Freeform Dynamics in association with CONFLUENT

Evolution of data architecture

Optimizing data analytics pipelines with shift-left

As expectations for fast, accurate and joined-up business insights continue to rise, how well are current data architectures keeping up, and what needs to change to meet emerging data analytics and AI requirements? A recent survey of 400 data professionals suggests a need to think differently.

Key takeaways

- **Data streaming is well established, but needs and expectations are changing**
  Beyond extreme scale mission-critical applications, the opportunity now is to extend the benefits of streaming to support modern analytics and AI use cases.
- **Shift-left transforms how streaming architectures deliver value**
  Moving processing and governance upstream creates smart, reusable data streams that dramatically reduce duplication while improving quality and consistency.
- **Modern platforms make shifting left practical, opening new possibilities**
  Data streaming platforms (DSPs) enable continuous processing to modernize batch pipelines, reduce storage costs through in-flight aggregation, and accelerate AI initiatives.
- **Strategic shift-left adoption correlates strongly with high performance**
  Organizations that shift-left strategically are nearly three times more likely to achieve superior outcomes from their data architecture.
- **Shift-left is becoming critical as streaming evolves into foundational infrastructure**
  This pattern offers a proven approach to reduce technical debt, improve governance, and accelerate time to value across the enterprise.

Why this research, and why now?

Many organizations have made significant investments in data streaming over recent years, often achieving impressive results for mission-critical applications. From major financial institutions processing millions of transactions to online retailers personalizing experiences in real time, streaming has proven its value at extreme scale.

Despite these successes, most organizations tell us they're struggling to extend streaming benefits more broadly across the enterprise. What works brilliantly for specific high-value use cases often becomes unwieldy when scaled to hundreds of applications and dozens of teams. The result is a growing tension between what's technically possible and what's practically achievable.

This tension is becoming more acute as business demands escalate, particularly in areas like business and operational analytics, and AI initiatives that require fresh, high-quality data streams. Customer expectations for continuously updated insights continue to rise, and the sheer volume of data flowing through organizations makes traditional warehouse-centric batch processing inadequate for many emerging use cases.

At the same time, the economics of data management are under scrutiny. Running disparate data processing infrastructures and stacks, maintaining duplicated pipelines, and managing sprawling technical complexity all carry significant costs. Many are realizing they need a more sustainable approach.

These converging pressures mean it's time for data teams to step back and examine how data streaming architectures and practices need to evolve within the broader information processing and management landscape. That's what prompted this research.

Shift-left architectural patterns can address persistent challenges and enable the creation of robust and highly reusable data products.

Time for a rethink and change of perspective

After years of streaming being primarily the domain of specialist engineering teams, we're witnessing a notable shift in the market. The tools are getting better, the barriers to entry are falling, and perhaps most importantly, the business case for modern streaming approaches is becoming impossible to ignore. What once took months of custom development can now be achieved in days, and deliver greater value.

Shift-left architectural patterns can address persistent challenges by moving processing and controls upstream to reduce duplication, improve quality, accelerate time to insight and enable the creation of robust and highly reusable data products.

But realizing these benefits needs a shift in thinking about how data and value flow through the organization. That's the perspective we explore in this research.

About the research and report structure

Methodology

This report discusses the findings of a survey-based research study completed in Q2 2025. The study was designed and executed by Freeform Dynamics. Data was collected from 400 participants (a mix of data leaders, specialist practitioners and cross-domain executives) during interactive interviews conducted via phone or web conference.

The sample was spread across the USA, EMEA and APAC (7 countries in all) and a cross section of industry sectors. 37% of organizations have more than 10,000 employees, with the remainder between 1,000 and 10,000 staff.

The work was sponsored by Confluent.

Intended audience and readership considerations

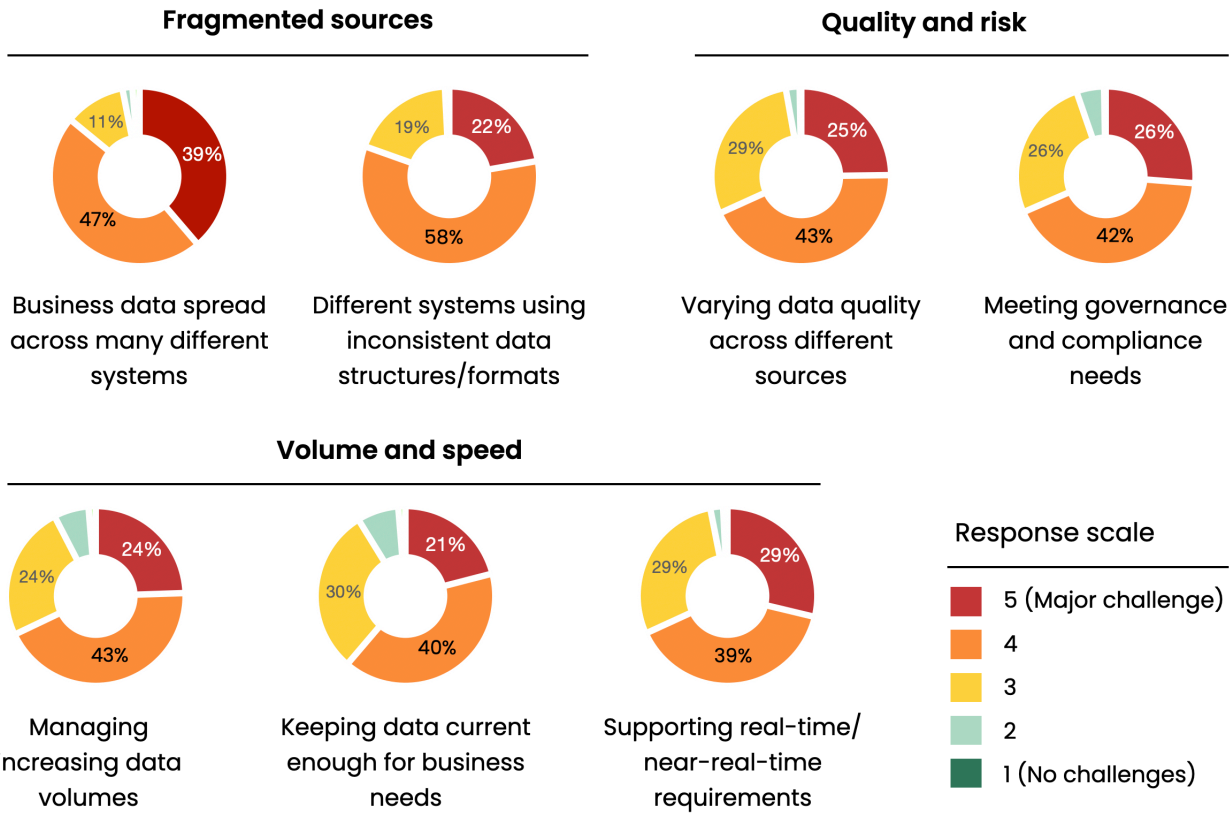
When writing this report, we had both senior leaders and experienced practitioners in mind - anyone, in fact, with an interest in how to meet evolving needs in the real-time data arena. This clearly presented some challenges with regard to the level of subject matter knowledge we should assume. On the whole, we have written the report to be accessible to stakeholders and relative newcomers who have a reasonable familiarity with the territory, but a limited or patchy understanding of the specifics in some areas. If you are more experienced with the technical detail, please therefore forgive us for abstracting or simplifying some elements of the discussion. Feel free to skip over these as you wish.

