Journal of Mathematical Modelling and Applied Computing

Volume No. 12

Issue No. 1

January - April 2024

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

Journal of Mathematical Modelling and Applied Computing

Aims and Scope

The Journal of Mathematical Modelling and Applied Computing is an Indian research journal, which publishes top-level original and review papers, short communications and proceedings on Interdisciplinary Integrative Forum on Modelling, Simulation and Scientific Computing in Engineering, Physical, Chemical Biological, Medical, Environmental, Social, Economic and Other Systems using Applied Mathematics and Computational Sciences and Technology.

Journal of Mathematical Modelling and Applied Computing

Journal of Mathematical Modelling and Applied Computing

(Volume No. 12, Issue No. 1, January - April 2024)

Contents

Big Data: Hadoop A Tool for Solution

[1]Wasim Ahmad Ansari, [2]Saeed Anwar Jamal Ansari
[1](M.Tech Scholar,MDU)
imca.wasim@gmail.com
[2](M.Tech Scholar,MDU)
saeedlko@gmail.com

ABSTRACT

There are various definitions offered for "big data," it means data may be very big, too fast, or too hard for existing tools to process. Here, "very big" means that organizations extremely must deal with huge scale collections of data that come from different unstructured data, transaction histories, sensors, and elsewhere. Another use of big data is increasingly popular in medical term and known as medical big data in healthcare services and clinical research. To get the logic behind medical big data shows motive in hospital information technology and shows great significance for hospital information systems that are designing and enhancing services. Big data has four characteristics –Volume, Variety, Velocity and Value (the 4 Vs) – that make traditional systems incapable of processing these data using standalones. To get over the big data problem Apache Hadoop Map Reduce is a crucial software framework for developing applications that process huge amounts of data in parallel with large clusters of commodity hardware in a reliable, fault-tolerant manner. These are the Hadoop framework and MapReduce application program interface(API) with those API we can more easily develop our own MapReduce applications to run on a Hadoop framework that can enhance from a single node to thousands of machines. This paper investigates a practical case of a Hadoop-based medical big data processing system. We developed this system to intelligently process medical big data and uncover some features of hospital information system user behaviors. This paper throws user behaviors regarding various data produced by different hospital information systems for daily work. Additionally, with medical big data analytics, we can design our hospital information systems to be much more intelligent and easier to use by making personalized recommendations.

Keywords—Big Data . Hadoop . MapReduce .Intelligent . Hospital Information System.

INTRODUCTION

It is notable that Health information technologies (HITs) have been motivating healthcare services and clinic research to quickly improve, but the medical data that are generated still needs to be maintained some inherent traits and complexities that are difficult to show. Medical data sets are continuously becoming bigger, hence furnishing it increasingly hard for standalone systems to process medical data. Getting the logic behind medical big data(MBD) has great importance for designing hospital information systems (HISs) that can be used for recommending services and may help overcome some

exploring better functions of those systems. Latest developments in open-source software (OSS) – namely, the Apache Hadoop Foundation and related projects – provide a framework for large scale data warehouses, thus enabling fault-tolerant parallelized analysis of MBD using the MapReduce [5, 6] programing model. A large variety of organizations and researchers have used Hadoop for healthcare services and clinical research projects [7–9].

Latest, analysis of user behaviors has become a popular approach to understand users, and Hadoop performs well at analyzing user behaviors by mining large-scale data sets such as petabyte scale. It can be said the universal usage of HISs, such as electronic medical records (EMRs), computerized physician order entry (CPOE), picture archiving and communication system (PACS), and other clinical and administrative systems, yields ultra-large volumes of data that are closely related and connected with users' behaviors. Secondary uses of such MBD are becoming increasingly popular in healthcare services and clinical research [24–26].

Harnessing collective intelligence and its related algorithms has been demonstrated to be a trustworthy method for studying medical data. Kim M et al. (2014) used collective intelligence to reduce the potential risk of misleading online information and the accompanying safety issues [19]. Alor-Hernandez G et al. (2012) developed a content-based image retrieval system to use collective intelligence; their system supports the medical community in providing differential diagnoses related to diseases of the breast [13].

Therefore, in this paper, we show a practical case of Hadoop-based applications to intelligently process MBD and uncover some features of the behaviors of users of HIS sand will see how Hadoop.

We analyzed user behaviors based on the structured, semi structured and unstructured data that are produced by HISs that are used for daily work..Compared with single-node algorithms, our distributed algorithms show promise for facilitating efficient MBD processing in healthcare services research and clinical research. Moreover, with MBD analytics, we can design our HISs to be much more intelligent and easier by making personalized recommendations, which may lead to better acceptance of HISs.

WHAT IS BIG DATA?

There are many definitions offered for “big data,” one is that it means data that's so large, too fast, or too hard for existing tools to process. Here, “to large” means that organizations increasingly must deal with large scale collections of data that come from unstructured data system, transaction histories, sensors, and elsewhere. “Too fast” means that not only is data big, but it must be processed quickly — for

example, to perform fraud detection at a point of sale or determine which ad to show to a user on a webpage.“Too hard” is a catchall for data that doesn't fit neatly into an existing processing tool or that needs some kind of analysis that existing tools can't readily provide. A similar breakdown is being promulgated by Gartner (which is probably a sign that I'm oversimplifying things),citing the “three Vs” — volume, velocity, and variety (a catchall similar to “too hard”).

HADOOP: AN ARCHITECTURE:

Now we can see that Hadoop-based MBD processing system (HMBDPS) has three components that each covers a particular application, as shown in Fig. 1. The components are a Sqoop-based Extract-Transform-Load (ETL) module; a Hadoop cluster for data storage, management and parallel processing. After Combined with the applications that create a user behavior log and a recommender interface server, all of the components form a closed loop system. We can have look on the functions as follows. Let's have a look on Data ETL module. As the history of the development of healthcare information technologies, clinical data are typically stored in relational databases. We all know that because relational databases are conventional information technology (IT) architectures, they cannot face the challenges and fix the problems caused by big data. On contrary, Hadoop is a reliable platform for processing big data. Essentially, it is made of two parts. One part is the Hadoop distributed file system (HDFS), which is a distributed file system that provides low-latency and high-throughput access to big data storage, and the other part is the Hadoop MapReduce, which is a software framework for easily developing applications with powerful computing capacities to process big data stored in the HDFS. However, structured data stored in relational databases are not easily accessible for analysis in Hadoop. Current methods of transferring data are inefficient. In our system, we developed a Sqoop-based ETL (SETL) module, which is designed to offer effective and efficient MBD transferring services to optimize our usage of MBD, as shown in Fig. 2. Sqoop is a tool for efficiently transferring bulk data between Hadoop and structured datasets, such as a relational database management system (RDBMS). To transfer data, the SETL module first identifies a connector to the source data through Java Data Base Connectivity (JDBC), looks up the metadata of the source data, and then transforms Structured Query Language (SQL)-type data to Java-type Sqoop records as an input format for MapReduce jobs.

Finally, appropriate numbers of map tasks and reduce tasks are launched to write the records to the HDFS, as shown in Fig.

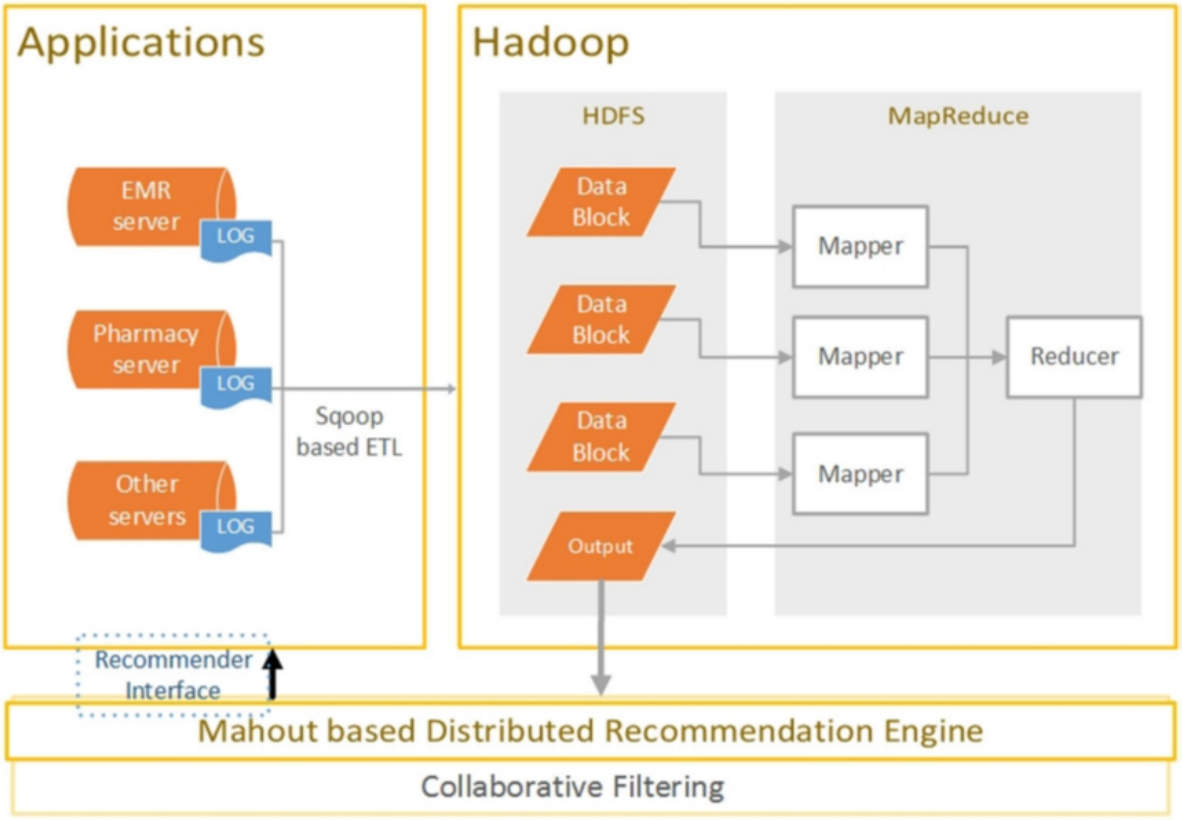

Applications
Hadoop
HDFS
MapReduce
EMR server
LOG
Pharmacy server
LOG
Other servers
LOG
Sqoop based ETL
Data Block
Data Block
Data Block
Output
Mapper
Mapper
Mapper
Reducer
Recommender Interface
Mahout based Distributed Recommendation Engine
Collaborative Filtering

