

Novel Dynamic and Scalable Retention Data Storage Management Architecture

Kunal Kumar¹, Sushil Kumar², Manish Shrivastava³

¹*Dept. of Information Technology LNCT, Bhopal, (India)*

²*Dept. of Information Technology LNCT, Bhopal, (India)*

³*Dept. of Information Technology LNCT, Bhopal, (India)*

ABSTRACT

In the present situation, stockpiling administration of Big Data is forcing worry for Grid Computing conditions, as a substantial scale circulated calculation System which can resolve the issue of asset sharing. In conventional approach there is superior processing machine comprising of devoted servers that are utilized to store information stockpiling and asset disclosure. This paper broke down the day and age and measure of preparing information that were utilized as a part of applying the forecast capacity to estimate the Stock Exchange of Thailand record (SET Index). The preparation information was assembled into one year furthermore, two years over the time of 2004-2009, and the testing information were gathered from February 2006 to April 2010. This paper thought about the consequences of utilizing distinctive eras preparing information with a specific end goal to locate the best preparing informational collection that will most precisely anticipate the SET file. The outcomes demonstrate that there is no huge distinction between utilizing one year and two years preparing information.

Keywords – SET, GOS, MAPE, Stock Market Forecasting, Evolution Strategies, Prediction Function, Data Analysis.

I. INTRODUCTION

Information is developing at a colossal speed making it hard to deal with such substantial measure of information (Exabyte's). The primary trouble in taking care of such huge measure of information is on account of that the volume is expanding quickly in contrast with the registering assets. The Big information term which is being utilized now a day is somewhat misnomer as it brings up just the span of the information not putting excessively of thoughtfulness regarding its other existing properties. [13] "Huge information" alludes to datasets whose size is past the capacity of average database programming devices to catch, store, oversee, and examine. Huge information analytics is the territory where best in class explanatory strategies work on huge informational indexes. It is truly around two things, Big information and Analytics and how the two have collaborated to make a standout amongst the most significant patterns in business knowledge (BI). [9] Big information can be characterized with the accompanying properties related with it:

Variety: Information being created isn't of single class as it incorporates the customary information as well as the semi organized information from different assets like website pages, Web Log Files, web-based social networking destinations, email, archives, sensor gadgets information both from dynamic detached gadgets. This

information is entirely unexpected comprising of crude, organized, semi organized and even unstructured information which is hard to be taken care of by the current customary scientific frameworks.

Volume: The huge phrase in big statistics itself characterizes the volume. At show the facts current is in petabytes and must increment to zettabytes in close-by means of destiny. The casual conversation locales existing are themselves handling over statistics arranged by using terabytes regular and this degree of statistics is honestly hard to be handled making use of the modern-day customary frameworks.

Velocity: Pace in massive statistics is an idea which manages the rate of the information originating from distinctive sources. This trademark isn't being restrained to the speed of approaching statistics yet in addition velocity at which the records streams. As an instance the data from the sensor devices could be constantly shifting to the database save and this sum won't be sufficiently little. In this manner, standard frameworks aren't sufficiently in shape on gambling out the exam at the information that is usually in movement.

Variability: Fluctuation considers the irregularities of the data flow. Information hundreds become enormously difficult to be stored up specially with the growth in use of the web-primarily based social networking which for the maximum component reasons pinnacle in information masses with unique occasions happening.

Complexity: It is a great enterprise to connect, coordinate, wash down and alternate information crosswise over frameworks originating from exclusive sources. It is also important to interface and correspond connections, revolutionary systems and one of kind records linkages or statistics can swiftly winding wild.

