

MEMO

QUESTION 3 (Higher order thinking question)

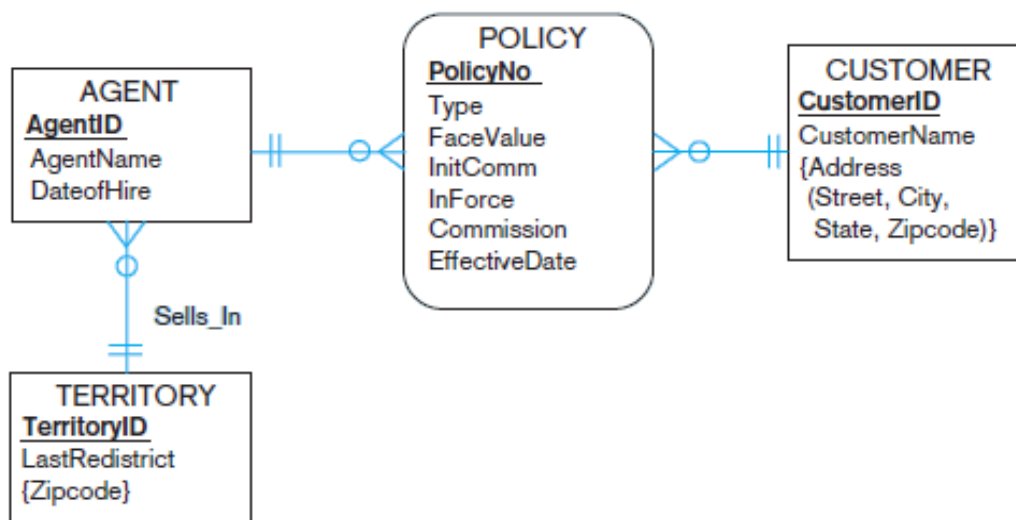
3. Analyse the case study below and answer the questions to follow:

WorkCube Insurance Company, which is primarily involved in the sale of annuity products, would like to design a data mart for its sales and marketing organization. Presently, the OLTP system is a legacy system residing on a Novell network consisting of approximately 600 different flat files. For the purposes of our case study, we can assume that 30 different flat files are going to be used for the data mart. Some of these flat files are transaction files that change constantly. The OLTP system is shut down overnight on Friday evening beginning at 6 P.M. for backup. During that time, the flat files are copied to another server, an extraction process is run, and the extracts are sent via FTP to a UNIX server. A process is run on the UNIX server to load the extracts into Oracle and rebuild the star schema. For the initial loading of the data mart, all information from the 30 files was extracted and loaded.

