International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 7 (2015) pp. 17469-17478

http://www.ripublication.com

Parallel Processing Agents For Data Mining With Cloud Computing & Multi-Agent Systems

Annie Sheryl S
[1]*PG student of DMI College of Engineering*
J.P Ananth
[2]*Professor DMI College of Engineering*
Department of Computer Science and Faculty of Computer Science,
DMI College of Engineering, Palanchur, Chennai, India.

Abstract

The cloud computing storage approaches have been increasingly adopted by several expertise's from various domains. The cloud computing system interacts with various systems and these services are offered from data centers all over the world. The Multi-Agent system (MAS) includes various data agent that interact with the Cloud Data Storage (CDS) which in turn interacts with the Data Mining Agents .The Data Mining Agents analyses data from different perspectives and summarizes it into a useful information . The main process in the distributed environment providing flexible and optimized access for data are, 1) the distributed data mining provides the in depth information of the data and provides interaction with various processing agents in parallel. 2) The Multi Agent system provides various source of information and also provides flexibility of communication within the agents using Java Agent Development Environment (JADE). 3) The distributed data mining systems provides improved access to the data with the analytical view of the data stored in Cloud Data Storage.

Key words: Cloud Computing; Cloud Data Storage; Security; Multi Agent System; Cloud Service Provider; Data Mining.

Introduction

The Cloud computing process provides the area of sharing or accessing resources irrespective of their location transparency. Based on certain protocols the cloud resources are accessed using the internet for communication. The access remains active on demand and provisioned by cloud provider. The Cloud Data Storage (CDS) is widely used for the storing data from various agents.

Cloud storage is storage where the digital data is stored in logical pools, the physical storage spans multiple servers and the physical environment is typically owned and managed by hosting client computers. These cloud storage providers are responsible for keeping the data available and accessible, and the physical environment protected and running. People and organizations buy or lease storage capacity from the providers to store user, organization, or data application. The major cloud architecture includes cloud services, cloud platform, cloud infrastructure, cloud storage. Cloud Data Storage (CDS) is composed of thousands of cloud storage devices clustered by network, distributed file systems and other storage middleware to provide cloud storage service for users. CDS provide cloud storage resources for all kinds of clients, and the fee can be based on CDS capacity or CDS bandwidth periodically.

Multi-agent systems (MAS) are heterogeneous and distributed software architecture, and that's made it the best solution for design, develop and implement the software platform for the distributed architectures. Multi-Agent System (MAS) stores its various data in the cloud data storage that is the clean and the extracted data from the data marts of various agents are stored in the CDS. The Data mining process provides a discovery of hidden values from the data stored in the cloud data storage. The preprocessed data is acquired by removing unwanted variables, and then analyze the useful information that can be used to represent the data.

Data mining is a computational process of discovering patterns in large data sets. The main aim of data mining process is to extract information and to transform it into understandable structure for future use. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records, unusual records and dependencies. The patterns are seen as a kind of summary of the input data, and used in further analysis. The analysis process includes in picking up the data that is available in the cloud, and brings out intelligence from the data in the required format. These formats are generally patterns that are discovered for business perspective.

The general classes on which the data are selected from the data base are clustering, classification, regression and summarization. Data mining can only uncover patterns actually present in the data, the target data set must be large enough to contain patterns while remaining concise enough to be mined within an acceptable time limit The general source of data for the data mining process is from data marts or preprocessed data store.

The main contributions of the paper are organized as follows:

1. Multi-Agent Systems (MASs) Development
2. Cloud Data Storage (CDSs) Development
3. Distributed Data Mining (DDMs) Development

Multi-Agent Systems (MASs) Development

A data agent is a piece of software that acts autonomously on behalf of human users to perform some set of tasks. Most advanced applications of agents, including the one discussed in this paper, employ "intelligent" software agents, which are not only autonomous but also reactive, proactive, and capable of interacting with each other in

a flexible manner. We decompose a system in multiple interacting and autonomous agents that have a defined set of objectives to achieve. Unlike the ill-defined linkages and privileges among Subsystems, the agents' inter-communication is significant.

