AWS Certified Solutions Architect - Associate

Tuesday, 2 April 2019 9:09 AM

- IAM
  - Users
  - Groups
  - Roles
  - **Policies** - a JSON document pertaining to permissions
  - Universal, it is not restricted in a specific region
  - "root account" the default account created when the AWS account is created. It has complete Admin access
  - Always setup MFA on your root account
  - "PowerUser" can access all AWS services but can't manage IAM
  - IAM policies can do IP matching
- S3
  - Objects size from 0 bytes to 5 TB
  - Unlimited storage
  - Files stored in Buckets
  - Bucket names should be unique
  - https://s3-eu-west-1.amazonaws.com/alvinator-files
  - Successful upload return HTTP 200
  - Can be configured to turn on access logs
  - Setup access control by:
    - Bucket Policies
    - Access Control Lists
  - Read After Write Consistency for PUTS for new Objects
  - Eventual Consistency for PUTS/DELETES for existing Objects
  - Storage Tier Types:
    - **Standard** - 99.99% availability, 99.(11 nine's) durability)
    - **S3-IA -** Infrequently-accessed. For data that is access less frequently but requires rapid ccess when needed. Lower fee than standard but you will be charged from retrieval.
    - **S3 One Zone IA -** Infrequency-accessed, one availability zone only. Only has 99.5% availability.

        - Only has 99.5% availability.
        - **S3 Intelligent Tiering** - uses machine learning process to automatically assign into a most cost effective storage tier types on objects stored in S3
        - **S3 Glacier** - Data Archiving. Retrieval may vary from minutes to hours
        - **S3 Deep Glacier** - Deeper Data Archiving. Retrieval may take 12 hours. First-byte latency around 3-5 hours
    - Encryption in Transit via SSL/TLSm
    - Encryption at Rest via:
        - S3 Managed Keys - (SSE-S3) where S3 managed all the keys
        - AWS KMS - (SSE-KMS) where you can use KMS to managed your encryption keys
        - Server Side Encryptipon with Customer Provided Keys - (SSE-C) where you can use your own encryption keys
    - Client-side Encryption is where your encrypt your data before uploading into S3
    - Cross-Region Encryption
        - Versioning must be enabled on both source and destination buckets
        - Regions must be unique
        - Files in an existing bucket are not replicated automatically. You have to manually copy them to destination
        - All subsequent files will be replicated automatically
        - Delete markers are not replicated
        - Deleting individual versions or delete markers will not be replicated
    - Lifecycle policy
        - are policies that you can set to automate moving your objects to different storage tiers
        - Can be used in conjunction with versioning
        - Can be applied on current and previous versions
    - S3 Transfer Acceleration - is a where your users can upload to your S3 buckets to the closest edge location as possible. This improves the upload performance quite dramatically.
    - S3 Signed URLS - can be used to generate S3 urls with access expiry
    - S3 Limts:
        - 3500 PUTS per second
        - Single PUT operation limit is 5GB
        - Multipart-upload limit up to 5TB
        - 100 Buckets per account
- CloudFront

    - There are more Edge locations than Regions.
    - Edge location are not just read-only, they are also writable
    - A distribution is the name of the CDN it can be:
        - Web Distribution - for caching web objects
        - RMTP Distribution - for caching media
    - CloudFront Origin supports:
        - S3
        - EC2
        - Route53
        - Elastic Load Balancer
        - Your own resources on you server as long as it is public
    - CloudFront Edge Locations are also writable
    - You can invalidate objects but you will be charged
- Snowball
    - Import/Export to S3
    - Snowball Supports 50TB or 80TB
    - Snowball Edge supports 100TB and local compuite (ie Lambda)
    - SnowMobile support 100PB per snowmobile
    - Secured by:
        - 256 bit encryption
        - Trusted Platform Module (TPM)
        - Tamper Resistant Enclosures
- Storage Gateway
    - Appliance or VM that can be used to replicate your onprem data into AWS
    - VM supports VMWare ESXi or Microsoft Hyper-V
    - Types:
        - **File Gateway (NFS)** - store objects into S3
        - **Volume Gateway (iSCSI)** - store volumes as EBS snapshots (asynchronously)
            - Stored Volumes - lets you store volume data locally while asynchronously backing up S3 as EBS snapshots. 1GB-16TB limit
            - Cached Volumes - store frequently accessed data locally in order for you to reduce the need to scale your storage locally. Up to 32TB
        - **Tape Gateway (VTL - Virtual Tape Library)** - leverage on your existing tape backup appliance in order to backup into S3. Supports backup systems such as NetBackup, Backup Exec, Veem, etc.
- EC2

