



SSAFIT NEWS

뉴스 데이터 수집 및 시각화 프로젝트



Contents

1.Introduction

- 프로젝트 목적 및 개요
- 핵심 기능 요약

2. Data Pipeline

- 실시간 데이터 처리 흐름
- 배치 기반 데이터 처리

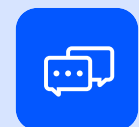
3. Frontend / Backend

- 백엔드 API 구성
- 프론트엔드 DEMO



Introduction

프로젝트 목적 및 개요



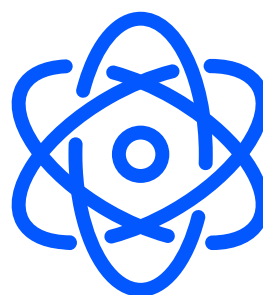
프로젝트 목적 및 개요



- 오늘날 정보의 양은 폭발적으로 증가하고 있음
- 그러나 사용자는 '중요한 뉴스'를 빠르게 탐색하기 어려움



- 실시간 수집된 뉴스를 AI로 분석하고, 유사 뉴스 추천 및 키워드 검색 기능 시스템을 구축

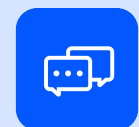


- 데이터 스트리밍 처리
- 벡터 임베딩
- 검색 최적화



Introduction

핵심 기능 요약



핵심 기능 요약



- 뉴스 실시간 수집
- 시간 단위 뉴스 수집 및 처리



