Working in Different Platforms of Cloud for the Online Signal and Trajectory Simplification

Chintamani Hindurao Kamble[1]
Department of Computer Engineering
JSPM, NTC, Rajarshi Shahu School of Engineering & Research,
Narhe, Pune, India

Prof. Vilas S. Gaikwad[2]
Department of Computer Engineering
JSPM, NTC, Rajarshi Shahu School of Engineering &
Research, Narhe, Pune, India

Abstract: This system describes how Complex Database in data mining is used in cloud computing. Data Mining is used for extracting potentially useful information from raw data. The integration of data mining techniques into normal day-to-day activities has become common place. mass data storage and distribution calculation of cloud computing, it provides a new method for data mining, effectively solving the problems of distribution of mass data mining and efficient storage computing. The cloud computing model brings many benefits and convenience. This paper introduces the cloud computing and data mining and then simply introduce some existing parallel data mining algorithms based on cloud computing and data mining service platforms. Finally it gives a simple description of the problems and prospects of data mining based on cloud computing. Every day people are confronted with targeted advertising, and data mining techniques help businesses to become more efficient by reducing costs. Data mining techniques and applications are very much needed in the cloud computing paradigm. The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage.

Key words: Cloud computing; Cloud Databases; Database architectures, trajectories, Ontology.

I. INTRODUCTION

Cloud computing (CC) becomes a natural and ideal choice for such organizations and customers. Cloud computing takes benefit of many technologies such as server consolidation, huge and faster storage, grid computing, virtualization, N-tier architecture and robust networks. It delivers highly scalable and expensive infrastructure with minimal set up and negligible maintenance cost. It provides IT-related services such as Software-as-a-Service, Development Platforms-as-a-Service and Infrastructure-as a-Service over the network on-demand anytime from anywhere on the basis of "pay-as-you-go" model. It is a fast growing concept changing the IT related perceptions of its users. Elasticity, scalability, high availability, price per-usage and multi-tenancy are the main features of Cloud computing. It reduces the cost of using expensive resources at the provider's end due to economies of scale. Quick provisioning and immediate deployment of latest Big Data is very beneficial when it comes to organization. It helps them to compete more effectively with other organization, better understanding of the customer, grow business revenue, and others. It has been found giving effective results in the field of: retail, banking, insurance, government, natural resource, healthcare, manufacturing, public sector administration, personal data location and other services. Applications at lesser cost are the benefits which force people to adopt Cloud computing. Cloud computing has brought a paradigm shift not in the technology landscape, but also in the database landscape. With more usage of Cloud computing, demand for provisioning of database services has raised. Provisioning of Cloud databases is known as Database-as-a-Service in Cloud terminology. The main objective of the paper is to explore the trends in Cloud databases and analyze the potential challenges to develop these databases. Data have hidden information in them and to extract this new information; interrelationship among the data has to be

achieved. Information may be retrieved from a hidden or a complex data set. Browsing through a large data set would be difficult and time consuming. [12][5]

II. RELATED WORK

1. Cloud Databases: Data mining environment produces a large amount of data that need to be analyzed; patterns have to be extracted from that to gain knowledge. In this new era with boom of data both structured and unstructured, in the field of genomics, meteorology, biology, environmental research and many others, it has become difficult to process, manage and analyse patterns using traditional databases and architectures. So, a proper architecture should be understood to gain knowledge about the Big Data. This paper presents a review of various algorithms from 1990-2013 necessary for handling such large data set. These algorithms define various structures and methods implemented to handle Big Data, also in the paper are listed various tool that were developed for analysing them.

2. A Comparative Study of Relational Databases and NoSQL Databases: In the earlier stages of computerization, there was more demand for transaction processing applications. As the database industry matured and people accepted computers as part and parcel of their lives, analytical applications became the focus of enterprises. Now they wanted to store they have emerged to address the requirements of data management in the cloud as they follow BASE (Basically Available, Soft state, eventually consistent) in contrast to the ACID guarantees. So, they are not suitable for update intensive transaction applications. They provide high availability at the cost of consistency Different soft computing tools can be used in different phases of Medical Data Handling Using Soft Computing. i.e. denoising, classification, clustering, filtering, searching, matching, customized searching and filtering. Soft computing (SC) is not a new term; we have gotten used to reading and hearing about it daily. Nowadays, the term is used often in computer science and information technology. It is possible to define SC in different ways. Nonetheless, SC is a consortium of methodologies which works synergistically and provides, in one form or another, flexible information processing capability for handling real life ambiguous situations. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness and low-cost solutions. SC includes fuzzy logic (FL), neural networks (NNs), and genetic algorithm (GA).methodologies. SC combines these methodologies as FL and NN (FL-NN), NN and GA. Among the large number of computational techniques used, SC, which incorporates neural networks, evolutionary computation, and fuzzy systems, provides unmatched utility because of its remonstrated strength in handling imprecise information and providing novel solutions to hard problems.
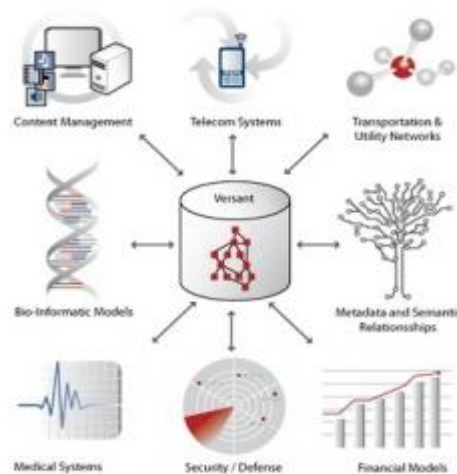

