Module 14. 프로젝트 방법론

01

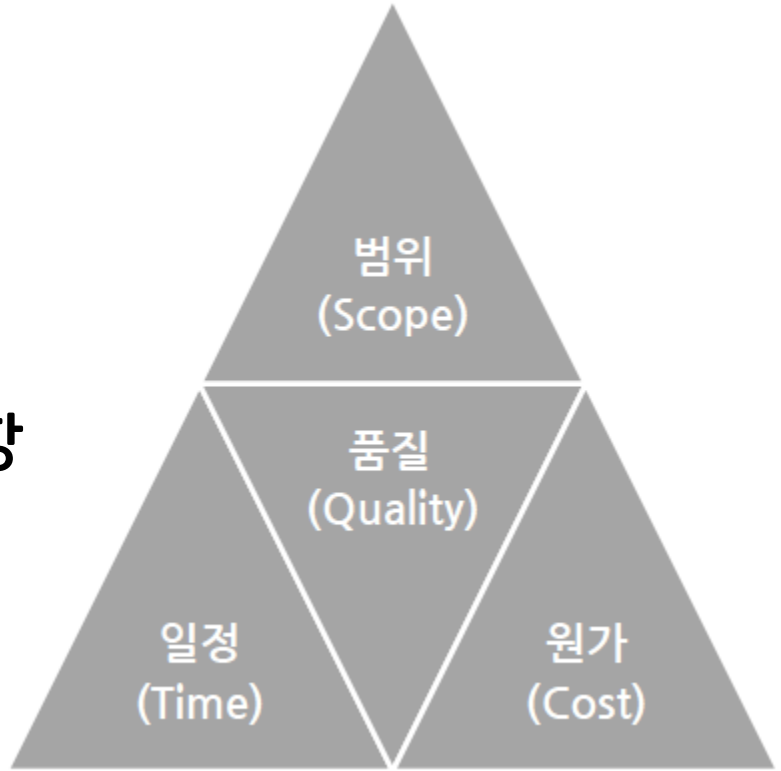
빅데이터 분석 프로젝트

경북대학교 배준현 교수
(joonion@knu.ac.kr)

프로젝트 관리란?

- 프로젝트 요구사항을 충족시키기 위하여
  - 지식, 기술, 도구, 기법 등을 프로젝트 활동에 적용하는 것

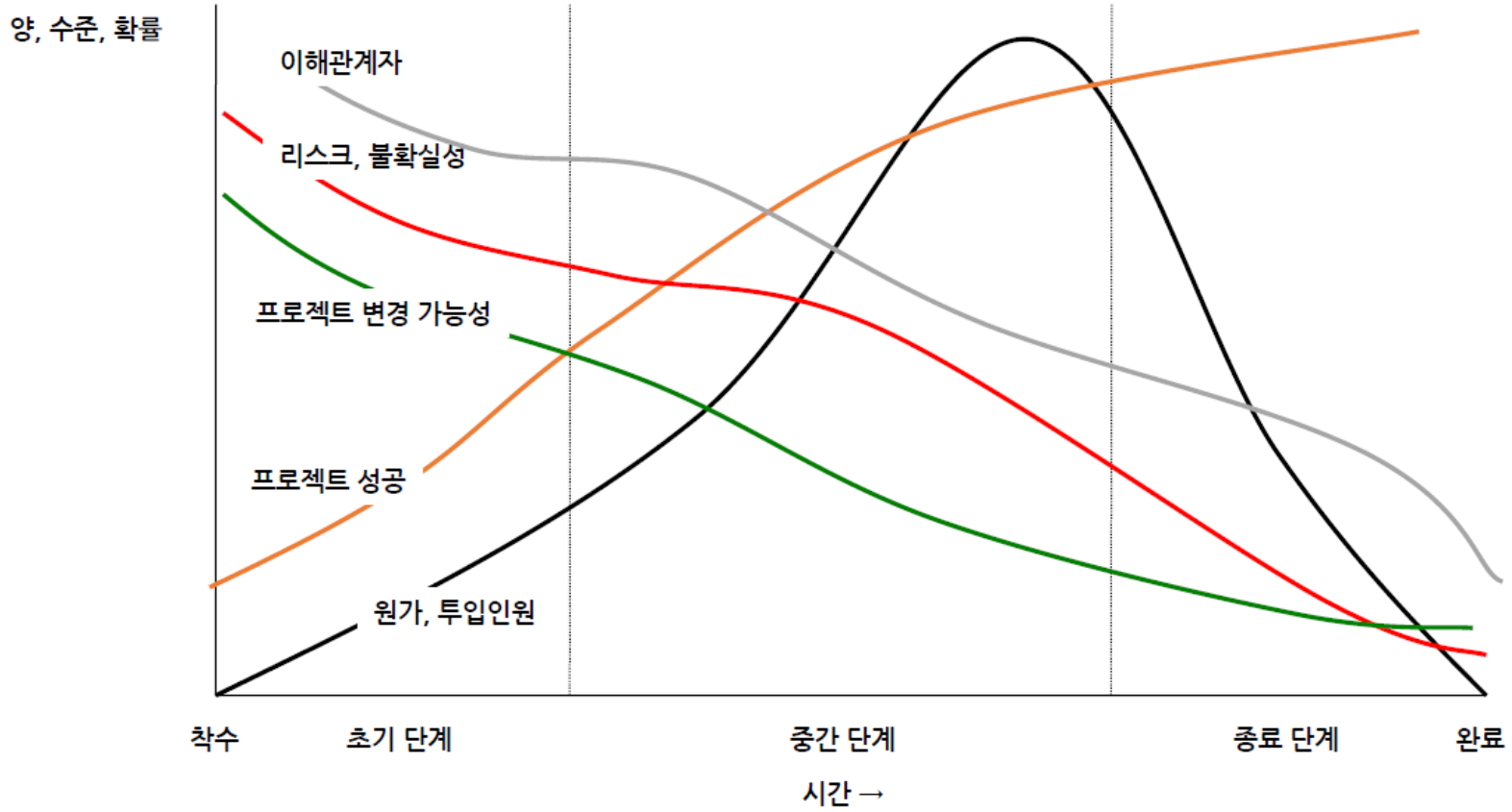
프로젝트 관리의 제약사항



- **프로젝트 관리 업무**
  - 요구사항의 식별
  - 프로젝트가 계획되고 실행됨에 따라 발생하는 이해관계자의
    - 다양한 요구사항, 관심사항, 기대사항의 처리 및 해결
  - 범위, 품질, 일정, 예산, 자원, RISK를 포함하여
    - 서로 경합하는 다양한 프로젝트의 제약사항들 사이에서의 균형 유지

프로젝트 라이프 사이클

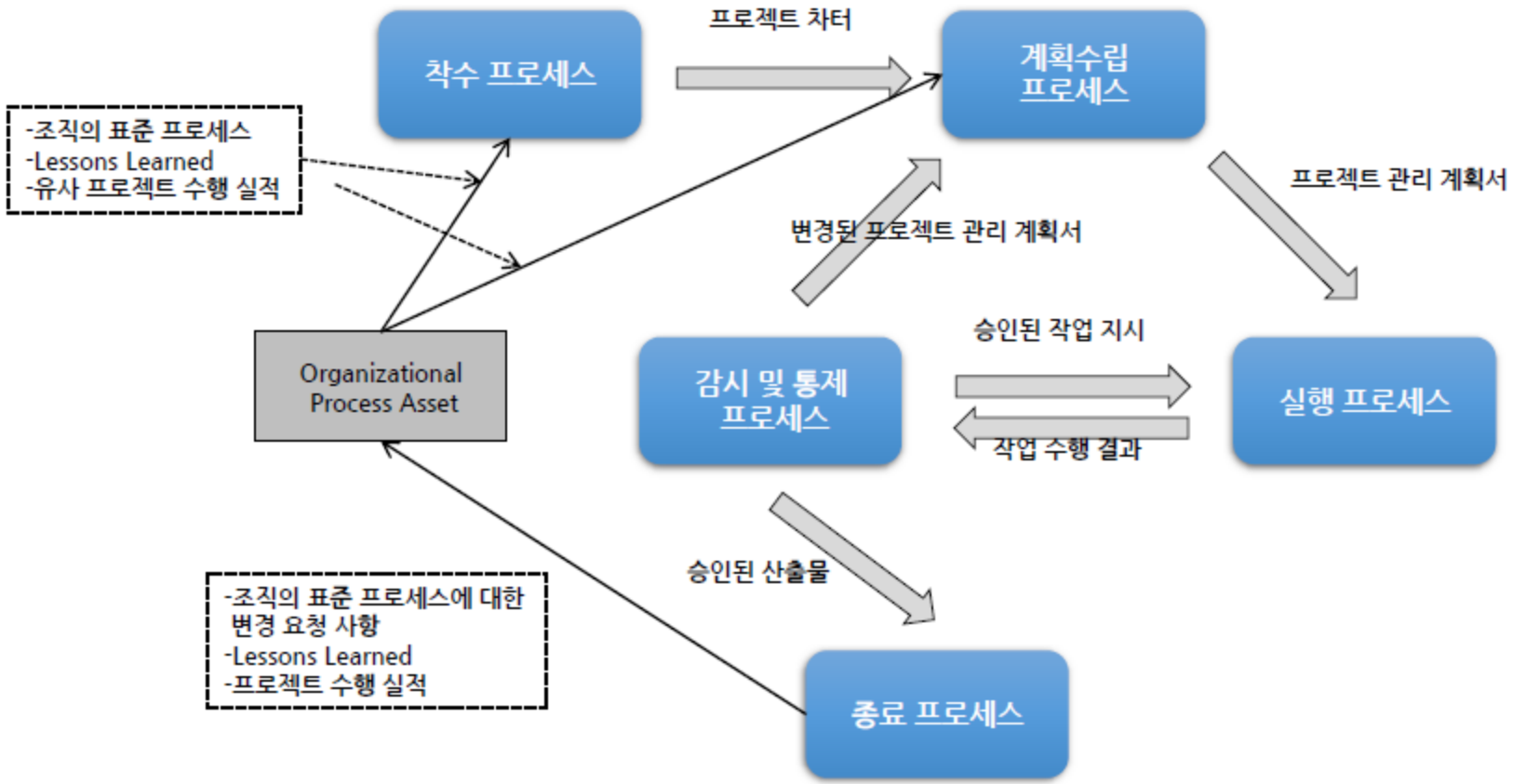
- 프로젝트는 착수일과 완료일이 정의된 라이프 사이클을 가진다.



