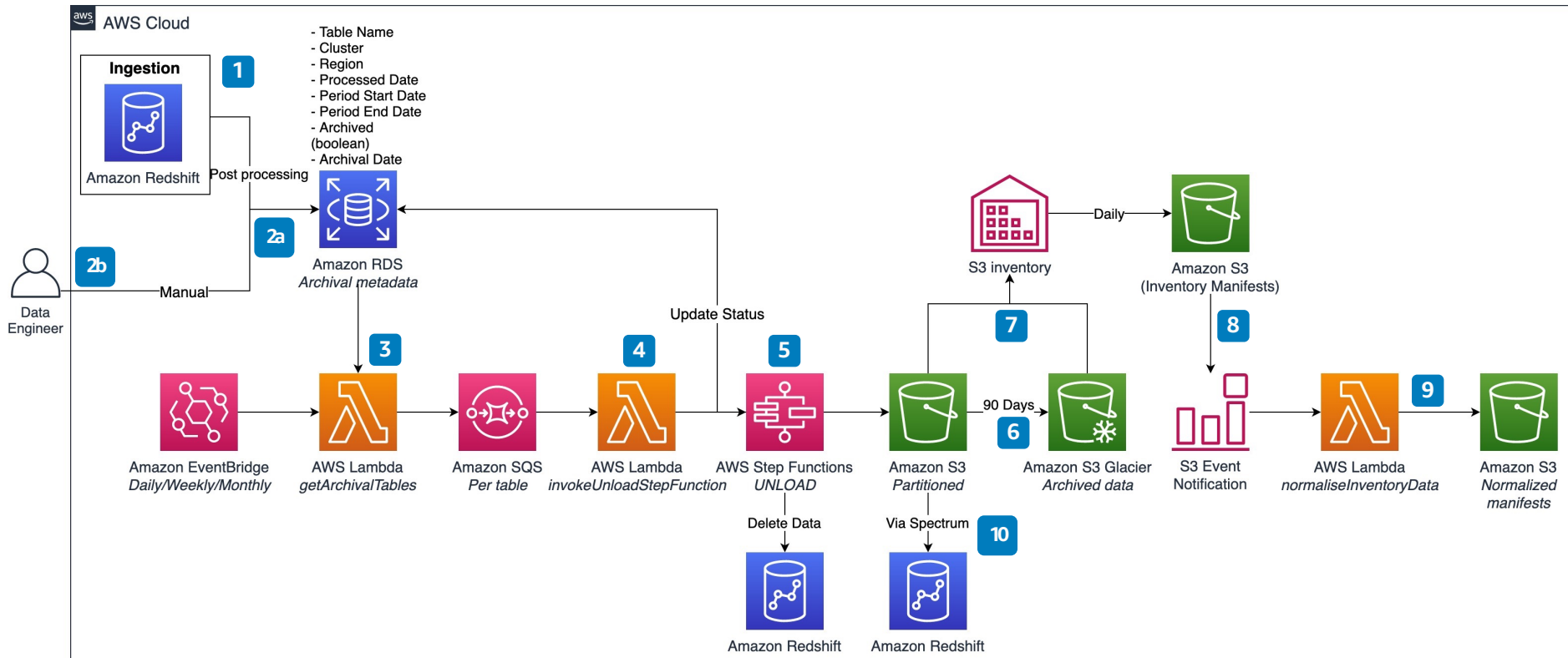


Automated Archival for Amazon Redshift

This architecture automates the periodic data archival process for an Amazon Redshift database. The cold data is available instantly and can be joined with existing datasets in the Amazon Redshift cluster using Amazon Redshift Spectrum.



- 1 Data is ingested into **Amazon Redshift** cluster at various frequencies.
- 2a After every ingestion load, the process creates a queue of metadata about tables populated into **Amazon Redshift** tables in **Amazon RDS**.
- 2b **Data Engineers** may also create the archival queue manually, if needed.
- 3 Using **Amazon EventBridge**, an **AWS Lambda** function is triggered periodically to read the queue from the RDS table and create an **Amazon SQS** message for every period due for archival. The user may choose a single SQS queue or an SQS queue per schema based on the volume of tables.
- 4 A proxy **Lambda** function de-queues the **Amazon SQS** messages and for every message invokes **AWS Step Functions**.
- 5 **AWS Step Functions** unloads the data from the **Amazon Redshift** cluster into an **Amazon S3** bucket for the given table and period.
- 6 **Amazon S3 Lifecycle configuration** moves data in buckets from **S3 Standard** storage class to **S3 Glacier** storage class after 90 days.
- 7 **Amazon S3 inventory** tool generates manifest files from the Amazon S3 bucket dedicated for cold data on daily basis and stores them in an S3 bucket for manifest files.
- 8 Every time an inventory manifest file is created in a manifest S3 bucket, an **AWS Lambda** function is triggered through an **Amazon S3 Event Notification**.
- 9 A **Lambda** function normalizes the manifest file for easy consumption in the event of restore.
- 10 The data stored in the S3 bucket used for cold data can be queried using **Amazon Redshift Spectrum**.

