AWS

AMAZON WEB SERVICES. COMPLETE GUIDE FROM BEGINNERS TO ADVANCED

STEVE FEDLER

AWS

Amazon Web Services. A complete guide from beginners to advanced

Steve Fedler

Download the Audio Book Version of This Book for FREE

If you love listening to audio books on-the-go, I have great news for you. You can download the audio book version of this book for **FREE** just by signing up for a **FREE** 30-day audible trial! See below for more details!

Audible Trial Benefits

As an audible customer, you will receive the below benefits with your 30-day free trial:

- FREE audible book copy of this book
- After the trial, you will get 1 credit each month to use on any audiobook
- Your credits automatically roll over to the next month if you don't use them
- Choose from Audible's 200,000 + titles
- Listen anywhere with the Audible app across multiple devices
- Make easy, no-hassle exchanges of any audiobook you don't love
- Keep your audiobooks forever, even if you cancel your membership
- And much more

Table of Contents

Introduction

I would like to congratulate you for purchasing our Book *“AWS: Amazon Web Services, Complete guide from beginners to advance”*. I am glad to see that you have decided to take a glimpse into the very effectiveness of AWS for modern solutions. AWS is like a boon for most of the computing and machine learning techniques of today.

AWS is meant to be used for all types of machine learning and data analytics systems, and can be used to speed up the entire process. The services provided by AWS are secure and can easily satisfy the most demanding needs.

AWS comes with the benefit of scalability that will allow its users to scale their requirements as per their needs. You will soon be discovering some of the most interesting facts and services of AWS, so that can speed up all your work. You will learn about how to create the best applications.

There are wide ranges of books available on AWS in the market, thank you again for choosing this book. Every effort behind this book has been made for making it as much use as possible for the users. Enjoy!

Chapter 1: What Is AWS? Things you need to know

Amazon Web Services also are known as AWS is regarded as the most comprehensive and the most accepted cloud platform worldwide. It offers more than 165 featured services from all the global data centers. Most of the leading companies use AWS for powering up the infrastructure and lower their overall costs. For all those who are unaware of the various miracles of AWS, it can be regarded as a boon to the cloud-computing sector. Cloud computing is very much necessary for all those businesses which provide on-demand storage and flexible services. It can give the user a completely new level of control over the database and information.

Types of Cloud Computing Process

Iaas or Infrastructure-as-a-Service is the first type of computing for the cloud, which allows the users in gaining internet-based access to all the information on cloud storage. IaaS allows users to sublet machines, virtual networks, servers, storage, etc. There is another type of computing process, which is known as PaaS or Platform-as-a-Service. This computing process allows users to host and create web and mobile applications with the use of internet servers. The last type of computing is the SaaS or Software-as-a-Service that lets the users access the same applications using the cloud storage data for all the devices.

Functionality

AWS provides a wide range of service applications that includes storage, compute, networking, databases, machine learning, analytics, artificial intelligence, application development, security, management, and deployment. Along with the widest range of services, AWS also comes with great functionality. For instance, EC2 is known for offering various sizes and types of computing instances than other providers that also includes the strongest GPU instances used for machine learning.

List of AWS Services

AWS offers some amazing computing services that include storage, security, mobile and email development, servers, networking and many more.

Storage

Amazon provides users with its storage service, which is known as S3 or Amazon Simple Storage Service. S3 offers the users with scalable storage for creating their backup for the data of up to 5 GB. In addition, the users can organize and store their files and data in the buckets of S3.

Amazon Glacier

This is one of the most low-cost services related to cloud storage for cold data. This allows the users to store up their accessed data which are accessed very infrequently for long times of retrieval.

Amazon Elastic Block Store

This is another storage service provided by Amazon. The service of the Elastic Block Store ensures that the users can store their data in block-level style storage that will also be available when the EC2 or Elastic Compute Cloud is shut down.

EC2 or Elastic Compute Cloud

It is a cloud service from the house of Amazon, which is web-based. It acts as a virtual server for businesses for running their applications on it. The servers are generally known as Instances that allows the developers to easily access the entire compute capacity on the AWS global data centers.

Database Management

The database management service provided by Amazon is known as the Amazon Relational Database Service. RDS is compatible with most of the database engines that ensure the users can recover, migrate and take a complete backup of their accessible data.

Migration of Data

AWS provides data migration services for all its users that allow them to migrate their data along with servers, applications, and databases on the public cloud of AWS. Users can easily manage the process of data migration to the storage cloud with the help of Migration Hub. Amazon also provides another service related to data migration known as AWS Snowball.

Networking

AWS provides various tools for balancing network traffic. With the use of VPS or Virtual Private Cloud, users can enjoy full control of a separate segment in the AWS cloud.

Configuration of Cloud and Management Tools

AWS provides tools such as AWS Config Rules and AWS Config for dealing with the configuration of cloud resources. AWS Trusted Adviser also helps the users in selecting the most recommended practices that will help the users in configuring their cloud resources based on performance enhancement, protection, and cost-effectiveness.

Security

IAM or AWS Identity and Access Management help in managing the overall access to cloud resources. AWS provides users with the power of controlling and creating custom policies for their multiple accounts at a time. AWS promises all-round security for all the user data with the help of its worldwide data centers.

Messaging Service

Amazon owns several useful Messaging Services like Amazon Simple Queue Service or SQS, Amazon Simple Notification Service or SNS along with Amazon Simple Email Service or SES. SQS is powerful, quick and is trustworthy while SNS is simple and provides flexible services with push notification. It allows the users to send messages at one go to single or multiple users. SES is an email service that lets all the users to seamlessly and easily send promotional, transactional along with other regulatory emails, which are required by the standards of business. SQS is the most used messaging service from Amazon and has also gained worldwide popularity in no time.

Chapter 2: Identity and Access Management

What is Identity and Access Management?

Identity and Access Management, also known as IAM is the framework for business policies, processes, and technologies. It facilitates the entire management of digital or electronic identities. With the use of the IAM framework, the IT managers can easily control the user access to all those critical information within the organization. IAM products come with a control-based role for information access. This, in turn, allows the system administrators to regulate the system or network access completely based on the individual user roles within the organization.

By access, means the ability of an individual to perform a particular task such as create, view or modify files. The roles are determined in accordance with the competency of job, responsibility, and authority within the organization. The systems, which are most widely used for identity and access management, are a single sign-on system, PAM or privileged access management and multi-factor authentication. Such technologies ensure that the identity is stored securely along with the profile data. It also secures the functions of data governance, which ensures that only the required is shared with the user. The systems based on IAM can be deployed within an organization, which needs to be provided by a third party through a model of cloud-based subscription or in a cloud, which is hybrid in nature.

The basic components of IAM

On the basic level, IAM comes with the following components:

- The way in which individuals are identified within a system.
- The way in which the roles are being identified in a system and how it is assigned to the individuals.
- Updating, adding and removing the users along with the roles in a system.
- Protecting all the sensitive data in the system and performing self-security.
- Assigning the access levels to the individuals.

What should IAM systems include?

IAM systems should include all the required tools and controls for recording and capturing the information for user login, removal of the privilege of access, management of the database of the users for the concerned enterprise and orchestrate the entire assignment. This means that the IAM systems should provide an overall directory service that will have an oversight along with visibility into all the necessary aspects of the user base.

The IAM systems need to balance the automation process and the speed of processing along with the administrator controls which is required for monitoring and modifying the rights to access. In order to manage the requests of access, the central system directory needs to have an access rights system that will automatically match the job titles of the employees and the locations to their levels of privilege.

The systems of IAM need to provide flexibility for establishing the various groups with individual privilege levels with specific roles. This will help in uniformly assigning the access rights completely based on the employee function.

Benefits of IAM

The IAM systems can bring about various benefits within an organization. It can be easily used for capturing, imitating, managing and recording the user identities along with the related permissions of access in a completely automated manner.

- The privileges of access are granted completely in accordance with the interpretation of the policy. All the services and individuals are authorized, authenticated and audited in the proper way.

- With automated systems of IAM, businesses can operate in a more efficient way by reducing the time, effort and money that would be necessary for managing all the accesses to the networks when done manually.
- Companies can have greater control over user access and reduce the overall risk of internal as well as external breaching of data.
- When it comes to security, employing the IAM framework

within an organization will make it easier for enforcing all the policies around the authentication of users, privileges, validation and addressing the issues of privilege sneak.

- The system of IAM helps the companies in better complying with the regulations, which are being set up by the Government by showing that the information related to the corporate employees is not being squandered. It also allows the companies to provide the required data on-demand at the time of auditing.

In addition to the benefits, with the implementation of IAM tools, a company can easily gain a competitive edge among all its competitors. For instance, IAM tools grant the businesses to give out to the users, who are outside the organization such as partners, customers, contractors and suppliers, easy access to its entire network across on-premises apps, mobile applications and SaaS apps without breaching the security of the company. This ultimately results in enhanced productivity for the business, better chances of collaboration, reduced costs of operation and increase on efficiency.

IAM and its risks

Implementation of access management and identity tools means that all the credentials and authorizations of a company are stored up in a single unified space. This whole system might turn out to be extremely risky when not secured properly. In case an attacker gains overall access to the system lacking in security, the overall database of digital identities will be at stake. Similarly, if any individual employee who is authorized with an IAM system fails to follow the recommended security practices, all the relevant information can get exposed easily.

IAM within an enterprise

It might turn out to be a daunting task for getting the funding for a dedicated IAM project because it will not directly affect the overall functionality or profitability of a company. However, the lack of effective management of access and identity might turn out to be a risky affair in relation to the overall security. The requirement of keeping up with the flow of business along with proper management of data access always requires attention at the administrative level. IAM systems can ease up the entire task of implementing and managing the user data along with proper security of the same.

IAM Users, Groups, and Policies

IAM framework is all about an organization and its functioning. It handles the various users, policies along with groups.

IAM Users

It is the identity of an individual that the owner can easily create for allowing the user to interact with the company database. An IAM user consists of the name and credentials.

Credentials and users

A user can access several services from AWS in various ways that depend completely on the provided credentials of the user.

- **Console password:** This is the user password with which the recorded user can easily sign in for the various interactive sessions, for example, the management console of AWS.
- **Access keys:** Access keys come with the combination of a secret access key and an ID of access key as well. You have the power to authorize one or more than one user at a time. The access keys can be easily used by the users for making programmatic changes in your services of AWS.
- **SSH keys:** SSH, the public key can be used for authenticating CodeCommit. The SSH keys will be in the OpenSSH format.

IAM user administration

You can take advantage of IAM user administration for administering the access keys, passwords and MFA devices.

- **Manage passwords:** You can create as well as change the user passwords that provide all-round access to the management console. You can set a policy for your passwords for enforcing the least complexity of the password.

- **Manage the access keys:** You can create as well as update the access keys from time to time for access to the programmatic resources.
- **Find and trace down unused access keys and passwords:** Any person who has access keys or password for the console account of yours or of any of the users of IAM can have complete access to the AWS resources. For practicing the best security measures, it is best to remove those access keys and passwords when the users are not using them anymore.
- **Download report of credentials:** You can easily generate a credential report and download the same. The report will list all the IAM users available within your own account along with various access keys and passwords.

IAM Groups

It is nothing but a compilation of all the users of IAM. IAM groups will allow the user to cite definite admissions for lots of users at a time. This, in turn, makes it easier for the administrator to manage the user permissions in an easier way. For instance, you can create an IAM group named Admin and assign the group with all types of permissions that any administrator needs typically. In case a new user joins the organization and required the privileges of an administrator, you can easily assign permissions for that user by including the new user in the Admin group. Exactly in the same way, if a person in the organization changes his/her role in the job, in place of editing the permissions of user access, you have the power to simply remove that person from the previous group and put them in the new appropriate IAM group.

It is to be noted that the IAM group is not a true identity in IAM as it is not possible to identify the group as the Principal in permission policy. In simple words, it is a method of attaching policies to various or multiple users at a time.

Characteristics of IAM groups

Every IAM group comes with various characteristics of its own.

- An IAM group can contain as many users as the administrator wants. A user can be a part of various groups at a time.
- IAM groups cannot be nested. The IAM groups can only contain users and not any other group.
- There is nothing like a group by default that will naturally include each of the users in an account of AWS. In case you are in need of a group of this sort, you are required to create a separate group and then assign each and every user into that group.
- The total quantity of the IAM groups comes with a limit that an AWS account can have. There is also a limit to the number of groups in which a user can be the part.

IAM Policies

IAM policy is the entity, which when attached to a resource or identity, helps in defining the permissions. The policies of IAM are stored in the form of JSON documents. You, as an admin, have the power to adhere to the identity-based policy for an identity or principal, just like a group, role or user in IAM. The policies based on identity include inline policies, managed policies along with policies, which are managed by the customers.

Creating IAM policies

You can easily create new IAM policies in the AWS management console as an admin by adhering to any of the mentioned methods.

- **Import:** As an owner, you have the power to easily import a policy that is already being managed within your account. You can customize that policy according to your desired requirements. While importing a policy that is managed, it can be either a policy that is managed by customers or a policy managed by AWS that you have created formerly.
- **Visual editor:** In case an owner does not want to keep anything from the previous policies and create a completely new policy from scratch, you can do it by using the visual editor. If you want to use a visual editor, there is no need of understanding the syntaxes of JSON.

- **JSON:** You can create an IAM policy using the syntax of JSON. You need to start by typing a fresh JSON document of policy. JSON policy document includes one or many statements. The statement needs to have all the necessary actions that will share a similar reaction along with supporting similar conditions and resources. In case, one statement requires specification of the admin for all the related resources and the other one supports ARN from a particular resource, the two need to be in completely separate JSON statements.

