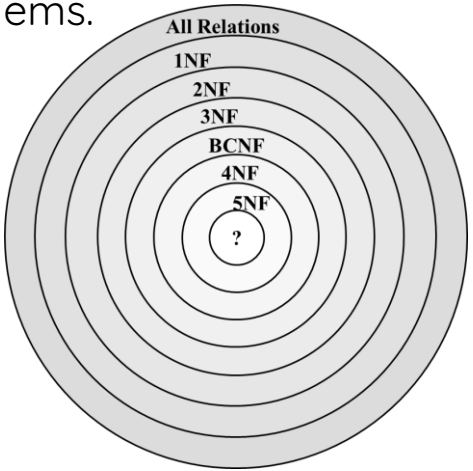
# Normalization

CE384: Database Design
Maryam Ramezani
Sharif University of Technology
maryam.ramezani@sharif.edu

# Introduction

- Normalization: systematic validation of participation of attributes in a relation schema from a data redundancy perspective.
- Normal Forms (NFs): stepwise progression towards attaining the goal of a fully normalized relation schema.
- A design that has a lower normal form than another design has more redundancy. Uncontrolled redundancy can lead to data integrity problems.
- We call candidate key  (C.K) in this slide as “key”.



# First Normal Form (1NF)

- No multi-valued attributes or composite attributes.
- By definition, a relation schema is in 1NF.
- Does the below relations satisfy 1NF?

EMPLOYEE

| Name | Age | Sex | Emp# |
|---|---|---|---|
| Anderson | 21 | F | 010110 |
| Decker | 22 | M | 010100 |
| Glover | 22 | M | 101000 |
| Jackson | 21 | F | 201100 |
| Moore | 19 | M | 111100 |
| Nakata | 20 | F | 111101 |

# First Normal Form (1NF)

- Does the below relations satisfy 1NF?

| EmpNum | EmpPhone | EmpDegrees |
|---|---|---|
| 123 | 233-9876 | |
| 333 | 233-1231 | BA, BSc, PhD |
| 679 | 233-1231 | BSc, MSc |

- Solution to make it 1NF:

Employee

| EmpNum | EmpPhone |
|---|---|
| 123 | 233-9876 |
| 333 | 233-1231 |
| 679 | 233-1231 |

EmployeeDegree

| EmpNum | EmpDegree |
|---|---|
| 333 | BA |
| 333 | BSc |
| 333 | PhD |
| 679 | BSc |
| 679 | MSc |

# Note

- A key attribute is any attribute that is part of a key.
- Any attribute that is not a key attribute, is a non-key attribute.

# Second Normal Form (2NF)

- If it is in 1NF and:
- At least one of the following conditions applies:
  - Primary key consists of a single attribute
  - No non-key attributes
  - Every non-key attribute depends on all of the **primary key** (fully functionally dependent)

**Pay Attention!!!!**

In some references 2NF is defined as:

A relation is considered to be in 2NF if it is in 1NF, and **every non-key attribute is fully dependent on each candidate key**.

# Second Normal Form (2NF)

- Example

NEW_ALBUM

| Album_no | Artist_nm | Price | Stock |
|---|---|---|---|
| BS123 | Britney Spears | 17.95 | 1000 |
| JT111 | Justin Timberlake | 17.95 | 1200 |
| BTL007 | John Lennon | 23.95 | |
| BTL007 | Paul McCartney | 23.95 | |
| BTL007 | George Harrison | 23.95 | |
| BTL007 | Ringo Star | 23.95 | |
| MJ100 | Michael Jackson | 17.95 | |
| JM456 | John Mayer | 16.95 | 1000 |
| JM151 | John Mayer | 16.95 | 1000 |
| MX789 | Madonna | 11.95 | 500 |
| DJM237 | John Denver | 11.95 | 2000 |
| DJM237 | Michael Jackson | 11.95 | 2000 |
| DJM237 | Madonna | 11.95 | 2000 |
| DR711 | Diana Ross | 12.95 | 1000 |
| PM137 | Paul McCartney | 19.95 | |

candidate key: {Album_no, Artist_nm}

Album_no → Price  Album_no → Stock

# Modification Anomalies

- **change the value of price or stock of Album_no BTL007 in NEW_ALBUM**
  - multiple tuples require update and failure to update some can cause an update anomaly
- **add a new tuple (Album_no: XY11, Price: 17.95 and Stock: 100) to NEW_ALBUM**
  - cannot insert without artist name, which is an insertion anomaly
- **delete Album_no BTL007 from NEW_ALBUM**
  - requires deletion of multiple tuples and failure to delete some can cause a deletion anomaly

