

EXERCISES & SOLUTIONS

1장 연습문제 풀이

데이터베이스 기본 개념

총 25문항 | 객관식 14 + OX 1 + 단답형 4 + 서술형 6

박상돈 조교수 | 대전대학교 컴퓨터공학과



Part A — 객관식 문제

Multiple Choice Questions (Q1 ~ Q14)



데이터와 정보에 대한 설명으로 옳지 않은 것은?

1

데이터와 정보를 구별하는 기준은 가공의 유무다.

2

데이터는 현실 세계에서 관찰이나 측정으로 수집한 사실이나 값이다.

3

정보는 의사 결정에 활용하기 위해 데이터를 처리한 결과물이다.

4

정보를 가공하면 데이터를 얻을 수 있다.

1 데이터와 정보를 구별하는 기준은 가공의 유무다.

2 데이터는 현실 세계에서 관찰이나 측정으로 수집한 사실이나 값이다.

3 정보는 의사 결정에 활용하기 위해 데이터를 처리한 결과물이다.

정보를 가공하면 데이터를 얻을 수 있다.
✓

해설

정답: ④ – 데이터를 가공하면 정보를 얻을 수 있습니다. 반대로 '정보를 가공하면 데이터를 얻는다'는 관계가 뒤바뀐 잘못된 설명입니다.

- 데이터(Data) → [가공/처리] → 정보(Information)
- 예: '서울 기온 35°C' (데이터) → '폭염 경보 발령' (정보)



데이터와 정보의 차이를 고려했을 때, 다음 중 정보의 예로 가장 적합하지 않은 것은?

1 제품가격

2 베스트셀러

3 우수고객

4 월별 평균 주문금액

제품가격 ✓

3 우수고객

2 베스트셀러

4 월별 평균 주문금액

해설

정답: ① — '제품가격'은 가공 없이 그대로 수집된 원시 데이터입니다.

- 베스트셀러: 판매량 데이터를 집계·순위화한 정보
- 우수고객: 구매 이력 데이터를 분석하여 등급을 부여한 정보
- 월별 평균 주문금액: 주문 데이터를 월별로 평균 계산한 정보
- 제품가격: 개별 제품에 매겨진 값 자체 → 데이터



데이터베이스의 정의와 거리가 먼 것은?

1

전용 데이터 (exclusive data)

2

통합 데이터 (integrated data)

3

저장 데이터 (stored data)

4

운영 데이터 (operational data)

전용 데이터 (exclusive data) ✓

3 저장 데이터 (stored data)

2 통합 데이터 (integrated data)

4 운영 데이터 (operational data)

해설

정답: ① – 데이터베이스의 4가지 정의는 공유 데이터(shared), 통합 데이터(integrated), 저장 데이터(stored), 운영 데이터(operational)입니다.

'전용 데이터(exclusive data)'는 데이터베이스의 정의에 포함되지 않습니다. 데이터베이스는 '전용'이 아니라 '공유(shared)' 데이터입니다.

? 다음 설명과 관련 있는 데이터베이스의 정의는?

"데이터베이스는 데이터의 중복을 최소화하고, 통제가 가능한 중복만 허용한다."

1 운영 데이터

2 저장 데이터

3 공유 데이터

4 통합 데이터

1 운영 데이터

2 저장 데이터

3 공유 데이터

통합 데이터 ✓

해설

정답: ④ – 통합 데이터(Integrated Data)는 데이터의 중복을 최소화한 데이터의 집합을 의미합니다.

- 완전히 중복을 제거하는 것이 아니라 '통제된 중복(controlled redundancy)'만 허용
- 같은 데이터가 여러 곳에 불필요하게 중복 저장되는 것을 방지
- 데이터 일관성(consistency) 유지에 핵심적인 역할

? 다음 설명과 관련 있는 데이터베이스의 정의는?

"데이터베이스는 조직을 운영하고 조직의 주요 기능을 수행하기 위해 꼭 필요한 데이터의 집합이다."

1 운영 데이터

2 저장 데이터

3 공유 데이터

4 통합 데이터

운영 데이터 ✓

3 공유 데이터

2 저장 데이터

4 통합 데이터

 해설

정답: ① — 운영 데이터(Operational Data)는 조직의 주요 기능을 수행하기 위해 반드시 필요한 데이터를 의미합니다.

