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

STG334-R

Build data mesh architectures using AWS Transfer Family & AWS Lake Formation

Blayze Stefaniak (he/him)
Senior Solutions Architect
AWS

Ken Howard (he/him)
Senior Solutions Architect
AWS

Shoeb Bustani (he/him)
Senior Solutions Architect
AWS

Kapil Shardha (he/him)
Principal Solutions Architect
AWS

Basiruddin Seikh (he/him)
Senior Storage Solutions Architect
AWS

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

Agenda

Why choose a modern data architecture?

Unified governance of data products by using AWS Lake Formation

Third-party data products with AWS Transfer Family managed workflows

Introduction to the session use case

Let’s start building!

Conclusion

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

Why choose a modern data architecture?

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

Customers want more value from their data



Growing exponentially



From new sources



Increasingly diverse



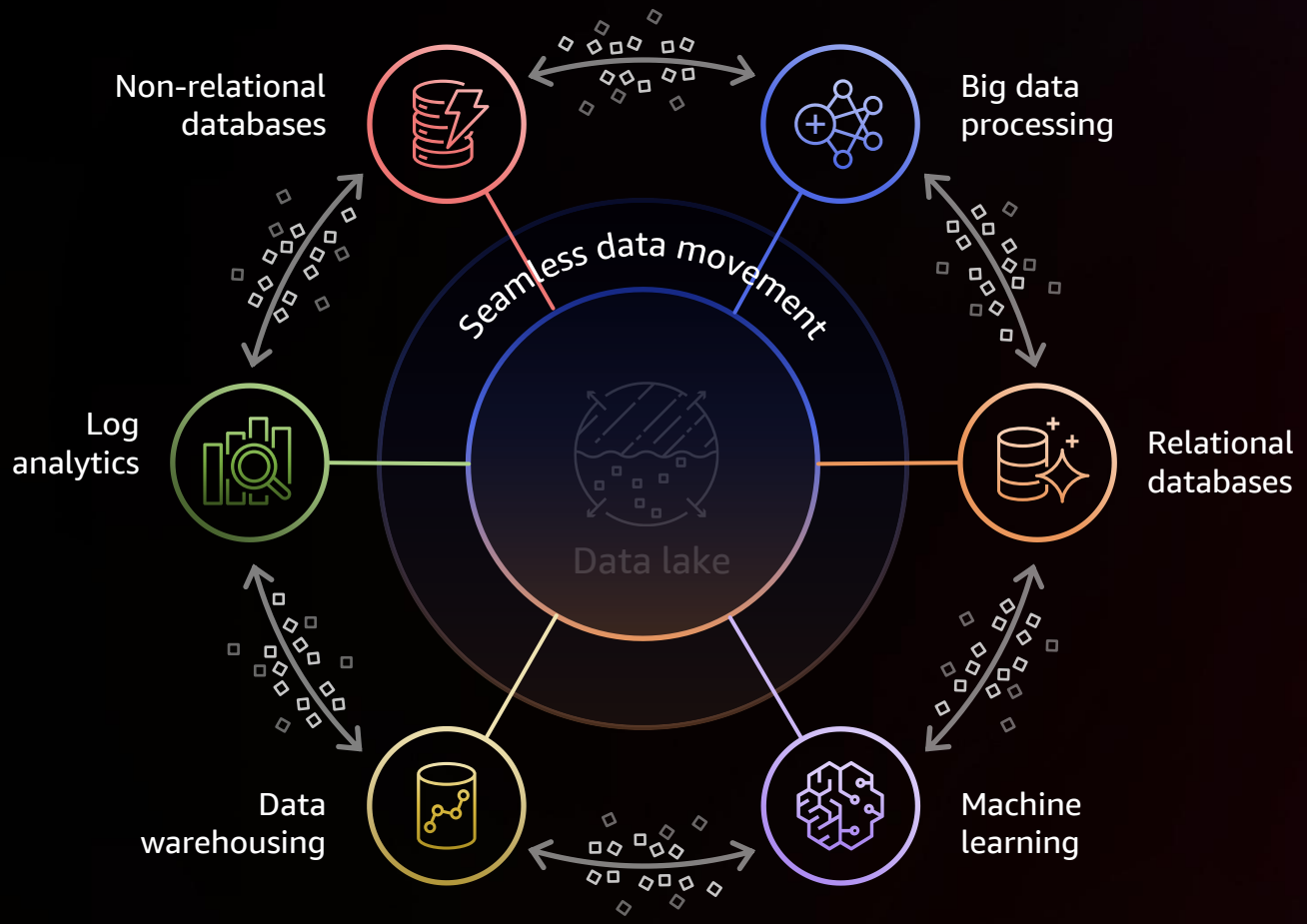
Used by many people



Analyzed by many applications

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

Key pillars of the modern data architecture



- Scalable data lake
- Purpose-built data services
- Seamless data movement
- Performant and cost effective
- Unified governance

