

National Institute of Technology Raipur

**DBMS
Term Paper Submission**

Big Data Management

By

Kunal Sachdeva

Roll No. 19115045

(5th Semester Computer Science & Engineering 2020-21)

Abstract

As businesses data stores have continued to grow exponentially, managing that big data has become more and more challenging. Businesses often find that the data they have has become outdated, that it is contrary to other data on their systems or that it is plain imprecise. In this paper, I am discussing the role of big data in businesses, The challenges faced while processing large quantity of data, The benefits and about various companies providing these services.

Introduction

Influence of that lack of trust may be significant. Businesses without precise data can miss out on the opportunities or even suffer reductions in brand value or consumer contentment. The Experian report said, “Nearly one in two organizations globally (52%) say that a lack of confidence in data contributes to an increased threat of non-compliance and regulatory penalties, and consequently, a downturn in customer loyalty (51%).”

To avoid those effects, businesses often turn to the discipline of managing the data. They set up data strategies and invest in a wide range of tools designed to help them deal with their stores of big data.

The term “big data” usually refers to data stores characterized by the “3 Vs”: high volume, high velocity and wide variety. Big data management is a wide-ranging concept that includes policies, procedures and technologies used for the collection, storage, management, organization, administration and delivery of big repositories of data. It can include data cleansing, migration, integration and preparation for use in reporting and analytics.

Big data management is closely linked to the idea of data lifecycle management (DLM). This is a policy-based method for determining which information should be stored where within an organization’s IT environment, as well as when data can be securely erased.

Within a typical company, people with several other job titles might be involved in big data management. They many include a chief data officer (CDO), chief information officer (CIO), data managers, database administrators, data architects, data modelers, data scientists, data warehouse managers, data warehouse analysts, business analysts, developers and others.

Big Data Management Challenges

In the IDG 2016 Data and Analytics Survey, 90 percent of those surveyed said they had experienced challenges related to big data management. Various factors make big data management more difficult than managing smaller repositories of data. Common issues include the following:

1. **Data silos:** Within most companies, various departments and business units use different applications and store data in separate databases. These separate databases may include similar data, but the data isn't always consistent from one database to another. For example, a retail outlet may store client addresses in a marketing database, a customer service database, an accounting database and an e-commerce website database. If just one of those databases has slightly different information for a particular customer — such as listing a customer's street address as an "avenue" when the other databases list it as a "street" — it could lead to problems like duplicate mailings, losing track of customer service records, double billing or inaccurate reporting. In addition, storing the same piece of information in several different locations eats up storage space — particularly when the problem is multiplied across an entire customer base.
2. **Growing data stores:** Managing big data is also difficult because of the sheer size of the data involved, compounded by the fact that it keeps getting bigger. In the customer address example above, fixing the customer records would be fairly easy for a very small company with only a hundred customers. In that case, someone could just look at the records involved and fix them. But for a national retailer with millions of customers and multiple petabytes worth of data, a different solution is necessary. In some cases, simply moving data around, say from a database into an analytics solution, can take a long time because of the large quantities involved. And performing any sort of operations on that data can slow performance to a crawl.
3. **Data and architectural complexity:** Not only is enterprise data stored in disparate siloes and constantly growing, today's data can be extremely complex. Enterprises often have both structured data (data that resides in a database) and unstructured data (data contained in text documents, images, video, sound files, presentations, etc.), and that data resides in a wide variety of different formats. A single enterprise may have thousands of applications on its systems, and each of those applications may read from and write to many different databases. As a result, simply cataloging what kinds of data an organization has in its storage systems can be a very

difficult job.

