

## STRATEGIES ON LARGE SIZED DATA IMPLEMENTING META-HEURISTIC APPROACH IN BUSINESS INTELLIGENCE

Jyoti Prakash MISHRA<sup>1</sup>, Zdzislaw POLKOWSKI<sup>2</sup>, Sambit Kumar MISHRA<sup>3</sup>, Samarjeet BORAH<sup>4</sup>

<sup>1</sup>Gandhi Institute for Education and Technology, Baniatangi, Bhubaneswar, affiliated to Biju Patnaik University of Technology, Rourkela, E\_mail : [jpmishra@gietbbsr.com](mailto:jpmishra@gietbbsr.com)

<sup>2</sup>Wroclaw University of Economics and Business, Poland, E\_mail: [zdzislaw.polkowski@ue.wroc.pl](mailto:zdzislaw.polkowski@ue.wroc.pl)

<sup>3</sup>Gandhi Institute for Education and Technology, Baniatangi, Bhubaneswar, affiliated to Biju Patnaik University of Technology, Rourkela, E\_mail : [sambitmishra@gietbbsr.com](mailto:sambitmishra@gietbbsr.com)

<sup>4</sup>Sikkim Manipal Institute of Technology, Majitar, East Sikkim, E\_mail : [samarjeet.b@smit.smu.edu.in](mailto:samarjeet.b@smit.smu.edu.in)

### Abstract

*In broad sense the term business intelligence is linked to enhancement of determination of performance, integration of databases along with planning as well as making decisions on the specified performance. In such a situation it is required to improve the decision support facilitating the process of data integration along with initiation of actions. The significance in this scenario is to carry out proper information management along with making decisions in different changed environments. It is also understood that the business intelligence is responsible towards making decisions in case of modification of business strategies even if there is an influence of information management towards decision making. It is clear that business intelligence is associated with the mechanisms implemented towards analysis of data of business sectors. Basically it has the provision of predictive as well as actual views of operations in business. The technologies associated with business intelligence can manage huge sized structured as well as unstructured data to build new strategic opportunities in business. The main intention in this case is to permit for accessing the large sized data easily. Considering the decisions related to strategic business, it is seen that the objective is achieved provided the involvement of decisions linked to internal or external data are consistent. During involvement of such activities, it is essential to focus on decisions on queries and continuous transactions to manage effectively the mechanism.*

**Key Words :** Business intelligence, Transactional Data, Meta-heuristic, Join operation, Pattern matching, Schema integration, Pheromone

## 1. INTRODUCTION

In general, it is required implement the proper system of transaction like enterprise resource planning initially prior to business intelligence, where it is required to update the transactional data in the relational databases. Anyway the solutions linked to business intelligence enable to make decisions over the large sized heterogeneous data and to obtain the result with optimality. It should be noted that to enhance and optimize the information stored in enterprise resource systems, the architectural concepts should be maximized to support the



This is an open-access article distributed under the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>).

decision capabilities with data analysis. The recent implementation of relational databases in such systems is well capable as well as provisioned with high level aggregated data.

In specific cases, there is the significance of management information system to accumulate data from various sources to support in report generation and making decisions. In many situations, the main aim is to enhance the value and profits of the business. Though there are several mechanisms for such activities, primarily it is required to compile the complete information to support solving the problems either structured or unstructured. There should be provision to accumulate reports at all levels. Also there should be the support to automate the workflow and facilitate the flow of functions within all routines in the business strategies providing the base level data. The major benefit in such scenario is to enhance the operational efficiency identifying the strength and aspects of business processes and to focus on tactical non programmed decisions. Similarly it is needed to accumulate the especially financial standardized information facilitating the system with stipulated regulations. Primarily it is focused on the basic design procedure along with size of associated databases with common automated generation features.

Very often it is also required to initiate the process to retrieve the patterns linked to large data sets to transform the extract information into comprehensible structure. Sometimes it is inclined to uncover the adequate knowledge during instantiation of database processes. As the databases are involved along with the aspects of data management, the primary focus in such situation is to extract the patterns linked with large scaled heterogeneous data applying the business intelligence and data extraction mechanisms. During retrieval and instantiation of large scaled data the dependencies among the data must be preserved and also sometimes it is desirable to implement decision making mechanisms to obtain more accurate prediction results. It is also acceptable that business intelligence can be implemented to identify focused data to support towards better management decisions. To focus on the activity it is required to include the key performance indicators as well as the bench marks of linked processes with data visualization to enhance the quality of data.

