ryft

The State of Apache Iceberg in the Enterprise 2026

Benchmarking adoption, production usage, and operational realities of Apache Iceberg at scale

Executive Summary

Apache Iceberg has moved from early adoption into broad production use across enterprise lakehouse environments.

Respondents no longer see Iceberg as an experiment or niche format; they rely on it as a core layer of their enterprise lakehouse architecture. Teams are adopting Iceberg to keep pace with rapid data growth, unlock new analytical, AI, and ML use cases, and regain control over performance and reliability as data volumes scale.

The research shows a gap between Iceberg's architectural promise and the day-to-day work required to operate it reliably at scale. Teams report high satisfaction and say Iceberg meets expectations, yet most organizations still depend on custom, homegrown solutions to handle critical operational requirements. This DIY approach is creating friction, risk, and uneven outcomes across teams.

Overall, respondents express confidence in Iceberg's design and are expanding usage, while recognizing that long-term success depends on operational maturity—not initial adoption.

What this indicates

Most organizations have already decided to adopt Iceberg; maintaining consistent operations now limits further progress. Teams are independently building similar tooling, signaling a gap in standardized operations. As environments scale, the primary risks shift from technology choice to governance, reliability, and cost control.

Ryft perspective

- Over the next 12–24 months, differentiation will shift from "Who adopted Iceberg?" to "Who can run it continuously and safely across engines, teams, and environments?"
- Manage Iceberg as a production platform: set reliability targets, automate data management, and standardize governance across the organization.

Survey Methodology

This research draws on a targeted survey of enterprise data and IT leaders who operate Apache Iceberg in production lakehouse environments. The goal was to capture real-world experience with Apache Iceberg at a meaningful scale, focusing on operational maturity rather than early-stage experimentation.

The survey was **conducted in January 2026** and **commissioned by Ryft**, with data collection and analysis carried out by **TrendCandy, an independent market research firm**. TrendCandy recruited 252 survey respondents through a private B2B IT paid panel, from the USA, who are responsible for data platform architecture, operations, and reliability.

To ensure relevance and credibility:

- Participation was limited to organizations **running Apache Iceberg in production**, not pilots or proofs of concept.
- Respondents were **hands-on owners or decision-makers**, including executives, VPs, directors, and managers.
- The survey examined **real operating environments**, covering scale, workloads, and day-to-day operational practices.

The research focused on four core areas:

- **Apache Iceberg operational maturity**
- **Operational complexity and maintenance effort**
- **Cost dynamics as data volumes scale**
- **Reliability and performance under production load**

The study reflects the perspectives of teams managing large, business-critical data platforms. Results are reported with a **±5 % margin of error at a 95% confidence level**, providing a statistically reliable view of current enterprise Iceberg adoption and operational realities.

Sampling consideration

This research reflects the perspectives of enterprise data teams actively operating Apache Iceberg in production environments. As with any targeted survey, results may over-represent organizations that are more advanced in their Iceberg adoption, more cloud-native, or more data-intensive than the broader market.

Readers should interpret these findings as representing current production Iceberg operators, not organizations that are still evaluating or experimenting with Iceberg. Within this defined population, the results provide a statistically reliable view of operational practices, challenges, and priorities.

About the Survey Respondents

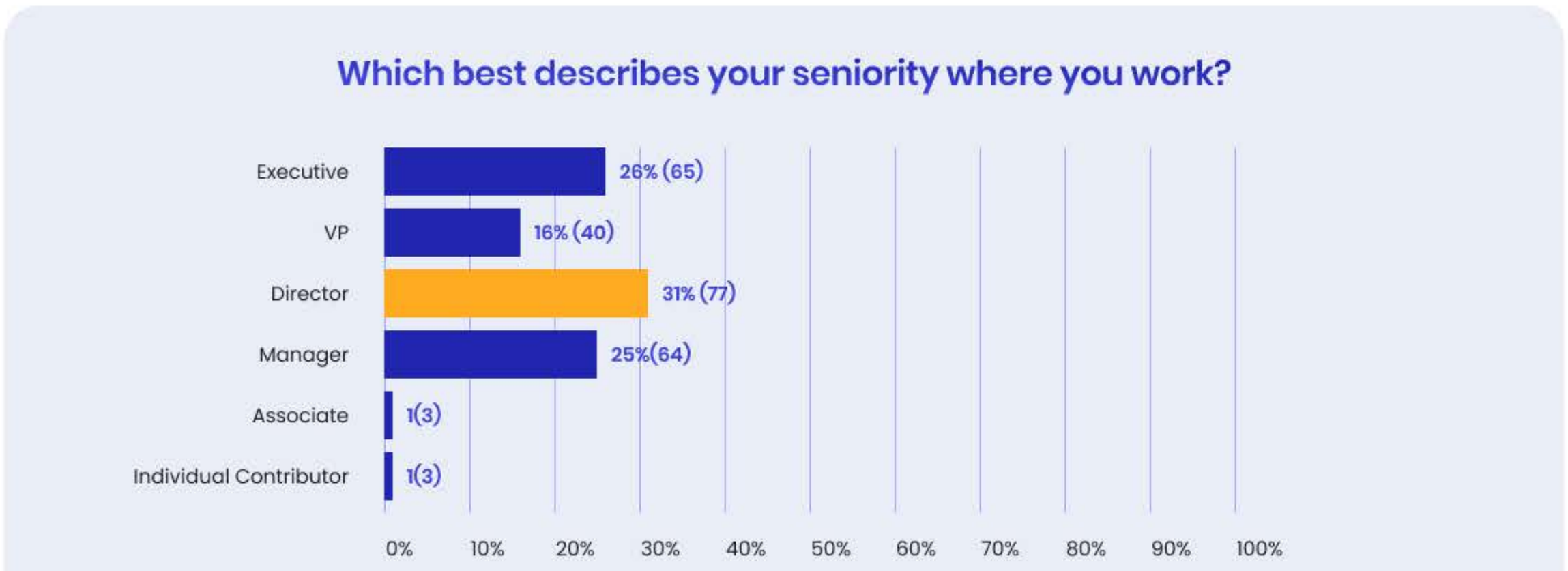
The survey reflects the perspectives of **senior IT and data leaders** responsible for production data platforms at enterprise organizations. Respondents hold **executive, VP, director, and manager-level roles** and are directly accountable for the reliability, performance, cost, and scalability of their data infrastructure. Respondents represent organizations with growing data estates and production workloads where reliability, cost, and governance are operational priorities.

Why respondent profile matters

These respondents reflect the **current mainstream of Apache Iceberg adoption**: teams that have moved beyond pilots and are using Iceberg as a core component of their data platform. Their experiences provide a realistic view of both the benefits Iceberg delivers at scale and the operational challenges that emerge as adoption deepens.

Seniority

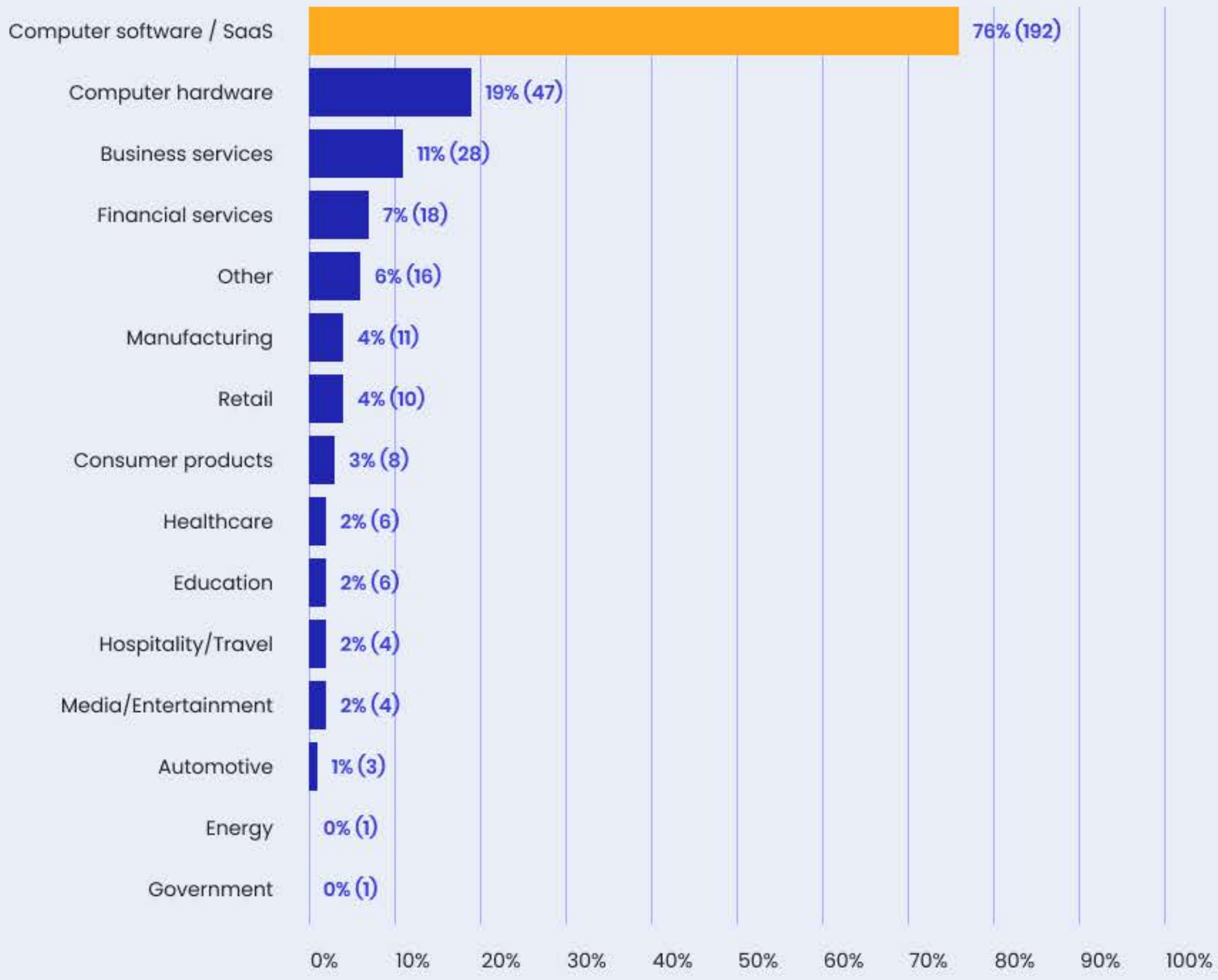
We surveyed executives, VPs, directors, and managers.



Industry

Participants represent a cross-section of **data-intensive enterprise sectors**, with the majority from the computer software/SaaS industry.

Which describes the industry or industries your company operates in? (Select all that apply)



Employee size

91% of respondents work for companies with 100-10,000 employees.

Approximately how many people does your company employ?

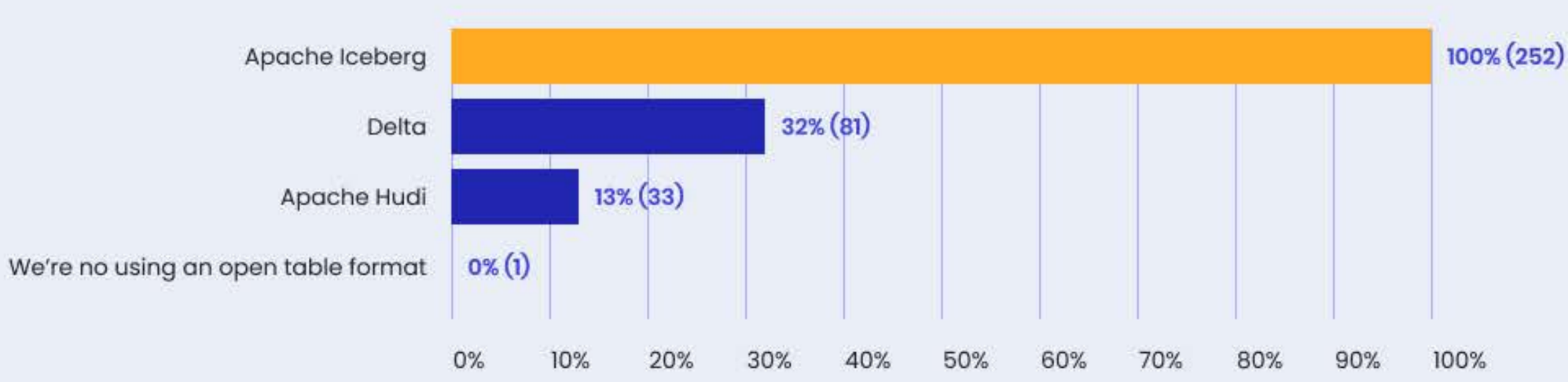
Technology environment

Respondents are actively operating modern lakehouse architectures in production and have adopted the following technology tooling:

Open table formats

Because the survey screened for production Iceberg usage, Iceberg is 100% by definition; Delta and Hudi appear as secondary formats in some environments.

