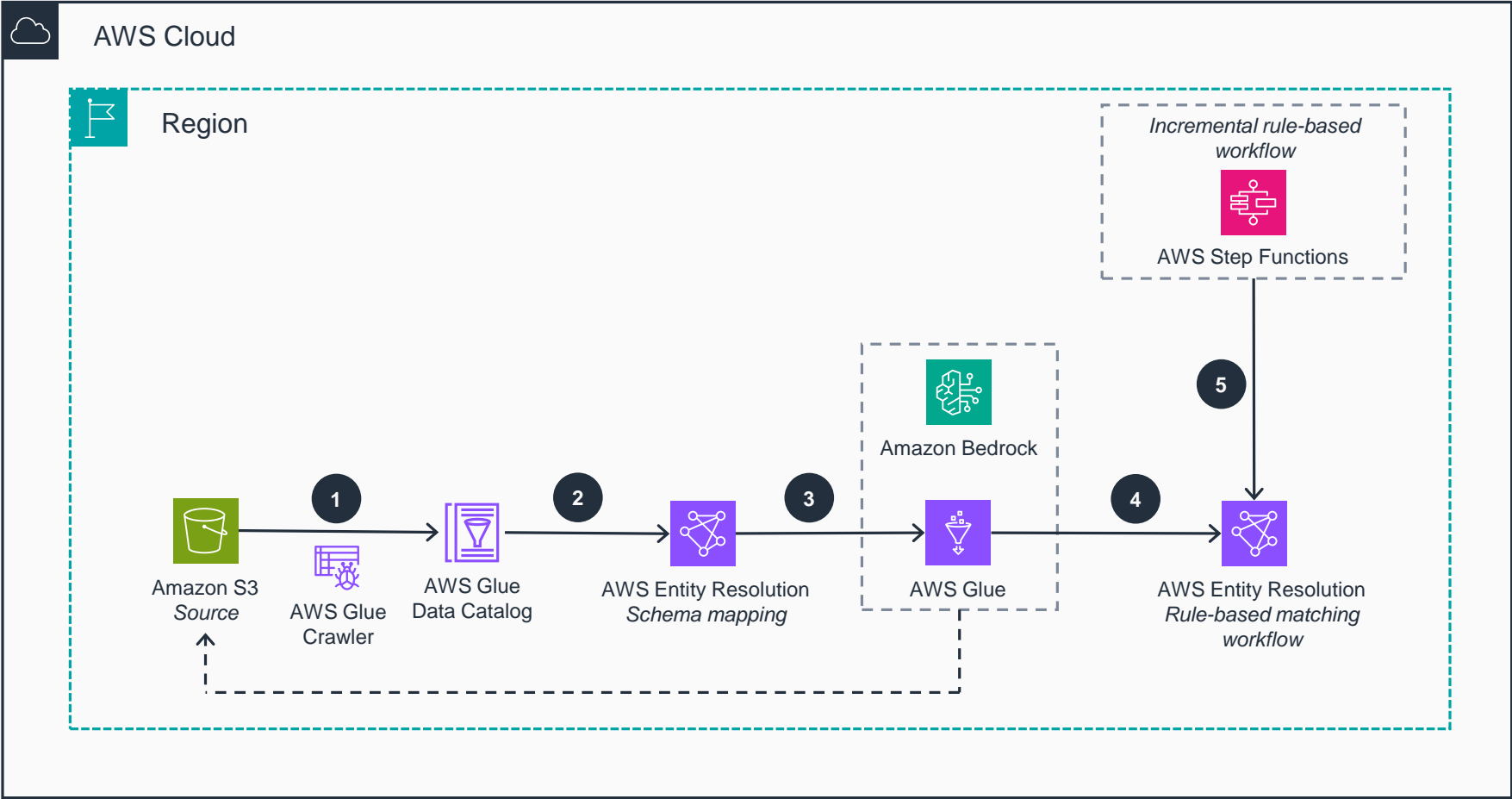


Guidance for Generating Rule Recommendations for Entity Resolution on AWS

Overview

This architecture diagram shows an overview of how to generate rule recommendations using an LLM hosted on Amazon Bedrock and an AWS Glue notebook and how to use these rules in a rule-based matching workflow in AWS Entity Resolution.



1 Load your input dataset (CSV/parquet) in an **Amazon Simple Storage Service (Amazon S3)** bucket and use an **AWS Glue Crawler** to create an **AWS Glue** table within the **AWS Glue Data Catalog**.

2 Create a schema mapping in **AWS Entity Resolution** using the **AWS Glue** table as the source.

3 Run the notebook in **AWS Glue**, which uses the **AWS Entity Resolution** schema mapping to understand the shape of the data. The notebook reads the data from **Amazon S3** and generates data quality metrics. It feeds these metrics to an LLM hosted on **Amazon Bedrock**. The LLM recommends rules to apply to an **AWS Entity Resolution** matching workflow for resolving entities.