- Pricing Philosophy:
  - Pay as you go
  - Pay less as you use more
  - Pay even less when you reserve instances
- It is much secure by attaching IAM roles to EC2 instead of hard coding AWS keys.
- You may now attach IAM roles to existing EC2 instances
- Bootstrap Scripts (User Data) are bash scripts that can be executed during the EC2 initialization.
- Instance meta-data can be collected from: curl http://169.254.169.254/latest/meta-data/
- Instance user-data can be collected from: curl http://169.254.169.254/latest/user-data/
- It is noyt possible to change the EC2 AZ after it is launched
- Limits
  - 20 instances per region for newly created accounts
- Underlying Hypervisors for EC2 are:
  - Nitro
  - Xen
- Types:
  - **Ondemand** - pay a fixed rate by hour (or by seconds) with no commitment
    - Uses:
      - spiky unpredictable workloads that cannot be interrupted
      - applications being tested on AWS for the first time
  - **Reserved** - provides you with capacity reservation with 1 or 3 year commitment with offer a significant discount on an hourly charge per instance
    - Uses:
      - applications with steady state or predictable usage
      - applications that requires reserved capacity.
    - Types:
      - **Standard** - up to 75% discount off ondemand instances. The more you pay upfront the and the longer the contract, the greater the discount
      - **Convertible** - up to 54% discount off ondemand instances but allows you to change the RI attributes as long as the exchange results in the creation of RI's are equal or greater

value. You cannot change the region though.
                - **Scheduled** - allows to schedule RI instances. Allows you to match your capacity requirements to a predictable recurring schedule that only requires fraction of day/week/month
                - Regional - ??
        - **Spot** - enables you to bid for whatever price you want for an instance type for even greater savings. Spot instances are taken back if you loose your bid.
            - Uses:
                - Applications that have flexible start/end times
                - Applications that are only feasible for low compute prices
                - Users with urgent computing needs for large amounts of additional capacity
            - Gotchas:
                - When AWS terminates your spot instance, you will not be charged by the partial hour. However if you terminate the sport instance yourself, you will be charged for the partial hour.
        - **Dedicated Hosts** - physical EC2 server dedicated for you to use. This can reduce costs by allowing you to use your existing server-bound software licenses. **Can be purchased On-demand (hourly)** or **can be purchased as a Reservation up which is up to 70% off the On-demand price.**
            - Uses:
                - Useful for regulatory requirements that may not support multi-tenant virtualization
                - Great for software licenses that may not support multi-tenant virtualization

- Security Groups
    - Security groups changes are applied in an instant
    - Security Groups are stateful while NACLS are stateless
    - All inbound traffic are blocked by default while all outbound traffic are allowed by default
    - You can have multiple security groups attached to an EC2 instance
    - You cannot block specific IP address using Security groups. If you need to, use NACL's instead.
    - If allow a specific inbound rule allowing traffic in, the outbound traffic

- If allow a specific inbound rule allowing traffic in, the outbound traffic will automatically be allowed back out again.
- You can only specify allow rules, but not deny rules. If you need ot use NACL's instead.E

- EBS
  - EBS volume is automatically replicated within its availability zone to protect you from component failure.
  - Root EBS volume is always in the same availability zone as the EC2 instance
  - EBS now support lifecycle policies for snapshots
  - You can change the storage size and type anytime on the fly.
  - You cannot delete a snapshot of an EBS volume that is being used as a Root volume for a registered AMI
  - In order to change the availability zone or replicate an EC2 instance, you may create a snapshot of the Root EBS volume and create an Image (AMI) on it. You can then use the Image to provision a new EC2 instance in a different availability zone. Do take note though that the Image is tied to the instance type.
  - You may also copy AMI to a different region.
  - Snapshots exist in S3.
  - Snapshots are incremental meaning it uses the delta from the previous snapshot.
  - To create a snapshot of the root volume, it is recommend to Stop the instance first.
  - You can create AMI's from volumes and snapshots
  - Types:
    - General Purpose SSD (gp2) - balance price/performance
      - Volume Size: 1GB - 16TB
      - Max IOPS: 16,000
      - API Name: gp2
      - Use Cases: Most Workloads
    - Provisioned IOPS SSD (io1) - highest performance designed for mission critical workloads
      - Volume Size: 4GB - 16TB
      - Max IOPS: 64,000
      - API Name: io1
      - Use Cases: databases
    - Throughput Optimized HDD (st1)
      - Volume Size: 500GB - 16TB

        - Max IOPS: 500
        - API Name: st1
        - Use Cases: Big Data and Data Warehouses
    - Cold HDD (sc1) - lowest cost HDD volume designed for less frequently access files
        - Volume Size: 500GB - 16TB
        - Max IOPS: 250
        - API Name: sc1
        - Use Cases: File Servers
    - EBS Magnetic HDD (standard) - previous generation HDD
        - Volume Size: 1GB-1TB
        - Max IOPS: 40-200
        - API Name: standard
        - Use Cases: Works loads with infrequently access data
- You can create encrypted snapshots from an existing unencrypted snapshot by copying.
- In order to lunch instances with encrypted root volumes you may:
    - Create a snapshot of the root volume
    - Copy the created snapshot and encrypt the copy
    - Create and AMI from the encrypted snapshot copy
    - Launch an instance using the AMI created from the encrypted snapshot copy.
- Snapshots of encrypted volumes are encrypted automatically
- Volumes restored from encrypted snapshots are encrypted automactically
- You can share snapshots but only if they are unencrypted
- Snapshots can be made public and can be shared with different AWS accounts