- 단순히 보관 목적이 아닌, 조직 운영에 필수적인 데이터
- 예: 은행의 고객 계좌 정보, 대학의 학생 수강 정보
- 일시적으로 필요하거나 단순 보관 데이터는 운영 데이터가 아님



다음 중 데이터베이스에 대한 설명으로 거리가 먼 것은?

1

조직의 고유한 업무를 수행하기 위해 반드시 필요한 데이터가 모여 있다.

2

컴퓨터가 접근할 수 있는 매체에 데이터를 저장한다.

3

검색의 효율성을 위해 모든 데이터를 반드시 여러 개 중복해서 저장한다.

4

여러 사용자가 공동으로 데이터를 소유하고 이용할 수 있다.

1 조직의 고유한 업무를 수행하기 위해 반드시 필요한 데이터가 모여 있다.

2 컴퓨터가 접근할 수 있는 매체에 데이터를 저장한다.

검색의 효율성을 위해 모든 데이터를 반드시 여러 개 중복해서 저장한다. ✓

4 여러 사용자가 공동으로 데이터를 소유하고 이용할 수 있다.

해설

정답: ③ — 데이터베이스는 '통합 데이터'로서 중복을 최소화합니다.

- '모든 데이터를 반드시 중복 저장한다'는 것은 오히려 파일 시스템의 문제점
- DB는 중복을 최소화하고, 필요한 경우 통제된 중복만 허용
- ①은 운영 데이터, ②는 저장 데이터, ④는 공유 데이터에 해당하는 올바른 설명



데이터베이스의 특성과 거리가 먼 것은?

1 실시간 접근성 (real-time accessibility)

2 계속 변화 (continuous evolution)

3 동시 공유 (concurrent sharing)

4 위치 기반 참조 (location reference)

1 실시간 접근성 (real-time accessibility)

2 계속 변화 (continuous evolution)

3 동시 공유 (concurrent sharing)

위치 기반 참조 (location reference) ✓

해설

정답: ④ – 데이터베이스의 4가지 특성은 실시간 접근성, 계속 변화, 동시 공유, 내용 기반 참조입니다.

- '위치 기반 참조'가 아니라 '내용 기반 참조(content reference)'
- 데이터의 물리적 위치(주소)가 아닌, 데이터 값(내용)을 기준으로 참조
- 예: WHERE 학과 = '컴퓨터공학과' ← 내용(값)으로 검색

? 다음 설명과 관련 있는 데이터베이스의 특성은?

"데이터베이스는 현실 세계의 상태를 정확히 반영하기 위해 데이터를 계속 삽입·삭제·수정하여 현재의 정확한 데이터를 유지해야 한다."

1 실시간 접근성

2 계속 변화

3 동시 공유

4 내용 기반 참조

1 실시간 접근성

3 동시 공유

계속 변화 ✓

4 내용 기반 참조

해설

정답: ② — 계속 변화(Continuous Evolution)는 데이터베이스가 동적으로 변한다는 특성입니다.

- 삽입(Insert), 삭제>Delete), 수정(Update)을 통해 항상 최신 상태 유지
- 현실 세계의 변화를 정확히 반영해야 함
- 예: 학생이 수강 신청 → 수강 데이터 삽입 / 졸업 → 데이터 갱신



데이터베이스의 특성으로 옳지 않은 것은?

1

같은 내용의 데이터를 여러 사람이 동시에 공유할 수 있다.

2

데이터베이스는 데이터의 삽입, 삭제, 수정으로 내용이 계속적으로 변한다.

3

수시로 이루어지는 사용자 질의에 대하여 실시간 처리로 응답할 수 있어야 한다.

4

데이터의 참조는 저장되어 있는 데이터의 주소나 위치에 의해 이루어진다.

1 같은 내용의 데이터를 여러 사람이 동시에 공유할 수 있다.

2 데이터베이스는 데이터의 삽입, 삭제, 수정으로 내용이 계속적으로 변한다.

3 수시로 이루어지는 사용자 질의에 대하여 실시간 처리로 응답할 수 있어야 한다.

데이터의 참조는 저장되어 있는 데이터의 주소나 위치에 의해 이루어진다. ✓

해설

정답: ④ – 데이터베이스는 '내용 기반 참조'이지, '주소/위치 기반 참조'가 아닙니다.