The SET Index forecast has been considered by numerous specialists. They have been utilizing different strategies for example, Neural Network, Autoregressive Integrated Moving Average (ARIMA) and the Evolution Strategies. Neural systems are extremely productive versatile anticipating models in light of their astounding execution in self learning ability. Not at all like different strategies that build practical structures to speak to connections of information, neural systems can take in examples or connections from information itself. Be that as it may, with the impacts of black-box, moderate union, nearby ideal they are not material for a few applications. ARIMA was presented by G. Box and G. Jenkins in the mid-1970s. This time arrangement examination can catch complex entry designs, including those that are stationary, non-stationary, and regular (occasional) ones. ARIMA approach is exquisite in principle yet has been of little down to earth use in business due to its unpredictability and constrained increment in precision over less complex strategies. Development Strategies (ES) are calculation which mirrors the standards of regular development as a technique to illuminate parameter enhancement issues. ES is one of the most prevalent transformative calculations. ES are for the most part connected to numerical enhancement for its genuine esteemed portrayal.

1.1 Sources of Big Data

There are different wellsprings of huge information. Among them the real commitment of information is from ventures, value-based, online networking, information created because of movement, open information, files and so forth.,

Venture information: Large volumes of information as E-sends, spreadsheets, archives, pdf and so forth.

Value-based information: The exchange that happens through web applications, versatile applications, here the information is organized as applications are standard.

Web-based social networking: The information produced from interpersonal organizations like Facebook,

Twitter, Skype and so on., this are unstructured information as it contains content, recordings, pictures and sound.

Action created: this information incorporates the information produced from mobile phones towers, sensor information, restorative gadgets, satellites, information from aviation ventures and so forth.,

Open Data: The distributions made by government associations, inquire about foundations, meteorological office for climate anticipating, prompts open information got to by open.

Documents: Data which isn't often gotten to shapes the chronicled information e.g. bank exchanges, duplicates of assertion, filtered reports and so forth.

1.2 Storage Methods

Systems of storage which need larger capacity are generally made in dual ways:

1. Storage Area Network (SAN)
2. Distributed File System (DFS)

In direct terms, a SAN is a specific, quick system that joins servers and limit devices. The SAN is here and there insinuated as the system behind the servers. A SAN allows any type of relationship over the system; by using interconnect segments, for instance, switches. The SAN detects the standard committed relationship between a server and limit, and the possibility that the server sufficiently claims and manages the limit devices.

The SAN in like manner takes out any restriction to the measure of data that a server can get to. Generally, a server is confined by the amount of limit devices that attach to the individual server. Or maybe, a SAN familiarizes the versatility of systems organization with engage one server or various heterogeneous servers to share a run of the mill stockpiling utility. A system may join various limit devices, including plate, tape, and optical stockpiling. In addition, the limit utility might be arranged far from the servers that it jobs. The SAN can be seen as an extension to the limit transport thought. This thought engages stockpiling contraptions and servers to interconnect by using tantamount parts, for instance, LANs and WANs (Wide range systems). In a SAN, a lot of capacity are put in a capacity hub, for example, a RAID server, and these capacity gadgets are associated together utilizing a committed system (i.e. SAN). Notwithstanding, the physical stockpiling system is generally Ethernet construct running with respect to conventions, for example, iSCSI or FCoE, which includes milliseconds of programming and system inactivity.

An option association is to appropriate the capacity among the application has and utilize the broadly useful system an option association is to disperse the capacity among the application has and utilize the universally useful system alongside a conveyed document to give a record level sharing reflection.

While an appropriated record system is shoddy and adaptable, the product overhead of simultaneousness control and the high-idleness blockage inclined universally useful system corrupts execution. By and by, generally, these system, programming and convention latencies are middle of the road since they are inconsequential contrasted with look for inactivity of discs made up of magnet.

1.3 Benefits of Big Data Processing

Businesses can use outside knowledge while taking choices: Access to social information from web search tools and destinations like Facebook, twitter are empowering associations to adjust their business procedures.

Improved client benefit: Traditional client input frameworks are getting supplanted by new frameworks outlined with 'Enormous Data' advancements. In these new frameworks, Big Data and regular dialect handling advances are being utilized to peruse and assess buyer reactions.

Better operational productivity: 'Enormous Data' advancements can be utilized for making arranging zone or landing zone for new information before recognizing what information ought to be moved to the information distribution center. What's more, such joining of 'Huge Data' advancements and information distribution center causes association to offload rarely got to information.

