StreamSets
A SOFTWARE AG COMPANY

Understanding, Governing, and Managing Data Pipelines: The Data "Wild West" Facing Financial Services

streamsets.com

Introduction

Bringing Order to Chaos in the Data "Wild West"

Heavy regulation creates a minefield of restrictions, requirements, and guidelines that financial services (FS) organizations must carefully tread. And these regulations are continually changing from region to region, so what applies in one bank's headquarters may not be the case in its other offices.

At the same time, financial institutions (FIs) operate in a data "wild west" of complex architectures spanning hybrid and multi-cloud environments, creating a web of legacy systems, point solutions, and custom-built tools that is difficult to navigate. Decision-makers often struggle to understand, govern, and manage a fragmented data supply chain to ensure compliance.

At the same time, line of business teams in FS organizations have become increasingly savvy about using data to inform their daily operations. These teams work independently to build their own data pipeline integrations to add value to the business. However, this often means pipelines are built without the IT department's knowledge, creating new data silos and more visibility gaps – leaving FS organizations in data darkness.

When data ecosystems are governed well and IT teams have visibility into pipeline creation, the way tools are used to create, process, store, share, and analyze data can be standardized. Standardization enables FS businesses to meet stringent governance policies as data is transported and processed through data pipelines. But many FS organizations lack the ability to implement and enforce these controls today.

FIs manage large volumes of sensitive data, from customer purchasing habits and insurance claims data to stock and share price tracking. This data provides an unmatched opportunity to inform risk-based decision-making – but only when used

safely and within regulatory guidelines. The financial sector must also meet fast-changing demands when it comes to environmental, social, and governance (ESG) reporting, as consumers and regulatory bodies hold FIs to higher sustainability standards.

Gathering accurate and real-time information to meet reporting requirements requires robust governance. But when consistent measures to safeguard data are absent, FS organizations increase the risk of fines, reputational damage, and security breaches. Without governance policies that establish data uniformity, FIs can't unlock the value of their information assets nor be confident they are working with reliable data.

To dig into modern FS organizations' issues in managing and governing data pipelines and operations, we surveyed data decision-makers and practitioners from large financial services enterprises in the US, UK, Germany, France, Spain, Italy, and Australia. In this report, we find out how FIs are navigating data governance challenges today.

What's inside:

- Closing the Governance Gap in the Data Wild West
- Confusion over Accountability Creates a Data Visibility Challenge
- The Multifaceted Data Privacy and Governance Challenge
- Meeting ESG Targets

- Reducing the Risk of Breaches
- Business Teams are Operating Outside of IT's Vision
- Establishing a Data "Mission Control" for the Building Blocks of Data Governance

Closing the Governance Gap in the Data Wild West

Financial services organizations have increasingly diverse modern data ecosystems. These ecosystems provide the foundations for innovation as analytics tools and projects can be decentralized and put in the hands of business users. Decentralization helps ease the burden on IT teams and enables line of business teams to create feature-rich financial products – such as spending alerts and categorization, personalized lending solutions, or promotional offers tailored to customer spending habits.

However, without clear governance policies, it also increases risk. Data is increasingly spread across the entire ecosystem, adding to the complexity. More than half (54%) of data leaders and practitioners in FS organizations say modern infrastructures that span on-premises and multiple cloud environments, combined with data decentralization between LOB teams, have created a data "wild west." Moreover, 63% say this fragmentation in the data supply chain has made it harder to understand, govern, and manage data in FS organizations.

Connecting many data sources and ensuring uniformity between definitions, formatting, and metadata is hugely complex. It makes initial deployments and later updates a painful management headache for data teams. They need a better way to enforce consistent rules to govern the

Five Critical Pillars of Data Governance in Financial Services

Good data governance requires a well-defined strategy.
Here are five fundamentals to consider when developing yours.

- **Identify your data**
  To design an effective strategy, you need to know your entire data landscape inside out, including types, structures, movements, locations, and points of data transformation.

- **Establish a governance body**
  The data governance body is a central control point around which all teams and departments can agree on consistent policies that align with business goals.

- **Ensure "privacy by design"**
  A privacy-first approach is central to good data governance. It involves collecting only necessary data, masking personally identifiable information (PII), and using data only for intended purposes.

