POLITECNICO
MILANO 1863

Last Name _____ First Name _____ Codice Persona _____

Theory (6 PT)

Choose whether the following statements are true or false by marking the CORRECT answers with an X.
Each correct answer awards 0,5 PTs. Each incorrect answer penalises -0,25 PTs. Missing answers count 0 PTs.

	STATEMENT	TRUE	FALSE
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			

SYSTEMS AND METHODS FOR BIG AND UNSTRUCTURED DATA
Prof. Marco Brambilla

POLITECNICO
MILANO 1863

August 30, 2022

Last Name _____ First Name _____ Codice Persona _____

Consider the following ER model describing the structure of a Music application.
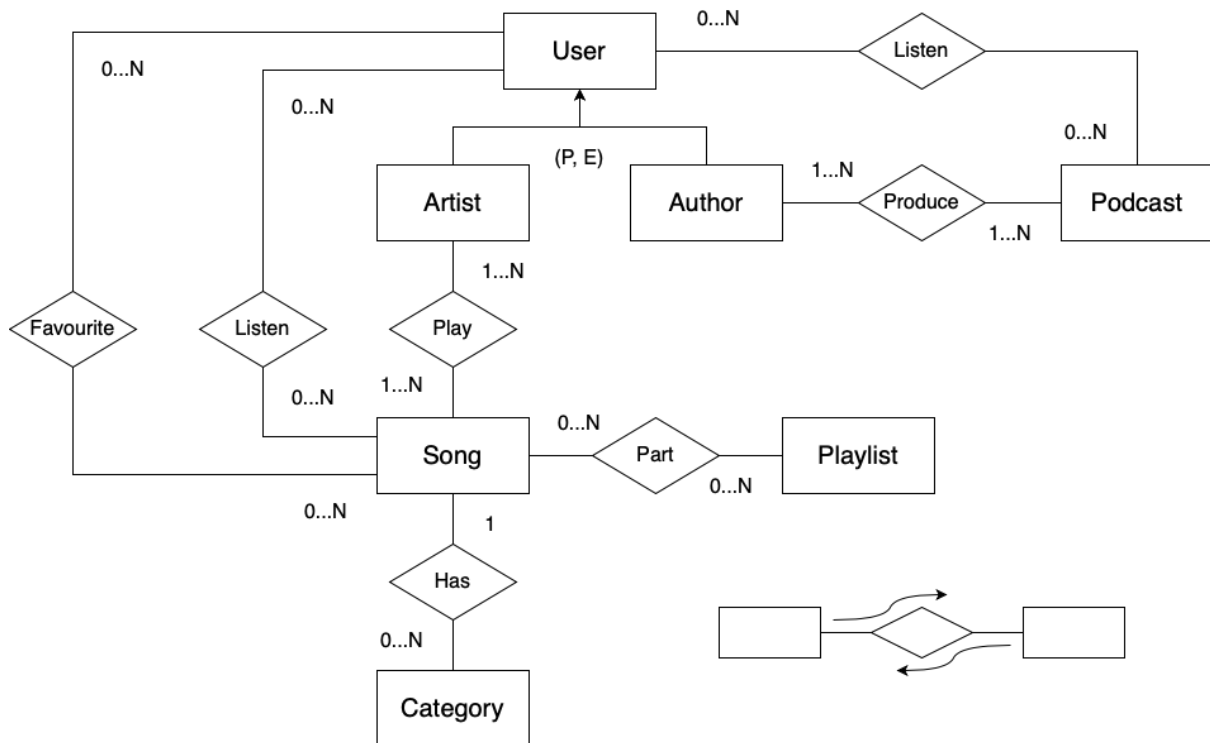
The following attributes describe the entities in the diagram
- User/Artist/Author – <u>Mail</u>, Username, Member Since, Premium (Yes/No)
- Podcast – <u>ID</u>, Title, Duration (in seconds), Publication Date
- Song – <u>ID</u>, Title, Duration (in seconds), Publication Date, Text
- Playlist – <u>ID</u>, Title, Public (Yes/No), Creation Date
- Category – <u>Name</u>, Description

The following attributes describe the relationships in the diagram
- Favourite – Since
- Listen (Song) – Date, Complete (Yes/No)
- Part – Since
- Listen (Podcast) – Date, Complete (Yes/No)

Exercise 1 (1 Point)

1.1. In the table below, identify 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			

Exercise 2 (5 PT)

2.1. Consider the entities Song, Playlist, Artist, and Category 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, main attributes, and edges. Either show an example graph or a graph with types. (1 PT)

2.2. Write a Cypher query to extract all the pairs of songs sung together by "Laura Pausini" and "Tiziano Ferro", whose publication date is after 01/01/2015, that do not belong to the same category. (2 PT)

2.3. Write a Cypher query to extract the total number of songs published for each category by artists who have been members since at least 01/01/2015. (2 PT)

SYSTEMS AND METHODS FOR BIG AND UNSTRUCTURED DATA
Prof. Marco Brambilla

POLITECNICO
MILANO 1863

August 30, 2022

Last Name _____ First Name _____ Codice Persona _____

Exercise 3 (6 PT)

Suppose you store in a documental database (MongoDB) the User/Artist/Author, Song, and Podcast. How many collections would you define? How would you implement the relations between the concepts? Provide a simple documental representation. (1 PT)

3.1. Write a query to extract all the authors that have been members since before 30/04/2020, for which at least one of their podcasts' duration is greater than 600 seconds. [If you opt for multiple collection in your model, write the query starting from **Author_Collection**.] (1 PT)

3.2. Write a query to count the number of songs released by each artist with a duration of at least 180 seconds, finally considering only the artists that published at least 25 of these songs. [Write the query starting from **Artist_Collection**.] (2 PT)

3.3. Write the query to find the number of podcasts for which at least one listener is a premium member and has been a member since at least 02/01/2023. [If you designed different collections, write the query starting from **Podcast_Collection**.] (2 PT)

Exercise 4 (4 PT)

Suppose you store an Elasticsearch index for the Songs.

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 the songs from the artist named "Pinguini Tattici Nucleari", whose text contains the words "Notte" and "Fuoco" while assigning a higher score to those that contain the word "Bergamo". (1.5 PT)

4.3. Write the complete query to extract the number of songs each artist publishes. (1.5 PT)