1.4 Retention Period for Big Data

The capacity to boost limit while limiting expense is basic for a capacity stage working at enormous information scale. A major information maintenance period will bolster information is to be promptly and cost-adequately scaled even as the endeavor stockpiling requests develop significantly. The capacity to boost limit while limiting expense is basic for a capacity stage working at enormous information scale. The settled limit of conventional stockpiling frameworks and the need to adjust future stockpiling needs with current capital uses, numerous organizations are compelled to relocate their information to new frameworks like clockwork. Tragically for these ventures, this relocation is costly and tedious.

The discrete frameworks are essentially detached from each other; they promptly move toward becoming information storehouses that repress an enterprise's capacity to see the 10,000 foot view. To have the capacity to completely misuse the chances of huge information, organizations must have the capacity to access and utilize the greater part of their information without impromptu mediations. Dissimilar to regular stockpiling gadgets, a major information stockpiling stage spans inheritance stockpiling storehouse, instead of essentially adding yet another framework to the blend. Before enormous information, making sense of how to enhance information access for comprehensively dispersed clients exhibited a huge test to IT associations. Presently, the information itself has turned out to be all inclusive appropriated, changing the IT challenge into one of making it promptly accessible to clients and applications over the worldwide undertaking.

The utilization of Big Data for organize arranging and enhancement was additionally observed as a key activity, given a high positioning by 49 for each penny of respondents generally and 46.9 for every penny of administrators. This echo talks that we had with industry more than 2013, in which a few players recommended that system unpredictability instead of transfer speed may quick turn into the hindrance to arrange development.

II. RELATED WORK

Rajeev Agarwal et al. (2014) portrays layer based capacity administration design and work process of following provenance. This backing both organized and semi structured provenance gathering. Provenance caught away layer is a blend of information and process provenance. It can be utilized for various utilize cases. This approach is helpful at times and misses the mark in others. They consider organized and semi-organized information.

Outline of proposed design is basic so it may not take much overhead. They give numerous information display support to deal with organized and semi organized information and acquaint representation with have a reasonable view on provenance data. MongoDB database is utilized which bolster versatility. They incorporate client explanation bolster which gives each client adaptability to include some custom documentation in provenance data. Every client can compose appropriate documentation as indicated by his need. Provenance start is client controlled however whole process is computerized. Their point is to catch provenance near crude configuration. They have acquainted access control with save security and protection of provenance information of every client. Head can picture provenance data of the considerable number of clients. Clients can just imagine provenance of their own. In spite of the fact that for ensuring provenance data just access control isn't sufficient, they may require some extra approaches. However, legitimate approval can give insurance of provenance data up to some degree. The enormous information blast has seen various NoSQL information stores for adaptability. In our paper, they utilize MongoDB which bolster both organized and semi organized provenance report. The way they oversee provenance information, it will be simple for the client to know where information objects are put away. Clients can check the NoSQL information store utilizing general articulations to scan for past occupation keeps running for an assortment of data [1].

Xueli Huang et al. (2014) proposed a productive plan to address the expanding worry of information protection in cloud for picture information to advance the distributed computing as an answer for huge information. Their plan isolates a picture into pieces and rearranges the squares with irregular begin position and arbitrary walk and works at the piece level rather than the pixel level, which incredibly accelerates the calculation. They changed over the picture security issue into the jigsaw perplex issue. To influence the jigsaw to perplex issue NP-finish, they adjusted every pixel of the picture information by utilizing arbitrary balanced mapping capacity. These operations make the pairwise linking un-dependable and make the rearranged picture un-unmistakable. We actualized our plan in genuine systems (counting the Amazon EC2) and tried the security and effectiveness. Both our investigation and trial comes about demonstrated that our plan is secure, proficient and has little overhead [2].

