Databases
-
The Conceptual approach in DB Design

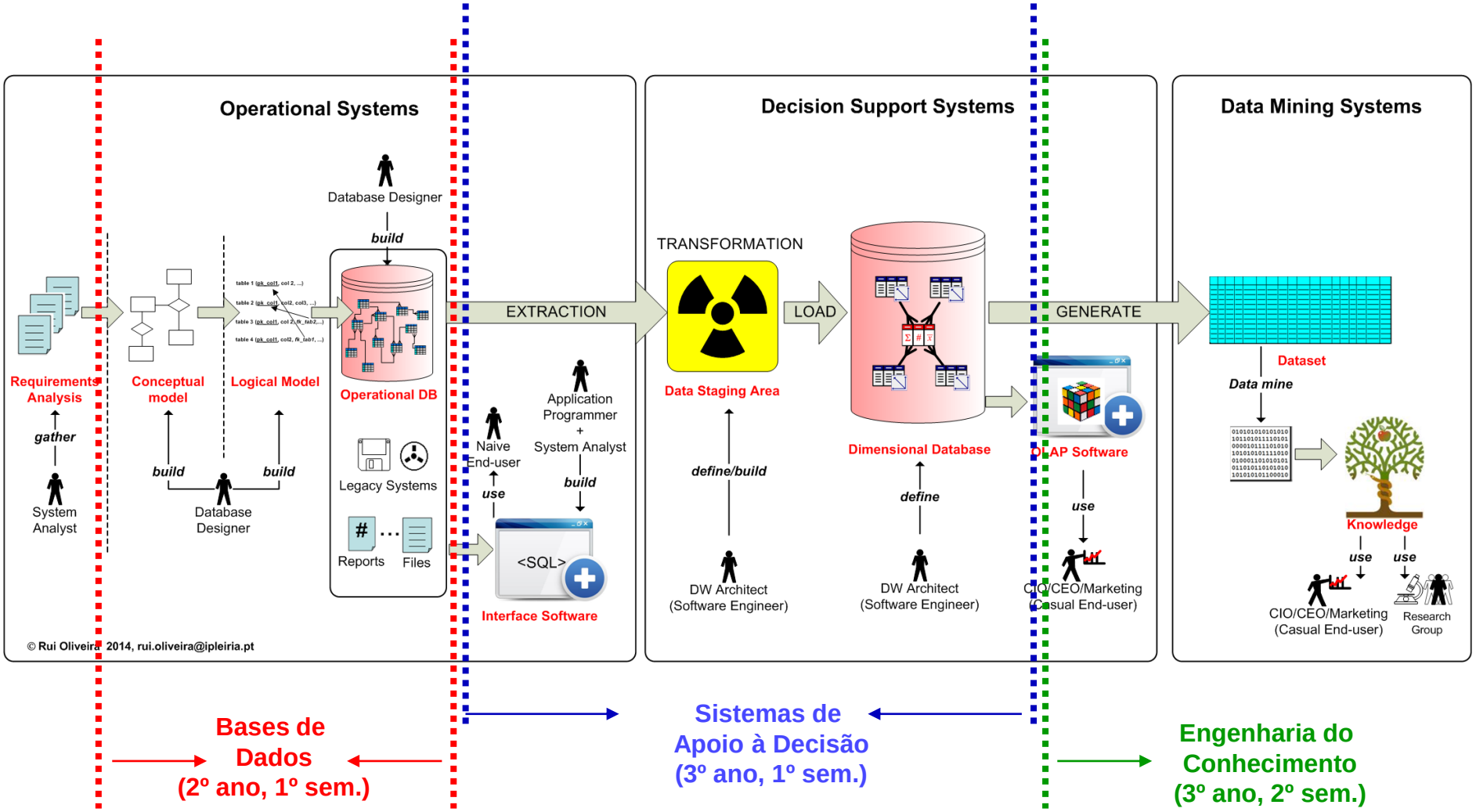
Rui Oliveira © rui.oliveira@ipleiria.pt

References

Further study required

- "Fundamentals of Database Systems", Elmasri & Navathe, 6th Edition, Adison Wesley, 2011:
  - *Chapter 7: Data Modeling Using the Entity-Relationship (ER) Model*
  - *Chapter 8: The Enhanced Entity-Relationship (EER) Model*
  - *Chapter 9: Relational Database Design by ER- and EER-to-Relational Mapping*

Databases: macro view



Relational DB Project: goals

A well modelled relational database should:

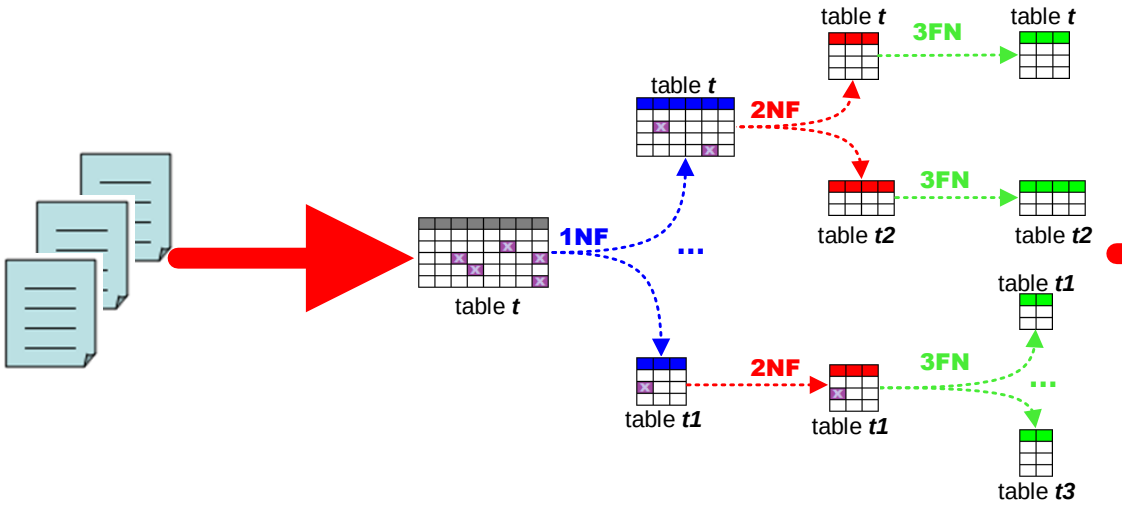
1. Answer only to end-users' needs
2. Avoid information redundancy
3. Avoid empty cells on tables, specially FKs
4. Avoid disk space waste
5. Simplify *update*, *insert* and *delete* operations on data

DB design using Normalization

(Previously on “Chapter 4”)

Scenario requirements

Logical Model



Physical Model

```
TABLE t
  name VARCHAR2(100)
  id NUMBER(5)
  nif NUMBER(9) UNIQUE
  ...

VIEW v1
  t.id
  t2.job
  ...

INDEX idx_t
  target: t.name
  type: BTree
...
```

DB design using a Conceptual approach

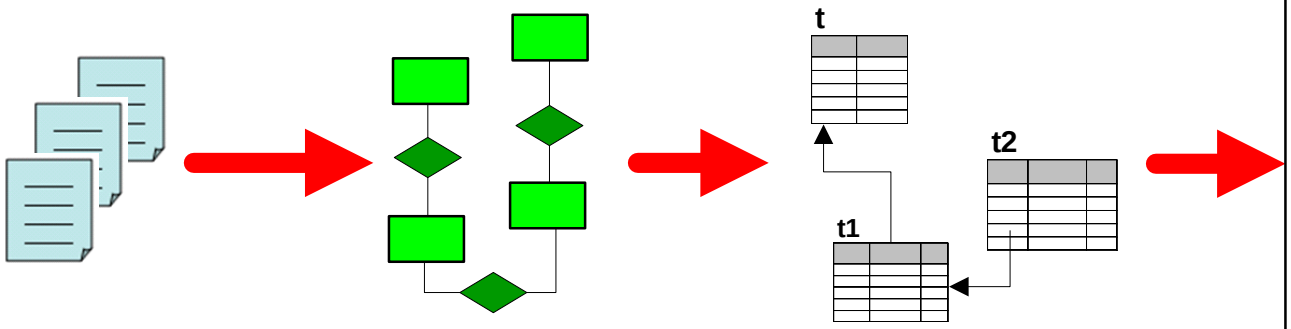
(and now)

Scenario requirements

Entity-Relationship Diagram

Logical Model

Physical Model




```
TABLE t
  name VARCHAR2(100)
  id NUMBER(5)
  nif NUMBER(9) UNIQUE
  ...

VIEW v1
  t.id
  t2.job
  ...

INDEX idx_t
  target: t.name
  type: BTree
...
```

Importance of the Conceptual design (in general)

Truth #1

A data model allows engineers to

share a standardized view of data relationships.

Truth #2

A Conceptual Model, in particular, defines data relationships in a more *natural* language than Normalization

Approach comparison

Programming languages	**Assembly** ----------→ **SQL**
DB Design	**Normalization** --------→ **Conceptual**

Rui Oliveira © rui.oliveira@ipleiria.pt

Why Conceptual Design (vs normalization)?

- Nearer to Human reasoning
- Easier to scale
- Easier to read

Conceptual DB Design: steps

1st, Requirements Analysis

1.1 Define main business processes (BP)

- What is the company all about?

1.2 Define business rules for the BP

- How does the real system works?

