**Essentials of Cloud Computing**

**Challenges with corporate datacenter environment:**

Capital Expenditure

Time to set up the infrastructure

Maintenance overhead

Resource under/over utilization

Implications of the on-premise solutions -

Under-utilisation of on-premise servers reduces ROI.

On-premise servers may be overloaded with the traffic and may go down.

Limited scalability in the on-premise environment, leads to loss of business continuity.

**Benefits of Cloud Computing**

To focus on business activities**.**

Lower capital Expenditure & TCO

User access management

Pay as you go pricing model

Secure storage management

Easy and agile deployment

Reliability, scalability and sustainability

Highly automated

proposed solution - deploy the application in the cloud environment so that capital expenditure can be reduced to a great extent, besides the ability to devise a highly available and rapidly scalable solution.

**Deployment Models:**

Cloud services are often delivered over the Internet. Based on workload patterns, security requirements, cloud resources can be deployed in any of 4 models.

explore which models are the most suitable for which workloads.

Public Cloud - Infrastructure owned by CSP and services will be delivered over the Internet.

Private Cloud - operated Solely for enterprise, infrastructure exists in on-premise or off-premie.

Retain control over building and delivering of all cloud based services.(To have better control over the infrastructure and for sensitive workloads, private Cloud is preferred).

As there is a need to keep the data within the premises of the organisation's firewall, private Cloud is suitable.

hybrid cloud - two or more cloud models. Leverage hybrid cloud deployment model and push non sensitive workload to cloud and retain sensitive workload governed by regulatory requirements, in the on-premise environment.

The on premise data center is running out of capacity during peak load and you decide to extend your data center to public cloud.

Community Cloud - good choice for collaboration of multiple enterprises, in which infrastructure lies within a designated and protected environment under the control of stakeholders of the community.

**Virtualization**

virtualize the infrastructure and improve the resource utilisation.

Virtualization reduces the infrastructure footprint.

Leverage bare metal or Type 1 hypervisor as it delivers at low latency.

**Containers**

Application to be containerized and orchestrated to manage the scalability and availability. Cost efficient and highly available solution.

Deploy the apps in a containerized environment and use these custom images to replicate it in the production environment. Containerized applications are highly portable.

**DevOps**

Leverage DevOps methodologies and automate the software release process through CI/CD pipeline.

Deploy the application in cloud infrastructure, setup CI/CD pipeline for quick release of features. Helps in releasing features quickly with the help of automation.

Played a part in microservices increased adoption -

Industry has shifted the focus to Agile culture and DevOps principles.

Allows you to choose a technology stack for developing the service.

**Cloud Service Models:**

Software as a Service - SaaS, readily available and adopts a pay as you go model.

Microsoft office, Google Apps or Gsuit, Microsoft Teams,slack channel

AWS - workmail,chime,connect

Platform as a Service - PaaS, provides a runtime and development environment for hosting your applications, without infrastructure management overhead.

Runtime, platform and tools by CSP. Control over application code. - .NET, Java, Apache, Python.

ex-

Google App Engine GAE from GCP

App cloud from Salesforce.com

oracle cloud platform from oracle

OpenShift from Red Hat Inc.

AWS - Elastic Beanstalk,

Infrastructure as a service - IaaS

Compute, Storage and network resources

Control over the OS and cloud storage device

Configure the development environment remotely.

Ex:- EC2 for virtual servers and S3 for scalable object storage from AWS, VM from Microsoft Azure and Google Compute Engine from Google Cloud Platform.

**Compute and Networking Services**

**AWS Elastic Compute Cloud - EC2 VM**

Provides scalable computing capacity.

Launch as many or as few virtual servers as you need.

Configure security and network.

Manage storage.

It is a IaaS

It can be accessed via the CLI or by using a programming language of choice using AWS SDK's.

Key features - Highly elastic,99.99%,Pay as you go model.

Amazon EC2 features

AMI - Amazon Machine Images - EC2, a region specific template that helps users to choose the operating system and required software packages to create EC2 Instance. User created AMIs can be shared with other users.

EC2 Instance types -

General Purpose Instance - general types of workloads such as POC,dev and test. ex:- t2.micro,m5.large among others

Compute Optimised - applications that require high performance processors.

ex:- Batch processing. c6g.large,c5.large ..

Memory Optimised - to deliver fast performance such as real time big data analytics, in memory db etc., r5.large, x1.16xlarge etc.,

Accelerated Computing Instance - provide hardware acceleration with the use of GPUs such as machine or deep learning. p3.2xlarge, P2.xlarge

and Storage Optimised - require high sequential read and write access or random IOPS with low latency disk performance such as nosql db such as i3.large,d2.xlarge etc.,

Storage Optimized: These instances are designed for workloads that require high amounts of sequential read and write operations, such as big data processing, data warehousing, and log processing.

Elastic IP addresses - provides static public ipv4 addresses for dynamic cloud computing. can attach to the EC2, Elastic IP can be quickly remap to a new instance if current instances fail.

Tags - it's metadata, you can define EC2 resources.

Key pairs -

AWS EC2 pricing model EC2 Instance/Launch types/purchasing options:

On-demand: pay as you go pricing model, no upfront payment or long-term commitment.

Reserved: compute for long term > 1,2,3 years, upto 70% discount compared to on demand.

3 types -

* Standard reserved instances - provides most discounts and are best suited for steady state usage.
* Convertible reserved instances - provides option to change instance families and other instance attributes during the reservation period.
* Scheduled reserved instances - when you want to launch instances within a specific time frame., weekly, monthly for the entire reservation period.

Spot Instances: AWS allow unused EC2 Instances in an availability zone, to be purchased at a price much lower than on demand pricing, upto 90% discount as compared to on demand pricing. Workloads which are stateless, fault tolerant, test and development workloads which are resistant to interruptions.

EC2 Instance Tenancy -

Tenancy defines how EC2 instances are distributed across physical hardware and affects pricing. There are three tenancy options available:

1.Shared ( default ) — Multiple AWS accounts may share the same physical hardware. But your instances run in full isolation with other instances. This is the most economical and cost-effective option.

2.Dedicated Instance ( dedicated ) — Your instance runs on single-tenant hardware.

3.dedicated host - Dedicated Host gives you additional visibility and control over how instances are placed on a physical server, and you can consistently deploy your instances to the same physical server over time.

It is suited for scenarios when software or applications running have a complex licensing model such as BYOL or companies with strict regularly and compliance needs.

