# Data Management (A.A. 2024/25) – exam of 11/07/2025

**Problem 1**
Consider a schedule $S$ with both shared and exclusive locks, that is legal and is such that each of its transactions enjoys the following properties: $(i)$ it is well-formed, $(ii)$ it has all the lock operations before the commit action, and $(iii)$ it has all the unlock operations after the commit action.

1.1 Provide the definition of view serializable schedule and prove or disprove that $S$ is view serializable.

1.2 Provide the definition of strict schedule, and prove or disprove that $S$ is strict.

**Problem 2**
Consider a system with processors $P_0, P_1, \ldots, P_{10}$, where $(i)$ $P_0$ has 10 buffer frames and stores the relation `Exam(student,course,mark)` with 1.000 pages, $(ii)$ each of the processors $P_1, \ldots, P_5$ have 100 buffer frames, and $(iii)$ each of the processors $P_6, \ldots, P_{10}$ have 20 buffer frames. Consider the query

```
select student,avg(mark)
from Exam
group by student
```

and answer the following questions:

2.1 Which algorithm would you use for computing the result of the query if you could use only processor $P_0$, and which is the cost of such algorithm in terms of number of page accesses?

2.2 Which algorithm would you use for computing the result of the query if you could use processors $P_1, \ldots, P_{10}$ for a parallel execution, and which is the cost of such algorithm in terms of both number of page accesses and elapsed time?

**Problem 3**
Consider the relations `R(A,B)` and `S(C,D,E)`, where $(i)$ both relations are stored in a heap with 9.000 and 30.000 pages, respectively, $(ii)$ the values appearing in the attribute `A` are the integers between 1 and 100 and such values are uniformly distributed over the tuples of `R`, $(iii)$ `C` is the primary key of `S`. Knowing that the buffer has 200 frames, illustrate the algorithm you would use to answer the query

```
select A,E
from R join S on R.B = S.C
where A > 50
order by E
```

and tell which is its cost in terms of number of page accesses.

**Problem 4**
Consider the relation `PRODUCT(code,category,supplier,cost)` that stores information about products sold by an e-commerce company, with the code, the category, the supplier and the cost of each product. The relation has 3.000.000 tuples, stored in 300.000 pages, and has 10.000 different values in the attribute `cost`, uniformly distributed in the tuples of the relation. We assume that all fields and pointers have the same length, independently of the attribute and that there is a sparse, clustering $B^+$-tree index on `PRODUCT` with search key `cost`, using alternative 2. Consider the query asking for the code and the category of the products whose cost is in a given range constituted by 10 values, and tell how many page accesses we need for computing the answer to the query.

**Problem 5** (only for students who opted for **option 1**, i.e., who do **not** do the project)
Consider an application storing data about customers, each with code, age, category and city of birth (with name, region and state) and the product they have bought, each with code, type, supplier (with fiscal code, name and city where the administration of the supplier is located), and the possible discount for the various categories of users. Choose an example of a data base suitable for such application (with instances of the various types of data mentioned above) and ilustrate the representation of such data base expressed in three data models::

5.1 the relational data model,

5.2 the document-oriented data model,

5.3 the graph data model.