Transformations of Big Data in Past, Present and Future

[1]V.Nirupa,
[1]Asst Professor
[1]Computer Science & Engineering, Madanapalle Institute of Science and Technology, Madanapalle, A.P,India

Abstract **- Owing to the self-improvement desire, the human being always tries to reach to the current information and generate new ones from the data on hand. The practices are realized by processing and transforming the data, whose existence is broadly accepted into information. Generating information from data is vitally important in terms of regulating the life. Especially firms need to store and transform data quickly and properly into information in order to achieve the objectives such as having a competitive edge, producing new products, moving the firm ahead and stabilizing the internal dynamics. The increase in the amount of data sources also increases the amount of the data acquired. Therefore storing and processing data become difficult and classical approaches remain incapable to do such transactions. By means of Big Data large amount of data with a wide range can be stored, managed and processed. Besides Big Data ensures proper information quickly and offers advantage and convenience to the firms, researchers and consumers by taking the properties of Volume, Value, Variety, Veracity and Velocity into consideration. This study consists of 5 parts. In the Introduction part the features, classification, the process, the areas of usage and the techniques of Big Data are explained. In the second part the appearance process and the advantages of the concept of Big Data are illustrated with examples. A detailed literature review is produced in the third part. The actual studies and the most interested areas of Big Data are told in this part. In the fourth part the future of the Big Data is evaluated. Besides the situation and distribution of the studies on Big Data in Turkey and all over the world is presented. In the Conclusion part, an overall assessment is included and probable troubles are mentioned.**

Index Terms**—Volume, Variety, Velocity**

1.Introduction

Big Data has various definitions in the literature. Some of those are specified below:
Big Data is the amount of data beyond the ability of technology to store, manage and process efficiently(Manyika et.al, 2011).

Big Data is a term which defines the hi-tech, high speed, high-volume, complex and multivariate data to capture,store, distribute, manage and analyze the information (TechAmerica Foundation, 2014).

Big data is 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 (Gartner, 2014;Gürsakal, 2014).

Big Data Technologies are new generation technologies and architectures which were designed to extract value from multivariate high volume data sets efficiently by providing high speed capturing, discovering and analyzing(Gantz and Reinsel, 2011).

Hashem et.al. define Big Data by combining various definitions in literature as follows:
The cluster of methods and technologies in which new forms are integrated to unfold hidden values in diverse, complex and high volume data sets (Hashem et.al., 2015).

As the definitions suggest, there are some points to take into consideration in big data sets. The data should be complex and multiple together with its size. Therefore conventional methods have difficulty in analyzing big data sets and new methods and technologies are needed.

1.1Characteristics of Big Data

Various studies in the literature show that big data has 3, 4 or 5 characteristics; 3 of whom are common at all: Volume, Velocity and Variety. Others are Veracity and Value (Hashem et.al, 2015; Elragal, 2014; Fadiya et.al.,2014; Yang et.al., 2014; López et.al., 2015). 5 characteristics of big data is shown in Figure 1.

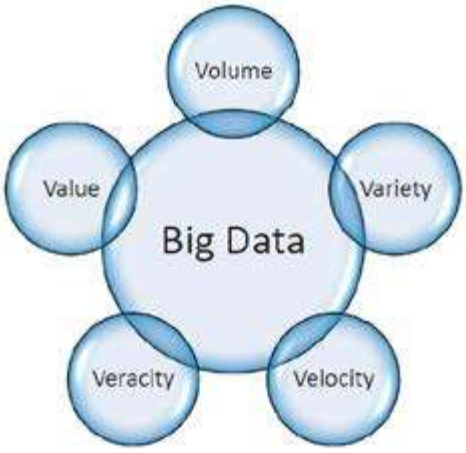

Fig 1. 1 Characteristics of Big Data (Elragal, 2014)

These 5 characteristics are explained as follows (Hashem et.al, 2015; Elragal, 2014; Fadiya et.al., 2014; Yang et.al., 2014; López et.al., 2015):

Volume: It is the most important characteristic of big data. It represents the size of the big data set.