1.3 Understand information needs regarding the BP

- Which questions need an answer?
- Distinguish operational from analytical information

Rui Oliveira © rui.oliveira@ipleiria.pt

Conceptual DB Design: steps

2nd, Entity-Relationship Diagram (ERD)

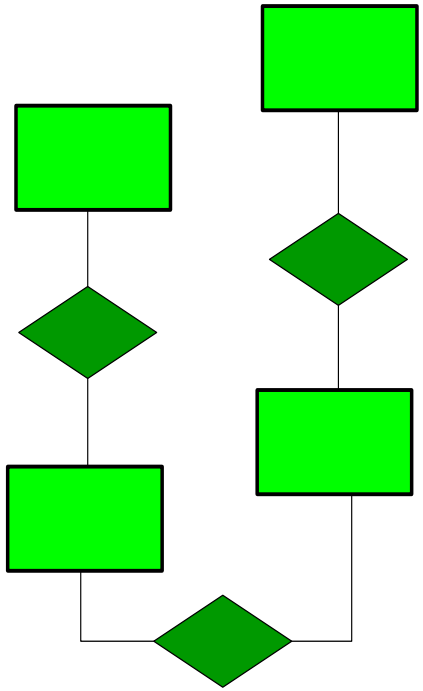
(*Diagrama de Entidade-Relationamento*)

2.1 Define entities

- Events
- Active actors
- Passive actors
- Types/categories

2.2 Document:

- Entities and Relationships
- Attributes
- Domains



Conceptual DB Design: steps

3rd, Logical Model

3.1 Transform the ERD --> Logical Model

(specific rules apply, see *Moodle*)

3.2 Validate the ERD

- `Test information retrieval`
- `Apply the Normalization Technique`

Conceptual DB Design: steps

4th, Physical Model

4.1 Define

- Datatypes
- Domains
- Candidate keys/Uniqueness rules

4.2 Document extra data integrity rules

- Data relationships (ex: col1>=col2, birth date<system date)

Rui Oliveira © rui.oliveira@ipleiria.pt

Quiz

Consider the following tasks:

1-Gather requirements **2**-Do Physical Model

3-Do Logical Model **4**-Do Conceptual Model

Which of the following sequences correctly describes the design of an operational database?

a) 1-2-4-3

b) 1-3-4-2

c) 1-4-3-2

d) 4-1-3-2

Example

Case study

- World's most visited institution: a **Retail Sales Company**

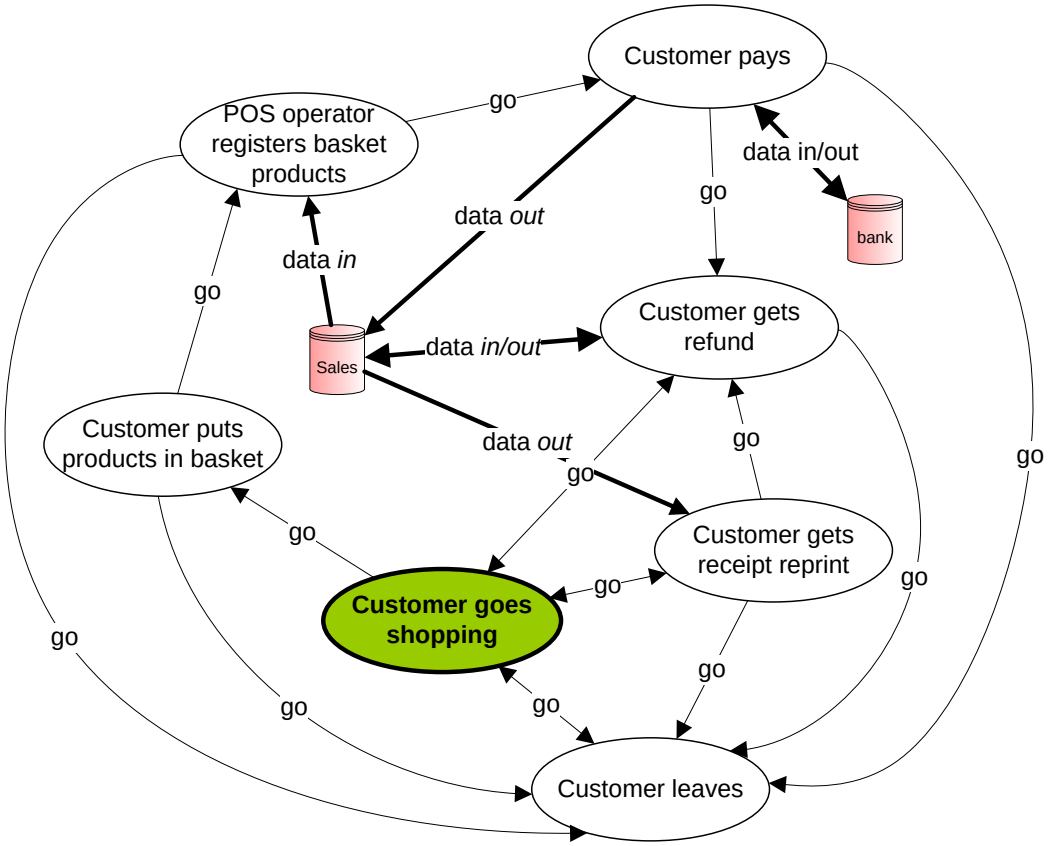
Exercise

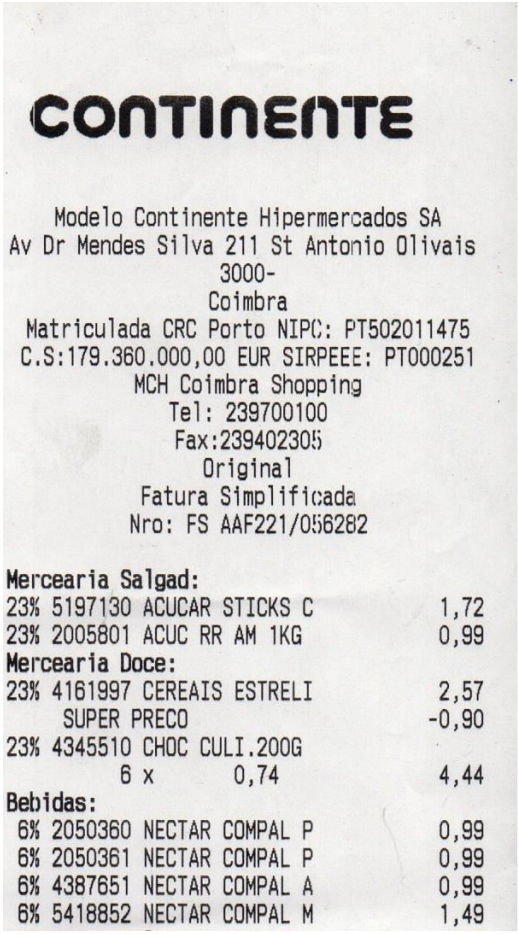
- Analyze given information requirements
- Build conceptual model
- Build logical model (tables)

Example

1st step, Requirements Analysis

REALITY





CONTINENTE

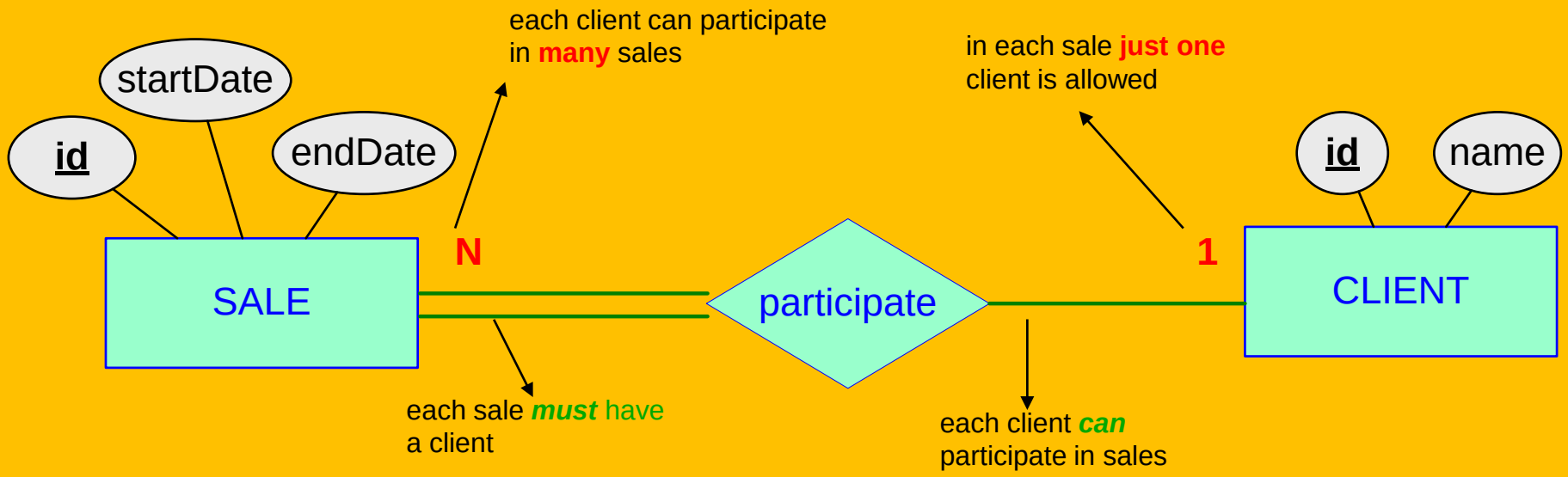
Modelo Continente Hipermercados SA
Av Dr Mendes Silva 211 St Antonio Olivais
3000-
Coimbra
Matriculada CRC Porto NIPC: PT502011475
C.S:179.360.000,00 EUR SIRPEEE: PT000251
MCH Coimbra Shopping
Tel: 239700100
Fax:239402305
Original
Fatura Simplificada
Nro: FS AAF221/056282

Mercearia Salgad:	
23% 5197130 ACUCAR STICKS C	1,72
23% 2005801 ACUC RR AM 1KG	0,99
Mercearia Doce:	
23% 4161997 CEREAIS ESTRELI	2,57
SUPER PRECO	-0,90
23% 4345510 CHOC CULI.200G	
6 x 0,74	4,44
Bebidas:	
6% 2050360 NECTAR COMPAL P	0,99
6% 2050361 NECTAR COMPAL P	0,99
6% 4387651 NECTAR COMPAL A	0,99
6% 5418852 NECTAR COMPAL M	1,49

Example, 1st step: Requirement Analysis

- **Main business process**: *sales*
- **Business rules:**
  - Clients can participate in several sales;
  - Each sale requires one client;
  - Some clients never buy products, but they are clients anyway;
- **Information needs:**
  - Each client's full name;
  - Which clients bought products and when;
  - In how many sales did a client participate;
  - Which clients were never involved in a sale.
  - How much time does a customer waste buying products.

