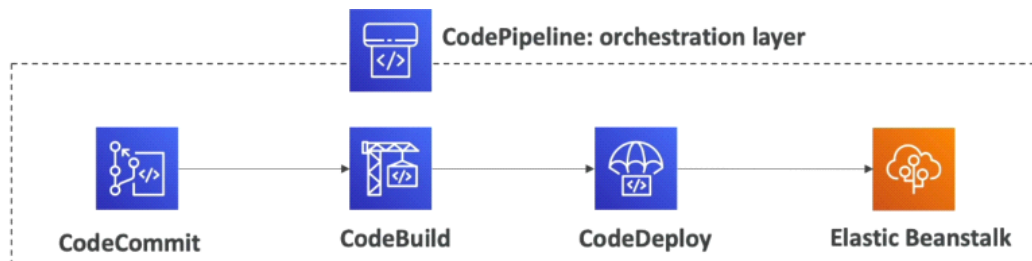
- It is discontinued
- Similar to GitHub, GitLab
- AWS managed VCS

CodeBuild

- Compiles source code, run tests and produces packages that are ready to be deployed (by CodeDeploy for example)
- AWS managed

CodePipeline

- Orchestrate the different steps to have the code automatically pushed to production
- Code -> Build -> Test -> Provision -> Deploy
- CI/CD



CodeArtifact

- From code a package (artifact) is created, that has to be stored somewhere i.e. artifact management system
- CodeArtifact is AWS managed artifact management system
- Works with all the dependency management tools

CloudStar

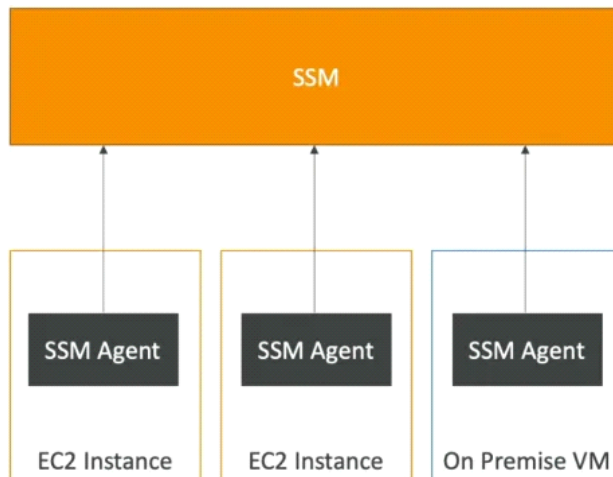
- Unified view for allowing developers to do CICD and code

Cloud9

- Cloud IDE with collab

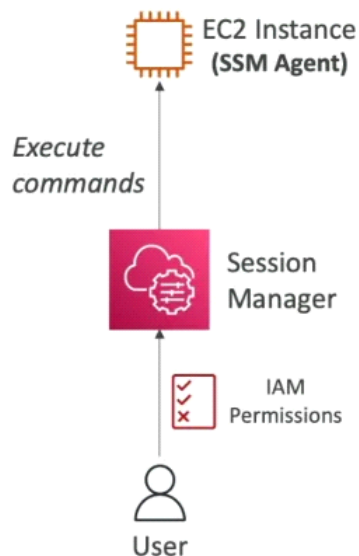
Systems Manager (SSM)

- Helps you to manage EC2 & on premises systems at scale
- Hybrid AWS service
- Operational insights
- Suite of 10+ products
- Features
 - o Patching automation for enhanced compliance
 - o Run commands across entire fleet of servers
 - o Store parameter configuration with SSM Parameter Store



SSM Session Manager

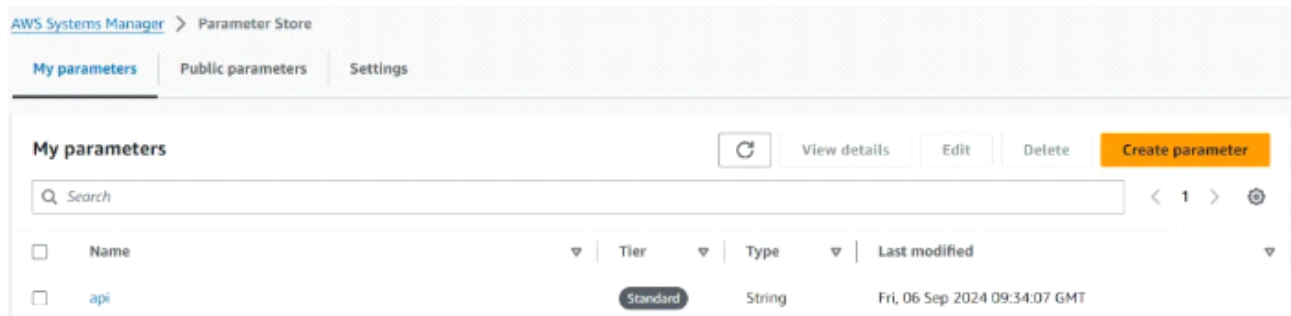
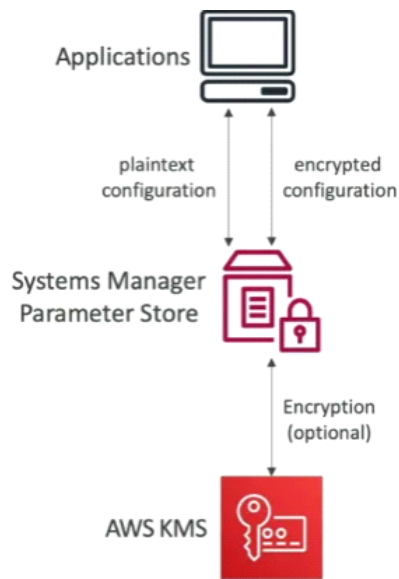
- Allows you to start a secure shell on your EC2 and on premise servers
- No SSH access, or SSH keys needed
- No port 22 needed
- But it needs IAM role



Systems Manager Parameter Store

- Secure storage for configuration & secrets

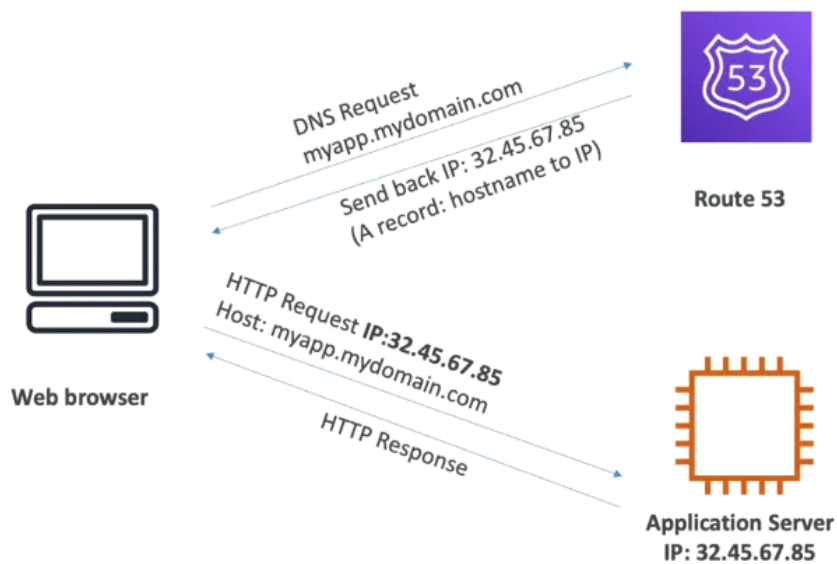
- API keys, passwords, configurations, etc.
- Control access permissions using IAM
- Version tracking & encryption for each parameter



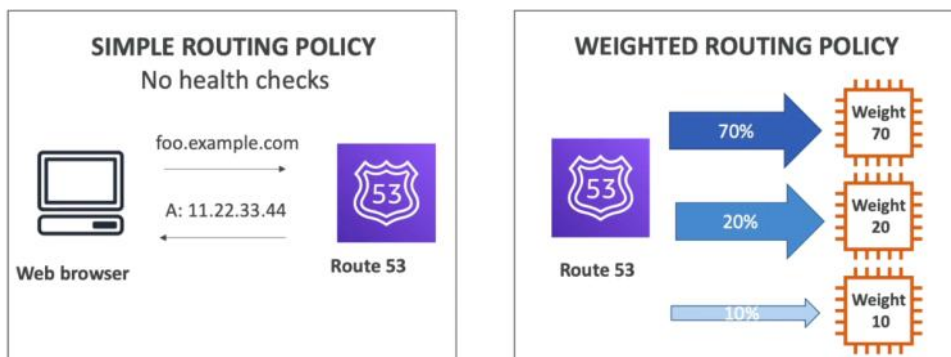
Route 53

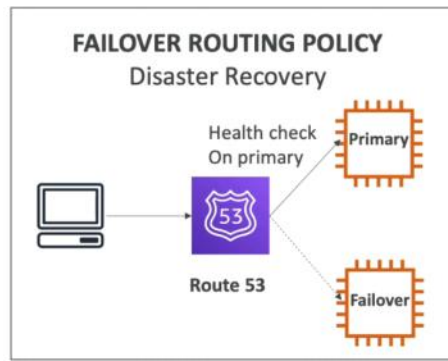
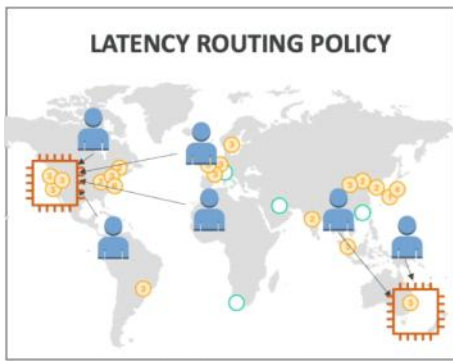
- Managed DNS (Domain Name System)
- DNS is collection of rules & records which helps client understand how to reach the server through URLs
- Most common records
 - o Domain to IPV4 => A
 - o Domain to IPV6 => AAAA
 - o Domain to Domain => CNAME
 - o Domain to AWS resource => Alias

Working of A record



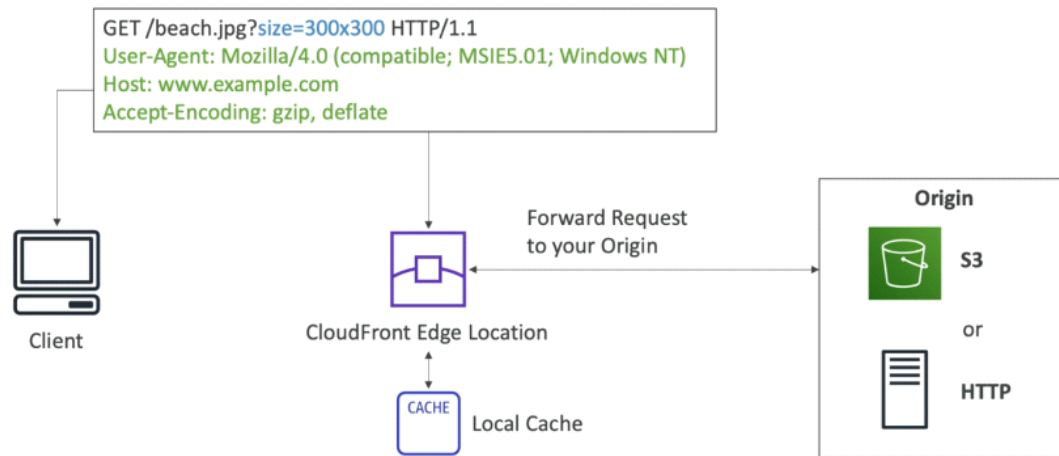
Routing Policies





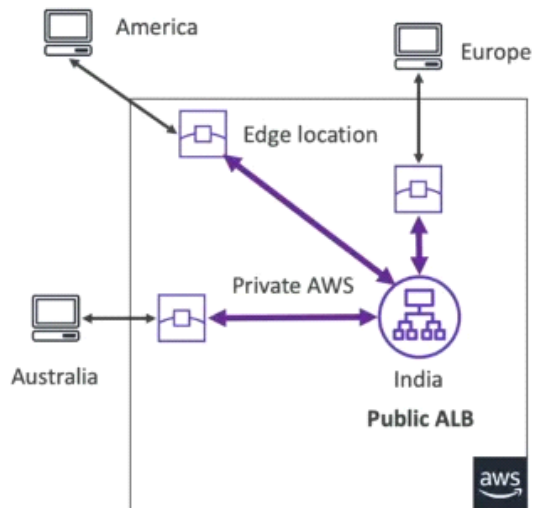
CloudFront

- CDN (Content Delivery Network)
- **Improves read performance**, by **caching content at different edge locations** which effects in **low latency**
- 216 edge locations
- DDoS protection, Web application firewall, integration with shield
- CloudFront can be used with S3, EC2, ALB

