



# Evolutionary Algorithms for Query Optimization in Distributed Database Systems: A review

Zulfiqar Ali<sup>1,2</sup>, Hafiza Maria Kiran<sup>1</sup> and Waseem Shahzad<sup>2</sup>

<sup>1</sup> Department of Computer Science and IT The University of Lahore, 1Km, Defence Road Lahore, Pakistan,

<sup>2</sup> Department of Computer Science National University of Computer and Emerging Sciences, A.K Brohi Road, H-10/4, Islamabad, Pakistan.

[zulfiqar.ali, hafiza.maria]@cs.uol.edu.pk, [zulfiqar.ali, waseem.shahzad]@nu.edu.pk

## KEYWORD

Query optimization;  
Distributed database;  
Evolutionary Computation;  
Bio-inspired algorithms

## ABSTRACT

*Evolutionary Algorithms are bio-inspired optimization problem-solving approaches that exploit principles of biological evolution, such as natural selection and genetic inheritance. This review paper provides the application of evolutionary and swarms intelligence based query optimization strategies in Distributed Database Systems. The query optimization in a distributed environment is challenging task and hard problem. However, Evolutionary approaches are promising for the optimization problems. The problem of query optimization in a distributed database environment is one of the complex problems. There are several techniques which exist and are being used for query optimization in a distributed database. The intention of this research is to focus on how bio-inspired computational algorithms are used in a distributed database environment for query optimization. This paper provides working of bio-inspired computational algorithms in distributed database query optimization which includes genetic algorithms, ant colony algorithm, particle swarm optimization and Memetic Algorithms.*

## 1. Introduction

Evolutionary computation and swarm intelligence based algorithms are bio inspired optimization techniques used for problems solving in the several fields of study i.e. Data Mining (Alcalá-Fdez *et al.*, 2009; Freitas, 2005), Bioinformatics (Das, Abraham, & Konar, 2008; Deutsch, 2003), Image Processing (Bounsaythip & Alander, 1997; Parker, 2010; Senthikumar & Rajesh, 2009), Medical Image Processing (Hill, Batchelor, Holden, & Hawkes, 2001; Klein, Staring, & Pluim, 2007; Maulik, 2009), Network Security (Banković, Stepanović, Bojanić, & Nieto-Taladriz, 2007; Burbank, 2008; Garcia-Teodoro, Diaz-Verdejo, Maciá-Fernández, & Vázquez, 2009), Business Analytics (Bose, 2009; Ghatasheh, 2014; Kaisler, Armour, Espinosa, & Money, 2013), Wireless Sensor Networks (Abbasi & Younis, 2007; Bara'a & Khalil, 2012; Jourdan & de Weck, 2004; Kleinschmidt, 2009), Routing Protocol Optimization (Akkaya & Younis, 2005; Ali & Shahzad, 2011, 2013), Emerging Pattern Mining (Ali & Shahzad, 2017; Novak, Lavrač, & Webb, 2009), Associative Classification (Jabbar, Deekshatulu, & Chandra, 2013; Shahzad, 2010; Shahzad & Baig, 2011; Soliman & Adly, 2012) and Distributed Database Systems (Allahverdi & Al-Anzi, 2006; Larrañaga & Lozano, 2001; Ulus & Uysal, 2003).



Evolutionary and Swarm Intelligence based approaches are also exploited in Distributed Databases Systems for the purpose of data fragmentation (Ahmad, Karlapalem, Kwok, & So, 2002; Goli & Rankoohi, 2012), data allocation and query optimization (Sevinç & Coşar, 2010).

Query optimization in the distributed environment is challenging task and NP-Hard optimization problem due to the several constraints (i.e. decentralized database, number of sites, the involvement of communication network, cost of data transmutation etc.) of the distributed system. There are various factors that to be considered during the design of distributed database system i.e. distributed query processing cost, memory used, CPU cost, input-output cost, data transmission cost, response time and total time for the execution of the query etc. However, the bio-inspired optimization approaches are more suitable for the distributed database systems and promising to be the more efficient and cost-effective for the overall system. The literature survey shows there are made some efforts to provide the information about the state-of-the-art query processing and query optimization approaches used in the field of Distributed Database Systems (DDS).

Ashok Kumar and Ankita Ahuja provided a collection of different methods for query optimization in distributed databases. Alaa Aljanaby, Emad Abuelrub, and Mohammed Odeh surveyed the different query optimization methods and models applied in the distributed database systems in (Aljanaby, Abuelrub, & Odeh, 2005). Pankti Doshi and Vijay Raisinghani provided the review of dynamic query optimization strategies in the distributed database in (Doshi & Raisinghani, 2011). Yasmeen R. M. Umar and Amit R. Welekar provide a review of the various techniques used for Query Optimization in Distributed Database in (Umar & Welekar, 2014). Ihab F. Ilyas, George Beskales, and Mohamed A. Soliman provided the survey of Top-k Query Processing Techniques in Relational Database Systems in (Ilyas, Beskales, & Soliman, 2008).