Table of Contents

Data quality and fragmentation remain persistent challenges

A good place to start our conversation is to reconsider a familiar range of data-related challenges in light of today's business and technology environment. Fragmentation and inconsistency of data sources, along with associated quality and risk issues, have dogged data teams for decades. And these remain issues, not least because of growing expectations among business users. No longer are they satisfied with retrospective analysis and reporting capabilities - the demand today is for ever richer and more holistic real-time business insights. Even with continuous improvement, it has therefore been hard to get ahead of existing problems, let alone address new ones.

How much of a challenge are the following?



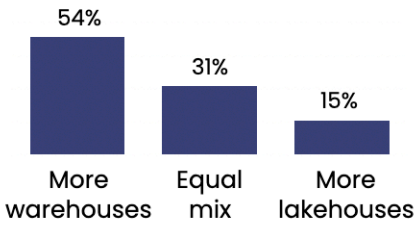
The continued role of centralization

It was in response to the challenge of data fragmentation that data warehousing was born around three decades ago, and the idea of pulling data together from disparate sources into a central repository for analysis purposes is still very much relevant today. This is despite efforts over the years by data and application teams to consolidate data stores and leverage technology that allows transaction processing and analytics workloads to coexist on the same platform. The truth is that while such tactical moves have been useful, if you zoom out to the bigger picture we've seen an explosion of data storage options, both on-premises and cloud based, and exploiting this choice has led to even more fragmentation. Central repositories are therefore still a requirement, though we've seen some big changes in the underlying enabling data architectures.

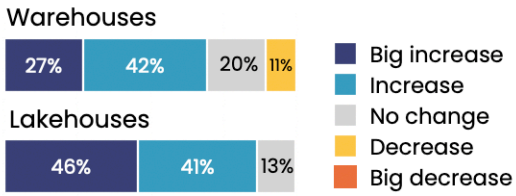
A clear trend towards data lakehouses and data streaming

In terms of repositories, data lakes burst onto the scene in the 2010s in response to the need to collate and analyze very high volumes of raw, unstructured data (a.k.a 'big data'). These were based on the highly flexible 'schema on read' model, and existed alongside traditional data warehouses that were still the solution of choice for the more prescriptive analysis and reporting needs. Since then, data lakes have evolved into data lakehouses, which notionally layer warehouse style structured schemas onto the underlying flexible data lake architecture, though with a lot more versatility. While warehouses still dominate, the direction of travel is clearly towards lakehouses, which are arguably better suited to meet high volume, real-time needs. Not surprisingly, we see a parallel increase in emphasis on stream-based pipelines, though batch remains important.

Current mix of repositories

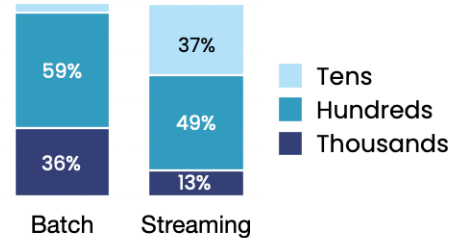


Repository dynamics (next 2 years)

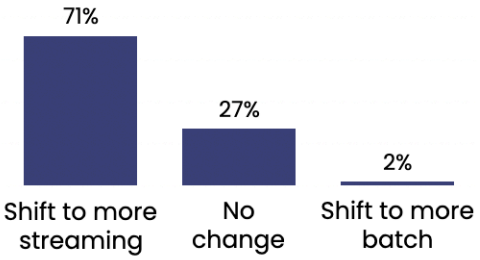


Direction of travel (more lakehouses, more streaming)

Current number of pipelines

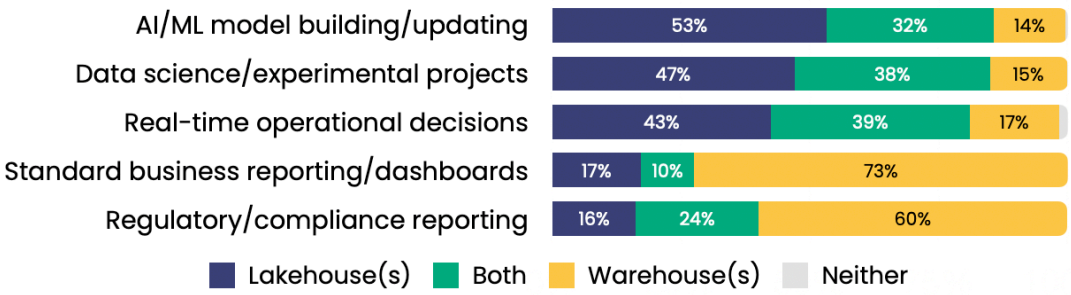


Pipeline dynamics (next 2 years)



On a specific point, today there's a lot of overlap between warehouse and lakehouse functionality so the difference is as much about design philosophy as is it is about technology - i.e. batch-driven data pooling and analysis versus continuous updates and real-time insights. This is evident when we look at activity through a use case lens.

