SPONSORED CONTENT

CIO
FROM IDG

CIO Guide

BUILDING a MODERN STRATEGY for ANALYTICS and MACHINE LEARNING SUCCESSS

aws

IDG COMMUNICATIONS, INC.

Introduction

Data and analytics have unleashed a new era of business opportunity. Fueled by the cloud and machine learning (ML), modern analytics deliver increased visibility across the business, helping to drive decisions that are paying big dividends.

Insights culled from real-time and historical data give leadership teams the ability to quickly adjust to unexpected change, improve customer experiences, become more predictive, and streamline processes for operational cost and time savings.

This e-book explores how the growing migration to cloud platforms and services, along with advances in artificial intelligence (AI) and ML, are accelerating analytics initiatives and helping overcome previous barriers to data availability, performance, and scale.

Turning data into business value

Here are a few examples of how organizations are already applying analytics for business benefit and the common good:

- **Supply chains** are gaining the ability to quickly pivot when market conditions or customer demands fluctuate. For example, if there's a disruption in the partner network, analytics powered by ML alerts network members so they can find an alternative supplier or shipper to keep delivery schedules on track.
- **Manufacturers** can predictively detect failing equipment and take corrective action early enough to avoid shop floor downtime. These capabilities boost productivity and business continuity, enabling manufacturers to improve production and delivery and reduce costs.
- **Telecommunications** companies and other data-intensive businesses are monetizing the data they generate and collect by packaging it and selling it to target vertical markets. In this way, they're creating new revenue streams for their businesses.
- **Marketing departments** are using analytics to learn about the likes, behaviors, and experiences of consumers based on their digital footprints. Equipped with this information, businesses can accurately target offers to micro-segmented groups and grow their customer bases. They can also learn customers' preferred ways of communicating and conducting transactions so that their businesses can interact with them accordingly, improving satisfaction and building loyalty.
- **The scientific community** is analyzing and sharing huge volumes of environmental data gathered by satellites, space stations, and robotic vehicles to inform us about actions we can take to improve human habitability on Earth and enable it on Mars.

Contents

Inhibitors to success

Monetizing data offers a potentially significant payback in business value. However, many organizations have struggled to harness their collective data to get the accurate and meaningful results needed to achieve their goals. Several issues stand in their way:

- **Data growth.** The amount of data created in the world continues to escalate: IDC predicts the amount of digital data created over the next five years—rising from 64 zettabytes in 2020 to 180 zettabytes in 2025—will be greater than twice the amount of all the data that's been created since the advent of digital storage. Rapidly rising volumes and types of data are increasingly difficult to manage using older on-premises infrastructure and manual processes.

"Few companies truly understand all the data that they have—whether in the cloud or stored locally," says Michael Gabriel, a Partner at Fortium Partners who has held global CIO positions at HBO and the National Basketball Association. "So knowing the data you have available or could have available, and how it needs to be utilized to support analytics, is problematic."

- **Outdated data infrastructure.** Three-tier, on-premises data infrastructure lacks the scale and performance capabilities needed to manage growing data—and requires constant configuration, management, and capacity planning. Aging infrastructure also makes it difficult to support ML and other advanced analytics capabilities that are critical to delivering actionable insights across the organization. Organizations need data systems that can scale as needs change and data volumes grow.

"We see a lot of customers who have an on-premises analytics system and have hit the limits of what they can do," says Rahul Pathak, Vice President for Analytics at AWS. "Their data warehouse might not be able to meet the growth and scale of data coming in. So they struggle with scale, performance, and the operational costs of maintaining it in their own data centers."

- **Data silos.** Corporate data tends to be walled off along departmental lines, accessible only to select groups of users. Data is also stored in multiple places, including your data warehouses, data lakes, and databases. Analytics projects relying on data that's incomplete because of these limitations often fail to produce the insights and the returns their designers anticipate. To gain new value from their data, organizations need the ability to break down silos so they can combine and analyze all relevant data regardless of where it lives.

