NORMALIZATION

Database Management System

Sujan Tamrakar

NORMALIZATION

- ► Result of ER analysis need further refinement.

- ► Appropriate decomposition can solve problems & underlying theory is called Normalization.

- ► It generates a set of relation schemas that allows us to store information without unnecessary redundancy, yet also allows us to retrieve information easily.

- ► The approach is to design schemas that are in an appropriate normal form.

- ► Normalization of data can be considered as a process of analysing the given relation schema based on their functional dependencies and primary keys to achieve the desirable properties of:

    - ► Minimizing redundancy

    - ► Minimizing the insertion, deletion & update anomalies (inconsistencies).

NORMALIZATION

► **Unsatisfactory relation schemas** that do not meet certain conditions (the normal form tests) are **decomposed into smaller relation schemas** that meet the tests & hence possess the desirable properties.

► Normal forms when considered in isolation from other factors do not guarantee a good database design.

► Types:
1. 1NF
2. 2NF
3. 3NF
4. BCNF (3.5NF)
5. 4NF
6. 5NF

rollno	name	branch	hod	office_tel
401	Akon	CSE	Mr. X	53337
402	Bkon	CSE	Mr. X	53337
403	Ckon	CSE	Mr. X	53337
404	Dkon	CSE	Mr. X	53337

- Data redundancy
- Memory issue
- Anomalies (Inconsistencies) in Insert, Update, Delete

NORMALIZATION – 1NF

► First Normal Form is now considered to be part of the formal definition of a relation in the basic relational model.

► It states that the domain of an attribute must include only atomic (simple, indivisible) values and that the value of any attribute in a tuple must be a single value from the domain of that attribute.

► Hence, 1NF disallows having a set of values, a tuple of values or a combination of both as an attribute value for a single value.

► The only attribute values permitted by 1NF are single atomic values.

NORMALIZATION – 1NF

Students

FirstName	LastName	Knowledge
Thomas	Mueller	Java, C++, PHP
Ursula	Meier	PHP, Java
Igor	Mueller	C++, Java

Startsituation

- -

Result after Normalisation

Students

FirstName	LastName	Knowledge
Thomas	Mueller	C++
Thomas	Mueller	PHP
Thomas	Mueller	Java
Ursula	Meier	Java
Ursula	Meier	PHP
Igor	Mueller	Java
Igor	Mueller	C++

(a)

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocations

(b)

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

(c)

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocation
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

NORMALIZATION – 1NF

DEPARTMENT			
Dname	Dnumber	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

There are 3 main techniques to achieve 1NF for such a relation;

1. Remove the attribute Dlocations that violates 1NF and place it in a separate relation Dept_locations along with the primary key Dnumber of Department. The primary key of this relation is the combination {Dnumber, Dlocation}.

2. Expand the key so that there will be a separate tuple in the original Department relation for each location of a Department. This solution has the disadvantage of introducing redundancy in the relation.

3. If a maximum number of values is known for the attribute (ex: maximum 3 locations is known), replace the Dlocations attribute by three atomic attributes: Dlocation1, Dlocation2, Dlocation3. This solution has the disadvantage of introducing NULL values if most departments have fewer than 3 locations.

NORMALIZATION – 1NF

► Of the aforementioned solutions, the first is generally considered the best because it does not suffer from redundancy & it is completely general, having no limit placed on a maximum number of values.

► In fact, if we chose the second solution, it will be decomposed further during subsequent normalization steps into the first solution.

Take the following table.

StudentID is the primary key.

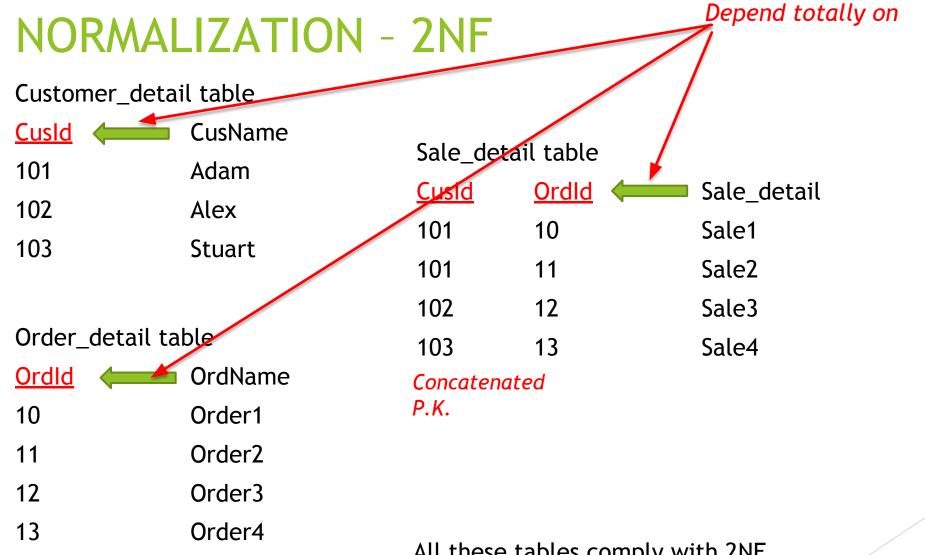| StudentID	StudentName	Address	HouseName	HouseColor	Subject	SubjectCost	Grade
19594332X	Mary Watson	10 Charles Street	Bob	Red	English	$50	B
					Maths	$50	A
					Info Tech	$100	B+

Is it 1NF?

