International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 03 | Mar 2023 www.irjet.net p-ISSN: 2395-0072

Study of Data Analysis Model Based on Big Data Technology

Shraddha Sanjay Paralkar[1], Shubham Sadashiv Dhage[2], Arshad Shaukat Mulani[3], Asst.prof.S.V. Thorat[4]

[1st]Shraddha Sanjay Paralkar, MCA YTC, Satara
[2nd]Shubham Sadashiv Dhage, MCA YTC Satara
[3rd] Arshad Shaukat Mulani,
[4] Prof. S.V. Thorat Dept. of MCA Yashoda Technical Campus,Satara-415003

Abstract — The traditional data analysis are grounded on the cause and effect relationship, formed a sample bitsy analysis, qualitative and quantitative analysis, the thinking mode of trend extrapolation analysis. Big data has a abecedarian impact on the traditional data analysis. Big data analysis grounded on correlation, formed global macro analysis ' data and specialized analysis ' correlation analysis and new thinking mode of correlation analysis. videlicet, from unproductive analysis to correlation analysis and knowledge discovery, from model fitting to data mining, from logical logic to association rules. Data analysis in the period of big data have taken great changes, videlicet, Big data analysis, from the analysis of objects, the mode of data processing, logical styles and tools, logical thinking.

Keywords - big data; data analysis; qualitative and quantitative analysis

1. INTRODUCTION

Big data is one of the world's hottest vocabularies after the Internet of effects and pall computing. Big data has brought a great impact. On allowing mode, education model, business operation model, scientific exploration model and medical individual model, etc. Big data has a abecedarian impact on all fields. Traditional data analysis has been developed from the analysis of the sample of " To see only one spot" into the time of overall analysis of " the overall situation ". Traditional data analysis of small data allowing model and fine model has been delicate to acclimatize to the data processing requirements of large data period. Chancing the knowledge, mining value, looking for association is the real need of data analysis in the period of bigdata.However, discarding the rubbish and elect the essential, But analysis of the age of big

If the traditional data analysis is the nuggets from the mine.

data is taking the gold from beach, discarding the false and retain the true beach began to see gold ". " Blowing beach only see gold " and " discover order from chaos " can be said that the most true depiction of data analysis of the period of big data analysis.

2. DATA ANALYSIS

A. Summary of Data Analysis

The connotation of data analysis Data analysis has a broad and narrow sense, Generalized data analysis refers to the sorting, sorting, sorting, organizing, storing, recycling, assaying and studying on the base of collecting and enwrapping the data, the whole process of discovering new knowledge. Narrow data analysis refers to the data analysis of the colorful links, similar as sorting, sorting, screening, association, storehouse, processing, analysis and exploration, etc. Data analysis is the identification, of the original data and the data collected through the collection.

Mining rules, intelligence and knowledge hidden in the data, which are give a prophetic , scientific, and comprehensive and vacuity conclusion or plan, for operation and decision making services

Data analysis has a different understanding of different disciplines. But the substance is the same. In the field of Statistics, Data analysis is generally interpreted as a data analysis or statistical analysis; In the field of information wisdom and data operation, Data analysis is generally understood as information analysis or information exploration; In the field of Computer Science, Data analysis is generally interpreted as data mining or knowledge discovery.

rudiments of data analysis From the view of the conception of data analysis, data analysis is an organic total that composed of a series of factors, similar as origin, substance, system, process, result and purpose. From the view of the substance, Data analysis is the discovery of the nature, characteristics, attributes, rules and associations from the data miracle. From the view of origin, data analysis comes from the demand of social data; From the view of process, data analysis needs a series of links and procedures to collect, sort, elect, organize, storehouse, processing, analysis and exploration, just draw a scientific and dependable conclusion; From the view of system, data analysis system can be divided into qualitative analysis system and quantitative analysis system, which composed

of scientific thinking system, statistical system, sociological system, information wisdom system. From the view of achievements, Data analysis process will produce new value- added products, videlicet knowledge, intelligence, scheme, report, etc; From the view of ideal, Data analysis is substantially for scientific operation and scientific decision-making services

Object of data analysis There are two main types of data analysis A class of numerical data, substantially refers to the original and deduced data, The purpose is to discover knowledge, intelligence, wisdom and law from the data through quantitative analysis system; A class of non-numerical data, substantially refers to effects and their marvels, the purpose is to find out the substance, trait, characteristic, rule and relation of the thing from the miracle through the qualitative analysis system.

