Databases

-

Chapters 3 & 4: The relational Model & The Normalization Technique

Rui Oliveira © rui.oliveira@ipleiria.pt

References

Further study required:

- ***"Fundamentals of Database Systems"***, Elmasri & Navathe, 6th Edition, Adison Wesley, 2016, *Chapter 3: The Relational Data Model and Relational Database Constraints*

- ***"Fundamentals of Database Systems"***, Elmasri & Navathe, 6th Edition, Adison Wesley, 2016, *Chapter 15: Basics of Functional Dependencies and Normalization for Relational Databases*

The Relational Model: concepts

Relation (*table*)

- Tuple (*row*)
- Attribute (*column*)

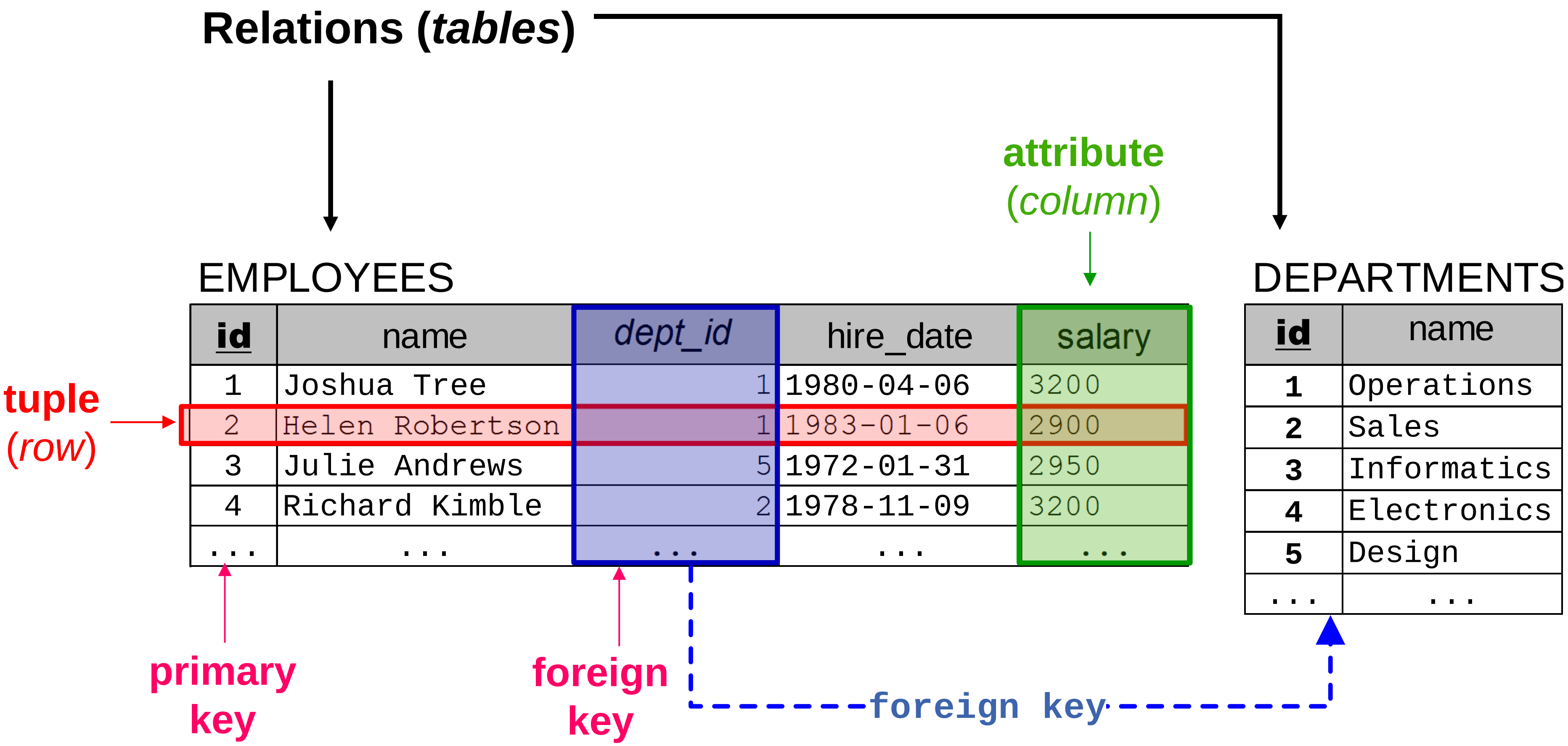
Keys

- Primary key
- Candidate key

Referential integrity

- Foreign key

The Relational Model: concepts



Relations (*tables*)

attribute
(*column*)

tuple
(*row*)

EMPLOYEES

id	name	*dept_id*	hire_date	salary
1	Joshua Tree	1	1980-04-06	3200
2	Helen Robertson	1	1983-01-06	2900
3	Julie Andrews	5	1972-01-31	2950
4	Richard Kimble	2	1978-11-09	3200
...	...	...	...	...

DEPARTMENTS

id	name
1	Operations
2	Sales
3	Informatics
4	Electronics
5	Design
...	...

