MURANG'A UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF IT

SIT407: CLOUD COMPUTING
CREDIT HOURS: 3 HOURS

COURSE NOTES

AMAZON WEB SERVICES (AWS)

Objectives

i) Explain Amazon web service (AWS)
ii) Elastic Block Storage
iii) Explain the meaning of S3 Storage

AMAZON WEB SERVICE (AWS)

Cloud storage is a web service where your data can be stored, accessed, and quickly backed up by users on the internet. It is more reliable, scalable, and secure than traditional on-premises storage systems.

Cloud storage is offered in two models:

1. Pay only for what you use
2. Pay on a monthly basis

Amazon Web Services is the gold standard or easily the best Cloud Service provider in the public cloud domain. It provides on-demand Cloud Computing services, that can be rented on metered usage and can be accessed across the globe by using the internet.

AMAZON WEB SERVICE (AWS) CONT'D

Amazon Web Services takes care of managing and monitoring resources so one as a consumer does not have to invest too much time in doing these activities.

It provides services in the following domains:

- Computation
- Storage
- Databases
- Security
- Networking
- Monitoring
- Messaging
- Migration
- Machine Learning
- DevOps
- IoT, etc

AMAZON WEB SERVICE (AWS) CONT'D

AWS serves in 245+ countries and offers 250+ services. It provides following features,

- Scalability
- Availability
- Metered Usage
- Flexibility
- Durability, etc

Cloud storage is nothing but an ability given to you as an individual to store your data on Cloud. Cloud storage lets you store data that can be,

- Files
- Pictures
- Processing Data
- Messages
- Logs
- Videos, etc

Cloud storage lets you store data at an affordable price and lets you take data backups ensuring your data is safe and secure on cloud. This data can be processed making sure you can put your data to use on top of Cloud.

AMAZON WEB SERVICE (AWS) CONT'D

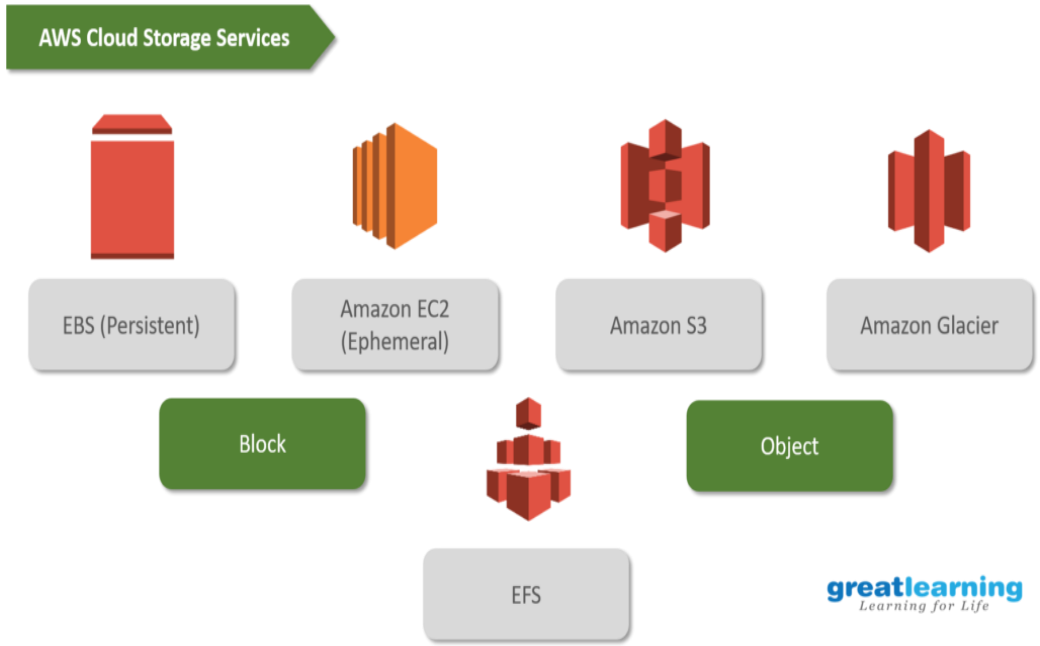
Amazon Web Services provides plenty of services that help you store data safely in secure manner on AWS Cloud Platform.
It provides following storage services,

