



INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS

ISSN 2320-7345

TRENDS IN BIGDATA ANALYSIS A SURVEY

R. Muthumeenakshi¹, A. Vanithakatherine², N. Vallileka³

¹PSNA College of Engineering and Technology, muthumeenakshi83@gmail.com

²PSNA College of Engineering and Technology, avanitha@yahoo.com

³PSNA College of Engineering and Technology, vallilekapsna@gmail.com

Author Correspondence: PSNA CET, 9952369562muthumeenakshi83@gmail.coms

Abstract: - The well-known three Vs of Big Data - Volume, Variety, and Velocity – are increasingly placing pressure on organizations that need to manage this data as well as extract value from this data deluge for Predictive Analytics and Decision-Making. Big Data technologies, services, and tools such as Hadoop, MapReduce, Hive and NoSQL/NewSQL databases and Data Integration techniques, In-Memory approaches, and Cloud technologies have emerged to help meet the challenges posed by the flood of Web, Social Media, Internet of Things (IoT) and machine-to-machine (M2M) data flowing into organizations.

Keywords: Big data analysis, data mining, velocity, Hadoop.

1. Introduction

Many companies have decided that big data is not just a buzzword, but a new fact of business life -- one that requires having strategies in place for managing large volumes of both structured and unstructured data. And with the reality of big data comes the challenge of analyzing it in a way that brings real business value. Business and IT leaders who started by addressing big data management issues are now looking to use big data analytics to identify trends, detect patterns and glean other valuable findings from the sea of information available to them.

It can be tempting to just go out and buy big data analytics software, thinking it will be the answer to your company's business needs. But big data analytics technologies on their own aren't sufficient to handle the task. Well-planned analytical processes and people with the talent and skills needed to leverage the technologies are essential to carry out an effective big data analytics initiative. Buying additional tools beyond an organization's existing business intelligence and analytics applications may not even be necessary depending on a project's particular business goals.

This Essential Guide consists of articles and videos that offer tips and practical advice on implementing successful big data analytics projects. Use the information resources collected here to learn about big data analytics best practices from experienced users and industry analysts -- from identifying business goals to selecting the best big data analytics tools for your organization's needs.

2. Business benefits

Technology selection is just part of the process when implementing big data projects. Experienced users say it's crucial to evaluate the potential business value that big data software can offer and to keep long-term objectives

in mind as you move forward. The articles in this section highlight practical advice on using big data analytics tools, with insights from professionals in retail, healthcare, financial services and other industries.

3. Trends of bigdata analysis

3.1 Smart Machines

Many private companies are using some form of smart machine technology with a limited number of employees to generate revenue. Gartner recently interviewed representatives of three of the fastest-growing private companies in the world. Each of them was using or piloting some form of smart machine technology, and each had fewer than 150 employees generating \$85 to \$160 million in annual revenue. Facebook's October 2014 purchase of the messaging platform WhatsApp (\$22 billion dollars for a company with only 55 employees) is a particularly dramatic example of this trend, and of a new pattern in valuations. Smart machines can help both enterprises and individuals in two important ways:

- By fully automating tasks and removing human control (for example, via algorithmic trading or “lights out” factories with no on-site employees).
- By augmenting the cognitive and physical performance of individuals in a manner that feels like an extension of their own abilities (for example, via decision support or wearable technology).

For instance, the non-profit online education platform edX has introduced a computer system that grades student essays and short answers on exams.

3.2. Customer Digital Assistants

While not quite in the league of true artificial intelligence, top technologies for customer digital assistants include face recognition, voice identification, emotion detection, natural language processing, and audience profile data. After decades of constant improvements in voice recognition and detection technology and face recognition technology from large, high-profile organizations like Adobe, Apple, Google, IBM and Microsoft and others — the technology is now here, and ready for prime time. Consider Microsoft's Xbox One and Sony's PS4, two leading gaming consoles that both have the ability (through Kinect and the PlayStation Camera, respectively) to recognize and sign in players by face, when they enter a room. Both systems respond to voice commands, and all of these capabilities are used in-game. Some of the promising mobile digital assistant technologies are, Cortana, Google Now, Siri, and Amazon Echo. Even though digital assistant technologies are used across various industries. Customers are willing to adopt these technologies to help them sift through increasing large amounts of information, choice and purchasing decisions.

3.3. Internet of things

Internet of things (IOT) is the concept of connecting devices with an on and off switch to the Internet or to each other through the internet. Industries. The above numbers clearly show that the market for IOT is growing at an incredible rate. GE's Industrial Internet Insights Report predicts that the “Internet of Things” will add \$10 to \$15 trillion to global GDP over the next 20 years. It is no surprise that employer competition for skills in this space is on the rise. Both technology and non-technology companies are hiring for IOT. Fastest growing IOT job positions include systems software developers (215% growth in the past year), information security analysts (113% growth), and computer systems engineers (110% growth). Adding to this, General Electric has advertised 2,104 jobs looking for skills needed to support their industrial internet initiative in the last two years.

