International Journal of Engineering & Technology, 7 (3.7) (2018) 363-366

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Challenges, Skills and Roles for Information Professionals in Managing Big Data

[1]Nur Adlina Khamis*, [2]Saiful Farik Mat Yatin, [3]Farah Nabilah Maidin, [4]Rasmira Mat Rashid, [5]Saidatul Akmar Ismail, [6]Norshila Saifuddin, [7]Yamin Kamis

[1234567]Faculty of Information Management, Universiti Teknologi MARA (UiTM) Selangor, Malaysia

Abstract

Big data is a now well used term designed to define the growing volume, variety, velocity, variability and value of information surging through organizations. The benefits that can be utilize from big data will only come about if the information is resourceful and been managed properly. As the amount and variety of data grows, so do the skills required to capture, manage, and analyze the data. The investigation for associations is making sense of how to get from existing data and application frameworks to flawlessly ingesting any key information sorts from all sources and quickly displaying bits of knowledge to chiefs. This article discusses about big data challenges faced by information professionals, the skills needs and roles for them in managing it.

Keywords: *Big data, Skills, Organization, Information, Information Professionals*

1. Introduction

Information is pouring in from numerous non-traditional sources, and associations are just start to see how to extricate an incentive from it. Take the Internet of Things (IoT). At more than 13 billion gadgets, a number that is climbing each day, the IoT is producing monstrous measures of information with the possibility to change business once it has been gathered and examined utilizing enormous information examination instruments, for example, Splunk and Hadoop. These innovations are intended to control monstrous informational collections with the goal that associations can increase important knowledge rapidly. Keeping in mind the end goal to achieve that point, associations must conquer overwhelming difficulties like stockpiling limit, execution, information uprightness and security. There's likewise the issue of adequately overseeing huge information over its lifecycle. Associations that have executed huge information examination realized rapidly that capacity dormancy is a genuine issue.

At the point when the information to be controlled is put away in different si-los, it is hard to understand the advantages of constant analyt-ics. Basically moving around and sharing huge pieces of information is a demoralizing errand. To conquer this bottleneck, crude information is dumped into a focal store called an information lake, where it is held in its unique arrangement until gotten to by investigation instruments. A data lake stores data using a flat architecture rather than hierarchical files or folders. A unique identifier is applied to each data element within the lake, along with extended metadata tags, allowing for the query and analysis of smaller data sets. This simplifies the integration of data from multiple sources and reduces data movement and the resulting latency.

Information lakes have developed as an intense engineering approach for dealing with the developing assortment and volume of information. Be that as it may, the capacity to deal with the information in view of age and pertinence has the effect between a valuable information lake and an exorbitant and wasteful one. What's more, the information converge in data isn't gathered in light of some biased arrangement of inquiries. It's difficult to realize what segments of the information may demonstrate profitable a suggestion that turns out to be more troublesome as the information ages. The big data environment has introduced significant changes to the information society. In this environment, the volume of user needs information is expanding rapidly, the forms of user needs are becoming diversified, the distribution of user needs is becoming fragmented and sparse and there is a dramatically sharpened focus on personal information protection [1]. In addition, these changes bring many challenges and impose many requirements on the big data in many organizations to fulfill user needs in the era of big data environment [2].

2. Big data challenges

World nowadays being explode with variety of data and existing of information as information's is a key success to organizations and people through a various type of technology whether hardware or software. Big data is a current trending research among researchers [3].

In this research big data explains as the data's that being produced in every minutes of the time and capture anywhere in type of structured or unstructured of data has merged the data to big data in organization or peoples. Comuzzi & Patel [4] defines big data as a relatively new term coined to label the exponential growth and availability of data, both structure and unstructured.

The expression "big data" is generally new, the demonstration of assemble ing and putting away a lot of data for inevitable investigation is ages old. The idea picked up energy in the mid 2000s when industry investigator Doug Laney explained the now-standard meaning of huge information as the four (4) V's. Another factor that is newly added is variability by Hilbert [5] to make it five (5) V's.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Volume: The quantity of generated and stored data. The size of the data determines the value and potential insight- and whether it can actually be considered big data or not. Organizations collect data from a variety of sources, including business transactions, social media and information from sensor or machine-to-machine data. In the past, storing it would've been a problem but new technologies (such as Hadoop) have eased the burden.
Velocity: The speed at which the information is produced and handled to address the requests and difficulties that lie in the way of development and improvement. Data streams in at an unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time.
Variety: The type and nature of the data. This helps people who analyze it to effectively use the resulting insight. Data comes in all types of formats from structured, numeric data in traditional databases to unstructured text documents, email, video, audio, stock ticker data and financial transactions.
Veracity: The quality of captured data can vary greatly, affecting accurate analysis. It refers to the degree of trustworthiness of the used information in order to take decisions. The vulnerability about the consistency and fulfillment of information and different ambiguities develops as the assortment and number of sources increments and keeps on expanding additional time. Absence of veracity may prompt off base correla-tions in huge information which further may prompt off base deci-sions for organizations.
Variability: Inconsistency of the data set can hamper processes to handle and manage it.