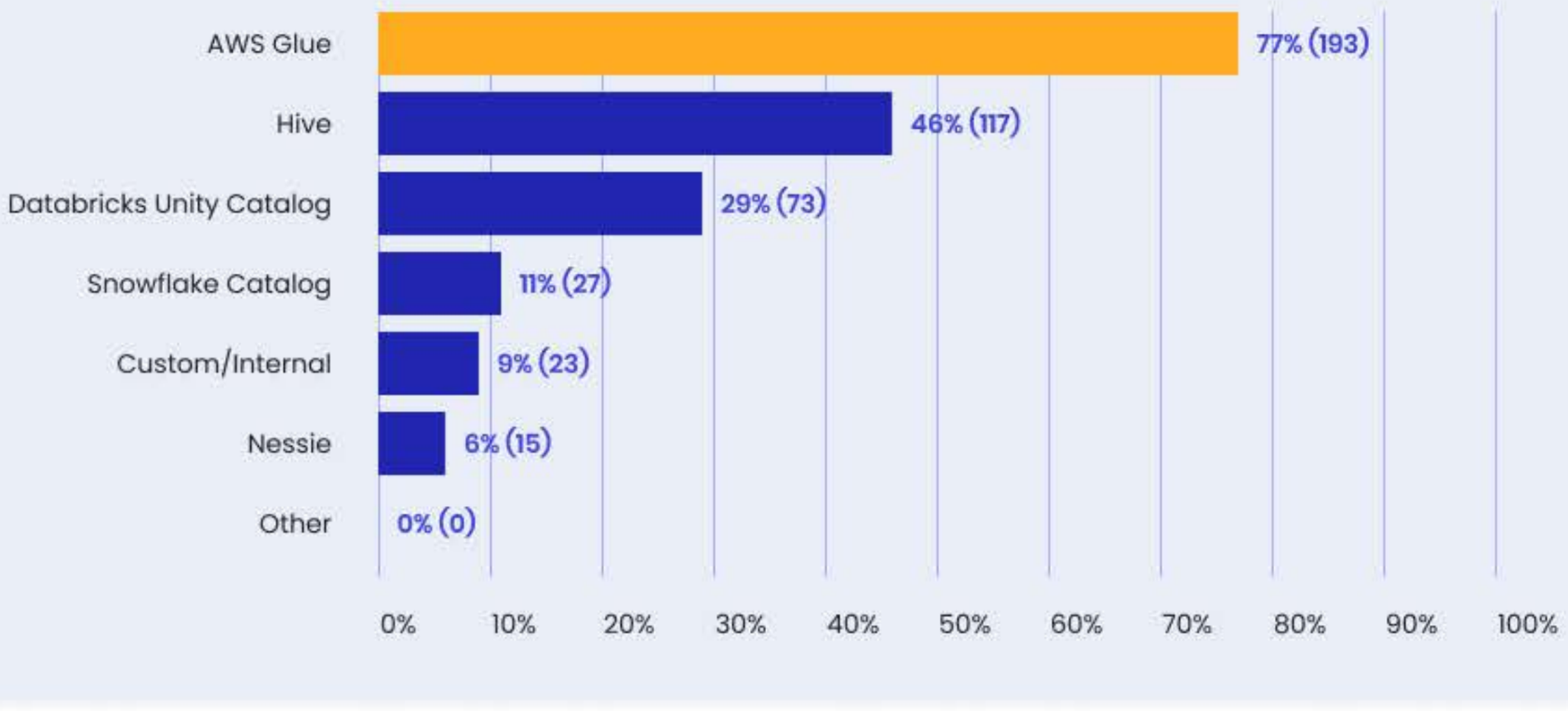
Which of the following open table formats are you using in production?



Apache Iceberg data

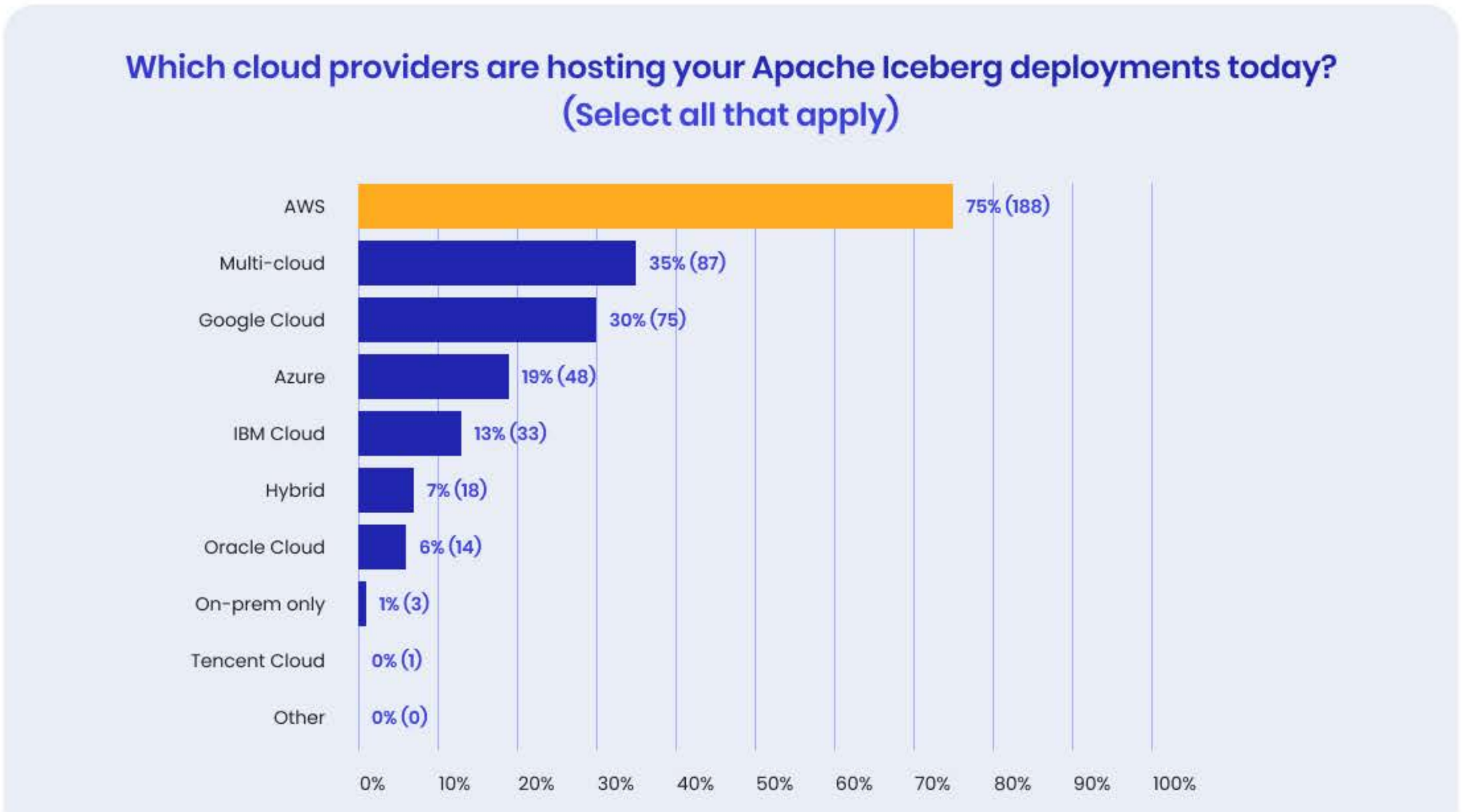
AWS Glue is the most commonly used catalog, followed by Hive Metastore and Databricks Unity Catalog.

Which catalogs are most used for Apache Iceberg in your organization today? (Select all that apply)



Cloud infrastructure

The majority are running Iceberg on AWS, with meaningful representation from multi-cloud deployments, as well as Google Cloud and Azure environments.

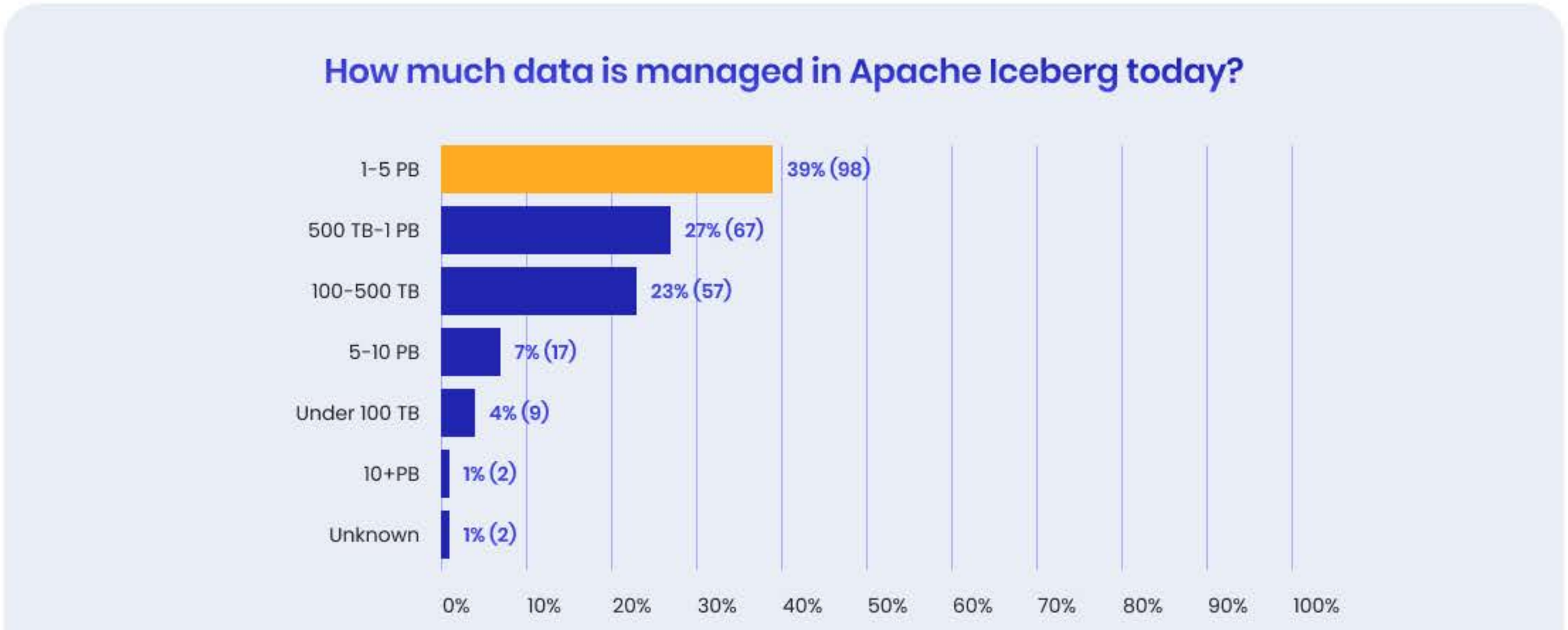


Scale and workload characteristic

Survey participants are not experimenting on a small scale. Most are operating Iceberg in **production-grade environments** with large (TB to PB) datasets across thousands of Iceberg tables, supporting **business-critical analytics, and running Iceberg on AI and ML workloads.**

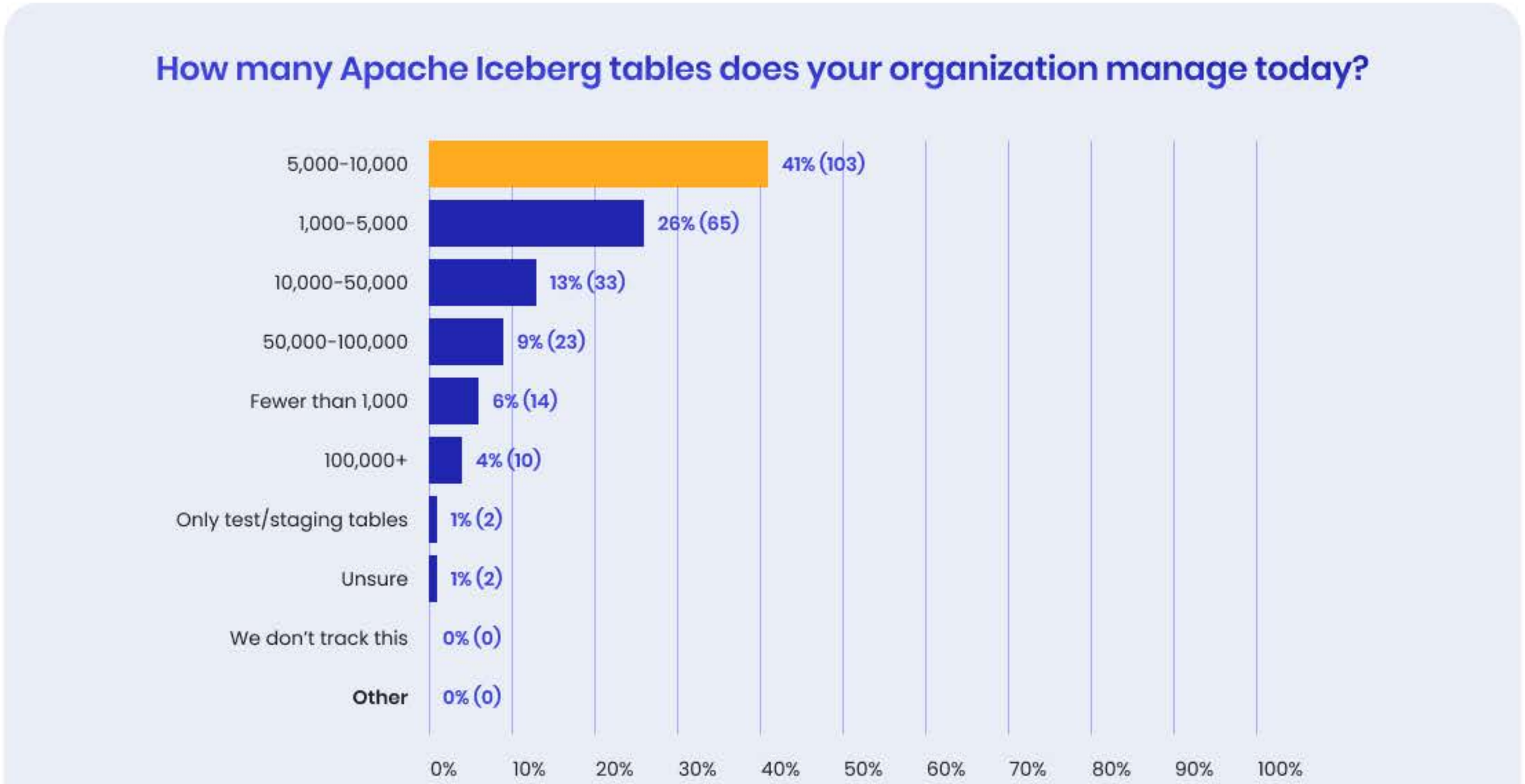
Data size

89% are managing between 100 terabytes and 5 petabytes of data.



