

Making data matter in real time



Preface

“Making data matter in real time” is an MIT Technology Review Insights report developed in partnership with DataStax. The report is based on interviews with senior tech executives and experts who specialize in data analytics, artificial intelligence, and information technology. The interviews were conducted between December 2022 and February 2023 to examine why enterprises should, and how they can, leverage real-time data to develop personalized digital experiences or applications for customers, nurture operational agility, and improve business performance. Robert Lemos was the writer, KweeChuan Yeo was the editor, and Nicola Crepaldi was the publisher. The research is editorially independent, and the views expressed are those of MIT Technology Review Insights.

We would like to thank the following executives for providing their time and insights:

Arshal Ameen, Global Head of Application Development, Rakuten Card

Arijit Bandyopadhyay, Chief Technology Officer of Enterprise Analytics and AI, Data Center and AI Group, Intel

SP Naidu, Director of Data Technologies Engineering, Digital River

Saurabh Saxena, Head of Technology and R&D, Uniphore

Jai Thomas, Vice President, IT Infrastructure & Operations, VerSe Innovation

CONTENTS

| | | |
|-----------|--|-----------|
| 01 | Executive summary | 4 |
| 02 | Need for real-time data | 5 |
| | Personalized content, right context | 5 |
| 03 | A real edge | 7 |
| | Data is all around..... | 7 |
| | Real-time data done right | 8 |
| 04 | Rethinking real-time data | 11 |
| | 'Contextual literacy' is key..... | 11 |
| | Using data to 'read the room' | 12 |
| 05 | Obstacles in real time | 13 |
| | As IoT expands, collection challenges grow | 14 |
| | Open-source technology, standards are helping..... | 15 |
| 06 | Conclusion | 16 |

01 Executive summary



As the world becomes increasingly networked and connected devices proliferate, organizations are producing a plethora of data. The potential to collect data is growing exponentially. From smart grids to mobile phones and from connected cars to the industrial internet of things, tens of billions of devices will act as sensors, delivering data to networks.

Whether data is created intentionally or as a by-product of operations, companies that want to grow beyond digitization and become data-driven enterprises must harness it to inform decisions and streamline operations. As experts predict that nearly every employee in the modern enterprise will use data to support their work, targeted curation and analysis of real-time data will be increasingly critical. While the fintech and health-care sectors are leaders in using real-time data, the trend is also picking up in manufacturing, city infrastructure, and other sectors where the push for smart technology is growing.

This report, “Making data matter in real time,” examines how organizations can rethink their data management strategies to leverage real-time data to develop personalized real-time digital experiences or applications that customers want, boost operational efficiency, and increase business performance.

The following are the key findings of this report:

- **Collecting and analyzing data in real time allows companies to create innovative applications.**

When done right, real-time data can be used to gauge customer sentiment and engagement, accurately deliver

targeted content, optimize supply chains, fight fraud, and offer alternatives to missed flights, among other things. However, experts say only a fraction of data from connected devices is collected and processed in real time due to hurdles including legacy technology architectures, signaling untapped potential for enterprises.

- **Organizations need to determine which business processes would benefit most from real-time data collection and analysis, because capturing all data and processing it in real time is expensive.**

Creating architectures to process captured data also becomes exponentially more difficult as the number of devices grows. Companies that create the right data architectures will be able to process the captured data in real time and store the information in a way that is most useful for future decisions.

- **Open-source solutions, cloud, and artificial intelligence systems could help organizations overcome the challenges of managing real-time data.** Experts we spoke to for this report recommend developing initial data pipelines with open-source technologies. That will help organizations to develop in-house expertise, avoid vendor lock-in, and establish a baseline for expectations. In addition, artificial intelligence (AI) is the next natural progression: Operational data platforms need to support real-time analytic functionality as AI and machine learning (ML) become increasingly embedded in enterprise operations. Real-time, AI-powered apps with the right data, at the right time, will deliver the biggest business impact for leading organizations.

02

Need for real-time data



Data is everywhere. Web requests, digital transactions, and myriad online and remote signals are being captured by devices, including mobile phones, cameras, routers, sensors, and even vehicles. According to a report by McKinsey, “only a fraction of data from connected devices is ingested, processed, queried, and analyzed in real time.”¹ The consultancy cites hurdles such as legacy technology structures, the challenges of adopting more modern architectural elements, and the high computational demands of intensive, real-time processing jobs.

Despite these tough challenges, for VerSe Innovation’s Jai Thomas, real-time data is a necessity. As the vice president of IT infrastructure and operations at VerSe, a video and data services firm, Thomas ensures the company is delivering the right content to the right users across its market in India. The world’s second-most-populated country has more than 700 million people connected to the internet,² but only about a quarter do so in English, while the rest use any of nearly two dozen major languages and about 1,700 regional dialects.

Personalized content, right context

VerSe’s short-video platform, Josh, uses customer data, ML, and AI to deliver personalized and relevant content to each of its 360 million users. While individual content creators make and curate the short videos that are at the heart of the technology platform, real-time data is just as important, allowing the company to deliver the three Cs – customers, content, and context.

“For you to understand the customer, you need to understand what kind of content to serve to each individual and the context in which the customer wants that content,” Thomas says. “The real-time data becomes very critical, because on average, VerSe ingests millions of pieces of content every day on Josh, and then it’s critical that we make sure that we are serving the right and relevant content to the user.”

That target will only become more challenging, says Thomas. By 2025, an estimated 900 million to 1 billion people³ in India will be connected to the internet, with about three-quarters using India’s native languages and only a quarter using English.

“To understand the customer, you need to understand what kind of content to serve to each individual and the context in which the customer wants that content.”

Jai Thomas, Vice President, IT Infrastructure & Operations, VerSe Innovation

Harnessing real-time data is not just a technological challenge. It also requires changes to management strategies, operations, and culture.