primary key

foreign key

foreign key

The NormalizationTechnique

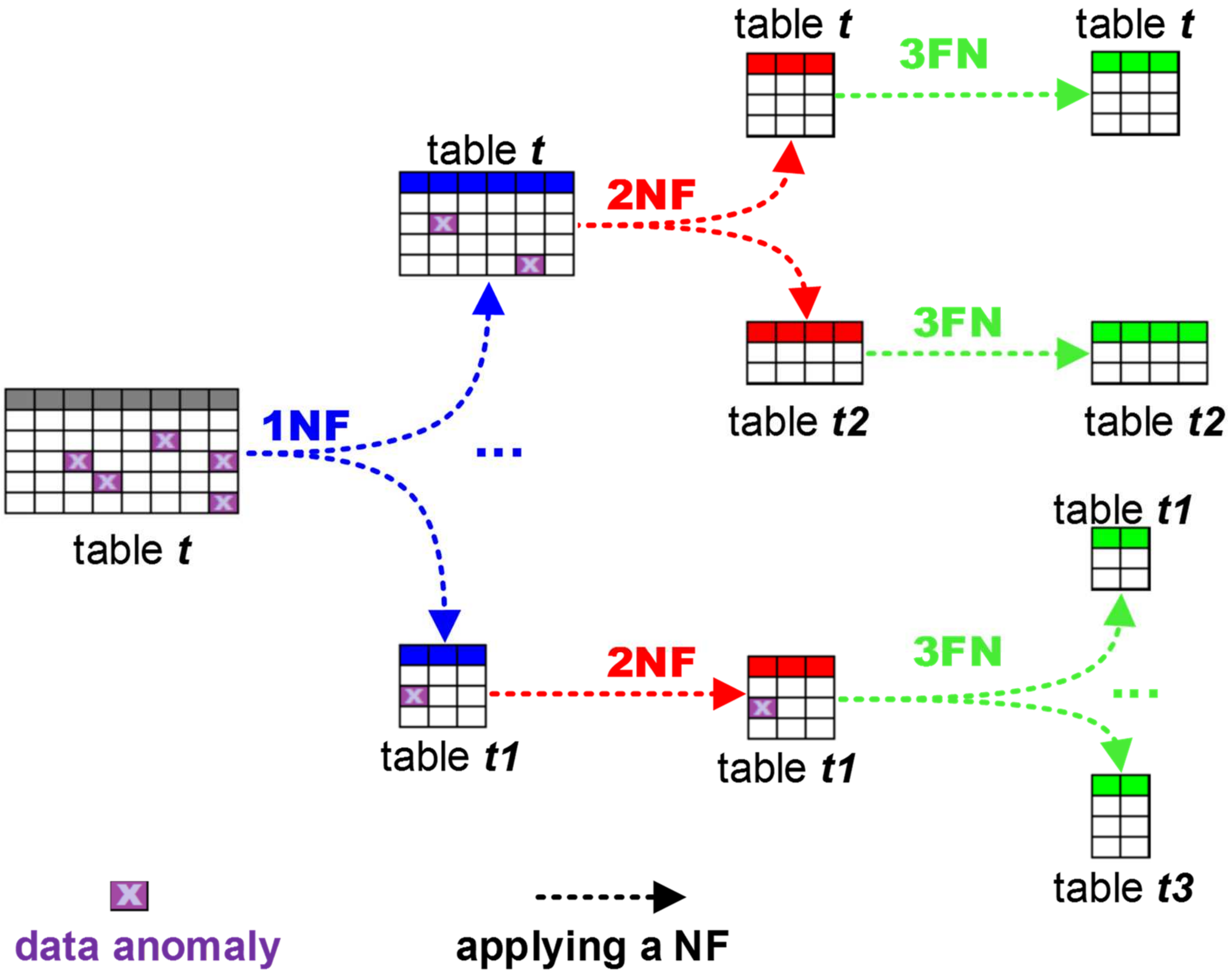
Goals: detect and solve anomalies found in data

- Redundant data
- NULL values
- Anomalies caused during inserting/updating/removing data

How does it works?

- Using a sequence of methods called ***Normal Forms*** **(NF)**
- Each **NF** splits a table ***t*** with an anomaly ***a*** by creating a new table, thus removing that anomaly ***a***
- Table ***t*** will still exist but with less data and/or with less columns

A possible scenario



Quiz

How many Normal Forms do exist?

a) 4

b) 5

c) 6

d) 7

Quiz

A Normal Form can: (choose three)

a) Enlarge a table structure.

b) Eliminate the structural anomalies of a table.

c) Add more columns to a table.

d) Create repeated data in a table.

e) Divide the table.

f) Eliminate an existing table.

The First Normal Form 1NF

step 1 - search for data anomalies

- Non-atomic attributes (+1 value per cell)
- Attributes representing the same characteristic

step 2 – solve data anomalies

- (*approach 1*) Move anomalous attributes to new table

or

- (*approach 2*) Restructure the current table

The First Normal Form 1NF

Scenario

CLIENTS

id	name	city	region	phoneNr1	phoneNr2	birthDate	taxPayerNr
1	António Freitas	Leiria	South	244101010		1980-04-06	1230009
2	Rita Marujo	Lisboa	South	210202020	931020300	1983-01-06	1230002
3	Carlos da Silva	Coimbra	Center			1972-01-31	1230004
4	Ana Oliveira	Leiria	South	244505050	244606060	1978-11-09	1230005
...	...	...	...	...	...	...	...

The First Normal Form **1NF**

step 1 - search for data anomalies

- Attributes representing the same characteristic

CLIENTS

the same characteristic

<u>id</u>	name	city	region	phoneNr1	phoneNr2	birthDate	taxPayerNr
1	António Freitas	Leiria	South	244101010		1980-04-06	1230009
2	Rita Marujo	Lisboa	South	210202020	931020300	1983-01-06	1230002
3	Carlos da Silva	Coimbra	Center			1972-01-31	1230004
4	Ana Oliveira	Leiria	South	244505050	244606060	1978-11-09	1230005
...	...	...	...	...	...	...	...

The First Normal Form **1NF**

step 2 – solve data anomalies

- (*approach 1*) Move anomalous attributes to new table

CLIENTS

id	name	city	region	~~phoneNr1~~	~~phoneNr2~~	birthDate	taxPayerNr
1	António Freitas	Leiria	South	~~244101010~~		1980-04-06	1230009
2	Rita Marujo	Lisboa	South	~~210202020~~	~~931020300~~	1983-01-06	1230002
3	Carlos da Silva	Coimbra	Center			1972-01-31	1230004
4	Ana Oliveira	Leiria	South	~~244505050~~	~~244606060~~	1978-11-09	1230005
...	...	...	...	~~...~~	~~...~~	...	...

turns into

CLIENTS

id	name	city	region	birthDate	taxPayerNr
1	António Freitas	Leiria	South	1980-04-06	1230009
2	Rita Marujo	Lisboa	South	1983-01-06	1230002
3	Carlos da Silva	Coimbra	Center	1972-01-31	1230004
4	Ana Oliveira	Leiria	South	1978-11-09	1230005
...	...	...	...	...	...

