AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

STG204

Data archiving and digital preservation solutions with AWS

Gayla Beasley
Senior Program Manager, Amazon S3 Glacier
Amazon Web Services

Marc Trimuschat
Head of WW Storage Go to Market
Amazon Web Services

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modernizing Archive

Why archive data?

Benefits of cloud archival

Amazon S3 archival storage classes

Archival storage use cases

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

New at re:Invent 2021

Amazon S3 Archive Launches

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Introducing S3 Glacier Instant Retrieval



NEW

What is it?

- For long-lived archive data that requires milliseconds retrieval
- 99.999999999% (11 9s) of durability
- Designed for 99.9% availability

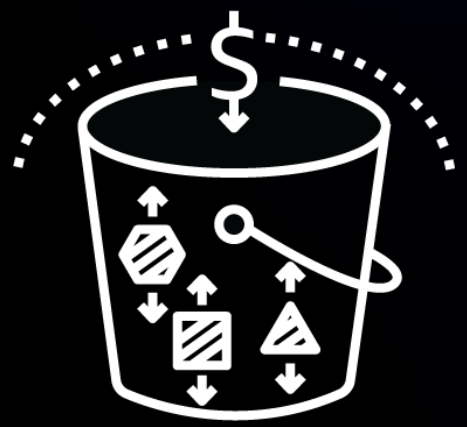
What are the use cases?

- Petabytes of archive data stored for indefinite periods of time
- Only a small percentage of this archive data is accessed each year
- Archive data must be immediately accessible when requested

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Introducing the Archive Instant Access Tier



NEW

What is it?

- A new access tier in S3 Intelligent-Tiering
- Corresponds to S3 Glacier Instant Retrieval
- Automatically reduces cost for data that lives for more than 90 days

What about existing data that is older than 90 days?

- Automatically tiers down to the new tier
- Results in an automatic storage cost reduction
- No action required by you

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Changes to the S3 Glacier storage class



New name! NEW

- S3 Glacier Flexible Retrieval (formerly S3 Glacier)
- No changes to APIs, Inventory Reports, etc.

New storage price! NEW

- Existing prices reduced
- Results in an automatic storage cost reduction

Free retrievals! NEW

- Bulk retrievals are now FREE
- No change to APIs or management

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Most of the world’s data is cold

Customers in every industry archive data in AWS



Media & Entertainment



Gaming



Healthcare & Life Sciences



Financial Services



Power & Utilities



Energy



Manufacturing



Retail



Telecom



Automotive



Education



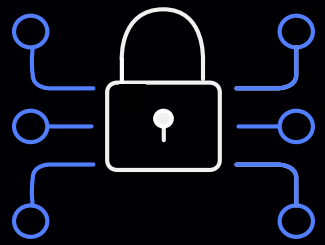
Government

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why archive to AWS?



Durability & Resilience



Security & Compliance



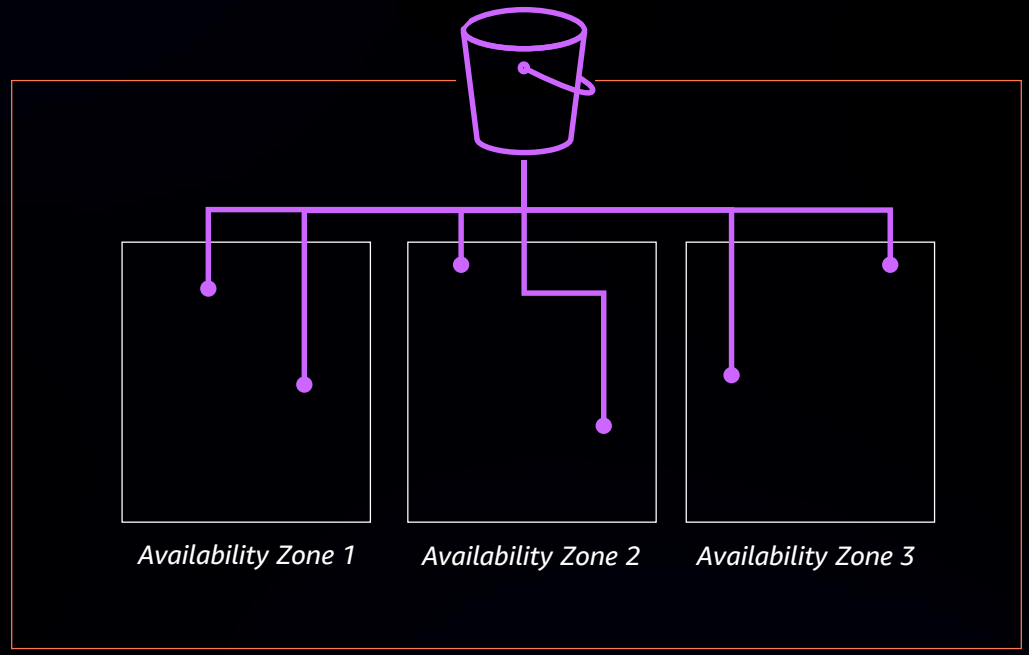
Lowest Cost

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Designed to provide
99.999999999%
of data durability

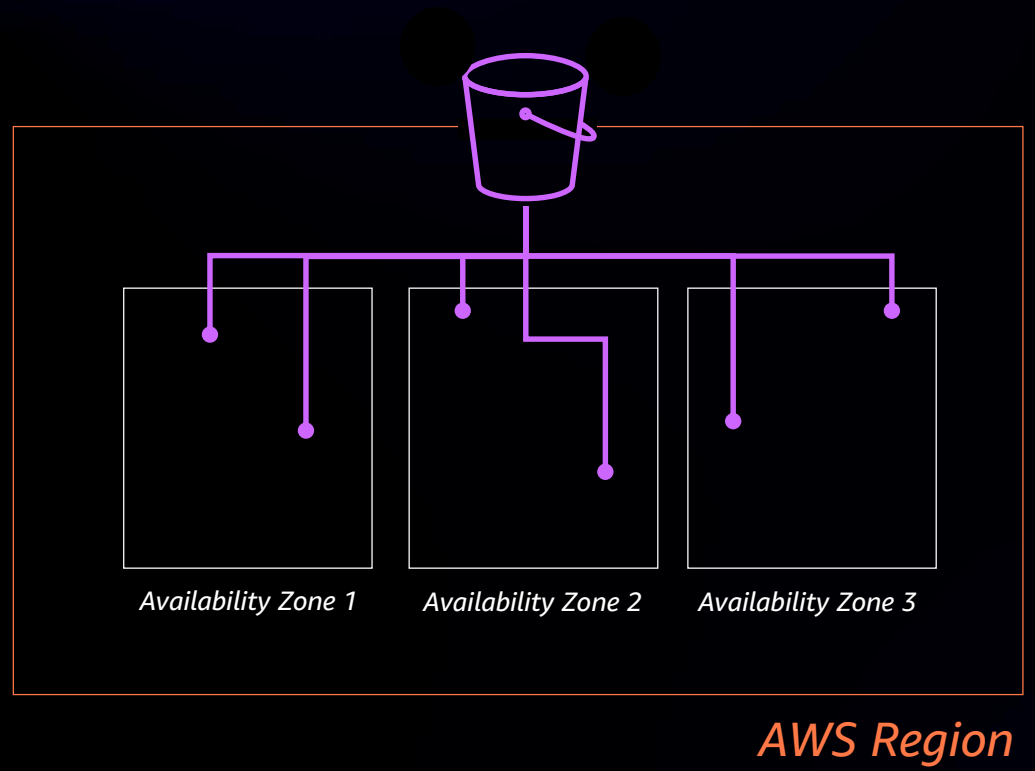
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The unique architecture of Amazon S3



