

# Master Degree Programme in Computer Science

## Enterprise Information Systems

### 12. Data Warehousing – Design Issues



SAPIENZA  
UNIVERSITÀ DI ROMA

Prof. Ing. Claudio CILLI

cilli@di.uniroma1.it

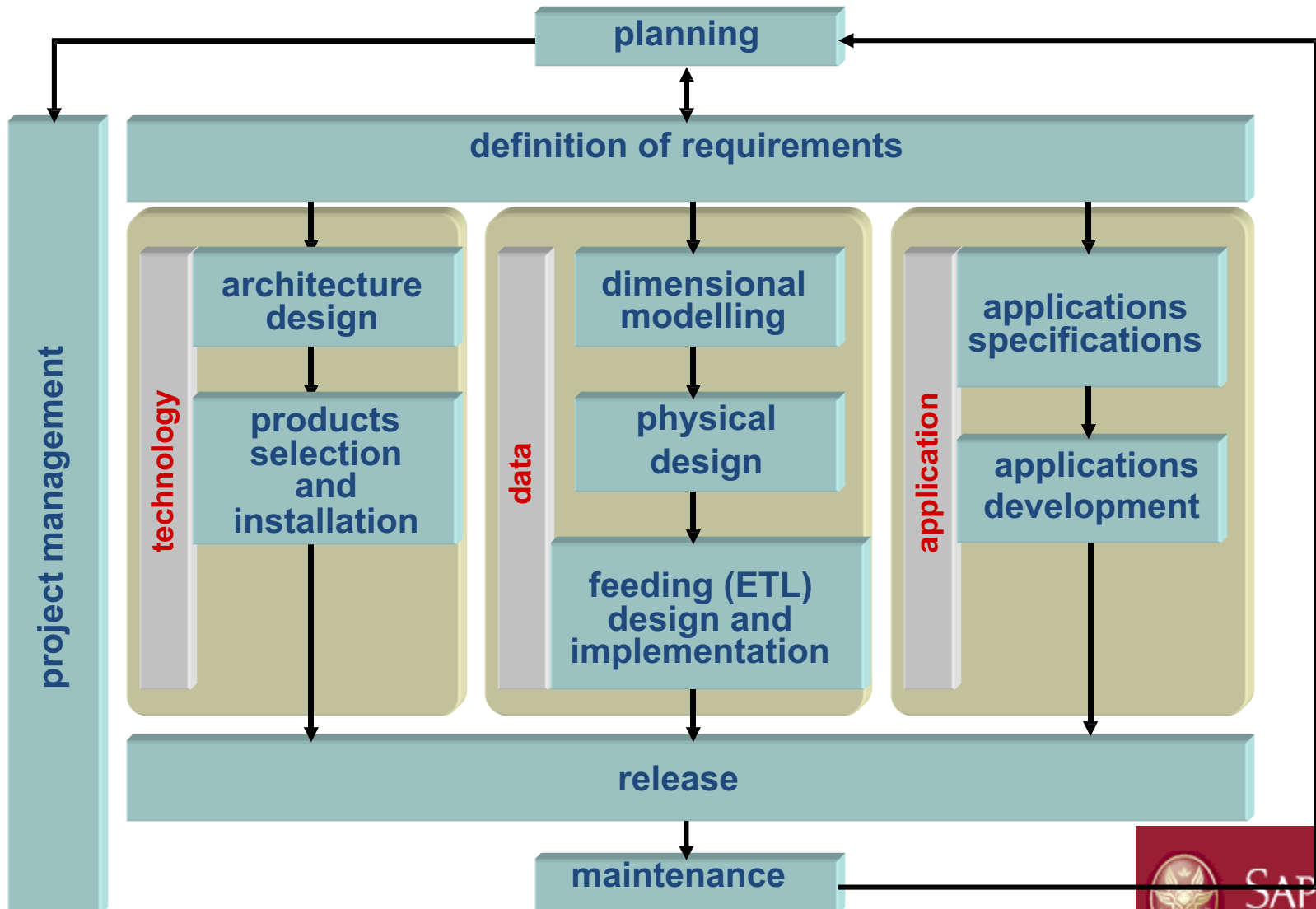
<http://dsi.uniroma1.it/~cilli>

# Data Warehouse Design

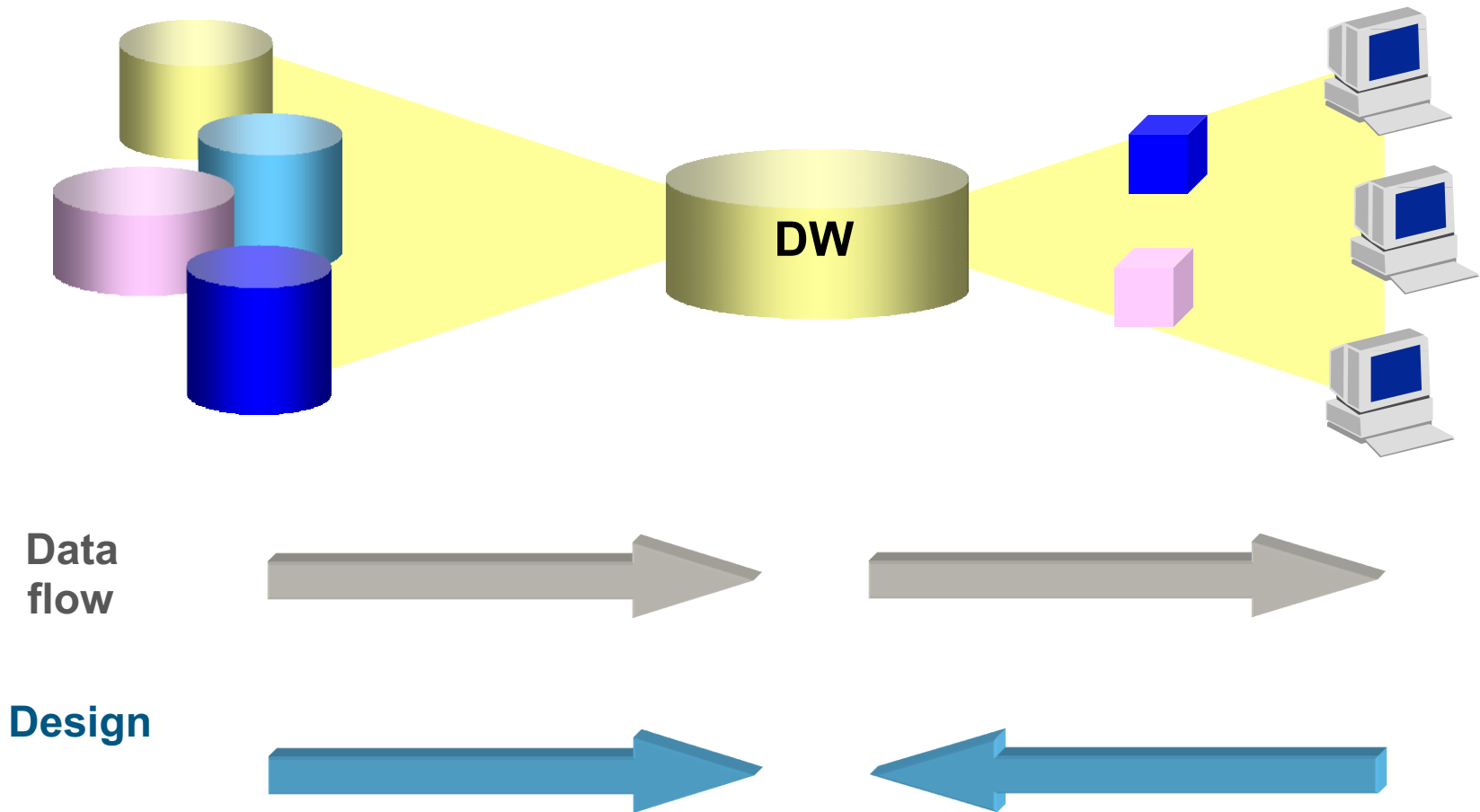
- Setting targets and planning
  - feasibility (border, size, sources, ...)
  - team
  - operating plan
- Design of the infrastructure
  - choice of architecture
  - choice of technologies
- Design of Data Marts
  - analysis with domain experts