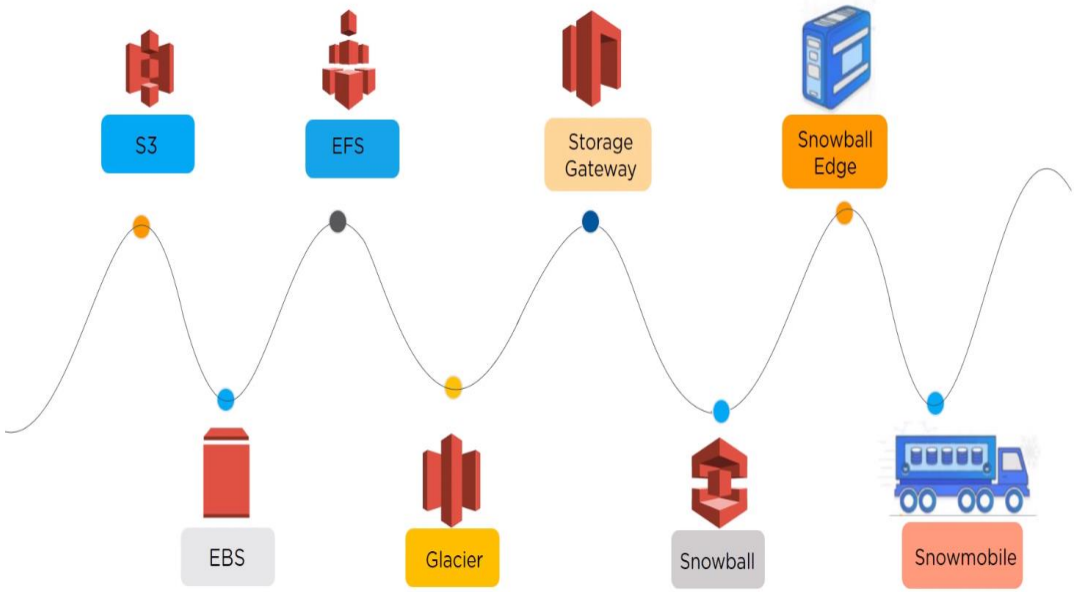
- Amazon EBS
- Amazon S3
- Amazon EFS
- Amazon Glacier
- Amazon Storage Gateway
- Amazon Snowmobile
- Amazon Snowball etc,

Amazon mostly classifies it storage in Block, Object and file kind of storage. Following image shows those kinds of storages

AMAZON WEB SERVICE (AWS) CONT'D

Following image shows those kinds of storages





Before AWS S3

Organizations had a difficult time finding, storing, and managing all their data

AMAZON WEB SERVICE (AWS) CONT'D

Not only that, running applications, delivering content to customers, hosting high traffic websites, or backing up emails and other files required a lot of storage. Maintaining the organization's repository was also expensive and time-consuming for several reasons.

Challenges included the following:

1. Having to purchase hardware and software components
2. Requiring a team of experts for maintenance
3. A lack of scalability based on your requirements
4. Data security requirements

These are the issues AWS S3 would eventually solve.

AMAZON WEB SERVICE (AWS) CONT'D

EBS

Elastic Block Store is a storage that comes in block form. In order to use storage. You have to attach it to a host virtual machine. It is similar to a hard disk drive where the storage can only be used when it is attached to a system or a laptop.

Amazon Elastic Block Store (Amazon EBS) provides block level storage volumes for use with EC2 instances. EBS volumes behave like raw, unformatted block devices. You can mount these volumes as devices on your instances

This is persistent storage that means if the system goes down the storage will still be available and can be used later when attached to a machine.

