International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 15 (2015) pp 35221-35225

Integrative Information River Model For Financial Market's Big Data Analytics

Krishna Kumar Singh

Research Scholar, Dept. of Computer Science,
GBPEC, pauri, Uttarakhand.
Email: singhkrishna80@gmail.com

Dr. Priti Dimri

Associate Professor and Head, Dept. of Computer Science,
GBPEC, Pauri, Uttarakhand
Email: pdimri1@gmail.com

ABSTRACT
Forecasting of financial market can be done with cloud computing based BIG DATA ANALYTICAL tools and methodologies but challenges in the cloud computing like ownership, privacy & security of data is the big hindrance in implementation of cloud computing in financial market. Tools like Hadoop can only be effective when sufficient, effective and accurate data is available at right time and place. Efficient financial market discounts each & every information. So, Demand of integrative data in the financial market needs customized framework under which BIG DATA ANALYTICAL tools will become effective. Financial market analysis can't be possible without integration of information placed at any corner of the globe. This paper deals with the design framework of scattered data and introduce Integrative Information River (IIR) model for the data required in the financial market forecasting under cloud computing. BIG DATA based tools like Hadoop and Hive will work on the top this model for the financial market forecasting.

Keywords: BIG DATA ANALYTICS; Integrative Information River (IIR) Model; Hadoop; Financial market forecasting;

I. INTRODUCTION

This work is the extension of our previous work in stock market analytics. We discussed about green database and design new model for the green database [5, 6, 7] in the stock market. Nature is the best teacher. Energy only changes its form and never destroys. If water flows within tightly bounded banks from source to destination is called river. Information from all corners of the globe can need to be integrated. Model for integration of information on the web can be well understood by natural river model and it is termed as Integrative Information River (IIR) model. Let us first describe the nature of rivers and its entire journey. Life of a river starts with the mountains where many small sources of water can come down from mountains and gradually merge to become a river. River flows form source to destination (either it merge with a big river or sea). Flood like situation is great devastation for the human being. Dams were created to control flood and uncontrolled flow of water. Dam water can be used in various methods for the benefit of human being and water is being release as per the need in control basis. Number of dams can be decided as per volume of water flow in the river. Volume of water can also be dependent upon life cycle stage of the river. Decisions about dams are based on need, necessity or compulsions. Water can reached to its final destination i.e. sea. Again water is re - circulated trough evaporation process and same cycle can repeat again & again. Pic01 shows first phase of life cycle i.e origin of the river, where multiple sources of water can meet and formed a continuous free flow of water with tight boundary.

Pic01: small rivers meet at the bottom of the mountain [13].

In Pic02, more than one river is merging and formed a big river. In both meetings, we are seeing the color of the water from two sources is different and after merging its color is different from earlier two.

Pic02: More than one river merged to become a big river [14].

In Pic03, a reservoir is created on the river. Water collected at this place is huge and label of water is maintained as per capacity of the dam. Dam needs a very large bounded place with one inlet and one outlet. Outgoing of the water is controlled but incoming. Release of water is decided as per flow of water and capacity of dam. Water can be used over the year without fail on various modes like irrigation, generation of electricity, drinking water etc.

Pic 03: Dam is created on the river [15].

In the next stage of the River's life is reached to the destination. Pictorial representation of this stage is shown in Pic04. Destination for some rivers is another river and some rivers are ocean.

Pic04: River reached to its destination [16].

Nature maintain re – circulation of water from sea to the mountains and river via evaporation & rain. Water is circulating and changing its form. Mode of flow of water is changed at every stage of its life cycle. Financial markets are so dynamic and it takes numerous calculations and algebra for decision making specially forecasting of the market. Decisions in the financial market are solely based on the integrative quality information. Supply of information at right time, place and person is the biggest challenge. Financial BIG DATA created on the web needs to segregate & linked properly, so that analytical tools can apply on integrative data in well planned manner. In this paper, we are discussing Integrative Information River (IIR) model for the financial data and try to provide cloud based architecture of financial BIG DATA for financial market forecasting. This model will become backbone for the financial BIG DATA analytical tools for forecasting.

