

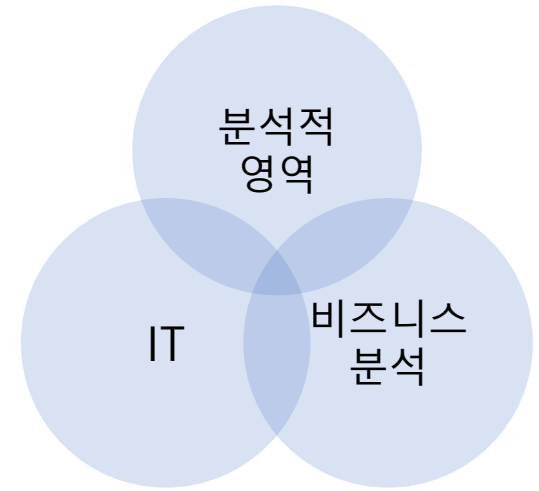
ADsP 데이터 분석 준전문가

2과목 데이터 분석 기획

데이터 사이언스

❖ 데이터로부터 의미 있는 정보를 추출해 내는 학문

↳ 정형·비정형을 막론하고 데이터 분석 (총체적 접근법)



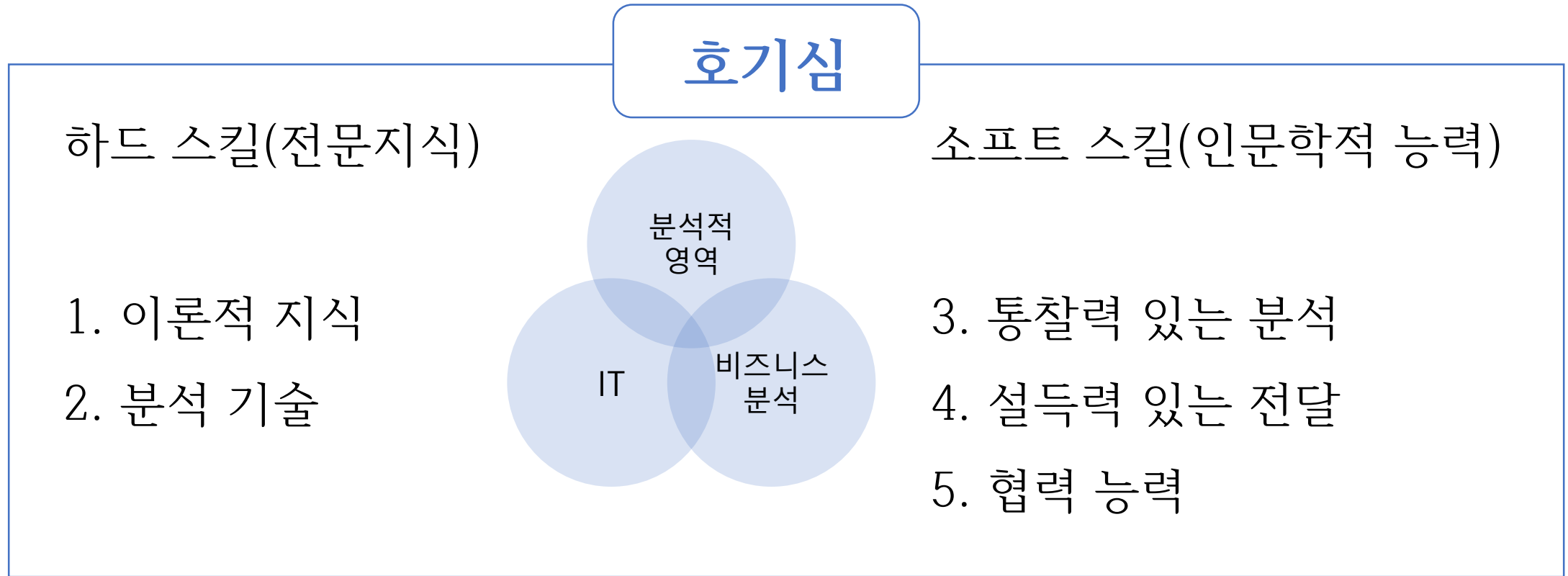
❖ 데이터 사이언스 영역

↳ 분석적 영역: 수학, 확률모델, 분석학 등등 과 같은 이론적인 지식

↳ IT: 프로그래밍, 데이터 엔지니어링, 고성능 컴퓨팅 과 같은 프로그래밍적 지식

↳ 비즈니스 분석: 커뮤니케이션, 시각화, 프레젠테이션 과 같은 비즈니스적 능력

데이터 사이언티스트 요구 역량



분석기획 방향성 도출

분석 방법 \ 분석 대상	분석 대상	
	Known	Un-Known
Known	최적화 (Optimization)	통찰 (Insight)
Un-Known	솔루션 (Solution)	발견 (Discovery)

목표시점 별 분석방식

목표 시점 별 분석 기획 방안		
과제 단위		장기적인 마스터 플랜
Speed & Test 테스트 위주로 빠르게	1차 목표 ↔	Accuracy & Deploy 정확하게 배포하기 위함
Quick & Win 빨리 & 이기기	과제의 유형 ↔	Long Term View 장기적인 시각
Problem Solving 문제 해결	접근 방식 ↔	Problem Definition 문제 정의