Example, 2nd step: the ERD



entity

relationship

simple attribute

primary key attribute

mandatory participation (***must*** do/have/go/etc.)

optional participation (***can*** do/go/have/etc.)

(**type** of the participation)

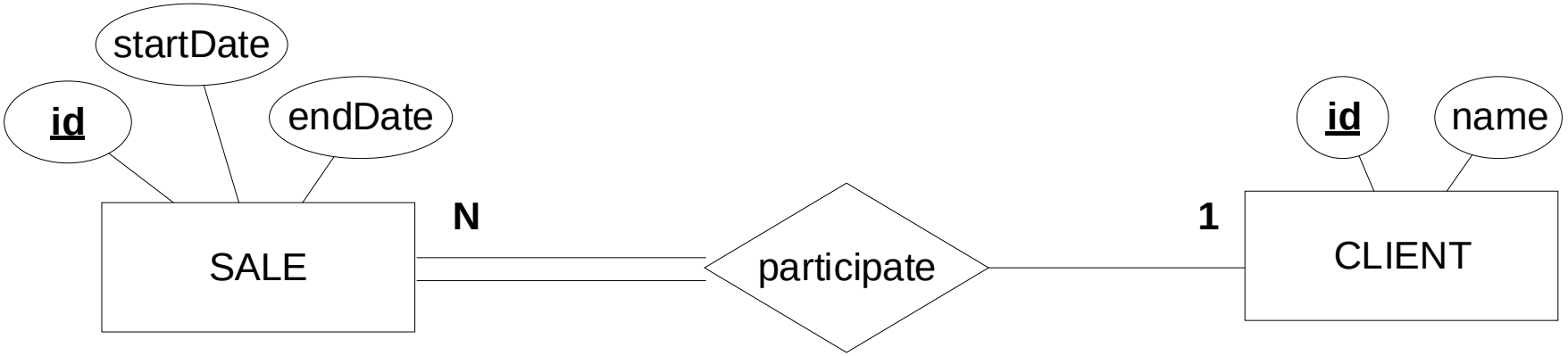
N = many

1 = just one

(**cardinality** of the relationship)

Example, 2nd step: the ERD

The ERD (clean version):



Example, 3rd step: Logical Model

3.1 Transform Conceptual Model --> Logical Model (see *Moodle*)

- Transform relationships **one by one** into logical tables

clients

id	name
1	António Freitas
2	Rita Marujo
3	Carlos da Silva
...	...

sales

id	startDate	endDate	*idClient*
...	...	...	...
625	05-07-2013 10:50	05-07-2013 10:54	1
626	04-08-2013 10:50	04-08-2013 10:50	1
627	07-08-2013 9:20	07-08-2013 9:20	3
...	...	...	...

OR

clients (**id**, name)

sales(**id**, startDate, endDate, *idClient*)

clients

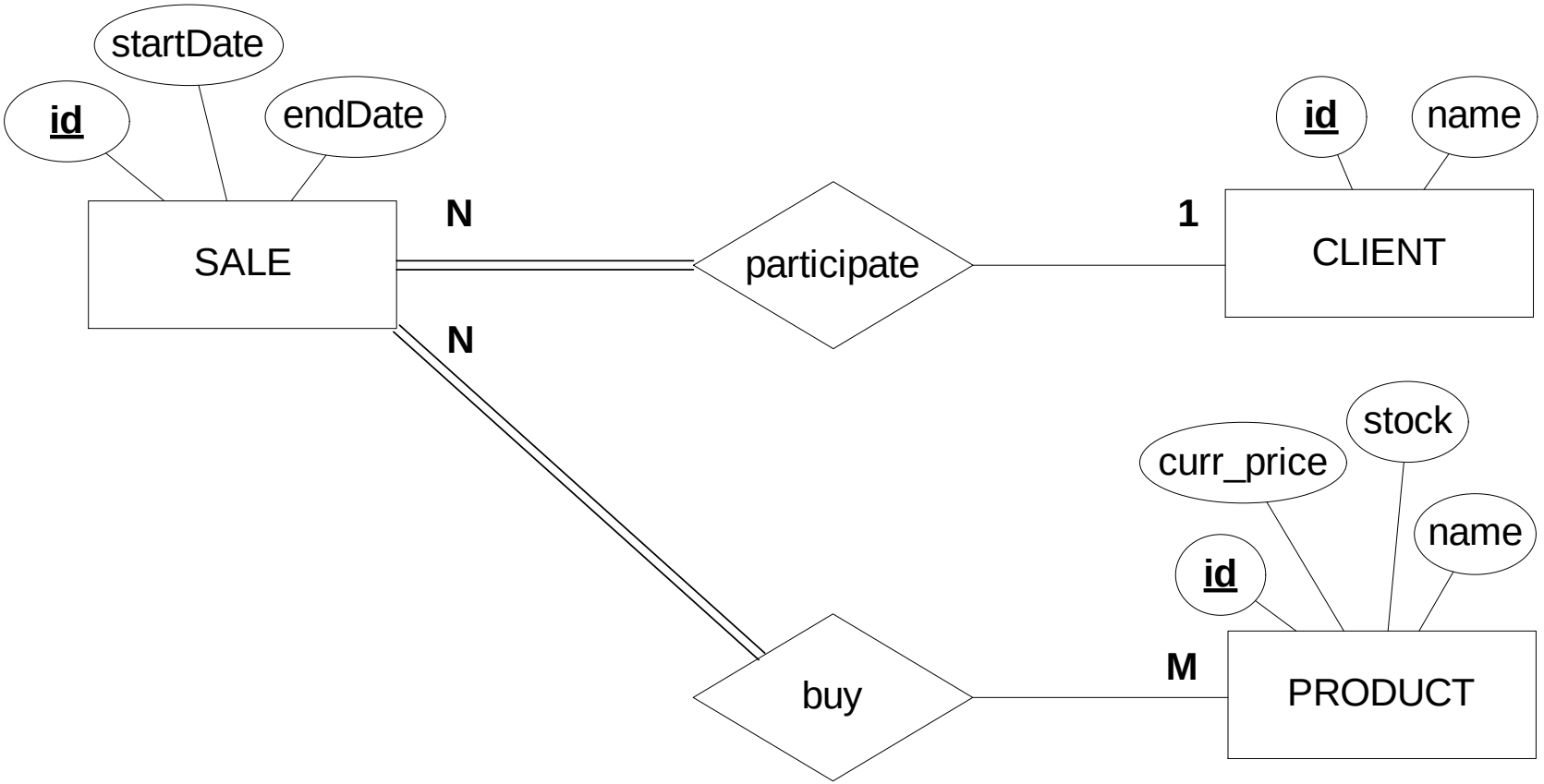
Bold = primary key
Italic = foreign key

Example (*refinement 1*), 1st step

- **(new) Business rules:**
  - A client can buy the same product several times;
  - A product can be sold many times, until the stock runs out;
- **(new) Information needs:**
  - What is the name and the current price of the products that each client bought;

Example (*refinement 1*), 2nd step

The ERD



Example (*refinement 1*), 3rd step

Logical Model

clients (**id**, name)

products (**id**, name, curr_price, stock)

product_sales (***product_id***, ***sale_id***)

sales(**id**, startDate, endDate, *idClient*)

OR

products

id	name	curr_price	stock
...	...	...	...
400	Frozen shrimp 60/80	12.99	450
401	Harry Potter (4)	14.99	100
402	Christmas socks - XXXL	3.49	200
...	...	...	...

clients

id	name
1	António Freitas
2	Rita Marujo
3	Carlos da Silva
...	...

product_sales

product_id	***sale_id***
...	...
400	625
401	625
401	626
402	626
...	...

