

빅데이터 분석기획 시 고려사항

정소영 투이컨설팅 이사
(jsy@2e.co.kr)

다시 생각하기

- 빅데이터 분석 기획이란?

빅데이터 분석을 기획한다는 것은...

- 무엇 때문에(Why)

: 전략 -> 비즈니스 모델

- 무엇을 해야 하고(What)

: 분석기회(과제)

- 그 무엇을 어떻게 할 것이고(How)

: 데이터, 모델링, 인프라

- 하고 나면 무엇이 좋아진다.(Business Outcom)

: 비즈니스 성과

빅데이터 분석 기획 CSF*

분석은 성과로 이어져야 한다.

분석은 분석기회를 찾는 것이 중요하다.

분석은 전체 최적화가 중요하다.

분석은 업무에 내재화 되어야 한다.

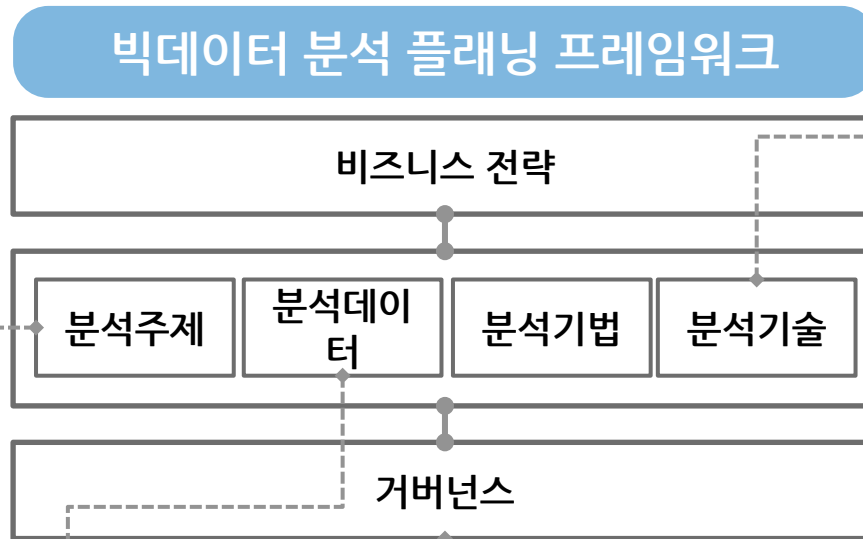
분석은 질문이 중요하다.

*CSF : Critical Success Factor

빅데이터 기획(Planning)의 4가지 핵심 고려요소

✓ 무엇을 분석할 것인가?

✓ 아키텍처 구성 시 고려요소는?



✓ 데이터 준비는 어떻게 해야하나?

✓ 빅데이터 관점의 거버넌스 특성요소는?

Q1.무엇을 분석할 것인가?

빅데이터 분석 플래닝 프레임워크



빅데이터 분석? 무엇이 다른가?

분석 주제
분석 주기
분석 기법
분석 데이터
분석 질문
분석 단위

기존 분석	빅데이터 분석	
	정보시스템이 하던 일 : 기존 분석 업그레이드, 사람이 하던 일 : 신규 빅데이터 분석 주제	
	<ul style="list-style-type: none"> 빅데이터 분석은 <p>기존의 정보시스템이 하던 분석을 보다 더 가치있게, 사람의 경험으로만 하던 일(프로세스)을 데이터 분석을 통해 지능화 하는 것임</p> 분석주제를 잘 찾아내서 정확한 질문으로 구체화 하는 것이 빅데이터 분석의 시작이고 핵심임 	
	군집 단위 분석 (Big)	개인 단위 분석 (Small) <ul style="list-style-type: none"> 고객 개인별 선호상품 예측 고객 개인별 이탈 가능성 예측 고객 개인별 선호업종/선호점포 예측 고객 개인별 웹/모바일 이동동선/패턴 분석

타사(금융사)에서는 어떤 분석을 하고 있나?



마케팅

[상품추천]

- 고객별 선호상품 예측
- 상품별 고객의 구매가능성 예측
- 고객별 상품구매 주기 예측
- 고객성향별 PF 추천 모델

[서비스 오퍼링]

- 고객별 업종 선호 예측
- 업종간 연관도 분석
- 고객별 가맹점 선호 예측
- 가맹점 간 연관도 분석
- 실시간 마케팅 시나리오 최적화
- 위치기반 상품/서비스 추천
- 고객별 선호 캠페인 유형 분석



고객

[고객 이탈예측]

- 유지계약 실효 위험도 예측
- 실효 원인 분석

[고객 라이프타임 가치(LTV) 예측]

[고객 발굴 분석]

- 신규 가망고객 발굴 분석
- VIP 잠재고객 예측
- 대면채널 전환 가망고객 발굴 분석

[비대면 채널 고객행동 분석]

- 웹/모바일 로그 기반 고객행동 분석
- 웹/모바일 로그 기반 채널전환 분석
- Kiosk 시스템 로그 분석
- CMS 거래적요 분석

[고객 다이내믹/마이크로 세그멘테이션]

타사(금융사)에서는 어떤 분석을 하고 있나?

[설계사 이탈예측]

- 설계사 이탈 위험도 예측
- 설계사 간 네트워크 관계 이탈 가능성 분석



직원

[설계사 성향 분석(코칭, 커리어패스)]

[고능률 설계사 성향분석 기반 리크루팅 유망고객 발굴]

[설계사 활동패턴 기반 매출 예측]

[설계사 다이나믹/마이크로 세그멘테이션]

[내부자 정보유출 탐지 분석]

[신계약 유지율 위험도 예측]

[불완전 판매 위험도 예측]

[진단적부 예측]

[이상거래 탐지 분석]



계약/거래

- 보험사기 탐지
- 실시간 카드 부정 사용 탐지
- 카드 이중 결제 탐지
- 은행 이상거래 탐지

[신용 리스크 분석]

- 머신러닝 기반 신용등급 평가 분석
- 심리 기반 신용등급 평가 분석(중금리 대출)
- 기술금융 평가 분석



타사(금융사)에서는 어떤 분석을 하고 있나?

[VOC 분석]



서비스 관리

- 녹취 STT 분석
- 상담메모 분석
- 이메일 분석
- 민원발생 위험도 예측
- 상담 자동 판별 분석

[고객-상담원 매칭 분석]

[챗봇]

[상품가격 최적화]

[운전습관 기반 보험료 산정 분석]

[상품수익성 예측]

[주가예측 분석]

- 소셜 감성어 사전 기반 주가예측
- 시장흐름 Regime Switch 분석



자산 운용

[로보어드바이저]

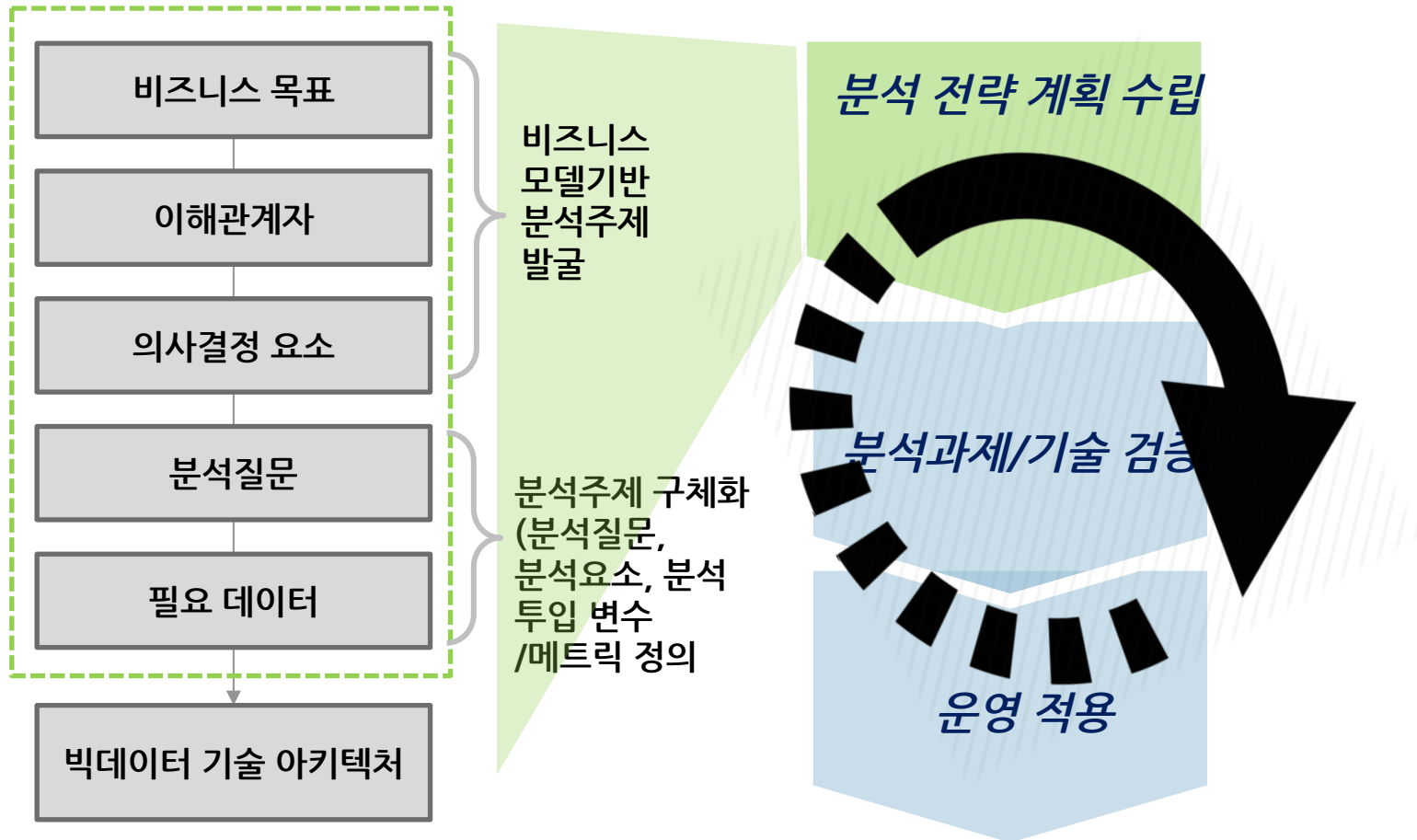
[자금세탁 방지 분석]



