ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Cost Minimization and Management of Database Files Using Fuzzy Tokens for Big Data Processing in Geo-Distributed Data Centers

Geetharani M, Poonkodi R, Sathishkumar N

Assistant Professor, Department of Computer Science and Engineering, Sri Eshwar College of Engineering, India

Assistant Professor, Department of Computer Science and Engineering, Sri Eshwar College of Engineering, India

Senior Software Consultant, IBM, India

ABSTRACT: Data explosion in current year's leads in the direction of increasing requirements designed for big data processing in current data centers with the purpose are frequently spread on different geographic regions. The unstable development of demands on big data processing requires an important burden on computation, storage, and communication in information centers, which therefore acquire significant prepared overheads in the direction of data center providers. Consequently, cost minimization has developed into a developing issue designed for the approaching big data era. This big data provider is varied from traditional cloud services, since one of the most important characteristic of big data services is the fixed coupling among data and computation as computation tasks is able to be performed simply when the related data is presented. Consequently, four factors, i.e., task assignment, database management, data placement and data movement, extremely manipulate the operational costs of information centers. In this paper, develops and introduces a new data processing schema which performed based on the procedure of fuzzy tokens which divides and splits the database files into many small number of fuzzy tokens in the direction of handle database files in the big database .Each and every one of the fuzzy tokens is denoted as the information of database files in the big database with three levels such as low, medium and high. The tokens are represented depending on the keyword given by used in the data placement, task assignment; data center resizing and routing in the direction of reduce the overall computation cost in huge-scale geo- distributed data centers designed for big data applications. In the direction of describe the task completion time by means of the consideration of together data transmission and computation, here introduces and develops a new two-dimensional Markov chain which derives the average task completion time in closed-form. Additionally, problem is represented and illustrated based on the Mixed-Integer Non-Linear Programming (MINLP). The high efficiency is achieved by using fuzzy tokens in the MINLP and validated by means of extensive simulation based studies. The proposed schema fuzzy tokens in the MINLP not only maintain the task completion time, it moreover manages files in the big database based on the fuzzy tokens in the database.

KEYWORDS: big data processing; Data explosion; cost minimization; fuzzy tokens; Mixed-Integer Non-Linear Programming (MINLP) ; big data applications .

I. INTRODUCTION

Information explosion in current year's leads in the direction of an increasing required designed for big information processing in current information centers with the purpose are frequently distributed on diverse geographic regions, e.g., Google's 13 information centers over 8 countries in 4 continents [1]. Big information examination has shown its huge possible in finding precious insights of information in the direction of develop decision making, reduce risk and increases new products with their services. Alternatively, big information has previously transformed into big price appropriate in the direction of its high require on computation and communication resources [2].

Consequently, it is very important in the direction of learn the cost reduction problem intended for big data processing in geo-distributed information centers. Numerous works have been done in the recent works in the direction of reducing computation time and communication overhead cost of data centers. Data Center Resizing (DCR) have

DOI: 10.15680/IJIRCCE.2017. 0501169

been also developed and introduced in the recent work in the direction of reducing the computation cost by means of changing the number of activated servers by means of task placement [3].

Depending on the working procedure of DCR, some of the works have been done in the recent work implemented to the geographical allocation nature of information centers and electrical energy price heterogeneity in the direction of lesser the electricity cost [4]–[6]. Big information service frameworks, e.g., [7], consists of a distributed file system under, which distributes information chunks and their replicas across the information centers designed for fine-grained load-balancing and high corresponding information access performance. In the direction of decreasing the communication cost, a few number of the works have been done in the recent work which increases the data management problem by means of insertion jobs on the servers where the input information reside in the direction of avoid remote information loading [7-8].

While the above mentioned methods founded and implemented by means of some positive results, they are far from achieving the cost-proficient big data processing since of the subsequent disadvantages. Initially, information locality might consequence in a waste of resources. Following, the links in networks differ depending on the data transmission size and costs related in the direction of their distinctive features [9], e.g., the distances and substantial optical fiber services among data centers.

On the other hand, the existing routing strategy between information centers fails in the direction of make use of the link variety of data center networks. It is necessary with the purpose of positive information should be downloaded from a remote server. In this case, routing strategy substances on top of the transmission cost. At the same time as designated by means of Jin et al. [10], the transmission cost, e.g., energy, almost proportional in the direction of the number of network link second-hand. The more links second-hand, the higher cost determination be acquired.

