

Secure Data Storage in Cloud Environment using MAS

T. P. Latchoumi^{1,2*} and Latha Parthiban³

¹Department of Computer Science and Engineering, Sathyabama University, Chennai - 600119, Tamil Nadu, India; lax.senthil@gmail.com

²Department of Computer Science and Engineering, Vignan's University, Vadlamudi, Guntur – 522213, Andhra Pradesh, India

³Department of Computer Science and Engineering, Community College, Pondicherry University, Pondicherry, – 605014, Tamil Nadu, India; lathaparthiban@yahoo.com

Abstract

Objectives: The objective is to introduce multi-agent systems to enhance the security rules through the access right to build a distributed warehouse in the cloud environment secure manner. The work aim is to take a smart decision making using MAS in a timely manner. **Analysis:** Cloud computing is a very powerful, and predictable in computing infrastructure for implementing complex agent-based applications. Cloud computing in Multi-Agent System (MAS) appears as an approach to current challenges in many areas. The distributed data warehousing is used to do based on how data's are distributed in the multiple servers. Its main problem show the Collaborative work of a multi-agent system designed for distributed data warehousing in the cloud environment. **Findings:** In the existing approach, a major issue is to retrieve the relevant data from the cloud storage is a very difficult task. Other issues in attributes are increasing the network loads (traffics) and response time, the need of secure data storage, data retrieval from the cloud environment and the database updating is very slow. Cloud infrastructures provide a platform to run the MAS in the real-time, because it takes large execution time by having a large amount of data processing and dynamic memory. **Improvement:** In this paper, we introduce data warehouse in the cloud computing through the multi-agent system technology. This enables cost and time saving. The technique based on data warehouse in the cloud environment using Multi-Agent Systems (MAS) technology to consider security and privacy in data storage and transmitted. The proposed system is the use of an MAS in the cloud environment, introduces autonomous decision making in the critical situation to speed up the execution time, response time, database updating and security enhancing. We apply this propose system in any application like e-banking, hospital management, election department, etc.

Keywords: Cloud Computing, Distributed Data Warehousing, Multi Agent Systems (MAS), Query Redirection Process, Security

1. Introduction

Our everyday activities we create a huge amount of data. A data warehouse is mainly designed for query processing and analysis. The data warehouse contains historical data's that extracts from other sources and derived from transaction data. Data shared across multiple data repositories. The individual Data Marts are built, managed, and maintained in the distributed data warehousing.

The cloud computing model enables secure computing resources, convenient that can be released with minimal

effort and service provider interaction. Normally, it stores and access the applications. It provides elastic services with high performance and large scalable data storage and increasing the number of users.

Some of the surveys show that some percentage of the Data Warehouses (DWs) are failing to meet the business objectives. Warehousing projects are long-term ones, and most of the requirements cannot be stated from the beginning because of their requirements are looked in real projects. DW applications are difficult to specify the information requirements since decision processes are

*Author for correspondence

structured flexibly, large organizations are poorly shared, managers are guarded jealously, and time to keep pace with evolving business processes are unstable. Decision making often refers to information requirements that does not exist in the expected form, and must be derived from other data sources. Nowadays, distributed data warehouses have more attention in the research use¹.

In addition, most of the distributed queries are normally poor¹ during their performance. Furthermore, to answer the queries related to its own subject area by designing the individual data mart, whereas its response time and network speed are very high.

An overview of distributed warehousing², it describes extraction can be done with the help of the back end tools, before going to load the data into a warehouse it should be cleaned it for that cleaning and loading takes place, and the querying and analysis are done by the front end client tools.

In some researches, several optimized query techniques are proposed. Optimized query techniques can be categorized into two²: Materialized views and indexes of redundant structure. It requires extra storage space. Horizontal partitioning is the non-redundant structure. Second techniques do not require an extra space like the first category.

This paper³ tells that the MAS in distributed data warehousing based on methods by using vertical partitioning.

