

01 빅데이터 이해와 분석 기획 · 데이터 수집·저장·제도

빅데이터 이해와 분석 기획

- 빅데이터의 규모·속도·다양성은 수집과 처리 방식의 선택에 영향을 준다.
- 분석 문제는 현상, 목표, 제약, 평가 기준과 예상 활용 주체를 포함해 정의한다.
- 분석 과제는 업무 가치, 시급성, 데이터 준비도와 실행 가능성으로 우선순위를 정한다.
- 로드맵은 과제의 선후관계, 공통 데이터와 플랫폼 의존성을 반영한다.

기획 요소	핵심 질문	대표 산출물
문제 정의	무엇을 왜 개선하는가	문제·목표·제약
성과 기준	성공을 어떻게 판단하는가	업무·모델 지표
우선순위	무엇부터 수행하는가	가치·시급성·가능성

데이터 수집·저장·제도

- 정형 데이터는 고정 스키마, 반정형 데이터는 태그·키, 비정형 데이터는 자유로운 형식을 가진다.
- ETL은 변환 후 적재하고 ELT는 원천을 적재한 뒤 목적에 맞게 변환한다.
- 데이터 웨어하우스는 통합 분석, 데이터 레이크는 다양한 원천의 유연한 보관에 초점을 둔다.
- 메타데이터와 계보는 데이터의 의미, 출처, 변환과 사용 위치를 추적하게 한다.

대상	특징	확인 사항
웨어하우스	정제·통합 분석	일관된 스키마
데이터 레이크	원형 중심 저장	검색·품질 통제
메타데이터	의미와 구조 설명	소유자·정의·버전