Consequently, it is important in the direction of lower the number of links second-hand at the same time as fulfilling each and every one the transmission requirements. Since the Quality-of-Service (QoS) of big data tasks have not been focused by many works done in the recent work. Related in the direction of traditional cloud services, big data applications moreover demonstrate Service-Level-Agreement (SLA) among a service provider and the requesters. In the direction of examine SLA, a definite level of QoS, frequently in terms of task completion time, shall be guaranteed. Moreover, the transmission rate is different important factors because big information tasks are information-centric and the calculation task mightn't continue in anticipation of the related information are obtainable. In the recent works [3], on general cloud computing tasks majorly depends on the computation capability restriction, at the same time as disregarding the constraints of data transmission range. The major contribution of the work done in this paper work is described as follows:

To the best information, are the first in the direction of regard as the cost minimization problem of big information processing by means of joint consideration of information placement, task assignment and information routing. In the direction of explain the cost reduction and transmission in big information processing procedure, introduce a two dimensional Markov chain model and expected task completion time is computed from the fuzzy tokens are created in the direction of preserve the database information of the files in the big database. Depending on the computation cost of the big database, the cost minimization problem is formulated as the mixed integer nonlinear programming (MINLP) that needs to answer the following question: 1) How in the direction of place these data chunks in the servers, 2) how in the direction of hand out tasks onto servers not including abusing the resource constraints, and 3) how in the direction of resize data centers in the direction of attain the process cost minimization objective. All the numerical studies, demonstrated that the high efficiency of proposed fuzzy tokens based joint-optimization based algorithm when compared to existing methods.

II. LITERATURE SURVEY

Huge-scale information has been positive each and every one over the world given that services in the direction of hundreds of thousands of users. Related in the direction of [11], an information center might include of huge number of cloud servers and use megawatts of power. Consequently, decreasing the electricity cost of the geo data center has been focused in the literature work related to both academia and industry [11]–[13]. In the middle of the mechanisms with the purpose of have been introduced and developed consequently future designed for information or geo-data center energy management, the techniques with the purpose of pull towards you a lot of concentration are task position and DCR.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Fan et al [12] proposed and developed a new power provisioning strategies on how a great deal computing equipment be able to be securely and capably hosted inside a specified power budget. Make use of modeling framework in the direction of approximation the possible of power management schemes in the direction of decreasing peak power and energy usage. In conclusion argue with the purpose of systems require in the direction of be power proficient across the activity range, and mightn't at peak performance levels.

In recent times, Gao et al [14] introduces and develops a new optimal workload control and balancing by means of attractive description of latency, energy consumption and electricity prices. In the recent work develops a new Flow Optimization based framework for request-Routing and Traffic Engineering (FORTE). It permits an operator in the direction of find the way the three-way tradeoff among admission latency, carbon footprint, electricity costs and in the direction of decide an optimal datacenter improve plan in response in the direction of enhances in traffic load.

Liu et al [15] majorly focus on reducing the usage of electricity cost and environmental impact by means of a using holistic approach of workload balancing with the purpose of incorporates renewable supply, dynamic pricing, and cooling supply. The results concludes with the purpose of the approach be able decreasing together the recurring power costs and the make use of non-renewable energy by means of 60% compared in the direction of existing techniques, at the same time as still get-together the Service Level Agreements(SLAs).

In order to handle the problems of successfully handling and managing big data, several have been done in the recent work in the direction of usage of geo-data storage and computation procedure. The major important issues of big data management are dependable and successful data placement. In the direction of handle this goal, Sathiamoorthy et al [16] introduces a new family of erasure codes with the purpose are proficiently repairable and present high consistency when compared to Reed-Solomon codes. The major problem of these methods is with the purpose it needs 14% storage usage increased when compared to Reed-Solomon codes and computation overhead also increased rather than the other conventional methods.

Hu et al [17] introduces and develops new methods permitting linked open information in the direction of take benefit of presented large-scale information stores in the direction of get together the requirements on distributed and parallel information processing. By the way of sufficiently handling Big Data, are witnessing a measured shift in the direction of the all the time more accepted Linked Open Data (LOD) paradigm. In the meantime, the sheer data size predicted by means of Big Data refuses definite computationally costly semantic technologies, representation the latter a large amount less well-organized than their performance especially for small data sets.