Phone_nrs

client_id	phone_nr
1	244101010
2	210202020
2	931020300
4	244505050
4	244606060
...	...

foreign key

The First Normal Form 1NF

step 2 – solve data anomalies

- (*approach 2*) Restructure the current table

CLIENTS

id	name	city	region	~~phoneNr1~~	~~phoneNr2~~	birthDate	taxPayerNr
1	António Freitas	Leiria	South	~~244101010~~		1980-04-06	1230009
2	Rita Marujo	Lisboa	South	~~210202020~~	~~931020300~~	1983-01-06	1230002
3	Carlos da Silva	Coimbra	Center			1972-01-31	1230004
4	Ana Oliveira	Leiria	South	~~244505050~~	~~244606060~~	1978-11-09	1230005
...	...	...	...	~~...~~	~~...~~	...	...

turns into

CLIENTS

id	name	city	region	phoneNr	birthDate	taxPayerNr
1	António Freitas	Leiria	South	244101010	1980-04-06	1230009
2	Rita Marujo	Lisboa	South	210202020	1983-01-06	1230002
2	Rita Marujo	Lisboa	South	931020300	1983-01-06	1230002
3	Carlos da Silva	Coimbra	Center		1972-01-31	1230004
4	Ana Oliveira	Leiria	South	244505050	1978-11-09	1230005
4	Ana Oliveira	Leiria	South	244606060	1978-11-09	1230005
...	...	...	...	...	...	...

The First Normal Form 1NF

A similar scenario

CLIENTS

id	name	city	region	phoneNrs	birthDate	taxPayerN
1	António Freitas	Leiria	South	244101010	1980-04-06	1230009
2	Rita Marujo	Lisboa	South	210202020, 931020300	1983-01-06	1230002
3	Carlos da Silva	Coimbra	Center		1972-01-31	1230004
4	Ana Oliveira	Leiria	South	244505050, 244606060	1978-11-09	1230005
...	...	...	...	...	...	...

The First Normal Form **1NF**

step 1 - search for data anomalies

- Non-atomic attributes (+1 value per cell)

CLIENTS

id	name	city	region	phoneNrs	birthDate	taxPayerNr
1	António Freitas	Leiria	South	244101010	1980-04-06	1230009
2	Rita Marujo	Lisboa	South	210202020, 931020300	1983-01-06	1230002
3	Carlos da Silva	Coimbra	Center		1972-01-31	1230004
4	Ana Oliveira	Leiria	South	244505050, 244606060	1978-11-09	1230005
...	...	...	...	...	...	...

the atribute is not atomic!

step 2 – solve data anomalies

- *...exactly as in the previous slides...*

Quiz

TASKS

id	description	lvl1	lvl2	lvl3	address	start_date	end_date
1	Automate; Deploy	x			11th Street, Washington, D.C	2018-01-12	2018-01-14
2	Automate	x	x		11th Street, Washington, D.C	2018-09-12	
3	Travel+Meeting	x	x		Reagan Institute, NY	2018-06-03	2018-06-04
4	Deploy;Gather logs	x		x	Lincon Avenue, 3rd floor, office A10	2018-10-12	
...	...	...	...	...	...	...	...

A. Which columns disrespect the 1NF?

B. Build two normalized versions of the scenario.

The Second Normal Form 2NF

step 1 – apply 1NF

step 2 – analyze data dependencies

- Find each table primary key (PK)
- Build the ***Functional Dependencies Diagram(s)***

step 3 – identify anomalous dependencies

- Dependencies on just **part** of the PK

step 4 – split table

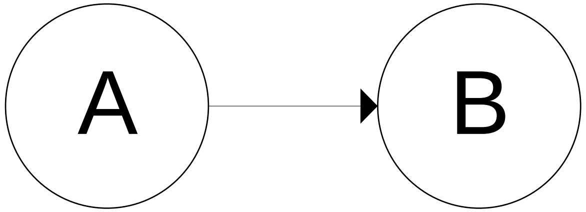
- Move anomalous dependencies to new table(s): one table for each determinant

The Second Normal Form 2NF

First, a core concept:

Functional Dependency

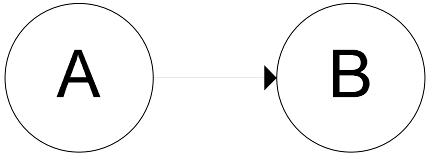
(*dependência funcional*)



or A -> B

The Second Normal Form 2NF

A functional dependency between two attributes ***A*** and ***B***



or A -> B

is read

- ***B*** is **functionally dependent** of ***A***
- ***A*** is the **determinant** of ***B***

and means

For each distinct value of *A*,
there is <u>only one</u> distinct value of *B*

The Second Normal Form 2NF

A typical schoolar scenario

Which functional dependencies may be true?

`student_nr -> student_name` **True**

false `student_name -> student_nr`

`student_nr -> student_birth_date` **True**

`student_nr -> student_citizen_id` **True**

false `student_citizen_nr -> student_nr`

Quiz

Consider the following database table.

column1	column2
101	770
101	808
202	770

Which of the following functional dependencies is true about this data?

a) column1 --> column2
b) column2 --> column1
c) None of the previous

Quiz

Fill the table with data about students so that it respects the following rules:

name --> address *and* **address --> name**

name	address

The Second Normal Form 2NF

Scenario

ROLES

emp_nr	dept_name	dept_city	dept_country	emp_name	emp_role
1001	Marketing	NY	USA	Chang	Manager
1001	Engineering	Vienna	Austria	Chang	Contractor
1008	Operations	NY	USA	Andrew	Contractor
1008	Engineering	Vienna	Austria	Andrew	Senior Engineer
2005	Engineering	Vienna	Austria	Selma	Junior Engineer
2005	Accounting	Heins	Austria	Selma	Junior Engineer
...	...	...	...	...	...

- Each department exists in only one city
- A city name won’t be repeated in different countries
- Each employee, in each department, will have only one role

The Second Normal Form 2NF

step 1 – apply the 1NF

- Non-atomic attributes (+1 value per cell) ✓
- Attributes representing the same characteristic ✓

no anomalies found

Table ROLES remains unchanged.

