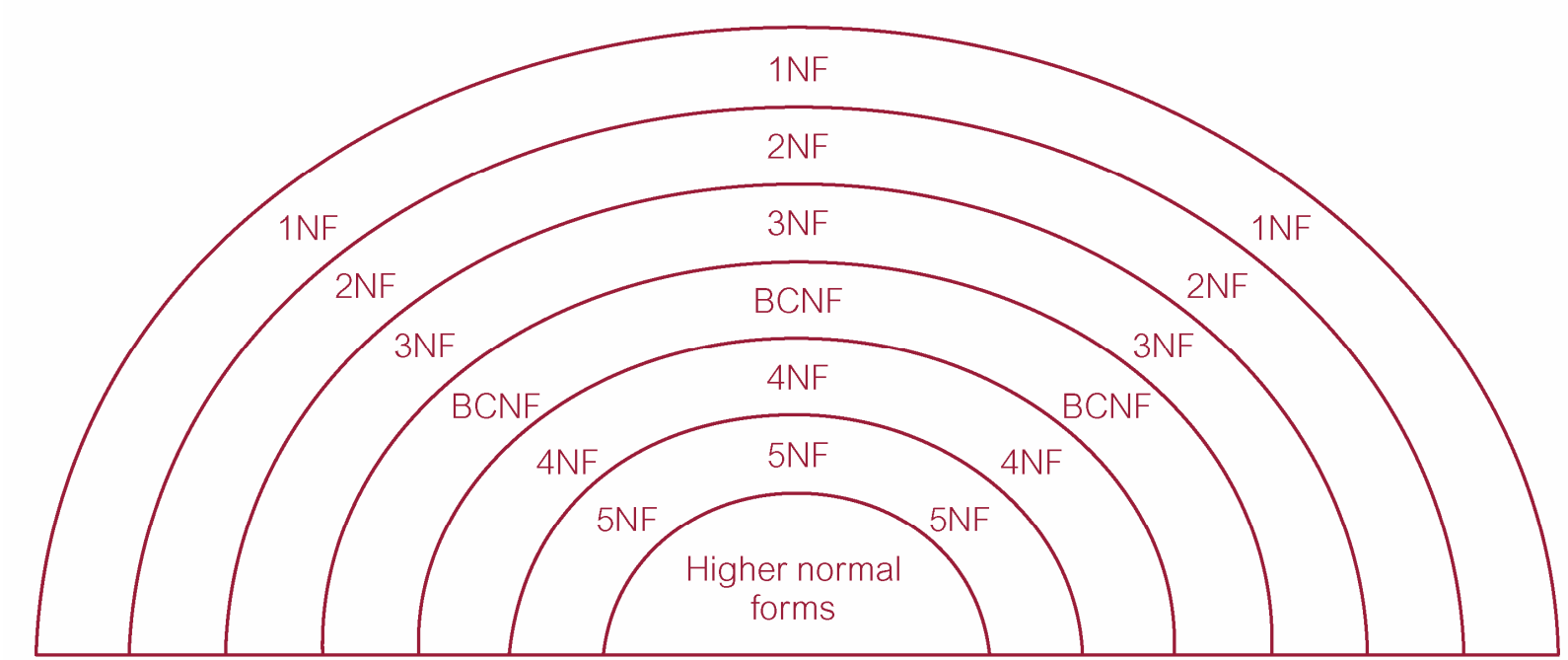
Logical Database Design
Normalisation

Part – 2

The Process of Normalization

- Formal technique for analyzing a relation based on its primary key and functional dependencies between its attributes.
- Often executed as a series of steps. Each step corresponds to a specific normal form, which has known properties.
- As normalization proceeds, relations become progressively more restricted (stronger) in format and also less vulnerable to update anomalies.

Relationship between the normal forms



Purchase Order - Attribute Analysis

ATTRIBUTE	TYPE	LEN-GTH	DESCRIPTION
PO-NO	N	3	Unique purchase order (PO) number. Many parts can be ordered in one PO
PO-DATE	D	8	DDMMYYYY date when PO written
EMP-CODE	C	2	Unique code of employee who wrote the PO
SUPP-NO	N	3	Unique number assigned to supplier
SUPP-NAME	C	20	Supplier name
PART-NO	N	2	Unique number assigned to each part
PART-DESC	C	10	Part description
PART-QTY	N	2	Quantity of parts ordered in given PO

Key PO-NO

Purchase Order Relation in 0NF

PO-NO	PO-DATE	EMP-CODE	SUP P-NO	SUPP-NAME	PART -NO	PART-DESC	PART -QTY
111	01012001	M2	222	AC Stores	P1	Nut	10
					P2	Bolt	5
					P3	Nail	3
					P5	Screw	6
112	01012001	S3	105	I Hardware	P2	Bolt	2
					P5	Screw	1
113	02012001	S1	111	BC Trading	P1	Nut	3
					P3	Nail	4
114	02012001	M2	150	DO Service	P6	Plug	5
115	03012001	S1	222	AC Stores	P7	Pin	8
116	04012001	S1	100	LM Centre	P8	Fuse	2

First Normal Form - 1NF

A relation is in First Normal Form (1NF) if **ALL** its attributes are **ATOMIC**. i.e.

