IDC

REGION FOCUS: NORTH AMERICA

Generative AI: Revolutionizing Healthcare and Life Sciences

Getting from here to there requires
a unified data, analytics, and AI strategy.

Nimita Limaye
Research Vice President, Life Sciences
R&D Strategy & Technology, IDC

Michael Townsend
Research Director, Life Science
Business System Strategies, IDC

Mutaz Shegewi
Research Director, Provider IT
Transformation Strategies, IDC

Table of Contents

CLICK BELOW TO NAVIGATE TO EACH SECTION IN THIS DOCUMENT.

Executive Summary

Healthcare and life sciences organizations have struggled with integrating data from disparate internal and external sources. The amount of data generated daily only continues to grow in volume, velocity, and variety. The future of healthcare relies on a connected ecosystem. Organizations that set themselves apart will possess the ability to ingest data from any source, share that data with internal and external stakeholders, and leverage AI-powered, real-time insights that hold the promise to transform healthcare and reshape R&D in the life sciences industry.

The data lakehouse is the path forward.
This IDC InfoBrief discusses the value of the data lakehouse, addresses the challenges associated with data warehouses and data lakes alone, touches on use cases healthcare and life sciences leaders are investing in today, and provides essential guidance to drive adoption.

Key takeaways:

- ✅ Descriptive analytics is no longer sufficient. Executives across healthcare and life sciences are investing in real-time analytics and artificial intelligence to drive more effective decision-making.

- ✅ The path to get from here to there requires breaking down data silos and unifying data, analytics, and AI on a single platform.

- ✅ Ideal data platform solutions are cloud-native and cloud-neutral.

- ✅ Data governance strategies should be robust and implement data security and data privacy controls. As these are highly regulated industries, companies must address data sovereignty issues and ensure data quality and integrity.

- ✅ Since a data lakehouse is based on open standards (i.e., Delta), it has the potential to prevent vendor lock-in by connecting to any data source.

Healthcare Has Not Tapped the Full Potential of All Data

Connecting data across the healthcare and life sciences ecosystem fuels data interoperability.

The healthcare and life sciences industries are struggling to deal with the exponential growth of structured and unstructured data. This explosion is fueled by the growth of precision medicine strategies, multiomics, imaging data, and the increased focus on real-world data, such as electronic health records, claims data, multimodal digital pathology data, social determinants of health (SDOH) data, Internet of Medical Things (IoMT) data, and data from mobile devices. Data is the bridge that connects the healthcare and life sciences ecosystem.

As these industries deal with patients' lives, shifting from merely measuring healthcare to driving real-time insights is spanning everything from discovering novel therapeutics faster, nudging patients with personalized recommendations, and determining the next best action for sales and marketing, to name a few use cases.

38% and **28%**

of healthcare organizations   **of life science organizations**
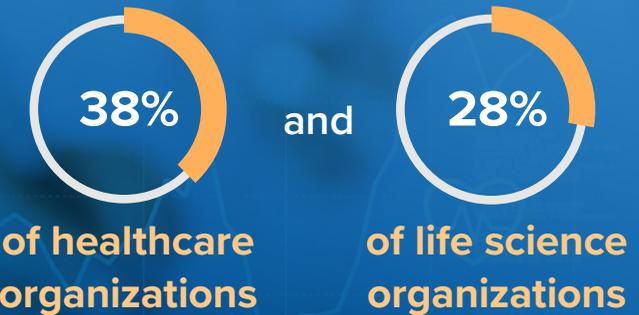
report that data growing beyond the capacity of existing systems is an important trigger leading them to use cloud services.

n = 200 for healthcare, n = 103 for life sciences
Source: IDC's *Worldwide Industry CloudPath Survey*, April 2022

What Does a Data-Driven Organization Look Like?

Addresses concerns regarding data privacy and security

Empowers data analysts, engineers, and scientists while ensuring control and compliance

Provides real-time AI-/ML-enabled insights, accompanied with robust data governance

Tackles data interoperability challenges and unifies data sources

Establishes process and cost efficiencies

Facilitates data sharing, collaboration, and innovation

68.5% of healthcare organizations thought that advanced data management capabilities were an important or very important criterion when evaluating cloud services.

