- A **symbolic model** is a subjective formal representation of ideas and knowledge about some aspects of the real world (*domain of discourse*), designed to serve an explicit purpose.

- A **data model** is a set of abstraction mechanisms to describe abstract knowledge

  What is the problem?

  - Conceptual data model: to analyse a problem, given user requirements
    - E.g., E-R or Entity-Relationship, **ODM or Object Data Model**

  How to solve it?

  - Logical model: to design a solution independently of actual DBMS
    - E.g., **Relational Data Model**

  How to implement a solution?

  - Physical model: to realize a project on a specific DBMS

To define the structure of a DW the following **data models** are used:

What is the problem?

The **Dimensional Fact Model** (DFM) is a graphical conceptual model.
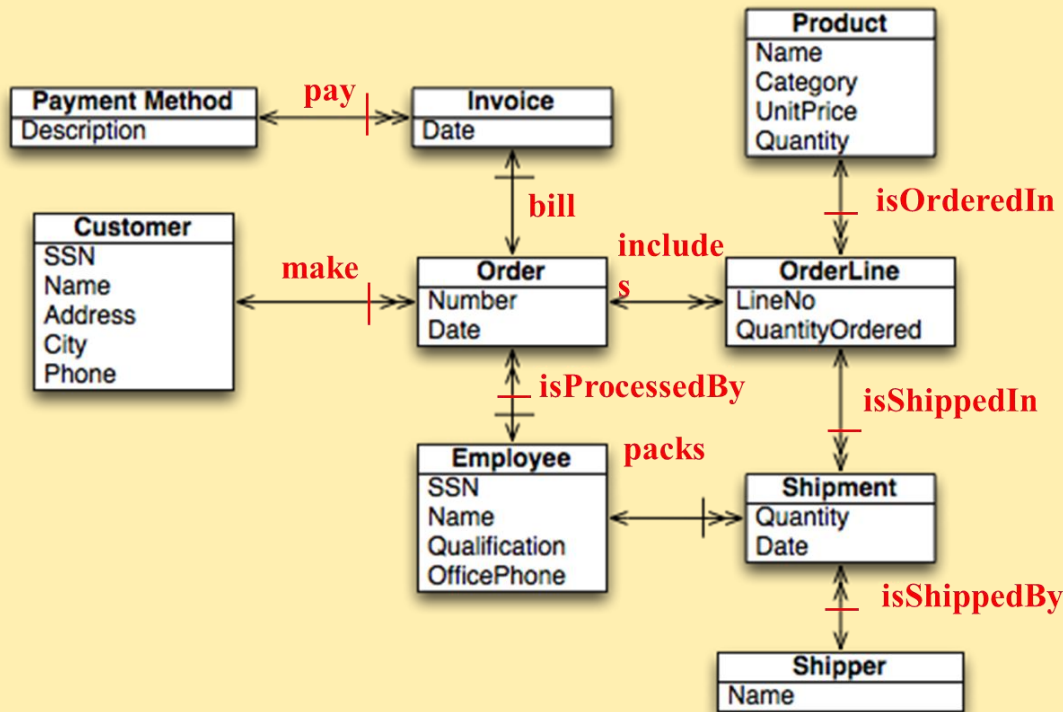
How to solve it?

The **Relational Data Model**, as a logical model

How to solve it?

The **Multidimensional Model** (called **Cube**), another logical data model useful to understand OLAP operations.

Today, we start looking at a conceptual data model for DW

GOAL: AN ORDER DATA MART

OPERATIONAL DATABASE

Number of products ordered,
by product, **by** customer, **by** month

Total revenue **by** product category,
by customer, **by** year

For customers of Italy, total revenue **by**
customer city, **by** year, **by** quarter

