

빅데이터 분석 실무과정

Part V : 프로젝트 실습

| Part V. 프로젝트 실습 - 목차

1. 데이터 분석 기획
2. 분석 마스터 플랜 - 생략
3. 데이터 분석 실습

1. 데이터 분석 기획

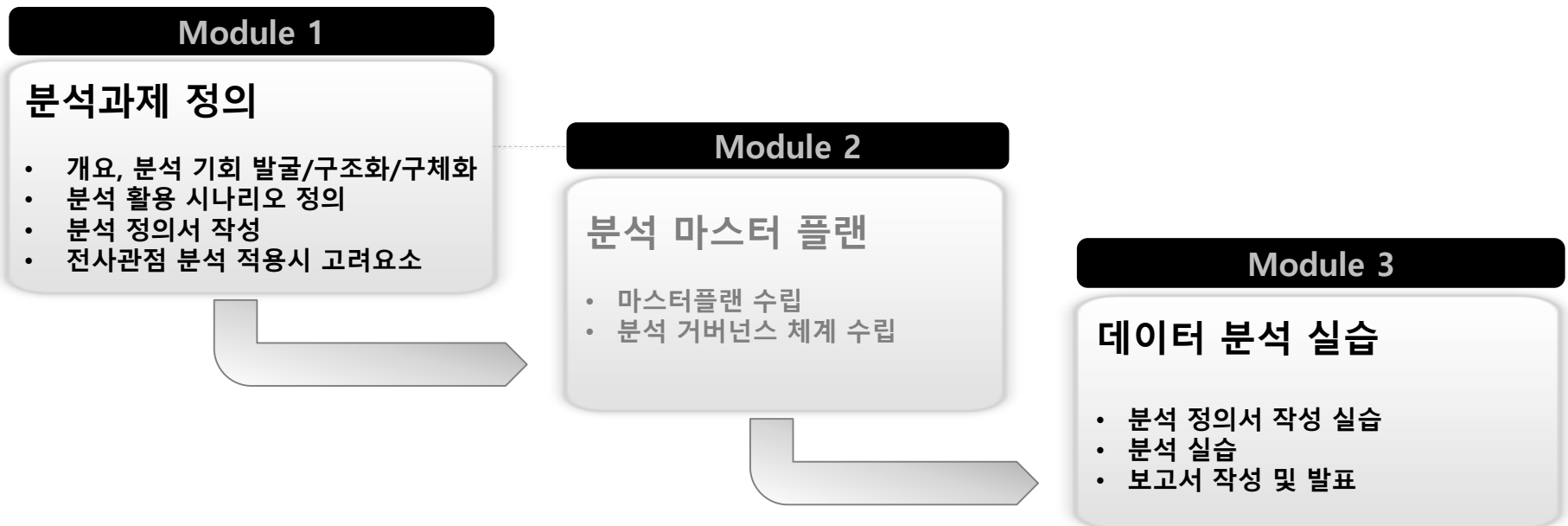
▶▶ Concept, Ecosystem

학습 개요

▶▶ 학습 목표

- 데이터 분석 기회 발굴 방법의 이해
- 데이터 분석 기회 구조화 및 구체화 방법의 이해
- 데이터 분석 활용 시나리오 정의 방법의 이해
- 전사관점 데이터 분석 적용 시 주요 고려요소의 이해

▶▶ 학습 구성

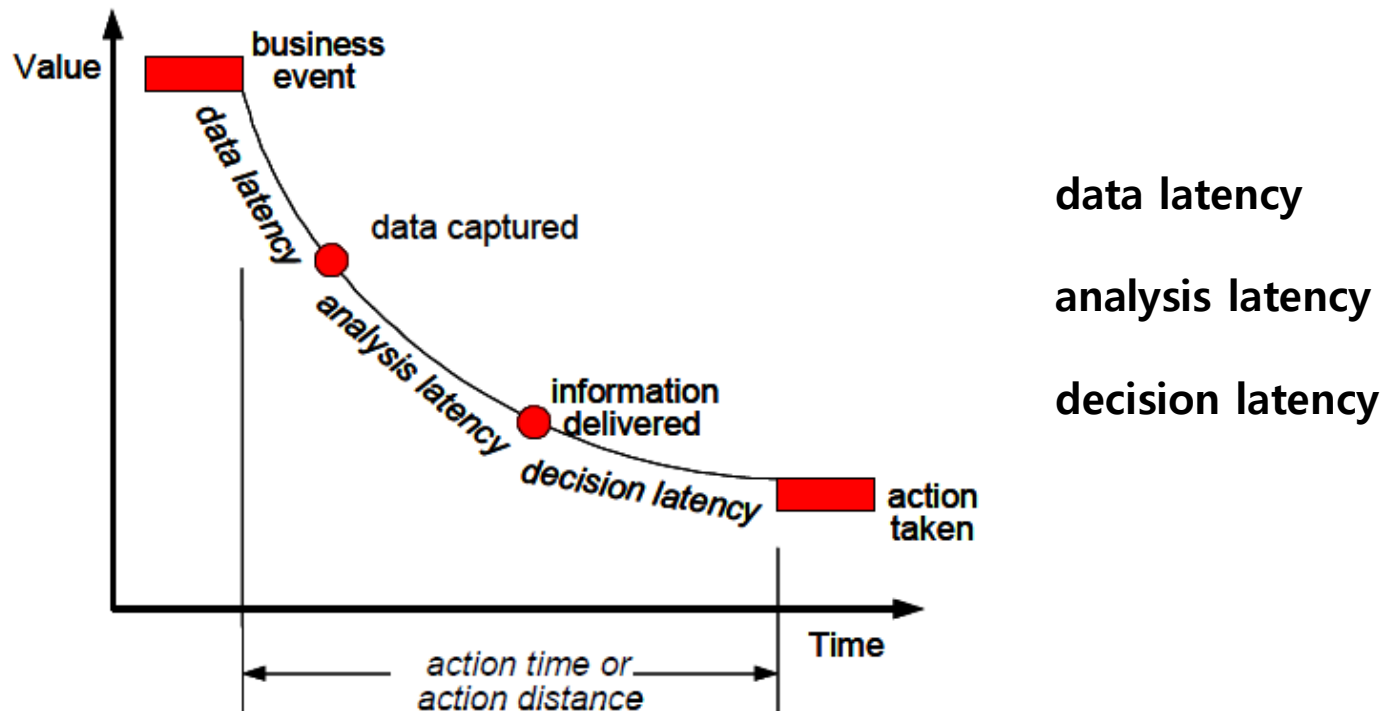


잠재 가치의 발굴

위협요인의 제거

| 데이터 분석의 개요

- ▶ 데이터의 핵심은 데이터 자체가 아닌 분석을 통한 의사결정의 최적화
 - 이벤트가 발생과 이에 대응하는 액션 간에는 Latency가 발생함
 - 데이터 분석, 의사결정 Latency Time이 운영 최적화의 장애요인으로 작용함
 - 빅데이터 기반의 분석 및 업무활용 체계를 도입을 통한 업무운영 최적화를 위해서는 3가지 Latency를 줄이기 위한 관점에서의 접근이 필요함