VerSe is not alone in its reliance on real-time data. The average company is awash in data, with the data-driven modern enterprise relying on employees to harness data on a daily basis to make decisions, inform interactions, and create new business strategies, according to McKinsey.⁴

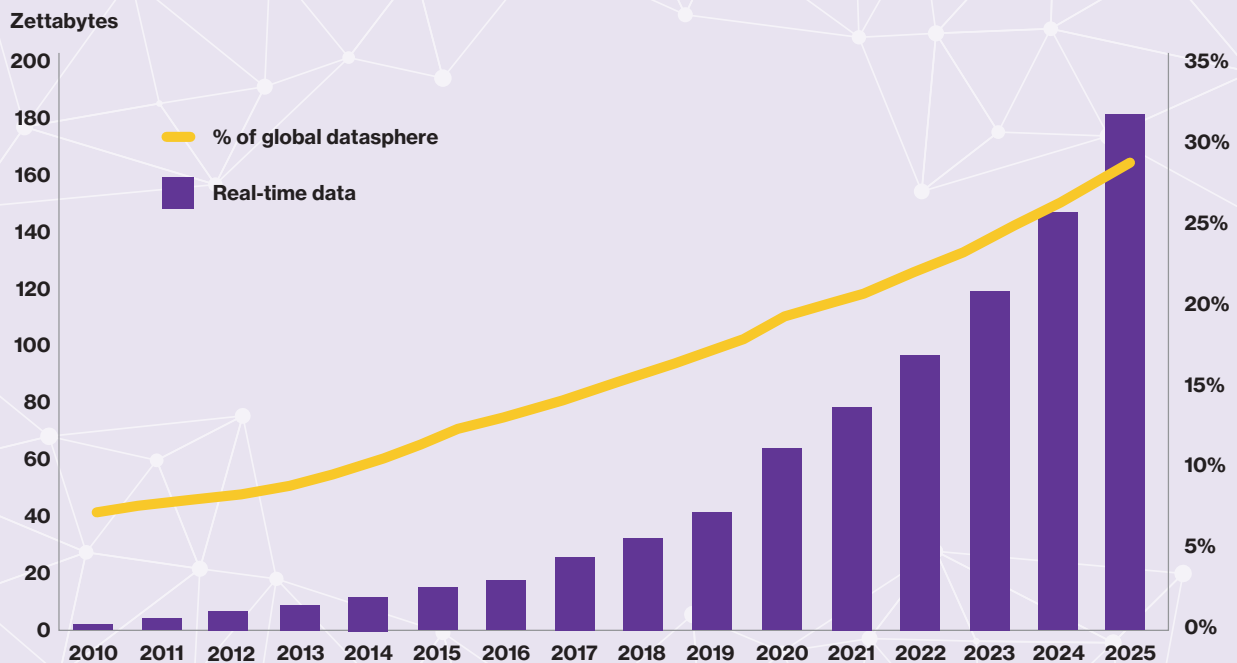
By 2025, McKinsey predicts that nearly every employee in modern enterprises will use data to support their work and inform their decisions. The consultancy also predicts that privacy and compliance will need to be embedded into their network architectures, and data will be delivered and processed in real time.⁵

Against this backdrop, companies need to leverage data to develop real-time digital experiences or applications that customers want, nurture operational agility for timely delivery of products and services, and improve business performance metrics – or risk getting left behind.

Harnessing real-time data is not just a technological challenge. It also requires changes to management strategies, operations, and culture. Companies that successfully turn real-time data into actions will have the advantage and deliver better growth and innovation.

Figure 1: Annual volume of real-time data produced and share of global datasphere

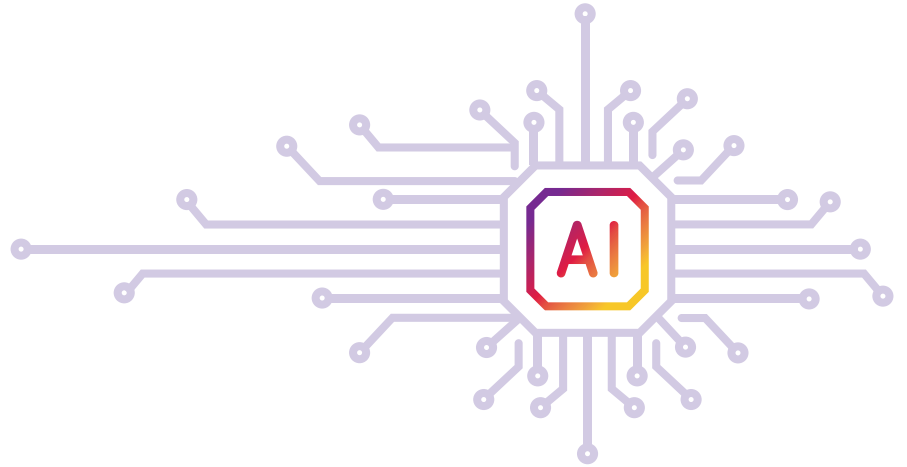
IDC predicts that by 2025 nearly 30% of data produced worldwide will consist of real-time data.



Source: Compiled by MIT Technology Review Insights, based on data from IDC, 2021.

03

A real edge



“Access to timely and accurate data can be used for analysis in real time for better planning and quicker decision-making for companies and their leadership team,” says Arijit Bandyopadhyay, chief technology officer of enterprise analytics and AI at Intel’s data center and AI group. “It also enables business to be proactive and allows them to seize the opportunity with the best contextual and environmental information for faster and more correct decision-making. In today’s world of generative AI and real-time augmented analytics, the right data at scale at the right time, with sub-millisecond processing, is absolutely critical for competitive advantages for [a] majority of customers across use cases.”