Chapter 3: AWS Networking

Virtual Private Cloud (VPC)

There are three types of clouds: private, public and hybrid. However, there is another type of cloud, which is available for business organizations, known as the virtual private cloud. It is somewhat related to the public cloud but both are not the same. Instead of just sharing space and resources within a public infrastructure, you will get the volatile allotment of the resources for configuration. There is slight isolation between the admin and the other users with the help of a private IP subnet and along with a virtual communication construct that is based on per user. This provides you with the privilege of ensuring a secure form of accessing your cloud resources remotely. This form of isolation resulted in the name, virtual private cloud because you can operate your very own private cloud within a private cloud.

Difference between VPC and private cloud

People often mix up the two things as being the same. However, that is not the case. The private clouds are dedicated only to your organization, which also combines the hardware. VPC or virtual private cloud does not have the dedication to hardware. It is dedicated to creating a more secure environment within a public cloud or infrastructure. It is somewhat like the VPN, you can use them for sending messages over the all-public internet in a very secure way just the way you would have done in your own personal network, but it is not actually your own network.

Benefits of VPC

Many people think that it is better at using a private cloud rather than using a VPC. However, it is not true in the actual sense. The private clouds are meant to be very expensive for operation. It is also because of the fact that the resources, as well as the hardware required for running the cloud, belongs only to you, so you cannot share the cost with anyone else. VPC comes with the best of both worlds. It provides you with a private cloud for compliance as well as for security purposes and comes with a reduced cost of infrastructure that you can only get with public clouds. The resources that you want to allot are yours and so do not need to share them with anyone else. You are only sharing up the entire infrastructure with others. It is commonly used for IaaS providers. In case you are looking out for a provider that can also provide you with cloud capabilities, VPC or virtual private cloud will be the best available option for your business.

Features of VPC

VPC comes loaded with various features.

- **Multiple options for connectivity:** You can choose from the various connectivity options. You can seamlessly connect VPC to the data center, to the internet or even to other VPCs. The mode of connectivity will be established on those particular resources that the owner would like to publicly expose along with the ones that the owner wants to have secure and private.
- **Hosting public-facing and simple website:** With the help of VPC, you can easily host one simple and basic nature of web application. These websites include the blog and/or just a nominal website in the cloud. You can also have add-on security and privacy layers, which are provided by AWS.
- **Hosting web applications of multi-tier:** You can seamlessly use VPC for hosting web applications, which are multi-tier. You can enforce overall security along with the access restrictions among the servers of application, databases along with web servers. The web servers can be launched in public accessible subnet, whereas, the databases and application servers can be launched in subnets which as non-publically accessible. However, the databases and application servers cannot be accessed directly from the internet.
- **Hosting web applications in the cloud that are scalable and have a connection to the datacenter:** You can easily create VPC in which the subnet instances just like the servers of the web, communicate with the internet and the instances in 0ther subnet like the servers of an application communicates the databases on the network. All the communications between the databases and application servers are secured by IPsec connection between the corporate network and the VPC. The application servers along with the web servers in the VPC can

clout Auto scaling and EC2 elasticity features to shrink and grow as required.

- **Stretching corporate network within the cloud:** With the help of VPC, you can launch the additional web servers, move the corporate applications into the cloud or add up more capacity of computing to the network simply by connecting your corporate network to the VPC. As it is possible to host VPC with the corporate firewall at the front, it is easy to move the resources of IT into your cloud with no need to even changing the way the users are accessing the applications.
- **Backup and disaster recovery:** It is possible to back up the critical data periodically from the data center which is critical to various EC2 instances which are small in size with the help of EBS or Amazon Elastic Block Store volumes. You can also import the images from the virtual machine into Amazon EC2. In case of any disaster or sudden data loss, you can seamlessly launch the replacement for computing capacity in the AWS for ensuring that the business continues at a regular pace. After the disaster period is over, you can simply transfer the critical data of your business mission back to the data center and then terminate the EC2 instances, which you do not require anymore. With the help of VPC for recovery during disasters, you can enjoy all the benefits of a disaster recovery site that too at a half cost of the normal costing.

Subnets, security groups, and NACLs

Subnets

When a user starts creating VPC, the user needs to specify the required IPv4 address range for VPC in the format of CIDR block or Classless Inter-Domain Routing, such as 10.1.1.1/15. VPC covers all the AZs of a region. Just after the creation of VPC, you can easily include one or many subnets in the availability zones. When you are creating a subnet, the user needs to specify the subnet CIDR block, which will act as the subset for the CIDR block VPC. Each of the subnets needs to reside within one of the availability zones and it is restricted from covering the zones. The AZs are the distinct locations that are specially engineered for being isolated from the Availability zone failures. You can easily protect the application failure of one single location by launching the instances in the separate availability zones.

Public, private and VPN-only subnet

In case the traffic of a subnet is being routed in a gateway of the internet, that specific subnet is called public subnet. If you need your required instance within a subnet that is public in nature to seamlessly communicate over IPv4 with the internet, it requires an elastic IP address/public IPv4 address. In case a subnet does not have an internet gateway route, it is called a private subnet. In case a subnet is not having internet gateway route but instead of that has the traffic routed to a gateway, which is virtual private in nature for a connection of SNS VPN, it is called a VPN-only subnet.

VPC and sizing of subnet

VPC offered by Amazon comes with IPv6 and IPv4 addressing and have a different block size limit of CIDR for each. You cannot change the behavior of a VPC having IPv4 CIDR blocks as it is by default. However, you can link a CIDR block, which is of IPv6 with the VPC. For IPv4 CIDR block, the block size which is allowed is between /16 netmask and /28 netmask. The IPv6 CIDR blocks use a fixed length of /64.

Security Groups

The security groups of AWS are linked with the instances of EC2 and are responsible for providing security at the port access and protocol level. The security groups work more or less like the firewall and contain a specific set of rules that can filter out the traffic that comes in and out of the EC2 instances. Each of the security group needs to have a name of its own so that it can be identified from the account menus. It is best to opt for a descriptive style name for the security groups that can tell you the purpose of the group at an instance. The security groups exist within each of the individual VPCs. While creating a security group you need to be assured that it is being created in the exact VPC of those resources it is bound to secure.

Rules of AWS security groups

The rules set which actually filters out the traffic are composed of two types of tables: outbound and inbound. The security groups of AWS are not required to have a similar set of rules for inbound and outbound traffic. Each of the rules is made up of four different fields.

- **Type:** This list lets you choose the protocols such as RDP or HTTP. If you want, you can select protocols by customizing them.
- **Protocol:** After selecting the type, you need to specify the protocol such as UDP, TCP, etc.
- **Port range:** This value comes pre-filled depending on the chosen protocol. You can opt for a custom port range as well.
- **Source:** This field can be either a particular IP address, subnet range or any other security group. If you want, you can leave the access open for the overall internet, which is using 0.0.0.0/0 as the value.

Before you start building up a complex plan that includes several security groups within one VPC, you need to note that only 100 security groups can be added per VPC.

NACLs

NACL or Network ACL is regarded as the network equivalent of the security groups that are connected to the instances of EC2. NACLs are responsible for providing a rule-based tool that helps in controlling the network traffic at the subnet and protocol level. In simple words, NACLs filter and monitor out the traffic, which is moving in and out of the network. If you want, you can attach an NACL to one or more than one subnets within your VPC. In case you do not want to create a custom NACL, the subnets will be automatically associated with the default ACL of the VPC that will allow all of the traffic to move in and out of your network.

Features

You can easily find out that the NACL rule works similarly to the rules of the security groups. However, the NACL rule comes with an additional field known as Rule #. It allows the user to number the rules. This set of rules is important because the NACL rules are always read in the ascending order, where each of the rules is applied to the matching packets regardless of the fact that another later rule might also match with the packet. That is why it is important for sequencing the rules by an organized system of numbering.

NACL limitations

While creating NACLs, you need to be aware of the fact that it comes with a default limit for both inbound and outbound rules per list, which amounts to 20 only for each. You can request higher limits but the maximum limit is 40. It is also to be noted that with each increase in number, it can affect the performance of the network as well. The maximum number of ACLs for each VPC is 200. All the configurations of NACL along with the modification of rules and the subnet associations can be applied with the help of PowerShell, AWS CLI, and AWS EC2 CLI. You can improve your own setup by customizing the ACLs according to your needs.

Chapter 4: AWS Compute

Building up your organization and running the same begins with computing, whether you are building up mobile or cloud-native apps, enterprise or running large clusters for sequencing human genome. AWS offers users with comprehensive computing services that will allow the users to develop, deploy, scale and run the applications along with the workloads in the most secure, innovative and powerful computing cloud.

Benefits of AWS compute

With AWS compute, you can enjoy a wide range of benefits with no compromise.

- **Deepest and broadest platform:** AWS offers the users with a wider selection of services along with various other functions within the computing services. You can select the services depending on your operating system, whether it is Windows or Linux. You can also get a wide selection of instances meant for either general purpose or the purpose, which is optimized for some particular needs such as big data, HPC and analytics.
- **Compute anywhere:** With the opportunity of 66 availability zones, you can start your computing process anywhere you want. You can get a wide range of choices for edge computing and hybrid cloud that also includes VMware cloud, Snowball Edge and AWS Outposts.

Elastic Cloud Compute (EC2) Instances

What is EC2?

Amazon EC2 or Elastic Cloud Compute is responsible for providing a scalable capacity of computing in the AWS cloud. Using Amazon EC2 comes with various benefits, such as it will eliminate your business need of investing in hardware, and thus allows you to develop the applications faster and deploy them for usage. You can use Elastic Cloud Compute for launching as many virtual servers you require, manage the storage and configure networking and its security. EC2 enables the users of AWS to scale up or down for handling the changes in the requirements and thus reduces your need to forecasting the network traffic.

Features of elastic cloud compute

Amazon EC2 comes with various features.

- It provides users with virtual environments for computing which are known as instances.
- Provides pre-configured templates for the instances of your network, known as AMI or Amazon Machine Images. It packs up the bits that your server needs.
- It provides memory, several CPU configurations, storage along with networking capacity required for the instances, called instance types.
- Secures your information on login for all the instances by the use of key pairs.
- It comes with instance store volumes that help in providing volumes for storage of all the temporary data, which has been deleted as you stop the instances.
- It provides persistent volumes for storage of your data by using Amazon EBS or Elastic Block Store.
- It provides several physical locations for resources such as EBS volumes and instances. The physical locations are known as regions.
- It comes with a firewall that will help the user in specifying the particular ports, protocols along with the source IP that can reach out to the instances by using the security groups.
- Provides Elastic IP Address for static IPv4 addresses meant for dynamic cloud computing.
- It provides metadata, which is known as tags that you can assign to the EC2 instances.

Accessing Amazon Elastic Cloud Compute

This service provides the users with an interface that is web-based, known as Amazon EC2 console. In case you have already signed for your account in AWS, you can easily access the EC2 console after simply signing into the Management Console of AWS and then select EC2 on the home page of the console. In case you want the CLI, you can use these options.

- **Amazon CLI or Command Line Interface:** The CLI provides users with the commands required for large set AWS related products. It is easily supported by major operating systems such as Linux, Windows, and Mac.
- **AWS Tools for PowerShell:** It provides the users with a wide range of commands for the products related to AWS for those who prefer scripting in the Windows PowerShell.

Amazon EC2 will provide you with a Query API. Such types of requests are generally of the request of HTTP or HTTPS that uses verbs of HTTP, which are, GET or POST along with a parameter for query known as Action. If you want to build your applications by using the APIs based on specific language in place of just submitting one simple request over HTTPS/HTTP, AWS provides the users with code sample, libraries, tutorials along with various resources meant for developing the software. Such libraries provide various normal functions for automating jobs like retrying requests, cryptographic signing of requests and responsible handling of the errors. This whole step will make it much easier for you to start with the process of software development.

Components of EC2

For starting with EC2, you need to be aware of the components, support, security, operating systems and pricing structures.

- **OS support:** Amazon EC2 supports some of the major operating systems such as Linux, UNIX, Mac, Windows, etc. All of these operating systems need to be enforced in conjunction with the VPC.
- **Security:** As a user, you can have complete control over the visibility of your account in AWS. In EC2, the system of security allows the users to create groups and also place the instances wherever required. You can also specify the groups with which the rest of the groups can communicate.
- **Pricing:** When it comes to pricing, AWS provides a wide range of pricing options that depends completely on the resource types

and the types of database and applications. AWS allows the users in configuring the resources accordingly to their needs and charges.

- **Migration:** The migration service will allow you to move your existing applications into Amazon EC2. It is best for those who have a huge amount of data for migrating.

Amazon EC2 is a very reliable service where you can work independently with your resources. It comes with flexible tools for the administrators and provides a secure network for the resources.

Chapter 5: AWS Storage Services

Cloud storage is one of the most critical components that come with cloud computing. It is necessary for holding all the necessary information, which is used by the applications. Data warehouses, databases, archive applications along with backup, all rely on the architecture of data storage. Cloud storage service is scalable, reliable and much more secured when compared to the traditional type of storage. AWS offers the users with various storage services for supporting both archival along with applicable compliance requirements.

S3 Storage Classes

Each of the objects in Amazon S3 comes with a class of storage associated along with it. For instance, if you list all the objects in the S3 bucket, you can see the storage class for all the objects included in the list. S3 comes with a wide range of storage classes for all your stored objects. You can choose the class for the objects depending on the performance and use case scenario access needs.

Storage classes for the objects, which are accessed frequently

For the data, which are accessed frequently, S3 provides users with these classes:

- **STANDARD:** This is regarded as the default class of storage. In case you do not specify the class while upload your objects, S3 will automatically assign the objects to the STANDARD class.
- **REDUCED_REDUNDANCY:** RRS or Reduced Redundancy Storage has been designed for all those data that are reproducible and noncritical, which can be easily stored with lesser redundancy when compared to the STANDARD class of storage.