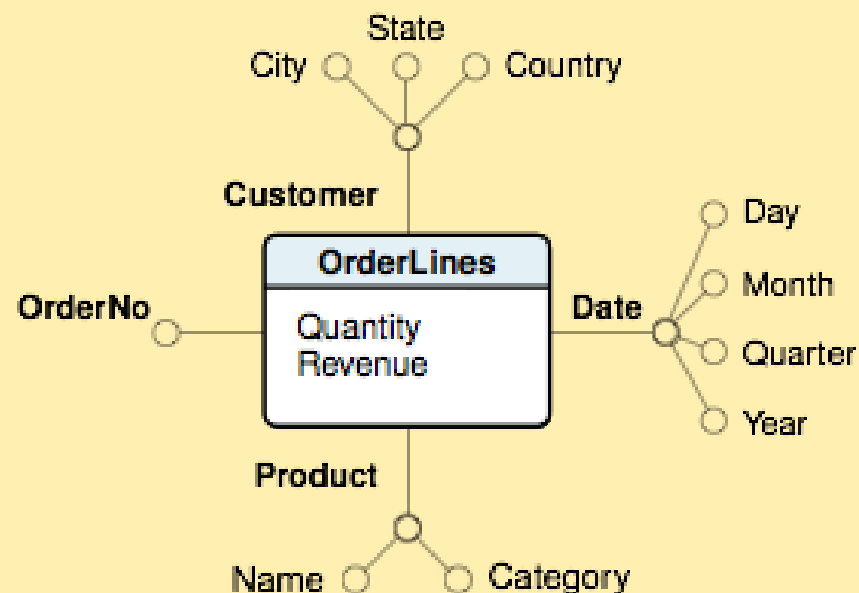**BUSINESS QUESTIONS**

Basics of a formalism to model

facts,

measures,

dimensions,

dimensional attributes.
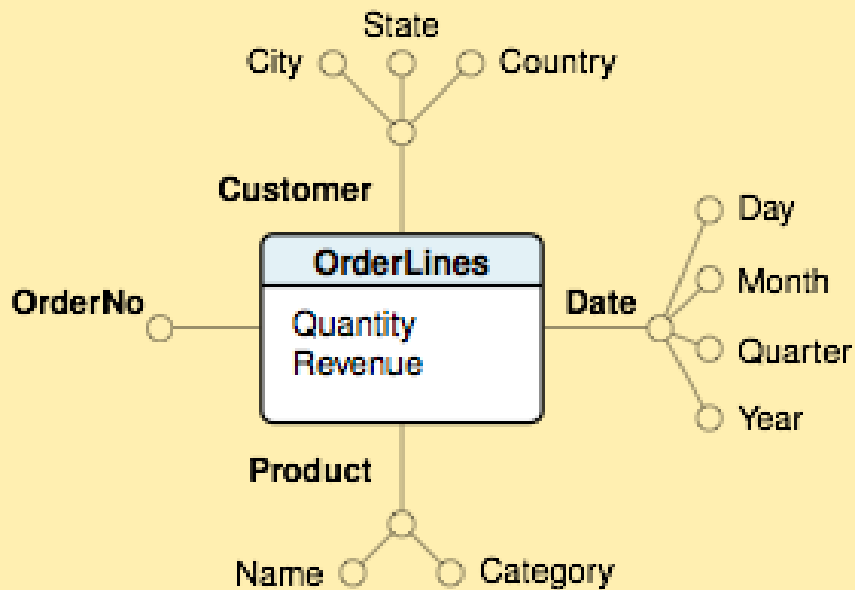
A dimension without **attributes** is called **degenerate**
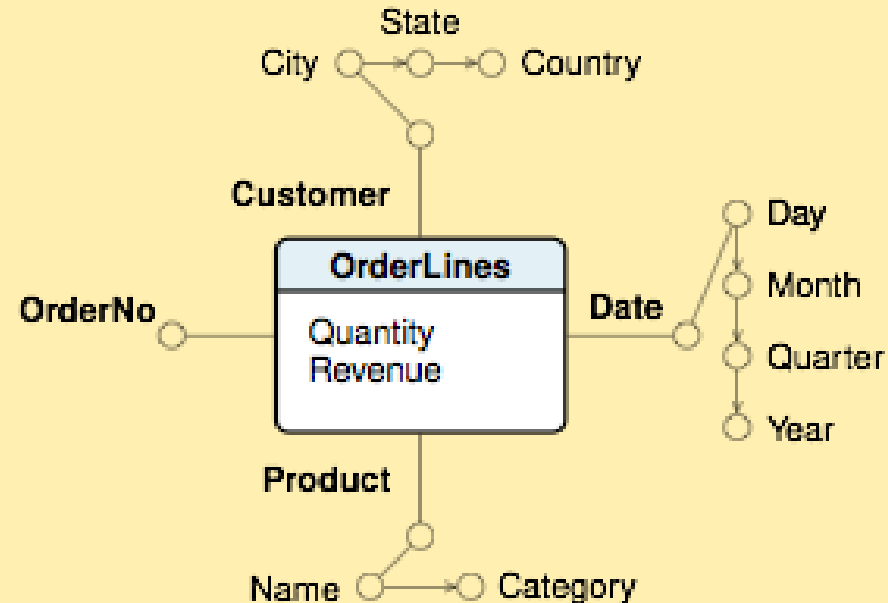
Later on, other formalism features and **how to model...**

Without hierarchies

With hierarchies

Let us assume that a key business process of interest has been identified together with a sample of analysis to perform to support decisions.