# Resolution of 2NF Violation

- Pull out the undesirable FD(s) from the target relation schema as a separate relation schema(s)
- Keep the determinant (left side of the FD equation) of the pulled-out relation schema as an attribute(s) in the leftover target relation schema

ALBUM_INFO

| Album_no | Price | Stock |
|---|---|---|
| BS123 | 17.95 | 1000 |
| JT111 | 17.95 | 1200 |
| BTL007 | 23.95 | |
| MJ100 | 17.95 | |
| JM456 | 16.95 | 1000 |
| JM151 | 16.95 | 1000 |
| MX789 | 11.95 | 500 |
| DJM237 | 11.95 | 2000 |
| DR711 | 12.95 | 1000 |
| PM137 | 19.95 | |

ALBUM_ARTIST

| Album_no | Artist_nm |
|---|---|
| BS123 | Britney Spears |
| JT111 | Justin Timberlake |
| BTL007 | John Lennon |
| BTL007 | Paul McCartney |
| BTL007 | George Harrison |
| BTL007 | Ringo Star |
| MJ100 | Michael Jackson |
| JM456 | John Mayer |
| JM151 | John Mayer |
| MX789 | Madonna |
| DJM237 | John Denver |
| DJM237 | Michael Jackson |
| DJM237 | Madonna |
| DR711 | Diana Ross |
| PM137 | Paul McCartney |

# Second Normal Form (2NF)

- Does the below relation satisfy 2NF?

EMPLOYEE

| Name | Age | Sex | Emp# |
|---|---|---|---|
| Anderson | 21 | F | 010110 |
| Decker | 22 | M | 010100 |
| Glover | 22 | M | 101000 |
| Jackson | 21 | F | 201100 |
| Moore | 19 | M | 111100 |
| Nakata | 20 | F | 111101 |

yes, because the primary key is one attribute

# Example

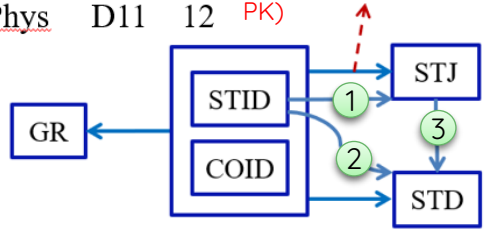
- In general, all attributes related to students, courses, and selections can be represented in a single relationship.
- Environmental Rules:
  - Each student studies one specific major (field of study) .
  - Each student is enrolled in one department.
  - Each field is offered in one department.
- What are anomalies?

**R (STID, COID, STJ, STD, GR)**

| | | | | |
|---|---|---|---|---|
| 777 | CO1 | Phys | D11 | 19 |
| 777 | CO2 | Phys | D11 | 16 |
| 777 | CO3 | Phys | D11 | 11 |
| 888 | CO1 | Math | D12 | 16 |
| 888 | CO2 | Math | D12 | 18 |
| 444 | CO1 | Math | D12 | 13 |
| 555 | CO1 | Phys | D11 | 14 |
| 555 | CO2 | Phys | D11 | 12 |

We have a non-complete FD => is not 2NF

FDs resulting from the primary key (left side of PK)




$$\begin{cases} (STID, COID) \rightarrow STJ \\ STID \rightarrow STJ \end{cases} \qquad \begin{cases} (STID, COID) \rightarrow STD \\ STID \rightarrow STD \end{cases}$$

# Example

$\Pi_{\langle STID,COID,GR\rangle}(R)$

**SCG (STID, COID, GR)**

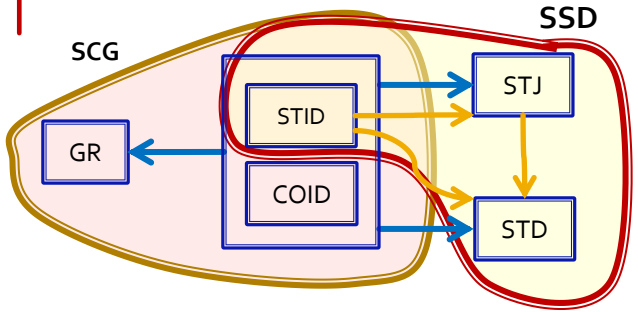
| | | |
|---|---|---|
| 777 | CO1 | 19 |
| 777 | CO2 | 16 |
| 777 | CO3 | 11 |
| 888 | CO1 | 16 |
| 888 | CO2 | 18 |
| 444 | CO1 | 13 |
| 555 | CO1 | 14 |
| 555 | CO2 | 12 |