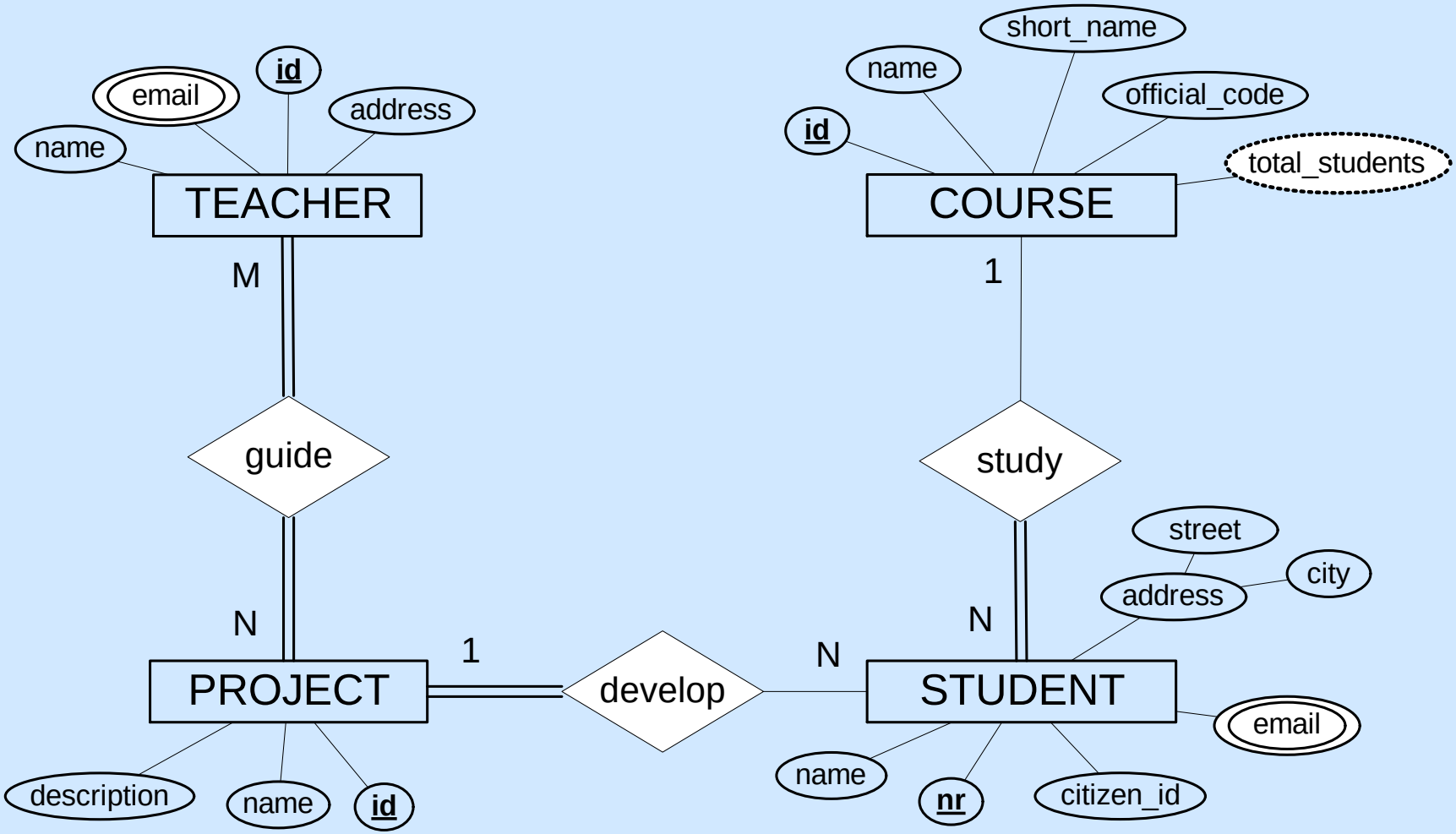
sales

id	startDate	endDate	*idClient*
...	...	...	...
625	05-07-2013 10:50	05-07-2013 10:54	1
626	04-08-2013 10:50	04-08-2013 10:50	1
627	07-08-2013 9:20	07-08-2013 9:20	3
...	...	...	...

Bold = primary key
Italic = foreign key

Quiz

Consider the following ER diagram:



Quiz

A. How many tables will the logical model have?

B. Which statements are true considering the ERD? (choose all that apply)

a) Each student has several email addresses.

b) Each student has, at least, one project.

c) Each student will allways have one course

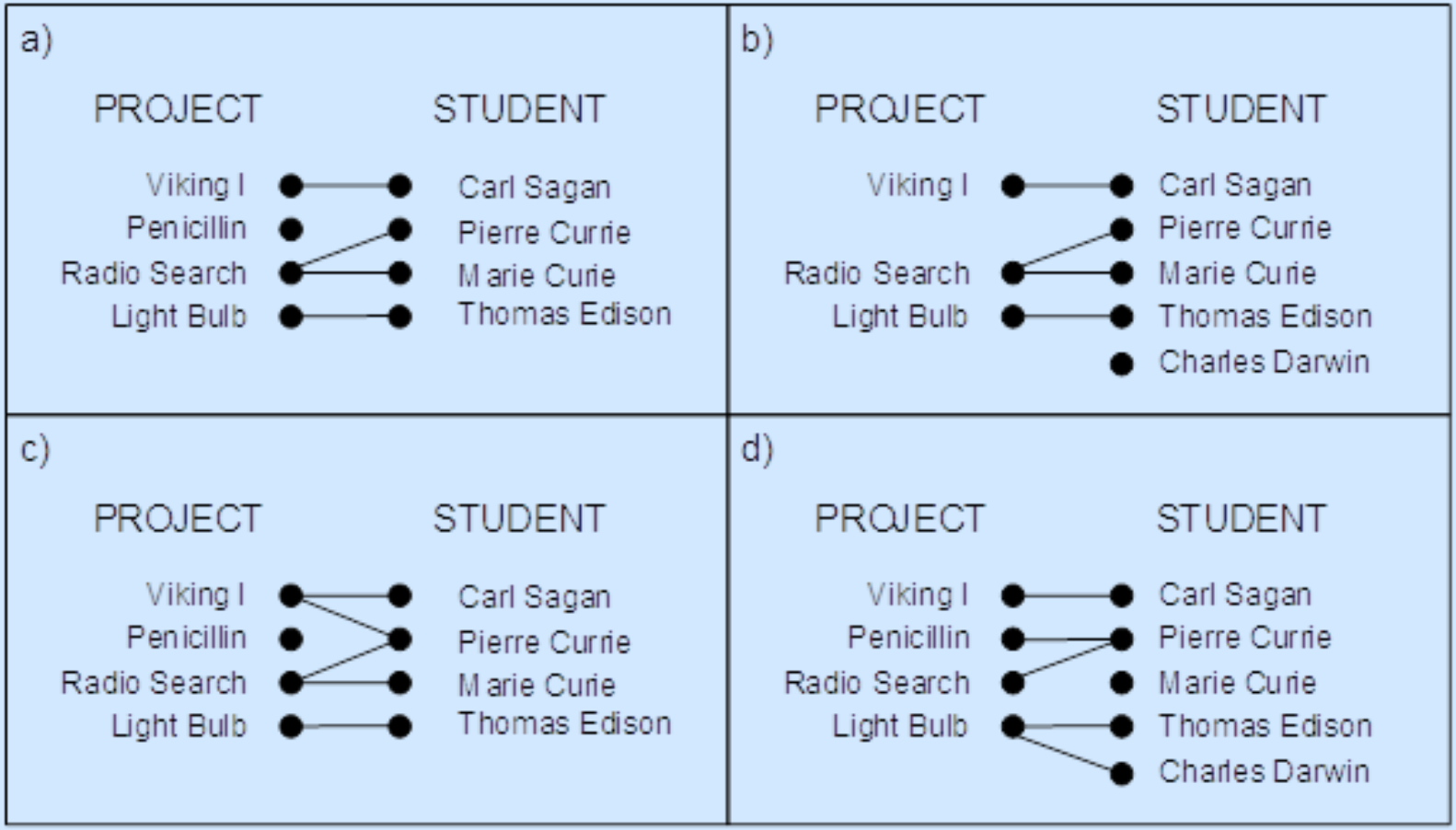
d) Each teacher can guide several projects.

e) Teachers and students can have the same email.

f) Each teacher may guide projects of diferent courses.

Quiz

C. Which of the following diagrams represents the *develop* relationship?

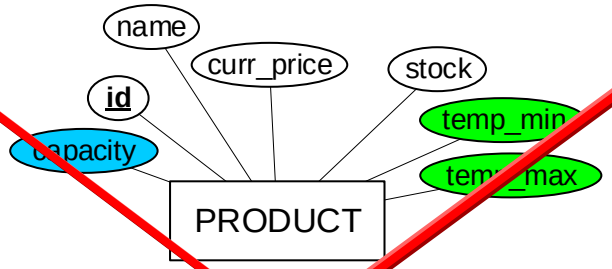


Example (*refinement 2*), 1st step

- **(new) Business rules:**
  - Some products have specific storage conditions, like juices, frozen products, fresh vegetables, fresh meat, etc.
- **(new) Information needs:**
  - What is the minimum and maximum temperature of frozen products?
  - What is the storage capacity (in liters) of each juice package?

Example (*refinement 2*), 2nd and 3rd steps

The ERD (?)

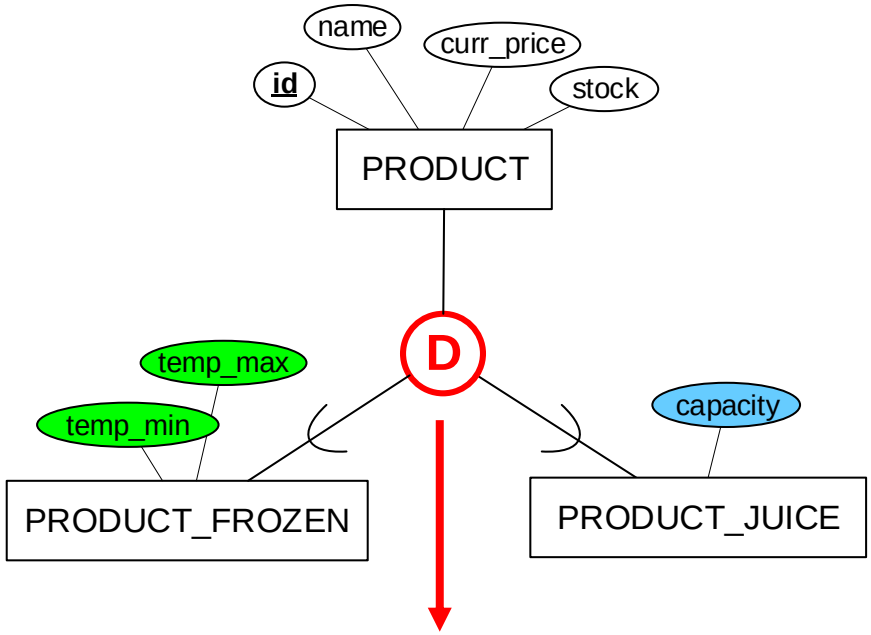


products

id	name	curr_price	stock	capacity	temp_min	temp_max
...	...	...	...	...	...	...
400	Frozen shrimp 60/80	12.99	450		-18	-12
401	Harry Potter (4)	14.99	100			
402	Christmas socks - XXXL	3.49	200			
403	Plum Nectar	1.49	250	1		
404	Beer 7%	5.00	1000	0.33		
405	Frozen shrimp 40/60	10.49	1000		-18	-12
...	...	...	...	...	...	...