## **2. REVIEW OF LITERATURE**

H. Chen et al.[1] in their work have signified the big data and business analytics considering the performance gains. As such these have been signaled linked to specific domain associated with data of most valuable organizational resources.

A. Ashrafi et al.[2] in their studies demonstrated the values associated with big data and business analytics along with their association with agility, innovation and competitive performance. In fact the it is required to focus on the recurring finding towards deriving values from big data, and build the organizational capacity to identify areas within the business and beneficiary towards strategically plan and data analytics tasks.

R.H. Chiang et al.[3] in their work have emphasized on translation of big data analytics and subsequent merits. They observed that it may not be sufficient to focus on the resources required to extract meaning out of data, somehow it may be essential to adopt a viewpoint of leveraging data.

I.O. Pappas et al.[4] in their work have focused on ecosystem towards generation of data and necessity to leverage this data towards digital transformation and sustainable societies.

G. Hindle et al.[5] during their studies have tried to develop a framework to place business analytics linked to the organizational and environmental sectors. In fact they have also tried to

expand various facets of big data analytics, focusing on the sources of data generation, aspects of sharing the data following the privacy and ethics.

P. Mikalef et al.[6] in their work focused on various aspects of big data with characterization to volume, variety, velocity, and veracity.

J.J. Seddon et al.[7] during their studies have discussed the various parameters as well as dimensions associated with the domain of big data focusing on attributes. In their work they have also tried to examine the implications linked to leveragability of data towards business outcomes.

K. Conboy et al.[8] in their work have studied the importance of analytics-based activities associated with sensing, seizing as well as reconfiguring processes. Also they have focused on complementary aspects of the data utilized for analytics projects.

M. Janssen et al.[9] in their studies have focused towards provisioning of big data with accuracy and accumulating the decisions. Also they have highlighted on the attributes of data influenced towards leveragability, and not explored, particularly in relation to value creation.

F.P.S. Surbakti et al.[10] in their work focused on the interdependencies of sparsely empirically researched data. In fact these are appropriately granular and suitable towards data resources.

R. Dubey et al.[11] during their studies have focused on some specific empirical work associated with driving forces of adoption considering the constraints towards initiation of gradual maturation of big data and business analytics. Mainly they want to focus on how do organizations develop their big data and business analytics capabilities and what are the constraining forces, pressures and enablers while adopting and maturing big data and business analytics? Somehow it is lined to the ability to leverage big data and business analytics based on the empirical examination on resource based views.

### **3. PROBLEM FORMULATION**

#### **3.1. Steps to prioritize data**

Step 1: Initially accumulate the data from different sources linked to the system

Step 2: Apply the pattern matching techniques with the data inclusive of multiple databases in the system

Step 3: Initiate the schema integration technique to process the integration of data

Step 4: Based on the results of query, the quality of data should be ascertained.

Step 5: Focusing on the process of preparation of data, it is required to transform and construct the data again and eradicate the noise or unnecessary facts

Step 6: Apply the aggregation operations on data and maintain the higher level hierarchy of data

Step 7: Apply the normalization techniques on attributes of each database and focus on additional construction of attributes if any within the process

#### **3.2. Steps to classify and associate data using descriptive mechanism**

Step1: Retrieve the relevant data and focus on training set as well as test set data. Specifically the test set data is responsible to achieve the accuracy on data.

Step 2: Identify the similarities and differences within the data and observe the similarities of attributes within each database

Step 3: Apply the mechanism to analyze and link the relationship within the data and to uncover the hidden patterns within the data.

Step 4: Apply the techniques within the data items in the data set along with outlier analysis

Step 5: Identify the similar patterns linked to transactional data

Step 6: Apply prediction technique on sequential patterns and data classification

### **3.3. Strategies and formulation and on optimization of join queries**

Considering the search spaces along with query evaluation plans it is required to focus on replication of databases to check the storage of relations in multiple locations. Initially the semi join operations are considered on the databases and based on response time the potential parallelism is to be focused. Generally the I/O cost is linked with the local processing cost; therefore cost of individual processing elements is also accompanied with the communication cost. It is known that the semi joins are implemented to minimize the cardinalities of the relations associated with cost of projections and join attributes. In such case, the cost of relations with the signified attributes can be evaluated considering the length of relations along with projection of join attributes. Similarly the response time can also be evaluated considering the execution of parallel operations. For example, while considering a particular relation during operation instead of focusing only the domain or tuple for the same, it can be observed that the joins to the specific location can be implemented in parallel as the joins can be performed at different locations not accompanied with the similar resources.