핵심 성공 요소

▶ 데이터 분석 도입의 성공요소

- Question First 방식 : 의사결정 -> 분석 -> 정보 -> 내/외부 데이터

데이터를 가공하면 정보가 산출된다는 접근이 아니라, 분석 질문이 무엇인지를 정의하고 이를 분석하기 위해 필요한 데이터가 무엇인지를 정의하는 방식으로 접근해야 한다.

- 선택과 집중 : 불보의 소비자 Needs 파악 방법

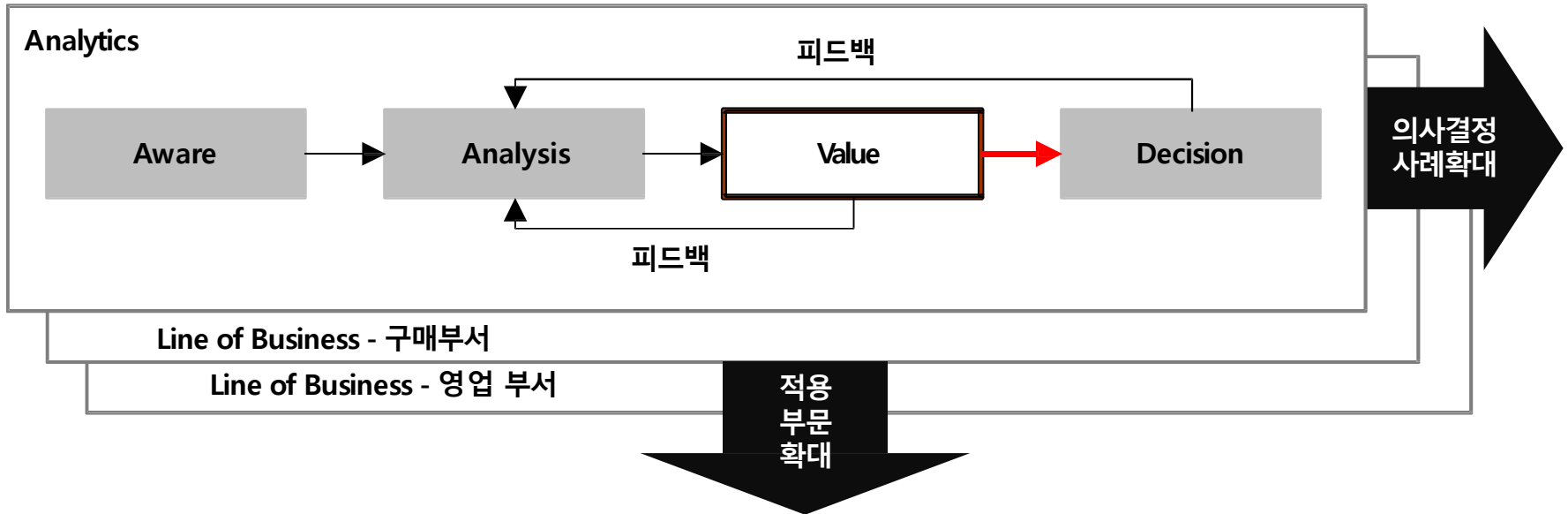
모든 것을 빠짐없이 리포팅, 분석하는 노력보다 비즈니스 모델을 강화할 수 있는 핵심 분석을 발굴하고 집중하면 경쟁사가 복제할 수 없는 핵심 경쟁력을 가질 수 있다.

- 내재화된 분석 업무 프로세스 : 5%의 직원 vs 95%의 직원

분석을 업무처리 프로세스 안에 내재화하여 분석결과에 따른 대응 액션을 즉각적으로 실행도록 업무 프로세스를 재설계해야한다..

핵심 성공 요소

▶ 데이터 분석 도입의 성공요소



볼보는 소비자의 운전과정에서 수집된 데이터를 본사 분석시스템에 자동 전송하도록 하여 빅데이터를 축적.

- 이를 통해 제품개발 단계에서 알기 어려운 다양한 결함과 소비자의 잠재 니즈를 빠르게 파악하여 대응
- 기존에는 50만 대의 차가 팔린 뒤에야 제기되었을 결함을 1,000대 판매 시점에서 포착하여 사후 관리 비용을 크게 절감시킴

| 1) 분석기회의 발굴

▶ 분석기회란?

