

## **Project Setup & Run Instructions if needed**

### **1. Start the project: (~25 min)**

**In a terminal, navigate to the project folder and run:**

```
bash ./orchestrate.sh --action start
```

- **If bash doesn't start or gives an error**, check that Docker Desktop is open and activated.
- Wait until the terminal shows “**API ready**”.
- Keep this terminal open to monitor logs.

Since logs are in the opened terminal , you will need to open a new terminal to run commands

### **2. Test PDF Ingestion**

Single File:

```
curl -X POST "http://localhost:8000/ingest/" -F "input=@path/name.pdf"
```

Multiple File:

```
curl -X POST "http://localhost:8000/ingest/" -F "input=@path/name.pdf" -F "input=@path/name.pdf"
```

Directory of PDFs (requires Docker mount):

```
curl -X POST "http://localhost:8000/ingest/" -F "input=/pdfs"
```

#### NOTE:

- The system limits concurrent PDF ingestion to **5 files at a time** using a semaphore-based concurrency control. If the limit is exceeded, the API returns a “**server busy**” error. This limit is configurable in the code.
- Due to Swagger UI limitations, the FastAPI /docs page cannot properly send the unified input field used by /ingest/ for both file uploads and directory paths. Testing should be done via curl or a similar client.

### **3. Test Semantic Search**

```
curl -X POST "http://localhost:8000/search/" -H "Content-Type: application/json" -d "{\"query\": \"What is embeddings?\"}"
```

### **4. Terminate**

**In a terminal, navigated to the project folder , run:**

```
bash ./orchestrate.sh --action terminate
```