International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 18
December 2016

Query Methodology for Advanced Structured Using Crowd sourcing Systems

Shaik Shahanaz [1], V.B.V.N.Krishna Suresh[2]

[1]PG Scholar, Dept of CSE, NRI Institute of Technology, Guntur, AP
Email: ssshahanaz2@gmail.com

[2] Assistant Professor, Dept of CSE, NRI Institute of Technology, Guntur, AP.
Email: vksuresh2000@gmail.com

ABSTRACT: Query optimization is an important aspect in crowdsourcing system. We consider the optimized crowdsourcing system uses SQL operators based crowdsourcesystem. The client is required to gift a SQL-like operator and framework assumes the liability of process in the inquiry making the execution stating and upload withthe crowd sourcingoriginations. The relational databasemodel , Query mythology is providing query structured is important for crowdsourcing. The propose system, a cost-based query optimization approach for crowdsourcing systems. The cost and time is consider in query optimization objective is proposed methods and generates query plans that give a good balance between the cost and latency.A main change in latest closed-world model for query processing is not hold for human input. From an implementation perspective, human-oriented query operators in solicit, integrate and cleanse crowdsourced data. The system develops efficient algorithms uses optimizing selection queries, join queries, and complex selection-join queries using parallel system. WSe are analyzing the various approaches that help to resolve the user queries in crowd sourcing modelis supports cost based query optimization, optimizing multiple crowd sourcing operators and access tradeoff in between cost from monetary point of view and latency.

Index Terms: Crowdsourcing, query optimization, Human Intelligence Tasks (HIT), SQL.

1. INTRODUCTION

Crowdsourcing system is web based activity in which thousands of people simultaneously post and edit work e.g., Yahoo! answers [2] where users take the review of people in question-answer format. Crowdsourcing is software which solves the complex tasks that computer cannot solve easily so with the equal intention crowdsourcing system combines human logic with computer power and gets the accurate result. Recently, crowdsourcing system has been adopted in database system. It will encourage the provision and coordinated effort of people, associations, and social orders[3]. we tend to trust that information Systems researchers area unit in associate one in all a form position to form large commitments to the present rising exploration zone and take into account it as another examination outskirts. Be that because it might, during this method number of studies has explained what are accomplished and what need to be finished[4]. This paper tries to gift a discriminating examination of the substrate of menstruation therefore on crowd exploration the scene of existing studies, as well as theoretic establishments, analysis strategies, and examination foci, and distinguishes a few vital exploration headings for IS researchers from 3 points of view—the member, association, and framework—and that warrant more study. This exploration adds to the IS writing and offers bits of information to scientists, fashioners, arrangement Creators and administrators to higher comprehend completely different problems in crowd sourcing frameworks[5]. Query optimization is a function of many relational database management systems. The query optimizer attempts to determine the most efficient way to execute a given query by considering the possible query plans [6]. Generally, the query optimizer cannot be accessed directly by users: once queries are submitted to database server, and parsed by the parser, they are then passed to the query optimizer where optimization occurs. A query is a request for information from a database. Queries results are generated by accessing relevant database data and manipulating it in a way that yields the requested information[7]. Since database structures are complex, in most cases, and especially for not-very-simple queries, the needed data for a query can be collected from a database by accessing it in different ways, through

different datastructures, and in different orders. Each different way typically requires different processing time. Query optimization find the best query plan in terms of estimated monetary cost[8].

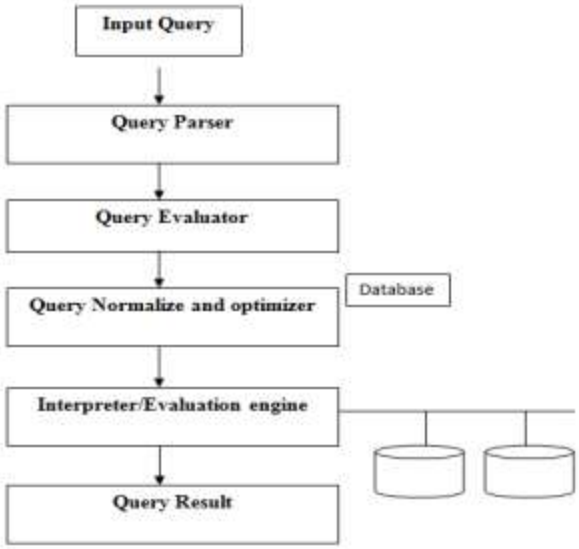

Fig 1: Query Optimization

Which includes the number of questions being asked difficulties of the questions and the monetary cost incurred It is therefore important to design an efficient crowd sourcing query optimizer that is able to consider all good query plans and select the "best" plan based on a cost model and optimization objectives [9]. The objective of proposed system is to evaluate the effectiveness of query optimization schemes for the crowd powered selection, join and complex queries in a simulated crowd sourcing environment using parallel system. System should give able to split the join query and run that query on two or more system parallel [10]

