Armazéns de Dados

Departamento de Engenharia Informática (DEI/ISEP)
Paulo Oliveira
pjo@isep.ipp.pt

1

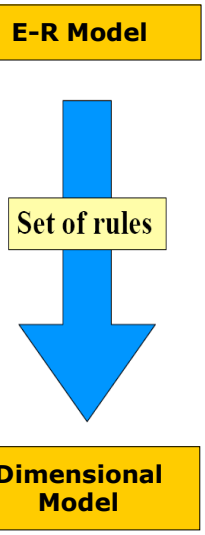
From Relational Models to Dimensional Models

Adapted from:

Moody, D. and Kortink, M (2000), "From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design". In *Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2000)*. Stockholm, Sweden, June 5-6.

2

Goal



3

3

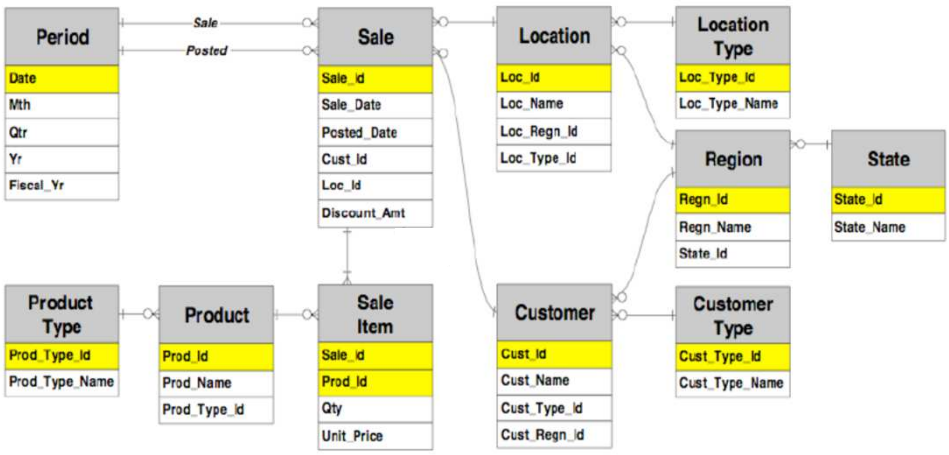
Method

- Step 1
  - Classify the entities of E-R model
- Step 2
  - Identify hierarchies
- Step 3
  - Produce the dimensional model

4

4

E-R Model Example



5

Types of Entities

- Entities of E-R model **classified into one of the following categories**:
  - Transaction Entities
  - Component Entities
  - Classification Entities

6

6

Transaction Entities

- Transaction Entities: **the business events**
  - Record **details about events** that occur at a point in time
  - Contain **measurements or quantities that may be analyzed** (e.g.: monetary amounts; weights; volumes; etc.)
  - **Base for fact tables**
- Examples: Orders; insurance claims; salary payments; hotel bookings
- Events that **decision makers want to analyze**
- Not all transaction entities will be of interest for decision support

7

7

Component Entities

- One which is **directly related to a transaction entity** via a one-to-many relationship
- Define the **details** for each transaction
  - Answer the "who", "what", "when", "where" and "why" of a business event
  - Sales transaction may be defined by several components:
    - Customer: who made the purchase
    - Product: what was sold
    - Location: where it was sold
    - Period: when it was sold
- Form the **basis for constructing dimension tables** in star-schemas

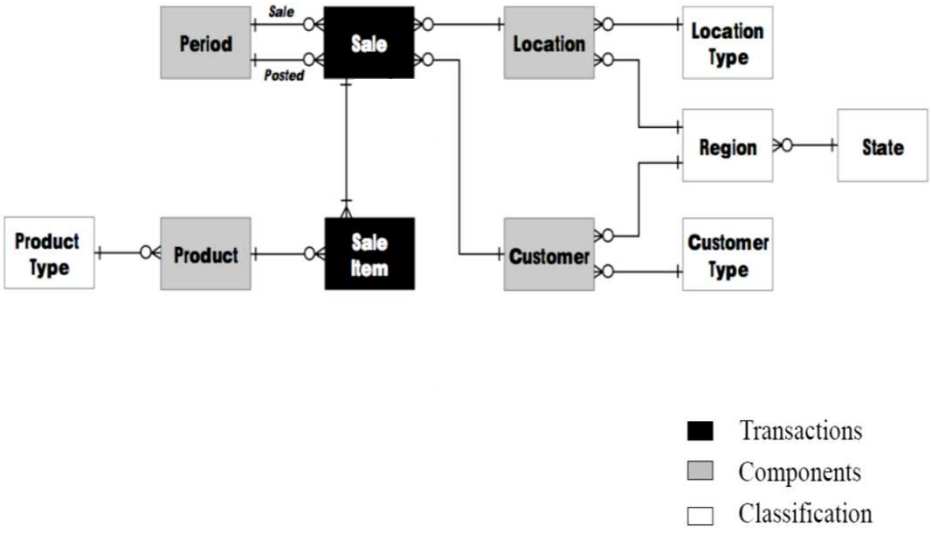
8

8

Classification Entities

- Entities which are **related to component entities by a chain of one-to-many relationships**
- **Functionally dependent on a component entity** (directly or transitively)
- Represent **hierarchies embedded in the data model**, which may be collapsed into component entities to form dimension tables

