
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

STP305-R

How to stay afloat in your data lakes

Winston Nwanne

Startup Solutions Architect
Amazon Web Services

Christine Samson

Manager, Startup Solutions Architecture
Amazon Web Services

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Data Lakes on AWS
- Security & Access Management
- Data Lake Cost Optimization
- Monitoring & Managing Your Data Lake Performance
- Q&A

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data Lakes on AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data lake recap

A data lake is a centralized repository that allows you to store all your structured and unstructured data at any scale

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3



Industry-leading scalability, availability, and durability

Easily manage data and access controls

Broad storage capabilities in the cloud

Industry-leading performance

Strong security, compliance, and audit capabilities

Wide range of cost-optimization capabilities

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How do I secure my data lake?

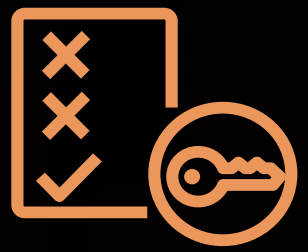
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Security & access management

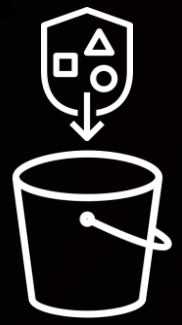
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Secure your data lake

DATA STORED IN AMAZON S3 IS DESIGNED TO BE SECURE BY DEFAULT



Access control with IAM and bucket policy



Amazon S3 Access Points



Amazon S3 data encryption



Amazon S3 object tagging



Amazon S3 Access Analyzer



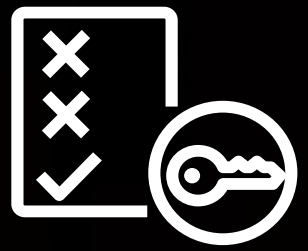
Amazon S3 Object Lock

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Secure your data lake

DATA STORED IN AMAZON S3 IS DESIGNED TO BE SECURE BY DEFAULT



Access control with IAM and bucket policy



Amazon S3 Access Points



Amazon S3 data encryption



Amazon S3 object tagging



Amazon S3 Access Analyzer



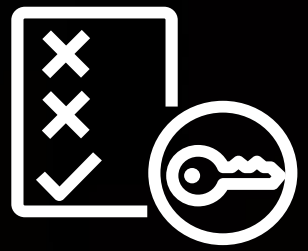
Amazon S3 Object Lock

aws

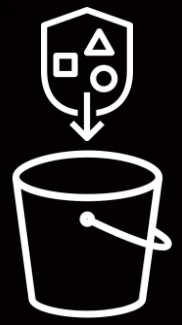
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Secure your data lake

DATA STORED IN AMAZON S3 IS DESIGNED TO BE SECURE BY DEFAULT



Access control with IAM and bucket policy



Amazon S3 Access Points



Amazon S3 data encryption



Amazon S3 object tagging



Amazon S3 Access Analyzer



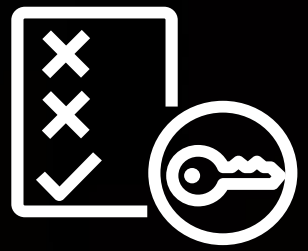
Amazon S3 Object Lock

aws

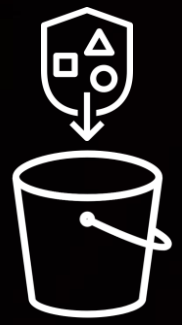
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Secure your data lake

DATA STORED IN AMAZON S3 IS DESIGNED TO BE SECURE BY DEFAULT



Access control with IAM and bucket policy



Amazon S3 Access Points



Amazon S3 data encryption



Amazon S3 object tagging



Amazon S3 Access Analyzer



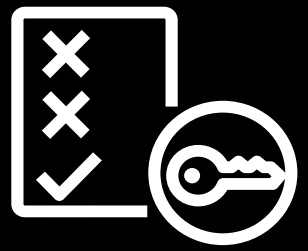
Amazon S3 Object Lock

aws

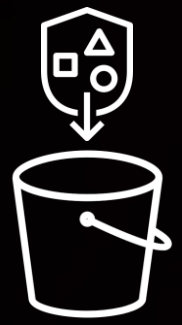
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Secure your data lake