Variety: Various data come to the companies from numerous resources (internal or external). These data entries from separate resources cause variance in data set. External data are hardly ever structural.

Velocity: The production rate of big data is notably high. The heavy increase in data means that the data should be analyzed more swiftly. The faster the data increases, the faster the need for the data increases; therefore the process shows increase as well.

Veracity: It is the accuracy of the data. The data should be acquired from correct resources and its security should be provided. Only authorized people should have the access permission.

Value: A result should be generated after all of the procedures and the result should enrich the process.

1.2. Classification of Big Data

The characteristic of big data can be understood better by dividing it into classes. These classes are Data Sources,Content Format, Data Stores, Data Staging and Data Processing (Hashem et.al, 2015).

Data Sources : Web & Social, Machine, Sensing, Transactions and IoT
Content Format : Structured, Semi-Structured and Unstructured
Data Stores : Document-oriented, Column-oriented, Graph based and Key-value
Data Staging : Cleaning, Normalization and Transform
Data Processing : Batch and Real time

1.3. Big Data Process

Big data process is shown below:
 Data Management
o Acquisition and Recording
o Extraction, Cleaning and Annotation
o Integration, Aggregation and Representation
 Analytics
o Modeling and Analysis
o Interpretation
Big data process is visualized by Gondomi and Haider as shown in Figure 2:

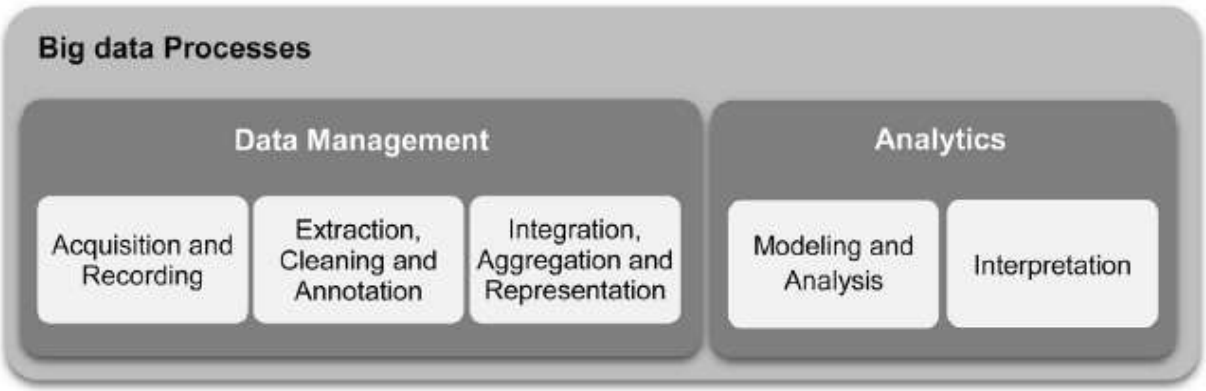

Fig2

Big data is used efficiently in numerous fields. Some of them are listed below:

- Automotive industry,
- High technology and industry,
- Oil and gas,
- Telecommunication sector,
- Medical field,
- Retail industry,
- Packaged consumer products,
- Media and show business,
- Travel and transport sector,
- Financial services,
- Social media and online services,
- Public services,
- Education and research,
- Health services,
- Law enforcement and defense industry.

1.5. Methods Used in Big Data

1.5.1. Text Analytics

Text analytics is used for information retrieval from data. E-mails, blogs, online forums, news and call center records are all examples of text data. Text analytics involve machine learning, statistical analysis and computational linguistics. Text analytics enable to extract meaningful summaries from large scale data (Gandomi and Haider,2015).

Information Extraction, Text Summarization, Question Answering and Sentiment Analysis are some of the techniques used in text analytics.

1.5.2. Audio Analytics

Audio analytics is used to extract information from unstructured audio data. Call centers and health services are commonly used utilization areas of audio analytics. Audio analytics can be used in numerous fields such as increasing the customer experience, the performance of customer representative and the sales rate; comprehending several tasks such as customer behaviors and the troubles of products (Gandomi and Haider, 2015).

