

MORE ABOUT DATA MART CONCEPTUAL MODELLING



UNIVERSITÀ DI PISA

Degenerate dimensions

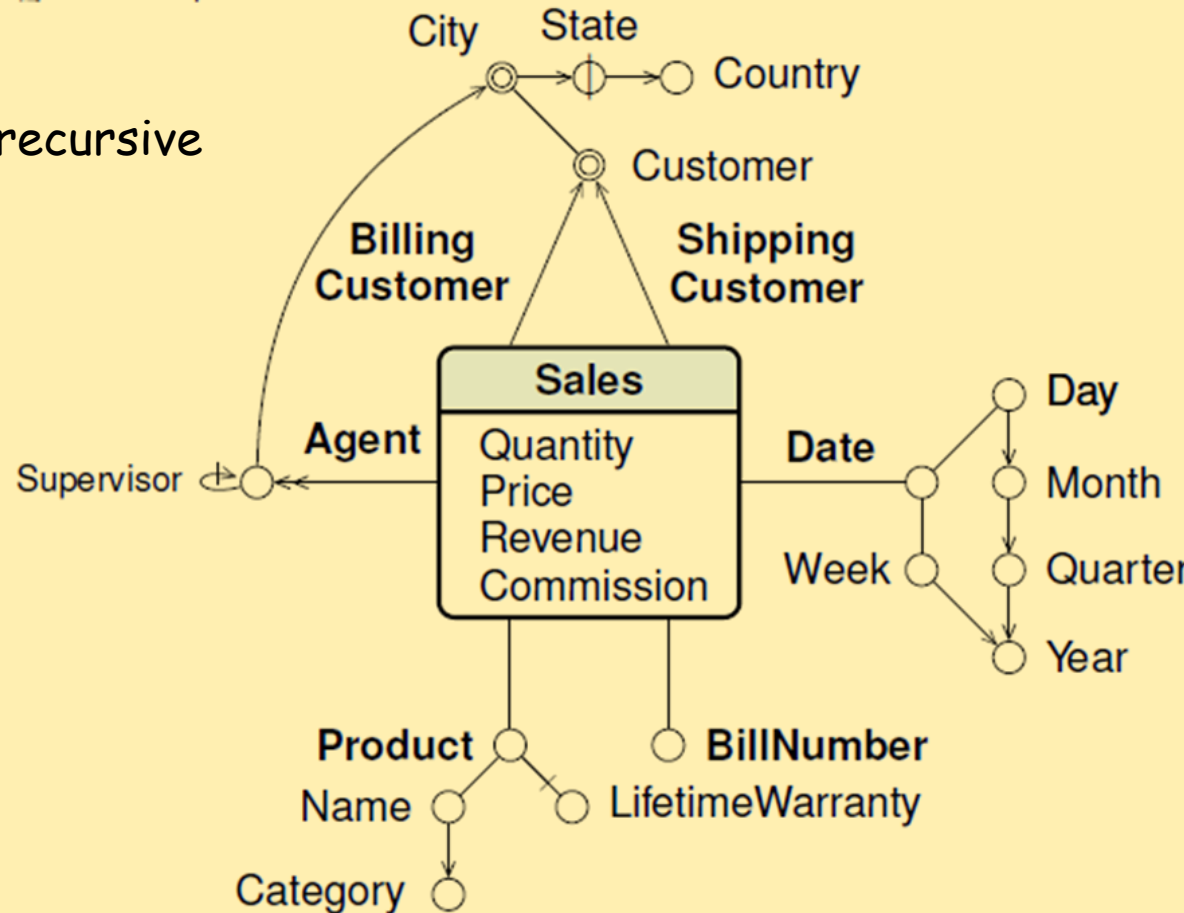
Optional dimensions or attributes

Hierarchy's types: balanced, ragged, recursive

Multivalued dimensions

Shared hierarchies and dimensions

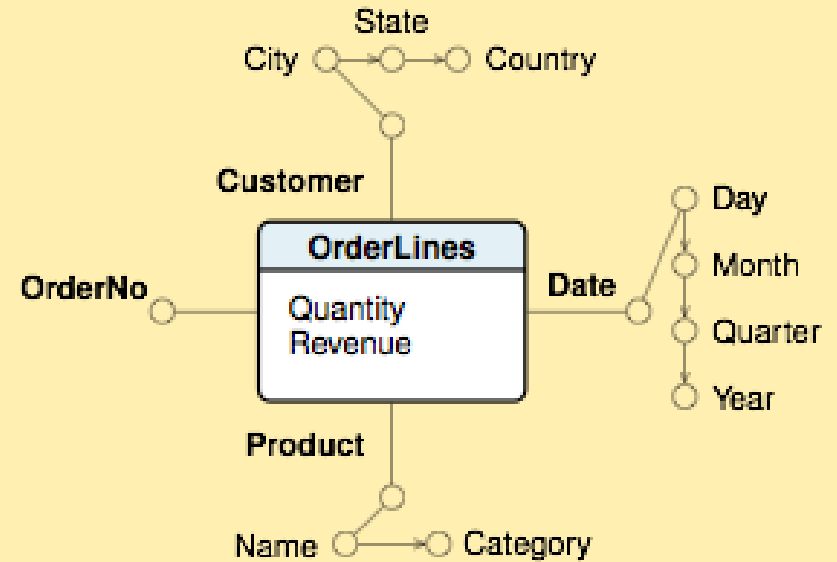
Descriptive attributes



Changing dimensions

Slowly changing dimensions

- TYPE 1 (overwriting the history)
- TYPE 2 (preserving all the history) .
- TYPE 3 (preserving one or more versions of history)
Not recommended



Fast changing dimensions

- TYPE 4

These aspects are not modelled in the conceptual schema

(will be considered at logical level) but they are reported in the documentation

Relational OLAP systems are relational DBMS extended with specific features to support business intelligence analysis.