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

Unified governance of data products by using AWS Lake Formation

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

Data mesh: Governed sharing of data products

Architecture that enables the **governed sharing** of **data products** across one or more data lakes



Data producers | Data mesh | Data consumers

aws

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

Third-party data products with AWS Transfer Family managed workflows

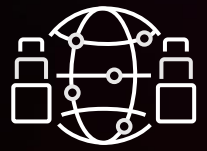
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Transfer Family

FRICTIONLESS FILE TRANSFERS ENABLING BUSINESS-TO-BUSINESS (B2B) CONNECTIVITY



Seamless migration of file transfers using SFTP, FTPS, and FTP



Fully managed, highly available across the globe



Secure and compliant



Native integration with AWS services



Cost effective



Simple to use

aws

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

FINRA uses AWS Transfer Family to provide secure access to their data lake on Amazon S3



Challenge

FINRA manages SFTP infrastructure for over 1,000 external users doing 6 million file transfers per month. This put a significant strain on their operations teams to manage and monitor.

Solution

FINRA has embraced using Amazon S3 as their data lake. As part of building their new FileX platform on AWS, FINRA used the AWS Transfer Family to provide their external users secure access to data stored in Amazon S3.

Benefits

By using AWS Transfer Family, FINRA was able to reduce operational burden, continue to use Amazon S3 as their data lake, and maintain connections to their existing authentication systems.

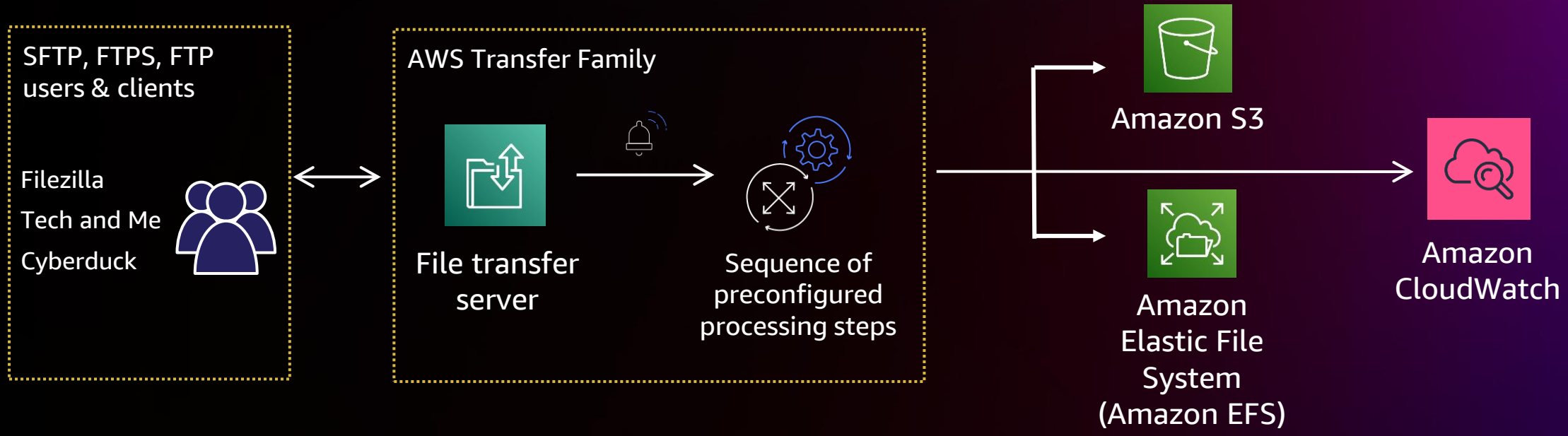
> "AWS Transfer for SFTP is helping us reduce our operational burden, keeping S3 as our data lake, while maintaining critical connection to our existing authentication systems for external users – so they don't experience any disruption as we complete our migration of SFTP services to AWS."
>
> – **Ranga Rajagopal**, Senior Director, Enterprise Data Platforms, FINRA

aws

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

Managed workflows for secure file transfers

AWS Transfer Family managed workflows for post-upload processing



Supported features

- Console/API/Amazon CloudFormation for defining steps
- Copy/tag/delete
- Custom option for Lambda invocation
- Triggered only on completed uploads for data quality
- Exception handler mode to catch erroneous data
- Metrics/logs consolidated in CloudWatch for auditing

