



ESSAY

It's time to modernize your big data management techniques

Thomas H. Davenport and Ashish Verma

THE LAST MAJOR period of data management innovation was in the 1980s. Companies began to realize then that they needed a permanent place to store the data used for business intelligence and analysis. Wells Fargo Bank took delivery, for example, of its first enterprise data warehouse (EDW) system in late 1983. This leading edge-system employed parallel processing of relational database data, and many other firms found it a useful technology.

But the data management technology used successfully for the last 30 years is not the most efficient and effective technology for today. Many forms of big data, including images, social media, and sensor data, can be difficult to put in the row-and-column relational format usually required for an EDW. Their volume also makes them expensive to store in a traditional EDW architecture.

Fortunately, over the last decade several new technologies have emerged that are radically changing what constitutes best practice in contemporary data management techniques, including Hadoop and other open-source projects, cloud-based architectures, approaches to managing streaming data, and new storage hardware environments. The price/performance of these tools is substantially better than for previous technologies, often by one or more orders of magnitude. Even mainstream vendors of the previous data management era are now offering a variety of products and services that incorporate these new technologies.

Let business objectives drive the change

But the availability of better technology is far from the only reason to modernize your data environment. Business needs are leading to substantial change in the data environment as well, and should be the ultimate driver of modernization initiatives. The business objectives that could motivate a new approach to data include an increased emphasis on understanding and predicting business trends through analytics, a desire for machine learning and artificial intelligence applications in key knowledge-based processes, the need to stream data from

and to machines using the Internet of Things, or increased security and privacy concerns. In many cases, these goals simply can't be accomplished without data modernization. A sound business case will be critical to organizations seeking to modernize their data; otherwise, the effort will feel like an abstraction.

At Disney, for example, the primary driver of a modernized data platform was a need for better analytics. Entertainment and media products were traditionally released into the market with little ability to measure their consumption, but now almost all of today's media offerings can be measured and their audiences analyzed. To enable a diverse range of analytical activities, Disney developed a road map for a sophisticated data and analytics capability, including a data lake, a new set of analytics tools, and a set of business use cases to take advantage of the new technologies.

Data lakes and their management

These types of projects typically result in the implementation of a data lake, or a data repository that allows storage of data in virtually any format. Data lakes are typically based on an open-source program for distributed file services, such as Hadoop. They allow large-scale data storage at relatively low cost. However, there are multiple approaches to data lakes; for example, some are based in the cloud, some on premise. Different data lake approaches also provide for different levels of security and governance. Therefore, it's important to plan a modernization effort carefully before implementing any particular technology.

Data lakes must also be carefully managed in order not to become "data swamps"—lakes with low-quality, poorly catalogued data that can't be easily accessed. And at some point, most unstructured data based in a data lake will need to be put in structured form in order to be analyzed. Data lakes, then, require that management approaches be defined in advance to ensure quality, accessibility, and necessary data transformations.

Deloitte helped one global technology firm, for example, transition from a 600 terabyte enterprise data warehouse to a data lake platform. The data is used by 2,800 employees, so the conversion process needed to involve minimal disruption. Lake storage still uses on-premise technologies, but the company now has a “consumption layer” in the cloud for easy and rapid access by users and automated processes. And instead of the time-honored “extract, transform, and load” (ETL) process, data is only transformed when necessary for analysis. In other words, it’s an ELT process.

Most organizations establishing data modernization approaches also try not to lift and shift existing data into the new data environment. Instead, they attempt to make improvements in the data at the same time, increasing integration and quality across the enterprise. Firms are increasingly using tools like machine learning to allow probabilistic matching of data; using this approach, data that is similar but not exactly the same as other data can be matched and combined with little human intervention. This bottom-up method of data integration can sometimes be faster and more effective than more top-down approaches to integration like Master Data Management.

The global pharmaceutical company GlaxoSmithKline, for example, used this approach to modernize and integrate its data for research and development. It was able to combine millions of data elements from three different domains—experiments, clinical trials, and genetic screenings—into a single Hadoop-based data lake. The company was able to incorporate 100 percent of the desired data into the lake within only three months. To work across the three domains, the data team created an integrated semantic layer on top of them with standardized definitions and meanings, and is now working on over 20 different use cases for data within the lake.¹

Making it happen

Companies we’ve seen that are successful at data modernization have several common attributes. They include:

- Involvement of key business executives—typically some combination of the CEO, chief information officer, chief financial officer, or chief marketing officer—to define data-related business goals and ensure resource commitments. They should be stakeholders throughout the entire process of modernization.
- A well-defined process, typically involving a set of “Imagine” activities to design the new data environment; “Implement” activities to design, develop, build, test, and roll out the modernized capabilities, and a set of “Operate” activities involving application and infrastructure maintenance, ongoing enhancements and new releases, and development of additional capabilities over time. Each of these steps works best in a series of agile sprints.
- Constant and consistent value delivery; each sprint should attempt to deliver measureable value. The enemy of successful data modernization is most commonly organizational fatigue and the inability of executives to see value, not technical limitations.
- New human capabilities are needed to implement and manage new data technologies. Since the skills to build and manage data lakes are in short supply, companies need to plan at an early stage for how they will source the talent for modernized big data management.

Business rewards are in store for the companies that succeed at these data modernization initiatives. Similarly, organizations that fail to undertake or succeed at modernization projects could find themselves at a competitive disadvantage from their inability to implement data-intensive business models and strategies.

1. Thomas H. Davenport and Randy Bean, “Biting the big data management bullet at GlaxoSmithKline,” *Forbes*, January 8, 2018.

Deloitte.

Insights

Sign up for Deloitte Insights updates at www.deloitte.com/insights.

 Follow @DeloitteInsight

Deloitte Insights contributors

Editorial: Ramani Moses, Abrar Khan, and Blythe Hurley

Promotion: Shraddha Sachdev

Artwork: Kevin Weier

About Deloitte Insights

Deloitte Insights publishes original articles, reports and periodicals that provide insights for businesses, the public sector and NGOs. Our goal is to draw upon research and experience from throughout our professional services organization, and that of coauthors in academia and business, to advance the conversation on a broad spectrum of topics of interest to executives and government leaders.

Deloitte Insights is an imprint of Deloitte Development LLC.

About this publication

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or its and their affiliates are, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your finances or your business. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

None of Deloitte Touche Tohmatsu Limited, its member firms, or its and their respective affiliates shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.