A DW is represented with a special kind of **relational schema**

A **star schema**,

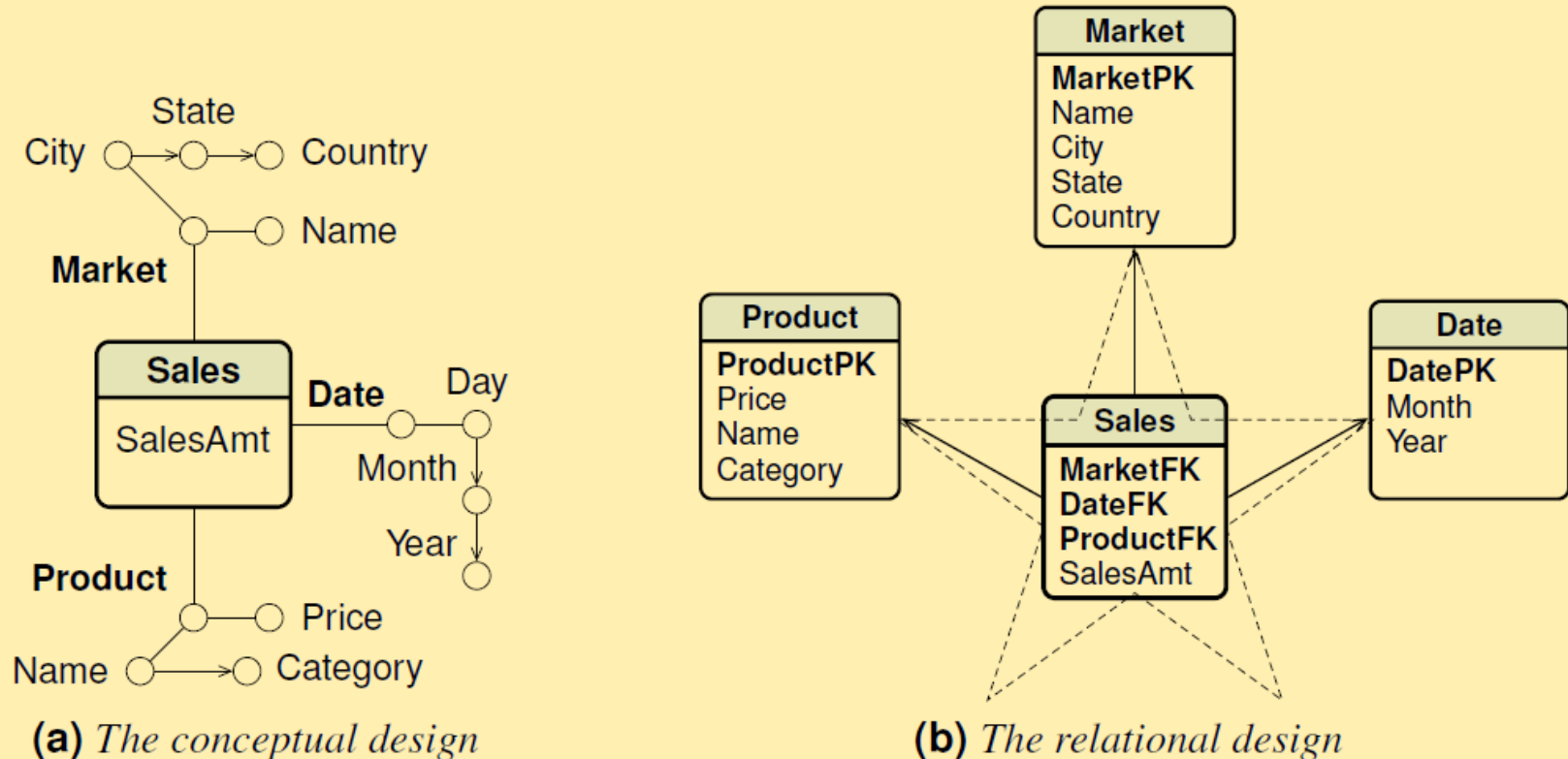
A **snowflake schema** or

A **constellation schema**.

