

Data Marts
Chapter 12

Class 10: Data Marts 1

Data Marts

- *Data Mart*: a subset of a data warehouse that supports the requirements of a particular department or business function.
 - *Limited scope*
 - *Not intended for operational reporting*
 - *Must less information than a data warehouse*

Class 10: Data Marts 2

Reasons for Creating a Data Mart

- Data tailored to department or function
- Lower cost than a full DW
- Lower risk project than a full DW
- Limited (usually 1) end user analysis tool
- Database placed physically near the department, reducing network delays

Class 10: Data Marts 3

Data Mart Approaches

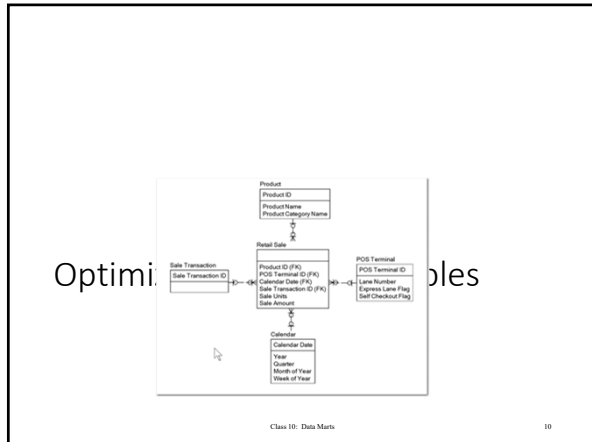
- Build enterprise DW to populate data marts
 - Data marts won't be done if DW project stalls
- Build several data marts and integrate later
 - Generally lower risk
 - Data marts may produce inconsistent results
 - Overall cost may be higher due to integration
- Build DW and data marts simultaneously
 - Practically guarantees a never-ending project from hell

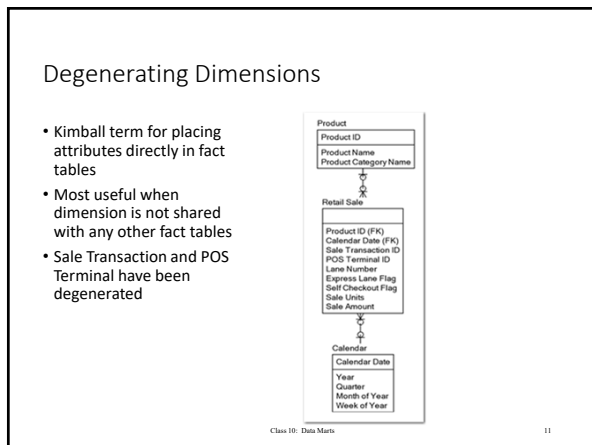
Multi-Dimensional OLAP

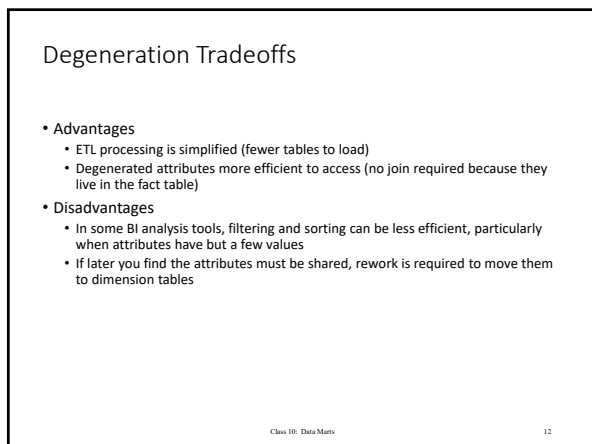
- Use multi-dimensional structures to store data and relationships with data.
- Facts are pre-summarized by each dimension pair
- Best visualized as cubes of data with cubes within cubes
- Each side of cube is a dimension
- Support for Analytical Operations:
 - Consolidation (aggregation of data)
 - Drill-down (reverse of aggregation)
 - Slicing and dicing (pivoting): look at data from different viewpoints

Commercial MOLAP Products

- Oracle Express
- Microsoft SQL Server Analysis Services
- IBM Cognos
- IcCube
- Palo
- Oracle Essbase





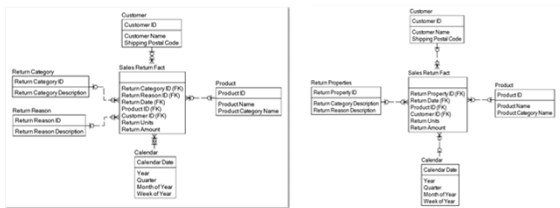


Junk Dimensions

- **Junk Dimension:** Kimball's term for combining dimensions with few attribute values into a single dimension.
- Rows formed as a Cartesian product of attribute values (rows represent every possible combination of data values)
- Goal is to reduce joins with fact table

Junk Dimensions

- Return Category and Return Reason (left) combined into Return Properties junk dimension (right)



Junk Dimension Tradeoffs

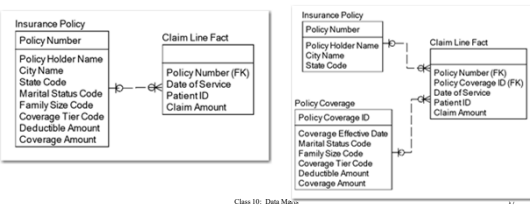
- **Advantages**
 - Fewer dimensions to load
 - Fewer joins between fact and dimensions
 - Related codes, flags and indicators can be bundled, making them more understandable in context
- **Disadvantages**
 - If you find later that you need to use the dimensions separately, it takes rework to redefine and reload them
 - Adds one step to ETL process (to combine the data into a Cartesian product)