EC2 Placement Groups - EC2 Instance is launched in AZ, the EC2 service itself places the instance in a way such that it is spread across the underlying hardware to reduce the risk of correlated failures. AWS allows users to create placement groups to modify the placement group of independent instances based on workload requirements. 3 types

1.Cluster - Cluster Compute Instances provide a large amount of CPU.

Cluster placement groups is a logical grouping of instances within a single Availability Zone. It is recommended for applications for minimum network latency, maximum network throughput.

2.Partition - Instances in a spread placement group generally share a common availability zone, but can span availability zones.

Partition placement groups help reduce the likelihood of correlated hardware failures for your application. When using partition placement groups, Amazon EC2 divides each group into logical segments called partitions. Amazon EC2 ensures that each partition within a placement group has its own set of racks.

Suited for distributed and replicated workloads such as Hadoop,Kafka,etc.

3.Spread - strictly places a group of instances across distinct underlying hardware to reduce correlated failures. spread placement group is limited to a maximum of seven running instances per availability zone.

Elastic Load Balancer ELB - manage application traffic between multiple instances and scale itself automatically to meet the growing network traffic of the application. (multiple targets,such as EC2, containers,IP addresses and Lambda function.)

Handle application traffic in single or multiple AZs

High availability

Automatic Scaling

Robust Security measures

1.Application Load Balancer - Layer 7 HTTP and HTTPS traffic

2.Network Load Balancer - Layer 4 TCP and UDP traffic

3.Gateway load Balancer - Layer 3 Listens for all IP packets

4.Classic Load Balancer - Previous GEN HTTP HTTPS and UDP traffic.

Auto-Scaling - allows to maintain **high availability** of your application by automatically adding or removing EC2 Instances according to the needs of the application or as defined by the user., scale in/out min, max, desired capacity.

With EC2 autoscaling you can significantly **improve the fault tolerance** of your application as unhealthy instances are replaced by a new one with autoscaling.

and improve **cost management**.

All instances will be terminated upon ASG deletion,

Manual scaling - change sizeof an ASG, adjust min,max, desired category.

Dynamic scaling - policy can be set automatically scale instance in ASG if avg. CPU utilisation go beyond 60%.

Scheduled scaling - settings a scale out policy to scale out on the specified schedule.

Avg.CPU Utilisation of instance seem to be approaching 100% ?

Create an AMI of the instance, define an ASG using AMI, deloy a load balancer under the ASG.

Serverless Services:

No infrastructure management (capacity provisioning, patching operating system).

Reduced TCO.

Highly available and fault tolerance.

Amazon Elastic Beanstalk - AWS Elastic Beanstalk is an easy-to-use service for deploying and scaling web applications and services developed with Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker on familiar servers such as Apache, Nginx, Passenger, and IIS.

You can simply upload your code and Elastic Beanstalk automatically handles the deployment, from capacity provisioning, load balancing, auto-scaling to application health monitoring. At the same time, you retain full control over the AWS resources powering your application and can access the underlying resources at any time.

Amazon Lambda - Run code without thinking about servers or clusters.

Run code without provisioning or managing infrastructure. Simply write and upload code as a .zip file or container image.

You need a serverless service to execute a piece of code in order to send notification to users as and when EC2 Instance is terminated.

AWS Glue: is a serverless data integration service that makes it easy to discover, prepare, and combine data for analytics, machine learning, and application development. AWS Glue provides all the capabilities needed for data integration so that you can start analyzing your data and putting it to use in minutes instead of months.

AWS managed services -all SaaS end to end managed by AWS, update

S3,Lambda

Customer managed services - IaaS and PaaS managed by customers.

EC2, Elastic containers,elastiche,RDS.

**Other compute services**

AWS Batch

AWS lightsail

AWS wavelength

AWS Outpost

**AWS networking and content delivery services:**

[**https://aws.amazon.com/products/networking**](https://aws.amazon.com/products/networking)**/**

Network foundation -Amazon VPC, AWS Transit gateway, AWS privateLink.

Application Networking - Amazon VPC Lattice, AWS appMesh, AWS API Gateway, AWS Cloud Map.

Edge networking - CloudFront, Route 53, AWS Global Accelerator.

Hybrid connectivity - AWS Direct Connect, AWS site-to-site VPN, AWS client VPN, AWS cloud WAN.

Network Security - AWS shield, AWS WAF, AWS network firewall, AWS Firewall Manager.

VPC and components

AWS VPC (Virtual Private Cloud)- allows you to create a customized own network in the AWS cloud to improve the security for deployed applications and resources.

Allows you to launch AWS resources in logically isolated networks which you can define and control. This virtual network environment gives you control and allows you to choose your own IP address Range, Subnets, Route Table, Network Gateway and more.

IPv4 CIDR (Classless Inter-Domain Routing) Block -is a method used to represent a range of IP addresses using a compact notation. In IPv4, an IP address is represented as a 32-bit number divided into four octets, separated by dots. For example, the IP address 192.168.1.1 is represented in binary as 11000000.10101000.00000001.00000001.

A CIDR block is represented by an IP address and a prefix length, separated by a slash (/) character. For example, the CIDR block 192.168.1.0/24 represents all the IP addresses from 192.168.1.0 to 192.168.1.255.

The prefix length specifies how many bits of the IP address are used to identify the network portion of the address. In the example above, the prefix length is 24, which means that the first 24 bits of the IP address represent the network portion, and the last 8 bits represent the host portion.

CIDR blocks are commonly used in routing tables and firewalls to specify which IP addresses are allowed or denied access to a network or service. They allow for more efficient use of IP addresses and make it easier to manage large networks with many devices.

172.31.0.0/24 for Subnet 256 ips, first five reserved for monitoring, rest of 251 ips available.

IPv4 is 4 parts, one is called optect,

each octets consist 8 bits, zeros and ones in binary format.

/24 - 32-24(n)=8 2^8=256-5=251

/22 - 2^10

/16 - 2^16

IP CIDR Range - IPv4(public and private ips) or IPv6(only public ips).

IPv4 CIDR Block - 172.31.0.0/16 for VPC 65,536 ips

Since 256 is a power of 2, we can use a /24 subnet mask, which provides 256 IP addresses in total. In a /24 subnet, the first 3 octets (24 bits) represent the network portion, and the last octet (8 bits) represents the host portion

\*The block size of AWS VPCs can be in between a /28 network and /16 network.

IPv6 CIDR Notation Explanation.