분석 기회란 비즈니스 목표 달성을 위해 기업의 전사 또는 개별 업무별 주요 의사결정 포인트에 활용하는 분석의 후보들이다

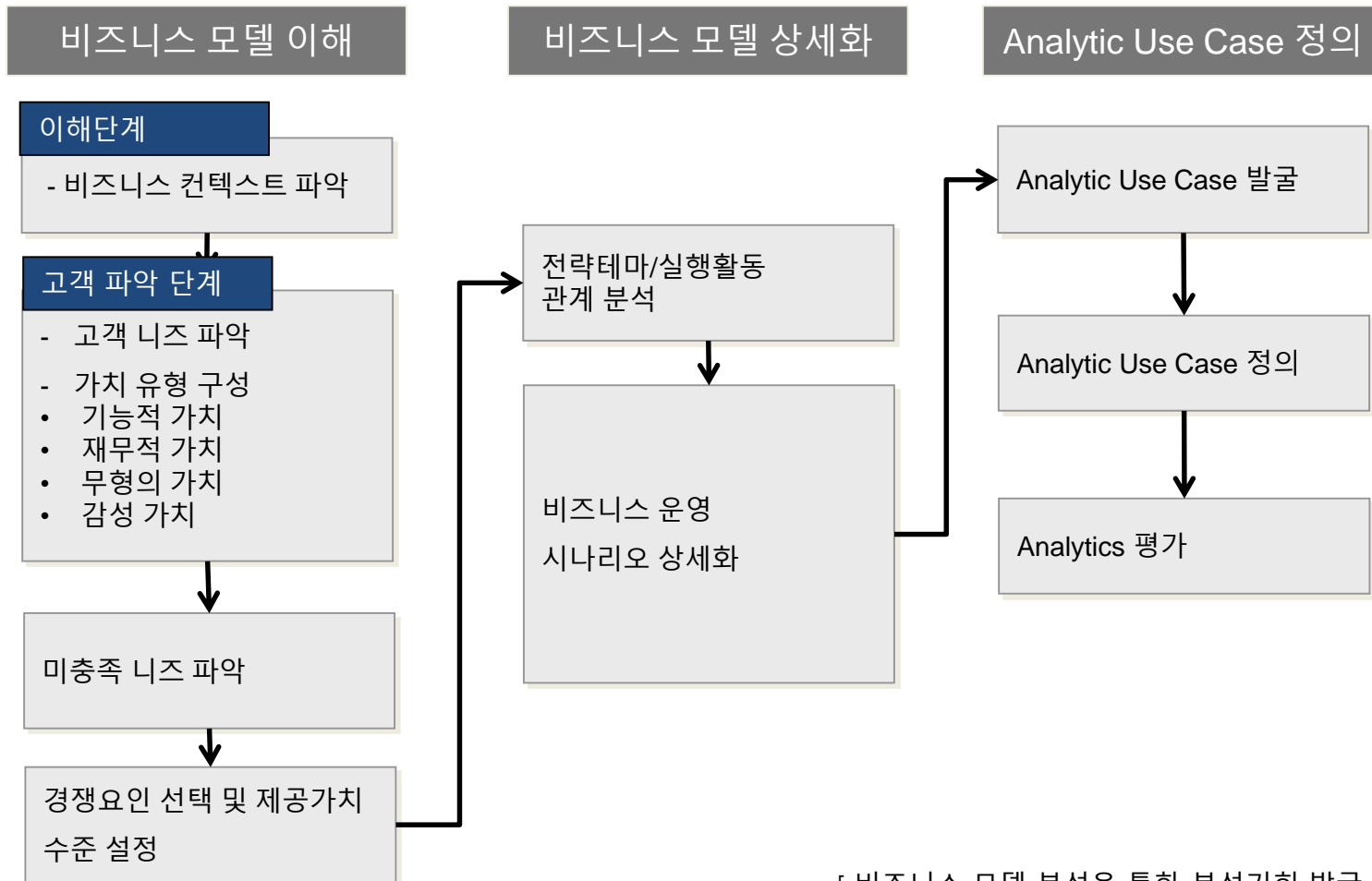
▶ 분석기회 발굴

분석 기회를 식별하는 활동이며, 데이터 분석 업무 프로세스에 도입하기 위한 접근 방식은 크게 3가지가 있다.

- Top Down Approach
 - 전사비즈니스 모델로부터 기업의 경쟁력 향상을 도모할 수 있는 분석 기회를 발굴하고 전사분석 체계 구현을 위한 조직, 인력, 기술 기반 거버넌스 체계 전반을 다루는 방식
- Path-Finding Approach
 - 업무 프로세스 레벨에서 특정 주제의 분석 기회를 발굴하고 주제별 분석 기회를 구현해 보는 Bottom Up 방식
- Bench-marking Approach
 - 산업별, 업무 서비스 별 분석테마 후보 풀의 벤치마킹을 통해 분석기회를 식별하는 방식

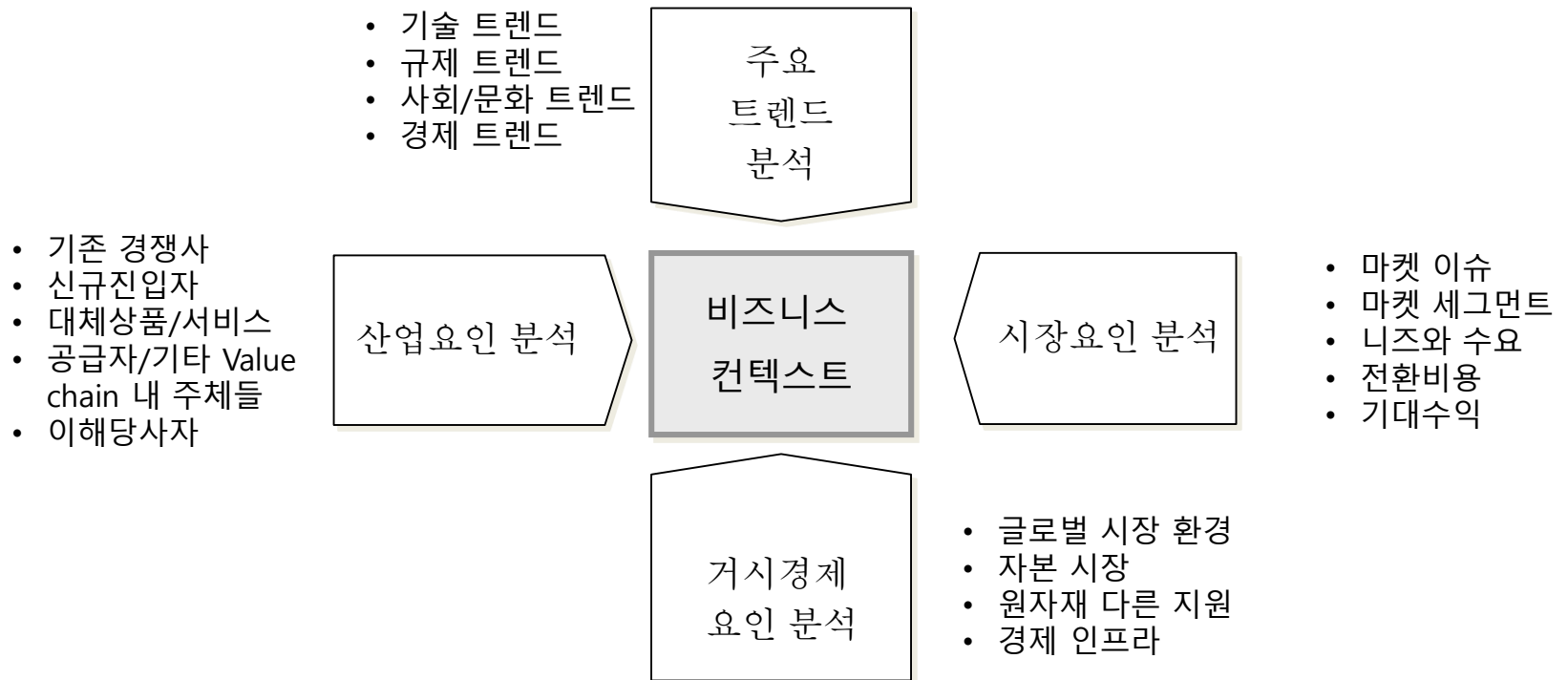
| Top-down approach

▶ 비즈니스 모델 분석



[비즈니스 모델 분석을 통한 분석기회 발굴 절차]