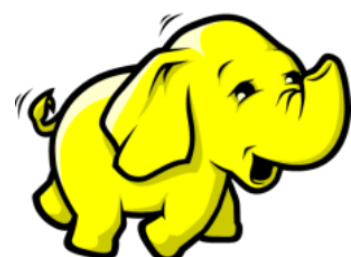
- 데이터 배치 처리
- 하루 단위 뉴스 분석 및 리포트 생성



- 워크플로우 스케줄링 자동화
- RSS 수집, Spark 분석, DB & ES 저장 자동화



- 실시간 데이터 처리 엔진
- 도커 실행만으로 Consumer 자동 구동, 즉시 전처리 수행



- 대용량 데이터 저장 및 리포트 보관
- Spark로 분석된 배치 결과를 HDFS에 저장



- 전문 검색 기능
- 키워드/날짜/카테고리 검색



Introduction

핵심 기능 요약



핵심 기능 요약



- 시스템 모니터링
- Flink Task 실시간 상태 확인



- Docker Compose 기반 컨테이너 구성
- Kafka, Spark, Flink, Airflow, PostgreSQL 등 통합 관리



- 사용자 인터페이스
- Vue 기반의 반응형 뉴스 탐색 UI



- Django REST Framework로 API 설계
- 뉴스 목록, 상세, 검색, 댓글, 좋아요 등 전 기능 제공

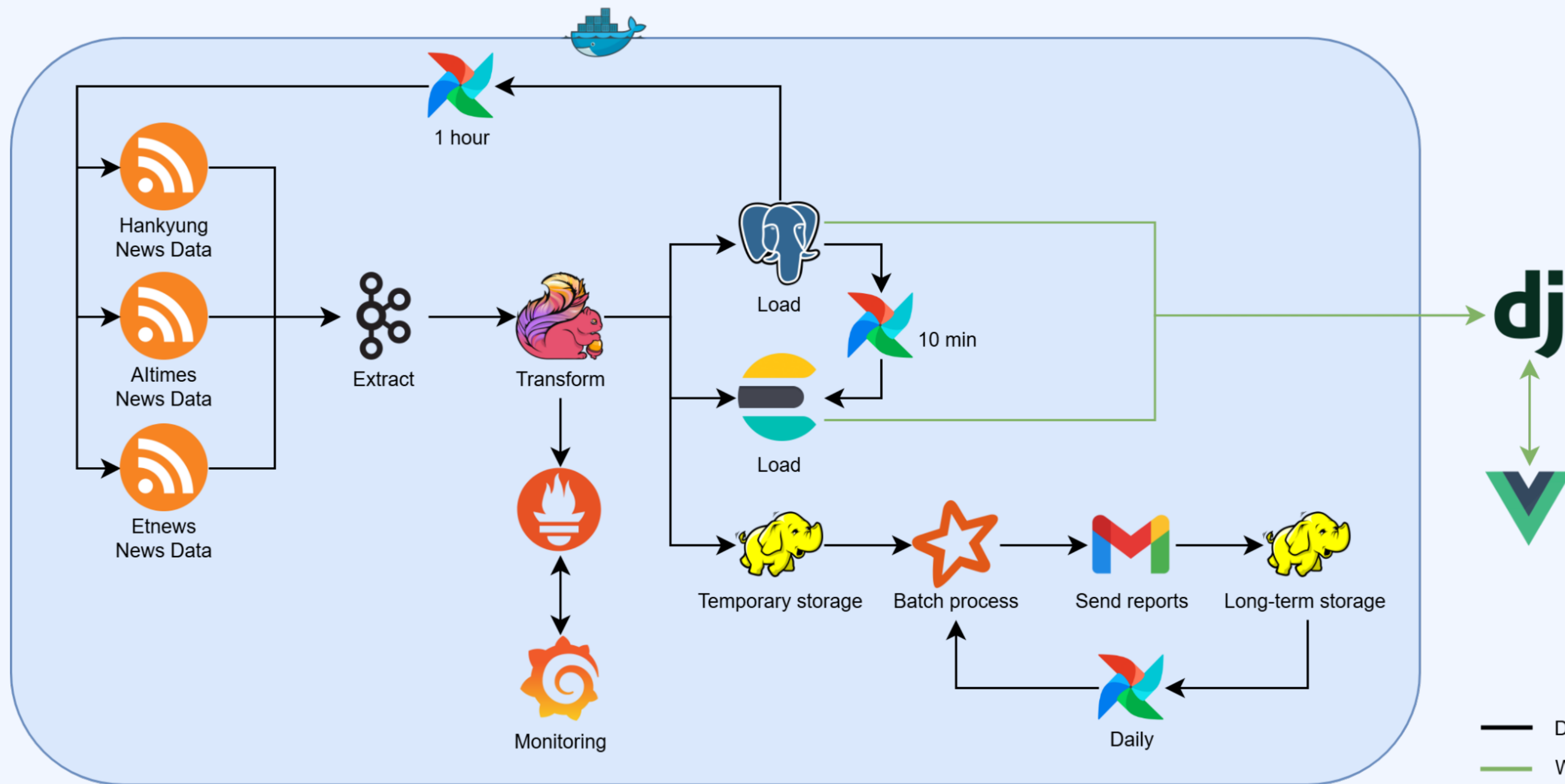