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

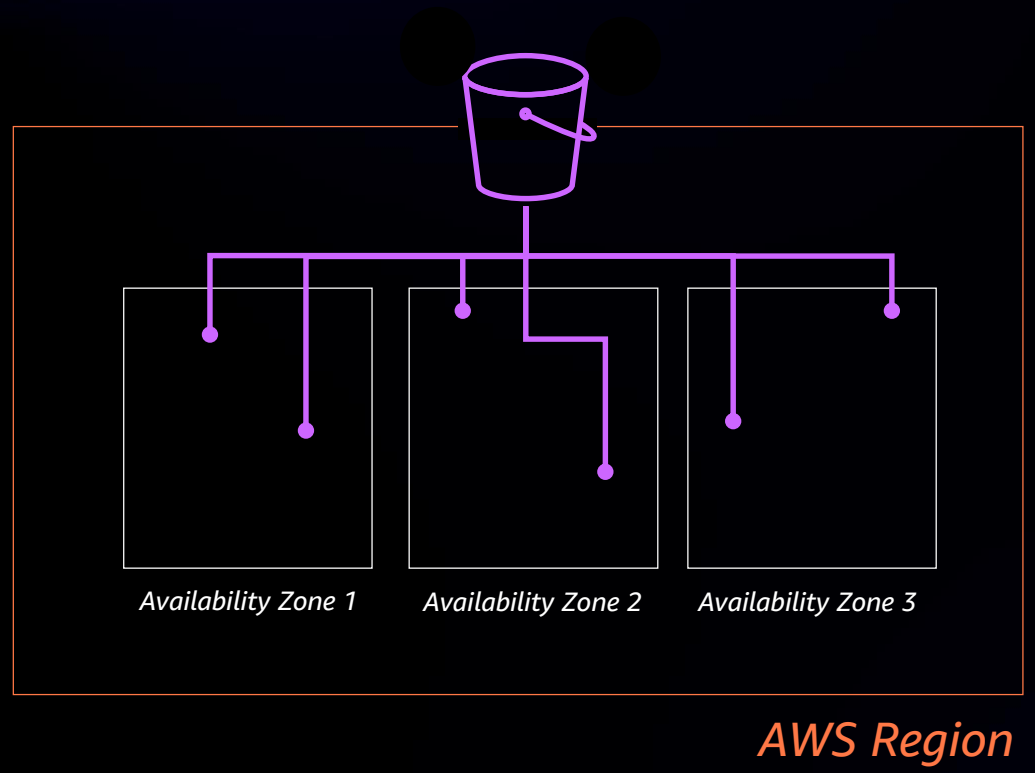
The unique architecture of Amazon S3



- Stored redundantly across a minimum of 3 Availability Zones
- Stored redundantly across multiple devices within an Availability Zone
- Designed to sustain concurrent device failures

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

A culture of durability



- Durability review & operational safeguards
- Robust integrity checking
- Auditors check and re-check data at rest

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

At AWS, security is Job Zero

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 features to secure and protect archives



S3 Block Public Access

Override S3 permissions that allow public access, with a centralized control to prevent variation in security configuration



S3 Encryption

Encrypt data in transit and at rest, using S3-managed or customer-managed keys; Manage client-side, or with bucket defaults



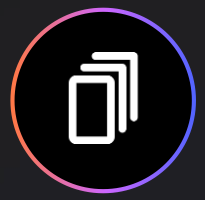
S3 Access Points

Scale access for hundreds of apps with permissions customized for each app



S3 Object Lock

Prevents an object from being deleted or overwritten for a fixed amount of time or indefinitely



S3 Versioning

For iterating on projects and protecting content from unintended deletes



S3 Replication

Protect archives by replicating in-Region or between Regions with different ownership; Distribute content to partners

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lowest cost storage in the cloud

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 storage classes



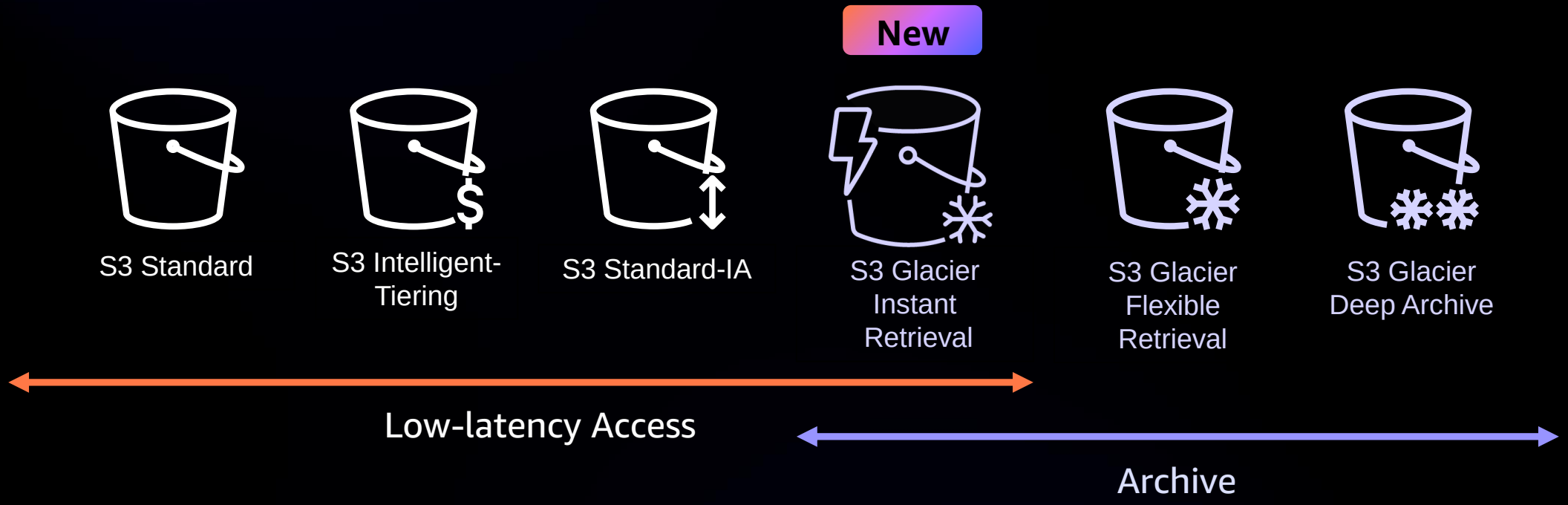
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 storage classes



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 storage classes



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Introducing S3 Glacier Instant Retrieval



NEW

What is it?

- For long-lived archive data that requires milliseconds retrieval
- 99.999999999% (11 9s) of durability
- Designed for 99.9% availability

What are the use cases?

- Petabytes of archive data stored for indefinite periods of time
- Only a small percentage of this archive data is accessed each year
- Archive data must be immediately accessible when requested

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 Glacier Flexible Retrieval



Bulk retrievals are now FREE

Price drop

What is it?

- For long-lived archive data and long-term backup
- 99.999999999% (11 9s) of durability
- Retrievals in 3-5 hours for standard
- Free Bulk retrievals in 5-12 hours

What are the use cases?

- Petabytes of archive data stored for indefinite periods of time
- Data accessed 1-2 times per year

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 Glacier Deep Archive



What is it?

- Archiving long-term data that which accessed infrequently
- 99.999999999% (11 9s) of durability
- Retrievals within 12 to 48 hours

What are the use cases?

- Archive data backups that are rarely accessed
- Data that needs to be retained for the long term

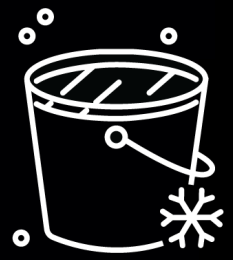
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Which archive storage class is right for me?