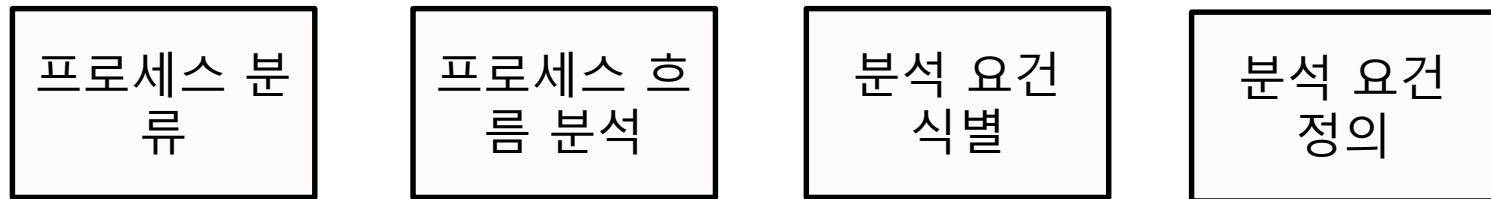
▶▶ 비즈니스 컨텍스트



| Bottom Up approach

▶▶ 프로세스 분류 체계에서 가치 발굴

- 프로세스 흐름 분석 : 프로세스별로 프로세스 맵을 통해 업무 흐름을 상세히 표현.
- 분석 요건 식별 : 각 프로세스 맵상의 주요 의사결정 포인트를 식별
- 분석 요건 정의 : 각 의사결정 시점에 무엇을 알아야 의사결정을 할 수 있는지, 즉 분석의 요건을 정의한다.



▶▶ 분석 유즈 케이스

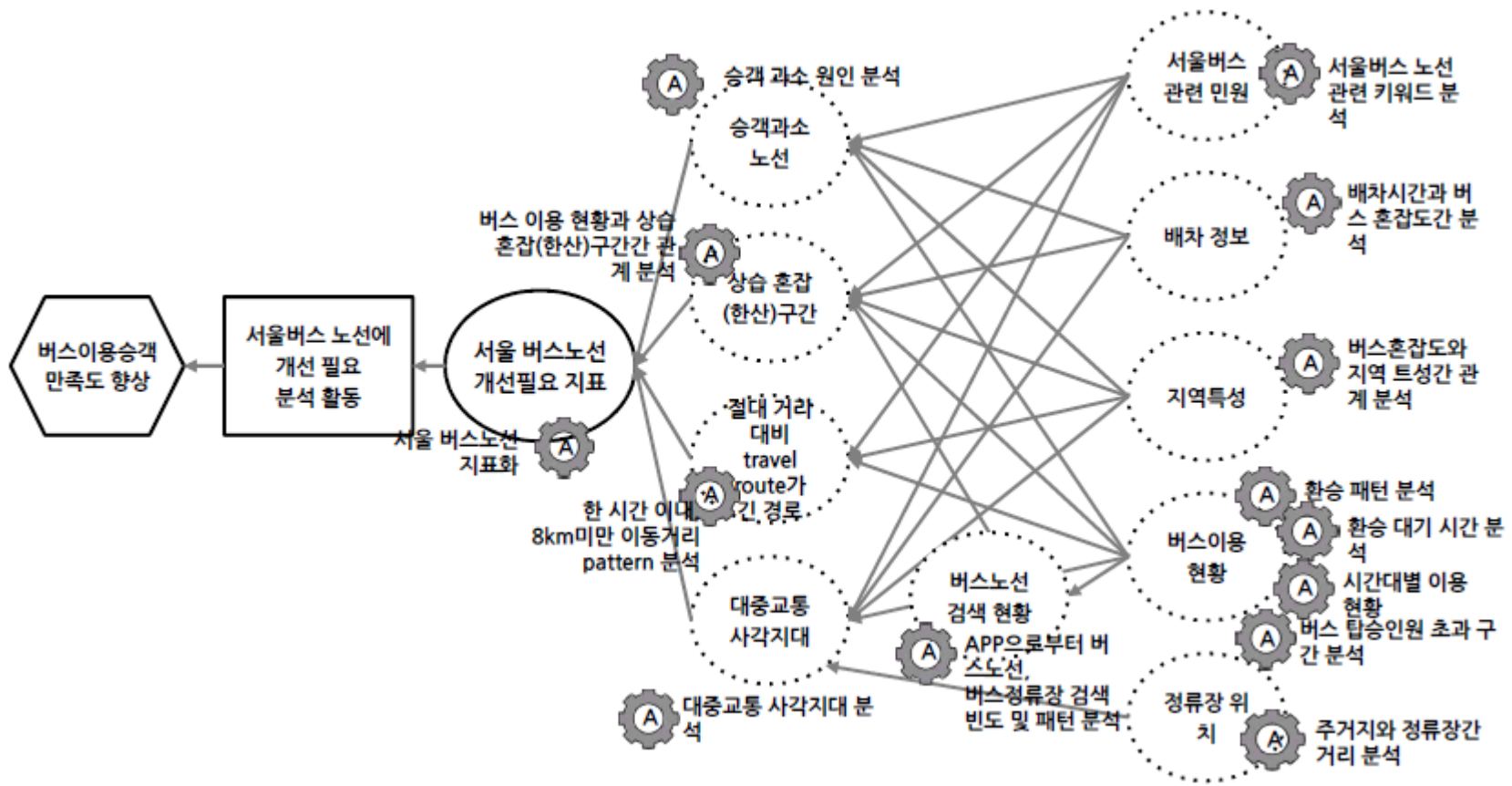
- 서비스 별, 부문 별, 직무 별 사례 분석/ 벤치마킹/ 브레인스토밍으로 도출
- 산업 및 업종을 불문하고 데이터 분석사례를 기반으로 한 분석테마 후보 그룹을 미리 정의할 수 있다.