상품 개발

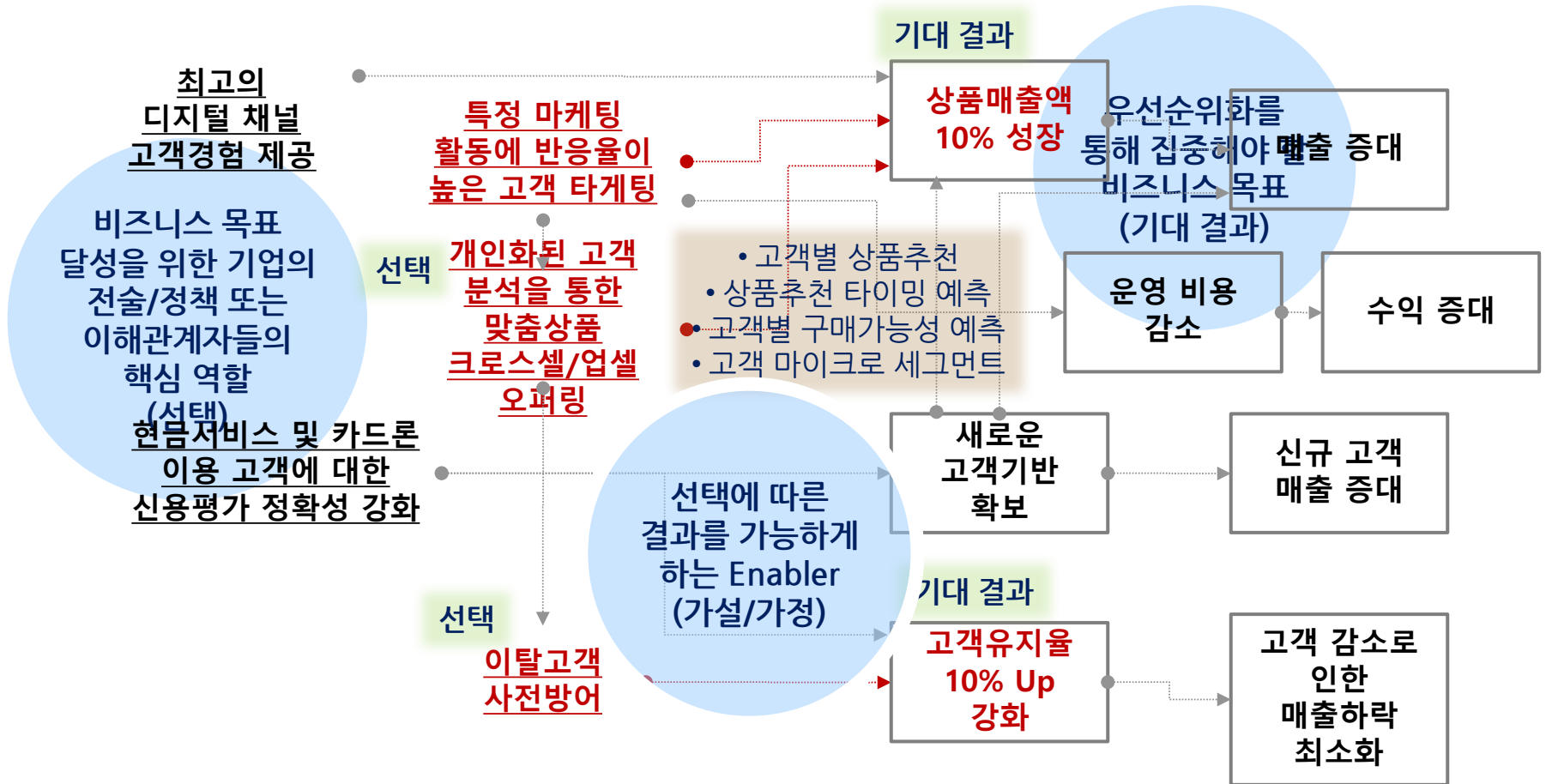
분석주제 발굴 및 분석 적용 절차

- 분석기회는 전사 비즈니스 모델 기반으로 도출 할 수 있음
- 질문이 좋아야 답변이 좋음. 분석기회는 좋은 질문을 잘 정의하는 것으로부터 시작됨



비즈니스 모델 기반 분석주제 발굴이란?

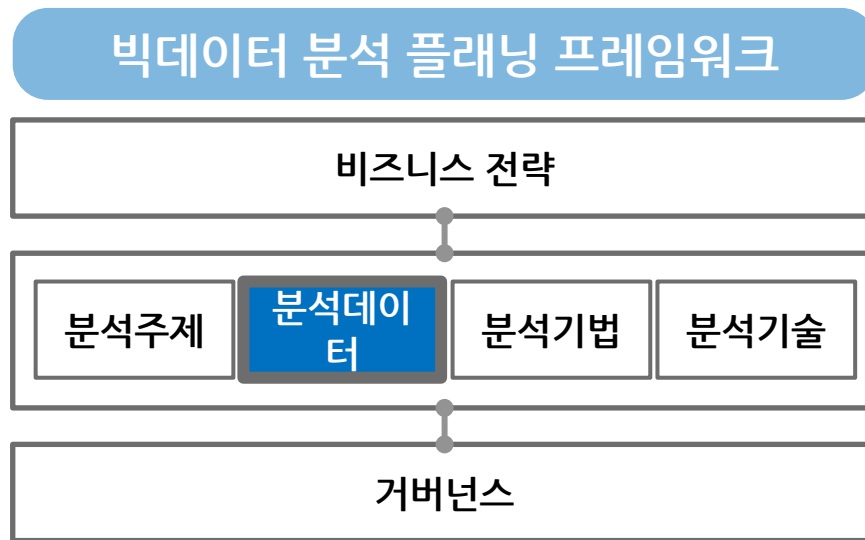
- 비즈니스 모델 기반으로 분석주제를 발굴한다는 것은 기업의 목표와 정책/전술/역할을 기반으로 비즈니스 모델을 정의하고, 해당 비즈니스 모델을 작동시킬 수 있는 핵심분석주제를 발굴하는 것임



정리 : 좋은 분석주제를 도출하는 방법은?

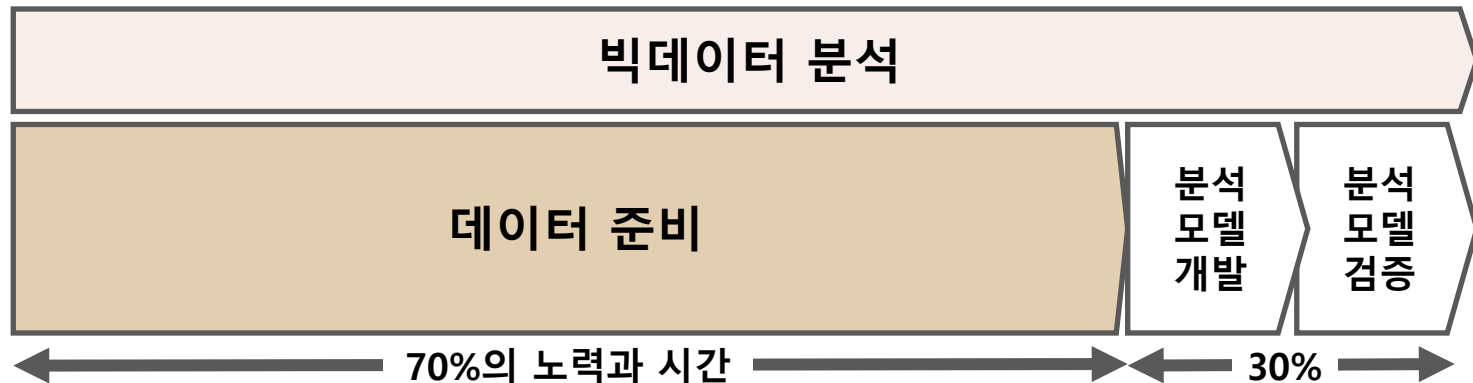
1. 비즈니스 목적 달성을 위해 필요한 **핵심 의사결정 요소**를 정의하고, 이를 위해 데이터 분석을 통해 **예측할 요소(핵심분석 기회)**를 정의하고, **분석질문을 구체화**한다.
2. 현 상황을 설명하는 질문에서부터 **예측과 처방을 위한 질문까지 도출**한다.
3. 비즈니스의 성과향상을 위해 보다 나은 예측결과를 만들어낼 수 있는 **다양한 데이터 소스(변수)와 지표를 고민하고 발굴**한다.
4. **분석이 어디서 어떻게 적용되어야** 고객행동에 좋은 경험을 제공하고, 사용자에게 보다 강력한 의사결정 지원을 할 수 있는지 **정의**한다.

Q2.데이터 준비는 어떻게 해야하나?



빅데이터 분석에서 데이터 준비의 중요성은?