Ipv6 cide block

An IPv6 address is a 128-bit address that is used to identify devices on a network. The address is typically represented as eight groups of four hexadecimal digits separated by colons (:), for example, 2001:0db8:85a3:0000:0000:8a2e:0370:7334.

Each group represents 16 bits of the address, and the entire address is divided into eight 16-bit blocks. These blocks are often referred to as "hextets."

IPv6 addresses can also be represented using the CIDR notation, which is used to indicate the network prefix. The network prefix is the number of bits that are used to identify the network portion of the address.

For example, the address 2001:0db8:85a3:0000:0000:8a2e:0370:7334/64 indicates that the first 64 bits of the address (the first four hextets) represent the network portion of the address, and the remaining 64 bits represent the device portion of the address.

In this case, the network prefix is /64, which means that the first 64 bits of the address are used to identify the network, and the remaining 64 bits are used to identify the device.

Route table - when a VPC is created a route table is called a main route table gets created along with it. All the Subnets in VPC are implicitly associated with this name route table. The route table comprises a set of rules, called this route that determine the flow of network traffic from your subnet. It is controlled by implicit VPC rather. Custom route tables can also be created and can be associated with the subnet of choice. One subnet can be associated with only a single route table. A route table comprises a set of rules called routes. Everywhere in the table has a destination and target. The destination is CIDR range of IP address

which specifies you want traffic to go. Target specifies whether a gateway or network interface or connection through which sends the destination traffic.

Ex:- IG, for communication within the VPC

Local route is added to the table by default.

Routes

Destination Targets

172.31.0.0/16 local

0.0.0.0/0 igw-id

Destination is for local route CIDR range

of the VPC, target is local, this route traffic which wants to reach destination within VPC is rather locally within VPC.

Subnets associated with the route table will allow ec2 instances launched within them to communicate with each other because of this local route. To provide public internet access to the VPC subnets by the IG however the destination is 0.0.0.0/0 is added and target is igw-id.

IG - establishes a communication between VPC and internet without any availability risks or bandwidth constraints on network traffic.

High available, Redundant, Scalable

Provides Internet route for VPC

Provide NAT for instances with Public IP.

VPC Subnets -is a subnetwork created inside a VPC. More than one subnet can exist in the same AZ. An AZ can have multiple subnets.

in AZ IP range subset of VPC range, scope of a Subnet restricted to a single AZ, can't span across zones.

public subnets - if the subnet is associated with a route table that has route to the internet by the IG.

If the VPC within IG attached to it has a subnet associated with the route table either a custom route table or main route table which directs traffic to the IG to enable Internet accessibility within the Subnets. AWS resources such as ec2 instances in the public subnet will be able to communicate with the Internet over IPv4. They have public IPv4 or Elastic IP is assigned to them.

Private subnets - opposite public subnet, private subnet not having a route to the Internet via the IG.

Even if the VPC has IG attached to it

If a Subnet in VPC is associated with a route table either custom route table or main route table and there is no route which directs traffic to the IG then the subnet becomes a private subnet. A single vpc in both public and private subnet by associating them with different route tables with appropriate routes.

Destination. Target

172.31.1.0/24. Local

NACLs - optional security layer, controlling inbound and outbound traffic for subnets in the VPC.

applied at subnet level, support Allow and Deny Rules, stateless, Rules evaluated by rule number, implicitly applied to all instances in subnet.

SG - control inbound and outbound traffic using rules. Resource level, EC2.

applied at instance level, support only allows rules,stateful,all rules evaluated to allow traffic, explicitly attached to the insurance.

Main route table and main NACL - default.

Availability Zones - AWS global infrastructure consists of one or more discrete data centres interconnected through low latency links.

Minimum 2 AZs should compute resources be provisioned to achieve high availability.

Egress and Ingress - is a network traffic.

Egress - exits an entry or a network boundary.

Ingress - enters the boundary of a network**.**

Stateless: Stateless Protocol is a network protocol in which Client sends a request to the server and server response back as per the given state. Ex:- HTTP,UDP,DNS

Stateful:Stateful Protocol is a network protocol in which if a client sends a request to the server then it expects some kind of response, in case of no response then it resends the request. Ex:- FTP, Telnet**.**

A NAT gateway is a Network Address Translation (NAT) service. You can use a NAT gateway so that instances in a private subnet can connect to services outside your VPC but external services cannot initiate a connection with those instances.

When you create a NAT gateway, you specify one of the connectivity types: public or private**.**

NAT Gateways are used to allow only Outbound internet Access.

[**https://docs.aws.amazon.com/vpc/latest/userguide/vpc-nat-gateway.html**](https://docs.aws.amazon.com/vpc/latest/userguide/vpc-nat-gateway.html)

Q:SSH, or remote connections to the instances are getting time-out.

a.Route tables of the subnet have a route to the IGW.

b.Inbound rules of the Security Group allow SSH or remote connections.

c.Inbound and Outbound rules of the Network ACL allow SSH or remote connections.

Route 53 - Domain Name System (DNS) web service. Multi-region availability using route 53, achieve LEAST administrative effort.

Routing Policies -

Latency routing policy - will evaluate the latency results and help your user get a DNS response that will minimize their latency.

Route 53 with routing policy(simple)- can be used to provide least administrative effort.

Weighted route 53 - allows you to redirect a part of traffic based on weight.

Failover routing policy -

AWS Global accelerator - Networking service which improves performance of user's traffic.

when creating elb, check it to create an accelerator to get static IP addresses and improve the performance and availability of your application. (Optional) -

2.AWS Transit gateway - A network transit hub that you can use to interconnect VPC and on-premises network.

3.AWS Direct connect - Establish a dedicated network connection from your on-premise to AWS. A consistent network than internet- based connections.

4.VPN - Establish secure connections between your on-premises networks, remote offices, client devices, to AWS global network. Highly-available, managed and elastic cloud VPN solution to protect your network.

5.CloudFront - an easy and cost effective way to distribute the content with low latency and high data transfer speed.

Secure,fast,Programmable Content Delivery Network (CDN). Co-reside at the edge locations/POP's.

Edge location - where data is cached and end users access CDN services faster in AWS.

6.App Mesh - Modern applications are composed of multiple microservices.

Run services by providing consistent visibility and network traffic controls, deliver secure services.

network and Content Delivery:

AWS Private link- establish connectivity between VPCs and AWS services without exposing data to the Internet.