- OpenAI API 기반 분석
- 키워드 추출, 카테고리 분류, 벡터 임베딩 생성



Data Pipeline

전체 WORKFLOW

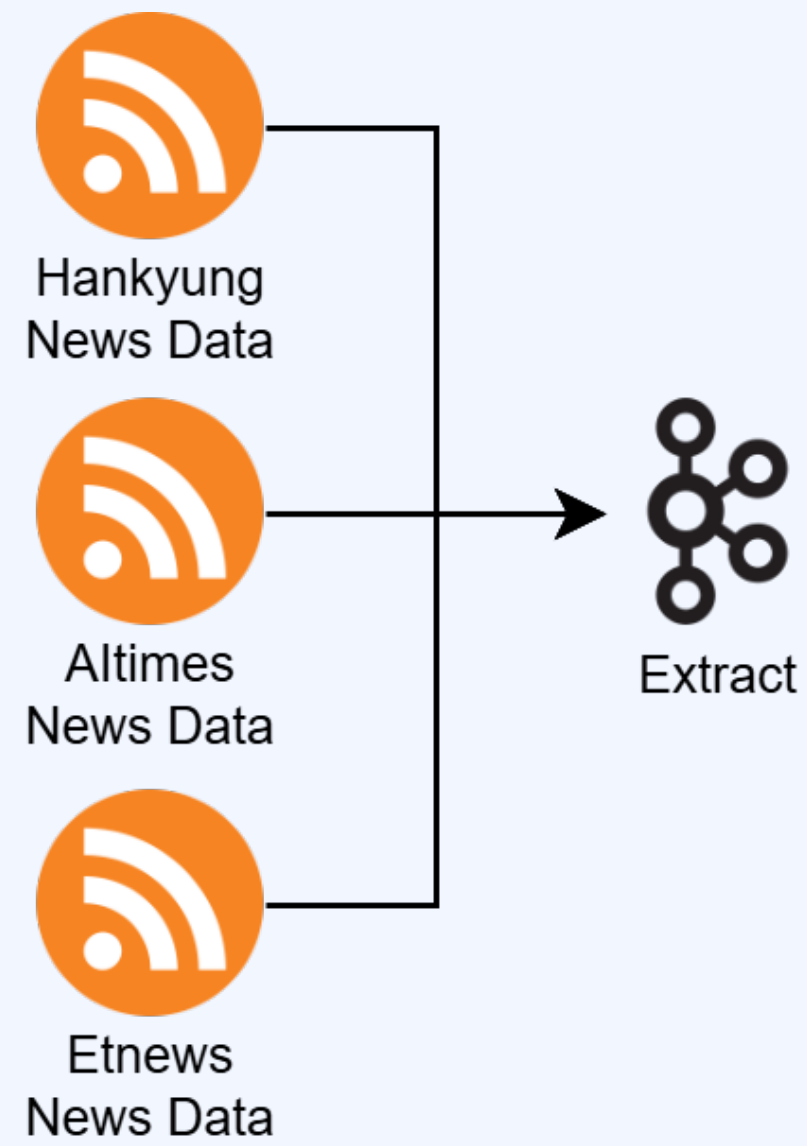




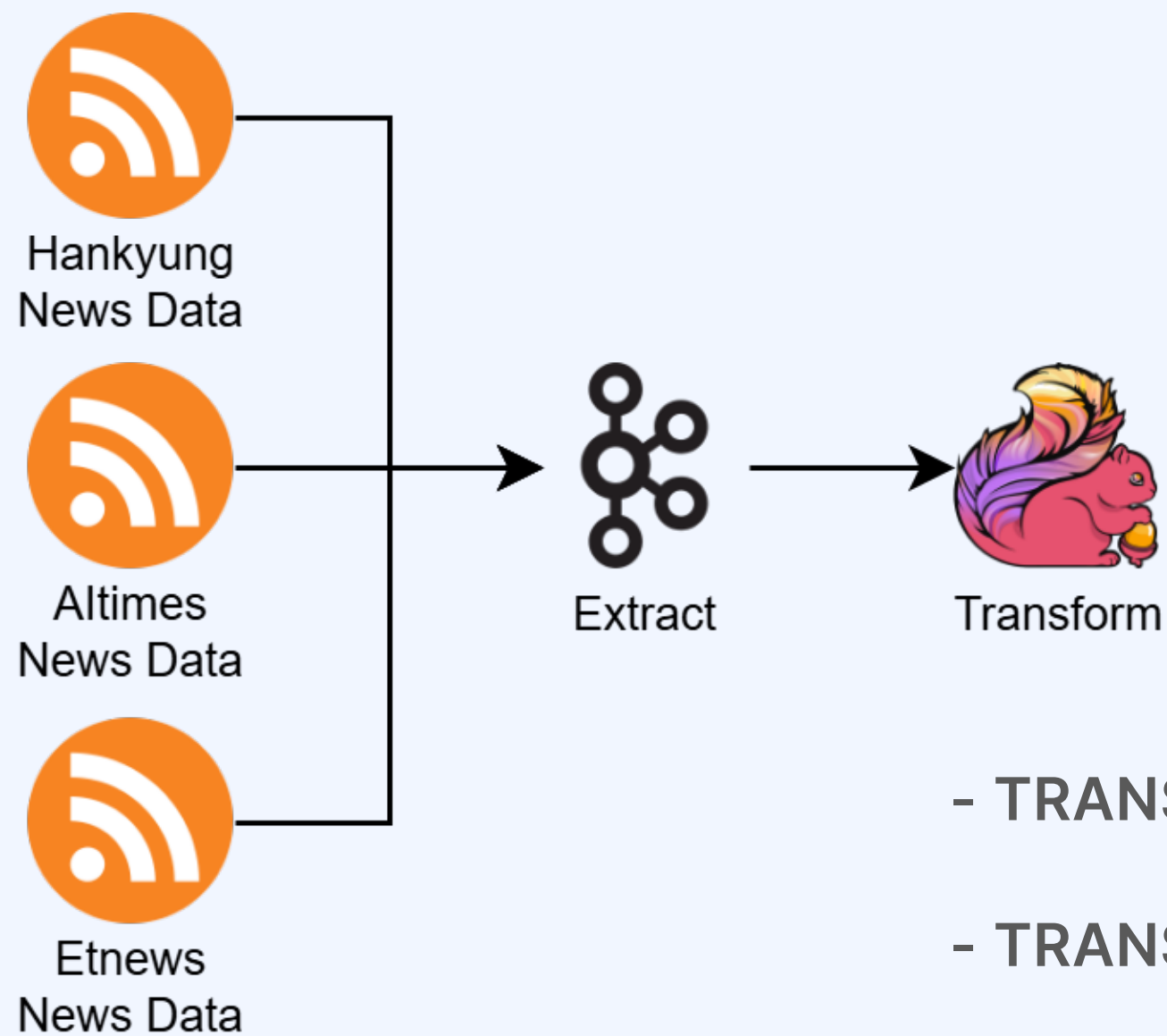
Data Pipeline

실시간 데이터 처리 흐름

STEP 01. KAFKA PRODUCER가 RSS 뉴스 기사 수집 → KAFKA TOPIC 발행

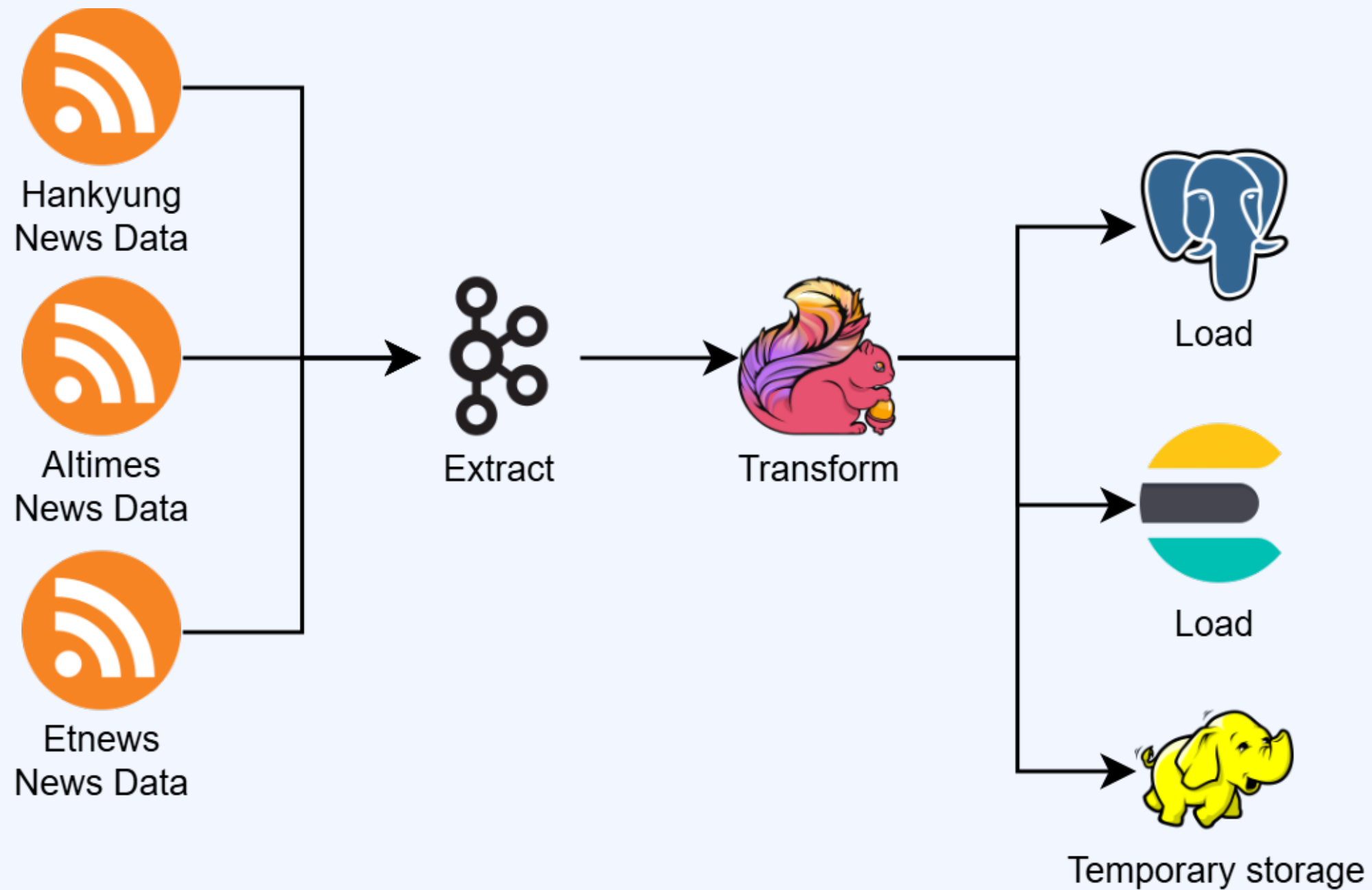


STEP 02. FLINK CONSUMER가 메시지 수신 후 전처리 작업

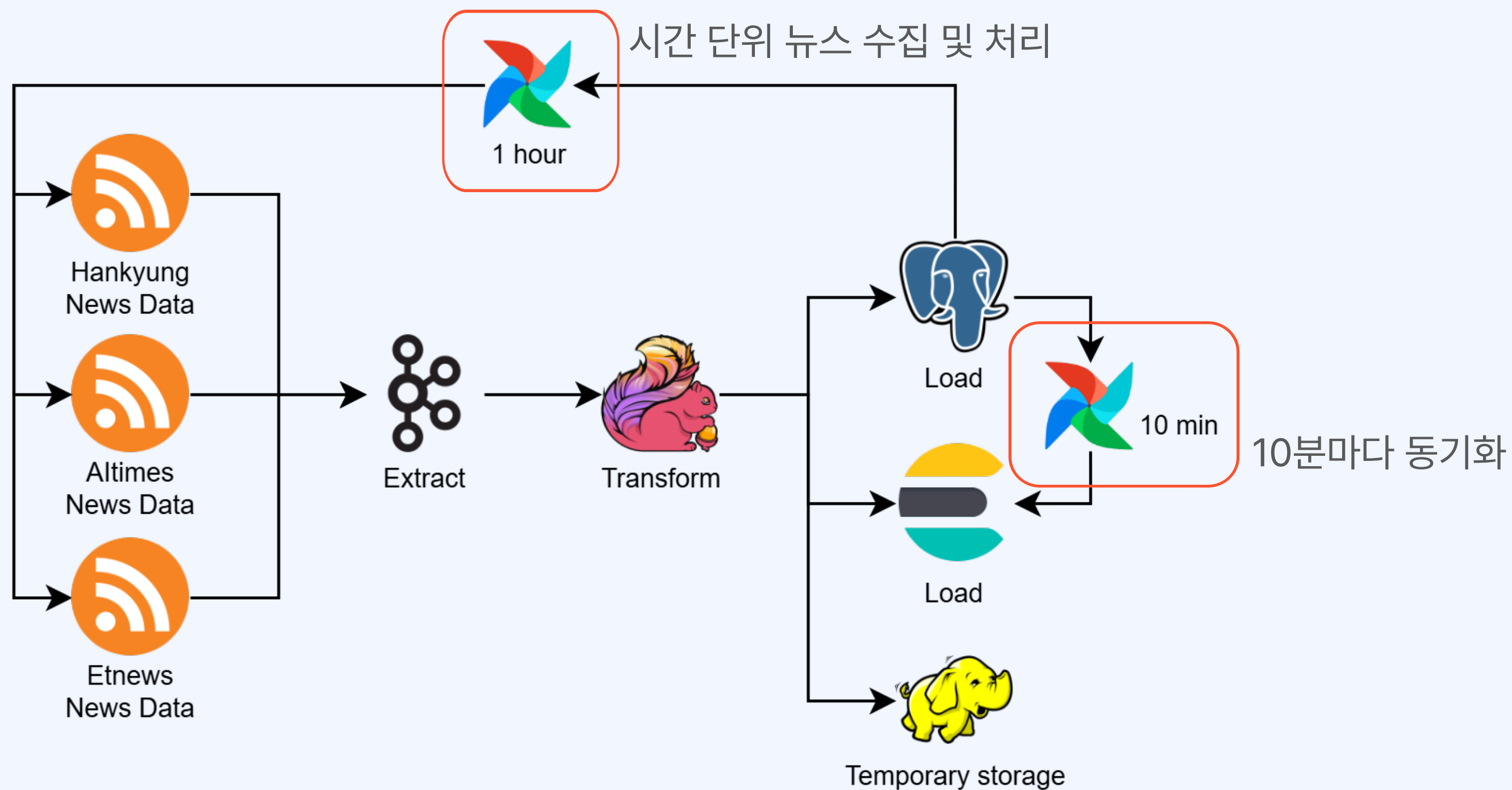


- TRANSFORM_EXTRACT_KEYWORDS() : 핵심 키워드 추출
- TRANSFORM_CLASSIFY_CATEGORY() : 기사 분류
- TRANSFORM_TO_EMBEDDING() : 문장 임베딩 생성

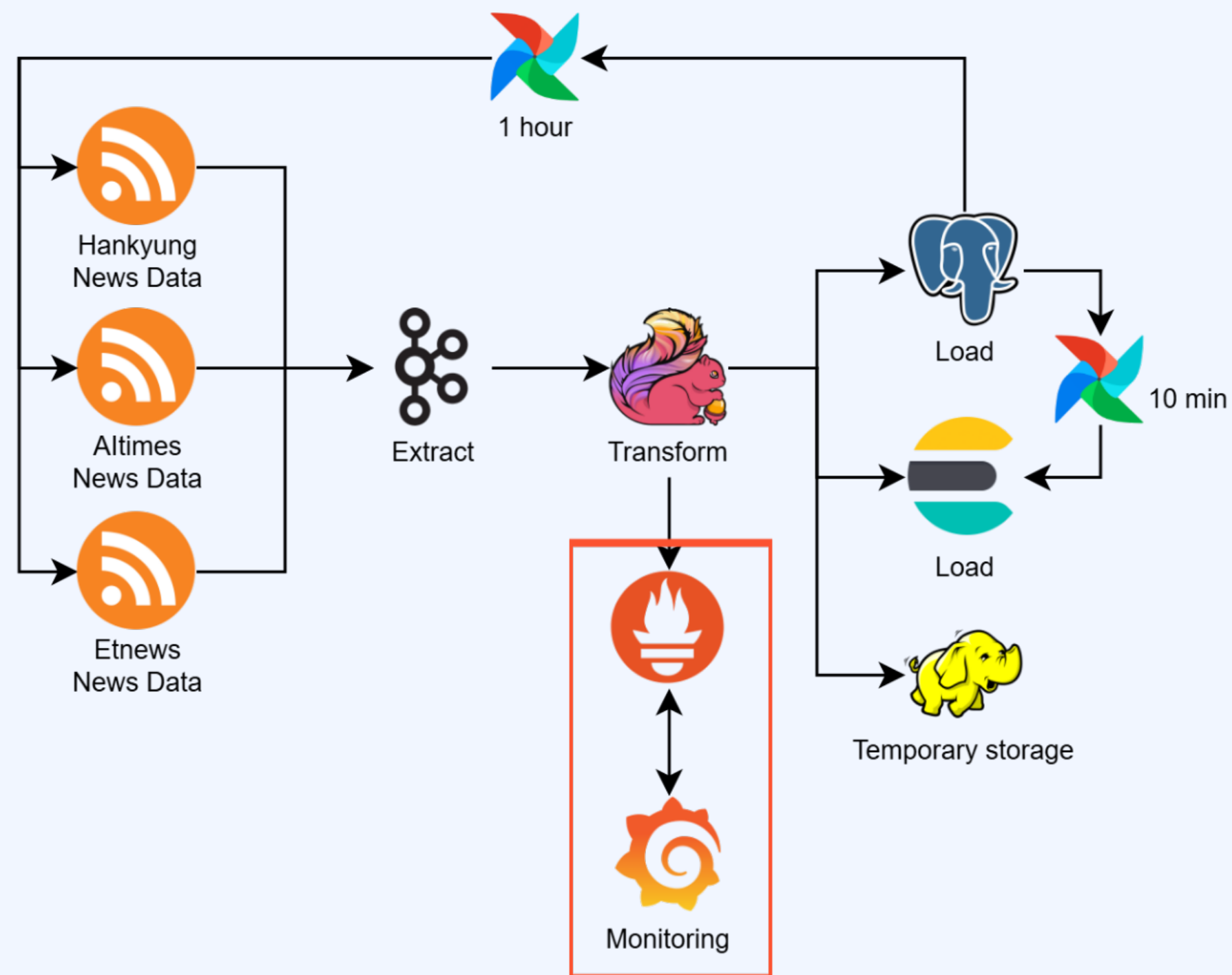
STEP 03. POSTGRESQL + ELASTICSEARCH + HADOOP 저장



