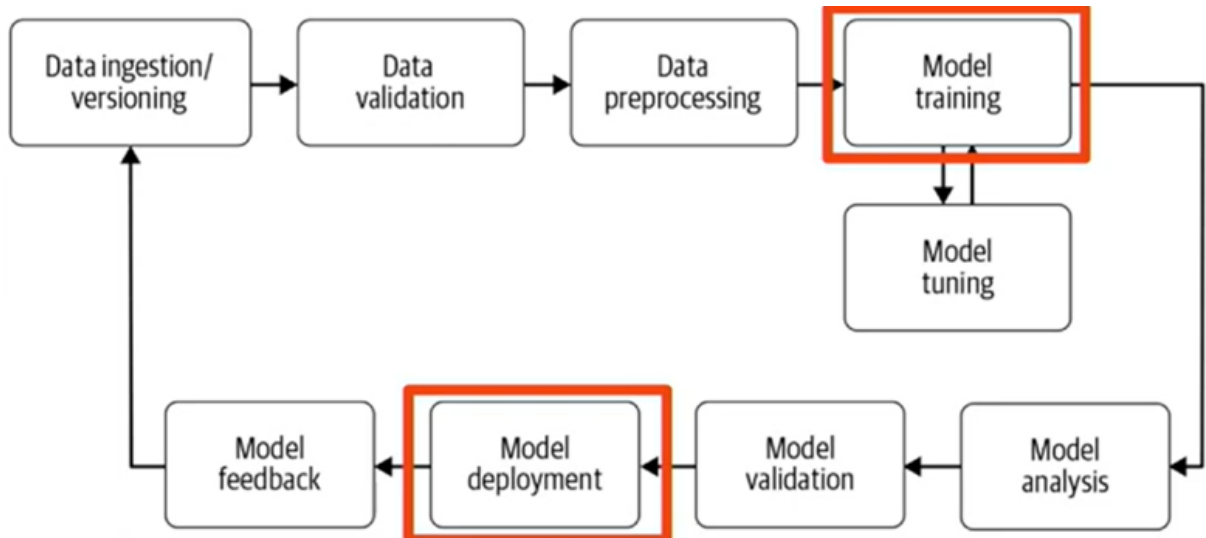


# 참고 논리 아키

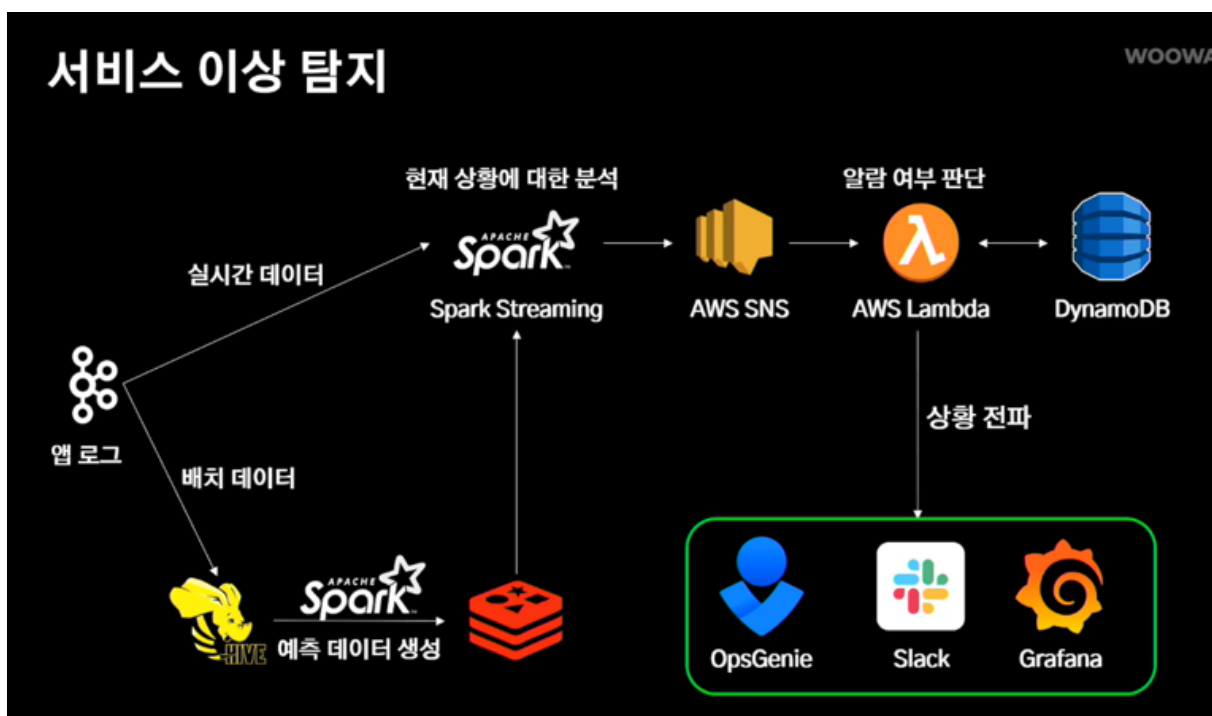
데이터 아키 고려사항  
데이터 생성속도  
데이터 보존기간  
데이터 중복여부  
Null 값 포함 여부  
스키마 변경 가능성  
  
데이터 누락 가능성  
읽기 쓰기 빈도  
전체 데이터 크기  
건당 데이터 크기데이터 변환 방식  
  
시스템 읽기 쓰기 속도  
프로세스의 병목  
저장시스템의 특징  
메타 데이터 관리  
작업 스케줄링  
  
배치 스트리밍  
실시간 정의  
데이터 변경 가능성  
  
데이터의 독립성  
개인정보 포함 여부  
  
장애 발생 시 전파 여부  
장애 발생 시 복구 방법  
  
데이터의 비즈니스적 가치  
고객 경험  
매출증진  
기타 목표 도달  
  
개발자 스킬  
개발 인원  
개발기간  
기존 시스템 설계  
  
운영효율성

업무영역	개발 Tool 사용 지원	
	사용 Tool	사용의 편리성
데이터 수집	§Apache NiFi §AWS S3 §AWS Athena §Amazon Kafka	§Apache NiFi : 유연한 확장성을 이용한 대규모 실시간 데이터 처리로 다양한 데이터 소스 형태와 대상에서 데이터를 수집/전송하며, 강력한 모니터링을 통한 직관적인 데이터 흐름 파악 용이 §AWS S3 : 손쉽게 데이터를 저장, 관리 및 백업할 수 있도록 직관적 인터페이스와 뛰어난 확장성을 제공 §AWS Athena : 서버리스 환경에서 SQL 쿼리를 통해 S3 데이터를 손쉽게 분석 §Amazon Kafka : 완전 관리형 Kafka 클러스터를 통해 손쉽게 대규모 실시간 데이터를 스트리밍하여 수집
데이터 통합	§AWS Glue §AWS Glue Crawler §AWS Glue Data Catalog §AWS Glue DQ §Managed Workflows for Apache Airflow §AWS Transfer Family §AWS Aurora RDS	§AWS Glue : 완전관리형 ETL 서비스로 다양한 데이터 소스를 쉽게 연결하고 통합 §AWS Glue Crawler : 다양한 데이터 소스를 자동으로 스캔하여 스키마, 메타데이터를 손쉽게 추출/업데이트 §AWS Glue Data Catalog : 다양한 데이터 소스의 메타데이터를 중앙 관리하고 검색할 수 있도록 지원하여 편리한 통합 데이터 관리 §AWS Glue DQ : 다양한 데이터 소스의 품질을 자동으로 검사하고 관리하여, 사용자가 신뢰할 수 있는 데이터를 통합 §Apache Airflow : 관리가 용이한 서버리스 아키텍처로, 강력한 데이터 파이프라인/워크플로우 구축과 편리한 DAG 모니터링, Amazon CloudWatch와 연계한 워크플로우 모니터링 및 로그 추적을 제공 §AWS Transfer Family : FTP, SFTP, FTPS를 통해 손쉽게 데이터를 전송하고 AWS 스토리지 서비스와의 통합을 지원하여 데이터 마이그레이션 및 통합 작업을 간소화 §Amazon Aurora PostgreSQL : 강력한 보안을 탑재하고 고성능의 데이터 처리와 적재, 관리를 통해 신속하고 안전한 데이터 통합을 제공하여 비즈니스의 요구를 효율적으로 충족
분석환경	§AICentro	§AICentro DL Studio : 딥러닝기반의 AI 서비스를 쉽고, 빠르게 구축, 배포, 운영 지원하는 KT DS의 §End-to-End AI 서비스플랫폼

