Data Engineering Technical Test

Here at Monese we deal with structured and unstructured data we extract from source database and load to Data Warehouse. This exercise is designed to test data modelling knowledge with data loading elements.

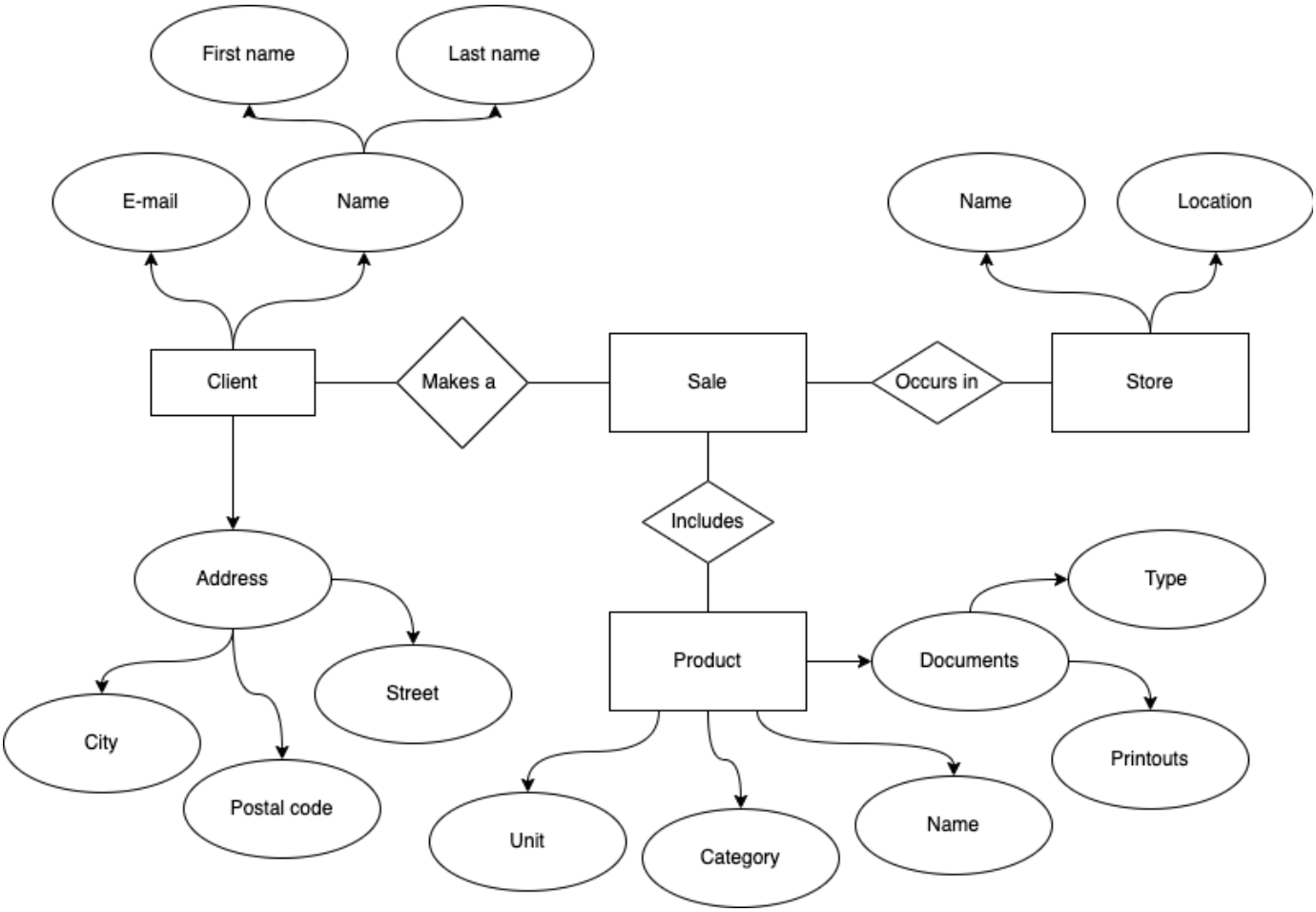
Data Modelling task

Basically the business case is - we need a data model that records data about sales transactions that contains information about product that was sold, client who bought a product and store where sale took place. Sales department is using a reporting tool to create dashboards and reports top of data stored in data warehouse, and they expect lag between sales system and data warehouse is not bigger than 1 hour (all sales that are happen more than 1 hour ago are presented in dashboards and reports).

Mandatory and known information about the sale event is as follows:

- Client name (first name and last name) and address (street, city and postal code).
- Product name, category, unit and related documents. Documents can be for instance user guide, safety instruction, sales license etc and document may have multiple printouts.
- Store name and location.
- Order has information about client, product and store.

Please create based on the conceptual data diagram a data model on third normal form as source for ETL process, Data Vault 2.0 mode as main model of data warehouse and a dimensional data model. There is no need to develop ETL code however please describe in general how data is transferred from one data layer to another, and what is purpose of added data layer.



Feel free to add any additional field You consider necessary. If anything seems to be unclear in the task description or you have an open question make your best guess and note it in your response (what the question was and what you assumed).

We are not looking for a perfect nor an absolutely correct solution. We are more interested instead of that how you approach the problem and what is Your style of thinking.

You can use for data modelling any tool you like.

Python task

Please have look into attached .zip with a python code. Your task is:

1. Comment/reorganise the main_function as you see fit
2. Describe what each function does
3. How would "sql_list" look like, roughly?
4. Return the code with any changes required for it to work in your computer

Summary

Save your response as a pdf file and Good Luck. We are looking forward to seeing your solution.