AMAZON WEB SERVICE (AWS) CONT'D

Amazon EC2 Ephemeral Storage

This is also block store but this volume or storage goes away with the instance. That means if the instance is terminated the storage also gets deleted by default.

Amazon Glacier

This is a type of archival storage that means we store that data which is not used frequently. Say for example data files that very rarely accessed. For example, hospital record, like birth certificates. This storage is hence called as cold storage. Since this data is not frequently used. We store it in cold storage something we refrain from accessing every now and then. It is cheaper and hence affordable.

AMAZON WEB SERVICE (AWS) CONT'D

EFS

EFS stands for Elastic File System and as the name suggests is used for Storing file data or managing file systems.

Amazon Snowball

This storage is used for moving data physically. Think of it as more of a Pendrive or Hard disk drive that can be used to move your data physically from one data centre to another.

What if the amount of data to be moved is large in size? In that case, we can use Amazon Snowmobile. Amazon Snowmobile lets us move heaps of data from one data centre to another physically.

S3 STORAGE

Amazon S3 (Simple Storage Service) provides object storage, which is built for storing and recovering any amount of information or data from anywhere over the internet.

It provides this storage through a web services interface. While designed for developers for easier web-scale computing, it provides 99.999999999 percent durability and 99.99 percent availability of objects. It can also store computer files up to 5 terabytes in size.

Amazon S3 support object storage and is hot storage in nature.

Let us understand Cold and hot storage first

Cold Storage

This is kind storage as the name suggests is cold or slow in nature. That means if you store your data here, it will take 4-48 hours for you to retrieve your data at least.

S3 STORAGE CONT'D

This data storage is affordable or cheaper. It is used by applications where data retrieval is not critical and data latency won't impact the business in general.

Hot Storage

Hot storage is as you might have guessed a storage that lets you retrieve data faster and is reliable in terms of processing data retrieval latency free.

Amazon S3 is hot cloud storage and is hence costlier because you get faster data retrieval.

S3 STORAGE CONT'D

Components of S3 storage service

Amazon S3 Buckets

In order to store your data in Amazon S3, you would need a container that can hold this data. These containers with respect to Amazon S3 are known as Buckets

Amazon S3 Objects

Amazon S3 objects are nothing but files that we store in the Amazon S3 bucket. Please not Amazon S3 is a key value store. You can store as many object in these buckets. These object can be as big as 5tb in size.

Key: The name you assign to the object

Version: It is the version Id of a specific version of a file. The version helps uniquely identify a particular object.

S3 STORAGE CONT'D

Value and Metadata: Value is nothing but a concept we are trying to store. Whereas Metadata is the information about the data we are trying to store

With S3 we can create multiple buckets in Different AWS regions and store or transfer our data there.

An object consists of data, key (assigned name), and metadata. A bucket is used to store objects. When data is added to a bucket, Amazon S3 creates a unique version ID and allocates it to the object.

Example of an object, bucket, and link address

S3 STORAGE CONT'D

How Does Amazon S3 work?

A user creates a bucket. When this bucket is created, the user will specify the region in which the bucket is deployed. Later, when files are uploaded to the bucket, the user will determine the type of S3 storage class to be used for those specific objects. After this, users can define features to the bucket, such as bucket policy, lifecycle policies, versioning control, etc.

Amazon S3 Storage Classes

Different storage classes using the example of a school:

1. Amazon S3 Standard for frequent data access: Suitable for a use case where the latency should below. Example: Frequently accessed data will be the data of students' attendance, which should be retrieved quickly.

S3 STORAGE CONT'D

2.Amazon S3 Standard for infrequent data access: Can be used where the data is long-lived and less frequently accessed. Example: Students' academic records will not be needed daily, but if they have any requirement, their details should be retrieved quickly.

3.Amazon Glacier: Can be used where the data has to be archived, and high performance is not required. Example: Ex-student's old record (like admission fee) will not be needed daily, and even if it is necessary, low latency is not required.