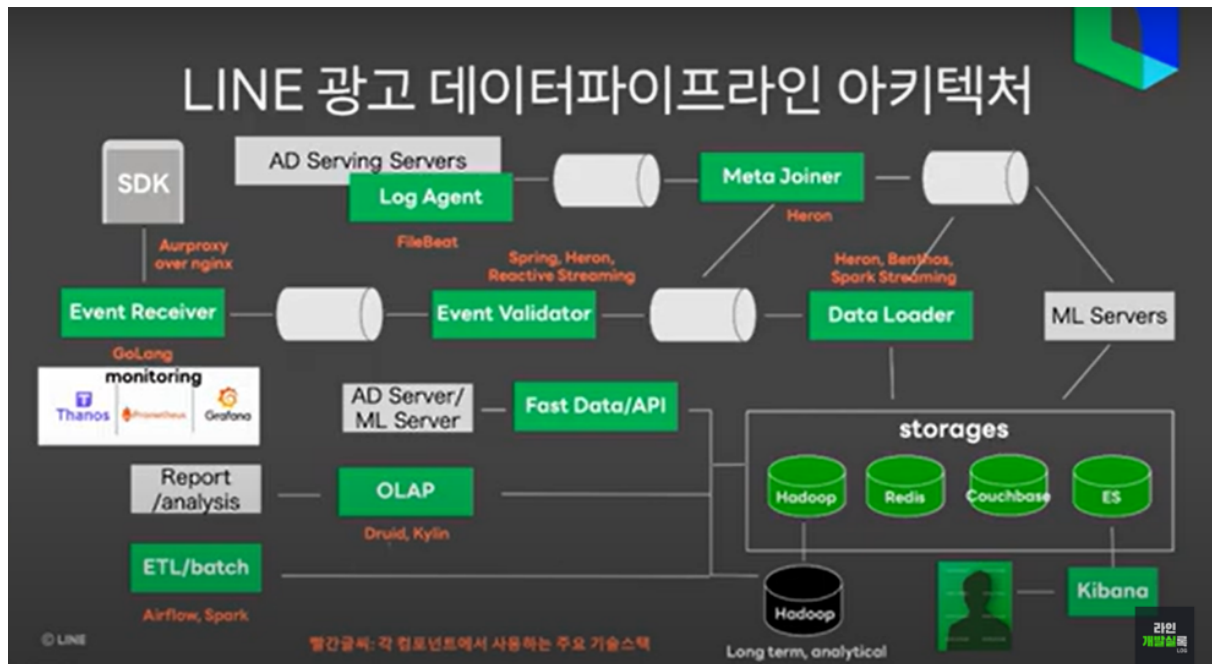
BI	§Tableau	§Tableau : 강력한 시각화 도구로, 데이터를 쉽게 분석하고 시각화하여 인사이트 도출
포털	§DataWorks	§Data Works : 데이터 포털 솔루션으로, 다양한 데이터 자산을 통합하고 사용자에게 데이터 접근과 분석 기능 제공