For each of the following use cases, which do you currently use?



Organizations still suffer from a range of stubborn systems issues

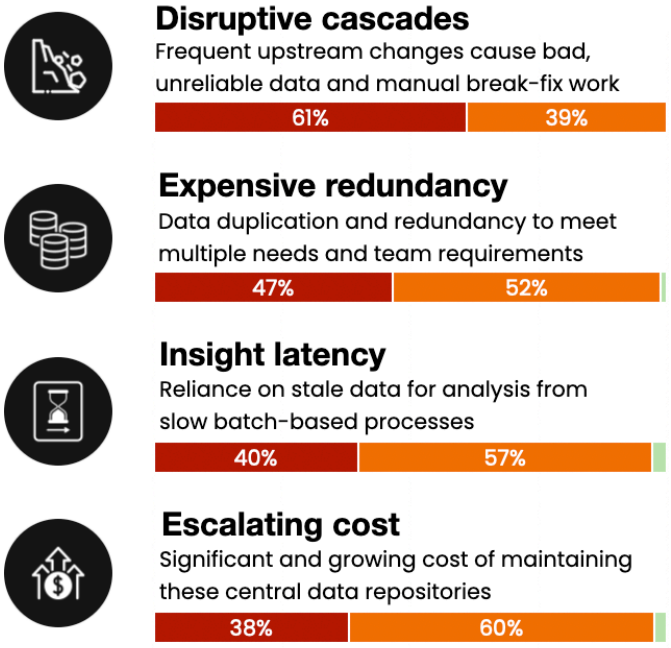
The greater emphasis on more flexible, performant and scalable architectures we've just been discussing bodes well for the future.

However, when we critically assess the readiness of current systems and processes, we still see a range of stubborn challenges.

Common issues

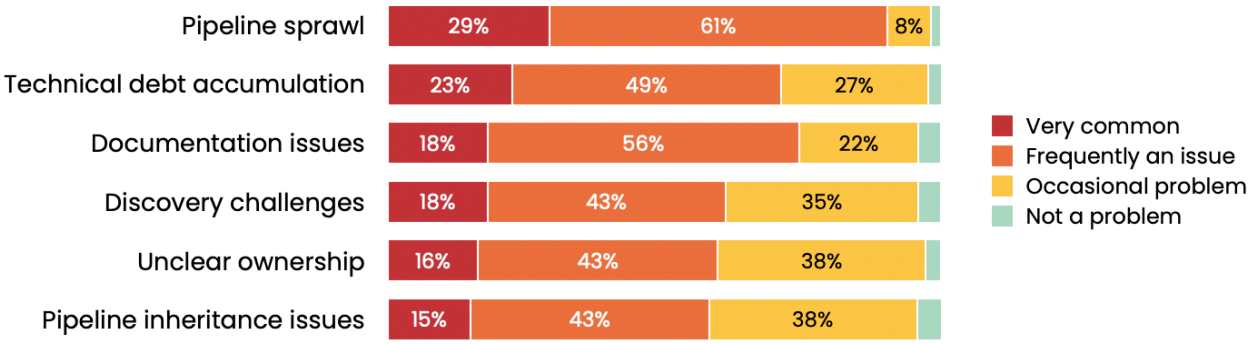
How much do these challenges apply to your current warehouse and lakehouse environments?

- Fully applies
- Partially applies
- No issues



We don't have space to embark on a full root cause analysis here, but these issues, at least in part, are likely to be down to a continued reliance on older technologies and methods. We can also infer a range of governance shortfalls, including a lack of coordination between projects and teams with regard to sharing, reuse and the overall management of complexity. This is corroborated by the issues reported in relation to data pipelines.

How common are these data pipeline related issues in your organization?



When considering this situation, it's tempting to think in terms of simply tightening things up and introducing new practices to define ownership, ensure the production of relevant documentation and monitor and track assets more effectively. However, when we consider recent technology advances, particularly in the area of data streaming, there's a strong argument for a more fundamental rethink of how pipelines are designed and built.

Taking stream-based pipelines to the next level by shifting left

Having discussed data architecture trends and the challenges faced by data teams today, let's drill down and consider options to both drive operational improvements and respond to evolving business needs.

An obvious initial action here is to revisit existing systems with a view to tightening them up from a compliance, control and efficiency perspective, perhaps updating existing technology stacks along the way.

However, whether it's modernizing aging ETL and warehousing infrastructure, or innovating around new requirements in areas such as AI, IoT and digital customer engagement, it's important to look at alternatives to established approaches, especially those that can potentially enable a step change in performance, agility and efficiency.

A key development worth calling out here is the introduction of shift-left processing into data streaming platforms (DSPs). This builds on the robust, high-scale, high-volume capabilities of familiar streaming frameworks such as Apache Kafka, enabling complex data manipulation to take place in-flight.

In essence, this model is about moving data processing upstream in the data pipeline (or 'shifting left'), allowing cleansing, filtering, transformation, governance and other operations to take place much closer to where data originates. This in turn allows the creation of smart, reusable streams that serve as governed data assets.

Beyond addressing many traditional challenges, this approach opens up a whole range of new possibilities.

