Starburst

The Iceberg Data Lakehouse

The perfect foundation for all your data

eBook

Why data matters more than ever in the age of AI

The business world runs on data. You use it to understand which areas of your organization are performing well and which need to improve.

Data drives performance across the business. It powers dashboards, applications, and the decisions that shape strategy.

As a leader, your goal is to ensure that trusted data informs every decision. Building this culture is not a one-time project but a continuous process, where better data leads directly to better outcomes and competitive advantage.

Data isn't just a part of your organization. It flows through it.

AI is changing the rules around data

With AI, data is more important than ever. It was already crucial for analytics. AI doubles down on this, adding both opportunities and risks.

How will you adapt?

To succeed, you need to approach the problem through a new lens. Specifically, you need to look at three areas:

Access

You need to be able to leverage all of your data, whether in the cloud or on-premises. That data needs to be ingested quickly, making it immediately usable.

Collaboration

You need to flatten data silos and allow all of your teams to share data and resources. Using multiple technologies is driving multiple methodologies and creating confusion.

Governance

You need to ensure that the right people access the right data in the right way at all times. This is important for any business; it's absolutely critical for organizations operating in high-compliance environments.

How will you get there?

The true promise of both data and AI isn't just technological; it's organizational.

Recognizing this means recognizing that the right technology isn't an end in itself. It's a means towards achieving business change. To be successful, it's best to outline your objectives and work backwards from there.

Starting point: **You want to revolutionize your business using AI.**

- **Improve access**
- **Facilitate collaboration**
- **Ensure data governance**

Starburst helps your data drive organizational change

Starburst is designed to solve this exact problem.

Our technology, powered by Apache Iceberg and Trino, allows for universal access to data at scale, wherever it lives.

This gives you one central thing.

Choice.

Teams can move data if it makes sense, but they can also access it from any location. This gives you options and infuses a sense of choice into both your technology and your business, unifying access to your cloud and on-premises data.

Roadmap

This book will be your roadmap for organizational change, led by Apache Iceberg and Starburst. It outlines how you can prioritize data in ways that actually work, adopting AI while maintaining robust data governance.

How to get from A to Z *with AI*

Your business is at a crossroads.

You want to get started with AI.

But how will you do this?

How will you make AI adoption actually work?

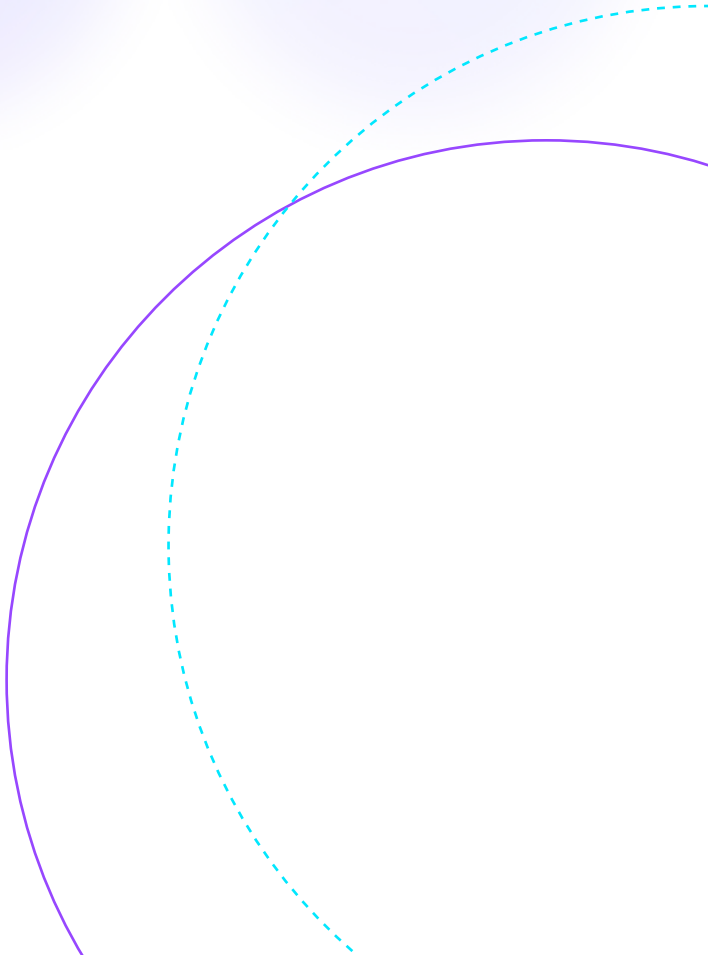
The state of data reflects the state of your organization

For most organizations, the starting point begins by recognizing that the current state of their data is fractured.