The contemporary literature shows that there are little bit review papers that provide the collection of evolutionary and swarm intelligence based query processing and optimization approaches. This review paper provides the collection of bio-inspired query optimization algorithm used in the distributed database environment. The problem of query optimization in a distributed database environment is one of the complex problems. There are several techniques which exist and are being used for query optimization in a distributed database. The intention of this research is to focus on how bio-inspired computational algorithms are used in a distributed database environment for query optimization. This paper briefly describes the working of most common and well known bio-inspired computational algorithms in distributed database query optimization which includes genetic algorithms (GA) (Holland, 1975), ant colony algorithm (ACO) (Dorigo, Maniezzo, & Coloni, 1996), particle swarm optimization (PSO) (Eberhart & Kennedy, 1995) and Memetic Algorithms (MA) (Moscato, 1989).

The remaining paper consists of following sections; Section II provides the introduction of the distributed database system and various components of the distributed environment. Section III explains the query optimization and importance of query optimization in the distributed database system. Section IV provides the prominent evolutionary approaches exploited for the query optimization in distributed database systems. Section V provides the review of the evolutionary algorithms used for the query optimization in the distributed systems. Section VI gives the taxonomy of the evolutionary approaches for DDS and discussion on the application of evolutionary algorithms for distributed systems.

## 2. Distributed Database System

The distributed database has an important influence on data processing. A distributed database is a blend of the computer network and database systems which are physically isolated but are centralized logically (Shahabi, Khan, & McLeod, 2000). In the distributed database, the data is distributed in several systems and they are connected via a network. In order to perform local application, each site is capable of independent processing in the network. Distributed database management system is basically a software system which allows the management of distributed database and ensures distribution transparency to users (Ozsu & Valduriez, 1994). So, the term distributed database system is basically the combination of both distributed database and distributed database management system. A distributed database is distributed among different sites and the reason behind the data distribution is nature of the data or performance which inherent. The data is not essentially an independent entity on each site in the distributed database but it can be relevant to the data which is stored on another site

(Stanchev, 2001). Distributed database ensures better reliability, good performance, and availability (P. Tiwari & S. V. Chande, 2013). It also improves file sharing, expandability, and security of an organization in which database is distributed geographically. In a distributed database, the distribution of data is done in two ways which are a partition and full replication (Dong & Liang, 2007). This distribution of the data causes difficulties to query processing. In a distributed database, query processing comprises much data communication among member sites. This communication of data is leading factor to control whole query optimization and processing.

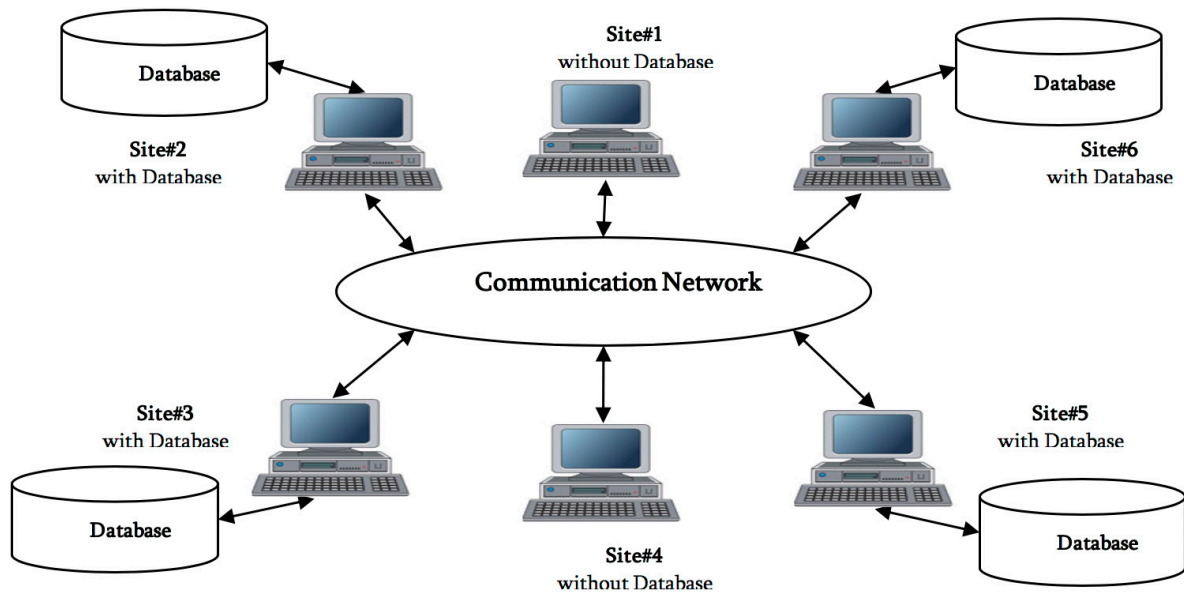


Figure 1: Distributed Databases System Architecture

### 3. Query Optimization

The process of selecting an appropriate plan for query evaluation from different strategies which can be probable for processing a query is called query optimization (Silberschatz, Korth, & Sudarshan, 1997). In grid databases, query optimization is one basic problem normally when there are the replication and storage of database on several network nodes. In large distributed databases, it is really very difficult to solve query optimization problem as it is the NP-Hard type of problem (Sohal, Singh, & Virk, 2015). The performance of the query is highly affected by the cost of communication for shifting the data across multiple sites. The concept of query optimization is well defined in (M. P. Tiwari & S. V. Chande, 2013) that the practice of discovering the finest execution plans for a particular query which will be the execution strategy for the query. In query processing, query optimization is one of the major features. Due to some aspects for example availability of memory, size of the database, data allocation, transmission of the data etc. the task of query optimization is very difficult in a distributed database (M. P. Tiwari & S. V. Chande, 2013). In order to create a query execution plan, a query optimizer is used which embodies query execution at the lowest cost (Aljanaby et al., 2005). There are a number of methods and optimization algorithms available to deal with several optimization problems (Salimi, 2015).