- **Metadata management**
  Properly managing metadata makes it easier to track data changes, control data access, and understand relationships between data to fulfill governance requirements.

- **Data quality management**
  An effective data governance strategy will establish consistent criteria and scoring to ensure data is high quality and reliable for use in analytics and AI/ML applications.

Once you've designed your strategy following these pillars, you are ready to implement it. Check out this article to learn how.

increasing number of apps, systems, data sources, and tools that get integrated into FS ecosystems.

The data leaders and practitioners in our research agree, with 88% saying they want consistent security measures to protect data as it flows between on-premises and cloud sources. Without consistency, visibility and control are lost, significantly increasing the risk of data breaches and fraud. Moreover, online attacks happen fast, and FS firms must be able to detect and respond to cybersecurity threats in real-time.

Life on the Data Frontier

54%
say that modern infrastructures and data decentralization have created a data "wild west" in FS organizations

63%
say data fragmentation makes it hard to

88%
want consistent security measures to protect data as it flows between on-premises and cloud sources
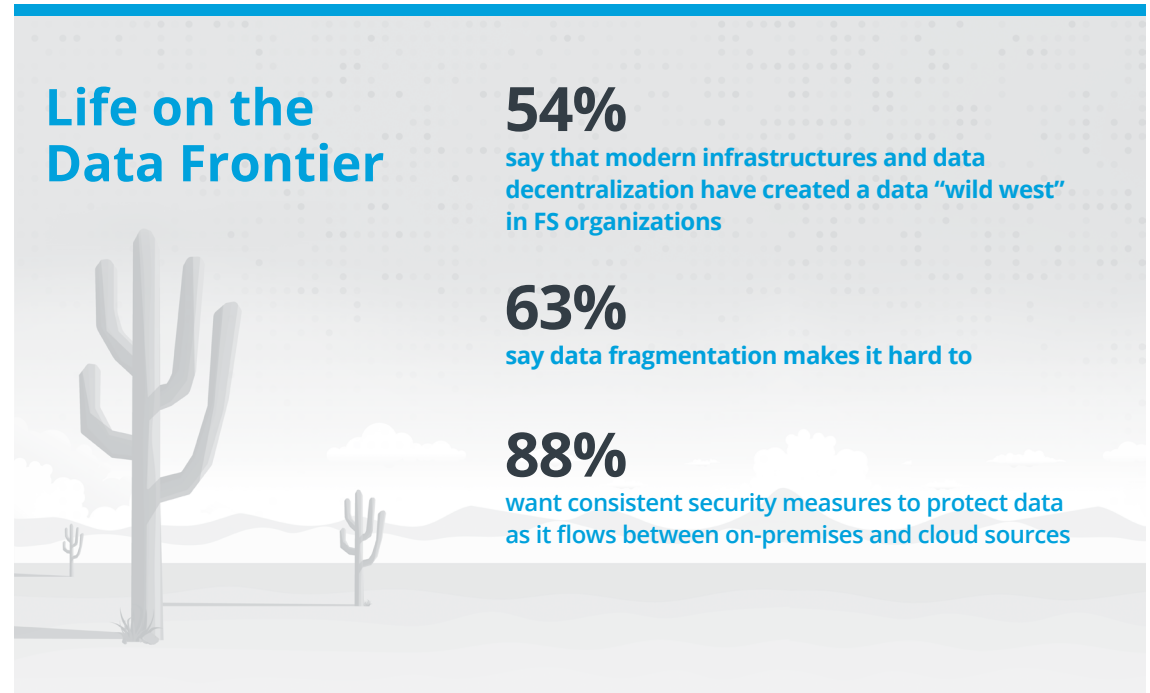
Confusion Over Accountability Creates a Data Visibility Challenge

An effective data governance strategy lays out clear roles and responsibilities to understand who is accountable for which area. However, in many FS organizations, there is often confusion over who is responsible for managing data. Plus, different business stakeholders have different priorities when it comes to pipeline building. Data leaders tend to focus more on compliance and risk management, whereas LOB teams prioritize the speed of data generation.

The research results show that responsibility, duties, and processes for data governance are typically not well-defined. 37% of respondents say the primary responsibility for managing data in FS organizations sits with the central IT team. However, a fifth (21%) say it lies with line of business teams. And 42% say the responsibility is split between line of business teams and IT.