- AMI's
    - Types:
        - EBS
            - The root device froman instance launched from the AMI is an Amazon EBS volume created from an EBS snapshot
        - Instance Store (or ephemeral volume)
            - The root device from an instance launched from trhe AMI is an instance store created from a template stored in S3.
            - EC2 instances with Instance Store root volumes can only

be Rebooted or Terminated. They cannot be stopped. On cases where the EC2 crashes, you will loose the data stored in the ephemeral volume. Rebooting the instance will keep the data in the ephemeral volume.

- Cloudwatch
  - Monitors performance
  - Can monitor events every **5minutes by default** unless you turn on Detailed Cloudwatch which allows events every **1minute.**
  - Host Level Metrics:
    - CPU
    - Network
    - Disk
    - Status Check - of the underlying hypervisor
  - AWS Cloudtrail is a monitoring tool for all AWS actions made on console and API's. Something like a cctv for your AWS account. You may see details such as source ip details from which the calls were made.
  - You can create cloudwatch alarms which can trigger notifications
  - Cloudwatch Dashboards can be global as well as regional
  - Cloudwatch Events can help you respond to state changes in your AWS resources
  - Cloudwatch Logs - helps you aggregate, monitor, and store logs.
  - Cloudwatch logs is keep up to 15 months for terminated instances
- EFS
  - (Elastic File System) allows you to create EBS volumes that can be attached to multiple EC2 instances.
  - You need to install amazon-efs-utils to your EDC2 instances in order to mount EFS volumes
  - Can be mounted with TLS
  - Support NFSv4 (Network File System)
  - You only pay for the storage that you use (no pre-provisioning required)
  - Can scale up tp petabytes
  - Can support thousands of NFS connections
  - Data is stored across multiple AZ's within a region
  - Read after Write consistency
  - Supports life-cycle just like S3
  - EC2 using Efs should have 2049 port open
- Placement Groups

- Placement Groups
  - Two types of strategy:
    - **Clustered Placement Group** - are groups of EC2 instances within a single availability zone. This is recommended for applications that need low network latency, high network throughput, or both. Can't span across multiple AZ's
    - **Spread Placement Group** - is a group of instances that are spread across distinct underlying hardware. Recommended for applications that have a small number of critical instances that should be kept separate from each other. Can span across multiple AZ's. Limited to only 7 instances per AZ
  - Name of placement group should be unique within the AWS account
  - Only certain types of instances can be placed in a placement group:
    - Compute Optimized
    - GPU Optimized
    - Memory Optimized
    - Storage Optimized
  - AWS recommend homogeneous instance types within a placement group
  - You can't merge placement group
  - You can't move an existing instance in a placement group. You may create an AMI from the existing instance and launch that within the placement group.
- Gotchas:
    - Availability Zones per account is randomized. It doesn't mean that us-east-1a is the same as us-east-1a from a different account
    - Cloudwatch detailed monitoring gives you 1minute interval updates
    - Root Volumes in EC2 only has General Purpose SSD, provisioned IOPS and Magnetic types only
    - Key Pairs are asymmetrical. Public keys are stored on the EC2 instance while the Private keys are downloaded locally and used in order to login to the instance.
    - On and EBS-backedr instance, the default action for the root EBS is to be deleted when the instance is terminated unless the default value of DeleteOnTermination is changed. The additional EBS volumes attached to the terminated instance are persisted
    - EBS root volumes cannot be encrypted by default unless you use

a third party tool (ie. Bitlocker). Addtiional volumes can be encrypted.

- Databases on AWS
  - Multi-AZ - failover automatically to healthy RDS instance. Accessed through the same domain.
  - Read-Replica - Replicates to another database instance. Does not do automatic failover, has different domain name. Can be used to offload reads from the main database. Read Replicas has **5 copies**
  - **RDS (OLTP):**
    - Supported Engines:
      - MS SQL
      - MariaDB
      - MySQL
      - PostgreSQL
      - Oracle
      - Aurora
    - Limits:
      - 16TB storage
      - 2 Availability Zone per Multi-AZ setup
    - Runs on Virtual Machines in which you don't have access
    - RDS has reserved instances as well which also has multi-az support
    - Patching of OS and DB on RDS servers is AWS responsibility
    - RDS is not Serverless
    - Aurora is Serverless
    - Types of Backups:
      - Automated Backups
        - Allows you to recover your database to any point in time within a "retention period" which can be between **1 to 35 days**
        - Full Daily Snapshot
        - Enabled by Default
        - Backups stored in S3 and you get free storage equal to the size of your database
        - Taken within a backup window. Expect elevated latency during the backup window.
      - Database Snapshots
        - Done manually (user initiated)