II. LITERATURE REVIEW

History of big data analytics is long. Many researchers have contributed a lot and some of them are discussed in this part of the paper. Bhavani et. al [2], discusses about secure data storage and retrieval in the cloud. In this paper, they focused upon large data generated from different sources and its interlinked connectivity. Hadoop and Hive like cloud tools are provides fine grains access to resources. Big data is commonly unstructured and need more real time analysis [3]. Big Dimensionality [12] is based on five V's and major challenge in the management of this is 'how to segregate and pour into a single main stream. According to Bhart et. al. [1], database analytics had its unique role in the big data scenario. Fetching unique data from big data needs configurable approach. It opens new discussions about database analytics and its differences from BIG DATA analytics. Big Data applications typically run as a set of Map Reduce jobs to take advantage of Hadoop's ease of service deployment and large-scale parallelism. Yet, Hadoop has not been adapted for multilevel secure (MLS) environments where data of different security classifications co-exist [9]. Analytical tools are further applied at different domains. Keith C.C chan [4] introduces use of analytical tools in the drug industry.

Discovery of drugs with the help of Hadoop based tools quite helpful. In spite of these advanced developments, BIG DATA analytics in the financial market needs more customization. There is big gap between demand of data set in the financial market and its supply. Financial market predictions based on neural network is quite common and Kuo proposed a neural network based Decision Support System (DSS) to provide investors with suggestions on transaction timing and transaction strategies of Taiwan 50 Index Exchange Traded Funds (ETF) [8]. BIG DATA processing model is based on demand – driven aggregation of information sources [10]. Learning data behavior with the certain depth is the active research area in machine learning. With the sheer size of data available today, big data brings big opportunities and transformative potential for various sectors. As the data keeps getting bigger, information extraction from the big data becomes more and more complex [11].

III. PROBLEM STATEMENT

Web is full of information and extraction of information from web is tedious job. It is hard to find out relevant integrative information needed for the particular time period. Efficient financial market follows efficient market hypothesis which reacts on every information. Tsunami of information is creating more confusion for the financial market investors. There are many tools available for forecasting of the financial market behavior by analyzing data. Efficiency of these tools is dependent upon many factors and quality of data is one of the key factors. Efficiency of the financial forecasting is largely dependent upon maximum inclusion of information in the forecasting processes. Major challenge in the financial market forecasting is the variety of data required from different sources and these are owned by different agencies. We required a model which can ease our process of streamline of these information's from different sources. How to use cloud based technologies and on what data?

IV. PROPOSED MODEL

This model is based on life cycle of Information River (IR). Information River (IR) model is based on believe that energy only changes it form and don't destroys. In the similar way, information once generated has linked with other information and it can't destroy but changed its forms. Information generated gradually from different sources on the globe is treated as birth of information. Like color of the water in the river, data has different nature and format. One information river (IR) can be merged with other information river (IR) same as two or more rivers can meet at place and became a single river. Small canals (small themes based like financial, political, economic etc) are created in the Information River (IR) based on the need of information e.g. stock market forecasting, GDP forecasting etc. These canals are again meets into river and flow towards its destination sea. Utility value of information is calculated based on its immediate domain based use and temporal value. Information River (IR) is based on particular theme. Various sources under one theme are like different sources of water in the mountain that's collectively formed Information River (IR). Reservoirs are servers or data centers. All information based on all themes collectively is WWW. Information under the different themes is best utilized by various methods. Reservoirs and canals are created again and again on the same river as per the use of information. Different school of thoughts has its own view about the same data and they can build Reservoirs and canals. Length and breadth of the reservoirs and canals can form according to the need and flow of information. Similarly, number of the reservoirs and canals can also be dependent upon the need and flow of information. Whole process of generation of information to the sea and again reaches to the rivers through evaporation is natural and well defined. Only it's our acumen and our expertise which counts. Information reservoirs must be integrated with other sources (including other reservoirs) and they have compulsion to release information continuously. Flow and connectivity will ensure distribution of integrated information at low cost for the peoples placed at the bottom. In spite of all reservoirs, this model is built in the porous format. Porous quality of information will ensure availability of information to all at low cost. Cloud servers (reservoirs) can be created in such a way that it must cater a defined area. Various tools are available to extract knowledge from data at www like distributed computing and cloud but specialized market like financial forecasting can't be address till now because it follows efficient market hypothesis. Financial forecasting is the very specialized market which requires information from almost all nook and corner of the globe. As per the definition of efficient market hypothesis, market reacts on each & every news. So integration of information from each source is important to us. As seen in the fig01, Information River (IR) model is drawn as a graph with several nodes and edges. Canals are the edges and servers are the nodes. Servers are of any type and it is based on different technology as per the use and need. Open source technologies for cloud and big data like Hadoop, Hive etc are extremely helpful. These technologies will be work on the top of this model.

