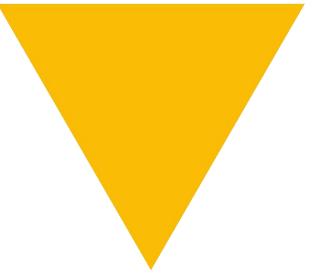


13



## Case study: GenAI solution design

# In this activity, you learn to ...

- 01 Evaluate customer requirements and propose a solution design
- 02 Architect a Retrieval Augmented Generation (RAG) solution
- 03 Extract text from multimedia content & generate embeddings
- 04 Define the work streams, pipelines, and components used in your solution
- 05 Articulate your design decisions in the context of SRE principles



# Activity Tasks:

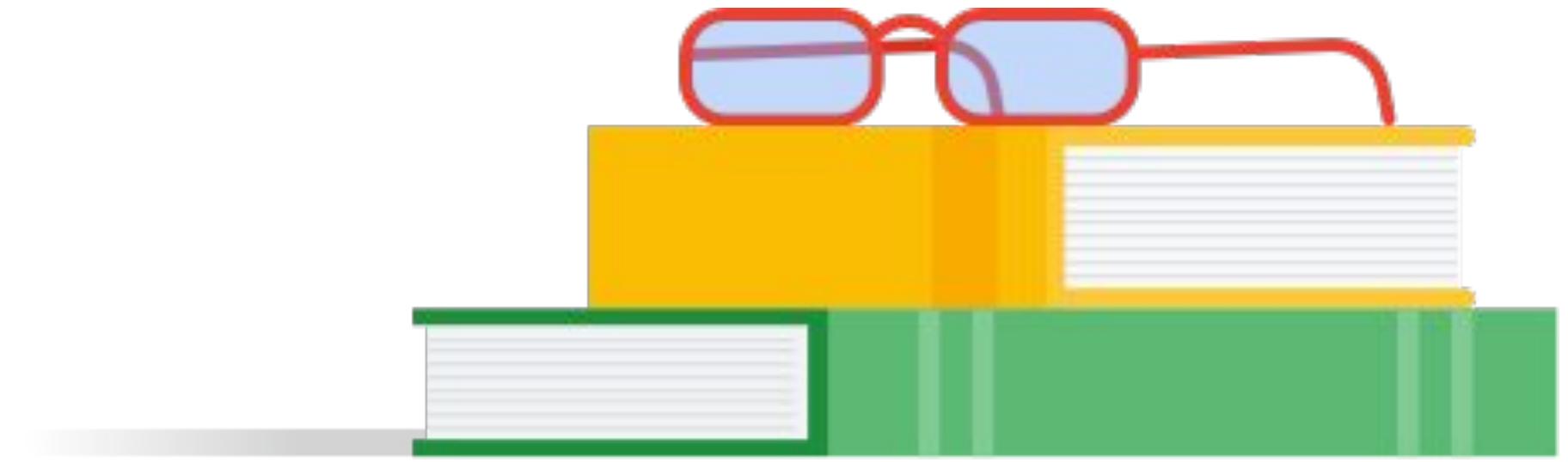
- 01 Read the scenario
- 02 Draft a conceptual design
- 03 Create a solution design
- 04 Review a sample solution



# Task 1

---

Read the scenario



# Activity introduction

## A real world design exercise

One of your customers is having a challenge with content discovery and searchability in their online learning platform.

The details of this case study are taken from a real customer engagement, but you must make the design decisions and present them with a solution.

### Action Item!

- Read the case study details

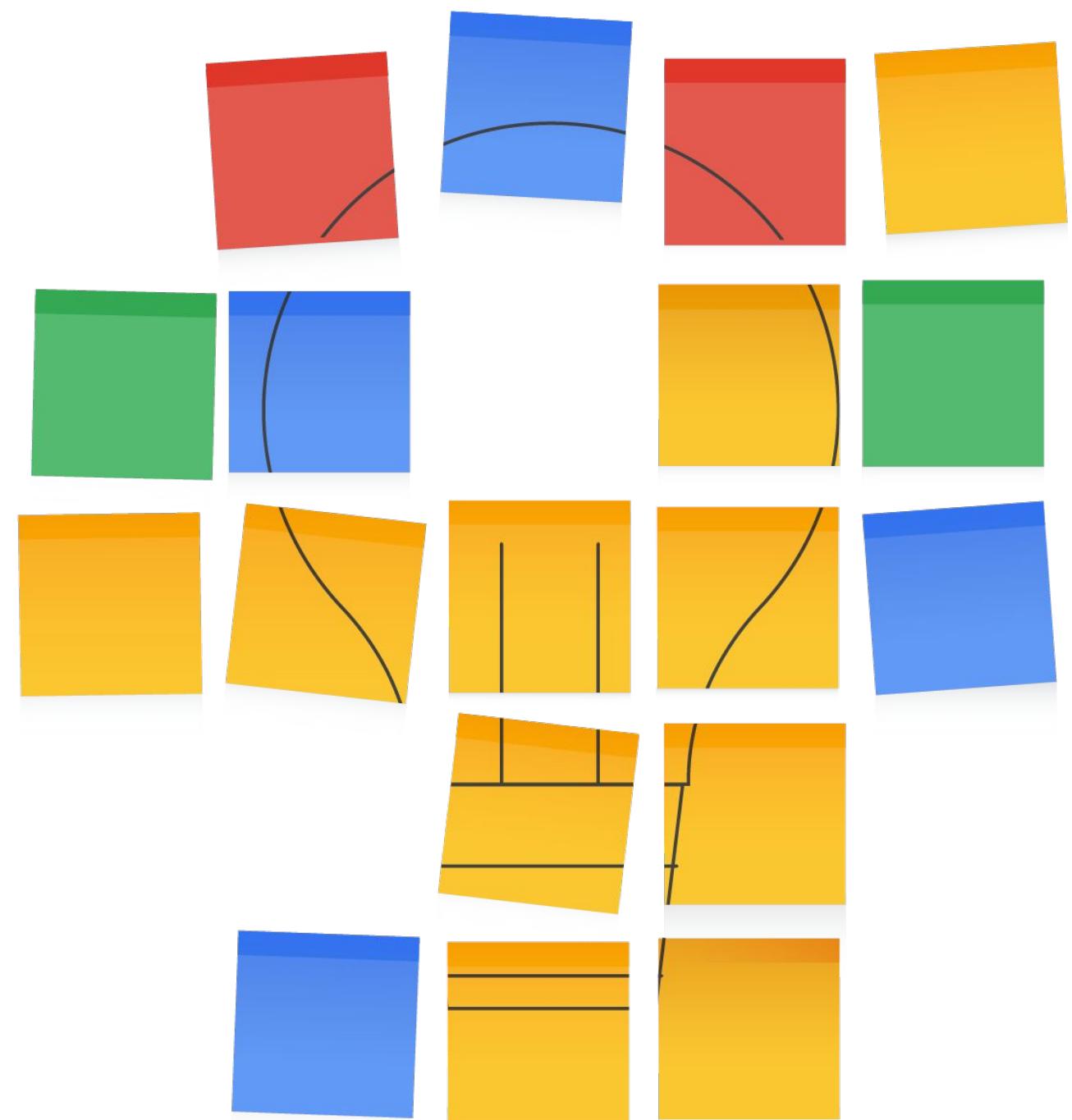


[Open the case study document](#)

# Task 2

---

Draft a conceptual design





DESIGN TIME!