There are many reasons for this.

Some are technological, but others are organizational. Different teams might use different technologies. Older methods of data collection, retrieval, and processing coexist alongside newer ones.

The result is a patchwork of technologies, but more importantly, a patchwork of strategies supporting those technologies.



How technology and organizational strategy are related

Change is ultimately a function of business processes. This means that understanding your current technology is the first step towards understanding how to leverage that technology in the service of your busness.

Often, organizations have created business practices around what their technologies allow, rather than the other way around.

Adopting a new strategy that prioritizes business choices is the first step toward true organizational change. And that requires a technology capable of offering choice.

So, where do you fit in the evolution of data technology?

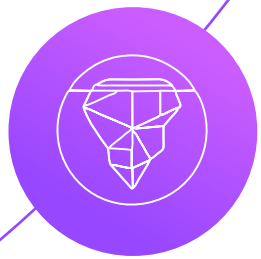
Let's look at three eras in the big data landscape.

Hadoop
The beginning of the data lake

Hadoop is the oldest of the big data technologies. Although many organizations still deploy Hadoop and HDFS, these technologies are slow, inefficient, and expensive. In terms of AI, they do not provide the correct foundation to run AI workloads.

Hive
The ultimate traditional data lake

Hive was an improvement over Hadoop. It was easier to use and faster. It also allowed insights from a wider variety of sources, including web click data. Although many businesses use Hive today, it is not always the most performant or efficient, especially for AI workloads.