#### 3.1. Evolutionary Techniques for Optimization

Some of the well-known computational systems which have been developed for finding the appropriate best solution of optimization problems are: ant colony optimization (ACO), particle swarm optimization (PSO), genetic algorithms (GAs), memetic algorithms (MAs) and shuffled frog leaping (SFL). The contemporary literature shows that there are four most prominent evolutionary algorithms used for the various problems in different fields: 1) Genetic Algorithm that is proposed by the Holland in (Holland, 1975), Genetic Programming by Koza

(Koza, 1992), Evolutionary Strategies by Rechenberg in (Back, 1996) and Evolutionary Programming by Fogel et al. in (Fogel, 1999). Moreover, the traditional differences between the approaches involve the nature of the representation schemes, the reproduction operators, and the selection methods.

The contemporary literature shows that GA is mostly exploited for the query optimization in distributed database systems from the evolutionary computing paradigm. However, this study focused on the query optimization approaches based on the genetic algorithms. Swarm Intelligence (SI) is a subfield of Computational Intelligence which provides a solution for complex optimization problems which are not easily tackled by other approaches. Swarm Intelligence mainly consists of Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO) and Honeybees paradigms. PSO is developed in 1995 by James Kennedy and Russell Eberhart (Eberhart & Kennedy, 1995) and ACO Colomni Dorigo and Maniezzo in early 1990 (Dorigo, Maniezzo, Colomni, & Maniezzo, 1991). This survey study includes the ACO and PSO exploitation for the query optimization in the distributed environment.

Furthermore, the basic working principle and Pseudo Code of algorithmic approaches of evolutionary computation (GA and MA) and swarm intelligence (ACO and PSO) are given in the next subsections.

### 3.1.1. Genetic Algorithm

Genetic algorithms are basically similar to biological systems and based on alteration, selection, and crossover (Elbeltagi, Hegazy, & Grierson, 2007). This algorithm is capable of providing multiple solutions of optimization problems which are intractable. Although the technique of genetic algorithms in query optimization is talented there is a trade-off in this technique between speed and excellence and this technique cannot be implemented for all the problems like multiple object problems, electric circuit problems, traveling salesman problems etc. A set of values for optimization variables is held by genes (Elbeltagi et al., 2007). Best chromosomes interchange the information through alteration or fitness for the purpose of pretending the natural existence of fitness process and generate offspring (children) chromosomes. These offspring chromosomes are then appraised and used to grow the population if they deliver better solutions than weak population members. This process is carried out to get the best fit solution for a large number of generations.

#### **Pseudo code for Genetic Algorithm:**

1. Selection of primary random population of individuals
2. Individuals' fitness evaluation
3. Repeat
  - a. Selection of best individuals used by genetic operators
  - b. Generation of new individuals by using mutation and crossover
  - c. New individuals' fitness evaluation
  - d. Replacement of worst individuals by the best individuals
4. Until stop criteria

### 3.1.2. Ant Colony Optimization

Ant colony algorithm (ACO) was first presented by M. Dorigo and fellows in 1990s as a unique nature-inspired meta-heuristics for the purpose of solving serious combinational optimization (CO) problems (Dorigo, 1992; Dorigo et al., 1991). Ant colony optimization belongs to approximate algorithms from a class of meta-heuristics for getting appropriate solutions to problems related to optimization in less computation time (Blum & Roli, 2003; Crainic & Toulouse, 2003). The purpose of the production of ant colony optimization is to make a model for searching shortest path with minimum cost (MMICT, , MMU, Garg, & Juneja, 2012). The artificial ants' behavior is quite similar to the real ants and these ants define their way from source to destination. A well-known application of ant colony optimization is shortest path problem which occurs mostly in the field of telecommunication networks. Ant colony optimization depends on the artificial system and it based on colonies of real

ants (Wagh & Nemade, 2017). In (Wagh & Nemade, 2017), modified ant colony optimization algorithm is used for query optimization and comparison with other query optimization algorithms is done. The results show that modified any colony algorithm performs well in terms of computation time as compared to other algorithms.

Ant colony algorithm is the best fit for the dynamic type of applications. The purpose of ant colony algorithm is to build a model for searching the shortest path with minimum cost. The behavior and performance of artificial ants are based on real ants. All ants define the path which covers the source point to the destination. Ants have the ability to discover the shortest path for the means of food. These ants release pheromone on the traveling track so that they can communicate with other ants. When they find the food, they start moving back by carrying food and the following pheromone and still releasing more pheromone on the traveling track (El-beltagi, Hegazy, & Grierson, 2005). All the quantitative information about the cost of the path and qualitative information like a load of traffic on the network is discovered by the ant while finding the path. Ant colony algorithm is best for telecommunication network problems.

