

Trino Gateway

Trino Summit 2023

Agenda

- Mission statement
- Benefits
- History
- Use Cases
 - Naver (presenter: Jaeho)
 - Bloomberg (presenter: Andy)
- Demo (presenter: Will)
- Trino Gateway Project Future
- Contribute!
- Q&A (presenter: Will, Vishal, Andy, Jaeho)

Mission Statement

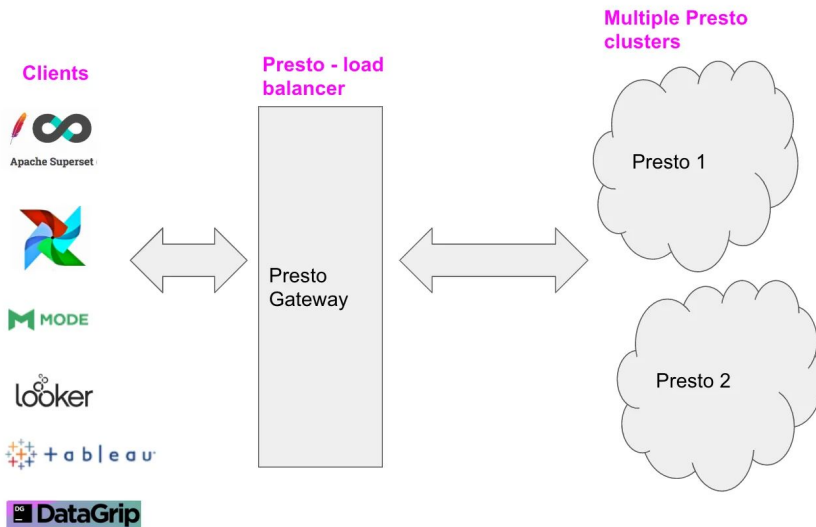
- Lightweight reverse proxy that **forwards requests** to distributed **Trino clusters**

Benefits to different stakeholders

- Benefits to cluster admins: configure an efficient and reliable infrastructure
 - Query Load Distribution
 - Consistent Querying Results
 - High Availability (HA) Support
 - Maintain Service Level Agreements (SLAs) for ETL Schedule Operations
 - Graceful Shutdown Support
- Benefits to end users: provide a reliable and fast query experience
 - Maximize performance and Service quality
 - Resource Failure Mitigation
 - Reliable Service During Deployment and Failures
 - Minimized Query Latency

History

- Show some love to Lyft
 - Originally [lyft/presto-gateway](https://github.com/lyft/presto-gateway)
 - First public release in Feb 16, 2019
 - [Presto Infrastructure at Lyft blog post](https://www.lyft.com/blog/presto-infrastructure-at-lyft)
- Starburst as new Maintainers
 - Special thanks to Manfred!
- Current contributors
 - Bloomberg
 - Naver
 - LinkedIn
 - Dune
 - Starburst



Naver Use Cases – Company

- [Naver Corporation](#) is a global ICT company, providing South Korea's number 1 search portal "NAVER".
- Offers various products, including a web search portal, a social media platform called [LINE](#), webtoon platform called [Line Webtoon](#) and diverse internet services.



WEBTOON, web novel, Blog, community, Knowledge iN, etc.

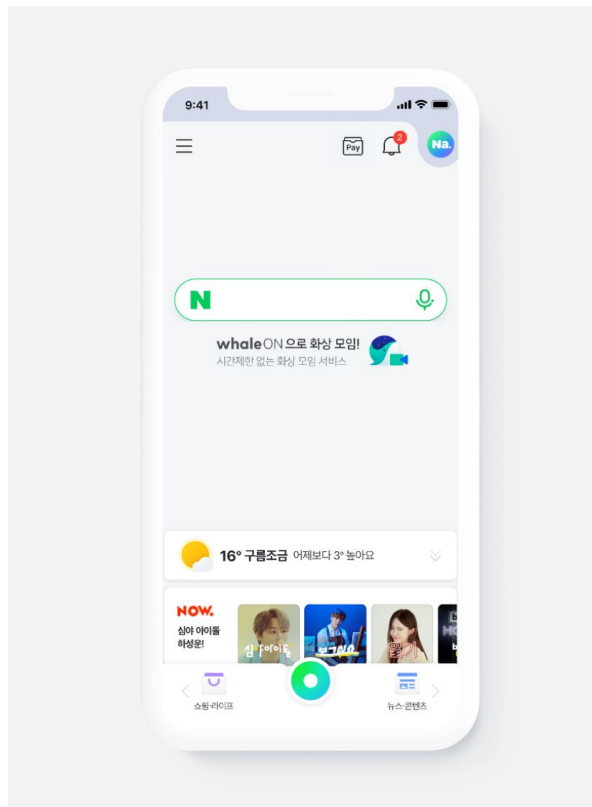
Naver Use Cases – Search Platform

Search Platform Team

- Provide optimal search results for
 - News, maps, cafe, blogs, images, videos
- Multi-ranking and collection-ranking technology

