AWS Associate Solutions Architect Study - ACloud.Guru

Ian Berry 09/07/2019 v1 AWS Associate Solutions Architect Study Notes

Timeline:

- **13/11/2018 - Signed up for A Cloud Guru**
- **March 2019 - Started practice exams**
- **23/05/2019 - Sat exam and passed with 91%**

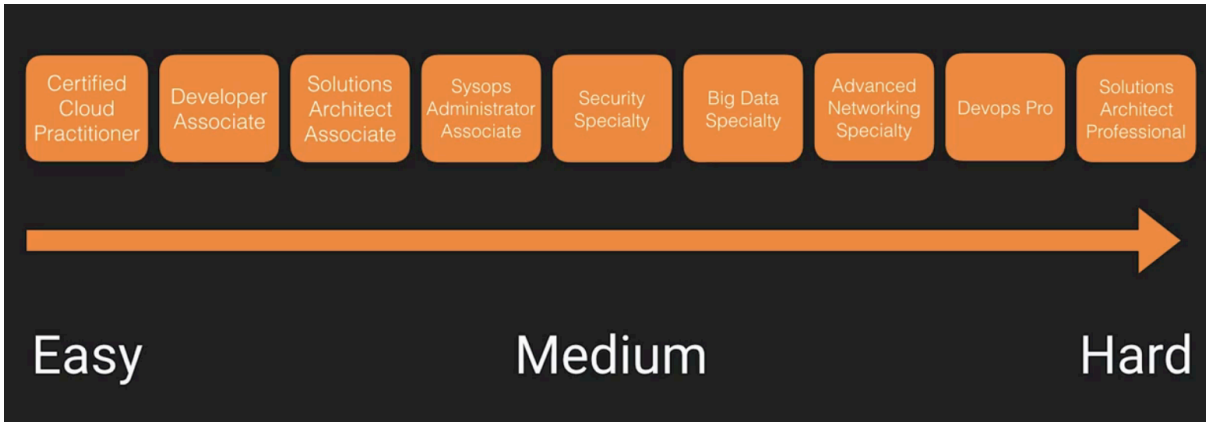
AWS certifications

https://aws.amazon.com/certification/certified-solutions-architect-associate/

- https://aws.amazon.com/whitepapers/

https://acloud.guru/course/aws-certified-solutions-architect-associate/dashboard

- AWS this week (5 min update): https://acloud.guru/aws-this-week

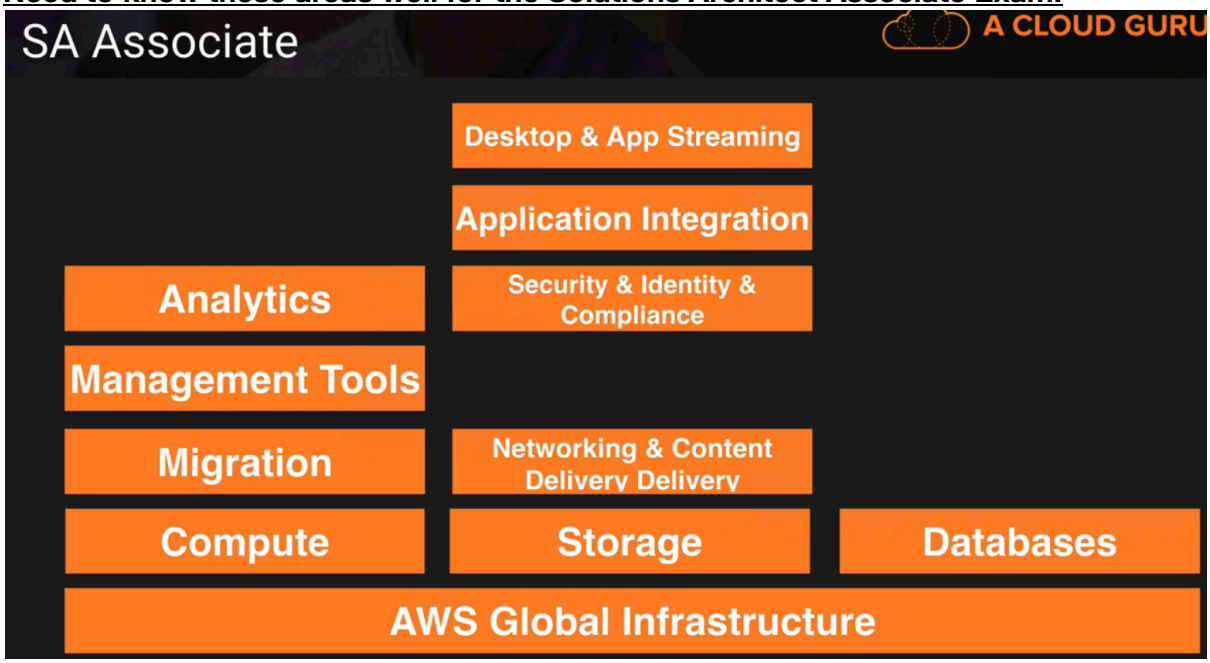


AWS Solutions Architect Associate Exam Layout & Preparation

- Approx. 60 questions in 120 minutes.
- Your results for the examination are reported as a score from 100-1000, with a minimum passing score of 720.
- AWS sample exam questions: https://d1.awsstatic.com/training-and-certification/docs-sa-assoc/AWS_Certified_Solutions%20Architect_Associate_Feb_2018_Sample%20Questions_v1.0.pdf

Domain	% of Examination
Domain 1: Design Resilient Architectures	34%
Domain 2: Define Performant Architectures	24%
Domain 3: Specify Secure Applications and Architectures	26%
Domain 4: Design Cost-Optimized Architectures	10%
Domain 5: Define Operationally Excellent Architectures	6%
TOTAL	**100%**

Need to know these areas well for the Solutions Architect Associate Exam:



** Need to know VPC's and IAM sections really well to pass all associate exams (SA, Dev, SysOps).
S3 FAQ: https://aws.amazon.com/s3/faqs/
Need to know VPC very well to pass all AWS associate exams (ass-architect, ass-dev, ass-sysops-admin) and need to understand how to build it out from scratch

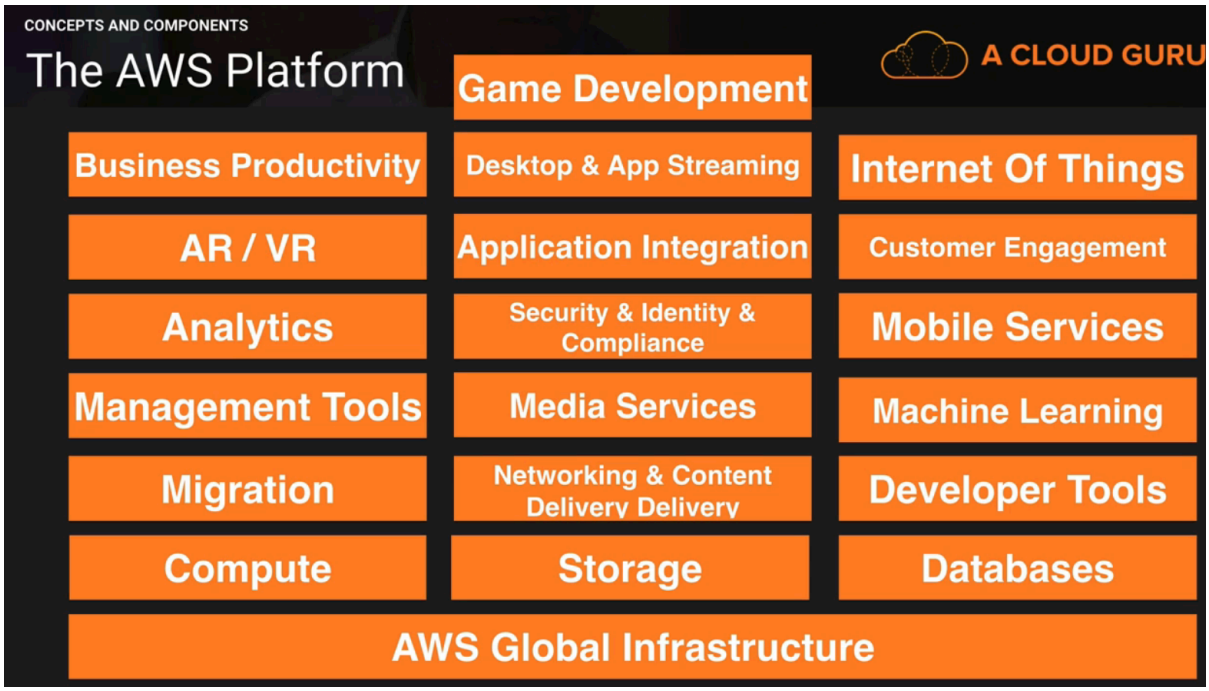
See peoples study experience:

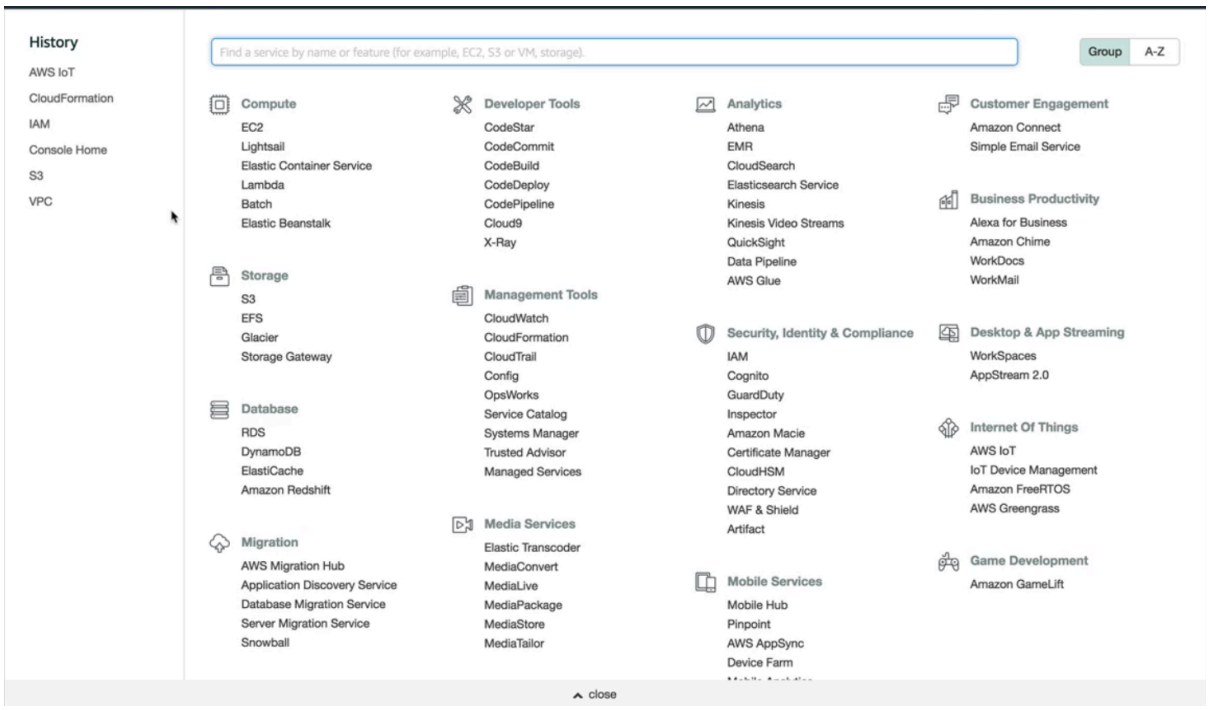
- https://medium.com/@LindaVivah/resources-i-used-to-pass-the-aws-solutions-architect-associate-certification-exam-1a9807f8d195
- https://medium.com/@annamcabee/guide-to-passing-all-3-aws-associate-level-certifications-73516bcef6e1

Considering using WhizLabs practice exams: https://www.whizlabs.com/aws-solutions-architect-associate/practice-tests/?gclid=Cj0KCQjwiJncBRC1ARIsAOvG-a43G1A1i7qNlMmngehmdJETb1gVM2-supXDdqazKd03duD_mv4QKhgaAieVEALw_wcB
Considering using Udemy Certified Solutions Architect Associate Practice Exams: https://www.udemy.com/aws-certified-solutions-architect-associate-amazon-practice-exams/?couponCode=ASSOCIATE

AWS Cheat Sheet: https://tutorialsdojo.com/aws-cheat-sheets/

10,000 Foot View of AWS





AWS Global Infrastructure

- 16+ Regions, 44+ Availability Zones and 96+ Edge Locations (CDN)

Compute

Storage

- S3 (simple storage service) - buckets
- EFS (elastic file system) - network file system attached to virtual machine
- Glacier - Data archiving & backup
- Snowball - Physically send hard drive to you, send back to AWS with data
- Storage Gateway - Virtual machines setup in your office or data centre to replicate to AWS

Databases

- RDS - relational DB
- DynamoDB - non relational DB
- Elasticache - cache
- Redshift - data warehousing / business intelligence

Migration

- AWS Migration Hub - Migration tracking service
- Application Discovery Service - Automatically detects & tracks running applications and their dependencies
- Database Migration Service - helps on-premises to AWS database migration
- Server Migration Service - helps migrate on-premesis servers to AWS

Networking & Content Delivery

- VPC - networking
- Cloudfront - CDN
- Route53 - DNS
- API Gateway - API's
- Direct Connect - Dedicated network from corporate office / data-centre into AWS

Developer Tools

Management Tools

- CloudWatch - Monitoring service
- Cloud Formation - Infrastructure as code
- CloudTrail - Logs changes to AWS environment (via APIs calls)
- Config - Monitors configuration of AWS environment over time
- OpsWorks - Chef & Puppet to automate configuration management
- Service Catalogue - Catalogue of IT services
- Systems Manager - Manage virtual servers and group resources (including Parameter Store)

- Trusted Advisor - Advice across security, cost, usage (save money). Also now across performance, fault tolerance & service limits.
- Managed Services - Offers managed services (not having to worry about AWS)

Media Services

- Elastic Transcoder - Resizes media (ie video)
- MediaConvert - File based video transcoding for multiscreen delivery
- MediaLive - Live video processing service
- MediaPackage - Prepare and protects video over internet
- MediaStore - Stores media
- MediaTailor - Targeted advertising of video

Machine Learning

- SageMaker - Easy deep learning
- Comprehend - Sentiment analysis
- DeepLens - Locally artificially aware camera (physical camera)
- Lex - Powers Alexa, voice interaction
- Machine Learning - Machine learning service
- Poly - Text to speech
- Rekognition - Upload image file, will provide meta-data about image
- Amazon Translate - Amazon translation service
- Amazon Transcribe - Automatic speech recognition to text

Analytics

- Athena - Run SQL queries against SQL bucket
- EMR - Process big data solutions
- CloudSearch - Search service
- ElasticSearch Service - Search service
- Kinesis - Ingesting large amounts of data into AWS
- Kinesis Video Streams - Ingest video into AWS
- QuickSight - Business intelligence tool
- Data Pipeline - Move data between different AWS services
- Glue - ETL

Security & Identity & Compliance

- IAM - Access management
- Cognito - Device authentication & customer sign-up, sign-in, access-control via IDAM / user pools
- Guard Duty - Monitors for malicious activity
- Inspector - Agent installed on virtual machines, analyses and reports on

vulnerabilities
- Macie - Scans S3 buckets for PII / PCI
- Certificate Manager - Manage SSL certificates
- CloudHSM - Hardware security module, dedicated for key storage
- DirectoryService - Microsoft Active Directory integration with AWS services
- WAF - Layer 7 Web application firewall
- Shield - DDOS mitigation
- Artefact - Audit & compliance portal for realtime AWS compliance reports

Mobile Services

- Mobile Hub - Management console for mobile apps to connect/setup to AWS services (use AWS mobile SDK)
- Pinpoint - Targeted push notifications (ie based on location etc)
- AWS AppSync - Realtime data-driven API (graphQL) for online/offline applications
- Device Farm - Mobile application testing on real devices
- Mobile Analytics - Mobile analytics service

Augmented Reality / Virtual Reality

- Sumerian - Augmented & virtual reality

Application Integration

- Step Functions - Managing state
- Amazon MQ - Message queues (managed activeMQ)
- SNS - Simple notification service
- SQS - Simple queue service
- SWF - Simple workflow service (can have humans in between too)

Customer Engagement

- Connect - Contact centre as a service
- SES - Simple email service

Business Productivity

- Alexa For Business - Does a whole bunch of business services like book a meeting room, log ticket for broken printer etc
- Chime - Video conferencing
- Work Docs - Store work files
- WorkMail - Work email

Desktop & App Streaming

- Workspaces - Virtual Desktop Infrastructure (VDI) solution
- AppStream 2.0 - Streaming applications to your device (ie like Citrix)

Internet Of Things

- iOT - Remote devices sending & receiving information
- iOT Device Management - Manage iOT devices at scale
- Amazon FreeRTOS - OS for micro-controllers
- Greengrass - Software which allows you to run local compute, caching, sync & machine learning on remote devices

Game Development

- Gamelift - Helps develop games and VR

AWS Premium Support

- https://aws.amazon.com/premiumsupport/faqs/
- https://aws.amazon.com/premiumsupport/pricing/
- Four plans:
    - Basic (default, unpaid)
    - Developer Plan
    - Business Plan
    - Enterprise Plan

	Developer Plan (Business hours*)	Business Plan (24x7)	Enterprise Plan (24x7)
General guidance	24 hours or less	24 hours or less	24 hours or less
System impaired	12 hours or less	12 hours or less	12 hours or less
Production system impaired		4 hours or less	4 hours or less
Production system down		1 hour or less	1 hour or less
Business-critical system down			15 minutes or less

Trusted Advisor

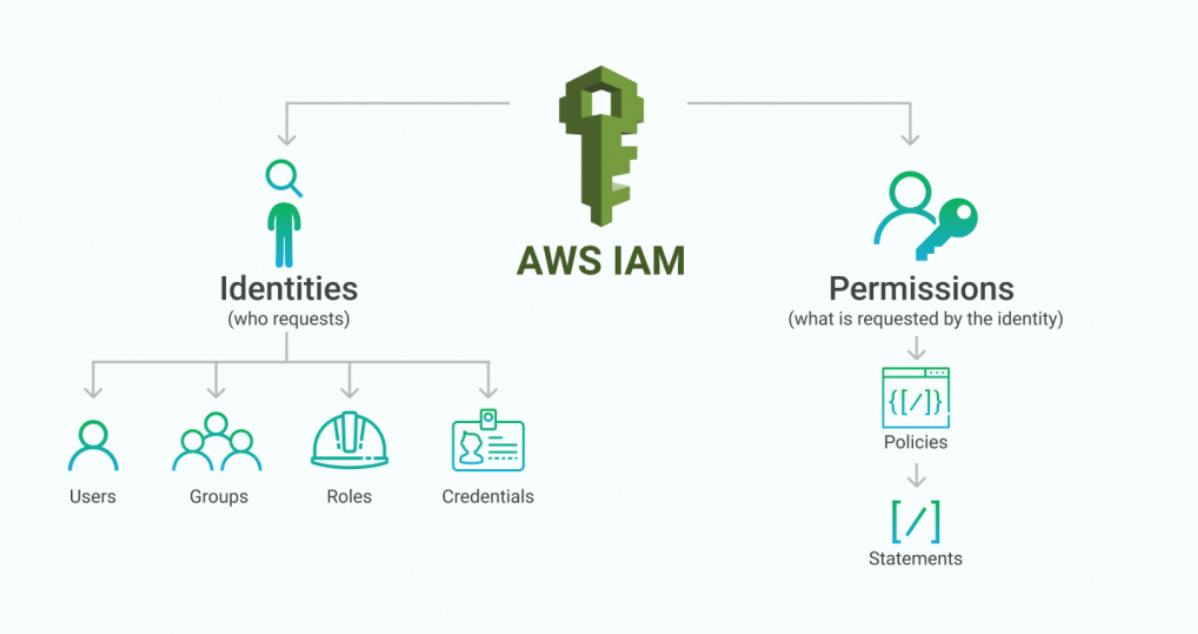
- AWS users have access to the data for seven checks. Users with Business or Enterprise-level Support can access all checks

---

Identity & Access Management (IAM)

Manages AWS users & access
Granular permissions
Identity Federation (ie use ActiveDirectory, LDAP etc)
Provides temporary access
Integrates with many AWS services
IAM is universal across AWS
New users get no permissions to start with
There is only one root account

User: end users or people
Groups: collections of users
Roles: create & assign roles to resources (i.e identify trusts & grant permissions)
Policies: document which defines one (or more) permissions. Can attach Policies to Users, Groups & Roles.

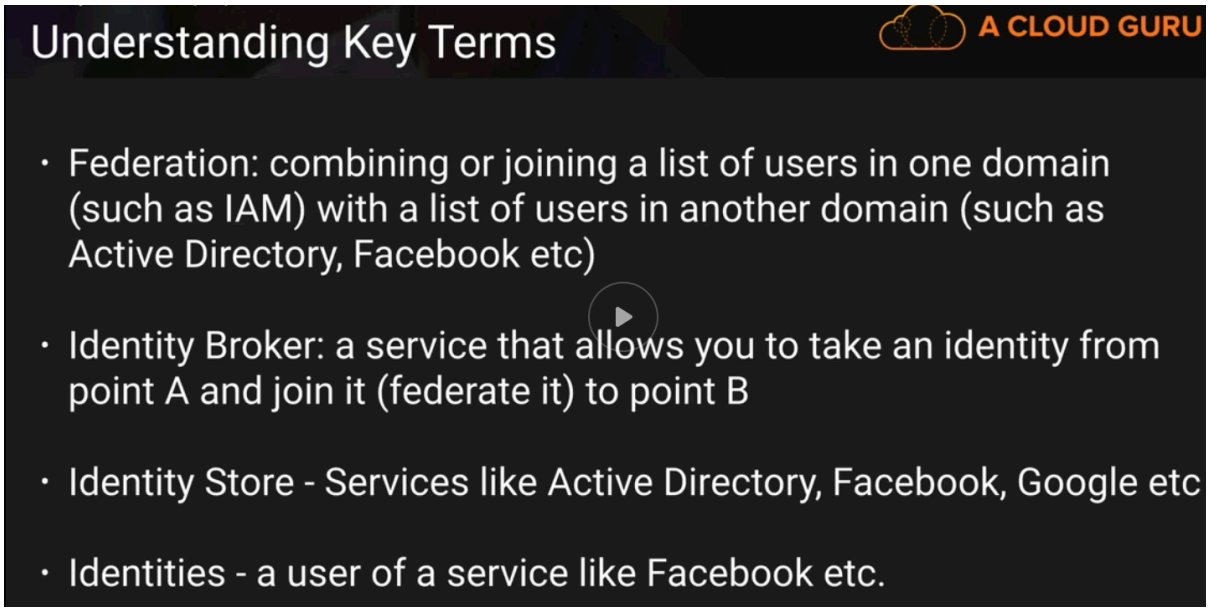


You assume an IAM role by calling the AWS Security Token Service (STS) AssumeRole APIs (in other words, AssumeRole, AssumeRoleWithWebIdentity, and AssumeRoleWithSAML). These APIs return a set of temporary security credentials that applications can then use to sign requests to AWS service APIs. There is no limit to the number of IAM roles you can assume, but you can only act as one IAM role when making requests to AWS services. An IAM role does not have any credentials and cannot make direct requests to AWS services. IAM roles are meant to be assumed by authorized entities, such as IAM users, applications, or an AWS service such as EC2.

