

Final Report

Real-time monitoring of volatile, retail investor-driven price fluctuations in the equity market

Team members :

Huai-Che Chang, Layne Pu, Wan-Yi Mao, Wen-Fang Lu

Outline

- **Problem Statement**
- **Project Objectives**
- **System Structure**
- **Data Source**
- **Data Flow**
- **Technology Used**
- **Front-end UI**

Team Member

Team Member	Role
Huai-Che Chang	Software Developer, Product Manager
Layne Pu	Team Leader, Software Developer
Wan-Yi Mao	Software Developer, DevOps
Wen-Fang Lu	Software developer, Scrum Master

Problem Statement

With the rise of social media, the influence of social media on financial markets is here to stay, as younger generations start saving and investing. Information sharing and discussion on internet platforms can improve market transparency and efficiency. On the other hand, social media platforms are known vehicles of disinformation and manipulation of human behavior.



reference [Carnegie](#)

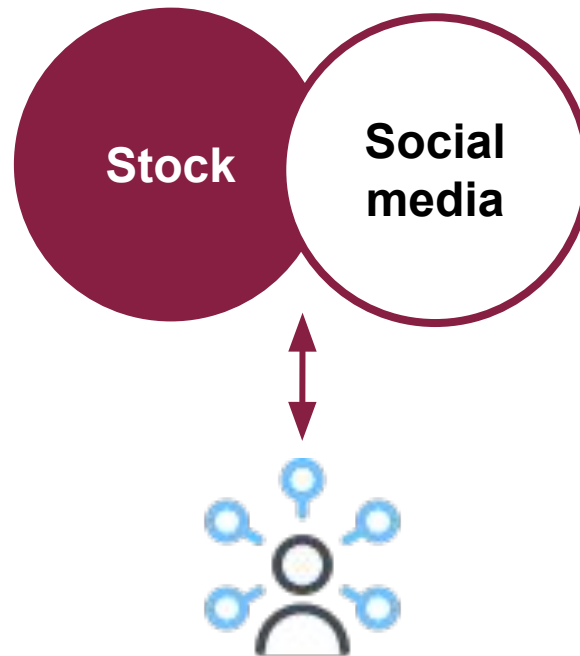


From GameStop To Tesla – How social media is driving the stock market

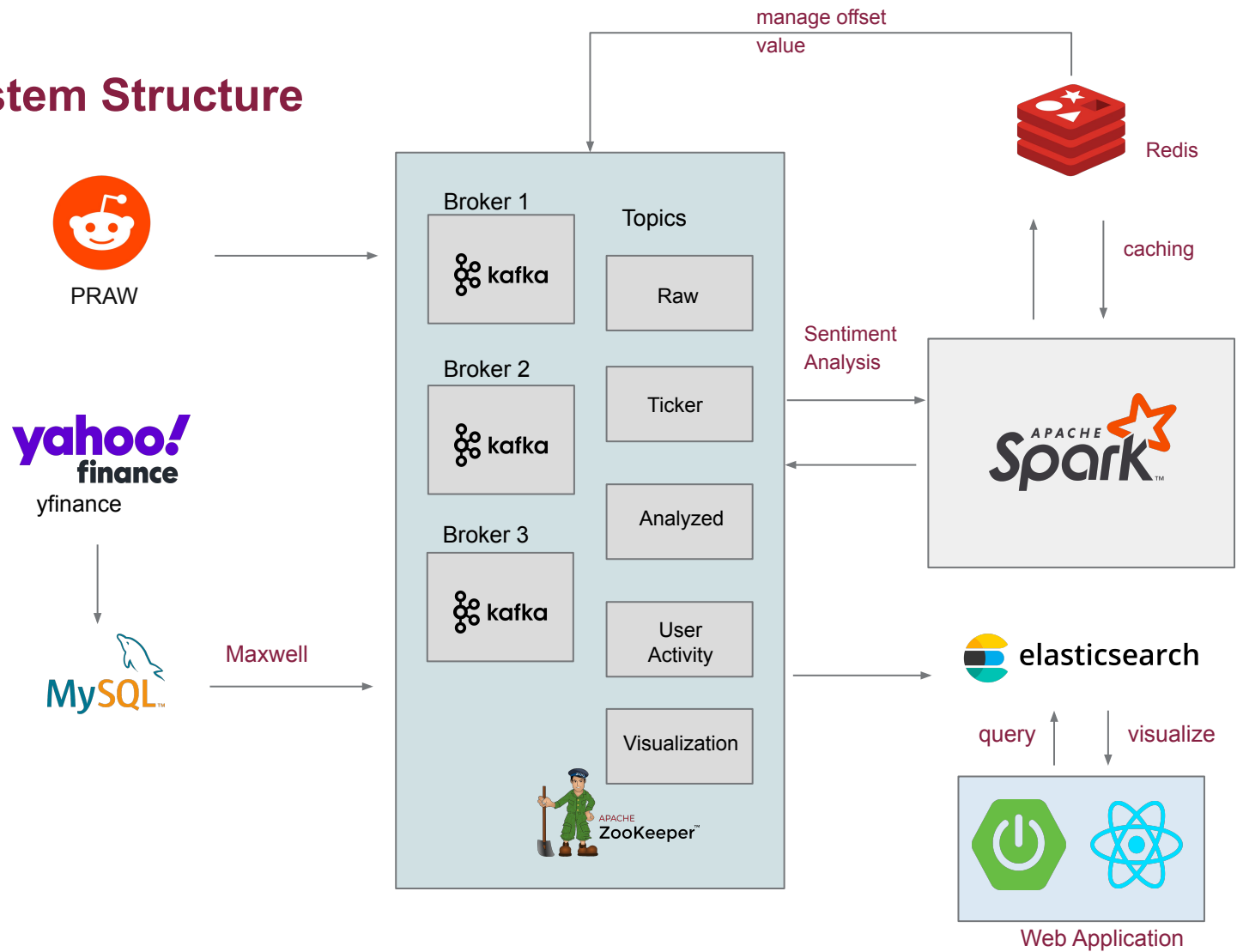
reference
[Maddyness](#)
[news](#)

