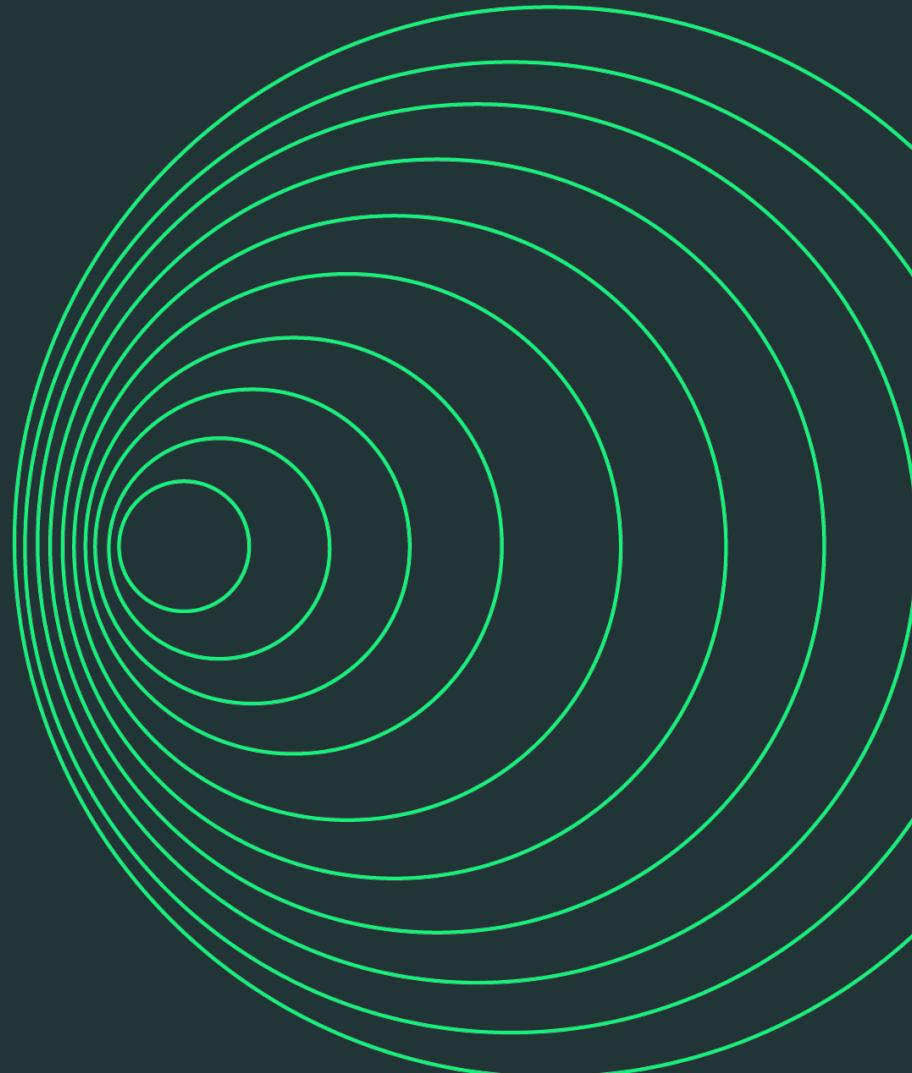




Enable, Adopt, Innovate

**Harnessing Trino to
supercharge data capability**



Welcome!

Who's who...



Jess Macleod
Senior Associate, Kubrick



Jay Lockwood
SVP North America, Kubrick



Brian Munz
Senior Director, Developer
Relations at Starburst



Monika Zajac
Senior Associate, Kubrick

What's what...

- ✓ **Let's get on the same page:** What is Trino and the common challenges
- ✓ **Enable to Adopt:** Building a data culture through our partnership with Starburst
- ✓ **Building strong capability:** The skills and tools that support an effective upskilling program
- ✓ **The Trino roadmap:** Upcoming announcements and how best to leverage the community

Ask your Qs!