When compared with the REDUCED_REDUNDANCY or RRS class, the STANDARD class costs less for the users.

Storage class, which optimizes the frequently and infrequently accessed

data or objects

The INTLLIGENT_TIERING class has been designed for optimizing the costs of storage by moving the data automatically to the storage access tier, which is the most cost-effective, without any kind of impact on the performance or the performance overhead. The INTELLIGENT_TIERING class delivers users with automated cost savings by moving their data on a linear object level between the two access tiers, a low-cost infrequent tier, and a frequent tier. This is done when there is a change in the patterns of access. The INTELLIGENT_TIERING class is a great option for you when you want to optimize your storage costs naturally for all those long-lived data when the patterns of access become unpredictable or unknown.

The INTELLIGENT_TIERING class stores the objects in two different access tiers: one-tier, which is completely optimized for the frequently accessed objects and the other tier, which is optimized for the objects, which are accessed infrequently. By providing a small fee for monitoring and automation for the objects, S3 monitors the patterns of access for the objects in the INTELLIGENT_TIERING class and moves the objects, which have not been accessed for a maximum of one month to the access tier of infrequently accessed data. In case you access an object, which is situated in the infrequent tier, it will be moved automatically into the frequent access tier. There are no additional fees for tiring when the objects within the INTELLIGENT_TIREING class are moved between the access tiers.

Storage classes for the objects, which are accessed infrequently

For the infrequently accessed objects, the storage classes, which are offered by S3, are STANDARD-IA along with ONEZONE_IA. In these classes, IA means infrequent access. The ONEZONE_IA along with the STANDARD_IA object is feasible for the millisecond access. However, S3 charges very less retrieval fee, so the classes are very convenient for the data, which are accessed infrequently. You can choose STANDARD_IA and ONEZONE_IA:

- Storing your data backups, which you do not need very often.
- All those old data that are infrequently accessed but which still require access to the millisecond.

Difference between STANDARD_IA in contrast to ONEZONE_IA

- **STANDARD_IA:** S3 stores up object data constantly across various availability zones. The objects of STANDRAD_IA are volatile to any kind of loss of any availability zone. STANDARD_IA offers a greater amount of resiliency and availability when compared to ONEZONE_IA. It is best for the primary or the copy only data that cannot be recreated again.
- **ONEZONE_IA:** S3 stores data of the object only in one of the availability zones and thus results in making this class much cheaper than STANDARD_IA. However, when it comes to the data, it is not volatile to the loss of availability zones that might result from any kind of disaster such as floods or earthquakes. ONEZONE_IA is very durable just like the STANDARD_IA, however, it is not much available. It is also less volatile. It is best for those types of data that be recreated even if the availability zone fails.

Storage classes for archived objects

For archiving of low-cost data, S3 comes with two types of storage classes, GLACIER and DEP_ARCHIVE. These storage classes for the archived objects come with the same level of resiliency and durability just like the STANDARD class.

- **GLACIER:** It is used for those archives where the data might be required to be retrieved within minutes. All the data stored in the GLACIER class comes with a minimum storage time period of 90 days. It can also be accessed within 1-5 minutes with the help of assisted retrieval.

- **DEEP_ARCHIVE:** It is used for those data that needs to be accessed on rare occasions. The data which is stored in the DEEP_ARCHIVE class comes with a minimum duration of the storage period of 180 days. It comes with a default time of retrieval which is of 12 hours. DEEP_ARCHIVE is regarded as the cheapest option of storage in AWS. You can easily reduce the cost of retrieval for DEEP_ARCHIVE by opting for bulk retrieval in which you can get the data within a time period of 48 hours.

You can easily set the object storage class to GLACIER or DEEP_ARCHIVE in the same way you do for the other object storage classes of Amazon S3.

Storage Gateway

The AWS services of storage gateway allow seamless storage of hybrid nature between on-premises storage space and the cloud. It comes with efficient connectivity of network to the services of cloud storage from Amazon. It delivers local performance at a virtually endless scale. The user can easily use it in their remote offices along with datacenters for backup, archive, and restore, tired storage and disaster recovery. The storage gateway connects directly to the local infrastructure as a virtual appliance and acts as a volume or also as VTL. The connection offered by storage gateway makes it easier for the organizations for augmenting the existing investments of on-premises storage along with high durability, high scalability and low cost for a cloud storage solution.

It is to be noted that the storage gateway does not function as an all-in-one solution for backup. The transfers and the backup are initiated by the existing host and are also managed by the same. It only acts as the portal for extending the infrastructure of data storage. This makes the storage gateway a flexible option as a portal. Storage gateway can be seamlessly integrated with various existing data storage configurations of an enterprise with no or minimal changes in hardware setup.

Gateway types of storage gateway

Storage gateway comes with a variety of gateway types that can make your data transfer and backup a very easy and seamless process.

- **Join as a file server or File Gateway:** The interface of files allows the applications and on-premises servers' easy access as a network-sharing file to the storage gateway. For boosting up the local performance of the servers, the data is cached. In Amazon S3, the data is also accessible as the objects in Amazon S3. For the protection of your S3 data with the help of native tools, you can use cross-region replication and versioning.

- **Connecting as local disk or Volume Gateway:** The storage gateway is presented by the volume interface to the application and on-premises servers as a local disk. The data available in these volumes is possible for transfer into the S3 cloud storage and can be accessed by the services of storage gateway. For the best performance, try to store the data locally along with backups in the cloud as a snapshot or blend upscale and latency by storing up the accessed data with the cooler data locally in the cloud.
- **Connecting as a VTL or virtual tape library or Tape Gateway:** The types of equipment of tape automation along with the backup tapes are replaced by the tape interface with cloud storage and local disk. The existing recovery software and backup write up the native jobs of backup to the virtual tapes, which are stored in the storage gateway. You can easily migrate virtual tapes into S3 and archive the same into Amazon Glacier at a very low cost. The backup application helps in accessing the data and the visibility of all the backup tapes and jobs is maintained by the backup catalog.
- **Transferring data in and out of the cloud:** Storage gateway eases up your job of transferring data into cloud storage. It automatically cushions the data present in the on-premises server and moves it into the cloud storage. It also helps in moving out your data out of the cloud. This whole thing ultimately reduces the cost and time, which goes in transferring data between the AWS cloud and the site. Various optimizations are available such as delta transfers, multipart management, bandwidth scheduling and bandwidth throttling which are standard for all the available interfaces.

How to use the AWS storage gateway?

A storage gateway is generally installed on your host in the data center as a virtual machine. When the storage gateway is activated, the management console of AWS is used up for provisioning the storage volumes. The provisioned volumes than can be mounted as iSCI devices on the on-premises servers. The volumes, which are mounted, can be like any normal storage volume by the local applications.

Key features of storage gateway

- **Managed cache:** The appliances of the local gateway maintain the cache for the recently accessed or written data so that the applications can lower latency access to the data, which is stored in the AWS cloud. The storage gateway uses a write-back and read-through cache.
- **Standard protocols for storage:** The gateway of storage seamlessly connects with the local backup applications with SMB, NFS, iSCCI-VTL or iSCSI. It helps in adopting the AWS cloud storage without any need of modifying the applications. The device emulation and protocol conversion provided by storage gateway enables the users to access the block data on the volumes, which are managed by the storage gateway.
- **Secured and optimized transfer of data:** The storage gateway helps in secure uploading of data and secure downloading of the data requested by the user encrypts any type of data that is in transit between any gateway appliance and AWS cloud by using SSL. Optimizations are applied for all the virtual and block tape data.
- **Integration of AWS:** Storage gateway is a native AWS service. It integrates with all other services of AWS required for backup, storage, and management. The storage gateway service stores the files as Amazon S3 objects, stores the EBS snapshots, which are generated by the volume gateway with EBS and archives the virtual tapes in GLACIER. The service of storage gateway also integrates with the backup service of AWS for seamless management of recovery and backup of the volumes of volume gateway, helps in meeting the regulatory and business backup requirements and simplifies the overall management of backup.

Adding to the features, storage gateway also provides seamless experience of management by using AWS Console service for monitoring and security with the help of other services from AWS like CloudTrail, CloudWatch, IAM and KMS.

Chapter 6: Elasticity and Scaling EC2

Elastic Load Balancing

ELB or Elastic load balancing helps in automatic distribution of the traffics of an application across various targets such as containers, instances of Amazon EC2, Lambda functions, and IP addresses. ELB can easily cope up with the fluctuating traffic application load across multiple or in a single AZ. Elastic load balancing or ELB comes with three different load balancers which features automatic scaling, high availability along with robust security which is very much necessary for making the applications tolerant of all kinds of faults.

- **Application Load Balancer:** This load balancer type is the best option for load balancing for HTTPS and HTTP traffic. It helps with an advanced type of request routing which is targeted to very delivery of the architectures of modern applications that also includes containers and microservices. This load balancer works at the individual layer of request and routes traffic within the Amazon VPC, which is based on the requested content.
- **Network Load Balancer:** This load balancer type is the best option for TCP or Transmission Control Protocol load balancing, TLS or Transport Layer Security and UDP or User Datagram Protocol where high-end performance is required. It operates at the level of connection and routes the traffic to several types of targets in the VPC. It has the capacity of easily handling a huge number of user requests every second along with maintenance of super-low latencies side by side. The network load balancer is optimized in a way so that it can also handle volatile and sudden patterns of traffic.
- **Classic Load Balancer:** This type of load balancer provides load balancing at the basic level across various instances of Amazon EC2. It operates both at the connection level and at the request level. This type of load balancer is best for those types of applications, which are built within the classic EC2 network.

Benefits of ELB or Elastic Load Balancing

ELB comes with various benefits for the users.

High availability

Along with the distribution of traffic across several targets, ELB also loads balance across the region by routing traffic to the healthy targets in various availability zones. The Amazon ELB commits for 99.99% of availability for a load balancer.

Secure

ELB or Elastic Load Balancing works hand in hand with Amazon VPC for providing potent features of security that also includes authentication of the user, integrated management of certificate and SSL/TLS decryption. By combing all of these, ELB provides you with the ultimate flexibility for managing the settings of TLS and offload the CPU workload from the applications.

Elastic

ELB can easily handle sudden and rapid changes in the patterns of network traffic. Additionally, Auto Scaling with deep integration makes sure of up to the marked capacity of the applications for meeting the varying application loads without the requirement of manual intervention.

Flexibility

Elastic Load Balancing or ELB allows users to use their IP addresses for routing the requests to the application targets. This, in turn, provides you with flexibility in the way you are going to virtualize the application targets and thus allows you to host more than one application on a similar instance. This also allows the applications to enjoy their own security groups respectively and use a similar port of network for further simplification of the inter-application communication in an architecture based on microservice.

Potent auditing and monitoring

ELB allows the users to monitor their applications along with their performances in real-time with the use of Amazon CloudWatch, logging, and metrics and tracing of requests. This helps in improving the overall visibility of your application behavior, identify the bottlenecks of performance and uncover the issues of your application stack on an individual request.

Hybrid balancing of load

Elastic Load Balancing offers the ability to balance the load for AWS along with on-premises source by using the same balancer of the load. This makes it easier for the users to migrate, failover or bursts the on-premises applications available on the cloud.

Features

Achieving a better level of fault tolerance for the applications

ELB provides an easy level of fault tolerance, which is required for the applications by automatic balancing all the network traffic to the available targets. In case not all the targets in the availability zone are healthy, ELB routes the traffic to the other health targets in another availability zone. Once the targets in the availability zone return to a healthy state, the load balancing automatically returns to the original targets.

Automated scaling of applications

ELB provides the users with the confidence that the application will effectively scale to the customer demands. With the feature of Auto Scaling for the instances of Amazon EC2 when the latency of any of the instances exceeds the threshold that was preconfigured, the application will be ready for serving the next request of the customer.

ELB and VPC

ELB makes the creation of an internet entry point into the user’s VPC an easy job. You can seamlessly assign the security groups to the load balancer for controlling which of the ports are in the open state for the allowed sources. As ELB comes integrated with the VPC, all the existing ACLs or Network Access Control Lists continue providing additional control of the network. When you create a load balancer for your dedicated VPC, you can choose whether the load balancer will be internal or internet-facing which is by default. In case you choose internally, you do not need to have an internet gateway for reaching out to the load balancer. The private IP address that comes with a load balancer will be used for the DNS record of the load balancer.

The user can configure the checking of the health of your targets so that the ELB is capable of request sending to healthy targets only. You can offload the overall work of encryption and decryption to the ELB for making the resources of computing to focus on the main job.

Auto Scaling

The AWS auto-scaling helps in monitoring the applications and automatically adjusts the capacity for maintaining predictable and steady performance and that too at a reasonable price. The auto-scaling comes with the easy setup process for scaling the applications for various resources beyond numerous services within a minute. Auto-scaling service provides a very powerful and simple interface for the users that will help you a lot in planning the plans of scaling of the resources that also include Spot Fleets and instances of Amazon EC2, Amazon DynamoDB indexes, and tables, Amazon ECS tasks along with Amazon Aurora Replicas. The service of Auto Scaling from AWS makes the step of scaling a simple job along with various recommendations, which will allow optimization of the overall costs, performance and perfect balance between all. It also helps in maintaining the availability of the applications and allows the users to remove or add instances of EC2 according to the defined conditions. The dynamic scaling system responds automatically to the change in the demands and the system of predictive scaling schedules a perfect number of instances of EC2, which is established on the demand that is being predicted.

Benefits of auto-scaling

The EC2 auto-scaling comes with various features that can help you in improving the overall functioning of your applications.

- **Improvement of fault tolerance:** The system of auto-scaling can easily detect when the EC2 instance is in an unhealthy condition, terminates the same and replaces the unhealthy instance with a new one.