Iceberg table count

89% manage between 1,000 and 100,000 Iceberg tables.

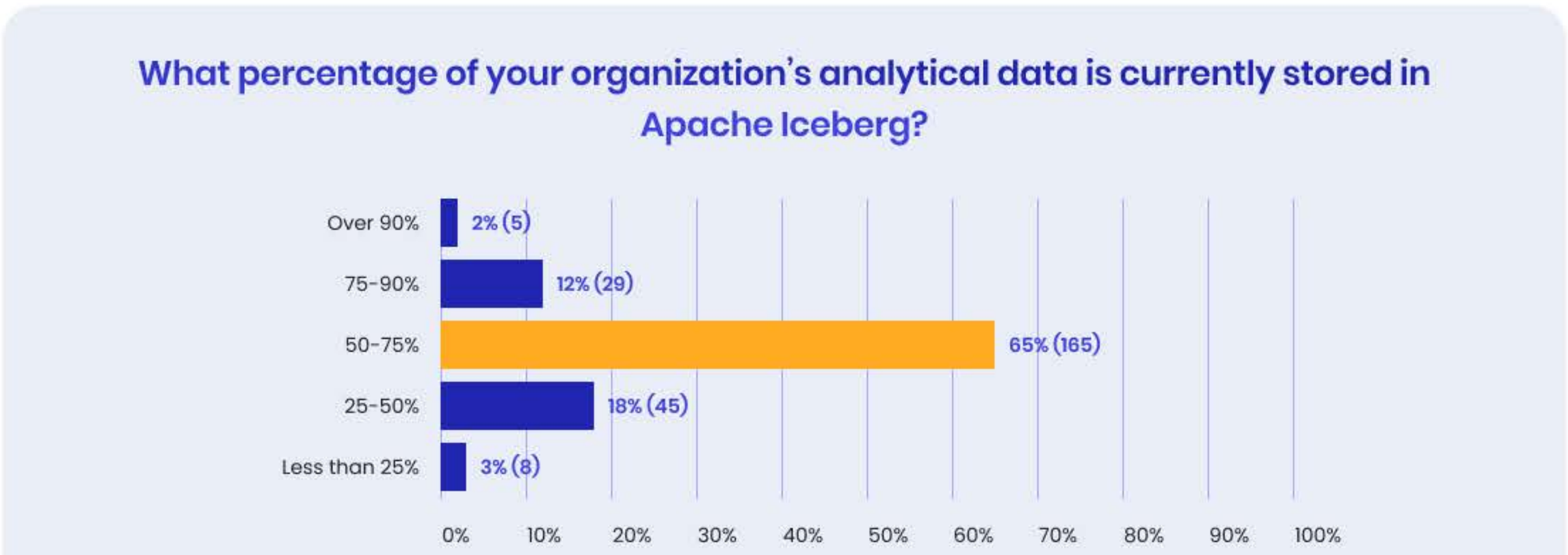


Iceberg workload type

58% are using Iceberg for business-critical analytics.

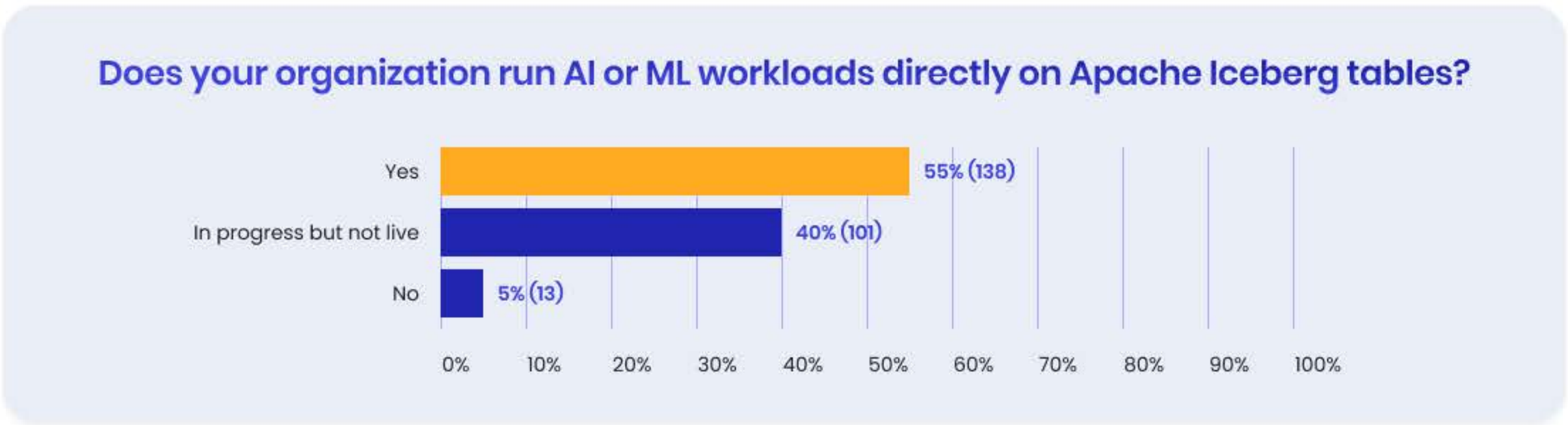
Analytical data usage

79% have more than half of their analytical data stored in Iceberg.



Active AI and ML workloads

The data shows that 95% of respondents report using or planning to use Iceberg to run AI or ML workloads.



What this indicates

Once Iceberg environments reach thousands of tables, manual processes no longer hold up. Planned AI and ML usage further increases load, data retention demands, and scrutiny of access and compliance.

Ryft perspective

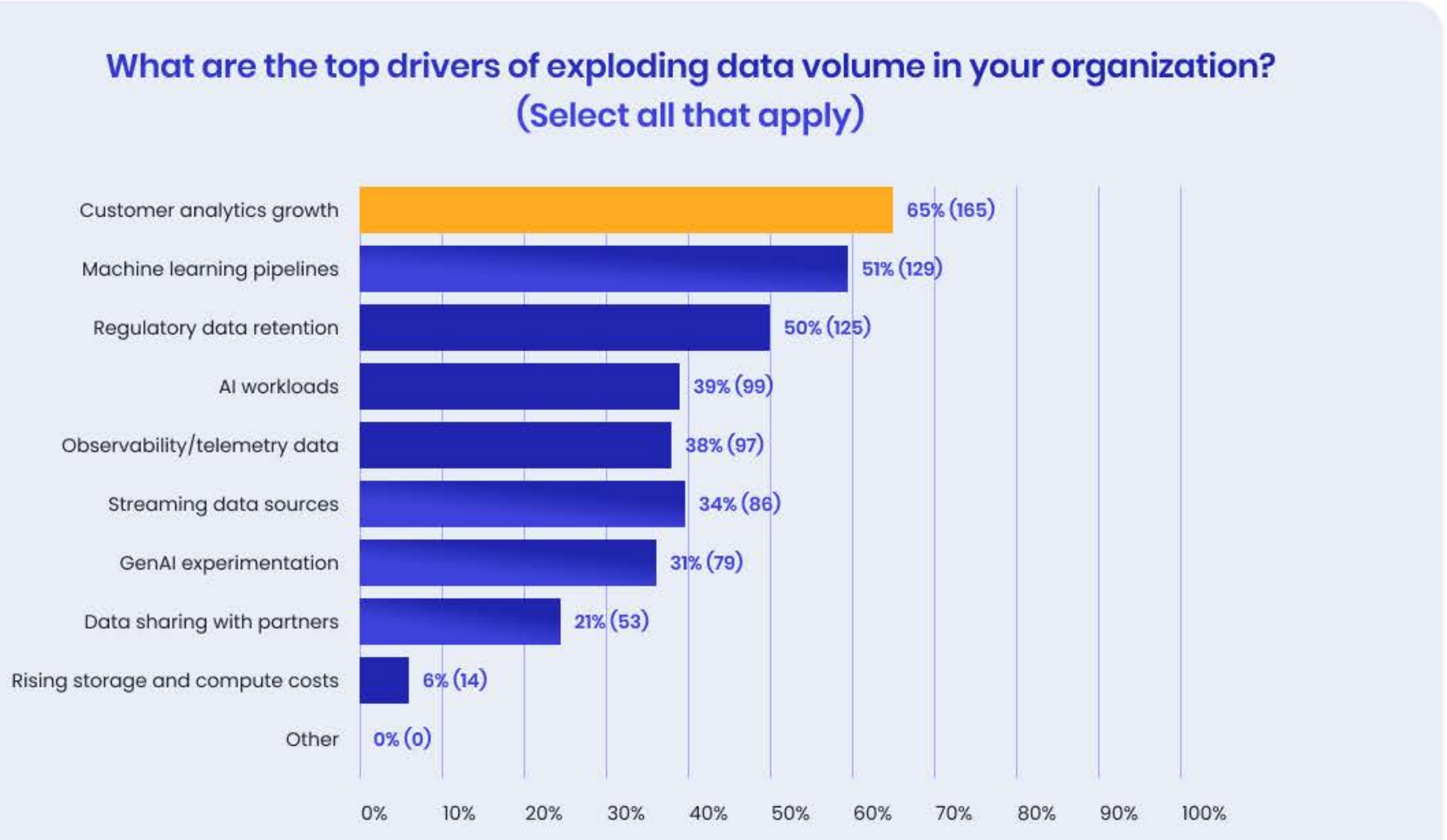
Iceberg operations typically need to move from ad hoc scripts to a managed platform once teams cross clear scale thresholds, such as thousands of tables, hundreds of terabytes of data, multiple query engines, or regulated data deletion requirements.

Trends: Why Data Is Growing Faster Than Teams

Respondents attribute the acceleration of data growth to several forces, led by customer analytics, ML pipelines, and regulatory retention.

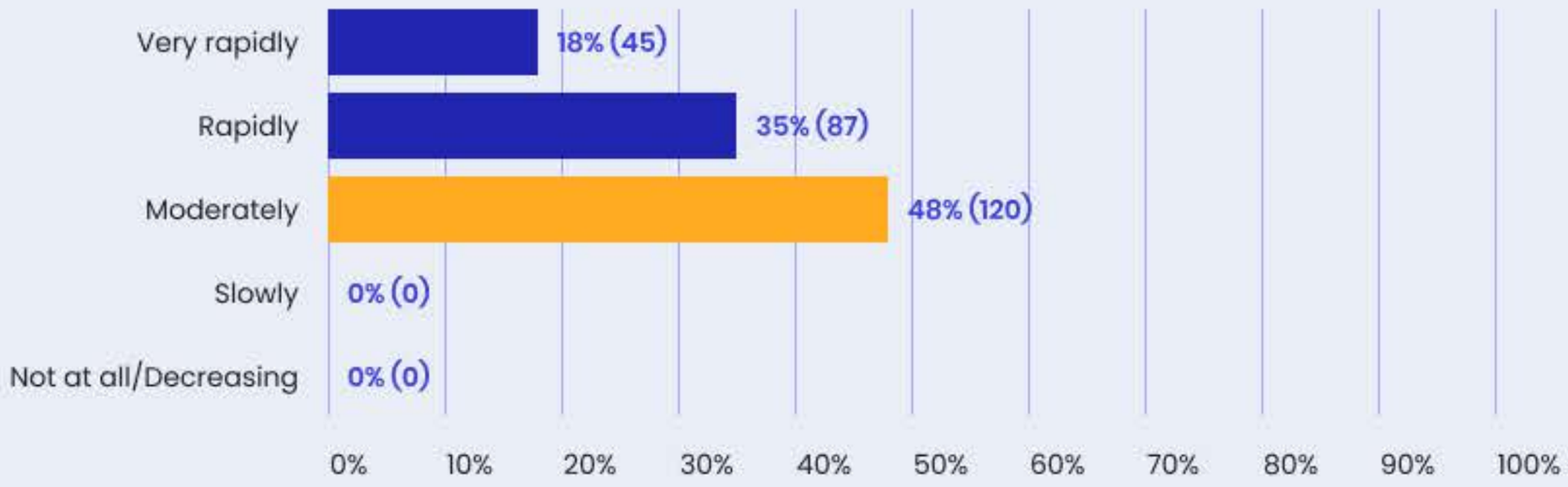
- **Customer analytics and product telemetry** are generating larger, more granular datasets.
- **Machine learning and AI workloads** are increasing both data retention needs and query intensity.
- **Regulatory and compliance requirements** are extending data retention windows and limiting aggressive deletion strategies.

Drivers of exploding data volume



Data volume growth expectations

How quickly do you expect your Iceberg-managed data volumes to grow over the next 12 months?



What this indicates

Teams use Iceberg to handle rapid data growth, not because they want more complexity, but because they need a format that scales without locking them into a single engine. The research shows Iceberg increasingly supports multi-engine analytics, real-time or near-real-time workloads, and long-term historical analysis in the same environment.

Data volumes are growing faster than teams can realistically manage by hand. Retention requirements and AI-driven workloads are compounding the problem.

