I am currently a final-year engineering student, and this write-up explains how I would approach building a research platform in a practical and scalable way. The focus of this approach is simplicity, clarity, and gradual improvement rather than overengineering from the beginning.

Context: The goal of this research platform is to manage large volumes of unstructured data such as PDFs, reports, and research documents, and convert them into useful insights for users. These documents may come in different formats and need to be processed, stored, and queried efficiently.

System structure

- ### Data ingestion

  The first step of the system is data ingestion. Users can upload documents such as PDFs, reports, or text files through a simple interface. These files are validated and stored securely for further processing.

- ### Storage

  Once uploaded, the documents are stored in a centralized storage system. Metadata such as file name, upload date, and document type are stored in a database to make retrieval easier.

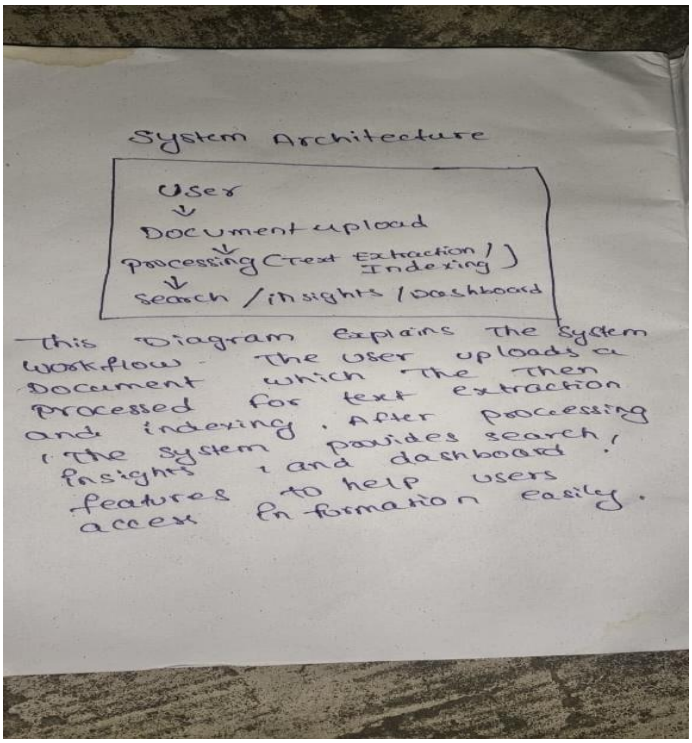
- ### Processing

  After storage, the system processes the documents to extract text and relevant information. Basic text extraction and indexing are performed so that the content can be searched and analyzed later.

- ### Insights & usage

  Processed data can then be used to generate insights, answer user queries, and display simple dashboards. Users can search documents, view summaries, or ask questions related to the stored data.

System architecture



A simple hand-drawn diagram has been included to explain the data flow in the system.

- Execution plan

I would build this system in phases. In the initial phase, the focus would be on basic document upload and storage. In the next phase, text processing and search functionality would be added. Finally, basic analytics and insight generation features would be implemented based on user requirements.

- Real-world constraints

  Some challenges include handling large file sizes, maintaining data quality, and ensuring good performance. Cost and security are also important factors, especially when dealing with sensitive documents. To manage these issues, I would start with simple solutions and optimize them over time.
- Conclusion

This approach focuses on building a practical and scalable research platform while keeping the design simple. As I gain more real-world experience, the system can be improved further, but this provides a strong foundation to start with.