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

ANT322-R

Simplifying ETL migration and data integration with AWS Glue

Andy Bette

Principal Product Manager
AWS Glue

Zach Mitchell

Senior Big Data Architect
AWS Glue

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

Agenda

ETL migration is a key part of cloud data migration

This interactive session will discuss how AWS Glue eases the transition

1. Connect to outside systems
2. Migrate ETL logic
3. Handle special data formats
4. Development tools
5. Q&A and discussion



aws

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

“AWS Glue’s native integration with other AWS Cloud services makes it the most suitable choice when selecting an AWS-native data ecosystem.”

Gartner, Inc.

Magic Quadrant for Data Integration Tools

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

Connect to outside systems

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

Connect with AWS Glue

80+ connectors for scalable and simple-to-use data movement



Built-in connectors

Out-of-box connectors **managed by** AWS Glue



Custom connectors

Flexibility to build domain-specific data access



Marketplace connectors

Broad range of subscription-based, **third-party options**

aws

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

Other ways to connect

Flexibility through Spark:
AWS Glue doesn't require you to use our libraries

Imports

AWS SDK for pandas (awswrangler)

Spark connector libraries:*

- spark-redshift
- spark-snowflake
- Excel
- Elasticsearch

External engines

AWS Lambda

Amazon S3 APIs

SFTP into Amazon S3

Amazon Kinesis Data Firehose

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

*This list is just a few examples

Migrate ETL logic

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

Partner-built converters to migrate from legacy ETL tools



"BladeBridge ETL converters use AWS Glue Studio APIs to convert legacy ETL code to AWS Glue Studio objects that can be **easily managed and maintained by non-Spark developers**."

Troy Clemente, Executive VP & Co-founder, BladeBridge



"Impetus LeapLogic is the leading migration automation solution for ETL, EDW, and Hadoop. We provide a modern, efficient, and fast approach to **migrate to AWS Glue** and power up a unified data platform that **better enables data scientists to provide value through data**."