Looking Beyond the Data Warehouse and Data Lake

Data comes in all sizes and shapes: structured, semistructured, and unstructured. Integration and access pose multiple challenges.

Conventional models extract data from operational repositories and other sources and store it in a raw data lake. Critical data subsets, including structured data, are then transferred to a data warehouse to generate business insights. This involves multiple rounds of ETL and duplicated data, increasing cost and complexity while delaying insights.
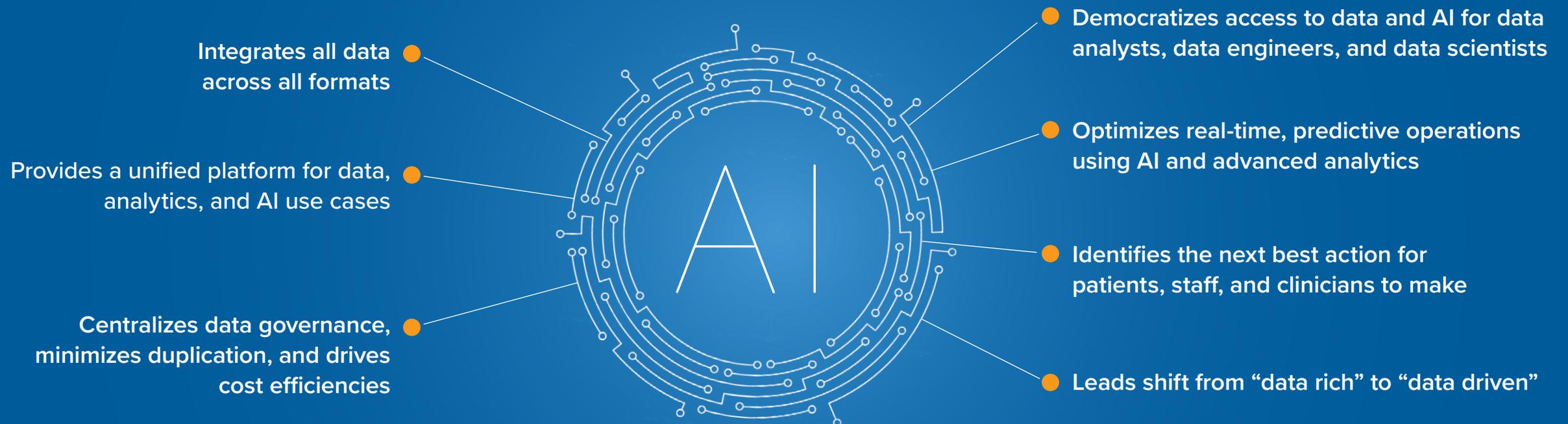
Additional challenges include:

- ✓ Consistency between data lake and data warehouse architecture
- ✓ Ensuring that data sets in both are synchronized and not stale
- ✓ The costs associated with maintaining multiple systems and maintaining data copies in the data lake and the data warehouse
- ✓ Advanced machine learning frameworks, such as PyTorch and TensorFlow, that aren't fully compatible with data warehouses
- ✓ Data governance and data sharing issues

By 2024, the proliferation of data will result in **60% of healthcare organizations' IT infrastructure** being built on a data platform that will use AI to improve process automation and decision making.

A Data Lakehouse Offers the Best of Both Worlds

This architectural strategy integrates structured and unstructured data and combines data, analytics, and AI on one cloud platform.

Integrates all data across all formats

Provides a unified platform for data, analytics, and AI use cases

Centralizes data governance, minimizes duplication, and drives cost efficiencies

AI

Democratizes access to data and AI for data analysts, data engineers, and data scientists

Optimizes real-time, predictive operations using AI and advanced analytics

Identifies the next best action for patients, staff, and clinicians to make

Leads shift from "data rich" to "data driven"

Empowering Patients Through Modern Platforms

By modernizing healthcare platforms and collaborating around the patient, data gets unlocked and patients are empowered to access and exchange medical information more easily, which enhances outcomes.

Healthcare providers must upgrade existing systems and technologies to keep up with the digital era, including AI-powered diagnosis and precision medicine tools. Platform modernization is a crucial step towards revolutionizing healthcare and ensuring that it remains efficient, effective, and accessible to all.