Statistics for Trino used by Analysts

- 4K queries, 300 TB data, 10 Trillion rows per day.
- 10 PB business data stored in our data centers



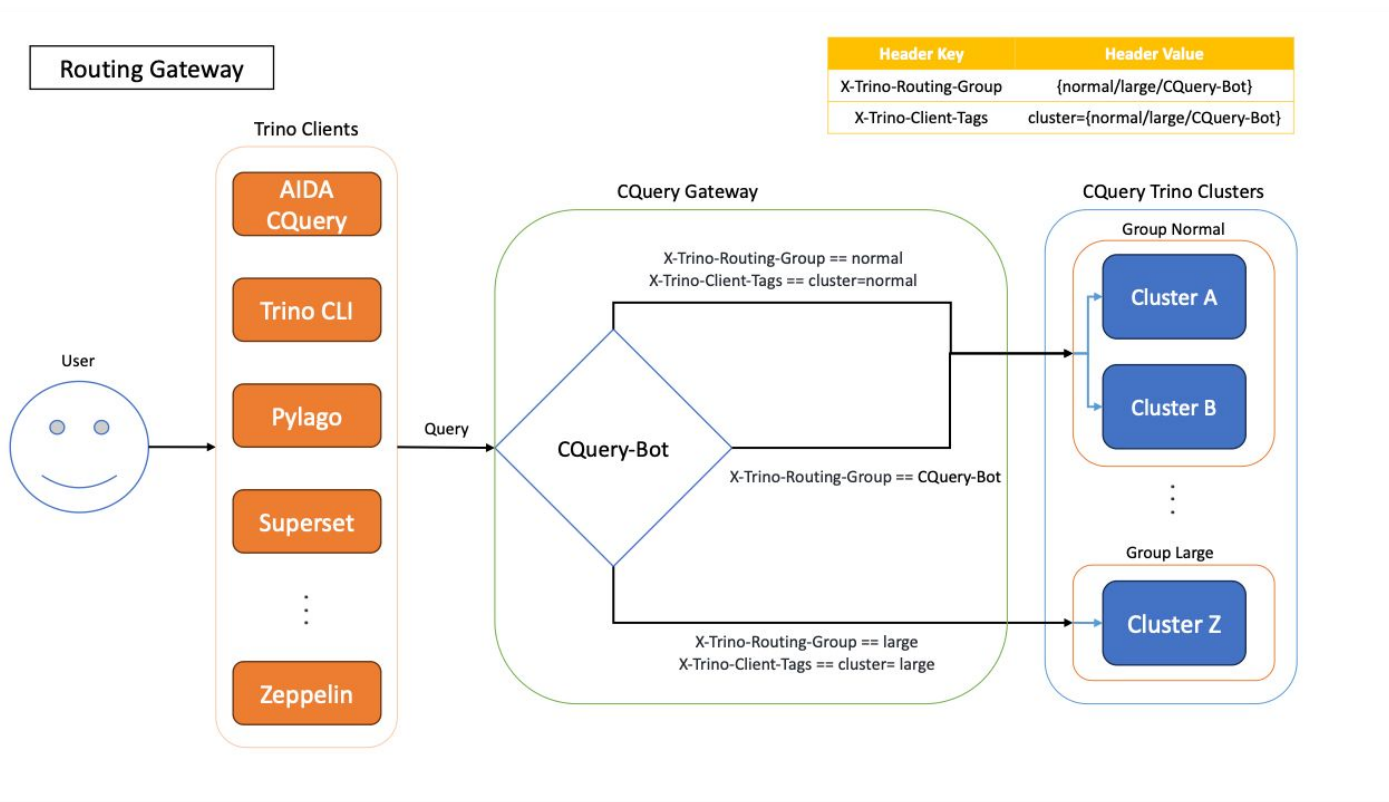
Naver Use Cases

- Managing Trino Clusters with Various Purposes
 - Problem
 - We provide trino clusters for various purposes: fast querying vs large datasets
 - Each cluster has their own url for different purposes
 - Any server updates forces users to switch to other clusters or wait until server was back up
 - Solution
 - Long and complex Trino Cluster URLs are combined into a single one.
 - Users now can query through a unified URL, avoiding disruptions during system updates.
 - Able to shutdown gracefully and update clusters