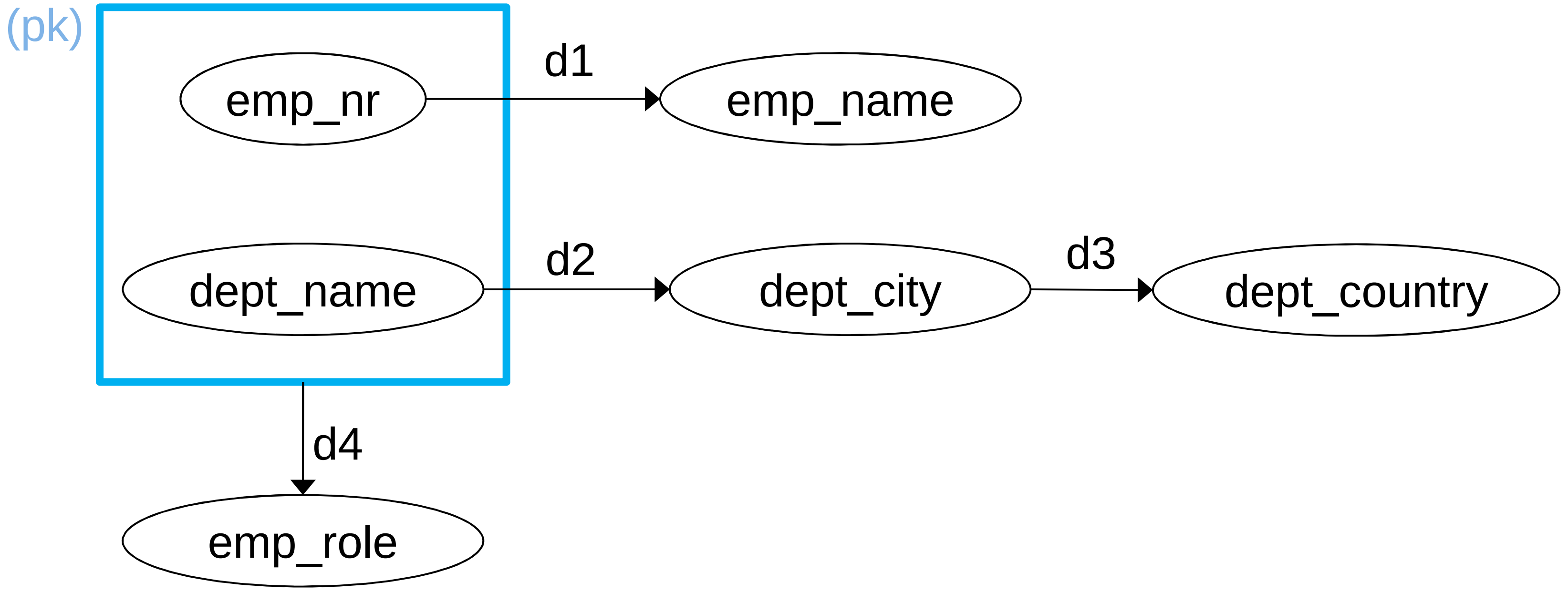
The Second Normal Form 2NF

step 2 – analyze data dependencies

- Find each table primary key (PK)

  PK its the pair <emp_nr, dept_name>

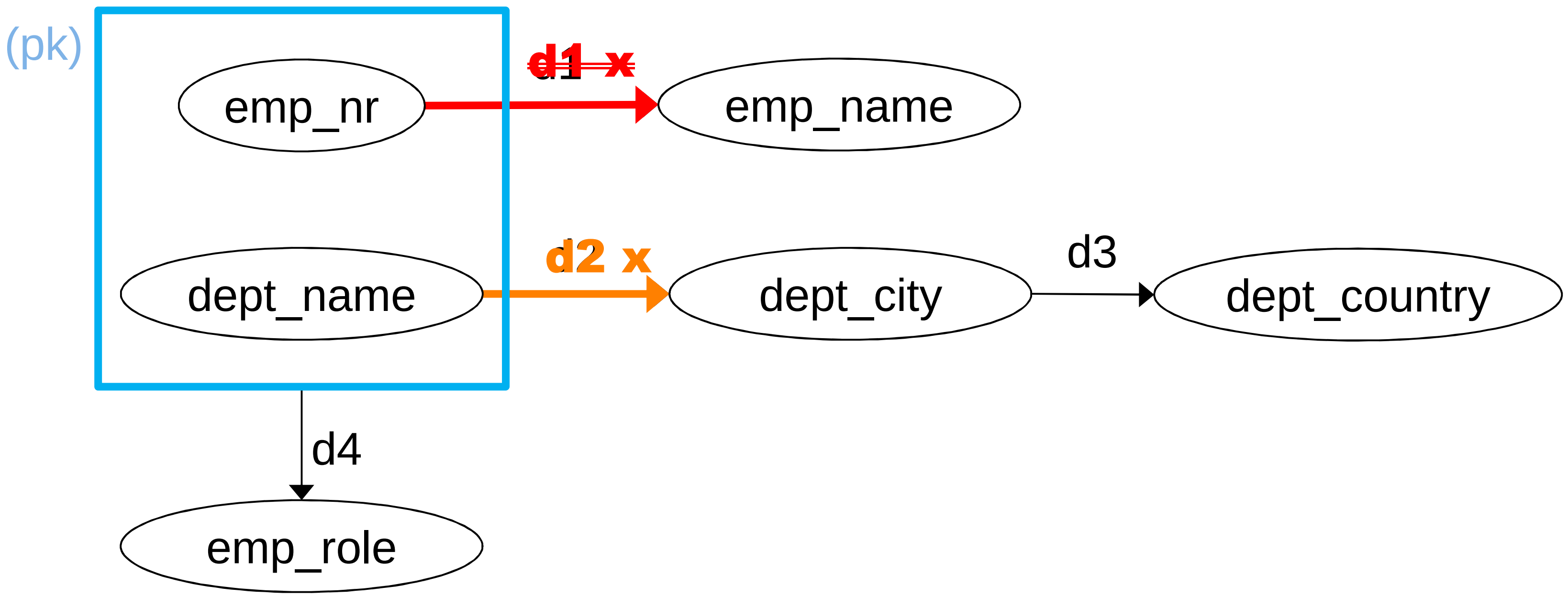
- Build the ***Functional Dependencies Diagram***