- 빅데이터 분석 프로젝트 사전에, 분석에 필요한 내외부의 정형/비정형 데이터를 조사하여 데이터를 수집하고 정비해 두는 노력이 필요함



- 어떤 데이터가 필요한가?
- 필요한 데이터소스는 어디인가?
- 존재한다면 사용할 수 있는 품질인가?
- 존재하지 않는다면 대체 데이터는 있는가?
- 대체 데이터가 없다면 어떻게 확보할 것인가?
- 그리고, EDA(데이터탐색)와 데이터 전처리

어떤 데이터를 이용? 내부 데이터 예시

정형데이터

Data Warehouse



CRM Mart



기타 유관시스템



비정형데이터

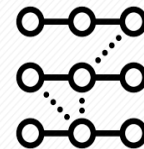


음성녹취
(STT : Speech to text)

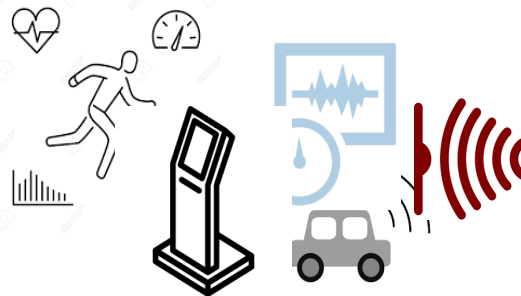
상담메모 텍스트



VOC 텍스트
이메일 텍스트
보고서 텍스트



고객행동 로그(웹/모바일 페이지 로그)
내부 사용자 로그(접근로그, 암호해제 로그,
출력로그, ...)



시스템 로그
(모바일/스마트워치 신체 로그,
Kiosk, 자동차 OBD,
시스템/생산설비 로그,
비콘 로그, ...)

어떤 데이터를 이용? 내부 데이터 예시

[기존분석]

예)이탈 분석

- 연체 → 실효 : 대부분 이탈
- 직납 → 이탈 ↑
- 직업 → 특정 직업군이 이탈 ↑
- 지역 (충북, 울산 : 이탈 ↓)

분석을 하지 않아도 짐작되는 당연한 결과!

이탈예측 모델로 충분한가 ?



내부 정형
데이터 분석



[콜센터] 해약환급금을 문의...



[홈페이지] 해약환급금 조회

→ 그 후 몇 개월 후 해약

비정형 데이터와
Advanced
Analytics(머신러닝,
딥러닝)
기법을 통해
**고객의 향후 행동
예측 가능**

고객의 “Real Voice”를 반영한 예측 모형 개발

비정형 데이터 분석 시스템 구축

정형/비정형 통합 분석체계 확립

예측 모델 개발

STT
(Speech to Text)

- 콜센터 녹취 파일의 Text화

Text
Analysis

- STT Data, 상담메모, 홈페이지 Q&A 등 비정형 데이터를 분석 가능한 정보로 자산화

로그데이터

- 고객의 웹로그 정보의 자산화

비정형 데이터
(Text
Analysis
Mart)

정형 데이터
(RDB)

후보 변수군 도출

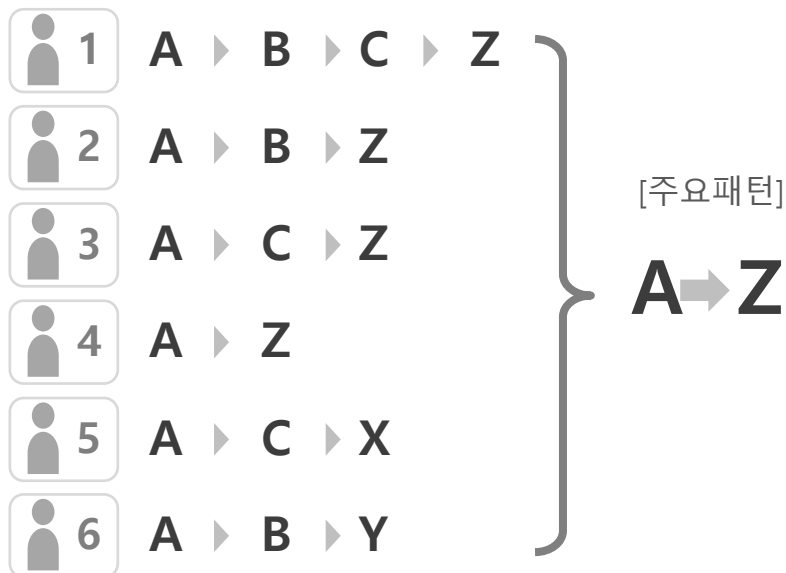
•고객의 Real Voice를
반영한
이탈예측 모형

어떤 데이터를 이용? 내부 데이터 예시

Sequential Pattern Mining

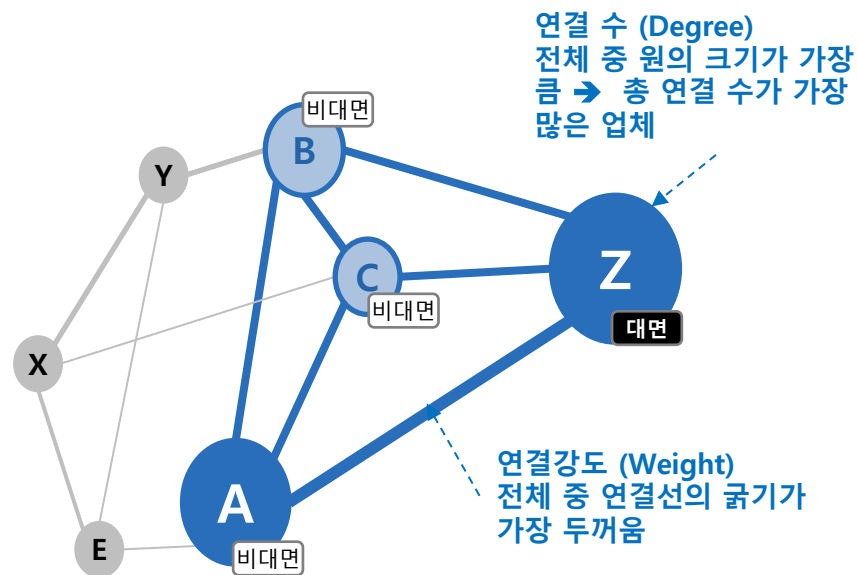
비대면에서 대면으로 채널 전환되는
시간 순서에 의한 주요 거래 패턴
선정

[고객 동선]



Garph Analytics

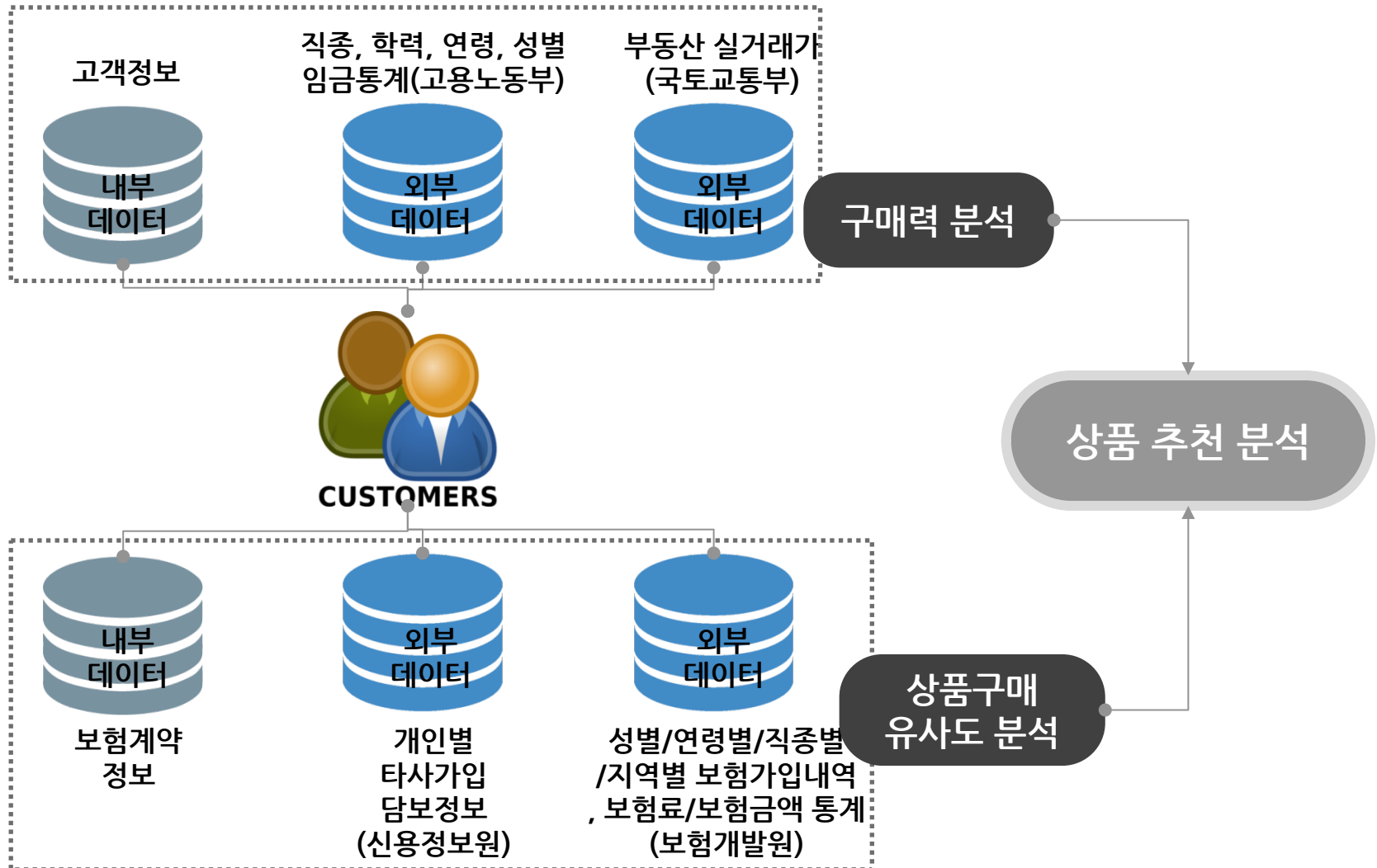
주요 거래 패턴에서 연결되는
거래들의
상세 관계 파악 (Degree, Weigth)