Fig.1

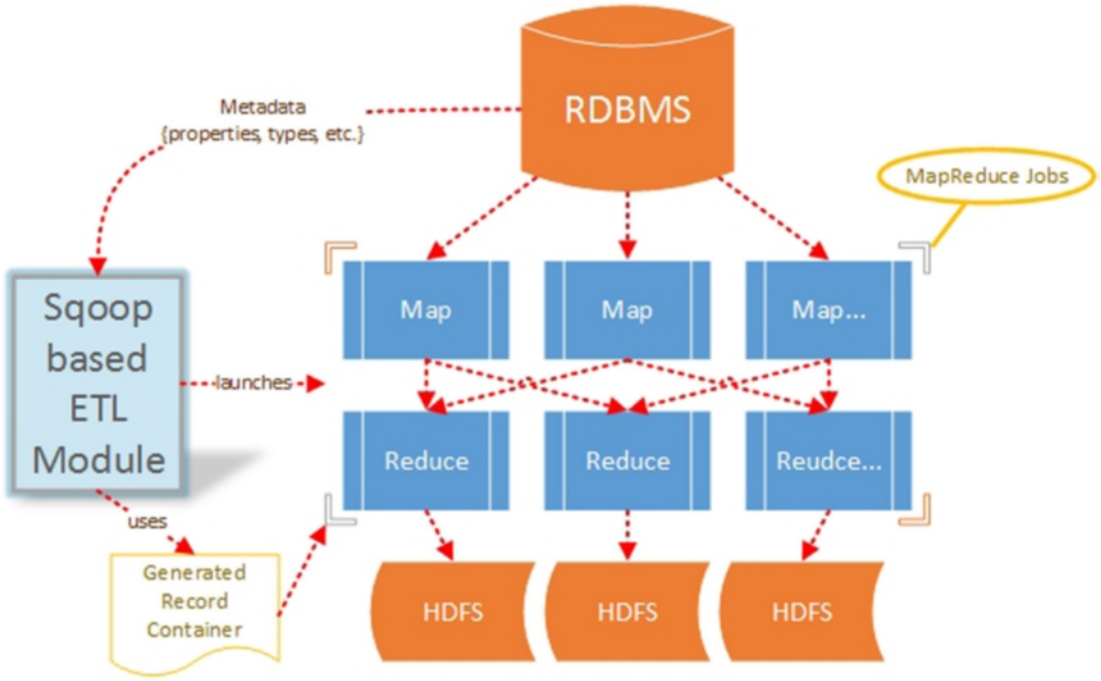

RDBMS
Metadata {properties, types, etc.}
MapReduce Jobs
Sqoop based ETL Module
launches
uses
Generated Record Container
Map
Map
Map...
Reduce
Reduce
Reudce...
HDFS
HDFS
HDFS

Fig.2

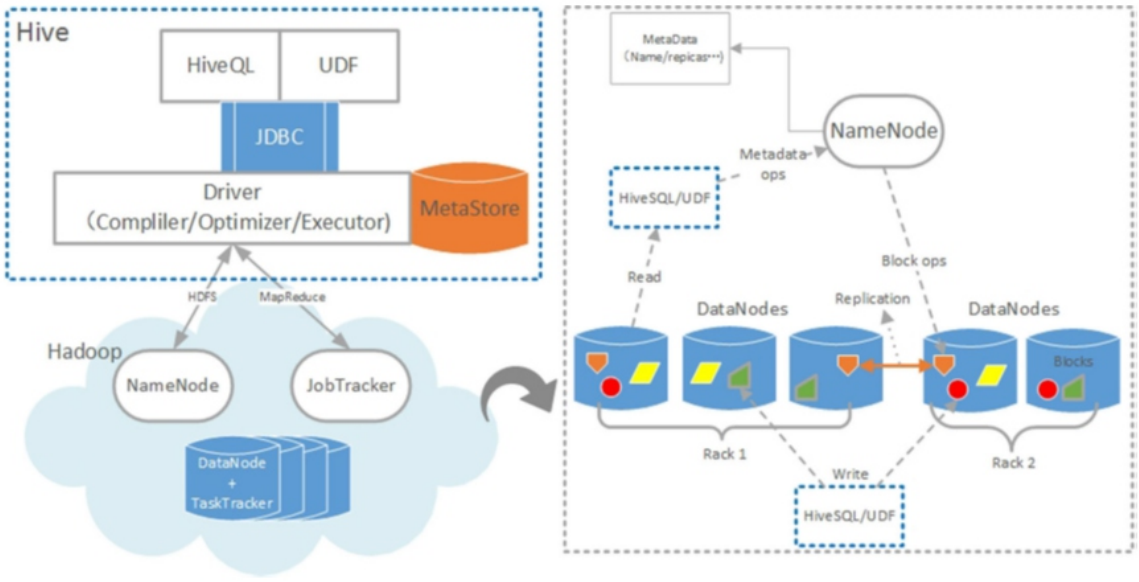


Fig.3

Core Component of engine.

CONCLUSIONS

This paper describes the design and development of a Hadoop-based MBD processing system that can be applied for secondary uses of MBD. This system solves problems of MBD collection, storage and analysis. Through this paper we have also seen how Hadoop work. Compared with non-distributed systems, Hadoop demonstrate that this system and its related distributed algorithms are capable of handling MBD more efficiently. Since Hadoop is a distributed system it work more efficient as it has master and slaves parts. This paper shows why big problem needs to be broken and solved individual and then their results combined to produce the required results which cannot be achieved by traditional system.

REFERENCES

1. Lin, C., Lin, I.-C., and Roan, J., Barriers to physicians' adoption of healthcare information technology: an empirical study on multiple hospitals. J. Med. Syst. 36(3):1965–1977, 2012.

2. Poon, E. G., Jha, A. K., Christino, M., Honour, M. M., Fernandopulle,R., Middleton, B., Newhouse, J., Leape, L., Bates, D. W., and Blumenthal, D., Assessing the level of healthcare information technology adoption in the United States: a snapshot. BMC Med. Inform.

Decis. Mak. 6(1):1, 2006.

Miller, R. H., and Sim, I., Physicians' use of electronic medical records:barriers and solutions. Health Aff. 23(2):116–126, 2004.

4. Blumenthal, D., Stimulating the adoption of health information technology.

N. Engl. J. Med. 360(15):1477–1479, 2009.

5. Dean, J., and Ghemawat, S., Mapreduce: simplified data processingon large clusters. Commun. ACM 51(1):107–113, 2008. doi:10.1145/

1327452.1327492.

6. Dean, J., and Ghemawat, S., MapReduce: a flexible data processingtool. Commun. ACM 53(1):72–77, 2010. doi:10.1145/1629175.

1629198.

7. Horiguchi, H., Yasunaga, H., Hashimoto, H., and Ohe, K., A userfriendlytool to transform large scale administrative data into widetable format using a mapreduce program with a pig latin based script.BMC Med. Inform. Decis. Mak. 12:8, 2012. doi:10.1186/1472-6947-

12-151.

8. Liu, B., Madduri, R. K., Sotomayor, B., Chard, K., Lacinski, L.,Dave, U. J., Li, J. Q., Liu, C. C., and Foster, I. T., Cloud-basedbioinformatics workflow platform for large-scale next-generation sequencinganalyses. J. Biomed. Inform. 49:119–133, 2014. doi:10.

1016/j.jbi.2014.01.005.

9. Santana-Quintero, L., Dingerdissen, H., Thierry-Mieg, J.,Mazumder, R., and Simonyan, V., HIVE-Hexagon: high-performance,parallelized sequence alignment for next-generation sequencingdata analysis. PLoS One 9(6):11, 2014. doi:10.1371/journal.

pone.0099033.

24. Hernan,M. A.,With great data comes great responsibility publishingcomparative effectiveness research in EPIDEMIOLOGY.Epidemiology 22(3):290–291, 2011. doi:10.1097/EDE.0b013e3182114039.

25. Weiss,N. S., The new world of data linkages in clinical epidemiologyare we being brave or foolhardy? Epidemiology 22(3):292–294,2011. doi:10.1097/EDE.0b013e318210aca5.

