# Take-Home Interview Exam – Franchisee Agentic Enrichment

## Project Context

Our team is building an automated data pipeline to extract and enrich structured data from Franchise Disclosure Documents (FDDs). These documents contain hundreds of franchisee location records that are critical for market planning, competitive analysis, and sales prospecting.

In the extracted data, the franchise owner may be listed as either an individual or an LLC. To enable deeper analysis and entity-level roll-ups, we enrich each record with the associated parent LLC, where applicable, along with relevant details about the business entity. The current workflow is too slow to scale with the growing volume of documents, making automation essential.

## Objective

### Task 1: Build a Python-Based Data Pipeline

Develop a Python pipeline that processes extracted franchisee records and enriches them with publicly available information about the business owners. The enrichment should involve matching the franchise owner (individual or LLC) to external data sources and retrieving relevant details.

Expected Output Fields:

* Franchise Owner Name
* Legal Corporate Name
* Corporate Address
* Corporate Phone Number
* Corporate Email
* Franchise Owner LinkedIn URL (if available)
* Source URLs used for enrichment

### Task 2: Design a Cloud Deployment Architecture

Provide a high-level cloud deployment architecture to support the data pipeline. The architecture should prioritize scalability, reliability, and automation. Google Cloud Platform (GCP) is preferred; AWS is acceptable.

The architecture should include components for:

* Document ingestion
* Parallel extraction and enrichment
* Quality control/exception handling
* Storage and access (e.g., dashboards, alerts)

## Deliverables

### 1. Owner Enrichment

* A Python module that:
  + Searches for public business owner data on the internet
  + Returns enriched records based on the extracted input

### 2. Deployment Architecture

* An architecture diagram (.png, .pdf, or .drawio)
* A brief description of the components and their interactions

### 3. Submission

* A GitHub repository or zipped package containing:
  + Python code for enrichment
  + A README.md with setup instructions
  + Sample PDFs (or links to publicly accessible documents)
  + Architecture diagram and accompanying description