According to Verma and Bhattacharyya [6], the researchers said that big data is a term to describe large and complex data sets which require a rapid growth of processing and the data are impossible to manage using standard database management and analytical tools. Based on their definition regarding big data, it shows that big data are the largest and fastest development of data that currently emerge from various way of development such as through social networking sites, through formal or informal data gathering, data sharing session and many more ways.
Comuzzi and Patel [4] said that big data is a comparatively new term to label the rapid growth and availability of data, can be capture and acknowledge during anytime and anywhere using various kind of method to capture the data whether structured or unstructured. People or organizations can capture the data using many ways within formal or informal ways to create the existing and upcoming big data. The researchers also mention that the availability of the massive amounts of data provides unprecedented opportunities for organizations. The existence of huge data can lead to the information overload and information explosion, and when this happen the relevancy of the data and information being judgement by the users or other researchers.

In addition, Feicheng, Chen and Zhao [7] said that the big data environment has brought substantial modifications to the information society. In these surroundings, the extent of consumer needs data is increasing swiftly, the forms of user wishes are becoming assorted, the distribution of person desires is becoming fragmented and sparse and there is a dramatically sharpened consciousness on private information safety. These modifications convey many challenges and impose many requirements on the organisation of person wishes records inside the large facts surroundings. There are techniques to analyse the BDA whether structured and unstructured data in real time have been categorized in five sub-categories by Gandomi and Haider [8]:

Text analytics: Refer to techniques that exact from information from textual data like emails, online forums, blogs, social network feeds, logs and such others [6].
Audio analytics (Speech analytics): It analyses and extracts information from unstructured audio data for example data from call center, healthcare and such others [9]. This techniques improves customers in providing better experience, evaluate and such others.
Video analytics: Refers to techniques to monitor, analyze and extract meaningful information from video stream [6].
Social media analytics: Refers to the analysis of unstructured and structured data from social media channels. The analytics techniques used to analyse and acquire intelligence from big data.

In this article, authors define big data as the explosion of data in various medium of data capture in formal or informal way. Based on previous research, the big data brought the meaning into the term that explains by the researches. The big data form by the huge data gather in the system or by any organizations. The data consists of the structured or unstructured data to perform and acknowledge people around the views.

Skills needed to become versatile IPs

Today, IP assuming a basic part in dealing with the perpetual stream of data created in a divided media scene. The test lies in curating and breaking down this mass of information to give important bits of knowledge that can be shared over the association [10]. The significance of big data doesn't spin around how much information associations have, yet what they do with it. Associations can take information from any source and dissect it to discover answers that empower:
>cost decreases;

>time decreases;
>new item advancement and enhanced contributions;
>smart basic leadership.

At the point when associations join enormous information with powerful examination, they can achieve business-related undertakings, for example,

- Determining underlying drivers of disappointments, issues and de-fects in close constant.
- Generating coupons at the purpose of offer in light of the client's purchasing propensities.
- Recalculating whole hazard portfolios in minutes.
- Detecting deceitful conduct before it influences your association.

Sum and assortment of information inside associations develops, so do the measures of abilities required to have the capacity to catch, oversee, and break down the information. Association need to get this privilege in the event that they need to advance and benefit from their utilization of enormous information, they require the correct mix of abilities and learning, require the suitable specialized capacities and furthermore require the profound business mastery to ensure that information is grilled and utilized as a part of ways that will convey new answer and potential outcomes for the business. As IPs in an organization, there are several skills need to be adopt in managing big data. The data is huge and come in variety form of data. With the upcoming of multiple technologies, the information professional need to get ready to cope with those new things.
The information professional need to do the assessments in the real-world organizations demonstrates to others subordinates. Comuzzi & Patel [4] explain that the evidence emerging from the assessments suggest that the creation of appropriate formal governance structures and sponsorships is an aspect that is often overlooked. This may lead Big Data initiatives to be technologically focused and, as a consequence, restricted to the concerns of the IT function and individual initiative within other functions. While Big Data technology is widely available in the market, organizations should also strive to secure resources to respond effectively to technology evolution.