# Lifecycle (Kimball, 1998)



# Data Flow & Project evolution



# Design of a Data Mart: phases

1. Analysis and reconciliation of data sources

schema of sources → reconciled schema

**entity-relationship**

2. Requirements analysis

reconciled schema → facts, work load

3. Conceptual Design

**Fact schema**

reconciled schema, facts, work load → fact schema

4. Logical design

**Star-schema, Snowflakes**

fact schema, work load → logical schema of Data Marts

5. Feeding (ETL) Design

schema of sources, reconciled schema, logical schema of DM → ETL procedures

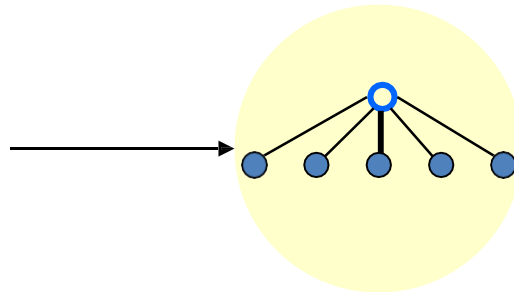
6. Physical design

logical schema of DM, work load, DBMS → physical schema of DM

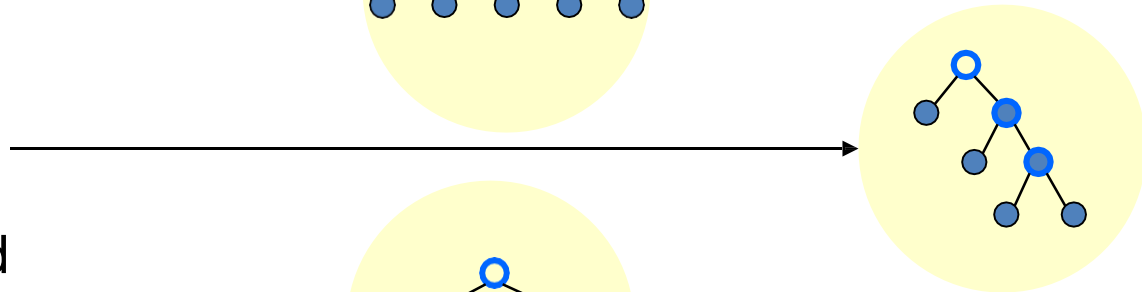
# Reconciling data sources

schema integration:

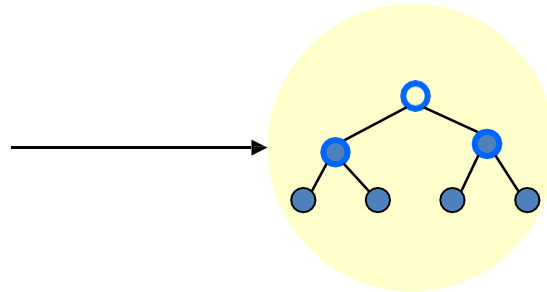
- one step



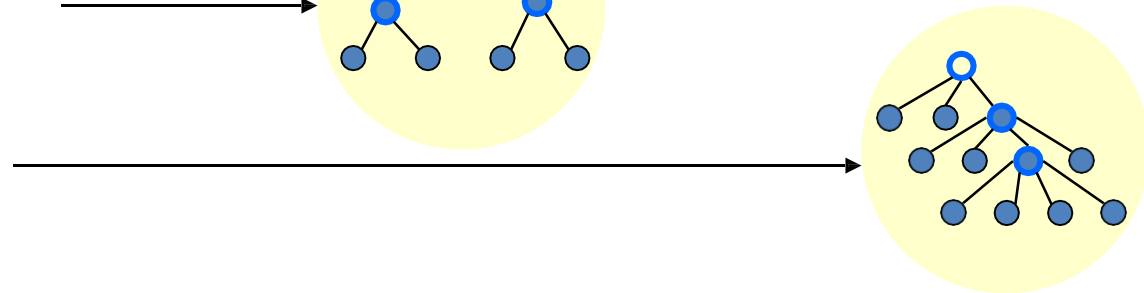
- steps



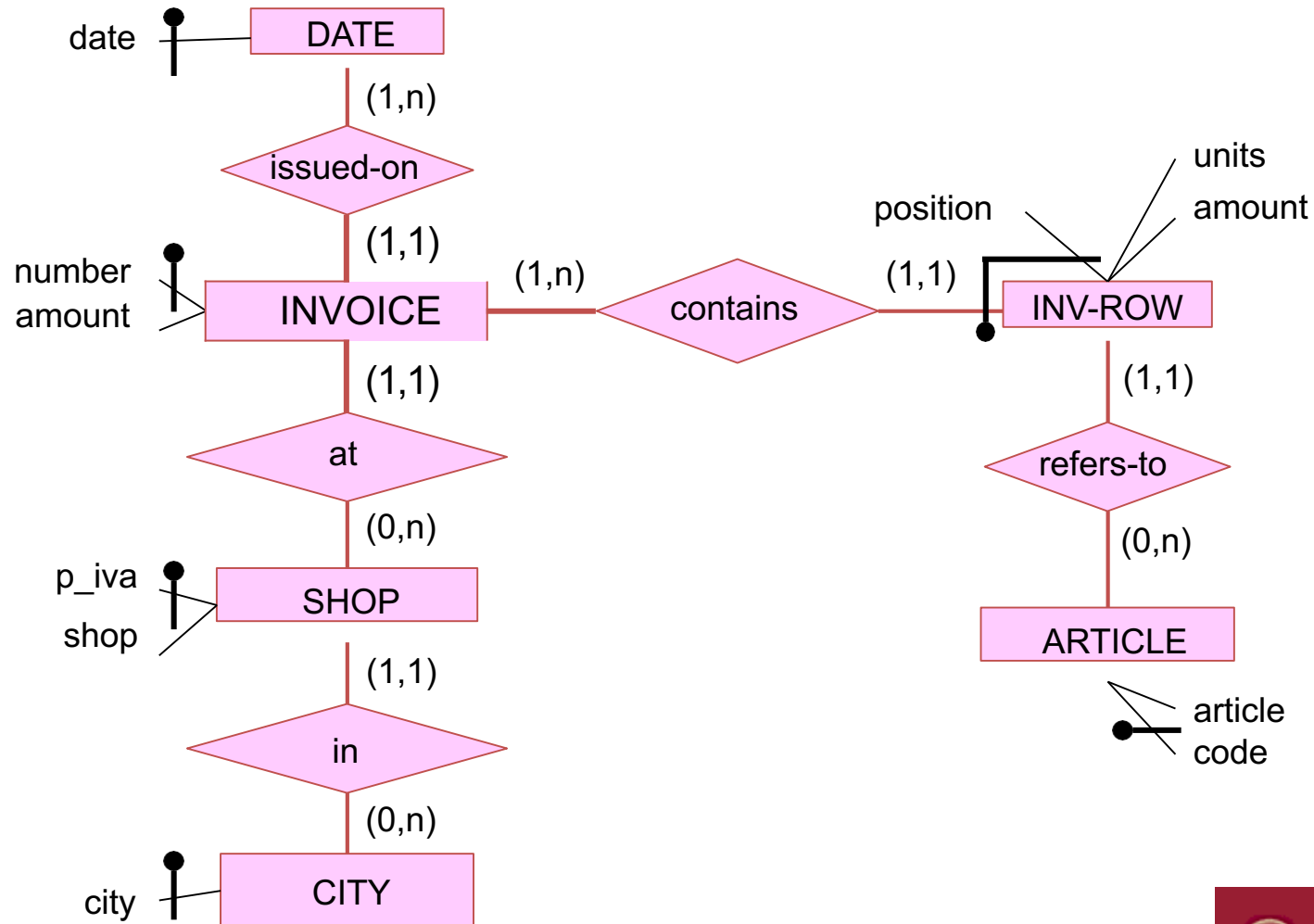
- balanced



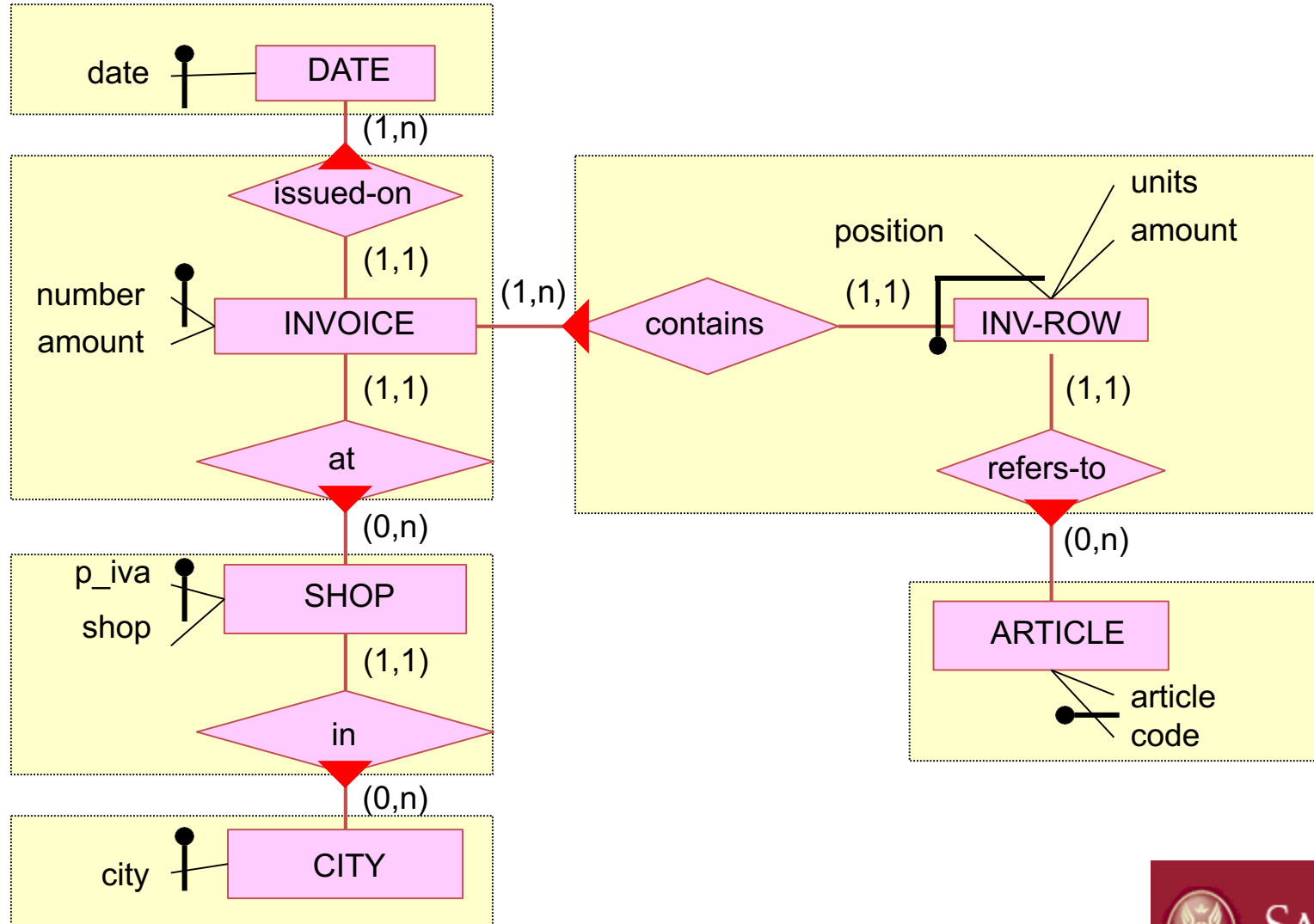
- iterative



# Operational data source: ER-Schema



# Operational data source: logical schema





# Operational data source: logical schema (rev.)

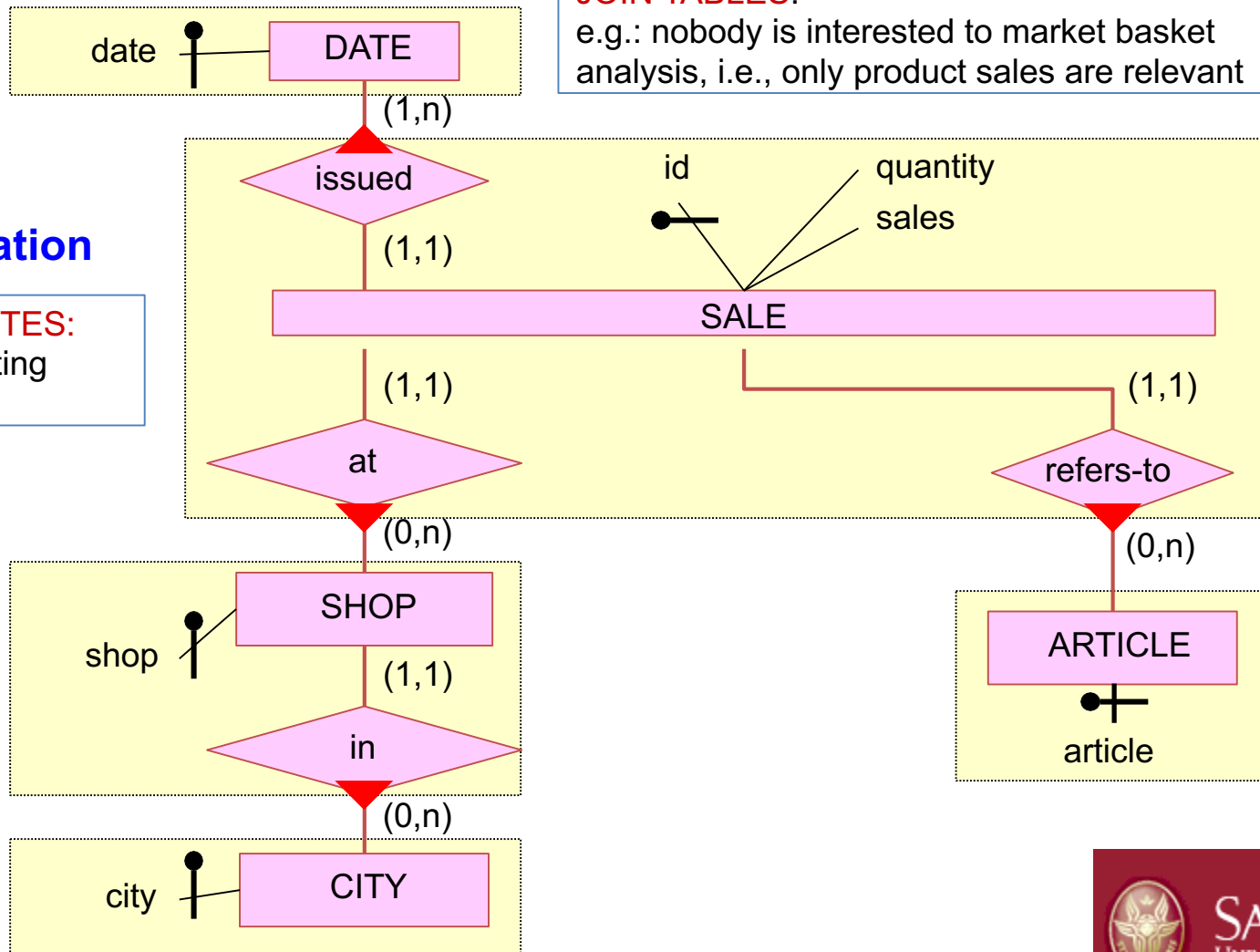
## denormalization

### JOIN TABLES:

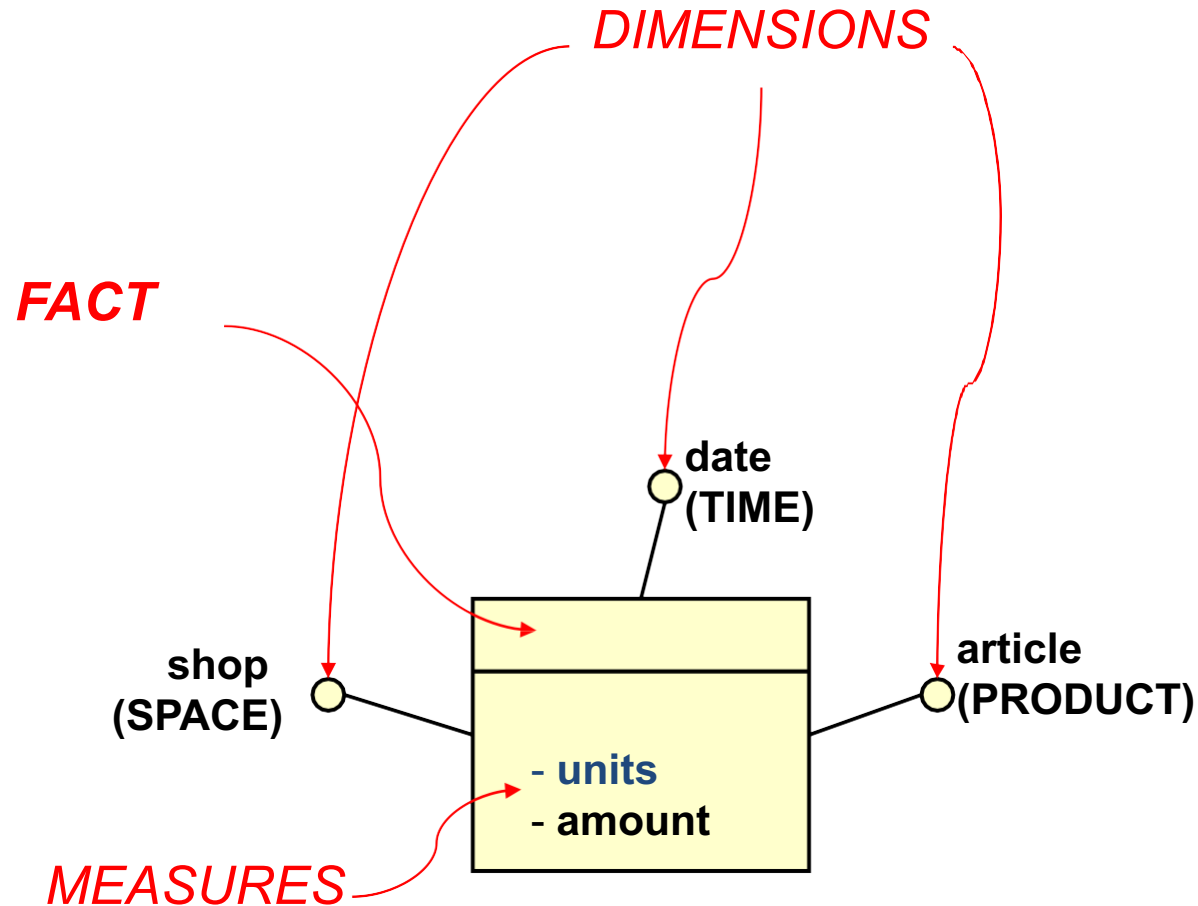
e.g.: nobody is interested to market basket analysis, i.e., only product sales are relevant