1.5.3. Video Analytics

Video analytics is the usage of various techniques to extract meaningful information, track and analyze video streams. Marketing and operations management is the main application area of video analytics (Gandomi and Haider, 2015).

1.5.4. Social Media Analytics

Social media analytics is the analysis of the structured and unstructured data on the social media channels. Social media can be categorized as follows (Gandomi and Haider, 2015):

- Social networks (Facebook, LinkedIn),
- Blogs (BlogSpot, WordPress),
- Microblogs (Twitter, Tumblr),
- Social news (Digg, Reddit),
- Social bookmarks (Delicious, StumbleUpon),
- Media sharing (Instagram, YouTube),
- Wiki (Wikipedia, Wikihow),
- Question-and-answer sites (Yahoo! Answers, Ask.com),
- Review sites (Yelp, TripAdvisor).

1.5.5. Predictive Analytics

Predictive analytics is based upon estimating future considering current or stale data. Predictive analysis is used to capture the relationships of data and discover the patterns. Predictive analytics which is primarily based on statistical methods, is highly applicable on many disciplines (Gandomi and Haider, 2015).

2.Yesterday in big data

In the academic field big data can be accepted as a new concept. The change in the interest in big data by years is shown in the Figure 3.

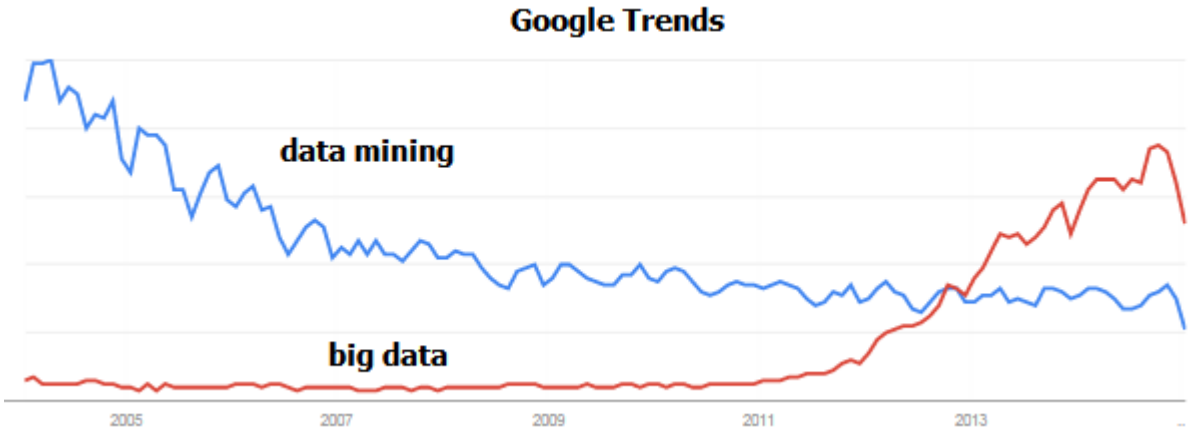

Fig 3

Changes in big data

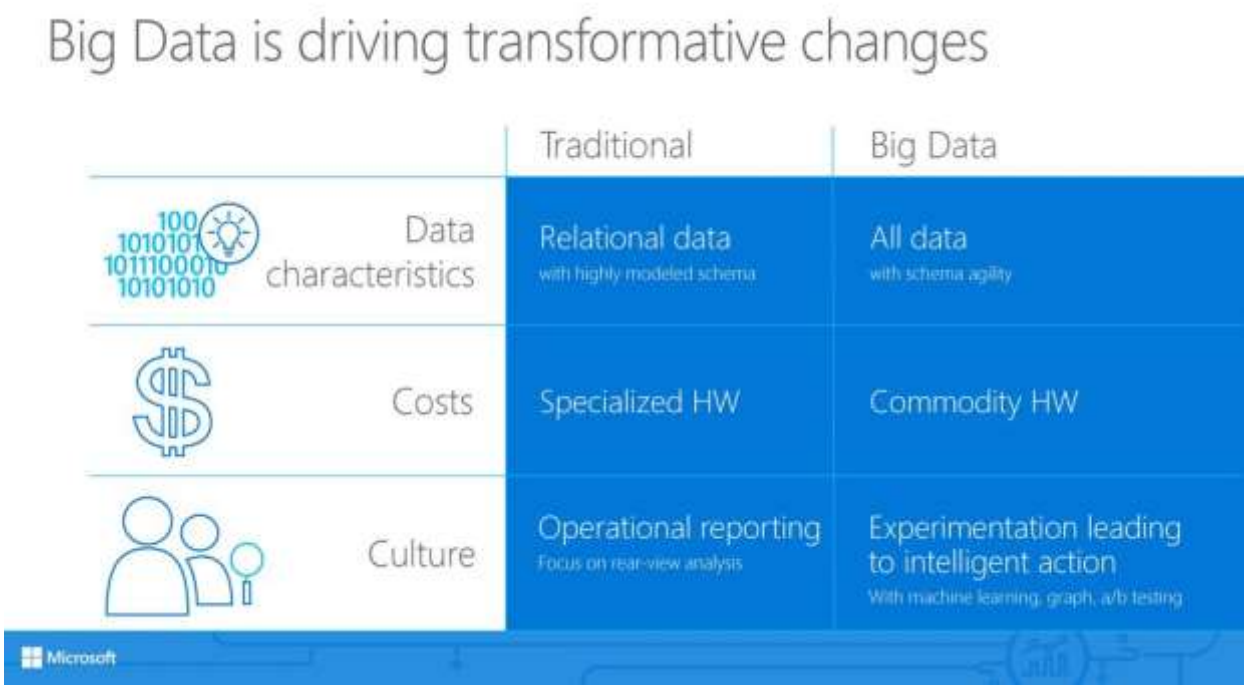

Fig: 4

3.Today in big data

Ackerman and Angus did a big data study to visualize spatial and temporal IP mobility. They visualized the temporal IP mobility in Los Angeles, New York, London, Moscow, Tokyo and Melbourne. Moreover, they plotted the IP map of Australia and showed the hourly variation of Melbourne's IP mobility (Ackerman and Angus, 2014).

Young focused on HIV and the importance of big data studied to be safe from HIV. Young claimed that a new HIV monitoring system could be set up by analyzing social media networking, thus new ideas could be generated for early intervention and disease control. According to Young, big data was applicable to various fields (Young, 2015).