- 데이터의 물리적 주소가 아닌, 데이터의 내용(값)으로 참조
- 예: `SELECT * FROM 학생 WHERE 이름 = '홍길동'`
→ 저장 위치가 아닌 '이름' 값으로 검색
- ①은 동시 공유, ②는 계속 변화, ③은 실시간 접근성으로 모두 올바른 설명



다음 중 비정형 데이터로 분류하기 어려운 것은?

1

CCTV 녹화 영상

2

이메일

3

회사의 인사 기록

4

SNS 활동 기록

1 CCTV 녹화 영상

2 이메일

회사의 인사 기록 ✓

4 SNS 활동 기록

해설

정답: ③ – 회사의 인사 기록은 정형 데이터입니다.

- 인사 기록: 직원번호, 이름, 부서, 직급 등 정해진 구조(테이블)에 저장 → 정형 데이터
- CCTV 영상: 영상 파일 → 비정형 데이터
- 이메일: 텍스트 + 첨부파일이 혼합된 → 비정형 데이터
- SNS 활동: 텍스트, 이미지, 동영상 혼합 → 비정형 데이터



다음 중 데이터 유형이 다른 것은?

1 개인 SNS에 올린 영화 관람 후기

2 물류창고 관리를 위해 온도와 습도를 시간 별로 기록한 센서 데이터

3 통화 내용을 녹음한 음성파일

4 항공권 구매를 위한 챗봇과의 대화 내용

1 개인 SNS에 올린 영화 관람 후기

3 통화 내용을 녹음한 음성파일

물류창고 관리를 위해 온도와 습도를 시간별로 기록한 센서 데이터 ✓

4 항공권 구매를 위한 챗봇과의 대화 내용

해설

정답: ② – 센서 데이터는 반정형 데이터이고, 나머지는 비정형 데이터입니다.

- 센서 데이터: 시간별 온도·습도 값이 일정한 구조로 기록 → 반정형 데이터
(JSON, XML 등의 형식으로 구조 정보를 함께 포함)
- SNS 후기, 음성파일, 챗봇 대화: 정해진 구조 없이 자유로운 형식 → 비정형 데이터



다음 중 범주형 데이터에 속하는 것은?

1 도서가격

2 도서장르

3 판매부수

4 도서평점

1 도서가격

도서장르 ✓

3 판매부수

4 도서평점


해설

정답: ② – 도서장르(소설, 시, 에세이 등)는 범주(카테고리)를 나타내는 범주형 데이터입니다.

- 도서가격: 숫자 값으로 크기 비교·연산 가능 → 수치형(연속형)
- 도서장르: 종류를 나타내는 분류 값 → 범주형(명목형)
- 판매부수: 개수를 세는 정수 값 → 수치형(이산형)
- 도서평점: 숫자 값으로 크기 비교 가능 → 수치형(연속형)



다음 중 이산형 데이터에 속하는 것은?

1 고객 멤버십 등급

2 가입 기간

3 고객 거주 지역

4 가입 고객 수

1 고객 멤버십 등급

2 가입 기간

3 고객 거주 지역

가입 고객 수 ✓


해설

정답: ④ – 가입 고객 수는 셀 수 있는 정수 값으로 이산형 데이터입니다.

- 이산형 데이터: 셀 수 있는(countable) 정수 값 (예: 고객 수, 주문 건수)
- 고객 멤버십 등급: 골드/실버/브론즈 → 범주형(순서형)
- 가입 기간: 연속적인 시간 값 → 수치형(연속형)
- 고객 거주 지역: 서울/부산 등 분류 → 범주형(명목형)



데이터를 정성적 데이터와 정량적 데이터로 구분했을 때, 정성적 데이터에 속하는 것은?

1 신문기사

2 주가

3 강수량

4 습도

신문기사 ✓

3 강수량

2 주가

4 습도

 해설

정답: ① – 신문기사는 텍스트로 된 정성적(Qualitative) 데이터입니다.

- 정성적 데이터: 수치로 표현하기 어려운 데이터 (텍스트, 의견, 감상 등)
→ 저장·분석 처리 비용이 상대적으로 높음
- 정량적 데이터: 객관적 수치로 표현된 데이터
→ 주가, 강수량, 습도는 모두 숫자로 측정 가능한 정량적 데이터

Part B — OX 문제

True or False (Q15)

Q15

문제

데이터베이스의 특성에 대한 설명이 맞으면 ○, 틀리면 ×를 표시하십시오.