1. Storage Cost
2. Retrieval Speed
3. Data Retention

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Choosing between S3 Glacier archive storage

	 S3 Glacier Instant Retrieval	 S3 Glacier Flexible Retrieval	 S3 Glacier Deep Archive
Storage cost	**$0.004 per GB-month**	**$0.0036 per GB-month**	**$0.00099 per GB-month**
Data retrieval	Instant	Expedited: 1-5 minutes Standard: 3-5 hours Bulk: 5-12 hours	Standard: Within 12 hours Bulk: Within 48 hours
Minimum object duration	**90 Days**	**90 days**	**180 days**

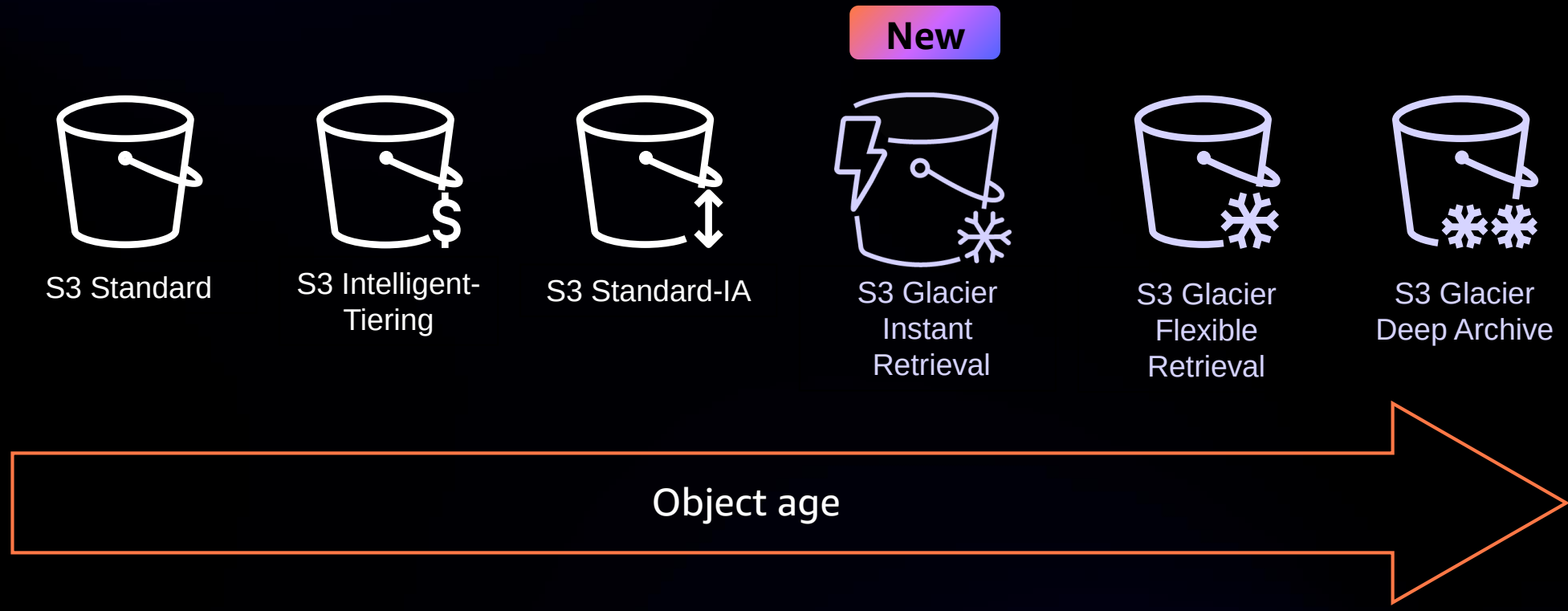
Bulk retrievals are now FREE

Price drop

aws

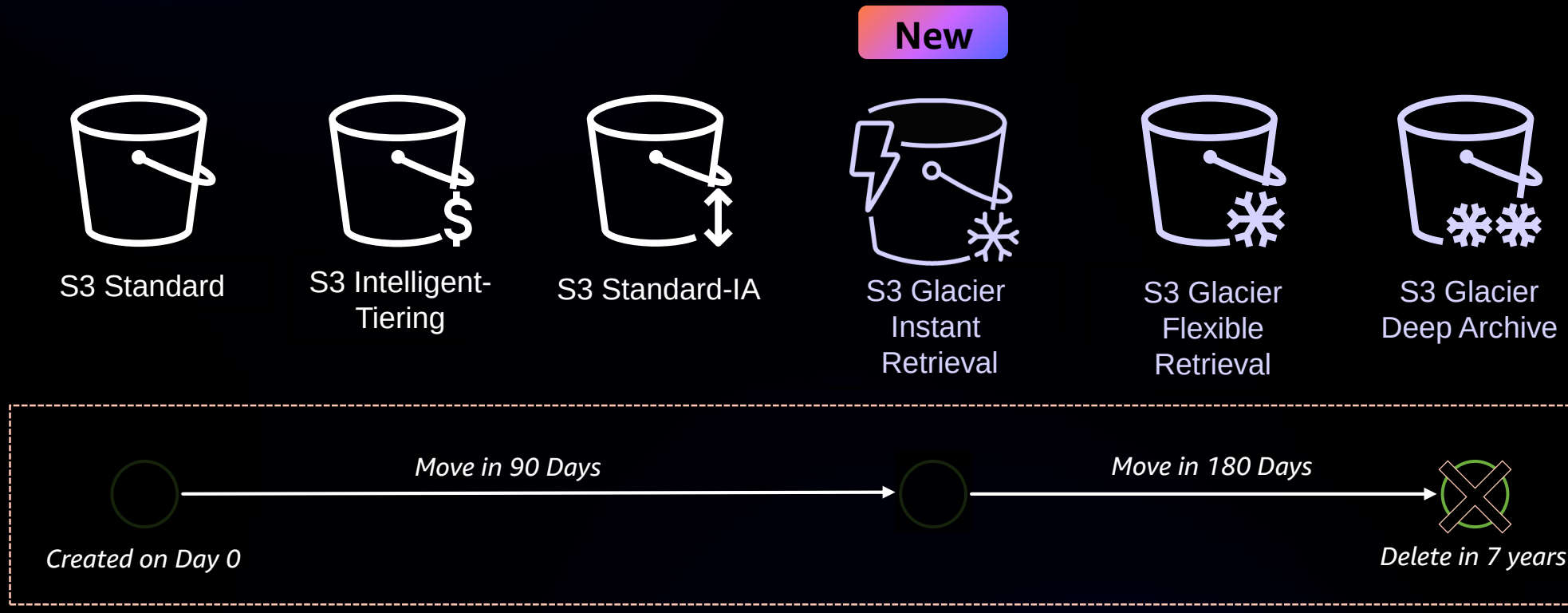
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tier data to optimize storage costs



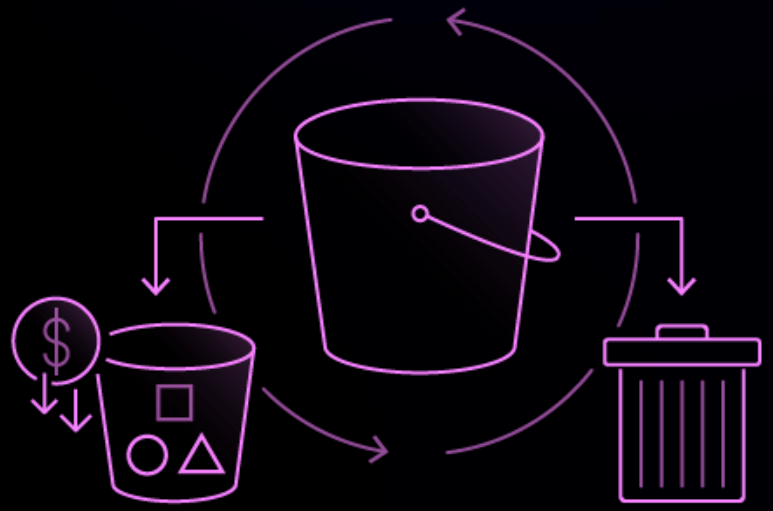
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tier data to optimize storage costs



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

S3 Lifecycle options



Transition and expire data based on…

- Bucket
- Prefix
- Object Tag
- Object Size **New**
- Number of Versions **New**

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What if my access patterns are unpredictable?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