However, despite this and contrary to earlier findings that show data professionals feel they are operating in a data "wild west," 75% of data leaders and practitioners in FS organizations say they are confident they have complete visibility and control over their data. But the gaps are apparent when we shine a light on the results. With new regulatory requirements, ESG concerns, and cybersecurity risks coming to the fore, many FS organizations are

Figure 2.
The causes of data friction integration within respondents' organizations
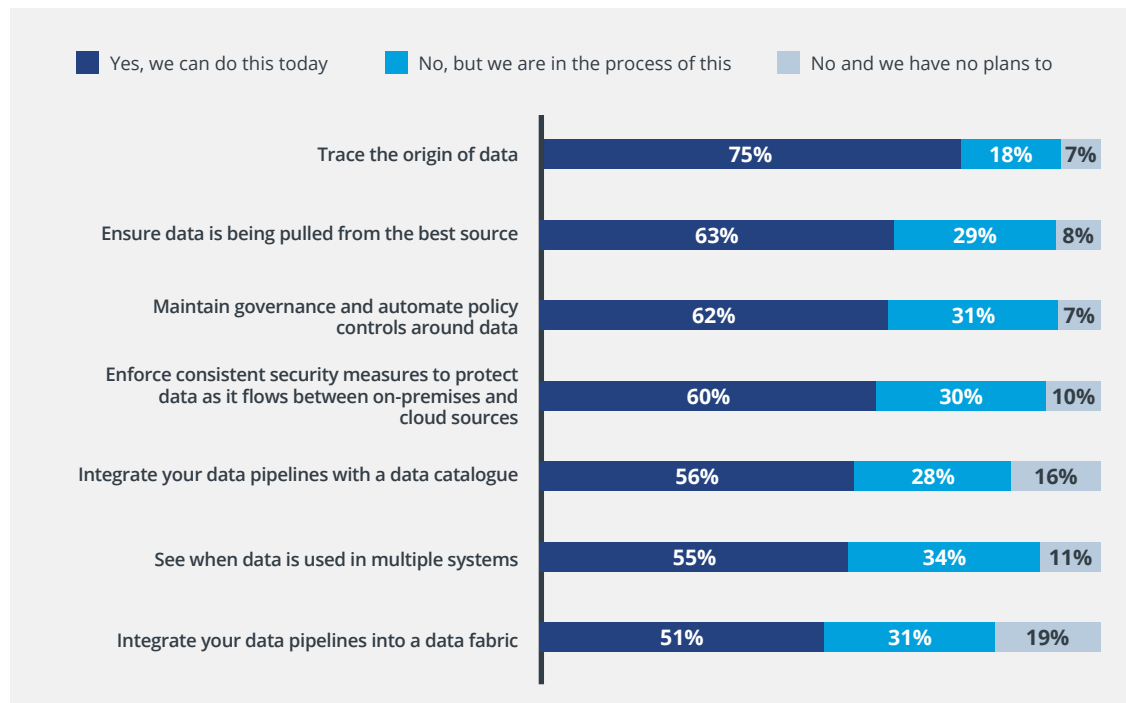
harboring a ticking time bomb of governance risks they are unprepared for.

The Multifaceted Data Privacy and Governance Challenge

The research further finds that 38% of FS

organizations cannot maintain governance and automate policy controls around data, and 40% cannot enforce consistent security measures to protect data as it flows between on-premises and cloud sources. This is a clear vulnerability. An inability to automate policy controls increases the chance that

employees who are not compliance experts or are not permitted to view certain data may inadvertently violate regulations.

This is not just a challenge for legacy institutions struggling with complicated architectures over mainframe, public, and private cloud environments. In fact, new entrants and challenger banks are typically less trusted with data. For instance, fintech organization **Plaid paid $58 million** in 2021 to settle allegations that it violated data privacy rules by obtaining and using customers' bank account credentials and financial information without consent. **The Open Finance Data Security Standard** (OFDSS) establishes common consumer data security, privacy, and control standards in the digital finance ecosystem. This robust framework for emerging cloud-native FS organizations ensures they align with existing enterprise requirements.