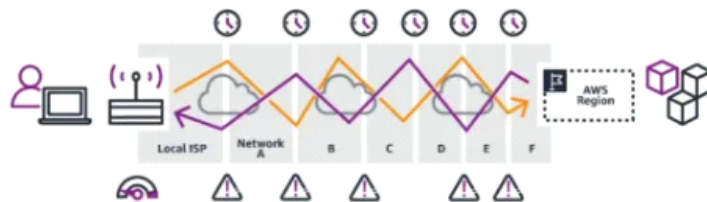


Global Accelerator

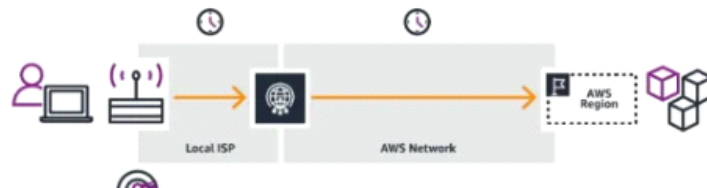
- Improve global application availability & performance using the AWS global network
- Leverage the AWS internal network to optimize the route to your application



Without Global Accelerator

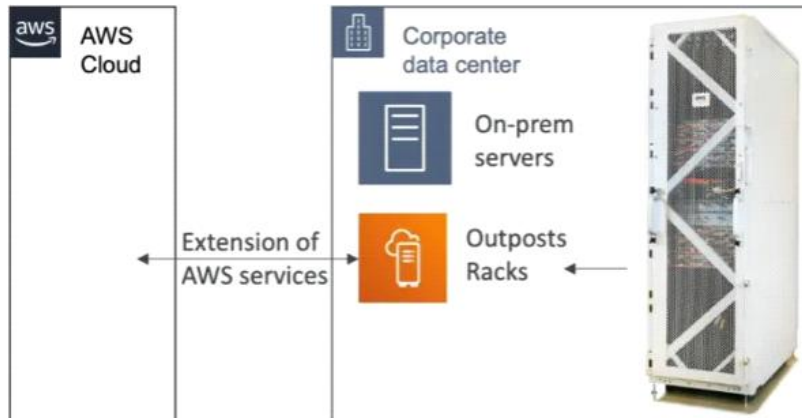


With Global Accelerator



Outposts

- AWS Outposts are server racks that offers the same AWS infrastructure, services, APIs & tools to build your own application on-premises just as in cloud
- AWS will setup and manage Outposts Racks within your on premise infrastructure & you can start leveraging AWS services on premises



- Easier migration from on premise to cloud
- Services that works on outposts are EC2, EBS, S3, EKS, ECS, RDS, EMR

Global Application Architecture

- Decreased latency
- Attack protection

Single Region, Single AZ

- ✗ High Availability
- ✗ Global Latency
- 🟢 Difficulty



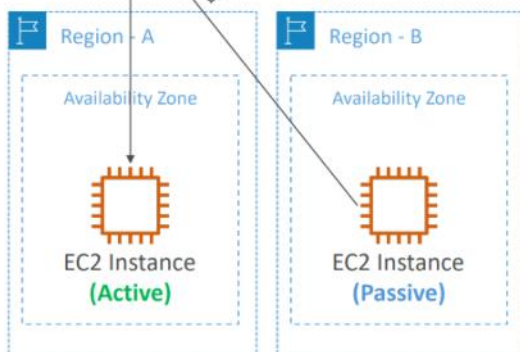
Single Region, Multi AZ

- ✓ High Availability
- ✗ Global Latency
- 🟡 Difficulty



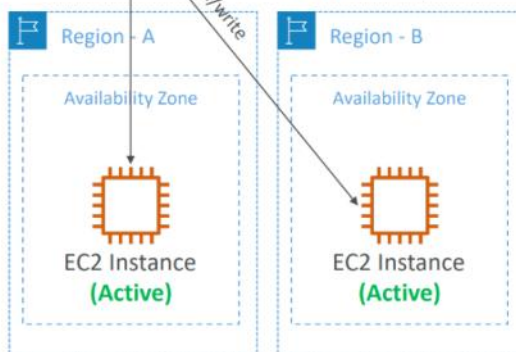
Multi Region, Active-Passive

- ✓ Global Reads' Latency
- ✗ Global Writes' Latency
- 🟡 Difficulty



Multi Region, Active-Active

- ✓ Reads' Latency
- ✓ Writes' Latency
- 🟡 Difficulty



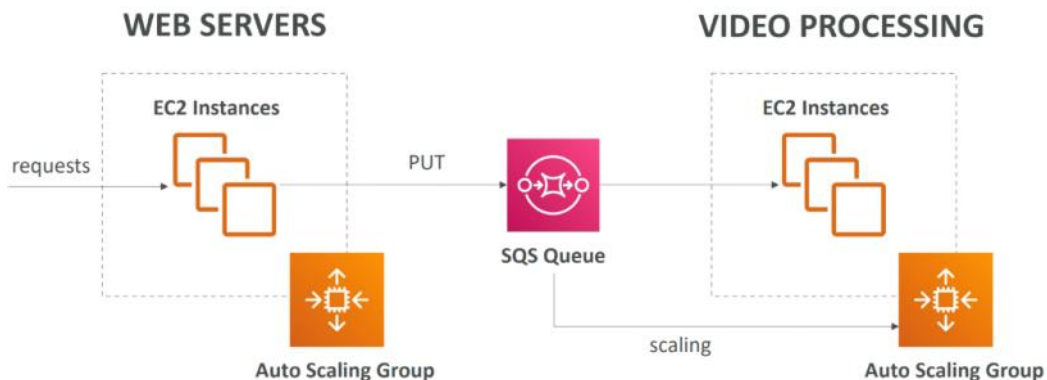
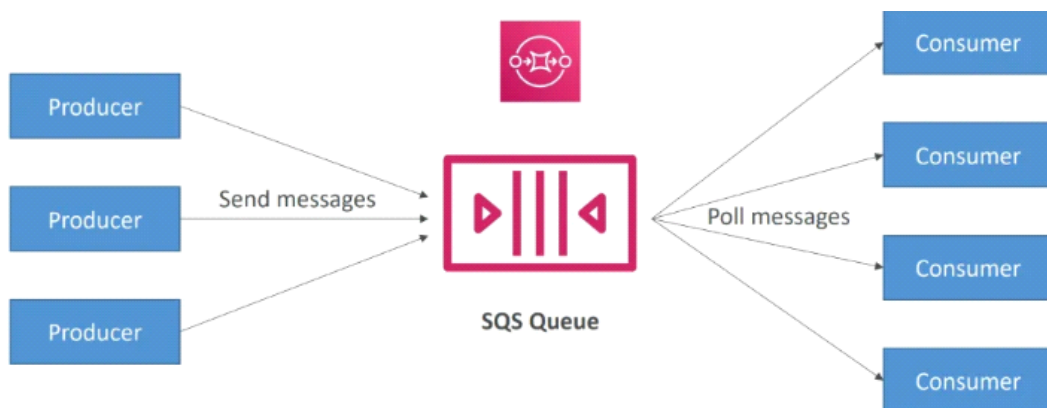
Inter Service Communication

- When we start deploying multiple applications, they will inevitably need to communicate with one another
- There are two patterns



Simple Queue Service (SQS)

- Queue model
- Standard & FIFO Queue
- Serverless, used to decouple applications
- Scales from 1 msg per second to 10,000 msg per second
- Low latency



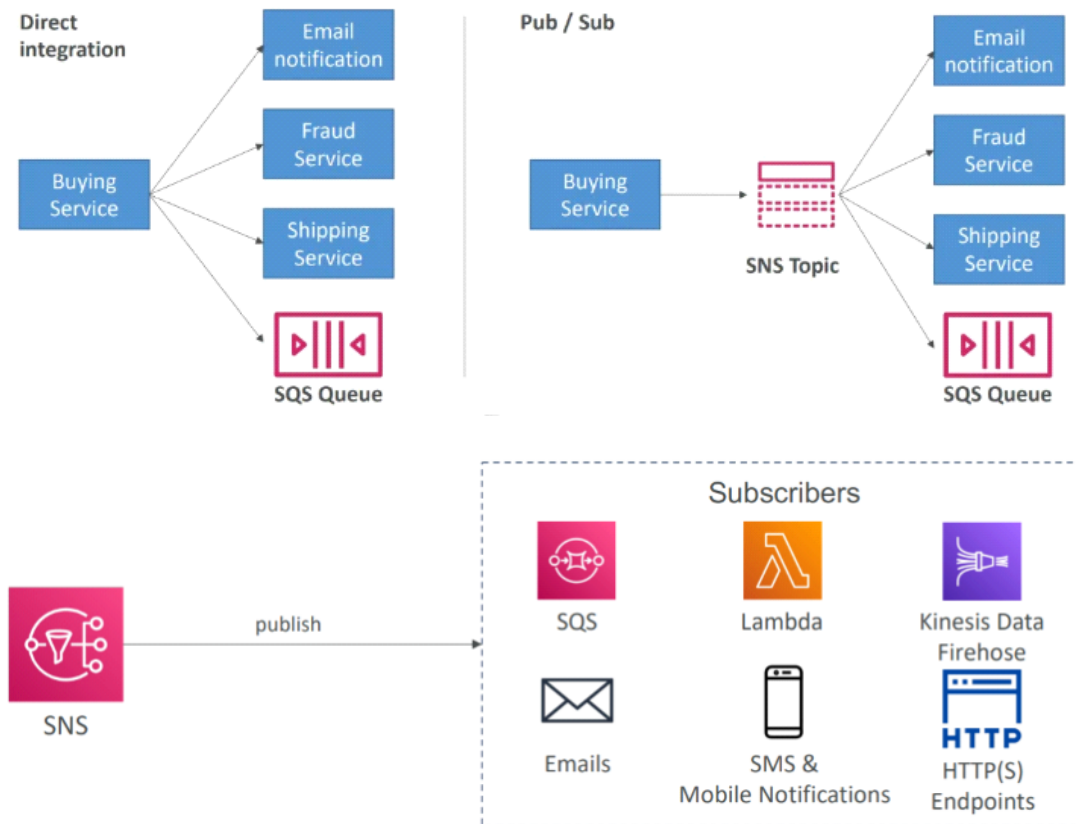
Kinesis

- Real time big data streaming
- Managed service to collect, process & analyze real time streaming data at any scale

Simple Notification Service (SNS)

- Pub/Sub model

- Event publishers only sends message to one SNS topic
- Many event subscribers can listen to SNS topic notifications
- Each subscriber to the topic will get all the messages
- Up to 12,500,000 subscriptions & 100,000 topics limit



MQ

- Traditional apps running from on premises may use open protocols like MQTT, AMQP, STOMP, Openwire, WSS
- When migrating to the cloud, instead of re-engineering application to use SQS & SNS, we can use Amazon MQ
- Message broker service for RabbitMQ, ActiveMQ
- Multi AZ support
- Queue (SQS) and Topic (SNS) feature

