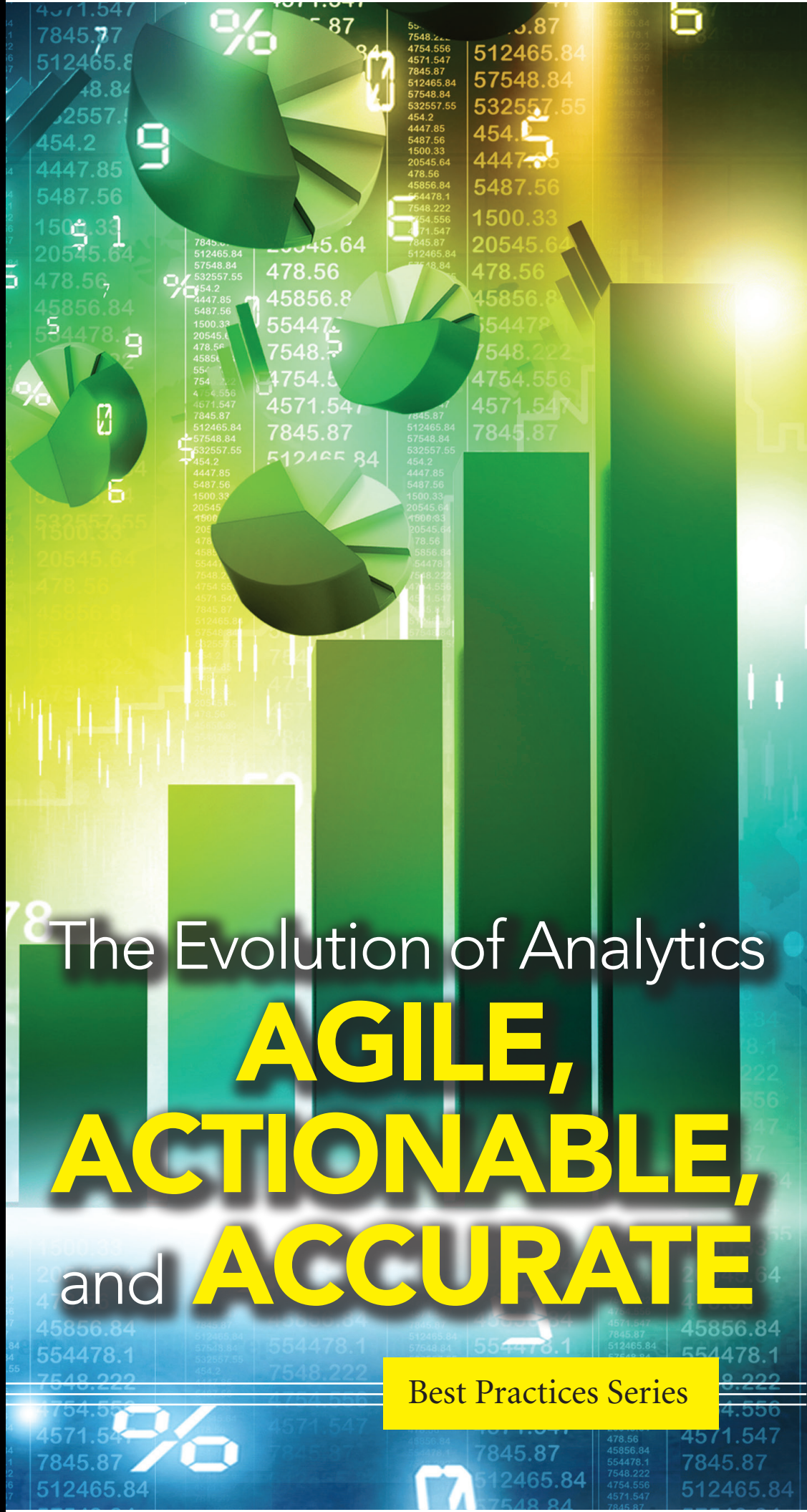


Collibra

MariaDB

Cloudera



The Evolution of Analytics

# AGILE, ACTIONABLE, and ACCURATE





**database**

TRENDS AND APPLICATIONS  
ONE COMPLETE MARKETING PROGRAM

# Eight Steps Toward AGILE, ACTIONABLE, and ACCURATE Data Analytics

Best Practices Series

IN THE HEYDAY of business intelligence tools, analysts and power users would pore over data to assess where their organizations stood in the preceding month or quarter. Companies held a few licenses for the desktop applications and often had multiple vendors' solutions onsite. Now, the emphasis is on analytics for all and analytics based on what's happening now or about to happen.

However, it's still going to take time to reach a data analytics-driven nirvana. Siloed data, a lack of tools, a general unawareness of data assets, and skills deficiencies need to be overcome. A survey of 314 senior executives by Forbes Insights and Dun & Bradstreet found

“analytical methods and tools trail both the appetite and ambition of most business leaders.” For example, 23% of analytics professionals are still using spreadsheets as their primary tool for data analysis. Only 38% of respondents felt strongly that business leaders took full advantage of their analytics initiatives.

There is clearly work to be done. Here are key considerations for turning data into agile, accessible, and actionable insights:

## DEVELOP AN ANALYTICS BUSINESS PLAN

Business context is everything—this is the essence of decision making. This is determined by looking at the data and

defining its implications and its role in the business. Work with business decision makers to determine where their pain points—or data blind spots—may be, and address those first. Developing an enterprisewide analytics infrastructure is not an overnight project—it requires investment and planning. Confer with business associates and develop a road map of where and when solutions will be activated. The analytics business plan should include a statement of the problem, the solution or technology proposed, the investment required, who will be involved, and potential return on investment. In addition, the business plan needs to be continually revisited

and revised as needed. Most importantly, the plan needs to be action-oriented, with clear steps defined.

### TAKE STOCK OF CURRENT DATA ASSETS

To plunge into an analytics initiative without awareness of the specific data assets that could be of benefit means many resources will be passed over. Many decision makers aren't even aware of the data that may be residing within their enterprises. In addition, it's important to determine exactly what types of analytics the business requires to proceed, and make recommendations when there is uncertainty.

