U1M6.LW.Star Schema Basics Part 2

Business Project

Shkrabatouskaya Vera

https://github.com/VeraShkrabatouskaya/DataMola_Data-Camping-2022

3. OLAP – Business analyses task

3.1. Task 03 – Solution concept – Business background

Overview

Data Warehouse Case Study (by Starcom)

Business Background

Starcom is a world-renowned media communications agency that architects connected human experiences to create value through precision marketing, content and technology solutions. With more than 5,000 employees worldwide, Starcom partners with the world's leading marketers and new establishment brands, including Airbnb, Bank of America, Kellogg Company, Kraft Heinz, Novartis, Samsung, Visa and more. Starcom is part of Publicis Media—the global media solutions group which also encompasses Zenith, Mediavest | Spark and Optimedia | Blue 449—a key division of Publicis Groupe, one of the world's leading communications groups.

Starcom harness the power of media, technology and data to create experiences people love and actions brands need. Data guides us to opportunity, helping us uncover what people want, need and expect. Technology makes personalization and relevance possible. Creativity makes it magic.

The company operates with a large amount of information and is therefore faced with the problem of data quality and lack of historical data in the right breakdowns.

Information from external and internal sources has a different structure and data is often duplicated or contradictory. The implementation of a Corporate Data Warehouse solves these problems. Data Warehouses accumulate cleansed and structured information about a company's business in a single source for external and internal users: management, employees and customers.

The problems that faced by Starcom is the agency personnel are increasingly facing a growing requirement to harness and leverage information so, it requires the ability to combine data virtually or in data warehouses with business intelligence analytic.

Benefit

We propose that Starcom create a data warehouse project:

- To structure this data and use it to help make time-sensitive decision, inform strategy and anticipate outcomes.
- To help it derive maximum potential from the mass of unstructured data that continues to grow each other.
- To have the potential to quickly generate a single view of all multi-channel retail information and convert big date into usable information.

Requirements

Business Requirements

ID	Description
BR-01	Calculation of revenue information (Gross, Net) on a daily, weekly, monthly, quarterly, annual basis by agency, department, client, brand, promotion.
BR-02	Calculation of cost information (Gross, Net) on a daily, weekly, monthly, quarterly, annual basis by agency, department, client, brand, promotion.
BR-03	Calculation of profit information (Gross, Net) on a daily, weekly, monthly, quarterly, annual basis by agency, department, client, brand, promotion.
BR-04	Calculation of employee salary information (Gross, Net) on a monthly, quarterly, annual basis.
BR-05	Calculation of information about the number of promotions by brands.
BR-06	Calculation of information on the number of clients served by the agency and the number of employees of the network agencies based on their location for the management of the head office.