In this paper⁴, the Writable Comparable interface is implemented by the key classes to facilitate the framework for mapping, reducing and shuffling.

Typically, the processors assigned tables in the range of a round-robin fashion (round-robin partitioning), hashing (hash partitioning), or values (range partitioning)⁴. The main challenge is to build a hashed partitioning using a similar search organization called as a Similarity Hashing (SH) technique. At query time, one partition to access at each level maximally is determined by the SH.

In^{4,5} papers they use the technique called Data Warehouse Striping (DWS). Three key algorithms are considered for the above-mentioned data distribution, respectively⁵.

Challenges related to distributed systems are it is desirable to design the systems that optimize the quality of services within the system and improve the throughput to give the optimal result. Multi-Agent System in distributed system will overcome these challenges like availability of data sources and improve the throughput time.

This paper⁶ describes how it is redirecting the query processing and providing security in the DWH systems. It tells which table is to deliver the effective answer. It may also used to monitor the actual query profile.

This paper⁷ tells how the data warehouse is implemented in the cloud environment. For storing the information and query generation of required data uses the combination of different databases in the data warehouse. The different sources will help us to access the related information, to save downloads and to update it. The responsibility of the cloud environment is to provide a better environment.

These agents are used for automatic storage of data, to collaborate with other agents, splitting operations on the distributed data warehousing^{7,8}. And it also describes how the response time delays. We found many issues in these papers; Difficult in directing queries to the appropriate server, high network load, high latency and the execution time delay, security is needed, databases are not updated regularly and it is not presented in the timely fashion and it stays unavailable for the protracted period, cloud storage has huge traffic data.

For above mentioned problems, we focus on security mechanism for distributed data warehousing based on MAS in the cloud environment.

Distributed DWH is using Multi-Agent System (MAS) provides a more reliable data with good quality of service. It minimizes the cost of communication, data movement and response time.

Cloud computing in DWH improves scalability, elasticity, deployment time, cost reduction and reliability. MAS are used to takes own decision making of transferring the data's from DWH into the cloud to solve security problems. It makes easier, controlling the environment of DWH over the cloud environment hypothetical.

The complexity is in the global schema makes a major problem in distributed DWH and most important issue is retrieving the relevant data from the large storage (cloud) is a very difficult task. The solution to this problem is complex, we implement a multi-agent systems concept in it.

The work aim is to take a smart decision making using MAS in a timely manner. The proposed system is the use of an MAS in the cloud environment, introduces autonomous decision making and critical thinking. To improve this existing approach by implementing the recycling of query process and to provide a security circulation of data for decreasing the network loads and get better execution time, to facilitate the cooperation, communication

between systems and to develop the equivalent execution of the client query. Our study focuses on 1. To develop a self-motivated system can handle the DW without human intervention (redirecting the query processing, storing the data, distribution of data on servers, accessing the data) by the multi-agent system through decision support system. 2. Getting a better response while retrieving data in the cloud storage and 3. providing the network security over the distributed data warehouse on the cloud environment.

This paper, prepared as follows: Sect. 2 gives the detail of distributed data warehousing background, Multi-agent systems and cloud computing in distributed DWH and the decision support systems. In section 3 gives the summary of related workings and discuss the issues associated with the query recycling process topics and secure over the network. In section 4, we express the existing issues, according to the works. During section 5, we illustrate the planned security mechanism in the distributed data warehousing by MAS. Finally, in this section a conclusion is made.

2. Background

2.1 Distributed Data Warehousing

Every organization needs a data warehouse. Data warehouses are a sophisticated breed of systems that would only benefit organizations with a certain degree of IT maturity. While several organizations have effectively deployed the distributed data warehouses and automatic decision supports system solutions that offer big business clients with a right to access in corporately, precise, and reliable information, mainly it ignored the advantages of a warehouse. It⁹ encompasses an absolute endeavor data warehouse. It provides users with access of relevant reports. A small number of criteria supposed to be measured as select the warehouse solution⁹: Real-time data incorporation capabilities, Master-to-Master Replication, and Synchronization check functionality. In a distributed data warehousing, the repository of Meta data is frequently replicated by means of every fragment of the warehouse, and the whole warehouse is administered centrally.