Ryft perspective

Growth itself isn't the risk. The risk is an unmanaged data lifecycle, where data, metadata, and costs accumulate without clear ownership or automation.

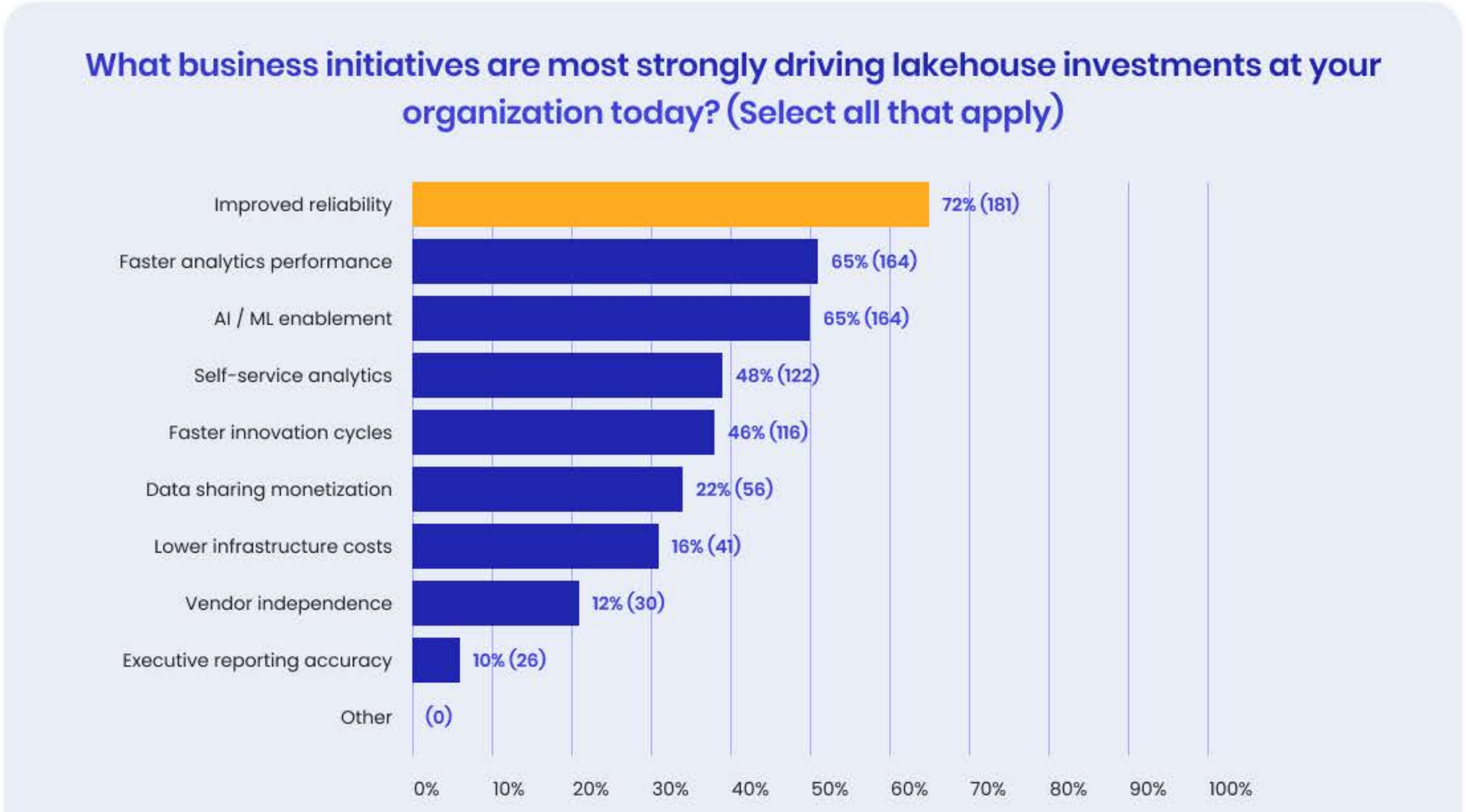
Business Drivers: Why Teams Chose Apache Iceberg

Adoption motivations skew toward operational outcomes: reliability, performance, and AI/ML enablement, more than ideological goals like vendor independence.

- **Reliability and predictability** under growing production load
- **Performance consistency** as data sizes and query concurrency increase
- **Team empowerment**, sharing consistent data across teams

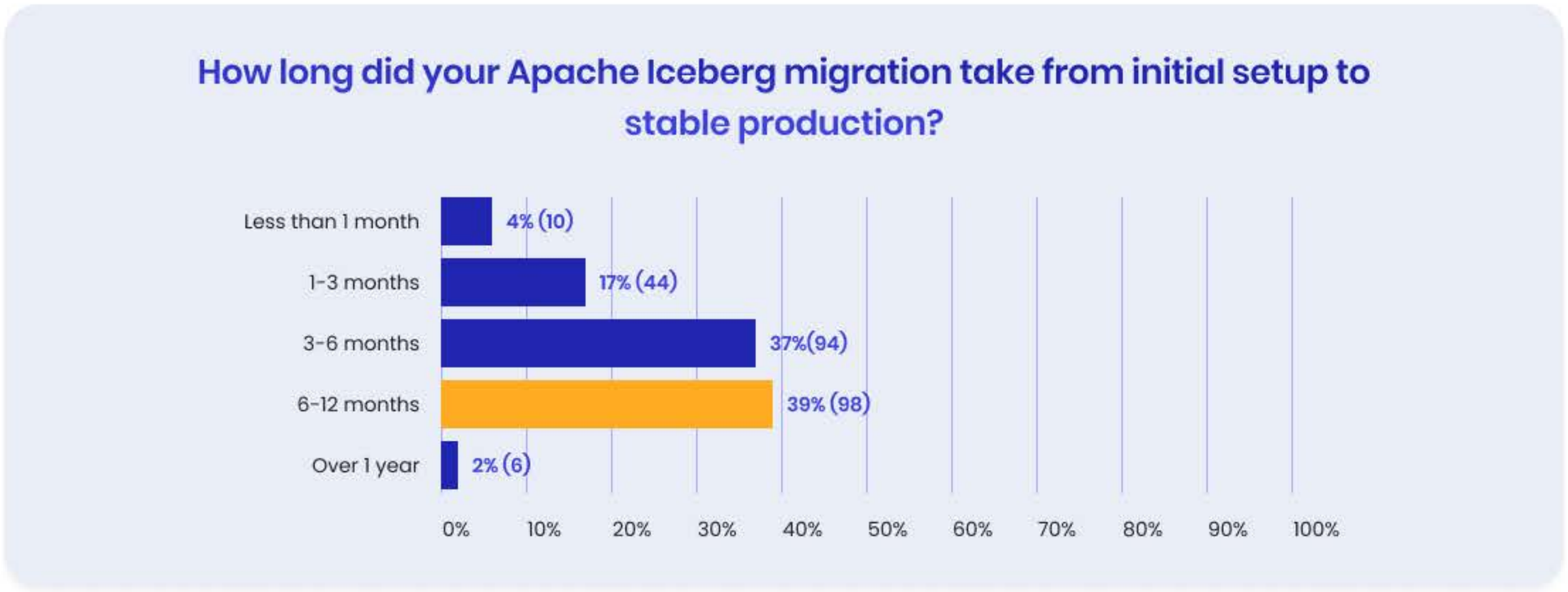
These same drivers are shaping lakehouse adoption more broadly, and Iceberg is emerging as the table format most closely aligned with them.

Business Initiatives driving lakehouse investments



Iceberg migration duration

76% of Iceberg migrations take from 3 to 12 months.



What this indicates

Organizations are choosing Iceberg to improve reliability, performance, and data access.

Ryft perspective

When teams justify Iceberg based on business outcomes, they must operate it with the same discipline as other production systems that carry SLAs.

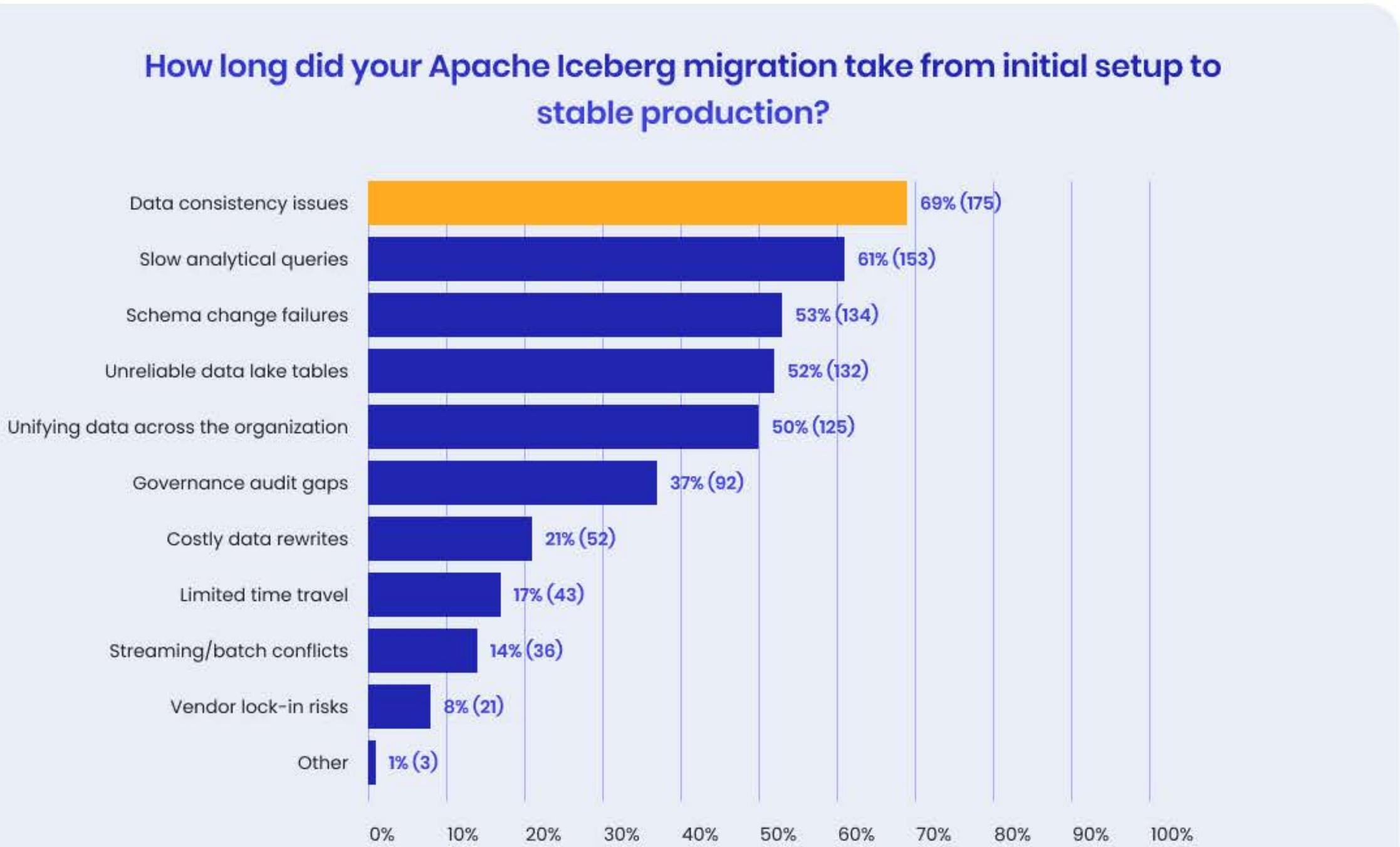
Market Sentiment: Iceberg Delivering on Its Core Promises

Overall sentiment toward Iceberg is positive. Most respondents report that Iceberg is delivering on its core promises:

- Addressing business problems
- Improving query performance
- Unlocking new use cases
- Providing access to trusted data to new teams
- Lowering costs

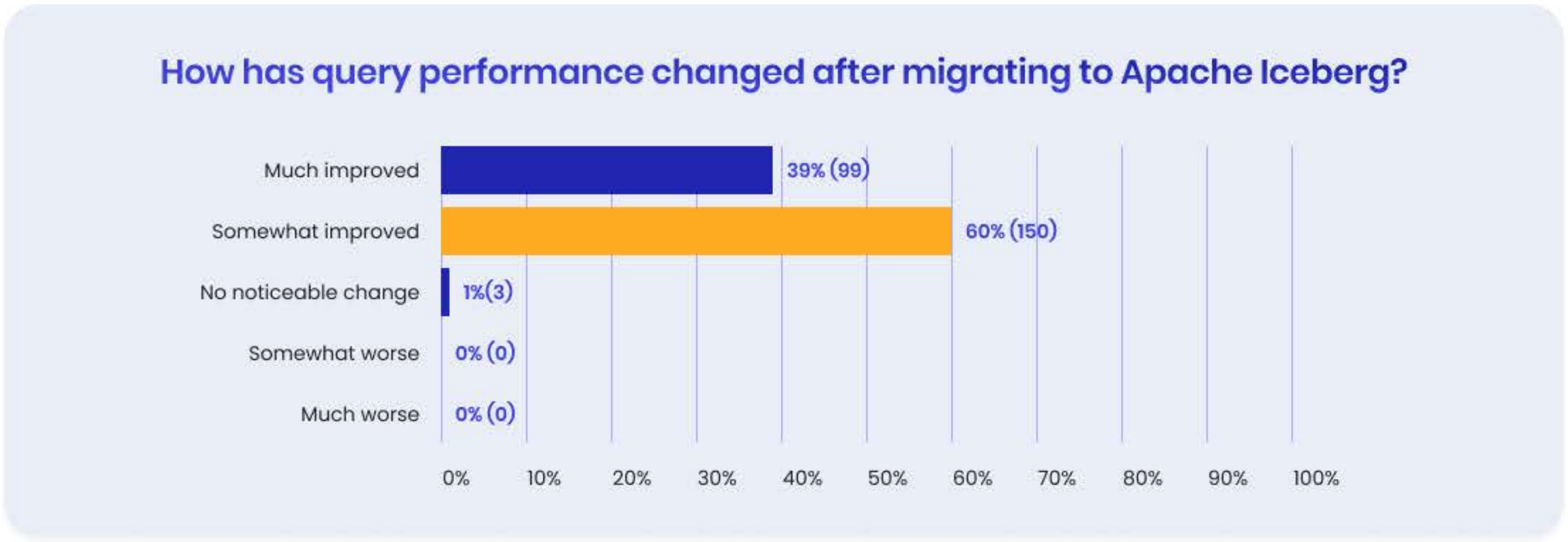
Business problems addressed

Respondents indicated that Iceberg addresses at least 10 business problems, with the top three being data consistency, query speed, and schema change failures.