Among the skills needed to become versatile IP is being able in managing the explosion in data sources. The sheer volume and recurrence of new news sources and substance implies that data

over-burdens are an undeniable issue. Presently notwithstanding conventional media sorts, for example, daily papers, business distributions and organization information, online sources and web-based social networking must be totaled and overseen. As a result of this blast of information sources with more data accessible on the web, there is a more prominent need to guarantee the nature of that data. IPs additionally needs to begin pondering how to enhance inquire about help administrations and need to gain from their associates. These abilities incorporate incredible tuning in and issue elucidating aptitudes, profound learning of what constitutes quality data, immense involvement in choosing data that is fitting for the organizations, seeing how to sort out data with the goal that it can be recovered and comprehension of how to perceive gatherings to complete function.

IPs also needs to have Leadership Skill. Leadership skills approach takes into account the knowledge and abilities that the leader has. A leader can learn certain skills and turn himself into a remarkable one such as human skill, conceptual skill and technical skill [11].

Human skill is the capacity to work adequately as gathering individuals and to fabricate agreeable exertion inside the group he or she leads [10]. Each administrative level requires cooperation with other individuals. A very created human aptitude and if IPs know about their states of mind, suppositions, and convictions about different people and gatherings, IPs can see their value and impediments and liable to acknowledge others' perspective, observations and convictions, which may be not quite the same as them. Human abilities can be created with no formalized preparing for a few. Numerous others are to be exclusively helped by their quick bosses who themselves ought to have the human aptitude so as to have the capacity to give that. With this adjustment in state of mind, there may likewise build up some dynamic expertise in managing human issues. IPs as a better may like than watch their subordinate's capacity to work viably with others.

Conceptual skills on the other hand are abilities to work with ideas and concepts [10]. These skills enable IPs to understand and better decide the actions and measures that has to be taken in a particular field of work.

Technical skill is the most important skills for IPs nowadays in managing Big Data [10]. Technical skill is primarily concerned with working with things. Technical skill involves process or technique knowledge and proficiency. A technical skill is knowledge about and competency and proficiency in a specific work or activity. For example, to use certain computer software packages (for example, Hadoop, NoSQL, Evangelist etc) is an advanced technical skill. The IPs need to be intelligent in order to cope with this new hardware or software upcoming to the organizations. They need to handles those hardware and software with manner and idea because they need to give an instruction to their subordinate.

Big data requires new abilities, for example, information warehousing experts who see how to catch the greater part of the information, comprehend the ideas driving the examination important to extrapolate significant experiences, and can index the information so others can utilize it all the more proficiently [10]. This is because in order to handle the explosion of data sources in the organizations. The IPs must have the skills to manage the big data issues in the future or in the meantime.

The last range of abilities which may encourage convey that edge to fulfilment will be the capacity to unite the conventional informational collections held in social databases with the new record based datasets and run that volume of data over another appropriated preparing foundation to pick up speed and assortment. The test for associations is making sense of how to get from existing information and application frameworks to flawlessly ingesting all key information sorts and aptitudes from all sources and quickly introducing experiences to leaders.

However, as a human towards information professionals it is irrelevant to cope with those listed skills above but as information professionals the critical thinking to handle those skills can be manage in order to achieve excellent and organizations goals. They need to learn how to do programming skills [10]. IPs need to figure out how to code the fundamental ability in the Big Data Analyst's. They have to code the lead of the numerical and measurable examination with gigantic informational collections. The system need to provide large amount of storage for the big data programming analysis.

The IPs also needs to have quantitative skill [10]. As a big data analyst, programming help to what the researcher wants. The quantitative skill need to be excellent because to answer those questions according to the related topic of big data such as statistics or numerical analysis. In many associations, an expansive dominant part of workers don't comprehend their own organization's information. Truth be told, most workers don't have a reasonable thought of where every one of the information is. These representatives frequently depend on preconfigured reports and dashboards to determine their experiences. Shockingly, this approach is unsafe. It doesn't give a comprehensive perspective of the information acquisition and examination process. As associations develop inorganically, extraordinary information storehouses blend, bringing about a befuddling mess.

Be that as it may, by asking the correct inquiries, IPs can set out on an appropriate investigation of the crude information. The correct inquiries and disclosures can change the course of business for an association. Associations requires better calculations to help with separating an incentive from it, which implies quants, those scientific wizards who make calculations to comprehend unstructured data, will keep on being perpetually profitable. At last, enormous information makes math experts cool, and maybe is the greatest achievement of all.

