

**Software Requirements Specification**  
for  
**Epoch Pulse.**

Version 1.0

June 01, 2026

# Contents

<b>1 Introduction</b> .....	<b>4</b>
1.1 Purpose .....	4
1.2 Scope .....	4
1.3 Definitions, acronyms, and abbreviations .....	4
1.4 References .....	5
1.5 Overview .....	6
<b>2 Overall Description</b> .....	<b>6</b>
2.1 Product Perspective .....	6
2.2 Product Functions .....	7
2.3 User Characteristics .....	7
2.4 Constraints .....	8
2.4.1 Hardware and Operating Environment .....	8
2.4.2 Operational and Design Constraints .....	8
2.5 Assumptions and Dependencies .....	8
2.6 Apportioning of Requirements .....	9
<b>3 Specific Requirements</b> .....	<b>9</b>
3.1 External Interface Requirements .....	9
3.1.1 User Interfaces .....	9
3.1.2 Hardware Interfaces .....	10
3.1.3 Software Interfaces .....	10
3.1.4 Communications Interfaces .....	11
3.2 Functional Requirements .....	11
3.2.1 실시간 시장 감시 및 이상 변동 탐지 .....	11
3.2.1.1 FR-1 Multi-Symbol Market Data Ingestion .....	12
3.2.1.2 FR-2 Anomaly Detection .....	12
3.2.2 이상 변동 시점의 뉴스 수집 및 요약 .....	12
3.2.2.1 FR-3 Real-Time News Collection and Summarization .....	13
3.2.3 뉴스 신뢰도 검증 및 감성 분석 기반 시장 인사이트 도출 .....	13
3.2.3.1 FR-4 NLI Relevance Filtering .....	13
3.2.3.2 FR-5 ABSA Sentiment Analysis .....	14
3.2.3.3 FR-6 Impact Score and News Ordering .....	14
3.2.3.4 FR-7 Event-Level Aggregation .....	15
3.2.4 데이터 영속성 및 실시간 서빙 .....	15
3.2.4.1 FR-8 Persistence and Serving .....	15
3.3 Performance Requirements .....	16
3.4 Logical Database Requirements .....	16
3.5 Design Constraints .....	17
3.6 Software System Attributes .....	18
3.6.1 Availability .....	18
3.6.2 ABSA Model Quality .....	18
3.6.3 Usability .....	18
<b>4 Appendices</b> .....	<b>18</b>
4.1 Appendix A: Kafka Topic 및 데이터 스키마 상세 규격 .....	18
4.1.1 1. 주가 데이터 (stock-prices) .....	19

---

4.1.2	2. 이상 탐지 이벤트 (anomaly-event)	19
4.1.3	3. 통합 뉴스 분석 객체 (Unified Analysis Object)	20
4.1.4	4. 토픽별 발행 및 진행 규칙	20
4.1.5	5. 이벤트 요약 결과 (Event Summary Result)	21
4.2	Appendix B: Event Summary Calculation	21
4.3	Appendix C: Impact Score Weight Matrix	21

# 1 Introduction

## 1.1 Purpose

이 문서는 이벤트 기반 실시간 금융 이상 탐지 및 뉴스 감정 분석 투자 보조 플랫폼 Epoch Pulse.의 소프트웨어 요구사항 명세서이다.

Epoch Pulse.는 실시간 주가 데이터에서 이상 변동을 탐지하고, 이상 이벤트 발생 시점의 금융 뉴스를 수집, 요약, 관련성 검증, 감정 분석, 이벤트 요약 산출까지 수행하여 웹 대시보드에서 시장 변동 원인을 사용자 스스로 해석할 수 있게 하는 투자 보조용 시장 인사이트 플랫폼이다.

## 1.2 Scope

Epoch Pulse.의 제품 범위는 다음과 같다.

- 포함 범위:
  - 실시간 시장 데이터 API에서 다중 심볼의 가격 데이터를 수집한다.
  - 가격 변동률과 지수 이동 평균 기반 이상 탐지 알고리즘으로 이상 이벤트 탐지 및 발행한다.
  - 이상 이벤트 시점 전후 20분 범위의 해당 심볼 관련 뉴스를 실시간 뉴스 API, RSS 계열 인터페이스, GDELT Project 등에서 수집한다.
  - OpenAI 호환 API 구조의 외부 LLM 요약 API를 사용하여 뉴스 요약을 생성한다.
  - NLI 모델로 뉴스와 심볼 간 관련성을 검증하고, 관련성이 낮은 뉴스는 분석 파이프라인에서 제외한다.
  - ABSA 모델로 뉴스가 대상 심볼에 대해 긍정, 중립, 부정 중 어느 감성을 갖는지 산출한다.
  - 개별 뉴스 분석 결과를 이상 이벤트 단위로 집계하여 종합 감정, 일관성, 신뢰도, 핵심 근거 지표를 제공한다.
  - 단일 사용자 웹 대시보드에서 가격 차트, 대표 이상치, 뉴스 분석, 이벤트 요약을 실시간으로 표시한다.
- 제외 범위:
  - 주식 주문, 매매 체결, 포트폴리오 리밸런싱, 투자 자문 자동 실행은 포함하지 않는다.
  - 다중 사용자 계정, 로그인, 권한 관리, 개인화 설정, 보안은 포함하지 않는다.
  - 고빈도 거래 수준의 초저지연 체결 또는 거래소 직접 연결은 포함하지 않는다.
  - Streamlit 기반 개발 및 데모 제어 화면은 최종 사용자 인터페이스로 간주하지 않는다.

## 1.3 Definitions, acronyms, and abbreviations

Term	Definition
Active Symbol (활성 심볼)	사용자가 대시보드에서 감시 대상으로 등록하고 실시간 가격 수집, 이상 탐지, 뉴스 분석 대상으로 처리되는 심볼.