Generally, a data warehouse scheme doesn't need that new data is added; to a certain extent, existing data needs reordering. The ability of key indicator of a successful OLAP application is to provide the needed information. All the OLAP¹⁰ applications, found in divergent functional

areas have subsequent key concepts: Multidimensional view of information, Calculating demanding capabilities, and intelligence. Three concepts must be evaluated for their usability of decision maintains and the association so-called the real data warehousing i.e. implicit data warehousing, ROLAP (Relational OnLine Analytical Processing) and (Multidimensional OnLine Analytical Processing) MOLAP. The difficulty in the global schema remains a major issue in distributed DWH¹⁰. We establish a multi-agent systems concept to solve the complex issues¹¹.

2.2 Cloud Computing and Warehousing

Cloud computing defines as the origin of several aspects like processing, resource storage and service-oriented interface¹¹. NIST defined "It is a model for enabling available resources, convenient, on-demand network access to a shared pool of secured computing resources (e.g., networks, servers, storage, applications, services) that can be released with minimum organizationaleffort or examine provider interaction." Moreover, "Cloud model promotes the availability with five key characteristics, four deployment models and three delivery models."

DWH refers to the grouping of several dissimilar databases are used to store the large number of data's and it generates a query concerning to the necessary information. The warehouse helps the user to access the data, save downloads and data updating viz.

Data warehouse on cloud computing is used to create a recycle and reprocess data, improve the data and updating the data multi-agent system.

2.3 Multi-Agent Systems

Agents contain a social ability¹² so as they can communicate with the clients, structure of resources and further agents as necessary in order to reach its goals and tendencies.

It is an autonomous problem solver i.e. MAS can have an ability to take decisions themselves and performs the actions on it to solve the problems. MAS are often distributed and they have proactive and reactive features which are very useful for taking own decision making. Each agent some limitations and it is decentralized. This limitation provides the security because the agents can't able to observe other surroundings. Characteristics of MAS are not only collaborating with other agents, they have an outside communication to permit new agents to come

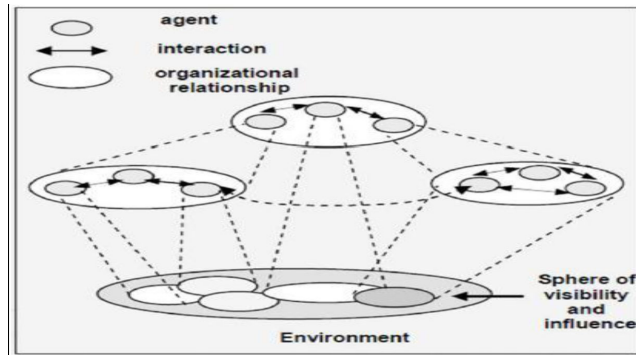


Figure 1. Multi-Agent Systems (MASs).

into the system and communicating with other agents to do the task through the communication protocol. It uses the platform such as JADE2, JACK3. In Figure 1 tells¹³ that how MAS interacts and communicates with each other.

This cleverness enable agents to act both reactively and proactively. The key in to deliver value-added application is to generate more smart software that preserves help out us hand over some of our responsibilities¹³. A MAS has a modular group when it is comprised of several modules, where every module can be supposed to do as a virtual standalone MAS. Normally, the division of the system into modules is complete along dimension such as a require for strong interaction between amongst agents and services in the same module. MAS are useful in increasing the task execution and reducing the communication fixed cost. By using a multi-agent system¹⁴ in the distributed data warehousing, it is doable to ease of management and monitoring the works, and the repairs cost of the warehouse system.

