# CITO RESEARCH

# 5 Reasons Why You Must Embrace Hybrid Data

actian™

# CONTENTS

## Reason 1:
## It's already here: Your data is hybrid data

The data landscape grows bigger, more complex and more diverse every day, posing both challenges and opportunities for your organization. Imagine fire hoses of different kinds of data pouring in from multiple sources.

The past 20 years of computing have chiefly focused on trying to standardize data. Technologies like relational databases, traditional data warehouses and master data management touted delivering a single version of the truth. Unfortunately, real world data has a dynamic life of its own, constantly evolving and flowing both within and around organizations. Forcing data into rigid systems of record has often resulted in costly and complex systems that delivered inflexible results of limited value for business leaders at all levels.

> Forcing data into rigid systems of record has often resulted in costly and complex systems.

Today's data landscape feels complex because it is, and standardizing it is problematic. Your data, whether you know it or not, is hybrid data. Hybrid data has many dimensions (the table on the following page captures at least a dozen, and it isn't exhaustive). These dimensions begin to mirror the complexity of the data we see and work with every day and the challenges you face with your data.

Like hybrid compute, hybrid data exists on-premise and in the cloud. Hybrid data also exists at the edge, in every smart device on the planet. Beyond value, source, accuracy, and location, hybrid data includes many other dimensions. It is structured, unstructured, and semi-structured. Hybrid data varies in its relationship to your organization. Hybrid data sometimes arrives raw in real time while other times it is created in well-defined batch modes. Hybrid data can take many forms, sometimes cleansed, prepared, and analyzed from established business systems. Other times it is formed by external sources such as business or technology partners, subscription data services, Twitter, Instagram, your own website, and open data sources. Hybrid data may come in transaction by transaction, in large batches, and in constantly flowing streams. In short, hybrid data encompasses every facet of all the data available to you today. Leveraging the power of all of that data can be game changing for organizations large and small.

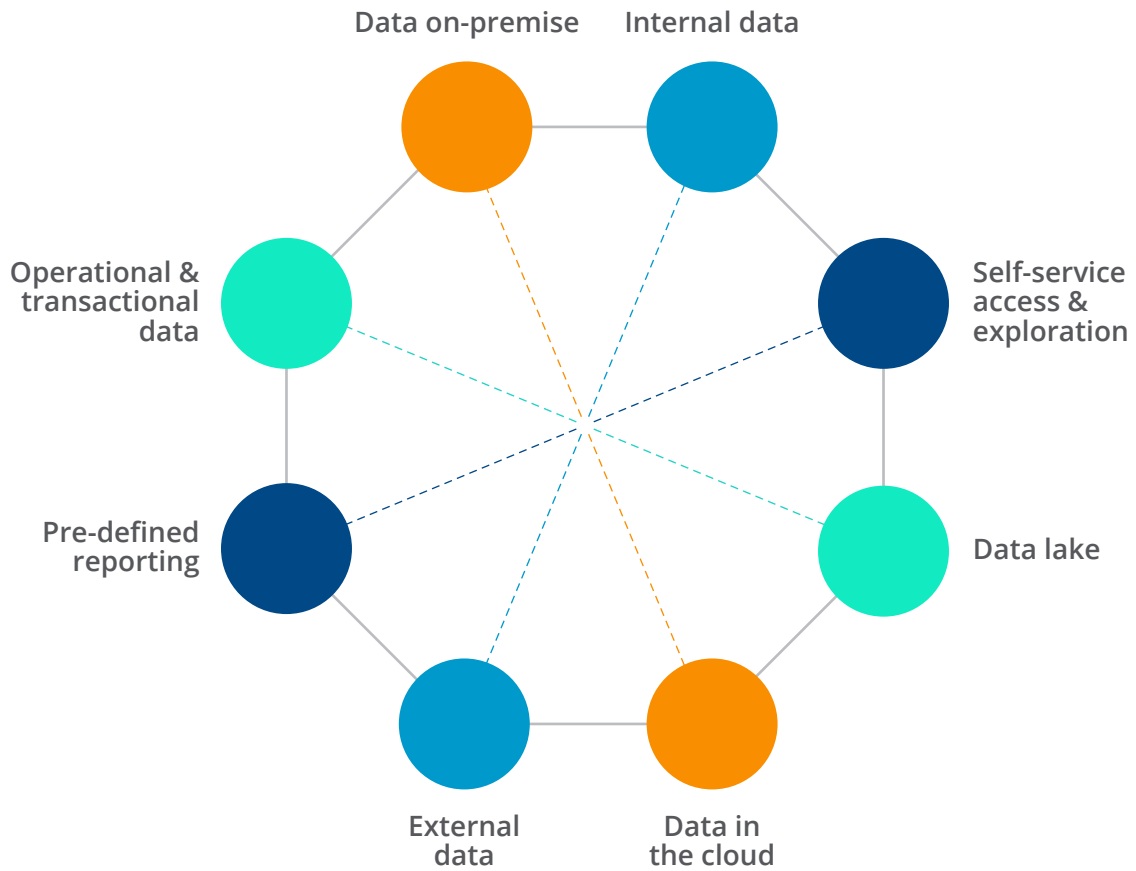Read how a constantly refreshed 360-degree view of customers radically reduced churn