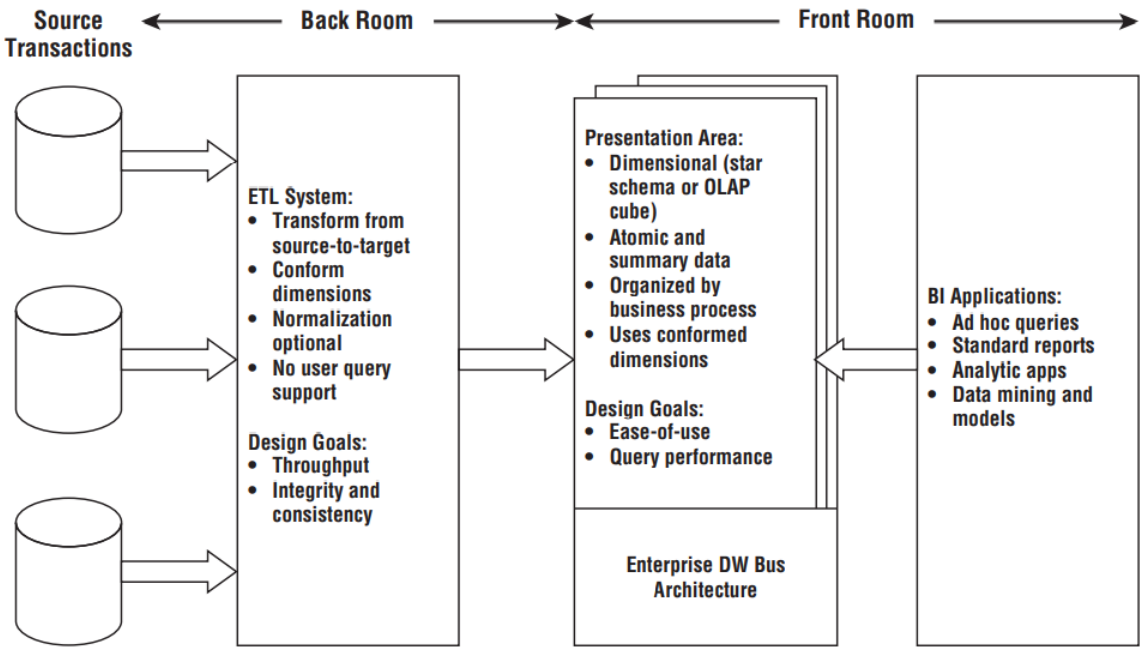
Technical Requirements

ID	Description
TR-01	Fast, round-the-clock access to data
TR-02	Automatic support of data consistency;
TR-03	Providing the ability to work with data slices
TR-04	Ensure data integrity and reliability
TR-05	Protect large volumes of data

Solution Sketch

Source Tables structure

We will use the Kimball DW/BI architecture model as a basis.



We will suggest that the customer consider 2 types of DWH: star and snowflake.

Summarize Data Plan

Expected result:

- A system will be designed and developed for Starcom DWH that can upload, analyze and import data from any data source.
- Any employee can link his data source to a Data Warehouse which will lead to quickly sharing important information within the company.

4. OLAP – Develop Star-Scheme and SnowFlake Scheme.

Working closely with representatives from the business and other DW/BI team members, the Data Warehouse Design Specialist developed a fact sheet with a grain of reporting for the customer, agency management and agency employees for each day.

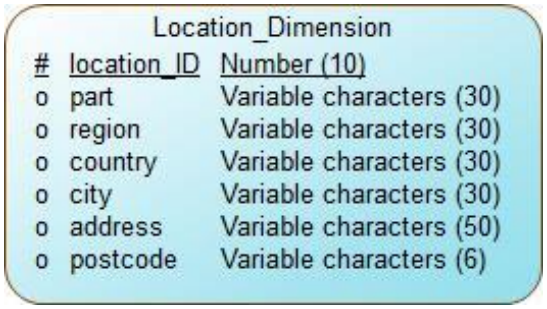
Based on the business problem, we needed to build a data warehouse to calculate gross/net cost and profit, employee salaries, number of clients, employees and promotion project. All fact table metrics go through a data aggregation procedure.

For the database, English was chosen for all locations in the agency's network. All monetary data is given in dollars.

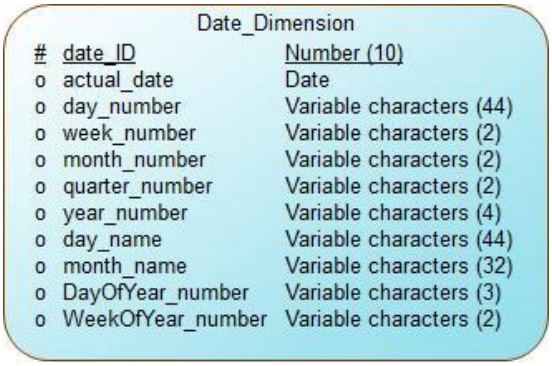
4.1. Task 04 – Develop Star-Scheme physical diagram

Dimensions tables of star scheme:

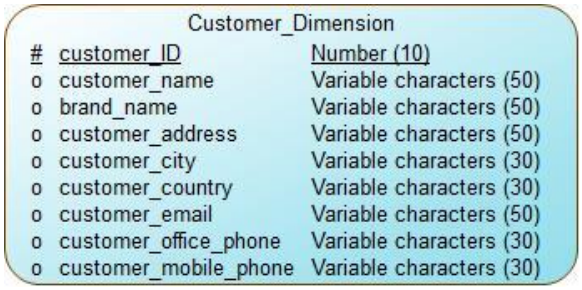
- Location_Dimension (PK location_ID)



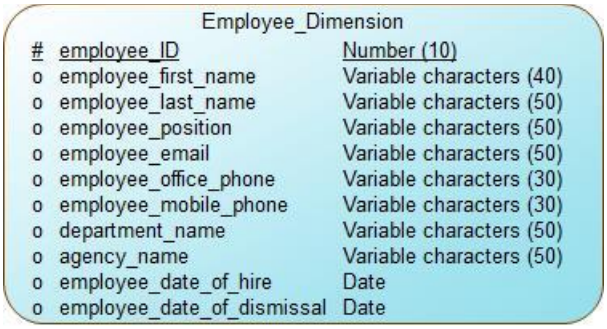
- Date_Dimension (PK date_ID)



