Basics of a formalism to model

facts,

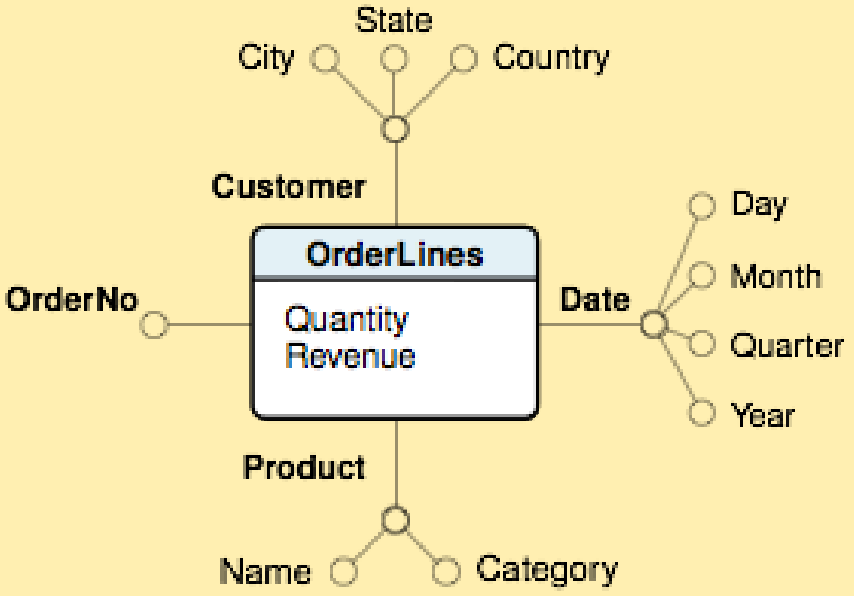
measures,

dimensions,

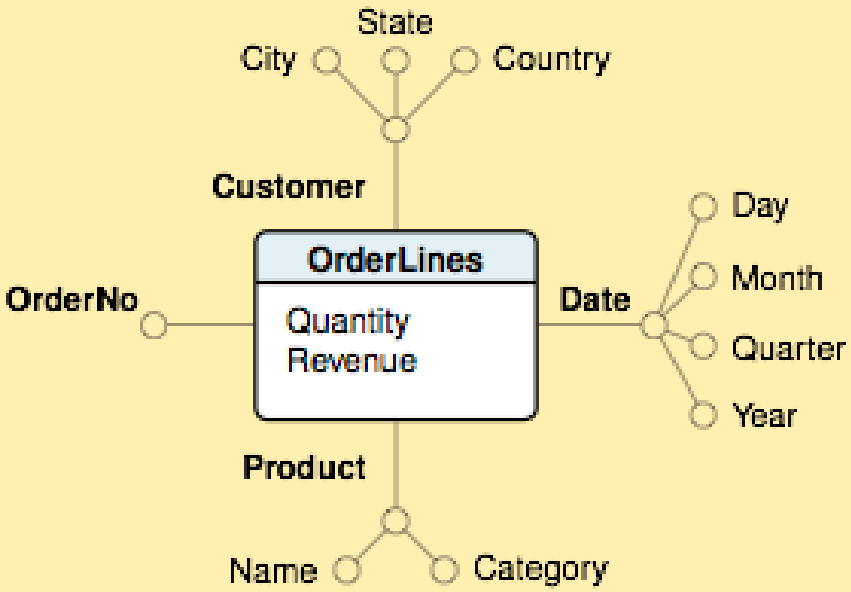
dimensional attributes.