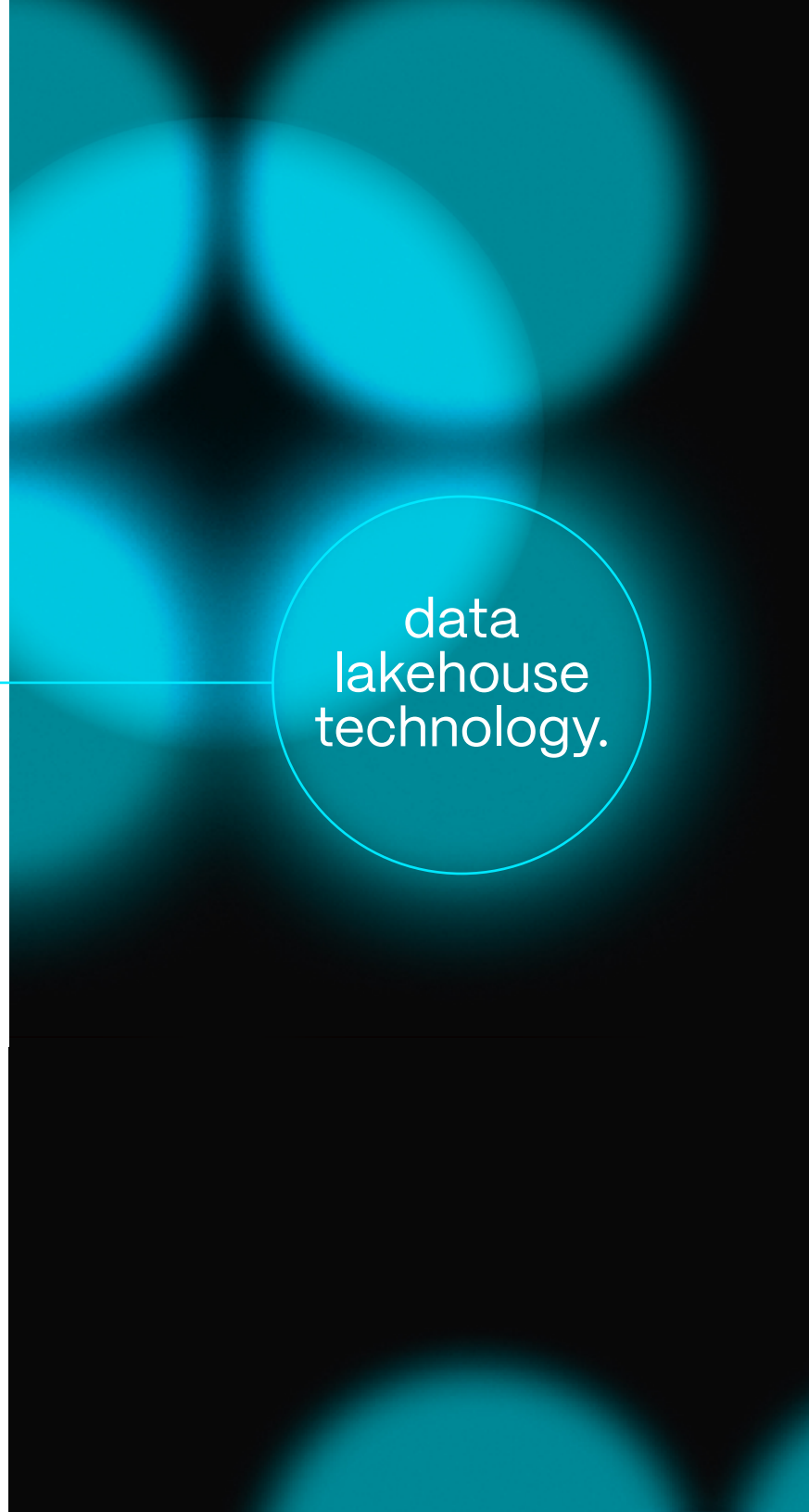


Iceberg
The data lakehouse era

Iceberg builds on everything that made Hive and data lakes valuable, while adding more flexibility and the ability to support a wider variety of data types.

Why Iceberg data lakehouses are now the default

Apache Iceberg is a



The data lakehouse is the latest iteration of data architecture, designed to make it easy to capture, process, and use all of your data, whether for analytics or AI. By breaking down silos, data lakehouses accelerate decision-making, ensure AI models have the clean and connected data they need, and create new value opportunities.

A lakehouse also improves efficiency by reducing duplication and lowering costs, while its open standards provide long-term flexibility. With a data lakehouse, your organization gains a single foundation that drives faster insights, supports innovation, and adapts to future change.

Why Apache Iceberg is the perfect foundation for your data

Especially the data that fuels your AI

Apache Iceberg is the key to unlocking your data's value. It provides an open standard that's scalable and adaptable, built for today's analytics and tomorrow's AI. Even more importantly, it is designed to handle the full complexity of your data, including structured, semi-structured, or unstructured data.

This means that Iceberg is perfectly positioned to unlock all of your organization's data.

Apache Iceberg as a driver for organizational change

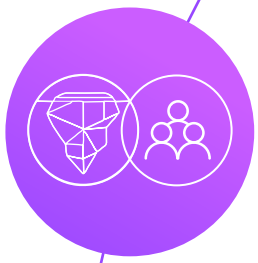
The challenges with traditional data lakes are not just technological. They are also organizational.

Limitations around access, ingestion, collaboration, and governance shape how teams work with data, often reinforcing silos instead of breaking them down.

Apache Iceberg addresses these gaps by providing flexibility for all data and all teams. This creates the perfect foundation where technology supports transparency, shared access, and organizational agility.

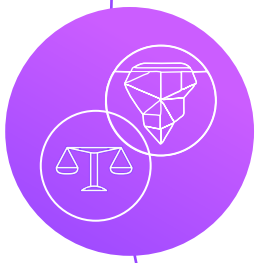
Iceberg and access

Apache Iceberg organizes your data lake like a database, giving you fast, consistent access across all your analytics and AI tools.



