Building Reliable & Scalable Data Replication Pipelines from Postgres on Kubernetes

Sanketh Balakrishna

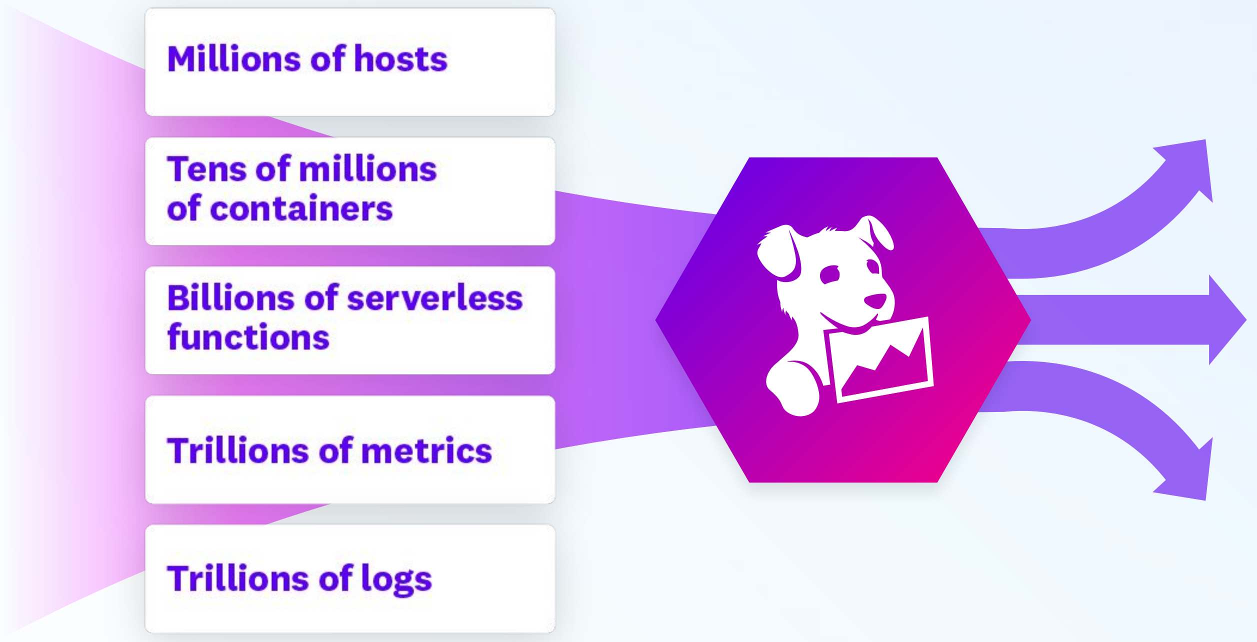
Postgres Conference Orlando March 2025

DATADOG

Hello, I’m Sanketh

- Bangalore, India -> Lowell, Massachusetts
- Engineering Manager at Datadog - MaRS
- Racket Sports (Pickleball!!)
- Traveling/Outdoors (Hiked Half Dome)

Enabling data analysis at scale


Millions of hosts
Tens of millions of containers
Billions of serverless functions
Trillions of metrics
Trillions of logs

PLATFORM SERVICES

- Dashboards
- Agents
- Collaboration
- Mobile
- Workflows
- Watchdog AI
- Open Telemetry

PRODUCTS / USE CASES

Infrastructure | APM | DBM
Log Management | Cloud SIEM
CI Visibility | Continuous Profiler
RUM | Network | Synthetics
Cloud Security Management
App Security Management
Observability Pipelines
Cloud Cost Management
... and more

Platform at Datadog

We deal with large amounts of Data

Platform teams enable faster iterations for Product

Core Database engines are integral to a reliable platform

Strong affinity to use and contribute to open source products

Platform - Going back a couple of years

Lots of teams had data in Postgres

It's just easy, reliable and open source

- Community support is strong
  - Teams had experience administering it

Platform - Going back a couple of years

Lots of teams had data in Postgres

It’s just easy, reliable and open source

- Community support is strong
  - Teams had experience administering it

Investing in PGK platform

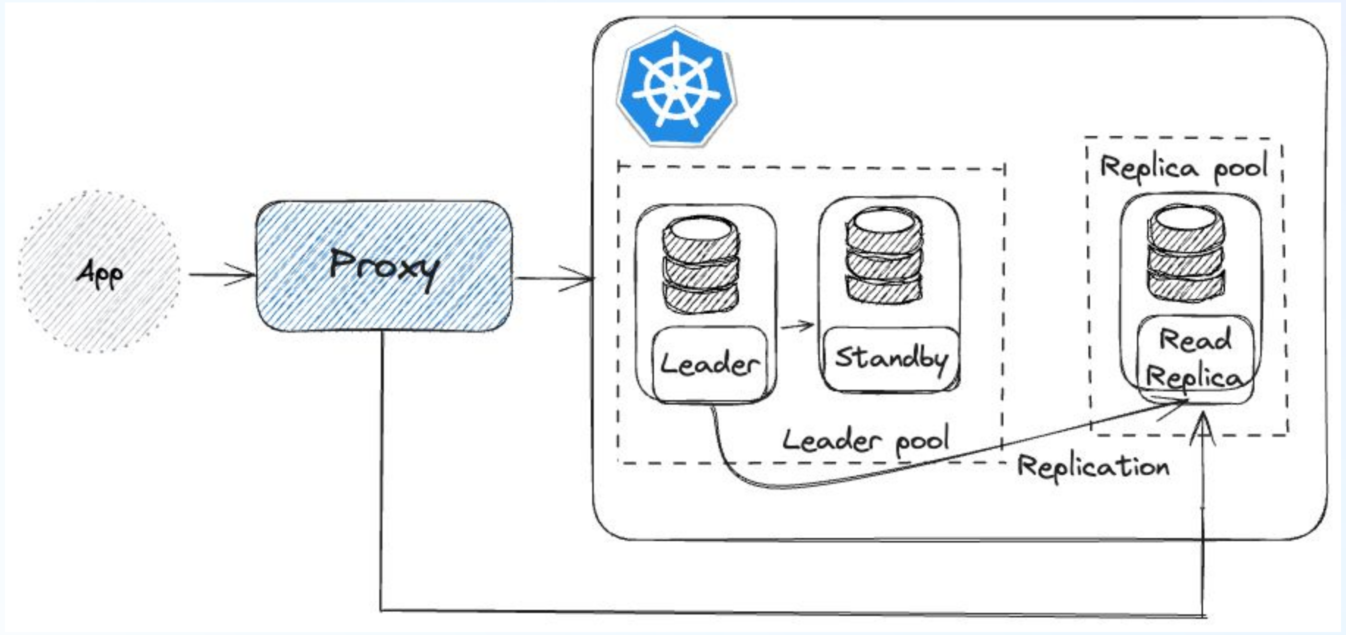
Platform - Going back a couple of years

Lots of teams had data in Postgres

It’s just easy, reliable and open source

- Community support is strong
  - Teams had experience administering it

Investing in PGK platform



Platform - Going back a couple of years

Lots of teams had data in Postgres

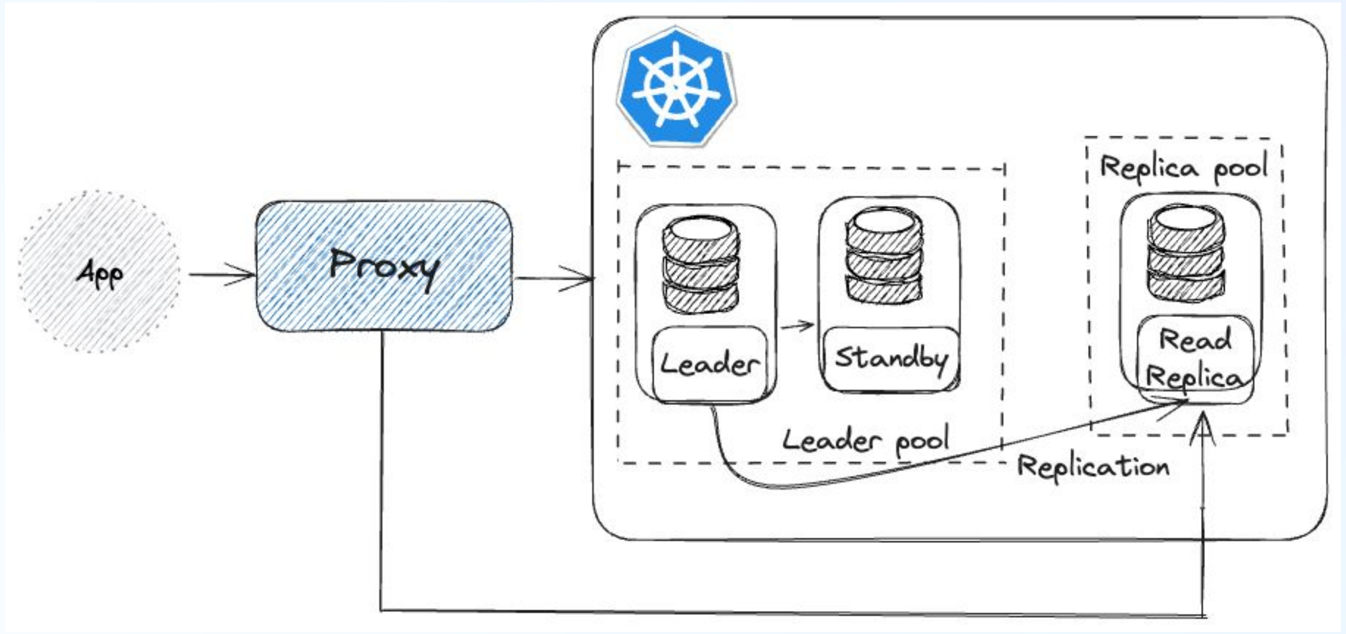
It's just easy, reliable and open source