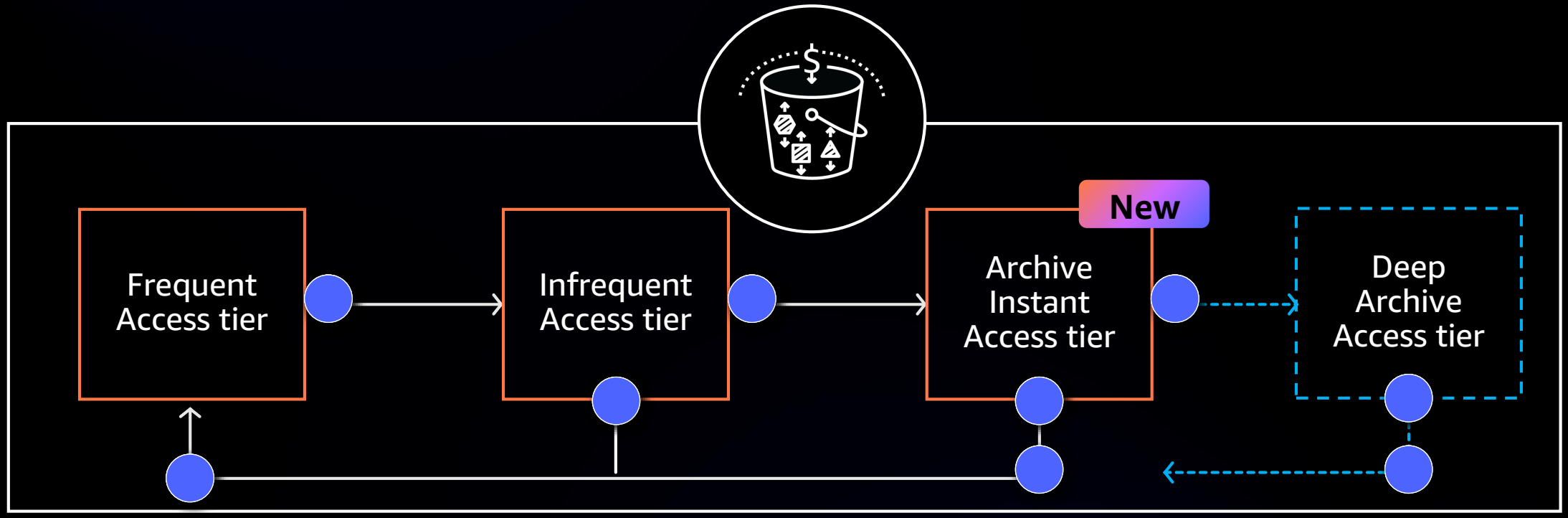
Amazon S3 Intelligent-Tiering



- Automatically moves objects between three access tiers
- Optional asynchronous archiving to realize lowest storage cost in the cloud
- No performance impact, operational overhead, lifecycle fees, or retrieval fees
- Designed for 99.9% availability and 99.999999999% durability

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Use S3 Intelligent-Tiering by default for data with unknown or changing access patterns



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Archiving best practices

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Companies in every industry archive data in AWS



Media & Entertainment



Gaming



Healthcare & Life Sciences



Financial Services



Power & Utilities



Energy



Manufacturing



Retail



Telecom



Automotive



Education



Government

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Getting from...

This



To this

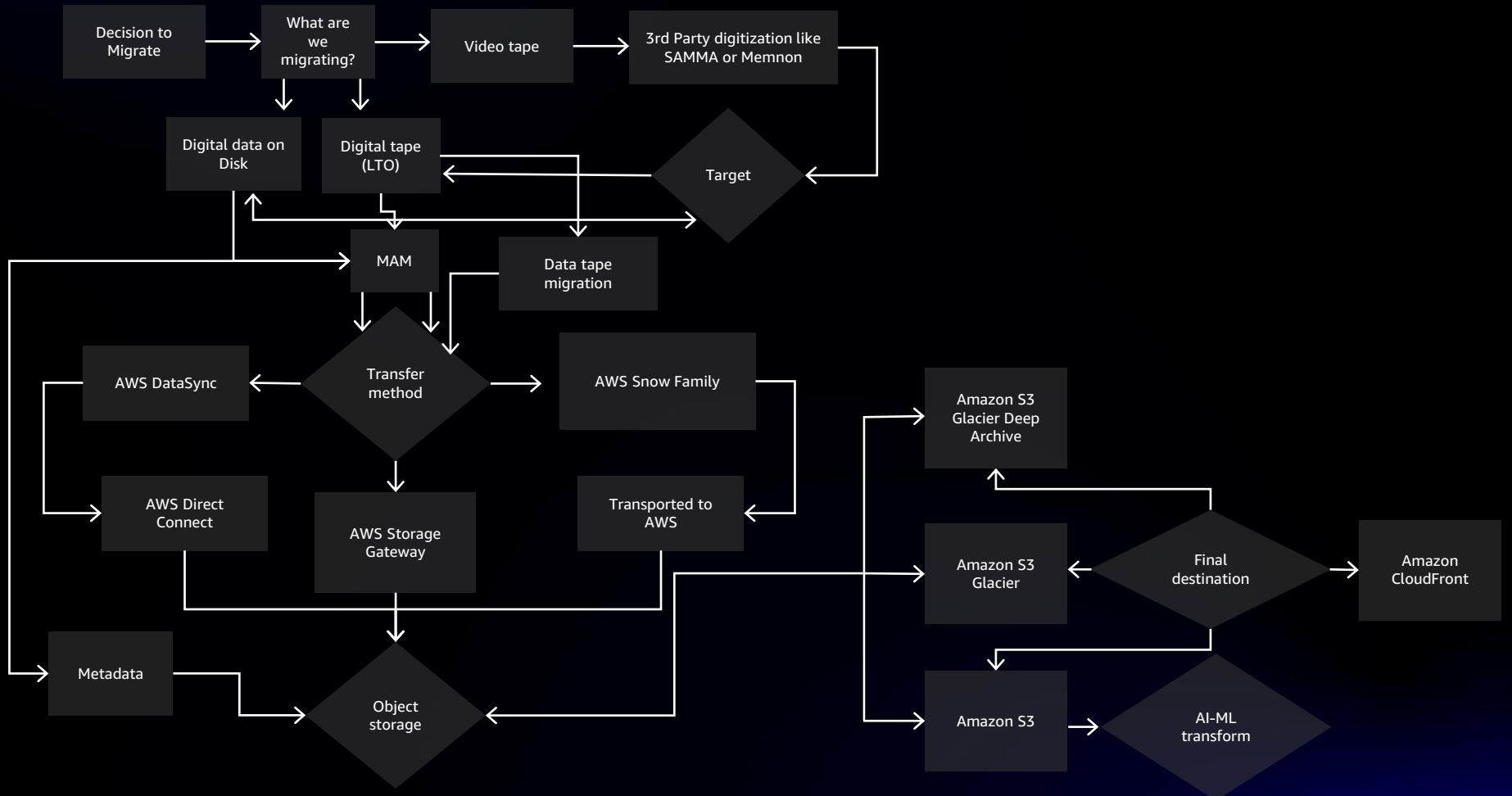
Amazon S3 Glacier Instant Retrieval storage class
Milliseconds retrieval of data in a low-cost archive S3 storage class

Amazon S3 Glacier Flexible Retrieval storage class
Minutes to 12 hours retrieval of data in a lower cost archive S3 storage class

Amazon S3 Glacier Deep Archive storage class
12 – 48 hours retrieval of data in the lowest cost archive S3 storage class

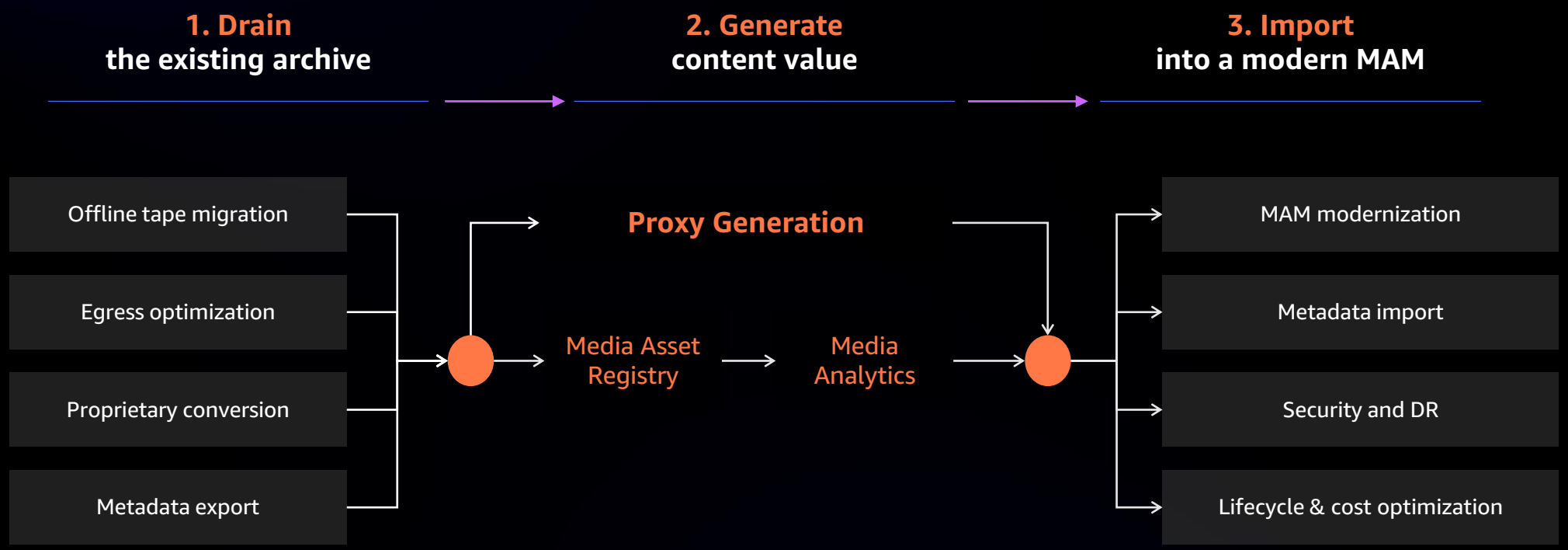
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Even basic media migration workflows are complex



