

DB 정리

시험 3일차

pdf ready

html ready

- [Latex](#) 문법을 사용합니다.

DB 시험 범위

1. 교과서 6, 7단원 (현재 페이지)
2. 유인물 1 ~ 7 (현재 페이지)
3. [회원가입 PPT](#)

시험 정리

6단원. 데이터모델링

1. 데이터베이스 생명 주기



데이터베이스는 최초 사용자의 요구에 의해 구축되어 사용되다가 필요에 따라 개선 또는 다시 구축되어 사용한다.

이러한 데이터베이스의 생성과 운영에 관련된 특징을 **데이터베이스 생명 주기**라고 한다.

생명주기 설명

1. 요구사항 수집 및 분석

사용자들의 요구사항을 듣고 분석하여 데이터베이스 구축의 범위를 정하는 단계이다.

모델링 과정 : 현실 세계의 대상 및 사용자의 요구사항을 정리하고 분석한다.

2. 설계

개념적 설계(중요 개념을 구분) ▶ 논리적 설계(각 개념을 구체화) ▶ 물리적 설계(데이터베이스 생성 계획에 따라 체계, 인덱스 등을 생성)

개념적 모델링: 개체를 추출하고 각 개체들 간의 관계를 정의하여 ERD를 만드는 과정까지를 말한다.

논리적 모델링 과정:

- i. 실제 필요한 모든 속성을 추출한다. (실제 DB 쿼리에 적용될 속성)

ii. 정규화를 수행한다. (7단원)

iii. 데이터의 표준화를 수행한다. (데이터의 형태 및 크기 등을 확정한다)

3. 구현

설계 단계에서 생성된 스키마를 실제 DBMS에 적용하여 DB를 만듦.

또한 관련 소프트웨어에 설계한 데이터베이스를 적용하여 서비스를 제공할 수 있도록 프로그램을 완성한다.

4. 운영

구현된 데이터베이스를 기반으로 소프트웨어를 구축하여 서비스를 제공

5. 감시 및 개선

데이터베이스가 지속적으로 운영될 수 있도록 변경 및 유지보수를 한다.

2. 학습지 1~3

데이터베이스 업무시나리오에서 엔티티 도출 (학습지 1~2)

도출 방법

1. 시나리오에서 명사를 구분한다.
2. 개념이 불분명한 것, 광범위한 것은 제거한다.
3. 엔티티타입의 특성이거나 속성의 값은 제거한다.
4. 포괄적인 업무 프로세스에 해당되는 명사는 제거한다.
5. 중복되는 명사는 제거한다.
6. 누락된 엔티티타입을 유추한다.

ex) 인터넷에 회원을 등록하는 정보로는 회원번호, 주민번호, 주소, 전화번호, 전자메일, 휴대폰번호, 결제 방법이 있다.

엔티 티	엔티티 설명	관련 속성
회원	회원 등록 정 보	회원번호, 주민번호, 전화번호, 전자메일, 휴대폰번호, 결 제 방법

데이터 모델링 - 속성(Attribute) (학습지 3)

속성 : 정보를 나타내는 최소의 단위, 엔티티의 성질, 수량, 상태, 특성 등을 나타냄

속성의 유형

- **기본 속성** : 해당 엔티티가 원래 가지고 있는 속성
- **설계 속성** : 원래 업무에는 존재하지 않지만 시스템의 효율성을 위해 임의로 추가되는 속성 (인덱스)
- **파생 (추출) 속성** : 다른 속성으로 계산이나 변형되어 생성되는 속성 (ex: 전과목 합계점수)

속성 정의 시 고려 사항

1. 엔티티가 관리할 특성들인가?

속성의 111 원칙

- 엔티티 타입 내에서 **한 속성을 한 시점에 한 값**을 가져야 한다. ※100% 시험에 나옴

파생 (추출) 속성의 너무 많은 값으의 중복은 **무결성 문제**가 발생할 수 있음

3. 학습지 4 (ER모델, ER다이어그램)

ER 모델은 데이터 모델링 과정 중 개념적 모델링에 사용하는 모델이다. 세상의 사물을 개체와 개체 간의 관계로 표현한다.

ER 모델은 개체와 개체 간의 관계를 **ER 다이어그램 (ERD)** 으로 표현한다.

1. 개체와 개체 타입

개체는 사람, 사물, 장소, 개념, 사건과 같이 유무형의 정보를 가지고 있는 독립적인 실체를 말한다.

개체는 비슷한 속성을 가진 **개체 타입**을 구성하며, 개체 집합으로 묶인다.

개체 타입은 프로그래밍 언어의 타입으로 생각하면 된다.