# CITO RESEARCH

## The many dimensions of hybrid data

| DIMENSION | WHAT IT MEANS |
|---|---|
| Structure and form | What kind of data is it? Audio, text, image, video, structured, unstructured, semi-structured, etc. |
| Content specific dimensions | Depending on the content type, we can learn many things from it. Consider audio. Who is speaking? What language are they speaking? What are they saying? What is their tone of voice and sentiment? What do they mean? What additional context is needed? How can we summarize all of those dimensions for machine learning? |
| State | Raw data generated by systems of record inside or outside the organization or "processed data": an output from analysis of one or more sources. |
| Relation-ship to organization | Internal data sources and external data sources (open data, subscription data, third party data, sensitive personally identifiable information (PII), partner data). |
| Time | Historical data, periodic reporting, point-in-time snapshots, time-series data, real time data. |
| Shelf life | Durable (desirable or required to keep/archive) or ephemeral (relevant only for a particular day's analytics; think traders and arbitrage). |
| Flow | Real-time, streaming or stored; think IoT sensor data that may stream but not be analyzed in real-time. It may be stored for later analysis or summarized to capture key information. |
| Edge | Edge systems (such as IoT devices, POS systems, or ATMs) are the opposite of centralized systems. Should data be stored at the edge? Analyzed at the edge so systems can react immediately? Analyzed at the edge and consolidated centrally? Should all interesting events be stored centrally? |
| Consumption | Who can use the data and how? Consider data specialists, data scientists, business analysts as well as machine learning based applications, authorized third parties, and so on. |
| Centralization | Is data centralized in a data lake or data warehouse, on-premise or in the cloud? Or are sources queried where they live (referred to as a federated query or as data virtualization)? |
| Analytical relevance | Profitability of an investment strategy might be determined through complex SQL queries, while fraud detection may require graph analysis on the very same transactional data. |
| Contextual relevance | Data's value and relevance is often determined by how it is employed and by who consumes it across the organization. For example, compliance, sales, and manufacturing functions may use the same data to answer fundamentally different questions. |
| Rigidity | Does the data support ad-hoc discovery and exploration or only predefined reporting? Increasingly the value of data is determined by the number of questions it enables one to ask to probe, understand, predict, and take action. |

# CITO RESEARCH

## Key dimensions of a hybrid data environment



Data on-premise

Internal data

Self-service access & exploration

Data lake

Data in the cloud

External data

Pre-defined reporting

Operational & transactional data

## Reason 2: You can't afford to ignore hybrid data. Your competition isn't

A hybrid data approach puts data at the center of a successful organization's strategy, operations, and execution. This is in stark contrast to traditional application-centric organizations where data's role is often secondary to an application's task, such as supply-chain planning. While enterprise software plays a critical role, systems of record and systems of engagement by design have focused more narrowly on tasks and users. Data by contrast is a currency that can be combined, analyzed, and acted on in myriad ways across an organization's ecosystem of employees, applications, partners, and customers.

Forward-thinking organizations are transitioning from an application-centric approach, where the data created is often secondary to the system of record application, to data-centric, where the key goal is the ability to harness actionable insight from diverse data sources to understand and serve a broad set of constituents and stakeholders. The ability to achieve this goal will determine an organization's ability to compete and win in the marketplace.

Initiatives such as machine learning and AI are data hungry and have the potential to yield material insights and make detailed predictions that traditional, "rear-view" window dashboard reporting analysis cannot. Further, regulatory requirements demand comprehensive, traceable handling of data.

It is vital to understand that data itself is key to delivering true competitive differentiation to forward-thinking organizations. This is because increasingly applications can be accessed by any organization, large or small, thus leveling the playing field. Even code itself is being commoditized, resulting in diminishing returns from custom application projects that were justified in the past by the competitive differentiation they provided. Only hybrid data, and the ability to harness it in all its forms, can truly deliver sustainable competitive advantage.

> Data itself is key to delivering true competitive differentiation to forward-thinking organizations.

Across virtually every industry, time to insight will be a key success factor. Gaining insight on shifts in industry trends or consumer preferences even one day, and in some cases many months, sooner than the competition can make a material difference in an organization's ability to compete and win in today's real-time marketplace. Your competition will be using hybrid data. Your shareholders will demand it.

