DATA MINING AS A SUPPORT FOR BUSINESS INTELLIGENCE APPLICATIONS TO BIG DATA

Satya Nagendra Prasad Poloju
SAP Business system engineer, Tek-Analytics LLC, USA

ABSTRACT

Cloud Computing plays a big function in the in data mining area of numerous sectors in today's culture. Building the data mining system based upon cloud computing is useful to accomplish effective data mining This paper evaluates the basic architecture of the big data mining platform based on cloud computing and the key technologies for its building on the basis of relevant concepts of cloud computing and also data mining.

I. INTRODUCTION

With the advent of the cloud age and the rapid growth of mobile Internet, China has gotten in the information age with as well big amount of information. According to a survey, By the end of 2014, the quantity of data has surpassed 3 ZB worldwide. Such a huge quantity of data has brought troubles to the application of data mining system, making huge data handling complicated. The computing power of the system can not meet the needs, nor can the computing resources of conventional stand-alone web servers. Thus, it is required to make use of distributed computing modern technology for mass calculation. The emergence of cloud computing makes the big data mining system have a new growth instructions and also makes its building and construction feasible. However, the big data mining system based on cloud computing has not yet been developed to excellence, which calls for constant scientific and also technological study on constructing a brand-new data mining system.

Cloud Computing and Data Mining

Data Mining. Also called data or knowledge discovery, data mining describes searching out prospective and also beneficial data from a multitude of blurry and random real data by computing. Data mining is closely pertaining to computer technology, which is realized through stats, internet analytical handling, information retrieval, machine learning, expert system as well as pattern recognition. Data mining is an essential technology in the field of understanding discovery, certain methods of which mostly include collection, extraction, warehousing, analysis as well as data; it is commonly utilized in locations such as internet, financing, telecoms as well as clinical research at present. Cloud Computing. Cloud computing is a computing approach based upon the Net, which shares software program as well as hardware sources and information to computers and also various other tools. Cloud refers to network, in particular the Web. In operation cloud computing, the individual does not need to recognize the information of the facilities in the "cloud", nor does he/she demand to have matching specialist expertise and also straight control over the whole computing process. Cloud computing primarily has the adhering to qualities. First of all, it achieves vibrant source circulation, various sources department according to the requirements of the user and also boost of readily available resources. Second of all, it realizes personalized services, that is, offer individuals with self-support resource solution, who do not require to communicate with the distributors. Finally, it is centered on network, through which it supplies individuals with services. Fourthly, solutions can be measurable as well as enhanced for the individual while the use of sources can be regulated. Fifthly, sources offered the customer are clear so that the individual does not need to recognize its inner framework.

Lately, the regard to Big Data has actually been created describing those difficulties as well as advantages originated from gathering as well as processing large amounts of data. This topic has actually looked like companies should deal with petabyte-scale collections of data. In fact, in the last 2 years we have created 90% of the total data created in history. The sources of such significant quantity of details are those applications that collect data from click streams, purchase histories, sensing units, as well as elsewhere. However, the first issue for the correct meaning of 'Big Data' is the name itself, as we might think that it is just associated with the data Volume. The heterogeneous framework, varied dimensionality, and also Variety of the data representation, additionally have value in this concern. Just think of the previous applications that execute the data recording: various software applications will certainly bring about various systems and also protocols.

Naturally it additionally depends upon the computational time, i.e., the efficiency and Speed in both obtaining and processing the data. Existing customers require a 'bearable elapsed time' for getting an answer. We have to place this term in relationship with the offered computational resources, as we can not contrast the power of a computer with respect to a computational server of a big corporation. Lastly, one main worry about applications that manages this

sort of data is to preserve the Honesty of the information.

Big data is when the typical application of present innovation does not make it possible for users to acquire prompt, cost-efficient, and high quality answers to data-driven questions. We have to point out that extra definitions consisting of as much as 9V's can be additionally located, adding terms such as Value, Stability, and Visualization, among others.

