

제 1 장

정규화 과정

학습 목표

- ▶ 관계형 데이터베이스의 정규화에 대해서 설명할 수 있습니다.



Written By Seo Y.H

◆ 논리적 모델과 데이터베이스의 오브젝트와 용어 비교

논리적 모델	개체(Entity)	속성(Attribute)	튜플(Tuple)	식별자	대표 식별자
데이터베이스	테이블	컬럼	행	Unique Key	Primary Key

▶ 정규화 과정이란 ?

고유한 식별자를 가지는 모든 개체(테이블)에 대해 더 이상 분리할 수 없는 상태로 나누는 과정을 말한다. 정규화 과정을 통해 불필요하게 중복된 데이터 제거 작업을 하게 된다.

정규화를 통한 개체의 분리는 관계가 있는 개체간의 참조 관계로 데이터를 유지 보수하게 된다. 정규화 수준을 높이면 데이터량이 줄고 데이터 갱신 속도가 빠르며 같은 자료가 여러 개체에 분산된 것보다 일관성을 유지하기 쉽다. 그러나 관계된 여러 속성을 동시에 조회하는 작업을 수행하기 위해서는 여러 개체를 JOIN 하여 작업해야 하는 어려움이 있을 수 있다.

또 갱신 작업에 있어 여러 개체에 동시에 반영해야 하는 작업의 경우는 명시적으로 트랜잭션 처리해야 하는 수고가 따른다. 따라서 먼저 정규화 과정을 수행하고 수행속도나 프로그램을 구현하기 어려운 경우에는 비정규화(역정규화) 과정으로 문제를 보완하게 된다.

정규화 과정에는 제 1 정규화, 제 2 정규화, 제 3 정규화, 제 4 정규화와 역정규화 과정이 있다.

▶ 제 1 정규화(1NF, First Normal Form)

반복되는 속성들을 다른 개체로 나누어 분리한다.

즉, 또 다른 개체로 판단할 수 있는 속성들을 분리하고 각 개체의 속성들의 유일한 식별자를 가지게 한다. **부모테이블의 식별자는 자식테이블의 외부키(Foreign Key)로 전이 되어짐.**

▶ 제 2 정규화(2NF, Second Normal Form)

모든 속성은 식별자에 직접적으로 의존적이어야 하며 이에 해당되지 않는 속성을 분리한다.

▶ 제 3 정규화(3NF, Third Normal Form)

식별자 이외의 속성은 식별자가 아닌 다른 속성에 종속적이지 않아야 한다.

식별자 이외의 속성에 종속적인 경우에 분리해 낸다.

▶ 제 4 정규화(4NF, Fourth Normal Form)

의존적인 다 대 다(Many to Many)관계를 제거한다.



Written By Seo Y.H

<<예제>>

성적에 관련한 다음과 같은 자료가 있다.

같은 과목에 대해 여러 교수가 동시에 진행하며 한 교수는 여러 과목을 강의할 수 있다.

과목번호	과목명	교수번호	교수명	강의실코드	강의실설명	학번	학생명	점수
DBO101	오라클기초	21	서영학	A301	공대301	9936720	김유신	B
DBO101	오라클기초	21	서영학	A301	공대301	9936721	엄정화	A
DBO101	오라클기초	21	서영학	A301	공대301	9936723	장나라	A+
DBO101	오라클기초	22	임훈	A302	공대301	9936722	김국진	A
DBO102	오라클고급	21	서영학	B201	외대201	9936723	장나라	A+
DBO102	오라클고급	22	임훈	B202	외대202	9936724	강감찬	B
DBM101	MS-SQL기초	21	서영학	C401	사대401	9936720	김유신	C+
DBM101	MS-SQL기초	22	임훈	C402	사대402	9936721	엄정화	A
DBM102	MS-SQL고급	21	서영학	C501	사대501	9936723	장나라	A+
DBM102	MS-SQL고급	22	임훈	C502	사대502	9936724	강감찬	B
OSM101	WIN2000기초	21	서영학	D201	자대201	9936720	김유신	A
OSM101	WIN2000기초	21	서영학	D201	자대201	9936721	엄정화	A+
OSM102	WIN2000고급	23	이민종	D301	자대301	9936723	장나라	A+
OSM102	WIN2000고급	23	이민종	D301	자대301	9936724	강감찬	C
OSL101	LINUX	24	이종오	D401	자대401	9936720	김유신	C
OSL101	LINUX	25	이창익	D402	자대402	9936721	엄정화	A



Written By Seo Y.H

※ 아래의 도표에서 음영은 식별자(Primary Key)를 나타내는 것이다.

▶ 제 1 정규화(1NF, First Normal Form)

위의 데이터에 대해 반복되는 과목번호, 과목명, 강의실코드, 교수번호, 교수명에 대해 다른 개체로 분리한다.

분리는 개설과목과 점수라는 개체(Entity)로 나눌 수 있을 것이다.

