Comparison of various storage services of AWS

Virtualization and Cloud Computing

APRIL 6, 2020

SUBMITTED BY: M. IFTIKHAR UDDIN KHAN SAMI

022-16-113275

*Submitted by: **M. Iftikhar Uddin Khan Sami (022-16-113275)***

Contents

Comparison of various storage services of AWS Assignment

Amazon Web Services (AWS) is one of the most competent cloud service providers in the world right now. Over the years, data storage has been diversified vastly to cater to varying needs. Ranging from the needs of a single person to a multinational company, data storage has become a must-have factor for everyone. Starting from 'Punch Cards' which are used to communicate information to equipment – even before computers evolved to 'Cloud Storage'; which is the most popular storage option currently available – data storage technologies have transformed and are still evolving day by day.

Among the hundreds of cloud service providers, Amazon Web Services (AWS) dominates the digital market and is a flexible, cost-effective, easy-to-use cloud computing platform.

Amazon Simple Storage Service (Amazon S3)

Amazon S3 is an object storage model that is built to store and retrieve any amount of data from any place such as websites, mobile apps, corporate applications, and data from IoT sensors or devices. Amazon S3 is the most supported storage platform available, with the largest ecosystem.

Usage –

In addition to object storing, Amazon S3 is particularly well suited for hosting web content that requires bandwidth along with high demand. S3 is also used to host entire static websites and storage for images, videos, and client-side scripts in formats such as JavaScript. You can easily move cold data (data that is not frequently accessed) to Amazon Glacier using lifecycle management rules on data stored in S3

Durability & availability –

Amazon S3 runs upon the world's largest global cloud infrastructure, and was built from the ground up to deliver a customer promise of 99.999999999% durability. Data is automatically distributed across a minimum of three physical facilities that are geographically separated within an AWS Region, and also automatically replicates data to any other AWS Region.

Security –

Amazon S3 is a highly secure storage service. S3 is the only cloud storage platform that supports three different forms of encryption, including server-side-encryption and client-side-encryption. You can manage access to Amazon S3 by granting other AWS accounts and users permissions to perform resource operations by writing an access policy.

Amazon Glacier

Amazon Glacier is a secure, durable, and extremely low-cost storage service for data archiving and long-term backup. Glacier provides 'query-in-place functionality', which allows you to run powerful analytics directly on archived data at rest. Glacier can make use of other AWS services such as S3, CloudFront etc. to move data in and out seamlessly for better and effective results.

Usage –

Amazon Glacier stores data in the form of archives. An archive can represent a single file, or you can combine several files to be uploaded as a single archive, and archives are organized in vaults. AWS Glacier is the only cloud archive storage service that allows you to query data in place and retrieve only the subset of data that you need from within an archive.

Durability & availability –

Since AWS Glacier is an archiving service, durability must be of utmost priority. Glacier is designed to provide average annual durability of 99.999999999% for archives. Data is automatically distributed across a minimum of three physical facilities that are geographically separated within an AWS Region.

Security –

By default, only the account owner can access Amazon Glacier data. If other people or services need to access the data, you can set up data access controls in AWS Glacier by using the AWS Identity and Access Management (IAM) service. Similarly, Glacier uses server-side encryption to encrypt all data at rest. Amazon Glacier allows you to lock vaults where long-term records retention is mandated, along with the use of lockable policies.

Amazon Elastic File System (Amazon EFS)

EFS delivers a simple, scalable, elastic, highly available, and highly durable network file system as-a-service to EC2 instances. Amazon EFS storage capacity is elastic and is capable of growing and shrinking automatically as you add and remove files without disrupting your EFS applications.

Usage –

EFS is designed to provide a highly scalable network file system that can grow to petabytes and allows massively parallel access from EC2 instances. EFS supports Network File System versions 4 (NFSv4) and 4.1 (NFSv4.1).

When mounted up on Amazon EC2 instances, the EFS file system provides a standard file system interface and file system access. Multiple Amazon EC2 instances can access an Amazon EFS file system (as a shared storage location). Thus, applications that scale beyond a single instance can access a file system.You can mount your EFS file systems on your on-premises datacenter servers when connected to your Amazon Virtual Private Cloud (VPC) with AWS Direct Connect service.