Data is all around

According to IDC, the creation and replication of data will grow 23% per year to more than 181 zettabytes in 2025, up from 64 zettabytes in 2020.⁶ Financial organizations and technology firms, as well as the health-care and industrial sectors, are leading in the

use of real-time data. At the same time, the push for more smart technology and management, using small sensors and controllers, has led to the growth of real-time data in city infrastructure, transportation, and other sectors where devices are increasingly connected to operations technology. Within five years, every new application will have a real-time component, predicts real-time data company DataStax. And according to IDC, almost one-third of the “global datasphere” – which it defines as the quantification of the amount of data created, captured, and replicated across the world – will be real-time information by 2025 (see Figure 1).

Utilizing this data is critical for businesses to be competitive in this decade. Collecting and analyzing data in real time allows companies to create innovative applications, automate operational decisions, and create more personalized digital experiences, as demonstrated by VerSe. Staying with traditional approaches, and not accounting for real-time data, means waiting weeks or months for information, which could lead to expensive failure.

“Access to timely and accurate data can be used for analysis in real time for better planning and quicker decision-making for companies and their leadership team.”

Arijit Bandyopadhyay, Chief Technology Officer of Enterprise Analytics and AI, Data Center and AI Group, Intel

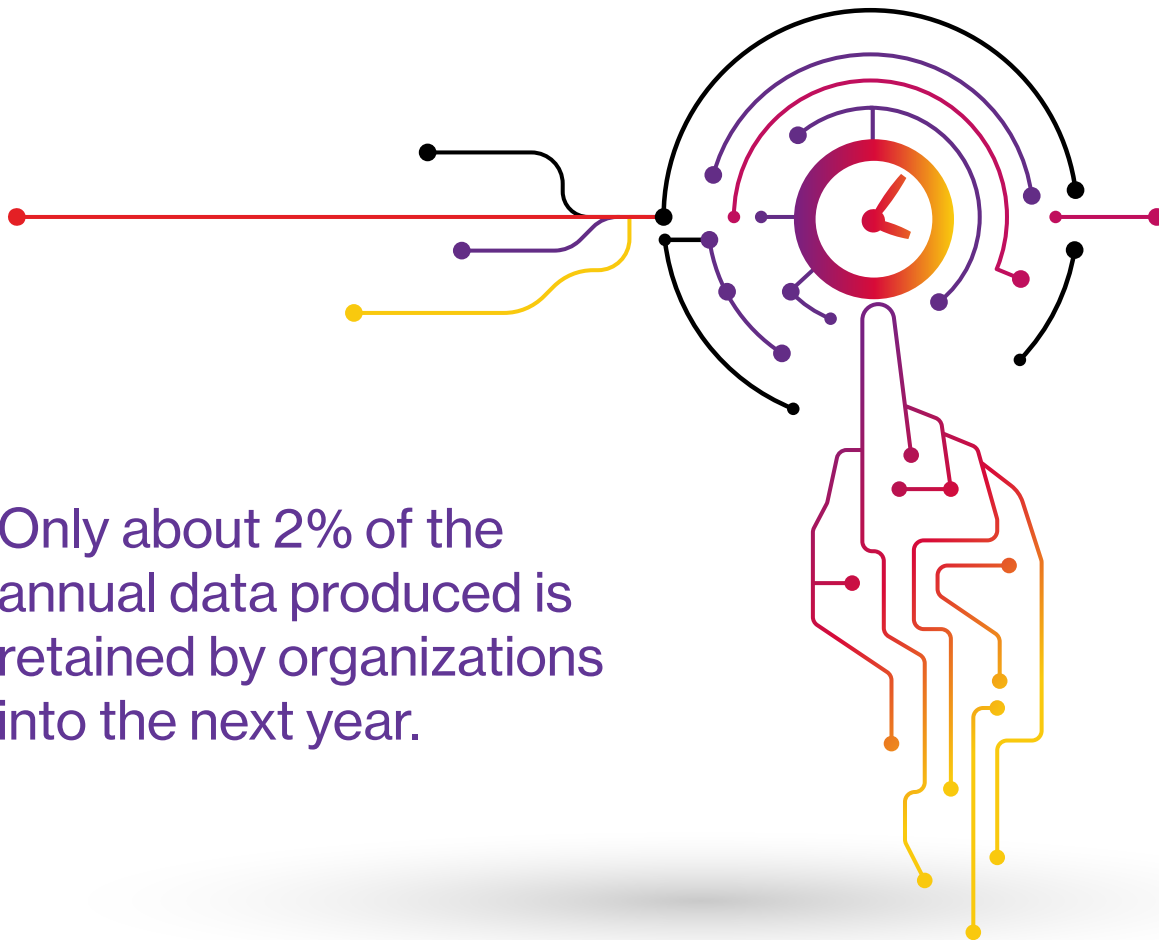
The operational meltdown of Southwest Airlines in December 2022 shows what can happen when outdated systems and a lack of access to real-time data combine with an unexpected event. A snowstorm and severe weather caused the airline to cancel thousands of flights between December 21 and December 31, 2022, causing customers and staff to be stranded at airports during the height of the Christmas and New Year's holidays. Overwhelmed by outdated systems and the absence of real-time information, the airline could not even locate its flight crews, while other airlines, experiencing the same weather impacts, were able to minimize disruptions. In the end, Southwest canceled more than 16,500 flights – accounting for seven out of every eight cancellations⁷ during that period – and lost an estimated \$835 million in revenue.⁸

Real-time data done right

Access to the right data could have saved Southwest a lot of pain. But the airline is not an outlier in failing to be selective about what data it keeps and how it can use that information. Only about 2% of the annual data produced is retained by organizations into the next year (see Figure 2), according to IDC,⁹ which means that executives at every firm need to determine the data most critical for their business. Done right, real-time data can be used to not only offer alternatives for missed flights, but also to optimize supply chains and fight fraud, among other applications.

