Data warehouse project

1. Business process

The business process for which the data warehouse is designed is estate brokerage service. The business process to which the proposed data warehouse applies is described in the Requirements specification for management and brokerage service of properties.
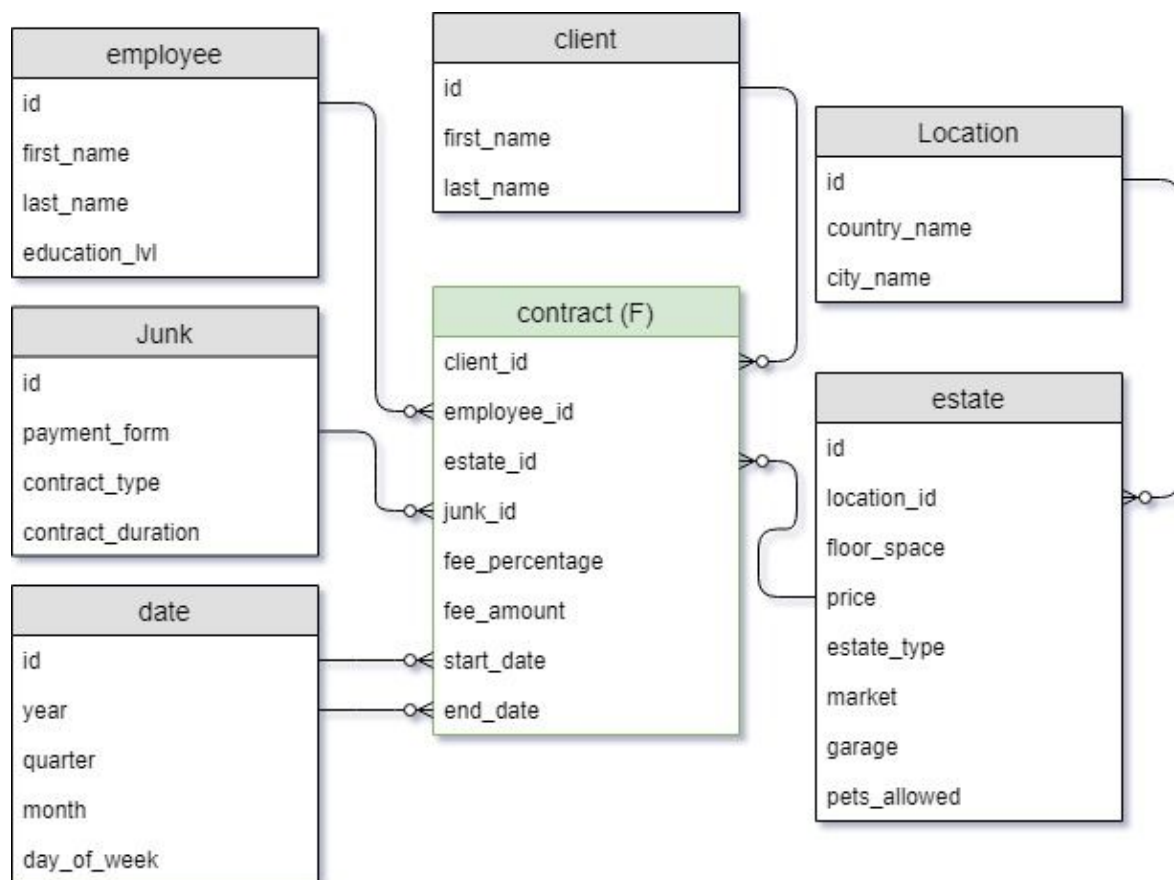
2. Relational Database Schema

Table name	Attribute	Attribute type	Description
contract (fact table)	Tuples correspond to the facts of contract sign.		
	client_id	numerical	FK client
	employee_id	numerical	FK employee
	estate_id	numerical	FK estate
	junk_id	numerical	FK junk
	fee_precentage	numerical	Value of the fee as a percentage (0% - 100%).
	fee_amount	numerical	Value of the fee in american dollars ($).
	start_date	numerical	FK date
	end_date	numerical	FK date
date (dimension table)	Tuples correspond to dates.		
	id	numerical	PK
	year	numerical	Describe year - 4 digits
	quarter	string (7 characters)	Describe quarter, range (first, second, third, fourth)
	month	string (10 characters)	Describe month, range (January, February, March, April, May, June, July, August, September, October, November, December)
	day_of_week	string (10 characters)	Describe day of week, range (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday)