### **3.4. Meta-heuristic approach**

Generally the meta-heuristic is a higher side heuristic or procedure to obtain the specific heuristic having a provision of better or optimal solution focused to specified optimization problems. It is confined to some few related assumptions based on the optimization problem towards achieving feasible solutions. In fact the meta-heuristics approach does not guarantee to obtain optimal solution globally rather some meta-heuristics also implement some version of stochastic optimization to observe dependency of generated random variables towards feasible solutions. In this scenario the strategies guide the search process having the common goal to obtain near optimal solution.

### **3.5. Application of ant colony optimization**

Basically there many techniques are available those are inspired through the behavior of ant colonies and focused to different optimization problems. The specific population based ant colony optimization is implemented to obtain feasible solutions for quite complex optimization problems. In this application, the artificial ants can act as software agents and apply search mechanisms to obtain better or feasible solutions for the desired optimization problem. To achieve this optimization problem can be transformed into problems of obtaining better paths and the ants can incrementally generate solutions by traversing the paths. Of course the solution mechanism in this case is stochastic based on pheromone concepts. Now comparing this issue with query optimization process it is observed that the queries linked to databases primarily are linked to cost sensitive process and gradually can be enhanced based on the success factors. In this case the join operators are responsible towards optimization process to minimize the runtime.

### 3.6. Determination of query plans using ant colony optimization techniques

Obviously specific consideration on individual ant is done in this approach to obtain the optimal strategy based on the prior decision.

- a. The ants should determine the sequence and specification of join indices and operations on join. The number of joins accumulated in the queries should focus towards unique operations on join.
- b. The ants should also determine the locations and in such case, the joins in the query should obtain uniform result in each location.
- c. The ants should be confined with each database within the relation and number of relations should be accompanied with join queries.

The solution linked to query evaluation plans is then obtained evaluating the ants and optimality can be occurred focusing the minimal cost during the process.

### 3.7. Algorithm

Step 1: Assign the maximum number of iterations, e.g. max\_it=500

Step 2: Define the population size, e.g. size of ants, 50 and initialize the constraint to 1

Step 3: Set the initial pheromone parameter, tpp<sub>0</sub>=10

Step 4: Set the pheromone exponential weight factor, ewf=0.3

Step 5: Set the evaporation rate, evpr=0.1

Step 6: Initiate the process and observe dynamism on ants

```
for i=1:max_it
for j=1:antq
ant(j).ewf=[];
for n=1:vt
q= tpp0(:,1).^ evpr;
q(ant(j).ewf)=0;
end
end
ant(i).Cost=CostFunction(ant(j).ewf);
if ant(j).Cost<optimal.Cost
optimal_value=ant(j);
end
end
```

Step 7: Initialize the parameters including the pheromone trails, t

Step 8: for counter=1 to upper-limit

for ant=1 to n

build feasible solution based on pheromones

Step 9: Evaluate the cost of each solution based on pheromones

Step 10: If the solution obtained has the solution at minimal cost, update the same as optimal solution, else reinitialize the pheromones

Step 11: Update the pheromone trails based on the optimal solution obtained and check the status of pheromones above the minimum criteria

