

# The Implications of Big Data in Indian Stock Market

Krishna Kumar Singh  
Research Scholar  
Dept. of Computer Science  
G.B.P.E.C, Pauri Garhwal  
Uttarakhand, India

Priti Dimri, Ph.D  
Associate Professor and Head  
Dept. of Computer Science  
G.B.P.E.C, Pauri Garhwal  
Uttarakhand, India

Krishna Nand Rastogi  
Research Scholar,  
Dept. of Computer Science  
G.B.P.E.C, Pauri Garhwal  
Uttarakhand, India

## ABSTRACT

Big, Unstructured, heterogeneous and temporal data is being generated every second in the stock market and it requires a new school of thought which can not only handles its complexities but also able to help in the future prediction and analytics of the market. Big data analytics is a very promising area and buzz word for the next generation information technologies. Knowledge discovery and future forecasting will not possible without handling the core challenges of big data. Stock market is one of the burning areas where data is growing day by day. Because of these heterogeneity and other complexities of data, big data architecture and design is needed which specifically deals with the stock market data and analyze these heterogeneous data for the future prediction of the market. This paper deals with nature of data generated and required for knowledge discovery & future prediction of the stock market. It also deals with the relevance of big data analytics in the stock market.

## Keywords

Big Data, Stock Market forecasting, stock market, temporal data, big data architecture

## 1. INTRODUCTION

Human reasoning is based on the decision maker's ability to process what they know, whether they learn by example or create new approaches. Reasoning is used by managers to discover patterns in data, helps to infer multiple meanings from a single input and generalize from diverse inputs. Intelligent information systems can aid and enhance decision making by applying specific reasoning tools: model-based reasoning, abstract reasoning, heuristic reasoning, procedural reasoning, analogical reasoning and case-based reasoning [7]. As we enter the new millennium, innovations and competitive pressures force organizations to rethink the manner in which they conduct business and to redefine the essential ingredients for success [8]. The basis of competition is being fundamentally altered through the introduction of the Internet and other advanced technologies (Sampler, 1998) [15]. The phrases 'big data' and 'open data', two particular manifestations of this sudden dawning of awareness, are just two terms which appear to be cropping up everywhere. Big data, obviously enough, refers to massive data sets, but quite what is meant by 'big' depends on the context. 'Big', in any case, is likely to grow over time. However, one quick definition is that it is a data set which is too large to fit into the computer's memory in one go. Open data are data which are released to the public. This will often be official data, collected, for example, by a National Statistical Institute, but the term can also refer to scientific data. The open data movement in science is being paralleled by other similar initiatives, such as open access publishing [2]. Big data means

- a) Data which are too extensive, heterogeneous and temporal to store and process by a single computer.
- b) 'big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze' [11];
- c) 'high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization' [5];
- d) 'a data set is large if it exceeds 20% of the RAM on a given machine, and massive if it exceeds 50%' [3].

Big data can also be understood by following example. It might seem pretty obvious from these examples, and from the quotes I opened with, that big data is a new phenomenon for the financial market. Or, at least, so one would think. But one would be mistaken. Some thirteen years ago, Hand et al listed [6]:

- a) Wal-Mart, with over 7 billion transactions annually;
- b) AT&T carrying over 70 billion long distance calls per year;
- c) Mobil Oil aiming to store over 100 terabytes of data;
- d) The NASA Earth Observing System, which was projected to generate around 50 gigabytes of data per hour.

It is true that big data provide us with analytic opportunities. These come in various shapes and forms. One is the possibility of detecting deviations, anomalies, or structural characteristics which are so small they would be imperceptible, or swamped by noise in anything but a large data set. The financial markets illustrate this [2]. According to the efficient market hypothesis, market reacts and discounts each and every news flow comes from every nook and corners of the globe. By considering efficient market hypothesis, it is quite natural that data comes at any time irrespective of the market conditions must be noticed for the market forecasting. Irrespective of the nature, format and behavior of the data, markets discount news by considering its implications in near future. After globalization, news and information is coming very fast and it is in the different format i.e. text, picture, image, matrix, word files, jpeg etc. It quite complicated to incorporate all information coming every second from all corners in the traditional database. Traditional database can't be able to cope up with the pace of the information and it is not efficient enough to handle it for the use of forecasting methodologies. To provide general implications of the information with great pace and dynamics, a new and contemporary architecture is required for the stock market data. New architecture should be design in such way that it saves data easily, changes its state dynamically and provides information with integration all data in desire format for market forecasting. Unavailability of Information has become one of the big hurdles to predict the market and due to it a great fear

about the market persist in the middle & small scale investors. This paper deals implications of the big data in the stock market and discuss new framework of integrated technologies required to handle this big data.