Security Token Service (STS)

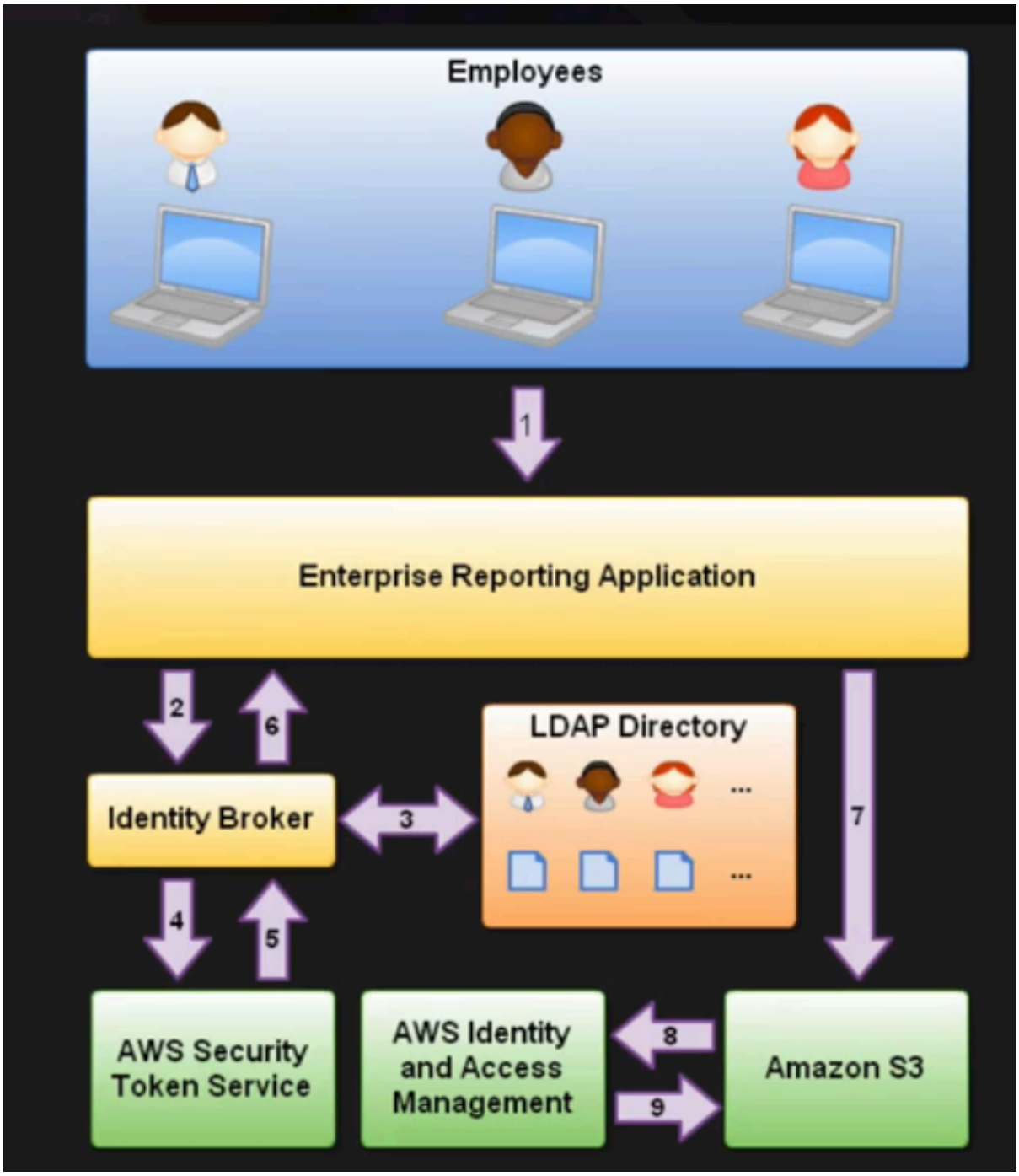
Grants users limited and temporary access to AWS resources who can come from:

- Federation (ie Active Directory) via SAML and can use Single Sign On (SSO) - user doesn't need to be a user in IAM
- Federation with Mobile Apps via OpenID (ie Facebook, Amazon, Google etc)
- Cross Account Access



Typical Scenario - User signs in to local system and wants access to their own Amazon S3 bucket
Can use:

- User Credentials (ie Active Directory) or
- AWS IAM Role


Employees
1
Enterprise Reporting Application
2
6
LDAP Directory
Identity Broker
3
7
4
5
AWS Security Token Service
AWS Identity and Access Management
8
9
Amazon S3

1. Employee enters their username and password
2. The application calls an Identity Broker. The broker captures the username and password
3. The Identity Broker uses the organization's LDAP directory to validate the employee's identity
4. The Identity Broker calls the new GetFederationToken function using IAM credentials. The call must include an IAM policy and a duration (1 to 36 hours), along with a policy that specifies the permissions to be granted to the temporary security credentials
5. The Security Token Service confirms that the policy of the IAM user making the call to GetFederationToken gives permission to create new tokens and then returns four values to the application: An access key, a secret access key, a token, and a duration (the token's lifetime)

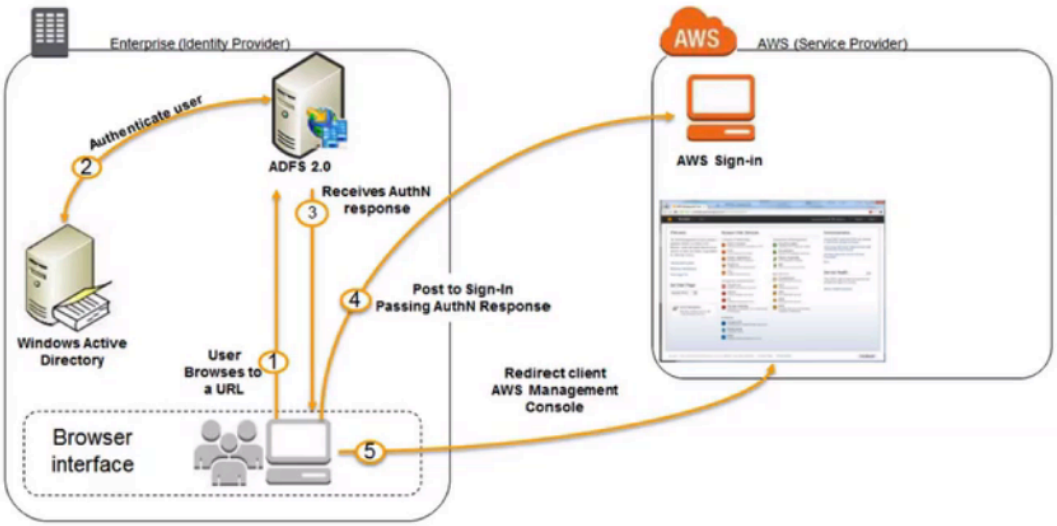
6. The Identity Broker returns the temporary security credentials to the reporting application.
7. The data storage application uses the temporary security credentials (including the token) to make requests to Amazon S3.
8. Amazon S3 uses IAM to verify that the credentials allow the requested operation on the given S3 bucket and key
9. IAM provides S3 with the go-ahead to perform the requested operation.

Active Directory Integration

Active Directory Service: https://docs.aws.amazon.com/directoryservice/latest/admin-guide/what_is.html

- **AWS Directory Service** for Microsoft Active Directory (Standard Edition or Enterprise Edition) if you need an actual Microsoft Active Directory
- **Simple AD** if you need a low-scale, low-cost directory with basic Active Directory compatibility that supports Samba 4–compatible applications, or you need LDAP compatibility for LDAP-aware applications.
- **Amazon Cloud Directory** if you need a cloud-scale directory to share and control access to hierarchical data between your applications.
- **Amazon Cognito** if you develop high-scale SaaS applications and need a scalable directory to manage and authenticate your subscribers and that works with social media identities.
- **AD Connector** if you only need to allow your on-premises users to log in to AWS applications and services with their Active Directory credentials

Integration uses SAML
Authenticate to ActiveDirectory first, granted a SAML token, passed to AWS Sign-In



1. The flow is initiated when a user (let's call him Bob) browses to the ADFS sample site (https://*Fully.Qualified.Domain.Name.Here*/adfs/ls/IdpInitiatedSignOn.aspx) inside his domain. When you install ADFS, you get a new virtual directory named adfs for your default website, which includes this page
2. The sign-on page authenticates Bob against AD. Depending on the browser Bob is using, he might be prompted for his AD username and password.
3. Bob's browser receives a SAML assertion in the form of an authentication response from ADFS.
4. Bob's browser posts the SAML assertion to the AWS sign-in endpoint for SAML (https://signin.aws.amazon.com/saml). Behind the scenes, sign-in uses the AssumeRoleWithSAML API to request temporary security credentials and then constructs a sign-in URL for the AWS Management Console.
5. Bob's browser receives the sign-in URL and is redirected to the console.

From Bob's perspective, the process happens transparently. He starts at an internal web site and ends up at the AWS Management Console, without ever having to supply any AWS credentials.

S3 101

Simple Storage Service
S3 FAQ: https://aws.amazon.com/s3/faqs/
Global service, although a bucket is stored in a single Region across at least 3 AZ's
Object based storage for files which are made up of:

- Key (name of file)
- Data (sequence of bytes of file)
- Version ID
- Metadata
- Subresources such as:
    - ACL's

    - Torrent

Files can be from 0 bytes to 5TB
Largest object that can be uploaded by a single put is 5GB.
For Objects larger than 100MB, consider using Multi-part Upload
Unlimited storage
Files stored in bucket (a folder in the cloud)

- Each bucket is unique in AWS, as s3 is a universal namespace
- Addresses are in this path format: https://s3-eu-west-1.amazonaws.com/TheUniqueBucketName
    - Can use a virtual-hosted-style name (if DNS setup)
      i.e.: http://TheUniqueBucketName.s3-eu-west-1.amazonaws.com
- HTTP 200 reply on successful upload

Data Consistency model:

- Read after Write consistency for PUTS of new Objects
- Eventual Consistency for overwrite PUTS and DELETES (takes time)

S3 Availability SLA is 99.9% (built for 99.99%)
S3 Durability SLA is 99.999999999% (for all Tiered Storage types)
S3 Performance - Provides increased performance to support at least 3,500 requests per second to add data and 5,500 requests per second to retrieve data.

S3 Tiered Storage

- S3 Standard (99.9% availability, designed for 99.99%)
- S3 Intelligent Tiering (unknown / changing data access patterns - moves files after 30 days no access)
- S3 IA (99% availability, designed for 99.9%, infrequently accessed data, can access immediately on demand. Is cheaper, ideal for backups, DR files etc.). **Cheaper to store, more expensive for access.**
- S3 One Zone IA (99% availability (designed for 99.95%, same as S3 IA including durability but less available, resilient & lost if Zone destructs . 20% cheaper)
    - Reduced Redundancy (not recommended)
- All tier options provide the same performance (latency & throughput) & durability, however different availability levels / resilience levels

Can store objects within a bucket at different tier levels
You can directly PUT into S3 by:

- Uploading via UI
- Specifying the x-amz-storage-class header: <<<tiered_option>>>

See difference between Availability and Durability: https://blog.westerndigital.com/data-availability-vs-durability/

Amazon Glacier tiered storage for data archiving:

- Expedited (Few minutes, expensive)
- Standard (3-5 hour restore, cheaper)
- Bulk (5-12 hour restore, cheapest)

S3 Secure using ACL's and Bucket policies
S3 Charge model:

- Storage
- Requests
- Storage Management pricing (meta-data)
- Data Transfer pricing (cross Region)
- Transfer Acceleration (S3 CDN enabled)

S3 Versioning:

- Stores all versions of the file (even if you delete, can restore by deleting delete-marker)
- Once enabled, can only be suspended (not removed)
- Integrates with lifecycle rules
- MFA delete (need MFA to delete for extra security)

S3 Cross Region Replication

- Need replication turned on in both buckets
- Region must be unique
- Files in existing bucket are not replicated automatically (you have to use CLI)
- Cannot replicate to multiple buckets or use daisy chaining
- Delete markers are replicated (however deleting individual versions or delete makers are not replicated)

S3 Notifications (events)

- Receive notifications when certain events happen in your bucket (Create, Update, Delete, Copy etc) and send to SQS, SNS, Lambda

S3 Lifecycle Management

- Create lifecycle rules which can
    - Can transition to other storage classes (S3 IA, or archive to Glacier)
    - Works with versions of Objects
    - Can expire (delete) after time

S3 Inventory

- S3 bucket audit report to help manage your files & storage on a daily or weekly basis available in CSV, ORC or Parquet format

S3 Batch Operations

- Automate the execution, management, and auditing of a specific S3 API requests or AWS Lambda functions across many objects stored in Amazon S3 at scale (update tag sets, update ACL's, copying objects between buckets, Glacier restore to S3 storage type, and execute Lambda functions etc)

S3 PUT request headers

- Cache-Control
- Content-Disposition
- Content-Encoding
- Content-Length
- Content-MD5
- Content-Type
- Expect
- Expires
- x-amz-meta-
- x-amz-storage-class
- x-amz-tagging
- x-amz-website-redirect-location
- x-amz-object-lock-mode
- x-amz-object-lock-retain-until-date
- x-amz-object-lock-legal-hold

S3 Security & Encryption

- By default all new buckets are private
- Customers may use four mechanisms for controlling access to Amazon S3 resources:
    - Identity and Access Management (IAM) policies,
    - Bucket policies,
    - Access Control Lists (ACLs),
    - Query String Authentication.
- Can setup access logging
- 4 different types of encryption:
    - In Transit via SSL/TLS
    - At Rest
        - Server side encryption (SSE)
            - S3 Managed Keys
            - AWS Key Management Service (KMS)
            - Server Side Encryption with customer provided keys

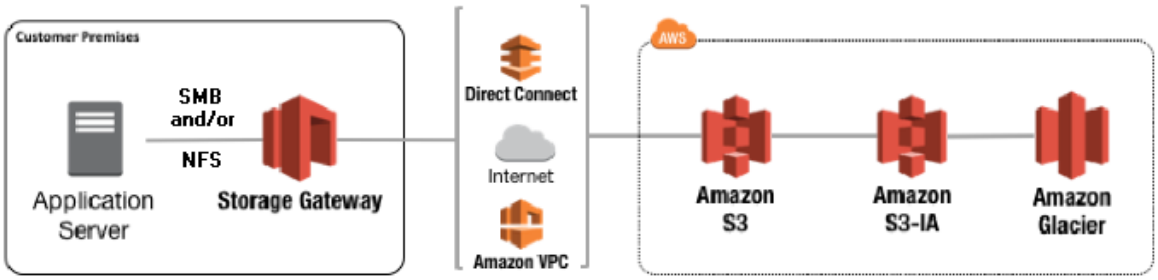
        - Client side encryption (CSE)

S3 Query In Place

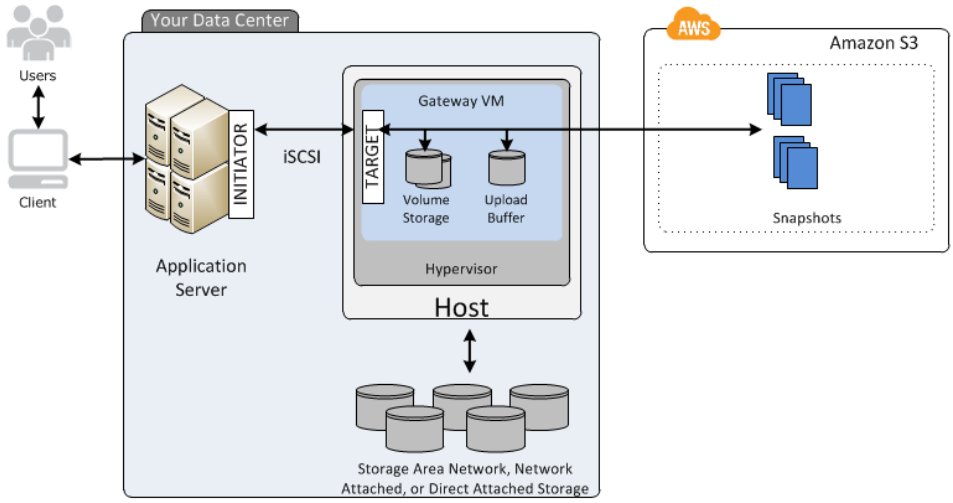
- Query your files without having to move them into an analytics platform (or other out of S3 for processing)
- S3 offers multiple query in place options:
    - S3 Select
    - Amazon Athena
    - Amazon Redshift Spectrum

Storage Gateway

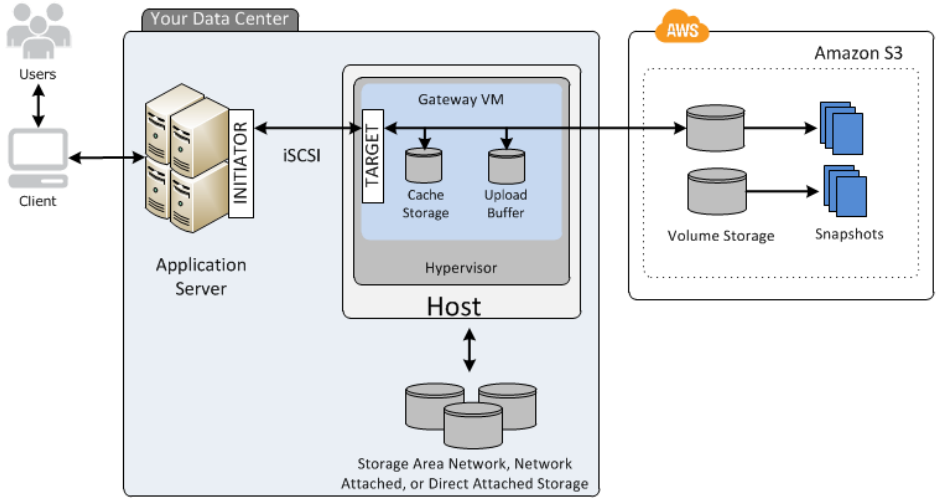
- A virtual appliance which sits within your own data centre and allows you to store data in AWS
- https://docs.aws.amazon.com/storagegateway/latest/userguide/StorageGatewayConcepts.html
- 4 types of AWS storage gateway
    - File Gateway (NFS) stores flat files in s3
    - Volumes Gateway (iSCSI) is block storage (ie a virtual hard disk to run OS, databases etc) copied asynchronously to S3:
        - Stored Volumes - All local storage, backed up to S3
        - Cached Volumes - All data in S3, local data cached in datacenter
    - Tape Gateway (VTL) to S3 via virtual tapes



Stored Volume Storage Gateway



Cached Volume Storage Gateway



Snowball

- Was called Import / Export before Snowmobile (you could send in your hard-drive, no standard format)
- Move large amounts of data into AWS using a physical portable storage device for transport (AWS standard device)
- 3 different types:
    - Snowball - storage only
    - Snowball Edge - storage and compute (AWS data centre in a box)
    - Snowmobile - Physical truck with shipping container - Petabyte / Exabyte level storage & transport
- Snowball can import to S3 and export from S3

S3 Transfer Acceleration

- Upload to CloudFront Edge Network (rather than directly to S3) via a distinct URL
- Distinct URL is in the format of https://BucketName.s3-accelerate.amazonaws.com
- Can do a speed comparison on all Edge Locations to see how much faster (or slower) per Region

S3 Static Website

- Will have the URL format of https://BucketName.s3-website-eu-west-1.amazonaws.com

---

CloudFront

- Edge Location - where content is cached (close to user)
    - Can Read and Write to them
    - Objects are cached for the life of the TTL (Time to Live)
    - Can clear cached objects (though charged $$)
- Origin - Original files the CDN will distribute, could be S3, EC2, ELB, Route53 or can cache non AWS hosted websites
- Distribution - A collection of edge locations
    - Web distribution (websites)
    - RTMP (media streaming, Flash Media protocol)
- Time to live (TTL), default is 24 hours, min TTL is 0 hours, max TTL is 365 days
- Can use Restrict Viewer Access for private (paid) viewing using Signed URLS or Signed Cookies
- Can use Geo-Restriction to Whitelist or Blacklist Countries

---

EC2 101

Elastic Compute Cloud (EC2)

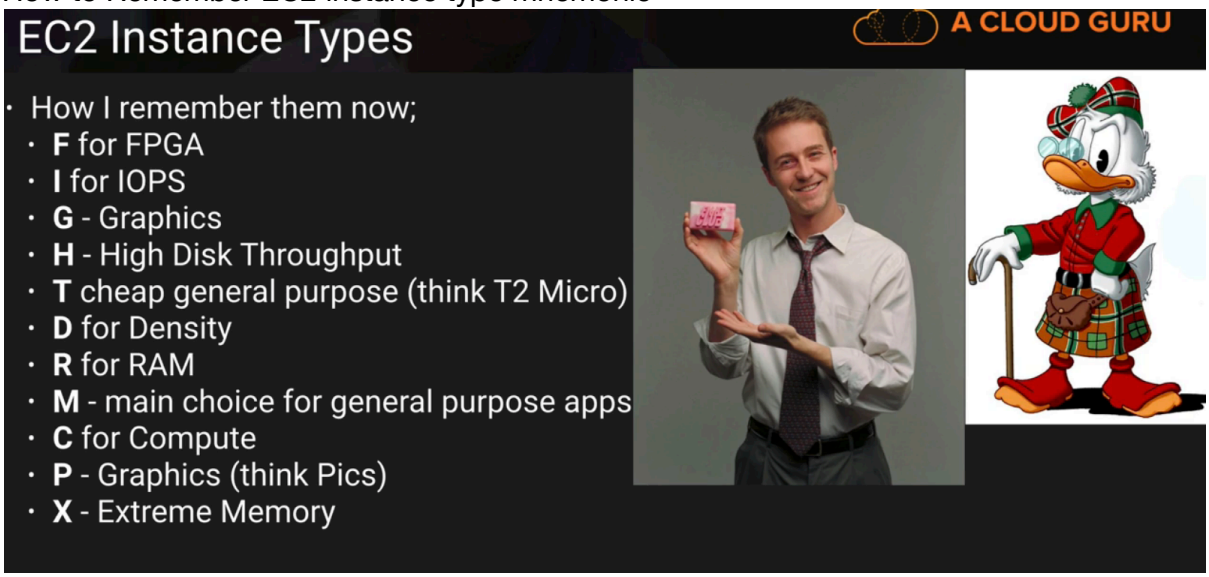
- On Demand - no commitment
- Reserved - capacity reservation (1 or 3 years)
    - Can move reserved instances across AZ's, but not across Regions
    - Can chose Standard Reserved Instance or Convertible Reserved Instances (exchangeable compute type)
    - Reserved Instance Marketplace provides an online marketplace to trade their Reserved Instances with other AWS customers
    - Types:
        - Standard Reserved instance - 1 or 3 year commitment, up to 75% off OnDemand
        - Zonal RIs provide the benefit of capacity reservation and a discount - 1 or 3 year commitment, up to 54% off OnDemanmd

