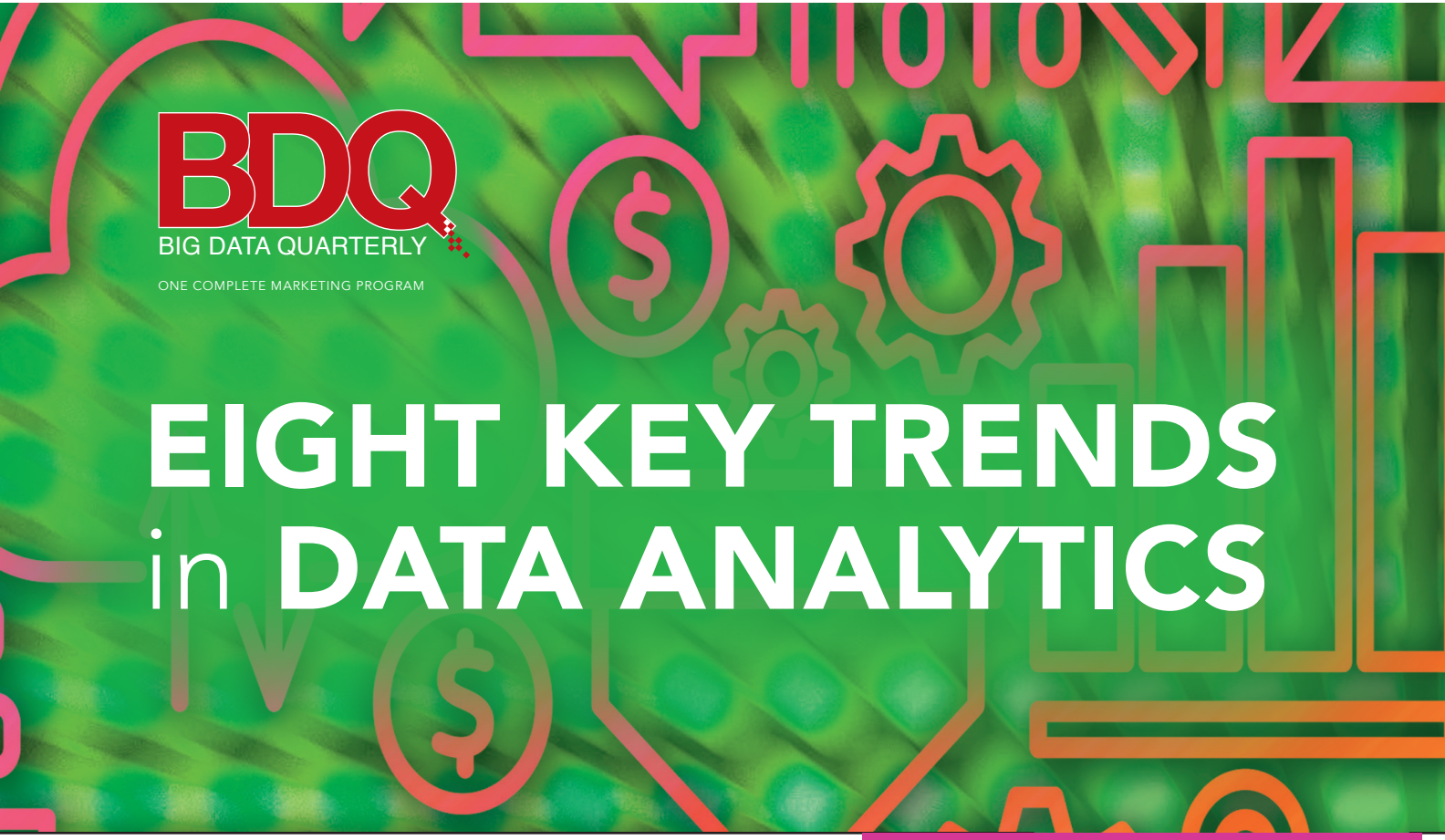
BDQ
BIG DATA QUARTERLY
ONE COMPLETE MARKETING PROGRAM

TOP TRENDS in ANALYTICS TODAY

Best Practices Series



EIGHT KEY TRENDS in DATA ANALYTICS

Best Practices Series

Data analytics is no longer the luxury of organizations with large budgets that can accommodate roving teams of analysts and data scientists. Every organization, no matter the size or industry, deserves a data analytics capability. Thanks to a convergence of technology and market forces, that's exactly what's happening.

Factors transforming data analytics in recent years—making it available across the board—include the cloud, the availability of open source platforms and frameworks, and the emergence of data-intensive technologies such as AI, machine learning, and the Internet of Things. This year, these trends are all converging, elevating data managers to expanded roles as advisors to and leaders of their businesses.

Trends shaping the data analytics landscape—as well as the careers of people working with data—include the following:

1. DATA IS BECOMING STRATEGIC

Data managers are seeing their mandates extend well beyond their original roles as administrators and troubleshooters in managing and securing day-to-day transactions. It's now about leveraging information to make strategic, operational, and tactical decisions that result in increased revenue, improve operational efficiency, and enhance customer experience.

