HIPAA QUESTIONS

Test Assignment

TASK DESCRIPTION

PRIMARY OBJECTIVE

Build a Retrieval-Augmented Generation (RAG) solution that effectively processes and queries HIPAA documentation which is represented as a single combined PDF file.
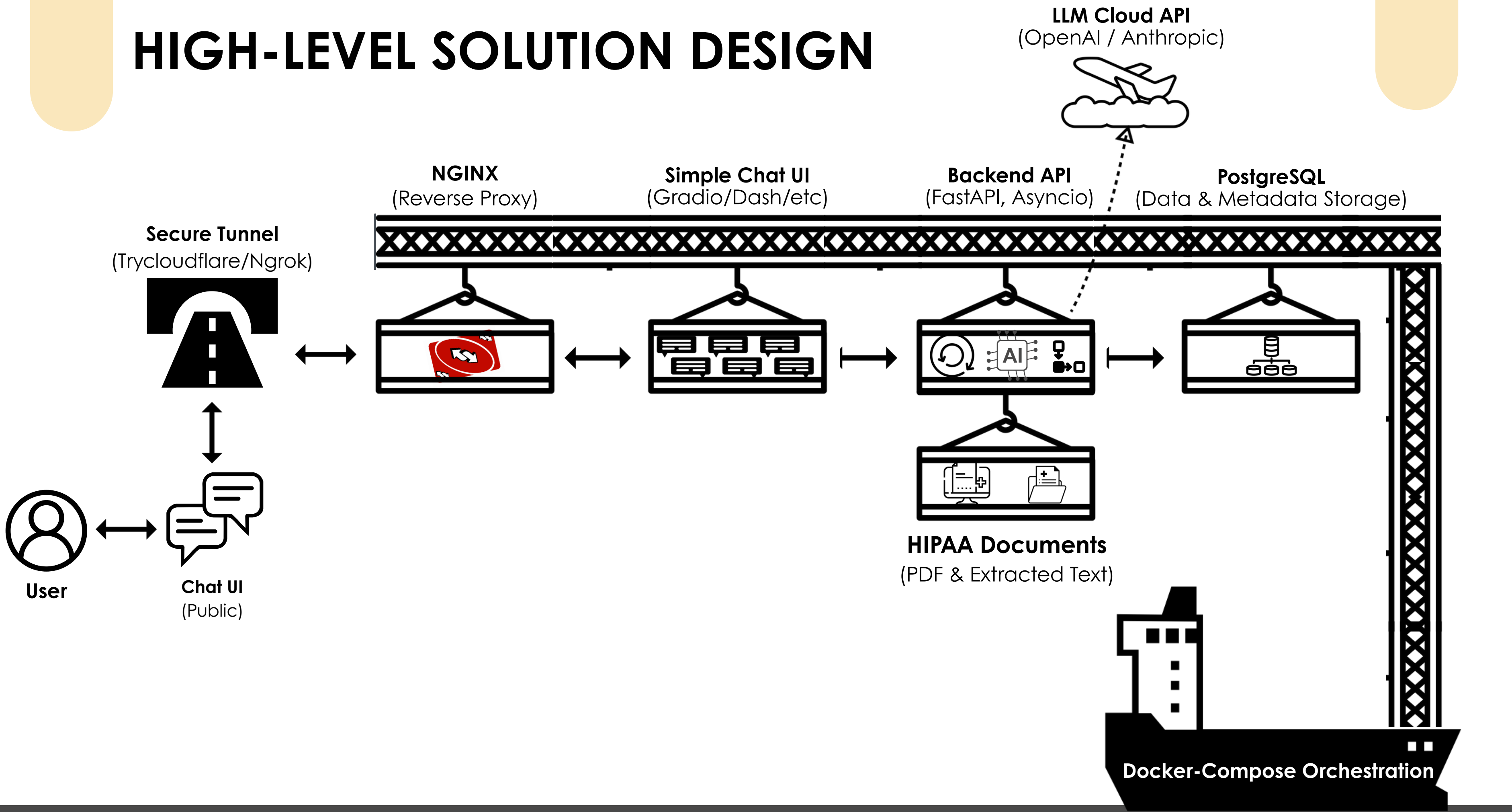
SOLUTION MANDATORY CAPABILITIES

The solution should be orchestrated under docker-compose and have following capabilities:

- Efficiently parsing, structuring, and indexing the HIPAA combined document to facilitate targeted information retrieval.
- Accurately answering user-submitted questions by returning precise, context-informed responses sourced from the provided HIPAA documentation.

