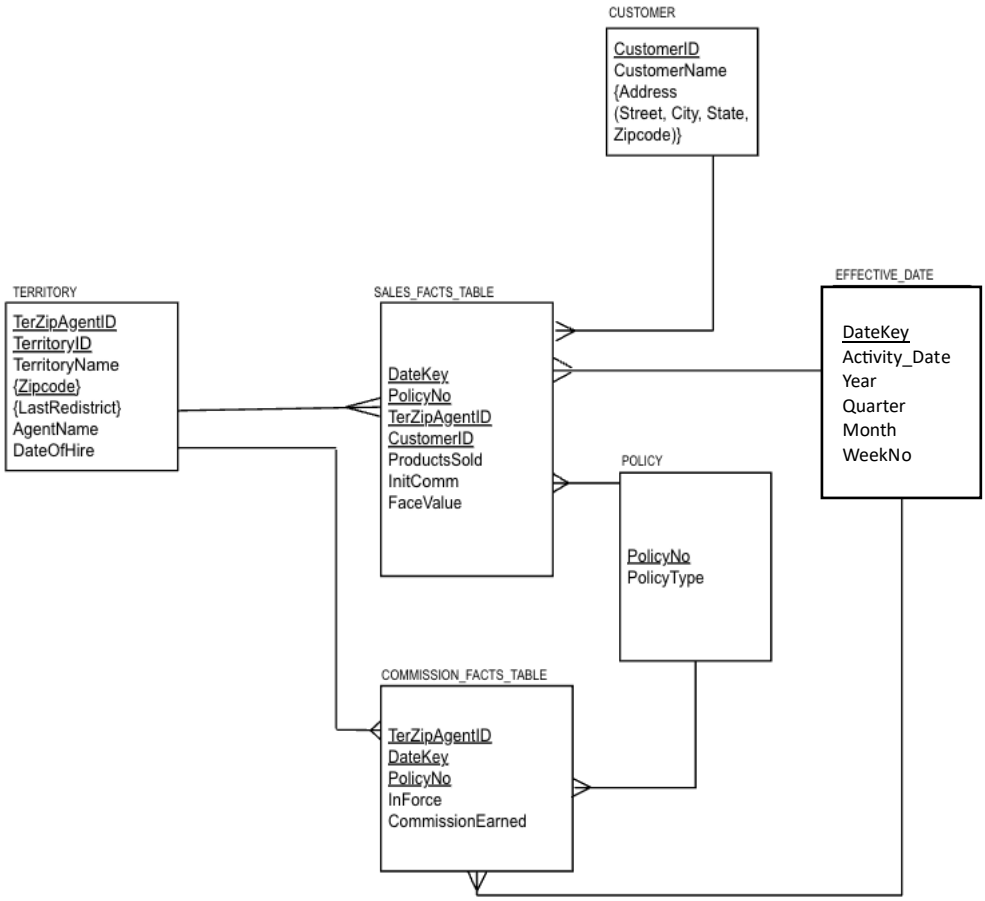
QUESTION 3A



For the purpose of this case study, I am assuming that the effective date is the date the products are sold/ policies signed/ commissions are earned.

