**EC2 Exam tips** -

1) Different pricing models - On demand, Reserved, Spot and Dedicated hosts

2) On Demand - Allows paying a fixed rate by the hour or by the second with no commitment

3) Reserved - Provided with a capacity reservation and offers a significant discount on the hourly charge for an instance. 1 Year or 3-year terms.

4) Different types - Standard RI, Convertible RI, Scheduled RI.

5) Spot instance - enables you to bid whatever price you want for instance capacity and provides for even more savings if applications have flexible start and end times. If a Spot instance is terminated by AWS, then there will not be a charge for the partial hour used. However, if we terminate the instance ourselves, then we would be charged for the complete hour in which the instance ran. For example, if we bid 100$ and the price then goes to 150$, the instance is terminated by AWS.

6) Dedicated hosts - Physical EC2 server dedicated for our own use. Can help reduce costs by allowing us to access existing server bound software licenses. Very useful for regulatory requirements that do NOT permit multi-tenancy virtualization.

**7) Volume types** –

SSD -

General Purpose SSD -

Balances price and performance for a wide variety of workloads

Provisioned IOPS SSD -

Highest performance SSD volumes for mission critical low latency or high throughput for DB etc.

Magnetic -

Throughput Optimized HDD - Low cost HDD volume designed for frequently accessed, throughput intensive workloads.

Cold HDD - Lowest cost HDD volume designed for less frequently accessed workloads

Magnetic (Standard) - Previous generation. Can be a boot volume.

\*\* Throughput = Throughput refers to how much data can be transferred from one location to another in a given amount of time. For example, a hard drive that has a maximum transfer rate of 100 Mbps has twice the throughput of a drive that can only transfer data at 50 Mbps. \*\*

8) Termination protection is turned OFF by default.

9) On an EBS backed instance, the default action is for the root instance to be deleted when the instance is terminated. We can modify this by unchecking the box.

10) EBS root volumes of our DEFAULT AMIs cannot be encrypted. However, we can make a copy of them and we can ENCRYPT that copy.

We can use a third-party tool (like bit locker) to encrypt root volume or this can be done when creating the AMI in the AWS console or using the API.

**Placement groups** –

1. Clustered placement group can't span multiple AZ.

2) The name specified for a placement group should be unique within your AWS account.

3) If in exam they are referring to placement group, assume it is clustered placement group.

4)Only certain types of instances can be launched into a placement group - Compute Optimized, GPU, Memory Optimized, Storage Optimized)

5) AWS recommends using homogenous instances within placement groups. (Homogenous = Same size and family)

6) Cannot merge placement groups

7) We cannot move an existing instance into a placement group.

8) We can, however, create an AMI from an existing instance and then launch a new instance using that AMI into a placement group.

Overall EC2 summary -

1) Pricing models -

On-demand

Reserved

Spot (We terminate, we pay. AWS terminates, no charge for the hour it was terminated in)

Dedicated

2) EC2 instance types -

FIGHTDRMCPX

F - Field Programmable Gate Array

I - High Speed Storage (IOPS)

G - GPU optimized

H - High Disk Throughput

T - General Purpose

D - Dense Storage

R - Memory Optimized

M - General Purpose

C - Compute Optimized

P - Graphics/General Purpose GPU intensive

X - Memory Optimized

3) **EBS - Elastic Block Storage**

Optimized for runing DB,OS etc.

Types –

SSD, General Purpose - GP2 - Upto 10,000 IOPS (can be root vol)

Provisioned IOPS - IO1 - more than 10,000 IOPS (can be root vol)

Magnetic -

Throughput Optimized - ST1 - Frequently Accessed Workloads (cannot be root vol)

Cold- SC1 - less frequently accessed data (cannot be root vol)

Magnetic regular - cheap, infrequently accessed data (can be root vol)

4) Cannot mount an EBS volume to more than one EC2 instance. Instead, use EFS (which is also block based storage or S3, which is object based)

5) Termination protection is turned OFF by default.

