

1과목 데이터모델링의 이해

데이터 모델링

- 정보시스템을 구축하기 위한 데이터 중심의 분석 및 설계 기법
- 현실 세계의 데이터를 약속된 표기법으로 표현
- 업무 내용을 정확하게 분석하고 설명, 단지 시스템 구현만을 위한 데스크가 아님

데이터 모델링 주요 개념

- 업무가 관여하는 어떤 것 (Things)
- 업무가 관여하는 어떤 것의 성격 (Attribute)
- 업무가 관여하는 어떤 것의 관계 (Relationships)

데이터 모델링 유의점 및 특성

- 유의점 : 중복, 비유연성, 비일관성
- 특징 : 추상화, 단순화, 명확성

ERD 작성 절차

엔티티 도출 -> 엔티티 배치 -> 관계 설정 -> 관계명 작성 -> 관계 참여도 표현 -> 관계 필수 여부 표현

데이터 모델링 3가지 관점

- 데이터 관점: 비즈니스 프로세스에서 사용되는 데이터를 의미, 구조분석
- 프로세스 관점 : 비즈니스 프로세스에서 수행하는 작업을 의미, 업무 시나리오 분석
- 데이터와 프로세스 : 프로세스와 데이터 간의 관계를 의미, CRUD 매트릭스

데이터 모델링 3단계

1. 개념적 모델링 : 가장 높은 수준의 추상화
 2. 논리적 모델링 : 정규화
 3. 물리적 모델링 : 실제 DB 설계
- * 논리 모델의 외래키는 물리 모델에서 반드시 구현할 필요 없음

데이터 베이스 스키마

- 외부 스키마 : 사용자 관점, 뷰
- 개념 스키마 : 통합된 사용자, 설계적 관점, 통합 DB
- 내부 스키마 : 내부(물리적 저장 구조)

독립성

논리적 독립성 - 개념 스키마가 변경되어도 외부 스키마에 영향 X

물리적 독립성 - 물리스키마가 변경되어도 논리스키마에 영향 X

- 파일 저장구조의 변경이 논리 스키마와 응용 프로그램에 영향 X

관계 표현 방식

존재 관계 / 행위 관계 - ERD에서 관계를 연결할 때에는 구분 없이 단일 표기

관계 표기법 - 관계명, 관계차수, 선택성(필수관계, 선택관계)

UML 클래스 다이어그램

연관 관계 - 항상 이용하는 존재적 관계 (실선)

의존 관계 - 상대방의 클래스의 행위에 의해 관계가 형성되는 행위적 관계 (점선)

엔터티

- 두 개 이상의 인스턴스 집합과 두 개 이상의 속성이 필요
- 다른 엔터티와 최소 1개 이상의 관계 필요
- 유일한 식별자에 의해 식별이 가능해야 함
- 엔터티는 업무 프로세스에 이용되어야 함, 업무에서 관리되어야 하는 집합
- 순수 개체이거나 행위 집합
- 저장에 필요한 어떤 것으로, 개념, 사건, 사람, 장소 등과 같은 명사이다.

두 개의 엔터티 사이에서의 관계를 도출할 때 확인해야 할 사항

- 두 엔터티 사이에 관심 있는 연관규칙이 존재하는가?
- 두 엔터티 사이에 정보의 조합이 발생하는가?
- 업무기술서, 장표에 관계연결에 대한 규칙이 서술되어 있는가?
- 업무기술서, 장표에 관계연결을 가능하게 하는 동사(Verb)가 있는가?

엔터티의 종류

키 엔터티 - 다른 엔터티의 도움없이 독립적으로 생성

메인 엔터티 - 업무처리에서 중심이 되는 엔터티

행위 엔터티 - 업무처리를 하는동안 발생하는 엔터티, 엔터티 종류 중 데이터 값이 가장 많을 것으로 예상

교차 엔터티 - M:N 관계를 해소하기 위해 인위적으로 만들어진 엔터티

발생시점에 따른 분류

기본 엔터티 - 그 업무에 원래 존재하던 정보, 다른 엔터티의 영향 X, 독립적으로 생성

중심 엔터티 - 기본 엔터티로부터 발생되고 그 업무에서 중심적인 역할, 기본 엔터티와 행위 엔터티의 중간

행위 엔터티 - 2개 이상의 엔터티로부터 발생

유무형에 따른 엔터티

유형 엔터티 - 업무에서 도출, 지속적으로 사용, 물리적으로 존재

개념 엔터티 - 물리적 형태 X

사건 엔터티 - 비즈니스 프로세스를 실행하며 생성

속성

- 사물의 성질, 특징, 본질적 성질을 의미
- 각각의 속성은 한 개의 속성값을 가져야 함, 1개 이상의 속성값을 가지면 정규화 필요

속성의 원자성

- 각 엔터티의 인스턴스가 해당 속성에 대해 단일하고 명확한 값을 가짐

특성에 따른 속성의 분류

기본속성 - 업무로부터 추출된 모든 속성

설계속성 - 기본 속성 외에 업무를 규칙화하기 위해 새로 만들어지거나 기본 속성을 변형해

만드는 속성

파생속성 - 다른 속성에 의해 만들어지는 속성 ex) 합계, 평균

- 자신의 속성이 없이도 다른 속성을 이용하여 결과를 도출할 수 있음
- 데이터를 조회할 때 성능을 빠르게 하기 위해 원래 속성의 값을 계산하여 저장할 수 있도록 만드는 속성