Iceberg and collaboration

Iceberg helps flatten data silos. Because it represents the best, most efficient table format for storing data of any type, it's the perfect format for organizations looking for a technology to carry them into the future.

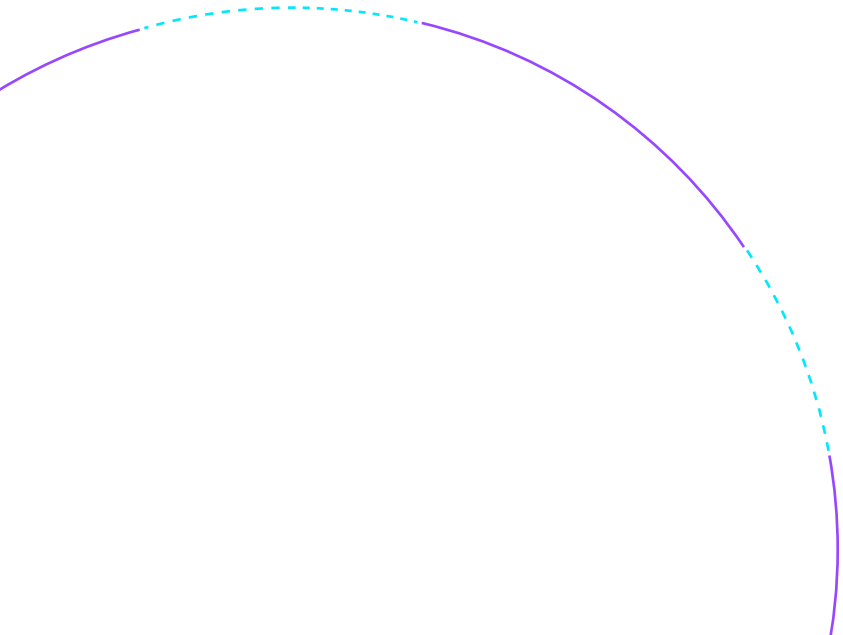


Iceberg and data governance

Useful data is governed data, and Iceberg achieves data governance better than any other technology on the market. Its enhanced metadata collection ensures that governance is baked into the technology, allowing for audit trails, version control, and a multitude of access controls.

Iceberg brings unique features that set your organization free

Iceberg lets you do more with your data, including AI workloads. The advantages of this aren't just technological, they're operational and organizational too. They bring true choice to your teams, stitching your patchwork of different data sources together.



Access transactional data easily

Traditional data lakes struggled with transactional data. In contrast, Iceberg provides full ACID guarantees, ensuring data consistency and reliability even as it evolves. With features that rival a traditional database, it supports both analytic and transactional workloads. This means teams can trust their data to always be accurate and usable.

Ensure that you're always ingesting the most recent data

Iceberg makes data ingestion seamless, whether through regular batch updates or real-time streaming data. Iceberg works with technologies like Starburst to ensure that new data is available immediately. This reduces latency and ensures that fresh data is always available for analytics and AI.

Maintain strong data governance with version control

Iceberg snapshots act like a time machine, allowing datasets to be rolled back to a previous state when needed. Every change is captured in an auditable trail, giving teams confidence in compliance and governance while reducing the risk of costly errors or failures.

Save money on storage costs

Sorted Iceberg tables organize data files by specific columns, making workloads faster and more efficient. The result is less time wasted scanning, delivering faster performance, and lower storage costs.

How Starburst and Iceberg work together

Starburst is the easiest way to leverage Iceberg at scale. Ingesting data into Iceberg tables and managing those tables can be complex and costly, but Starburst simplifies it with seamless ingestion, automated maintenance, and strong governance, helping data teams unlock value faster and more reliably.

The two technologies were built together and fit complementary roles for both analytics and AI workloads. Here, Iceberg operates as the table format and Starburst as the engine and platform, handling data ingestion and processing.

Using Starburst and Iceberg together provides universal access, enhanced collaboration, and strong governance.

Starburst and Iceberg work so well together that we call this data architecture by its own name.