2.4 Multi-Agent Systems in Data Warehousing

Intelligent software agent technology¹⁵ plays a significant position in designing and growth of a data mart for decision support system. A clever agent helps in discovering, locating and reporting the information. When applied agent technology in data warehousing, it allows to improve its presentation, creation of feasible solution to improve searching mechanism by means of self-adaptation abilities into the everyday enterprises¹⁶. By allotting the warehouse behavior of an agent society, it reduces management activities, monitoring costs and preservation cost of a warehousing system. And also it helps to optimize collection, extraction and assimilation process

which is used to recover the information excellence and ensures advanced levels of assurance in decision-making processes¹⁷. Extraction, Transformation, and Load (ETL) method is carried out with ETL agents¹⁵ i.e. spread on the databases. ETL agents identify the relevant sources, extract it, then do the essential data cleansing and transforming. The extracted source is followed by loading the data for use. Decision-maker agent¹⁷ has an understanding of the domain capabilities and to produce the final result. A report/visualization agent contains preformed report and expresses the visual representation of reports. The data guide agent is in charge of caring the track of data storage, and where it is going to be stored.

An Intelligent Agent¹⁸ is a standalone/autonomous entity (program) that can observe the problem and able to react flexibly in a dynamic and uncertain environment for achieving solutions to that problem. In a Multi-agent system contains multiple agents together to interact with each other for solving a particular problem. For successful interaction of agents, they need cooperation, coordination, and negotiation among each other.

3. Roposed Method

To conquer the above mentioned issues, our idea is to treat this existing approach by implementing the recycling method and provide security using multi-agent systems. To satisfy the security access rights for platform and agents to improve the three characteristics of agents adaptively, mobility and flexibility in the data warehouse with cloud computing.

Recycling is the procedure of diverting a request from one data mart to another one (i.e. If the searching data is not in the essential data mart after that, this agent redirects that call to the other data marts automatically until it establish).

Iteration agent is worn to preserve that how many times the request data's are to be redirected. After that we have to confer how our model is proposed to work in distributed data warehousing in Figure 2.

Agent 1: Client Agent- The user utilizes the customer to send the data request and to do the access operations (queries). All client agents have its individual buffer. That buffer is used for searching queries result storage, which is arrived from the query result processor. While in the existing approach, they preserve to send the outcome simply when the client gets connected or else, the searched query result will be put it in the queue until the client agent will

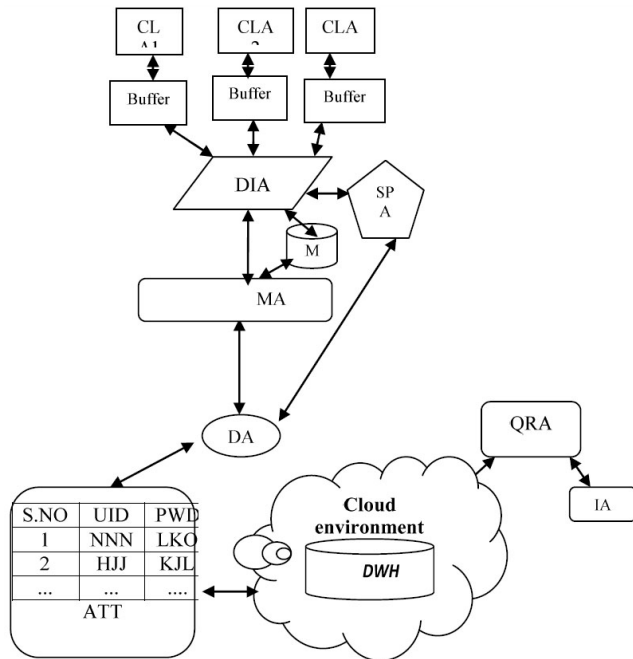


Figure 2. Shows the proposed architecture for distributed data warehousing on the cloud environment using MAS

CLAn-Client Agent, DIA – Dispatcher Agent, MB – MetaBase, DA-Domain Agent, QRA-Query Redirection Agent, IA-Iteration Agent, SPA-Splitting Agent, MA-Messenger Agent, DWH-Data Warehouse, VA-Verifying Agent, ATT – Agent Trust Table

connect again. If the queue gets full after that the result of the query which is arriving from the sender agent will become overflow. Thus we utilize a buffer in our planned approach.