Durability & availability –

Each Amazon EFS file system object (such as a directory, file or link) is redundantly stored across multiple availability zones within a region. Amazon EFS is designed to be as highly durable and available as Amazon S3.

Security –

There are three main levels of access controls to consider when it comes to EFS file system.

1. IAM permissions for API calls
2. Security groups for EC2 instances and mount targets
3. Network File System-level users, groups, and permissions.

Amazon groups play a critical role in establishing network connectivity between EC2 instances and EFS file systems. You can associate one security group with an EC2 instance and another security group with an EFS mount target associated with the file system. These security groups act as firewalls and enforce rules that define the traffic flow between EC2 instances and EFS file systems.

Amazon Elastic Block Store (Amazon EBS)

EBS volumes provide durable block-level storage for use with EC2 instances in the AWS cloud. Volumes are automatically replicated within Availability Zones for high availability and durability.

Usage –

EBS volumes are network-attached storage that persists independently from the running life of a single EC2 instance. After an EBS volume is attached to an EC2 instance, you can use the EBS volume similar to a physical hard drive – typically by formatting it with the file system of your choice and using the file I/O interface provided by the instance operating system. Multiple EBS volumes can be attached to a single EC2 instance and it allows you to dynamically increase capacity, tune performance, and change the type of any new or existing current generation volume with no downtime or performance impact. Furthermore EBS provides the ability to save point-in-time snapshots of your volumes. Each separate volume can be configured as EBS General Purpose (SSD), Provisioned IOPS (SSD), Throughput Optimized (HDD), or Cold (HDD) as needed.

Durability & availability –

Amazon EBS cloud service is designed to be highly available and reliable. As mentioned earlier EBS volumes data is replicated across multiple servers within availability zones. Taking snapshots of your EBS volumes increases the durability of the data stored on your EBS volumes. Furthermore, EBS volumes are designed for an Annual Failure Rate (AFR) of between 0.1 and 0.2 percent, where failure refers to a complete or partial loss of the volume, depending on the size and performance of the volume.

Security –

IAM service enables access to EBS volumes, allowing you to specify who can access which EBS volumes. EBS encryption enables data-at-rest and data-in-motion security. It offers seamless encryption of both EBS boot volumes and data volumes as well as snapshots. Access control plus encryption offers a strong defense-in-depth security strategy for your data.

Amazon EC2 Instance Storage

EC2 Instance store provides temporary block-level storage for EC2 instances. This storage is located on disks that are physically attached to the host computer. Instance store is ideal for temporary storage of information that changes frequently such as buffers, caches and scratch data.

Usage –

In general, Instance storage volumes are ideal for temporary use as mentioned above. This can only be used from a single EC2 instance during that instance's lifetime. Unlike EBS volumes, instance store volumes cannot be detached or attached to another instance. Some instance types use NVMe or SATA-based solid state drives (SSD) to deliver high random I/O performance.

Durability & Availability –

EC2 instance store volumes are not intended to be used as durable disk storage. Applications using instance storage for persistent data generally provide data durability through replication, or by periodically copying data to durable storage. Data on instance store volumes persist only during the life of the associated EC2 instance. EC2 instance volumes are not recommended to be used for any data that must persist over time, such as permanent file or database storage.

Security –

IAM service helps to gain secure control over which users can perform operations such as launch and termination of EC2 instances in your account. When you stop or terminate an instance, the applications and data in its instance store are erased, and thus no other instance can have access to the instance store in the future.

AWS Storage Gateway

AWS Storage Gateway is a hybrid storage service that enables on-premises applications to implement seamless and secure storage integration with AWS cloud storage. AWS Storage Gateway supports industry standard storage protocols that work with existing applications as well.

Usage –

AWS Storage Gateway is recommended for backing up and archiving, disaster recovery, cloud bursting, storage tiering, and migration. Storage Gateway software can be downloaded as a Virtual Machine (VM) image that can be installed on a host in your data center or as an EC2 instance. Once you have installed the gateway and associated it with your AWS account through the AWS activation process, you can use the AWS Management Console to create gateway-cached volumes, gateway-stored volumes, or a gateway-virtual tape library (VTL).