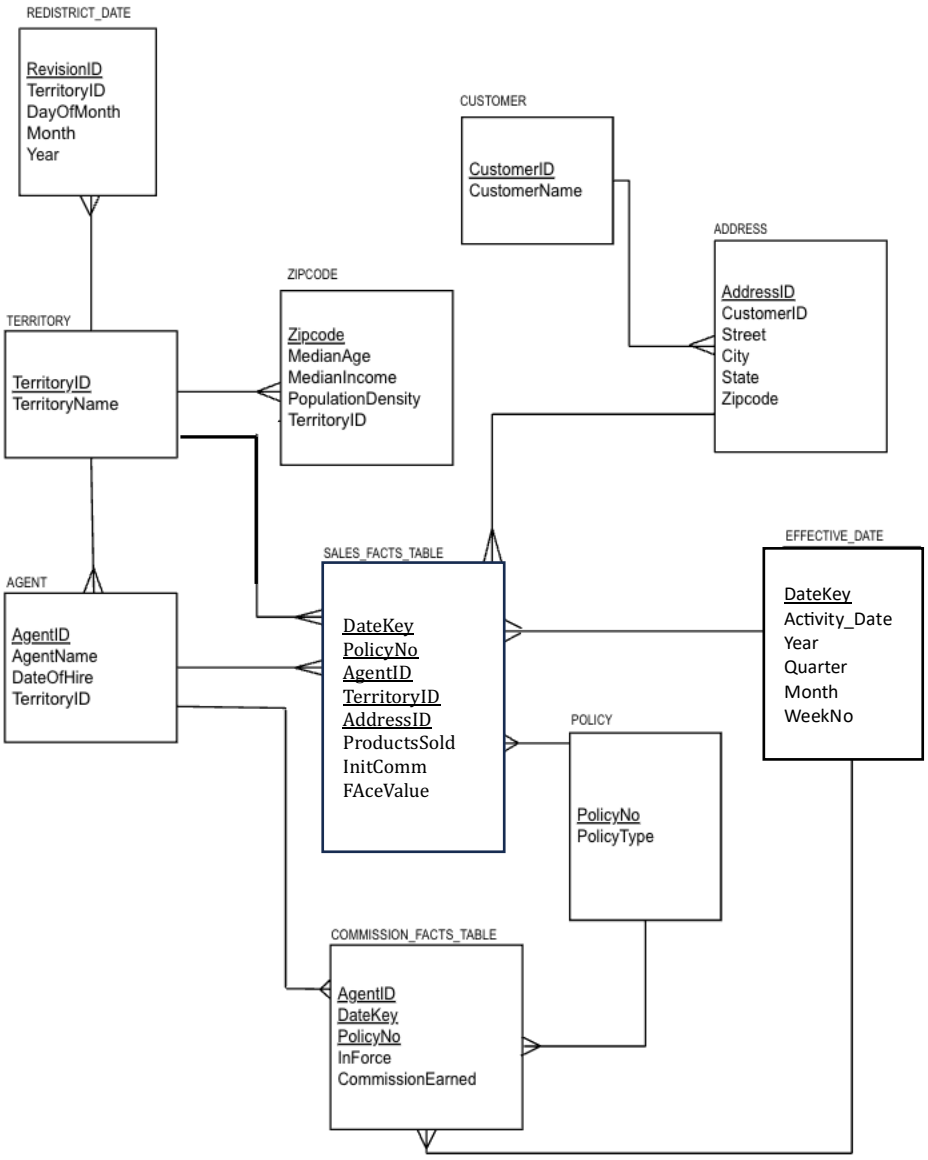
The date dimension (i.e. the effective date) is populated with WeekNo, Month, Quarter and Year variations in order to serve the business purpose of querying the dataset on a monthly, quarterly as well as yearly basis. It also includes week number to allow for easier upload into MySQL as the sales team extracts and loads **only** additions and updates on a weekly basis (Fridays). This is included in the figure below.

FIG 3.1

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 MySQL 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

Subsequently, the data mart time dimension can be improved to include day of the week, day in month, week of month, day of year etc. as the business may require.

