International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8171
***Article Received:** 29 June 2023 **Revised:** 26 August 2023 **Accepted:** 06 September 2023*

A Novel Hybrid Optimization With Ensemble Constraint Handling Approach for the Optimal Materialized Views

Nagagopiraj Vullam[1], Chinta Venkata Murali krishna[2*], Suneetha Davuluri[3*], Muvva Venkateswara Rao[4], Venugopal Boppana[5], Shobana Gorintla[6], Popuri Srinivasarao[7]

[1]Computer Science and Engineering, Chalapathi institute of engineering and technology, guntur, India.
[2*]Computer Science and Engineering, NRI Institute of Technology, Vijayawada, India.
[3]Computer Science and Engineering, NRI Institute of Technology, Vijayawada, India.
[4]Computer Science and Engineering, NRI Institute of Technology, Vijayawada, India.
[5]Computer Science and Engineering, NRI Institute of Technology, Vijayawada, India.
[6]Computer Science and Engineering, NRI Institute of Technology, Vijayawada, India.
[7]Computer Science and Engineering, RVR & JC College of Engineering, Guntur, India.
Email:srinivas.popuri4@gmail.com

Abstract —The datawarehouse is extremely challenging to work with, as doing so necessitates a significant investment of both time and space. As a result, it is essential to enable rapid data processing in order to cut down on the amount of time needed to respond to queries that are sent to the warehouse. To effectively solve this problem, one of the significant approaches that should be taken is to take the view of materialization. It is extremely unlikely that all of the views that can be derived from the data will ever be materialized. As a result, view subsets need to be selected intelligently in order to enable rapid data processing for queries coming from a variety of locations. The Materialized view selection problem is addressed by the model that has been proposed. The model is based on the ensemble constraint handling techniques (ECHT). In order to optimize the problem, we must take into account the constraints, which include the self-adaptive penalty, the Epsilon ()-parameter, and the stochastic ranking. For the purpose of making a quicker and more accurate selection of queries from the data warehouse, the proposed model includes the implementation of an innovative algorithm known as the constrained hybrid Ebola with COATI optimization (CHECO) algorithm. For the purpose of computing the best possible fitness, the goals of "processing cost of the query," "response cost," and "maintenance cost" are each defined. The top views are selected by the CHECO algorithm based on whether or not the defined fitness requirements are met. In the final step of the process, the proposed model is compared to the models already in use in order to validate the performance improvement in terms of a variety of performance metrics.

Keywords: Materialization, Ensemble approach, Stochastic ranking, COATI Optimization, Optimal view selection.

I. Introduction

In general, a view is a representation of the data set returned by a query. It is referred to as a materialised view (MV) if the query data is occasionally updated in the base stable. MVs are typically implemented in settings in which the data is frequently accessed [1, 2]. In order to reduce the load on the network, MVs are frequently utilised in the data warehousing industry. In addition, the MVs are frequently utilised to enhance the performance of query operations [3]. To improve the efficiency of analytical query processing in data warehouses, MV is a strategy that is increasingly being implemented. Memory space is the most significant challenge presented by the MV process due to the fact that it utilises a significant amount of space. The cost of maintaining MVs is another challenge [4, 5]. It is essential to make an appropriate choice of views in order to achieve a faster response to the queries. Materialised views are typically more compact than the data warehouse, and they can provide responses to queries in a shorter amount of time [6, 7]. Materialised views are regarded as the best views to select from the data warehouse in order to improve query data analytics. In order to respond to online queries with approximately one million reports in a shorter amount of time, a DW (data warehouse) is utilised. The most important challenge is to shorten the amount of time needed to process online queries in comparison to the other methods [8].

The views were selected with the goal of enhancing the query process and reducing the cost. Materialised views have the goal of reducing the amount of time it takes for analytical queries to be executed when they are posted next to a DW [9, 10]. In most cases, a DW is made up of various points of view and gives responses to questions. Through the use of view selection, the total amount of time spent responding to queries can be cut

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8171
***Article Received:** 29 June 2023 **Revised:** 26 August 2023 **Accepted:** 06 September 2023*

down significantly compared to other approaches [11]. The most significant drawback of a data warehouse is the increased amount of time needed to process queries. MV reduces the amount of time required for a response by reporting views rather than the entire table. The MV selection process improves the effectiveness of query processing as well as the performance of decision making [12]. The primary factors contributing to a lengthy response time are the complex nature of analytical query processing and large amounts of data. The processing time of analytical queries is something that the MV process is working to cut down on [13].

Studies being conducted at the moment are concentrating on automatic data view creation as well as view identification. Because of the query-intensive nature of data warehouse environments, MVs are appealing because they are very effective at speeding up the query process. This makes MVs attractive in data warehouse environments. A data warehouse (DW) is the result of compiling a sizable amount of information drawn from a wide variety of distinct sources (14). An efficient selection of materialised views is the method for reducing the amount of time required for processing. A materialised view is made up of data that has been aggregated and pre-calculated beforehand. The problem with the NP card can be solved by selecting the best possible views. The process of materialisation allows for the selection of appropriate viewpoints. It has the potential to cut down on the amount of time needed for response to queries [15].

For the purpose of MV selection, existing randomised methods [16], evolutionary methods [17], and meta-heuristic methods [18] are used respectively. In this case, randomised schemes are used to select the set of views in a near-optimal manner. In addition, genetic algorithms [18], PSO (Particle swarm optimisation) [19], and greedy-based algorithms are utilised to carry out the process of selecting the optimal view. Deterministic algorithms, randomised algorithmic approaches, and constraint programming are three of the most common methods utilised in MV selection [20]. In order to achieve optimised view selection, the techniques that were utilised in earlier works require further development. The utilisation of a creative strategy in the selection of MVs is the primary purpose of the work that is proposed. The following is a description of the most important contributions made by the proposed methodology:

An approach to the handling of ensemble constraints is presented in this article in order to develop an optimal materialisation of views.