To serve the above purposes, Gateway connects to AWS storage services, such as Amazon S3, Amazon Glacier, and Amazon EBS, providing storage for files, volumes, and virtual tapes in AWS.

Durability & availability –

Storage Gateway durably stores your on-premises application data by uploading it to Amazon S3 or Amazon Glacier. This service is designed to provide an average annual durability of 99.999999999 percent (yes, 11 nines).

Security –

Similar to most AWS storage services all data transferred through AWS Storage Gateway is secure and encrypted when at rest. IAM services provide security in controlling access to AWS Storage Gateway and when integrating with other storage services, access control can be used in combination with other services.

AWS Snowball

Snowball is a petabyte-scale data transport solution that uses a secure Snowball device to transfer large amounts of data in and out of the AWS cloud. The Snowball device is purpose-built for efficient data storage and transfer between Amazon and local premises.

Usage –

Snowball addresses common challenges with large-scale data transfers such as high network costs, long transfer times, and security concerns. All AWS regions have 80 TB Snowballs while US Regions have both 50 TB and 80 TB models.

With Snowball, you do not need to write any code or purchase any hardware to transfer your data. Simply create a job in the AWS Management Console and a Snowball device will be automatically shipped to the local premises. Once the device arrives, it must be attached to the premises local network. Download and run the Snowball client to establish a connection, and then use the client to select the file directories that you want to transfer to the device. The client will then encrypt and transfer the files to the snowball device at high speed.

Durability & availability –

Once the data is imported to AWS, the durability and availability characteristics of the target storage applies just like any other AWS services.

Security –

Snowball can be integrated with AWS Identity and Access Management (IAM) services to control access over which actions a user can perform. Similarly, an IAM user that creates a Snowball job must have permissions to access the Amazon S3 buckets that will be used for the import operations.

All data loaded onto a Snowball appliance is encrypted using 256-bit encryption. Snowball is physically secured using an industry-standard Trusted Platform Module (TPM) that uses a dedicated processor designed to detect any unauthorised modifications to the hardware, firmware, or software. In addition, Snowball is included in the AWS HIPAA compliance program so you can use Snowball to transfer large amounts of Protected Health Information (PHI) data in and out of AWS.

Amazon CloudFront

Amazon CloudFront is a global Content Delivery Network (CDN) service which securely delivers website's dynamic, static, and streaming content by making it available from a global network of edge locations. Amazon CloudFront supports all types of files that can be served over HTTP.

Usage –

CloudFront makes use of Edge locations to deliver content to end-users. Edge locations are located in most of the major cities around the world and are specifically used by AWS CloudFront (CDN) for distribution of content. CloudFront is seamlessly integrated with other AWS services including AWS Shield for DDoS mitigation, Amazon S3, Elastic Load Balancing or Amazon EC2 as origins for your applications, and AWS Lambda to run custom code close to your viewers.

When a user requests for content that is served with Amazon CloudFront, the user is routed to the edge location that provides the lowest latency (time delay). As a result that content is delivered with better performance than if the user had accessed the content from a data center farther away.

If the content is already in the edge location with the lowest latency, Amazon CloudFront delivers it immediately. If the content is not currently in that edge location, Amazon CloudFront retrieves it from an Amazon S3 bucket or an HTTP server that was identified as the source for the definitive version of your content.

Amazon CloudFront caches content at edge locations for a period of time specified by the provider.

Amazon CloudFront supports all files that can be served over HTTP. These files include dynamic web pages, such as HTML or PHP pages along with any popular static files that are part of your web application, such as website images, audio, video, media files or software downloads.

Durability & availability –

Since CloudFront is an edge cache, Amazon CloudFront does not provide durable storage. The origin server, such as Amazon S3 or a web server running on Amazon EC2, provides the durable file storage needed. But CloudFront provides high availability by using a distributed global network of edge locations. Amazon constantly monitors and optimizes the network paths which provide content for both availability and performance.

Security –

CloudFront is a highly secure CDN that provides both network and application level protection. CloudFront customers benefit from the automatic protection of AWS Shield (DDoS protection service) Standard, at no additional charge. CloudFront is also seamlessly integrated with AWS WAF (Web Application Firewall) and AWS Shield Advanced to help protect your applications from more sophisticated threats and DDoS attacks.