6) On an EBS backed EC2 instance, the default action is foe the root EBS volume to be deleted when the instance is terminated.

7) EBS backed root volumes can now be encrypted using AWS API or the console. Also, third-party tools (like bitlocker) can be used to encrypt the root volume.

8) Any additional volumes can also be encrypted.

9) **Volumes vs Snapshots** –

Volumes = Virtual HDD in the cloud. These exist in EBS.

Snapshots = Point-in-time snapshots/images of volumes that are stored in S3.

Snapshots are incremental - This means that only the blocks that have changed since the last snapshot are moved to S3.

First snapshots always take some time to be created.

Snapshots of encrypted volumes are automatically encrypted.

Volumes restored from encrypted snapshots are encrypted automatically.

Snapshots can be shared but only if they are unencrypted.

These snapshots can be shared with other AWS accounts or made public.

10) It is very strongly recommended that you stop the instance when taking a snapshot of the root device volume.

11) EBS vs Instance Store -

Instance store volumes are called Ephemeral Storage

Instance Store volumes cannot be stopped. If the underlying host fails (the hypervisor), then we lose that data.

EBS backed volumes can be stopped. We will not lose data if it is stopped.

However, both EBS and Instance store can both be re-booted and no data will be lost.

By default, the root device volumes will be deleted on termination of instance. However, with EBS, we have the choice to configure the instance in such a way that the EBS volumes doesn't get deleted even after instance termination.

12) **Taking snapshots of RAID arrays** –

Problem - If we take a snapshot, the snapshot excludes the data held in the cache by applications and OS. This is not a problem for a single volume; However, in a RAID setup, this can be a problem owing to interdependencies of the array.

Solution - Take an application consistent snapshot:

Stop applications from writing to disk

Flush all caches to the disk

How can this be achieved?

Freeze the file system

Unmount the RAID array

Shutting down EC2 instance

**Security Groups** – Firewall

1) All inbound traffic is blocked by default.

2) We can only allow traffic using security groups.

3) All outbound traffic is allowed by default.

4) Changes to security groups takes place immediately.

5) We can have any number of EC2 instances within a security group.

6) We can have multiple security groups associated with one EC2 instance.

7) Security groups are STATEFUL. If we create an inbound rule allowing traffic in, that traffic is allowed back out by default.

8) Security groups are stateful. Network Access Control lists are stateless.

9) Cannot block specific IP addresses using security groups, instead use Network Access Control Lists. Security groups don’t have a deny or block option

10) We can specify ALLOW rules, specifying DENY rules is not possible using Security Groups.

**EBS** –

1) Volumes exist on EBS - Similar to Virtual HDD

2) Snapshots of EBS volumes created are stored in S3.

3) Snapshots are point-in-time copies of volumes.

4) Snapshots are incremental - Only the blocks that have changed since the last snapshot are moved to S3.

5) If this is the first snapsot, then it might take some time to be created.

6) To create snapshot for EBS volumes that service as root devices, it is a best practice to stop the instance before takin the screenshots (However, you can still take a snapshot when the instance is running).

7) We can create AMIs from both volumes and snapshots.

8) We can change the EBS volume sizes on the fly including changing the size and storage types (Again, the best practice is to stop the instance before taking snapshots).

9) Volumes will always be in the same AZ as the EC2 instance.

10) To move an EC2 volume from one AZ to another AZ/Region, take a snapshot or an image of it and then copy it to new AZ/Region.

**Differences between EBS volumes and ephemeral (Instance store) volumes**

Instance store volumes are also called Ephemeral Storage.

Instance store volumes cannot be stopped. If the underlying host fails, we will lose data.

Rebooting an Instance store volume preserves data. However, stopping or terminating them will destroy data.

EBS backed instances can be stopped and we will not lost data if the instance is stopped.

By default, both ROOT volumes will be deleted on termination, however, with EBS volumes, you can always instruct AWS to keep the root volume.