STEP04. AIRFLOW



STEP05. MONITORING

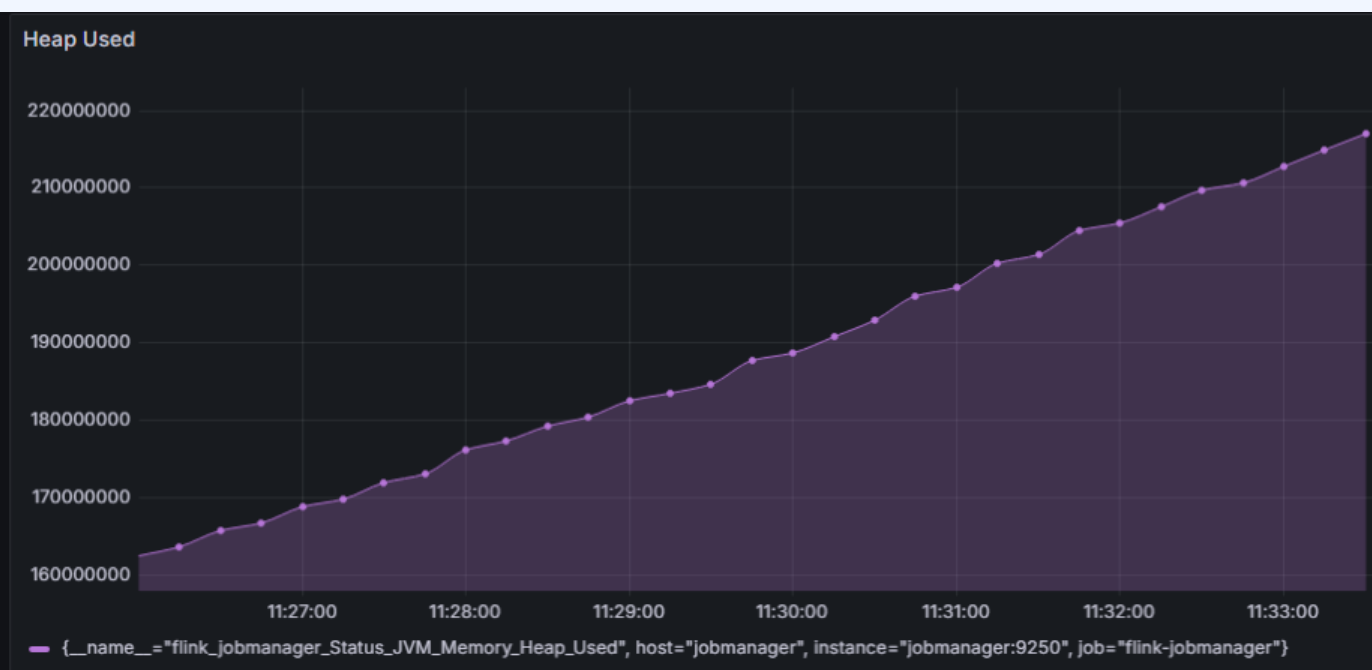
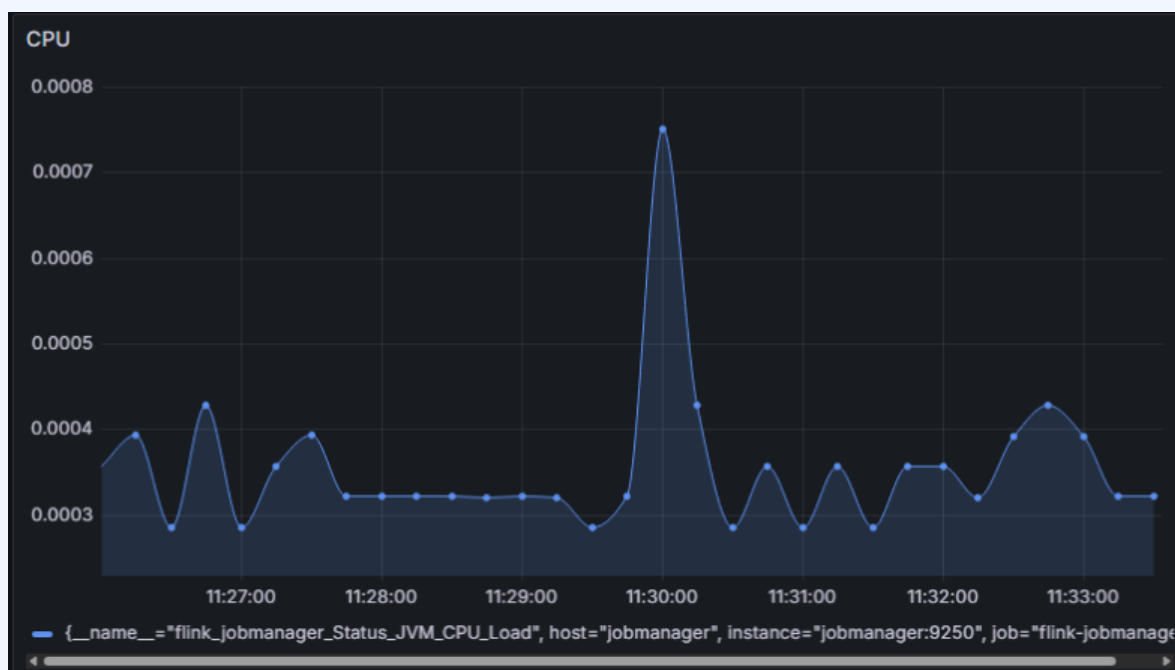




Data Pipeline

실시간 데이터 처리 흐름

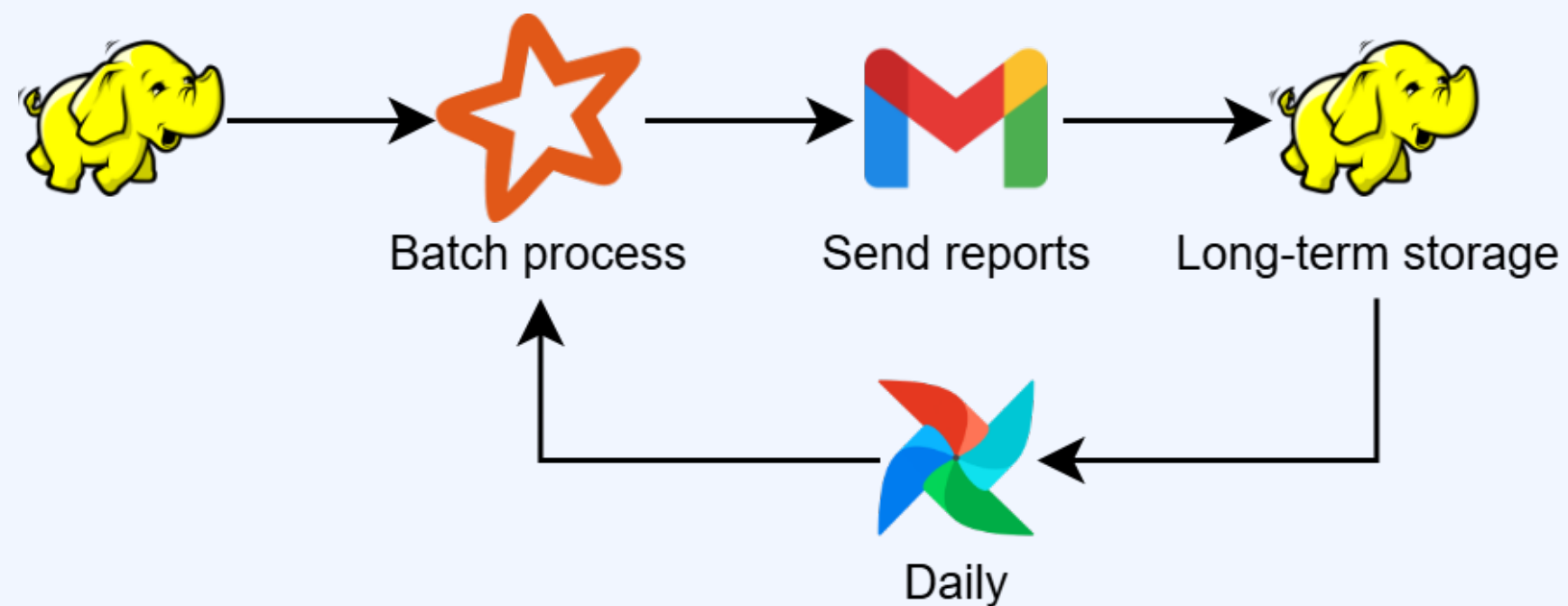
STEP05. MONITORING



Data Pipeline

배치 기반 데이터 처리

매일 새벽 1시에 DAG 자동 실행



분석 항목:

- 키워드 상위 N개 추출
- 카테고리 상위 N개 추출
- 시간대별 기사 분포 시각화

결과 출력:

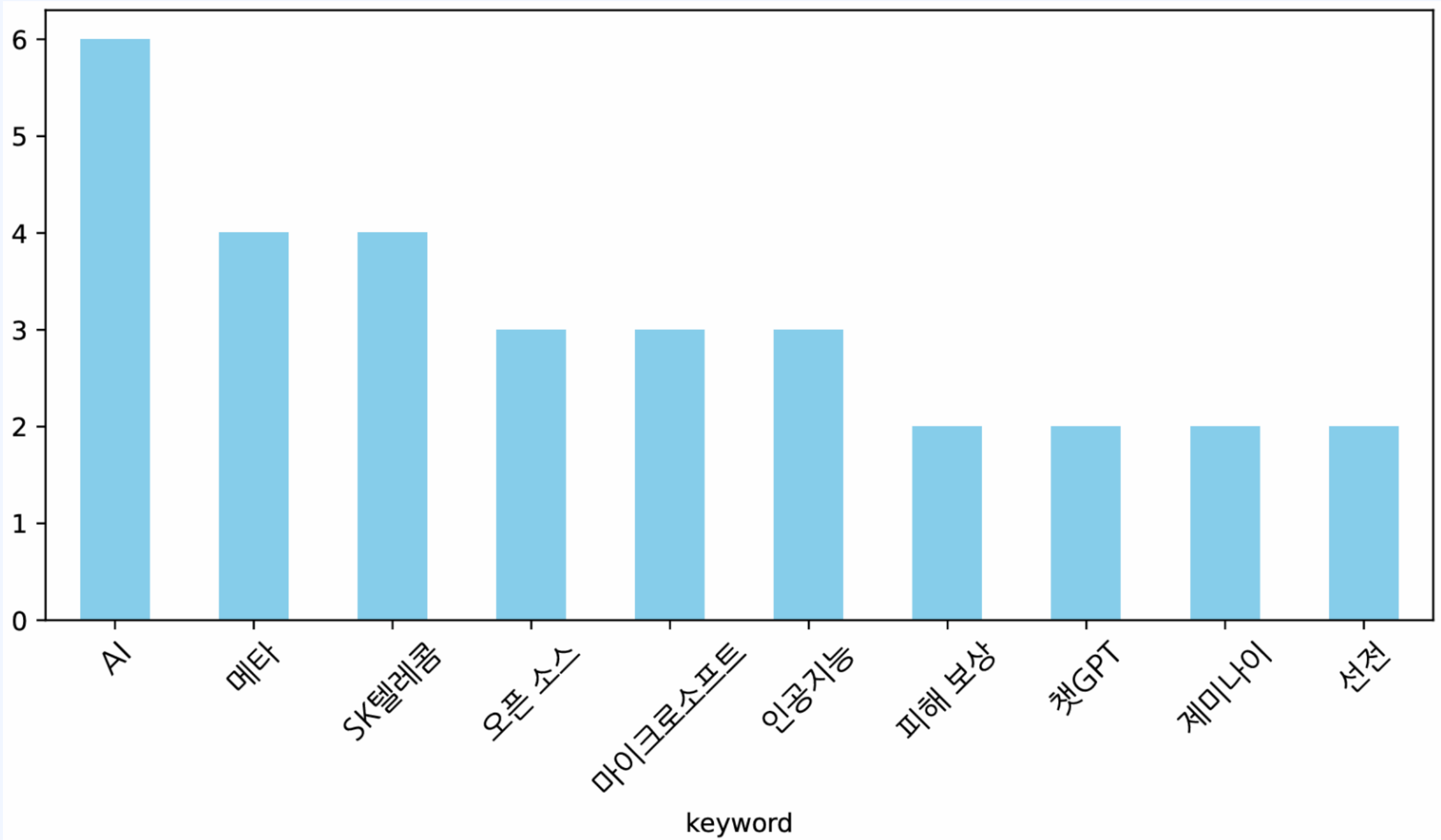
- /NEWS_ARCHIVE/에 JSON 백업
- /REPORTS/에 PDF 리포트 저장 (HDFS 저장 포함)