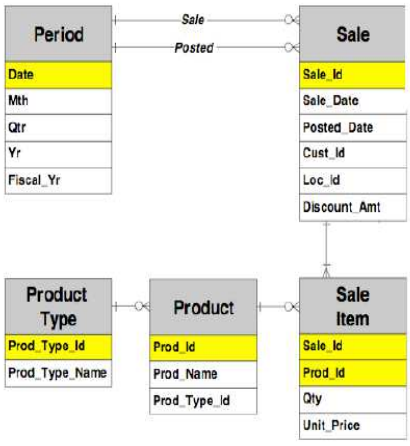
Classification of Entities



Resolving Ambiguities

- In some cases, **entities may fit into multiple categories**
- **Precedence hierarchy** for resolving such ambiguities
  - 1. Transaction entity (highest precedence)
  - 2. Classification entity
  - 3. Component entity (lowest precedence)

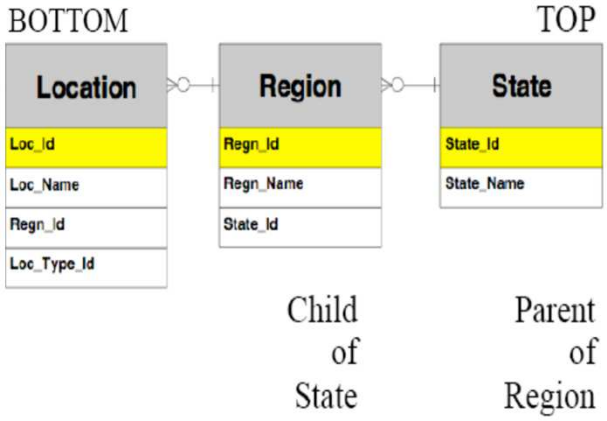
Resolving Ambiguities



- Sale item is considered as a ***Transaction Entity***
- Since Sale is directly related to Sale Item by a 1:N relationship, **Sale can be considered as a *Component Entity*** (of Sale Item)
- Considering the rules for resolving ambiguities, **Sale is classified as a *Transaction Entity***

Hierarchies

- Hierarchy in an E-R model is any **sequence of entities joined together by one-to-many relationships**, all aligned in the same direction



13

13

Production of Dimensional Models

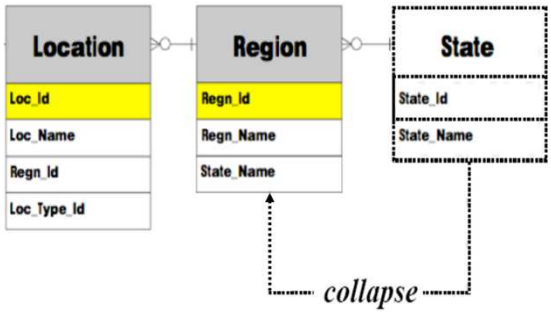
- Two operators to produce dimensional models from E-R models:
  - **Collapse Hierarchy**
    - Higher-level entities can be "collapsed" into lower-level entities within hierarchies
  - **Aggregation**
    - Can be applied to a transaction entity to create a new entity containing summarized data

14

14

Collapse Hierarchy

- **Higher level entities can be "collapsed" into lower-level entities** within hierarchies



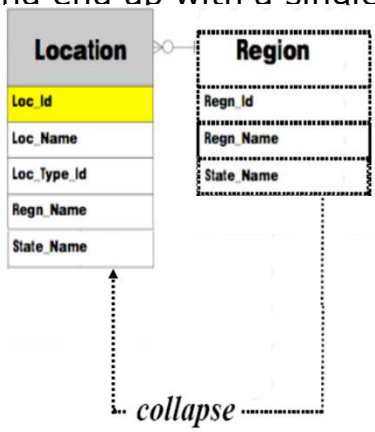
- State entity is being collapsed into the Region entity
- Region entity contains its original attributes plus the attributes of the collapsed table, except the key
- Is therefore a **form of denormalization**

15

15

Collapse Hierarchy

- Continue doing this until the bottom of the hierarchy is reached, and end up with a single (dimension) table



- Number of rows/records remains the same before the operation (i.e., the number of rows of table Location)

