AMAZON-AWS SOLUTION ARCHITEC SAA-01

By Rovol Keng

- **Below sheet is created while preparing this dumps.**

ECS : Amazon Elastic Container Service (Amazon **ECS**)

Amazon Elastic Container Service (Amazon ECS) is a highly scalable, high-performance container orchestration service that supports Docker containers and allows you to easily run and scale containerized applications on AWS. Amazon ECS eliminates the need for you to install and operate your own container orchestration software, manage and scale a cluster of virtual machines, or schedule containers on those virtual machines.

ECS cluster: An Amazon ECS cluster is a logical grouping of tasks or services. When you first use Amazon ECS, a default cluster is created for you, but you can create multiple clusters in an account to keep your resources separate.

ECS CLUSTER :

Amazon ECS is a regional service that simplifies running application containers in a highly available manner across multiple Availability Zones within a Region. You can create Amazon ECS clusters within a new or existing VPC. After a cluster is up and running, you can define task definitions and services that specify which Docker container images to run across your clusters. Container images are stored in and pulled from container registries, which may exist within or outside of your AWS infrastructure.

API : application program interface (**API**)

An application program interface (**API**) is a set of routines, protocols, and tools for building software applications. Basically, an **API** specifies how software components should interact. Additionally, **APIs** are **used when** programming graphical user interface (GUI) components.

What is AWS CloudWatch agent?
Amazon **CloudWatch** monitors your Amazon Web Services (**AWS**) resources and the applications you run on **AWS** in real time. You can use **CloudWatch** to collect and track metrics, which are variables you can measure for your resources and applications.

What is SSM in AWS?
AWS Systems Manager Agent (**SSM** Agent) is Amazon software that can be installed and configured on an Amazon **EC2** instance, an on-premises server, or a virtual

machine (VM). **SSM** Agent makes it possible for Systems Manager to update, manage, and configure these resources.

What is detailed monitoring?

The basic **monitoring** level collects **monitoring** data in 5 minute periods. To increase this level and make the **monitoring** data available at 1-minute periods, you must specifically enable it for your instance(s). With **detailed monitoring**, you can also get aggregated data across groups of similar EC2 instances

---

What is the EBS in AWS?

Amazon Elastic Block Store (Amazon **EBS**) provides block level storage volumes for use with **EC2** instances. **EBS** volumes behave like raw, unformatted block devices. ... **EBS** volumes are highly available and reliable storage volumes that can be attached to any running instance that is in the same Availability Zone.

Launching an **EC2 Fleet**. An **EC2 Fleet** contains the configuration information to launch a **fleet**—or group—of **instances**. In a single API call, a **fleet** can launch multiple **instance** types across multiple Availability Zones, using the On-Demand **Instance**, Reserved **Instance**, and Spot **Instance** purchasing options together.

Does AWS have MongoDB?

MongoDB on **AWS**. **MongoDB is** an open source, NoSQL database that provides support for JSON-styled, document-oriented storage systems. ... **AWS** enables you to set up the infrastructure to support **MongoDB** deployment in a flexible, scalable, and cost-effective manner on the **AWS** Cloud.

PostgreSQL is an advanced, enterprise class open source relational database that supports both SQL (relational) and JSON (non-relational) querying.
... **AWS** supports **PostgreSQL** through a fully managed database service with Amazon Relational Database Service (RDS).

Amazon **Redshift Spectrum** is a feature within Amazon Web Services' **Redshift** data warehousing service that lets a data analyst conduct fast, complex analysis on objects stored on the AWS cloud. With **Redshift Spectrum**, an analyst can perform SQL queries on data stored in Amazon S3 buckets

What is ElastiCache used for?

Amazon **ElastiCache** is a Caching-as-a-Service from Amazon Web Services. AWS simplifies setting up, managing, and scaling a distributed in-memory cache environment in the cloud. It provides a high-performance, scalable, and cost-effective caching solution.

Amazon Elastic File System - EFS

Description

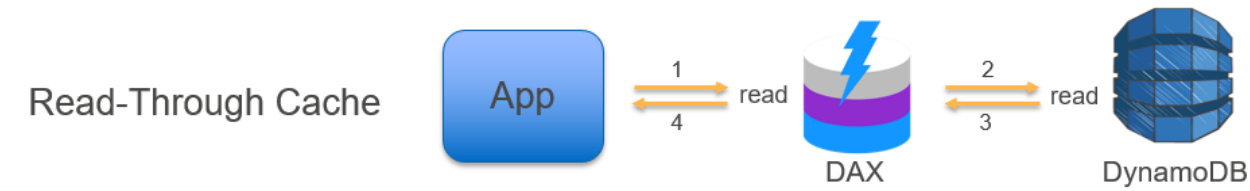
Amazon Elastic File System is a cloud storage service provided by Amazon Web Services designed to provide scalable, elastic, concurrent with some restrictions, and encrypted file storage for use with both AWS cloud services and on-premises resources.

Redis replication group. A **replication group** is a collection of cache clusters, where one of the clusters is a primary read-write cluster and the others are read-only replicas.

Route Origin Authorization (ROA) The ROA authorizes Amazon to advertise an address range under a specific AS number. You can simply create a Route Origin Authorization (ROA) then once done, provision and advertise your whitelisted IP address range to your AWS account.

---

amazon DynamoDB Accelerator (DAX) is a fully managed, highly available, in-memory cache that can reduce Amazon DynamoDB response times from milliseconds to microseconds, even at millions of requests per second.



References:
https://aws.amazon.com/dynamodb/dax

https://aws.amazon.com/device-farm

---

since the company is using Microsoft Active Directory which implements **Security Assertion Markup Language (SAML),** you can set up a SAML-Based Federation for API Access to your AWS cloud. In this way, you can easily connect to AWS using the login credentials of your on-premises network.

AWS supports identity federation with SAML 2.0, an open standard that many identity providers (IdPs) use. This feature enables federated single sign-on (SSO), so users can log into the AWS Management Console

---

The word ***ephemeral*** means *"short-lived"* or *"temporary"*

AWS, always consider this as just a temporary memory or a short-lived storage.

However, data in the instance store is lost under the following circumstances:

- The underlying disk drive fails

- The instance stops

- The instance terminates

Since you are using an EC2 instance with an Instance store, its data is ephemeral which means that it will be erased

---

LOAD BALANCER -When the load balancer determines that an instance is unhealthy, it stops routing requests to that instance. The load balancer resumes routing requests to the instance when it has been restored to a healthy state.

The question refers to one of the common scenarios for temporary credentials in AWS. Temporary credentials are useful in scenarios that involve identity federation, delegation, cross-account access, and IAM roles. In this example, it is called **enterprise identity federation** considering that you also need to set up a single sign-on (SSO) capability.

In Amazon Redshift, you use workload management (WLM) to define the number of query queues that are available, and how queries are routed to those queues for processing. WLM is part of parameter group configuration. A cluster uses the WLM configuration that is specified in its associated parameter group.

Amazon Elastic File System (Amazon EFS) provides simple, scalable, elastic file storage for use with AWS Cloud services and on-premises resources. When mounted on Amazon EC2 instances,

-AWS **CloudTrail** is a service that enables governance, compliance, operational auditing, and risk auditing of your AWS account. With CloudTrail, you can log, continuously monitor, and retain account activity related to actions across your AWS infrastructure.----------
---

Amazon **ElastiCache** is a web service that makes it easy to deploy, operate, and scale an in-memory data store or cache in the cloud. The service improves the performance of web applications by allowing you to retrieve information from fast, managed, in-memory data stores, instead of relying entirely on slower disk-based databases.

In case that one of the EC2 instances failed a health check, the Application Load Balancer stops sending traffic to that instance.

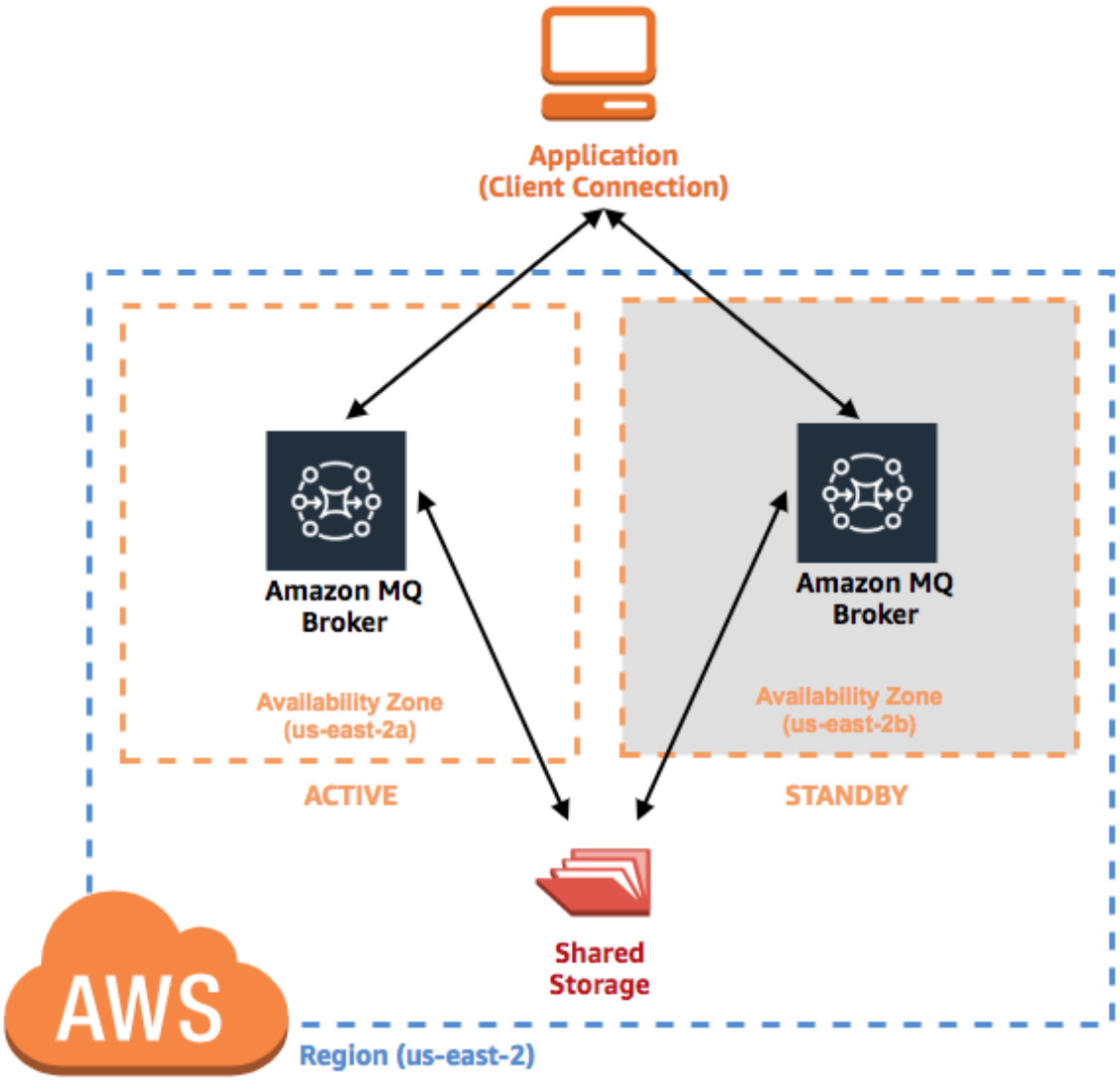
Your Application Load Balancer periodically sends requests to its registered targets to test their status. These tests are **called *health checks*.**

Amazon RDS Multi-AZ deployments provide enhanced availability and durability for Database (DB) Instances, making them a natural fit for production database workloads. When you provision a Multi-AZ DB Instance, Amazon RDS automatically creates a primary DB Instance and synchronously replicates the data to a standby instance in a different Availability Zone (AZ).