Shin and Choi's study was on ecology. In this sense, the socio-ecological effects of big data transactions; such as social dynamics, political rhetoric technological choices were analyzed. Korea's big data studies were also
mentioned and some hints and tips were offered for Korean big data entrepreneurs (Shin and Choi, 2015).

4.Future in big data

Big data studies have been included in numerous fields. The importance of big data studies which have been used in a wide range of industries from automotive and communication to finance and health will increase in the future. As is seen from Figure 3, the interest in big data have increased day by day whereas the interest in data mining have diminished in importance by 2000s.

Data volumes will continue to grow. There's absolutely no question that we will continue generating larger and larger volumes of data, especially considering that the number of handheld devices and Internet-connected devices is expected to grow exponentially.

Ways to analyse data will improve. While SQL is still the standard, Spark is emerging as a complementary tool for analysis and will continue to grow, according to Ovum.

More tools for analysis (without the analyst) will emerge. Microsoft MSFT +0.63% and Salesforceboth recently announced features to let non-coders create apps to view business data.

Prescriptive analytics will be built in to business analytics software. IDC predicts that half of all business analytics software will include the intelligence where it's needed by 2020.

In addition, real-time streaming insights into data will be the hallmarks of data winnersgoing forward, according to Forrester. Users will want to be able to use data to make decisions in real time with programs like Kafka and Spark.

Machine learning is a top strategic trend for 2016, according to Gartner. And Ovum predicts that machine learning will be a necessary element for data preparation and predictive analysis in businesses moving forward.

