CrowdOp: College Finder Web Application using Query Optimization

Sunny Asknani[1], Khushal Mohite[2], Pibalen Naorem[3],Bhagwat Wadhe[4] , Prof. Chaitali Patil[5]

[1] *Student, Computer, K. K. Wagh Institute Of Engineering Education & Research Nashik, Maharashtra, India*
[2] *Student, Computer, K. K. Wagh Institute Of Engineering Education & Research Nashik, Maharashtra, India*
[3] *Student, Computer, K. K. Wagh Institute Of Engineering Education & Research Nashik, Maharashtra, India*
[4] *Student, Computer, K. K. Wagh Institute Of Engineering Education & Research Nashik, Maharashtra, India*
[5] *Professor, Computer, K. K. Wagh Institute Of Engineering Education & Research Nashik, Maharashtra, India*

ABSTRACT

In Academic Education system while applying for Bachelor of Engineering Education, students lack knowledge of colleges that they are eligible for, according to their JEE marks. Many times students filled incorrect option form with blind approximation and missed chance to get best eligible college that they deserve. This web based application will predict more accurate colleges according to students score, rank, category and cut-off marks of last year admission process. Using query optimization technique system will provide results in very less time, and with more efficiency. Crowdsourcing is a sourcing model in which individuals or organizations use contributions from Internet users to obtain needed services or ideas. Crowd - sourcing enables user not having SQL knowledge to get desired result by _ring query in his terms and his expected conditions. The system provides the capability to optimize and provide a near optimal query execution plan for each query. CROWDOP is a query optimization technique that selects the optimal query plan among several plans for execution of a query considering both cost and latency. CROWDOP is capable of finding the query plan with low latency given a user- defined budget which nicely balances the cost and time requirement of users. A cost-based query optimization, a cheapest execution plan is created for each SQL condition (cost is with respect to predefined function). These plans use least amount of resources to get desired output. Supporting multiple crowdsourcing operators gives a cost based optimization for all operators involved in the query and derives best query plan for execution.

Keyword :- *Concurrency, Distributed databases, Multimedia databases, Object-oriented databases, Parallel databases, Query processing, Constrained optimization, Convex programming, Global optimization*

1. INTRODUCTION

To work on a education based data set for query optimization purpose. The web based application will provide an optimized result, a list of colleges to the user by taking academic information as input from the user. Proper optimization algorithm will be implemented along with joint and selection query operations. The system will process data quickly using CWORDOP and give useful and accurate result.

One of the major issues of any web application giving result as huge amount of data is time required to process the data. Web application based on academic system give output as vast amount of data in which some data is unnecessary though it is relevant to the result that the user seeks.

CROWDOP is capable of finding the query plan with low latency given a user- defined budget which nicely balances the cost and time requirement of users. A cost-based query optimization, a cheapest execution plan

is created for each SQL condition (cost is with respect to predefined function). These plans use least amount of resources to get desired output. Supporting multiple crowdsourcing operators gives a cost based optimization for all operators involved in the query and derives best query plan for execution.

2.LITRATURE SURVEY

We study the research challenges that naturally arise in the system design of CROWDOP. The first challenge is the formalization of our optimization objectives that consider both monetary cost and latency. To address this challenge, we introduce two optimization objectives. The first minimizes cost without considering latency, and the second uses budget-bounded latency minimization to judiciously trade off cost and latency. The second challenge is to efficiently select the best query plan with respect to the defined optimization objectives. To this end, we develop a class of optimization algorithms. For selection queries, we study how to balance between the cost and latency when selecting items by using multiple SELECT operators (e.g., *R*3.color=" *black*" , *R*3.make =" *Volvo*" , etc). We devise an algorithm to determine which SELECT operators should be crowdsourced in parallel for reducing the latency, and which ones should be applied over the results filtered by other operators in order to save cost. For join queries, we introduce a hybrid framework that combines FILL and JOIN operators: we leverage the crowd to first fill some missing attributes of items and then join the items having the same attributes. For example, to match cars' images with their reviews, we can first ask the crowd to fill makes of cars and then only match the cars with the same make. A key challenge in this framework is how to balance the costs from FILL and JOIN. To this end, we propose a partition-tree based strategy to select the most appropriate attributes to fill, and devise efficient algorithms for building the partition tree under latency constraints. Finally, we study the optimization of complex queries that involve all the three aforementioned operators with latency constraints.