▪ 프로젝트 관리 기능 및 세부 프로세스

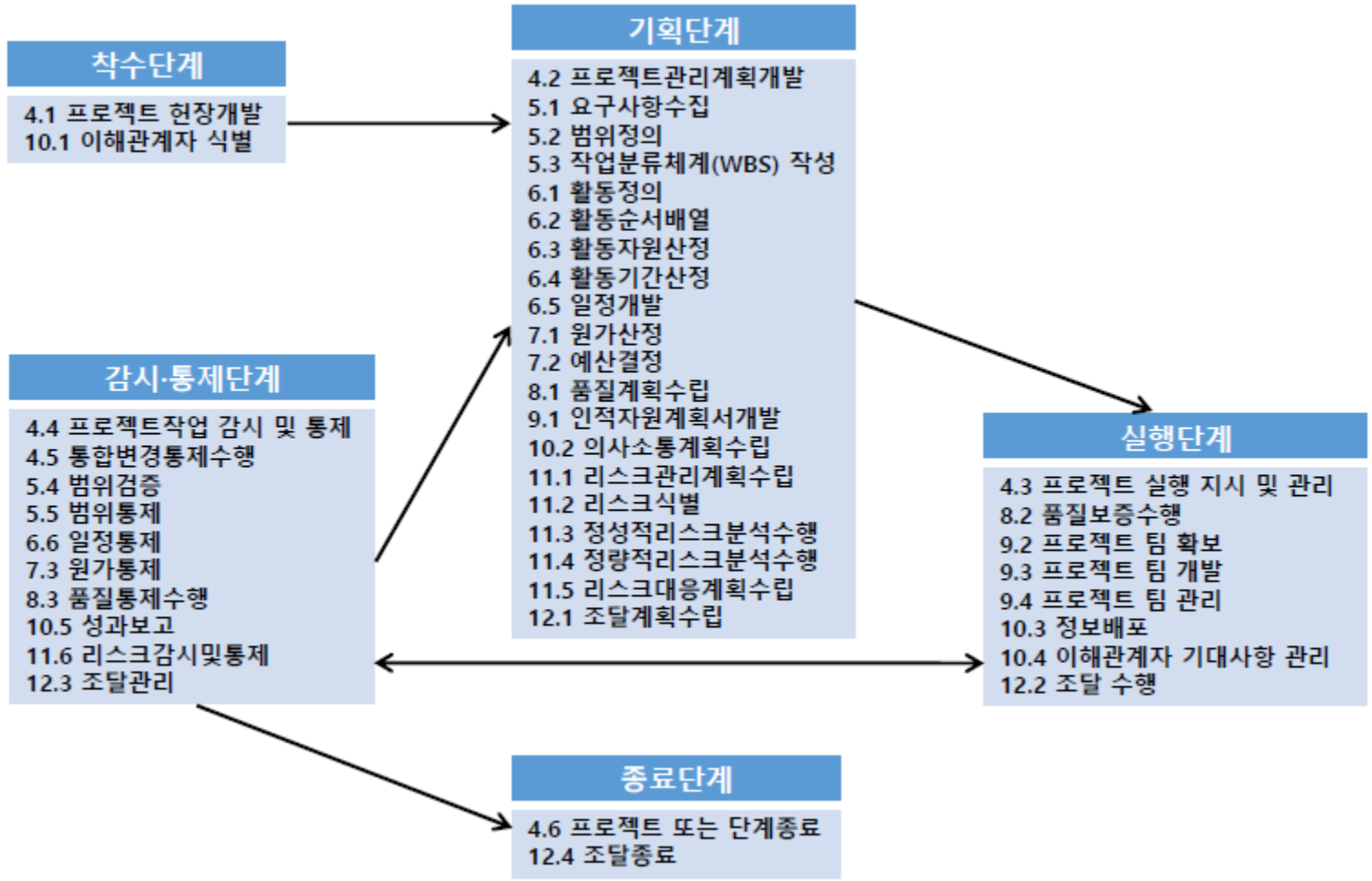
Project Scope Management	Project Schedule Management	Project Finacial Management
• 개시 • 계획 수립 • 범위 확정 • 범위 검증 • 범위 변경 컨트롤	• Activity 정의 • Activity 배열 • 작업 기간 산정 • 일정 개발 • 일정 컨트롤	• 자원 계획 수립 • 원가 산정 • 원가 예산 수립 • 원가 컨트롤
Project Quality Management	**Project Resource Management**	**Project Communications Management**
• 품질 계획 수립 • 품질 보증 • 품질 컨트롤	• 조직 계획 수립 • 인원 획득 • 팀 개발	• 의사소통 계획 수립 • 정보 배포 • 성과 보고 • 의사소통 종료
Project Risk Management	**Project Contract Management**	**Customer Relationship Management**
• 리스크 관리 계획 수립 • 리스크 식별 • 리스크 정성적 분석 • 리스크 정량적 분석 • 리스크 대처 계획 수립 • 리스크 모니터링 및 컨트롤	• 계약 관리 계획 수립 • 계약 협상 • 계약 체결 • 계약 이행 컨트롤 • 계약 종료	• 고객 정보 수집 • Partnership 구축 • Commitment 획득 • 관계 유지

- **프로젝트 관리는 프로세스 그룹 간의 상호관계로 진행된다.**



프로세스 그룹의 태스크

- 각 프로세스 그룹에서는 세부 태스크를 처리해야 한다.



▪ 프로젝트 관리 영역 내용 및 태스크

범위관리	원가관리	품질관리	자원관리	의사소통관리	리스크 관리
• 프로젝트를 성공적으로 완료하는데 반드시 필요한 작업만을 빠짐없이 프로젝트에 포함시키기 위해 필요한 프로세스 • 프로젝트 업무 범위 설정 및 승인을 받아 프로젝트 목표에 맞도록 관리하는 기능	• 승인된 예산 범위 내에서 프로젝트를 완료할 수 있도록 원가를 산정하고, 예산을 책정하고, 원가를 통제하는 프로세스	• 프로젝트가 요구사항을 충족할 수 있도록 품질정책, 품질목표, 품질 책임사항을 결정하는 프로세스	• 프로젝트 팀을 구성하고, 관리하고, 프로젝트 팀을 이끄는 프로세스 • 프로젝트 수행요원들의 활동을 조직하고 조정하는 기능	• 프로젝트 정보의 생성, 수집, 배포, 저장, 검색 그리고 최종처리가 적시에 적절히 수행되도록 하기 위해 필요한 프로세스 • 프로젝트의 이해관계자간에 효율적인 정보 전달체계를 계획, 조직, 관리하는 기능	• 프로젝트에 대한 리스크의 관리 기획, 식별, 분석, 대응 기획, 감시 및 통제를 수행하는 프로세스 • 프로젝트 수행과정에서 일어날 수 있는 리스크 요인을 발견하고 분석하여 대책을 수립하는 기능
• 요구사항 수집 • 범위 정의 • 작업분류체계 (WBS) 작성 • 범위 검증 • 범위 통제	• 원가 산정 • 예산 결정 • 원가 통제	• 품질 계획 수립 • 품질 보증 수행 • 품질 통제 수행	• 자원 계획서 개발 • 프로젝트 팀 확보 • 프로젝트 팀 개발 • 프로젝트 팀 관리	• 이해관계자 식별 • 의사소통계획 수립 • 정보배포 • 이해관계자 기대사항 관리 • 성과보고	• 리스크 관리 계획 수립 • 리스크 식별 • 정성적 리스크 분석 수행 • 정량적 리스트 분석 수행 • 리스크 대응 계획 수립 • 리스크 감시 및 통제

소프트웨어 개발 방법론: *methodology*

- 소프트웨어 개발 방법론이란
  - 소프트웨어를 만들 때 어떻게 만들어야 하는 지를 정의하는 방법

작업절차
- 소프트웨어를 개발할 때 진행하는 작업의 순서

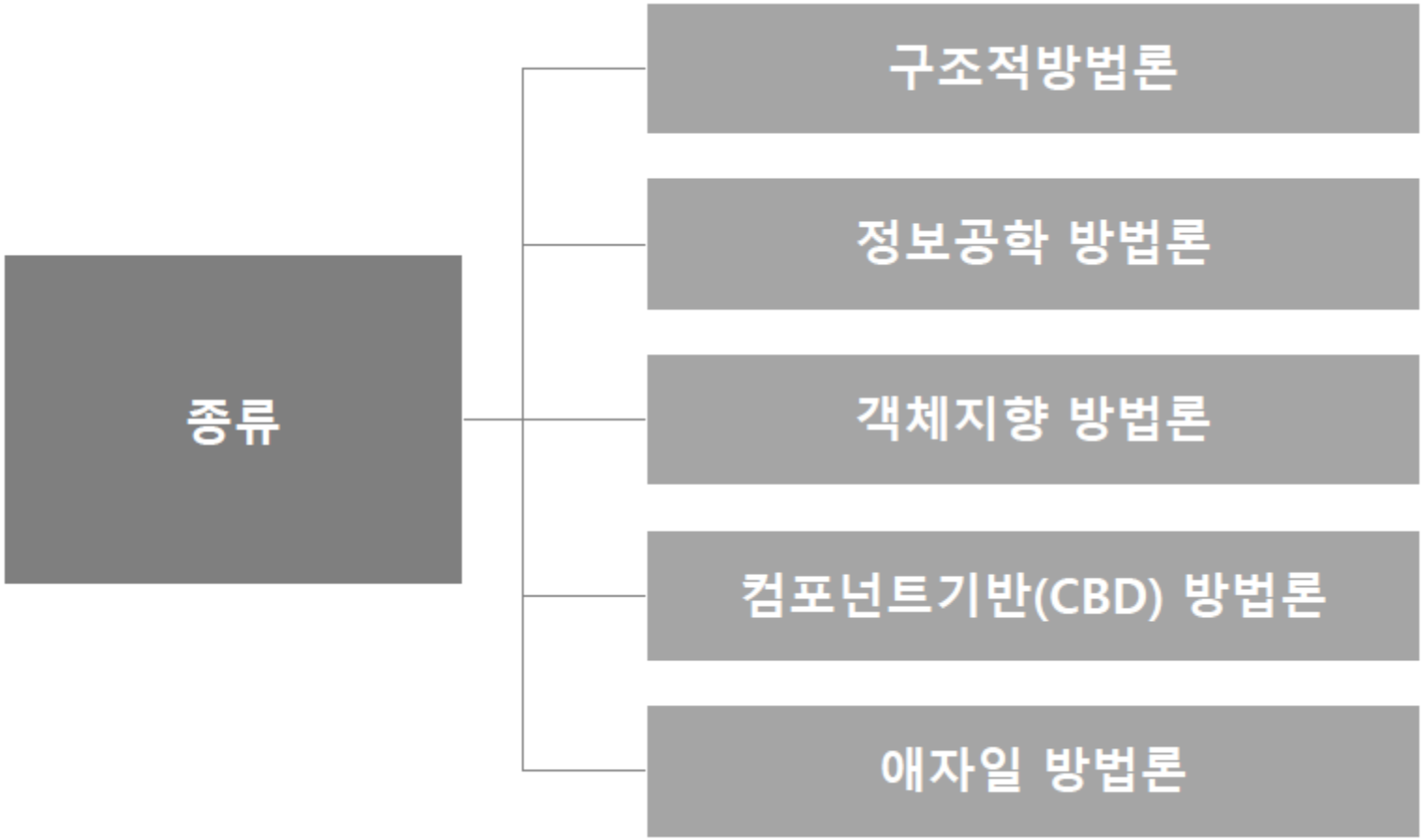
작업방법
- 각 단계 마다 수행해야 할 일(Task, 누가 언제, 무엇을)

