PACKT BOOK NOTES

**STS**

STS is a web service that enables an application to dynamically generate *temporary security credentials* with restricted permissions based on an IAM Role. This temporary credential canbe generated either for an IAM user or for a federated user as we have seen in the previoussection for web identity federation.Temporary security credentials generated using AWS STS for a trusted user can controlaccess to your AWS resources.

**AWS STS and AWS regions**

Temporary security credentials are generated by AWS STS. By default, AWS STS is a global service with a single endpoint at https:/ / sts. amazonaws. com. However, you can also choose to make AWS STS API calls to endpoints in any other supported region. This can reduce latency (server lag) by sending the requests to servers in a region that is geographically closer to you. No matter which region your credentials come from, they work globally.

**Using temporary credentials in Amazon EC2 instances**

You do not have to explicitly get the temporary security credentials—the AWS SDKs, AWS CLI, and tools for Windows PowerShell automatically get the credentials from the EC2 instance metadata service and use them.

**AWS SDKs**

To use temporary security credentials in code, you programmatically call an AWS STS API such as AssumeRole, extract the resulting credentials and session token, and then use those values as credentials for subsequent calls to AWS.

**ENI**

ENI is a virtual network interface. It is a communication hub for an EC2 instance, that enables network communication on an instance. An EC2 instance can have one or more network interfaces. When any EC2 instance is created inside a VPC, by default, a network interface is also created and attached to it. The default network interface created while launching an instance is called a primary network interface of the instance. This primary network interface also gets one primary IPv4 address from the subnet's available IP range.

You cannot detach a primary network interface from an EC2 instance and attach to another. However, you cannot detach a primary network interface, AWS allows us to create additional network interfaces that can be attached to the EC2 instance. Additional network interfaces are also called secondary network interfaces. Secondary network interfaces can be detached from one EC2 instance and attached to another EC2 instance. During this transition from one EC2 instance to another, its attributes remain intact.

**NAT**

NAT can be defined as a virtual router or a gateway in a VPC, that enables instances in a private subnet to interact with the internet. It's an important stopping point for data on its way from private subnets to the internet without directly exposing the instances to the internet. It acts as a firewall, dynamically assigns a temporary public address to an instance, and routes the traffic between the requesting instances and the internet.

There are two types of NAT devices:

NAT gateway: It is the gateway service provided and managed by AWS

NAT instance: It is a custom-provisioned EC2 instance hosting NAT services

**DHCP option sets**

DHCP is a network protocol, that dynamically assigns IP addresses to instances in a VPC from the respective subnet's CIDR block

**S3 Lifecycle management**

Lifecycle management is a mechanism in S3 that enables you to either automatically transition an object from one storage class to another storage class or automatically delete an object, based on configuration.

**S3 Cross-region replication**

Amazon S3 enables you to automatically and asynchronously copy objects from a bucket in one AWS region to another AWS region. This is a bucket level feature, which can be configured on source bucket. In the replication configuration, you can specify destination bucket where you want your source bucket objects to be replicated. In the configuration, you can specify a key-name prefixes. S3 replicates all the objects starting with the specific key prefixes to destination bucket

**AWS Storage Gateway**

is a **hybrid storage service** which connects onpremise environments with cloud storage using a software appliance. It seamlessly connects on-premise environments with Amazon's block-level and object-level storage services such as EBS, S3, and Glacier. Storage Gateway uses standard storage protocols such as NFS and iSCSI. It provides low-latency for exchanging data from on-premise to S3, Glacier, or EBS volumes and vice versa. Storage Gateway can provide high performance for frequently accessed data by caching them at source in on-premise environment.

**Common usage**: Storage Gateway can be configured for use as a file server in conjunction with S3. It can also be used as a virtual tape library for backup on S3 and virtual tape shelf for archival on Glacier. It can also be configured to be used as a local iSCSI volume. Storage Gateway can also be handy for transferring data from on-premise environments to AWS or transferring the data from AWS to on-premise environments.

**1.File gateways**

A file gateway creates a file interface into Amazon S3. It allows you to access S3 using the

NFS protocol. When you opt for a file gateway, a software appliance is hosted in the onpremise

environment on a virtual machine running on VMware ESXi. Once the file gateway

is created, it enables you to directly access S3 objects as files using an NFS volume mounted

on a server

**2.Volume gateways**

When you create a volume gateway, it creates a cloud-backed storage volume, which you

can mount as iSCSI devices on your on-premises servers where iSCSI stands for **Internet**

**Small Computer System Interface**. Volume gateways stores all data securely on AWS.

There are two types of volume gateway, which determine how much data is stored onpremises

and how much data is stored on AWS storage and are discussed as follows:

**>Gateway–cached volumes**

Cached volumes enable you to store complete data on S3 and cache a copy of only

frequently used data on on-premise. By reducing the amount of data stored on on-premise

environment, you can reduce the overall storage cost. It also boosts performance by

providing low-latency access to frequently accessed data using a cache.