Anomaly Event (이상 이벤트)	가격 변동률 또는 이동 평균 이탈이 설정된 임계값을 초과하여 생성된 원시(Raw) 이상 탐지 단위. 각 이상 이벤트는 전후 20분 내의 뉴스들과 1:N으로 매핑되어 뉴스 수집 및 분석의 트리거가 된다.
Anomaly Cluster (이상치 클러스터)	발생 시간 간격이 15분 이내이고 가격 변동 방향(상승/하락)이 동일한 연속된 Anomaly Event들의 논리적 그룹.
Representative Anomaly (대표 이상치)	하나의 Anomaly Cluster 내에서 이상 점수(Score)가 가장 높은 단일 Anomaly Event. 차트 마커 표시 및 대시보드의 뉴스 필터링 기준점이 된다.
Directional Alignment (방향성 일치)	Representative Anomaly의 주가 변동 방향(상승/하락)과 해당 뉴스의 감성(Positive/Negative)이 일치하는 상태. Impact Score 산출 시 가중치에 반영된다.
Confidence Score (확신도)	NLI 또는 ABSA 모델이 특정 라벨로 분류할 때 산출한 예측 확률 값(0.0 ~ 1.0).
Impact Score (영향도 점수)	NLI 관련성 점수, ABSA 감성 점수, 그리고 Directional Alignment(방향성 일치 여부)를 결합한 개별 뉴스의 종합 영향도 점수.
Event Summary (이벤트 요약)	Representative Anomaly에 연결된 여러 개별 뉴스 분석 결과들을 집계하여 산출한 종합 감성, 일관성, 신뢰도 등 수치 기반 종합 지표 및 통계.
Unified Analysis Object (통합 뉴스 분석 객체)	뉴스 수집부터 최종 분석까지 이어지는 파이프라인 전반에서 뉴스 정보, 분석 결과(NLI/ABSA), 처리 통계 등을 공유하기 위해 사용되는 통일된 JSON 스키마 구조.
Choreography-based MSA	중앙 오케스트레이터 없이 각 마이크로서비스가 Kafka 이벤트를 자율적으로 소비 및 발행하여 파이프라인을 구성하는 아키텍처 패턴.
EMA (지수이동평균)	Exponential Moving Average. 최근 데이터에 더 높은 가중치를 두어 주가 추세를 파악하는 이상 탐지 알고리즘.
Robust Z-score (로버스트 Z-스코어)	중앙값(Median) 등을 활용하여 극단치에 의해 평균과 표준편차가 왜곡되는 것을 방지하는 통계적 이상 탐지 방식.
LLM (대규모 언어 모델)	Large Language Model. 크롤링된 뉴스 원문에서 시장 변동과 관련된 핵심 맥락만을 1~2문장으로 추출(요약)하는 데 사용된다.
NLI (자연어 추론)	Natural Language Inference. 뉴스가 해당 심볼의 가격 변동과 의미적으로 관련되는지 검증(Entailment/Neutral/Contradiction)하는 자연어 추론 방식.
ABSA (속성 기반 감성 분석)	Aspect-Based Sentiment Analysis. 특정 심볼(Aspect)에 대해 해당 뉴스가 긍정(Positive), 중립(Neutral), 부정(Negative) 중 어느 감성을 갖는지 분류하는 분석 방식.

## 1.4 References

1. IEEE Std 830-1998, IEEE Recommended Practice for Software Requirements Specifications.
2. Intel Corporation, Intel Processor N100 official specifications.
3. Jakob Nielsen, Response Times: The 3 Important Limits.
4. Apache Kafka Documentation.

- 5. PostgreSQL and TimescaleDB Documentation.
- 6. Socket.IO Documentation.

### 1.5 Overview

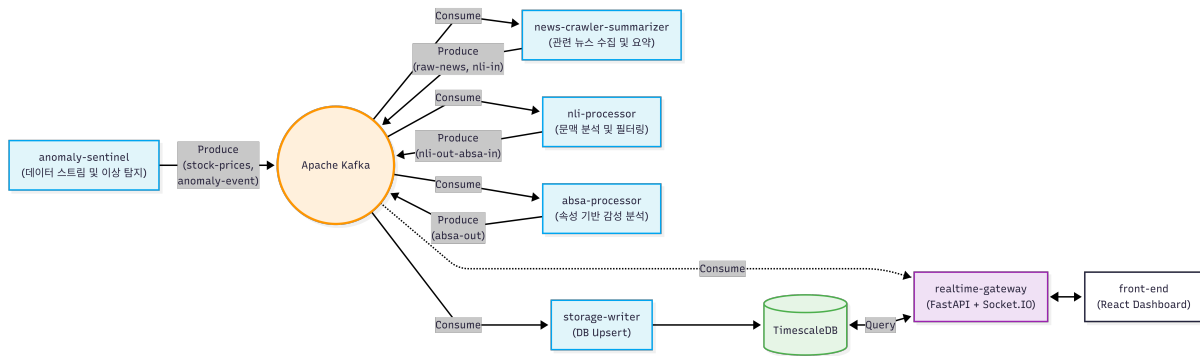
제1장은 본 문서의 목적과 범위를 정의한다. 제2장은 제품의 거시적인 관점, 사용자 특성, 운영 환경 및 제약 사항을 설명하여 시스템의 전체적인 윤곽을 그린다. 이를 바탕으로 제3장에서는 외부 인터페이스, 기능 요구사항, 성능 요구사항, 설계 제약 및 품질 속성 등 구체적이고 기술적인 명세를 정의한다. 마지막으로 부록(Appendices)은 본문의 이해와 요구사항 추적을 돕기 위해 주요 데이터 계약(Kafka Schema)과 이벤트 집계 공식을 제공한다.

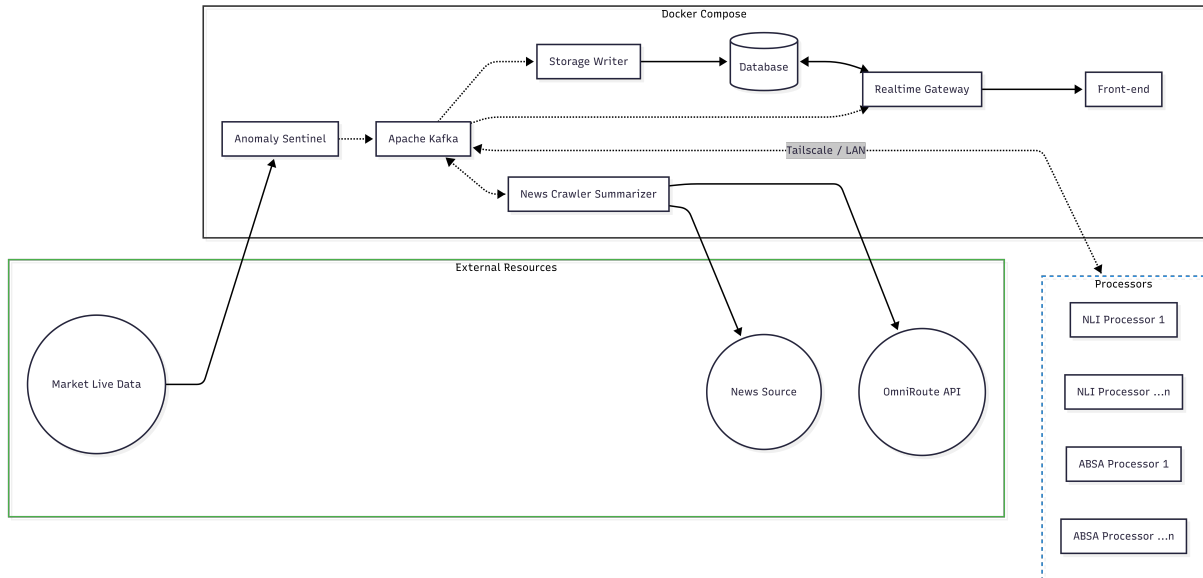
## 2 Overall Description

### 2.1 Product Perspective

Epoch Pulse.는 이벤트 기반 마이크로서비스 시스템이다. 시스템 내부 서비스는 Kafka 토픽을 통해 비동기 이벤트 흐름으로 연결되고, 데이터 조회와 실시간 브로드캐스트는 Realtime Gateway를 통해 웹 대시보드에 제공된다.

시스템 바운더리와 내부 아키텍처는 다음과 같다.





Boundary Area	Responsibility
Inside System	시장 데이터 수집 어댑터, 이상 탐지, 뉴스 수집 어댑터, 뉴스 요약 요청, NLI/ABSA 분석, 이벤트 집계, DB 저장, REST/Socket.IO 제공, 웹 대시보드.
External Systems	시장 데이터 API, 실시간 뉴스 API/RSS, 외부 LLM 요약 API, 운영체제, 컨테이너 runtime, 네트워크, 브라우저.
Explicitly Outside	주식 주문 실행, 사용자 인증, 다중 사용자 권한, 세금/회계 처리, 법적 투자 자문.