이상탐지



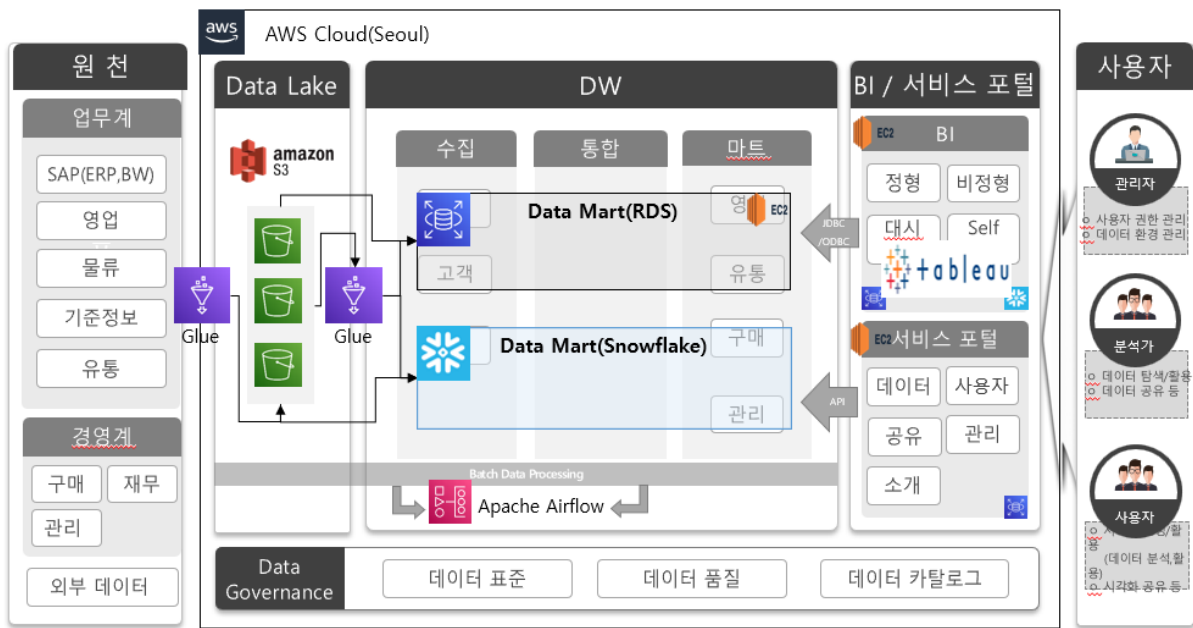
## 대용량 데이터파이프라인 아키텍처



### 1. 구축 사례

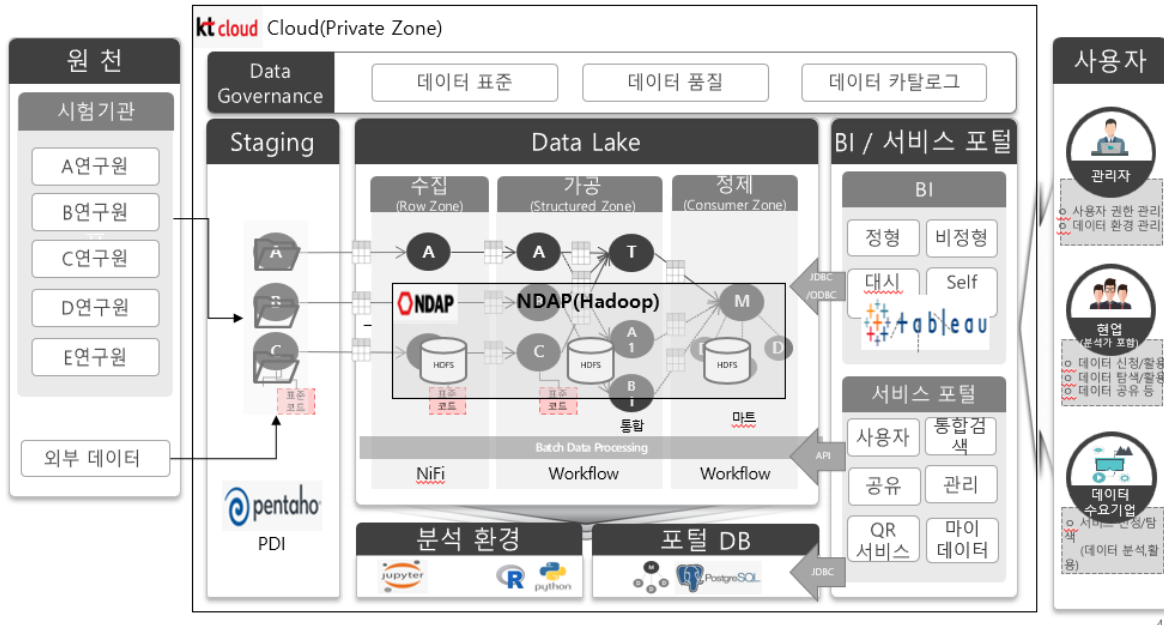
#### 1.1 시스템 구축 사례 - A사

AWS 기반의 원천 데이터 수집을 통해 분석계 데이터를 Data Lake에 저장하고, 해당 데이터에 대해서 데이터 마트 구성을 통해 제공



## 1.2 시스템 구축 사례 – B사

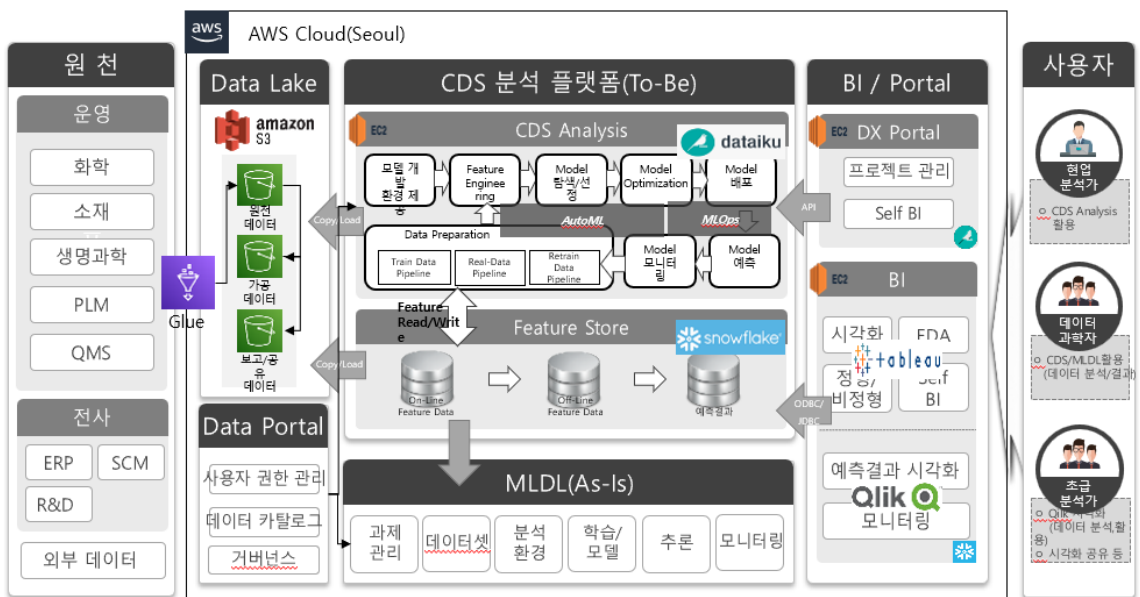
시험 참여 기관 데이터 수집을 통해 Data Lake 및 분석 환경, 포털 DB를 통한 빅데이터 플랫폼 구축