- Community support is strong
  - Teams had experience administering it

Investing in PGK platform

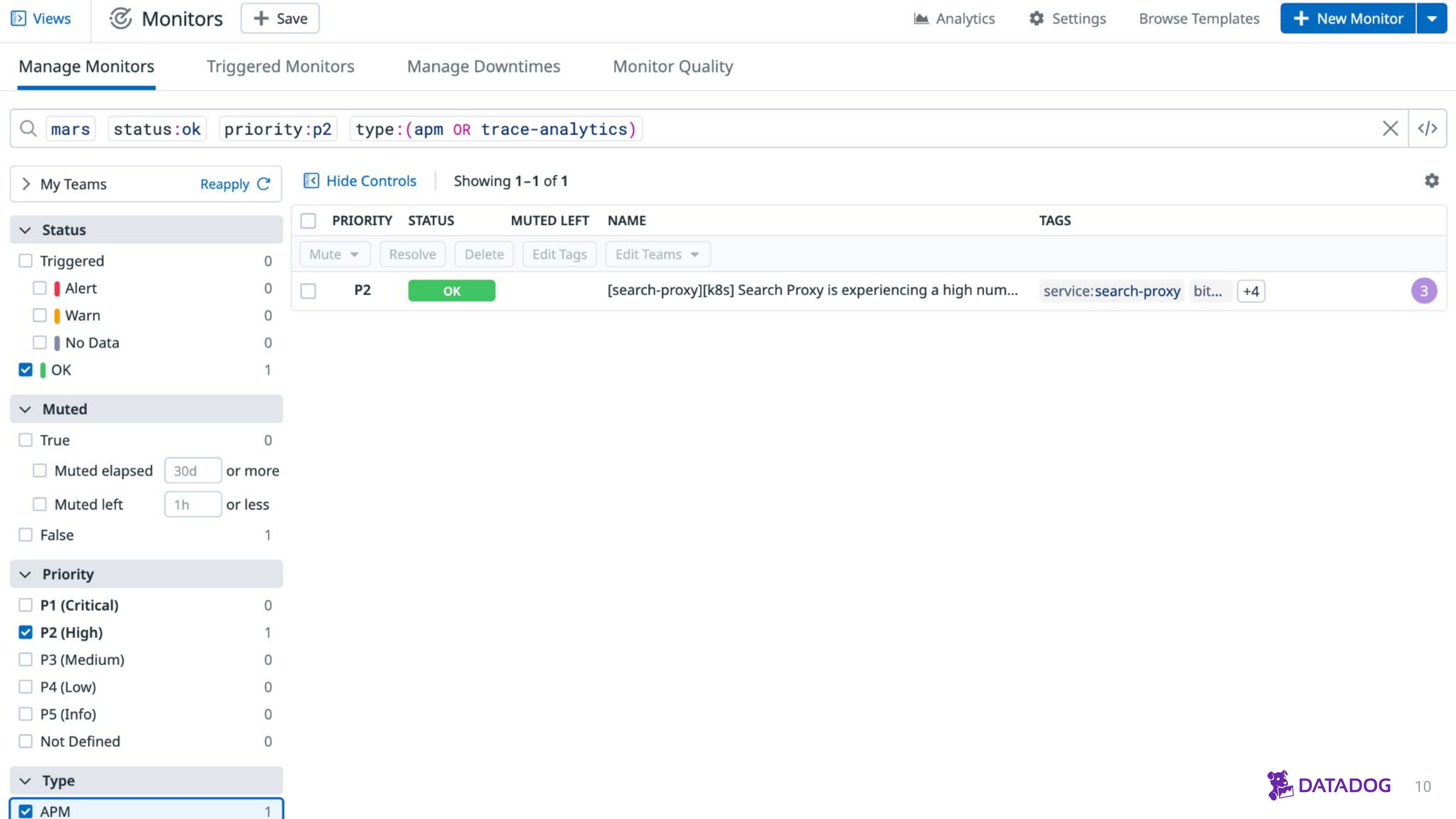
Lot's of flexibility with architecture

- Growth necessitated structure
  - Upgrade legacy and break monoliths



Platform - Going back a couple of years

Search was increasingly important



Views | Monitors | + Save | Analytics | Settings | Browse Templates | + New Monitor

Manage Monitors | Triggered Monitors | Manage Downtimes | Monitor Quality

```
mars  status:ok  priority:p2  type:(apm OR trace-analytics)
```

My Teams Reapply

Status

☐	Triggered	0
☐	Alert	0
☐	Warn	0
☐	No Data	0
☑	OK	1

Muted

☐	True	0
☐	Muted elapsed 30d or more	
☐	Muted left 1h or less	
☐	False	1

Priority

☐	P1 (Critical)	0
☑	P2 (High)	1
☐	P3 (Medium)	0
☐	P4 (Low)	0
☐	P5 (Info)	0
☐	Not Defined	0

Type

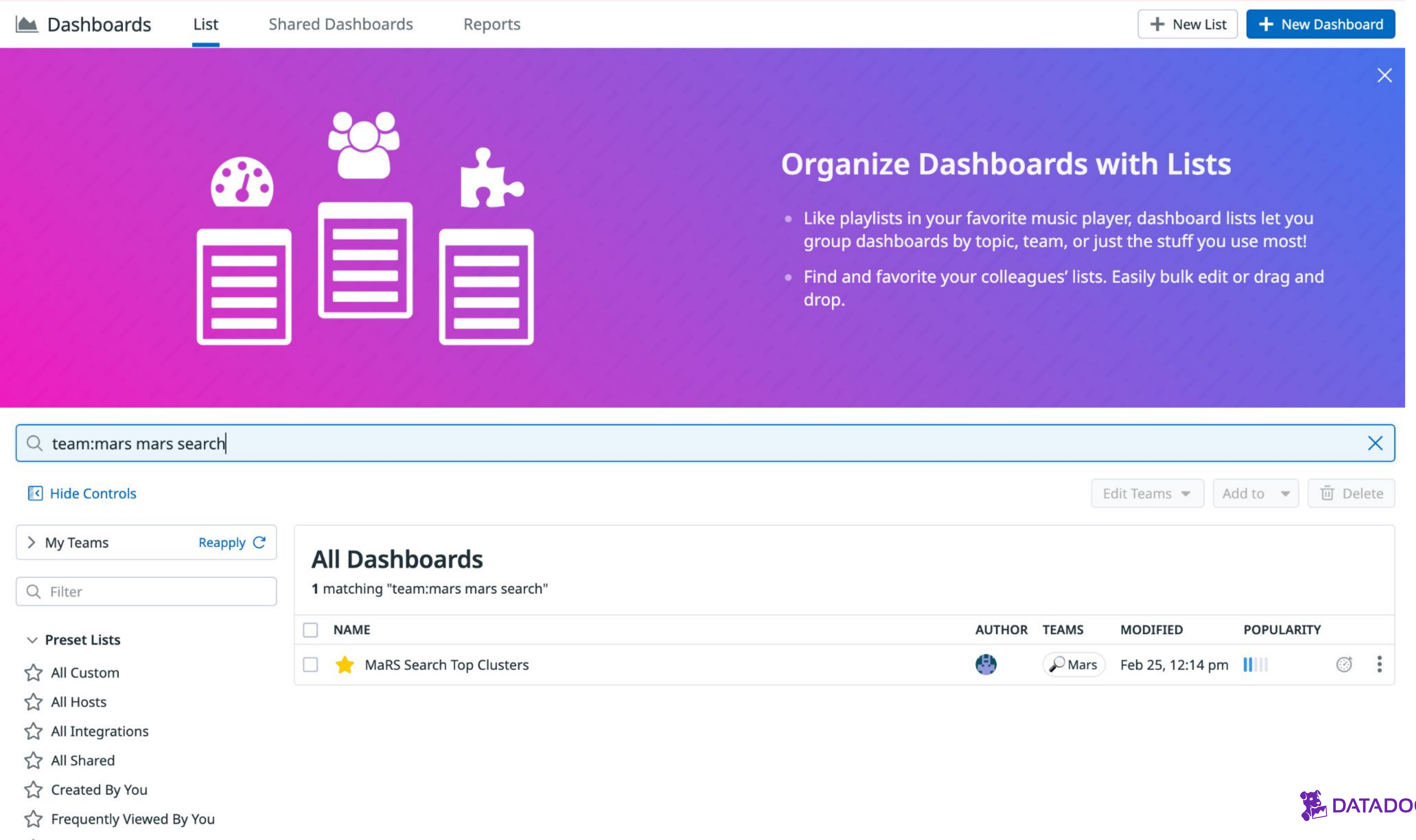
☑	APM	1

Hide Controls | Showing **1–1** of **1**

Mute | Resolve | Delete | Edit Tags | Edit Teams

PRIORITY	STATUS	MUTED LEFT	NAME	TAGS
P2	OK		[search-proxy][k8s] Search Proxy is experiencing a high num...	service:search-proxy bit... +4

DATADOG 10


Dashboards List Shared Dashboards Reports

New List New Dashboard

Organize Dashboards with Lists

- Like playlists in your favorite music player, dashboard lists let you group dashboards by topic, team, or just the stuff you use most!
- Find and favorite your colleagues' lists. Easily bulk edit or drag and drop.