(1) 데이터베이스는 실시간으로 접근이 가능하다.



(2) 데이터베이스는 변화가 허용되지 않는다.



(3) 데이터베이스는 동시 공유가 가능하다.



(4) 데이터베이스의 데이터는 주소로 참조된다.





(1) 데이터베이스는 실시간으로 접근이 가능하다.

→ 실시간 접근성 – 사용자의 질의에 즉시 응답



(2) 데이터베이스는 변화가 허용되지 않는다.

→ 계속 변화 – 삽입·삭제·수정으로 항상 최신 상태 유지



(3) 데이터베이스는 동시 공유가 가능하다.

→ 동시 공유 – 여러 사용자가 동시에 같은 데이터 이용 가능



(4) 데이터베이스의 데이터는 주소로 참조된다.

→ 내용 기반 참조 – 주소가 아닌 데이터 값(내용)으로 참조

Part C — 단답형 문제

Short Answer Questions (Q16 ~ Q19)



다음 설명에서 A와 B가 각각 무엇인지 답하십시오.

데이터를 구조화된 형태에 따라 분류했을 때, 엑셀의 스프레드시트나 관계 데이터베이스의 테이블과 같이 미리 정해진 구조에 따라 저장된 데이터를 A라 한다. 그리고 소셜 데이터의 텍스트, 영상, 이미지와 같이 정해진 구조가 없이 저장된 데이터를 B라 한다.



정답

A : 정형 데이터 (Structured Data)

B : 비정형 데이터 (Unstructured Data)



해설

데이터는 구조화 정도에 따라 세 가지로 분류됩니다.

- 정형 데이터: 미리 정해진 구조(스키마)에 따라 저장 (예: 엑셀, RDBMS 테이블)
- 반정형 데이터: 구조 정보가 데이터 내에 포함 (예: HTML, XML, JSON)
- 비정형 데이터: 정해진 구조 없이 저장 (예: 텍스트, 영상, 이미지)



다음은 범주형 데이터 중에서도 어떤 유형의 데이터에 대한 설명인가?

성별, 혈액형, 음식 메뉴, 제품 색상, 거주도시



정답

명목형 데이터 (Nominal Data)



해설

명목형 데이터는 단순히 분류·구분을 위한 이름이나 라벨을 나타내는 데이터입니다.

- 순서나 크기 관계가 없음 (서울 > 부산 같은 비교 불가)
- 빈도 계산은 가능하지만, 산술 연산은 불가
- 비교: 순서형 데이터(학점 $A > B$, 만족도 상 > 중 > 하)는 순서 관계가 존재



다음은 수치형 데이터 중에서도 어떤 유형의 데이터에 대한 설명인가?

온도, 길이, 몸무게, 시력, 점수



정답

연속형 데이터 (Continuous Data)



해설

연속형 데이터는 연속적인 범위 내에서 어떤 값이든 가질 수 있는 데이터입니다.

- 소수점 이하의 값도 의미가 있음 (예: 36.5°C, 172.3cm)
- 측정(measurement)을 통해 얻음
- 비교: 이산형 데이터(학생 수, 판매 건수)는 셀 수 있는(countable) 정수 값만 가능



데이터를 정성적 데이터와 정량적 데이터로 구분했을 때 다음은 무엇에 대한 설명인가?

"판매량, 재고량, 매출액과 같이 객관적인 수치로 표현된 데이터로 저장하고 검색이나 분석 처리를 수행하는데 상대적으로 비용이 적게 든다."



정답

정량적 데이터 (Quantitative Data)



해설

정량적 데이터는 객관적인 수치(숫자)로 표현할 수 있는 데이터입니다.

- 저장·검색·분석의 처리 비용이 상대적으로 적음
- 예: 판매량, 재고량, 매출액, 주가, 온도 등
- 반대로 정성적(Qualitative) 데이터는 텍스트·의견 등 수치화가 어렵고 처리 비용이 높음

Part D — 서술형 문제

Essay Questions (Q20 ~ Q25)



데이터와 정보의 차이를 설명하시오.

 정답

데이터는 현실 세계에서 관찰이나 측정을 통해 수집한 사실이나 값이고, 정보는 데이터를 의사 결정에 활용할 수 있도록 처리한 결과물이다.