Buffer – It is used for result storage, which is expected by the sender agent. So that it reduces the work of the dispatcher agent. Also, every time the client agent gets connected, then the buffer sends the appropriate results to the clients.

Agent 2: Dispatcher Agent - Arranges the received operations, based on their arrival order. After that the messenger agent sends the results to the dispatcher agents. Later, the result moves to the client agent's buffer.

Agent 3: Splitting Agent -splitting operations take place after the data mart reaches its storage limit.

Agent 4: Messenger Agent - This agent is responsible for executing every operation which is in the waiting queue.

Agent 5: Redirection agent -It plays a vital role. Redirection is the process of requesting data to be redirected from one data mart to another one (i.e. If the pointed query is not in one data mart in that case this agent redirects that request to the other data marts auto-

matically until it establishes). This concept makes decreasing the network loads and speeds up the execution time.

Agent 6: Domain Agent – This agent is in charge of sending the operations to the DW which is in the cloud environment. Then they gather the replies sent by the cloud agent and transmit the final result to the messenger agent.

Agent 7: Iteration Agent – To preserve that how many times it redirects the query.

Agent 8: Security agent provides 2 services

Verifying Agent: It allows only the authorized users request for the search a data's in distributed data warehouse and the result will be sent back to the users again verifying back using this agent. Its static information contains the name and the size of the information that it knows how to transport. Its dynamic information is stored, accessed and secured the data over the network.

Trust Degree Analysis: It maintains trust in agent interaction when communication happens.

While a new user comes to the cloud environment, then it goes to the Security Agent (SA) which authenticates this new user by checking their User_Id (UID) in the trusted third party.

Agent 9: Server Agent -This agent is responsible for received operations and sends back to the domain agent. Its static information consists of the name and the address. Dynamic information is used for storing the received operations from the domain agent.

In the cloud environment, we have agent layers which are responsible for receiving the request which is coming from the dispatcher agent and another queue is responsible for sending and receiving the response from the cloud through the cloud agent. It is split into

- Analysis Agent – Data analyzer and updating
- Controller Agent – Center machine
- Monitor Agent - Monitoring role

3.1 Algorithm for Checking the New user

Input: New user

Output: Adding or removal of new user

1. First new user enters into the cloud environment
2. New users go to the Security Agent (SA)
3. SA checks User_id (UID) in Agent Trust Table (ATT)
4. If it is not available, then SA authenticates the new user by using the threshold value
5. If the threshold value of the new user is valid then provide new UID and password.

6. SA enters the new UID into the ATT.
7. If the threshold value of the new user is not valid then discard the new user.
8. End

3.2 Statistical Analysis

In Figure 3, the graph shows that traditional request time Vs data size per request.

In Distributed Data warehouse using Multi- Agent System (MAS) provides a more reliable data with good quality of service.

In Figure 4, the graph shows that traditional response time Vs number of users.

It minimizes the cost of communication, data movement and response time

4. Results and Discussion

The registration can be done by using Security Agent SA and give UID for each registered client shown in Figure 5. The pie chart shows the security of the system. Security is high in the MAS in DWH using cloud environment when compared with other 2 technologies is shown in Figure 6. By using that IDs Security Agent check during login into the Data Warehouse shown in Figure 7.

