E5.2

Star vs. Snowflake

DWH Presentation

L. Katzmaier & M. Marhofen

Structure

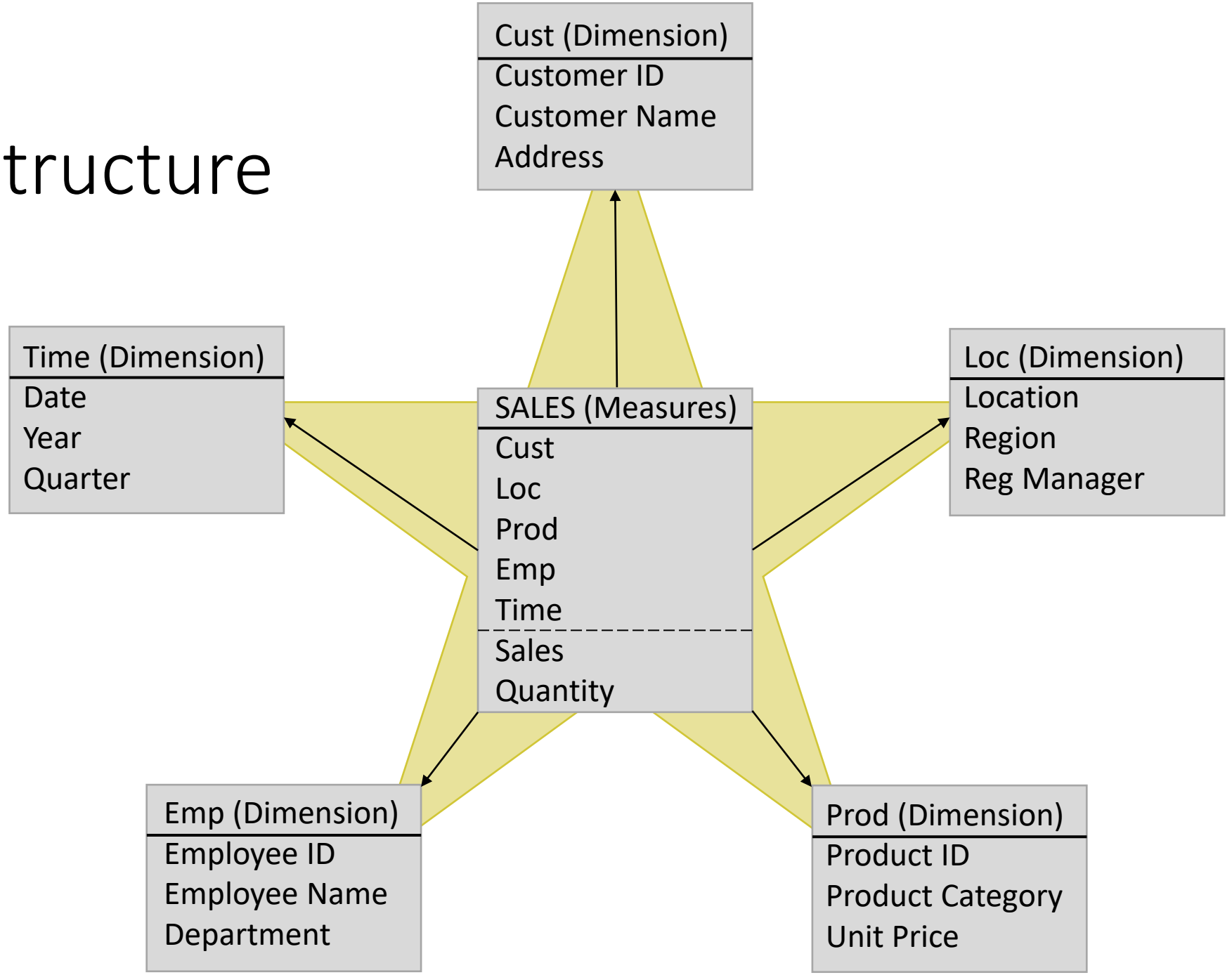
- STAR
  - Model
  - Structure
  - Join
  - Advantages & Disadvantages
- SNOWFLAKE
  - Model
  - Structure
  - Join
  - Advantages & Disadvantages
- COMPARISON

Star Schema

Star – Model

- Application area
  - multidimensional data structures in relational databases
  - Analytical applications
- Attempt to minimize number of tables
- Measures: express important relationships in a quantitatively measurable and condensed form
- Dimensions: enable different views of the measures