Function of data analysis Data analysis plays a data collation, objective evaluation, trend vaticination, data feedback, and other introductory functions in scientific operation and scientific decision- timber, which Plays an important part in the identification and selection, arrangement and sequencing, monitoring and early warning, as well as staff and navigation.

B) ***Data analysis model***

Principle of data analysis Data analysis is grounded on the attributes, characteristics, nature, law and correlation of data to expand the qualitative and quantitative analysis, in order to discover new knowledge. thus, Data analysis is grounded on the unproductive relationship or correlation between effects, marvels and data. Relationship refers to the correlation between effects due to time, order, structure, movement and so on, including time, space, circumstance and development sense. The relationship between effects, marvels and detail is veritably Complex and different. But it can be classified as two kinds of query relation and certainty relation. query relation is substantially the affiliated relationship, which is the base of qualitative analysis; while the certainty relation is substantially quantitative relation, which is the base of quantitative analysis.

Dialectical materialism tells us the world is universal and no independent actuality of the miracle and effects. Small world miracle(six degrees of separation proposition) and social network analysis system tell us that between people is generally and through a variety of connections to forming social networks. Meanwhile, everything always happens and develops in a certain time and space, which has egregious heritage and development and show a logicalrelationship.The universal actuality of effects, marvels and data is the base of data analysis. Although some connections are direct and significant and easy to find, and some connections are circular and implicit relations, it's delicate to find. Because of time, these connections may have a cause and effect relationship.

The thinking mode of data analysis For a long time, the data analysis substantially follows three introductory ideas, videlicet sample and population, qualitative and quantitative, trend extrapolation, which formed a set of allowing mode and has played an important places in the" small data" analysis of the times.

a) Sample bitsy analysis Data analysis takes the data and the miracle as the objects, It's generally named from the whole or part of the overall samples for analysis and be called sample analysis or slice analysis.

b) Qualitative and Quantitative analysis Its grounded on correlation. Sample's nature, law, characteristic, trait and relation of sample are anatomized by qualitative system; Its grounded on cause and effect, the characteristics, laws and relations of the samples were quantitatively described or fitted by fine and statistical models. Quantitative connections between samples are generally not rigorously functional, but the approximate function relationship, which need to use function relation to roughly describe the relationship.

Trend extrapolation analysis Grounded on the qualitative and quantitative analysis, the nature, the rule, the characteristic, the trait and the relation of the samples are attained, and the tendency is decided to the whole or the population, and the overall vaticination or estimation is carried out.

system and tool for data analysis Data analysis styles are substantially deduced from the sense system, system analysis system, quantitative, sociological system, statistical system, fine system, which generally divided into three situations of philosophical styles, general styles and specific styles. Concrete analysis system is generally also divided into three types qualitative system, quantitative system and semi quantitative system. The qualitative styles substantially have logical thinking and scientific thinking system, which Included bracket and comparison, analysis and conflation, induction and deduction, analogy and imagination, etc. The quantitative styles substantially have multivariate analysis system(similar as correlation analysis, retrogression analysis, cluster analysis,etc.), Time series analysis(similar as moving average,

exponential smoothing, direct trend, seasonal indicator,etc.), Literature dimension system, etc. Semi quantitative system substantially includes content analysis system, logical scale process, Delphi system, etc. There are four main types of tools for data analysis First, social check and expert check tools second, logical thinking tool; third, Mathematical and statistical models; Forth, Date base and computer data mining tools.

