|  |
| --- |
|  |

Support levels: Enterprise, Business, Developer

Premium support: Basic, Developer, Business, Enterprise

Max response time for Business level Premium support case: 1 hour

Resource Groups: collection of resources that share one or more tags. labelling purposes and helps managing, organizing resources, used for Cost allocation to categorize and track and Conditional Access Control policy to define permission to allow or deny access on resources based on tags

Links:

BluePrint and stuff

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

<https://aws.amazon.com/certification/certification-prep/>

Certification Account:

<https://www.certmetrics.com/amazon/>

Faq:

<https://aws.amazon.com/faqs/>

Exam Simulator:

<https://acloud.guru/exam-simulator/start?courseId=aws-csa>

Blogs, services, tips:

<http://jayendrapatil.com/>

<http://jayendrapatil.com/tag/cheat-sheet/>

<https://digitalcloud.training/>

left off at Storage/EFS

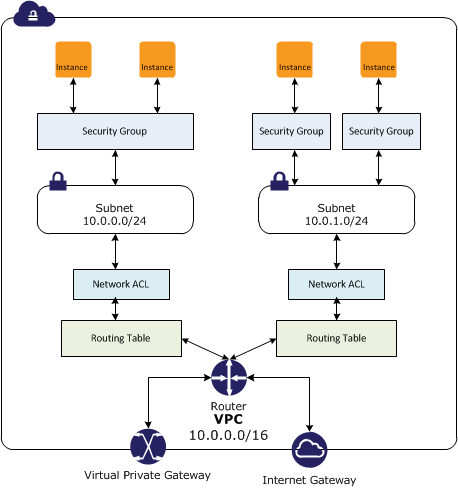
https://linuxacademy.com/amazon-web-services/training/course/name/aws-certified-solutions-architect-associate-level

Most(4-6) questions on: S3, EC2, VPC, Route 53, RDS, KMS, IAM, ELB (NLB + ALB), Auto scaling

Few (2-4) questions on: Aurora, DynamoDB + DAX, Lambda, API GW, CloudFront + OAI, CloudWatch, SQS, Cognito, EBS, EFS, PrivateLinks (VPC endpoint)

Least (1-2) questions on: CloudFormation, SNS, ElasticCache, Redshift, ECS, Elastic Beanstalk, Systems Manager, DirectConnect, VPN, Kinesis

# VPC


      Using a default VPC
     

per vpc

5 EIPs

5 VPC per region (request more)

1 IGW per VPC

200 subnets per VPC

500 SG per VPC

200 NACLs per VPC

CIDR: AWS reserve both the first 4 and last 4 IP addresses in each subnet’s CIDR block. /28 is the smallest possible subnet in an AWS VPC and /16 is the largest

Default VPC includes an Internet gateway and each default subnet is a public subnet in each AZ. Each instance that you launch into a default subnet has a private IP address and a public IP address

if a subnet in the a non-default VPC is not explicitly associated with a routing rule it uses the main rule set which does not have a route to the internet.

You can allocate your own EIP, and associate it to your instance after launch. Can move EIP to running instances in different AZs (same region) for fault tolerance with ELB. .

Every **non default VPC** comes with a default SG to which newly launched instances is associated in case no other security group had been specified for them. Allows no inbound traffic and all outgoing traffic. It can’t be deleted

Also created a default NACL

Also creates a default route table with only a local route setup

No subnets, IGW, NAT Gateway are created

SGs supports only allow rules. ACL supports allow and deny

SG evaluates all rules before deciding whether to allow traffic. ACL process rules in number order when deciding whether to allow traffic

SG are stateful,: return traffic is automatically allowed, regardless of any outbound rules. A network ACL and traffic is stateless: Return traffic must be explicitly allowed by rules.

Need to allow both inbound and outbound traffic in order for EC2 instances in a network ACL to be able to communicate over a particular protocol.

SGs can span across subnets and AZs!

SGs operate at the instance level

Default VPC vs Custom VPC:

All subnets are public in default and Each ec2 has private and public IP in default

You create new subnets in a custom VPC, by default they can communicate with each other, across availability zones

ONE route table per subnet (uses main route table by default). route table can be associated with MANY subnets

each subnet is mapped to one AZ and CANNOT span across multiple AZ

each subnet can communicate by default with the others in the VPC

then build the IGW and Attach to VPC

then need to create route table to connect to the IGW

so add a route in the table with target the IGW THEN associate subnets to the route. HAS TO BE ROUTED TO THE IGW to connect to internet

*VPC endpoint: end points to s3 buckets:*

private connection between your VPC and another AWS service without requiring access over the Internet, a NAT device, a VPN connection, or AWS Direct Connect.

## Gateway Endpoints

**https://docs.aws.amazon.com/vpc/latest/userguide/vpce-gateway.html**

are virtual devices.

private connection between your VPC and another AWS service using its private IP address. Does not leave the AWS network

oes not require a public IP address, access over the Internet, NAT device, a VPN connection or AWS Direct Connect

Gateway Endpoints are single region only supporting S3 and DynamoDB currently

*Scenario for public instance:*

create the endpoint to s3: this will add a route to the route table that we use for our instance pointing to the endpoint GW

change the SG of control instance to sshSG that only allows inbound ssh and outbound nothing. This will REMOVE the access to S3

add an outbound rule to this SG to allow HTTPS ( s3 uses that) to the s3 service prefix name. can get that from the [describe-prefix-lists](https://docs.aws.amazon.com/cli/latest/reference/ec2/describe-prefix-lists.html) command or

look at the route the endpoint created

rule: HTTPS TCP 443 pl-63a5400a

Once a VPC is set to dedicated hosting, it is not possible to change the VPC or the instances to Default hosting. You must re-create the VPC.

----------------------

NAT

need an SG that allows public to private inbound and outbound rules. So the SG needs rules to allow the private subnet ( source) access to http, etc inbound and http, etc outbound to the same ports

need to configure the main route table or a custom route table for the nat VPC: add route for all traffic to target the NAT instance with destination 0.0.0.0/0

the SG of the instance in the private subnet must have a rule in the SG that is assigned to allow inbound port 80, etc with the source being the private subnet CIDR or anywhere

Now a days its all NAT gateways and **easy**:

* Create the NAT gateway on the public subnet( creates an EIP)
* Add a route in the default route table that your private sub uses and add public (0/0) to the nat gateway

They are not associated with SGs like NAT instances and you want them in multil AZs for redundancy. They are more secure than NAT instance

BASTION:

Deploy a Bastion host within each Availability Zone for HA

allows you to login to instances in the Private subnet securely without having to store the private keys on the Bastion host (using ssh-agent forwarding or RDP gateways)

all the Instances in the private subnet should be hardened to accept SSH/RDP connections only from the Bastion host

Bastion and NAT should be in the public subnet. As Web Server has direct access to Internet, the subnet should be public and a Route pointing to IGW. Another route for private subnets should point to NAT for outgoing internet access

ACL:

numbered list of rules that evaluate in order, starting with the lowest numbered ( LOWER HAS PRECEDENCE) , to determine whether traffic is allowed in or out of any subnet associated with the network ACL.

Each network ACL includes a rule \*. This rule ensures that if a packet doesn't match any of the other numbered rules, it's denied. You can't modify or remove this rule.

