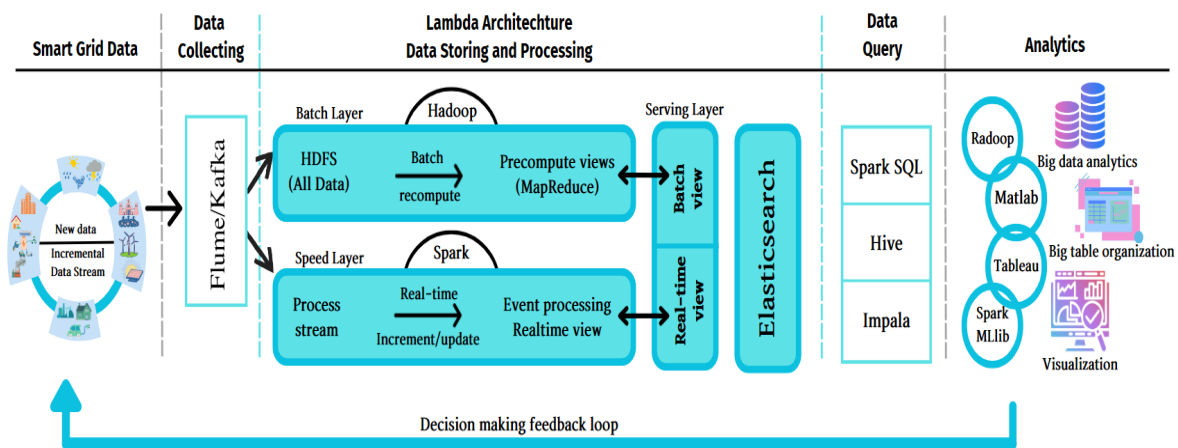# Final Project

1. Requirements

- Architecture: Building a lambda architecture for storing and processing big data including the following components: data crawling, storage, analytics, and visualization.

System architecture

- o Data Collecting: Kafka, ..
- o Database: GG bucket, Elasticsearch, MongoDB, …
- o Data query and process: Spark, …
- o Analytics: Kibana, …
- Problems: Select one of the problems in Section **5**

2. Registration

Fill in the following Google docs file. Note that up to 4 groups are allowed to register the same problem. A group can have up to 4 members.

https://docs.google.com/spreadsheets/d/1mlcJQvIl8HuWKldoK1HAr621WtRdTnmP/edit?usp=sharing&ouid=106996255610935127545&rtpof=true&sd=true

Deadline: Oct 29, 2024

3. Checkpoints

There will be a few checkpoints. On those days, each group must present the work in progress. You must pass all the checkpoints before having the final presentation.
The schedule for checkpoints will be announced later.

4. Final submission and presentation.

You are required to submit the report, source code, and presentation slides. The submission link will be announced later. On the presentation day, all members of the group must attend the class and each member is recommended to present his/her own part.
Deadline: Dec 22, 2024

5. Problems
5.1 Evaluation of Twitter Account Quality for KOLs in the Web3 Space

In Web3 marketing, Key Opinion Leaders (KOLs) on social media platforms like Twitter are crucial for driving trends, building communities, and shaping opinions. This project aims to develop a robust framework to assess the quality of Twitter accounts operated by KOLs in the Web3 space, providing insights into their influence, effectiveness, engagement, and credibility.

Requirements:

- **Identification of KOL Twitter Accounts:** Compile a comprehensive list of Twitter accounts belonging to KOLs being actively involved in the Web3 space.
- **Data Collection Workflows:** Develop and implement automated workflows for collecting relevant data from Twitter.
- **Deployment of Machine Learning, AI, and Deep Learning Models:** Utilize advanced techniques in machine learning, artificial intelligence, and deep learning to evaluate the quality of each KOL's Twitter account.

5.2 Evaluation of Reputation Score and Trending Projects for NFTs, Tokens, and Web3 Projects on Social Media

In the rapidly evolving Web3 ecosystem, the reputation and popularity of NFTs, tokens, and projects play a significant role in driving adoption and investment. This project evaluates the reputation scores and trending status of various NFTs, tokens, and projects based on their social media presence and community engagement.

Requirements:

- **Data Collection from Web3-related Social Media Platforms:** Implement processes to gather comprehensive data from various social media platforms where discussions about NFTs, tokens, and Web3 projects take place.
- **Integration of Data from Multiple Platforms:** Combine and integrate the collected data from different social media sources to create a unified dataset.
- **Evaluation of Popularity and Reputation in Communities:** Analyze the integrated data and apply ML or DL models to assess the popularity and reputation of specific NFTs, tokens, and projects.

5.3 Twitter KOLs and Telegram Contributors Recommendation System

To drive growth and engagement in Web3 projects, identifying and connecting with the right influencers and potential contributors is essential. This project aims to suggest and link Key Opinion Leaders (KOLs) and potential collaborators from platforms like Twitter and Telegram to specific projects, helping to boost visibility and foster partnerships.