NORMALIZATION – 2NF

► A table that is in 1NF must meet additional criteria if it is to qualify for 2NF.

► A table to be normalized to 2NF should meet all the needs of 1NF and there must not be any partial dependency of any column on primary key.

► It means that for a table that has concatenated primary key, each column in table that is not part of the primary key, must depend upon the entire concatenated key for its existence.

► If any column depends only on one part of the concatenated key, then the tables fails 2NF.

NORMALIZATION – 2NF

Customer table

CusId	CusName	OrdId	OrdName	SaleDetail
101	Adam	10	Order1	Sale1
101	Adam	11	Order2	Sale2
102	Alex	12	Order3	Sale3
103	Stuart	13	Order4	Sale4

- ➤ In customer table concatenation of CusId & OrdId is the primary key.

- ➤ This table is in 1NF but not in 2NF because there are partial dependencies of columns on primary key.

- ➤ CusName is dependent on CusId, OrdName is dependent on OrdId. There is no link between the SaleDetail and CusName.

- ➤ To reduce Customer table to 2NF, break the table into following 3 tables;
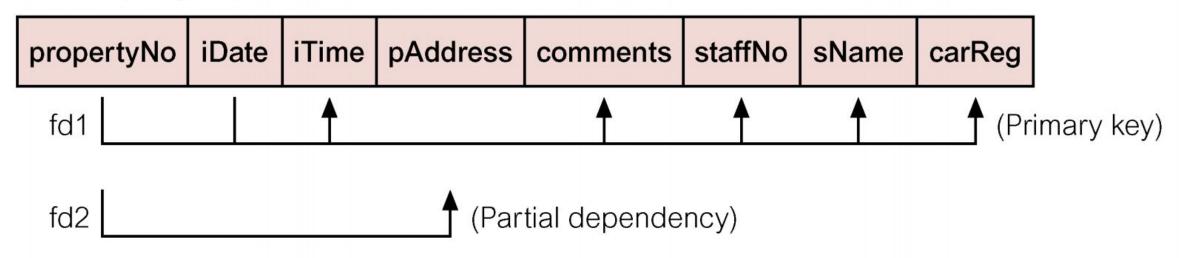
NORMALIZATION – 2NF

Depend totally on

Customer_detail table

CusId	CusName
101	Adam
102	Alex
103	Stuart

Sale_detail table

CusId	OrdId	Sale_detail
101	10	Sale1
101	11	Sale2
102	12	Sale3
103	13	Sale4

Concatenated P.K.

Order_detail table

OrdId	OrdName
10	Order1
11	Order2
12	Order3
13	Order4

All these tables comply with 2NF.

NORMALIZATION – 3NF

► 3NF applies that every non-prime attribute of table must be dependent of primary key.

► The transitive functional dependency should be removed from the table.

► The table must be in 2NF.

× **Transitive Dependency** – when a non-key attribute determines another non-key attribute.

Transitive Dependency

EMPLOYEE

Emp_ID	F_Name	L_Name	Dept_ID	Dept_Name
111	Mary	Jones	1	Acct
122	Sarah	Smith	2	Mktg

NORMALIZATION – 3NF

StaffPropertyInspection

propertyNo	iDate	iTime	pAddress	comments	staffNo	sName	carReg

fd1 (Primary key)

fd2 (Partial dependency)

fd3 (Transitive dependency)

NORMALIZATION – 3NF

Student_detail table						
StdId	StdName	DOB	Street	City	State	Zip

- ► StdId is Primary Key. But Street, City, State depends on Zip.

- ► The dependency between zip & other fields is called Transitive Dependency.

- ► Hence, to apply 3NF, we need to move street, city & state to new table with zip as primary key.

Student_detail table			
StdId	StdName	DOB	Zip

Address_detail table			
Zip	Street	City	State

- ► Pros:
  - ► Amount of data duplication is reduced
  - ► Data integrity achieved

NORMALIZATION – BCNF

- Boyce-Codd Normal Form

- An extension of 3NF, hence sometimes termed as 3.5 NF

- A table is in BCNF if and only if there are no non-trivial functional dependencies of attributes on anything other than a superset of a candidate key.