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

Builders’ session use case

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

AWS Transfer Family data product personas

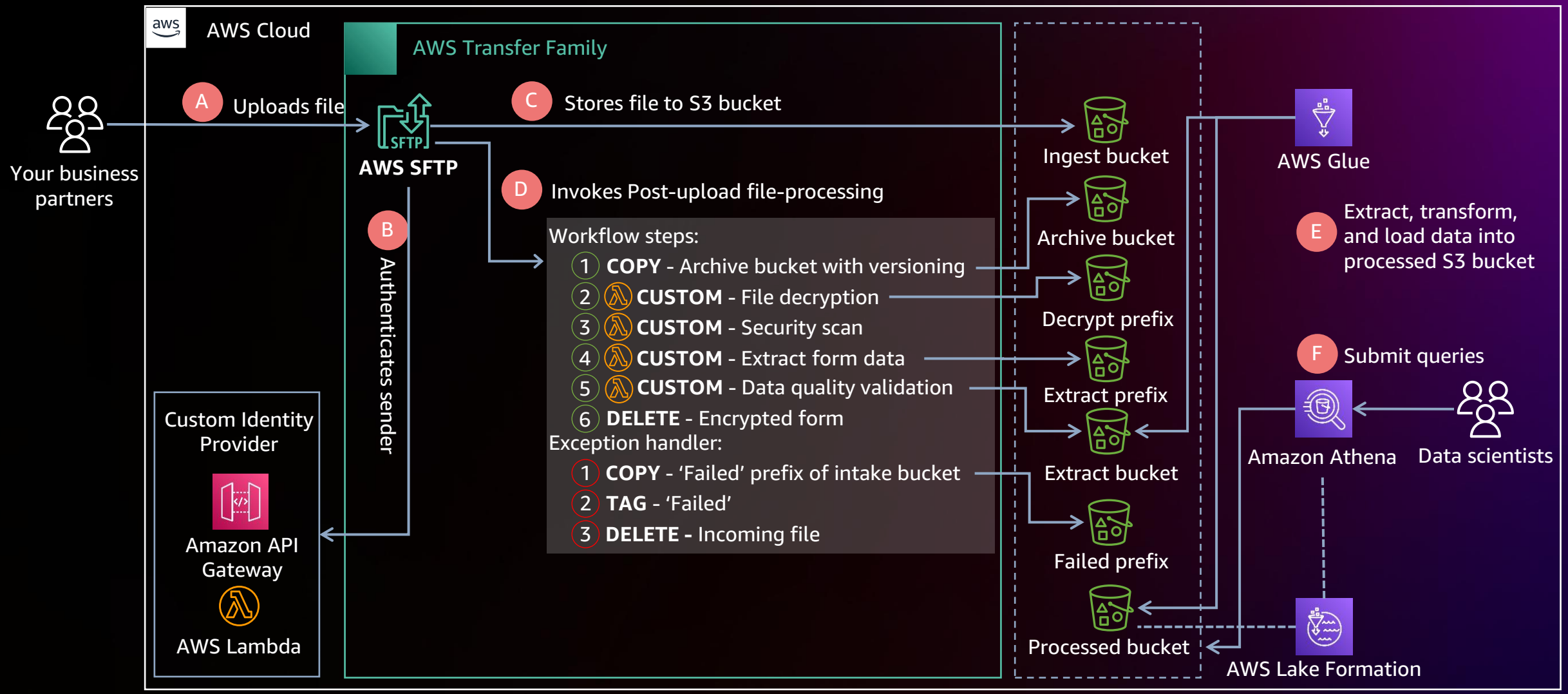
YOU VALIDATE FORMS YOU RECEIVE FROM THIRD PARTIES AND MAKE THEM AVAILABLE TO YOUR CONSUMERS