What are the steps for designing a DFM schema?

UNIVERSITÀ DI PISA

Step 0: Requirements gathering.

Requirements gathering focuses on the study of business processes and on analysis relevant for decision making.

A not useful requirement analysis (a business question to answer):

Why is my business not meeting the targets?

A useful business question:

Total revenue, **by** customer, **by** product

Aggregation (metric)   **Measure**   **Dimensions**
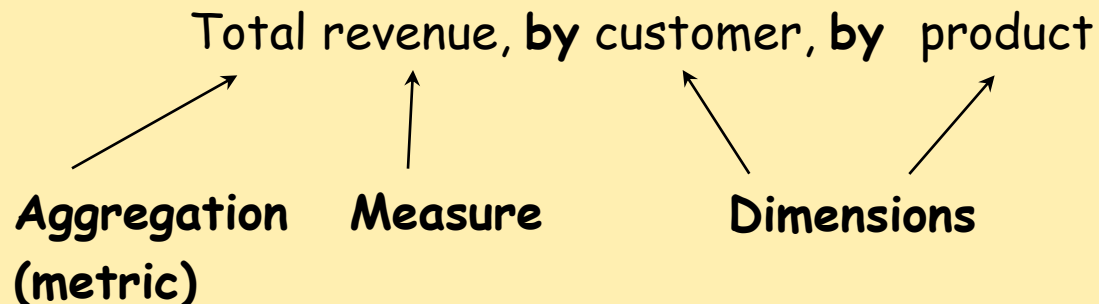
Alternative: A report example

Step 1: Identify the Granularity of the Fact

The first fundamental decision to be taken is the meaning of the fact.

What is the **grain ?**

Identifying the grain also means deciding on the level of detail you want to be made available in the dimensional model. The more detail there is, the lower the level of granularity.
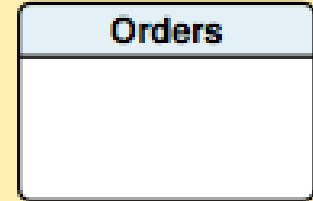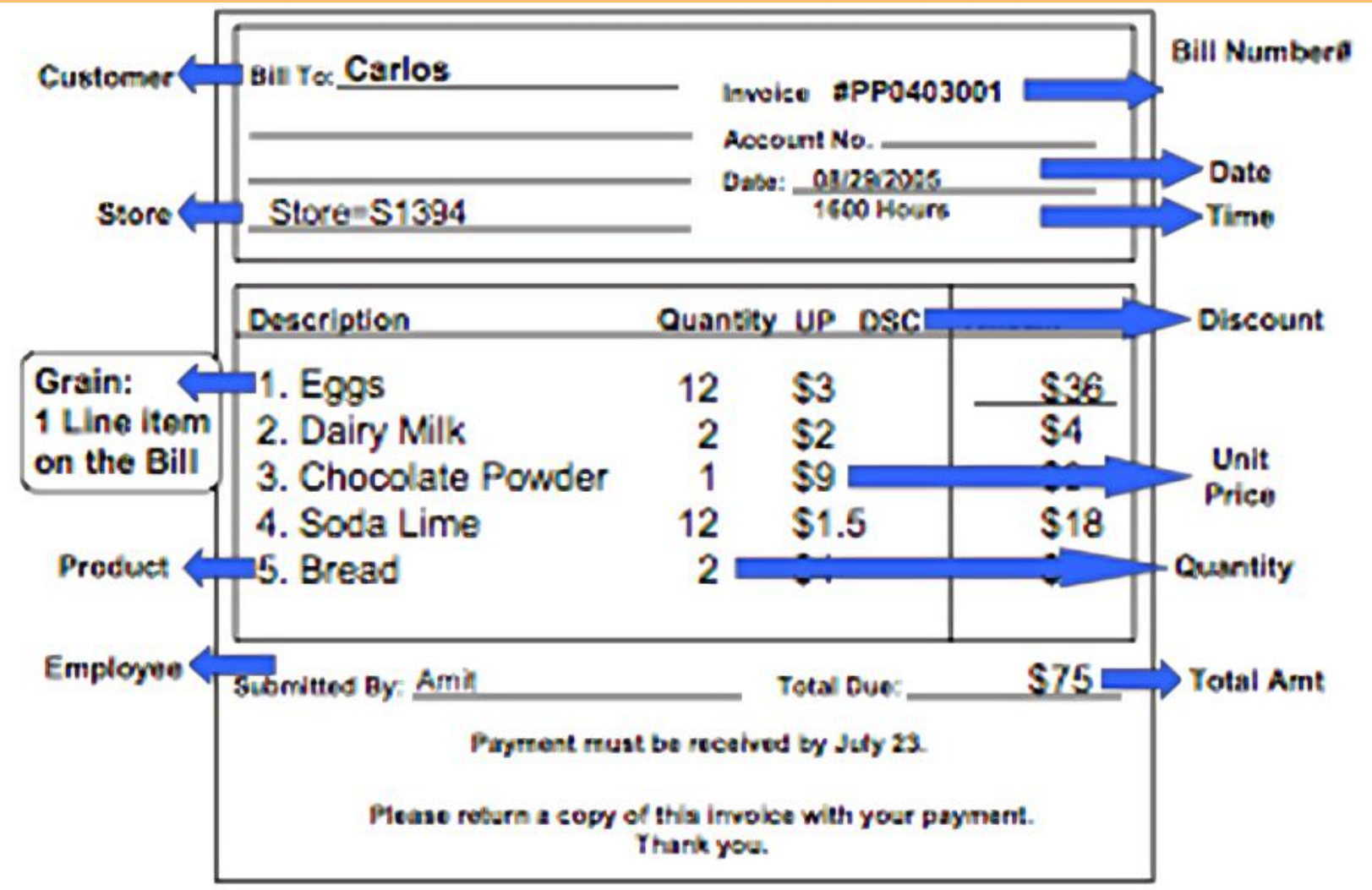
Remember:
1. **Grain is the precision with which the measurements are taken.**