CloudWatch

Metrics

- CloudWatch provides metrics for every service in AWS like CPU utilization, Network In with timestamps
- Can create CloudWatch dashboards for metrics

Alarm

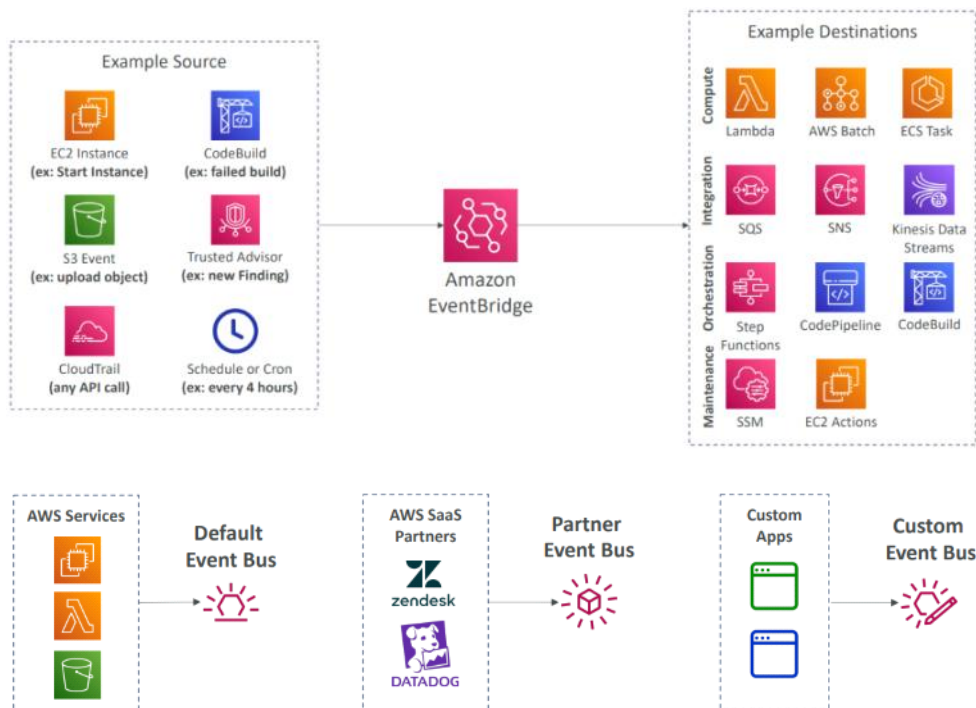
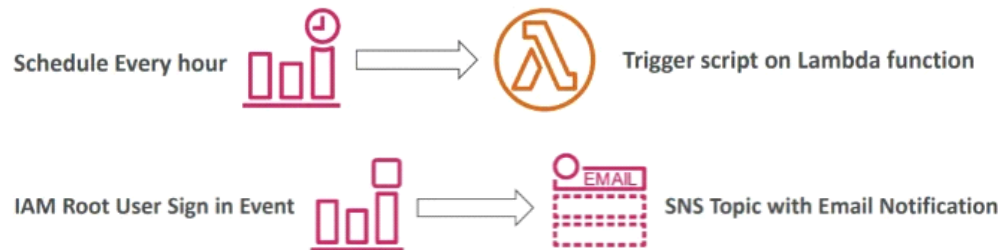
- CloudWatch alarms are used to trigger an action like Auto Scaling, EC2 actions, SNS notification, etc. based on the metrics
- Alarm states are OK, INSUFFICIENT_DATA, ALARM
- Billing alarm is available only in us-east-1

Logs

- Collects log from almost all the services, real time monitoring logs
- For EC2, it won't happen by default, you have to setup CloudWatch Logs Agent on EC2 or on premise server

EventBridge

- Formerly known as CloudWatch Events
- Run some script or action on some event
- EventBridge Scheduler helps in defining the schedules
- Examples like



CloudTrail

- Provides governance, compliance & audit for your AWS account
- Enabled by default
- Get history of all the events / API class in your account by
 - o Console
 - o SDK
 - o CLI
 - o AWS services
- Can be applied to all regions (default) or single region
- You can store all the data in S3



X-Ray

- Visual analysis of your application
- Helps in debugging the application better in production
- Find errors and exceptions
- Pinpoint service issues

CodeGuru

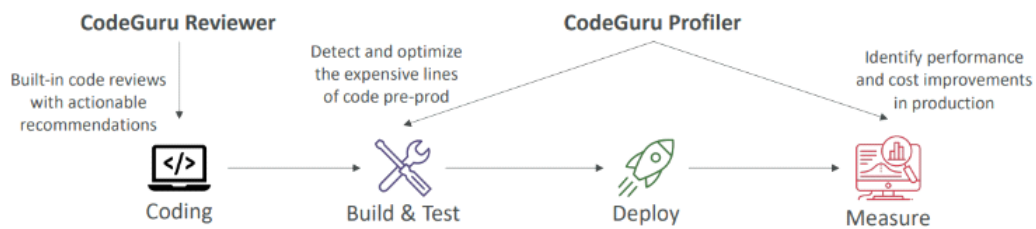
- ML powered service for automated code review and application performance recommendations

CodeGuru Reviewer

- Automated code review for static code analysis (development)
- It looks at commits and do review
- Integrates with GitHub, Bitbucket & CodeCommit

CodeGuru Profiler

- Visibility / recommendations about application performance during runtime (production)
- Identify and remove code inefficiencies
- Improve application performance
- Decrease compute costs
- Provides heap summary
- Anomaly detection
- Supports application running on AWS or on premise
- Minimal overhead on application



Health Dashboard

Service Health

- Previously called AWS Service Health Dashboard
- Shows all regions, all services health
- Shows historical information for each day
- RSS feed you can subscribe to

Your Account Health

- Previously called AWS Personal Health Dashboard
- Provides alerts and remediation guidance when AWS is experiencing events that may impact you directly
- Global service

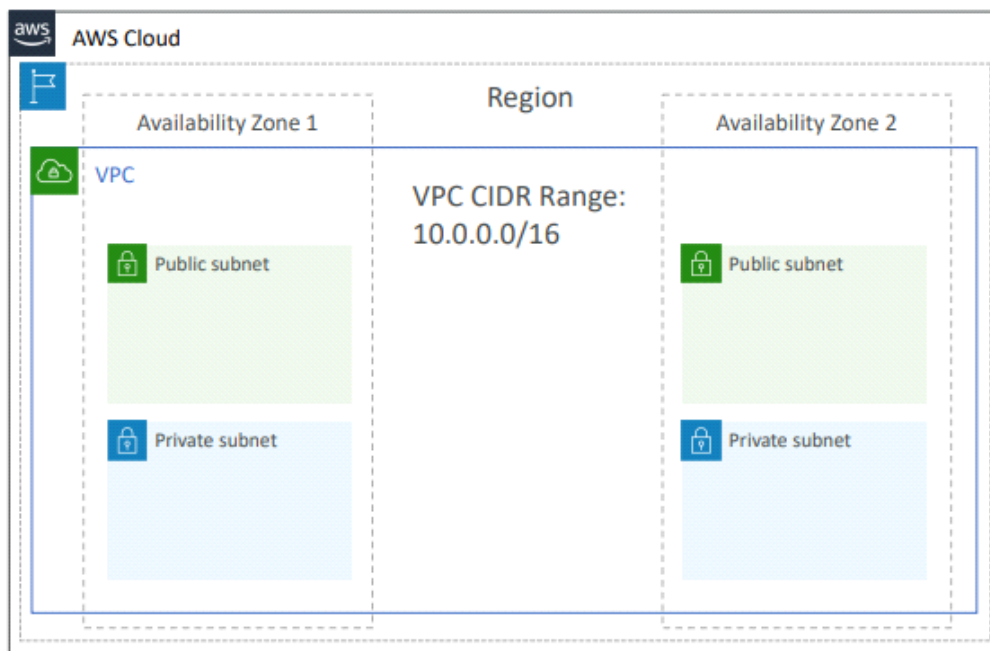
Virtual Private Cloud (VPC)

IP Address

- IPv4 (4.3 Billion)
- IPv6 ($3.4 * 10^{38}$)
- Public and Private IP
- EC2 instance gets new public IP address every time you stop then start
- Private IP is fixed for EC2 instances
- Elastic IP allows you to attach a fixed public IPv4 to EC2
- All Public IPv4 on AWS are chargeable
- IPv6 are free

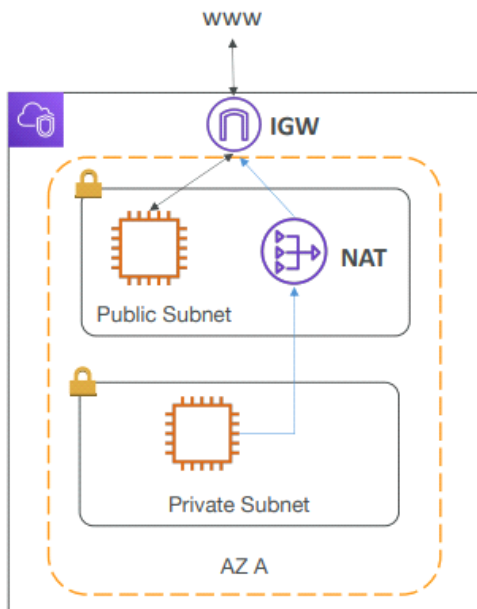
VPC & Subnets

- Private network to deploy your resources
- Subnets allow you to partition your network inside your VPC
- Subnets are tied to AZ
- Public subnet is accessible from internet whereas Private is not
- To define access to internet and between subnets, we use Route tables



Internet Gateway & NAT Gateway

- IG helps VPC to connect with internet
- Public subnet will have route to IG
- NAT helps private subnets to access internet while remaining private

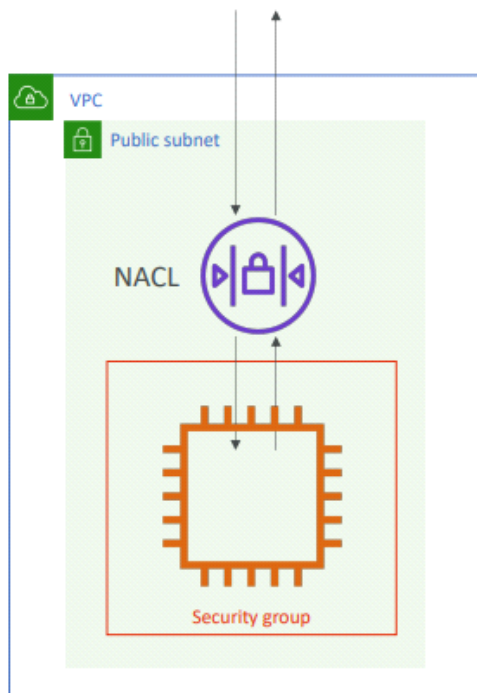


Network ACL

- Firewall that controls traffic from and to subnet
- Attached at subnet level
- Can have ALLOW and DENY rules
- Rules only include IP address

Security Groups

- Firewall that controls traffic from and to EC2
- Can have ALLOW rules
- Rules include IP address and other security groups



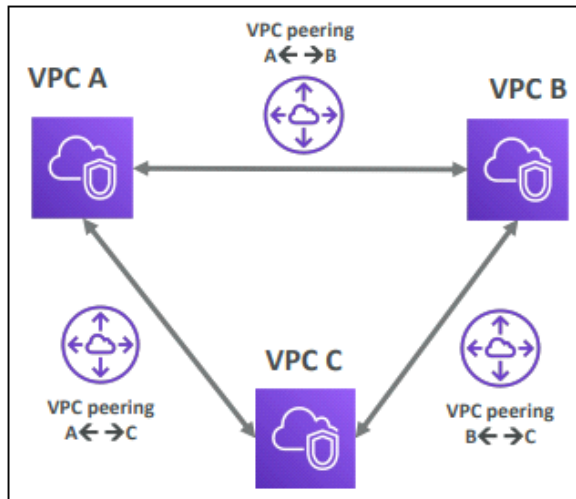
VPC Flow Logs

- Log of all IP traffic : VPC, Subnet, Elastic Network Interface

- Useful in monitoring : subnet to internet, subnet to subnet, internet to subnet