If there are no repeating groups.

If each attribute is a primitive.

e.g. integer, real number, character string, but not lists or sets

non-decomposable data item

single-value

Purchase Order Relation in 0NF

PO(PO-NO, PO-DATE, EMP-CODE, SUPP-NO, SUPP-NAME, PARTS-ORDERED{PART-NO, PART-DESC, PART-QTY})

Within a single purchase order we could find several part numbers, part descriptions and part quantities. Hence, parts ordered can be decomposed.

Purchase Order Relation in 0NF

PO-No	PO-Date	Emp-Code	Supp-No	Supp-Name	Part-No	Part-Desc	Part-Qty
111	01012001	M2	222	AC Stores	P1	Nut	10
					P2	Bolt	5
					P3	Nail	3
					P5	Screw	6
112	01012001	S3	105	I Hardware	P2	Bolt	2
					P5	Screw	1
113	02012001	S1	111	BC Trading	P1	Nut	3
					P3	Nail	4
114	02012001	M2	150	DO Service	P6	Plug	5
115	03012001	S1	222	AC Stores	P7	Pin	8
116	04012001	S1	100	LM Centre	P8	Fuse	2

1NF - Actions Required

1) Examine for repeat groups of data
2) Remove repeat groups from relation
3) Create new relation(s) to include repeated data
4) Include key of the 0NF to the new relation(s)
5) Determine key of the new relation(s)

Purchase Order Relations in 1NF

PO

PO-NO	PO-DATE	EMP-CODE	SUP P-NO	SUPP-NAME
111	01012001	M2	222	AC Stores
112	01012001	S3	105	I Hardware
113	02012001	S1	111	BC Trading
114	02012001	M2	150	DO Service
115	03012001	S1	222	AC Stores
116	04012001	S1	100	LM Centre

PO-PART

PO-NO	PART-NO	PART-DESC	PART-QTY
111	P1	Nut	10
111	P2	Bolt	5
111	P3	Nail	3
111	P5	Screw	6
112	P2	Bolt	2
112	P5	Screw	1
113	P1	Nut	3
113	P3	Nail	4
114	P6	Plug	5
115	P7	Pin	8
116	P8	Fuse	2

Problems - 1NF

1.INSERT PROBLEM

do not know available parts until an order is placed (e.g. P4 is bush)

2. DELETE PROBLEM

loose information of part P7 if we cancel purchase order 115 (i.e. Delete PO-PART for Part No P7)

3. UPDATE PROBLEM:

to change description of Part P3 we need to change every tuple in PO-PART containing Part No P3

Second Normal Form - 2NF

- 2NF is based on the concept of *full functional dependency*.

- A functional dependency X→ Y is a **full functional dependency** if removal of any attribute A from X means that the dependency does not hold any more.

- A functional dependency X→ Y is a **partial dependency** if some attribute A Є X can be removed from X and the dependency still holds.

Second Normal Form - 2NF

- {Empid, Pnumber} → Hours

  is a full dependency (neither Empid → Hours nor Pnumber → Hours holds).

However, the dependency

{Empid, Pnumber} → Ename is partial as Empid → Ename holds.

Second Normal Form - 2NF

A relation is in 2NF if it is in 1NF and every non-key attribute is dependent on the whole key

i.e. Is not dependent on part of the key only.

PO-PART Relation (Parts Ordered) in 1NF

PO-PART(PO-NO, PART-NO, PART-DESC, PART-QTY)

Part Description is depended only on Part No, which is part of the key of PO-PART.

Parts Ordered Relation in 1NF

PO-No	Part-No	Part-Desc	Part-Qty
111	P1	Nut	10
111	P2	Bolt	5
111	P3	Nail	3
111	P5	Screw	6
112	P2	Bolt	2
112	P5	Screw	1
113	P1	Nut	3
113	P3	Nail	4
114	P6	Plug	5
115	P7	Pin	8
116	P8	Fuse	2

Second Normal Form - 2NF

Deals with the relationship between non-key and key fields

It is relevant when the key is composite, i.e. consists of several fields

2NF - Actions Required

If entity has a concatenated key

1) Check each attribute against the whole key
2) Remove attribute and partial key to new relation
3) Optimise relations

Parts Ordered Relations in 2NF

PO-PART

PO-NO	PART-NO	PART-QTY
111	P1	10
111	P2	5
111	P3	3
111	P5	6
112	P2	2
112	P5	1
113	P1	3
113	P3	4
114	P6	5
115	P7	8
116	P8	2