Amazon MQ

has a message broker service

Amazon MQ, AmazonSQS and Amazon SNS.all are messaging service



Amazon **API Gateway** provides **throttling** at multiple levels including global and by a service call. Throttling limits can be set for standard rates and bursts. For example, API owners can set a rate limit of 1,000 requests per second for a specific method in their REST APIs, and also configure Amazon API Gateway to handle a burst of 2,000 requests per second for a few seconds.

https://aws.amazon.com/api-gateway/faqs/#Throttling_and_Caching

---

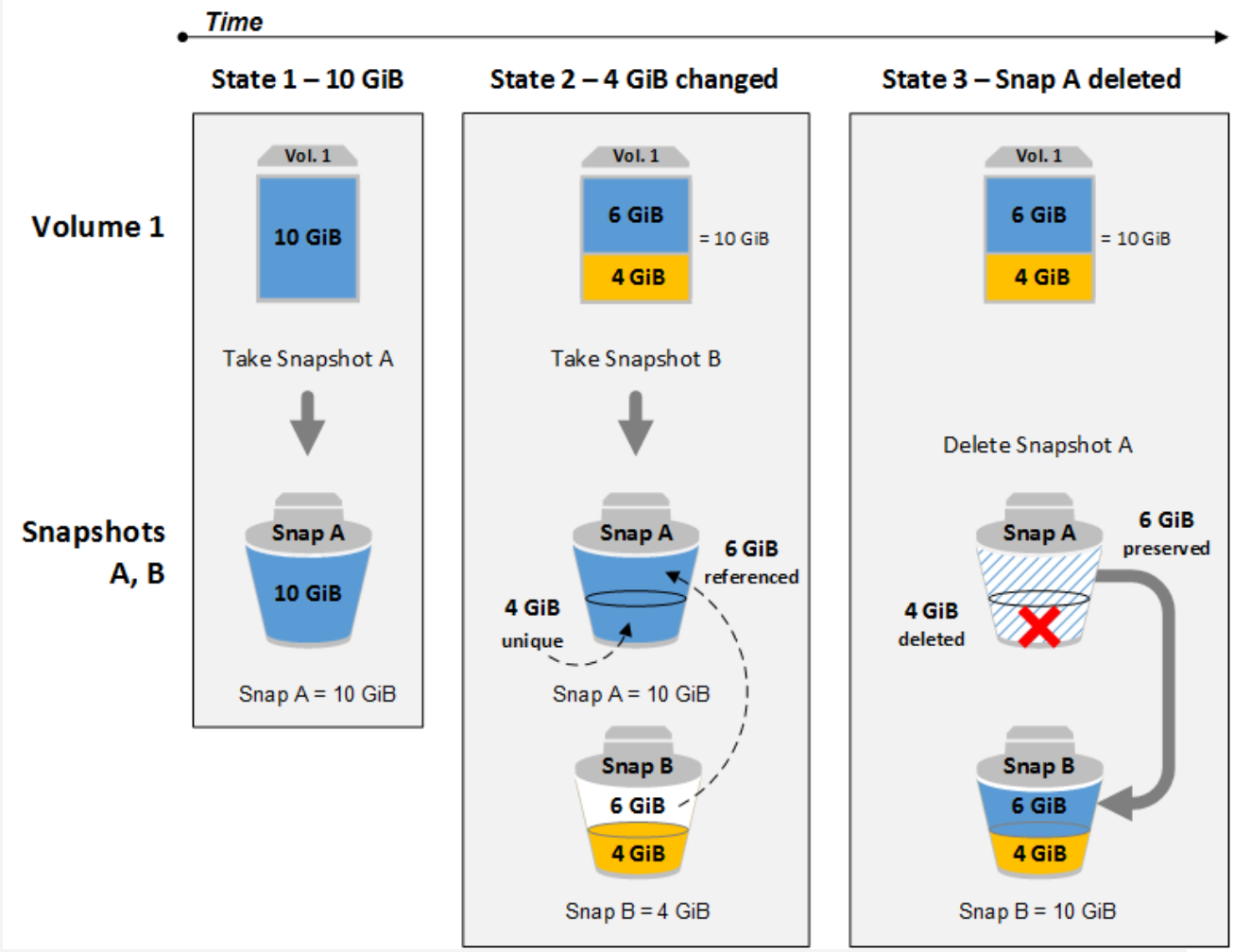
Amazon Aurora typically involves a cluster of DB instances instead of a single instance. Each connection is handled by a specific DB instance. When you connect to an Aurora cluster, the host name and port that you specify point to an intermediate handler called an ***endpoint*.**

AWS Security Token Service (STS) is a web service that enables you to request temporary, limited-privilege credentials for AWS Identity and Access Management (IAM) users or for users that you authenticate (federated users). It is not used for encrypting data unlike KMS.

For higher levels of protection against attacks targeting your applications running on Amazon Elastic Compute Cloud (EC2), Elastic Load Balancing(ELB), Amazon CloudFront, and Amazon Route 53 resources, you can subscribe to AWS Shield Advanced. In addition to the network and transport layer protections that come with Standard, **AWS Shield** Advanced provides additional detection and mitigation against large and sophisticated DDoS attacks, near real-time visibility into attacks, and integration with AWS WAF, a web application firewall.

A single **Amazon SQS message** queue can contain an unlimited number of messages. However, there is a 120,000 limit for the number of inflight messages for a standard queue and 20,000 for a FIFO queue. Messages are inflight after they have been received from the queue by a consuming component, but have not yet been deleted from the queue.

You can back up the data on your **Amazon EBS volumes** to Amazon S3 by taking point-in-time snapshots. Snapshots are *incremental* backups, which means that only the blocks on the device that have changed after your most recent snapshot are saved.

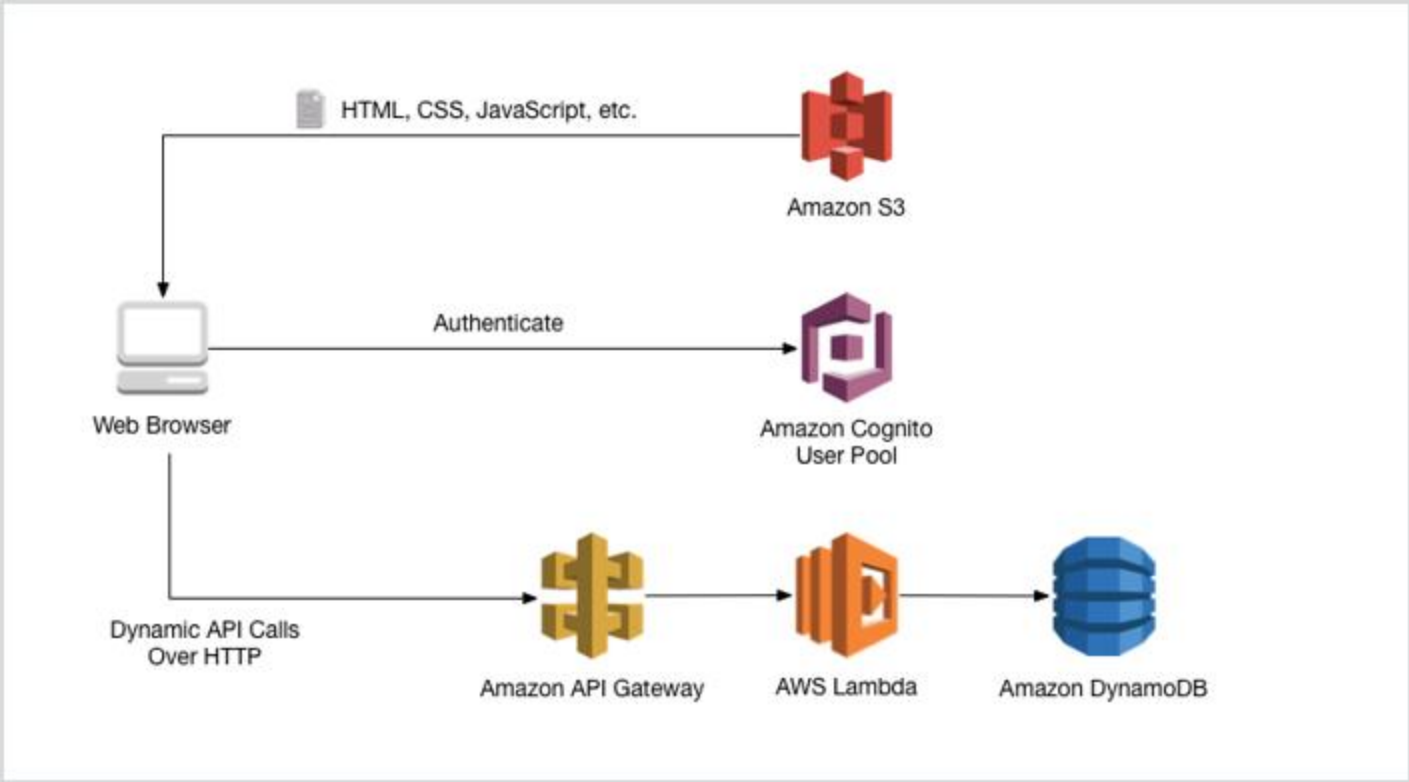


This minimizes the time required to create the snapshot and saves on storage costs by not duplicating data. When you delete a snapshot, only the data unique to that snapshot is removed. Each snapshot contains all of the information needed to restore your data (from the moment the snapshot was taken) to a new EBS volume.

https://tutorialsdojo.com/aws-cheat-sheet-amazon-ebs/

---

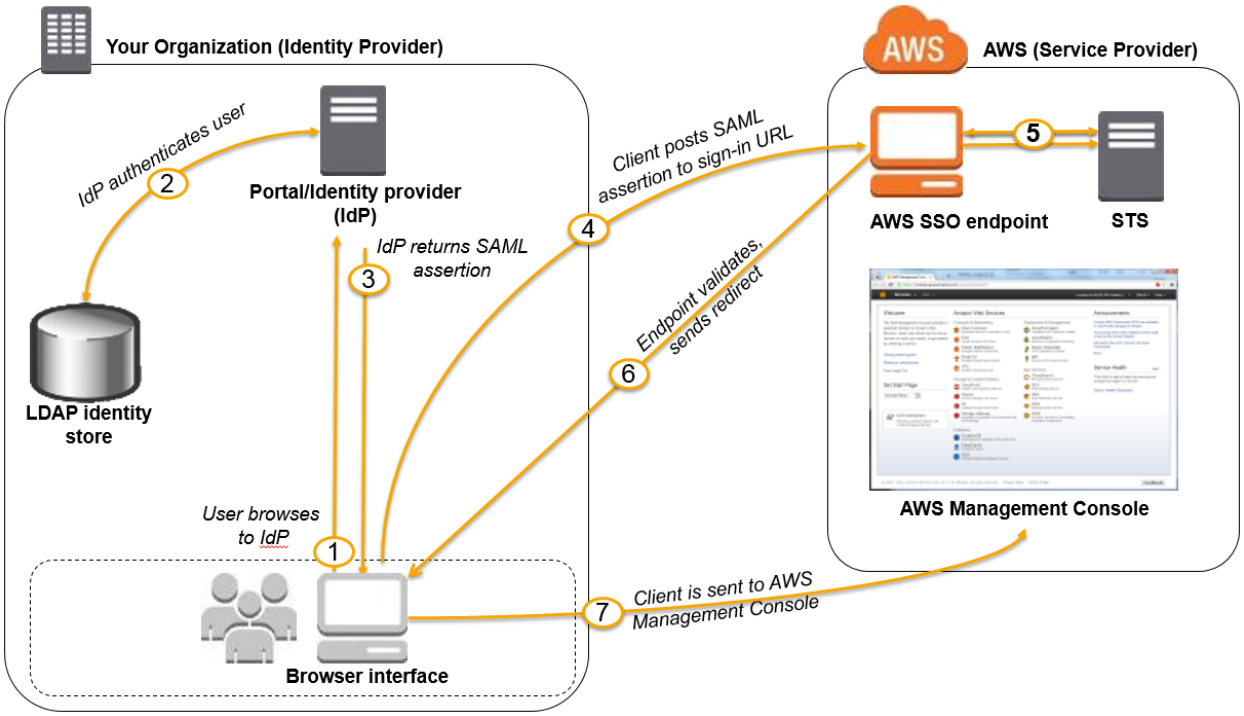
amazon API Gateway is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale. With a few clicks in the AWS Management Console, you can create an API that acts as a “front door” for applications to access data, business logic, or functionality from your back-end services, such as workloads running on Amazon Elastic Compute Cloud (Amazon EC2), code running on AWS Lambda, or any web application. Since it can use AWS Lambda, you can run your APIs without servers.