team:mars mars search

Hide Controls

Edit Teams Add to Delete

My Teams Reapply

Filter

Preset Lists

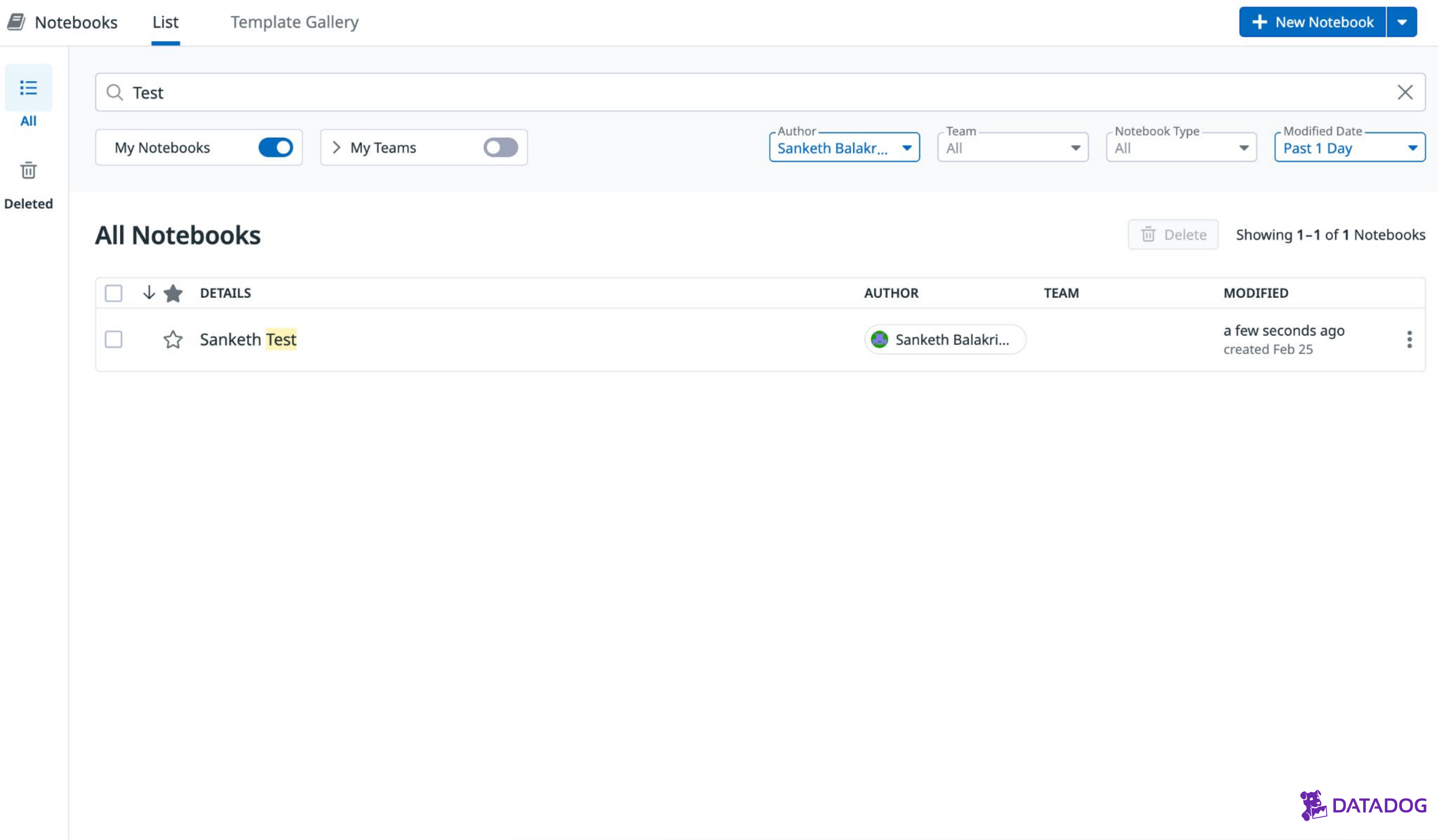
- All Custom
- All Hosts
- All Integrations
- All Shared
- Created By You
- Frequently Viewed By You

All Dashboards

1 matching "team:mars mars search"

NAME	AUTHOR	TEAMS	MODIFIED	POPULARITY
★ MaRS Search Top Clusters		Mars	Feb 25, 12:14 pm	

DATADOG


Notebooks | List | Template Gallery

New Notebook

All

Deleted

Test

My Notebooks [✓ on]

My Teams [off]

Author: Sanketh Balakr...

Team: All

Notebook Type: All

Modified Date: Past 1 Day

All Notebooks

Delete

Showing **1–1** of **1** Notebooks

DETAILS	AUTHOR	TEAM	MODIFIED
Sanketh Test	Sanketh Balakri...		a few seconds ago created Feb 25