어떤 데이터를 이용? 외부 데이터 예시

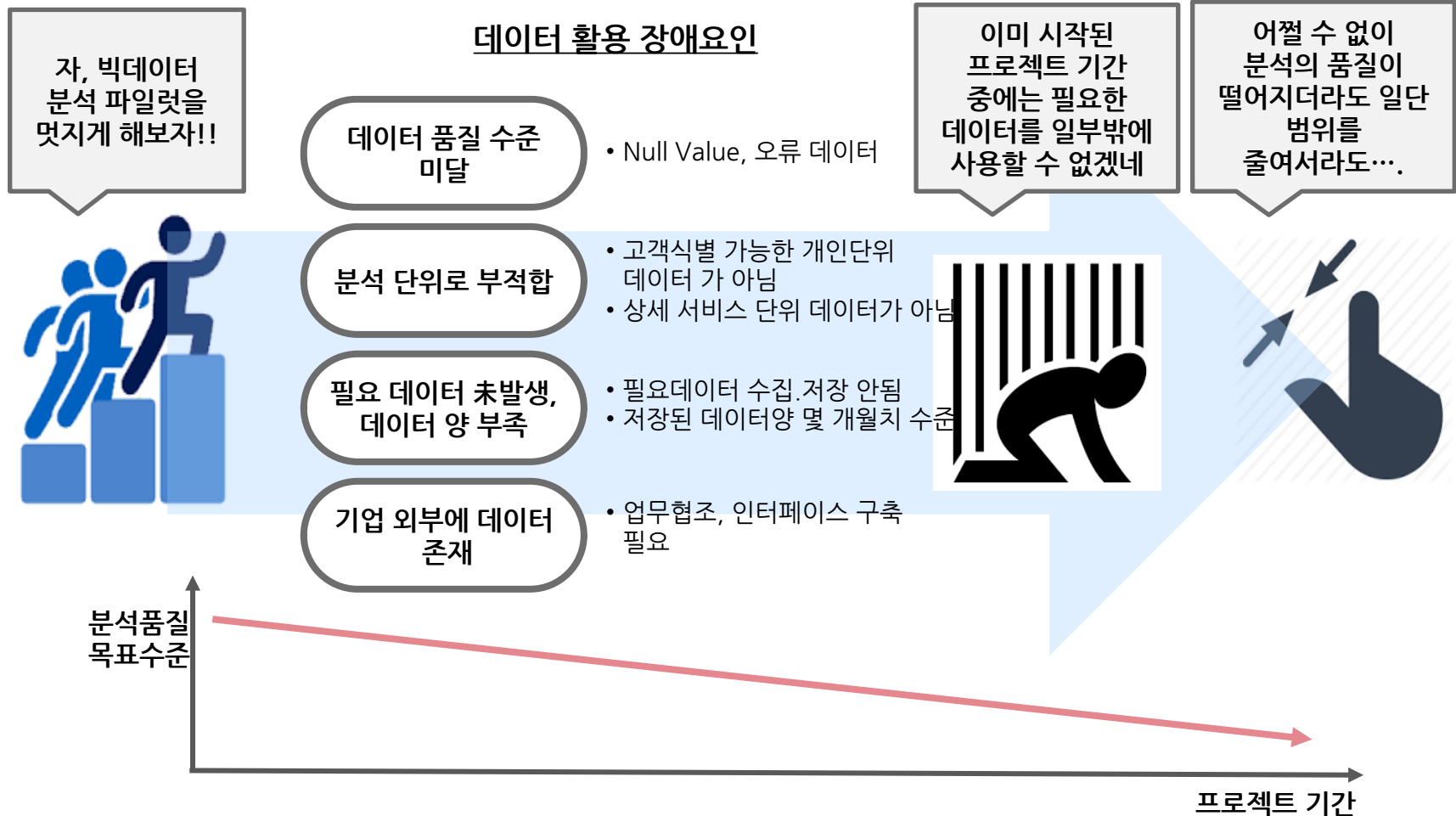
보유기관	제공정보	분석활용 예	보유기관	제공정보	분석활용 예
생명보험협회 (→한국신용정보원)	개인별 타사가입 담보정보	개인별 상품추천	NICE 신용평가정보	기업신용평가 정보	기업리스크예측
보험개발원	성별/연령별/직종별/지역별(시군구) 보험가입내역/보험료/보험금액 통계	군집별 상품보유/계약특성 파악 군집별 선호 상품발굴	한국은행	한국은행 총재 금리정책/기조 발표문 , 통화신용정책보고서, 금융안정보고서 , 경제전망보고서, 기업경영분석 통계	금융리스크예측
고용노동부	임금통계(직종, 학력/연령/성별)	경제력/구매력 추정	한국거래소	주가정보	금융리스크예측 주가예측
국토교통부	부동산 실거래가	경제력/구매력 추정	건강보험공단 (건강보험심사평가원)	진료/처방/건강검진내역, 질병통계, 보건의료빅데이터	적부심사 상품개발분석
통신사	위치정보, 결제정보, 상점정보, 유동인구정보, 상권분석정보, TV/Mobile/Navi 이용행태 정보	고객행동패턴, 고객성향 분석	로이터	세계 금융시장의 실시간 뉴스 및 마켓데이터, 경제동향 및 시장분석, 리스크 및 컴플라이언스 정보	금융시장예측 주가예측
...	
...	...		소셜네트워크	VOC	평판분석 고객행동패턴분석

어떤 데이터를 이용? 외부 데이터 예시



데이터준비에 대한 사전고려가 없을 경우...

