

Notes

Customer Examples for Macie Custom Data Identifiers

@<https://aws.amazon.com/blogs/security/discover-sensitive-data-by-using-custom-data-identifiers-with-amazon-macie/>

Macie FAQ

@https://w.amazon.com/bin/view/Amazon_Macie_Frequently_Asked_Questions

Macie Configuring Replication

@<https://docs.aws.amazon.com/AmazonS3/latest/dev/replication-walkthrough-2.html>

What is a Data Lake?

<https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/>

Services in AWS that creates Data Lake? separate permissioning

DL allows you to have unlimited data warehouse. easier for you to streamline where the data is going.

Data Sync: Lake is data goes to / Faucet is sources of data, manual processing batch uploads from. Could be chat history.. random kinds of data.

Upload to S3

Have that kick off a Macie Job. Triggered to run. *How to trigger services from S3.

Data is scanned and flagged.

Triggers an event bridge to be written to DynamoDB

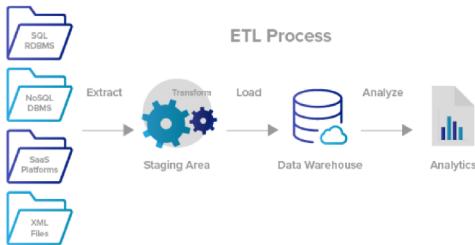
Encrypt from reading it in S3, then encrypting it client side in DynamoDB

ETL vs. ELT

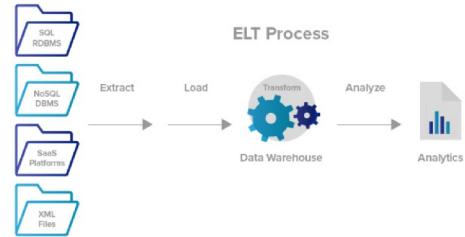
The five critical differences of ETL vs ELT:

1. ETL is the Extract, Transform, and Load process for data. ELT is Extract, Load, and Transform process for data.
2. In ETL, data moves from the data source to staging into the data warehouse.
3. ELT leverages the data warehouse to do basic transformations. There is no need for data staging.
4. ETL can help with data privacy and compliance by cleaning sensitive and secure data even before loading into the data warehouse.
5. ETL can perform sophisticated data transformations and can be more cost-effective than ELT.

ETL Process in Detail



ELT Process in Detail



Data transformation is the process where you extract data, sift through data, understand the data, and then transform it into something you can analyze. ↗

write Lambda to DynamoDB

<https://sysadmins.co.za/interfacing-amazon-dynamodb-with-python-using-boto3/>

AD examples

<https://creately.com/blog/diagrams/aws-templates-for-architecture-diagrams/>

Event Patterns

<https://docs.aws.amazon.com/eventbridge/latest/userguide/filtering-examples-structure.html>

Macie UserGuide

<https://docs.aws.amazon.com/macie/latest/user/macie-user-guide.pdf#%5B%7B%22num%22%3A553%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2C72%2C503.447%2Cnull%5D>

acloud.guru

so [accloud.guru](#)

log in with your @amazon email, and your Midway token (I believe)

and then take the aws certified Solutions architect course

it's very good

and has hands-on labs you can do at your own pace

IAM 101

IAM allows you to manage users and their level of access to the AWS Console.

Identity Access Management (IAM) offers the following features;

- Centralised control of your AWS account
- Shared Access to your AWS account
- Granular Permissions
- Identity Federation (including Active Directory, Facebook, LinkedIn etc)
- Multifactor Authentication
- Provide temporary access for users/devices and services where necessary
- Allows you to set up your own password rotation policy
- Integrates with many different AWS services
- Supports PCI DSS Compliance



1

Users

End Users such as people, employees of an organization, etc.

2

Groups

A collection of users. Each user in the group will inherit the permissions of the group.

3

Policies

Policies are made up of documents, called Policy documents. These documents are in a format called JSON and they give permissions as to what a User/Group/Role is able to do.

4

Roles

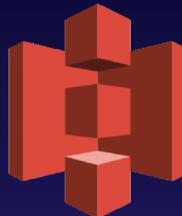
You create roles and then assign them to AWS Resources.