Amazon API Gateway handles all the tasks involved in accepting and processing up to hundreds of thousands of concurrent API calls, including traffic management, authorization and access control, monitoring, and API version management. Amazon API Gateway has no minimum fees or startup costs. You pay only for the API calls you receive and the amount of data transferred out.

Since the company is using Microsoft Active Directory which implements Security Assertion Markup Language (SAML), you can set up a SAML-Based Federation for API Access to your AWS cloud. In this way, you can easily connect to AWS using the login credentials of your on-premises network.

AWS supports identity federation with SAML 2.0, an open standard that many identity providers (IdPs) use. This feature enables federated single sign-on (SSO), so users can log into the AWS Management Console or call the AWS APIs without you having to create an IAM user for everyone in your organization. By using SAML, you can simplify the process of configuring federation with AWS, because you can use the IdP's service instead of writing custom identity proxy code.



Amazon Elastic File System (Amazon EFS) provides simple, scalable, elastic file storage for use with AWS Cloud services and on-premises resources. When mounted on Amazon EC2 instances, an Amazon EFS file system provides a standard file system interface and file system access semantics, allowing you to seamlessly integrate Amazon EFS with your existing applications and tools. Multiple Amazon EC2 instances can access an Amazon EFS file system at the same time, allowing Amazon EFS to provide a common data source for workloads and applications running on more than one Amazon EC2 instance.

AWS OpsWorks is a configuration management service that provides managed instances of Chef and Puppet.

Amazon S3. It's a simple storage service that offers a highly-scalable, reliable, and low-latency data storage infrastructure at very low costs.

Amazon DynamoDB is a fast and flexible NoSQL database service for all applications that need consistent, single-digit millisecond latency at any scale. It is a fully managed cloud database and supports both document and key-value store models. Its flexible data model, reliable performance, and automatic scaling of throughput capacity makes it a great fit for mobile, web, gaming, ad tech, IoT, and many other applications.

Lambda function traffic shifted

-Canary: Traffic is shifted in two increments. You can choose from predefined canary options that specify the percentage of traffic shifted to your updated Lambda function version in the first increment and the interval, in minutes, before the remaining traffic is shifted in the second increment.

Amazon RDS automatically performs a failover in the event of any of the following:

1. Loss of availability in primary Availability Zone
2. Loss of network connectivity to primary
3. Compute unit failure on primary
4. Storage failure on primary

Instance metadata is the data about your instance that you can use to configure or manage the running instance. You can get the instance ID, public keys, public IP address and many other information from the instance metadata by firing a URL command in your instance to this URL:
http://169.254.169.254/latest/meta-data/

IAM group is a collection of IAM users. Groups let you specify permissions for multiple users, which can make it easier to manage the permissions for those users.

you can manage **identity providers** using IAM Dashboard instead of creating IAM users in your AWS account. With an identity provider (IdP), you can manage your user identities outside of AWS and give these external user identities permission to use AWS resources in your account.

Explanation

Amazon S3 Standard - Infrequent Access (Standard - IA) is an Amazon S3 storage class for data that is accessed less frequently, but requires rapid access when needed. Standard - IA offers the high durability, throughput, and low latency of Amazon S3 Standard, with a low per GB storage price and per GB retrieval fee.

This combination of low cost and high performance make Standard - IA ideal for long-term storage, backups, and as a data store for disaster recovery. The Standard - IA storage class is set at the object level and can exist in the same bucket as Standard, allowing you to use lifecycle policies to automatically transition objects between storage classes without any application changes.

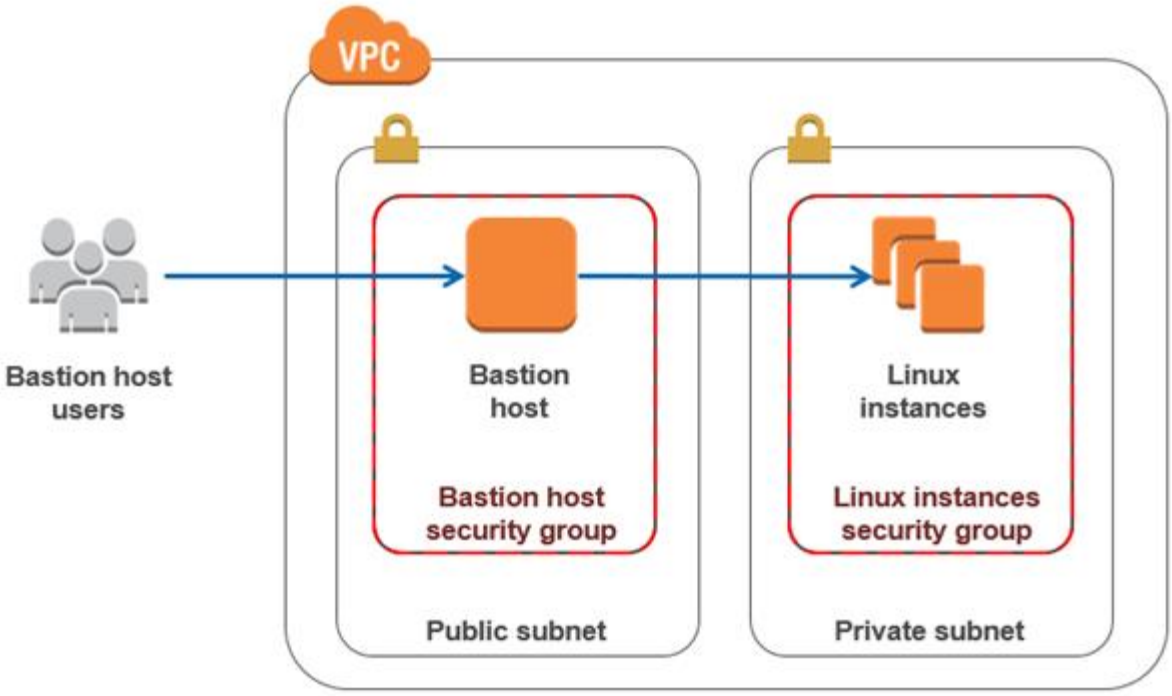
Key Features:

- - Same low latency and high throughput performance of Standard
- - Designed for durability of 99.999999999% of objects
- - Designed for 99.9% availability over a given year
- - Backed with the Amazon S3 Service Level Agreement for availability
- - Supports SSL encryption of data in transit and at rest
- - Lifecycle management for automatic migration of objects

Bastion Host

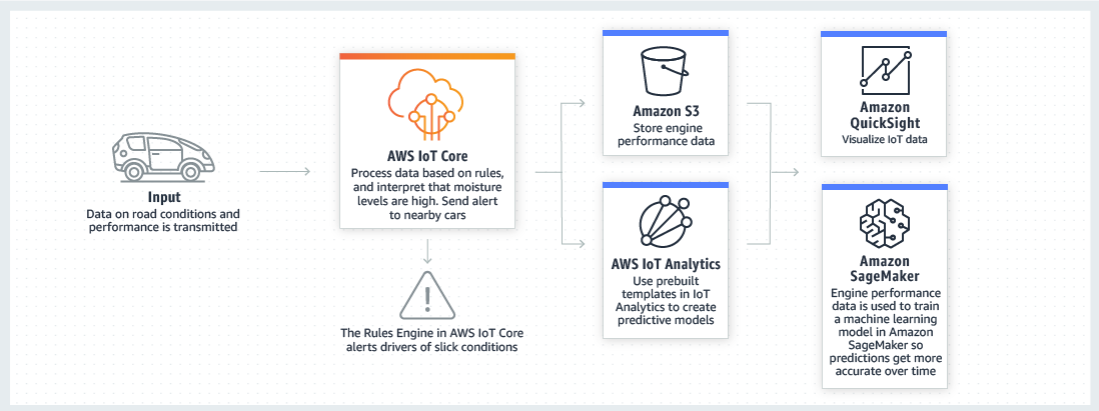
A bastion host is a server whose purpose is to provide access to a private network from an external network, such as the Internet. Because of its exposure to potential attack, a bastion host must minimize the chances of penetration. For example, you can use a bastion host to mitigate the risk of

allowing SSH connections from an external network to the Linux instances launched in a private subnet of your Amazon Virtual Private Cloud (VPC). a **bastion h**ost is to create a small EC2 instance which should only have a security group from a particular IP address for maximum security.



.

AWS IoT Core is a managed cloud service that lets connected devices easily and securely interact with cloud applications and other devices. AWS IoT Core provides secure communication and data processing across different kinds of connected devices and locations so you can easily build IoT applications.



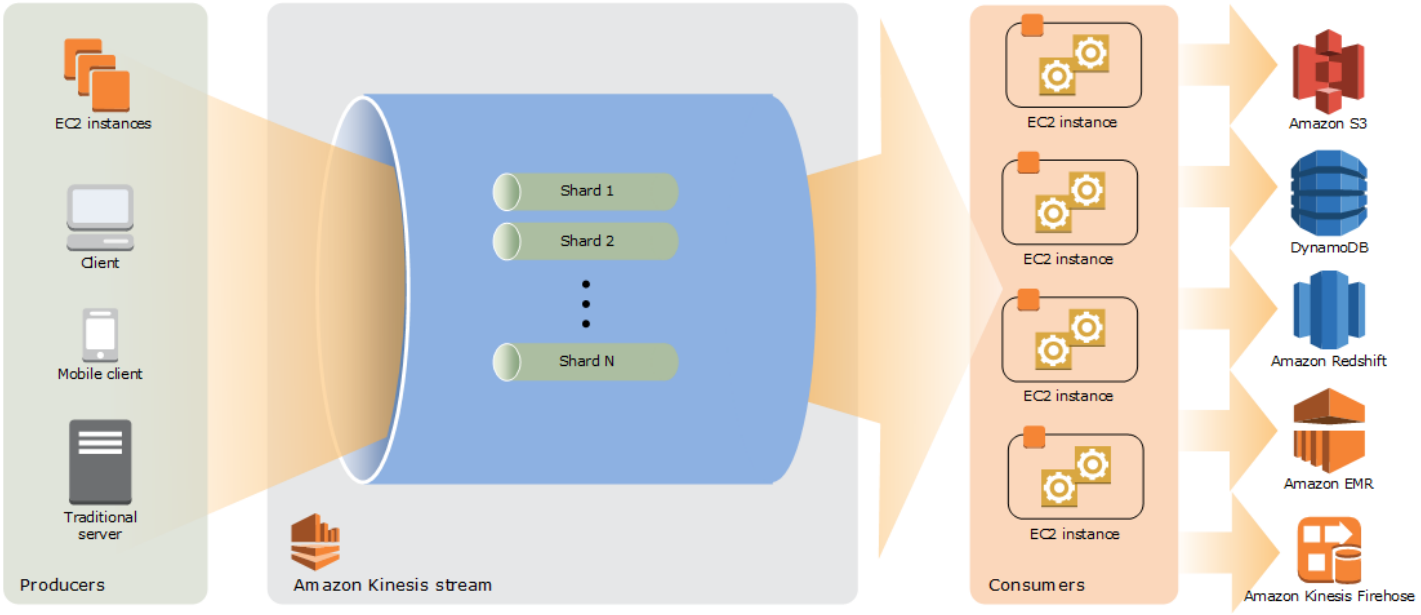
====----------

Perfect Forward Secrecy is a feature that provides additional safeguards against the eavesdropping of encrypted data, through the use of a unique random session key. This prevents the decoding of captured data, even if the secret long-term key is compromised.