## Reason 3:
## Many challenges, great value in overcoming them

Organizations face challenges as they make their businesses more digital and leverage the power of hybrid data. One challenge is the varying skillsets of different users and stake-holders. Consider the expertise and computing comfort level of a data scientist, a business analyst, and a salesperson. Regardless of skill level, stakeholders must typically know what data exists, where it is, how to access it, and even how to bring multiple data sources together. They need to be able to envision what data can tell them and how that data might drive business decisions.

Another challenge is data analysis, both where and how to conduct it. Some data must be analyzed near its original source, in place, often on-premise; other data will be analyzed in central data reposito-ries on-premise and increasingly in the cloud. Some data must be analyzed while it is in motion in a small-footprint, low power device while other data can be built up in a petabyte size data lake and analyzed using the latest in columnar, in-memory compute resources. It is critical to be able to use the right tool to answer the right question when searching, querying, joining, analyzing, alerting, and predicting insights from diverse data sources.

Once stakeholders know what data is available and its potential, many will want to access and analyze data simultaneously: can the system handle this challenge? Other challenges include keeping data secure, meeting regulatory rules, and ensuring that data is authoritative and can be trusted.

All of these challenges will only grow over time.

Organizations face many challenges with their data (and gain competitive advantage by overcoming them).

Learn how numerous hybrid data sources help airlines manage thousands of flights per day

### Types of data challenges:

- Consumption
- Integration
- Analysis
- Computing environment
- Size/completeness
- Comprehension
- Security, governance, compliance and trust

## Reason 4: Hybrid data use cases abound, limited only by your imagination

Once you accept that your data is hybrid data, with numerous dimensions, you'll begin to envision more and more use cases.

Many companies want to achieve a 360-degree view of the customer, but to fuel this view, you'll need hybrid data that enables you to correlate social media feeds with customer IDs and activities of website visitors and app users. Others want to detect fraud in real time by blending transactional data with graph-based relationship data to flag and isolate potential rogue actors. Digital transformation requires effective use of all the data you have, and quite possibly adding more sources.

> Digital transformation requires effective use of all the data you have, and quite possibly adding more sources.

Consider how hybrid data changed the game for one financial services institution in the UK. New regulations require that such institutions produce a report at the end of each trading day that shows they didn't expose risk over the course of the day. The report, covering 3 billion risk data points across 30 different risk portfolios, must be produced in one hour. The institution knew that its current systems, which ran 30 separate reports, couldn't meet the compliance reporting requirement. A hybrid data system of record enables the institution to run a single report integrating all 30 risk categories in seconds.

**CITO RESEARCH**

## Reason 5: Hybrid data will fundamentally change your business

As data volumes have grown, so has the time it takes existing systems to analyze that data. Many organizations have coped with increasing volume by performing their analysis on a sampling of the data, with some making decisions using samples of as little as 10% of the data in order to meet user latency expectations or system performance limitations. Yet sampling, while pragmatic, may result in missed breaking trends and inaccurate results. The ramifications can be significant when mission critical decisions involving finances or healthcare are involved.

*Some business decisions are made based on analysis of as little as 10% of the data.*

Analyzing the latest data, refreshed from relevant sources, yields timely and accurate insights. With hybrid data in hand, you can gain a true 360-degree view of your customer. You will learn more about true customer behavior by including the key information you have on purchases, social media sentiment, website clickstream data, customer support data, and more. You can use that multi-dimensional intelligence to drive personalized ads and next best offers.

*With hybrid data in hand, you can gain a true 360-degree view of your customer*

Expanded, comprehensive data sources and larger data volumes can power more innovation such as the use of machine learning and AI. Network optimization enables you to analyze billions of data points across a mobile or transportation network to determine how to best optimize throughput while delivering committed levels of quality of service. Connected IoT solutions that can combine and harness data from edge sensors with systems of record and systems of engagement can deliver a next generation of data-driven services across nearly any industry, from healthcare to transportation to supply chain optimization.

Embracing hybrid data empowers you and takes you from data as a pain point to data as capital to drive growth, innovation, and revenue.

Read about how a major retailer galvanized hybrid data for marketing insights

# CITO RESEARCH

These stories are only the beginning. There are many stories to tell, including yours.

To learn more about hybrid data and the technology to truly harness it and to find out how your organization can embrace the reality of hybrid data, visit actian.com

**This paper was created by CITO Research and sponsored by Actian Corporation**

## CITO Research

CITO Research is a source of news, analysis, research and knowledge for CIOs, CTOs and other IT and business professionals. CITO Research engages in a dialogue with its audience to capture technology trends that are harvested, analyzed and communicated in a sophisticated way to help practitioners solve difficult business problems.

Visit us at http://www.citoresearch.com