A dimension without **attributes** is called **degenerate**

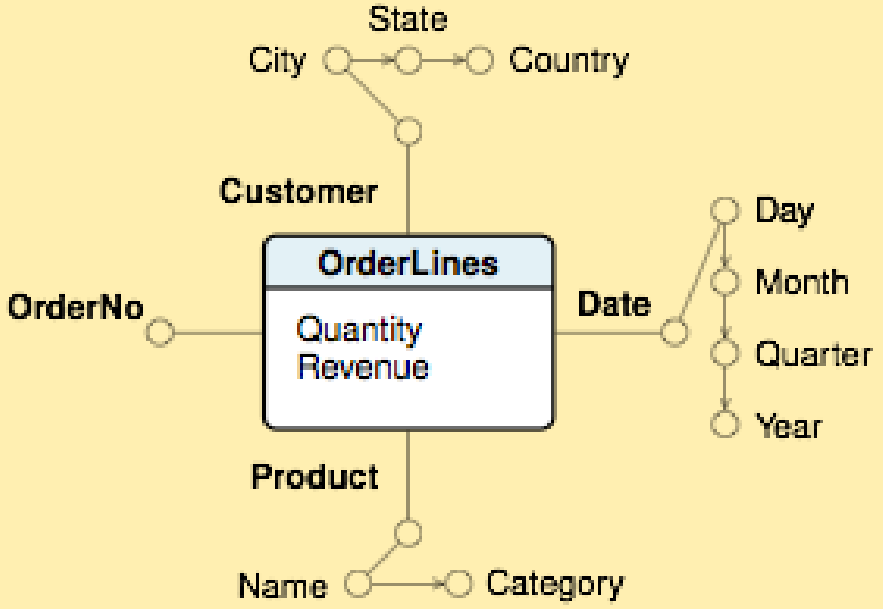
Later on other formalism features and **how to model**...



A DATA MODEL FOR CONCEPTUAL DESIGN: DIMENSIONAL ATTRIBUTES WITH HIERACHIES



Without hierarchies



With hierarchies

CASE STUDY: University Exams

UNIVERSITÀ DI PISA

A university plans to build a data mart that would help them in analyzing the exams performance of the students in **master degree programs** in different academic sessions.

Courses have a code, which is unique, a name, whether it is mandatory or not, the teacher and department name, the credits and the semester in which a course is offered.

Students have a number, which is unique, the gender, the university name that awarded the bachelor degree, the name of the master degree program, the year of enrollment.

Exams have a grade, a value between 1 and 31, considered passed if the grade is greater than 17, the exam session, the academic year. Failed exams are registered too.

1. Number of exams passed, and number of exams failed, **by** course name, **by** academic year, and **by** session.
2. Number of exams failed, by the course name, by academic year, **by** session, and **by** bachelor university name.
3. For a specified master degree program and student' s enrollment year, the average grade of passed exams and the total number of credits given, **by** student gender.
4. For the current academic year, average exams grade, number and the percentage of students who passed the exam, **by** the course name, and **by** session.
5. For a specified master degree program and courses with a number of exams passed of less than 3, the number of exams, **by** the course name, **by** academic year.