It's not just about better ways to move and process data, it's about opening opportunities to drive substantially higher business value.

Opportunity to rethink data movement patterns

While real-time analytics dashboards and operational monitoring remain core streaming use cases, DSP-enabled shift-left capabilities are expanding what's possible. Organizations are discovering that continuous, in-flight processing offers compelling advantages even when real-time isn't the primary requirement.

Consider the opportunity to modernize aging batch pipelines. Rather than simply replacing like-for-like, you can upgrade to a continuous, hands-off delivery model that processes data as it arrives. This isn't just about speed - it's about efficiency, reduced overhead, and better resource utilization.

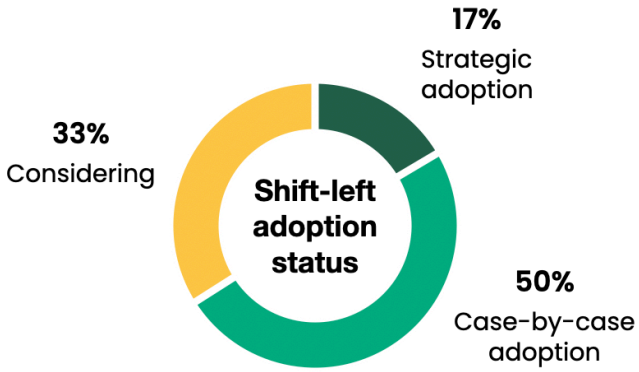
We're also seeing innovative approaches to data lake/lakehouse population. Instead of traditional ELT-style batch feeds dumping raw data into bronze tier storage, you can use stream processing to aggregate, filter, and enrich data in-flight. For industrial IoT or clickstream scenarios, this means what lands in your repository is already silver-tier data, dramatically condensed, pre-validated, and ready for consumption. The impact on cost efficiency can be substantial.

Use cases like theses illustrate how shift-left isn't just solving old problems - it's enabling better ways to build data architectures, which lead to a whole range of tangible benefits.

Shifting left to improve data quality and foster collaboration

During the study, we found that data professionals varied in their familiarity with a term that's recently emerged in data and analytics circles - 'shift-left'. A few hadn't come across the term at all, some knew it in the context of software development, and others clearly had a good understanding of what it meant in the context of data pipelines. Once defined, however, no one had trouble with the general idea; indeed, most told us they had already started to adopt shift-left to one degree or another.

To what degree have you adopted a shift-left approach?

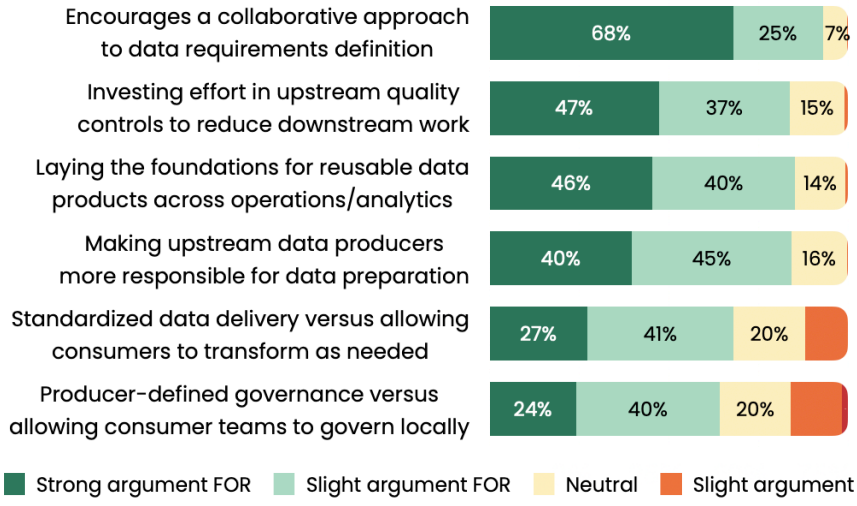


Definition provided to respondents

In the context of data pipelines, 'shift-left' refers to the movement of processing and governance controls upstream, i.e closer to where the data originates.

With a few exceptions, the data professionals we spoke with were generally very positive about the benefits of shift-left. The only real push back came from a few who seemed to have concerns about shifting left potentially restricting the freedom of downstream users.

From your own professional perspective, how much do you see the following as arguments FOR or AGAINST implementing shift-left?



Research Note

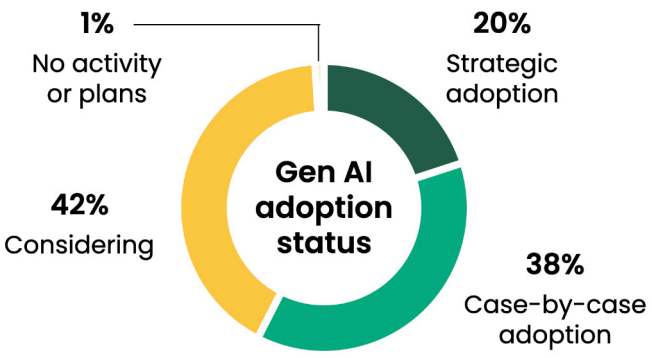
The idea of this question was to flush out concerns and objections as well as to capture the positives, hence the way this question and the options are phrased.

On a particular point, the strong emphasis on collaboration gels with what we see when shift-left is used in a software development context. Even if there's initial reticence, most come to realize that beyond better overall productivity, encouraging more cross-team communication and understanding removes many of the frustrations of working in a fast-paced environment, making everyone's job both easier and more satisfying.

Shift-left as an accelerator of success with generative AI

In addition to the benefits already mentioned, we specifically explored the intersection of shift-left with an area that's receiving a great deal of attention at the moment. Well over half of the respondents in our study said they were active with generative AI at the moment, particularly with AI applications that leveraged company data.

To what degree have you adopted generative AI to leverage company data?