The main difficulty when resolving Big Data is related to two highlights:

- The storage space as well as administration of huge quantities of details. This issue is related to DBMS, and also the typical entity-relation version. Industrial systems report to scale well, being able to deal with multi-petabyte databases, yet in addition to their 'expense' in regards to cost and also equipment resources, they have the restriction of importing data into a native depiction. On the other hand, commonly taken on open-source systems, such as MySQL, are far more minimal in regards to scalability than their commercial analytics counterparts.

- The procedure for carrying out the expedition of these large volumes of data, which intends to uncover helpful information as well as knowledge for future activities. The standard logical processing is led by an entity-relation system, where inquiries were created making use of the SQL language. The first hitch of these type of systems is the necessity of preloading the data, as specified previously. In addition, there is not much assistance for in-database data as well as modeling, as well as several DM designers might not fit with the SQL declarative design. Even in the event that engines supply these performances, as repetitive algorithms are not conveniently expressible as parallel procedures in SQL, they do not work well for substantial quantities of data

II. ARCHITECTURE OF BIG DATA MINING PLATFORM BASED ON CLOUD COMPUTING

The development of network cloud brings brand-new issues as well as obstacles in addition to a brand-new instructions of development to data mining. Cloud computing belongs to a commercial estimation version with the combination of network computing, parallel computing as well as distributed computing, the power of which recognizes terrific performance of big data mining. With the realization of standardization as well as normalization of the SaaS feature of cloud computing, big data mining based upon cloud computing of SaaS is gradually comprehended and also put into application. This write-up constructs a SaaS system of big data mining from 3 perspectives, namely, solution of big data mining based on cloud computing, parallelization of data mining algorithm and componentization of data mining formula.

The total framework of the big data mining platform based on cloud computing is revealed as in Fig. 1. The bottom level of the structure is supported by cloud computing, embracing cloud computing to provide dispersed storage and also computing capability for data mining performance. The style of the data mining platform depends on the middle. Ultimately the top degree data mining capability is accessed via a third party algorithm capacity, and after that so exposed that it can be called according to the requirement of business system.

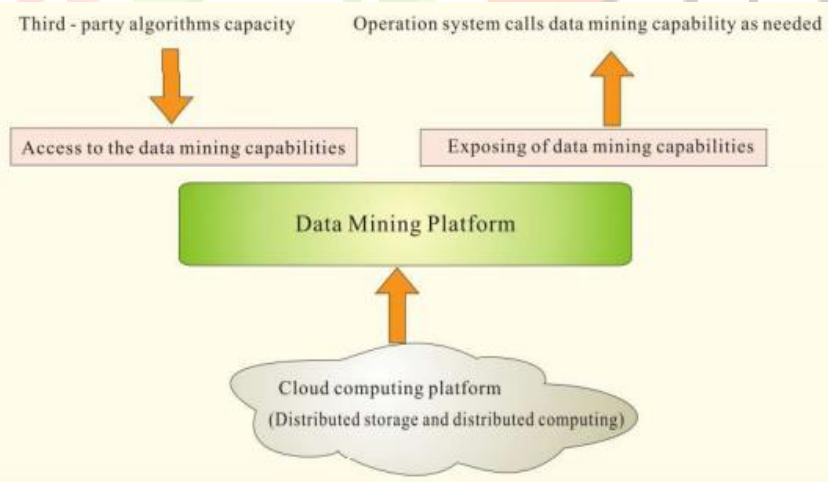

Figure 1 : Architecture of the Big Data Mining Platform Based on Cloud Computing

The typical architecture of data mining innovation is improved the basis of relational data source, which is incapable to satisfy the computing of substantial quantity of data. Cloud computing utilizing its distributed storage and computing kinds constructs a brand-new kind of cloud computing data mining system, as shown in Fig. 2, which mainly consists of three layers from the bottom to the top: the third layer of cloud computing assistance system, the 2nd data mining capability layer, and also the first layer of cloud solution of data mining.