16

16

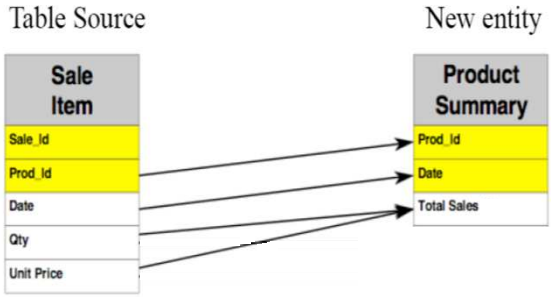
Aggregation

- Can be applied to a **transaction entity to create a new entity containing summarized data**
- Aggregation attributes **must be numerical**
- Another **subset of attributes chosen to aggregate by** (the grouping attributes)

17

Aggregation

- Aggregated entity shows for each product the total sales amount (quantity*price) on a daily basis



- Aggregation attributes - Qty, Unit Price
- Grouping attributes – Prod_ID, Date

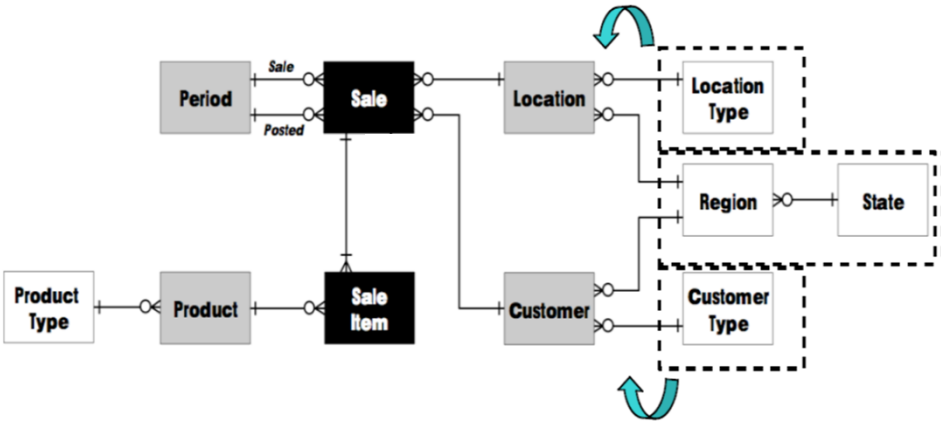
18

Star Schema

- Dimension table is formed for the component entity, by **collapsing hierarchically related classification entities into it**
- Hierarchies provide the **ability to "drill down" between transaction levels**
- **Numerical attributes within transaction entities can be aggregated by key attributes**

19

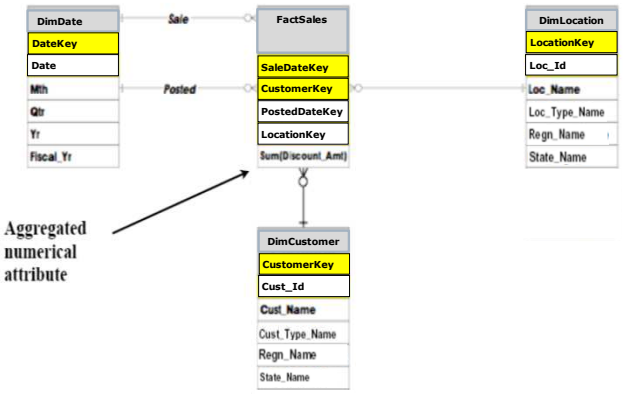
Derivation of Sale Star Schema



20

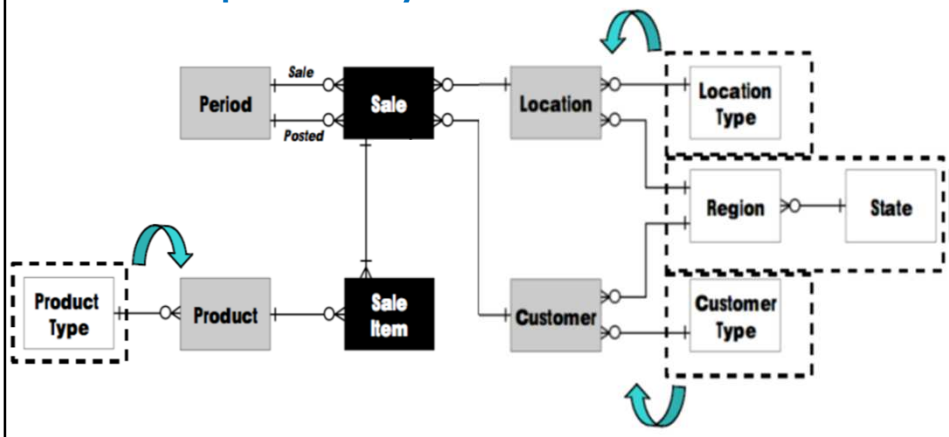
Derivation of Sale Star Schema

- Star schema has three dimensions, each of which contains embedded hierarchies



Derivation of Sale Item Star Schema

- Where hierarchical relationships exist between transaction entities, the **child entity inherits all the key attributes from the parent entity**



Derivation of Sale Item Star Schema

- This star schema has four dimensions, including three dimensions from its "parent" transaction entity (Sale)