و

$\Pi_{\langle STID,STJ,STD\rangle}(R)$

**SSD (STID, STJ, STD)**

| | | |
|---|---|---|
| 777 | Phys | D11 |
| 888 | Math | D12 |
| 444 | Math | D12 |
| 555 | Phys | D11 |



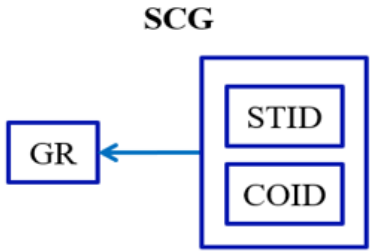

We must decompose the relation in such a way that no partial functional dependencies remain in any resulting relation, and the original set of functional dependencies can be derived from the union of the dependencies of the decomposed relations.

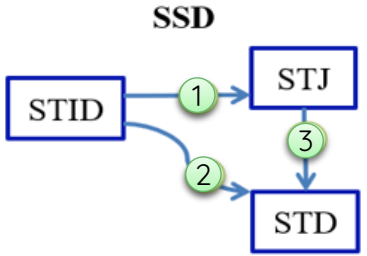
# Example

- The new relations do not exhibit anomalies in R:
  - 1. Insert: ('666', 'Chem', 'D16')
    - This is inserted into SSD without any issue.

| SSD (STID, | STJ, | STD) |
|---|---|---|
| 777 | Phys | D11 |
| 888 | Math | D12 |
| 444 | Math | D12 |
| 555 | Phys | D11 |
| 666 | Chem | D16 |

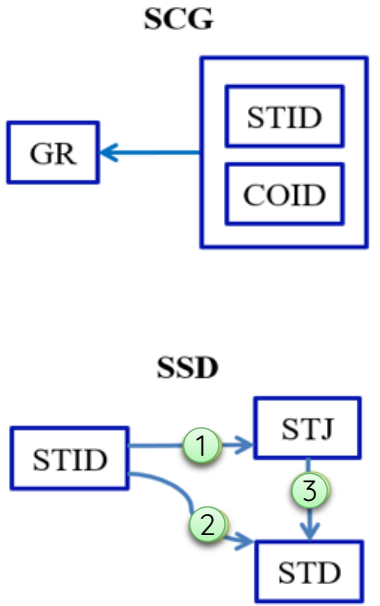
  - 2. Delete: ('444', 'CO1', 13)
    - This is removed from SCG without any issue.
  - 3. Update: Change the major of student 777 to Cher
    - This update is reflected in SSD without any issue.



# Example

- The new anomalies:
  - 1. Insert: Add “software” field in Dep=200.
    - This needs STID for inserting in SSD.
  - 2. Delete: ('444', ‘IT’, 200) from SSD
    - This will remove IT in 200, too.
  - 3. Update: Change the Dep of Math:
    - This update needs change multiple rows in SSD.



# Third Normal Form (3NF)

- A relation is in third normal form if:
  - It is in second normal form
  - It has no transitive dependencies
- **Solution:**
  - Decompose and set up a new relation that includes the nonkey attribute(s) that functionally determine(s) the other nonkey attributes.
  - The common attribute be a CK in at least one of them.
  - All FDs of main relation be in union FD of decomposed relations or can be inferred.

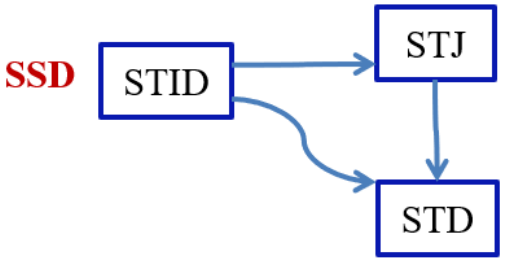
# Example

- Which one is a good decomposition?

I SS (STID, STJ) SD (STJ, STD) ★

II SS (STID, STJ) SD (STID, STD)

III SS(STID, STD) SJ (STJ, STD)



# Note

- Your designed relations must be in at least Third Normal Form (3NF).