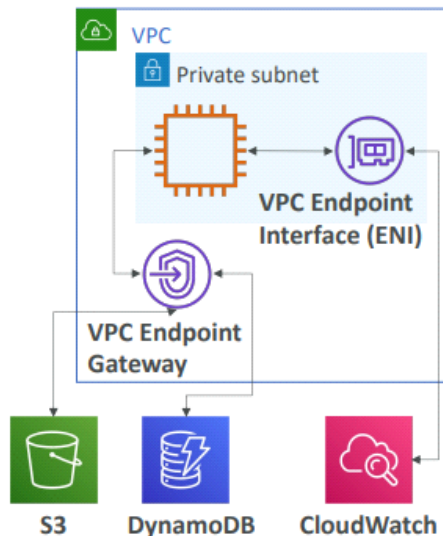
VPC Peering

- Connect two VPC, privately using AWS network
- Make them behave as if they were in same network
- Not transitive



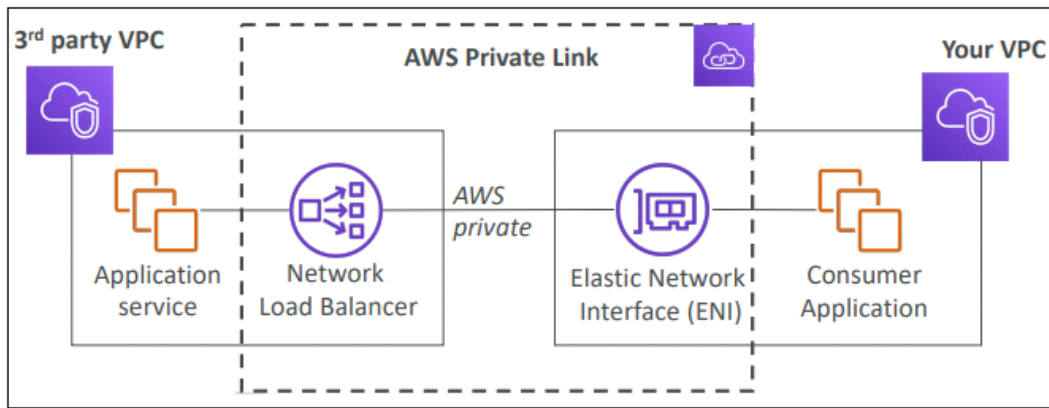
VPC Endpoints

- This allows you to connect to AWS services using a private network instead of public www network
- Enhanced security and lower latency
- VPC Endpoint Gateway : S3 and DynamoDB
- VPC Endpoint Interface : rest services



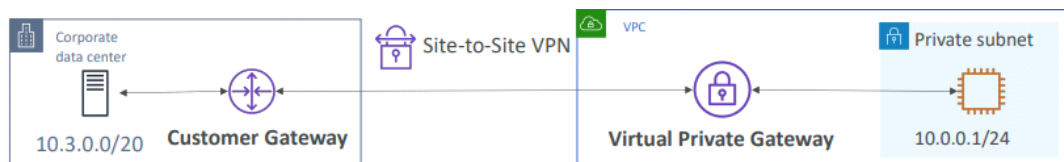
AWS PrivateLink (VPC Endpoint Services)

- Allows you to connect services running in your VPC to some other 3rd party VPC directly and privately
- Requires network load balancer and ENI

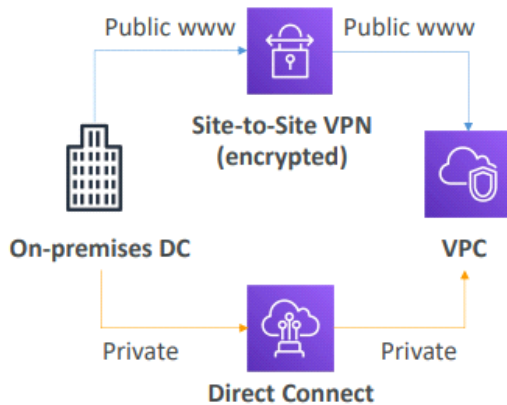


Site to Site VPN & Direct Connect

- Connecting on premise DC to cloud VPC
- Site to Site VPN
 - o Automatically encrypted
 - o Goes over public internet



- Direct Connect (DX)
 - o Physical connection
 - o Private, secure, fast
 - o Goes over private network



Client VPN

- Connect your computer using OpenVPN to your private network in AWS and on premises
- Allow you to connect your EC2 over private IP (as if you were in private VPC network)
- Goes over public network

Computer with
AWS Client VPN (OpenVPN)



Internet WWW



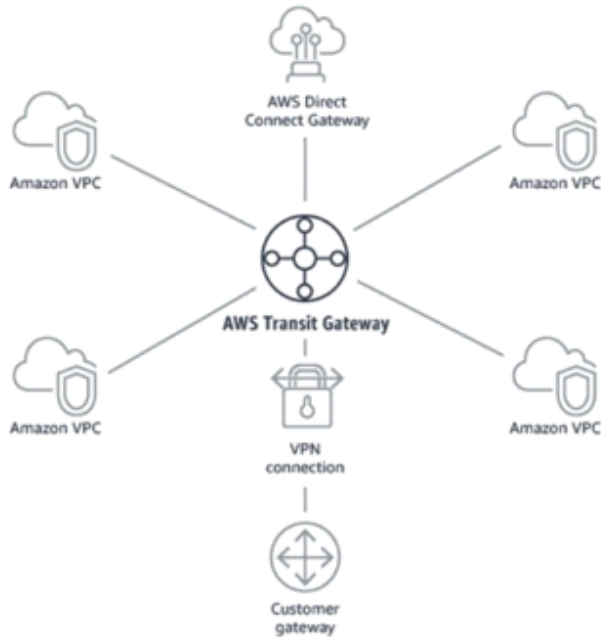
Site-to-Site VPN



On-Premises
Data Center

Transit Gateway

- For having transitive peering between thousands of VPC and on premises, star connection
- One single gateway



Encryption

Data at rest vs Data in transit

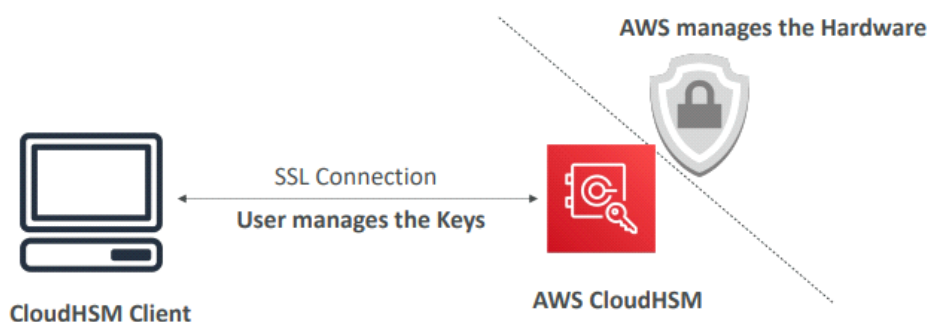
- **At rest**
 - o Data stored or archived on device
- **In transit**
 - o Data being moved from one location to another
 - o Transferred on network
- To encrypt data in both states we use encryption keys

AWS Key Management Service (KMS)

- AWS manages encryption keys for us
- Encryption opt in
 - o EBS volumes
 - o S3 buckets : SSE-S3 enabled by default, SSE-KMS opt in
 - o Redshift DB
 - o RDS DB
 - o EFS drives
- Encryption automatically enabled
 - o CloudTrail Logs
 - o S3 Glacier
 - o Storage Gateway

CloudHSM

- AWS provisions encryption hardware
- Dedicated hardware (Hardware Security Module)
- You manage your own encryption keys entirely



Types of KMS keys

- **Customer Managed Key**
 - o Create, manage, used by customer, can enable or disable
 - o Possibility of rotation policy
 - o Possibility of bring your own key
- **AWS Managed Key**
 - o Created, managed, used on customer behalf by AWS
 - o Used by AWS services
- **AWS Owned Key**

- Collection of CMKs that service owns and manages to use in multiple accounts
 - You can't view the keys
- CloudHSM Keys (custom keystore)
 - Keys generated from your own CloudHSM hardware device

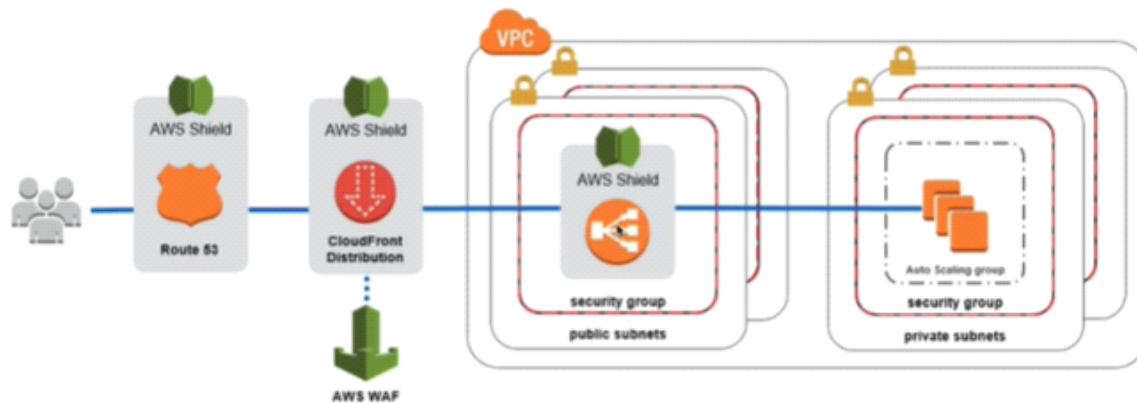
Secrets Manager

- Store secrets
- Force rotation of secrets every x days
- Automate generation of secrets on rotation using lambda
- Integration with RDS
- Encrypted using KMS

Firewall Services

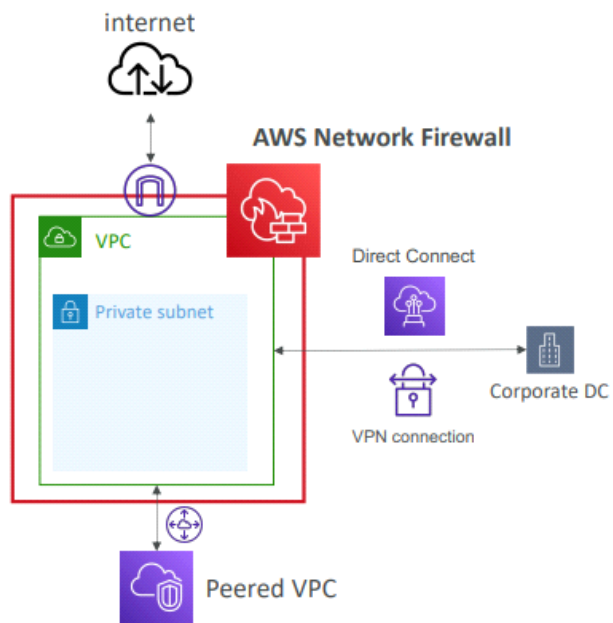
Distributed Denial of Service (DDoS)

- **AWS Shield Standard**
 - o Protects against DDoS attack for your website and app, no additional costs
- **AWS Shield Advanced**
 - o 24/7 premium DDoS protection
- **AWS WAF**
 - o Filter specific requests based on rules
 - o Protects your web app from web exploits (http layer 7)
 - o Deploy on ALB, API Gateway, CloudFront
 - o Define Web ACL
- **CloudFront and Route 53**
 - o Protection using global edge network
 - o Combined with AWS Shield, provides attack mitigation at the edge