    - Regional RIs automatically apply the RI's discount to instance usage across AZs and instance sizes in a region (however Regional RIs don't provide capacity reservation)
    - On-Demand Capacity Reservations allow you to reserve capacity for any duration without a commitment - similar to Reserved Instances without the discount
    - Schedule Reserved Instance - Purchase capacity reservations that recur on a daily, weekly, or monthly basis, with a specified start time and duration, for a one-year term
- Spot - bid on available compute (if terminated by AWS, you will not be charged for a partial hour. You will be charged full hour if you terminate)
    - AWS can interrupt spot instances with 2 minute notification; suitable for fault-tolerant flexible workloads. Charged by the Spot Instance hour, set every hour.
    - Can set a "maximum Spot price" and if the Spot price rises above this, your instance will be reclaimed with a two-minute notification
    - Cannot use Paid 3rd party AMIs with Spot instances (such as IBM software packages)
    - With Spot hibernate, Spot instances will pause and resume around any interruptions so your workloads can pick up from exactly where they left off (ensure enough HDD space on EBS Root Volume to store memory (RAM) on hibernation). You have no control over hibernate stop/start or hibernate/resume cycles (AWS controlled).
    - A Spot Fleet allows you to automatically request and manage multiple Spot instances
- EC2 Fleet lets you provision compute capacity across different instance types, Availability Zones and across On-Demand, Reserved Instances (RI) and Spot Instances
- Cluster Compute Instances (EC2 Cluster Placement Group) combine high compute resources with a high performance networking for High Performance Compute (HPC) applications and other demanding network-bound applications - specifically engineered to provide high performance networking
    - Inter-instance traffic within the same region can utilize 5 Gbps for single-flow and up to 100 Gbps for multi-flow traffic. When launched in a placement group, select EC2 instances can utilize up to 10 Gbps for single-flow traffic
- Amazon EC2 allows you to choose between Fixed Performance Instances (e.g. C, M and R instance families) and Burstable Performance Instances (e.g. T2). T2 instances use CPU credits (accumulated during idle) to use for burst cycles.
- Dedicated hosts - physical EC2 server
- Uses Xen and Nitro Hypervisors
- Limited to running 20 On-Demand instances, 20 Reserved Instances & dynamic # of Spot Instances within an AWS account (ask for extension to default limit from AWS support instance request form)
- Each EC2 instance type is charged differently and any data transferred within & between AWS Regions will be charged as Internet Data Transfer on both sides of transfer

EC2 Instance Types

Family	Speciality	Use case
F1	Field Programmable Gate Array	Genomics research, financial analytics, real-time video processing, big data etc
I3	High Speed Storage	NoSQL DBs, Data Warehousing etc
G3	Graphics Intensive	Video Encoding/ 3D Application Streaming
H1	High Disk Throughput	MapReduce-based workloads, distributed file systems such as HDFS and MapR-FS
T2	Lowest Cost, General Purpose	Web Servers/Small DBs
D2	Dense Storage	Fileservers/Data Warehousing/Hadoop
R4	Memory Optimized	Memory Intensive Apps/DBs
M5	General Purpose	Application Servers
C5	Compute Optimized	CPU Intensive Apps/DBs
P3	Graphics/General Purpose GPU	Machine Learning, Bit Coin Mining etc
X1	Memory Optimized	SAP HANA/Apache Spark etc

How to Remember EC2 instance type mnemonic



FIGHT DR MC PX

Elastic Block Storage (EBS)

- A virtual disk, can only be mounted to one EC2 instance (use EFS for shared)
- General Purpose SSD (GP2) use for < 10,000 IOPS
- Provisioned IOPS SSD (IO1) use for > 10,000 IOPS

- Throughput Optimised HDD (ST1)
    - Magnetic drive
    - Cant be a boot volume
- Cold HDD (SC1) use for lowest cost
    - Magnetic drive
    - Cant be a boot volume
- Magnetic (Standard)
    - Lower cost, only bootable magnetic drive
- EBS must be in same AZ as EC2 Instance
- Can modify Volumes (size, type) on the fly (without downtime), however not on older magnetic drives
- Can copy a Volume to another Region by creating a Snapshot (Snapshots are stored on S3 - incremental on update)
    - Shutdown EC2 instance before taking Snapshot of Root Volume
    - If taking Snapshot of RAID Array (with application / OS cache):
        - Stop all applications
        - Flush all caches to the disk
        - Unmount RAID array
        - Shutdown EC2 instances
        - Take Volume Snapshot
- Can copy a Snapshot & Image (AMI) to a different Region
- Can create an Image (AMI) from a Snapshot
- You can share Snapshots (but only if they're unencrypted).

- Can have two Root Volume device types:
    - EBS
        - Can stop & reboot
    - Instance Store (Ephemeral Storage) - cannot STOP EC2 instance
        - Can reboot
- Both Root Volumes will be deleted in EC2 termination (however with EBS you can select not to delete)
- You can attach multiple volumes to a single instance; **you can not attach multiple instances to one volume.**

SSD-backed volumes are **designed for transactional** IOPS-intensive database workloads, boot volumes, and workloads that require high IOPS. SSD-backed volumes include Provisioned IOPS SSD (io1) and General Purpose SSD (gp2).
HDD-backed volumes are **designed for throughput-intensive and big-data workloads**, large I/O sizes, and sequential I/O patterns. HDD-backed volumes include Throughput Optimized HDD (st1) and Cold HDD (sc1).

Amazon Machine Image (AMI)
- https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/virtualization_types.html

- Virtual machines of many operating systems & configurations
- HVM (hardware virtual machine) - Great performance, provides the ability to run an operating system directly on top of a virtual machine without any modification, as if

it were run on the bare-metal hardware
- PV (paravirtual machine) virtual machine types - Uses a special boot loader called PV-GRUB, but they cannot take advantage of special hardware extensions such as enhanced networking or GPU processing
- AMI's are Region based (how you can copy an image to a different Region)

EC2 Management

- Termination Protection is turned off by default
- When EC2 instance is terminated, default action is to delete root EBS Volume (other Volumes won't be deleted, only root Volume)
- Cannot encrypt EBS Root Volumes of default AMIs (need to copy the AMI and encrypt your own)
- Additional Volumes can be encrypted

Security Groups

- A virtual firewall for EC2
- Security Groups operate at a VPC level & do not span VPC's (or Regions)
- All Inbound traffic is blocked by default - you can only Allow traffic (not Deny traffic)
- All Outbound traffic is allowed by default
- Any rule you make to Security Groups are applied immediately
- EC2 instances can share Security Groups and an EC2 instance can have multiple Security Groups attached
- Rules are stateful (so for an Inbound rule on port 80 any incoming connection HTTP connection is allowed to respond, even if you don't have corresponding Outbound rule)
- You cannot blocked specific IP addresses using Security Groups (use NACL's for this)

Elastic Load Balancers

- Balance load of traffic across many servers
- Three different types of LBs:
    - Application Load Balancer (Layer 7) - most likely to be used - HTTP, HTTPS protocols
    - Network Load Balancer (Layer 4) - very high network performance required - TCP, TLS protocols
    - Classic Load Balancer (Known as ELB, Layer 4 & some Layer 7 such as X-Forwarded-For header) - old generation, TCP, SSL/TLS protocols, HTTP, HTTPS protocols
- A HTTP 504 error (Gateway timeout) means the application (ie EC2 Webserver) is not responding
- Instances monitored by ELBs are either InService or OutofService
- ELBs dont have pre-defined IP addresses, must use DNS name

- For ALBs - at least two AZ's / public subnets must be specified on ALB creation (and only one subnet per AZ). ELB & NLB do not have this requirement.

Feature	Application Load Balancer	Network Load Balancer	Classic Load Balancer
Protocols	HTTP, HTTPS	TCP, TLS	TCP, SSL/TLS, HTTP, HTTPS

Application Load Balancer - can associate multiple certificates for the same domain to a secure listener. Supports:

- ECDSA and RSA certificates
- Certificates with different key sizes (e.g. 2K and 4K) for SSL/TLS certificates
- Single-Domain, Multi-Domain (SAN) and Wildcard certificates

Network Load Balancer only supports RSA certificates with 2K key size. Do not support RSA certificate key sizes greater than 2K or ECDSA certificates

Elastic IPs

- A reserved, static public IPv4 which can be assigned to an EC2 instance, and automagically re-assigned to another EC2 instance in case of failure
- Elastic IPs form a core part of the dynamic cloud and allow you to mask and recover from failure
- Limited to 5 EIP per Region (can ask AWS support for extension)
- https://docs.aws.amazon.com/vpc/latest/userguide/vpc-eips.html
- You can have one Elastic IP (EIP) address associated with a running instance at no charge. If you associate additional EIPs with that instance, you will be charged for each additional EIP associated with that instance per hour on a pro rata basis. To ensure efficient use of Elastic IP addresses, we impose a small hourly charge when these IP addresses are not associated with a running instance or when they are associated with a stopped instance or unattached network interface. There is no charge for Elastic IP addresses you create from an IP address prefix you brought into AWS using Bring Your Own IP.

CloudWatch

- CloudWatch Default Monitoring is every 5 minutes
    - Can change to Detailed Monitoring every 1 minute (additional charge)
    - Can store metrics from 3 hours @1m resolution to 15 months@1hr resolution (depending on resolution/data-point type)
    - Stores metrics for up to 2 weeks after monitoring disabled or instances terminated
- Can create custom dashboards
- Free tier has basic resource monitoring, captured every 5 minutes (or 1 minute for

detailed monitoring at $$ cost) such as:
    - CPU based
    - Disk based
    - Network based
    - StatusCheck based
    - Need to create a Custom Metric for RAM, EBS data etc
- Alarms - to alert on key events / thresholds
- Events - State changes in resources and setup corresponding rules/actions
- Logs - Capture application and OS logs, aggregate and store. Can setup agent on OS to send logs to CloudWatch service
- Metrics - Enables you to view metrics in realtime (rather than use Dashboards)

IAM Roles with EC2

- Use Roles to connect to other AWS services rather than storing AWS access keys (access key ID & secret access key) on EC2 for AWS CLI - works automagically with correct IAM Roles - no keys required.
- Role Type (ie AWS Service Role)
- Establish Trust (ie EC2)
- Attach Policy (ie permissions)
- Roles are global (not Region/AZ based)

EC2 SSH
->$ ssh ec2-user@xxx-ec2-instance-ip -i ~/Documents/AmazonKeys/xxx-you-ssh-key-xxx.pem

EC2 Bash Scripts

- Can run scripts as root when EC2 instance starts to help configure instance. As an example:

==========
#!/bin/bash
yum update -y
yum install httpd -y
service httpd start
chkconfig httpd on
aws s3 cp s3://mywebsitebucket-acloudguru/index.html /var/www/html
==========

EC2 Metadata

- Can retrieve meta-data about your EC2 instance from http://169.254.169.254/latest/meta-data/
- Can retrieve user-data about your EC2 instance from http://169.254.169.254/latest/user-data/
    - User data is made up of what you provide when you start the instance

- https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-metadata.html

==========
->$ curl http://169.254.169.254/latest/meta-data/
->$ curl http://169.254.169.254/latest/meta-data/reservation-id
==========

EC2 Launch Configurations & Auto Scaling Groups

- https://docs.aws.amazon.com/autoscaling/ec2/userguide/LaunchConfiguration.html
- https://docs.aws.amazon.com/autoscaling/ec2/userguide/scaling_plan.html
- You can only specify one launch configuration for an Auto Scaling group at a time, and you can't modify a launch configuration after you've created it.
- If you want to change the launch configuration for your Auto Scaling group, create a launch configuration and then update your Auto Scaling group with the new launch configuration.
    - New instances are launched using the new configuration parameters, but existing instances are not affected.
- Can setup ASG lifecycle hooks to put instance into a wait state (so you can perform custom activities) - default wait period is 1 hour
- Amazon supports the following auto-scaling policies:
    - Manual Scaling - Manually attach / detach instances to your ASG
    - Scheduled Scaling - A schedule allows you to set your own scaling schedule for predictable load changes
    - Dynamic scaling:
        - Target Tracking Scaling Policies - Select a scaling metric and set a target value
        - Simple and Step Scaling Policies - Choose scaling metrics and threshold values for the CloudWatch alarms that trigger the scaling process
        - SQS based scaling - In response to changing demand from an Amazon Simple Queue Service (Amazon SQS) queue
- Instance Warm up: Specify the number of seconds that it takes for a newly launched instance to warm up. Until its specified warm-up time has expired, an instance is not counted toward the aggregated metrics of the Auto Scaling group.
- Cooldown periods: Prevent the initiation of additional scaling activities before the effects of previous activities are visible. Does not support cooldown periods for step scaling policies.

Script to setup Apache and create custom index.html from AWS meta-data with instance details on EC2 instance startup
==========

```
#!/bin/bash
yum update -y
yum install httpd -y
service httpd start
chkconfig httpd on

echo '<html><head><title>Hello World!</title></head><body><h1>Hello World!</h1>
<table>' >> /var/www/html/index.html
echo '<tr><td>InstanceID</td><td>' >> /var/www/html/index.html
curl http://169.254.169.254/latest/meta-data/instance-id >> /var/www/html/index.html
echo '</td></tr><tr><td>Region & AZ</td><td>' >> /var/www/html/index.html
curl http://169.254.169.254/latest/meta-data/placement/availability-zone >>
/var/www/html/index.html
echo '</td></tr><tr><td>Public Hostname</td><td>' >> /var/www/html/index.html
curl http://169.254.169.254/latest/meta-data/public-hostname >>
/var/www/html/index.html
echo '</td></tr><tr><td>Public IP</td><td>' >> /var/www/html/index.html
curl http://169.254.169.254/latest/meta-data/public-ipv4 >> /var/www/html/index.html
echo '</td></tr></table></body></html>' >> /var/www/html/index.html

aws s3 cp s3://acloudguru-aws-associate-solutions-architect-study /var/www/html/ --
recursive
```

==========

EC2 Placement Groups

- https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html

- Placement Groups are used to keep compute & data very close to each other to reduce network time
- The placement group can only have 7 running instances per Availability Zone
- Three types of Placement Groups:
    - Clustered Placement Group - All instances within one AZ (low network latency, high network throughput. Only certain instance types allowed).
    - Spread Placement Group - All instances placed in distinct underlying hardware, can spread across multiple AZ's (Only certain instance types allowed)
    - Partition Placement Group - All instances placed across logical partitions - ensuring partitions do not share underlying hardware

Elastic File System (EFS)

- Essentially a shared, mounted file server. Storage capacity is elastic, growing & shrinking automatically and is block based storage
- Supports Network File System (NFS)
- Data is stored across multiple AZ's within a single region
- Read after Write consistency
- Can only mount an EFS on one VPC - however can use VPC peering to share EFS to other VPCs

- Can apply OS user-level, directory level & file-level permissions

https://docs.aws.amazon.com/efs/latest/ug/performance.html#performancemodes

- Choose one of two modes when creating EFS for Performance:
    - General Purpose Performance Mode - General Purpose is ideal for latency-sensitive use cases, like web serving environments, content management systems, home directories, and general file serving
    - Max I/O Performance Mode - Can scale to higher levels of aggregate throughput and operations per second with a tradeoff of slightly higher latencies for file operations. Highly parallelized applications and workloads, such as big data analysis, media processing, and genomics analysis

https://docs.aws.amazon.com/efs/latest/ug/performance.html#throughput-modes

- Choose one of two modes when creating EFS for Throughput:
    - Bursting Throughput - Throughput on Amazon EFS scales as the size of your file system in the standard storage class grows
    - Provisioned Throughput - You can instantly provision the throughput of your file system (in MiB/s) independent of the amount of data stored.

Lambda

- Function based compute service, don't have to worry about any underlying hardware, server, OS concerns or scaling
- Supports the following 5 programming languages:
    - Node
    - Python
    - Java
    - Go
    - .NET
- Supports the following Triggers (events):
    - API Gateway
    - AWS IoT
    - Alexa
    - CloudFront
    - CloudWatch (Events & Logs)
    - CodeCommit
    - Cognito Sync Trigger
    - DynamoDB
    - Kinesis
    - S3
    - SNS
    - SQS
- Lambda is charged based on:
    - Number of requests:

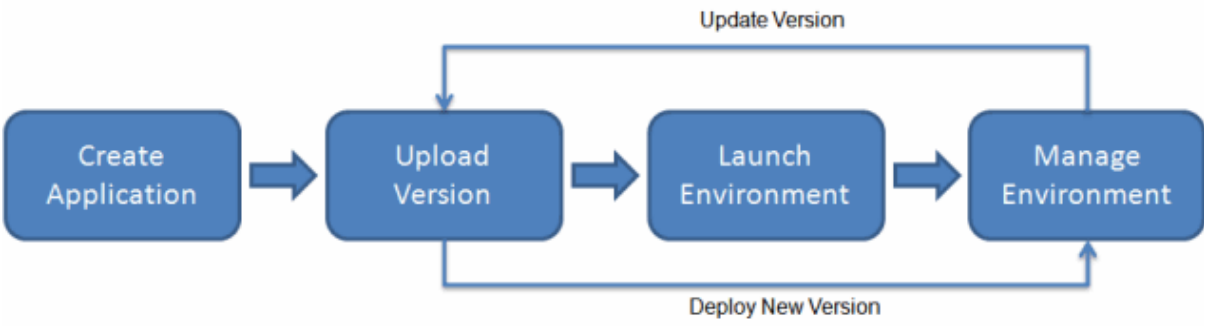
        - First 1m requests free
        - $0.20 per 1m requests thereafter
    - Duration
        - Code execution time rounded to nearest 100ms, dependent on how much memory you allocate to your function
        - $0.00001667 for every GB-second
    - Function cannot execute for longer than 5 minutes

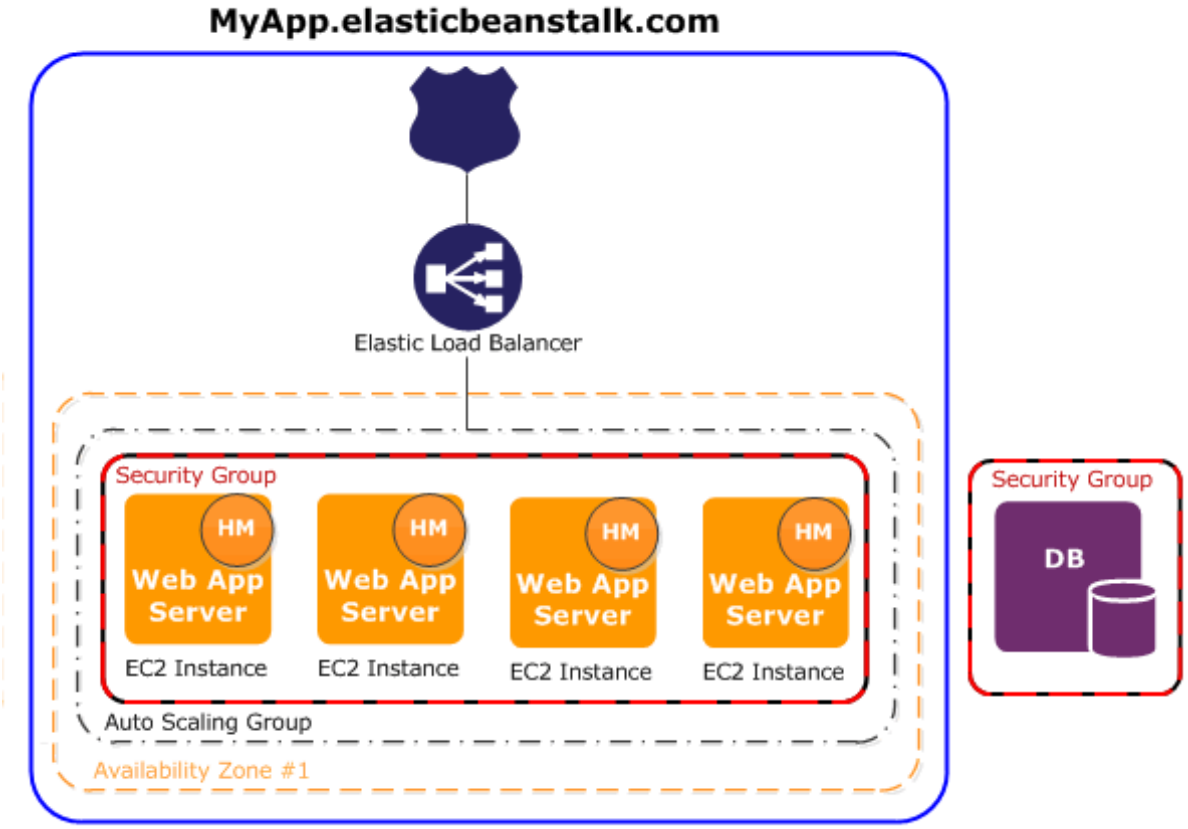
EC2 Connection Timeout debugging

- https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/TroubleshootingInstancesConnecting.html#TroubleshootingInstancesConnectionTimeout
    - If getting EC2 connection timeouts check:
        - Security Group rules
        - Route Table Subnet config
        - Network ACLs Subnet Config
        - Check local network firewall rules (inbound & outbound) for blocking port 22 / 3389
        - Check your instance has public Ip4 address (or Elastic IP)
        - Check EC2 CPU load (could be overloaded)
-

AWS Beanstalk

- Elastic Beanstalk - quickly deploy and manage applications in the AWS Cloud without having to learn about the infrastructure that runs those applications
- Elastic Beanstalk supports applications developed in Go, Java, .NET, Node.js, PHP, Python, and Ruby
- Elastic Beanstalk supports Docker
- To use Elastic Beanstalk, you create an application, upload an application version in the form of an application source bundle (for example, a Java .war file) to Elastic Beanstalk, and then provide some information about the application.
- Elastic Beanstalk automatically launches an environment and creates and configures the AWS resources needed to run your code
- Great for web server and worker workloads
- Elastic Beanstalk uses **nginx** as the reverse proxy to map your application to your Elastic Load Balancing load balancer on port 80. Elastic Beanstalk provides a default nginx configuration that you can either extend or override completely with your own configuration.