### CAST A WIDE NET

Powerful and insightful analytics may come from places not even considered yet—such as stored documents or log data from a particular system. An active analytics effort should encompass all possible data types—both structured and unstructured. Content-style data—such as documents and pictures—should also be a prime candidate for this initiative. Bring your structured data and content into a single view. Tools and platforms such as data lakes, Hadoop, and Spark help ensure that data of all types is available on demand. Ultimately, enterprise-wide integration is the key. What kind of databases are supporting the environment, and how can data from these various sources be brought together into a single view? How can organizational silos be opened up?

### KEEP TRUST FRONT AND CENTER

Trust is the glue of the analytics-driven enterprise. If an organization is to compete on analytics, decision makers need to be able to trust the data they are working with—that it is timely,

*An analytics-driven enterprise **requires people** who are skilled at building and managing such systems, as well as being **able to use** such approaches.*

cleansed, complete, and secure. Ensuring this a traditional role for database administrators, but with the overwhelming volume, variety, and speed of data, more automated methods—and ultimately machine learning—needs to be put in place to guarantee this quality.

### ENGAGE IN STORYTELLING

A good story is compelling and also has illustrations to match. Decision makers need to know, clearly and concisely, how given insights are affecting their business or the way they do their jobs. For many executives, the ability to quickly grasp concepts through simple and clear visualization makes it all worthwhile. The dashboard, scorecard, or graphics should be so clear that decision makers can spot anomalies, outliers or trends just by a quick glance.

### DEVELOP AND NURTURE THE RIGHT SKILLS

An analytics-driven enterprise requires people who are skilled at building and managing such systems, as well as being able to use such approaches. Importantly, people both inside and outside the data management department need to be trained in critical thinking and how to expand business thinking to accommodate innovations not possible before the analytics-driven infrastruc-

ture was put in place. Plus, data analysts and data scientists need to develop their storytelling skills so they can take statistics and present it to the business.