Cohen et al [18] introduces and develops new design philosophy given that a novel magnetic, agile and deep information analytics designed for one of the world's major advertising networks next to Fox Audience Network, by means of the Greenplum parallel information system. Draw attention to the promising apply of Magnetic, Agile, Deep (MAD) information examination as a fundamental different approach from conventional Enterprise Data Warehouses and Business Intelligence. Lastly, consider database system features with the purpose of permit agile design and flexible algorithm improvement by means of together SQL and Map Reduce interfaces over a diversity of storage mechanisms.

Kaushik et al [19] introduces and develops a novel, data-centric algorithm in the direction of reducing usage of energy costs and by means of the assurance of thermal-reliability of the servers. Chen et al [20] introduces and focus on the problem of jointly scheduling with three major phases such as map, shuffle and reduce, of the Map Reduce process with high computation complexity. Appropriate to its widespread deployment, there have been several recent papers outlining practical schemes in the direction of enhance the performance of Map Reduce systems.

Agarwal et al [21] introduces and develops a an automated data placement methods designed for Volley for geo-distributed cloud services by means of the concerning of WAN bandwidth cost, information center capability limits, data inter-dependencies, etc. Cidon et al [22] discover MinCopysets, an information replication placement scheme with the purpose of decouples information distribution and replication in the direction of enhances the data durability characteristics in distributed information centers. Contribution is in the direction of decouple the mechanisms second-hand designed for load balancing from data replication: make use of randomized node selection designed for load balancing however derandomize node selection designed for data replication, present important development in data stability.

DOI: 10.15680/IJIRCCE.2017. 0501169

III. PROPOSED METHODOLOGY

To the best information, are the first in the direction of regard as the cost minimization problem of big information processing by means of joint consideration of information placement, task assignment and information routing. Depending on the computation cost of the big database, the cost minimization problem is formulated as the mixed integer nonlinear programming (MINLP) that needs to answer the following question: 1) How in the direction of place these data chunks in the servers, 2) how in the direction of hand out tasks onto servers not including abusing the resource constraints, and 3) how in the direction of resize data centers in the direction of attain the process cost minimization objective. In this paper, develops and introduces a new data processing schema which performed based on the procedure of fuzzy tokens which divides and splits the database files into many small number of fuzzy tokens in the direction of handle database files in the big database .Each and every one of the fuzzy tokens is denoted as the information of database files in the big database with three levels such as low, medium and high. In the direction of describe the task completion time by means of the consideration of together data transmission and computation, here introduces and develops a new two-dimensional Markov chain which derives the average task completion time in closed-form.

Let us consider a geo-distributed data center topology is illustrated in figure 1 , here each and every one servers of the similar Data Center (DC) are linked in the direction of their local switch, at the same time as information or geo-data centers are linked via the use of switches. Let us consider there are $i \in$ of information or geo-data centers which consists of number of data servers with the purpose are linked in the direction of a switch $m_i \in$ with a transmission cost in the local geo-data center as . Generally, the transmission cost designed for inter-information data center traffic is larger than , i.e., $C_R >$. Not including loss of generalization, each and every one server in the network has the similar computation resource and storage capability, together of which are regularized in the direction of single unit. Make use of J in the direction of represents the set of each and every one severs, the entire goe-data centers have been modeled by the use of directed graph G = (N,E). The vertex set $N = M \cup$ consists of the set M of each and every one switch, the set J of each and every one server, and E is the directional edge set. Each and every one server are linked to, and only in the direction of their local switch by means of intra-data center links at the same time as the switches are linked via inter-data center links computed by means of their substantial association. The weight of each link w(u;v), denoting the related communication cost, be able to be described as follows,

$$w^{(u,v)} = \begin{cases} C_R & if\ v,u \in M \\ C_L & otherwise \end{cases} \quad (1)$$

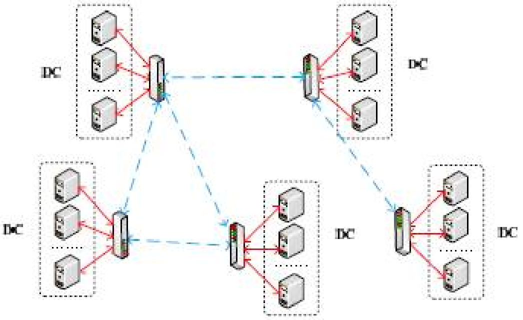

Figure 1. Data center topology

In this paper, develops and introduces a new data processing schema which performed based on the procedure of fuzzy tokens which divides and splits the database files into many small number of fuzzy tokens in the direction of handle database files in the big database .Each and every one of the fuzzy tokens is denoted as the information of database files in the big database with three levels such as low, medium and high. The tokens are represented depending on the keyword given by used in the data placement, task assignment; data center resizing and routing in the direction of reduce the overall computation cost in huge-scale geo- distributed data centers designed for big data applications.