        - Restoring backups will create a new RDS instance with a different endpoint
        - Changes to the backup windows are applied immediately
        - Encryption at east is supported by all six database engines. Encryption is done via KMS.
        - Once RDS instance is encrypted: read replicas, automated backups, and snapshots are also encrypted
        - Multi-AZ
            - In an event of Database Maintenance, DB instance failure, AZ failure, RDS will point to the backup instance automatically
            - For disaster recovery only
        - Read-replica
            - Asynchronously replicates data from master database to replica database
            - Read replicas can also have read replicas as well
            - Read replicas as read only
            - Supported by:
                - MySQL
                - MariaDB
                - PostgreSQL
                - Aurora
            - Used for scaling
            - You can have up to **5 read replicas** for a given database
            - Read replicas have their own DNS endpoint
            - Read replicas can also have Multi-AZ
            - Read replicas can be promoted to master but this will break the replication
            - Read replicas can be Aurora or MySQL
            - You can have a read replica in a second region
            - You cannot turn on read replica if the automated backups are turned off
    - DynamoDB (NoSQL)
        - Stored on SSD storage
        - Spread across 3 geographically distinct data centers
        - Eventual Consistent Reads (default)
            - Consistency usually reached within a second
            - Best Read Performance
        - Strongly Consistent Reads
            - Returns a result that reflects all writes that received a successful

        - Returns a result that reflects all writes that received a successful response prior to the read
    - Limits:
        - 400KB for combined Name and Value attributes

- Redshift (OLAP)
    - Fully managed petabyte-scale data warehouse. Used for Business Intelligence.
    - Can be configured as follows:
        - Single Node (160gb)
        - Multi-Node
            - Leader Node - which receives trhe queries
            - Compute Node - store data and perform queries and computations. Up to **128 compute nodes**
    - Uses columnar data store
    - Doesn't require indexed and materialized views
    - When storing data, Redshift automatically selects the best compression algorithm
    - Automatically distributes data and query across all nodes. It also makes it easy to add new nodes for performance as your data warehouse grows.
    - Backups
        - Enabled by Default with a 1 day period
        - Max retention is 35 days
        - Always attempt to maintain three copies of data backed up in S3
    - Priced by compute node hours, backup and data transfer
    - You are not charged for the leader node
    - Security
        - Encrypted in transit using SSL
        - Encrypted at rest using AES-256 encryption
        - Keys managed by Redshift by default
            - Manage your key through HSM
            - Manage your key through KMS
    - Currently only available in 1 AZ
    - Can restore snapshots to a different AZ
- Elasticache
    - In-memory database which can be used for caching frequently access objects to increase database and web application performance
    - Engines supported:

        - Memcache
            - Can scale horizontally
            - Multi-threaded performance
            - No Multi-AZ support
            - No Persistence
        - Redis
            - Has Multi-AZ support
            - Has Persistence
            - Has PubSub
            - Can do backups and restores

    - Aurora
        - Provides up to fives time more performance than MySQL
        - Compatible for both MySQL and PostgreSQL engines
        - Starts with 10GB, scales in 10GB increments to 64TB (Storage Autoscaling)
        - Compute resources can scaled up to 32vCPUs and 244GB memory
        - Maintains 2 copies of your data contained in each availability zone, with minimum 3 availability zones. 6 copies of your data.
        - Designed to handle the loss of two copies of your data without affecting the database write availability and up to three copies without affecting the read availability
        - Aurora storage is self-healing
        - Supports Multi-Master (Active-Active)
        - Types of Replicas:
            - Aurora Replicas (15)
                - Automated Failover: Yes
                - Failover Data Loss: None
            - MySQL Replicas (5)
                - Automated Failover: No
                - Failover Data Loss: potentially minutes of data loss
        - Backups:
            - Always enabled
            - No performance impact on backups, snapshots
            - Snapshots can be shared across different AWS accounts
        - You can migrate your existing MySQL RDS instances to Aurora by creating Aurora Read Replicaes and promoting it.
- **Route53**