## 2.2 Product Functions

Epoch Pulse.는 다음 상위 기능을 제공해야 한다.

1. 다중 심볼 가격 수집 및 정규화.
2. 가격 이상 탐지 및 이상 이벤트 발행.
3. 이상 이벤트 기반 실시간 뉴스 수집.
4. 뉴스 요약 및 진행 상태 제공.
5. NLI 기반 관련성 필터링.
6. ABSA 기반 심볼 감성 분석.
7. 이벤트 단위 집계 분석.
8. 시계열 데이터와 분석 결과의 영속 저장.
9. 웹 대시보드의 메인 요약, 심볼 상세 차트, 이상치, 뉴스, 이벤트 요약 표시.

## 2.3 User Characteristics

최종 제품의 사용자 클래스는 단일 일반 사용자로 정의한다. 일반 사용자는 별도 로그인 없이 하나의 브라우저 세션에서 메인 페이지의 대시보드 요약을 확인하고, 관심 심볼을 검색하여 심볼별 상세 페이지에

서 차트와 뉴스 분석 결과를 확인한다. 시스템은 동시 다중 사용자 운영, 역할별 권한 분리, 사용자별 설정 저장을 요구하지 않는다.

Streamlit 제어 화면, Kafka UI, DB UI는 제품 사용자 클래스에 포함하지 않는다.

## 2.4 Constraints

### 2.4.1 Hardware and Operating Environment

본 문서에서 Base Node는 Intel Processor N100급 4코어/4스레드 CPU와 16GB 메모리를 갖춘 단일 PC 운영 환경을 의미한다. 전체 시스템은 Docker Compose로 실행되며, 다음 환경 제약을 따른다.

- 시스템의 코어 서비스(Realtime Gateway, Anomaly Sentinel, News Crawler/Summarizer, Storage Writer, TimescaleDB, Front-end) 및 Kafka Broker는 Base Node에서 단일 인스턴스로 운영한다.
- NLI와 ABSA 추론 프로세서는 CPU-only 실행을 허용하며, GPU 가용성을 필수로 요구하지 않는다.
- 단일 구성 시에는 Base Node의 CPU와 메모리 한계를 고려하여 추론 프로세서의 동시 실행 수를 제한할 수 있다.
- NLI Processor와 ABSA Processor는 외부 컴퓨팅 자원을 활용한 수평 확장을 허용한다.

### 2.4.2 Operational and Design Constraints

- 심볼별 가격 수집 주기는 1분 이상이어야 한다.
- 하나의 이상 이벤트에 대해 분석 후보 뉴스는 최대 20건으로 제한한다.
- Kafka 메시지는 JSON UTF-8 형식으로 교환해야 한다.
- 외부 확장이 허용된 토픽(nli-in, nli-out-absa-in)의 파티션 수는 12개를 상한으로 한다.
- DB에는 최종 뉴스 분석 상태(DONE, DROPPED)만 저장해야 하며, 중간 상태는 실시간 스트림으로만 제공해야 한다.
- 외부 LLM 요약 API 장애 시 시스템은 원문 또는 스니펫을 대체 본문으로 사용하고 파이프라인을 계속 진행해야 한다.
- 실시간 시장 데이터 API와 실시간 뉴스 API는 벤더 교체 가능하도록 어댑터 패턴으로 분리, 구현해야 한다.
- 최종 사용자 인터페이스는 웹 대시보드 하나로 제한한다.
- 뉴스 요약 처리는 다수의 이상 이벤트가 동시 발생하여 처리할 뉴스 건수가 급증하더라도, 전체 파이프라인 지연 기준을 초과하지 않도록 병목을 완화할 수 있어야 한다.

## 2.5 Assumptions and Dependencies

시스템은 다음 가정에 의존한다.

- 시장 데이터 API는 최소 20개 심볼에 대해 1분 이상의 수집 주기를 허용한다.
- 뉴스 API 또는 RSS는 이상 이벤트 시각 기준 전후 뉴스 검색을 지원하거나, 수집 후 시스템 내부에서 시간 필터링 가능하다.

- 외부 LLM API는 OpenAI 호환 인터페이스 또는 동등한 chat completion 인터페이스로 요약 요청을 처리할 수 있다.
- NLI와 ABSA 모델은 로컬 모델 파일로 로드 가능하며, CPU-only 추론으로도 동작한다.
- 사용자는 결과를 투자 참고 정보로만 사용하며, 시스템 출력은 매매 권고 또는 자동 주문 신호가 아니다.
- 본 시스템의 성능 보장은 외부 뉴스 API 및 외부 LLM API가 정상적인 응답 속도와 가용성을 제공하는 가정하에 성립한다.
- NLI/ABSA Processor가 외부 서버에 배치될 경우, 해당 서버는 Base Node의 내부 Kafka Broker와 네트워크로 통신할 수 있다.

## 2.6 Apportioning of Requirements

모든 기능 요구사항, 외부 인터페이스 요구사항, 데이터 요구사항, 성능 요구사항, 품질 요구사항은 Epoch Pulse. 시스템의 필수 요구사항이다. 본 명세에서 차기 버전으로 배분하거나 보유하는 요구사항은 없다.

# 3 Specific Requirements

## 3.1 External Interface Requirements

### 3.1.1 User Interfaces

UI-1. 시스템은 React 18.3.1 기반 웹 대시보드를 최종 사용자 인터페이스로 제공해야 하며, 대시보드는 메인 페이지와 심볼별 상세 페이지로 구성되어야 한다.

UI-2. 메인 페이지는 활성 심볼 수, 분석된 뉴스 수, 최근 발견된 이상 이벤트 (최대 10건), 심볼 검색 및 심볼별 상세 페이지 이동 기능을 제공해야 한다.

UI-3. 심볼별 상세 페이지는 선택된 심볼의 가격 데이터를 시간순 라인 차트로 표시하고, 동일 분 단위의 중복 가격 데이터는 최신값으로 병합해야 한다.

UI-4. 심볼별 상세 페이지는 15분 이내의 동일 방향 이상 이벤트를 이상치 클러스터로 묶고, 점수의 절대값이 가장 높은 이벤트를 대표 이상치로 선택하여 차트 마커로 표시해야 한다.

UI-5. 심볼별 상세 페이지는 차트 마커 표시 시 시각적 위계를 두어 일반 이상치는 빨간 점, 대표 이상치는 노란 원으로 구분하되, 대표 이상치와 같은 방향의 DONE 뉴스가 존재할 경우 해당 마커를 주황색으로 강조 표시해야 한다.

UI-6. 심볼별 상세 페이지는 선택된 심볼과 관련된 뉴스의 수집, 요약, 관련성 검증, 최종 분석 상태를 실시간으로 표시해야 한다.

UI-7. 심볼별 상세 페이지는 NLI 분석 결과를 통해 해당 뉴스가 대상 심볼과 관련성이 있는지 확인할 수 있어야 한다.

UI-8. 심볼별 상세 페이지는 ABSA 분석 결과를 Positive, Neutral, Negative 중 하나의 감성 라벨과 신뢰도 점수로 표시해야 한다.

UI-9. 심볼별 상세 페이지는 차트의 이상치 마커(대표 이상치 또는 일반 이상치) 클릭 시 해당 마커가 속한 이상치 클러스터를 선택 또는 선택 해제할 수 있어야 하며, 클러스터 선택 시 뉴스 목록은 해당 클러스터의 시간 범위(시작 시각 종료 시각) 전후 20분 내의 뉴스 항목으로 즉각 필터링되어야 한다.