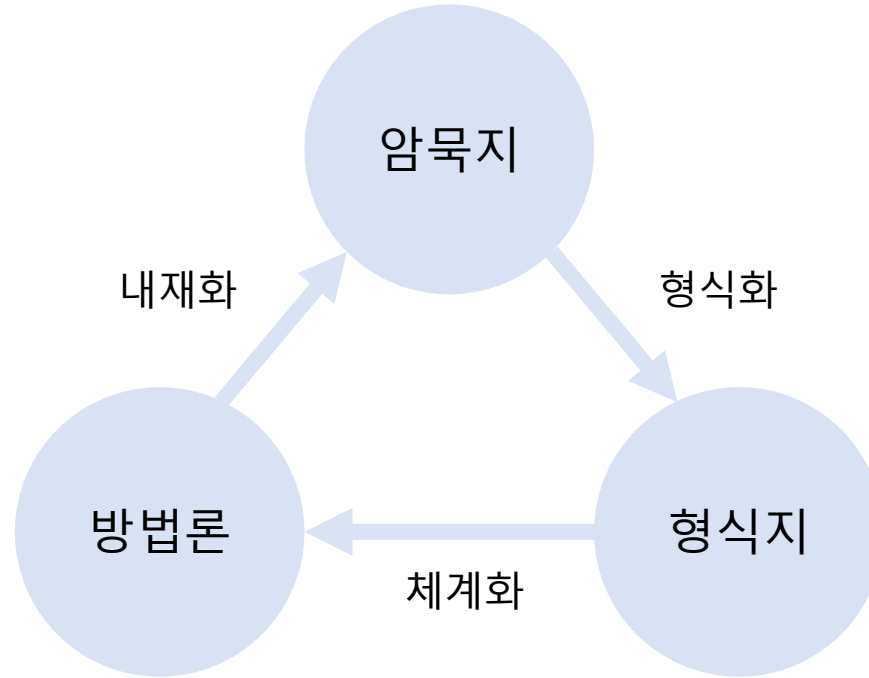
분석 기획시 고려사항

- ❖ 가용 데이터 : 분석의 기본이 되는 데이터 확보 및 파악
- ❖ 적절한 유스 케이스 : 기존에 잘 구현되어 활용되고 있는 유사 분석 시나리오 및 솔루션 최대한 활용
- ❖ 장애요소들에 대한 사전계획 수립 : 일회성 분석에 그치지 않고 충분하고 지속적인 교육을 통한 조직의 역량으로 내재화

데이터 기반 의사결정 필요성

- ❖ 경험과 감에 따른 직관적인 의사결정 → 데이터 기반의 의사결정
- ❖ 합리적 의사 결정을 가로막는 장애요소: 프레이밍 효과, 고정 관념, 편향된 생각
 - ↳ 프레이밍 효과 : 문제의 표현 방식에 따라서 동일 상황임에도 개인의 판단이나 결정이 달라지는 경우가 있음

분석 방법론 생성 과정



❖ 암묵지와 형식지 사이의 상호작용

↳ 공통화 → 표출화 → 연결화 → 내면화



분석 방법론

❖ 폭포수 모델: 단계를 순차적으로 진행, 이전 단계가 완료되어야 다음 단계로 진행 가능

↳ 요구분석 및 정의 → 설계 → 구현 및 단위 테스트 → 통합 테스트 → 운영 및 유지보수

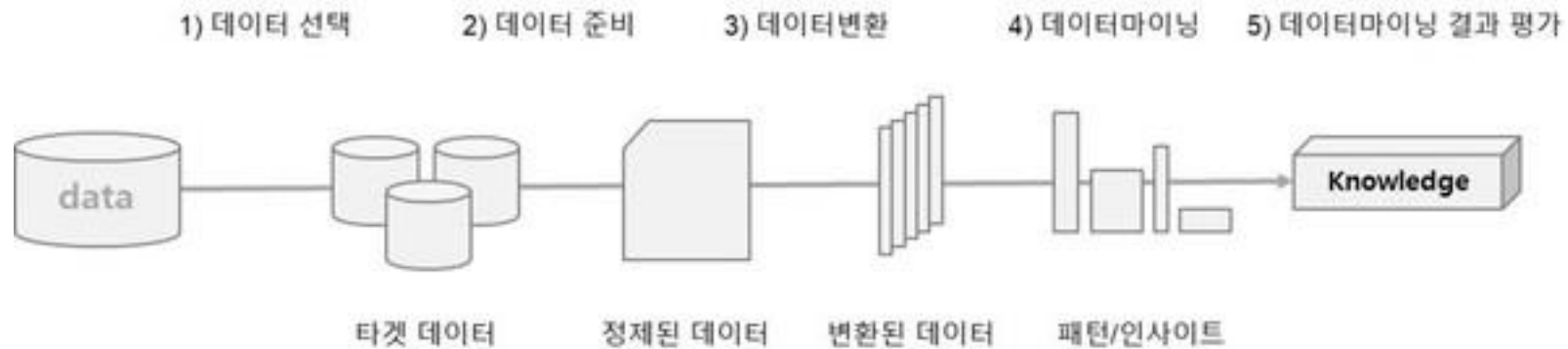
❖ 나선형 모델: 계획 및 정의 → 위험 분석 → 개발 → 고객 평가를 거쳐 점진적으로 프로젝트 완성

↳ 관리체계를 효과적으로 갖추지 못한 경우 복잡도가 상승한다.

❖ 프로토타입 모델: 일부분을 우선 개발하고 개선 작업을 거치는 것 중요한 기능들이 포함되어 있는 시스템의 초기 모델

↳ 요구사항 분석 → 시제품 설계 → 시제품 개발 → 시제품 평가 → 시제품 정제 → 완제품 생산

KDD 분석 방법론



출처 : <https://www.2e.co.kr/news/articleView.html?idxno=301010>

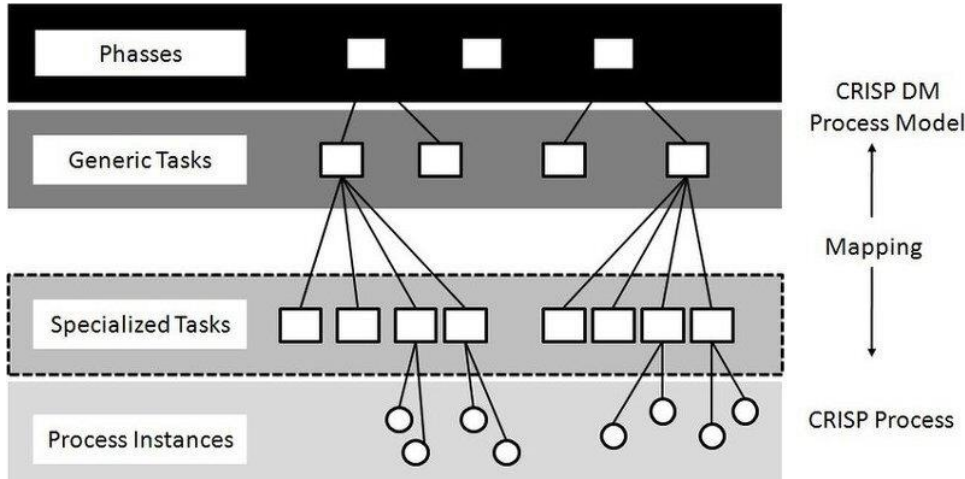
❖ 데이터로부터 통계적 패턴이나 지식을 찾기 위한 데이터 마이닝 프로세스