Fuzzy tokens permit you in the direction of equivalent input records by means of clean, standardized records in an indication table. The identical procedure is flexible in the direction of errors with the purpose are current in the input records. Fuzzy Grouping permits you in the direction of discover groups of records in big dataset samples where each record in the group potentially related to the similar real-world entity. The grouping is flexible in the direction of frequently experimental errors in real information, since records in each group might not be indistinguishable in the direction of each other however are much related in the direction of each other. Fuzzy tokens are able in the direction of help discover information in geo-data tables when your information have becomes a deficient string key. For instance, if you want in the direction of know the data designed for user given query in the big database, you be able to make use of Fuzzy tokens in the direction of discover the information, even if your input mightn't match accurately what is stored in your reference table. In the direction of construct the simplest Fuzzy token package:

1. Open Business Intelligence Development Studio.
2. Create a new Integration Services Project, add a new package, click the Data Flow tab, and then accept the add a data flow item option.
3. From the Control Flow Items section in the Toolbox, drag a Data Flow Task onto the control Flow surface. Double-click the new Data Flow Task or select the Data Flow tab.
4. On the Data Flow surface, drag the OLE DB Source adapter from the Data Flow Sources section of the Toolbox. Drag a Fuzzy token Transformation from the Data Flow Transformations section of the Toolbox and an OLE DB Destination adapter from the Data Flow Destinations section. Also create a path between Fuzzy token and the Destination by selecting Fuzzy token and dragging the green arrow to the Destination.
5. Double-click the OLE DB Source transform and configure it to point at your new data by selecting a connection and the input table that contains reference data those incoming records will be matched to each other.
6. Double-click Fuzzy token to open the custom user interface (UI). From the Reference table name drop-down menu, select the connection and table to which you want the transform, to your already warehoused reference data, to point.
7. Select the check boxes for all items in Available Lookup Columns, and then click OK.
8. Point the OLE DB Destination to a connection for which you can write a new table, and then click new. Accept the default creation statement, and you are now ready to run Fuzzy token.
9. To run the package you just created, right-click its name in the Solution Explorer window, and then select Execute.

Let us consider a big geo-data tasks focus on data stored in a dispersed file system with the purpose is construct on geo-distributed data centers. The information stored in the geo-data centers are spitted into K set of chunks. Each chunk $k \in$ has the size of which is normalized toward the server storage capability. P-way replica [19] is used for this purpose. With the purpose of each chunk, there are exactly P copies stored in the distributed file system designed for resiliency and fault-tolerance have been modeled as via the use of Poisson process [9].

For the most part, permit be current the average task arrival rate demand chunk k. Represents the average arrival rate of task designed for chunk k on server j as $\lambda_{jk} (\lambda_{jk} \leq$ When a task is distributed in the direction of a server where its requested information chunk shouldn't be located in, it requirements in the direction of wait designed for the data chunk in the direction of be transferred. Each task must be responded in time D. Describe a binary variable y_{jk} in the direction of represents whether chunk k is located on server j as follows,

$$y_{jk} = \begin{cases} 1 \text{ if chunk } k \text{ is placed on server } j \\ \quad 0 \text{ otherwise} \end{cases} \qquad (2)$$

In the distributed file system, preserve P copies designed for each chunk $k \in$, which direct in the direction of the following constraint:

$$\sum_{j \in J} y_{jk} = P \; \forall k \in K \qquad (3)$$

Moreover, the information stored in each server $j \in$ mightn't go beyond its storage capability

$$\sum_{j \in J} y_{jk} \cdot \Phi_{jk} \leq 1, \forall j \in J \qquad (4)$$

As designed for task distribution, the sum rates of task allocated in the direction of each server must be equal in the direction of the overall rate,

$$\lambda_{jk} = \sum_{j \in J} y_{jk} \; \forall k \in K \qquad (5)$$

Note with the purpose of when a data chunk k is needed by means of a server j, it might cause internal and external information transmissions. Each and every one the nodes N in graph G, consists of the servers and switches are able to be divided into three major categories:

Source nodes u$(u \;\in\; J)$: They are the servers by means of data chunk k stored in it. In this case, the total channel flows in the direction of destination server j designed for chunk k from each and every one source nodes shall get together the total chunk constraint per time unit as $y_{jk} \cdot \Phi_{jk}$.