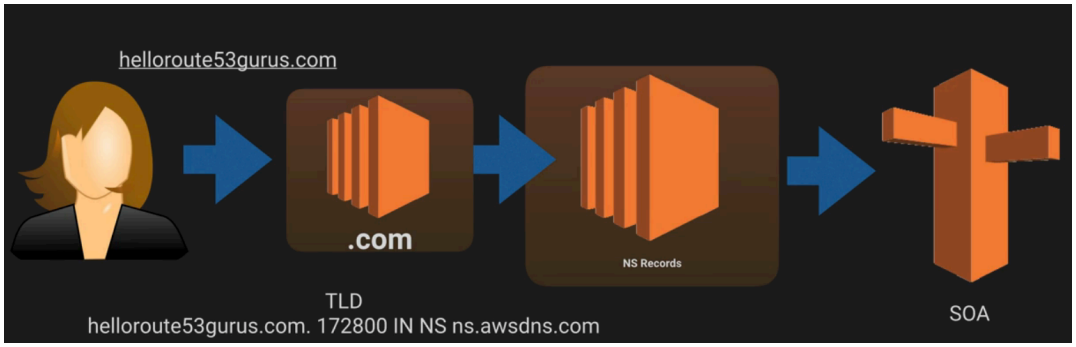
DNS 101

- Convert human friendly domain names to Internet Protocol (IP) address (ie google.com to 216.58.196.142)
- IP address can be either IPv4 (32 bit) and IPv6 (128 bit)
- Top level domains are .com, .gov, .net etc, second level domains are the second word such as .com.au (.com), .co.uk (.co)
- Can view all top level domain names at IANA: http://www.iana.org/domains/root/db
- Root Zone database & Domain Registers - control registration of domain names

- DNS servers are responsible for managing/owning select DNZ zones (ie your domain name)
- Start Of Authority record (SOA) - contains zone data of server, administrator, version, time to live (TTL)
    - Each zone contains only a single SOA record
- Name Server record (NS) - Used by top level domain servers to direct to DNS server which contains authoritative records (details of dns record)
- A Record - Domain name to translate to IP address (i.e. example.com to 135.14.32.14)
- Within AWS, always chose Alias record over CNAME if given the choice
- Can manage up to 50 Domains with Route53 (before requesting an increase from AWS support)

Alias vs CNAME records

- https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/resource-record-sets-choosing-alias-non-alias.html
- Canonical Name (CNAME) - Resolves one domain name to another (i.e. blah.test.com to boo.test.com), **can't be used** for naked domain name.
- Alias Record (AWS Route53 custom dns extension) - Maps resource record to a dns name, similar to CNAME and can be used for naked domain names (also known as Zone Apex). Amazon can change Alias records on the fly (if say an ELB IP changes) and update Route53 immediately without any changes to the hosted zone



Route53 Routing Policies

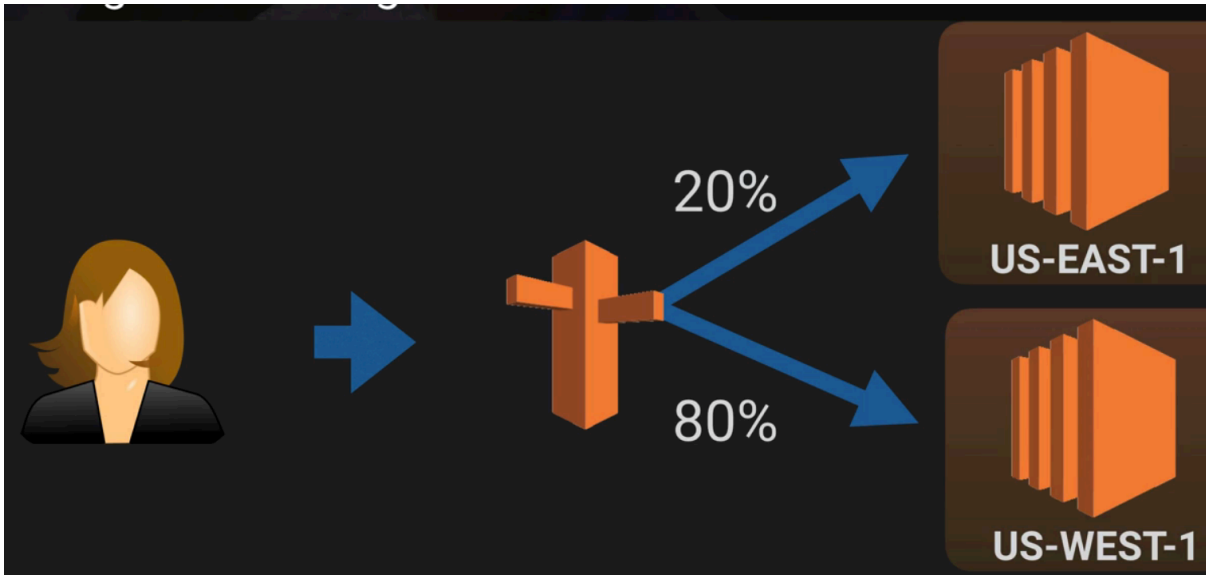
- **https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-policy.html**
- Simple Routing - Single resource
- Weighted Routing
- Latency-based Routing
- Failover Routing - Active/passive failover
- Geolocation Routing
- Geoproximity Routing
- Multivalue Answer Routing - Up to 8 healthy records chosen at random

Simple Routing

- Default routing policy
- Single A record but can specify a single record (with multiple IPs)
    - Route53 provides random server resolution to dns resolvers (if specified multiple IP addresses)

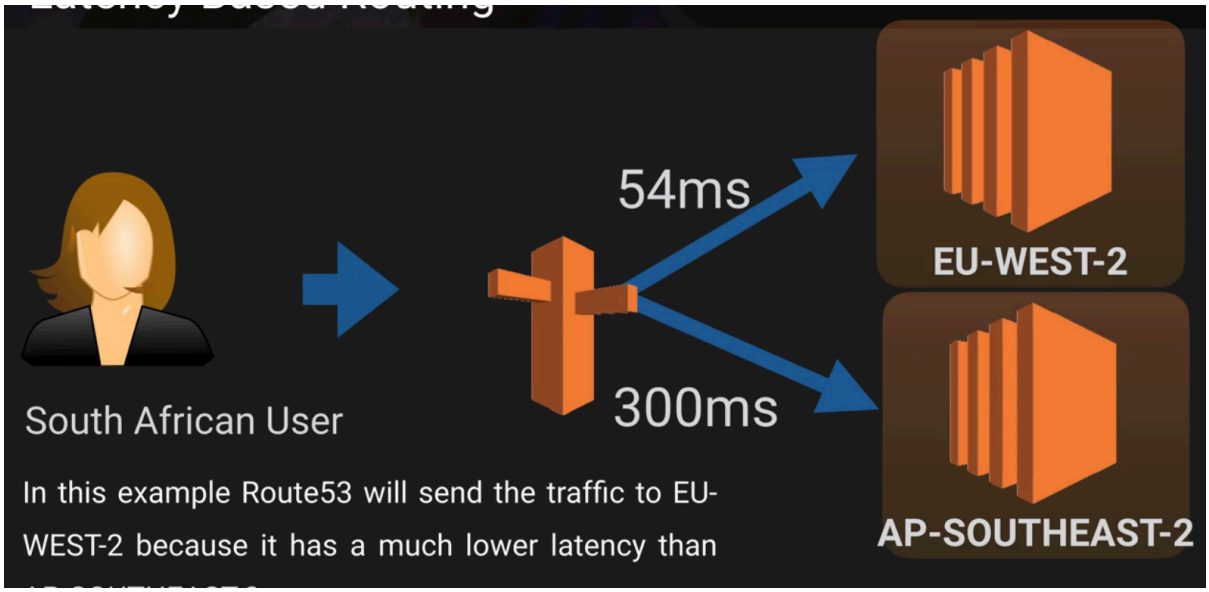
Weighted Routing

- Specify weight of traffic to select AZ's
- Multiple A Record specifying weight and servers



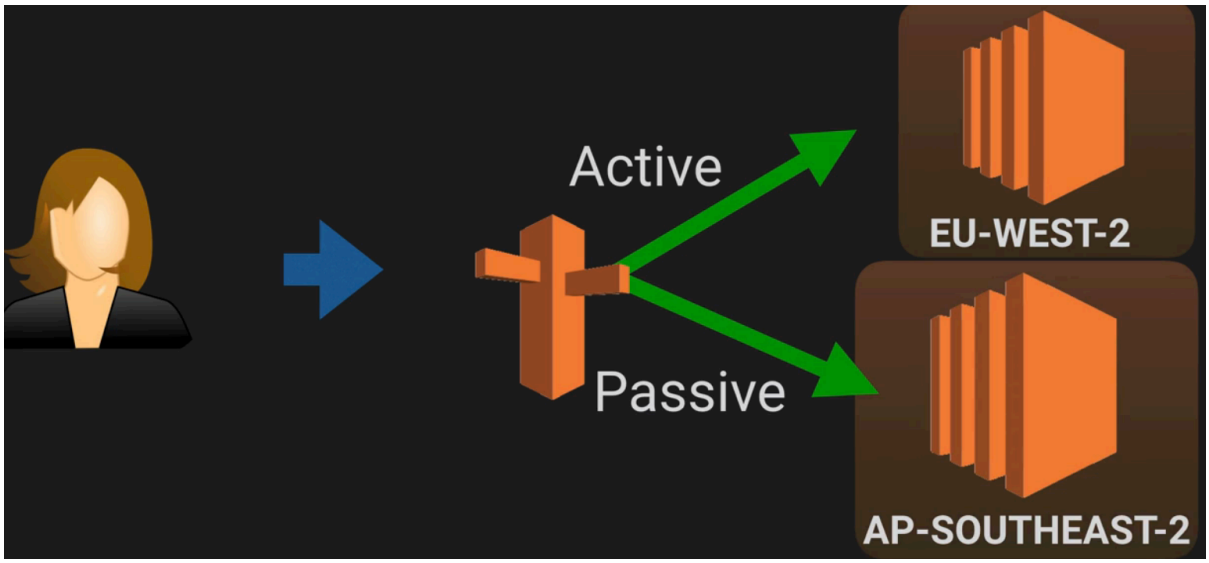
Latency-based Routing

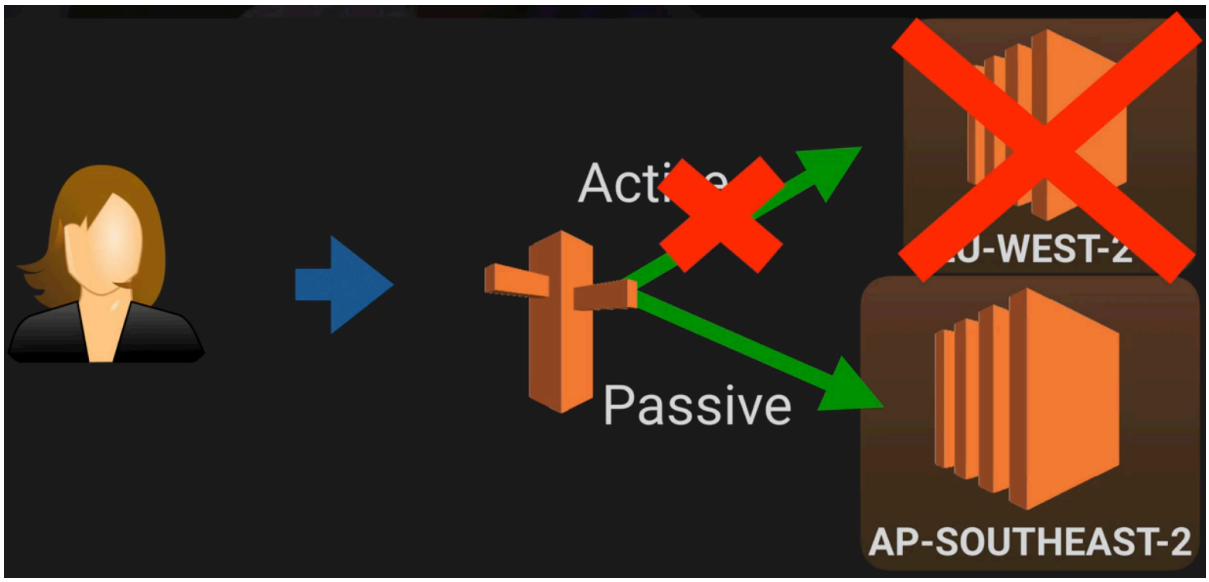
- Route traffic based on lowest network latency for your end user
- Multiple A Record specifying Regions and one or many servers in each
- Can use VPN to test different Regions and latency



Failover Routing

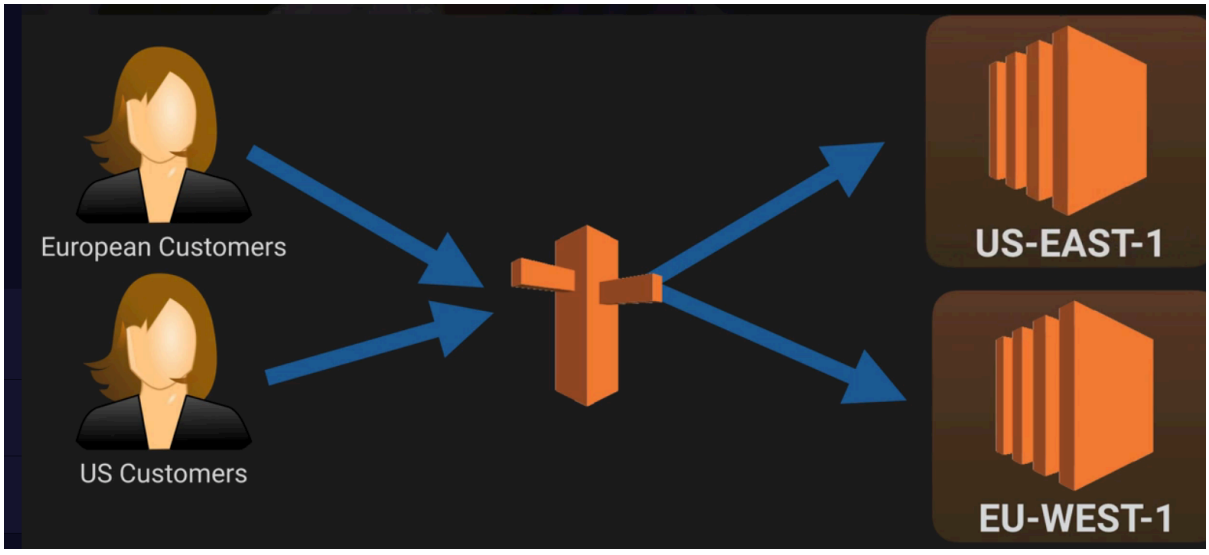
- Create an active / passive setup, good for DR with primary & secondary locations
- Route53 will use a health check monitoring to determine if failover required
- Create Health Check in Route53 which can monitor:
    - Endpoint
    - Status of other health checks
    - State of CloudWatch alarm





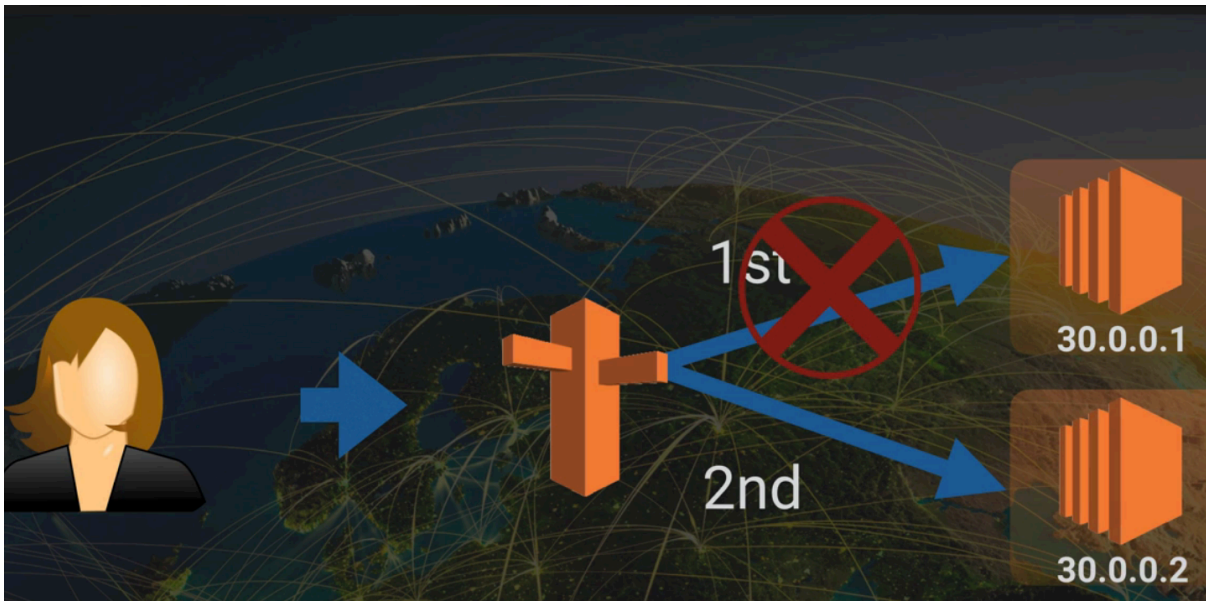
Geolocation Routing & Geoproximity Routing

- Allows you to route traffic based on geographic location of your users
- Geoproximity routing allows you to route traffic from based on location of resources and shift traffic from one resource to another in different locations
- Can have US customers routed to US-EAST-1 & European customers routed to EU-WEST-1
- Can use VPN to test different geolocations



Multivalue Answer Routing

- Route traffic (approx.) randomly across web servers using health checks
    - Route53 provides different server resolutions to dns resolvers
- Allows multiple A record each with their own IP addresses



Databases 101

Relational database

- Consist of Database, Tables, Rows, Fields (columns)
- https://aws.amazon.com/rds/
- AWS RDS DB's:
    - Microsoft SQL Server
    - Oracle
    - MySQL
    - PostGreSQL
    - Aurora
    - MariaDB
- With RDS, you pay for On-demand or Reserved DB instances (no direct DB licensing fees)
- On creation of your DB, you're always given a DNS endpoint (not an IP)
- Need to add your EC2 Webserver instance Security Group to the RDS instance Security Group for access

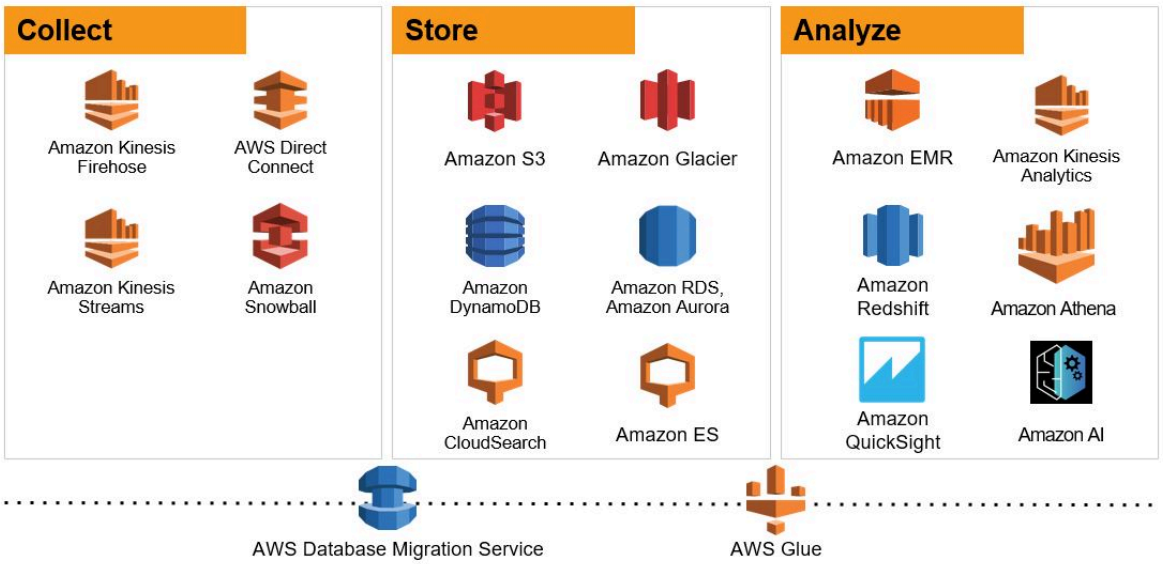
Non-Relational database (NoSQL)

- Consist of Database, Collection (i.e Table like), Document (i.e Row like), Key-value pairs (i.e Field like)
- https://aws.amazon.com/nosql/
- AWS NoSQL DB:
    - DynamoDB

- Other NoSQL DB's not directly on AWS platform (ie MongoDB, Cassandra, Couchbase etc)

Data Warehousing

- Used for business intelligence, very large data-sets
    - Tools like Cognos, Jaspersoft, SQL Server Reporting Services, Oracle Hyperion, SAP NetWeaver etc
- Data-lakes are used to store all data (structured & unstructured)
- https://aws.amazon.com/data-warehouse/
- AWS Data Warehousing:
    - Redshift
- Online Transaction Processing (OLTP - direct querying) vs. Online Analytics Processing (OLAP - complex combination, summation & analytic querying). Data warehousing is designed for OLAP



Elasticache

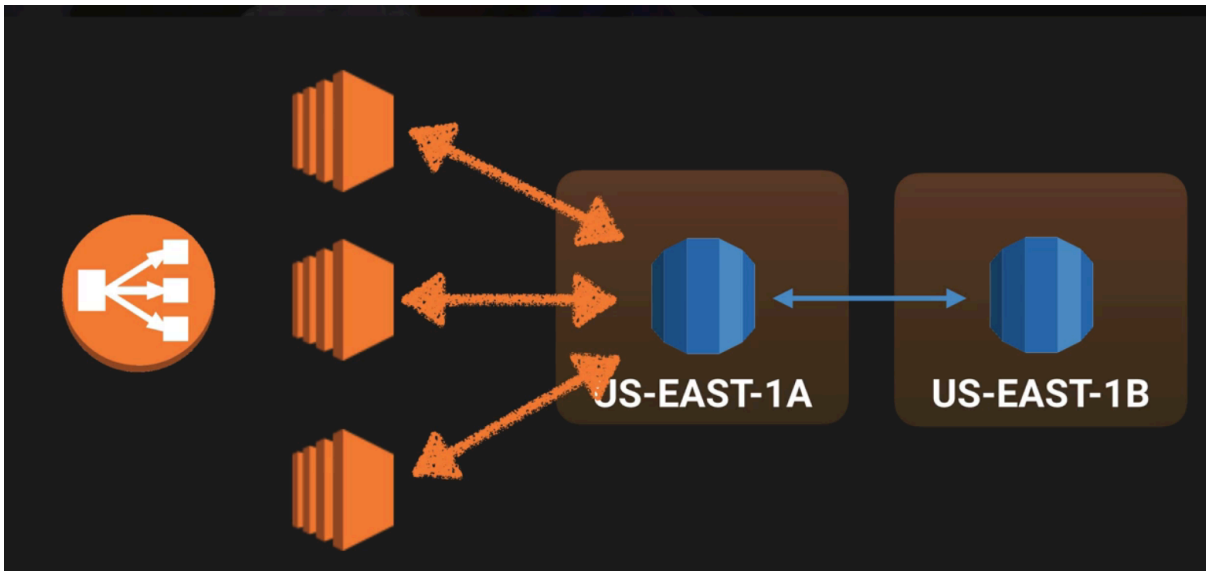
- A webservice that makes it easy to run and consume an in-memory cache (improves latency and throughput).
- Ideal for read-heavy applications
- https://aws.amazon.com/elasticache/
- AWS ElastiCache's:
    - Memcached
        - No Multi AZ capability
    - Redis
        - Multi AZ capability

RDS Backups

- Two types of backups:
    - Automated Backups
        - Full daily snapshots and store transaction logs throughout the day, recover your DB to any point in time within the retention period (1-35 days)
        - RDS Automated Backups are enabled by default, occur during a defined window and your storage I/O maybe suspended (and experience elevated latency)
        - RDS backups are stored in S3 and you get free storage up to the size of your DB.
        - RDS Automated Backups are deleted when you delete your RDS instance
    - Database Snapshots
        - Are done manually by the user
        - They are stored even after you delete the original RDS instance
- Restoring Backups
    - Whenever you restore a DB (either Automated Backup or Snapshot), the restored DB will be a new RDS instance and have new DNS endpoint
- Encryption
    - All RDS databases support encryption at rest.
    - Encryption is done using AWS Key Management Service (KMS)
    - Once your RDS instance is encrypted, all data stored, its backups, read replicas and snapshots are encrypted
    - You can encrypt an existing DB, have to create a snapshot, make a copy and encrypt the copy