Naver Use Cases

- Need for Dynamic Routing depending on Trino Cluster's Status
 - Problem
 - Too many queries are requested at once, leading poor performance due to resource starvations.
 - A query might need more memory than the query limit set for each worker node
 - A query might need more memory than the workers' physical memory.
 - Failing after running for certain period of time is a waste of resources
 - Solution
 - Before routing
 - Phase 1 - consider current cluster's queued status and use [Resource Groups](#) to force CPU/Memory/Concurrency limits
 - Phase 2 - (TBD) Calculate expected memory & cpu usage for each query and dynamically route to appropriate cluster

Naver Use Cases - Architecture



Bloomberg



	2018 Q4	2019 Q1
Revenue	348,468,032	348,468,032
Operating Profit	296,598,349	288,763,237
Net Income	255,316,824	255,316,824
Adjusted Earnings	190,650,277	190,650,277
Operating Expenses	100,999,544	100,999,544
Depreciation	64,636,164	64,636,164
Amortization	61,063,812	61,063,812
Goodwill Impairment	59,278,139	59,278,139
Other	47,848,838	47,848,838
Non-Operating Income	45,675,438	45,675,438
Other	41,861,529	41,861,529
Income Tax Expense	41,987,370	41,987,370
Other	41,095,884	41,095,884
Net Income	37,295,930	37,295,930
Other	32,000,700	32,000,700

TechAtBloomberg.com

Why and How we use Trino and Trino Gateway at Bloomberg

- Query federation with custom access controls
 - Access different data sources such as Postgres or SQL
 - Access control via Trino OPA plugin
- Trino during ETL process
 - Create intermediate datasets or final products
- Store high volume data products
 - SQL on S3 via Apache Iceberg/Hive for data products
- Served **tens of millions** of Trino queries to date
- High availability is our priority

Bloomberg Use Cases

- App Security
 - Supports OAuth authentication
 - Supports Form authentication
 - Basic
 - LDAP

TechAtBloomberg.com

© 2023 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Bloomberg Use Cases

- Health checks
 - Problem
 - Clusters were inactivated programmatically without turning back on after they became healthy
 - Solution
 - Introduced Healthy, Pending, and Unhealthy states to prevent a cluster from automatically being turned off
 - Achieved high availability

Bloomberg Use Cases

- Dependency modernization
 - Problems
 - Core dependencies were too old and no longer supported
 - Difficult to run integration tests because we could not use the DropwizardExtensionSupport extension support from the library
 - Solution
 - Upgraded dependency versions and dropped duplicate dependencies
 - Dropwizard v1.3.29 -> v4.0.2
 - Java 11 -> Java 17
 - Jetty v9.4.48 -> v11.0.15
 - Javax -> jakarta
 - TestNG/junit5 -> junit5