개선과제 도출

Mega Process	Process Chain	Process	개선과제	실행과제	
싸이월드 커뮤니티	전체				
	1.마케팅				
		싸이월드 마케팅전략	예산부문의 독립채산제가 필요 대외단일창구 마련(대외접촉Rule 작성 및 운영 방안) - 본부과제	대외접촉 Rule 작성	
				Rule에 따라 하도록 공지 관련 톨이 있음(nae.com만 있음)-이 톨의 활용	외부업체
				SKT(딜을 함)	SKT(딜을 함)
				내부분부는 규칙을 지킴(잡무 로 인식)	내부분부는 규칙을 지킴(잡무 로 인식)
			전결규정관련 본부 지원담당 선정		
		Co-Promotion		12	
		광고상품기획 및 관리	싸이월드 광고탑 기능 신설	마케팅, 외부R & R 정립 기여매출로 반반 잡힘	
			업무범위 재정립(광고탑은 배너만 수행, 프로모 션은 수행말고) - 실적평가 재평가(브랜드 미니 홈피도 마케팅실적으로 안잡힘)	마케팅, 외부R & R 정립	
				실적 재평가	
				싸이월드 가치제고 기여방안	
		C/S	SKT 담당자(황승준대리) 선정하여 해결중		
		기타	디자인 담당 인력충원		
	2.운영				
		운영정책기획	인력충원	운영정책기획 지침 수립방안 ->인력충원되어 이지영대리 가 담당으로 해결	
			작성, 실행, 통제로 나누어짐 R & R 정의	법무팀 피드백 개선 요청(정 리해서 요청)	
				벤치마크	
				운영정책기획 방안 (진행중-신동하대리)	
		위기관리	매뉴얼 세분화 방안		
		CS관리	CS관리 지표 개선		
			센터통제방안 수립	커뮤니케이션 제도 보완	
			센터 통제 강화		
			센터 매뉴얼 작성 유도 및 이에 대한 일정관리 필요		
		공문처리	공문처리 프로세스 개선	운영파트에 전담인력 배치 법무팀에 전담인력 배치	
		기타(정산)	정산 프로세스 확인		
			외부와 커뮤니케이션 힘듦	12 처리	
				1.1	
			채용프로세스 개선		
			직군, 직무 정의		

2) 분석기회의 모형화

분석 기회가 발굴되면 분석목표를 명확히 한 후, 구체적인 분석질문을 찾아내어 의사결정 요소를 모형화 하는 방식으로 구체화하는 단계가 필요



분석별로 현재 기업(기관)내에 보유하고 있는 데이터와 외부에서 확보해야 할 데이터를 정의 데이터 확보의 가능성과 비용요인을 고려하여 분석의 우선순위와 범위를 조정

분석명	필요 정보(데이터)	
	서울시 보유 데이터	외부 데이터
서울버스 노선 관련 민원분석	<ul style="list-style-type: none"> • 다산콜센터 민원정보 • 홈페이지 민원정보 	<ul style="list-style-type: none"> • 소셜 정보
배차시간과 버스 혼잡도간 분석	<ul style="list-style-type: none"> • 서울버스 배차 정보 • 동일버스 시간대별 실 운행 간격 정보 • 시간대별 버스이용 승/하차인원 수 정보 	
버스혼잡도와 지역 특성간 관계 분석	<ul style="list-style-type: none"> • 버스/정류소별 승/하차 인원 수 정보 • 정류소/시간대별 승/하차 인원 수 정보 • 정류소 요일별 승/하차 인원 수 정보 	<ul style="list-style-type: none"> • 정류소별 지역 특성 정보
APP으로부터 버스노선, 버스정류장 검색 빈도 및 패턴 분석	<ul style="list-style-type: none"> • 버스 앱 검색 정보 	<ul style="list-style-type: none"> • 길찾기 앱 검색 정보 (민간기업)
버스 이용 현황과 상습혼잡(한산) 구간간 관계 분석	<ul style="list-style-type: none"> • 버스 정류소별 승하차 인원 수 분석 	
버스노선내 지하철 존재가 버스혼잡도에 미치는 영향 분석	<ul style="list-style-type: none"> • 버스 노선정보 • 지하철 노선정보 • 지하철역 인근 버스정류소를 지나는 • 버스의 승하차 인원 수 정보 	
대중교통 사각지대	<ul style="list-style-type: none"> • 버스정류장 주소 정보 • 마을버스 노선 정보 • 마을버스 승하차 인원 수 정보 	<ul style="list-style-type: none"> • 주소정보 (안전행정부)
승객 과소 원인 분석	<ul style="list-style-type: none"> • 버스별 정류소별 승하차 인원 수 분석 	

| 3) 분석 시나리오/ 분석 정의서 작성

▶ 분석 시나리오 정의

- 분석 컨텍스트를 기반으로 도출된 분석 체계를 종합적으로 고려
- 분석으로 인한 프로세스 지능화는 기업의 업무 프로세스 재설계를 동반.