Ganapathy Mani et al. (2014) proposed a moderate and versatile PC association configuration to treat a mental imbalance, where the advance observing can dispense with the risk of fixation. The proposed design depends on the idea of the Big Data computational model that chooses just a little piece of data to work expressly while whatever remains of data adds to this determination just certainly. Understanding the data handling model of the cerebrum is key with a specific end goal to regard mental clutters, for example, extreme introvertedness. Developmental science attests that the obviousness is key for the usefulness of the mind. The human mind comprises of billions of neurons and trillions of neurotransmitters. Its dynamic memory is only 1 gigabit. However, it can forms and recovers data speedier than a PC. Given the little dynamic memory however productive handling, they guess that the mind's data preparing is impacted by the obviousness or foundation data. Mental disarranges happen because of twists in this foundation data. Essential twists are nearness of over the top and undesirable data out of sight, and nonattendance of key data. Discouragement is a case for the first (e.g. horrible accidents put away in foundation). Extreme introvertedness is a case of the later. Their recommended approach will have an Oracle machine—a qualified therapist or chief PC program—will give particular recommendations to the individual in view of their sort of mental issue and regulate their choices. The Oracle

machine will indicate individual the perfect decision in the event that they have chosen a wrong decision. Given the data handling model of the cerebrum, these recommended determinations and perfect decisions among the choices would probably influence the obviousness of the mind in this manner embeddings new data in the relevant foundation of the mind. After a specific timeframe, they hope to see the upgrades in the conduct of the person with a mental imbalance [3].

Yu-Neng Fan et al. (2013)proposes an Incremental Rough Set-based Rule Induction Agent (IRSRIA). Manage acceptance depends on making specialists for the primary displaying forms. Moreover, an incremental engineering is composed, to address extensive scale dynamic database issues. A contextual analysis of a Home shopping organization is utilized to demonstrate the legitimacy and productivity of this technique. The primary reason for this exploration is to execute manage acceptance on incremental information, utilizing shrewd specialists inside the fundamental procedures of govern enlistment, for example, reduct estimation, noteworthiness record (SI) calculation, run enlistment and control pruning. In this investigation, five specialists are utilized: the status assurance operator, the reduct age operator, the SI calculation specialist, the administer acceptance operator and the run pruning specialist. These five operators must be utilized as a part of arrangement. Since the proposed strategy is a multi-specialist framework, the operators are accessible to finish the errands in parallel, despite the fact that these operators must be executed in succession, the specialists can at present run at the same time. This parallel property of the operator makes control acceptance more effective. The aftereffects of examinations demonstrate that the IRSRIA can extensively lessen the calculation time for initiating choice standards, while keeping up a similar nature of principles [4].

Shankar Ganes h Manikandan et al. (2014)proposed different strategies for taking into account the troubles nearby thru Map Reduce structure over Hadoop Distributed File System (HDFS). Guide Reduce is a Minimization technique which makes usage of report ordering with mapping, arranging, rearranging ultimately lessening. The need to process large quantities of facts has in no way been more prominent. Not exclusively are terabyte-and petabyte-scale datasets fast attending to be evidently normal, but there may be accord that tremendous esteem lies blanketed in them, holding up to be opened by using the privilege computational devices. Huge Data research apparatuses like Map Reduce over Hadoop and HDFS, ensures to enable associations to better comprehend their clients and the economic centre, ideally prompting higher business choices and top fingers. For engineers building records making ready apparatuses and programs, considerable and heterogeneous datasets which can be developing ceaseless flow of data, prompt greater successful calculations for an in-depth variety of assignments, from machine translation to spam detection. The capacity to examine massive amounts of information may provide the way to opening the secrets of the universe or the puzzles of life. MapReduce can be abused to take care of an assortment of issues identified with content handling at scales that would have been inconceivable a couple of years back [5].

Panagiotis D.Diamantoulakis et al. (2015) summarized the state-of-the-art in the exploitation of big data tools for dynamic energy management in smart grid platforms. In order to deal with the extreme size of data, the smart grid requires the adoption of advanced data analytics, big data management, and powerful monitoring techniques. Next, they elaborated on the utilization of the most commonly used smart grid data mining and predictive analytics methods, focusing on the smart meter data that are necessary for the accurate and efficient power consumption/supply forecasting. They proceeded with a brief survey on the works dealing with high

performance computing, insisting on cost efficiency and security issues in the context of SG control. Finally, they discussed several interesting techniques and methods that have to be further explored into the framework of a real-time monitoring and forecasting system, and provided promising research directions for future research in the field [6].