CloudFront and Elastic Load Balancing are the two AWS services that support Perfect Forward Secrecy

---

A new shard iterator is returned by every **GetRecords** request (as `NextShardIterator`), which you then use in the next **GetRecords** request (as `ShardIterator`). Typically, this shard iterator does not expire before you use it. However, you may find that shard iterators expire because you have not called **GetRecords** for more than 5 minutes, or because you've performed a restart of your consumer application.



If the shard iterator expires immediately before you can use it, this might indicate that the DynamoDB table used by Kinesis does not have enough capacity to store the lease data. This situation is more likely to happen if you have a large number of shards. To solve this problem, increase the write capacity assigned to the shard table

---

Amazon Simple Queue Service (SQS) and Amazon Simple Workflow Service (SWF) are the services that you can use for creating a decoupled architecture in AWS.

An **elastic network interface (ENI)** is a logical networking component in a VPC that represents a virtual network card. You can attach a network interface to an EC2 instance in the following ways:

1. When it's running (hot attach)

2. When it's stopped (warm attach)
3. When the instance is being launched (cold attach).

Amazon Simple Notification Service (Amazon SNS) topic - A web service that coordinates and manages the delivery or sending of messages to subscribing endpoints or clients.

Amazon Simple Queue Service (Amazon SQS) queue - Offers reliable and scalable hosted queues for storing messages as they travel between computer.

AWS Lambda - AWS Lambda is a compute service where you can upload your code and the service can run the code on your behalf using the AWS infrastructure. You package up and upload your custom code to AWS Lambda when you create a Lambda function

---

You can create a new Route 53 with the failover option to a static S3 website bucket or CloudFront distribution as an alternative.

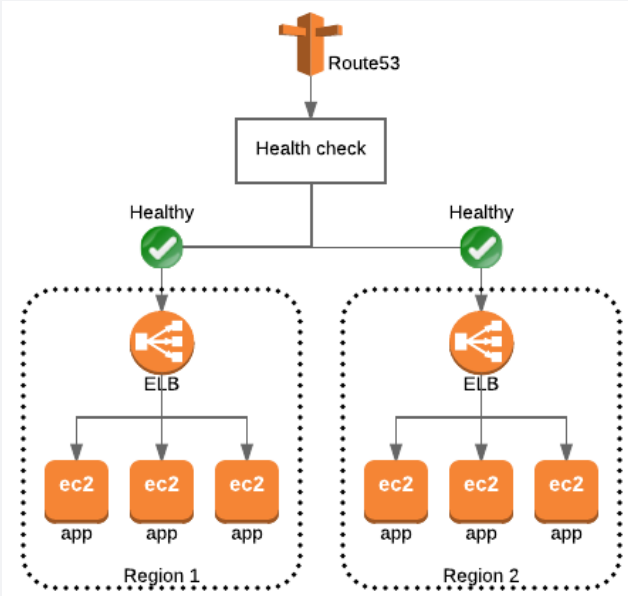


Figure 1 - Both regions operating normally

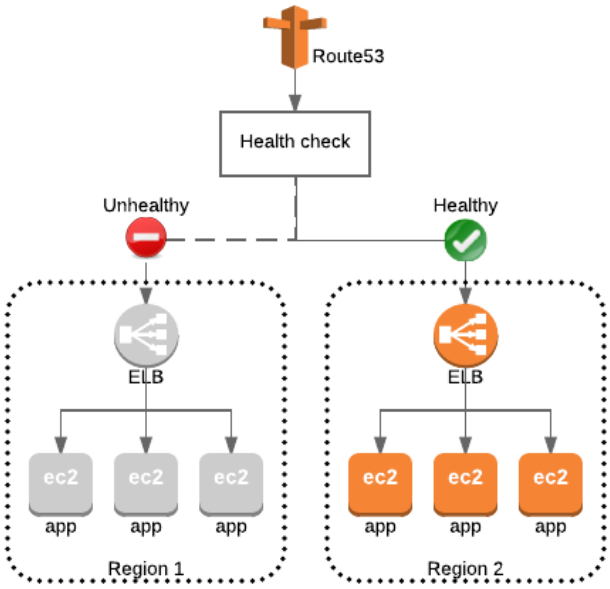


Figure 2 - region 1 experiencing issues

WS Security Token Service (AWS STS) is the service that you can use to create and provide trusted users with temporary security credentials that can control access to your AWS resources. Temporary security credentials work almost identically to the long-term access key credentials that your IAM users can use

---

You can use **Amazon Data Lifecycle Manager (Amazon DLM)** to automate the creation, retention, and deletion of snapshots taken to back up your Amazon EBS volumes. Automating snapshot management helps you to:

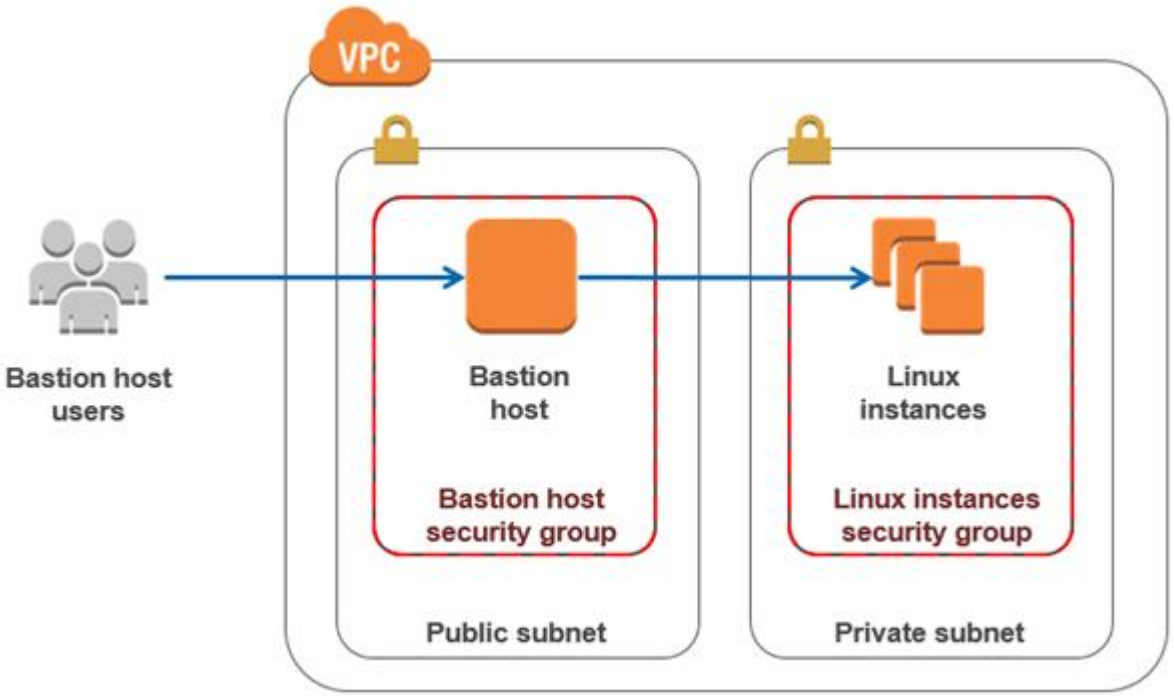
- -Protect valuable data by enforcing a regular backup schedule.
- -Retain backups as required by auditors or internal compliance.
- -Reduce storage costs by deleting outdated backups.

---

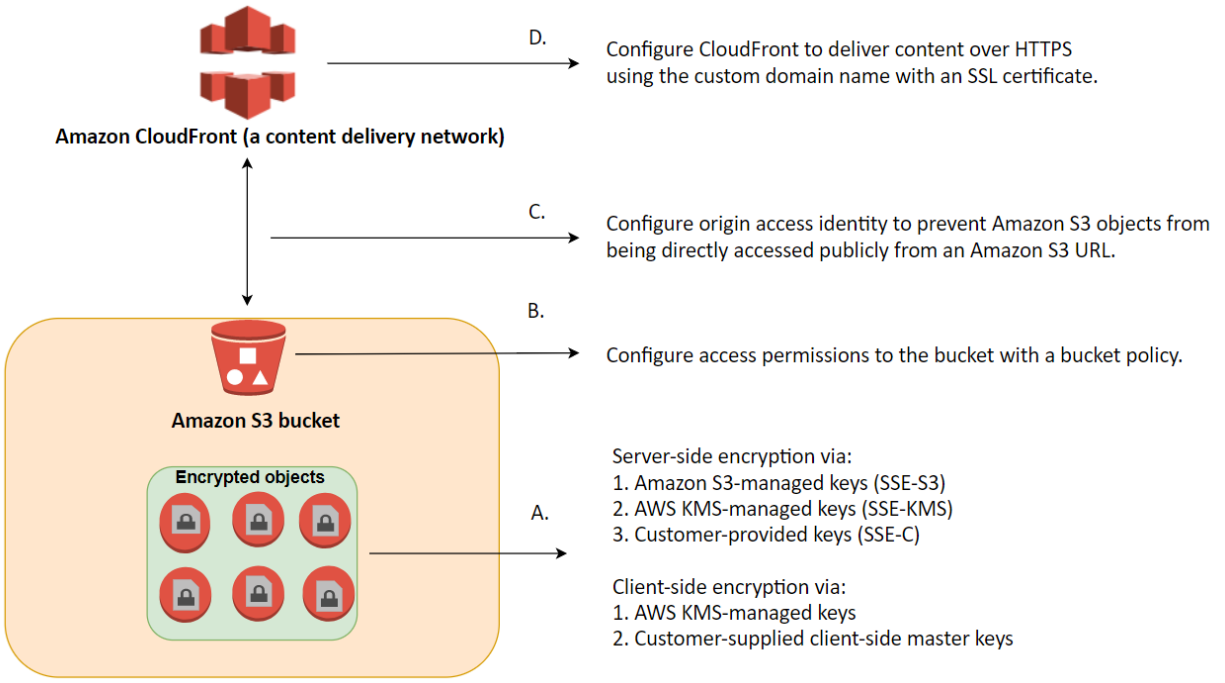
deploy a **Windows Bastion host** with an Elastic IP address in the public subnet and

allow RDP access to bastion only from the corporate IP addresses.