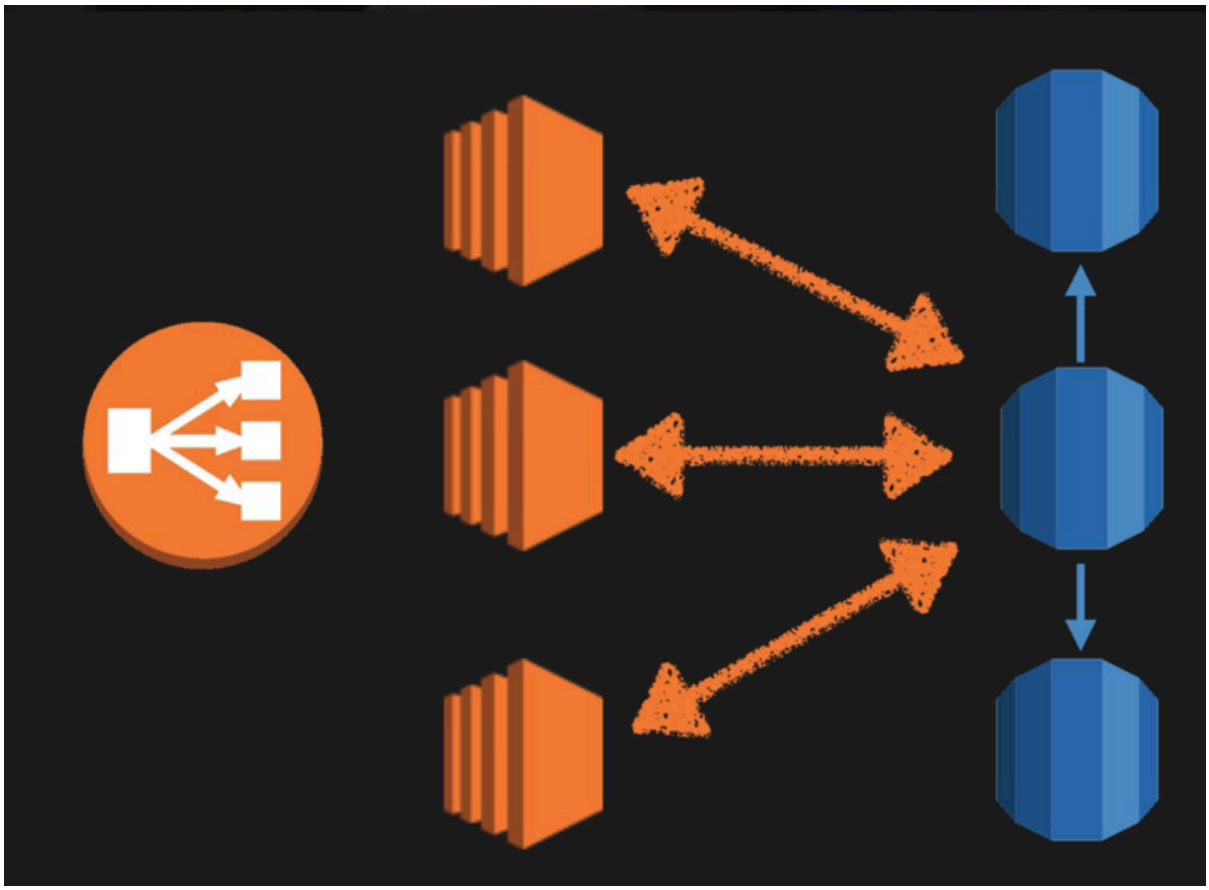
Multi-AZ RDS

- For disaster recovery only, an RDS instance can automatically fail-over to secondary AZ using Multi-AZ RDS
- Good for DR, DB maintenance etc but it doesn't have any performance benefit as EC2 instances always point to primary RDS (or secondary on fail).
- Data written to US-East-1A is Synchronously replicated to US-East1B (an exact copy)



RDS Read Replicas

- For improved performance only (not used for DR), especially in high read applications you can use RDS Read Replica. Must have Automatic Backup turned on and can have up to 5 Read Replica copies of a DB
- EC2 instances write to a primary RDS instance, and then these writes are distributed to other replica RDS instances to read from (read-only)
- Great way to scale your DB tier. You can have Read Replicas of Read Replicas in different AZs or Region (will increase latency)
- Achieved using Asynchronous replication from primary RDS instance to Read Replica instance.
- You can have Multi-AZ and Read Replicates turned on the same DB
- Read Replicas can be promoted to their own DB, however this breaks the replication
- Can create a Route53 weighted routing policy for all read replica DNS URLS



DynamoDB

- Fast & flexible fully managed No-SQL DB, document and key-value data models.
- Stored on of SSD
- Spread across 3 geographically distinct data-centres (multi-region)
- Read settings:
    - Eventual Consistent Reads (best performance) or
    - Strongly Consistent Reads
- Provides ACID transactions
- Provides Point in time recovery (up to 35 days)
- The combined Value and Name combined fields must not exceed 400 KB.
- Pricing:
    - Write Throughput $0.0065 per hour for 10 units
    - Read Throughput $0.0065 per hour for 50 units
    - Storage cost of $0.25GB per month
- Can reserve capacity for DynamoDB (1 or 3 years)
- Streams https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Streams.html
    - A DynamoDB stream is an ordered flow of information about changes to items in an Amazon DynamoDB table. When you enable a stream on a table, DynamoDB captures information about every modification to data items in the table and stores this information in a log for up to 24 hours

RedShift

- Fast & flexible fully managed data warehouse for OLAP
- Redshift Configuration
    - Single Node (160GB) or
    - Multi-Node
        - Leader Node (manages client connection / receives queries)
        - Compute Node (data storage, computation & executes queries)
- Columnar Data Storage - organises data by columns for improved query & computation performance, better compression
- Pricing:
    - Storage: $0.25 per hour for up to petabyte of data and then $1000 per terabyte per year
    - Compute Node Hours: 1 unit per node per hour
    - Backup Storage
    - Data transfer (within VPC, not outside it)
- Encryption
    - Encrypted in transit
    - Encrypted at rest by Redshift, however can manage yourself via:
        - Hardware Security Modules (HSM) or
        - KMS
- Only available in 1 AZ (low availability), can restore via snapshots

Aurora

- AWS RDS relational DB, only runs on AWS infrastructure
- Compatible with MySQL and 5x better performance than MySQL
- Storage auto-scales, starting with 10GB and increments by 10GB up to 64TB
- Compute resources can scale up to 32vCPU's and 244GB of memory
- 2 copies of your data is contained within each AZ (minimum of 3 AZ's - so you have 6 copies in total. Can lose up-to 2 copies and still maintain write, and 3 copies and still maintain read)
- 2 types of Replicas:
    - Aurora Replicas (up to 15)
    - MySQL Read Replicas (up to 5)

Virtual Private Cloud (VPC)

https://aws.amazon.com/quickstart/architecture/vpc/

- See AWS VPC quick start reference architecture: https://docs.aws.amazon.com/quickstart/latest/vpc/welcome.html

- See AWS VPC quick start reference Cloud Formation: https://fwd.aws/px53q

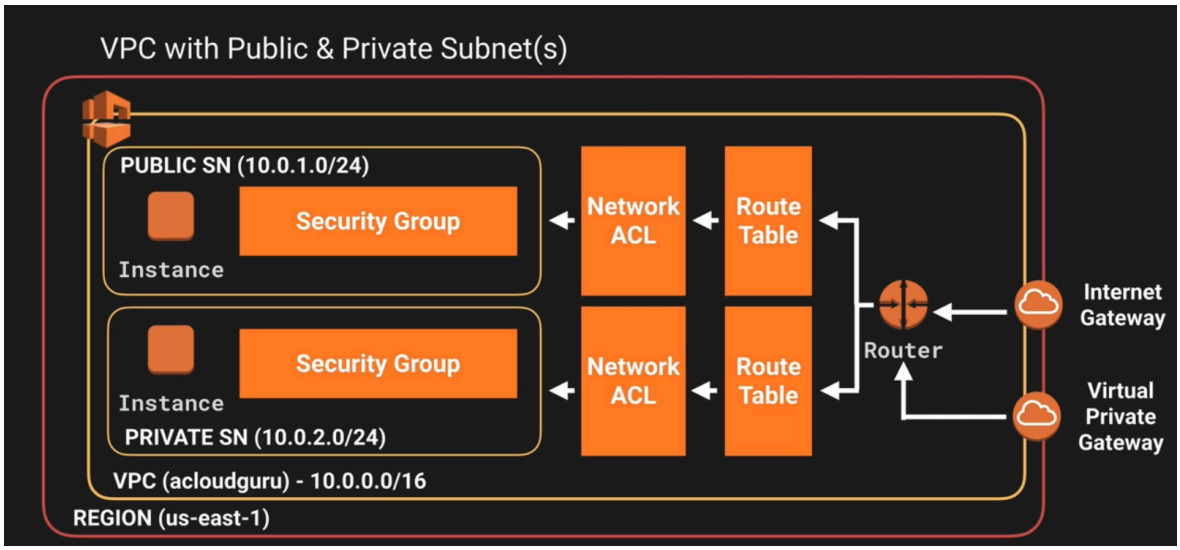
VPC is like a virtual data centre in the cloud
Every Region in the world has a default VPC

Amazon Virtual Private Cloud (Amazon VPC) lets you provision a logically isolated section of the Amazon Web Services (AWS) Cloud where you can launch AWS resources in a virtual network that you define. You have complete control over your virtual networking environment, including selection of your own IP address range, creation of subnets, and configuration of route tables and network gateways.

You can easily customize the network configuration for your Amazon Virtual Private Cloud. For example, you can create a public-facing subnet for your webservers that has access to the Internet, and place your backend systems such as databases or application servers in a private-facing subnet with no Internet access. You can leverage multiple layers of security, including security groups and network access control lists, to help control access to Amazon EC2 instances in each subnet.

Additionally, you can create a Hardware Virtual Private Network (VPN) connection between your corporate datacenter and your VPC and leverage the AWS cloud as an extension of your corporate datacenter.

VPC Diagram



AWS CIDR allowed address range
See www.cidr.xyz for CIDR range calculator

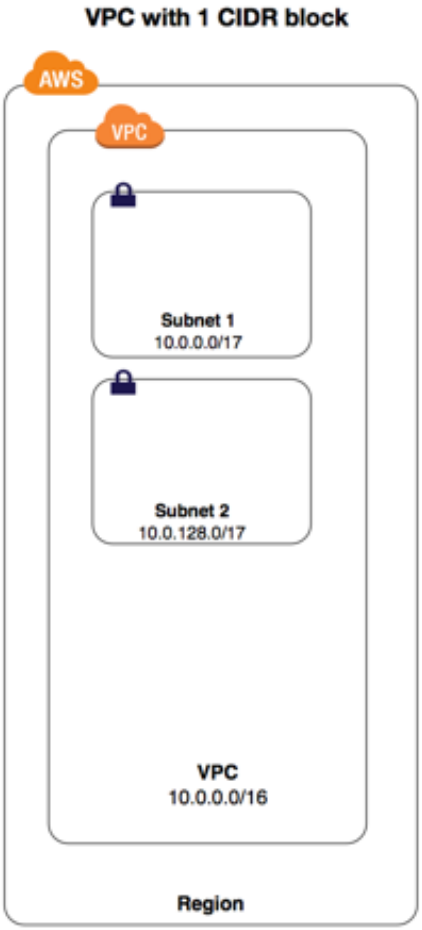
A VPC is located within a single Region
Each AZ has its own subnet
Within a VPC you can:

- Launch instances into a subnet
- Assign custom IP address ranges within a subnet
- Configure Route Tables between subnets
- Create an Internet Gateway and attached to your VPC (there can only be one Internet Gateway per VPC. IG are highly available)
- Use Network Access Control Lists (NACL's)
- Setup Instance Security Groups (Security Groups can span AZ's)

Amazon sets up a Default VPC per Region

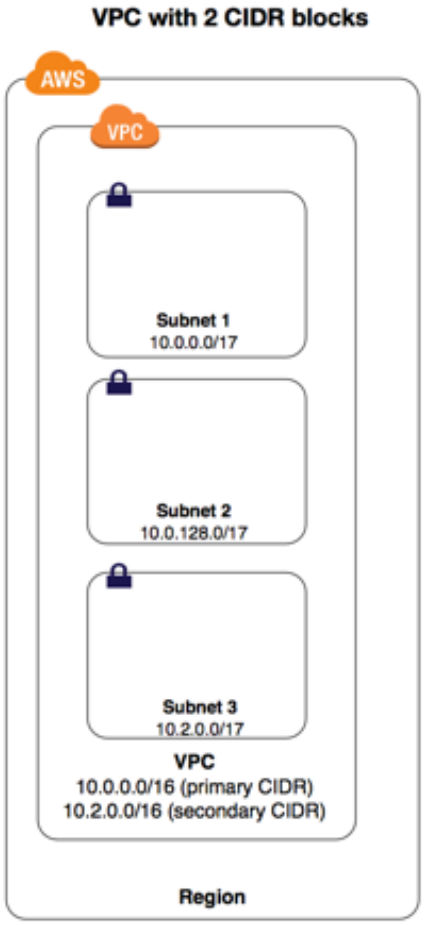
- All Subnets in default VPC have a route to the internet (and have a public & private IP)

VPC allowed block size is between a /16 netmask (65,536 IP addresses) and /28 netmask (16 IP addresses).
You can add up to four (4) secondary CIDR blocks after creation of the VPC



Main route table

Destination	Target
10.0.0.0/16	local

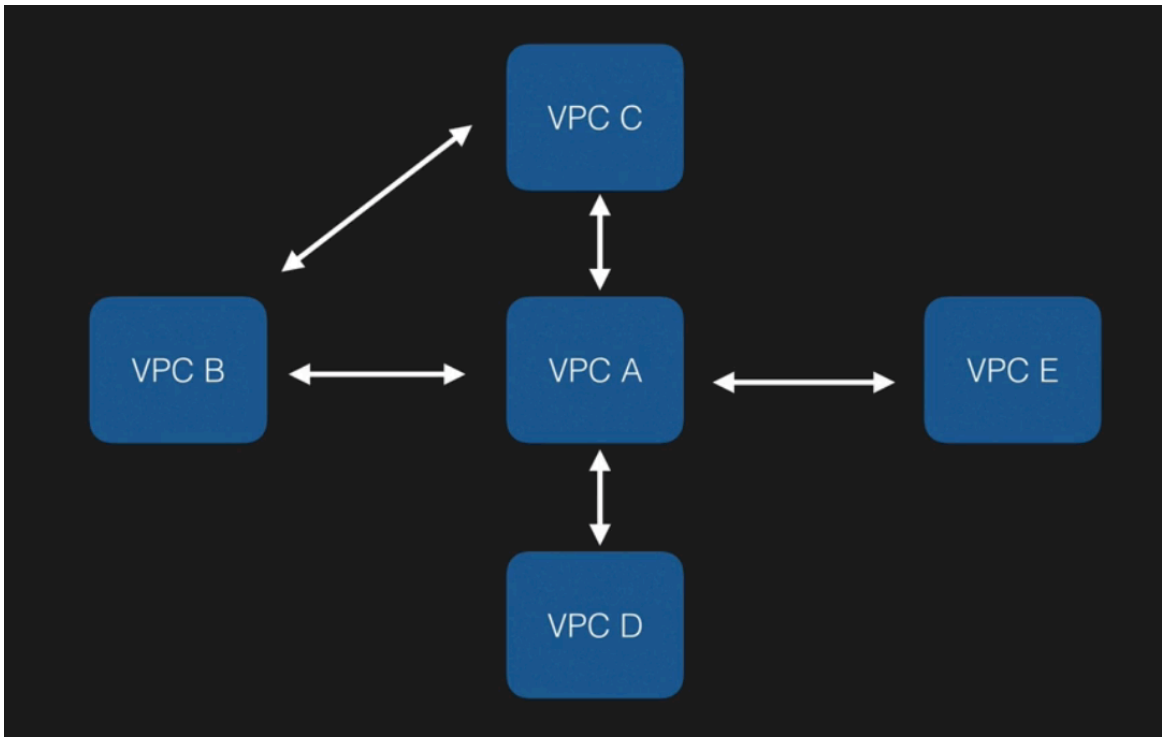


Main route table

Destination	Target
10.0.0.0/16	local
10.2.0.0/16	local

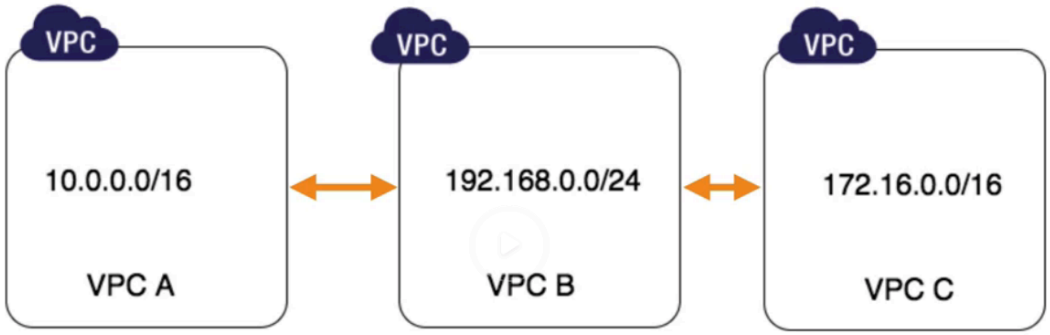
https://docs.aws.amazon.com/vpc/latest/peering/vpc-peering-basics.html
Can have multiple VPC's (up to 5 per account, can extend via AWS support request)

- VPC Peering allows you to connect one VPC with another using private IP addresses
- VPC Peering also allows you to connect to other AWS accounts
- There is no transitive peering, you must connect each VPC to each other to communicate - i.e:



Connect two VPCs and route traffic between them using private IP addresses and Instances appear as if they're in the same network
Can connect two VPCs within an AWS account or cross AWS accounts **within a single region**.

VPC Peering need to use different CIDR ranges when connected (not allowed matching or overlapping ranges)
Transitive Peering is not supported (so below, VPC A cannot communicate to VPC C)



Tenancy

- When creating a VPC, you can chose:
  - Default (shared hardware)
  - Dedicated (on dedicated hardware, much more expensive)

Creating a VPC

- When creating a VPC you also get created corresponding:
    - Route Table
    - Network ACL
    - Security Group

Creating a Subnet

- https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Subnets.html
- Created with:
    - Name: 10.0.1.0 - us-east-1a
    - AZ: us-east-1a
    - 10.0.1.0/24
- If public subnet, click on subnet and select Modify *auto-assign public IPv4 addresses*
- Security Groups do not span VPCs
- Note Amazon reserve 5 IP's from this range for each subnet (first 4, last 1):

The first four IP addresses and the last IP address in each subnet CIDR block are not available for you to use, and cannot be assigned to an instance. For example, in a subnet with CIDR block `10.0.0.0/24`, the following five IP addresses are reserved:

- `10.0.0.0`: Network address.
- `10.0.0.1`: Reserved by AWS for the VPC router.
- `10.0.0.2`: Reserved by AWS. The IP address of the DNS server is always the base of the VPC network range plus two; however, we also reserve the base of each subnet range plus two. For VPCs with multiple CIDR blocks, the IP address of the DNS server is located in the primary CIDR. For more information, see Amazon DNS Server.
- `10.0.0.3`: Reserved by AWS for future use.
- `10.0.0.255`: Network broadcast address. We do not support broadcast in a VPC, therefore we reserve this address.

Currently you can create 200 subnets per VPC (support request to extend)
The minimum size of a subnet is a /28 (or 14 IP addresses.) for IPv4. Subnets cannot be larger than the VPC in which they are created.
An IP address assigned to a running instance can only be used again by another instance once that original running instance is in a "terminated" state (not "stopped" state).

Creating an Internet Gateway

- Created an Internet Gateway
- Attached to my VPC
- Cannot have multiple IG on VPC

Internet Connectivity

- You can use public IP addresses, including Elastic IP addresses (EIPs), to give instances in the VPC the ability to both directly communicate outbound to the Internet and to receive unsolicited inbound traffic from the Internet (e.g., web servers)
- Instances without public IP addresses can access the Internet in one of two ways:
    - Route their traffic through a NAT gateway or a NAT instance to access the Internet (doesn't allow machines on the Internet to initiate a connection to internal instances)
    - For VPCs with a hardware VPN connection or Direct Connect connection,

instances can route their Internet traffic through their internal data-centre

When using public IPs, traffic between two instances within a Region stays within the AWS network. When between instances in different Region with inter-region VPC peering will stay within AWS network however when but not using peering, no guarantee will stay within AWS network.

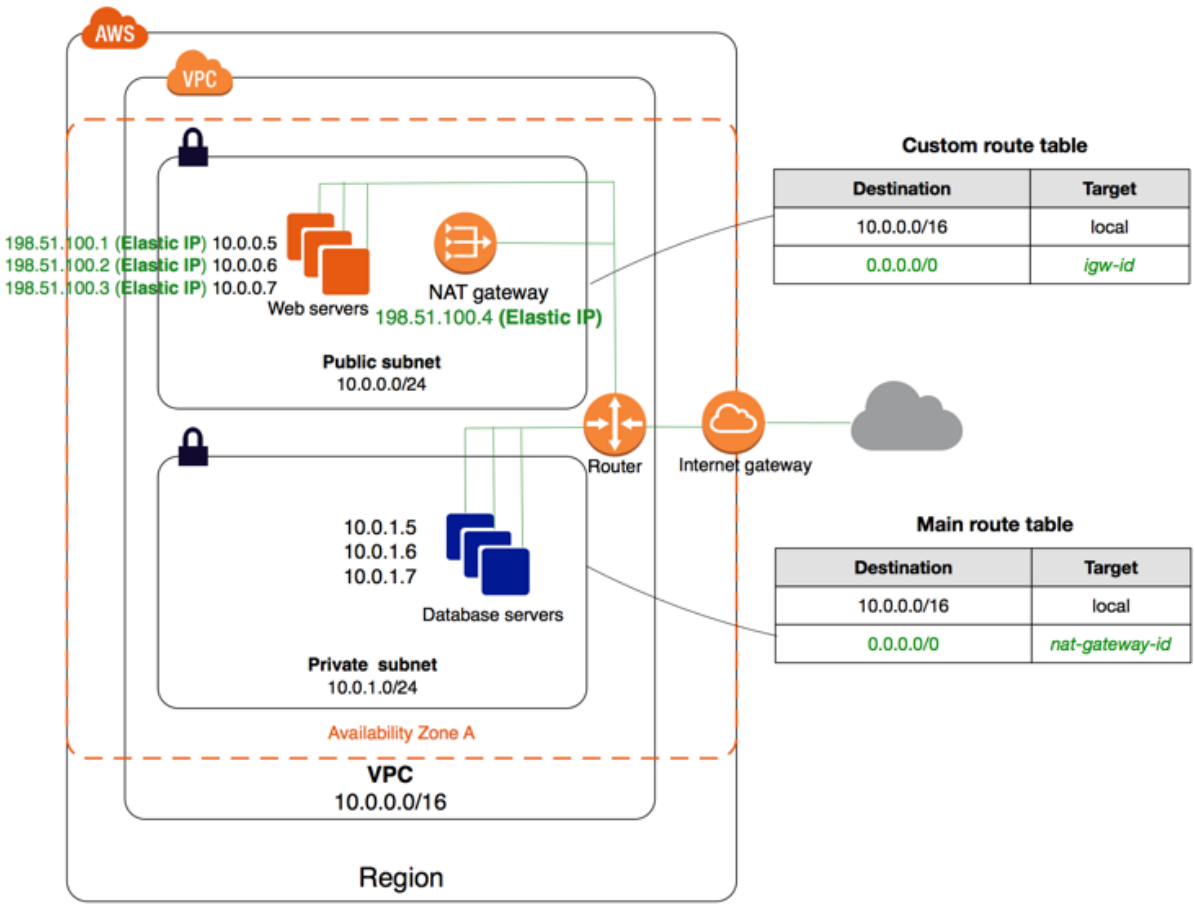
Creating a Route

- Two types of Route Tables:
  - Main
  - All others
- By default, all subnets are associated with the Main Route Table (and therefore Internet accessible if setup with Internet Gateway)
- Add a secondary (public) Route Table, associate with:
  - Destination: 0.0.0.0/0
  - Target: The Internet Gateway
- Whenever you create subnets now, by default they'll get associated with the Main Route Table (private subnet)
- You can create a default route for each subnet to egress the VPC via the Internet gateway, the virtual private gateway, or the NAT gateway