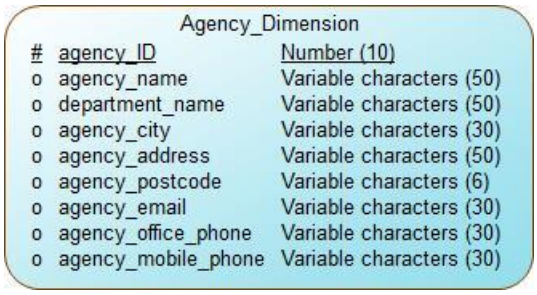
- Customer_Dimension (PK customer_ID)



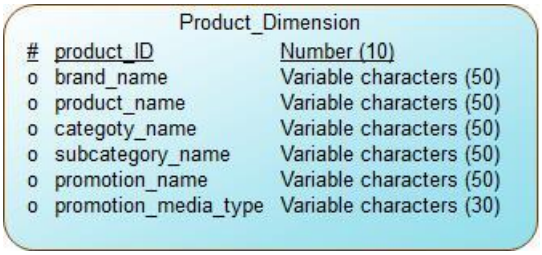
- Employee_Dimension (PK employee_ID)



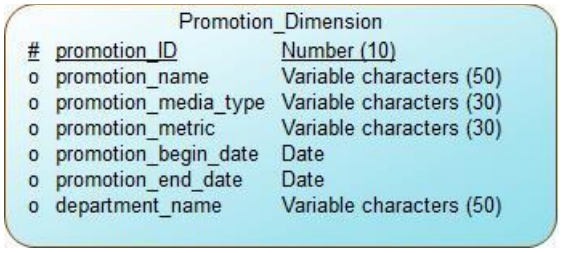
- Agency_Dimension (PK agency_ID)



- Product_Dimension (PK product_ID)

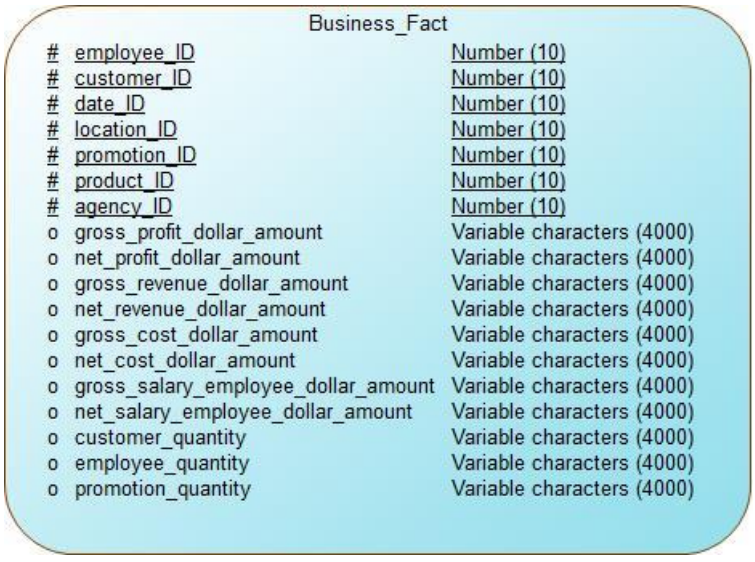


- Promotion_Dimension (PK promotion_ID)

