

A Survey on the Challenges and Implementation of Data Management

Mac-Donald .C. Ilondu,
MSc. (I.T) Final Year Student, Jain University,
Bangalore- 69, India.

Dr. Suchitra Suriya,
H.O.D, MSc. (I.T) Department, Jain University,
Bangalore-69, India.

Abstract:- This paper aims to discuss about the overview of Data Management, its challenges as well as some implementations on how to manage big data. Data management is more-or-less a process or a method whereby data can be processed, stored and managed. It is widely consisted of data governance, data warehousing, and data integration. It involves the collection and storage of data, new big data or both. Data has been one of the most relevant and sensitive part of our technological lives. For us to manage data efficiently and sufficiently, it would require some efforts and resources to achieve that. This paper discusses about how it affects and restricts data management from producing optimal solutions.

Keywords: Data management, data warehousing, data integration, big data.

1. INTRODUCTION

First and foremost, it is essential that knowledge of Data Management is known. This paper surveys that Data Management is the process of calculating or controlling the information generated during a research project. It further surveys that some studies will require some level of data management, and companies with huge sponsorships are increasingly requiring scholars to plan and execute good data management methods[4]. It is actions that contribute to effective storage, preservation and reuse of data and documentation throughout the research lifecycle[3].

Management of data generally focuses on the defining of the data element, how it is structured, stored and moved. Management of information is more concerned with the security, accuracy, completeness and timeliness of multiple pieces of data. These are all concerns that accountants are trained to assess and help manage for an organization[7].

Managing data is an integral part of the research process. Managing data helps you as a researcher or analyst to organize research files and data for easier access and analysis. It helps ensure the quality of your research[4]. It supports the published results of your work and, in the long term, helps ensure accountability in data analysis. Effective data management practices include – designating the responsibilities of every individual involved in the study, determining how data will be stored and backed up, implementing the data management plan and deciding how data will be dealt with through each modification of the study.

Big data has created opportunities like never before. Professionals who can analyze the huge amount of data present & create useful information are highly sought after

by companies across the world. It is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from big data, you need optimal processing power, analytics capabilities and skills [1].

Big data is changing the way people within organizations work together. It is creating a culture in which business and IT leaders must join forces to realize value from all data. Insights from big data can enable all employees to make better decisions—deepening customer engagement, optimizing operations, preventing threats and fraud, and capitalizing on new sources of revenue. But escalating demand for insights requires a fundamentally new approach to architecture, tools and practices.

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves[2].

2. LITERATURE REVIEW ON DATA MANAGEMENT

The definition provided in the Data Management Association (DAMA) Data Management Body of Knowledge (DAMA-DMBOK) is: "Data Management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets."[6]

This is where data management disciplines, tools and platforms (both old and new) are applied to the management of big data. Traditional data and new big data can be quite different in terms of content, structure and intended use, and each category has many variations within it. To accommodate this unique diversity, software solutions for Big Data Management (BDM) tend to include multiple types of data management tools and platforms, as well as diverse user skills and practices[5].

Big data management is about two things – big data and data management – plus how the two work together to achieve business and technology goals[5]. To get us all on the same page, let us start with definitions of both, and then bring them together.

Data management - Definition

This involves the collection and storage of data, plus its processing and delivery – whether traditional data, new big data, or both[5]. Processing can be extensive, especially when data is repurposed for a use differing from that of its origin (as is common in business intelligence, data warehousing and analytics).

Big data- Definition

Big data is first and foremost about data volume, namely large data sets measured in tens of terabytes, or sometimes in hundreds of terabytes or petabytes. Before the term *big data* became common parlance, we talked about very large databases (VLDBs); these, usually contain exclusively structured data, managed in a database management system (DBMS). In many organizations, big data and its management follow the VLDB paradigm[5]. In these large data sets, big data can also be an eclectic mix of structured data (relational data), unstructured data (human language text), semi-structured data (XML, RFID) and streaming data (from sensors, social media, Web applications and machines).