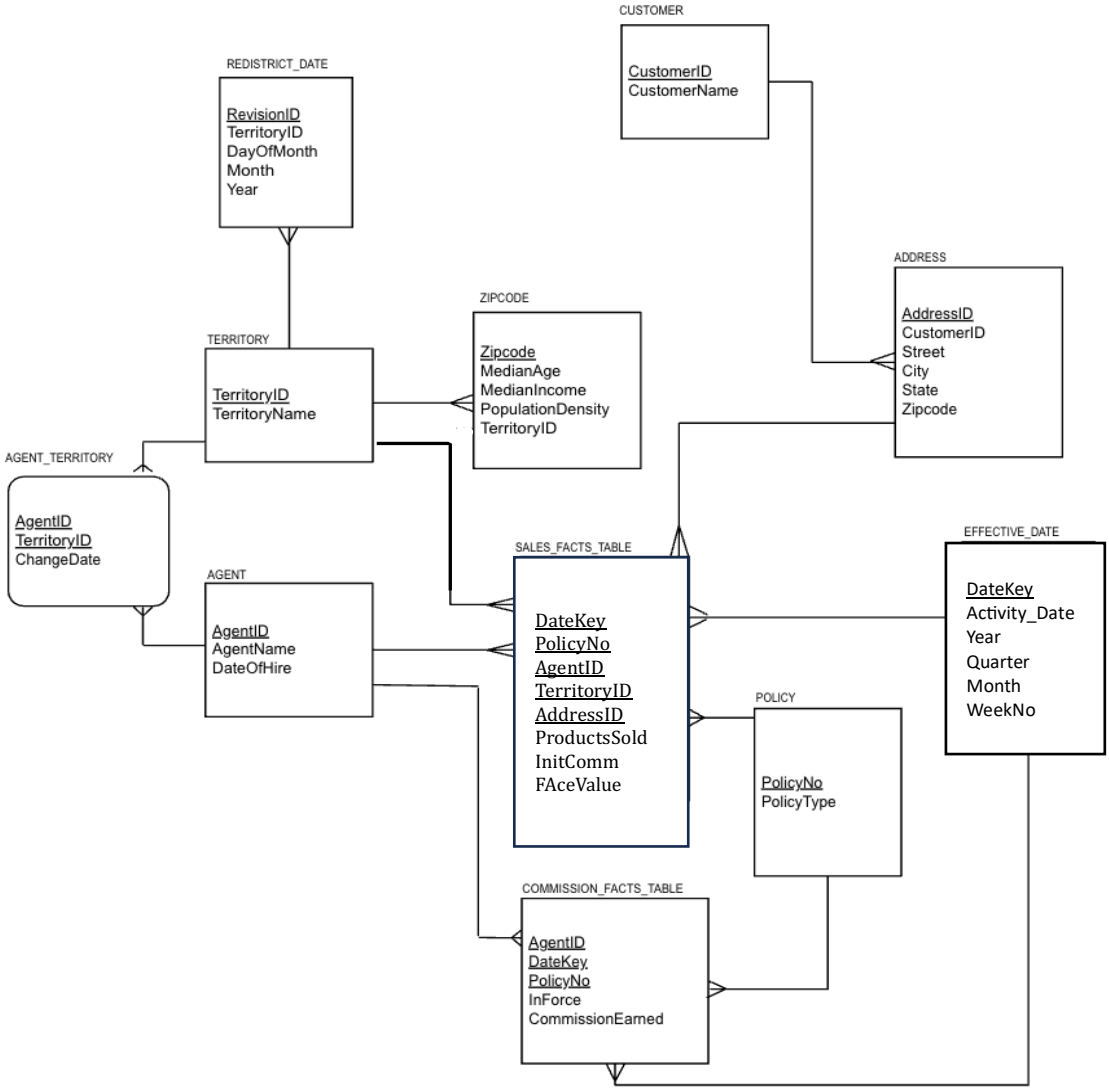
QUESTION 3B



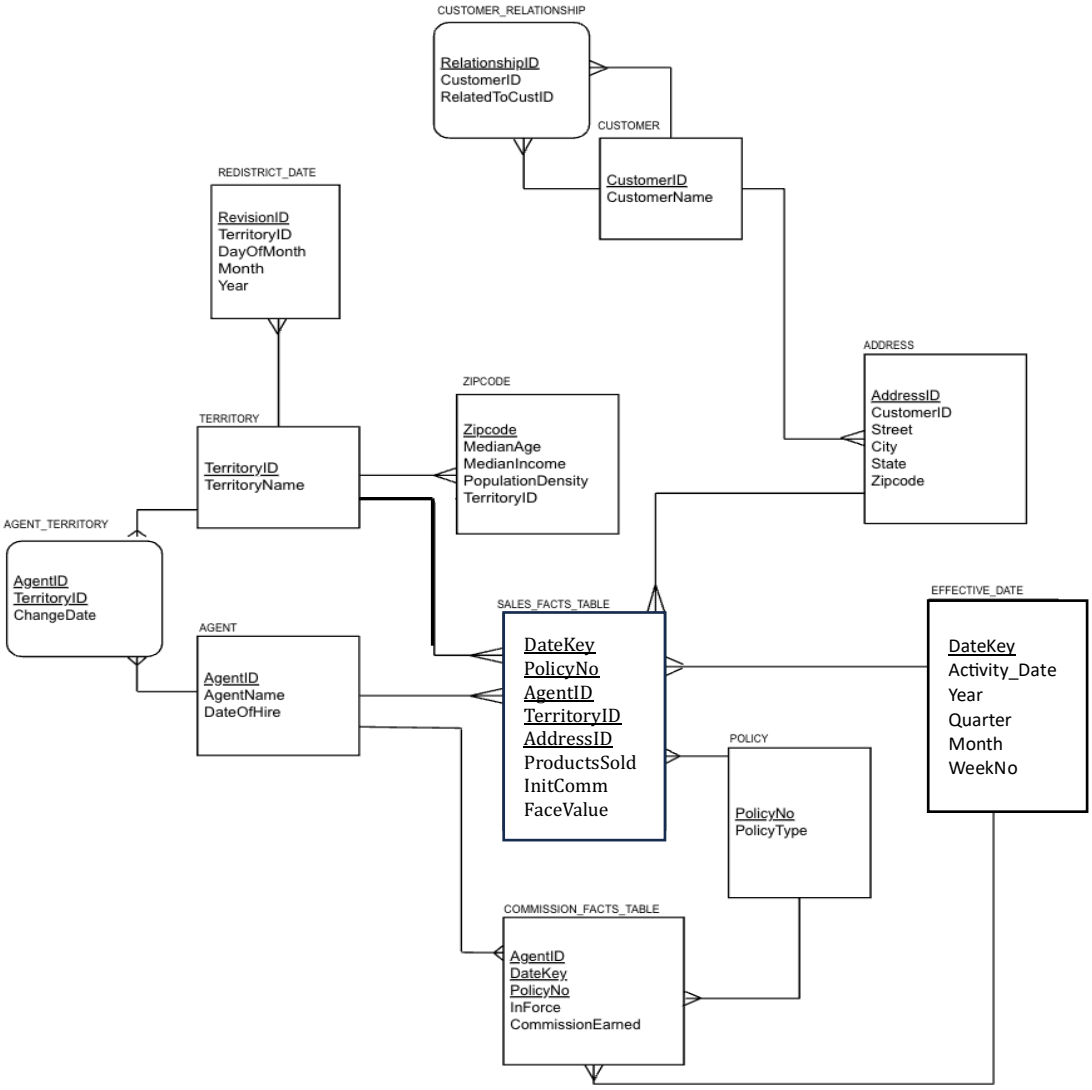
a. I prefer to normalize the star schema because it provides more granularity to the facts table, especially for the Agents dimension since that seems to be one of the major interests of the sales and marketing team of the Atlantic Insurance Company