2. RELATED WORK

Now a days, many works have been proposed to perform database operations in a crowdsourcing system such as select, join and aggregation operators such as count, max and sort. Recently many crowd sourcing systems have been developed to provide SQL like query interface to the crowd such as Crowd DB system solves the queries by using human answer which cannot be solved by database and uses SQL as query language[13]. It is rule based optimization crowdsourcing system. Rule based is easy to implement but it generates very less effective evaluation plans. It uses crowdprobe, crowdjoin and crowdcompare operators in query optimization process. Deco is declarative crowdsourcing system in which SQL queries are solved. Deco works on missing value of database. Deco uses only fill operator. CrowdOP is the optimized crowdsourcing system. A considerable amount of research on Query optimization for crowd sourcing has been done over the past few years. Some important techniques

are discussed here. Various techniques have been proposed and research has been done in the field of query optimization. Also many advanced methods have been introduced for query optimization and crowd sourcing Chien-Ju Ho, ShahinJabbari and Jennifer Wortman Vaughan [11], state that Crowdsourcing markets have gained attractiveness as a tool for reasonably collecting data from diverse populations of workers. classify tasks, in which workers provide labels for in- stances are among the most common tasks posted, but due to human error and the rate of spam, the labels collected are often noisy. They examine the problem of task assignment and label deduction for diverse classification tasks. By applying online primal-dual techniques, derive a provably near-optimal adaptive assignment algorithm. They show that adaptively assigning workers to tasks can lead to more accurate predictions at a lower cost when the accessible workers are diverse. Adam Marcus David Karger Samuel Madden Robert Miller Sewoong Oh [12], proposed a several techniques for using workers on a crowdsourcing stage like Amazon's Mechanical Turk. This is important in crowdsourced query optimization to support predicate ordering and in query estimate, when performing a GROUP BY operation with a COUNT or AVG aggregate. They develop techniques to remove spammers and collude attackers trying to slant selectivity estimates when using count estimation approach and find that for images, counting can be much more effective than sampled labeling, reducing the amount of work necessary to arrive at an estimate that is within 1% of the true fraction by up to an order of magnitude, with lower worker latency.