A STAR SCHEMA EXAMPLE

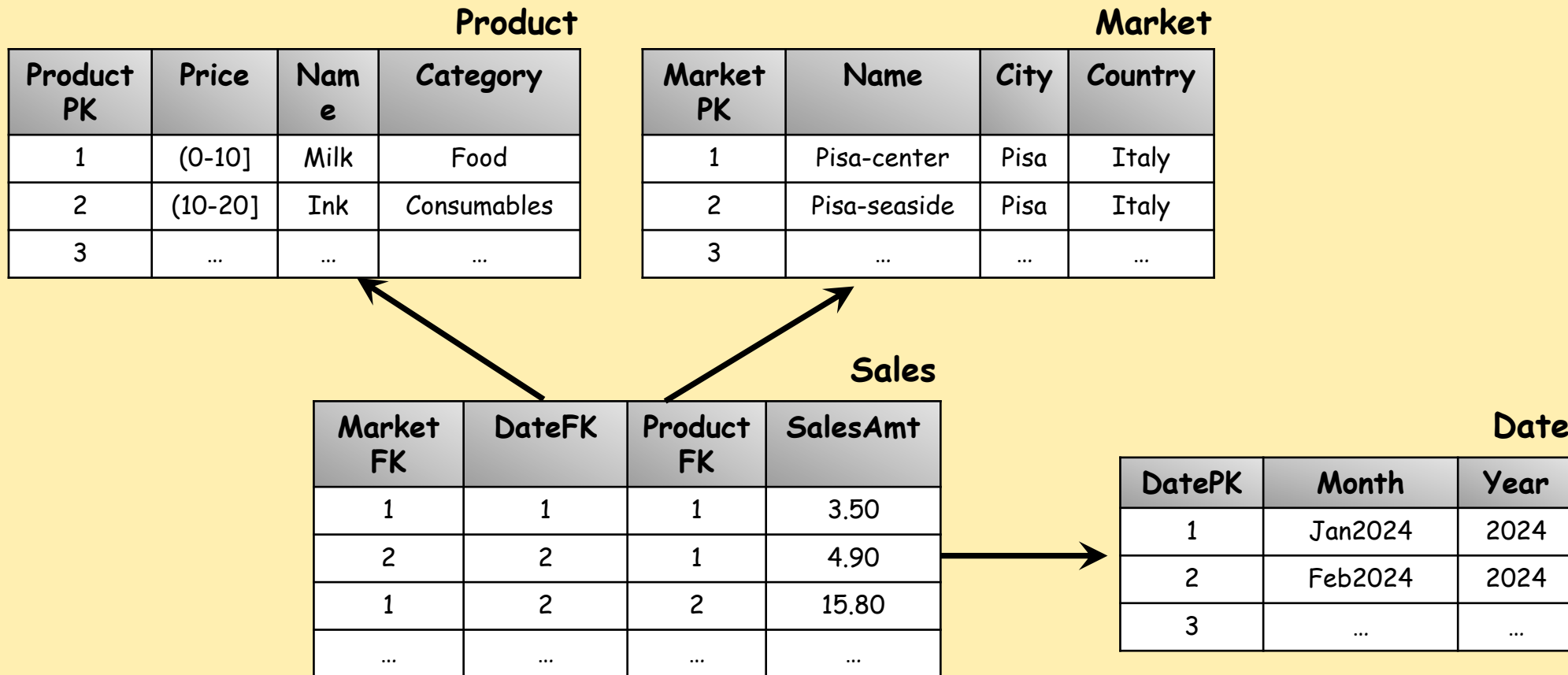


N.B. In the relational schema, the hierarchy info is lost!

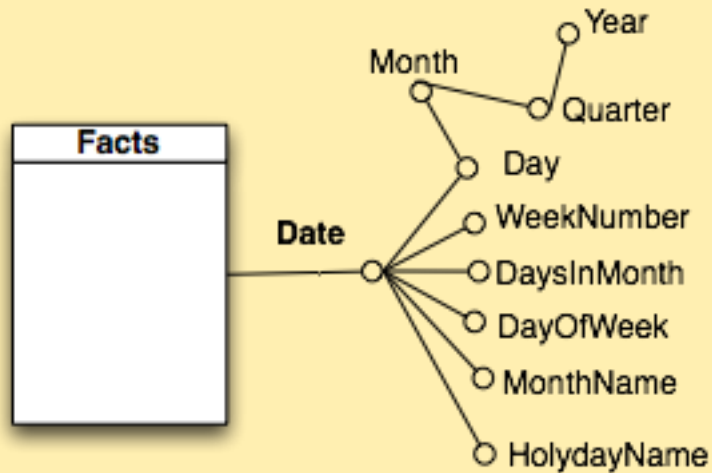


Facts stored as rows in a fact table. Measures are columns of the fact table. Dimensions stored in dimension table. Attributes are columns of the dimension table. A dimension table always uses a system-generated primary key, called a Surrogate Key, to support attribute updates (see later).

A STAR SCHEMA EXAMPLE



Hyp: Date at daily grain

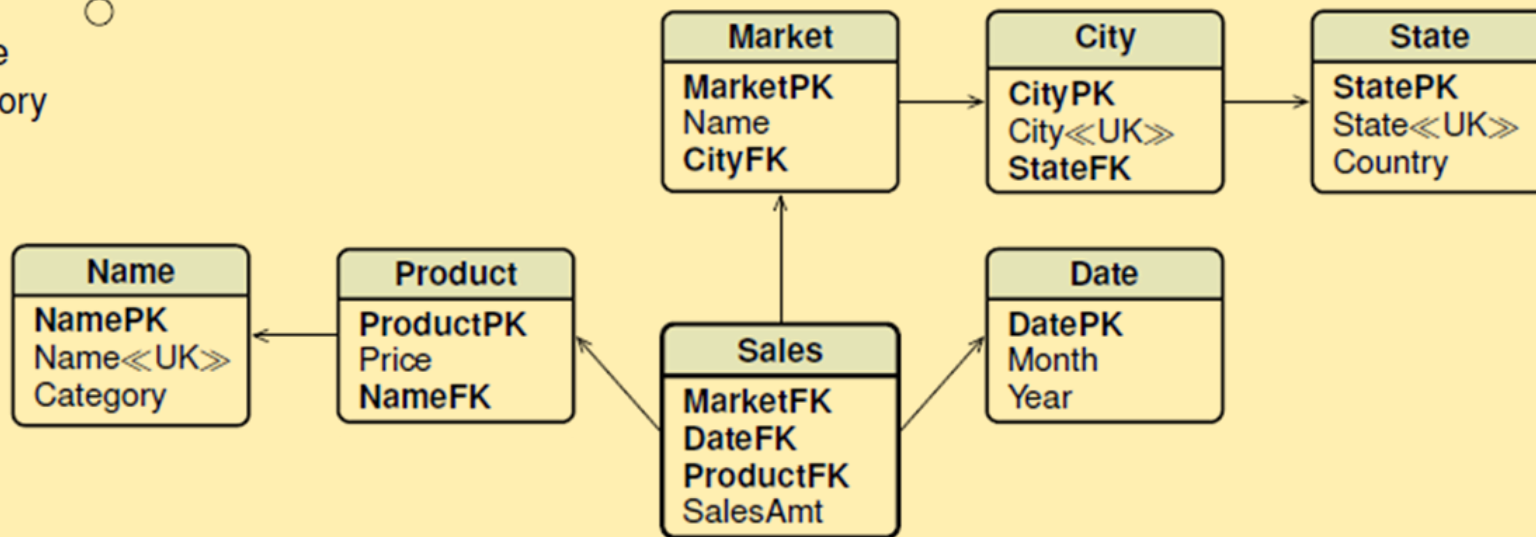
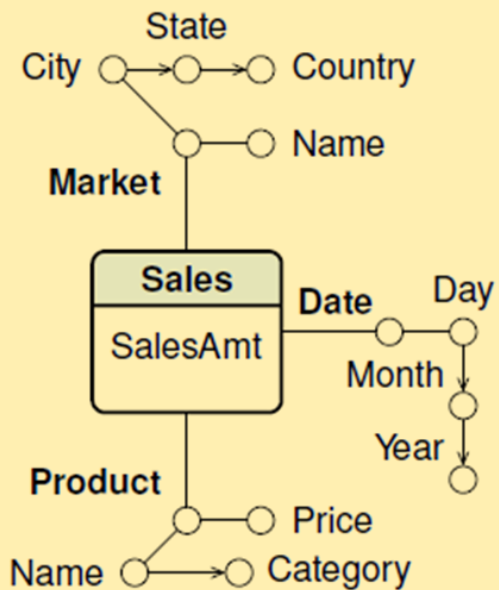