UI-10. 심볼별 상세 페이지는 특정 이상치 클러스터가 선택된 경우 해당 클러스터 내 대표 이상치의 방향성을 기준으로, 선택된 것이 없는 경우 개별 뉴스와 연결된 원천 이상 이벤트의 방향성을 기준으로 뉴스들의 영향도 점수를 산출하여 정렬 제공해야 한다. 또한 클러스터 선택 시, 해당 대표 이상치에 연결된 여러 뉴스 분석 결과를 종합하여 산출된 이벤트 요약 지표를 함께 제공해야 한다.

### 3.1.2 Hardware Interfaces

HW-1. 시스템은 별도 전용 하드웨어 없이 x86-64 PC에서 동작해야 한다.

HW-2. 기준 운영 환경(Base Node)은 Intel Processor N100급 CPU, 16GB RAM, 100GB 용량 이상의 SSD 저장소를 갖춘 단일 PC이다.

HW-3. GPU는 선택 사항이며, GPU가 없더라도 NLI와 ABSA 워커는 CPU-only 모드로 동작해야 한다.

HW-4. 시스템은 단일 서버 환경을 기본 운영 환경으로 하되, NLI Processor와 ABSA Processor는 별도 서버 또는 외부 컴퓨팅 환경으로 확장 가능해야 한다.

### 3.1.3 Software Interfaces

SI-1. 시스템은 시장 데이터 API로부터 RESTful HTTP 또는 WebSocket을 통해 심볼, timestamp, price, previous price, volume을 포함한 가격 이벤트를 수신해야 한다.

SI-2. 시스템은 뉴스 API 또는 RSS 인터페이스로부터 HTTP GET 등을 통해 title, content 또는 summary, source, url, published timestamp를 포함한 뉴스를 수신해야 한다.

SI-3. 시스템은 외부 LLM 요약 API에 뉴스 title, content, target symbol을 전달하고 1-2문장 요약을 수신해야 한다.

SI-4. LLM 요약 인터페이스는 특정 공급자에 종속되지 않아야 하며, OpenAI 호환 API 또는 동등한 추상 인터페이스로 교체 가능해야 한다.

SI-5. 시스템은 PostgreSQL 17 기반 TimescaleDB에 가격, 이상 이벤트, 최종 뉴스 분석, 이벤트 요약 데이터를 저장해야 하며, 모든 소프트웨어 간 데이터 교환은 JSON 포맷을 사용해야 한다.

SI-6. NLI Processor는 Hugging Face Transformers 라이브러리를 사용하여 DeBERTa 기반 NLI 모델과 토큰나이저를 로드하고 추론을 수행해야 한다.

SI-7. ABSA Processor는 Hugging Face Transformers 라이브러리를 사용하여 DeBERTa 기반 ABSA 모델과 토큰나이저를 로드하고 추론을 수행해야 한다.

SI-8. 시스템은 Python 3.10+ 환경에서 Anomaly Sentinel, Realtime Gateway 서비스를, React 18.3.1 환경에서 Front-end 서비스를 구동하며 해당 런타임 표준을 준수해야 한다.

SI-9. 시스템은 Apache Kafka 3.9.2 (KRaft Mode)를 메시지 미들웨어로 사용하여 서비스 간 이벤트를 소비 및 발행해야 한다. 뉴스 분석 파이프라인 전반(raw-news absa-out)에서 교환되는 메시지는 통합 뉴스 분석 객체(Unified Analysis Object) 스키마를 사용하여 데이터 일관성을 유지해야 하며, 각 분석 단계를 거칠 때마다 해당 필드의 데이터를 점진적으로 채워 넣어야 한다.

SI-10. Kafka 파이프라인의 진행 상태를 제어하기 위해 다음 정규화된 status 값을 사용해야 한다: COLLECTED(수집 직후), SUMMARIZED(요약 완료), NLI\_PASSED(연관성 통과), NLI\_DROPPED(연관성 탈락), DONE(분석 완료).

SI-11. 모든 마이크로서비스는 Docker 27.x 및 Docker Compose 2.x 환경의 컨테이너 인터페이스 내에서 실행 및 관리되어야 한다.

### 3.1.4 Communications Interfaces

CI-1. 내부 서비스 간 이벤트 교환은 Kafka 프로토콜을 사용하며, 내부 통신은 TCP 9092, 외부 확장을 위한 노출은 TCP 20000(LAN) 및 20001(Tailscale) 포트를 사용해야 한다.

CI-2. Realtime Gateway는 TCP 20004 포트를 통해 대시보드에 REST API와 Socket.IO 이벤트를 제공해야 한다.

CI-3. Realtime Gateway는 다음 Socket.IO 이벤트를 통해 대시보드에 실시간 데이터를 전송해야 한다. 각 이벤트의 페이로드 형식은 부록 A의 데이터 스키마 상세 규격을 준수해야 한다.

1. price:tick: 부록 A.1의 주가 데이터 스키마를 사용하여 실시간 틱 데이터를 브로드캐스트한다.
2. anomaly:detected: 부록 A.2의 이상 탐지 이벤트 스키마를 사용하여 탐지 정보를 알린다.
3. news:progress: 부록 A.3의 통합 뉴스 분석 객체 스키마를 사용하며, status가 COLLECTED, SUMMARIZED, NLI\_PASSED인 중간 상태 데이터를 전송한다.
4. news:final: 부록 A.3의 통합 뉴스 분석 객체 스키마를 사용하며, status가 DONE 또는 NLI\_DROPPED인 최종 상태 데이터를 전송한다.
5. event:summary: 부록 A.5의 이벤트 요약 결과 스키마를 사용하여 특정 대표 이상치에 대한 종합 분석 지표를 전송한다.
6. symbol:clear: 특정 심볼의 실시간 스트림 데이터 초기화가 필요할 때 전송한다.

CI-4. Kafka 메시지와 REST 응답 데이터는 서로 다른 시스템 간에 원활하게 통신할 수 있도록 모두 JSON 텍스트 포맷으로 변환하여 전송해야 하며, 모든 타임스탬프는 ISO8601 UTC 포맷을 준수하여 정규화되어야 한다.

## 3.2 Functional Requirements

### 3.2.1 실시간 시장 감시 및 이상 변동 탐지

사용자는 관심 심볼의 실시간 주가 흐름을 모니터링하고, 비정상적인 가격 변동을 즉시 인지할 수 있어야 한다.

### 3.2.1.1 FR-1 Multi-Symbol Market Data Ingestion

FR-1.1. 시스템은 주식 시장의 활성 거래 시간에 한하여 설정 파일로 지정된 모든 활성 심볼의 가격 데이터를 1분 주기로 수집해야 한다.

FR-1.2. 시스템은 각 가격 이벤트를 `symbol`, `event_time`, `price`, `previous_price`, `change_percent`, `volume` 필드로 정규화해야 한다.

FR-1.3. 시스템은 정규화된 가격 이벤트를 `stock-prices` 토픽에 발행해야 한다.

FR-1.4. 시스템은 수신된 가격 이벤트를 시계열 DB에 저장해야 한다.

FR-1.5. 시스템은 동일 심볼과 동일 timestamp 가격 데이터가 중복 수신될 경우 최신 payload로 `upsert`해야 한다.