DATADOG

Platform - Going back a couple of years

Search was increasingly important

Lot’s of Datadog UI’s have search bars

- Search limitations with Postgres
  - Search queries not intuitive

Platform - Going back a couple of years

Search was increasingly important

Lot’s of Datadog UI’s have search bars

- Search limitations with Postgres
  - Search queries not intuitive

Same data was needed by multiple teams

Varying requirements

- Different targets
  - Challenges with tightly coupled integrations

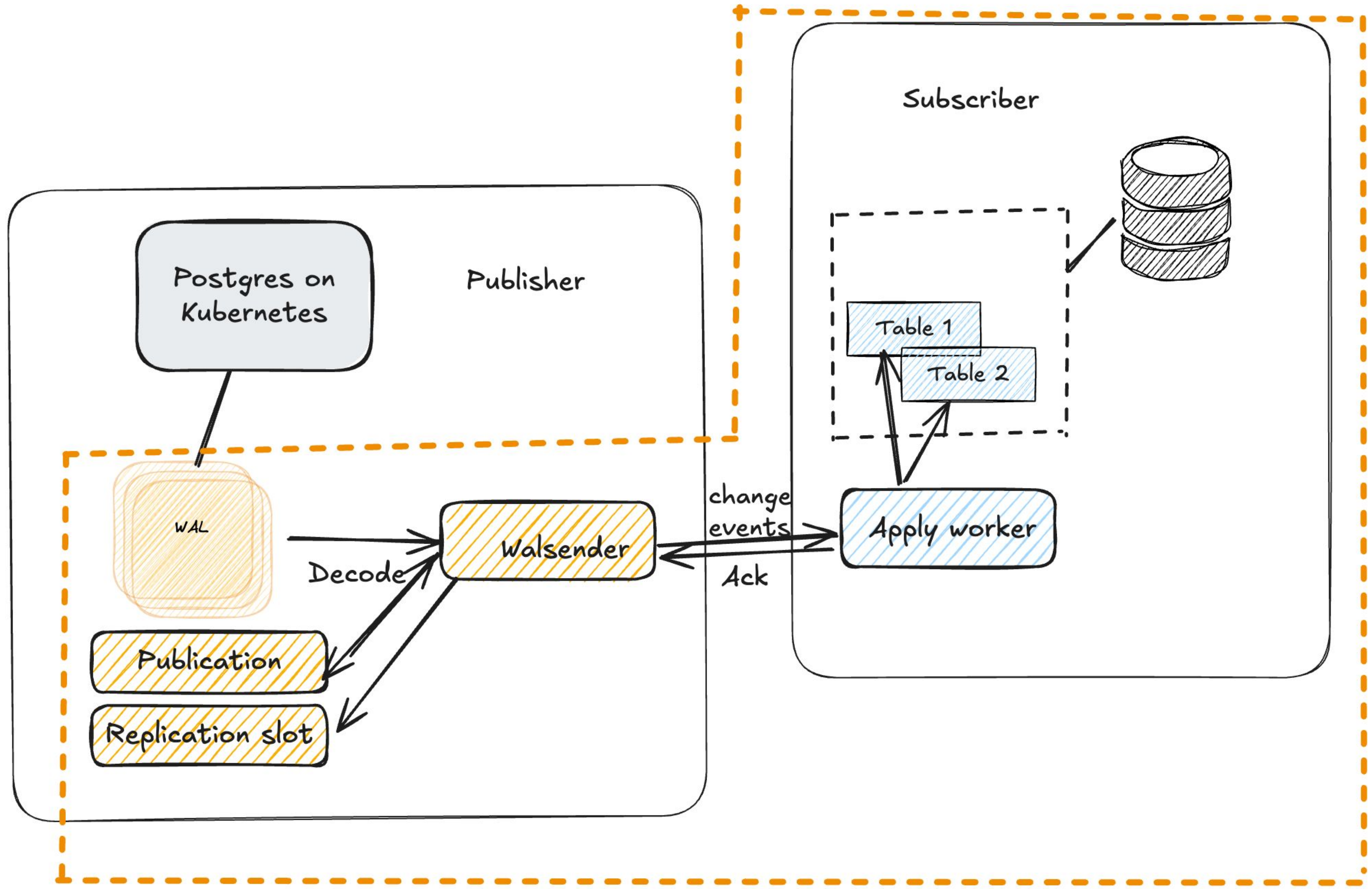
Requirements

Replicate data from Postgres

Scalable & Reliable

Target systems can vary

Logical Replication - An obvious choice



Data Replication - Postgres

Synchronous

Data Replication - Postgres

Synchronous Replication

Primary and secondary systems are synchronous

- Strongly consistent
- Less components to manage

- Less Flexible
- More prone to flakiness

Ex: PG -> PG Replication, PGSync

Data Replication - Postgres

Synchronous

Asynchronous

Data Replication - Postgres

Synchronous Replication

Data flow between primary and secondary systems is synchronous

- Strongly consistent
- Less components to manage

- Less Flexible
- More prone to flakiness

Ex: PG -> PG Replication, PGSync

Asynchronous Replication

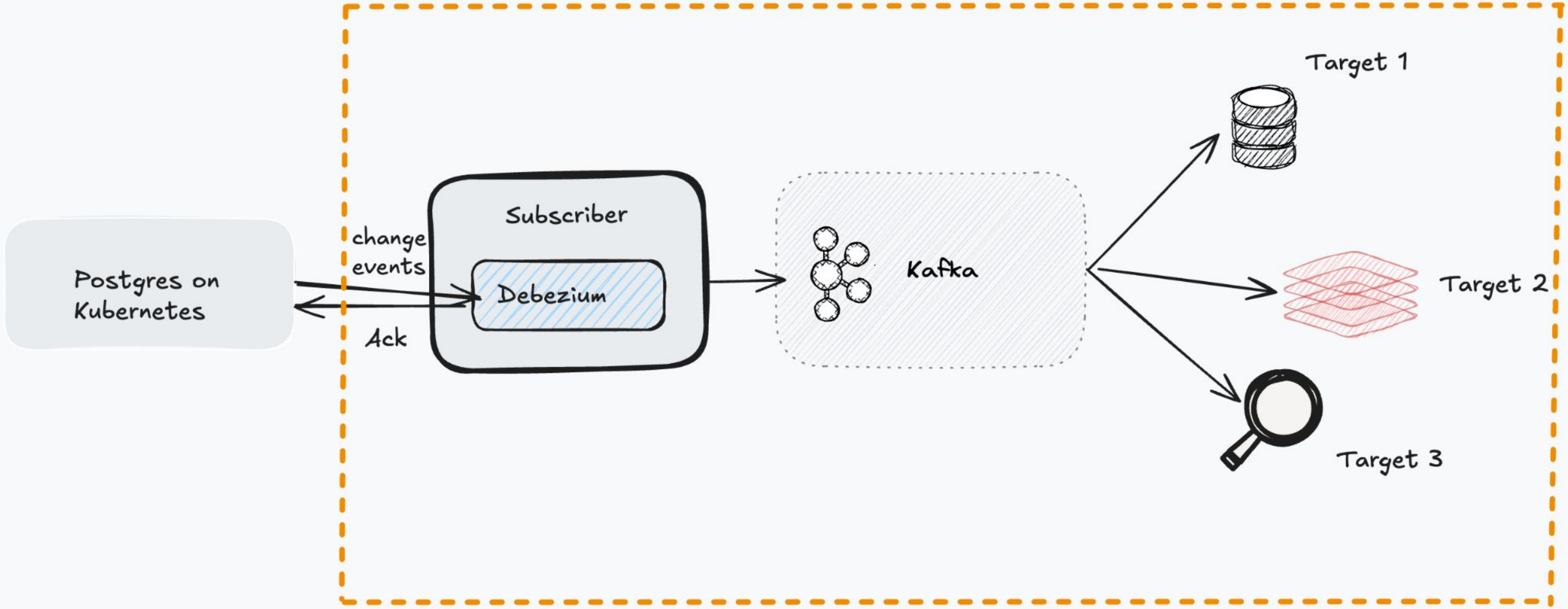
The subscriber is closer to Postgres

- Scalable
- Reliable

- More moving pieces
- Less consistent

Ex: Debezium based replication

Data Replication - Using Debezium



Requirements

✓ Replicate data from Postgres

✓ Scalable & Reliable

✓ Target systems can vary

Data Replication - Platform Challenges

Provisioning

Metrics/Traces

Schemas

Customization

Provisioning

1. PG Operations

High margin for error

```
      CREATE USER sanketh_debezium_user WITH PASSWORD
'<password>';

      ALTER USER sanketh_debezium_user CREATEDB REPLICATION;

      GRANT USAGE ON SCHEMA sanketh_schema TO
sanketh_debezium_user;

      // For all tables that need to be replicated
      GRANT SELECT ON sanketh_schema.my_table TO
sanketh_debezium_user;

      GRANT SELECT, UPDATE ON dbz_heartbeat IN SCHEMA
sanketh_schema TO sanketh_debezium_user;
```

DATADOG

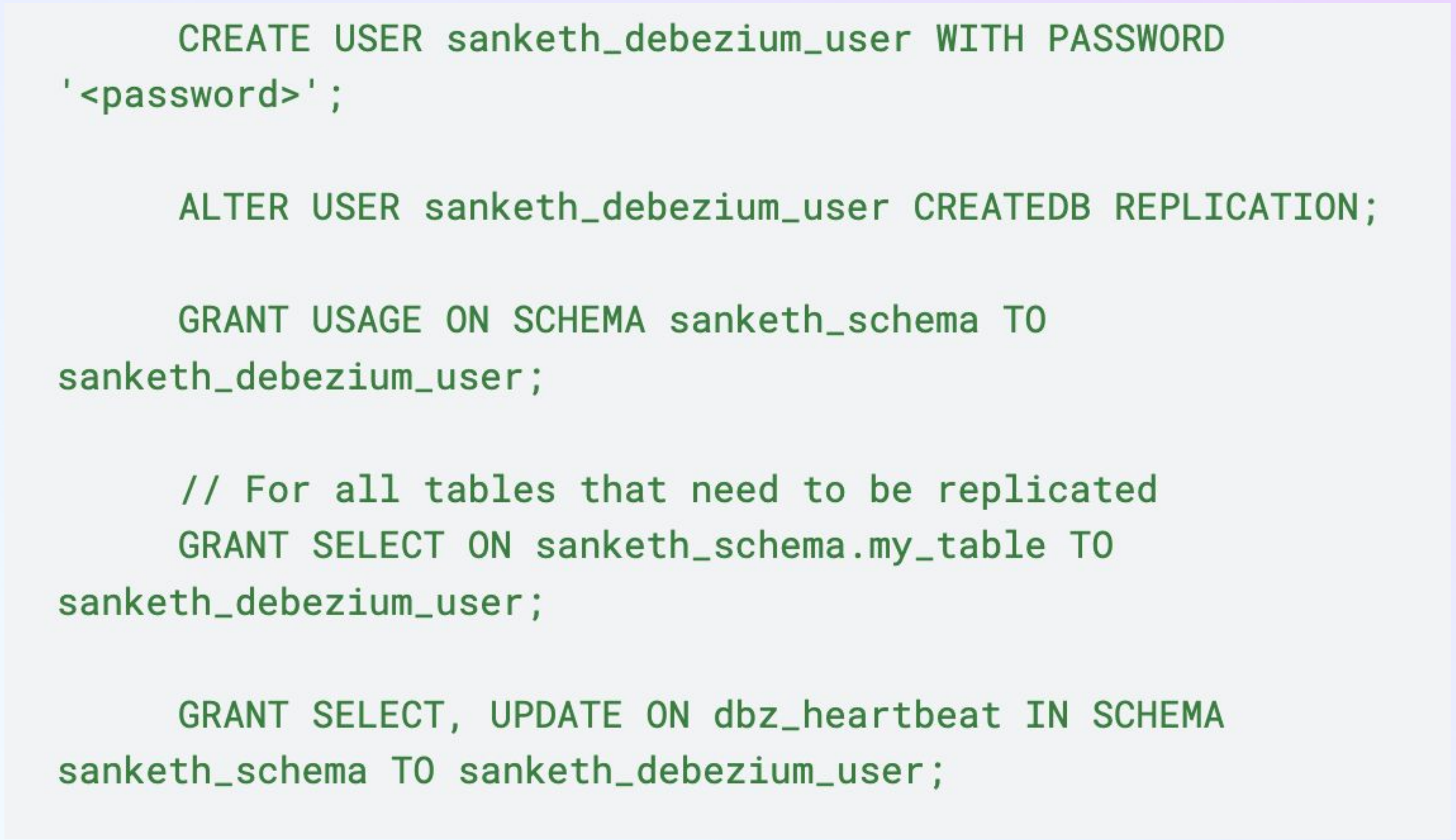
Provisioning

1. PG Operations

High margin for error

2. Repetitive User Asks

Region isolation -> provision everything everywhere



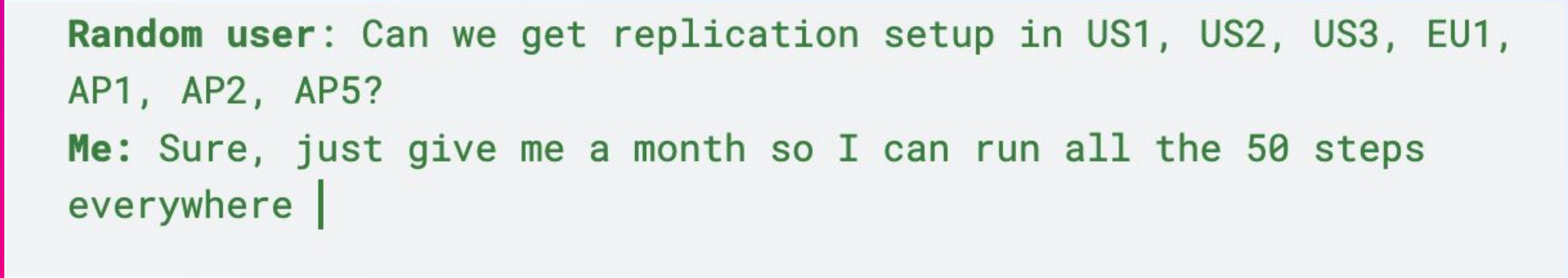
```
      CREATE USER sanketh_debezium_user WITH PASSWORD
'<password>';

      ALTER USER sanketh_debezium_user CREATEDB REPLICATION;

      GRANT USAGE ON SCHEMA sanketh_schema TO
sanketh_debezium_user;

      // For all tables that need to be replicated
      GRANT SELECT ON sanketh_schema.my_table TO
sanketh_debezium_user;

      GRANT SELECT, UPDATE ON dbz_heartbeat IN SCHEMA
sanketh_schema TO sanketh_debezium_user;
```



Random user: Can we get replication setup in US1, US2, US3, EU1, AP1, AP2, AP5?