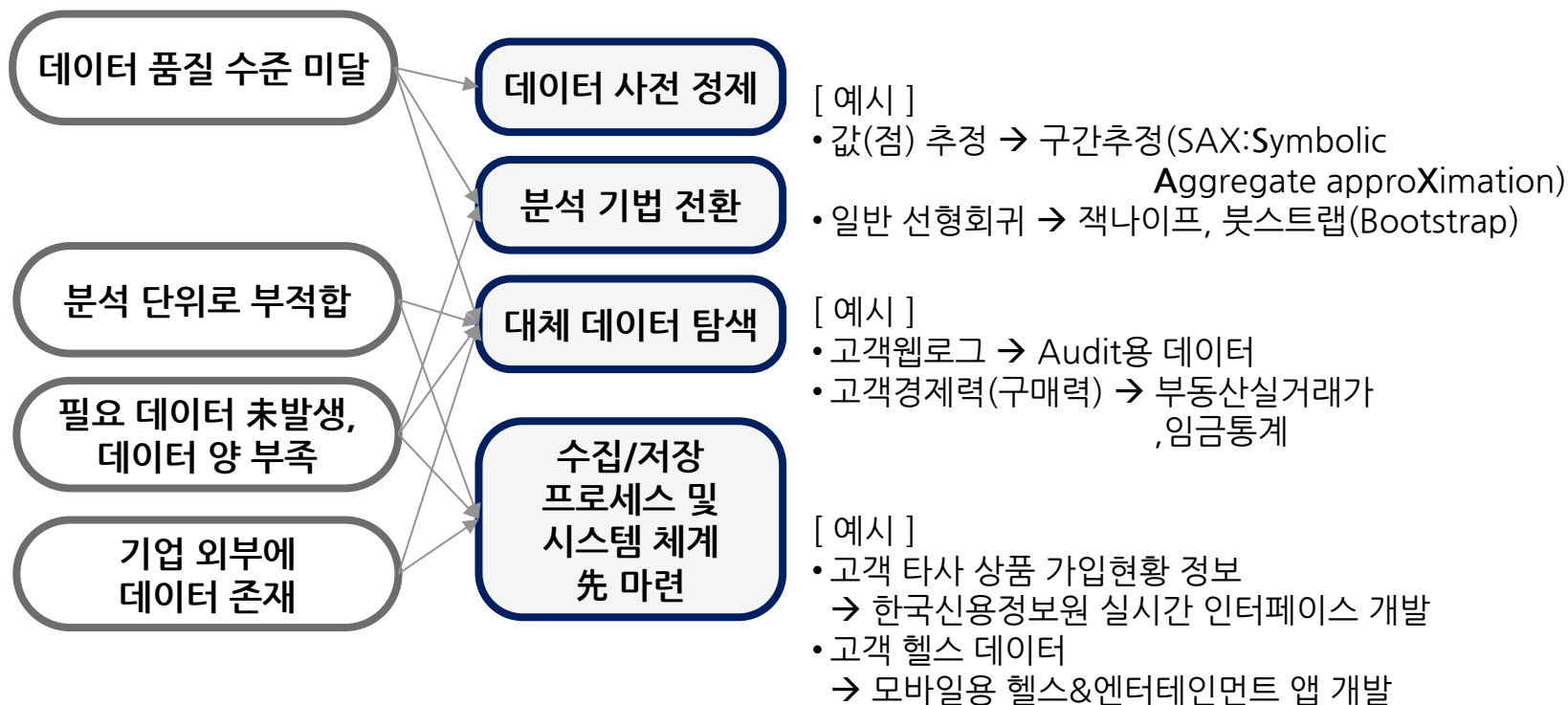
- 빅데이터 분석 프로젝트의 품질저하와 목표수준 하향조정으로 이어짐



데이터활용 장애요인에 대한 대안마련

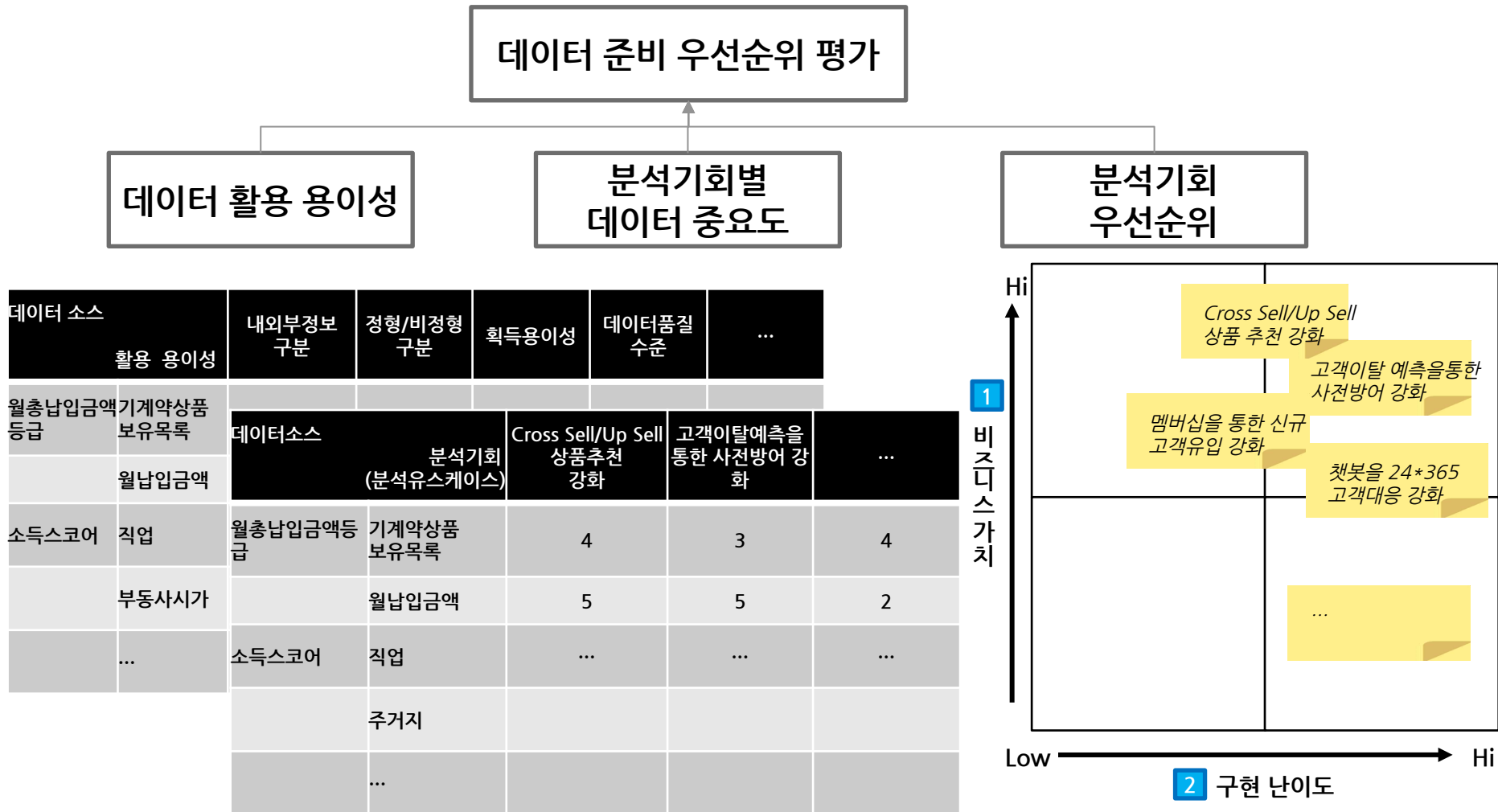
- 데이터활용 장애요인에 따라 적절한 대안을 사전에 마련해야 함

Time



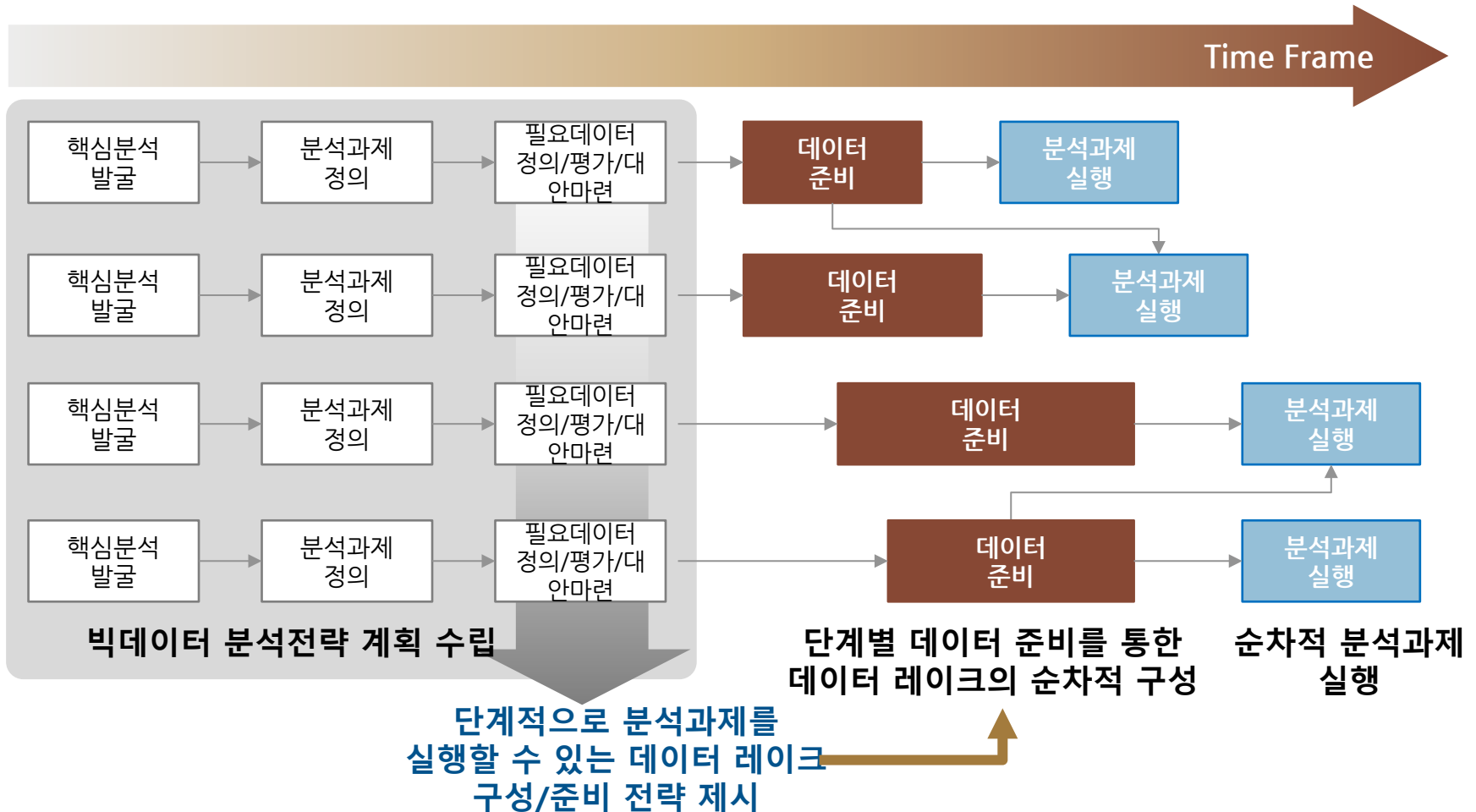
데이터 준비를 위한 우선순위 평가

- 데이터 준비를 위한 우선순위를 평가하고, 이에 따른 단계적 준비방안 수립이 필요함

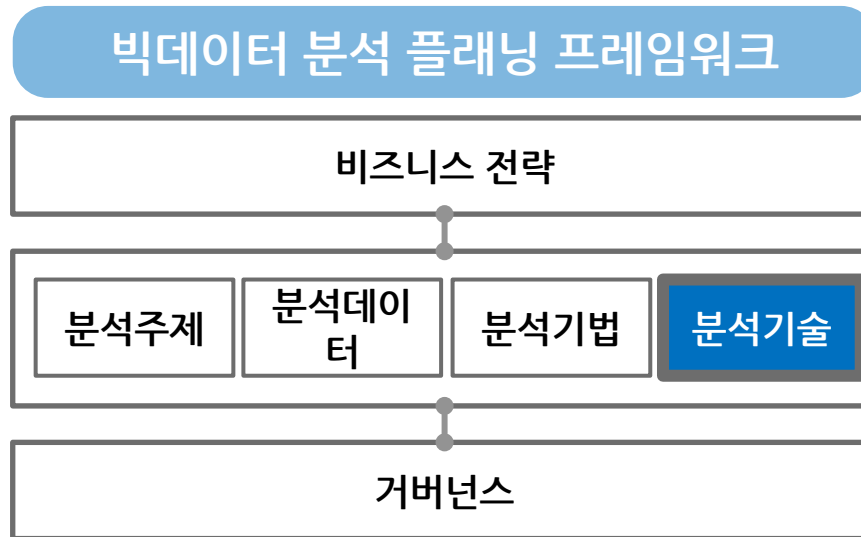


정리 : 데이터 준비는 어떻게 해야 하는가?