2. ER 다이어그램 (ERD)

요구 분석 상황을 **그림**으로 그려 **관계**를 도출한 것

1. 개체 관계 모델 (Entity Relationship Model)

- 구조화된 데이터에 대한 일련의 표현
- 설계 기법 중 하나가 개체-관계 모델링

2. 데이터베이스에서 관계란?

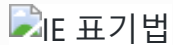
- 두 엔티티 사이의 논리적인 관계
- 두 엔티티간의 업무 연관성)

3. ERD를 그리는 2가지 방법

- 피터 첸 표기법 (책에는 나와있으나 학습지에는 없다)



- IE 표기법 (학습지에 존재하는 표기법)



관계 이름짓기 (위 그림에서)

- 학생은 학과에 소속된다
- 학과는 사원을 포함한다

ERD 표기 방법은 학습지 참고

4. 관계의 차수 - 두 엔티티간 관계 사이의 참여 수를 표현

- 1:1 관계
 - 하나의 구매신청은 하나의 구매주문을 신청한다
 - 한 개의 구매주문에 대해 한 개의 구매 신청서 내용을 작성한다
- 1:N 관계
 - 한 부서에 여러명의 사원이 포함된다
 - 한명의 사원은 한 부서에 소속된다
- N:N 관계
 - 한 주문은 여러개의 제품을 포함한다
 - 한 제품은 여러개의 주문에 주문된다

※ 0...1, 1, 1...N, 0...N 4가지 경우에 기호 외워놓을 것

5. 관계의 종류

1. 기본 관계

엔티티와 엔티티가 독립적으로 분리되어 있으며 상호간의 한 가지 관계만 존재하는 형태

2. 재귀 관계

하나의 엔티티 내에서 엔티티와 엔티티가 관계를 맺고 있는 형태

3. 베타적 관계 (원 안에 X 기호)

슈퍼타입의 엔티티가 반드시 하나의 서브타입에는 속하는 형태

대구 소프트웨어 고등학교 2학년 학생은 소프트웨어과, 임베디드과 반드시 하나에 속해야 한다.

슈퍼타입: 대소고 학생 / 서브타입: 소프트웨어과 학생, 임베디드과 학생

4. 포함 관계 (원 기호)

슈퍼타입의 엔티티가 두 개 이상의 서브타입에 포함될 수 있는 관계

접수를 방문접수, 인터넷 접수 두 가지로 할 수 있다면 인터넷 접수를 한 사람은 다시 방문접수를 할 수 있다.

슈퍼타입: 접수 / 서브타입: 방문접수, 인터넷 접수

5. 식별 관계 (실선)

i. 부모 없이는 자식이 존재할 수 없는 관계이다

ii. 부모의 **식별자**가 자식의 **식별자**로 상속된다

여기서 식별자는 PK(Primary Key)로 보면 된다

6. 비식별 관계 (점선)

i. 부모 없이도 자식이 존재할 수 있는 관계이다

ii. 부모의 **식별자**가 자식의 **속성**으로 상속된다.

회원 없이도 주문이 가능할 수 있을 때 (비회원 주문이 존재), 회원과 주문은 비식별 관계이다.

3. 학습지 5 (관계 - RelationShip)

마당서점 엔티티타입 도출

엔티티명	속성
도서	도서번호, 도서이름, 출판사이름, 도서단가
출판사	출판사이름, 담당자이름, 전화번호
고객	고객번호, 고객이름, 주소, 전화번호
주문	고객번호, 도서번호, 주문일자, 주문금액

관계정의서 작성 (정답은 아닐 수 있다.)

하나/각각	기준 엔티티	하나/여러	관련 엔티티	선택/필수	관계 참여 방법
각각의	도서는	하나의	출판사에	항상	속할 수 있습니다
각각의	출판사는	여러개의	도서를	선택	소속되어 있습니다
하나의	주문은	하나의	고객이	항상	소속되어 있습니다
하나의	주문은	하나의	도서를	항상	소속되어 있습니다
하나의	도서는	여러개의	주문에	선택	속할 수 있습니다
하나의	고객은	여러개의	주문에	선택	속할 수 있습니다

4. 학습지 6 (이상현상, 함수 종속성)

이상현상 : 잘못 설계된 테이블에 삽입, 삭제, 수정의 데이터 조작시 발생하는 현상

- 삽입이상: 투플을 삽입할 때 `NULL` 값이 입력
- 연쇄삭제가 됨
- 데이터의 일관성이 훼손

이상현상이 발생하지 않게 테이블을 분리하는 것이 필요하다

함수 종속성

릴레이션 R 과 속성의 집합 X, Y 가 있을 때 X 각각의 값이 Y 의 값 한 개와 대응이 될 때,

X 는 Y 를 함수적으로 종속한다

라고 하고 $X \rightarrow Y$ 라고 표기한다.