Example (*refinement 2*), 2nd and 3rd steps

ERD Solution = Hierarchy



Disjoint =
Each product can be a frozen product or a juice, but not both

products

id	name	curr_price	stock
...	...	...	...
400	Frozen shrimp 60/80	12.99	450
401	Harry Potter (4)	14.99	100
402	Christmas socks - XXXL	3.49	200
403	Plum Nectar	1.49	250
404	Beer 7%	5.00	1000
405	Frozen shrimp 40/60	10.49	1000
...	...	...	...

products_juices

product_id	capacity
403	1
404	0.33
...	...

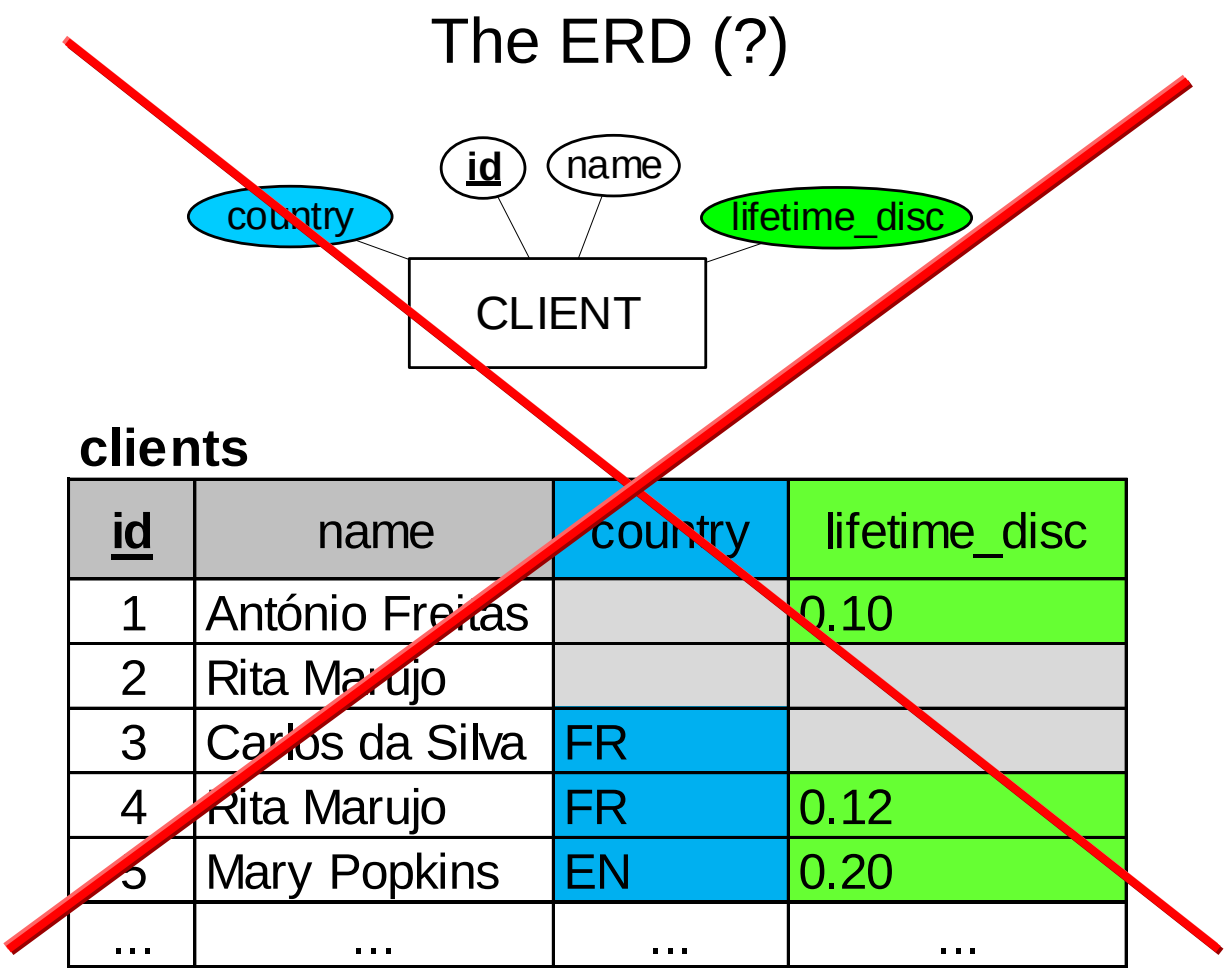
products_frozen

product_id	temp_min	temp_max
400	-18	-12
405	-18	-12
...	...	...

Example (*refinement 3*), 1st step

- **(new) Business rules:**
  - Foreigner clients will be asked for their residence country name and country-specific documentation at the time of the first payment.
  - Clients (domestic and foreigner) can be considered premium and, in such cases they will have a lifetime discount in every sale.
- **(new) Information needs:**
  - Which clients live in another country?
  - How many domestic and foreign clients are premium?
  - How many domestic and foreign clients ***are not*** premium?
  - What is the ratio of clients per country?
  - What is the average lifetime discount of premium clients?
  - What is the average lifetime discount of premium clients per country?

Example (*refinement 3*), 2nd and 3rd steps

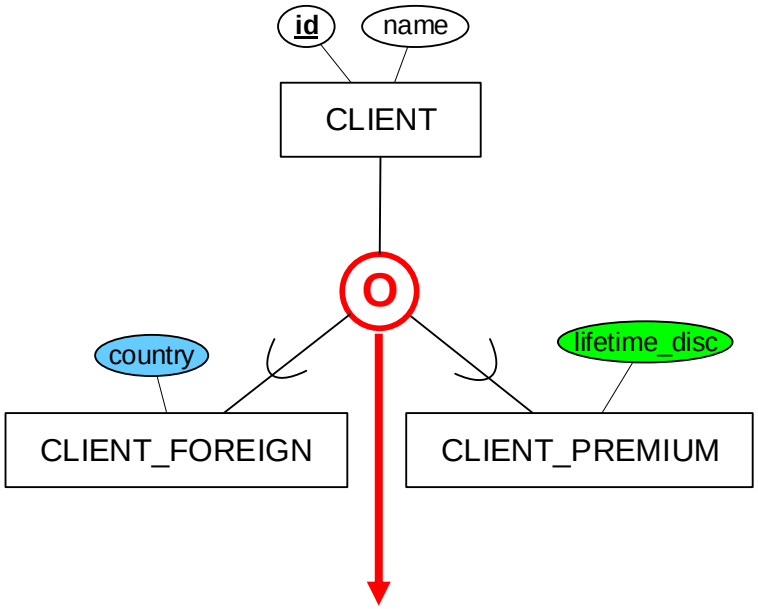


clients

id	name	country	lifetime_disc
1	António Freitas		0.10
2	Rita Marujo		
3	Carlos da Silva	FR	
4	Rita Marujo	FR	0.12
5	Mary Popkins	EN	0.20
...	...	...	...

Example (*refinement 3*), 2nd and 3rd steps

ERD Solution = Hierarchy



Overlap =
Each client can be a *foreign client*, a *premium client* or *both*

clients

id	name
1	António Freitas
2	Rita Marujo
3	Carlos da Silva
4	Rita Marujo
5	Mary Popkins
...	...

foreign_clients

client_id	country
3	FR
4	FR
5	EN
...	...

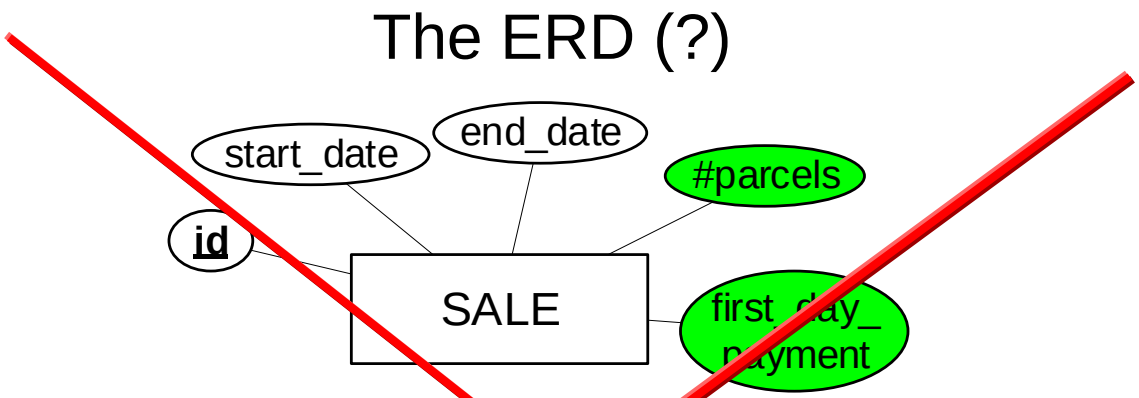
premium_clients

client_id	lifetime_disc
1	0.10
4	0.12
5	0.20
...	...

Example (*refinement 4*), 1st step

- **(new) Business rules:**
  - The company allows sales to be payed in two, three or six parcels.
  - In such authorized cases, the client will choose the number of parcels and the day in which the first parcel will be charged from its account.
- **(new) Information needs:**
  - Which clients regularly use the parcel-option in their payments?
  - What is the most chosen type of payment? (full or in two/three/six parcels)
  - When will each client pay all its debts?