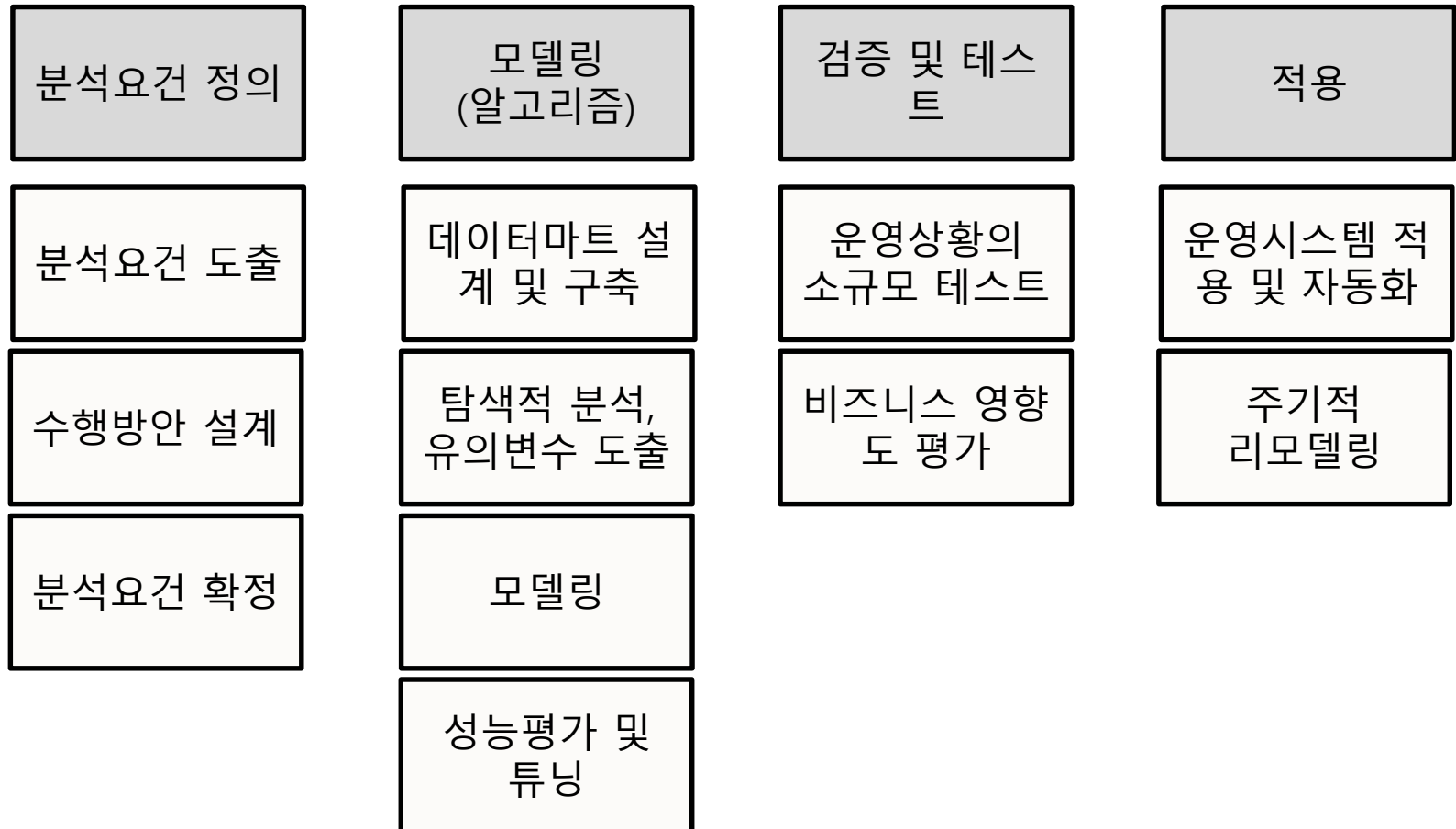
▶ 분석 정의서

- 분석 별로 필요한 소스 데이터 확보/분석 방법, 데이터 입수 및 분석의 난이도,
- 분석 수행 주기, 분석 결과에 대한 검증 오너십 등 상세 분석 과정을 정의.

분석명		분석정의	
해지상담 접촉패턴 분석		기 해지 해약건 발생 고객의 해지 시점 상담정보 분석을 통해 해지 고객의 상담 특성을 발굴	
소스 데이터	데이터 입수 난이도	분석방법	
접촉 채널/건수, 평균 접촉 시간, 최종접촉이후 해지까지 시간, 상담인력 업무 능숙도	하	해지로 이어지는 해지 상담의 유의미한 속성을 요인분석을 통해 발굴하고, 클러스터링을 통해 영향 요인을 그룹핑하고, 그룹 요인이 해지에 미치는 영향도를 회귀분석함	
	데이터 입수 사유		
	N/A		
분석 적용 난이도	분석적용 난이 사유	분석 주기	분석결과 검증 Owner
중	비구조적 데이터(접촉 로그 등) 분석 필요	월별 업데이트	해지방어팀

| 4) 분석 프로세스 설계

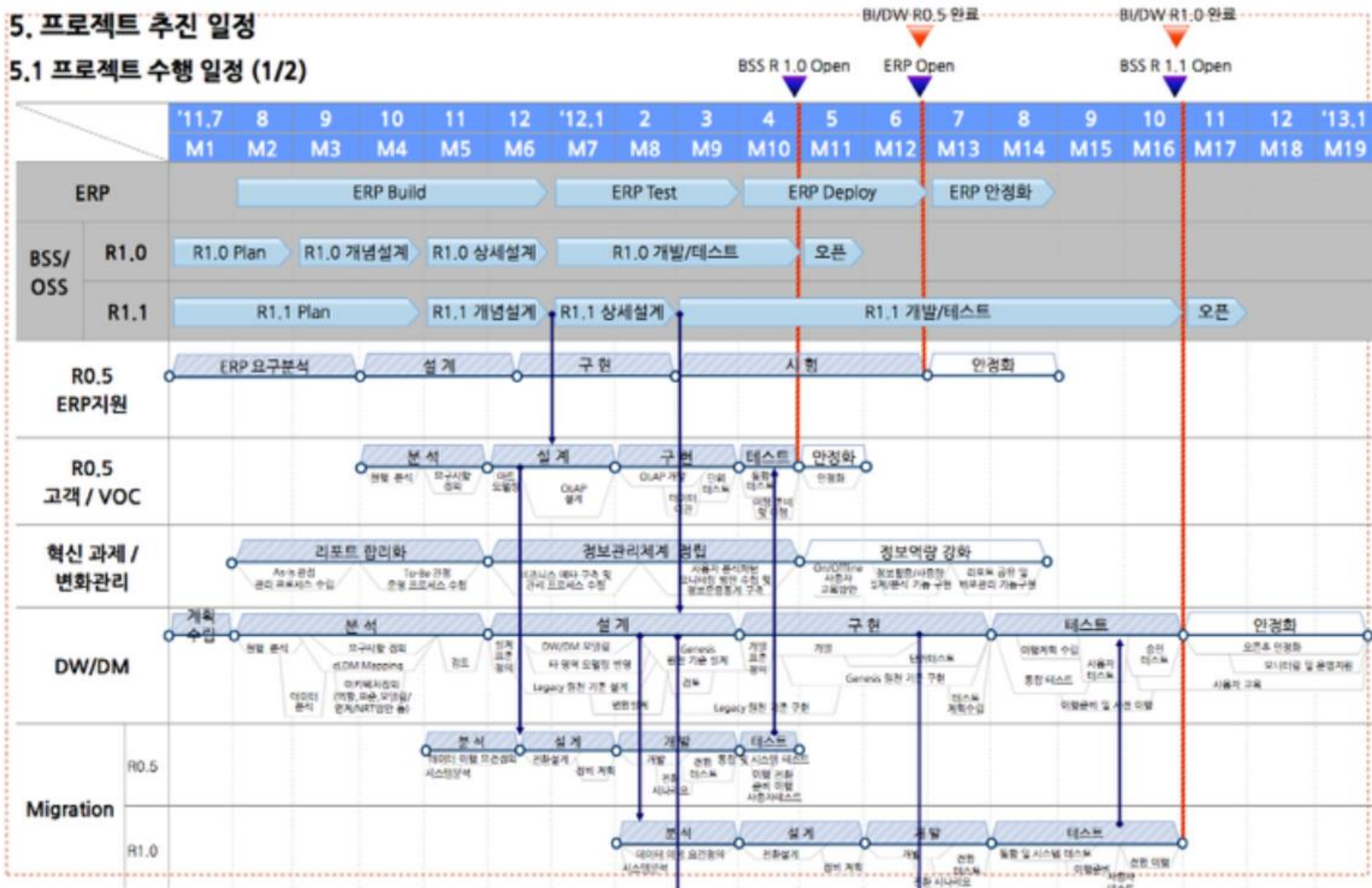
▶ 데이터 분석 프로세스



▶ 수행방안 설계

5. 프로젝트 추진 일정

5.1 프로젝트 수행 일정 (1/2)