2. **Grain determines measures and dimensions and dimensions determine grain !**

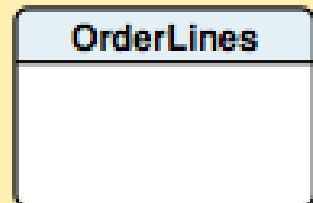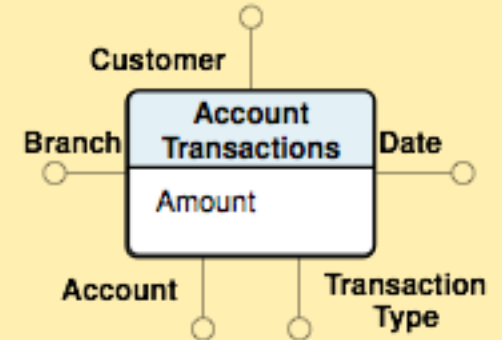Example: Analyses are about customer orders. **What is an Order?**

AN ORDER EXAMPLE

Transaction

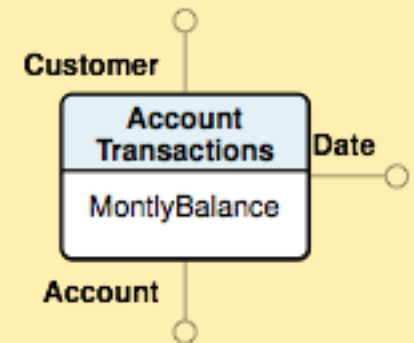One fact per transaction (an event that occurs at a specific point in time)

Example: A transaction for an individual account of
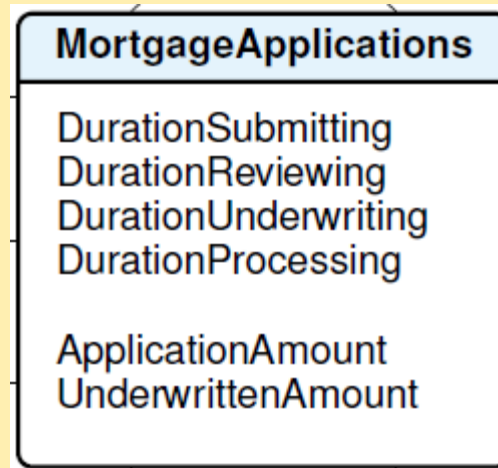
a customer of a bank

Periodic

One fact for a group of transactions made over a period of time.

Example: The **amount** is the **monthly balance** for all transaction against an individual account of a customer of a bank.