AWS Network Firewall

- Protect your entire VPC from layer 3 to 7
- All directions can be inspected

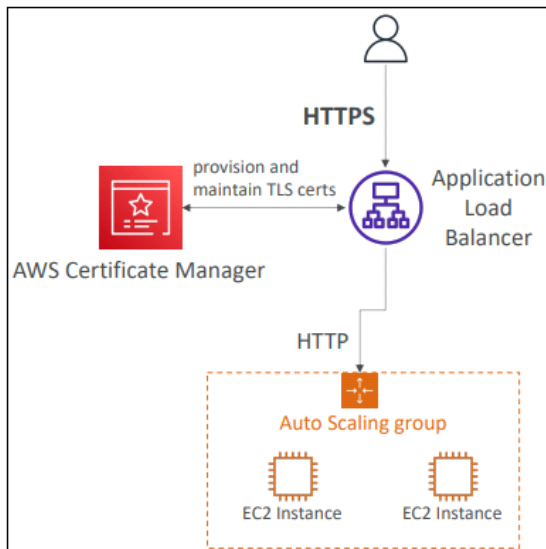


AWS Firewall Manager

- Manage security rules in all accounts of an AWS organization
- Rules are applied to new resources as they are created across all and future accounts

AWS Certificate Manager (ACM)

- Provision, manage, deploy SSL/TLS certificates
- Provide in flight encryption for websites (HTTPS)
- Supports both public and private TLS certificates
- Free of charge for public TLS
- Automatic TLS certificate renewal
- Integrates with ELB, CloudFront, API Gateway



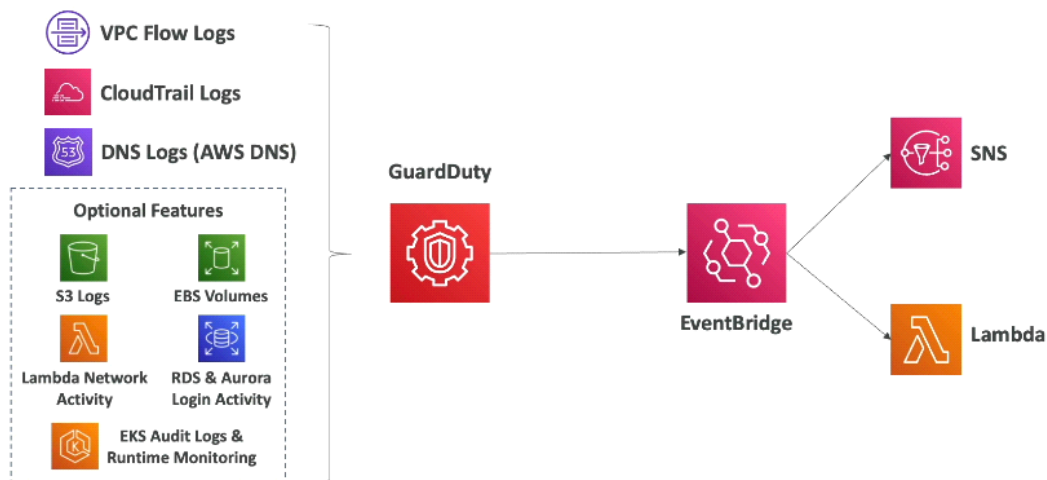
Security Services

AWS Artifact

- Not really a service
- Portal provides customers with on demand access to AWS compliance documentation and AWS agreements
- Third party reports

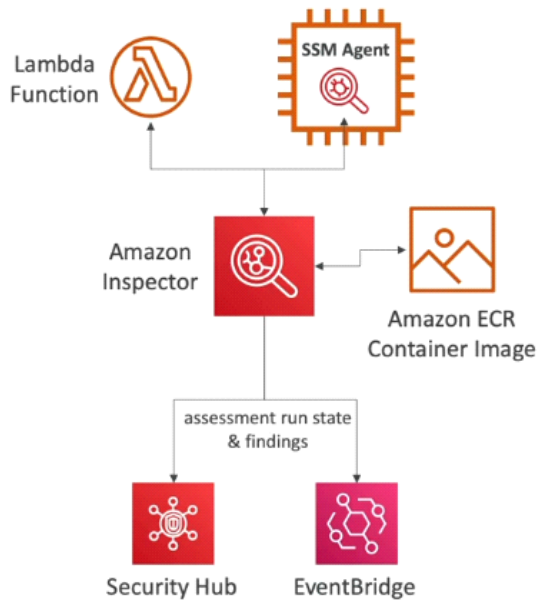
Amazon GuardDuty

- Intelligent threat discovery to protect your AWS account
- Uses ML algorithms, anomaly detection, 3rd party data
- Can setup EventBridge rules to be notified
- Can protect against Cryptocurrency attacks



Amazon Inspector

- Automated security assessment for vulnerabilities
- EC2 : Uses SSM agent for assessment
- ECR : assessment of images that are pushed
- Lambda functions

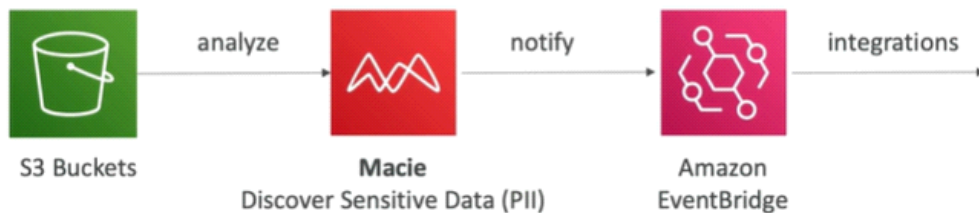


Config

- Helps in auditing and recording compliance of AWS resources
- Record configuration changes over time and compliance against rules
- Possibility of storing the configuration data in S3

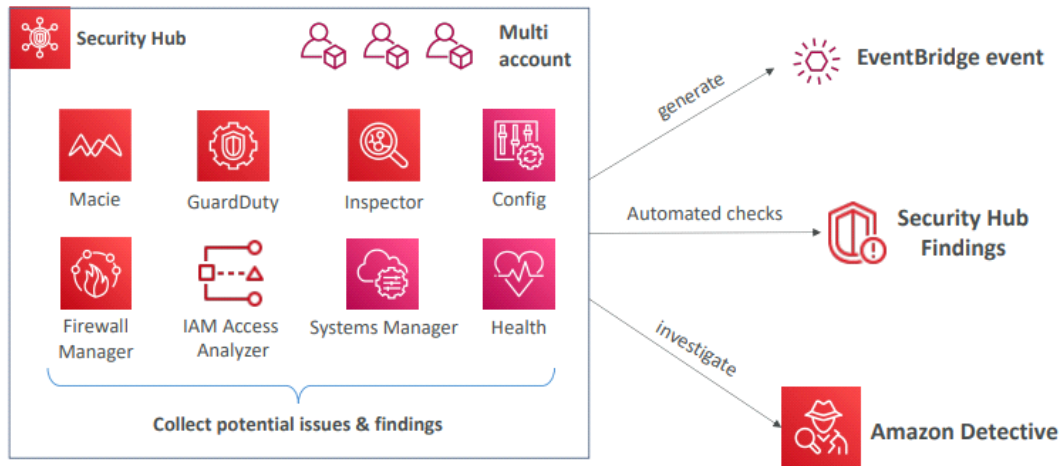
Macie

- Data security and data privacy service that uses ML and pattern matching to discover and protect your sensitive data in AWS
- Helps identify and alert you to sensitive data such as PII



AWS Security Hub

- Central security tool to manage security across several AWS accounts and automate security checks
- Automatically aggregates alerts in predefined or personal findings formats from various AWS services



Amazon Detective

- Analyzes, investigates and quickly identifies root cause of security issue or suspicious activities (using ML and graphs)
- Automatically collects and processes events from VPC logs, CloudTrail, GuardDuty and create unified view

AWS Abuse

- Report suspected AWS resources used for abusive or illegal purposes

Trusted Advisor

- High level AWS account assessment
- 6 categories
 - o Cost optimization
 - o Performance
 - o Security
 - o Fault tolerance
 - o Service limits
 - o Operational Excellence

Machine Learning

Amazon Rekognition

- Find objects, people, text, scenes in images and videos using ML
- Facial analysis & Facial search to do user verification, people counting
- Features
 - o Face liveness
 - o Face compare and search
 - o Face detection and analysis
 - o Content moderation
 - o Custom labels
 - o Text detection
 - o Labels
 - o Video segment detection
 - o Celebrity recognition

Amazon Transcribe

- Converts speech to text
- Uses deep learning process called Automatic Speech Recognition (ASR)
- Automatically removes PII using Redaction
- Multi-lingual audio

Amazon Polly

- Converts text to speech

Amazon Translate

- Natural and accurate language translation
- Allows you to localize content such as websites and applications
- Realtime or Batch

Amazon Lex

- Build chatbots quickly for your application using voice and text
- Multi-lingual
- Integration with Lambda, Connect, Comprehend, Kendra
- Bot Automatically understand the user intent to invoke the correct lambda function and also it will ask for inputs if necessary

Amazon Comprehend

- For NLP
- Uses ML to find insights and relationships in text
 - o Language of the text
 - o Extracts key phrases, places, people, brands or events
 - o Understands how positive and negative text is
 - o Parts of speech

Amazon SageMaker

- Fully managed service for developers of data scientist to build ML models

- End to end ML service

Amazon Forecast

- Uses ML to deliver highly accurate forecasts
- 50% more accurate than looking at the data itself

Amazon Kendra

- Document search service
- Extract answers from within a document of any type

Amazon Personalize

- Realtime personalized recommendations

Amazon Textract

- Automatically extracts text, handwriting, and data from any scan documents

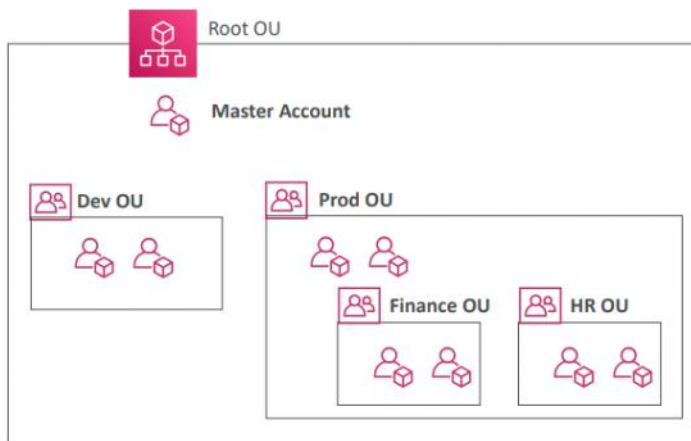
Account Management

AWS Organizations

- Global service
- Allows to manage multiple AWS accounts
- Main account is also known as **Master Account / Management Account**
- Consolidated billing, cost benefits
- API to automate AWS account creation
- Restricts account privileges using service control policies
- **Audit environment** for compliance
- **Secure environment** with policies

Organizational Units

- Department based
- Env based
- Project based
- Has pre-packaged SCP



Service Control Policies (SCP)

- Whitelist or blacklist IAM actions on OU or Account level
- SCP doesn't affect service linked roles
- Must have explicit ALLOW

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowsAllActions",
      "Effect": "Allow",
      "Action": "*",
      "Resource": "*"
    },
    {
      "Sid": "DenyDynamoDB",
      "Effect": "Deny",
      "Action": "dynamodb:*",
      "Resource": "*"
    }
  ]
}
```

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:*",
        "cloudwatch:*"
      ],
      "Resource": "*"
    }
  ]
}
```

Consolidated Billing

- Combined usage of all AWS accounts in AWS organization to share volume pricing, reserved instances and savings plan discounts
- One single bill