What Have We Learnt So Far?

- **IAM is universal.** It does not apply to regions at this time.
- The “**root account**” is simply the account created when first setup your AWS account. It has complete Admin access.
- New Users have **NO permissions** when first created.
- New Users are assigned **Access Key ID & Secret Access Keys** when first created.
- **These are not the same as a password.** You cannot use the Access key ID & Secret Access Key to Login in to the console. You can use this to access AWS via the APIs and Command Line, however.
- **You only get to view these once.** If you lose them, you have to regenerate them. So, save them in a secure location.

S3 101

S3 provides developers and IT teams with secure, durable, highly-scalable object storage. Amazon S3 is easy to use, with a simple web services interface to store and retrieve any amount of data from anywhere on the web.



The basics of S3 are as follows;

- **S3 is a universal namespace.** That is, names must be unique globally.
- <https://aclougdguru.s3.amazonaws.com/>
- <https://aclougdguru.eu-west-1.amazonaws.com/>
- When you upload a file to S3, you will receive a
- **HTTP 200 code** if the upload was successful.

S3 is Object based. **Think of Objects just as files.**

Objects consist of the following:

- Key (This is simply the name of the object)
- Value (This is simply the data and is made up of a sequence of bytes).
- Version ID (Important for versioning)
- Metadata (Data about data you are storing)
- Subresources;

Access Control Lists
Torrent



5

S3 Glacier

S3 Glacier is a secure, durable, and low-cost storage class for data archiving. You can reliably store any amount of data at costs that are competitive with or cheaper than on-premises solutions. Retrieval times configurable from minutes to hours.

6

S3 Glacier Deep Archive

S3 Glacier Deep Archive is Amazon S3's lowest-cost storage class where a retrieval time of 12 hours is acceptable.

- Not suitable to install an operating system on.
- Successful uploads will generate a **HTTP 200** status code.
- You can turn on **MFA Delete**

1

S3 Standard

99.99% availability
99.999999999% durability,
stored redundantly across
multiple devices in multiple
facilities, and is designed to
sustain the loss of 2 facilities
concurrently.

2

S3 - IA

(Infrequently Accessed):
For data that is accessed
less frequently, but requires
rapid access when needed.
Lower fee than S3, but you
are charged a retrieval fee.

3

S3 One Zone - IA

For where you want a
lower-cost option for
infrequently accessed data,
but do not require the
multiple Availability Zone
data resilience.

4

S3 - Intelligent Tiering

Designed to optimize costs by
automatically moving data to the
most cost-effective access tier,
without performance impact or
operational overhead.

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier	S3 Glacier Deep Archive**
Designed for durability	99.999999999% (11.9's)	99.999999999% (11.9's)	99.999999999% (11.9's)	99.999999999% (11.9's)	99.999999999% (11.9's)	99.999999999% (11.9's)
Designed for availability	99.99%	99.9%	99.9%	99.5%	N/A	N/A
Availability SLA	99.9%	99%	99%	99%	N/A	N/A
Availability Zones	≥3	≥3	≥3	1	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128KB	128KB	40KB	40KB
Minimum storage duration charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval fee	N/A	N/A	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds	select minutes or hours	select hours

- Remember that S3 is **Object-based**: i.e. allows you to upload files.
- Files can be from 0 Bytes to 5 TB.
- There is unlimited storage.
- Files are stored in Buckets.
- **S3 is a universal namespace**. That is, names must be unique globally.
- <https://aclouddguru.s3.amazonaws.com/>
- <https://aclouddguru.eu-west-1.amazonaws.com/>

The Key Fundamentals of S3 Are;

- Key (This is simply the name of the object)
- Value (This is simply the data and is made up of a sequence of bytes).
- Version ID (Important for versioning)
- Metadata (Data about data you are storing)
- Subresources;

Access Control Lists

Torrent

S3 Security and Encryption

By default, all newly created buckets are **PRIVATE**. You can setup access control to your buckets using;

- **Bucket Policies**
- **Access Control Lists**

S3 buckets can be configured to create access logs which log all requests made to the S3 bucket. This can be sent to another bucket and even another bucket in another account.