4

## 1.3 시스템 구축 사례 – C사

GUI 기반의 AutoML, MLOps가 적용된 DX수행 을 위한 사용 친화적인 CDS 분석 플랫폼 구성을 통한 데이터 분석 확대

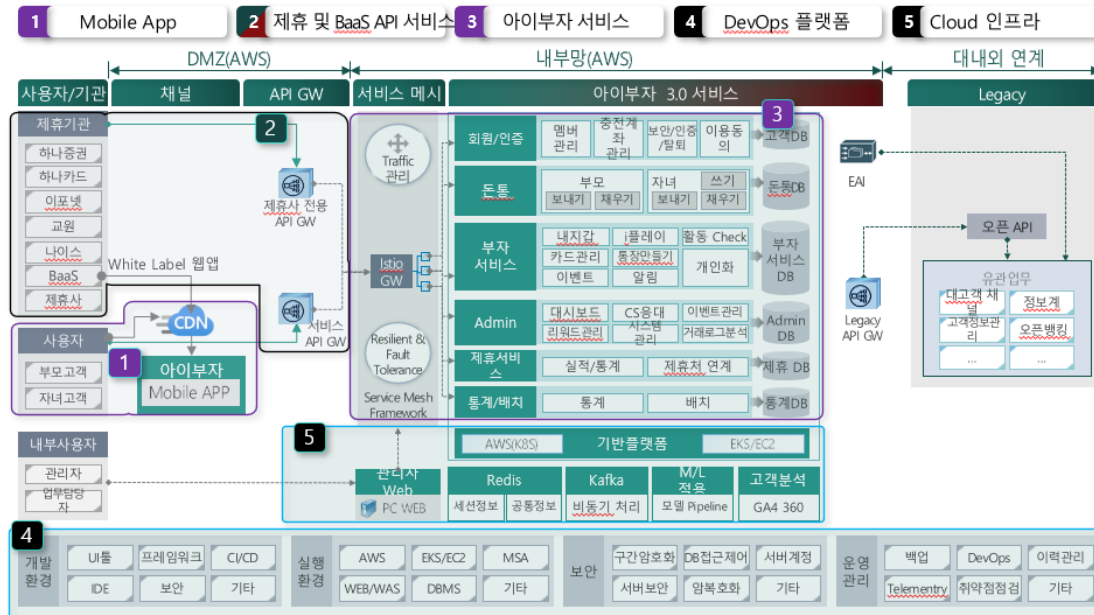


3

## 서비스메시

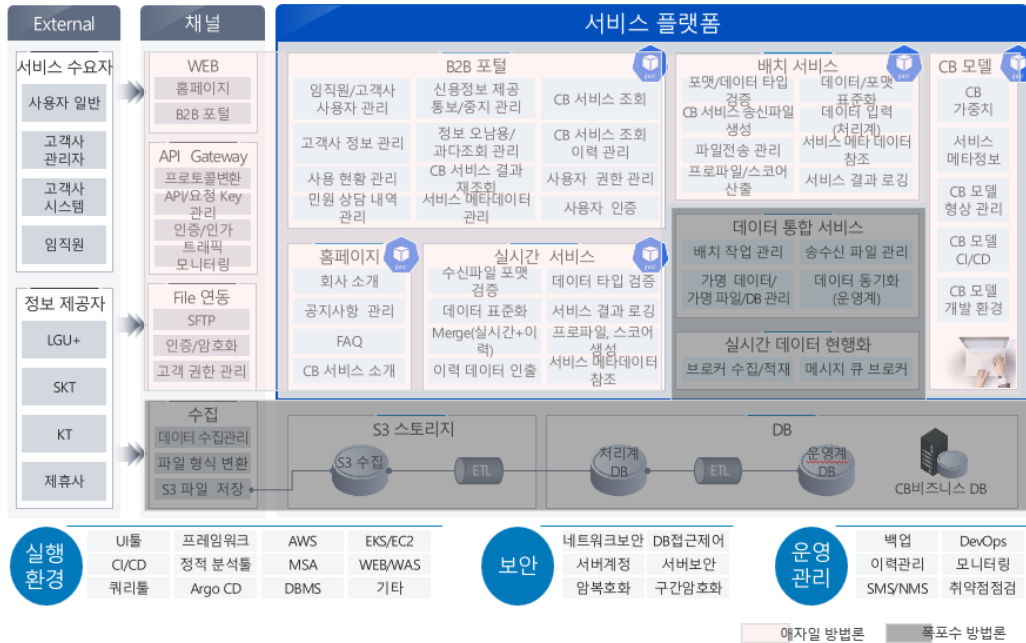
### 1.4 시스템 구축 사례 – D사

마이크로서비스 아키텍처 기반 금융권 구축 프로젝트



## 1.5 시스템 구축 사례 – E사(1/2)

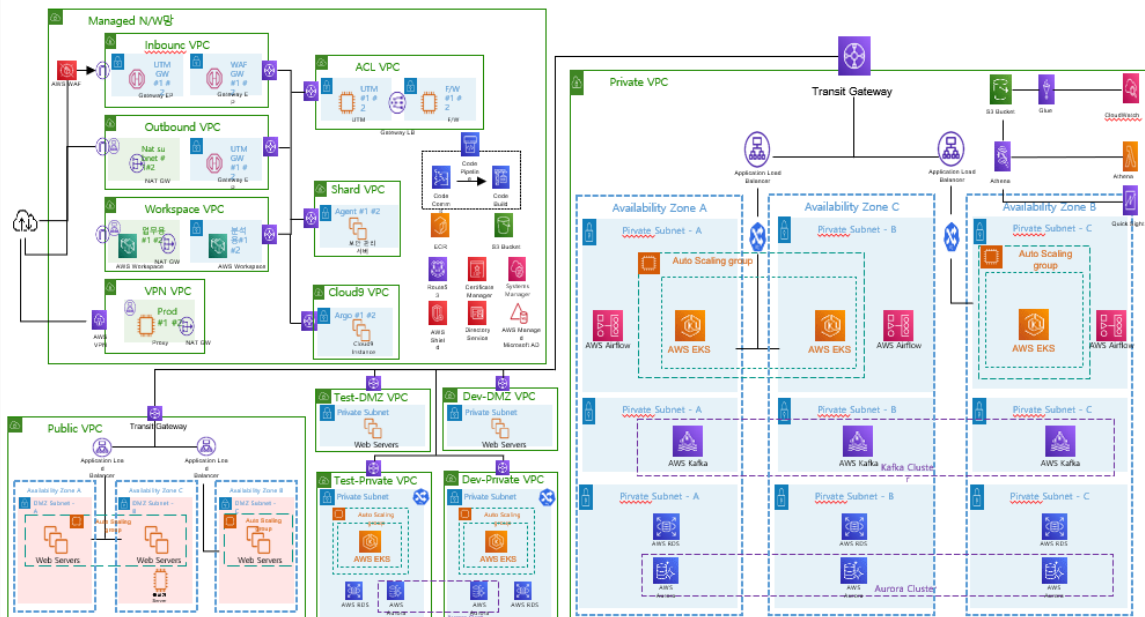
Hybrid 방법론(폭포수 + 애자일) 기반 프로젝트 수행



5

## 1.5 시스템 구축 사례 – E사(2/2)

AWS 기반 플랫폼 구축



6

## 1.6 시스템 구축 사례 – H사(1/3)

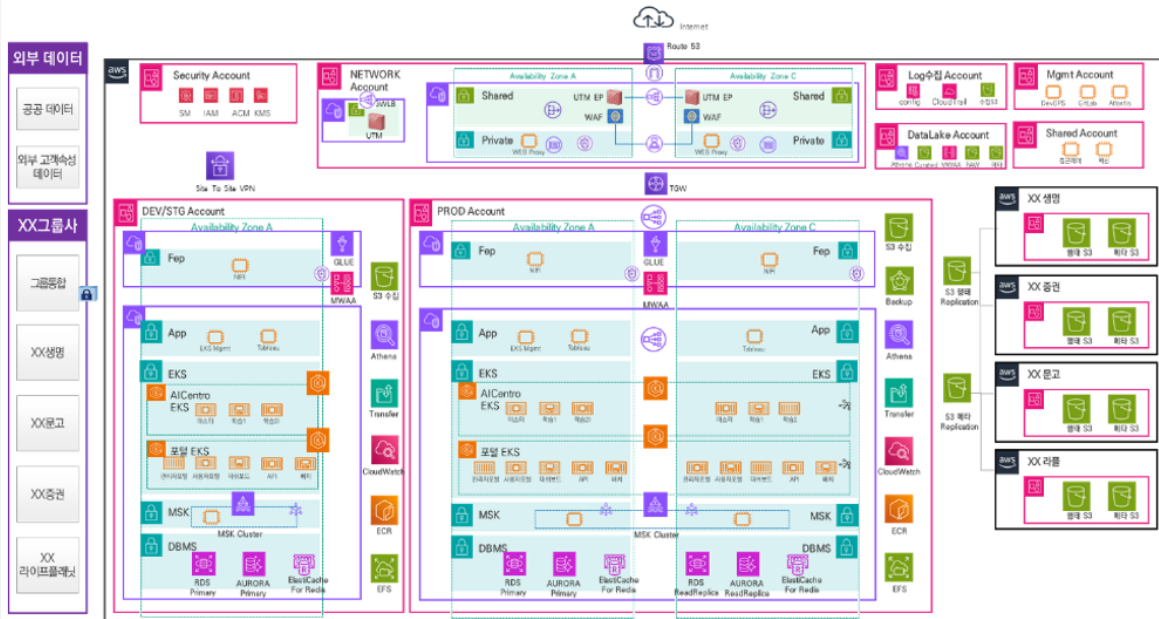
그룹 데이터 통합 데이터 서비스 포털을 구축을 위한 AWS 환경에 구성





## 1.6 시스템 구축 사례 – H사(3/3)

그룹 데이터 통합 데이터 서비스 포털을 구축을 위한 AWS 환경에 구성



8

## 1.7 시스템 구축 사례 – DataLake/MLOps

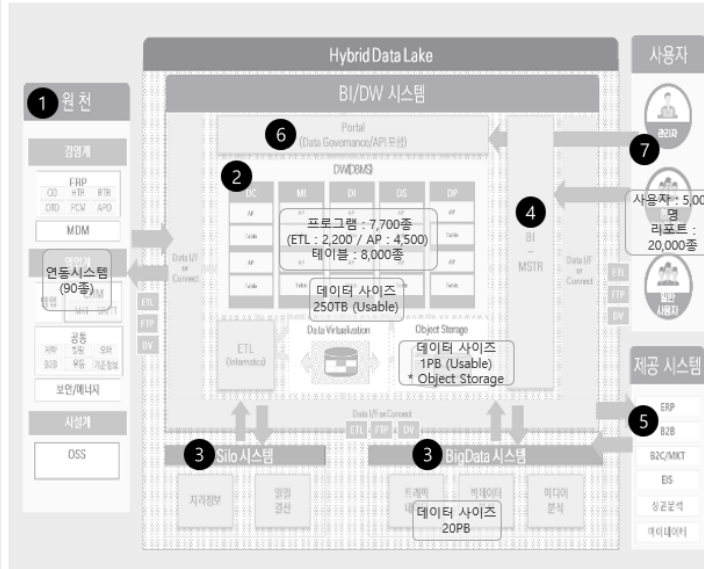
최신 AI모델 개발을 위한 AI분석/빅데이터 플랫폼 및 인프라 구축을 통한 분석환경 재구축