4.One Zone-IA Storage Class: It can be used where the data is infrequently accessed and stored in a single region. Example: Student's report card is not used daily and stored in a single availability region (i.e., school).

5.Amazon S3 Standard Reduced Redundancy storage: Suitable for a use case where the data is non-critical and reproduced quickly. Example: Books in the library are non-critical data and can be replaced if lost.

S3 STORAGE CONT'D

A comparison of all storage classes

Amazon S3 Standard for frequent data access	Amazon S3 Standard for infrequent data access	Amazon Glacier	One Zone-IA Storage Class	Amazon S3 Standard Reduced Redundancy storage
• For frequently accessed data • It is a default storage class • Can be used for cloud applications, dynamic websites, content distribution, gaming applications, and Big data analytics	• For infrequently accessed data • Demands rapid access • Suitable for backups, disaster recovery and lifelong storage of data	• Suitable for archiving data where data access is infrequent • Vault-lock feature provides a long term data storage • Provides the lowest cost availability	• Suitable for infrequently accessed data • Unlike other classes, this new storage class stores the data in a single AWS Availability Zone • Data that doesn't require any high level of security can be stored here	• For frequently accessed data • Stores reproducible and non crucial data at lower cost • A highly available solution designed for sharing or storing data that can be reproduced quickly

S3 STORAGE CONT'D

Technical comparison between classes

Storage Class	Durability	Availability	SSL support	First byte latency	Lifecycle Management Policies
STANDARD	99.999999999%	99.99%	Yes	Milliseconds	Yes
STANDARD_IA	99.999999999%	99.99%	Yes	Milliseconds	Yes
ONEZONE_IA	99.999999999%	99.5%	Yes	Milliseconds	Yes
GLACIER	99.999999999%	99.99%	Yes	Minutes or Hours	Yes
RRS	99.99%	99.99%	Yes	Milliseconds	Yes

S3 STORAGE CONT'D

AWS S3 Features

Lifecycle Management

In lifecycle management, Amazon S3 applies a set of rules that define the action to a group of objects. You can manage and store objects in a cost-effective manner. There are two types of actions:

1. Transition Action

With this action, you can choose to move objects to another storage class. With this, you can configure S3 to move your data between various storage classes on a defined schedule. Assume you've got some data stored in the S3 standard class. If this data is not used frequently for 30 days, it would be moved to the S3 infrequent access class. And after 60 days, it is moved to Glacier. This helps you to migrate your data to lower-cost storage as it ages automatically.

S3 STORAGE CONT'D

2. Expiration Actions

Here, S3 removes all objects within the bucket when a specified date or time period in the object's lifetime is reached.

Bucket Policy

Bucket policy is an IAM policy where you can allow or deny permission to your Amazon S3 resources. With bucket policy, you also define security rules that apply to more than one file within a bucket.

For example: If you do not want a user to access the "Simplilearn" bucket, then with the help of JSON script, you can set permissions. As a result, a user would be denied access to the bucket.

S3 STORAGE CONT'D

3. Data Protection

Amazon S3 provides IT teams with a highly durable, protected, and scalable infrastructure designed for object storage.
Amazon S3 protects your data using two methods:

i) Data encryption
ii) Versioning
iii) Cross-region Replication
iv) Transfer Acceleration

i) Data Encryption

This refers to the protection of data while it's being transmitted and at rest. It can happen in two ways, client-side encryption (data encryption at rest) and server-side encryption (data encryption in motion).

S3 STORAGE CONT'D

ii) Versioning

It is utilized to preserve, recover, and restore an early version of every object you store in your AWS S3 bucket. Unintentional erases or overwriting of objects can easily be managed with versioning. For example, in a bucket, it is possible to have objects with the same key name but different version IDs.

iii) Cross-region Replication

Cross-region replication provides automatic copying of every object uploaded to your buckets (source and destination bucket) in different AWS regions. Versioning needs to be turned on to enable CRR

iv) Transfer Acceleration

This enables fast, easy, and secure transfers of files over long distances between your client and S3 bucket. The edge locations around the world provided by Amazon CloudFront are taken advantage of by transfer acceleration. It works by carrying data over an optimized network bridge that keeps running between the AWS Edge Location (closest region to your clients) and your Amazon S3 bucket.