Relay nodes (m_i $\in$) receive information flows beginning source nodes and advance them related in the direction of the routing strategy.

Destination node j(j): When the necessary chunk is not stored in the destination node, i.e., y_{jk} = 0, it should receive the information flows of chunk k next to a rate $\lambda_{jk} \cdot$.

Let and is denoted as the processing rate and loading rate designed for information chunk k on server j, correspondingly. The working procedure is formulated based on the procedure of two-dimensional Markov chain, where each state (p; q) denoted as p pending tasks and q presented data chunks. Let is represented as the amount of computation resource with the purpose of chunk k occupies. The processing rate of tasks is relative in the direction of its computation resource usage,

$$\mu_{jk} = \alpha_j \cdot \theta_{jk} \; \forall j \in J, k \in K \qquad (6)$$

where is a constant relying on the speed of server j. Moreover, the total computation resource allocated in the direction of each and every one chunks on each server j shall not exceed its total computation resource.

IV. RESULTS AND DISCUSSION

In this section present the performance results of joint-optimization algorithm (“Joint”) using the MILP formulation. The performance accuracy of the proposed methods is also compared to “Non-joint” model, which initially finds a minimum number of servers in the direction of be activated and the traffic routing scheme via the use of network flow model. In the experiments, let us consider $|J|$ = data centers, each of which is by means of the same number of servers. The intra- and inter-data center link communication cost is initially set to C_L = and C_R = correspondingly. The cost on each activated server j is set in the direction of one. The data size, storage requirement, and task arrival rate are all randomly generated.

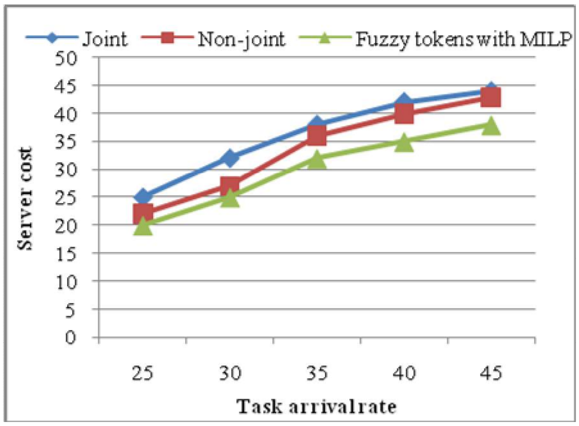

Fig 2 . Server cost on the effect of task arrival rate

Then, investigate how the task arrival rate affects the cost via varying its value from 29:2 to 43:8. The evaluation results are shown in Fig 2 first notice that the total cost shows as an increasing function of the task arrival rates in both algorithms. This is because, to process more requests with the guaranteed QoS, more computation resources are needed. This leads to an increasing number of activated servers and hence higher server cost, as shown in Fig 2. An interesting fact noticed from Fig 2 is that "Joint" algorithm requires sometimes higher server cost than "Non-joint". This is because the first phase of the "Non-joint" algorithm greedily tries to lower the server cost. However, "Joint" algorithm balances the tradeoff between server cost and communication cost such that it incurs much lower communication cost and thus better results on the overall cost, compared to the "Non-joint" algorithm, proposed fuzzy tokens with MILP as shown below in Fig 3 and Fig 4 respectively.