These styles and tools can dissect data, data and marvels from different perspectives and position and give the necessary qualitative and quantitative base for scientific operation and scientific decision making

3. BIG DATA ANALYSIS

A. *Big Data Overview*

Generation and development of large data The generation and development of big data has endured three stages of development. From 1980s to middle of the 90s, that's the embryonic stage of big data. In 1980, The visionary of Alvin Toffler of America thinks that big date will be praised as" the third surge of the cadenza in the" third surge". In the middle of 1990s to the first 10 times of twenty-first Century, Big date is extensively concerned stage. Big data has come a hot content in the field of colorful diligence and disciplines. Big data, this language can be traced back to the org Apach's open source design — Nutch. At that time, big data was used to describe a large number of data t sets that need to be reused or anatomized at the same time to modernize the network hunt. September 2008, Nature magazine published" Big Data Science in the petabyte period Big" series of Special papers and the conception of" big data" was put forward. Then'the big data has come popular word in the IT assiduity. Academia, assiduity and government have given a high degree of concern. President of the United States Science and Technology Advisory Committee gave President Obama and Congress a report that entitled" The future of digital planning". In 2011, wisdom also launched Special columns about" Dealing with Date ", which bandied the significance of in scientific exploration and operation of data. In June of the same time, McKinsey & company released a detailed report about big data, videlicet" Big Data The coming frontier for invention, competition, and productivity(big data invention, competition and the coming frontier of productivity), which was carried out a detailed analysis in impact on big data, crucial technology and operation fields, etc. IBM, Microsoft, Apple of IT titans have enforced big data plans and systems, which are trying to enthrall the commanding elevation in the field of large data. After 2012, big data pours into the rapid-fire development stage. The United States, Japan, other countries and the European Union have put forward the response measures about the development of large data. China is also laboriously involved in. Of February

2012, The United States Obama government published " big data exploration and development proffers ", planned to use big data in the field of biology, technology, drug and other fields. March, Davos World Economic Forum released" big data, big impact"; In May, the United Nations Secretary General's office has issued" big data to promote development challenges and openings"; June, The ninth session of the OECD Statistics Committee issued a exploration report- Use big data for decision timber; In July, The Japanese Ministry of internal affairs put new comprehensive strategy for CIT, videlicet" the exertion of CIT in Japan, the focus on big data operations. In January 2013, the British government blazoned that it would invest1.89 Billion pound in the field of observation, Medical and health work of large data and energy saving calculating technology. The development and exploration of big date got into the climax and in our country is also a hot. The time of 2011, is China's first time of big data. 2012 is China's big data important time. colorful kinds of big data forum held constantly and a variety of large data systems, planning, reporting, and strategy were surfaced One after another. 2013 is named by the first time of China's big data statistics. In November 2013, The National Bureau of Statistics, Ali, Baidu and other 11 companies inked a big data strategic cooperation frame agreement, which has put big data to the peak. At the morning of 2013, The Ministry of wisdom and technology of China blazoned the time 2014" National crucial introductory exploration and development plan(videlicet 973 Plan, including major scientific exploration design", among this," the Research on the base of large data calculating" come an important direction to support.

The characteristics of the data The computer wisdom and artificial intelligence laboratory at the Massachusetts Institute of Technology professor Sam Madden first summarizes the" 3v" characteristics of big data, videlicet the volume, variety, haste. IDC holds that the characteristics should also add value. IBM considers that big data should also include veracity. Forrester critic Brian Hopkins and WeiErSong epitomize the characteristics of the big data as mass, diversity, high speed and variability. Overall, big data has the characteristics of " 6v 1c", videlicet the large volume of data(Volume), the variety of type(Variety), the fast processing haste(haste), the large operation value(Value), carrying and transferring freely and flexibly(Vender), the veracity(Veracity), Great difficulty in processing and analysis(Complexity). presently, colorful diligence have different interpretation on the characteristics of big data. The" 4v" characteristics of big data, videlicet the volume(large capacity), variety(colorful types), haste(high speed) and the most important value(low viscosity), are widely honored