10

## 1.8 데이터 활용 사례 – X사(1/2)

영업, 시설, 경영계 데이터를 수집을 통해 DW 및 빅데이터 시스템 구성을 통해 다양한 분야에 데이터 활용

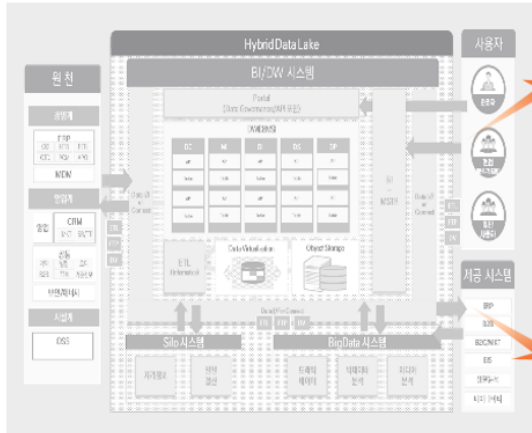


## BI/DW 구성 현황

- 1 원천 데이터 수집(경영, 영업, 시설)
  - ☑ 필요 데이터 기간계 시스템 데이터 수집
- 2 데이터 표준화 및 통합, 마트 구성(DW)
  - ☑ 데이터 표준화 및 도메인, 레이아웃 구성 관리
  - ☑ 데이터 보관주기 및 이력 관리(Object Storage)
- 3 데이터 용/복합 분석을 위한 데이터 인터페이스
  - ☑ 빅데이터, 지리정보, 경영정보(결산) 등
- 4 데이터 시각화를 위한 리포트 구성 및 제공
  - ☑ 비정형, 정형, 대시보드, Self BI, 스케줄링
- 5 데이터 활용을 위한 로우 및 통계 데이터 제공
  - ☑ 경영, 고객, 마케팅, 상권분석, 마이데이터
- 6 포털 서비스 제공(데이터 거버넌스 및 API 포함)
  - ☑ 메타 정보 및 BI 활용, 권한 관리, 챗봇 등 적용
- 7 관리자 및 사용자 그룹별 시스템 접근
  - ☑ 사용자 그룹별, 사용자단위 시스템 활용

## 1.8 데이터 활용 사례 – X사(2/2)

영업, 시설, 경영계 데이터를 수집을 통해 DW 및 빅데이터 시스템 구성을 통해 다양한 분야에 데이터 활용



**데이터 활용(내부)** 내부 활용(조직, 상품 단위 가입자 및 매출 실적..등)

- ☑ 영업 관리를 위한 B2C, B2B CRM을 위한 데이터 제공(VOC, 캠페인..등)
- ☑ 전자 손익 관리를 위한 원가, 매출, 이익 등 통계 데이터 제공
- ☑ AI 모형 및 모델 반영을 통한 고객 해지방어, 미납 채권 등 데이터 활용
- ☑ 고객 계약 및 시설 정보를 통합 마트 구성을 통한 마케팅 데이터 활용

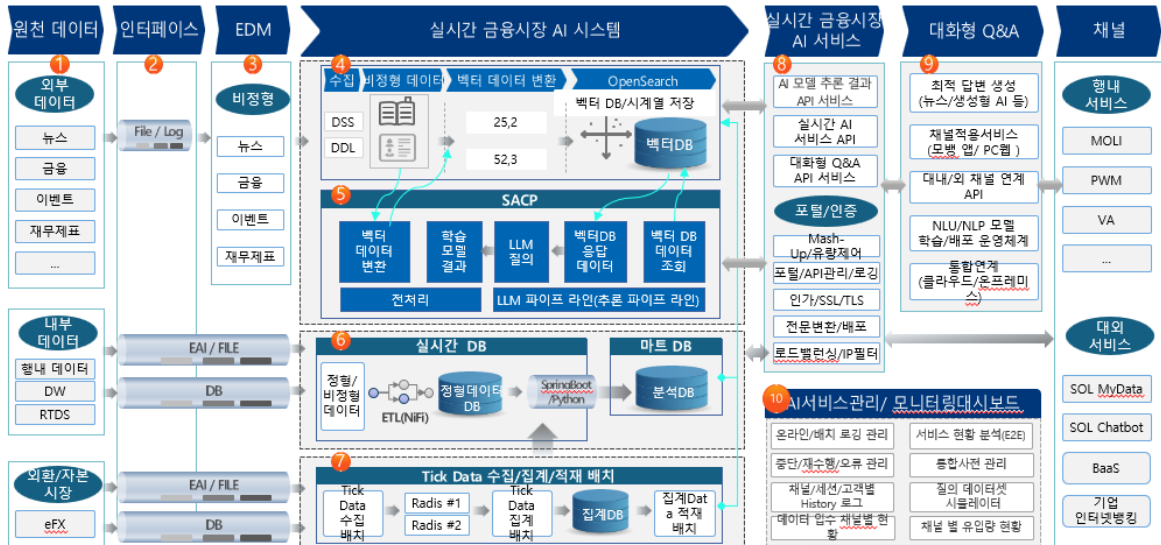
**데이터 활용(외부)**

- ☑ 타사 데이터를 활용한 타사 단위 정보 활용을 통한 상권 분석 정보 제공
- ☑ 기업간 데이터 제휴를 통한 마이 데이터 서비스 확대
- ☑ 공공 데이터 및 보유 고객 데이터를 활용한 유동인구 분석 정보 제공
- ☑ 보유 고객 정보를 활용한 타겟 마케팅 서비스 제공
- ☑ 데이터 융/복합 분석을 위한 빅데이터 키 중심 융합 데이터 제공 (성별, 연령, 날짜, 업종, 지역, 위치..등)

12

## 2.1 제안 사례 – A사 (1/6)

On-Premise 및 Public Cloud기반 AI 플랫폼 구축 구성도



- ① 원천데이터 수집 대상 – 내/외부 데이터 및 외환 eFX Tick 데이터
- ② 인터페이스 – 비정형/정형으로 구분, File, EAI, DB 등 인터페이스를 분류
- ③ EDM – 외부데이터 수집 시스템, 행내 제공 시스템
- ④ 비정형 분석마트 구성 – DSS/DDI 등 비정형 데이터를 벡터분석DB에 시계열 저장
- ⑤ SACP – 전처리, 벡터 변환/조화, 학습모델 결과 및 LLM 파이프 라인 제공

- ⑥ 정형데이터 분석마트 구성 – 실시간 정형 데이터를 집계 및 통합 후 분석마트 DB 구성
- ⑦ Tick Data 집계DB – Tick Data를 수집하여 집계DB를 생성 후 분석DB와 벡터DB에 적재
- ⑧ 금융 실시간 API 서비스 – API G/W(BEAST) 서버를 통해 실시간 금융시장 AI 서비스 제공
- ⑨ 대화형 Q&A – OpenSearch 및 API G/W를 통해 최적 답변을 생성하여 채널에 답변 피드백
- ⑩ 서비스관리/모니터링 – E2E, 질의 데이터 셋 시뮬레이터 등 관리자 포털 제공