- **A bastion host** is a special purpose computer on a network specifically designed and configured to withstand attacks. If you have a bastion host in AWS, it is basically just an EC2 instance. It should be in a public subnet with either a public or Elastic IP address with sufficient RDP or SSH access defined in the security group. Users log on to the bastion host via SSH or RDP and then use that session to manage other hosts in the private subnets.
-



-



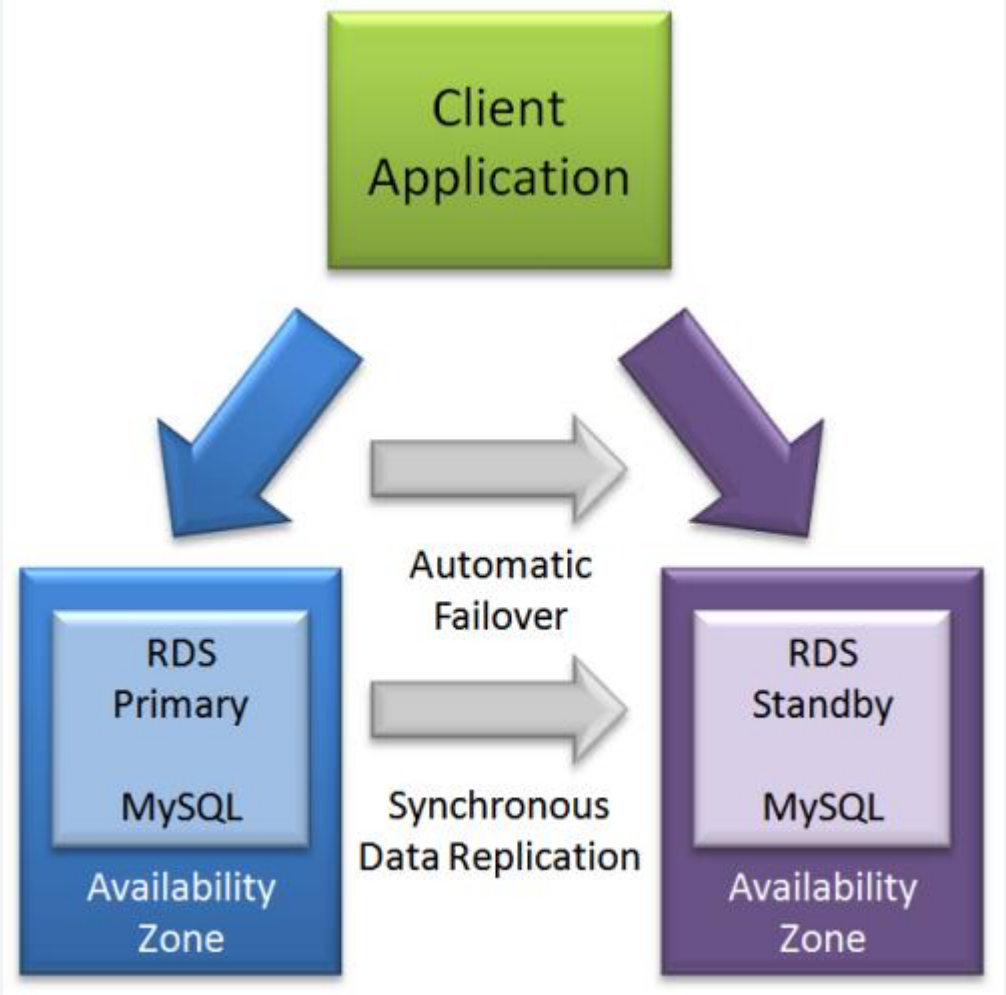
Server-side encryption is about data encryption at rest—that is, Amazon S3 encrypts your data at the object level as it writes it to disks in its data centers and decrypts it for you when you access it. As long as you authenticate your request and you have access permissions, there is no difference in the way you access encrypted or unencrypted objects. For example, if you share your objects using a pre-signed URL, that URL works the same way for both encrypted and unencrypted objects.

Pilot Light.

The term pilot light is often used to describe a DR scenario in which a minimal version of an environment is always running in the cloud. The idea of the pilot light is an analogy that comes from the gas heater. In a gas heater, a small flame that's always on can quickly ignite the entire furnace to heat up a house. This scenario is similar to a backup-and-restore scenario.

For example, with AWS you can maintain a pilot light by configuring and running the most critical core elements of your system in AWS. When the time comes for recovery, you can rapidly provision a full-scale production environment around the critical core.

In **Amazon RDS, failover** is automatically handled so that you can resume database operations as quickly as possible without administrative intervention in the event that your primary database instance went down. When failing over, Amazon RDS simply flips the canonical name **record (CNAME) for** your DB instance to point at the standby, which is in ***turn promoted to become the new primary.***

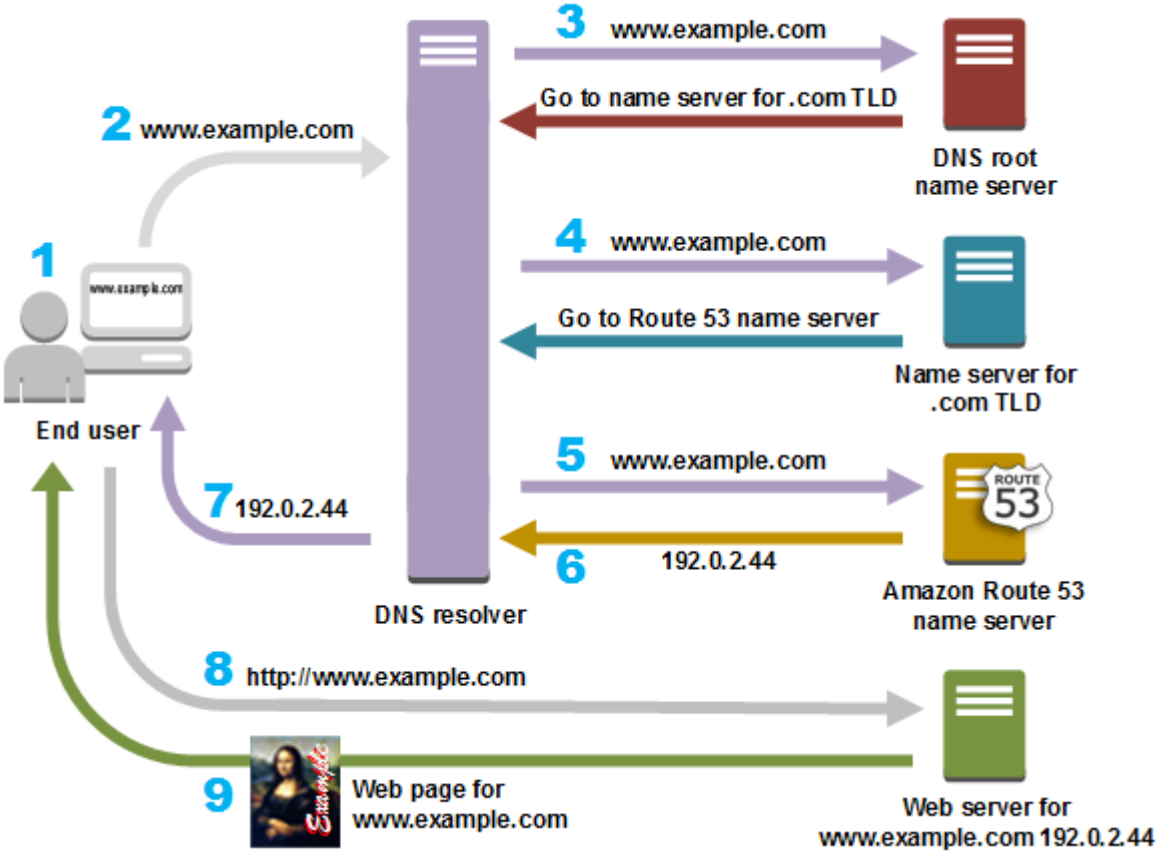


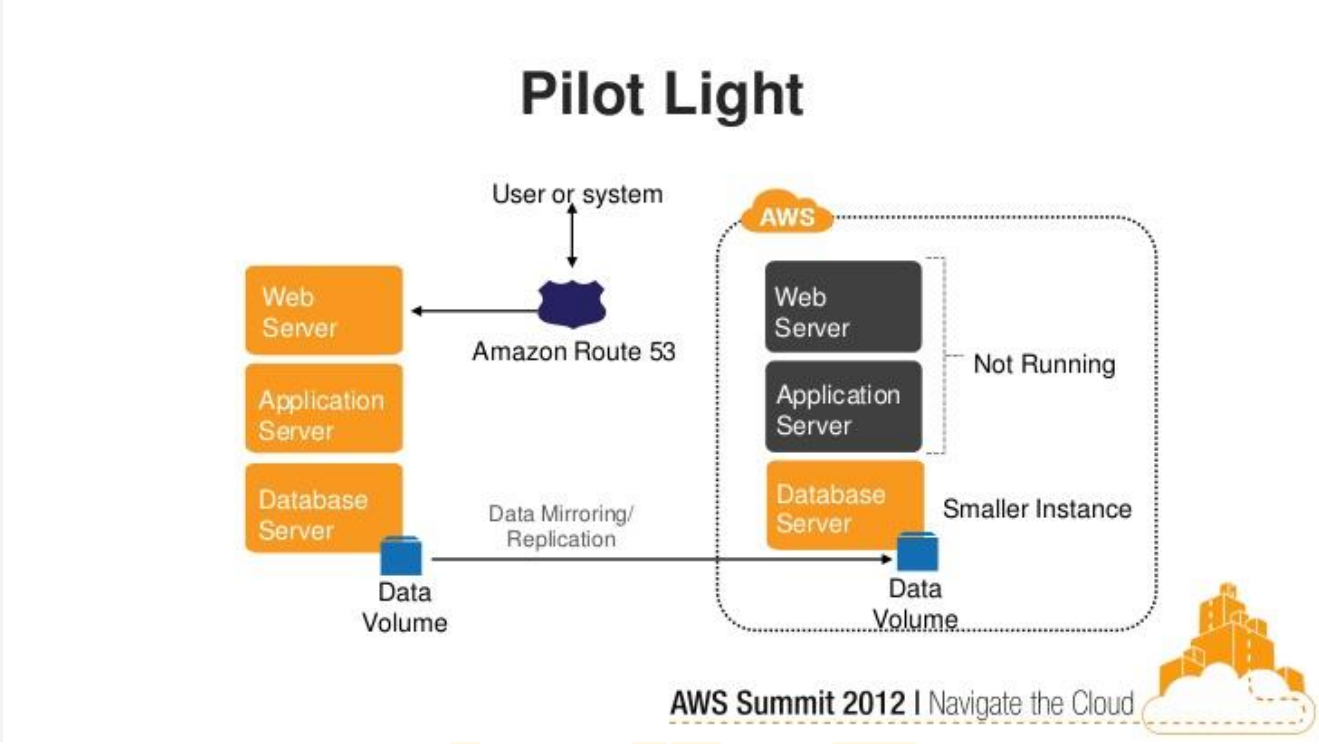
You can use **Amazon Cognito** to deliver temporary, limited-privilege credentials to your application so that your users can access AWS resources. Amazon Cognito identity pools support both authenticated and unauthenticated identities. You can retrieve a unique Amazon Cognito identifier (identity ID) for your end user immediately if you're allowing unauthenticated users or after you've set the login tokens in the credentials provider if you're authenticating users.

Geolocation routing lets you choose the resources that serve your traffic based on the geographic location of your users, meaning the location that DNS queries originate from. For example, you might want all queries from Europe to be routed to an ELB load balancer in the Frankfurt region.

When you use geolocation routing, you can localize your content and present some or all of your website in the language of your users. You can also use geolocation routing to restrict distribution of content to only the locations in which you have distribution rights. Another

possible use is for balancing load across endpoints in a predictable, easy-to-manage way, so that each user location is consistently routed to the same endpoint.