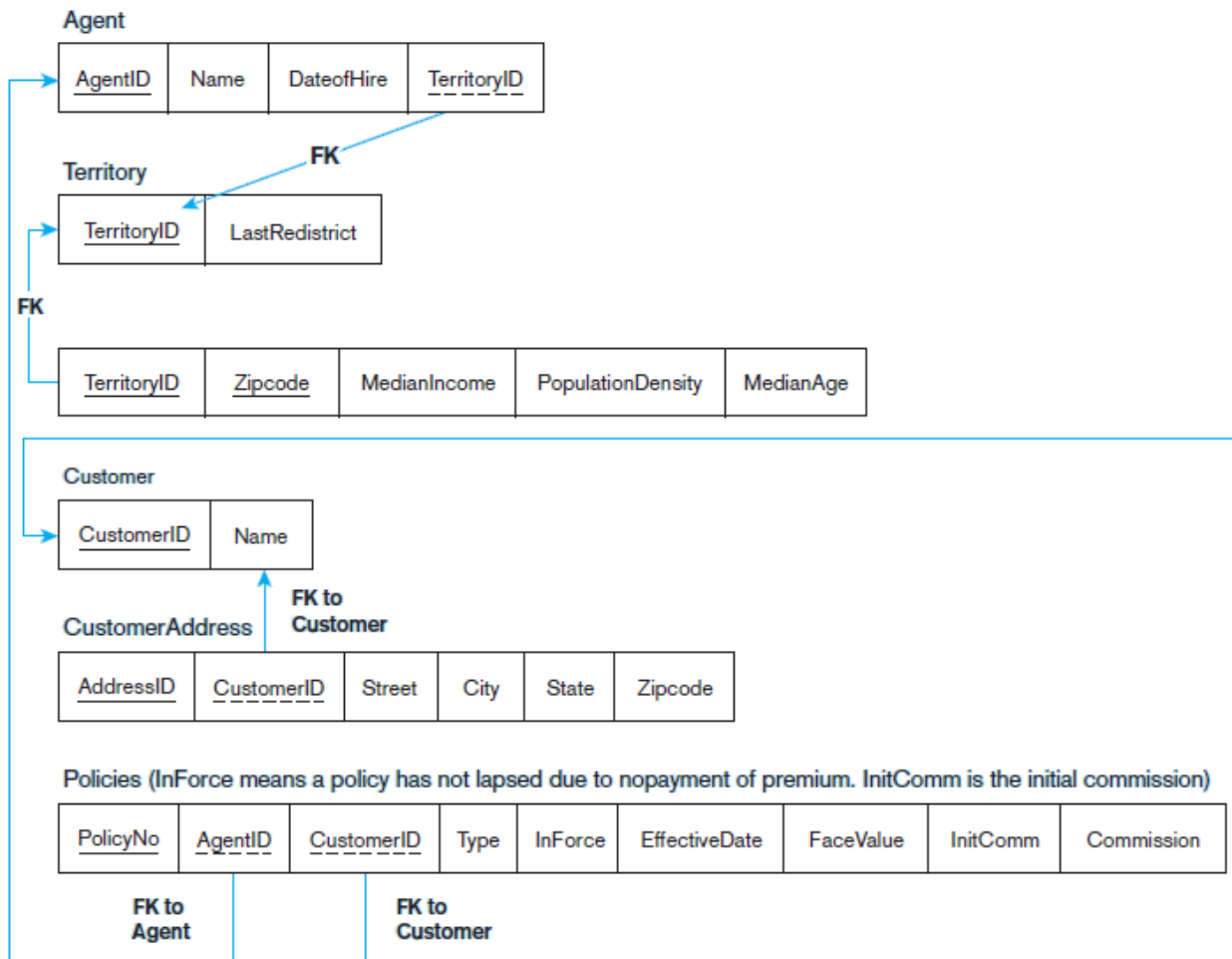
On a weekly basis, only additions and updates will be included in the extracts.

Although the data contained in the OLTP system are broad, the sales and marketing organization would like to focus on the sales data only. After substantial analysis, the ERD shown in the figure below was developed to describe the data to be used to populate the data mart. Sales and marketing is interested in viewing all sales data by territory, effective date, type of policy, and face value. In addition, the data mart should be able to provide reporting by individual agent on sales as well as commissions earned. Occasionally, the sales territories are revised (i.e., zip codes are added or deleted). The Last Redistrict attribute of the Territory table is used to store the date of the last revision. Some sample queries and reports are listed here:

- Total sales per month, by territory, by type of policy.
- Total sales per quarter by territory, by type of policy.
- Total sales per month by agent, by type of policy.
- Total sales per month by agent, by zip code.
- Total face value of policies, by month of effective date.
- Total face value of policies by month of effective date, by agent.
- Total face value of policies, by quarter of effective date.
- Total number of policies in force, by agent.
- Total number of policies not in force, by agent.
- Total face value of all policies sold by an individual agent.
- Total initial commission paid on all policies to an agent.
- Total initial commission paid on policies sold in a given month by agent.
- Total commissions earned by month, by agent.
- Top-selling agent by territory, by month.