### MEASURE, MEASURE, AND MEASURE SOME MORE

As the saying goes, “What you don't measure, you can't manage.” Data being fed into analytics systems should be continually evaluated for its impact on the business. There are many types of metrics or key performance indicators that can be employed, depending on place within the business. Marketing managers would require different measures of success from IT managers or production managers.

### KEEP THINKING CRITICALLY

There will still be a regular need for human review or intervention to continually assess the value of this data to the business. Algorithms can go awry, and some experts, such as Cathy O'Neil, a mathematician and data scientist, caution against “blind faith” in analytics. In a recent TED Talk, she cautioned that many algorithms are not well understood, and may even reflect the biases of their creators. Algorithms and the analytics that support them need to be constantly reviewed. ■

—Joe McKendrick



# Closing the Data Engagement Gap

It's NO SECRET THAT data powers modern business. It personalizes consumer experiences, drives competitive advantage, and sparks new product innovation.

But as the volume of data available grows, a significant challenge emerges: The data engagement gap.

According to a recent study by MIT Sloan Management Review, the gap between more access to useful data and the ability to develop practical insights doubled, from 14% in 2012 to 28% in 2017. And worse, the ability to use data to guide future strategy dropped 6% from 2016 to 2017. How did this happen?

## A BRIEF HISTORY OF DATA ENABLEMENT

Using data to drive decision making is not new. For many years, organizations relied on centralized reporting to meet their data needs. These reports delivered consistent information about the business to controlled audiences. These select few studied the reports and used the information to drive decision making. But these reports had a flaw: They failed to provide insight to the masses. There was too much control over access.

Fast-forward a few years, and the focus shifted. Self-service BI and analytics tools entered the market, promising broader data access for more people. Now, business analysts and data scientists could access data, develop reports and visualizations, and glean insight from data that was previously restricted to them. This approach, while expanding the audience to analysts and data scientists, had flaws: It (still) excluded the majority of business users.

So under the guise of access for all, data lakes emerged. Finally, a place where the organization can place all their data so

everyone can access it! However, even the best laid plans can go awry. What started off as a pristine lake quickly turned into a murky, data mess. Data entered the lake without any organization or context, making the data virtually unusable for the business. The organization solved the issue of too much control, but now they tipped the scale too far in the other direction, with unfettered freedom causing data chaos.

Striking a balance between control and freedom is critical for organizations as they transform their organization to be data-driven. But to do so, they must think differently about data access, data usage, and data reach.

## MIND THE (ENGAGEMENT) GAP

Organizations that have successfully mastered the art of data engagement know a secret. To enable the business and close the data engagement gap, they must first embrace data engagement with governance. But this is not the controlling, locked-down data governance of the past. Rather, it's governance that is transparent, collaborative, and accessible. It enables everyone across your organization to **find**, **understand**, and **trust** the data. And it's the key to (finally) turning the corner to become data-driven.

## FIND THE DATA

Business users want to find the data they need, fast. And they want the technology they use to mimic the apps they use in everyday life. Enable the business by providing an online shopping experience where all users can find—and consume—the information they need to do their jobs. Think of it like Amazon, but for data. Users can browse the data aisles, explore recommendations and ratings,

and collaborate with others to determine which data meets their needs.

## UNDERSTAND WHAT THE DATA MEANS

Finding data is the first step in closing the data engagement gap, but understanding what it means is equally critical. Using a data catalog, users can explore the lineage of the data to understand where it comes from, who has access to it, and how they've used it. They can tap into the knowledge of experts through crowdsourcing. Then, with this understanding in hand, they can assess if the data is fit for purpose.