When the word durability pops out, the first service that should come to your mind is Amazon S3. Since this service is not available in the answer options, we can look at the other data store available which is **Amazon DynamoDB**

AWS

Environment Variable Encryption

When you create or update **Lambda functions** that use environment variables, AWS Lambda encrypts them using the AWS Key Management Service. When your Lambda

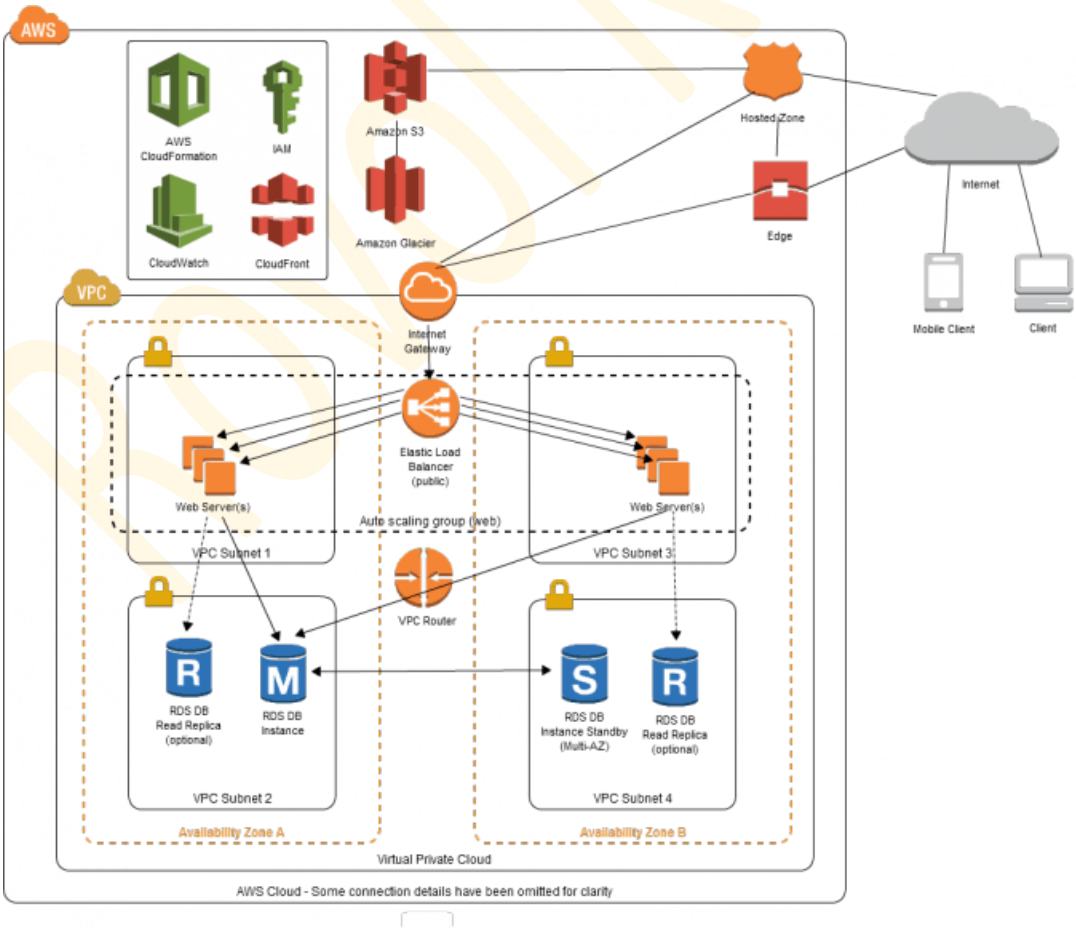
function is invoked, those values are decrypted and made available to the Lambda code.

The first time you create or update Lambda functions that use environment variables in a region, a default service key is created for you automatically within AWS KMS. This key is used to encrypt environment variables

TWO-TIER APPLICATION DEPLOYMENT

In our reference design, we have spread all resources across two availability zones (AZ) to provide for redundancy and resilience to cater for unexpected outages or scheduled system maintenance.

Our web tier consists of two web servers (one in each availability zone) that are deployed on Elastic Compute Cloud (EC2) instances. We balance external traffic to the servers using Elastic Load Balancers



(ELB).

What is Amazon Kinesis streams?

Amazon Kinesis Data **Streams** (KDS) is a massively scalable and durable real-time data **streaming** service. ... The data collected is available in milliseconds to enable real-time analytics use cases such as real-time dashboards, real-time anomaly detection, dynamic pricing, and more.

Default Amazon Cloud Watch metrics for ec2 ? memory comes is Custom Cloud watch

- CPU utilization
- Network utilization
- Disk performance
- Disk Reads/Writes
- Memory utilization, disk swap utilization, disk space utilization, page file utilization, log collection

what is cloudwatch tunnel state metric

The VPN **tunnel state** is reported as a Boolean value in the **CloudWatch metric** TunnelState, where 0 indicates that the **tunnel** is down and 1 indicates that the **tunnel** is up. You can set up a **CloudWatch** alarm based on this **CloudWatch metric** to notify you when one or both VPN **tunnels** are down

Throughput Optimized HDD (st1) Volumes

Throughput Optimized HDD (`st1`) volumes **provide low-cost magnetic storage** that defines performance in terms of throughput rather than IOPS. This volume type is a good fit for large, sequential workloads such as Amazon EMR, ETL, data warehouses, and log processing. Bootable `st1` volumes are not supported.

Throughput Optimized HDD (`st1`) volumes, though similar to Cold HDD (`sc1`) volumes, are designed to support *frequently* accessed data.

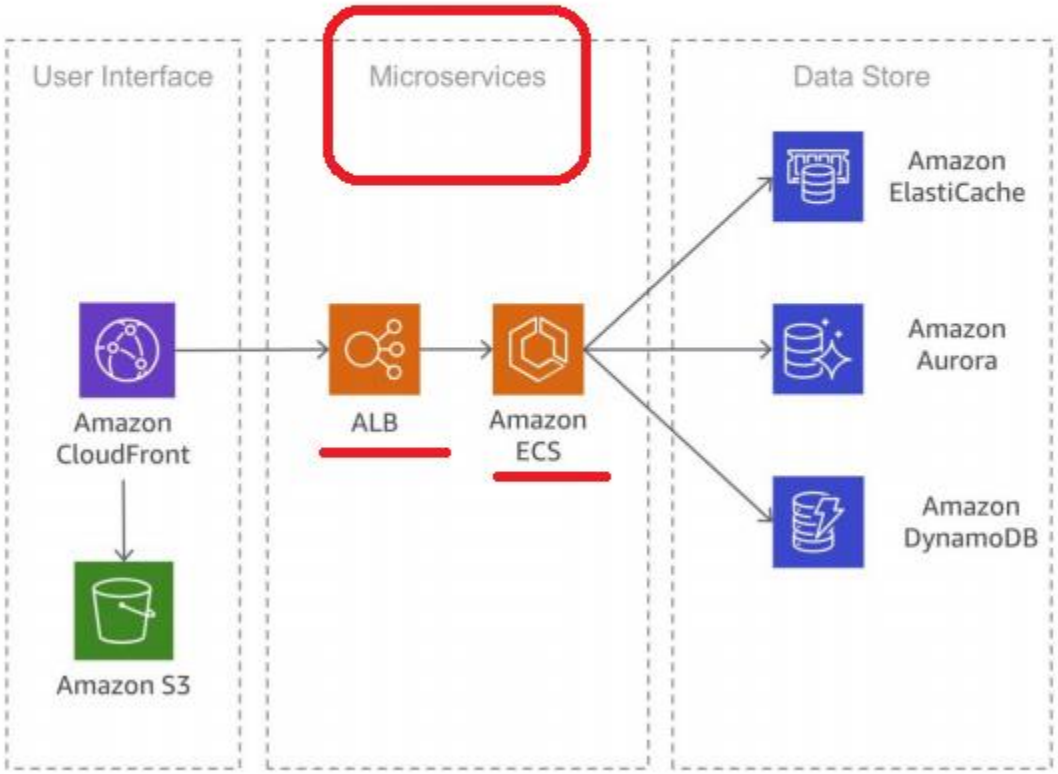
This volume type is optimized for workloads **involving large, sequential I/O**, and we recommend that customers with workloads performing small, random I/O use `gp2`.

How do I create a Microservice in AWS?

Create an API Using Amazon API Gateway

1. Sign in to the AWS Management Console and open the AWS Lambda console.
2. Choose Create **Lambda function.**
3. Choose Blueprint.
4. Enter microservice in the search bar. Choose the microservice-http-endpoint blueprint and then choose Configure.
5. Configure the following settings.

domain. Figure 1 depicts a reference architecture for a typical microservices application on AWS.



Scheduled Reserved Instances

Scheduled Reserved Instances (Scheduled Instances) enable you to purchase capacity reservations that recur on a daily, weekly, or monthly basis, with a specified start time and duration, for a one-year term. You reserve the capacity in advance, so that you know it is available when you need it. You pay for the time that the instances are scheduled, even if you do not use them.

Amazon Kinesis Data Analytics is the easiest way to analyze streaming data, gain actionable insights, and respond to your business and customer needs in real time. Amazon Kinesis Data Analytics reduces the complexity of building, managing, and integrating streaming applications with other AWS services. SQL users can easily query streaming data or build entire streaming applications using templates and an interactive SQL editor. Java developers can quickly build sophisticated streaming applications using open source Java libraries and AWS integrations to transform and analyze data in real-time.

Adding Multi-Factor Authentication (MFA) to a User Pool

Multi-factor authentication (MFA) increases security for your app by adding another authentication method, and not relying solely on user name and password. You can choose to use SMS text messages, or time-based one-time (TOTP) passwords as second factors in signing in your users.

- You can only choose MFA as **Required** when you initially create a user pool.
- Phone numbers must be verified if MFA is enabled and **SMS text message** is chosen as a second factor.
- Advanced security features require that MFA is enabled, and set as optional in the Amazon Cognito user pool console

What Is Indexing in DynamoDB?

Amazon DynamoDB provides fast access to items in a table by specifying primary key values. But if you want to fetch the data of attributes other than the primary key, indexing comes into the picture.

DynamoDb provides two types of indexing:

1. Global secondary index
2. Local secondary index

AWS KMS is integrated with AWS CloudTrail to record all API requests, including key management actions and usage of your keys.

Encryption of Data at Rest

Managing Keys

Amazon EFS is integrated with AWS KMS, which manages the encryption keys for encrypted file systems. AWS KMS also supports encryption by other AWS services such as Amazon Simple Storage Service (Amazon S3), **Amazon Elastic Block Store (Amazon EBS),** Amazon Relational Database Service (Amazon RDS), Amazon Aurora, Amazon Redshift, Amazon WorkMail, Amazon WorkSpaces, etc. To encrypt file system contents, Amazon EFS uses the Advanced Encryption Standard algorithm with XTS Mode and a 256-bit key (XTS-AES-256).

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

Synchronous AWS Decoupling Solutions

In order to achieve synchronous decoupling an Elastic Load Balancing (ELB) could be used. An ELB distributes incoming application traffic across multiple EC2 instances, in multiple Availability Zones. You can add and remove instances from your load balancer as your needs change, without disrupting the overall flow of requests to your application.

What is API gateway used for?
Summary. For most microservices-based applications, it makes sense to implement an **API Gateway**, which acts as a single entry point into a system. The **API Gateway** is responsible for request routing, composition, and protocol translation. It provides each of the application's clients with a custom **API**

Amazon Elastic File System (Amazon EFS) now supports 512 locks per file. With Amazon S3 object lock, you can store objects using a write-once-read-many (WORM) model. You can use it to prevent an object from being deleted or overwritten for a fixed amount of time or indefinitely.