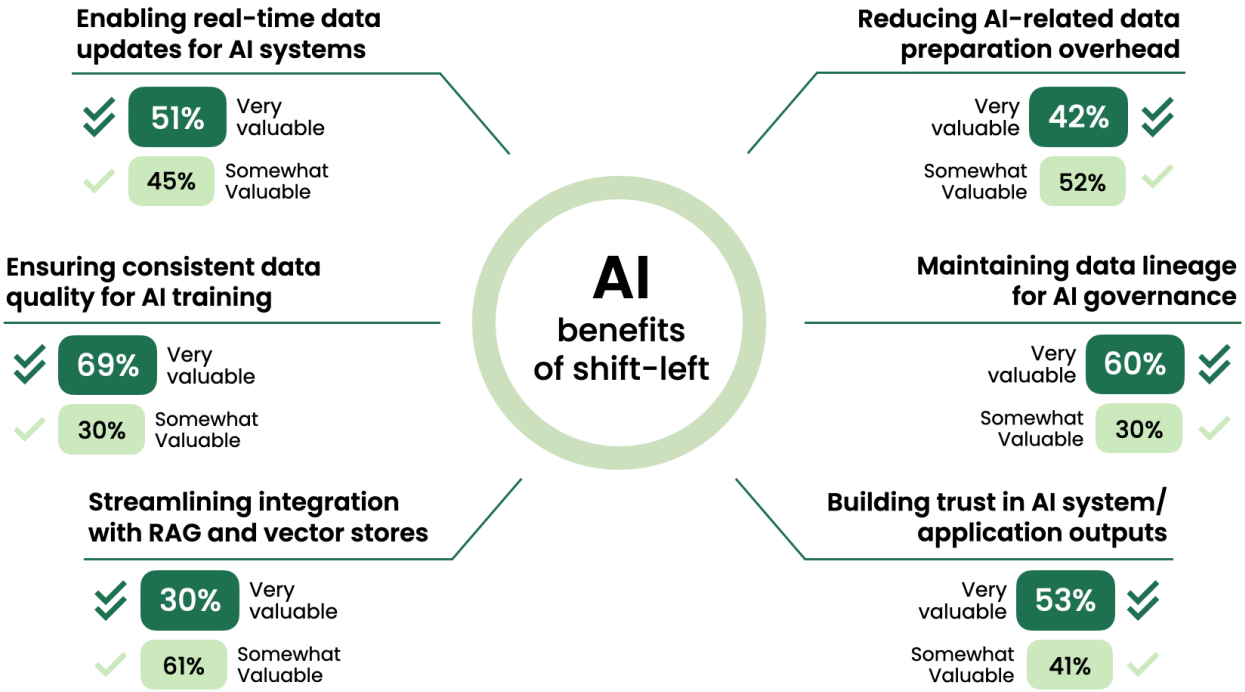


Research Note

This question relates to the use of generative AI to leverage company data, which typically means AI capabilities embedded in enterprise applications and/or customer developed solutions (as opposed to general purpose AI chatbots).

When asked whether shift-left could help with their initiatives in this area, quite a few potential benefits were acknowledged, ranging from ensuring AI applications operate with the highest quality and most up-to-date data, through to tracking lineage and maintaining good governance to assure trust in the output from AI systems.

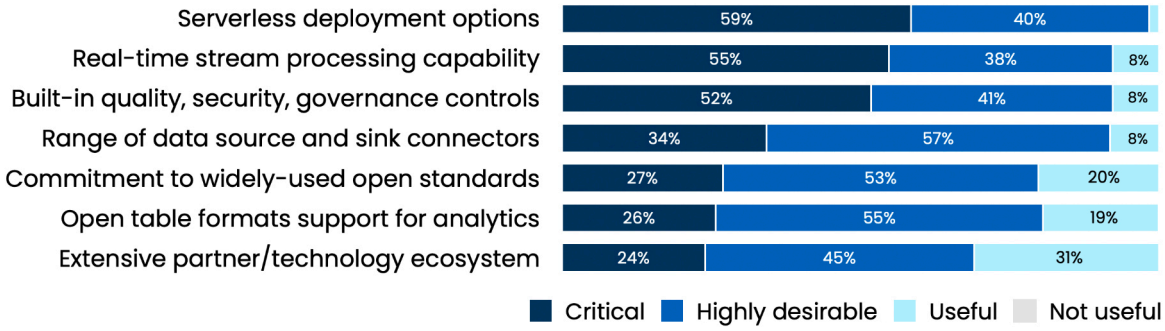
How valuable do you consider the following shift-left capabilities in supporting AI initiatives?



Technology requirements for modern shift-left streaming

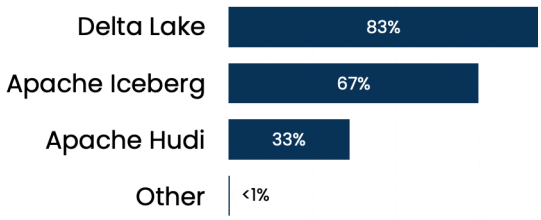
If you are currently relying on a native Kafka framework (or similar foundational technology), further investment is likely to be required to implement shift-left. One route you could take is to manually integrate a stream processing engine and do the integration work on a DIY basis. The downside of this is the time and effort needed, plus there's an obvious ongoing support and maintenance burden to consider. If you wanted a simpler and more accessible option then a modern data streaming platform with an embedded stream processing engine might be a better way to go. Beyond enabling shift-left, such a solution would also offer a range of other benefits, most of which were acknowledged in our research.

When considering a data streaming platform, how important do you see the following?

