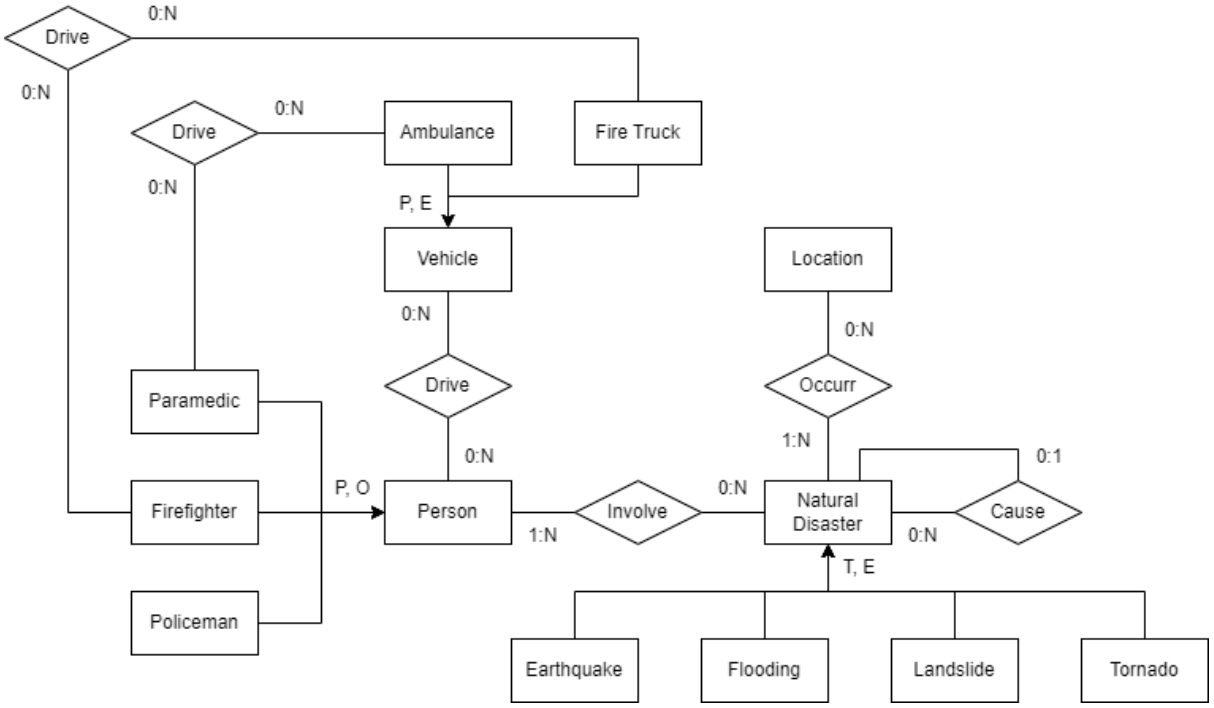
SYSTEMS AND METHODS FOR BIG AND UNSTRUCTURED DATA - Prof. Marco Brambilla - June 6, 2024

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

Consider the ER Diagram representing a data structure for tracking natural disasters.



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

- **Natural Disaster** - ID, Start Time, Start Date, End Time, End Date, Disaster Description
    - **Earthquake** - Magnitude (1 to 10), Mercalli Scale (1 to 12)
    - **Flooding** - Water Height
    - **Landslide** - Landslide Class (1 to 5)
    - **Tornado** - EF Scale (0 to 5)
- **Location** - Postal Code, City, Country, Description
- **Person** - Personal ID, Name, Surname, Postal Code,
    - **Paramedic** - Paramedic ID, Training Level (Novice, Intermediate, or Expert)
    - **Firefighter** - Firefighter ID
    - **Policeman** - Policeman ID, Rank
- **Vehicle** - Plate, Number of Seats
    - **Ambulance** - Stretcher (Yes/No), Defibrillator (Yes/No)
    - **Fire Truck** - Water Capacity, Ladder (Yes/No)

Note → A paramedic can be a firefighter (the same applies to the other entities involved in that ISA relationship).

The following attributes describe the relationships.

- **Drive** (Firefighter – Fire Truck) - Driving License Number
- **Drive** (Paramedic – Ambulance) - Driving License Number
- **Cause** (A Natural Disaster may cause another Natural Disaster) - Causal Description

Exercise 1 - Unstructured data models (1 PT)

Which DB technology would be the best to store the entities/rel. of the ER diagram? (1 PT)

ENTITIES / REL.	DB TYPE	MOTIVATION

Exercise 2 - Neo4j (5 PT)

Consider the entities Person, Paramedic, Firefighter, Vehicle, Ambulance, and Fire Truck 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 number of people who drive at least 5 different vehicles, one of which should either be an Ambulance or a Fire Truck (2 PT)

Write a Cypher query to return the collection of Fire Trucks with a water capacity greater than 100 litres that were driven by at least 5 Firefighters named John, who also drove 10 non-Fire Truck vehicles each. (2 PT)

Exercise 3 - MongoDB (7 PT)

Consider Location and Natural Disaster (with 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 list of earthquakes with a magnitude greater than 8 that also caused at least one flooding with the flooding's causal description equal to "very strong earthquake", which in turn caused two landslides. (2 PT)

Write a query to count the number of natural disasters that started after 01:00 AM on 19 April 2023 and ended before 7:30 PM on the same day. Consider only the ones that happened in Italy that affected more than 10 locations. (2 PT)

Write a query to return the list of Tornadoes with an EF Scale greater than 3 that affected at least one location whose Country is Italy and whose postal code is 23563 or that affected Milan. (2 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 all the locations whose description includes the words “Beautiful” and “Town”, removing the ones whose name includes the word “Milano”. The condition on the word “Town” should not affect the final score. (1.5 PT)

Write a query to return the count of the locations for each country. (1.5 PT)

Exercise 5 - Cassandra (3 PT)
Consider the Policeman table. Write a Cassandra script to perform the operations listed below.

Create the Policeman table with Personal_ID as the partition key and Policeman_ID as the clustering key. (1 PT)

Write a CQL query to collect the Policemen whose Surname is “Rossi”. If any further operation is required, write its CQL code. (1 PT)

Write the CQL script to be written in an “operations.txt” file to insert two new policemen (you can add attribute values of your choice) and write the CQL operation to run the whole script from the file. (1 PT)