# Boyce-Codd Normal Form (BCNF)

- A relation is in BCNF if every determinant of Non-Trivial irreducible FD is a candidate key.
- Note: If a relation is in BCNF, then it is also in 1NF, 2NF, and 3NF.

- **Solution:**
  - Decompose and set up a new relation that includes the non-candidate key attribute(s) that functionally determine(s) the other nonkey attributes.

# Notes

- Boyce-Codd Normal Form (BCNF) is a stronger version of the Third Normal Form (3NF).
  - Why??

# Notes

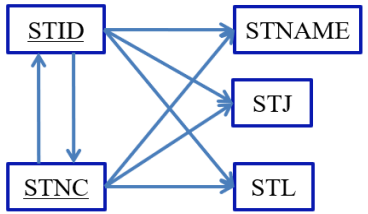
- Case I: The relation R has only one CK, which is the same as the primary key (PK). → If R is in 3NF, then it is also in BCNF.
- Case II: The relation R has more than one CK.
  - Case II-I: The CKs are disjoint (i.e., they have no attributes in common).

    → If R is in 3NF, then it is also in BCNF

**Why this is 3NF?**

ST (STID, STNAME, STNC, STJ, STL, …)

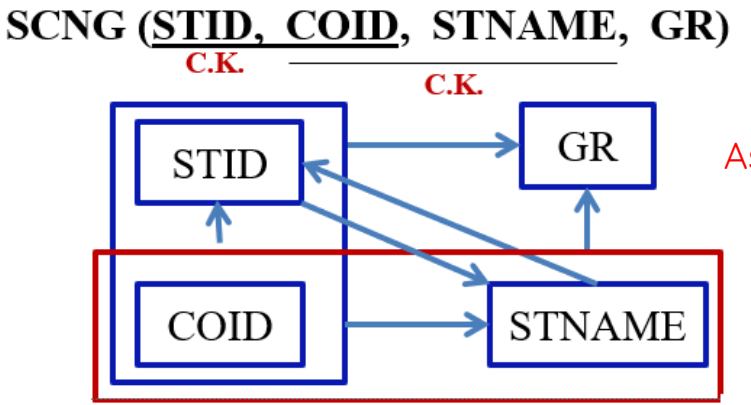
C.K. P.K. C.K.




There are two determinants, and both are CKs.

# Notes

- Case II-II) The candidate keys are overlapping. → If R is in 3NF, **it is not necessarily in BCNF**.




Assumption: No two students have the same name

# Example

| Patient | Hospital | Doctor |
|---|---|---|
| Smith | Methodist | D. Cooley |
| Lee | St. Luke's | Z. Zhang |
| Marks | Methodist | D. Cooley |
| Marks | St. Luke's | W. Lowe |
| Lou | Hermann | R. Duke |

Candidate keys:
{Patient , Hospital},
{Patient, Doctor}

{Patient, Hospital} → Doctor

✖ Doctor → Hospital

# Example

PAT-DOC (Patient, Doctor)

| Patient | Doctor |
|---|---|
| Smith | D. Cooley |
| Lee | Z. Zhang |
| Marks | D. Cooley |
| Marks | W. Lowe |
| Lou | R. Duke |

DOC-HOS (Doctor, Hospital)

| Doctor | Hospital |
|---|---|
| D. Cooley | Methodist |
| Z. Zhang | St. Luke's |
| W. Lowe | St. Luke's |
| R. Duke | Hermann |

# Conclusion

| Normal Form | Requirements | Decomposition Rules |
|---|---|---|
| First | No multi-valued attributes | Form new relations for each multivalued attribute or repeating group |
| Second | Satisfy at least one of the following three conditions:<br>Primary key consists of a single attribute<br>No non-key attributes<br>No non-key attribute should be functionally dependent on part of the primary key (every non-key attribute should be fully functionally dependent on the primary key) | Decompose and setup a new relation for each partial key with its dependent attribute(s). Make sure to keep a relation with the original primary key and any attributes that are fully functionally dependent on it |
| Third | No transitive dependencies. Relation should be in second normal form and should not have a non-key attribute functionally determined by another non-key attribute (or a set of non-key attributes) | Decompose and set up a new relation that includes the nonkey attribute(s) that functionally determine(s) the other nonkey attributes |
| BCNF | Every determinant is a candidate key | Decompose and set up a new relation that includes the non-candidate key attribute(s) that functionally determine(s) the other nonkey attributes. |