Convert Relations Into 3NF

Matthew Pike & Yuan Yao

University of Nottingham Ningbo China (UNNC)

Overview

Learning Outcomes

By the end of this lecture, you should be able to

- Understand what is **2NF**.
- Convert 1NF relations into **2NF**.
- Know the problems of **2NF** relations.
- Understand what is **3NF**.
- Convert 2NF relations into **3NF**.
- Know the problems of **3NF** relations.

Second Normal Form

Second Normal Form

2NF:

- It is in 1NF.
- There is no non-key attributes B is partially dependent on a candidate key.
- No $C \rightarrow B$, where B is a set of non-key attributes, C is a strict subset of candidate key.

Attributes:

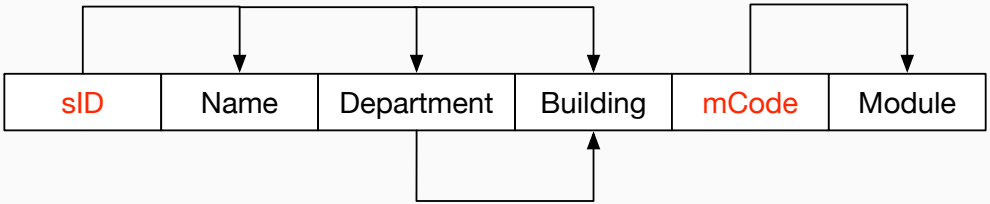
- Key attributes: part of some candidate key.
- Non-key attributes: the rest of the attributes.

Exercise 2: 2NF

sID	Name	Department	Building	mCode	Module
...	...	...	...	...	...

Is this relation in 2NF? If not, find the partial FDs on a candidate key.

Exercise 2: 2NF



- Candidate keys: $\{sID, mCode\}$
- Partial FDs:
  - $\{sID, mCode\} \rightarrow \{Module\}$
  - $\{sID, mCode\} \rightarrow \{Name, Department, Building\}$

Removing FDs

Suppose we have a relation R with schema S, a FD $A \rightarrow B$, where:

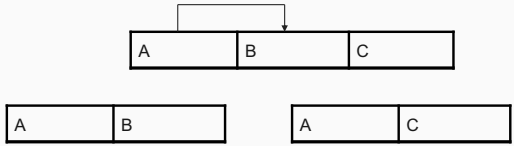
- A is a key attribute.
- $A \cap B = \{\}$

Let $C = S - (A \cup B)$, i.e., all other attributes.

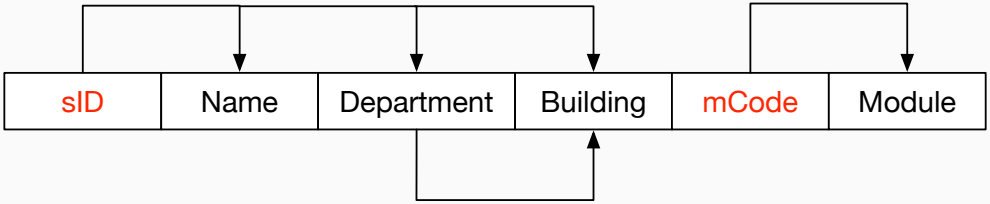
We could divide S into two parts:

- R_1, with schema $A \cup C$
- R_2, with schema $A \cup B$

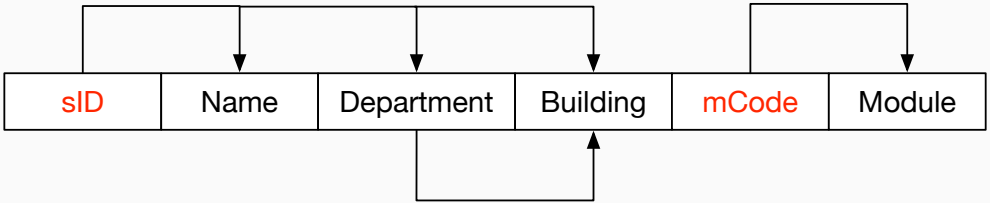
$S = R_1 \bowtie R_2$



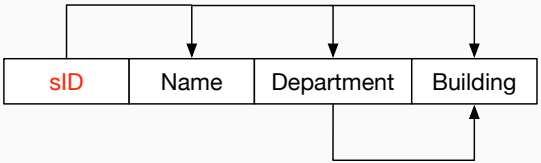
Exercise 3: Removing FDs



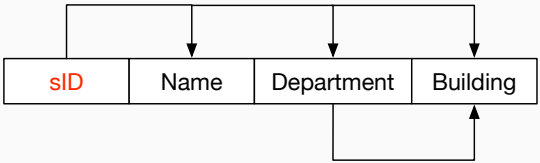
Exercise 3: Removing FDs



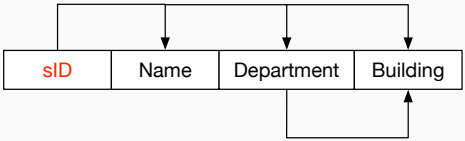
$A = sID$, $B = \{Name, Department, Building\}$, $C = \{mCode, Module\}$

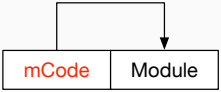


Exercise 3: Removing FDs



$A = mCode, B = \{Module\}, C = \{sID\}$





Example Database in 2NF

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

Recall the problems we have in 1NF

If we add a new staff David Ford in Computer Science with ID 5.