Fig 2: A Comparative Study of Relational Databases and NoSQL Databases

A database accessible to clients from the cloud and delivered to users on demand via the Internet from a cloud database provider's servers. Also referred to as Database-as-a-Service (DBaaS), cloud databases can use cloud computing to achieve optimized scaling, high availability, multi-tenancy and effective resource allocation. While a cloud database can be a traditional database such as a MySQL or SQL Server database that has been adopted for cloud use, a native cloud database such as Xeround's MySQL Cloud database tends to better equipped to optimally use cloud resources and to guarantee scalability as well

as availability and stability. Cloud databases can offer significant advantages over their traditional counterparts, including increased accessibility, automatic failover and fast automated recovery from failures, automated on-the-go scaling, minimal investment and maintenance of in-house hardware, and potentially better performance. At the same time, cloud databases have their share of potential drawbacks, including security and privacy issues as well as the potential loss of or inability to access critical data in the event of a disaster or bankruptcy of the cloud database service provider.

Security and privacy issues as well as the potential loss of or inability to access critical data in the event of a disaster or bankruptcy of the cloud database service provider.

Thus, the basic models of providing CC services are:

• SaaS (software as a service) - a technology platform that allows access to applications via the Internet in the form of services that are hired as needed, instead of buying a separate software programs that must be installed on the user (office and /or home) computers;

•PaaS (Platform as a Service) - model is a variation of SaaS structure that, as a service delivers environment development. Allows the user to build his own applications that run on the provider's infrastructure. Applications are delivered to users through the servers' interface accessible via the Internet.

•IaaS (Infrastructure as a Service) - provides the ability to use computer infrastructure (mainly virtual platforms). Users do not buy servers, software, data storage or network equipment, but they buy these resources as an external service.

3: A Mathematical Model for Database design in cloud Computing.

Database, warehousing and mining the data in database, with an emphasis on their new requirement. That will discuss here bank end tools for managing the data extracting, cleaning and loading data into a data warehouse and front end client tools for querying and data analysis server extensions for processing the efficient query, tools for metadata management and for managing the warehouse. This overview gives us **the** basic knowledge of various database tools and techniques. A data warehouse is a collection of integrated databases designed to support a DSS. It is a collection of integrated, subject-oriented databases designed to support the DSS function, where each unit of data is non-volatile and relevant to some moment in time.

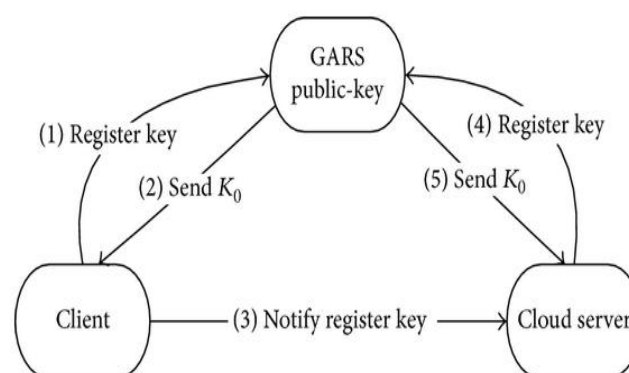**3.1: Optimization Model & Algorithm:**

Fig: Optimization model of cloud computing.

A simple GA is composed of three operations: selection, genetic operation, and replacement. The advantage of this technique is that it can handle a vast search space, applicable to complex objective function and can avoid being trapping into local optimal solution

1. Initial population generation: GA works on fixed bit string representation of individual solution. So, all the possible solutions in the solution space are encoded into binary strings. From this an initial population of ten (10) many chromosomes are selected randomly.

2. Crossover: The objective of this step is to select most of the times the best fitted pair of individuals for crossover. The fitness value of each individual chromosome is calculated using the fitness function.

3. This pool of chromosomes undergoes a random single point crossover, where depending upon the crossover point, the portion lying on one side of crossover site is exchanged with the other side. Thus it generates a new pair of individuals.

