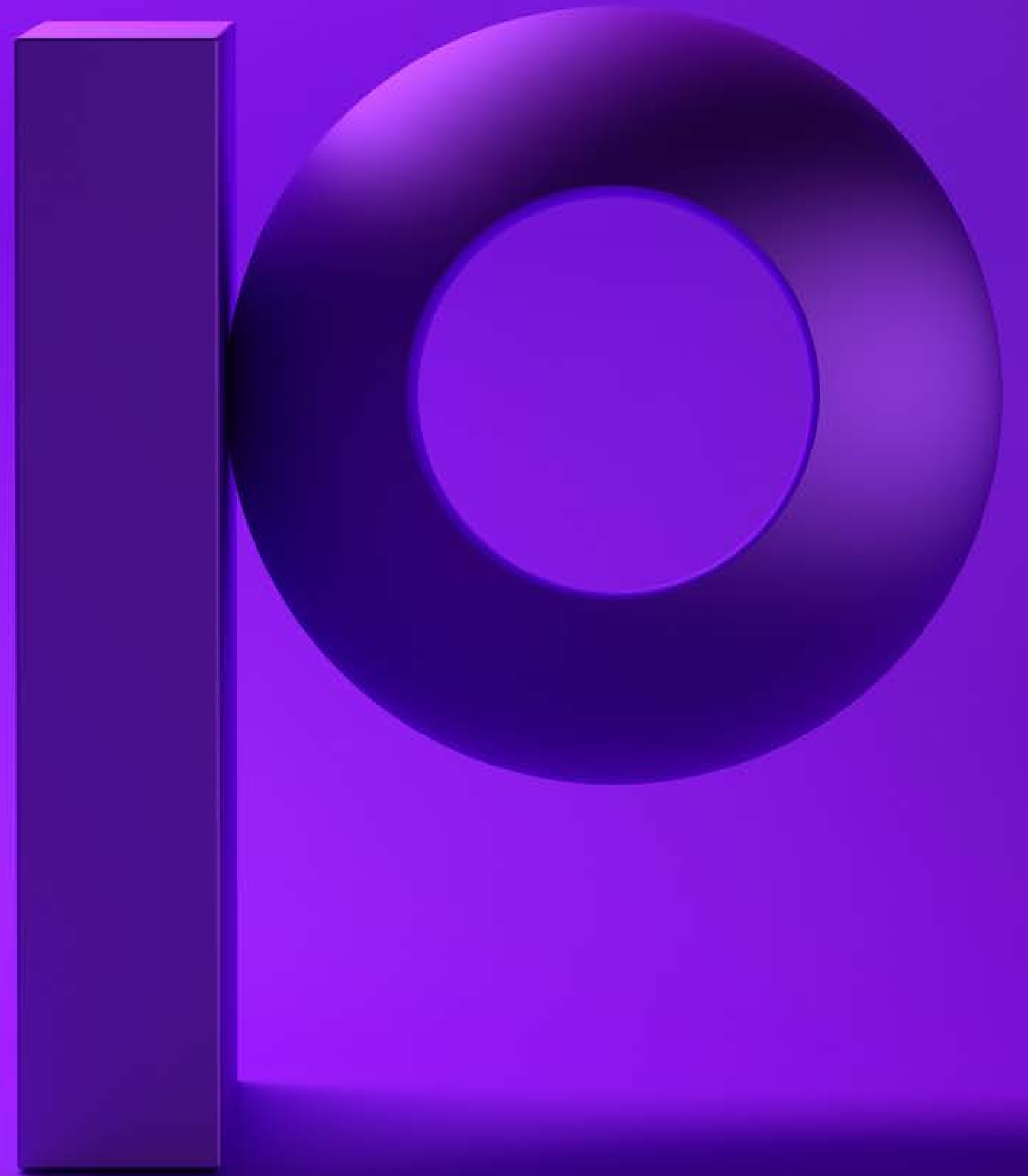precisely

Streaming Legacy Data for Real-Time Insights

Secret Sauce: Streaming Data for Real-Time Insights

Every business is being hit with tidal waves of data across the organization from different sources. To derive value from the 2.5 quintillion bytes of data created each day, most businesses have started initiatives around big data and analytics. However, the challenge with big data projects is that most businesses struggle to find the right data to analyze, or they find that by the time the data is extracted, it has already become outdated. How can organizations solve this problem? Enter streaming data.

Streaming data helps process large amounts of data from many different sources across the business. They help to connect data that drive a variety of use cases from analytics to artificial intelligence and machine learning. Data streaming is now becoming the "secret sauce" of helping businesses to remain competitive within their market through real-time insights.