26. Sturmer, T., Funk, M. J., Poole, C., and Brookhart, M. A.,Nonexperimental comparative effectiveness research using linkedhealthcare databases. Epidemiology 22(3):298–301, 2011. doi:10.1097/EDE.0b013e318212640c.

29. Kim, M., Jung, Y., Jung, D., and Hur, C., Investigating the congruenceof crowdsourced information with official government data: thecase of pediatric clinics. J. Med. Internet Res. 16(2):12, 2014. doi:10.2196/jmir.3078.

30. Alor-Hernandez, G., Perez-Gallardo, Y., Posada-Gomez, R., Cortes-Robles, G., Rodriguez-Gonzalez, A., and Aguilar-Laserre, A. A.,iPixel: a visual content-based and semantic search engine for retrievingdigitized mammograms by using collective intelligence. Inform.

Health Soc. Care 37(3):159–176, 2012. doi:10.3109/17538157.2012.654840.

Big Data Issues And Challenges Based On Implementation Prospective

[1]Deepti Khanna, [2]Sanjive Saxena,
[1]Associate Professor, Jagan Institute of Management Studies, Sector-5, Rohini, New Delhi-85
dipudeepti@gmail.com
[2]Associate Professor, Jagan Institute of Management Studies, Sector-5, Rohini, New Delhi-85
Sanjive.saxena@gmail.com

ABSTRACT

The business world of today is hyper competitive and this hyper competitiveness has forced the business entities to optimize their business and operational processes. Big data is one such means and mechanism which the business units are pro-actively utilizing so as to maintain their leadership position on account of the fact that big data is responsible for generating trends and patterns which is in turn lead to the formulation, development and subsequent implementation of the operational business processes. This is paper is designed to address the issues and the challenges involved in the management of big data from the implementation perspective. The genesis of the paper is the fallout of the demand of big data in the current scenario and its integration with the information technology. The design of the paper follows a structured approach. It starts with the basics of big data gradually moving onto the process of identifying the challenges and the subsequent formulation of the implementation program to address these challenges. In order to address these granular aspects a live practical scenario based on the experience of one of the authors in IT industry is taken into consideration. Finally the paper concludes with limitations and recommendations for further research process.

Keywords—Big Data, Business Units, Challenges, Implementation

INTRODUCTION

The galloping advancements in information technology has resulted in hyper competitiveness of the business units. For, business units have tried and are still trying to inject information technology into their core operational processes and the results of this integration is astounding. Further, since the executive management of business units rely heavily on data to manage their business, they pondered on the possibility of the utilizing data of yesteryears. This possibility of the utilizing the data of yesteryears provided a encouraging results when the trends which were generated by the data of yesterday, was subsequently implemented to achieve competitiveness with the peers.

DEFINITION OF BIG DATA

In plain common parlance, big data refers to a term to include collection of data sets which are large and complex and hence are difficult to process by using conventional or traditional data processing software. Further, when dealing with big data, the business units ardently require for uncovering of hidden trends and patterns, unknown correlations and regressions, trending comprising of marketing patterns, consumer behavior and other vital information which directly or indirectly provides value to the management in executing the various operations of the business. A common example of big data usage and utilization is the weather forecasting of Indian meteorological department, which utilizes the data of past 150 years to predict the trends of monsoon season in India.

Practical Scenario of big data

In order to understand the concept of big data and the challenges involved in the implementation of the same let us consider the case of a company which was started in the year 1963 in Noida city (Fig.1).

Noida office(estd. 1963)

Records Entered Manually in software capable to work in stand alone PC

Product Code	Item Name	Frequency of Purchase	Qty of Material purchased	Client
MHL1234	Raw Iron	Monthly	20 tonnes	H Lever
DHL456	Raw Iron	Daily	40 kgs	HL
FTM12	G. Iron	Fortnightly	60 tonnes	Tata Motor
BF13	Steel	Bi-Monthly	50 kgs	Fiat
MHL1234	Raw Iron	Monthly	20 tonnes	Hind Lever

And So on. Approximately 5 lacs records are maintained through this system.

Figure 1: A company established in 1963 in Noida

The above figure shows the processing details of the raw materials along with additional details such as product code, frequency of purchase, client and the quantity of material purchased.

As the company was doing very well, it expanded its business operations and subsequently established another office in Chennai in the year 1972. (Fig.2)

Chennai Office (Estd. 1972)				
Records Entered Manually in software capable to work in stand alone PC				
Qty of Material purchased	Frequency of Purchase	Client	Product Code	Item Name
20 tonnes	30 days	H Lever	MHL12	Raw Iron
40 kgs	Everyday	HL	DHL46	Raw Iron
60 tonnes	15 days	Tata Motor Corporation	FTM21	G. Iron
150 kgs	Bi-Monthly	Premier Padmni	BF18	Steel
20 tonnes	30 days	Hind Lever	MHL234	Raw Iron
Approximately 10 lacs records are maintained through this system.				

Figure 2: A company another office in Chennai established in 1972

From fig. 2 it is observed that though the columns are the same that is the type of information being maintained in fig. 1 and fig.2 is the same yet the mechanism is different. For example, the first column in fig. 2 is different from fig. 1 and that the frequency of purchase is entered as combination of number and alphabets eg. 15 days whereas in fig. 1 it was fortnightly. Apart from that it is observed that the client's name too is changed from Tata Motors in fig.1 to Tata Motors corporation in fig.2. Thus it is observed that the change has occurred not only in the structure and format but also in the contents of the columns.

Again consider the scenario for the year 2014. The company has been doing very well and it has opened several offices across the country. Each of the office is maintaining similar type of information but in different software (fig.3).

Kolkatta Office

Product Code	Product Name	Client	Quanitity ordered	Date	Amount (in Lakhs)
GI1234	Galvanised Iron	Hindustan Motors	200 Kgs	21/03/2013	Rs. 4.50
ST1080	Steel	Premier Padmini	490 Kgs	22/03/2014	Rs. 5.5
And so on. Approximate number of records 1 crore					

Noida Office

Client	Product Name	Product Code	Quanitity ordered	Amount (in Lakhs)	Date
Hindustan Motors	Galvanised Iron	GI1234	200 Kgs	Rs. 4.50	03/21/2013
Premier Padmini	Steel	ST1080	490 Kgs	Rs. 5.5	03/24/2014
And so on. Approximate number of records 2 crore					

Year 2014

Mumbai Office

Product Code	Product Name	Client	Quanitity ordered	Date	Amount (in Lakhs)
GI1234	Galvanised Iron	Hindustan Motors	200 Kgs	21/03/2013	Rs. 4.50
ST1080	Steel	Premier Padmini	490 Kgs	22/03/2014	Rs. 5.5
And so on. Approximate number of records 1 crore					

Delhi Office

Date	Product Code	Product Name	Quanitity ordered	Amount (in Lakhs)	Client
21/03/2013	GI1234	Galvanised Iron	200 Kgs	Rs. 4.50	Hindustan Motors
22/03/2014	ST1080	Steel	490 Kgs	Rs. 5.5	Premier Padmini
And so on. Approximate number of records 2 crore					

Figure 3:Company's office in Mumbai, Delhi, Kolkata and other places

It is observed in fig. 3 that the type of information is similar due to business processes, yet the scenario has changed completely. In the changed scenario, it is evident that the information is now available in other format apart from the differences in the content and the structure of the information source. This dissimilarity is due to the fact that technology has made rapid advancements and as such it possess a challenge for IT professionals as well as for the executive management of various business units.

CHALLENGES AND ISSUES OF BIG DATA

Having set the background for big data, let us now move to the issues and the challenges which are faced by the Information Technology professionals.

The greatest challenge and issue when dealing with big data is the process of setting the objective which in essence means what is required to be derived when dealing with big data. If the objective is inadequately or vaguely defined then the big data program would result in a failure.

After the objective is set the next challenge or issue moves to the process of developing an implementation program along with the timelines. This implementation program covers the work breakdown structure, the timelines and the resources required to execute the program. This stage of the ensures that nothing is missed out which will defeat the purpose of the program.

Once the plan is finalized, the execution of the plan begins. This phase throws up several challenges to professionals who are responsible for developing the system. The notable challenges that are faced are enumerated below:

- Estimating the size and the volume of data which is required to be processed in the big data management program
- Determining the type of software tools and technology which will assist the IT professional in carrying out big data management program

- Data extraction from different sources and which is in different formats
- Data transformation of data so extracted. Meaning that the original extracted data is required to be standardized. For examples, Tata Motors is required to be changed to Tata Motors Corporation. Similarly, product code is required to be standardized to say 12 digits or 10 digits instead of any arbitrary numbering scheme which has generated 4 digits product code or three digits product code for data of different cities
- Developing a mapping scheme for transforming the data used in standardization of the same. For example what is mechanism for developing the product code 12 digits or 10 digits and so on.
- Designing the structure of tables which to store transformed data

LIMITATIONS

Having understood the various issues and challenges involved in big data implementation program the following are the limitations which are widely experienced

- Every business unit has different operational processes which have evolved over a period of times. This results in the generation of data which may not assist the management in improvisation of the business processes
- When it comes to the process of implementation of big data technology the experience of the human persons plays a key role. For example, an inexperienced IT professional may not be able to understand the dynamics of the business processes and thus will be unable to capture the vital informational requirements of the client resulting in time delays and cost over run
- Due to the fact that big data is all about data any wrong or invalidated data in the data sources will defeat the whole objective of big data program. This in essence means that validation and verification needs to be thoroughly executed

RECOMMENDATIONS

Having understood the various issues and challenges involved in the big data management program, the following recommendations are provided below which will ensure a sound and robust implementation of big data program.