Join us at
Slido.com

#2444 211



Section 01

What is Trino

‘Trino is a ludicrously fast, open source, distributed, massively parallel processing, SQL query engine designed to query large data sets from many disparate data sources.’

BRIAN MUNZ



What is Trino?

kubrick

Trino (*formerly known as Presto*) is a fast distributed SQL query engine designed to query large data sets distributed over one or more heterogeneous data sources.



Open Source

- User-driven development
- Large and active community
- Huge variety of users
- Apache license, version 2.0



Highly Performant

- ANSI SQL MPP Query Engine
- Cost-Based Query Optimizer
- High Concurrency
- Proven Scalability



Flexible Query Engine

- Scale storage and compute independently
- No ETL or data integration necessary to get to insights from multiple sources
- SQL-on-Anything



No Vendor Lock-In

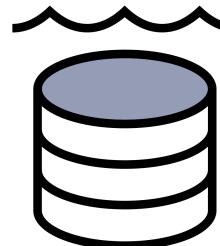
- No Hadoop distro vendor lock-in
- No storage engine vendor lock-in
- No data/file format lock-in
- No cloud vendor lock-in
- No database lock-in

Where should I use it?

Interactive data analytics

Enter a SQL query for Trino to process and return results as quickly as possible.

- Query large amounts of data
- Test hypotheses
- Run A/B testing
- Build visualizations



High performance data lake analytics

Trino enables users to run SQL based analytics on HDFS/Hive and cloud object storage

- Run petabyte scale analytics
- Capitalize on scale and performance benefits

Federated analytics

Create a single point of access by using Trino to query disparate data sources.

- Object storage
- Relational systems
- Streaming systems
- NoSQL systems



Batch ETL processing

Run resource intensive ETL processes in batches without fear of failure with Trino.

- Use SQL with every data source
- Work with numerous data sources and targets all in the same system
- Ensure speed and reliability

What Challenges does it solve?

With the growing demand for huge volumes of data, multi tech stack approaches and the rise of generative AI we have seen numerous challenges arise.



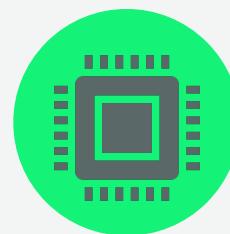
Data Integration & Interoperability

Challenge

Increasingly complex data landscapes with multiple sources and formats.

Solution

Trino's ability to query data across diverse sources and formats without data movement or transformation supports seamless integration and interoperability.



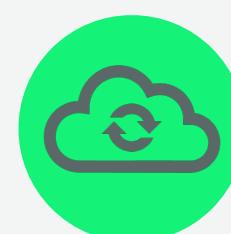
Scalability and Performance

Challenge

Handling large volumes of data and complex queries in a scalable manner.

Solution

Trino's architecture is designed for high performance at scale, efficiently managing large datasets and complex analytical queries.



Data Silos

Challenge

Accessing and combining data from varied sources like Hadoop, AWS S3, relational and NoSQL databases.

Solution

Trino facilitates federated queries across these sources without the need for data movement, ensuring data integrity and reducing complexity.



Accessibility

Challenge

Leveraging SQL for querying non-SQL databases.

Solution

Trino's 'SQL on everything' approach allows SQL querying even on non-SQL databases, enhancing accessibility and usability.

Exploring Trino's Impact

In our exploration of Trino's capabilities, we delve into two compelling use cases:

1. Distributed caching on a data lake
2. Anomaly detection at Salesforce

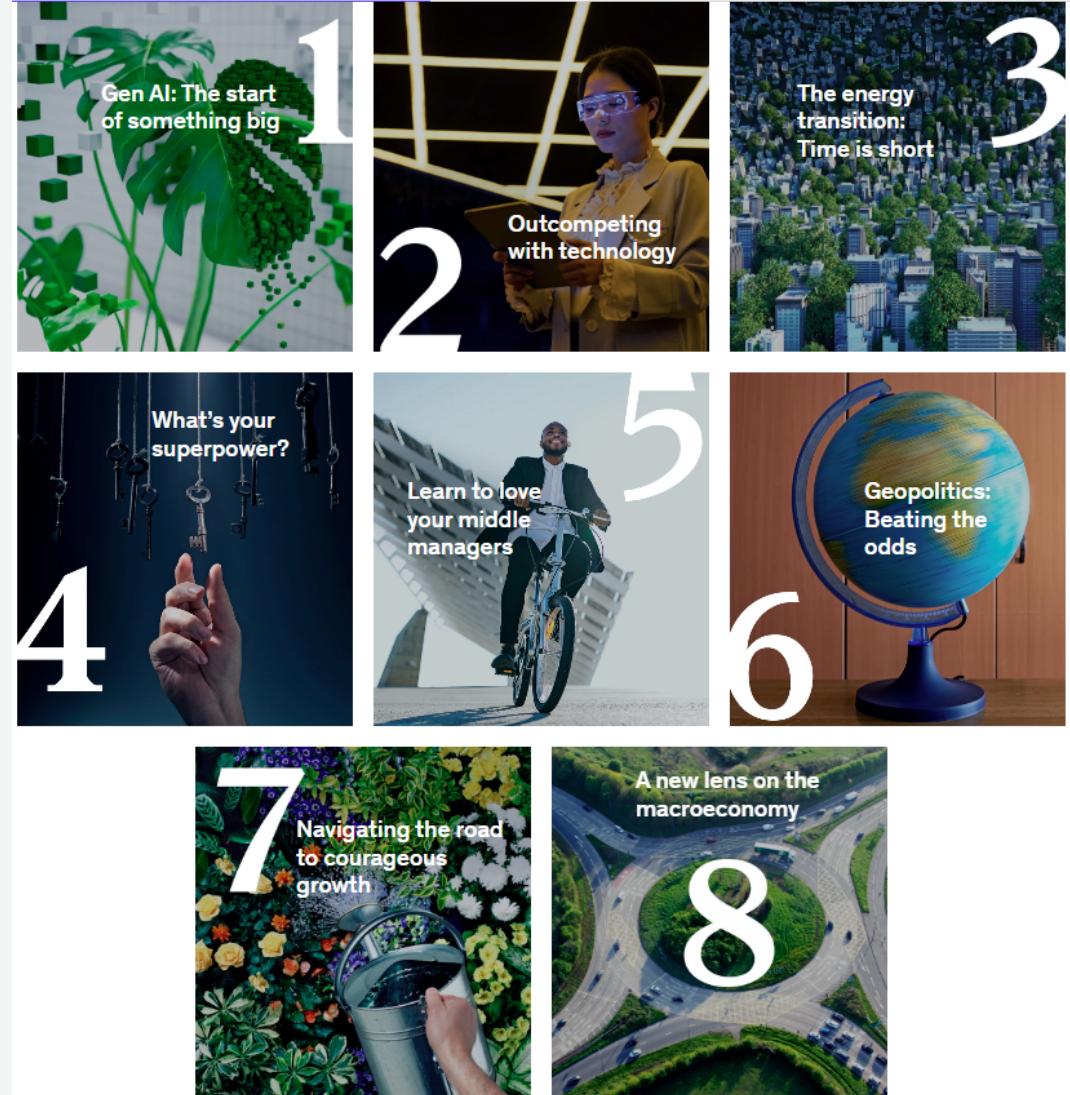


Section 02

Enable to Adopt

The C-Suite know tech needs to be a priority but are burdened with more external challenges than ever.