Big data will face huge challenges around privacy, especially with the new privacy regulation by the European Union. Companies will be forced to address the 'elephant in the room' around their privacy controls and procedures. Gartner predicts that by 2018, 50% of business ethics violations will be related to data.

More companies will appoint a chief data officer. Forrester believes the CDO will see a rise in prominence — in the short term. But certain types of businesses and even generational differences will see less need for them in the future.

Predicting the future of big data (Source: Shutterstock)

"Autonomous agents and things" will continue to be a huge trend, according to Gartner, including robots, autonomous vehicles, virtual personal assistants, and smart advisers.

Big data staffing shortages will expand from analysts and scientists to include architects and experts in data management according to IDC.

But the big data talent crunch may ease as companies employ new tactics. The International Institute for Analytics predicts that companies will use recruiting and internal training to get their personnel problems solved.

The data-as-a-service business model is on the horizon. Forrester suggests that after IBMIBM -0.71%'s acquisition of The Weather Channel, more businesses will attempt to monetize their data.
Algorithm markets will also emerge. Forrester surmises that businesses will quickly learn that they can purchase algorithms rather than program them and add their own data. Existing services like Algorithmia, Data Xu, and Kaggle can be expected to grow and multiply.

Cognitive technology will be the new buzzword. For many businesses, the link between cognitive computing and analytics will become synonymous in much the same way that businesses now see similarities between analytics and big data.

"All companies are data businesses now," according to Forrester. More companies will attempt to drive value and revenue from their data.

Businesses using data will see $430 billion in productivity benefits over their competition not using data by 2020, according to InternationalInstitute for Analytics.

"Fast data" and "actionable data" will replace big data, according to some experts. The argument is that big isn't necessarily better when it comes to data, and that businesses don't use a fraction of the data they have access too. Instead, the idea suggests companies should focus on asking the right questions and making use of the data they have — big or otherwise

5. Conclusion and Recommendations

Several difficulties may show up in the acquisition, storage and processing of data. As the interest in big data increases, such difficulties will decrease or will be solved in shorter time. Turkey has fallen behind the world in terms of the academic interest and the academic studies in big data. There are plenty of academic studies in big data which can be taken by the researchers as a model.Data providers bear tremendous responsibility as much as the researchers in big data. The data providers which cannot process data will get harmed in terms of competition as they hide their data. Moreover, they prevent the data transform into knowledge. The data providers and the researchers should be in cooperation perspicuously based on the basis of trust. This will increase the reliability of the results and keep the firms advantageous in competition. Big firms like Facebook, Flickr, YouTube, Academia, The Marker, Google and Amazon have already gone towards big data because they all have broad visions. Those firms each hold its own market, expand their dominance at the same time ensure customer satisfaction. They keep their leadership and increase their market values day by day.As the studies in big data increase; technological developments and customer satisfaction will increase, diseases will be cured or precautions will be taken earlier and the general price level will decrease.

References

[1] Ackermann, K., & Angus, S. D. (2014). A Resource Efficient Big Data Analysis Method for the Social Sciences: The Case of Global IP Activity. Procedia Computer Science, 29, 2360-2369.

[2] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. MIS quarterly, 36(4), 1165-1188.

[3]Cook, S., Conrad, C., Fowlkes, A. L., & Mohebbi, M. H. (2011). Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PloS one, 6(8), e23610.

[4]Du, D., Li, A., & Zhang, L. (2014). Survey on the Applications of Big Data in Chinese Real Estate Enterprise. Procedia Computer Science, 30, 24-33.

[5]Dunham, M. H. (2006). Data mining: Introductory and advanced topics. Pearson Education India.

[6]Elragal, A. (2014). ERP and Big Data: The Inept Couple. Procedia Technology, 16, 242-249.

[7]Fadiya, S. O., Saydam, S., & Zira, V. V. (2014). Advancing big data for humanitarian needs. Procedia Engineering, 78, 88-95
.
[8] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37.

[9]Fire, M., Tenenboim, L., Lesser, O., Puzis, R., Rokach, L., & Elovici, Y. (2011, October). Link prediction in social networks using computationally efficient topological features. 2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing,73-80. IEEE.