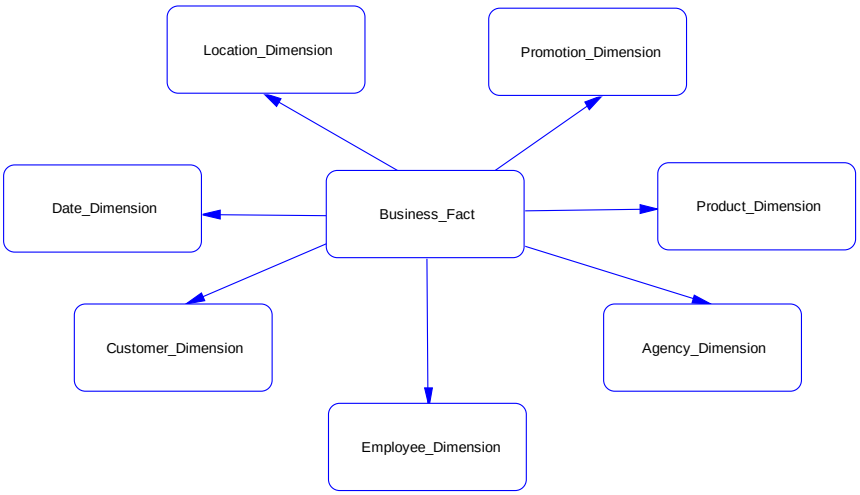


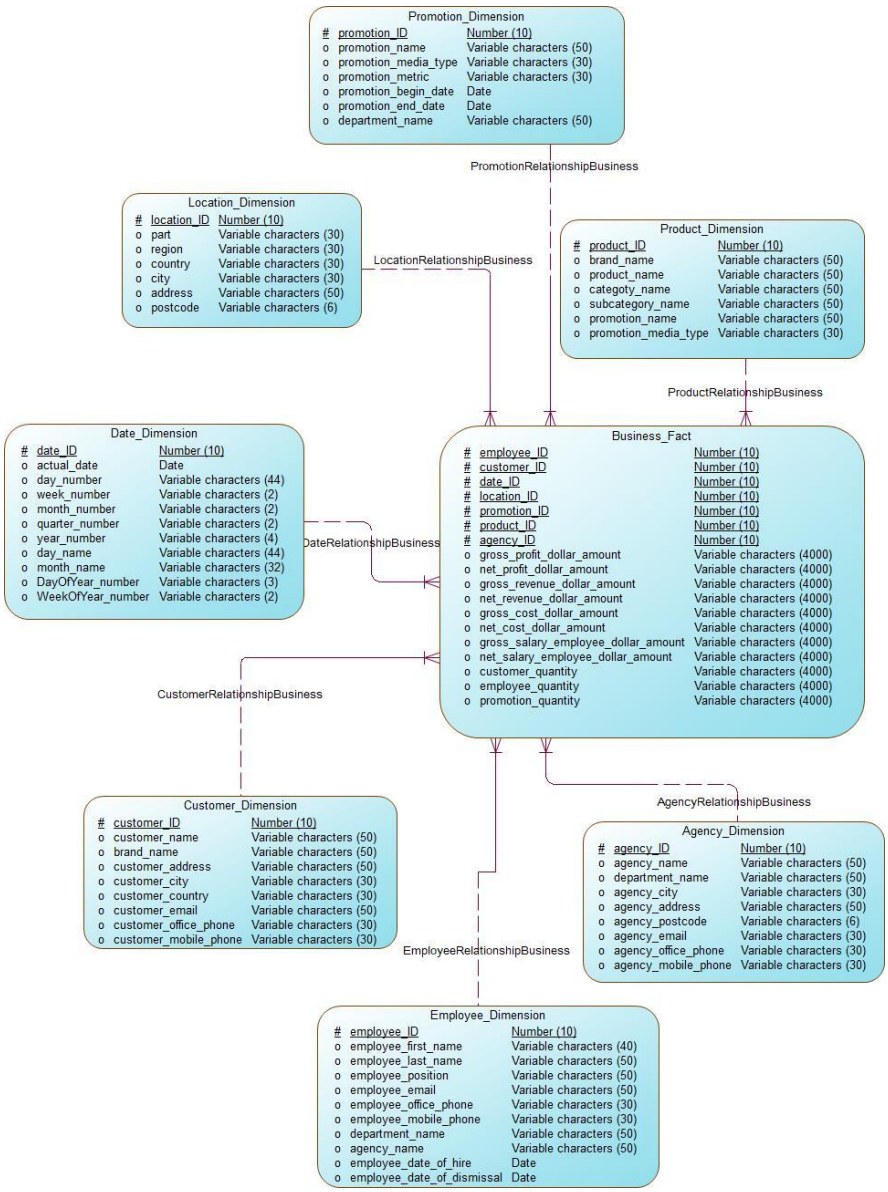
Fact table of star scheme:

- Business_Fact (FK employee_ID, FK customer_ID, FK date_ID, FK location_ID, FK promotion_ID, FK product_ID, FK agency_ID)



Star Scheme:

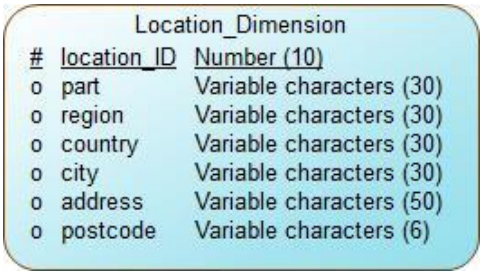

Location_Dimension
Promotion_Dimension
Date_Dimension
Business_Fact
Product_Dimension
Customer_Dimension
Agency_Dimension
Employee_Dimension