Me: Sure, just give me a month so I can run all the 50 steps everywhere |

DATADOG

Provisioning - Solutions

Build automation for repetitive tasks

Focus on this right after initial 1-2 users

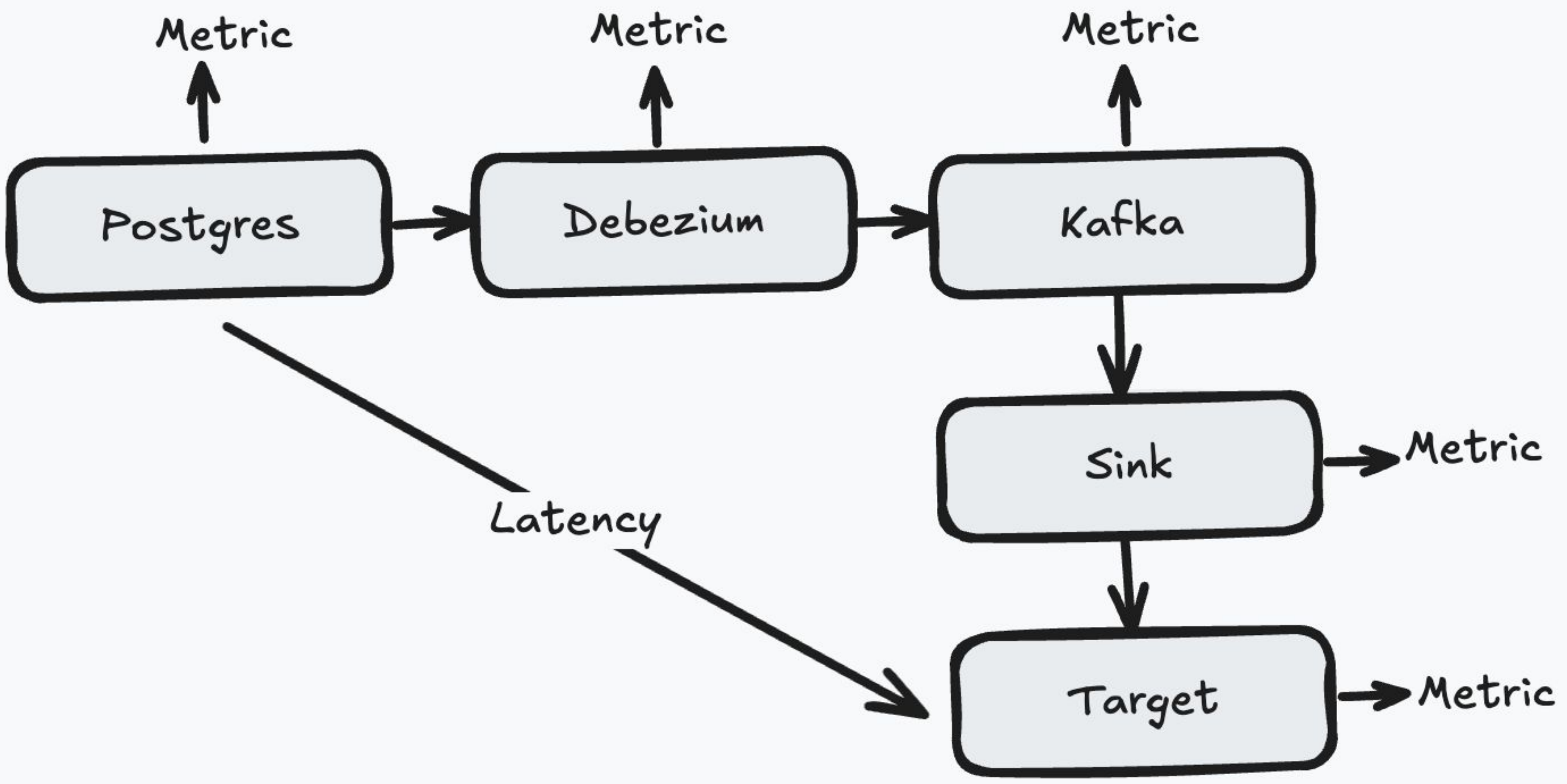
Start small - Doesn’t need to be perfect

Metrics/Traces

1. Unified metrics view is useful

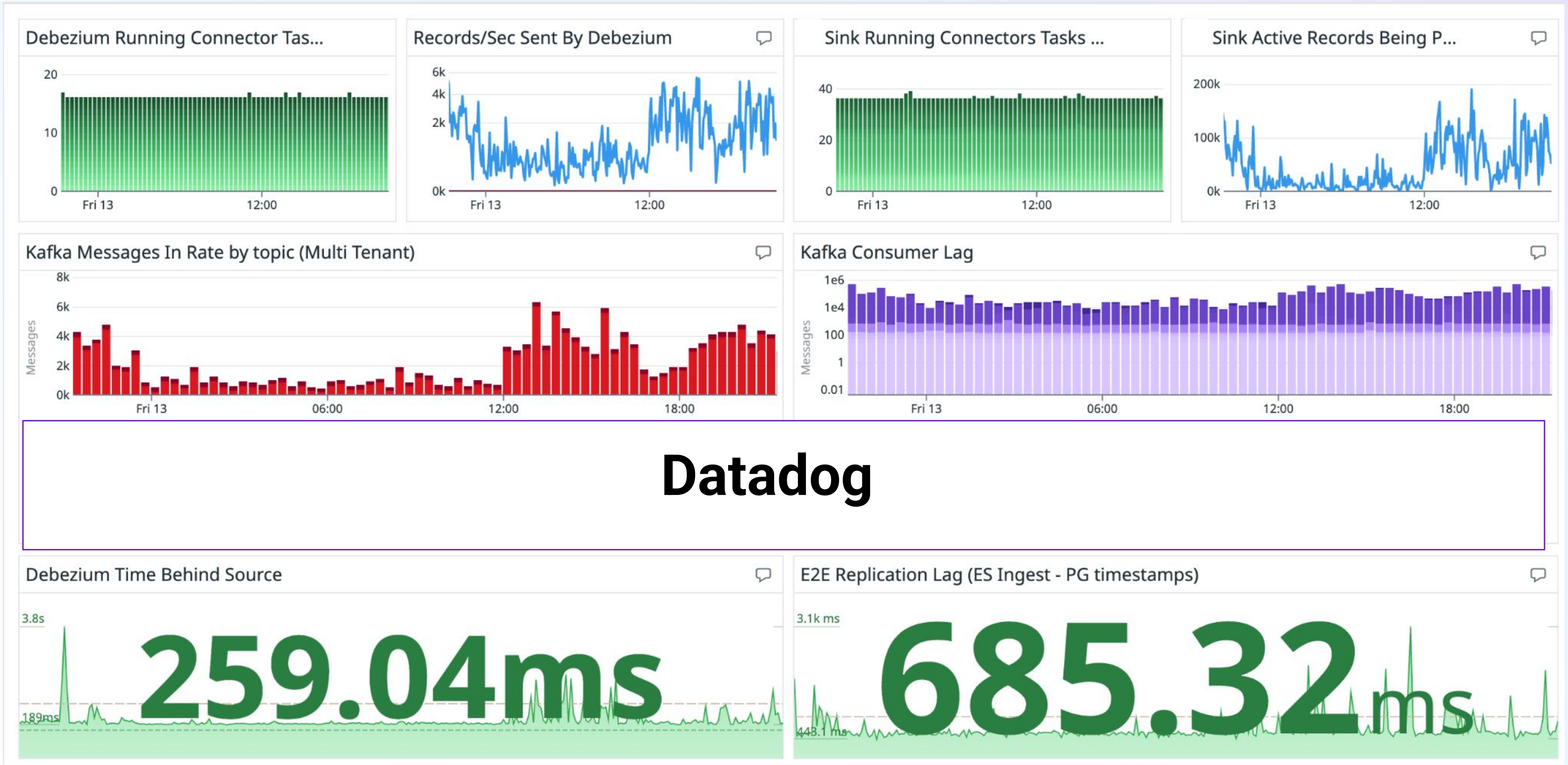
2. Users care about E2E latency

3. Tracing a request through the pipeline



Metrics/Traces - Solutions

1. Metrics Unified View



Debezium Running Connector Tas...
Records/Sec Sent By Debezium
Sink Running Connectors Tasks ...
Sink Active Records Being P...
Kafka Messages In Rate by topic (Multi Tenant)
Kafka Consumer Lag
Messages
Fri 13
06:00
12:00
18:00
Debezium Time Behind Source
3.8s
189ms
259.04ms
E2E Replication Lag (ES Ingest - PG timestamps)
3.1k ms
443.1 ms
685.32ms

Datadog

Metrics/Traces - Solutions

2. Measuring latency

End to End latency is not trivial to measure

One way is to get the time diff through a field

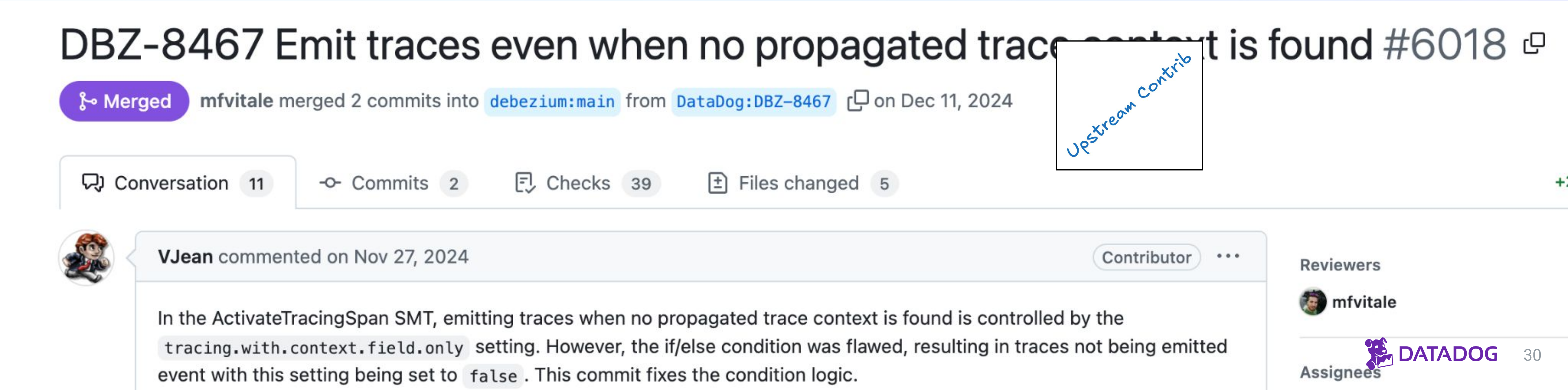
You could update or use external datastore to update latency

