

2024-2 기말고사 (데이터베이스)

DB에서 잘못된 설계를 바로잡는 과정: 정규화

이상현상 3개: 삽입 이상, 갱신 이상, 삭제 이상

정규화 과정에서 이상 문제를 해결하려면 연관성이 (높은) 속성들로만 릴레이션을 구성해야 함

릴레이션 속성 사이의 연관성을 평가하기 위한 척도: 함수 종속성

함수 종속 관계를 이해하기 쉽도록 표현한 그림: 함수 종속 다이어그램, FD Diagram

특정 속성이 결정자인 둘 이상의 전체 속성 조합에는 함수 종속이면서 결정자의 어떤 일부 속성에도 함수 종속이 아닐 때: 완전 함수 종속

특정 속성이 결정자인 둘 이상의 전체 속성 조합에도 함수 종속이면서, 결정자의 일부 속성에도 함수 종속일 때: 부분 함수 종속

정규화 과정에서 릴레이션이 만족해야 하는 특정한 함수 종속성의 충족 조건: 정규형

어떤 릴레이션이 제1 정규형이고 기본키에 속하지 않는 모든 속성이 기본키에 **완전 함수 종속**이면: 제 2 정규형

어떤 릴레이션 R이 제2 정규형이고 기본키에 속하지 않는 모든 속성이 기본키에 **이행적 함수 종속**이 아니면: 제 3정규형

어떤 릴레이션에 속한 모든 속성의 도메인이 **원자값**이라면: 제 1정규형

어떤 정보도 손실되지 않게 동등한 릴레이션들로 분해: 무손실 분해

복잡한 식별자 관계에 의한 문제를 해결하기 위한 것: 보이스코드 정규형

성능 저하가 문제될 경우, 분해된 릴레이션을 역으로 통합하는 법: 반정규화

E-R다이어그램의 구성 요소: 개체, 관계, 속성, 링크

E-R다이어그램 한국어: 개체-관계 모델



현실 세계에서 저장할 가치가 있는 데이터와 관련된 독립적 존재: 개체

특정 개체의 개체 인스턴스를 모아놓은 그룹: 개체집합

개체가 갖는 속성 값이 여러 개이면: 다중 값 속성

특정 속성이 갖는 값이 하나이면: 단일 값 속성

둘 이상의 속성으로 이루어져 의미적으로 더 작은 단위로 분해가 가능한 속성: 복합속성

의미적으로 더 이상 분해할 수 없는 속성: 단순속성

실제 값을 저장하는 속성: 저장속성

값을 저장하지 않아도 다른 속성 값에서 계산되거나 유도될 수 있는 속성: 유도속성

각 개체를 유일하게 식별할 수 있는 고유한 값을 갖는 속성: 키 속성

개체와 개체 사이에 맺어지는 연관성을 의미: 관계

관계 유형 4개: 일대일, 일대다, 다대일, 다대다,

차수가 1이며 개체가 자기 자신과 스스로 맺는 관계: 1진 관계 || 순환 관계

차수가 2인 가장 일반적인 관계유형: 2진 관계

3개의 개체가 함께 맺는 관계: 3진 관계

독립적인 두 개체가 대등한 관계를 맺을 때: 비식별 관계

두 개체가 대등한 관계가 아닌 종속적 관계를 맺는 경우: 식별 관계

자신을 고유하게 식별할 수 있는 속성을 갖는 개체: 강 개체

자신을 고유하게 식별할 수 있는 속성을 갖지 못하는 개체: 약 개체

개체 사이의 상하관계: 일반화 관계

기호	의미	기능
	(강) 개체	고유한 키 속성을 갖는 개체
	약 개체	키 속성을 갖지 못하는 개체
	(비식별) 관계	강 개체와 강 개체 사이의 대등한 관계
	식별 관계	강 개체와 약 개체 사이의 종속적 관계
	(단일, 저장, 단순) 속성	의미적으로 분해되지 않는 값 하나를 저장하는 속성
	키 속성	개체를 고유하게 구별짓는 속성
	부분키 속성	키의 일부에 속할 수 있는 속성
	다중 값 속성	값 여러 개를 가질 수 있는 속성
	복합 속성	의미적으로 더 분해 가능한 속성
	유도 속성	다른 속성들로부터 값을 유도 또는 계산 가능한 속성
	전체참여 개체	관계에 빠짐없이 참여해야하는 개체
	부분참여 개체	관계에 참여하지 않을 수도 있는 개체
	일반화 관계	개념을 포함하는 상위 개체와 하위 개체와의 관계

