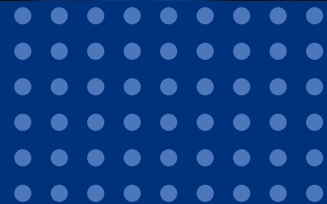


# ENTITY EXTRACTION USING GEN AI

---

MAY 2024

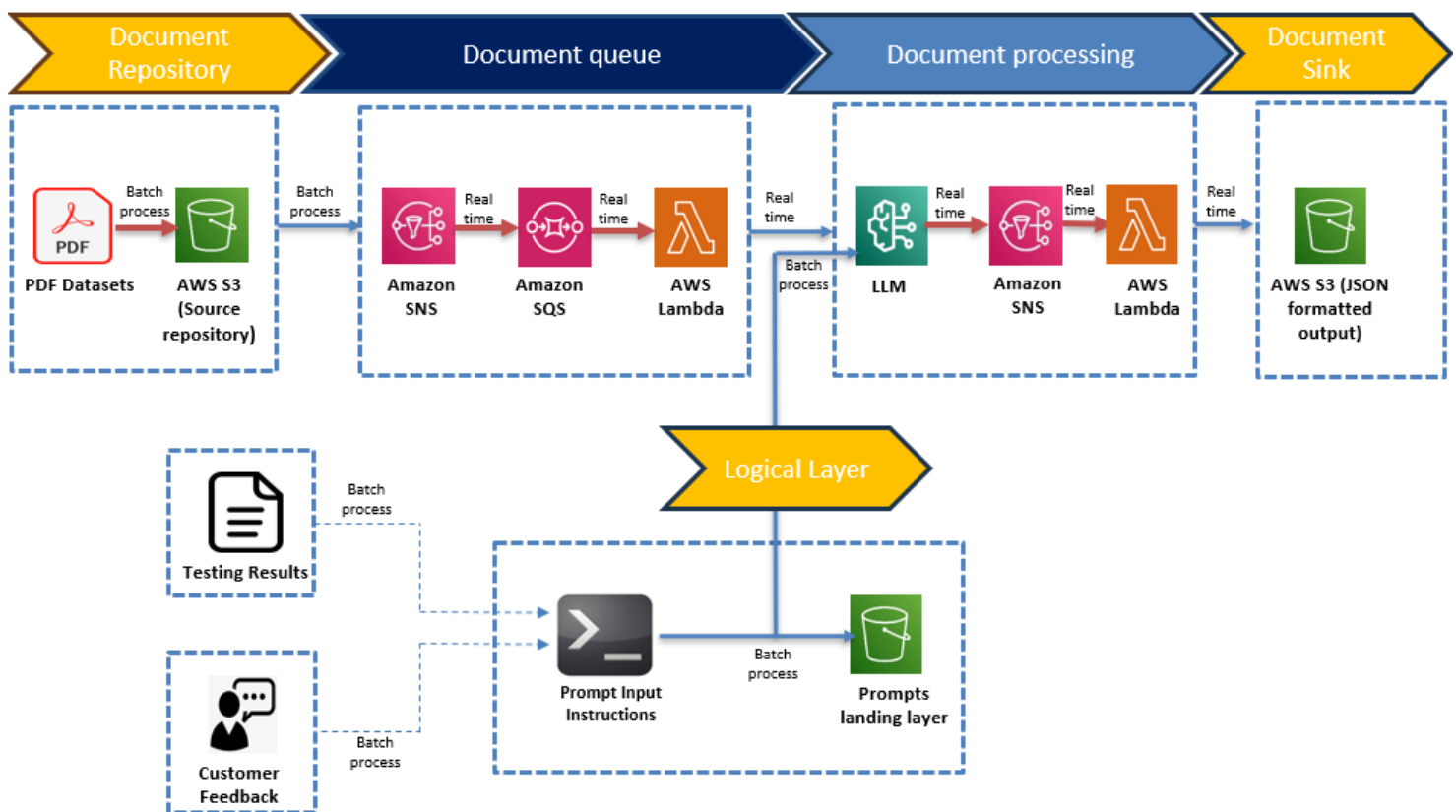


## Request and Guidelines Provided

- Client: Private Equity Firm
- Currently, entity data extraction is done manually, where the necessary entities are extracted from the source file, and the details are captured in the Excel file and finally converted into a JSON file
- As it involves more time and is error-prone, we implemented the automation process to classify the valid files, extract the necessary entities, and create the JSON file where the end-user application would consume it