Promotion_Dimension
promotion_ID Number (10)
o promotion_name Variable characters (50)
o promotion_media_type Variable characters (30)
o promotion_metric Variable characters (30)
o promotion_begin_date Date
o promotion_end_date Date
o department_name Variable characters (50)
PromotionRelationshipBusiness
Location_Dimension
location_ID Number (10)
o part Variable characters (30)
o region Variable characters (30)
o country Variable characters (30)
o city Variable characters (30)
o address Variable characters (50)
o postcode Variable characters (6)
LocationRelationshipBusiness
Product_Dimension
product_ID Number (10)
o brand_name Variable characters (50)
o product_name Variable characters (50)
o categoty_name Variable characters (50)
o subcategory_name Variable characters (50)
o promotion_name Variable characters (50)
o promotion_media_type Variable characters (30)
ProductRelationshipBusiness
Date_Dimension
date_ID Number (10)
o actual_date Date
o day_number Variable characters (44)
o week_number Variable characters (2)
o month_number Variable characters (2)
o quarter_number Variable characters (2)
o year_number Variable characters (4)
o day_name Variable characters (44)
o month_name Variable characters (32)
o DayOfYear_number Variable characters (3)
o WeekOfYear_number Variable characters (2)
DateRelationshipBusiness
Business_Fact
employee_ID Number (10)
customer_ID Number (10)
date_ID Number (10)
location_ID Number (10)
promotion_ID Number (10)
product_ID Number (10)
agency_ID Number (10)
o gross_profit_dollar_amount Variable characters (4000)
o net_profit_dollar_amount Variable characters (4000)
o gross_revenue_dollar_amount Variable characters (4000)
o net_revenue_dollar_amount Variable characters (4000)
o gross_cost_dollar_amount Variable characters (4000)
o net_cost_dollar_amount Variable characters (4000)
o gross_salary_employee_dollar_amount Variable characters (4000)
o net_salary_employee_dollar_amount Variable characters (4000)
o customer_quantity Variable characters (4000)
o employee_quantity Variable characters (4000)
o promotion_quantity Variable characters (4000)
CustomerRelationshipBusiness
AgencyRelationshipBusiness
Customer_Dimension
customer_ID Number (10)
o customer_name Variable characters (50)
o brand_name Variable characters (50)
o customer_address Variable characters (50)
o customer_city Variable characters (30)
o customer_country Variable characters (30)
o customer_email Variable characters (50)
o customer_office_phone Variable characters (30)
o customer_mobile_phone Variable characters (30)
Agency_Dimension
agency_ID Number (10)
o agency_name Variable characters (50)
o department_name Variable characters (50)
o agency_city Variable characters (30)
o agency_address Variable characters (50)
o agency_postcode Variable characters (6)
o agency_email Variable characters (30)
o agency_office_phone Variable characters (30)
o agency_mobile_phone Variable characters (30)
EmployeeRelationshipBusiness
Employee_Dimension
employee_ID Number (10)
o employee_first_name Variable characters (40)
o employee_last_name Variable characters (50)
o employee_position Variable characters (50)
o employee_email Variable characters (50)
o employee_office_phone Variable characters (30)
o employee_mobile_phone Variable characters (30)
o department_name Variable characters (50)
o agency_name Variable characters (50)
o employee_date_of_hire Date
o employee_date_of_dismissal Date

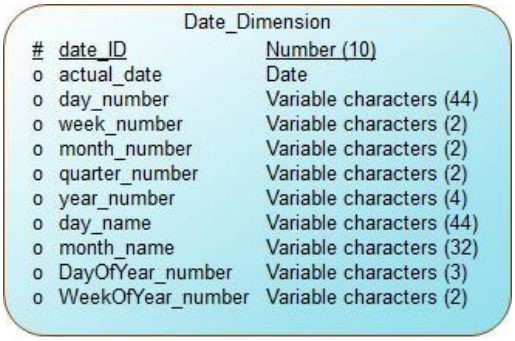
4.2. Task 05 – Develop SnowFlake physical diagram

Dimensions tables of snowflake scheme:

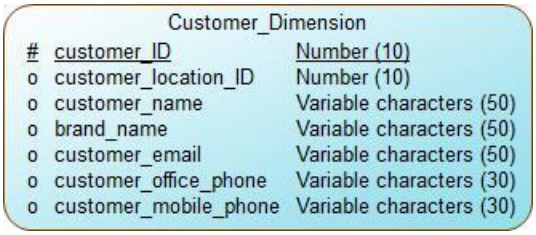
- Location_Dimension (PK location_ID)