3. WHY BIG DATA AND DATA MANAGEMENT ARE COLLIDING

Today, we no longer use the term *eBusiness* (which was trendy during the 1990s) because we assume that an organization of any size or complexity should have numerous applications for the sake of efficiency and competitiveness. A consequence of the post-eBusiness era is that many organizations now have massive volumes of application data to manage and to leverage for business value. Although organizations have the skills for structured data (which is what comes out of most operational applications), today's unprecedented data volume and speed of generation make big data management a challenge.

“ Structured data from applications is a common form of big data, although it is not new”

Big data comes from many sources, in many formats. Some industries have large, valuable stores of unstructured data, typically in the form of human language text. For example, the claims process in insurance generates many textual descriptions of accidents and other losses, plus the related people, locations, and events. Most insurance companies process this unstructured big data using technologies for Natural Language Processing (NLP), often in the form of text analytics. The output from NLP may feed into older applications for risk and fraud analytics or actuarial calculations, which benefit from the larger data sample provided via NLP.

“ Big data can be industry specific, such as unstructured text in insurance, healthcare and government”

Sensors are coming online in great numbers as a significant source for big data. For example, robots have been in use for years in manufacturing, but now they have additional sensors so they can perform quality assurance as

well as assembly. For decades, mechanical gauges have been common in many industries (such as chemicals and utilities), but now the gauges are replaced by digital sensors to provide real-time monitoring and analysis. GPS and RFID signals now emanate from mobile devices and assets ranging from smart phones to trucks to shipping pallets—so all these can be tracked and controlled precisely.

“ Sensor data and other machine data are new and large, and they enable new applications”

4. TECHNOLOGY DRIVERS BEHIND BIG DATA MANAGEMENT

Big data just gets bigger. It's important to beef up data management infrastructure and skills as early as possible. Otherwise, an organization can get so far behind from a technology viewpoint that it's difficult to catch up. From a business viewpoint, delaying the leverage of big data delays the business value. Similarly, capacity planning is more important than ever, and should be adjusted to accommodate the logarithmic increases typical of big data.[5]

Leverage big data, don't just manage it. It costs money to collect and store big data, so don't let it be a cost center. Look for ways to get business value from big data. As you select data platforms for managing big data, consider low-cost new ones and open source.[5]

Joining big data with traditional data is another path to value. For example, so-called 360-degree views of customers and other business entities are more complete and bigger when based on both traditional enterprise data and big data. In fact, some sources of big data come from new customer touch-points (mobile apps, social media) and so belong in your customer view.[5]

Big data can enable new applications. For example, in recent years, a number of trucking companies and railroads have added multiple sensors to each of their fleet vehicles and train cars[5]. The big data that streams from sensors enables companies to more efficiently manage mobile assets, deliver products to customers more predictably, identify noncompliant operations, and spot vehicles that need maintenance.

Big data can extend older applications. This includes any application that relies on a 360-degree view, as mentioned above. Big data can also beef up the data samples parsed by many analytic applications, especially those for fraud, risk, and customer segmentation.[5]

5. PROBLEMS AND OPPORTUNITIES FOR BIG DATA MANAGEMENT

In recent years, TDWI has seen many organizations adopt new vendor platforms and user best practices that enabled them to overcome some of the performance issues with big data that dogged them for years, especially data volume scalability and real-time data processing. With that

progress in mind, this report’s survey asked: “Is the management of big data mostly a problem or mostly an opportunity?” [5] (Figure 1.)

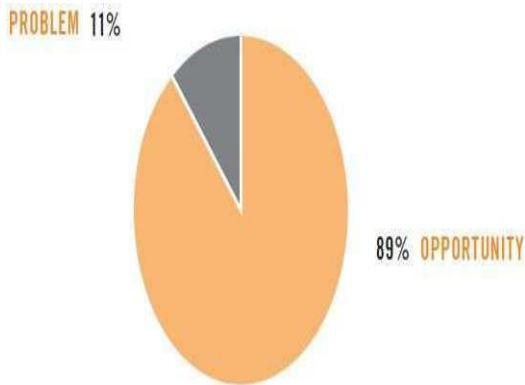


Figure 1: Based on 461 respondents [5]

A tiny minority consider BDM a problem (11%). No doubt, big data presents technical challenges due to its size, speed, and diversity. Data volume alone is a showstopper for a few organizations.