Cloud Data Storage (CDSs) Development

Most systems plan to utilize cloud storage to augment their existing file storage infrastructure. Files in active use will be stored on local systems, while files that are unlikely to be accessed will be stored in the cloud. However, users and applications must be able to access files stored on local systems and in the cloud. Theses cloud data storage will help for storing data in the cloud systems and those data are generally used for data analysis. Performance for outsourced storage is likely to be lower than local storage but the data stored for analysis purpose requires a regular load process rather than frequent access. Reliability and availability depends on wide area network availability and on the level of precautions taken by the service provider. The publically available cloud is generally used to store and retrieve data in organization wide.

Distributed Data Mining (DDMs) Development

The distributed data mining is a process of analyzing large datasets for extracting intersecting patterns. The mining process is done in data sources. DDM is widely seen as a means of addressing the scalability issue of mining large data sets. The task of distributed data mining is to extract and integrate knowledge from different sources. Solving such tasks requires a special approach and tools, different from those applied to learning from data located in a single database. One of the approaches suitable for the DDM is to select relevant local patterns from the distributed databases. Such patterns, often called prototypes, are merged to create a compact representation of the data repositories.

Narrative Study

The major difference among various systems is that the proposed system provides the integration of cloud data storage and multi agent system with distributed data mining. The parallel processing systems are not available widely. Cloud computing has become a popular high performance computing model where resources are provided as services over the Web. In order to isolate users from the complexity of cloud programming, a middleware is used where these systems exhibits easier access and not an optimized access. This existing system provides the cloud structure for data storage and often failed to provide parallel processing systems.

Multi-Agent system (MAS) is a system composed of a lot of Agents, and generally these Agents exchange information through network infrastructure. Management Agent It receive user's request from the user's graphic interface, then look over whether there is user's information in Agent information storehouse. User Agent is responsible for communicating with management Agent and receiving and passing user's order, to the mining Agent, finally analyses and coordinates the mining result

transmitted by mining Agent to upgrade user's model storehouse and preset information storehouse and submit to the management Agent at the same time.

The internal agent communication is available for agent interaction Knowledge Query Manipulation Language (KQML) is an agent communication language based on speech acts, which is used widely at present. KQML is both a message format and a message handling protocol proposed to support run-time knowledge sharing among agents. The core of KQML is a group of extensional per formatives defining permissible speech acts of knowledge base and objective base among agents. Top model such as contract nets and consultation mechanism can be established based on KQML per formatives. Moreover, KQML language provides a basic architecture of knowledge sharing through a special class of agents called communication facilitators which coordinate the interactions with other agents.

Distributed Data Mining aims at extraction useful pattern from distributed heterogeneous data bases in order, for example, to compose them within a distributed knowledge base and use for the purposes of decision making. A lot of modern applications fall into the category of systems that need DDM supporting distributed decision making. Applications can be of different natures and from different scopes Most of the system fails to provide a distributed data mining environment. These include information retrieval, user interface design, robotics, electronic commerce computer mediated collaboration, computer games, education and training ,smart environments, ubiquitous computers and social simulation Individual tasks can be completed by individual agent and, if necessary, through communication can work together to complete certain global tasks. When considered for data mining system exist with heterogeneous data to be analyzed. . The use of multi-agent-systems creates a framework which allows the inter-operation of a vast set of heterogeneous solutions to carry out the complex supply chain finance context-aware distributed data mining tasks, many data mining tasks to connect heterogeneous resources, as data sources, processing nodes and end user applications. It considers a data warehousing that supports context-aware OLAP queries, ensuring the interoperability of all data sources, and then focuses on distributed clustering algorithms and some potential applications in multi-agent-based problem solving scenarios.

Overview of the Proposed Work

MAS consist of various agents and their virtual environment and it also provides the self organized system. The agents are defined as location based agent. These agents communicate with one another using KQML (Knowledge Query and Manipulation Language).The agents store the data in Cloud Data Storage (CDS). These summarized data are gathered for mining process.

