

**2024**

**Total Cost of Ownership**

**6/13/2024**

Learning Management Architecture Service



**Contents**

[**1.** **Introduction:** 3](#_Toc169193395)

[**2.** **Software/Hardware (Clouding) Assumptions:** 4](#_Toc169193396)

[**3.** **Requirements on Architecture:** 4](#_Toc169193397)

[Plan Overview 5](#_Toc169193398)

[Server Requirements 6](#_Toc169193399)

[**4.** **Project Cost Working Breakdown Structure** 6](#_Toc169193400)

[**4.1** **Hardware Requirements on Servers for LLMs and Django Hosted Server** 7](#_Toc169193401)

[AWS Pricing Estimate 7](#_Toc169193402)

[Azure Pricing Estimate 7](#_Toc169193403)

[Flexible Server Option (e.g., DigitalOcean, Linode) 8](#_Toc169193404)

[**5.** **Total Cost of Ownership (TCO)** 8](#_Toc169193405)

[Total Estimated Cost: $3000 USD 9](#_Toc169193406)

# **Introduction:**

1. Overall needs:

It is required to implement an artificial intelligence teaching assistant system that can use local knowledge base and general AI large model to realize students' independent question answering and resource recommendation functions.

2. Backend functions:

1). You can upload local knowledge base, including text, video and other resources, which can be uploaded, deleted and updated.

2). Support student information management, including registered student names, classes, student numbers, and question and answer records

3). Students can give positive or negative evaluations of AI answers, and teachers can see them in categories in the background and evaluate the correctness of the system's answers. For negative evaluations, teachers can update and supplement the knowledge base to gradually improve their question-answering capabilities.

3. Front-end functions:

Implement student Q&A in three ways

1). WeChat chatbot uses a separate WeChat ID to interact with students

2). WeChat Mini Program

3). Web page

4. Sample page:

Knowledge base upload page, where you can upload text, video and other resources

Web page Q&A interface:

5. Deployment requirements: Cloud deployment or local deployment, capable of being used by at least 500 people at the same time

6. Technology stack requirements

1). The development language uses Python or node.js

2). Knowledge enhancement using langchain framework

3). Use faiss for vector database

4). Graph database uses neo4j

5).AI model uses Tongyi Qianwen

7. Other needs

# **Software/Hardware (Clouding) Assumptions:**

For our chosen server we consider different expected user growth over a year’s period by providing a composite and flexible Server performance and service strategies as well as a choice of connectivity options. We have assumed that the location of Data center is in Republic of China. The first question in choosing a storage is the amount of space required for LLMs and Video Online Courses. This need to consider the number of expected users and current users for the next years. In order to face with the reliability of the server the Azure or other cost effective options so the resources should be kept below the engineering limits according to layer 2/3 server specification with value of as follow:

* **CPU – 80%**
* **Disk – 75%**
* **Memory – 80%**
* **Bandwidth – 75%**

# **Requirements on Architecture:**

Flexibility is a key performance of designing of a simple and cost effective clouding system. Our mentioned architecture is including of the following levels and tasks requirements:

# Plan Overview

1. **Project Setup and Environment Preparation (10 hours)**
   * Set up development environment (Python/Django or Node.js/Express.js).
   * Clone and review the repositories:
     + [FlowiseAI/Flowise](https://github.com/FlowiseAI/Flowise)
     + [redianmarku/MesoOn](https://github.com/redianmarku/MesoOn)
2. **Backend Development (30 hours)**
   * Implement local knowledge base management (upload, delete, update).
   * Develop functionalities for student information management and question-answer records.
   * Integrate evaluation and improvement features for AI answers.
3. **Front-end Development (30 hours)**
   * Create WeChat chatbot integration with a separate WeChat ID.
   * Build a WeChat Mini Program for student access.
   * Design and develop the web page interface for Q&A.
4. **Testing and Debugging (20 hours)**
   * Conduct thorough testing of all functionalities.
   * Debug and fix any issues or bugs.
5. **Deployment and Server Setup (10 hours)**
   * Choose a cloud provider (e.g., AWS, Google Cloud Platform, Azure) based on your server requirements.
   * Configure server instances and databases.
   * Deploy the application and ensure scalability and security.
6. **Documentation and Finalization (10 hours)**
   * Document the project, including setup instructions and usage guidelines.
   * Finalize any remaining tasks and prepare for project delivery.

# Server Requirements

In this project, it is needed to determine how much size are needed for a local knowledge baser, looking at video streaming bandwidth and storage. As the result the server requirements as well as AI model implementation and other costs of ownership are determined to provide simple and sustainable investment standards according to the 500 expected users.