DATA STORED IN AMAZON S3 IS DESIGNED TO BE SECURE BY DEFAULT



Access control with IAM and bucket policy



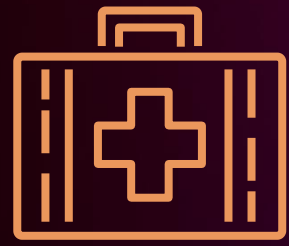
Amazon S3 Access Points



Amazon S3 data encryption



Amazon S3 object tagging



Amazon S3 Access Analyzer



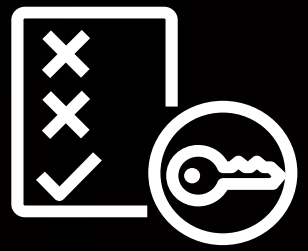
Amazon S3 Object Lock

aws

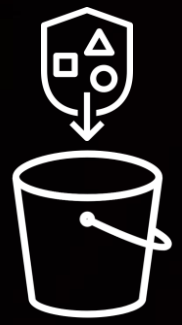
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Secure your data lake

DATA STORED IN AMAZON S3 IS DESIGNED TO BE SECURE BY DEFAULT



Access control with IAM and bucket policy



Amazon S3 Access Points



Amazon S3 data encryption



Amazon S3 object tagging



Amazon S3 Access Analyzer

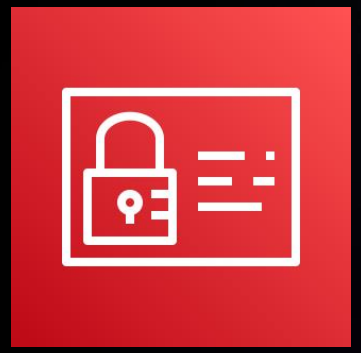


Amazon S3 Object Lock

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Additional services for S3 security



AWS IAM



Amazon Cognito



AWS Lake Formation

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

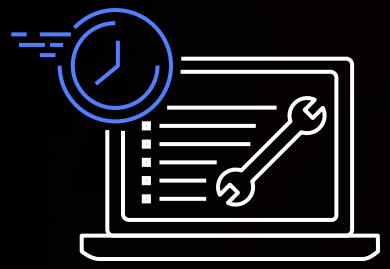
How do I cost optimize?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data lake cost optimization

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Choose a storage class for your data



Access pattern



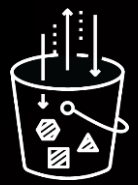
Duration of storage

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Choose a storage class for your data



S3 Intelligent-Tiering



S3 Standard



S3 Standard-IA



S3 Glacier Instant Retrieval



S3 Glacier Flexible Retrieval (formerly S3 Glacier)



S3 Glacier Deep Archive



S3 One Zone-IA



S3 Outposts

AWS Region ≥ 3 Availability Zones

Changing access patterns

- Milliseconds access
- No retrieval charge
- **Archive Instant Access tier**

Frequently accessed data

- Milliseconds access
- No retrieval charge

Infrequently accessed data

- Milliseconds access
- Per-GB retrieval charge

Rarely accessed data

- Milliseconds access
- Per-GB retrieval charge

Archive data

- Retrieval options from minutes to hours
- **Free bulk retrieval**

Long-term archive data

- Retrieval in hours

AWS AZ

Re-creatable, infrequently accessed data

- Milliseconds access
- Per-GB retrieval charge

AWS Outposts

On-premises data

- Milliseconds access

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Match the use case

Changing data access patterns, no retrieval charge, millisecond access

Frequent data access patterns, no retrieval charge, millisecond access

Infrequent access patterns, per-GB retrieval charge, re-creatable data

Rare data access patterns, per-GB retrieval charge, millisecond access

Archive data access patterns, no **bulk** retrieval charge, min to hrs access

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Choose a storage class for your data

Raw, pre-processed, or processed data (non-production)



