



DW Design Methodology

- Database design methodology for data warehouses
 - Nine steps methodology



Designing Data Warehouses

- For many enterprises the way to avoid the complexities associated with designing a data warehouse is to start by building one or more data marts.
- Data marts allow designers to build something that is far simpler and achievable for a specific group of users.



Database Design Methodology for Data Warehouses

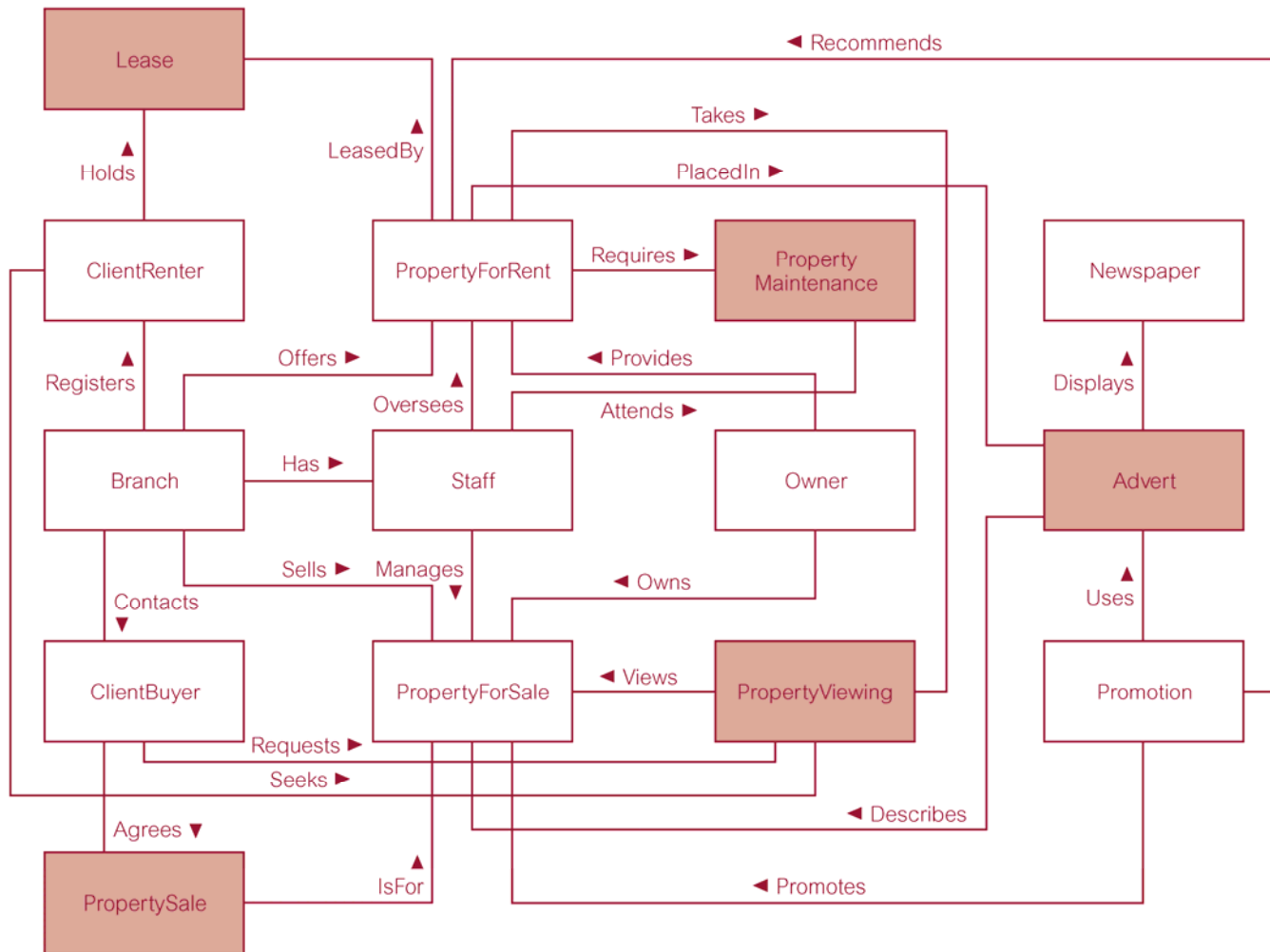
- **‘Nine-Step Methodology’ includes following steps:**
 - Choosing the process
 - Choosing the grain
 - Identifying and conforming the dimensions
 - Choosing the facts
 - Storing pre-calculations in the fact table
 - Rounding out the dimension tables
 - Choosing the duration of the database
 - Tracking slowly changing dimensions
 - Deciding the query priorities and the query modes