3.4. Automated composition engines

Significant advances in machine learning, data integration, and predictive analytics applications have empowered computers to write business content types like reports, white papers, press releases articles, etc. For example, automated insights have already been employed to robot-write thousands of stories for fantasy football newsletters based on relatively rigid templates.

Content management vendors are adding better content composition and analytic capabilities to their systems. Solutions that combine the ability to find analyze and assemble data into formats that can be read in natural language, will improve both the speed and the quality of business content delivery. Personal assistants such as Apple's Siri and Microsoft's Cortana, and IBM Watson, with its cognitive technology — provide richer and

more interactive content. Drones, sensors, wearables, and cameras will create huge volumes of data that needs to be accessed and assimilated into understandable content.

3.5. Robo-Boss

Robo-bosses will increasingly make decisions that previously could only have been made by human managers. Companies will develop smart machine “contract managers” to measure and supervise the performance of particular worker classes. Evaluation of employee performance will become even more granular as smart machines become the primary means of analyzing performance. Smart machines can effectively track activities and events that are too difficult for human managers to measure. Robo-bosses use advanced machine learning techniques like deep learning, to automatically learn and improve with experience. A good example of deep learning is Google’s self-driving car project.

4. Conclusion

These five trends mark the beginning of a new era for big data analytics. Some, like customer digital assistants and smart machines, threaten the groundwork of big data analytics in general; while others, like Robo-boss and automated composition engines, can be adapted to and taken advantage.

REFERENCES

- [1] Agneeswaran VS, Tonpay P, Tiwary J (2013) Paradigms for realizing machine learning algorithms. *Big Data* 1(4):207–214
- [2] Zaharia M, Chowdhury M, Franklin MJ, Shenker S, Spark SI (2010) Cluster Computing with Working Sets. In: *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing.*, pp 10–10
- [3] Milojevic DS, Kalogeraki V, Lukose R, Nagaraja K, Pruyne J, Richard B, Rollins S, Xu Z (2002) Peer-to-peer computing. Technical Report HPL-2002-57, HP Labs
- [4] Steinmetz R, Wehrle K (2005) *Peer-to-Peer Systems and Applications*. Springer Berlin, Heidelberg
- [5] Sievert O, Casanova H (2004) A simple MPI process swapping architecture for iterative applications. *Int J High Perform Comput Appl* 18(3):341–352
- [6] Hadoop. <http://hadoop.apache.org/>
- [7] Borthakur D (2008) HDFS architecture guide. HADOOP APACHE PROJECT. http://hadoop.apache.org/docs/r1.2.1/hdfs_design.pdf
- [8] Vavilapalli VK, Murthy AC, Douglas C, Agarwal S, Konar M, Evans R, Graves T, Lowe J, Shah H, Seth S (2013) Apache hadoop yarn: Yet another resource negotiator. In: *Proceedings of the 4th annual Symposium on Cloud Computing.*, p 5
- [9] Dean J, Ghemawat S (2008) MapReduce: simplified data processing on large clusters. *Commun ACM* 51(1):107–113
- [10] Lee K-H, Lee Y-J, Choi H, Chung YD, Moon B (2012) Parallel data processing with MapReduce: a survey. *ACM SIGMOD Record* 40(4):11–20
- [11] Olston C, Reed B, Srivastava U, Kumar R, Tomkins A (2008) Pig latin: a not-so-foreign language for data processing. In: *Proceedings of the ACM SIGMOD international conference on Management of Data*. ACM, pp 1099–1110
- [12] Thusoo A, Sarma JS, Jain N, Shao Z, Chakka P, Anthony S, Liu H, Wyckoff P, Murthy R (2009) Hive: a warehousing solution over a map-reduce framework. *Proceedings of the VLDB Endowment* 2(2):1626–1629
- [13] Yu Y, Isard M, Fetterly D, Budiu M, Erlingsson Ú, Gunda PK, Currey J (2008) DryadLINQ: A System for General-Purpose Distributed Data-Parallel Computing Using a High-Level Language. In: *OSDI.*, pp 1–14
- [14] Owen S, Anil R, Dunning T, Friedman E (2011) *Mahout in Action*. Manning
- [15] Bu Y, Howe B, Balazinska M, Ernst MD (2010) HaLoop: efficient iterative data processing on large clusters. *Proceedings of the VLDB Endowment* 3(1–2):285–296