In order to make the query selection process go more quickly, hybrid optimisation strategies have been developed. In this case, Ebola and COATI are combined as optimisation strategies in order to achieve better results in the materialisation processes.

In order to speed up the process of responding to queries, the suggested methodology took into account a variety of cost-based fitness evaluations. The combination of methods that was suggested provides an optimal view selection and results in a shorter amount of time for the query response.

The remaining sections of the paper are organised as follows: section 2 discusses the recent works that are associated with the topic; section 3 gives a clear explanation of the methodology that has been proposed; section 4 illustrates the results and their discussion; and section 5 provides a summary of the findings of the paper.

II. Related Work

Jay Prakash and Vijay Kumar [21] introduced a multi-objective based MV selection. Here, optimal MV selection is performed using a non-Pareto based genetic optimization approach. The multi-dimensional lattice view was generated for the finest chosen of MVs. The most important K-views were selected for the finest selection of MVs. The performance of query execution time was improved than the compared approaches. However, the performance of the proposed scheme can be improved by considering recent optimization approaches.

Archana Bachhav et al. [22] developed a proficient query optimizer with MVs in a dispersed location. A large amount of resources needs a query optimizer to decrease the response time and increase the utilization of resources. The proposed innovative query optimization scheme enhances the query processing with selected materialized views. Moreover, it decreases the payment overhead of customers. The performance of the presented approach can be improved by utilizing an improved combination of methodologies.

AnjanaGosain and KavitaSachdeva [23] introduced a stochastic ranking based cuckoo search (CS) optimization. The optimal selection of MV improves the efficiency of the query process. Here, constraints handling was performed by utilizing the stochastic ranking process, and the CS optimization was utilized for view selection. The incorporation of both ranking and optimization schemes improved query processing. The combination of approaches solves the scalability and price of the query process. However, the MV selection process can be improved using recent approaches.

Santanu Roy et al. [24] developed an optimal materialized view creation. Here, MVS is created in the data space of non-binary space. Different weight values were considered for selecting the particular queries from many queries. The developed scheme creates the weight-based MV selection process to choose the views significantly. The developed scheme was validated with

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8171
***Article Received:** 29 June 2023 **Revised:** 26 August 2023 **Accepted:** 06 September 2023*

different datasets. The validation results illustrate the efficiency of the developed scheme. However, the MV selection was not achieved good results, and it needs further improvements.

Mustapha ChabaMouna et al. [25] introduced a proactive selection of MVs. Here, the RE-selection scheme was considered for an optimal selection of views[29]. The online and offline features were considered for the analysis process. The threshold was selected for the optimal selection MVs. Afterwards, a scheduling scheme was considered for an optimal choice of views. The performance efficiency of the system was improved, and improved approaches need to be used to improve the process of MV selection.

The genera Nasua [28]and Nasuella, which are both found in the family Procyonidae, are home to the coatis, which are also known as coatimundis. They are mammals that are active during the day and can be found in their natural habitat in the southwestern United States, Mexico, Central America, and South America. A slender head that has a All coatis have the same characteristics, including a flexible, extended, and slightly upward-turned nose, black paws, tiny ears, and a long, non-prehensile tail that is used for signalling and maintaining balance. The length of an adult coati can range anywhere from 33 to 69 centimetres (cm) from its head to the tip of its tail, which can be as long as its body. The average coati weighs between 2 and 8 kilogrammes and stands approximately 30 centimetres tall at the shoulder. Their size is comparable to that of a large house cat. Males have the potential to grow to be nearly twice as large as females and have larger, more pointed canine teeth. These are the measurements for the South American coati and the white-nosed coati. The mountain coati is the smaller of the two coati subspecies.Coatis are classified as omnivores, meaning that in addition to eating invertebrates like tarantula, they also consume small vertebrate prey like birds, lizards, rodents, and even eggs from birds and crocodiles. A coati's go-to meal is an iguana, especially one that is green in colour. Since these large lizards, known as iguanas, are frequently found in trees, coatis must hunt them in large groups.While some of them are able to climb trees and scare the iguana into jumping to the ground, others of the coatis are able to quickly attack it. Despite this, coatis are susceptible to attacks from a variety of dangerous animals. There are a number of animals that prey on coatis, including jaguars, ocelots, tayras, jaguarundis, dogs, foxes, maned wolves, boa constrictors, and anacondas. Large raptors, such as harpy eagles, black-and-chestnut eagles, and ornate hawk-eagles, are also known to hunt them.

The analysis of existing works is important for deliberating the necessity of materialized view selection. They focused on materialized view selection using different approaches. However, the existing approaches are failed to get the query response in a lesser amount of time. Moreover, an effective MV selection approach is needed to improve the performance using materialized views. Therefore, this work presented an optimal MV selection approach using an ensemble of constraint handling approaches and optimization approaches.

III. Proposed methodology

The MVS problem is addressed by the model that was proposed, and it does so using ensemble constraint handling techniques (ECHT). In order to optimize the problem, the ensemble constraints are taken into consideration. The proposed model includes the implementation of an innovative algorithm known as the constrained hybrid Ebola with COATI optimization (CHECO) algorithm, which enables a more expedient and effective selection of queries from the data warehouse. The fitness is determined by taking into account multiple objectives, and the proposed algorithm seeks to minimize this fitness metric while adhering to the ensemble constraints. The top views are selected by the CHECO algorithm on the basis of whether or not the defined fitness criteria are met. In the end, the proposed model is validated against other models to determine whether or not there has been an improvement in performance. Figure 1 depicts the overall plan for the proposed method in the form of a diagram.