employee (dimension table)	Tuples correspond to personal information of employees.		
	id	numerical	PK
	first_name	string (32 characters)	First name of employee
	last_name	string (32 characters)	Surname of employee
	education_lvl	string (10 characters	Education level (primary, medium, higher)
client (dimension table)	Tuples correspond to personal information of clients.		
	id	numerical	PK
	first_name	string (32 characters)	First name of a client.
	last_name	string (32 characters)	Last name of a client.
estate (dimension table)	Tuples correspond to personal information of real estates.		
	id	numerical	PK
	location_id	numerical	FK location
	floor_space	string (8 characters)	Floor space in m^2, range ['(0; 20>', '(20; 40>', '(40; 60>', '(60; 90>', '(90;150>, (150;250), '(250;...) ']
	price	string (16 characters)	Price in american dollars ($), range ['(0; 50 000>', '(50 000; 100 000>', '(100 000; 150 000>', '(150 000; 200 000>', '(200 000; 300 000>', '(300000; ....)']
	estate_type	string (16 characters)	Type of real estate, range (House, Studio Apartment, Bed Apartment)

	market	string (16 characters)	Type of market, range (primary market, secondary market)
	garage	boolean	Is there a garage? True/False
	pets_allowed	boolean	Can pets live there? True/False
location (dimension table)	Tuples correspond to location (country, city).		
	id	numerical	PK
	country_name	string (32 characters)	The name of the country.
	city_name	string (32 characters)	The name of the city.
Junk (dimension table)	Tuples correspond to junks.		
	id	numerical	PK
	payment_form	string (32 characters)	The form of payment used, range (check, cash, blik, bank transfer, direct debit, debit card)
	contract_type	string (32 characters)	Type of signed contract, range (sale brokerage, purchase brokerage)
	contract_duration	string (16 characters)	Duration of contract in days, range ['(1;7>', '(7; 14>', '(14; 30>', '(30; 90>', '(90;365>, (365;...)']

3. Multidimensional Data Model
3.1. Fact definition

Fact 1: The signing of a successful contract on a specified date, with a specified client, involving a specified employee and regarding a specified estate.
Fact table: contract
Granularity:
- specific signing
- specific employee with a certain education
- specific client
- specific date
- specific estate with specific details regarding the property and existing in a specific location

Measures and aggregation functions:
Number of signed contracts - COUNT(1)
Number of clients - COUNT(client_id)
Number of employees - COUNT(employee_id)
Fee amount - fee_amount
Profit - SUM(fee_amount)
Average fee amount - AVG(fee_amount)
Average profit - (Profit)/(Number of signed contracts)

3.2. Dimension definition

Dimensions of fact 1:

Dimension / Dimension attribute	Table / Field in table	Type
Percentage of the fee.	contract.fee_percentage	Degenerate dimension
Amount of the fee.	contract.fee_amount	Degenerate dimension
Employee	employee	Dimension
First name of the employee.	employee.first_name	Dimension attribute
Last name of employee.	employee.last_name	Dimension attribute
Education of employee.	employee.education_lvl	Dimension attribute
Start date - hierarchy	● date.year ●● date.month ●●● date.day_of_week	Hierarchical Dimension
End date - hierarchy	● date.year ●● date.month ●●● date.day_of_week	Hierarchical Dimension
Quarter of signing	● date.year ●● date.quarter	Hierarchical Dimension
Start date	date	Dimension
Year of start of contract.	date.year	Dimension attribute
Quarter of year of start of contract	date.quarter	Dimension attribute
Month of start of contract.	date.month	Dimension attribute
Day of week of start of contract.	date.day_of_week	Dimension attribute
End date	date	Dimension
Year of end of contract.	date.year	Dimension attribute
Quarter of year of end of contract	date.quarter	Dimension attribute

Month of end of contract.	date.month	Dimension attribute
Day of week of end of contract.	date.day_of_week	Dimension attribute
Client	client	Dimension
First name of the client.	client.first_name	Dimension attribute
Last name of the client.	client.last_name	Dimension attribute
Estate location - hierarchy	● location.country_name ●● location.city_name	Hierarchical dimension
Estate location	location	Dimension
Country name.	location.country_name	Dimension attribute
City name.	location.city_name	Dimension attribute
Other	Junk	Dimension
The form of payment.	junk.payment_form	Dimension attribute
The type of contract.	junk.contract_type	Dimension attribute
The duration of the contract.	junk.contract_duration	Dimension attribute

4. Checking the feasibility of queries based on a multidimensional model

Analytical problem: What affected the change in the number of customers, when compared to the previous month?

1. Check 10 the best clients in previous month.

Measure: Fee amount

Dimension: client (attributes: id, first_name, last_name)

Dimension: date (attributes: id, year, month)

2. Compare the number of clients in different quarters, months and days of week.

Measure: Number of clients

Dimension: date (attributes: year, quarter, month, day_of_week)

3. Check which type of payment form was the most frequently chosen depending on type of signed contract in previous month.