The Second Normal Form 2NF

step 3 – identify anomalous dependencies

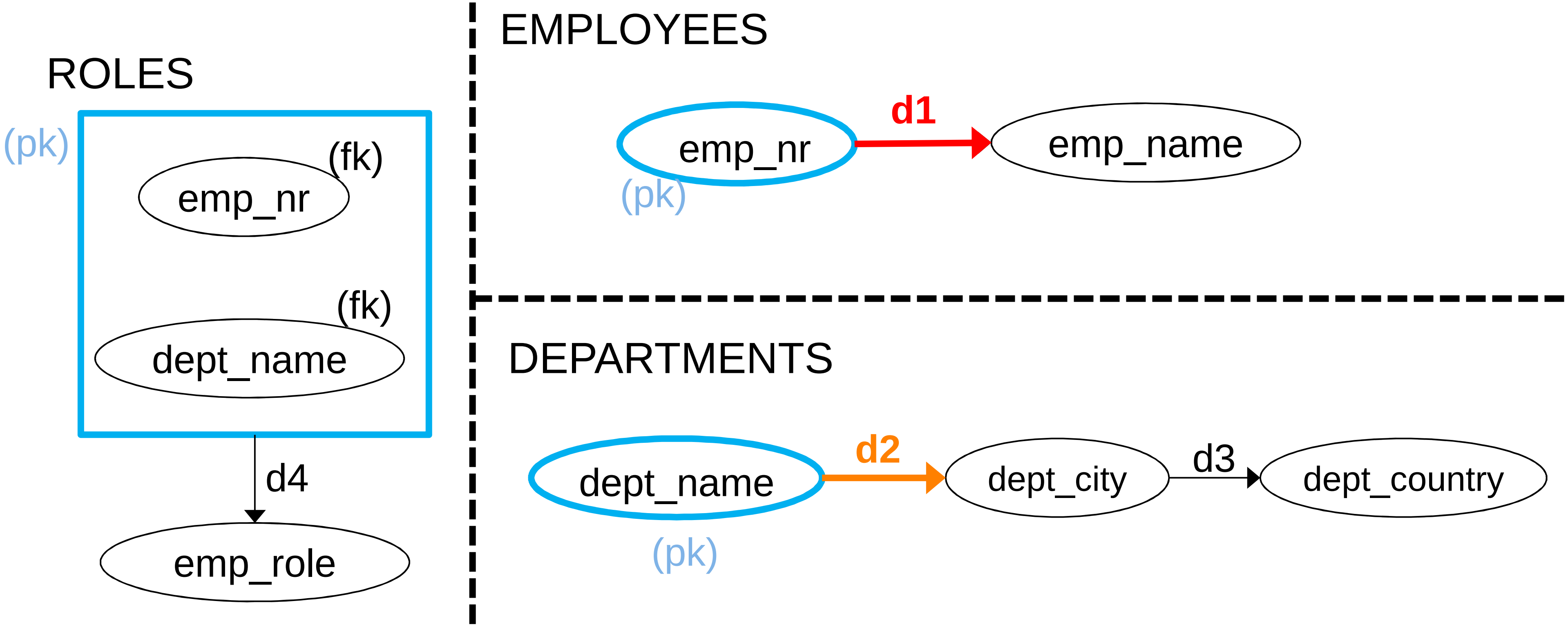
- Dependencies on just **part** of the PK



The Second Normal Form **2NF**

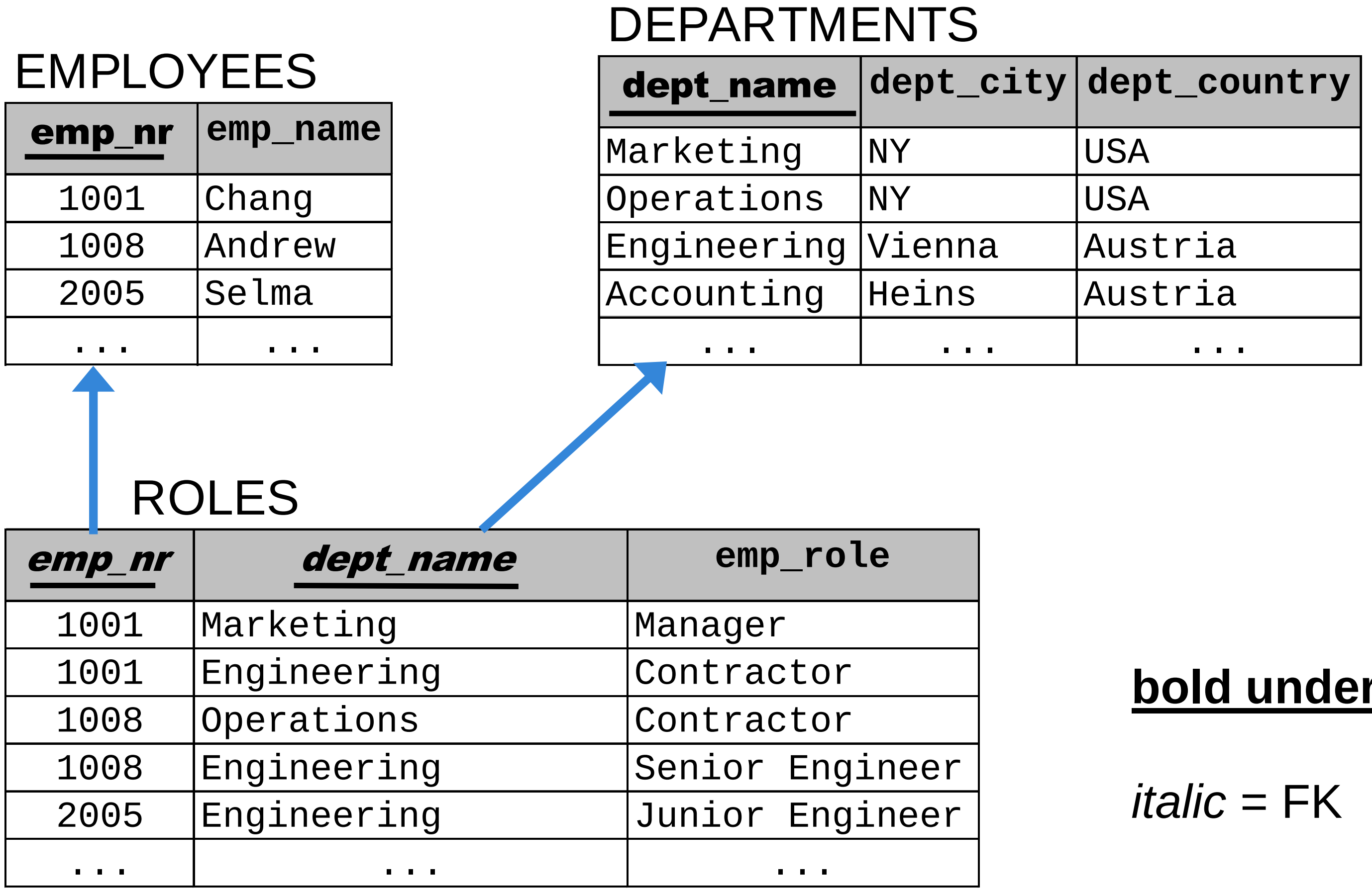
step 4 – split table

- Move anomalous dependencies to new table(s): one table for each determinant



The Second Normal Form 2NF

Final result after 2NF



EMPLOYEES

emp_nr	emp_name
1001	Chang
1008	Andrew
2005	Selma
...	...

DEPARTMENTS

dept_name	dept_city	dept_country
Marketing	NY	USA
Operations	NY	USA
Engineering	Vienna	Austria
Accounting	Heins	Austria
...	...	...

ROLES

emp_nr	***dept_name***	emp_role
1001	Marketing	Manager
1001	Engineering	Contractor
1008	Operations	Contractor
1008	Engineering	Senior Engineer
2005	Engineering	Junior Engineer
...	...	...

bold underlined = PK

italic = FK

The Second Normal Form 2NF

...or, using a more formal representation

EMPLOYEES (**emp_nr**, emp_name)

DEPARTMENTS (**dept_name**, dept_city, dept_country)

EMPLOYEES ↗ DEPARTMENTS ↗

ROLES (***emp_nr***, ***dept_name***, emp_role)

bold underlined = PK *italic* = FK

Quiz

BANK_ACCOUNTS

IBAN	location	account_extras	client_id	client_name	account_holder	balance	account_type	interest
0035 20202020	London	credit card health ensurance	11001	Carlos Sousa	Yes	123.03	Pentions	2.50%
0035 20202020	London	credit card health ensurance	12004	Jorge Ferreira	No	123.03	Pentions	2.50%
0035 30303030	Glasgow	primary salary	13006	Miguel Carmo	Yes	298	Current	0.40%
0035 30303030	Glasgow	primary salary	11001	Carlos Sousa	No	298	Current	0.40%
0035 40404040	London	primary salary life ensurance	11009	Pedro Mico	Yes	1148	Current	0.40%
0035 50505050	Glasgow	credit card life ensurance primary salary	11001	Carlos Sousa	Yes	329	Special1	2.50%
...	...	...	...	...	...	...	...	...

Assume that:

- Each account is identified by its IBAN (International Bank Account Number).
- Each account has one account holder.
- Each account type has one interest rate.

Q: Apply the 2NF to table BANK_ACCOUNTS

Simplifying a Functional Dependencies Diagram

Rules

- Transitivity (*transitividade*)

  IF A->B and B->C are TRUE

  THEN A->C is also TRUE but unnecessary

- Augmentation (*aumentatividade*)

  IF A->B is true

  THEN (A,C) -> B is also TRUE but unnecessary

Don’t forget...

«Most practical design projects acquire existing designs of databases from previous designs, designs in legacy models, or from existing files.»

«(...) database design as practiced in industry today pays particular attention to normalization only up to 3NF, BCNF, or at most 4NF.»

in Fundamentals of Database Systems, 6th edition

Another core concept

Candidate key

(*chave candidata*)

Any atribute which could be chosen as primary key

(only the simpler ones are considered)

Candidate keys: example

CLIENTS

id	name	city	region	phoneNrs	birthDate	taxPayerNr
1	António Freitas	Leiria	South	244101010	1980-04-06	1230009
2	Rita Marujo	Lisboa	South	210202020, 931020300	1983-01-06	1230002
3	Carlos da Silva	Coimbra	Center		1972-01-31	1230004
4	Ana Oliveira	Leiria	South	244505050, 244606060	1978-11-09	1230005
...	...	...	...	...	...	...

Assuming that every client has a **non nullable unique** *taxPayerNr*, then:

- candidate keys = **id** , **taxPayerNr**
- primary key = **id**

Quiz

What types of keys may exist in a relational table? (choose all that apply)

a) Nullable key
b) Primary key
c) Candidate key
d) Optional key
e) Foreign key
f) Foreigner key
g) Relational key

Quiz

Choose the sentences which are always correct (choose two).

a) A candidate key is a primary key.
b) A candidate key is nullable.
c) A table may have two or more primary keys.
d) A primary key is a candidate key.
e) A foreign key can have repeated values.

The Third Normal Form 3NF

step 1 – apply 2NF

step 2 – analyze data dependencies

- Find each table candidate key
- Build the ***Functional Dependencies Diagram(s)***

step 3 – identify anomalous dependencies

- Non candidate keys depending on non candidate keys

step 4 – split table

- Move anomalous dependencies to new table(s): one table for each determinant

The Third Normal Form 3NF

step 1 – apply the 2NF

- Already done! ✓

step 2 – analyze data dependencies

- Find each table candidate key
- Build the ***Functional Dependencies Diagram(s)***

The Third Normal Form 3NF

step 2 – analyze data dependencies

- Find each table candidate key
  - EMPLOYEES: candidate key = **emp_nr**
  - DEPARTMENTS: candidate key = **dept_name**
  - ROLES: candidate key = (**emp_nr**, **dept_name**)

- Build the ***Functional Dependencies Diagram(s)***
  - Already done! ✓

The Third Normal Form 3NF

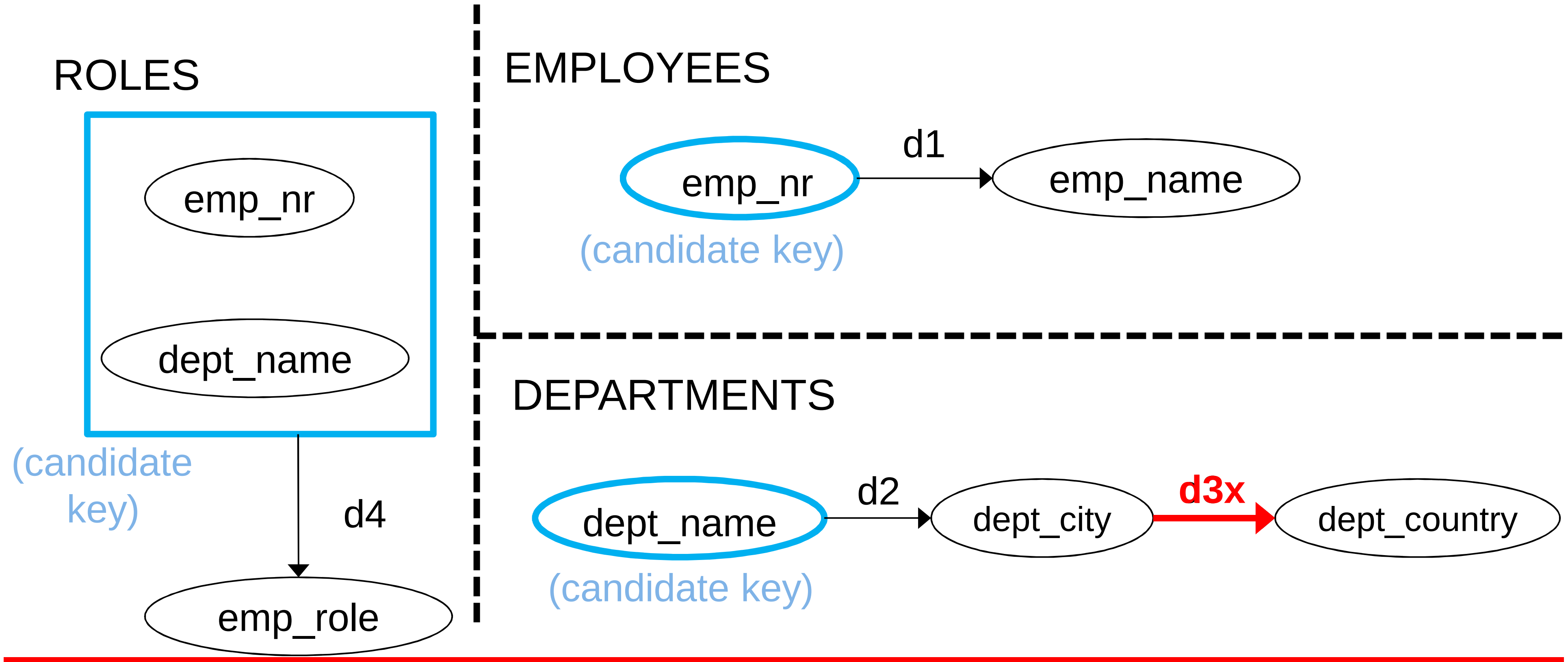
step 3 – identify anomalous dependencies

- Non key attributes depending on non key attributes

The Third Normal Form 3NF

step 3 – identify anomalous dependencies

- Non candidate keys depending on non candidate keys

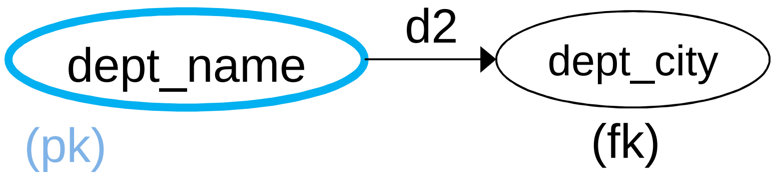


The Third Normal Form 3NF

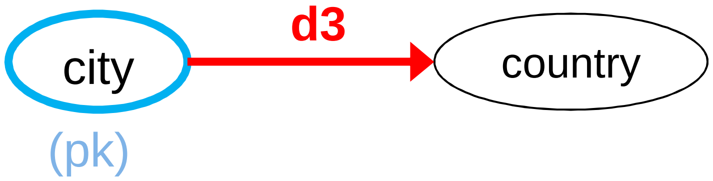
step 4 – split table

- Move anomalous dependencies to new table(s): one table for each determinant

DEPARTMENTS

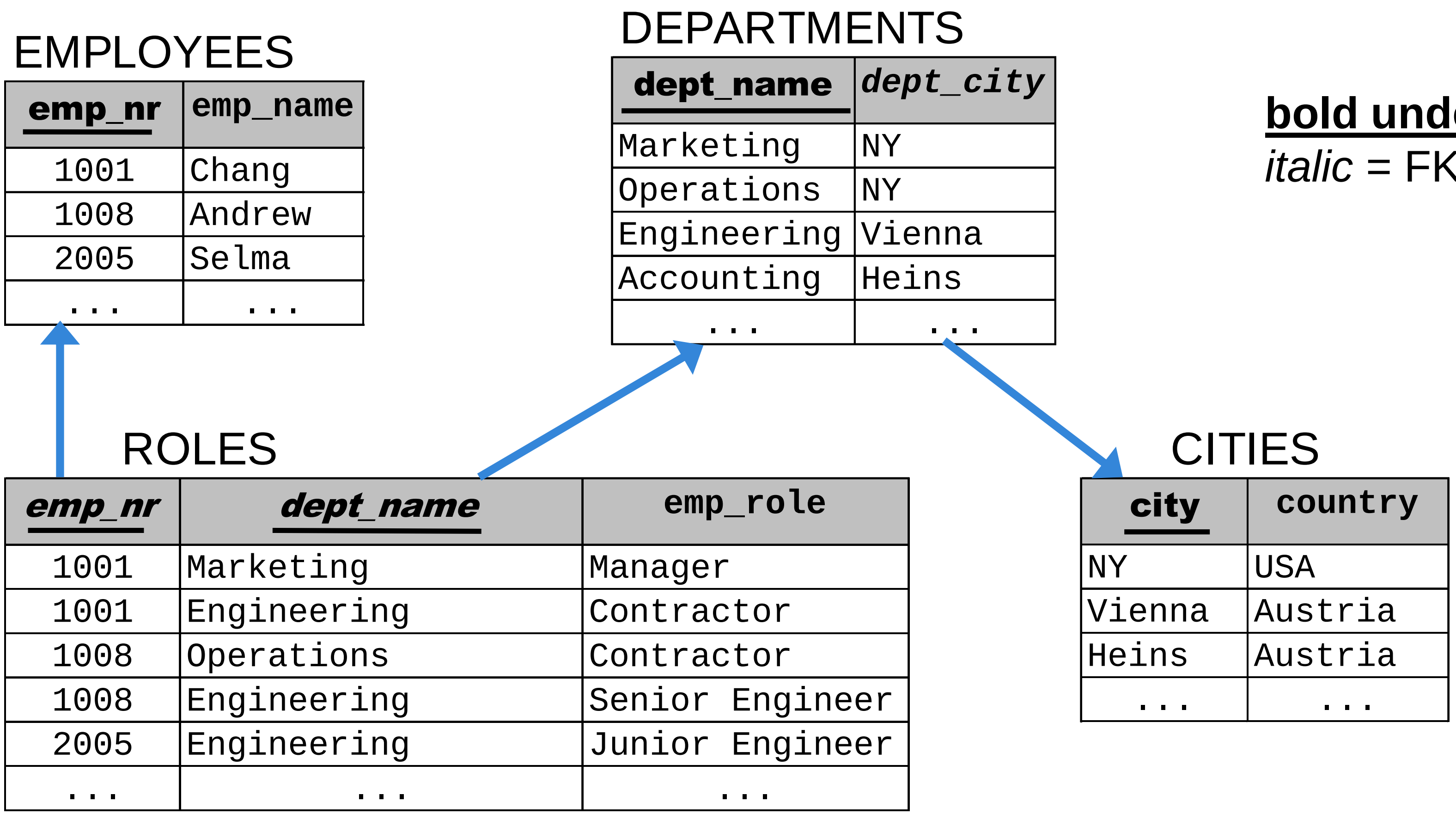


CITIES



The Third Normal Form 3NF

Final result after 3NF



EMPLOYEES

emp_nr	emp_name
1001	Chang
1008	Andrew
2005	Selma
...	...