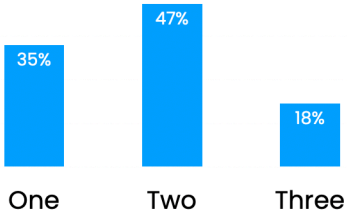


Picking up on the mention of open table formats, considered critical or highly desirable by over 4 in 5 of respondents, these are particularly relevant if stream-based pipelines are used to populate and update modern data lakehouses. Formats such as the ones shown below are increasingly used to publish streams directly into such environments. ACID transaction support ensures availability to the latest up-to-date data, with 'time travel' functionality for accessing historical versions of data and enabling rollbacks when needed. Support for metadata management, including rich table-level and field-level metadata, leverages the governance and data lineage tracking capabilities of DSPs operating in a shift-left manner.

Which of these open table formats are you currently using or considering?



How many open table formats are your currently using or considering?

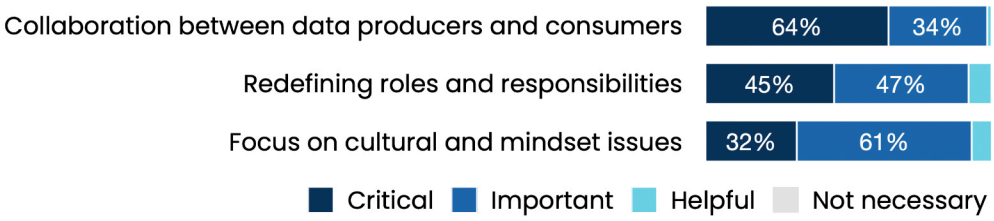


But it's not just about technology enablers. To get the full benefit of shift-left adoption, it's also necessary to address a number of other key requirements.

Beyond the technology to people and cultural imperatives

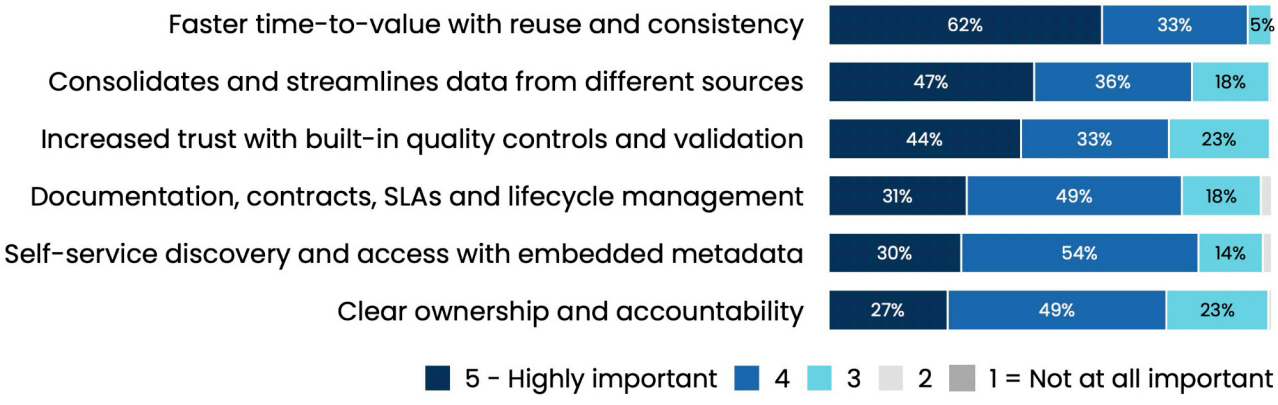
While technology such as modern data streaming platforms and the latest data lakehouse capabilities are critical to implement shift-left successfully at an enterprise level, just like any other strategic transformation, it's also important to deal with people-related impacts and requirements for change.

How important are the following for success with shift-left in a data architecture context?



As we can see from the above chart, one of the most important considerations is how teams and individuals collaborate, particularly across the traditional producer-consumer divide, If you are going to move processing and controls upstream, for example, data engineers creating smart, governed streams or pipelines need to speak with business analysts and application owners working downstream to synchronize requirements. This may seem like it's constraining the separation of concerns, but it's no different to teams working in any other service oriented environment or with a microservice architecture. In simple terms, it's about figuring out where best to draw the line between data-centric processing and controls that are largely application independent, and logic that's specific to the downstream application or service. It's a short step from here to adopting a data product approach to packaging data assets.

How important are these benefits and capabilities from adopting a data product approach?



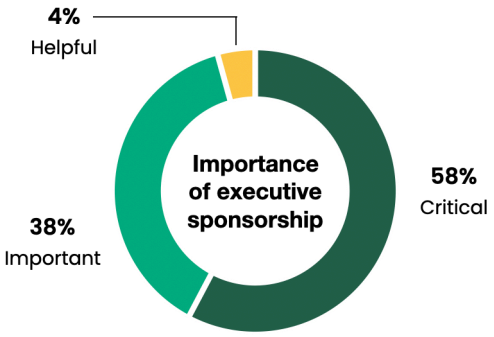
While it's unlikely to make sense to define all stream-based pipelines as full data products, e.g. those that purely serve the needs of a specific use case, this is a useful discipline to consider as you move to a more enterprise level approach to sharing and reuse. And the good news is that a modern DSP will provide most if not all of what's needed to do this.

The compelling case for shift-left investment

First-generation streaming adoption often began as grassroots engineering-led initiatives. The shift-left approach, however, with its emphasis on smart, governed pipelines, generally requires broader organizational commitment. It's therefore not surprising to see executive sponsorship highlighted as important e.g. to secure budget, release resources, and provide a mandate for transformational change.