Star – Structure

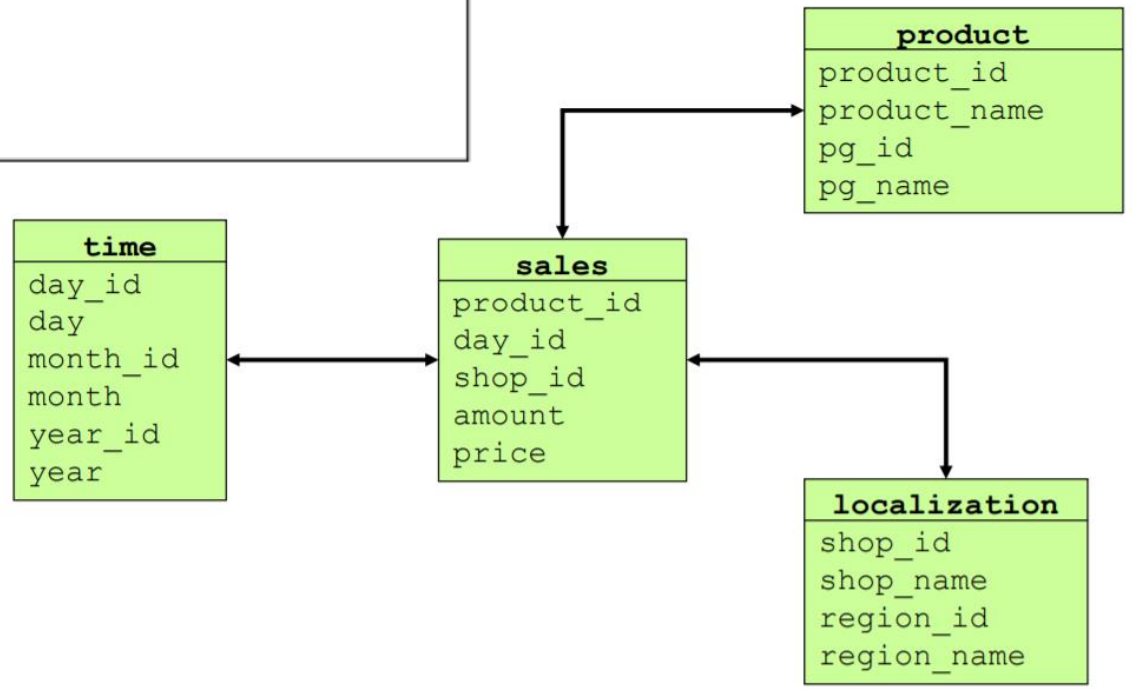


Star – Join

Source of the example: https://www.informatik.hu-berlin.de/de/forschung/gebiete/wbi/teaching/archive/ws_1718/vl_dwhdm/05_rolap_molap.pdf (page 11 and 14)

```
SELECT    L.shop_name, T.year, sum(amount*price)
FROM      sales S, product P, time T, localization L
WHERE     P.pg_name=„Wasser“ AND
          P.product_id = S.product_id AND
          T.day_id = S.day_id AND
          L.shop_id = S.shop_id
group by  L.shop_name, T.year
```

- 3 Joins
- Number of joins independent of aggregation paths in the request
- Number of joins increases linearly with the number of dimensions in the request.



Star – Advantages & Disadvantages

- Advantages
  - Intuitive data model
  - Simpler queries
  - Simplified Business Reporting Logic
  - Optimizes navigation
- Disadvantages
  - Poor response behavior with large dimension tables
  - Lack of flexibility
  - Data integrity
  - Large number of redundancies