- **Increases availability of application:** Autoscaling makes sure that the applications have the perfect amount of computing all the time automatically and provisions the capacity of the applications with Predictive Scaling.
- **Lowers the costs:** Autoscaling helps in adding up the instances when required and helps in optimizing the overall costs and performance by scaling across the various purchase options.

How does auto-scaling work?

Autoscaling works at different levels with different methods:

Fleet Management

No matter if you are using only single instance EC2 or hundreds of it, you can opt for auto-scaling for detecting the impaired instances of EC2 along with the unhealthy applications. It also helps in replacing the unhealthy instances without any kind of manual intervention. This makes sure that the application is having all the capacity to compute that you require and expect. Auto-scaling performs three major functions for automating the fleet management for the instances of Amazon EC2.

- It monitors the current health of all the instances, which are running. It makes sure that the application of the user is in the state of receiving all the required traffic and that the instances of EC2 are working in the proper condition. The EC2 auto-scaling performs periodic health checks for identifying an unhealthy instance.
- It helps in replacing the impaired instances without any form of intervention. When an instance fails the health check, it is terminated and replaced by a new one by EC2 auto-scaling. Thus, you need not respond when there is any need for instance replacement.
- EC2 auto scaling balances the capacity beyond various AZs. The auto-scaling system can naturally balance the instances of EC2 across different zones and launch the instances, which are new so it is possible to balance them between the various zones in an

even manner.

Scheduled Scaling

It is always helpful when you can schedule a certain job as it automates the whole process and functions without manual intervention. The same thing is associated with scheduled scaling. It allows you to scale based on your required schedule and scale the applications much ahead of the known changes in load. For instance, every 7 days, the application traffic starts increasing on Monday, remains at the peak on Tuesday and begins to decrease again on Wednesday. Therefore, with the help of scheduled scaling, you are able to easily plan your activities of scaling completely based on application traffic patterns known to you.

Dynamic Scaling

The auto-scaling EC2 feature enables the users to closely track the curve of demand for the applications, thus reducing the requirement for the manual provision of the capacity of EC2 beforehand. For instance, you can also use the policies of target scaling for selecting the required load metric, which is necessary for the application like utilization of CPU. You can also set the overall value of your target by using the feature of request count per metric target, which is available with an application load balancer, which is an option for the service of Elastic Load Balancing or ELB. EC2 auto scaling adjusts the number of EC2 instances automatically as required for maintaining the target.

Predictive Scaling

It is a feature provided by the service of auto-scaling that uses the technique of machine learning for scheduling the required amount of instances of EC2 within the prospect of the approaching changes in traffic of your application. With predictive scaling, you can easily predict the future application traffic that also includes the daily occurring spikes and sets up the right number of instances beforehand. The machine-learning algorithm of predictive scaling detects even the slightest changes in weekly or daily traffic patterns and adjusts the forecasts automatically. This, in turn, discards the requirement for adjusting manually the parameters of auto-scaling as the cycle changes with time. Thus, it makes the process of auto-scaling much simpler. Auto-scaling paired with the service of Predictive Scaling can deliver simpler, faster and much more accurate provision of capacity, which will result in much more active applications and lowers the cost.

Choose how and when to scale

You can scale based completely on the metrics of Amazon CloudWatch or also according to a predictable schedule that you can define. You can also receive notifications that can alert you when to use Amazon CloudWatch alarm for initiating the actions of auto-scaling via Amazon SNS. You can also get notifications when auto-scaling completes its job.

ELB and Auto Scaling

ELB and auto-scaling are the two most major components of EC2 that triggers and maintains the overall functioning. Autoscaling is dedicated to increase or decrease the total number of containers or virtual machines that work as per the policy of scaling. The policy of scaling can be easily triggered by various events such as a schedule, metric alarms of CloudWatch or anything else that can make API call. This whole thing helps in scaling the capacity of your application, which is completely based on the planned usage or real-time demand. The capacity can be registered as well as deregistered, but the nodes, which are being removed or added needs to be registered or deregistered with the help of a load balancing solution. For the proper functioning of auto-scaling, each and every aspect of this process needs to be automated. That is where ELB or elastic load balancing comes into play.

The consideration of load balancing for the auto-scaling groups

Auto-scaling indicated that the nodes have already been created or removed. This whole thing might result from the schedules or utilization metrics. The containers or virtual machines that are being added will not do anything more for serving the load unless the entity is notified in some way which is feeding them. When the containers or machines are detached due to an event of auto-scaling, is the nodes are not deregistered, the ELB or load balancer will continue directing the traffic to those nodes. While virtual machines are loaded or removed from the load balancer, it is also very important to consider if the session persistence is in use and how the overall load is being assigned.

- **Adding the nodes**

Whenever a new node is added to the capacity of the application, it is required to register the same with the solution of load balancing otherwise none of the traffic will be pushed to the new capacity. Many people might also think that adding up the capacity and then not using the same might be harmless. However, it might also result in some adverse effects. The first thing that comes in the list is the cost, no need to pay for something, which you are not even using. The second thing is the metrics. It is a very common thing to use the CPU utilization statistical average for determining is the capacity is required to be added or detached. While using average, it is always assumed that the load is evenly distributed among all the virtual machines. However, when the load is in an unbalanced state, this whole assumption might result in some serious issues. The whole statistic might jump back from a high point of average to a lower one. This thing is called a rubber banding. It takes all the action to serve the overall demand but does not provide any intended effect.

- **Deregistering the nodes**

While auto-scaling, you also need to consider the deregistering of the nodes from the load balancer. The nodes need to deregister from the load balancer whether they are checking the health actively or not. In case a node suddenly gets unavailable, your clients will be experiencing worse session loss or session timeouts if the application is dependent on the persistence of the session. For removing a node from the pool of an application in the clean state, you need to drain the connections first along with the persistent sessions and then opt for deregistering.

- **Load balancing algorithm**

You need to consider the algorithm of load balancing which is being used by the solution. It is a very important thing to understand how the removing or adding of a node from the application pool will distribute the overall load. There are various algorithms such as the round-robin, which aim in distributing the load evenly completely based on a provided metric. The round-robin algorithm does load balancing based on the sum metric, which is being requested. Removal or addition of nodes while using algorithms like the round-robin will have no or less impact on the overall distribution. While you are load balancing, you have various things to consider. The first thing that you need to consider is the way in which the machines are registered or deregistered from the solution of load balancing. The second thing that you need to consider is the session persistence impact along with the changes in the load distribution.

Approaches to load balancing

The considerations mentioned above can be approached easily with a little bit of foresight and automation.

Proactive approach

The best way of approaching session persistence is by moving the session. For appropriate load balancing, your clients need to be able to hit any of the nodes in the application without any kind of issue. By giving all the servers of your application access to the shared and centralized memory, you do not need session persistence any longer. In case the session of moving is being restricted, you can easily handle the sessions with the help of a ported load balancer.

Reactive approach

You can opt for a reactive approach by letting your load balancer query AWS API and then update the load balancer when nodes are removed or become online. This whole approach is reactive as the load balancer is updated right after the node is gone or is alive. Most of the reactive approaches fall short as the load balancer face timeouts just before the health checks notify the node, which is missing as unhealthy. There are various features of load balancing that handles easily the upstream timeouts and then proxies the request to the available upstream node. This approach is very much essential for preventing your clients from seeing the errors of HTTP. Reactive approaches can also be carried on using SNS or Simple Notification Service. Such notifications trigger the arbitrary code, which is meant to be run by an API that looks after and controls the process of registration as well as deregistration.

Chapter 7: Content Delivery and Domain Name System (DNS)

DNS or Domain Name System is the regulation by following which all the names, which are used by the users on the internet, are concluded with the respective IP addresses. The hostname is a unique name that names a computer in a unique way. DNS hostname is composed of the domain name along with one hostname. The DNS servers are responsible for resolving the DNS hostnames with their respective IP addresses.

CloudFront and DNS

For speeding up the delivery speed of your web content, the best way is to use the CloudFront. CloudFront is AWS content delivery network. CloudFront is capable of delivering your whole website that also includes the static, dynamic, interactive and streaming content. In doing so, CloudFront uses a system of a global network comprised of all the edge locations. The request for the web content is routed in an automatic way to edge location that in turn gives your users the lowest amount of latency. It is also to be noted that you can only route the traffic to CloudFront distribution only for the hosted zones that are public.

For using CloudFront in order to distribute the web content, you need to start by creating a web distribution and then specify all the settings such as HTTP server or Amazon S3 bucket from where you want CloudFront to get all the content. You also need to specify whether you want only one selected user group for accessing the content or whether you want all the users to employ HTTPS. When you construct a web distribution, you will be assigned a domain name to the distribution by CloudFront, for example, d111abcde258.abcfront12.net. The user can also use the domain name for URLs of content like:

http://d111abcde258. abcfront12.net/logo.jpg

However, in case the user wants to use their own created domain name for the URLs, they can do that as well, for example:

http://exampledomain.com/logo/jpg

In case you want your own domain name in the URLs of your content, you need to use Amazon Route 53 for creating an alias record that ultimately points out the CloudFront distribution. Alias record is an extension of Route 53 to the DNS. It is completely similar to the CNAME record. However, you can create alias record for both root domain, such as examp.com and for the subdomains, such as www.examp.com. When Route 53 will receive a query from the DNS that also matches the type and name of alias record, Route 53 counters along with the name of the domain, which is linked with the distribution.

Requirements

Before you start with the process, you require the following:

- You need a CloudFront web distribution first. The whole distribution must have an alternate name of your domain that needs to match with the domain name that you are going to use for the content URLs in place of the domain name, which was assigned by CloudFront for your distribution. For example, if you want your content URLs to contain this domain name: examp.com, the field for the alternate name for your domain for the whole distribution needs to include examp.com.
- You must have a domain name that needs to be registered as well. You can choose any registrar you like or can also use the service of Amazon 53 as the registrar of your domain.
- You need to have Route 53 as the primary DNS service of the domain. When you register the name of your domain with the help of Route 53, it will be configured automatically as the primary DNS service of the domain.

Configuring Route 53 for traffic routing to CloudFront distribution

For configuring Route 53 for traffic routing to CloudFront distribution, you need to follow the following procedure.

For traffic routing to CloudFront web distribution

- Get the name of the domain that was assigned by CloudFront for the web distribution and make sure whether IPv6 is enabled or not:

1. Sign-in into the management console and then open the console of CloudFront.
2. Select the distribution name on which you want to route the traffic.
3. Under the general tab, determine the value of the domain name.

- Under the IPv6 field, check whether it is enabled for your distribution or not. In case IPv6 is enabled, you need to create 2 alias records, one for routing IPv4 traffic and one to route the traffic of IPv6.
- Add one or more than one domain name to the CloudFront distribution. These domain names are the ones that you will be used as the main domain name and the subdomain name.
- Sign in again into the management console and then open the console of Route 53.
- Choose Hosted Zones at the navigation panel.
- Select the hosted zone name for your selected domain, which you will be used for routing traffic to the CloudFront distribution.
- Select create a set of records.

After you are done with all the mentioned steps, you need to specify these values:

- **Name:** Enter the name of the domain, which you want to be used for traffic routing to the distribution of CloudFront.
- **Type:** Choose the IPv4 address. In case you have enabled IPv6 for your distribution, choose AAAA IPv6 address.
- **Alias:** Select yes for this section.
- **Alias target:** In this section of CloudFront distributions, select that name which was assigned by CloudFront to the distribution when it was created. This will be the value that you have received in step 1.

- **Routing policy:** Choose the routing policy, which is applicable for your distribution.
- **Evaluate target health:** Select the default value, which is, No.

After you are done with all these steps, you just need to select create.

Route 53 and DNS

Amazon Route 53 is a system of scalable and available DNS or Domain Name System. You can use the services of Route 53 for performing three different types of functions and that too in any type of combination: registration of your domain, health checking and routing of DNS. If you want to use Route 53 for all the mentioned function, you need to follow these steps:

- **Register the names of the domain**

  When you want to host a website of your own, the website needs a specific name such as abcd.com. With Route 53, you can easily register the name for the web applications or website, which is known as the domain name.

- **Route the traffic to your domain resources**

  When any user opens the browser and the name of your domain or the name of your subdomain in the designated address bar, Route 53 helps by connecting the browser with your designated web application or website.

- **Checking the health of the website resources**

  Route 53 helps in checking the health of the resources. It does so by sending automatic requests to the resources over the internet such as web server, for verifying it is readily available, reachable and also functional. You can also select the option for receiving notifications when any of the resources becomes unreachable or unavailable and then choose to route away from the traffic from the unhealthy resources to a fresh one.

Registration of domain

You need a domain name while you create a web application or website for your content. Your domain name will look like abcd.com, which the users will be entering in the address bar for viewing the website. The registration of the domain name is an easy job.

- You need to start by choosing a domain and check, whether it is available or not. In case the domain name you want is already taken, you can try by choosing other names or by changing the domain like.com to another domain like .hockey or .ninja. There are various top-level domains that you can find in the list of Route 53.
- After choosing the name of the domain, you need to register it using Route 53. While registering, you need to provide various information about the domain such as the name of the owner and contact information of the owner of the domain. When you register your domain with the help of Route 53, it will automatically make itself the service of DNS for that domain after performing the following:

  1. Create a hosted zone that will have a similar name just like the domain.
  2. Assigns four name sets of the servers for the hosted zone.
  3. Gets the name of the server's right from the hosted zone and will add them to your domain.

- After the registration process, your information is sent to the domain registrar. The domain registrar can either Amazon Registrar Inc. or something else.
- Your information will be sent out for the registry of the domain. A registry is nothing but the company, which sells registration for the domains.
- The registry will be storing your domain information in the database.

In case you have already registered the domain name with any other registrar, the user can easily transfer the same into Route 53. For this, it is not necessary to use the features of Route 53.

How is the traffic routed to a website using Route 53?

