Data science and database technology

Politecnico di Torino

Exam policy A.Y. 2020/2021

(Covid-19 emergency)

The exam consists **exclusively of a written part**. The written test takes place with a PC using the exam university framework integrated with the Respondus proctoring tool.

- The correct execution of the exam on the Respondus tool will be verified and irregular behaviors will be prosecuted following the university rules.

Rules for the written test

- Only students who have regularly booked for the exam through the Didactic portal (Portale della didattica) are admitted to the written test.
- Students have to provide an **identity document including a photograph**.
- The written test lasts **80 minutes**.
- Books and notes **are NOT allowed**. Electronic devices of any kind (PCs, laptops, mobile phones, …), apart from the PC used to take the written test, **are NOT allowed**.
- 2 white sheets on both sides can be used for a draft (brutta copia). These white sheets must be shown in front of the camera before the written test begins.
- The desk must be empty except for the PC used to take the exam, the 2 white sheets used for the draft (brutta copia), and a pen.
- In case of connectivity problems (or other technical problems) the university rules in the "Guidelines for online exams" will be followed.

Structure of the written test

- 4-6 multiple-choice **theory** questions (max. total points 6/30).
- 1-3 multiple-choice questions on **physical design** (max. total points 5/30). The use of drawing tools is not required.
- 1-2 exercises on **trigger design** (max. total points 9/30). The exercises are structured as open questions. The student has to provide textual answers by writing the SQL code in a textual box.
- 1 exercise on **data warehousing** including:
    - 1-3 multiple-choice questions on **data warehouse design** (max. total points 3/30) The use of drawing tools is not required.

    - 2 **extended SQL queries** against a given relational schema (max. total points 8/30). The exercises are structured as open questions. The student has to provide textual answers by writing the SQL code in a textual box.
- In multiple-choice questions a penalty (negative score) is assigned for an incorrect answer. The negative score is equal to 15% of the question score. Missing answers are worth zero.

Topics of the multiple-choice theory questions

1. Data preparation: discretization, normalization, distance measures, etc.
2. Association rules: extraction algorithms (also on examples), types of itemsets (closed etc), quality indices
3. Classification: algorithms, model validation
4. Clustering: algorithms, quality indices (those presented in lessons)
5. Recovery: centralized and distributed recovery algorithms (also applied to examples of log files)
6. Concurrency control management: description of CSR, VSR, 2PL techniques (also applied to examples of schedules)
7. Distributed databases: data fragmentation and transparency levels, 2 phase commit, parallelism, benchmarking, non-relational databases

Homework

- Each homework exercise, delivered by the deadline, gives 0.5/30.
- The points on the homework exercises will be valid until the exam session of September 2021 (included).

Homework discussion

- Before each written exam, we will notify the names of the students involved in the discussion of the homework exercises, and when homework will be discussed (usually immediately after the written exam).

Final evaluation

- The exam is passed if a grade ≥ 18/30 is achieved on the written test.
- If in the written test a grade ≥ 18/30 is achieved, the final grade is obtained by summing the grade obtained on the written test and the points obtained on the homework assignments. A final grade greater than or equal to 31/30 without rounding corresponds to 30 with honors.
- If students fail the exam, the exam failure is recorded.