To start with, the cloud computing assistance platform can be developed in a fast, simple and also extensible means to take care of huge and also complex IT facilities, which is generally composed of cloud system generally to save data, cloud computing platform provided priority to data handling as well as thorough cloud computing platform handling data storage and also processing at the same time.

Secondly, the layer of data mining capability primarily gives infrastructural power of data mining, the primary feature of which is parallelizing data handling formulas, sending off service administration structure, providing interior system data mining processing, suggesting mathematical library and making it possible for the third party data mining formula into the layer. This layer is the basis of mining providing and the core of the entire data mining system.

Third, the main function of the cloud solution layer of data mining is to give cloud solution and also pertinent engines for language and declaration access so as to promote automatic use of cloud service The interface forms solution ability encapsulation are branched out, mostly consisting of internet service under basic object accessibility method, scalable markup language (XML), hypertext transfer procedure (HTTP) and also regional application shows user interface (API). According to different scenarios, each organisation system of the cloud service layer can restructure as well as call data mining cloud solution.

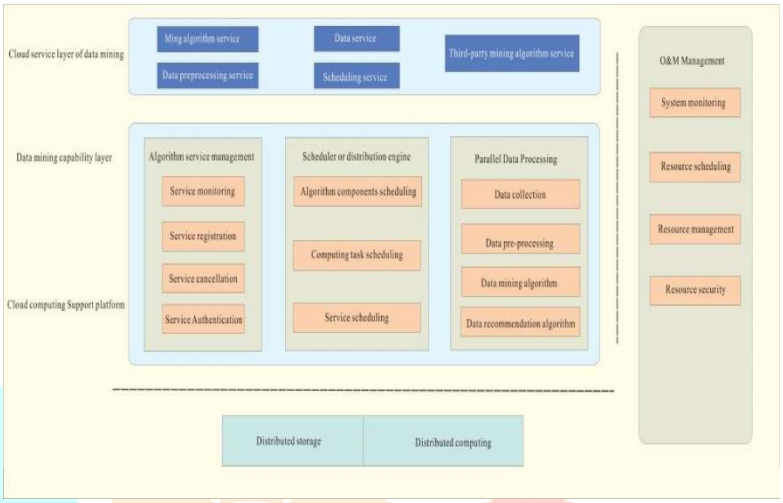

Figure 2 : Framework of Data Mining Platform Based on Cloud Computing

III. KEY TECHNOLOGIES TO CONSTRUCTING BIG DATA MINING PLATFORM BASED ON CLOUD COMPUTING

The building and construction of big data mining platform based upon cloud computing depends on the assistance from sophisticated science as well as innovation, for which the vital technologies required are provided as follows.

Cloud Computing Technology. The very first is distributed storage space modern technology, which takes use disk room on the computer system through the network and also constitutes an online storage device by scattered sources, thus accomplishing data storage in a dispersed manner. It symbolizes the reliability and economic performance of cloud computing in data, handling as well as helps to give the individual with an economical as well as excellent way of data mining.
The 2nd is virtualization innovation, which describes computer parts operating on the basis of the digital atmosphere. It assists to broaden the capability of the equipment as well as to streamline the software application reconfiguration procedure also enables running multiple operating systems on a solitary system, where the programs are independent of each other and also prevent mutual impact, eventually substantially enhancing the efficiency of the computer system.

The 3rd is parallel cloud computing technology, which contributes to reliable implementation of data mining technology and also to envelop the information of cloud computing such as task similarity, task organizing, job mistake resistance, system fault resistance or data distribution, etc. The user does not require to appreciate these information so regarding improve the development performance.

Data Collection Control Center. Its function is accumulating various kinds of data, that is, completing the collection job of all the business data that have been accessed to the cloud computing data mining platform, and to fix the contradictions important to appropriate stipulations as well as protocols between various data, making them adapt to various source data styles.