Mini Dimensions

- Mini Dimension: Kimball's term for a dimension split from a larger one to accommodate:
 - Attributes that must be Type 2 (while others remain Type 1)
 - Dimension tables with too many wide columns (which can affect performance)
- Kimball uses the term Type 4 for Dimensions that include both Type 1 and Type 2 attributes

Mini Dimension Example

- Six attributes from Insurance Policy dimension (left) split to Policy Coverage dimension (right)
- These attributes tend to change when life events occur



Mini Dimension Tradeoffs

- Advantages
 - Ability to track history for slowing changing dimensions (Type 2) without including Type 1 attributes
 - Potentially simpler history tracking, especially if mini dimension can be changed to a junk dimension
- Disadvantages
 - Additional join when data from both the original and the mini dimension
 - If some attributes change more rapidly, you may later have to split the mini dimension further (see Kimball Group articles on Type 4, 5, 6 and 7 dimensions).

Bridge Tables

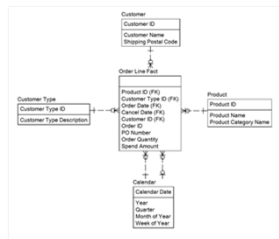
- Many-to-many relationships cannot be directly implemented in star schemas
- There are two possible alternatives:
 - Implementing a fact table at the many-to-many grain
 - Using an intersection table, known as a **bridge table**, between two dimension tables (or a fact table and a dimension table)

Class 10: Data Marts

19

Many-to-Many Example

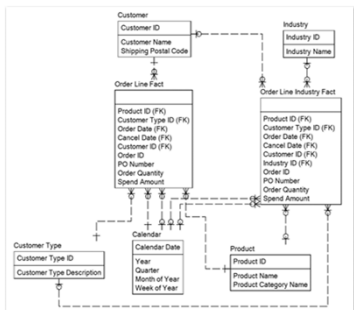
- Business wishes to add Customer Industry dimension
- Some customers are in many industries



Class 10: Data Marts

20

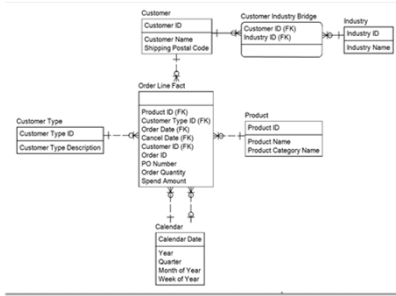
Alternative: Add Fact Table



Class 10: Data Marts

21

Alternative: Add Bridge Table



Class 10: Data Marts

22

Bridge Table Tradeoffs

- Advantages:
 - Better support for many-to-many between dimension tables
 - Somewhat more intuitive
- Disadvantages
 - Queries more complex because of need to join "through" the bridge table.
 - In some tools, you must hand code the SQL for these joins
 - If the bridge table has too many rows, performance can suffer

Class 10: Data Marts

23

Fact Table Optimization

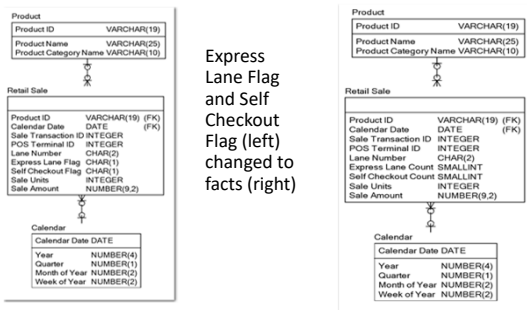
Class 10: Data Marts

24

Transform Attributes to Facts

- Attributes Which Are Candidates for Facts:
 - Few possible values
 - Used more often for counting fact rows
 - Seldom/never used for filtering or sorting fact rows
- For example:
 - Y/N flag can be turned into a counter set to 1 for "yes" or 0 from "no"
 - Gender Code can be turned into two counters: Male Count and Female Count with one set to 1 and the other set to 0 for any given fact table row

Attribute to Fact Example

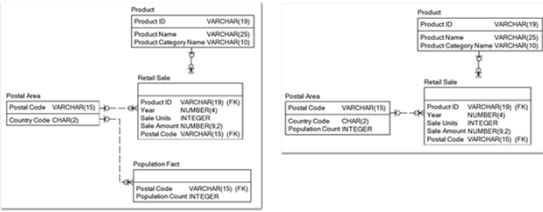


Lifted Facts

- When the grain of a fact table is restricted to a single dimension, the fact(s) can be "lifted" into the dimension
- Advantage: Reduces join to obtain the fact data (it lives in the dimension)
- Disadvantage: Only applies to facts that have a single dimension (a rare circumstance)

Lifted Fact Example

Population Count lifted from Population Fact (left) to Postal Area dimension (right)



Class 10: Data Marts

28

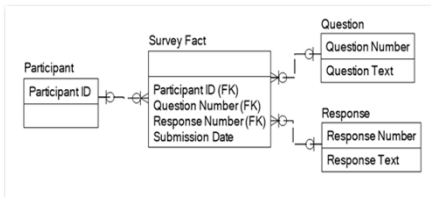
Factless Fact Tables

- In relatively rare situations, we need to record events that have no obvious fact(s) to record.
- Example: Survey where each question has the same set of responses (1 = Strongly Disagree, 2 = Somewhat Disagree, ... 5 = Strongly Agree)
 - Recording the number of the response is meaningless because summing them provides nothing useful
 - However, we do need to know what the responses were so we can count how many answered one way versus others
- Alternative is to add a dummy fact that is always 1
 - Most bridge tables can be turned into dummy fact tables

Class 10: Data Marts

29

Factless Fact Table Example



Class 10: Data Marts

30