Example (*refinement 4*), 2nd and 3rd steps

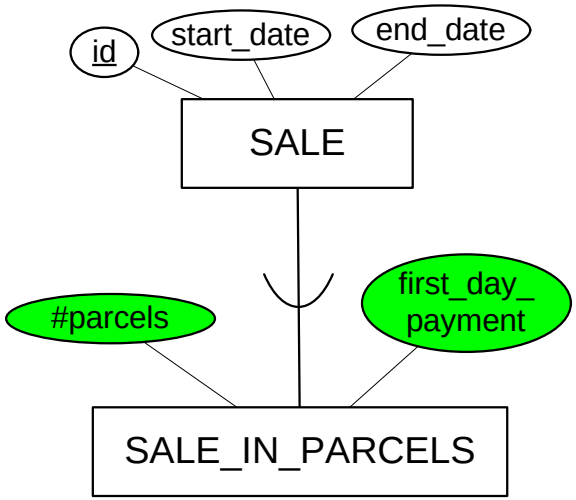


sales

id	startDate	endDate	#parcels	first_day_payment
…	…	…	…	…
625	05-07-2013 10:50	05-07-2013 10:54	6	05-10-2013
626	04-08-2013 10:50	04-08-2013 10:50		
627	07-08-2013 9:20	07-08-2013 9:20		
…	…	…	…	…

Example (*refinement 4*), 2nd and 3rd steps

ERD Solution = Hierarchy



sales

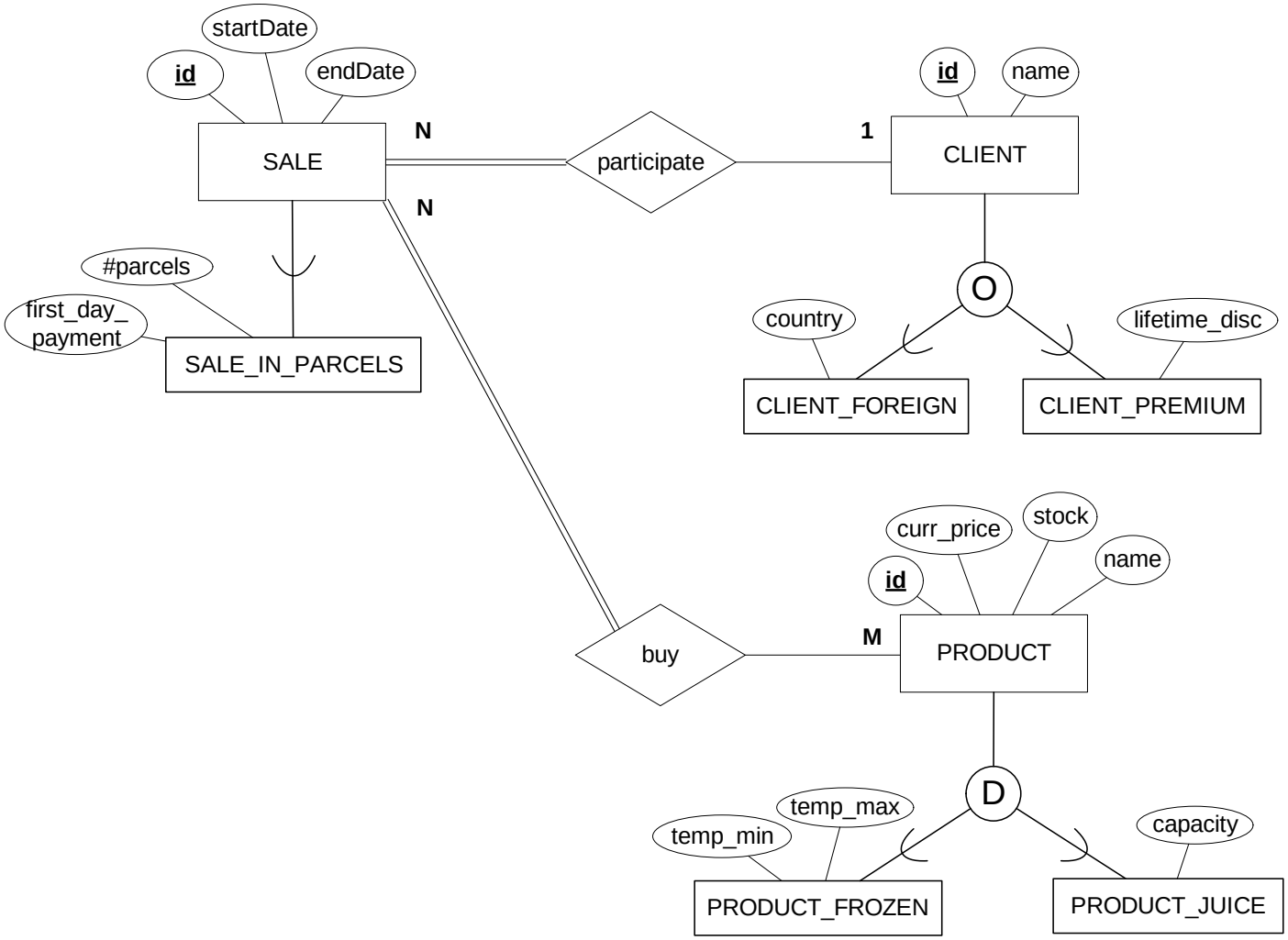
id	startDate	endDate
...	...	...
625	05-07-2013 10:50	05-07-2013 10:54
626	04-08-2013 10:50	04-08-2013 10:50
627	07-08-2013 9:20	07-08-2013 9:20
...	...	...

sales_in_parcels

sale_id	#parcels	first_day_payment
...	...	...
625	6	05-10-2013
...	...	...

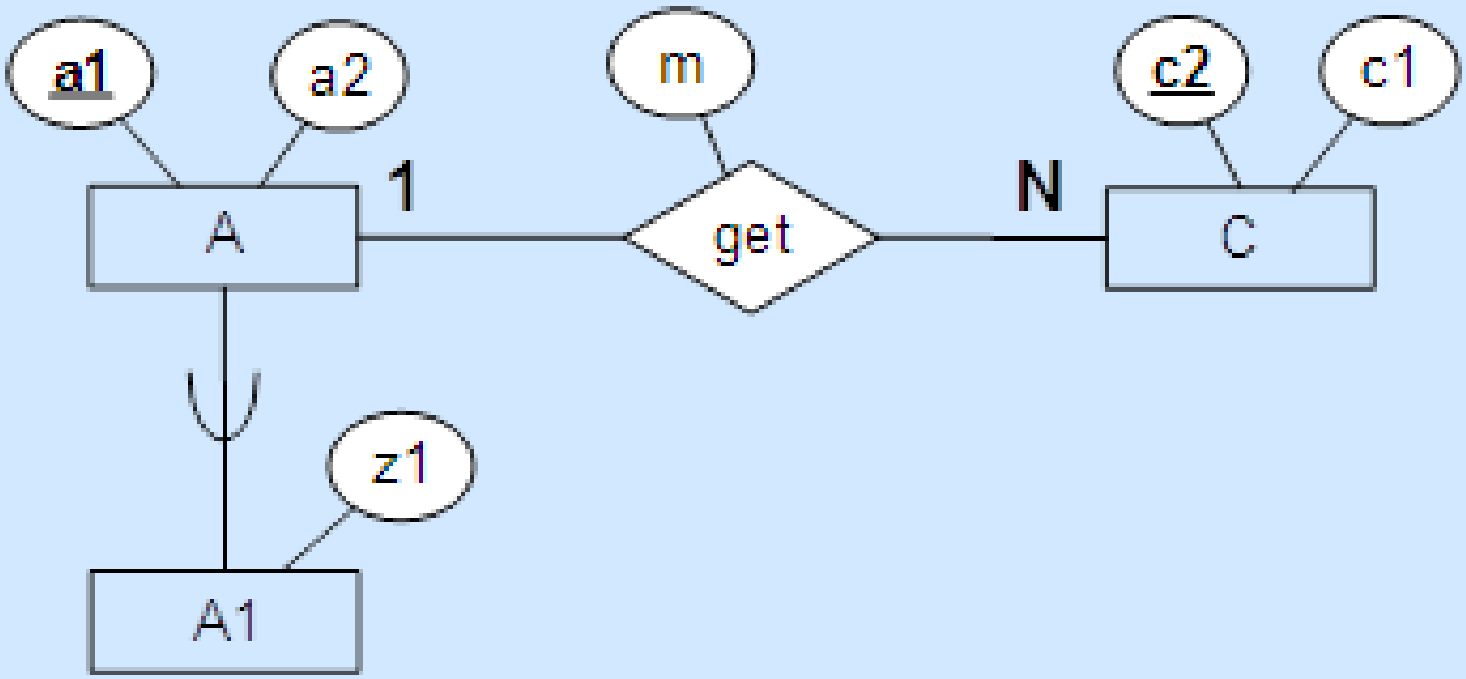
Rui Oliveira © rui.oliveira@ipleiria.pt

ERD (*after refinement 4*)



Quiz

Consider the following ER diagram:



Quiz

A. Which of the following options shows the corresponding logical model (choose one):

a) A (**a1**, a2) A1 (***a1***, z1) → A C (**c2**, c1) A_C (*a1*, ***c2***, m) → A, C	b) A (**a1**, a2, z1) C (**c2**, c1) A_C (*a1*, ***c2***, m) → A, C
c) *A (**a1**, a2)* *A1 (**a1**, z1)* → A *C (**c2**, c1, m, a1)* → A	d) None of the previous.

Quiz

B. Choose the answers which are correct from reading the ERD (choose all that apply):

a) Entity A represents a subset of entity A1.

