

# Data Management for Data Science – 31/5/2017

LAST NAME: .....

FIRST NAME: .....

ID (MATRICOLA): .....

Autorizzo la pubblicazione del mio voto di questo esame sul sito web <http://www.dis.uniroma1.it/~rosati/gd>, secondo quanto prevede il decreto legislativo 196/2003 (codice in materia di protezione dei dati personali) che dichiaro di conoscere. In fede,

.....

**Exercise 1** We want to build a relational database about the domain of movies. In particular, we want to store information about **movies**, **directors**, **actors**, **location** in three distinct relations. Moreover, a relation **isDirectorOf** stores information about the directors of movies, a relation **playsIn** stores information about the actors of movies, and a relation **filmedIn** stores information about the locations where a movie has been filmed.

1. Write SQL statements that define the schema of the above described database;
2. Write SQL statements that insert at least two tuples in each of the tables defined at the previous point;
3. Write SQL statements that express the following queries:
  - (a) return the titles of the comedies filmed between 2010 and 2016;
  - (b) return the name of every directors who has directed herself/himself as an actor in a movie;
  - (c) return the title and year of the movies directed by Italian directors and played by at least one non-Italian actor;
  - (d) for every actor, return the number of movies played by that actor and filmed in Italian locations.

## Exercise 2

1. Define a file organization for the relations of the schema of Exercise 1 such that the execution of the SQL query relative to point (a) of Exercise 1 is optimized;
2. define a file organization for the relations of the schema of Exercise 1 such that the execution of the SQL query relative to point (b) of Exercise 1 is optimized.

**Exercise 3** We want to store a multidimensional structure containing the following information about products:

- quantity (number of items sold)
- type (entertainment, industrial, electronics, food, furniture, hardware)

over the following dimensions:

- Time (datetimestamp, day, week, month, quarter, year)
- Customer (name, city, country, subregion, region, continent)

1. Define a star schema to represent the above multidimensional structure;
2. define a snowflake schema that reduces on both dimensions the redundancy of the star schema defined at the previous point;
3. write an SQL query over the star schema defined at point 1 that returns the quantity sold and the customer name of the food items sold in January 2016;
4. write an SQL query over the snowflake schema defined at point 2 that returns the quantity sold and the customer name of the food items sold in January 2016;
5. write an SQL query over the star schema defined at point 1 that returns, for every European country, the number of electronic items that were sold in that country from January 2014 to December 2015.

## Exercise 4

- (a) Write an RDF model representing the following statements about URIs **Person**, **Director**, **Actor**, **Country**, **Movie**, **Comedy**, **Drama**, **Male**, **Female**, **filmedIn**, **isDirectorOf**, **playsIn**, **bornIn**, **Ingrid**, **Cary**, **Alfred**, **Notorious**, **USA**, **Brazil**, **UK**.
  1. Alfred is the director of Notorious;
  2. Ingrid and Cary play in Notorious;
  3. Notorious was filmed in the U.S.A. and in Brazil;
  4. Alfred was born in the U.K.;
  5. Ingrid is female;
  6. Cary is male.
- (b) Write SPARQL queries corresponding to the following requests: (a) return the titles of the comedies filmed in the U.K.; (b) return the name of every director who has directed herself/himself as an actor in a movie. (c) return all the countries where at least one comedy or one horror movie were filmed.

**Exercise 5** Consider the information requested in point 2 of Exercise 1. Express such an information using a graph database (e.g., Neo4J).

**Exercise 6** Consider the information requested in point 2 of Exercise 1. Express such an information using the MongoDB data model (JSON). Moreover, express the query (a) at point 3 in Exercise 1 as a query in the MongoDB system.