3. Data Warehouse logical design

Based on the relation model

Inputs

- **conceptual fact schema**
- **workload**
- **data volume**
- **system constraints**

Outputs

Relation logical schema

Star schema

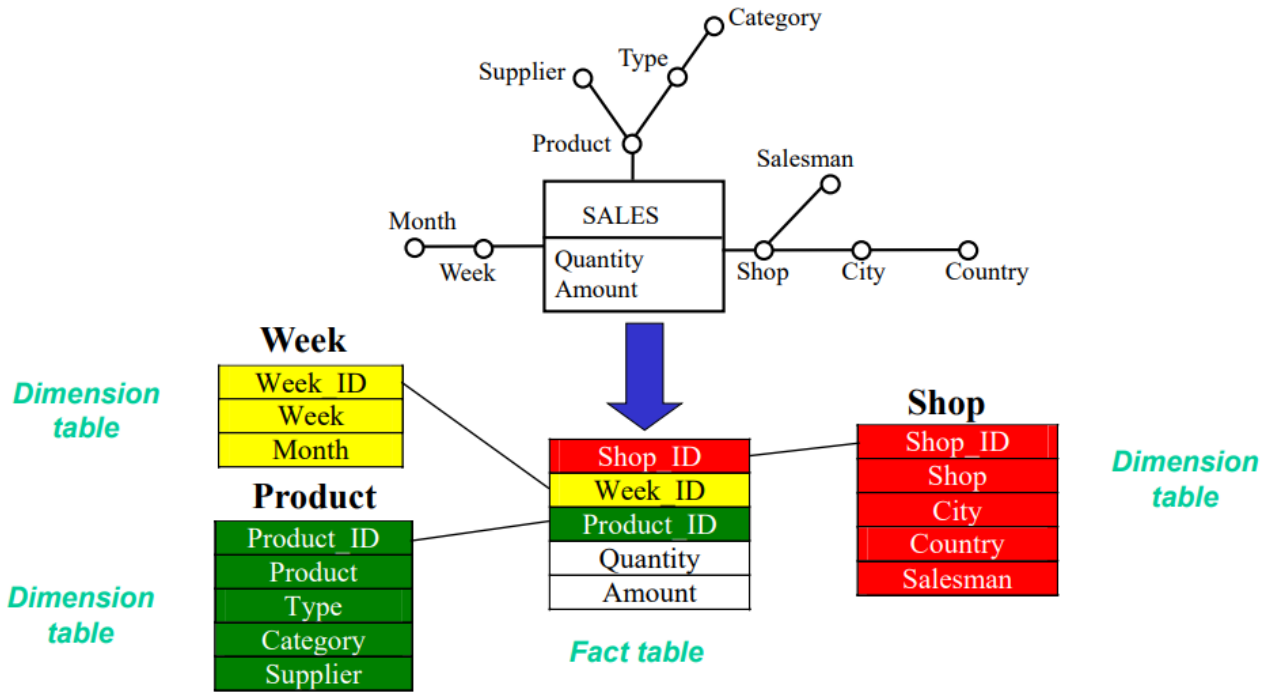
Dimensions

- One table for each schema
- A surrogate(generated) primary key
- It contains all dimensions attributes
- Not normalized since there may be a functional dependency from an attribute to another attribute that's not a primary key(violates the BCNF).

Facts

- One fact table for each fact schema
- Primary key composed by foreign keys of all dimensions

- Measures are the fact schema attributes

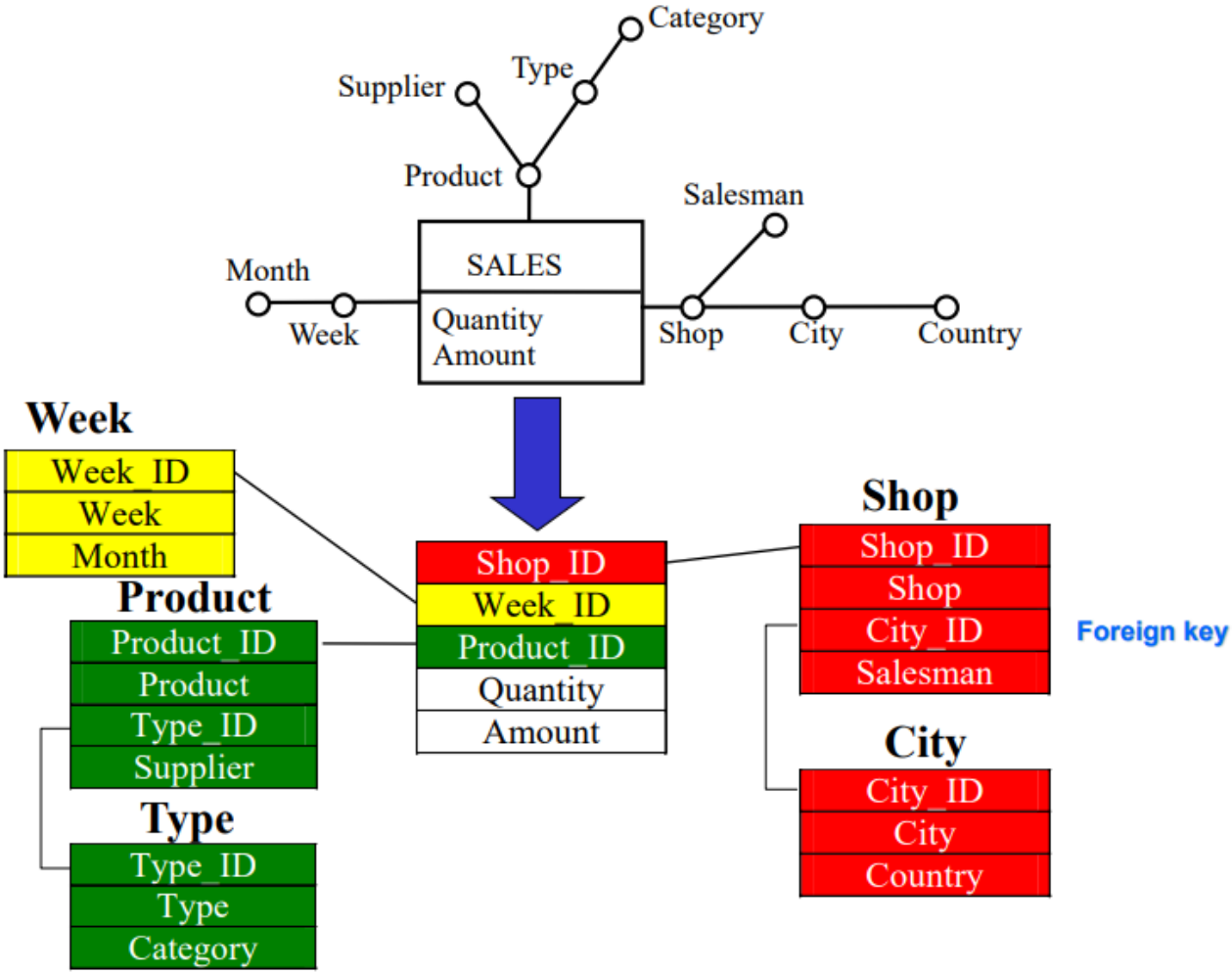


The fact table is kinda like a cross-table on a typical ER

Snowflake Schema

Certain functional dependecies can be broken down to multiple tables. This way, data redundancy is reduced.
However, this reduced redundancy comes at a price: joins are needed in order to retrieve data and joins might be very expensive, especially when used with a fact table.

Thus, a snowflake schema is rarely used.

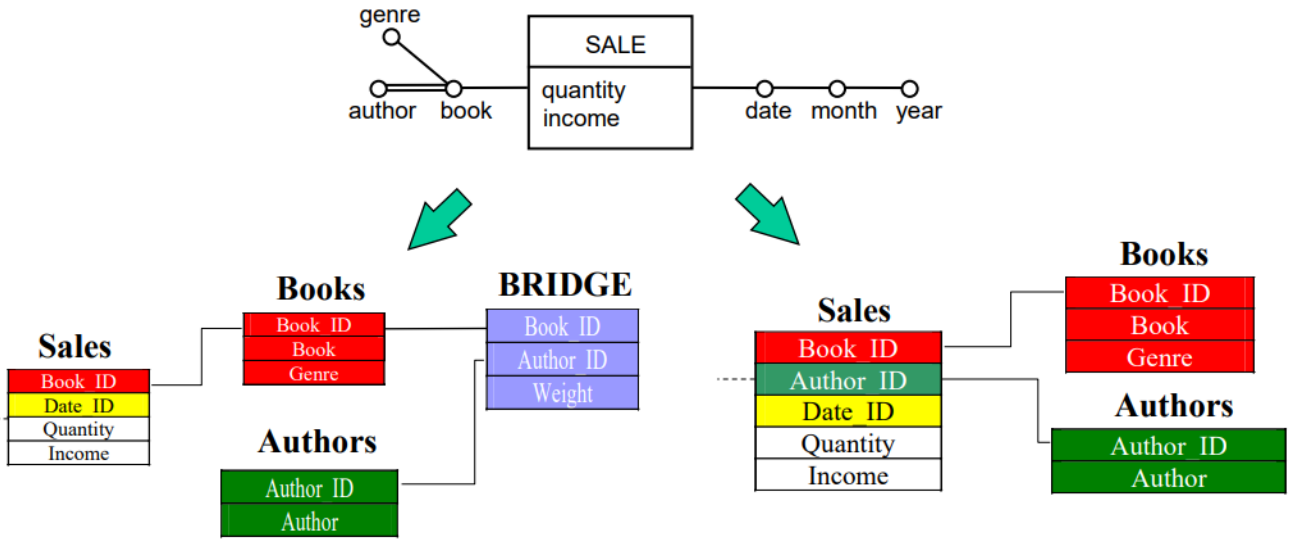


Multiple edges

Multiple techniques:

- **Bridge table:** a new table which models N:N relationship and a new attribute that represents the weight(e.g. how much an author contributed to a book) of tuples in the relationship. It's the tipically adopted solution
- **Push down:** multiple edge fully integrated in the fact table. A new dimension is thus created. The weight is coded in the fact table.

Less joins so query execution is faster however data redundancy is introduced.



Queries with multiple edges

- **Weighted query:** takes into account the weight of each tuple
- **Impact query:** does not take into account the weight. Harder to perform with a pushdown approach.

Degenerate dimensions

Dimensions with only 1 attribute.
They can be integrated into the fact table as its attributes; feasible only with small size attributes.
Otherwise, a **junk dimension** can be created, containing several degenerate dimension. There must no function dependency between attributes of a junk dimension. Feasible only when attributes cardinality is low.