All the computers, smartphones or laptop that serves content for the massive websites of retail, communicates with each other with the use of numbers. These numbers are known as IP addresses. The IP addresses are generally very long, however, when you visit a site you do not need to remember the IP address. You just need to enter the name of the domain such as abcd.com. DNS services such as Route 53 helps in establishing a connection between the IP addresses and the names of the domains.

Process of traffic routing

- When you enter a domain name and press enter, the domain name request is routed to DNS resolver, which is managed by the user's ISP.
- The resolver forwards the user request to the DNS root name server.
- The resolver again forwards the request for a domain to a TLD name server for the domains with .com. The name server then responds to the request of the user with four name servers of Route 53, which are associated with the domain. The resolver then caches the name servers of Route 53.
- The resolver chooses a name server of Route 53 and then forwards the request for the domain such as abcd.com to the name server.
- Route 53 looks into the hosted zone of abcd.com and gets the associated IP address for the webserver.
- The DNS resolver gets the IP address and returns the resulting webpage for abcd.com

How is the health checking of resources done by Route 53?

Route 53 checks the health status for resources such as email servers or web servers.

- You start by creating a health check with Route 53 and specify the values that will determine how you want the health checking to function.

- Route 53 will start by sending requests to an endpoint at the specified interval. If the endpoint responds to the request of Route 53, the endpoint is regarded as healthy.
- In case the endpoint fails to respond, Route 53 starts counting the total number of endpoint requests that have failed.
- When Route 53 specifies the endpoint to be unhealthy, it notifies CloudWatch.
- CloudWatch will trigger the alarm and takes the help of SNS for sending out notification to the recipients who are specified.

You can also perform a health check using Route 53 for checking the status of the CloudWatch alarm.

Chapter 8: Monitoring, Logging, and Notifications

Monitoring and Logging

AWS provides users with monitoring services that have been built for the DevOps developers, engineers, IT managers and the SREs or site reliability managers. The monitoring service from AWS is provided by CloudWatch that provides the users with various actionable data and insights for monitoring the applications, optimizing the utilization of the resources, responding to the system-wide changes in performance and for getting a unified view of the health of application operation. The monitoring service provided by CloudWatch collects the operational as well as monitoring data in the form of metrics, logs, and events. It also provides users with a consolidated view of the applications, resources along with services, which run on the servers, situated on-premises. With this service, you can easily detect abnormal behavior of your applications, visualize the logs, set alarms, take actions, discover the application insights and troubleshoot various issues.

Benefits of AWS monitoring

AWS monitoring can offer you with a wide array of benefits.

- **Power of observation across one single platform**

  A modern application that runs on the architecture of the microservices generates huge data volume in forms of logs, events, and metrics. CloudWatch allows the users to gather access and then correlate the data, all in one platform from all the resources of AWS. It helps in breaking down the functioning of your applications so that you can have wider visibility of your overall system.

- **Gathering operational insights and visibility**

  For the ultimate optimization of your resources and performance, all you need is an undivided view of the operation, historical reference, and real-time data. CloudWatch can help you in regards to this by providing data with a granularity of 1 second. It also offers the storage of metrics and retention of the same for 15 months.

- **Collect AWS metrics**

The task of monitoring your AWS resources and the applications becomes easier with the help of CloudWatch. It integrates along with 70 or more services of AWS such as EC2, DynamoDB, ECS, S3, Lambda, and EKS. It helps by publishing metrics of 1 minute so that you can gain greater insights into your logs.

- **Improve the optimization of resources along with operational performance**

With CloudWatch, you can set alarms and automate the actions completely based on the algorithms of machine learning or thresholds, which are predefined for identifying any kind of abnormal behavior in your metrics. You can also use this service for triggering the workflows with various services like Amazon SNS, AWS Lambda and AWS CloudFormation.

- **Deriving actionable and useful insights from your logs**

With CloudWatch, you can analyze, explore and visualize all your application logs, which help in troubleshooting the various problems of operation easily. With the insights of the logs, you need to pay only for those queries that you need and the ones that you run. It provides the users with answers in no time by scaling the log volume and complexity of the query. You can also publish the metrics of your logs for all-round visibility of the operation.

Logging

The logging service provided by AWS enables the organizations to clearly understand the relationship between operation and security along with changing the events management. It also helps in maintaining a comprehensive outline of the infrastructure as well. The customers of AWS enjoy specific access to the log files and metrics for gaining greater insight into the operation of each AWS service, configuration changes, API calls, and billing events. The log files from the applications, web servers, and the operating systems provide precious data in various formats and in a well-distributed manner.

Features of logging

AWS logging service comes with various features that can help you in gaining a greater insight into the functioning of the applications.

- **Implementation of centralized logging reference**

  You can deploy a centralized solution of logging by using AWS CloudFormation. The template of CloudFormation automatically launches and configures the various components, which are necessary for uploading the log files from multiple accounts. It also uploads the log files from AWS regions to the Amazon ES for visualization and completes analysis in a user-friendly and customizable dashboard.

- **Easy access to dashboards with the help of Amazon Cognito**

  You can easily control the overall access of your dashboards with the help of Amazon Cognito that also helps in simplifying the process of authentication to Amazon ES.

- **The capability of logging beyond the default service logs of AWS**

  You can extend the capabilities of logging beyond the default level of services logs of AWS. This super flexible solution comes with various examples that are useful for capturing the log files and host-level along with the VPC log flows. It has been designed for scaling according to the growth of your business.

- **Visualization of data with the help of built-in support from Amazon ES**

  You can simplify the visualization process for your data by using the built-in support from Amazon ES for Kibana. It also includes a set of dashboards that are preconfigured by default and gives you the first visualization into the capabilities of Kibana for customization.

The architecture of centralized logging

The primary template of logging deploys a domain of Amazon ES, which is actually the exposed software, hardware, and data by the endpoints of Amazon ES. A custom function of AWS Lambda is deployed for loading the log data from the CloudWatch to a domain of Amazon ES. It is being configured with one default set of dashboards from Kibana as the starting point for the visualization of data. A pool of Amazon Cognito provides user authentication for the Kibana dashboard. One secondary template allows the customers for indexing the logs from the second region and accounts for the domain of Amazon ES in the prime region or account. There is also one demo template that deploys the sample logs, which can be used by the customers for the purpose of testing.

Cloud Watch (Monitoring, Metrics, and Logs)

Monitoring

You can easily monitor all the instances of EC2 with the help of Amazon CloudWatch. It collects and then processes up the raw data collected from Amazon EC2 ad transforms the same into real-time, readable metrics. The statistics are stored for 15 months that in turn provide you with historical information about your instances. It also allows the users to get a better visualization of their web application and the way it is performing. Amazon EC2 sends out metric data to the CloudWatch in a period of 5 minutes, which is set by default. You can also enable detailed monitoring for your instances in a 1-minute period for sending out metric data of your instances to CloudWatch. The console of Amazon EC2 displays metrics of raw data in the form of graphs, which is based on the data shared by Amazon CloudWatch. You can also choose to gather data for the instances from CloudWatch in place of the graphs in the EC2 console, depending on your preference and needs.

Enabling detailed monitoring for the instances

The instances are enabled for detailed monitoring by default setting. However, if you want you can also enable or disable monitoring on your own. After you enable it by yourself, the console will be displaying the graphs of the monitoring of the 1-minute period. You can enable the monitoring for your instances when you launch it or when the instance has been stopped. When you enable detailed monitoring for the instances, it will not be affecting the EBS volume monitoring which is attached to the instances. For enabling monitoring for existing instance, follow these steps:

- Open up the console of Amazon EC2.
- Choose instances to option from the navigation pane.
- Choose the desired instance followed by Actions, then Monitoring and then Enable Monitoring.
- A dialog box will pop up, select enable.
- After you are done, the instances will be monitored for a 1-minute period.

While launching any instance from the Management Console, select the option of Monitoring Check.

Metrics

Metrics is all about the data about the performance of the systems. Several services provide the users with free metrics for the resources by default such as EBS, EC2, and RDS DB. You can enable detailed monitoring for some of the resources by yourself such as for the instances of EC2 or you can also publish the metrics of your own application. CloudWatch can help by loading all the metrics data in your own account for the purpose of graphing, searching and alarms. CloudWatch loads all the metrics that include resource metrics of AWS and the application metrics provided by you. The data of the metrics can be recorded for 15 months that enables the users to check updated data along with historical data of the applications.

How to view available metrics?

The metrics are first grouped by namespace and then it is grouped by a different combination of the dimensions within each of the namespace. You can view all the metrics of EC2, the metrics of EC2, which are grouped by the instances, and the metrics of EC2 grouped by the Auto Scaling group. For viewing the available metrics by dimension and namespace, you need to follow these steps:

- Open the console of CloudWatch.
- Choose Metrics from the navigation pane.
- Select the namespace for the metric.
- Select the dimension of the metric.

The All Metrics tab will display all the available metrics for the namespace dimension. You can also perform the following functions:

- You can sort down the table by using the column heading.
- You can also graph a metric by selecting the checkbox, which is placed right next to a metric. For selecting all the metrics, you need to select the checkbox situated in the row header of the table.
- You can filter out the resources by simply choosing the resource ID.
- You can also filter out the metrics by choosing the name of the metric and then select search.

Logs

You can use CloudWatch logs for monitoring, accessing and storing the log files from EC2, Amazon Route 53, CloudTrail and various resources. With the help of CloudWatch logs, you can centralize all the logs from all your available applications, AWS service and other services that you are using for a scalable service. You can easily view such logs, search for specific patterns or error codes, archive them for analysis in the future or filter them on the basis of particular fields. You can view all the logs, regardless of the source. You can also sort the logs and query the same based on different dimensions.

Features of CloudWatch logs

- You can monitor all the logs from EC2 instances. CloudWatch logs can easily track the total number of errors that occurred in the application and then send the user the required notification when the error rate exceeds the specified threshold. There is no requirement of code change as CloudWatch logs use the data of logs for monitoring. You can also monitor the logs for particular terms or count the total occurrence number of a term at any particular position of the log data.

- You can easily monitor the logged events of CloudTrail. You can set the alarms and receive particular notifications on API activity and then use the notification for troubleshooting various problems.
- The logs are kept in an indefinite form that never expires. This setting is by default. You can easily adjust the policy of retention for each of the log groups by either choosing an indefinite detention or retention period of one to ten days.
- You can store all the log data in durable storage. This helps in accessing the required data whenever required.
- CloudWatch Logs can be used for logging various information about the queries of DNS that is received by Route 53.

Chapter 9: Notification Services

SNS

The notification service from AWS is known as Simple Notification Service or SNS. It is a very durable, available, highly secure and fully managed messaging service that allows the users to decouple microservices along with the serverless applications. SNS from Amazon provides various topics for push-based, high-throughput, many too many messaging. With the help of Amazon SNS, the publisher systems of the users can easily fan out their messages to large endpoint numbers of subscribers for the purpose of parallel processing along with AWS Lambda functions, Amazon SQS queues, and HTTP/S webhooks. Also, you can use SNS for fanning out various notifications directly to the end-users by using mobile email and SMS. The users can easily get started with SNS service with the help of CLI, Management Console or SDK or Software Development Kit.

Benefits of SNS

The Amazon SNS comes with various benefits that are surely going to help your applications.

- **Reliable delivery of messages with high durability:** SNS provides high durability of messages by using the storage of cross availability zone messages. The topics of Amazon SNS are available for the applications whenever you require them as it runs within the proven network datacenters and infrastructure of Amazon. The messages, which are published, to SNS are stored across numerous servers along with data centers that are separated geographically. SNS delivers all your messages reliably to all the AWS endpoints that are valid such as AWS Lambda functions and Amazon SQS queues.
- **Automatic scaling of workload:** SNS clouts have a proven cloud of AWS for dynamically scaling with the application. It is a service that is managed for taking complete care of the heavy lifting associated with provisioning, capacity planning, patching,

and monitoring. The whole service has been designed in such a way so that it can handle bursty patterns of traffics along with high-throughput. The best thing about Amazon SNS is that there is no requirement of upfront cost along with no requirement of installation, configuration or up-gradation of the messaging software.

- **Simplify the architecture with the help of filtering:** SNS helps the users to simplify their architecture of pub/sub messaging simply by offloading the filtering logic for messages from the systems of the subscribers and routing logic of messages from the publisher systems. With the help of SNS filtering, the subscribed endpoints will receive only those messages, which are of their interest in place of all other messages, which are published for the topic. CloudWatch provides visibility for the activity of filtering of the messages. CloudFormation enables the users for deploying subscription policies related to filtering in a secure and automated manner.
- **Keep all the messages secure and private:** The owners of Amazon SNS topics can secure various sensitive data simply by setting policies for the topics. The topic policies restrict access to who all can publish along with subscribing to the topics. SNS makes sure that all the data is in encrypted form while transit by applying certificates of Amazon ATS for supporting the HTTPS API. It can also easily encrypt those data, which are at the state of rest with the use of KMS keys. Additionally to this, the user can also publish messages privately to the topics of SNS from the subnets of Amazon VPC without even crossing the public internet. It can support the use cases which are available in the markets, those are regulated and those which are in extension with the various programs of compliance, along with PCI, ISO, HIPAA, SOC, FedRAMP and FIPS.

Push Notification Services

AWS offers the users with various services with which push notification can be sent to the applicants. The platform that you use will depend completely on which app store your customers handle the most for getting your app. The most common push notification platforms are APN or Apple Push Notification Service, Google Cloud Messaging, Baidu Cloud Push, ADM or Amazon Device Messaging and Firebase Cloud Messaging.

Apple Push Notification or APN service

APN was the first service related to push notifications, which were designed, for allowing the various third-party developers to send out messages directly to the users. The application developers can use the service of APN for sending out notifications to the users of mobile apps on all iOS devices. APN can also send out notifications to the apps of macOS desktop.