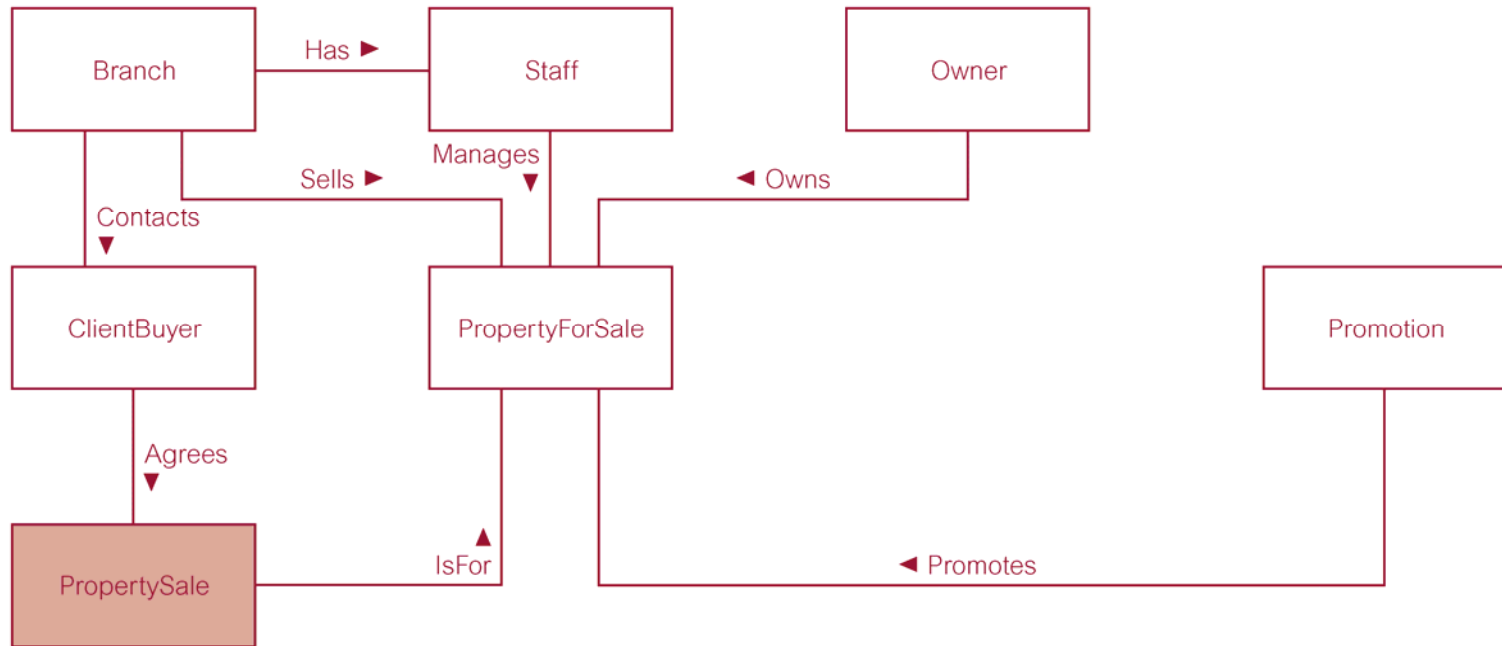
Step 1: Choosing the process

- The process (function) refers to the subject matter of a particular data mart.
 - E.g. *Lease, PropertySale, ...*
- First data mart built should be the one that is most likely to be delivered on time, within budget, and to answer the most commercially important business questions.