**Storage, Database and Migration Services**

**Storage -**

1.Simple Storage Service - S3 -

object/durable storage option for retaining data backup for immediate retrieval. Low cost option for hosting static websites.

Store in buckets.

Buckets - logical containers for storing data in S3.

99.999999999% of durability,99.99% of availability.

Free teir - 5GB of standard storage class.

Standard - offers high durable,availability and performance object store for frequently accessed data and available immediately. Used for content distribution, big data analytics, dynamic website and gaming applications deliver high throughput and low latency.

Standard IA - data is critical and can't be last. The data is accessed once in a quarter.

One Zone IA - in all other storage classes, your data replicated across multiple AZ. However in One Zone IA, which is present in one AZ only, its charges are 20% less than S3 standard IA class. Lower cost access for frequent data but don't require multiple AZ only one AZ.

Intelligent Tiering - when not sure about frequency of the data access.

S3 Glacier - Archival storage solution, capacity of single archive - 40TB.

AWS snowball and Direct Connect Integration.

Vault lock and access control.

Integrated lifestyle management.

Glacier Deep Archive - retaining data for very long time and data is very very rarely accessed. Public sectors, financial services and health care which need to store data for 7 to 10 years for compliance requirements should use this class. It takes 48 hours for retrieval.

2.EFS - Storage solution for File System, Scalable and Elastic.

Auto scale to petabytes of data.

Regional service.

High throughput and IOPS.

data sharing between multiple EC2 instances.

Standard storage class: Frequent access, pay for file storage/month.

EFS IA: Infrequent access, low cost.

3.FSx - For Lustre and Amazon Simple Storage Service. Can spin up an Amazon FSx for Lustre file system in 5 min, and it's managed by AWS.

4.EBS - Primary storage, persistent block storage, 1 to 16 TB, 99.99% availability and durability, attached only one instance at a time, volumes can be backed-up as snapshots, and support data encryption.

Types of EBS volumes

SSD-Backed volumes

EBS General purpose.

EBS Provisioned IOPS.

HDD-Backed volumes

Throughput optimized HDD.

Cold HDD

EBS Snapshots - Save point-in-time snapshots of your volumes to Amazon S3.

Incremental backups of EBS volumes.

New volumes and instances can be created from snapshots.

Region specific.

Can be shared across AWS accounts.

Automates the creation, retention and deletion of EBS snapshots.

Automation is achieved by creating a life-cycle policy.

5.Instance store Volumes - provides temporary block storage for the EC2 Instances.

6.AWS Storage Gateway - Move backups to the cloud,low latency access for applications to cloud, use on-premises file shares backed by cloud storage, data protection and disaster recovery.

AWS Backup - Automates backup, configure, monitor backup.

Integrated services are EBS volumes, EC2 instances,RDS databases,DynamoDB tables,EFS file systems, storage gateway volumes.

**Databases -**

1.RDS - supported DB Engines are MySQL, Microsoft SQL server, Amazon Aurora - a relational database and Highly available,Oracle,postgresQL, mariaDB.

2.DynamoDB - fast and reliable NOSQL database, global tables in DynamoDB is a fully managed, multi-region and multi-active database which will automatically replicate the tables across your choice of region, read replicas to improve the performance. Document and key value database.

Who is looking to offload the Administrative burdens of operating and scaling distrusted databases to AWS. The db team doesn't want to perform hardware provisioning,setup and configuration, throughput capacity planning, replication, software patching, or cluster scaling.

High performance,Serverless,and key- value NoSQL DB.

3.Neptune - graph database.

4.Redshift - is a scalable data warehouse service, to analyze data for DWH.

5.DocumentDB -

6.ElastiCache - fully managed in-memory data store. Store accumulated game results, ideal for storing the results of typical database queries.

7.Amazon QLDB - Amazon Quantum Ledger Database (QLDB) is a fully managed ledger database that provides a transparent, immutable, and cryptographically verifiable transaction log.

8.Keyspaces - Amazon Keyspaces (for Apache Cassandra) is a scalable, highly available, and managed Apache Cassandra–compatible database service. With Amazon Keyspaces, you can run your Cassandra workloads on AWS using the same Cassandra application code and developer tools that you use today.

**Migration -**

Cloud native features,reduced cost and increased performance.

AWS migration hub - single location to track progress, metrics and tools, improve visibility.

AWS Transfer family - file transfer solution, compatible with SFTPFTP,FTPS.

Server Migration Service - Agentless migration service, migrate large no. of workloads to AWS, reduces downtime.

CloudEndure Migration - Automated lift and shift, expedites Re-hosting, agent based solution.

VMWare on Cloud - vSphere workload solution,faster, simple and cost-effective,modernize on the go.

Snow Family - physical devices and capacity points to run operations in austere setup.

AWS snowball - most cost effective mechanism to move data and meet the migration deadline.

Using a snowball device is the fastest method of getting the data into Amazon Glacier.

As a couple of snowball devices (80TB) is enough to move 150 TB easily.

Snowball Edge - data migration service, want to migrate 100TB of data to AWS cloud.

AWS DataSync - Transfer to S3,EFS,FSx for Windows file server. 10x faster.

AWS Data Migration Service - simplifies the migration of the database to AWS.

Migration Strategies

**6 R's** refer to the 6 application Migration Strategies(Re-hosting, Replatforming, Repurchasing, Refactoring/Re-architecting, Retire, Retain). Reprocess is not part of migration strategy**.**

**Monitoring and Auditing Services**

Cloud trail - You can troubleshoot operational and security incidents over the past 90 days in the CloudTrail console by viewing Event history. You can look up events related to creation, modification, or deletion of resources (such as IAM users or Amazon EC2 instances) in your AWS account on a per-region basis. Events can be viewed and downloaded by using the AWS CloudTrail console. You can customize the view of event history in the console by selecting which columns are displayed and which are hidden. You can programmatically look up events by using the AWS SDKs or AWS Command Line Interface. You can also compare the details of events in Event history side-by-side.

Trusted Advisor - provides recommendations to update architecture in areas like cost optimization, performance security, fault tolerance and service limits. Helps you ensure whether best practices are in place in your architecture.

AWS Config - keep track history of changes made in AWS resources like SG attached to EC2 Instances where your app deployed.

Monitoring and logging

Core features

CloudWatch Metrics -

Data points collected against time

AWS services send metrics to CloudWatch.

CloudWatch Alarms

CloudWatch Events

CloudWatch Logs

Cloud watch - is one stop solution for monitoring, logging and event handling in your account.

