AUTOMATE RECEIPT PROCESSING

Objective:

To automate the extraction and storage of receipt data using AWS services like S3, Lambda, Textract and DynamoDB.

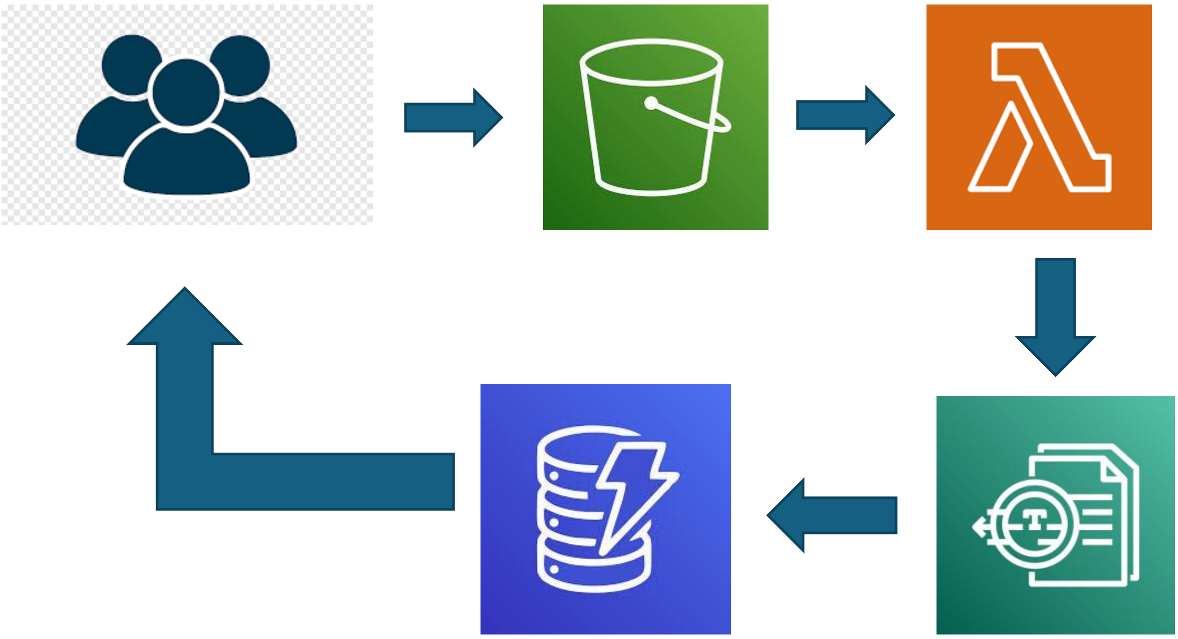
Tools & Services Used:

- Amazon S3 – Storage for receipt images.
- AWS Lambda – Event-driven backend logic.
- Amazon Textract (AnalyzeExpense API) – To extract structured data from receipts.
- Amazon DynamoDB – To store the extracted receipt data.

Work Flow:

- User uploads receipt (image) to an S3 bucket.
- S3 triggers a Lambda function via event notification.
- Lambda calls Textract's analyze_expense() to extract fields like Vendor Name, Total, Date, etc.
- Extracted data is converted properly (e.g., using Decimal) and saved to a DynamoDB table.

Architecture Diagram:



STEP 1: CREATE S3 BUCKET

1. Bucket Name: “your-bucket-name"
2. Trigger: Enabled for object-created events (no prefix needed).
3. File format must be supported by Textract (.png, .jpeg, .pdf).
4. The file format mostly preffered is .pdf. Because the textract need a clear format of image or document to extract the needed text.

STEP 2: CREATE DYNAMODB TABLE

1. Table Name: “your-table-name".
2. Partition key: “receipt_id”. Keep other things as default and create table.

STEP 3: CREATE LAMBDA FUNCTION

1. Author from scratch > Lambda function name: “your-function-name".
2. Runtime: Python-3.13.
3. Create LambdaFunction.

STEP 4: PERMISSIONS ON LAMBDA FUNCTION

1. Go to Permission on Lambda function > Attach Policies.
   - 1.1. DynamoDB FullAccess
   - 1.2. S3 ReadonlyAccess
   - 1.3. Textract FullAccess
2. Go to Permission on Lambda function > Create Inline Policy which is provided as source code in JSON format and Attach policy.
   - 2.1. Make sure to change your S3-Bucket name. DynamoDB table name, your account id and your Region before applying the json source code on permission.

STEP 5: DEPLOY CODE ON LAMBDA FUNCTION

1. Copy and paste the source code which is given on Python format in the code section of your Lambda Function.
2. Deploy the code.

STEP 6: TRIGGER S3 ON LAMBDA FUNCTION

1. Click Add Trigger > Lambda Function.
2. On Select source > Choose S3 Bucket.
3. On Bucket name > Click on “your-bucket-name".
4. On Event types > Apply **“PUT”** and **“**All object creates events**”.**
5. Click Add.
6. Now your S3 Bucket will be connected with your Lambda Function.

DEPLOYMENT

- Add a high pixel or clear jpg receipt in your S3-Bucket.
- Check on your DynamoDB Table.