The challenges include:

- ✓ Integrating outdated legacy system technologies with newer platforms
- ✓ Taking robust-enough measures to align patient data protection and security concerns and stay compliant with regulatory requirements
- ✓ Affording platform modernization, including ongoing maintenance and support
- ✓ Ensuring user-friendliness and enough benefits to alleviate any resistance from healthcare providers and patients to adopting new technologies
- ✓ Training healthcare providers and staff on new technologies and processes to ensure effective and efficient use of new platforms

31% of healthcare providers have modern analytics data platforms that enable sharing and collaboration across the enterprise.

Transforming Patient Outcomes With Responsible AI and Large Language Models

Modernizing data platforms and democratizing data via sharing expands opportunities for AI and large language models (LLMs) to improve patient care.

With proper guardrails, **AI and LLMs can be fueled by modern data platforms** and generate powerful gains and insights that help improve patient outcomes.

Breaking down data **silos and facilitating data exchange create a comprehensive view of patient information**, allowing AI and LLMs to shine when it comes to more accurate diagnoses and personalized treatment plans.
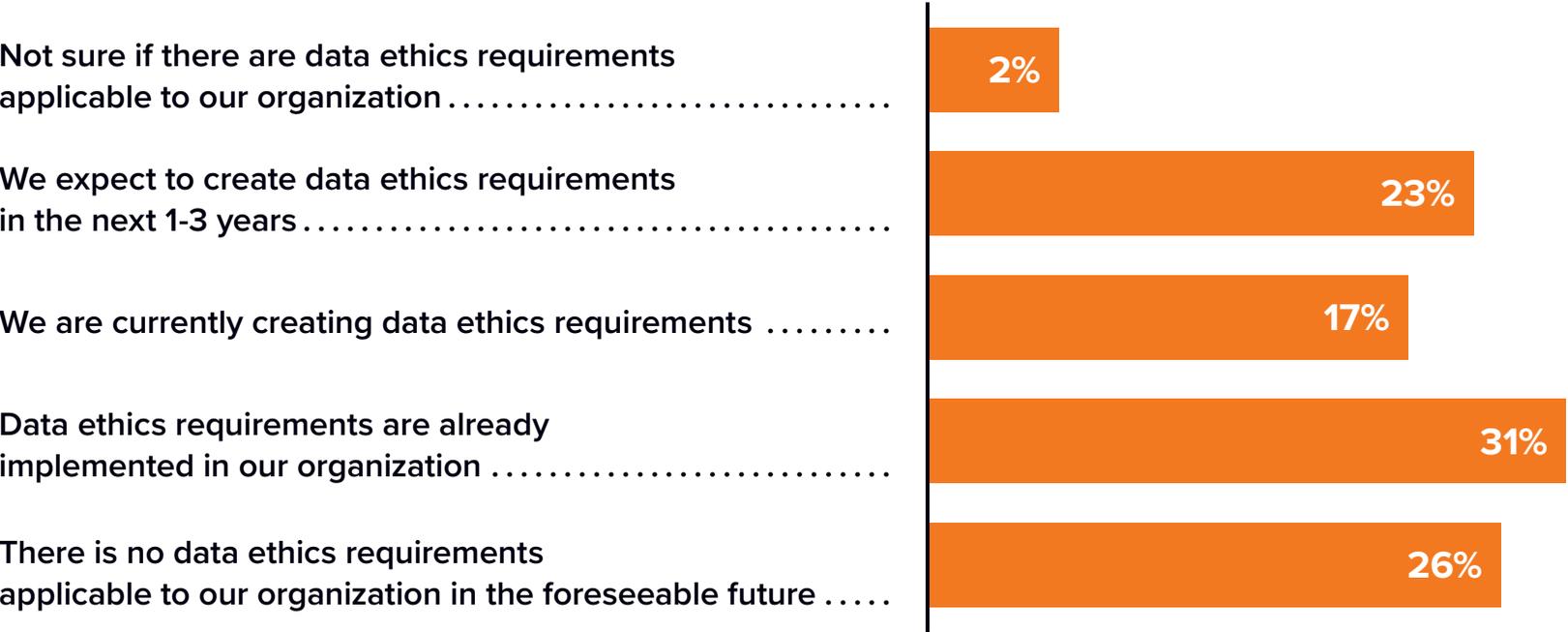
Access to diverse and rich data sets helps **AI identify patterns and leverage LLMs** in ways like content summarization, biomedical Q&A, and information retrieval.