ELASTIC BLOCK STORAGE

Amazon Elastic Block Store (EBS) is user-friendly block storage service that runs with very high performance and used with Amazon Elastic Compute Cloud (EC2) for both throughput and intensive transactions. A broad range of workloads, such as relational and non-relational databases, enterprise applications, containerized applications, big data analytics engines, files systems, and media workflows can be deployed on Amazon EBS.

The options are divided into two major categories:

Transactional workloads, such as databases and boot volumes (performance depends primarily on IOPS) have SSD-backed storage. Throughput intensive workloads, such as MapReduce and log processing (performance depends primarily on MB/s) have disk-backed storage.

ELASTIC BLOCK STORAGE CONT'D

Block level storage volumes for use with EC2 instances is provided by EBS. EBS volumes are like raw, unformatted block devices.

- Multiple volumes can be mounted on the same instance, but each volume can be mounted to only one instance.
- File system can be created on top of these volumes or use them in any way you would use a block device (like a hard drive). Dynamically changes can be made to the configuration of a volume attached to an instance.
- EBS volumes are termed as highly available and reliable storage volumes that can be attached to any running instance that is in the same Availability Zone. EBS volumes attached to an EC2 instance are exposed as storage volumes that independently persist from the life of the instance. In Amazon EBS, you pay only for what you use.

ELASTIC BLOCK STORAGE CONT'D

Multiple volumes to the same instance can be attached within the limits specified by your AWS account. Your account has a limit on the number of EBS volumes that you can use, and the total storage available to you.

Amazon EBS is recommended when data must be quickly accessible and requires long-term persistence. EBS volumes are particularly well-suited for use as the primary storage for file systems, databases, or for any applications that require fine granular updates and access to raw, unformatted, block-level storage.

Amazon EBS is well suited to both database-style applications that rely on random reads and writes, and to throughput-intensive applications that perform long, continuous reads and writes.

ELASTIC BLOCK STORAGE CONT'D

Benefits Of EBS

Performance for any workload

Most demanding workloads, including mission-critical applications such as SAP, Oracle, and Microsoft products are ideal case scenarios for EBS volumes. Volumes designed for high performance applications and a general-purpose volume that offers strong price/performance for most workloads are included in SSD-backed options. Volumes designed for large, sequential workloads such as big data analytics engines, log processing, and data warehousing are included in HDD-backed volumes. Multiple volumes together can be used for higher storage performance per instance.

ELASTIC BLOCK STORAGE CONT'D

Easy to Use

Easy to create, use, encrypt, and protect are the features of Amazon EBS volumes. It allows increasing storage, tune performance up and down, and change volume types without any disruption to your workloads.

EBS Snapshots allow you to easily take backups of your volumes for geographic protection of your data.

Data Lifecycle Manager (DLM) is an easy-to-use tool for automating snapshot management without any additional overhead or cost.

ELASTIC BLOCK STORAGE CONT'D

Highly Available and Durable

Reliability for mission-critical applications is offered by Amazon EBS architecture. Volume are designed to protect against failures by replicating within the Availability Zone (AZ), offering 99.999% availability and an annual failure rate (AFR) of between 0.1%-0.2%. For simple and robust backup, use EBS Snapshots with Amazon Data Lifecycle Manager (DLM) policies to automate snapshot management. To avoid disruption to your critical workloads Amazon EBS enable you to increase storage. You can build applications that require as little as a single GB of storage, or scale up to petabytes of data in just a few clicks.