Encryption In Transit is achieved by

- **SSL/TLS**

Encryption At Rest (Server Side) is achieved by

- **S3 Managed Keys - SSE-S3**
- **AWS Key Management Service, Managed Keys - SSE-KMS**
- **Server Side Encryption With Customer Provided Keys - SSE-C**

Client Side Encryption



S3 Performance

S3 LIMITATIONS WHEN USING KMS

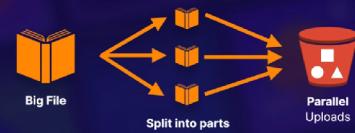
- If you are using **SSE-KMS** to encrypt your objects in S3, you must keep in mind the **KMS limits**.
- When you **upload** a file, you will call **GenerateDataKey** in the KMS API.
- When you **download** a file, you will call **Decrypt** in the KMS API.

S3 Limitations When Using KMS

- ✓ Uploading/downloading will count toward the **KMS quota**.
- ✓ Region-specific, however, it's either **5,500, 10,000, or 30,000 requests per second**.
- ✓ Currently, you **cannot** request a quota increase for KMS.

Multipart Uploads

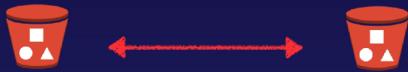
- Recommended for files **over 100 MB**
- Required for files **over 5 GB**
- Parallelize uploads (increases **efficiency**)



Sharing S3 Buckets Across Accounts

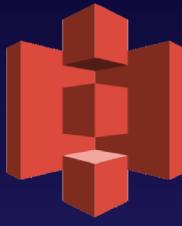
3 different ways to share S3 buckets across accounts

- Using Bucket Policies & IAM (applies across the entire bucket). Programmatic Access Only.
- Using Bucket ACLs & IAM (individual objects). Programmatic Access Only.
- Cross-account IAM Roles. Programmatic AND Console access.



S3 Transfer Acceleration

S3 Transfer Acceleration utilises the CloudFront Edge Network to accelerate your uploads to S3. Instead of uploading directly to your S3 bucket, you can use a distinct URL to upload directly to an edge location which will then transfer that file to S3. You will get a distinct URL to upload to:



Cloudfront

A content delivery network (CDN) is a system of distributed servers (network) that deliver webpages and other web content to a user based on the geographic locations of the user, the origin of the webpage, and a content delivery server.



Amazon CloudFront can be used to deliver your entire website, including dynamic, static, streaming, and interactive content using a global network of edge locations. Requests for your content are automatically routed to the nearest edge location, so content is delivered with the best possible performance.



- Edge locations are not just READ only — you can write to them too. (ie put an object on them.)
- Objects are cached for the life of the **TTL (Time To Live.)**
- You can clear cached objects, but you will be charged.

- **Edge Location** - This is the location where content will be cached. This is separate to an AWS Region/AZ.
- **Origin** - This is the origin of all the files that the CDN will distribute. This can be an S3 Bucket, an EC2 Instance, an Elastic Load Balancer, or Route53.
- **Distribution** - This is the name given the CDN which consists of a collection of Edge Locations.



- Web Distribution - Typically used for Websites.
- RTMP - Used for Media Streaming.

	Edge Location This is the location where content will be cached. This is different than an AWS Region/AZ .		Web Distribution Typically used for websites.
	Distribution This is the name given the CDN , which consists of a collection of edge locations.		RTMP Used for media streaming.
	Origin This is the origin of all the files the CDN will distribute. This can be either an S3 bucket, an EC2 instance, an Elastic Load Balancer, or Route 53.		

Create a CloudFront Signed URLs and Cookies

Use CloudFront Signed URLs or Signed Cookies

- 1 A signed URL is for individual files.
1 file = 1 URL.