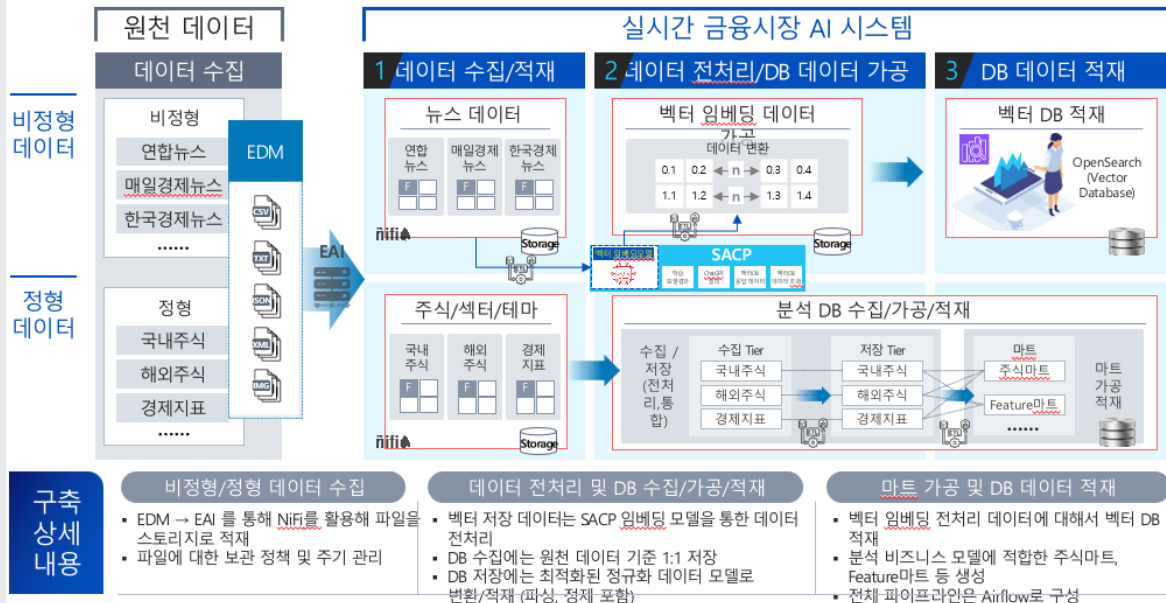
## 2.1 제안 사례 – A사 (2/6)

실시간 AI 시스템에서 OpenSearch와 대화형Q&amp;A는 AWS에서 서비스되며, DB는 On-Premise 서비스



## 2.1 제안 사례 – A사 (3/6)

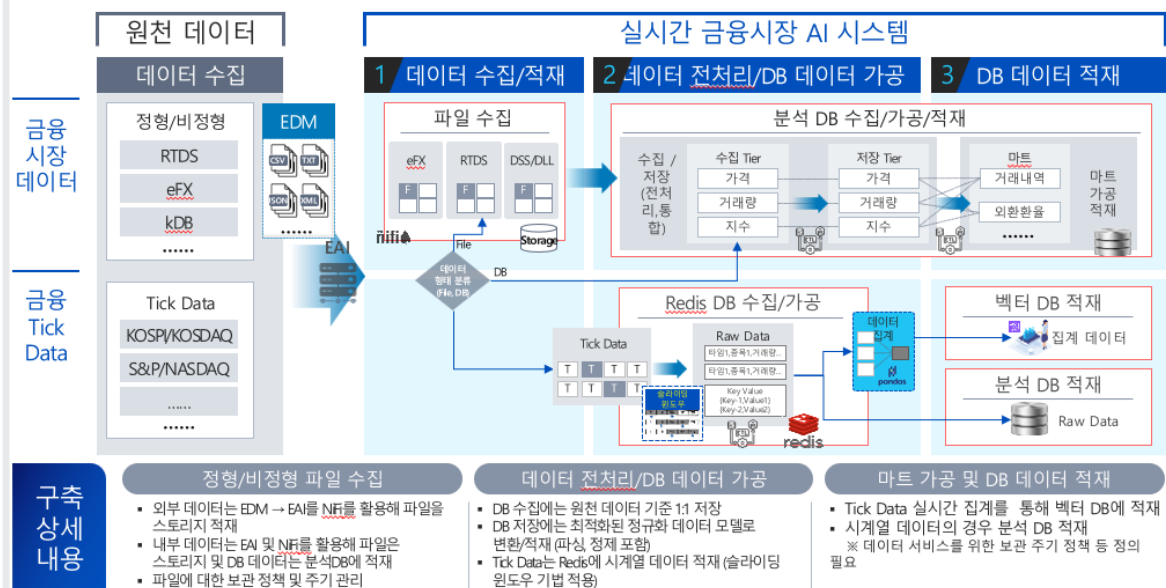
외부 비정형/정형 데이터를 파일 수집, 전처리 하여 분석, 벡터 DB 저장 구성



15

## 2.1 제안 사례 – A사 (4/6)

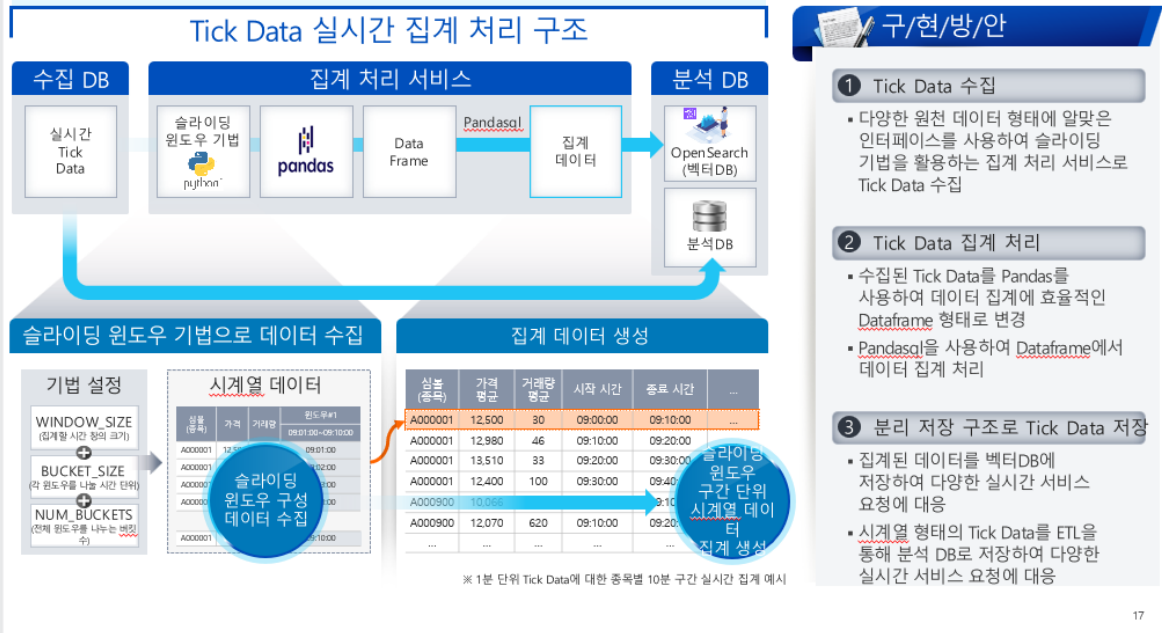
대내/외 데이터 수집/적재는 Tick Data를 고려한 분석, 벡터 DB 저장 구성