Uniphore uses real-time data to analyze meetings and presentations, allowing organizations to gauge customer sentiment and engagement, generating

Only about 2% of the annual data produced is retained by organizations into the next year.



insights on how team leaders, sales associates, and customer support teams can do better. “If you pay attention to your customer sentiment and engagement, generally it will result in good outcomes,” says Saurabh Saxena, head of technology and R&D at Uniphore. “We have data that shows that if the customer sentiment improves, your win percentage improves, and your deal cycle reduces, meaning your deals are closing faster at a higher percentage.”

With the proper data and software architectures, real-time data can speed up iterations, improve reliability, and minimize inventory and bottlenecks in hardware and software products.

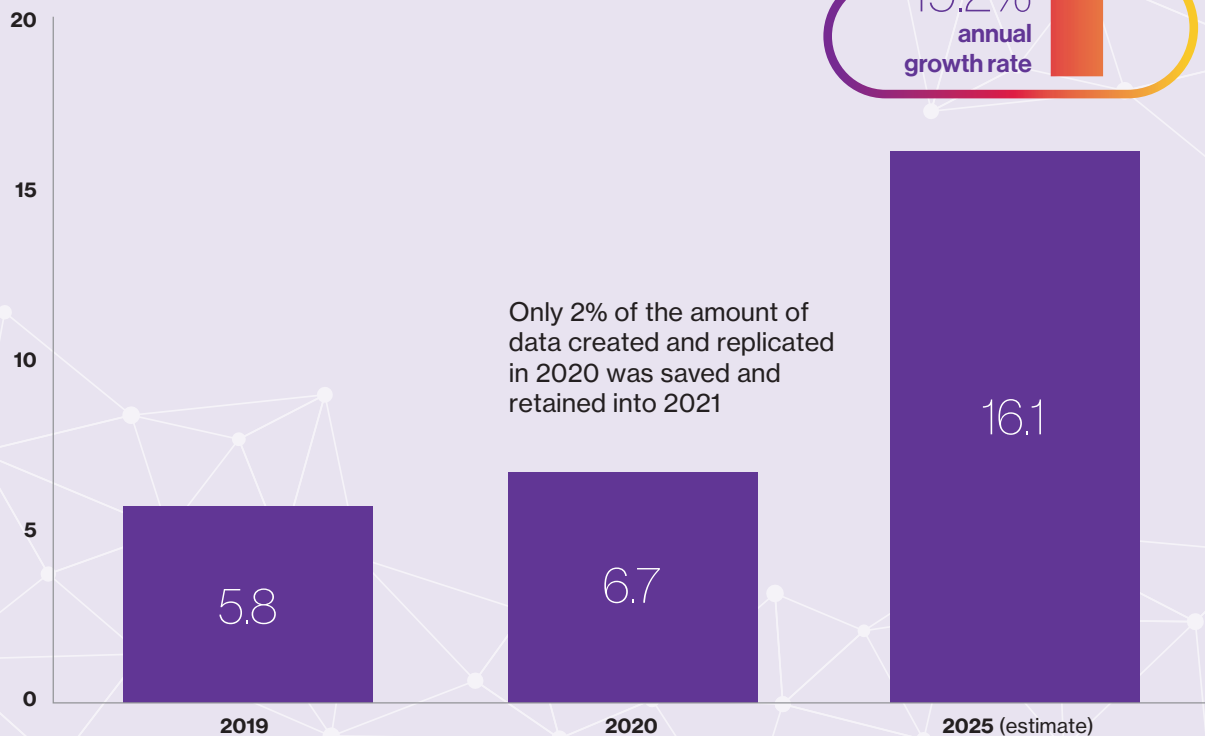
Global e-commerce provider Digital River has already moved much of its critical infrastructure to real-time, or near real-time, data collection. Inventory management, for example, must be real time so that an online shop using Digital River’s services can avoid two customers purchasing the last remaining product. In addition, the company has moved its batch processing of transactions for fraud evaluation toward near real-time decisions so that transactions can be accepted or rejected at the point of sale, says SP Naidu, director of data technologies engineering at Digital River.

“When you lack real-time data, your customer can run out of inventory – not a good experience from their

Figure 2: Global data storage capacity (zettabytes), 2020–2025

Organizations are rapidly creating and replicating more data each year, but they are saving less of the data generated annually, according to IDC.

Global installed base of storage capacity for data reached 6.7 zettabytes in 2020. IDC forecasts it will grow at an annual rate of 19.2% by 2025.



Source: Compiled by MIT Technology Review Insights, based on data from I-SCOOP, 2021.

standpoint,” he says. “Both that and handling fraud in real time actually improves the reliability of our orders. Catching duplicate reservations – or fraud – at the time of fulfillment is a much better experience for the customer.”

For companies that use real-time data efficiently, the payoff can be significant. Data analytics products and services drive companies toward becoming insights-driven businesses, which are nearly three times more likely to report double-digit year-over-year growth, according to Forrester Research.¹⁰

“When you lack real-time data, your customer can run out of inventory – not a good experience from their standpoint.”

SP Naidu, Director of Data Technologies Engineering, Digital River

Partner perspective

DataStax and Intel collaborate to advance Apache Cassandra- and Pulsar-based technologies on Intel’s Xeon scalable processors.

In October 2022, DataStax and Intel joined forces to enhance the performance of the powerful Apache Cassandra database and the Apache Pulsar streaming technologies enabled by Intel’s Xeon family of scalable processors. This collaboration builds upon the long-standing partnership between the two companies, with Intel and DataStax working together for at least four to five years to optimize the open-source database on Intel’s central processing units.

Apache Cassandra is an open-source database management system used by companies such as Apple and Netflix to process and manage user data worldwide. Massively scalable databases like Apache Cassandra run most effectively on hardware created to underpin database workloads.