The vast majority consider BDM an opportunity (89%). Conventional wisdom today says that big data enables data exploration and predictive analytics to discover new facts about customers, markets, partners, costs, and operations.[5]

6. CHALLENGES OF BIG DATAMANAGEMENT

Big data management has benefits, as we just saw. Yet, it also has barriers. To get a sense of which problems are more likely than others, this report’s survey asked respondents: “What problems hinder the successful management of big data in your organization?” [5](Figure 2.)

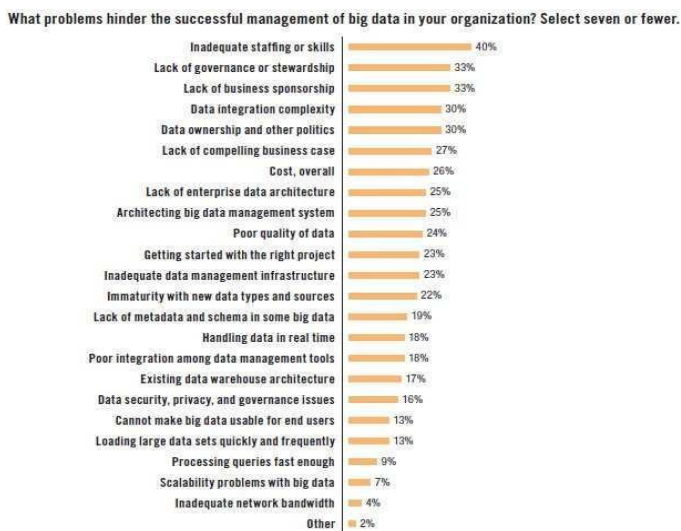


Figure 2: Based on 2,287 responses from 461 respondents; 5 responses per respondent, on average [5]

Being new to big data and its management is the biggest challenge users face. When an organization is new to big data, it typically has (relative to managing big data) inadequate staffing or skills (40%), inadequate data management infrastructure (23%), and immaturity with new data types and sources (22%)[5]. The cure is to dive in with training and new hires (or consultants, more likely), then work through the learning curve, as with any new project type.

Serious BDM efforts are unlikely without proper business support. It’s difficult for any new project type to get off the ground when it lacks governance or stewardship (33%), business sponsorship (33%), or a compelling business case (27%). [5]

Solution design and architecture can be challenging, but not a showstopper. It takes time and angst to work through data integration complexity (30%) and the architecture of a big data management system (25%), but it’s doable for teams with solid data management experience[5]. In a related issue, it’s difficult to determine big data’s role in enterprise data architecture if you don’t have one (25%). This is one reason why many BDM solutions are silos.

Some problems aren’t much of a problem at all. A few issues ranked so low in the survey that we should consider them non-issues, namely loading large data sets (13%), fast processing of queries (9%), scalability with big data (7%), and network bandwidth (4%).

7. IMPLEMENTATION – VOLUMES OF BIG DATA BEING MANAGED

Everyone wants to know: How big is big data? What’s the volume managed today? How will that change in the future? To quantify these issues, this report’s survey asked: “What’s the approximate total volume of big data (by any definition of big data) that your organization manages, both today and in three years?”[5] (Figure 3.)