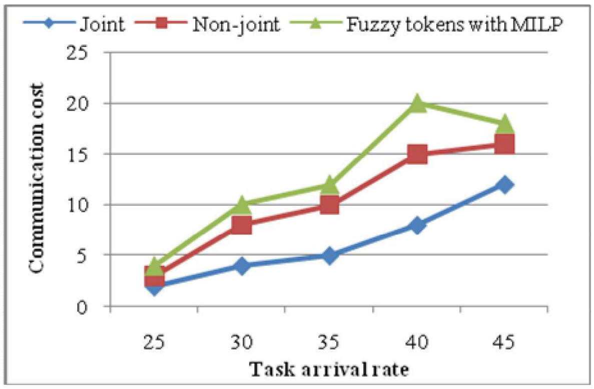

Fig 3. Communication cost on the effect of task arrival rate

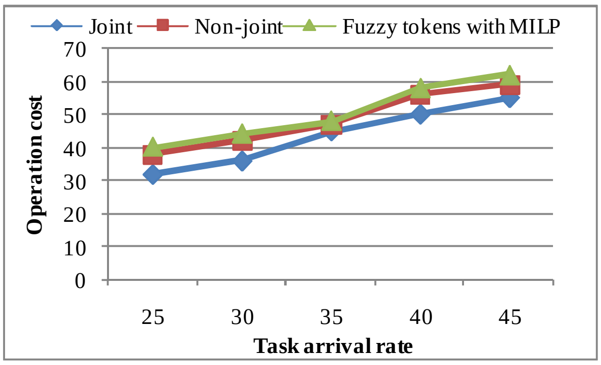

Fig 4. Overall cost on the effect of task arrival rate

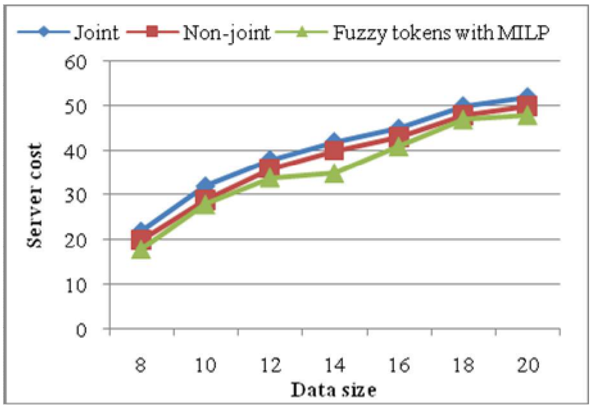

Fig 5.Server cost on the effect of data size

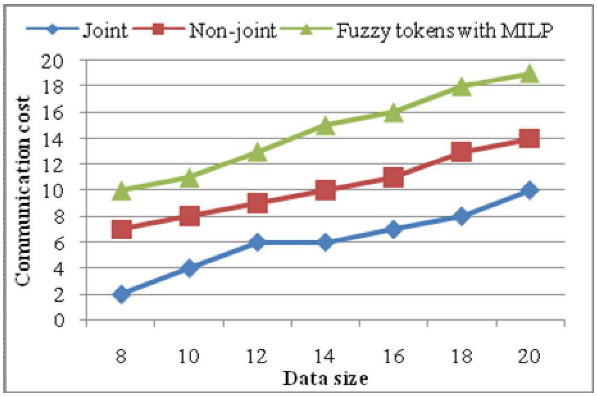

Fig 6. Communication cost on the effect of data size

Fig 5 illustrates the cost as a function of the total data chunk size from 8:4 to 19. Larger chunk size leads to activating more servers with increased server cost as shown in Fig 4. At the same time, more resulting traffic over the links creates higher communication cost as shown in Fig 5. Finally, Fig 6 illustrates the overall cost as an increasing function of the total data size and shows that proposal outperforms proposed fuzzy tokens with MILP under all settings.

