The Limitless Applications of Analytics

The amount of data we use and manage doubles annually, and performing analytics on that data can uncover heretofore unknown insights that lead to competitive advantage.

By Craig S. Mullins

There is no denying that big data and analytics are now entrenched and vital components of modern business and the information technology systems that power today's organizations. Although the big data trend has cooled off, the reason for amassing all of that data and performing analytics on it for business insight has become a permanent fixture in most organizations.

Analysts estimate that the amount of data we use and manage doubles annually, and performing analytics on that data can uncover heretofore unknown insights that lead to competitive advantage. Furthermore, the big data used to power analytics is being adapted for use by AI and machine learning software that will further improve the return on our computing investment through automation of processes and tasks, thereby increasing productivity and operational efficiencies.

Types of Analytics

Before moving on to discuss the trends, let's briefly examine the four types of analytics: descriptive, diagnostic, predictive, and prescriptive.

- Descriptive analytics focuses on what has already occurred.
- Diagnostics analytics focuses on understanding why things happened.
- Predictive analytics focuses on what could happen next.
- Prescriptive analytics attempts to inform us as to what should happen next.

All are important and growing within organizations today. Nevertheless, there are clear trends that resonate with current analytics efforts today.

KEY TRENDS IN ANALYTICS TODAY

Continuing Impact of the COVID-19 Pamdemic

Any discussion of the current state of analytics has to begin with the way the pandemic has significantly altered many aspects of not just business but also everyday life across the globe. The disruption to business as usual imposed on organizations as healthcare and government officials attempted to slow the spread of COVID-19 caused many significant and, possibly permanent, changes to how we work.

The impact of the pandemic has been to speed up the need for digital transformation. Improved data analytics is one aspect of this.

Social distancing and work-from-home (WFH) efforts advanced the notion of "distributed everything"—data, devices, and people. Organizations had to adapt to this distribution and have been aided by data analytics.

With a WFH workforce, HR analytics can be an important tool for monitoring the health of both companies and employees. Analytics can be used to increase the effectiveness of onboarding, employee engagement, morale, and productivity. Of course, this type of data should be anonymized to protect the privacy of individual employees before it is shared within the company.

Analytics also has played a role in understanding and reacting to the COVID pandemic. Organizations that have adopted analytics to monitor consumer spending data, coupled with other data (such as vaccination rates, COVID hospitalization cases, and governmental rules), are better prepared to survive the continuing pandemic. By spotting trends and predicting changes in upcoming economic patterns, companies can make better decisions about staffing, supply chain management, stocking, and so on.

Most consumers are not capable of predicting their own behavior shifts, but data analytics can sift through multiple types of data and deploy models to predict changing consumer activities. This can have a far-reaching effect on the economy as businesses need to remain nimble to react to the complexities of the health crisis caused by COVID and its variants.

Furthermore, the application of analytics to the pandemic is not just for commerce. Government agencies can rely on analytics to recommend and set policies based on reliable, data-based predictions. And school districts can use analytics to better plan for when schools should be shut down, reopened, or conducted online.

AI-Augmented Analytics

Perhaps the most significant trend impacting data and analytics is the integration of AI into the process, commonly referred to as augmented analytics. This was first identified by Gartner and defined as using enabling technologies such as AI and machine learning (ML) to bolster data preparation as well as the generation and explanation of insights derived from data.

Organizations are increasingly deploying augmented analytics to improve the description and connection of multiple internal and external data sources and to find relationships within the data. The difference is that augmented analytics automates many aspects of the process, thereby making it possible for non-experts to perform work such as developing, managing, and deploying models, which previously required skilled data scientists.

Of course, as is the case with most AI capabilities, adoption of augmented analytics is nascent, but growing. As data volumes continue to expand, automating analytics with AI will continue to escalate in importance as it becomes impossible to manually deal with the incredible volume of data.