❖ 데이터셋 → 데이터 전처리 → 데이터 변환 → 데이터 마이닝 → 분석 결과 평가

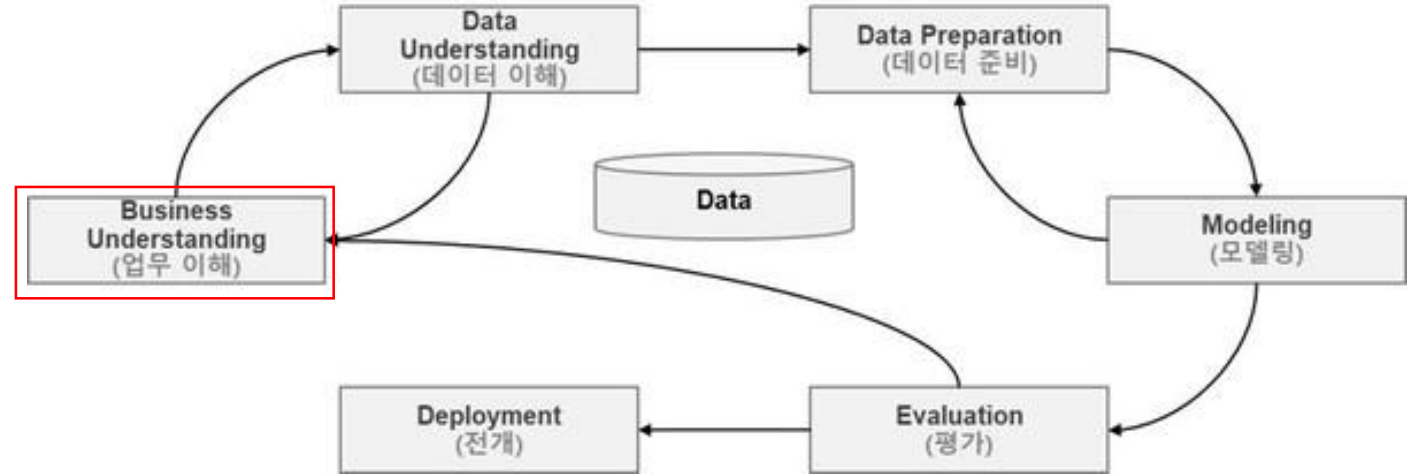
↳ **전처리 과정**에서 이상 값, 잡음을 식별

↳ **데이터 변환 과정**에서 분석 목적에 맞는 변수 선택 및 차원 축소 과정을 거친다.

CRISP-DM 분석 방법론



출처 : <https://kun-hee.tistory.com/entry/CRISP-DM-%EB%B6%84%EC%84%9D-%EB%B0%A9%EB%B2%95%EB%A1%A0>



