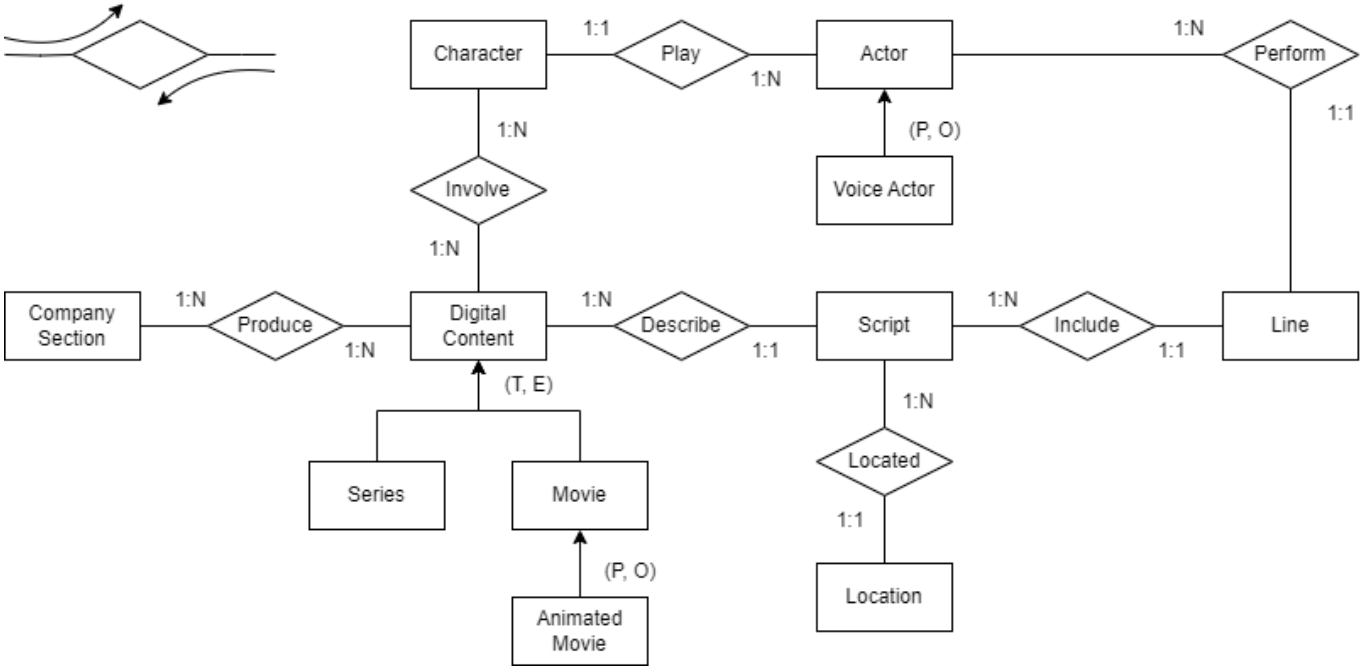
SYSTEMS AND METHODS FOR BIG AND UNSTRUCTURED DATA - Prof. Marco Brambilla - June 24, 2024

Last Name (Surname, Cognome) _____ First Name _____ Codice Persona _____

Consider the ER Diagram representing a data structure for tracking Disney's digital products.



The following attributes describe the entities. The primary keys are underlined.

- **Company Section** - ID, Name, Description, Number of Employees
- **Digital Content** - Title, Release Date, Duration, Production Cost, Minimum Age
    - **Series** - Number of Episodes, …
    - **Movie** - Sequence Number, Nominated for Award (Yes/No), …
        - **Animated Movie** - …
- **Character** - Name, Nickname, Age
- **Actor** - Personal ID, Name, Surname, Age, Birth Date, Career Description
    - **Voice Actor** - Voice Tone, Voice Strength (from -5 to 5)
- **Script** - ID, Length, Duration, Title
- **Location** - ID, Name, Description, Furniture (Text), Indoor (Yes/No)
- **Line** - ID, Sequence Number, Text

The following attributes describe the relationships.

- **Perform** - Recording Date
- **Produce** - Starting Date, End Date

Exercise 1 - Unstructured data models (1 PT)

Which DB technology would be the best to store the entities/rel. of the ER diagram? [you may not need all the rows of the table] (1 PT)

ENTITIES / REL.	DB TYPE	MOTIVATION

Exercise 2 - Neo4j (5 PT)

Consider the entities Character, Actor, Voice Actor, and Line from the ER model. 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)

Write a Cypher query to collect the total number of lines that were performed by actors named “Mark” who performed at least 100 lines with at least three characters (2 PT)

Write a Cypher query to return the collection of voice actors who played at least 15 characters nicknamed “Jack” and who acted at least once as “Morpheus”, performing at least 1000 lines in the latter role. (2 PT)

Exercise 3 - MongoDB (7 PT)

Consider the entities Company Section, Digital Content, Series, Movie, Animated Movie and Script (and all its child entities). How many collections would you define within a documental database? How would you implement the relations between the concepts? Provide a simple documental representation. (1 PT)

Write a query to return the number of digital content generated by each company section, keeping only the ones who produced at least 20 of them. (1.5 PT)

Write a query to count the number of series produced by a company section named “Digital Artist & Co” with ID “908686” and for which at least one script includes the word “Pier” in its title. (2 PT)

Write a query to return the total count of the scripts of animated movies produced by at least one company section named “Animation”, for which at least one script has a duration greater than 100 minutes and a title equal to “Cinderella”. (2.5 PT)

Exercise 4 - Elasticsearch (4 PT)

Consider the Location entity.

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

PUT …

Write a query to return the list of Locations whose description includes the words “Stadium” and “Basketball”, assigning a higher score to those including the words “Planet” in their furniture and “Space” in their description. (1.5 PT)

Write a query to return the count of indoor and outdoor locations. (1.5 PT)

Exercise 5 - Redis (3 PT)
Consider the Line table.

Write the Redis code to create a new list of lines involving the lines “Hey you”, “Watch out”, and “Hello there”. (1 PT)

Write the Redis operation to collect the first three lines in the list. (1 PT)

Write the Redis operations to create a new key-value pair named “Line #0001” with the value “A long time ago in a galaxy far, far away” and check its lifespan. (1 PT).