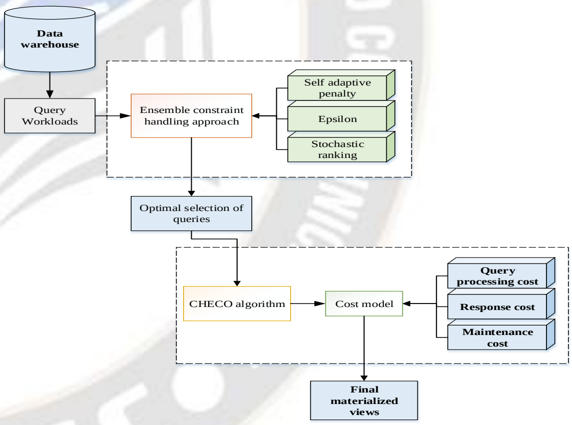

Figure 1: Schematic diagram of the proposed methodology

3.1 Processing of materialized views by means of ensemble constraint handling techniques (ECHT)

The approaches for handling constraints make the MV selection process more effective and provide the best possible solution. In this case, ensemble constraint handling is accomplished by integrating the various constraints listed below.

3.1.1 The self-adaptive cost of failure

In this step, two penalty types are united for each individual to identify non-viable individuals. The higher penalty value is considered for the infeasible views by using the constraint handling scheme. Similarly, a lower penalty value is considered

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8171
***Article Received:** 29 June 2023 **Revised:** 26 August 2023 **Accepted:** 06 September 2023*

a feasible solution. Then, the threshold value is utilized for ranking the feasible and infeasible solutions, and thus optimal views are obtained. The ranking process of views with the generated fitness value is expressed in condition (1),

$$f(z) = dist(z) + P_y(z) \qquad (1)$$

This represents the fitness value, this represents the distance measure, and this represents the penalty value.

3.1.2 Epsilon (ε)-constraint

In this method of data processing, the parameter is responsible for handling the constraints. An accurate selection of views requires careful consideration of the appropriate value. The counter has reached the control generation counter, which is still the value that has been revised. If the counter is greater than the, then must be set to zero in order to find a solution that satisfies all of the constraints while still producing the desired result. The expression of this is found in condition (2),

$$\varepsilon(0) = w(Z_t) \qquad (2)$$

$$\varepsilon(n) = \begin{cases} \varepsilon(0) = \left(1 - \dfrac{n}{K_m}\right)^{np}, & 0 < n < K_m \\ 0, & n \geq K_m \end{cases} \qquad (3)$$

In this instance, stands for the best individual, and also for the limit of parameters that is thought of as the threshold. When taking this constraint-based approach, the view selection is thought to be in a good state if the total number of violated views is fewer than the threshold value.

3.1.3 Stochastic ranking (SR)

The SR constraint handling scheme searches for the best possible solution by striking a balance between the objective function and the penalty function. It does this by comparing two individuals according to the probability of individuals and then providing a rank for each individual based on that comparison. If both individuals produce optimal results, then the person who has a lower objective value should be regarded as having the higher value. In addition to this, if one individual achieves a non-viable result and the other achieves a viable outcome, then the individual who achieved the viable outcome is awarded the highest rank. The overall performance of the view selection process is being helped along by this constraint handling approach's ensemble.

3.2 Optimal query selection achieved through the combination of constrained hybrid Ebola and COATI optimization (CHECO)

This section presents a hybrid Ebola with COATI optimization (CHECO) for a faster and optimal selection of queries from the data warehouse. At first, arbitrarily create the index from all individuals. Then, fix the index as the current best and the global best. Subsequently, compute the fitness value based on the global and current best one. If the maximum iteration is not reached, there will be an optimal individual. Henceforth, every individual updates their position and displacement. According to this update of individuals, optimal solutions are attained. The processing steps of the CHECO approach are described in subsequent subsections.

3.2.1 initialization of data

The COA method is a view population-based metaheuristic, and the coatis are treated as population members within the context of this algorithm. The values for the decision variables are determined based on where each coati(view) is situated within the search space. Therefore, according to the COA, the position of coatis is an example of a possible solution to the issue. The position of the coatis in the search space is initially seeded with a random value using equation (4) when the COA algorithm is first put into practice.

$$X_i: x_{ij} = l_{bj+r.(ubj-lb_j)}, i = 1,2,3.....N, j = 1,2,.....m, \qquad (4)$$

where Xi represents the location of the ith coati in search space and xi,j represents the value of the jth decision variable, where N is the number of coatis, m is the number of decision variables, r is a random real number in the range [0, 1], l_{bj} and u_{bj} are the lower bound and upper bound of the jth decision variable, respectively, and r is a random real number in the interval [0, 1].

The current best positions are located, as well as kept up to date, and the information is expressed in condition. (5),

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N\times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{1,1} & \cdots & x_{1,1} & \cdots & x_{1,1} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{1,1} & \cdots & x_{1,1} & \cdots & x_{1,1} \end{bmatrix}_{N\times m} \qquad (5)$$

The following equation provides a mathematical representation of the coati population in the COA:

3.2.2 Iguana hunting methods and attack strategy

The iguana represents the best performing member of the view population in the COA layout. Half of the coatis are thought to ascend the tree in search of the iguana, while the other half wait below. Therefore, using Eq(6), we can update the position where the coatis will be as it climbs the tree.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8171
***Article Received:** 29 June 2023 **Revised:** 26 August 2023 **Accepted:** 06 September 2023*

$$X_i^{P1} : x_{i,j}^{P1} = x_{i,j} + r\,.\left(Iguana_j - I\,.x_{i,j}\right), for\ i = 1,2,\ldots,\left\lfloor \frac{N}{2} \right\rfloor \ \ and\ j = 1,2,\ldots,m. \quad (6)$$

When the iguana hits the ground, it is placed somewhere in the search space at random. Coatis on the ground use this arbitrary launch point to explore the modeled-equation-space search.