To have a successful streaming data approach, it's important to note and understand several key trends that proliferate the market:

- While most businesses are using streaming data platforms to build out projects, using data streaming for analytics is still a growing application

- Scaling data integration across the organization is still a challenge — most businesses struggle with data quality, having the right staff in place and data governance

- There are many options for streaming data from coding to open source; however, there is a lot of confusion within organizations around choosing the right set of solutions that meets their needs

- Organizations that have successfully implemented streaming data capabilities have seen the benefits of real-time insights and increased ability to meet shortening SLAs

- Most organizations are creating data integration strategies that do not include the full picture by not connecting legacy platforms such as mainframes to the pipelines feeding their data lakes and enterprise data hubs

Value of Streaming Data

Streaming data pipelines bring a lot of value to the business, such as:

Real-time insight: Streaming data enables organizations to access data as soon as it becomes available. This type of up-to-the minute feedback helps businesses to make the best decisions that they can with the most up-to-date information and keep a competitive edge.

Break down data silos: In Precisely's *2019 Data Trends Survey*, more than two-thirds (68%) of respondents indicated that siloed data negatively impacts their organization. Building data pipelines helps to make all enterprise application data accessible and usable — and in turn begins to break down the data silos that commonly exist within large environments.

Evolve business applications: In scenarios where data is generated on a continuous basis, streaming data can help drive the collection of that data. As data collection grows over time, the data can then be fed into more complex processes helping to create the foundation for next generation business applications that are driven by AI or machine learning.

Improve customer experiences: Streaming data when delivered in real-time can determine correlations, changes, and common behaviors over time, helping organizations to gain insights into the needs of their customers. By leveraging these insights, organizations can begin to create more customized customer experiences.
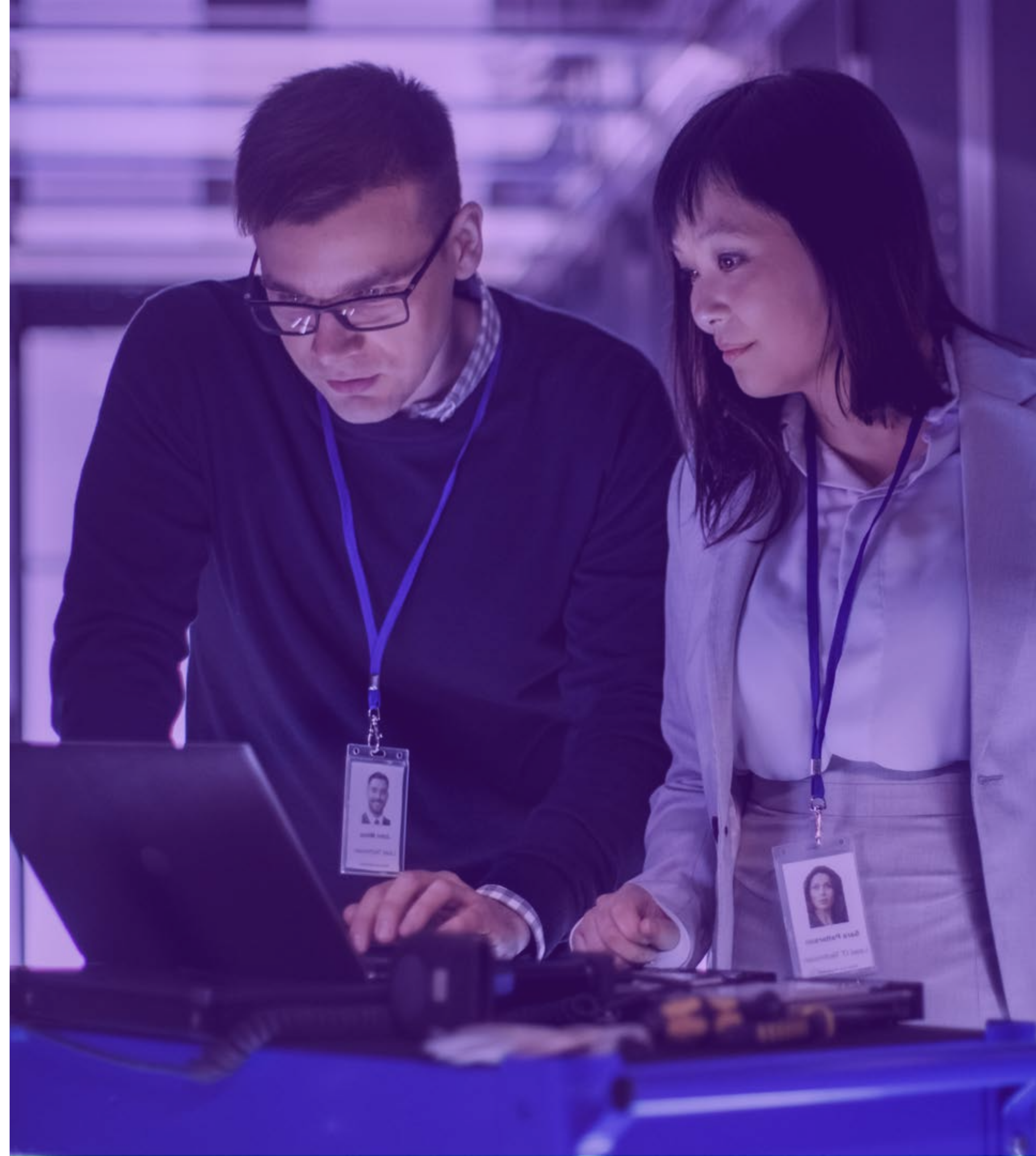
Importance of Legacy Data