The Intel-DataStax collaboration is a boost to the long-running success of Apache Cassandra with processors that have the capacity for this usage.

Apache Pulsar is a cloud-native and flexible messaging and streaming platform designed for modern distributed systems. The strength of Apache Cassandra and Apache Pulsar, combined with built-to-last Intel processors, will pave the way for smooth database operations and instant access to critical data.



04 Rethinking real-time data



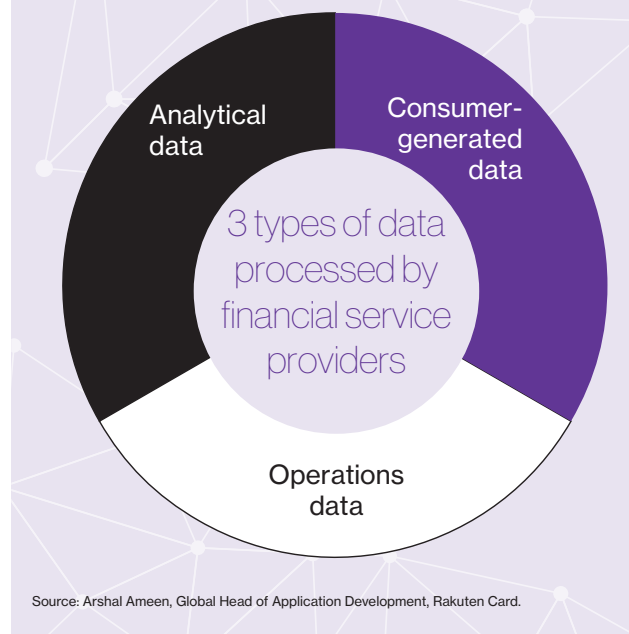
As the amount of data produced grows, companies need to reassess their data management strategies. The first challenge is the volume of data produced by operations and creating an architecture to find the right data. India-based VerSe and its short-video platform, Josh, produce more than 250 billion events every day, says Thomas. “All that data turns into 100 petabytes a month at the cache,” he says. “Now there are all these events, and the question is how do we use this data for meaningful decision-making? To do that while serving them in real time takes a lot of effort.”

In VerSe’s case, the company invested heavily in creating ML pipelines to improve recommendations, dramatically increasing user engagement with its platform. In 2019, before the covid-19 pandemic, the average user spent about 10 minutes on the company’s short-video platform. Today, that figure has grown to 30 minutes. “I think we were able to achieve all those gains primarily because of the ability to look at the data in real time,” Thomas says.

‘Contextual literacy’ is key

VerSe is not alone. Financial service providers process three types of data to increase customer visibility and improve operations: consumer-generated data, operations data, and analytical data. Finding ways to use that data to enrich real-time telemetry and add context to make better decisions is critical, says Arshal Ameen, global head of application development at Rakuten Card.

Figure 3: Data processed by financial service providers



Companies need to foster “contextual literacy,” where each person who uses data can also manage the integrity of the data within their specific context, Ameen says. Companies need to create a culture where everybody shares equal responsibility for ensuring that data is clean and consistent and maintaining data integrity as efficiently as possible. For that, data literacy is most important, he says.

“Technology literacy is not the key here – that’s the easy part. The difficult part is having that contextual knowledge with the actual data that you have in hand.”

Arshal Ameen, Global Head of Application Development, Rakuten Card

“Almost all companies have an outdated approach on how to manage data,” Ameen says. “Technology literacy is not the key here – that’s the easy part. The difficult part is having that contextual knowledge with the actual data that you have in hand.” To develop contextual literacy, companies need to develop their own processes, and then reinforce them with a combination of culture and training, he adds.

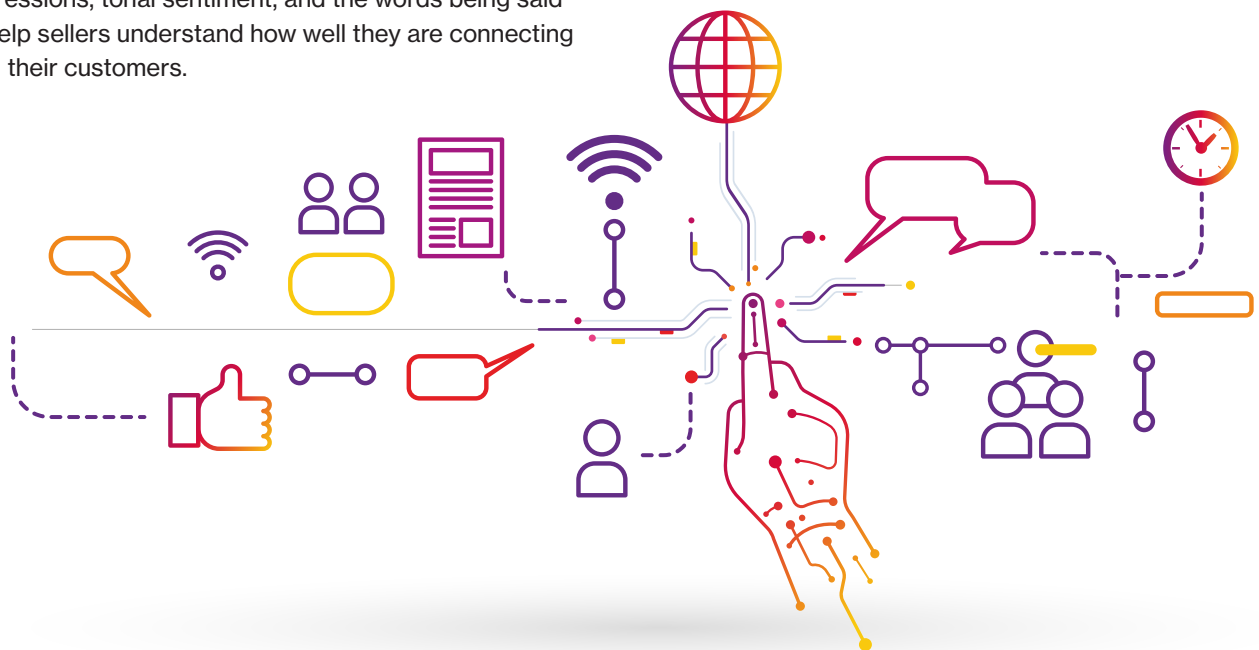
Using data to ‘read the room’