4. **Ensuring data quality:** All of these challenges make it very difficult for enterprises to ensure that their data is reliable and accurate. As already mentioned above, the lack of synchronization across data silos can make it difficult for managers to know which piece of data is correct. But data quality is also affected by another big problem — human error. In the Experian study, 56 percent of those surveyed said that human error was the biggest challenge that affected their data accuracy. Everyone makes mistakes when typing. But when a data-driven organization is using information typed by humans as the basis for major business decisions, simple typos could have potentially disastrous consequences.
5. **Inadequate staffing:** Another big issue complicating big data management is a lack of trained staff. There simply aren't enough data scientists and other big data professionals to fill all the available positions. As a result, salaries tend to be quite high. According to Indeed.com, the current average salary for a data scientist is \$130,235, while a data warehouse engineer typically makes \$112,607. By comparison, software engineers, which are generally some of the best paid employees in an IT department, earn an average of \$100,512. Clearly, the demand for big data skills exceeds the available supply.
6. **Lack of executive support:** Another potential challenge for big data management efforts is senior managers who do not understand the importance and value of good data management. Flashier technologies like predictive analytics and artificial intelligence may get a lot of attention — and budget — while the mundane processes of moving and cleaning data don't generate as much excitement. However, this problem appears to be diminishing somewhat. In the Experian study, the number of respondents citing inadequate senior management support as a big challenge to data management diminished from 21 percent in 2016 to 19 percent in 2017. And in the NewVantage Partners Big Data Executive Survey 2017, 52.5 percent of executives said that data governance was critically important to big data business adoption.
7. **Establishing a data-friendly culture:** For any organization, moving from a culture where people made decisions based on their gut instincts, opinions or experience to a data-driven culture marks a huge transition. The NewVantage study found that 52.5 percent of executives pointed to “organizational impediments” as a reason they had failed to achieve the goals of their big data projects, and only 27.9 percent of those surveyed said that they had been successful in their efforts to establish a data-driven

culture. Changing the mindset of employees and managers takes time, but most experts agree that it is necessary for big data management to be effective.

Big Data Management Benefits

Companies with successful big data management programs report a number of benefits, including the following:

1. **Increased revenue:** In Experian study, 61 percent of those surveyed said their data management efforts were helping their organizations increase revenue. In addition, 82 percent said specifically that data quality solutions helped grow revenues.
2. **Improved customer service:** In the same study, improved customer service was the second most-common benefit of big data management, cited by 56 percent of respondents. And in the IDG survey, customer service was the number one business objective of big data initiatives with 55 percent of respondents selecting it as a top goal.
3. **Enhanced marketing:** Although it is not as common a benefit as sales and customer service, marketing also sees a boost from big data management. Experian found that 39 percent of respondents said big data management enhanced their marketing efforts, and a whopping 85 percent said they saw “improvements in timely and personalized customer communications” as a result of data quality initiatives.
4. **Increased efficiency:** According to the Experian survey, 57 percent of those surveyed said maintaining high-quality contact data helped them increase efficiency. And the 2017 MIT Sloan Management Review report “Analytics as a Source of Business Innovation”, which was sponsored by SAS, found that some customers expected millions of dollars per year in efficiency gains as a result of big data projects.
5. **Cost savings:** Closely related to those efficiency gains is cost savings, experienced by 38 percent of those who took part in the Experian study. Similarly, in the NewVantage survey, 49.2 percent of those surveyed said their big data efforts had helped them decrease expenses.
6. **Enabling new applications:** Anecdotally, organizations report that when they have more confidence in their data it increases their innovation and inspires them to create new applications. In the NewVantage survey, 64.5 percent of executives surveyed said they were looking to big data to create new avenues for innovation, and 44.3 percent had achieved some success in this area.
7. **Improved accuracy for analytics:** And the most obvious benefit of good data management practices is that it increases the accuracy and reliability

of big data analytics. Good data coming in to the analytics solution sets the organization up for quality business insights coming out of the solution.

8. **Competitive advantage:** According to the MIT report, 57 percent of organizations using analytics reported a competitive advantage. Quality big data management practices enable those analytics, which in turn allow the companies to outperform their peers.

Big Data Management Best Practices

So how can organizations overcome the challenges of big data management and maximize the benefit from their efforts? Experts recommend several best practices:

- Involve team members from all the relevant departments in your big data management efforts. Big data management involves writing strategy, creating policies and transforming the organizational culture — not just investing in technology. In order to be successful in those efforts, it helps to have as many of the stakeholders involved in the process as possible. This includes members of the IT team as well as participants from the business side and, of course, an executive sponsor.
- Create a written strategy and policies for data lifecycle management. Having a written policy makes it much more likely that the policy will be implemented throughout the organization. In addition, many organizations need to have their data lifecycle management practices in writing for compliance purposes.
- Identify and protect sensitive data. With cyberattacks and data breaches in the news seemingly every day, organizations are more aware than ever of the need to protect corporate and customer information. Data management teams need to make sure that all the sensitive data in their systems is adequately secured and that data security teams are keeping up with the latest defensive strategies and techniques.
- Deploy strong identity and access management controls that include an audit trail. A key part of any data security plan is making sure that only authorized personnel can view or interact with sensitive data, as well as keeping track of who has seen or used the data and when. Again, these controls can also be important for compliance purposes.