Project Objective

- Discover and analyze social media impact on stock market trend
- Provide an overview of public view about a specific stock



System Structure



└ Data Source

1. Reddit

- Subreddit
- Post information

2. Yahoo Finance

- Basic stock market information
- Stock history prices
- Stock news

└ Data Flow

1. Reddit Posts

- praw API -> Kafka topics -> sentiment analysis
-> Spark -> Elasticsearch -> web application

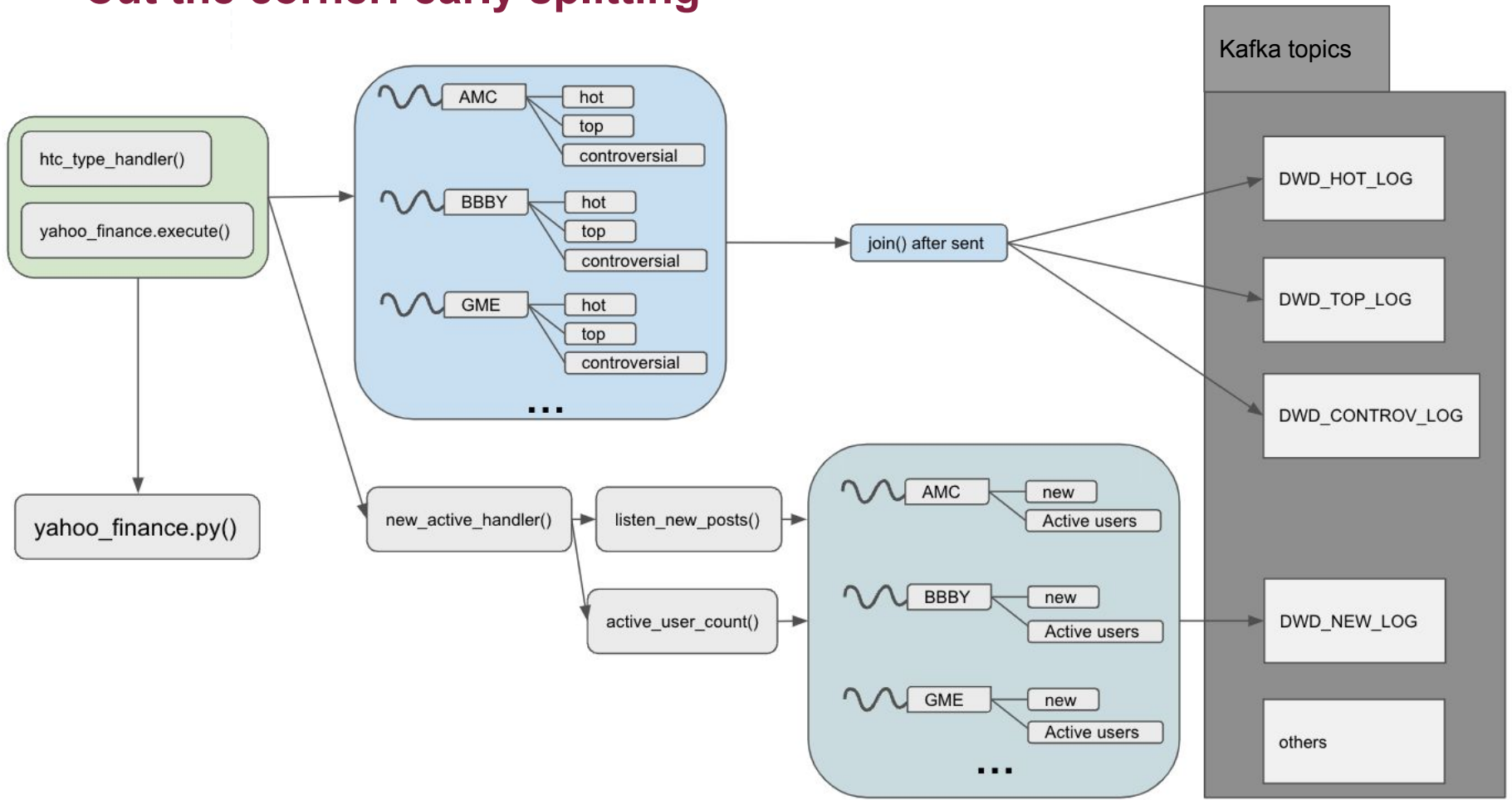
2. Yahoo Finance data

- yfinance API -> MySQL -> web application

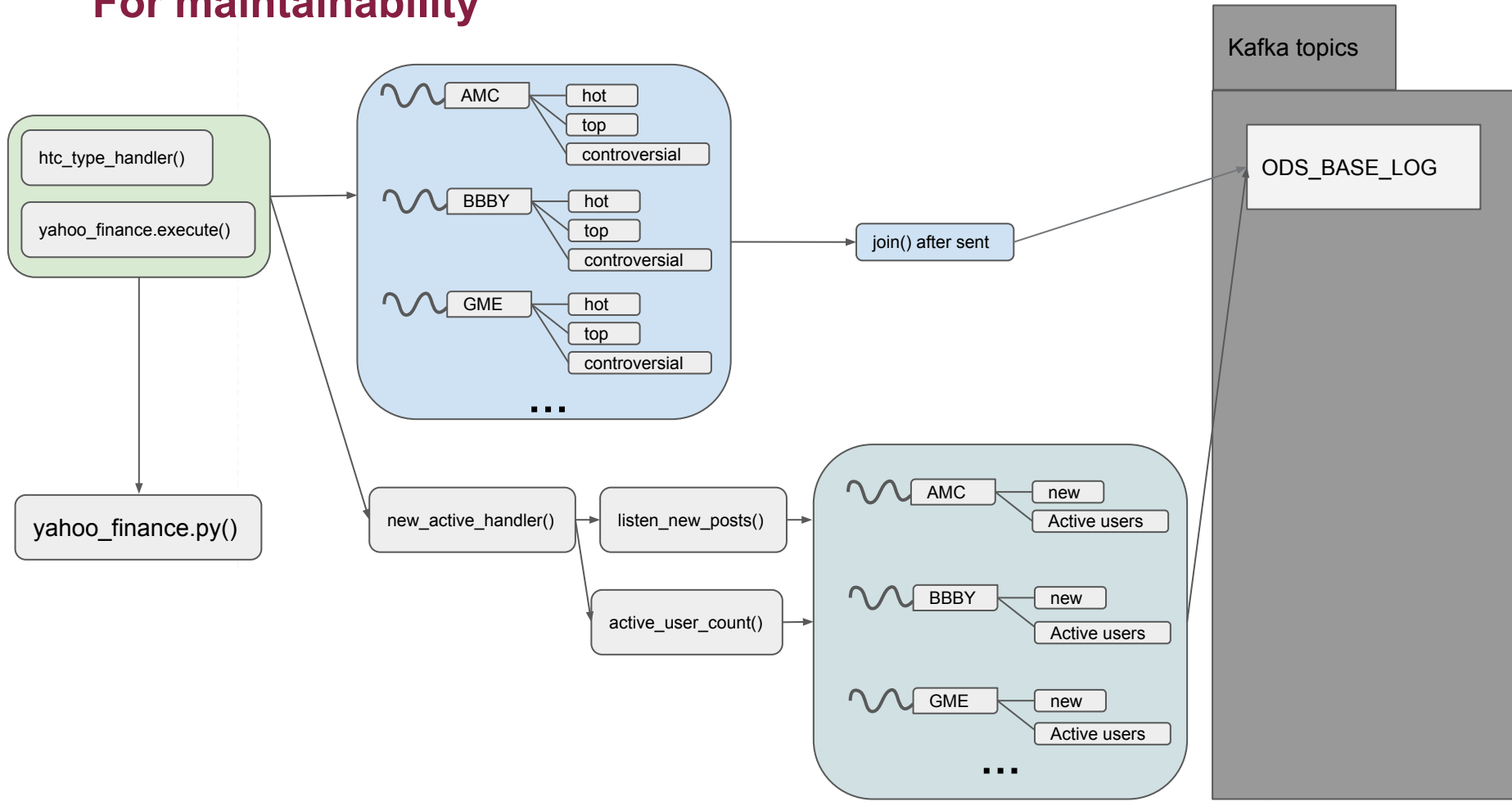
Technologies Used

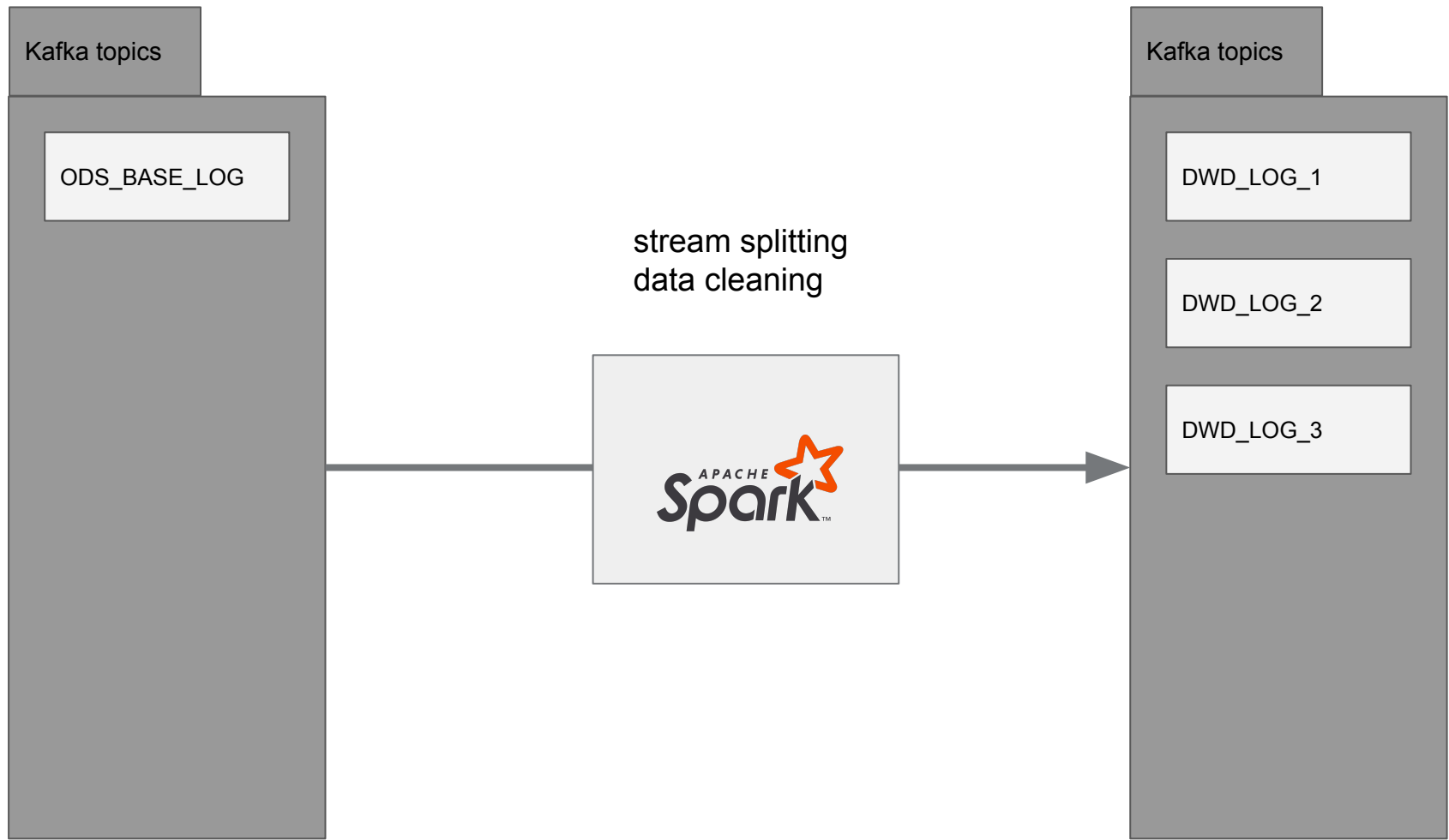
1. yfinance API - scrape data from Yahoo finance
2. praw API - scrape reddit post from target subreddits
3. Spark - operate data streaming and real-time analytics
4. Redis - manage offset to prevent duplicate data
5. Kafka - offer scalable, fault-tolerant and real-time data streaming for data processing
6. Elasticsearch - support full-text search functionality
7. NLP - sentiment analysis using VADER
8. Spring Boot - backend API
9. React - frontend web page