ER model of an extended version of *DreamHome*



ER model of **property sales business** **process** of *DreamHome*





Step 2: Choosing the grain

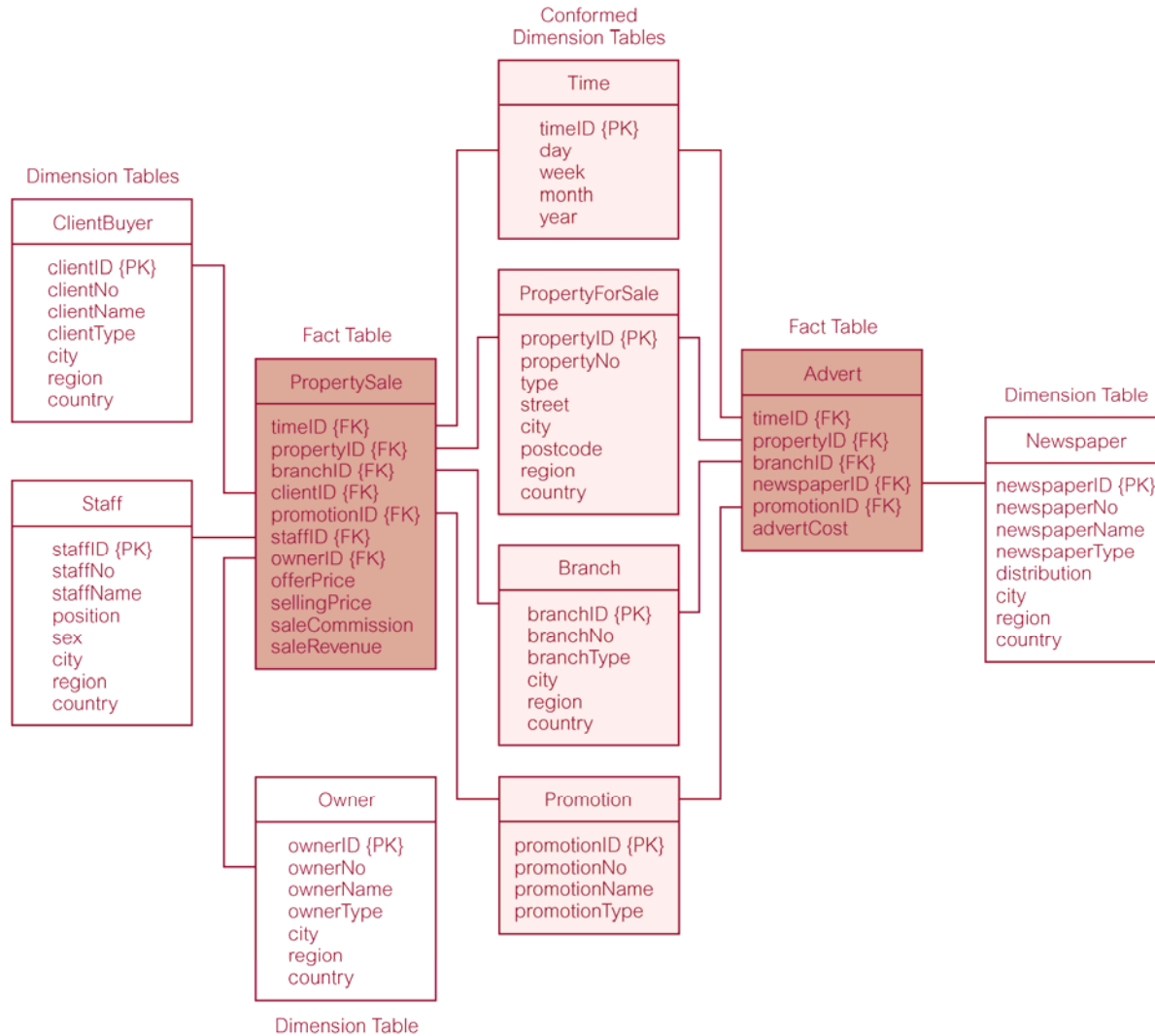
- Decide what a record of the fact table is to represent.
- Identify dimensions of the fact table. The grain decision for the fact table also determines the grain of each dimension table.
- Also **include time** as a core dimension, which is always present in star schemas.



Step 3: Identifying and conforming the dimensions

- **Dimensions set the context for asking questions about the facts in the fact table.**
- **If any dimension occurs in two data marts, they must be exactly the same dimension, or one must be a mathematical subset of the other.**
- **A dimension used in more than one data mart is referred to as being conformed.**