또한 X 는 결정자, Y 는 종속자라고 한다. (BCNF 설명에서 나온다.)

ex) 학생수강성적 릴레이션 (밑에서 많이 쓰인다)

학생번호	학생이름	강좌이름	강의실	성적
501	박지성	데이터베이스	공학관101	3.5
401	김연아	데이터베이스	공학관101	4.2
501	박지성	자료구조	공학관111	4.5

종속 관계에 있는 속성값들

학생번호 \rightarrow 학생이름	강좌이름 \rightarrow 강의실	(학생번호, 강좌이름) \rightarrow 성적

함수 종속성 규칙

적용규칙	함수 종속성 규칙
부분집합 규칙	if $Y \subseteq X$ then $X \rightarrow Y$
증가 규칙	if $X \rightarrow X$ then $XZ \rightarrow YZ$
이행 규칙	if $X \rightarrow Y$ and $Y \rightarrow Z$ then $X \rightarrow Z$
결합 규칙	if $X \rightarrow Y$ and $X \rightarrow Z$ then $X \rightarrow YZ$
분해 규칙	if $X \rightarrow YZ$ then $X \rightarrow Y$ and $X \rightarrow Z$
유사이행 규칙	if $X \rightarrow Y$ and $WY \rightarrow Z$ then $WX \rightarrow Z$

학습지 7 (정규화(NF))

정규화 : 이상현상이 발생하는 릴레이션을 분해하여 이상현상을 제거하는 과정

NF는 Normal Form의 약자이다.

정규형의 7가지 종류

1. 제 1정규형(1NF): 릴레이션의 모든 속성 값이 원자값일 때 릴레이션의 조건을 만족하는 것
- 비정규 릴레이션

이름	취미
김연아	영화, 쇼핑 (원자값이 아니다!)

■ 1차 정규형으로 변환

이름	취미
김연아	영화 (원자값이다!)
김연아	쇼핑

2. 제 2정규형(2NF): 제 1정규형 + 후보 키가 아닌 속성이 기본키에 완전 함수 종속일 때

후보 키 : 슈퍼 키 중 더이상 줄일 수 없는 형태를 가진 키.

슈퍼 키 : 행(튜플, row)를 식별해낼 수 있는 속성들

원래 학습지에 있는 예제를 쓰려고 했지만 학습지에 있는 2정규형과 3정규형의 구분이 애매해서 차이점이 드러나게 작성하겠다.

완전 함수 종속: 한 릴레이션에서, $A \rightarrow B$ 종속성이 성립 시, B가 A의 속성 전체에 함수 종속하고 부분 집합 속성에 함수 종속하지 않는 경우

즉 후보 키가 아닌 속성이 기본키의 모든 속성을 같이 종속해야 한다.

○ 1차 정규형 (학생수강강좌 릴레이션)

학생번호	학과	강좌이름	등록금	성적
501	컴퓨터 공학	데이터베이스	300	B+
401	스포츠 경영학	데이터베이스	250	A
501	컴퓨터 공학	자료구조	300	B

여기서 후보키는 학생번호, 강좌이름이다.

이 때, 학과랑 등록금은 학생번호 속성 하나에만 종속하기 때문에 완전 함수 종속이 안된다.

즉 학과랑 등록금을 별도의 릴레이션으로 나눠 주는것이 제 2정규형의 역할이다.

○ 2차 정규형으로 변환

■ 학생 릴레이션

학생번호	학과	등록금
501	컴퓨터 공학	300
401	스포츠 경영학	250

■ 강좌 릴레이션

학생번호	강좌이름	성적
501	데이터베이스	B+
401	데이터베이스	A
501	자료구조	B

2차 정규형이 완성된 모습이다. 하지만 여기서 학과, 강좌이름, 등록금에 대해 **이상현상**이 발생한다.

- 삽입 이상 : 새로운 학과가 생기면 학생번호가 **NULL** 이 된다.
- 갱신 이상 : 컴퓨터 공학의 등록금이 300에서 400으로 오르면 컴퓨터 공학 학과에 있는 학생들의 등록금을 전부 바꿔주지 않으면 데이터 불일치 문제가 발생한다.
- 삭제 이상 : 501번 학생이 자퇴를 하면 컴퓨터 공학에 대한 정보가 같이 사라진다.

위와 같은 문제점을 해결하기 위해 **제 3정규형**이 등장하였다.