Cut the corner: early splitting

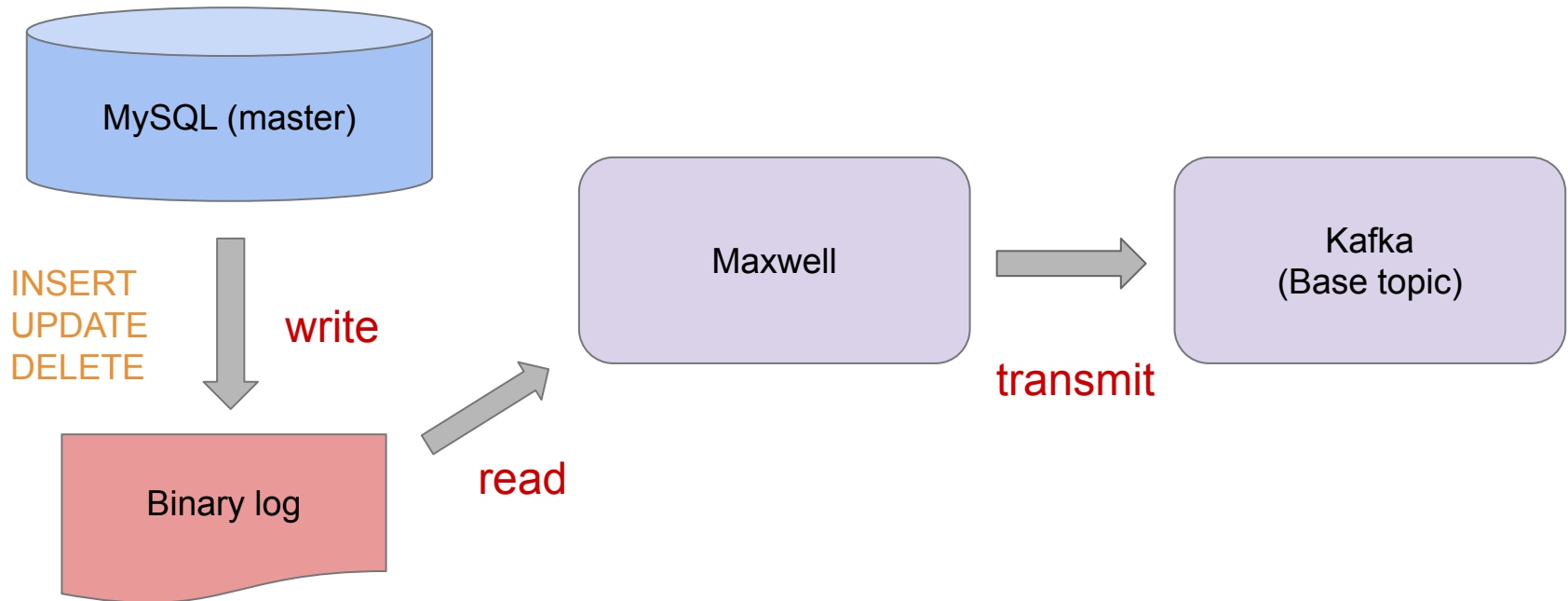


For maintainability



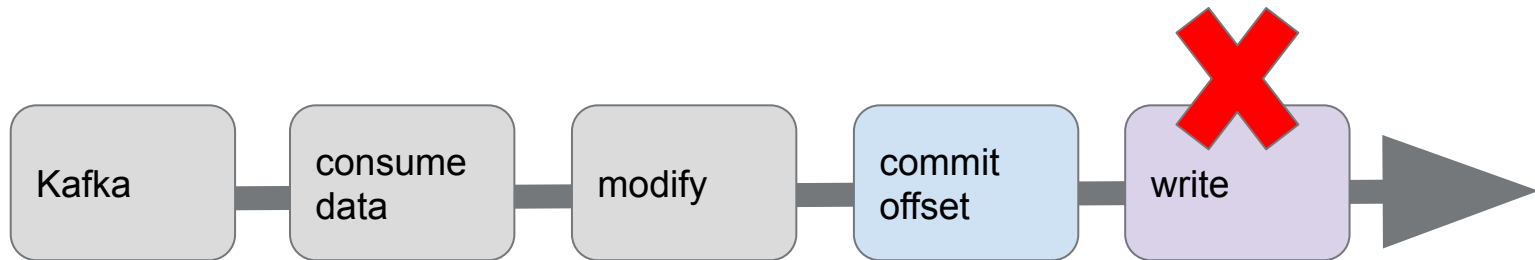


Maxwell: MySQL Data Monitoring and Transmission to Kafka

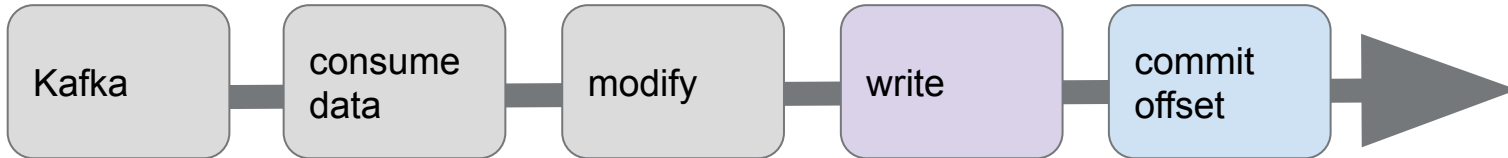


- Incremental synchronization

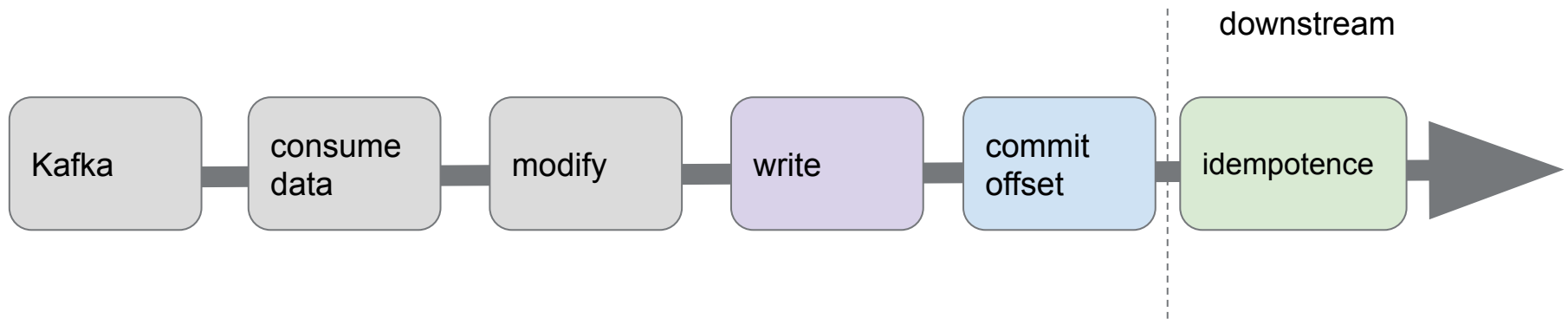
It's all for the consumer - data loss



It's all for the consumer - duplicated consumption



It's all for the consumer - solution



1. read offset from redis
2. use these offset values to consume kafka data
3. record the partition + end of offset
4. use the consumed data
5. submit the offset in step 3 to Redis

Save the offset to Redis

Question: Where is the offset from?