16

## 2.1 제안 사례 – A사 (5/6)

실시간 Tick Data의 시계열/집계데이터 분리 저장구조로 데이터 활용 가치 강화



17

## 2.1 제안 사례 – A사 (6/6)

정확도 높은 검색 결과를 최단시간에 LLM에 제공

## 구현 방안

## 1 사용자 검색 요청

- 사용자의 자연어 질문을 입력 받아 WAS로 전달

## 2 사용자 검색 요청

- 검색 품질 향상을 위해 검색 대상 데이터의 선별 및 언어 모델 추론에 적합한 형태로 전처리

## 3 사용자 검색 요청

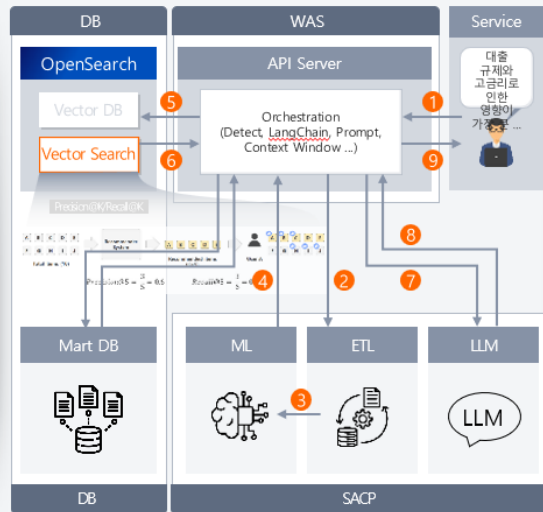
- 추론을 수행할 데이터의 정규화 및 리랭킹 후 언어 모델로 전달

## 4 사용자 검색 요청

- 정제된 데이터의 벡터 임베딩
- 임베딩 결과를 WAS에 전달

## 5 사용자 검색 요청

- 이전 생성된 답변을 재활용하기 위해 질문이 동일하거나 유사한지 확인하고 캐싱된 데이터가 아닌 경우 벡터DB 적재된 데이터를 대상으로 문서 검색을 수행



## 구현 방안

## 6 검색 결과 전달

- 키워드 검색 및 의미 검색 결과를 정렬해서 WAS로 전달
- 검색 성능 향상을 위한 튜닝
  - Boosting
  - Text Analyze
  - Pagination
  - Re-ranking 등

## 7 LLM에 답변 생성 요청

- Context window에 맞춰 준비된 Prompt에 검색 결과를 담아 LLM에 자연어 답변 생성을 요청

## 8 자연어 답변 전달

- LLM이 생성한 자연어 답변을 WAS로 전달

## 9 검색 결과 반환

- 사용자에게 자연어 답변을 제공

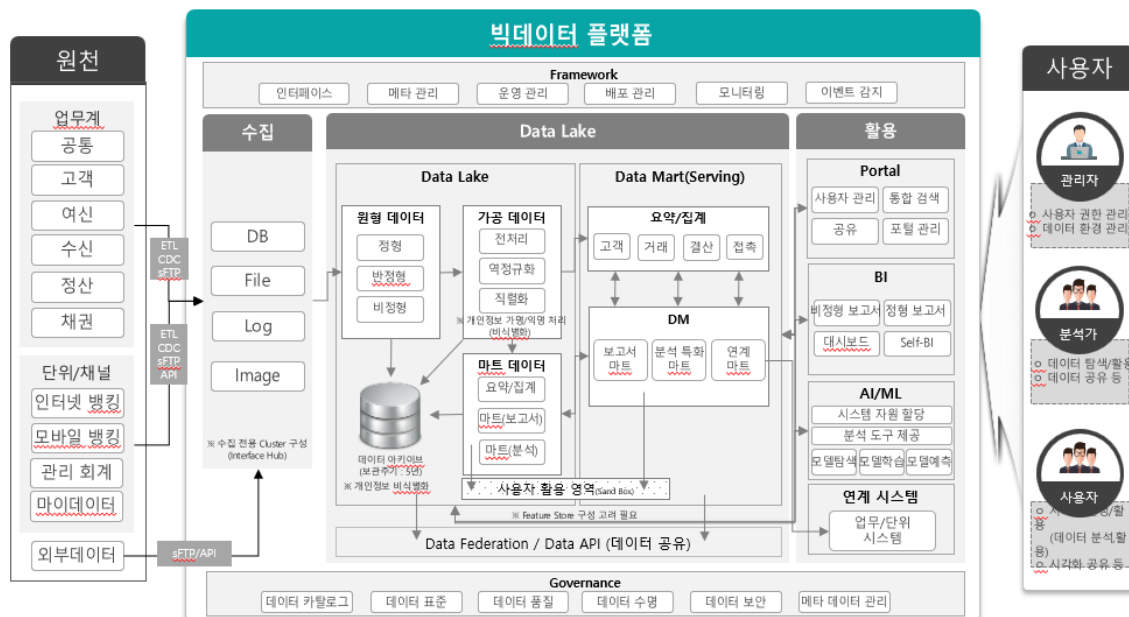
18

18

## 3. 목표 시스템 아키텍처(안)

## 3.1 목표 시스템 아키텍처 – Data Lake

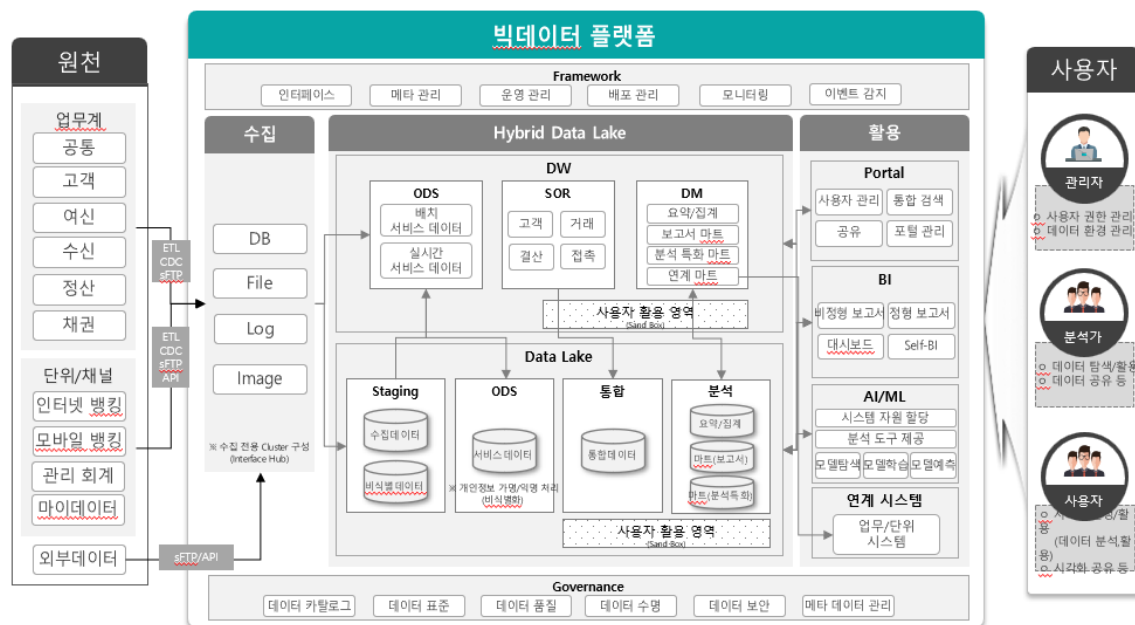
통합 데이터 플랫폼 구성을 위한 데이터 수집, 변환, 적재, 활용에 대해서 Data Lake 기반 저장소 구축