#### 4. EXPERIMENTAL ANALYSIS

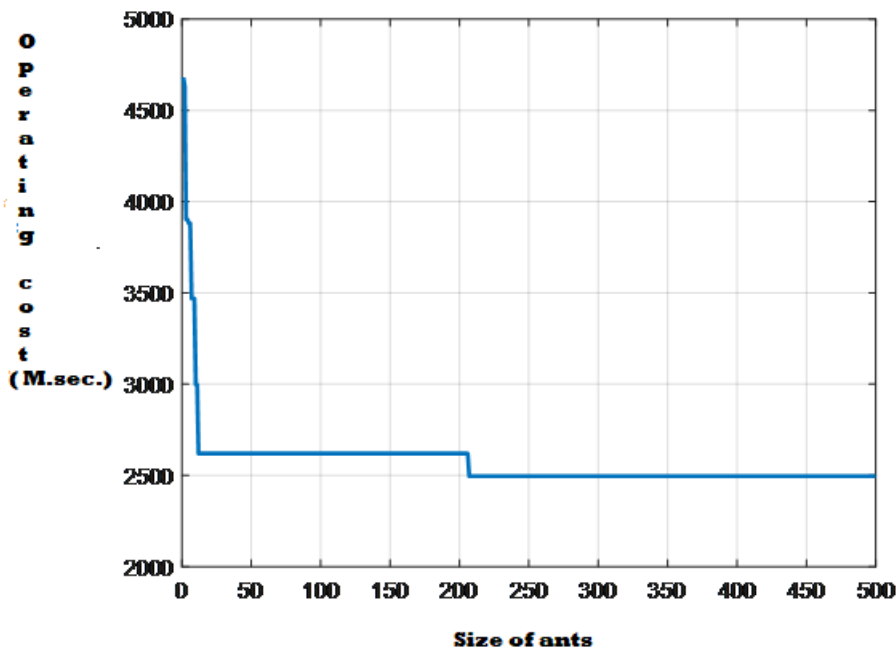


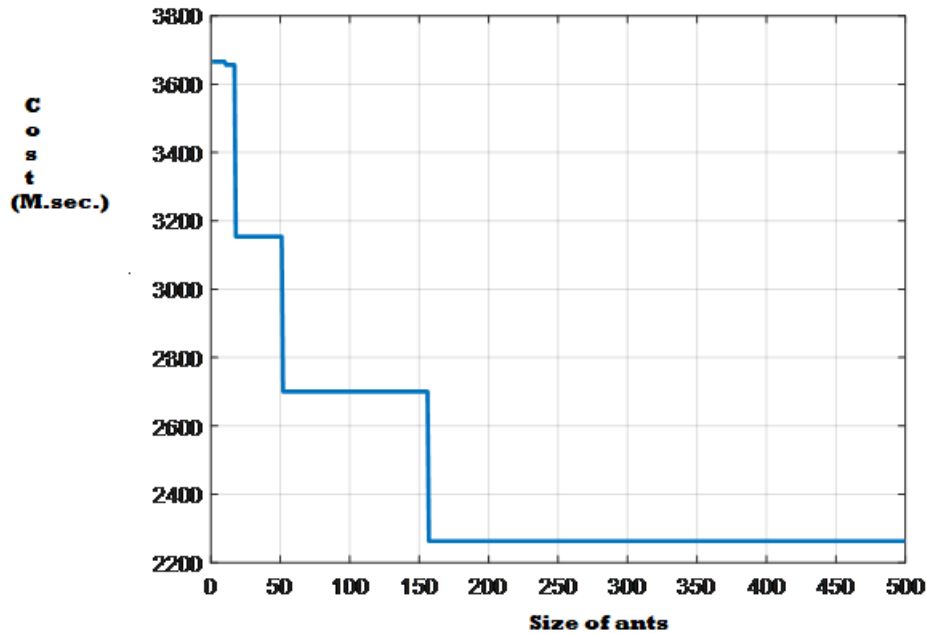
Figure-1 (Size of ants VS Operating Cost range(msec.))

Table 1 : Size of ants with Operating Cost

| Size of ants          | Operating Cost range(msec.) |
|-----------------------|-----------------------------|
| [5,6,7,8,19,20]       | 2627.1660172488             |
| [51,56,67,74,83,92]   | 2627.1660172488             |
| [151,164,173,182,191] | 2627.1660172488             |
| [209,218,227,236,245] | 2500.0000                   |
| [254,263,272,281,290] | 2500.0000                   |
| [308,317,326,335,344] | 2500.0000                   |
| [407,416,425,434,443] | 2500.0000                   |

The ants in this situation behave as simulation agent and focus to obtain the optimal solutions by traversing the parameter space prioritizing all possible solutions. The pheromones in such situation direct towards the resources to explore the desired paths. Accordingly the simulated ants monitor their positions as well as measure the solutions towards obtaining the better solutions. As shown in figure-1, the consistency is maintained on the operating cost of ants after certain range. For example, while the size of ants are within the range [5,6,7,8,19,20] or

[51,56,67,74,83,92] or [151,164,173,182,191], the consistency is maintained in the operating costs. Similarly in the higher range of ants i.e. [209,218,227,236,245] or [308,317,326,335,344] or [308,317,326,335,344] or [407,416,425,434,443], also the consistency is maintained in the operating costs.



