

Let it SNOW for Trino

Teng YU & Erik Anderson

<https://github.com/trinodb/trino/pull/17909>





First part highlights



01

Introductions

02

Introduce Snowflake

03

**Why ForePaaS needed
Snowflake Connector**



04

Features of the connector



Who we are?



- I am Teng YU, a Full-Stack Developer at ForePaaS
 - <https://www.linkedin.com/in/tyu-fr/>
- ForePaaS is a modern data platform for creating, deploying and scaling analytics & AI projects on the cloud
 - **2014:** ForePaaS is founded in Paris, France
 - **2018:** ForePaaS expands in the United States with a 1st SF office
 - **2022:** ForePaaS is acquired by OVHcloud, the leading European cloud provider



Who we are?



Current organization: **Development** EN ⚙️ ⓘ 🖨️

Platform Home

FAVORITES

- Control Center (MUTUALIZED)
- Identity Access Manager (MUTUALIZED)
- Data Manager (DEVELOPMENT)

ALL SERVICES

- Data Manager
- Data Processing Engine
- Analytics Manager
- Machine Learning Manager
- Applications
- APIs
- Billing Explorer
- Storage Engines
- Control Center
- Identity Access Manager

All services

End-to-end Projects	<ul style="list-style-type: none"> Projets End-to-end & ready-to-use data environments 	<ul style="list-style-type: none"> End-to-end Projects Analytics
Analytics	<ul style="list-style-type: none"> Data Manager Poly-engine data lake and data warehouse Data Processing Engine Create & run ETL data pipelines Analytics Manager Data visualization for enterprise analytics 	<ul style="list-style-type: none"> Artificial Intelligence Application Services Admin Tools
Artificial Intelligence	<ul style="list-style-type: none"> Machine Learning Manager Train and deploy machine learning models 	
Application Services	<ul style="list-style-type: none"> Applications Create and deploy applications featuring analytics and AI APIs Dedicated APIs to implement your custom code logic 	
Admin Tools	<ul style="list-style-type: none"> Billing Explorer 	











Introduce Snowflake





Introduce Snowflake



- Cloud-based data warehousing and analytics platform
 - Scalability, flexibility and performance
 - Separating compute and storage, allowing organizations to scale their computing resources independently
- 
- 
- 
- 
- 
- 



03

Why ForePaaS needed Snowflake connector

Select a storage engine to start with

The storage engine is the technology powering your data lake & data warehouse. You can always use others later on.



Use PostgreSQL

Dedicated storage engine running on PostgreSQL. You can specify the location and resources to allocate.



Use Snowflake

Dedicated storage engine running on Snowflake. You can specify the location and resources to allocate.

You don't have any instance here.



NEW STORAGE ENGINE



CONNECT AN ACCOUNT Coming soon!



Sandbox

This is a basic testing environment. Not recommended for a production-ready Dataplant.



Use your organization's default configuration

This is your default storage engine configuration. You can customize it in your organization's settings.

Queries > Preview: chicago_calendar_full

VISUAL **SQL**

All modifications are saved

CONFIGURATION

PREFERENCES

METRICS

Version: v1

Hide table details

CHICAGO_CALENDAR_FULL

Search for an attribute

- autumn_holidays NUM
- bridge NUM
- christmas_holidays NUM
- cloud_cover NUM

```

1 SELECT
2 *
3 FROM chicago_calendar_full

```



DISPLAY 5000 ROWS

RUN



autumn_holidays	summer_holidays	spring_holidays	public_holiday	week_day_label	month	wind_speed	humidity	christmas_holid
0	0	0	2	Thursday	1	6.12	0.56	1
0	0	0	0	Friday	1	6.83	0.61	1
0	0	0	0	Saturday	1	1.01	0.7	1
0	0	0	0	Sunday	1	0.73	0.9	1
0	0	0	0	Monday	1	5.82	0.85	0
0	0	0	0	Tuesday	1	4.63	0.71	0
0	0	0	0	Wednesday	1	5	0.77	0
0	0	0	0	Thursday	1	7.62	0.62	0
0	0	0	0	Friday	1	7.27	0.77	0
0	0	0	0	Saturday	1	7.77	0.69	0



OVERVIEW



QUERIES



DASHBOARDS



CONSUMERS



HISTORY



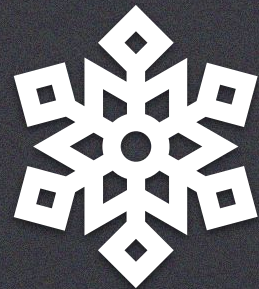
RESOURCES



SETTINGS







04

Connector Features





Connector Features

- A full feature trino connector
 - Bridge between Trino and Snowflake
 - Leveraging the power of Trino's distributed query engine and the scalability, flexibility, and security features of Snowflake.
 - Using Snowflake's optimized architecture and distributed computing capabilities, ensuring high performance and scalability for processing large datasets.
 - Thanks for Trino community
- 
- 
- 
- 
- 
- 



trino



Second part highlights



01

Introduce self and Bloomberg

02

Bloomberg's need for
Snowflake

03

Demo time !