- ◆ 개설과목의 식별자 : 과목번호, 교수번호
- ◆ 성적의 식별자 : 과목번호, 교수번호, 학번
- ◆ 참조 키 : 과목번호, 교수번호

표 개설과목

과목번호	과목명	교수번호	교수명	강의실코드	강의실설명
DBO101	오라클기초	21	서영학	A301	공대301
DBO101	오라클기초	22	임훈	A302	공대301
DBO102	오라클고급	21	서영학	B201	외대201
DBO102	오라클고급	22	임훈	B202	외대202
DBM101	MS-SQL기초	21	서영학	C401	사대401
DBM101	MS-SQL기초	22	임훈	C402	사대402
DBM102	MS-SQL고급	21	서영학	C501	사대501
DBM102	MS-SQL고급	22	임훈	C502	사대502
OSM101	WIN2000기초	21	서영학	D201	자대201
OSM102	WIN2000고급	23	이민종	D301	자대301
OSL101	LINUX	24	이종오	D401	자대401
OSL101	LINUX	25	이창익	D402	자대402

표 성적

과목번호	교수번호	학번	학생명	점수
DBO101	21	9936720	김유신	B
DBO101	21	9936721	엄정화	A
DBO101	21	9936723	장나라	A+
DBO101	22	9936722	김국진	A
DBO102	21	9936723	장나라	A+



Written By Seo Y.H

DBO102	22	9936724	강감찬	B
DBM101	21	9936720	김유신	C+
DBM101	22	9936721	엄정화	A
DBM102	21	9936723	장나라	A+
DBM102	22	9936724	강감찬	B
OSM101	21	9936720	김유신	A
OSM101	21	9936721	엄정화	A+
OSM102	23	9936723	장나라	A+
OSM102	23	9936724	강감찬	C
OSL101	24	9936720	김유신	C
OSL101	25	9936721	엄정화	A

▶ 제 2 정규화(2NF, Second Normal Form)

모든 속성은 식별자에 직접적으로 의존적이어야 하며 이에 해당되지 않는 속성을 분리한다. 즉, 복합키로 구성된 식별자를 가진 경우에 식별자의 일부분에 종속적인 속성을 분리한다.

과목명은 개설 과목의 식별자(과목번호, 교수번호)에서 과목번호에 부분적 종속 속성을 가지고 있으므로 개설과목, 과목으로 분리할 수 있다. 교수명 역시 개설과목의 식별자(과목번호, 교수번호)에서 교수번호에 부분적 종속 속성을 가지고 있으므로 개설과목, 과목으로 분리할 수 있다.

표 과목

<u>과목번호</u>	과목명	배당학점
DBO101	오라클기초	2
DBO102	오라클고급	3
DBM101	MS-SQL기초	2
DBM102	MS-SQL고급	3
OSM101	WIN2000기초	2
OSM102	WIN2000고급	3
OSL101	LINUX	3



Written By Seo Y.H

표 교수

교수번호	교수명	주소
21	서영학	서울 송파구
22	임훈	서울 동대문구
23	이민종	서울 영등포구
24	이종오	서울 은평구
25	이창익	서울 양천구

표 개설과목

과목번호	교수번호	강의실코드	강의실설명
DBO101	21	A301	공대301
DBO101	22	A302	공대301
DBO102	21	B201	외대201
DBO102	22	B202	외대202
DBM101	21	C401	사대401
DBM101	22	C402	사대402
DBM102	21	C501	사대501
DBM102	22	C502	사대502
OSM101	21	D201	자대201
OSM102	23	D301	자대301
OSL101	24	D401	자대401
OSL101	25	D402	자대402

학생명은 성적의 식별자(과목번호, 교수번호, 학번)에서 학번의 부분적 종속 속성을 가지고 있으므로 성적과 학생으로 나눌 수 있다.

표 학생

학번	학생명	주소
9936720	김유신	서울
9936721	엄정화	서울
9936723	장나라	서울
9936722	김국진	경기 성남
9936724	강감찬	경기 수원



Written By Seo Y.H

표 성적

과목번호	교수번호	학번	점수
DBO101	21	9936720	B
DBO101	21	9936721	A
DBO101	21	9936723	A+
DBO101	22	9936722	A
DBO102	21	9936723	A+
DBO102	22	9936724	B
DBM101	21	9936720	C+
DBM101	22	9936721	A
DBM102	21	9936723	A+
DBM102	22	9936724	B
OSM101	21	9936720	A
OSM101	21	9936721	A+
OSM102	23	9936723	A+
OSM102	23	9936724	C
OSL101	24	9936720	C
OSL101	25	9936721	A

▶ 제 3 정규화(3NF, Third Normal Form)

식별자 이외의 속성간에 종속적인 경우에 분리해 낸다.

제 2 정규화로 생성되어진 위의 개설과목 테이블에서 “강의실설명” 컬럼은 개설과목의 식별자(과목번호, 교수번호)에서 전혀 종속적이지 않고 식별자가 아닌 “강의실코드” 컬럼에 종속되어 있으므로 개설과목과 강의실코드로 분리할 수 있다.