## TRUST THAT THE DATA IS RIGHT

Finding and understanding data is good, but trusting data is better. Trust is the key to turning the corner on digital transformation. It separates industry leaders from industry giants. And it is the critical element to closing the data engagement gap. When users trust the data, they use the data. They develop practical insights and apply them to guide future strategy. Trust begins not just with secure data, but with data that is governed transparently and flexibly.

Achieving this new reality for data governance isn't simple. But focusing on governance as an enabler for data engagement is critical to your success. Shift your mindset and you'll see how governance—and a governed data catalog—can boost operational efficiency, spark innovation, and drive competitive edge. Then, you'll be ready to change the way your organization thinks about governance to finally overcome the data engagement gap. ■





# How Modern Analytics is Changing Database Requirements

USE CASES FOR modern analytics have expanded beyond BI and reporting to include, for example, recommendation engines and decision support systems. The evolution from descriptive analytics to prescriptive and predictive analytics is changing database requirements.

Database requirements are changing because analytics is becoming a competitive differentiator, with faster time-to-action, and thus faster time-to-insight, underpinning the business goals of modern analytics strategies. To realize faster time-to-insight, data analysts and data scientists must be able to analyze near-real-time data on demand, without being constrained by the schedules of batch processes and ETL jobs, predefined questions, or schemas optimized for specific queries.

## INTRODUCING MARIADB AX

MariaDB AX is an enterprise open source database optimized for modern analytics—fast, powerful and scalable, engineered to run on commodity hardware or cloud infrastructure, and easy to use. It enables data analysts and data scientists to import and analyze data on demand, with the full power of SQL, and without the limitations and constraints of traditional data warehouses.

### Distributed for performance and scalability

It is built on MariaDB ColumnStore, an analytical storage engine. It uses distributed data and parallel query processing to provide scalable, high-performance analytics—data analysts and data scientists don't have to wait hours for a query to complete. Data is partitioned, with different partitions stored on

different nodes, and queries are turned into sets of jobs executed on one or more nodes, with a node executing one or more jobs.

### Columnar for efficiency and flexibility

Data is stored in a columnar format for efficiency and flexibility—data analysts and data scientists are no longer limited to predefined queries. There is no need for indexes, and while common data warehousing schemas (e.g., star/snowflake) can be used, they are not required. Data can be stored in a single table with hundreds of columns, and the schema does not have to be optimized for specific queries—the same data can be queried in different ways and with the same performance.

### APIs for simplified data ingestion

In addition to standard JDBC/ODBC drivers, MariaDB AX includes C++, Java, and Python clients for collecting and importing data on demand, and an Apache Kafka consumer for streaming data ingestion—data analysts and data scientists no longer have to wait for the next scheduled batch process or ETL job to complete. REST services, written in Java by a developer, can be deployed to collect and import user-generated data (e.g., shopping carts) while a machine learning program, written in Python by a data scientist, can persist predictions (e.g., product recommendations). To support analysis of near-real-time data, it can consume messages from an Apache Kafka topic (e.g., clickstream data).

### SQL for ease of use

MariaDB AX supports standard SQL, including JOINS—data analysts

and data scientists do not need to learn a new language or API, or work around the limitations of non-relational databases. However, if necessary, stored procedures and user-defined functions (aggregate and window) can be used to implement custom analytical algorithms.

## MARIADB AX IN THE REAL WORLD

MariaDB AX is used in financial services, healthcare, telecommunications, and manufacturing industries. In financial services, it is used to identify trade patterns and anomalies for fraud detection, store historical data for regulatory compliance, and forecast investments. In healthcare, it is used to find genetic matches, recommend treatments, and predict viral outbreaks. It is used to improve call and network quality in telecommunications, and in manufacturing, it is used to predict machine failures.

Why? MariaDB AX can scale to 100+ terabytes of data and billions of rows (years of data), ingest tens of millions of rows per day (hypergrowth), and meet performance expectations. It is cost-effective, easy to use, and supports standard SQL, making it the enterprise open source alternative to expensive, proprietary databases and data warehouses.