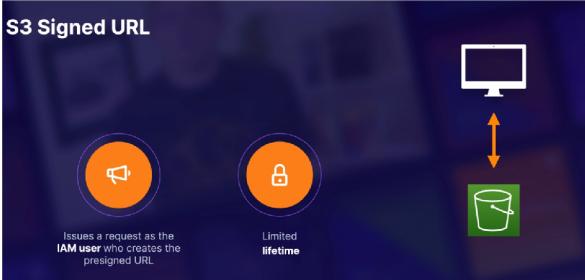
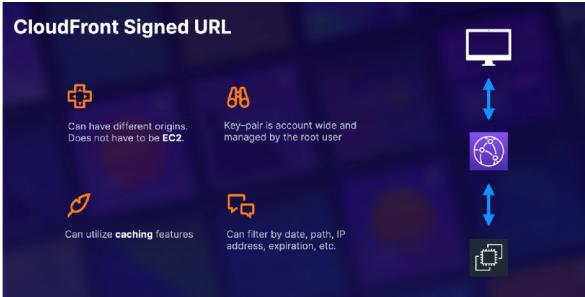
- 2 A signed cookie is for multiple files.
1 cookie = multiple files.

When we create a signed URL or signed cookie, we attach a policy.

The policy can include:

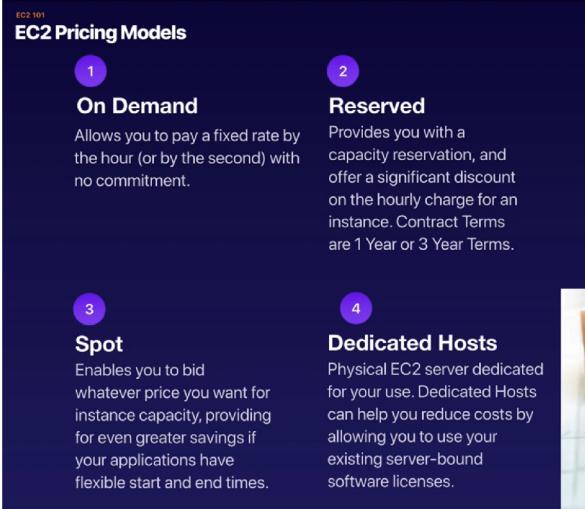
- URL expiration
- IP ranges
- Trusted signers (which AWS accounts can create **signed URLs**)

- ✓ Use signed URLs/cookies when you want to secure content so that only the people you authorize are able to access it.
- ✓ A signed URL is for individual files. **1 file = 1 URL**.
- ✓ A signed cookie is for multiple files. **1 cookie = multiple files**.
- ✓ If your origin is EC2, then use CloudFront.



EC2 101 Elastic Compute Cloud

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change.



On Demand pricing is useful for:

- Users that want the low cost and flexibility of Amazon EC2 without any up-front payment or long-term commitment
- Applications with short term, spiky, or unpredictable workloads that cannot be interrupted
- Applications being developed or tested on Amazon EC2 for the first time

Lambda

Lambda Is The Ultimate Abstraction Layer;

- Data Centres
- Hardware
- Assembly Code/Protocols
- High Level Languages
- Operating Systems
- Application Layer/AWS APIs
- AWS Lambda

AWS Lambda is a compute service where you can upload your code and create a Lambda function. AWS Lambda takes care of provisioning and managing the servers that you use to run the code. You don't have to worry about operating systems, patching, scaling, etc.

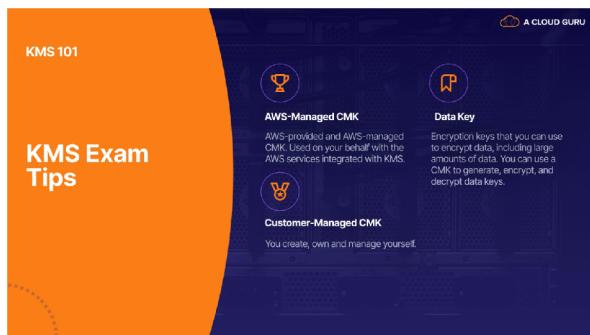
You can use Lambda in the following ways;

- As an event-driven compute service where AWS Lambda runs your code in response to events. These events could be changes to data in an Amazon S3 bucket or an Amazon DynamoDB table.
- As a compute service to run your code in response to HTTP requests using Amazon API Gateway or API calls made using AWS SDKs. This is what we use at A Cloud Guru.