By leveraging modern data platforms and promoting data sharing, **healthcare provider and life sciences organizations can collaborate more effectively**, sharing insights and best practices to optimize care delivery and drive positive patient outcomes.

Transforming Patient Outcomes With Responsible AI and Large Language Models (continued)

Which of the statements best describes your organization's approach to data ethics (responsible use of data) for artificial intelligence/machine learning (AI/ML)?

Not sure if there are data ethics requirements applicable to our organization **2%**

We expect to create data ethics requirements in the next 1-3 years **23%**

We are currently creating data ethics requirements **17%**

Data ethics requirements are already implemented in our organization **31%**

There is no data ethics requirements applicable to our organization in the foreseeable future **26%**
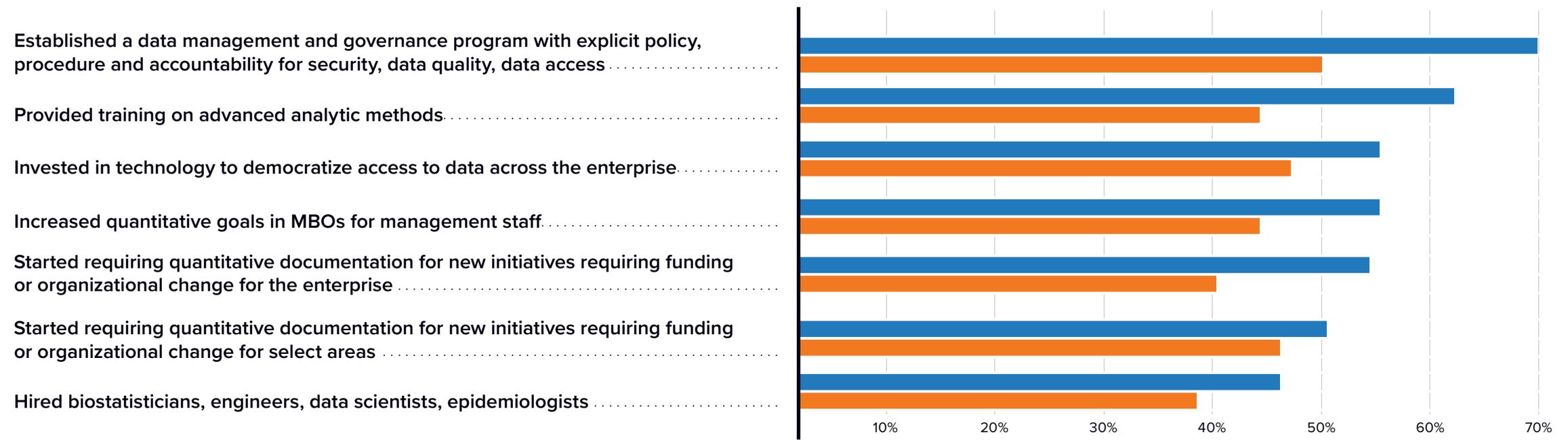
By the end of 2025

65%

of healthcare organizations will have data governance frameworks in place, prioritizing the ethical and explainable use of AI for predictive, preventive, and personalized care.

Source: IDC's *Future Enterprise Resiliency and Spending Survey, Wave 2,* March 2023

What Does a Data-Driven Culture Look Like?

What has your organization done in the past 12 months to become more data driven?

■ Data driven   ■ Low data driven

Established a data management and governance program with explicit policy, procedure and accountability for security, data quality, data access . . . .

Provided training on advanced analytic methods . . . .

Invested in technology to democratize access to data across the enterprise . . . .

Increased quantitative goals in MBOs for management staff . . . .

Started requiring quantitative documentation for new initiatives requiring funding or organizational change for the enterprise . . . .

Started requiring quantitative documentation for new initiatives requiring funding or organizational change for select areas . . . .