Fig 01: Integrative River Model.

This model is helpful in the data definition and segregation of the financial BIG DATA. Integrative Information River (IIR) model is designed for the financial market forecasting but it may be used in the other stream of data. Cloud based technologies are still encountering problems like security, privacy, ownership of data etc. Majority of data owning agencies like govt. are conservative to share its data to the market player. This model is trying to address these challenges. Data stored at the servers/data centers (or dams) are owned by different agencies and they have full control over it. Security of data flow through canals can be address as same as river banks. Data or information flow with tightly bounded and well define banks. We can apply security features along the banks, so that there is no breach of security or trust. Different financial forecasting agencies demand data from these owners as per their need and servers are free to provide data as per their preferences. After flow of data from one dam to the river or canal, another dam can built as per their mutual or exclusive agreement. Here rivers and dams are connected to each other, so it is easy to flow data in the river but accessibility of data is dependent upon excess permission. Canals can be built as per the need of the like minded agencies with mutual agreement between parties. These parties can ensure the security issues of the information flow in the canals. In this way ownership and security issues in the cloud based framework can be address properly.

V. GREEN INITIATIVES

Concerns of cloud based technologies like security and ownership of the data is the big hindrance in the implementation of these technologies. Integrative Information River (IIR) model can deal with these challenges and now it's easy to implement cloud based technologies with IRR model in the real time systems. Stock markets are still struggling to implement cloud computing especially in the third world economies. There are more than 96, 000 intermediaries (data from SEBI, India, 2013) registered in Indian stock market and many more are working without registration. Means same data is stored at least 96, 000 times in India. Middle class population is the backbone of Indian economy and they are perceiving stock market as gambling ground (Because of non availability of integrative information to the investors at affordable cost). Cloud computing will save space, time and cost in these areas. IIR model will be helpful in generating more confidence in the implementation of cloud computing and address issues of ownership and privacy of data. Finally, it can save resources.

VI. FINANCIAL FORECASTING ISSUES WITH INTEGRATIVE INFORMATION RIVER (IIR) MODEL

We are living in the Global Village. Each and every activity is connected to each other. Irrespective of its nature and stream, information generated at one part of the globe has relationship with other. Forecasting of financial markets is the long and continuous process. Markets generally follows efficient market hypothesis and efficient financial market discounts each & every information. Efficient forecasting can only be possible after considering and analyzing all possible information in the channelized manner. Integrative Information River (IIR) model is the way of channelizing financial BIG DATA. Financial BIG DATA need to be considered in the connected manner. Here no information is kept into isolation (whether it is political, economic, social or any other). Integrative Information River (IIR) model is design to address issues of privacy and ownership of information under cloud computing. Through this, information is linked with each other as rivers are and helpful in analyzing all connected information. Information generated at any walks of life and streams has capability of changed the investment scenario altogether. Cloud based technologies can be run on the top of Integrative Information River (IIR) model. Different analytical and forecasting tools can be use to analyze integrative data based on this design. Data available through this process is cost effective and integrative.