3.2.3 Getting away from predators (exploitation phase).

The process of updating where coatis like a views are in the search space is modelled mathematically based on how coatis act when they see predators and try to get away from them. When a predator tries to eat a coati, the animal moves away. Coati's moves in this strategy put it in a safe place close to where it is now. This shows the COA's ability to use local search to its advantage

The perspectives are adjusting their positions and moving closer and closer to the ideal position.

$$X_i = \begin{cases} X_i^{P2}, & F_i^{P2} < F_i \\ X_i \end{cases} \quad (7)$$

Here X P2 i is the new position calculated for the ith coati, based on the second phase of COA, x P2 i,j is its jth dimension, F P2 i is its objective function value, r is a random number in the interval.

3.2.4 Repetition process

After the first and second phases of a COA's iteration have changed where all the coatis are in the search space, the iteration is done. The process of keeping track of people. The locations of the individuals are adjusted so that we can arrive at the best possible spot. The update to the best possible position is expressed in the given conditions.

The proposed hybrid optimization scheme attains the optimal materialized views. The example diagram of materialized view selection is depicted in figure 2.

Data collection

Order ID	Name	Account	Item ID	Stock	Quantity
1	Shirts	A	40	125	2
2	Sarees	B	41	143	3
3	Pants	C	42	115	3

Materialized views

Item ID	Name	Quantity
40	Shirts	2
41	Sarees	3
42	Pants	3

Figure 2: Example diagram of materialized view selection

An optimal choice of views for materialization is necessary for a lesser processing time of queries. A data warehouse is a huge data storehouse that supports query decision making in an incorporated environment. A data warehouse has many data records, and it is necessary to decrease the online query processing time. Furthermore, fitness is evaluated to update the optimal position using the condition (14).

3.2.5 Multi objective-based fitness evaluation

In this section, different processing cost measures are considered for computing the performance of the developed system.

3.2.5.1 Query processing cost

Accessing the views of queries that provide information about their execution frequency constitutes the processing cost. It is communicated through subsequent conditions (8),

$$Qp_k = \sum fq \times Ca_k^q \quad (8)$$

Here, Qp_k signifies the query processing cost, fq represents the frequency count of queries and Ca_k^q signifies the accessing cost of k queries.

3.2.5.2 Maintenance cost

It is the cost that must be incurred in order to restore the views whenever the respective base associations undergo reorganisation. The formula for determining maintenance costs is expressed by the following condition (9),

$$Mc_k = \sum uq \times Ca_N^q \quad (9)$$

In this case, represents the cost of maintenance, denotes the updated frequency of queries, and denotes the cost of maintaining a view for an updated base association.

3.2.5.3 Response cost

It refers to the expense incurred in providing responses to the queries. The decreased response cost contributes to an improvement in the system's overall performance. The calculation takes place in the following condition (13),

$$Rc_k = \sum_{q \in Q} C(Mt) \qquad (10)$$

This represents the cost of the query response when materialised views are used, and this represents the total cost of the response. Condition 14 depicts the results of the fitness assessment based on these cost estimations.

$$F(q) = Min(Qp_k + Mc_k + Rc_k) \qquad (11)$$

In this case, denotes the evaluated fitness value, which ought to be lower for views that have been optimally selected. The performance of the system can be improved by lowering the costs associated with the various processing steps. The reduction of the total weighted processing cost is the primary objective of the MV selection process. If the MVs are selected properly, they will be able to provide accurate responses to the queries in a shorter amount of time. The CHECO method's pseudocode can be found in Algorithm 1, which describes the method.

Input: Query workloads (W_Q), number of views (V_k), maximum number of iterations ($I_{\max}$).

Output: Optimal materialized view selection

Begin

Initialization of variables (number of queries and views)

For each data population k do

Arbitrary selection of queries in views

Compute the fitness evaluation using different costs in (11)

Identify the finest position of views as global best (G_{best})

If the finest solution is not attained,

Then, update position using condition (5) to (9),

For, $K = 1 to N$ do

Calculate the fitness value using condition (14)

If $F(q) < 0.5$

Then update the position towards optimal

Else

If $F(q) \geq 0.5$

Then, update the positions using condition (9)

End

End

If the fitness position equals to G_{best} at $I = I_{\max}$, then

Return optimal materialized views

End

End

Algorithm 1: Pseudocode of CHECO algorithm

For quicker and more effective selection, the queries in the data warehouse are chosen using hybrid Ebola with COATI optimization (CHECO) schemes. Here, the best MV is chosen in order to respond to queries quickly and efficiently.

IV. RESULTS AND DISCUSSION

The performance of the suggested materialised view selection is assessed in this section using ensemble approaches. In terms of various performance metrics, such as query processing cost, maintenance cost, total cost, execution time, and maintenance cost, the presented methodology is compared to various existing schemes. ACOMVS (Ant-Colony optimization-based MVS), CROMVS (Coral reefs optimization-based MVS), PRoREs, PHAN [25], EA (evolutionary algorithm), SRCSAMVS (Stochastic ranking based cuckoo search algorithm for MV selection) [23], and YANG's [25] algorithm are some of the current methods. The well-known dataset TPC-H is used to assess the effectiveness of the proposed model for MVS [27].

4.1 Performance metrics

The effectiveness of the suggested approach is validated in this section using a variety of performance metrics. The following subsections provide descriptions of them.

4.1.1 Execution time

The number of queries that are executed during the materialised view selection process takes time. The subsequent condition (15) computes it.

$$T_{exe} = t_{End} - t_{initial} \qquad (15)$$

Here, T_{exe} represents the execution time, $t_{initial}$ represents the initial time and t_{End} represents the ending time.