출처 : <https://www.2e.co.kr/news/articleView.html?idxno=301010>

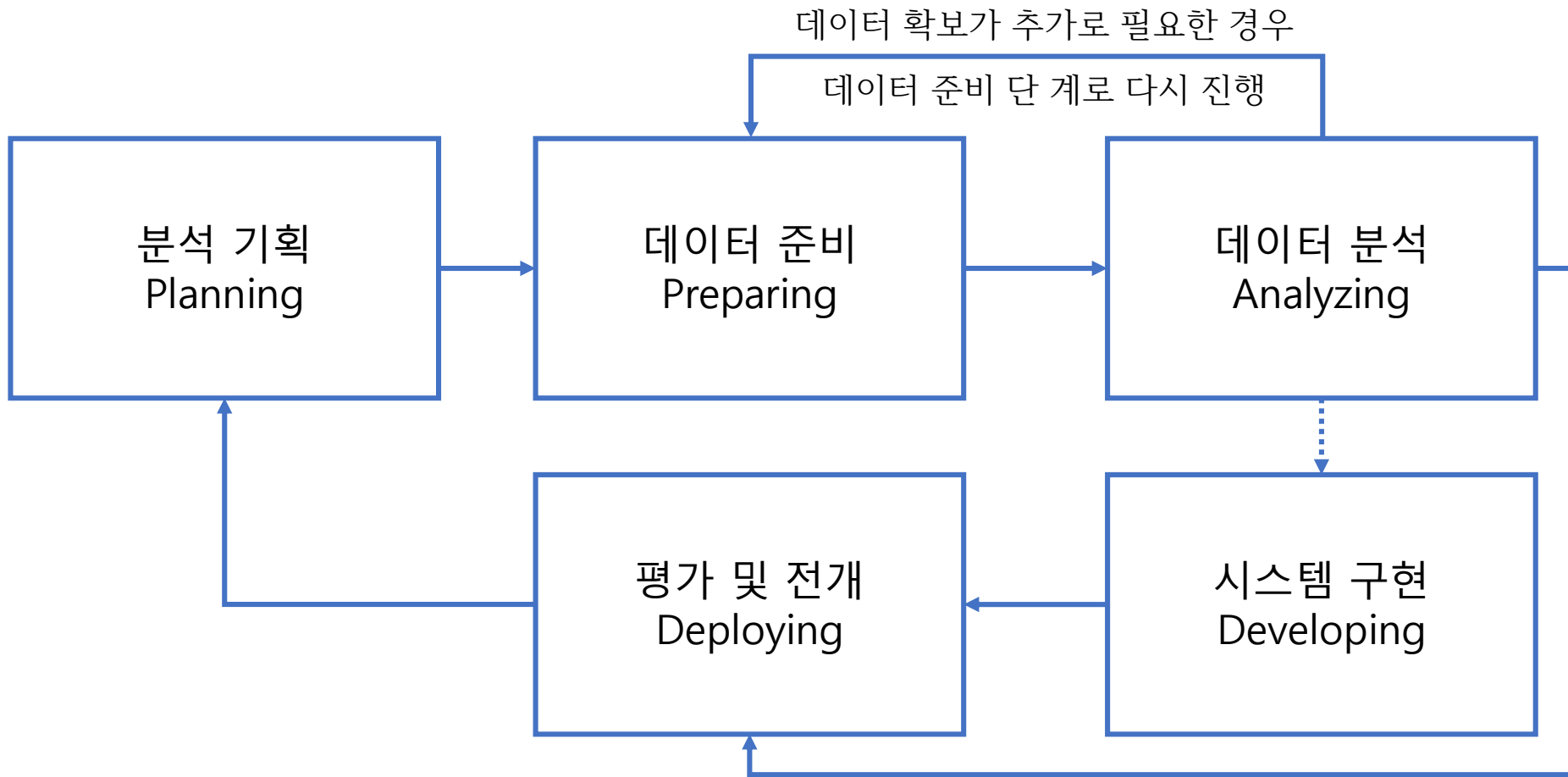
❖ 4개 레벨로 구성된 계층적 프로세스 모델

❖ 업무 이해 → 데이터 이해 → 데이터 준비 → 모델링 → 평가 → 전개

↳ 데이터 준비 과정은 KDD 분석에서의 데이터 변환과정과 유사

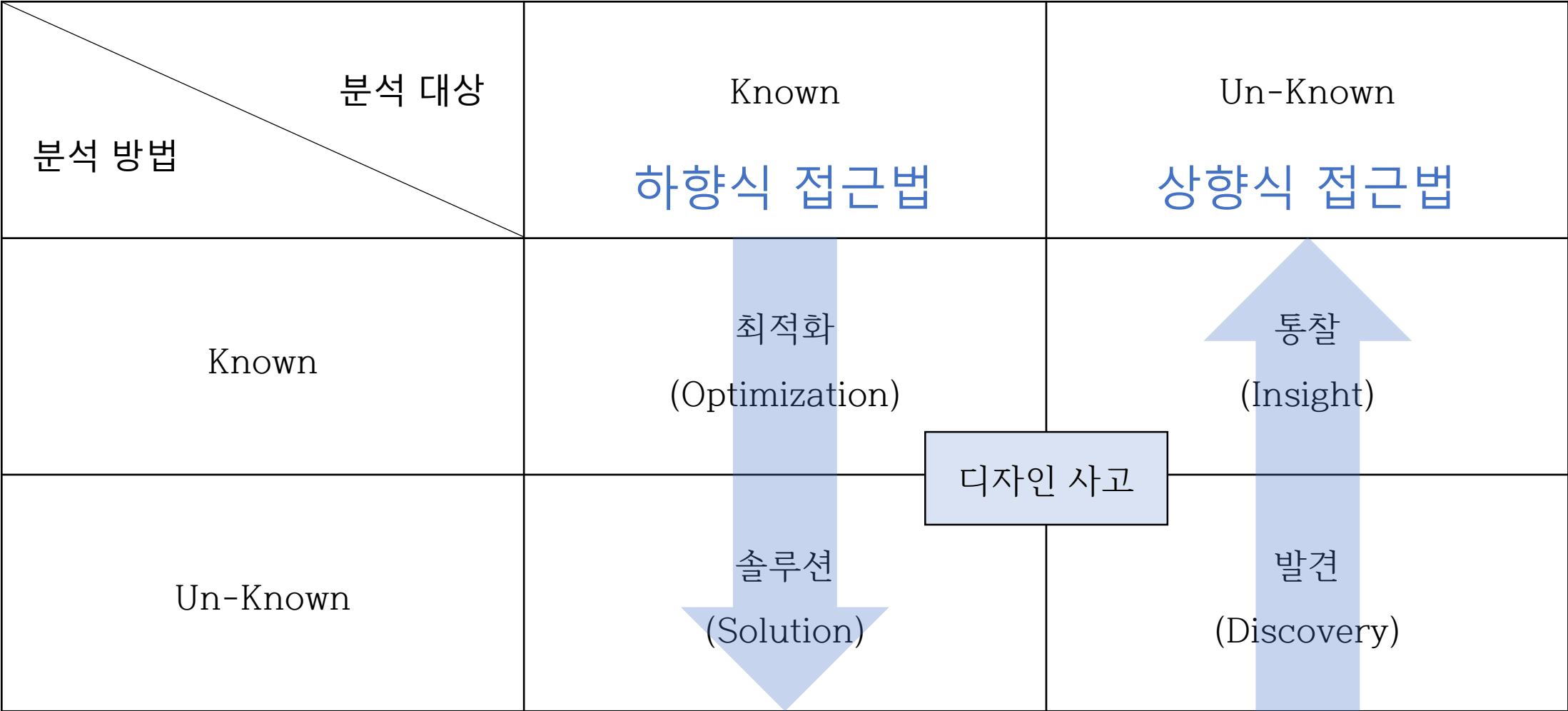
↳ 모델링 단계에서 **모델 평가** 수행, **모델 적용성 평가**는 평가 단계에서 진행함

빅데이터 분석 방법론



❖ 분석 기획 단계 : 범위 설정 → 프로젝트 정의 & 계획 → **위험 식별** & 대응 계획 수립

상향식 접근법 vs 하향식 접근법



하향식 접근

❖ 문제가 주어지고 이에 대한 해법을 찾기 위해 과정이 진행(지도 학습)

↳ 문제 탐색 → 문제 정의 → 해결방안 → 탐색 타당성 검토

❖ 문제탐색

↳ 문제를 해결함으로써 발생하는 가치에 중점을 두는 것이 중요

↳ 비즈니스 모델 기반 : 업무·제품·고객·규제와 감사·지원 인프라

↳ ‘지원 인프라’ ‘업무’중에 ‘고객’ 이 ‘제품’을 ‘규제와 감사’ 했다

↳ 거시적 관점: STEEP(사회·기술·경제·환경·정치)