### simplification

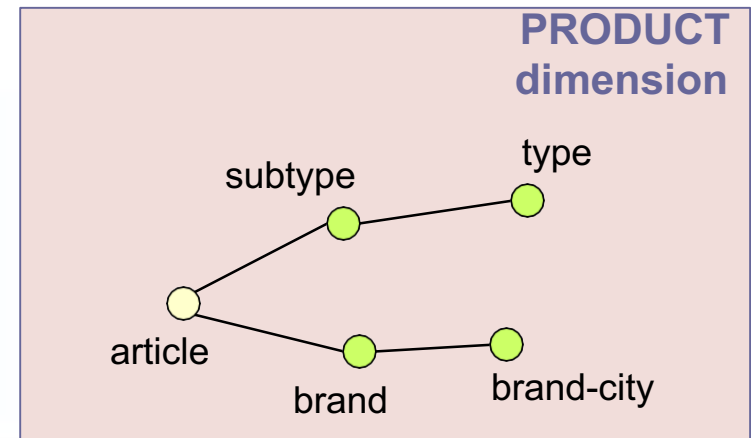
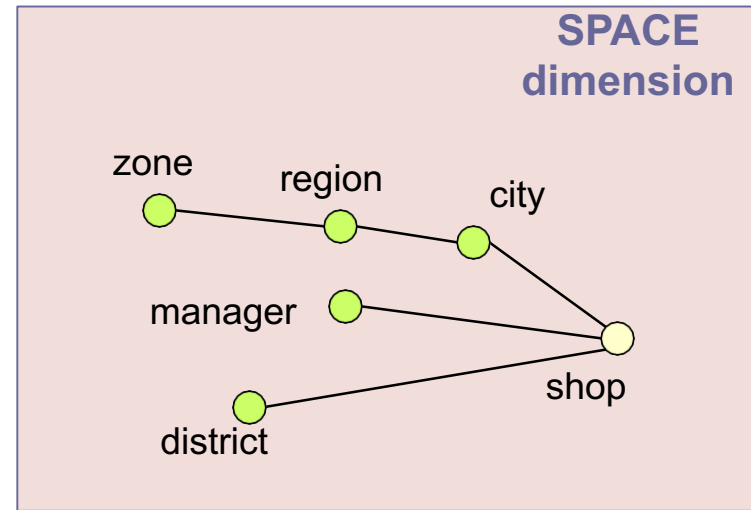
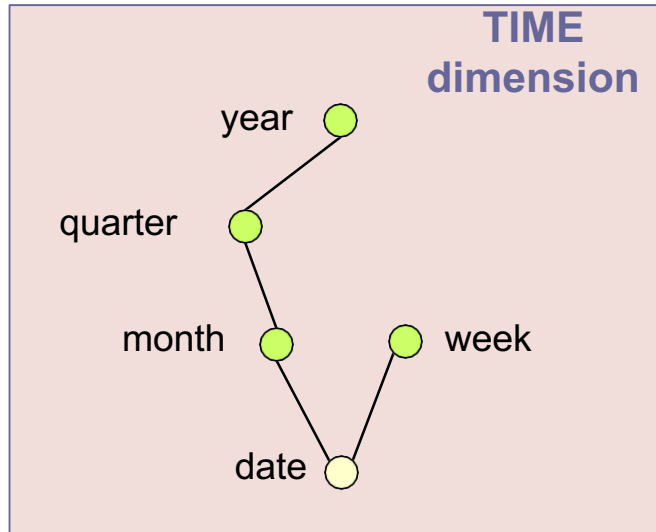
**DROP ATTRIBUTES:**  
delete uninteresting attributes