4.1.2 Query processing cost

It is the cost taken to process the queries in the proposed methodology. It is computed by expressed condition (16),

$$Qp(\cos t) = \sum Q \times C_k^q \qquad (16)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8171
***Article Received:** 29 June 2023 **Revised:** 26 August 2023 **Accepted:** 06 September 2023*

Here, $Qp(\cos t)$ represents the query processing cost, Q denotes the frequency count of queries, and C_k^q represents the accessing cost of k queries.

4.1.3 Maintenance cost

Every time one of its respective base associations is reorganised, this cost is necessary to restore the views. The following condition (17) serves as an expression for the maintenance calculation

$$M(\cos t) = \sum Q_{updated} \times Ca_N^q \qquad (17)$$

Here, Mc_k denotes the maintenance cost, $Q_{updated}$ denotes the updated frequency of queries, and Ca_N^q signifies the maintenance cost of view k for an updated base association N

4.1.4 Total cost

It represents the total cost of all MV processes, including the processing, answering, and maintaining of queries. It is expressed in the following circumstances (18),

$$T(\cos t) = \sum C_K \qquad (18)$$

Here, $T(\cos t)$ represents the total cost and C_K represents the consumed cost of each process.

4.2 Performance analysis

This section compares various contemporary methodologies to the performance of the methodology that is being presented. Figure 3 shows the query process comparison of the suggested approach is analysed in execution time.

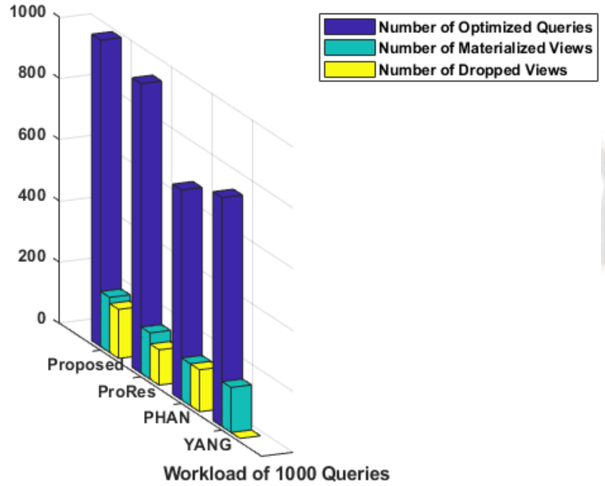

Figure 3: Comparison of query processing

In figure 3, the performance of the proposed methodology in a number of queries materialized views and dropped views for 1000 queries are illustrated. In this scenario, the time it takes to put the proposed scheme into action is noticeably less than that required by the competing approaches. The proposed execution time is compared with the execution times of existing approaches such as ProRes, PHAN, and YANG [25]. The time needed to complete computations using the proposed method is significantly less than that required by the competing methods. Then, the comparative analysis of the suggested plan using a range of different numbers of queries is shown in figure 4, as shown above.

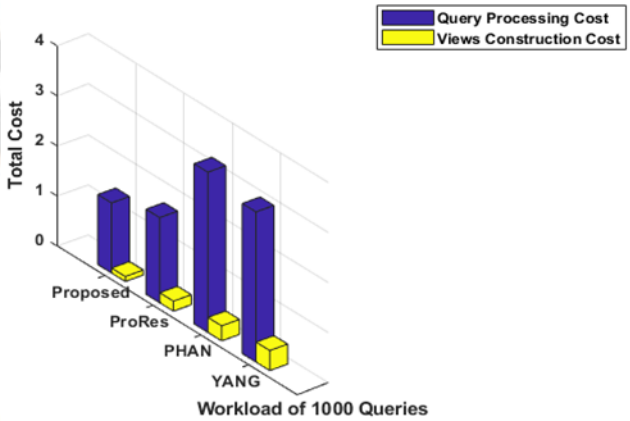

Figure 4: Comparison of proposed approach by varying number of queries

Because of the high construction cost and the cost of processing queries, Figure 4 examines the developed system using a total of one thousand queries. The comparison of the proposed method's performance to that of ProRes, YANG, and PHAN [25] approaches demonstrated that the proposed methodology is capable of achieving superior performance. Figure 5 presents the total costs incurred for a variety of query counts, which can be compared side by side.

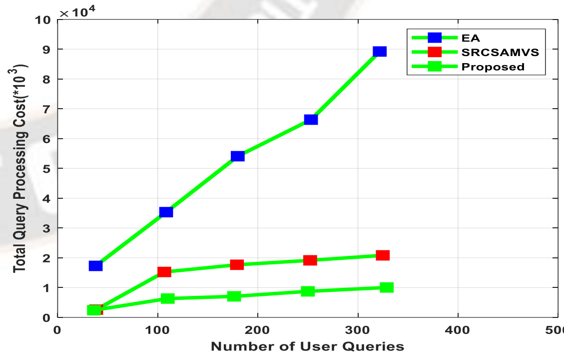

Figure 5: Comparison of total query processing cost by varying queries

Figure 5 is an illustration of a comparison examination of the total process cost. Several of the currently available methods, such as EA and SRCSAMVS, are utilised in its investigation [23]. The cost analysis demonstrated that the processing cost of the proposed scheme is significantly lower than the costs of the approaches that were compared for different numbers of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8171
***Article Received:** 29 June 2023 **Revised:** 26 August 2023 **Accepted:** 06 September 2023*

queries, specifically 100, 200, 300, 400, and 500, respectively. In addition, the comparative analysis of he cost of query processing with other methods that are currently in use is shown in figure 6.