Companies should also look for opportunities to create valuable real-time data from their products and services, not just use extant data. Uniphore says its “Q for Sales” product is an enterprise-level solution that equips sales organizations with an emotional intelligence-based solution to “drive positive, engaging, and trusted customer interaction and outcomes.”

According to Saxena, it uses conferencing call data to understand the participant’s attention, facial expressions, tonal sentiment, and the words being said to help sellers understand how well they are connecting with their customers.

On the flip side, for the seller, it can help focus on conversational techniques like empathy, hesitation, talking speed, talking ratio, and confidence in communication. “The real-time part is we focus on reading the room – a high-level signal,” Saxena explains. “We focus on sentiment engagement of key people: Are you, as a conversationalist, talking too fast? Those are the things that we will do in real time.”

Uniphore pairs those real-time signals with post-meeting near real-time analysis, he says. The result can be used to help train sales and support staff or give presenters feedback on their audience’s reactions. “While the meeting is still top of mind, five to 10 minutes after the meeting, we give you a full summary of the meeting from an emotional standpoint,” Saxena says. “The listener was engaged. They were happy. There were negative moments at this point. Here are the critical points that you should be focusing on.”



05

Obstacles in real time



Using real-time data comes with a host of challenges. Companies need to determine which business processes would benefit most from collecting and analyzing real-time data, because capturing all data and processing it in real time is too costly. Organizations must prioritize the most important data; develop and deploy a robust architecture for data collection, analysis, and storage; and hire or develop the right skillsets within the organization to manage the infrastructure, says Digital River's Naidu.

Without these, real-time data can overwhelm a company's resources, he says. "When you are implementing real-time solutions, you have to address high availability and auto scalability, and have a mechanism to fail gracefully – either during

“When you are implementing real-time solutions, you have to address high availability and auto scalability, and have a mechanism to fail gracefully.”

SP Naidu, Director of Data Technologies Engineering, Digital River

Event and data streaming platforms

Collecting data asynchronously from thousands, or millions, of sensors, devices, or products can quickly become a communications quagmire. To that end, event streaming systems allow the capture of data in real time from a variety of sources, such as mobile devices, cloud applications, monitoring sensors, and operational technology.

An event and data streaming platform stores and routes the real-time data, or events, from these sources and helps tame the continuous flow of information. Apache Kafka is one of the best-known event and data streaming platforms, because it is open source. Originally developed by LinkedIn in 2011 to handle real-time data feeds, Apache Kafka enables users to subscribe to or publish streams of data and process the real-time data as it is produced, ingesting it into a data lake or other backend systems.

Another open-source data streaming platform is Apache Pulsar. Originally developed by Yahoo, Apache Pulsar is a distributed messaging system that provides a “publish-subscribe” model for the efficient exchange of messages between components of a system in real time. Among its many advantages are ease of modernizing legacy applications and lower cost of ownership.¹

1. Robert Kimani, “13 Reasons Why Enterprises Should Use Apache Pulsar,” The New Stack, March 1, 2023, <https://thenewstack.io/13-reasons-why-enterprises-should-use-apache-pulsar/>.

asynchronous processing or at a later point in time, so that it won't disrupt the business itself," he says. "Otherwise, you are at a much bigger risk of your customers suffering a business disruption."

As IoT expands, collection challenges grow

From smart grids to mobile phones, and from connected cars allowing over-the-air updates to industrial internet of things (IoT) networks designed to monitor factories, tens of billions of devices will be connected to networks and act as sensors, delivering data.¹¹ With the increase in connected devices, the potential to collect data also grows exponentially. Creating network architectures, processing the captured data, and storing it in real time becomes increasingly difficult.

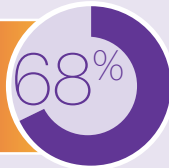
"If you have IoT devices – 5,000 of them churning out information in a smart context at a very high frequency – then you need to read from [them, and] the network has to be appropriate to transmit [the data]," says Bandyopadhyay. "Then, when you're intercepting it and storing it into an intermediary database or whatever, they have to be able to commit fast, in real time, in the (same) sequence of how this data is coming in." An intermediate database, such as Apache Cassandra or MongoDB, allows for the storing of intermediate data. In addition, event streaming platforms, such as Apache Kafka, allow data to be collected as it is produced.

Companies need to determine which business processes would benefit most from collecting and analyzing real-time data, because capturing all data and processing it in real time is too costly.

Figure 4: Why organizations use more open source during tough times.

Key reasons selected by 239 respondents for why their organizations' leaders were more likely to encourage the use of more open source for application development during the downturn triggered by covid-19.

Save money and development time by using existing open source components vs. writing new code



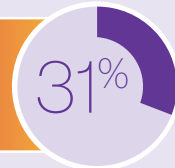
Increase efficiency of application development and maintenance



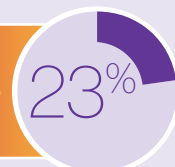
Eliminate vendor lock-in



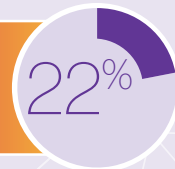
Increase developer satisfaction



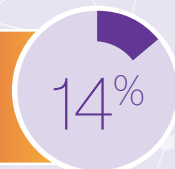
Part of an initiative to migrate more application workloads to open source



Increase security of applications



Part of a new or existing initiative to migrate to microservice architecture



Source: Compiled by MIT Technology Review Insights, based on data from Tidellift, 2020.