#### **Pseudo code for ACO:**

1. *Initialization of parameters*
2. *Initialization of array heuristics*
3. *Initialization of pheromone matrix*
4. *Do, while stopping conditions not satisfied*
5. *Build solution*
  - a. *Apply local search*
  - b. *Updating pheromone*
6. *End while*
7. *Outlook best solution*
8. *Stop*

#### **3.1.3. Particle Swarm Optimization**

Particle swarm optimization (PSO) was first presented by Eberhart and Kennedy in 1950s. It is basically optimization and computational search method that is biologically inspired and based on the behaviors of fish schooling or birds flocking (Rini, Shamsuddin, & Yuhanz, 2011). For the purpose of improvement in quality and the speed, some variations are developed in particle swarm optimization. But basic simple particle swarm optimization is more suitable for solving simple optimization problems. There are following three steps in particle swarm optimization which are particles generation, position and velocity and last one is position and velocity update (MMICT et al., 2012).

#### **Pseudo code for PSO:**

1. *Begin*
2. *Set parameter  $t=0$*
3. *Initialization of particles  $p(t)$*
4. *Evaluation of particles  $p(t)$*
5. *While*
  - a. *Until termination conditions are not satisfied*
  - b. *Begin*
    - i. *Set  $t=t+1$*
    - ii. *Update weights*
    - iii. *Selection of  $pBest$  for each particle*

- iv. Selection of  $gBest$  from  $p(t-1)$
- v. Calculation of particle's velocity  $p(t)$
- vi. Update particle's position
- vii. Evaluation of particles  $p(t)$
- c. End
- 6. End

### 3.1.4. Memetic Algorithms

The term memetic algorithms (MAs) (Moscato, 1989) was acquainted in the late 80s with mean a group of met heuristics that have a central topic the hybridization of various algorithmic methodologies for a given problem. Unique accentuation was given to the utilization of a populace based approach in which an arrangement of collaborating and contending operators was occupied with times of individual change of the arrangements while they sporadically cooperate.

The term memetic algorithm was first introduced by Moscato in (Moscato, 1989) in a technical report. Memetic algorithms are inspired by both of the principles of natural evolution Dawkins' and Darwinian notion of a meme. It was first proposed as a mean of passing on the message that, albeit moving for some, organic advancement ought not to compel that creative energy to create populated based strategies. Different types of advancement might be quicker, with social development being one of those less-prohibitive cases. The major functional area in the memetic algorithms is the process of generational step. Evolution of solution is done in this part of the algorithm. The generational step consists of three main components which are selection, reproduction, and update. It is the responsibility of first step and the third step for competition aspects of individuals in the population. The effectiveness and efficiency of individuals are evaluated by utilizing the information which is given by problem dependent guiding function. In order to create new solutions, a part of individuals is chosen on the basis of efficiency and effectiveness measure. The individual selection is done by using non-proportionate and fitness-proportionate methods.

#### Pseudo code for Memetic Algorithms:

1. Begin
2. Initialization of population
3. Evaluation of each candidate
4. Repeat until a termination condition is not satisfied
  - a. Parents selection
  - b. Recombination to produce offspring
  - c. Mutate offspring
  - d. Offspring's improvement via local search
  - e. Evaluation of offspring
5. Selection of individuals for next generation
6. End

## 4. Evolutionary Techniques for Query Optimization in Distributed Database Systems

The web has gained much attention and is being a source of attraction for its users. It has become the important means of switching information, exchanging ideas and sharing services. In this context, there are a lot of tools

developed for information sharing over the internet. So, in order to store, manage and model a large amount of information, the technology of database management system (DBMS) plays a significant role in efficient query processing on a large amount of data. The DBMS technology also provides the facility to ensure the integrity of the data, data manipulation in a reliable manner, synchronization of the data for concurrent access and data security(Ouzzani & Bouguettaya, 2004). There was need of development of such systems which can retrieve the information from the web by using some efficient techniques. In this regard, information retrieval (IR) systems play a vital role. But the most difficult and important task of these information retrieval systems was query generation which correctly recognizes related stuff and rejects the others(Wang & Sun, 2008).

To understand the behavior and the structure of the web content, researchers have proposed a lot of models and studies for efficient presentation of relevant data on the web. There are a number of evolutionary algorithms which have been proposed by researchers for the purpose of near optimum solution searching to the mathematical type of problems. Evolutionary algorithms (EAs) are basically included searching methods which totally based on the living behavior of the species and the comparison of the natural biological evolution of these species (Holland, 1975). For example, how the birds discover their destination when they move and how the ants explore (Dorigo et al., 1996) the shortest route to move towards destination or food. Some of the well-known computational systems which have been developed for finding appropriate best solution of optimization problems are: ant colony optimization (ACO), particle swarm optimization (PSO), genetic algorithms (GAs), memetic algorithms (Mas) and shuffled frog leaping (SFL)(Elbeltagi et al., 2007).

#### 4.1. Application of GAs for query optimization in distributed database

For query optimization in a distributed database, dynamic programming is not practicable. A new genetic algorithm (NGA) technique which is basically a genetic algorithm based query optimizer is proposed in (Sevinç & Coşar, 2010). The performance of the proposed technique is then compared with exhaustive and random algorithms. The experimentation of the proposed technique has done on a synthetic database which includes replicated relations. The results of the experimentation and comparisons show that the proposed new genetic algorithm (NGA) helped to achieve improvement over earlier genetic algorithms.

Several genetic algorithms and heuristic algorithms are explored in (Zhou, 2007) for query optimization for the large-scale distributed database. In this study, the problem of query optimization is decreased to a join ordering problem which is somehow similar to traveling salesman problem. The experimentation was done on these algorithms and the results were compared. Experimentation shows that genetic and heuristic algorithm is a feasible approach for the solution of query optimization problem on the large-scale distributed database system. In distributed database query optimization, a genetic algorithm is a more preferred and viable technique because of its flexible nature and robustness. The GA is also capable to explore and achieve search spaces.