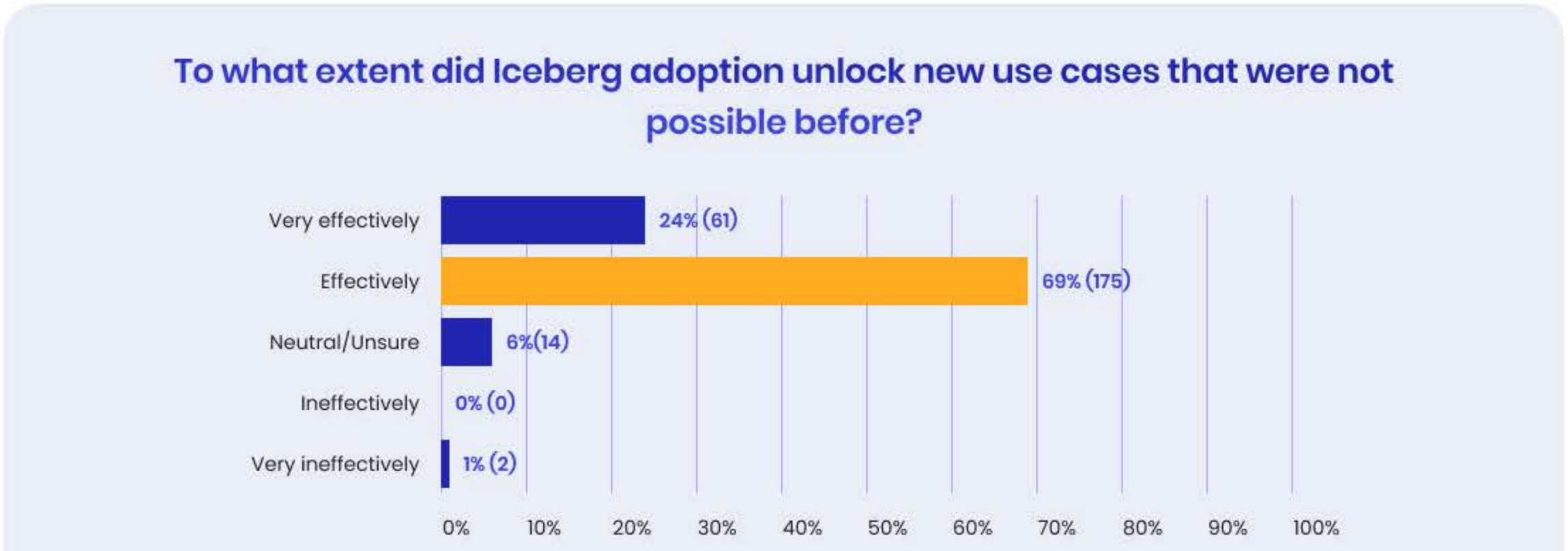
Query performance improvements

99% of respondents report improved query performance after migrating to Apache Iceberg.



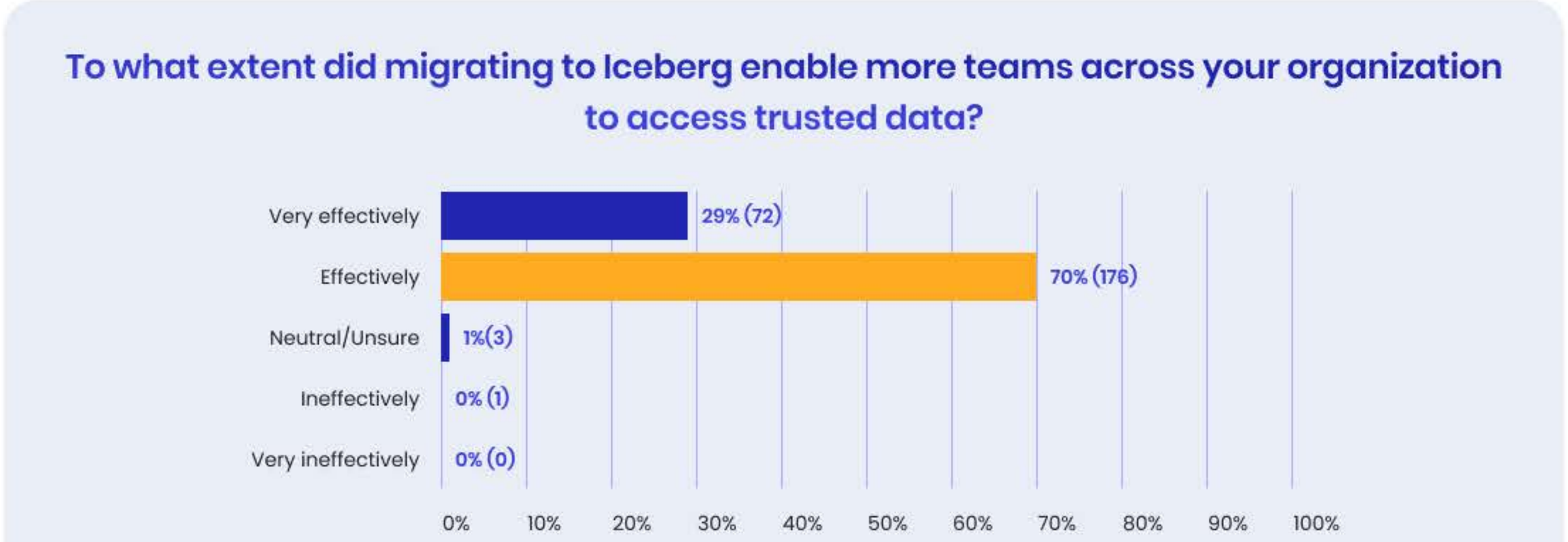
New use cases unlocked

93% of respondents report that Iceberg adoption unlocked new use cases.



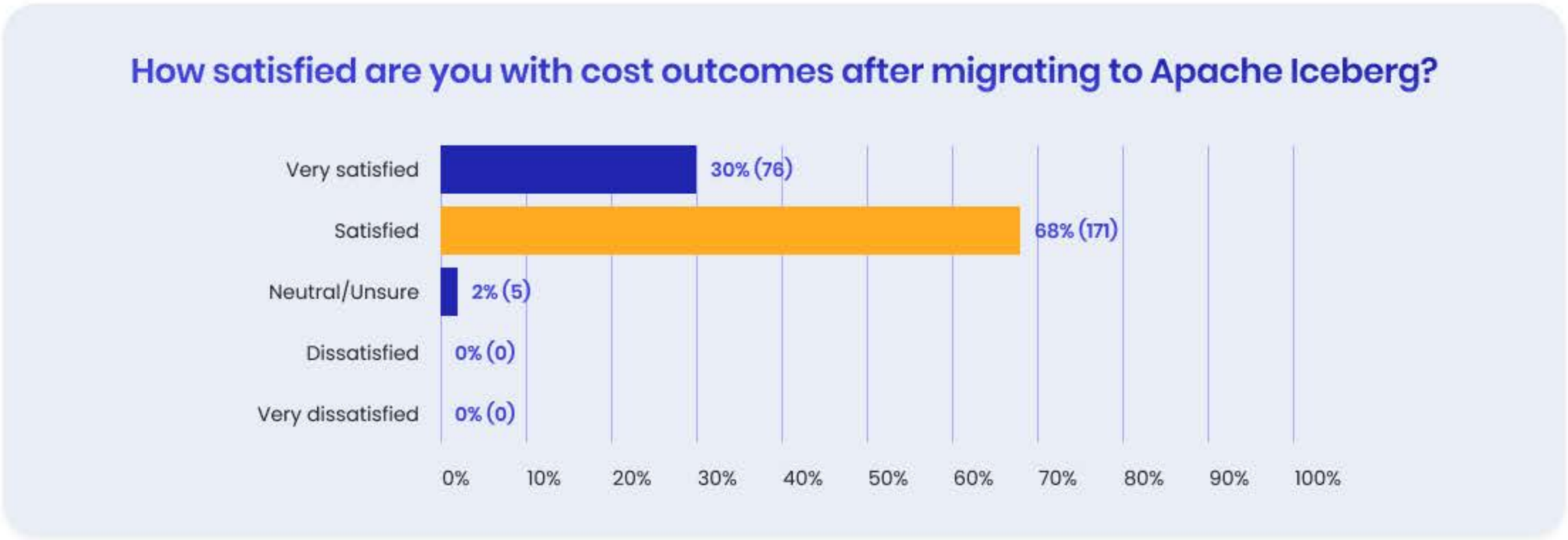
Data access to more teams

99% report that Iceberg enables more teams to access trusted data.



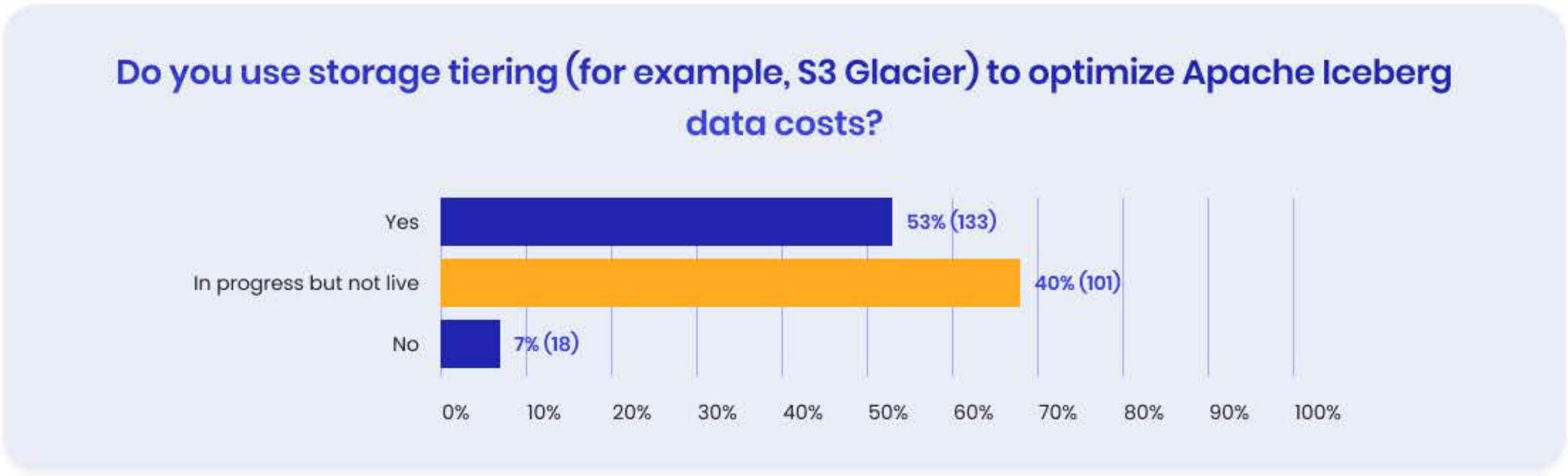
Cost outcome satisfaction

98% of respondents are either satisfied or very satisfied with cost outcomes after migrating to Apache Iceberg.



Storage tiering usage

93% of respondents are either using or plan to use storage tiering to optimize Iceberg data costs



What this indicates

Iceberg is meeting expectations for performance, flexibility, and new use cases. Satisfaction is high across production users.

Ryft perspective

Strong satisfaction raises the bar for operations. As Iceberg becomes more trusted, failures and governance gaps become less acceptable.