Amazon S3 vs EFS vs EBS Comparison

In summary, we distinguished a few specific features of all three storage services:

AMAZON S3	AMAZON EBS	AMAZON EFS
Can be publicly accessible Web interface Object Storage Scalable Slower than EBS and EFS	Accessible only via the given EC2 Machine File System interface Block Storage Hardly scalable Faster than S3 and EFS	Accessible via several EC2 machines and AWS services Web and file system interface Object storage Scalable Faster than S3, slower than EBS
Good for storing Backup	Is meant to be EC2 drive	Good for shareable applications and workloads

Limits for Amazon EFS File Systems

The following are limits specific to the Amazon EFS file systems:

- Maximum name length: 255 bytes.
- Maximum symbolic link (symlink) length: 4,080 bytes.
- Maximum number of hard links to a file: 177.
- Maximum size of a single file: 52,673,613,135,872 bytes (47.9 TiB).

• Maximum directory depth: 1,000 levels deep.

• **Any one particular file can have up to 512 locks across all instances connected and users accessing the file.**

• In General Purpose mode, there is a limit of 7,000 file system operations per second. This operations limit is calculated for all clients connected to a single file system.

VPC with Public and Private Subnets (NAT)

this scenario can also be optionally configured for IPv6—you can use the VPC wizard to create a VPC and subnets with associated IPv6 CIDR blocks. Instances launched into the subnets can receive IPv6 addresses, and communicate using IPv6. Instances in the private subnet can use an egress-only Internet gateway to connect to the Internet over IPv6, but the Internet cannot establish connections to the private instances over IPv6. For more information about IPv4 and IPv6 addressing

Enabling VPC Flow Logs will help you detect security and access issues like overly permissive security groups and network ACLs and alert abnormal activities triggered within your Virtual Private Cloud network such as rejected connection requests or unusual levels of data transfer

Dynamic Scaling for Amazon EC2 Auto Scaling

When you configure dynamic scaling, you must define how to scale in response to changing demand. For example, you have a web application that currently runs on two instances and you want the CPU utilization of the Auto Scaling group to stay at around 50 percent when the load on the application changes. This gives you extra capacity to handle traffic spikes without maintaining an excessive amount of idle resources. You can configure your Auto Scaling group to scale automatically to meet this need.

Throughput Optimized HDD (st1) Volumes

Throughput Optimized HDD (st1) volumes provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. This volume type is a good fit for large, sequential workloads such as Amazon EMR, ETL, data warehouses, and log processing. Bootable st1 volumes are not supported.

Throughput Optimized HDD (st1) volumes, though similar to Cold HDD (sc1) volumes, are designed to support *frequently* accessed data.

What is AWS LAMP stack?
Launch a **LAMP stack** web application using **AWS** Elastic Beanstalk and Amazon Relational Database Service (RDS). Elastic Beanstalk provisions and manages the underlying infrastructure (e.g., Amazon EC2 instances) and **stack** components (e.g., OS, web server, language/framework) for you. RDS provides the MySQL databas

Kinesis Streams

Kinesis Streams is capable of capturing large amounts of data (terabytes per hour) from data producers, and streaming it into custom applications for data processing and analysis. Streaming data is replicated by Kinesis across three separate availability zones within AWS to ensure reliability and availability of your data.

Kinesis Streams is capable of scaling from a single megabyte up to terabytes per hour of streaming data. You must manually provision the appropriate number of shards for your stream to handle the volume of data you expect to process. Amazon helpfully provides a shard calculator when creating a stream to correctly determine this number. Once created, it is possible to dynamically scale up or down the number of shards to meet demand, but only with the AWS Streams API at this time.

It is possible to load data into Streams using a number of methods, including HTTPS, the Kinesis Producer Library, the Kinesis Client Library, and the Kinesis Agent.

By default, data is available in a stream for 24 hours, but can be made available for up to 168 hours (7 days) for an additional charge.

Monitoring is available through Amazon Cloudwatch.

Kinesis Firehose

Kinesis Firehose is Amazon's data-ingestion product offering for Kinesis. It is used to capture and load streaming data into other Amazon services such as S3 and Redshift. From there, you can load the streams into data processing and analysis tools like Elastic Map Reduce, and Amazon Elasticsearch Service. It is also possible to load the same data into S3 and Redshift at the same time using Firehose.

Firehose can scale to gigabytes of streaming data per second, and allows for batching, encrypting and compressing of data. It should be noted that Firehose will automatically scale to meet demand, which is in contrast to Kinesis Streams, for which you must manually provision enough capacity to meet anticipated needs.

As with Kinesis Streams, it is possible to load data into Firehose using a number of methods, including HTTPS, the Kinesis Producer Library, the Kinesis Client Library, and the Kinesis Agent. Currently, it is only possible to stream data via Firehose to S3 and Redshift, but once stored in one of these services, the data can be copied to other services for further processing and analysis.

Monitoring is available through Amazon Cloudwatch.

VPC Flow Logs. **VPC Flow Logs** is a feature that enables you to capture information about the IP traffic going to and from network interfaces in your **VPC**. **Flow log** data can be published to **Amazon CloudWatch Logs** and **Amazon** S3. After you've created a **flow log**, you can retrieve and view its data in the chosen destination

Web Identity Federation allows you to simplify authentication and authorization for large user groups. You can skip the creation of individual accounts, and require users to login to an **identity** provider to get temporary credentials or tokens. It uses AWS Security Token Service (STS) to manage credentials.

AMAZON ELASTICACHE

ElastiCache is a web service that makes it easy to deploy and run Memcached or Redis protocol-compliant server nodes in the cloud.

The in-memory caching provided by ElastiCache can be used to significantly improve latency and throughput for many read-heavy application workloads or compute-intensive workloads .

Amazon S3	Scalable storage in the cloud
Amazon Glacier	Low-cost archive storage in the cloud
Amazon EBS	Persistent block storage volumes for Amazon EC2 virtual machines
Amazon EC2 Instance Storage	Temporary block storage volumes for Amazon EC2 virtual machines
AWS Import/Export	Large volume data transfer
AWS Storage Gateway	Integrates on-premises IT environments with cloud storage
Amazon CloudFront	Global content delivery network (CDN)
Amazon SQS	Message queue service
Amazon RDS	Managed relational database server for MySQL, Oracle, and Microsoft SQL Server
Amazon DynamoDB	Fast, predictable, highly-scalable NoSQL data store
Amazon ElastiCache	In-memory caching service
Amazon Redshift	Fast, powerful, full-managed, petabyte-scale data warehouse service
Databases on Amazon EC2	Self-managed database on an Amazon EC2 instance

create an encrypted volume from the encrypted snapshot

S3 can achieve at least **3,500 PUT/COPY/POST/DELETE** and **5,500 GET/HEAD requests per second per prefix** in a bucket.

AWS OpsWorks for **Chef** Automate provides a fully managed Chef Automate server and suite of automation tools that give you workflow automation for continuous deployment, automated testing for compliance and security, and a user interface that gives you visibility into your nodes and their status. The Chef Automate platform gives you full stack automation by handling

operational tasks such as software and operating system configurations, continuous compliance, package installations, database setups, and more. The Chef server centrally stores your configuration tasks and provides them to each node in your compute environment at any scale, from a few nodes to thousands of nodes. OpsWorks for Chef Automate is completely compatible with tooling and cookbooks from the Chef community and automatically registers new nodes with your Chef server.

CloudFront Benefits

- CloudFront eliminates the expense and complexity of operating a network of cache servers in multiple sites across the internet and eliminates the need to over-provision capacity in order to serve potential spikes in traffic
- CloudFront also provides increased reliability and availability because copies of objects are held in multiple edge locations around the world
- CloudFront keeps persistent connections with the origin servers so that those files can be fetched from the origin servers as quickly as possible.
- CloudFront also uses techniques such as collapsing simultaneous viewer requests at an edge location for the same file into a single request to the origin server reducing the load on the origin
- CloudFront integrates with AWS WAF, a web application firewall that helps protect web applications from attacks by allowing rules configured based on IP addresses, HTTP headers, and custom URI strings

- supports a live event, such as a meeting, conference, or concert, in real time. For live streaming, distribution can be created automatically using an AWS CloudFormation stack.
- origin servers can be either an Amazon S3 bucket or an HTTP server, *for e.g., a web server or an AWS ELB etc*

What Are CloudTrail Events?

An event in CloudTrail is the record of an activity in an AWS account. This activity can be an action taken by a user, role, or service that is monitorable by CloudTrail. CloudTrail events provide a history of both API and non-API account activity made through the AWS Management Console, AWS SDKs, command line tools, and other

AWS services. There are two types of events that can be logged in CloudTrail: management events and data events. By default, trails log management events, but not data events.

Both management events and data events use the same CloudTrail JSON log format.

Note

CloudTrail does not log all AWS services. Some AWS services do not enable logging of all APIs and events. Even if you configure logging all management and data events in a trail, you will not create a log with all possible AWS events. For details about which APIs are logged for a specific service, see documentation for that service in CloudTrail Supported Services and Integrations.

[An application program interface (API)]

ELB- Elastic Load Balancing automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, IP addresses, and Lambda functions. It can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones. Elastic Load Balancing offers three types of load balancers that all feature the high availability, automatic scaling, and robust security necessary to make your applications fault tolerant.

What is the use of placement group in AWS?
Specifies a **placement group** in which to launch instances. The strategy of the **placement group** determines how the instances are organized within the **group**. A cluster **placement group** is a logical grouping of instances within a single Availability Zone that **benefit from low network latency, high network throughput.**

a. **Internet Gateway (IGW)**
b. **Virtual Private Gateway (VGW)**

Cross-Zone load Balancing

- By default, the load balancer distributes incoming requests evenly across its enabled Availability Zones *for e.g. If AZ-a has 5 instances and AZ-b has 2 instances, the load will still be distributed 50% across each of the AZs*
- Enabling Cross-Zone load balancing allows the ELB to distribute incoming requests evenly across all the back-end instances, regardless of the AZ
- Cross-zone load balancer reduces the need to maintain equivalent numbers of back-end instances in each Availability Zone, and improves application's ability to handle the loss of one or more back-end instances.
- It is still recommended to maintain approximately equivalent numbers of instances in each Availability Zone for higher fault tolerance.

NAT device Configuration Key Points

- needs to be launched in the Public Subnet
- needs to be associated with an Elastic IP address (or public IP address)
- should have the **Source/Destination flag disabled** to route traffic from the instances in the private subnet to the Internet and send the response back
- should have a Security group associated that
  - allows Outbound Internet traffic from instances in the private subnet
  - disallows Inbound Internet traffic from everywhere
- Instances in the private subnet should have the Route table configured to direct all Internet traffic to the NAT device

amazon EBS volume types

The following table shows use cases and performance characteristics of current generation EBS volumes:

	Solid State Drives (SSD)		Hard Disk Drives (HDD)	
Volume Type	**EBS Provisioned IOPS SSD (io1)**	EBS General Purpose SSD (gp2)*	Throughput Optimized HDD (st1)	Cold HDD (sc1)

Short Description	Highest performance SSD volume designed for latency-sensitive transactional workloads	General Purpose SSD volume that balances price performance for a wide variety of transactional workloads	Low cost HDD volume designed for frequently accessed, throughput intensive workloads	Lowest cost HDD volume designed for less frequently accessed workloads
Use Cases	**I/O-intensive NoSQL and relational** databases	Boot volumes, low-latency interactive apps, dev & test	Big data, data warehouses, log processing	Colder data requiring fewer scans per day
API Name	io1	gp2	st1	sc1
Volume Size	4 GB - 16 TB	1 GB - 16 TB	500 GB - 16 TB	500 GB - 16 TB
Max IOPS**/Volume	64,000	16,000	500	250
Max Throughput***/Volume	1,000 MB/s	250 MB/s	500 MB/s	250 MB/s
Max IOPS/Instance	80,000	80,000	80,000	80,000
Max Throughput/Instance	1,750 MB/s	1,750 MB/s	1,750 MB/s	1,750 MB/s
Price	$0.125/GB-month $0.065/provisioned IOPS	$0.10/GB-month	$0.045/GB-month	$0.025/GB-month
Dominant Performance Attribute	IOPS	IOPS	MB/s	MB/s

Route 53

Amazon Route 53 is a highly available and scalable Domain Name System (DNS) web service.