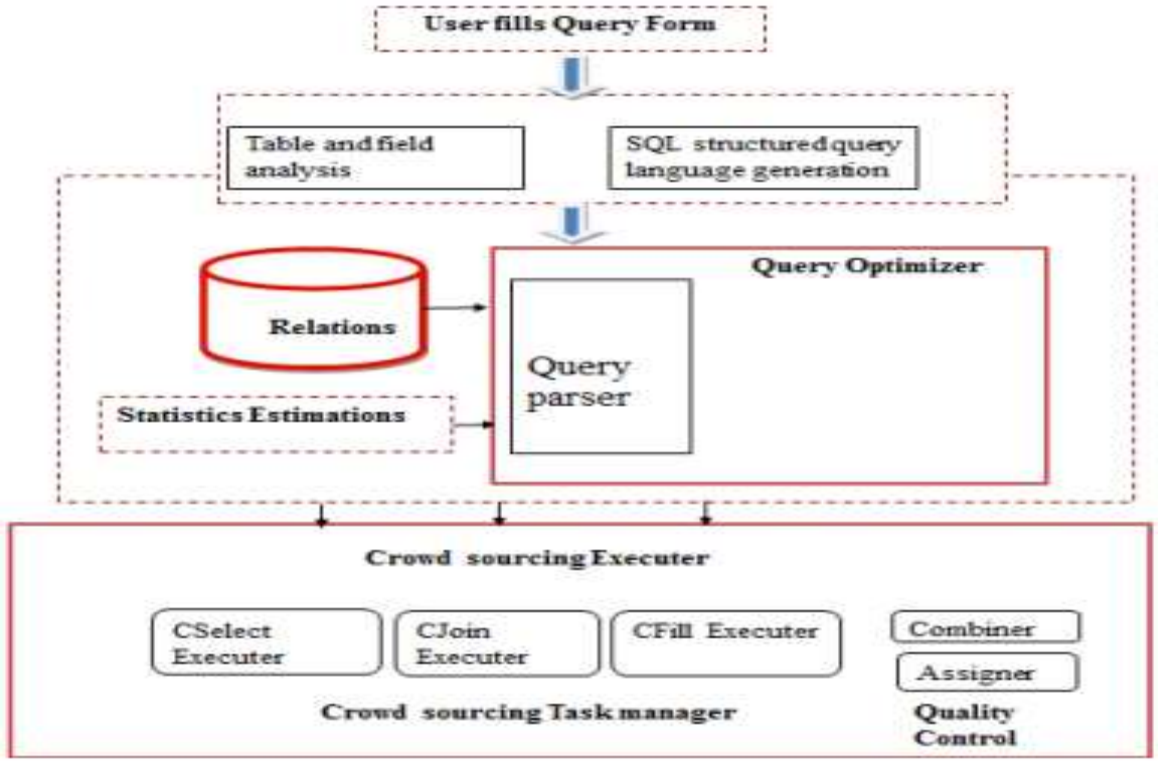

Figure 2: Block diagram of proposed system

3. Proposed System

The proposed system uses database operations in query optimization process so that there is need to use database query language in the system. In this section, we introduce the data model and query language that is used by the proposed system and also architecture of proposed system is described.

Data Model The Proposed system uses relational data model. In this system data is considered as schema that consists of a set of relations. These relations are designated by schema designers. SQL query can be executed by relations and queries can be given by requester

Query Language The proposed system goes through the SQL query language over the database and result of the query is obtained by executing query by using the relations used in the data model. The proposed system considers following two types queries

(1) Selection Query: A selection query uses one or more selection conditions over the records in single relation. It has many

applications in crowdsourcing platform like filter item [14]. A simple example is find out cars having width is less than equal to 66.5 and length is greater than equal to 165.3 and wheel base is less than equal to 96.9. For this, we express the query as Q1 for three conditions, Q1: select * from car where width <= 66.5 AND length >= 165.3 AND wheel_base< 96.9 We take another example of selection query for two conditions and express the query as Q2, Q2: Select * from car where wheel_base< 96.1 AND stroke >= 3.12

(2) Join Query: Join query is used to combine records from two or more relations according to certain conditions. An example join query Q3 over relations car and image which combines records from car relation to image relation, which is presented as follows. Q3: select c1.*, c2.* from car c1, image c2 where c1.carid=c2.carid and c1.symboling >= 2 and c1.bore >= 3.31 and c1. highway_mpg> 34

Architecture of Proposed System The proposed system is based on crowd sourcing platform which has some optimization in sense of SQL queries. The proposed system includes SQL queries like select, join, fill, count, and max. In this section, we describe overall architecture of proposed system.

The proposed system incorporates the traditional query compilation, optimization and execution components. We briefly describe each component as follows [15].

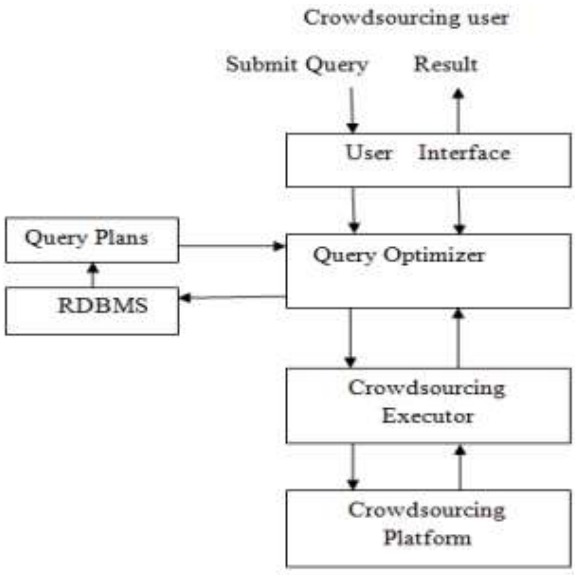

Figure 3: Block Diagram of the OptimizedCrowd System

1) **User Interface:** By using user interface, user submits SQL query.

2)Query Optimizer: Submitted query is processed by query optimizer. In query optimizer, query first compiles and generates the optimized query plans which is based on latency i.e. execution time of query. The query optimizer selects the low latency query plan for execution purpose.

3) **Crowd sourcing Executor:** The selected query plan is evaluated by crowdsourcing executor and generates the tasks. Then publish these tasks on crowdsourcing platform. These tasks are solved by human workers and answers are given to these tasks. The system collects the answer from the workers. By using these answers, crowdsourcing executor, executes the query and accurate result is given to the user.

