Quick Look at Normalization

Dr. Bassam Hammo

Normalization

- 1NF
- 2NF
- 3NF
- BCNF
- Decomposition

Anomalies (Review)

- Anomalies are inconsistencies in data that occur due to unnecessary redundancy.
- Update anomaly
  - Some copies of a data item are updated, but others are not.
- Insertion anomaly
  - Can't insert "real" data without also inserting unrelated or "made up" data
- Deletion anomaly
  - Can't delete some data without also deleting other, unrelated data

Data Normalization

- Primarily a tool to validate and improve a logical design so that it satisfies certain constraints that avoid unnecessary duplication of data.
- The process of decomposing relations with anomalies to produce smaller, well-structured relations.

Anomaly Examples

OID	ODate	CID	CName
1	8/10/2004	2	ABC Inc
2	8/10/2004	2	ABC Inc
3	8/11/2004	3	XZY Co
4	8/11/2004	4	QWE Inc
5	8/12/2004	4	QWE Inc

Add a customer: Must also add an order

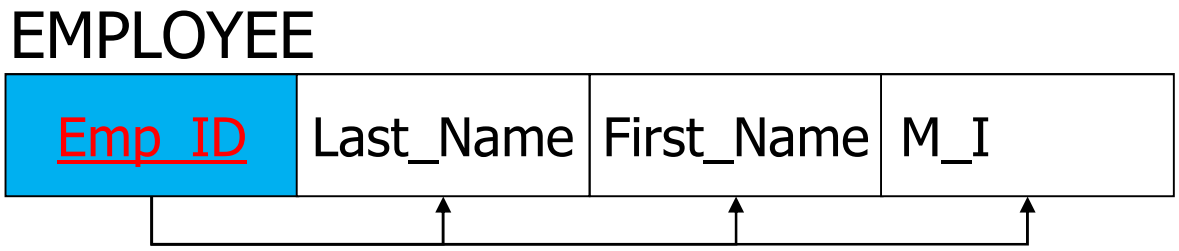
Update the name of Customer #2: Must update multiple rows

Delete Order #3: Must also delete info about Customer #3

Functional Dependencies and Keys

- Functional Dependency: The value of one attribute (the *determinant*) determines the value of another attribute.
- Candidate Key: A possible key.
  - Each non-key field is functionally dependent on every candidate key.
  - No attribute in the key can be deleted without destroying the property of unique identification.