REQUIREMENTS SPECIFICATION

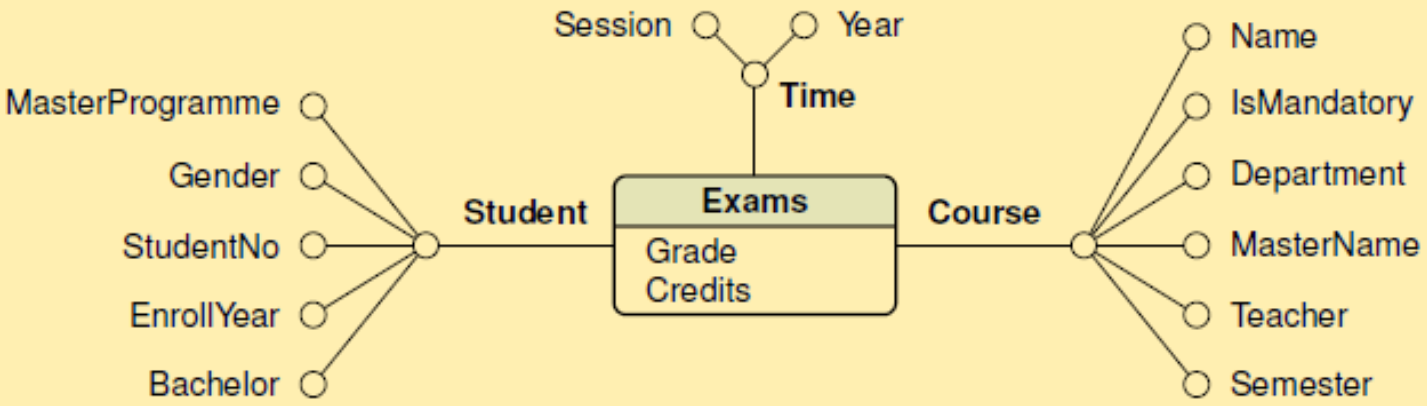
Passed is a calculated measure

UNIVERSITÀ DI PISA

Requirements analysis	Dimensions	Measures	Metrics
1. Number of exams passed, and number of exams failed, **by** course name, **by** academic year, and **by** session.	Course(name), Time(a.y., session)	Grade Passed=CASE WHEN Grade>17 THEN 1 ELSE 0 END	SUM(Passed), SUM(1-Passed)
2. Number of exams failed, by the course name, by academic year, **by** session, and **by** bachelor university name.	Course(name), Time(a.y., session), Student(BachelorUniv)	Grade Passed	SUM(1-Passed)
3. For a specified master degree program and student's enrollment year, the average grade of passed exams and the total number of credits given, **by** student gender.	Student(Master,enrr.year, gender),	Grade, Credits	AVG(Grade), SUM(Credits)
4. For the current academic year, average exams grade, number and the percentage of students who passed the exam, **by** the course name, and **by** session.	Course(name), Time(a.y.,session)	Passed	SUM(Passed), COUNT(*)
5. For a specified master degree program and courses with a number of exams passed of less than 3, the number of exams, **by** the course name, **by** academic year.	Course(Master,name), Time(a.y.)	Passed	COUNT(*)