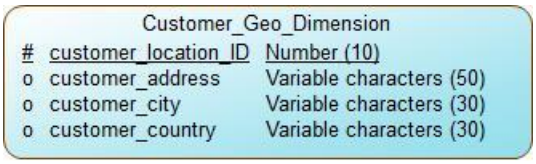
- Date_Dimension (PK date_ID)



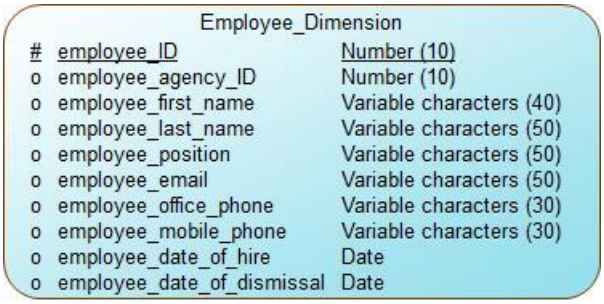
- Customer_Dimension (PK customer_ID)



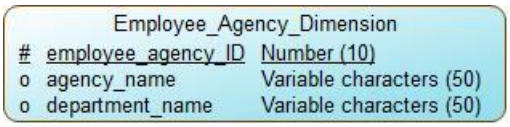
- Customer_Geo_Dimension (PK customer_location_ID, Child table of Customer_Dimension)



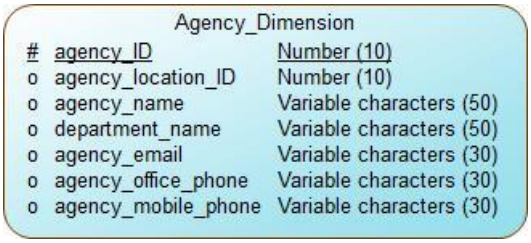
- Employee_Dimension (PK employee_ID)



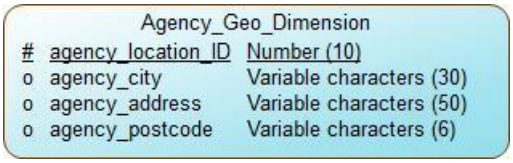
- Employee_Agency_Dimension (PK employee_agency_ID, Child table of Employee_Dimension)



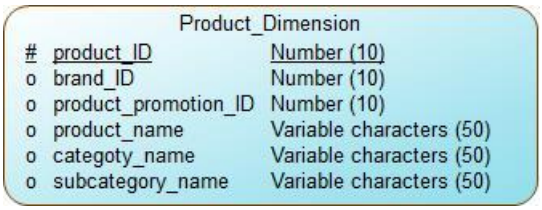
- Agency_Dimension (PK agency_ID)



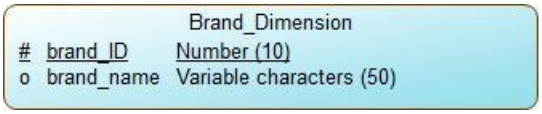
- Agency_Geo_Dimension (PK agency_location_ID, Child table of Agency_Dimension)



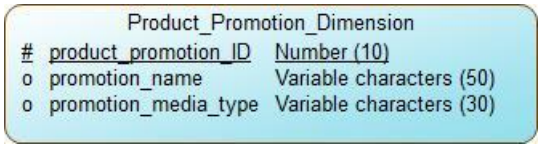
- Product_Dimension (PK product_ID)



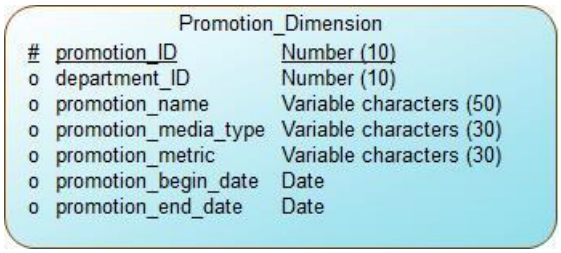
- Brand_Dimension (PK brand_ID, Child table of Product_Dimension)



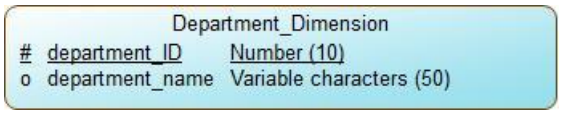
- Product_Promotion_Dimension (PK product_promotion_ID, Child table of Product_Dimension)



- Promotion_Dimension (PK promotion_ID)