Supported DNS Resource Record Types

- A (Address) Format
  - is an IPv4 address in dotted decimal notation *for e.g. 192.0.2.1*
- AAAA Format
  - is an IPv6 address in colon-separated hexadecimal format
- CNAME Format
  - is the same format as a domain name
- SOA (Start of Authority) Format
  - SOA record provides information about a domain and the corresponding Amazon Route 53 hosted zone
- SPF (Sender Policy Framework) Format
  - SPF records were formerly used to verify the identity of the sender of email messages, however is not recommended
  - **Alias record is similar to a CNAME record, but can create an alias record both for the root domain or apex zone, *such as example.com*, and for subdomains, *such as www.example.com*. CNAME records can be used only for subdomains**

- **Reduced Redundancy Storage (RRS) storage** class is designed for noncritical, reproducible data stored at lower levels of redundancy than the STANDARD storage class, which reduces storage costs
- Designed for durability of 99.99% of objects
- Designed for 99.99% availability over a given year
- This durability level corresponds to an average annual expected loss of 0.01% of objects.

Reduced Redundancy Storage (RRS) is an Amazon option that enables customers to store noncritical, reproducible data at lower levels of standard storage. ... Reduced Redundancy Storage is: Backed with the Amazon Service Level Agreement for availability

New Amazon Cloud-Watch **Custom** Monitoring Scripts

Available metrics include:

- Memory Utilization (%)
- Memory Used (MB)
- Memory Available (MB)

- Swap Utilization (%)
- Swap Used (MB)

- Disk Space Utilization (%)
- Disk Space Used (GB)

Amazon S3 Server Access Logging

Server access logging provides detailed records for the requests that are made to a bucket. Server access logs are useful for many applications. For example, access log information can be useful in security and access audits. It can also help you learn about your customer base and understand your Amazon S3 bill.

The main function of DynamoDB is **to** store **data.** Where as the main function of "*Kinesis*" *is* to **analyze** data in real-time. The requirement in the question is to find an AWS service which provides highly available "datastore".

Restricting Access to Amazon S3 Content by Using an Origin Access Identity- OAI

To restrict access to content that you serve from Amazon S3 buckets, you create CloudFront signed URLs or signed cookies to limit access to files in your Amazon S3

bucket, and then you create a special CloudFront user called an origin access identity (OAI) and associate it with your distribution. Then you configure permissions so that CloudFront can use the OAI to access and serve files to your users, but users can't use a direct URL to the S3 bucket to access a file there. Taking these steps help you maintain secure access to the files that you serve through CloudFront.

To ensure that your users access your files using only CloudFront URLs, regardless of whether the URLs are signed, do the following:

1. Create an origin access identity, which is a special CloudFront user, and associate the origin access identity with your distribution. You associate the origin access identity with origins, so that you can secure all or just some of your Amazon S3 content. You can also create an origin access identity and add it to your distribution when you create the distribution. For more information, see Creating a CloudFront Origin Access Identity and Adding it to Your Distribution.

Amazon EBS–Optimized Instances

An Amazon EBS–optimized instance uses an optimized configuration stack and provides additional, dedicated capacity for Amazon EBS I/O. This optimization provides the best performance for your EBS volumes by minimizing contention between Amazon EBS I/O and other traffic from your instance.

EBS–optimized instances deliver dedicated bandwidth to Amazon EBS, with options between 425 Mib/s and 14,000 Mib/s, depending on the instance type you use. When attached to an EBS–optimized instance, General Purpose SSD (gp2) volumes are designed to deliver their baseline and burst performance 99% of the time, and

Provisioned IOPS SSD (`io1`) volumes are designed to deliver their provisioned performance 99.9% of the time. Both Throughput Optimized HDD (`st1`) and Cold HDD (`sc1`) guarantee performance consistency of 90% of burst throughput 99% of the time. Non-compliant periods are approximately uniformly distributed, targeting 99% of expected total throughput each hour

AWS CloudHSM

- AWS CloudHSM provides secure cryptographic key storage to customers by making **hardware security modules (HSMs)** available in the AWS cloud
- AWS CloudHSM helps meet corporate, contractual and regulatory compliance requirements for data security by using dedicated HSM appliances within the AWS cloud.

- Q: How frequently does the CloudWatch Logs Agent send data?
- The CloudWatch Logs Agent will send log data every five seconds by default and is configurable by the user.

What are lifecycle hooks?

Lifecycle hooks let you take action before an instance goes into service or before it gets terminated. This can be especially useful if you are not baking your software environment into an Amazon Machine Image (AMI). For example, launch hooks can perform software configuration on an instance to ensure that it's fully prepared to handle traffic before Amazon EC2 Auto Scaling proceeds to connect it to your load balancer. One way to do this is by connecting the launch hook to an AWS Lambda function that invokes RunCommand on the instance. Terminate hooks can be useful for collecting important data from an instance before it goes away. For example, you could use a terminate hook to preserve your fleet's log files by copying them to an Amazon S3 bucket when instances go out of service.

Why isn't CloudFront following a cache behavior that I created?

Last updated: 2019-06-21
I created a custom cache behavior for a specific URL path pattern on my distribution, but Amazon CloudFront isn't following the cache behavior. Why?

Resolution

If your CloudFront distribution isn't following a cache behavior that you created, check the following:

- Review your distribution's behaviors to be sure that there's a cache behavior for your requested path pattern. If the requested path pattern isn't correctly defined in a cache behavior, CloudFront uses the default cache behavior. For example, if a cache behavior has Path Pattern set to test/, then a request to example.com/test/file1.jpg follows the default cache behavior. The request won't follow the behavior specified for test/ because the path pattern is missing a wildcard at the end of the path (test/*).
- Review the path pattern specified in the cache behavior to be sure that the capitalization is correct. Path patterns specified in CloudFront are case-sensitive. References to buckets or objects in Amazon Simple Storage Service (Amazon S3) are also case-sensitive. If a path pattern has incorrect capitalization and your default cache behavior doesn't cache the requested content, users might receive a 404 Not Found error.

Note: CloudFront doesn't generate 404 responses. If a requested object isn't found in a CloudFront cache, the request is sent to the origin and the origin generates the 404 response.

IAM Roles for Tasks

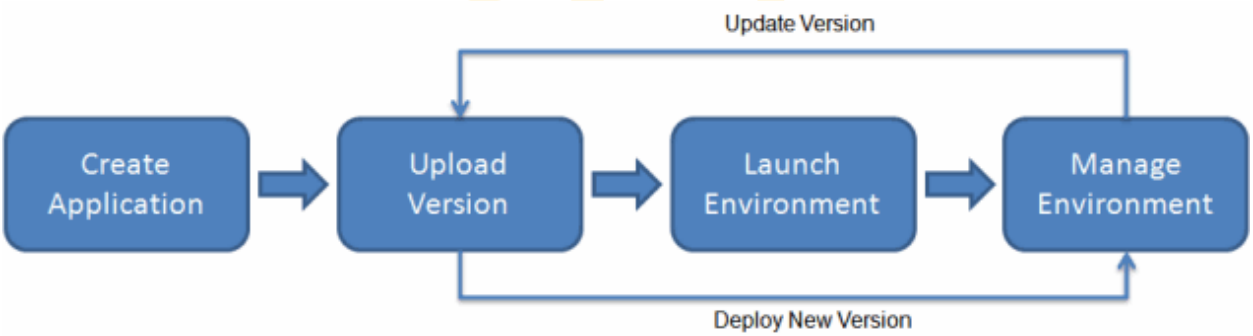
With IAM roles for Amazon ECS tasks, you can specify an IAM role that can be used by the containers in a task. Applications must sign their AWS API requests with AWS credentials, and this feature provides a strategy for managing credentials for your applications to use, similar to the way that Amazon EC2 instance profiles provide credentials to EC2 instances. Instead of creating and distributing your AWS credentials to the containers or using the EC2 instance's role, you can associate an IAM role with **an ECS task definition** or `RunTask` API operation. The applications in the task's containers can then use the AWS SDK or CLI to make API requests to authorized AWS services.

Standard Reserved Instances provide you with a significant discount (up to 75%) compared to On-Demand **instance** pricing, and can be purchased for a **1-year or 3-year term**. Customers have the flexibility to change the Availability Zone, the **instance** size, and networking type of their **Standard Reserved Instances**

What is end point in AWS?
A VPC **endpoint** enables you to create a private connection between your VPC and another **AWS** service without requiring access over the Internet, through a NAT device, a VPN connection, or **AWS** Direct Connect. **Endpoints** are virtual devices

The anomaly detection use case demonstrates how you can detect outliers in streaming data using **Amazon Kinesis**. As raw **vehicle sensor data streams** to Amazon S3 through an ***Amazon Kinesis Firehose*** delivery stream, an Amazon Kinesis Analytics application analyzes each record, extracts anomalies, and sends those records to an Amazon Kinesis Stream, triggering an AWS Lambda function that stores the data and notifies the driver of any issues.

Elastic Beanstalk



AWS Lambda Limits

AWS Lambda limits the amount of compute and storage resources that you can use to run and store functions. The following limits apply per-region and can be increased. To request an increase, use the Support Center console.

Resource	Default Limit
Concurrent executions	1,000
Function and layer storage	75 GB

For details on concurrency and how Lambda scales your function concurrency in response to traffic, see AWS Lambda Function Scaling.

The following limits apply to function configuration, deployments, and execution. They cannot be changed.

Resource	Limit
Function memory allocation	128 MB to 3,008 MB, in 64 MB increments.
Function timeout	900 seconds (15 minutes)
Function environment variables	4 KB
Function resource-based policy	20 KB
Function layers	5 layers
Function burst concurrency	500 - 3000 (varies per region)
Invocation frequency (requests per second)	10 x concurrent executions limit (synchronous – all sources) 10 x concurrent executions limit (asynchronous – non-AWS sources) Unlimited (asynchronous – AWS service sources)
Invocation payload (request and response)	6 MB (synchronous) 256 KB (asynchronous)
Deployment package size	50 MB (zipped, for direct upload) 250 MB (unzipped, including layers)

	3 MB (console editor)
Elastic network interfaces per VPC	160
Test events (console editor)	10
`/tmp` directory storage	**512 MB**
File descriptors	1,024
Execution processes/threads	1,024

What is Amazon redshift in AWS?

Amazon Redshift is a fully managed, petabyte-scale data warehouse service in the cloud. ... The first step to create a data warehouse is to launch a set of nodes, called an **Amazon Redshift** cluster. After you provision your cluster, you can upload your **data set** and then perform data analysis queries.

How is Amazon SQS different from Amazon SNS?

Amazon SNS allows applications to send time-critical messages to multiple subscribers through a “push” mechanism, eliminating the need to periodically check or “poll” for updates. Amazon SQS is a message queue service used by distributed applications to exchange messages through a polling model, and can be used to **decouple** sending and receiving components.

DynamoDB a NoSQL **DB which means you can change the schema easily**. It's also the only DB in the list that you can scale without any downtime Amazon Aurora, RDS MySQL and RedShift all require changing instance sizes in order to scale which causes an outage. They are also all relational databases (SQL) so changing the schema is difficult

Creating Resources for Amazon **EFS**. Amazon **EFS** provides elastic, shared file storage that is **POSIX-compliant**. The file system you create supports concurrent read and write access from multiple Amazon EC2 instances and is accessible from all of the Availability Zones in the AWS Region where it is created

Hashtag data : #amazon

Amazon Kinesis Data Analytics