Process overview:

Work stream #1

- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_

## A conceptual design should:

- Restate/paraphrase their challenges and business goals.
- Use simple, non-technical language and easy to draw diagrams (flow charts, boxes, arrows) to describe the solution.
- Divide the challenge space into discrete work streams.
- Divide each work stream into tasks and describe their purpose/function.
- Walkthrough changes to their critical user journeys.
- Explain how the components in your solution satisfy (or mitigate) each of the customer's business goals.





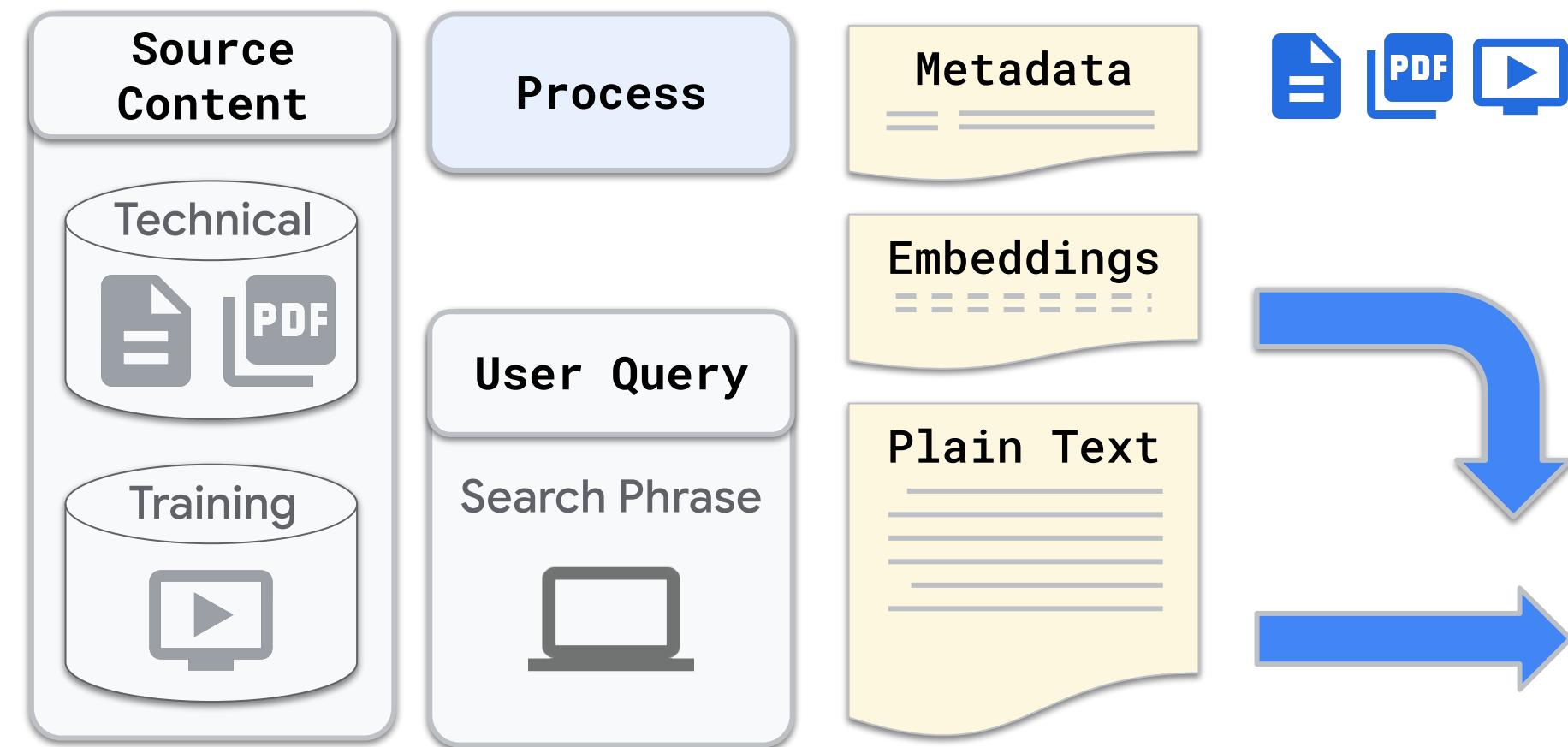
# DESIGN TIME!

## Process overview:

# Work stream #1

- This image shows a template for handwriting practice. It features ten identical sets of horizontal lines arranged vertically. Each set includes a solid top line, a dashed middle line, and a solid bottom line. To the left of the first set, there is a vertical margin with ten small, dark circular dots, one positioned above each set of lines. These dots are intended to help guide the placement of letters when writing.

*Some artwork to get you started...*





# DESIGN TIME!



## Process overview:

## Work stream #1



# Create a conceptual design



# Task 3

---

Create a solution design





DESIGN TIME!

Process overview:

Work stream #1

- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_

Your solution design should demonstrate your ability to:

- Identify the candidate products/services for use by the processes within each work stream.
- Evaluate the applicability and suitability of each candidate option against SRE design best practices.
- Interconnect dependent products and services to process data.
- Explain where your code executes, and where their data will reside.





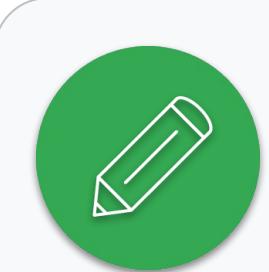
# DESIGN TIME!