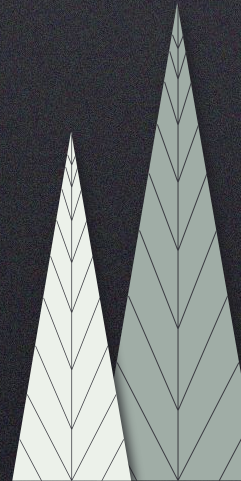
Erik Anderson, Bloomberg Engineering

01

- Erik Anderson, Software Engineer. Bloomberg Enterprise Data/API's, Services, and of course Trino ecosystem!
- <https://linkedin.com/in/erikanderson>
 - Please mention this TrinoFest talk in your request or
- Software Engineer for over 30 years (aka a Senile Engineer)
- Been working in open source for 20+ years and Trino open source for ~3 years
- Standard Financial Services Disclaimer: My MANY opinions are my own

<https://www.bloomberg.com/company/values/tech-at-bloomberg/open-source/>



"We believe open source software is a key driver of innovation not only within our business, but across the global tech industry. Over the last two decades, Bloomberg has undergone a journey and cultural shift in becoming an 'open source first' company. Today, our engineers are actively engaged in the open source ecosystem as both users and contributors. Hundreds of global employees, from senior leadership down, are encouraged to get involved in our open source efforts, with Bloomberg's Open Source Program Office (OSPO) within the Office of the CTO spearheading our efforts to produce, publish and support open source software."





02

Bloomberg's need for

- Its open source!
 - I dont discuss Bloomberg business use cases but to make up for that I will show an interactive demo of how to use the Snowflake connector to solve an interesting business data analytics challenge
 - I will use that connector to publish the new derived content to Snowflake cloud services
- 
- 



03



Demo time !

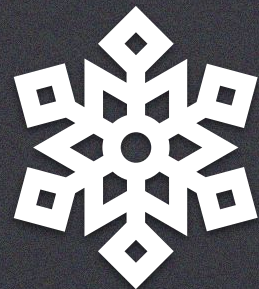
- https://github.com/bloomberg/trino/tree/trinofest_june_2023
- TrinoFest Let it Snow SQL's
 - All SQL for this demo is in this gist link
 - <https://gist.github.com/dprophet/62fa4b569ee253adb125da25c1979be6>



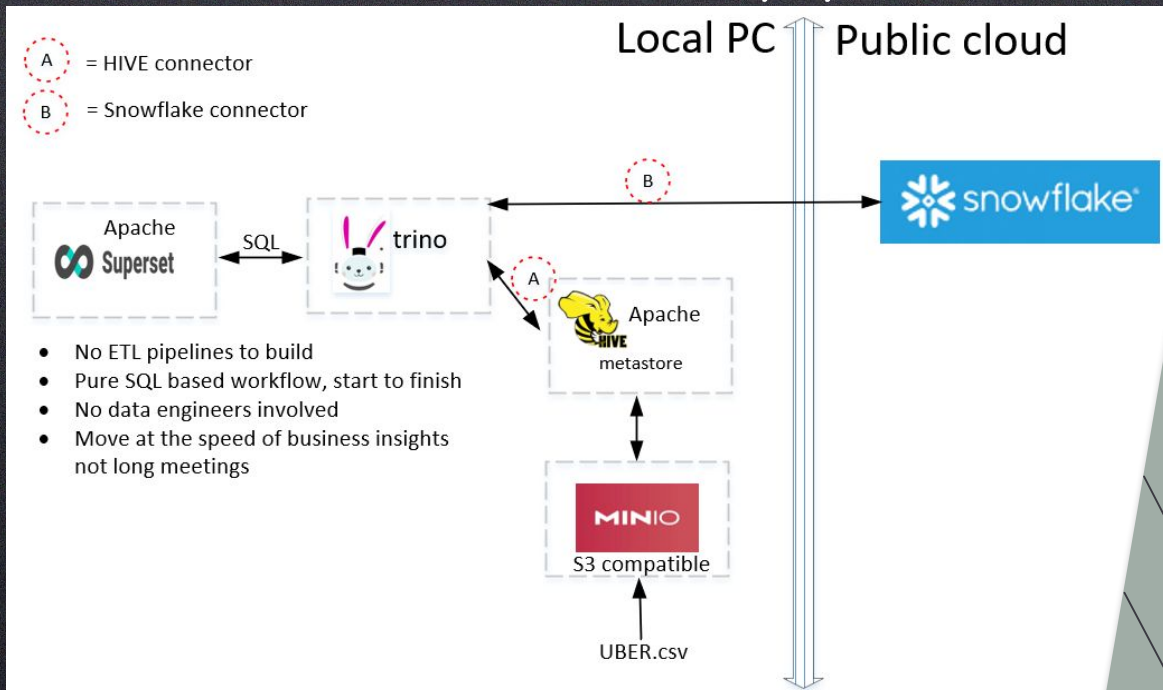
Demo:



open source demo



- Local PC
 - Download data set <https://www.kaggle.com/datasets/varpit94/uber-stock-data?resource=download>
 - Load the data into MinIO+Apache HIVE Metastore
 - Using Apache Superset we will create new derived dataset known as trading envelopes for stock trading triggers. All SQL driven workflow. No long meetings, no data engineers involved.
- Public Cloud
 - Publish that new dataset from PC to Snowflake cloud using pure SQL
 - Create a dashboard of the new Snowflake dataset to visually verify our work





THANKS!

Does anyone have any questions?