Lambda Exam Tips

- Lambda scales out (not up) automatically
- Lambda functions are independent, 1 event = 1 function
- Lambda is serverless
- Know what services are serverless!
- Lambda functions can trigger other lambda functions, 1 event can = x functions if functions trigger other functions

KMS 101



Accolades

<https://accolades.corp.amazon.com/>

Architecture Accelerating

<https://kiku.aws.training/SessionSearch?pageNumber=1&courseId=37002&trainingProviderId=1>

AWS Cloud Economics (3.5 hrs)

<https://kiku.aws.training/Details/eLearning?id=36904>

AWS Technical Professional (3 hrs)

<https://kiku.aws.training/Details/Curriculum?id=45423>

Solutions Architecture

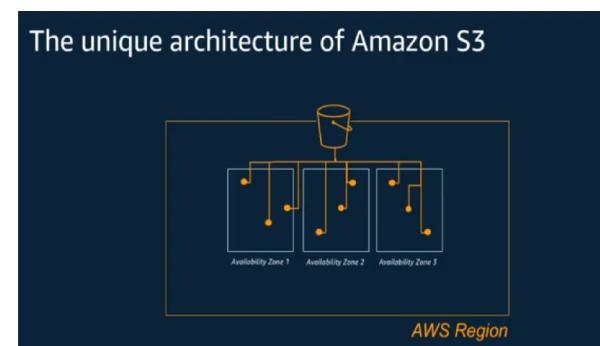
- SA's responsibilities:
 1. Remove technical barriers and move the sales cycle forward
 2. 1 to many evangelism
 3. Make the Platform Better
 4. Hire and Develop the Best
- Focus on the 3 Components of your 90-day Ramp
 1. Culture
 2. Technology
 3. Customer Facing Techniques

Amazon Confidential – Page 3



Ramp up

- Getting started with S3: https://www.youtube.com/watch?v=vFfY_-TL-pc



Amazon S3 storage classes

S3 Standard	S3 Intelligent-Tiering	S3 Standard-IA	S3 One Zone-IA	S3 Glacier	S3 Glacier Deep Archive
Access frequency					Inrequent →
Active, Frequently accessed data Milliseconds access ≥ 5 AZ \$0.0210/GB	Data with changing access patterns Milliseconds access ≥ 3 AZ \$0.0210 to \$0.0125/GB	Infrequently accessed data Milliseconds access > 3 AZ \$0.0125/GB	Re-creatable, less accessed data Milliseconds access 1 AZ \$0.0100/GB	Archive data Minutes or hours access > 3 AZ \$0.0040/GB	Archive data Hours to access > 3 AZ \$0.0009/GB

```

1. bash
bash                               head-object (bash)          put-object (bash)
pmeighan:demo $ aws s3api put-object --bucket a-demo-bucket-101 --key kids.jpg --body ./kids.jpg
{
  "ETag": "\"f40089a3393e810e3030d05b5f9c522d\""
}
pmeighan:demo $ aws s3api head-object --bucket a-demo-bucket-101 --key kids.jpg
{
  "AcceptRanges": "bytes",
  "LastModified": "Tue, 18 Feb 2020 23:40:06 GMT",
  "ContentLength": 2055626,
  "ETag": "\"f40089a3393e810e3030d05b5f9c522d\"",
  "ContentType": "binary/octet-stream",
  "Metadata": {}
}
pmeighan:demo $

```

Object tags work with lifecycle policies

Perform automated actions on a subset of your data with object tags

Lifecycle

Specify a tag filter to transition or expire objects
Use S3 batch operations to apply object tags at scale
Ex: Transition all objects tagged "Project : Delta" to S3 Glacier



Lifecycle policies use rules to manage your storage

Use lifecycle policies to transition objects to another storage class

Lifecycle rules take action based on object age. Here's an example:

1. Move objects older than 30 days to S3 Standard — Infrequent Access
2. Move objects older than 365 days to S3 Glacier Deep Archive



S3 Standard



S3 Standard —
Infrequent Access



S3 Glacier
Deep Archive

Block Public Access

Applies blanket protection against accidental public access

Can be applied to ACL access, bucket policy access, or both

Can be applied to new data or to all data

Set at the bucket level or at the account level

Amazon S3 default encryption



One-time
bucket-level
setup



Automatically
encrypts all new
objects



Simplified
compliance



Supports SSE-S3
and SSE-KMS

Provides Amazon S3 encryption-at-rest support for applications that do not otherwise support encrypting data in Amazon S3

Amazon S3 inventory

A managed alternative to using the LIST API



Regularly generates a list of objects for [analytics](#) and [auditing](#).