**Load Balancers -**

1) ALB, NLB, CLB (Application Load Balancers, Network Load balancers and Classic Load balancers)

2) 504 => gateway has timed out. Application is not responding within the idle timeout period. Troubleshoot application.

3) Use X-forwarded-for (XFF) header to get the IPv4 address of the end user for ALB or use some proxy service for classis ELB

ELB and Health checks -

1) Instances monitored by ELB are reported as -

In service or Out of service

2) Health Checks check the instance health just by talking to it

3) Elastic Loadbalancers have their own DNS names. We are never given an IP address for a load balancer. Amazon handles the DNS resolution on their own in the case of load balancers.

4) Read ELB FAQ for Load Balancers.

5) There is a deep dive course on A Cloud Guru platform.

**S3 -**

1) Object based. For files. Not for DB or OS.

2) 0B to 5 TB.

3) Unlimited storage.

4) Files are stored in buckets.

5) S3 is universal namespace. Hence, the bucket name should be unique globally.

6) Write to S3 - HTTP 200 return code for successful write (can be seen in CLI)

7) We can upload files to S3 much faster by enabling multipart uploads. Big files are broken into smaller ones and are then written to S3.

**S3 - Encryption** -

1) Bucket policies and Access control lists to secure buckets

2) Turn logging on to see what the bucket is being queried for (This log can be written to another bucket or another AWS account).

3) Types of encryption and how they work on a high level.

S3 - Static website hosting -

Sample use-case :

When you have a move preview and we don't know how many people are goign to watch it. So, we can add that video file to the static web page hosting bucket and S3 automatically scales according to the traffic. We need not worry about load balancing etc.

The URL names are exam question. Static website with S3 URL looks like this -

http://bucketname.s3-website-region.amazonaws.com

Example - http://acloudguru.s3-website-us-east-1.amazonaws.com

**Versioning** -

1) Stores all versions of object (including all writes and deletes)

2) Great backup tool

3) Once enabled, cannot be disabled, only suspended

4) Integrates with Life cycle rules

5) Has MFA delete capability

**Cross Region replication -**

1) Versioning MUST be enabled for CRR.

2) Regions MUST be unique (cannot use a different availability zone in the same region).s

3) Already existing files in the source bucket are not automatically replicated.

4) You cannot replicate to multiple buckets or use daisy chaining.

5) Delete markers are replicated.

6) Deleting individual versions or delete markers will NOT be replicated.

**Lifecycle management** -

1) Can be used in conjunction with versioning.

2) Can be applied to both current and previous versions.

3) Transfer to S3-IA or One Zone IA can occur only after 30 days.

4) The object MUST stay in S3-IA at least for 30 days before it can be moved to glacier. HOWEVER, there is no such 30 day restriction for objects to move from IA-OneZone to Glacier.

**Cloudfront** -

1) Edge location = This is the location where the content will be cached. This is different from AWS Region/AZ. They are both READ/WRITE (We can PUT object on Edge locations and they can transfer them back to the server).

2) Origin = The origin of all the files that the CDN will distribute. This can be anything - S3 bucket, an EC2 instance, an Elastic Load balancer or Route 53. This is basically where the original content is. It doesn't even have to be an AWS service (could be a private DC).

3) Distribution - This is the name given to the CDN that consists a collection of edge locations.

4) Objects are chached based on the specified TTL.

5) We can clear cached objects but we would be charged.

**Storage gateway** -

1) File Gateway (NFS)- For flat files (word, pdf, image, videos etc), stored directly on S3.

2) Volume Gateway -

Stored Volumes - Entire dataset is stored on-site and is asynchronously backed up to S3

Cached Volumes - Entre dataset is stored on S3 and the most frequently accessed data is stored on-site.

3) Virtual Tape Library (VTL) - Used for Backup applications like NetBackup, Backup Exec, Veeam etc.

4) Scenario based questions for exam should be expected.