## Process overview:

## Work stream #1



*Additional product artwork to get you started...*



Sample diagrams and Google Cloud icons:  
<https://cloud.google.com/icons>

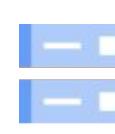


# LLM\_Process

## text-bison



- Name/purpose
- Firestore



- Name/purpose
- Cloud Storage



# Process Name

## Pipeline Component

## Copy and label the icons:

## Generic



- Name/purpose
- Cloud Storage

# Meaningfu



- Temp Copy of Assets
- Cloud Storage



# DESIGN TIME!



## Process overview:

## Work stream #1



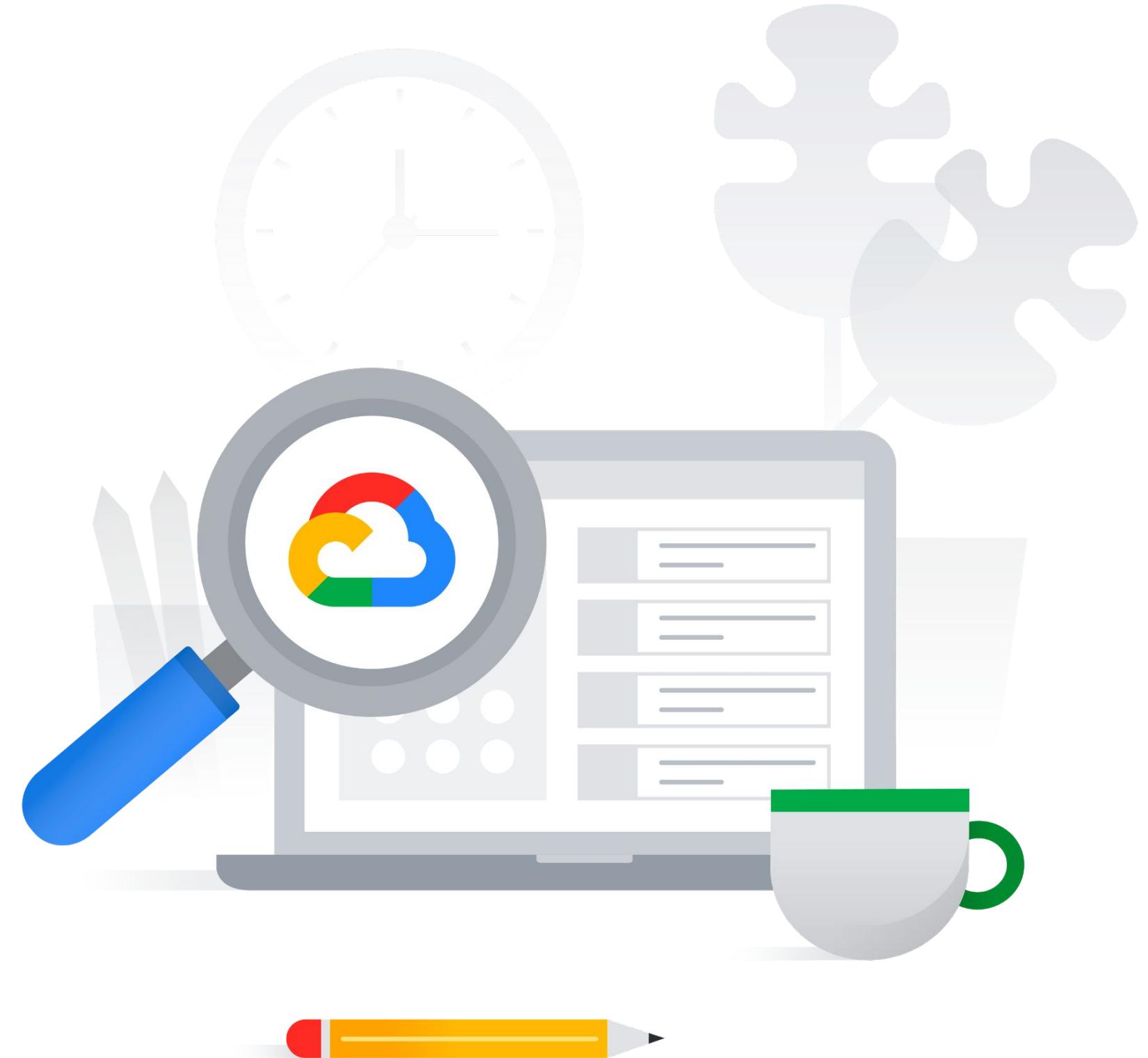
# Create a solution design



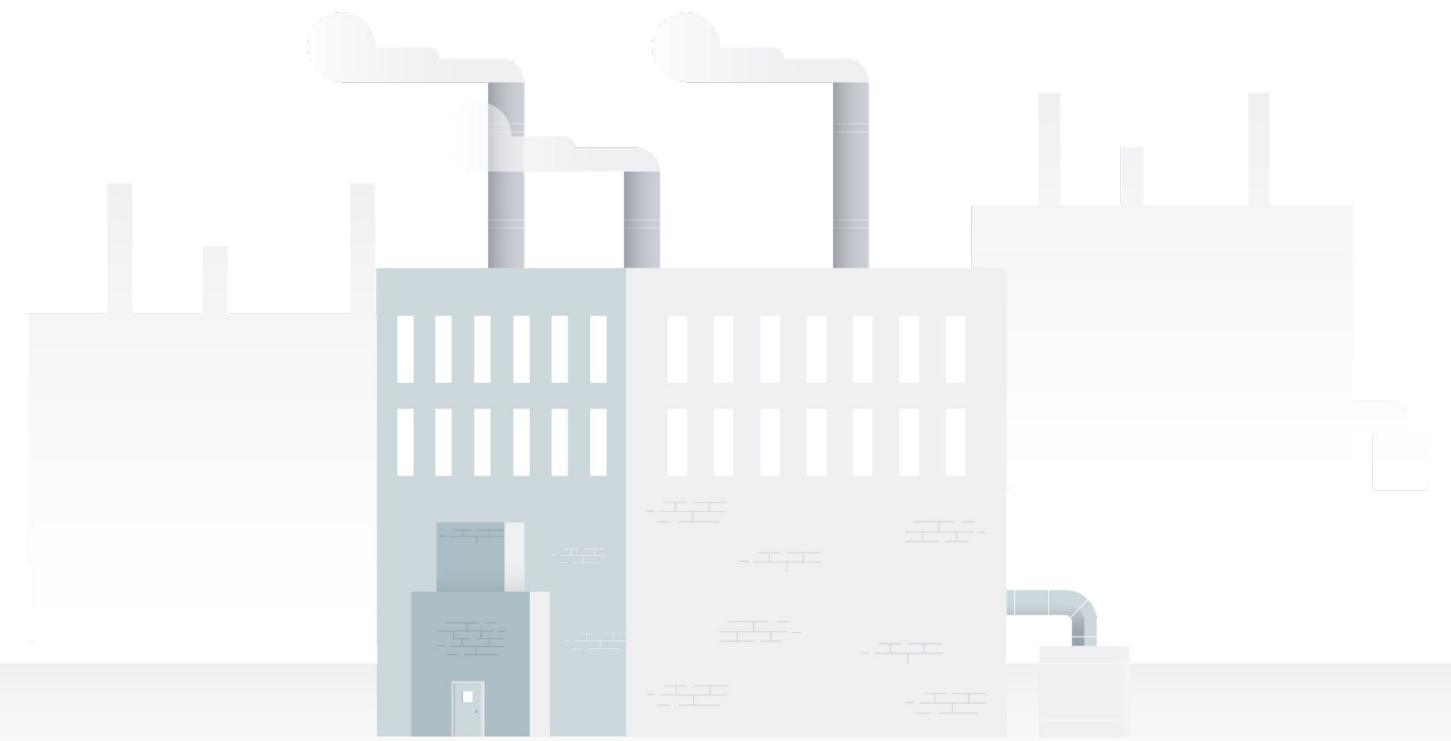
# Task 4

---

Review a sample solution



# Getting to know your customer



- Global manufacturer of compute and network infrastructure products.
- Provides their business partners free access to technical and training resources via their Partner University.
- Learner utilization of the Partner University is a KPI for your customer's staff.

# Focusing on their pain points

Your customer identified these areas of concern:

## Content Management

- Labor-intensive and error prone content management process
- Requires humans to label and summarize the content

## Content Discovery

- Relies on outdated keyword search indices
- Users report frustration locating relevant content

## Viewership / Engagement

- Increase in enrollments suggests users are interested
- But video viewership statistics indicate users aren't engaged

# Desired outcomes

Your customer stated their goals for the project are to:

Content Management

Minimize human involvement in the content ingestion process.