Star schemas for property sales and property advertising

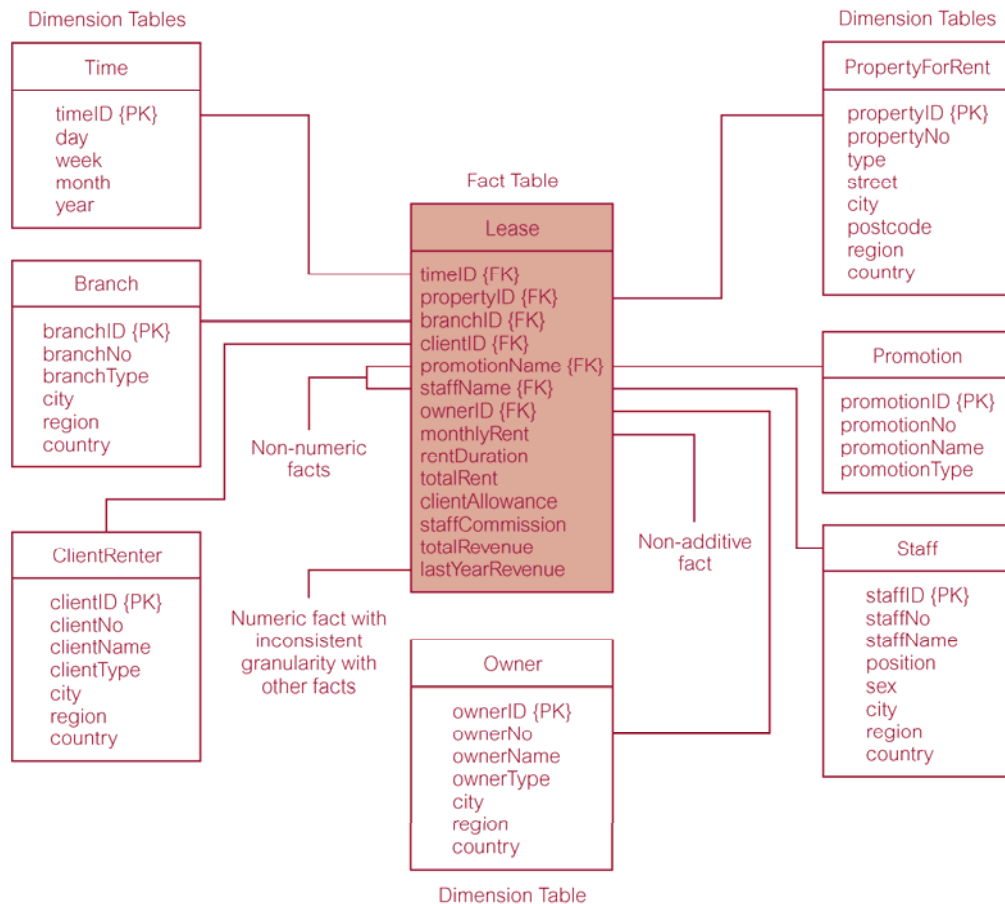




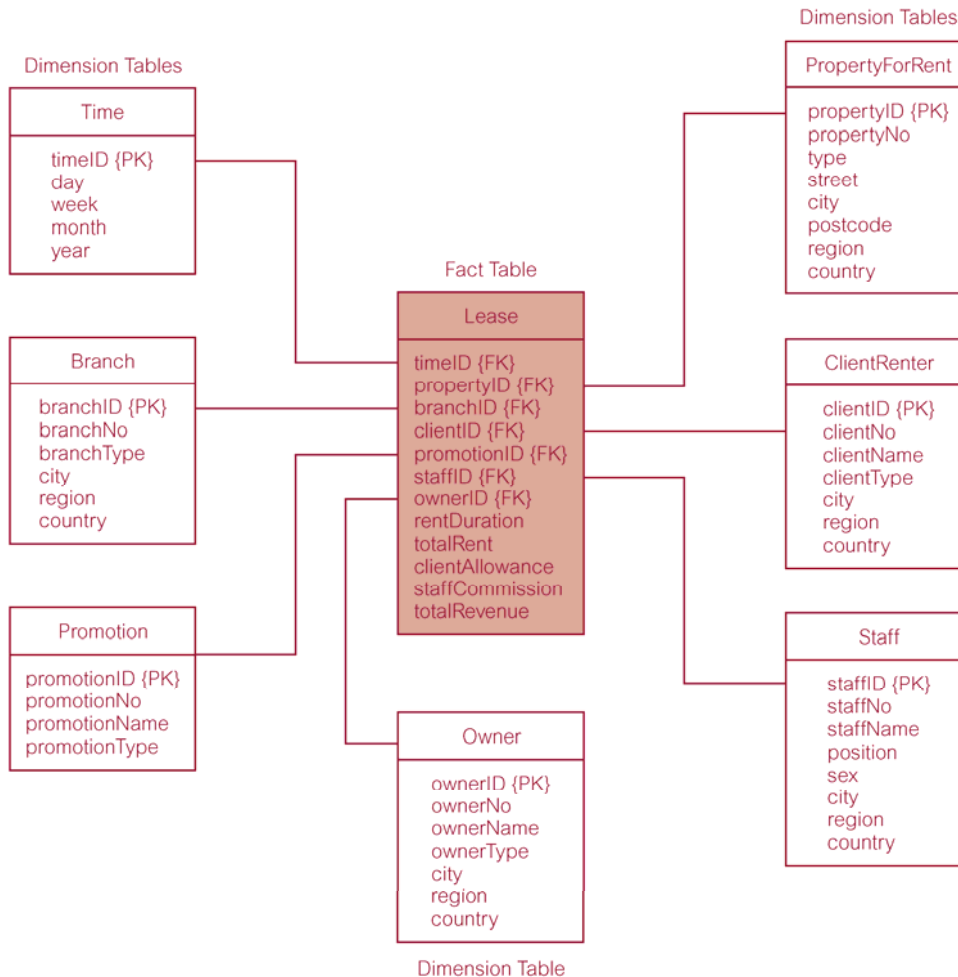
Step 4: Choosing the facts

- The grain of the fact table determines which facts can be used in the data mart.
- **Facts should be numeric and additive.**
- **Unusable facts include:**
 - non-numeric facts
 - non-additive facts
 - fact at different granularity from other facts in table

Property rentals with a badly structured fact table



Property rentals with fact table corrected





Step 5: Storing pre-calculations in the fact table

- Once the facts have been selected each should be re-examined to determine whether there are opportunities to use pre-calculations.



Step 6: Rounding out the dimension tables

- Text descriptions are added to the dimension tables.
- Text descriptions should be as intuitive and understandable to the users as possible.
- Usefulness of a data mart is determined by the scope and nature of the attributes of the dimension tables.

Step 7: Choosing the duration of the database



- Duration measures how far back in time the fact table goes.
- Very large fact tables raise at least two very significant data warehouse design issues.
 - Often difficult to source increasing old data.
 - It is mandatory that the old versions of the important dimensions be used, not the most current versions. Known as the ‘Slowly Changing Dimension’ problem.

Step 8: Tracking slowly changing dimensions



- **Slowly changing dimension problem means that the proper description of the old dimension data must be used with the old fact data.**
- **Often, a generalized key must be assigned to important dimensions in order to distinguish multiple snapshots of dimensions over a period of time.**

Step 8: Tracking slowly changing dimensions



- **There are three basic types of slowly changing dimensions:**
 - **Type 1, where a changed dimension attribute is overwritten**
 - **Type 2, where a changed dimension attribute causes a new dimension record to be created**
 - **Type 3, where a changed dimension attribute causes an alternate attribute to be created so that both the old and new values of the attribute are simultaneously accessible in the same dimension record**



Step 9: Deciding the query priorities and the query modes

- **Most critical physical design issues affecting the end-user's perception includes:**
 - **physical sort order of the fact table on disk**
 - **presence of pre-stored summaries or aggregations**
- **Additional physical design issues include administration, backup, indexing performance, and security.**