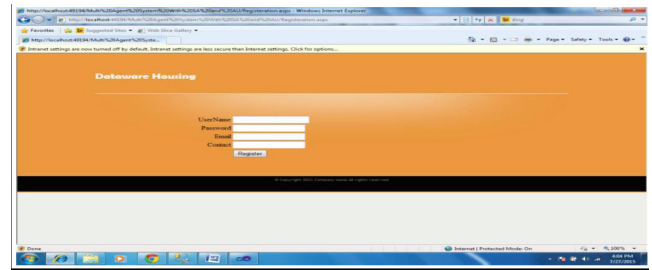


Figure 5. Client Agent should register into the data warehouse.

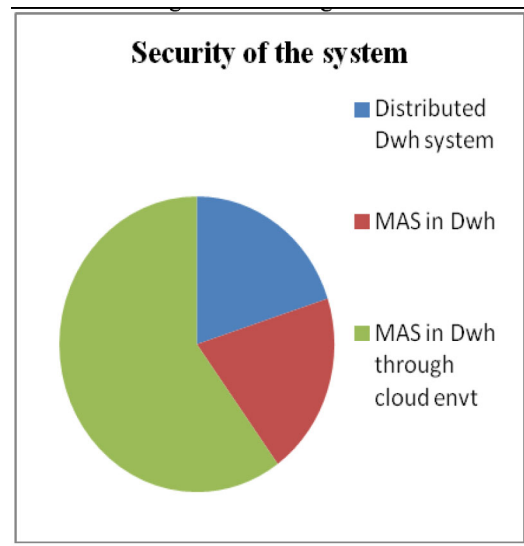


Figure 6. Security of the system compared in 3 technologies.

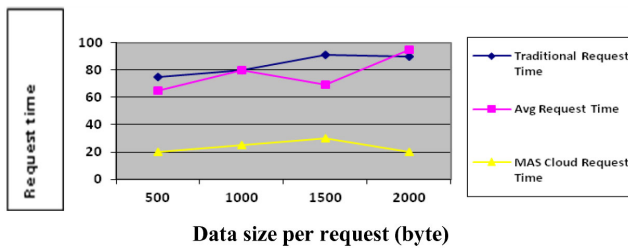


Figure 3. Average request time using MAS in cloud environment.

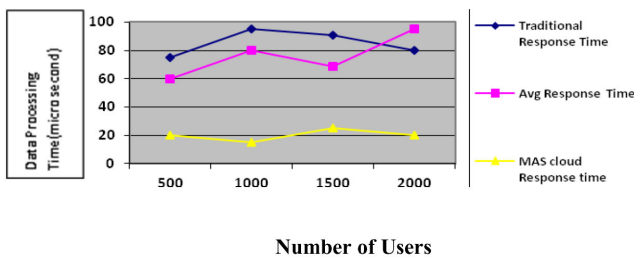


Figure 4. Average Response Time using MAS in cloud environment.

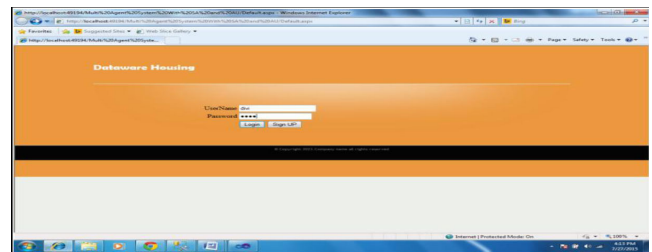


Figure 7. Client Agent login the data warehouse.

id	username	password	emailid	contact
1	apple	apple	NULL	NULL
2	divi	divi	lax.sureth@gmail.com	8754830690
3	surea	surea	surea@gmail.com	8056595669
4	lax	lax	lax@gmail.com	9876534210
5	surea	surea	surea@gmail.com	8056595669
6	lakshmi	sural	lkj.sur	8909873241
7	divi	divi	ldiaye	7890986373
8	eswat	eswat	lkntrapeir	2345768903

Figure 8. Agent Trust Table ATT contains user details.