Solution Scheduling as well as Management Technology. In the big data mining platform based on cloud computing, corresponding service scheduling and also administration technology is so essential that it enables different service systems to use the computing system. Service organizing assists to guarantee safety and security and dependability of cloud solution while management technology combines the features of service exposing as well as enrollment. The last makes it viable for the third party data mining as well as contributes to broadening the influence of the solution platform.

Mining Formula Parallelization Modern Technology. Data mining algorithm parallelization under cloud computing efficient utilize the standard abilities the cloud computing system gives, primarily including algorithm parallelism, parallel approach picking, etc. Parallelization of data mining algorithms of contributes to make complete use of the sources of each workstation and to execute unified organizing and coordinate processing, hence lastly accomplishing effective parallel computing.

IV. DATA MINING AS A SUPPORT FOR BUSINESS INTELLIGENCE APPLICATIONS TO BIG DATA

The building and construction of big data mining system based on cloud computing depends on the support from innovative science and also technology, for which the essential modern technologies required are listed as adheres to.

Cloud Computing Technology. The initial is distributed storage space technology, which takes use disk space on the computer system with the network as well as comprises a digital storage device by spread sources, therefore attaining data storage in a distributed way. It personifies the dependability as well as financial efficiency of cloud computing in data, handling and aids to offer the user with an inexpensive and exceptional method of data mining.
The 2nd is virtualization modern technology, which describes computer parts operating on the basis of the virtual environment. It helps to broaden the ability of the equipment as well as to simplify the software program reconfiguration process also allows running numerous operating systems on a single platform, where the programs are independent of each other and also avoid shared influence, eventually significantly improving the effectiveness of the computer.

The 3rd is parallel cloud computing modern technology, which contributes to reliable application of data mining innovation and to envelop the information of cloud computing such as job similarity, job scheduling, task mistake tolerance, system mistake resistance or data circulation, and so on. The individual does not require to care about these details so as to boost the growth performance.

Data Collection Nerve Center. Its function is collecting various kinds of data, that is, completing the collection job of all the business data that have been accessed to the cloud computing data mining system, as well as to solve the contradictions relevant to appropriate stipulations and procedures between various data, making them adapt to numerous resource data styles.

Service Scheduling and also Monitoring Modern Technology. In the big data mining platform based upon cloud computing, equivalent service scheduling and also management innovation is so vital that it enables different service systems to make use of the computing system. Solution organizing helps to make sure safety and security and integrity of cloud solution while administration modern technology merges the functions of service subjecting and also enrollment. The latter makes it practical for the 3rd party data mining and also is conducive to increasing the impact of the service platform.

Mining Algorithm Parallelization Innovation. Data mining formula parallelization under cloud computing reliable make use of the basic capacities the cloud computing system offers, mostly including algorithm similarity, parallel approach choose, and so on. Parallelization of data mining algorithms of is conducive to make full use of the resources of each workstation and also to execute unified scheduling as well as coordinate handling, therefore ultimately attaining effective parallel computing.

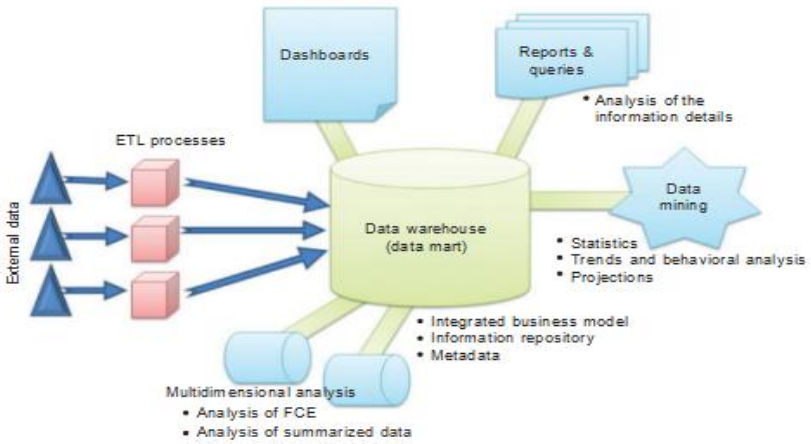

FIGURE 3 : Business Intelligence structure.