Using AI and ML to augment data preparation is another aspect of augmented analytics adoption that deserves to be mentioned. The manual tasks required to prepare data before it can be modeled or analyzed are some of the most time-consuming aspects of data analytics and data science. Automating data preparation with AI and ML to profile and enrich the data, define metadata, catalog the data, and perform the other mundane aspects of getting data ready for analysis is being used to accelerate analytics processes. As these AI and ML capabilities improve over time, expect to see a reduction in the amount of upfront planning and preparation time required. In addition, data quality may be improved also as the AI/ML techniques improve.

Obviously, the problem of integrating AI into analytics has not been completely solved. Many issues remain, such as a lack of skilled technicians, the cost of adoption, software immaturity, and the length and difficulty of pilot projects. Nevertheless, AI is burgeoning and will continue to be embraced as augmented analytics improves and succeeds.

As more AI capabilities are embedded into hardware, the velocity of augmented analytics adoption will increase. One such effort to keep an eye on is Telum, IBM's latest processor for its IBM Z mainframe computer. Telum contains on-chip acceleration for AI inferencing while a transaction is taking place. Announced in August 2021, the first Telum-based mainframe is expected to be available in the first half of 2022.

Adoption of Data Fabric Architecture

A data fabric architecture is deployed to simplify and integrate data management in a hybrid environment spanning both on-prem and cloud implementations. The data fabric integrates data and connecting processes using analytics over a set of data services to provide consistent capabilities for accessing and using data across the entire hybrid environment.

Most organizations rely on a myriad of operational systems, both on-prem and in the cloud, that are based on multiple technologies and requirements. The data fabric confers the following benefits on the analytics processes that take advantage of it:

- Access to the business, operational, technical, and social metadata that defines the data through the data fabric, as well as the continuous analysis, modification, and depiction of metadata assets.
- Continuous learning with predictions for data management and integration requirements and capabilities.
- The ability to create and curate knowledge graphs that enrich data with semantics to make it more intuitive for data analytics usage.
- Built-in compatibility for multiple data integration techniques, such as ETL, replication, streaming, and so on.

These benefits make it easier to access data that exists in multiple silos throughout the organization. As such, deploying a data fabric architecture can help to alleviate vexing problems such as poor data quality, missing data, and inaccessible data. Building upon a single data fabric eliminates problems that arise from using multiple and different types of database systems with disparate data management tools.

Although adoption of a data fabric architecture is only just emerging, prescient organizations are deploying data fabric architectures to modernize data management and data integration for analytics. Organizations that store copies of data sourced from multiple (often thousands) of systems can benefit greatly from the data fabric architecture.

Marriage of the Cloud and Analytics

Less of a current trend and more of an ongoing one, deploying analytics in the cloud continues to gain traction. Cloud analytics—or carrying out analysis using cloud computing—requires a range of analytical tools and techniques to help companies extract information from big data and present it for consumption, typically via a web browser.

Until recently, cloud analytics was generally considered to be analytics performed exclusively on cloud data, but this is no longer the case. For medium-to-large enterprises, the trend is toward hybrid analytics, where cloud analytics solutions (those that run in the cloud) can be used to analyze any data, whether it is on-prem, in the cloud, or both. Analytics services and tools are expected to be able to satisfy such scenarios, and usage of those that cannot is waning.

Adoption of Graph Technology and Techniques

The adoption and usage of graph techniques and database systems is accelerating because of their strong ability to map and analyze relationships. Graph theory is based on the mathematical graph structure, which is used to model relations between objects.

When massive amounts of data are stored throughout an organization, in the cloud, and distributed across multiple environments, the use of graph techniques makes it possible to discover and manage multiple connections between data sources. Furthermore, when organizations need to perform analytics primarily on relationships between data, rather than on the data itself, graph technology can be invaluable.

The Economics of Analytics

Organizations are increasingly implementing data analytics projects with the goal of understanding, measuring, and monetizing these projects instead of treating analytics as experimentation. The actionable insights achieved by data analytics projects can contribute to the bottom line of an organization in many significant ways.

By applying economic principles to the deployment of information assets, the insights from analytics can be used to improve innovation, make better decisions, generate increased value, and thereby improve organizational success.