"Without sharing, there can be redundancy of data, redundancy of cost, and an inability to see the full picture," says Gabriel. "This creates a disadvantage to the company as well as to the individual department."

- **Data gravity.** As data continues to grow, it becomes harder to move around. This data "gravity" limits the comprehensiveness of analytics run against a particular domain and can degrade reliability of the insights and decision-making based on them.

"To make decisions with speed and agility, customers need to acknowledge data gravity by easily moving the data they need between data stores in a secure and governed way," says Pathak.

Inconsistent data governance. Some organizations have pieced together governance policies over time, resulting in stove-piped access control policies that can lead to stale data, security vulnerabilities, and regulatory noncompliance. Other organizations have overcompensated with restrictive governance policies that impede productivity.

"A common mistake we've seen is that organizations end up creating overly rigid governance scenarios, where the central teams become a bottleneck to those trying to work with data to improve the business," says Pathak. "The key to good governance is figuring out how to define access, then getting out of the way. By that I mean creating exception processes, rather than taking an approach that anytime you need data, you have to ask someone in a central organization for it."

What's the right mix of analytics?

There are many kinds of data analytics, and your business will likely want to apply a mix of them to achieve various outcomes. Here are the basic analytics types and their most typical use cases.

- **Real-time analytics** turns data into insights as it's being collected. This type of analytics is used for highly time-sensitive applications such as online trading or vehicle control systems. Real-time analytics can predict when equipment is about to fail, help a self-driving vehicle avoid an accident, and detect credit card fraud before a transaction is complete. There are two main types of real-time analytics: On-demand analytics waits for users or applications to send a query before delivering a result, while streaming analytics continuously delivers alerts or results.
- **Log analytics**, also called operational analytics, is the assessment of event data that might be captured from a computer, network, application, operating system, or another IT component. An organization can use log analytics to uncover patterns in user behaviors, identify trouble spots, audit security activities, manage regulatory compliance, and plan for capacity or other IT infrastructure changes.
- **Big data analytics** involves running advanced analytics against very large, diverse data sets that might include structured, semi-structured, and unstructured data from different sources and in different sizes. The data might originate from sensors, computing or communications devices, video/audio, networks, log files, transactional applications, web content, and social media.
- **Data warehousing analytics** performs queries against large amounts of historical data gathered from many sources such as application log files and transaction applications. This type of analytics lets users run queries based on subject and assesses changes over time.
- **Machine learning** analyzes and interprets patterns in the data to enable learning and decision-making without human interaction. It can help create entirely new revenue opportunities, enable better and faster decisions, and improve operational efficiencies.

