CSCI312 Big Data Management

Assignment 2

Report

Task 1

Question (1)

Hierarchy

CALANDER

- TIME
    - YEAR
    - MONTH
    - DAY

LOCATION

- ADDRESS
    - COUNTRY
    - CITY
    - STREET

- OWNER
    - COUNTRY
    - CITY

- BUYER
    - COUNTRY
    - CITY

Fact table

- PROPERTY-TRANSACTION
- PROPERTY-INTEREST

Dimension

- PROPERTY-TRANSACTION
    - AGENT
    - OWNER
    - REAL-ESTATE-PROPERTY
    - TIME
    - ADDRESS

- PROPERTY-TRANSACTION
    - AGENT
    - REAL-ESTATE-PROPERTY
    - TIME
    - ADDRESS

MEASURE

- PROPERTY-TRANSACTION
    - price-asked
    - price-final
    - period-on-market

- PROPERTY-TRANSACTION
    - Interest-count

The total number of real estate sold is calculated through the aggregate function count(*) and grouped by the relevant dimensions. Therefore, it doesn't need to be included in the fact table.

The average asked price of real estate can be determined by using the aggregate function avg(price-asked) and grouping by the respective dimensions. Since the data within the REAL-ESTATE-PROPERTY table gets deleted after transactions, the "price-asked" measure is replicated in the fact table.

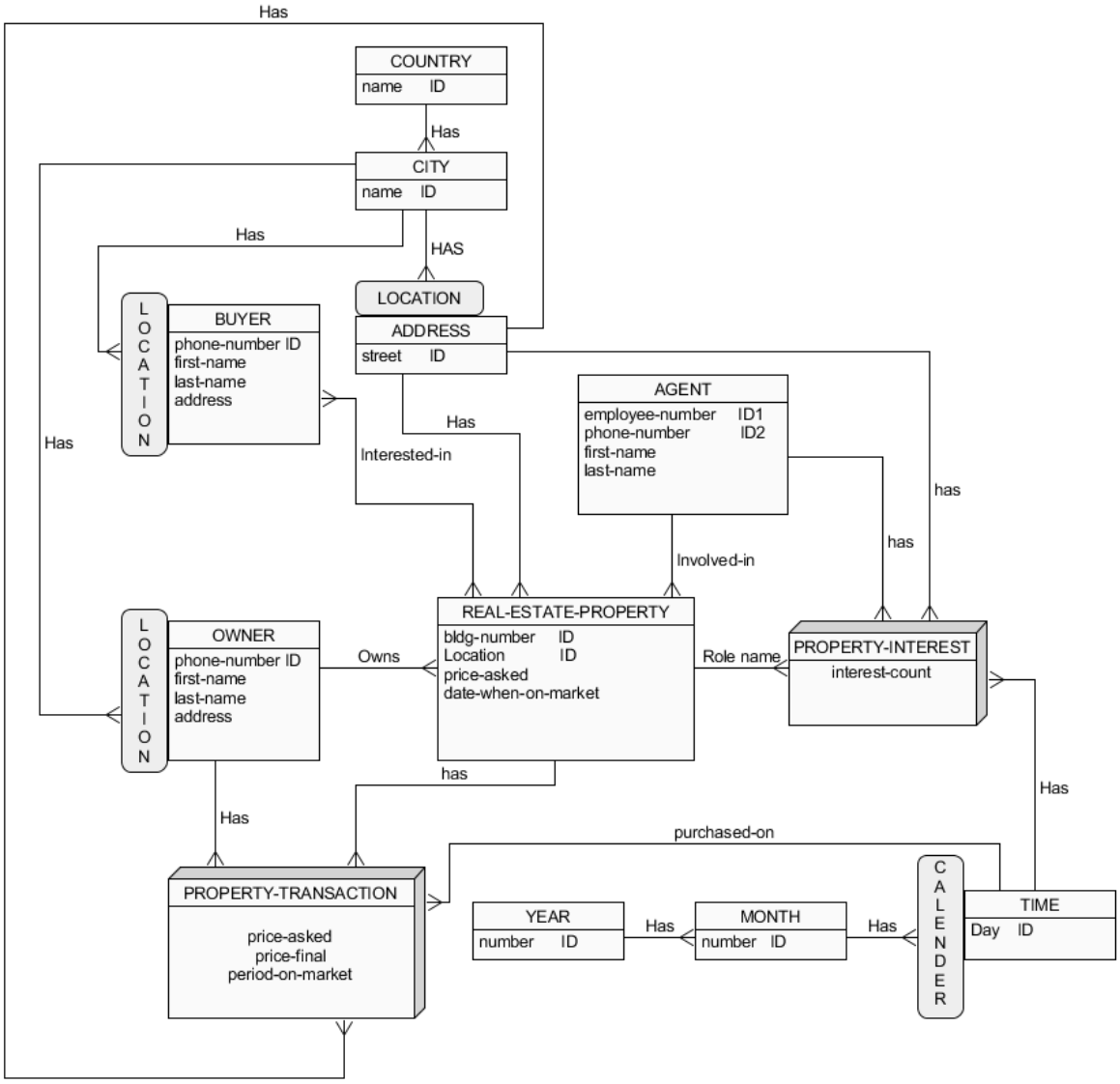
Similarly, the average final price of real estate can be computed using the aggregate function avg(price-final) and grouped by the relevant dimensions. This requires adding "price-final" to the fact table as a new measure.

The average period of time on the market can be calculated through the aggregate function avg(period-on-market) and grouped by the respective dimensions. This requires adding "period -on-market" to the fact table as a new measure. This can be derived by subtracting from the date of transaction complete to date-when-on-market. But the value inside the REAL-ESTATE-PROPERTY will get deleted so a new measure is added to save this.

The total number of times each real estate property sold within a specific timeframe can be derived by using the aggregate function count(*) on the PROPERTY-TRANSACTION table and grouping by the relevant dimensions. Thus, it's unnecessary to include this measurement in the fact table.

Finally, the total number of buyers interested can be computed by utilizing the aggregate function SUM (interest-count) and grouping by the respective dimensions. When someone is interested in the property, the count will be stored as a new measure in the ‘PROPERTY-INTEREST” fact table. Therefore, the total number of interest count to that property can be derived by using sum(interest-count) and grouped according to respective dimensions.

Conceptual Schema including the fact table and hierarchies.



Question (2) Three-dimensional data cube

Dimension 1 – TIME Dimension

Dimension 2 – AGENT Dimension

Dimension 3 – REAL-ESTATE-PROPERTY Dimension

Measure – Total number of property sold