#### 4.2. Application of ACO for query optimization in distributed database

A technique of multi-colony ant algorithm is proposed in (Padia, Khulge, Gupta, & Khadilkar) for optimization of join queries in a distributed database environment in which relations can be replicated. The proposed ACO algorithm is based on MAX-MIN ant system. For query plan creation and execution of the query, four types of ant colonies work together in this proposed algorithm. This results in the growth of scalability and modularity. In the process of finding an optimal plan, separately each type of ant makes a significant decision. The quality of the plan which is generated is then evaluated by using two cost models which based on response time and total time. The proposed algorithm is then compared with two other genetic algorithms and the result shows that the quality of proposed technique is much better than previous techniques and approximate 80% of optimization time is saved in this algorithm.

There are a lot of query optimization techniques available for the distributed database but it has been observed that when ant colony optimization algorithm is combined with some other query optimization algorithm, then the performance of distributed query optimization enhanced (Padia et al.). The experimentation results of the hybrids of ACO show that this hybrid approach is really effective and feasible for query optimization problems. The implementation of proposed hybrid technique has shown that it can produce worthwhile solutions

when a number of joins and the size of query grow in the distributed database as well as in relational database management system.

As query optimization is really a complex task in a distributed database environment, there are so many techniques which are used for query optimization purpose. A modified version of ant colony algorithm (ACO) is proposed in (Wagh & Nemade, 2017) in order to find an appropriate solution for query optimization. A comparison is then done among the existing ant colony algorithm and modified ant colony algorithm and the results show that modified ant colony algorithm consume less computation time as compared to others and simple ant colony algorithm.

### 4.3. Application of PSO for query optimization in distributed database

A collection of hybrid particle swarm optimization algorithms and canonical algorithms are developed for modeling the problems of distributed database query optimization as Bare Bones particle swarm optimization (Dokeroglu, Tosun, & Cosar, 2012). Particle swarm optimization algorithms proficiently work for distributed database query optimizations and also generates a plan for query execution which can be compared with the plans generated by genetic algorithms. In particle swarm optimization algorithm, the movement of particles depends on probability distribution instead of velocity parameters manipulation. The experimentation of this work was done on different synthetic databases.

A substantial number of questions are posted on databases spread over the globe. Keeping in mind the end goal to process these inquiries effectively, ideal inquiry handling methodologies that create productive question preparing plans are being concocted. In spite of the fact that it isn't computationally achievable to investigate all conceivable inquiry designs in such a vast hunt space, the question arranges for that gives the most financially savvy choice for question preparing is viewed as important and ought to be produced for a given inquiry. Regarding this, an endeavor has been made to produce such ideal question designs utilizing Set based Comprehensive Learning Particle Swarm Optimization (S-CLPSO) (Kumar, Kumar, & Singh, 2013). Exploratory examinations of this calculation with the GA based dispersed question design age calculation demonstrates that for a higher number of relations, the S-CLPSO based calculation can create nearly better quality Top-K inquiry designs.

### 4.4. Application of Memetic Algorithms for Query Optimization in Distributed Database

Memetic algorithms are somehow same as genetic algorithms but in memetic algorithms, the element is known as memes instead of genes which produce chromosome. The word meme describes the concept of a unit of cultural transmission. Memetic algorithms include local search process in order to refine the solution before involving an evolutionary process. Memes also have self-replicating property same as genes but the difference is they are not inherited as genes. They are spread through imitation. The novel feature of memetic algorithms is that before getting involved in the evolutionary search, offspring and chromosomes are allowed to increase the experience through local search (Kumar et al., 2013).

A novel technique improved memetic algorithm (IMA) is proposed in (Deulkar & Narvekar, 2015) which is basically based on a memetic algorithm (MA) for searching in case of complex queries. A hybrid selection policy is used in the proposed technique in order to improve the search results. Moreover in IMA, in order to make the population diversity preserved, chromosomes are personalized to be different. The performance of proposed technique improved memetic algorithm is then evaluated by comparing the results of improved IMA and basic memetic algorithm. And the comparison shows that improved IMA provides better and high-quality solutions as compared to basic MA.

Adaptive cellular memetic algorithm (ACMA) technique is proposed in (Melita, Gopinath, & Sebsibe, 2015) for optimization of query results. The performance of the proposed adaptive cellular memetic algorithm is then analyzed by comparing the results produced by ACMA with the results produced by other genetic algorithms. The comparison exposed that the proposed algorithm improved the efficiency than other genetic algorithms. Regardless of the enhancement of data, the algorithm's performance improved retrieval time.



## 5. Discussion

The Web is persistently pulling in floods of new users and specialist co-ops. It is currently the true medium for trading data and services. This globalization has additionally impelled the advancement of apparatuses and helps to explore and share data in corporate intranets that were beforehand open online just prohibitively and at restrictive expenses. The data age unrest has featured the part of the database management system (DBMS) as a key empowering innovation. DBMSs are right now the innovation of decision for displaying, putting away, overseeing, and productively questioning a lot of data. They likewise give capacities to ensuring information respectability, encouraging solid and adaptable information access and control, synchronizing simultaneous gets to from applications, and securing data.