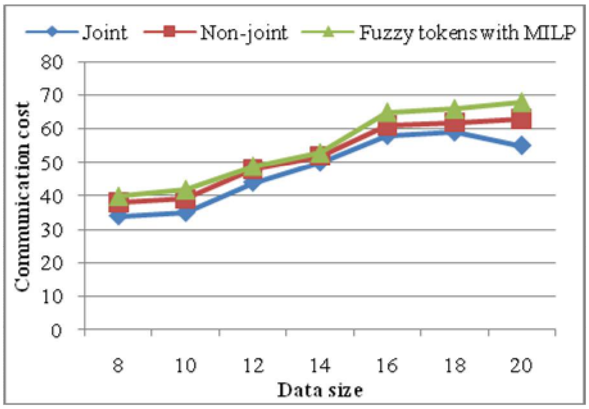

Fig 7. Operation Cost on the effect of data size

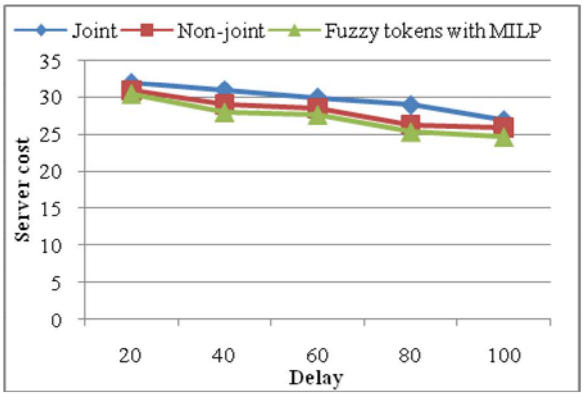

Fig 8. Server cost on the effect of expected task completion delay

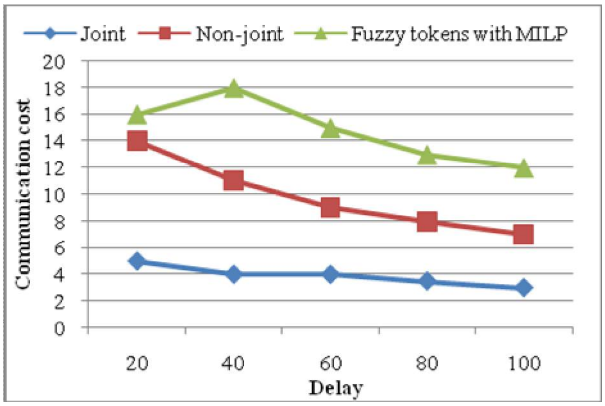

Fig 9. Communication cost on the effect of expected task completion delay

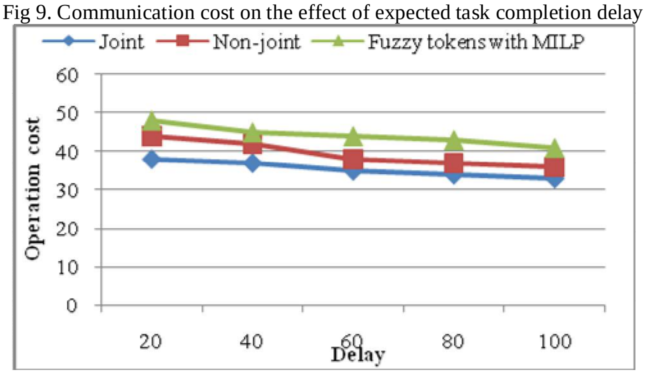

Fig 10 .Operation cost on the effect of expected task completion delay

Fig 8 the results when the expected maximum response time D increases from 20 to 100. From Fig 8 can see that the server cost is a non-increasing function of D. The reason is that when the delay requirement is very small, more servers will be activated to guarantee the QoS. Therefore, the server costs of both algorithms decrease as the delay constraint increases. A looser QoS requirement also helps find cost efficient routing strategies as illustrated in Fig 9. Moreover, the advantage of proposed fuzzy tokens with MILP can be always observed in Fig 10.

V. CONCLUSION AND FUTURE WORK

In this paper, develops and introduces a new data processing schema which performed based on the procedure of fuzzy tokens which divides and splits the database files into many small number of fuzzy tokens in the direction of handle database files in the big database .Each and every one of the fuzzy tokens is denoted as the information of database files in the big database with three levels such as low, medium and high. The tokens are represented depending on the keyword given by used in the data placement, task assignment; data center resizing and routing in the direction of reduce the overall computation cost in huge-scale geo- distributed data centers designed for big data applications. Subsequently distinguish the information processing work by the use of two-dimensional Markov chain and at the same time as the expected task completion time is computed from joint optimization, it is formulated as an MINLP problem. From the experimentation results it concluded that the proposed joint-optimization with fuzzy tokens produces better results than the three -step separate optimization. Several interesting phenomena are also observed from the experimental results. In future, intend in the direction of regard as multiple hierarchies designed for performing fuzzy tokens and duplicate elimination difficulty of distinguished several tuples in the big database is moreover discovered, which explain the related real world entity, are significant information cleaning problem. Future algorithm designed for removing duplicates in dimensional tables in a data warehouse, which are frequently connected by means of hierarchies.

