**Assignment 4**

Documenting the brainstorming and discussion process for deciding on the type of business organization for developing and deploying the generative AI system:

Team Members Present:

1. [Azhan Saleem]
2. [Erik Lorimor]
3. [Tarun Krishna]

Date: [4/01/2024]

Location: [Zoom]

**Agenda:**

1. Discuss potential types of business organizations.
2. Evaluate the suitability of each type for developing and deploying the generative AI system.
3. Reach a consensus on the type of business organization.

**Discussion Summary:**

1. **Introduction:** The meeting began with an introduction to the task at hand: to decide on the type of business organization for which we will develop and deploy the generative AI system.
2. **Brainstorming**: Each team member contributed ideas regarding various business organizations. Some of the suggestions included:
   * Technology startup
   * Financial institution
   * Healthcare provider
   * Legal firm
   * Retail corporation
3. **Evaluation Criteria**: We discussed the criteria for evaluating the suitability of each type of business organization. Factors considered included:
   * Availability of proprietary documents for training and testing the AI system
   * Potential impact and usefulness of the generative AI system within the organization
   * Compatibility with Google Cloud Platform (GCP) Vertex AI services
   * Accessibility of relevant expertise within the team or the organization
4. **Analysis and Discussion:** After considering the various options, we analyzed the pros and cons of each type of business organization. We discussed how each organization type aligns with the requirements and objectives of the project.
5. **Decision**: After thorough deliberation, the team agreed to develop and deploy the generative AI system for a **technology startup.** We chose this option because:
   * Technology startups often deal with large amounts of proprietary documents and can benefit significantly from efficient content searches and knowledge retrieval.
   * The fast-paced and innovative nature of startups aligns well with the cutting-edge AI technologies we plan to implement.
   * Startups typically have a culture of experimentation and adoption of new technologies, making it an ideal environment for deploying advanced AI systems.
   * Our team members have experience or expertise relevant to the startup environment, which will facilitate development and deployment processes.
6. Next Steps: We outlined the next steps, which include formalizing the decision in our project documentation, informing stakeholders, and proceeding with the development process tailored to the needs of a technology startup.

**PART II: Business and Technical Requirements of the System**

The decision to develop a generative AI system stems from the recognition of its potential to significantly enhance various aspects of business operations and productivity within the chosen technology startup. The following are the key business values associated with the implementation of the generative AI system:

1. **Enhanced Knowledge Management:** The generative AI system will enable the organization to efficiently manage and access its proprietary documents, thereby streamlining knowledge retrieval processes. This will lead to improved decision-making, faster problem-solving, and enhanced collaboration among team members.
2. **Increased Productivity:** By automating content searches and providing quick and accurate answers to questions, the generative AI system will boost employee productivity. Time spent on manual information retrieval tasks can be redirected towards more strategic activities, ultimately driving business growth and innovation.
3. **Competitive Advantage**: Implementing cutting-edge AI technologies reflects the organization's commitment to innovation and staying ahead of the competition. The generative AI system will equip the startup with a powerful tool for extracting valuable insights from its data, enabling it to make informed decisions and respond swiftly to market changes.
4. **Scalability and Flexibility:** As the startup grows, the generative AI system can scale alongside the business, accommodating increasing volumes of data and evolving user needs. Its flexible architecture will ensure seamless integration with existing workflows and systems, supporting the organization's long-term growth objectives.

