

SPARKLE CBT · 빅데이터분석기사 필기 · ORIGINAL STUDY GUIDE

빅데이터분석기사 필기 종합 학습서

분석기획 - 탐색 - 모델링 - 결과해석을 하나의 분석 생명주기로 연결한 시험
대비서

KDATA 빅데이터분석기사 공식 시험 주요내용 기준
최종 확인일 2026-08-19

00 이 책의 범위와 학습 방법

독립 제작 학습자료

이 자료는 공식 교재, 공식 문제 또는 복원문제가 아닌 Sparkle CBT 자체제작 학습자료입니다. 공식 출제범위를 학습 목적으로 사용했으며 사례, 설명, 점검 문항은 새로 작성했습니다.

적용 기준: KDATA 빅데이터분석기사 공식 시험 주요내용 기준

최종 확인일: 2026-08-19

공식 안내: https://www.dataq.or.kr/www/sub/a_07.do

시험 구조

구분	구성	학습 판단
빅데이터 분석기획	20문항·25점	이해·계획·수집·저장
빅데이터 탐색	20문항·25점	전처리·탐색·통계
빅데이터 모델링	20문항·25점	모형 설계·기법 적용
빅데이터 결과 해석	20문항·25점	평가·개선·활용
합격	80문항·120분	과목별 40점 이상, 전 과목 평균 60점 이상

권장 학습 루틴

- 4과목을 분석 문제 정의부터 운영 활용까지 하나의 흐름으로 먼저 연결한다.
- 통계와 모델은 공식, 입력 조건, 결과 해석과 대표 함정을 한 묶음으로 정리한다.
- 평가 지표는 정의만 외우지 말고 어떤 업무 오류를 줄이는지 사례로 판단한다.
- 120분 모의고사에서 과목별 과락을 먼저 방지하고 반복 오답 원인을 제거한다.

규정 확인

시험 일정, 프로그램 버전, 출제기준과 운영 규정은 바뀔 수 있습니다. 접수 전 반드시 주관기관의 최신 공고를 다시 확인하세요.

01 빅데이터 이해와 분석 기획

분석은 기술을 고르는 일보다 해결할 업무 문제와 성공 기준을 정의하는 일에서 시작한다. 빅데이터의 특징과 가치, 조직 역할을 분석 과제의 목적과 연결해 판단한다.

핵심 구조

- 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.
- 분석가, 데이터 엔지니어, 업무 담당자와 책임자의 역할을 구분한다.

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
로드맵	어떤 순서로 확장하는가	단계·의존성·책임

자주 틀리는 지점

정확도가 높은 모형이라는 말만으로 분석 목표가 되지 않는다. 어떤 업무 의사결정을 얼마나 개선할지와 허용 가능한 오류를 함께 정의해야 한다.

적용 점검

- 이탈 예측 과제의 업무 목표와 모델 지표를 각각 적는다.
- 과제 두 개의 우선순위를 가치와 데이터 준비도로 비교한다.
- 분석 로드맵에서 선행 데이터 과제를 표시한다.

02 데이터 수집·저장·제도

분석 가능한 데이터는 출처와 수집 방식, 저장 구조, 품질과 법적 허용 범위를 함께 관리해야 한다. 정형·반정형·비정형 자료의 특성을 플랫폼 구성과 연결한다.

핵심 구조

- 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.
- 개인정보 활용은 목적, 최소 수집, 안전조치와 재식별 위험을 함께 검토한다.

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
데이터 계보	변환 경로 추적	원천·처리·사용처

자주 틀리는 지점

가명정보는 추가정보 없이는 특정인을 알아볼 수 없게 처리한 정보이지, 모든 보호 의무가 사라지는 익명정보와 같지 않다.

적용 점검

- ETL과 ELT를 변환 시점으로 비교한다.
- 데이터 레이크에 품질 통제가 필요한 이유를 설명한다.
- 개인정보 활용 계획에 포함할 보호 조치를 적는다.