Functional Dependencies

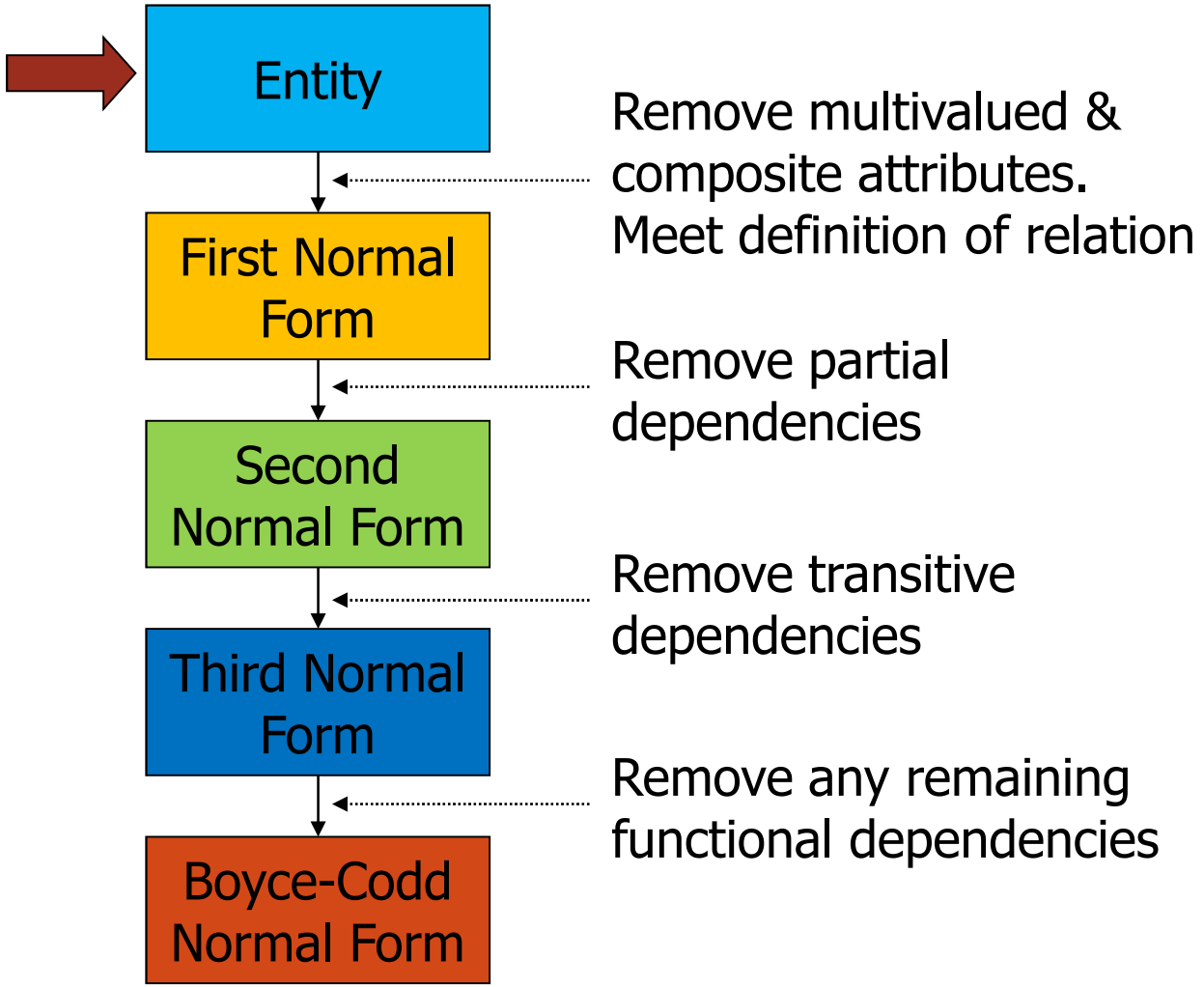


Alternate way to show dependencies

Emp_ID ➔ Last_Name, First_Name, M_I

Steps in Normalization

(doesn't meet the definition of a relation)



First Normal Form (1NF)

If a table of data meets the definition of a relation, it is in first normal form.

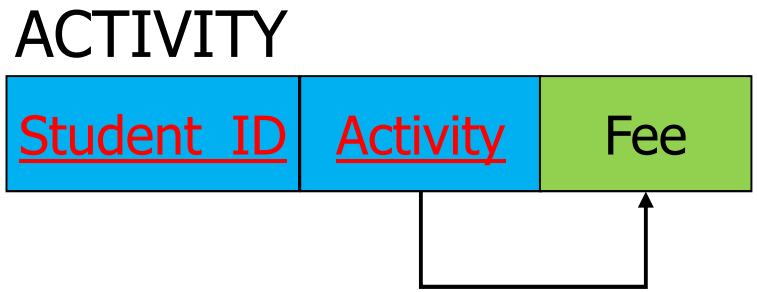
- Every relation has a unique name.
- Every attribute value is atomic (single-valued).
- Every row is unique.
- Attributes in tables have unique names.
- The order of the columns is irrelevant.
- The order of the rows is irrelevant.

Second Normal Form (2NF)

- 1NF and no partial functional dependencies.
- Partial functional dependency: when one or more non-key attributes are functionally dependent on part of the primary key.
- Every non-key attribute must be defined by the entire key, not just by part of the key.
- If a relation has a single attribute as its key, then it is automatically in 2NF.

Second Normal Form (2NF)

A relation that is not in 2NF



Key: Student_ID, Activity

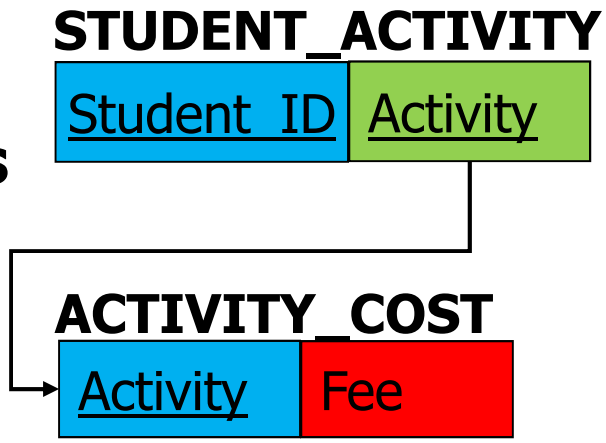
Activity → Fee

Fee is determined by Activity

Student_ID	Activity	Fee
222-22-2020	Swimming	30
232-22-2111	Golf	100
222-22-2020	Golf	100
255-24-2332	Hiking	50

Second Normal Form (2NF)

Divide the relation into two relations that now meet 2NF



Key: Student_ID and Activity

Key: Activity

Activity → Fee

Student_ID	Activity
222-22-2020	Swimming
232-22-2111	Golf
222-22-2020	Golf
255-24-2332	Hiking

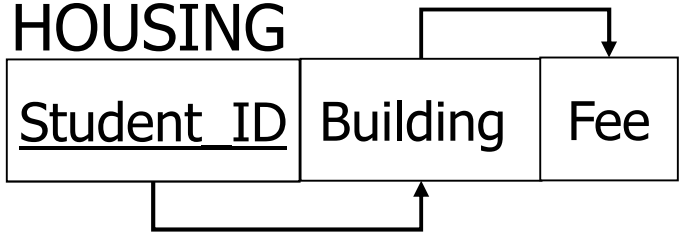
Activity	Fee
Swimming	30
Golf	100
Hiking	50

Third Normal Form (3NF)

- 2NF and no transitive dependencies
- Transitive dependency: a functional dependency between two or more non-key attributes.