The VPC default ACL allows ALL inbound and outbound unless you change it

have to associate subnets with custom ACLS!

when you create your own ACL ALL is DENIED!

when associate a subnet to the ACL it will dissociate that subnet from default ACL that has it associated

ACLs can be across multiple subnets but subnets can only have one ACL

A network ACL and traffic is stateless: Return traffic must be explicitly allowed by rules. Need to allow both inbound and outbound traffic in order for EC2 instances in a network to be able to communicate over a particular protocol

Egress-Only Internet Gateway" is to allow IPv6 based traffic within a VPC to access the Internet, whilst denying any Internet based resources the possibility of initiating a connection back into the VPC. t does not support IPv4 data communication nor does it allow a Security Group to be associated with it

VPC Flow Logs can be created at the VPC, subnet, and network interface levels

<https://docs.aws.amazon.com/vpc/latest/userguide/flow-logs.html>

AN ENI in one subnet can be attached to an instance in the same or another subnet, in the same AZ and the same VPC

## VPN


      Using a virtual private gateway
     

You can enable access to your network from your VPC with a VPN by attaching a virtual private gateway (each connection is secured by a pre-shared key in conjunction with the IP address of the customer gateway device) to the VPC, creating a custom route table, and updating your security group rules. customer gateway is a physical device or software application on your side of the VPN connection. Assign a public IP address on the customer gateway for the on-premise network. When you create a VPN connection, the VPN tunnel comes up when traffic is generated from your side of the VPN connection. If your VPN device supports BGP, specify dynamic routing when you configure your VPN connection. If your device does not support BGP, specify static routing. When you use a BGP device, you don't need to specify static routes to the VPN connection because the device uses BGP to advertise its routes to the virtual private gateway. Best practice to setup redundant VPN connections: Each VPN connection has 2 tunnels, with each tunnel using a unique virtual private gateway public IP address. It is important to configure both tunnels for redundancy. When one tunnel becomes unavailable (for example, down for maintenance), network traffic is automatically routed to the available tunnel for that specific VPN connection. To protect against a loss of connectivity in case your customer gateway becomes unavailable, you can set up a second VPN connection to your VPC and virtual private gateway by using a second customer

gateway.

New service: Private Link

Roles and VPC endpoints (PrivateLinks are preferred instead of Access keys.


     Using an interface endpoint to access an AWS service
    

---------------

The maximum number of Security groups per network interface an instance in a VPC can belong to is 5 (changed ?? )

## vpc peering

can connect VPCs to each other or private network using each other’s private IP. Can create connection cross account and cross region. Not transitive, always 1 to 1. Max of 50 VPCs. No single point of failure because it uses the existing AWS network infrastructure.

Instances behave as they were on one VPC star configuration. Can do cross account peering as well. Can’t overlap CIDR blocks.

Cannot do edge to edge routing:

If either VPC in a peering relationship has one of the following connections, you cannot extend the peering relationship to that connection:

A VPN connection or an AWS Direct Connect connection to a corporate network

An internet connection through an internet gateway

An internet connection in a private subnet through a NAT device

A VPC endpoint to an AWS service; for example, an endpoint to Amazon S3.

software based VPN fully customer managed. Could be a bottleneck

difference between Hardware IPsec and direct connect. direct connect 1 or 10 GB speed

iPsec appliance uses internet so no guarantee of performance. cloudHub goes on top of these. Spoke and hub

## Direct Connect

[**https://docs.aws.amazon.com/directconnect/latest/UserGuide/Welcome.html**](https://docs.aws.amazon.com/directconnect/latest/UserGuide/Welcome.html)

1 or 10 GB speed using industry standard 802.1q VLANs, this dedicated connection can be partitioned into multiple virtual interfaces.

links your internal network to an AWS Direct Connect location over a standard Ethernet fiber-optic cable using a Virtual Private Gateway. One end of the cable is connected to your router, the other to an AWS Direct Connect router. With this connection, you can create *virtual interfaces* directly to public AWS services (for example, to Amazon S3) or to Amazon VPC, bypassing internet service providers in your network path while maintaining network separation between the public and private environments

An AWS Direct Connect location provides access to AWS in the Region with which it is associated. You can use a single connection in a public Region or AWS GovCloud (US) to access public AWS services in all other public Regions.

Data xfer billed at lower rate. Reduce latency while reducing cost


    AWS Direct Connect
  ­

# Storage

## S3

Can have 100 buckets per account

object: <https://s3.amazonaws.com/bucket-name/object-name> or

bucket URLS:

<https://s3-eu-west-1.amazonaws.com/acloudguru1234>

http://mynewbucket.s3-aws-region.amazonaws.com

hosted web site: <http://s3-aws-stuff.s3-website-us-east-1.amazonaws.com> or you can use RT53 alias record. If you use custom domain name, bucket must use that same name

Data transferred to amazon EC2 from Amazon S3 in the same region does not cost

CORS: on s3 hosted web site a browser would normally block JavaScript from allowing those requests, but with CORS, you can configure your bucket to explicitly enable cross-origin requests

read after write for consistency for puts of new objects   
eventual consistency for overwrite puts and deletes (will be eventually consistent across facilities)   
can lose 2 AZs and still recover  
cross region replication new objects and new VERSIONs will replication from source to target bucket. Use an IAM role for this with a policy attached  
life cycle mgmt- can move objects to other tiers after so many days.

S3 standard storage provide 99.99% availability  and 99.999999999% durability and to sustain the concurrent loss of data in two facilities.

S3-IA (infrequent access) tier, cheaper, charges per retrieval retrieve right away

~~Store the images on S3 RRS, and create a lifecycle policy to delete the image after 24 hours~~ ( DEPRECATED). Replaced with One Zone IA. costs 20% less than S3 Standard-IA with slight less availability than standard IA

ideal for customers who want a lower-cost option for infrequently accessed data but do not require the availability and resilience of S3 Standard or S3 Standard-IA. It’s a good choice for storing secondary backup copies of on-premises data or easily re-creatable data

To retrieve data from Glacier, you initiate a retrieval job via S3 REST API or can use the S3 console.

E.g: aws s3api restore-object

Once it completes, your data will be available to download for **24 hours**

* Expedited is 1-5 minutes retrieval (most expensive)
* Standard is 3.5 hours retrieval (cheaper, 10GB data retrieval free per month)
* Bulk retrieval is 5-12 hours (cheapest, use for large quantities of data)

S3 Storage Classes: <https://aws.amazon.com/s3/storage-classes/?nc=sn&loc=3>

Can retrieve up to 5% of your average monthly storage, pro-rated daily, for free each month. Retrieve data using http get request. It automatically encrypts the data using AES256 and stores it durably in an immutable form

Vault names must be unique within an account and the region

Full S3 $0.023 per GB for the lowest band. S3 standard IA is $0.0125 per GB, S3 OneZone-IA is $0.01 per GB, and Legacy S3-RRS is around $0.024 per GB for the lowest band. Of the offered solutions S3 One Zone-IA is the cheapest suitable option. S3 has an availability of 99.99%, S3-IA has an availability of 99.9% while S3-1Zone-IA only has 99.5%