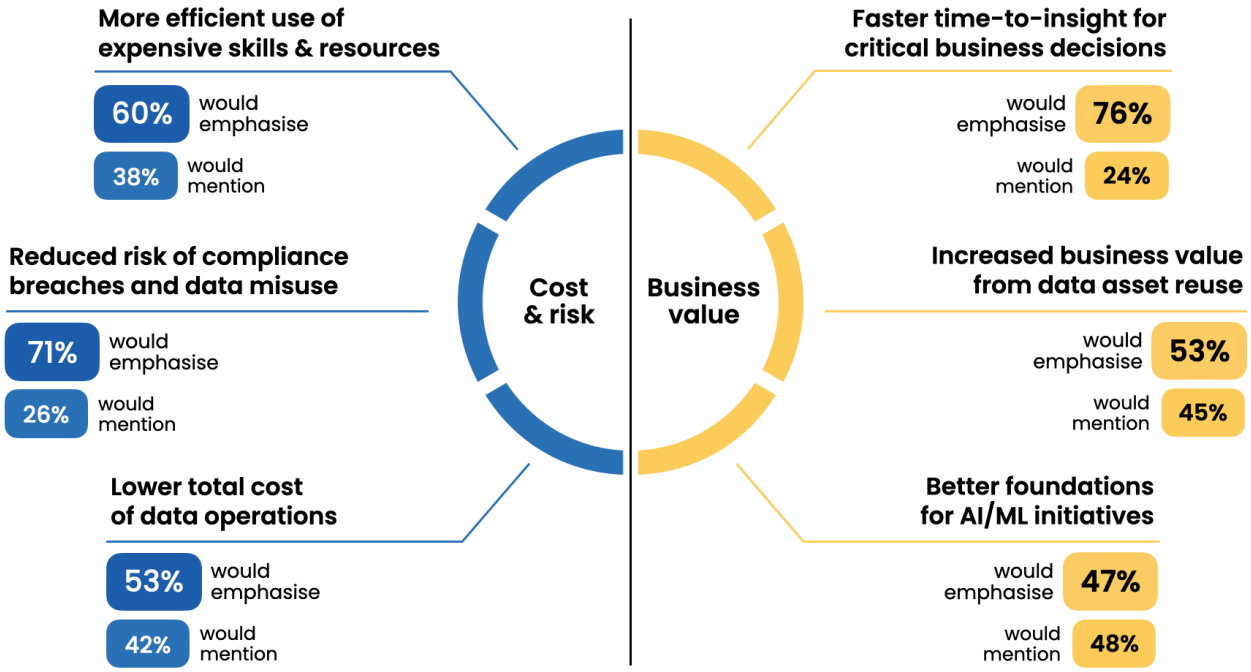
How important is clear executive sponsorship to success with shift-left

(No one responded "Unhelpful")



As part of this, technical leaders must make a business case that translates the practical benefits discussed earlier into business value statements that resonate with a senior business audience. The chart below provides some insight into how our respondents would frame shift-left benefits in way that will resonate with board-level executives.

If you had to justify investing in the shift-left approach to a board-level executive, how much would you emphasize the following potential benefits?

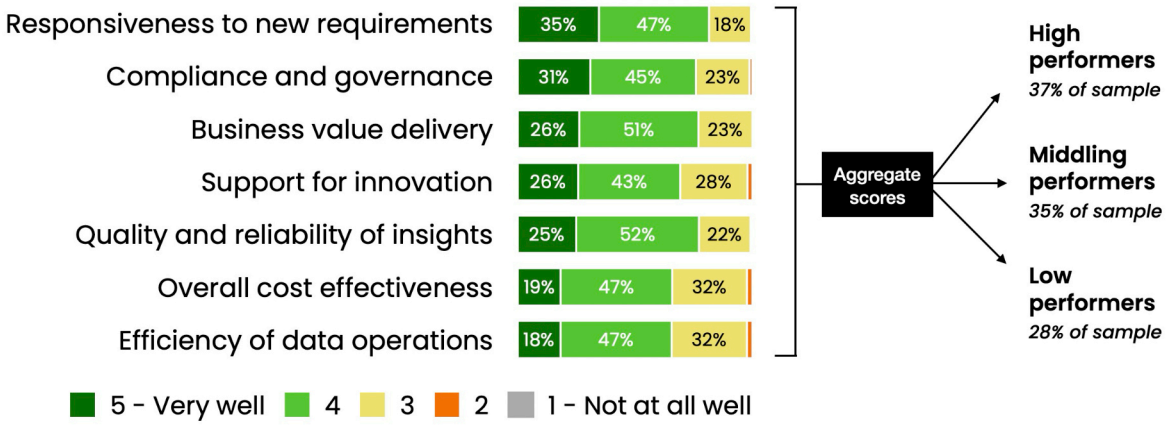


Focusing on both business value and cost/risk benefits makes absolute sense, but did anything come out of the research that corroborates the link between shift-left and successful delivery? The short answer is 'yes' - read on to see the evidence.

Shifting left and positive business outcomes go hand-in-hand

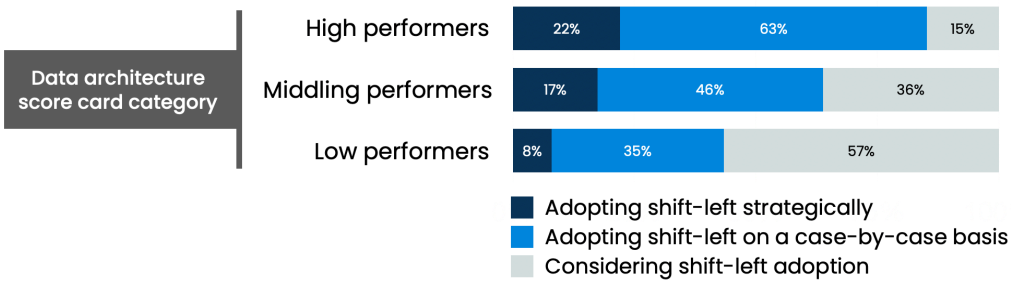
As part of the study, we included a question in 'scorecard' format that captured how well each respondent's data architecture was meeting needs and expectations across a range of key indicators. Scores were then aggregated and averaged for each individual respondent, allowing us to place them into one of three performance groups - High, Middling and Low - as shown in the graphic below.

How well does your current data architecture meet needs and expectations in the following areas?



From here we were able to cross tabulate between performance categories and the degree of shift-left adoption (strategic, case-by-case or just considering), and when we did this, the resulting correlation that leapt out was clear and very striking.