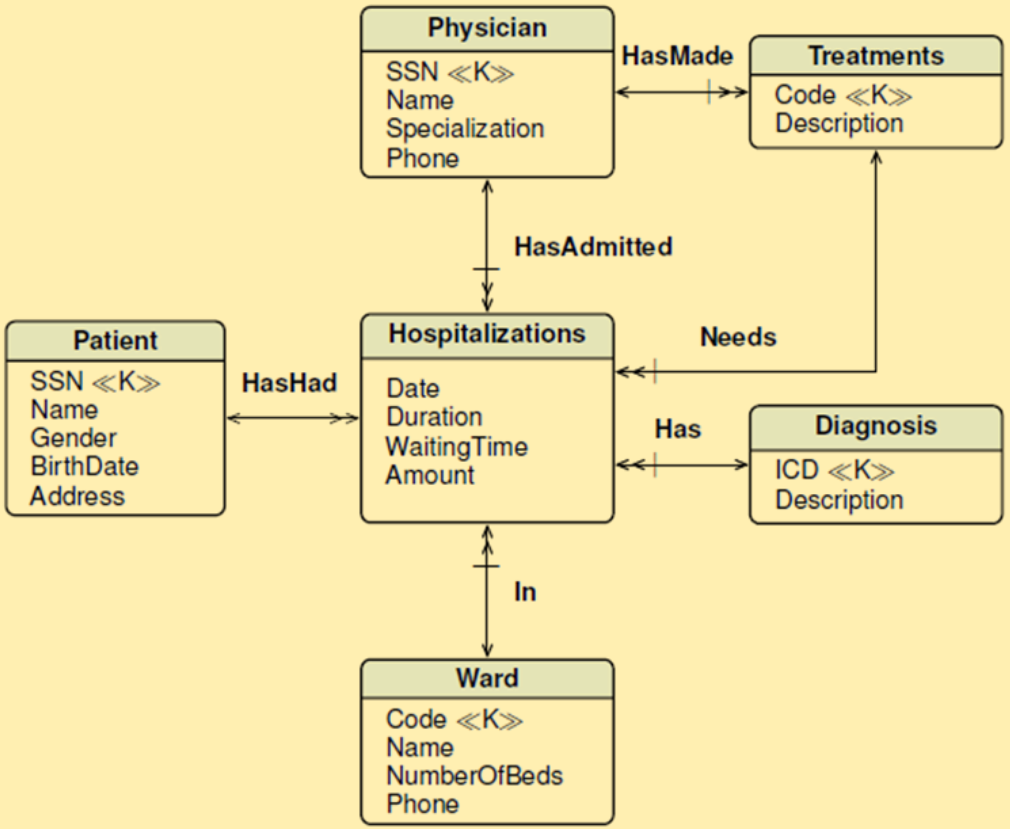
CASE STUDY: University Exams

Fact granularity	
Description	A fact is the occurrence of an exam
Preliminary dimensions	Student, Time, Course
Preliminary measures	Grade, Credits



CASE STUDY: HOSPITAL (Appendix A1)

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).