- IPv4 32bit field space - 4billion addresses.
- IPv6 128bit field space - 340 undecillion addresses.
- .com.ph
  - Top level domain name: .com (also known as root zone)
  - Second level domain name: .ph
- Top level domain names are controlled by IANA: https://www.iana.org/domains/root/db
- Domain Registrar is an authority that assign domain namesdirectly under one or more top level domain names. These domains are registered in InterNIC which is a service of ICANN which enforces the uniqueness of domain names across the internet. Each domain name becomes registered in a central database known as WhoIS database.
  - Amazon
  - GoDaddy
  - Name.com
- DNS Types
  - **SOA** - (Start of Authority) Record
    - Name of the server that supplied the data for the zone
    - The administrator of the zone
    - The current version of the data file
    - The default number of seconds for the time-to-live file on resource records.
  - **NS** - (Name Server Records)
    - They are used by Top Domain servers to direct traffic to the Content DNS Server which contains the authoritative DNS records
  - **A -** is the fundamental type of DNS record. The "A" Record stands for "Address". This is used by a computer to translate a domain name into IP address
  - **TTL** - "Time to Live" is the length of the DNS record to be cached on a resolving server or local computer. The lower the TTL, the faster the IP gets propagated throughout the internet.
  - **CNAME** - "Canonical Name" can be used to resolve one domain name to another
  - **MX** - "Mail Exchanger" records. Specifies the mail server responsible of accepting the email messages on behalf of a domain name.
  - **PTR -** Are used for "Reverse DNS". Used to find out the DNS for a given IP.
  - **Alias** - records that are used to map resource record sets in your

        - **Alias** - records that are used to map resource record sets in your hosted zone to ELB, S3, and CloudFront distributions that are configured as websites. They pretty much work like CNAME records (ie. www.adelagon.com -> elb12345.elb.amazonaws.com).
    - Difference between CNAME and Alias: CNAME can't be used for naked domain names (zone apex record). You can't have a CNAME for http://adelagon.com, it must be either an A record or an Alias.
    - ELB's doesn't have a pre-defined IPv4 address ; you resolve to them using DNS name.
    - You can now buy domain names directly from AWS
    - It can take up to three days to register depending on the circumstances
    - You can configure Route53 to automatically remove a record if the serving resource fails the health check.
    - Minimum ttl you can set in Route53 is 60s
    - Limits:
        - 50 domain names per account
    - Route53 Routing Policies:
        - Simple Routing
            - You can only have A record with multiple IP addresses.
            - If you specify multiple values in a record, Route53 will return all the IP address in random order.
        - Weighted Routing
            - Allows you to split your traffic based on a different weights assigned.
        - Latency-based Routing
            - Allows you to route your traffic on a specific region based on the lowest network latency or your end user
        - Failover Routing
            - Are used hen you want to create and active/passive setup. Route53 will monitor the health of your primary site using health check.
        - Geolocation Routing
            - Allows you to choose where your traffic ill be sent based on the geographic location of your users (ie the location of the DNS queries originated)
        - Geoproximity Routing
            - Allows you to route traffic to your resources based on the geographic location of your users and your resources. You can also optionally choose to route more traffic or less o a given

resource by specifying "bias" value.
        - To use geoproximity routing, you must use Route53 Traffic Flow
    - Multi-value Answer Routing
        - Very similar to Simple Routing except that it does healthchecks on each record set provided to make sure that the records returned are healthy records.

- **VPC**
    - Lets you provision a logically isolated section of AWS cloud where you can launch AWS resources a virtual network at you define. You have complete control over your virtual networking environment such as IP ranges, subnets, etc.
        - Launch instances on the specific subnet
        - Assign custom IP ranges in each subnet
        - Configure route tables between subnets
        - Create Internet gateway and attach it to your VPC
        - Much better security control over your AWS resources
        - Assign security groups per instance
        - Have subnet network access control lists (ACLS)
    - Consists of:
        - IGW (Internet Gateways) or Virtual Private Gateways
        - Route Tables
        - Network Access Control Lists
        - Subnets
        - Security Groups
    - 1 Subnet = 1 Availability Zone
    - Private IP ranges:
        - 10.0.0.0 - 10.255.255.255 (10/8 prefix)
        - 172.16.0.0 - 172.31.255.255 (172.16/12 prefix)
        - 192.168.0.0 - 192.168.255.255 (192.168/16 prefix)
    - TIP: Visit http://cidr.xyz
    - Limits:
        - 200 subnets per VPC
        - 125 peering connections per VPC
        - Allowed block size /16 netmask (65536) to /28 netmask (16). Do take note that AWS reserves 5 IP addresses. Ie. For 10.0.0.0:
            - 10.0.0.0 - network address
            - 10.0.0.1 - reserved for VPC router
            - 10.0.0.2 - reserved by AWS. By default is the DNS server

        - 10.0.0.3 - reserved by AWS for future use
        - 10.0.0.255 - network broadcast address. AWS **Do Not Support Broadcast in VPC**
    - 1 Internet Gateway per VPC
    - 5 VPC per region
- Default VPC
    - Is a preconfigured VPC when you create an AWS account. Should not be used for production VPC
    - All subnets in Default VPC have route out to the internet
    - Each EC2 instance has both private and public address
- VPC Peering
    - Allows you to connect one VPC to another
    - Instances behave as if they were on the same private network
    - You can also peer VPC's from other accounts as well as other VPC's in the same account
    - You can peer across different regions
    - Peering is a Star configuration: ie. 1 Central VPC peers with 4 others. No transitive peering.