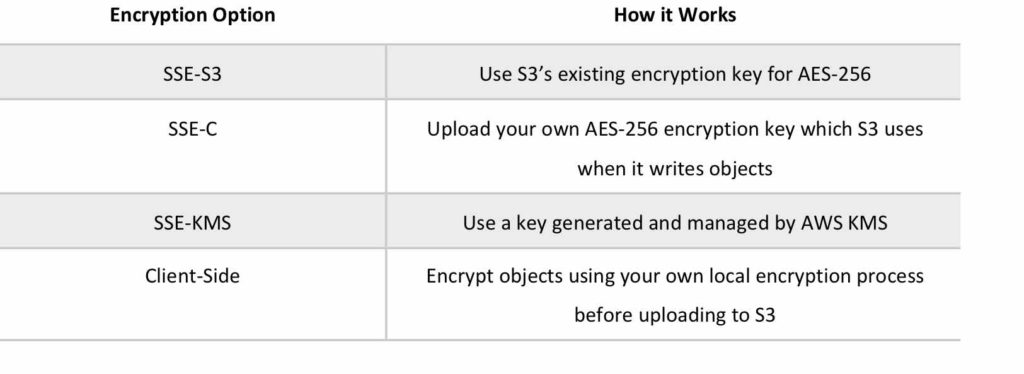
The valid encryption on S3 are Server Side Encryption at rest are S3 managed keys ( SSE-S3), encryption with keys stored in S3’s key management service (SSE-KMS) which is a bit more secure and has logging/audit features, and encryption using customer-provided and managed keys (SSE-C),  or a Client side client library such as Amazon S3 Encryption Client.

server-side encryption: S3 encrypts the object data before saving it on disks in its data centers and decrypts it when the user downloads the objects

Client-side encryption refers to encrypting data before sending it to Amazon S3. You have the following two options for using data encryption keys: Use an AWS KMS-managed customer master key and Use a client-side master key

You can Import your own keys, disable and re-enable keys and define key management roles in IAM

You can view the default and custom keys in the KMS console



data is spread across multiple devices and facilities(AZs)

5TB limit per object, 0 byte min, unlimited storage. A single PUT limited to 5GB

Charges are storage, requests, data xfer

S3 is a key (name) value(data) store

Event Notifications enable you to send alerts or trigger workflows. Notifications can be sent via SNS or SQS or to a Lambda function (depending on the bucket location).

Version ID: all objects have an ID, so all versions of a file take up X space

delete the "delete marker" file to restore after versioning turned on

restore later versions to go back to previous versions

S3 also can use MFA

create rules under LifeCYcle section to send shit to other tiers

have to wait 30 days before sending to IA. files of at least 128k, and 30 days after that for glacier

Can now use MFA for delete for: Change the versioning state of your bucket and Permanently delete an object version

*secure:*

bucket policies

You grant permission to an AWS account by the email address or the canonical user ID

Access Logs log to another bucker or even another account

Ways to control access to buckets: IAM policy, ACL, bucket policy, and query string authentication: allows them to share Amazon S3 objects through URLs that are valid for a predefined period of time

*Transfer acceleration*

Fast, easy, and secure transfers of files over long distances

Leverages Amazon CloudFront’s globally distributed AWS Edge Locations

To use, need to enable transfer acceleration on the S3 bucket

## Import/Export

allows for the import of large data sets, using external hard disks which are sent directly to amazon

disk: move big data sets into cloud with portable storage devices using amazon networks no internet

max 16TB device to send to AWS

import to EBS, S3, glacier and export From only S3

for transfers to S3, you specify the specific buckets to which the data should be loaded and ensure that the account doing the loading has write permission for the buckets. You should also specify the access control list to be applied to each object loaded to Amazon S3.

For transfers to EBS, you specify the target region for the EBS import operation. If the storage device is less than or equal to the maximum volume size of 1 TB, its contents are loaded directly into an Amazon EBS snapshot. If the storage device’s capacity exceeds 1 TB, a device image is stored within the specified S3 log bucket. You can then create a RAID of Amazon EBS volumes and copy the image from S3 to this new volume.

snowball: a device from amazon where u put your shit up to 50TB. uses 256 bit encryption, export and import s3

Snowball Edge: run lambda functions

SnowMobile: 100 TB in a fucking truck

## Storage Gateway

*File Gateway* – Flat files stored directly on S3 retain a copy of frequently accessed data subsets locally like gateway cached

Connect on premise appliance to AWS you can download the storage GW software appliance.

*Gateway-cached* ( where cloud is primary) volumes allow you to utilize Amazon S3 for your primary data, while retaining some portion of it locally in a cache for frequently accessed data. These volumes minimize the need to scale your on-premises storage infrastructure, while still providing your applications with low-latency access to their frequently accessed data. You can create storage volumes up to 32 TBs in size and mount them as iSCSI devices from your on-premises application servers.

*Gateway-stored* (where cloud is backup) volumes store your primary data locally, while asynchronously backing up that data to S3 over SSL as EBS snaps. These volumes provide your on-premises applications with low-latency access to their entire datasets, while providing durable, off-site backups. You can create storage volumes up to 1 TB in size and mount them as iSCSI devices from your on-premises application servers.

*Virt Tape Lib*: collection of virtual tapes backed by s3 or glacier. used for apps like netbackup etc. with up to 10 virtual tape drives per gateway, 1 media changer and up to 1500 virtual tape cartridges. Each virtual tape drive responds to the SCSI command set, so your existing on-premises backup applications (either disk-to-tape or disk-to-disk-to- tape) will work without modification. VTL shelf goes to glacier

## EFS

storage capacity is elastic, growing and shrinking automatically as you add and remove files, so your applications have the storage they need, when they need it. pports the Network File System version 4.1 (NFSv4.1) protocol, so the applications and tools that you use today work seamlessly with Amazon EFS. pay only for the storage used by your file system. EFS file systems store data and metadata across multiple Availability Zones in a region and can grow to petabyte scale, drive high levels of throughput, and allow massively parallel access from Amazon EC2 instances to your data.

# IAM

<https://685804732721.signin.aws.amazon.com/console>

IAM objects are global. Key pairs are NOT IAM objects

By Default a new user cant do anything, cant even change his pw

access keys and secret are only used for API an CLI.

attach policies to groups

SO when we assign the role to the new EC2 instance (s3 admin). NO NEED TO RUN aws configure DONT NEED TO store sec creds

roles easy to manage. IAM user, groups, roles are universal, use in any region. cant add users to roles can only assume roles. Eg need to change s3 access for 50 users: Use the IAM groups and add users as per their role to different groups and apply the policy to group

*AD integration:*

The user navigates to ADFS webserver, 2) The user enters in their single sign on credentials, 3) The user's web browser receives a SAML assertion from the AD server, 4) The user's browser then posts the SAML assertion to the AWS SAML end point for SAML and the AssumeRoleWithSAML API request is used to request temporary security credentials. 5) The user is then able to access the AWS Console

<http://docs.aws.amazon.com/IAM/latest/UserGuide/images/saml-based-sso-to-console.diagram.png>

Paths in IAM have informative purposes only , thus no boundaries are enforced between users or groups based on their paths

If u want URL of the AWS IAM sign-in page to have a company name instead of the AWS account ID, can create an alias for his AWS account ID.

Each user in Amazon IAM can be part of no more than 10 groups.