NAT Instances & NAT Gateways

- NAT allows private subnets egress traffic (ie allows instances in subnet to get to the internet), but does not allow any ingress traffic (ie traffic from the internet into the subnet)
- NAT Instances are on their way out, replaced by NAT Gateways
  - https://docs.aws.amazon.com/vpc/latest/userguide/vpc-nat-comparison.html
- NAT Instances
  - Are launched like any EC2 instance, using Community AMIs (NAT instance)
  - Add Instance to your VPC, and into the public subnet
  - Add a Security Group with SSH, HTTP, HTTPS access (so private subnet can initiate connections to internet over those protocols)
  - On NAT Instance, in *Networking* dropdown, *Change Source/Dest. Check* to disable
    - Each EC2 instance must always be either source or destination of traffic, NAT Instance are neither!
  - On your main Route Table, add a Route from 0.0.0.0/0 to your NAT Instance
  - Problems with NAT Instances
    - Single point of failure, limited compute capacity & network throughput, single OS etc. You can setup Auto Scaling Groups, across multiple AZ's but gets complicated and you have to manage them all yourself - this is the benefit of NAT Gateway
  - Can use as Bastion Server and setup Port Forwarding

- NAT Gateways:
    - https://docs.aws.amazon.com/vpc/latest/userguide/vpc-nat-gateway.html



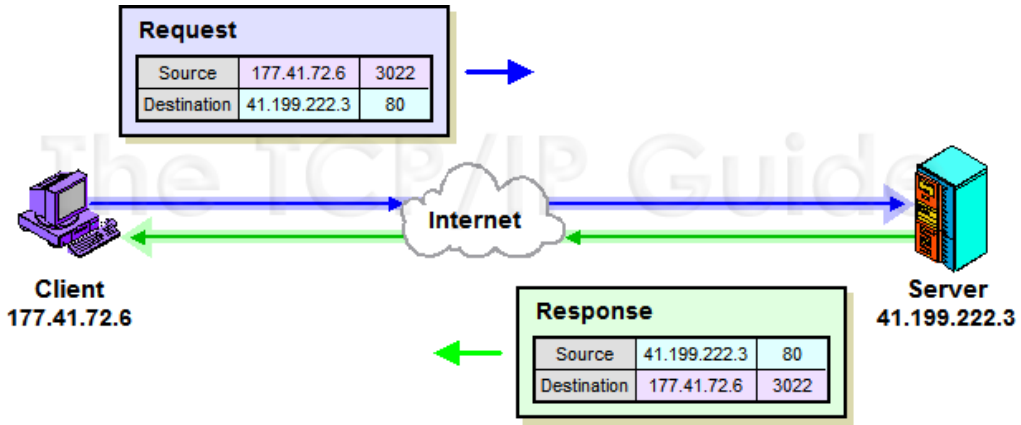
- Create NAT Gateway from VPC Dashboard
    - Egress-Only Internet Gateway is for IPv6 traffic only
    - NAT Gateway is for IPv4x
    - Add to the public Subnet
    - Allocate (create) an Elastic IP
    - Edit Route Table, on your main Route Table, add a Route from 0.0.0.0/0 to your NAT Gateway
- For availability, create a NAT Gateway in each AZ (ie public subnet)
- NAT Gateways support 5Gbps of bandwidth, **automatically scales up to 45Gbps (limitation)**
- Can associate exactly one Elastic IP with a NAT Gateway (ie NAT Gateways get assigned a public IP)
- NAT Gateways support TCP, UDP & ICMP protocols
- You cannot associate a Security Group with a NAT Gateway (only on the instances within the private subnets)
- You can associate Network ACLs with private subnet that apply to Internet

Gateway

- NAT Gateways are more secure than a NAT Instance

Network Access Control Lists (NACL)

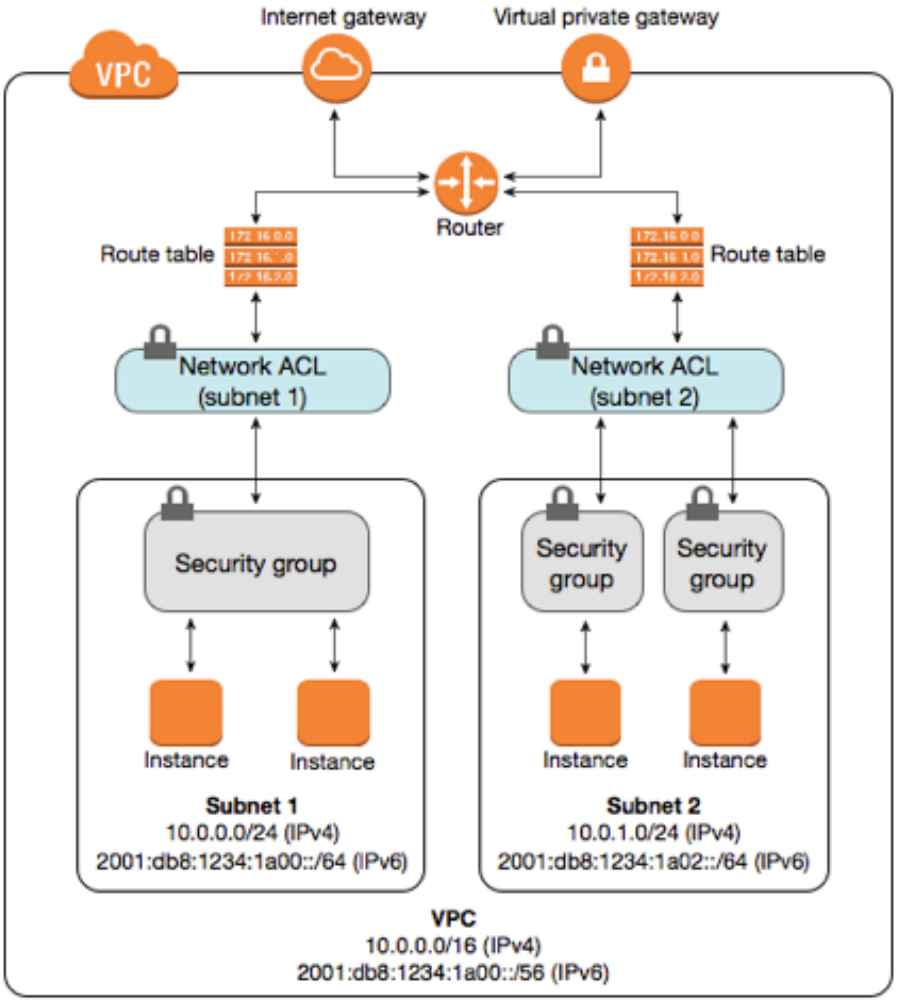
- https://docs.aws.amazon.com/vpc/latest/userguide/vpc-network-acls.html
- Network Access Control List adds an extra layer of security to your VPC and acts as a firewall for controlling traffic in/out of your subnets. Can Allow or Deny INGRESS and EGRESS traffic.
- A default NACL is created for your VPC, allowing all INGRESS & EGRESS traffic
- NACL's operate and are applied at a Subnet level (but do not filter traffic between instances within a subnet)
- Can create many NACL's per VPC, which default to denying all INGRESS & EGRESS traffic
- Can only associate a subnet with a single NACL, however a NACL can be associated with many subnets
- NACLs are stateless, you must specify INGRESS and EGRESS rules for every network connection
- For NACL Rules, start at Rule #100, and increment each Rule# by 100 (ie Rule#100: HTTP; Rule#200 HTTPS; Rule#300 SSH etc)
- Rules are applied in numerical order (earlier deny rules take precedent)
- NACL rules are applied BEFORE Security Groups
- Can block IP ranges with NACL
- For EGRESS, add Ephemeral Port rules:
    - Custom TCP Rule; Allow; Port Range 1024-65535
    - https://en.wikipedia.org/wiki/Ephemeral_port
    - https://docs.aws.amazon.com/vpc/latest/userguide/vpc-network-acls.html#nacl-ephemeral-ports
    - Short lived transport protocol (ie TCP/UDP) port that is specified by the client OS, so the remote server (well known port) can communicate back to the client. As an example for ssh connection might look like:
        - 192.168.1.102:37852 ---> 192.168.1.105:22



NACL vs Security Group

- https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Security.html#VPC_Security_Comparison

Security Group	Network ACL
Operates at the instance level	Operates at the subnet level
Supports allow rules only	Supports allow rules and deny rules
Is stateful: Return traffic is automatically allowed, regardless of any rules	Is stateless: Return traffic must be explicitly allowed by rules
We evaluate all rules before deciding whether to allow traffic	We process rules in number order when deciding whether to allow traffic
Applies to an instance only if someone specifies the security group when launching the instance, or associates the security group with the instance later on	Automatically applies to all instances in the subnets it's associated with (therefore, you don't have to rely on users to specify the security group)



VPC Flow Logs

- Capture information about the IP traffic to network interfaces within in your VPC, stored within CloudWatch
- Flow logs can be created at 3 levels:
  - VPC
  - Subnet
  - Network Interface Level
- Need to setup Flog Log IAM role
- Need to setup CloudWatch log group

    - Can stream events to Lambda (react in real-time)
    - Can stream events to Elasticache
    - Can export logs to S3
- Cannot enable VPC Flow Logs on peered VPC accounts (VPC must be in your AWS account)
- Cant tag a Flow Log
- Cant change configuration (i.e. IAM role) after created
- Not all IP traffic is monitored i.e.:
    - AWS DNS server traffic (your own DNS server is logged)
    - Windows instances for Amazon Windows license activation
    - Traffic to/from Amazon Metadata (i.e. 169.254.169.254)
    - Traffic to the reserved IP address for the default VPC router

Bastion / Jump boxes

- Public exposed instances setup to securely access your internal AWS network for admin (i.e. using SSH or RDP).
- Highly secure / locked down instance
- Normally accessed via VPN / secure channel
- Can setup highly available Bastion servers using Auto-Scaling-Groups, two public subnets / AZ's and use Route53 health checks

VPC End Points (also know as AWS Private Link)
An internal gateway within the Amazon network to attach your VPC to another Amazon service (i.e. S3, dynamoDB, EC2 etc) i.e prevents your EC2 instances having to connect to S3 over the public internet.
Two types of VPC End Points:

- Interface - Elastic network interface (ENI) attached to your EC2 instances
- Gateway - Creates a target in your Route table for the associated service

As an example, If you setup a VPC Endpoint to S3 from a private subnet, you no longer need corresponding NAT Gateway to access S3
You can now setup your application as accessible over a AWS Private Link / VPC endpoint, other consumers can connect their VPC to your via a VPC endpoint

---

Direct Connect

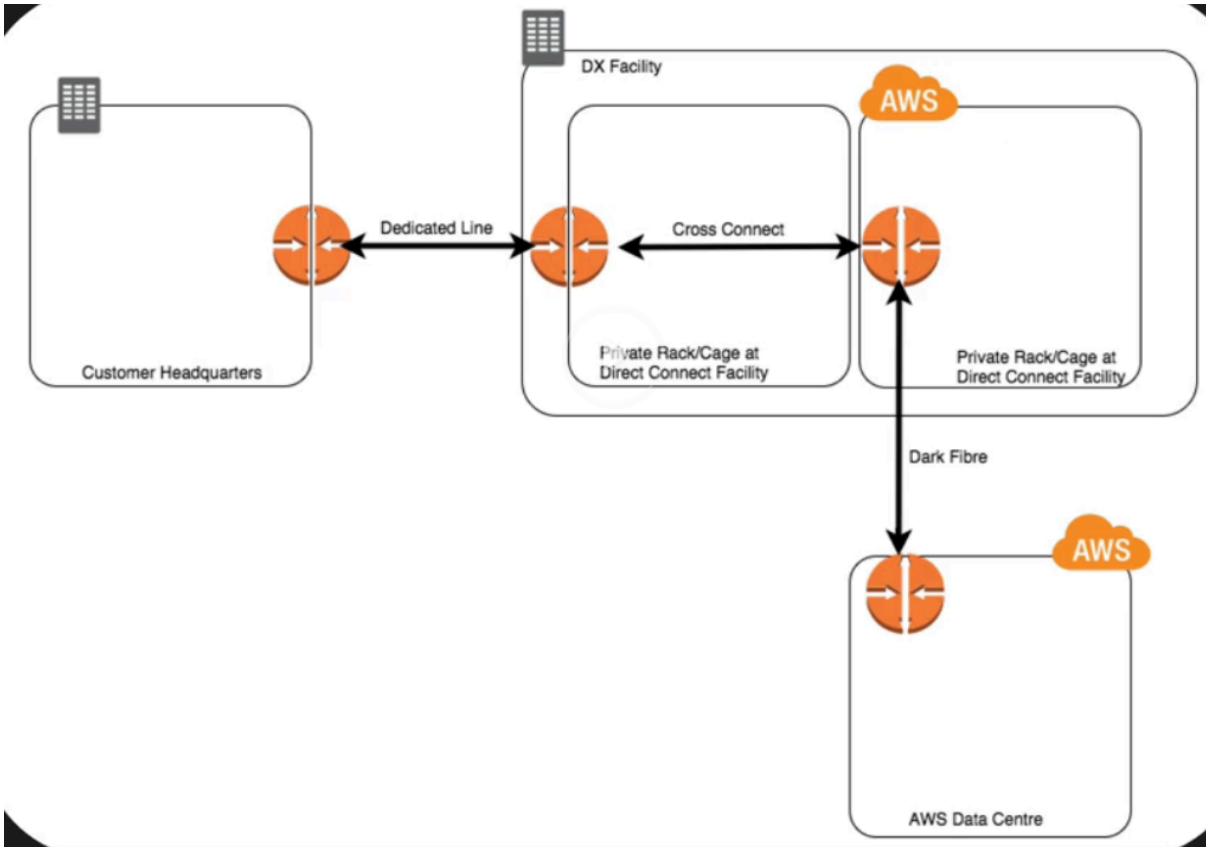
Establishes a dedicated private network connection from your on premises data-centre to AWS
Improves network performance and can reduce network costs (over internet based VPN costs) for large volumes of data. Available in:

- 10Gbps
- 1Gbps
- Sub 1 Gbps through Direct Connect partners

Uses Ethernet VLAN trunking (802.1Q)

VPN Connections can be configured quickly and provide modest performance over the public Internet
vs.
Direct Connect does not use the internet; but a private dedicated network connection direct to AWS. Can take 1 to 4+ months to setup dedicated network.

For Direct Connect, your Data Centre connects via a dedicated line (provided by your telco - ie Telstra / Optus etc) to a Direct Connect facility, which has a Dark Fibre connection into the AWS Data Centre/ backbone/network.



VPN

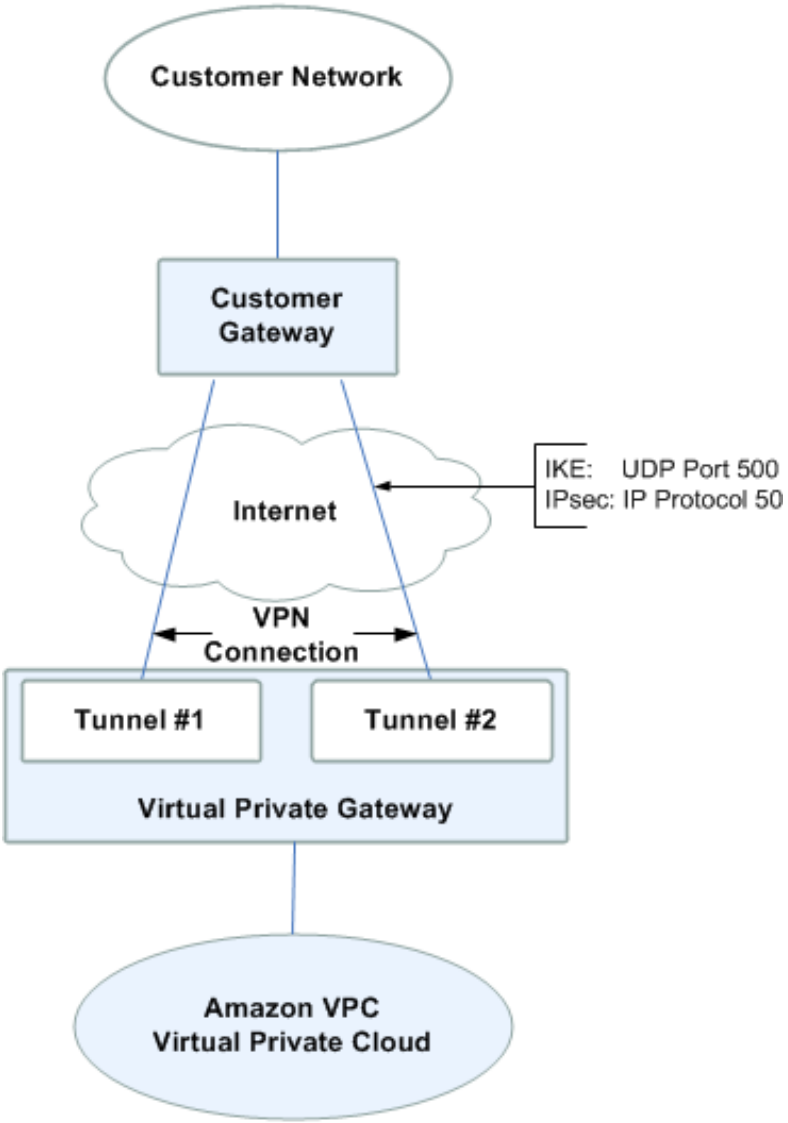
AWS VPN is comprised of two services:

- AWS Site-to-Site VPN - connect your on-premises network or branch office site to your Amazon Virtual Private Cloud (Amazon VPC)
- AWS Client VPN - connect users to AWS or on-premises networks.

AWS Site-to-Site VPN
Customer Gateway is on the customers side
Virtual Private Gateway is the Amazon side of a VPN connection
An AWS Site-to-Site VPN connection connects your datacenter to Amazon VPC, Amazon supports Internet Protocol Security (IPSec). An internet gateway is not required to establish an AWS Site-to-Site VPN connection. To use this service, you must have an internet-routable IP address to use as the endpoint for the IPsec tunnels connecting your customer gateway to the virtual private gateway. If a firewall is in place between the internet and your gateway, the rules in the following tables must be in place to establish the IPsec tunnels.

Virtual gateway supports IPSEC VPN throughput up to 1.25 Gbps
You can have:

- Five virtual private gateways per AWS account per AWS Region
- Fifty customer gateways per AWS account per AWS Region
- Ten IPsec VPN Connections per virtual private gateway



https://docs.aws.amazon.com/vpc/latest/userguide/VPC_Scenario4.html

- VPC with a Private Subnet Only and AWS Site-to-Site VPN Access
    -

Simple Queue Service (SQS)

https://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/sqs-how-it-works.html
Distributed Message queue provided via web service
A pull based system (i.e. Messages are pulled from the Queue)
Used to decouple architecture and enable fail-safe & asynchronous communication
Messages can contain upto 256KB of text in any format (i.e. plain text, JSON, XML,

custom DSL etc)
Messages kept in Queue for anywhere from 1minute to 14 days (default retention is 4 days)
SQS WaitTimeSeconds is the duration (in seconds) for which the call waits for a message to arrive in the queue before returning
SQS Visibility Timeout is the amount of time the Message is invisible when being processed by a job. If job longer than timeout, Message will become visible for other jobs to process it.

- Default Visibility Timeout is 30 seconds
- Maximum Default Visibility Timeout is 12 hours

SQS Long Polling is a way to retrieve Messages where a request to the Message Queue doesn't respond immediately, but waits until a Message arrives in the Queue (or the Long Poll times out). Long Polling can save you money. Set *Receive Message Wait Time* or *WaitTimeSeconds* to setup Long Polling
Two different types of queues

- Standard Queue - Default queue, nearly unlimited amount of transactions, guaranteed at-least-once delivery, best effort ordering (sometimes more than one Message delivered and out of order)
- FIFO Queue (First in First Out) - Guaranteed delivery in order, processed exactly once (lives until deleted) and limited to 300 transactions per second

JavaScript NodeJS SQS examples: https://docs.aws.amazon.com/sdk-for-javascript/v2/developer-guide/sqs-examples.html

---

Simple Workflow Service (SWF)

SWF is a web service which coordinates work across distributed application components, represented in processing steps.

- SWF *Domains* is a scoped context (isolated) which contains your workflow, tasks & executions; Workers & Deciders.
- *Workflow Starters* - An application which initiates a workflow
- *Decider* is a program that controls the coordination of tasks (ordering, concurrency and scheduling)
- *Activity Workers* are programs which interacts with SWF to get tasks, process and return results

Workers and Decider can run on AWS or your remote data centre (internal machines behind firewall) compute and SWF stores tasks, assigns to Worker, monitors progress and ensures a task is **assigned once** (and never duplicated) - SWF maintains the applications state durably

Maximum workflow duration is 1 year

SWF presents a task-orientated API, where SQS offers message-oriented API

When should I use Amazon SWF vs. AWS Step Functions?
AWS Step Functions is a fully managed service that makes it easy to coordinate the components of distributed applications and microservices using visual workflows. Instead of writing a Decider program, you define state machines in JSON. If Step Functions does not fit your needs, then you should consider Amazon Simple Workflow (SWF). Amazon SWF provides you complete control over your orchestration logic, but increases the complexity of developing applications.