Hospitalization

Requirements analysis	Dimensions	Measures	Metrics

Fact granularity

Description	
Preliminary dimensions	
Preliminary measures	

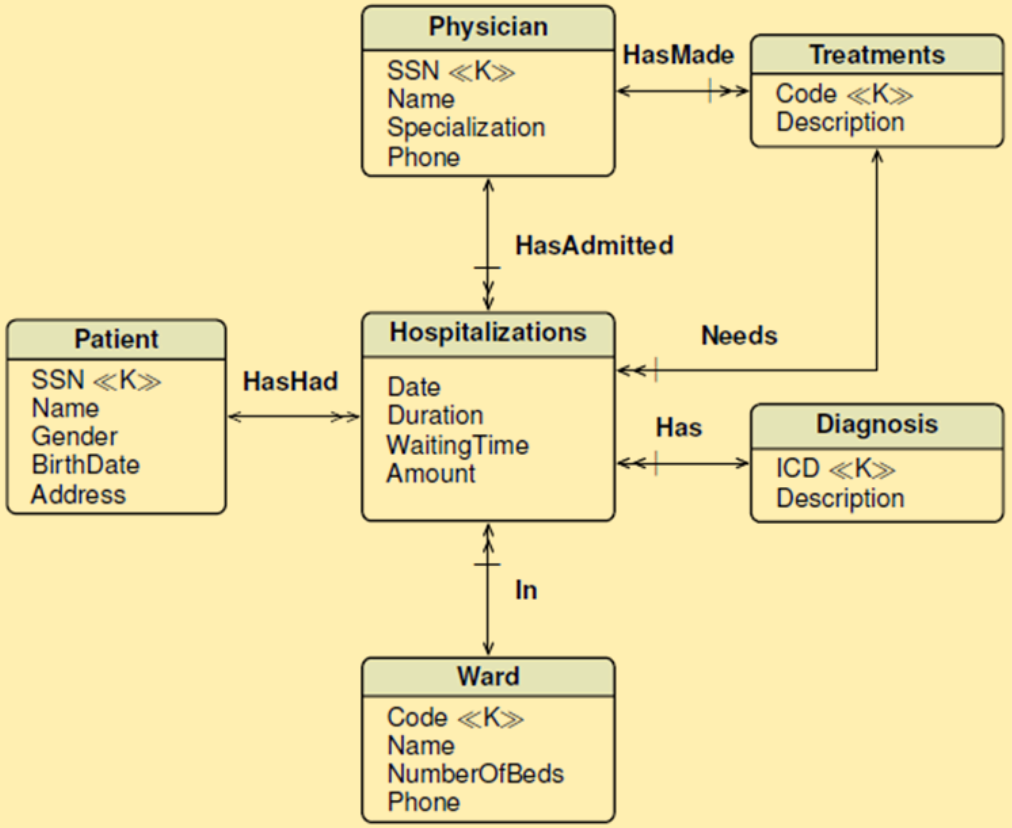
Data Mart Conceptual Schema

Requirements analysis	Dimensions	Measures	Metrics
Total billed amount for hospitalizations, by diagnosis code and description, by month (year).	Diagnosis (ICD, Description), Date (Month, Year)	Amount	Total Amount
Total number of hospitalizations and billed amount, by ward, by patient gender (age at date of admission, city, region).	Ward, Patient (Gender, Age, City, Region)	Amount	Total number Total Amount
Total billed amount, average length of stay and average waiting time by diagnosis code and description, by name (specialization) of the physician who admitted the patient.	Diagnosis (ICD code, Description), Physician (Name, Specialization)	Amount, Duration, WaitingTime	Total Amount Average Duration Average WaitingTime
Total billed amount, and average waiting time for admission by patient age (region), by treatment code (description).	Patient (Age, Region), Treatment (Code, Description)	Amount, WaitingTime	Total Amount Average WaitingTime

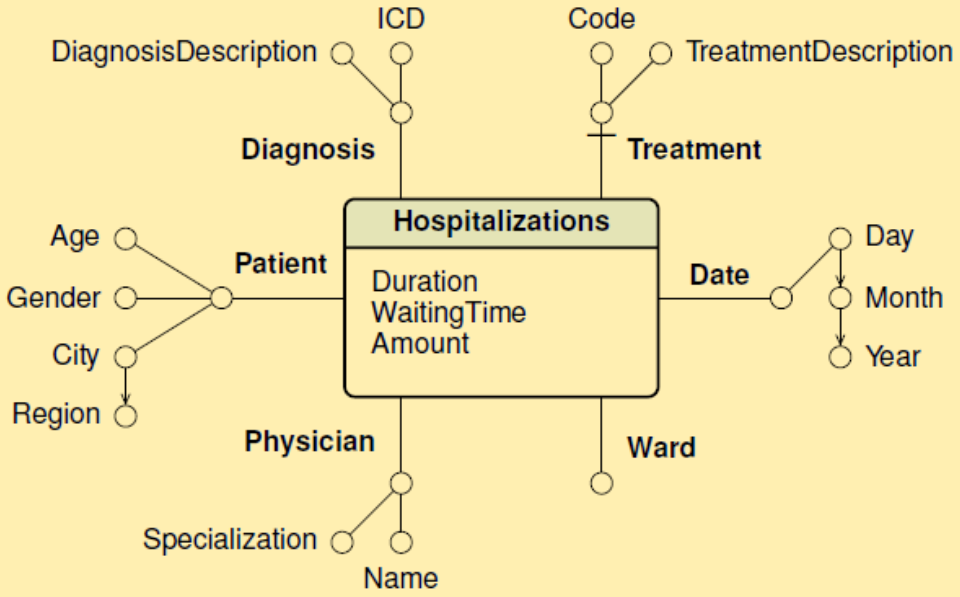
REQUIREMENTS SPECIFICATION

	Fact granularity
Description	A fact is a hospitalization of a patient, assuming that they may require one treatment only
Preliminary dimensions	Patient, Date, Ward, Diagnosis, Treatment, Physician
Preliminary measures	Duration, WaitingTime, Amount

HOSPITALIZATIONS DATA MART CONCEPTUAL SCHEMA



DATA BASE



DATA MART

SUMMARY

The **analysis-driven** design of a data mart.

Business questions

For a data subsets to use,

the metrics to compute,

grouping data **by dimensions (attributes)**,

how the result should be presented.

SELECT D, SUM(M) **FROM** … **WHERE** C **GROUP BY** D **ORDER BY** O

Facts granularity, measures and their types, dimensions

Data availability