## Methodology and Final Deliverable

- AWS services were leveraged to establish an automated entity extraction pipeline, enabling the extraction of data from source files upon their upload to S3, finally producing the entity-extracted results in JSON format
  - Once we receive the notification of files arrival in S3, we read the files sequentially and classify the files using Amazon Textract
  - After the classification, we process the valid files into the LLM (Anthropic Claude 2) to extract the necessary entities based on the provided final output structure
  - During fine-tuning, we used prompt engineering to have the valid prompts processed to LLM to avoid the hallucinated results being generated
  - We leverage the services of SQS, SNS, Lambda, CloudWatch, etc.. as part of the pipeline to have seamless integration from the source to the output being generated
  - The final JSON output is converted to CSV equivalent files and format to MS Excel reporting standards and shared over E-mails



Enabling our clients to achieve resource optimization through leveraging an offshore capacity model

## Our Highlights



Worked with  
1,000+ Clients



From Emerging to  
Established Firms



Servicing Clients  
in 80+ Countries



Across Industries  
and Asset Classes



Client-Integrated  
Operating Strategy

## Why Partner With Us



### HR Solution

- Flexible Staffing Model
- Dedicated Tiered Team Support



### Quality

- Industry-Leading Talent
- Multiple Layers of Quality Checks



### Operating Leverage

- Centralized Resource Pool
- Integration across Value Chain



### Cost Efficiencies

- Lower Indirect Costs
- Meaningful Direct Cost Savings

## What We Do

### Private Equity and Portfolio Companies

- Data Sourcing and Engineering
  - Capturing Data from Portfolio Companies
  - Creating Data Pipeline Enabling Automated Data Flows
- Reporting and Automation
  - Dashboard and Reports for Performance of Portfolio Companies
  - Fund Summary Reports for Investors
  - Process Optimization and Automation

### Capital Markets

- Data Sourcing and Engineering
  - Capturing and Processing Disparate Datatypes from Multiple Data Sources
  - Data Sanitization, Curation, and Data Pipeline Creation
  - Creating Databases, Data Lakes, and Data Warehouses & Ingesting using ETL/ELT Jobs
- Data Analytics, Reporting, and Automation
  - Customized Analytics
  - Key Indicators, Thresholds, Alerts, and Controls
  - Robotic Process Automation

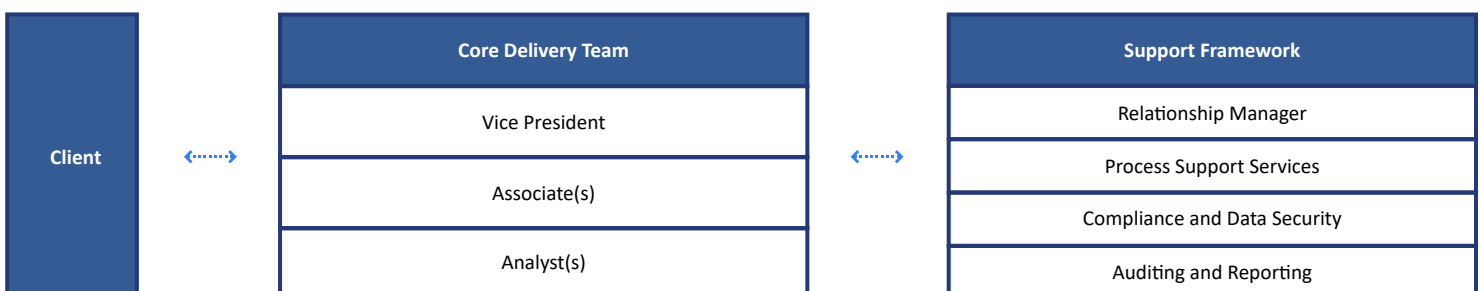
### Corporates

- Data Sourcing, Design, Architecture, and Storage
  - Capturing, Sanitizing, and Ingesting Disparate Datatypes from Multiple Data Sources using Data Model and Data Flow Diagrams
  - Creating and Maintaining Databases, Data Lakes, Data Warehouses, and Data Marts
- Data Analytics, Reporting, Governance, and Automation
  - Customized Analytics and OLAP Processing
  - Key Indicators, Thresholds, Alerts, Controls, and Data Governance
  - Robotic Process Automation and Artificial Intelligence

### Data Vendors

- Data Sanitization and Enrichment
  - Creating Rules and Alerts
  - Enrichment of Data using Fuzzy Logic
- Data Reporting, Automation, and Governance
  - Dashboard and Reports for Performance of Portfolio Companies
  - Master/Meta Data Management, Data Lineage, and Access Controls

## How We Do It





[salesupport@tresvista.com](mailto:salesupport@tresvista.com) |  
[www.tresvista.com](http://www.tresvista.com)