Data Pipeline

배치 기반 데이터 처리

키워드 상위 N개 추출





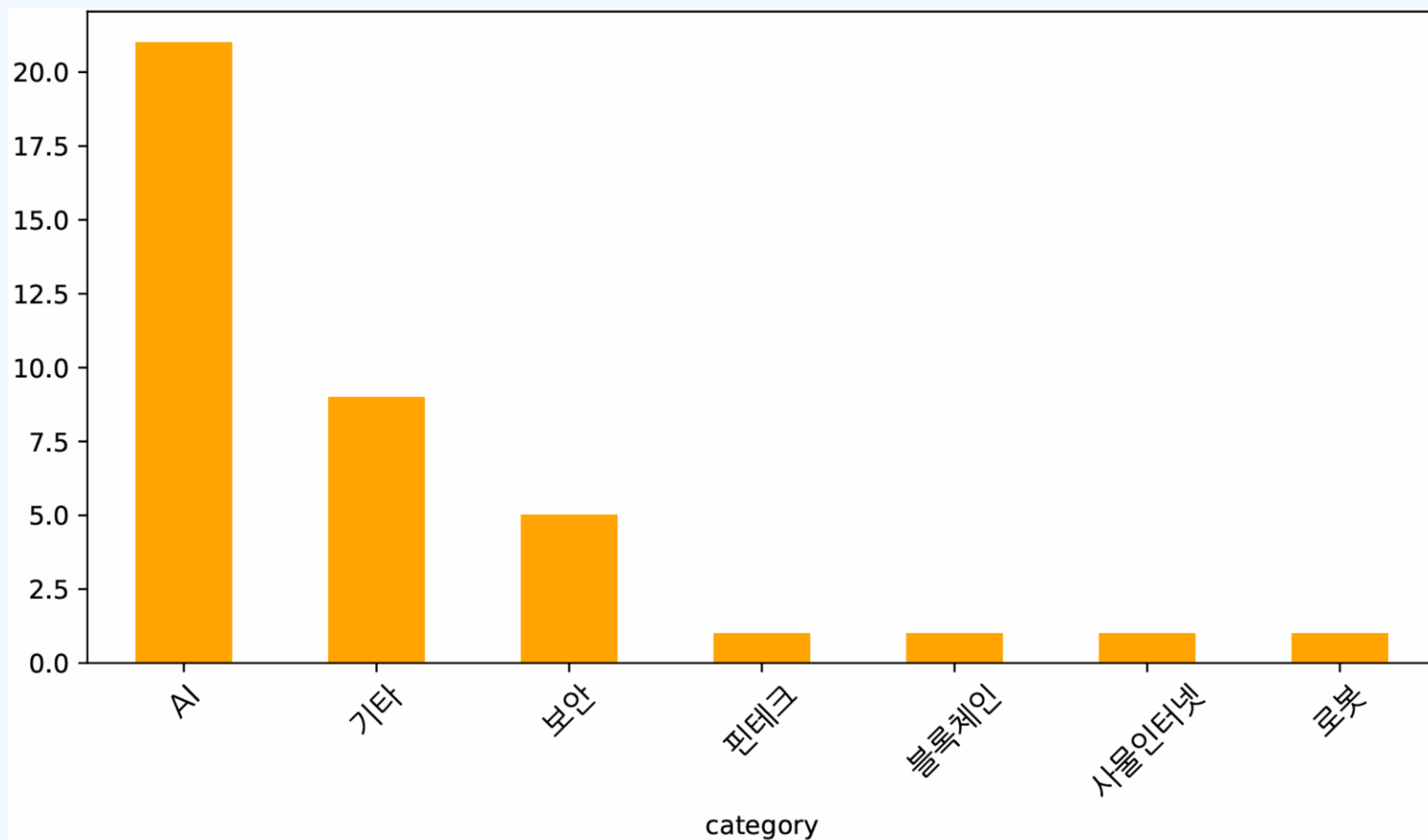
Data Pipeline



SSAFIT NEWS

배치 기반 데이터 처리

카테고리 상위 N개 추출





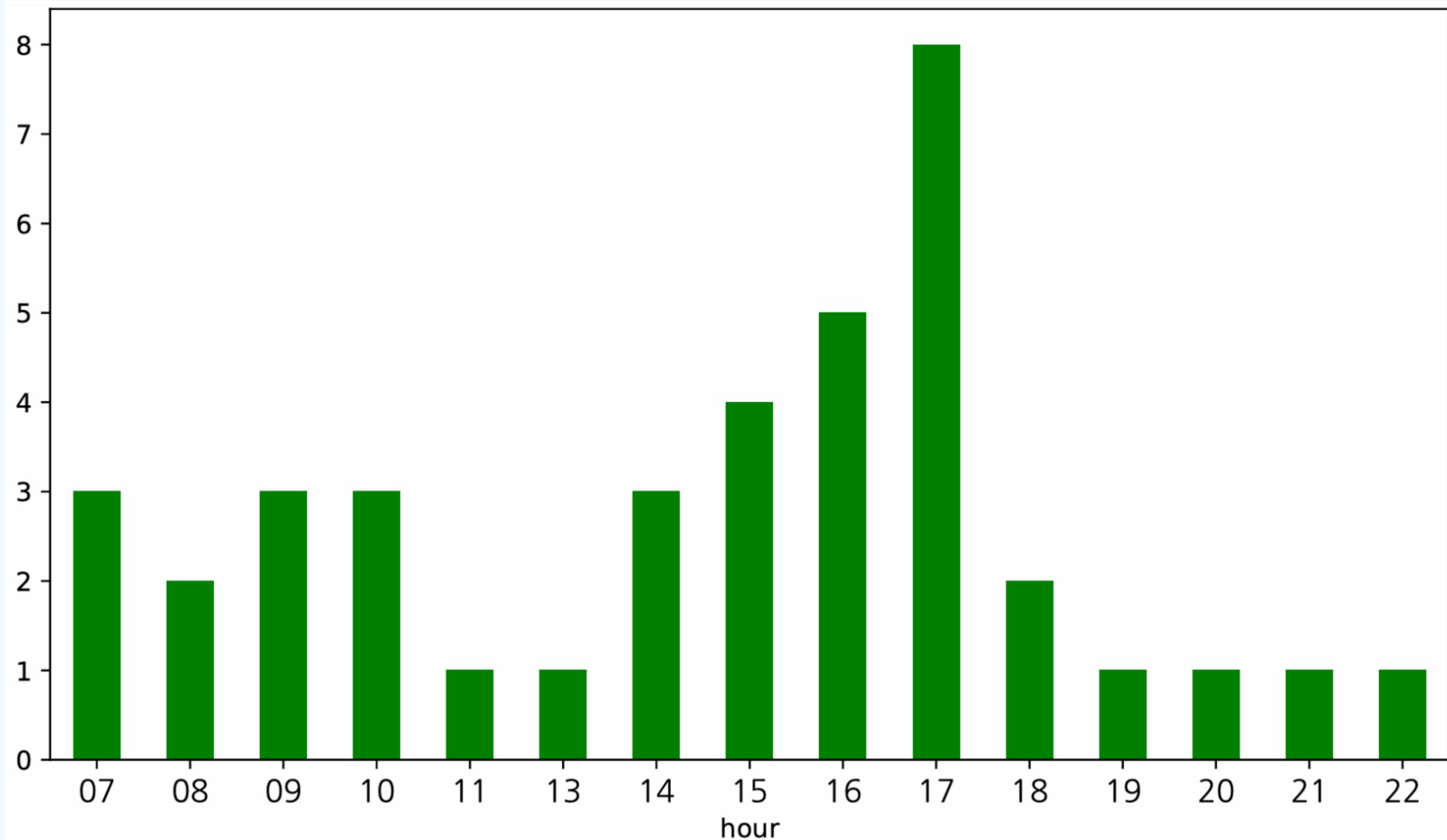
Data Pipeline

배치 기반 데이터 처리

시간대별 기사 분포 시각화



SSAFIT NEWS





Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO



회원가입

POST /auth/registration/

http://127.0.0.1:8000/auth/registration/

회원가입 기능

HEADERS

Content-Type application/json

Body raw (json)

```
json
{
  "username": "test_name1",
  "email": "test1@gmail.com",
  "password1": "Ssafy10000",
  "password2": "Ssafy10000"
}
```



SSAFIT NEWS

SSAFIT NEWS

맞춤형 AI 뉴스 큐레이션

회원가입

SSAFY 뉴스 서비스에 오신 것을 환영합니다

이름

이메일

비밀번호

비밀번호 확인

가입하기 →

이미 계정이 있으신가요? [로그인하기](#)

[이용약관](#) [개인정보처리방침](#) [로그인](#)

© 2024 SSAFY NEWS. All rights reserved.



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO



로그인

POST /auth/login/

http://127.0.0.1:8000/auth/login/

로그인 기능 (Access token 발급용)

HEADERS

Content-Type application/json

Body raw (json)

json

```
{
  "email": "admin@admin.com",
  "password": "admin"
}
```

SSAFIT NEWS

맞춤형 AI 뉴스 큐레이션

로그인

SSAFY 뉴스 서비스에 오신 것을 환영합니다

이메일

비밀번호

로그인하기 →

아직 계정이 없으신가요? [가입하기](#)

[이용약관](#) [개인정보처리방침](#) [회원가입](#)

© 2024 SSAFY NEWS. All rights reserved.



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO

뉴스 리스트

GET /news-list/

http://127.0.0.1:8000/news-list/

뉴스 목록 조회 기능

(로그인 없이 Header에 access token 보내면 맞춤형 추천 X)

PARAMS

sort_by latest

SSAFIT NEWS

검색어를 입력하세요



SSAFIT NEWS

IT 트렌드를 한눈에, 기술 뉴스 큐레이션 서비스

전체

AI

빅데이터

블록체인

사물인터넷

클라우드

로봇

자율주행

반도체

스타트업

핀테크

보안

메타버스

양자컴퓨팅

AR/VR

기타

최신순

추천순

AI

박수빈 기자 · 2025.05.22.

클라우드웍스, '아시아테크 싱가포르'서 해외 비즈니스 전개

(사진=클라우드웍스) 인공지능(AI) 테크 기업 클라우드웍스(대표 김우승)는 '아시아테크 싱가포르 2025(ATxSG 2025)'에 참가해 에이전틱 AI 기반의 핵심 기술과 솔루션을 선보인다고 22일 밝혔다. 27일~29일(현지시간) 열리는 ATxSG는 지난해 110개국 이상에서 2만2000명 이상의 업계 관계자들이 참가했다. 올해는 정부 및 기관 고위급 인사들이 참여해 AI 발전 방향을 모색하는 'ATx서밋'과 비즈니스 중심 'ATx엔터프라이즈' 등으로 구성됐다. 클라우드웍스는 이번 전시회에서 에이전...

1 97

#클라우드웍스

#에이전틱 AI

#알피

#ATxSG 2025

#동남아 시장

보안

박수빈 · 2025.05.22.

SKT "고유식별번호 유출 안됐다" 해명에도...'유심 복제' 가능성 우려

인천 중구 인천국제공항 1터미널 SK텔레콤 로밍센터 앞에서 한 여행객이 유심을 교체하고 있다. 사진=임형택 한국경제신문 기자 SK텔레콤이 "고유식별번호(IMEI) 정보가 유출됐다 해도 불법 단말 복제는 사실상 불가능하다"면서 진화에 나섰다. 하지만 일각에선 불법 단말 복제보다 '유심 복제' 가능성이 커진 게 문제라는 우려가 나온다. IMEI 정보가 유출되면 유심보호서비스만으로는 유심 불법 복제를 막을 수 없어 세밀한 대응책을 만들어야 한다는 지적이다. 22일 회사 측에 따르면 ...

0 26

#IMEI

#유심 복제

#SK텔레콤

#유출

#인증키



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO



뉴스 상세 페이지



GET /news/{article_id}/

http://127.0.0.1:8000/news/720/?sort_by=latest

뉴스 목록 조회 기능

(로그인 없이 Header에 access token 보내면 맞춤형 추천 X)



기타

'원료의약품 사업 순항' 유한양행, 길리어드에 800억원대 공급계약

이지현 2025.05.22.