Marcos D. Assunção et al. (2015) talks about methodologies and situations for doing investigation on Clouds for Big Data applications. It rotates around four imperative zones of investigation and Big Data, to be specific (i) information administration and supporting structures; (ii) show improvement and scoring; (iii) representation and client collaboration; and (iv) plans of action. The measure of information presently created by the different exercises of the general public has never been so enormous, and is being produced in a consistently expanding speed. This Big Data incline is being seen by enterprises as a method for acquiring advantage over their rivals: on the off chance that one business can comprehend the data contained in the information sensibly snappier, it will have the capacity to get more costumers, increment the income per client, streamline its operation, and decrease its expenses. In any case, Big Data investigation is as yet a testing and time requesting errand that requires costly programming, vast computational foundation, and exertion. Distributed computing helps in lightening these issues by giving assets on-request costs corresponding to the real use. Moreover, it empowers frameworks to be scaled here and there quickly, adjusting the framework to the real request. There are a lot of answers for Big Data identified with Cloud processing. Such a substantial number of arrangements have been made in light of the extensive variety of investigation prerequisites, however they may, at times, overpower non-experienced clients. Examination can be elucidating, prescient, and prescriptive; Big Data can have different levels of assortment, speed, volume, and veracity. In this way, it is vital to comprehend the necessities so as to pick suitable Big Data apparatuses; It is likewise evident that examination is a mind-boggling process that requests individuals with aptitude in tidying up information, understanding and choosing legitimate techniques, and breaking down outcomes [7].

Vijay Gadepally et al. (2014) proposed a procedure to comprehend the auxiliary qualities of a database called dimensional information examination. Utilizing DDA, a specialist can take in an extraordinary arrangement about the concealed examples, basic qualities and conceivable blunders of a huge obscure database. DDA comprises of speaking to a dataset utilizing affiliated clusters and playing out an examination between the constituent acquainted exhibits and planned perfect database exhibits. Deviations from the expected model can feature essential points of interest or wrong data. They prescribe that the DDA system be the initial step of an investigative pipeline. The basic following stages in a scientific pipeline, for example, foundation displaying, highlight extraction, machine learning and visual investigation depend vigorously on the nature of information. Following stages to this work incorporate building up a robotized component to perform foundation displaying of enormous datasets, and use of identification hypothesis to huge informational collections [8].

III. PROPOSED WORK

The proposed architecture named “Novel Dynamic and Scalable Retention Data Storage Management (NDSSM)” to sort out Grid stockpiling gadgets into an expansive scale and topographically disseminated capacity framework to meet the necessities forced by a wide range of Grid applications.

Building up an information maintenance approach for holding data is imperative for operational or administrative consistence and in addition for meeting lawful and business information recorded prerequisites. Here there is a case SET Index for huge information, on the grounds that in stock trade there is different sorts of message exchange is going to put. The Thai securities exchange some time ago known as the Security Exchange of Thailand. In 1991, its names were changed to Thai Stock Exchange of Thailand (SET). The file of the stock trade of Thailand is known as SET file. SET record is composite market capitalization-weighted value list, which looks at the present market estimation of all recorded regular stock with its incentive on a solitary base date when the stock trade was built up. In this trial, the information was getting a handle on from the different dependable sources, the Bank of Thailand, which contain the chronicled information of SET file, the Dow Jones list (New York), the Nikkei file (Japan), the Hang Seng (Hong Kong), and Minimum Loan Rate (the conversion standard of the Thai Baht and the US dollar, day by day powerful finished night government subsidize rates in the USA). Since the crude information is from various securities exchange in the different nations, there exchanging days and occasions was extraordinary. Along these lines, those holes were loaded with the information from the earlier day with no critical contrasts. The supposition hidden this investigation was that the missing information on non-exchanging day will be satisfied with the past information. The MAPE (Mean Average Percentage Error) of two years is more helpful and smoother concerning MAPE of one year. The more drawn out time of preparing information can yield bring down blunder, with the exception of the most reduced mistake level. The mistake of your information is smidgen lower than two years preparing information, yet it isn't noteworthy.