- You can specify VPC to be hosted on Dedicated rather than Multi-tenancy but there will be costs.
- When you create a VPC, AWS will automatically create the following:
    - A route table
    - A Network ACL
    - Security Group
- NAT Instances & NAT Gateways
    - NAT Instances are individual EC2 instances serving NAT (soon to be deprecated)
        - Single point of failure
        - If overwhelmed, increate the instance size or HA via Autoscaling groups
        - Always behind a security group
        - You have top explicitly disable the source/destination checking for it to work.
    - NAT Gateways are highly-available managed by AWS instances serving NAT (recommended)
        - Redundant inside the availability zone
        - Starts at 5Gbps then autoscales up to 45Gbps

        - Not associated with Security Groups
        - No need to patch server as you don't manage an EC2 instance
        - Automatically assigned with an IP address
        - Remember just to update the route tables
        - NAT Gateways is assigned to a specific availability zone. In order to improve resiliency, you may opt to create multiple NAT Gateways and assign the resources on that specific availability zone to use that NAT Gateway.
    - NACL
        - Newly created NACL's are Deny by default
        - Newly created VPC's automicatically comes with a default NACL which by default allow all traffic inbound/outbound
        - Each subnet must be associated with NACL. If not, it will be automatically associate with the default NACL.
        - NACL 's allow you to block specific IP address as opposed to Security Groups
        - You can associate NACL with multiple subnets although a subnet can only be associated with one NACL. If you associate a subnet with a new NACL, it will loose the its associate with the previous NACL.
        - You have to create ephemeral port for outgoing for resources serving web
        - NACL rules are done in a chronological order. If you allow before deny, the rule will follow allow.
        - NACL rules take effect immediately
        - It is recommended to create rules in increments of 100
        - When create LoadBalancers for resources in VPC, it should have at least **2 public subnets**
    - VPC Flowlogs
        - Is a feature that enables you to capture information bout the IP traffic going to and from network interfaces within your VPC.
        - Flow log data is stored using Cloudwatch logs or S3 bucket
        - Can be created in three levels:
            - VPC
            - Subnet
            - Network Interface
        - You cannot enable Flow Logs for VPC's peered to your VPC unless the peer VPC is in your account
        - You cannot tag a flow log
        - After you create a flow log, you cannot change it configuration

        - After you create a flow log, you cannot change it configuration
        - Not all IP traffic are monitored
            - Traffic generated by instances when they contact the AWS DNS server. If you use your own DNS server then that traffic will be logged
            - Traffic generated by a windows instance for amazon windows license activation
            - Traffic to 169.254.169.254
            - DHCP traffic
            - Traffic to the reserved IP address or the default VPC router
    - Bastion Host
        - A special purpose computer in a network that is designed to withstand attacks. This computer generally host applications that are only needed to access internal computers in the network
        - It is hardened due to its purpose and often sits outside of a firewall or in a demilitarized zone (DMZ)
        - Primarily an EC2 server that is accessible by trusted users and used to administer computes inside the network
        - There's bastion server AMI's on the AWS marketplace.
        - You cannot use a NAT gateway as a bastion host
    - Direct Connect
        - Establish a private connectivity between your onpremise servers to AWS
        - Provide more consistent network performance
        - Useful for high throughput network workloads
    - VPC Endpoints
        - Enables you to privately connect your VPC supported AWS services and VPC endpoint services powered by PrivateLink withour requiring an Internet Gateway, NAT device, VPN connection, or AWS direct connect. Instances in your VPC do not require public IP addresses to communicate with resources in the service. Traffic between our VPC the other service does not leave the Amazon network.
        - Are virtual devices. They are horizontally scaled, redundant, and highly available VPC components that allow communication between instances in your VPC and services without imposing availability risks or bandwidth constraints.
        - Types:
            - Interface Endpoints
                - An Elastic Network Interface (ENI) that serves as an entry

point for traffic destined to a supported AWS Service (ie. API Gateway, Cloudwatch, SQS, etc.
        - Gateway Endpoints
            - Supports S3 and DynamoDB
    - Gotchas:
        - Newly created Subnets are associated to the main route table which is public (may be a security concern). Therefore it is recommended to reconfigure the main route table to be private and the create a separate route table for public.
        - Security Groups **Do Not Span VPC's**
        - You should disable source/destination checks on a NAT instance
        - A VPN Connection consists of Customer Gateway and Virtual Private Gateway
        - "E-gress only Internet Gateway" allow IPv6-based traffic within a VPC to access internet while denying internet-based resources to connect to the VPC
        - AWS now allows you to do pen tests within your VPC without alerting them anymore
        - In AWS VPC, instances does retrain their Private IP
        - By Default new instances in new subnets can communicate to each other across AZ's
- **ELB (Elastic Load Balancer)**
    - Types:
        - ALB (Application Load Balancer)
            - Best suited to load balance HTTP/HTTPS traffic.
            - Application-aware
            - Operates at Layer-7
        - NLB (Network Load Balancer)
            - Best suited for load balancing TCP traffic where extreme performance is required
            - Operates at Layer-4
            - Capable of handling millions of requests per second while maintaining ultra-low latencies
        - ELB (Classic Load Balancer)
            - Legacy load balancer that both operates as Layer 7 and Layer 4 with strict limitations
            - Responds with 504 error when the backend server times out
            - Support X-Forwarded-For header so that the backend services

