

Intelligent Invoice Processing System using Amazon Textract & various other AWS services.

About MATE

Motherson Automotive Technologies & Engineering (MATE) is the polymer division of Motherson Sumi Systems Ltd. MATE units are located across India in different locations. MATE manufactures a wide range of plastic components for the automotive and the non-automotive industry. Motherson Automotive Technologies & Engineering (MATE) specializes in large size injection molding, blow molding, compression molding and vacuum forming, which are supported by post molding operations.

AWS Services used



About Challenge

MATE's back-office operations team, process Invoices as a part of back-office operations for their customers, all these invoices which they receive from their customers are manually processed.

In this manual processing they pick information manually from scanned invoice documents and put in system, below information is extracted from invoices –

- Header information (e.g. Supplier Name, Invoice No., Invoice Date, GSTN No etc.)
- Footer information (e.g. Taxes & Invoice Amount etc.)
- Line level information: Item level information (e.g. part no., Qty, amount etc.)

This is quite time consuming, error prone and tedious for them to accomplish for nearly 30K Invoices every month, a team of 12 people are engaged in this processing leads to lot undifferentiated activity which creates scope of automation. Other OCR solutions which were being tried had their own shortcomings so couldn't be adopted as a solution.




Proposed Solution

It is required to have an automated serverless solution to address this requirement. So, all the invoices are copied to a pre-decided system location and our system picks from that location and extract the required information from Invoice documents using Amazon Textract, as it is more than a simple optical character recognition (OCR) to also identify the contents of fields in forms and information stored in tables, and put into database system and process all invoices one by one. Steps involved are as follows -

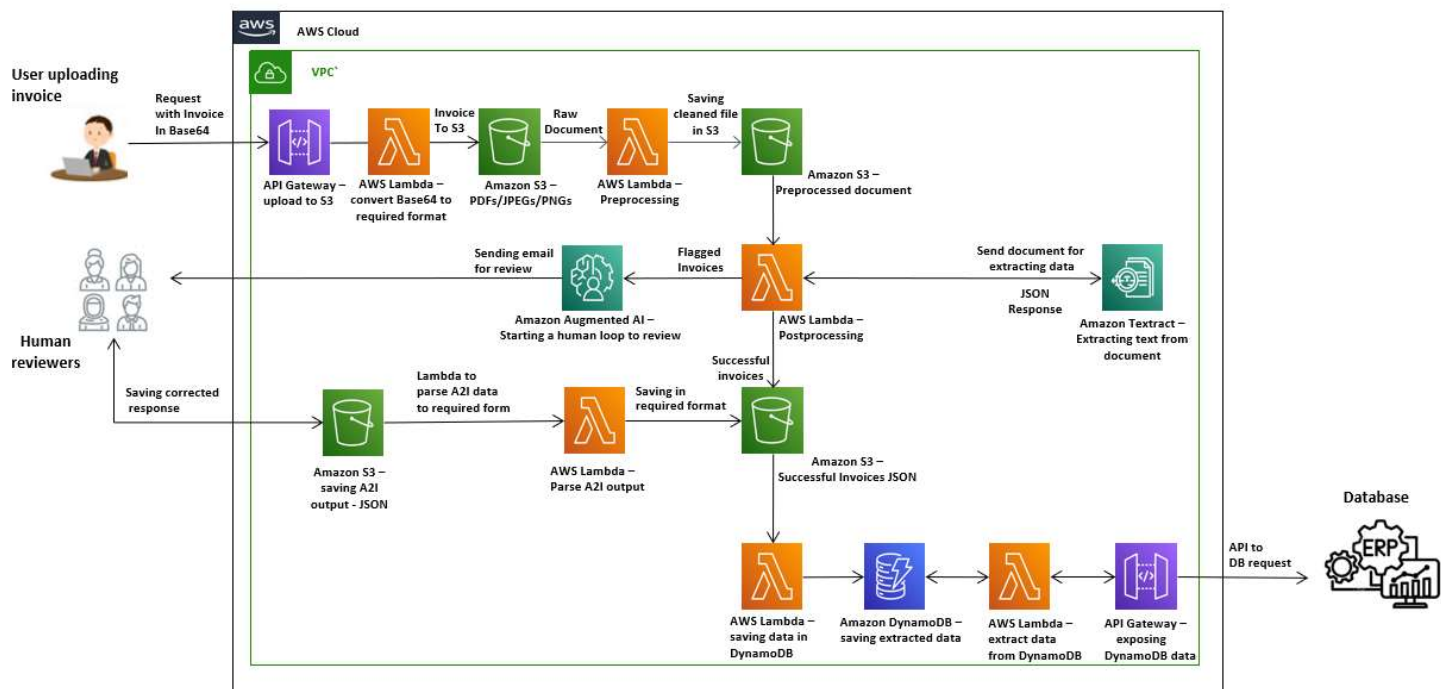
- Pre-processing of Invoices – to clean invoice documents which increases readability
- Post-processing of Invoices – to extract information and validate data to further improve quality of solution and made it error-proof also, line/word confidence level to make it complete solution
- Creating a human loop – using Amazon A2I to review and edit flagged invoices
- Updating data to ERP – sending DynamoDB data to SQL server where all the information is priorly stored

Solution Outcome

All in all, we successfully created a solution to process invoices with accuracy of 99% approx. This resulted in manual effort saving of about 40% of team (FTE's) initially and thus lead to

-  Reduction in effort and cost. (40% team reduced)
 -  Reduction in cycle time
 -  Improvement in quality of processing
- There is increased customer satisfaction which possibly will be translated to more orders from same customer.

Architecture Diagram



How AWS services helped in the digitization of the invoices

Amazon S3 to store files in different stages

It is an object storage service that offers industry-leading scalability, data availability, security, and performance. In this solution it helped to store raw invoice documents, preprocessed invoice documents and for storing final JSON output files.

AWS Lambda to run code serverless for preprocessing and post processing

It let us run code without provisioning or managing servers. Lambda allowed us to run code for virtually any type of application or backend service - all with zero administration. Just upload the code and Lambda takes care of everything required to run and scale the code with high availability. Here, preprocessing lambda and post processing lambdas were set to be automatically triggered from different S3 buckets.

DynamoDB to store key-value pairs from processed JSON

It is a key-value and document database that delivers single-digit millisecond performance at any scale. It's a fully managed, multiregional, multi-master, durable database with built-in security, backup and restore, and in-memory caching for internet-scale applications. It helped to store processed JSON for invoice documents.

Amazon Textract to extract data from PDF

It is a fully managed machine learning service that automatically extracts printed text, handwriting, and other data from scanned documents that goes beyond simple optical character recognition (OCR) to identify, understand, and extract data from forms and tables. Here, it helped to extract data from scanned invoice documents format. It uses machine learning to instantly read and process any type of document, accurately extracting printed text, handwriting, forms, tables and, other data which helped in quickly automating manual document related activities.

About the Partner



MothersonSumi INfotech & Designs Ltd.

MothersonSumi INfotech & Designs Limited (MIND), a SEI CMMI Level 5 IT services company and the IT back bone of Motherson group. MIND is a trusted technology partner to over 200 clients globally. Our value proposition is in our strength in specific Industry segments and years of experience in the areas of intelligent warehousing, Supply chain enablement, software application development, smart ERP customization, infra managed services, cloud, IoT & Analytics. MIND is serving customers in 41+ countries with a strong team of 1500+ professionals.