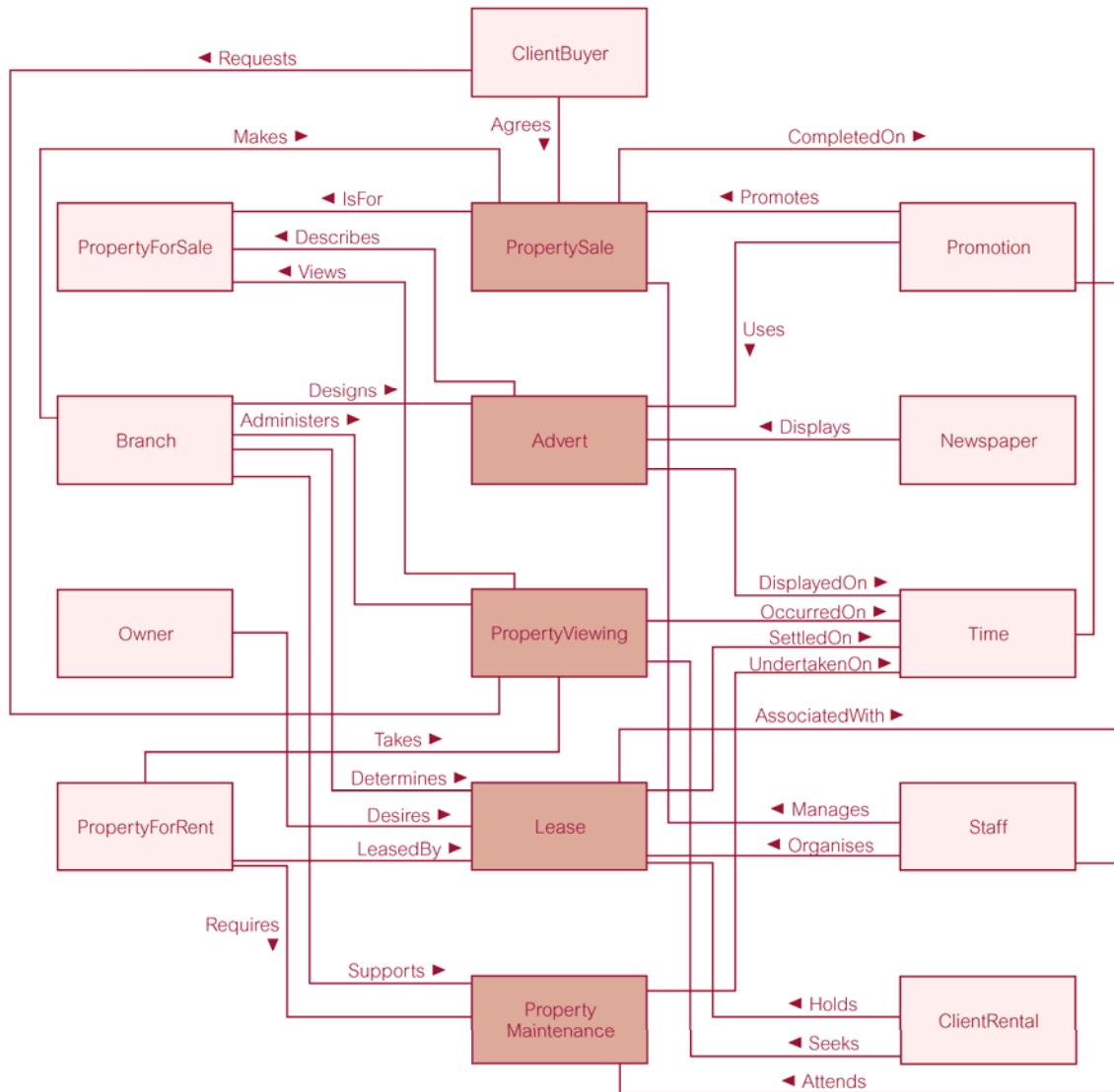
Database Design Methodology for Data Warehouses

- Methodology designs a data mart that supports the requirements of a particular business process and allows the easy integration with other related data marts to form the enterprise-wide data warehouse.
- A dimensional model, which contains more than one fact table sharing one or more conformed dimension tables, is referred to as a fact constellation.

Fact and dimension tables for each business process of *DreamHome*

Business process	Fact table	Dimension tables
Property sales	PropertySale	Time, Branch, Staff, PropertyForSale, Owner, ClientBuyer, Promotion
Property rentals	Lease	Time, Branch, Staff, PropertyForRent, Owner, ClientRenter, Promotion
Property viewing	PropertyViewing	Time, Branch, PropertyForSale, PropertyForRent, ClientBuyer, ClientRenter
Property advertising	Advert	Time, Branch, PropertyForSale, PropertyForRent, Promotion, Newspaper
Property maintenance	PropertyMaintenance	Time, Branch, Staff, PropertyForRent

Dimensional model (fact constellation) for the *DreamHome* data warehouse



Criteria for assessing the dimensionality of a data warehouse



- **Criteria proposed by Ralph Kimball (2000) to measure the extent to which a system supports the dimensional view of data warehousing.**
- **Twenty criteria divided into three broad groups: architecture, administration, and expression.**

Criteria for assessing the dimensionality of a data warehouse

Group

Criteria

Architecture

Explicit declaration
Conformed dimensions and facts
Dimensional integrity
Open aggregate navigation
Dimensional symmetry
Dimensional scalability
Sparsity tolerance

Administration

Graceful modification
Dimensional replication
Changed dimension notification
Surrogate key administration
International consistency

Expression

Multiple-dimension hierarchies
Ragged-dimension hierarchies
Multiple valued dimensions
Slowly changing dimensions
Roles of a dimension
Hot-swappable dimensions
On-the-fly fact range dimensions
On-the-fly behavior dimensions

Criteria for assessing the dimensionality of a data warehouse



- **Architectural criteria describes the way the entire system is organized.**
- **Administration criteria are considered to be essential to the ‘smooth running’ of a dimensionally-oriented data warehouse.**
- **Expression criteria are mostly analytic capabilities that are needed in real-life situations.**