## 2. LITERATURE REVIEW

A stock exchange is an entity which provides “trading” facilities for stock brokers and traders to trade stocks and other securities. How to invest in stock exchange is one of the important issues in investment, and one of the factors that can help investors in the process of investment is the efficiency of the corporation under consideration. In traditional point of view, the most important objective of efficiency appraisal was a review of the historical efficiency [9]. Although big data concepts in the financial market are relatively new and are introduced in the last few years, internet and related technologies are not new for it. The realities of the electronic marketplace make Internet technology the most used form of IT (Ferguson, 1996)[4]. Stock investment has been a major investment tool for personal wealth. Stock investors have been eager to know how to select stocks to gain profit. Generally speaking, there are two most often used analytical methods when one selects stocks to invest: fundamental analysis and technical analysis. Fundamental analysis mainly focuses on the listed companies’ operation and financial status to forecast future profit/loss and one can then select stocks accordingly. Technical analysis focuses on historical stock price movement, which might show a pattern or trend used by investors to forecast the future possibility of the rise/fall of the price, and the investor can then decide whether to invest or not [1]. Researchers are continuously trying to explore new methodologies for the stock market prediction. Volume of stock prices traded per session is crucial in the forecasting methodologies[10]. One of the original works describing the application of evolutionary algorithms to stock trading can be found in Oussaide’ne et al. (1996, 1997) [13][14]. The usefulness of a model of this type depends upon its ability to forecast future changes in stock market valuations [12]. There is dearth of literature available on the big data for the stock market. An ample amount of research evidence is available for the big data in financial market but there is hardly any evidence of customized big data architecture over Hadoop for the stock market.

## 3. NATURE AND BEHAVIOR OF DATA IN STOCK MARKET

According to the data from SEBI (Security and Exchange Board of India), there are more the 83,000 intermediaries registered in India and its growing day by day. Intermediaries are developing its own repositories and due to it, same data has been saved at all places. That is using power, space, energy and cost. In current market scenario, future forecasting requires mammoth of data comes from every corners of the globe. Data repetition of data at all databases of interested parties is the great concern for green computing practitioners. Indian as well as world market had witness many down falls in the past. Mayhem in the financial market downgrades the confidence of the middle class population. In year 2008, subprime lending in the America had impacted all over the world market and a global recession or slowdown had witnessed. Similarly in 2008 – 09, information about Satyam created a vacuum in the market and small investor’s committed suicide. Little and slow access of information became big reason for the big loss in the market for small investors. All information need to integration very fast and provide information to the small and medium investors at very reasonable & affordable cost. Information Traditional

database can’t be able to handle and integrate this type of heterogeneous, temporal, unstructured and dynamic data stored at different places, so a new architecture is required for the integration of the heterogeneous database. Data can’t be available at one place or one database and it is in the different format. It requires an integrated mechanism which provides all temporal data at one place and easy to process for the decision making process in desired format. Glimpse of data and its formats are given in Fig. 01.

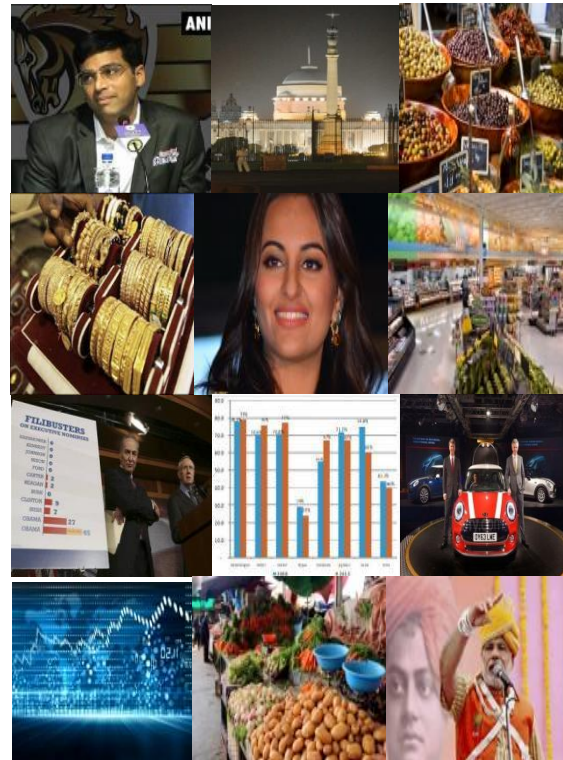


Fig. 01.