유한양행 본사. 유한양행 제공

유한양행

이 미국 길리어드사이언스와 800억원대 인간면역결핍바이러스(HIV) 원료의약품 공급 계약을 맺었다.

올해부터 수출이 잇따를 것으로 예상돼 실적 개선에 도움이 될 것이란 평가다.

유한양행은 길리어드와 888억1494만원(6377만달러) 규모의 HIV 치료제 원료의약품 공급 계약을 맺었다고 22일 공시했다. 계약 기간은 내년 12월31일까지다. 길리어드는 유한양행이 대금을 청구한 날로부터 한달 안에 청구액을 지급할 예정이다.

유한양행은 자회사 유한화학을 통해 길리어드와 베링거인겔하임 등의 원료를 공급하고 있다. 앞서 지난해 9월 유한양행은 길리어드와 8089만달러(당시 환율 기준 1076억6965만원) 규모의 HIV 치료제 원료 의약품 공급 계약을 맺었다고 공시했다.

업계에선 유한양행이 원료를 공급하는 길리어드사이언스의 HIV 치료제를 '선렌카'로 추정하고 있다. 선렌카는 '항바이러스제 명가'로 꼽히는 길리어드에서 2022년 미국에 출시한 HIV 치료용 주사제다. 길리어드는 이 약을 연 1회 투여하는 예방용으로 개발하고 있다. 특별한 문제가 없으면 올해 6월께 미국 식품의약국(FDA) 승인을 받을 수 있을 것으로 예상된다. 올해부터 유한화학의 원료의약품 공급 매출도 급격히 늘 것이란 전망이 나오는 이유다.

선렌카는 글락소스미스클라인(GSK)의 보카브리아+레캄비스와 함께 HIV 치료제 시장 판도를 장기지속형 주사제 시장으로 전환하고 있다. 장기 지속형 주사제는 HIV 환자가 매일 약을 먹는 불편을 줄여줘 치료 패러다임을 바꿀 것이란 평가를 받고 있다.

이지현 기자 bluesky@hankyung.com

#유한양행

#길리어드사이언스

#HIV

#원료의약품

#선렌카

0

12



좋아요

관련 기사

기타

유한양행 '렉라자', 21일 日 출시...후생성, 1250억 판매 전망

2025.05.20.

0

0

기타

난치병에 집중된 재생의료..."항노화로 넓혀야"

2025.04.17.

0

0

기타

동아에스티 "비만 치료 신약, 4주만에 체중 4.3% 감량"

2025.04.17.

0

0

기타

美 식도염약 시장 '한일전' 열린다...HK 이노엔, 최종임상 성공

2025.04.25.

0

0

기타

유바이오로직스, 질병청 mRNA 백신 개발 정부과제 선정

2025.04.19.

0

0



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO

뉴스 검색 / 자동완성

GET /news/search/

http://127.0.0.1:8000/news/search/?q=AI

PARAMS

q AI

GET /news/suggestions/

http://127.0.0.1:8000/news/suggestions/?q=AI

PARAMS

q AI

SSAFIT NEWS

AI

AI

AI 워커

AI 인프라

AI 체험존

AI 혜택



SSAFIT NEWS

IT 트렌드를 한눈에, 기술 뉴스 큐레이션 서비스

전체

AI

빅데이터

블록체인

사물인터넷

클라우드

로봇

자율주행

반도체

스타트업

핀테크

보안

메타버스

양자컴퓨팅

AR/VR

기타

최신순

추천순

AI 박수빈 기자 · 2025.04.17.

SKT, AI로 결과 분석하는 AI 서베이 리포트 공개

(사진=SKT) SK텔레콤(대표 유영상)은 인공지능(AI) 기반의 설문 조사 결과와 분석 리포트를 제공하는 'SKT AI 서베이' 서비스를 통해 분석 결과를 무료로 공개한다고 17일 밝혔다. SKT 사용자들은 T멤버십 앱의 '미션' 탭에 들어가 정기적으로 올라오는 다양한 주제의 설문에 응하고, T플러스포인트를 지급받는다. SKT의 AI와 빅데이터를 활용, 특정 설문대상에 맞춤형 설문 조사도 가능하다. 설문은 SKT가 자체 제작하거나, 기업과 연구기관·소상공인 등 외부 설문 의뢰를 받아 설계한다. 문항...

0 2

#SK텔레콤

#AI

#설문 조사

#빅데이터

#대형언어모델

AI 박찬 기자 · 2025.04.21.

오픈AI·구글, AI 안전 정보 미공개로 빈축

(사진=오픈AI) 오픈AI와 구글이 최근 최신 인공지능(AI) 모델을 공개하며 세부적인 안전 정보를 공개하지 않아 빈축을 사고 있다. 숨 가쁘게 모델이 쏟아지는 가운데, 시스템 카드 공개가 늦어지는 것은 물론, 이제는 아예 건너뛰는 사례까지 발생하고 있다. 구글은 17일(현지시간) '제미니 2.5 프로' 안전 평가 결과를 담은 기술 보고서를 공개했다. 이는 3월26일 모델 출시 후 거의 3주 만이다. 하지만, 세부 정보 부족으로 인해 실질적인 위험 평가가 어렵다는 지적이 잇따르고 있다. AI 모델의 위...

0 0

#오픈AI

#구글

#AI 모델

#안전 보고서

#위험 평가



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO

상호작용 (좋아요 기능)

POST /news/like/

http://127.0.0.1:8000/news/like/?article_id=622

DELETE /news/like/

http://127.0.0.1:8000/news/like/?article_id=622

GET /news/like/

http://127.0.0.1:8000/news/like/?article_id=622



기타

'원료의약품 사업 순항' 유한양행, 길리어드에 800억원대 공급계약

이지현 2025.05.22.

유한양행 본사. 유한양행 제공

유한양행

이 미국 길리어드사이언스와 800억원대 인간면역결핍바이러스(HIV) 원료의약품 공급 계약을 맺었다. 올해부터 수출이 잇따를 것으로 예상돼 실적 개선에 도움이 될 것이란 평가다.

유한양행은 길리어드와 888억1494만원(6377만달러) 규모의 HIV 치료제 원료의약품 공급 계약을 맺었다고 22일 공시했다. 계약 기간은 내년 12월31일까지다. 길리어드는 유한양행이 대금을 청구한 날로부터 한달 안에 청구액을 지급할 예정이다.

유한양행은 자회사 유한화학을 통해 길리어드와 베링거인겔하임 등의 원료를 공급하고 있다. 앞서 지난해 9월 유한양행은 길리어드와 8089만달러(당시 환율 기준 1076억6965만원) 규모의 HIV 치료제 원료 의약품 공급 계약을 맺었다고 공시했다.

업계에선 유한양행이 원료를 공급하는 길리어드사이언스의 HIV 치료제를 '선렌카'로 추정하고 있다. 선렌카는 '항바이러스제 명가'로 꼽히는 길리어드에서 2022년 미국에 출시한 HIV 치료용 주사제다. 길리어드는 이 약을 연 1회 투여하는 예방용으로 개발하고 있다. 특별한 문제가 없으면 올해 6월께 미국 식품의약국(FDA) 승인을 받을 수 있을 것으로 예상된다. 올해부터 유한화학의 원료의약품 공급 매출도 급격히 늘 것이란 전망이 나오는 이유다.

선렌카는 글락소스미스클라인(GSK)의 보카브리아+레캄비스와 함께 HIV 치료제 시장 판도를 장기지속형 주사제 시장으로 전환하고 있다. 장기 지속형 주사제는 HIV 환자가 매일 약을 먹는 불편을 줄여줘 치료 패러다임을 바꿀 것이란 평가를 받고 있다.

이지현 기자 bluesky@hankyung.com

#유한양행

#길리어드사이언스

#HIV

#원료의약품

#선렌카

♥ 1

👁 12



♥ 좋아요



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO



댓글 기능

GET /news/{article_id}/comments/

http://127.0.0.1:8000/news/1798/comments/

POST /news/{article_id}/comments/

http://127.0.0.1:8000/news/1798/comments/

PUT /comments/{comment_id}/

http://127.0.0.1:8000/comments/1/

DELETE /comments/{comment_id}/

http://127.0.0.1:8000/comments/1/

댓글 2개

댓글을 작성해 주세요 ...

Ctrl + Enter 로 작성

작성하기

admin

2025.05.22.



테스트 수정 댓글

답글 달기

admin

2025.05.22.



