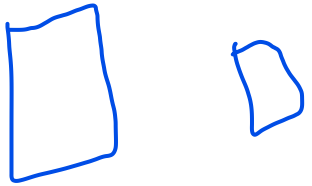
정보시스템 = DB 설계

→ 요구사항분석 → E-R모델링 (ER 다이어그램) →

(스키마정의) (스키마변경)

Table 변환 → 정규화

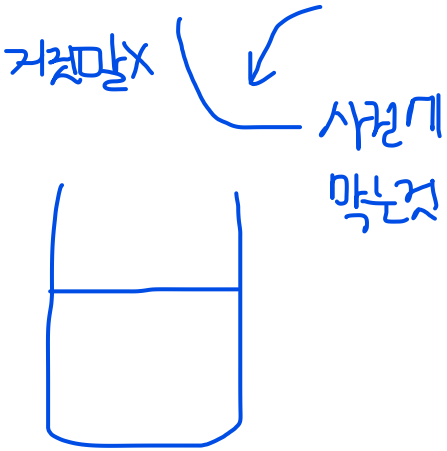


6. Normalization

정규화

목적: DB 상태를 정상으로 유지

비정상 (Anormaly - 이상)



변경 ← Select - 조회
Insert
Delete
Update
} 변화 과정중으로

분해 - Decompsition
1종류씩만 알도록 만든 과정

1NF 2NF 3NF BCNF 4NF 5NF

저장되어질 → 주제 - P.K (개체무결성)

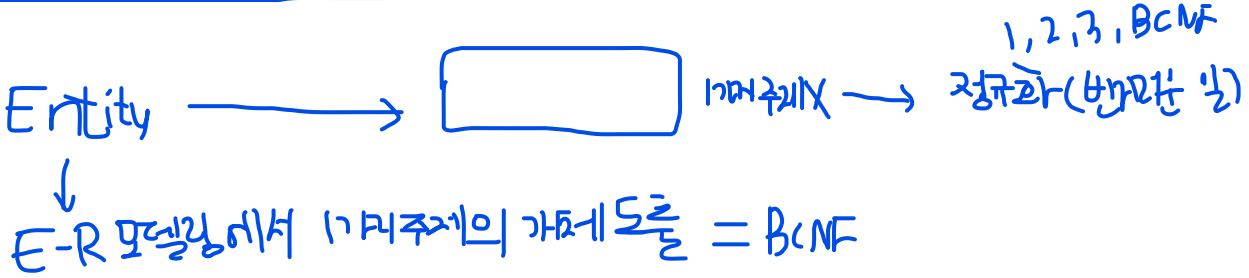
주제가 2개이상 조재 (부분)
↓
주제가 1개만존재 (완전성)

Intro.

- It should become obvious that good, thoughtful design of a conceptual model will result in databases that are either already normalized or can be easily normalized with minor changes

ERD

- This truth illustrates the beauty of the conceptual modeling approach to database design, in that the experienced relational database designer will develop a natural gravitation toward a normalized model from the beginning



When the entire database is defined as a single table

- Result in a large amount of redundant data
- Lengthy searches for just a small number of target rows
- Long and expensive updates, and deletions
  - Can result in the elimination of useful data as an unwanted side effect

Sales

product_name	order_no	cust_name	cust_addr	credit	date	sales_name
vacuum cleaner	1458	Dave Bachmann	Austin	6	1-3-03	Carl Bloch
computer	2730	Qiang Zhu	Plymouth	10	4-15-05	Ted Hanss
refrigerator	2460	Mike Stolarchuck	Ann Arbor	8	9-12-04	Dick Phillips
DVD player	519	Peter Honeyman	Detroit	3	12-5-04	Fred Remley
radio	1986	Charles Antonelli	Chicago	7	5-10-05	R. Metz
CD player	1817	C.V. Ravishankar	Mumbai	8	8-3-02	Paul Basile
vacuum cleaner	1865	Charles Antonelli	Chicago	7	10-1-04	Carl Bloch
vacuum cleaner	1885	Betsy Karmeisool	Detroit	8	4-19-99	Carl Bloch
refrigerator	1943	Dave Bachmann	Austin	6	1-4-04	Dick Phillips
television	2315	Sakti Pramanik	East Lansing	6	3-15-04	Fred Remley

Single table database

- Products, Salespersons, Customers, Orders are all stored in a single table

First Normal Form

- ***Definition:*** A table is in *first normal form (1NF)* if and only if all columns contain only atomic values
- The most basic level of normalized tables
- Standard SQL assumes only atomic values and a relational table is by default in 1NF

multi (composite) → atmic (변환)

AB CD

Super Keys, Candidate Keys, and Primary Keys

- Super Key → 유일성 ABCD, ABD ...
  - A set of one or more attributes, which, when taken collectively, allows us to identify uniquely an entity or table
- Candidate Key → 유일성+최소성 AB ...
  - Any subset of the attributes of a super key that is also a super key, and not reducible to another super key
- Primary Key → 1개 ⟷ 대체키 - 나머지
  - Selected arbitrarily from the set of candidate keys

Functional dependence

- The property of one or more attributes that uniquely determine the value of one or more other attributes
- A -> B
  - Given a table (R), a set of attributes (B) is functionally dependent on another set of attributes (A) if, at each instant of time, each A value is associated with only one B value

Report

report_no	editor	dept_no	dept_name	dept_addr	author_id	author_name	author_addr
4216	woolf	15	design	argus1	53	mantei	cs-tor
4216	woolf	15	design	argus1	44	bolton	mathrev
4216	woolf	15	design	argus1	71	koenig	mathrev
5789	koenig	27	analysis	argus2	26	fry	folkstone
5789	koenig	27	analysis	argus2	38	umar	prise
5789	koenig	27	analysis	argus2	71	koenig	mathrev

  - ***report:***
    - report_no -> editor, dept_no
    - dept_no -> dept_name, dept_addr
    - author_id -> author_name, author_addr

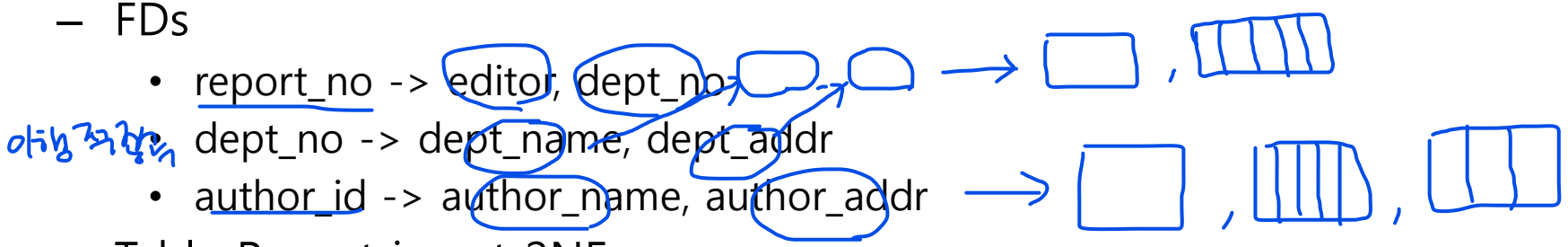
Second Normal Form

- Definition: A table is in second normal form (2NF) if and only if it is in 1NF and every nonkey attribute is fully dependent on the primary key

Report

report_no	editor	dept_no	dept_name	dept_addr	author_id	author_name	author_addr
4216	woolf	15	design	argus1	53	mantei	cs-tor
4216	woolf	15	design	argus1	44	bolton	mathrev
4216	woolf	15	design	argus1	71	koenig	mathrev
5789	koenig	27	analysis	argus2	26	fry	folkstone
5789	koenig	27	analysis	argus2	38	umar	prise
5789	koenig	27	analysis	argus2	71	koenig	mathrev

  - (report_no, author_id), is the only candidate key and is therefore the primary key
  - FDs
    - report_no -> editor, dept_no
    - dept_no -> dept_name, dept_addr
    - author_id -> author_name, author_addr
  - Table Report is not 2NF



Disadvantages of 1NF in table report

Report

report_no	editor	dept_no	dept_name	dept_addr	author_id	author_name	author_addr
4216	woolf	15	design	argus1	53	mantei	cs-tor
4216	woolf	15	design	argus1	44	bolton	mathrev
4216	woolf	15	design	argus1	71	koenig	mathrev
5789	koenig	27	analysis	argus2	26	fry	folkstone
5789	koenig	27	analysis	argus2	38	umar	prise
5789	koenig	27	analysis	argus2	71	koenig	mathrev

Overcoming the disadvantages

Report

report_no	editor	dept_no	dept_name	dept_addr	author_id	author_name	author_addr
4216	woolf	15	design	argus1	53	mantei	cs-tor
4216	woolf	15	design	argus1	44	bolton	mathrev
4216	woolf	15	design	argus1	71	koenig	mathrev
5789	koenig	27	analysis	argus2	26	fry	folkstone
5789	koenig	27	analysis	argus2	38	umar	prise
5789	koenig	27	analysis	argus2	71	koenig	mathrev

- By transforming the 1NF table into two or more 2NF tables
  - Report1{report_no, editor, dept_no, dept_name, and dept_addr}
    - report_no -> editor, dept_no
    - dept_no -> dept_name, dept_addr
  - Report2{author_id, author_name, and author_addr}
    - author_id -> author_name, author_addr
  - Report3{report_no and author_id}
    - report_no, author_id is a candidate key (no FDs)

Overcoming the disadvantages

Report

report_no	editor	dept_no	dept_name	dept_addr	author_id	author_name	author_addr
4216	woolf	15	design	argus1	53	mantei	cs-tor
4216	woolf	15	design	argus1	44	bolton	mathrev
4216	woolf	15	design	argus1	71	koenig	mathrev
5789	koenig	27	analysis	argus2	26	fry	folkstone
5789	koenig	27	analysis	argus2	38	umar	prise
5789	koenig	27	analysis	argus2	71	koenig	mathrev

Report 1

report_no	editor	dept_no	dept_name	dept_addr
4216	woolf	15	design	argus 1
5789	koenig	27	analysis	argus 2

Report 2

author_id	author_name	author_addr
53	mantei	cs-tor
44	bolton	mathrev
71	koenig	mathrev
26	fry	folkstone
38	umar	prise
71	koenig	mathrev

Report 3

report_no	author_id
4216	53
4216	44
4216	71
5789	26
5789	38
5789	71

Disadvantage of Second Normal Form

- The 2NF tables still suffer from the same types of anomalies as the 1NF tables ← Reasons associated with transitive dependencies
- If a transitive (functional) dependency exists in a table, it means that two separate facts are represented in that table

Report 1

report_no	editor	dept_no	dept_name	dept_addr
4216	woolf	15	design	argus 1
5789	koenig	27	analysis	argus 2

  - report_no -> editor, dept_no
  - dept_no -> dept_name, dept_addr
- The side effect of deleting

Third Normal Form

- Definition: A table is in *third normal form (3NF)* if and only if for every nontrivial functional dependency X->A, where X and A are either simple or composite attributes, one of two conditions must hold. Either attribute X is a superkey, or attribute A is a member of a candidate key. If attribute A is a member of a candidate key, A is called a prime attribute.

Report 1

report_no	editor	dept_no	dept_name	dept_addr
4216	woolf	15	design	argus 1
5789	koenig	27	analysis	argus 2

- report_no -> editor, dept_no
- dept_no -> dept_name, dept_addr

Report 11

report_no	editor	dept_no
4216	woolf	15
5789	koenig	27

Report 12

dept_no	dept_name	dept_addr
15	design	argus 1
27	analysis	argus 2

- report11: report_no -> editor, dept_no
- report12: dept_no -> dept_name, dept_addr

Normalized

Report

report_no	editor	dept_no	dept_name	dept_addr	author_id	author_name	author_addr
4216	woolf	15	design	argus1	53	mantei	cs-tor
4216	woolf	15	design	argus1	44	bolton	mathrev
4216	woolf	15	design	argus1	71	koenig	mathrev
5789	koenig	27	analysis	argus2	26	fry	folkstone
5789	koenig	27	analysis	argus2	38	umar	prise
5789	koenig	27	analysis	argus2	71	koenig	mathrev

Report 11

report_no	editor	dept_no
4216	woolf	15
5789	koenig	27

Report 12

dept_no	dept_name	dept_addr
15	design	argus 1
27	analysis	argus 2

또 정규화가 필요

Report 2

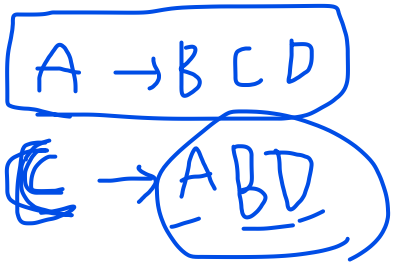
author_id	author_name	author_addr
53	mantei	cs-tor
44	bolton	mathrev
71	koenig	mathrev
26	fry	folkstone
38	umar	prise
71	koenig	mathrev

Report 3

report_no	author_id
4216	53
4216	44
4216	71
5789	26
5789	38
5789	71

Boyce-Codd Normal Form

- Definition: A table **R** is in *Boyce-Codd normal form (BCNF)* if for every nontrivial FD X->A, X is a superkey
- BCNF is considered to be a strong variation of 3NF
  - 3NF, which eliminates most of the anomalies known in databases today, is the most common standard for normalization in commercial databases and CASE tools
  - The few remaining anomalies can be eliminated by the Boyce-Codd normal form (BCNF)
  - Every table that is BCNF is also 3NF, 2NF, and 1NF, by the previous definitions
- BCNF is a stronger form of normalization than 3NF
  - Because it eliminates the second condition for 3NF, which allowed the right side of the FD to be a prime attribute



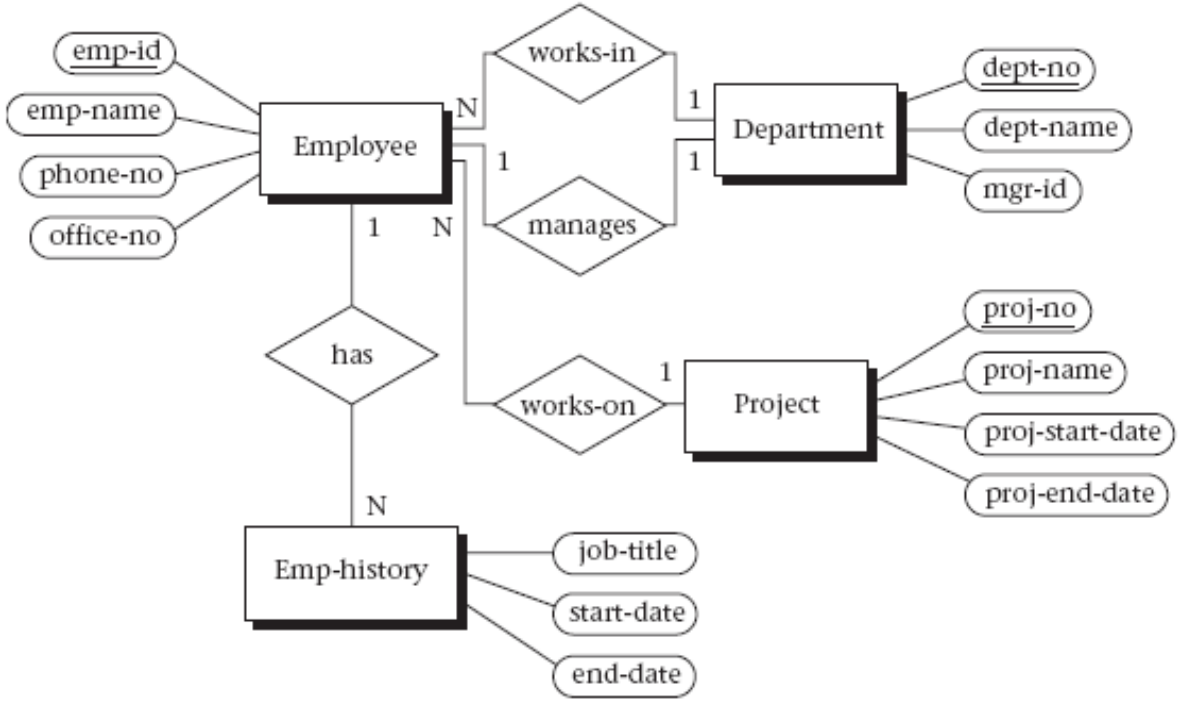
A 3NF table that is not BCNF

- Such tables have delete anomalies similar to those in the lower normal forms
- Assertion 1
  - For a given team, each employee is directed by only one leader. A team may be directed by more than one leader.
  - emp_name, team_name -> leader_name
- Assertion 2
  - Each leader directs only one team.
  - leader_name -> team_name

team:

emp_name	team_name	leader_name
Sutton	Hawks	Wei
Sutton	Condors	Bachmann
Niven	Hawks	Wei
Niven	Eagles	Makowski
Wilson	Eagles	DeSmith

Example



1. emp_id, start_date -> job_title, end_date
2. emp_id -> emp_name, phone_no, office_no, proj_no, proj_name, dept_no
3. phone_no -> office_no
4. proj_no -> proj_name, proj_start_date, proj_end_date
5. dept_no -> dept_name, mgr_id
6. mgr_id -> dept_no

we can define the following tables

- emp_hist: emp_id, start_date -> job_title, end_date
- employee: emp_id -> emp_name, phone_no, proj_no, dept_no
- phone: phone_no -> office_no
- project: proj_no -> proj_name, proj_start_date, proj_end_date
- department: dept_no -> dept_name, mgr_id, mgr_id -> dept_no