SYSTEMS AND METHODS FOR BIG AND UNSTRUCTURED DATA

Prof. Marco Brambilla

POLITECNICO MILANO 1863

June 9, 2022

Last Name _____ First Name _____ Codice Persona _____

The municipality of Milan would like to implement a database to keep track of the people borrowing books from the city library network.
The system stores books (described by ISSN, title, description, publication date, and number of pages), their author (described by name, surname, birthdate, and a unique identifier). Each book has one author and a publisher (described by an identifier, a name, and an email). For each book, one or more copies can be available in one or more libraries of the network. Each book copy has an ID and a status (borrowed or available) and is associated with the respective book, the library and the shelf where it is stored. Each shelf has a book category (described by a name, a small description and an identifier). A book of a specific category can only be placed on a shelf with the same category. People can borrow books (up to 5 at a time). Notice that people search and want to borrow books, but actually one specific copy is associated with the borrowing. A person is described by name, surname, address, email, and library card number. When someone borrows a book copy, an entry is added to the system, storing the date in which the book copy has been borrowed and the date in which it should be returned. When a book copy is returned, the return date is also stored and the status is updated from “booked” to “returned”.

Exercise 1 (3 PT)

1.1. Describe the conceptual model of the data using an Entity-Relationship model (in the next page). Focus on conceptualising the problem and avoiding redundancy. Add appropriate entities, relationships, cardinalities, and attributes (2 PT).

1.2. Highlight which parts of the model you would implement in different database solutions (relational or non-relational, specifying the type of non-relational). Briefly motivate the choices. (1 PT)

#	ENTITIES / RELATIONSHIPS	DB TYPE	MOTIVATION
1			
2			
3			

1.1. ER MODEL: (make sure it’s readable and tidy)

Exercise 2 (5 PT)
Pick the entities Author, Book, Copy, Shelf and Library (and their relationships) from the ER model and suppose you want to store the respective data instances in a graph database. Sketch a graph model example describing the nodes and edges of the resulting graph model you would implement for them. Either show an example graph or a graph with types. (1PT)

2.2. Write a Cypher query to extract the books whose author's name is "Michel" that are currently misplaced (i.e., they are currently placed on a shelf with a book category different from the book one). If needed you may add details to the above graph model. (2 PT)

2.3. Find the list of all the books that have more than two copies assigned to at least one library of the network. (2 PT)

POLITECNICO MILANO 1863

February 10, 2022

Last Name _____ First Name _____ Codice Persona _____

Exercise 3 (5 PT)

Suppose you store in a documental database (MongoDB) the lists of Shelves, Copies and Books. Provide a simple documental representation. How many collections would you define? How would you implement the relations between the concepts (shelves, copies and books)? For writing the queries, provide a name to your collections and be coherent with your ER diagram. (1 PT)

3.1. Write a query to extract all the shelves containing at least one misplaced book with a publication date before the “21/08/1992”. (1 PT)

3.2. Write a query to count the number of books with more than 100 pages. (1 PT)

3.3. Write the query to find the average number of pages of the books placed on a shelf whose (i.e., of the shelf) category is “Action”. (2 PT)

Exercise 4 (3 PT)
Suppose you store an Elasticsearch index of the books, coherently with your ER diagram.

4.1. Provide the complete mapping of the index (i.e., field name, field type, the structure of the mapping, etc.) (1 PT)

PUT …

4.2. Write the complete query to extract all the books with category “Thriller”, whose title contains the terms “history”, prioritizing those whose title contains the word “spy”. (2 PT)