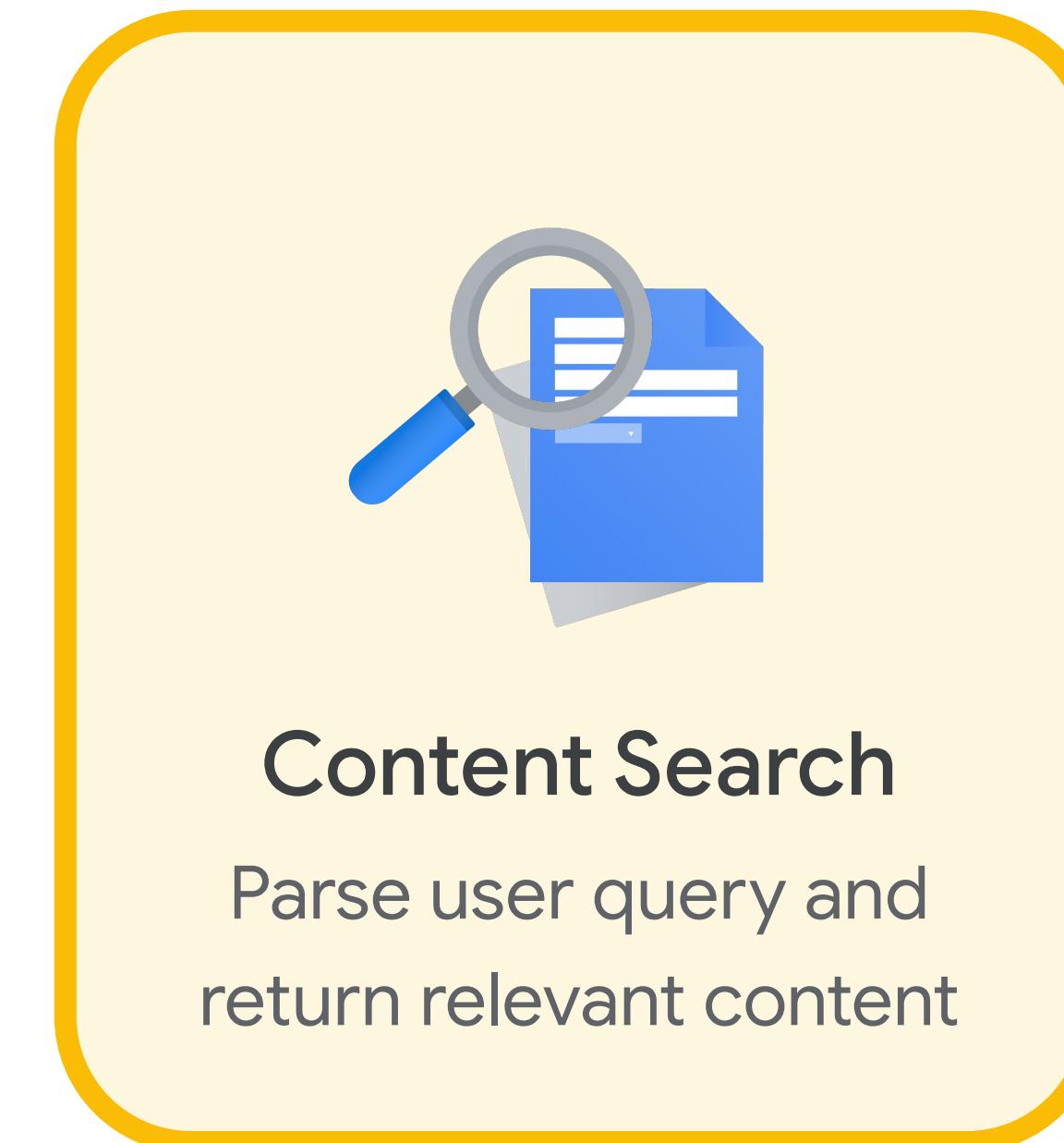
Content Discovery

Increased searchability and discoverability of our content.

Viewership / Engagement

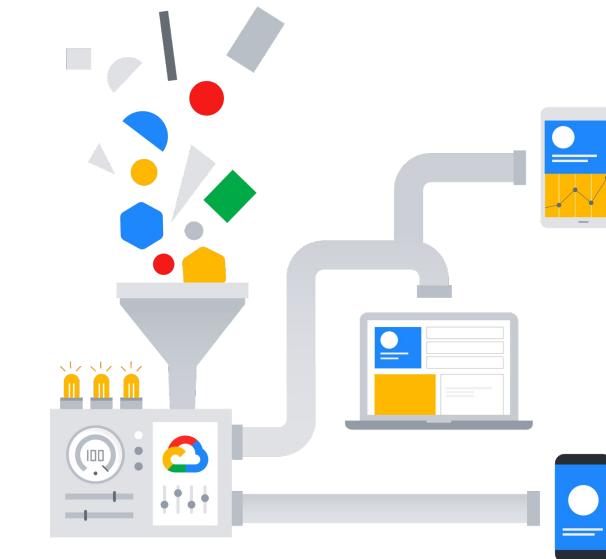
Increased content viewership and engagement.

# We split the challenge into two workstreams:



# Content Ingestion

Ingest and index the documents and videos



Related functional requirements:

- Minimize human involvement in the content ingestion process.
- Index PDFs, docs, and videos.

## Content Ingestion

Process Overview

- Copy the content
- Extract the text
- Transcribe the videos
- Generate metadata
- Generate embeddings
- Store the metadata and embeddings

# Content Search

Parse user query and return relevant content



Related functional requirements:

- Increased discoverability of content.
- Increased relevance of search results.
- Increased content viewership and engagement.

## Content Ingestion Process Overview

- Generate embedding for the user query
- Search vector store for Nearest Neighbors
- Retrieve documents
- Return results to user