Addition of de-duplication and Encryption in cloud AWS Azure are considered as the framework of a AI based architecture. These requirements are the direct result of the growing reliance of users on IT-based services, the need for service integration, and strengthening IT governance over dedicated servers. As a result, the amount of data that needs to be transferred or sent is dramatically reduced. By eliminating duplicate data records, we need between less space to store data in data center but needs more bandwidth required fewer client resources on source (client side) de-duplicated.

Based on your deployment needs for at least 500 concurrent users, consider the following server requirements:

* CPU: Dual-core or higher (e.g., AWS EC2 t3.medium or equivalent)
* RAM: 4GB or higher
* Storage: SSD-based storage for improved performance
* Database: Utilize managed database services for scalability and reliability (e.g., AWS RDS, Google Cloud SQL)

# **Project Cost Working Breakdown Structure**

Because the above and all requirements mentioned and designed a, therefore we need to use the adjusted server location and also estimate the prospected Active Users and their behavior in pick time to run LLM and video online courses to run smoothly. My plan excluding any hardware needs for programming and handling are estimated as below:

(10+30+30+20+10) hours ×20USD/hour = 2000 USD

## **4.1 Hardware Requirements on Servers for LLMs and Django Hosted Server**

### AWS Pricing Estimate

1. **Server Instances:**
   * AWS EC2 T3 instances (2 vCPUs, 4 GB RAM) - Approximately $480 per year ($40 per month).
2. **Managed Database:**
   * AWS RDS (Small instance, MySQL or PostgreSQL) - Approximately $360 per year ($30 per month).
3. **Storage:**
   * AWS S3 for data storage (100 GB) - Approximately $12 per year.
4. **Bandwidth:**
   * AWS Data Transfer (100 GB outgoing) - Approximately $20 per year.

Total AWS Estimated Cost: $872 per year.

### Azure Pricing Estimate

1. **Server Instances:**
   * Azure Virtual Machines B-series (2 vCPUs, 4 GB RAM) - Approximately $480 per year ($40 per month).
2. **Managed Database:**
   * Azure SQL Database (Basic tier) - Approximately $360 per year ($30 per month).
3. **Storage:**
   * Azure Blob Storage (100 GB) - Approximately $12 per year.
4. **Bandwidth:**
   * Azure Data Transfer (100 GB outgoing) - Approximately $20 per year.

Total Azure Estimated Cost: $872 per year.

### Flexible Server Option (e.g., DigitalOcean, Linode)

1. **Server Instances:**
   * DigitalOcean Droplets (2 vCPUs, 4 GB RAM) - Approximately $360 per year ($30 per month).
2. **Managed Database:**
   * Managed database service from the provider (e.g., DigitalOcean Managed Databases) - Approximately $300 per year ($25 per month).
3. **Storage:**
   * Additional block storage (100 GB) - Approximately $60 per year ($5 per month).
4. **Bandwidth:**
   * Outgoing bandwidth included in the plan, minimal additional cost.

Total Flexible Server Estimated Cost: $720 per year.

# **Total Cost of Ownership (TCO)**

As have shown above, indirect and overhead costs are not exclusively and specifically measurable in related to cost objects. In the dedicated server in addition to the general server costs we should calculate the costs object which are mostly related to software and programs licenses.

I have considered the project to be finished with total 3000 USD and duration of two month. I would need a programmers and AI Engineer.(I have considered myself AI Engineer). The minimum deposit for start of the project is 500 USD. The second installment payment are to be fulfilled by the demonstration. Finally should be completed cleared by test and delivery.

The important characteristics of the selected server are as follow:

1. **Dedicated Server Hardware:**
   * Estimated Cost: $1000 - $1500 USD (depending on specifications and provider)
2. **Operating System License:**
   * Windows Server License: Approximately $500 - $800 USD (one-time cost, varies based on edition)
   * Linux (open-source): No additional license cost
3. **Database Software License:**
   * Microsoft SQL Server Standard Edition: Approximately $2000 - $4000 USD (one-time cost, varies based on edition)
   * PostgreSQL (open-source): No additional license cost
4. **Web Server and Application Frameworks:**
   * Cost varies based on specific software used (e.g., Apache, Nginx, Django, Node.js).
   * Estimate: $100 - $500 USD (one-time or subscription-based)
5. **Monitoring and Security Tools:**
   * Cost varies based on specific tools and services selected.
   * Estimate: $100 - $300 USD (annual subscription)
6. **Backup and Disaster Recovery:**
   * Cost varies based on backup frequency, storage capacity, and disaster recovery solutions.
   * Estimate: $200 - $500 USD (annual subscription)

### Total Estimated Cost: $3000 USD

By carefully selecting software licenses, opting for open-source solutions where possible, and choosing cost-effective providers for services like monitoring and backup, you can stay within your budget of $3000 USD for the dedicated server setup.