Accumulating   One fact for the entire lifetime of an evolving event that has a duration and change over time

Example: The lifetime of a mortgage application.

MortgageApplications

DurationSubmitting
DurationReviewing
DurationUnderwriting
DurationProcessing

ApplicationAmount
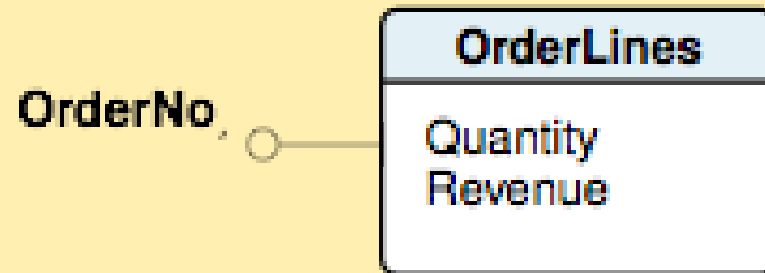UnderwrittenAmount

Facts and measures only…

Step 2: Identify the Fact Measures

The **measures** of interest are **numeric values** that make sense to add.

Not everything that is numeric is a measure!
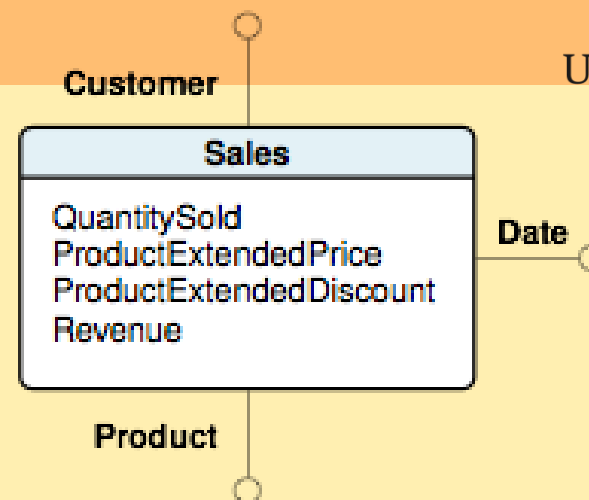
Remember:
A measure is an observation of the performance of a business process

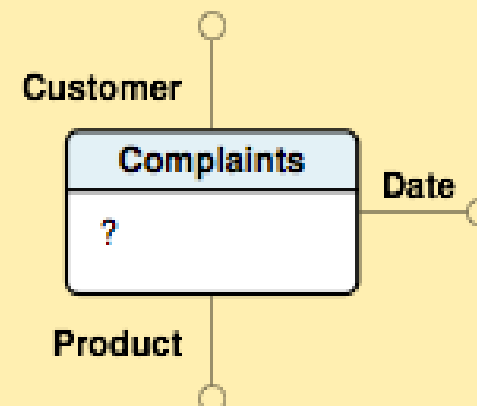**It is important to specify a measure Type.**

MEASURE TYPES

Numeric (calculated) **additive**.

The measures may be missing !

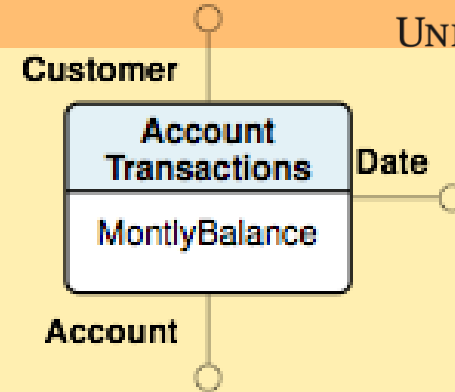**Factless** (better **Measureless**)

Numeric **non-additive**.

Gross Margin = Margin/Revenue ?

Unit Price ?

Customer

Sales

QuantitySold
ProductExtendedPrice
ProductExtendedDiscount
Revenue

Date

Product

Customer

Complaints

?

Date

Product

Numeric **semi-additive, with respect to a dimension**

Definition A measure M is **semi-additive with respect to a dimension D1** when it **can not be aggregated** with the function **SUM** for groups of data **with different values of D1**.

Total MontlyBalance, by Customer (?)

Total MontlyBalance, by Account (?)

Total MontlyBalance, by Date   (?)

Total MontlyBalance in d1, by Customer (?)