- Need to develop sound analytical skills at various levels of management. This step will ensure that more and more analytics will go into the efficient and effective utilization of data in the data repository. This approach will infuse a new means and insight into the manner in which the business operational processes can be executed so as to gain competitive advantage
- Need to be aware of happenings at global and international level with regard to handling of data so that the same can be capitalized and executed in their processes

CONCLUSION

This paper has discussed the various issues and challenges from implementation perspective of big data. However, the challenges so discussed are generic due to the fact that each business has its own mechanism for handling big data. However, certain basic criterion and approaches are covered in the paper and by judiciously following the limitations and recommendations a business unit stands to gain immensely.

REFERENCES

- *Kumar Praveen ,Rathore Vijay, "Efficient Capabilities of Processing of Big Data using Hadoop Map Reduce", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 6, June 2014*
- *Katal Avita, Wazid Mohammad, Goudar R H, "Big Data: Issues, Challenges, Tools and Good Practices, 978-1-4799-0192-0/13/$31.00 ©2013 IEEE*
- *Madden Sam, "From Databases to Big Data", IEEE, Internet Computing, May-June 2012.*
- *Bakshi Kapil, "Considerations for Big Data: Architecture and Approach", IEEE , Aerospace Conference, 2012*
- *Kaisler Stephen, Armour Frank, J. Espinosa Alberto, Money William "Big Data: Issues and Challenges Moving Forward", IEEE, 46th Hawaii International Conference on System Sciences, 2013*

An Improved Join Algorithm for Result Rate Maximization

Pratik A. Patel[1], Arpit M. Rana[2]
[1]Dept of Computer Science, Gujarat Technological University,
Chandkheda, Ahmedabad – 382424 - India
[2]Dept of Computer Science, Gujarat Technological University,
Chandkheda, Ahmedabad – 382424 - , India

ABSTRACT

MODERN information processing is moving into a reality where we often need to process data that are pushed or pulled from autonomous data sources through heterogeneous networks. Adaptive join algorithms have recently attracted a lot of attention in emerging applications where data are provided by autonomous data sources through heterogeneous network environments. This work is focused on a new join operator for such situations. Its main advantage over traditional join techniques is that they can start producing join results as soon as the first input tuples are available, thus, improving pipelining by smoothing join result production and by masking source or network delays. In this paper, the first propose Double Index NEsted-loops Reactive join (DINER), a new adaptive two-way join algorithm for result rate maximization. DINER combines two key elements: an intuitive flushing policy that aims to increase the productivity of in-memory tuples in producing results during the online phase of the join, and a novel re-entrant join technique that allows the algorithm to rapidly switch between processing in memory and disk-resident tuples, thus, better exploiting temporary delays when new data are not available. Then extend the applicability of the proposed technique for a more challenging setup: handling more than two inputs. Dynamic Multiple Index Join (DMIJ) is a multi-way join operator that inherits its principles from DINER.

1. INTRODUCTION

The word Join is a means for combining fields from two tables by using values common to each. In Real systems, it is difficult to maintain all the data is stored in one large table. To do so would require maintaining several duplicate copies of the same values and could threaten the integrity of the data .Instead, IT department everywhere almost always divide their data among several different tables. Because of this, a method is needed to simultaneously access two or more tables by using join operation. Adaptive join algorithms were created in order to lift the limitations of traditional join algorithms in such environments. By being able to produce results whenever input tuples become available, adaptive join algorithms overcome situations like initial delay, slow data delivery, or busty arrival, which can affect the efficiency of the join. Here, main focus is to show how an improved join algorithm works efficiently in real systems.

2. A JOIN PROCESS OVERVIEW

Joins are one of the basic constructions of SQL and databases such as, they combine records from two or more database tables into one row source, one set of rows with the same columns and these columns can originate from either of the joined tables as well as be formed using an expressions and built-in or user-defined functions.

Joins are used for joining records or fields from two or more tables in a database by using a value common to both the tables and the result set can be stored or saved in a table [2].

There are four types of joins and they are specified by ANSI (American National Standard Institute) and they are INNER, OUTER, LEFT, and RIGHT. Inner join are further classified into equi join, natural join and cross join. Outer join are further classified as left outer join, right outer join and full outer join. Two tables are used as an example of joins; they are Dept ID column of the Emp table and Dept table.

Emp Table

LastName	DeptId
Aa	11
Bb	13
Cc	13
Dd	14
Ee	14

Dept Table

Dept Id	DeptName
11	Sales
13	Engineering
14	Clerical
15	Marketing

Figure 1 Example of Join

Inner join are considered as a common operation of join and they are also a default type of join based on the predicate. They combine the values of two tables and the results are kept in new table. Inner join has both explicit join notation and implicit join notation.

Outer join does not expect any matching record and they does not require each record in two tables to

be joined to have a matching record. Outer join does not have Implicit join notation. Explicit join notation and implicit join notation are the ways of expressing join syntax and they are specified by SQL explicit join notation uses the keyword "JOIN" and "On" [2]

Select * from Emp INNER JOIN Dept On Emp.DeptID = Dept.DeptID;

Implicit join notation list the join table and they use select statement:-

Select * from Emp, Dept Where Emp.DeptID = Dept.DeptID;

Adaptive Join: - Adaptation schemes for join queries are significantly more complicated to design and analyze compared to those for selection ordering for several reasons.

The key performance of adaptive joins is rapid availability of first results and a continuous rate of tuple production. It overcomes the situation like initial delay, slow data delivery or bursty arrival, which can affect the efficiency of join [1]

3. EXISTING WORK-LITERATURE REVIEW

3.1 EXISTING WORK-LITERATURE REVIEW

The main three categories of join algorithms are

3.1.1) Nested-loop join algorithm

3.1.2) Sort-merge join algorithm

3.1.3) Hash-based join algorithm

We have gone through other join also in review below:

3.1.1 Nested-Loop Join Algorithm [2]:-

Nested-loop join is considered as a one of the simplest algorithm of join where, for each record of the first table the entire records of the second table has to be scanned. This process is repeated for each and every record of the first table that is for all the first table records. The loop is of two levels and they are outer loop and the inner loop. First table loop is called as outer loop and the second table loop are called as inner loop. As this, Nested loop join algorithm has a repeated input/output scans of one of the table. They are considered as inefficient.

Let the two tables be A and B, then the algorithm of Nested-loop algorithm are as for each record of table A

Read record from table A

For each record of table B

Read record from table B

Compare the join attributes

If matched

Then

Store the records

Example: - Consider schema of two tables “Customers” and “Sales”

```
Create Table Customers (Cust_Id int, Cust_Name varchar (10))
  Insert Customers values ( 1,'PRATIK')
  Insert Customers values (2,'HEMANT')
  Insert Customers values (3,'MONALI')
```

Another table is 'Sales',

```
Create Table Sales (Cust_Id int, Item varchar (10))
  Insert Sales values ( 2,'Camera')
  Insert Sales values (3,'Computer')
  Insert Sales values (3,'Monitor')
  Insert Sales values (4,'Printer')
```

`Query is written as:-

```
Select * from Sales S inner join Customers C on S.Cust_Id = C.Cust_Id
```

In above example, the outer table is “Customers” while the inner table is “Sales”. Thus, it begins by scanning the “Customers” table. It takes one customer at a time and, for each customer, it scans the “Sales” table. Since, there are 3customers; it executes the scan of the “Sales” table 3 times. Each scan of the “Sales” table returns 4 rows. It compares each “Sales” to the current “Customers” and Evaluate whether the two rows have the same Cust_Id.Return these rows of Cust_Id. It has 3 customers and 4 sales. So, it performs this comparison a total of 3*4 or 12 times. Only 3of these comparison results in a matching.

3.1.2 Sort-Merge Join Algorithm [2]:-

Sort merge algorithm are considered as an efficient join algorithm when compared to Nested loop join algorithm. Sort merge join algorithms have two operations and they are sorting and merging. In sorting operation, the two tables to be joined are sorted in ascending order. In merging operation, the two sorted

tables are merged. Sort records of table a based on the join attribute Sort records of table B based on the join attribute.

Let i = 1 and j =1

Repeat

Read record A (i)

Read record B (j)

If join attribute A (i) < join attribute B (j) Then

i++

Else

If join attribute A (i) > join attribute B (j) Then j++

Else Put records A (i) and B (j) into the Query.

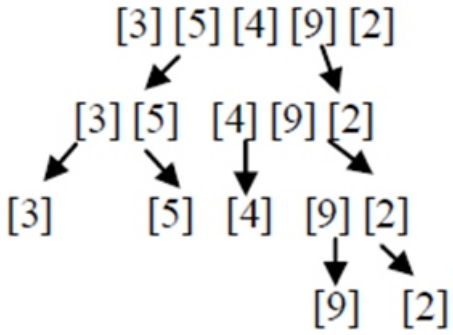


Figure 2. Example of SMJ.