As far as attempting to deliver the correct approach, it might be best to concentrate on the capacity of extensive informational indexes to yield comes about that exceed the legitimate and security worries, notwithstanding deciding the maintenance time, documented principles, information designs, and the allowable methods for capacity, access, and encryption, for example. Covering approach territories may cover matters on development and advancement, to add understanding to the maintenance and administration abilities. For imagining and associating with information we required a maintenance period. We propose the maintenance time of Big Data Management is 2 years. In light of, before this day and age, information isn't applicable and it's not useless to the administration of the information administration.

So, there must be a maintenance period for Big Data. Information before the period ought to be in the compacted frame. So, seeking ought to be quicker. On the off chance that, somebody needs the past information, there is a component to decompress them.

IV. EXPERIMENTAL RESULTS

The exploratory information was gathered from a dependable source, the Bank of Thailand, which comprised of recorded information of the SET list, the Dow Jones file, the Nikkei list, the Hang Seng file, and Minimum Loan Rate. Since the crude information were gotten from various securities exchanges in distinctive nations, accordingly a few information was absent since every nation has distinctive securities exchange occasions or non-exchanging days. Be that as it may, those holes can be filled by utilizing the information from earlier day with no measurably critical contrast. In this way, the suspicion basic this investigation was that the missing

information on non-exchanging days will be loaded with earlier day's information.

The preparation information in this examination was isolated into two diverse time ranges: one year and two years preparing information. The one year preparing information comprised of six diverse eras: 2004, 2005, 2006, 2007, 2008 and 2009, and the two years preparing information comprised of five distinctive eras: 2004-2005, 2005-2006, 2006-2007, 2007-2008 and 2008-2009. The test information utilized as a part of this analysis was gathered from February 2006 - April 2010, 1540 days in add up to.

The explored time arrangement began from February 2006 to April 2010. It contained 1540 days of test information. The MAPE at different $b(0)$ from February 2006 to April 2010 with one year training data scale from 0.90 to 1.04 is shown in figure 1 and MAPE at different $b(0)$ from February 2006 to April 2010 with two years training data scale from 0.90 to 1.04 is shown in figure 2.

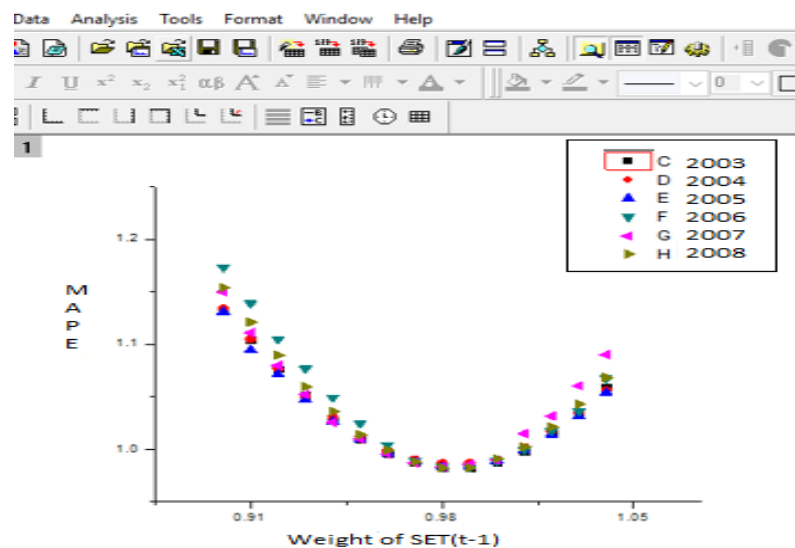


Figure 1: MAPE at different $b(0)$ from Jan 2003 to April 2008 with one year training data scale from 0.90 to 1.04

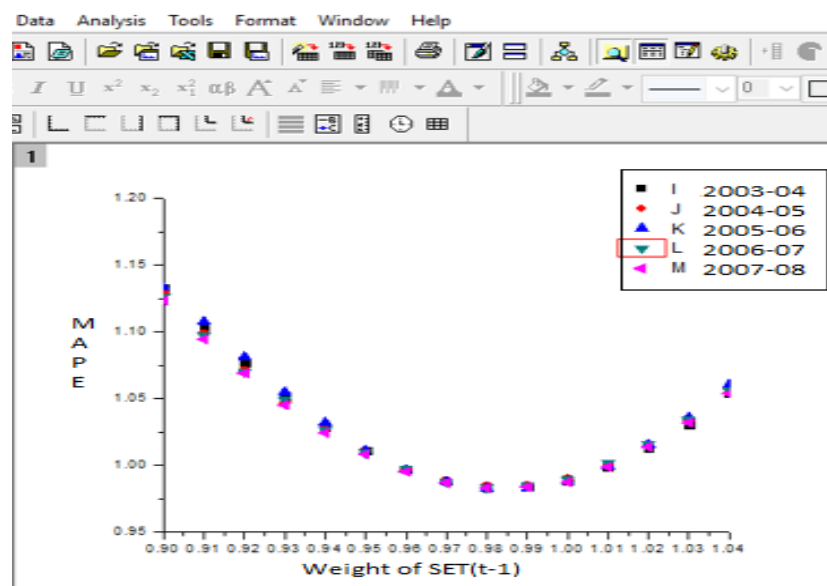


Figure 2: MAPE at different b(0) from Jan 2003 to April 2008 with two years training data scale from 0.90 to 1.04

V. CONCLUSION

This paper investigated the day and age and measure of preparing information that were utilized as a part of applying the forecast capacity to estimate the Stock Exchange of Thailand file (SET Index). The investigation was directed by utilizing one year and two years preparing information over the time of 2003 to 2008. The outcomes demonstrate that there is no noteworthy distinction between utilizing distinctive years of preparing information, despite the fact that many emergencies occurred in Thailand, for example, the overthrow in 2006, the counter government dissent and the subprime contract emergency in 2008. What's more, the outcomes additionally demonstrate that there was no noteworthy contrast between utilizing one year and two years preparing information. Be that as it may, utilizing two years preparing information required more calculation time than one year preparing information, so one year preparing information is extensively adequate. All in all, the preparation information utilized for SET file forecast can be gotten from whenever period amid 2003-2008 and can be it is possible that one year or two years with no huge contrast as far as forecast precision.

REFERENCES

- [1] Rajeev Agrawal, Ashiq Imran, Cameron Seay, Jessie Walker, "A Layer Based Architecture for Provenance in Big Data", IEEE International Conference on Big Data, 2014.
- [2] Xueli Huang, Xiaojiang Du, "Achieving Big Data Privacy via Hybrid Cloud", IEEE INFOCOM Workshop on Security and Privacy in Big Data, 2014, pp. 512-517.
- [3] Ganapathy Mani, Simon Berkovich, Duoduo Liao, "Adaptive and Interactive Design based on Big Data Computational Model for Treating Autism", IEEE Fifth International Conference on Computing for Geospatial Research and Application, 2014, pp. 121-122.
- [4] Yu-Neng Fan, Ching-Chin Chern, "An Agent Model for Incremental Rough Set-based Rule Induction: A Big Data Analysis in Sales Promotion", IEEE 46th Hawaii International Conference on System Sciences, 2013, pp. 985-994.
- [5] Shankar Ganes h Manikandan, Siddarth Ravi, "Big Data Analysis using Apache Hadoop", IEEE, 2014.
- [6] Panagiotis D.Diamantoulakis, Vasileios M.Kapinas, George K.Karagiannidis, "Big Data Analytics for Dynamic Energy Management in Smart Grids", Science Direct, Big Data Research 2, 2015, pp. 94-101.
- [7] Marcos D. Assunção, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, Rajkumar Buyya, "Big Data computing and clouds: Trends and future directions", Science Direct, J. Parallel Distrib. Comput. 79-80, 2015, pp. 3-15.
- [8] Vijay Gadepally & Jeremy Kepner, Big Data Dimensional Analysis, IEEE, 2014.
- [9] Avita Katal, Mohammad Wazid, R H Goudar, "Big Data: Issues, Challenges, Tools and Good Practices", IEEE, 2014, pp. 404-409.
- [10] Changqing Ji, Yu Li, Wenming Qiu, Uchechukwu Awada, Keqiu Li, "Big Data Processing in Cloud Computing Environments", IEEE International Symposium on Pervasive Systems, Algorithms and Networks, 2012, pp. 17-23.