Metrics/Traces - Solutions

3. Tracing a request - Distributed Tracing

Debezium offers to create traces - ActiveTracingSpan SMT

Set parent ID in the span for each component



DBZ-8467 Emit traces even when no propagated trac[illegible]t is found #6018

Upstream Contrib

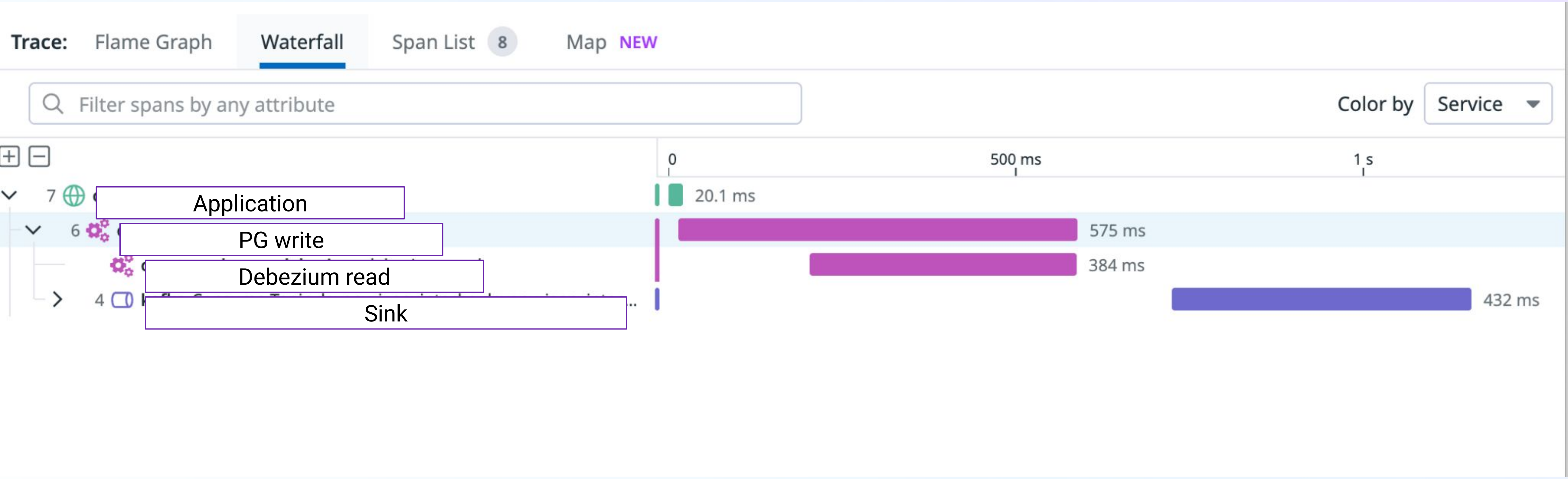
Merged mfvitale merged 2 commits into `debezium:main` from `DataDog:DBZ-8467` on Dec 11, 2024

VJean commented on Nov 27, 2024 Contributor

In the ActivateTracingSpan SMT, emitting traces when no propagated trace context is found is controlled by the `tracing.with.context.field.only` setting. However, the if/else condition was flawed, resulting in traces not being emitted event with this setting being set to `false`. This commit fixes the condition logic.

Metrics/Traces - Solutions

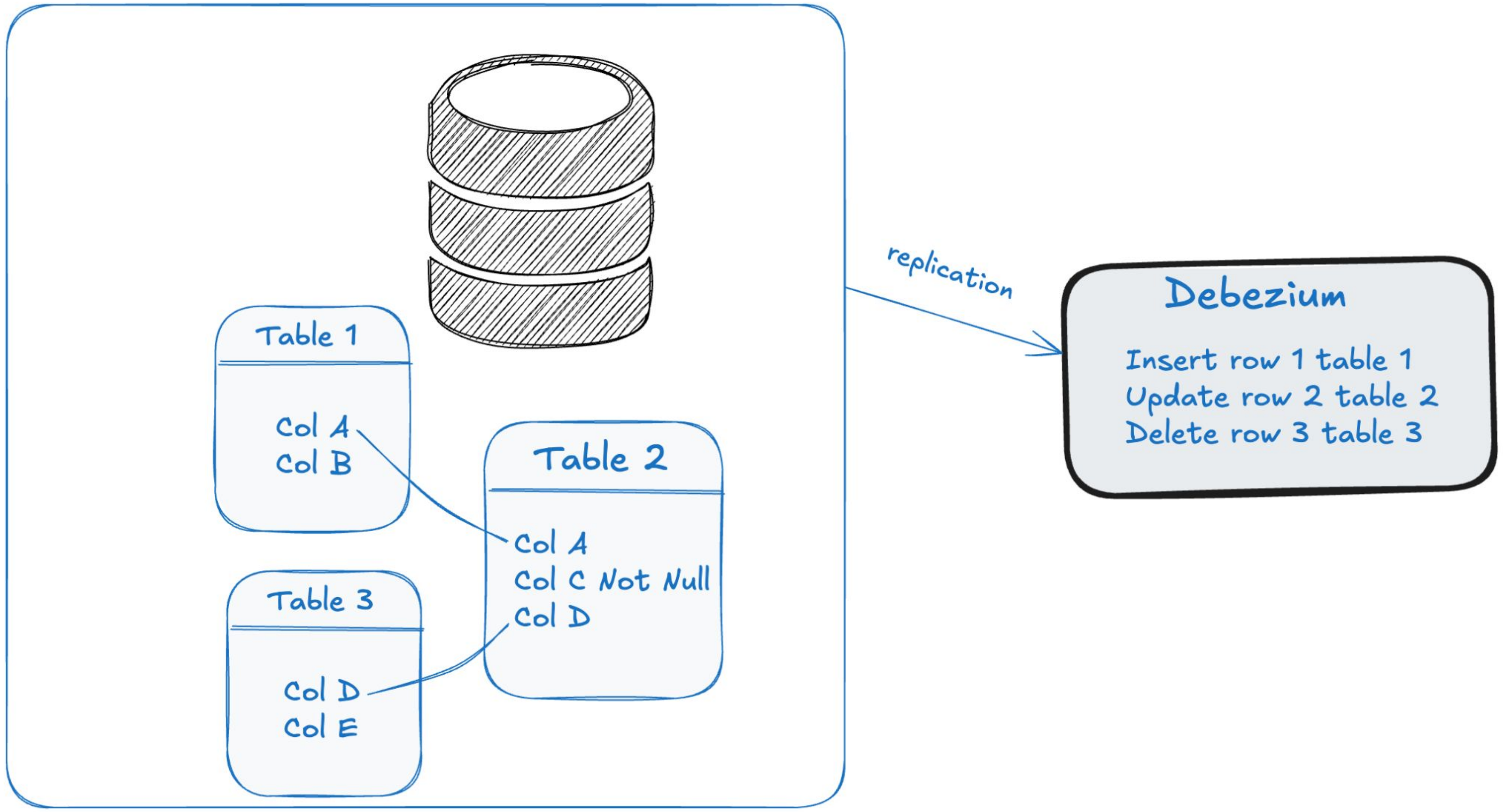
3. Tracing a request - Datadog Traces



Schemas

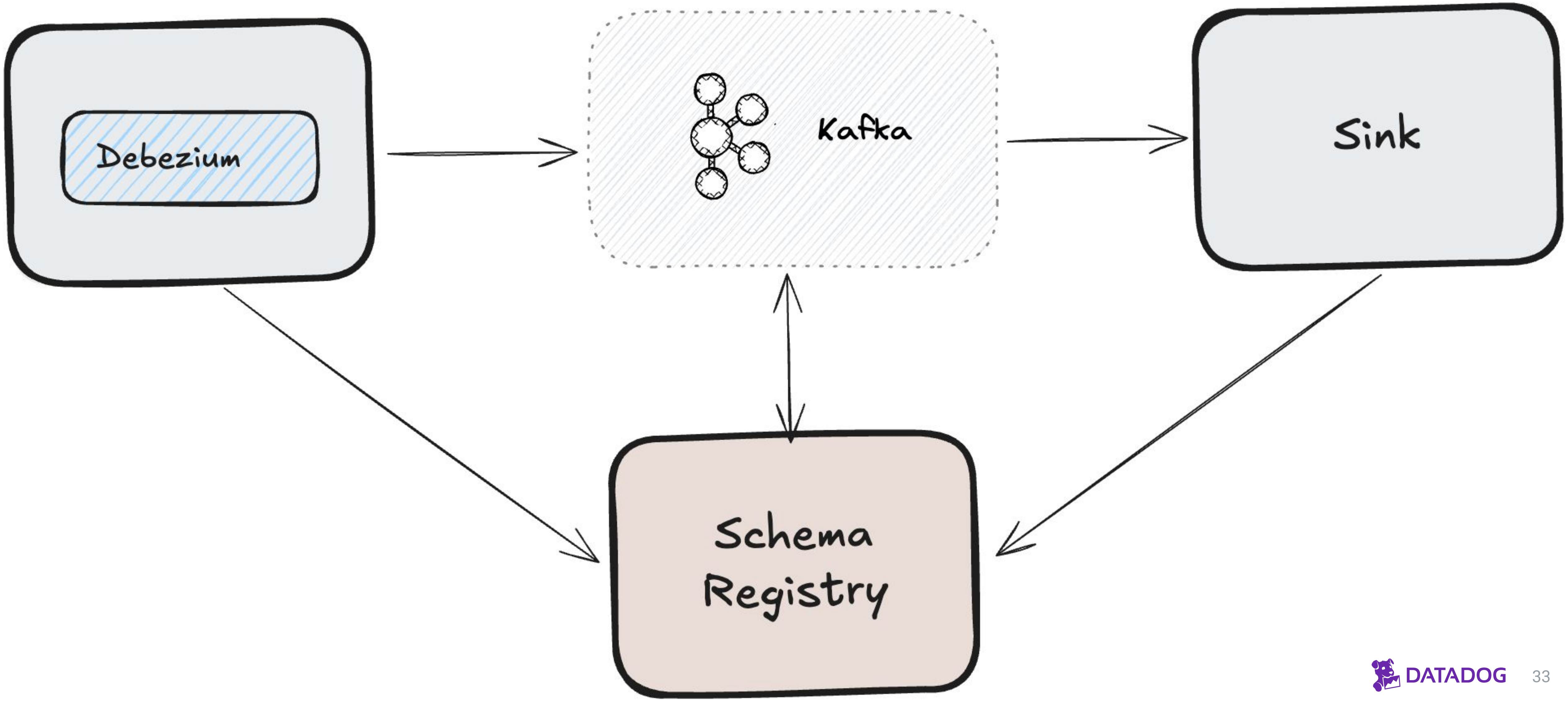
1. DDL's not supported through replication

