Large Scale and Multi-Structured Data Bases (9CFU)
University of Pisa
Academic Year 2024-2025

Written Test Example 04/12/2024

1) First, write the student name and surname on the text and on the answer sheets (both must be returned to the teacher).
2) Answer the questions clearly and concisely.
3) It is forbidden to use any type of material other than sheets and pens.
4) At the end of the time, the examination will be corrected by the teacher who will assign the assessment.
5) During the correction phase, the teacher may ask the student to clarify any aspects of the answers given.
6) Answers not provided will not be assessed and the student may not request to provide the answers during the correction phase.
7) Answers partially provided will be assessed based on the amount of information contained. The student will not be allowed to add more details to the provided answers during the correction phase.
8) Do not copy (*obvious, but it always better to repeat it!).*

Answer the following questions (10 points each):

1. Let's suppose to consider a one-to-many relationship between the ***student entity*** and the ***exam entity,*** namely one student can be associated to several exams and one specific exam instance is associated to a specific student. The most important attributes for the student entity are: 1) Name, 2) Surname, 3) DOB. As regards the exam entity, the most important attributes are: 1) Date, 2) Mark, 3) Subject . We want to design an application for allowing the student to manage his/her exams.
   We request to:
   a) Depict the UML analysis class diagram
   b) Define some functional requirements supporting the ***document embedding*** strategy for implementing a document DB for the application
   c) Define some functional requirements supporting the ***document linking*** strategy for implementing a document DB for the application
   d) Provide the structure of the collections considering the two alternative document DB implementations, including an example (in correct JSON format) with the documents storing the following information:

      Student1:
      Name: Pippo
      Surname: Pluto
      DOB: 23, Jan 2023
      Exam1:
      Date: 23, Feb 2022
      Mark: 23/30
      Subject: Data Mining
      Exam2:
      Date: 25, Dec 2022
      Mark: 30/30
      Subject: Large Scale

2. What is the key space of columnar DB? Provide details on Row Keys, Column and Columns families.

3. Discuss the differences between sharding and replication? How do the contribute to satisfy non-functional requirements of massive web-scale applications.