   It also helps to reduce data integrity violation within the dimension tables.

QUESTION 3C


CUSTOMER
CustomerID
CustomerName
REDISTRICT_DATE
RevisionID
TerritoryID
DayOfMonth
Month
Year
ADDRESS
AddressID
CustomerID
Street
City
State
Zipcode
ZIPCODE
Zipcode
MedianAge
MedianIncome
PopulationDensity
TerritoryID
TERRITORY
TerritoryID
TerritoryName
AGENT_TERRITORY
AgentID
TerritoryID
ChangeDate
SALES_FACTS_TABLE
DateKey
PolicyNo
AgentID
TerritoryID
AddressID
ProductsSold
InitComm
FAceValue
EFFECTIVE_DATE
DateKey
Activity_Date
Year
Quarter
Month
WeekNo
AGENT
AgentID
AgentName
DateOfHire
POLICY
PolicyNo
PolicyType
COMMISSION_FACTS_TABLE
AgentID
DateKey
PolicyNo
InForce
CommissionEarned

QUESTION 3D



QUESTION 3E

Each record of each flat file will have to be reviewed critically with the intention to understand the structure and content of the data available.

Identifying the classes of entities, attributes as well as the relationships between various entity classes is also an important step in extracting the data from the flat files to create an ERD. The instances of each supposed entity should be analyzed and grouped according to their provided attributes. The data should then be cleaned to remove blanks or non-repeating attributes among entity instances.

At this point, the dataset is cleaned to a substantial amount and can now be converted into an ERD with the choice of normalization, depending on the company's preferences.

The size of the company's dataset as well as the amount of change in the dataset should determine the frequency of its extraction process. Initially, like it is stated in the case study, all the information should be extracted and loaded. However, if the changes to the dataset are not very frequent, an incremental extract is recommended either daily or weekly depending on the size of the data acquired by the company periodically, otherwise a static extraction is recommended to ensure uniformity in its data mart.

In the case of Atlantic Insurance, I would recommend an incremental extraction as the only stated constant changes are transactional and those do not require a general database update else there will be loss of the previous transactional information.

QUESTION 3I

No, it will not change the way I would extract or load data into the data mart because the data mart is time variant and as such accommodates changes that are made on a timely basis.

QUESTION 4A

I would recommend the Platform as a Service for the EDW implementation because of its provision of data management and configuration resources. This will allow the company developers focus on building and optimizing data warehouse applications without bothering about the infrastructure. Scalability is also an advantage when using the PaaS because as data increases in volume, the provider can scale the resources to ensure that a data warehouse can handle the load efficiently and effectively.

QUESTION 4B

Atlantic Insurance would need to ensure that all its employees are informed of the importance of data protection as well as the roles they play in achieving this. The employees should also be educated on the need to be threat sensitive and ways to avoid/avert it.

To protect its hardware from loss, procedures such as user authentication, uniform software installation, data encryption on storage devices as well as hardware maintenance schedules should be put in place to reduce the amount and severity of such data losses if and when they occur.

In consideration of losses that are not accidental, the company must ensure that the locations where sensitive data is stored are properly physically secured, allowing only authorized personnel access to those locations and to specific information.

Data encryption is an important aspect of data security be it hardware, software or via communication links. Encrypted data should be stored safely and only be accessed by authorized persons.

Establishing a firewall and will also help protect the company's data against communication links which are unauthorized to gain access to its database.

Authentication and authorization are software security protocols that should be put in place to confirm that a person trying to gain access to a specific dataset is who they say they are and that they are authorized by the company to do so.

QUESTION 4D

Role	Privilege
Agent	INSERT, UPDATE
Sales and marketing	SELECT