Building an Analytics Stack for ImageRegistrars

# Overview

ImageRegistrars is seeking to develop a modern analytics stack to support data-driven decision-making across multiple departments. The project will focus on centralizing data from different sources, performing sentiment analysis where applicable, and visualizing performance metrics on Apache Superset. The proposal outlines the technical architecture, departmental use cases, costs, and the project timeline.

# 1. Proposed Analytics Stack

Each department has specific data requirements and sources:  
  
1. Front Office (Walk-in Dashboard Analytics): Data from customer interactions and agent performance logs.  
2. Registry (Certificate Replacements, Transfers): Data on registry activities such as certificate replacements and agent performance.  
3. Call Center (3CX Integration): Call logs and transcripts from the 3CX phone system running on a Linux-based server (Debian).  
4. Emails/e-Correspondence (FreshDesk Integration): Email and support ticket data from FreshDesk.  
5. Finance/Admin (QuickBooks Integration): Financial data from QuickBooks for summaries and reports.

**ETL Process (Data Processing Pipeline):**  
- Tool: Mage-AI for ETL  
- Extract: Pull data from MSSQL, MongoDB, 3CX, FreshDesk, and QuickBooks.  
- Transform: Data cleansing, aggregation, and sentiment analysis (where required).  
- Load: Store transformed data into MSSQL and MongoDB for Superset visualizations.  
  
**Sentiment Analysis:**  
- Language: Python-based NLP libraries (e.g., NLTK, TextBlob, Hugging Face).  
- Call Center: Analyze customer sentiment from 3CX call transcripts.  
- Emails/e-Correspondence: Perform sentiment analysis on tickets and emails from FreshDesk.  
  
**Database Layer:**  
- MSSQL: For structured data, such as performance metrics and financial transactions.  
- MongoDB: For unstructured data, including sentiment analysis results and email logs.  
  
**Visualization Platform:**  
- Apache Superset: This tool will be used to create interactive dashboards for each department, visualizing key metrics, performance indicators, and sentiment analysis results.  
  
**APIs and Integration:**  
- FastAPI will be used to integrate external platforms like 3CX, FreshDesk, and QuickBooks with the analytics stack.  
  
**Unstructured Data Handling:**  
- Apache Iceberg/Trino: For large-scale management and querying of unstructured data, including transcripts and email logs.