Google Cloud Messaging or GCM service

With the help of the GCM service, the users can easily send out notifications to the Android app users. All the third-party developers can use this service for sending out direct notifications to all the android app users.

Baidu Cloud Push

Google Play Store is no more available in China. Therefore, it is not possible to send out push notifications from Google’s platform to all the users of Android devices in China. That is Baidu Cloud Push was launched for sending out push notifications to all the users in China who downloads your app from the app store of Baidu.

Amazon Device Messaging or ADM service

The Amazon Kindle Fire Tablets uses Android as its primary OS. However, it uses up the app store of Amazon itself, pushing aside the Google Play Store. You can easily use the Amazon Device Messaging or ADM service for sending out push notifications to all the users who use the app store of Amazon for your app. ADM is supported by all the devices of Kindle.

Firebase Cloud Messaging

Firebase is a company that serves as a mobile and web application development platform. FCM or Firebase Cloud Messaging is a component of Firebase that acts as a messaging service for all the Android apps.

Additionally, FCM also provides the users with an interface that allows the developers to send out push notification messages to the iOS apps. It was later purchased by Google in the year 2014.

Chapter 10: Database Services

AWS provides users with the largest selection of databases that have been built according to the purpose of the users for their application needs. You can choose from the 14 databases which are purpose-built that also includes key-value, relational, in-memory, ledger, time series and graph databases. The database service from AWS supports the models of diverse data and allows the users to build up highly scalable, use case driven and well-distributed applications. You can pick up the best database for solving some specific or group of problems and break apart from the monolithic style of databases. You can focus on application building for meeting the business needs with the perfect database of your choice.

Benefits of using AWS Database services

- **High-performance scale:** You can get relational databases that are three to five times faster than the various popular database alternatives. It will provide you with a latency of sub-millisecond or microsecond. You can start small and scale the performance with the growth of your applications. You can scale the storage resources and compute the databases with minimal or no downtime.
- **Management:** With the AWS database service, you are free from the worry of database management such as patching, server provisioning, configuration, setup, recovery or backups. AWS monitors the clusters continuously for keeping up the workloads and running with auto-healing storage along with automated scaling. It allows you to focus on the higher values of your application development.
- **Enterprise-class:** The databases served by AWS have been built for the enterprise and business-critical workloads that offer a higher percentage of reliability, availability along with high security. The databases can easily support multi-master and multi-region applications and provide complete oversight of user data with different security levels. It also comes with the feature of network isolation, which is carried on with the help of Amazon VPC and encryption at rest by using the keys created by

the user.

SQL and NoSQL

The application developers need to deal with two of the most common things: RDBMS and SQL. NoSQL is the term that is used for describing the nonrelational database systems, which are highly available. It is highly scalable as well and can be optimized by the user for the highest level of performance. In place of the relational model, the NoSQL databases such as DynamoDB use various alternate models for the management of data such as document storage or key-value pairs.

Which one to choose: SQL or NoSQL?

The applications of today come with more demands in requirements when compared to the applications of the past. For instance, an online game starts out with only a few users along with a very little amount of data. However, when the game becomes highly successful, it can very easily outstrip all the resources of the underlying system of database management. It is a very common thing for the applications based on the web to have thousands or millions of circumstantial users with terabytes of new data, which is generated every day. The NoSQL databases such as DynamoDB are better suited for managing these types of workloads. The application developers can begin with a small amount of throughput, which is being provisioned, and then gradually increase the same as the application becomes popular with time. NoSQL databases can seamlessly handle higher amounts of data along with a huge number of users.

The relational databases or SQL are divided into various sets of rows and columns for storing data, which are often called tables. The NoSQL databases are more document-oriented with distributed storage that functions without any type of structured table. SQL databases come with a pre-defined schema that is well designed for better functioning whereas the NoSQL databases come with a dynamic style of the schema. You can scale the SQL databases vertically whereas the NoSQL databases can only be scaled horizontally. The SQL databases can be easily scaled by increasing the hardware strength, which is not the case with NoSQL databases. For scaling the NoSQL databases, you need to expand the servers of the database within the pool of assets for reducing the burden.

For the purpose of defining the data, SQL databases use the structured language of the query but NoSQL databases unstructured query language, also known as UnQL. NoSQL databases come with the hierarchical style of data storage, which is not the case with SQL. The users can easily add up new data in NoSQL without the requirement of any prior steps. But, SQL requires some changes like altering the schemas or backfilling the data for adding up new data. SQL comes with a standard interface, which is a great option for dealing with complex queries that are not possible with NoSQL as it lacks in any form of the standard interface.

Why opt for SQL?

SQL is a great option for protecting the integrity of the databases with the help of ACID compliance. As it comes with a structured style of data, you will not be requiring any support for an integrated system for dealing with different types of data. It is a highly preferable database option for businesses due to its predefined schemas and structure.

Why opt for NoSQL?

The NoSQL databases are gaining popularity day by day as it allows the users to store various types of data altogether. It also makes the scaling process easier by spreading across various servers. If you need to develop an application within a fixed amount of time, opting for the NoSQL database is the best option as it will speed up the performance with the help of the rapid development phase.

Bottom line

Each and every business comes with its own set of preferences, which are based on the type of project requirements. Each of the databases comes with their specific style of functioning that you can choose according to your needs. Therefore, it is very important to specify your requirements before opting for any of the two databases for the development of your applications.

Relational Database Services (RDS) and DynamoDB

AWS database services come with a wide range of database choices that you can choose depending on your application requirement.

RDS or Relational Database Services

RDS was also has known as relational Database Service is the cloud-hosted solution from Amazon which is managed by RDBMS. With the usage of RDS, users are not required to install, configure or manage the various relational system of databases like Microsoft SQL Server, Oracle, MariaDB, MySQL or PostgreSQL. With RDS, you can spin any of the database instances that you choose with a minimal amount of input from your side. In simple words, the users need to make some fundamental choices including:

- Which database software will they like to install for their applications.
- The overall capacity of the database instances like RAM, CPU and disk space.
- The master password and username that they want for the database instances.
- Schedules of backup and preference of maintenance.
- Any settings of the non-default parameter of configuration.
- The VPC or network and the region where the database instance is supposed to run.

Once the user chooses the required options, RDS plans an instance of the database and changes any required settings in the configuration. After all of these are done, RDS makes the database instance completely available for the user. The users will not be able to access the underlying host directly as RDS is a managed solution. This is also because of the fact that there is no form of SSH access or remote desktop at the system level of operation. AWS takes care of all the process of patching, installation, security, maintenance, snapshots, failover, etc. The users have the choice of either bringing own licenses for the software of the database or buy the database software license as a part of the overall instance cost.

With the facility of automated backup that comes with RDS, it is possible for the users to restore any instance within a period of 5 minutes at any point in time during the retention period of backup. The overall retention can date back to the maximum last 35 days. The best feature that makes RDS so popular is the ability of scaling. RDS instance can be as limited as having only 2 GB of RAM with 1 vCPU or as large as having 488 GB of RAM and 64 vCPUs. In case, any of the instances require more power, it is possible to easily upgrade the same to any high-end server without any kind of hassle. The storage, which is in the underlying state, can be made ready to perform for a specific number of I/O or Input/Output in a second with the provisioned IOPs. Achieving this level of scalability in any form of a traditional data center will have cost as well as time intensive unless it is prohibited.

Use cases

- Back-end database for the applications on the web
- Data marts and small warehouses of data
- Back-end database for the application of enterprise
- Source system for the warehouses of data

DynamoDB

DynamoDB is a database service designed by Amazon, which is of the NoSQL category. It has been designed for faster processing of small amounts of data, which changes and grows dynamically. It is non-relative in nature. The main feature of DynamoDB is its unstrict table structure, which consists of attributes and items. The mutability of database and faster rate of I/O is powered by the use of an SSD as the only hardware for storage.

When it comes to DynamoDB, there are no instances of hardware on which the billing and capacities depend upon. The primary value is the throughput of reading/write which is used by the database. The best part is that there is no limit on the storage of resources. The storage grows in size as the database also grows without any type of instance replication or other types of cloud scaling. The multi-AZ feature, which you require to pay a fee with RDS, comes with the box with DynamoDB. The data is replicated automatically among 3 AZ or availability zones within a region selected by the user. DynamoDB becomes super durable due to the absence of replication of data, activities of administration and scaling models of final-performance. However, DynamoDB cannot support the functions, which are complex in nature such as advanced transactions and queries. As the data is partitioned in

DynamoDB for the durability, the re-writing process takes a lot of time in each replica after the successful writing operation in the main one. Read Consistency is the ratio between the capacities of writing and reading.

- The option of Consistent Reads gives the overall priority to the operation of reading which forwards the data in case it is already modified but has not been replicated to the local availability zone. This option helps in speeding up the performance of reading but the requests of reading need to be performed again for getting the updated data.
- The option of Strongly Consistent Reads targets at getting the data that has been updated and is the latest. It takes up more amount time but returns a result that reflects the successful writings which have been made right before the initialization of reading.

Use cases

The NoSQL databases are not used for applications based on the web or for a modern cloud system. It can be used for storing up the preference of the users, streaming data and gaming software.

- Processing and systemization of the data blocks

- Gaming: World changes, high-scores, statistics, the status of the player, etc.
- Advertisement services: Collecting data from the customer base, creating trend-charts, etc.
- Blogging and messaging: Building up the blog list entries of the author, message selections, etc.
- Other cases where the processing of data is required instead of just storing the data. The data needs to be highly available instead of just being available for the transaction.

ElastiCache and RedShift

ElastiCache

The technique, which is used for storing the information, which is accessed frequently in a temporary location on the server, is known as caching. In this world of today, which is driven by the web, catering to the requests of the users within a fixed time is the one and only goal of the websites. For delivering the requests of the users within time, speed and performance are required. That is why the caching layer like the ElastiCache from Amazon is the tool that is used by the websites for serving the most frequently accessed and static data. With ELastiCache, you can store all the frequently accessed HTML pages, information and images.

ElastiCache is a web service for caching from AWS. This service from AWS simplifies the task of setting up, scaling and managing an environment of in-memory cache, which is distributed within the cloud. It comes with a highly scalable, great performance and cost-effective solution for caching. ElastiCache is capable of removing all the complexities, which are associated with the deployment and management of a well-distributed environment of the cache. ElastiCache comes with various features that can easily enhance the reliability of various critical deployments of production along with:

- Automatic recovery and detection from the failures of the cache node.
- Automated failover of a primary cluster, which has failed, to a replica, read in the replication groups of Redis.
- Flexible placement of availability zones for the clusters and nodes.
- Integration with the various AWS services such as CloudWatch, EC2, SNS, and CloudTrail for providing a caching solution that is secure and is capable of high performance.

The ElastiCache service from Amazon provides two engines for caching, Redis and Memcached. You can shift your already existing Redis or Memcached implementation of caching to ElastiCache without any kind of effort. All you need to do is to just change the Redis/Memcached endpoints in the application.

ElastiCache Node

ElastiCache nodes are the smallest blocks of the architecture of ElastiCache service. The nodes are nothing but network-attached RAMs.

ElastiCache Cluster

The clusters in ElastiCache are the logical collection of nodes. If the ElastiCache cluster is having Memcached nodes, you can then have several nodes in various AZs or availability zones for implementing high-availability. However, in the case of the Redis cluster, it is always a single node. As a user, you can have various replication groups across the availability zones. The Memcached cluster comes with various nodes of which the cached data is partitioned horizontally across each and every node. All the nodes in a cluster are capable of write and read. Redis cluster comes with one node which acts as the master node. It also does not support the partitioning of data.

ElastiCache Memcached

It is a very simple model for caching. Memcached is very helpful for those who are in need of running large nodes with various threads or cores. You can also scale out or scale in various nodes according to the demand and requirement. It allows users to handle data partitioning across several shards. Memcached handles objects of cache such as a database. The nodes in the Memcached cluster come with an individual endpoint.

Elasticache Redis

Redis is suitable for supporting various complex types of data such as hashes, strings, sets, and lists. It is capable of ranking the in-memory sets of data. You can also get persistence for the key store. It is responsible for replicating the cached data from primary to more than one read replicas just for making the applications read-intensive. In case the primary node fails, Redis comes with the capability of automated fail-ver. It also comes with restore and backup capabilities.

Redshift

Redshift is used for extending the queries of the data warehouse to the data lake of the user with no requirement of loading. You can easily run a request for analytic queries for the huge piles of data that are stored in Redshift and also directly against the huge pile of data stored which are stored in S3. It is very easy to use setup and also automates most of the administrative tasks along with the delivery of fast performance at any required scale.

Parallel

Redshift can easily delivery query on the datasets which ranges in size of gigabytes to exabytes. It uses data compression, columnar storage and zone mapping for reducing the I/O amount, which is required for performing the queries. Redshift uses MPP or massively parallel processing architecture of the data warehouse for distributing and parallelizing the operations of SQL in order to take full advantage of all the available resources.

Machine learning

Redshift uses machine learning for delivering high throughput, which is irrespective of the concurrent usage or workload. It utilizes various sophisticated algorithms for predicting the run times of incoming queries and then assigns all the queries in a queue for faster processing.

Result caching

It uses result caching for delivering sub-second time of response for the repeated queries. When a query is executed, Redshift searches the cached data to check if there is any cached result from the previous run.

Automated backup

Redshift continuously backups your available data into Amazon S3. It can also replicate the snapshots to Amazon S3 to another region for the purpose of disaster recovery. By using the management console, you can use any system for restoring the cluster.