Stock market is affected by all information’s. As per the efficient market hypothesis, stock market must react on all of the information’s. Small and Middle class investors haven’t such type of resources to know all happenings instantaneously and take their positions respectively. By boys have edge over it and they are always making money. Loss fear factor had been played crucial role for the small and middle class population and that’s why they are keeping distance from the market. Easy availability of this integrated data from all corners of the globe is the key issue. Due to temporal nature of the data required in the stock market, big data architecture is being implemented in such a way that all data must be updated and integrated on timely basis and reports should be generated based on the current information. To achieve efficient market hypothesis, these information must reached to all participates of the market. Due to shared resources and other above discussed benefits of big data technologies, small investors too have equal opportunity to excess relevant data. This will increase participation, participation will increase more investments, investments will increase more stable market with market hypothesis, more realistic market hypothesis will leads to more accurate prediction of the market and finally it leads to stable and mature market. These will create investment cycle as shown in Fig. 02[10].

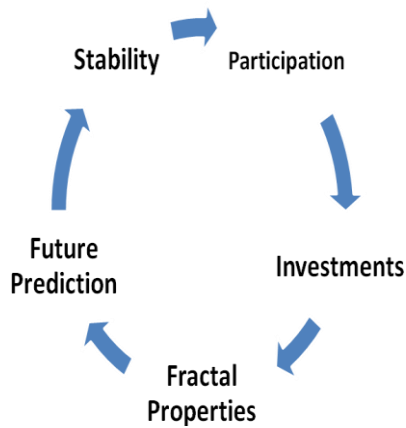


Fig.02

Participation will increase investment in the market. Investment will increase fractal behavior and that leads to more accurate prediction of the market. Final this process will forward towards more stable market in the third world countries like India also. Behavior of the stock market will entirely change after implementation of the big data technologies in the stock market.

#### 4. BIG DATA ARCHITECTURE DESIGN FOR STOCK MARKET DATA

Data required for the stock market prediction is placed and generated at different and distinct places of the globe. Distinguished agencies are maintaining this data in its own format. So, to cater need of current scenario, big data architecture for the stock market is framed as in Fig. 03. Big data architecture for stock market is divided into three parts. News and Information flows from any corner have impact on investment. First job is to identify different data and its sources required for the future prediction of the market. Store, acquire and process of heterogeneity, unstructured and temporal data is the second major challenge. Last part of the architecture is to attain its goal. In this section, analyze, visualize, reporting and decision making has done. Every second data is generated for the market and its format is different from previous dataset. To store and integrate these temporal data for the stock market is the biggest challenge. In current scenario, cost required for these data is affordable by few big investors. Investors from third and fourth world countries have acute scarcity of resources. Big data architecture with latest technologies will provide easy access of information flow and it is relatively affordable to all investors.

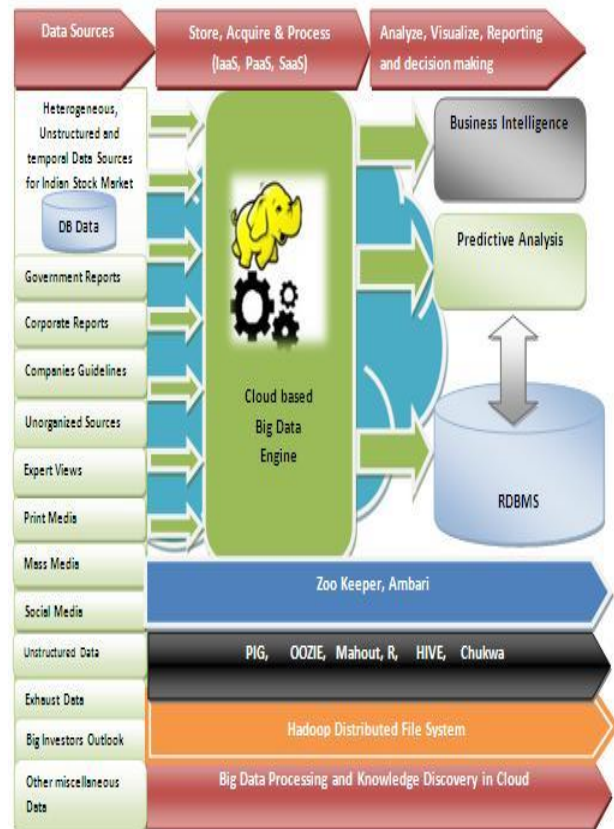


Fig. 03

Cloud based big data engine is used to integrate different heterogeneous and unstructured data. Cloud based technology will provide all analytical services to the clients at reasonable costs. Distributed file systems like HDFS (Hadoop Distributed File System) and parallel processing in the cloud will increase efficiency in the processing of data with quick response.

#### 5. RESULTS AND CONCLUSION

Big data integration will impact greatly on the growing markets like India. There is no need to store same data again and again. Intermediaries and other interested parties will be benefited by two ways. First, they can get data without wasting time and cost as per their own desire format. Secondly, they exploit the benefit of other existing benefits of Big or cloud computing like SaaS, Paas, IaaS etc. Implementation of big data technologies and above design architecture will reduce use of resources with same efficiency and effectiveness. Implementation of green computing will also helpful in the forecasting technologies or methods. Indian stock markets are still working with the traditional database and only upper class of investors has capacity to afford information from all database maintained by different agencies. Because of it, market theories and definitions like efficient market hypothesis, fractal market hypothesis etc are not applied properly in the third world markets. Extensive, heterogeneous, unstructured and temporal data required for the market prediction will handled only by big data technologies, i.e. cloud computing, parallel computing, high performance computing, distributed computing etc. Unique feature of cloud computing like Sharing of resources must be incorporated in the market for more sustainability of the investment process and economies. In spite of all these benefits there are few hiccups also which must be deal. Few of them are

1. Ownership of big data is still not defined and it is big hurdle in the framework design and implementation of such architecture.
2. Security issues of the data exist.
3. Data redundancy at various places by intermediaries is the greater concern for the greener technologies.

## 6. REFERENCES

- [1] Chien-Jen Huang, Peng-Wen Chen, Wen-Tsao Pan, "Using multi-stage data mining technique to build forecast model for Taiwan stocks", Springer-Verlag London Limited 2011.
- [2] David J. Hand, "Data, Not Dogma: Big Data, Open Data, and the Opportunities Ahead", LNCS 8207, pp. 1-12, 2013.
- [3] Emerson, J.W., Kane, M. J.: Don't drown in data. *Significance* Vol. 9(4), pp. 38–39, 2012.
- [4] FergusonGT. Strategy in the digital age: Role of information technology in corporate strategic planning. *Journal of Business Strategy*, Vol. 17(6), pp. 28–31, 1996.
- [5] Gartner (2012), <http://www.gartner.com/DisplayDocument?id=2057415&ref=clientFriendlyUrl>.
- [6] Hand, D.J., Blunt, G., Kelly, M.G., Adams, N.M.: Data mining for fun and profit. *Statistical Science* 15, pp. 111–131, 2000.
- [7] Hand, D.J.: Mining the past to determine the future: problems and possibilities. *International Journal of Forecasting* 25, pp. 441–451, 2008.
- [8] Hoplin HP. Re-engineering information technology: An enabler for the new business strategy. *Industrial Management & Data System*, Vol. 95(2), pp. 24–27, 1995.
- [9] Kamran Rezaie, Vahid Majazi Dalfard, Loghman Hatami-Shirkouhi, Salman Nazari-Shirkouhi, "Efficiency appraisal and ranking of decision-making units using data envelopment analysis in fuzzy environment: a case study of Tehran stock exchange", *Neural Comput & Applic* DOI 10.1007/s00521-012-1209-6.
- [10] Krishna Kumar Singh, Dr. Priti Dimri and Madhu Rawat, "Fractal Market Hypothesis in Indian Stock Market", *IJARCSSE*, Vol. 3(11), pp. 739-743, Nov 2013.
- [11] Manyika, J., Chui, M., Brwon, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, R.H.: Big data: the next frontier for innovation, competition, and productivity (2011), <http://www.mckinsey.com/insights/business-technology/big-data-the-next-frontier-for-innovation>.
- [12] Patrick Thomas, "A relationship between technology indicators and stock market performance", Jointly published by Kluwer Academic Publishers, Dordrecht Scientometrics, and Akadémiai Kiadó, Budapest Vol. 51, No. 1, pp. 319–333, 2001.
- [13] Oussaide`ne M, Chopard B, Pictet OV, Tomassini M Parallel genetic programming: an application to trading models evolution. In: Proceedings of the first annual conference on genetic programming, GECCO '96. MIT Press, Cambridge, pp 357–362, 1996.
- [14] Oussaide`ne M, Chopard B, Pictet OV, Tomassini M, Parallel genetic programming and its application to trading model induction. *Parallel Comput* 23, pp. 1183–1198, 1997.
- [15] Sampler JL, Redefining industry structure for the information age. *Strategic Management Journal* Vol. 19(4), pp. 343–355, 1998.