Meeting ESG Targets

The spotlight is increasingly turning to ESG. Regulations calling for greater transparency and disclosure of ESG data mean that FS organizations face mounting pressure to meet evolving ESG reporting demands. For instance, the **Sustainable Finance Disclosure Regulation** (SFDR) in the EU, which aims to increase transparency around

sustainability claims made by financial market participants, is bringing new expectations for FIs to collect and report on their ESG data. There are similar requirements in the US too, as the US Securities and Exchange Commission launched its **Climate and ESG Task Force** to protect investors from material gaps or misstatements in issuers' disclosures. FS organizations need to leverage data to support their ESG journey. Still, given what the data indicate about the current lack of accountability for data governance, many will struggle to navigate this continuously shifting sustainability landscape.

Reducing the Risk of Breaches

FIs need visibility into where data flows to and from to protect themselves from breaches. But the research reveals that 45% of FS organizations can't see when data is being used in multiple systems, and 37% cannot ensure data is being pulled from the best source. Moreover, nearly half (49%) cannot integrate pipelines into a data fabric, and 44% cannot integrate pipelines with a data catalog.

Maintaining compliance around data access, use, and storage is challenging, particularly as many large FS organizations can have thousands of data pipelines in operation. FIs are held much more accountable for their actions than almost any other industry, and protecting customer data is paramount. Breaches

can knock consumer confidence, such as the major breach that hit **JPMorgan Chase** in 2014, impacting over 76 million households in the US. Having complete visibility over data flow, storage, and use helps to tighten cyber defenses and regain control – minimizing the risk of a costly and damaging breach.

CASE STUDY

Leveraging Data to Enhance Business Operations & Customer Experience

Aon is a leading risk consulting firm providing a broad range of solutions across Commercial Risk, Health, Reinsurance, and Wealth. To maintain its position as a leader in the financial services industry, the organization knew it had to adopt modern DataOps practices that would enhance internal operations and customer experience. Aon needed to establish a global environment to store and process strategic proprietary data, in batch or near real-time, to create differentiating value for its clients.

Before adopting the StreamSets platform, lead time to acquire data from various sources for projects could take months. The teams that needed that data had to go through a project intake process, buy and install ETL tools, spin up SQL databases, and then begin their work. With StreamSets, Aon's time for data ingestion has been significantly reduced, and the organization now also has a team onboarding process that takes weeks or less.

As Aon ingests data from various disparate sources, StreamSets pipelines help clean and augment the data in flight. Not only has this helped the organization centralize the data, but they can also run, better connect, and expose it for analytics through Snowflake. By helping data engineers build pipelines to supply strategic client data of varying types and from various sources, StreamSets is helping Aon build its view of the client, including relevant data such as industry risk attributes, insurance programs, internal revenue, commercial claims, health benefits, and investment portfolios, to name a few.

# Business Teams are Operating Outside of IT's Vision

As outlined, the democratization and decentralization of data have delivered several benefits to FS organizations. When LOB teams can access and handle data through a comprehensive governance framework, teams can innovate safely within centralized guardrails. But without those guardrails, blind spots are unintentionally created.

Business teams in FS organizations often move fast and develop new integrations and pipelines without IT's oversight. The research finds that more than two in three (68%) data leaders and practitioners in FS organizations say that line of business teams and users 'go rogue' and create datasets independently without telling IT or data teams. And almost all respondents (96%) say this creates problems and business risks.

Respondents reported several risks stemming from business users operating outside IT's field of vision, including opening up the risk of that data not being used in a compliant way (41%), creating the risk of data breaches (41%), and causing the organization to use inaccurate data (41%).

Figure 3.