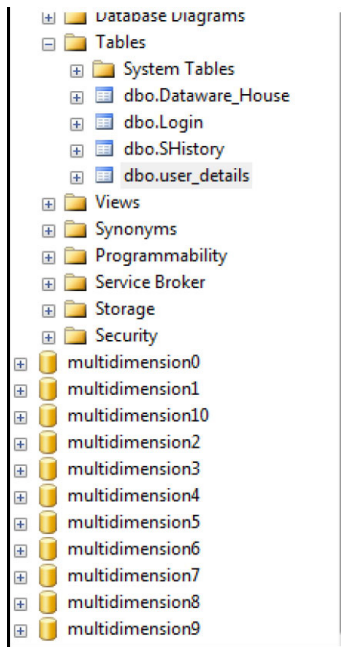


Figure 9. 11 Data Marts using in this proposed work.

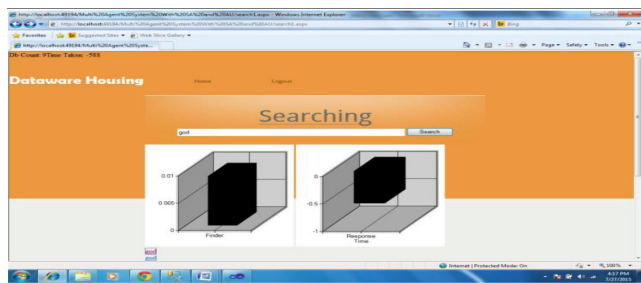


Figure 10. Searching the 11 Data Marts and produce the result with response time quickly.

Table 4.1. Comparison work of Attributes and Technology

Technology \ Attribute	MAS in DWH	DWH in Cloud Environment	DWH in Cloud Environment using MAS
Response Time	Slow	Very Slow	Very Fast
Network Load	Very High	Low	Very Low
Security	Very Low	Low	Very High
Latency	Slow	Very Slow	Very Fast
Execution Time	Slow	Very Slow	Very Fast
Data Base Updation	Fast	Very Fast	Very Fast

Agent Trust Table ATT contains user details of Agent ID, username, password, email and contact numbers of users shown in Figure 8. In Figure 9, shows 11 Data Marts are used in the proposed work. If the 11 Data Marts exceed its size limit, then new Data Mart will be created.

In Figure 10, Searching takes place in 11 Data Marts and produce the result with response time quickly.

A comparison study onMAS and cloud computing in Data warehouse is shown in Table 4.1.

This corresponding table shows the comparison of attributes like Response time, Network load (traffic), Security, latency, Execution time, Database update) and technology.

5. Conclusion and Future Work

Currently, the data warehouse is used as the managerial repository to carry big business conclusion making. Frequently the data warehouse systems utilize a centralized approach in advance. The overview of the agents incorporates in the data warehousing were discussed. We have presented some research work that deals with the data warehousing context with the multi-agent systems. We described at this time is redirecting the query using Multi-Agent System in distributed data warehousing and providing the secure network over the system. In this paper, we include proposed solutions to hold the security problems in the cloud environment. There is an opportunity that malevolent agent can engage in the method and send the incorrect service to the user.

The future enhancement of this system can be implemented Distributed Data warehousing in the cloud environment, implementing of all the agents as being specified in the proposed architecture. And also to implement Total numbers of queries are searched by the user in a day, in a week and the last three months, to calculate how many times the same query has been searched, to find out the word length, More than one user can submit the query simultaneously, Give the input as a string instead of using the single word, To add Map reduces techniques and Data garage in distributed data warehousing in a cloud environment. We are still doing research and reading so many survey papers improving the performance and strengthen the work of our system.