will still get the originating IP
        - Routes each request independently to the registered EC2 instance with the smallest load
    - Limits:
        - A target group can only be assigned to one LB
        - Can only Load Balance within a region
        - Default TTL is **60 seconds**
    - You can create internal load balancers within a VPC
    - Instances monitored by ELB are either:
        - InService
        - OutOfService
    - Healthchecks checks the instance health by visiting it
    - Target groups arew used to group together instances for LoadBalancer HealthChecks and routing
    - Load Balancers always have DNS names and never an IP address
    - **Sticky Sessions** allows you to bind user's session to a specific EC2 instance. This is both supported in ALB and Classic LB
    - Cross-Zone Load Balancing
        - Without - Load Balancers are limited to balance within a specific AZ
        - With - Load Balancers can balance across AZ
    - **Path-based Routing** - you can create a listener with url path pattern rules that can route to specific target groups
- Autoscaling Groups
    - When you delete an autoscaling group, the instances are terminated as well
- CloudFormation
    - Is a way of completely scripting your cloud environment
    - AWS Quick Start is a bunch of CloudFormation Template created by AWS Solution Architrects
- ElasticBeanstalk
    - Quickly Deploy and Manage applications. Just upload the code and Elasticbeanstalk will automatically handles the infrastructure provisioning.
- SNS (Simple Notification Service)
    - Limits:
        - 10,000 email sent
        - 10 million subscribers per topic is a soft limit
- SQS (Simple Queue Service)
    - Is a distributed queue system that enabled web applications to quickly and reliably queue messages that one component generates to be consumed by

another component.
    - A queue is a temporary repository for messages at are awaiting processing
    - With SQS you can decouple your applications they run independently
    - You can trigger autoscaling based on the SQS queue
    - SQS is pull-based not push-based (use SNS instead)
    - **VisibilityTimeOut** is the amount of time the message is invisible in the SQS queue when a consumer picks up the message. If the consumer finishes before the timeout expires, the messages will be deleted. If job is not processed within the timeout setting, the message will be visibile again and can be picked up by other consumers.
    - SQS using http long polling. Long polling doesn't return a response until a message arrives the message queue or the long poll times out.
    - Message-oriented API
    - Types:
        - Standard
            - Guarantees delivery at least once
            - Occasionally, more than one copy of a message might be delivered out of order.
            - Provided best-effort ordering
        - FIFO
            - Complements the standard queue. Delivers First-In First-Out and exactly-once processing;
            - The order of messages re sent and received strictlt preserved and a message is delivered once and remains until a consumer processes and deletes it;
            - Duplicates not introduced in the queue
    - Limits:
        - Message size up to 256kb of text in any format (adjustable)
        - FIFO queues are limited to 300 transactions per second.
        - Standard Queues are nearly unlimited in terms of tps
        - Messages in the queue can be kept from 1minute up to 14 days; the default retention period is 4 days.
        - Maximum VisibilityTimeOut is 12 hours
- SWF (Simple Workflow Service)
    - SWF makes it easy to coordinate work across distributed application components
        - Business workflows
        - Media processing
    - Task-oriented API

    - Task-oriented API
    - Ensures at a task assigned only once and is never duplicated
    - SWF keeps track of the tasks and events an application unlike SQS in which you may have to develop yourself using multiple queues
    - Can coordinate between executable code, scripts, web service calls, even human actions
    - SWF Actors:
        - Workflow Starters - an application that initiates (start) of a workflow
        - Decider - controls the flow of activity tasks a workflow execution and decides what to do next
        - Activity Workers - carry out activity tasks
    - Limits:
        - Workflow executions can last up top 1 year
- SNS (Simple Notification Service)
    - Push-based delivery (no polling)
    - Makes it easy to send out notifications from the cloud
    - Supports:
        - Push Notifications
        - SMS messages
        - Email
        - HTTP endpoint
    - Allows you to group multiple recipients using **topics**
    - Messages are stored redundantly across multiple-AZ's
- **Elastic Transcoder**
    - Media transacoder in the cloud
    - Provides transcoding presets depending on the target device
    - Pay based on the minutes that you transcode and the resolution at which you transcode
- **API Gateway**
    - Fully managed service that makes it easy for developers to publish, maintain, monitor, and secure API's at any scale
    - Frontdoor for:
        - Lambda
        - Web Applications
        - DynamoDB
    - Can send each API endpoint to a different target
    - Scales effortlessly
    - Track control usage by API key
    - Throttle requests, prevent attacks