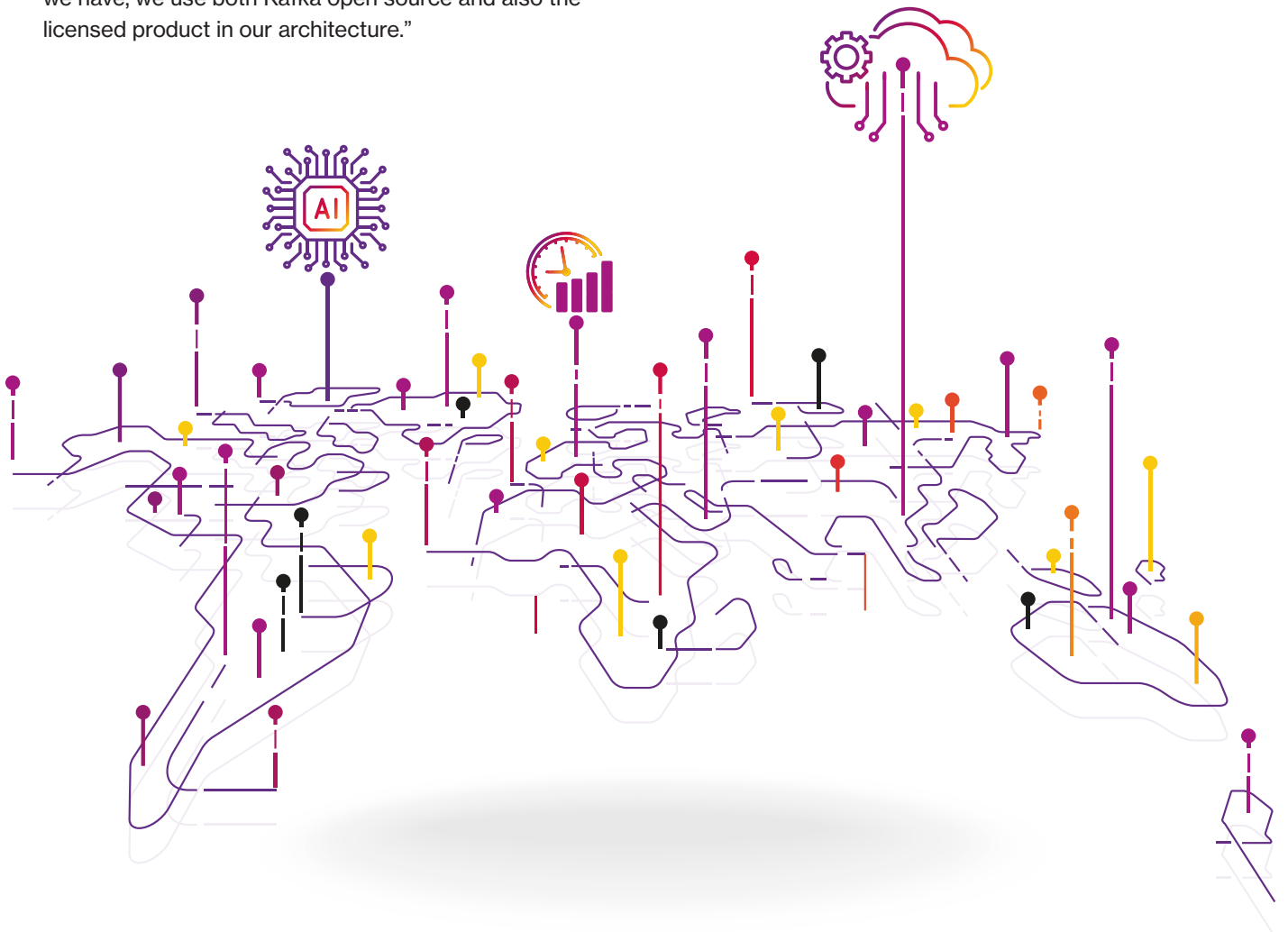
Open-source technology, standards are helping

Open standards that are freely available for adoption, implementation, and updates are increasingly being established. By using open-source technologies, organizations can avoid vendor lock-in and develop real-time architectures that are less likely to be made obsolete, experts say.

Generally, each part of a typical business data architecture has an open-source technology component. Using open source is not mandatory, but Rakuten Card's Ameen recommends developing initial data pipelines with open-source technologies, because not only does this allow organizations to avoid vendor lock-in, it also lets them develop in-house expertise and establish a baseline for expectations. In addition, whether to use open-source technology will depend on the criticality of the data and the need for support. "The key here is to choose based on your use case," says Ameen. "In the case of a streaming platform, which we have, we use both Kafka open source and also the licensed product in our architecture."

As the architecture grows, using a licensed product or cloud platform may make more sense, as many of the complexities can essentially be outsourced to companies that have committed to building on top of open-source platforms. VerSe originally built its platform completely on open-source technologies but has had to replace parts of it with cloud-native services and infrastructure to make the technologies more manageable, Thomas says.

"Everything pretty much on the backend was open source, but I think there is a transition point where you need to look at your overall TCO (total cost of ownership), because managing an open-source infrastructure at scale becomes an issue," he says. VerSe brought in a vendor, whose product is still based on the open-source software the company uses but is easier to manage.



06 Conclusion



Organizations produce a deluge of data, especially if they are running mobile applications, managing operational technology, servicing IoT devices, or creating telemetry. As the number of connected devices continues to grow, the volume of data that will be collected will also increase exponentially.