6. References

1. Kumaravel A, Sudha M. Performance comparison based on attribute selection tools for data mining. *Indian Journal of Science and Technology*. 2014 Nov; 7(S7). Doi: 10.17485/ijst/2014/v7iS7/60459.
2. Bernardino J, Furtado PS, Madeira HC. Approximate query answering using data warehousing striping. *Journal of Intelligent Information Systems*. 2002; 19(2):145–67.
3. Brintha Rajakumari S. Data Quality Mining in Electronic News Paper. *Indian Journal of Science and Technology*. 2014 Jun; 7(S5). Doi:10.17485/ijst/2014/v7iS5/50403.
4. Park C-S, Kim M H, Lee Y-J. Finding an efficient rewriting of OLAP queries using materialized views in data warehouses. *Decision Support Systems*. Elsevier Science B.V. 2002; 32:379– 99.
5. Brintha Rajakumari S, Nalini C. An efficient data mining dataset preparation using aggregation in relational database. *Indian Journal of Science and Technology*. 2014 Jun; 7(S5). Doi:10.17485/ijst/2014/v7iS5/50381.
6. Madeira H, Almeida R , Vieira J , Vieira M , Bernardino J. Efficient Data Distribution for DWS, CISUC, Dept of information Engineering, Univ of Coimbra, Coimbra, Portugal, 2002.
7. Kezunovic M, Popovic T. Data Warehouse and Analysis Agents Mladen Kezunovic is with Texas A&M University, College Station, Texas, USA (e-mail: kezunov@ece.tamu.edu) Tomo Popovic is with the Test Laboratories International, Inc., College Station, Texas, USA (e-mail: tomo@tli-inc.com)
8. Kolsi N, Abdellatif A, Ghedira K. Data warehousing access using Multi agent systems. *Distributed Parallel databases*, Springer Science. 2009 Feb; 25:29–45.
9. Somu N, Gangaa A, Shankar Sriram VS. Authentication service in hadoop using one time pad. *Indian Journal of Science and Technology*. 2014 Apr; 7(S4). Doi: 10.17485/ijst/2014/v7i4/50062.
10. John R, Saravanan V. Vertical partitioning in object oriented databases using intelligent agents. *IJCSNS International Journal of Computer Science and Network Security*. 2008 Oct; 8(10).
11. Manickam R, Boominath D, Bhuvanewari V. An analysis of data mining: past, present and future. *International journal of Computer Engineering and Technology (IJCET)*. 2012; 3(1):1–9. ISSN Print: 0976 – 6367, ISSN Online: 0976 – 6375.
12. Verma. Data warehousing on cloud computing. *International Journal of Advanced Research in Computer Engineering and Technology (IJARCET)*. 2013 Feb; 2(2).
13. Zhang W-R. Concepts, challenges, and prospects on Multiagent Data Warehousing (MADWH) and Multiagent Data Mining (MADM), Department of Computer Science, Georgia Southern University, USAE-mail: wrzhang@georgiasouthern.edu. *Int J Intelligent Information and Database Systems*. 2008; 2(1).
14. Shihab Al_Doori MM, AL_Obaidy AT. The future for adaptive software development in cloud computing environment using multi agent system. Received on: 23/4/2014 & Accepted on: 4/12/2014. 2015; 33:B, No.1.
15. Lobo C, Smyl S, Nath S. Data Garage: Warehousing Massive Performance Data on Commodity Servers, Microsoft Corporation, proceedings of the VLDB Endowment. 3(2). Copyright 2010 VLDB Endowment 21508097/10/09.
16. Datta S, Bhaduri K, Giannella C, Wolff R, Kargupta H. Distributed Data Mining in Peer-to-Peer Networks. *IEEE Internet Computing*. 2006 July/Aug; 10(4):18–26. Doi:10.1109/MIC.2006.74.
17. Boussaid O, Bentayeb F, Darmont J. A MAS-Based ETL Approach for Complex Data ERIC/BDD, Universite Lumiere Lyon 2 5 avenue ierre-Mendes-France, Bron Cedex FRANCE hal-00321977, 2008 Sept 16, version 1,
18. Available from: <http://data-warehouses.net/architecture/index.html> Data accessed: 03/08/2015