To learn more, please visit: <https://mariadb.com/products/solutions/olap-database-ax>. ■



# Break Free of Data Analytics Barriers

BUSINESS INTELLIGENCE AND ANALYTICS continue to be the lifeblood of any business. However, many enterprises are stuck just trying to keep the status quo. With a proliferation of siloed data marts and piecemeal insights, it can be challenging to tackle new business needs with existing data warehouse architectures.

Now, self-service data exploration and cloud are also considerations. Technology has evolved to where anyone within your organization is capable of exploring data to solve problems they care about. But how do you enable this self-service exploration? And how do you bring cloud into the mix in a secure, integrated way?

## TODAY'S CHALLENGES

Traditional databases and data warehouses are a long-standing part of IT architectures. However, the increase in amount and types of data is testing the limits of these systems—causing frustrations to the business users and IT. Frustrations like:

- **How to tap into more types of data?** New data types such as clickstream or sensor data are challenging for these traditional systems to handle.
- **How to enable self-service analytics?** End-users want faster, more flexible access to data without always having to go through IT.
- **How to take advantage of real-time insights?** Streaming data opens new possibilities. Being able to report and act on data as-it-happens is a strategic advantage for any business, but can be impossible with existing architectures.

## SOUND LIKE YOU?

**Are you only able to analyze a subset of your data at a time?**

Costs and scale limits end up defining what data is collected and kept. It's difficult to incorporate new datasets and or support new types of workloads and reports.

## Do queries seem to “take forever?”

Legacy systems are focused on meeting SLAs for pre-canned dashboards and reports, meaning exploration and ad hoc query performance can often be unpredictable. This results in users finding workarounds.

**Are you maintaining multiple data silos and data copies across your environment?**

Copying data between these systems takes time, limits volume, and leads to operational inefficiencies. It can also lead to data security, governance, and version-control issues.

## A MODERN SOLUTION

Cloudera's modern analytic database is designed to help you tap into the full value of your data. As a high-performance analytic database, it opens up BI and exploratory analytics over more data—using the skills analysts already rely on—to derive instant value. It's a complete solution all built across a shared data lake—ensuring you'll not only better address the business needs of today, but that you can quickly evolve to address the needs of tomorrow. With Cloudera's analytic database solution, you get:

- **High-performance, high concurrency:** Built with cutting-edge technologies, Cloudera's platform provides high-performance analytic SQL—even while supporting high user concurrency—so all analysts across the business have exploratory access to data. And, through integrations with the leading BI tools, your business has immediate access using the tools and skills they already know.
- **Self-service flexibility:** Bring in more data from new sources, actively adjust models based on changing needs, and iteratively design for a variety of use cases. With Cloudera's analytic database, IT can empower the business and make self-service a reality.

- **Unprecedented scale:** Confidently scale your system to tackle any amount of data and support new use cases and users without a major maintenance operation. Cloudera's platform lets you elastically and cost-effectively scale your cloud or on-premises deployments—as and when you need them.
- **Open architecture for SQL and beyond:** Built around open storage and formats, Cloudera's shared data platform means all data is available not just for SQL users, but also for your data scientists and data engineers to work against the same data using their preferred tools, all without any data movement. Easily power end-to-end analytic workflows and enable collaboration across your whole business.
- **On-prem or cloud-native:** Whether you're all on-premises or looking to move to the cloud, Cloudera's solution gives you portability to run the same powerful analytics in any environment. And with native AWS S3 and Microsoft ADLS integration, you can take full advantage of the cloud with Cloudera.

## HOW IT'S USED TODAY

Organizations like NYSE, Experian, Quest Diagnostics, and Cox Automotive optimized their data warehousing environments with Cloudera.

- **NYSE:** Gains real-time insights from 20+ petabytes of data
- **Experian:** Powers customer insight with 50x performance increase
- **Quest Diagnostics:** Delivers insights to better manage diseases
- **Cox Automotive:** On-the-fly dashboards and insights at 50% lower costs ■

CLUDERA

www.cloudera.com