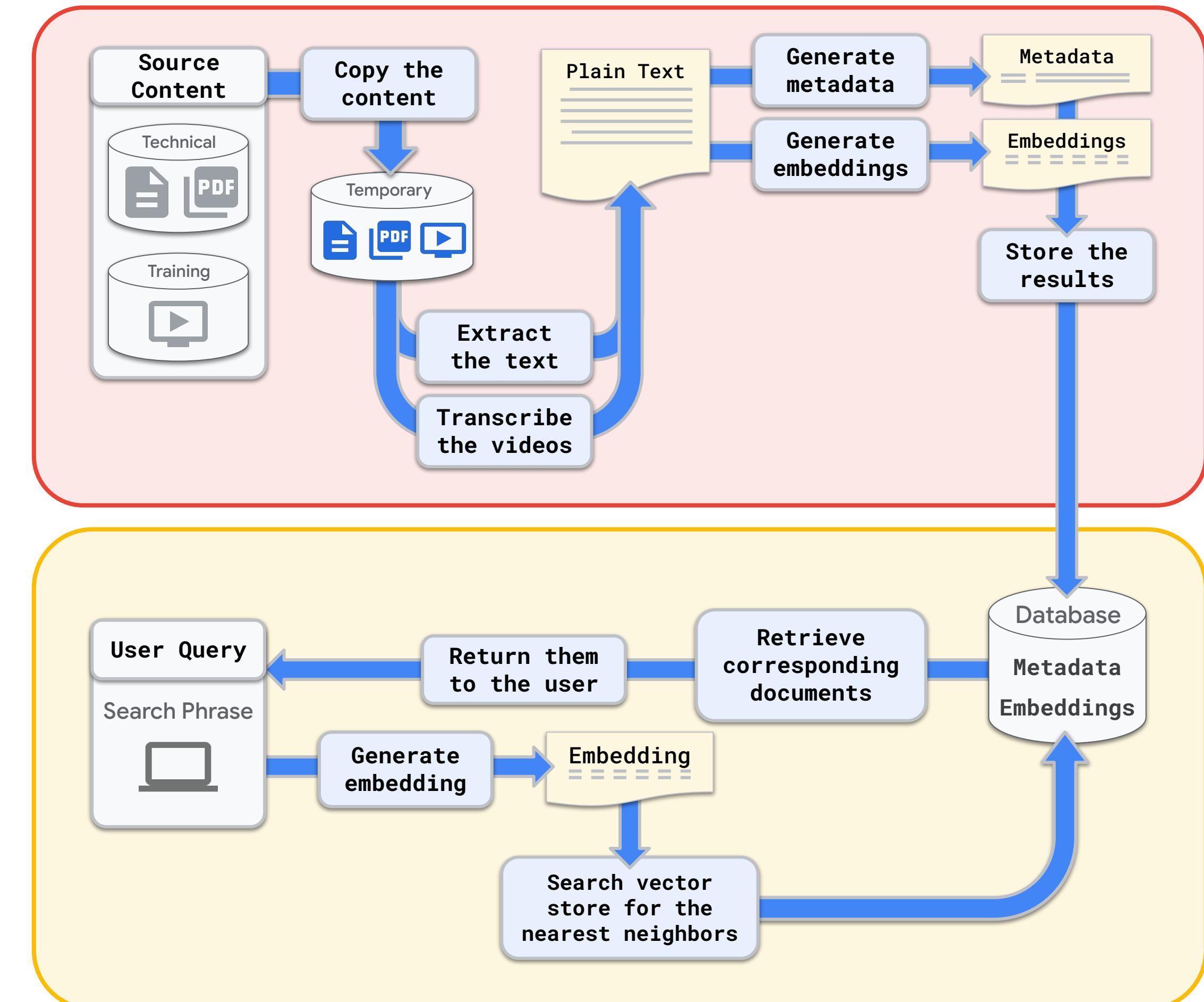
# Conceptual design

## Business Goals:

 Content Management  
Minimize human involvement in the content ingestion process

 Content Discovery  
Increased searchability and discoverability of our content

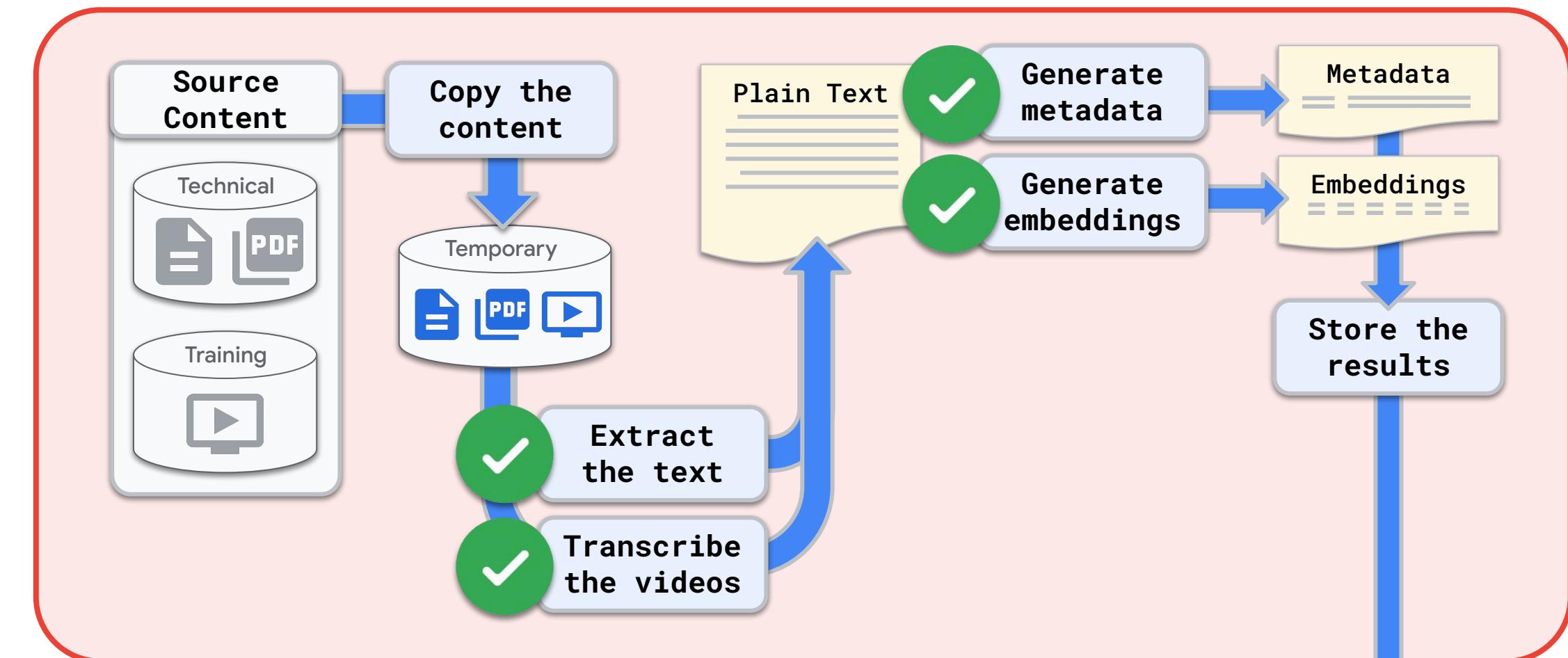
 Viewership / Engagement  
Increased content viewership and engagement



# Conceptual design

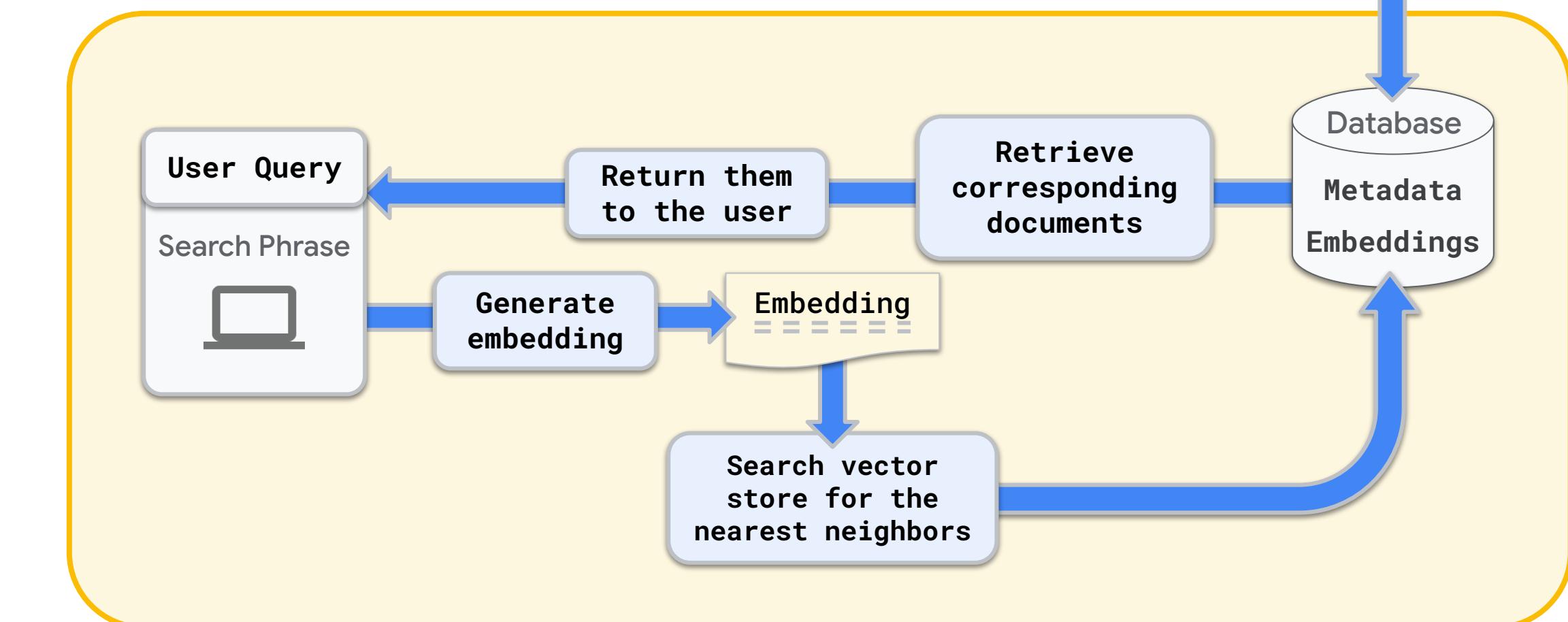
## Business Goals:

- Content Management
- Minimize human involvement in the content ingestion process