▶▶ 모델링 마트 설계와 구축

■ 수행 준거

- 다양한 원천 데이터로부터 분석 대상 데이터를 획득
- 획득된 데이터를 탐색, 정제, 요약하여 변수 식별
- 분석 대상 데이터를 구조화하여 모델 마트 설계
- 설계를 기초하여 모델 마트 구축

■ 고려사항

- 데이터 원천 : RDBMS, DW, 시스템 로그, 비정형 텍스트 데이터 등
- 데이터 유형
 - 연속형 : 연속되는 실수로 구성
 - 범주형 : 수치형, 텍스트형, 명목형 변수, 순위형 변수

■ 기존 정보시스템 내의 데이터를 최대한 활용, 확장

■ 데이터 정제 단계

- 데이터 요약 -> 파생 변수 도출 -> 변수 확대
- 유의미한 변수들로 분석 모형 구축

▶▶ 탐색적 분석과 유의 변수 도출

- 데이터 마이닝에 해당하는 업무
- EDA(Exploratory Data Analysis) : 탐구 데이터 분석
 - 비즈니스 이해와 분석요건에 대한 팩트를 발견해 통찰을 얻는 과정
- 유의미한 변수 파악
 - 목표값 별로 변수값의 분포도 분석
- 결과 공유 단계에서 시각화 표현은 관련 부서의 이해를 도울 수 있다
 - 팩트와 통찰을 전달할 수 있는 것에 중점
- 수행 준거
 - 분석 목적과 요건, 데이터 특성을 기반으로 적합한 데이터 분석기법 선정
 - 선정된 분석기법을 기준으로 분석 모형 설계
 - 설계한 분석 모형을 기준으로 유의성 분석 -> 유의미한 변수 식별

▶▶ 모델링

- 데이터 마이닝, 시뮬레이션, 최적화 기법 등을 이용하여 분석 모델을 구축하는 단계
- 수행 준거
 - 다양한 모델링 기법을 다뤄 업무 특성에 맞는 기법을 선택, 결합 적용할 수 있어야 한다
 - 선택된 모델링 기법을 이용하여 모델링
 - 데이터 모델링 : 미래값 예측 시 프로세스적인 측면이 없을 경우
 - 시뮬레이션 기법 : 프로세스 및 자원에 대한 제약, 입력값이 확률분포를 갖는 경우
 - 최적화 기법 : 프로세스 및 자원에 대한 제약, 입력값이 상수값을 가질 때
- 고려사항
 - 데이터 마이닝 모델링은 지나치게 통계적 가설이나 유의성에 집착하지 않는다
 - 충분한 시간이 있을 때 다양한 옵션을 줘서 시도
 - 테스트 데이터를 20~40% 섞어서 실시하여 예상 성능을 만족하면 중단
 - 과도한 성능에 대한 집착을 버린다 (주목적 실무 적용, 일정 준수)

▶▶ 비즈니스 영향도 평가

- 재무 효과 증명 가능해야 기업의 입장에서 의미 있음
- 비즈니스 영향도 산출 : 업무 전문가의 도움이 절대적
- 투자 대비 효과 정량화 기법
 - 총 소유 비용 (TCO, Total Cost of Ownership)
 - 투자대비효과 (ROI, Return On Investment)
 - 순현재가치 (NPV, Net Present Value)
 - 내부수익률 (IRR, Internal Rate of Return)
 - 투자회수기간 (PP, Payback Period)

▶▶ 운영 시스템에 적용과 자동화

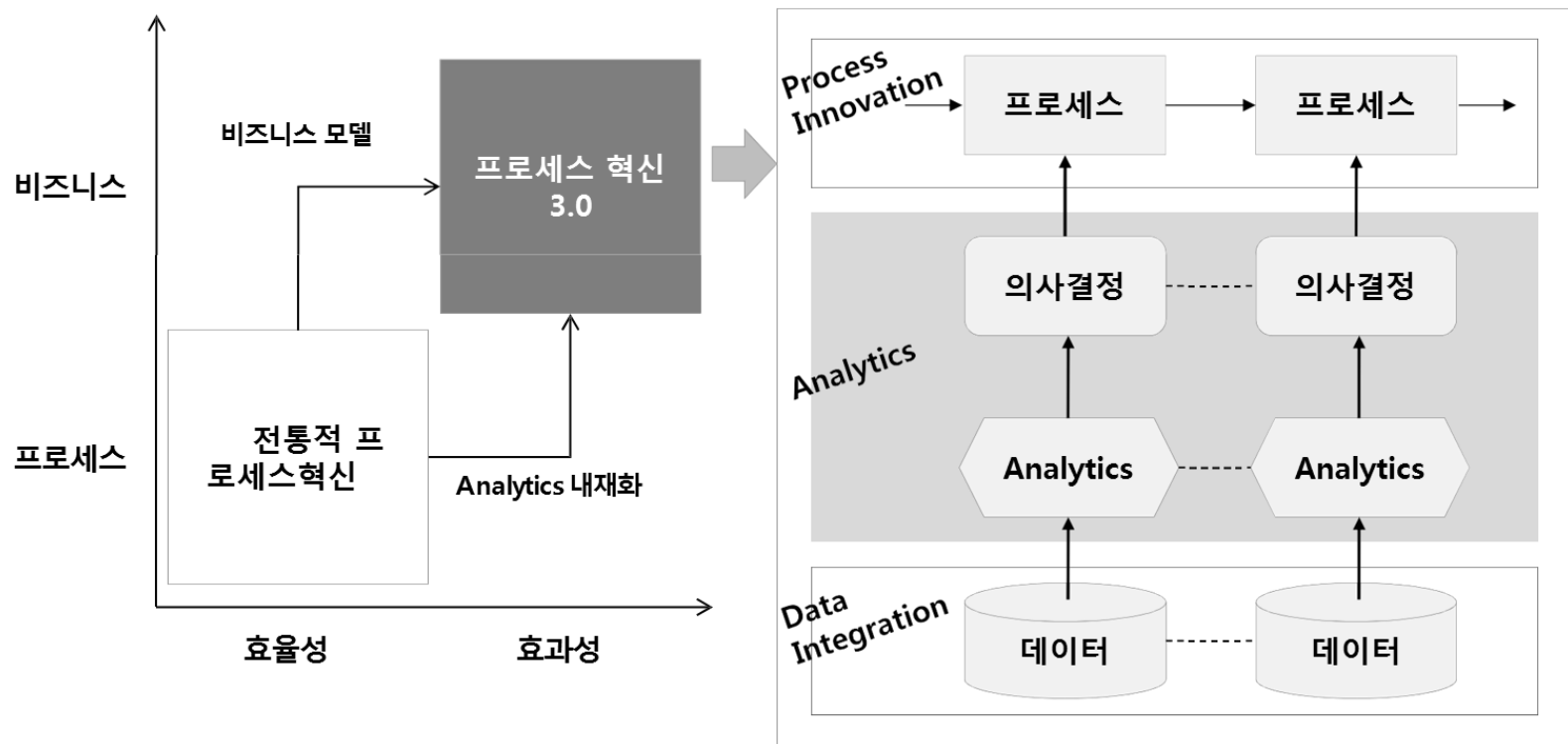
- 실시간 or Batch Scheduler에 적용
 - 주기별 모니터링
 - 이상 현상 발생 시 확인 Process 수립
 - 정기적인 분석 모형 재평가 실시
- 수행 준거
 - 업무 적용에 따른 기존 업무와 개선 기회 분석
 - 개선 기회를 바탕으로 목표 업무 프로세스 설계