4. Mutation: Now a very small value (0.05) is picked up as mutation probability. Depending upon the mutation value the bits of the chromosomes, are toggled from 1 to 0 or 0 to 1. The output of this is a new mating pool ready for crossover.

3.2: Experiment and Results Analyze:
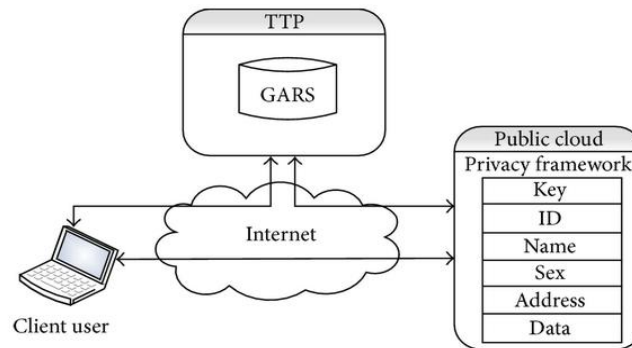
Case Study – Face book

Intensive data mining on Face book is impossible without the use of CC solution. With about 500 million users and an average of one billion page views per day, this most popular social network (and Cloud application) daily generates and accumulates huge amounts of data. One of the biggest challenges, practically from the start, was (and still is) solving the problem of efficient storage, processing and analyzing (mining) of this data.

III. CONCLUSION

Cloud computing is an emerging computing paradigm that is increasingly popular. Leaders in the industry, such as Microsoft, Google, and IBM, have provided their initiatives in promoting cloud computing. However, the public literature that discusses the research issues in cloud computing are still inadequate. In a study of the research literature surrounding cloud computing, I found that there is a distinct focus on the needs of the scientific computing community. By developing cloud based data mining solutions accessing data mining services every time and everywhere and from various platforms and devices will be made possible. Ultimately, the application of CDM solutions can provide a sort of knowledge discovery eco-system built of a large numbers of decentralized data analysis services. Also, a significant moment that should be noted is that the creation and giving the service of data mining in the Cloud, today a critical business activity, which, otherwise, requires significant financial and technical resources, becomes accessible to the less affluent, small and medium-sized companies that have not used so far the advantages.

References

1.   Mell, P. and Grance, T. (2011). The NIST Definition of Cloud Computing (Draft). Recommendations of the National Institute of Standards and Technology, NIST.

2.   Pejić Bach, M. (2005). Data mining in the banking industry. Proceedings of Faculty of Economics, University in Zagreb.

3.   Z. Li, B. Ding, J. Han, and R. Kays, "Swarm: Mining relaxed temporal moving object clusters," VLDB Endow., vol. 3, no. 1, pp. 723–734, Sep. 2010.

4.   G. Rote, "Computing the minimum Hausdorff distance between two point sets on a line under translation," Inform. Process. Lett. vol. 38, no. 3, pp. 123–127, 1991.

5.   P. Joia, D. Coimbra, J. Cuminato, F. Paulovich, and L. Nonato, "Local Affine Multidimensional Projection," IEEE Trans. Visualization and Computer Graphics, vol. 17, no. 12, pp. 2563-2571, Dec. 2011

6.   K. Svanberg, "A Class of Globally Convergent Optimization Methods Based on Conservative Convex Separable Approximations," SIAM J. Optimization, vol. 12,   no. 2, pp. 555-573, 2002.

7.   H. Jeung, H. Shen, and X. Zhou, "Convoy queries in spatiotemporal databases," in Proc. ICDE, Cancun, Mexico, 2008,pp. 1457–1459.

8.   Z. Li, B. Ding, J. Han, and R. Kays, "Swarm: Mining relaxed temporal moving object clusters," VLDB Endow., vol. 3, no. 1, pp. 723–734, Sep. 2010.

9.   G. Rote, "Computing the minimum Hausdorff distance between two point sets on a line under translation," Inform. Process. Lett. vol. 38, no. 3, pp. 123–127, 1991.

10. P. Kalnis, N. Mamoulis, and S. Bakiras, "On discovering moving clusters in spatio-temporal data," in Advances in Spatial and Temporal Databases. Berlin, Germany: Springer, 2005, pp. 364–381.

11. T. Nguyen and J. Zhang, "A Novel Visualization Model for Web Search Results," IEEE Trans. Visualization and Computer Graphics, vol. 12, no. 5, pp. 981-988, Sept./Oct.   2006.

12. A. Spoerri, "RankSpiral: Toward Enhancing Search Results Visualization," Proc. IEEE Symp. Information   Visualization, pp. 208-214, 2004.

13. K. Zheng, Y. Zheng, N. J. Yuan, and S. Shang, "On discovery of gathering patterns from trajectories," in Proc. ICDE, Brisbane, QLD, Australia, 2013, pp. 242–253.