Icehouse architecture

How Starburst Icehouse architecture is the perfect pairing

Starburst empowers organizations to discover and explore data across diverse sources, then seamlessly ingest the most critical datasets into an Iceberg lakehouse for fast, efficient, and reliable analytics at scale.

Iceberg was designed with Trino in mind, offering a natural fit between a fast, scalable query engine and a flexible, open table format.

Together, they form a powerful foundation for open lakehouse architectures.

Instantly make new data usable, automatically

Getting data into the right format quickly is essential for speed and accuracy.

With Starburst, data ingestion is easy. This ensures that new data is immediately usable and ingested, no matter whether it arrives through batch uploads or real-time streaming.

By ingesting data in Iceberg using Starburst, organizations reduce latency, improve reliability, and make fresh data instantly available for analytics and AI workloads.

Safely collaborate across teams

Starburst enables the creation of governed data products, allowing teams to share datasets in a compliant way.

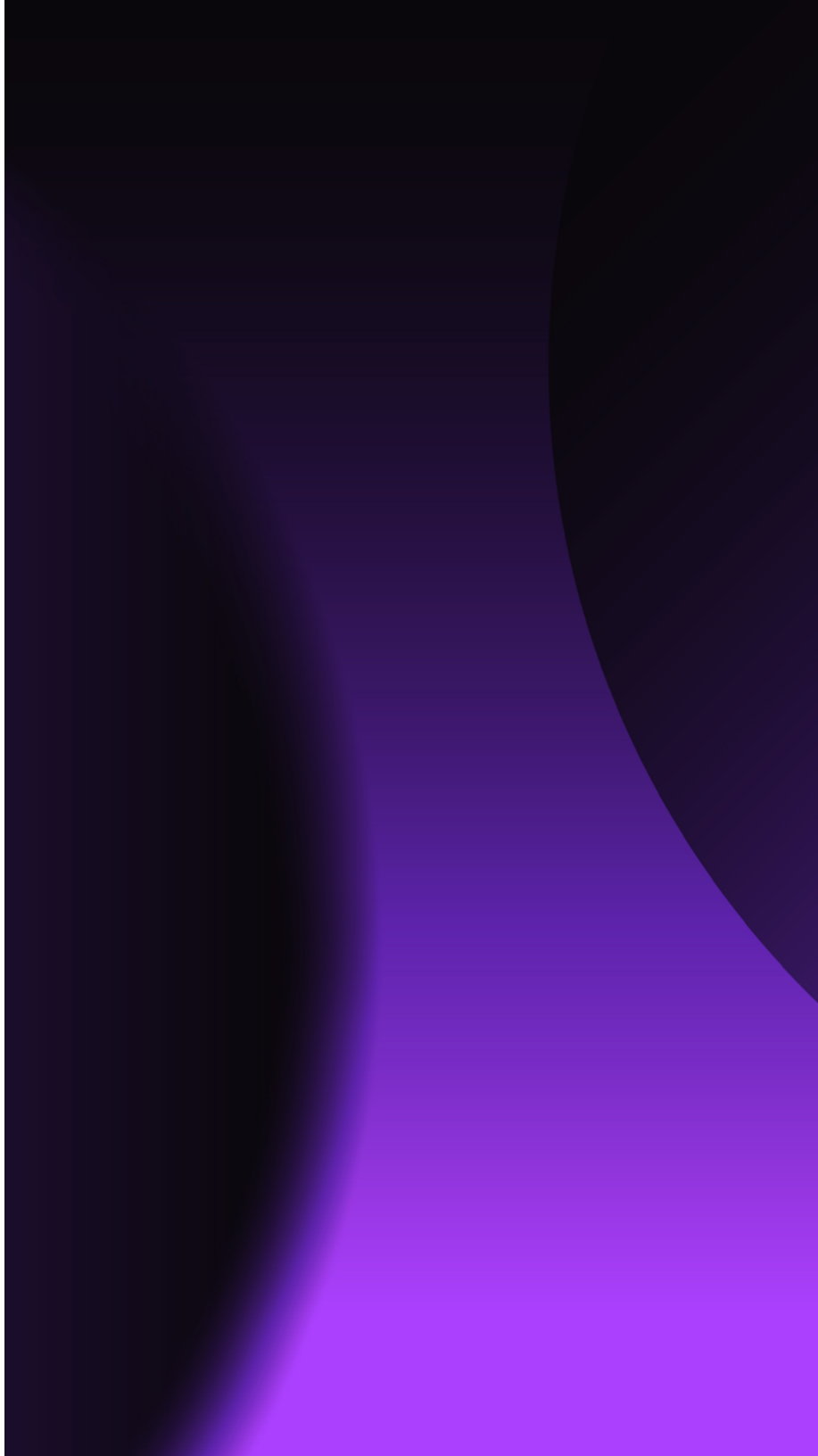
To do this, Starburst's robust data governance works alongside Iceberg's own strong guarantees around schema evolution, access control, and versioning.