| 5) 전사관점 고려사항

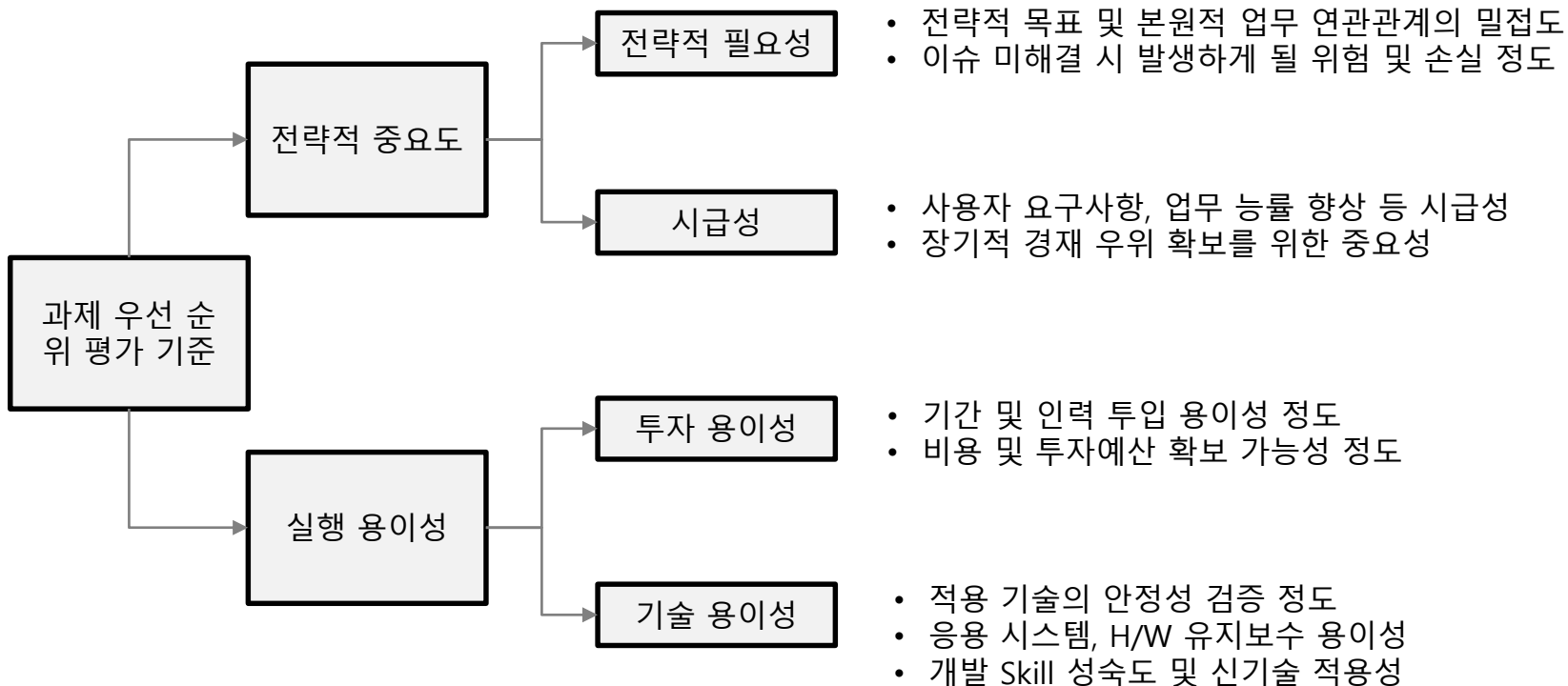
▶ 분석의 선순환 구조

- 특정 부서에서 수행한 분석 결과는 다른 조직의 분석에도 영향을 미친다.
- 개별 분석 결과에 따른 의사결정이 조직별로 상충 여부를 파악하고, 최적화한다.

▶ 분석 내재화 프로세스 정의



▶ 전사 최적화



빅데이터의 특징을 고려한 분석 ROI 요소

투자 비용 요소 (Investment) : Volume (데이터 규모/양), Variety (데이터 종류/유형), Velocity (데이터 생성/처리 속도)
 비즈니스 효과 (Return) : Value (분석 결과 활용 및 실행을 통한 비즈니스 가치)

3. 데이터 분석 실습

- ▶▶ 팀 구성
- ▶▶ 분석 정의서 작성
- ▶▶ 보고서 작성
- ▶▶ 발표

구성원	전화번호	이메일	참여동기	관심분야

| 분석 정의서 작성

- ▶▶ 제공하는 ppt 파일에 작성

| 보고서 작성 및 발표

▶▶ 보고서는 다음 사항을 유의하여 작성

- 분석 목표
- 분석 결과와 분석 목표이 일치성
- 분석 개념도
- 분석 목표를 위한 소요 데이터
- 데이터의 소스 및 입수에 있어서의 문제점
 - 필요 데이터 입수가 불가능할 경우 대체 데이터 확보 노력도
- 데이터의 전처리
- 분석 모형
- 평가 지표
- 개선 방안
- 팀원들의 소감