데이터베이스 구조를 생성하는 절차적 과정: 데이터 모델링

DB 설계 과정

요구사항 분석> 개념적 설계> 논리적 설계> 물리적 설계> 구현

요구사항 분석 시 작성하는 것: 요구사항 명세서

개념적 설계 시 작성하는 것: E-R 다이어그램

개체	관계	속성
독립적 존재	종속적 존재	종속적 존재
명사(주어,목적어)로 표현	동사(서술어)로 표현	명사(수식어)로 표현
상위 개념	상위 개념	하위 개념
고유한 명칭(이름,번호) 보유	인위적 명칭(이름,번호) 부여	해당사항 없음
지속적	일시적	지속적/일시적

의사결정을 위한 효율적인 분석 기반을 제공하는 통합 데이터 저장소: 데이터 웨어 하우스

위에서 설명한 용어의 특성 4가지: 주제 지향성, 통합성, 시계열성, 비휘발성

전사적 데이터 웨어하우스를 구축한뒤, 필요시 순차적으로 데이터 마트를 하나씩 추가하여 전체 시스템을 완성하는 방식: 하향식

소규모의 데이터 마트들을 순차적으로 먼저 구축한 후, 나중에 이를 통합하여 전사적 데이터하우스를 완성하는 방식: 상향식

기획과 설계는 하향식 구축방식으로, 실제 구축은 상향식 구축 방식처럼 가장 우선시 되는 주용한 분야부터 진행하는 방법: 절충식

다차원적인 분석을 위해서 데이터 웨어 하우스는 (다차원) 구조로 구성

관계형 DB의 한계점을 극복하기 위해 개발된 분석 전용 DB: 다차원 데이터베이스

큐브의 일부분을 둘로 쪼개어 자신이 원하는 큐브 단면을 분석하는 것: 슬라이싱

큐브중에서 보고자 하는 부분만을 선별함으로써 작은 큐브 형태의 영역을 분석하는 것: 다이싱

큐브의 축 위치를 바꾸어 같은 분석 결과를 확인: 피보팅

다차원 모델의 기본 구조: 데이터 큐브

관계형 스키마 구조에 차원성을 부여하는 설계방법: 차원 모델링

차원 모델링의 대표적 모델: 스타 스키마

비교	스타 스키마	스노우플레이크 스키마
장점	구조가 간단함 적은 조인으로 응답시간이 빠름	정규화로 데이터 중복이 거의 없음 적은 기억장소 요구 유연성 및 확장성 높음
단점	비정규화로 데이터 중복에 따른 데이터의 비일관성 높음 많은 기억장소 요구 유연성 낮음	구조가 복잡함 많은 조인으로 응답시간이 느림

최종 사용자가 직접 다양한 분석 관점의 흐름에 따라 대규모의 다차원 데이터를 온라인에서 동적으로 분석하고 보고서를 만드는 모든 과정: OLAP

OLAP 특성 5가지: 다차원성, 직접성, 대화성, 비정형성, 자동성

항목	관계형 OLAP	다차원 OLAP	혼합형 OLAP
저장소	RDB	MDB	MDB + RDB
모델	차원 모델링	데이터 큐브	차원 모델링과 데이터 큐브
분석 수준	간단한 분석	복잡한 분석	복잡한 분석
조작	읽기	읽기/쓰기	읽기/쓰기
용량	대용량	중 · 소용량	대 · 중용량
속도	느림	빠름	보통
구축 기간	중기	단기	중기
개발 주체	정보부서	사용자	정보부서 + 사용자
확장성	높음	낮음	높음
대상	전사적 데이터웨어하우스	데이터마트	데이터웨어하우스, 데이터마트

OLAP 인터페이스 연산 5가지 (드릴 다운, 롤업, 슬라이스와 다이스, 피봇, 드릴 쓰루)

물리적으로는 분산된 데이터베이스를 컴퓨터 네트워크로 연결하여 하나의 데이터베이스처럼 사용할 수 있도록 저장. 관리하는 데이터베이스: 분산 데이터베이스

분산 데이터 독립성의 투명성 5가지

(위치 투명성, 중복 투명성, 단편화 투명성, 병행 투명성, 장애 투명성)

