

ANT002

Discover, prepare, and integrate your data with AWS Glue

Sean Ma

Principal Product Manager

Noritaka Sekiyama

Principal Big Data Architect

© 2022, Amazon Web Services, Inc. or its affiliates.

Agenda

Introduction to AWS Glue

How customers use AWS Glue

Demo: Enrich and cleanse incremental data for analytics

Demo: Interactive feature engineering for machine learning

© 2022, Amazon Web Services, Inc. or its affiliates.



AWS Glue

Discover, prepare, and integrate all your data at any scale



All-in-one data integration service



Cost effective, serverless, and scalable



Tailored tools to support all data users



Support all workloads in one place

© 2022, Amazon Web Services, Inc. or its affiliates.

Customers

Hundreds of thousands of AWS Glue customers

Financial services











Software & internet











Healthcare











Manufacturing











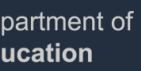
Government













aws

© 2022, Amazon Web Services, Inc. or its affiliates.

How customers use AWS Glue

COMMON USE CASES

Migrate from traditional ETL

Gain flexibility and reduce cost

Load and process on data lakes

Easiest way to load and transform data

Self service data integration

Allow various users in the company have access to data

ETL/ELT from data warehouse

Clean, blend and secure data to share insights

Leverage powerful Apache Spark engine

Process PB of data without the maintenance burden

Data preparation for ML

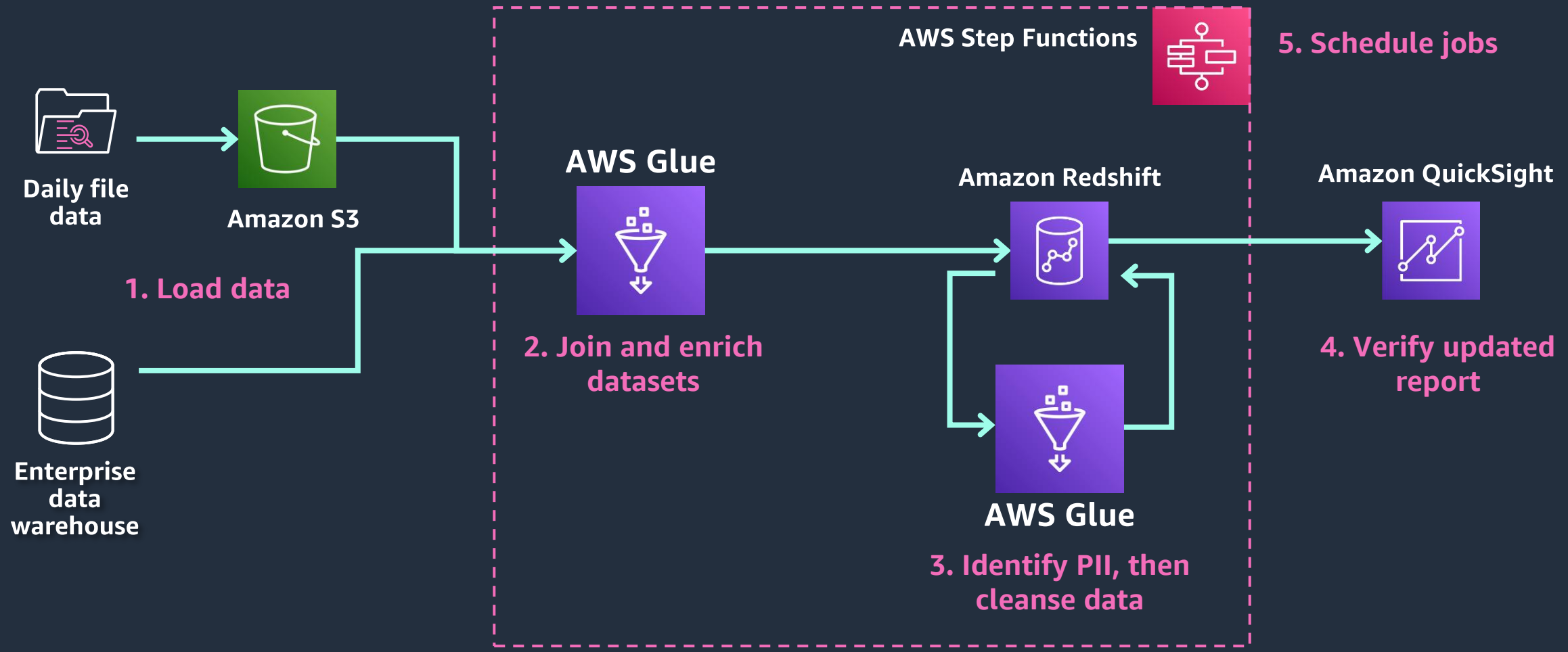
Flexible tools for interactive data exploration

© 2022, Amazon Web Services, Inc. or its affiliates.