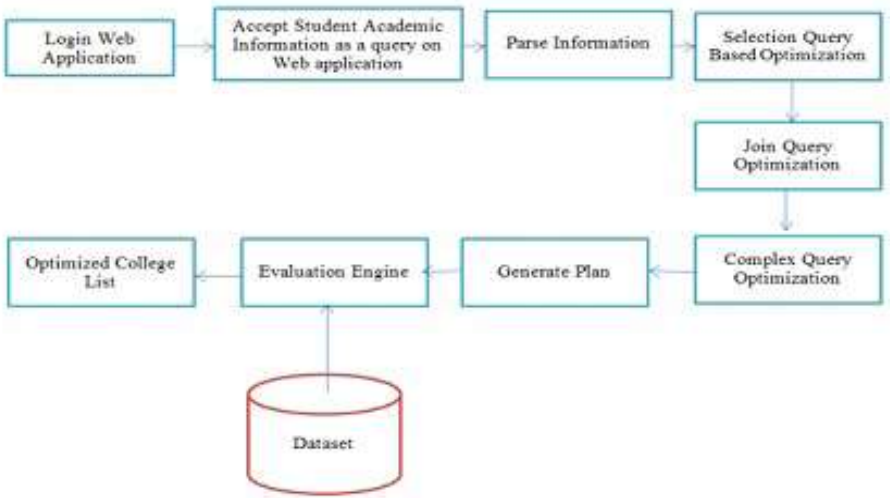
3.PROPOSED SYSTEM

We propose a novel optimization approach CROWDOP to finding the most efficient query plan for answering a query. Compared to the query optimization techniques proposed in recent crowdsourcing systems and algorithms, CROWDOP has the following fundamental differences in its design principle.

There are few software quality attributes which describes the quality of proposed system and they are as mentioned below:

1. Adaptability: Proposed System can be adapted easily to various software data patterns.
2. Availability: Can easily execute on currently available minimum configuration of hardware and software
3. Correctness

4.SYSTEM ARCHITECTURE DIAGRAM



5. CONCLUSIONS

Query optimization in relational databases is a well-studied problem [7]. Some of their techniques can be applied to the crowdsourcing scenario, such as pushing down the select predicates and utilizing selectivity to determine the select/join order. However, some inherent properties of crowdsourcing makes its query optimization a new and challenging problem. For instance, monetary cost is quite di_erent from computation cost in RDBs, and latency, which is an important criteria in crowdsourcing, is not a serious problem in RDBs.

6. REFERENCES

1. S. B. Davidson, S. Khanna, T. Milo, and S. Roy, Using the crowd for top-k and group-by queries, in Proc. 16th Int. Conf. Database Theory, 2013, pp. 225236.

2. J. Fan, M. Lu, B. C. Ooi, W.-C. Tan, and M. Zhang, A hybrid machine crowd- sourcing system for matching web tables, in Proc. IEEE 30th Int. Conf. Data Eng., 2014, pp. 976987.

3. J. Gao, X. Liu, B. C. Ooi, H. Wang, and G. Chen, An online cost sensitive decision-making method in crowd- sourcing systems, in Proc. ACM SIGMOD Int. Conf. Manage. Data, pp. 217228, 2013.

4. M. J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, and R. Xin,CrowdDB: Answering queries with crowd- sourcing, in Proc.ACM SIGMOD Int. Conf. Manage. Data, 2011, pp. 6172.

5. Marcus, D. R. Karger, S. Madden, R. Miller, and S. Oh, Counting with the crowd, Proc. VLDB Endowment, vol. 6, no. 2, pp. 109120, 2012.

6. H. Park and J. Widom, Query optimization over crowd-sourced data, Proc. VLDB Endowment, vol. 6, no. 10, pp. 781792, 2013.