Data Management for Data Science – 7/6/2016

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 (name, ID, year, director-ID, country), directors (name, ID, birthdate, nationality), actors (name, ID, birthdate, nationality). Moreover, a relation isDirectorOf (director-ID, movie-ID) stores information about the directors of movies, and a relation playsIn (actor-ID, movie-ID) stores information about the actors of movies.

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 names of all the Italian directors;
   (b) return the names of the movies filmed in the U.S.A. and directed by Italian directors;
   (c) return the names of the actors who played in movies directed by Italian directors;
   (d) return the number of movies played by Robert De Niro.

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 (c) of Exercise 1 is optimized.

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

- quantity (number of millimeters)
- type (rain, snow, hail, sleet)

over the following dimensions:

- Time (hour, day, week, month, quarter, year)
- Location (city, region, country, 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 types of precipitations in Rome from January 2014 to June 2014;
4. write an SQL query over the snowflake schema defined at point 2 that returns the types of precipitations in Rome from January 2014 to June 2014;
5. write an SQL query over the star schema defined at point 1 that returns, for evry Italian city, the quantity (number of millimeters) of snow that fell in that city from January 2000 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`, `actsIn`, `bornIn`, `Joe`, `Mary`, `Ann`, `Paul`, `Italy`, `France`, `ABC`, `XYZ`.

1. Ann is the director of movie XYZ;
2. Joe and Paul act in movie ABC;
3. ABC was filmed in France;
4. Paul was born in Italy;
5. Ann is of type female;
6. Paul is of type male.

(b) Write SPARQL queries corresponding to the following requests: (b1) "return every comedy that was filmed in France by an Italian director; (b2) "return all the countries where at least one comedy and one drama were filmed"; (b3) "return the directors who directed at least one movie that was filmed in Italy or in France".

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 a noSQL document data model (e.g., JSON).