- Encryption status
- Storage class
- Creation date
- Replication status
- Object size, and more
- S3 Intelligent-Tiering access tier [new!](#)

S3 Data Protection capabilities

GOAL

- Replicate data for compliance and bad actor protection

- Protect data from accidental deletes

- Protect data for governance and compliance purposes

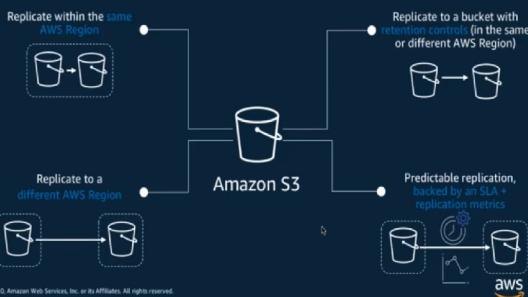
AMAZON S3 FEATURES

- Use [S3 Replication](#) with [Replication Time Control](#) and [ownership override](#)

- Use [bucket versioning](#) while reducing cost with [Lifecycle policies](#)

- Use [S3 Object Lock](#) to store objects as write-once-read-many (WORM)

S3 replication



Amazon S3 Replication time control [new!](#)

Designed to replicate 99.99% of objects within 15 minutes



15 minute replication time backed by an [AWS Service Level Agreement \(SLA\)](#)



Monitor replication using [Amazon CloudWatch metrics](#) and event notifications

Amazon S3 Replication time control

Designed to replicate 99.99% of objects within 15 minutes

Monitor your replication with 3 new CloudWatch metrics

Optional: Set up alarms on your metrics



Amazon S3 Object Lock

Use [Object Lock](#) to store objects as write-once-read-many (WORM)

Compliance mode
Store compliant data

Governance mode
Store data in WORM format; privileged users can modify retention controls

Legal hold
If you're unsure how long you want your objects to stay immutable

```

1. bash
bash                         361          head-object (bash)           362          put-object (bash)           363
"ContentType": "binary/octet-stream",
"Metadata": {}
}
pmeighan:demo $ aws s3api put-object --bucket a-demo-bucket-101 --key kids.jpg --body ./kids.jpg
{
    "ETag": "\"f40089a3393e810e3030d05b5f9c522d\"",
    "ServerSideEncryption": "AES256"
}
pmeighan:demo $ aws s3api head-object --bucket a-demo-bucket-101 --key kids.jpg
{
    "AcceptRanges": "bytes",
    "LastModified": "Tue, 18 Feb 2020 23:53:56 GMT",
    "ContentLength": 2055626,
    "ETag": "\"f40089a3393e810e3030d05b5f9c522d\"",
    "ContentType": "binary/octet-stream",
    "ServerSideEncryption": "AES256",
    "Metadata": {}
}
pmeighan:demo $ aws s3api put-object --bucket a-demo-bucket-101 --key kids.jpg --body ./kids.jpg
{
    "ETag": "\"f40089a3393e810e3030d05b5f9c522d\"",
    "ServerSideEncryption": "AES256",
    "VersionId": "bdb5NpAUkgKQNB76eBDnKSDPMcLYXF6E"
}

```

- Introducing the New Macie: <https://www.youtube.com/watch?v=I-ewoQekdXE> (the Macie service you are using is new, we had an older version which is now called Macie Classic)

Amazon Macie

Discover and protect your sensitive data at scale

Gain visibility and evaluate	Discover sensitive data	Centrally manage at scale	Automate and take actions
<ul style="list-style-type: none"> Bucket inventory Bucket policies 	<ul style="list-style-type: none"> Inspection jobs Flexible scope 	<ul style="list-style-type: none"> AWS Organizations Managed & custom data detections 	<ul style="list-style-type: none"> Detailed findings Management APIs