There are various bios inspired evolutionary algorithms which are used for query optimization in a distributed database environment. According to the literature survey and previous studies, the most common and well-known techniques and algorithms are discussed in this paper and also mentioned in figure 2. The major techniques and algorithms which have been used include genetic algorithms (GAs), ant colony optimization (ACO) and particle swarm optimization (PSO). Figure 2 shows a hierarchical structure of the algorithms and their extended form. In a distributed database environment, the problem of query optimization is solved by using a simple genetic algorithm and new genetic algorithm. The technique of ant colony algorithm is used for query optimization as generic ant colony algorithm, multi ant colony algorithm, hybrid ant colony algorithm and modified ant colony algorithm. Similarly, particle swarm optimization is also used for query optimization in the distributed database as simple particle swarm optimization algorithm and S-CLPSO.

The literature and previous studies show that these evolutionary and bio-inspired algorithms are significant in the distributed database for solving the task of query optimization. These evolutionary algorithms are much better and efficient as compared to other query optimization algorithms. Experiments and the comparisons of GAs with other algorithms show that it produces better quality results than other algorithms because of its flexible nature and robustness. Ant colony algorithm also generates better solutions as compared to other techniques and algorithms. It also consumes less computation time improve results.

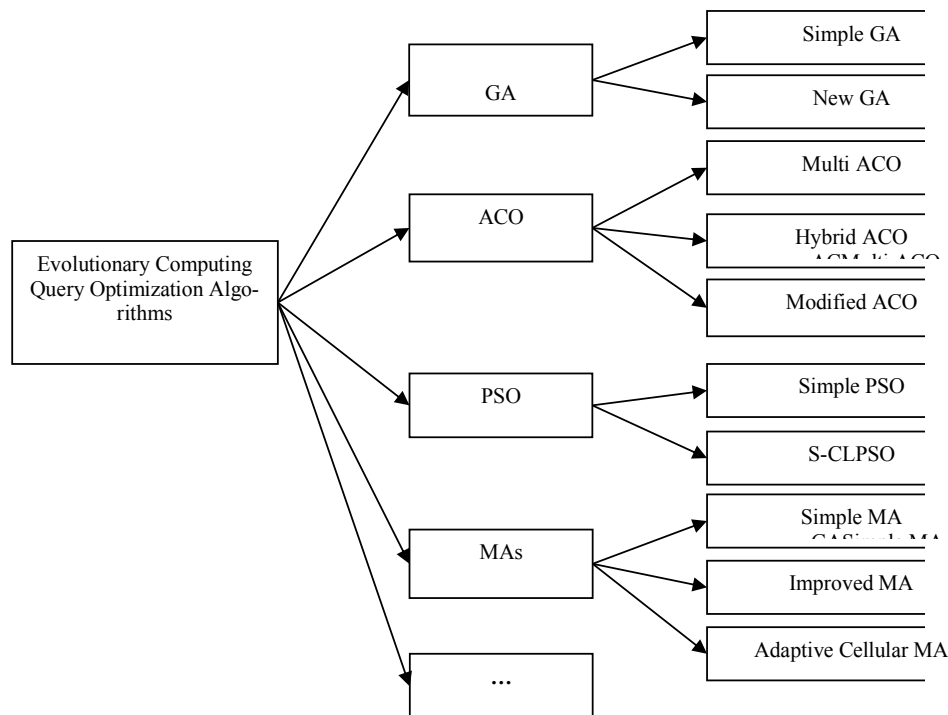


Figure 2: Evolutionary algorithms for query optimization in a distributed database environment

## 6. Conclusion

This review article investigate the application of evolutionary and swarms intelligence for creation of optimized query plans in the distributed environments i.e. ant colony algorithm, particle swarm optimization, genetic algorithm and memetic algorithm. This review paper elaborates the problem of query optimization in distributed databases and explore the application of evolutionary computation for the solution of this problem. The working principle of bio-inspired computational algorithms in distributed database query optimization is explained. The contemporary studies reveals that bio-inspired approaches provide promising results in query optimization for distributed database systems as compared to the conventional techniques.