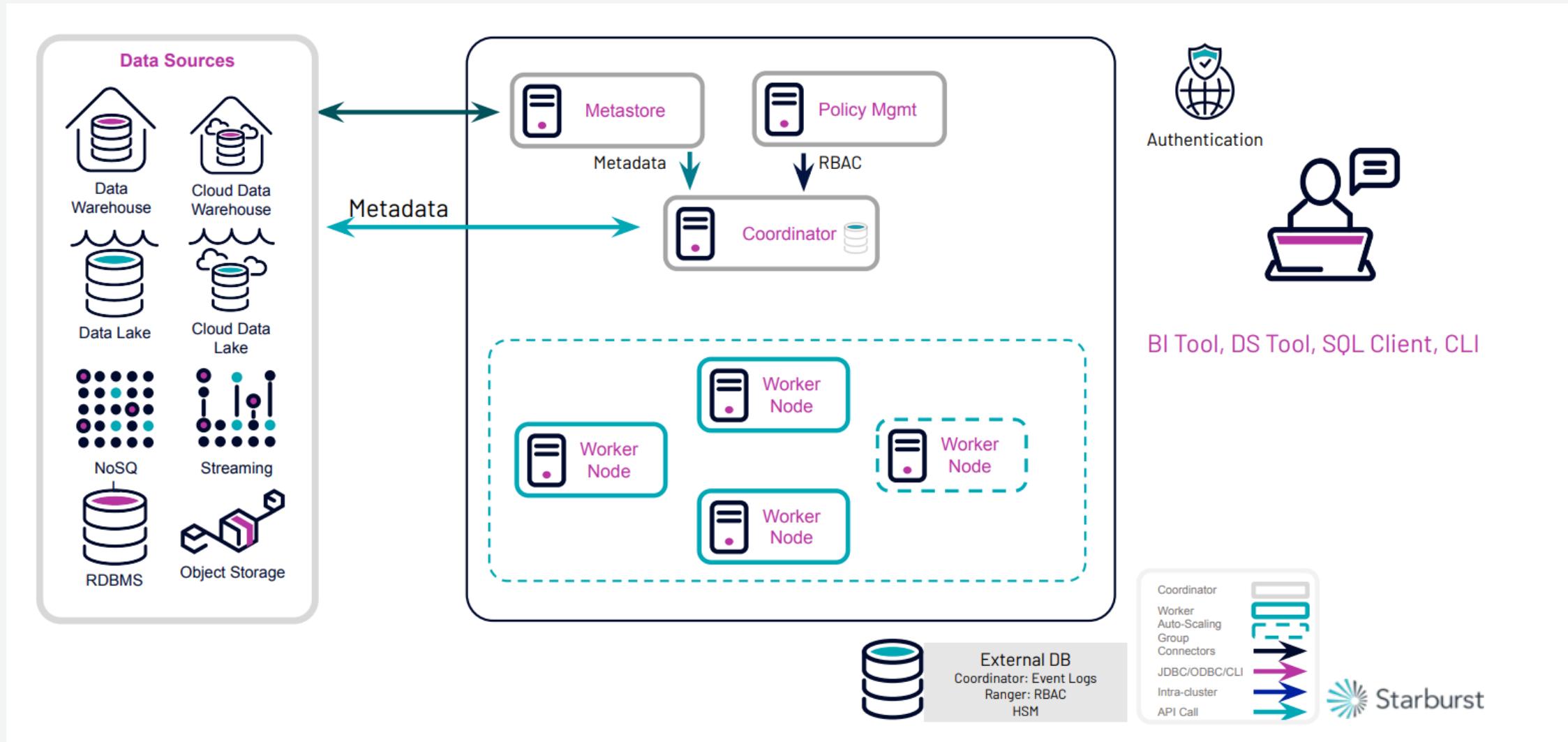
- They want GenAI to deliver value – but they lack the data governance and democratization to make it a reality.
- Tech is the secret to getting ahead – but they need to align their tech agendas with their strategic agendas.
- Environmental and geopolitical challenges threaten business models and performance – but data is the secret to mitigating risk.