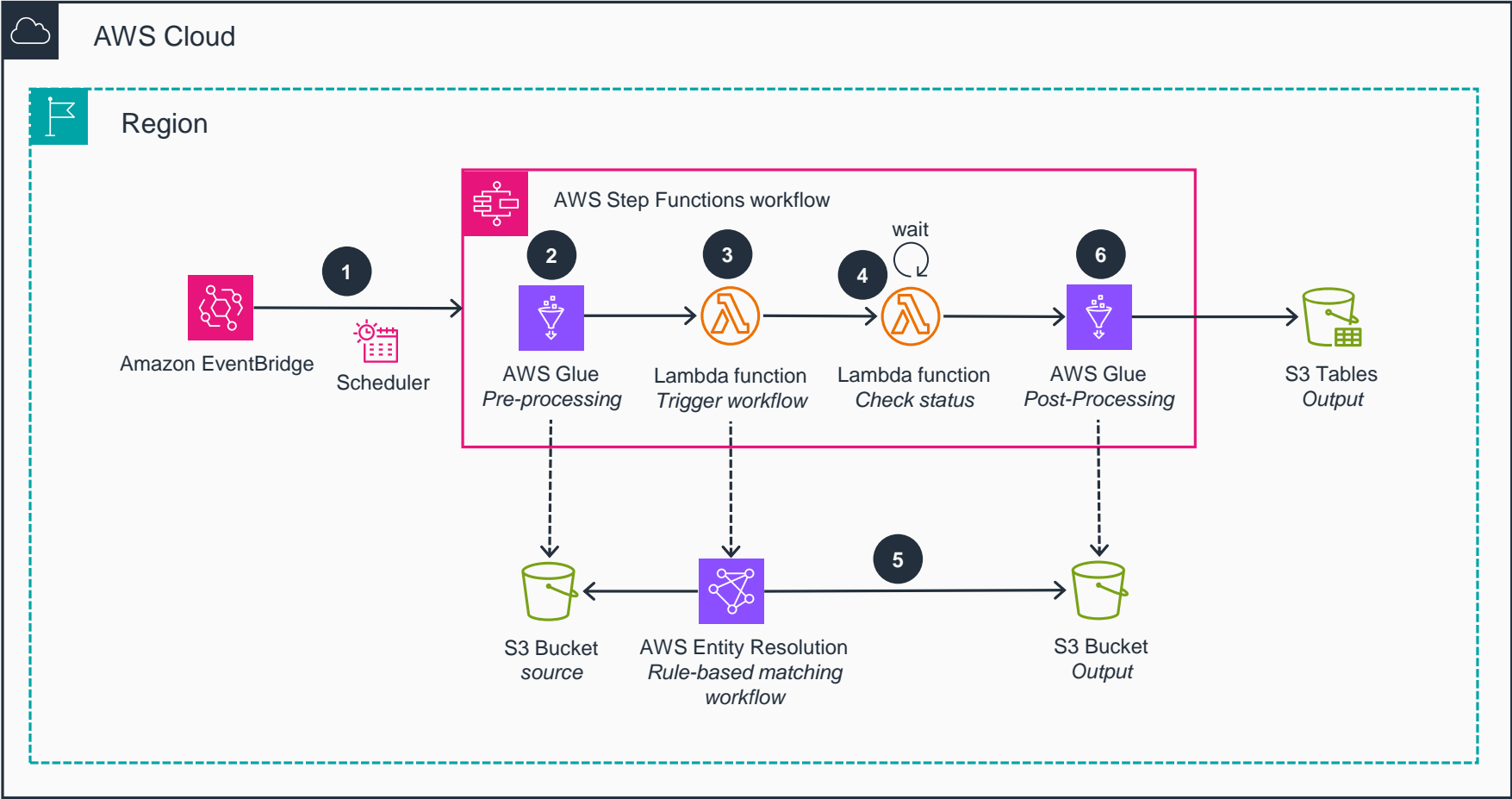
4 The recommended rules generated by the **AWS Glue** notebook are used to create a rule-based matching workflow within **AWS Entity Resolution**.

5 An **AWS Step Functions** workflow orchestrates the execution of the rule-based matching workflow to process the incremental source data.

Guidance for Generating Rule Recommendations for Entity Resolution on AWS

Incremental rule-based workflow

This architecture diagram shows how to run an incremental rule-based matching workflow in AWS Entity Resolution using an AWS Step Functions workflow.



- 1** Create a schedule in **Amazon EventBridge** to trigger **Step Functions** at a desired frequency.
- 2** **Step Functions** triggers an **AWS Glue** extract, transform, load (ETL) job that pre-processes the incremental source data and prepares it for **AWS Entity Resolution** rule-based matching workflow.
- 3** An **AWS Lambda** function triggers the rule-based matching workflow in **AWS Entity Resolution**. The workflow reads the incremental data from the source **Amazon S3** bucket and processes it.
- 4** The **Lambda** function checks the status of the matching workflow running in **AWS Entity Resolution** until the job status changes to *Completed*.
- 5** Upon completion, the **AWS Entity Resolution** matching workflow writes the output to an **S3** output bucket.
- 6** The **AWS Glue** post-processing ETL job reads the output from **AWS Entity Resolution** and writes it to an **Amazon S3** table. The **Amazon S3** table is chosen as the destination because it supports Atomicity, Consistency, Isolation, Durability (ACID) transactions.
- 7** The **AWS Entity Resolution** incremental matching workflow has the capability to merge or split records. Given this ability, a datastore that supports ACID transactions is an ideal choice to help ensure data integrity and consistency.

