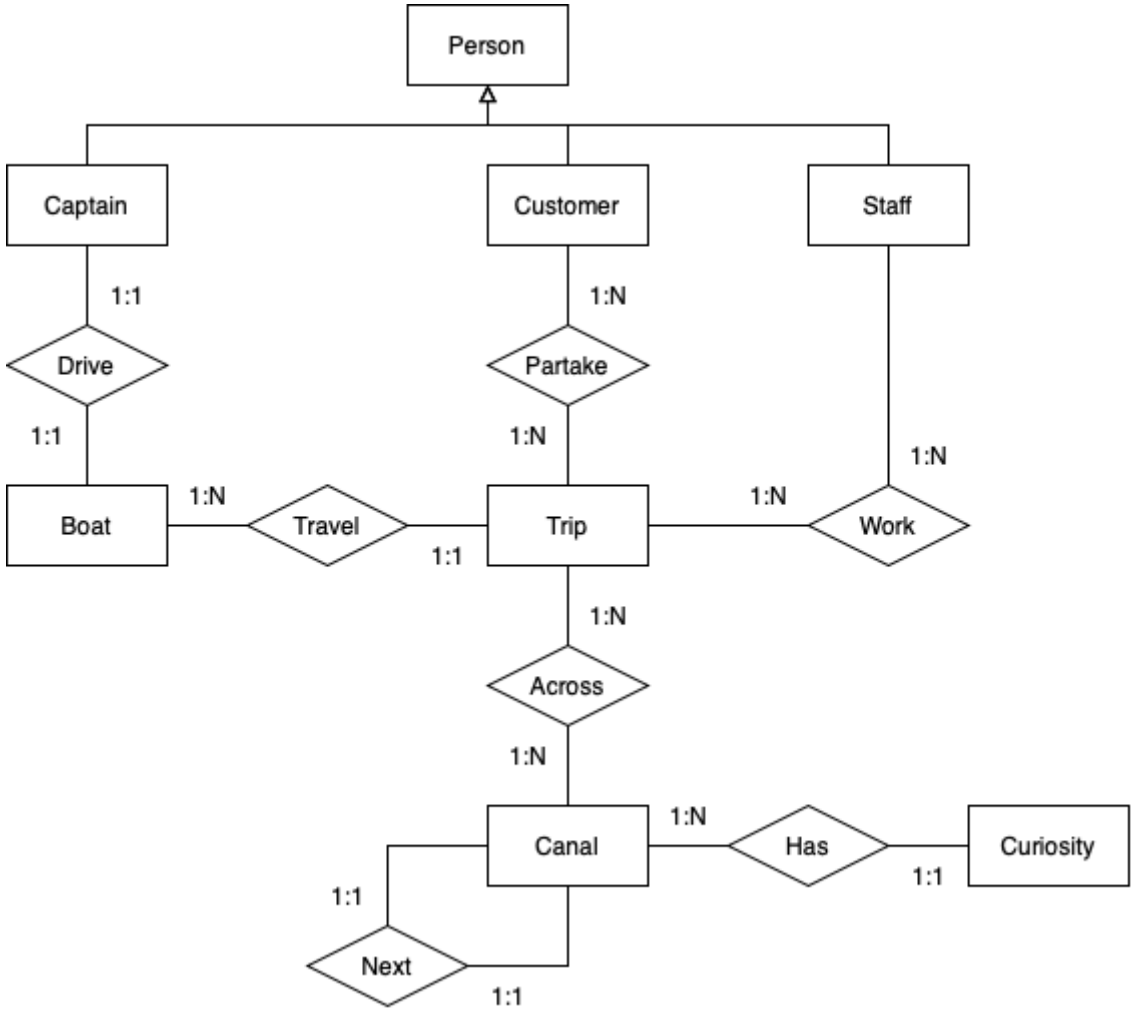
SYSTEMS AND METHODS FOR BIG AND UNSTRUCTURED DATA
Prof. Marco Brambilla

POLITECNICO MILANO 1863

September 6, 2023

Last Name _____ First Name _____ Codice Persona _____

Consider the following ER model describing the structure of a Dutch Canal Cruise company.



The following attributes describe the entities in the diagram

- **Captain** – ID Card Number, Name, Surname, Birth Date, Hiring Date, Boat License Number
- **Staff** – ID Card Number, Name, Surname, Birth Date, Hiring Date
- **Customer** – ID Card Number, Name, Surname, Birth Date,
- **Boat** – ID, Number of Seats, Length, Width, Last Review Date
- **Trip** – ID, Date, Begin Time, End Time
- **Canal** – Name, Width, Length, Depth
- **Curiosity** – ID, Description

The following attributes describe the relationships in the diagram

- **Next** – Travel Time

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 Customer, Trip, Canal, and Boat 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 the number of boats that cruise each canal, considering only the trips whose begin time is later than 12:00. Consider that a trip is associated with each of the canals the boat cruises. (2 PT)

2.3. Write a Cypher query to extract the number of customers with the same name that partook in a trip on the same day for each name. (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 Staff Members, Customers, and Trips. How many collections would you define? How would you implement the relations between the concepts? Provide a simple documental representation. [Ignore the “write the query starting from…” conditions enforced by the following requests when performing **this part of the exercise**] (1 PT)

3.1. Write a query to extract the number of customers participating in each trip, returning only the trips whose customer count exceeds 50. [Write the query starting from **Trip_Collection**.] (1.5 PT)

3.2. Write a query to count the number of trips performed by each staff member named “Margherita” or who was born after 12/09/98. [Write the query starting from **Trips_Collection**.] (1.5 PT)

3.3. Write the query to find the customers who partook in at least one trip that left the dock after 10:00 and at least one trip that returned at the dock before 15:00. [Write the query starting from **Customer_Collection**.] (2 PT)

Exercise 4 (4 PT)
Suppose you store an Elasticsearch index for the Curiosities.

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 curiosities whose text contains the words “People”, “Dirt”, and “Canal”, excluding the ones that contain the “Excavation” word. (1.5 PT)

4.3. Write the complete query to extract the number of curiosities about each canal. [Just for this request, consider the existence of a further attribute named “canal_id”] (1.5 PT)