- A 3NF table which does not have multiple overlapping candidate keys is guaranteed to be in BCNF.

- BCNF is based on the concept of determinant.

- A determinant is any attribute on which some other attribute is fully functional dependent.

- "If B is functionally dependent on A, that is, A functionally determines B then in such case A is called determinant."

- A table is in BCNF if and only if every determinant is a candidate key.

NORMALIZATION – BCNF

► Consider following relation & determinants;

  R($\underline{a,b}$,c,d)

1. a,c ▫ b,d
2. a,d ▫ b

► First determinant suggests that p.k. of R could be changed from a,b to a,c.

► If this change was done all of the non-key attributes present in R could still be determined & therefore this change is legal.

► However, second determinant indicates that a,d determines b, but a,d could not be the key of R as a,d doesn't determine all of non-key attributes of R (doesn't determine c).

► We would say that 1st determinant is a candidate key, but 2nd determinant is not a candidate key & thus this relation is not in BCNF (but in 3NF).

NORMALIZATION – BCNF

Example of BCNF Decomposition

StudentProf

sNumber	sName	pNumber	pName
s1	Dave	p1	MM
s2	Greg	p2	MM

FDs: pNumber → pName

Student

sNumber	sName	pNumber
s1	Dave	p1
s2	Greg	p2

Professor

pNumber	pName
p1	MM
p2	MM

FOREIGN KEY: Student (PNum) references Professor (PNum)

NORMALIZATION – 4NF

► It is concerned with multivalued dependency.

► "A multi-valued dependency is said to occur when there are two attributes in a table which depend on a third attribute but are independent of each other."

► A table is in 4NF if and only if for everyone on its non-trivial multi-valued dependencies X ▫ Y, X is a superkey i.e. X is either a candidate key or a superset thereof.

► A table is in 4NF if and only if following conditions are satisfied;

  ► R is already in 3NF or BCNF

  ► If it contains no multivalued dependencies

► BCNF to 4NF involves removal of the multivalued dependencies from the relation by placing the attribute(s) in a new relation along with a copy of the determinants.

NORMALIZATION – 4NF

Normalization
BCNF to 4NF Relations

Branch_Staff_Client relation

Branch_No	SName	CName
B3	Ann Beech	Aline Stewart
B3	David Ford	Aline Stewart
B3	Ann Beech	Mike Richie
B3	David Ford	Mike Richie

↓

Branch_Staff relation

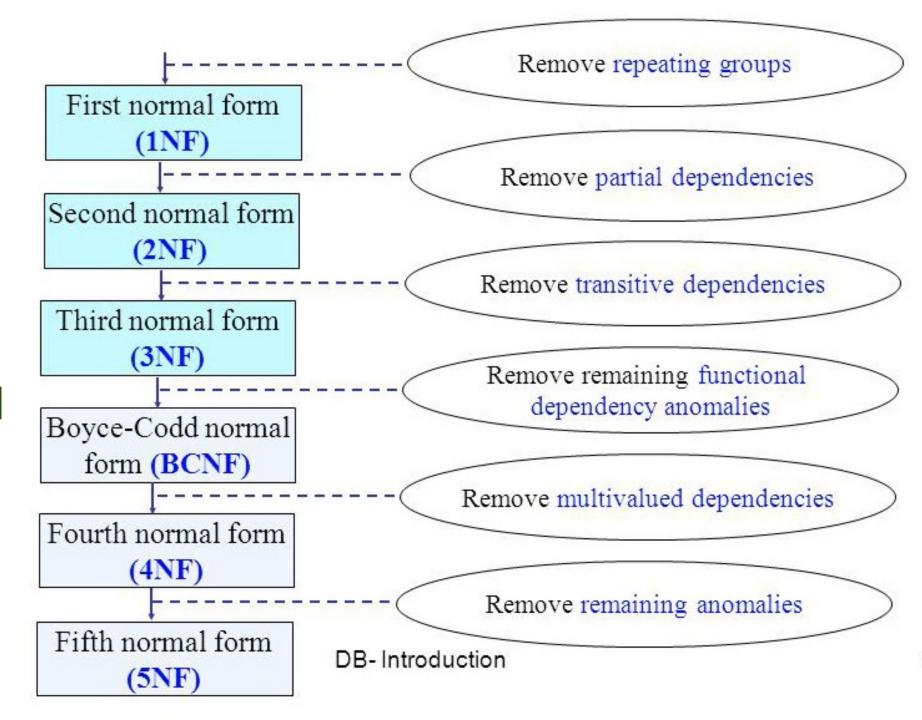| Branch_No	SName
B3	Ann Beech
B3	David Ford