하향식 접근

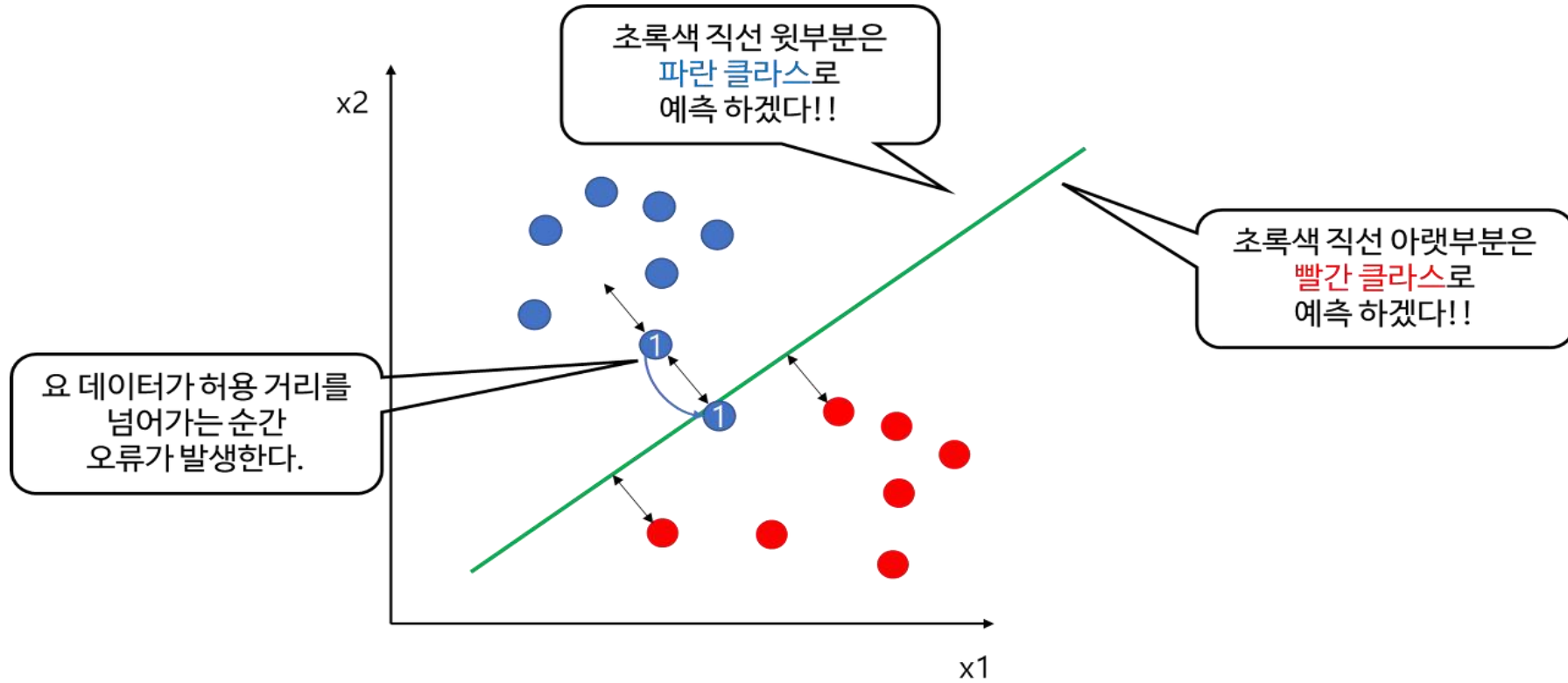
❖ 탐색 타당성 검토 : 대안을 과제화 하기 위해서 다각적인 타당성 분석이 수행

↳ 경제적 타당성 : 비용대비 편익 분석 관점의 접근이 필요

↳ 데이터 타당성 : 데이터 존재 여부, 분석 역량이 필요

↳ 기술적 타당성 : 역량 확보 방안을 사전에 수립

하향식 접근(분류 분석)

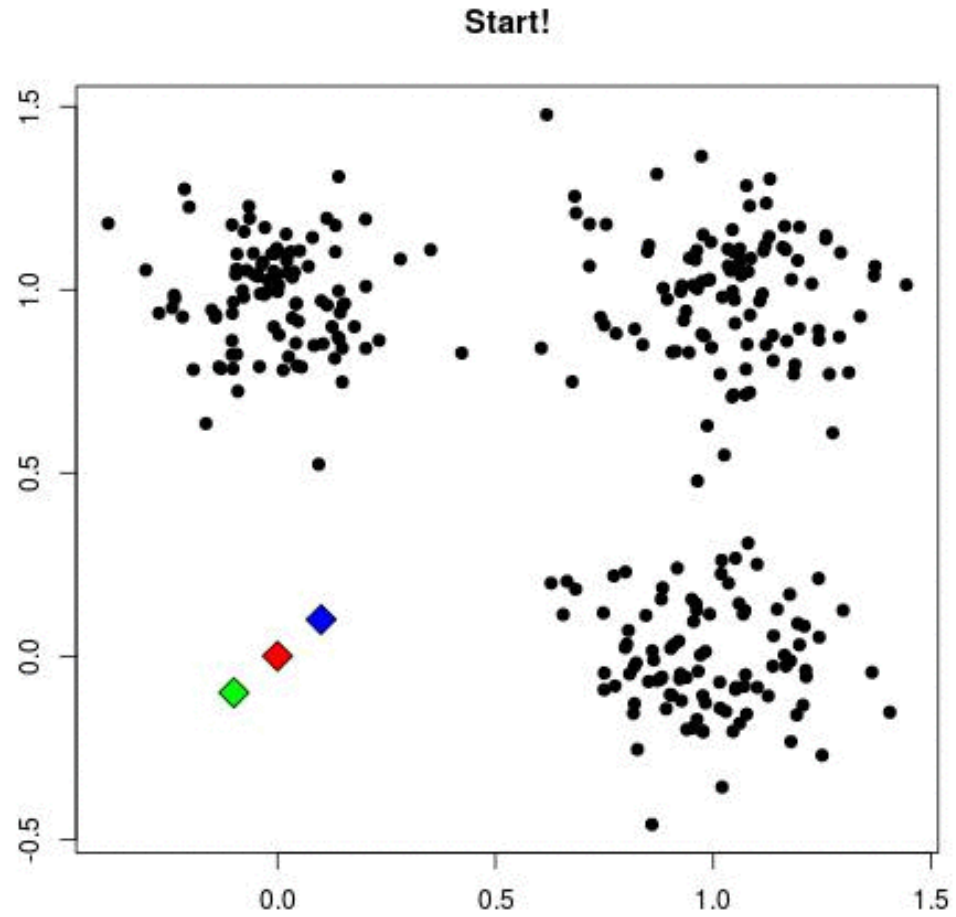


상향식 접근

❖ 비지도 학습 방법에 의한 데이터 분석

- ↳ 하향식 접근 방법의 한계를 극복하기 위한 분석 방법론
- ↳ 문제의 정의 자체가 어려운 경우 데이터를 기반으로 문제를 탐색

상향식 접근(군집 분석)



지도 학습과 비지도 학습

지도 학습	비지도 학습
<ul style="list-style-type: none">• 명확한 목적(원인과 결과)을 가지고 데이터 분석을 실시• 미리 정의된 데이터에 대한 라벨(label)을 바탕으로 학습 및 예측	<ul style="list-style-type: none">• 명확한 목적을 알지 못할 때 데이터 분석을 실시• 데이터 자체의 결합, 연관성을 중심으로 데이터의 상태를 표현
자료가 입력변수와 출력변수로 주어짐 예측 모형을 얻을 때 사용	데이터 마이닝에서 자료가 출력변수 없이 입력변수만 주어지는 경우

분석과제 정의서

- ❖ 분석별로 필요한 [소스데이터](#)
- ❖ 분석방법
- ❖ 데이터 입수 및 분석 난이도
- ❖ 상세 분석
- ❖ 분석 수행 주기

분석 과제 관리를 위한 주요 요인

- ❖ 데이터 크기

- ❖ 데이터 복잡도

- ❖ 속도

- ❖ 분석 복잡성

- ❖ 정확성·정밀도

↳ 정확도를 생각해 분석을 실행하면 복잡해지고, 정밀도를 생각해 분석을 실행하면 정확도가 떨어지는 경우가 많다. (trade off 관계)

분석 프로젝트 관리 항목 10가지

❖ 범위, 시간, 원가, 품질, 통합, 조달 자원, 리스크, 의사소통, 이해관계자

범위통합이해관계자 조달리스크품질 시간원가 의사소통자원

범통이 조리품 시원 의자!

