DISTRIBUTED QUERY PROCESSING

OBJECTIVES

- Learn about distributed query processing. Namely:
    - The differences between centralized and distributed query processing,
    - Phases of the distributed query processing

ACTIVITIES

During the week, read and understand the following parts of the "Lab Material: Distributed Query Processing" document (you will find it in the additional material provided for the "Advanced Notions on Distributed Databases" lab):

- If needed, refresh how centralized query processing works (Section 1: "Basics of Query Optimization"),
- Read the section about distributed query processing (Section 2: from page 19 to page 30).

DELIVERABLES

Solve the following exercise (you can do it with one of your classmates), and hand in the solution at the beginning of the lecture on "Distributed Query Processing":

Consider the following distributed schema for the Project relation (primary key underlined).

Project(pno, name, head, budget, city):

- P1 = $\sigma_{budget<100000}(Project)$
- P2 = $\sigma_{100000\leq budget\leq 500000}(Project)$
- P3 = $\sigma_{budget>500000}(Project)$

You can assume that this fragmentation strategy is correct (i.e., complete, disjoint and reconstructible). Given the following query:

SELECT * FROM Project WHERE budget > 90000 AND budget < 200000;

Reproduce the data location (i.e., express the query in terms of fragments) and determine which steps the reduction phase would follow.