3.1.3 Hash Based join algorithm [2]: -

In hash based join algorithm, hashing and probing are the two processes. A hash table is created by hashing all records of the first table using a particular hash function. Records from the second table are also hashed with the same hash function and probed. If any match is found, the two records are concatenated and placed in the query result. A decision must be made about which table is to be hashed and which table is to be probed. Since a hash table has to be created, it would be better to choose the smaller table for hashing and the larger table for probing. The hash join algorithm is given as

Let H be a hash function

For each record in table B

Read a record from table B

Hash the record based on join attribute value

using hash function H into hash table

For each record in table A

Read a record from table A

Hash the record based on join attribute value using H

Probe into the hash table

If an index entry is found then

Compare each record on this index entry

With the record of table S

If matched then

Put the pair into query.

3.1.4 Double Pipelined Hash Join (DPHJ) [5]:-

The double Pipelined Hash Join (DPHJ) is another extension of the symmetric hash join algorithm [7]. DPHJ has two stages. The first stage is similar to the in-memory join in the symmetric hash join and XJoin[5]. In the second stage, pairs that are not joined together in the first phase are marked and are joined in disk. DPHJ [6] is suitable for moderate size data, but does not scale well for large data sizes.

3.1.6 MJoin [9]: -

The basic idea of the MJoin algorithm is simple: generalize the symmetric binary hash join and the XJoin algorithms to work for more than two inputs. Our primary goal is to maximize the output rate during the memory-to-memory phase of the MJoin. In MJoin, the disk to-memory phase is intended to allow the system to generate outputs while its inputs are blocked, while the disk to-disk phase is intended to generate any final answers after the inputs have terminated. Interestingly, for the MJoin, how we handle memory overflow determines the output rate of the memory-to-memory phase.

4. NEW PROPOSED JOIN TECHNIQUE

4.1 DINER: -

In this work, First algorithm is Double Index Nested-Loop Reactive join (DINER), an adaptive two-way join algorithm. DINER [4] combines two key elements: an intuitive flushing policy that aims to increase the productivity of in-memory tuples in producing results, and a novel reentrant join technique that allows the algorithm to rapidly switch between processing in-memory and disk-resident tuples, thus, better exploiting temporary delays when new data are not available. .

Consider two finite relations RA and RB, which may be stored at potentially different sites and are used to our local system. Incoming tuples from both relations share the available memory. A separate index on the join attribute is maintained for the memory resident part of each input relation .A separate index is maintained for the memory resident part of each input relation. When incoming tuples arrive, compute the memory occupied by the buffer and in-memory tuples being processed by the algorithm. Each relation is associated with a disk partition, which stores the tuples from the relation which do not fit in the memory and have been flushed to disk. Every pair of tuples between the two relations will be joined exactly once and produced result. Each tuple is residing in main memory has a join bit, which is actually part of the index. This is initially set to 0. Whenever an in-memory tuple produce a join result, its join bit set to1.[2]

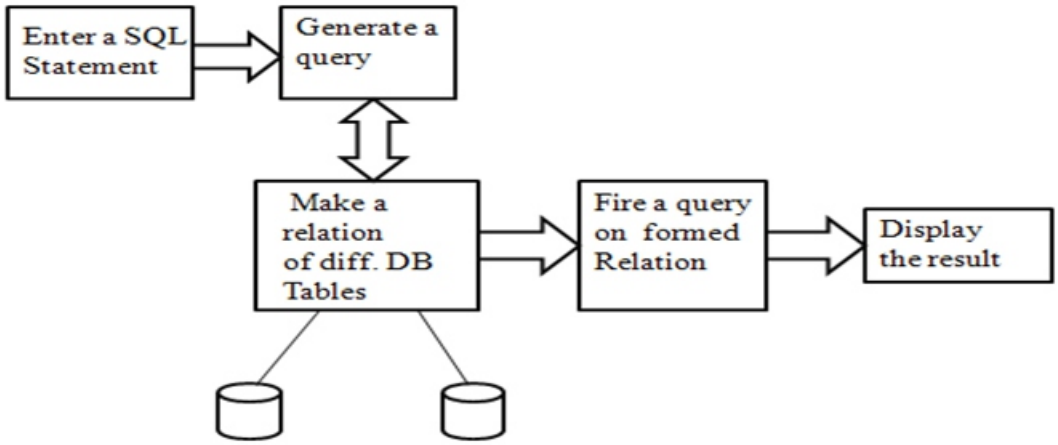


Figure 3. The General Flow Diagram of Proposed System

4.2 DMIJ: -

Dynamic Multiple Index Join (DMIJ) is used for optimization of data in homogeneous environment .our experiments using two tables demonstrate that DMIJ [1] outperforms previous adaptive join algorithms in producing result tuples at a significantly higher rate, while making better use of the available memory. Our experiments also show that in the presence of multiple inputs, DMIJ manages to produce a high percentage of early results. . Consider two finite relations. It maintains for each joined relation a separate index on each join attribute. When a new tuple arrives, it is stored and indexed in the memory space of its relation, based on the relations join attribute. This new tuple that need to be joined with all the matching in-memory tuples belonging to all other relations participating in the joins. Flushing is applied when memory gets full.

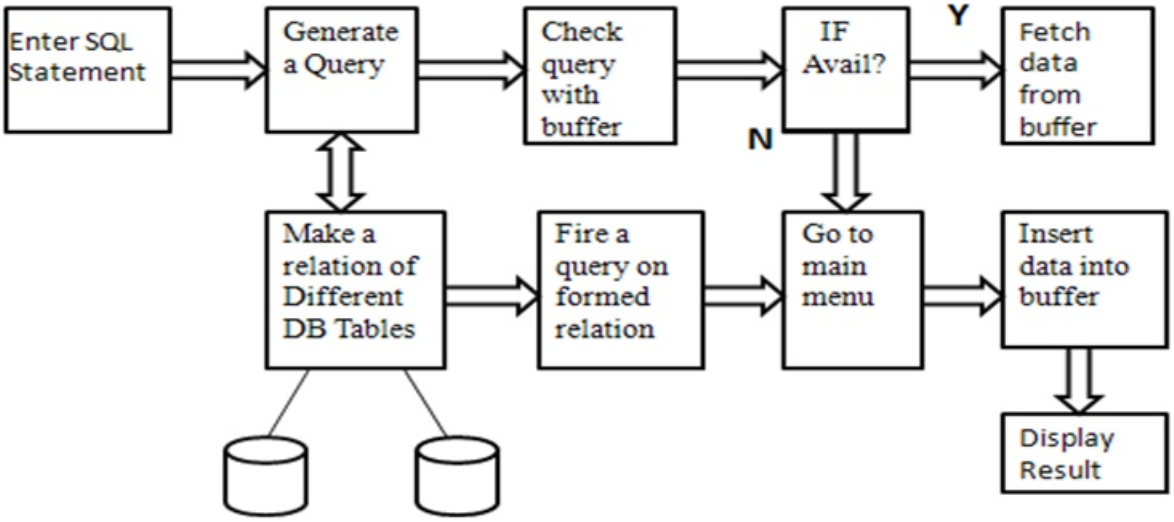


Figure 4. The General flow Diagram of DMIJ

4.3 Difference between DINER and Existing algorithm:

1. DINER supports equi-joins and range queries. PMJ also supports range queries but it has some limitation due to its poor blocking behavior.
2. DINER will introduces flushing policy, is used to create and maintained three overlapping value regions.
3. DINER will introduces a more responsive phase that allows the algorithm to quickly move into processing tuples when both data sources block.
4. In Leaner Algorithm, DINER improves its relative performance compared to the existing algorithm.[1]

5. CONCLUSIONS

In this work, to introduce DINER, a new adaptive join algorithm for maximizing the output rate of tuples, when two relations are being streamed to and joined at a local site. The advantages of DINER stem from 1) its intuitive flushing policy that maximizes the overlap among the join attribute values between the two relations, while flushing to disk tuples that do not contribute to the result and 2) a novel re-entrant algorithm for joining disk resident tuples that were previously flushed to disk. Moreover, DINER can efficiently handle join predicates with range conditions, a feature unique to this technique. To also present a significant extension to this framework in order to handle multiple inputs. The resulting algorithm, DMIJ addresses additional challenges, such as determining the proper order in which to probe the in-memory tuples of the relations, and a more complicated bookkeeping process

during the Reactive phase of the join. Through this experimental evaluation, we have demonstrated the advantages of both algorithms on a variety of real and synthetic data sets, their resilience in the presence of varied data and network characteristics and their robustness to parameter changes.

6. REFERENCES

[1] Mihaela A.Bornea, Vasilis Vassalos, Yannis Kotidis, Antonios Deligiannakis: Adaptive Join Operators for Result Rate Optimization on Streaming Inputs. IEEE Trans. Knowl. Data Eng. 22(8): 1110-1125 (2010)

[2] J.Jayashree and C.Ranichandra "Join Algorithm for Efficient Query Processing For Large Datasets" Asian Journal of Computer Science and Information Technology 2: 3 (2012) 31 –35.

[3] J. D. Ullman, H. Garcia-Molina, and J. Widom. Database Systems: The Complete Book. Prentice Hall, 2001

[4] M. A. Bornea, V. Vassalos, Y. Kotidis, and A. Deligiannakis.DoubleIndex Nested-loop Reactive Join for Result Rate Optimization. In ICDEConf., 2009

[5] David Taniar, Clement H.C. Leung, Wenny Rahayu, Sushant Goel. (2008) "High-Performance Parallel Database Processing and Grid Databases" A John Wiley

[6] Z. G. Ives, D. Florescu, and et al. An Adaptive Query Execution System for Data Integration.In SIGMOD, 1999.