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

Insertion anomalies in 2NF

If we add a new staff David Ford in Computer Science with ID 5.

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1
5	David Ford	Computer Science	?

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

Insertion anomalies in 2NF

If we add the department Chemistry that is in building C1.

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1
Null	Null	Chemistry	C1

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

Deletion anomalies in 2NF

If we want to delete module MCS

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

Deletion anomalies in 2NF

If we want to delete module MCS

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms

Deletion anomalies in 2NF

What if we want to delete David Jones?

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

Deletion anomalies in 2NF

Mary Jones is now transferred to the department of Mathematics

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

Deletion anomalies in 2NF

Mary Jones is now transferred to the department of Mathematics

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Mathematics	B1
4	David Jones	Mathematics	A1

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

3NF

Third Normal Form

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1

- A special type of FDs: transitive FD.
  - $A \to C$ is a transitive FD, if there is some set B such that $A \to B$ and $B \to C$.
  - E.g., $\{sID\} \to \{Department\}$, $\{Department\} \to \{Building\}$
- Third Normal Form (3NF):
  - It is in 2NF.
  - No non-key attribute is transitively dependent on a candidate key.

Removing Transitive FDs

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1

- If $A \rightarrow B$ and $B \rightarrow C$, S is the schema.
- Divide S into two parts:
  - R_1, with schema $B \cup C$.
  - R_2, with schema $S - C$.

Exercise 4: Removing transitive FDs

sID	Name	Department	Building
1	John Smith	Computer Science	B1
2	Mark Brown	Computer Science	B1
3	Mary Jones	Computer Science	B1
4	David Jones	Mathematics	A1

Exercise 4: Removing transitive FDs

sID	Name	Department
1	John Smith	Computer Science
2	Mark Brown	Computer Science
3	Mary Jones	Computer Science
4	David Jones	Mathematics

Department	Building
Computer Science	B1
Mathematics	A1

Example in 3NF

sID	Name	Department
1	John Smith	Computer Science
2	Mark Brown	Computer Science
3	Mary Jones	Computer Science
4	David Jones	Mathematics

Department	Building
Computer Science	B1
Mathematics	A1

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

Insertion in 3NF

We want to add the department of Chemistry that is in building C1

sID	Name	Department
1	John Smith	Computer Science
2	Mark Brown	Computer Science
3	Mary Jones	Computer Science
4	David Jones	Mathematics

Department	Building
Computer Science	B1
Mathematics	A1

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

Insertion in 3NF

We want to add the department of Chemistry that is in building C1

sID	Name	Department
1	John Smith	Computer Science
2	Mark Brown	Computer Science
3	Mary Jones	Computer Science
4	David Jones	Mathematics

Department	Building
Computer Science	B1
Mathematics	A1
Chemistry	C1

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

Deletion in 3NF

We want to delete David Jones.

sID	Name	Department
1	John Smith	Computer Science
2	Mark Brown	Computer Science
3	Mary Jones	Computer Science
4	David Jones	Mathematics

Department	Building
Computer Science	B1
Mathematics	A1

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

Deletion in 3NF

We want to delete David Jones.

sID	Name	Department
1	John Smith	Computer Science
2	Mark Brown	Computer Science
3	Mary Jones	Computer Science

Department	Building
Computer Science	B1
Mathematics	A1

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI

Update in 3NF

Mary Jones is now transferred to the department of Mathematics.

sID	Name	Department
1	John Smith	Computer Science
2	Mark Brown	Computer Science
3	Mary Jones	Computer Science
4	David Jones	Mathematics

Department	Building
Computer Science	B1
Mathematics	A1

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

Update in 3NF

Mary Jones is now transferred to the department of Mathematics.

sID	Name	Department
1	John Smith	Computer Science
2	Mark Brown	Computer Science
3	Mary Jones	Mathematics
4	David Jones	Mathematics

Department	Building
Computer Science	B1
Mathematics	A1

mCode	Module
DBI	Database and Interfaces
FAI	Foundation of Artificial Intelligence
PGA	Programming and Algorithms
MCS	Mathematics for Computer Scientists

sID	mCode
1	DBI
1	FAI
2	FAI
3	PGA
3	DBI
4	MCS

Summary

Normalization and Design

Normalization is related to Database Design

- A database should normally be in 3NF at least.
- If your design leads to a non-3NF database, then you might want to revise it.

When you find you have a non-3NF database

- Identify the FDs that are causing problems.
- Think if they lead to any insert, update or deletion anomalies.
- Try to remove them.

Summary

Normalization

- Definition
- Functional Dependencies
- Normal Forms
- 1NF, 2NF and 3NF