B. The Model of Big Data Analysis

The arrival of the period of big data has changed the thinking mode of traditional data analysis. In the period of big data, we need not only the traditional, micro data analysis grounded on a sample, but also the ultramodern, macro data analysis grounded on the overall. 1) The proposition of big data analysis The data analysis of the period of big data can be called the big data analysis, which

substantially follows three introductory generalities. First, concentrate on all not slice Big data analysis is the macro data analysis, which needs to completely observe the substance, characteristics, attributes, laws and contact of the overall, rather than Sample to ramify the connection between detail or marvels. Second, concentrate on correlation not reason In the period of big data, face the challenge of huge quantities of data, knowing what's more important than knowing why. similar as stock data, it's easy to know whether it rises or falls according to the big data analysis, but it's hard to know why it can rise or fall. The typical task of big data analysis is to realize pattern mining and vaticination analysis through correlation. Big data analysis emphasizes set up we should find the new patterns we do not know in advance and the unknown correlation. Third, concentrate on effectiveness not delicacy In the period of big data, time and cost is more meaningful than the accurate results. Because of big data analysis to all or overall as an object, it's nearly insolvable to find a suitable statistical or fine model to describe the all or overall characteristic, chronicity, andcontact. However, time and cost must be amazing, If any. At the same time, it's delicate to directly or intimately set up all or overall substance, parcels, characteristics, chronicity, and contact. 2) Big data logical allowing mode Big data analysis focuses on data analysis, onmulti-source data emulsion, emphasizes on correlation analysis as the core, has formed a new mode of thinking. That's from unproductive analysis to the correlation analysis and knowledge discovery, from model fitting to data mining, from logical logic to association rule timber. a) The whole and macroscopic analysis Big data takes all the data or overall as the analysis object, and the data is the core and key. The nature, trait, characteristic, rule and relation of big data should be observed on the whole and macro. b) Data and specialized analysis Big data takes data and technology(computer technology and network technology) as the core, takes database, data mining and knowledge discovery algorithm as tools. The emphasis is association discovery(18). c) Correlation analysis and knowledge discovery Big data is grounded on the correlation relationship rather than reason, and focuses on the retired rules, links and values of the data. 3) crucial technologies of large data analysis The core of big data analysis is big data technologies, which is a collection of precious data from colorful types of massive data. The crucial technology of big data analysis substantially include data accession, data access technology, structure, data processing, statistical analysis, data mining technology, model vaticination technology, and the present technology.

4.NEW TREND OF DATA ANALYSIS MODEL AND DEVELOPMENT BASED ON BIG DATA

A. Data Analysis Model Based on Big Data

Due to the large data analysis and traditional data analysis has the difference in the analysis of the object, foundation, Patterns and analysis of the results and other aspects. thus, in the period of big data needs tore-build a large data analysis model. Big data analysis model includes large number of accession and collection, processing and processing, dispersion and sharing of analysis, service and application, and so on. Data sources stem from mortal conditioning, computer, network and the physical world leaves the track. At present, these large data substantially through hunt machine and the data inflow machine, database machine or middleware, or ETI machine accession and collection to form a set of target data. also, it uses the big data platform to carry on the real time processing(including the static data and the dynamic on- line data batch processing or the structure, the semi structure, thenon-structure batch processing). Eventually, it shows the visual display, to give services and uses. Data analysis model with large data platform grounded by data correlation and data association mining algorithm to deal with comprehensive data, and visual display, to give support for the operation and decision- timber. As shown Figure 1.