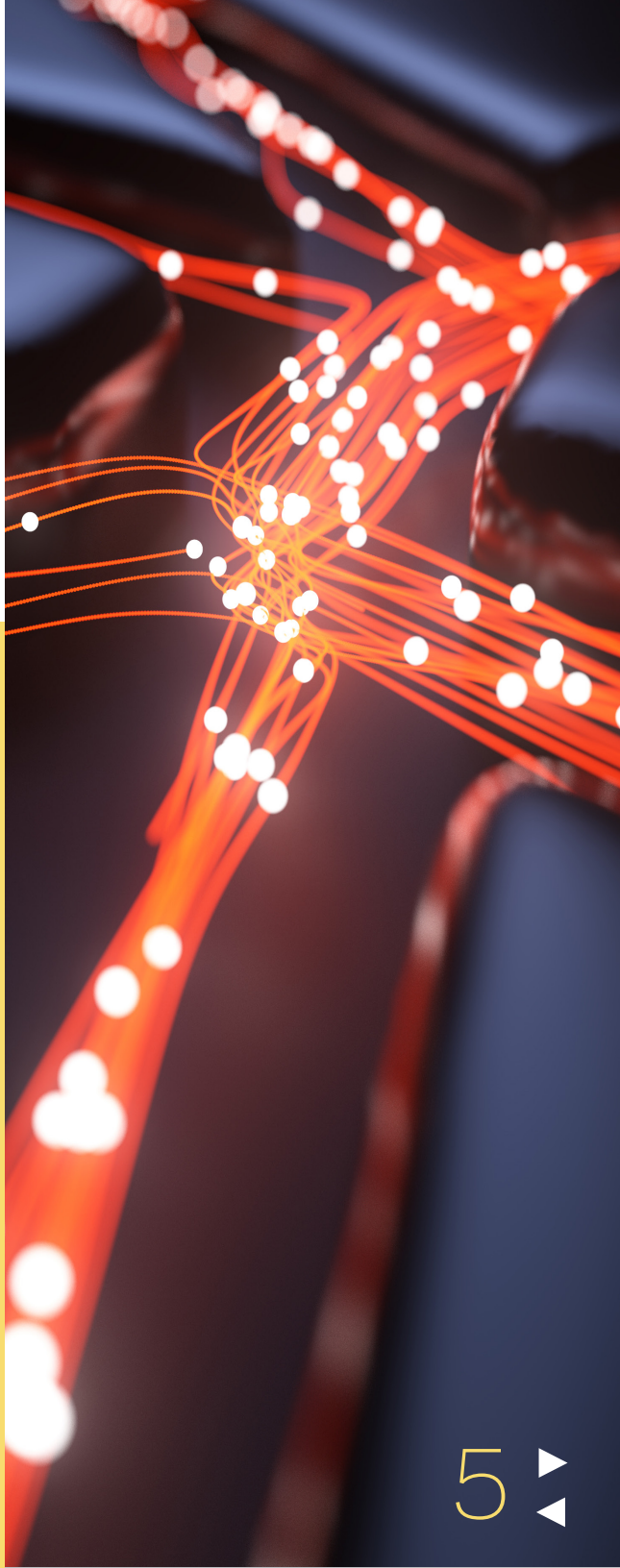
Overcoming the obstacles

Succeeding with analytics requires clear business alignment on expected outcomes and an IT infrastructure that supports those goals, rather than hindering them.

"It's difficult to understand the infrastructure you need to have in place until you know what you're trying to achieve and where your current infrastructure needs improvement," says Gabriel.

An analytics modernization strategy has two fundamental components: 1) data must be integrated and universally accessible to authorized users; and 2) the infrastructure must support a centralized approach to data management, security, and governance.

Centrally managed data and aggregation allow organizations to run analytics against the full breadth of data required for smart decision-making and optimal business results.



Machine learning: Capabilities for all

AI services

Increasingly, AI is being infused into applications and business processes. Ready-made intelligence through AI services addresses common use cases such as personalized recommendations, contact center intelligence, document processing, intelligent search, business metrics analysis, and more.

ML infrastructure and frameworks

Expert practitioners want the most powerful infrastructure to run the ML frameworks of their choice like TensorFlow, MXNet, and PyTorch.

ML services

Data scientists, ML developers, and, increasingly, data analysts are looking for a tool that makes it easier and faster to build and implement machine learning. Like an integrated development environment for software, they need a service that covers every step of ML development, including labeling, data preparation, feature engineering, statistical bias detection, auto-ML, training, tuning, hosting, explainability, monitoring, and workflows.

How to implement a modern analytics approach

To harness the insights and innovation that a modern analytics approach enables, consider these four foundational steps.

1. Aggregate data using data lakes

An important step in unifying siloed data is to aggregate it into one or more data lakes. The greater the amount of data there is to store, manage, and analyze, the more beneficial it is to create a data lake on a public cloud foundation with infinitely scalable processing resources.

Once in the data lake, unified data sets can be consumed again and again, reducing costs and maximizing data value. With this holistic approach, analytics programs account for the whole data picture as they compile and return insights for optimal results.

NuData, a Mastercard company, is a testament to the benefits of aggregated, holistic data analytics. It has built fraud detection services that run specialized analytics against very large datasets it stores in Amazon S3 data lakes, powered by ML. NuData runs 26 microservices, each based on a customer use case for detecting a particular type of fraud. While each microservice has its own data lake, it can share data with others. These capabilities help the company correlate petabytes of data each day to identify malicious or erroneous login attempts that might compromise user accounts, explains Justine Fox, NuData's Director of Software Engineering.