The Reality Check: Operating Iceberg

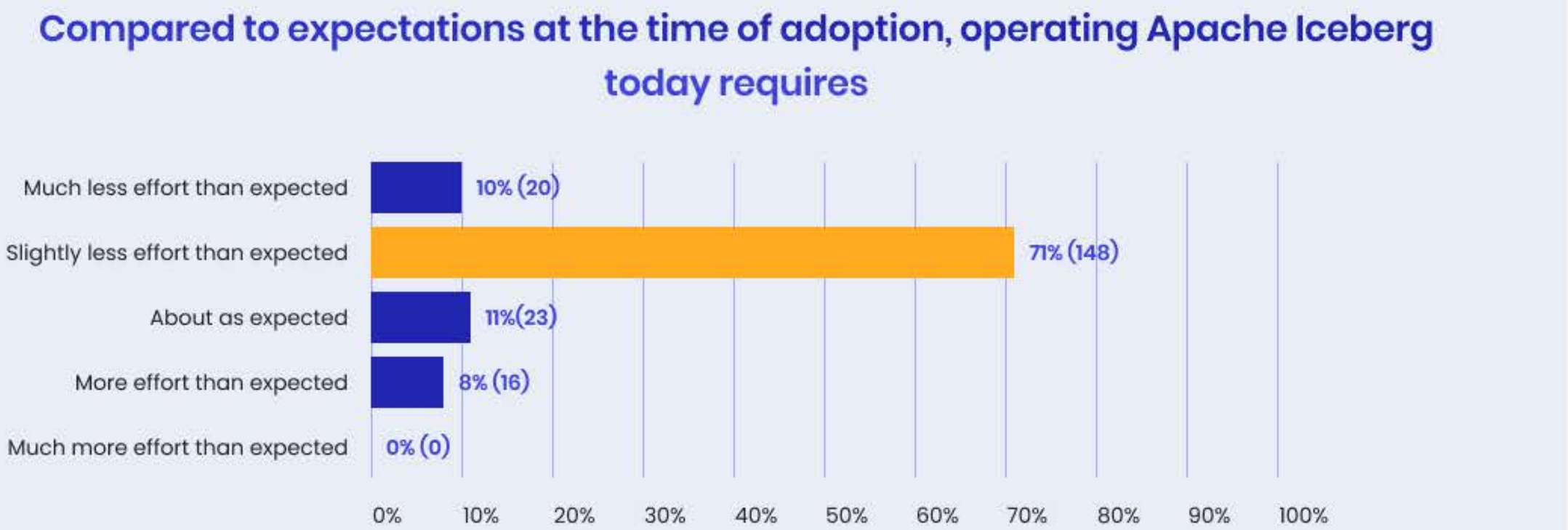
While many respondents say operating effort is 'slight less effort than expected,' most teams are still using DIY solutions to operate Iceberg and are reporting challenges.

Also, respondents report a wide variety of approaches to implementing the following enterprise requirements:

- GDPR compliance
- RBAC implementation
- Data retention policy, ownership, and enforcement
- Back-up and disaster recovery

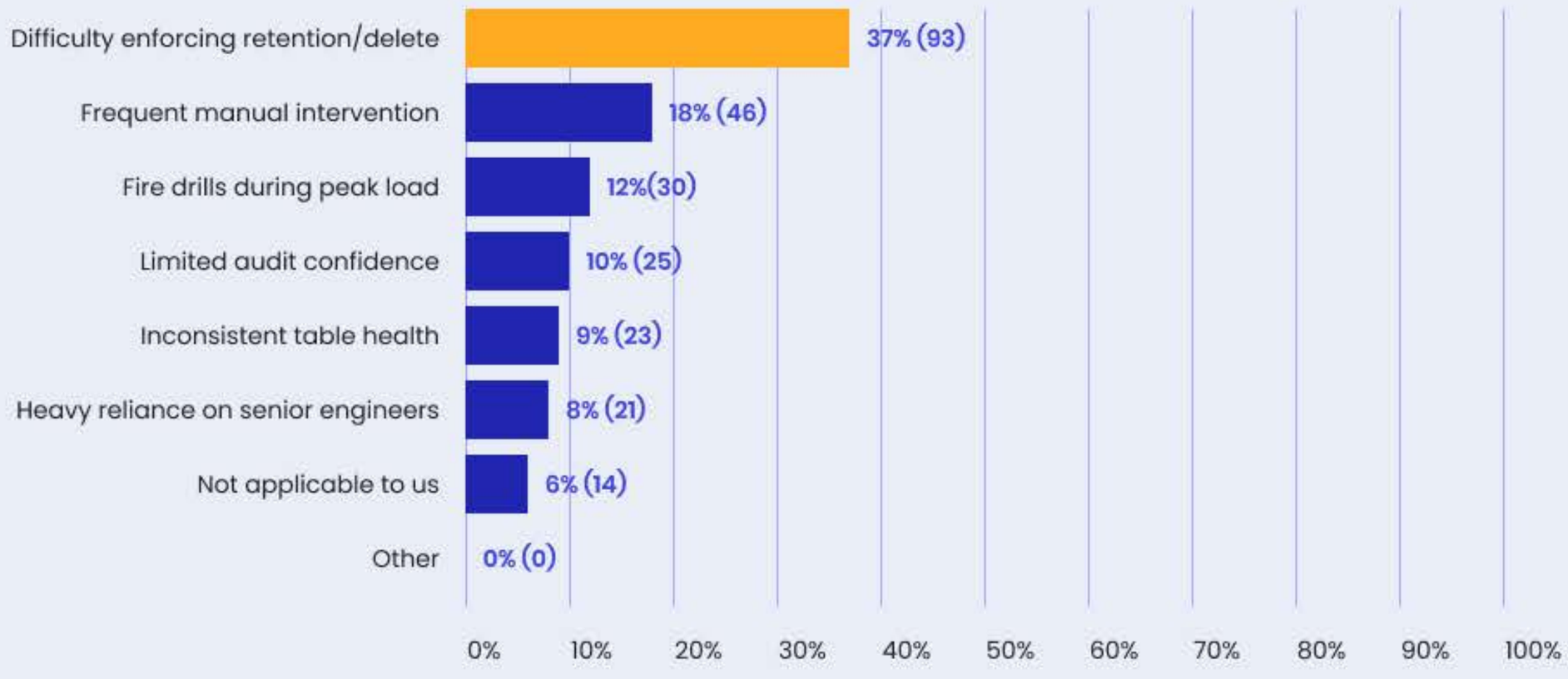
Engineering effort to operate Iceberg versus expectations

71% of respondents indicate that operating Iceberg requires about slightly less engineering effort as expected.



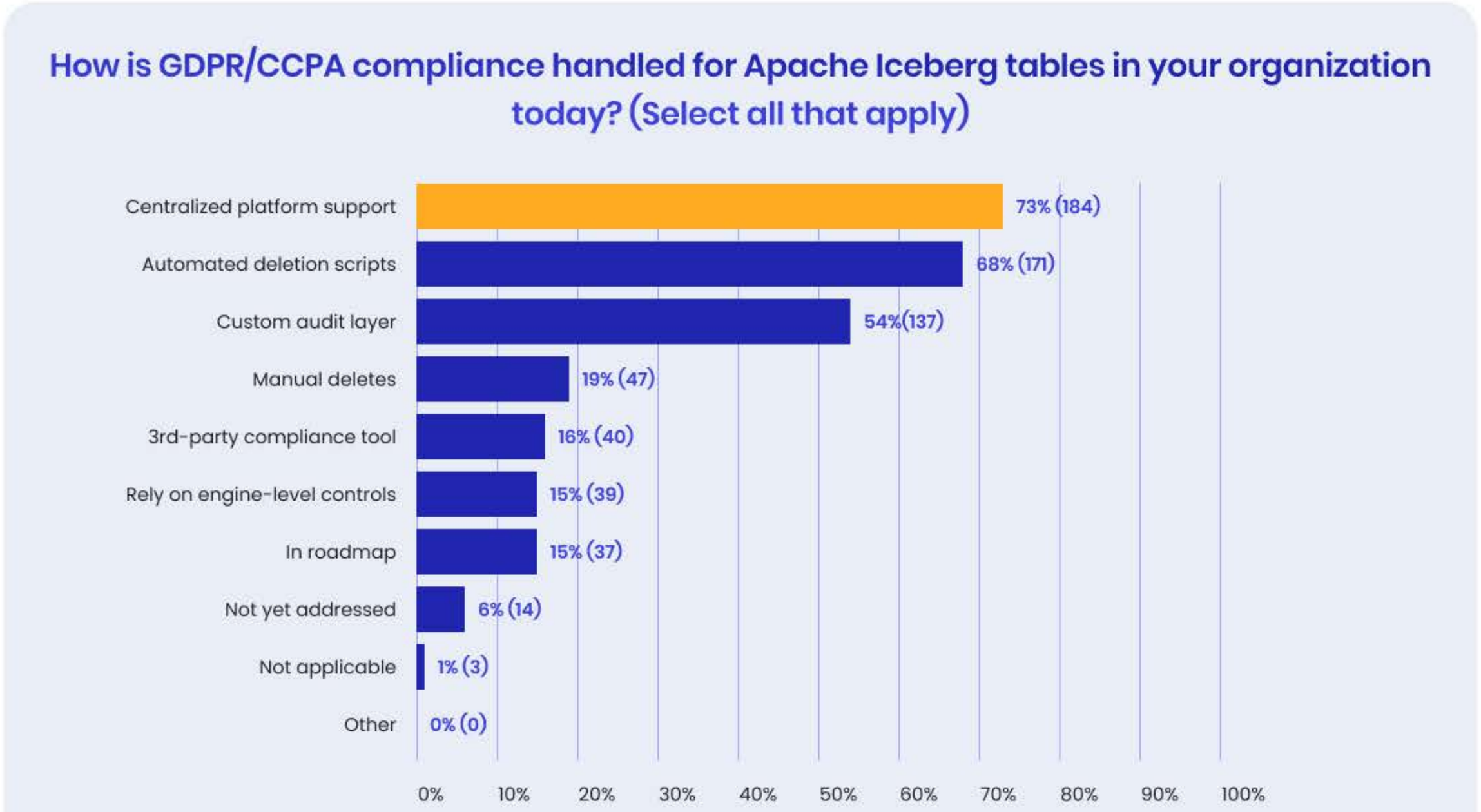
Challenges with DIY or internal tooling

Respondents report at least six challenges with DIY tooling, with difficulty enforcing retention or deletes as a top challenge.



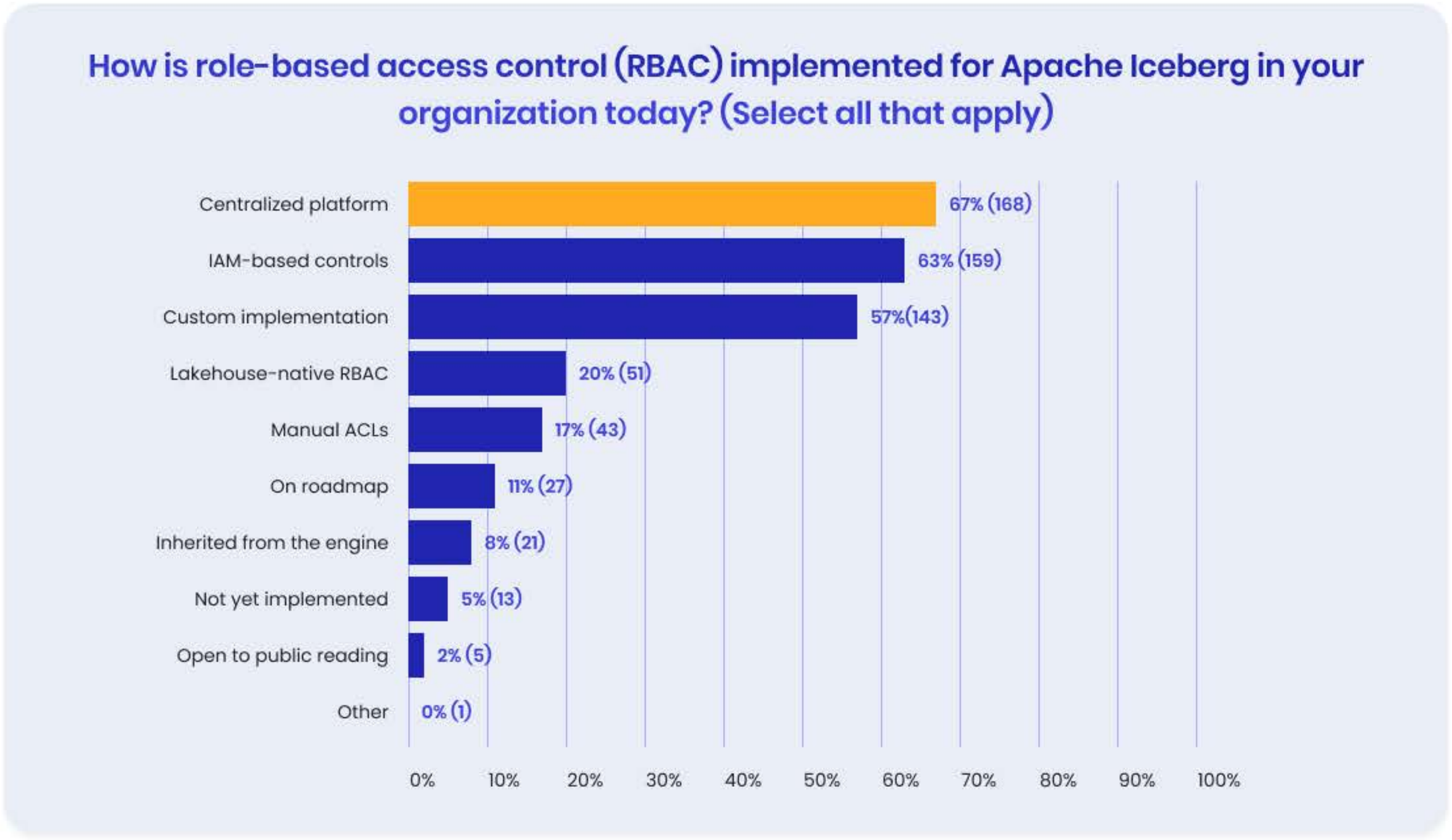
GDPR compliance handling

There are at least six ways data engineering teams are addressing GDPR compliance, indicating a lack of standardization.