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

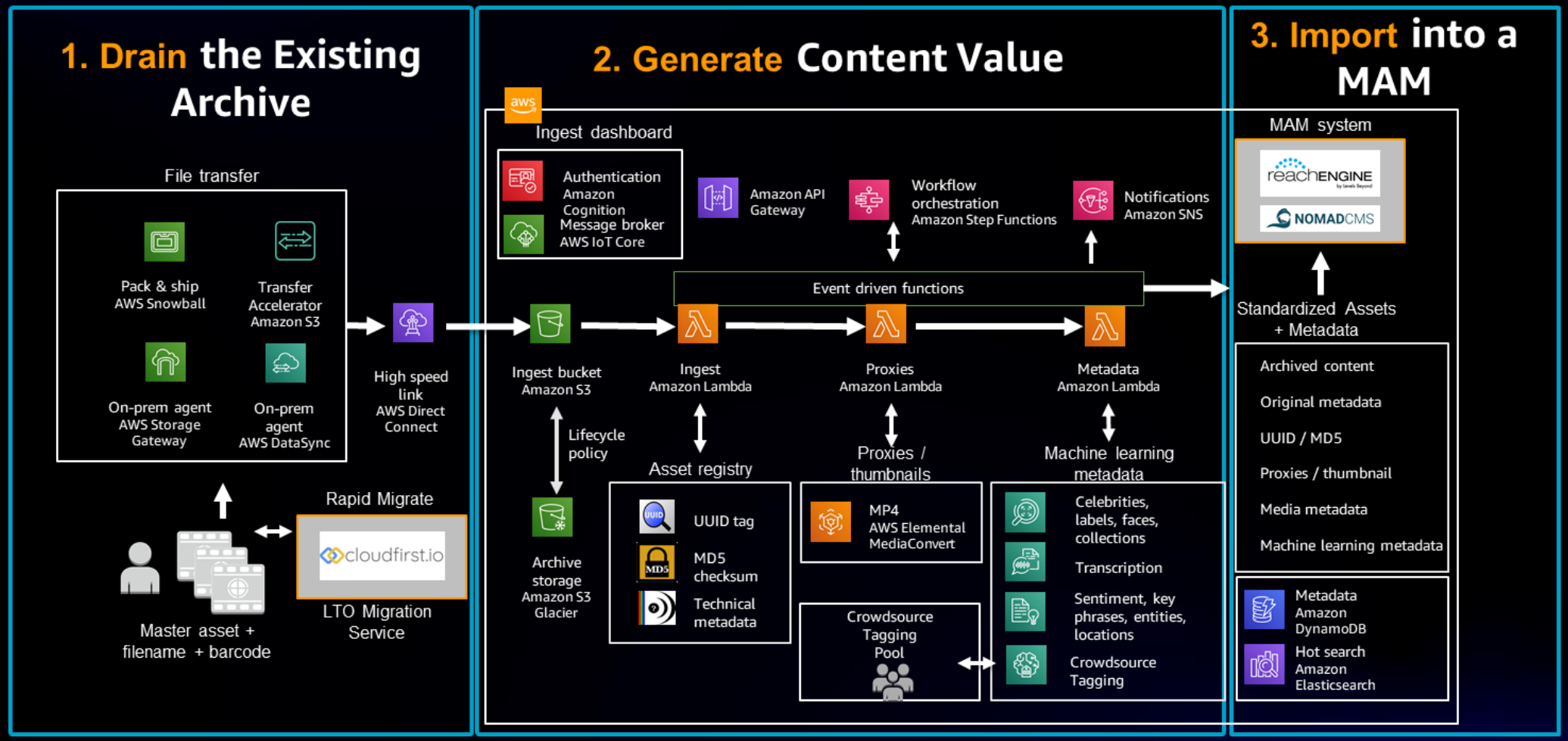
For media archives, a simpler solution: Media2Cloud



- 3-phased project template: Export assets with metadata, process/analysis, and import to MAM
- Flexible to accommodate different workflows, content types, and partner components
- Professional services to provide optional assessment, and project management
- Partners to provide offline tape migration

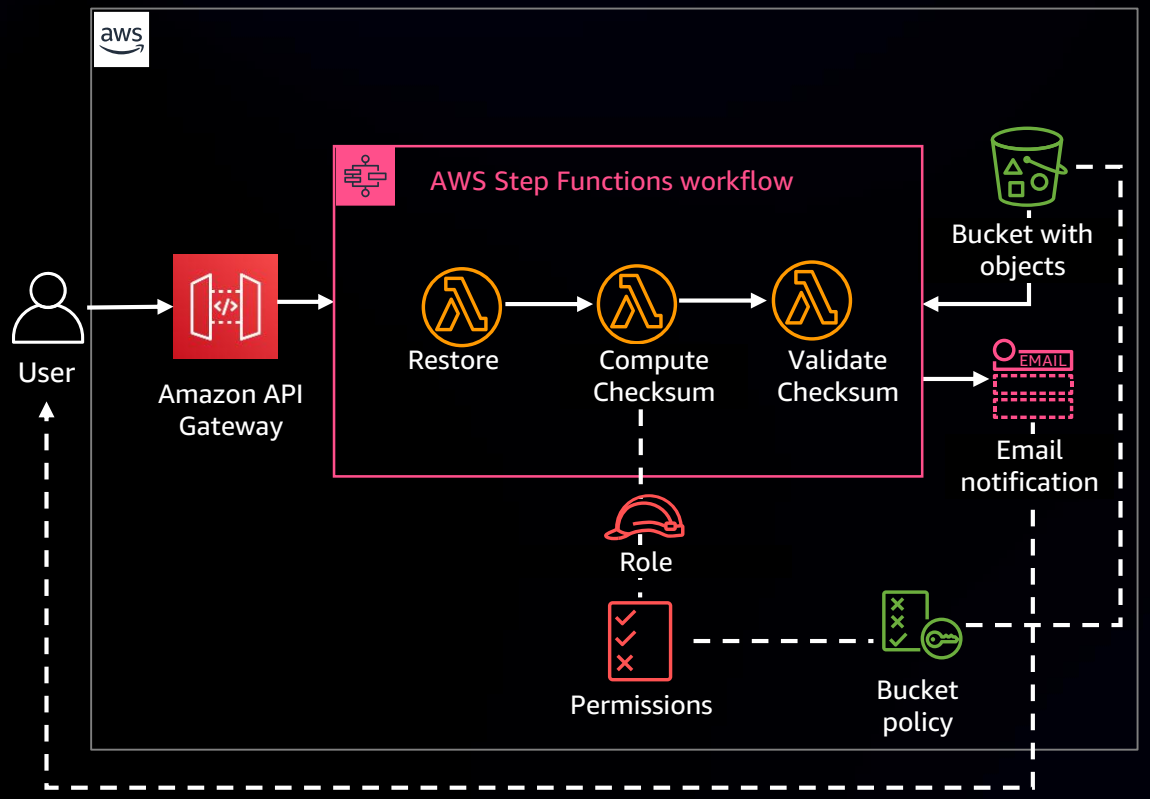
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Media2Cloud: architecture overview



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Solution: Serverless fixity



- Easily calculate checksums of S3 objects on-demand for digital preservation compliance and regulatory reasons
- Perform serverless checksums using MD5 or SHA1 on objects stored in any Amazon S3 storage class, including S3 Glacier or S3 Glacier Deep Archive
- Provides a CloudFormation template, reference architecture, and deployment guide to simplify solution deployment
- New AWS Solution available today at aws.amazon.com/solutions