Chetan Kalanki, VP of Global Engineering Services, Impetus

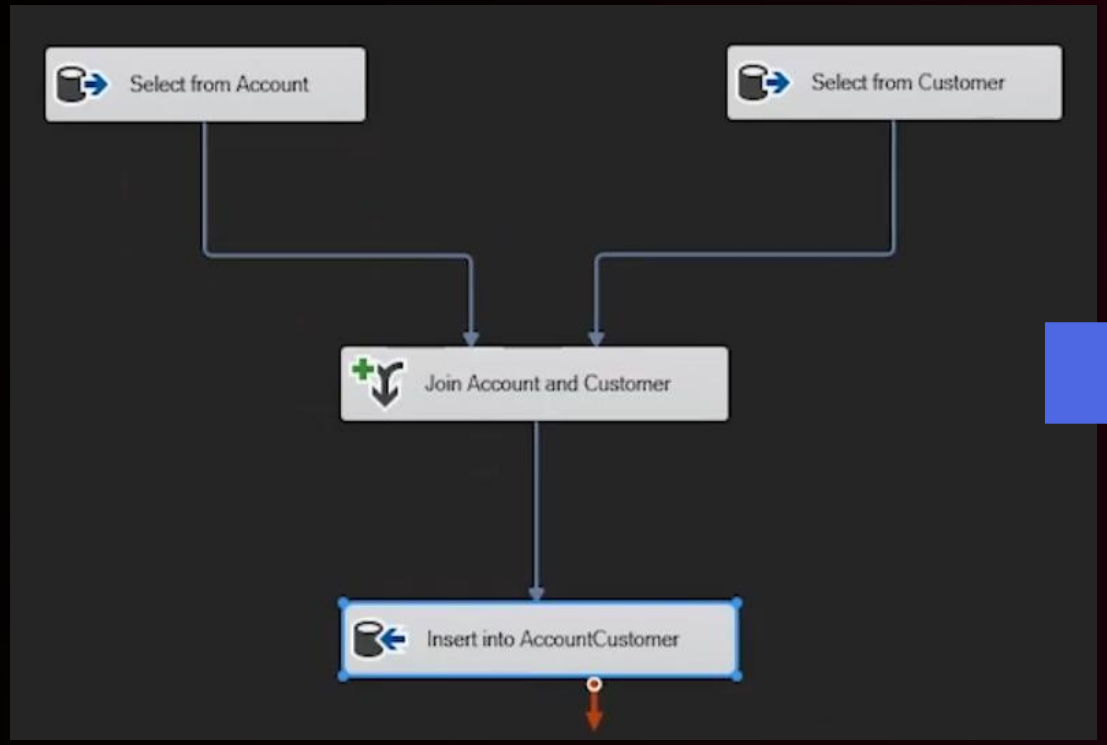
aws

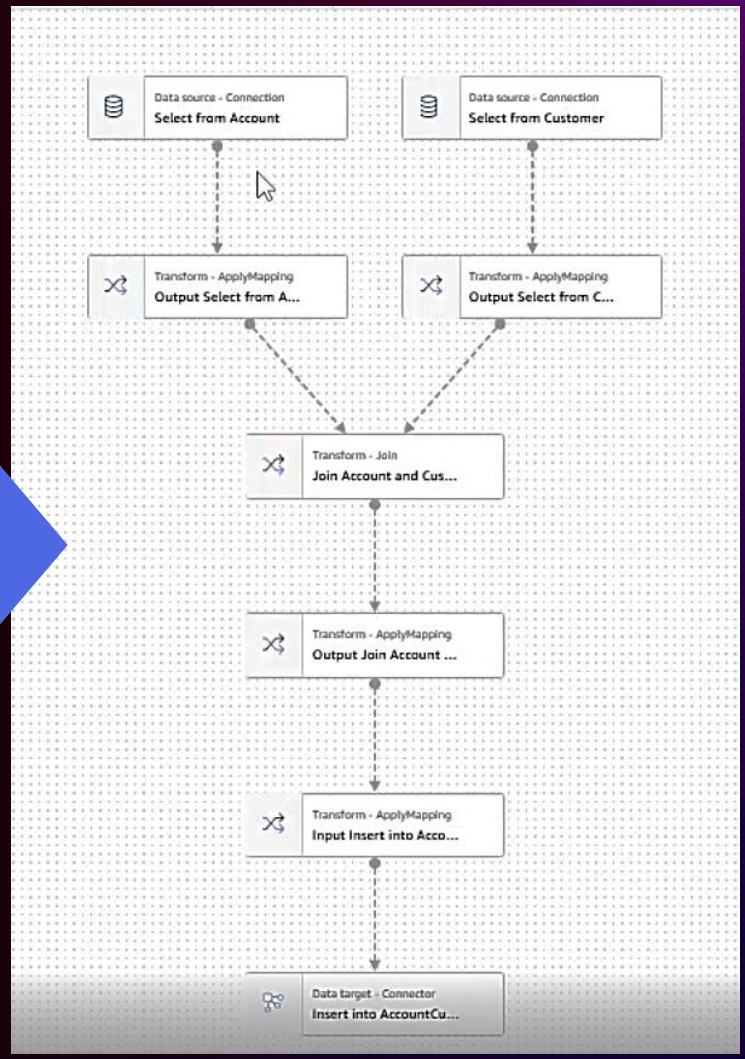
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Migrate your ETL logic

Automated translation tools convert 80% of your ETL jobs

Example: Conversion from SSIS using AWS Schema Conversion Tool





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

Video: Microsoft SSIS to AWS Glue conversion using AWS Schema Conversion Tool



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

Migrate your ETL logic

ETL Migration Program: Get help migrating your ETL processes to AWS Glue

Converters available for

Informatica

Talend

SAS

Microsoft SSIS

IBM DataStage

Dozens of SI Partners help migrate your ETL pipelines

AGILISIUM
BIG ON CLOUD. BIG ON DATA.

mission™
CLOUD SERVICES

quantiphi
Solving What Matters

slalom

wavicle
DATA SOLUTIONS

rackspace
technology

TEKsystems
Global Services

This is not a complete list. To view all AWS Partners for this category, visit AWS Partner Solutions Finder. This list of partners is current as of November 27, 2022.