분석 마스터 플랜 수립



출처: <https://velog.io/@zinu/ADsP-%EB%8D%B0%EC%9D%B4%ED%84%B0-%EB%B6%84%EC%84%9D-%EA%B8%B0%ED%9A%8D-%EB%B6%84%EC%84%9D-%EB%A7%88%EC%8A%A4%ED%84%B0-%ED%94%8C%EB%9E%9C>

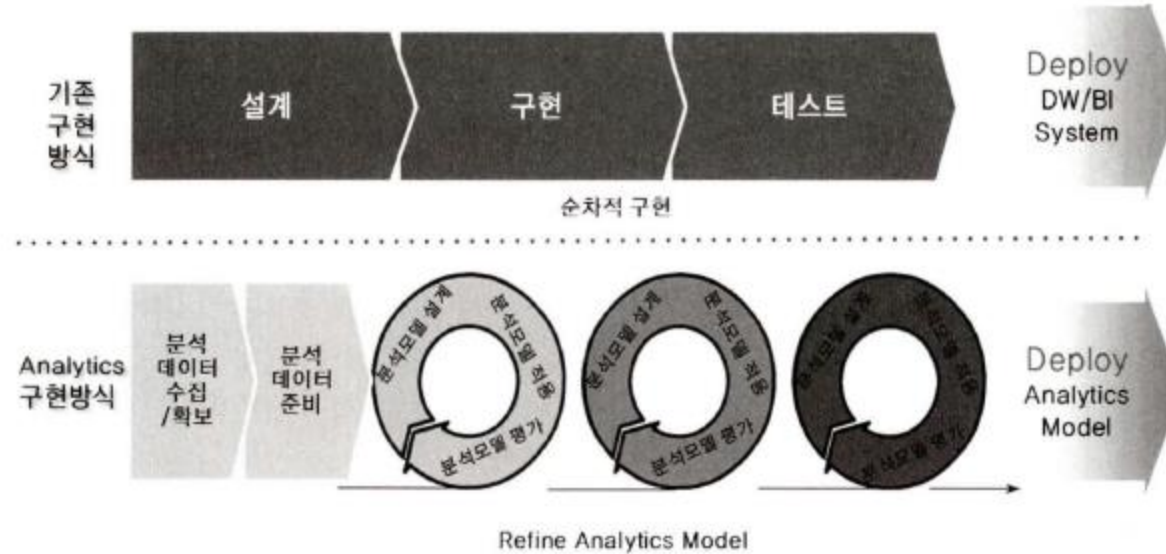
- ↳ 전략적 중요도 : 전략적 필요성, 시급성
- ↳ 실행 용이성 : 투자 용이성, 기술 용이성

ROI 관점 빅데이터 특징

난이도 Difficulty	Difficult	2, 3순위 I	4순위 II
	Easy	1순위 III	2, 3순위 IV
		현재	시급성 미래

- ❖ ROI : 투자 수익률
- ❖ 투자 비용 요소(**난이도**) : 3V= Volume, Variety, Velocity
- ❖ 비즈니스 효과(**시급성**) : Value
- ❖ 가장 쉽고 현재에 시급한 III부터 해결,
- ❖ 가장 어렵고 시간적 여유가 있는 II가 마지막