In the logical schema,
the dimension **Date** has the surrogate key
with the integer value
YYYYMMDD

DATE

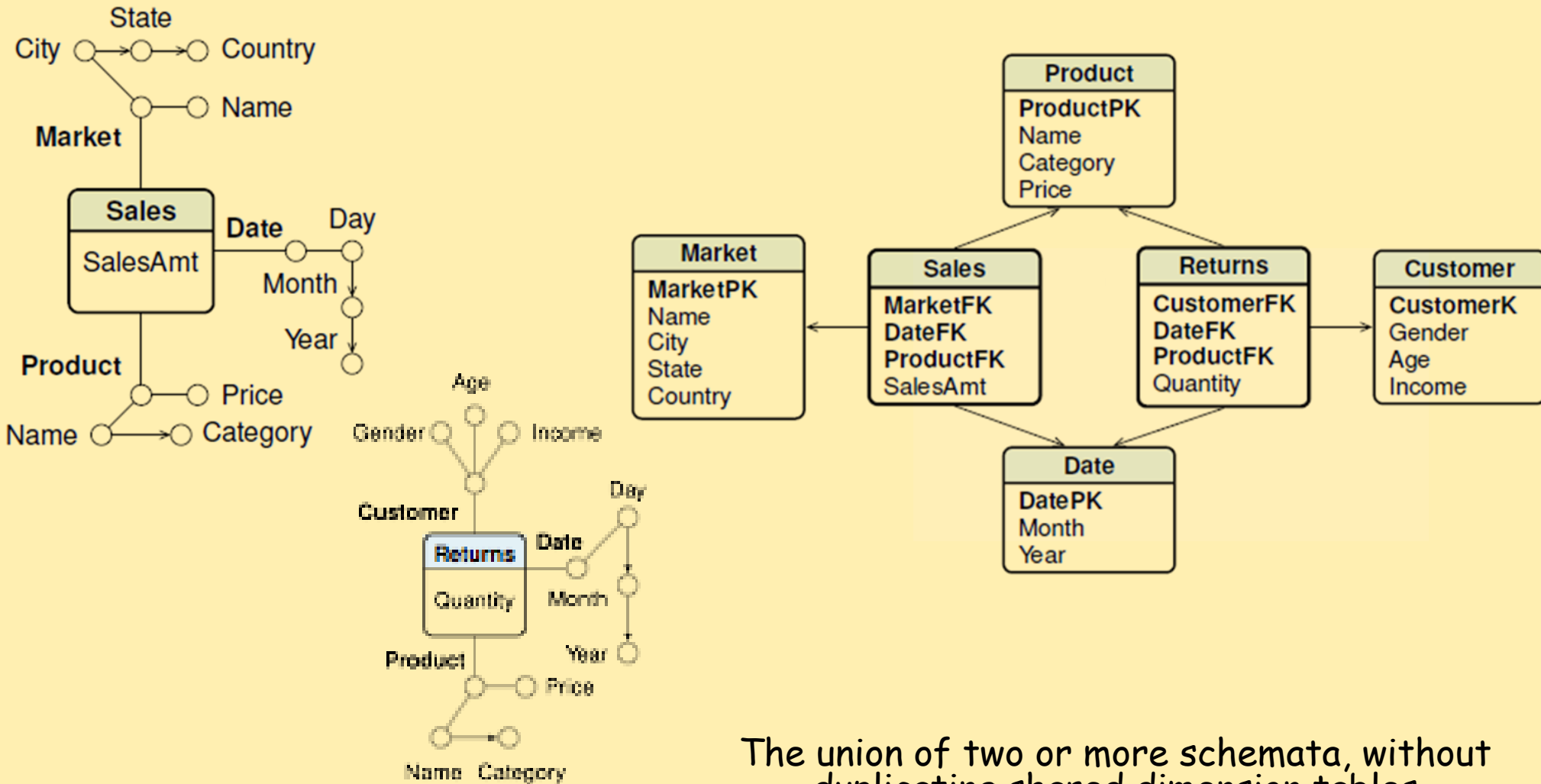
Attribute Name	Type	Format/Example
DatePK	int	YYYYMMDD
Month	int	YYYYMM
Quarter	int	YYYYQ
Year	int	YYYY
WeekNumber	int	1 to 52 or 53
DayInMonth	int	1 to 31
DayOfWeek	string	Monday
MonthName	string	January
HolydayName	string	Easter

SNOWFLAKE SCHEMA



Dimension tables split into two or more.