분해여부에 따른 속성

단일속성 - 하나의 의미로 구성

복합속성 - 여러 개의 의미로 구성

다중값 속성 - 여러 개의 값이 들어가는 속성

도메인 - 속성이 가질 수 있는 값의 범위

식별자

주식별자 - 대표성, 유일성, 불변성, NOT NULL

보조식별자 - 유일성은 있으나 대표성 없음(참조 불가)

본질식별자 - 업무적으로 의미 부여됨

외부식별자 - 다른 엔티티로부터 상속됨

인조식별자 - 업무적으로 만들어지지 않지만 원조식별자가 복잡한 구성을 가지고 있기 때문에
인위적으로 만들어진 식별자

- 인조식별자를 사용하면 데이터의 중복이 발생할 수 있고 별도의 인덱스 생성 등이 필요

관계에 따른 식별자

식별 관계 - 자식이 부모 식별자 포함, 강한 연결 관계, 실선

비식별 관계 - 자식이 부모 식별자를 일반 속성으로 상속, 약한 연결 관계, 점선

생성여부에 따른 식별자 분류

내부식별자 - 엔티티 내부에서 스스로 생성되는 식별자

외부식별자 - 다른 엔티티의 관계로 인하여 만들어지는 식별자

대표성에 따른 식별자 분류

주식별자 - 최소성, 유일성, 불변성, 대표성, NOT NULL을 만족하고 엔티티를 대표해야 한다.

보조식별자 - 대표성 X, 참조관계를 가지지 X

대체여부에 따른 식별자 분류

본질 식별자 - 비즈니스 프로세스에서 만들어지는 식별자

인조 식별자 - 인위적으로 만들어지는 식별자, 유일한 값을 만들기 최대한 범용적인 값 사용

조인, 관계

조인(Join) - 데이터를 결합하는 방법으로, 식별자를 사용하여 테이블을 연결

관계(Relationship) - 테이블 간의 구조적 연결을 정의하며, 식별자를 통해 매핑키를 활용하여
데이터의 연관성을 유지

존재관계 - 엔티티 간의 상태

행위관계 - 엔티티 간의 어떤 행위가 있는 것

선택도와 카디널리티

선택도 - 특정 조건에 해당하는 레코드 비율

카디널리티 = 선택도 * 전체 레코드 수

NULL - 아직 정의되지 않은 값, 값의 부재 (0이나 공백과는 다름)

- NULL과의 모든 비교(IS NULL제외)는 알 수 없음(Unknown)을 반환

정규화

정규화 절차	설명
제 1 정규화	속성의 원자성을 확보
제 2 정규화	부분 함수 종속 제거
제 3 정규화	이행 함수 종속 제거 <u>주식별자와 관련성이 제일 낮음</u>
BCNF	다수의 후보키 분리
제 4 정규화	다치 종속 제거
제 5 정규화	조인 종속 제거

반정규화

- 성능 향상을 위해 중복 허용, 조인 최소화
- 적용 시점 : 범위 조회가 많거나 집계 정보가 자주 필요한 경우

반정규화 절차

반정규화 절차	설명
대상 조사 및 검토	데이터 처리 범위, 통계성 등을 확인해서 반정규화 대상 조사
다른 방법 검토	반정규화를 수행하기 전에 다른 방법이 있는지 검토 뷰 또는 클러스터링 기법 적용
반정규화 수행	테이블, 속성, 관계 등을 반정규화

중복관계 추가 - 데이터 무결성을 깨뜨릴 위험을 갖지 않고서도 데이터 처리의 성능을 향상시킬 수 있는 방법

슈퍼타입과 서브타입 변환방법

종류	설명
OneToOne Type	슈퍼타입과 서브타입을 개별 테이블로 도출
Plus Type	슈퍼타입과 서브타입 테이블로 도출
Single Type	슈퍼타입과 서브타입을 하나의 테이블로 도출

분산DBMS 장단점

장점

- 신뢰성, 가용성 향상
- 병렬 처리로 빠른 응답

단점

- 관리 및 보안 복잡
- 데이터 무결성 관리 어려움
- 설계가 복잡함