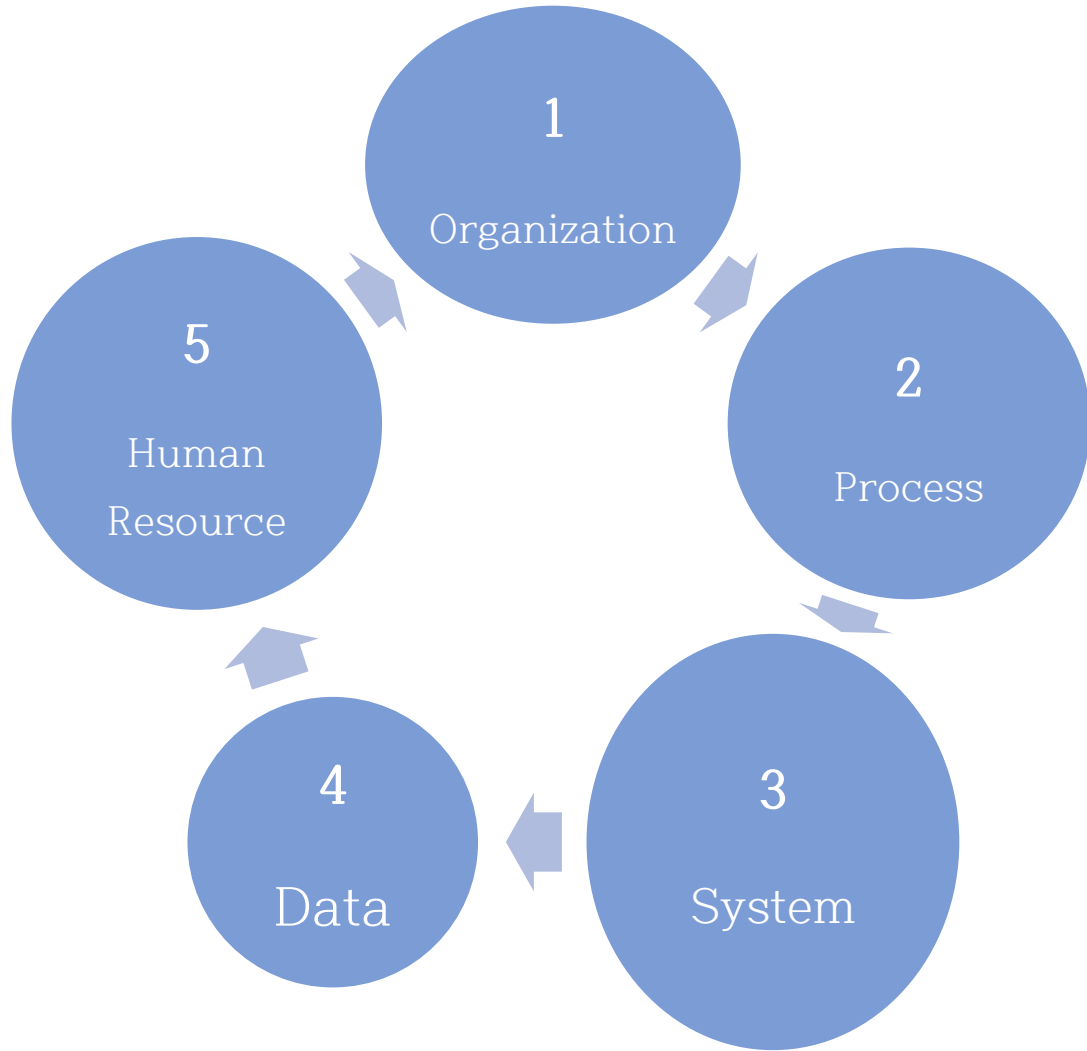
분석 마스터 플랜 세부 이행 계획 수립



출처 <https://m.blog.naver.com/jdhuppy/221343786927>

- ❖ 폭포수 방식도 있으나 반복적인 정련과정을 통하여 프로젝트의 완성도를 높이는 방식을 주로 사용
- ❖ 반복적인 분석체계라고 모든 단계를 반복하는 것이 아닌 모델링 단계를 중점으로 반복하여 진행하고 세부적인 일정계획도 수립.

분석 거버넌스 구성 요소



1. 조직(Organization)

2. 프로세스(Process)

3. 시스템(System)

4. 데이터(Data)

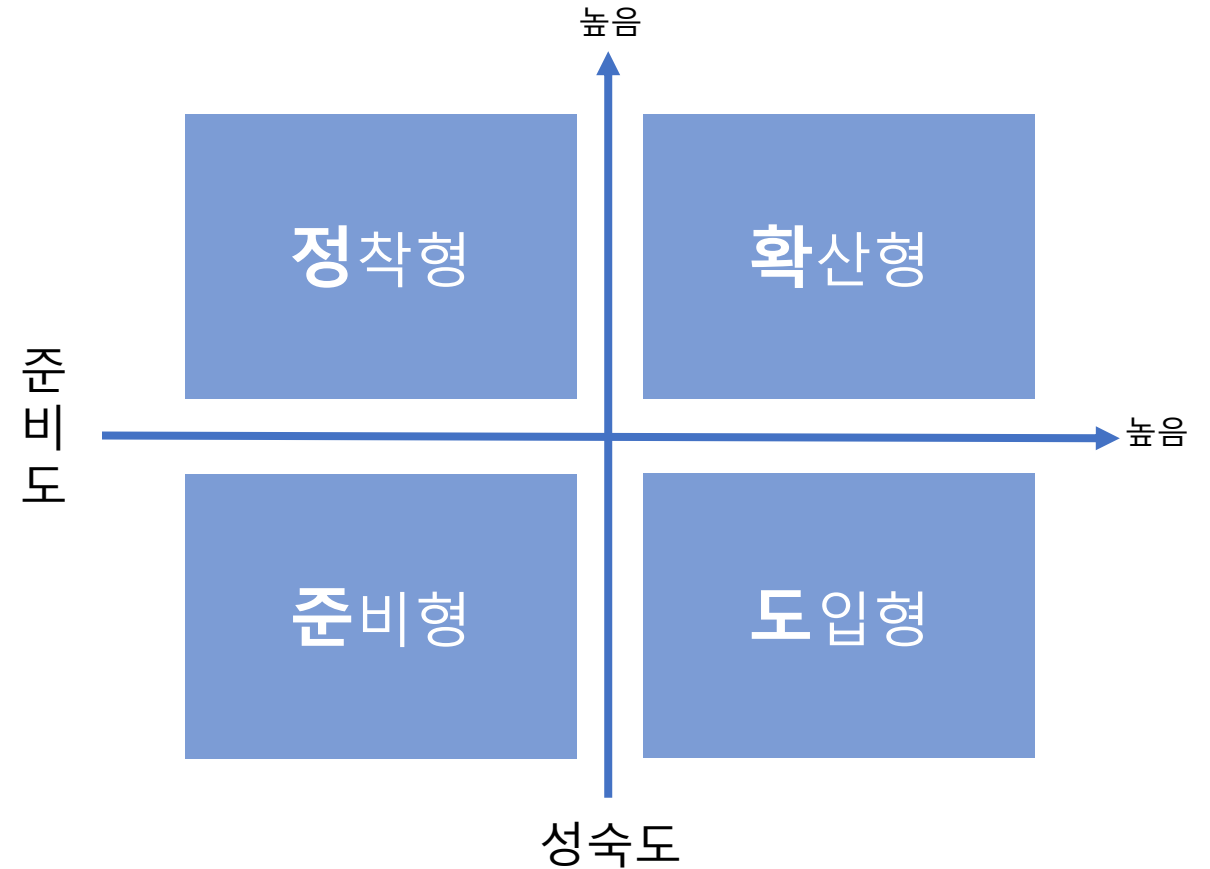
5. 분석 관련 교육 및 마인드 육성 체계
(Human Resource)