Cloud watch logs - real time monitoring and adjustable retention.

VPC Flow logs - to/from VPC traffic.

**AWS Cloud Security and Compliance**

1)IAM - Identity Access Management, enables you to manage access to AWS services and resources securely.

IAM User - the person who uses an IAM account for interacting with AWS services and resources.

IAM Group -

IAM Role -

Service Control Policies - centrally manage all users and role permissions in your team.

AWS Cognito - provides authentication, Authorization and user management for your web and mobile application. Basically if you use the service, directly sign in username and password through third party applications such as FB,AWS, Google. Two main components - user pool, identify pool.

AWS Single-Sign-On(SSO) - makes it easy to centrally manage access to multiple accounts and business applications.

AD Connector - is a directory gateway with which you can redirect derectory requests to your on-premise Microsoft Active Directory.

AWS Directory Service - provides AWS cloud Directory and Microsoft AD (active directory) with other AWS services.

AWS Resource Access Manager - It is a services enable you to easily and securely share AWS resources with any AWS account or within your AWS Organization.

2) Infrastructure protection -

AWS Shield - is a managed (DDoS) protection service that safeguard application is running on AWS.

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

AWS firewall manager - it is securely management service which allows you to certainly configure and manage firewall rules across your accounts and applications in AWS Organization.

AWS WAF (web application firewall) - protect your web application or API against common web exploits that may affect availability, compromise, security or consume excessive resources.

A good practice is to use a Web Application Firewall (WAF) against attacks, such as SQL injection or cross-site request forgery, that attempt to exploit a vulnerability in your application itself. Additionally, due to the unique nature of these attacks, you should be able to easily create customized mitigations against illegitimate requests which could have characteristics like disguising as good traffic or coming from bad IPs, unexpected geographies, etc. At times it might also be helpful in mitigating attacks as they happen to get experienced support to study traffic patterns and create customized protections.

3) Detective controls

AWS Security Hub - can automatically aggregate security findings data from supported AWS Partner Network (APN) security solutions, so you can have a comprehensive view of security and compliance across your AWS environment. It consolidates, organizes and prioritizes security warnings and results from several AWS services.

Amazon GuardDuty - identifies threats by continuously monitoring the network activity, data access patterns, and account behavior within the Amazon Web Services environment.

Amazon GuardDuty is a continuous security monitoring service that analyzes and processes data sources, such as AWS CloudTrail data events for Amazon S3 logs, CloudTrail management event logs, DNS logs, Amazon EBS volume data, Amazon EKS audit logs, and Amazon VPC flow logs. It uses threat intelligence feeds, such as lists of malicious IP addresses and domains, and machine learning to identify unexpected, potentially unauthorized, and malicious activity within your AWS environment. This can include issues like escalation of privileges, use of exposed credentials, or communication with malicious IP addresses, domains, or presence of malware on your Amazon EC2 instances and container workloads. For example, GuardDuty can detect compromised EC2 instances and container workloads serving malware, or mining bitcoin. It also monitors AWS account access behavior for signs of compromise, such as unauthorized infrastructure deployments, like instances deployed in a Region that has never been used, or unusual API calls like a password policy change to reduce password strength.

GuardDuty informs you of the status of your AWS environment by producing security findings that you can view in the GuardDuty console or through Amazon CloudWatch events.

AWS Inspector - an automated vulnerability management service that continually scans AWS workloads for software vulnerabilities and unintended network exposure.

4) Data protection

AWS macie - is fully managed data security and data privacy service that uses machine learning and patterns matching to discover and protect your sensitive data in AWS.

AWS Key Management Service (AWS KMS) makes it easy for you to create and manage cryptographic keys and control their use across a wide range of AWS services and in your applications**.**

Setup an IAM role and access policy for each customer application. Encrypt objects using server side encryption with a key stored in AWS KMS (SSE-KMS) and configure the DMS to use a separate AWS KMS key for each customer.

AWS cloudHSM - is a cloud base Hardware Security Module that enables you to easily generate and use your own encryption key on the AWS cloud.

AWS certificate manager - is a service that let you easily provision, manage and deploy public and private (SSL/TLS) certificates for use with AWS services and your internal connected resources.

**Compliance** -

gives customers a comprehensive view of their compliance status.

AWS Artifact - No cost, self-service portal for on-demand access to AWS compliance reports.

Protect against DDoS attacks?

There are several AWS services that can help mitigate DDoS attacks, but two of the most commonly used ones are:

**AWS Shield:** AWS Shield is a managed DDoS protection service that is available to all AWS customers at no additional cost. It provides protection against common network and transport layer DDoS attacks by automatically detecting and mitigating them. AWS Shield also offers a number of additional features for customers who require more advanced DDoS protection, including 24/7 monitoring and support, advanced visibility and reporting, and the ability to customize protection policies.

**Amazon CloudFront:** Amazon CloudFront is a content delivery network (CDN) that can help mitigate DDoS attacks by distributing content across a global network of edge locations. This can help reduce the impact of DDoS attacks by spreading the traffic across multiple locations, rather than allowing it to overwhelm a single server or data center. Additionally, CloudFront provides a number of security features that can help protect against DDoS attacks, including SSL/TLS encryption, IP whitelisting and blacklisting, and access control lists.

CloudFront - Using a content delivery network (CDN) sure as Amazon CloudFront to cache and serve static text and images or downloadable objects such as media files and documents is a common strategy to improve webpage load times, reduce network bandwidth costs, lessen the load on the web servers, and mitigate DDOS attacks**.**

AWS Shield standard protects you from 96% of today's most common DDOS attacks through mitigation.

Route53 - reduces DDoS Risks through it's features of scalling and fault tolerance.

Data protection at rest and transit?

AWS Certificate Manager - helps in providing protection to data at rest and in transit.

VPN connectivity - helps in providing protection to data at rest and transit.

AWS KMS - helps in providing protection to data at rest and transit by encrypting data.

Amazon S3 generates and manages the encryption keys in default (SSE-S3) server side encryption.

Patching servers is done by the customer itself as needed and under AWS security best practices.

EC2 key pairs credentials required to SSH directly into EC2 instances.

Root users will have unrestricted access to your AWS resources.

