Project Phase 3 - <u>Query Optimizer and Executor</u>

Description:

- Support for aggregate (sum, min, max, average) and Group-By and having clauses. The data must be in three sites.
- Support for semi-join. You should use semi-joins for executing joins where data is at different sites.
- Support for executing select-project-join queries where data is located in three sites.
- Need selectivity factors for selections and joins.
- For selections, assume selectivity to be uniform (0.5 or 0.3 etc)
- For joins, maintain a file with the selectivity factors -
  - Join Selectivity Factor is the number of rows in the result of joining of two relations.
- Use standard heuristics (push selects and projects down and joins up) to reduce the time complexity
- Optimize the fragment queries generated based on
  - Allocation information
  - Statistics about fragments size of fragments – selectivity factors
  - Communication factors – use ping numbers for finding data transfer rates.

Deliverables:

- Upload the zip file with all code files on the portal.
- The report should be in pdf format. In report:
  - Explain the optimization algorithm/heuristics in detail
  - For a few queries, in detail include the final optimized query tree and output.
- Submission file name format - TeamName_phase3.zip
- Only one person from each team needs to submit on the portal.

Deliverables contain:

- Code
- A detailed report Report.pdf file describing data structure, any algorithms, and outputs for a few queries.

Evaluation:

- An application database will be provided to you from us for which you have to populate the system catalog and do the same stuff using the script you have written in the previous phase. We will run queries and check for the proper execution of it.
- A distributed DBMS software that can execute queries using both semi-joins and regular joins. The query result and time it took to execute the query should be provided.

Deadline: 11:59 pm, 1ˢᵗ April 2022