# Fact Schema (preliminary)

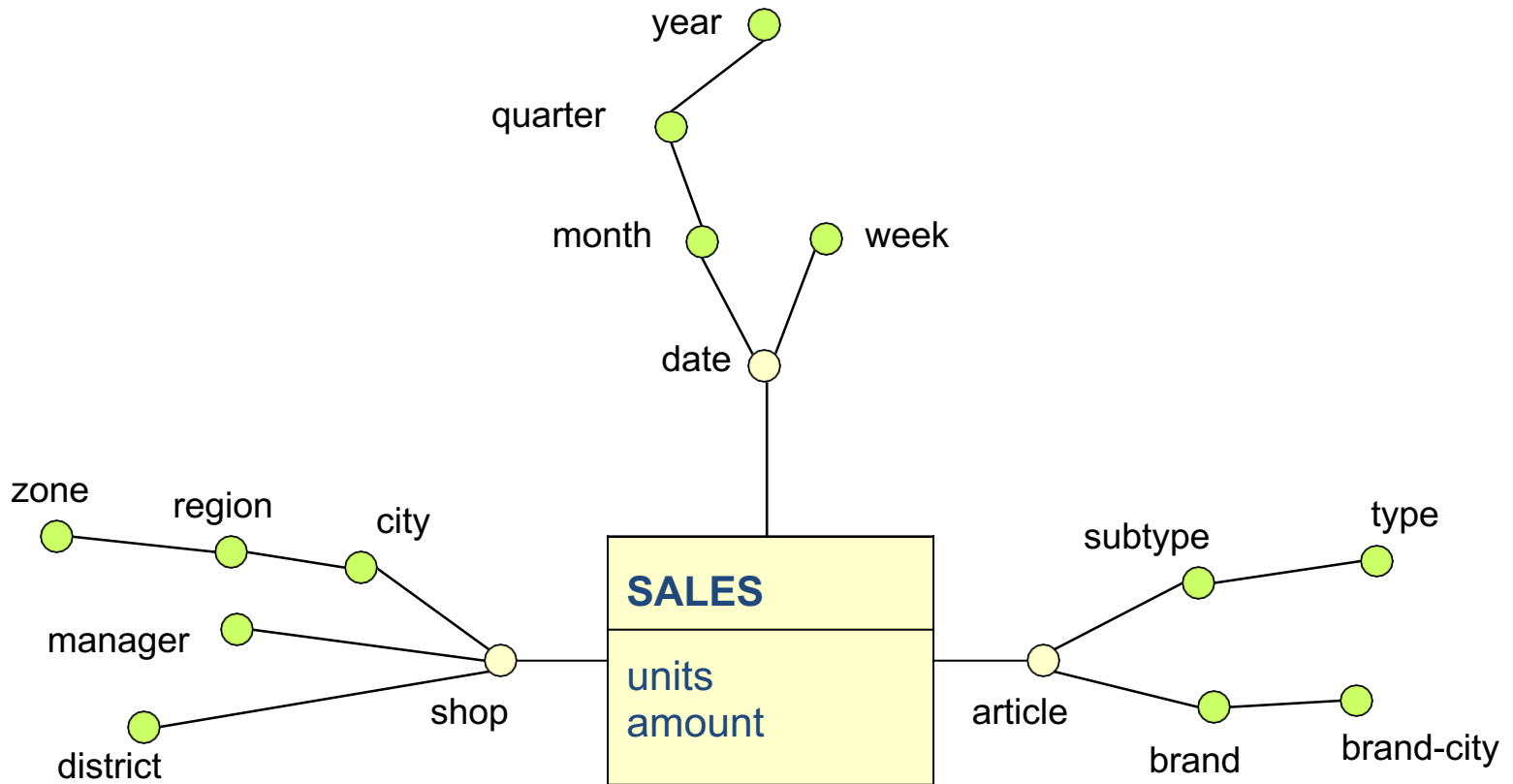


# Dimensional Hierarchies

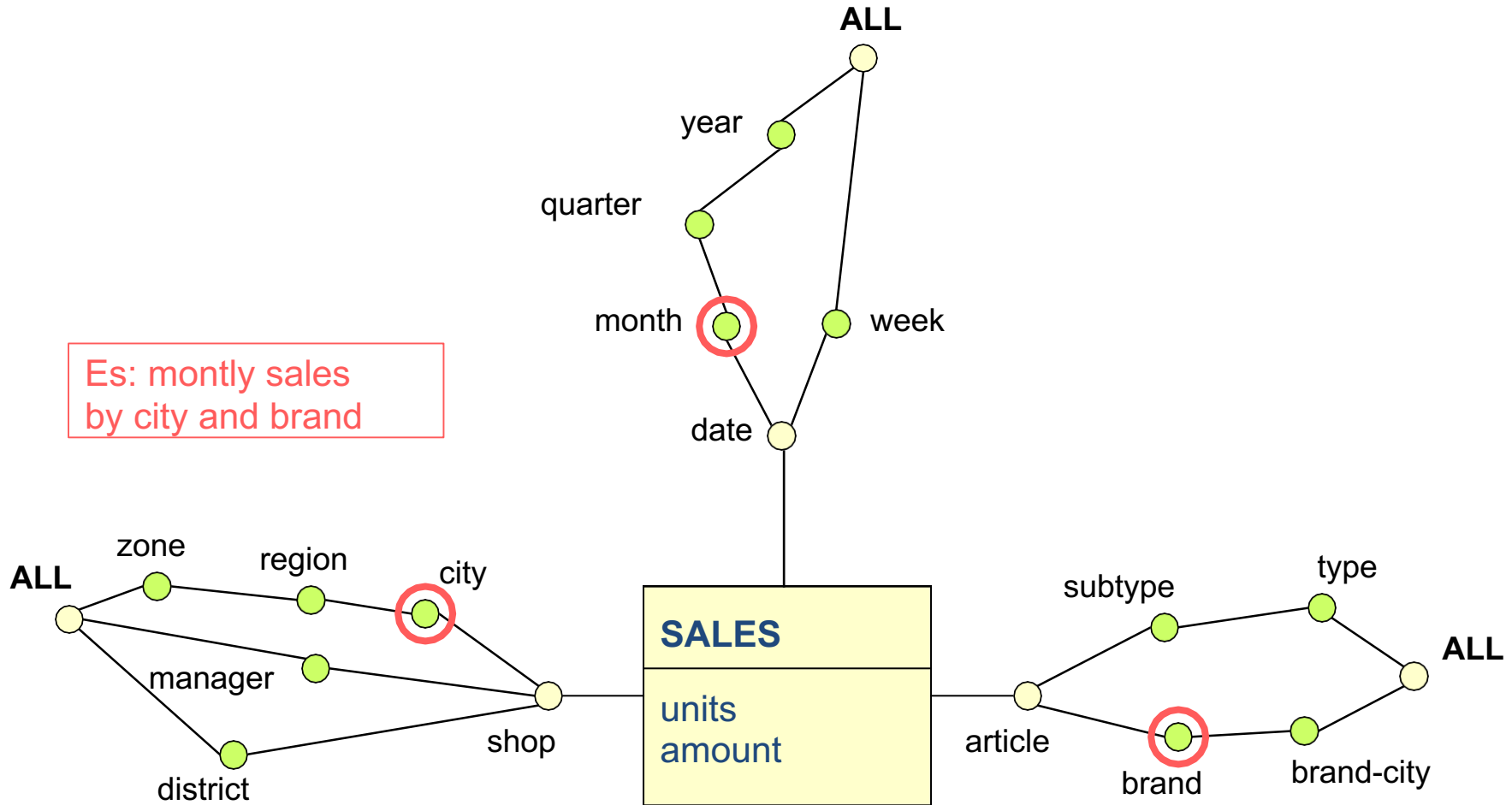


**details based on user requirements**

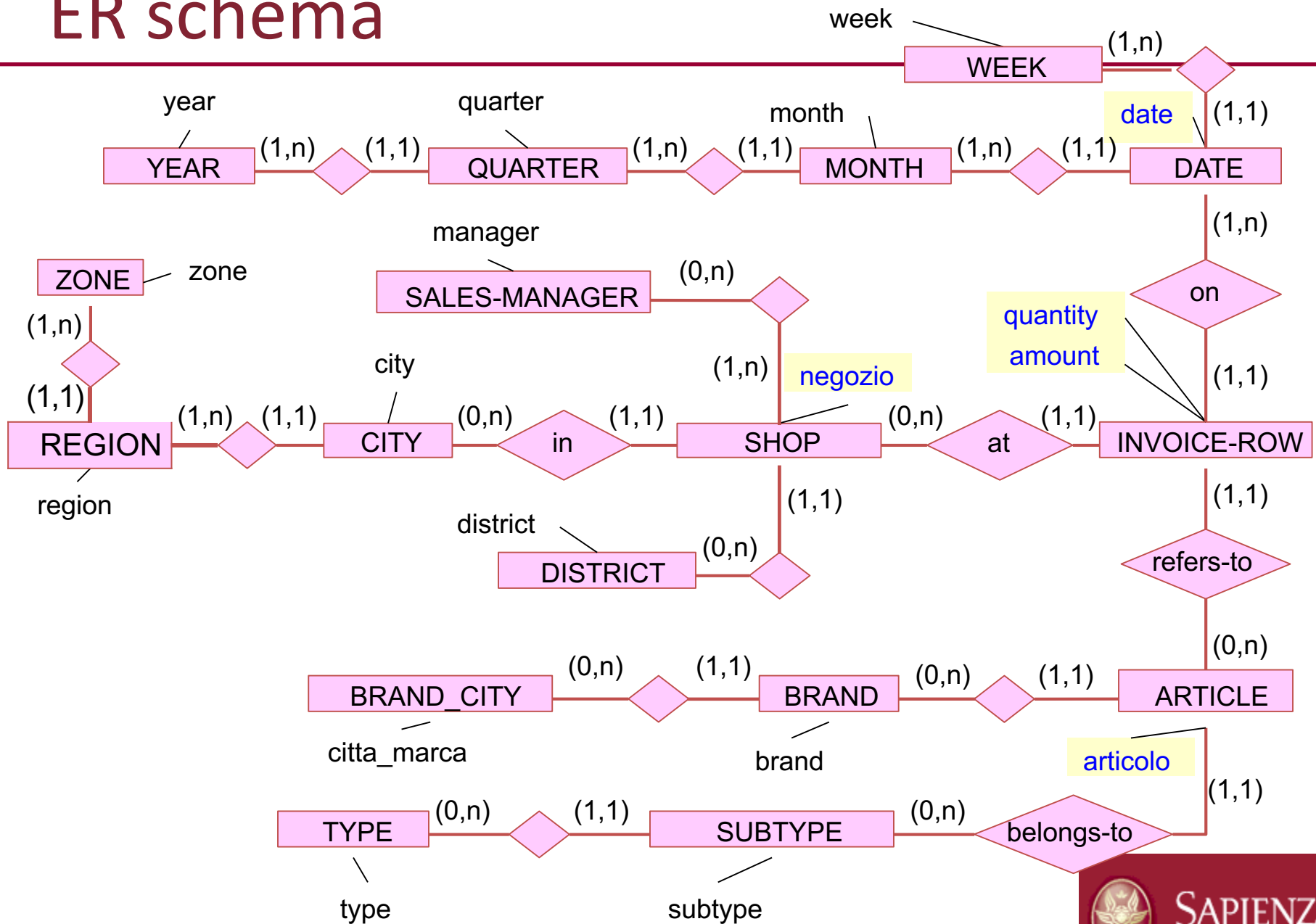
# Fact Schema: DFM (Dimensional Fact Model)