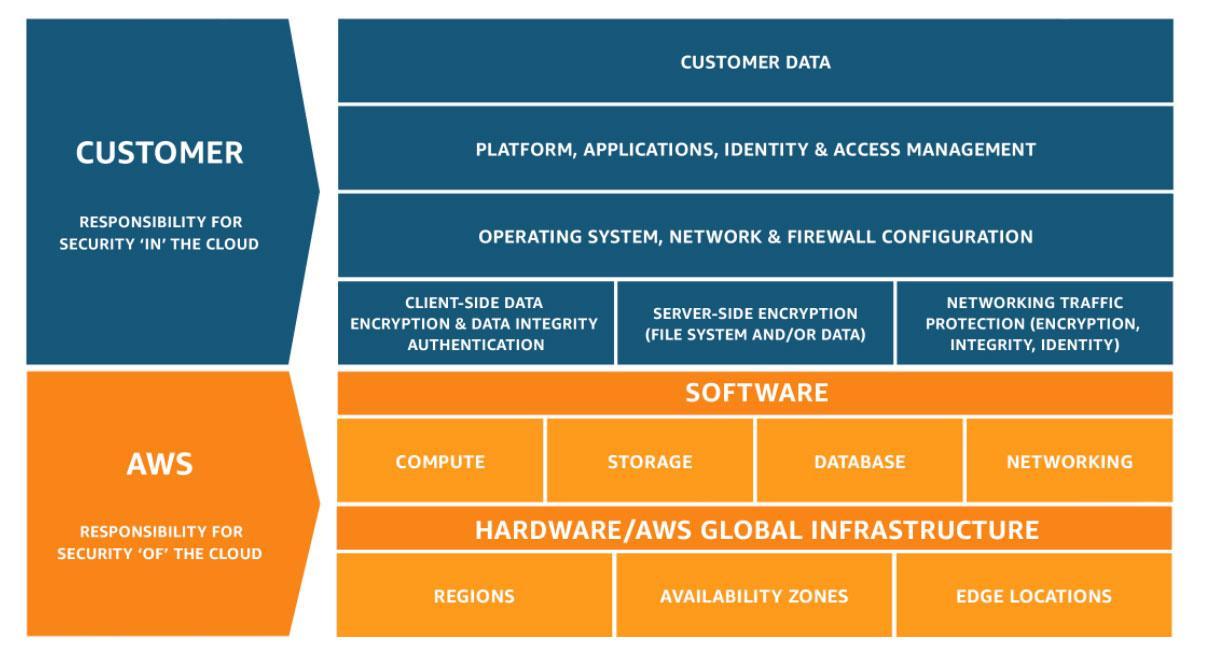
AWS Security Model, AWS performed -

Decommissioning storage device according to NIST 800-88.

**Shared Responsibility Model:Security and Compliance is a shared responsibility between AWS and the customer -**

AWS responsibility “Security of the Cloud”

Customer responsibility “Security in the Cloud”



Patch Management – AWS is responsible for patching and fixing flaws within the infrastructure, but customers are responsible for patching their guest OS and applications.

Configuration Management – AWS maintains the configuration of its infrastructure devices, but a customer is responsible for configuring their own guest operating systems, databases, and applications.

Awareness & Training - AWS trains AWS employees, but a customer must train their own employees.

**AWS Pricing and billing**

Cloud Cost optimization: is a method of reducing total cloud cost by optimising the resources and their cost for example, managing unused resources,opting for reserved capacity for greater discounts.

AWS TCO calculator - Estimating the AWS cloud cost over on-premise setup.

AWS Budget - Improve planning and cost control with flexible budgeting and forecasting. With AWS Budgets, set custom budgets to track your costs and usage, and respond quickly to alerts received from email or SNS notifications if you exceed your threshold.

Use cases

Monitor costs and usage

Set your preferred budget period to daily, monthly, quarterly, or annually, and create specific budget limits.

Create scheduled reports

Stay informed on how actual or forecasted costs and usage progress toward your budget threshold.

Respond to thresholds

Set up custom actions to run automatically or through an approval process when a budget target is exceeded

AWS Cost Explorer - Visualize, understand, and manage your AWS costs and usage over time. Get started quickly by creating custom reports that analyze cost and usage data. Analyze your data at a high level (for example, total costs and usage across all accounts) or dive deeper into your cost and usage data to identify trends, pinpoint cost drivers, and detect anomalies.

**AWS Cost Explorer Resource Optimization** to get a report of EC2 instances that are either idle or have low utilization. You can reduce costs by either stopping or downsizing these instances.

AWS Compute Optimizer - is a service that analyzes your EC2 instance utilization metrics, such as CPU and memory utilization, to recommend optimal instance types and sizes. It can also identify under-utilized instances that may be candidates for downsizing or termination to reduce your AWS bill.

AWS Cost and Usage Report - The AWS Cost and Usage Reports (AWS CUR) contains the most comprehensive set of cost and usage data available. You can use Cost and Usage Reports to publish your AWS billing reports to an Amazon Simple Storage Service (Amazon S3) bucket that you own. You can receive reports that break down your costs by the hour, day, or month, by product or product resource, or by tags that you define yourself. AWS updates the report in your bucket once a day in comma-separated value (CSV) format. You can view the reports using spreadsheet software such as Microsoft Excel or Apache OpenOffice Calc, or access them from an application using the Amazon S3 API.

Pricing/Simple Monthly Calculator - Estimate your monthly bill.

Cloud watch billing alarm - to optimize cloud cost.

Billing alerts - triggers when your actual account billing exceeds the threshold you specify.

QuickSight - creating a new dashboard for any billing problem detected from CUR.

**Features of AWS Cloud**

**Global reach:** will support international company's requirement for low latency to all of its customers**.**

Consolidated billing feature from **AWS Organization**, customer is using multiple AWS accounts with separate billing, take advantage of volume discounts with minimal impact to the aws resources.

Amazon allows you to convert your bill to prefered currency to pay your invoices.

1.your organisation is looking for a solution to estimate the cost savings and compare the the cost of your application in an onpremise or traditional hosting environment to AWS. Which of the following service is appropriate to use?

AWS cost explorer

AWS computer optimizer

AWS Budget

AWS TCO Calculator - yes

2.which analysis calculates the total cost of ownership for running an ene-to-end traditional IT environment versus deploying to AWS?

Depreciation

Cost savings - yes

Deployment

Customer satisfaction

3.which of the following represents the best value of a cost savings analysis for a customer?