2. CUSTOMER SERVICE IS BEING DRIVEN BY DATA

Superior customer experience, or CX, has increasingly been acknowledged as a competitive differentiator and is driving a closer alliance between marketing, customer service, and IT or data managers. That's because CX almost wholly depends on the intelligence and insights provided by data. In addition, personalization—in which data is sourced to deliver targeted customer experiences—has emerged as a leading priority on business leaders' agendas. The "experience" aspect doesn't just stop at customers either—data needs to deliver highly personalized experiences to end users such as employees.

3. MORE SELF-SERVICE

There is a growing emphasis on enabling end users to create their own queries to ask any question at any time of their data—without having to requisition reports from their IT departments. Vendors recognize the need for such flexibility, as well as the challenge of finding data across enterprises, and are responding with user-empowerment tools and offerings. From an organizational point of view, enterprises need to ensure that the process for acquiring and integrating this data is well-managed and well-governed, with policies and frameworks that assure the information users generate is trustworthy and reliable.

4. THE DRIVE TOWARD DATA QUALITY INTENSIFIES

As data has become ever-more critical to the business, the need for decision makers to be able to trust their data has grown. Data quality is a process that begins as data sources are identified and accessed, extending to managing and storing the data. There have long been robust tools and applications

on the market to help ensure data quality at the granular level, in terms of deduplication and cleanliness, but an effective data quality effort also needs to encompass its timeliness and the governance policies that surround it. Decision makers need to be assured that their real-time analytics and AI systems are employing the highest-quality data available.

5. DATAOPS AND DEVOPS

Not only does data need to be readily available for decision making, but a process needs to be put in place that ensures that it is moved and processed on a continuous basis, as automatically as possible. That's why many enterprises are turning to emerging practices such as the DataOps and DevOps models of continuous integration and continuous delivery. DevOps is concerned with the process for developing and delivering application releases, while DataOps is an automated, process-oriented methodology to improve the quality and reduce the cycle time of data analytics. Lately, both of these methodologies are being applied to enable organizations to move quickly to have access to the latest algorithms and data to stay on top of their markets.

6. AI AND MACHINE LEARNING

No discussion of the power of data analytics in 2019, of course, can take place without including AI and machine learning. For starters, AI is replacing standard BI reporting as we've known it for decades, with real-time insights and updates on changes or developments within selected areas of the business. With AI, of course, many analytics-driven decisions can be made and executed without human intervention. At the same time, at a higher level, AI is assisting decision makers in understanding what data is telling them.

7. BOTS AND DIGITAL WORKERS

Another trend on the horizon for data analytics is the use of robotic process automation and the rise of digital workers. As more "bots" take on the day-to-day tasks of back-office work—such as managing workflows or searching files—data analytics is increasingly laying the groundwork for their intelligent performance.

8. DATA ECOSYSTEMS EXPANDING

Data environments are no longer insular systems contained within corporate walls. The ability to deliver and act on data-driven insights is increasingly amplified by connected ecosystems of partners, customers, and other constituencies. Data-driven enterprises are learning to bring together expertise and knowledge from both inside and outside their corporate walls to deliver growth and innovation.

One thing is clear: When it comes to all the possibilities data analytics offers, 2019 is a year of transformation. Data has never been more closely tied to the success of businesses, which means new opportunities for growth and advancement among the data professionals who are leading the way.

—Joe McKendrick

Today's Analytics Challenge:

Bringing Real-Time Insight to Transactional Applications

Like everyone else, you have databases for transaction processing (OLTP) and data warehouses for analytics processing (OLAP). The databases support customer-facing applications. The data warehouses support internal BI/analytics teams. However, while analytics has become a competitive differentiator in every industry, it is often restricted to data warehouses and the internal BI/analytics teams using them (and other tools).

In the beginning, businesses began creating competitive differentiation by using analytics to derive valuable insight, and using it to improve operational efficiency, mitigate risk, and more. Now, the challenge is using analytics within customer-facing applications to create competitive differentiation by improving customer engagement and experience—by providing customers, rather than the business, with valuable insight (directly and/or indirectly).

This type of analytics, needed by customer-facing applications, falls somewhere between databases and data warehouses. It does not rise to the level of needing a data warehouse, but it exceeds the analytical capabilities of databases. This is the problem. For example, you may use a data warehouse for market basket analysis and the database to look up product information and store purchases, but what if you want to alert customers about "soon-to-be-sold out" products based on the products in shopping carts and recent purchases and their current inventory?

The database is great for standard CRUD operations such as viewing a product, adding it to a shopping cart, and completing a purchase—as well as for searching, filtering, and sorting products. These queries read or write most, if not all, columns in a row, and use indexes. The data warehouse is great for performing a market basket analysis to improve product recommendations. This type of query reads a small number of columns in a row, and it will not benefit from indexes—it accesses every row.