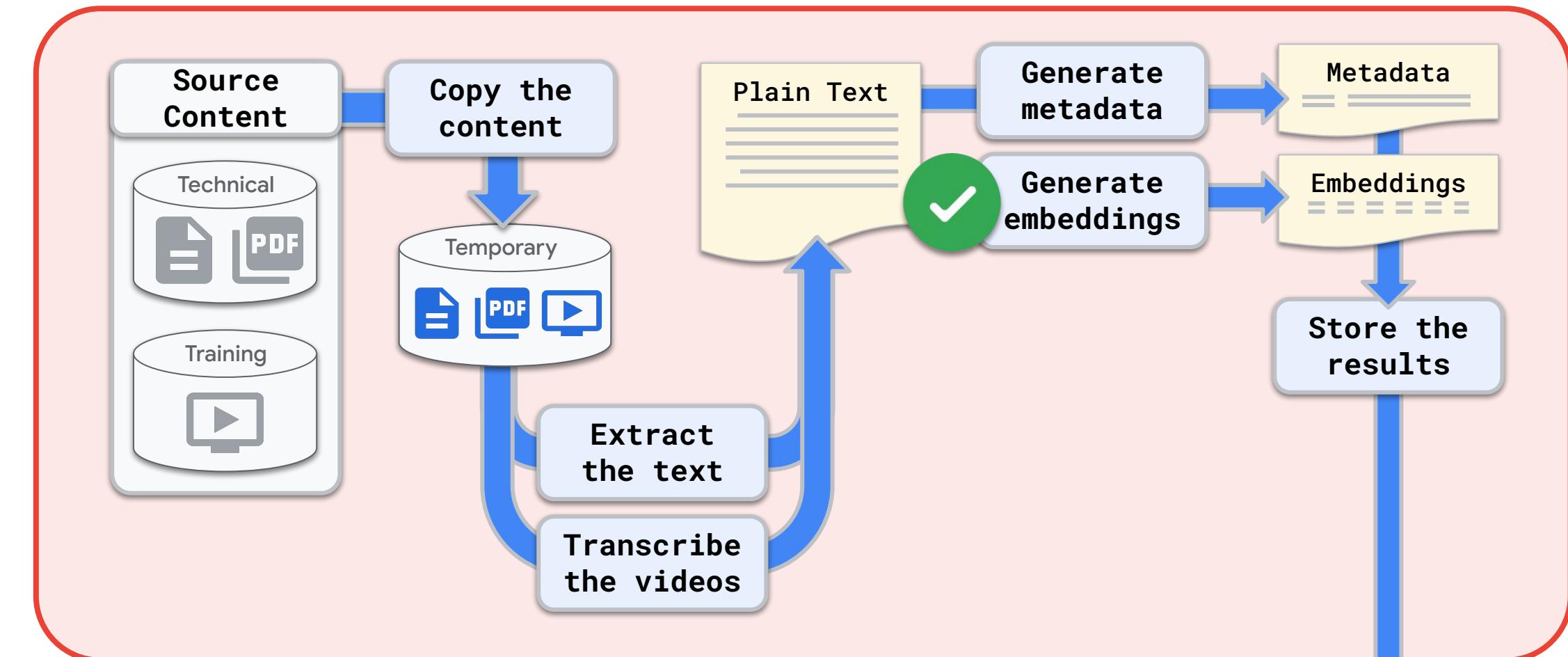
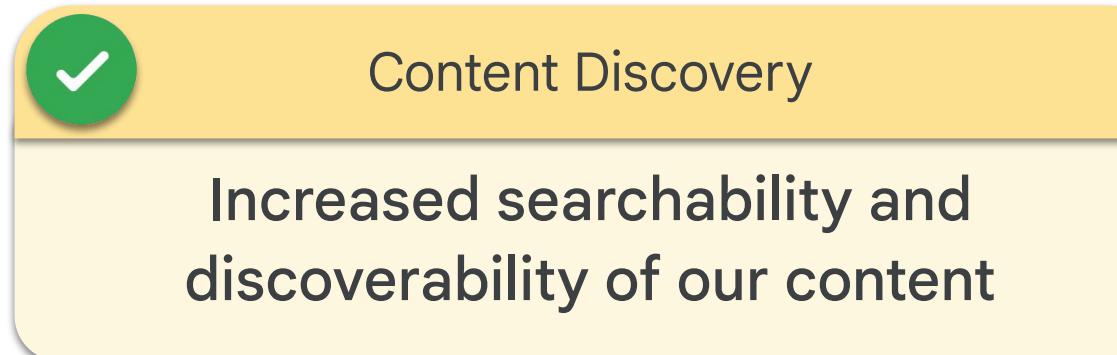
## Achieved by:

- Automating the text parsing, extraction, and segmentation.
- Using an LLM to generate the metadata and embeddings.



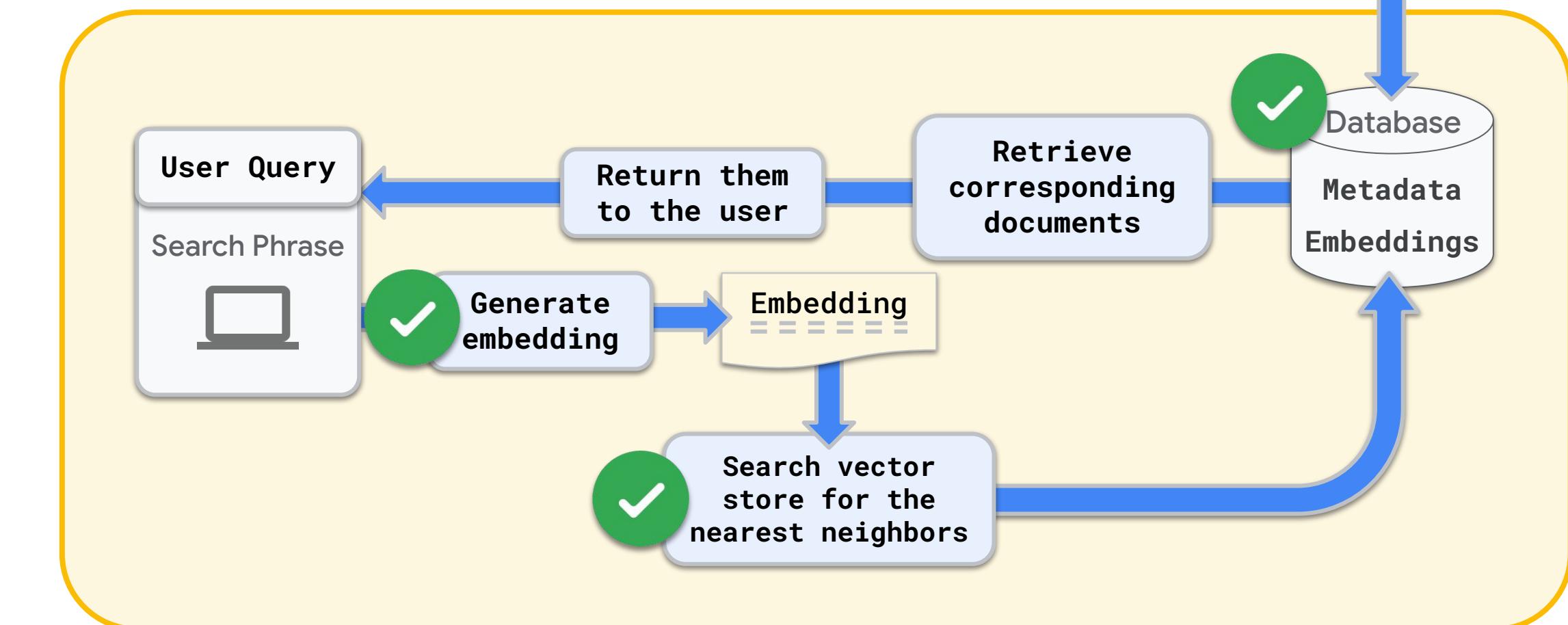
# Conceptual design

## Business Goals:



## Achieved by:

- Indexing the content by its text embeddings value.** (marked with a green checkmark)
- Generating the text embeddings value for the user's query.** (marked with a green checkmark)
- Returning the list of 'nearest neighbors' from the vector store.** (marked with a green checkmark)



