Name: Bharghav Srikhakollu

CS549 – ST/Scalable Querying & ML – Spring 2022

Topic: CAvSAT: Answering Aggregation Queries over Inconsistent Databases via SAT Solving

https://arxiv.org/pdf/2103.03314.pdf

What is the problem discussed in the paper?

In almost every field, we can see that we can inconsistency in the database data where we can see that the data violates the integrity constraints of the database. Such databases can be referred to as inconsistent databases. In databases, inconsistencies are unavoidable in the real world.

To handle the inconsistency in the databases, primarily we have two approaches.

- ➔ Data Cleaning: Inconsistencies are removed by adding, deleting, updating the tuples in the relations. It is one of the main approaches followed in the industry which is more engineering oriented than science.
- ➔ Database Repairs: It is kind of framework where we will be coping with the inconsistent databases without cleaning the dirty data.

When we have multiple repairs on the database, the intersection of the queries on the repairs gives us the final output. Here another accurate approach is to find the range consistent answers where for a tuple we can represent the result as [glb, lub] where glb is the greatest lower bound and lub is the lowest upper bound which indicates the restriction on the potential output answer.

For a fixed Select-Project-Join aggregation query, computing the range consistent answers over a database instance is an NP-Hard problem. So, the paper discusses about addressing this problem in a better approach by working on the range consistent answers.

Why is it important?

Aggregation queries are very important and essential on real time databases. So, to handle the inconsistency over the databases, the repairing of databases should be done properly to get most desired answers from the database. There are many approaches to consistent query answering but not every approach can handle the aggregation queries.

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

The paper talks about the evaluation of performance of AggCAvSAT (Aggregate Consistent Answers via Satisfiability Testing) which is an enhanced version of the CAvSAT and can compute the range consistent answers to all aggregation queries involving SUM, COUNT, COUNT(*) with or without grouping. Input query, Integrity constraints are passed to the query pre-processor where it will go to inconsistent database, then it will work on Query re-writing and SAT solving steps to produce the consistent answers. MS SQL Server, Web Server and GUI are used for the implementation of CAvSAT.

Reference citation: CAvSAT: Answering Aggregation Queries over Inconsistent Databases via SAT Solving | Proceedings of the 2021 International Conference on Management of Data (acm.org)