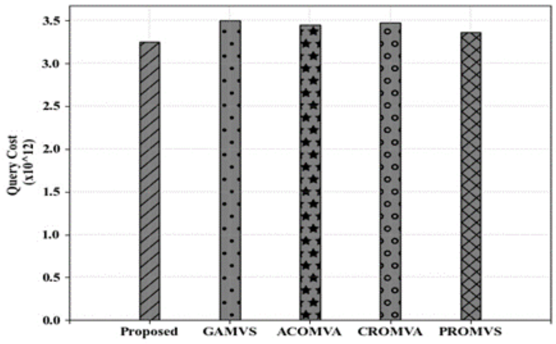

Figure 6: Comparison analysis of query cost

Figure 6 presents a comparison of the query cost of the proposed scheme with that of several other existing approaches, including GAMVS, ACOMVS, PSOMVS, and CROMVS [26]. The developed method has significantly lower costs associated with the processing of queries in comparison to other approaches that are currently in use. Figure 7 presents the results of a comparison of performance in terms of total cost.

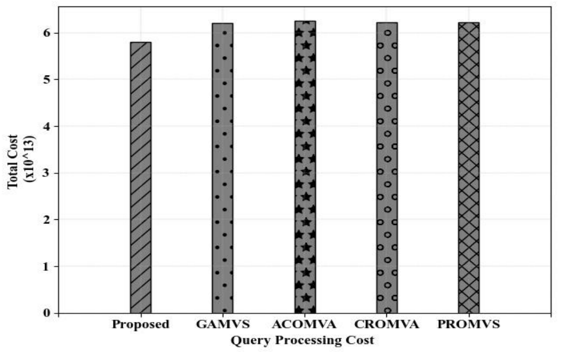

Figure 7: Comparison of the total cost

Figure 7 compares the total cost of the newly developed method to the cost of other methods already in use. In this context, the total cost value of the proposed scheme is significantly lower than that of the approaches that are being contrasted. Table 1 contains information about an examination that compares and contrasts various cost values.

Table 1 Comparative analysis of various cost values

Methods	Maintenance cost (*10^{12})	Query Processing cost (*10^{12})	Total cost (*10^{12})
ACOMVS	6.32935	3.52286	9.85221
GAMVS	6.32935	3.52286	9.85221
CROMVS	6.32935	3.52286	9.85221
PSOMVS	6.32936	3.52286	9.85222
Proposed	**6.32825**	**3.46944**	**9.79769**

The comparisons of the various monetary values for costs are presented in table 1. It is mentioned that the cost of the proposed scheme is lower than the costs of the various approaches that are compared. Figure 8 presents a comparison of the overall performance in terms of cost for different sized datasets.

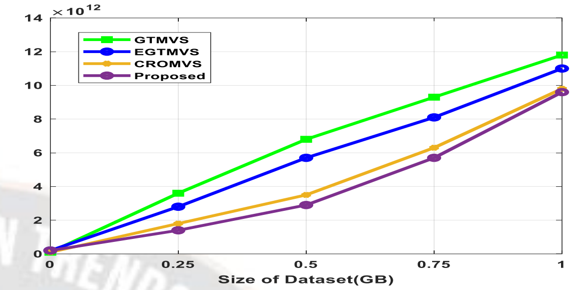

Figure 8: Comparison of the total cost for varying dataset size

Figure 8 compares the total cost of the developed scheme to several different approaches that are currently in use for a variety of datasets. The performance of the newly developed methodology is superior to that of the approaches that were analysed previously. The existing CROMVS, EGTMVS, and GTMVS approaches will be used as benchmarks for the performance comparison. The performance comparison with regard to total cost is significantly lower than the approaches that are being compared. Figure 9 provides a visual representation of the results of a comparison of the costs of maintenance.

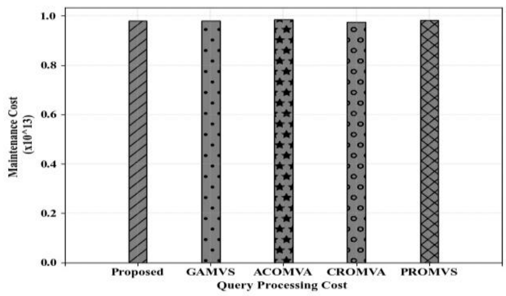

Figure 9: Performance comparison of maintenance cost

A comparison of the upkeep costs of the suggested method with those of several other methods already in use can be found in figure 9. In this scenario, the proposed method incurs fewer expenses in terms of upkeep in comparison to the other methods examined, such as GAMVS, ACOMVS, CROMVS, and PSOMVS [26]. Figure 10 presents the findings of an analysis comparing the costs of maintenance for different sized datasets.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8171
***Article Received:** 29 June 2023 **Revised:** 26 August 2023 **Accepted:** 06 September 2023*

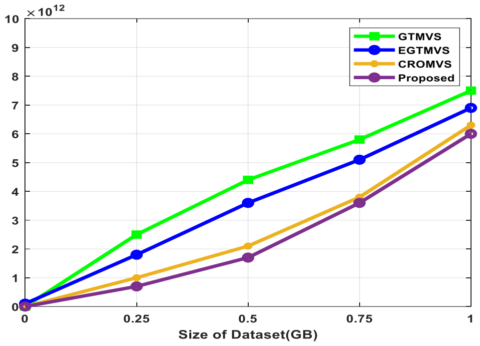

Figure 10: Maintenance cost for varying sizes of the dataset

Figure 10 presents a comparison of the performance of different dataset sizes in terms of their associated maintenance costs. The developed method has lower costs associated with its ongoing maintenance in comparison to other methodologies currently in use, such as CROMVS, EGTMVS, and GTMVS [26] approaches. Table 2 presents the results of a comparison of performance metrics regarding execution speed.

Table 2: Execution time comparison

Data size (GB)	Proposed (seconds)	ACOMVS (seconds)	PSOMVS (seconds)	GAMVS (seconds)	CROMVS (seconds)
0.25	**10**	14	13	15	13
0.50	**61**	85	70	90	81
0.75	**99**	175	120	185	160
1	**144**	250	165	265	225

The performance comparison for the amount of time it takes to execute is provided in table 2. In this case, the proposed approach achieves a faster execution time than the approaches that are being compared. In addition, the comparison of the amount of time required for execution is shown in figure 11.

