



By Alan Drummer | 4 minute read

## Extracting insights from complex, unstructured big data

According to projections from IDC, 80% of worldwide data will be unstructured by 2025.<sup>1</sup> Even though most enterprises already use [data warehouses](#) to analyze structured data, more are now turning to data lakes to leverage unstructured and semi-structured data in its native format. This includes the growing volume of streaming audio and video, social media, and clickstream, sensor and log data.

[Data lakes](#) bring organizations new opportunities by providing access to a greater variety and volume of data that can fuel more accurate analytic predictions and models. They also deliver rich, actionable insights by enabling organizations to find interesting relationships, trends, patterns and anomalies that wouldn't be visible otherwise.

A [Ventana research study](#) showed that among enterprises with a data lake:

- 41% reported gains in competitive advantage
- 37% noted lowered costs
- 35% enjoyed improved customer experiences
- 33% believed it helped them respond better to opportunities and threats
- 8% revealed an increase in sales

## Harness more data for better, data-driven decisions

Organizations in almost every industry are benefiting from data lakes. Let's look at three industries where value has grown especially fast.

### Financial services

Financial firms are using the ability to organize and accommodate unstructured data to include location, IoT, sensor, clickstream and social media data in their analytics. As a result, they're able to deliver personalized insurance offerings, fight fraud more accurately, and gain 360-degree customer views. Read the report and watch the webinar to learn how they're getting results like these:

- 300% higher conversion rate
- 30% fewer fraud incidents
- Nearly USD 4 million reduction in expenditure
- 90% quicker time-to-value for big data analytics

Read TechTarget's assessment of these financial services accomplishments in the report: [It's a new era in advanced analytics and AI](#).

[Watch the webinar](#) on how global bank ING accelerated time to market for new products and shortened the selling process from years to months.

### Healthcare

Healthcare firms depend on unstructured data such as doctors' notes, X-rays, CT scans and research articles that empower front-line caregivers with real-time data and predictions. Results include:

- 25% possible annual savings<sup>2</sup>
- 31% reduction in 30-day readmission for certain patients<sup>3</sup>
- USD 500,000 savings per year<sup>4</sup>

Read TechTarget's [Welcome to the healthcare revolution](#) to learn more.



### Communication service providers

Communication service providers are analyzing video feeds, clickstreams and third-party data to predict and prevent churn, optimize networks, and detect fraud in real time. Results include:

- 350% more fraud incidents discovered<sup>5</sup>
- Up to 20% revenue loss reduction<sup>6</sup>
- 5% – 10% reduction in customer churn<sup>7</sup>

Read TechTarget's [Data Analytics and AI are empowering transformation](#) report to learn more about these advancements.



## What elements make a data lake successful?

Governance is essential for a data lake to succeed. Without it, a data lake can quickly turn into an unmanageable data swamp in which users can't find, trust, or use the data they need.

A [governed data lake](#) contains clean, trusted data from structured and unstructured sources that can easily be found, accessed, managed and protected. It enables self-service access to help users find relevant information through simple search interfaces. Read the ebook "[Governed data lakes for business insights](#)" to explore the key building blocks of effectively delivering trusted data.

An **enterprise data catalog** is the foundation for data lake governance. A catalog organizes a data lake by automating data discovery, metadata generation, and the building of machine-learning-extracted business glossaries. It can also perform automated scanning and risk assessments of unstructured data, as well as track data lineage. Learn more about these capabilities when you read the ebook "[Build a better data lake.](#)"

**AI-led automated data integration** can efficiently cleanse and deliver trusted data anywhere, at any scale and complexity, on and across multicloud and hybrid cloud environments.

In-flight data quality and active metadata and policy enforcements help ensure trusted delivery to data lakes. [Read how to save on data movement and storage costs and boost business productivity using IBM DataStage.](#)

**Having a single source** for hardware, software, services and multivendor solutions can make it easier to build, manage, govern and secure the data lake. IBM provides that source. Learn more about [the IBM and Cloudera partnership](#), with an ecosystem of offerings designed for faster analytics results at scale. See how

freedom from vendor lock-in, enhanced self-service and optimized integration are some of the drivers that "provide unprecedented flexibility, choice and value for clients" in the race to implement AI insights. Read the [analyst report on total value of ownership](#).

## Start leveraging semi-structured and unstructured data today

When properly governed with an enterprise data catalog, and unified with data virtualization so data can be queried from a single source, a data lake can transform the fast-growing volume of new data types from a burden into a benefit that helps fuel new and actionable insights.

Over the next few years, as semi-structured and unstructured data become as much as 80% of the world's information, companies that can include these types of data as they look for relationships, trends, patterns, and anomalies will have a growing advantage over those who can't. To get started, [read this ebook on how to build a better data lake.](#)

© Copyright IBM Corporation 2020. IBM, the IBM logo and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

- [1] Timothy King, 80 Percent of Your Data Will Be Unstructured in Five Years, Data Management Solutions Review, March 28, 2019. Accessed November 24, 2020.
- [2] Sabyasachi Dash et. al., Big data in healthcare: management, analysis and future prospects. SpringOpen, June 19, 2019. Accessed November 19, 2020.
- [3] Marina Turea, Ultimate Guide to Big Data in Healthcare, Healthcare Weekly, September 26, 2020. Accessed November 19, 2020.
- [4] Ibid.
- [5] Communication Service Providers: Data Analytics and AI Are Empower Transformation, TechTarget Custom Media. Accessed November 24, 2020.
- [6] Cloudera, Reducing Revenue Loss from Fraud by up to 20 Percent, Cloudera.com. Accessed November 19, 2020.
- [7] Ibid.