Even with the growth of next-gen technologies, legacy systems (i.e. mainframes and IBM i) still play an important role within many businesses. More than 70% of Fortune 500 enterprises continue to use mainframes for their most crucial business functions. Mainframes often hold critical information — from credit card transactions to internal reports.

Most large enterprises have made major investments in mainframe data environments over a period of many years and will not be leaving these investments anytime soon. It is estimated that 2.5 billion transactions are run per day, per mainframe across the world. This high volume of data is one that organizations cannot choose to ignore or neglect. Additionally, mainframes often have no peer when it comes to the volume of transactions they can handle and cost-effectiveness. As a result, these environments contain the data that organizations run on, and in turn, power the strategic initiatives driving the business forward — machine learning, AI and predictive analytics.

Business insights, artificial intelligence and machine learning efforts are only as good as the data that is being fed in and out of them. Leaving mainframe data out of the equation when building strategic initiatives risks omitting critical information that could greatly influence business outcomes. Specifically, neglecting mainframe data from strategic initiatives results in:

- The value of an organization's big data investments being diminished
- Analytics that are not accurate or complete
- Large, rich enterprise datasets that never even get analyzed

Legacy Data + Streaming – Easier Said Than Done

With so much valuable data on the mainframe, the most logical thing to do would be to connect legacy to big data applications. However, many complexities can occur when businesses attempt to stream data between new and legacy sources. As a result, the plans made to connect these two areas are often easier said than done. Common challenges of streaming legacy data are:

Data bottlenecks: Integrating mainframe data into newer systems, such as Hadoop, is problematic because there is no native connectivity and processing capabilities in these new systems for mainframe data. The time and effort it takes to load hundreds, or even thousands, of database tables into a big data platform – combined with inefficient use of system resources – can create a data bottleneck that hampers your streaming data projects from the start.

Complex copybooks: Most tools cannot handle variable length records from legacy systems without padding to max length, grinding processing to a crawl and often requiring multiple redefines. Additionally, legacy systems are often not readily compatible with affordable open-source frameworks and data formats for analysis.

Data Governance: When moving data to the cloud, most tools must unpack, expand and convert data to human readable format. However, performing these conversions causes several different issues to the original mainframe data, such as:

- Copybook no longer matches data.
- Metadata is no longer reliable.
- Record descriptors are lost.
- No copy kept of data before modification.
- Data is bloated so processing chokes

Real-time Change Data Capture: With data flowing between mainframe and business applications, the other challenge becomes the need for ongoing real-time change data capture. Just like any other system feeding a business application, the data from the mainframe must remain fresh. This means that there is a need to track and detect when these changes occur. The challenge then becomes how fast these changes are captured as data from the mainframe often changes very rapidly. Most of the time, businesses are not able to capture these changes in real-time and, therefore, operate on a delay of information.

Stream Your Legacy Data

Realizing the full value of your organization's data for actionable insights and a competitive edge requires the ability to connect to legacy systems. To do this, the business needs to invest in solutions that enable these connections and help drive the most value to business applications.

Precisely's change data capture solution, Connect, is designed to help businesses stream real-time application data from legacy systems, such as mainframes, to mission critical business applications and analytics platforms that demand the most up-to-date information for accurate insights (e.g., fraud detection). Connect enables businesses to:

- Have a reliable transfer of data from legacy systems to business applications, even if connectivity fails on either side.
- Keep data in sync in real-time without overloading networks, affecting the source database performance or the need for specialized skills.
- Support both cloud and hybrid architectures through a variety of application topologies that can be combined, including: one-way, two-way, bidirectional, consolidated, distributed, and cascade.

With Connect real-time data replication software, businesses have the power to build stream and share application data across the enterprise – from mainframes to the cloud – to drive the business forward.

Learn how Precisely can support your enterprise.

precisely

Precisely is a global leader in data integrity, providing accuracy and consistency in data for 12,000 customers in more than 100 countries, including 90 percent of the Fortune 100. Precisely enables companies to integrate, verify, locate, and enrich their data to power better business decisions. To learn more, visit www.precisely.com.

www.precisely.com