Choosing, Using and Extending Trino as a Primary Datastore
Hello!

I am Rob Dickinson
CTO at Resurface Labs

You can find me at @robfromboulder
Our Use Case

Purpose-built database for API traffic
To see API failures, threats & slowdowns
Because guessing is lame

Workload characteristics:
Writes – sequential appends (lots)
Deletes – drop partition (periodically)
Updates – only through views (as needed)
Reads – query all columns / all rows 😐
Super-columnar queries (all columns)
Let’s talk about Trino...
What’s Trino
(formerly known as PrestoSQL)

- If you know Presto…
- If you’ve been using PrestoSQL…
- … same software, same people, same community
- … under a shiny new name and a cute bunny
Why We ❤ Trino

Trino hides data sources behind SQL
Trino offers advanced SQL features
Trino pushes queries down to data
Trino is easy to extend
Trino scales up and down well
Trino has advantages over Presto
Trino hides many sources of data behind SQL

Behind the curtain:
Accumulo, BigQuery, Cassandra, ClickHouse, Druid, Elastic, Iceberg, Hive, Kafka, Kinesis, Kudu, Mongo, MySQL, Oracle, Phoenix, Pinot, Postgresql, Prometheus, Redis, SQL Server
Trino supports advanced SQL

- Not just some reduced subset of SQL
- Ouch-that-hurts-my-brain SQL
- Inner/outer joins and subqueries (even across datastores)
- Built-in JSON and GEO functions
- Histogram function is awesome 😍
- Recent additions: OFFSET, WINDOW, RANGE, UNNEST
Trino pushes queries down to data

Not just proxying!
Distributed query planning
Predicate pushdown
Minimize data pulled across network
Trino is easy to tweak and extend

Not just connectors!
Query functions
Views, baby, views
Apache licensed
Trino scales up and down

- Trino deployments in the wild scale to 20k+ nodes
- Which is great if you’re Facebook (or trying to be)

- Most of us need cost efficiency with least caretaking
- Trino runs on a single node, runs on cheap nodes
- Doesn’t have the weight of Hadoop, Zookeeper
Resurface Pilot Edition

By the numbers:
- 4 GB all-in-one container
- 250 MB download
- Run any query in ~1 sec

Diagram:
- SQL Clients
- Trino Coordinator/Worker
- Microservice
- Loggers
- Single container
- Binary files
- Connector
Scaling up to many workers
Trino has the upper hand over Presto

- Majority of the original founding contributors
- 461 current contributors
- 4036 Slack members (500+ weekly)
- 55 releases since split (one every two weeks)
- 8043 commits since split
Trino is getting more commits
Trino community growth after rename
The word is out about Trino
THANK YOU!

PERCONA
LIVE ONLINE
MAY 12 - 13th
2021
Thank you!

Any questions?

You can find me at @robfromboulder or rob@resurface.io