Measure: Number of signed contracts

Dimension: junk (attributes: payment_form, contract_type)

Dimension: date (attributes: year, month)

4. Check the most popular duration of contracts signed in previous month.

Measure: Number of signed contracts

Dimension: junk (attributes: contract_duration)

Dimension: date (attributes: year, month)

5. Check the most common type of contract made in the previous month.

Measure: Number of signed contracts

Dimension: junk (attributes: contract_type)

Dimension: date (attributes: year, month)

6. Which type of offer was the most popular among the clients in previous month?

Measure: Number of signed contracts

Dimension: estate (attributes: id, floor_space, price, estate_type, garage, market, garage, pets_allowed)

Dimension: date (attributes: year, month)

Analytical problem: What affected the change in the profit, when compared to the previous month?

1. Calculate and compare the average fee amount in different quarters, months and days of week.
Measure: Fee amount
Dimension: date (attributes: year, quarter, month, day_of_week)

2. Calculate and compare profits in different quarters, months and week days
Measure: Profit
Dimension: date (attributes: year, quarter, month, day_of_week)

3. Check 10 the best employees in previous month.
Measure: Fee amount
Dimension: employee (attributes: id, first_name, last_name)
Dimension: date (attributes: year, month)

4. Check which estate type was the most profitable in previous month.
Measure: Average profit
Dimension: estate (attributes: estate_type)
Dimension: date (attributes: year, month)

5. Compare the average profits gained from customers who are looking to sell their properties and those who are having difficulty purchasing one in previous month.
Measure: Average profit
Dimension: junk (attributes: contract_type)
Dimension: date (attributes: year, month)

6. Compare the average profits gained by employees with higher education level and those with medium education level in previous month.
Measure: Average profit
Dimension: estate (attributes: estate_type)
Dimension: date (attributes: year, month)

7. Compare the average fee percentage in different cities in previous month.
Measure: Average fee percentage
Dimension: location (attributes: country_name, city_name)
Dimension: date (attributes: year, month)

5. Checking if there are data in the data sources that can fill the data warehouse

Table name	Attribute	Data source
contract (fact table)	Tuples correspond to the facts of contract sign.	
	client_id	Foreign key from dimension table. Its value comes from table client in main database.
	employee_id	Foreign key from dimension table. Its value comes from table employee in main database.
	estate_id	Foreign key from dimension table. Its value comes from table estate in main database.
	junk_id	Foreign key - generated by the database.
	fee_percentage	Its value comes from table contract in main database.
	fee_amount	Fee amount is calculated by multiplication of price(from estate table) and fee_percentage (from contract table).
	start_date	Foreign key from dimension table. Its value comes from table contract in main database.
	end_date	Foreign key from dimension table. Its value comes from table contract in main database.
date (dimension table)	Tuples correspond to specific dates. All data in this table are generated tuple after tuple on the basis of any calendar before the ETL process.	
employee (dimension table)	Tuples correspond to personal information of employees.	
	id	Surrogate key - generated by the database.

	first_name	First name comes from Column C of the Employee excel sheet.	
	last_name	Last name comes from Column D of the Employee excel sheet.	
	education_lvl	Education level comes from Column F of the Employee excel sheet.	
client (dimension table)	Tuples correspond to personal information of clients.		
	id	Surrogate key - generated by the database.	
	first_name	First name of client comes from client table in main database. It is the part before the last whitespace of the client_name attribute.	
	last_name	Last name of client comes from client table in main database. It is them part after the last whitespace of the client_name attribute.	
estate (dimension table)	Tuples correspond to information of real estates.		
	id	Surrogate key - generated by the database.	
	location_id	Foreign key from dimension table. Its value comes from table estate in main database.	
	floor_space	Floor space comes from estate table in main database. It is classified into one of the allowed ranges.	
	price	Value of price comes from contract table in main database. It is classified into one of the allowed ranges.	
	estate_type	Estate type comes from estate_type table in main database.	

		market	Market comes from estate table in main database.
		garages	Garages comes from number_of_garages table in main database. It is converted to a boolean value.
		pets_allowed	Information on whether pets are allowed comes from estate table in main database.
location(dimension table)	Tuples correspond to location (country, city).		
		id	Surrogate key - generated by the database.
		country_name	Name of country comes from country table in main database.
		city_name	Name of city comes from city table in main database.
junk (dimension table)	Tuples correspond to junk attributes.		
		id	Surrogate key - generated by the database.
		payment_form	Form of payment comes from payment_form table in main database.
		contract_type	Type of contract comes from contract_type table in main database.
		contract_duration	The duration comes from the difference between start_date and end_date in contract table in main database. It is classified into one of the allowed ranges.