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

New checksum options

Coming Soon

New checksum

`x-amz-checksum = CRC32C`



SHA-256 | **SHA-1** | **CRC32** | **CRC32C**

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Accelerate integrity checks by up to 92%

Coming Soon

1. Trailing checksums allow you to check data while you stream it in

2. Parallel checksums allow you to break large objects up, and leverage many cores

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Digital preservation with Amazon S3 Glacier



The Rock & Roll Hall of Fame museum documents the history of rock music and the artists, producers, engineers, and other notable figures who have influenced its development

OPPORTUNITY

New strategic plan to build RRHOF of the future, making the 21st century Rock Hall accessible worldwide and providing critical preservation of one of a kind digital media

SOLUTION

Partnered with AWS and Tape Ark to migrate all media into Amazon S3 via AWS Snowball, and set lifecycle policies in S3 to move media into Amazon S3 Glacier Deep Archive

BENEFITS

- Cost-effective storage to preserve estimated 350 TBs of data or approximately 2,800 hours of video material
- Recovered original preservation files through tape backups

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customers in every industry archive data in AWS



Media & Entertainment



Gaming



Healthcare & Life Sciences



Financial Services



Power & Utilities



Energy



Manufacturing



Retail



Telecom



Automotive



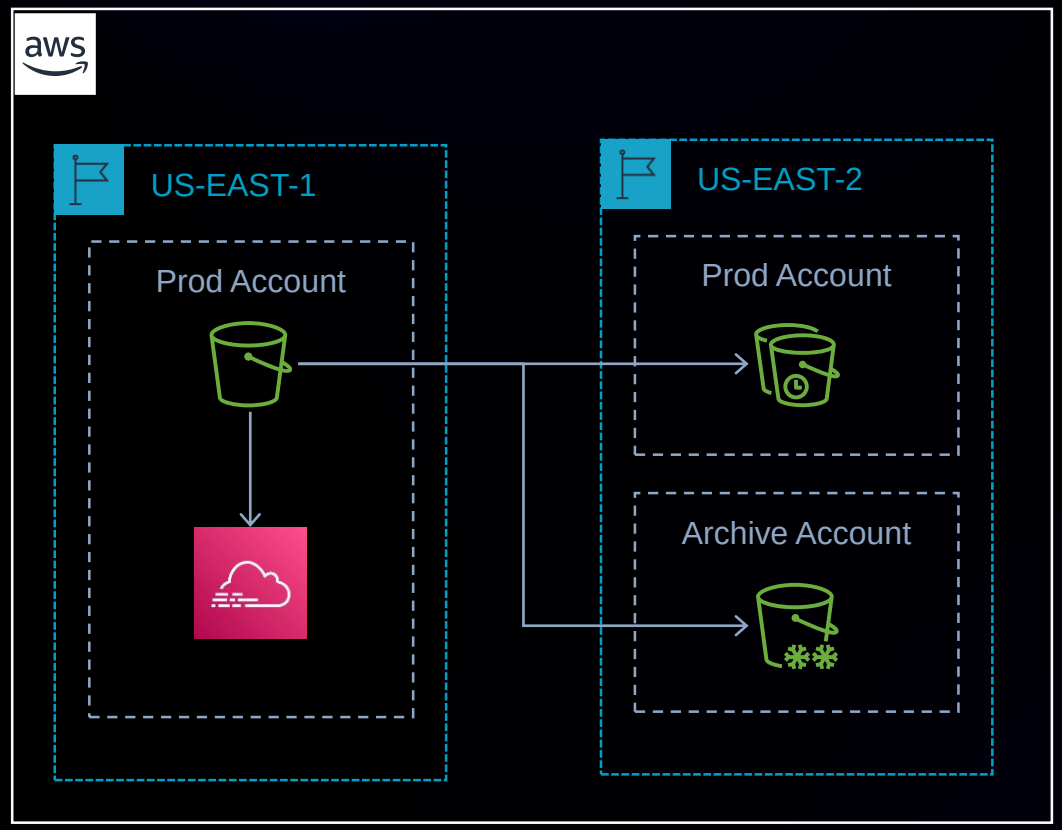
Education



Government

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

FSI Requirements on security & resilience

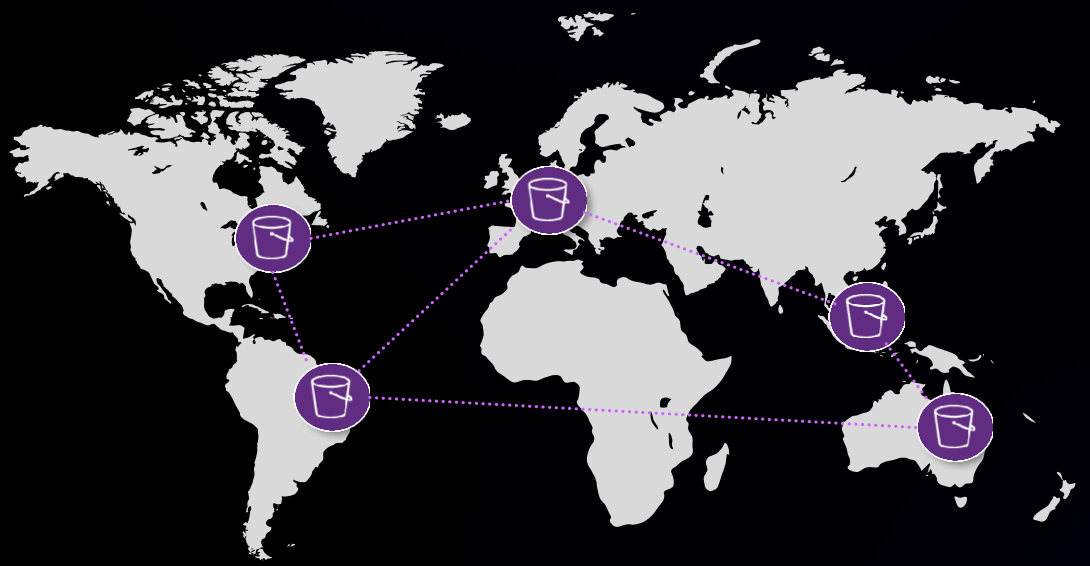


Critical datasets are…

- Replicated according to SLA
- Rendered immutable
- Encrypted at rest and in flight
- Meticulously logged

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-Region S3



- Multi-Region Access Points accelerate and simplify multi-Region apps
- First of its kind SLA on replication latency
- Robust metrics monitor replication status down to the object level

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

More replication options

Replicate within the same AWS Region



Replicate to a bucket with retention controls (in the same or different AWS Region)



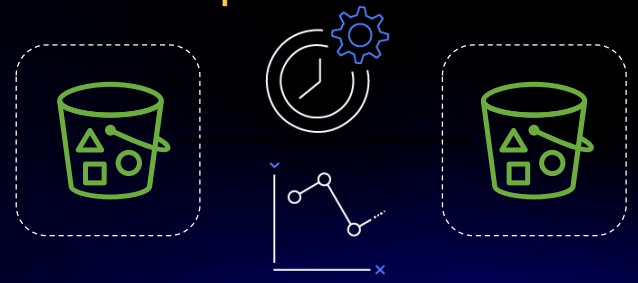


Amazon S3

Replicate to a different AWS Region



Replicate faster to a different AWS Region, backed by an SLA + replication metrics



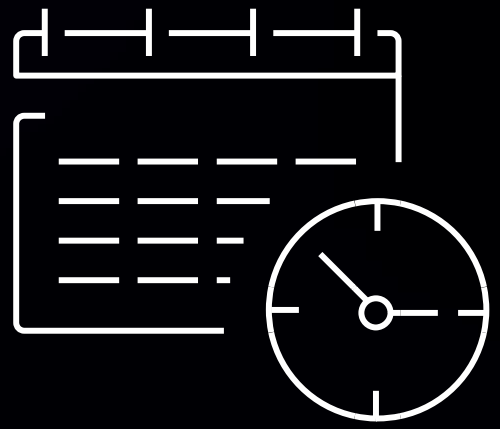
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 Object Lock for immutability