**Business Requirements of the Generative AI System:**

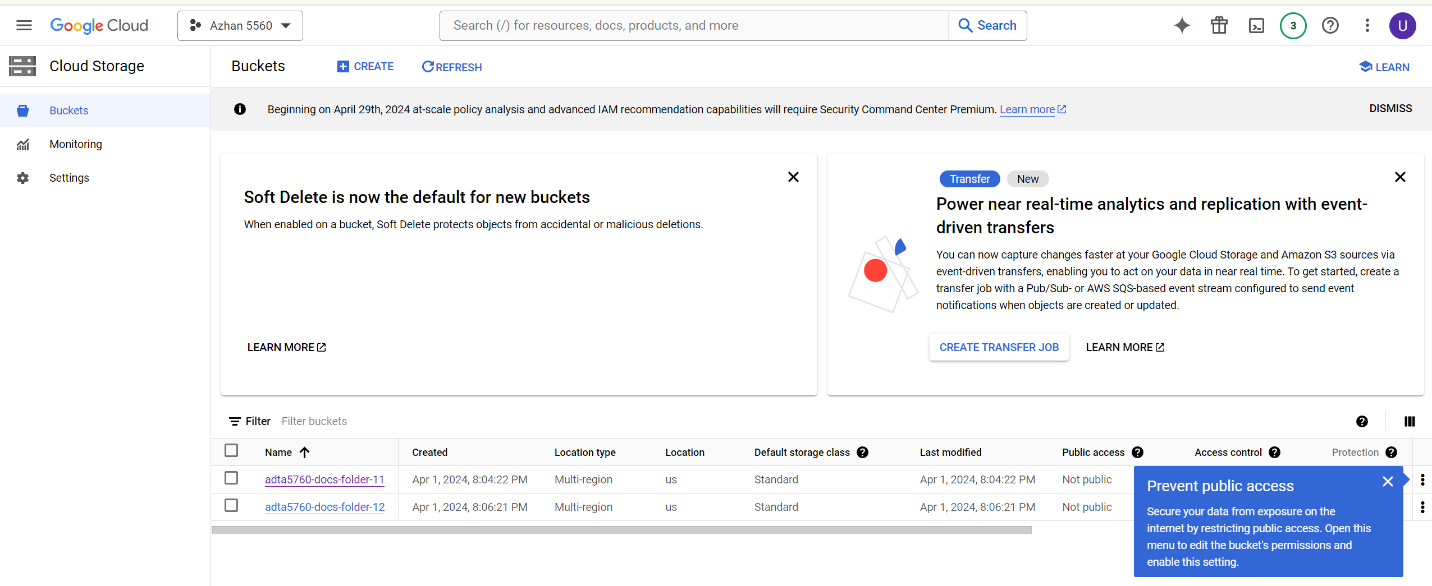
The business goals and objectives that the organization aims to achieve with the generative AI system include:

1. Improved Information Retrieval: The primary objective is to develop a system that facilitates quick and accurate retrieval of information from the organization's proprietary documents. This includes text-based searches, question-answering capabilities, and contextual understanding of user queries.
2. Knowledge Sharing and Collaboration: The system should encourage knowledge sharing and collaboration among employees by providing easy access to relevant information. It should support features such as document tagging, content recommendation, and collaborative filtering to enhance information discovery and dissemination.
3. Data Security and Compliance: Ensuring the security and confidentiality of proprietary documents is paramount. The system must comply with data protection regulations and implement robust security measures to safeguard sensitive information from unauthorized access or disclosure.

**Technical Requirements of the Generative AI System:**

1. AI Platform: The generative AI system will be developed using Google Cloud Platform (GCP) Vertex AI services, leveraging its comprehensive suite of tools and capabilities for machine learning and AI development.
2. Large Language Model (LLM): The system will utilize state-of-the-art large language models, such as GPT (Generative Pre-trained Transformer) models, for natural language understanding and generation tasks.
3. Generative AI Platform: The development will make use of popular generative AI techniques, including Retrieval Augmented Generation (RAG), Sentence Transformer, and tools provided by generative AI platforms like LangChain and Hugging Face. These platforms offer pre-trained models, fine-tuning capabilities, and APIs for seamless integration into the system.
4. Cloud Storage: The system will leverage cloud storage solutions provided by GCP for storing and accessing proprietary documents and trained model artifacts securely.
5. Vector Embeddings Generation and Management: Vector embeddings will be generated using techniques such as Word2Vec or BERT embeddings to represent documents and queries in a high-dimensional vector space. These embeddings will be managed efficiently using vector databases to enable fast and accurate similarity searches.
6. Advanced Vector Search Technologies: The system will incorporate advanced vector search technologies for efficient retrieval of relevant documents based on semantic similarity and contextual relevance.