Amazon S3 Standard or S3 Intelligent-Tiering

Production data lake



Amazon S3 Intelligent-Tiering

Historical data



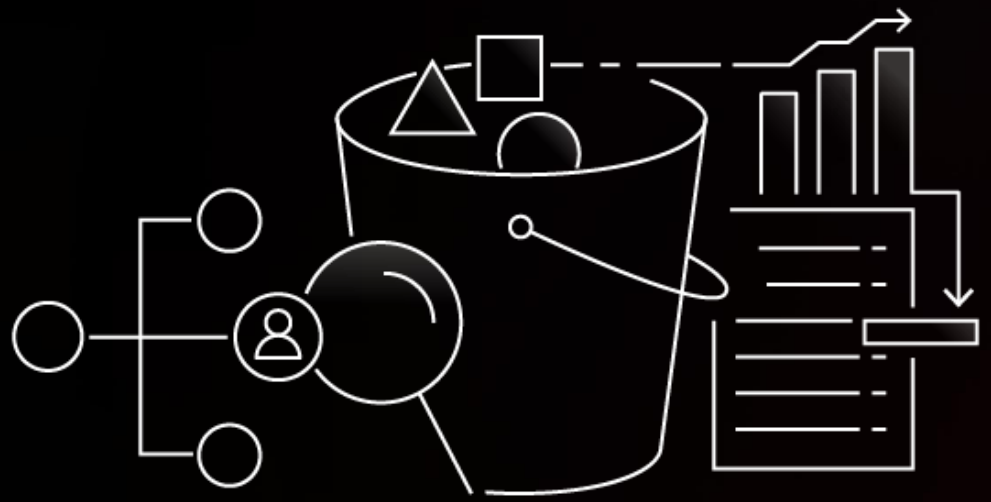
Amazon S3 Glacier Flexible Retrieval, S3 Glacier Deep Archive, or S3 Intelligent-Tiering

Optimize storage cost and performance for all stages of a data lake workflow

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 Storage Lens



Organization-wide visibility

29+ storage usage & activity metrics

Drill down by Region, storage class, bucket, and prefix

Identify incomplete multi-part uploads

Discover ”cold” buckets

Amazon CloudWatch publishing

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Size and formats matter

SIZE

Aim for large objects that are 128MB+

PARTITION

Things typically used while querying

FORMAT

Carefully choose your formats

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How do I monitor and manage performance?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Manage & monitor your data lake

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Manage your data lake

HIGH AVAILABILITY & DISASTER RECOVERY

What do I want to prevent or allow?	How do I achieve this?
Prevent data loss	Inherent durability with multiple AZs
Prevent accidental deletion or corruption	S3 Versioning
Allow data copies in different regions	S3 Cross-Region Replication
Prevent updates after writing	S3 Object Lock

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Monitor your data lake



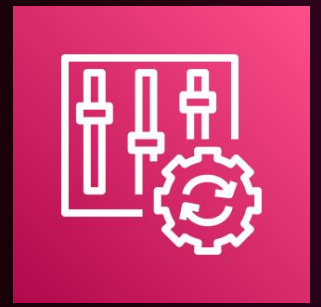
Amazon CloudWatch



AWS CloudTrail



Amazon Macie



AWS Config



AWS Trusted Advisor

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Monitor your data lake



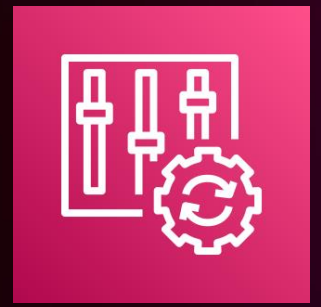
Amazon CloudWatch



AWS CloudTrail



Amazon Macie



AWS Config



AWS Trusted Advisor

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Monitor your data lake



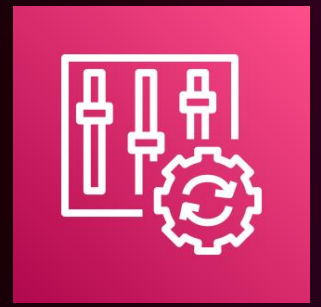
Amazon CloudWatch



AWS CloudTrail



Amazon Macie



AWS Config



AWS Trusted Advisor

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Winston Nwanne
nwannw@amazon.com

Christine Samson
chrsamso@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.