This pairing creates a foundation where data is consistent, secure, and easy to reuse, enabling safe collaboration across analytics, data science, and AI projects.

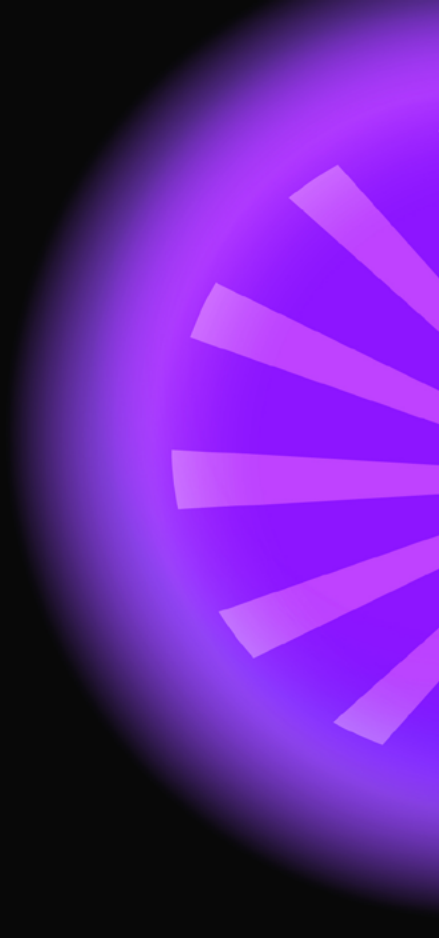
Manage your data, automatically

Iceberg was built to keep data organized and efficient as it grows. Its built-in maintenance features ensure tables stay healthy over time, without constant manual effort.

Alongside this, Starburst adds automated management, ensuring your data remains fast, reliable, and ready for use across the business.



Transform your business with Iceberg and Starburst



Modernizing your data architecture starts with choosing the right foundation. This foundation should work with the analytics you use today and the AI you will use tomorrow. It is able to access all of your data, in the cloud, on-premises, or across hybrid environments.

Starburst and Iceberg provide this foundation.

Together, they enable organizations to evolve beyond rigid, centralized systems and move toward an architecture that balances speed, governance, and choice.

About Starburst

Starburst is the data platform built for flexibility, delivering fast, secure access to all your data, wherever it lives. Whether on-premises, across clouds, or in hybrid environments, Starburst provides choice and control to your architecture. Built on an open data stack with Trino and Apache Iceberg, it unifies distributed data without complex or costly migrations, unleashing the full power of the data lakehouse for analytics and AI.

With our Lakeside AI architecture, enterprises gain federated access, governed collaboration, and full data lineage, laying the foundation for scalable, compliant AI innovation. Starburst empowers data-intensive and security-conscious organizations to unlock the full potential of their data while ensuring performance, governance, and control.

Enterprises in 60+ countries, including Comcast, Citigroup, and 4 of the top 5 global banks, trust Starburst to maximize data value. Our strategic partnerships with AWS, Dell Technologies, and top cloud providers ensure seamless interoperability across environments.

From insights to action to AI, Starburst fuels innovation at every level.

Learn more at **starburst.ai**

Starburst
Rocket fuel for AI