### 3.2.1.2 FR-2 Anomaly Detection

#### 탐지 알고리즘

FR-2.1. 시스템은 가격 변동률에 대해 Robust Z-score 기반 이상 탐지를 수행해야 하며, 최소 10개의 누적 데이터를 기준으로 임계값  $3.0\sigma$ 를 초과할 때 이상으로 판별해야 한다.

FR-2.2. 시스템은 가격에 대해 EMA 기반 이동 평균 이탈 탐지를 수행해야 하며, EMA(기간 20틱) 대비 3.0%를 초과하여 이탈할 때 이상으로 판별해야 한다.

FR-2.3. 이상 점수(score)는 알고리즘별로 다음과 같이 정의한다: Robust Z-score의 경우 표준편차 배수( $\sigma$ )를, EMA 기반 탐지의 경우 이동평균 대비 이탈 비율(%)을 점수로 사용한다.

FR-2.4. Z-score와 EMA 알고리즘이 동시에 이상을 탐지할 경우, 시스템은 이상 점수의 절대값이 더 높은 결과를 최종 채택하여 단일 이벤트로 처리해야 한다.

#### 이상 이벤트 식별 및 발행

FR-2.5. 시스템은 각 이상 이벤트를 고유하게 식별하기 위해 고유 식별자인 UUID를 발급하고, 심볼, 이벤트 시각, 탐지 시각, 가격, 이전 가격, 변동률, 알고리즘, 점수, 메시지를 함께 포함해야 한다.

FR-2.6. 시스템은 최종 판별된 이상 이벤트를 `anomaly-event` 토픽에 발행해야 한다.

FR-2.7. 시스템은 발생한 이상 이벤트를 시계열 DB에 저장해야 한다.

#### 이상치 클러스터링

FR-2.8. 시스템은 발생 시간 간격(15분 이내)과 변동 방향이 동일한 이상치들을 하나의 클러스터로 그룹화하고, 그중 점수의 절대값이 가장 높은 것을 대표 이상치로 선정해야 한다.

### 3.2.2 이상 변동 시점의 뉴스 수집 및 요약

사용자는 주가 이상 변동이 발생한 시점 전후의 관련 뉴스를 수집하고, 핵심 내용으로 요약된 형태로 신속히 파악할 수 있어야 한다.

### 3.2.2.1 FR-3 Real-Time News Collection and Summarization

#### 뉴스 수집 및 상태 발행

FR-3.1. 시스템은 이상 이벤트 발생 시 해당 심볼과 이벤트 시각을 기준으로 실시간 뉴스 수집을 시작해야 한다.

FR-3.2. 시스템은 이벤트 시각 전후 20분 범위의 뉴스만 해당 이벤트 후보로 유지해야 한다.

FR-3.3. 시스템은 하나의 이상 이벤트에 대해 최대 20건의 후보 뉴스만 분석 대상으로 전달해야 한다.

FR-3.4. 시스템은 수집 직후 COLLECTED 상태의 뉴스 진행 이벤트를 raw-news 토픽에 발행해야 한다.

#### 뉴스 요약 처리

FR-3.5. 시스템은 외부 LLM 요약 API를 통해 뉴스 요약을 생성하고, 성공 시 SUMMARIZED 상태로 갱신하여 raw-news와 nli-in 토픽에 각각 재발행 및 전파해야 한다.

FR-3.6. 요약 API가 실패하더라도 시스템은 원문 또는 스니펫을 그대로 사용하여 raw-news와 nli-in 토픽에 발행함으로써 후속 분석 처리를 계속 진행해야 한다.

#### 처리 제약 및 무결성

FR-3.7. 시스템은 하나의 이상 이벤트에 할당된 최대 20건의 후보 뉴스를 정해진 성능 기준 내에서 처리할 수 있어야 한다.

FR-3.8. 시스템은 동일 세션 내 중복 뉴스 수집 및 중복 분석을 방지할 수 있는 안정적인 뉴스 식별 체계를 제공해야 한다.

#### 처리 통계 기록

FR-3.9. 시스템은 뉴스 수집 및 요약에 소요된 시간을 metadata.processing\_stats.news\_duration에 기록해야 한다.

FR-3.10. 시스템은 뉴스 요약 처리가 완료된 시각을 metadata.trace\_timestamps.summarized\_at에 기록해야 한다.

### 3.2.3 뉴스 신뢰도 검증 및 감성 분석 기반 시장 인사이트 도출

사용자는 연관성과 감성 분석을 거친 신뢰도 높은 뉴스 정보를 선별하고, 종합 지표를 통해 시장 변동 원인을 해석할 수 있어야 한다.

#### 3.2.3.1 FR-4 NLI Relevance Filtering

##### NLI 입력 구성

FR-4.1. 시스템은 뉴스 콘텐츠 필드를 입력으로 대상 심볼 설명문 사이의 NLI 추론을 수행해야 한다.

FR-4.2. 시스템은 뉴스 content 필드와 symbol을 이용하여 NLI 입력 문장쌍을 생성해야 한다.

FR-4.3. NLI 입력 문장쌍은 뉴스 content와 “This news is about {symbol}.” 형식의 대상 설명문으로 구성되어야 한다.

FR-4.4. 시스템은 NLI 입력 생성 시 대상 심볼의 표시명과 분석용 명칭을 일관되게 정규화해야 한다.

##### 추론 및 결과 라우팅

FR-4.5. 시스템은 NLI 모델을 통해 entailment, neutral, contradiction 점수를 산출해야 한다.

FR-4.6. NLI entailment 점수가 0.5 이상인 뉴스는 NLI\_PASSED 상태로 nli-out-absa-in 토픽에 발행해야 한다.

FR-4.7. NLI entailment 점수가 0.5 미만인 뉴스는 스트림 상에서 NLI\_DROPPED 상태로 실시간 대시보드에 알림을 위해 absa-out 토픽에 발행하고, DB 저장 시 DROPPED 최종 상태로 정규화해야 한다.

FR-4.8. NLI 결과는 label, score, all\_scores를 포함해야 한다.

#### 처리 통계 기록

FR-4.9. 시스템은 NLI 처리 시간을 metadata.processing\_stats.nli\_duration에 기록해야 한다.

FR-4.10. 시스템은 NLI 처리 완료 시각을 metadata.trace\_timestamps.nli\_processed\_at에 기록해야 한다.

### 3.2.3.2 FR-5 ABSA Sentiment Analysis

#### ABSA 입력 구성

FR-5.1. 시스템은 NLI를 통과한 뉴스에 대해 대상 심볼을 분석 속성으로 하는 ABSA를 수행해야 한다.

FR-5.2. 시스템은 뉴스 content 필드와 symbol을 이용하여 ABSA 입력 문장을 생성해야 한다.

FR-5.3. ABSA 입력 문장은 다음 형식으로 구성되어야 한다: [CLS] content [SEP] symbol [SEP]

FR-5.4. 시스템은 ABSA 입력 생성 시 대상 심볼의 표시명과 분석용 명칭을 일관되게 정규화해야 한다.

#### 추론 수행 및 결과 출력

FR-5.5. 시스템은 ABSA 모델을 통해 Negative, Neutral, Positive 중 하나의 감성 label을 산출해야 한다.

FR-5.6. ABSA 결과는 label, score, all\_scores를 포함해야 한다.

FR-5.7. ABSA 완료 뉴스는 DONE 최종 상태로 처리해야 한다.

FR-5.8. 시스템은 ABSA 결과를 absa\_result 필드에 저장해야 한다.

