Name: Bharghav Srikhakollu

CS549 – ST/Scalable Querying & ML – Spring 2022

Topic: Scalar Aggregation in Inconsistent Databases

https://www.sciencedirect.com/science/article/pii/S0304397502007375

What is the problem discussed in the paper?

The paper talks about the issue of obtaining consistent information from inconsistent databases. The inconsistent databases are basically the databases which violate the given integrity constraints. They talk about the practical scenarios in which inconsistent databases occur like integration of autonomous data sources, unenforced integrity constraints, temporary inconsistencies, and conflict resolution.

The paper concentrates on scalar aggregation queries in databases that violate the functional dependencies. They defined consistent answers to such queries and showed how to compute such answers. The computational complexity of the problem is discussed in detail. It showed how tractability can be improved in special cases. They presented a practical hybrid query evaluation method to those cases.

Why is it important?

In practical scenarios, although there are defined set of functional dependencies and integrity constraints on the databases, we were expected to have inconsistent data in the databases which create problems when aggregated or complex queries are run on the database. So, this paper approach is very much critical to address those issues in best possible way. Also, it involves huge computational cost for fixing or working on such databases. Hence, it is important to explore and understand the approach in reducing the computational cost for computing such answers.

What are the main ideas of the proposed solution for the problem?

The paper talk about the graph theoretical representation of database repairs, which are specifically related to functional dependencies. The complexity of query processing as a function of the number of tuples in the database instance is well studied by the data complexity notion. The data complexity of checking repairs, the data complexity of computing consistent query answers is defined separately.

Since it is computationally not feasible to evaluate a scalar aggregation query in every repair, they proposed two ways of computing consistent answers by querying the given, possibly inconsistent database instance without having to compute all the repairs. The original query is modified into a new query by query transformation. The set of all repairs of an instance is represented as the conflict graph.

The computational cost of computing answers to aggregation queries are showed by decomposing the computation into 2 parts. One, that is involving standard relational query evaluation and second, that computes the consistent answers in a smaller instance.