- [11] M. Fazio, A. Celesti, A. Puliafito, M. Villari, "Big Data Storage in the Cloud for Smart Environment Monitoring", Science Direct, The 6th International Conference on Ambient Systems, Networks and Technologies (ANT2015), pp. 500-506.
- [12] Jenny Weisenberg Williams, Kareem S. Aggour, John Interrante, Justin McHugh, Eric Pool, "Bridging High Velocity and High Volume Industrial Big Data Through Distributed In-Memory Storage & Analytics", IEEE International Conference on Big Data, 2014, pp. 932-941.
- [13] Katarina Grolinger, Michael Hayes, Wilson A. Higashino, Alexandra L'Heureux David S. Allison, Miriam A.M. Capretz, "Challenges for MapReduce in Big Data", IEEE 10th World Congress on Services, 2014, pp. 182-189.
- [14] Anita Gupta, Abhay Mehrotra, P. M. Khan, "Challenges of Cloud Computing & Big Data Analytics", IEEE, 2015, pp. 1112-1115.
- [15] David C. Chou, "Cloud computing risk and audit issues", Science Direct, Computer Standards & Interfaces 42, 2015, pp. 137-142.
- [16] Imran Khan, Mansaf Alam, S. K. Naqvi, S. N. A Rizvi, "Data Model for Big Data in Cloud Environment", IEEE, 2015, pp. 582-585.
- [17] Omar Y. Al-Jarrah, Paul D. Yoo, Sami Muhaidat, George K. Karagiannidis, Kamal Tahaa, "Efficient Machine Learning for Big Data: A Review", Science Direct Big Data Research 2, 2015, pp. 87-93.
- [18] FAN ZHANG, WEI TAN, KEQIN LI, "Evolutionary Scheduling of Dynamic Multitasking Workloads for Big-Data Analytics in Elastic Cloud", IEEE, 2014, pp. 338-351.
- [19] Du Zhang, "Inconsistencies in Big Data", IEEE, 2013, pp. 61-67.
- [20] Nagarathna N, Indiramma M and Jyothi S Nayak, "Optimal Service Selection Using Trust Based Recommendation System for Service-Oriented Grid", IEEE International Symposium on Cloud and Services Computing, 2012, pp. 101-106.
- [21] Xiaoyun Zhou, Xiongpai Qin, Keqin Li, "Parallel Techniques for Large Data Analysis in the New Version of a Futures Trading Evaluation Service", Science Direct Big Data Research 2, 2015, pp. 102-109.
- [22] Shweta Pandey, Dr. Vrinda Tokekar, "Prominence of MapReduce in BIG DATA Processing", IEEE Fourth International Conference on Communication Systems and Network Technologies, 2014, pp. 555-560.
- [23] Ekaterina Olshannikova, Aleksandr Ometov, Yevgeni Koucheryavy, "Towards Big Data Visualization for Augmented Reality", IEEE 16th Conference on Business Informatics, 2014, pp. 33-37.
- [24] Cameron Seay, Rajeev Agrawal, Anirudh Kadadi, Yannick Barel, "Using Hadoop on the Mainframe: A Big Solution for the Challenges of Big Data", IEEE 12th International Conference on Information Technology - New Generations, 2015, pp. 765-769.