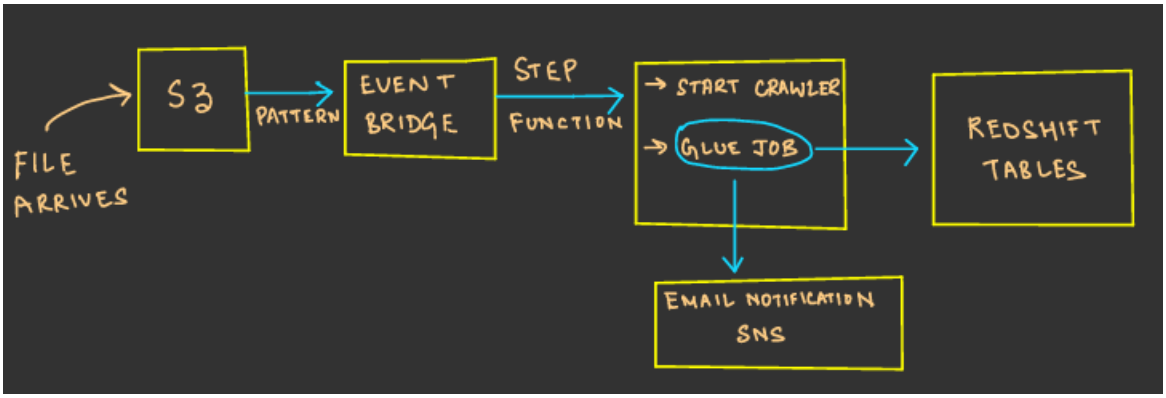
- **Airline data ingestion**
  Tech Stack – *SQL, SAP Hana, Redshift, Glue, S3, Step function, EventBridge*

Project design:



Project execution steps:

- Daily flights data would come to s3 bucket. 1 historical master table is present for airports.
- Created 2 tables in redshift for flights and historical master table for airports.
- Ingested data to master Airport table in Redshift from S3 using COPY command as it's a one time load.
- Created 2 crawlers for redshift table and 1 crawler for one time initial data load.
- Created airlines database in glue.
- Created glue visual etl job which would take data source from s3 raw flight data and redshift historical airport table, perform multiple join operations and dump to target source redshift table. Incremental load will be there as job bookmarking has been enabled.
- Orchestrating different aws services based on sdk and api calls through step functions. Glue job has been automated through step functions.
- Created new trail in CloudTrail and added in AWS CloudTrail data events in S3 so that when any file gets added to S3, step function could be triggered and through EventBridge rule.
- Created eventbridge rule when any files gets added to s3 then step function needs to be triggered by cloudtrail. The process is fully automated.
- After glue job execution gets completed, user notified though email notification.

Summary - Crawler will get executed, it will crawl the new files and new partition will get updated in the metadata in data catalogue, step function will get triggered, glue job will read the new file based on the job bookmarking and will load the new records in the redshift table.

Step function execution: