Assignment: Research Platform Thinking & Execution

Context

You are building a **research platform** that manages and extracts insights from **large volumes of unstructured data** (research papers, reports, filings, PDFs in mixed formats) on companies and industries.

The goal is to build a **research platform** that allows to efficiently organise and explore these documents, extract insights, make dashboards and answer questions around it.

Your Task

In **2-3 pages (or a concise structured write-up)**, describe how you would approach this problem. Focus on **system thinking, execution planning, and real-world constraints**.

Please cover the following:

1. **System structure**
   Describe the key components of the system and how data flows between them (ingestion, storage, processing, retrieval, usage). Submit a hand drawn diagram for the system architecture showing the flow of data.

2. **System design decisions**
   Call out a few important system design decisions you would make early, and explain *why*. Also clearly mention **at least one reasonable approach you would *not* choose**, and why you would avoid it.

3. **Tech stack choice**
   Choose a specific tech stack for this and give your reasoning on the choice of the particular stack over the alternatives.

4. **Use of AI**
   Where would you apply AI in this system, and where would you deliberately avoid it? Explain your reasoning.

5. **Execution plan**
   What would you build first, and what would you defer? How would you sequence the work? And why?

6. **Scale & failure points**
   What do you expect to break first as data volume or usage grows? How would you address those issues? Identify the **top 2–3 technical or product risks** you expect in

months 1–6 as data or usage grows. Describe how you would mitigate or monitor each of them?

7. **Cost breakdown**

   A. What would the rough infrastructure costs(Server, LLM, Parsing, Storage, etc) for each part of the stack in the below scenarios(Please give a breakup of costs for each tool used).

Scenario	Number of Companies	Size in Gb	Pages
1	1000	2,970	1,08,00,000
2	3000	8,910	3,24,00,000
3	5000	14,850	5,40,00,000

B. Now if you had to limit the infrastructure cost to $400 per month in the 3rd scenario.

- What specific system design changes would you make(for example, the stack, processing or storage sequences) to meet these constraints(mention the exact changes you would want to make and its impact).
- What would be the tradeoffs and risks introduced by these changes?
- What would be the best way to handle the specific tradeoffs and risks?

Guidelines

- Prioritize **clarity of thought, structure, and decision-making**
- Explain tradeoffs and assumptions where relevant

What We're Evaluating

- How you break down an ambiguous problem
- How can you plan and make system level decisions
- How you think about workflows and data flow
- Your judgment around scale, constraints, and tradeoffs
- Your ability to communicate complex ideas clearly

Note

There are no “right answers.” We are interested in **how you think**, not whether your approach matches a specific solution.