Requirements:

- **Data Collection from Twitter, Discord, and Telegram:** Implement workflows to gather relevant data from social media platforms such as Twitter, Discord, and Telegram.
- **Labeling Accounts as KOLs or Potential Contributors**
- **Building a Recommendation System for KOLs and Contributors:** Develop a recommendation system to suggest KOLs and potential collaborators for Centic's list of projects.

5.4 Post Suggestions for Web3 Influencers and Marketers

Creating engaging and informative content is essential for Web3 influencers and marketers to build credibility and grow their audience. This project focuses on generating impactful post suggestions based on blockchain data analysis and market trends, helping influencers and marketers maintain a strong presence on platforms like Twitter and Telegram.

Requirements:

- **Blockchain Data Analysis Provided by Centic:** Use the blockchain data provided by Centic to perform an in-depth analysis of market trends, user behavior, token movements, and project performance.

- **Utilize NLP and LLM Models for Automated Post Generation:** Employ Natural Language Processing (NLP) and Large Language Models (LLMs) to automatically generate well-structured posts for Twitter, Telegram, and other platforms.
- **Integrate Analysis with Post Generation and Suggest Content for Influencers and Marketers:** Combine the analytical insights with automated content generation to create comprehensive and insightful posts.

5.5 Tracking and Evaluating the Effectiveness of KOLs and Contributors with Corresponding Projects

To ensure that Key Opinion Leaders (KOLs) and contributors are effectively supporting specific Web3 projects, this project focuses on tracking and evaluating their performance based on social media engagement. The goal is to assess how well these influencers and collaborators promote and contribute to the growth of Centic's projects.

Requirements:

- **Data Collection from Twitter and Telegram:** Implement systems to collect relevant data from Twitter, Telegram, and other social media.
- **Mapping KOLs and Contributors to Projects Provided by Centic:** Match KOLs and contributors to the specific projects they are promoting or associated with.
- **Evaluating the Effectiveness of Tweets and User Interactions with Projects:** Analyze the performance of KOLs and contributors by evaluating the effectiveness of their tweets, posts, and interactions with users regarding the projects.

5.6 Cash Flow Analysis and Investment/Trading Strategy Recommendations on DEXs

In the decentralized finance (DeFi) space, understanding cash flow patterns and developing effective trading strategies are key to success. This project focuses on analyzing cash flow and providing investment and trading strategy recommendations, leveraging data from Centic and additional sources in the Web3 ecosystem.

Requirements:

- **Utilize Centic Data and Collect Additional Web3 Data**: Analyze the financial and transactional data provided by Centic, and supplement it with additional data collected from decentralized exchanges (DEXs) and relevant blockchain platforms.

- **Analyze Cash Flow Patterns on DEXs**: Perform in-depth analysis of cash flow trends on DEXs by tracking large transactions, liquidity movements, and the behavior of prominent wallets.
- **Recommend Investment and Trading Strategies:** Based on the cash flow analysis and other relevant data, provide strategic recommendations for investment and trading.

5.7 Collaboration Suggestions Between Projects

To foster growth and innovation within the Web3 space, collaboration between similar projects can be highly beneficial. This project focuses on identifying opportunities for partnerships between projects by analyzing their data and communities, both from Web3 and Web2 platforms.

Requirements:

- **Collect and Use Data from Web3, Web2, or Both:** Gather data from Web3 sources (such as blockchain data and decentralized apps) as well as Web2 platforms (social media, websites, user communities) for various projects.
- **Evaluate the Similarities Between Projects and Their Communities:** Analyze the data to assess the degree of similarity between different projects.
- **Suggest Collaborations Between Projects with High Similarity:** Based on the analysis, recommend collaboration opportunities between projects that share significant similarities.

5.8 Automated Chatbot System Built from Website Content
Description:
This system automatically analyzes, classifies, and extracts information from any given website. Based on this data, a vector database is created, allowing for the rapid development of a question-answering chatbot. The goal is to generate a chatbot version that can efficiently respond to user inquiries in a short time frame.

Related technique:
- Language models
- Retrieval-Augmented Generation
- Prompt engineering
- Web scraping
- Vector database

5.9 Recommendation system for a chatbot in the blockchain domain

Description:

Building a recommendation system for suggesting follow-up queries to users based on asset information, query history, and the system's goals, such as encouraging users to perform transactions or take actions in a certain direction. This aims to balance between user experience and the system's profit/revenue goals.

Related technique:
- Knowledge graph
- Multi-turn recommendation
- Sequential conversational target-oriented recommend system

5.10 Automated system for building real-time entity data in the blockchain domain
Description:

Entities in the blockchain domain, such as tokens, dApps, etc., are vast in number and are continuously added or updated daily. Currently, there is a significant lack of quality datasets related to blockchain in general and blockchain entities in particular, which are necessary for natural language processing tasks. Therefore, developing an automated system that constructs diverse datasets about entities specifically, and other related data in general, is essential to support the implementation of various additional tasks.

Related technique:
- Language models
- Multi-agent LLM system
- NoSQL and SQL database systems
- Data Integration and Processing