Solution Architecture **Group ExpertFinder**

Process (People)

- Scrap LinkedIn profiles and Scholar articles
- EDA
- Data processing
- Model training/tuning
- Build App
- User can query the app: ask for an expert
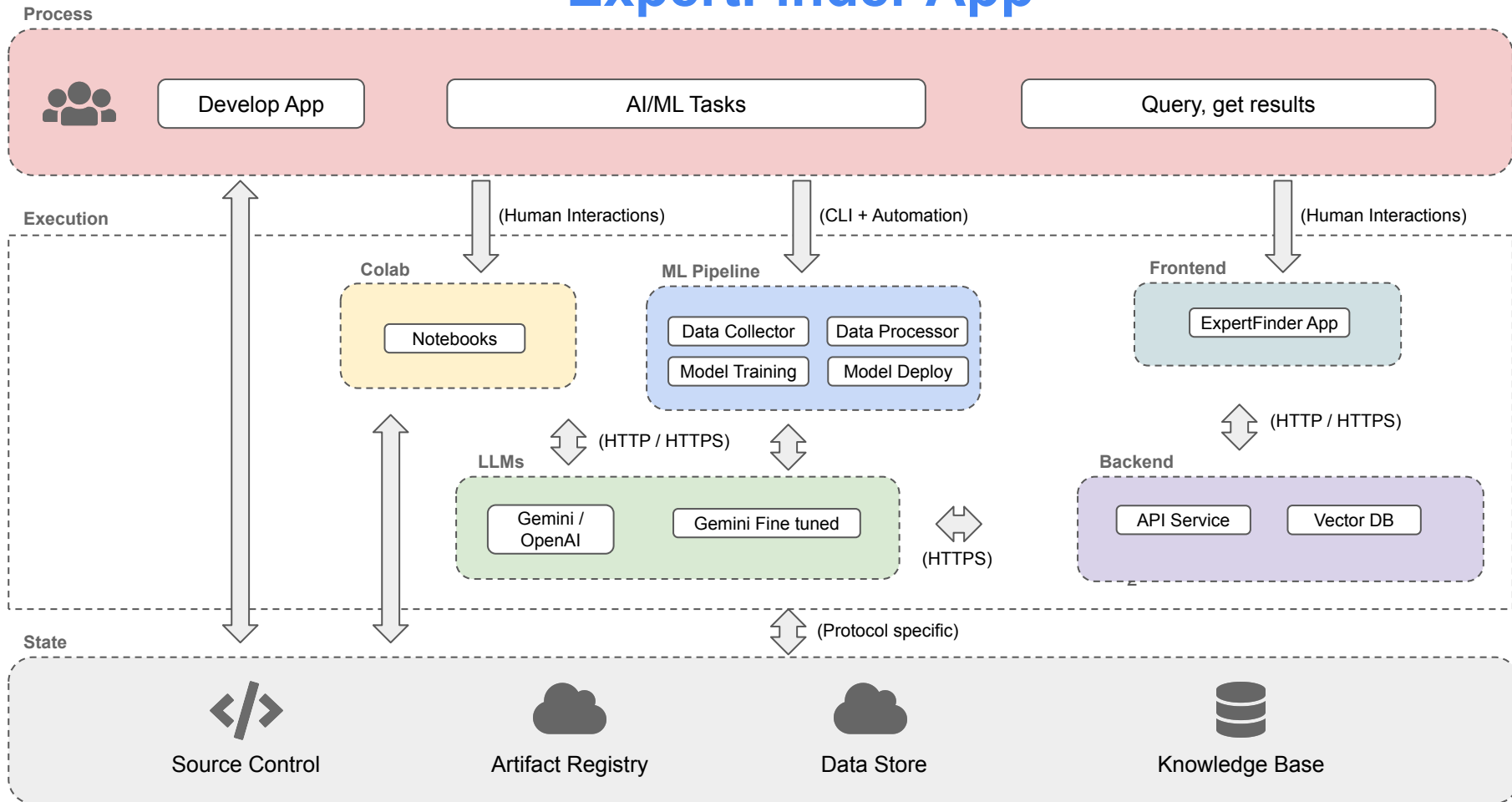- View profiles and save search history

Execution (Code)

- Preprocess new searches
- RAG finds and ranks experts
- Call LLM for chat
- RAG for QA
- Get attributes and present results

State (Source, Data, Models)

- Save APIs raw data and processed data
- Save model weights
- Save text chunks
- Save embeddings
- Information on pre processing
- Source code repo for collaboration

Solution Architecture

ExpertFinder App

Process

- Develop App
- AI/ML Tasks
- Query, get results

(Human Interactions)

(CLI + Automation)

(Human Interactions)

Execution

Colab
- Notebooks

ML Pipeline
- Data Collector
- Data Processor
- Model Training
- Model Deploy

Frontend
- ExpertFinder App

(HTTP / HTTPS)

(HTTP / HTTPS)

LLMs
- Gemini / OpenAI
- Gemini Fine tuned

(HTTPS)

Backend
- API Service
- Vector DB

(Protocol specific)

State

- Source Control
- Artifact Registry
- Data Store
- Knowledge Base

Technical Architecture

ExpertFinder App

Developers / Data Scientists

IDE/ CLI

Containers

Browser

Users

Browser

HTTPS 443

HTTPS 443

HTTP/HTTPS 80

Source Control

GitHub

Colab

Notebooks

HTTPS 443

Vertex AI

Data Collector

Data Processor

Model Training

Model Deploy

Gemini

Single Compute Instance/ Kubernetes Cluster

NGINX Container

HTTP 3000

HTTP 9000

API Service Container

ExpertFinder App Container

HTTP 8000

Vector DB

HTTPS 443

GCP

Google Artifact Registry

LinkedIn Data Collector

Google Scholar Data Collector

LinkedIn Data Processor

Google Scholar Data Processor

Bucket Storage & DVC control

Expert Ranking Image

API Service Image

ExpertFinder App Image

HTTPS 443

GCS Bucket

Data & Models

HTTPS 443

HTTPS 443

3

GCE Persistent Volume

Persistent Disk

NFS

HTTPS 443

HTTPS 443