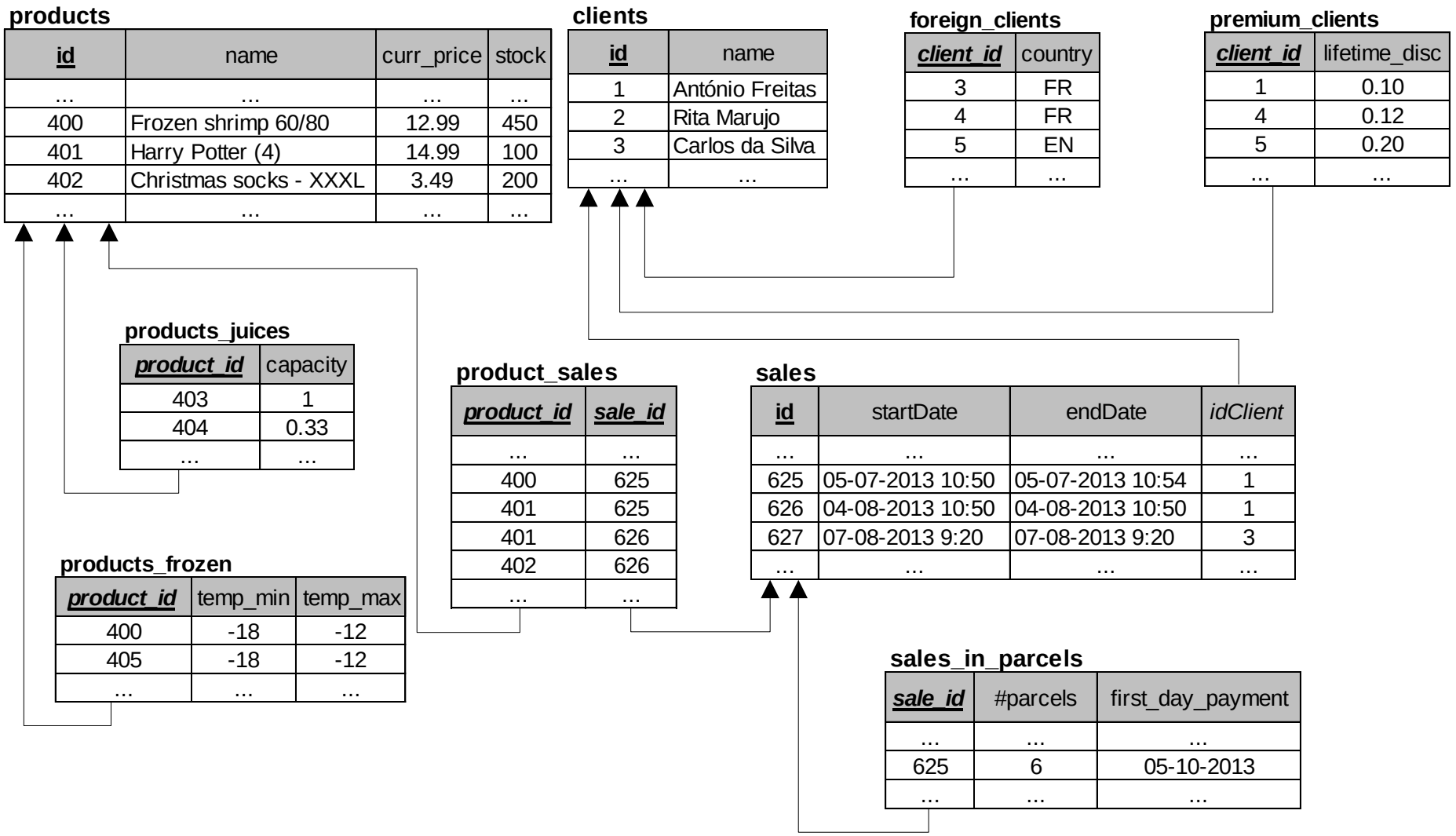
b) The elements of A1 can be the same elements of A.

c) Each element of C can relate to several elements of A.

d) Each element of A1 can relate to several elements of A.

e) For each element of C there is at least one element of A.

Logical Model (*after refinement 4*)



products

id	name	curr_price	stock
...	...	...	...
400	Frozen shrimp 60/80	12.99	450
401	Harry Potter (4)	14.99	100
402	Christmas socks - XXXL	3.49	200
...	...	...	...

clients

id	name
1	António Freitas
2	Rita Marujo
3	Carlos da Silva
...	...

foreign_clients

client_id	country
3	FR
4	FR
5	EN
...	...

premium_clients

client_id	lifetime_disc
1	0.10
4	0.12
5	0.20
...	...

products_juices

product_id	capacity
403	1
404	0.33
...	...

product_sales

product_id	sale_id
...	...
400	625
401	625
401	626
402	626
...	...

sales

id	startDate	endDate	idClient
...	...	...	...
625	05-07-2013 10:50	05-07-2013 10:54	1
626	04-08-2013 10:50	04-08-2013 10:50	1
627	07-08-2013 9:20	07-08-2013 9:20	3
...	...	...	...

products_frozen

product_id	temp_min	temp_max
400	-18	-12
405	-18	-12
...	...	...

sales_in_parcels

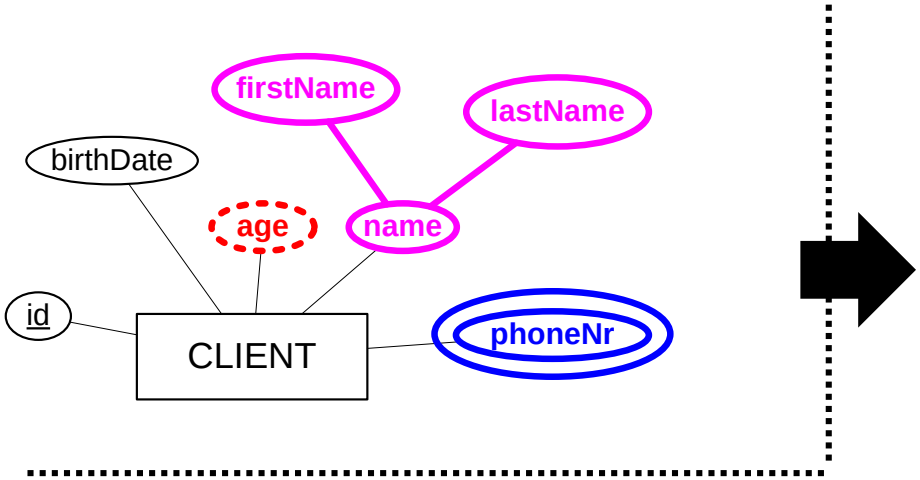
sale_id	#parcels	first_day_payment
...	...	...
625	6	05-10-2013
...	...	...

Example (*refinement 5*), 1st step

- **(new) Business rules:**
  - Each client can have multiple telephone numbers
  - Clients can share telephone numbers
- **(new) Information needs:**
  - The first and last name of each client
  - The telephone number(s) of each client
  - The age of each client

Example (*refinement 5*), 2nd and 3rd steps

ERD Solution = new types of attributes



clients

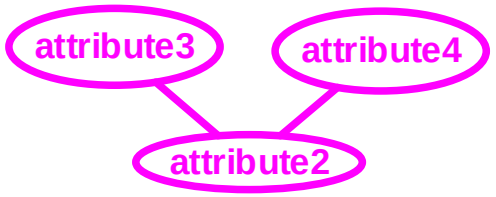
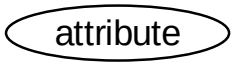
id	birthDate	firstName	lastName
1		Anónimo	
100298847	21-10-1998	Rui	Oliveira
100298848	02-09-1987	Rita	Marujo
100298849	31-01-1975	João	Andrade
...	...	...	...

phoneNumbers

idClient	phoneNr
100298847	244000001
100298848	244000022
100298849	244000001
100298849	960004444
...	...

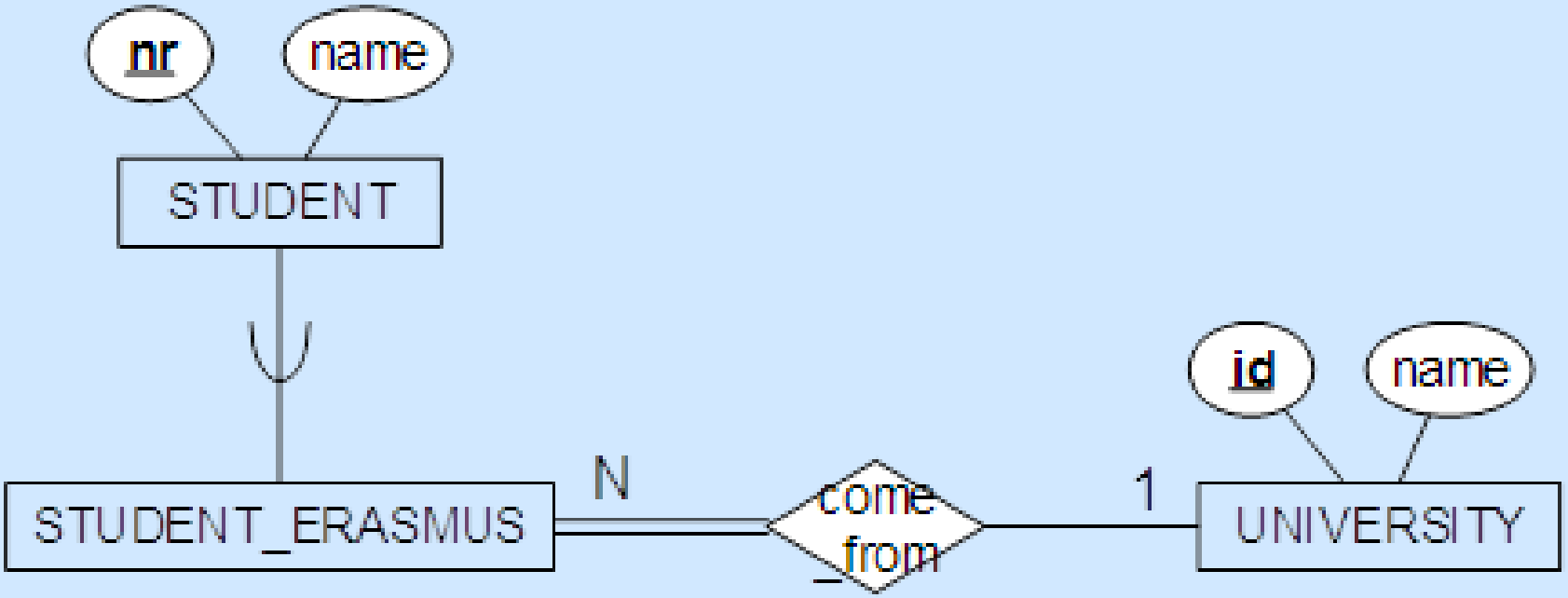
```
SELECT c.id, c.lastName||', '|| c.firstName, AS name,
       p.phoneNr AS "telefone_numbers",
       MONTHS_BETWEEN(SYSDATE,c.birthDate)/12) AS age,
FROM clientes c
     LEFT JOIN phoneNumbers p ON c.id= p.idClient
ORDER BY 2 ASC;
```

Types of attributes

ERD representation	Type of attribute
	Derived (in PT, *calculado/derivado*)
	Composite (in PT, *composto*)
	Simple (in PT, *simples*)
	Multivalued (in PT, *multivalor*)

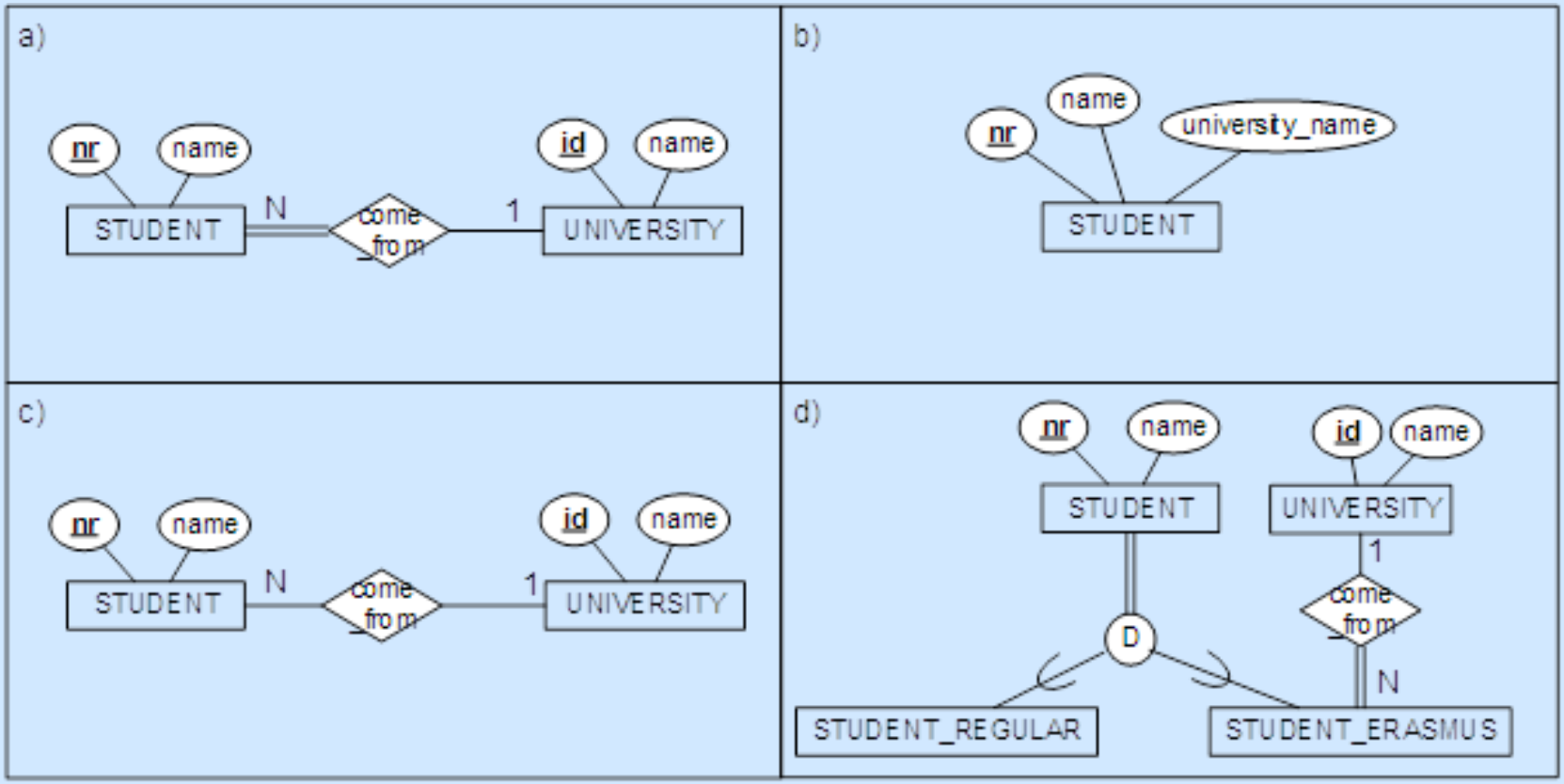
Quiz

Consider the following ER diagram:



Quiz

A. Which of the following diagrams is the most similar to the previous ERD?



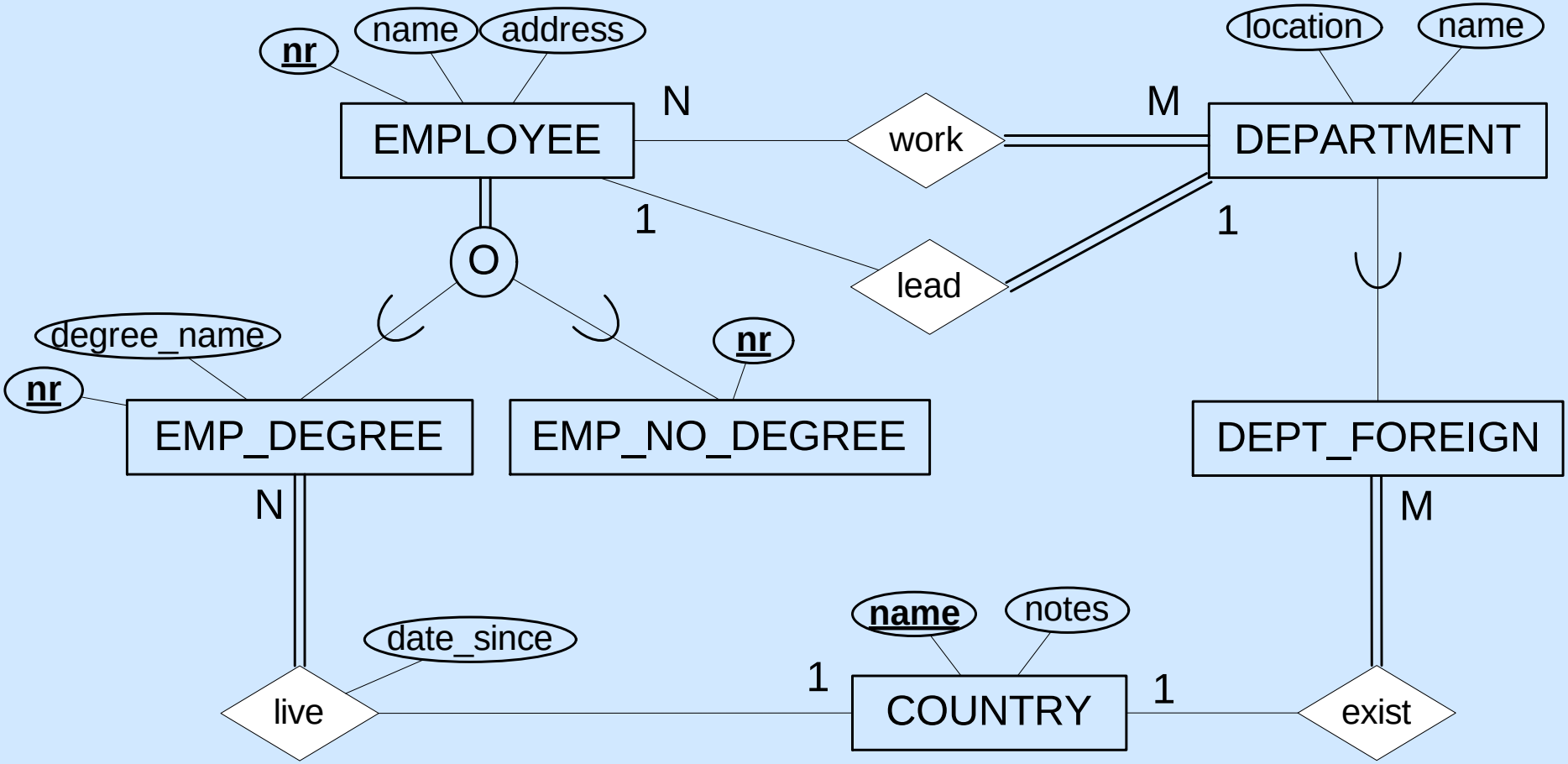
Quiz

B. How/where would you represent the total number of students coming from each university?

C. How/where would you represent the total number of Erasmus students?

Quiz

Check the following ERD for representation errors.



The end
(for now).