**Figure-2 ( Size of ants VS Operating Cost)**

**Table 2 : Size of ants with Operating Cost**

| Size of ants          | Operating Cost   |
|-----------------------|------------------|
| [19,29,38,47]         | 3127.1660172488  |
| [56,74,119,128,137]   | 2729.56284548253 |
| [155,164,173,182,191] | 2267.6469234659  |
| [254,263,272,281,290] | 2267.6469234659  |
| [317,326,335,344,353] | 2267.6469234659  |
| [416,425,434,443,452] | 2267.6469234659  |
| [461,470,485,494]     | 2267.6469234659  |

Basically the ants are the simple computational agents to search for good solutions linked to specific problems and also can be used to obtain the shortest path somehow stochastically and update the pheromone levels at each stage. Each ant will definitely focus its current position along with the corresponding pheromone level and will move within the states to obtain the intermediate solutions. Generally the cost minimization is the basic optimization criteria towards constraining the ants. But somehow the ants can obtain a symmetric balance while focusing the criteria as shown in figure-2. The system based on these criteria can be more specific while the ants change the level of pheromones applying the local pheromone updates at each level.

## 5. DISCUSSION AND FUTURE DIRECTION

During accumulation of large scaled data and focusing to business analytics, it is observed that positioning specific resources to gain the performance is somehow due to the ability of these

resources towards optimality. Moreover, focusing on large scaled data with business analytics, the strategies should be maintained to find the optimal paths of sourcing and to generate analytics solutions to mobilize the linked resources towards operational objectives. Also some specific aspects linked to large scaled data along with business analytics approaches with the help of database triggers can be beneficial towards general performance. The large scaled data and big data analytics are partially connected and somehow share common operations. Also there can be provision to manage large scaled data identifying the large scaled data and the relationship within the business intelligence and analytical data in such situation will be optional.

## 6. CONCLUSION

Somehow in this work it has been focused on analysis of large scaled data with implementation mechanism and its linkages with business intelligence. It is observed that the analysis of large scaled data is viable towards enhancing the business intelligence along with the system implementation. In such situation, there should be provision to model the relativity of intelligence linked to business intelligence towards common implementation in large scale data. Again it will be possible to focus on similar intelligent mechanisms to focus on specified data with acceptability towards business intelligence.

## REFERENCES

- [1] H. Chen, R.H. Chiang, V.C. Storey, *Business intelligence and analytics: from big data to big impact*, *Mis Q.* 36 (4) (2012) 1165–1188.
- [2] A. Ashrafi, A.Z. Ravasan, P. Trkman, S. Afshari, *The role of business analytics capabilities in bolstering firms' agility and performance*, *Int. J. Inf. Manage.* 47 (2019) 1–15.
- [3] R.H. Chiang, V. Grover, T.-P. Liang, D. Zhang, *Strategic Value of Big Data and Business Analytics*, Taylor & Francis, 2018.
- [4] I.O. Pappas, P. Mikalef, M.N. Giannakos, J. Krogstie, G. Lekakos, *Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies*, *Inf. Syst. E-business Manag.* (2018).
- [5] G. Hindle, M. Kunc, M. Mortensen, A. Oztekin, R. Vidgen, *Business Analytics: Defining the Field and Identifying a Research Agenda*, Elsevier, 2019.
- [6] P. Mikalef, I.O. Pappas, J. Krogstie, M. Giannakos, *Big data analytics capabilities: a systematic literature review and research agenda*, *Inf. Syst. E-business Manag.* 16 (2018) 1–32.
- [7] J.J. Seddon, W.L. Currie, *A model for unpacking big data analytics in high-frequency trading*, *J. Bus. Res.* 70 (2017) 300–307.
- [8] K. Conboy, P. Mikalef, D. Dennehy, J. Krogstie, *Using business analytics to enhance dynamic capabilities in operations research: a case analysis and research agenda*, *Eur. J. Oper. Res.* (2019).
- [9] M. Janssen, H. van der Voort, A. Wahyudi, *Factors influencing big data decisionmaking quality*, *J. Bus. Res.* 70 (2017) 338–345.
- [10] F.P.S. Surbakti, W. Wang, M. Indulska, S. Sadiq, *Factors influencing effective use of big data: a research framework*, *Inf. Manag.* (2019).
- [11] R. Dubey, A. Gunasekaran, S.J. Childe, C. Blome, T. Papadopoulos, *Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture*, *Br. J. Manag.* 30 (2) (2019) 341–361.