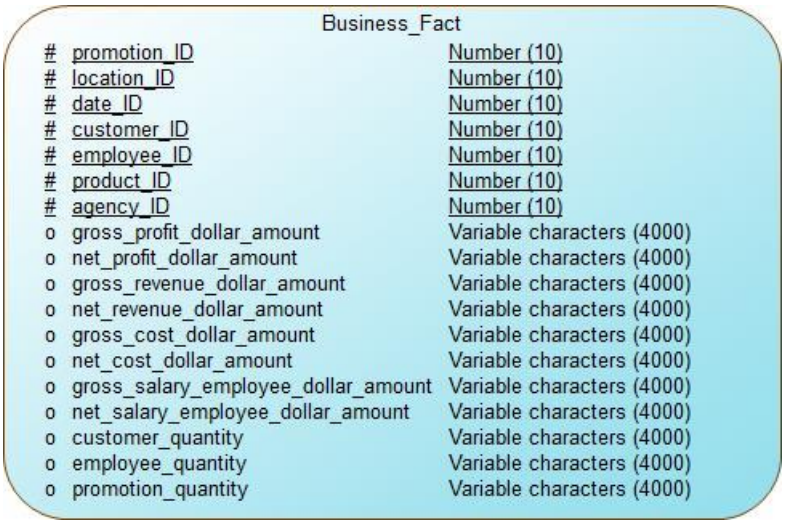


- Department_Dimension (PK department_ID, Child table of Promotion_Dimension)