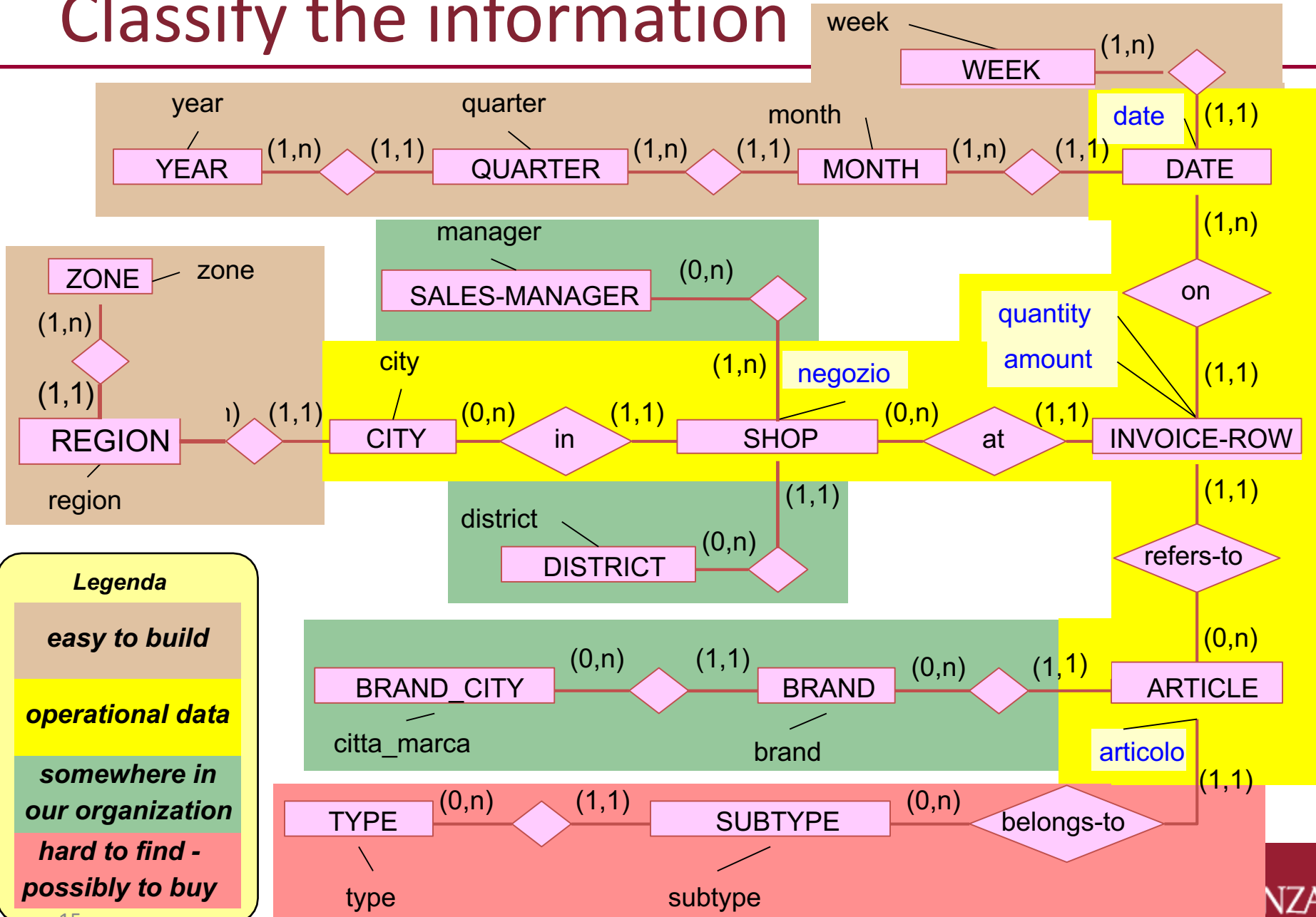
# Fact Schema (an interpretation for OLAP)