Automated provisioning

Redshift is very easy to set up and also operates easily. You can deploy new warehouse data with some simple clicks in the console of AWS and Redshift will automatically plan out the infrastructure for you. Most of the tasks of administration such as replication and backup are automated for allowing the user to focus more on the data and not in administration. You can also take control in your own hands with the help of the options provided by Redshift for making necessary adjustments for tuning the specific workloads. The new capabilities are transparently released which eliminates the requirement of scheduling and applying the patches and upgrades.

Chapter 11: Serverless Compute: Lambda

Lambda is a service of serverless computes from AWS that runs the codes of the users in reply to the events. It also manages automatically the underlying resources of computing. AWS Lambda can be used for extending the other services of AWS with customized logic or by creating own back-end services that will operate at the AWS security, scale, and performance. Lambda is capable of running the codes automatically in response to various events such as requests of HTTP via API Gateway, updates in the table in DynamoDB, state transitions Step Functions and object modification in S3 buckets.

Lambda will run all your codes on computing infrastructure, which is of high-availability and performs all related administrative tasks of the resources. It also includes maintenance of operating system and server, automatic scaling, capacity provisioning, deployment of security patch and code and logging and monitoring of code. You are only required to supply your code.

Lambda functions

The code, which is run by the user on AWS, is known as Lambda Function. After the creation of the Lambda function, it is ready for running as soon as it is triggered. It is somewhat similar to the spreadsheet formula. Each of the functions will be including your code along with some associated information on configuration. It also includes the requirements of resource and function name. The Lambda functions are in stateless condition with no sign of inclination to the underlying infrastructure. This allows Lambda to quickly launch as many numbers of copies of the function, which is required for scaling the overall rate of the incoming events.

After you are done with uploading the code to Lambda, you have the facility to associate the function with any specific resource of AWS such as S3 bucket, SNS notification or DynamoDB table. After the resource has been changed, the function will be executed by Lambda and will also manage the resources of computing as required for keeping up with the incoming requests.

Features of AWS Lambda

Extension of AWS services with the help of custom logic

With AWS Lambda, you can easily add up custom logic to the AWS resources such as DynamoDB tables and S3 buckets. It makes it easier for applying to compute to the data as it moves through or enters the cloud. It is a very easy task to get started with Lambda. All you need to do is to first create the function simply uploading the code or you can also build the code in the console of Lambda. Then you need to choose the timeout period, memory and AWS IAM or Identity Access Management role. After that, you need to specify the required AWS resource for triggering the function, which can be either an S3 bucket, Kinesis stream or DynamoDB table. After the resources are changed, Lambda will start to run your function and will launch as well as manage the computing resources.

Building up of customized back-end services

AWS Lambda can be used for creating new back-end services for the applications that have been triggered using Lambda API or customized endpoints of API, which has been built by using API Gateway from Amazon. You can easily avoid the platform variations of clients, enable easy updates and reduce draining of battery by using Lambda for the processing of the custom events in place of directly servicing them on your clients.

The automated system of administration

Lambda single-handedly manages all infrastructures for running your code on the fault-tolerant and highly available infrastructure. This makes you free to completely focus on the building up of differentiated services related to the back-end. Lambda comes with an additional facility where you are not required to update your underlying operating system when the patch is released. There is no need to worry about addiction or the resizing of new servers with the growth of usage as well. Lambda deploys the code seamlessly, takes care of the maintenance, administration and security patches. It also provides built-in monitoring and logging with the help of CloudWatch.

A built-in feature of fault tolerance

Lambda comes with a built-in feature of fault tolerance. Lambda maintains the capacity of computing across various AZs or availability zones in every region for protecting the code from data center facility or individual machine failures. Lambda along with the functions running on it provides the reliable and expected performance of the operation. It has been designed for providing high availability for its service and the operational function. There is no scheduled feature of downtimes or maintenance windows.

Automated scaling

Lambda conjures the code only when it is required and scales it automatically for supporting the income request rate without any kind of manual intervention. The number of requests that can be handled by the code has no limit. It starts to run the code within microseconds of an event. As Lambda comes with the auto-scaling feature, the rate of performance remains high consistently with an increase in the frequency of events. As the code used in Lambda is stateless, it can start with as many numbers of instances it requires without any kind of delay in configuration and deployment.

Coordination of multiple functions

By building up workflows with the help of AWS Step Functions, you can easily coordinate several functions of Lambda for the long-running or complex tasks. Step Functions allow the users to define the workflows that, in turn, triggers a large collection of Lambda functions by using parallel, sequential, error-handling and branching steps. With Lambda and AWS Step Functions, you can build up long-running and stateful processes for the applications.

Security model

Lambda allows the user code to access the other AWS services securely with its in-built AWS SDK along with integration with AWS IAM or Identity Access Management. Lambda runs the code within a secure VPC by default. As a user, you can configure Lambda for accessing the various resources behind your VPC that will allow you to clout the custom groups of security.

Chapter 12: Security and Compliance

Security and Compliance Services

AWS provides various organizations with security and compliance services for uninterrupted services. Security and compliance services can be linked with other AWS services as well. The cloud computing security service is a very rapid growing service that functions similarly to the traditional style of IT security. This includes the protection of all the critical information from data leakage, deletion, and theft.

Security

Security is regarded as the highest priority service of AWS. As a user of AWS services, you will be gaining profit from a data center and a network architecture, which has been built to meet all the requirements of the highest organizations, which are security-sensitive. While using the cloud, securing it from all possible sides is of utter importance. Cloud security is somewhat like the security of your on-premises data center. It is of much more important than the cost of the hardware and maintenance facilities. You are not required to manage the storage devices or physical servers in the cloud. Instead of doing it yourself, you use security tools that are based on software for protecting and monitoring the overall flow of information that goes in and out of the cloud resources.

One of the most interesting and advantageous features of AWS Cloud is that it allows the users to scale along with innovating. It can be done while maintaining a super-secure environment, without paying for those services that you do not use at all. This means that you can enjoy premium security and that too at a lesser cost when compared to the security of your on-premises data center environment. While using the services from AWS, you need to inherit all the practices regarding AWS policies, operational processes, and architecture which have been built for satisfying the ultimate requirements of those customers who are the most security-sensitive. You can enjoy the agility along with flexibility with security services from AWS for your data centers.

AWS Cloud comes with a shared responsibility model. While the security of the cloud is managed by AWS, you are the one who is completely responsible for the security in the cloud. In simple words, you can retain the security control which you choose for protecting your content, applications, platform, networks, and systems in the same way that you would have done for an on-premises data center. You will also receive the required guidance and competence via online personnel, resources, and partners. AWS provides users with various advisories for current issues. You can also enjoy the opportunity to work along with AWS whenever you come across any kind of security issue.

For the purpose of meeting up with your objectives regarding cloud security, you also get access to various features and tools. AWS provides the users with tools and features specific to security across network configuration, access control, management and encryption of data. The AWS environments are audited at regular intervals with certifications from various accreditation bodies across verticals and geographies. You can take full advantage of the automated tools in the AWS environment for the purpose of access reporting and asset inventory.

Benefits of AWS security

AWS security comes with various benefits that can make your cloud working much more convenient and secure.

- It helps in keeping all your data safe. The infrastructure of AWS puts in place strong safeguards for protecting the privacy of the users. All of your data is stored in super-secure data centers of AWS.
- You can meet compliance requirements with AWS security. AWS is known for managing various programs of compliance within its infrastructure. In simple words, the segments of your required compliance have been completed already.
- You can save lots of money with AWS security services. You can cut down your costs with the help of AWS data centers. You can maintain the highest security standards without any need of managing your facility.
- You can scale quickly and conveniently with AWS security services. The security services also scale with the usage of your

AWS Cloud. Whatever may be the size of your business, the infrastructure of AWS has been designed for keeping all your data safe.

Infrastructure security

AWS provides users with various capabilities in security and services for improving the control and privacy of network access. These also include:

- Customer-controlled encryption, which is in transit with all the TLS across all the services of AWS.
- The network firewalls which are built into VPC from Amazon and firewall capabilities of the web applications in AWS WAF lets you create various private networks and also control access to the applications and instances.
- Options of connectivity that enables dedicated or private connections from your dedicated on-premises environment or office.
- Automated encryption of all your application traffic on the regional and AWS global networks between the secured AWS facilities.

Encryption of data

With the AWS security service, you can get the facility of adding an extra security layer to all the data at rest inside your cloud along with providing efficient encryption and scalable features. This also includes:

- Flexible options of key management along with AWS KMS or key management service. This service will allow you to choose if you want AWS to manage all the encryption keys or you want to have complete control over the keys.
- Capabilities of data encryption available in database and AWS storage services such as S3, EBS, Glacier, SQL RDS, Redshift and Oracle RDS.
- Queues for encrypted messages required for the transmission of highly sensitive data using SSE or server-side encryption for

SQS.

- Dedicated cryptographic key storage based on hardware using CloudHSM that will allow you to satisfy all your compliance needs.

Additionally, AWS also provides APIs for integrating data protection and encryption with any kind of service that you deploy or develop in the AWS environment.

Configuration and inventory

AWS offers the users with a wide range of useful tools that will allow you faster movement while still safeguarding the resources of your cloud with the best practices and organizational standards. This service includes:

- Various tools for deployment and for managing the decommissioning and creation of the AWS resources according to the standards of the organizations.
- Amazon Inspector, which is a security assessment service from AWS that assesses all the applications automatically for deviations or susceptibility from the best practices that also include OS, impacted networks and attached storage.
- Tools for configuration and inventory management along with AWS Config that will identify the resources of AWS and then manage and track all the changes to all those resources with time.
- Tools for template management and definition along with AWS CloudFormation for creating preconfigured and standard environments.

Logging and monitoring

AWS security provides you with various features and tools with which you can have a complete visualization of what is happening in the environment of AWS. This also includes:

- Options for log aggregation, compliance reporting and streamlining the investigations.

- Usage of AWS CloudTrail for having deep visibility into the API calls that also includes who, when, what and from where all the calls were made.
- Alert notifications with the help of Amazon CloudWatch whenever thresholds exceed or any specific event occurs.

All these tools will provide you with the ultimate visibility that you need for spotting any kind of issues right before they can impact the functioning of the business. It also allows the users to improve the posture of security along with reducing the risk of the environment.

Access and identity control

AWS security services offer you all the capabilities for defining, managing and enforcing the policies of user access across all the services of AWS. This also includes:

- AWS IAM or Identity and Access Management for defining the respective user account with proper permissions across the resources of AWS.
- AWS Directory Service that allows you to federate and integrate with the various corporate directories for improving the experience of end-user and reducing the overhead of administration.
- AWS Multi-Factor Authentication for all the privileged accounts along with options for the authenticators, which are based on hardware.

AWS provides the customers with native access and identity management integration across most of its services that also includes the integration of API with any of your services or applications.

Penetration testing

AWS regularly tests its own infrastructure, the results of which can be found in the compliance reports. The customers of AWS security services that easily carry out assessments of security and penetration testing against the AWS infrastructure of their own without any kind of prior approval for the core number of services.

DDoS Mitigation

Availability is of preeminent importance for the cloud. The customers of AWS benefits from the services of AWS and technologies, which has been built for providing flexibility during the time of DDoS attacks. With a combination of the AWS security services, it is possible to implement a well-built defense with an in-depth strategy for countering the DDoS attacks. The services, which have been built for DDoS help, help in minimizing the overall time of mitigation and also reduces the impact.

AWS Compliance

With the help of AWS compliance, you can easily understand the various robust controls, which are in place at AWS for maintaining the overall data protection and security in the AWS cloud. As the systems or applications are, built on top of the infrastructure of AWS Cloud, the responsibilities of compliance need to be and will be shared. By tying altogether the audit-friendly and governance-focused service features along with applicable audit or standards of compliance, the enablers of AWS Compliance builds upon the traditional programs. This, in turn, helps all the customers to operate and establish in an environment that is controlled by AWS security.

The IT infrastructure provided by AWS to its customers is managed and designed in alignment with the best practices of security along with a variety of standards in IT security. The following are the partial assurance programs, which are satisfied by AWS:

- FISMA, FedRAMP, and DIACAP
- SOC 2, SOC 1/ISAE 3402 and SOC 3
- ISO 9001, ISO 27017, ISO 27001 and ISO 27018
- PCI DSS Level 1

AWS provides its customers with a wide variety of information on the IT control environment in reports, whitepapers, accreditations, certifications and other types of third-party attestations.

Attestations and Certifications

The certifications of attestations and compliance are determined by an independent, third-party auditor, which results in an audit report, certification or compliance attestation. The auditors include:

- C5
- ASIP HDS
- DoD SRG
- FIPS
- ISO 9001
- ISO 27017
- ISO 27001
- ISO 27018
- TISAX
- K-ISMS
- PCI DSS Level 1

Privacy and regulations

The customers of AWS remain responsible for adhering to the compliance regulations and laws. In some of the cases, AWS also offers enablers, functionality and legal agreements such as AWS Business Associate Addendum for supporting the compliance of the customers. No kind of formal certification is feasible to the providers of cloud service within these regulatory and law domains.

- CCPA
- CLOUD Act
- CISPE
- FERPA
- GLBA
- HIPAA
- GDPR
- ITAR

- VPAT / Section 508
- PIPEDA
- PHIA
- PDPA – 2010
- Privacy Act (New Zealand)
- IRS 1075

Frameworks and alignments

The frameworks and alignments of compliance include compliance or published security requirements for a particular purpose such as a particular function or industry. AWS provides users with enablers and functionality for these kinds of programs. The requirements under particular frameworks and alignments might not be subject to attestation or certification. However, some of the frameworks and alignments are covered under other programs of compliance.