2. Enable unified data governance and access

Every person in your organization who needs access to data should have it—wherever and whenever they need it. This requires secure, seamless access to data lakes and purpose-built data stores and services and being able to access your data no matter where it lives. Data also needs to move seamlessly among applications, systems, and services. For instance, clickstream data from web applications can be collected directly in a data lake, and a portion of that data can be moved to a data warehouse for daily reporting. Or you may want to move sales data from a warehouse into a data lake where it can be stored and analyzed using ML.

In tandem with data integration comes the need for centralized data governance and security. Unified data governance involves setting consistent policies across data, services, and applications. It's important to strike the right balance between data security and worker productivity. Consider these steps:

- **Take an enterprise-wide inventory** of data resources, who's responsible for them, and who requires access. Create a map of which resources should be made accessible to whom. "One of the benefits of doing this in the cloud is that you can track all these interactions," says Pathak.
- **Adopt a zero-trust security approach** to access control, which involves limiting access to only those users and applications that require it. Unifying data across silos while tightly controlling access for security reasons requires a delicate data management balance.

Unified data governance involves setting consistent policies across data, services, and applications. It's important to strike the right balance between data security and worker productivity.

2. continued

- **Deploy a data catalog** or other centralized management mechanism that automatically discovers, tags, and catalogs data so you can manage and audit policies all in one place. This enables you to provide fine-grained access to data to the right user at the right time, and effectively meet regulatory governance and compliance requirements.

- **Work with your cloud service provider** (CSP) to help you manage compliance across different geographies. Specifically, make sure your CSP has a way to control where data physically resides, since the cloud uses virtual machines that could theoretically be located anywhere. Creating and maintaining a compliance database can help; mapping out digital compliance standards by country creates a clear, active structure for compliance.

3. Deploy purpose-built data and analytics services for the best price/performance

The exploding volume of data points to be analyzed and correlated is driving many enterprises to migrate more of their data and analytics infrastructure to the cloud. A cloud foundation has the infinitely scalable compute and storage resources required to analyze mass quantities of data, deliver meaningful, actionable insights, and provide the rich training data needed for accurate ML-based automation.

Organizations are using purpose-built databases, analytics, and ML services to better solve analytics use cases by storing or processing data in a way that is optimized for each particular use case. For example, a document database would be apt for a mobile application that requires great scalability and performance, while a graph database could help developers explore hidden connections between tightly related records. This ensures organizations can use best-in-class functionality for all workloads, meaning there is no compromise on performance, scale, or cost. With purpose-built data services, organizations get the best price/performance for all of their applications and analytics needs.