SOLUTION COMPONENTS

AI Model	OpenAI / Anthropic
Database	PostgreSQL
Frontend	Simple Chat UI with Gradio, Dash, or other UI Builder
Backend	– Async FastAPI service – Nginx as reverse proxy – TryCloudflare/ngrok for public access
Orchestration	Docker Compose

HIGH-LEVEL SOLUTION DESIGN

LLM Cloud API
(OpenAI / Anthropic)

NGINX
(Reverse Proxy)

Simple Chat UI
(Gradio/Dash/etc)

Backend API
(FastAPI, Asyncio)

PostgreSQL
(Data & Metadata Storage)

Secure Tunnel
(Trycloudflare/Ngrok)

HIPAA Documents
(PDF & Extracted Text)

User

Chat UI
(Public)

Docker-Compose Orchestration

PDF FILE STRUCTURE

HIPAA documentation is represented as a single PDF that combines three HIPAA parts: 160, 162, and 164.

Within this PDF:
• Each of the three parts retains its outline and references to subdivisions (e.g., sub-paragraphs, sections).
• All text-based knowledge is splitted betwen paragraphs & sub-paragraphs which might contain information or references to another paragraphs
• All table-based information (if any) is out of scope for this assignment and should be ignored
• All picture-based information (if any) is out of scope for this assignment and should be ignored

EXPECTED RESULTS

1. Answer any HIPAA-related queries, providing accurate contextual information derived directly from the provided document.
2. When users explicitly request the full text of a specific HIPAA requirement or policy, the solution must directly quote exact content and precisely reference the original source sub-paragraph or section contained within the HIPAA combined PDF.
3. **Nice-to-have:** Appropriately handle broader user inquiries, such as requests for lists of inquiry-related HIPAA paragraphs.
4. **Nice-to-have:** Include exact quotes & relevant sub-paragraphs references

LIMITATIONS

Please consider following limitations during development:

- **No user management**: Do not implement registration, credentials management, user authentication, or roles and permissions.
- **Minimal security implementation**: Beyond basic encryption (HTTPS via tunnel/proxy), you are not required to implement any security controls, alignment protocols, HIPAA compliance infrastructure, or any audits.
- **Single-document focus**: Limitation to one provided combined HIPAA document (Parts 160, 162, and 164); multi-document ingestion and complex document management outside this core document are out of this task scope.
- **Low usage/concurrency:** Assume minimal concurrency (max a few simultaneous users). Scalability, high-availability, clustering, and performance optimizations are outside the scope.
- **No analytics/telemetry**: Any usage tracking, logging, analytics, and telemetry are out of the scope.

EVALUATION

The quality of your HIPAA RAG solution will be evaluated against these core indicators:

1. **Information Accuracy & Completeness**: Ability to deliver correct and precise answers to straightforward HIPAA-related questions, clearly reflecting the factual content of the provided documentation.
2. **Precision in Source Attribution & Referencing:** Ability to provide exact quotes from the text whenever specific requirements or policy details are requested, accompanied by clear references to exact source sections/subparagraphs.
3. **Robustness & Code Quality**: The solution should be robust, the code should be clear, easy to understand & follow best practices of Python development.
4. **Optional: Reference Gathering**: Ability to provide structured lists of all reference sub-paragraphs related to user inquiry.
5. **Optional: Quoting**: Ability to provide exact quotes and reference sub-paragraphs numbers within the answer to user inquiry.

SAMPLE EVALUATION QUESTIONS

Some of possible evaluation questions include:

1. What is the overall purpose of HIPAA Part 160?

2. Which part covers data privacy measures?

3. What does "minimum necessary" mean in HIPAA terminology?

4. Which entities are specifically regulated under HIPAA?

5. What are the potential civil penalties for noncompliance?

6. Does HIPAA mention encryption best practices?

7. Can I disclose personal health information to family members?

8. If a covered entity outsources data processing, which sections apply?

9. Cite the specific regulation texts regarding permitted disclosures to law enforcement.

WE ARE CONFIDENT IN YOUR ABILITIES

Wishing you success in this final step.