#### 처리 통계 기록

FR-5.9. 시스템은 ABSA 처리 시간을 metadata.processing\_stats.absa\_duration에 기록해야 한다.

FR-5.10. 시스템은 ABSA 처리 완료 시각을 metadata.trace\_timestamps.absa\_processed\_at에 기록해야 한다.

FR-5.11. ABSA 완료 결과는 absa-out 토픽으로 발행되어야 한다. 해당 토픽에는 DONE 상태의 ABSA 완료 알림과 NLI\_DROPPED 상태의 탈락 알림 두 종류의 메시지가 함께 담긴다.

### 3.2.3.3 FR-6 Impact Score and News Ordering

FR-6.1. 시스템은 뉴스별 영향도 점수를 다음 공식으로 계산해야 한다. 최종 점수는 1.0을 상한으로 한다.

$$\text{Impact} = \min(1.0, (\text{EntailmentScore} * 0.4) + (\text{SentimentScore} * 0.6 * \text{AlignmentWeight}))$$

여기서 EntailmentScore는 NLI 모델이 산출한 ‘함의(entailment)’ 라벨에 대한 확신도를 의미하며, SentimentScore는 ABSA 모델이 산출한 해당 감성 라벨에 대한 확신도(Confidence Score)를 의미한다. AlignmentWeight는 FR-6.2에 정의된 방향성 일치 가중치이다.

FR-6.2. AlignmentWeight는 가격 변동 방향과 감성 label을 대조하여 산출한다. 클러스터 선택 시는 클러스터 대표 이상치의 변동률을, 클러스터 미선택 시는 해당 뉴스와 연결된 이상 이벤트의 변동률을 기준으로 한다.

- 감성 label이 neutral이거나 분류가 없는 경우: 0.5
- 가격 상승 및 Positive 뉴스, 또는 가격 하락 및 Negative 뉴스(방향 일치): 1.0
- 가격 상승 및 Negative 뉴스, 또는 가격 하락 및 Positive 뉴스(방향 상충): 0.1

FR-6.3. 뉴스 목록은 DONE, COLLECTED, SUMMARIZED, NLI\_PASSED 등 진행 상태, DROPPED(NLI\_DROPPED 포함) 순으로 우선 정렬하고 같은 상태 안에서는 영향도 점수 내림차순으로 정렬해야 한다.

### 3.2.3.4 FR-7 Event-Level Aggregation

FR-7.1. 시스템은 이상치 클러스터 내 점수 절대값이 가장 높은 대표 이상치의 anomaly\_id에 연결된 DONE 상태의 뉴스 분석 결과를 모아 이벤트 단위 요약을 생성해야 한다.

FR-7.2. 시스템은 NLI를 통과한 뉴스들의 ABSA 점수와 NLI entailment 점수를 가중치로 사용하여 종합 감성 점수와 감성 일관성을, 통과 뉴스 수와 NLI 및 ABSA 예측 평균 확신도를 추가 변수로 사용하여 신호 신뢰도를 산출해야 한다. (상세 산출 공식 및 변수 정의는 Appendix B 참조)

FR-7.3. 시스템은 이벤트 요약에 종합 감성 점수, 감성 label(종합 감성 점수가 양수이면 Positive, 음수이면 Negative, 0이면 Neutral), 감성 일관성, 신호 신뢰도, 영향도 점수 기준 상위 3개 뉴스로 구성된 핵심 근거 목록을 포함해야 한다.

FR-7.4. 이벤트 요약은 개별 뉴스 카드와 분리되어 대표 이상치 내역 상단에 종합 분석 지표 형태로 표시되어야 한다.

FR-7.5. 이벤트 요약은 분석 후보 뉴스가 모두 최종 상태로 완료되거나 이상 이벤트 발생 후 180초가 경과하면 생성되어야 한다.

### 3.2.4 데이터 영속성 및 실시간 서빙

시스템은 수집 및 분석된 모든 데이터를 안정적으로 저장하고, 대시보드에 실시간으로 제공할 수 있어야 한다.

#### 3.2.4.1 FR-8 Persistence and Serving

FR-8.1. 시스템은 가격 데이터, 이상 이벤트, 최종 뉴스 분석 결과를 DB에 저장해야 한다.

FR-8.2. 중간 뉴스 상태(COLLECTED, SUMMARIZED, NLI\_PASSED, NLI\_DROPPED 등)는 그 자체로 저장되지 않으며, 최종 상태인 DONE 또는 DROPPED로 확정된 경우에만 DB에 저장해야 한다.

FR-8.3. Realtime Gateway는 심볼 목록, 가격 히스토리, 이상 탐지 및 뉴스 분석 이력, 대시보드 통계 요약 등 시스템의 모든 저장된 데이터를 조회할 수 있는 인터페이스를 제공해야 한다.

FR-8.4. Realtime Gateway는 Kafka 이벤트를 Socket.IO 스트림으로 변환하여 실시간으로 브로드캐스트해야 한다.

FR-8.5. Storage Writer은 Kafka 또는 DB 일시 장애 후 서비스 재시작 시 동일 key(anomaly\_id, news\_id 등) 데이터가 중복 저장되지 않도록 방지해야 한다.

FR-8.6. NLI 결과에 의해 분석에서 제외된 뉴스는 NLI\_DROPPED 상태를 거쳐 최종 DROPPED 상태로 DB에 보존되어 어떤 뉴스가 제외되었는지 확인 가능해야 한다.

### 3.3 Performance Requirements

PR-1. Base Node 단일 구성 시 시스템은 최대 20개 활성 심볼을 지원해야 하며, 파티션 상한을 최대로 활용한 확장 구성(NLI, ABSA Processors를 Kafka 파티션 상한인 각 12개씩 구성했을 때) 시에는 기본 구성 대비 6배인 최대 120개 활성 심볼 처리를 지원해야 한다.

PR-2. 시스템은 심볼당 최대 1건/분, 기본 구성 시 전체 20건/분 이하의 가격 이벤트를 정상 수집, 발행, 저장해야 한다. (최대 확장 시, 전체 120건/분)

PR-3. 가격 이벤트 수신 후 대시보드 price:tick 표시까지의 지연은 p95 기준 500ms 이하여야 한다.

PR-4. 이상 이벤트 생성 후 대시보드 anomaly:detected 표시까지의 지연은 p95 기준 1초 이하여야 한다.

PR-5. 단일 심볼 상세 화면의 초기 REST 데이터 조회는 p95 500ms 이하여야 한다.

PR-6. 이상 이벤트 발생 후 첫 뉴스 진행 상태(COLLECTED)는 외부 뉴스 API가 정상 응답할 경우 응답 후 p95 3초 이내에 웹 대시보드에 표시되어야 한다.

PR-7. 대상 뉴스가 20건 이하이고 외부 LLM API가 정상 응답하며 시스템이 Base Node 내 최대 구성(NLI, ABSA Processors 각 2개)으로 작동할 때, 이벤트 요약 지연은 p95 180초 이내여야 한다. (확장 구성 시 지연 시간은 Processors 수에 반비례하여 단축되어야 하며, 최대 확장 시 p95 30초 이내여야 한다.)

PR-8. 브라우저 내 UI 상호작용은 INP 기준을 준수하여 (INP: Interaction to Next Paint, 사용자 입력 후 화면이 실제로 갱신될 때까지의 응답성을 측정하는 지표) p95 200ms 이내에 시각적 피드백을 제공해야 한다.