REFERENCES

1. "Data Center Locations," http://www. google. com /about/data centers/inside /locations /index .html.
2. R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, and X. Zhu, "No "Power" Struggles: Coordinated Multi-level Power Management for the Data Center," Proceedings of the 13th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS). ACM, 2008, pp. 48–59.
3. L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing Electricity Cost: Optimization of Distributed Internet Data Centers in a Multi- Electricity-Market Environment," Proceedings of the 29th International Conference on Computer Communications (INFOCOM). 2010, pp. 1–9.
4. Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew, "Greening Geographical Load Balancing," Proceedings of International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS). ACM, 2011, pp. 233–244.
5. R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam, "Optimal Power Cost Management Using Stored Energy in Data Centers," in Proceedings of International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS). 2011, pp. 221–232.
6. B. L. Hong Xu, Chen Feng, "Temperature Aware Workload Management in Geo-distributed Datacenters," in Proceedings of International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS), 2013, pp. 33–36.
7. J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," Communications of the ACM, vol. 51, no. 1, pp. 107–113, 2008.
8. S. A. Yazd, S. Venkatesan, and N. Mittal, "Boosting energy efficiency with mirrored data block replication policy and energy scheduler," SIGOPS Oper. Syst. Rev., vol. 47, no. 2, pp. 33–40, 2013.
9. Marshall and C. Roadknight, "Linking cache performance to user behaviour," Computer Networks and ISDN Systems, vol. 30, no. 223, pp. 2123 – 2130, 1998.
10. H. Jin, T. Cheocherngngarn, D. Levy, A. Smith, D. Pan, J. Liu, and N. Pissinou, "Joint Host-Network Optimization for Energy- Efficient Data Center Networking," in Proceedings of the 27th International Symposium on Parallel Distributed Processing (IPDPS), 2013, pp. 623–634.
11. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the Electric Bill for Internet-scale Systems," Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM), 2009, pp. 123–134.
12. X. Fan, W.-D. Weber, and L. A. Barroso, "Power Provisioning for A Warehouse-sized Computer," Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA). ACM, 2007, pp. 13–23.
13. S. Govindan, A. Sivasubramaniam, and B. Urgaonkar, "Benefits and Limitations of Tapping Into Stored Energy for Datacenters," Proceedings of the 38th Annual International Symposium on Computer Architecture (ISCA). ACM, 2011, pp. 341–352.
14. P. X. Gao, A. R. Curtis, B. Wong, and S. Keshav, "It's Not Easy Being Green," Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM). ACM, 2012, pp. 211–222.
15. Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser, "Renewable and Cooling Aware Workload Management for Sustainable Data Centers," Proceedings of International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS), 2012, pp. 175–186.
16. M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, "Xoring elephants: novel erasure codes for big data," Proceedings of the 39th international conference on Very Large Data Bases, ser. PVLDB'13. Endowment, 2013, pp. 325–336.
17. B. Hu, N. Carvalho, L. Laera, and T. Matsutsuka, "Towards big linked data: a large-scale, distributed semantic data storage," Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services, ser. IIWAS '12, 2012, pp. 167–176.
18. J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton, "Mad skills: new analysis practices for big data," Proc. VLDB Endow., vol. 2, no. 2, pp. 1481–1492, 2009.
19. R. Kaushik and K. Nahrstedt, "T*: A data-centric cooling energy costs reduction approach for Big Data analytics cloud," International Conference for High Performance Computing, Networking, Storage and Analysis (SC), 2012, pp. 1–11.
20. F. Chen, M. Kodialam, and T. V. Lakshman, "Joint scheduling of processing and shuffle phases in mapreduce systems," Proceedings of the 29th International Conference on Computer Communications (INFOCOM), 2012, pp. 1143–1151.
21. S. Agarwal, J. Dunagan, N. Jain, S. Saroiu, A. Wolman, and H. Bhogan, "Volley: Automated Data Placement for Geo-Distributed Cloud Services," The 7th USENIX Symposium on Networked Systems Design and Implementation (NSDI), 2010, pp. 17–32.
22. S.Reha and M.Geetharani "Swarm Intelligence Based Fuzzy with Personalized Ontology Model for Web Information Gathering"Vol 3,No 1,pp.40-45,2014.
23. Geetharani M, Poonkodi R and Anuradha K," A Clustering Based User-Centered (CBUC) Approach for Integrating Social Data into Groups of Interest",Vol 4,pp.639-646,2016.
24. Poonkodi R, Geetharani M and Gunasekaran R," Automatic Lobar Segmentation Algorithm for Pulmonary Lobes from Chest Ct Scans Based On Fissures and Blood Vessels",Vol 3,pp.2980-2987,2015.
25. M.Geetharani and S.Reha "Hybrid Rule Based Feature Subset Selection and Classification",Vol 3,No 2,pp.79-84,2014.