산출물
- 단계별로 만들어지는 문서(목록, 설계서, 정의서, 명세서)

관리
- 개발 진행을 어떻게 관리하고 제어할 것인가에 대한 방법

기법
- 단계별 일을 진행할 때 사용하는 기술 이나 기법(DFD, ERD, User Case)

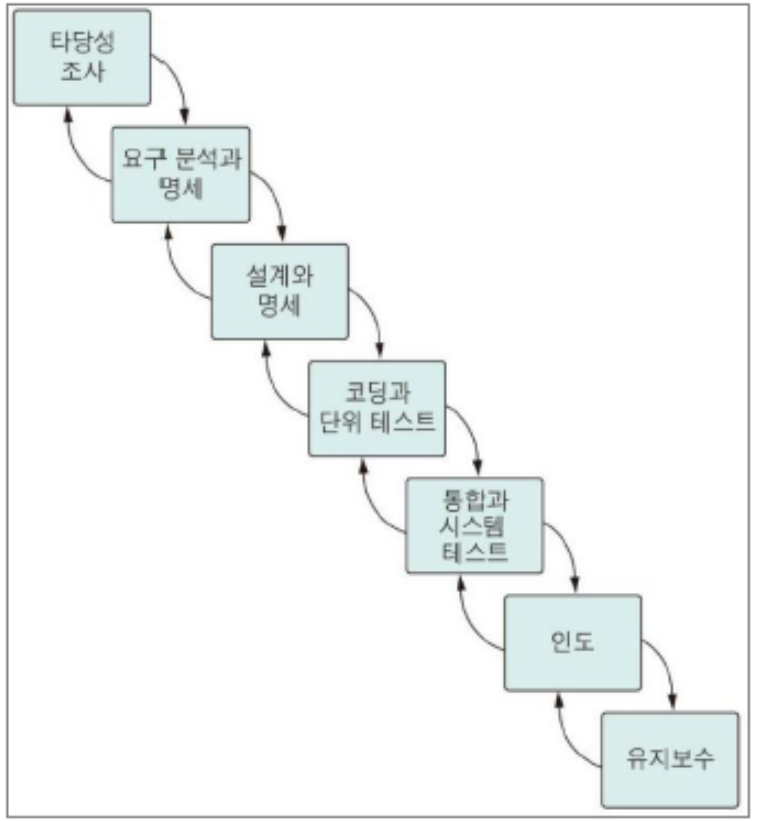
도구
- 사용되는 기법별 지원도구(Power Point, Excel, ERWin, Git, Draw IO)


종류
구조적방법론
정보공학 방법론
객체지향 방법론
컴포넌트기반(CBD) 방법론
애자일 방법론

▪ 구조적 방법론

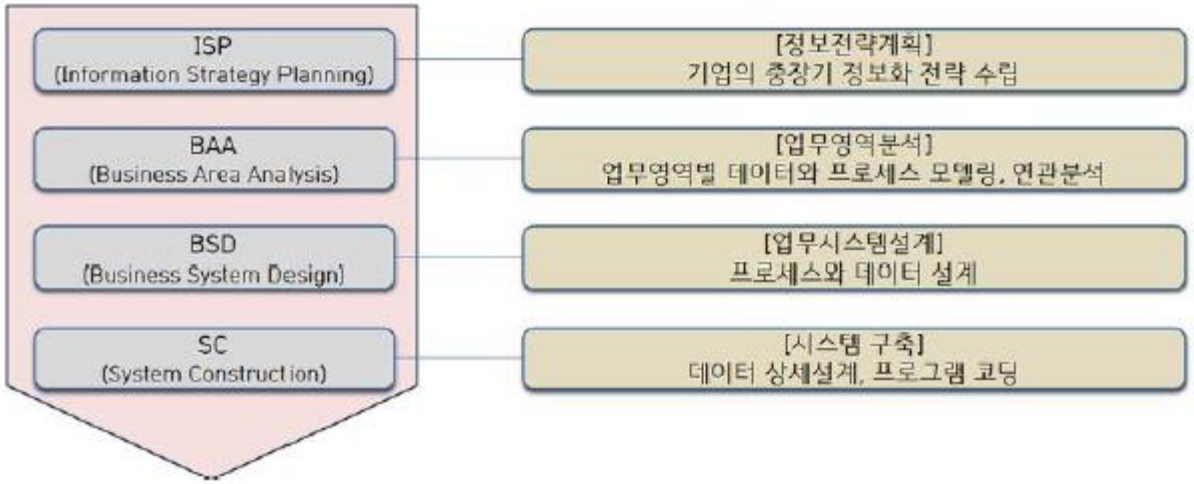
구조적방법론 개념도

폭포수 모델 (Waterfall Model)