AccountTransactions

Customer	Account	Date	MontlyBalance
c1	a1	d1	100
c2	a2	d1	200
c1	a3	d1	500
c2	a4	d1	800
c1	a1	d2	400
c2	a2	d2	300
c1	a3	d2	400
c2	a4	d2	600

Step 3: Identify the Fact Dimensions

> Identify the **dimensions to** give fact measures their **context.**

The **Five Ws and one H questions, or the Six Ws** (?)

(from Wikipedia) are questions whose answers are considered basic in information-gathering. They are often mentioned in journalism, research, and police investigation. They constitute a formula for getting the complete story on a subject. According to the principle of the **Six Ws**, a report can only be considered complete if it answers the following questions:

Who is it about?   **What** happened?   **When** did it take place?

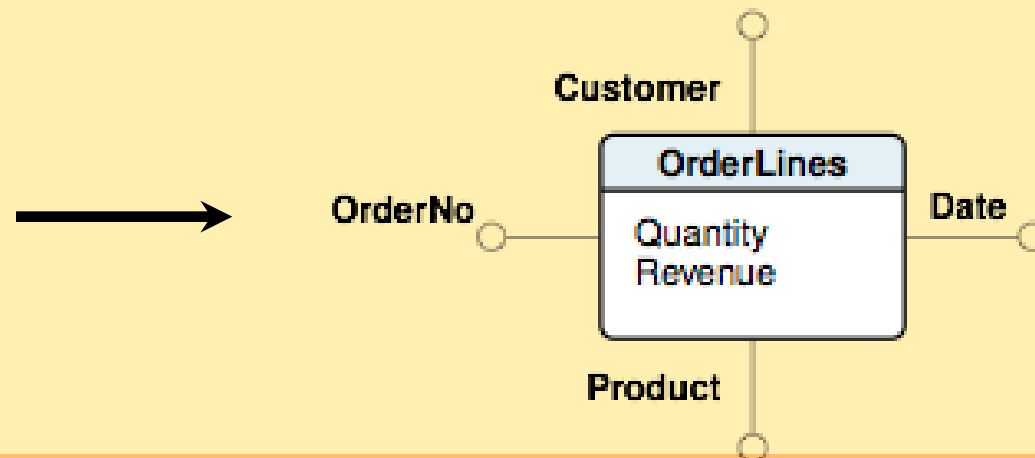**Where** did it take place?   **Why** did it happen?   **How** did it happen?
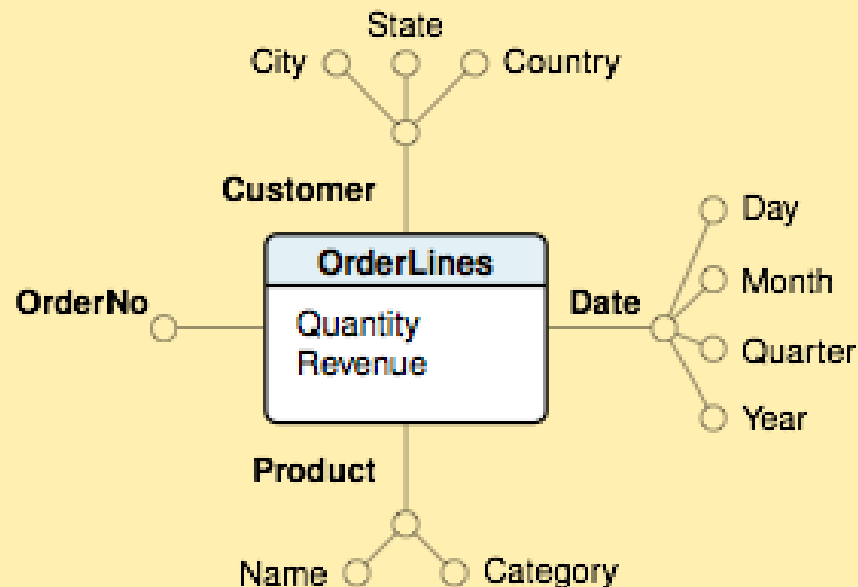
Step 3: Identify the Fact Dimensions

Identify the **dimensions to** give fact measures their **context.**

The **Six Ws** questions aim to identify the **variables determining the measures** and possible **intervention levers**.

Step 4: Identify Dimensional Attributes

The dimensional attributes are important for analysis and for reports.

UNIVERSITÀ DI PISA

Step 5: Identify the Dimensional Attribute Hierarchies

Attribute hierarchies is a natural way to support interactive exploration of facts. Users understand them intuitively, because they are used to look at a summarized report and then to decide to look at a more detailed one.