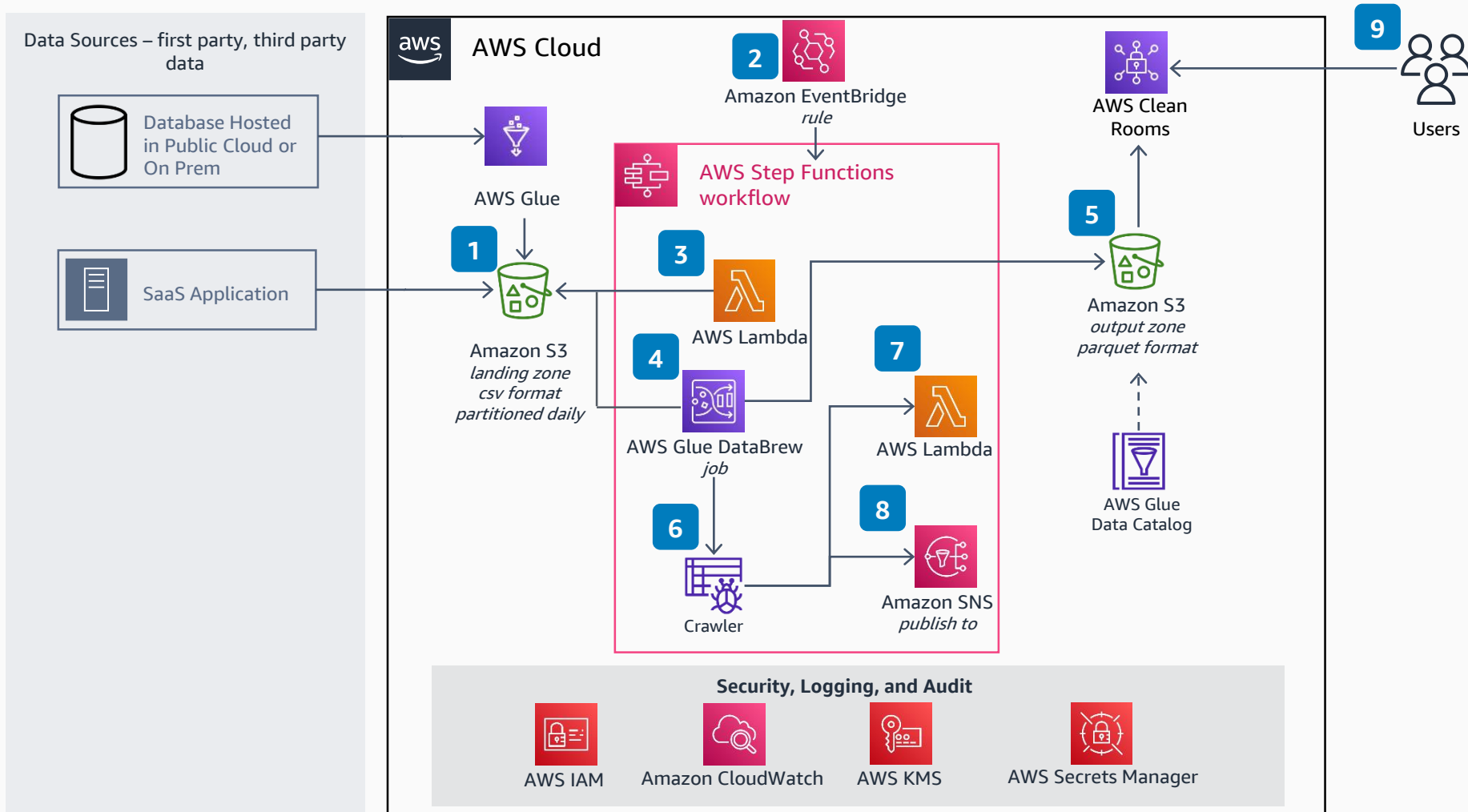


Guidance for Connecting Data to AWS Clean Rooms

This Guidance shows how an AWS customer can import data stored from a variety of data stores into an AWS storage service and prepare it for AWS Clean Rooms data collaboration.



1

Data stored in external applications needs to be ingested into **Amazon Simple Storage Service** (Amazon S3). Either export data directly from a SaaS application that supports a native **Amazon S3** connector. Use an **AWS Glue** extract, transform, and load (ETL) service to pull data from relational databases.

2

Create a rule in **Amazon EventBridge** to schedule the data processing in **AWS Step Functions**. The function includes data ingestion and downstream processing steps.

3

Use the **AWS Lambda** function to decrypt the files from the source **Amazon S3** bucket using **AWS Key Management Service** (AWS KMS) and place them in a different prefix for **AWS Glue DataBrew** to pick up and process.

4

Use **AWS Glue DataBrew** recipe to transform the data from the decrypted source **Amazon S3** location. Use this step to normalize, and secure Personal Identifiable Information (PII) data using the SHA256 hashing algorithm.

5

The output of the **AWS Glue DataBrew** recipe is written to the target **Amazon S3** bucket:prefix location in parquet format.

6

An **AWS Glue Crawler** job is initiated to "refresh" the table definition and its associated meta-data.

7

After the **AWS Glue Crawler** job concludes, a **Lambda** function moves the source data files to an "archive" prefix location as part of clean-up activity.

8

An event is published to **Amazon Simple Notification Service** (Amazon SNS) to inform the user that the new data files are now available for consumption within **AWS Clean Rooms**.

9

The user can use the latest data within the **AWS Clean Rooms** service to collaborate with other data producers.

Security, Logging, and Audit

The solution uses the following AWS services to promote security and access control:

AWS Identity and Access Management (IAM): Least-privilege access to specific resources and operations

AWS KMS: Provides encryption for data at rest and data in transit (using PGP encryption of data files)

Secrets Manager: Provides hashing keys for PII data

Amazon CloudWatch: Monitors logs and metrics across all services used in this solution



Reviewed for technical accuracy January 12, 2023
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Reference Architecture