분산 DB의 모든 데이터 구조와 제약 조건을 정의: 전역 스키마

전역 스키마를 논리적으로 분할한 단편들을 정의: 단편 스키마

각 단편들을 물리적으로 저장할 지역 위치를 정의: 할당 스키마

지역별로 저장하는 데이터의 구조와 제약조건을 정의: 지역 스키마

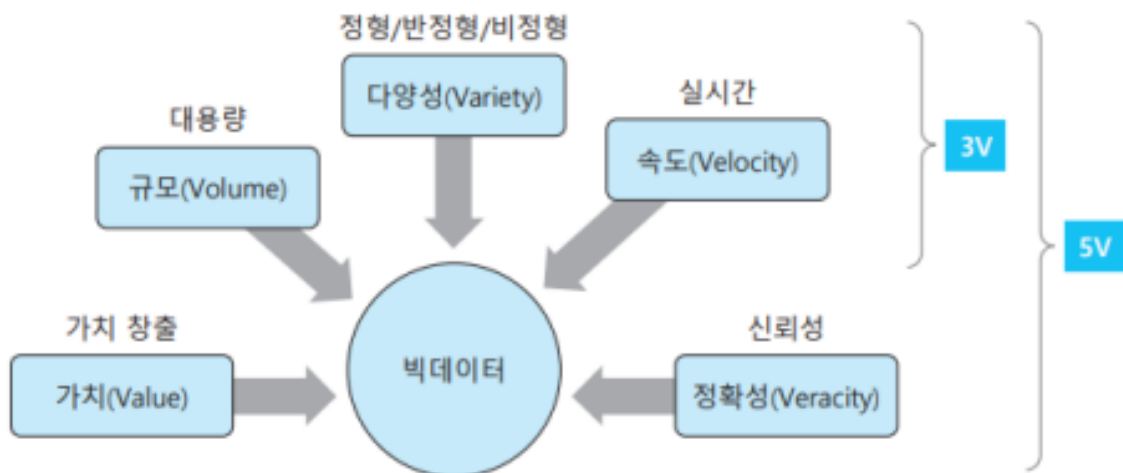
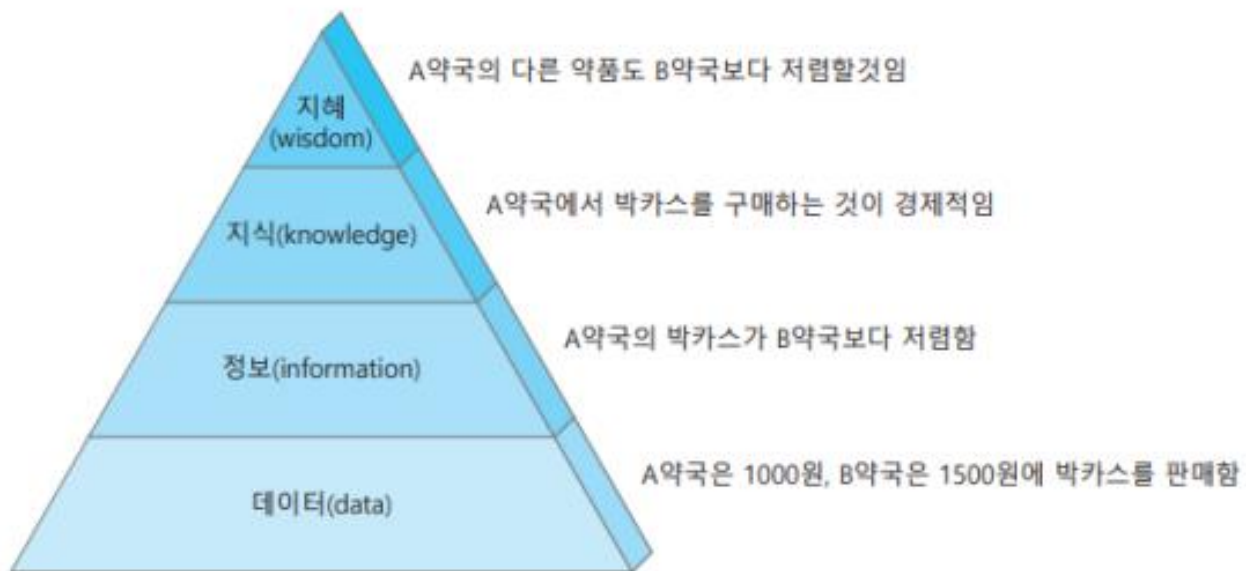
객체, 속성, 메소드, 클래스등 객체지향 개념을 지원함으로써 객체 단위로 데이터를 저장 및 관리하는 DB: 객체지향 데이터베이스

