Em xin báo cáo với anh các task đã hoàn thành trong ngày 25/7/2024

• Cài đặt và viết guideline cho công cụ Redis cluster (Docker) (Đã hoàn thành)

• Cài đặt và viết guideline cho công cụ Elastic (Docker) (Đã hoàn thành)

• Cài đặt và viết guideline cho công cụ Kafka (Docker) (Đã hoàn thành)

Em xin báo cáo với anh các task dự kiến trong ngày hôm nay (26/7/2024)

• Vẽ lại luồng biểu mẫu chính (Đã hoàn thành)

• Đọc, setup về Kafka Consummer (Đã hoàn thành)

• Đọc, setup về Logstash (Đã hoàn thành)

Em xin báo cáo với anh các task dự kiến trong ngày hôm nay (27/7/2024)

* Thiết kể lại luồng xử lý (Đã hoàn thành)
* Hiểu rõ hơn bài toán xử lý (Đã hoàn thành)

Em xin báo cáo với anh các task dự kiến trong ngày hôm nay (28/7/2024)

* Thiết kể lại luồng xử lý (Đã hoàn thành)
* Hiểu rõ hơn bài toán xử lý (Đã hoàn thành)

Em xin báo cáo với anh các task dự kiến trong ngày hôm nay (29/7/2024)

* Thiết kể lại luồng xử lý (Đã hoàn thành)
* Hiểu rõ hơn bài toán xử lý (Đã hoàn thành)

Em xin báo cáo với anh các task dự kiến trong ngày hôm nay (30/7/2024)

    1. Viết lại mô tả hoàn chỉnh về sản phẩm và gửi lại cho anh Thế Anh, nội dung mô tả bao gồm  
        + Mô tả mục tiêu, mục đích của sản phẩm (Đã hoàn thành)  
        + Mô tả đầu vào, đầu ra của sản phẩm (Đã hoàn thành)

        + Mô tả các thành phần cấu thành nên sản phẩm và công nghệ sử dụng (Đã hoàn thành)

        + Mô tả các giao diện của sản phẩm (Đã hoàn thành)

        + Mô tả luồng xử lý tổng quan của sản phẩm (Đã hoàn thành)

        + Mô tả các thành phần chức năng có trong sản phẩm (Đã hoàn thành)

    2. Sau khi đã gửi bản tài liệu cho anh Thế Anh, thảo luận với anh Thế Anh để thực hiện các công việc sau:  
        + Chỉnh sửa các nội dung chưa đúng hoặc chưa hợp lí liên quan đến sản phẩm (Đã hoàn thành)  
        + Thảo luận bắt đầu triển khai từng thành phần của sản phẩm (Đã hoàn thành)

Em xin báo cáo với anh các task dự kiến trong ngày hôm nay (31/7/2024)

1. Hoàn thành tiếp bản mô tả sản phẩm bao gồm:  
            \_ Mô tả và thiết kế luồng kết nối cho công cụ Query to RDBMS gồm có:

                    + MySQL    + PostGreSQL + MSSQL + Oracle (Đã hoàn thành)  
            \_ Mô tả và thiết kế luồng kết nối cho công cụ Query to NoSQL gồm có:  
                    + Cassandra    + HBase    + Elastic Search    + MongoDB (Đã hoàn thành)  
            \_ Mô tả và thiết kế luông kết nối cho công cụ Query to Visualize gồm có:  
                    + Excel to Tableau    + RDBMS to Tableau (Dự kiến)  
            \_ Mô tả và thiết kế luồng cho công cụ Recommend gồm có:

                    + Planning Advisor (Dự kiến)  
            \_ Mô tả và thiết kế luồng cho công cụ Extract from Kafka (dự kiến)

            \_ Mô tả và thiết kế luồng cho công cụ Extract from File (dự kiến)

            \_ Mô tả và thiết kế luồng cho công cụ Extract from Tabular (dự kiến)  
  
    2. Dựng cơ sở hoàn chỉnh:  
            \_ Mục tiêu: Đảm bảo các cơ sở hạ tầng phục vụ cho sản phẩm được dựng lên hoàn chỉnh và tương tác được với nhau bao gồm các thành phần sau:  
                    + Source Data Layer

                            \* RDBMS: MySQL, PostgreSQL, MSSQL, Oracle   
                            \* NoSQL: Cassandra, HBase, ElasticSearch, MongoDB  
                    + Injestion Layer:

                            \* Batch processing: Nifi

                            \* Stream processing: Kafka  
                            \* Log processing: Logstash

                    + Computing/ Processing Layer:

                            \* Spark

                    + Storage Layer:

                            \* Hot data: Redis Cluster

                            \* Cold data: Elastic Search

                    + Serving:

                            \* Framework: Langchain

                            \* LLM model: Gemini, GPT

                            \* Visualization: Tableau

Em xin báo cáo với anh các task dự kiến trong ngày hôm nay (1/8/2024)

1. Kham thảo demo của Mai Anh về tool NLP2SQL

2. Xây dựng hệ thống cho công cụ đầu tiên gồm:  
\_ Thiết kế ui cho bảng công cụ tổng quan  
\_ Thiết kế ui cho công cụ MySQL query  
\_ Dựng cơ sở hạ tầng gồm:  
    + MySQL,  
    + Nifi,

    + Spark,

    + Langchain ,

    + Gemini & gpt api,  
    + Redis with Rag,  
    + Backend & Frontend  
    + Hive

Em xin báo cáo các công việc sẽ hoàn thành trong ngày hôm nay(2/8/2024):

    1. Xây dựng hệ thống cho công cụ đầu tiên:

        \_ Xây dựng luồng lưu trữ metadata trên Hive để đảm bảo hiệu suất truy vấn  
        \_ Tìm hiểu và triển khai các kĩ thuật RAG và lưu trữ trên Redis bao gồm:

            + Rank RAG

            + Grapgh RAG

            + Advance RAG  
        \_ Thêm LLMA3 và Mítral7B vào model LLM  
        \_ Xây dựng nhanh ui cho trang công cụ tổng quan và tool sử dụng reactjs và Springboot.

em xin báo cáo các công việc sẽ hoàn thành trong ngày hôm nay(5/8/2024):

1. Xây dựng hệ thống cho công cụ đầu tiên:

\_ Xây dựng React Frontend cho trang công cụ tổng quan và Công cụ NLP to MySQL Query đầu tiên

\_ Xây dựng Spring Boot phục vụ cho backend

\_ Xây dựng Jenkin để kết nối giữa Spring Boot và các module khác của Docker Compose  
  
  
  
6/8/2024:  
\_ Frontend:

+ Dựng id session  
 + Dựng params DB để nhập input  
 + Dựng chatbot  
 + Dựng luồng kết nối về backend  
  
\_ Backend:  
 + Dựng CORS để kết nối giữa front end & backend  
 + Dựng API model

7/8/2024

\_ Datahub:  
 + Viết luồng xử lí  
 + Viết GraphSQL để trích xuất metadata  
\_ Backend:  
 + Tích hợp luồng xử lí datahub  
 + Fix lỗi tích hợp luồng xử lí  
  
  
22/8/2024  
Backend:

+ Modularize datahub ingestion service

+ Modularize DBconnection

+ Modularize LLM model api call