Energies: Power usage forecast

Utility business utilize clever meter to determine gas and electrical power intake. These devices produce big volumes of data. A big data facilities requires to keep track of as well as evaluate power generation and also usage using wise meters.

Social Network: Belief evaluation

Social networking companies such as Twitter requires to determine what individuals are saying and topics which are trending in order to do view evaluation.

Telecommunication: Anticipating analytics

Telecommunication offers requirement to construct churn versions which relies on the customer profile data characteristics. Predictive analytics can predict churn by evaluating the customers calling patterns.

Client Service: Call monitor

Call facility big data solutions make use of application logs to improve efficiency. The log files demands to be combined from various styles prior to they can be used for analysis.

Financial: Scams Detection

Bank ought to be able to avoid scams on a transaction or an individual account. Big data solutions must evaluate purchases in real time and also provide referrals for immediate activity and also stop fraud.

Merchants: Product referral

Retailers can keep track of customer surfing patterns and also history of items acquired as well as provide a solution to advise products based on it. Merchants need to make privacy disclosures to the individuals before implementing these applications.

V. CONCLUSION

It is of terrific value to build the big data platform based on cloud computing. The big data mining system based upon cloud computing is valuable to resolve the issues relevant to conventional data mining modern technology such as low efficiency, backwards function, post ponement as well as lag of info and also high expense. The platform aids to attaining high performance, reliability and economic situation in data mining. This paper starts to create the big data mining system from three angles of servitization of cloud computing of large data mining, the parallelization of data mining formulas, the componentization data of mining formulas, as well as evaluates three core cloud computing technologies significant to data collection control center, service organizing as well as monitoring and mining formula parallelization. It tries to supply recommendation for the construction of big data mining system.

REFERENCES

[1] D.-H. Tran, M. M. Gaber, K.-U. Sattler, "Change discovery in streaming data in the age of big data: designs and also concerns," ACM SIGKDD Explorations, Vol. 16, No. 1, pp. 30-38, 2014

[2] W. Follower, A. Bifet, "Mining Big Data: Existing Status, and also Forecast to the Future," ACM SIGKDD Explorations, Vol. 14, No. 2, pp. 1-5, December 2012.

[3] Y. Demchenko, P. Grosso, C. D. Laat, P. Membrey, "Resolving Big Data Issues in Scientific Data Infrastructure," 2013 International Conference on Cooperation Technologies as well as Equipments (CTS), 20-24 May 2013, San Diego, CA, UNITED STATES, pp. 48-55, 2013.

[4] D.E. O'Leary, "'Big Data', the 'Web of Points' and also the 'Internet of Indicators'," Intelligent Equipment in Audit, Money and Monitoring, Vol. 20, pp. 53- 65, 2013.

[5] H.V. Jagadish, A. Labrinidis, Y. Papakonstantinou, et al., "Big Data and also Its Technical Obstacles," Communications of the ACM, Vol. 57, No. 7, pp. 86- 94, 2014.

[6] S. K. Markham, M. Kowolenko, and also T. L. Michaelis, "Disorganized Text Analytics to Support New Item Advancement Decisions," Research Innovation Management, pp. 30-38, March-April, 2015.

[7] S. B. Boddu, "Eliminate the loud data from websites using data mining strategies," Computer technology and Telecom, Vol. 38, No. 2, pp. 39-46, 2013.

[8] C.L. P. Chen, C.-Y. Zhang, "Data-intensive applications, obstacles, strategies and also innovations: A survey on Big Data," Details Sciences, Vol. 275, No. 10, pp. 314-347, 2014.