3. 제 3정규형(3NF): 제 2정규형 + 기본키가 아닌 속성이 기본키에 **비이행적으로 종속**할 때

이행적 함수 종속: 함수 종속성 규칙의 **이행 규칙**을 뜻한다.

테이블 내에 모든 속성이 **기본 키**에만 의존하며 다른 **후보 키**에 의존하지 않는다.

기본 키: 테이블 내에 여러 후보 키가 존재할 때, 기본 키는 후보키들 중 하나를 선택한 것이다.

즉 $[X, Y, Z]$ 릴레이션에서 $X \rightarrow Y, Y \rightarrow Z$ 일 때 $X \rightarrow Z$ 가 일어나면 $[X, Y], [Y, Z]$ 릴레이션으로 분리한다.

제 2정규형에서 썼던 예제를 제 3정규형으로 다시 고쳐보겠다.

기존에 학생 릴레이션을 살펴보면

학번 \rightarrow 학과 ($X \rightarrow Y$)

학과 \rightarrow 등록금 ($Y \rightarrow Z$)

학번 \rightarrow 등록금 ($X \rightarrow Z$)

가 성립한다. 그러면 [학번, 학부], [학과, 등록금] 으로 분리해 보자.

- 학생 학과 테이블

학번	학과
501	컴퓨터 공학
401	스포츠 경영학

■ 학과 테이블

학과	등록금
컴퓨터 공학	300
스포츠 경영학	250

이렇게 변경하면 제 3정규형을 만족한다. 이 릴레이션에는 이상 현상이 발생하지 않지만 후보키를 여러개 가지고 있는 릴레이션에서는 제 3 정규형을 만족하더라도 이상현상이 발생할 수 있다.

이를 해결하기 위해 **BCNF**라는 정규형을 사용한다.

4. BCNF(Boyce-Codd Normal Form): $X \rightarrow Y$ 가 성립될 때 모든 결정자 X 가 후보키일 때

결정자란? 각 X 값에 대해 반드시 한개의 Y 가 대응할 때 X 는 **결정자** 이고 Y 는 **종속자** 이다.

실질적으로 많이 일어나지 않는 경우이며 일어나는 경우의 수는 $AB \rightarrow C, C \rightarrow B$ 일 때 BCNF가 일어난다.

더 풀어서 얘기하자면 **일반 컬럼이 후보키를 결정하는** 경우에 발생한다.

학생번호	특강이름	담당교사
2101	Unity	박주희
2201	Unity	박주희
2302	정보처리기능사	이하늘
2101	정보처리기능사	마호돌

특수한 점은 한 특강을 여러 담당교사가 한다는 점이다. (**중요**)

위 예제에서 정의를 해줘야 하는 부분이 있다.

- (학생번호, 특강이름)이 기본키이다.
- 담당교사는 오직 한개의 특강만을 가르친다
- 담당교사가 결정이 되면 특강이름이 결정이 된다.

이 상황에서 함수 종속성은

- i. (학생번호, 특강이름) \rightarrow 담당교사
- ii. 담당교사 \rightarrow 특강이름

이다.

제 2정규형이 성립하는 이유는

후보키는 (학생번호, 특강이름), (학생번호, 담당교사)가 존재하기 때문에 후보키가 아닌 속성이 존재하지 않는다.

그러면 제 3정규형을 보자.

기본 키는 (학생번호, 특강이름) 이고 담당교사는 (학생번호, 특강이름)에만 의존한다.

하지만 여기서는 이상현상이 발생한다.

- i. 삽입 이상 : 홍길동이 담당하는 해킹 수업이 열렸을 때 학생번호는 `NULL` 로 삽입이 된다.
- ii. 갱신 이상 : 박주희 쌤이 담당하는 강의가 바뀌게 될 경우 수강생의 수만큼 갱신해 줘야 하므로 데이터 불일치 문제가 발생할 수 있다.
- iii. 삭제 이상 : 2302번 학생이 자퇴를 하면 이하늘 쌤이 담당하는 정보처리기능사 반이 없어진다.

이를 분해하는 방법은

- i. 일반 컬럼이 후보키를 결정하는 경우 $X \rightarrow Y$ 를 찾는다.
- ii. 두 개의 릴레이션으로 분해한다.
 - XY 로 구성된 릴레이션 하나
 - X 와 나머지 속성들로 구성된 릴레이션 하나

이렇게 분리하면 된다.

그러면 나눠보자.

- 학생번호-담당교사 릴레이션

학생번호	담당교사
2101	박주희
2201	박주희
2302	이하늘
2101	마호돌

◦ 담당교사-특강이름 릴레이션

담당교사	특강이름
박주희	Unity
이하늘	정보처리기능사
마호돌	정보처리기능사