Power user: access to all AWS services except the management of groups and users

*Credential report*: CSV lists all users in your account and the status of their various credentials, including passwords, access keys, MFA devices, and signing certificates

Walk through having a different user use the S3 instance role and attach to an EC2 with no other rights

# EC2

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html>

curl http://169.254.169.254/latest/meta-data/

curl <http://169.254.169.254/latest/user-data>

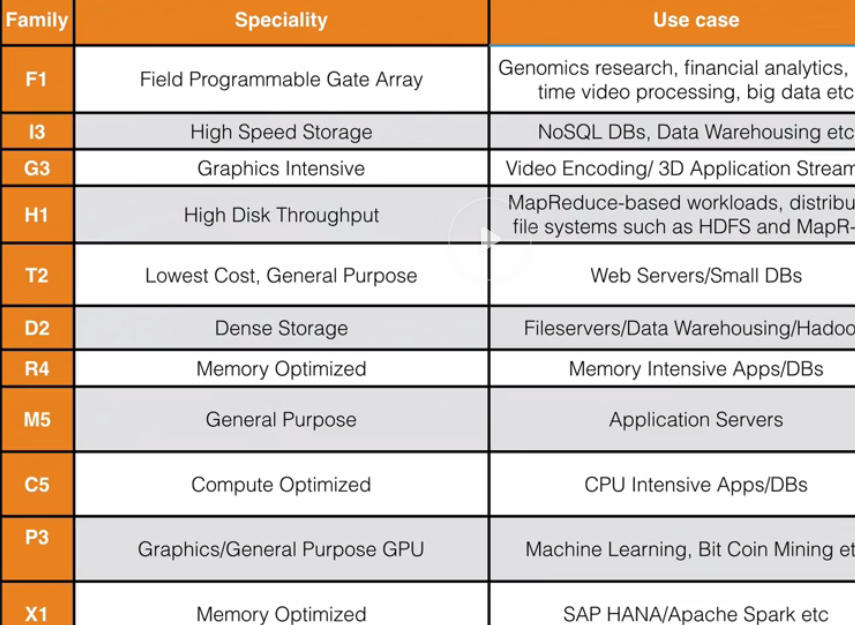
<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-metadata.html>

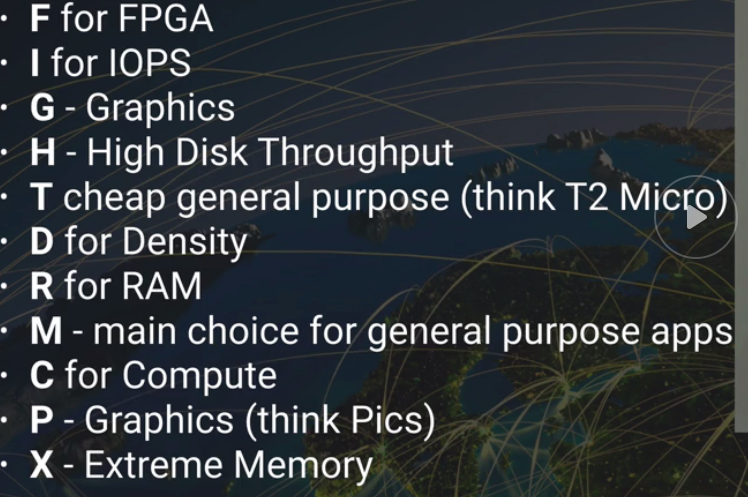
logs go here for user data: /var/log/cloud-init.log

HyperVisor: Xen and now Nitro

Get charged for EIP if not associated with a running instance, or if it is associated with a stopped instance or an unattached network interface

Instance types:





## EBS

Vol Types:

General SSD GP2, 99.999% avail

Increase volume size, adjust performance, or change the volume type while the volume is in use

Provisioned IOPS for IO intensive server. For applications that require sustained IOPS performance, or more than 16,000 IOPS or 250 MiB/s of throughput per volume

Study this table: [snap](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html)

Know the vol size ranges

You can now encrypt root volumes a few different ways.

EBS optimized: optimized configuration stack and provides additional, dedicated capacity for Amazon EBS I/O. certain instance types support it

When you create an EBS volume in an AZ, it is automatically replicated within that zone. When you create an EC2 instance, a **public** snapshot is created for the root volume. When you terminate the EC2, **that snapshot remains**!

Creating a volume manually does not create a snapshot by default.

Can attach a volume to any instance provided they are both in the same AZ.

Snapshot will asynchronously copy the data modified on the EBS volume. It does not copy the whole volume or all the data written to it, but just the modified one. For the first time it will copy all the data written on the volume and not copy the blank / empty blocks.

Snapshots are stored in S3 and you can copy it from one AWS region to another and encrypt while copying

Snapshot of a volume is available across all the Availability Zones within a region and you can use an Amazon EBS snapshot to create one or more new Amazon EBS volumes in any Availability Zone in the region

Recommended to stop instance if creating snap of the root device but not mandatory.

Create an encrypted volume from a snapshot of another encrypted volume. Create an encrypted snapshot from an unencrypted snapshot by creating an encrypted copy of the unencrypted snapshot.

RAID

amazon HATES RAID5 and RAID6. Parity writes cause bad performance.

RAID10 - striped and mirror, good performance

snapshot of RAID array 3 ways to flush cache:

-freeze file system

-unmount it

-shut down the instance

## AMI

If you create image from instance it will create a snapshot and stored in s3

Delete any key info from public AMI, stop bad services, etc., certs, accounts etc.

Get rid of key files, bash history

AMIs are regional but you can copy to other regions

Must deregister AMI before deleting root device

## EC2 Instance Info and Options

The keys that EC2 uses are 2048-bit SSH-2 RSA keys

There is a soft limit of 20 instances per region for new accounts.

*On demand* - fixed rate by the hour

*Reserved* - 1 or 3 year term. Capacity reserve, nice discount for paying up front

It is possible to transfer a reserved instance from one Availability Zone to another but NOT region

You can terminate reserved instance to avoid any data transfer charges that the instance might incur and sell the reserved instance in the AWS Reserved Instance Marketplace to recuperate cost.

2 Types: Standard and Convertible: <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/reserved-instances-types.html>

*SPOT:* set bid price for what u want to pay and wait for the SPOT price. No guarantee. When spot price is higher than bid price you lose instance. if your instance is terminated, you won’t be charged for that hour.

If u terminate, you do pay for the hour. 2 minute warning if the instance is to be terminated to save any unsaved work

## Dedicated Hosts

Instance runs on a Dedicated Host, an isolated server with configurations that you can control. Complete isolation

Cannot change the tenancy of a default (shared) instance after you've launched it. You can change the tenancy of an instance from dedicated to host after you've launched it, and vice versa (must stop instance first) .

Each VPC has a related instance tenancy attribute. You can't change the instance tenancy of a VPC after you create it.

Difference between a Dedicated Host and a dedicated instance is that a Dedicated Host gives you additional visibility and control over how instances are placed on a physical server, and you can consistently deploy your instances to the same physical server over time. As a result, Dedicated Hosts enable you to use your existing server-bound software licenses and address corporate compliance and regulatory requirements.