DEPARTMENTS

dept_name	*dept_city*
Marketing	NY
Operations	NY
Engineering	Vienna
Accounting	Heins
...	...

ROLES

emp_nr	***dept_name***	emp_role
1001	Marketing	Manager
1001	Engineering	Contractor
1008	Operations	Contractor
1008	Engineering	Senior Engineer
2005	Engineering	Junior Engineer
...	...	...

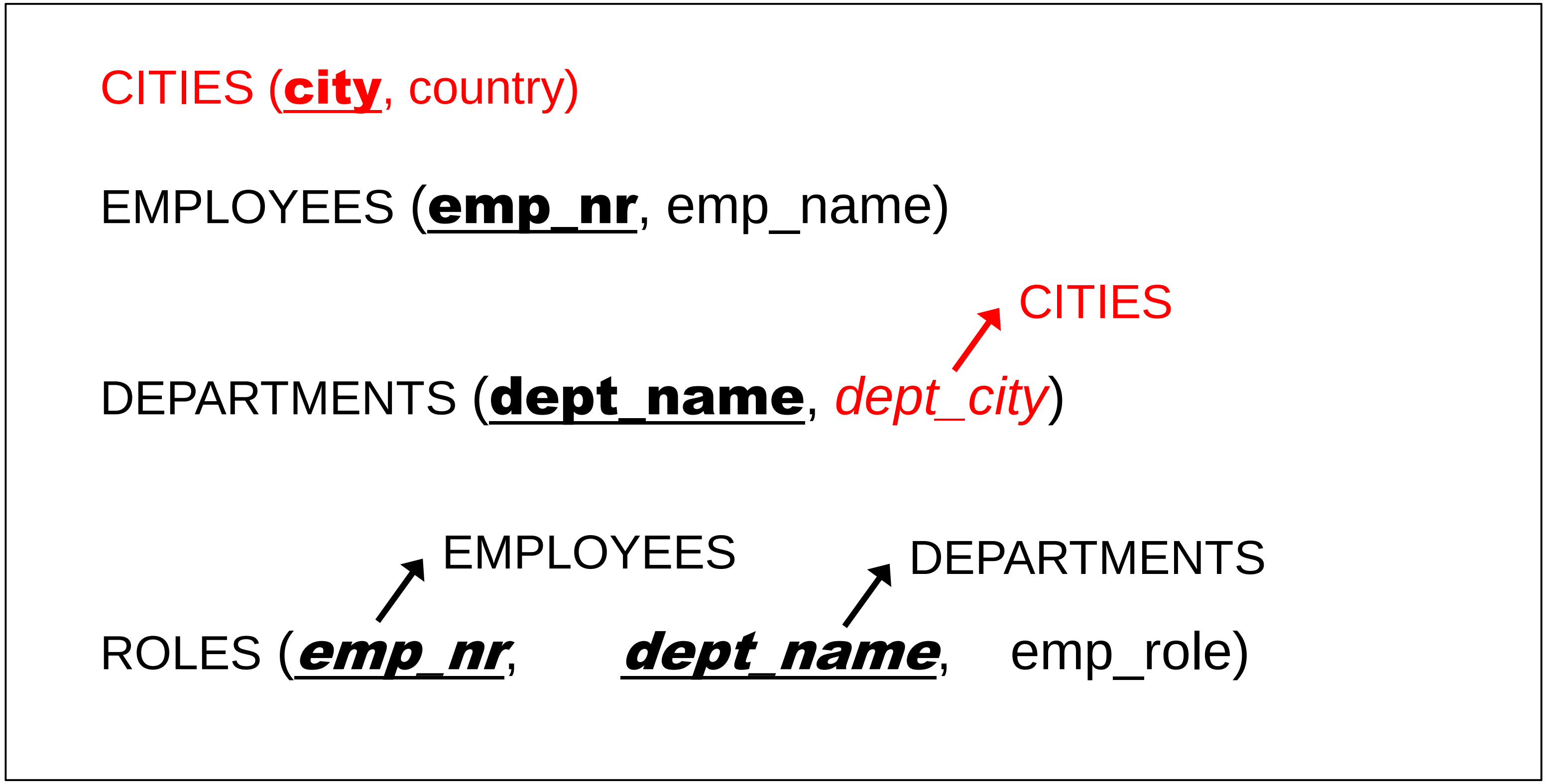
CITIES

city	country
NY	USA
Vienna	Austria
Heins	Austria
...	...

bold underlined = PK
italic = FK

The Third Normal Form 3NF

...or, using a more formal representation



CITIES (**city**, country)

EMPLOYEES (**emp_nr**, emp_name)

DEPARTMENTS (**dept_name**, *dept_city*)

ROLES (***emp_nr***, ***dept_name***, emp_role)

bold underlined = PK   *italic* = FK

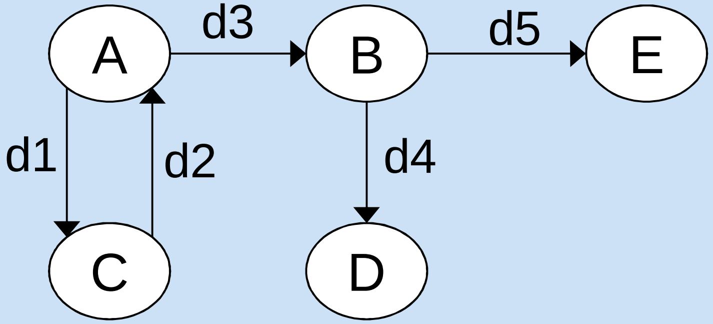
Quiz

In order to minimize resource consumption and also to minimize data updating operations in the database, how would you optimize the scenario tables after applying the 3FN?

Quiz

Consider the following FDD.

We know that A and C are candidate keys.



One or more dependencies disrespect the 3NF. Which? (choose one)

a) Just d1

b) d1 and d2

c) Just d3

d) d4 and d5

e) Just d5