    - Throttle requests prevent attacks
    - Can be connected to Cloudwatch to log all requests for monitoring
    - Maintain multiple version of your API
    - You can set request/response transformations
    - You can use your custom domain and now supports AWS Certificate Manager: free SSL/TLS certs
    - You can also enable API caching in order to reduce the workload of your API servers. It caches it by setting a defined TTL in seconds
    - CORS (Cross-origin Resource Sharing) is a mechanism hat allows restricted resources on a web page be requested from another domain outside the domain from which the resource was served. You can enable CORS on API gateway
- **Kinesis**
    - Streaming Data is a data that is generated continuously by thousands of data resources in small sizes. Kinesis is a platform on AWS to send your streaming data to.
    - Makes it easy to load and analyze streaming data
    - Provides you with the capability to build your custom applications for your business needs
    - Types:
        - Kinesis Streams
            - By default can store data up to 24 hours can be adjusted up to 7 days
            - Data is contained in a **shard**:
                - 5 transactions per second for reads
                - Maximum total data read rate of 2MB per second
                - Up to 1000 records per second for writes
                - Maximum total data write rate of 1MB per second (including partition keys)
            - Consumers then processes the data stored in a **shard**
            - Data Capacity of your stream a function of number of shards that you specify on a stream (Sum)
        - Kinesis Firehose
            - Optional to have lambda functions
            - No Data Persistence
        - Kinesis Analytics
            - Allows to create analytics on the data incoming on the fly
- **Web Identity Federation & Cognito**
    - Web Identity Federation allows your users to access AWS resources after

- Web Identity Federation allows your users to access AWS resources after they have successfully authenticated with a web-based identity provider (ie. Amazon, Facebook, or Google). Following a successful authentication with the provider, you get an authentication code that can be traded for temporary AWS credentials
    - Cognito features:
        - Sign-up/Sign-in
        - Access for guest users
        - Acts as identity broker between your application Web ID providers
        - Synchronizes user data or multiple devices
        - Recommended for all mobile apps
    - **User Pools**
        - Users can sign-up/sign-in directly to the User Pool or using any WebID provider.
        - Successful authentication generates a **jwt** (JSON Web Token)
    - **Identity Pools**
        - Able to provide temporary AWS credentials to access AWS Services
    - Cognito uses SNS to send a notification to all the devices associated with a given user identity whenever there's a change in the data in the cloud
- **Lambda**
    - Is a compute service wherein AWS already takes take of the server that will host your function. No need to deal with managing servers anymore, just upload your code.
    - Is an event-driven service where AWS Lambda runs our code in response to events. 1 event = 1 function
    - Lambda functions may also trigger other lambda functions
    - Can also be used to run as a compute service to run your code response to HTTP requests sing API Gateway
    - Lambda Scales out automatically.
    - You may use AWS X-ray to debug what is happening to your lambda functions
    - Priced by:
        - number of requests ($0.20 per 1 miliion requests thereafter)
        - Duration - calculated to the time until your function returns or terminates rounded up to the nearest 100ms. The price depends on the amount of memory you allocate per function. $0.00001667 for every GB-second used
    - Limits:
        - 15 minutes execution time

        - Memory up to 128MB to 3,008MB in 64MB increments
        - 250MB unzipped size of the code
        - 5 layers of application?

- Amazon Guard Duty
    - Anomaly Detection
    - Powered by Machine Learning
- Amazon Macie
    - Data Security Automation
- AWS Support Plans
    - Basic
    - Developer
    - Business - Includes TrustedAdvisor
    - Enterprise - Includes TrustedAdvisor
- AWS Trusted Advisor
    - Give insights to the following areas:
        - Cost Optimization
        - Performance
        - Security
        - Fault Tolerance
        - Service Limits
- AWS Organizations
    - Central Management of multiple accounts
        - Policies
        - Consolidated Billing
        - Categorize Workloads using Groups
        - Helps with security, audit, and compliance
- AWS Shared Security Responsibility Model
    - AWS Responsibilities:
        - AWS is responsible for the security of the infrastructure and foundation services.
        - Reduces the operational burden to customers as AWS operates, manages, controls the components from the host operating system and virtualization layer, down to the physical security of the facilities in which the services operate.
    - Customer Responsibilities:
        - Customer is responsible with the security of his virtual environment, data, and applications

        - Responsible with anything that runs on the guest environment.

TODO

- https://aws.amazon.com/whitepapers/
- Understand OLTP vsa OLAP
- Read FAQ's for:
    - S3
    - EC2
    - RDS
    - ELB
    - DynamoDB
- Visit AWS Quick Starts
- Read: https://d1.awsstatic.com/whitepapers/AWS_Cloud_Best_Practices.pdf
- Read: https://d1.awsstatic.com/whitepapers/architecture/AWS-Operational-Excellence-Pillar.pdf