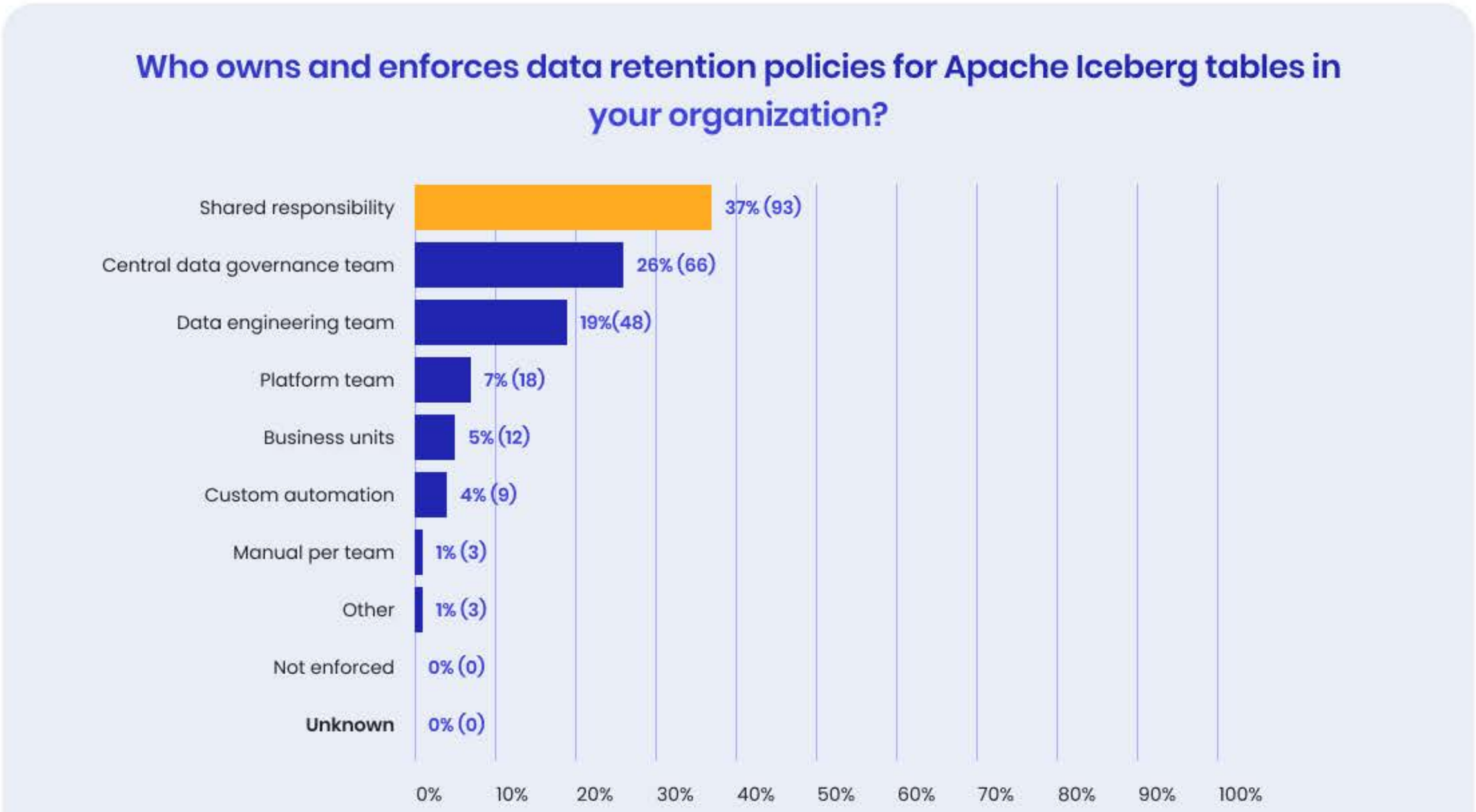
RBAC implementation

Respondents report at least seven approaches to implementing role-based access controls.



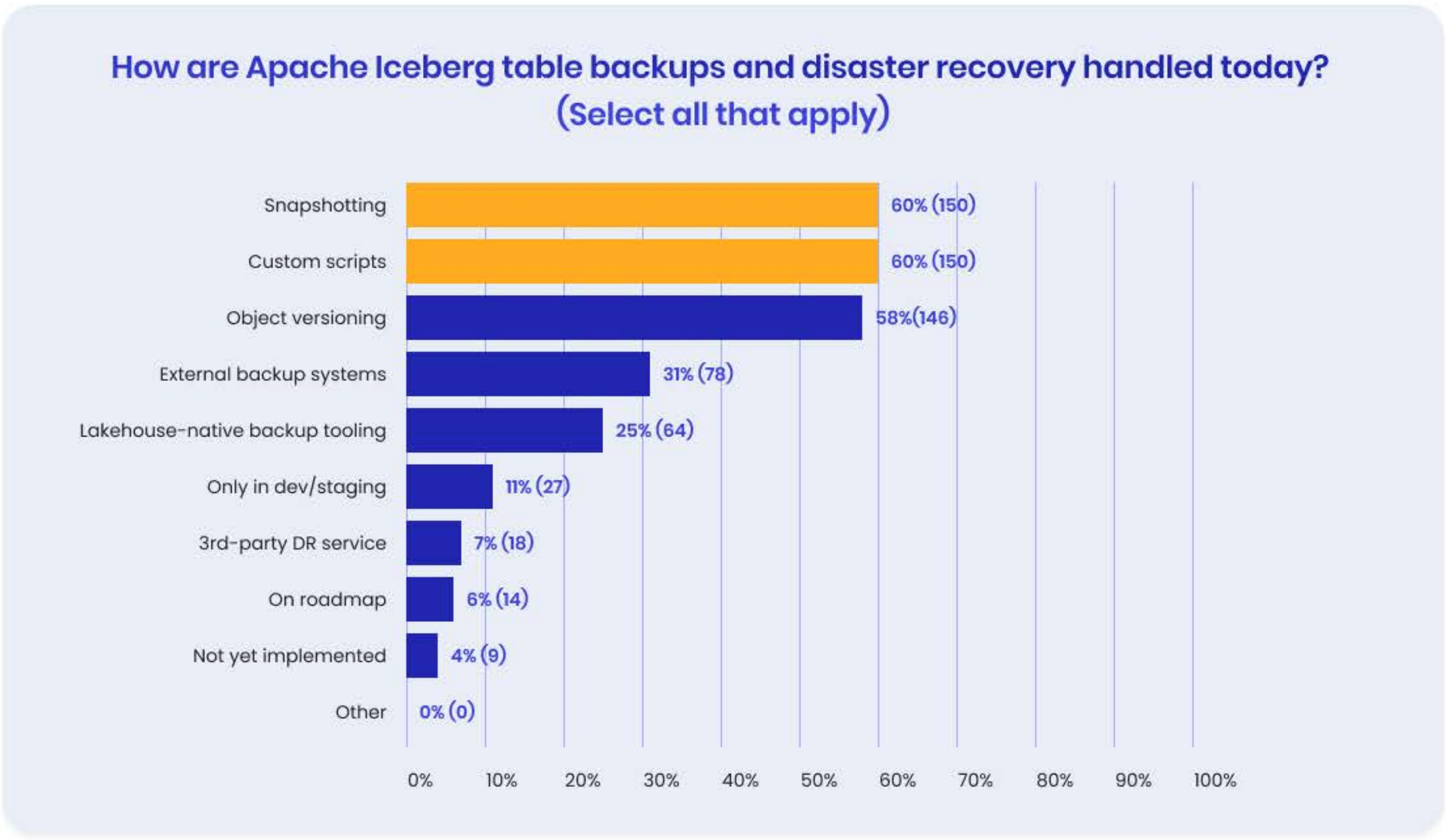
Ownership and enforcement of data retention policies

There are at least seven different approaches for data retention policy ownership and enforcement across respondent organizations.



Approaches to back-up and disaster recovery

There are a variety of approaches teams use for table backups and disaster recovery.



Key Challenges Reported by Apache Iceberg Adopters

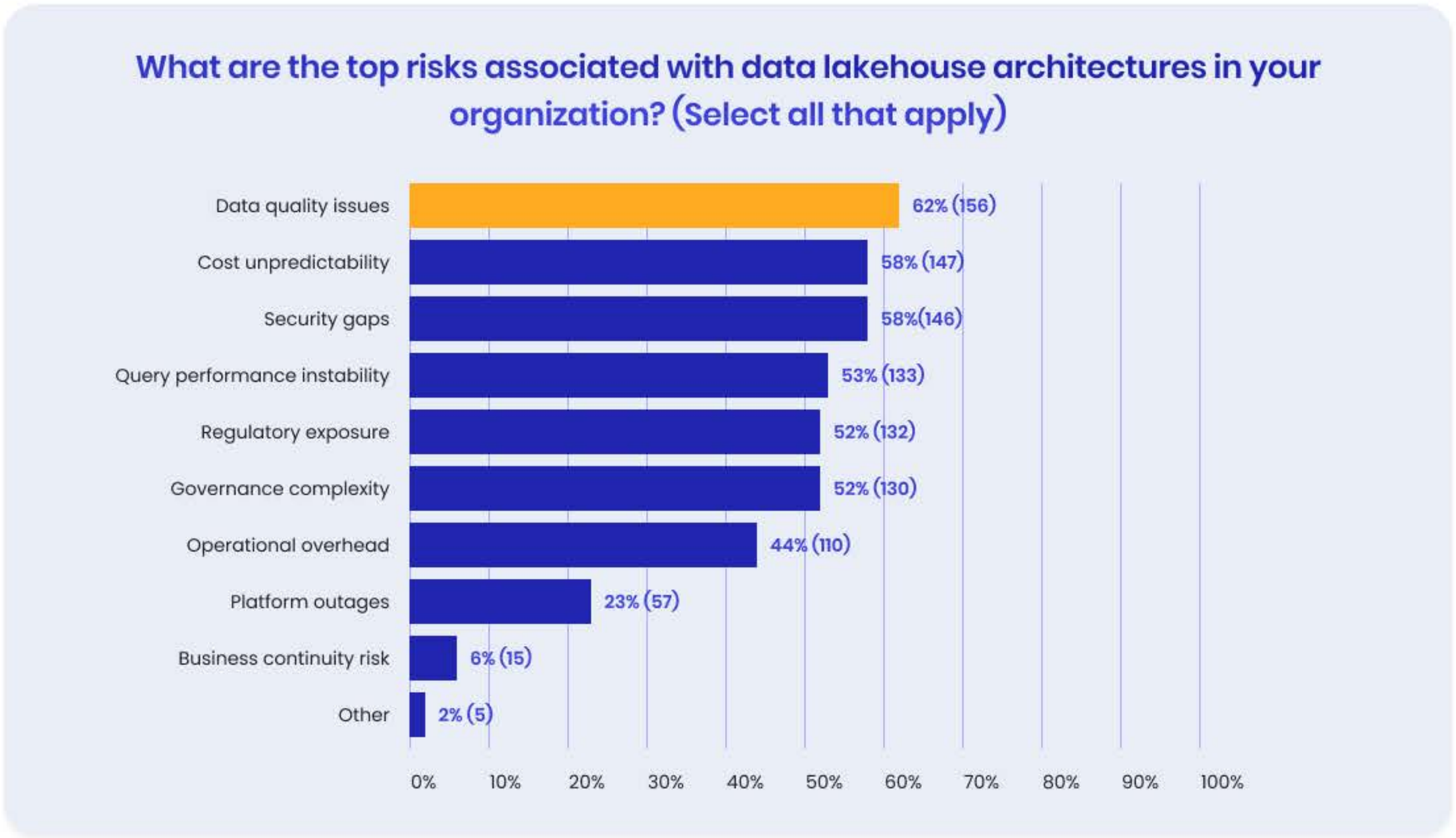
Respondents repeatedly surface similar challenges:

- **Operational complexity** as table counts and workloads grow
- **Limited visibility** into table health, performance, and cost drivers
- **Inconsistent governance** across teams and environments
- **Difficulty standardizing best practices** as adoption spreads

Teams do not view these challenges as flaws in Iceberg itself; they see them as gaps in their production management of Iceberg.

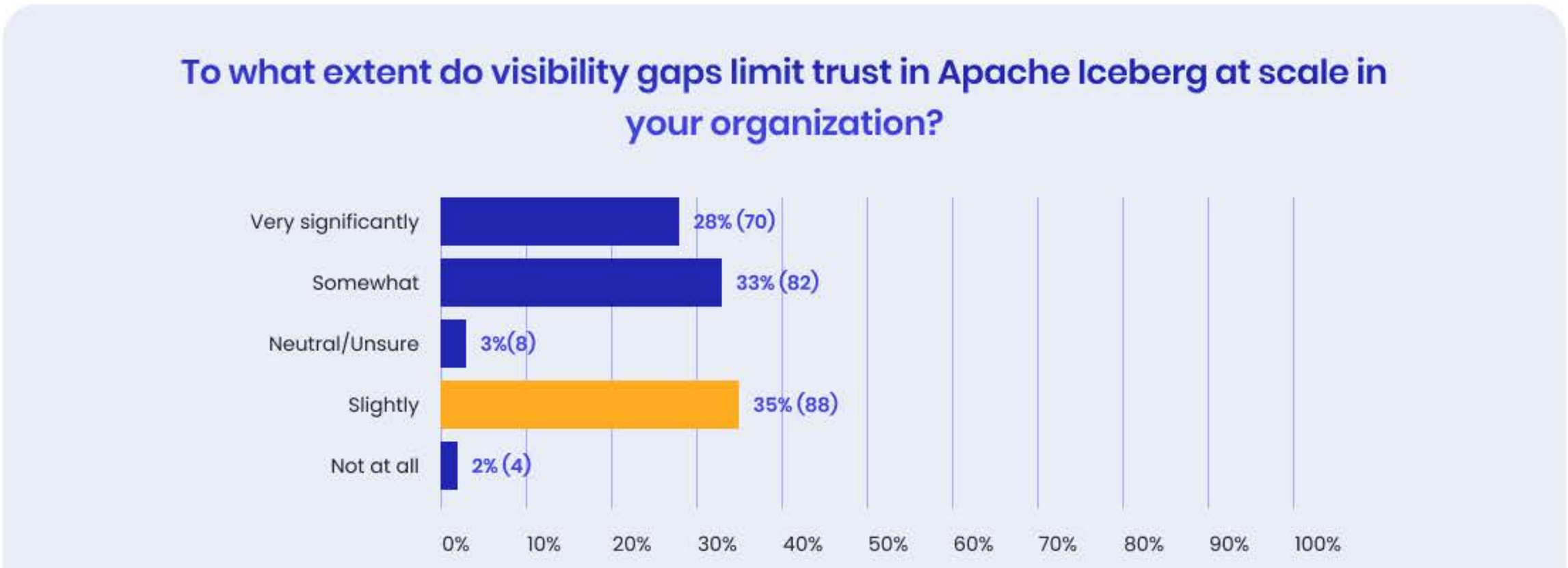
Data lakehouse architecture risks

Respondents cite a wide variety of risks associated with data lakehouse architectures.