Simple Notification Service (SNS)

SNS is a web service which allows pushed-based notifications to be sent to subscribers
SNS follows publish/subscribe (pub-sub) messaging standard with a push mechanism (avoiding having to constantly poll)
Data type/format is JSON
Can deliver notifications via:

- Mobile - iOS, Android, Google, Windows, Fire OS
- SMS
- Email
- Email JSON
- SQS
- Any HTTP / HTTPS endpoint
- Lambda

Pay as you go model:

- $0.50 per 1m SNS requests
- $0.06 per 100K notifications over HTTP
- $0.75 per 100 notifications over SMS
- $2.00 per 100K notifications over Email

All notification messages are stored across multiple AZ's (redundantly)
SNS is made up of:

- A Topic (an access or endpoint to subscribe)
- Subscribers of the Topic

One topic can support deliveries to multiple endpoint types

Elastic Transcoder

Media transcoder in the cloud; converts media files from original source to a different format to play on other devices / channels
Supports popular output formats

Pay based on minutes that you transcode and the resolution you transcode

---

API Gateway

AWS fully managed API Gateway to publish, maintain, monitor and secure APIs at scale
Supports API caching based on TTL period (in seconds)
Fully scalable, low cost
Can throttle requests
Can connect to CloudWatch and log traffic
Supports client-side COR configuration for same-origin-policy relaxation - enable CORS in API Gateway

---

Kinesis 101

Streaming data is data that is generated continuously by thousands of data sources, simultaneously and in small sizes (KBs) - think online stores, stock prices, online game data, social network data, geospatial data, IoT data etc

Kinesis is an AWS platform to send streaming data too, enabling easy load and analysis of streams of data. Three core Kinesis Services:

- Kinesis Streams
    - *Producers* are data producers / sources like EC2, Mobile device, desktop web, games, IoT devices etc
    - Data is stored in a *Shard* by default for 24 hours, and up to 7 days
        - Supports 5 transactions per second for reads, up to 2MB read rate per second
        - Supports up to 1000 records per second for write, up to 1MB write rate per second
        - Can have multiple Shards per Stream
    - Consumers (EC2 instances) subscribe to a Shard and consume data, analyse and stored if required (i.e. in DynamoDB, S3, EMR, Redshift etc)
- Kinesis Firehose
    - Have *Producers*, but you don't have to worry about *Shards* or *Streams.* Scaling is automated, can analyse incoming data via Lambda and send to S3
    - Does not support data storage/retention like Streams. Can send to S3 and Redshift (via S3) or ElasticSearch Cluster
- Kinesis Analytics
    - Run SQL queries on Streams and Firehouse, and can store data into S3, Redshift, Elasticsearch Cluster

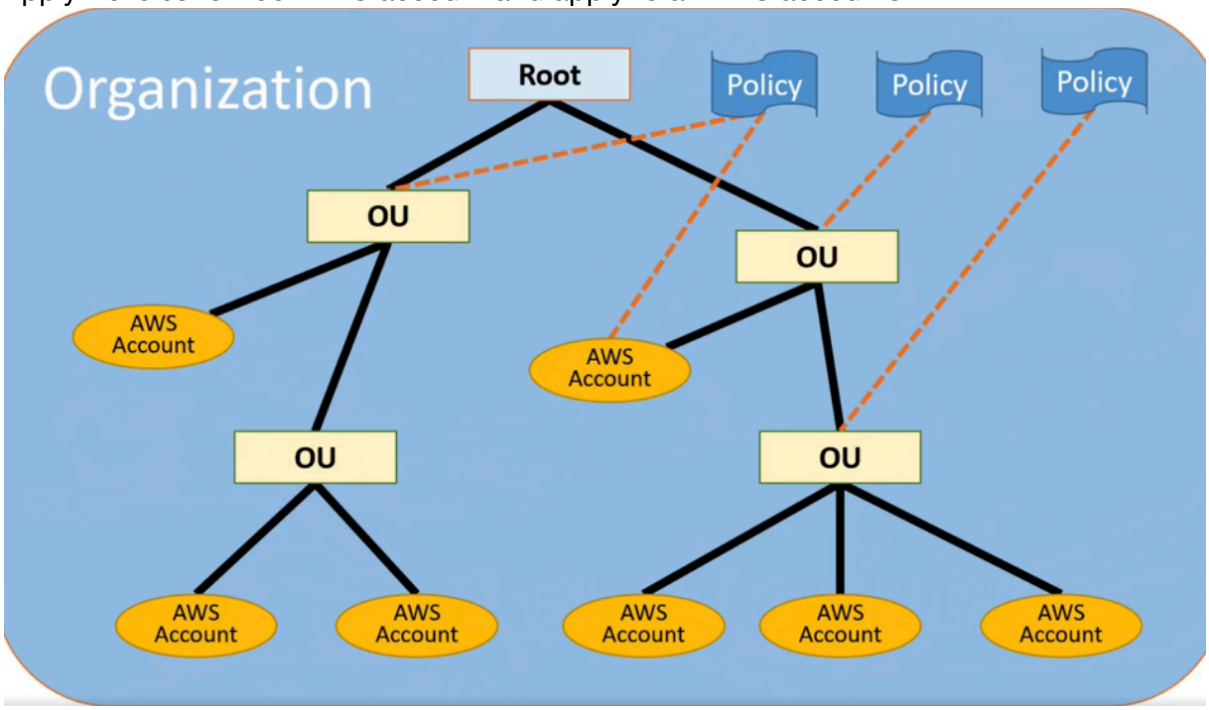
---

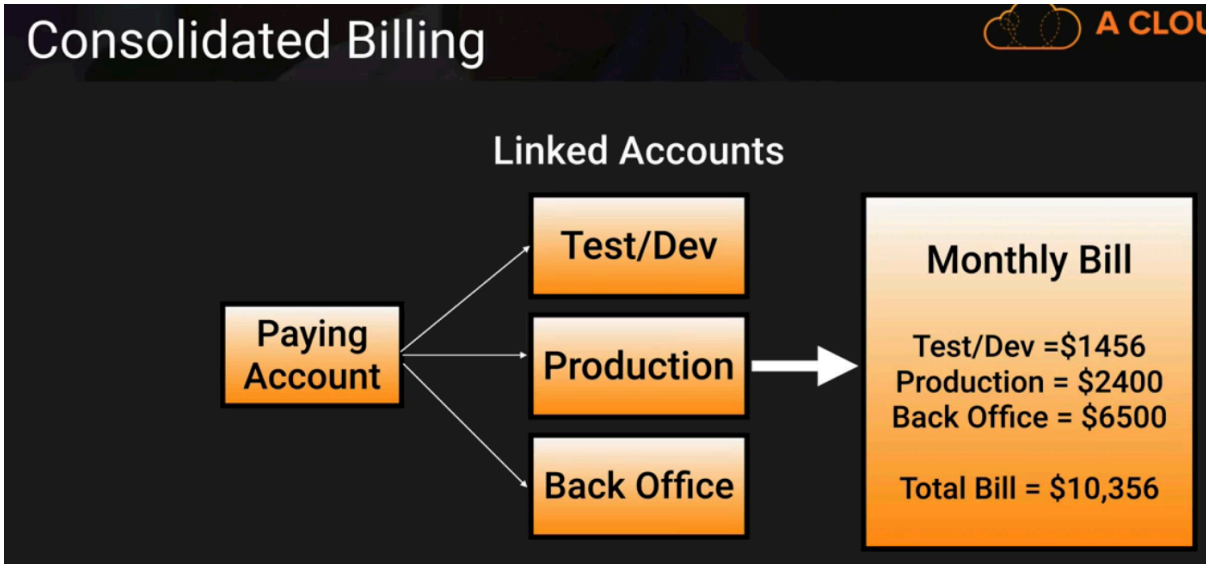
AWS Organisations

Enables you to manage and group multiple AWS accounts into a single view to centrally manage. Has two feature sets:

- Consolidated Billing
- All Features

Apply Policies to Root AWS account and apply to all AWS accounts





Always use multi-factor authentication on root account
Root account / paying account should be used for billing purposes only (do not deploy resources)
Can only link 20 accounts (by default)

Can setup Billing alerts for all accounts
Consolidated billing allows for volume discounts across your AWS accounts (including Reserved Instances usage)
Cloud Trail still remains at a per AWS account level (but can be aggregated into a single S3 bucket)
Can control access to AWS Services using Service Control Policies (allow or deny ie Blacklisting or WhiteListing
)
Can create Organisational Units (groups) of AWS accounts and attach Policies

---

Cross Account Access

For organisations with multiple AWS accounts, enables you to manage accounts (ie via Console) without having to re-login using different credentials

- To delegate permission to access a resource, create an IAM role that has two policies attached (within Trusting Account):
    - The permissions policy grants the user of the role the needed permissions to carry out the desired tasks on the resource.
    - The trust policy specifies which trusted accounts are allowed to grant its users permissions to assume the role
- The other half is a permissions policy attached to the user in the trusted account that allows that user to switch to, or assume the role.

Tagging & Resource Groups

Tagging is key/value pairs attached to an AWS resource i.e. resource meta-data
Resource groups allow you to group your resources based on their tags and are either:

- Classic Resource Groups
- AWS Systems Manager (Region based)

With Systems Manager you can see Insights / Compliance / Inventory and Execute Automation Steps on group resources

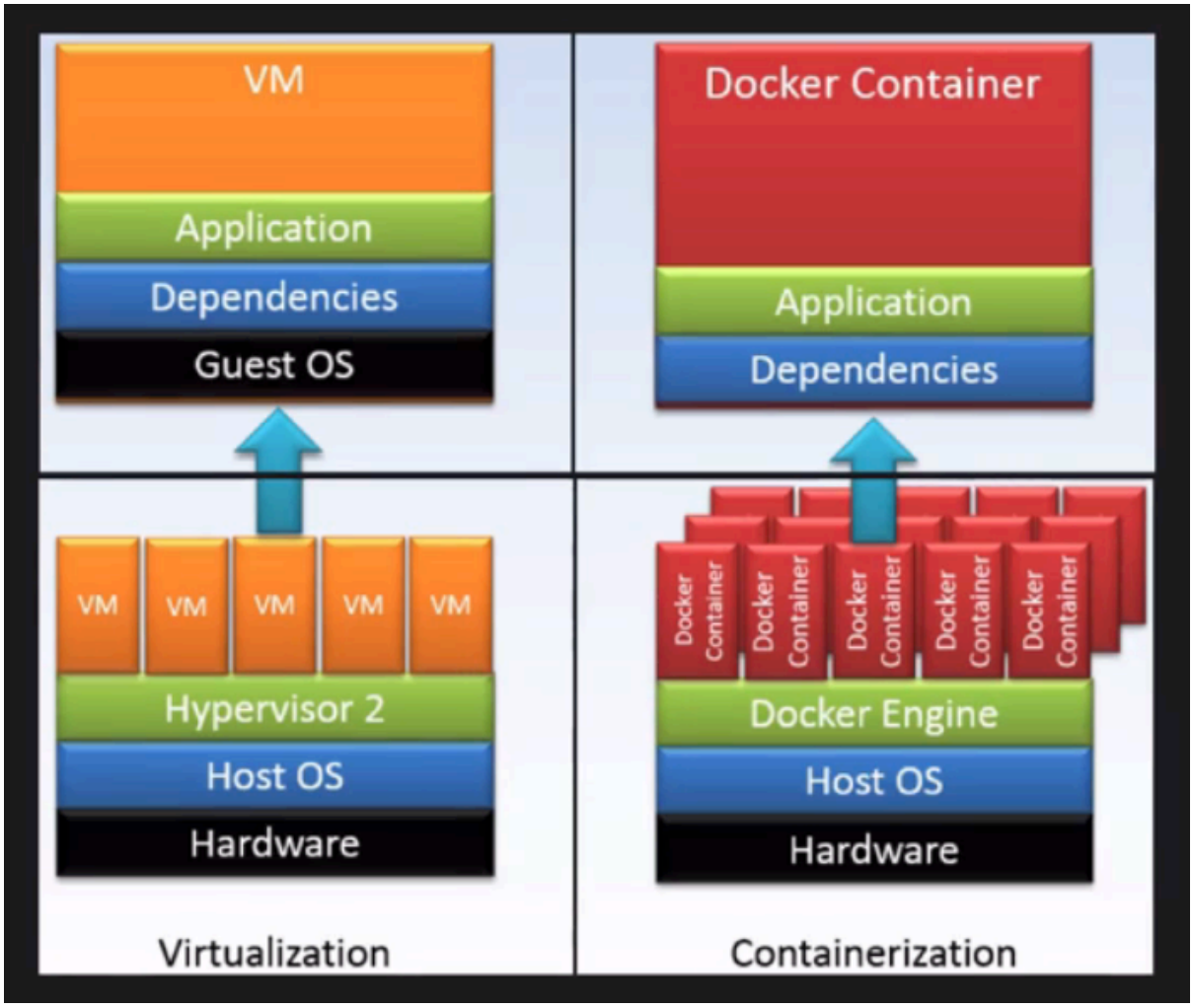
Workspaces

Virtual Desktop Infrastructure
Replaces local PC with cloud-based PC/services such as compute, storage, operating system & applications
Connect to your VDI via a local (dumb) device such as PC, Tablet etc
Can use Workspace credentials or ActiveDirectory credentials to login (***you do not need*** an AWS account to login to Workspaces)
Runs Windows 7 experience on top of Windows Server 2008 R2
By default, users are given local admin access to install their own applications

Workspaces are persistent and users can personalise
All data on D:\ drive is back-up every 12 hours

ECS

Elastic Container Service
Build, test & deploy applications in a standardised way using Containers
A standard for packing & versioning a applications code, configuration and dependencies into a single deployable unit



Traditional Virtualisation vs. Containerisation

Docker Components

- Docker Image - contains a read-only template of an executable package of O/S files, application & dependencies required to run
- Docker Layers - Many docker images over-layed to create Docker Container using Union File System
- Docker Container - a instance of the Docker Image ready to run (created from

Docker Image)
- DockerFile - Specification for a Docker Image, used to create a Docker Image (using Docker Layers to implement)
- Docker Daemon/Engine - Runs Docker Containers
- Docker Client - Docker CLI to interact with Docker environment (ie Daemon / Containers)
- Docker Registries / Docker Hub - Register and upload Docker Images / Docker Layers

ECS is a highly scalable, fast container management system, running on top of a cluster of EC2 instances.
ECS is a Region based service aiding in management of Docker based applications (don't have to worry about cluster management, configuration management or scaling)
Elastic Container Registry (ECR) is a private, managed Docker Registry to store your Docker Images / Layers

ECS Task Definition are required to run Docker Containers on ECS.

- JSON format that describes on or more Docker Containers
- Can specify Docker Images, CPU, Memory, Networking, Port Mappings, failure recovery/termination, boot-strap command, environment variables, data volumes and IAM permissions

Amazon ECS Container Agent allows Docker Containers instances to connect to your cluster.

- Runs on all ECS optimised AMI's, but you can install on any EC2 instances which supports the ECS specification
- Will not work with Windows

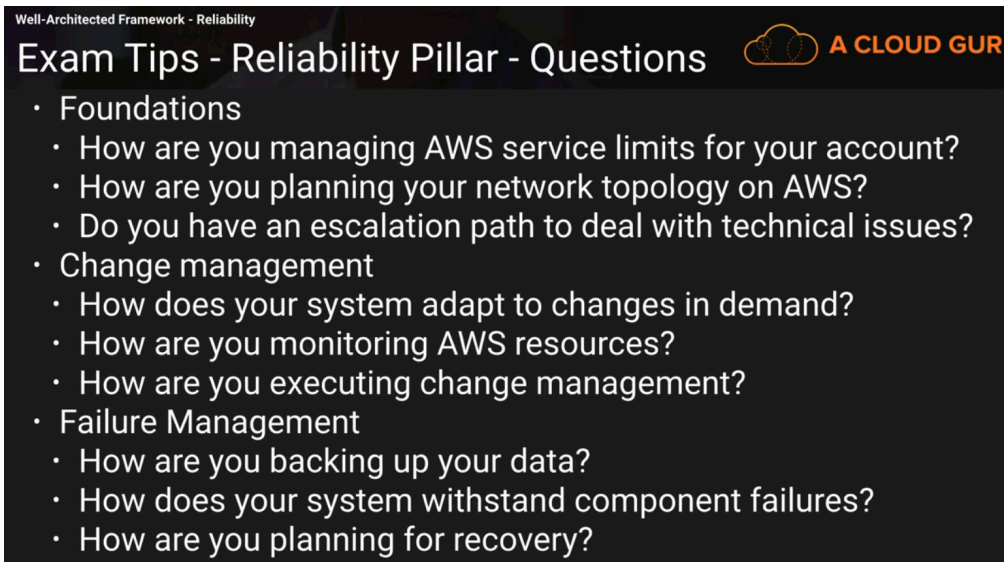
ECS Security & IAM

- ECS Tasks use an IAM role to access AWS services & resources
- Security Groups attached at the ECS Instance-level
- EC2 instances use IAM roles to access ECS
- You an access the underlying OS on the EC2 instance running your ECS cluster

ECS Scheduling via:

- Service Scheduler
- Custom Scheduler

Well Architected Framework



Well-Architected Framework - Performance Efficiency

Exam Tips - Performance Efficiency Pillar

Performance Efficiency in the cloud consists of 4 areas;

- Compute
- Storage
- Database
- Space-time trade-off

Well-Architected Framework - Performance Efficiency

Exam Tips - Performance Efficiency Pillar - Questions

A CLOUD GURU

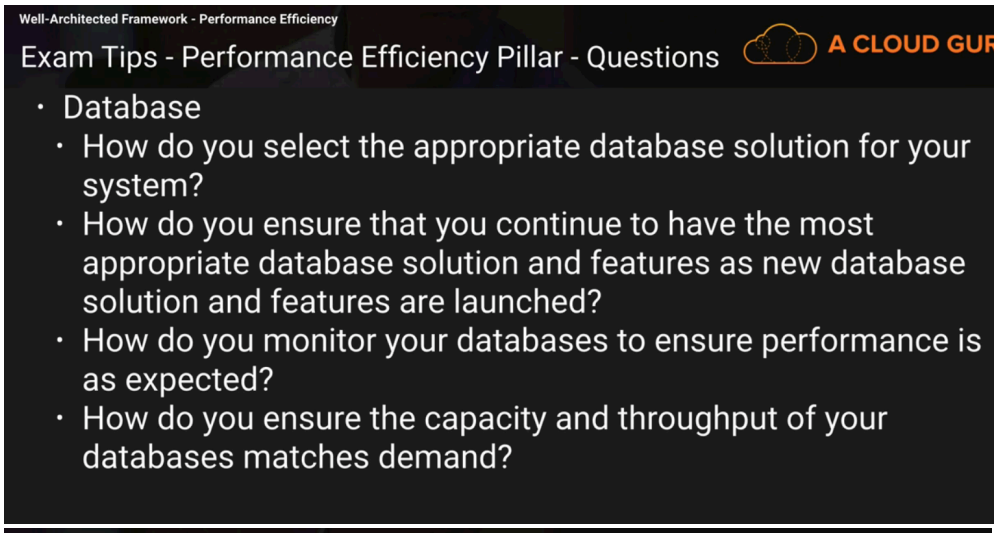
- Compute
  - How do you select the appropriate instance type for your system?
  - How do you ensure that you continue to have the most appropriate instance type as new instance types and features are introduced?
  - How do you monitor your instances post launch to ensure they are performing as expected?
  - How do you ensure that the quantity of your instances matches demand?

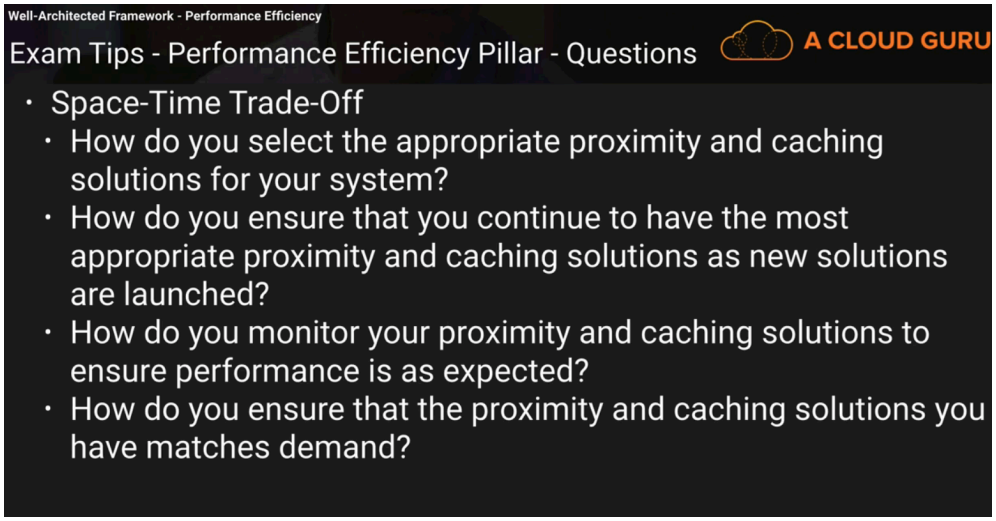
Well-Architected Framework - Performance Efficiency

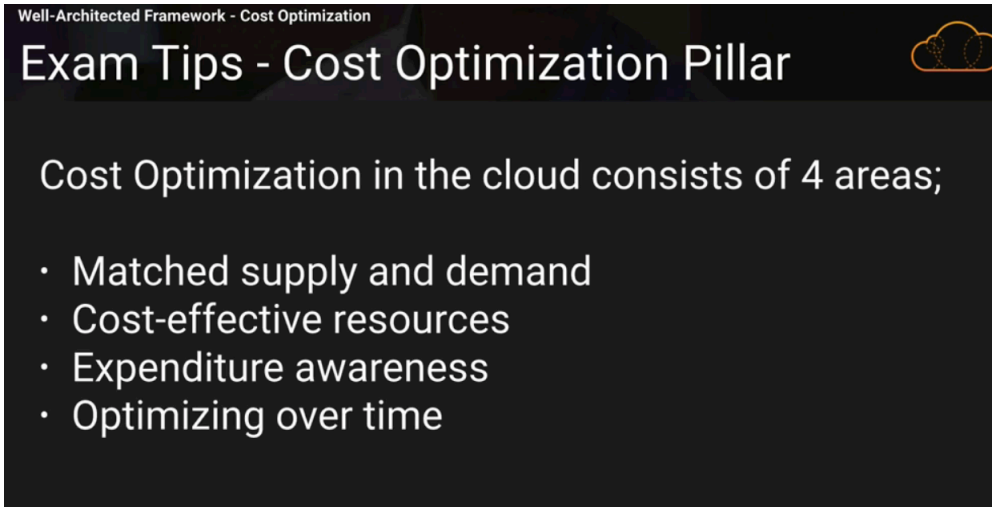
Exam Tips - Performance Efficiency Pillar - Questions

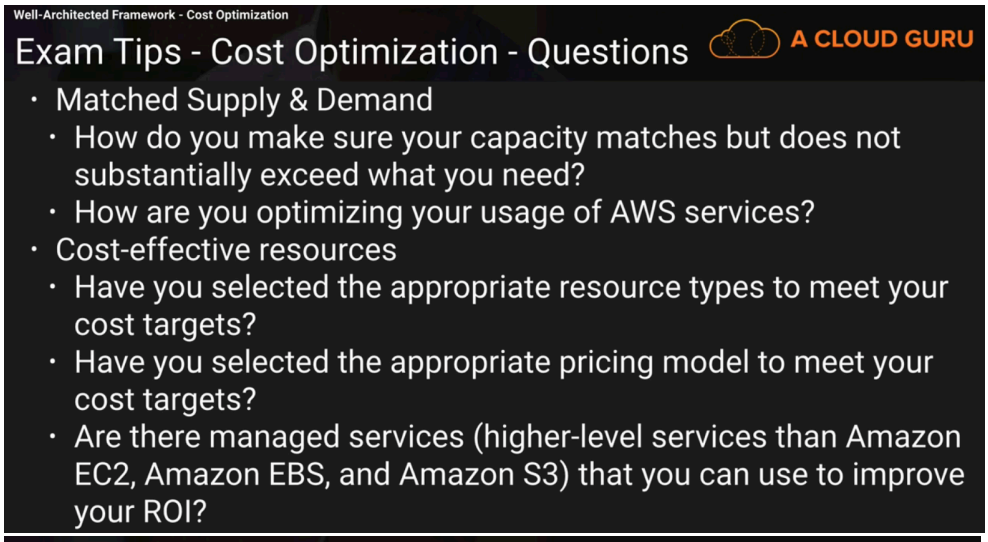
A CLOUD GU

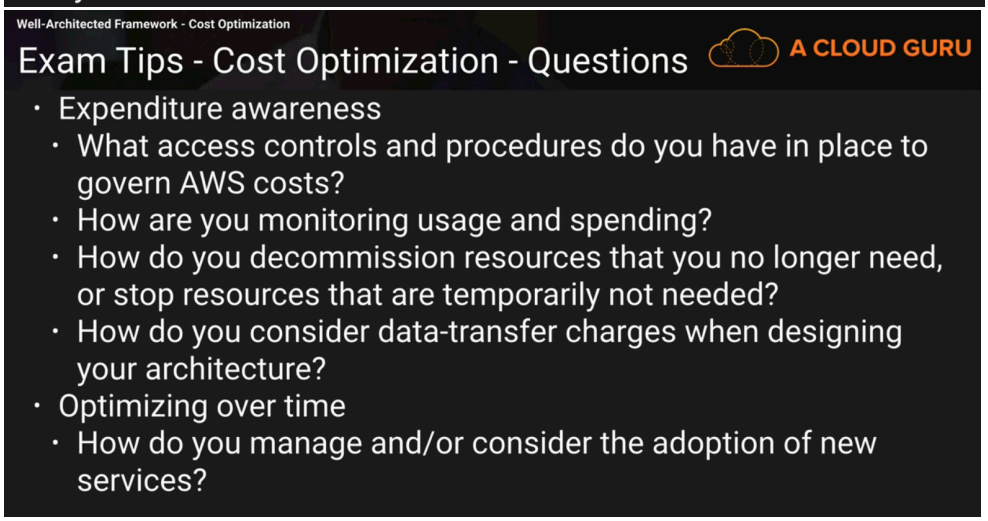
- Storage
  - How do you select the appropriate storage solution for your system?
  - How do you ensure that you continue to have the most appropriate storage solution as new storage solutions and features are launched?
  - How do you monitor your storage solution to ensure it is performing as expected?
  - How do you ensure that the capacity and throughput of your storage solutions matches demand?