Enrich and cleanse incremental data for analytics

© 2022, Amazon Web Services, Inc. or its affiliates.

Enrich and cleanse incremental data for Analytics



© 2022, Amazon Web Services, Inc. or its affiliates.

AWS Glue Studio

Jobs
Monitoring
Connectors
Sensitive data detection
What's new

Glue console
Glue catalog
Crawlers
Security configurations
Interactive sessions

Marketplace
Documentation

Analytics

AWS Glue Studio

Visually create job flows and monitor their performance

AWS Glue Studio is an easy-to-use graphical interface for creating, running, and monitoring AWS Glue extract, transform, and load (ETL) jobs.

Getting started



Create and manage jobs

Visually author, run, view, and edit your AWS Glue jobs.

View jobs



Monitor job runs

Diagnose, debug, and check the status of your AWS Glue jobs.

View monitoring



Connect external data stores

Create and manage connections to external data stores for AWS Glue Studio.

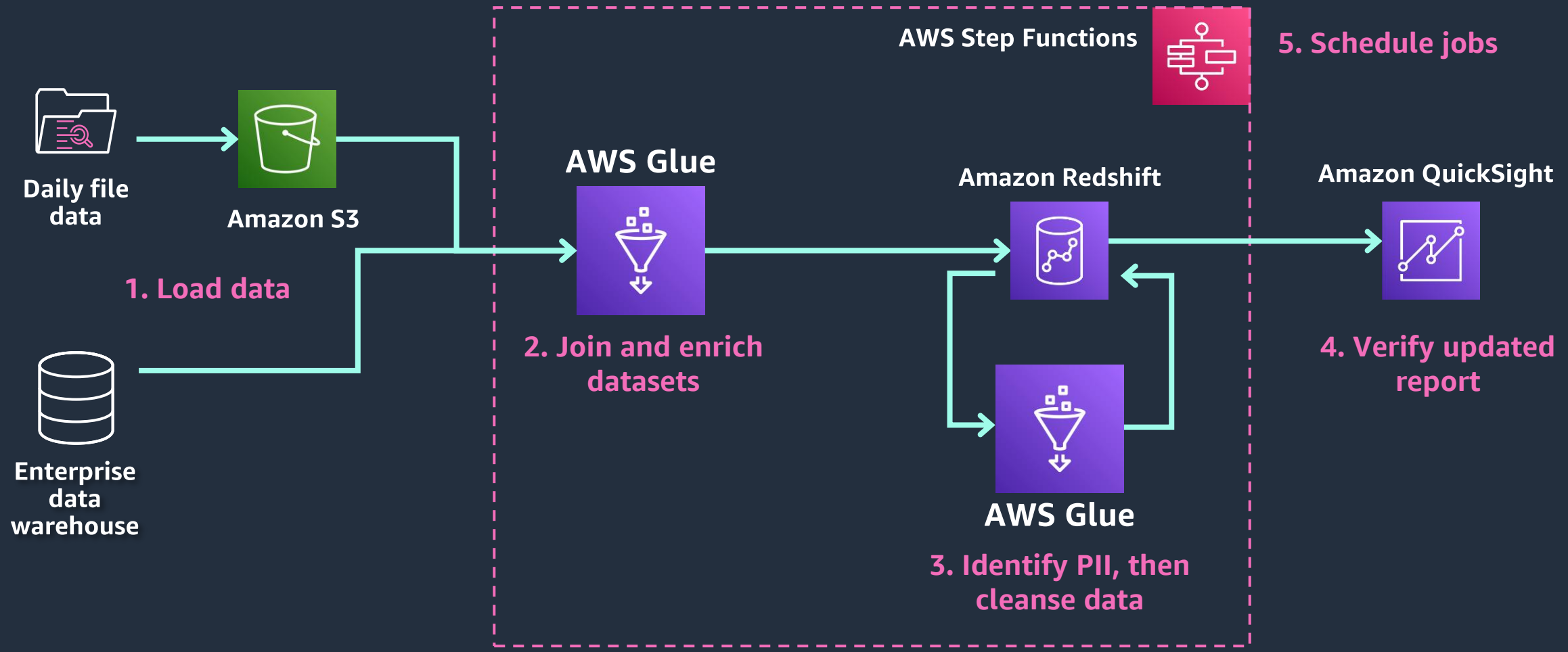
View connectors

Tutorials

What is AWS Glue Studio?
Getting started with Glue Studio
Video: Working with AWS Glue Studio, Part 1
Video: Working with AWS Glue Studio, Part 2
Video: Working with AWS Glue Studio, Part 3
Using the open-source elasticsearch Spark Connector
Adding an AWS Glue Crawler
Using connectors and connections

© 2022, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Enrich and cleanse incremental data for Analytics



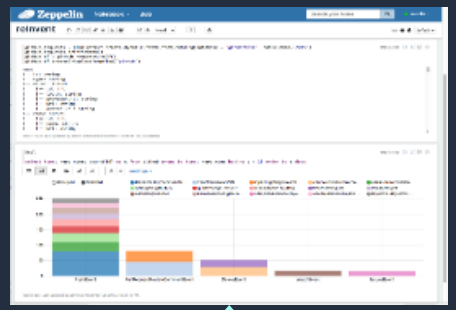
© 2022, Amazon Web Services, Inc. or its affiliates.