Branch_Client relation

Branch_No	CName
B3	Aline Stewart
B3	Mike Richie

Teacher_Name	Teacher_Subject	Teacher_Language
Narendra	Science	Hindi
Narendra	Maths	Hindi
Narendra	Science	English
Narendra	History	English
Alok Sharma	Science	Hindi
Alok Sharma	Physical Education	English

Teacher_Name	Teacher_Subject
Narendra	Science
Narendra	Maths
Narendra	History
Alok Sharma	Science
Alok Sharma	Physical Education

Teacher_Name	Teacher_Language
Narendra	Hindi
Narendra	English
Alok Sharma	Hindi
Alok Sharma	English

STAGES OF NORMALIZATION

First normal form (1NF)

Remove repeating groups

Second normal form (2NF)

Remove partial dependencies

Third normal form (3NF)

Remove transitive dependencies

Boyce-Codd normal form (BCNF)

Remove remaining functional dependency anomalies

Fourth normal form (4NF)

Remove multivalued dependencies

Fifth normal form (5NF)

Remove remaining anomalies

DB- Introduction

Full Names	Physical Address	Movies rented	Salutation	Category
Janet Jones	First Street Plot No 4	Pirates of the Caribbean, Clash of the Titans	Ms.	Action, Action
Robert Phil	3rd Street 34	Forgetting Sarah Marshal, Daddy's Little Girls	Mr.	Romance, Romance
Robert Phil	5th Avenue	Clash of the Titans	Mr.	Action

Full Names	Physical Address	Movies rented	Salutation
Janet Jones	First Street Plot No 4	Pirates of the Caribbean	Ms.
Janet Jones	First Street Plot No 4	Clash of the Titans	Ms.
Robert Phil	3rd Street 34	Forgetting Sarah Marshal	Mr.
Robert Phil	3rd Street 34	Daddy's Little Girls	Mr.
Robert Phil	5th Avenue	Clash of the Titans	Mr.

Table 1: In 1NF Form

FULL NAMES	PHYSICAL ADDRESS	MOVIES RENTED	SALUTATION
Janet Jones	First Street Plot No 4	Pirates of the Caribbean	Ms.
Janet Jones	First Street Plot No 4	Clash of the Titans	Ms.
Robert Phil	3rd Street 34	Forgetting Sarah Marshal	Mr.
Robert Phil	3rd Street 34	Daddy's Little Girls	Mr.
Robert Phil	5th Avenue	Clash of the Titans	Mr.

Table 1: In 1NF Form

MEMBERSHIP ID	FULL NAMES	PHYSICAL ADDRESS	SALUTATION
1	Janet Jones	First Street Plot No 4	Ms.
2	Robert Phil	3rd Street 34	Mr.
3	Robert Phil	5th Avenue	Mr.

Table 1

MEMBERSHIP ID	MOVIES RENTED
1	Pirates of the Caribbean
1	Clash of the Titans
2	Forgetting Sarah Marshal
2	Daddy's Little Girls
3	Clash of the Titans

In 2NF

Table 2

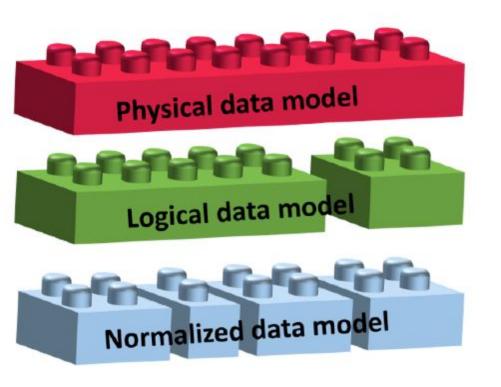| MEMBERSHIP ID	FULL NAMES	PHYSICAL ADDRESS	SALUTATION ID
1	Janet Jones	First Street Plot No 4	2
2	Robert Phil	3rd Street 34	1
3	Robert Phil	5th Avenue	1

TABLE 1

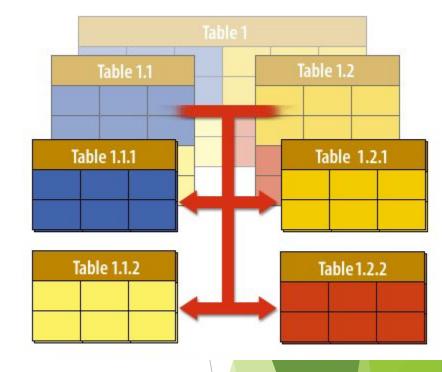| MEMBERSHIP ID	MOVIES RENTED
1	Pirates of the Caribbean
1	Clash of the Titans
2	Forgetting Sarah Marshal
2	Daddy's Little Girls
3	Clash of the Titans

Table 2

SALUTATION ID	SALUTATION
1	Mr.
2	Ms.
3	Mrs.
4	Dr.

Table 3

In 3NF

Thank you.

Assignment : Solve UT1 Both Questions
Due: Wednesday 30th June 2022