[7] W. Hong and M. Stonebraker. Optimization of Parallel Query Execution Plans in XPRS. In PDIS, 1991

[8] T.Urhan and M.J.Franklin.Xjoin: A Relatively scheduled pipilined join operator.IEEE Data Eng.Bull,23920,2000

[9] S.D Viglas,J.F.Naughton and J.Burger.Maximizing the output rate of multiway join queries over streaming information sources.In VLDB 2003: proceeding of the 29th international

[10] J. Dittrich, B. Seeger, and D. Taylor. Progressive merge join: A generic and non-blocking sort-based join algorithm. In Proceedings of VLDB,2002.

[11] M. F. Mokbel, M. Lu, and W. G. Aref. Hash-Merge Join: A Non blocking Join Algorithm for Producing Fast and Early Join Results. In ICDE Conf., 2004.nal conference on very large databases 2003.

[12] Y. Tao, M. L. Yiu, D. Papadias, M. Hadjieleftheriou, and N. Mamoulis. RPJ: Producing Fast Join Results on Streams Through Rate-based Optimization. In Proceedings of ACM SIGMOD Conference, 2005.

BOOKS

[14] J. D. Ullman, H. Garcia-Molina, and J. Widom. Database Systems: The Complete Book. Prentice Hall, 2001

[15] David Taniar, Clement H.C. Leung, Wenny Rahayu, Sushant Goel. (2008) "High-Performance Parallel Database Processing and Grid Databases" A John Wiley

WEBSITES

[13] http://en.wikipedia.org/wiki/Join_(SQL)

[14] http://www.softwaretestinghelp.com/types-of-software-testing/

[15] http://www.microsoft.com/isapi/redir.dll?prd=ie&pver=6&ar=msnhome

Data Warehouse

[1]Shilpa Seth

A B S T R A C T

The use of data warehousing technology is increasing in companies with time. In the data mining process, the first step is the data construction, where the data must be organized and stored in the data warehouse (DW). The construction of a DW is a complicated and challenging task because it needs a high knowledge of the technologies such as data integration and data warehousing and the contents of data sources .The prime thing where one can lack is the architecture of DW which can be perceive by reading this topic. This paper presents an overview of Data Warehouse and its architecture which helps in understanding its working.

Keywords—Data Warehouse, Metadata, ETL.

INTRODUCTION

An elementary concept of a data warehouse is the difference between data and information.

Data: Data is a simple matter obtained before any operations or processing. It is an unprocessed form like any facts or figures etc. Data is made up of observable facts that are found in operational or transactional systems. In a data warehouse environment, data only comes to have importance to end-users when it is modified and shaped as information.

Information: Information is what we get after applying operations on data. Information is a combined collection of facts and is used as the basis for decision-making.

DATA WAREHOUSE

Data warehouse is a place where all the data is stored obtained from different sources. These are the main repositories of combined data from one or more separate sources. It may include current and past data which can be used for reporting and data analysis. For example: A business man is having a godown where he stores all his product items which he has to sell to different users. So that godown act as a data warehouse and those products act as a data or information and the customers play a role of business users or analysts.

Data warehouses stores current and historical data and are used for creating trending reports such as annual and quarterly comparisons. It is a relational database that is drawn for query and analysis rather than transaction work. It usually contains previous data inherited from transaction data, but it can also have data from different sources. It enables an organization to coalesce (combine) data from several sources. It provides a data environment that serves as the unified source of data for processing information.

The data warehouse has specific characteristics that include the following:

Subject-Oriented: Information is presented according to specific subjects or areas of interest, not simply as originated. Data is processed to provide information related to a particular subject.

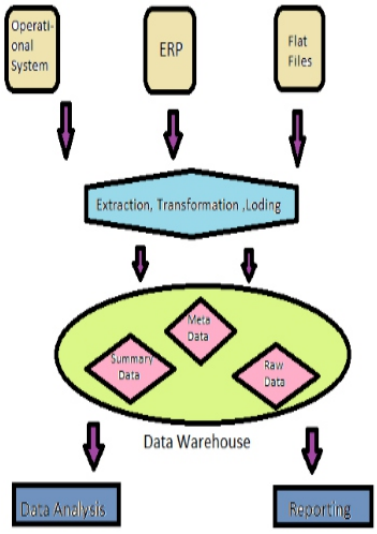
Integrated: The information from distinct sources is integrated to form data warehouse. It contains information about variety of subjects.

Non-Volatile: Information doesn't modify each time an operational process is executed. Information remains stable regardless of when the warehouse is accessed.

Time-Variant: It contains a history of the subject, as well as present information. Historical information is a valuable component of a data warehouse.

Accessible: The main idea of a data warehouse is to provide easily accessible information to users.

Process-Oriented: Data warehousing must be viewed as a process for conveyance of information. The maintenance of a data warehouse is ongoing process.



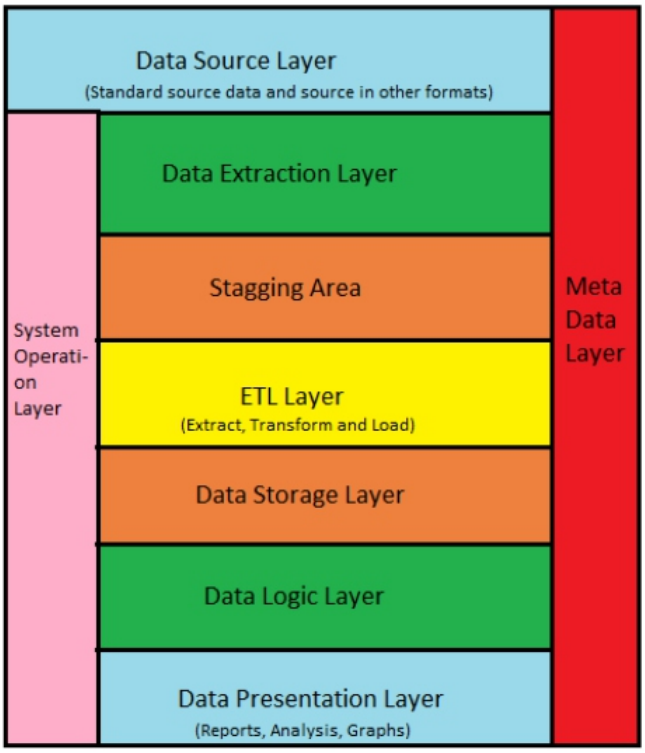
ARCHITECTURE OF DATA WAREHOUSE

Different data warehousing systems have different designs. Some may have an operational data store, while some may have numerous data marts.

Some may have a few data sources, while some may have dozens of data sources. In view of this, it is far more reasonable to present the different layers of data warehouse architecture rather than discussing the specifics of any one system.

In general, all data warehouse systems have the following layers:

1. Data Source Layer
2. Data Extraction Layer
3. Staging Area
4. ETL Layer
5. Data Storage Layer
6. Data Logic Layer
7. Data Presentation Layer
8. Metadata Layer
9. System Operations Layer



Each component is discussed below:

1. **Data Source Layer.** The Data Source layer represents all data sources that are used as information suppliers to DW. The data source can be of any form like reports, excel sheets, survey data, democratic data etc.

2. **Data Extraction Layer:** Data gets fetched from the data source into the data warehouse system. It is responsible for extracting data from multiple sources.

3. **Staging Area:** A data staging layer maintains a boundary between the processes that deliver data to the data warehouse and the way in which data is presented within the data storage layer.

4. **ETL Layer:** This layer involves the functioning of extraction, transformation and loading. The module of extraction, transformation and load collects relevant data, cleans, integrates, transforms and stores these data in the staging area.

5. **Data Storage Layer:** This layer is the actual physical data model for data warehouse tables. It provides the brief description of the entities, attributes, and relationships present in the data warehouse.

6. **Data Logic Layer:** This layer depicts some transformation rules which don't affect the data but it affects its designing.

7. **Data Presentation Layer:** This refers to the information that is presented to the users. It can be in any form like reports, analysis etc. Generally OLAP tool or a reporting tool is used in this layer.

8. **Metadata Layer:** This is where information about the data stored in the data warehouse system is stored. The information contained in this repository is crucial for overall operation and maintenance of the components.

9. **System Operations Layer:** This layer shows information on how the data warehouse system works, such as system status, and user access history.

REFERENCES

1. *http://www.1keydata.com/datawarehousing/datawarehouse.html*
2. *http://www.tutorialspoint.com/dwh/dwh_overview.htm*
3. *http://www.webopedia.com/TERM/D/data_warehouse.html*
4. *http://datawarehouse4u.info/*

Strengthening Collaborative Mind Map Tool To Facilitate Requirement Elicitation Process

[1]Shahnawaz Ahmad, [2]Yahiya Gazi
[1,2]M. Tech. Scholar, Department of Computer Science and Engineering,
Faculty of Engineering and Technology, Al-Falah University,
Dhauj, Faridabad, Haryana, India.
E-mail: shahnawaz98976@gmail.com, yahiyaghazi@gmail.com

ABSTRACT

Requirements elicitation is a crucial task in any software development process. It is notable as a major contributor to the project failure. To minimize and subsequently improve the process, folding Mind Map concept in the elicitation process and implementing it collaboratively is being proposed in this study. The aim of the study is to study the impact of the Mind Map and groupware in facilitating requirements elicitation. A prototype based on Power Meeting is developed to illustrate the idea of mind map groupware called Collaborative Mind Map Tool (CMMT) which to be analyzed its impact towards requirement elicitation process.