User/persona	Job duties	AWS service usage
 Ana – tax accountant (third-party business partner)	Provides tax services to the population	**Submits W2 form data over Mateo's AWS Transfer Family SFTP server**
 Mateo – IT operations (data product owner)	Validates the integrity of files processed by managed file transfer (MFT) ingestion service	**Writes AWS Transfer Family workflow to validate authenticity and data quality of submitted forms**
 Zhang – data engineer (data product owner)	Builds and manages extract, transform, load (ETL) jobs to write raw data into a consumable format	**Authors AWS Glue job, which reads from an Amazon S3 bucket and AWS Glue table, and writes to columnar format on Amazon S3**
 Saanvi – data scientist (data consumer)	Studies income trends across the population to influence data-driven public policy proposals	**Is granted access via AWS Lake Formation, and queries financial data on Amazon S3 with Amazon Athena**

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

AWS Transfer Family data product use case

YOU VALIDATE FORMS YOU RECEIVE FROM THIRD PARTIES AND MAKE THEM AVAILABLE TO YOUR CONSUMERS



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

Let’s start building!

Workshop dependencies have been deployed with infrastructure as code

1. Run lab prerequisite steps from an AWS Cloud9 environment
2. View your AWS Transfer Family managed workflow data pipeline
3. Set up AWS Lake Formation
4. Create an extract, transform, load (ETL) job with AWS Glue Studio
5. Register your data product with AWS Lake Formation, and grant permissions
6. Query your data product with Amazon Athena
7. Upload forms to your data pipeline Transfer Family server, and run Glue job
8. Retrieve processed form data from Athena

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

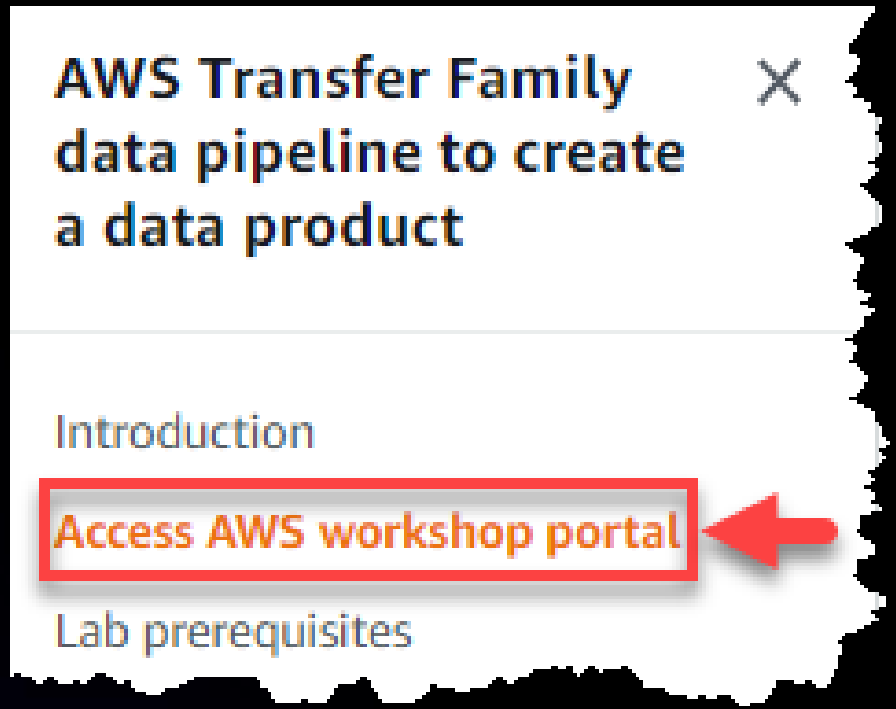
Getting started

Go to https://tinyurl.com/yw64fnxr

Choose **Access AWS workshop portal**

- Choose the Workshop Studio join page link
- Register with your email
- Enter the event hash (below)
- Review Terms and Conditions
- Choose Open AWS Console (bottom left)

Event hash: 8fab-0e26a1-da



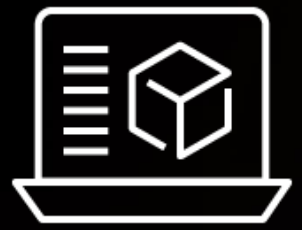
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Conclusion

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

Continue your AWS Storage learning

Build a learning plan



Set your AWS Storage Learning Plans via **AWS Skill Builder**

Increase your knowledge



Use our **Ramp-Up Guides** to build your storage knowledge

Earn AWS Storage badges



Demonstrate your knowledge by earning **digital badges**

aws.training/storage

aws

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

Thank you!

Blayze Stefaniak Ken Howard Shoeb Bustani Kapil Shardha Basiruddin Seikh



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

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