**>Gateway–stored volumes**

You can use gateway-stored volumes when you need low-latency access to your entire data

set. It stores all your data locally first and then asynchronously takes a point-in-time backup

of this data as a snapshot to S3. It is generally used as an inexpensive off-site backup option

for **DISASTER RECOVERY**.

**3.Tape-based storage solutions**

A tape gateway serves as a replacement for an on-premise tape drive for backup purposes.

It stores data on Amazon Glacier for long-term archival. It provides a virtual tape, which

can scale based on requirement. It also reduces the burden of managing the physical tape

infrastructure.

There are two types of tape-based storage solutions: **Virtual Tape Library** (**VTL**) and

**Virtual Tape Shelf** (**VTS**), In VTL, data is stored on S3 whereas VTS stores data

in Glacier

**Elastic Beanstalk Version lifecycle**

Elastic Beanstalk creates a newer application version upon uploading a newer source code

bundle. Creating a newer version and not deleting the old unwanted application version

leads to hitting the application version limit. As a result, it does not allow us to create any

newer web application version.

The default Elastic Beanstalk limits are as follows:

**Resource Default limit**

Applications 75

Application versions 1,000

Environments 200

With the help of the application version lifecycle policy for an application, hitting an

application version limit can be avoided. Consistently, it will manage the number of

available application versions at any given time. Once the lifecycle policy is enabled, it will

keep either the total count of recent versions (that is, the last 200 versions of the application)

or the versions which are not older than the specified age in the terms of days (that is, 180

days). Application lifecycle can be configured using the web console, CLI, or API.

**Other notes**

Dynamo DB - throttling gvs provisionexceedexception and 400Http error , bad request

SQS -

>Unlimited no. of msgs n queues per account

>queue naming - max 80chars and unique accross account

>you can delete queue even if u have msgs to b processed

EC2-

If you need to host multiple websites(with different IPs) on a single EC2 instance, the following is the suggested method from AWS: -

-Launch a VPC instance with two network interfaces

- Assign elastic IPs from VPC EIP pool to those interfaces (Because, when the user has attached more than one network interface with an instance, AWS cannot assign public IPs to them)

- Assign separate Security Groups if separate Security Groups are needed This scenario also helps for operating

A user has launched an EBS backed Linux instance. How can a user detach the root device and attach it to another instance as a secondary volume?

  Ans. Stop the first instance and then attach instance's root volume as a new volume to the other instance

Explanation: If an Amazon EBS volume is the root device of an instance, it cannot be detached unless the instance is in the stopped state.

SWF

For asynchronous applications

-Worker

 Tasks are processed by workers which are programs that interact with Amazon SWF to get tasks, process them, and return their results. A worker implements an application processing step

-Decider

To coordinate the application execution across workers, you write a program called the decider in your choice of programming language.

RDS

>If RDS Sec Grp are changed, ingress rule is applied again with "authorizing" status and later "authorized" status

>If the user is launching RDS with Multi AZ the user cannot provision the Availability Zone. RDS is launched automatically instead.

Glacier

>Size of each archive can range from 1byte to 40 TB. >With the help of the S3 lifecycle rules, objects from S3 canbe automatically transferred to Glacier.

>These archives can be logicallyisolated in containers called vaults. A maximum of 1,000 vaults peraccount per region can be created.

EFS

>Ituses the Network File System versions 4.1 (NFSv4.1) protocol.

>When using any EFS volume for the first time, you simply need to mount andformat it to the desired filesystem. Subsequently, you can mount thisvolume on other EC2 instances directly and start using it.

>EFS volumescan also be accessed from on-premise environments using DirectConnect.

>You cannot access it from an on-premise environment overVPN connectivity.

>EFS is available in two modes, General Purpose modeand Max I/O mode.

>Commonusage: EFS is designed to provide very high disk throughput.It can be used for big data and analytics, media, content management,web serving, and home directories.

NAT Gateways

You can use a network address translation (NAT) gateway to enable instances in a private subnet to connect to the internet or other AWS services, but prevent the internet from initiating a connection with those instances.

XRay

Xray SDK provides:

1.interceptors to add to ur code to trace incoming HTTP req

2.client handlers to instrument aws sdk clients thats ur application uses to call other aws services

3.http client to instrument to other internal and external http web services

Lambda

>if your write a code that uses other 3rd party resources or u want to use aws cli instead of console, u need to frst create lambda function deployment package and then use the console or cli to upload package

S3

>prefixes enhance PUT operations

>cloudfront enhances GET operations

API gateway

>type:aws\_proxy

to send client api req directly to lambda type backend

>type:http\_proxy

to send http backend endpoint

>type:mock

for testing api without sending to backend