19

### 3.2 목표 시스템 아키텍처 – Hybrid Data Lake

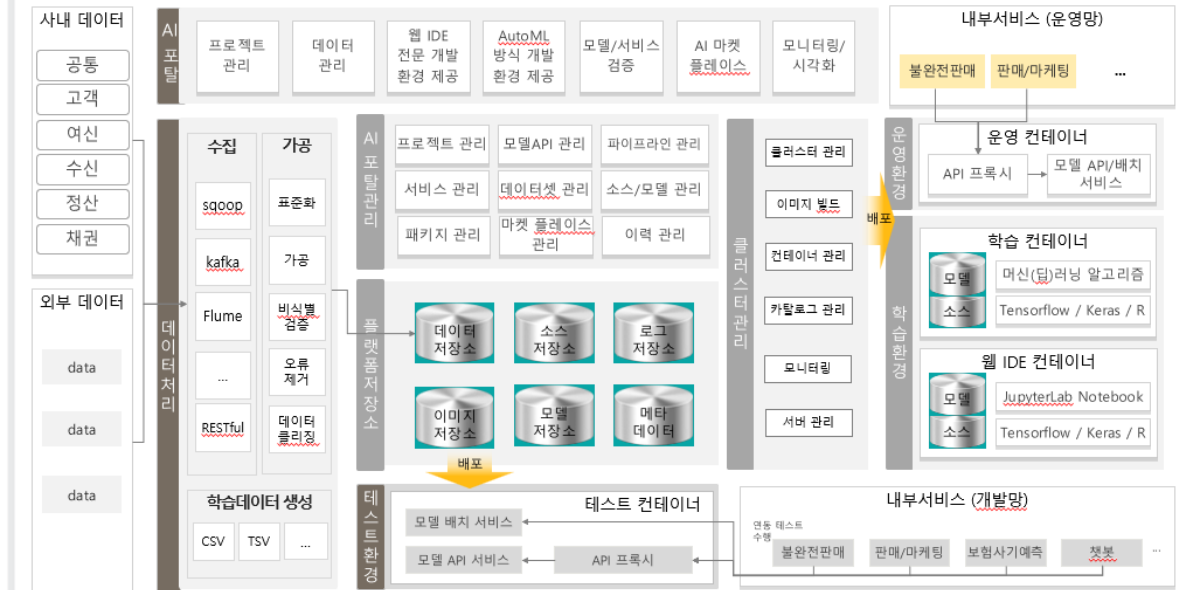
통합 데이터 플랫폼 구성을 위한 데이터 수집, 변환, 적재, 활용에 대해서 DW 및 Data Lake에 대해서 구성





## 3.3 목표 시스템 아키텍처 – MLOps

AI플랫폼 기반의 MLOps 구축으로 AI서비스를 쉽고, 빠르게 구축, 배포/운용을 지원하는 AI End-to-End AI통합 플랫폼 구축



바쁘신데 자료들 공유해 주셔서 감사합니다.

DevOps(MLOps포함) 가 공통주제가 될수 있으니 이부분도 같이 점검 한번해주세요.

이상탐지는 제가 리포트 이상 할 때 하던게 있어서 보니까 뭐 이런 자료는 많을 듯...

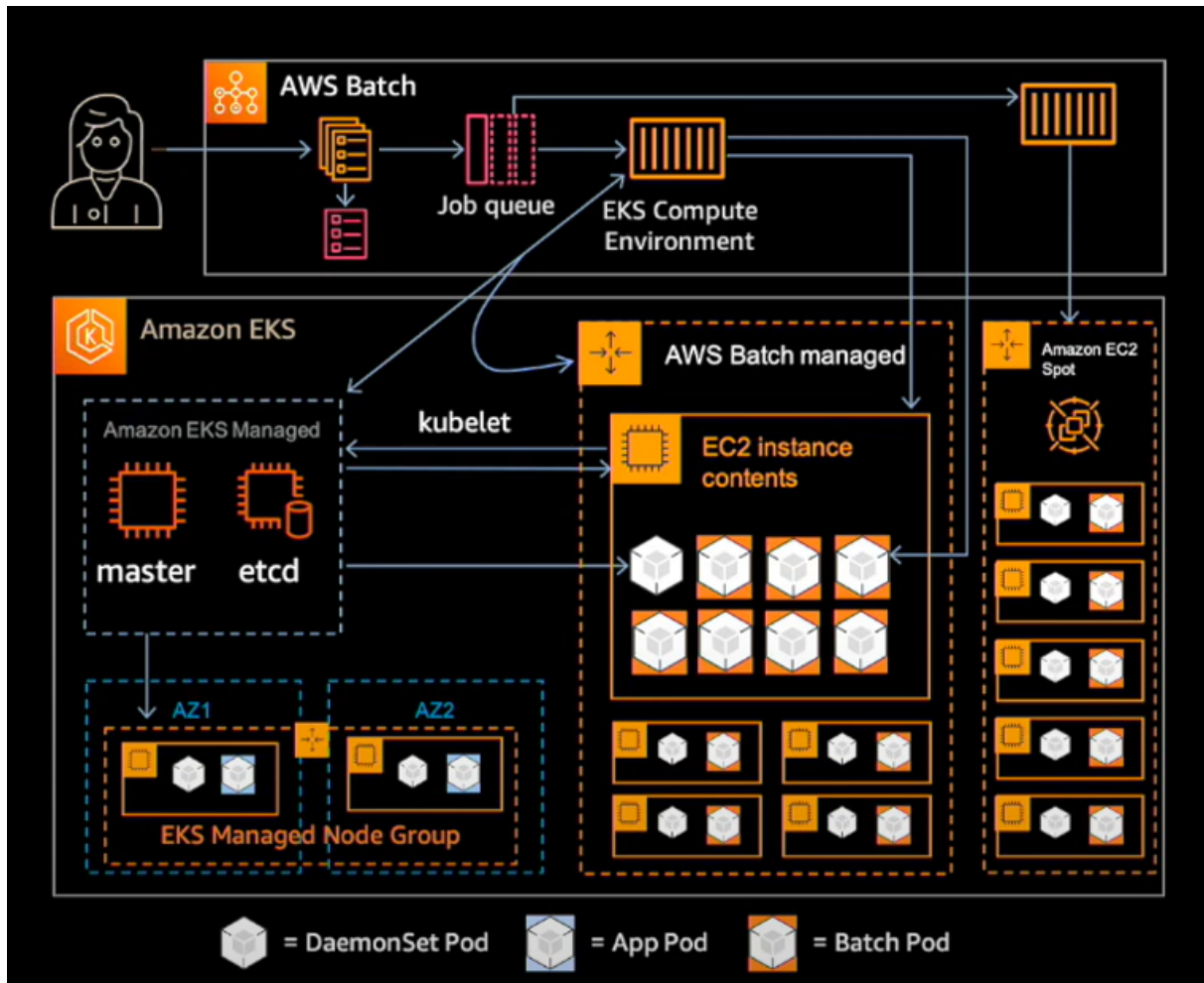
그리고 데이터브릭스 AL/ML 보니까 이상탐지도 있더라고요.

이상 감지 서비스 개발 방안

## 온/오프라인 연계 실시간성 CDP(Customer Data Platform)

온/오프라인 마케팅 관련 사례 조사

- 다른 기업들은 어떤 데이터를 사용해서 어떻게 초개인화 마케팅을 하고 있는지?
- 우리도 이런 데이터가 있으면 좋겠다 or 이런 분석을 하면 좋겠다



# 데이터 분리

WOOW