5) If a financial firm has very time specific processing SLA and they should have very minimal latency, then go for stored volumes.

6) If a media firm has a LOT of videos being uploaded/read and they want to minimize some cost by not scaling their on-site infrastructure, then go for cached storage.

**Snowball** -

1) Understand what snowball is

2) Understand what Import/Export is (legacy - still an option but not used)

3) What snowball can do -

Import from S3

Export to S3

If the data is in glacier, restore it into S3 and then transfer it.

**CloudWatch** -

Standard Monitoring = 5 minutes

Detailed monitoring = 1 minute

Dashboards (Line, Number, Text, Stacked area) = Creates awesome dashboards to see what is happening with AWS environment

Alarms (Billing, Metric tresholds) = Allows setting of alarms that notify when particular tresholds are hit

Events = Cloud watch events that helps you to respond to changes in AWS RSS.

Logs (Log groups, log streams) = Helps to aggregate, monitor and store logs. We need to install log agent on EC2 server.

Difference between cloudwatch and cloudtrail -

Cloud watch = Logging and Monitoring

Cloudtrail = auditing

**Lambda** -

1) Lambda scales out (NOT up) automatically

2) Lambda functions are independent => 1 event = 1 function

3) Lambda is serverless

4) Understanding what services are serverless is important for exam (S3, API Gateway, Lambda, DynamoDB etc). EC2 is NOT serverless.

4) Lambda functions can trigger other lambda functions.

5) Architectures can get EXTREMELY COMPLICATED. AWS X-ray allows us to debug when there is a confusion/issue

6) Lambda can do things globally - We can use Lambda to backup S3 buckets to other S3 buckets etc.

**Route53** -

1) Elastic Load balancers don't have a ppre-defined IPv4 addresses. We have to resolve them using the DNS endpoint provided.

2) Difference between Cname and Alias record. When we are making a request to route53 for a DNS record, we are charged for that request if we are using Cnames. However, if we use alias record instead of cname, we are NOT going to be charged for that.

3) ELB do not have pre-defined IPv4 addresses, you resolve to them using DNS name.

4) Difference between Cname and Alias record. When presented a choice in exam, go for Alias instead of Cname.

5) Routing policies and use cases -

Simple, weighted, latency, failover, geolocation

6) Route53 supports MX reccords.

7) There is 50 domain names available by default, however it is a soft limit and can be raised by contacting AWS support. http://aws.amazon.com/route53/faqs/

**Elasticache** -

1) Memcached and Redis

Typical exam question "Database is under a lot of stress/load. Which service can you use to alleviate this?" -

Elasticache is a good choice if the DB is particularly read heavy and not prone to frequent changing.

Redshift is a good choice if the stress is because of management running OLAP queries on it.

Other possible answers = read replicas or using dynamoDB table

**VPC -**

1) Think of VPC as logical DC in AWS

2) Consists of IGW or VPGs, route tables, NACLS, subnets and security groups

3) 1 subnet = 1 AZ

4) Security groups are stateful, NACLs are stateless.

5) Meaning, with security groups, if you open port 80 in the inbound rules list, then there is no need to explicitly open port 80 on the outbound list. It is automatically done. However, with NACLs, we have to explicitly open both inbound and outbound ports (outbound ports are NOT opened automatically).

6) No Transitive peering. (If A connects to B and B connects to C, it doesn't mean A can communicate with C)

7) When creating a NAT instance, disable source/dest check on the instance

8) NAT instances must be on a public subnet

9) There must be a route out of the private subnet to NAT instance

10) The amount of traffic a NAT instance can support depends on the instance size. If you are bottlenecking, increase the instance size.

11) We can create high availability using autoscaling groups, multiple subnets in different AZs and a script to automate failover.

12) NAT instances are behind a security group.