텍스트, 그래픽, 이미지, 비디오, 오디오 등이 복합적으로 구성된 다양한 미디어를 효율적으로 저장·관리하기 위한 데이터베이스: 멀티미디어 데이터베이스

다차원 공간 안의 객체와 객체간의 공간 관계를 공간적 특성과 함께 저장하고 관리하는 데이터베이스: 공간 데이터베이스

현대 사회의 복잡성을 분석하고 예측하기 위해 꼭 필요한 데이터 관리 기술: 빅데이터 기술

기존 하드웨어나 소프트웨어의 처리 능력을 넘어서는 대용량 데이터의 집합: 빅데이터



고정된 필드로 저장되지만 데이터의 형식과 구조가 변경될 수 있는 데이터: 반정형 데이터

음성이나 동영상, 블로그, SNS 메시지 등과 같이 정해진 형식이나 구조가 없는 정형화되지 않은 데이터: 비정형데이터

정해진 형식과 구조에 따라 고정된 필드에 저장되도록 구성된 데이터: 정형 데이터

데이터웨어하우스나 데이터마이닝에서 수집된 데이터를 분석이 쉽도록 통일된 형태로 변환하는 과정: ETL

분류 기준	종류
데이터 유형	정형 데이터, 반정형 데이터, 비정형 데이터
데이터 특성	정량적 데이터, 정성적 데이터
데이터 위치	내부 데이터(조직 안의 데이터), 외부 데이터(조직 밖의 데이터)
수집 빈도	일회성, 반복성
수집 주기	배치(일괄), 실시간
데이터 형식	RDB, 스프레드시트, HTML, XML, 센서 데이터, 웹로그, 이미지, 동영상, 문서 등

데이터 처리 방식 3개: 대화형, 일괄형, 실시간형

아파치: 하둡

HDFS: 하둡 분산파일 시스템

대용량 데이터의 분산 병렬처리를 위해 구글에서 만든 소프트웨어 프레임워크: 맵리듀스

MapReduce 처리단계: map단계 > shuffle 단계> reduce 단계

미리 정의된 스키마를 가진 구조화된 정형 데이터를 저장: SQL 데이터베이스

고정된 스키마 없이 자유롭게 필드를 추가할 수 있어 스키마가 유연하거나 생략될 수 있는 반정형 및 비정형데이터를 저장: NoSQL 데이터베이스

주로 내부 구조도 없는 비정형 데이터를 분산 파일 형태로 확장 가능하도록 저장: 분산 파일 시스템

NoSQL DB의 장점 4가지: 유연성, 확장성, 경제성, 가용성

NoSQL 데이터베이스 중 가장 단순하고 기본적인 형태: 키-값 데이터베이스

위 데이터베이스의 발전된 형태, NoSQL 데이터베이스 중 가장 인기: 문서 데이터베이스

구조면에서 가장 복잡한 유형, 관계형 데이터베이스와 비슷함: 컬럼 패밀리 데이터베이스

데이터를 데이터 간의 관계와 함께 표현하는 특수한 유형: 그래프 데이터베이스

기법	적용 사례
회귀 분석	고객의 성별이 구매 상품의 선택에 어떤 영향을 미치는지 분석
분류	고객이 어떤 구매성향을 갖는 고객 집단에 속하는지 분석
연관규칙 발견	상품 A를 구매한 고객이 상품 B를 같이 구매하는지 분석
기계 학습	상품 구매 패턴을 분석하여 고객에게 새로운 상품을 추천
감정 분석	신규 상품에 대한 고객의 반응을 분석
소셜미디어 분석	특정 상품에 대한 소셜 미디어에서의 평판을 분석
유전 알고리즘	특정 상품에 대한 최적의 마케팅 전략을 수립

데이터 집합으로부터 의미있는 규칙이나 패턴을 찾아냄: 데이터 마이닝

정형 데이터마이닝 기법: 분류, 군집, 연관, 예측

데이터 마이닝 알고리즘 3개: 의사결정 나무, 인공 신경망, 사례기반 추론

비정형 데이터마이닝 기법: 텍스트 마이닝, 웹 마이닝, 오피니언 마이닝, 소셜 마이닝

빅데이터 시각화 분류 5가지: 시간 시각화, 분포 시각화, 관계 시각화, 비교 시각화, 공간 시각화