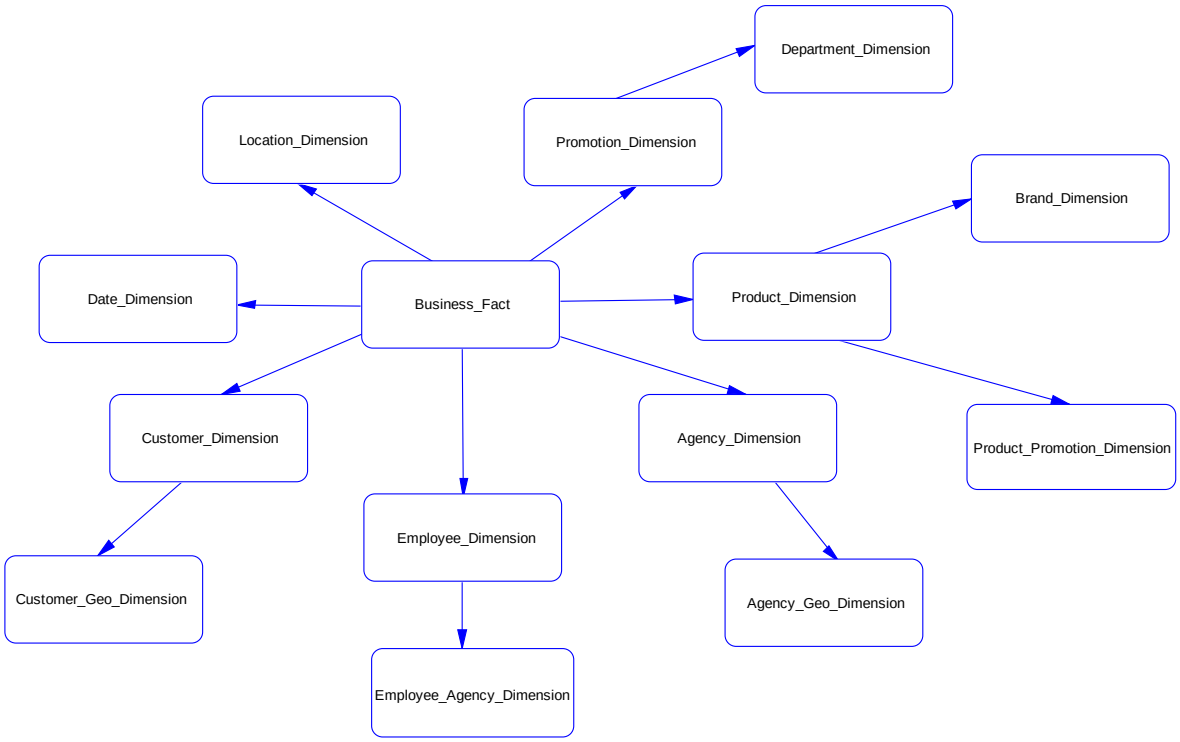


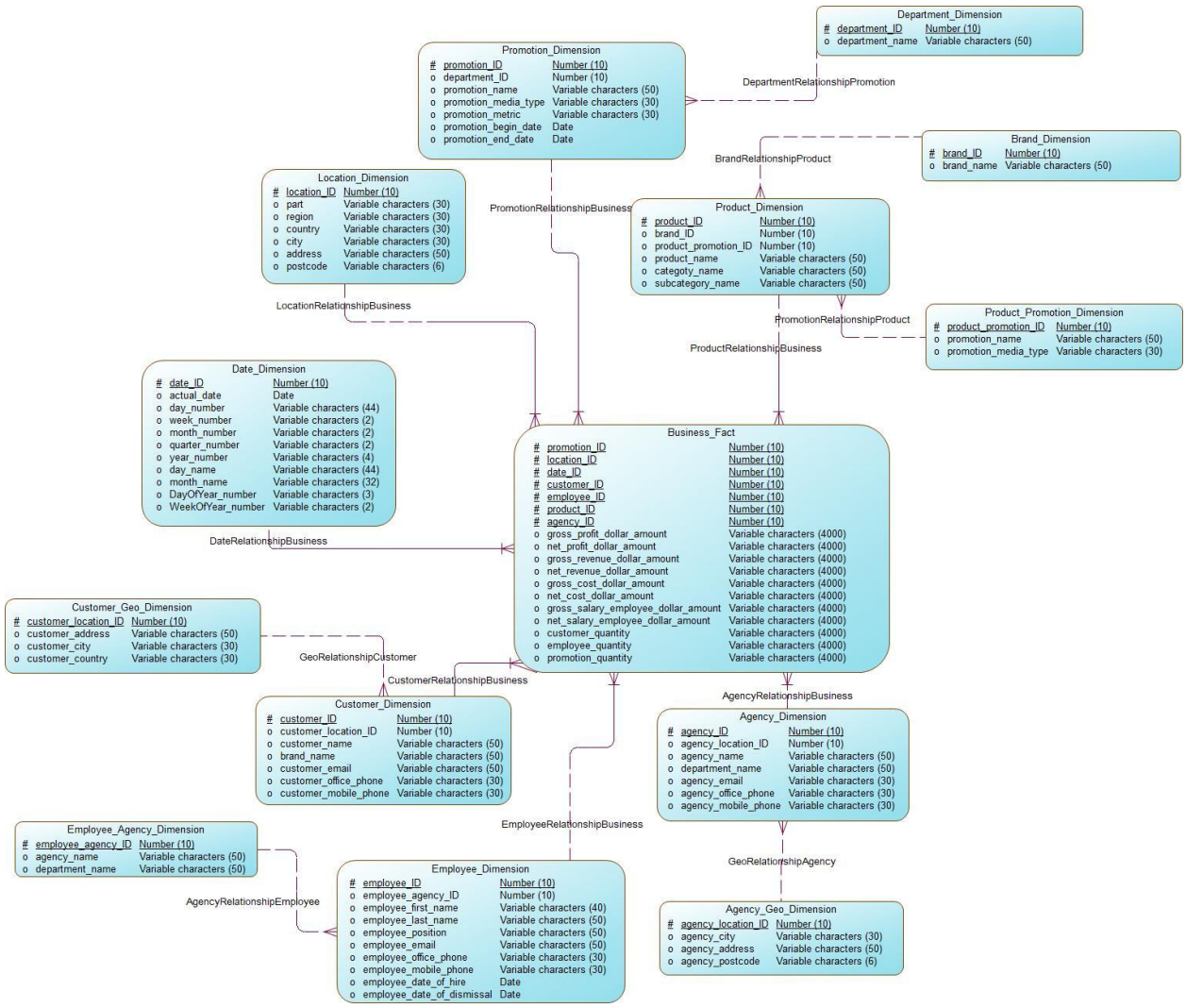
Fact table of snowflake scheme:

- Business_Fact (FK employee_ID, FK customer_ID, FK date_ID, FK location_ID, FK promotion_ID, FK product_ID, FK agency_ID)



Snowflake Scheme:





Summary:

As we can see, the main difference between the two relational database models is normalisation. The dimension tables in the star schema are not normalised, which means that the business model will use relatively more space to store the dimension tables, and more space means more redundant records, which will eventually lead to inconsistency. The snowflake scheme, on the other hand, minimizes data redundancy because the measurement tables are normalized, which explains the much smaller number of redundant records. The business hierarchy and its dimensions are preserved through referential integrity, which means that links can be updated independently in the data stores.

The star schema has fewer links between the dimension table and the fact table compared to the snowflake schema, which has many links, which explains the lower query complexity. Because measurements in a star type scheme are linked through a central fact table, it has clear connection paths, which means fast query response time, and fast response time means better performance. The snowflake scheme uses more connections, so the query response time increases, resulting in more complex queries, which ultimately reduces performance.

The choice between these two models must be made based on the customer's business objectives.