4. ALGORITHM

Optimization framework

The construction modeling of query handling in CROWDOP is outlined. A SQL query is issued by a crowd sourcing client what is additional is foremost handled by query OPTIMIZER, that parses the query and produces an increased question arrangement. The inquiry arrangement is then dead by CROWDSOURCING executor to supply human data assignments (or HITs) and distribute these HITs on crowdsourcing stages, as an example, Amazon Mechanical Turk (AMT). Taking into account the HIT answers gathered from the cluster, CROWDSOURCING executor assesses the question and returns the acquired results to the client. CROWDOP employs relational data model, like previous work on crowdsourcing systems [16].

Input: Query Q, Cost C

Output: Query Q, Optimized plan

Step 1: Initialize database and tables, load tables

Step 2: Initialize C =nil

Step 3: Calculate Latency Min (Lmin)

Step 4: Execute Query SELECT

Step 5: Calculate Latency Max (Lmax)

Step 6: Compute Query cost Lmax – Lmin

Step 7: Do Step 3 to 6 for JOIN and COMPLEX

Step 8: Compare Latency.

Our proposed optimization schemes for the crowd-powered selection, join and complex queries in a simulated crowd sourcing environment, and then examine the latency model and query optimization via experiments on the real crowdon Amazon Mechanical Turk (AMT).We develop efficient and effective optimization algorithms for select, join and complex queries. Our experiment on both simulated and real crowd demonstrates the

effectiveness of our query optimizer and validates our cost model and latency model [17].

Marcus,Karger,Madden,Miller,[18], proposed a several techniques for using workers on a crowdsourcing platform like Amazon's Mechanical Turk to estimate the fraction of items in a dataset that satisfy some property or predicate (e. and do this without explicitly iterating through every item in the dataset. This is important in crowdsourced query optimization to support predicate ordering and in query evaluation, when performing a GROUP BY operation with a COUNT or AVG aggregate. They compare sampling item labels, a traditional approach, to showing workers a collection of items and asking them to estimate how many satisfy some predicate. Additionally, they develop techniques to eliminate spammers and colluding attackers trying to skew selectivity estimates when using this count estimation approach and find that for images, counting can be much more effective than sampled labeling, reducing the amount of work necessary to arrive at an estimate that is within 1% of the true fraction by up to an order of magnitude, with lower worker latency.

5. Experimental Results

In this section, initially we examine the effectiveness of our proposed optimization method for the select, join queries. For our experiments purpose, we use a real auto import dataset which consists of specification of 205 cars to generate the car relation. This dataset is in text format. In preprocessing step, we convert text format into SQL database format so that system can easily process database queries. Then we manually generate relation image by adding image related with each car present in car relation. This image is taken from yahoo auto we validate optimization approach for join query. For experimental purpose, we take query with three conditions. We join two relations car and image on basis of attribute carid. Query evaluate in three ways on the basis of conditions present in the query.

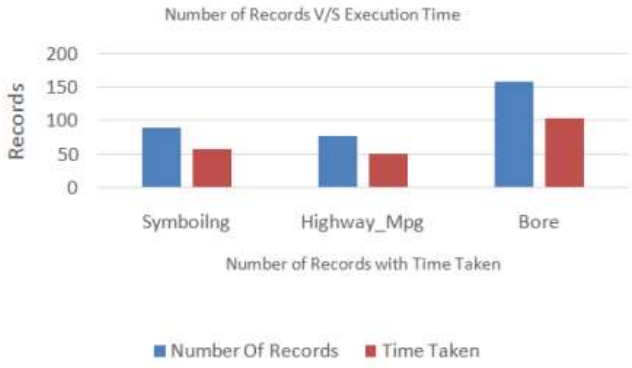

Figure 4: Performance of Query with execution time

From the above discussion, we have observed that selection of best query plan improves the performance of system in terms of execution time hence query optimization plays an important role in developed system.

6. CONCLUSION

We propose a cost-based query optimization that considers the cost-latency tradeoff and supports multiple crowd sourcing operators. We develop efficient and effective optimization algorithms for select, join and complex queries. Our experiments on both simulated and real crowd demonstrate the effectiveness of our query optimizer and validate our cost model and latency model. Different approaches of query optimization for crowdsourcing have been discussed in detail. The efficient and effective optimization algorithm develops for select, join, complex query. In the present scenario, simulated and real crowd demonstrate the effectiveness of our query optimizer and take review of query optimization objective and generates query plans that provide a good balance between the cost and latency in query optimization purpose and generates query plans that provide a good balance between the cost and latency. sThe future plan is to extend the proposed system for more advance SQL operation and run on big data in a cloud environment.

7. Future Work

We have presented a latency based query optimization which helps to optimization in sense of SQL queries. The system includes SQL queries like select, join, fill operations. In addition to these three operations some aggregate operators are used in query optimization purpose such as count and max. This optimization is effective for SQL query processing. In future, we intend to study complex join query for query optimization and also we use cost and latency combination in query optimization technique.