13) **NAT Gateways** –

Preferred by Enterprise

Not behind a security group

Fully AWS managed

They scale up to 45 Gbps

Automatically assigned a public IP address

Remember to update route tables

Should be created in each AZ for redundancy and availability

No Need to disable source/destination checks

More secure than NAT instance

14) VPC automatically comes with a default NACL and it allows all inbound and outbound traffic by default

15) We can create custom ACLs. By default, each custom ACL denies all inbound and outbound traffic.

16) Each subnet in a VPC MUST be associated with a NACL. If we do not explicitly associate a subnet with a NACL, then it is associated with the default NACL for that VPC.

17) We can associate a NACL with multiple subnets; However, a subnet can be associated with only one NACL at a time. Wheen we associate a NACL with a subnet, the previous association is removed.

18) NACLs can span AZs. Subnets cannot (subnets are basically the AZs themselves).

19) NACLs contain a numbered list of rules that is evaluated in order, starting from the lowest numbered rule.

20) NACLs have separate inbound and outbound rules and each rule can either allow or deny traffic.

21) NACLs are stateless, we have to configure both inbound and outbound rules (outbound rules are NOT automatically created for their inbound counterparts; Unlike security group).

22) Ephemeral ports are to be opened only for outbound rules and not the inbound rules.

23) We can block specific IP addresses using NACLs and not using security groups.

**VPC Flowlogs** -

1) We cannot enable flow logs for VPCs that are peered with our VPC unless the peer VPC is in our account.

2) We cannot tag a flow log.

3) After creating a flow log, we cannot change its configuration; for instance, we cannot associate a different IAM role with the flow log.

4) Not all IP traffic is monitored.

Traffic generated by instances when they contact AWS DNS server is NOT monitored. However, if we use our own DNS server, then all traffic to that DNS server is logged.

Traffic generated by a windows instance for Amazon windows license activation is NOT monitored.

Traffic to and from 169.254.169.254 for instance metadata is NOT monitored.

DHCP traffic is NOT monitored.

Traffic to the reserved IP addresses for the default VPC is NOT monitored.

5**) NAT vs Bastion -**

A NAT instance is used to provide Internet traffic to EC2 instances in private subnets.

A bastion is used to securely administer EC2 instances (using SSH or RDP) in private subnets. In Oz, this iscalled "jump boxes".

**VPC Overall –**

1a) NAT Instances

2) When creating a NAT instance, disable source/destination checks

3) NAT instances must be on a public subnet

4) There must be a route out of the private subnet to the NAT instance for this to work

5) The amount of traffic that NAT instances can support depends on the instance size. If you are bottlenecking, increase the instance size.

6) We can create high availability using autoscaling groups, multiple subnets in different AZs and a script to automate failover.

7) NAT instances are always behind a seccurity groupp.

1b) NAT Gateways -

1) Preferred by the Enterprise

2) Scale automatically up to 45Gbps

3) No need to patch

4) Not asssociated with security groups

5) Automatically assigned with a public IP address

6) Remember to update route tables with the ARN of NAT gateway

7) No need to disable source/destination checks

8) More secure than a NAT instance

1c) NACLs -

1) VPC automatically comes with a default NACL

2) You can create a custom NACL. By default, each NACL automatically denies all traffic.

3) Each subnet in a VPC MUST be associated with a NACL. If we do not explicitly associate a subnet with NACL, then it is automatically associated with the default NACL.

4) We can associate a NACL with multiple subnets. However, a subnet can be associated only with a single NACL at any given time. When we associate a NACL with a subnet, the previous association (if any) is removed.

5) NACLs contain a numbered list of rules that is evalated in order, starting with the lowest numbered rule.

6) NACLs are stateless; Inbound and outbound rules have to be configured explicitly. Each rule can either allow or deny traffic.

7) We can block IP addresses using NACLs but NOT using SGs.

**SQS -**

1) SQS is pull based, not push based.

2) Messages are 256 KB in size.

3) Messages can be kept in queue from 1 minute to 14 days. Default = 4 days.