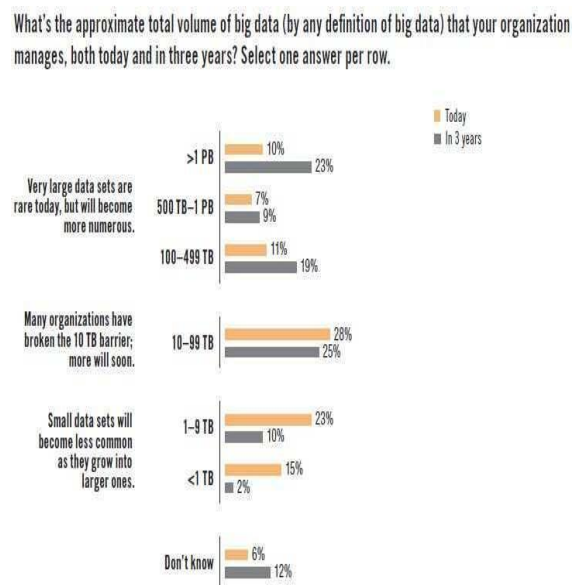


Figure 3: Based on 188 respondents who have experience managing big data.[5]

Many organizations have broken the 10-terabyte barrier.

In fact, the 10-to-99 TB range received more survey responses than other ranges, indicating that it's the norm for today's big data volumes. Within three years, 100 TB will become the norm.[5]

“ 10 to 99 terabytes is the big data norm today”

Smaller data sets will become less common as they grow into larger ones. In forecasting big data volumes for three years from now, survey respondents project far fewer data sets in the 1 TB and 1-to-9 TB ranges[5]. This is natural as big data repositories mature into greater volume. TDWI surveys on big data analytics (mid-2011) and high-performance data warehousing (mid-2012) showed near-identical declines in sub-10-TB data volumes.

“ Many firms anticipate breaking the one-petabyte barrier within three years”

Conversely, very large data sets are rare today, but will become more numerous. Looking at data volumes in the 100 TB and greater range, many more organizations will manage big data volumes in this range within three years (51%) as compared to today (28%). Furthermore, almost a quarter of users surveyed (23%) anticipate breaking the one-petabyte barrier within three years[5].

8. CONCLUSION

From this study, it is safe to say that data management plays a very vital role in our everyday life. From the uploading of multimedia files on social platforms to the amount of data bombarded into an organization's database. Data management is so unique in a way, so much so, that everything around us is typically data and can be managed – with the right guidance and resources. It has come to our knowledge that no matter how massive data is, it can be managed efficiently and effectively – with the right resources, by the right people and also, with the right technology. The knowledge of big data is now fast trending in the world of information technology. With the help of HADOOP, which is an open-source framework, it can store and process any amount of massive data in a distributed environment across clusters of computers using simple programming tools. Alongside HADOOP, are some main components (such as MapReduce, HDFS, HBase, ZooKeeper, Hive) which help in making sure that massive amounts of data are stored and can be processed even in so many years to come.

9. REFERENCES

- [1] IBM – Big Data – United States. Source: <http://www.ibm.com/big-data/us/en/>
- [2] S.A.S Institute. Source: http://www.sas.com/en_us/insights/big-data/what-is-big-data.html
- [3] Amanda L.W., Steven V.T, OSU Libraries. Slides 6-7. Source: www.slideshare.net/amandawhitmire/introduction-to-data-management-an-abbreviated-orientation-workshop
- [4] PennState University Libraries. Source: https://www.libraries.psu.edu/psul/pub cur/what_is_dm.html
- [5] TDWI Best Practices Report “*Managing Big Data – by Philip Russom*”, (October 1, 2013), pp 4-21. Source: <http://www.pentaho.com/resource/tdwi-best-practices-report-managing-big-data>
- [6] Strategic Sustainability Consulting – “*Data Management Concepts for Sustainability – Part 1*”. Source: <http://www.sustainabilityconsulting.com/blog/2015/8/11/data-management-concepts-for-sustainability-part-1>
- [7] American Institute of CPAs – Data Management. Source: <http://www.aicpa.org/INTERESTAREAS/INFORMATIONTECHNOLOGY/RESOURCES/DATAANALYTICS/Pages/default.aspx>