EIGHT CEO PRIORITIES FOR 2024 | SOURCE: MCKINSEY

Building a data culture

Adaptability and speed with adopting new technologies such as Starburst is critical to building a strong data culture



Section 03

Building Strong Capability

Building a Center of Excellence



Guiding the transformation from technology novice to expert



Challenges:

- Lack of focus
- Lack of knowledge
- Dependency on other teams and tools
- Building an enterprise-wide standard

1. Start small

- You don't need to know every feature.
- Focus on the business-critical use cases and grow as needs arise

2. Build a repository of resources

- Build your knowledge of how things work and then configure generic documentation accordingly.
- Document your knowledge and use cases as you go to make it repeatable elsewhere.
- Find a structure - if it's not digestible, it won't be used.

3. Understand the surrounding technology

- Trino/Starburst is not the challenge - how it integrates with other systems is.
- Understand what you are working with to engage the right stakeholders at the start.

4. Be approachable and helpful

- Make the business want to come to *you* to seek your advice.
- It will help ensure alignment across teams with the goal of the CoE.

Section 04

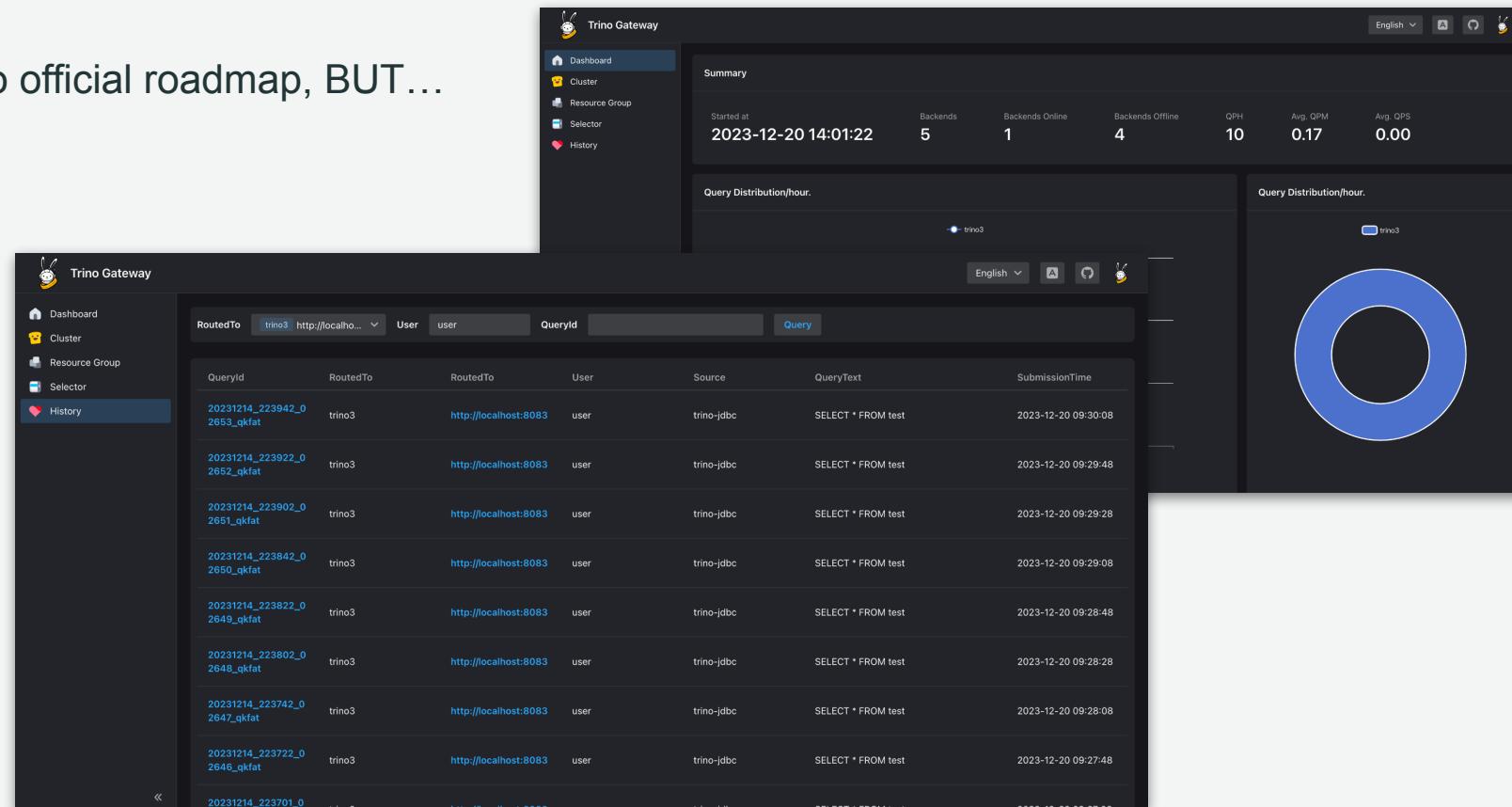
Trino Roadmap & Community

Trino Roadmap

Remember this is open source so there's no official roadmap, BUT...

Potential Roadmap Highlights