CONSTELLATION SCHEMA

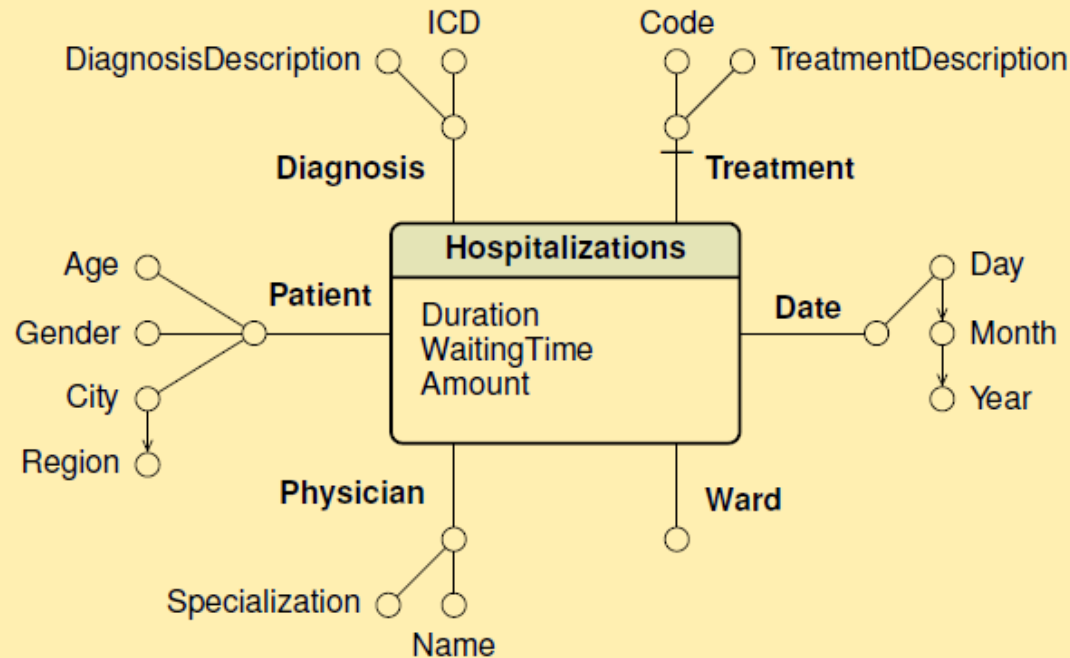


The union of two or more schemata, without duplicating shared dimension tables (conformed dimensions).

HOSPITALIZATIONS DATA MART CONCEPTUAL SCHEMA



UNIVERSITÀ DI PISA

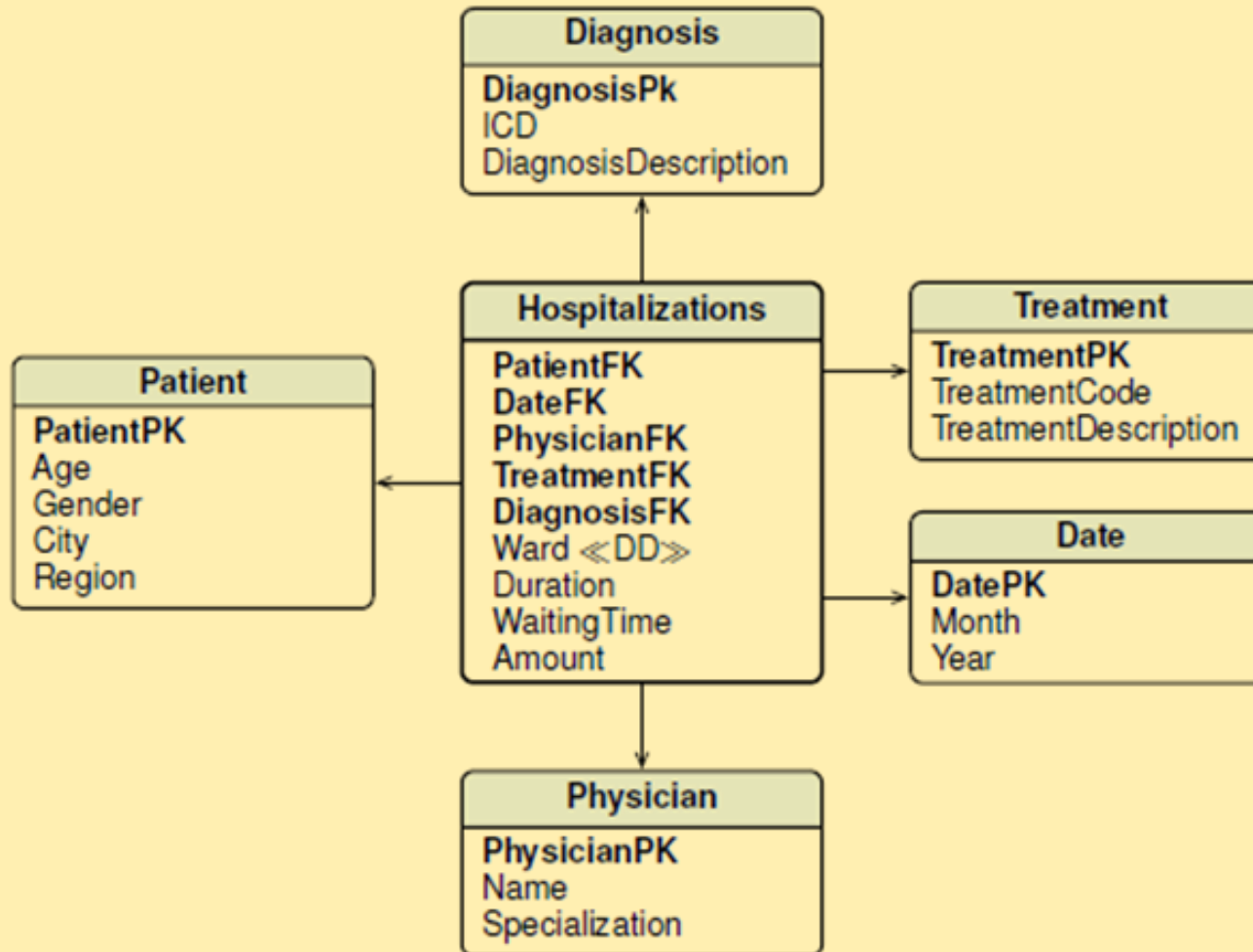


DESIGN THE LOGICAL SCHEMA

HOSPITALIZATIONS: INITIAL LOGICAL SCHEMA



UNIVERSITÀ DI PISA



MISSING VALUES IN FACT TABLE



How to code facts where a dimension value, eg. Treatment, is missing?

Treatment

TreatmentPK	TreatmentCode
1	T1
2	T2

Hospitalizations

...	TreatmentFK	...
...	1	...
...	1	...
...	2	...
...	NULL	...

Number of hospitalizations by treatment

```
SELECT TreatmentCode, COUNT(*) AS N
FROM Sales, Treatment
WHERE TreatmentFK = TreatmentPK
GROUP BY TreatmentCode
```

TreatmentCode	N
T1	2
T2	1

Number of hospitalizations

N
4

MISSING VALUES IN FACT TABLE



How to code facts where a dimension value, eg. Treatment, is missing?

Treatment

TreatmentPK	TreatmentCode
0	No treatment
1	T1
2	T2

Number of hospitalizations by treatment

```
SELECT TreatmentCode, COUNT(*) AS N
FROM Sales, Treatment
WHERE TreatmentFK = TreatmentPK
GROUP BY TreatmentCode
```

Hospitalizations

...	TreatmentFK	...
...	1	...
...	1	...
...	2	...
...	0	...