Organizations must harness a rapidly growing volume of data to improve operations, while allowing every employee to use data to support their work and inform their decisions. Turning data into a business advantage in real time takes knowledgeable design, planning, and a resilient architecture. While many companies base their architectures on open-source software and open standards, the complexity may grow beyond what many enterprises consider to be their business, leading to necessary partnerships with data-specific products or service providers.

Companies that create the right data and network architectures will be able to process the captured data in real time and store information that will benefit their future decisions. The faster companies can capture, process, and access data, the more quickly they can make those decisions.

At the same time, as AI becomes increasingly ubiquitous, organizations are introducing AI and ML models at many operational levels to derive intelligence from business and customer data. Adding AI can strengthen data analytics and enable regular users to better harness the data. For example, in early 2023, DataStax acquired Kaskada, an ML company specializing in managing, storing, and accessing time-based data to train ML models.¹²

According to Gartner's estimates, by 2027 over 90% of new business software applications will contain ML models or services as enterprises utilize massive amounts of data.¹³ With real-time interactivity becoming a necessity, traditional processes that rely on batch extraction, transformation, and loading of data from operational data platforms into analytic data platforms will no longer be adequate. Real-time, AI-powered apps with the right data, at the right time, will drive the biggest impact.

"Businesses must operate in real time, leveraging data to power operations and fuel informed decision-making," said Chet Kapoor, DataStax chairman and CEO, in an article about machines' increasing use of AI and the need for real-time competencies.¹⁴ "The future of real-time data is AI – soon, all applications will leverage real-time data and AI to provide the next-best offer, recommendation, or course of action."

About MIT Technology Review Insights

MIT Technology Review Insights is the custom publishing division of *MIT Technology Review*, the world's longest-running technology magazine, backed by the world's foremost technology institution – producing live events and research on the leading technology and business challenges of the day. Insights conducts qualitative and quantitative research and analysis in the U.S. and abroad and publishes a wide variety of content, including articles, reports, infographics, videos, and podcasts. And through its growing MIT Technology Review [Global Insights Panel](#), Insights has unparalleled access to senior-level executives, innovators, and entrepreneurs worldwide for surveys and in-depth interviews.

About DataStax

DataStax is the real-time AI company. With DataStax, any enterprise can mobilize real-time data and quickly build smart, high-growth applications at unlimited scale, on any cloud. DataStax delivers the Astra DB cloud database built on Apache Cassandra® and the Astra Streaming event streaming technology built on Apache Pulsar™. Hundreds of the world's leading enterprises, including Verizon, Audi, ESL Gaming, and many more, rely on DataStax to unleash the power of real-time data to win new markets and change industries. Learn more at DataStax.com.



References

1. "The data-driven enterprise of 2025," McKinsey & Co., January 28, 2022, <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>.
2. Javed Farooqui, "Report says over 700 million active internet users in India as of December 2022," *The Economic Times of India*, March 16, 2023, <https://economictimes.indiatimes.com/tech/technology/report-says-over-700-million-active-internet-users-in-india-as-of-december-2022/articleshow/98673654.cms>.
3. "India to have 900 million active internet users by 2025, says report," *The Economic Times of India*, June 3, 2021, <https://economictimes.indiatimes.com/tech/technology/india-to-have-900-million-active-internet-users-by-2025-says-report/articleshow/83200683.cms>.
4. "The data-driven enterprise of 2025," McKinsey & Co., January 28, 2022, <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>.
5. Ibid.
6. International Data Corp., "Global DataSphere Forecast, 2021–2025," Redgate Blog, September 8, 2021, <https://www.red-gate.com/blog/database-development/whats-the-real-story-behind-the-explosive-growth-of-data>.
7. Craig Richardson, "What Deeper Lessons Can We Learn from Southwest Airlines' Meltdown?" *American Institute for Economic Research*, January 4, 2023, <https://www.aier.org/article/what-deeper-lessons-can-we-learn-from-southwest-airlines-meltdown/>.
8. Giulia Heyward, "Southwest Airlines' holiday chaos could cost the company as much as \$825 million," *National Public Radio*, January 6, 2023, <https://www.npr.org/2023/01/06/1147445280/southwest-airlines-cancellations-costs-825-million>.
9. Alex Woodie, "Big Growth Forecasted for Big Data," *Datanami*, January 11, 2022, <https://www.datanami.com/2022/01/11/big-growth-forecasted-for-big-data/>.
10. Boris Evelson, "Data, Analytics, and Insights Investments Produce Tangible Benefits – Yes, They Do," *Forrester Research*, May 18, 2020, <https://www.forrester.com/blogs/data-analytics-and-insights-investments-produce-tangible-benefits-yes-they-do/>.
11. The exact number varies widely. A new forecast from International Data Corporation (IDC) estimates that there will be 41.6 billion connected IoT devices, or "things," generating 79.4 zettabytes (ZB) of data, in 2025 (<https://www.businesswire.com/news/home/20190618005012/en/Growth-Connected-IoT-Devices-Expected-Generate-79.4ZB>). Cisco, meanwhile, expects 500 billion devices connected by 2030 (<https://connect.comptia.org/blog/internet-of-things-stats-facts>).
12. Adrian Bridgwater, "The Real-Time AI Data Race Is On," *Forbes*, January 12, 2023, <https://www.forbes.com/sites/adrianbridgwater/2023/01/12/the-real-time-ai-data-race-is-on/?sh=2d04467d4c24>.
13. Ibid.
14. Ibid.

Illustrations

Illustrations assembled by Tim Huxford with elements from Shutterstock.

While every effort has been taken to verify the accuracy of this information, MIT Technology Review Insights cannot accept any responsibility or liability for reliance by any person in this report or any of the information, opinions, or conclusions set out in this report.



MIT Technology Review Insights

www.technologyreview.com

insights@technologyreview.com