Immutable Objects in S3

Assessed for SEC compliance by Cohasset Associates



Retention Controls

Down to the object and object version levels



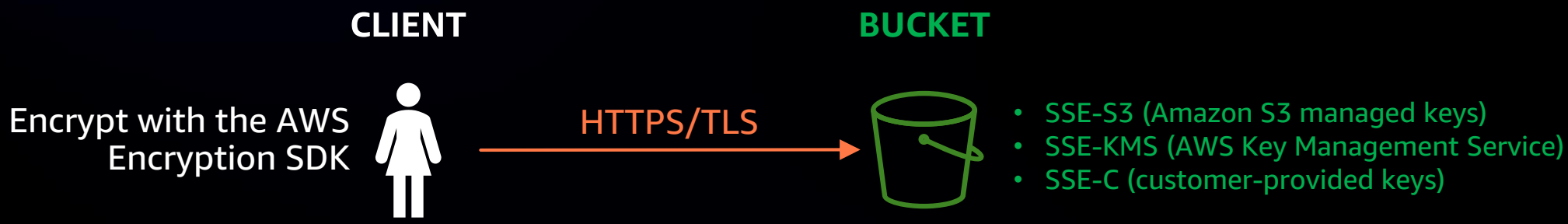
Auditing & Visibility

Track, report, and alert on access & retention changes

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 encryption support



1 Client-side

2 In transit

3 At rest

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Companies in every industry archive data in AWS



Media & Entertainment



Gaming



Healthcare & Life Sciences



Financial Services



Power & Utilities



Energy



Manufacturing



Retail



Telecom



Automotive



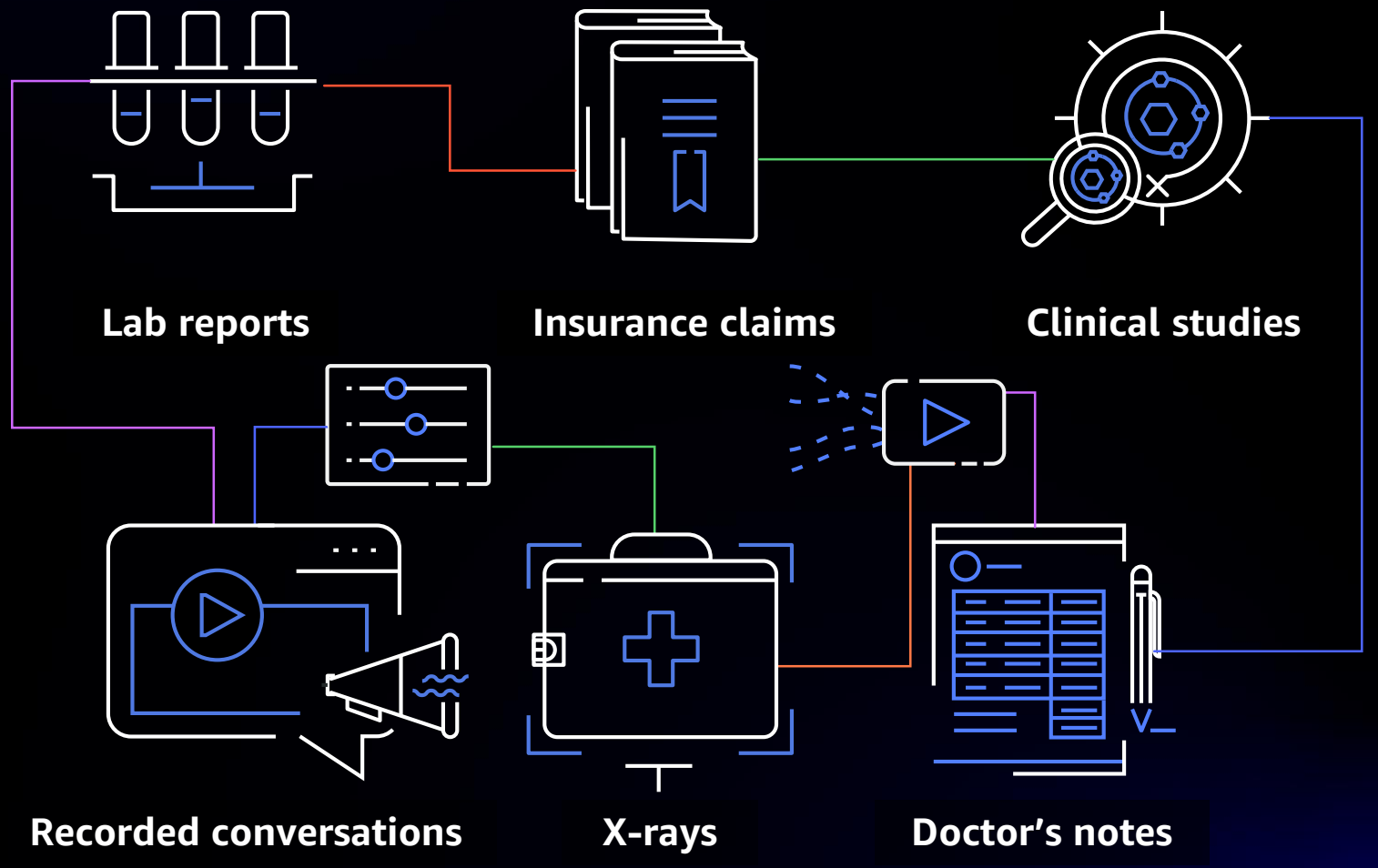
Education



Government

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Health data is inconsistent, unstructured, and complex



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon HealthLake

A HIPAA-eligible service that enables healthcare providers, health insurance companies, and pharmaceutical companies to store, transform, query, and analyze health data in the AWS Cloud at petabyte scale



IMPORT

- Quickly and easily import clinical notes, lab reports, insurance claims, and more



STORE

- Stored in AWS in a secure, compliant, and auditable way



TRANFORM

- Tag and index unstructured data using specialized ML models



QUERY & SEARCH

- Powerful query and search capabilities



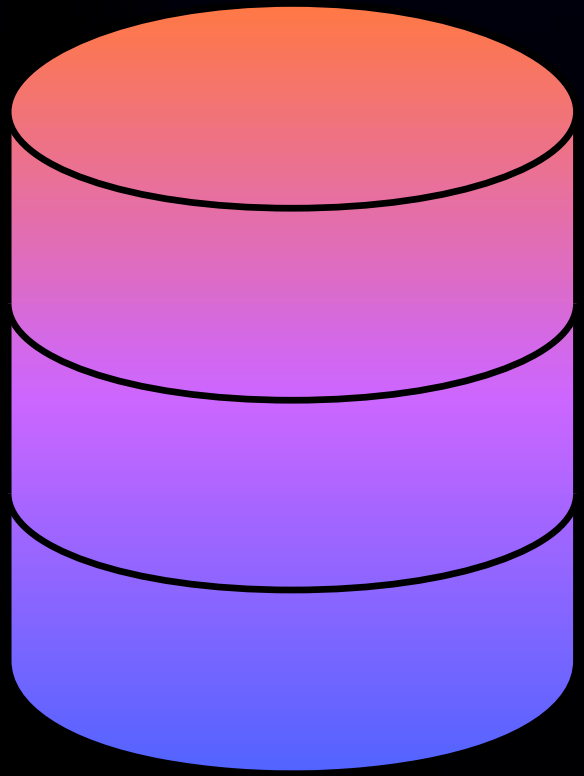
ANALYZE

- Understand relationships in the data with pre-built ML models

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Identify trends and make predictions