TreatmentCode	N
No treatment	1
T1	2
T2	1

Number of hospitalizations

N
4

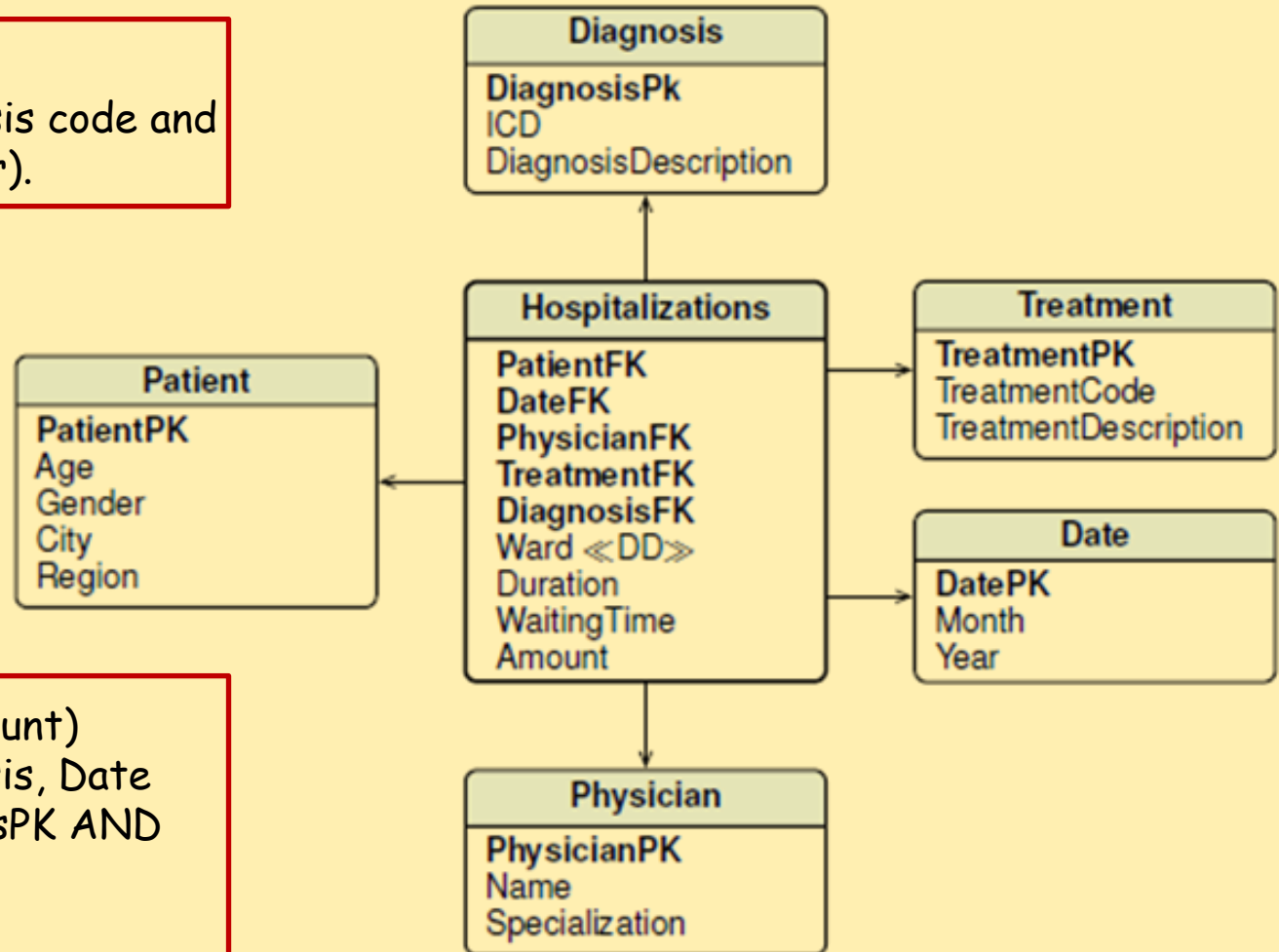
An hospital needs a DM to extract information from their operational database with information about inpatients treatments.

1. Total billed amount for hospitalizations, **by** diagnosis code and description, **by** month (year).
2. Total number of hospitalizations and billed amount, **by** ward, **by** patient gender (age at date of admission, city, region).
3. Total billed amount, average length of stay and average waiting time, **by** diagnosis code and description, **by** name (specialization) of the physician who has admitted the patient.
4. Total billed amount, and average waiting time of admission, **by** patient age (region), **by** treatment code (description).

HOSPITALIZATIONS: INITIAL LOGICAL SCHEMA



1. Total billed amount for hospitalizations, **by** diagnosis code and description, **by** month (year).



```
SELECT ICD, Month, SUM(Amount)
FROM Hospitalizations, Diagnosis, Date
WHERE DiagnosisFK = DiagnosisPK AND
      DateFK = DatePK
GROUP BY ICD, Month
```