## 7. References

- Abbasi, A. A., & Younis, M. (2007). A survey on clustering algorithms for wireless sensor networks. *Computer communications*, 30(14-15), 2826-2841.
- Ahmad, I., Karlapalem, K., Kwok, Y.-K., & So, S.-K. (2002). Evolutionary algorithms for allocating data in distributed database systems. *Distributed and Parallel Databases*, 11(1), 5-32.
- Akkaya, K., & Younis, M. (2005). A survey on routing protocols for wireless sensor networks. *Ad hoc networks*, 3(3), 325-349.
- Alcalá-Fdez, J., Sanchez, L., Garcia, S., del Jesus, M. J., Ventura, S., Garrell, J. M., . . . Rivas, V. M. (2009). KEEL: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 13(3), 307-318.
- Ali, Z., & Shahzad, W. (2011). Critical analysis of swarm intelligence based routing protocols in adhoc and sensor wireless networks. Paper presented at the Computer Networks and Information Technology (ICCNIT), 2011 International Conference on.
- Ali, Z., & Shahzad, W. (2013). Analysis of routing protocols in ad hoc and sensor wireless networks based on swarm intelligence. *International Journal of Networks and Communications*, 3(1), 1-11.
- Ali, Z., & Shahzad, W. (2017). EPACO: a novel ant colony optimization for emerging patterns based classification. *Cluster Computing*, 1-15.
- Aljanaby, A., Abuelrub, E., & Odeh, M. (2005). A Survey of Distributed Query Optimization. *Int. Arab J. Inf. Technol.*, 2(1), 48-57.
- Allahverdi, A., & Al-Anzi, F. S. (2006). A PSO and a Tabu search heuristics for the assembly scheduling problem of the two-stage distributed database application. *Computers & Operations Research*, 33(4), 1056-1080.
- Back, T. (1996). *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*: Oxford university press.
- Banković, Z., Stepanović, D., Bojanić, S., & Nieto-Taladriz, O. (2007). Improving network security using genetic algorithm approach. *Computers & Electrical Engineering*, 33(5-6), 438-451.
- Bara'a, A. A., & Khalil, E. A. (2012). A new evolutionary based routing protocol for clustered heterogeneous wireless sensor networks. *Applied Soft Computing*, 12(7), 1950-1957.
- Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys (CSUR)*, 35(3), 268-308.
- Bose, R. (2009). Advanced analytics: opportunities and challenges. *Industrial Management & Data Systems*, 109(2), 155-172.
- Bounsaythip, C., & Alander, J. T. (1997). Genetic algorithms in image processing-a review. Paper presented at the Proceedings of the Third Nordic Workshop on Genetic Algorithms and their Applications (3NWGA).
- Burbank, J. L. (2008). Security in cognitive radio networks: The required evolution in approaches to wireless network security. Paper presented at the Cognitive Radio Oriented Wireless Networks and Communications, 2008. CrownCom 2008. 3rd International Conference on.
- Crainic, T. G., & Toulouse, M. (2003). Parallel strategies for meta-heuristics *Handbook of metaheuristics* (pp. 475-513): Springer.

- Das, S., Abraham, A., & Konar, A. (2008). Swarm intelligence algorithms in bioinformatics Computational Intelligence in Bioinformatics (pp. 113-147): Springer.
- Deulkar, K., & Narvekar, M. (2015). An Improved Memetic Algorithm for Web Search. *Procedia Computer Science*, 45, 52-59.
- Deutsch, J. (2003). Evolutionary algorithms for finding optimal gene sets in microarray prediction. *Bioinformatics*, 19(1), 45-52.
- Dokeroglu, T., Tosun, U., & Cosar, A. (2012). Particle Swarm Intelligence as a new heuristic for the optimization of distributed database queries. Paper presented at the Application of Information and Communication Technologies (AICT), 2012 6th International Conference on.
- Dong, H., & Liang, Y. (2007). Genetic algorithms for large join query optimization. Paper presented at the Proceedings of the 9th annual conference on Genetic and evolutionary computation.
- Dorigo, M. (1992). Optimization, learning and natural algorithms. Ph. D. Thesis, Politecnico di Milano, Italy.
- Dorigo, M., Maniezzo, V., & Coloni, A. (1996). Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(1), 29-41.
- Dorigo, M., Maniezzo, V., Coloni, A., & Maniezzo, V. (1991). Positive feedback as a search strategy.
- Doshi, P., & Raisinghani, V. (2011). Review of dynamic query optimization strategies in distributed database. Paper presented at the Electronics Computer Technology (ICECT), 2011 3rd International Conference on.
- Eberhart, R. C., & Kennedy, J. (1995). Particle swarm optimization. Paper presented at the Proceedings of the IEEE international conference on neural networks.
- Elbeltagi, E., Hegazy, T., & Grierson, D. (2005). Comparison among five evolutionary-based optimization algorithms. *Advanced engineering informatics*, 19(1), 43-53.
- Elbeltagi, E., Hegazy, T., & Grierson, D. (2007). A modified shuffled frog-leaping optimization algorithm: applications to project management. *Structure and Infrastructure Engineering*, 3(1), 53-60.
- Fogel, L. J. (1999). *Intelligence through simulated evolution: forty years of evolutionary programming*: John Wiley & Sons, Inc.
- Freitas, A. A. (2005). Evolutionary algorithms for data mining *Data mining and knowledge discovery handbook* (pp. 435-467): Springer.
- Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *computers & security*, 28(1-2), 18-28.
- Ghatasheh, N. (2014). Business analytics using random forest trees for credit risk prediction: A comparison study. *International Journal of Advanced Science and Technology*, 72(2014), 19-30.
- Goli, M., & Rankoohi, S. M. T. R. (2012). A new vertical fragmentation algorithm based on ant collective behavior in distributed database systems. *Knowledge and Information Systems*, 30(2), 435-455.
- Hill, D. L., Batchelor, P. G., Holden, M., & Hawkes, D. J. (2001). Medical image registration. *Physics in medicine & biology*, 46(3), R1.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems. An introductory analysis with application to biology, control, and artificial intelligence*. Ann Arbor, MI: University of Michigan Press, 439-444.
- Ilyas, I. F., Beskales, G., & Soliman, M. A. (2008). A survey of top-k query processing techniques in relational database systems. *ACM Computing Surveys (CSUR)*, 40(4), 11.
- Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013). Heart disease prediction system using associative classification and genetic algorithm. *arXiv preprint arXiv:1303.5919*.
- Jourdan, D. B., & de Weck, O. L. (2004). Layout optimization for a wireless sensor network using a multi-objective genetic algorithm. Paper presented at the Vehicular technology conference, 2004. VTC 2004-Spring. 2004 IEEE 59th.
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. Paper presented at the System sciences (HICSS), 2013 46th Hawaii international conference on.
- Klein, S., Staring, M., & Pluim, J. P. (2007). Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE transactions on image processing*, 16(12), 2879-2890.
- Kleinschmidt, J. H. (2009). Genetic Algorithms for Wireless Sensor Networks *Encyclopedia of Artificial Intelligence* (pp. 755-758): IGI Global.
- Koza, J. R. (1992). *Genetic Programming II, Automatic Discovery of Reusable Subprograms*: MIT Press, Cambridge, MA.