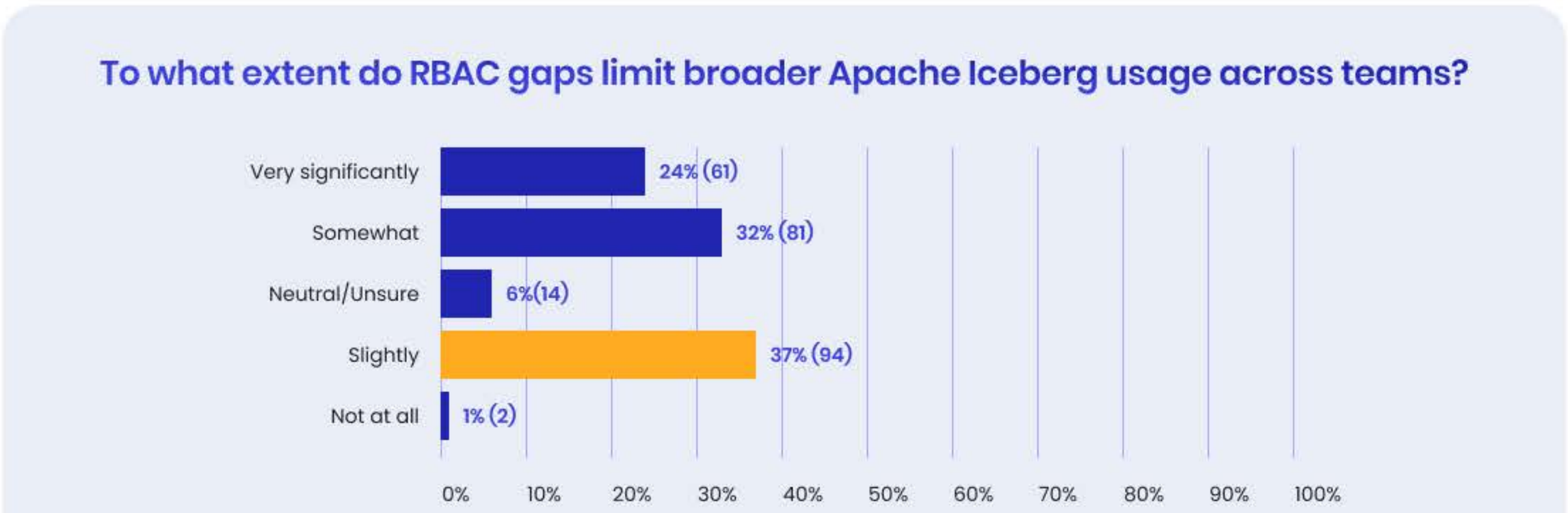
Visibility gaps limiting trust at scale

61% report visibility gaps that limit trust in Iceberg at scale.



RBAC gaps are limiting broader usage

56% report that RBAC gaps limit broader Iceberg usage across teams



What this indicates

Operational complexity, limited visibility, and governance gaps are the most common blockers to broader Iceberg usage.

Ryft perspective

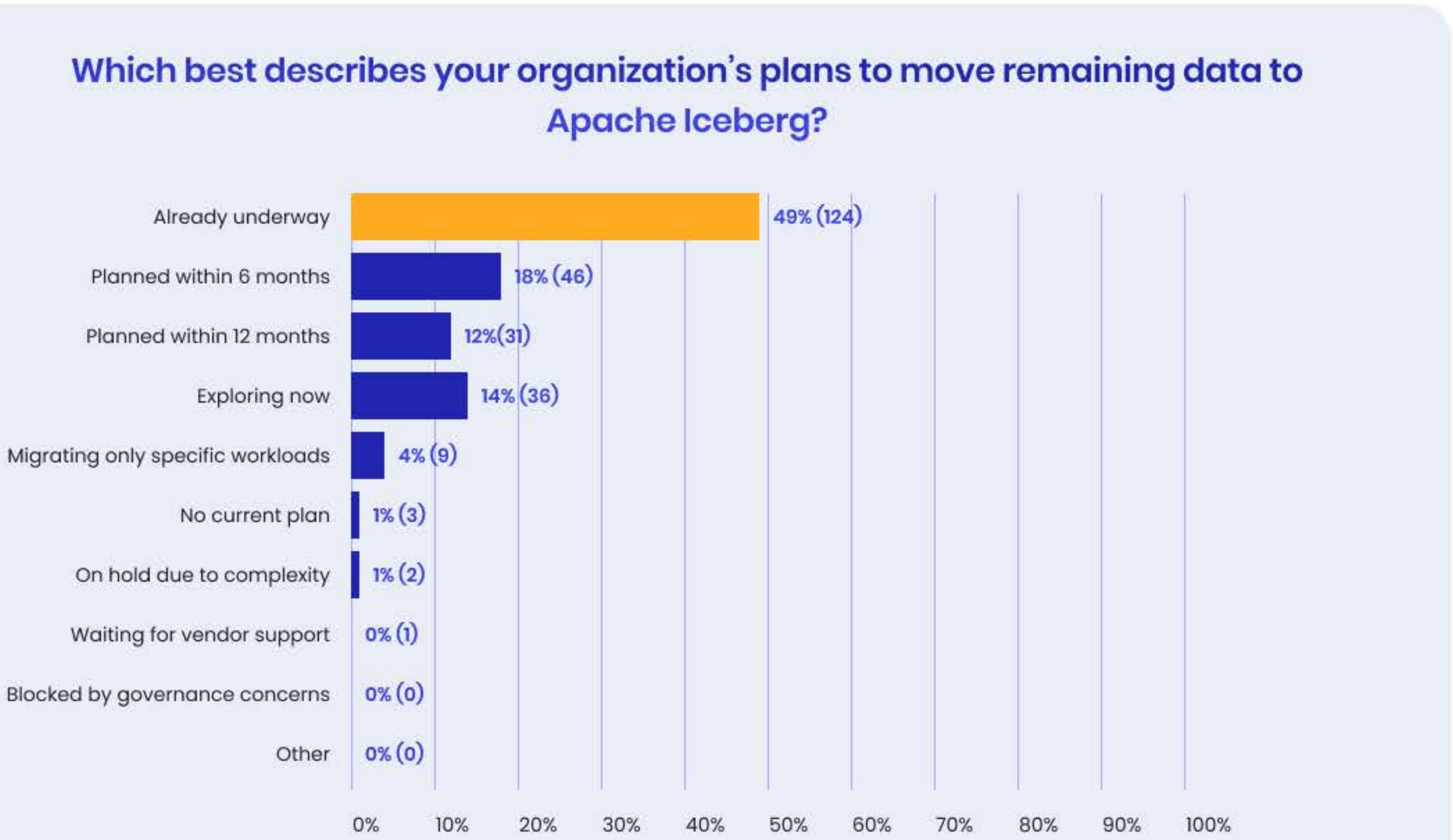
Trust at scale depends on visibility. Without clear insight into table health, performance, and policy compliance, teams hesitate to expand usage.

Future Plans

Importantly, Iceberg adoption is still expanding: many respondents plan to migrate additional datasets over the next 12 months to take advantage of Iceberg-enabled use cases, with data growth projections ranging from moderate to very rapid. Iceberg is becoming the default destination as legacy systems reach their limits.

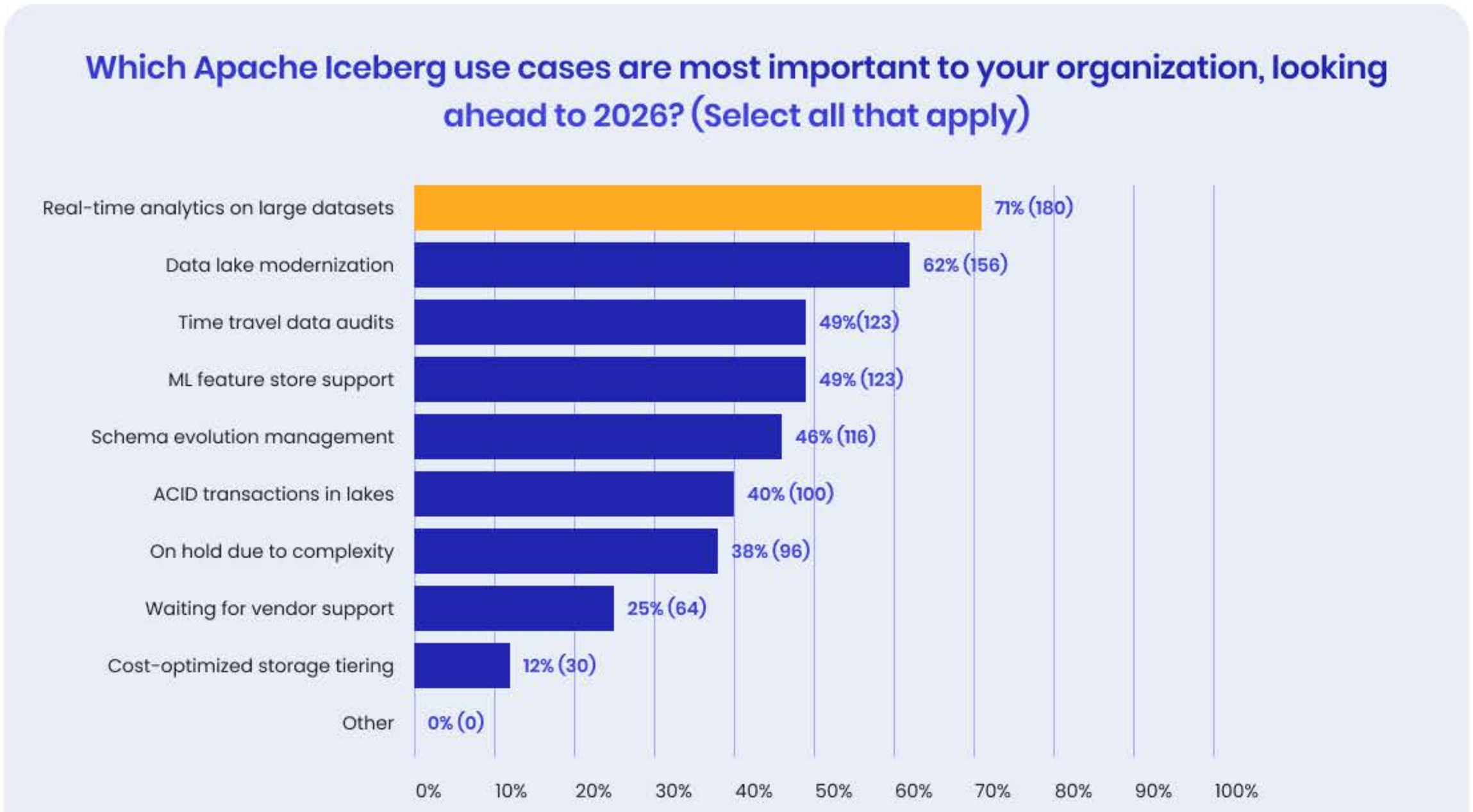
Plans to move the remaining data to Iceberg

79% are moving or plan to move their remaining data to Iceberg in the next 12 months.



Use cases for 2026

Organizations are looking forward to at least nine Iceberg-enabled use cases over the next year.



What this indicates

Most organizations plan to move more data and workloads onto Iceberg over the next year, increasing operational demands.

Ryft perspective

Expansion amplifies existing weaknesses. Teams that don't strengthen operations early will face more incidents as usage grows.

How to Benchmark Your Apache Iceberg Operations

If you are adopting Iceberg today, the market data suggests you should expect:

- Strong performance and flexibility benefits early
- Growing operational overhead as usage scales
- Increasing pressure to professionalize Iceberg management

Teams that plan for this transition early are reporting smoother scaling, fewer incidents, and greater confidence in expanding Iceberg to new workloads and business-critical use cases.

What This Means for the Market

Iceberg's momentum is real and accelerating. Adoption is expanding, satisfaction is high, and use cases are broadening.

The next phase of value creation is less about selecting Iceberg and more about operationalizing it: standardizing lifecycle management, enforcing governance, and improving visibility across workloads.

This research makes it clear that Apache Iceberg is becoming a core enterprise data primitive, and the teams that win with it are those that treat operations, governance, and visibility as strategic capabilities.

About Ryft

Ryft (ryft.io) is the AI Data Lake, built for Apache Iceberg. Ryft helps companies build a fully autonomous Iceberg data lake for ML & AI workloads, optimizing data in real-time based on usage patterns, automating compliance, managing access controls from a single place and ensuring that data access is efficient, reliable and secure.