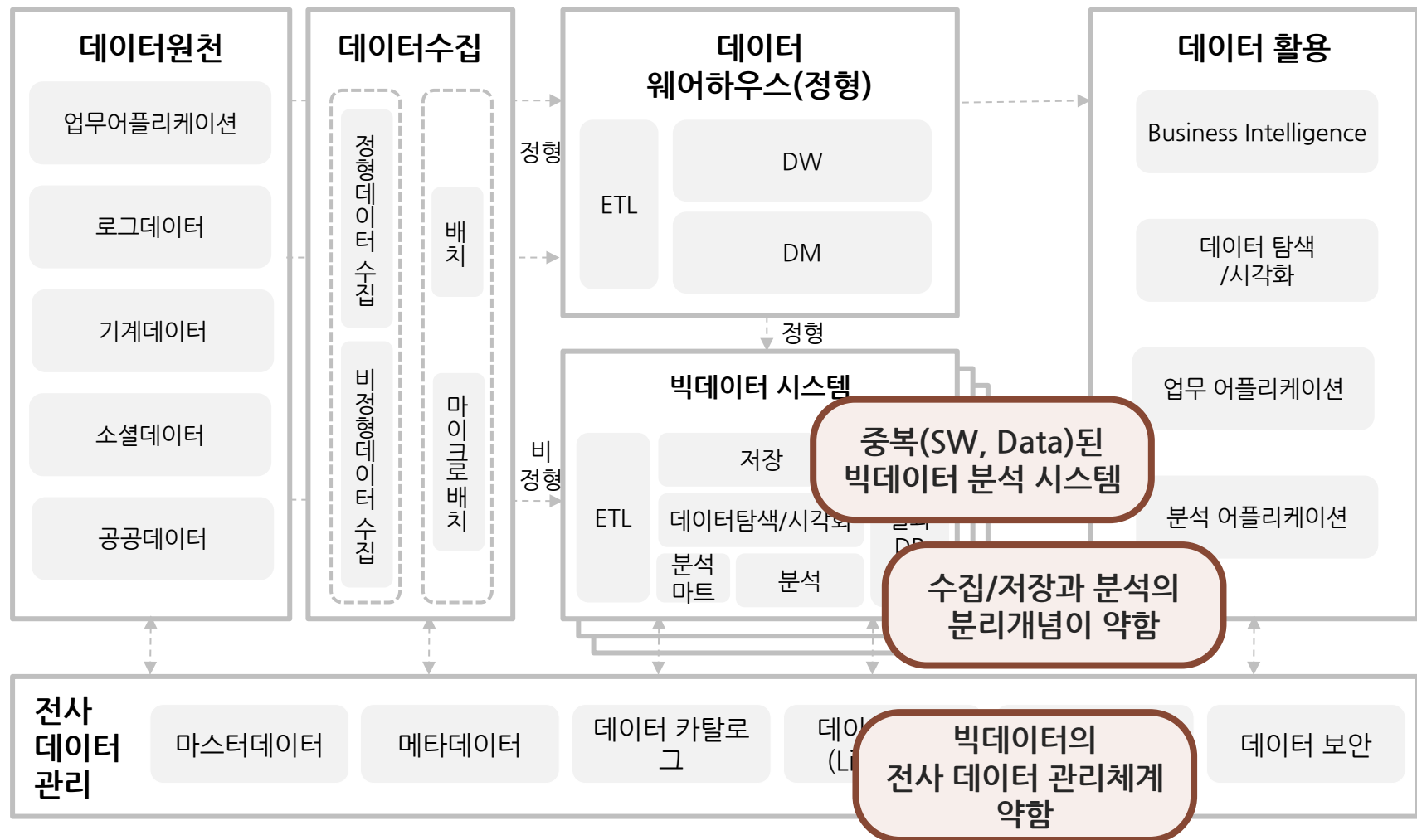
- 분석과제의 실행의 로드맵은 단계별 데이터 준비 로드맵을 고려하여 수립되어야 시행착오를 최소화 하고 분석품질을 극대화 할 수 있음



Q3.아키텍처 구성 시 고려요소는?

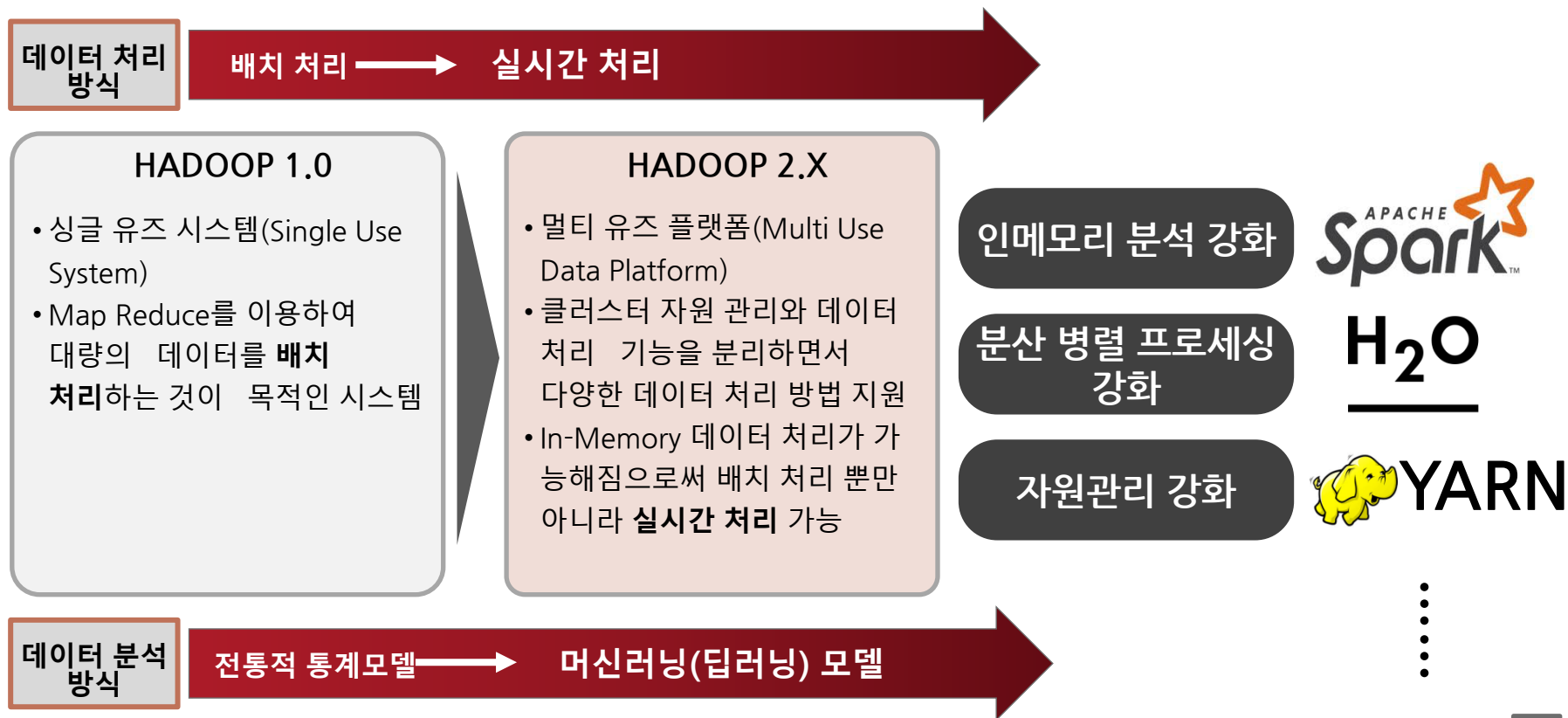


현행 빅데이터 아키텍처



실시간 처리, 자원관리 강화

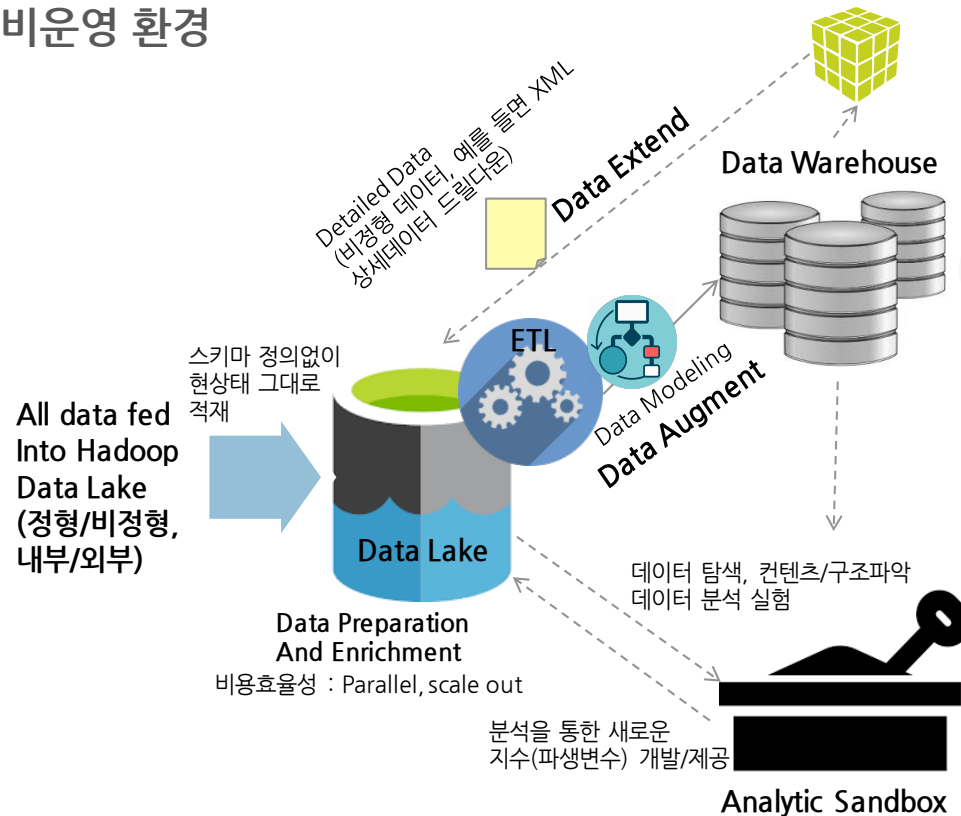
- 실시간 분석 요건 강화, 분석 알고리즘 복잡도 증대로 인해 인메모리 분석/분산 병렬 처리/컴퓨팅 자원관리를 강화하기 위한 기술요소들을 고려해야 함