## Instance store

cant stop/start the instance, reboot or terminate only.

root device for instance launched from the AMI is an instance store volume created from template stored in s3.

if host fails, u lose the data, cant move the root device. Can share unencrypted snapshots only

faster then EBS

ideal for temporary storage of information that changes frequently, such as buffers, caches, scratch data

## Placement group

2 types: Clustered and Spread: placed on distinct underlying hardware and span multiple AZ, and you can have a maximum of seven running instances per AZ per group

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

Logical grouping of instances. Traditional Clustered groups only in a single AZ

Enables applications to participate in a low-latency, 10 Gbps network. Placement groups are recommended for applications that benefit from low network latency, high network throughput, or both

Good for stuff like Hadoop or Casandra

Only certain types of AMI can be in placement group

Recommend homogenous instances

Cant merge placement groups

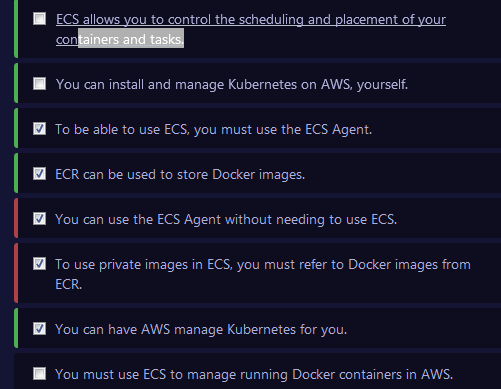
~~can’t move existing instance into placement group~~

update: Existing instance can now be moved to a placement group, or moved from one placement group to another, or removed from a placement group, given it is in the stopped state**.**

can span across peered VPCs for the same AZ

## ECS

No additional charge for Amazon ECS. You pay for AWS resources (e.g. EC2 instances or EBS volumes) you create to store and run your application



## Auto Scaling


   An illustration of a basic Auto Scaling group.
  


    The lifecycle of instances within an Auto Scaling group.
   

Scaling groups cannot span multiple regions only AZs in a region. You can create a launch config from an existing EC2 instance ( instance settings -> Attach to As..) . You can't modify a launch configuration after you've created it. However, you can change which launch configuration is associated with an Auto Scaling group. specify launch configuration, minimum num of instances, and max number of instances. You can optionally specify a desired capacity, which is the number of instances that the group must have at all times. If you don't specify a desired capacity, the default desired capacity is the minimum number of instances that you specified. Cool down is period of time which AS ignores scaling actions default is 300 secs( 5 mins) . should have two policies, one for **scaling in** (terminating instances) and one for **scaling out** (launching instances), for each event to monitor. example, if you want to scale out when the network bandwidth reaches a certain level, create a policy specifying that Auto Scaling shouldn’t start a certain number of instances. But you may also want an accompanying policy to scale in by a certain number when the network bandwidth level goes back down

when creating AS group, choose multiple subnets so you span across multiple AZs

Scaling Options:

* Maintain – keep a specific or minimum number of instances running
* Manual – use maximum, minimum, or a specific number of instances
* Scheduled – increase or decrease the number of instances based on a schedule
* Dynamic – scale based on real-time system metrics (e.g. CloudWatch metrics)

By default, an Auto Scaling group determines the health state of each instance by periodically checking the results of the EC2 instance status checks but you can instead associate an ELB health check.

Health check grace period: The length of time that Auto Scaling waits before checking an instance's health status. The grace period begins when an instance comes into service

Simple scaling policies must wait for the cooldown period to expire after a scaling activity or health check replacement before they can respond to alarms that are breached.

If an Auto Scaling group is launching more than one instance, the cool down period for each instance starts after that instance is launched. The group remains locked until the last instance that was launched has completed its cool down period.

To use with ELB, remember to check “receive traffic from ELB “ when creating AS group

Know this: http://jayendrapatil.com/aws-auto-scaling-elb/

Proactive Cycle: scale out based on peak periods

Proactive Event-based: scale out anticipation of increased demand

Demand: scale out based on metrics: CPU, network, etc ( horizontal) use monitoring service such as cloud watch

AutoScaling attempts to distribute instances evenly between the Availability Zones

**Scaling out** rather then up means adding instances instead of increasing instance size

THE ELB SG must have port 80 open AND the instances launched in the AS group must have port 80 open. The SG is configured in the launch config

AS will automatically register targets in ELB target group if you are using AS with EIB

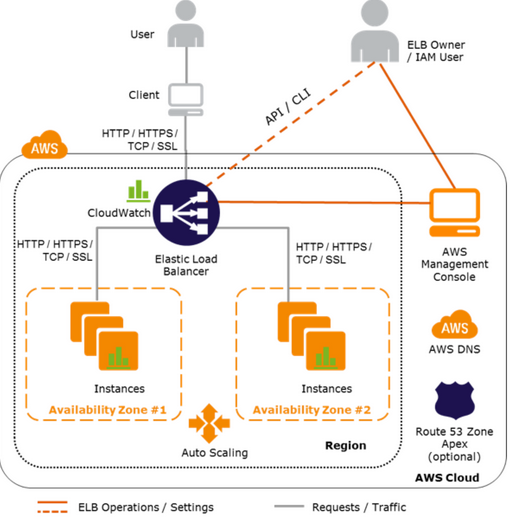
There is now AWS AS and EC2 AS: so can use with other service such as DynamoDB and Aurora

Example command:

as-put-scaling-policy scale-up-pol --auto-scaling-group my-as-group --adjustment=1 --type ChangeInCapacity --cooldown 300

consist of an LC that defines instances that are created in response to demand and an AS group that defines when to use an LC to create new instances and in which AZ and and ELB context they should be created in.

Cooldown is period of time which AS ignores scaling actions



## ELB

A distributed system that is fault tolerant and actively monitored by AWS and scales it as per the demand

Given a DNS host name – any requests sent to this host name are delegated to a pool of EC2 instances for multi AZ (Cross Zone load balancing to help route traffic evenly across all EC2 instances regardless of the AZ) and single region

With ALB (layer 7) and Network Load Balancers (Layer 4), you register **targets in target groups**, and route traffic to the target groups synchronously. With Classic Load Balancers, you register instances with the load balancer.

ALB has better app support. Web sockets, sticky sessions, health checks, path based routing. CAN NOT do TCP/IP requests. Better logging and monitoring. More expensive than classic

Sticky Sessions:

Session affinity to bind a user’s session to a specific application instance

It is not fault tolerant, if an instance is lost the information is lost

Requires HTTP/HTTPS listener and does not work with TCP

Requires SSL termination on ELB as it users the headers

ELB gives you a single DNS name for addressing and AS ensures there is always the right number of healthy Amazon EC2 instances to accept requests

Because the set of IP addresses associated with a LoadBalancer can change over time, you should never create an "A" record with any specific IP address.

If you want to use a friendly DNS name for your load balancer instead of the name generated by the Elastic Load Balancing service, you should create an ALIAS record for the LoadBalancer DNS name

Health checks: default checks every 30 seconds 10 times = 300 seconds ( 5 mins)

Unhealthy threshold: number of failures before declaring unhealthy. it will check the path that many times

Healthy threshold: # of checks before declaring healthy'

Response time out: ttl when receiving response from health check'

Health check interval: time between health checks so if you change the interval to 10 it cuts down the time to declare it dead

Change healthy threshold to 3 so it declares healthy faster

After setting up, make sure the instances u add are 'in service'

X-Forwarded-For request header takes the following form: X-Forwarded-For: clientIPAddress, previousRequestIPAddress, LoadBalancerIPAddress.

Supports load balancing of applications using HTTP, HTTPS, SSL and TCP

For SSL termination at backend instances or support for Client Side Certificate use TCP for connections from the client to the ELB, use the SSL protocol for connections from the ELB to the back-end application, and deploy certificates on the back-end instances handling requests

When you de register an instance, status will be “draining” to bleed off open requests

An ELB-Classic Load Balancer in an EC2-Classic (Legacy, nonVPC) environment it can have an associated IPv4, IPv6, and dualstack (both IPv4 and IPv6) DNS name, and supports IPv6 on the External/public interface. However inside a VPC IPv6 is not supported on the external or internal interface(s).