4) Visibility timeout = amount of time that the message is invisible in the queue after a reader picks up that message. Provided the job is processed before the visibility time out expires, the messages will then be deleted from the queue. If the job is not processed within that time then the message will become visible again and another reader will process it. This could result in the same message being delievered twice.

5) Maximum visibility timeout = 12 hours.

6) SQS gurantees that the message will be delievered at least once.

7) Amazon SQS long polling is a way to retrieve messages from SQS queues. While regular short polling returns immediately, even if the queue being polled is empty, long polling doesn't return a response until a message arrives in the queue or the poll times out. This can save money because EC2 instance will not have to continuously poll the instance.

**SWF -**

1) API gateway has caching capabilities to increase performance

2) API gateway is low cost and scales automatically

3) We can throttle API gateway to prevent attacks

4) We can log results to cloudwatch

5) If we are using javascript or AJAX that uses ultiple domains with API gateway, we need to ensure that we have enabled CORS on the API gateway.

**Additional exam tips** -

1) Kinesis => Way to consume Big data, stream large amounts of data from social media, news feed logs etc.

Process large amount of data => RedShift for BI, EMR for BigData processing

2) **EC2 - EBS Backed vols vs Instance stored vols**

3) EBS -

EBS backed volumes are persistent (They continue to live even after the EC2 instance has died)

EBS volumes can be detached and re-attached to other EC2 instances

EBS volumes can be stopped and data will persist.

EBS is for long term data storage

4) Instance Store -

Instance store backedup vols are not persistent (ephemeral)(They can live only as long as the EC2 instance lives)

IS volumes cannot be detached and reattached to other instances

IS volumes cannot be stopped. If we do this, data will be wiped.

IS Volumes NOT for long-term data storages

5) **OpsWorks** -

Orchestration service that uses Chef

Chef consists of recepies to maintain a consistent state

Look for term chef or recepies or cookbooks and think OpsWorks

6) **Elastic Transcoder –**

Transcodes media files from one format to another

Charged based on time time taken to encode and the resolution at which the file is being encoded.

7) SWF Actors –

WF Starter - Application that can start a WF.

Decider - Control the flow of activity tasks in WF execution

Activity Workers - Carry out activity tasks

8) **EC2 - Get public IP address -**

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

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

This is instance's metadata and NOT user data.

9) Consolidated billing and AWS organizations -

Consolidated billing allows to get volume discounts on all accounts.

Unused RIs for EC2 are applied across the group.

Cloud trail is on a per account and per region basis but can be aggregated into a single bucket in the paying account.

10) AWS Organizations is under "Helpful tips" - "Create an Organization".

**ECS -**

1) ECS = Amazon's managed EC2 container service. Allows us to manage docker containers on a cluster of EC2 instances.

2) Containers are a method of operating system virtualization that allows us to run an application and its dependencies in rss-isolated processes.

3) Containers are created from read-only templates called images.

4) An image is a read-only template with instructions for creating a Docker container.

5) Images are stored in registry such as Dockerhub or AWS ECR.

6) Amazon EC2 container registry (ECR) is a managed AWS docker registry service.

7) A task definition is required to run docker containers in ECS.

8) Task definitons are text files in JSON format that describe one or more containers that form our application.

9) Think of task definition as a cloud formation template but for docker. Configures things such as CPU, RAM etc.

10) ECS service allows us to run and maintain a specofied number (or "desired count") of instances of a task definition simultaenously in an ECS cluster.

11) Think of services like auto-scaling groups for ECS.

12) An ECS cluster is a logical grouping of container instances that we can place tasks on.

13) Clusters can contain multiple and different container instance types.

14) Clusters are region-specific.

15) Container instances can only be a part of one cluster at a time.

16) We can create IAM policies for clusters to allow or restrict users access to specific clusters.

17) We can schedule ECS in 2 ways -

Service Scheduler

Custom Scheduler

18) ECS agent to connect EC2 instancesto your ECS cluster. LINUX only.

19) IAM with ECS to restrict access.

20) Security groups operate at the instance leve and NOT at the task or container level.