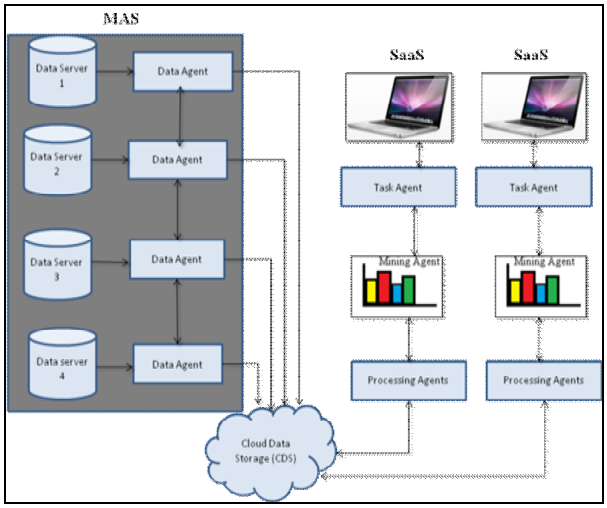

Figure 3.1: System Architecture

The processing agent retrieve the data from CDS execute in parallel. These processing agents provide the optimized accessing of data's from cloud storage. The data mining agents interact with its own processing agents and choose the data mining algorithm to discover hidden patterns and used in DSS. The hidden patterns are the analyzed data from heterogeneous databases. The mining agents act as instructed by the task agent in choosing the data for analysis. The business view of data remains helpful for expecting future predictive systems. These analysis help in providing a decision support system finally.

Multi Agent System

Implementation is the stage of the project when the theoretical design is turned out into a working system. The working system includes the multi agent system development and interaction with its corresponding data agent. The data agents provide the local storage of data in the data base. Location based multi agents are developed. Multi agent system composed of multiple interacting intelligent agents. A generalized login is available for all the agents from a location. Once after the agent login is done the system prompts for the agent selection window. Individual location based agent admin is provided and each agent is given a unique agent ID. The agent ID helps in identifying every individual agent and helps during agent communication. The system allows every individual agent to enter data on daily basis. Theses day to day information is loaded into the database

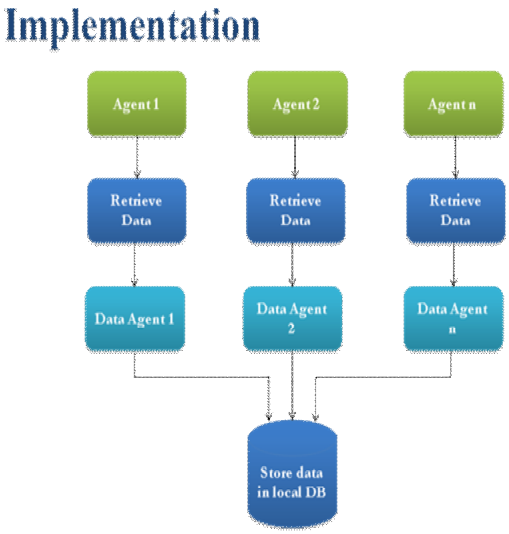

Figure 4.1: Multi Agent System Architecture

The MAS is a set of agents that interact together in one society to resolve a common problem by using the resources and the knowledge of each agent. MAS' architecture is a structure that depicts the various families of agents and their relationships. A configuration is the instantiation of architecture with a chosen combination and an appropriate number of agents of each type. For a given architecture, we can generate several configurations, and a given configuration is closely related to the topography and context of where it is deployed (organizational structure, characteristics of the Intranet, location of interest, etc. . .). Thus, the architecture must be designed so that all possible configurations cover the different organizational contexts possible.

Data Agent: The data agent will connect to the data sources and pre process and clean data are stored in the local database. The data agent remain exist individually for every data sources. The number of data agent for each task is the number of the data sources.

Agent Communication

The working system includes the multi agent based on each location. Multi agent system composed of multiple interacting intelligent agents. The agent communication happens through Agent Communication Language. These agents interact with each other to check if the data for each location for certain number of days were entered or not. JADE provides a platform for communicating between agents either on same or different platforms. These agents are placed under the agent containers. The agent containers include one java process per container and remain transparent to every individual agent. The communication happens through messages. Platform routes messages between agents.