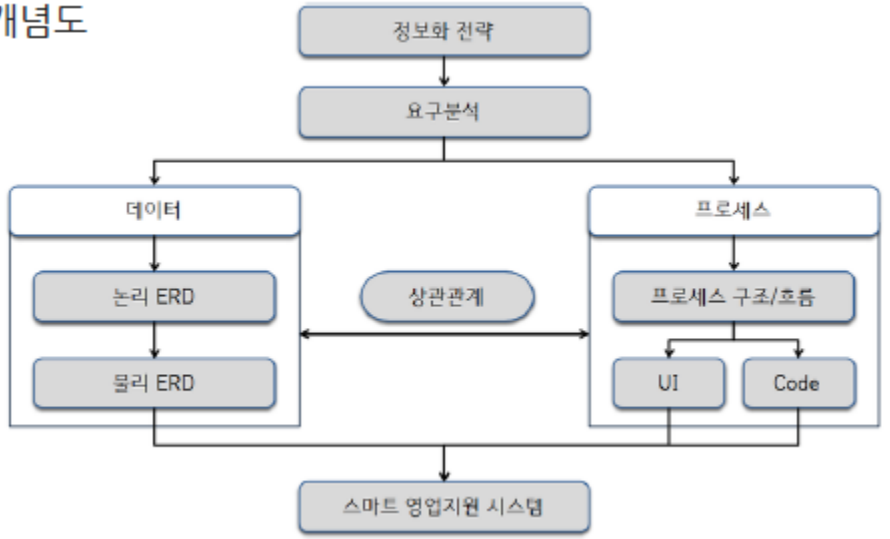


▪ 정보공학 방법론

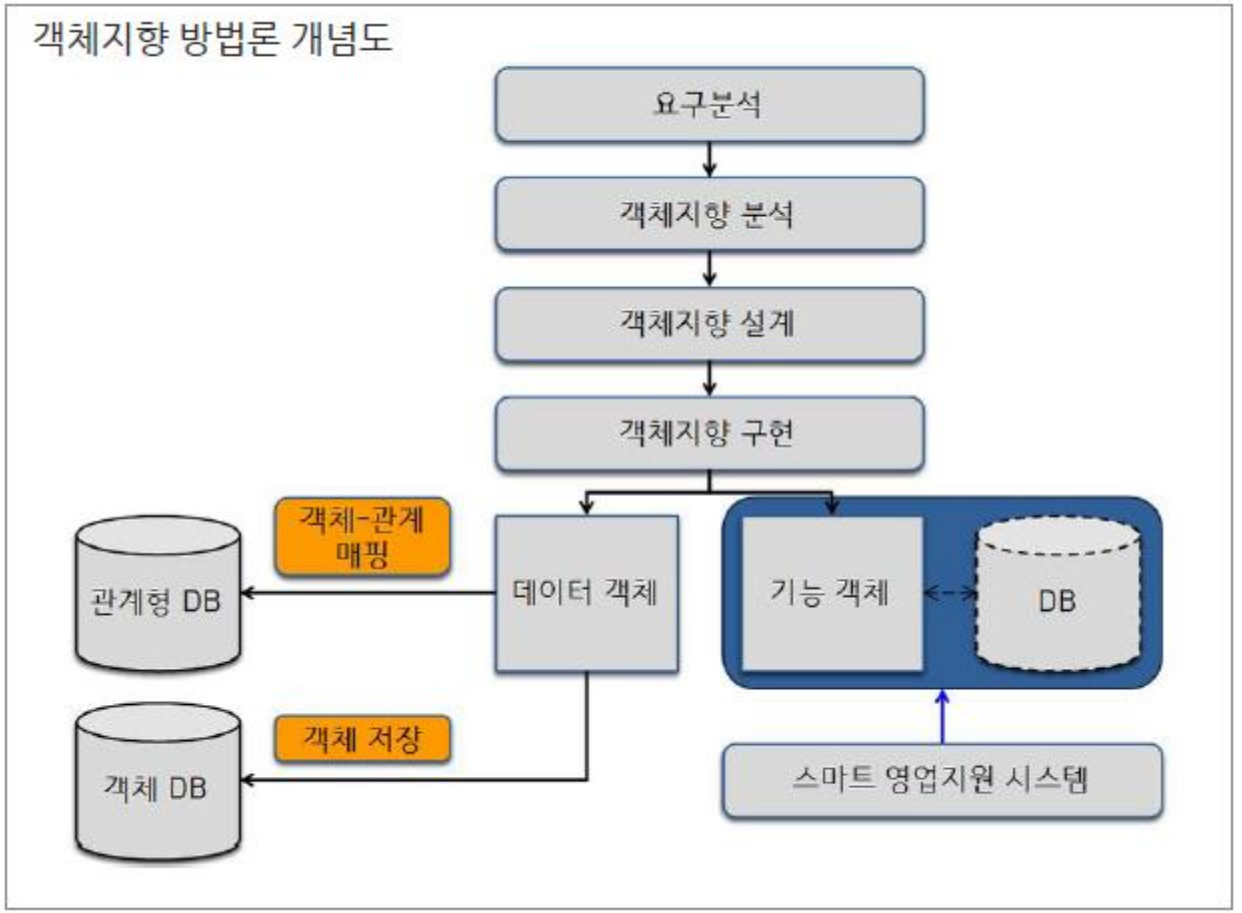
정보공학 개발 방법론 절차



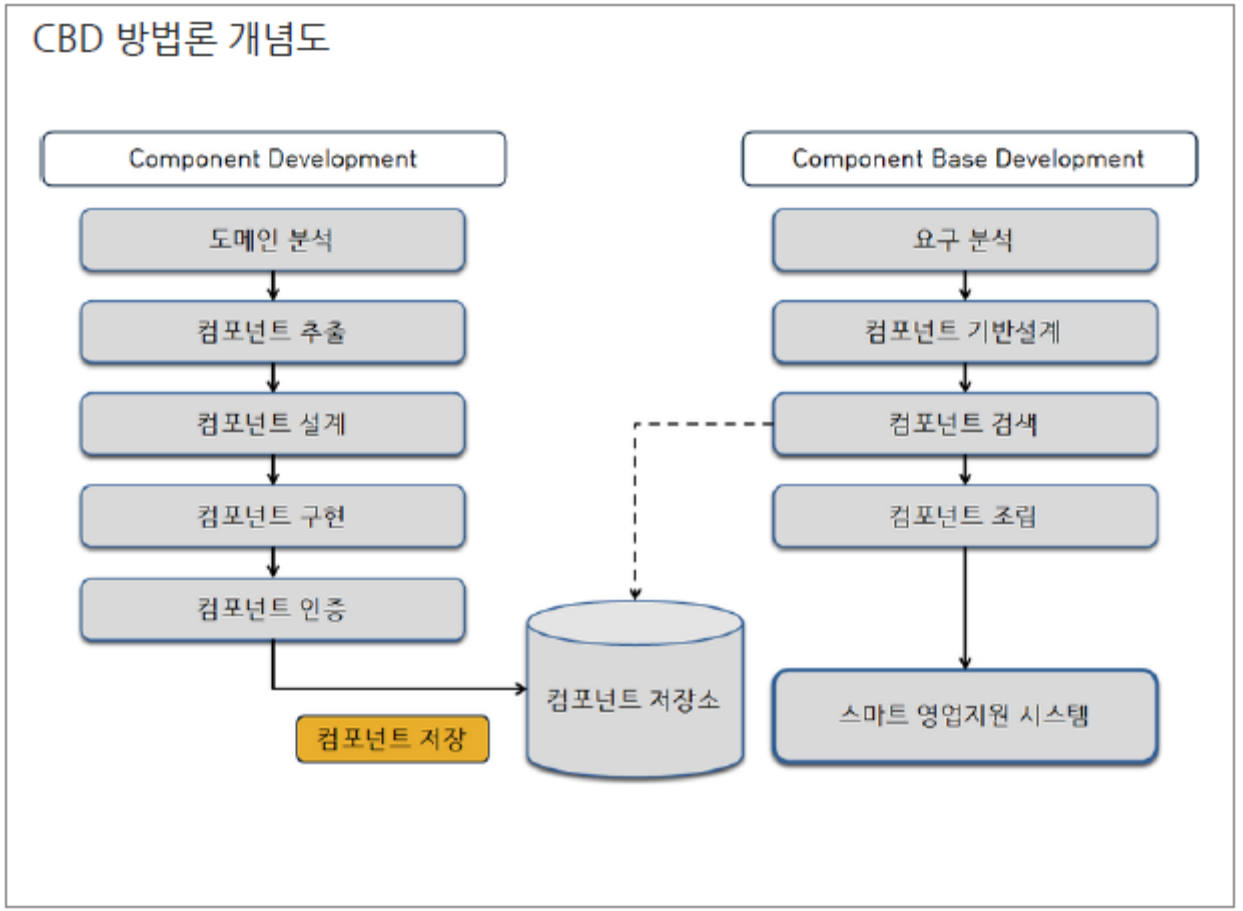
개념도



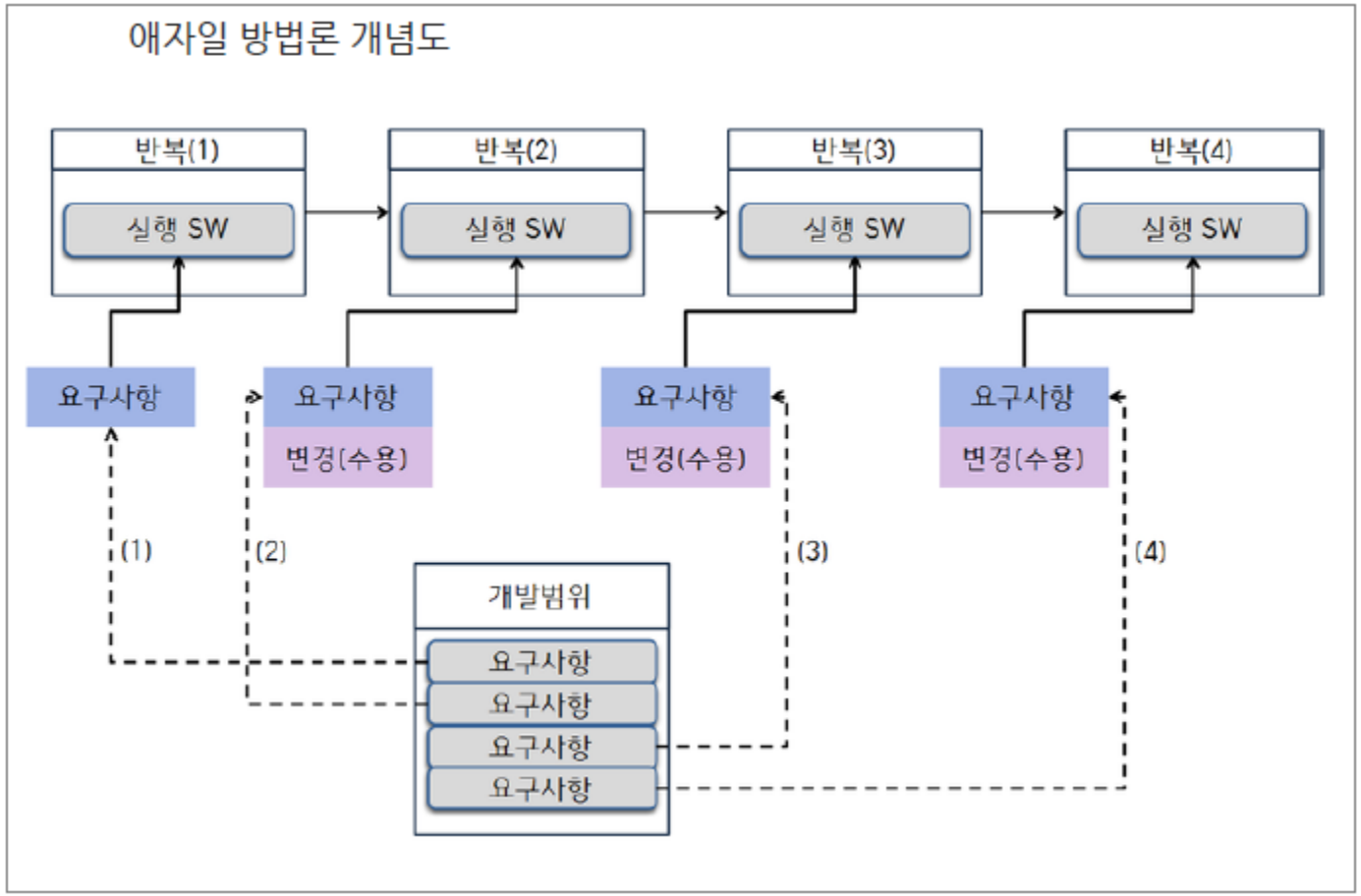
▪ 객체지향 방법론


객체지향 방법론 개념도
요구분석
객체지향 분석
객체지향 설계
객체지향 구현
객체-관계 매핑
관계형 DB
데이터 객체
기능 객체
DB
객체 저장
객체 DB
스마트 영업지원 시스템

- CBD 방법론: CBD, *component based* development



■ 애자일 방법론


애자일 방법론 개념도
반복(1)
반복(2)
반복(3)
반복(4)
실행 SW
실행 SW
실행 SW
실행 SW
요구사항
요구사항
변경(수용)
요구사항
변경(수용)
요구사항
변경(수용)
(1)
(2)
(3)
(4)
개발범위
요구사항
요구사항
요구사항
요구사항

▪ 빅데이터 분석 방법론

KDD

- 1996년 Fayyad가 프로파일링 기술을 기반으로 통계적 패턴이나 지식을 찾기 위해 체계적으로 정리한 데이터 마이닝 프로세스
- 데이터 마이닝, 기계학습, 인공지능, 패턴인식, 데이터 시각화 등에서 응용될 수 있는 구조

CRISP-DM

- 1996년 유럽연합의 ESPRIT의 프로젝트에서 시작
- 주요한 5의 업체(Maimler - Chryster, SPSS, NCR, Teradata, OHRA)가 주도
- 계층적 프로세스 모델로서 4개 레벨로 구성됨

- 단계 (Phases)
- 일반과제(Generic Tasks)
- 세부과제 (Specialized Tasks)
- 실행 process instances