- Invest in training for employees. Because big data experts are expensive and in short supply, it makes sense to develop big data talent from within. Helping current staff obtain big data skills can be a win-win for the company and the employee.
- Enable data sharing across your organization. According to the MIT report, “Companies that share data internally get more value from their analytics. And the companies that are the most innovative with analytics are more likely to share data beyond their company boundaries.”
- Consider appointing a chief data officer (CDO). This executive role is becoming increasingly common in large enterprises. The NewVantage survey found that 55.9 percent of executives surveyed said that their organization had a CDO. When asked about what the CDO should do, 48.3 said he or she should drive innovation and a data culture, while 41.4 percent said the CDO should manage data as an enterprise asset. Less than 4 percent said the role was unnecessary.

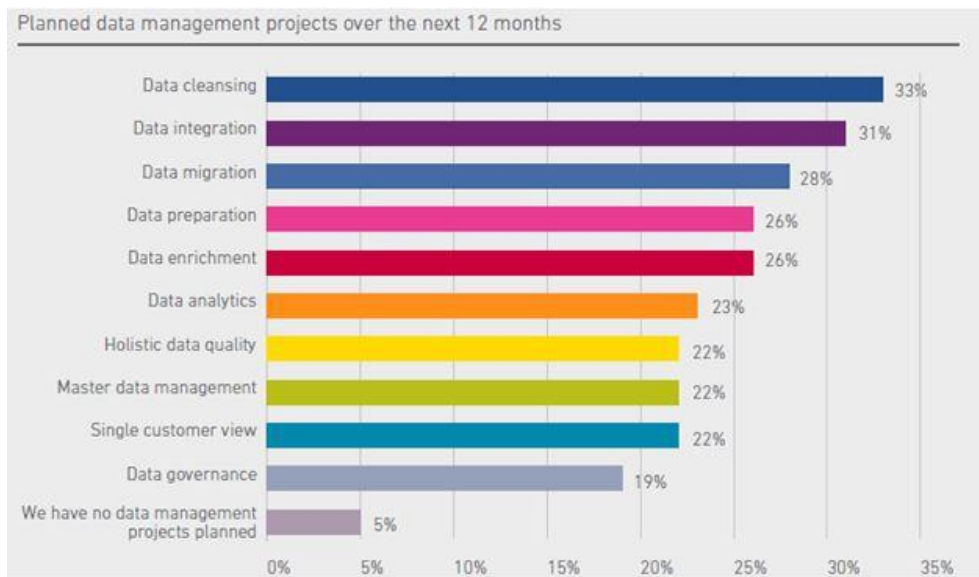
Big Data Management Services

When it comes to technology, organizations have many different types of big data management solutions to choose from. Vendors offer a variety of standalone or multi-featured big data management tools, and many organizations use multiple tools. Some of the most common types of big data management capabilities include the following:

- **Data cleansing:** finding and fixing errors in data sets
- **Data integration:** combining data from two or more sources
- **Data migration:** moving data from one environment to another, such as moving data from in-house data centers to the cloud
- **Data preparation:** readying data to be using in analytics or other applications
- **Data enrichment:** improving the quality of data by adding new data sets, correcting small errors or extrapolating new information from raw data
- **Data analytics:** analyzing data with a variety of algorithms in order to gain insights
- **Data quality:** making sure data is accurate and reliable
- **Master data management (MDM):** linking critical enterprise data to one

master set that serves as the single source of truth for the organization

- **Data governance:** ensuring the availability, usability, integrity and accuracy of data
- **Extract transform load (ETL):** moving data from an existing repository



into a database or data warehouse

Source: Experian Data Quality 2017 Global Data Management Benchmark Report

Well-known big data management solution vendors include the following companies: IBM, Microsoft, Oracle, Orchestra Networks, SAP, SAS, Zoho, etc.