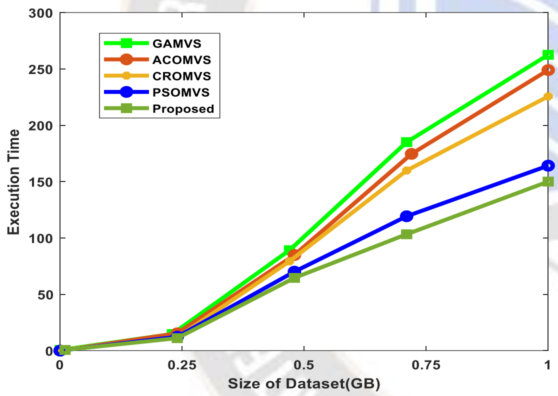

Figure 11: Comparison of execution time

Figure 11 presents the results of an analysis of the performance of the execution time. The proposed method is evaluated alongside various other methodologies already in use, such as GAMVS, ACOMVS, CROMVS, and PSOMVS [26]. The time needed to carry out the proposed strategy is significantly less than that required by the competing methods. The proposed method offers significant improvements in a variety of performance aspects in comparison to the approaches that were considered.

Evaluation of Statistical analysis

In this part of the report, we conduct an evaluation of the statistical aspects of the proposed methodology. To begin, the analysis of variance, or ANOVA, is considered for the purpose of determining the exact difference between the methods and the results. The ANOVA test is evaluated using datasets of varying sizes and a number of different queries. As a result, the ANOVA test is utilised in order to investigate the various results. In this section of the proposed work, the TPC-H dataset is taken into consideration for the ANOVA test. Table 3 contains information about the test analysis performed for various queries using the TPC-H dataset.

Table 3: ANOVA test analysis of execution time for some queries using the TPC-H dataset

Methods	Count	Sum	Average	Variance
EA	4	485	102	7687
GAMVS	4	497	124.25	11, 148.25
ACOMVS	4	470	117.5	10, 443.66
PSOMVS	4	269	67.25	3334.25
SRCSMAVS	4	365	87	6745
ProRes	4	324	92	4567
CROMVS	4	434	108.5	9209
FSAMVS	4	250	60.10	2,748.2
Proposed (CHECO)	**4**	**235**	**56**	**2135**

The results of the analysis of variance performed on the investigation into the amount of time required to complete a task using the TPC-H database are presented in Table 3. For the purposes of carrying out the ANOVA test, the significance level of confidence has been set to. In addition, the analysis of the ANOVA test using different sized datasets, as presented in table 4, was carried out.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8171
***Article Received:** 29 June 2023 **Revised:** 26 August 2023 **Accepted:** 06 September 2023*

Table 4: ANOVA test analysis of execution time for data size using the TPC-H dataset

Methods	Count	Sum	Average	Variance
EA	4	578	98	9845
GAMVS	4	568	142	11, 610
ACOMVS	4	540	135	10, 574.6
PSOMVS	4	352	88	3236
SRCSMAVS	4	435	92	5679
ProRes	4	410	78	8764
CROMVS	4	509	127.25	9215.583
FSAMVS	4	325	80	2764
Proposed (CHECO)	**4**	**296**	**72**	**2267**

The results of an ANOVA test of time analysis based on the amount of data are presented in Table 4, which makes use of the TPC-H database. According to the analysis of the data, the strengths and weaknesses are computed before being compared with various other approaches that already exist. In Table 5, we compare and contrast the advantages and disadvantages of the proposed method with those of the existing one.

Table 5 Comparison of strengths and weaknesses of proposed with existing approaches

Methods	Overall cost	Execution time
EA	Weak	Applicable
GAMVS	Applicable	Weak
ACOMVS	Applicable	Weak
PSOMVS	Weak	Applicable
SRCSMAVS	Moderate	Weak
ProRes	Applicable	Moderate
CROMVS	Moderate	Applicable
FSAMVS	Applicable	Applicable
Proposed	**Applicable**	**Applicable**

Table 5 makes it abundantly clear that the proposed method is applicable to any and all applications. The currently available models are not optimal for all of the available applications.

V. CONCLUSION

In this paper, an optimal selection of materialized views was presented by making use of an efficient combination of ensemble approaches. To begin, an ensemble combination of constraint handling approaches is presented for an optimal selection of queries. This is followed by the presentation of the results. For the purpose of making an optimal selection, various constraints, such as stochastic ranking, epsilon constraint, and self-adaptive penalty, are taken into consideration here. After that, a combination of Ebola and COATI optimization is applied in order to facilitate a faster and more accurate selection of queries in views. In this context, various fitness parameters such as maintenance cost, query processing cost, and response cost are taken into consideration in order to improve performance. The performance of the developed MV selection[30] is validated with various current approaches in terms of various performance metrics including query processing cost, maintenance cost, total cost, execution time, and maintenance cost respectively. In the future, it will be possible to enhance the materialized view selection by using further enhanced processes, and it will also be possible to analyze it using a variety of benchmark datasets containing a large number of records.

REFERENCES

[1] Adnan, R., & Abbas, T. M. (2020). Materialized Views Quantum Optimized Picking for Independent Data Marts Quality. Iraqi Journal of Information and Communications Technology, 3(1), 26-39.

[2] Gjengset, J. F. R. (2021). Partial state in dataflow-based materialized views (Doctoral dissertation, Massachusetts Institute of Technology).

[3] Raipurkar, A. R., &Chandak, M. B. (2021). Optimized execution method for queries with materialized views: Design and implementation. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-15.

[4] Sohrabi, M. K., & Azgomi, H. (2019). Evolutionary game theory approach to materialized view selection in data warehouses. Knowledge-Based Systems, 163, 558-571.

