COMP 4111/5111 Data Management for Business Intelligence
Fall 2018 Assignment 1

Name: XXX   **Student Number: YYY**

Instructions:

1. **For your solution use this template file, and follow the instructions that follow. "Ignore" the instruction in the final file, i.e. put them in environment** `\ignore{ .... }`

   In particular: (a) Include at the top of the first page: full name, student number, and email address. (b) Assignments have to be created with Latex, and submitted in pdf format. (c) Every problem solution MUST include the problem statement. The source file for this assignment is provided.

   Latex has to be used as such, not as you would use a text editor, such as Notepad. In particular, formulas have to be written using Latex's mathematical features, and then compiled.

2. Assignments are individual, no groups.

3. Submit by email to the instructor, with "Assignment "Number", DMBI" in the subject. **Include your last name in the file name!** For example, in the subject: "Assig. 1 DMBI". The file name: "bertossi-1.pdf".

   Only a single pdf file will be accepted as submission. No tar, zip, Latex, etc. files (or anything like that) will be accepted. Keep your Latex source files in case you are requested to show them.

4. Explain your solution very carefully, but still be succinct with your answers. No unnecessary verbose arguments, please. Go to the point. Make explicit all your assumptions.

5. **Not following the instructions above or the solution template file will make you lose points.**

1. Assume you are interacting with a relational database with schema $\mathcal{R} = \{R(A, B), S(B, D)\}$ about which *you know* that it satisfies the integrity constraint (IC):

$$\varphi: \ \forall x \forall y (R(x, y) \rightarrow \exists z S(y, z)). \tag{1}$$

(a) Assume that, for security issues, you are allowed to pose queries expressed only in terms of predicate S. What query $\mathcal{Q}_0'$ can you pose as a "relaxed" version of the query $\mathcal{Q}_0(y): \ \exists x R(x, y)$, i.e. a query that allows you to obtain at least all the answers to $\mathcal{Q}$. Explain carefully and in detail, but to the point. [4 points]

Solution: ZZZZZZ

(b) How would you semantically optimize your query $\mathcal{Q}_1(x)\colon \exists y(R(x,y) \wedge \neg\exists z S(y,z))$, by rewriting in into a semantically equivalent query $\mathcal{Q}_1'$. Idem. [4 points]

(c) You are allowed to pose queries in terms of the view (predicate defined by) $V(x)\colon \exists y R(x,y)$. What query $\mathcal{Q}_3'$ would you pose to obtain answers to the query: $\mathcal{Q}_3(x)\colon \exists y \exists z(R(x,y) \wedge S(y,z) \wedge x > 5)$. Idem. [4 points]

(d) Give an instance for the schema that respects the IC, and illustrate each of the three items above. [6 points]

2. For the IC in (1):

(a) Indicate the *relevant* updates, i.e. those that could potentially lead to a violation of the IC (that this really happens may depend on the instance, but here we have only the schema). They can be insertions, deletions and changes of attribute values of a particular kind. Explain why for each case. [4 points]

(b) Define in relational calculus (predicate logic) and SQL a violation view for the IC. [4 points]

(c) Define in SQL an active rule (or trigger) that clearly contains the three components for the maintenance of the IC. It has to use the view you defined in the previous item. Before specifying it, explain in words what maintenance policy you decided to specify. [4 points]

Deadline: October 5, at 23:55