PR-9. 외부 LLM API 장애 또는 지연은 가격 수집, 이상 탐지, 기존 DB 조회, 차트 표시의 가용성을 저해하거나 분석 파이프라인의 메시지 스트림을 차단하지 않아야 한다.

PR-10. 활성 심볼당 약 98,280건(1건/분 수집 기준, 6.5시간/거래일 × 252 거래일)의 1년치 틱 데이터가 적재된 상태에서도, 특정 1시간 범위의 시계열 데이터 조회 쿼리는 p95 1초 이내에 완료되어야 한다.

PR-11. Tailscale망을 통해 연결된 외부 Processor와의 통신 왕복 지연(RTT)은 p95 500ms 이내를 유지해야 한다.

PR-12. NLI 및 ABSA Processor는 단일 인스턴스 기준 뉴스 1건당 각 모델의 추론 시간을 3초 이내로 유지해야 한다.

### 3.4 Logical Database Requirements

DB-1. stock\_prices 테이블은 (symbol, timestamp)를 기본 키로 가져야 하며, 가격, 거래량, 변동률, 그리고 원본 메시지 전체를 JSONB 형식의 payload 필드로 저장해야 한다.

DB-2. anomalies 테이블은 anomaly\_id를 기본 키로 가져야 하며, 심볼, 이벤트 시각, 탐지 시각, 가격, 이전 가격, 변동률, 알고리즘, 점수, 메시지, 그리고 알고리즘 상세 정보를 포함한 전체 메시지를 JSONB 형식의 payload 필드로 저장해야 한다.

DB-3. news\_analysis 테이블은 news\_id를 기본 키로 가져야 하며, anomaly\_id를 외래 키로 참조해야 한다. NLI 및 ABSA 분석 결과와 메타데이터는 JSONB 형식으로 저장해야 한다.

DB-4. news\_analysis.status는 DONE 또는 DROPPED만 허용해야 한다.

DB-5. DB는 심볼별 timestamp 내림차순 조회를 효율적으로 수행할 수 있도록 인덱스를 제공해야 한다.

DB-6. 시스템은 가격 데이터, 이상 이벤트, 뉴스 분석 결과 등 모든 시계열 및 분석 데이터를 최소 1년간 보존해야 한다.

### 3.5 Design Constraints

DC-1. 시스템은 Apache Kafka 기반의 Choreography 방식 이벤트 흐름을 유지해야 한다.

DC-2. 모든 마이크로서비스는 Docker 컨테이너 환경에서 독립적으로 실행 및 배포 가능해야 한다.

DC-3. 시장 데이터 API, 뉴스 API, LLM API는 어댑터 패턴을 적용하여 공급자 변경 시 핵심 로직 수정 없이 교체 가능해야 한다.

DC-4. NLI와 ABSA 모델은 각 프로세서 서비스 시작 시 1회만 로드되어야 하며, 메시지 처리마다 모델을 재로드하는 오버헤드를 방지해야 한다.

DC-5. 최종 사용자 인터페이스는 웹 대시보드로 제한하며, React 기반 웹 애플리케이션으로 메인 페이지와 심볼별 상세 페이지를 제공해야 한다.

DC-6. 시스템은 단일 사용자 운영을 가정하며, 인증/인가(Auth) 레이어를 설계에 포함하지 않는다.

DC-7. Kafka Broker 및 코어 서비스(Sentinel, Gateway, Storage Writer 등)는 Base Node에서 단일 인스턴스 운영을 기본으로 한다.

DC-8. 시스템은 독립적으로 확장 및 배포 가능한 MSA 기반 프로세서 구조를 사용해야 한다.

DC-9. NLI Processor는 nli-in 토픽을 입력으로, nli-out-absa-in 토픽을 출력으로 사용해야 한다.

DC-10. ABSA Processor는 nli-out-absa-in 토픽을 입력으로, absa-out 토픽을 출력으로 사용해야 한다.

DC-11. 시스템은 Hugging Face Transformers 라이브러리와 DeBERTa 기반 NLP 모델을 사용해야 한다.

DC-12. ABSA의 Base Model은 yangheng/deberta-v3-base-absa-v1.1을 사용해야 한다.

DC-13. 모델 파일(.safetensors) 및 토큰라이저 설정은 로컬 파일 시스템에서 로드할 수 있어야 한다.

DC-14. 실시간 추론에 사용되는 ABSA 모델은 선정된 Base Model을 금융 데이터셋으로 파인튜닝하여 구축된 최종 모델을 사용해야 한다.

DC-15. Kafka Consumer는 메시지 처리 완료 후 Offset을 커밋하는 방식을 사용하여 데이터 처리 보장을 관리해야 한다.

DC-16. 시스템은 실시간 스트리밍 처리 중에도 DB 저장 기능을 수행할 수 있어야 한다.

DC-17. 심볼 표시명과 분석용 명칭이 다른 경우, 해당 매핑은 설정 가능한 메타데이터로 관리해야 한다.

DC-18. NLI 및 ABSA 프로세서가 외부 서버에 배치될 경우 해당 서버는 Tailscale 또는 LAN 전용 네트워크를 통해 Base Node의 Kafka Broker와 통신해야 한다.

DC-19. 시스템 내 모든 마이크로서비스 간 교환되는 데이터의 타임스탬프는 ISO8601 UTC 형식을 준수하여 정규화되어야 한다.

DC-20. 확장성 및 성능 보장을 위해 Kafka 토픽의 파티션 수는 다음 상한 규격을 따른다.

- nli-in, nli-out-absa-in: 12 (상한)
- stock-prices, anomaly-event, raw-news, absa-out: 1

DC-21. 외부 API 키, DB 접속 비밀번호, Kafka 설정 정보 등은 소스 코드에 하드코딩하지 않고 환경 변수나 보안 설정 파일을 통해 주입 가능해야 한다.

DC-22. NLI 모델과 ABSA 모델은 유지보수를 위해 각각 독립된 모델 디렉터리로 관리되어야 한다.

## 3.6 Software System Attributes

### 3.6.1 Availability

QA-1. 시스템의 개별 마이크로서비스는 장애 발생 시 30초 이내에 정상 가용 상태로 복구되어야 하며, 반복 장애 주입 시 복구 성공률은 100%를 달성해야 한다.

### 3.6.2 ABSA Model Quality

QA-2. 사용되는 ABSA 모델은 금융 뉴스 도메인 테스트셋 기준 Accuracy 90% 이상을 만족해야 한다.

QA-3. 사용되는 ABSA 모델은 금융 뉴스 도메인 테스트셋 기준 Macro F1 0.85 이상을 만족해야 한다.

QA-4. 사용되는 ABSA 모델은 금융 뉴스 도메인 테스트셋 기준 Weighted F1 0.90 이상을 만족해야 한다.

QA-5. 사용되는 ABSA 모델은 파인튜닝 전 Base Model 대비 뉴스 테스트셋 기준 Accuracy가 5%p 이상 향상되어야 한다.

### 3.6.3 Usability

QA-6. (Learnability) 사전 교육이 없는 사용자가 심볼을 검색하여 상세 페이지 도달 및 차트를 확인하기까지의 소요 시간은 60초 이내여야 한다.

QA-7. (Operability) 사용자가 클러스터 마커를 클릭하여 해당 시점의 뉴스 목록을 필터링하기까지 필요한 조작 단계는 3단계 이내여야 한다.

