### **1. Job Scraping & Data Ingestion**

**Responsibilities:**

* Integrate the Ultra JSEARCH API to fetch finance job postings, store in SQL database all key fields.
* Extract key fields: job title, company, location, description, and ticker (if available).
* Validate and preprocess data. (Ensure no missing values in data set).
* Store the raw job data in your SQL database (using the jobs table from the schema).

**Team Tasks:**

* Develop a cron job to schedule mechanisms to run exactly 1 month after the first date. (To continually pull a new 50K job searches and store in DB).
* Build error-handling and logging into the process.
* Document API usage and data storage conventions.

*For ease, it is simpler to ask Chatgpt to build this into a .csv file first, and then integrate it after into the database for permanent storage.*

### **2. Resume Parsing & Keyword Matching**

**Responsibilities:**

* Load and parse a general-purpose finance resume (in text and PDF format compatible).
* Use NLP techniques (Regex, TF-TDM model) to extract key descriptive words and skills from job postings, as well as the resume, to use TF-TDM model to weight the words and generate a similarity score.
* Compare and rank job postings against the resume using similarity scores. Validate that you agree with the resulting weights.
* Link this model to the database for permanent storage of our TF-TDM model.

**Team Tasks:**

* Create pipelines for text extraction and cleanup.
* Develop and fine-tune the TF-TDM model to compute similarity scores.
* Integrate the similarity scores back into the job data (update the SQL database with similarity scores to the resume in a new SQL table)

### **3. Sector Classification & Company Data Enrichment**

**Responsibilities:**

* Extract ticker symbols from job postings. Utilize the database with company information to link this to tickers and extract relevant information such as closing price and sector name.
* Integrate with the YFinance API to fetch company sector and industry information.
* Categorize job postings into finance sub-sectors (investment banking, asset management, fintech, etc.)
* Update the companies table to maintain consistent company information.
* Compute total sector counts from job postings pull and weight each sector against the whole to find the total % of each sector.

**Team Tasks:**

* Build functions to query YFinance and map ticker data to sectors.
* Handle cases where ticker data is missing or ambiguous.
* Ensure data consistency between job postings and company records. (Test a few company records against the related sector information to ensure that these are consistent with expectations).

### **4. Historical Job Data Tracking & Forecasting**

**Responsibilities:**

* Maintain a SQL database of past job postings to monitor trends.
* Track historical trends based on keywords, sectors, and posting dates.
* Develop and train an ARIMA forecasting model to predict future job opportunities based upon previous data.
* Generate insights on hiring demand and market shifts.

**Team Tasks:**

* Develop data-cleaning routines and historical data pipelines.
* Build forecasting models and validate their performance.
* Document assumptions, model parameters, and forecast results.

### **5. Visualization & Dashboard Development**

**Responsibilities:**

* Create visualizations of the data via Streamlit to display:
  + Current job postings and sector classifications.
  + Resume-job similarity scores.
  + Historical job trends and sector-wise hiring patterns.
  + Forecasted job opportunities and trends.

**Team Tasks:**

* Design the visual layouts and user interface.

### **6. General Oversight and Infrastructure, Logging & Deployment (Ryan)**

**Responsibilities:**

* Set up the project’s infrastructure, including a virtual environment (venv), configuration management, and logging.
* Develop continuous integration and deployment (CI/CD) pipelines.
* Implement unit tests and code quality checks for robustness.
* Assist in the development of the pipeline and integration with the SQL database

**Team Tasks:**

* Configure global settings and environment variables (e.g., API keys, database paths).
* Create deployment scripts and set up version control (Git).
* Maintain documentation for infrastructure setup and deployment procedures.
* Maintain a back-up copy in case of user error