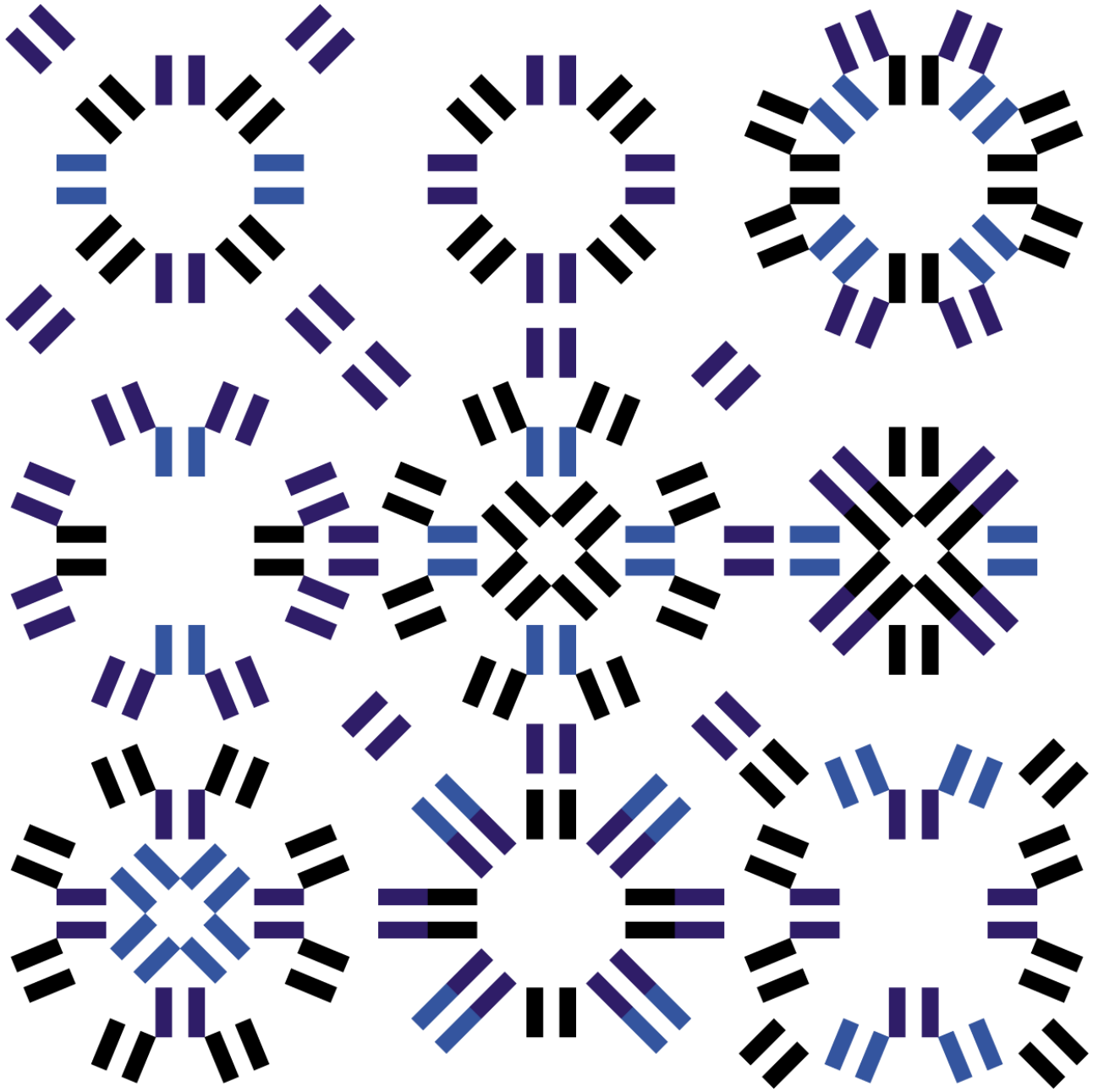


Data Bases
Design of Relational Database Schemas

Laurent D’ORAZIO

Université de Rennes

Introduction

- Many designs possible
- Some design better than others

- How to choose?

Running example – college application

Source: legacyca.com

Running example

- Information on college application
  - SSN
  - Name
  - Colleges applying to
  - High schools attended (with city)
  - Hobbies

- Proposed schema

```
Apply(SSN, sName, cName, HS, HScity, hobby)
```

- Example of application

123 Ann from PAHS and GHS, playing tennis and Trumpet, applying for Stanford, Berkley and MIT.

Design anomalies
Redundancy

- Problem
  - Information multiple time

- Example

```
123, Ann, Stanford, PAHS, PA, Tennis
123, Ann, Berkley, PAHS, PA, Tennis
123, Ann, MIT, PAHS, PA, Tennis
...
123, Ann, Stanford, GHS, PA, Tennis
...
123, Ann, Stanford, PAHS, PA, Trumpet
```

Design anomalies
Update anomaly

- Problem
  - Modify a record but not another

- Example

```
...
123, Ann, Stanford, PAHS, PA, ~~Trumpet~~Cornet
123, Ann, Berkley, PAHS, PA, Trumpet
...
```

Design anomalies
Deletion

- Problem
  - Delete a record and loss of information

- Example

```
...
456, Bob, Stanford, PAHS, PA, Surfing
...
```

Alternative

- Proposed schema

```
Student(SSN, sName)
Apply(SSN, cName)
HighSchool(SSN, HS)
Located(HS, HScity)
Hobbies(SSN, hobby)
```