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

Migrate your ETL code

AWS Glue jobs support direct migration:

- Vanilla Apache Spark
  - Scala
  - PySpark
- Vanilla Python
  - C-based Python libraries
- pandas



Scala





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

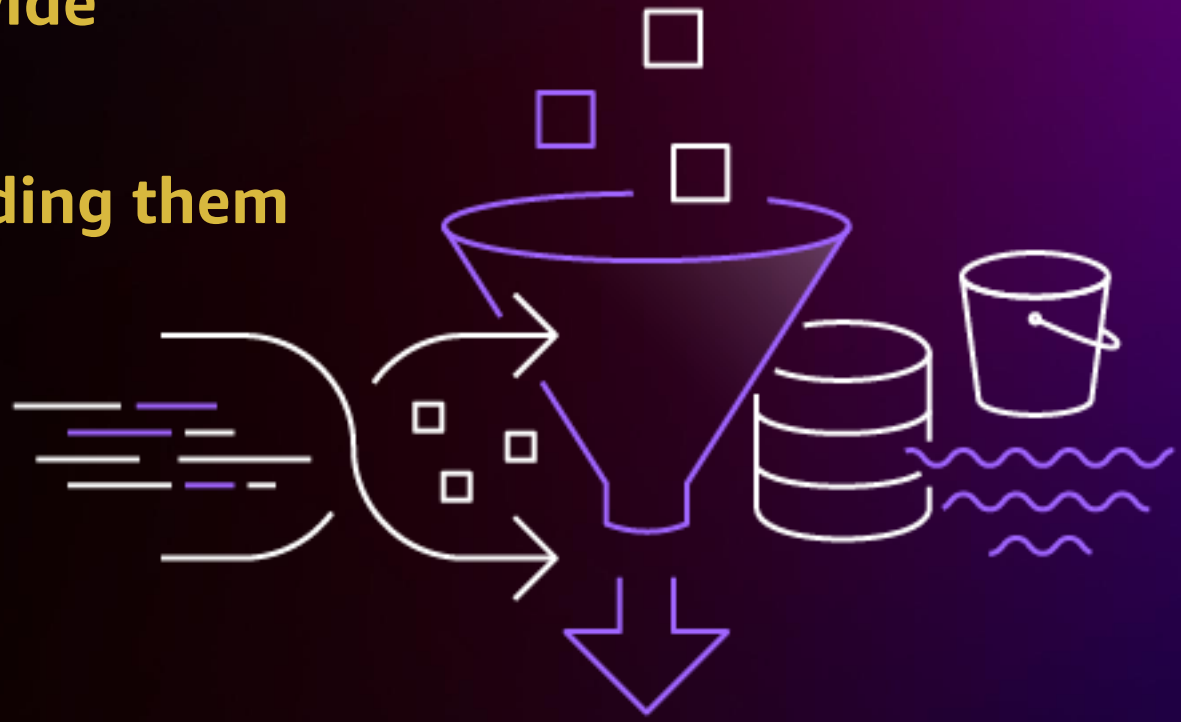
Handle special data formats

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

Handle special data formats

Your existing systems produce data in a wide variety of formats

AWS Glue has a variety of options for reading them



Built-in format support

- Structured: Parquet, ORC, AVRO
- Semi-structured: JSON, CSV, XML
- Custom classifiers for complex CSV, JSON, and more
- Grok patterns for other text structures – e.g., mainframe formats