빅데이터 분석 방법론

- 계층적 프로세스 모델- 3계층으로 구성
① 단계 (Phase) : 프로세스 그룹을 통하여 완성된 단계별 산출물 생성
② 태스크 (Task) : 단계를 구성하는 단위활동, 물리적 또는 논리적 단위
③ 스텝 (Step) : WBS의 워크패키지에 해당되고, 입력자료, 처리 및 도구, 출력자료로 구성된 단위 프로세스

빅데이터 분석 방법론의 3계층

- **계층적 프로세스 모델**: Stepwised Process Model

Phase (계층)

- 프로세스 그룹을 통하여 완성된 단계별 산출물 생성
- 기준선(Baseline)으로 설정 관리
- 버전관리(Configuration Management) 등을 통한 통제

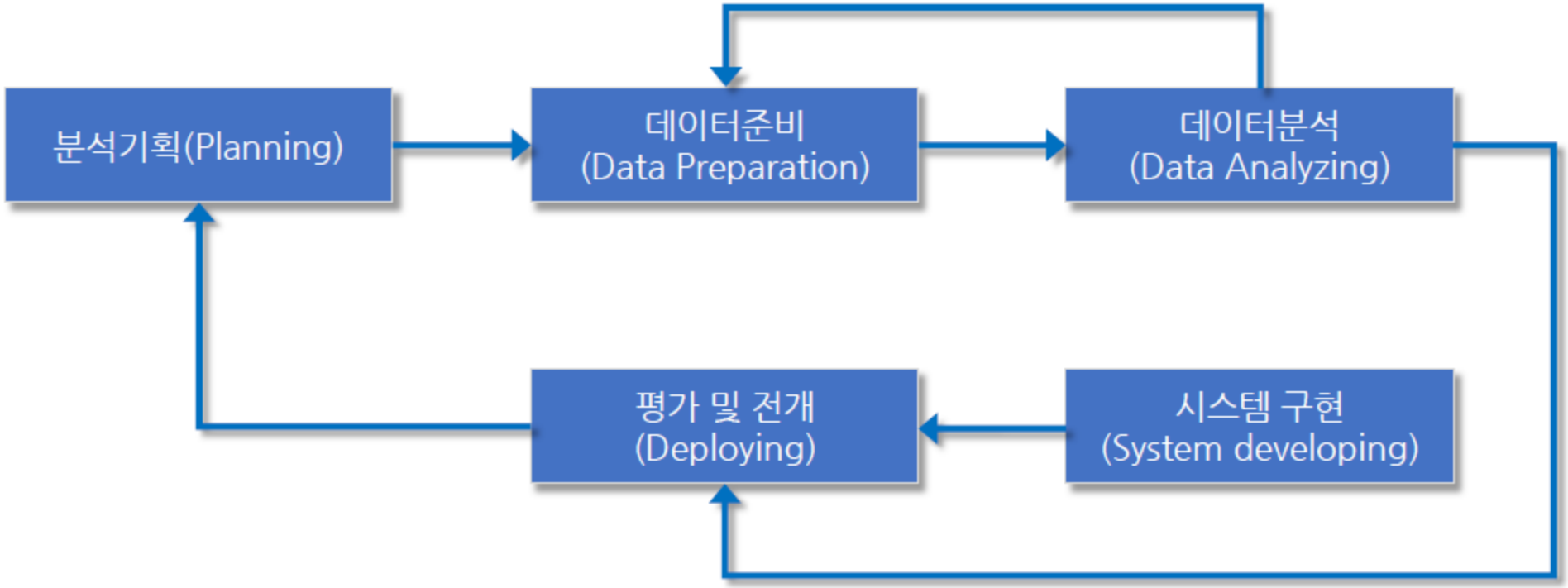
Task (테스크)

- 단계를 구성하는 단위 활동
- 물리적, 논리적 단위
- 품질 검토의 항목이 될 수 있음

Step(스텝)

- WBS(Work Breakdown Structure)의 워크패키지(Work Package)에 해당
- 입력(Input), 처리 및 도구(Process & Tool), 출력(Output)으로 구성
- 단위 프로세스(Unit Process)

▪ 빅데이터 분석 방법론의 단계



▪ 빅데이터 분석 방법론의 단계별 태스크

분석기획 (Planning)	데이터준비 (Data Preparation)	데이터분석 (Data Analyzing)	시스템 구현 (System developing)	평가 및 전개 (Deploying)
비즈니스 이해 및 범위설정	필요 데이터 정의	분석 데이터 준비	설계 및 구현	모델 발전계획 수립
프로젝트 정의 및 계획수립	데이터 스토어 설계	텍스트 분석	시스템 테스트 및 운영	프로젝트 편가 및 보고
프로젝트 위험 계획 수립	데이터 수집 및 정합성 점검	탐색적 분석		
		모델링		
		모델 평가 및 검증		
		모델적용 및 운영방안수립		

▪ 빅데이터 분석 방법론 – 분석기획 단계 (Planning)

Phase	Task	Step	내용	입력	처리및도구	출력
분석기획 (Planning)	비즈니스 이해 및 범위 설정		• 비즈니스에 대한 충분한 이해와 도메인 문제점 파악한다. • 업무 매뉴얼 및 업무 전문가 도움이 필요하다. • 구조화된 명세서 작성한다.			
		1) 비즈니스 이해	• 내부 업무 매뉴얼과 관련 자료 조사 • 외부의 관련 비즈니스 자료 조사 • 향후 프로젝트 진행을 위한 방향 설정	• 업무 매뉴얼 • 전문가의 지식 • 빅데이터 분석 대상 도메인에 대한 관련 자료	• 자료 수집 및 비즈니스 이해	• 비즈니스 이해 및 도메인 문제점
		2) 프로세스 범위 설정	• 비즈니스에 대한 이해와 프로젝트 목적에 부합하는 범위(Scope)를 명확하게 설정 • 이해 관계자(Stakeholders)의 이해를 일치 시키기 위해 구조화 된 프로젝트 범위 정의서(Statement Of Work, SOW)를 작성	• 중장기 계획서 • 빅데이터 분석 프로젝트 지시서, • 비즈니스 이해 및 도메인 문제점	• 자료 수집 및 비즈니스 이해 • 프로젝트 범위 정의서 작성 절차	• 프로젝트 범위 정의서 (SOW)
	프로젝트 정의 및 계획 수립		• 모델의 운영 이미지를 설계하고 모델 평가 기준을 설정하고, 프로젝트의 정의를 명확하게 한다. • WBS를 만들고, 데이터 확보 계획, 빅데이터 분석 방법, 일정계획, 예산계획, 품질계획, 인력구성계획, 의사소통계획 등을 포함하는 프로젝트 수행 계획을 작성한다.			
		1) 데이터 분석 프로젝트 정의	• 프로젝트의 목표 및 KPI, 목표 수준 등을 구체화하여 상세 프로젝트 정의서를 작성하고 • 프로젝트의 목표를 명확히 하기 위해 모델 운영 이미지 및 평가 기준을 설정	• 프로젝트 범위 정의서 • 빅데이터 분석 프로젝트 지시서	• 프로젝트 목표 구체화 • 모델 운영 이미지 설계	• 프로젝트 정의서 • 모델 운영 이미지 설계서 • 모델 평가 기준
		2) 프로젝트 수행 계획 수립	• 프로젝트의 목적 및 배경, 기대효과, 수행방법, 일정 및 추진조직, 프로젝트 관리 방안을 작성 • WBS는 프로젝트 산출물 위주로 작성하고 프로젝트 범위를 명확히 함	• 프로젝트 정의서 • 모델 운영 이미지 설계서 • 모델 평가 기준	• 프로젝트 수행 계획 작성 • WBS 작성 도구 • 일정 계획 수립 도구	• 프로젝트 수행 계획서 • WBS
	프로젝트 위험계획 수립		• 발생 가능한 모든 위험(Risk)를 발굴하고, 사전에 대응 방안을 수립함으로써 프로젝트 진행의 완전성을 높인다.-> 리스크 관리			
		1) 데이터 분석 위험 식별	• 프로젝트 산출물과 Lesson Learned를 참조하고 전문가의 판단을 활용함 • 식별된 위험은 위험의 영향도와 빈도, 발생가능성 등을 평가하여 우선순위 설정	• 프로젝트 정의서 • 프로젝트 수행 계획서 • 선행 프로젝트 산출물 및 정리자료	• 위험 식별 절차 • 위험영향도 및 발생가능성 분석 • 위험 우선순위 판단	• 식별된 위험 목록
		2) 위험 대응 계획 수립	• 식별된 위험은 정량적/정성적 항목을 상세히 분석하여 위험 대응 방안을 수립 • 대응은 "회피(Avoid), 전이(Transfer), 완화(Mitigate). 수용(Accept)"으로 구분하여 작성	• 식별된 위험 목록 • 프로젝트 정의서 • 프로젝트 수행 계획서	• 위험 정량적/정성적 분석	• 위험관리 계획서

▪ 빅데이터 분석 방법론 – 데이터 준비 단계 (Preparing)

Phase	Task	Step	내용	입력	처리 및 도구	출력
데이터 준비 (Preparing)	필요 데이터 정의		• 정형/비정형/반정형 등의 모든 내/외부 데이터 포함하고 데이터 속성, 오너, 담당자 등을 포함하는 데이터 정의서를 작성한다. • 구체적인 데이터 획득방안을 상세하게 수립하여 프로젝트 지연을 방지한다.			
		1) 데이터 정의	• 내/외부 원천 데이터 소스(Raw Data Source)로 부터 분석에 필요한 데이터를 정의	• 프로젝트 수행 계획서 • 시스템 설계서 • ERD • 메타 데이터 정의서 • 문서 자료	• 내/외부 데이터 정의, • 정형/비정형/반정형 데이터 정의	• 데이터 정의서
		2) 데이터 획득방안 수립	• 부서간 업무협조와 개인정보보호 및 정보보안과 관련한 문제점을 사전 점검 • 외부 데이터 획득은 인터페이스 및 법적인 문제점 고려해야함	• 데이터 정의서 • 시스템 설계서 • ERD • 메타데이터 정의서 • 문서 자료 • 데이터 구 입	• 데이터 획득 방안 수립	• 데이터 획득 계획서
	데이터 스토어 설계		• 획득 방안 수립 후 전사 차원의 데이터 스토어(Data Store)를 설계한다.			
		1) 정형 데이터 스토어 설계	• 일반적으로 RDMS(관계형 데이터베이스)를 사용하고, 데이터 스토어의 논리적, 물리 적 설계를 구분하여 설계	• 데이터 정의서 • 데이터 획득 계획서	• 데이터베이스 논리설계 • 데이터베이스 물리설계 • 데이터 매핑(Data Mapping)	• 정형데이터 스토어 설계서 • 데이터 매핑 정의서
		2) 비정형 데이터 스토어 설계	• 하둡(Hadoop), NoSQL 등을 이용한 논리적, 물리적 데이터 스토어 설계	• 데이터 정의서, • 데이터 획득 계획서	• 비정형/반정형 데이터 논리 및 물리 설계	• 비정형데이터 스토어 설계서 • 데이터 매핑 정의서
	데이터 수집 및 정합성 점검		• 데이터스토어에 크롤링(Crawling), 실시간 처리(Real Time), 배치 처리(Batch) 등으로 데이터를 수집한다. • DB 연동, API 활용, ETL 도구등을 활용하여 데이터 수집을 진행한다.			
		1) 데이터 수집 및 저장	• ETL, API, Script 프로그램 등을 이용하여 데이터를 수집하고 데이터 스토어에 저장함	• 데이터 정의서 • 데이터 획득 계획서 • 데이터 스토어 설계서	• 데이터 크롤링 도구 • ETL 도구 • 데이터 수집 스크립트	• 수집된 분석용 데이터
		2) 데이터 정합성 검증	• 데이터 스토어에 저장된 데이터의 정합성을 확보하고 • 품질개선이 필요한 부분은 보완 작업을 수행	• 수집된 분석용 데이터	• 데이터 품질 확인 • 데이터 정합성 점검 리스트	• 데이터 정합성 점검 보고서