High performers are significantly more likely to be adopting shift-left

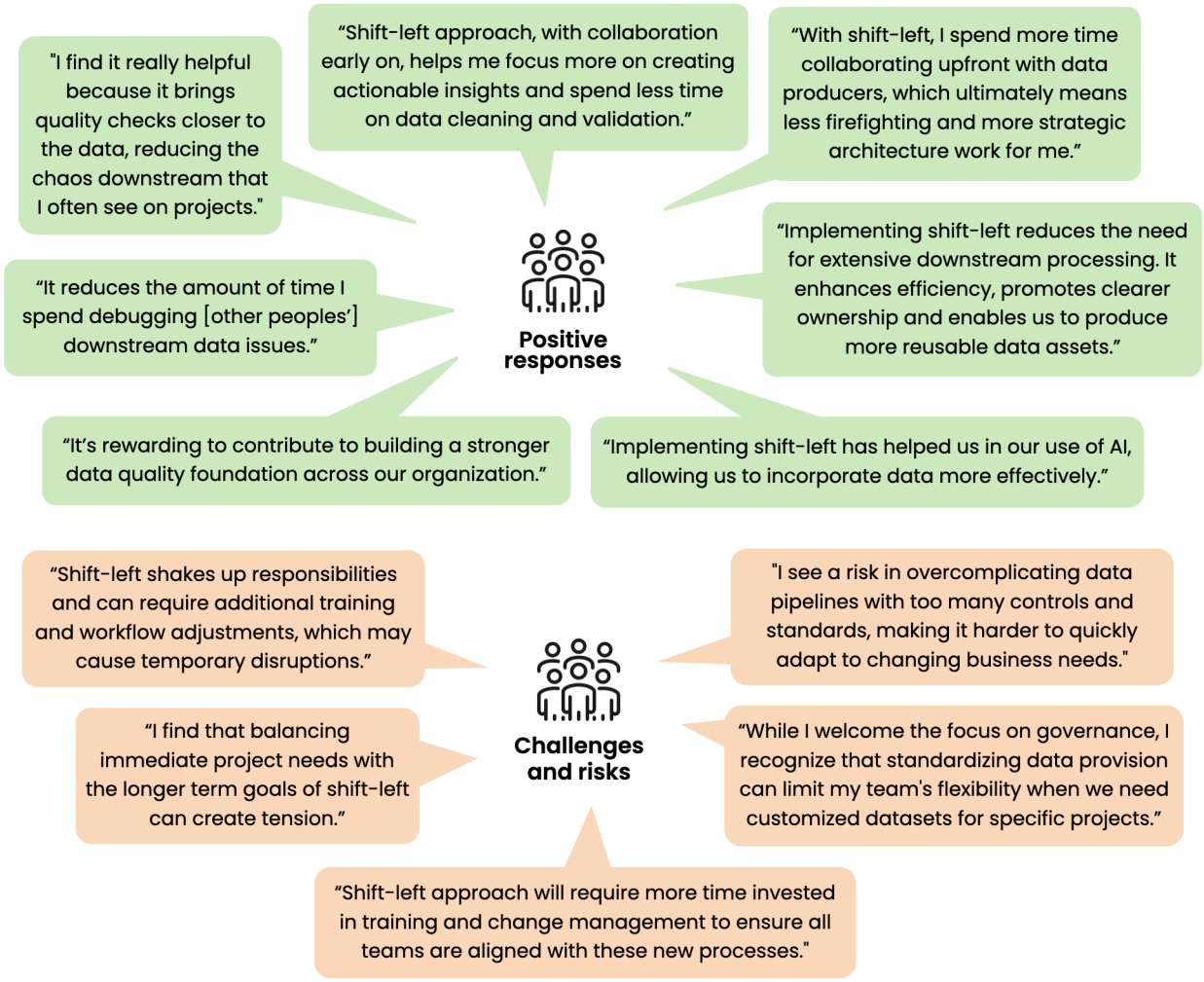


As a word of warning, correlation is not the same as causation, i.e. just because high performers are nearly three times more likely to be using shift-left strategically compared to lower performers, we cannot directly infer that shift-left is a performance driver. What we can say is that shift-left seems to go hand in hand with modern, effective data architectures, and with all of the other research findings we've reviewed, there would appear to be a strong case for making sure shifting left is on your agenda.

Final thoughts

It's easy to forget when looking at the kind of statistics we've been discussing that behind every number is input from real humans based on their own unique experiences and opinions. Let's therefore finish by looking at some of the comments we received during our interviews. Here's a selection to provide a flavor of personal perspectives on shift-left.

How does the use of shift-left affect you in your current role?



Sentiment analysis revealed that just over half of the comments we received were positive, a little over a third were neutral or balanced, with the rest reflecting a more skeptical view.

This is a reminder that the transition to shift-left can have great benefits but is not trivial, especially if you are already significantly invested in data streaming. Our advice for most organizations is therefore not to rush things. A pragmatic approach worth considering is to try the latest DSP offerings, which can typically be provisioned with little up front cost via a fully-managed or self-managed model, and take it from there.

In the meantime, we hope this report has provided some useful insights to help you with your next steps.

About

About Freeform Dynamics

Freeform Dynamics is an IT industry analyst firm. Through our research and insights, we help busy IT and business professionals get up to speed on the latest technology developments and make better-informed investment decisions.

For more information and access to other research, please visit www.freeformdynamics.com.

About Confluent

Confluent is creating the foundational platform for data-in-motion. With Confluent, organizations can harness the full power of continuously flowing data to innovate and win in the modern digital world.

For more information, please visit www.confluent.io/shift-left.

Research note

When viewing the charts in this report, please note that percentages may not always sum to exactly 100% due to rounding to the nearest whole number. This is a standard presentation convention and does not affect the validity of the data presented.

Terms of Service