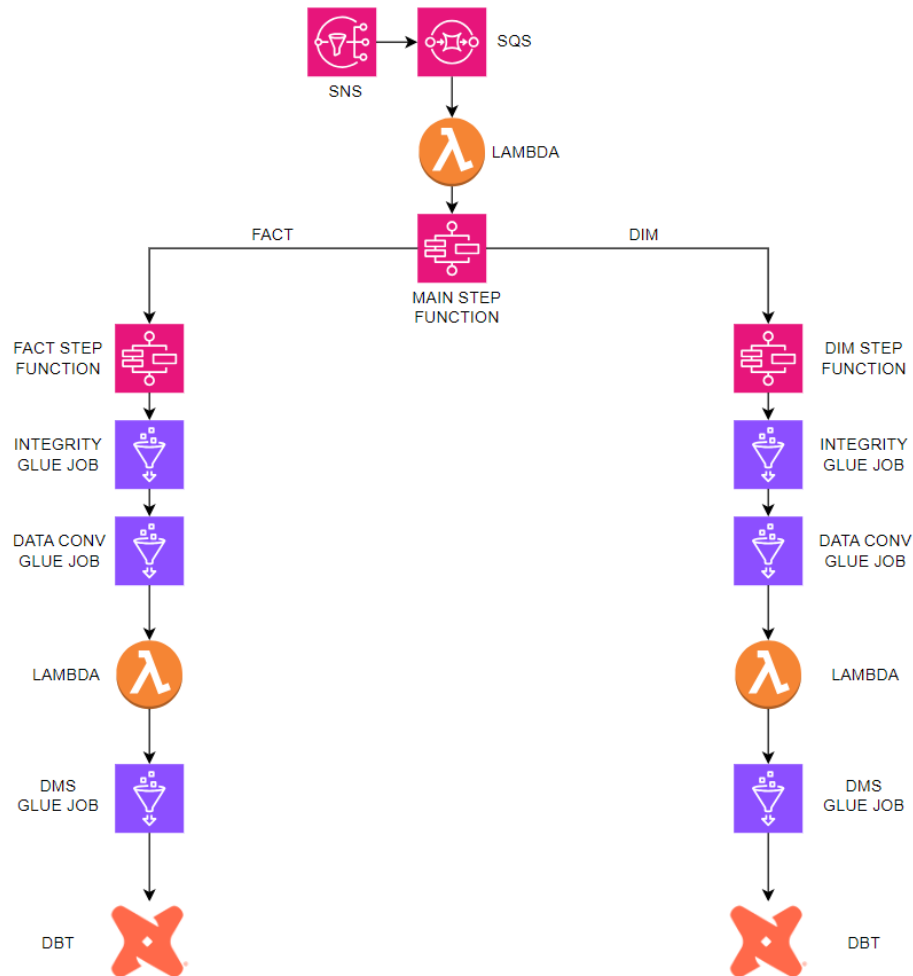


Process flow Diagram

(considering e.g. for **sae sp rid**)

Flow Diagram:



Buckets/folder details:

- Source s3 Bucket : unfi-data-landing-dev-s3
- Destination s3 Bucket : unfi-data-prepared-dev-s3
- Final prepared folder : s3://unfi-data-prepared-dev-s3/dev/sae_sp_rid/stage/
- Archive Folder : s3://unfi-data-prepared-dev-s3/dev/sae_sp_rid/dms/stage/

Flow details:

❖ SNS/SQS

Both works as notification services or to raise the event whenever is the new incremental file is loaded in the particular folder of the table

❖ Step Function's

Used for Orchestration

❖ Glue Job

- Each Glue job is operated on Schema level
- Which carries out below 3 operations on each tables within that schema
 1. Integrity – Validate the count in the files
 2. Data Conversion – Conversion (Datetime, Null, Numeric, Data, Column), Drop extra columns etc.
 3. DMS – Load the data to Redshift Staging table

❖ DBT

Once the glue job moves the data to the Staging table then, DBT task is invoked which move the data to main/master table.