데이터 분석 수준 진단

❖ 분석 준비도 : 분석 업무, 인력, 기법, 데이터, 문화, 인프라

❖ 분석 성숙도

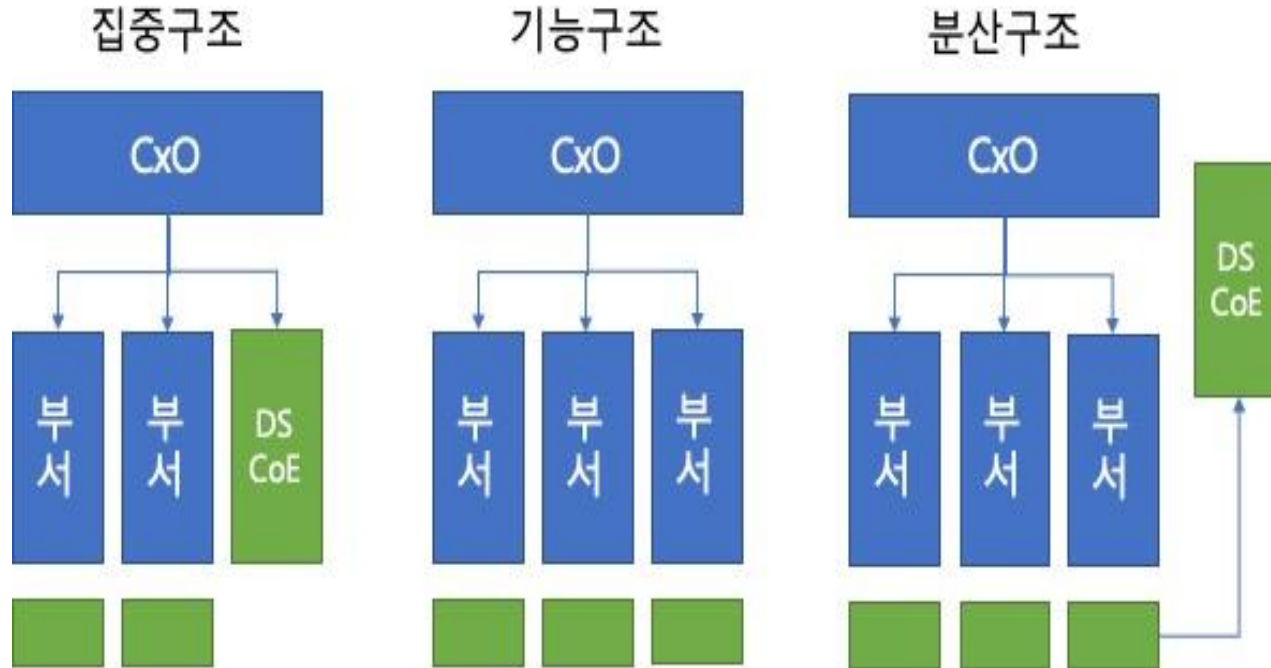
↳ 도입 → 활용 → 확산 → 최적화



데이터 거버넌스

- ❖ 전사 차원의 모든 데이터에 대하여 표준화된 관리체계를 수립하고 운영을 위한 프레임 워크 및 저장소를 구축하는 것
- ❖ 마스터 데이터, 메타 데이터, 데이터 사전은 데이터 거버넌스의 중요한 관리 대상임
- ❖ 데이터 거버넌스 구성 3요소 : 원칙, 조직, 프로세스
- ❖ 데이터 거버넌스 체계 : 데이터 표준화(구축), 데이터 관리 체계, 데이터 저장소 관리, 표준화 활동(점검)

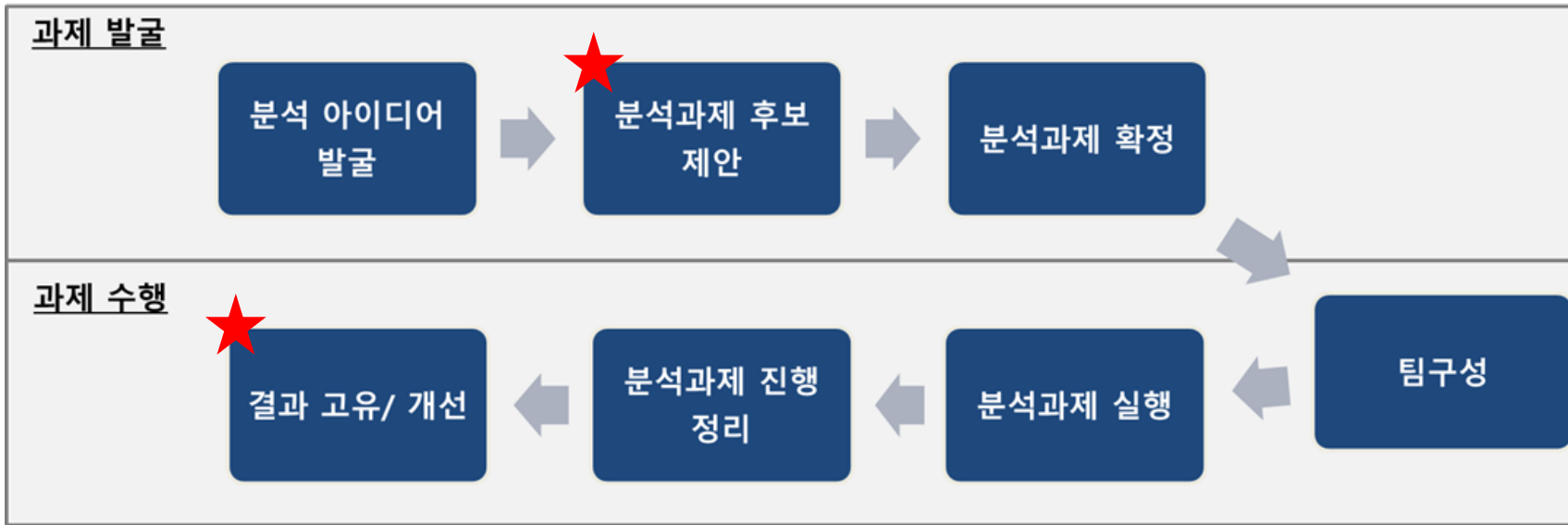
분석을 위한 조직 구조



출처: <https://wikidocs.net/>

- ❖ DSCoE : 분석 조직
- ❖ 집중 구조 : 분석을 위한 부서 존재
- ❖ 기능 구조 : 분석 조직이 없음
- ❖ 분산 구조 : 각각의 부서에 분석 인력 배치

분석 과제 관리 프로세스



❖ **과제 발굴** : 분석 idea 발굴 → 분석 과제 후보 제안 → 분석 과제 확정

❖ **과제 수행** : 팀 구성 → 분석과제 실행 → 분석 과제 진행 관리 → 결과 공유 & 개선

↳ 분석 과제 후보 제안, 결과 공유& 개선 단계에서만 저장공간(Pool)에 축적 및 관리함

문제풀이

2과목 문제풀이 정답

1	④	11	①	21	①	31	②	41	③
2	③	12	①	22	②	32	①	42	④
3	③	13	③	23	②	33	④	43	②
4	②	14	④	24	②	34	④	44	④
5	①	15	①	25	①	35	②	45	①
6	②	16	④	26	①	36	②	46	④
7	④	17	③	27	④	37	③	47	③
8	①	18	①	28	②	38	②	48	④
9	②	19	②	29	③	39	①	49	④
10	②	20	③	30	③	40	②	50	③