※ ETL: Extract Transform Load

▪ 빅데이터 분석 방법론 – 데이터 분석 단계 (Analyzing)

Phase	Task	Step	내용	입력	처리 및 도구	출력
데이터 분석 (Analyzing)	분석용 데이터 준비		• 프로젝트 목표와 도메인을 이해하고 비즈니스 룰(Business Rule)을 확인한다. • 데이터 스토어에서 분석용 데이터 셋을 추출하고 구조화된 데이터 형태로 편성한다.			
		1) 비즈니스 룰 확인	• 프로젝트의 목표를 정확하게 인식, 이해 함 • 세부적인 비즈니스 룰을 파악하고 데이터 범위를 확인함	• 프로젝트 정의서 • 프로젝트 수행 계획서 • 데이터 정의서 • 데이터 스토어	• 프로젝트 목표 확인 • 비즈니스 룰 확인	• 비즈니스 룰 • 분석에 필요한 데이터 범위
		2) 분석용 데이터셋 준비	• 분석을 위해 추출된 데이터는 DB나 구조화된 형태로 구성하고 • 필요 시 분석을 위한 작업 공간과 전사 차원의 데이터스토어로 분리할 수 있음	• 데이터 정의서 • 데이터 스토어	• 데이터 선정 • 데이터 변환 • ETL 도구	• 분석용 데이터셋
	텍스트 분석		• "웹페이지, 로그, 텍스트 자료" 등을 이용하여, • "어휘/구문 분석(Word Analysis), 감성 분석(Sentimental Analysis), 토픽 분석(Topic Analysis), 오피니언 분석(Opinion Analysis), 소셜 네트워크 분석(Social Network Analysis)" 등을 실시하여 적절한 모델 구축을 한다.			
		1) 텍스트 데이터 확인 및 추출	• 비정형 데이터를 데이터스토어에서 확인하고 필요한 데이터를 추출함	• 비정형 데이터 스토어	• 분석용 텍스트 데이터 확인 • 텍스트 데이터 추출	• 분석용 텍스트 데이터
		2) 텍스트 데이터 분석	• 텍스트 데이터를 분석 도구로 적재하여 다양한 기법으로 분석하고 • 모델 구축, 용어 사전(유의어, 불용어 등)을 확보하고 도메인에 맞도록 작성 • 구축된 모델은 텍스트 시각화 도구를 이용하여 모델의 의미 전달	• 분석용 텍스트 데이터, • 용어사전(유의어, 불용어 등)	• 분류체계 설계 • 키워드 도출 • 형태소 분석 • 토픽 분석 / 감성 분석 / 오피니언 분석 / 네트워크 분석	• 텍스트 분석 보고서
	탐색적 분석		• 분석용 데이터셋에 대한 정합성 검토, 데이터 요약, 데이터 특성 파악 및 모델링에 필요한 데이터 편성한다. • EDA는 다양한 데이터 시각화를 활용하여 가독성을 높이고 형상 및 분포 등을 파악한다.			
		1) 탐색적 데이터 분석	• 기초 통계량(평균, 분산, 표준편차, 최대값, 최소값 등)을 산출하고 • 분포와 변수 간의 관계 등 데이터 자체의 특성 및 • 통계적 특성을 이해하고 모델링을 위한 기초 자료 활용	• 분석용 데이터셋	• EDA 도구, 통계 분석 • 변수간 연관성 분석 • 데이터 분포 확인	• 데이터 탐색 보고서
		2) 데이터 시각화	• 탐색적 분석을 위한 도구로 활용, 모델의 시스템화를 위한 시각화를 목적으로 할 경우 시각화 기획, 시각화 설계, 시각화 구현 등의 별도 프로세스를 따라 진행	• 분석용 데이터셋	• 시각화 도구 및 패키지 • 인포그래픽 • 시각화 방법론	• 데이터 시각화 보고서

※ EDA: Exploratory Data Analysis(탐색적 데이터 분석)

01. 데이터 분석 프로젝트

Phase	Task	Step	내용	입력	처리 및 도구	출력
데이터 분석 (Analyzing)	분석용 데이터 준비		• 프로젝트 목표와 도메인을 이해하고 비즈니스 룰(Business Rule)을 확인한다. • 데이터 스토어에서 분석용 데이터 셋을 추출하고 구조화된 데이터 형태로 편성한다.			
	텍스트 분석		• "웹페이지, 로그, 텍스트 자료" 등을 이용하여, • "어휘/구문 분석(Word Analysis), 감성 분석(Sentimental Analysis), 토픽 분석(Topic Analysis), 오피니언 분석(Opinion Analysis), 소셜 네트워크 분석(Social Network Analysis)" 등을 실시하여 적절한 모델 구축을 한다.			
	탐색적 분석		• 분석용 데이터셋에 대한 정합성 검토, 데이터 요약, 데이터 특성 파악 및 모델링에 필요한 데이터 편성한다. • EDA(Exploratory Data Analysis)는 다양한 데이터 시각화를 활용하여 가독성을 높이고 형상 및 분포 등을 파악한다.			
	모델링		• 가설 설정을 통해 통계 모델을 만들거나 기계학습(지도학습, 비지도학습 등)을 이용하여 모델을 만드는 과정이다. • 훈련용(Training)과 테스트용(Testing)으로 분할하여 과적합(Over-Fitting) 방지하고 모델의 일반화에 이용한다.			
		1) 데이터 분할	• Training과 Testing 용으로 분할, 교차검증(Cross Validation) 수행하거나 • 앙상블 (Ensemble) 기법을 적용할 경우 데이터 분할 또는 검증 횟수, 생성모델 개수 등을 설정 하여 분할 기법을 응용함	• 분석용 데이터셋	• 데이터 분할 패키지	• **훈련용 데이터** • **테스트용 데이터**
		2) 데이터 모델링	• 분류(Classification), 예측(Prediction), 군집(Clustering) 등의 모델을 만들어 가동 중인 운영 시스템에 적용함.	• 분석용 데이터셋	• 통계 모델링 기법 • 기계학습, 모델 테스트	• **모델링 결과 보고서**
		3) 모델 적용 및 운영 방안	• 운영에 적용하기 위해선 상세한 알고리즘 설명서 작성 필요 • 필요 시 의사코드 (Pseudocode) 수준의 상세한 작성 필요	• 모델링 결과 보고서	• 모니터링 방안 수립 • 알고리즘 설명서 작성	• **알고리즘 설명서** • **모니터링 방안**
	모델 평가 및 검증		• 프로젝트 정의서의 평가 기준에 따라 모델의 완성도를 평가하고 • 검증은 분석용 데이터셋이 아닌 별도의 데이터셋으로 검증하며 • 목표에 미달하는 경우 모델링 태스크를 반복하는 등 모델 튜닝 작업을 수행			
		1) 모델 평가	• 모델 평가를 위해 모델 결과 보고서 내의 알고리즘을 파악하고 • 테스트용 데이터나 필요 시 모델 검증을 위한 별도의 데이터를 활용함	• 모델링 결과 보고서 • 평가용 데이터	• 모델 평가, 모델 품질관리 • 모델 개선작업	• **모델 평가 보고서**
		2) 모델 검증	• 운영 데이터를 확보한 검증용 데이터를 이용해 • 모델 검증 작업을 실시하고 모델링 검증 보고서 작성함	• 모델링 결과 보고서 • 모델 평가 보고서 • 검증용 데이터	• 모델 검증	• **모델 검증 보고서**