Should create a secure SSL listener and you will need an SSL cert:

<http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/elb-add-or-delete-listeners.html>

# CloudFormation

The only required section in template is resource. can see events for a deleted stack for up to 90 days.

# CloudTrail

captures AWS API calls and related events made by or on behalf of an AWS account and delivers log files to an Amazon S3 bucket that you specify. Optionally, you can configure CloudTrail to deliver events to a log group to be monitored by CloudWatch Logs. You can also choose to receive Amazon SNS notifications each time a log file is delivered to your bucket

# CloudFront

CDN: content cached at over 50 edge locations read and write. A distribution is a network of edge locations. CloudFront sends the distribution configuration to all the edge locations

origin is where the original content lives. Can be s3,ec2,elb,route53 or an on premise server.

If a local copy of the file is not there (at edge location), it will get a copy from the origin

https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/private-content-restricting-access-to-s3.html

URL per edge location where it caches it where it has a TTL

Origin server can be configured to limit access protocols, caching behavior, add headers to the files to add TTL

you can clear cached objects but you will be charged

Signed URLs and Signed Cookies are ways to ensure that users attempting access to files in S3 can be authorized. One method generates URLs and the other generates special cookies but they both require the creation of an application and policy to generate and control these items. **Origin Access Identity** is a virtual user identity that is used to give the CloudFront distribution permission to fetch a private object from an S3 bucket

after setup you can reference objects with the CDN url prefix e.g. dan2en6xtms8f.cloudfront.net

disabled distribution is no longer functional and Amazon stops billing it.

To remove an object from a cache, must invalidate it in the cloudfront console, API, or CLI

How to get to a CDN object by default: <http://d6nw2h20x5scj.cloudfront.net/IMG_1792.JPG>

**REVIEW the Cloudfront origin lab and setup OAI . setup an signed URL**

http://jayendrapatil.com/aws-cloudfront/

# Lambda

AWS Lambda lets you run code without provisioning or managing servers. You pay only for the compute time you consume - there is no charge when your code is not running.

AWS Lambda is a compute service that runs your code in response to events and automatically manages the underlying compute resources for you

node.js, java, python, GO, ruby, .NET, and PowerShell

NO SERVERS

CONTINUOUS SCALE

CHEAP

Scales out automatically

Can trigger other lambda functions, 1 event = x functions if functions trigger other functions

Global , can use to backup s3 bucket to bucket

Must know which services can trigger Lambda

**RUN THROUGH LAMBDA LAB LECTURES A FEW TIMES**

# CloudWatch

Choosing detailed monitor at ec2 creation switches from 5 mins to 1 mins but cost $

Memory is not available in this because it grabs hyperV stats. Metrics exist only in the region in which they are created.

Need 3rd party script for metrics: memory, disk space, swap utilization

U get a cpu credit bal that builds up if your utilization is low

Can’t use IAM to control access to CloudWatch data for specific resources. For example, you can't give a user access to CloudWatch data for only a specific set of instances or a specific LoadBalancer.

You can however specify which CloudWatch actions a user in your AWS Account can perform. For example, you could create an IAM policy that gives only certain users in your organization permission to use GetMetricStatistics

CloudWatch stores metrics for terminated Amazon EC2 instances or deleted Elastic Load Balancers for 15 months and alarm history for 14 days.

Alarm Actions:

* Recover the instance (only supported on specific instance types and can be used only with StatusCheckFailed\_System)
* Stop the instance (only applicable to EBS-backed volumes)
* Terminate the instance (cannot terminate if termination protection is enabled)
* Reboot the instance

# SNS

SNS has the following subscribers; Lambda, SQS, HTTPS, Email, Email-j**son**, SMS..  **Push** instant notifications to clients.  Pay as you go pricing Push to apple, google, android., etc Stored across multiple AZs. Can also deliver by SMS or email or SQS.

# Route53

Limit of 50 domain names but can be increased

No charge for route Internet traffic to CloudFront, Elastic Beanstalk, ELB, or S3.

zone apex" (sometimes called the "root domain" or "naked domain") to be an "A Record," can not be a CNAME. A R53 alias same as CNAME but can point to subdomains AND zone apex

zone apex record to point to the load balancer: A record aliased to the load balancer DNS name

Alias resource record sets can save you time because Amazon Route 53 automatically recognizes changes in the resource record sets that the alias resource record set refers to. because the set of IP addresses associated with a LoadBalancer can change over time, you should never create an "A" record with any specific IP address. Best to create an alias record to the ELB name

If an alias resource record set points to a CloudFront distribution, a load balancer, or an S3 bucket, the time to live (TTL) **can’t be set;** Route 53 uses the CloudFront, load balancer, or Amazon S3 TTLs.

Every request you make to the Amazon Route 53 control API must be authenticated.

--

*Routing Policies:*

<https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/routing-policy.html>

Simple: one record with multiple IP addresses

Failover: route traffic to a healthy resource or to a different resource when the first resource is unhealthy. Active/Passive using health checks

Geolocation: example, you might want all queries from Europe to be routed to an ELB load balancer in the Frankfurt region. Routes traffic based on originating location to specified region. Always have a default location!. Geolocation is based on national boundaries and will meet the needs well. Geoproximity is based on Latitude & Longitude

Weighted route policy: example: if you want to send a tiny portion of your traffic to one resource and the rest to another resource, you might specify weights of 1 and 255. The resource with a weight of 1 gets 1/256th of the traffic and the other resource gets 255/256ths . You can gradually change the balance by changing the weights. If you want to stop sending traffic to a resource, you can change the weight for that record to 0.

Latency based routing: example, you have web servers for example.com in data centers in Ireland and in Tokyo. When a user browses to example.com, Amazon Route 53 chooses to respond to the DNS query based on which data center gives your user the lowest latency. *e.g. if the latency from the user in Singapore to Ireland improves, the user can be routed to Ireland*

Multivalue answer: return multiple values, such as IP addresses for your web servers, in response to DNS queries. You can specify multiple values for almost any record, but multivalue answer routing also lets you check the health of each resource, so Route 53 returns only values for healthy resources. It's not a substitute for a load balancer, but the ability to return multiple health-checkable IP addresses is a way to use DNS to improve availability and load balancing.

--

CNAME record maps a name to another name. It should only be used when there are no other records on that name. can point to any DNS record hosted anywhere

ALIAS record maps a name to another name, but in **turn it can coexist with other records on that exact name**. Can only point to a CloudFront distribution, an Elastic Beanstalk environment, an ELB load balancer **OR** records you already created for that zone!!

CNAME charges for queries, not so with Alias record.

Common to create an A record [www.urdomain.com](http://www.urdomain.com) then create an alias to that using the naked domain name. so urdomain.com will point to www

There is also Geoproximity routing wtf

# DB

**DynamoDB**

*Eventually Consistent Reads by deault*: When you read data from a DynamoDB table, the response might not reflect the results of a recently completed write operation. The response might include some stale data. If you repeat your read request after a short time, the response should return the latest data.