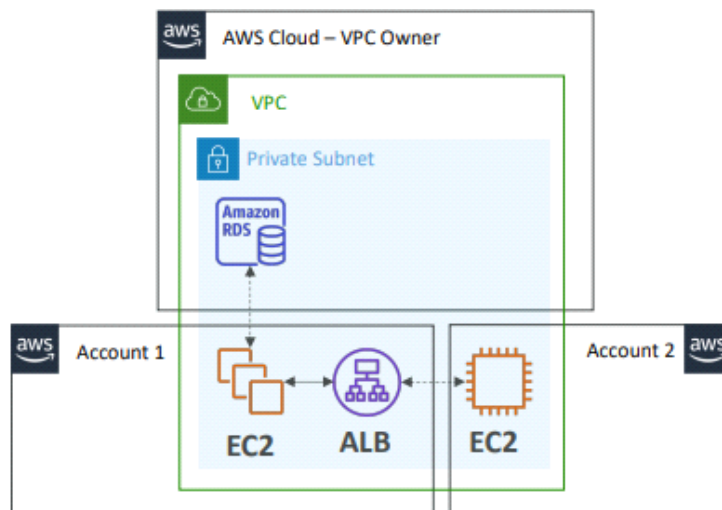


Control Tower

- Easy way to setup, govern secure and compliant multi account AWS env
- Automate policy managements using guardrails
- Runs on top of AWS Organizations
- Implements SCP

AWS Resource Access Manager (RAM)

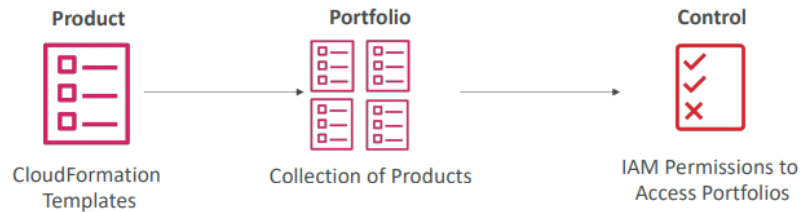
- Share AWS resources that you own with other AWS accounts, with any account or within your organization



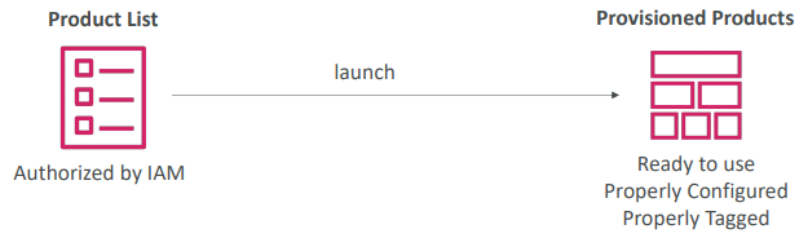
AWS Service Catalog

- Quick self-service portal to launch set of authorized products pre-defined by admins

ADMIN TASKS



USER TASKS



Billing & Support

Pricing Models

- Pay as you go
 - o Pay for what you use, remain agile, responsive, meet scale demands
- Save when you reserve
 - o Long term requirements
- Pay less by using more
 - o Volume based discounts
- Pay less as AWS grows

Savings Plan

- Commit certain amount per hour for 1/3 years
- Easiest way to setup long term commitments
- EC2 savings plan
- Compute savings plan
- ML savings plan

Compute Optimizer

- Reduce costs and improve performance by recommending optimal AWS resources for your workloads
- Helps in optimal configurations
- Uses ML to analyzer resources configuration and utilization CloudWatch metrics

Billing & Costing Tools

- **Estimating costs in cloud (Future)**
 - o Pricing Calculator :
 - Estimate the cost for your solution architecture
- **Tracking costs in cloud (Past)**
 - o Billing Dashboard
 - Get costs actually for the month, the forecast, and month to date
 - Free tier dashboard
 - o Cost Allocation Tags
 - Track your AWS costs on detailed level, organize and group them
 - AWS generated tags (aws:) / User defined tags (user:)
 - AWS Resource Groups
 - o Cost and Usage Reports
 - Dive deeper into AWS costs and usage, additional metadata
 - o Cost Explorer
 - Visualize, understand and manage your AWS costs and usage over time
 - Monthly, hourly, resource level
 - Forecast usage upto 12 months based on previous usage
 - Optimal savings plan
- **Monitoring against costs plans (Present)**
 - o Billing Alarms
 - Billing data metric is stored in us-east-1
 - It's for actual cost, not for projected cost

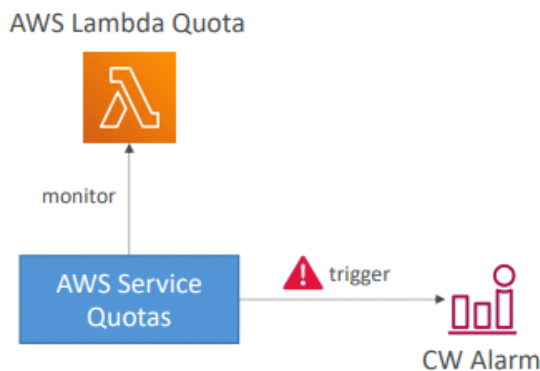
- Intended for simple alarm
- Budgets
 - Create budget and send alarms when costs exceeds the budget
 - 4 types of budgets : Usage, Cost, Reservation, Savings Plans
 - Upto 5 SNS notification per budget

AWS Cost Anomaly Detection

- Continuously monitor your cost and usage using ML to detect unusual spends
- Learn from unique, historic spend patterns to detect one time cost spike
- Sends report with analysis
- Get notified

AWS Service Quotas

- Notify you when you're close to service quota value threshold
- Create CloudWatch alarms on service quotas console



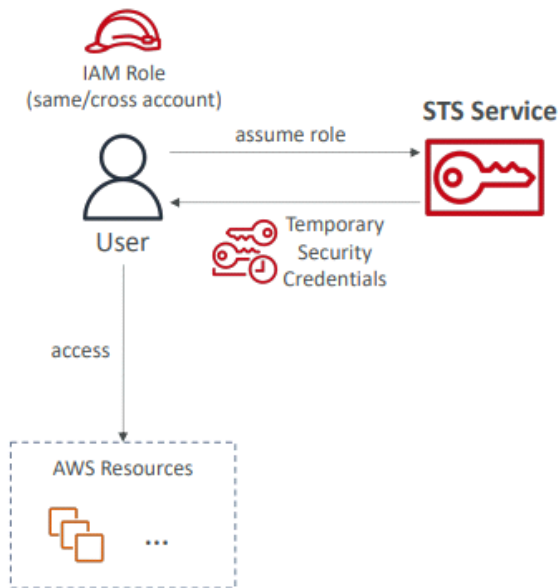
Support Plans

- Basic Plan
 - Free
 - Customer service and communication
 - AWS Trusted Advisor
 - Personal Health Dashboard
- Developer Plan
 - Business hours email access
 - Unlimited cases
- Business Plan
 - 24*7 phone, email and chat to cloud support engineers
 - For production workloads
- Enterprise Plan
 - Business-critical system down response under 15 minutes
 - Access to a Technical Account Manager & Concierge Support Team (billing & account experts)
 - Support to third party software integration to AWS

Advanced Identity

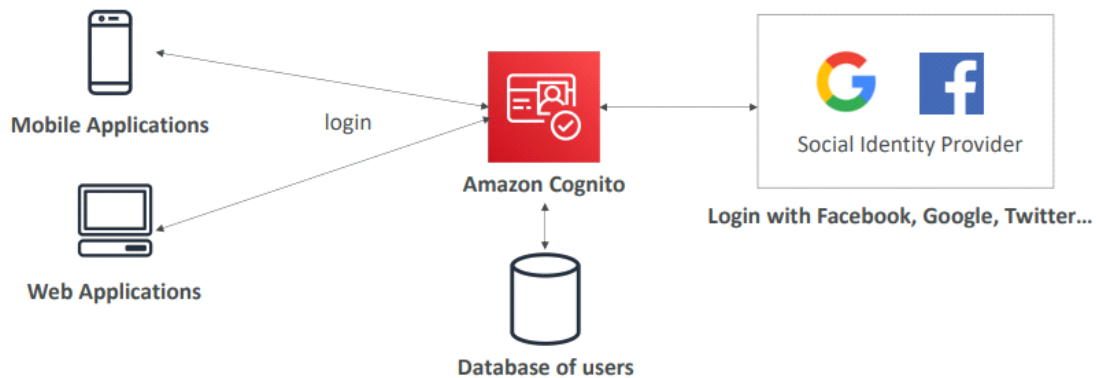
AWS Security Token Service (STS)

- Enables you to create temporary limited privileges creds to access resources
- Short term creds: configure expiration period



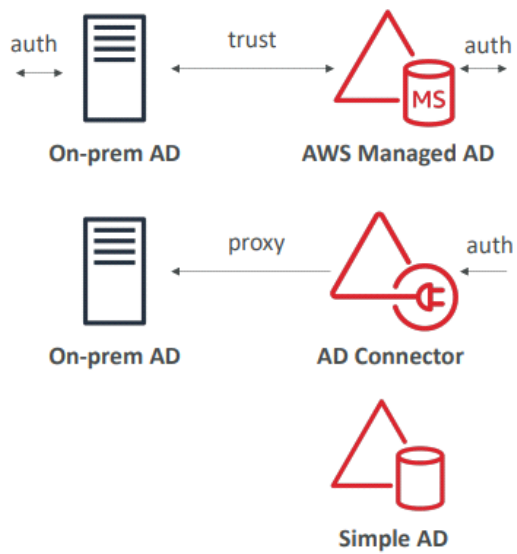
Amazon Cognito

- Identity for web and mobile app users



AWS Directory Services

- You can extend the directory services to managed Microsoft AD, AD Connector and Simple AD



IAM Identity Center

- One login for multiple AWS accounts & applications
- Successor to AWS SSO
- Identity providers store in IAM or 3rd party (AD, OneLogin, Okta)

Migration Services

Cloud Migration Strategies

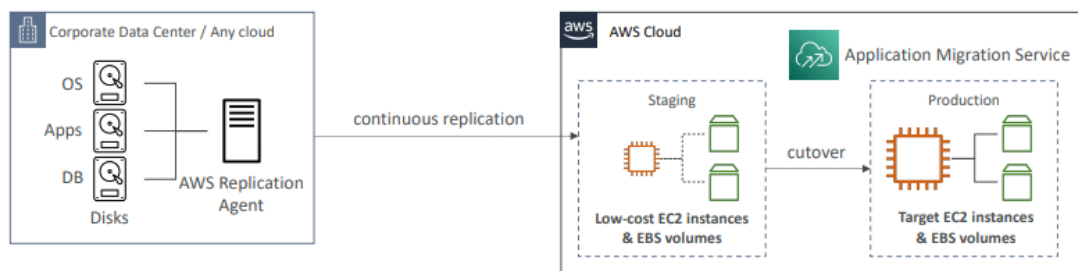
- The 7 R's
- **Retire**
 - o Turn off things you don't need
- **Retain**
 - o Do nothing for now
- **Relocate**
 - o Move apps from on premise to Cloud
 - o Move EC2 to different VPC, AWS account or region
- **Rehost "lift and shift"**
 - o Simple migrations by rehosting on AWS
 - o Migrate machines (physical, virtual, another cloud) to AWS
- **Replatform "lift and reshape"**
 - o Not changing core architecture but leverage some Cloud optimizations
- **Repurchase "drop and shop"**
 - o Moving to different product while moving to cloud
 - o Often to move to a SaaS platform
- **Refactor / Rearchitect**
 - o Reimagining how the app is architected using Cloud Native features
 - o Move from monolithic to microservices

Application Discovery Service

- Plan migrating projects by gathering info about on prem data center
- Agentless Discovery (AWS Agentless Discovery Connector)
 - o VM inventory, configuration and performance history
- Agen-based Discovery (AWS Application Discovery Agent)
 - o System configuration, system performance, running processes
- Resulting data can be viewed in AWS Migration Hub

Application Migration Service (AMS / MGN)