# ER schema



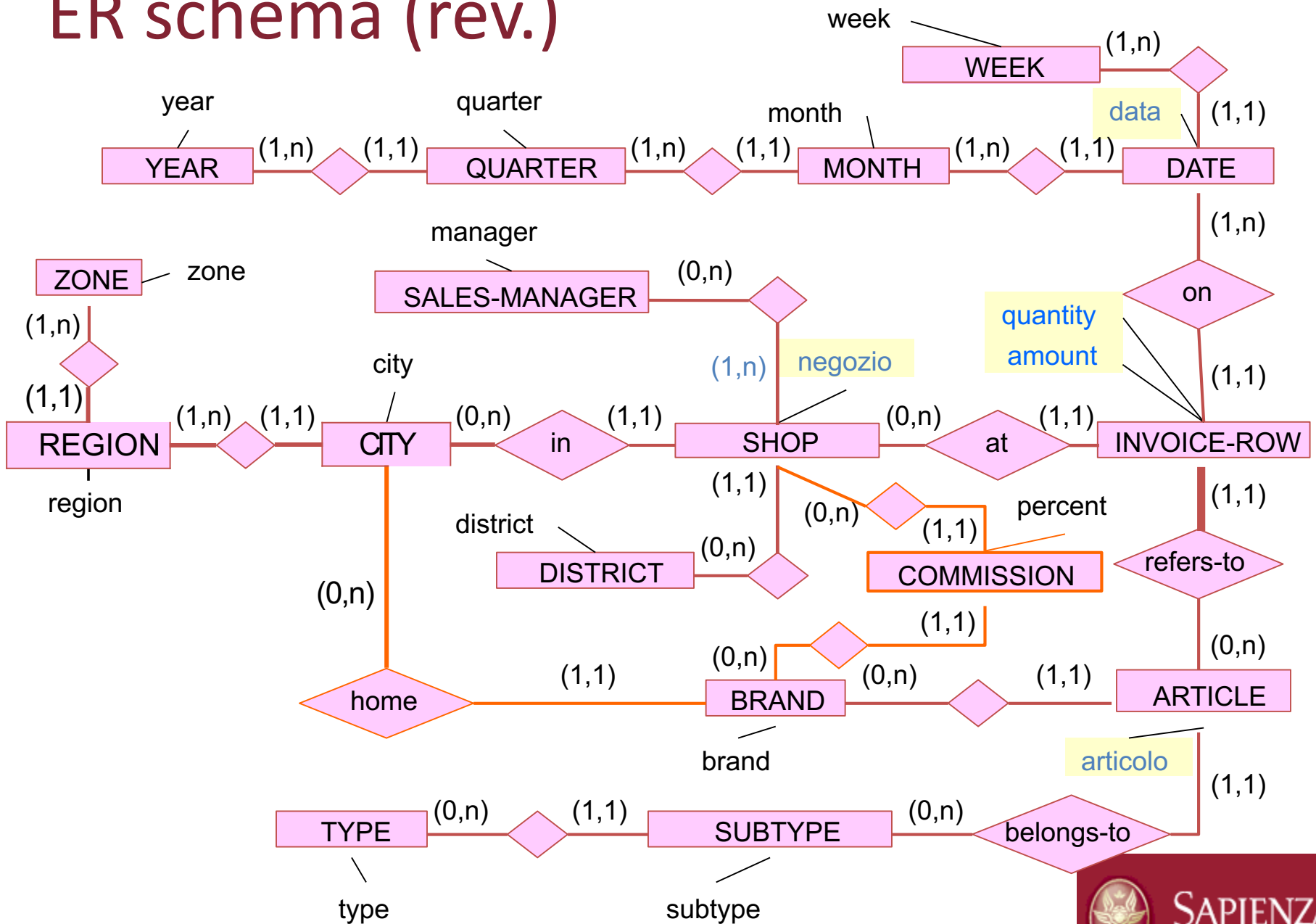
# Classify the information



**Legenda**

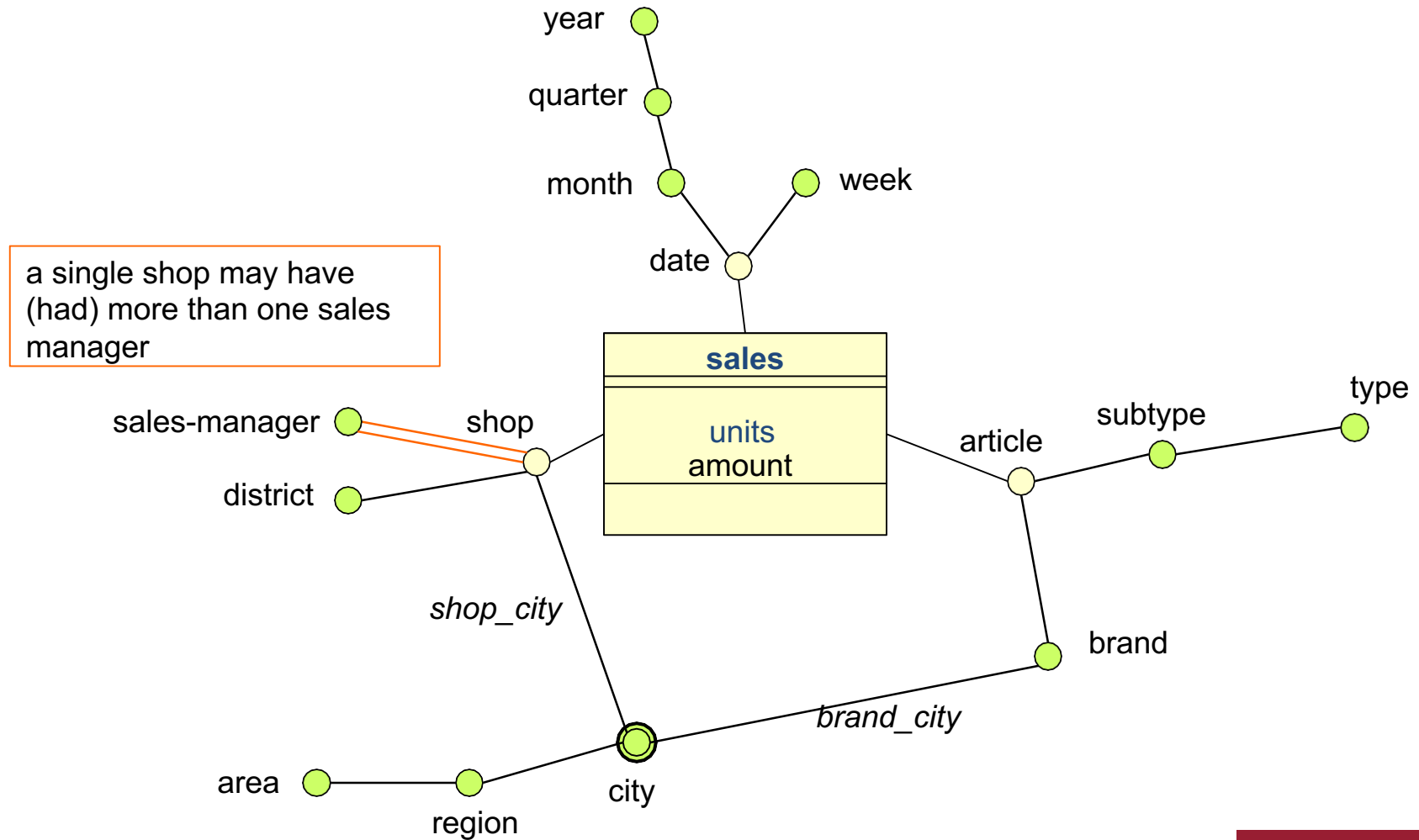
- easy to build
- operational data
- somewhere in our organization
- hard to find - possibly to buy

# ER schema (rev.)





# Multiple arcs (n:n relation)



# Cross-dimensional Attributes