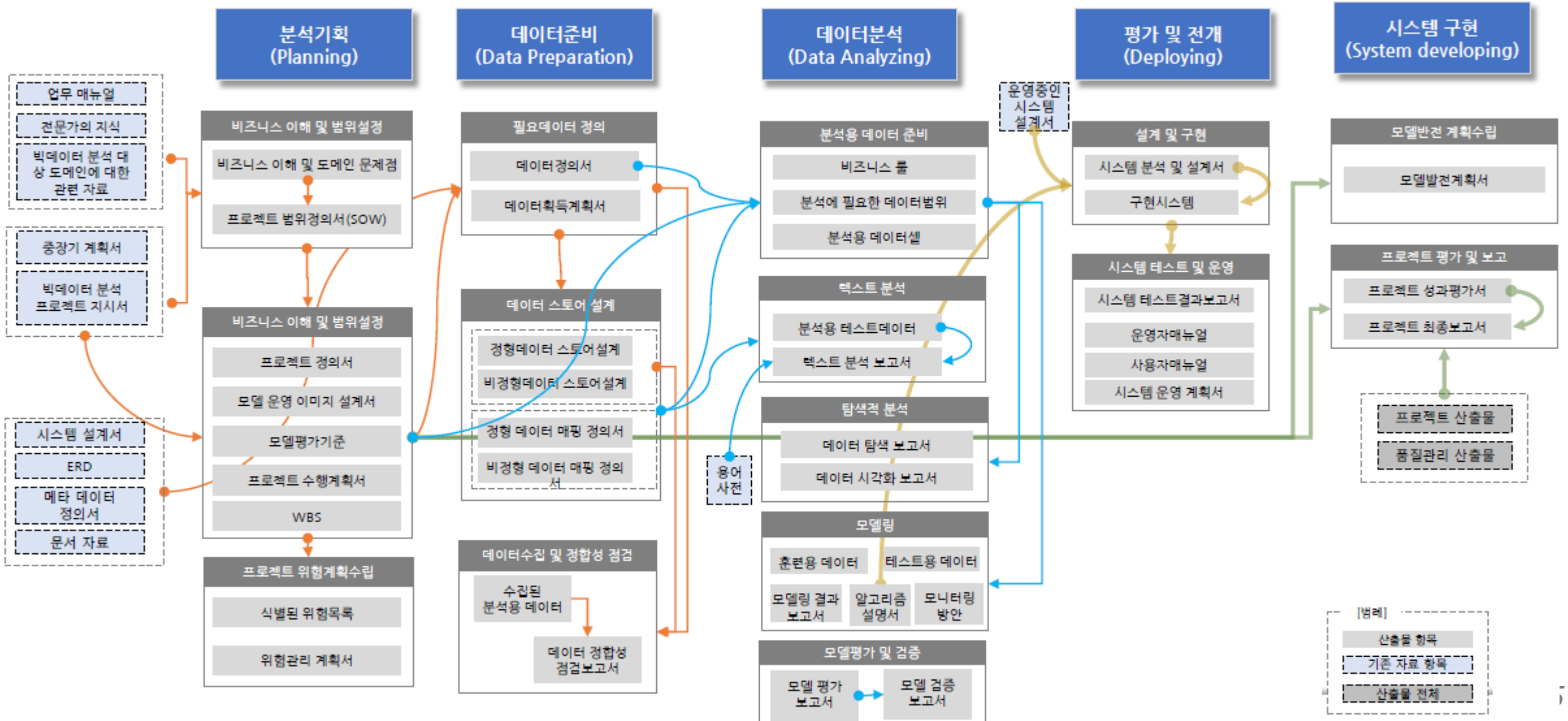
▪ 빅데이터 분석 방법론 – 시스템 구현 단계 (Developing)

Phase	Task	Step	내용	입력	처리 및 도구	출력
시스템 구현 (Developing)	설계 및 구현		• 모델링 태스크에서 작성된 알고리즘 설명서와 데이터 시각화 보고서를 이용하여 • 시스템 및 데이터 아키텍쳐 설계, 사용자 인터페이스 설계를 진행한다. • 설계서를 바탕으로 BI(Business Intelligence) 패키지를 활용하거나 새롭게 프로그램을 코딩하여 구축한다			
		1) 시스템 분석 및 설계	• 가동중인 시스템을 분석하고 • 알고리즘 설명서에 근거하여 응용시스템 구축 설계 프로세스를 진행	• 알고리즘 설명서, 운영중인 시스템 설계서	• 정보시스템 개발방법론	• 시스템 분석 및 설계서
		2) 시스템 구현	• 시스템 분석 및 설계서를 따라 BI 패키지를 활용하거나 • 새롭게 시스템을 구축함	• 시스템 분석 및 설계서 • 알고리즘 설명서	• 시스템 통합개발도구(IDE) • 프로그램 언어, 패키지	• 구현 시스템
	시스템 테스트 및 운영		• 시스템에 구현된 모델은 테스트를 통해 가동중인 시스템에 적용하고 효율적인 운영을 위한 프로세스를 진행한다.			
		1) 시스템 테스트	• 구축된 시스템의 검증(Verification & Validation)을 위해 • 단위테스트, 통합테스트, 시스템 테스트 등을 실시함	• 구현 시스템 • 시스템 테스트 계획서	• 품질관리 활동	• 시스템 테스트 결과보고서
		2) 시스템 운영 계획	• 시스템 운영자, 사용자를 대상으로 필요한 교육을 실시하고 • 시스템 운영계획 수립함	• 시스템 분석 및 설계서 • 구현 시스템	• 운영계획 수립 • 운영자 및 사용자 교육	• **운영자 매뉴얼** • 사용자 매뉴얼 • 시스템 운영 계획서

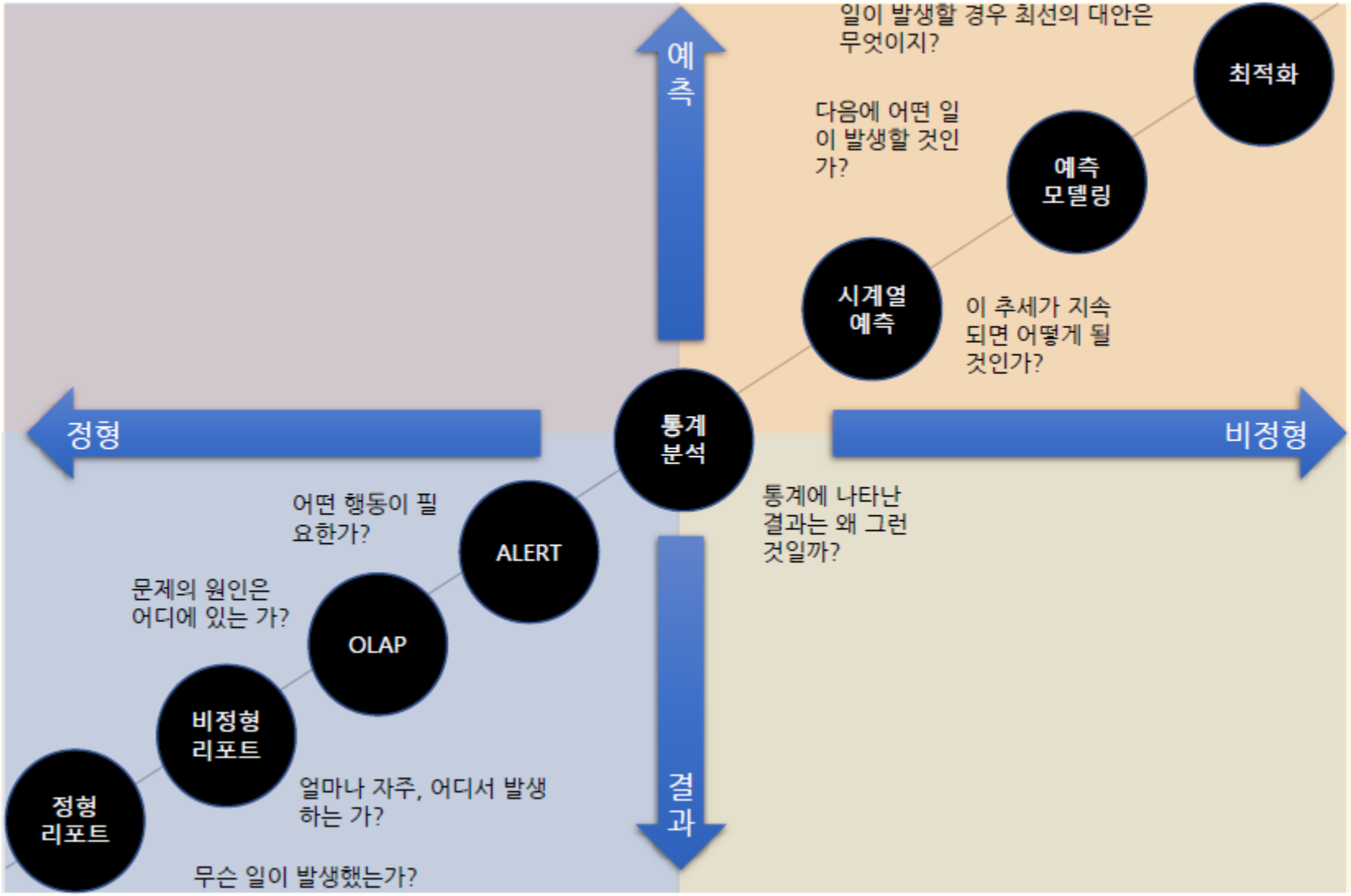
▪ 빅데이터 분석 방법론 – 평가 및 전개 단계 (Deploying)

Phase	Task	Step	내용	입력	처리 및 도구	출력
평가 및 전개 (Deploying)	모델 발전 계획 수립		• 모델의 생명 주기(Life Cycle)를 설정, 주기적인 평가를 실시하여 유지보수 하거나 재구축 방안을 마련한다. • 모델의 특성을 고려하여 모델 업데이터를 자동화하는 방안을 수립하고 적용할 수 있다.			
		1) 모델 발전 계획	• 발전계획을 상세하게 수립하여 모델의 계속성 확보함	• 구현 시스템 • 프로젝트 산출물	• 모델 발전 계획 수립	• 모델 발전 계획서
	프로젝트 평가 및 보고		• 기획 단계에서 설정된 기준에 따라 프로젝트의 성과를 정량적, 정성적 평가하고 • 프로젝트 진행과정에서 지식, 프로세스, 출력자료를 지식 자산화하고 • 프로젝트 최종 보고서를 작성한 후 의사소통계획에 따라 프로젝트를 종료한다.			
		1) 프로젝트 성과 평가	• 프로젝트의 정량적 성과와 정성적 성과로 나눠 성과 평가서 작성함	• 프로젝트 산출물 • 품질관리 산출물 • 프로젝트 정의서 • 프로젝트 수행 계획서	• 프로젝트 평가 기준 • 프로젝트 정량적 평가 • 프로젝트 정성적 평가	• 프로젝트 성과 평가서
		2) 프로젝트 종료	• 진행과정의 모든 산출물 및 프로세스를 지식자산화하고 • 최종 보고서를 작성하여 의사소통 절차에 따라 보고하고 • 프로젝트를 종료함	• 프로젝트 산출물 • 품질관리 산출물 • 프로젝트 정의서 • 프로젝트 수행 계획서 • 프로젝트 성과 평가서	• 프로젝트 지식자산화 작업 • 프로젝트 종료 활동	• 프로젝트 최종 보고서