2. Relational structure broken after PG



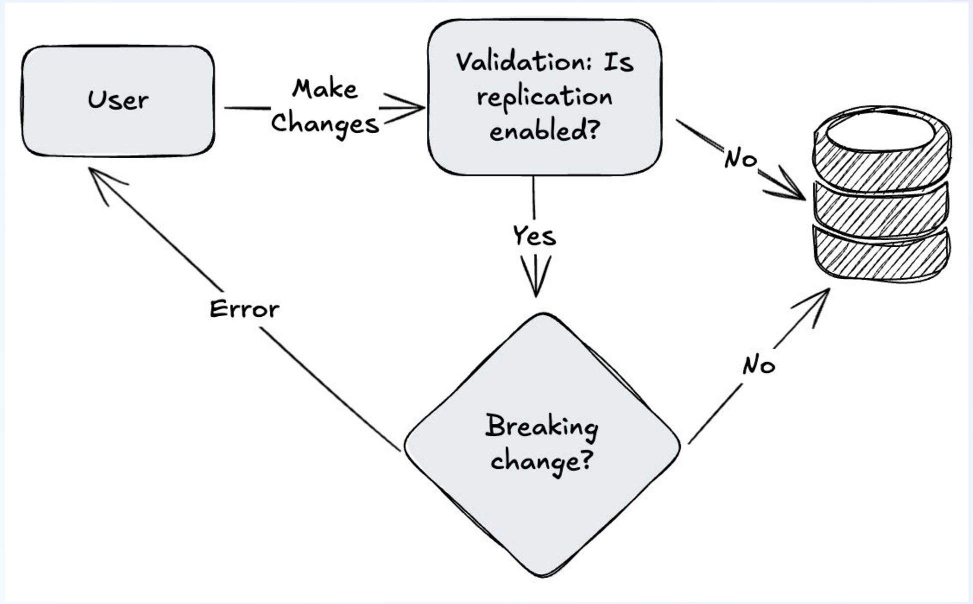
Schemas - Solutions

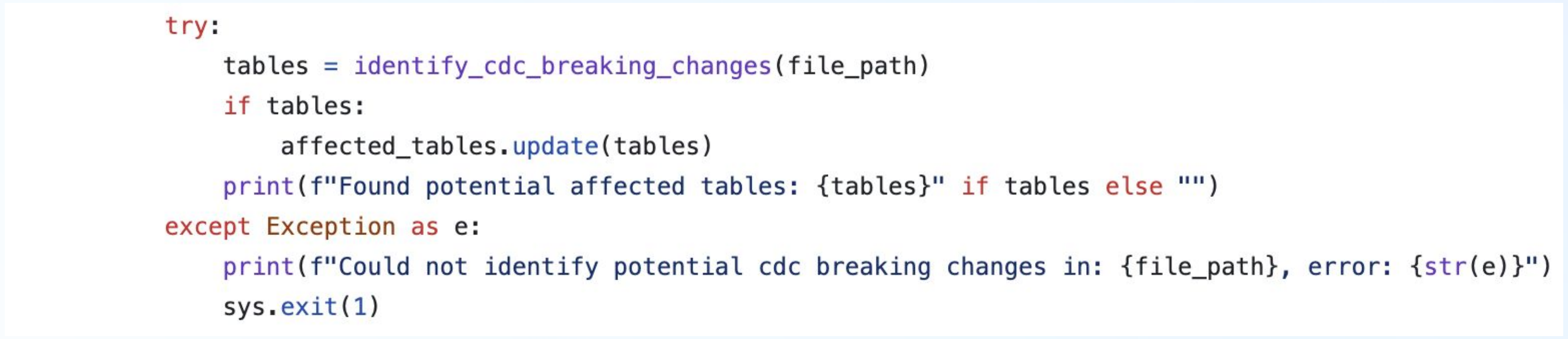
1. Use a Schema Enforcer



Schemas - Solutions

2. Enforce Validations



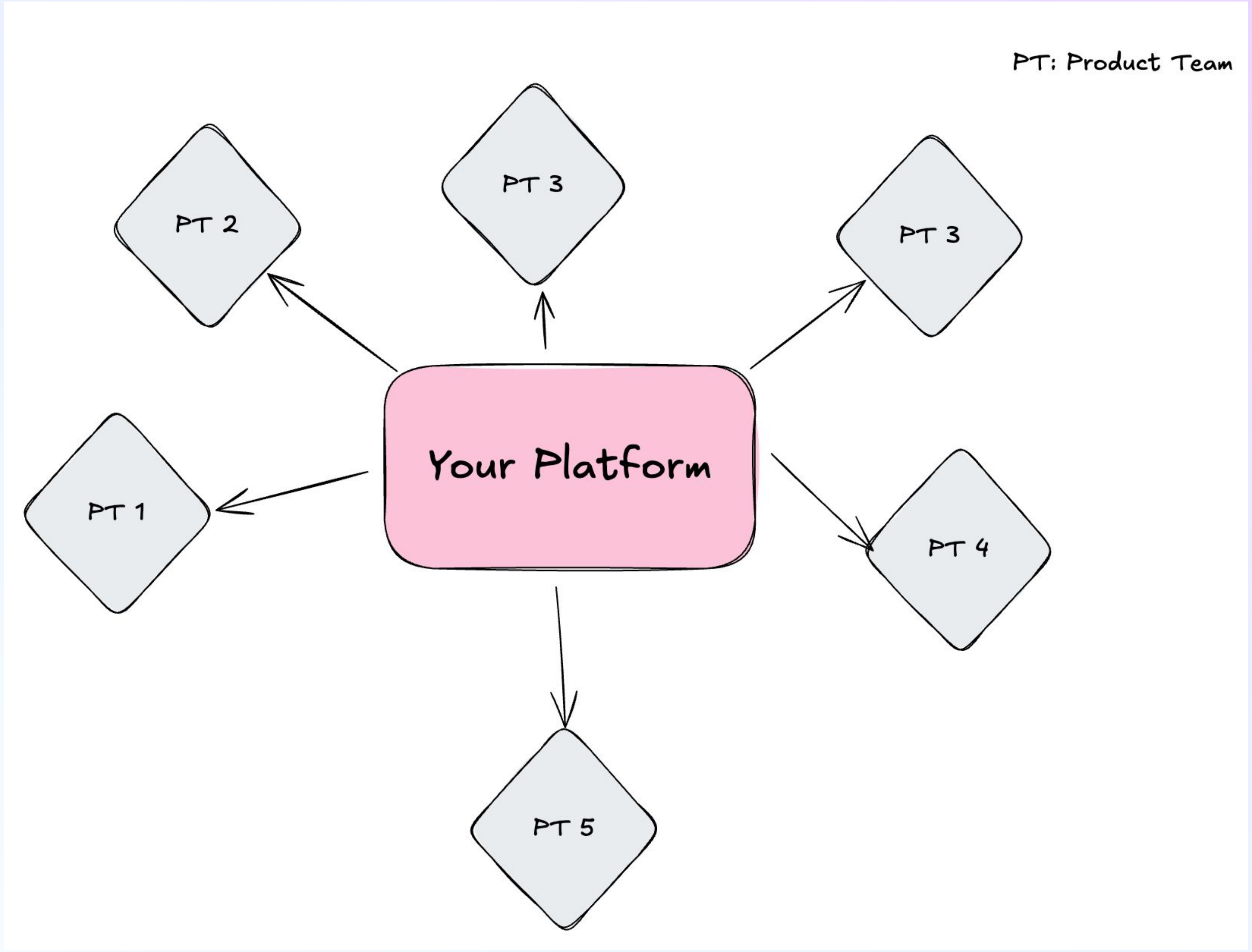


```
try:
    tables = identify_cdc_breaking_changes(file_path)
    if tables:
        affected_tables.update(tables)
    print(f"Found potential affected tables: {tables}" if tables else "")
except Exception as e:
    print(f"Could not identify potential cdc breaking changes in: {file_path}, error: {str(e)}")
    sys.exit(1)
```