MEMO

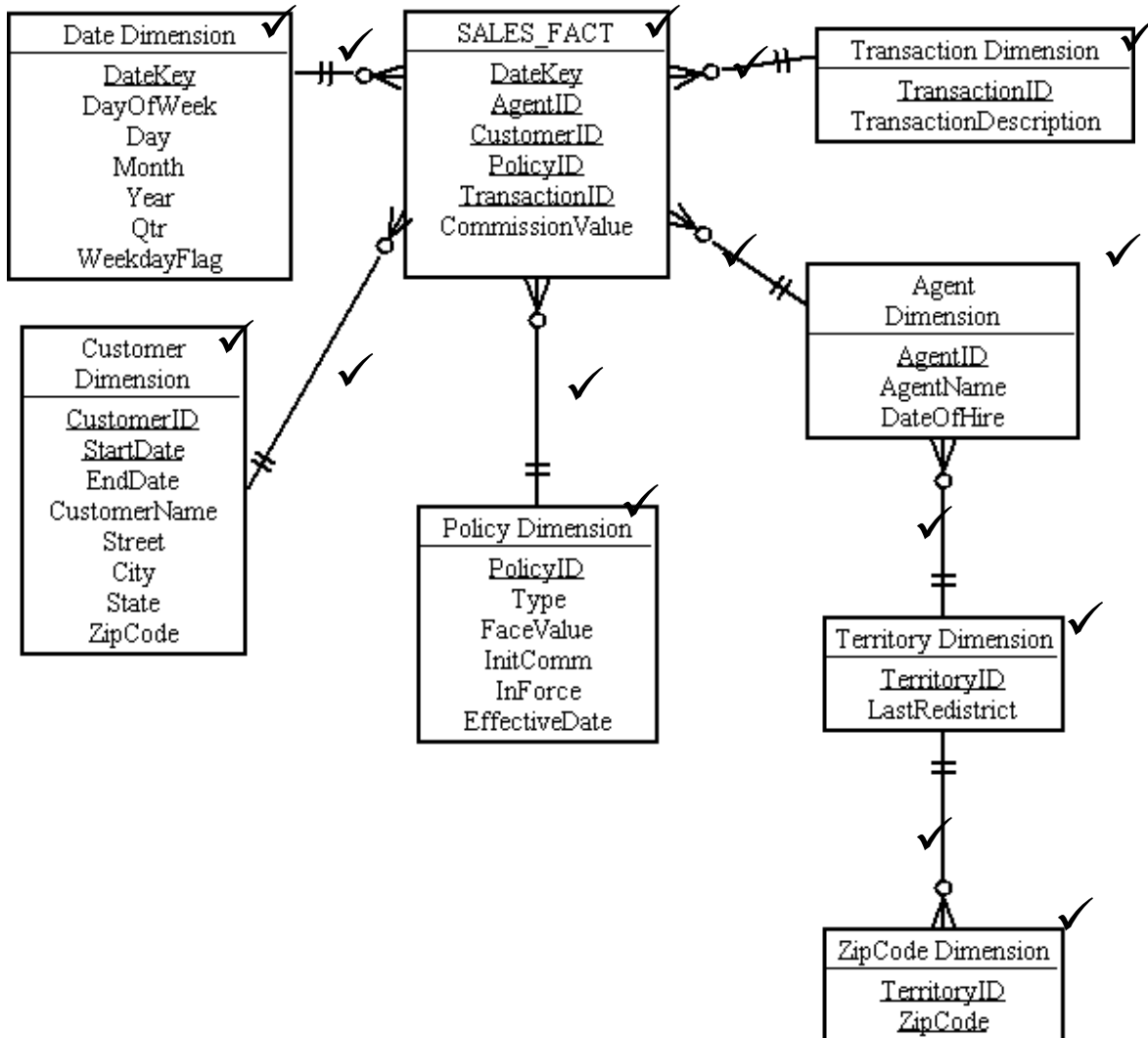


Commissions are paid to an agent upon the initial sale of a policy. The InitComm field of the policy table contains the percentage of the face value paid as an initial commission. The Commission field contains a percentage that is paid each month as long as a policy remains active or in force. Each month, commissions are calculated by computing the sum of the commission on each individual policy that is in force for an agent.

3.1. Create a star schema for this case study. How did you handle the time dimension? (19)

- The time dimension is handled with the Date Dimension table, which includes a surrogate key, and attributes of DayOfWeek, Day, Month, Year, Qtr, and WeekdayFlag. ✓
- The multivalued attribute of customer address is handled in this schema with a slowly changing dimension (SCD) element of a start date as part of the surrogate key for the customer dimension table. ✓
- Territory is snowflaked off of Agent. ✓
- TransactionID indicates whether the transaction is initial commission, monthly commission, write policy, or discontinue policy. ✓

MEMO



- 1 Mark for entity with attributes and keys show.
- 1 Mark for relationship with name.

3.2. Would you prefer to normalize (snowflake) the star schema of your answer to question 3.1? If so, how and why? Redesign the star schema to accommodate your recommended changes.

(2)

The answer to question 3.2 handles the changing customer address as a slowly changing dimension (SCD), thus there is no reason to snowflake to normalize further for the slowly changing dimension.

✓✓

[21]