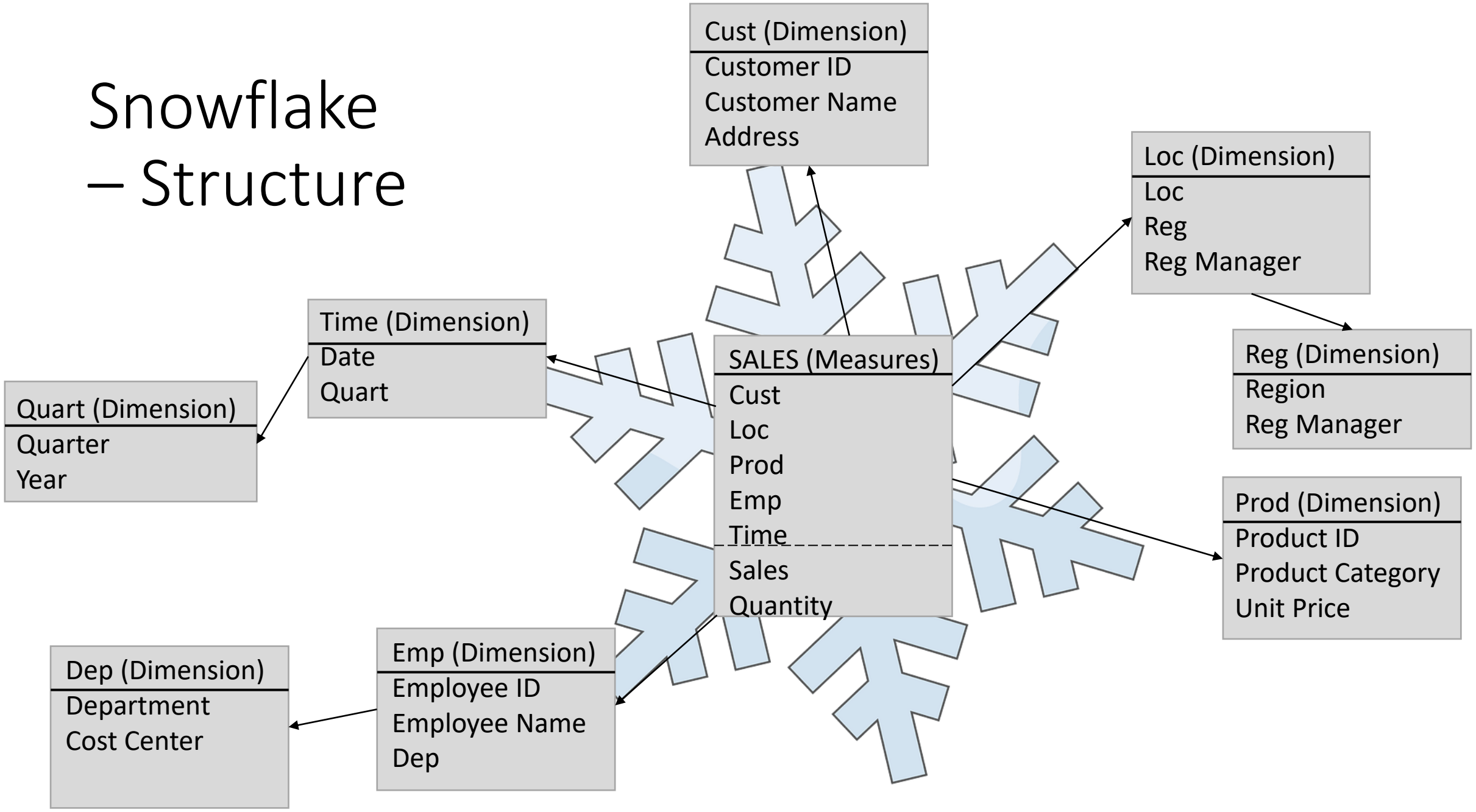
Snowflake – Schema

Snowflake – Model

- Normalized version of star schema
  - Fact tables remain
  - Storing of dimensions in normal form

- Use cases:
  - Large amount of data
  - Many users with different scenarios

Snowflake – Structure



Snowflake – Join

- Relations are defined with foreign keys
- Often many join-operations are necessary
- Joins put more load on the performance
- Advantage due to non-redundancy might be lost

Snowflake – Advantages & Disadvantages

- Advantages
  - More structured data
  - Less redundancy, improved integrity
  - Needs less storage
  - Easier maintenance
  - Performance improvements due to less data being retrieved
- Disadvantages
  - Harder to design
  - More complex join-operations
  - More joins result in worse performance
  - Requires higher skills

Comparison

	STAR	SNOWFLAKE
Database design	Simple	Complex
Hierarchies	Stores the hierarchies for the dimensions	Separated into different tables
Joins	Single join to create relation between dimension and fact table	Many joins required to fetch data
Dimension tables	Dimension tables surround fact tables	Every fact table surrounded by a dimension table, which is surrounded by more dimension tables
Data Redundancy	High-level	Low-level
Cube Processing	Fast	Might be slower (due to complex join)
Data in dimension tables	Single one contains aggregated data	Data split into different tables
Amount of retrieved data in query	All data is retrieved	Only the requested tables are retrieved

Thank you for your Attention

DWH Presentation

L. Katzmaier & M. Marhofen

Sources

- https://www.datenbanken-verstehen.de/data-warehouse/data-warehouse-grundlagen/data-warehouse-datenmodell/star-schema/
- https://www.tecchannel.de/a/business-intelligence-teil-3-datenmodellierung-relationale-und-multidimensionale-modelle,1744994,10
- https://www.geeksforgeeks.org/star-schema-in-data-warehouse-modeling/
- https://hackr.io/blog/star-schema-in-data-warehouse
- https://www.ibm.com/support/knowledgecenter/en/SSEPEK_11.0.0/perf/src/tpc/db2z_starschemaaccess.html
- https://www.informatik.hu-berlin.de/de/forschung/gebiete/wbi/teaching/archive/ws_1718/vl_dwhdm/05_rolap_molap.pdf
- https://dbs.uni-leipzig.de/file/DWH_WS0910_Kap3_Modellierung.pdf
- https://www.vertabelo.com/blog/data-warehouse-modeling-the-snowflake-schema/?utm_source=dbweekly&utm_medium=email
- https://www.upsolver.com/blog/difference-between-star-schema-and-snowflake-schema
- Buch datenmodellierung (isbn: 3-89842-535-5)
- Sein Skript (S. 120)