8. 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. 225–236.

[2]A.D.Sharma, A. Parameswaran, H. Garcia- Molina, and A. Halevy, "Crowd-

powered find algorithms", in Proc. IEEE 30th Int. Conf. DataEng., 2014, pp. 964-975.

[3] J. Fan, M. Lu, B. C. Ooi, W.-C.Tan, and M. Zhang. A hybrid machine-crowdsourcing system for matching net tables‖ .In ICDE Conference, 2014.

[4] M. J. Franklin, D. Koss Mann, T. Kraska, S. Ramesh, and R. Xian. ―Crowd dB: responsive queries withCrowdsourcing.‖ In SIGMOD Conference, pages 61–72, 2011.

[5] J. GAO, X. Liu, B. C. Ooi, H. Wang, and G. Chen. ―An online cost sensitive decision-making methodology in crowdsourcing systems.‖ In SIGMOD Conference, pages 217–228, 2013.

[6] C.-J. Ho, S. Jabbari, and J. W. Vaughan, "Adaptive task assignment for crowdsourced classification," in Proc. 30th Int. Conf. Mach. Language, 2013, vol. 1, pp. 534–542.

[7] X. Liu, M. Lu, B. C. Ooi, Y. Shen, S. Wu, and M. Zhang, "CDAS: A crowdsourcing data analytics system," Proc. VLDB Endowment, vol. 5, no. 10, pp. 1040– 1051, 2012.

[8] R. Karger, S. Madden,R. Miller, and S. Oh, "Counting with the crowd," Proc. VLDB Endowment, vol. 6, no. 2, pp. 109–120, 2012.

[9] J. Fan, M. Lu, B. C. Ooi, W.-C. Tan, and M. Zhang,"A hybrid machine crowdsourcing system for matching web tables," in Proc .IEEE 30th Int. Conf. Data Eng., 2014, pp. 976–987

[10] Ju Fan, Meihui Zhang, Stanley Kok, Meiyu Lu, and Beng Chin Ooi, "CrowdOp: Query Optimization for Declarative Crowdsourcing Systems", IEEE transactions on knowledge and data engineering, VOL. 27, NO. 8, AUGUST 2015

[11] C.-J. Ho, S. Jabbari, and J. W. Vaughan, "Adaptive task assignment for crowdsourced classification," in Proc. 30th Int. Conf. Mach. Language, 2013, vol. 1, pp. 534–542.

[12] A. Marcus, D. R. Karger, S. Madden,R. Miller, and S. Oh, "Counting with the crowd," Proc. VLDB Endowment, vol. 6, no. 2, pp. 109–120, 2012.

[13] A.G.Parameswaran, H.Park, H.Garcia-Moline, N.Polyzotis, and J. Widom, "Deco: Declarative crowdsourcing", in Proc. 21st

ACM Int. Conf. Inf. Knowl. Manage, 2012, pp. 1203-1212

[14] A.G. Parameswaran, H. Garcia-Milina, H. Park, N. Polyzotis, A. Ram and J. Windom, "CrowdScreen: Algorithms for filtering data with humans," in Proc ACM SIGMOD Int. Conf. Manage. Data, 2012, pp. 361-372.

[15] A. Marcus, D.R. Karger, S. Madden, R. Miller, and S. Oh, "Counting with the crowd", Proc. VLDB Endowment, vol. 6, no. 2, pp. 109-120, 2012.

[16] M. J. Franklin, D. Koss Mann, T. Kraska, S. Ramesh, and R. Xian. —Crowd dB: responsive queries withCrowdsourcing.|| In SIGMOD Conference, pages 61–72, 2011.

[17] J. GAO, X. Liu, B. C. Ooi, H. Wang, and G. Chen. —An online cost sensitive decision-making methodology in crowdsourcing systems.|| In SIGMOD Conference, pages 217–228, 2013.

[18] A. Marcus, D. R. Karger, S. Madden,R. Miller, and S. Oh, "Counting with the crowd," Proc. VLDB Endowment, vol. 6, no. 2, pp. 109–120, 2012.

Student details:-

Shaik Shahanaz is studying M.tech in Dept of CSE, NRI Institute Of Technology,Visadala(P), Medikonduru(M), Guntur,AP

Email: ssshahanaz2@gmail.com

Guide Details:

V.B.V.N.Krishna Suresh is working as a Assistant Professor, in Dept of CSE,NRI Institute Of Technology,Visadala(P), Medikonduru(M), Guntur,AP.

Email: vksuresh2000@gmail.com