Agent communication happens through ACL (Agent Communication Language). The messages sent across agent include Sender, Receiver, Content required for

communication. The sniffer agent communication shows how the agent interacts with all the agents present in the main container.

Cloud Data Storage

The cloud data storage provides a working environment for storing data that are gathered by the data agents.

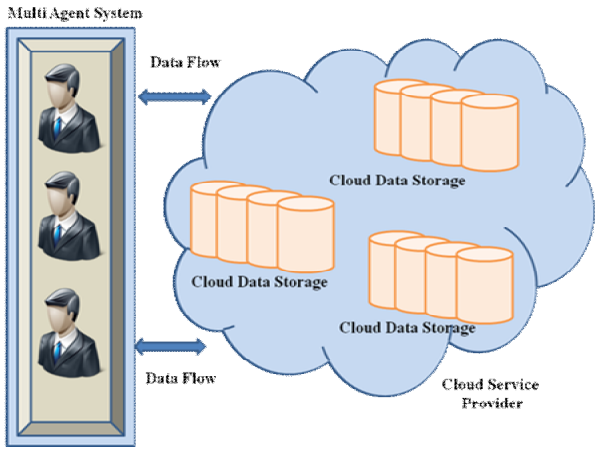

Figure 6.1: Cloud Data Storage

The data agents provide the data that are entered by the individual agents respective to their locations. The data are stored in the cloud storage space that is provided by the service providers. These services are publically available services that are maintained individually by the organizations in order to avail the storage space irrespective to the user. The cloud services are provided whenever they are required by the system for data storage.

Data Mining

The mining agent is composed of agents that analyses the data that are stored in the cloud data store. . This level is composed of queries and reports generating tools, dynamic analysis tools. Business Intelligence portals hold the most important position in creating specialized, flexible, user-friendly and accessible interfaces, allowing users a good end-to-end experience, nice graphical appearance, report integration options. The decision support system provides a business required consolidation of production and prediction system

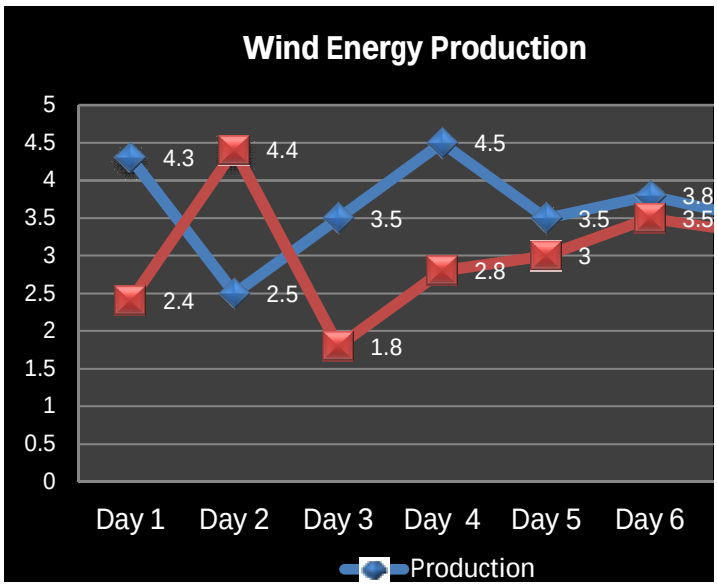

Figure 7.1: Predictive System For Wind Energy

Results

1. Web based application:

RRB application based system provides an idea or knowledge about the entire system. Decision support system provides a business perspective output from analyzed data.

2. Generalized Login:

A generalized login is provided for various location based data agents. This login provides the overall admin facility for the system.

3. MultiAgent Selection Window:

The generalized login redirects the pages to the multi agent selection window. The multi agent selection window displays all the agents that are provided in the system. The system contains all the location based agents. These agents provide a individual login.

4. Agent1 login: Palladam

Every individual agent provides a agent ID which is Unique. For Agent palladam a separate login is given along with ID 001.

5. Agent2 login: Poonamallee

Every individual agent provides a agent ID which is Unique. For Agent poonamallee a separate login is given along with ID 002.