테스트 대댓글



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO

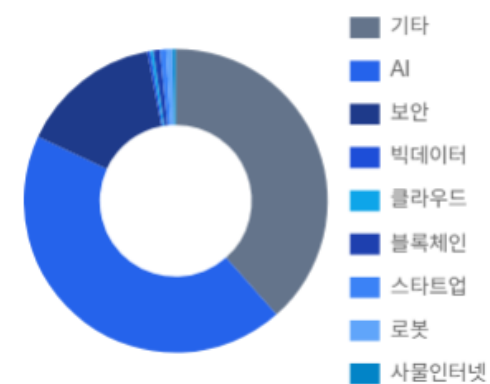
개인 대시보드

GET /dashboard/

http://127.0.0.1:8000/dashboard/

🐣 나의 관심 카테고리

주로 읽은 기사들을 분석하여 AI, 보안, 스타트업 등 가장 관심 있는 기술 분야를 한눈에 보여드립니다.



1순위: AI (130개)

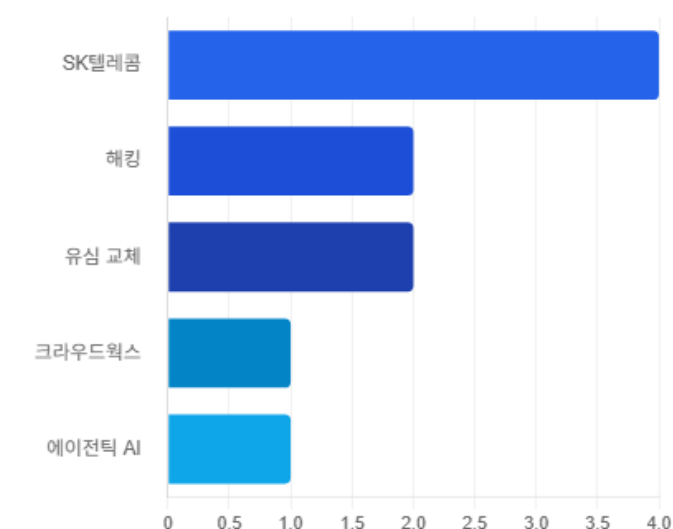
2순위: 기타 (115개)

3순위: 보안 (45개)

4순위: 블록체인 (2개)

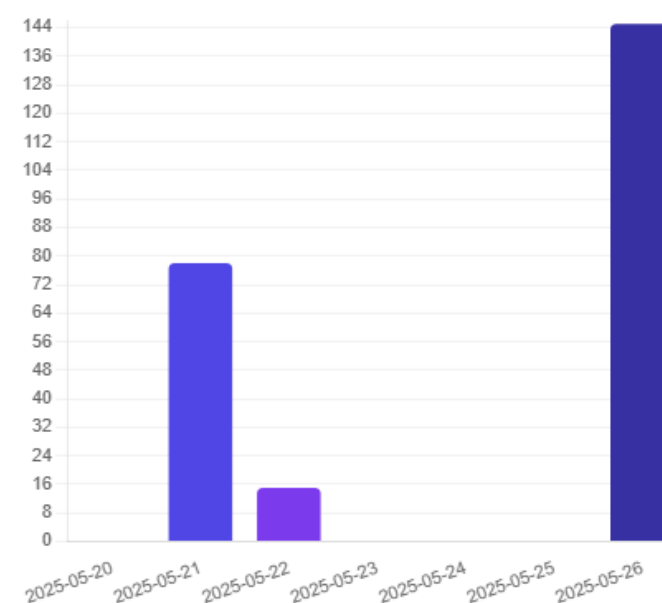
★ 주요 키워드

읽은 기사들의 핵심 키워드를 분석하여 현재 관심 있는 기술 트렌드를 보여드립니다.



📰 주간 읽은 기사

최근 일주일간의 기사 조회 기록을 통해 기술 뉴스 학습 패턴을 분석합니다.



❤️ 좋아요 누른 기사

관심 있는 기술 기사들을 모아보고 나만의 학습 자료로 활용하세요.

AI 클라우드웍스, '아시아테크 싱가포르'서 해외 비즈니스 전개

2025.05.22.

보안 SK텔레콤, 3년간 악성코드 탐지 못했다..."보안 미흡" 인정

2025.05.20.

보안 SKT 가입자, 1인당 고작 '3000원'...정보보호 투자 '부실'

2025.05.20.

기타 SK바사, 폐구균 백신 원액 수출 가능해졌다...화이자



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO



AI 챗봇

POST /news/chat/

http://127.0.0.1:8000/news/chat/

AI 뉴스비서 뉴비

이 기사에 대해 궁금한 점을 물어보세요!

안녕하세요! 저는 AI 뉴스비서 뉴비입니다. 이 기사에 대해 궁금한 점이 있으시다면 언제든지 물어보세요! 😊

IMEI가 뭐야?

IMEI는 "International Mobile Equipment Identity"의 약자로, 모바일 기기의 고유 식별 번호입니다. 이 번호는 각 휴대폰에 고유하게 부여되어 있으며, 기기를 식별하고 추적하는 데 사용됩니다. IMEI 번호는 보통 15자리 숫자로 구성되어 있으며, 기기가 도난당했거나 분실된 경우에도 이를 통해 기기를 차단하거나 추적할 수 있습니다.

메시지를 입력하세요...

전송

AI 뉴스비서 뉴비

이 기사에 대해 궁금한 점을 물어보세요!

되어 있으며, 기기를 식별하고 추적하는 데 사용 됩니다. IMEI 번호는 보통 15자리 숫자로 구성되어 있으며, 기기가 도난당했거나 분실된 경우에도 이를 통해 기기를 차단하거나 추적할 수 있습니다.

기사 짧게 요약해줘

SK텔레콤의 가입자 유심 정보 유출 사고가 발생했으며, 전문가들은 이 사건을 심각하게 받아들여야 한다고 경고하고 있습니다. 유심 정보 유출은 금융 사기에 연루될 가능성이 있어 주의가 필요하며, SK텔레콤은 민감한 개인정보는 유출되지 않았다고 밝혔습니다. 전문가들은 SK텔레콤 가입자들에게 각종 계정 비밀번호를 변경할 것을 권고하고 있습니다.

메시지를 입력하세요...

전송



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO



AI 챗봇 - 프롬프트 엔지니어링

POST /news/chat/

`http://127.0.0.1:8000/news/chat/`

프롬프트 설계 기준:

1. 뉴스 기사 기반의 사실 기반 응답
2. 쉽고 친절한 문체로 대답
3. 추측 없이, 기사 정보에 집중
4. 배경 설명과 추가 정보도 함께 제공

SYSTEM_PROMPT = (

"당신은 친절한 뉴스 비서 '뉴비'입니다. "

"주어진 뉴스 기사에 대해 정확하고 상세하게 답변해주세요. "

"답변할 때는 다음 규칙을 따라주세요:\N"

"1. 기사 내용을 바탕으로 객관적이고 사실에 기반한 답변을 제공하세요.\N"

"2. 복잡한 내용은 쉽게 설명해주세요.\N"

"3. 관련 배경 정보나 추가 설명이 필요한 경우 제공하세요.\N"

"4. 답변은 친절하고 자연스러운 대화체로 작성하세요.\N"

"5. 기사 내용에 없는 정보는 추측하지 마세요."

)



Frontend / Backend

백엔드 API 구성 및 프론트엔드 DEMO



AI 챗봇 - LANGCHAIN

POST /news/chat/

<http://127.0.0.1:8000/news/chat/>

OPENAI EMBEDDINGS로 뉴스 문단 벡터화

CHROMA 벡터스토어에 저장

LANGCHAIN CONVERSATIONALRETRIEVALCHAIN

- 질문 → 벡터 유사 검색 → 답변 생성
- 유사도 기반 뉴스 3건 검색 (K=3)
- 이전 대화 이력 기반 응답 (MEMORY 유지)

```
EMBEDDINGS = OPENAIEMBEDDINGS()
VECTORSTORE = CHROMA.FROM_TEXTS(
    TEXTS=TEXTS,
    EMBEDDING=EMBEDDINGS,
)
```

```
QA_CHAIN = CONVERSATIONALRETRIEVALCHAIN.FROM_LLM(
    LLM=LLM,
    RETRIEVER=VECTORSTORE.AS_RETRIEVER(
        SEARCH_TYPE="SIMILARITY",
        SEARCH_KWARGS={"K": 3}
    ),
    MEMORY=MEMORY,
    RETURN_SOURCE_DOCUMENTS=TRUE
)
```