The LoB teams creating their own datasets without notifying the data / IT teams creates problems for 96% of respondents
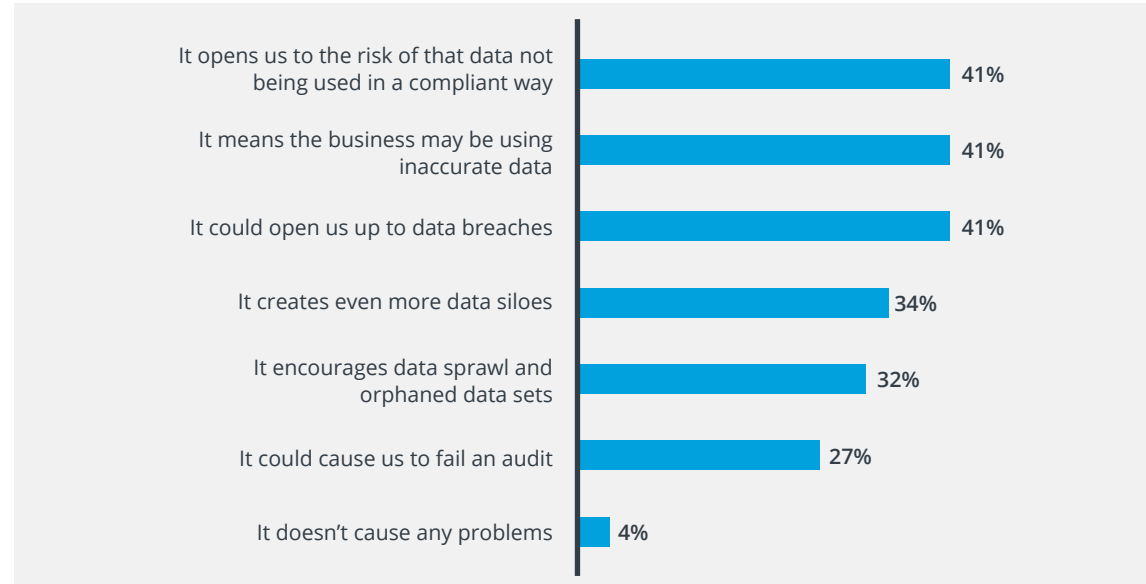
The Problems Caused when LoB Teams Create Datasets Without Notifying the Data / IT Teams

The creation of data silos was also cited as a challenge (34%), as was the spread of data sprawl and orphaned datasets (32%). In fact, more than half (59%) of respondents identified orphaned datasets that users have created and then forgotten as one of the most significant risks to their business.

As customers constantly raise the bar in financial services, FS organizations need to strike a balance between control and compliance and enabling innovation. Everyone in the business wants access to more data. To drive transformation and delight customers, FIs must empower LOB teams with the data for innovation, prototyping, and experimentation – but they must do so safely. There is an appetite for this way of working in FS firms; the research reveals

that 78% of data leaders and practitioners want to empower line of business teams to use data while maintaining visibility and control, and 81% want to enable a self-service data model for end-users.

However, considering what we've learned about the lack of visibility that FS firms have over their data ecosystems, many will struggle to empower business users to deliver transformation. IT teams need transparency to allow LOB teams to independently access and utilize data while ensuring governance policies are in place. Data leaders need a 360-degree view of how data flows across the enterprise, including where new integrations or pipelines are built, when, and by whom. This visibility will shine a light on any blind spots that cause risk to the business while driving innovation in business teams and maintaining compliance.

Establishing a Data "Mission Control" for the Building Blocks of Data Governance

To achieve the level of visibility needed to empower safe innovation, data leaders and practitioners must establish a centralized management console that acts as a data "mission control." And given the chaos of modern data ecosystems – and the watchful eye of global regulators – any solution needs the capability to "see" across all environments.