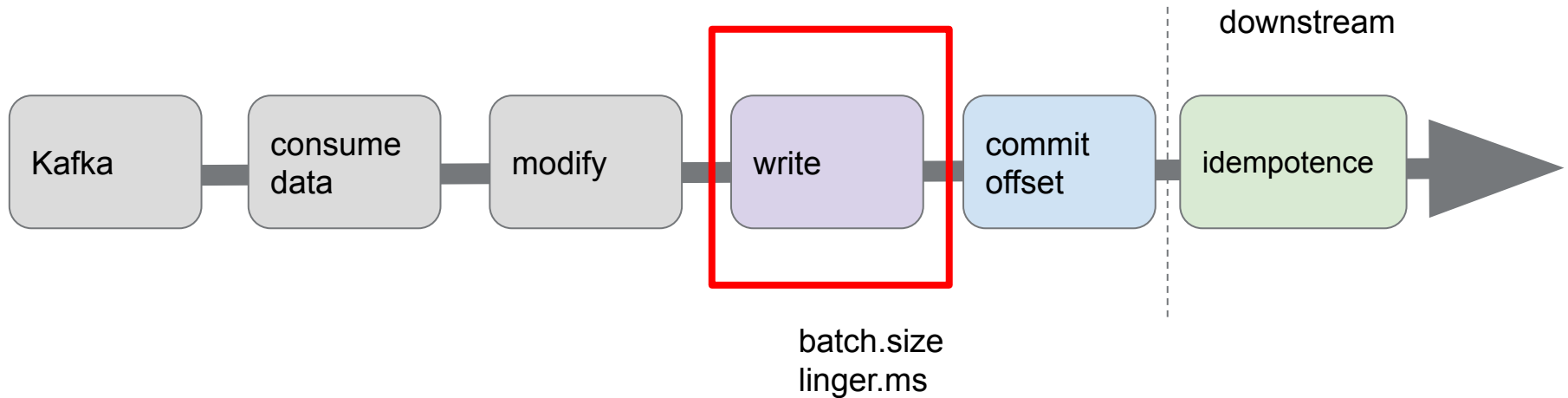
from the consumed cata. The structure in Kafka is key -> val pair
key: group ID + topic + partition -> val: offset value

Question: How to save in Redis?

key: group ID + topic

value: partition + offset, partition + offset,

Inside the Kafka - still linger in cache?



solution1: send one msg after one synchronously (slow)
solution2: **flush** before commit offset. Flush for each partition

Frontend UI - search by ticker symbol

The image shows a financial data page for TSLA. On the left, there is a summary section with the following data:

Previous Close:	
Open:	
Day High:	
Day Low:	
Previous Close:	
Volume:	107440900
Market Cap:	539005145936

In the center, a search dropdown menu is open, showing the search term "TSLA" and a list of recent searches:

- Form Built... on-demand video
- Access Music on mobile and TV
- Product Update - Q4 2021
- Master Digital Marketing Strategy course
- Dedicated forms for products
- Product Update - Q4 2021

Below the search dropdown, there are two sections: "RECENT PAGES" and a table of financial data.

RECENT PAGES:

- Messages - Conversation / ... / Mike Mills
- Messages - Conversation / ... / Eva Patrick

Table of financial data:

	314.6667
	101.81
	-0.3518
	182.448
	208.4498
	122445310
3-Month Average Volume:	122445310

Frontend UI - summary, analysis, active users, price history

Summary of TSLA

Previous Close:	161.9991
Open:	163.97
Day High:	170.79
Day Low:	163.51
Previous Close:	161.9991
Volume:	107440900
Market Cap:	539005145936

Analysis

Year High:	314.6667
Year Low:	101.81
Year Change:	-0.3518
50-Day Average:	182.448
200-Day Average:	208.4498
10-Day Average Volume:	122445310
3-Month Average Volume:	122445310

Active Users ?

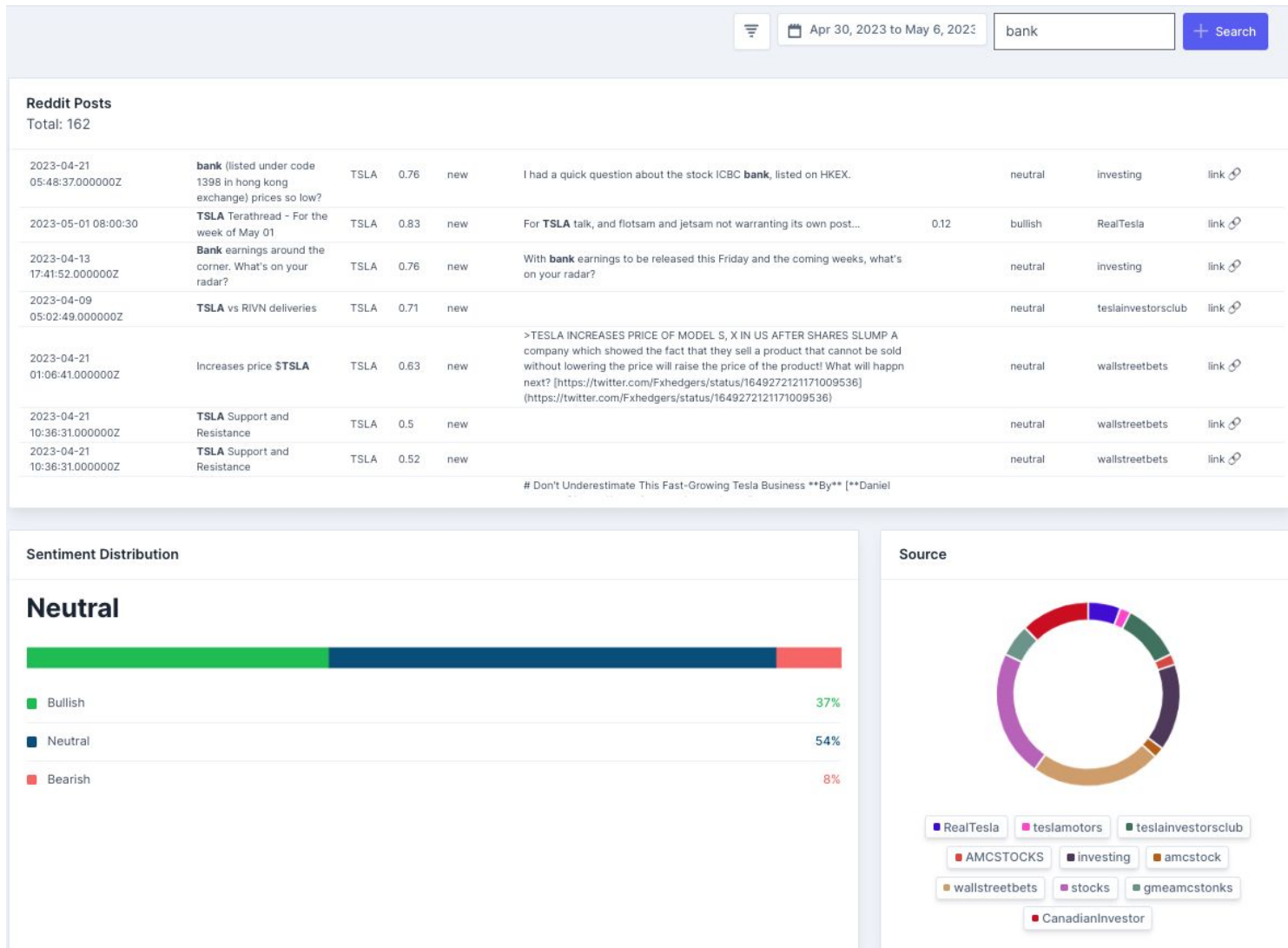
3005 +0.87%



Price History



Frontend UI - real time reddit posts, sentiment analysis



Frontend UI - recent news related to this stock

Recent News



Elon Musk Responds to a White House Snub

TheStreet.com

[STORY](#)

5/6/2023



Tesla Is Doing Exactly What You Wouldn't Think It Would Do

Motley Fool

[STORY](#)

5/6/2023



2 Reasons to Buy Tesla Stock Before it Skyrockets

Motley Fool

[STORY](#)

Thank you