ELASTIC BLOCK STORAGE CONT'D

Snapshots can be used to quickly restore new volumes across a region's Availability Zones, enabling rapid scale.

Secure

It is built to be secure for data compliance. New EBS volumes can be encrypted by default with a single setting in your account. Volumes support encryption of data at-rest, data in-transit, and all volume backups. Encryption is supported by all volume types, includes built-in key management infrastructure, and has zero impact on performance.

ELASTIC BLOCK STORAGE CONT'D

Cost-effective

Four different volume types are offered by EBS at various price points and performance benchmarks. It enable you to optimize costs and invest in a precise level of storage for your application needs. Highly cost-effective dollar per gigabyte volumes to high performance volumes with high IOPS and high throughput designed for mission critical workloads are the option to choose from. EBS Snapshots are incremental and save on storage costs by not duplicating data.

ELASTIC BLOCK STORAGE CONT'D

Features of Amazon EBS

Specific Availability Zone can be used for EBS volumes, and then can be attached to any instances in that same Availability Zone. Volume can be made available outside of the Availability Zone, you can create a snapshot and restore that snapshot to a new volume anywhere in that Region. Snapshots can be copied to other Regions and then can be restored to new volumes making it easier to leverage multiple AWS Regions for geographical expansion, data center migration, and disaster recovery.

AMAZON NETWORK SECURITY TOOLS

Most businesses collect, process, and store sensitive customer data that needs to be secured to earn customer trust and protect customers against abuses. Regulated businesses must prove they meet guidelines established by regulatory bodies.

There several AWS security tools including:
AWS Identity and Access Management (IAM)
AWS IAM is essential for controlling access to AWS resources.

AMAZON NETWORK SECURITY TOOLS CONT'D

- Amazon GuardDuty
- Amazon Macie
- AWS Config
- AWS CloudTrail
- Security Hub
- Amazon Inspector
- AWS Shield

AMAZON NETWORK SECURITY TOOLS CONT'D

AWS Shield

AWS Shield is a managed DDoS protection service. Shield can protect EC2, Load balancers, CloudFront, Global Accelerator, and Route 53 resources.

While DDoS protection may not seem revolutionary, consider that Amazon claims that 99 percent of all infrastructure flood attacks detected by shield are mitigated in less than one second on CloudFront.

Sometimes attacks are simply designed to prevent a company from doing business.

Having a tool that allows you to stay up, without engaging your security teams, can be a significant competitive advantage.

AWS shield can even protect websites that are not hosted inside AWS.

AMAZON NETWORK SECURITY TOOLS CONT'D

GuardDuty

GuardDuty is the "watcher on the wall". GuardDuty is a managed threat detection service that is simple to deploy, and scales with your infrastructure.

It will analyze logs across all of your accounts and services, making sure that nothing is left unprotected. Amazon boasts that GuardDuty analyzes tens of billions of events across AWS — and leverages machine learning to ensure you get accurate and actionable alerts.

There are very few other companies that can boast that kind of data set.

AMAZON NETWORK SECURITY TOOLS CONT'D

GuardDuty is capable of detecting activities related to reconnaissance, instance compromise and account compromise. This encompasses things like, port scanning, data exfiltration, malware, unusual API calls, and attempts at disabling logging.

CloudWatch

CloudWatch is the AWS monitoring tool for, well, everything. CloudWatch ingests logs, events, and metrics across your AWS infrastructure to ensure you have visibility into everything going on in your environment.

AMAZON NETWORK SECURITY TOOLS CONT'D

Having a tool that can aggregate a ton of data and make it accessible to engineers is crucial.

Because CloudWatch integrates with GuardDuty, and can provide a huge amount of surrounding information, it can also make it easier to troubleshoot security incidents.

Aside from its security applications, CloudWatch also aggregates performance and resource utilization data.

It can be used to set up auto scaling for EC2 instances to automatically add or remove compute resources to make sure organizations get the best value out of their spend for AWS services.

The bottom line: CloudWatch provides a single pane of class for visibility into log events and other security services.