Every environment has unique deployment and governance challenges, so end-to-end visibility is crucial. Data teams in FS organizations must ensure that cloud-based applications can securely leverage data from a different environment and vice versa. The research shows that respondents want to achieve this, with 78% saying a single platform that can handle the complexity of data spanning across cloud and on-premises worlds would be a huge benefit.
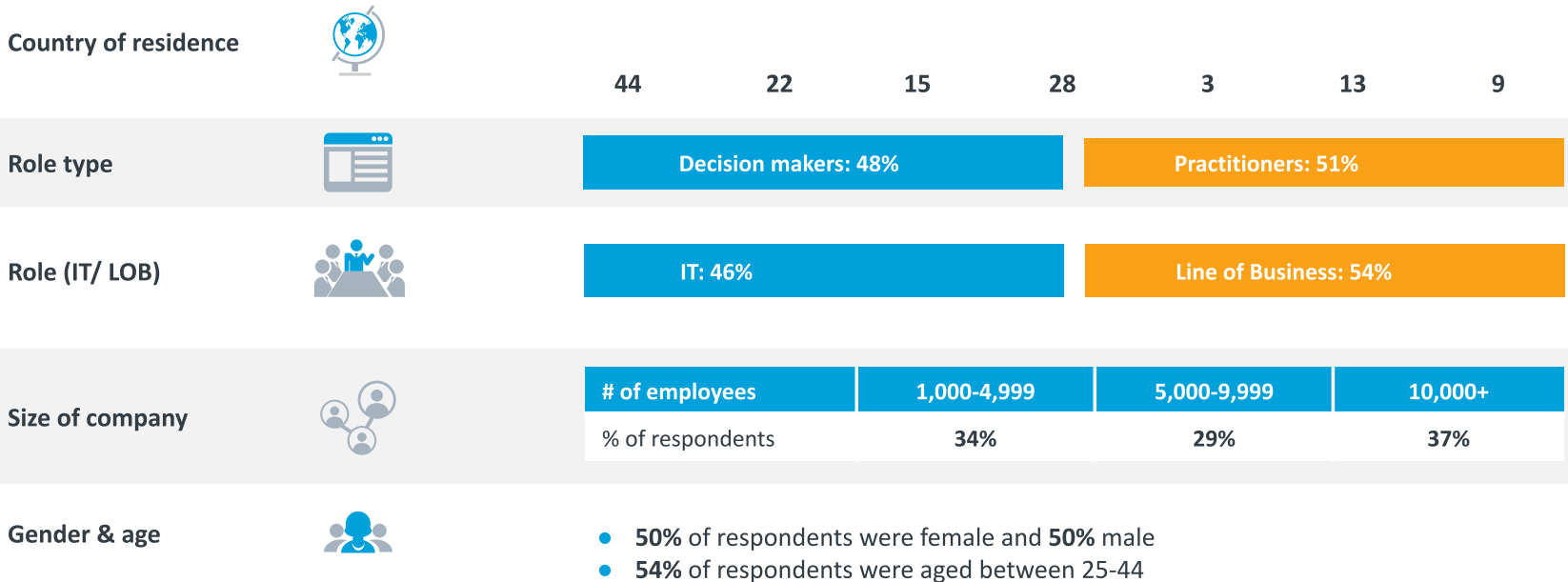
Data leaders and practitioners who can establish a centralized data console will equip LOB teams to extract maximum value from data. At the same time, they will lower costs, reduce the headache of managing a fragmented data supply chain, and embed good governance throughout the organization.

FS organizations will continue to face stringent data governance requirements in ESG, security, and regulatory compliance. StreamSets helps businesses create order from chaos to reduce the compliance burden. Our single, fully managed, end-to-end platform becomes an organization's mission control. Sophisticated topologies deliver deep visibility into how systems are connected, allowing data leaders and practitioners to see how data flows across financial services enterprises.

With StreamSets, you can be certain that policies and procedures governing how data is created, processed, and distributed are in place throughout the entire data lifecycle, ensuring access to reliable data and complying with ever-changing privacy, ESG, and data safety laws at all times. With StreamSets, FS businesses have centralized guardrails that allow line of business users to explore the art of the possible, with the confidence that data is compliant and secure.

Demographics   Total respondents: 653

Country of residence

44	22	15	28	3	13	9

Role type

Decision makers: 48%	Practitioners: 51%

Role (IT/ LOB)

IT: 46%	Line of Business: 54%

Size of company

# of employees	1,000-4,999	5,000-9,999	10,000+
% of respondents	34%	29%	37%

Gender & age

- **50%** of respondents were female and **50%** male
- **54%** of respondents were aged between 25-44

Methodology and Demographics

The survey was commissioned by StreamSets; it was conducted among 134 financial services decision makers for data tools and practitioners who use data tools in the UK, US, Germany, France, Spain, Italy and Australia.
The interviews were conducted online by Sapio Research using an email invitation and an online survey.

About StreamSets

StreamSets, a Software AG company, eliminates data integration friction in complex hybrid and multi-cloud environments to keep pace with need-it-now business data demands. Our platform lets data teams unlock data—without ceding control—to enable a data-driven enterprise. Resilient and repeatable pipelines deliver analytics-ready data that improve real-time decision-making and reduce the costs and risks associated with data flow across an organization. That's why the largest companies in the world trust StreamSets to power millions of data pipelines for modern analytics, data science, smart applications, and hybrid integration.

streamsets.com