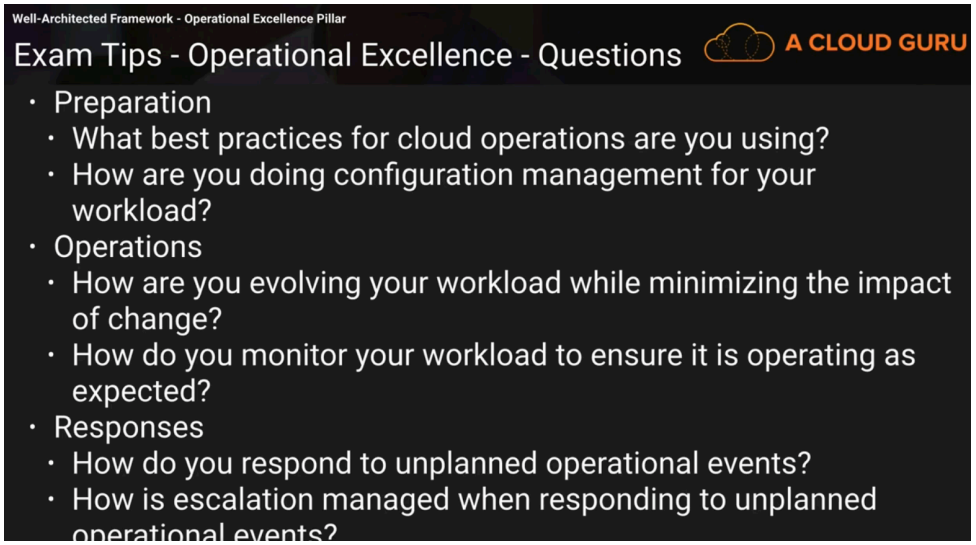

Well-Architected Framework - Performance Efficiency
Exam Tips - Performance Efficiency Pillar - Questions
A CLOUD GUR
• Database
• How do you select the appropriate database solution for your system?
• How do you ensure that you continue to have the most appropriate database solution and features as new database solution and features are launched?
• How do you monitor your databases to ensure performance is as expected?
• How do you ensure the capacity and throughput of your databases matches demand?


Well-Architected Framework - Performance Efficiency
Exam Tips - Performance Efficiency Pillar - Questions
A CLOUD GURU
• Space-Time Trade-Off
• How do you select the appropriate proximity and caching solutions for your system?
• How do you ensure that you continue to have the most appropriate proximity and caching solutions as new solutions are launched?
• How do you monitor your proximity and caching solutions to ensure performance is as expected?
• How do you ensure that the proximity and caching solutions you have matches demand?


Well-Architected Framework - Cost Optimization
Exam Tips - Cost Optimization Pillar
Cost Optimization in the cloud consists of 4 areas;
• Matched supply and demand
• Cost-effective resources
• Expenditure awareness
• Optimizing over time





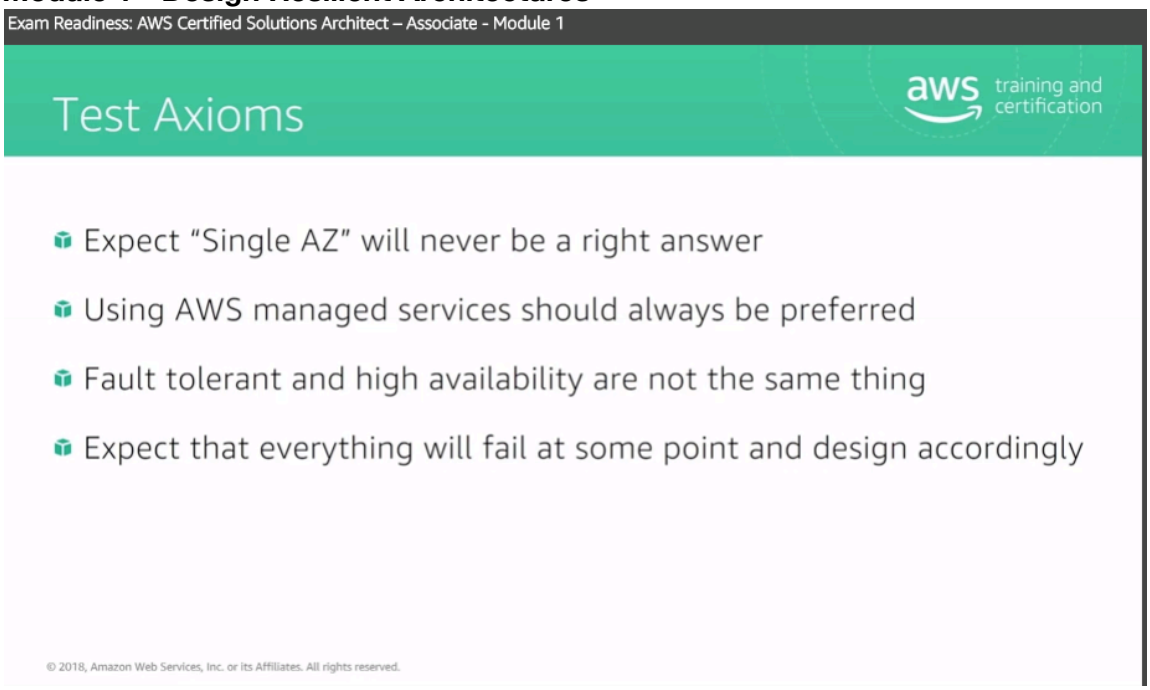


AWS Service Limits

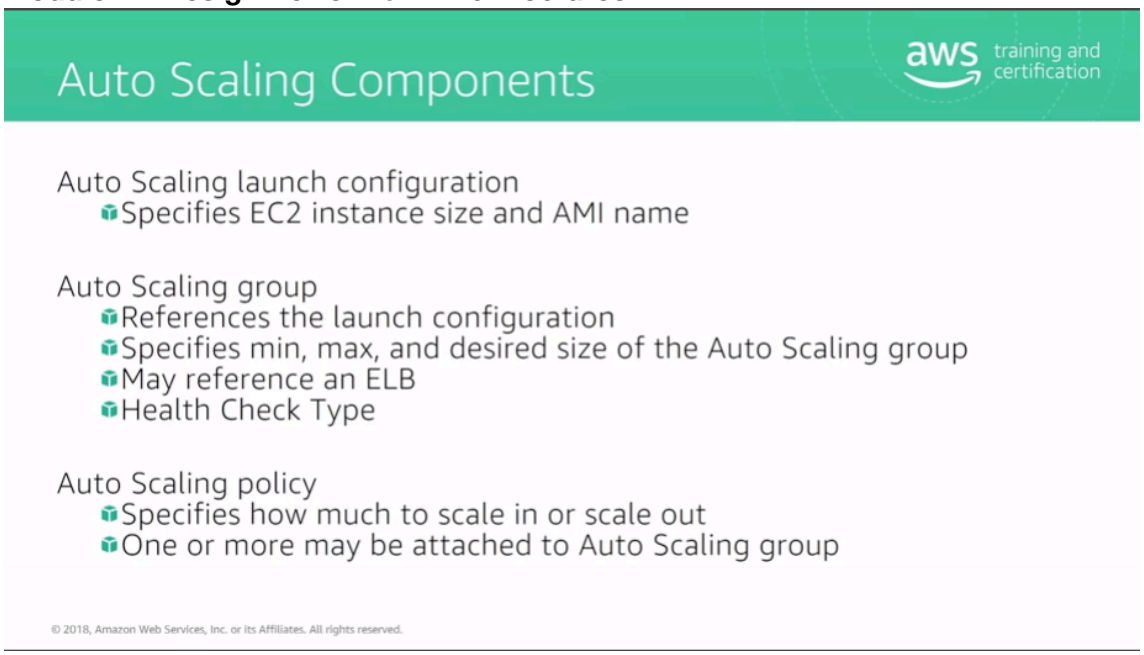
https://docs.aws.amazon.com/general/latest/gr/aws_service_limits.html

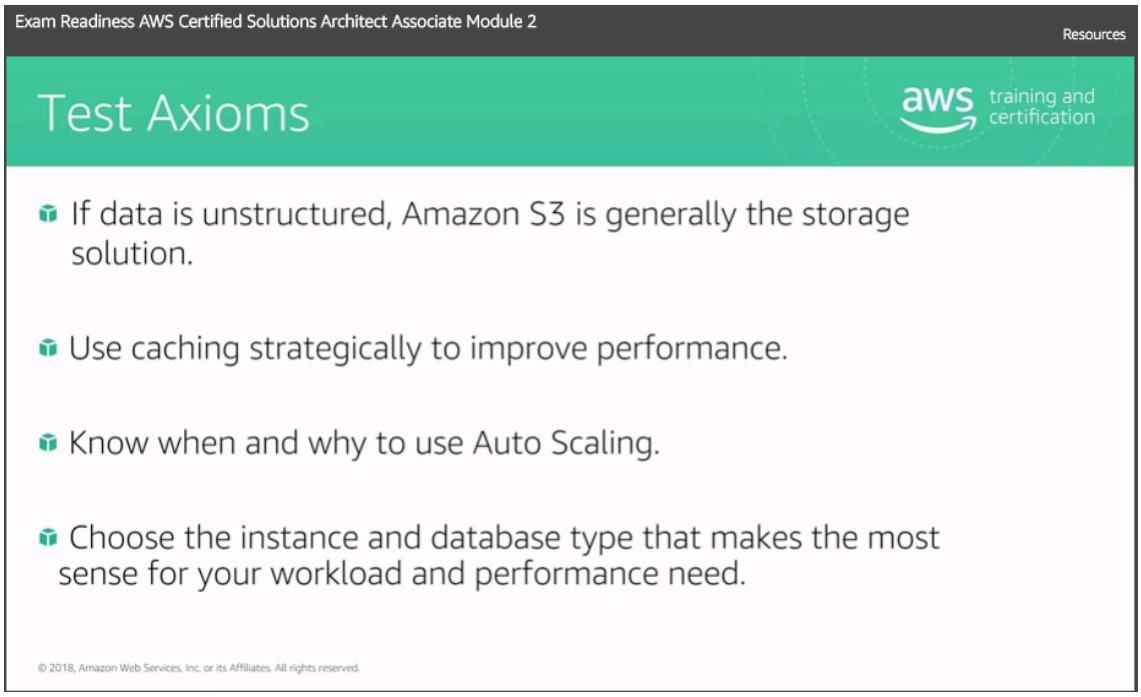
AWS Training - Exam Readiness: AWS Certified Solutions Architect - Associate

Module 1 - Design Resilient Architectures

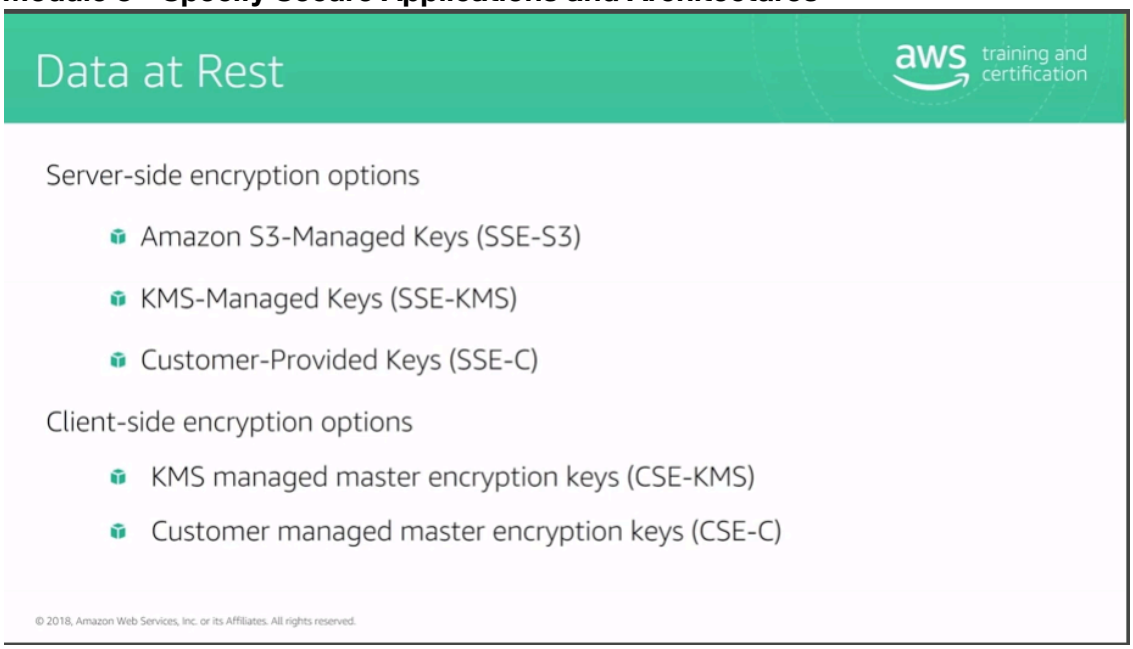


Module 2 - Design Performant Architectures





Module 3 - Specify Secure Applications and Architectures



Test Axioms

aws training and certification

- Lock down the root user
- Security groups only *allow*. Network ACLs allow explicit *deny*.
- Prefer IAM Roles to access keys

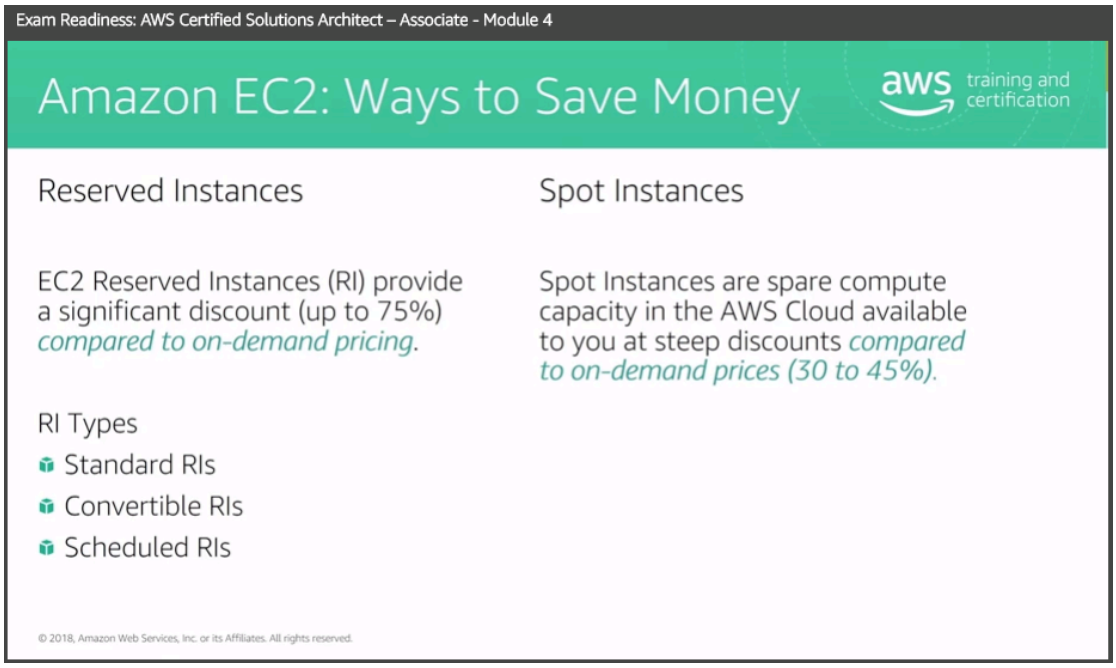
Module 4 - Design Cost-Optimized Architectures

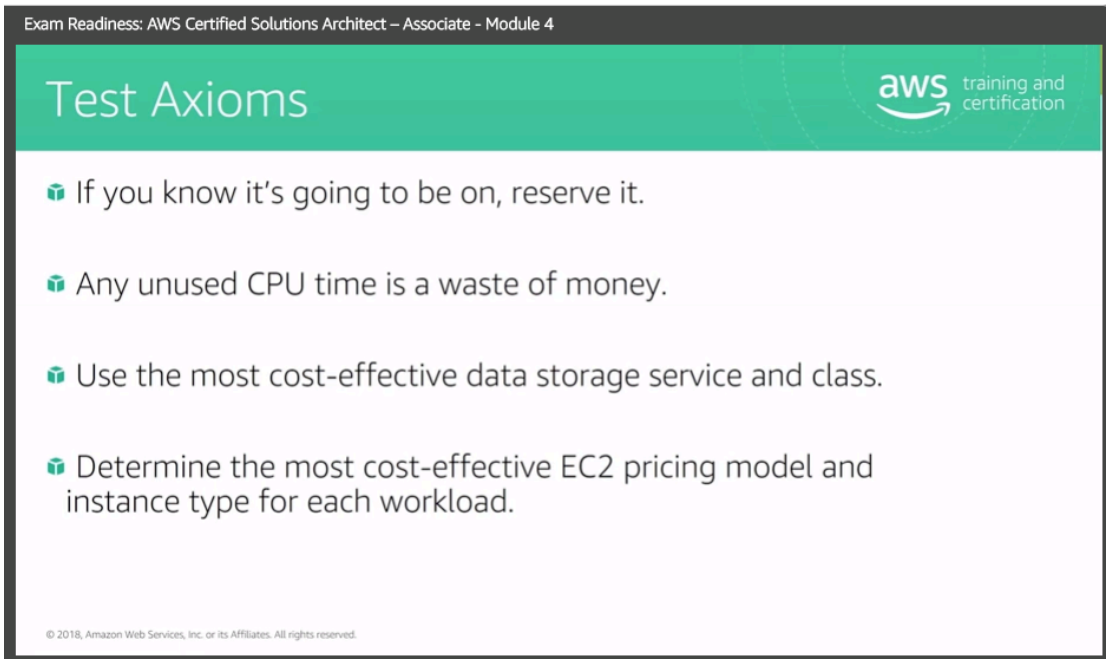
Fundamental Pricing Characteristics

aws training and certification

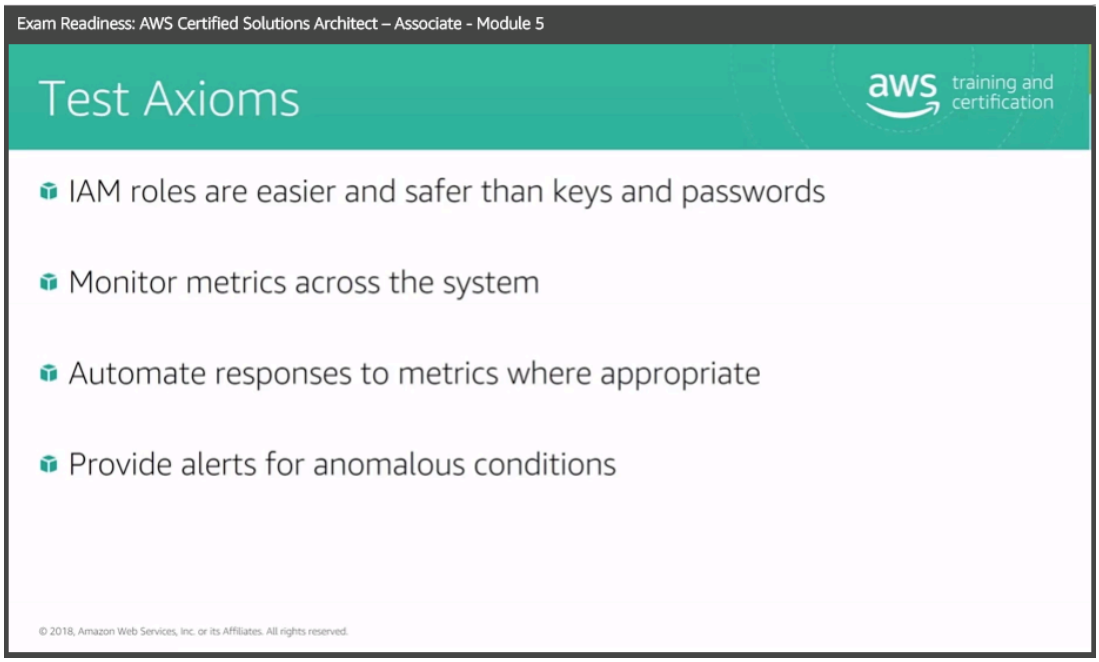
Three fundamental characteristics you pay for with AWS:

- Compute
- Storage
- Data transfer





Module 5 - Define Operationally-excellent Architectures



AWS Frequently Asked Questions (FAQs)

Read these before the exam https://aws.amazon.com/faqs/

- EC2: https://aws.amazon.com/ec2/faqs/
- VPC: https://aws.amazon.com/vpc/faqs/
- VPN: https://aws.amazon.com/vpn/faqs/
- S3: https://aws.amazon.com/s3/faqs/
- ELB: https://aws.amazon.com/elasticloadbalancing/faqs/
- SQS: https://aws.amazon.com/sqs/faqs/
- IAM: https://aws.amazon.com/iam/faqs/
- Lambda: https://aws.amazon.com/lambda/faqs/
- RDS: https://aws.amazon.com/rds/faqs/
- DynamoDB: https://aws.amazon.com/dynamodb/faqs/

AWS Whitepapers

- Overview of Security Processes: https://aws.amazon.com/whitepapers/overview-of-security-processes/
- Architecting for the Cloud: https://aws.amazon.com/whitepapers/architecting-for-the-aws-cloud-best-practices/
    - https://d0.awsstatic.com/whitepapers/AWS_Cloud_Best_Practices.pdf
- AWS Well Architected Framework: https://d1.awsstatic.com/whitepapers/architecture/AWS_Well-Architected_Framework.pdf
- AWS Storage Service Options: https://aws.amazon.com/whitepapers/storage-

options-aws-cloud/
- Building Fault Tolerant Apps: https://d0.awsstatic.com/whitepapers/aws-building-fault-tolerant-applications.pdf

---

Glossary / Terms

Redundancy: Available in two or more places
Durability: Data survivability - storing for a long time with no corruption or loss
Reliability: Consistently perform according to its specifications
Availability: System / Service uptime
RTO: Recovery Time Objective - how long to restore service
RPO: Recovery Point Objective - how much data loss is acceptable, in what point in time can we restore too

Interesting By Not Important for Associate Solutions Architect exam

- AWS Data Lake: https://d0.awsstatic.com/whitepapers/Storage/data-lake-on-aws.pdf

TODO:

- [ ] Study
- [ ] Finish reading notes
- [ ] Finish reading AWS FAQs & AWS White Papers
- [ ] Review ACloudGuru forums for exam tips
- [ ] Review AWS Cheat Sheets: https://tutorialsdojo.com/aws-cheat-sheets/
- [ ] Complete ACloudGuru exercise/setup for public/private VPC
- [ ] Sit ACloudGuru exam
- [ ] Sit Multiple WizLabs practice exams
- [ ] Complete AWS exam readiness course: https://www.aws.training/training/schedule?courseId=10006
- [ ] Book AWS exam