PART

PART-NO	PART-DESC
P1	Nut
P2	Bolt
P3	Nail
P5	Screw
P6	Plug
P7	Pin
P8	Fuse

Purchase Order Relations in 2NF

PART

PART-NO	PART-DESC
P1	Nut
P2	Bolt
P3	Nail
P5	Screw
P6	Plug
P7	Pin
P8	Fuse

PO-PART

PO-NO	PART-NO	PART-QTY
111	P1	10
111	P2	5
111	P3	3
111	P5	6
112	P2	2
112	P5	1
113	P1	3
113	P3	4
114	P6	5
115	P7	8
116	P8	2

PO

PO-NO	PO-DATE	EMP-CODE	SUPP-NO	SUPP-NAME
111	01012001	M2	222	AC Stores
112	01012001	S3	105	I Hardware
113	02012001	S1	111	BC Trading
114	02012001	M2	150	DO Service
115	03012001	S1	222	AC Stores
116	04012001	S1	100	LM Centre

Second Normal Form - 2NF

- Emp_Proj

Empid	Pnumber	Hours	Ename	Pname	Plocation

Problems - 2NF

1. INSERT PROBLEM

   cannot know available suppliers until an order is placed (e.g. 200 is hardware stores)

2. DELETE PROBLEM

   loose information of supplier 100 if we cancel purchase order 116 (i.e. Delete PO for Supplier No 100)

3. UPDATE PROBLEM

   to change name of Supplier 222 we need to change every tuple in PO containing Supplier No 222

Third Normal Form - 3NF

A relation is in 3NF if it is in 2NF and each non-key attribute is only dependent on the whole key, and not dependent on any non-key attribute.

i.e. no transitive dependencies

Third Normal Form - 3NF

PO(PO-NO, PO-DATE, EMP-CODE, SUPP-NO, SUPP-NAME)

Supplier name is a non-key field depended on another non-key field (i.e. the supplier no) in addition to be depended on the key purchase order no.

Third Normal Form - 3NF

Deals with the relationship between non-key fields

A non-key field cannot be a fact about another non-key field

3NF - Actions Required

1)Check each non-key attribute for dependency against other non-key fields

2)Remove attribute depended on another non-key attribute from relation

3)Create a new relation comprising the attribute and non-key attribute which it depends on

4)Determine key of the new relation

PO and SUPPLIER Relations in 3NF

PO

PO-NO	PO-DATE	EMP-CODE	SUPP-NO
111	01012001	M2	222
112	01012001	S3	105
113	02012001	S1	111
114	02012001	M2	150
115	03012001	S1	222
116	04012001	S1	100

SUPPLIER

SUPP-NO	SUPP-NAME
100	LM Centre
105	I Hardware
111	BC Trading
150	DO Service
222	AC Stores

Purchase Order Relations in 3NF

SUPPLIER

SUP P-NO	SUPP-NAME
222	AC Stores
105	I Hardware
111	BC Trading
150	DO Service
222	AC Stores
100	LM Centre

PART

PART-NO	PART-DESC
P1	Nut
P2	Bolt
P3	Nail
P5	Screw
P6	Plug
P7	Pin
P8	Fuse

PO-PART

PO-NO	PART-NO	PART-QTY
111	P1	10
111	P2	5
111	P3	3
111	P5	6
112	P2	2
112	P5	1
113	P1	3
113	P3	4
114	P6	5
115	P7	8
116	P8	2

PO

PO-NO	PO-DATE	EMP-CODE	SUP P-NO
111	01012001	M2	222
112	01012001	S3	105
113	02012001	S1	111
114	02012001	M2	150
115	03012001	S1	222
116	04012001	S1	100

Further Normalization

- BCNF or Boyce–Codd Normal form
- 4th Normal form
- 5th Normal form

  In a normal situation normalization up-to 3NF is quite sufficient. Certain relations may even be de-normalized on account of efficiency. The Normalizations which are discussed next are not practically enforced most of the time.
- But a relation in 3NF does not guarantee that all anomalies have been removed, hence the additional normalizations.

Boyce Codd Normal Form

- Consider the below relation

Student	Course	Instructor

- It also has the following dependencies

{student,course}⟶ Instructor

Instructor⟶ course

Boyce Codd Normal Form

- This relation has more than one candidate key.(student,course) or (student,instructor)
- We have arbitrarily chosen the previous. Choosing the latter would also be fine
- The above relation is in 3NF, but not in BCNF
- But this has anomalies
  - If we need to add an instructor ‘silva’ for ‘database’ we have to wait till an appropriate student is present.
  - Deleting a record may also delete other data, specially if an instructor is in one students record (deleting the student will also erase the instructors data)

solution

- We divide the relation into 2 .
- The attribute that is a determinant but not a key (instructor) must be placed in a separate relation and becomes its key.

Instructor	Course

Instructor	Student

Definition of Third and Boyce-Codd Normal Form

3NF

- A relation schema R is in third normal form (3NF) if, whenever a nontrivial functional dependency X→ A holds in R, either

  (a) X is a superkey of R or

  (b) A is a prime attribute of R

BCNF

- A relation schema R is in BCNF if whenever a nontrivial functional dependency X→ A holds in R, then X is superkey of R