Customization

1. Teams want a flavor of your platform

2. How do you handle multiple requests from teams?



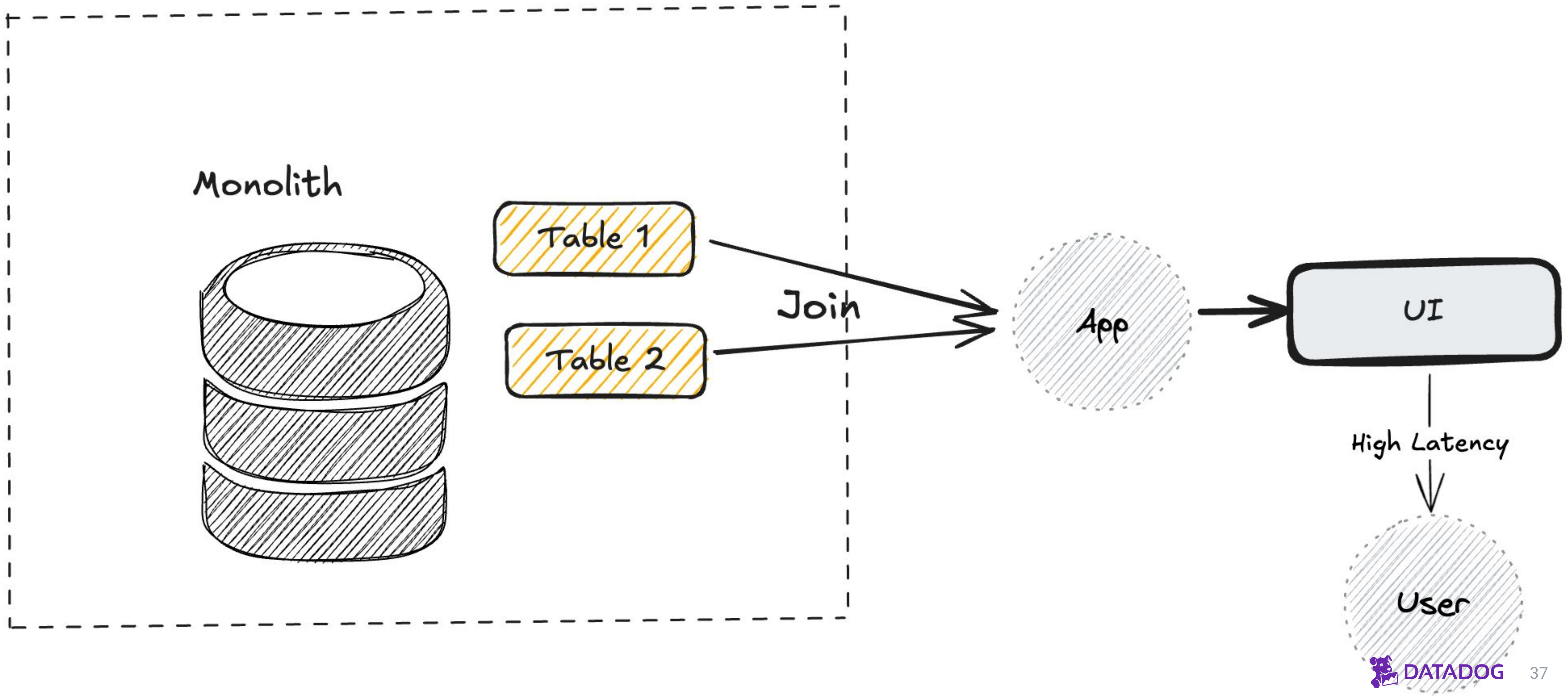
Customization - Solutions

Generalize your platform

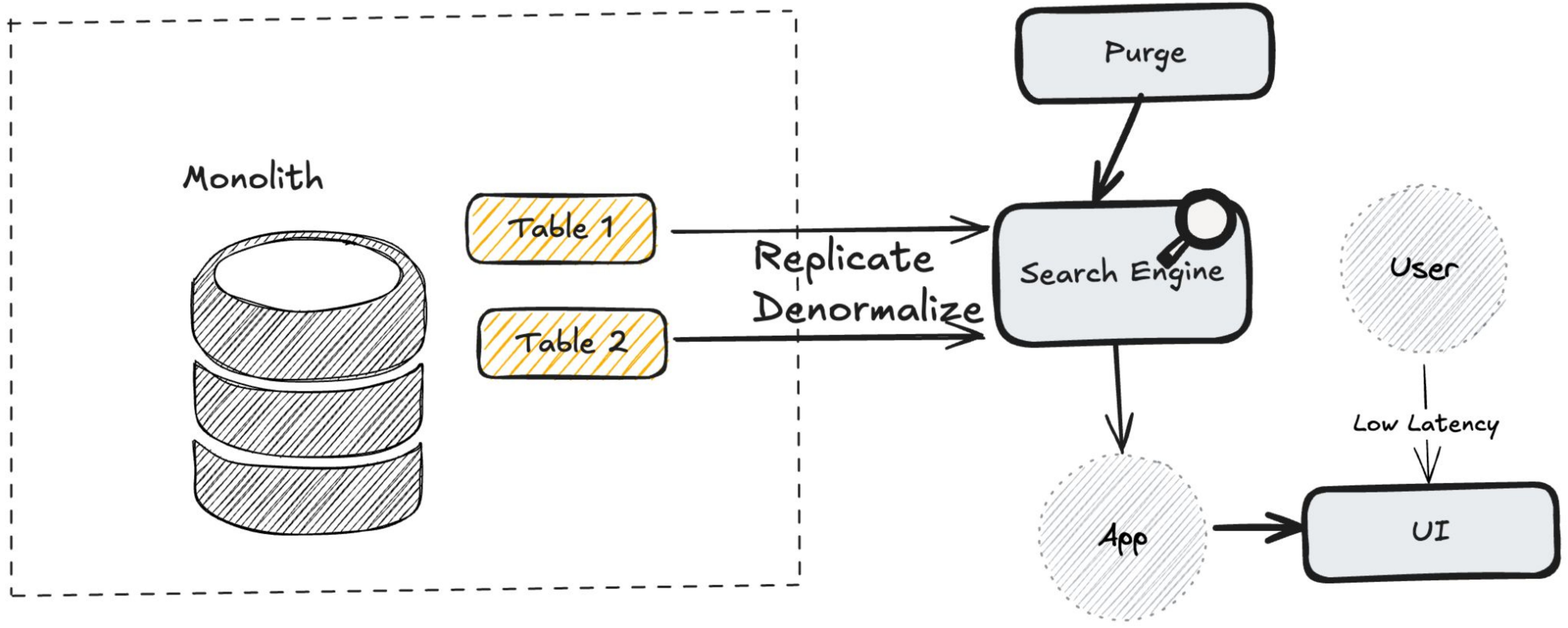
White gloved support is unavoidable

Think about if a feature is a right fit for your platform

A Use Case at Datadog - Problem

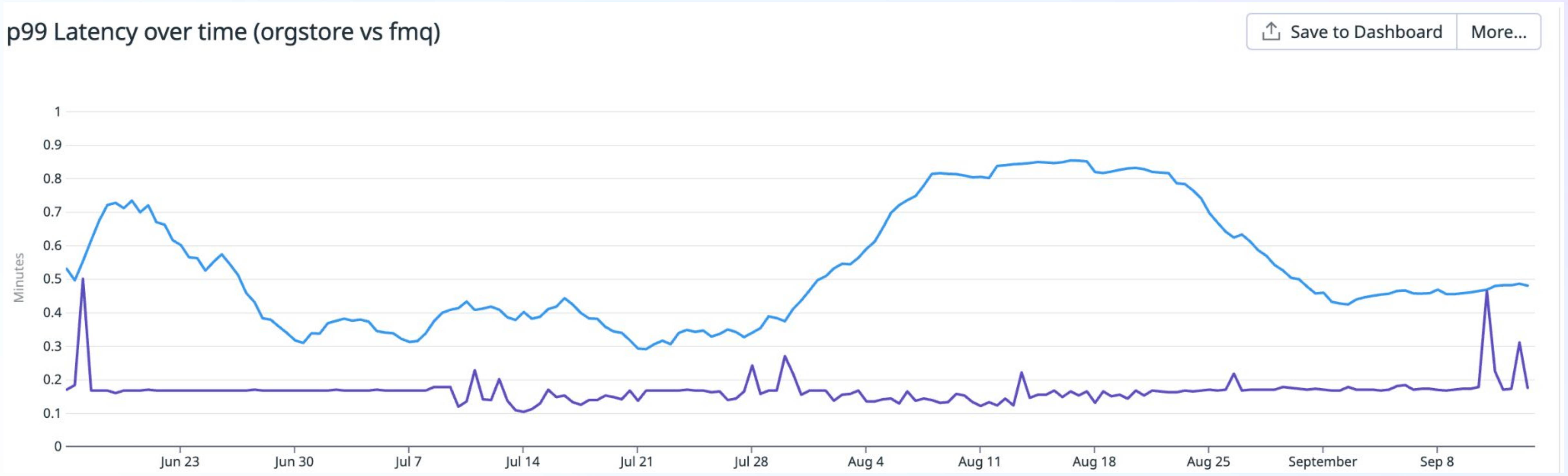


A Use Case at Datadog - Solution

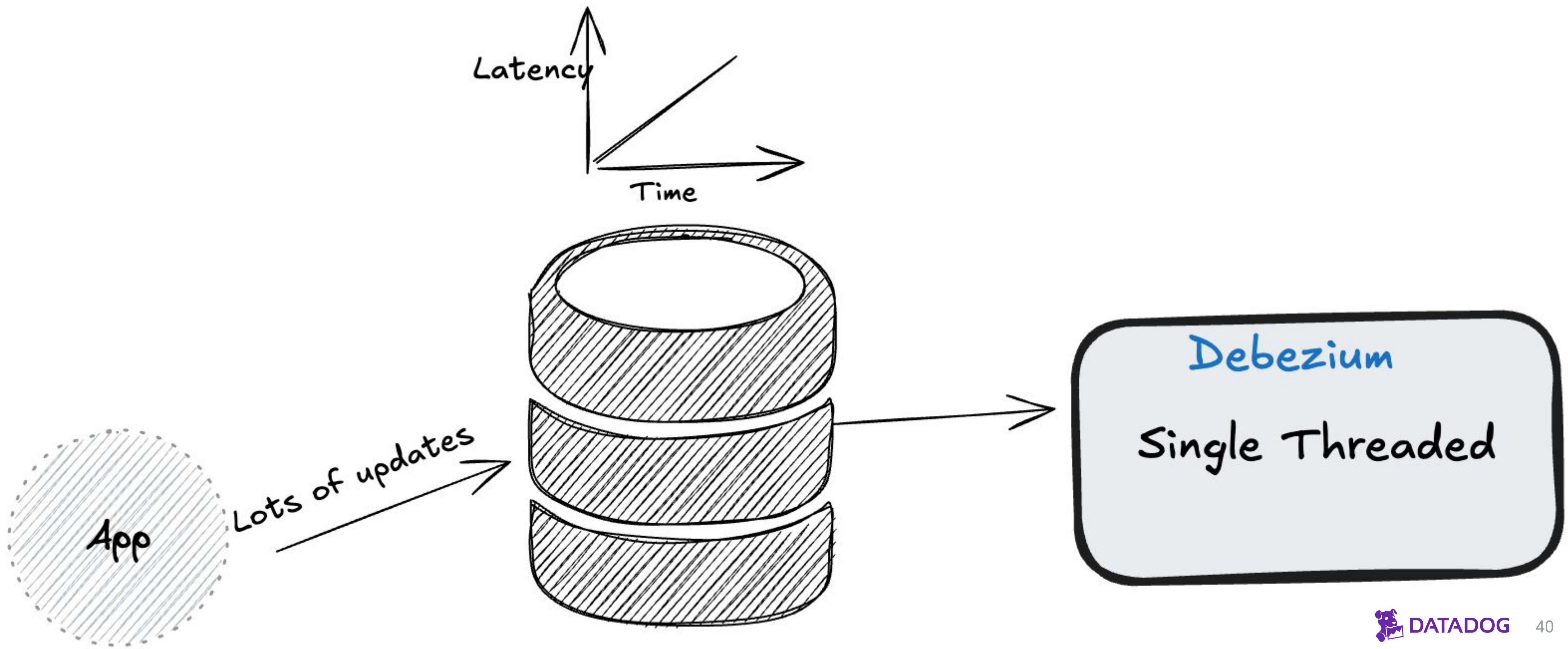


Results

- **41%** Average decrease in p90 load times
- Customers saw a drop in latency up to **95%+** with some loads dropping from ~**27s to ~1s**



An Architecture That Didn't Work



Takeaways

Architecture is important

Shared WAL is a bottleneck we live with

Build Generalizable solutions that can scale

Thank you!