To show lower price is the matters

To convince the customer that the cloud is more secure than traditional IT.

providing a complete view of the cost to deliver a equal improved performance with AWS when compared to an on-premise or colocation infrastructure. - yes

4.you have been hired to enhance overall cloud storage cost myntl.in. considering which of the the below three points are most appropriate and will have impact on the storage?

Clean incomplete uploads. - yes

store only the critical files in the cloud.

Storing data in a compressed format. -yes

Changing the storage class of the data over time. - yes

Keep the frequently accessed data on-premise and use cloud storage just to back up archive data

5.Your company hosts a social media site supporting users in multiple countries. As a cloud solution architect you have been asked to provide a low latency design to the application that leverage multiple regions for the most recently accessed content and latency sensitive portions of the website. Which of the below capability would you recommend to meet your organisation requirement?

Elasticity

Scalability

Caching -yes

Availability

6.you are tasked with moving a legacy application from a virtual machine running inside your datacenter to the cloud. As a cloud architect which approach would you recommend to optimize the cloud spending?

Move only few servers to cloud

Multi cloud approach

Single cloud approach - yes

Leveraging all possible services from CSP to reduce maintenance.

7.A social blogging company has hosted their application on AWS EC2 instances and wants to set a budget that alerts when it exceeds their budgeted cost or usage amount. Which of the following service is appropriate to use?

AWS cost explorer

AWS compute optimizer

AWS budget - yes

AWS TCO Calculator

8.Your organisation is seeking recommendation from you for optimising system performance, cost availability. As a solution architect you have asked your organisation to leverage Trusted Advisor from AWS. On which of the below categories, Trusted Advisor provides insights on?

Secured fault tolerance,high availability and connectivity.

Performance, cost optimization, security, fault tolerance and service limits. - yes

Security, access control,high availability and performance.

Performance, cost optimization,access control, connectivity,and service limits.

9.Identify the best practices of cost optimization?

Instance right sizing - yes

Instance elasticity -yes

Choosing the right pricing model -yes

Optimising the storage - yes

Increasing security standards -yes

10.which of the following is not one of the four key areas of cloud financial management?

Establishing cost transparency to ensure visibility into and accountability for spend.

Cost optimization.

Planning and forecasting.

Licence pricing negotiation. -yes

Cloud financial operations.

11.As a solution architect you are asked to train your team members in optimising cloud cost. Which of the below principles would you share to optimise cloud spend? select 3.

Use a small size instance.

Taking care of resource utilisation. -yes

Distribute servers across the globe.

Use caching. -yes

Clean up incomplete uploads. yes

12.Tending market is hosting a shopping website application on AWS EC2 instances. They want to discover the under-utilised instances, so as to analyse the impact on your AWS bill and consider the savings plans. Which of the following is appropriate to use?

AWS compute optimizer. -yes

AWS cost explorer.

AWS budgets

AWS cost and usage report.

13.You want to access several products to help you understand, analyse and control your AWS costs. You also want detailed information on usage trends, and take action on their insights. Which of the following services serves as a solution for the above?

AWS cost explorer.

AWS budgets.

AWS cost and usage report.

All of the above. - yes

14.A famous web design company has decided to host their servers in the cloud. Incorporating which of the below principles will help the organisation to save their cost and to maintain availability?

Only run minimal resources to optimise cost.

Always Procure resources only on demand basis so that you can shut down when not required.

Reserve instances when it is required for a longer duration across multiple data centres. -yes

Distribute servers across the globe to maintain availability.

15.Your organisation is looking for solution to explore your AWS costs and usage as per AWS service, Region etc., and default reports to analyse costs. Which of the following is an appropriate solution?

AWS cost explorer -yes

Billing dashboard

AWS budgets

AWS TCO Calculator

16.Your organisation has launched a periodic image analysis application in AWS cloud. To optimise the cloud spending, you are asked to set up a billion alarm which will notify you whenever your account bill goes beyond 50000. Which among the below services is most appropriate to use?

AWS cloud watch - yes

AWS cost explorer

AWS trusted advisor

AWS cloud trail.

17.Which of the following are pillars of the cloud value framework ? select 4.

Business Agility - yes

Cost savings - yes

Effective Security -yes

Operational resilience -yes

Smart purchasing

Staff productivity

18.GameHere is hosting a gaming website on AWS EC2 instances and want to understand the costs and usage associated with Reserved Insurances. Which of the following options can serve as a solution?

RI reports in cost explorer.

Reservation related data in cost and usage report.

AWS cost explorer's RI purchase recommendation.

All the above. - yes

19. Library Anywhere is hosting a website application on AWS EC2 instances and want to improve performance and understand the risks to evaluate the price performance trade-off for their workloads. They are looking for recommendations for downsizing and upsizing instance type. Which of the following service is appropriate to use?

AWS cost explorer - yes

AWS budgets

AWS compute optimizer

AWS TCO Calculator

**Cloud9 IDE for Developers.**

**Cloud9 is a cloud-based integrated development environment (IDE) that allows developers to write, run, and debug their code from anywhere with an internet connection. It was originally founded in 2010 and acquired by Amazon Web Services (AWS) in 2016.**

**Cloud9 provides a complete development environment with features such as code editing, debugging, version control, and collaboration tools. It supports several programming languages including Python, JavaScript, PHP, Ruby, and more. Users can choose from a range of pre-configured environments or create their own custom environments.**

**One of the key benefits of using Cloud9 is that it allows developers to work collaboratively on the same codebase, regardless of their location. Users can share their IDE with others and work together in real-time on the same codebase, making it ideal for remote teams. Additionally, Cloud9 offers seamless integration with other AWS services, making it easy to deploy and manage applications in the cloud.**

**Key words**

Reliability,High Availability, fault tolerance

Upfront cost / No upfront cost.

Scale in and out.

migration smooth and hassle free.

single point of failure, so host on HA zones

Loosely coupled / decoupled arch or mechanism

LAMP stack - is one of the most common software stacks for many of the web's most popular applications. However, LAMP now refers to a generic software stack model and its components are largely interchangeable.

Monitoring tools by CSP can in both cloud and on premise and create a dashboard?

Design principles?

Design of failure. . . .

Granular billing features?

Why and when do we need to use ALB?

Difference between nlb vs alb

How did you configure alb for your application?

What methods have you used to configure ALB?

How do you manage certificates?

1.which of the following are benifits of the AWS cloud? Select Two

Companies needs increased IP staffs.

Capital expenses are replaced with variable expenses. - yes

Customer receives the same monthly bill regardless of which resource they use.