Third Normal Form (3NF)

A relation with a transitive dependency



Key: Student_ID

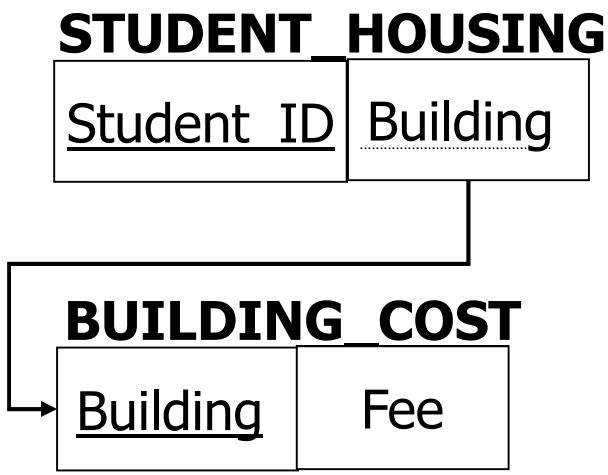
Building → Fee

Student_ID → Building→ Fee

Student_ID	Building	Fee
222-22-2020	Dabney	1200
232-22-2111	Liles	1000
222-22-5554	The Range	1100
255-24-2332	Dabney	1200
330-25-7789	The Range	1100

Third Normal Form (3NF)

Divide the relation into two relations that now meet 3NF



Key: Student_ID

Student_ID → Building

Key: Building

Building → Fee

Student_ID	Building
222-22-2020	Dabney
232-22-2111	Liles
222-22-5554	The Range
255-24-2332	Dabney
330-25-7789	The Range

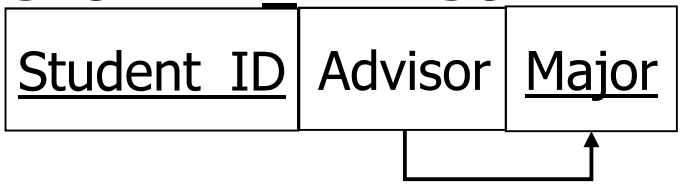
Building	Fee
Dabney	1200
Liles	1000
The Range	1100

Boyce-Codd Normal Form (BCNF)

- 3NF and every determinant is a candidate key.

Boyce-Codd Normal Form (BCNF)

STUDENT_ADVISOR



Primary Key: Student_ID, Major

Candidate Key: Student_ID, Advisor

Advisor $\rightarrow$ Major

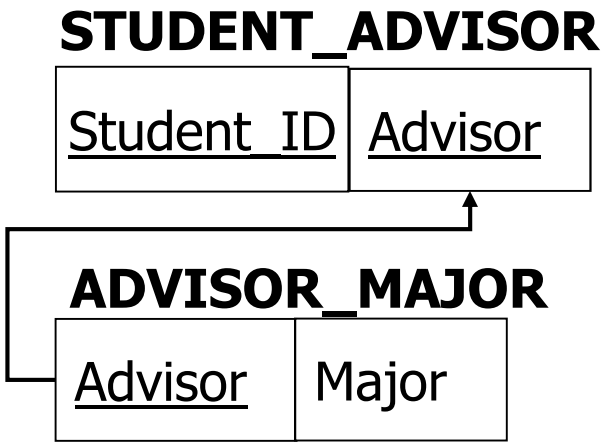
A relation where a determinant is not a candidate key

Student_ID	Major	Advisor
222-22-2020	MIS	Leigh
232-22-2111	Management	Gowan
222-22-2020	MIS	Roberts
222-22-2111	Marketing	Reynolds
255-24-2332	Marketing	Reynolds

Note: Students can have a double major and have an advisor for each major. An advisor works only with students in their assigned area.

Boyce-Codd Normal Form (BCNF)

Divide the relation into two relations that meet BCNF



Key: Student_ID, Advisor

Key: Advisor

Advisor $\rightarrow$ Major

Student_ID	Advisor
222-22-2020	Leigh
232-22-2111	Gowan
222-22-2020	Roberts
222-22-2111	Reynolds
255-24-2332	Reynolds

Advisor	Major
Leigh	MIS
Gowan	Management
Roberts	MIS
Reynolds	Marketing