Keywords—requirements elicitation, folding Mind Map, Collaborative Mind Map Tool (CMMT)

INTRODUCTION

Requirements elicitation is a process of identifying the requirements of the problem to be resolved by gathering, uncovering and understanding the needs of stakeholders. It involves the process of extracting and understanding the needs of stakeholders, and defining the application domain and problem context with explicit and precise descriptions. Requirements are obtained from the stakeholders as the primary resources, careful analysis of the organization, the application domain and business process where the system will be deployed (Kotonya, et al., 1998). Obviously those activities require intensive communication, collaboration and cooperation between stakeholders and requirement engineers. During this process, it is essential to capture the "right" understanding of the problem and interpret it correctly in agreeing representation. It has well been recognized that requirements elicitation is the most crucial and important aspect in any software development prior to analysis and specification stage.

Despite its crucial process, it is notable that requirement gathering is a major contributor to the failure in any software development project (Macaulay, 1996; Bohem, 1981; Rahman, 2004). Studies indicate

that among the main problems in the requirements elicitation process are: gaps in communication and understanding, inadequacy of management; lack of knowledge and skilled people in approaching the requirement process; and process with wrong techniques and methods (Al-Rawas, et al.,1996; Weigers, 2003; Tsumaki, et al., 2005).

To overcome the situations, a Mind Map, which is a thinking "tool" that reflects how the information is stored and retrieved in a more organized and systematic way inside a human brain, will be bring into the requirement elicitation. Mind Map is a technique which uses graphical illustration in expressing thoughts and idea based on the concept of Radiant Thinking- a natural function of human mind (Buzan, et al., 1995). The subject of attention is placed in a central image, and radiates the ideas from the central image and hierarchically expands and associates the branches and its sub-branches with keywords. Information is categorized and classified, and subject is layout in a "snapshot" and organized way.

Furthermore, incorporating objects which are more stimulating than words such as shapes, images and colors in Research done by Ralph Haber, and later by R.S. Nickerson, have proven that images are more stimulating than words and thus generate more creative ideas and encourage better memorization. The Mind Map is used widely by many individuals and organizations and can be applied in most activities or situations, i.e. in decision making, analysis, problem solving, to-do list, note taking, brainstorming, presentation or even to write down simple notes.

Exploitation of Mind Map will generally multiply individual personal creative capabilities when collaborate in a group (Buzan, et al., 1995). Psychologically, group mind map encourages information sharing and teamwork (Morris, et al., 1998). It generates the sense of contribution which then encourages the team members to commit on job execution by viewing and understanding ultimate objective of the subject matter. This is essential criterion in teamwork to have same vision and mission (Hughes, 2008).

Greater exploitation can be achieved when incorporates the group mind map into groupware. Groupware is a technology that supports systematic and structured group collaboration (RAMA, et al., 2006). Generally, groupware is categorized into two primary dimensions of perspectives; the time and the space. The time illustrates the collaboration happens at same time (synchronous) or different time (asynchronous), whereas the space view on the perspective of where the collaboration is executed; whether it is in the same place (co-located) or different place (none co-located).

There are many significant advantages of groupware compared to a single user system, i.e. 1) To facilitate communication; make it faster, clearer and more persuasive 2) As a mechanism to transfer and share knowledge 3) To help in motivating individual to perform better 4) To collaborate a group of people of the same interest 5) A way to form structured group coordination and proactive collaboration 6) To save time and cost.

However, building groupware system is considered to be more challenging compared to individual system due to distinguish features in groupware system such as 1) presentation of groupware- high level of usability criteria; easy to learn, easy to use, error tolerant and subjectively pleasant, 2) collaboration and communication mechanism-classify the private and sharing information and maintain the consistency of the data throughout the instance, 3) work coordination- a defined process to be followed to encourage work accomplishment 4) openness- easily be integrated with other applications 5) group composition- management and control of the accessibility of groupware participants (Wells, 1996; Volksen, 1992).

COLLABARATIVE MIND MAP TOOL (CMMT) IN FACILATATING REQUIREMENT ELICITATION

Many studies have been done to find ways in improving requirement elicitation mainly due to the cognizance of the requirement engineering as one of the crucial factors contributing to the success of a software project. The process is more about people, communication and collaboration. A few studies and attempts to use Mind Map tool in various phase of software development process has been encountered (Eric T. Blue, 2006). However, very little usage of Mind Map in requirement activities as compared to project management activities. One convincing example that practically uses Mind Map in requirement elicitation is by Kenji Hiranabe. In his article, he has outlined few benefits and illustrated the example with specific template in using Mind Map in requirement process. He also provides a way on how to map mind map into UML diagrams using JUDE-a design and communication tool (Hiranabe, 2007; Change Vision, Inc, 2006).

Surveyed by Chucked Frey (2007) shows that collaboration is the most beneficial feature in Mind Map tool, and according to Nikos Drakos, a Gartner Inc. analyst, Mind Map will become more interesting and beneficial if it is move towards collaborative online (Gilhooly, 2006). this indicates that collaborative Mind Map tool has potential significant impact to people and the process of software development in particular.

REQUIREMENT AND SPECIFICATION

Conceptually, the proposed CMMT hybrid the features from Mind Map theories and the distinguish features of groupware system; with the objective to improve the current problem facing in requirement elicitation.

The following are features proposed to be in CMMT:

Presentation of the tool: Inherited from the concept of Mind Map, the tool should tag with high level of usability. Conceptually, Mind Map is a simple and easy concept as it is based on how our brain works. Thus, the tool should be easy to integrate into practice and simple to use. Additionally, color, shapes and images will be used to trigger information which make the tool more attractive and motivate users to use it.

Communication mechanism: Radiant concept of mind map should articulate the information in a better way and present them in a more structure, clearer and more focus. With categorization and prioritization of information help to better understand the requirement and reduce ambiguity and inconsistency of requirements. Enriching the value of the tool, it should support both "synchronous" and "asynchronous" features. Having these features, existing requirement elicitation techniques (e.g. introspection, task analysis, brainstorming, laddering, requirements workshops, goal-based approach, scenarios, viewpoints and etc) can easily blend with mind map.

Collaboration mechanism: Information can be classified as private or sharing depending on the scenarios. A control mechanism is provided to manage the accessibility of information among the participants with centralize architecture to maintain data consistency. Nowadays, increasing in globalization has demanded multisite software development organizations (Damian, et al., 2003). Making the tool as a web based, it able to support distributed geographical requirement gathering activities which obviously save time and cost.

Work coordination: A floor control or session control is used to control the session and to ensure the process is executed in defined and structured manner. It also controls the accessibility of the participants in each elicitation session. This is to make the elicitation process more systematic, focus and efficient.

Requirement elicitation specific features: 1) Data Dictionary – A feature where crucial term or glossary can be defined. One of the problems in requirement elicitation is not sharing common understanding

of concepts and terms. This ensures all crucial terms and concepts can be defined and agreed between stakeholders. 2) Record and Playback - A feature to record and replay each of the session. This can be used to further understand and review on the agreed requirements or decision. 3) Attachment Uploading – A feature where each requirement/information can be attached with supporting resources (e.g. organization plan, business process, legacy system manuals and etc.). This is to cater for which requirements may originate from different sources and formats. 4) Prioritization and Status updates – A feature where each requirement can be attached with priority number and status. This is to prioritize on the crucial requirements and to set the status of requirements (e.g. to revisit and decide on the specific requirement.). 5) Chat – A feature where users can discuss on any issues rose during the requirement elicitation session.

IMPLEMENTATION STRATEGY

The prototype of CMMT is built on top of Power Meeting framework. Power Meeting, a work-piece (on going research) by Dr. Weigang Wang (2008), is a web based synchronous groupware framework which offers flexible, customizable and extensible groupware development environment. It basically provides with basic collaborative functionalities such as floor session control, user management, transaction management, text messaging, voice chat, session control, and also other existing groupware plug-in tools such as calendar tool, pincardboard and presentation slides.

From a high level technical point of view, the PowerMeeting framework is modeled based on Model (Shared Model) – ViewController (MVC) architecture and transactional replicate architecture employed from CommonGround toolkit, to provide with the basic groupware services such as user management, session and group management, replication and transaction management and persistence management (Wang, 2008). The implementation is built using AJAX technologies offered by Google Web Toolkit (GWT). AJAX is used to gain the benefit of instantaneous and faster response time, asynchronous partial updates and other rich user experience characteristics.

The flexibility and collaborative features furnish by PowerMeeting framework has made the prototype implementation and deployment of CMMT easier. Figure 1, illustrates the CMMT on the PowerMeeting interface.