Companies gain increased agility. -yes

2.which AWS cloud architecture design principal support the distribution of workloads across multiple Availability Zone?

Implement automation.

Design for agility.

Design for failure. - yes

Implement electricity.

3.which of the following is a responsibility of AWS under the shares responsibility model?

Design a customer's application for disaster recovery.

Update the guest OS on deployed Amazon EC2 instances.

Configure new resources within a AWS account.

Secure the physical infrastructure. -yes

4.which of the following describes a security best practice that can be implemented by using AWS Identify and Access Management (IAM)?

Turn off AWS management console access for all users.

Generate secret keys for every IAM user.

Grant permission to use who are required to a specific task only. - yes

Store the AWS credentials within Amazon EC2 instances.

5.A company needs to monitor and receive alerts about AWS management console sign-in events that involve the AWS account root user. Which AWS service can the company use to meet these requirements?

Amazon CloudWatch. -yes

AWS Config.

AWS trusted advisor.

AWS Identify and Access Management (IAM).

6.A company has an application server that runs on a Amazon EC2 instance. The application server needs to access contents within a private Amazon S3 bucket. What is the recommended approach to meet this requirement?

Create an IAM role with appropriate permissions. Associate the role with the EC2 instance. - yes

Configure a vpc peering connection to allow private communication between the EC2 instance and S3 bucket.

Create a shared access key. Configure EC2 instance to use the hard-coded key.

Configure the application to read an access key from a secured source.

7.which security related services of features does AWS offer? select Two.

Complete PCI compliance for customer applications that run on AWS.

AWS a Trusted Advisor security checks. -yes

Data encryption. -yes

Automated penetration testing.

8.which recommendations are included in the AWS Trusted Advisor checks? Select Two.

Amazon S3 bucket permissions. -yes

AWS service outage for services.

Multi-factor authentication (MFA) use on the AWS account root user. -yes

Available software patches for Amazon EC2 instances.

Number of users in the account.

9.A company wants a dedicated private connection to the AWS cloud from its on-premises operations. Which AWS service or feature will provide this connection?

AWS VPN.

AWS Private Link.

VPC endpoint.

AWS Direct Connect. -yes

10.Which aspect of AWS infrastructure provides global deployment of compute storage?

Multiple Availability Zones in an AWS Region.

Multiple AWS Region. -yes

Tags.

Resource Groups.

11.which AWS servers or features support data replication across AWS Regions? select Two.

Amazon S3. -yes

Amazon EBS.

Amazon EC2 instance store.

AWS Storage gateway.

Amazon RDS. -yes

12.A company is hosting a static website from a single Amazon S3 bucket. Which AWS service will achieve lower latency and high tranfer speed?

Amazon Elastic Beanstalk.

Amazon DynamoDB Accelerator (DAX).

Amazon Route 53.

Amazon Cloud Front. - yes

13.Which AWS service provides a simple and scalable shared file storage solution for use with Linux-based Amazon EC2 instances and on-premises servers?

Amazon S3.

Amazon Glacier.

Amazon Elastic Block Store (EBS).

Amazon Elastic File System (EFS). -yes

14.which AWS storage service allows content to be accessed via a web browser without using a compute service?

Amazon Elastic File System.

Amazon Elastic Block Store.

Amazon simple Service. -yes

AWS snowcone.

15. Your organisation has multiple application running in one AWS account. You have been asked to apportion AWS costs by application. What can you used to identify cost related to each application?

Use multiple Amazon VPCs.

Tag all resources. - yes

Assign a different IAM Role to each application.

Purchase Reserved Insurances.

16.What type of database capability is provided by Amazon Aurora?

Time-series database.

NoSQL databases.

Relational database. -yes

Graph database.

17.Which of the following pricing models provides lowest cost for storage in Amazon S3?

Savings Plan.

Standard Storage. -yes

Instance Store.

Spot Instances.

18.An application running on an EC2 instance needs to access resources in Amazon S3. How can you assign permissions so that the application can access S3?

Using Identify Federation to login to the instance.

Provide an MFA Token to the application.

Generate a keypair for the instance.

Assign a IAM Role to the instance. -yes

19.which of the following AWS Compute Services is good for workloads that only run for brief periods in response to the infrequent events?

Amazon Elastic Container Services.

AWS Lambda. -yes

Amazon Elastic Compute Cloud.

Amazon CloudWatch.

20.In the AWS shared responsibility Security Model, what is the responsibility of AWS?

Assigning IP address ranges to VPC subnets.

Operating system patch management.

Security Group Configuration.

Edge location. -yes

21.AWS service offer pau-as-you-go pricing, so you pay pay for only what you use each month. Each AWS service ha it's own pricing model. Which among the below of tools can be used to evaluate the expenses of operating a web application on AWS with a conventional hosting environment?

AWS cost explorer.

AWS budgets.

AWS cost and usage report.

AWS total cost of ownership (TCO) calculator. -y

22.A user must audit and review the configuration of all AWS resources on a regular basis, detect non-compliant accounts and be alerted when a resource changes. Which AWS offering may be utilised to fulfil these requirements?

AWS trusted advisor.

AWS config. -y

AWS resource access manager.

AWS system manager.

23.You are using the Lambda function for your application. Which duty is your responsibility while managing AWS Lambda function under the AWS responsibility model?

What problems are solved by AWS CloudFront?

AWS CloudFront provides a globally distributed network of edge locations that cache and deliver content to end-users with low latency and high data transfer speeds.

1. Improved Latency: CloudFront reduces the latency experienced by end-users by caching content at edge locations closer to them. When a user requests content, CloudFront delivers it from the nearest edge location, reducing the round-trip time and improving overall performance.

2. Scalability: CloudFront is designed to handle high traffic and scale automatically. It can distribute content across multiple edge locations, ensuring that the load is balanced and providing consistent performance even during traffic spikes.

3. Global Reach: With CloudFront, you can serve content to users worldwide. It has a large network of edge locations distributed across different regions, allowing you to deliver content with low latency to users in various geographic locations.

4. Security: CloudFront integrates with other AWS services, such as IAM, to provide security features like access control and authentication. It supports SSL/TLS encryption, helping secure content delivery.

5. Reduced Server Load: By caching content at edge locations, CloudFront reduces the load on origin servers. This is particularly beneficial when serving static content

6. Cost Optimization: CloudFront offers cost-effective content delivery by reducing data transfer costs. By caching and delivering content from edge locations, it minimizes the amount of data transferred from the origin server.