Hired biostatisticians, engineers, data scientists, epidemiologists . . . .

n = 101 for data driven, n = 102 for low data driven; Source: IDC's *Healthcare Business Intelligence Payer and Provider Survey*, August 2021
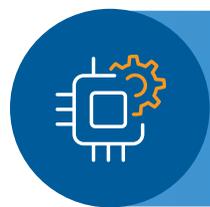
A data-driven culture has an established organizational structure for data management and governance with **explicit accountability and authority for data quality and access**. The data management team consists of individuals with **expertise in data forensics, data science, and advanced analytics**. Ideally, some team members also have **clinical and epidemiological skills.**

Benefits of Unifying Data, Analytics, and AI on a Single Platform

60% of life sciences companies considered cloud and health data platforms as very important for driving their digital transformation initiatives according to IDC's *Life Science DX Survey* **(January 2022).**

Intelligent health data management enables life sciences organizations to generate real-time, data-driven insights to accelerate drug discovery, analyze the safety and efficacy of a drug, and power precision medicine strategies.

| Accelerates compute to deliver real-time insights | Drives transparency; ensures traceability | Improves customer experience | Enables collaboration and co-innovation | Decreases capex and opex | Reduces IT burden; ensures compliance | Simplifies infrastructure; streamlines workflow |

Essential Guidance

In the healthcare and life sciences industries, critical decisions need to be made in real time. One cannot afford delayed insights. In a world where data is exploding, cloud-based platforms that can seamlessly integrate multimodal data from disparate sources and leverage AI and large LLMs models to generate real-time insights are the need of the hour.

- ✓ **Drive data interoperability** across all data formats. **Democratize access** to data and insights.

- ✓ **Focus responsibly** on the future of AI and LLMs for healthcare and life sciences.

- ✓ Don't focus only on integrating the data. **Unify data, analytics, and AI use cases** on one platform.

- ✓ **Develop a data-driven culture.** Innovation across the healthcare and life sciences industries is being led by data-driven organizations.

- ✓ Look at the big picture; **keep the patient 360-degree view in focus.**

- ✓ **Create a roadmap** of how your organization will use data to achieve its goals. Crawl, walk, and then run.

- ✓ **Establish strong data governance models,** keeping the end users in mind: the data analysts, engineers, and scientists.

About the IDC Analysts

Nimita Limaye
Research Vice President,
Life Sciences R&D Strategy
& Technology

Dr. Nimita Limaye is a research vice president with IDC Health Insights and provides research-based advisory and consulting services, as well as market analysis on R&D strategy and technology in the life sciences industry. She addresses aspects such as the role of digital transformation in discovery research, e-clinical ecosystems, the role of natural language processing, artificial intelligence, machine learning, deep learning, and robotic process automation in transforming drug development, precision medicine, pharma R&D execution, and strategic outsourcing models.

More about Nimita Limaye

Michael Townsend
Research Director, Life Science
Business System Strategies, IDC

Michael Townsend is a research manager for IDC Health Insights responsible for commercial life sciences and life sciences commercial strategies. He provides research-based advisory and consulting services, as well as market analysis on key topics within the commercial life sciences industry. Michael's research expertise includes sales and marketing, supply chain, manufacturing systems, and emerging technologies and market trends.

More about Michael Townsend

Mutaz Shegewi
Research Director, Worldwide Healthcare
Provider Digital Strategies, IDC

Mutaz Shegewi leads the provider research practice at IDC Health Insights, covering topics of most relevance to healthcare organizations looking to digitally transform and become more digitally native. Mutaz advises the executive, clinical, and technical leadership of the world's foremost health information technology suppliers and buyers by producing data-driven research and thought leadership to navigate strategic challenges and transform complexity in decision-making to decrease costs, enhance quality, optimize access, improve patient safety, and champion patient experience.

More about Mutaz Shegewi

Message from the Sponsor

databricks

Databricks is the Data and AI company founded by the original creators of Apache Spark™, Delta Lake and MLflow.

Databricks is on a mission to help data teams solve the world's toughest problems, unifying their data, analytics and AI on a single platform.

To learn more about Databricks, visit our website:

www.databricks.com

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. This IDC material is licensed for external use and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.