표 개설과목

과목번호	교수번호	강의실코드
DBO101	21	A301
DBO101	22	A302
DBO102	21	B201
DBO102	22	B202
DBM101	21	C401



DBM101	22	C402
DBM102	21	C501
DBM102	22	C502
OSM101	21	D201
OSM102	23	D301
OSL101	24	D401
OSL101	25	D402

표 강의실코드

강의실코드	강의실설명
A301	공대301
A302	공대301
B201	외대201
B202	외대202
C401	사대401
C402	사대402
C501	사대501
C502	사대502
D201	자대201
D301	자대301
D401	자대401
D402	자대402

▶ 제 4 정규화(4NF, Fourth Normal Form)

의존적인 다 대 다(Many to Many)관계를 제거하는 것이 제 4 정규화 이다.

아래의 그림과 같이 ‘과목’ 테이블과 ‘교수’ 테이블이 있다고 가정하자.

이 두 개의 테이블은 다 대 다 관계가 형성 되어진다. 즉, 특정 1 과목에 대해 1 명 이상의 교수들이 강의를 할 수 있고, 1 명의 교수가 1 과목 이상의 과목을 강의 할 수 있기 때문이다. 이러한 의존적인 다 대 다(Many to Many)관계를 해소하면 개 설과목이라는 새로운 실체(Entity)가 생성 되어진다. 이때 생성되어지는 새로운 실 체(Entity)의 Primary Key 는 부모 테이블(여기서는 과목, 교수)의 Primary Key 가 전 이되어 복합 Primary Key 로 생성되어진다.



Written By Seo Y.H

표 과목

표 교수

과목번호	과목명	교수번호	교수명
DBO101	오라클기초	21	서영학
DBO102	오라클고급	22	임훈
DBM101	MS-SQL기초	23	이민종
DBM102	MS-SQL고급	24	이종오
OSM101	WIN2000기초	25	이창익
OSM102	WIN2000고급		
OSL101	LINUX		

부모테이블에 있던 식별자는 자식테이블의 외부키(Foreign Key)로 전이되어진다.

표 개설과목

과목번호	교수번호		
DBO101	21		
DBO101	22		
DBO102	21		
DBO102	22		
DBM101	21		
DBM101	22		
DBM102	21		
DBM102	22		
OSM101	21		
OSM102	23		
OSL101	24		
OSL101	25		

<<문제 1>>

어떤 교육원에서 진행중인 각 과정들의 데이터를 과정이라는 테이블에서 관리한다고 해보자. 그리고 각 과정에 소요되는 교재를 과정 테이블에서 컬럼으로 관리한다면 아래의 그림과 같이 테이블이 만들어지고 데이터가 들어갈 것이다.

테이블명 : 과정

컬럼명	과정 코드	과정명	교육 내용	교육 기간	강의 시간	수강료	교재1	교재2	...
키 형태	PK								

위 테이블에서 교재 컬럼의 개수는 교육원의 과정들 중에서 가장 많은 교재를 사용하는 과정을 기준으로 설정되어야 하며 아래의 견본 데이터에서는 5 권이라 가정했다.

테이블명 : 과정

과정 코드	과정명	교육내용	교육기간	강의시간	수강료	교재1	교재2	교재3	교재4	교재5
PK										
A001	웹 프로그래밍	웹 구현에 필요한 기본적인 내용과..	5개월	8시간	1800000	SQL Server	HTML 활용	ASP 기초	ASP 활용	실무 구축
A002	웹마스터	웹 사이트를 구축할 수 있는 기본적인 내용과...	4개월	4시간	700000	Windows 2000	SQL Server	NULL	NULL	NULL
A003	ASP 프로그래밍	ASP를 활용한 웹사이트 구축에...	3개월	8시간	1400000	HTML 활용	ASP 활용	NULL	NULL	NULL
J0001	JAVA 프로그래밍	Java 프로그래밍...	9개월	8시간	2000000	SQL Server	JAVA 기초	JSP	NULL	NULL



Written By Seo Y.H

이렇게 데이터가 들어갈 경우 교재가 5 권인 웹프로그래밍의 컬럼들은 모두 채워지지만 그 외의 과목은 최소 1 개 이상의 NULL 값이 들어가지게 된다. 테이블에 NULL 값이 많다는 것은 쓸데없는 데이터 저장공간의 낭비이다.
위의 테이블을 정규화 하시오

<<문제 2>> 다음의 테이블을 정규화 하시오.

테이블명 : 주문

컬럼명	주문ID	주문일	회원ID	회원명	회원등급	상품ID	상품명	단위	수량	단가
키형태	PK									
건본 데이터	1	20020101	HONG	홍길동	일반	A101	숫팩	1-팩	2	20,000
	1	20020101	HONG	홍길동	일반	B102	연필	타스	3	11,000
	1	20020101	HONG	홍길동	일반	C202	전화기	1대	4	15,000
	2	20020101	SONG	송종국	우수	A101	숫팩	1-팩	2	20,000
	3	20020103	KIM	김남일	우수	C202	전화기	1대	1	15,000
	4	20020105	HONG	홍길동	일반	B102	연필	타스	2	11,000
	4	20020105	HONG	홍길동	일반	D203	오이팩	2-팩	3	10,000