aws

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

Development tools

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

Tools to develop and test

Local and on-premises

- Install the AWS Glue libraries
- Run our Docker containers

Hybrid

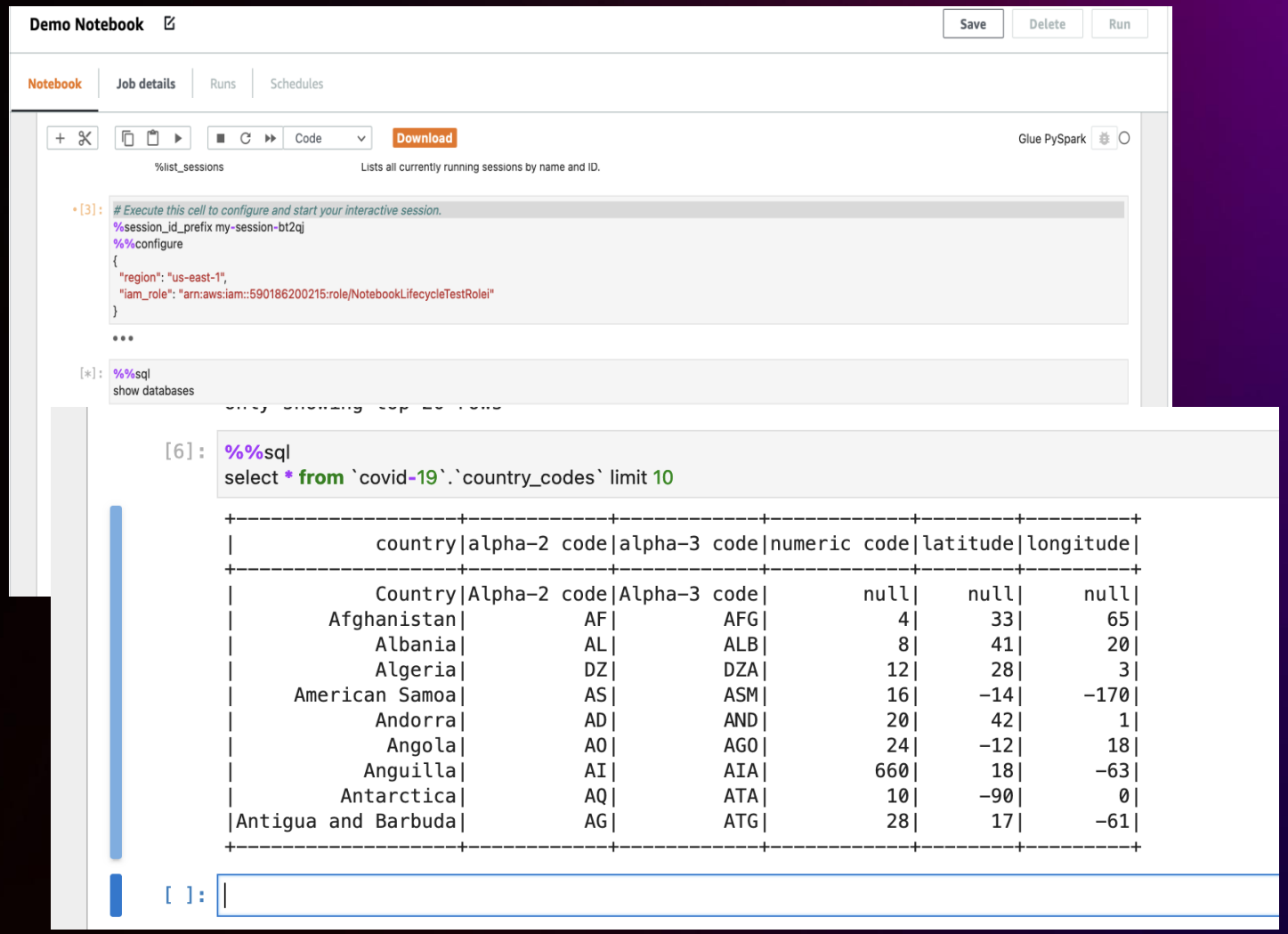
- Connect your notebook or IDE to AWS Glue interactive sessions*

Cloud

- Built-in notebook, code editor, and visual ETL interface

Deployment

- Built-in Git integration



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

*Interactive sessions replaced Dev Endpoints – use them instead!

Thank you!



Contact us through your account manager



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

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