Clearly, a comprehensive approach to the economics of analytics requires organizational effort and buy-in, a firm understanding of both economic and data management principles, and much more. For an in-depth study of such efforts, consult the recent books, *Infonomics* by Douglas Ladley and *The Economics of Data, Analytics, and Digital Transformation* by Bill Schmarzo.

As part of this effort, organizations are shifting data and analytics to become a core business function, instead of a component of IT that supports the business. The long-awaited shift to treating data as a business asset appears to be closer to becoming a reality than it ever has been.

Edge Analytics and DataOps

The growth in the amount of data being created and stored is a trend that has continued unabated for some time now. And with the increasing number of sensors being connected to IoT, more data is being generated and processed every day.

The growth of edge analytics has resulted from this confluence of trends. Edge analytics refers to the process of gathering, analyzing, and creating insight directly from the devices connected to IoT, instead of passing the data from the device to a central computing environment. Edge analytics goes one step beyond edge computing by performing the analytics on the data where it is captured.

Yet another trend resulting from data growth is that more organizations are deploying DataOps, a methodology for automating data management to improve data quality and reduce the time needed for preparing and utilizing data for analytics. DataOps is not just DevOps for data. DevOps is a philosophy and set of practices that integrate development and operations with the goal of improving software delivery and quality. DataOps encompasses practices, workflows, procedures, and cultural norms to bolster and improve the lifecycle of data. DataOps requires collaboration between and among all pertinent IT professionals, from data architects to DBAs and data scientists to developers. It improves all aspects of data management through processes such as data creation and ingestion, data quality steps, data preparation, data accessibility, reporting, and beyond. As the amount of data continues to grow, DataOps is becoming a crucial component of being able to glean insight from analytics.

As might be expected, all of these trends have resulted in increased demand for expertise and solutions. Organizations are willing to expand their investment in analytics talent, software, and training. For example, the U.S. Bureau of Labor Statistics forecasts that the number of data scientist jobs will increase by approximately 28% through 2026, which equates to about 11.5 million new jobs.

WHAT'S AHEAD FOR THE APPLICATION OF ANALYTICS

The applications of analytics are as varied as the number of businesses and requirements that exist. Any business that can benefit from a clearer view into its operations can benefit from descriptive analytics. And what business would not be interested in using analytics to understand why things happen, what may come next, and what should be done?

For example, consider the range of applications for advanced analytical capabilities for operational applications such as logistics, security, and more.

Analytics on logistics data can determine the most appropriate shipping routes and predict approximate delivery times. Using GPS systems and sensors in delivery vehicles, actual delivery data can then be funneled back into the analytics system to improve upon routes and delivery estimates.

Improving security is another key application of analytics. Fraud accounts for billions of dollars in losses annually, particularly for the banking, insurance, and healthcare sectors, but for many others as well. Continuing improvements to fraud analytics systems are helping to identify and prevent fraud via statistical analysis techniques that range from simple statistics such as averages and quantiles to more sophisticated techniques such as regression analysis, probability distributions and models, data matching, and time-series analysis.

There is even an Israeli startup company, BeeWise, that has built a robotic beehive system that uses analytics and AI to provide around-the-clock care to analyze beehives for disease, monitor them for pesticides, and report in real time any hazards threatening a colony. The company claims that its big data analytics can be used to rescue up to 80% of bees that would have otherwise perished.

Indeed, the applications of data analytics are nearly limitless. ■

Craig S. Mullins is president and principal consultant with Mullins Consulting, Inc. (www.MullinsConsulting.com.). He has more than 3 decades of experience in all facets of data management and database systems development. Mullins is the author of three best-selling books, *DB2 Developer's Guide, Database Administration: The Complete Guide to DBA Practices & Procedures,* and *A Guide to Db2 Application Performance for Developers*. Mullins is also an IBM Champion for Analytics, a DB2 Gold Consultant, and a member of the IDUG Volunteer Hall of Fame.