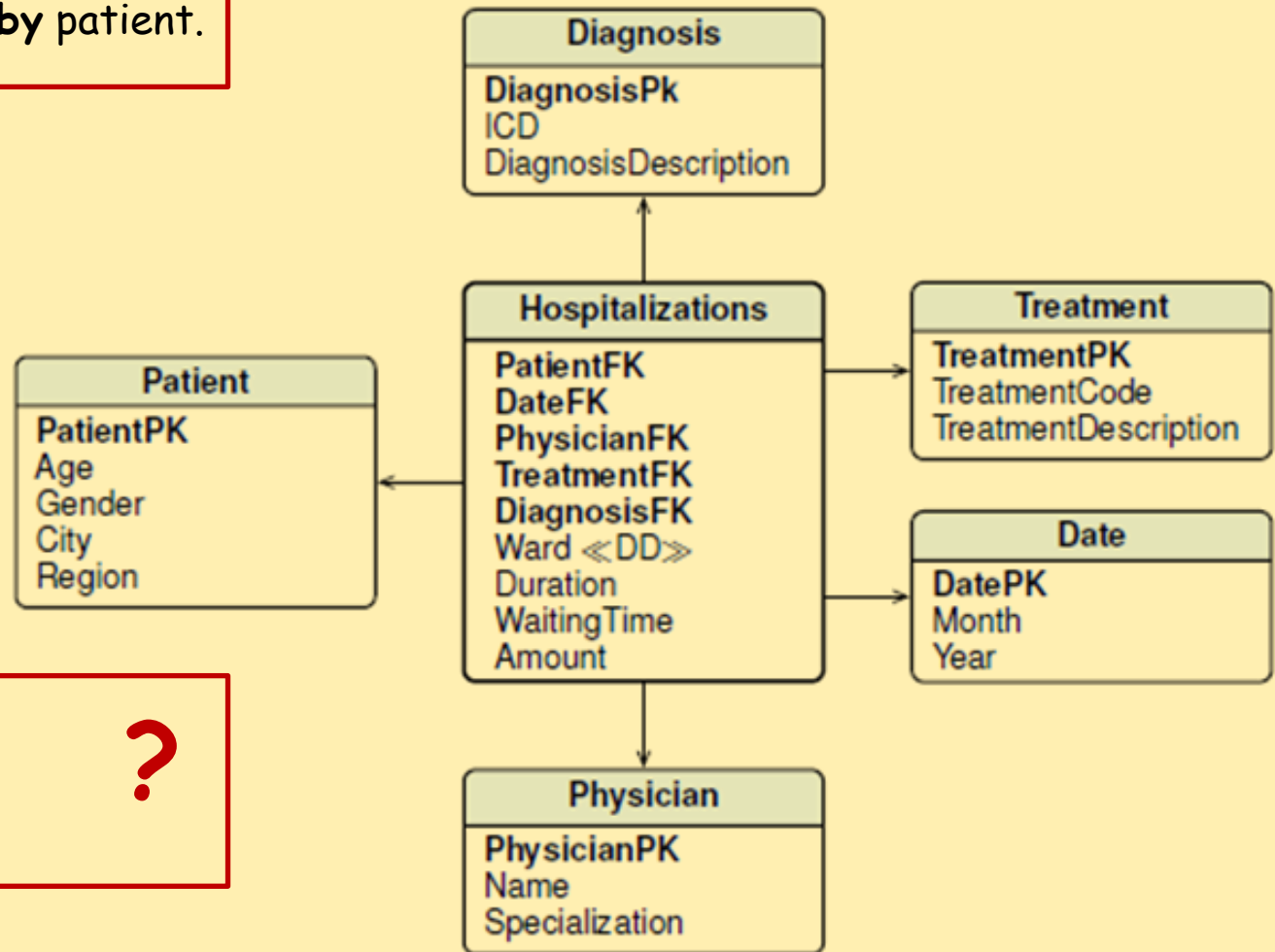
An hospital needs a DM to extract information from their operational database with information about inpatients treatments.

1. Total billed amount for hospitalizations, **by** diagnosis code and description, **by** month (year).
2. Total number of hospitalizations and billed amount, **by** ward, **by** patient gender (age at date of admission, city, region).
3. Total billed amount, average length of stay and average waiting time, **by** diagnosis code and description, **by** name (specialization) of the physician who has admitted the patient.
4. Total billed amount, and average waiting time of admission, **by** patient age (region), **by** treatment code (description).
5. Number of hospitalization **by** patient.

HOSPITALIZATIONS: INITIAL LOGICAL SCHEMA



5. Number of hospitalization by patient.

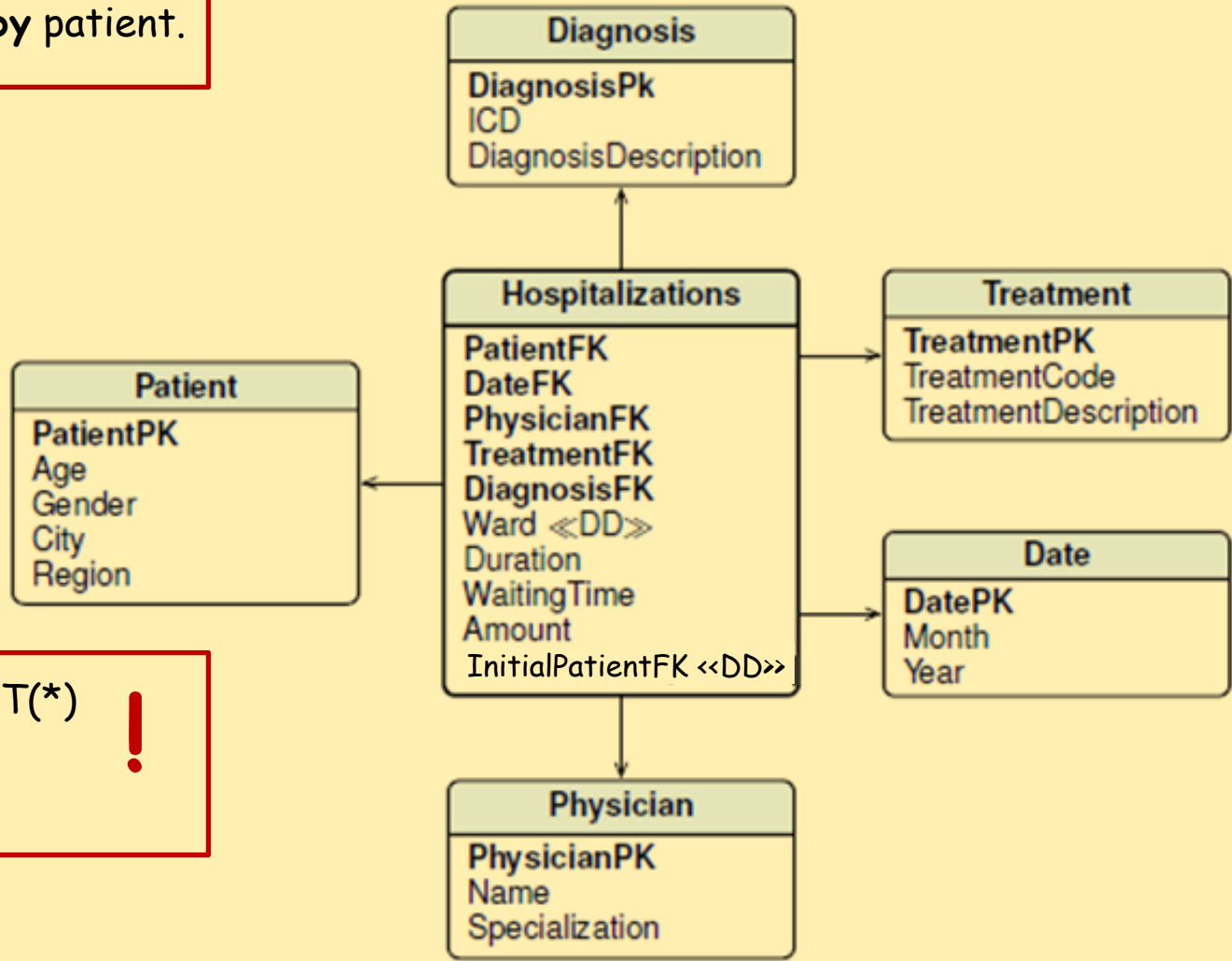


```
SELECT PatientFK, COUNT(*)  
FROM Hospitalizations  
GROUP BY PatientFK
```