Large-scale feature engineering for machine learning

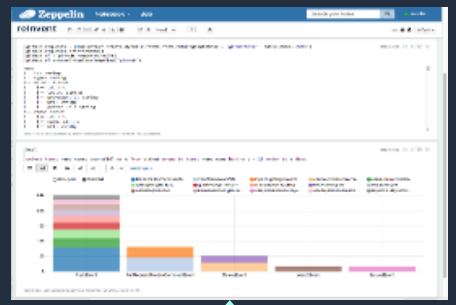
© 2022, Amazon Web Services, Inc. or its affiliates.

Large scale feature engineering for machine learning

2. Explore and cleanse data with Glue Studio Notebooks



4. Extract key features for model training



1. Ingest raw data

Logs, app data

AWS Glue

Cleaned data

Amazon S3

AWS Glue

Extracted features

Amazon S3

Amazon SageMaker

5. Train and deploy ML model

AWS Lake Formation

3. Set up access controls with Lake Formation

© 2022, Amazon Web Services, Inc. or its affiliates.

AWS Glue Studio

Jobs
Monitoring
Connectors
Sensitive data detection
What's new

Glue console
Glue catalog
Crawlers
Security configurations
Interactive sessions

Marketplace
Documentation

Analytics

AWS Glue Studio

Visually create job flows and monitor their performance

AWS Glue Studio is an easy-to-use graphical interface for creating, running, and monitoring AWS Glue extract, transform, and load (ETL) jobs.

Getting started



Create and manage jobs

Visually author, run, view, and edit your AWS Glue jobs.

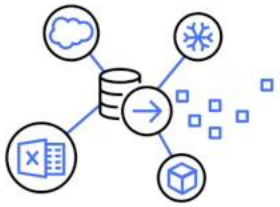
View jobs



Monitor job runs

Diagnose, debug, and check the status of your AWS Glue jobs.

View monitoring



Connect external data stores

Create and manage connections to external data stores for AWS Glue Studio.

View connectors

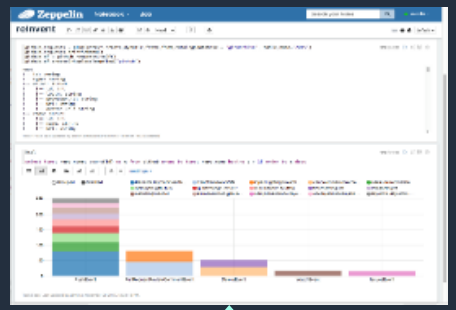
Tutorials

- What is AWS Glue Studio?
- Getting started with Glue Studio
- Video: Working with AWS Glue Studio, Part 1
- Video: Working with AWS Glue Studio, Part 2
- Video: Working with AWS Glue Studio, Part 3
- Using the open-source elasticsearch Spark Connector
- Adding an AWS Glue Crawler
- Using connectors and connections

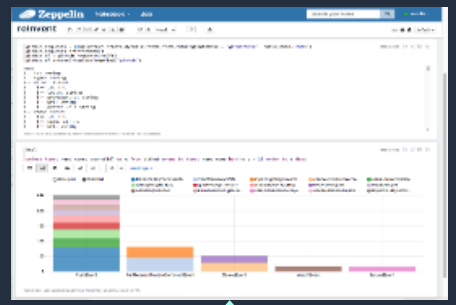
© 2022, Amazon Web Services, Inc. or its affiliates.

Large scale feature engineering for machine learning

2. Explore and cleanse data with Glue Studio Notebooks



4. Extract key features for model training



1. Ingest raw data

Logs, app data

AWS Glue

Cleaned data

Amazon S3

AWS Glue

Extracted features

Amazon S3

Amazon SageMaker

5. Train and deploy ML model

AWS Lake Formation

3. Set up access controls with Lake Formation

© 2022, Amazon Web Services, Inc. or its affiliates.



Thank you!

Sean Ma
seanrma@amazon.com

Noritaka Sekiyama
sekiyama@amazon.com

© 2022, Amazon Web Services, Inc. or its affiliates.