6. Agent3 login: Tenkasi
Every individual agent provides a agent ID which is Unique. For Agent Tenkasi a separate login is given along with ID 003.

7. Agent4 login: Thirunelvelli
Every individual agent provides a agent ID which is Unique. For Agent Thirunelvelli a separate login is given along with ID 004.

8. Data Entry:
The data based on date are collected and stored temporarily in the database. For every agent a Agent ID is provided by default, Employee name, Date for which the value is entered

9. Data Retrieval:
The data entered for 24 hours based on date are stored in database for further load process that happens to the Cloud Data Storage (CDS)

10. Data Mining:
The Mining process happens based on the analysis of data for particular period of time. The decision support system provides a predictive analysis of data.

Conclusion & Future Works

The main motive of this paper is to present a combined service of multi agent system and cloud data storage in order to provide a knowledge Discovery in Database. It is specially designed for agents who interact with each other frequently for setting up the business standards. The Cloud storage service is provided over the internet. . The contributions of this work include: 1) Constructing the Multi agent system 2) Designing the warehouse architecture in cloud data storage 3) Processing agent retrieve data from the cloud and producing a mining output. It is the first attempt in building an agent-based system that gives a decision support system when retrieving information about Cloud services. The inbuilt security features for the system can be implemented as a part of enhancement.

Bibliography Reference

[1] Xiangyu Luo, Mengmeng Zou, Lingjie Luo: A modeling and verification method to multi-agent systems based on KQML , 2013 IEEE International Conferences on Web Intelligence (WI).
[2] Prospects. Architectural issues of Webenabled electronic business pages 113-128 Hershey, PA : Idea Group Pub., c2003, ISBN: 1591400813.
[3] Hellol Kargrupta, Ilker Hamzaoglu, Brian Stafford. Scalable, Distributed Data Mining Using An Agent Based Architecture. Proceedings the Third

International Conference on the Knowledge Discovery and Data Mining, AAAI Press, Menlo Park, California (1997).

[4] Vladimir Gorodetsky, Oleg Karsaeyv, Vladimir Samoilov. Software Tool For Agent-Based Distributed Data Mining. Sequence diagram of the agents' interaction 115 Conference on Computational Intelligence and Multimedia Applications 2007 International Conference on (2007) Volume: 2, Publisher: IEEE, Pages: 18-24.

[5] Octavio Gutierrez-Garcia, Kwang-Mong Sim. Self-Organizing Agents for Service Composition in Cloud Computing. CLOUDCOM '10 Proceedings of the 2010 IEEE Second International Conference on Cloud Computing Technology and Science.

[6] Fabien Gandon, Rose Dieng-Kuntz, Olivier Corby, Alain Giboin. Web Sémantique et Approche Multi-Agents pour la Gestion d'une Mémoire Organisationnelle Distribuée. INRIA - Projet ACACIA, 2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis, FRANCE.

[7] Amir Mohamed Talib, Rodziah Atan, Rusli Abdullah, Masrah Azrifah Azmi Murad. CloudZone: Towards an integrity layer of cloud data storage based on multi agent system architecture. Open Systems (ICOS), 2011 IEEE Conference: Pages 127 – 132.

[8] J.O. Gutierrez-Garcia and K.M. Sim, "A Family of Heuristics for Agent-based Elastic Cloud Bag-of-Tasks Concurrent Scheduling," Future Generation Computer Systems, doi: 10.1016/j.future.2012.01.005 (published online: 7 February 2012).

[9] J.O. Gutierrez-Garcia and K.M. Sim, "GAbased Cloud Resource Estimation for Agent-based Execution of Bag-of-tasks Applications," Information Systems Frontiers, doi: 10.1007/s10796-011-9327-8 (published online: 16 September 2011).

[10] Myoungjin Kim, Hanku Lee, Hyogun Yoon, Jee-In Kim, HyungSeok Kim, IMAV: An Intelligent Multi-Agent Model Based on Cloud Computing for Resource Virtualization, ICIEE 2011, May 28-29, IPCSIT Vol.6, pp. 199-203, ISBN 978-981-08-8637-0.