HOSPITALIZATIONS: FINAL LOGICAL SCHEMA



5. Number of hospitalization by patient.



```
SELECT InitialPatientFK, COUNT(*)  
FROM Hospitalizations  
GROUP BY InitialPatientFK
```

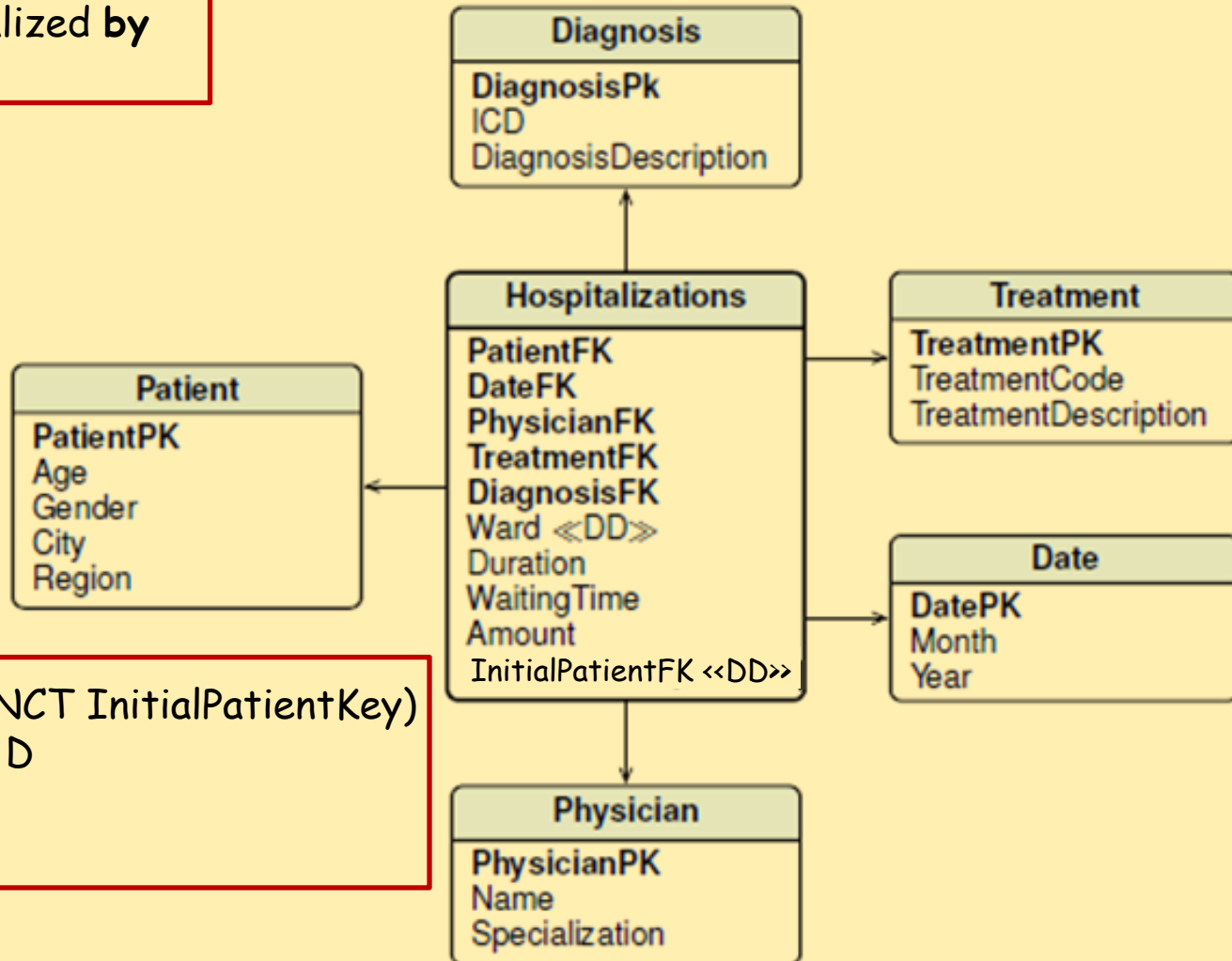
An hospital needs a DM to extract information from their operational database with information about inpatients treatments.

1. Total billed amount for hospitalizations, **by** diagnosis code and description, **by** month (year).
2. Total number of hospitalizations and billed amount, **by** ward, **by** patient gender (age at date of admission, city, region).
3. Total billed amount, average length of stay and average waiting time, **by** diagnosis code and description, **by** name (specialization) of the physician who has admitted the patient.
4. Total billed amount, and average waiting time of admission, **by** patient age (region), **by** treatment code (description).
5. Number of hospitalization **by** patient.
6. Number of patients hospitalized **by** month.

HOSPITALIZATIONS: FINAL LOGICAL SCHEMA



6. Number of patients hospitalized by month.



```
SELECT Month, COUNT(DISTINCT InitialPatientKey)
FROM Hospitalizations H, Date D
WHERE H.DateFK = D.DatePK
GROUP BY Month
```