3. Role of IPs

The role of IPs as knowledge managers was also explored in the context of big data management in an organization. The benefits for IPs are substantial in terms of advancing their role [3]. The information professionals focused on exploring the core characteristics that should obtain to be successful in their work. There are four general classes of chances for data experts, to be specific:

Facilitating a situation helpful for information sharing, dealing with the corporate memory;

Transferring data administration to another setting connected to business procedures and center tasks and creating corporate data education;

Has to be advance and futuristic mind that will most likely be updated on big data management such hybrid of computing, media specialization, and instructional technologies.

Need to be analytical. With access to advanced technology, big data has to be developed skills that enable them to organize and analyze data. It would help to communicate those data in a manner that can drive impactful data driven decisions. IPs need to be able in simplifying a problem and break it into smaller analysis segments and provide recommendations accordingly.

4. Conclusion

The paper discusses the challenges of big data, skills needed to become versa tile's IP and the roles of IP's. Today IPs assumes a basic part in dealing with the interminable stream of data produced in a divided media scene. The test lies in curating and dissecting this mass of information to give significant bits of knowledge that can be shared over the association. The IPs and each of organizations need to understand the meaning of big data. The big data definitions are different according to each organizations or experts. The works to handling big data in organizations also may be different. There must have a standard in order to standardize the implementation of big data in the organizations. The organization

must adapt and adopt the skills needed to become versatile IPs. In order to be versatile, they need to be creative in order to solve issues regarding big data. The creativity must be outside the box of other people ideas. The creativity will cope to the big data issues and make them be versatile in their own way. The big data are hard and complicated to handles. The organization must have strategic plan to handle the process of capture the data until the interpretation of the data. The data will be valuable and easy to retrieve and can transform to valuable information. They must learn to cope with multiple technologies, programming, hardware and software to handle capture and storage process and able to interpret data. They need to cope with those entire requirements in order to become an excellent data analyst.

References

[1] Hajar Mousannif. (2016). Big data projects: just jump right in. 12 (2), 260-288, Retrieved from http://dx.doi.org/10.1108/IJPCC-04-2016-023

[2] The Next Big Data Phase: Business knowledge expansion. (2013). Database trends an application. Retrieved from http://www.dbta.com/Editorial/Think-About-It/The-Next-Big-Data-Phase-Business-Knowledge-Expansion-88652.aspx?PageNum=2

[3] Big Data & Investment Management: The potential to quantify traditionally qualitative factors. (2013). *Citi Business Advisory Services*. Retrieved from https://www.cmegroup.com/education/files/big-data-investment-management-the- potential-to-quantify-traditionally-qualitative-factors.pdf

[4] Comuzzi, M. & Patel, A. (2016). How organisations leverage big data: a maturity model. *Industrial Management & Data Systems*, *116*(8), 1468-1492. Retrievedfrom https://doi.org/10.1108/IMDS-12-2015-0495

[5] Hilbert, M (2015). Big Data for Development: A Review of Promises and Challenges. Development Policy Review. Retrieved martinhilbert.net.

[6] Verma, S & Bhattacharyya, Sekhar, S (2017). Perceived strategic value based adoption of big data analytics in emerging economy: a qualitative approach for Indian firms. Journal of Enterprise Information Management, 30(3), 1-37. doi: 10.1108/JEIM-10-2015-0099

[7] Feicheng Ma, Ye Chen, Yiming Zhao. (2017). Research on the organization of user needs information in the big data environment. *The Electronic Library*, *35*(1), 36-49. Retrieved from https://doi.org/10.1108/EL-07-2015-0130

[8] Gandomi, A. and Haider, M. (2015). Beyond the hype: big data concepts, methods, and Analytics. *International Journal of Information Management*, *35*(2), 137-144.

[9] Hirschberg, J., Hjalmarsson, A., & Elhadad, N. (2010). "You're as sick as you sound": Using computational approaches for modeling speaker state to gauge illness and recovery. In A. Neustein (ed.), Advances in speech recognition (pp. 305–322). United States: Springer.

[10] Madden, M. J. (2008). Today's UK professional association library and information service: challenges and best practice. *60*(6), 556-569. Retrieved from http://dx.doi.org/10.1108/00012530810924267

[11] Katz, R (2017) Skills approach: leadership model. Retrieved on 2017-06-11 at http://www.tlu.ee/~sirvir/Leadership/Leadership%20Models/skills_approach_robert_katz.html