- Trino Kubernetes Operator
- Trino Gateway UI
- Parquet column encryption
- Snowflake connector
- Json_table
- Project Hummingbird
- Improved testing & CI



Sources

Most Recent Contributor Meeting <https://github.com/trinodb/trino/wiki/Contributor-meetings#trino-contributor-call-1-feb-2024>

Trino Summit Keynote

<https://www.youtube.com/watch?v=pXdZqpwgdxA>

Trino Community

Getting started, talking shop, learning more, and enabling yourself.

kubrick

Trino Blog

<https://trino.io/blog/>

Helpful articles and updates relevant to progress and usage of Trino

Trino Community Broadcast

<https://trino.io/broadcast/>

A monthly-ish broadcast highlighting exciting activity, use cases, contributors

Trino Slack

<https://trino.io/slack.html>

A monthly-ish broadcast highlighting exciting activity, use cases, contributors



Trino Community



Getting started, talking shop, learning more, and enabling yourself.

Trino Events

Trino Summit, Trino Fest, Meetups, and more.

<https://www.starburst.io/learn/events-webinars/>
<https://trino.io/community.html>

Starburst Galaxy

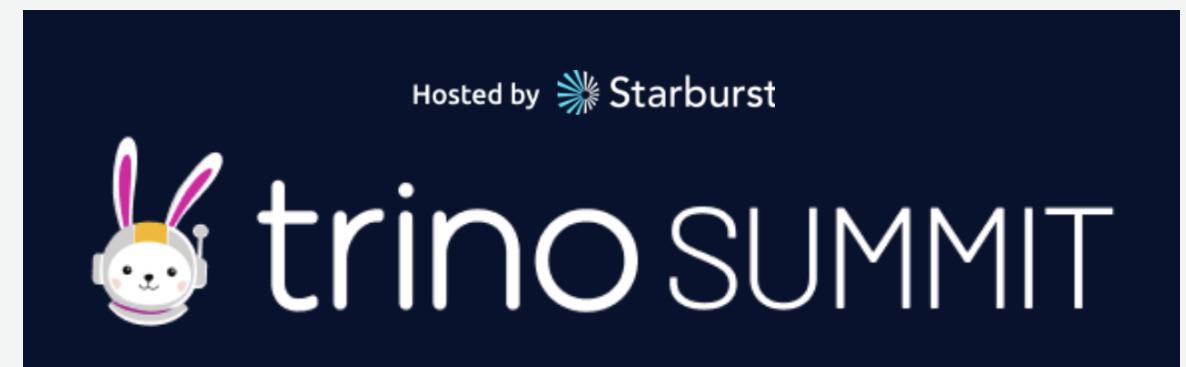
<https://www.starburst.io/platform/starburst-galaxy/>

Explore some Trino capabilities in a hosted platform.

Starburst Community

Starburst was founded by the founders of Trino and create useful content and events.

<https://www.starburst.io/community/>



Q&A



www.kubrickgroup.com
joinus@kubrick.com