- CJIS
- CIS
- CSA
- FFIEC
- EU-US Privacy Shield
- FISMA
- ICREA
- MPAA
- NIST
- MITA 3.0
- UK Cloud Security Principles
- G – Cloud
- Uptime Institute Tiers

AWS Shared Responsibility Model

AWS functions with a shared responsibility model that includes security and compliance. Security and compliance is the shared responsibility model between the services of AWS and the customers. The shared responsibility model might help a lot in relieving the customers from the burden of operation as AWS manages, operates and controls all the components from the layer of virtualization and operation system right to the overall physical security of those facilities in which the AWS services are operated. The customer needs to be responsible and manage the guest OS that also includes security and update patches, other linked software applications along with the configuration of the security firewall group, which is being provided by AWS.

All the customers need to consider all the services that they choose. This is because, the responsibilities of the customers vary depending on the services they use, applicable regulations and laws and the integration of all the used services into the IT environment of the customers. This nature of the shared responsibility model provides the ultimate flexibility in usage along with customer control that, in turn, permits service deployment. In simple words, the shared responsibility model is also referred to as the security of the cloud in contrast to the security in the cloud.

Responsibility on the part of AWS: Security of the Cloud

AWS is solely responsible for the overall protection of its infrastructure that runs all the services, which are offered, in the AWS cloud. The infrastructure consists of software, hardware, networking along with all the facilities that run the services of AWS cloud.

Responsibility on the part of the customer: Security in the Cloud

The responsibility on the part of the customer is determined by the services of the AWS cloud that is selected by the customer. This ultimately determines the overall amount of configuration work that the customer needs to perform as a part of the security responsibilities. For instance, any service of AWS such as Amazon EC2 or Elastic Compute Cloud is being categorized as IaaS or Infrastructure as a Service and requires performance on the part of the customer regarding all the necessary management tasks and configuration of security. Customers who deploy EC2 instance from Amazon are liable for the complete management of the guest OS or operating system that also includes the security and update patches, configuration of the firewall provided by AWS which is known as security group and any application utilities or software which is being installed by the customer on the EC2 instances.

For the services of AWS, which are abstracted, such as DynamoDB or Amazon S3, the infrastructure layer is operated by AWS itself along with the platforms. The customers can easily access the endpoints for the purpose of retrieving and storing the data. The customers are solely responsible for the management of their data that also includes the options of encryption, classification of the assets and use of IAM tools for application of the required permissions.

This AWS and customer shared responsibility model also extends itself to the IT controls. Just like AWS and its customers share the responsibility of the operations in the IT environment, the operation, management along with verification of the IT controls are also shared between the two. AWS can help the customers by relieving the burden of the customers to operate the controls by managing all the required controls, which are associated with the infrastructure, which is deployed in the environment of AWS that might have been managed, by the AWS customers previously.

As each and every customer is deployed in different ways in AWS, the customers have the opportunity to shift the overall management of some specific IT controls to the AWS which ultimately results in a brand new environment of distributed control. Then the customers can use the controls of the AWS along with documentation of compliance that are available to them for performing the evaluation of control and procedure of verification as needed. The following are the examples of the controls, which are managed by the AWS, customers of AWS and/or by both.

- **Inherited controls:** These are the controls that are inherited by the customers from AWS fully. It comes with environmental as well as physical controls.
- **Shared controls:** These controls are applied to both the customer layers and the infrastructure layer. It is being applied in a completely different perspective or context. In shared control, AWS provides all the objectives required for its infrastructure and the customer is required to provide their own implementation of controls within the use of their required services from AWS. The examples are:

1. **Patch management:** AWS is solely responsible for the

fixing and patching of flaws within its infrastructure whereas the customers are responsible for the patching of their own databases, applications and guest operating system along with its applications.

2. **Configuration management:** AWS is responsible for maintaining the configuration of the devices of infrastructure whereas the customers are responsible for the configuration of their own databases, applications and guest operating system along with its applications.
3. **Training and awareness:** AWS is responsible for the training of its employees whereas the customers are responsible for the training of their employees.

- **Customer-specific:** This includes all those controls, which are completely the responsibility of the customers, which is based on those applications that they are deploying within the services of AWS. The examples are:

    1. Communication and Service Protection or Security of the Zone, which might require the customers to zone or route the data within the particular environments of security.

To sum it up, Amazon, like most other cloud service providers, focuses on the overall security of its offering in the cloud. When the customers begin with the usage of the AWS services, AWS shares the overall responsibility of data security and compliance with the customers that ultimately makes the security and compliance of AWS a shared responsibility.

AWS Key Management Service

AWS KMS or Key Management Service is a service from Amazon that makes it easier for the users to create and then manage the keys. It also allows the users to control encryption usage across a wide variety of AWS services and in the applications. It is a very volatile and secure service that uses up the security modules of the hardware, which has been validated under FIPS 140 – 2 or is in the validation process for protecting the keys. KMS is integrated with CloudTrail for providing the users with all the key usage logs for meeting all kinds of compliance and regulatory requirements. With AWS

KMS, you can easily add the functionality of encryption to your code of applications either through direct encryption and decryption service APIs or through the integration with AWS SDK.

Features of KMS or Key Management Service

The AWS LMS comes with a variety of features that can help you a lot with your AWS services.

Centralized Management of Keys

With AWS KMS, you can have centralized control of all the encryption keys. CMKs or Customer Master Keys are being used to control the access of the data encryption keys that helps in encryption and decryption of your data. As a user, you have the facility to create new master keys, whenever you want. You can also manage access to the master keys. You can select the services with which the master keys can be used. If you want, you can import the keys from your key management infrastructure directly into AWS Key Management Service or else use the leys stored up in the cluster of CloudHSM and then manage the same using AWS KMS. AWS KMS provides the users with a facility with which they can manage the master keys and audit its usage from Management Console or by the use of SDK or CLI.

The keys in KMS, which are whether created by the user within KMS, cluster of CloudHSM or imported by the user himself, are stored up in highly secure and durable storage in an encrypted format so that the keys can be used only when required. You can select to automatically rotate the master keys, which are created within KMS, once every year without any requirement of re-encrypting the data, which has been encrypted with the master key already. Also, you are not required to keep detailed track of the old master keys.

AWS Service Integration

AWS Key Management Service is integrated with most of the services from AWS. Integration with KMS means that the user can easily use the master keys of KMS for controlling the encryption of the stored data within the services. While you decide to encrypt your data in any service, you can select a master key which is managed by AWS and controlled by KMS automatically. You can also track key usage but it is managed by the KMS service on the behalf of the user. In case you require direct control over the master key lifecycle or you want the other accounts to use the same as well, you can easily create and manage the master keys that you can use for the AWS services. The customer controlled master keys come with full control over the permissions required for access that determines who can use the keys and under what type of condition. The AWS services, which are integrated with KMS, include Aurora, EMR, Glacier, S3, Lex, Connect, DynamoDB, Lambda, Glue and many more.

Capabilities of Audit

If your AWS account is linked with CloudTrail, each of the requests that you make to KMS is properly recorded within a log file that is forwarded to the S3 bucket that you have already specified while enabling CloudTrail. The recorded information includes all the details of the user, date, time, API action and when any relevant key has been used.

Durability, scalability, and availability

KMS is a managed service. The encryption grows with an increase in usage. KMS has the capability to scale with needs automatically. KMS allows the users to manage several master keys in their accounts and use them whenever the user wants. KMS comes with a default limit of keys along with request rates that can be increased if needed. The master keys that the user creates in KMS or the keys, which have been created by the other AWS services on behalf of the user, cannot be readily exported from the KMS service. Thus, KMS handles the durability of the keys. For ensuring that the data along with the keys are highly available, multiple copies of the encrypted key versions are stored in the system by KMS that has been designed for durability of 99.99999%.

If the user imports the keys into KMS, the user can easily maintain a secure copy of all the master keys so that they can be re-imported when they are not available when the user needs to use them. When you use KMS for creating the master keys using the custom key feature, the encrypted key copies are backed up automatically and the user can have complete control over the process of recovery.

Secure

KMS has been designed in such a way so that no one along with the employees of AWS can fetch the plaintext keys from KMS. The service of KMS uses HSMs or hardware security modules for protecting the integrity along with confidentiality of the keys regardless of the fact that whether you request KMS for the creation of your keys, you create them in the cluster of CloudHSM or import the keys within the service. The plaintext keys are not written on the disk. The keys which are created by KMS itself are not transmitted outside the AWS region. The KMS HSM firmware updates are controlled by the multi-party control of access that is audited and reviewed by an Amazon independent group.

Chapter 13: Other AWS Services

AWS provides users with all types of cloud storage along with database solutions. It is well known for its secure, durable, volatile and highly available services. Apart from the services, which we have already discussed in the previous chapters, AWS provides several other services that can help you with all your needs. Let us have a look at some of them.

Amazon Elastic Container Registry

Amazon ECR or Elastic Container Registry is a Docker container registry system that is fully managed that makes it easier for all the developers to manage, store and deploy the various images of Docket container. The service of ECR is integrated with ECS or Elastic Container Service that helps in simplifying the development of production workflow. ECR discards the need of operating the container repositories by the user himself along with scaling of the underlying base. ECR hosts the images in a highly scalable and available structure that allows the users to deploy the containers for their applications. The integration of ECR with IAM or Identity and Access Management grants control of each of the repository at the resource-level. The users need to pay only for the amount of data they store in the repositories along with the amount of data that is transferred to the internet.

Amazon Elastic Container Service

Amazon ECS or Elastic Container Service is a high performing and highly scalable service for container management also supports Docker containers. It allows the users to run the application easily on a cluster of EC2 instances, which is fully managed as well. The service of ECS discards the requirement for the user to operate, install and scale the infrastructure of cluster management. You can stop and launch the applications which are Docker-enabled with the help of API calls and also access the familiar features such as security groups, EBS volumes, Elastic Load Balancing, IAM roles and query the complete cluster state.

ECS can be used for scheduling the container placements across the clusters, which are based on the needs of the resources and requirements of availability. If you want, you can also integrate your won third-party types of schedulers for meeting the requirements of your application and business.

AWS Elastic Beanstalk

AWS Elastic Beanstalk is a very easy-to-use service which can be used for scaling and deploying the web applications along with the services which are developed using .NET, JAVA, Node.js, PHP, Ruby, Docker, Python and Go on regular servers such as Nginx, Apache, IIS, and Passenger. All you need to do is to just upload the code and Elastic Beanstalk will automatically handle the job of deployment, starting from provisioning of capacity, auto-scaling, load balancing to the monitoring of application health. You can also retain your full control over the resources of AWS and can also access the base resources any time you want.

Amazon Elastic Block Store

Amazon EBS or Elastic Block Store provides the users with storage volumes of block-level for using the instances of EC2. The EBS volumes are attached to the network and prevail freely from the instances. EBS provides storage volumes that are reliable, available and predictable. EBS is suited for those applications that require a file system, database or access to storage of raw block level.

AWS Snowball

AWS Snowball is a data transport solution that uses various security devices for transferring huge amounts of data in and out of the services of AWS. The process of data transfer using Snowball is very simple, secure, and fast and is low cost as well.

Amazon Aurora

Amazon Aurora is a relational database that is Postgre SQL and MySQL compatible. It is built for the cloud that functions by combining the availability and performance of the high-end databases along with the cost-effectiveness and simplicity of the databases, which are open source. Aurora, when compared to the standard MySQL databases, is five times faster than MySQL and is three times faster than the PostgreSQL databases. Aurora provides the customers with the availability, security along with the reliability of the databases, which are of the commercial-grade at half the cost. RDS or Relational Database Services, which also automates the various tasks of administration such as setting up a database, provisioning of hardware, backups along with patching, manage Aurora fully.

Chapter 14: AWS Billing and support services

AWS is full of features and some awesome tools. Some of the tools even come as complimentary with the basic services as well. It has turned out to be a necessary service that all of the organizations need for speeding up their functioning and for getting greater insights into their business.

All the users of AWS have complete access to the billing and account support services from Amazon at no extra cost. Only the personalized support of technical issues requires a support plan. You can also visit the AWS support website for more information. The fastest way of contacting with AWS support is by opening up a support case for the inquiry of your billing. It is a direct method. AWS support does not require any direct contact number for reaching out to the support representative. For opening support case, you need to sign in into your own AWS account or else you need to have permissions of IAM for an opening support case.

Contacting the AWS support

- Sign in to your AWS account and navigate to the Support Center. In case any dialog box is prompted, enter your email address and password.
- Select the option “Open new case.”
- On the page of an open new case, you need to select Account and Billing Support and then fill in the required forms.

After you have completed filling the form, you can either choose the Web for getting an email response or Phone for requesting a support call from the support representative of AWS. The facility of instant chat is not available for the cases of billing inquiries.

AWS Business Support

You can get business support from AWS in case you are having heavy workloads on AWS. You can also enjoy the facility of 24*7 support along with architectural guidance. The AWS trusted advisor helps in checking all the available resources and provisions the resources for reducing the cost and improving performance, security and fault tolerance. The AWS health dashboard helps in analyzing the resources and alerts the user when any of the resources are affected. You can also get 24*7 access to the engineers with the help of chat, phone, and email.

Conclusion

After you are done reading the whole Book, you will be able to develop a clear image of the concept of AWS services. By now, you should be geared up and all set for incorporating AWS services to your business. If you keep on thinking, you can never start and this is a perfect time, to begin with, the services from Amazon.

With AWS services, you can have complete control over all the choices and decisions regarding your data centers and cloud. The main concept behind this is to make you aware of the benefits of AWS that can allow your business and applications to touch new heights. The key benefit of being a user of AWS services is that you no need to think again about the various sectors of cloud computing as most of the tasks are done by AWS itself. Therefore, you can enjoy automated services without any manual intervention.

Your web applications are all about the requirements of your clients. Therefore, it is your duty to keep up with all the requests and deliver what they need and it can be easily done with the help of AWS services. If you are going to create something, which already exists over the internet, why will the clients choose you? The only goal is to focus on your data and leave everything else on AWS.

If you find this book helpful for your business in any way, kindly leave a review on Amazon.