QA-8. (User interface aesthetics) 웹 대시보드 전반에 걸쳐 동일한 상태(예: 분석 완료, 분석 제외)에 대해 서로 다른 색상이나 아이콘이 혼용되지 않고 일관되게 표시되어야 한다.

## 4 Appendices

### 4.1 Appendix A: Kafka Topic 및 데이터 스키마 상세 규격

관련 요구사항: SI-9, SI-10, CI-1, CI-3, CI-4

프로젝트의 실시간 데이터 파이프라인에서 사용되는 모든 이벤트는 **ISO8601 UTC** 시간 포맷을 준수하며, **JSON** 형식을 기반으로 한다.

#### 4.1.1 1. 추가 데이터 (stock-prices)

추가 시뮬레이터(anomaly-sentinel)가 발행하는 실시간 추가 틱 또는 과거 데이터 재생 이벤트이다.

```
{
  "symbol": "string",
  "event_time": "string (ISO8601 UTC)",
  "detected_at": "string (ISO8601 UTC)",
  "source": "string",
  "data": {
    "price": "float",
    "previous_price": "float",
    "change_percent": "float",
    "volume": "int | null"
  },
  "metadata": {
    "stream_session_id": "string (UUID)"
  },
  "message": "string"
}
```

#### 4.1.2 2. 이상 탐지 이벤트 (anomaly-event)

추가 이상 탐지 시스템이 감지한 최초 이상 이벤트를 정의하며, 뉴스 분석 파이프라인의 트리거(Trigger)로 사용된다.

```
{
  "anomaly_id": "string (UUID)",
  "symbol": "string",
  "event_time": "string (ISO8601 UTC)",
  "detected_at": "string (ISO8601 UTC)",
  "source": "string",
  "data": {
    "price": "float",
    "previous_price": "float",
    "change_percent": "float"
  },
  "analysis": {
    "algorithm": "string (Z-SCORE | EMA)",
    "score": "float",
    "details": "dict"
  },
  "metadata": {
    "stream_session_id": "string (UUID)"
  },
  "message": "string"
}
```

### 4.1.3 3. 통합 뉴스 분석 객체 (Unified Analysis Object)

raw-news, nli-in, nli-out-absa-in, absa-out 토픽에서 공유되는 통일된 스키마이다. 파이프라인을 거치며 null 필드가 점진적으로 분석 데이터로 채워진다.

```
{
  "news_id": "string (UUID5 - URL+Session 조합)",
  "anomaly_id": "string (UUID - 원천 이상치 식별자)",
  "symbol": "string",
  "title": "string",
  "content": "string (원문 스니펫 또는 AI 요약문)",
  "status": "string (COLLECTED | SUMMARIZED | NLI_PASSED | NLI_DROPPED | DONE)",
  "nli_result": {
    "label": "string | null",
    "score": "float | null",
    "all_scores": "dict | null"
  },
  "absa_result": {
    "label": "string | null",
    "score": "float | null",
    "all_scores": "dict | null"
  },
  "metadata": {
    "summary_model": "string | null",
    "timestamp": "string (ISO8601 UTC)",
    "source": "string",
    "url": "string",
    "processing_stats": {
      "news_duration": "float | null",
      "nli_duration": "float | null",
      "absa_duration": "float | null"
    },
    "trace_timestamps": {
      "collected_at": "string | null",
      "summarized_at": "string | null",
      "nli_processed_at": "string | null",
      "absa_processed_at": "string | null"
    }
  }
}
```

### 4.1.4 4. 토픽별 발행 및 진행 규칙

1. **raw-news (Phase 1/2):** 수집 직후(COLLECTED) 발행 후, LLM 요약 완료 시 content를 갱신하여 (SUMMARIZED) 재발행한다.
2. **nli-in:** SUMMARIZED 상태의 객체를 소비하여 NLI 관련성 검증을 시작한다.
3. **nli-out-absa-in:** 관련성 통과 시 NLI\_PASSED 상태로 ABSA에 전달한다. 탈락 시 NLI\_DROPPED 상태로 최종 토픽(absa-out)으로 우회(Bypass)한다.
4. **absa-out:** ABSA 분석 완료 시 DONE 상태로 발행한다. NLI에서 우회된 NLI\_DROPPED 메시지도 이 토픽에 함께 담기며, 대시보드와 DB는 이 토픽을 구독하여 최종 상태를 처리한다.

### 4.1.5 5. 이벤트 요약 결과 (Event Summary Result)

부록 B의 공식에 따라 산출된 특정 이상치 클러스터의 종합 분석 결과 스키마이다.

```
{
  "anomaly_id": "string (UUID - 대표 이상치 식별자)",
  "symbol": "string",
  "summary_metrics": {
    "sentiment_score": "float (S)",
    "consistency_score": "float (C)",
    "data_density": "float (D)",
    "model_confidence": "float (M)",
    "reliability_score": "float (R)"
  },
  "overall_sentiment": "string (Positive | Negative | Neutral)",
  "key_evidence_count": "int",
  "timestamp": "string (ISO8601 UTC)"
}
```

## 4.2 Appendix B: Event Summary Calculation

관련 요구사항: FR-7.2

각 뉴스의 감성 점수  $s_i$ 는 PositiveScore - NegativeScore로 계산한다. 각 뉴스의 가중치  $w_i$ 는 NLI entailment score로 둔다. 본 공식은 FR-7.1에 따라 DONE 상태로 판정된 뉴스만을 대상으로 수행한다.

$$S = \frac{\sum(w_i * s_i)}{\sum(w_i)}$$

$$C = \frac{\text{abs}(S)}{\text{mean}(\text{abs}(s_i))} \text{ (단, 분모가 0인 경우 0으로 처리)}$$

$$D = \min(N_{\text{passed}} / 3, 1.0)$$

$$M = (\text{NormalizedMeanNliConfidence} + \text{NormalizedMeanAbsaConfidence}) / 2$$

$$R = 0.2 * D + 0.3 * M + 0.5 * C$$

S는 종합 감성 점수, C는 감성 일관성, D는 데이터 충분성, M은 모델 확신도, R은 신호 신뢰도이다.

변수 상세:

- $N_{\text{passed}}$ : NLI를 통과하여 DONE 상태로 분석 완료된 뉴스의 개수.
- NormalizedMeanNliConfidence: 분석 완료된 뉴스들의 NLI entailment 점수 평균.
- NormalizedMeanAbsaConfidence: 분석 완료된 뉴스들의 ABSA 신뢰도(Confidence) 점수 평균.

종합 감성 label은 S를 기준으로 다음과 같이 파생한다.

- $S > 0$ : Positive
- $S < 0$ : Negative
- $S = 0$ : Neutral

핵심 근거 뉴스 목록은 해당 이벤트에 연결된 DONE 상태 뉴스 중 영향도 점수 기준 상위 항목으로 구성된다.

## 4.3 Appendix C: Impact Score Weight Matrix

관련 요구사항: FR-6.1, FR-6.2

Condition	Alignment Weight	Description

Direction Match	1.0	가격 상승과 Positive 뉴스 또는 가격 하락과 Negative 뉴스가 일치하는 경우
Neutral News	0.5	감성 분석 결과가 Neutral이거나 방향 판단이 불확실한 경우
Direction Mismatch	0.1	가격 상승과 Negative 뉴스 또는 가격 하락과 Positive 뉴스가 상충하는 경우