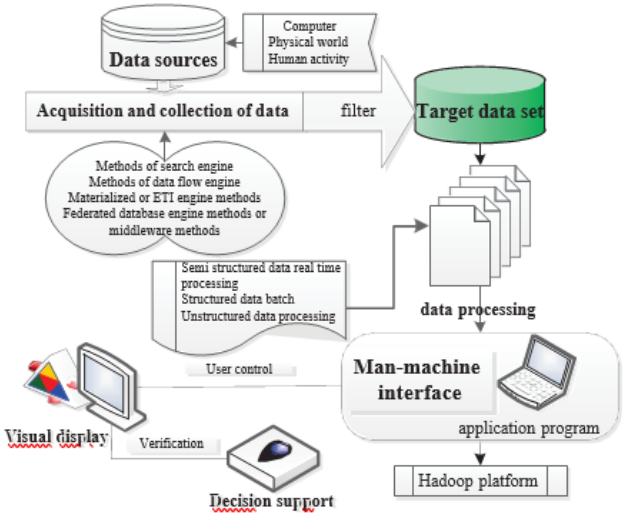

Figure 2. Data analysis model based on big data

B. New Trends in the Development of Large Data Analysis

Data analysis is the introductory direction of the period of big data analysis of the development and with nonstop expansion of large data capacity. Big data analysis process and analysis technology, analysis styles and analysis models showing some new trends

First, large data accession including the accurate selection of data sources, high quality raw data accession styles,multi-source data processing styles, data form and automatic correction styles.

Alternate, large data processing including a larger quantum of data analysis and mining styles, large data real- time processing, big data analysis and mining algorithm to ameliorate.

Third, large data visualization including image analysis, mortal computer commerce, scalability andmulti-level issues, visualization and automatic data mining combined with the visualization tool for the millions.

Fourth, big data security contains APT attacks, social network sequestration protection, threat adaptive access control, data accession, storehouse, analysis of 3 independent process.

There are other big data effective high- speed transmission system, large data virtual machine exploration, super computer links to join, big data gift training,etc.

5. CONCLUSION

Data analysis is an important process of research or simply discovering information related to any work. Data derived from the observation, experiment, and other primary and secondary data collection methods is large and cannot be taken as it is. Not all data is relevant, neither can it directly signify any trends, relations, facts, and associations within the data. To find out those required trends and relations, the data needs to be reconstructed in the relevant form and modified. This process is called data analysis. Data analysis and conclusion take forward the research.

6. REFERENCES

1. Huang Yihua. Deep understanding of big data: big data processingand programming practice. Machinery Industry Press, 2014.

2. Li Baodong , Song Hantao. Research status and development of data mining language. computer engineering and application, pp.78- 81,2003.

3. Davenport series, Wu Junshen translation. Big data analysis: data driven enterprise performance optimization, process management and operation decision. Machinery Industry Press, pp.67-68,2015

4. Luan Wenpeng, Yu Yixin, Wang Bing. AMI data analysis method.. proceedings of the Chinese society of electrical engineering, pp.178- 189,2015.

5. Zhang Xiaoyu, Zou Kai. The research progress of the big data in the field of Library and Information Science in China . library science research, pp.45-51,2015.

6. Agnes Wa. Subversion big data analysis: Based on Storm, Hadoop and other Spark alternative technology in real time applications, the electronics industry press, pp.76-80,2015.

7. Sun Qiunian, Rao yuan. Research on network data visualization technology based on association analysis]. computer science, pp.67- 86,2015.

8. Party Qian Na, Luo Tianyu. Multidimensional data evolution in the field of technological innovation, frontier and characteristics . Science science and management of science and technology. pp.34-40,2015.

9. Guo Chong. Based on large data analysis of online shopping customer loyalty modeling simulation . computer simulation, pp.56- 67,2015.

10. Yu Xiaoji. Research on the construction of personalized teaching information service platform based on big data application. information science pp.66-71,2015.

11. pan fan. Big data concept is not no solution -- rambling data of three [EB/OL].[2016-01-06]. China information news network version http://www.zgxxb.com.cn/ppsd/201409020016.shtml any of the ten major enterprises in the practice of big data. Internet Weekly, pp.56-59,2014

12. Nature. Big Data [EB/OL].[2016-01-06].http://www.nature.com/.

13. Wikipedia [EB/OL]. [2016-01-06]. http://zh.wikipedia.org/wiki/ big data big data.

14. Wu Fatih, mu Zhijia. Ebook package based on the data of students' individual analysis model construction and realization path. The Chinese audio-visual education, pp,62-65,2014.

15. Gao Zhipeng, Niu Kun. Analysis of big data oriented technology. Journal of Beijing University of Posts and Telecommunications, pp,2-9,2015.