빅데이터 시스템의 역할 정립

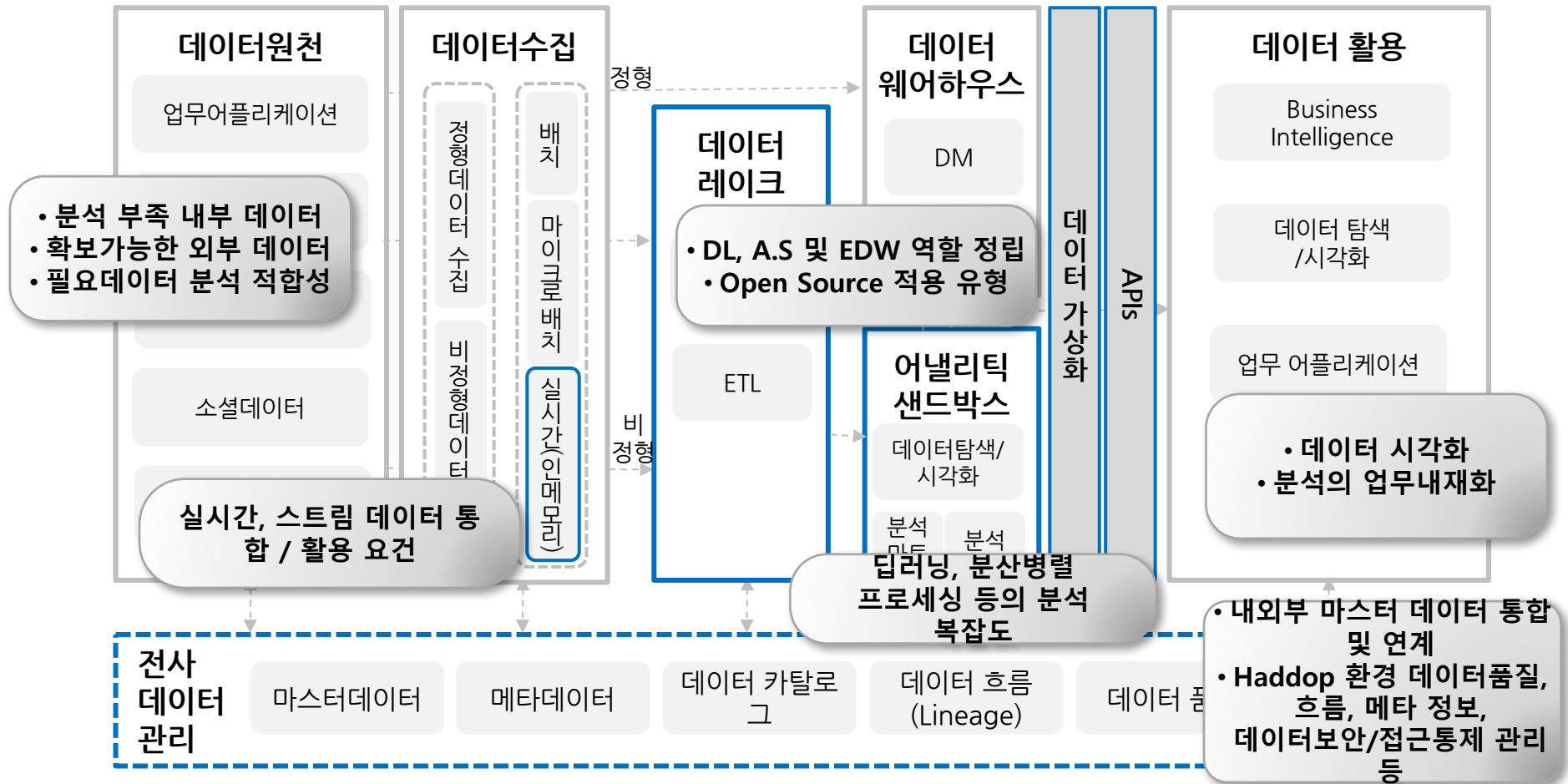
- EDW는 데이터 사용을 위해 정제되고 구조화 된 저장소
- Data Lake는 데이터를 사용하기 전에 다양한 종류(정형/비정형, 내부/외부)의 대용량 데이터를 저장해 두는 객체 기반 저장소
- Analytic Sandbox는 데이터사이언티스트들이 아이디어 테스트, 데이터가공, 모델테스트를 할 수 있도록 준비된 비운영 환경

- ✓ All data is loaded from source systems.
No data is turned away.
- ✓ Data is stored at the leaf level in an untransformed or nearly untransformed state.
- ✓ Data is transformed and schema is applied to fulfill the needs of analysis.

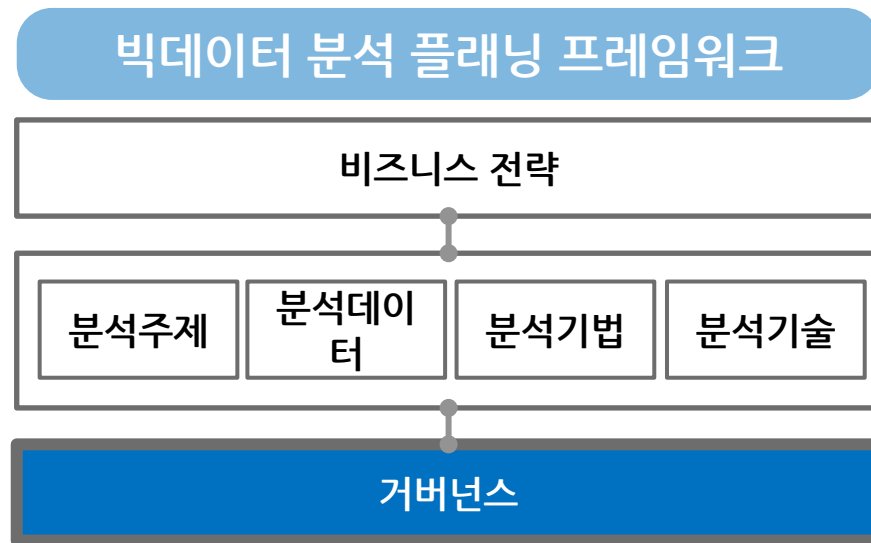


정리 : 빅데이터 아키텍처

- 실시간 분석, 분석복잡도, 분석업무의 운영시스템 내재화, 빅데이터 시스템의 역할 정립을 고려한 아키텍처 구성이 필요함

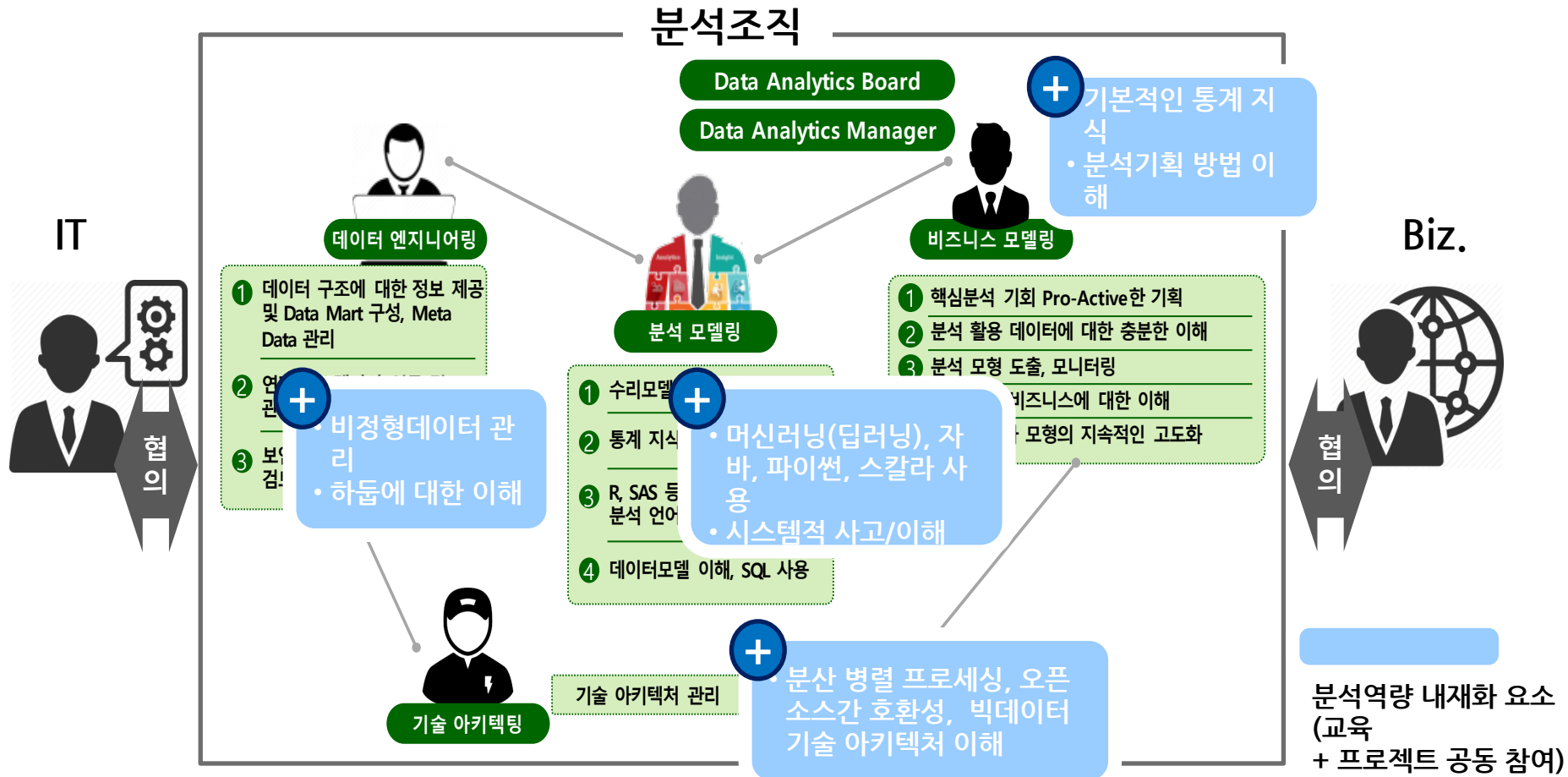


Q4.빅데이터 관점의 거버넌스 특성요소는?