*strongly Consistent Reads*: When you request DynamoDB returns a response with the most up-to-date data, reflecting the updates from all prior write operations that were successful. A strongly consistent read might not be available if there is a network delay or outage. Consistent reads are not supported on global secondary indexes (GSI)

uses SSD storage

creates the table across 3 AZs automatically and synchronously, u can’t select it. Scan operations are always eventually consistent

stores structured data in tables, indexed by primary key, and allows low-latency read and write access to items ranging from 1 byte up to 64 KB. supports three data types: number, string, and binary, in both scalar and multi-valued sets. Tables do not have fixed a schema, so each data item can have a different number of attributes. The primary key can either be a single-attribute hash key or a composite hash-range key. Local secondary indexes provide additional flexibility for querying against attributes other than the primary key. provides both eventually-consistent reads (by default), and strongly-consistent reads (optional), as well as implicit item-level transactions for item put, update, delete, conditional operations, and increment/decrement

pay only for what you use and there is no minimum fee. *3 pricing components*: provisioned throughput capacity (per hour), indexed data storage (per GB per month), data transfer in or out (per GB per month, However: no charge for transfer into DynamoDB, within a single region)

good for storing BLOB, JSON, and web session data

*Elasticache*: in memory cache cloud, good for repetitive queries. Uses memcached or redis.

REDSHIFT (DW)

OLAP: does analytics on large amount of records. Provides MPP

uses replication and continuous backups to enhance availability and improve data durability and can automatically recover from node and component failures

single node or multi-node: leadernode (manage client connect and receive querey) Compute nodes: store data and perform query,up to 128 compute nodes

columnar data storage. not like rows. easier to compress. doesn’t require indexes. SUPER FAST

Massive Parallel Processing: easy to load across nodes

pricing: billed 1 unit per node. billed for compute node hours. not leader node hours. pay for backup

and data xfer as well

encrypted transit SSL,encrypted as rest using AES-256

only available in 1 AZ. Can restore snapshot to new AZ

block size: 1024KB / 1MB

Analyze global sales data for multiple products

 Store historical stock trade data

 Analyze ad impressions and clicks

 Aggregate gaming data

 Analyze social trends

 Measure clinical quality, operation efficiency, and financial performance in the health care space

***RDS:***

RDS instances are regional can be copied to different regitions via snapshot or promoting read replica

supports MySQL, MariaDB, PostgreSQL, Oracle, SQL Server

Aurora is amazons RDS: 2 copies of your data contained in each AZ with min of 3 = 6 copies

can handle the loss of 2 copies data without affecting writes and 3 without affecting reads. storage is self healing, auto scanned and repaired

RDS (OLTP):: transaction record based queries

By default, the maximum provisioned IOPS capacity on an Oracle and MySQL RDS instance (using provisioned IOPS) is 30,000 IOPS.

For multi-AZ, Updates to your DB Instance are synchronously replicated across Availability Zones to the standby in order to keep both in sync and protect your latest database updates against DB Instance failure

There is no charge to replicate data, Multi-AZ can be turned on after DB is created. transaction success is returned only if the commit is successful both on the primary and the standby

to allow access to mysql port from the SG, have to do through VPC

Restore DB Instance (snapshot) : You are creating a NEW DB Instance(with new DNS endpoint) from a source DB Instance at a specified time. DB Snapshots are user-initiated.

By default and at no additional charge, RDS enables automated backups of your DB Instance with a 1 day retention period. backup retention max 35 days. n RDS automatically performs a full daily snapshot of your data to S3

To scale beyond the capacity constraints of a single DB Instance for read-heavy database workloads, Read Replicas. You can create a Read Replica of a given source DB Instance using the AWS Management Console or the CreateDBInstanceReadReplica API. Once the Read Replica is created, database updates on the source DB Instance will be propagated to the Read Replica asynchronously. You can create multiple Read Replicas for a given source DB Instance and distribute your application’s read traffic amongst them. Backups for replicas are not configured by default.

Can be cross AZ and region and can have 5 max RRs copies of a DB

mysql -h jimmydb.czlxgi6wj1oj.us-east-1.rds.amazonaws.com –P 3306 -u jimmy -p

nifty: instead of using the long ass RDS endpoint name, create internal zone and then create a CNAME to the endpoint: eg mysql.internal.-> CNAME wordpressdb.czlxgi6wj1o…

default DB subnet group is created for your account by default with the default VPC ONLY

DB subnet group is a collection of subnets (typically private) that you create in a VPC and that you then designate for your DB instances.

make sure create a DB subnet group in the same VPC as the EC2 SG. then use the EC2 SG as the source when creating rule for DB Subnet Group

upgrade DB instance: apply the changes immediately, choose Apply immediately. Choosing this option can cause an outage in some cases. For auto updates: an enable auto minor version upgrades. For modifications If you don't choose to apply changes immediately, the changes are put into the pending modifications queue. During the next maintenance window, any pending changes in the queue are applied

RDS automatically sends metrics to CloudWatch every minute for each active database: standard monitoring provides metrics ranging from CPU, memory, and disk usage, to more specific ones such as DB connections, I/O/latency/throughput

RDS encryption is done via KMS and SSL . once the instance is encrypted, so are the backups, replicas, and snapshots.

Walkthrough: creating DB instance and connecting:

1. create private Sec G in the VPC where your EC2s will connect. add rule for mysql port with the EC2 instance SG as the source

2. will ask for DB  subnet grp, leave default OR if you choose a VPC that does not have one need to create it. (Each DB subnet group should have subnets in at least 2 Availability Zones)

you can create the DBSG in non default VPC  manually and it has to cover at least 2 AZ.  If you don’t, the RDS console will create when u launch RDS instance and will add all subnets in the VPC to it

3. leave AZ no preference and launch it

4. now should be able to connect from EC2 instances from step1. each DB instance creates a default snap shot, delete it after delete instance

**Create and play with Arora and replica b4 test**

**EMR(Hadoop):**

utilizes a hosted **Hadoop** framework running on the web-scale infrastructure of EC2 and S3. launches all nodes for a given cluster in the **same Availability Zone**, which improves performance

supports Persistent and Transient cluster types

Cluster states are:

STARTING – The cluster provisions, starts, and configures EC2 instances.  
BOOTSTRAPPING – Bootstrap actions are being executed on the cluster.  
RUNNING – A step for the cluster is currently being run.  
WAITING – The cluster is currently active, but has no steps to run.  
TERMINATING - The cluster is in the process of shutting down.  
TERMINATED - The cluster was shut down without error.  
TERMINATED\_WITH\_ERRORS - The cluster was shut down with errors.