- Lift and shift (rehost)



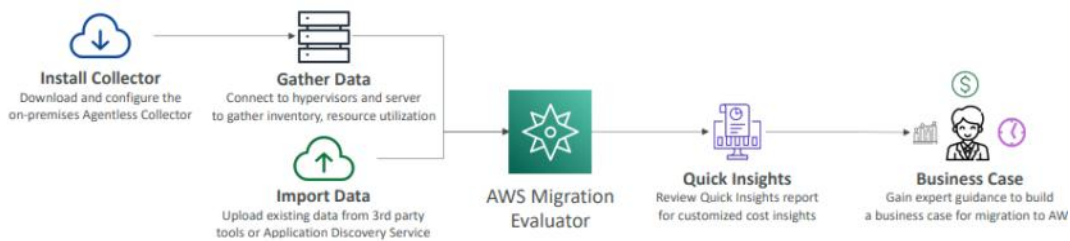
Database Migration Service (DMS)

- Quickly & securely migrate database to AWS, resilient, self-healing
- The source database remains available during the migration
- Types
 - o Homogenous migrations: source and target are same
 - o Heterogenous migrations: source and target are different



Migration Evaluator

- Helps in build data driven business case for migration
- Install agentless collector to conduct broad based discovery
- Take snapshot of on prem foot print, server dependencies
- Analyze current state, defined target state, then develop plan

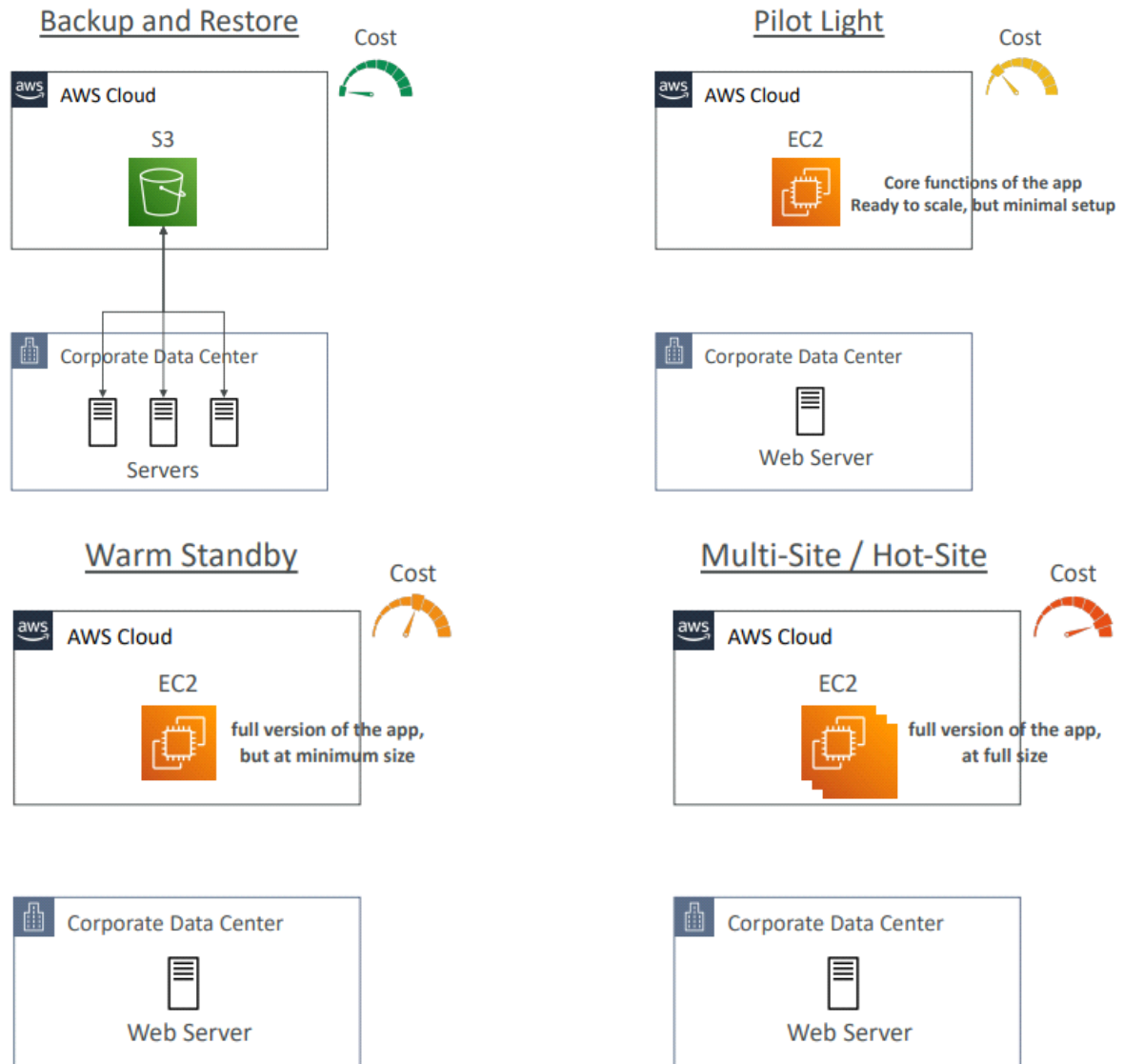


Migration Hub

- Central location to collect servers and apps inventory data for assessment, planning and tracking
- Migration Hub Orchestrator : provides pre built templates to save time and effort
- Supports migrations status updates from MGN and Database Migration Service (DMS)

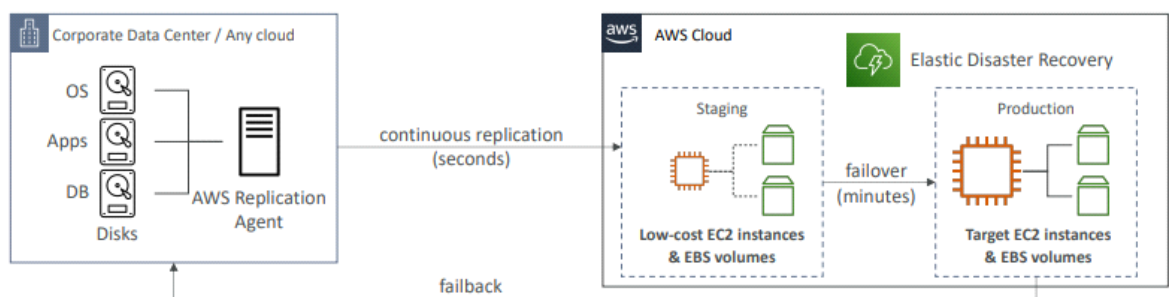
Disaster Recovery

Disaster Recovery Strategies



AWS Elastic Disaster Recover (DRS)

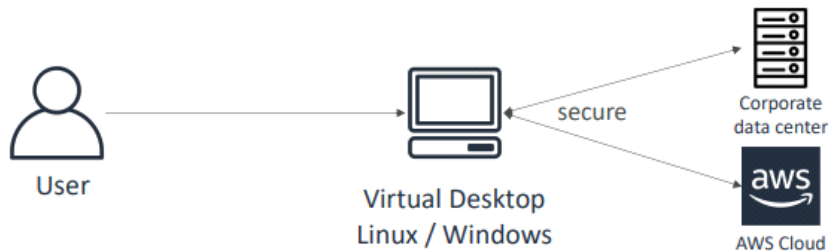
- Quickly and easily recover your physical, virtual and cloud based servers into AWS
- Continuous block level replication for your servers



AWS Other Services

Amazon WorkSpaces

- Desktop as a Service (DaaS) solution to easily provision Windows/Linux desktops
- Eliminate on premise VDI (Virtual Desktop Infrastructure)



AppStream 2.0

- Stream desktop application to web browser

IoT Core

- Easily connect IoT devices to AWS Cloud

Amazon Elastic Transcoder

- Convert media files in S3 into media files in formats required by consumer playback devices (phones, etc.)

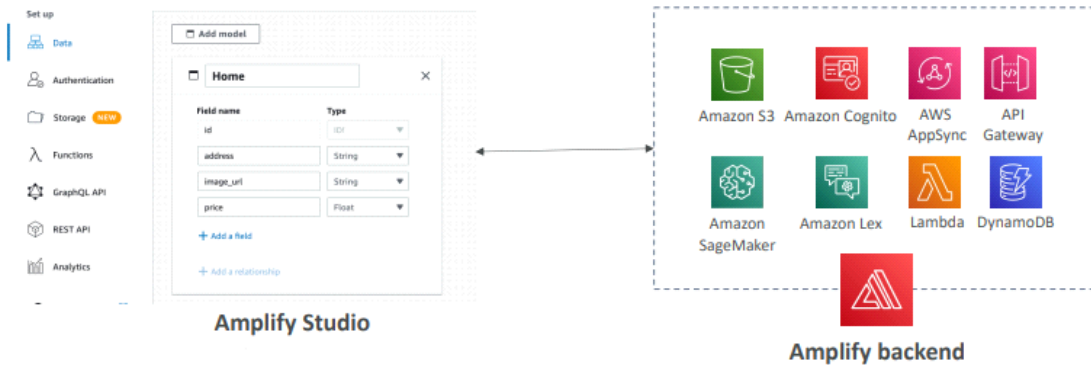


AWS AppSync

- Store and synch data across mobile and web apps in real time
- Uses GraphQL
- Real time subscriptions

AWS Amplify

- Set of tools and services to develop and deploy scalable full stack web and mobile applications
- Amplify Studio
- Authentication, Storage, REST API, CI/CD, PubSub, AI ML, etc.

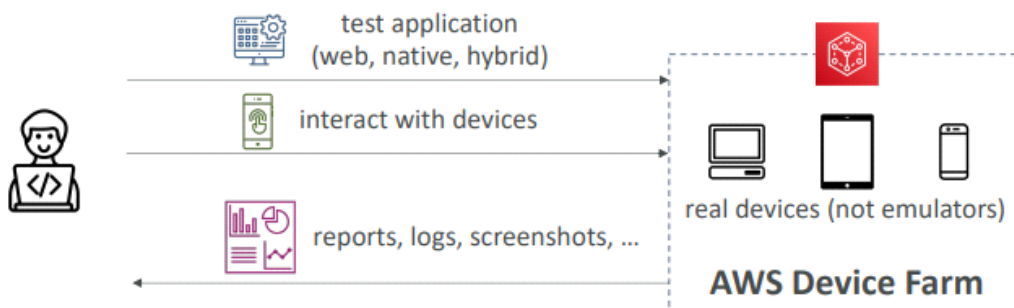


AWS Infrastructure Composer

- Visually design and build serverless applications quickly on AWS
- Generates IaC using CloudFormation
- Existing CloudFormation templates can be imported

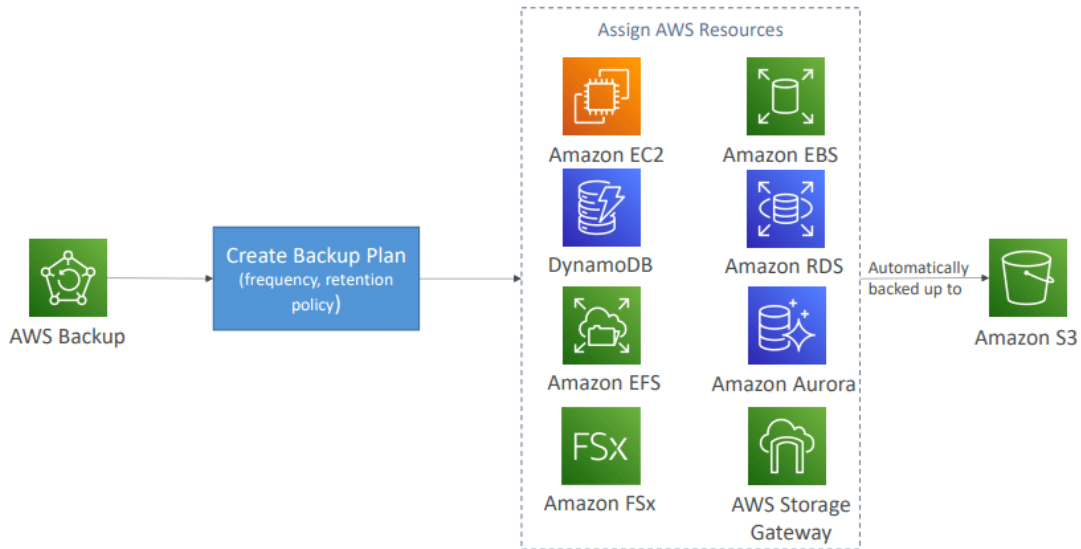
AWS Device Farm

- Tests your web and mobile apps against real mobile devices, desktop browsers
- Run tests concurrently on multiple devices
- Configure device settings