VII. CONCLUSION

Integrative Information River (IIR) model is helpful in the definition and segregation of financial BIG DATA. It can be used as base for financial BIG DATA based technologies like Hadoop, Hive etc. This model is generated in view of the different data set required for the financial market forecasting. It will helpful in filtering process of the data in cloud. Used of technologies in the cloud computing will become much more easy.

VIII. FUTURE WORK

Integrative Information River (IIR) model addresses the framework issues of financial BIG DATA. Statistical and other tools are used in the forecasting of financial market is still not tested on all type of data flow but few. Data required for financial forecasting in the short and long run is different. Technical and fundamental analyst's views are quite different. So testing on all sets of data flow from on all parameters or views is needed to be done.

REFERENCES

[1] Bharat Sukhwani. et. al. "Database analytics: are configurable – computing approach", Published by the IEEE Computer Society, pp. 19 – 29, Feb 2014.

[2] Bhavani Thuraisingham, Vaibhav Khadilkar, Anuj Gupta, Murat Kantarcioglu, Latifur Khan, "Secure Data Storage and Retrieval in the Cloud".

[3] Han Hu, Yonggang Wen, Tat – Seng Chau and Xuelong, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial", IEEE Access, pp. 652 - 687, July 8 2014.

[4] Keith C.C. Chan, "Big Data Analytics for Drug Discovery", 2013 IEEE International Conference on Bioinformatics and Biomedicine, 2013.

[5] Krishna Kumar Singh, Dr. Priti Dimri and Madhu Rawat, "Green Data Base for Stock Market: A case study of Indian Stock Market", IEEE Xplore digital library, pp. 848 – 853, DOI:

10.1109/CONFLUENCE.2014.6949306, 25 – 26, September 2014.

[6] Krishna Kumar Singh, Dr. Priti Dimri and J. N. Singh, "Green Data Base Management System for intermediaries of the Indian stock market", IEEE Xplore digital library, pp. 1 - 5, ISBN number: 978-1-4799-3063-0, DOI: 10.1109/CSIBIG.2014.7056996, March 2014.

[7] Krishna Kumar Singh, Dr. Priti Dimri, and Somitra Chakraborty, "Green Referential Data Base Management System for Indian Stock Market", International Journal of Computer Application, Vol. 89(3), pp. 8 - 11**,** ISBN: 973-93-80880-18-3, DOI:10.5120/15480-4197, March 2014.

[8] Kuo, Mu-Hsing Chen, Chih-Lung, "An ETF Trading Decision Support System by Using Neural Network and Technical Indicators", "2006 International Joint Conference on Neural Networks", pp. 2394 - 2401,July 16-21, 2006.

[9] Thuy D. Nguyen, Mark A. Gondree, Jean Khosalim, Cynthia E. Irvine, "Towards A Cross-Domain MapReduce Framework", "2013 IEEE Military Communications Conference", pp. 1436 – 1441, 2013.

[10] Xindong Wu,Fellow, Xingquan Zhu, Gong-Qing Wu, and Wei Ding, "Data Mining with Big Data", "IEEE Transactions on Knowledge and Data Engineering", Vol. 26, NO. 1, January 2014.

[11] Xue – Wen Chen, and Xiaotong Lin, "Big Data Deep Learning: Challenges and Perspectives", http://www.ieee.org/ publications_standards/publications/rights/index.html , pp. 514 – 525, May 28 2014.

[12] Yiteng Zhai, Yew-Soon Ong and Ivor W. Tsang, "The Emerging Big Dimensionality", IEEE Computational Intelligence magazine, pp. 14 - 16, Aug 2014.

[13] http://www.panoramio.com/photo/63452141

[14] http://iheartsudan.tumblr.com/image/79743187313

[15] http://en.wikipedia.org/wiki/Srisailam_Dam

[16] http://www.earthtimes.org/newsimage/smallnew-eco-battery -potential-rivers-meets-sea_303.jpg