AMAZON NETWORK SECURITY TOOLS CONT'D

Macie

Macie is all about protecting data. It is a machine learning service that watches data access trends and finds anomalies to spot data leaks and unauthorized data access.

It can send all of its alerts to Cloudwatch to leverage all of the automation and custom alerting.

It is a fully managed service. It's always nice to be able to add additional visibility and alerting without any additional work. It currently only supports monitoring S3 buckets.

It seems like it is a simple service, but quickly identifying unusual data access or data exfiltration can be incredibly important to containing breaches.

AMAZON NETWORK SECURITY TOOLS CONT'D

In 2017, Uber reported that it had a breach that affected the personal information of 57 million of its users. The breach was not a result of a misconfiguration or a failure of its AWS security, but a hacker accessing a private GitHub repo that contained its AWS credentials.

Uber paid the hackers $100,000 to keep the breach quiet until Uber itself ultimately revealed it to the public.

It's unknown whether the attackers approached Uber or Uber detected the attack themselves, but this is an effective illustration of Macie's value proposal.

The bottom line: Macie lets you know if your data is compromised.

AMAZON NETWORK SECURITY TOOLS CONT'D

AWS Inspector

It is always nice to be proactive. AWS inspector is a security assessment service that does vulnerability and best-practice scanning for AWS applications. The best part about AWS Inspector is that administrators get consistent improvements, as the AWS security team consistently updates best practices. Building security compliance and standards into infrastructure and application deployment gives organizations a massive head start to staying secure.

The bottom line: AWS inspector is always up to date.

AMAZON NETWORK SECURITY TOOLS CONT'D

Compliance and Configuration Scanners

Because AWS is a haven of DevOps engineers, it's no surprise that some of the best security tools are third party tools. Scoutsuite and Prowler are two of the best compliance and configuration scanners that have been developed by the open source community.

Prowler

Prowler describes itself as an AWS Security best practices assessment, auditing, hardening, and forensics readiness tool. It has 89 pages that spans configuration areas like identity management and networking, as well as configurations related to GDPR and HIPAA.

The bottom line: Prowler features extensive documentation.

AMAZON NETWORK SECURITY TOOLS CONT'D

Scoutsuite

Scoutsuite is also an auditing tool. The major differentiator between these tools is that Scoutsuite is multi-platform. It supports AWS, Microsoft Azure, and Google Cloud Platform.

While auditing tools may not be as exciting as some of the other tools on the list, the importance of them cannot be overstated. Some of the worst data breaches on AWS have been a result of simple misconfigurations. Allowing public read/write access to AWS S3 buckets have been responsible for data breaches of epic scale.

In 2017, Accenture, a corporate consulting firm mistakenly left four S3 buckets publically available. A security researcher discovered the buckets and alerted the company. In a display of just how easy this would have been to prevent, the buckets were secured the next day.

AMAZON NETWORK SECURITY TOOLS CONT'D

It's reported that there was 137Gb of data on the buckets, including plaintext client passwords, credentials for AWS, and other cloud platforms, decryption keys, certificates etc. If a malicious attacker had accessed the data, the damage they could have done to Accenture and its clients could have been catastrophic.

Again in 2017, a third party partner of Verizon, NICE systems, left an S3 bucket publicly available that contained names, addresses, account details, and PINS of upward of 14 million Verizon customers.

The scale of these breaches illustrates how important an auditing tool could be to keeping your data safe.

The bottom line: Start with a solid security foundation.

AMAZON NETWORK SECURITY TOOLS CONT'D

Security at Scale

AWS is all about scale — being able to grow quickly has never been easier. Many organizations host their entire application in AWS, from web front end, backend databases, compute resources, and massive amounts of data. The ease of this scaling can also mean that it is easy to build large, poorly configured and insecure deployments quickly.

Following the AWS published best practices and taking advantage of the available security services should allow companies to both grow quickly and securely.

The end.

Thank you.