So, where do customer-facing applications send queries to sort, filter, and aggregate data in many rows without the help of indexes? These queries may need to aggregate a column for many rows in one table and use the results to look up a small number of rows in a different table. They are both column-oriented and row-oriented. For example, to alert customers of the top five "soon-to-be-sold out" products based on the quantity in active shopping carts, the quantity in recent purchases (say the last 24 hours), and quantity in current inventory. This query will use sorting, filtering, and aggregation to find the product ids for the top five "soon-to-be-sold out" products—it's column-oriented and analytical. It may access thousands, if not millions, of shopping carts and purchases—all to get the ids of five products. It will then use them to look up five rows in the products table by product id—row-oriented and transactional.

The solution is a database capable of four things. First, storing table data in both row and columnar storage. Second, synchronizing row storage with columnar storage so all writes to row storage are replicated to columnar storage. Third, joining a table with row storage to a table with column storage. And fourth, routing queries to tables based on query type (transactional or analytical) and the underlying storage format (row or columnar) of the tables. For example, a query to get to display a shopping cart should access the row storage while a query to analyze the contents of many shopping carts should access the columnar storage.

This is the architecture of MariaDB Platform, an enterprise open source database for transactional, analytical, and hybrid transactional/analytical workloads. It uses the InnoDB storage engine for transaction processing with row storage and the MariaDB ColumnStore storage engine for analytics processing with columnar storage, synchronizes the two via streaming change-data-capture, supports cross-engine joins, and performs dynamic query routing based on configurable rules and syntax. It's engineered to provide transactional applications with modern analytics—to deliver real-time insight to customer-facing applications.

LEARN MORE ABOUT MARIADB PLATFORM:
https://mariadb.com/products/mariadb-platform/

Information Builders

AI, ML, and Analytics at Scale

The data science business is booming, but it's less clear whether businesses are booming because of data science. Companies need to focus on maximizing the value that data scientists provide to justify their cost and ensure they fulfill the promises of AI and ML, and that means scaling AI and ML out to the enterprise and beyond.

Data scientists can create sophisticated AI or machine learning models, but they often do so in programming environments such as R, python, and TensorFlow—and they don't like to change environments without good reason. That can make them very useful for small-scale projects, but harder to capitalize on at scale. This in turn means they only achieve a fraction of their potential impact on an organization.

Take a predictive model for customer churn, for instance. It's useful if five people in marketing can use it to create a campaign to target these customers with cost-effective offers. The data scientist can get a database extract and run the model against it, and then hand the results to the marketing team.

That's great, but it has far less value than if five hundred people in a contact center can use this model as part of their day-to-day interactions with customers. If every one of those interactions is smarter, the overall impact on customer retention could be far higher than the marketing campaign might hope to achieve.

Bringing the impact of AI and ML to scale, then, depends greatly on two key factors: embedding AI and ML into other analytics, and embedding them into other applications.

EMBEDDING AI AND ML INTO OTHER ANALYTICS

First, consider data-intensive positions with data-savvy employees: that marketing team, for instance. They probably already run their businesses off of dashboards and analytical tools. If they can create their own charts, reports, and dashboards, they can leverage the data scientist's customer churn model without using the data scientist's time.

There's a prerequisite, though: They need to be able to call the model as if it were an ordinary field or transformation. In other words, the model has to be integrated into the analytics platform in a way that makes it seamless to the data-savvy employees who use their tools.

Doing this saves time for the data scientist, allowing him or her to work on new projects, and makes the marketing team more efficient.

EMBEDDING AI AND ML INTO OTHER APPLICATIONS

Next, consider those customer service representatives, who don't have the experience or tools available to create their own analytics, reports, or dashboards.

The vast majority of them don't really want "data science" or "predictive analytics." Instead of data science, they need information. Instead of predictive analytics, they need predictions. They just need help deciding what to do.

For them, a simple stoplight on their screen—red, yellow, or green, according to the likelihood of churn—will make a big difference. Even better? The three offers they can make to this customer that are most likely to prevent churn.

To be effective, the stoplight can't be in some other dashboard they have to turn to in time of need. It should be on their screen from the moment they answer an incoming call. And that means embedding AI and ML into an application, not just a dashboard.

In many cases, such as Salesforce.com, there's a portal into which analytics can be embedded. In other cases, there are APIs that can be used to embed analytics. Regardless, bringing the results of AI and ML directly into the workflow of non-technical users is one of the most important ways to leverage AI and ML at scale.

INFORMATION BUILDERS: DATA AND ANALYTICS AT SCALE

Information Builders provides a complete data and analytics platform that makes data easy to integrate, govern, and share. This provides the trusted data foundation needed for creating AI and ML models, while also enabling business users—data-savvy and less data-oriented alike—to get the tools or apps needed to make better decisions.

A key part of this platform is the tight integration of WebFOCUS, Information Builders' highly scalable analytics and BI product, with R, Python, and other languages. Data scientists don't need to give up their preferred tools to contribute AI and ML models to applications that can scale from ten users to ten million.

And WebFOCUS is designed to work standalone or to embed into other applications, whether it's a Salesforce portal or a custom app.

INFORMATION BUILDERS
www.informationbuilders.com