따라서 데이터와 정보의 차이는 가공의 유무이다.

 해설

- 데이터(Data): 가공하지 않은 원시 상태의 사실·값
→ 예: 개별 판매 기록 (제품A, 3개, 30,000원)
- 정보(Information): 데이터를 가공·처리한 결과물
→ 예: 월별 매출 보고서, 베스트셀러 목록
- 핵심 관계: Data → [가공/처리] → Information



정보 시스템이 무엇인지 설명하시오.



정답

조직을 운영하기 위해 필요한 데이터를 수집하여 저장해 두었다가 필요할 때 처리하여 유용한 정보를 만들어 주는 수단이다.



해설

정보 시스템(Information System)의 핵심 구성:

- 데이터 수집 → 저장 → 처리 → 정보 생성 → 의사결정 지원
- 대표 예시: 경영 정보 시스템(MIS), 의사 결정 지원 시스템(DSS)
- 정보 시스템의 핵심에는 데이터베이스가 존재합니다.



데이터베이스를 한 문장으로 정의하시오.



정답

데이터베이스는 특정 조직의 여러 사용자들이 공유해서 사용할 수 있도록 통합해서 저장한 운영 데이터의 집합이다.



해설

정의에 포함된 4가지 핵심 키워드:

- 공유 데이터(Shared Data): 여러 사용자가 함께 사용
- 통합 데이터(Integrated Data): 중복을 최소화하여 통합
- 저장 데이터(Stored Data): 컴퓨터가 접근 가능한 매체에 저장
- 운영 데이터(Operational Data): 조직 운영에 반드시 필요한 데이터



데이터베이스의 네 가지 특성을 설명하시오.

 정답

- ① 실시간 접근성: 사용자의 질의에 즉시 응답
- ② 계속 변화: 삽입·삭제·수정으로 항상 최신 상태 유지
- ③ 동시 공유: 여러 사용자가 동시에 같은 데이터 이용 가능
- ④ 내용 기반 참조: 데이터의 주소가 아닌 값(내용)으로 참조

 해설

- 실시간 접근성(Real-time Accessibility): 질의에 대해 실시간으로 응답
- 계속 변화(Continuous Evolution): 삽입(C)·조회(R)·수정(U)·삭제(D)로 동적 변화
- 동시 공유(Concurrent Sharing): 여러 사용자가 동시에 같은 데이터에 접근 가능
- 내용 기반 참조(Content Reference): 물리적 위치가 아닌 데이터 값으로 검색
→ 예: WHERE 학과 = '컴퓨터공학과'



반정형 데이터에 대해 설명하시오.



정답

구조에 따라 저장된 데이터이지만 정형 데이터와 달리 데이터 내용 안에 구조에 대한 설명이 함께 존재한다. 따라서 구조를 파악하는 파싱(parsing) 과정이 필요하다.



해설

반정형 데이터(Semi-structured Data)의 특징:

- 데이터 + 구조 설명이 함께 존재 (self-describing)
- 대표 예시: HTML, XML, JSON 문서, 웹 로그, 센서 데이터
- 파싱(Parsing)을 통해 구조를 파악한 후 정형 데이터로 변환하여 활용 가능
- 정형 데이터보다 저장·검색·분석 처리가 복잡함



범주형 데이터와 수치형 데이터의 차이를 설명하시오.

 정답

범주형 데이터는 성별이나 학년처럼 범주로 구분할 수 있는 값을 가진 데이터이며, 크기 비교와 산술 연산이 불가능하다.

수치형 데이터는 키, 몸무게, 판매량처럼 크기 비교와 산술 연산이 가능한 숫자 값을 가진 데이터이다.

 해설

범주형 vs 수치형 비교:

- 범주형(Categorical): 종류·분류를 나타내는 값 (주로 문자 타입)
 - 명목형(성별, 혈액형): 순서 없음 / 순서형(학점, 등급): 순서 있음
- 수치형(Numerical): 양적 크기를 나타내는 숫자 값
 - 이산형(학생 수, 판매 건수): 셀 수 있는 정수
 - 연속형(온도, 몸무게): 연속 범위의 실수



수고하셨습니다!

박상돈 조교수 | 대전대학교 컴퓨터공학과