▪ 빅데이터 분석 방법론 - 산출물

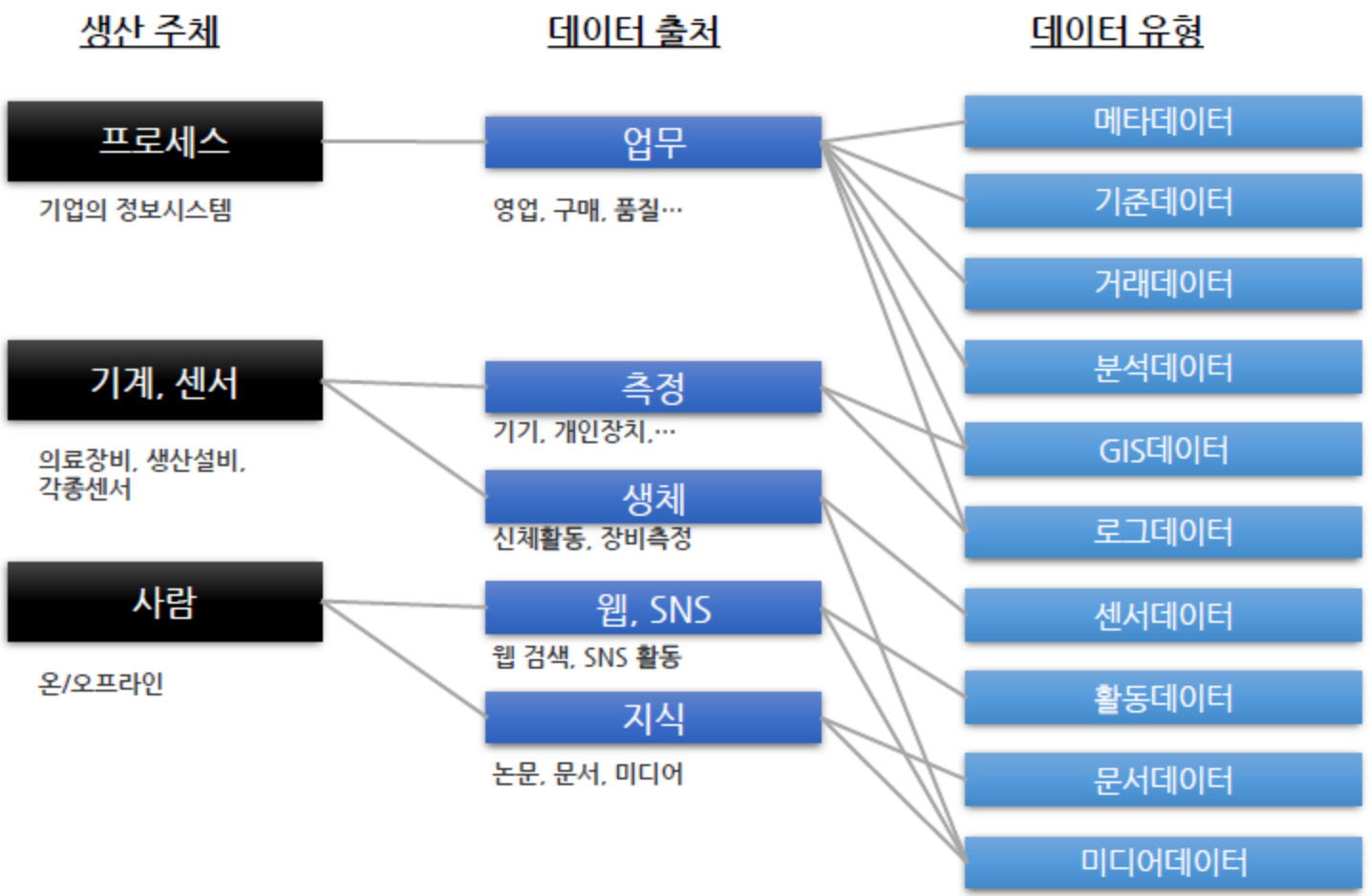


▪ 빅데이터 분석 프로젝트의 목적

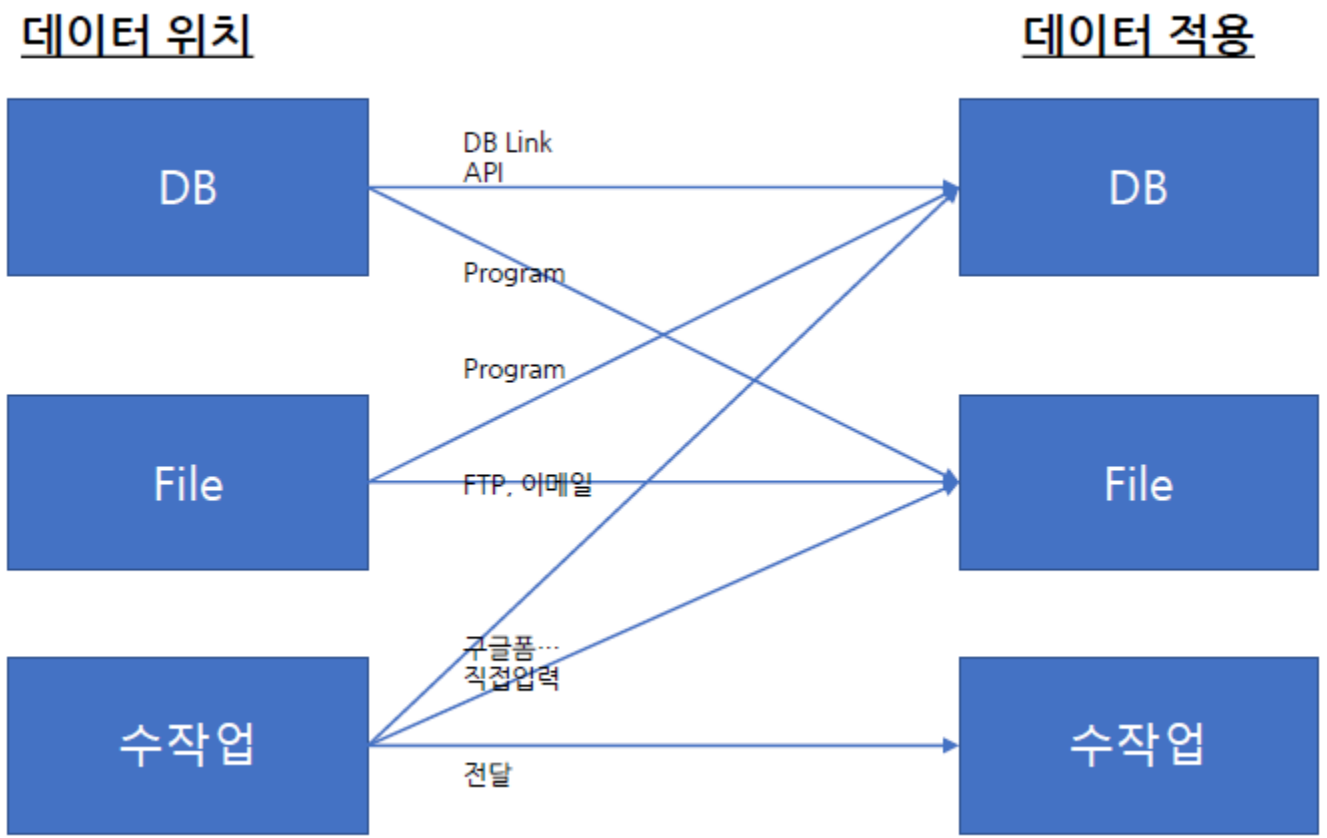


▪ 데이터의 유형 분류:

• 데이터의 생산 주체 및 자료의 출처에 따른 분류



▪ 데이터 수입



- DB Link
- DB Connect
- API
- FTP
- XML
- 로그 데이터
- 센서 데이터
- 구글폼
- ETL

01. 데이터 분석 프로젝트

위치	형태	종류	수집방법 정의
외부	정형 데이터	DBMS	DBMS 벤더가 제공하는 API를 통해 정형 데이터에 접근해 데이터를 수집하고 시스템에 저장
		이진 파일	ftp 프로토콜을 사용해 파일을 수집 시스템에 다운로드하고 해당 파일의 API를 통해 데이터 처리
	반정형 데이터	스크립트 파일	http 프로토콜을 사용해 파일의 텍스트를 스크랩하고 데이터에 저장된 메타정보를 읽어 파일을 파싱해 데이터 처리
		이진 파일	스트리밍을 사용해 파일의 텍스트를 스크랩하고 데이터에 저장된 메타정보를 읽어 파일을 파싱해 데이터 처리
	비정형 데이터	이진 파일	ftp 프로토콜을 사용해 파일을 수집 시스템에 다운로드하고 해당 파일을 API를 통해 데이터 처리
		스크립트 파일	http 프로토콜을 사용해 파일의 텍스트를 스크랩하고 내부 처리에서 텍스트를 파싱해 데이터 처리
내부	정형 데이터	DBMS	DBMS 벤더가 제공하는 API를 통해 정형 데이터에 접근해 데이터를 수집하고 시스템에 저장
		이진 파일	ftp 프로토콜을 사용해 파일을 수집 시스템에 다운로드 하고 해당 파일의 API를 통해 데이터 처리
	반정형 데이터	스크립트 파일	http 프로토콜을 사용해 파일의 텍스트를 스크랩하고 데이터에 저장된 메타정보를 읽어 파일을 파싱해 데이터 처리
		이진 파일	스트리밍을 사용해 파일의 텍스트를 스크랩하고 데이터에 저장된 메타정보를 읽어 파일을 파싱해 데이터 처리
	비정형 데이터	파일	ftp 프로토콜을 사용해 파일을 수집 시스템에 다운로드 하고 해당 파일을 API를 통해 데이터 처리

※ 이진 파일(영어: binary file 바이너리 파일)은
텍스트 파일이 아닌 컴퓨터 파일. "바이너리 파일"이라는 용어는 종종 "논-텍스트 파일(non-text file)"을 의미하는 용어로 사용
컴퓨터 파일로 컴퓨터 저장과 처리 목적을 위해 이진 형식으로 인코딩된 데이터를 포함한다

데이터 수집을 위한 개방형 공공데이터 목록

- Kaggle: www.Kaggle.com
- **공공데이터 포털**: www.data.go.kr
- DACON: dacon.io
- AI Hub: aihub.or.kr
- **서울 열린데이터광장**: data.seoul.go.kr
- **경기 데이터 드림**: data.gg.go.kr
- D-**데이터허브**: data.daegu.go.kr
- **보건의료 빅데이터**: opendata.hira.or.kr

Any Questions?