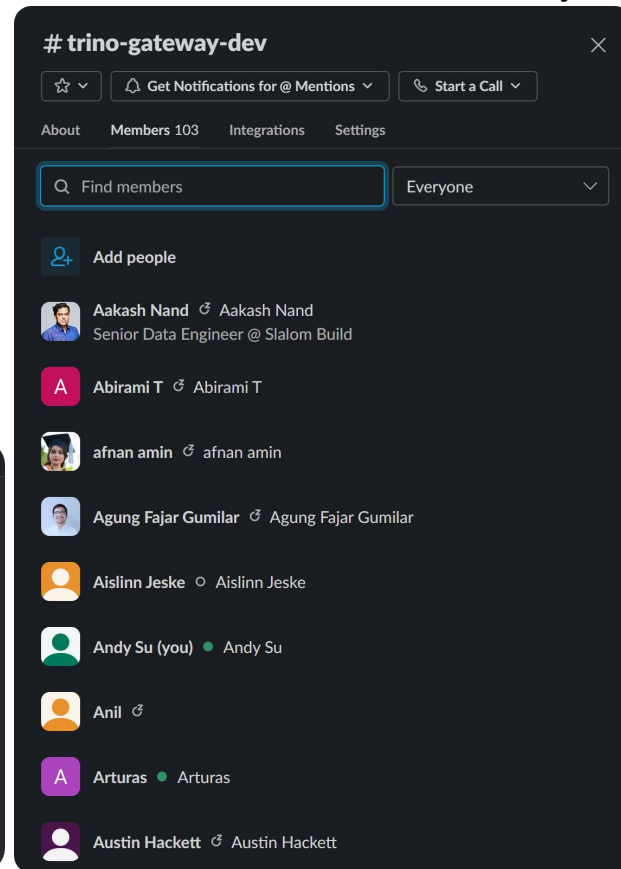
Demo

Trino Gateway Project Future

- Short-term goals
 - Modernizing dependencies continued
 - Improving test suite
 - Aligning code styles and patterns with Trino
 - Supporting modern deployment tools
 - Tracing
 - Containerization
- Medium-term and long-term goals for capabilities and improvements
 - Automatic query retry to other clusters if backend goes offline or due to lack of resources (memory)
 - Integration with group providers for routing
 - Supporting for parallel testing (e.g. duplicate a query to a second cluster and consume result)

Contributions

- Trino Gateway Homepage
 - [trinodb/trino-gateway \(github.com\)](https://trinodb/trino-gateway)
- Where to find us
 - <https://trinodb.slack.com/archives/C059KUNPTSP>



#trino-gateway-dev

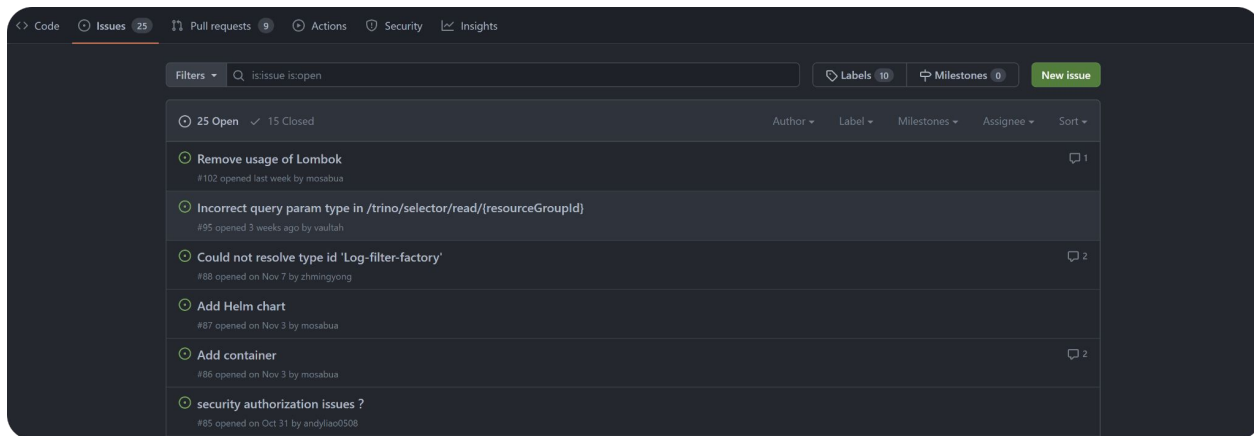
☆ Get Notifications for @ Mentions Start a Call

About Members 103 Integrations Settings

Find members Everyone

Add people

- Aakash Nand Aakash Nand Senior Data Engineer @ Slalom Build
- Abirami T Abirami T
- afnan amin afnan amin
- Agung Fajar Gumilar Agung Fajar Gumilar
- Aislinn Jeske Aislinn Jeske
- Andy Su (you) Andy Su
- Anil
- Arturas Arturas
- Austin Hackett Austin Hackett



Code Issues 25 Pull requests 9 Actions Security Insights

Filters is:issue is:open Labels 10 Milestones 0 New issue

Issue	Author	Label	Milestones	Assignee	Sort
Remove usage of Lombok #102 opened last week by mosabua					1
Incorrect query param type in /trino/selector/read/{resourceGroupId} #95 opened 3 weeks ago by vaullah					
Could not resolve type id 'Log-filter-factory' #88 opened on Nov 7 by zhmingyong					2
Add Helm chart #87 opened on Nov 3 by mosabua					
Add container #86 opened on Nov 3 by mosabua					2
security authorization issues ? #85 opened on Oct 31 by andyliao0508					