# SQS

Amazon SQS is ideal for multiple application components must communicate and coordinate their work in a loosely coupled manner. This occurs particularly in producer-consumer scenarios where some components may work faster or slower than others, or where the number of interacting components changes with time or load. serve as the “software glue” that enables components to communicate reliably without being tightly coupled or highly dependent upon synchronous operation, or on a fixed number of components. supports multiple readers and writers interacting with the same queue as the same time

Messages can contain up to 256kb of text in any format

Messages can be kept in que from 1 min to 14 days

Default retention period is 4 days

2 types of queues:

Standard: (Default) Unlimited transactions per second. Guaranteed Delivers message at least once but cannot guarantee it will not create duplicates and can’t guarantee order . It will make best effort to keep messages in order

FIFO: guarantee order and no duplicates. Messages are delivered once and is there until consumed. Limited to 300 trans/sec

To illustrate, suppose you have a number of image files to encode. In an Amazon SQS worker queue, you create an Amazon SQS message for each file specifying the command (jpeg-encode) and the location of the file in Amazon S3. A pool of Amazon EC2 instances running the needed image processing software does the following:

1. Asynchronously pulls the task messages from the queue (ALWAYS a pull based system, never push)

2. Retrieves the named file

3. Processes the conversion

4. Writes the image back to Amazon S3

5. Writes a “task complete” message to another queue

6. Deletes the original task message

7. Checks for more messages in the worker queue

The visibility timeout (where the message once read by a consumer is not visible to the other consumers till the timeout occurs) clock starts ticking once Amazon SQS returns the message. During that time, the component processes and deletes the message. But what happens if the component fails before deleting the message? If your system doesn't call [DeleteMessage](http://docs.aws.amazon.com/AWSSimpleQueueService/latest/APIReference/API_DeleteMessage.html) for that message before the visibility timeout expires, the message again becomes visible (is hidden after retrieve) to the [ReceiveMessage](http://docs.aws.amazon.com/AWSSimpleQueueService/latest/APIReference/API_ReceiveMessage.html) calls placed by the components in your system and it will be received again. If a message should only be received once, your system should delete it within the duration of the visibility timeout. Default visibility timeout is 30 seconds but can be increased to 12 hours max

'DelaySeconds' attribute: When a new message is added to the SQS queue, it will be hidden from consumer instances for a fixed period

When the consumer instance polls for new work, the SQS service will allow it to wait a certain time (WaitTimeSeconds) for one or more messages to be available before closing the connection.

While the regular SQS short polling returns immediately, even if the queue being polled is empty, SQS long polling polls periodically and doesn’t return a response until a message arrives in the queue, or the long poll times out. SQS long polling makes it easy and inexpensive to retrieve messages from your SQS queue as soon as they are available.

Enable long polling by setting the ReceiveMessageWaitTimeSeconds to a number > 0

Short does not query all the servers that the SQS messages can reside on, so multiple queries of the queue may be needed to retrieve all messages in the queue.

# SWF

TASK oriented NOT message oriented

Workers : app that can start a workflow. Like ordering from web site, they get tasks, process them, and return results

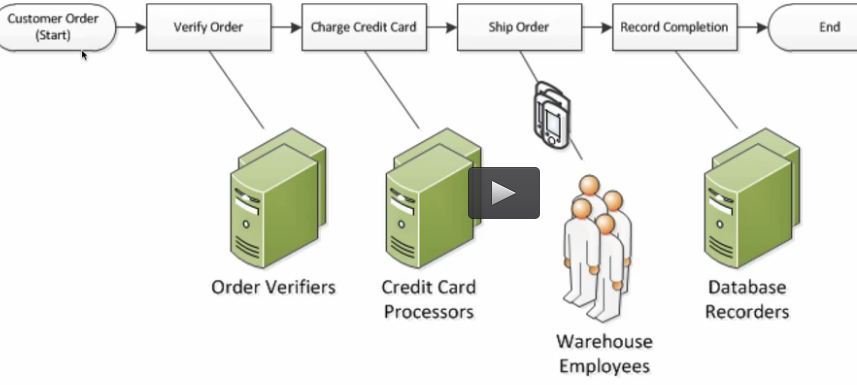
Deciders: control flow of activity. Verify address, cred card etc. tells the decider the state of the work flow and the next thing to be done

SWF **ensures that a task is assigned only once and is never duplicated** and guarantees order and has retention of 1 year for workflow executions

Uses long polling

Can be used with on prem servers

SWF Domains: collection of related work flows ( written in json)



# MISC

**Athena** is an interactive query service that makes it easy to analyses data in Amazon S3, using standard SQL. It will work with "JSON", "Apache Parquet", "Apache ORC" amongst others, but "XML" is not supported

**API Gateway:**

Same origin policy: web app permits scripts contained in one web site to access data in a second, ONLY if they are in the same origin ( domain) . Can use CORS to override this

Caching used to increase performance and can throttle it back if needed

**Kinesis**:

Streams: Streaming Data from lots of sources. Data records are only accessible for a default of 24 hours from the time they are added to a stream. Can increase to 7 days. Stored in Shards( partitions). The data is consumed by something like EC2 -> S3, Redshift, etc.

Provides ordering of records, as well as the ability to read and/or replay records in the same order

Firehose: data producers -> into firehose ( no shards, it does it all 4 u) , analyze the data( optional) then send to something like s3

Analytics: can run sql queries in firehose and streams then store in s3, redshift etc

**SES**: email service

**OpsWorks** = Chef

**White Paper:**

**Watch the summary lesson before exam**

<https://aws.amazon.com/architecture/well-architected/>

best practices:

<https://aws.amazon.com/whitepapers/>

5 pillars of well architected frame work: *Operational Excellence, Security, Reliability, Performance efficiency, Cost Optimization*

Scale out instead of up. Automated elasticity and scalability

*Security Pillar:*

Automate responses to security events

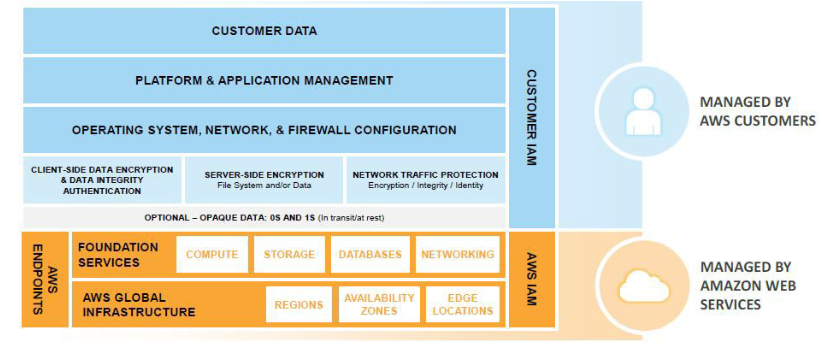
Security pillar applies to:

Data protection

Privilege management ( IAM)

Infrastructure Protection (VPC)

Detective Controls ( Cloud Trail)



*Reliability Pillar:*

Auto recover from failures

Scale out to increase system availability

Managing AWS account limits

Foundations (IAM, VPC)

Change Management ( CloudTrail)

Failure Management ( CloudFormation ?, backup and recovery )

*Performance efficiency Pillar:*

Compute ( choose correct type, or lambda, monitor instances, demand match quantity(AUTO SCALING)?

Storage:

Database: read replicas, caching, proximity, monitor proximity and caching.

Space/Time trade off: Cloudfront, elasticache, direct connect, RDS read replicas

*Cost Optimization Pillar:*

Trade capex for operating exp. Economies of scale

Match supply to demand ( AS)

Cost effective resources ( EC2) , Trusted Advisor

Expense aware ( cloudwatch, sns)

Optimizing over time ( trusted advisor,

*Operational Excellence Pillar:*

Preparation: AWS config good tool to use for inventory of resources and AWS service catalog.

Operations: COdecomit, codedeploy, pipeline, SDKs, cloudtrail

Responses: cloudwatch alarms, sns,

**Other:**

cloud HSM: Locating HSM appliances near your EC2 instances decreases network latency, which can improve application performance. Its an appliance which u get to via ssh or rdp

RTO is the time it takes after a disruption to restore a business process to its service level and RPO acceptable amount of data loss measured in time before the disaster occurs

**REDUNDANT SITE**