STRUCTURED, NORMALIZED DATA

Amazon SageMaker

Amazon QuickSight

3rd party applications

IDENTIFY TRENDS

Easily search and query the data

Develop interactive dashboards

MAKE PREDICTIONS

Create predictive models, like predicting Sepsis in a hospital ICU

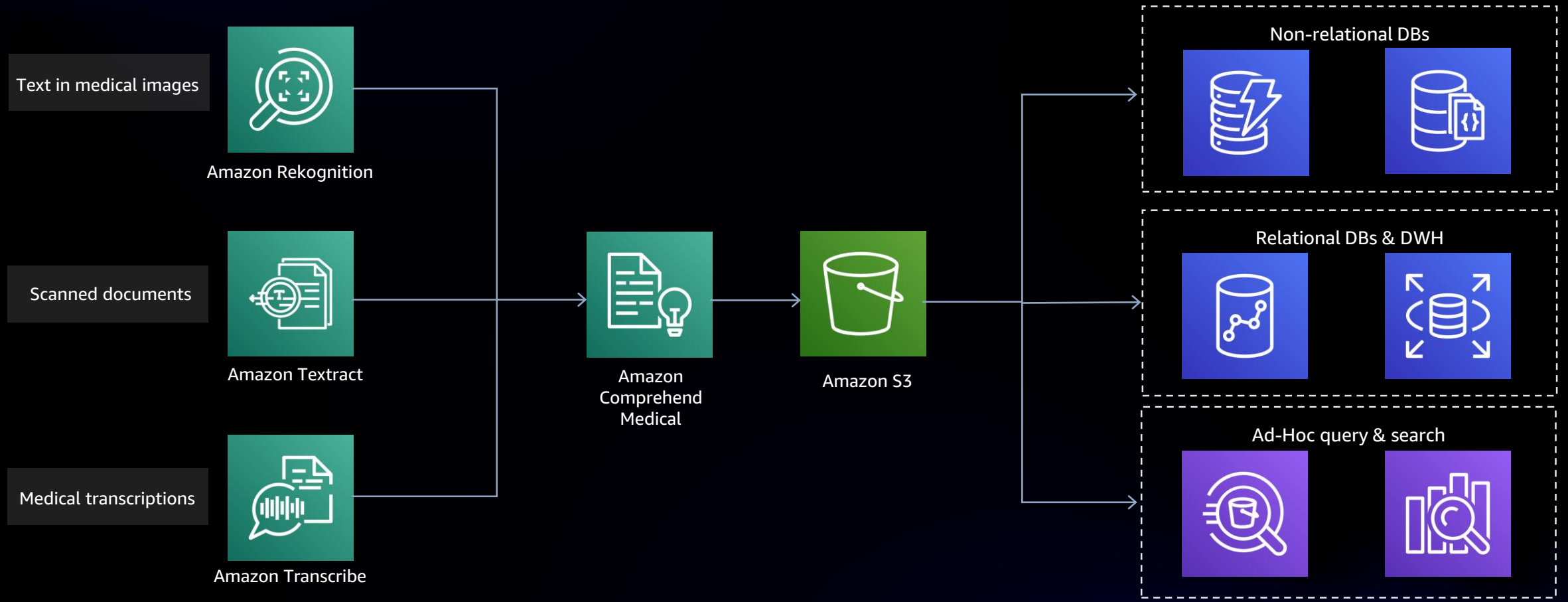
SHARE DATA SECURELY

Healthcare providers and patients alike can easily share and access their data

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

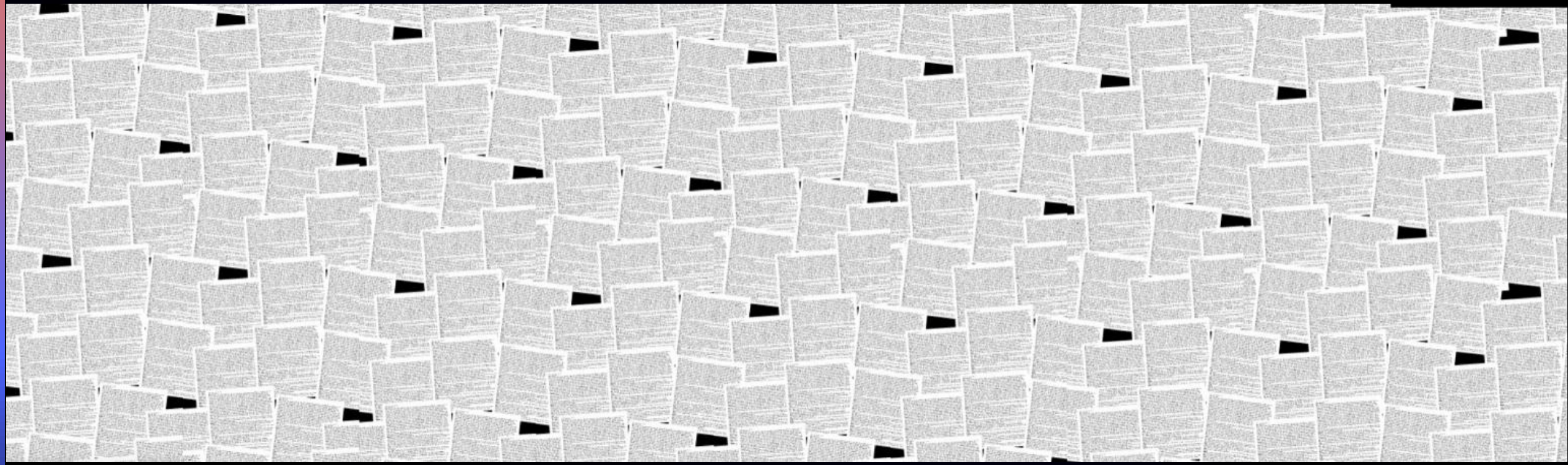
An example architecture



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Unstructured Medical Documents

1.2B created / year, critical information trapped, difficult to extract insights



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Extracting from images – Amazon Rekognition

Enables customers to build their own specialized machine learning based image analysis capabilities using Auto ML. No machine learning experience required.

SIMPLIFY DATA LABELING

- The Rekognition Custom Labels console provides a visual interface to make labeling your images fast and simple.

AUTOMATED MACHINE LEARNING

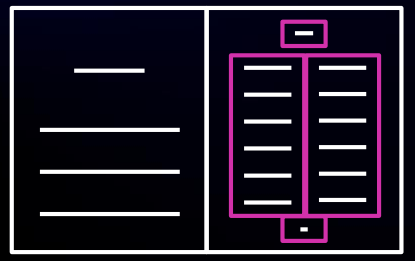
- No machine learning expertise is required to build your custom model. Rekognition Custom Labels includes AutoML capabilities that take care of the machine learning for you.

SIMPLIFIED MODEL EVALUATION, INFERENCE AND FEEDBACK

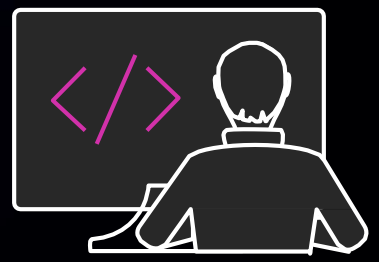
- Evaluate your custom model's performance on your test set. For every image in the test set, you can see the side by side comparison of the model's prediction vs. the label assigned.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

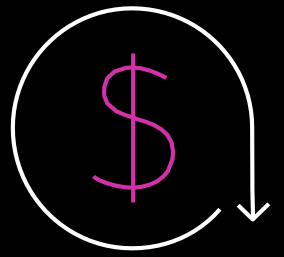
Extracting from medical forms – Amazon Textract



Extract data quickly and accurately



Eliminate manual effort



Lower document processing costs

KEY FEATURES

Optical Character Recognition (OCR)	Key-value pair detection	Table detection	Adjustable confidence thresholds	Bounding box coordinates	No ML experience required

Alfresco®

Roche

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Storing petabytes of PHI in AWS



- Philips set up Attunity CloudBeam for Amazon Redshift in less than one minute to upload 37M health records to AWS in 90 minutes
- Store and optimize any size dataset within two hours, allowing its consultants and analysts to provide solutions quickly to customers
- Philips HealthSuite Data Platform on AWS houses over 40+ PB of PHI, and growing at ~1 PB a month

"Thanks to AWS, our business was able to upload over 37 million records into Amazon Redshift in 90 minutes"

—Doug Ranaham
Manager, Philips Healthcare

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Summary of launches

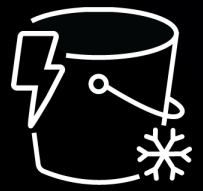


S3 Intelligent-Tiering

Optimizes storage costs by moving objects between different access tiers

Archive Instant Access tier **NEW**

NEW



S3 Glacier Instant Retrieval

Milliseconds retrieval



S3 Glacier Flexible Retrieval

(formerly S3 Glacier)

Minutes to hours retrieval

Free bulk retrievals **NEW**

Lower storage price **NEW**



S3 Glacier Deep Archive

12–48 hours retrieval

Amazon S3 is also reducing storage prices up to 31% for S3 Standard-Infrequent Access across various Regions

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Gayla Beasley

gybsl@amazon.com

Marc Trimuschat

mtrims@amazon.com

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.