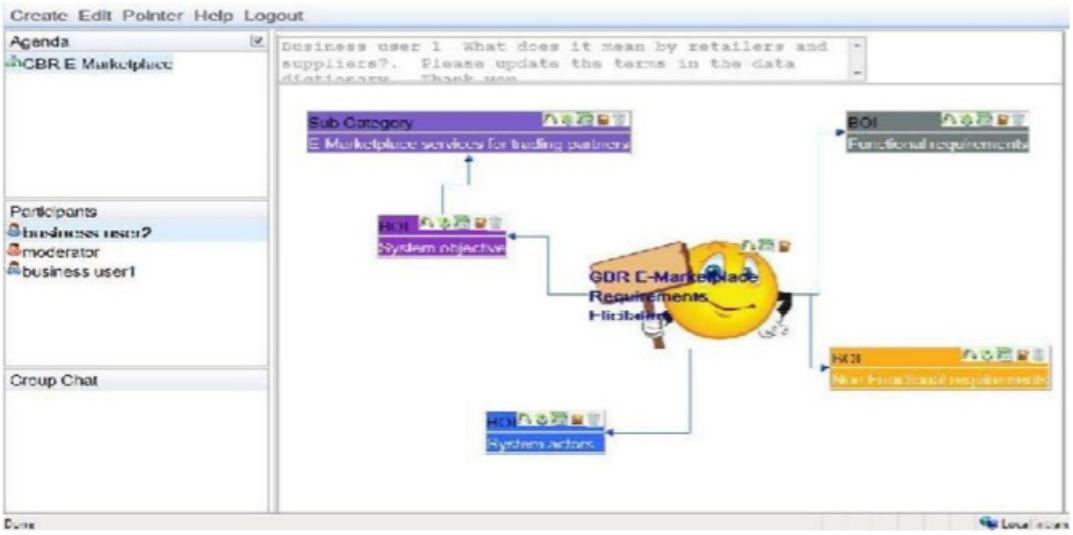


Figure 14: CMMT on *PowerMeeting*

MIND MAP TOOL ARCHITECTURE

Mind Map tool is plugged-in into the PowerMeeting framework through the creation of Model/Shared Model objects and View-Controller objects. Model objects are used to hold and process the data and information in CMMT while the View-Controller objects hold the responsibilities to interpret the events initiated by the users, convey the instructions to the model for data processing or application specific processing and notify the view of any changes in the model object.

Below figure shows the object models of CMMT.

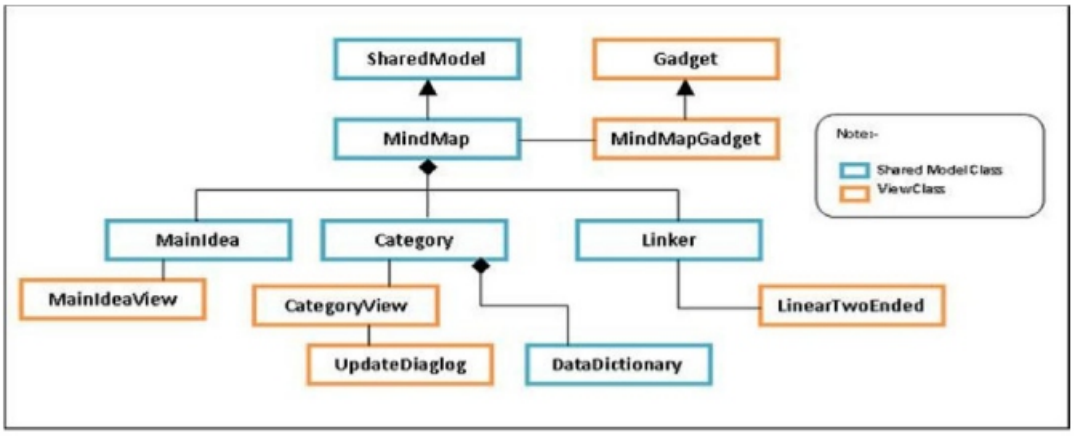


Figure 15: CMMT Object Model

Model objects are defined mainly on the basis of mind map theory by which it has the central subject matter (MainIdea) and radiates the idea into branches (Category) through associations (Linker). CMMT model objects need to extend SharedModel object provided by the framework to inherit the services to

handle replication and transaction. Having this, CMMT model object instances can be made available and shared by all users in real time.

EVALUATION AND ANALYSIS

The prototype-CMMT was successfully built which furnishes with three basic essential elements outlined to be achieved i.e. features that supporting requirement elicitation activities, Mind Map concept, and collaboration. A scenario of a software project development is given to a small group of participants to be executed for evaluation. The selected participants are those who are already familiar with elicitation process but having minimal background knowledge on mind map. Having this focus group we aim to get more reliable insight on the impact of CMMT towards elicitation process.

Feedback and comments were gathered through questionnaire at the end of the elicitation activities. The goal are to capture responses on usefulness of CMMT in requirement elicitation, the effortlessness in using CMMT, the layout and the performance of CMMT. Scale of 1(Strongly Agree) to 5(Strongly Not Agree) were used in capturing the responses. As an addition, participants can also plainly comments on a free text space provided.

More than 70% of participants agree that CMMT overlays a structured way of capturing requirements resulting to better understanding of the gathered requirements. However, only 30 % of respondents agree that CMMT promotes group work and encourages more detailed requirements capturing. In terms of the layout, over 60% responses in the group agree that the CMMT layout is easy to use and learn yet they are not sure on the effectiveness of the layout to generate creativity and memorization. On the other hands, the response time of CMMT is generally bearable to the participants. Since it was only tested by a small group of people, this finding would not be a strong claim to support the hypothesis. However, the work presented has provided an insight for future works in collaborative groupware and requirement engineering field.

CONCLUSION AND FUTURE WORKS

Observing the result from the evaluation and analysis, we believe that CMMT has demonstrated its potential capabilities in supporting requirement elicitation process.

In future, CMMT prototype is to be tested and validated in real software development project situation in order to get an actual insight of CMMT in terms of the significance offered and the improvement

elicitation process. Also, further refinement should focus on implementing all the proposed features and subsequently extending the feature list so that it can fit to other requirement engineering processes; i.e. requirement modeling and analysis, requirement validation and requirement management.

REFERENCES

Al-Rawas, A. & EasterBrook, S. (1996). Communication Problems in Requirement Engineering : A Field Study. London : Proceeding of the First Westminster Conference on Professional Awareness in Software Engineering. Royal Society.

Bohem, B. W. (1981). Software engineering economics. New Jersey : Prentice-Hall. Buzan, T, & Buzan, B. (1995). The Mindmap Book. London : BBC Books.

Change Vision, Inc. (2006). JUDE/Professional Design & Modeling. Retrieved from http://jude.change-vision.com/jude-web/product/professional.html#mindmap_convert.

Damian, Daniela E. & Zowghi, Didar. (2003). Requirement Engineering challenges in multi-site soft ware developmet organization. s.l. : Requirement Engineering Journal, 8.

Eric T. Blue. (2006). Mind Mapping and the Software Development Life Cycle. Eric's Blue Blog : Technology, Philosophy, and Personal Development. Retrieved from http://eric-blue.com/2006/12/16/mind-mapping-and-the-software-development-life-cycle/

Gilhooly, K . (2006) Business on the map: Mind-mapping tools are finding a home in corporate IT. Computerworld Inc. Retrieved from:http://www.computerworld.com/action/article.do?command=viewArticleBasic&taxonomyName =Software&articleId=112247&taxonomyId=18&pageNumber=1

Hiranabe, K. (2007) Agile Modelling with Mind Map and UML. Retrieved from StickyMinds.com website: http://www.stickyminds.com/sitewide.asp?ObjectId=11861&Function=DETAILBROWSE&ObjectType=ART

Hughes, Thomas B. (2008). What is Team Work and Why is it Important to Your Organization. Retrieved from EzineArticles.com website: http://ezinearticles.com/?What-is-Team-Work-and-Why-is-it-Important-to-Your-Organization&id=1693070

Kotonya, G. & Sommerville, I. (1998). Requirements Engineering: Process and Techniques. NewYork : Willey.

Macaulay, Linda A. (1996). Requirement Engineering. London : Springer.

Morris, S, & Smith, J. (1998). Understanding Mind Maps in a week. London : Hodder & Stoughton.

Rahman, M. (2004). Requirement Engineering : UK e-Science Usability Task Force. [Internet] Oxford: Oxford e-Science Centre Review, Oxford University.

Rama, J. & Bishop, J. (2006). Survey and comparison of CSCW Groupware applications. Proceedingsof SAICSIT. s.l. : SAICSIT.

Tsumaki, T. & Tamai, T. (2005). A Framework for Matching Requirement Engineering Techniques to Project Characteristics and Situation Changes. Paris : Proceedings of SREP'05, Paris, France.

Völksen, G. (1992). Approach Strategies to Groupware. Retrieved from http://www.mathematik.uni-marburg.de/~voelksen/publications/wmwork/www/approach_1.html

Wang, W. (2004). PowerMeeting. Retrieved from http://sites.google.com/site/meetinginbrowsers/Home

Wang, W. (2008). PowerMeeting on CommonGround: web based synchronous groupware with rich user experience. In Proceedings of the Hypertext 2008 Workshop on Collaboration andCollective intelligence, Pittsburgh, PA, USA. WebScience '08. ACM, New York, NY, 35-39. Retrieved from http://doi.acm.org/10.1145/1379157.1379166

Wells, D. (1996). Groupware & Collaboration Support. OBJS. Object Services and Consulting, Inc. Retrieved from http://www.objs.com/survey/groupware.htm

Wiegers, K.E. (2003). Software requirements : practical techniques for gathering and managingrequirements throughout the product development cycle. [e-Book] s.l. : Redmond, Wash. :Microsoft Press, 2003. Available at: The John Rylands University Library Catalogue. Retrieved from http://www.netlibrary.com/Reader/

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

-

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Address of the Editorial Office:

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005

Notes: