

# 실시간 tableau 대시보드 구축 (진행중)

# INDEX

|      |               |
|------|---------------|
| ch 1 | 프로젝트 주제       |
| ch 2 | 아키텍처          |
| ch 3 | Tableau UI 구성 |
| ch 4 | 결론 및 느낀점      |

## Chapter 1

---

# 프로젝트 주제



# 내가 좋아하는 축구 데이터로 의미있는 시각화를 해보자

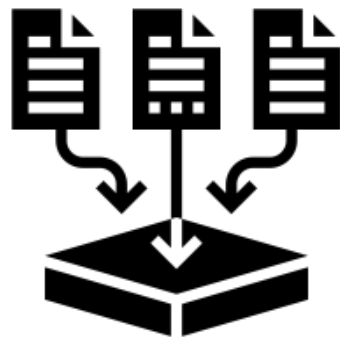
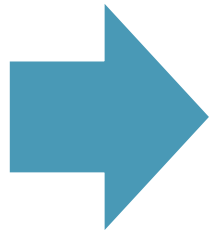
- 국내 축구 뉴스 기사를 보다보면, 기자가 가져오는 기사를 수동적으로 보는 느낌이 강함
- 축구 뉴스 기사 소스로부터 데이터를 직접 가져와서 이러한 불편함을 해소해보자
- 이때 시각화를 실시간으로 구현하여 자동으로 결과를 업데이트 해보자

# 프로젝트 주제

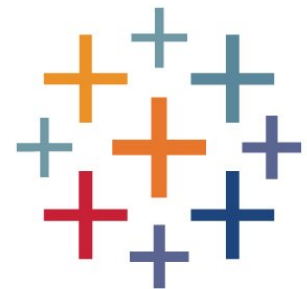
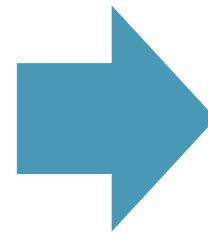
## 프로젝트 흐름



- 잉글랜드 프리미어 리그 (EPL)의 한 팀인 첼시와 관련된 뉴스 기사를 수집함



- 여러 데이터 출처를 활용함



- tableau 시각화

## 수집하는 데이터 종류

- 첼시와 관련된 뉴스 기사
- 첼시 감독의 인터뷰 내용
- 첼시 선수들의 경기 관련 통계치



## Chapter 2

---

# 아키텍처

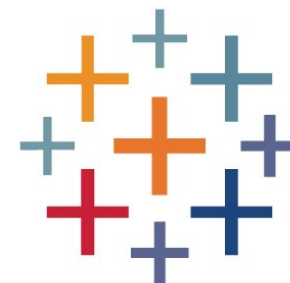
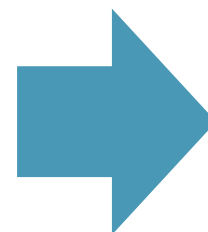
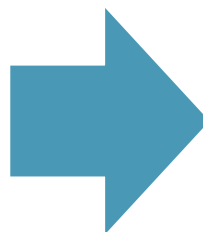


# 프로젝트 아키텍처

수집된 데이터의 소스를 토픽으로 구분하여 카프카에 저장



crontab을 통해 크롤링  
python script를 스케줄링  
(producer 역할)



kafka connect를 이용해서  
Rockset과 연결

실시간으로 카프카 토픽에 있는 메시지를  
JDBC에 저장함  
(consumer 역할)

Tableau을 통해서 실시간  
시각화





## crontab

- 스케줄링을 통해서 일정 간격으로 python script를 실행시킴
- 기사가 업데이트 되는 시점을 파악하여 스케줄링을 하면 경기 진행 후, 업데이트되는 기사를 크롤링할 수 있음
- apache airflow로도 스케줄링 작업을 할 수 있지만, 스케줄링 작업이 처음인만큼 가장 간단한 crontab으로 구현해봄



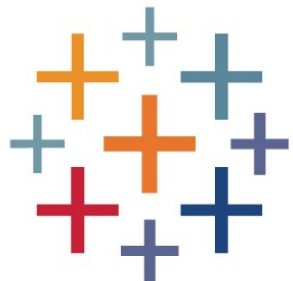
## Kafka & Kafka Connect

- Kafka를 로컬에 구축하여 크롤링된 내용을 처리해봄
- 사실 대용량 데이터가 아니고 초 단위로 데이터가 업데이트 되는 것은 아니기 때문에 Kafka를 사용하지 않아도 됨
- 그럼에도 불구하고 Kafka를 사용하여 구축 경험을 쌓고자 하였음



## Rockset

- Tableau의 데이터 소스인 JDBC를 실시간으로 구축하기 위해서 사용함
- Kafka와 Tableau을 연결해줌
- Kafka뿐만 아니라 mongo DB, AWS S3 등 데이터 소스로부터 자동으로 데이터를 가져와줌
- SQL 쿼리문을 작성하여 원하는 테이블 형태로 실시간 저장할 수 있음



## Tableau

- 자유도가 매우 높은 BI 툴
- 매우 다양한 데이터 소스를 사용할 수 있어서 편리함
- 데이터 소스를 자동으로 업데이트하게 설정하여 자동으로 업데이트되는 시각화를 만들고자 하였음

## Chapter 3

---

# Tableau UI 구성



# Tableau UI 구성

- *WHERE WE ARE*

현재 프리미어리그의 팀 순위를 보여줌. 특히, 관심있는 팀인 첼시의 순위를 노란색으로 강조하였음.

- *SEE DETAIL*

이 버튼을 클릭하면 주별 팀 순위의 변동을 시각화로 이동할 수 있도록 구축할 예정

- *LATEST NEWS WORD CLOUD*

첼시와 관련된 기사들 중, 최근 3일간의 기사 내용에 대한 wordcloud 그림

- *NEWS TITLE & HASH TAG*

아직 완성하지 못함. 뉴스 기사의 내용을 알고리즘을 통해서 요약하여, HASH TAG로 시각화할 계획임

## Chelsea Football Club Newspeed



Click other tab you are interested in !

Latest News

Coach Interview

Recent Game Result

Player History

### WHERE WE ARE

SEE DETAIL

|  | Play | Points | Won | Drawn | Lost | For | Against | GD  |
|--|------|--------|-----|-------|------|-----|---------|-----|
|  | 25   | 59     | 18  | 5     | 2    | 50  | 15      | 35  |
|  | 25   | 49     | 14  | 7     | 4    | 53  | 32      | 21  |
|  | 25   | 49     | 15  | 4     | 6    | 44  | 27      | 17  |
|  | 25   | 45     | 13  | 6     | 6    | 39  | 29      | 10  |
|  | 25   | 43     | 12  | 7     | 6    | 41  | 25      | 16  |
|  | 25   | 40     | 11  | 7     | 7    | 45  | 34      | 11  |
|  | 24   | 40     | 12  | 4     | 8    | 37  | 33      | 4   |
|  | 23   | 36     | 11  | 3     | 9    | 37  | 26      | 11  |
|  | 24   | 36     | 10  | 6     | 8    | 37  | 27      | 10  |
|  | 25   | 35     | 11  | 2     | 12   | 43  | 43      | 0   |
|  | 25   | 34     | 10  | 4     | 11   | 31  | 26      | 5   |
|  | 25   | 33     | 9   | 6     | 10   | 26  | 32      | -6  |
|  | 25   | 32     | 9   | 5     | 11   | 29  | 43      | -14 |
|  | 25   | 30     | 8   | 6     | 11   | 31  | 43      | -12 |
|  | 25   | 28     | 7   | 7     | 11   | 18  | 30      | -12 |
|  | 25   | 26     | 5   | 11    | 9    | 26  | 32      | -6  |
|  | 25   | 25     | 7   | 4     | 14   | 26  | 43      | -17 |
|  | 25   | 22     | 4   | 10    | 11   | 21  | 32      | -11 |
|  | 25   | 14     | 2   | 8     | 15   | 19  | 55      | -36 |
|  | 25   | 11     | 3   | 2     | 20   | 15  | 41      | -26 |

### LATEST NEWS WORD CLOUD

see clubgood  
odoi thing match goal get mountleg year time  
firstsecondplayer come teamgive league  
gamehudson chelsea tuchel atletico  
say champion season take make play newsletter  
madrid onethink follow know side big backalso  
last bluewell win like

### NEWS TITLE & HASH TAG

# Tableau UI 구성

- Coach Interview

이 버튼을 클릭하면 가장 최근 경기에 대한 첼시 감독의 인터뷰를 요약한 시각화를 볼 수 있음 (계획).

- Recent Game Recent

이 버튼을 클릭하면 가장 최근 경기의 결과를 요약한 시각화를 볼 수 있음 (계획).

- Player History

이 버튼을 클릭하면 첼시 선수들의 경기 통계치에 대한 시각화를 볼 수 있음 (계획).

- 4:34

이 버튼은 4분 34초 후에 데이터 소스가 자동으로 업데이트 된다는 뜻임. 이에 따라 시각화 내용도 업데이트 됨

## Chelsea Football Club Newspeed



Click other tab you are interested in !

Latest News

Coach Interview

Recent Game Result

Player History

### WHERE WE ARE

SEE DETAIL

|  | Play | Points | Won | Drawn | Lost | For | Against | GD  |
|--|------|--------|-----|-------|------|-----|---------|-----|
|  | 25   | 59     | 18  | 5     | 2    | 50  | 15      | 35  |
|  | 25   | 49     | 14  | 7     | 4    | 53  | 32      | 21  |
|  | 25   | 49     | 15  | 4     | 6    | 44  | 27      | 17  |
|  | 25   | 45     | 13  | 6     | 6    | 39  | 29      | 10  |
|  | 25   | 43     | 12  | 7     | 6    | 41  | 25      | 16  |
|  | 25   | 40     | 11  | 7     | 7    | 45  | 34      | 11  |
|  | 24   | 40     | 12  | 4     | 8    | 37  | 33      | 4   |
|  | 23   | 36     | 11  | 3     | 9    | 37  | 26      | 11  |
|  | 24   | 36     | 10  | 6     | 8    | 37  | 27      | 10  |
|  | 25   | 35     | 11  | 2     | 12   | 43  | 43      | 0   |
|  | 25   | 34     | 10  | 4     | 11   | 31  | 26      | 5   |
|  | 25   | 33     | 9   | 6     | 10   | 26  | 32      | -6  |
|  | 25   | 32     | 9   | 5     | 11   | 29  | 43      | -14 |
|  | 25   | 30     | 8   | 6     | 11   | 31  | 43      | -12 |
|  | 25   | 28     | 7   | 7     | 11   | 18  | 30      | -12 |
|  | 25   | 26     | 5   | 11    | 9    | 26  | 32      | -6  |
|  | 25   | 25     | 7   | 4     | 14   | 26  | 43      | -17 |
|  | 25   | 22     | 4   | 10    | 11   | 21  | 32      | -11 |
|  | 25   | 14     | 2   | 8     | 15   | 19  | 55      | -36 |
|  | 25   | 11     | 3   | 2     | 20   | 15  | 41      | -26 |

## LATEST NEWS WORD CLOUD

see clubgood  
odoi thing match goal get mountleg year time  
90firstsecondplayer come teamgive league  
gamehudson chelsea tuchel atletico  
say champion season take make play newsletter  
madrid onethink follow know side big back also  
last bluewell win like

## NEWS TITLE & HASH TAG

## Chapter 4

---

# 결론 및 느낀점





# 결론 및 느낀점

- 실시간으로 데이터를 수집하는 파이프라인을 구축해봄
- 시각화는 완성하지는 못했고 50% 정도까지만 진행함
- 이번 프로젝트를 통해서 데이터를 동적으로 바라보는 새로운 관점을 가질 수 있었음
- 또한 Tableau을 사용하여 동적인 데이터를 실시간으로 시각화하는 과정까지 경험하여 추후 프로젝트에서 유용하게 관련 지식을 사용할 수 있을 것이라 생각됨.

감사합니다