[5] Gosain, A., &Sachdeva, K. (2020). Random Walk Grey Wolf Optimizer Algorithm for Materialized View Selection (RWGWOMVS).In Novel Approaches to Information Systems Design (pp. 101-122).IGI Global.

[6] Azgomi, H., &Sohrabi, M. K. (2021). MR-MVPP: A map-reduce-based approach for creating MVPP in data warehouses for big data applications. Information Sciences, 570, 200-224.

[7] Solanki, S. S. (2018). Incremental Maintenance of a Materialized View in Data Warehousing: An Effective Approach. Global Journal of Computer Science and Technology.

[8] Verma, A., Bhattacharya, P., Bodkhe, U., Ladha, A., &Tanwar, S. (2020, March). Dams: Dynamic association for view materialization based on rule mining scheme. In The International Conference on Recent Innovations in Computing (pp. 529-544).Springer, Singapore.

[9] Mohseni, M., &Sohrabi, M. K. (2020). MVPP-based materialized view selection in data warehouses using simulated annealing. International Journal of Cooperative Information Systems, 29(03), 2050001.

[10] Shantinee Thasan, Abdul Rahman, Parameswaran Subramanian, Maria Josephine Williams. (2023). An Intelligent Decision Support System to Aid Profit Planning in Manufacturing Companies. International Journal of Intelligent Systems and

Applications in Engineering, 11(4s), 345–356. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2674.

[11] Gosain, A., &Sachdeva, K. (2019). Selection of materialized views using stochastic ranking based Backtracking Search Optimization Algorithm. International journal of system assurance engineering and management, 10(4), 801-810.

[12] Gosain, A., & Madaan, H. (2018). Query prioritization for view selection. In Progress in Intelligent Computing Techniques: Theory, Practice, and Applications (pp. 403-410). Springer, Singapore.

[13] Ordonez-Ante, L., Van Seghbroeck, G., Wauters, T., Volckaert, B., & De Turck, F. (2020). A workload-driven approach for view selection in large dimensional datasets. Journal of Network and Systems Management, 28(4), 1161-1186.

[14] Fadhil, Z. M. (2020). Human behavior based particle swarm optimization for materialized view selection in data warehousing environment. Periodicals of Engineering and Natural Sciences, 8(4), 2367-2378.

[15] Ordonez-Ante, L., Van Seghbroeck, G., Wauters, T., Volckaert, B., & De Turck, F. (2019). Automatic View Selection for Distributed Dimensional Data.In IoTBDS (pp. 17-28).

[16] Berkani, N., Bellatreche, L., & Ordonez, C. (2018, May). ETL-aware materialized view selection in semantic data stream warehouses. In 2018 12th International Conference on Research Challenges in Information Science (RCIS) (pp. 1-11). IEEE.

[17] Betouati, F., & Rahal, S. A. (2019). A scalable approach to model big and interacted queries for materialized view through data mining. Multiagent and Grid Systems, 15(2), 137-154.

[18] Kumar, S., & Vijay Kumar, T. V. (2018). A novel quantum-inspired evolutionary view selection algorithm. Sādhanā, 43(10), 1-20.

[19] Yusoh, Z. M., Gan, K. B., &Emran, N. A. (2020, March). Materialized view selection problem using genetic algorithm for manufacturing execution system. In Journal of Physics: Conference Series (Vol. 1502, No. 1, p. 012044). IOP Publishing.

[20] Kumar, A., & Kumar, T. V. (2018). Materialized view selection using set based particle swarm optimization. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 12(3), 18-39.

[21] Sohrabi, M. K., &Ghods, V. (2016). Materialized View Selection for a Data Warehouse Using Frequent Itemset Mining. J. Comput., 11(2), 140-148.

[22] Prakash, J., & Kumar, T. V. (2021). A multi-objective approach for materialized view selection. In Research Anthology on Multi-Industry Uses of Genetic Programming and Algorithms (pp. 512-533). IGI Global.

[23] Kharat, V., &Shelar, M. (2021). An Efficient Query Optimizer with Materialized Intermediate Views in Distributed and Cloud Environment. Tehničkiglasnik, 15(1), 105-111.

[24] Sahoo, D. K. . (2021). Improved Routing and Secure Data Transmission in Mobile Adhoc Networks Using Trust Based Efficient Randomized Multicast Protocol. Research Journal of Computer Systems and Engineering, 2(2), 06:11. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/25.

[25] Gosain, A., &Sachdeva, K. (2020). Materialized View Selection for Query Performance Enhancement Using Stochastic Ranking Based Cuckoo Search Algorithm. International Journal of Reliability, Quality and Safety Engineering, 27(03), 2050008.

[26] Roy, S., Shit, B., Sen, S., &Cortesi, A. (2021). Construction and distribution of materialized views in Non-binary data space. Innovations in Systems and Software Engineering, 17(3), 205-217.

[27] Mouna, M. C., Bellatreche, L., &Boustia, N. (2022). ProRes: Proactive re-selection of materialized views. Computer Science and Information Systems, (00), 3-3.

[28] Azgomi, H., &Sohrabi, M. K. (2019). A novel coral reefs optimization algorithm for materialized view selection in data warehouse environments. Applied Intelligence, 49(11), 3965-3989.

[29] https://www.tpc.org/tpch/

[30] Dehghani, M., Montazeri, Z., Trojovská, E., & Trojovský, P. (2023). Coati Optimization Algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems. Knowledge-Based Systems, 259, 110011.

[31] Srinivasarao, P., & Satish, A. R. (2023). Multi-objective materialized view selection using flamingo search optimization algorithm. Software: Practice and Experience, 53(4), 988-1012.

[32] Srinivasarao, Popuri, and Aravapalli Rama Satish. "A Novel Hybrid Optimization Algorithm for Materialized View Selection from Data Warehouse Environments." Computer Systems Science & Engineering 47.2 (2023).