

## The Question-Answering Problem

Author: Luana Bezerra Batista

### Definition of the Problem and Storing Strategy

In a Question-Answering problem, a question needs to be stored with its respective answers. When a new question arrives, ideally, we want to check if a similar question was already asked before. If so, the question will not be stored as a new one. Instead, the answers for the similar question will be output to the user. This is a NLP (Natural Language Processing) solution for avoiding storing all questions asked in question-and-answer sites such Quora and Stack Overflow. Nevertheless, even using this NLP solution, the set of questions/answers might become very big overtime. Therefore, a Big Data architecture would be the most suitable way of dealing with the increasing amount of data to be stored and accessed in this type of system.

Regarding the database definition, different options have been considered. For instance, a Key-value-based NoSQL would be a good choice if the system allowed only one answer per question. However, a same question can have more than one answer. In a document-based NoSQL (or in a simple file-based system), a question with its respective answers could be grouped in a same data structure (e.g., a json file), under a same key. However, adding new answers to a question (as well as retrieving answers from a question) wouldn't be trivial, since the answers need to be ranked by using a voting scheme. Therefore, a traditional Relational DB (MySQL) has been chosen for our proposed Question-Answering problem.

Regarding the data properties, despite the choice of using a Relational DB, this system can behave well with a BASE Consistency Model. In the following, we define our conceptual (Figure 1) and logical models (Figure 2).

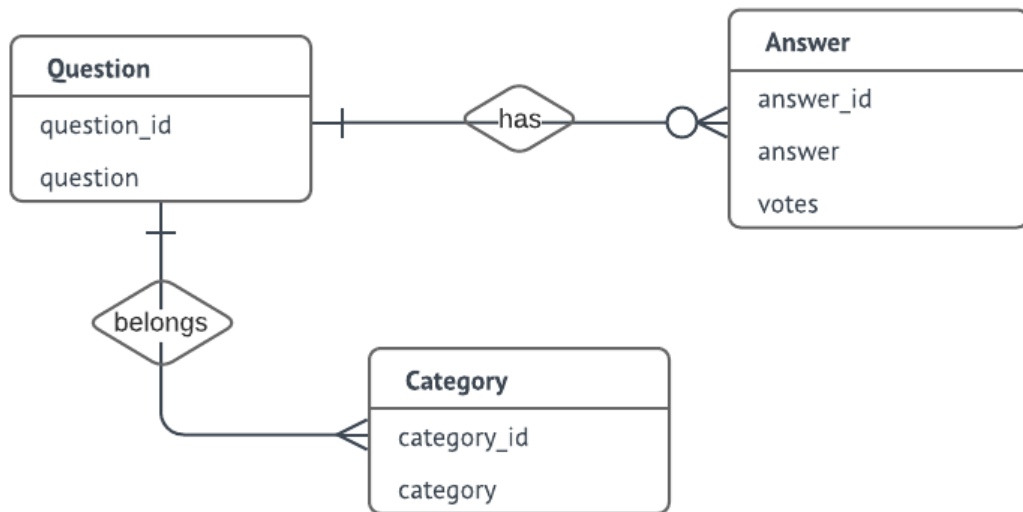


Figure 1 – Conceptual Model

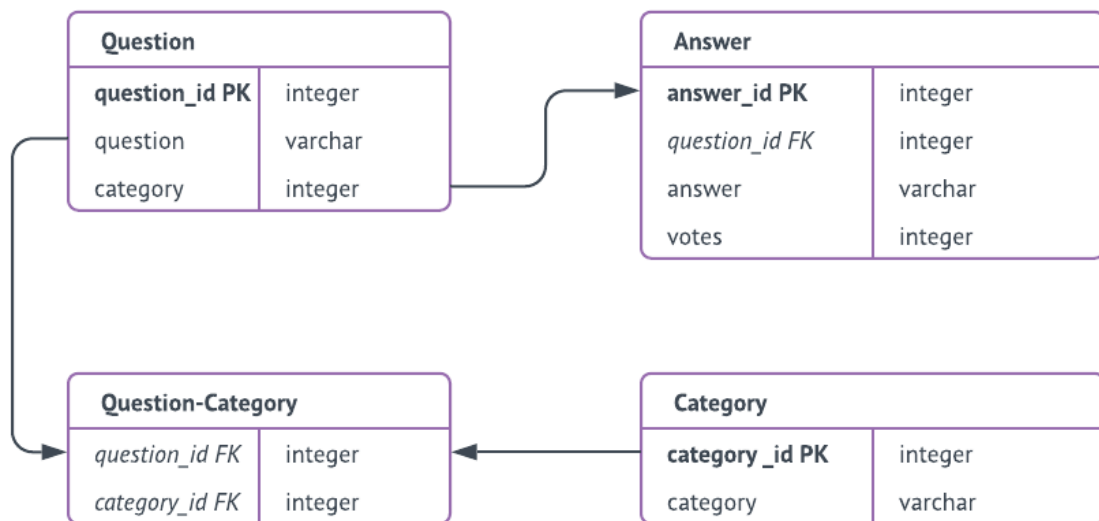


Figure 2 – Logical Model, where all attributes are mandatory (not null)

As for the DDLs, please refer to the SQL script:

[https://github.com/luana-be/CEB1250\\_repo/blob/master/QUESTION-ANSWERING.sql](https://github.com/luana-be/CEB1250_repo/blob/master/QUESTION-ANSWERING.sql)