**PART III: Data and Cloud Data Storage Requirements**

****

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* Created two buckets on Google Cloud Platform:
  + BUCKET 1: Name = adta5760-docs-folder-11
  + BUCKET 2: Name = adta5760-docs-folder-12
* Captured screenshots of both buckets:
  + Bucket 1: adta5760-docs-folder-11
  + Bucket 2: adta5760-docs-folder-12

**PART IV: Data (PDF Documents) Requirement 1**

**A screenshot of a computer

Description automatically generated**

**Step 1: Download and Extract ZIP File**

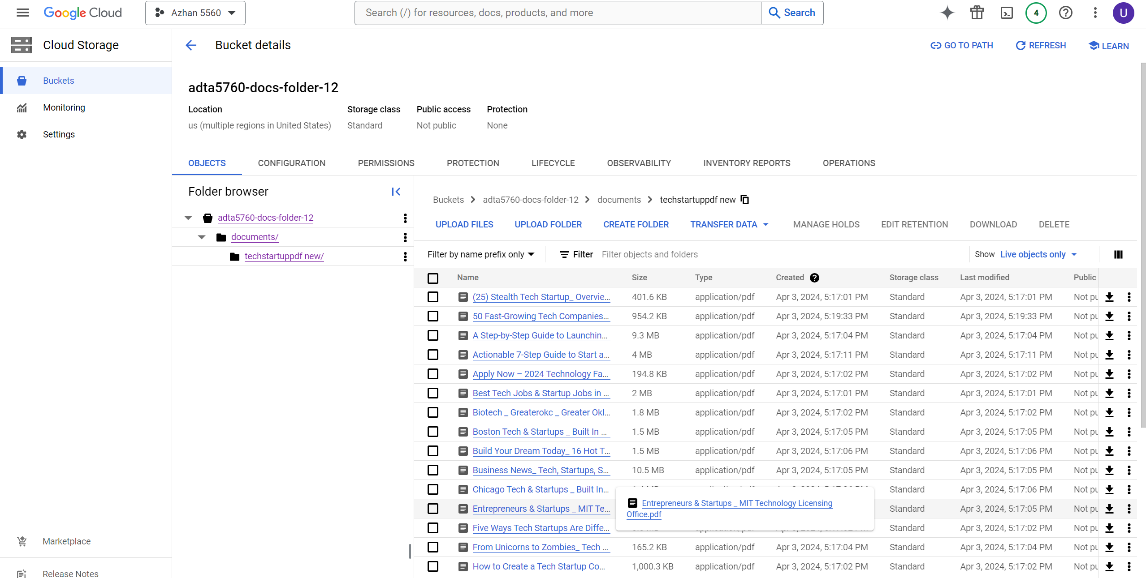
1. Accessed the designated location to download the ZIP file named "AAA\_FOUNDATIONALS.zip."
2. Downloaded the ZIP file to the local machine.

**Step 2: Upload Contents to Cloud Folder**

1. Accessed the Google Cloud Platform Console using the provided credentials.
2. Navigated to the bucket named "adta5760-docs-folder-11."
3. Located the folder structure within the bucket and identified the path "documents/pdfs" for uploading the contents.
4. Extracted the contents of the downloaded ZIP file on the local machine.
5. Uploaded the extracted contents to the "pdfs" folder within the bucket on Google Cloud Platform.

**PART V: Data (PDF Documents) Requirement 2**

* The team selected "Technology Startups" as the topic for this task.
* Relevant keywords and phrases related to technology startups were identified to aid in the search process. Examples include "startup ecosystem," "entrepreneurship," "innovation," "venture capital," "startup success factors," and "startup funding."
* The team conducted searches in academic databases, online repositories, and credible websites using the identified keywords and phrases.
* The team verified that a minimum of 30 documents were collected, and each document met the quality standards required for the task.
* Accessed the Google Cloud Platform Console using the provided credentials.
* Navigated to the bucket named "adta5760-docs-folder-12."
* Located the folder structure within the bucket and identified the path "documents/pdfs" for uploading the contents.
* Uploaded the 30 docs within the bucket on Google Cloud Platform.



A screenshot of a computer

Description automatically generated

**PART VI: Teamwork Evaluation (10 Points)**

1. Group Number: Team 2
2. Group Members: Azhan, Erik, Tarun.
3. Organized Meetings: Yes, all members participated in organizing meetings both online and in-person to work on HW 4.
4. Attendance: All members, including the student himself, attended the meetings.
5. Active Participation: All the members made good efforts to actively participate in the group work.
6. Opinions: Everybody collaborated.