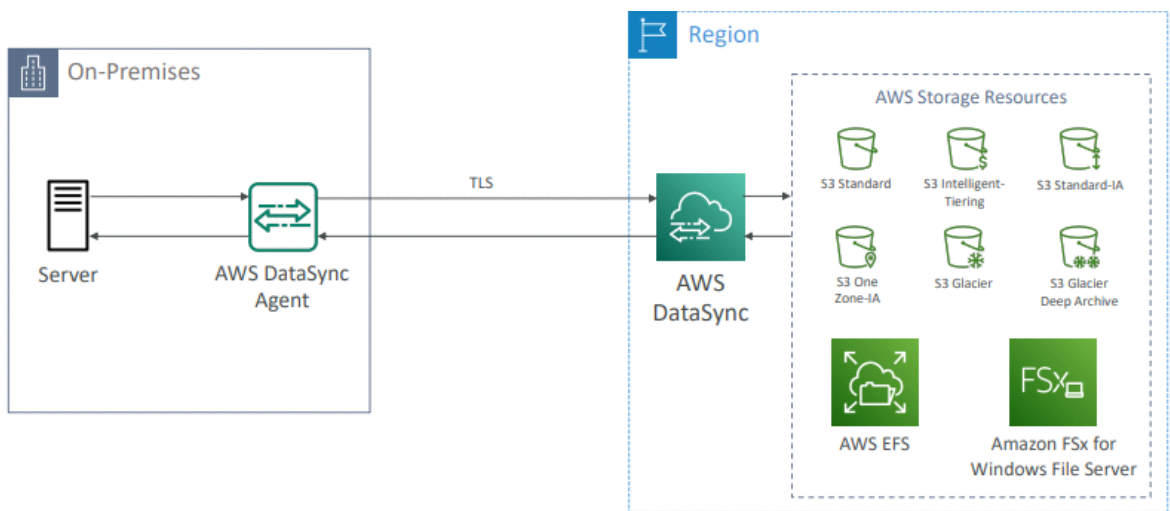
AWS Backup

- Centrally manage and automate backups across AWS service
- On demand and scheduled backups
- Point in time recovery
- Retention periods, lifecycle management, backup policies
- Cross region backup
- Cross account backup using AWS Organizations



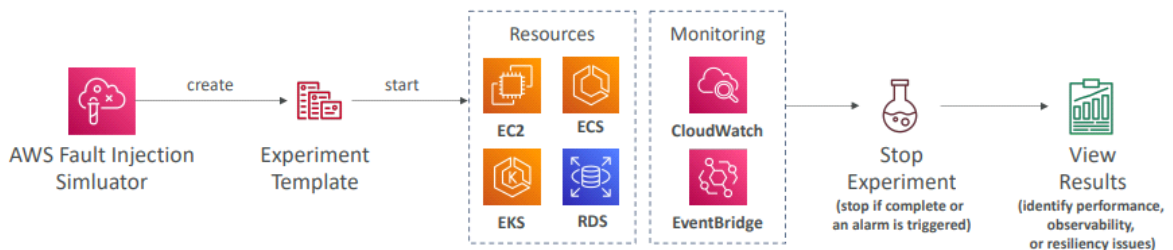
AWS DataSync

- Move large amount of data from on premise to AWS
- Can synchronize to : S3, EFS, FSx for Windows
- Replication tasks can be scheduled hourly, daily, weekly
- Replication tasks are incremental after the first full load



AWS Fault Injection Simulator (FIS)

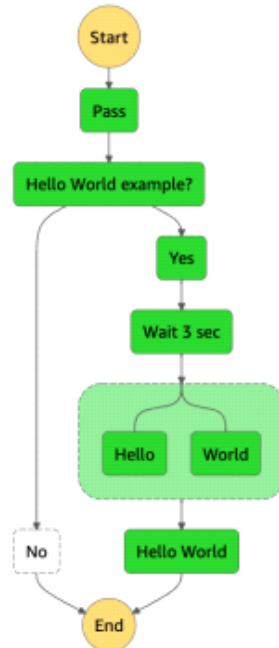
- Running fault injection experiments on AWS workloads
- Stressing application by creating disruptive events, observing how system responds and implementing improvements
- Pre-built templates



AWS Step Functions

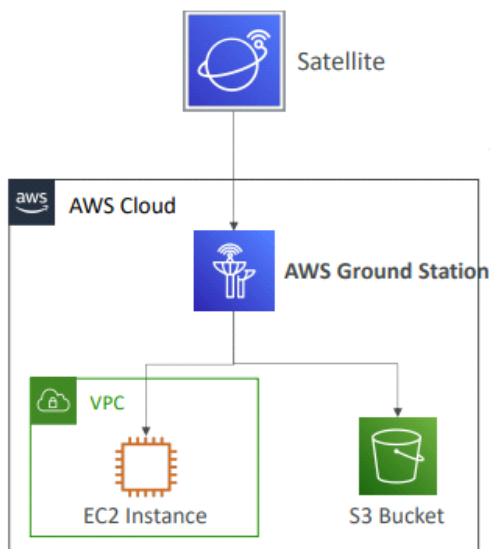
- Build serverless visual workflow to orchestrate lambda functions
- Sequence, parallel, conditions, timeout, error handling

■ In Progress ■ Succeeded ■ Failed ■ Cancelled ■ Caught Error



AWS Ground Station

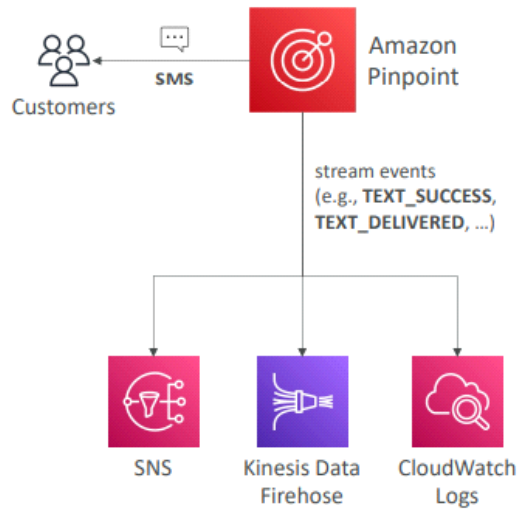
- Control satellite communications, process data
- Provides global network of satellite ground stations near AWS regions
- Download satellite data to your AWS VPC in seconds



Amazon Pinpoint

- Scalable 2way (inbound/outbound) marketing communications service
- Supports email, SMS, push, voice and in app messaging
- Personalize messages with right content to customers
- Receive replies

- Templates, schedules, full campaigns



AWS Glue

- Event driven
- Serverless computing

AWS Data Pipeline

- Web service helps process and move data between AWS compute and storage services

AWS Data Exchange

- Subscribe to 3rd party data sources

AWS Cloud9

- Cloud based IDE

AWS Architecting & Ecosystem

Well Architected Framework Guiding Principles

- Stop guessing your capacity need instead use auto scaling
- Test systems at production scale
- Automate to make architectural experimentation easier
- Drive architecture using data and need
- Improve through game days (simulate apps for flash sale days)
- Loosely coupled components

6 Pillars of AWS Well Architected Framework

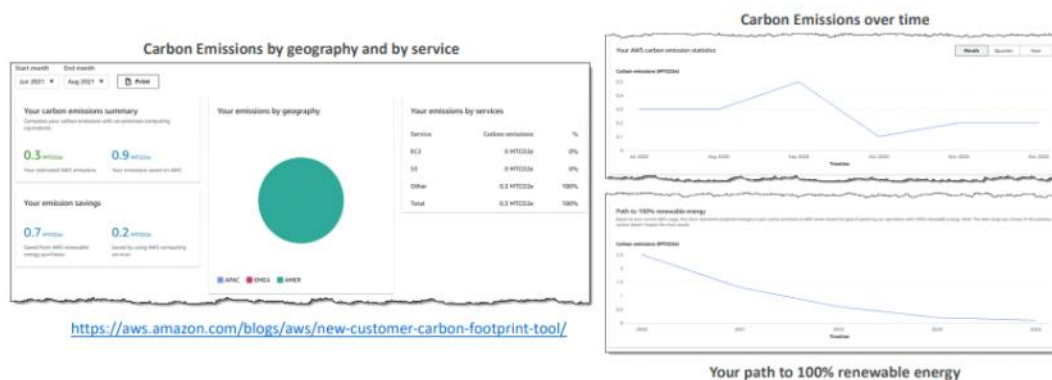
- Operational Excellence : running and monitoring systems, improving process
- Security
- Reliability
- Performance Efficiency : structured and streamlined allocation of resources
- Cost Optimization
- Sustainability

AWS Well Architected Tool

- Free tool to review your architectures against 6 pillars

AWS Customer Carbon Footprint Tool

- Track, measure, review and forecast the carbon emissions generated from your AWS usage



AWS Cloud Adoption Framework (AWS CAF)

- Helps you build and execute comprehensive plan for digital transformation through innovative use of AWS
- CAF groups its capabilities in 6 perspectives
 - o Business
 - o People : serves bridge between technology and business
 - o Governance
 - o Platform
 - o Security
 - o Operations

AWS Right Sizing

- Process of matching instance types and sizes to your workload performance and capacity requirements at lowest possible cost
- Process of looking at deployed instances and identifying opportunities to eliminate or downsize without compromising capacity or other requirements which results in lower costs

AWS Ecosystem

- AWS Blogs
- AWS Forums (community)
- AWS Whitepapers and Guides
- AWS Solutions Library
- AWS Training
- AWS Professional Services
- Partner Network
 - o Technology Partners
 - o Consulting Partners
 - o Training Partners

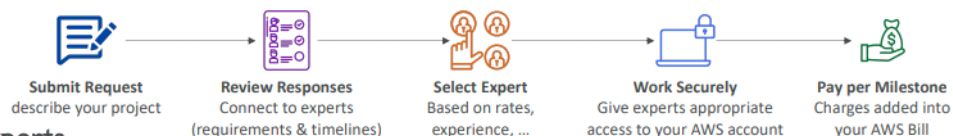
AWS Marketplace

- Digital catalog with thousands of software listings from independent software vendors
- Billed in AWS
- Can sell own solutions

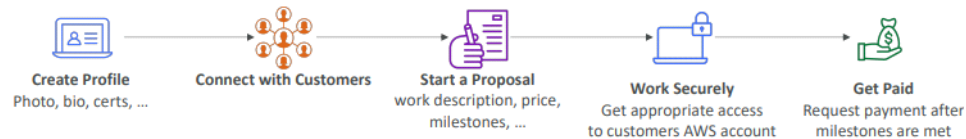
AWS IQ

- Quickly find professional help for your AWS projects
- Engage and pay AWS certified 3rd party experts for on demand project work
- Video conferencing, contract management, secure collaboration, integrated billing

• For Customers



• For Experts



AWS re:Post

- Q&A service offering
- AWS Forums

AWS Managed Services (AMS)

- Provides infrastructure and application support on AWS
- AMS offers a team of AWS experts who manage and operate your infrastructure
- AMS business hours are 24*7