Amazon Macie - Gain visibility and evaluate

- Provides customers visibility into S3 bucket inventory
 - Number of buckets
 - Storage size
 - Object count
- Monitors changes to S3 bucket policies
 - Publicly accessible
 - Unencrypted
 - Shared outside of the account
 - Replicated to external accounts

Across multiple accounts and automatically includes new buckets

Amazon Macie - Discover sensitive data

- Ongoing evaluation of your Amazon S3 environment and data

Select target for data discovery	Define the scope	Review status (complete, cancelled, idle)
Create and schedule jobs	Scheduled frequency (one-time, daily, weekly, monthly)	Take actions (Cancel, copy)
<ul style="list-style-type: none"> Select target for data discovery Create and schedule jobs 	<ul style="list-style-type: none"> Scheduled frequency (one-time, daily, weekly, monthly) Object criteria (Tags, modified time, extension type, size) 	<ul style="list-style-type: none"> Take actions (Cancel, copy)

Amazon Macie - Centrally manage at scale

Master/Member setup

- Multi-account with up to 1,000 member accounts
- AWS Organizations support up to 5,000 accounts

Macie master can create jobs on behalf of members

- One-click deployment with no upfront data source integration

With a few more clicks in the console, you can enable Macie across multiple accounts. Once enabled, Macie generates an ongoing Amazon S3 resource summary across accounts that includes bucket and object counts as well as the bucket-level security and access controls.

© 2020, Amazon Web Services, Inc. or its affiliates.
aws

Amazon Macie - Centrally manage at scale

Fully managed sensitive data types

- Amazon Macie maintains a growing list of sensitive data types that include common personally identifiable information (PII) and other sensitive data types as defined by data privacy regulations, such as GDPR, PCI-DSS, CCPA and HIPAA.



Data identifiers

- Financial (card, bank account numbers...)
- Personal (names, address, contact...)
- National (passport, ID, driver license...)
- Medical (healthcare, drug agency ...)
- Credentials & secrets (AWS secret keys, private keys ...)

Amazon Macie - Centrally manage at scale

Custom-defined sensitive data types

- Amazon Macie provides you the ability to add custom-defined data types using regular expressions to enable Macie to discover proprietary or unique sensitive data for your business.



- Regular expression that defines the pattern to match
- Keywords that define specific text to match
- Ignore words that define specific text to exclude

Amazon Macie - Centrally manage at scale

Supported file and storage formats in Amazon Macie

- When Amazon Macie analyzes data in an S3 bucket, it performs a deep inspection that factors the file or storage format for the data. Macie can analyze and detect sensitive data in many different formats, including commonly used compression and archive formats.



File and storage formats

Big Data - Apache Avro object containers and Apache Parquet files
Compression or archive - .gz, .gzip, .tar, .zip
Document - .doc, .docx, .pdf, .xls, .xlsx
Text - .csv, .htm, .html, .json, .tsv, .txt, .xml, and others (depending on the type of non-binary text file)

Amazon Macie - Automate and take actions

Finding types

- Bucket policy findings

- Sensitive data discovery findings

Findings categorized by

- By bucket
- By type
- By job



Detailed and actionable security and sensitive data discovery findings

- Findings sent to CloudWatch Events
- Bucket policy findings sent to Security Hub

Amazon Macie - Automate and take actions

Export results to S3 bucket

- Store all of your sensitive data discovery results in an S3 bucket



Automated actions on alerts

- Simplify with Lambda

Management APIs

- Integrate with additional services and tools
- Support for CloudFormation

Amazon Macie – Use Cases

Assessing your data privacy and security



An important aspect in maintaining the right level of **data security** is to be able to continuously **identify** your **sensitive data** and evaluate security and access controls.

NOTES:

- Built-in Data Identifier, Custom-Defined Sensitive Data Types
- Classification Findings
- Detailed Discovery Reports to Separate S3 bucket for customers to configure
- Delegated Administrator
- Getting started with Identity & Access Management (IAM): <https://www.youtube.com/watch?v=Zvz-qYYhvMk>
- Using KMS for data protection: <https://www.youtube.com/watch?v=hxWvbNvj2lg>
- Introduction to Cloud9: https://pages.awscloud.com/Introducing-AWS-Cloud9_0106-DEV_OD.html
- Attached sample encryption requirements doc