Alternative
Some issues

- Proposed schema

```
Student(SSN, sName)
Apply(SSN, cName)
HighSchool(SSN, HS)
Located(HS, HScity)
Hobbies(SSN, hobby)
```

`HS` is not a key, `(HS,Hscity)` required to identify a college

`hobby` should be specific to the college
Example: Stanford and surfing

Discussion
Better design

- Proposed schema

```
Student(SSN, sName)
Apply(SSN, cName, hobby)
HighSchool(SSN, HS, city)
~~Located(HS, HScity)~~
~~Hobbies(SSN, hobby)~~
```

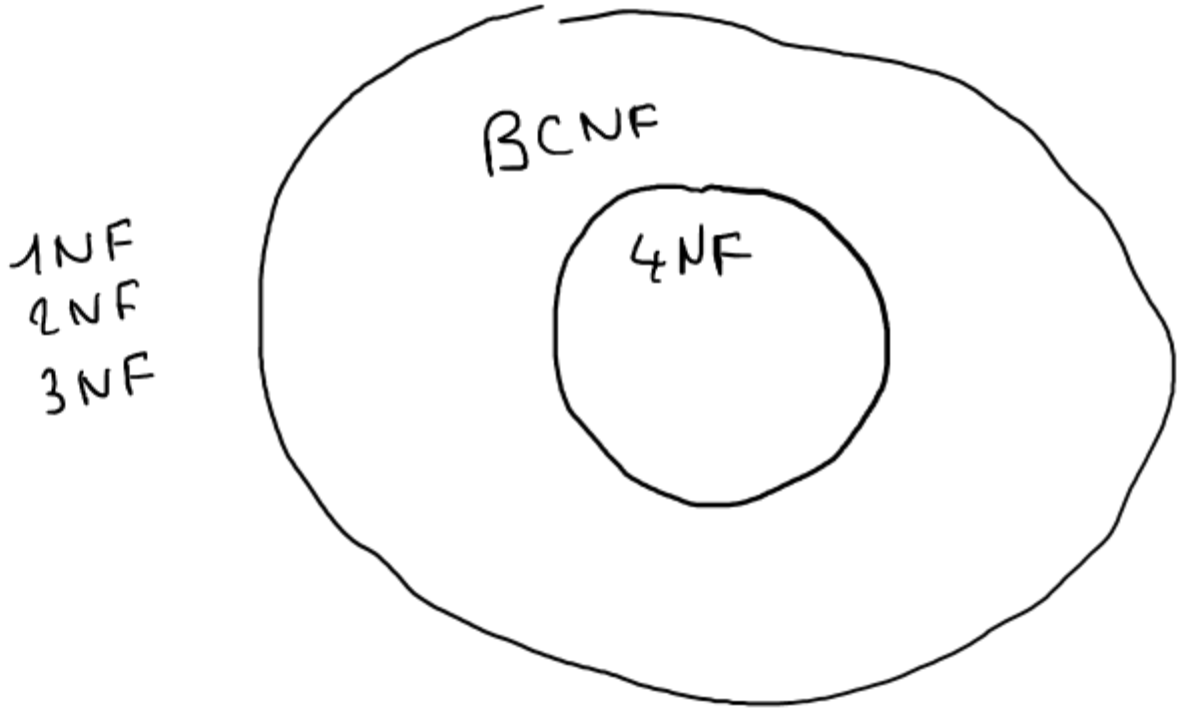
Design by decomposition

- Process
  - Star with “mega” relations with everything
  - Decompose into better and smaller relations with same information

- Automatic decomposition into set of relations satisfying normal form
  - Mega relations + properties of the data => no anomalies, no loss of information

Properties and Normal Forms

- Functional dependencies – Boyce Codd Normal Form
- Multivalued dependencies – Fourth Normal Form



Normal forms

- Properties
  - Good relations
  - No anomalies

Normal form
1NF

- Definition
  - Table in 1NF if each column contains atomic (indivisible) values

- Example (non compliant with 1NF)

Id	Name
123	Ann Collin

- Example (compliant with 1NF)

Id	Fname	Lname
123	Ann	Collin

Source: Wikipedia

Normal form
2NF

- Definition
  - 1NF and
  - None key attributes cannot be functionally dependant of a strict subset of the key

- Example (non compliant with 2NF)

CommandId	ArticleId	ArtDescription
1	10	HD TV...

ArticleId -> ArtDescription

- Example (compliant with 2NF)

CommandId	ArticleId
1	10

Source: Wikipedia

Normal form
3NF

- Definition
  - 2NF and
  - None key attributes do not functionally depend other non key attributes

- Example (non compliant with 3NF)

CommandId	ClientID	ClientName
1	100	Alice

ClientID -> ClientName

- Example (compliant with 3NF)

CommandID	ClientId
1	100

Id	Name
100	Alice

Source: Wikipedia

Normal form
BCNF

- Definition
  - 3NF and
  - Key attributes are not functionally dependant of a none key attribute

- Example (non compliant with BCNF)

SSN	Country	Name	Region
123	France	Ann	Bretagne
456	France	Bob	Bretagne

Region -> Country

- Example (compliant with BCNF)

SSN	Region	Name
123	Bretagne	Ann
456	Bretagne	Bob

Region	Country
Bretagne	France

Source: S. Crozat

Bibliography

- Database Systems: The Complete Book: International Edition : Garcia-Molina, Hector, Ullman, Jeffrey D., Widom, Jennifer