# Conceptual design

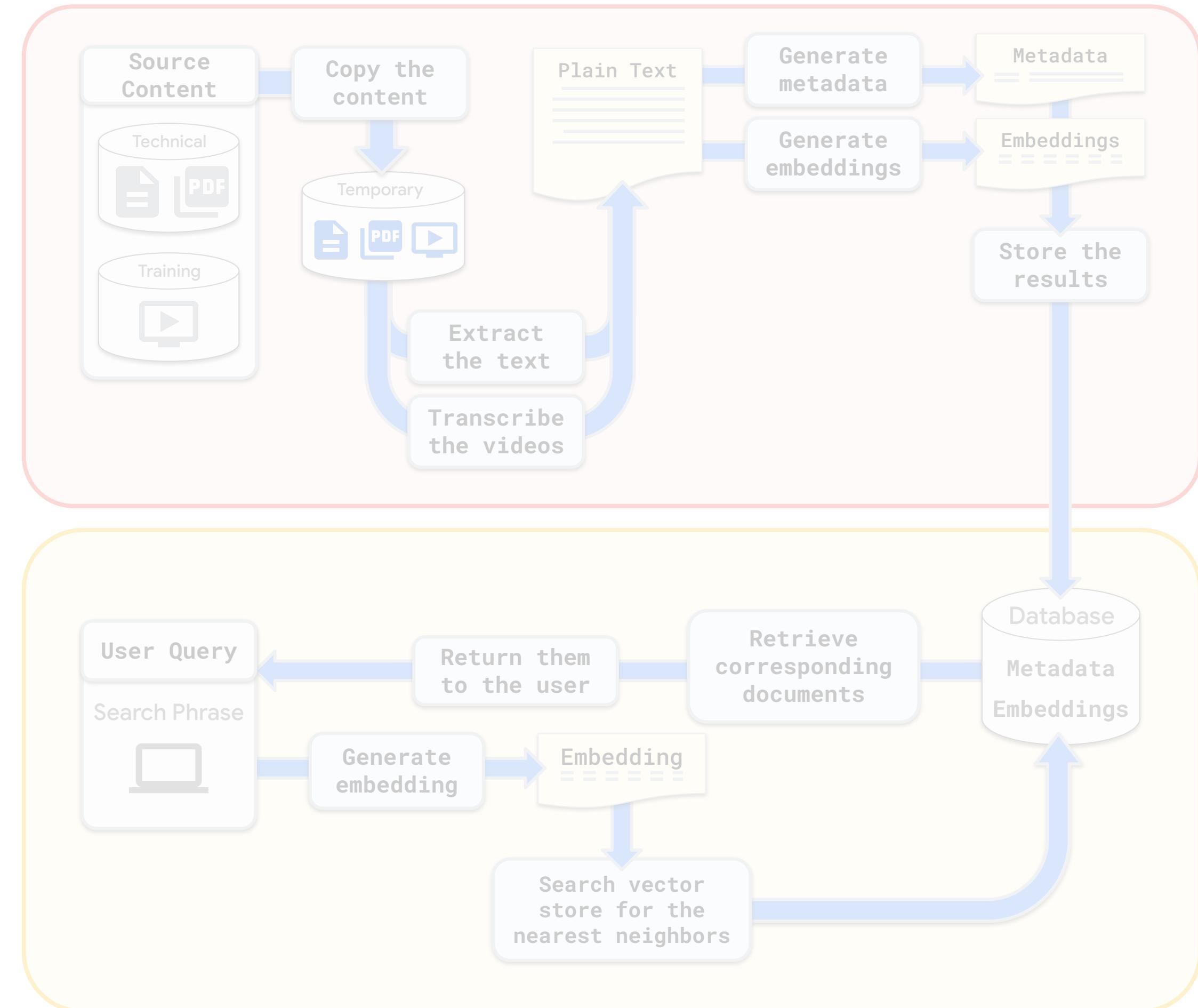
## Business Goals:



## EXCLUDED FROM SCOPE

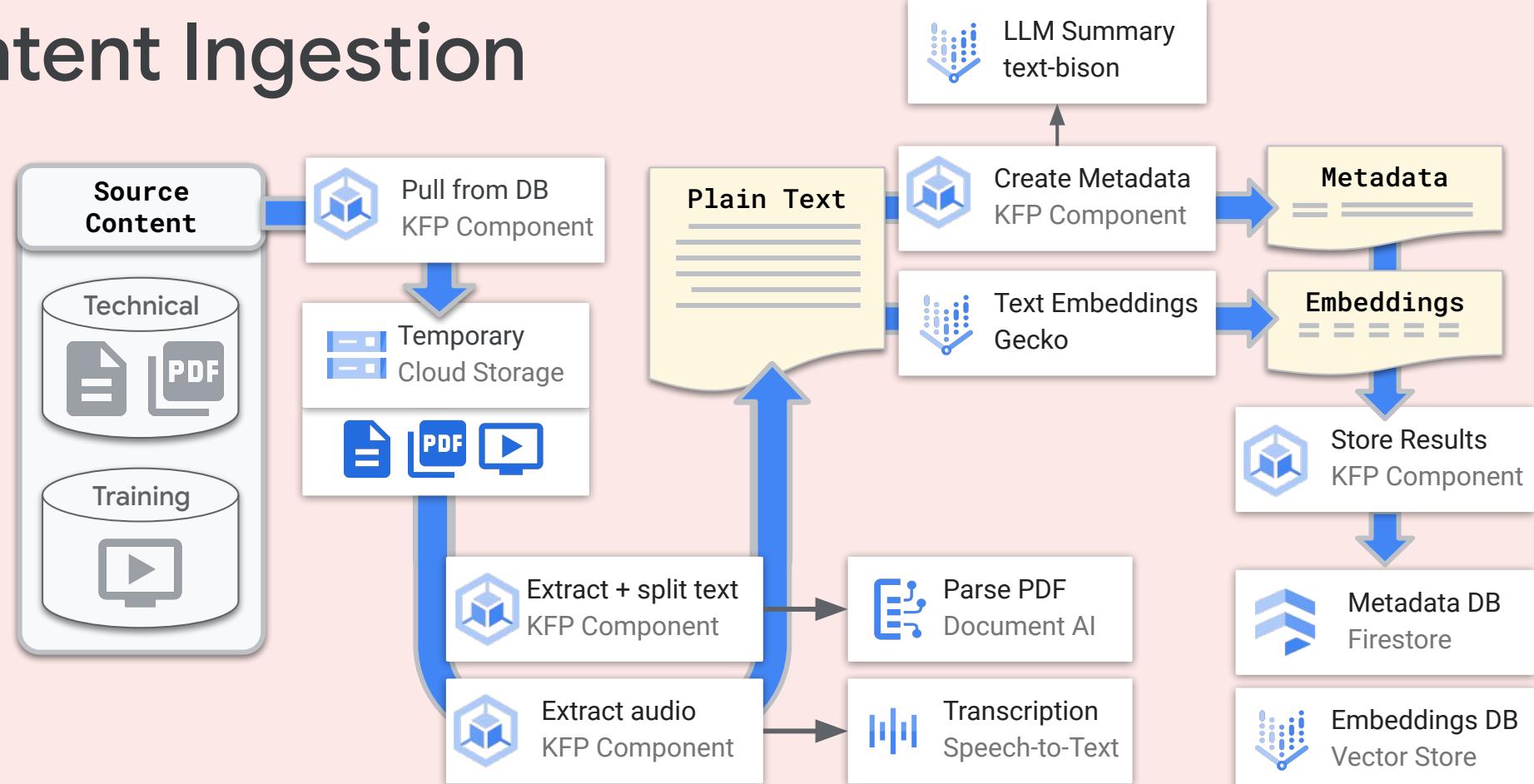
## Use recommendations engine to:

- + Suggest additional content based on similarly profiled users' activity (collaborative filtering).
- + Recommend content also based on the user's viewing history.

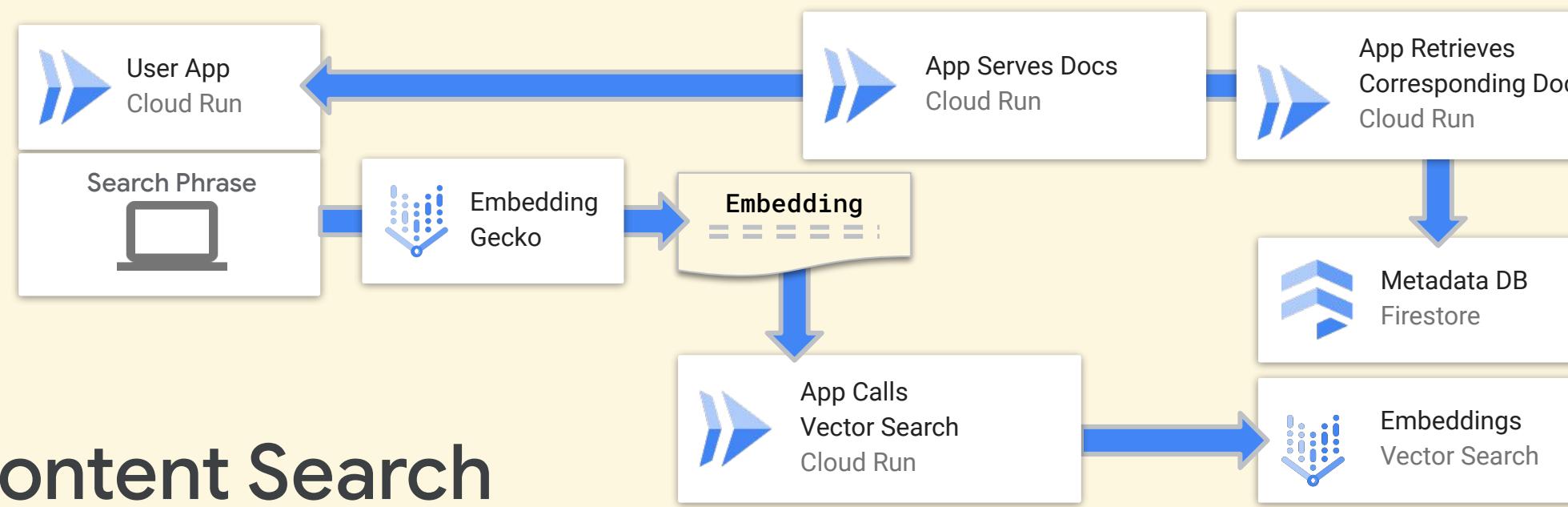


# Component design

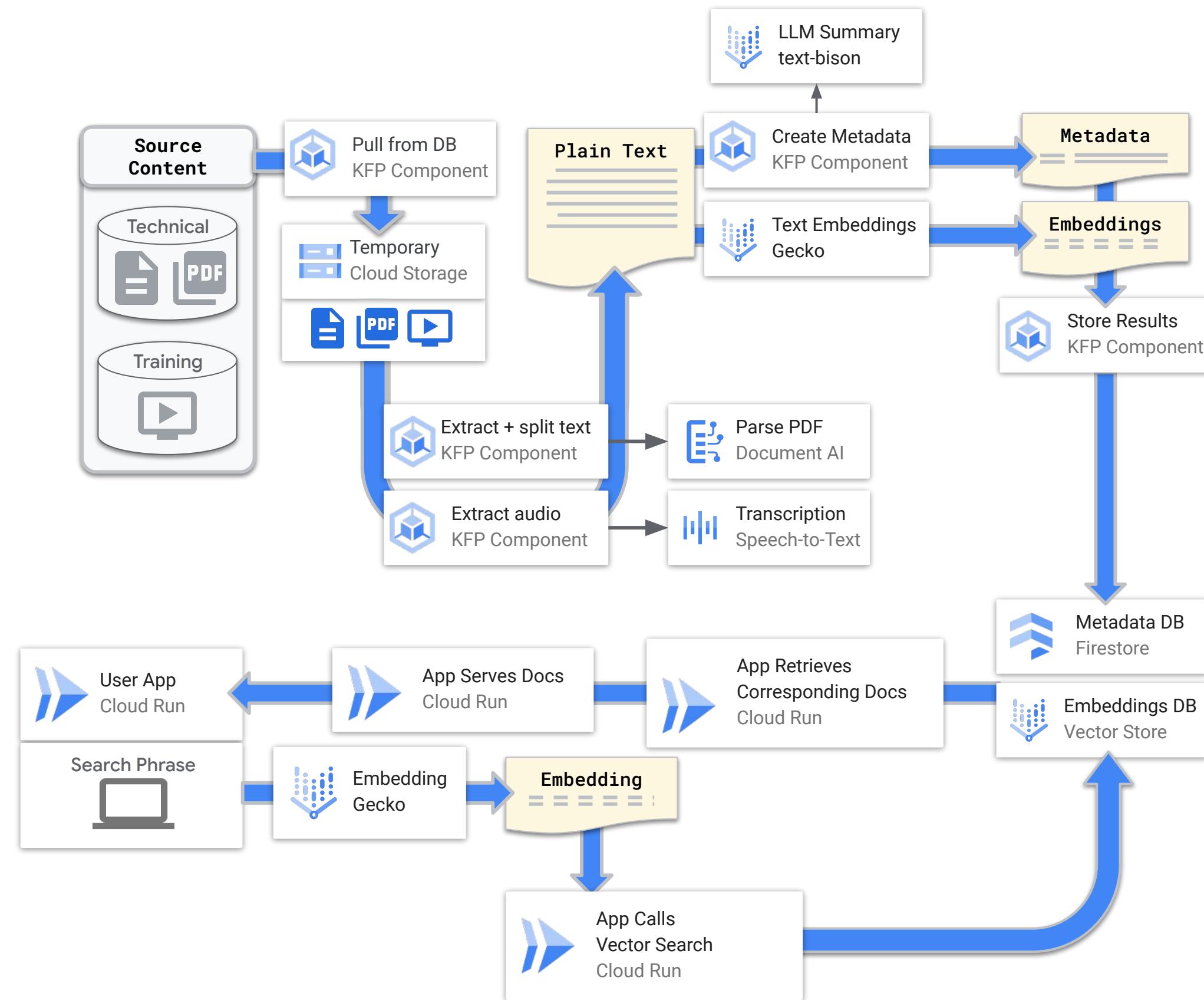
## Content Ingestion



## Content Search



# Solution overview



# In this module, you learned to ...

- 01 Evaluate customer requirements and propose a solution design.
- 02 Architect a Retrieval Augmented Generation (RAG) solution
- 03 Extract text from multimedia content & generate embeddings.
- 04 Define the work streams, pipelines, and components used in your solution.
- 05 Articulate your design decisions in the context of SRE principles.



**Google** Cloud