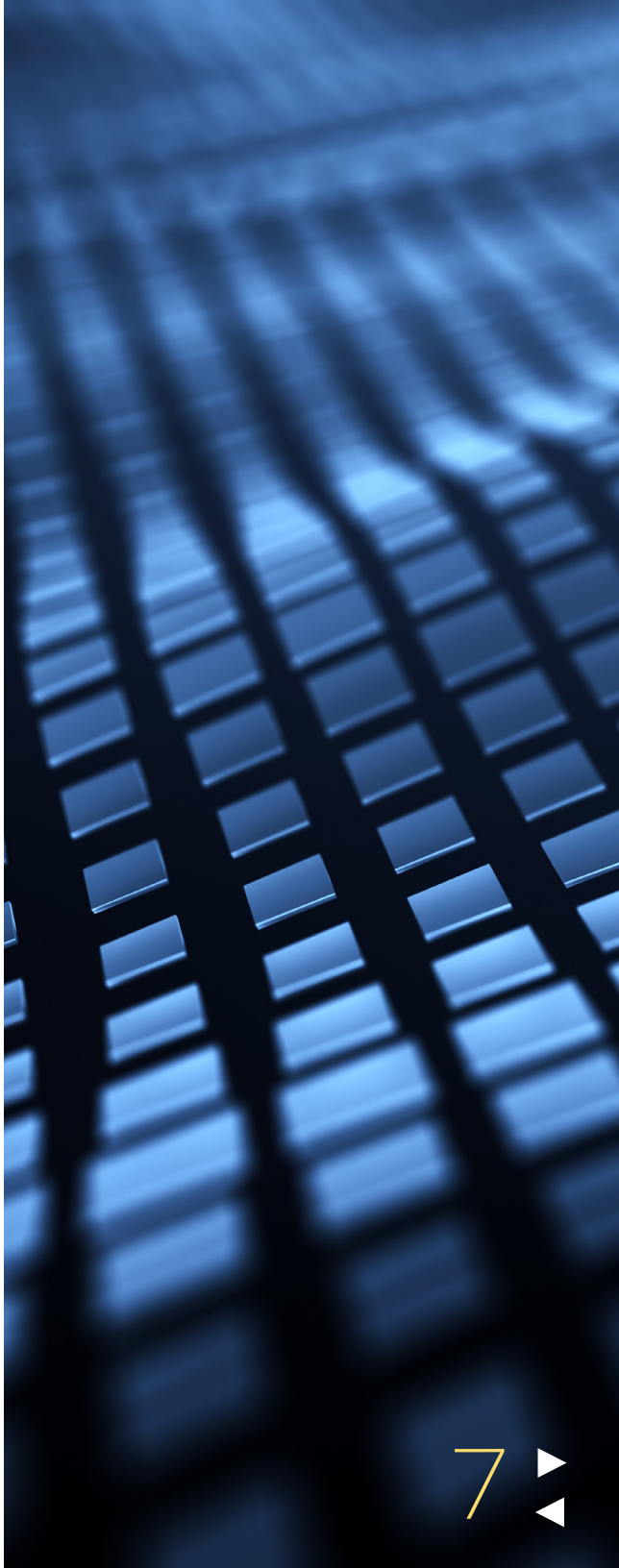
4. Use ML and AI to solve business challenges

Whether organizations want to enhance their customer experience, improve productivity and optimize business processes, or speed up and scale innovation, they can access ML and AI services to meet their business needs.

AI and ML technologies enable organizations to do more with data sets that were previously almost unusable. For example, unstructured data found in content such as PDFs, audio, video, earnings transcripts, and reports can now be run through ML processes for fresh insights.

"Instead of having analysts read hundreds of thousands of documents, we can start to have machine learning go through those documents, create structured data, and build applications on top of it," says Michael O'Rourke, Senior Vice President and Head of AI/Technology, Investment Intelligence at Nasdaq, which has embraced the cloud, data, and AI/ML as foundational elements for innovation and growth. AI and ML play an increasingly important role not just for Nasdaq's data business, but across the entire organization.

"In the financial industry, the opportunity for AI is enormous," says O'Rourke. "Within Nasdaq, every single business line is looking at how they can utilize machine learning and AI to make better products, improve productivity, and create new solutions."



Put your data to work

In today's digital environment, data can be overwhelming to organizations large and small. A modern analytics approach will help make data accessible to your entire organization, opening up new ways to put your data to work and reinvent your business.

Sponsor viewpoint

With a modern approach to data strategy based on a portfolio of purpose-built services from AWS, you can help your organization plan for the future, lead in your industry, and come out on the other side with limitless possibilities for reinvention. Modernizing your analytics will give you the ability to rapidly gain insights from all of your data and deliver access to all of your users.

AWS provides the broadest and deepest portfolio of purpose-built analytics services to realize a modern data and analytics strategy. Our portfolio includes the most scalable data lakes, the broadest set of purpose-built analytics and ML services, seamless data access, and unified governance—all delivered with the best performance at the lowest cost. Tens of thousands of organizations have already built data lakes using AWS, and we make it easier than ever to get started.

▷ **Learn more**