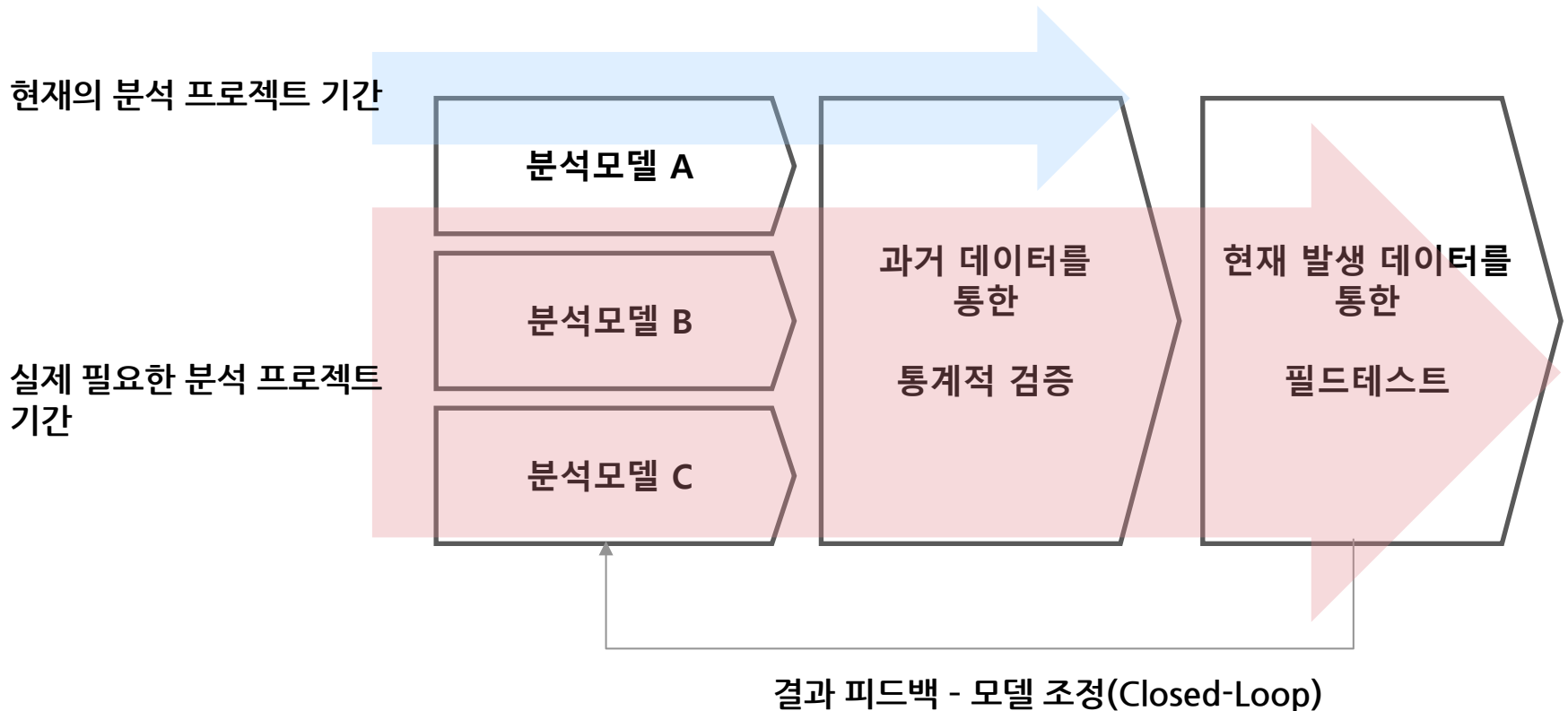
분석조직 역량

- 기존 조직의 각 영역 전문가(Biz. IT)들로 분석조직 구성 시, 분석역량을 내재화 하기 위한 방안 수립 필요



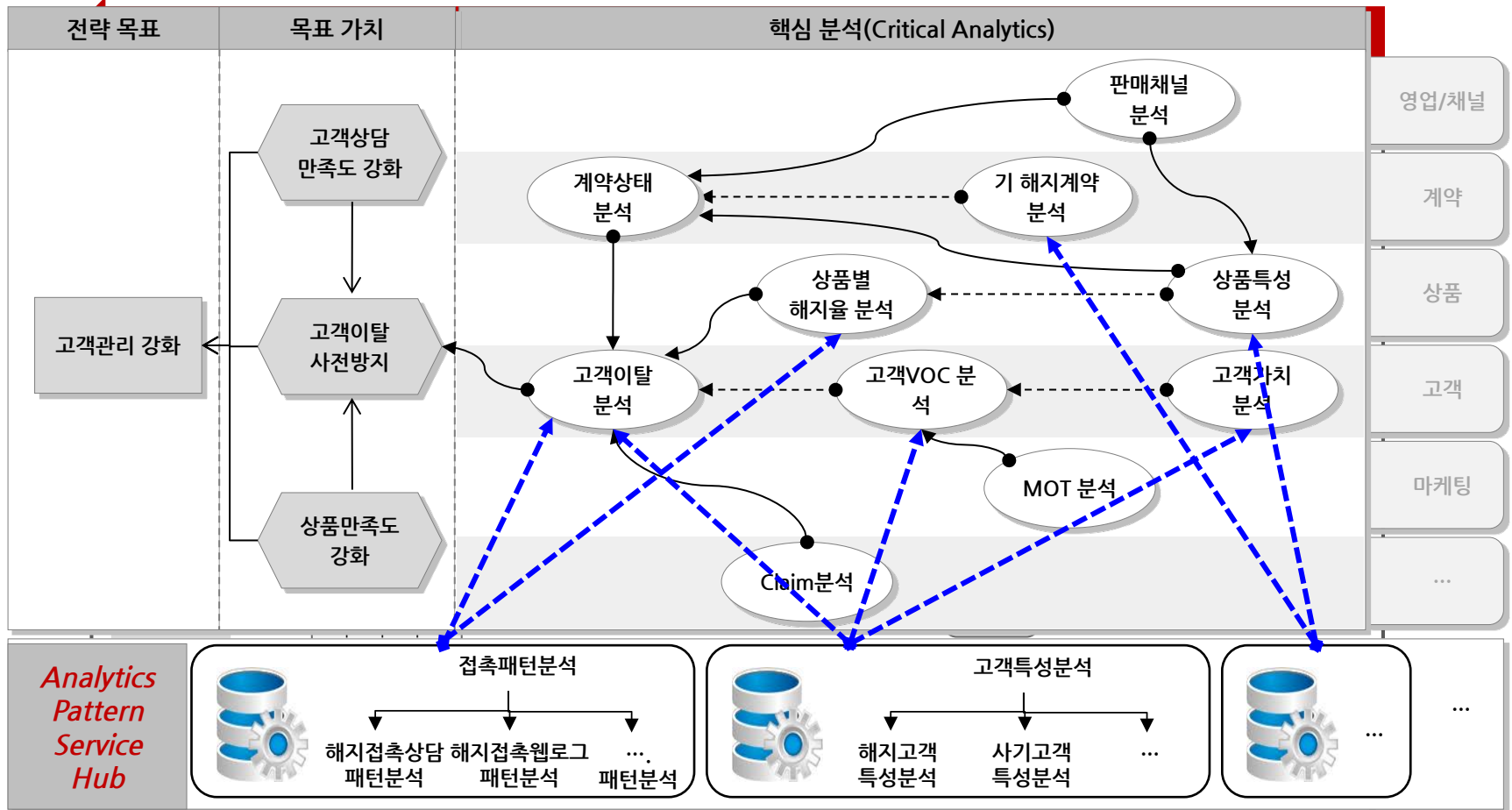
분석 모델 평가

- 분석 모델 개발은 과거의 업무시스템/BI 시스템 개발과 적용 사이클이 다름
- 다양한 분석방식 적용, 필드테스트 수행 및 피드백을 고려한 프로젝트 기간에 대한 고려가 필요함



분석의 전사연계 관리

- 분석(분석 유스케이스, 분석모델, 데이터)을 자산화 하여 관리하고, 사일로化 되지 않도록 연계성을 관리하는 것이 필요함



정리 : 빅데이터 플래닝의 4가지 핵심요소

분석주제

빅데이터 분석주제를 비즈니스 모델과 연계하여 발굴 및 구체화 한 후, 이를 기반으로 기술인프라와 추진로드맵을 수립하는 **분석전략계획 수립** 과정이 필요합니다.

분석데이터

분석 질문에 따라 필요한 데이터를 사전 정의하고, 데이터의 가치와 활용 가능성을 평가하여, **단계적인 데이터 준비 전략을 수립**해야 합니다.

분석기술

실시간 분석 요건, 분석 복잡도 증대, 분석의 운영업무 내재화, 빅데이터 시스템의 역할 세분화를 고려한, **업그레이드 된 빅데이터 아키텍처 구성**이 필요합니다.

거버넌스

분석역량 내재화 프로그램 운용, 분석모델의 평가/운영 방식, 데이터 모델 관리범위의 확장, 분석의 자산화 및 전사연계 관리라는 **빅데이터의 특징을 고려한 거버넌스**가 필요합니다.

END

Jeong So Young
Data Science
Associate Director

Mobile : +82.10.2028 0312
Phone : +82.2.786.7281
Email : jsy@2e.co.kr