- Kumar, T. V., Kumar, A., & Singh, R. (2013). Distributed query plan generation using particle swarm optimization. *International Journal of Swarm Intelligence Research (IJSIR)*, 4(3), 58-82.
- Larrañaga, P., & Lozano, J. A. (2001). *Estimation of distribution algorithms: A new tool for evolutionary computation (Vol. 2)*: Springer Science & Business Media.
- Maulik, U. (2009). Medical image segmentation using genetic algorithms. *IEEE Transactions on information technology in biomedicine*, 13(2), 166-173.
- Melita, L., Gopinath, G., & Sebsibe, H. (2015). Web Search Query Result Optimization based on Memetic Algorithms: A Comparative Study. *International Journal of Computer Science Issues (IJCSI)*, 12(3), 240.
- MMICT, B., MMU, H., Garg, I. A., & Juneja, I. D. (2012). A Comparison and analysis of various extended techniques of query optimization.
- Moscato, P. (1989). On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms. *Caltech concurrent computation program, C3P Report*, 826, 1989.
- Novak, P. K., Lavrač, N., & Webb, G. I. (2009). Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research*, 10(Feb), 377-403.
- Ouzzani, M., & Bouguettaya, A. (2004). Query processing and optimization on the web. *Distributed and Parallel Databases*, 15(3), 187-218.
- Ozsu, M. T., & Valduriez, P. (1994). Distributed data management: unsolved problems and new issues. *Readings in Distributed Computing Systems*, 512-544.
- Padia, S., Khulge, S., Gupta, A., & Khadilkar, P. Query Optimization Strategies in Distributed Databases.
- Parker, J. R. (2010). *Algorithms for image processing and computer vision*: John Wiley & Sons.
- Rini, D. P., Shamsuddin, S. M., & Yuhaniz, S. S. (2011). Particle swarm optimization: technique, system and challenges. *International Journal of Computer Applications*, 14(1), 19-26.
- Salimi, H. (2015). Stochastic fractal search: a powerful metaheuristic algorithm. *Knowledge-Based Systems*, 75, 1-18.
- Senthilkumaran, N., & Rajesh, R. (2009). Edge detection techniques for image segmentation—a survey of soft computing approaches. *International journal of recent trends in engineering*, 1(2), 250-254.
- Sevinç, E., & Coşar, A. (2010). An evolutionary genetic algorithm for optimization of distributed database queries. *The Computer Journal*, 54(5), 717-725.
- Shahabi, C., Khan, L., & McLeod, D. (2000). A probe-based technique to optimize join queries in distributed internet databases. *Knowledge and information systems*, 2(3), 373-385.
- Shahzad, W. (2010). *Classification and Associative Classification Rule Discovery Using Ant Colony Optimization*. National University.
- Shahzad, W., & Baig, A. (2011). Hybrid associative classification algorithm using ant colony optimization. *International Journal of Innovative Computing, Information and Control*, 7(12), 6815-6826.
- Silberschatz, A., Korth, H. F., & Sudarshan, S. (1997). *Database system concepts (Vol. 4)*: McGraw-Hill New York.
- Sohal, M., Singh, A., & Virk, R. S. (2015). A Framework For Optimizing Distributed Database Queries Based On Stochastic Fractal Search. *Int. J. Comp. Sc. and Mob. Computing (JCSMC)*, 4(6), 544-551.
- Soliman, O. S., & Adly, A. (2012). Bio-inspired algorithm for classification association rules. Paper presented at the Informatics and Systems (INFOS), 2012 8th International Conference on.
- Stanchev, L. (2001). *Semantic Data Control in Distributed Database Environment*. Survey Paper for CS748T Distributed Database Management, University of Waterloo, 1-25.
- Tiwari, M. P., & Chande, S. V. (2013). Query optimization strategies in distributed databases. *International Journal of Advances in Engineering Sciences*, 3(3), 23-29.
- Tiwari, P., & Chande, S. V. (2013). Optimization of distributed database queries using hybrids of Ant colony optimization algorithm. *International Journal of Advanced Research in Computer Science and Software Engineering*.
- Ulus, T., & Uysal, M. (2003). Heuristic approach to dynamic data allocation in distributed database systems. *Pakistan Journal of Information and Technology*, 2(3), 231-239.

- Umar, Y. R., & Welekar, A. R. (2014). Query Optimization in Distributed Database: A Review. Query Optimization in Distributed Database: A.
- Wagh, A., & Nemade, V. (2017). Query Optimization using Modified Ant Colony Algorithm. International Journal of Computer Applications, 167(2).
- Wang, Z., & Sun, X. (2008). An efficient web query optimization algorithm based on LDA and MA. Paper presented at the MultiMedia and Information Technology, 2008. MMIT'08. International Conference on.
- Zhou, Z. (2007). Using heuristics and genetic algorithms for large-scale database query optimization. Journal of Information and Computing Science, 2(4), 261-280.



