Streaming ETL on the Shoulders of GIANTS

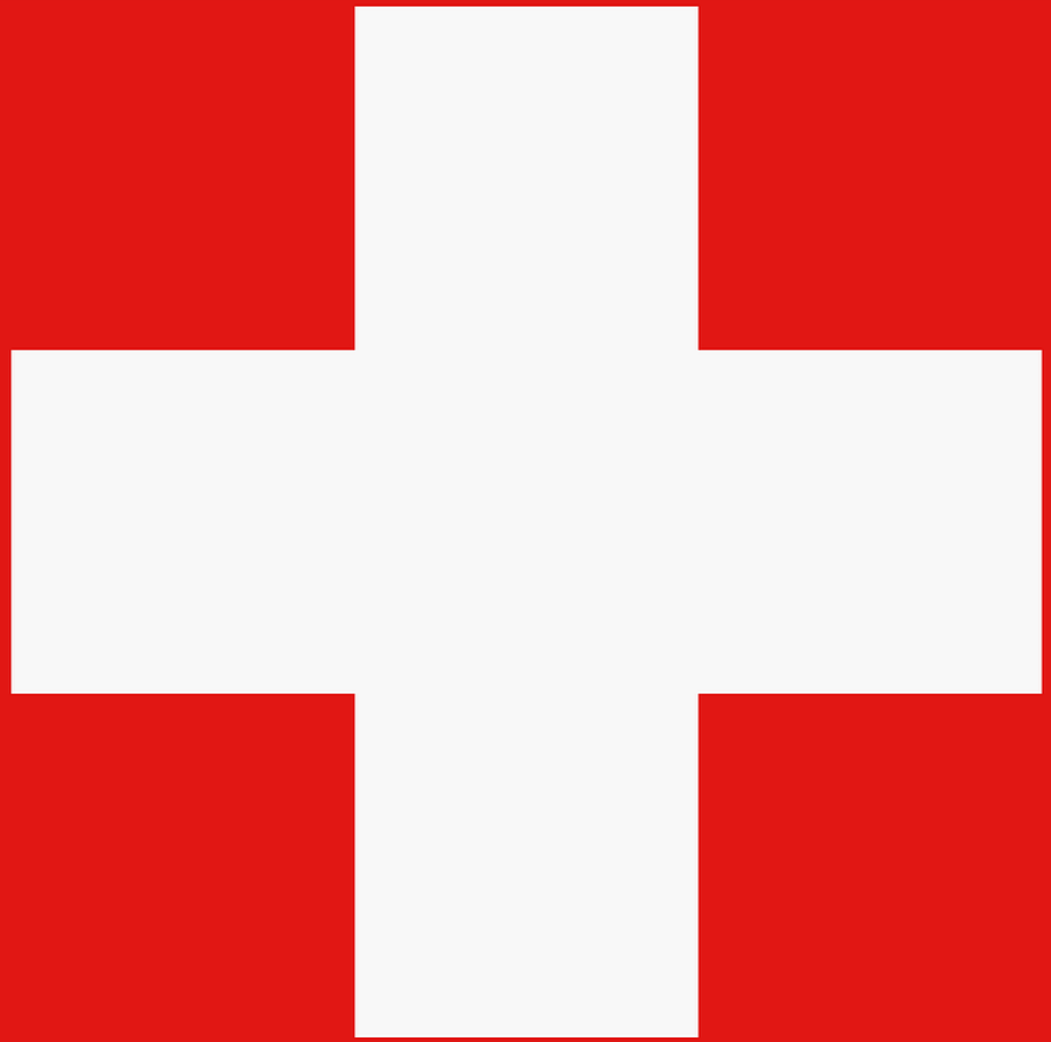
Hans-Peter Grahsl

- working & living in Graz 🇦🇹
- technical trainer at NETCONOMY
- independent consultant & engineer
- associate lecturer
- 🗣️ occasional conference speaker 📣

Speed & Agility

For businesses to **stay relevant** they must **deliver value** at a **breakneck pace** and be constantly seeking **new sources of value** ...

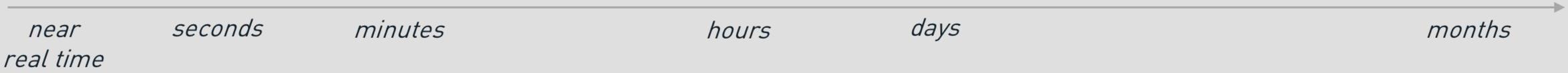

100
10
SCHWEIZERIS
BANCA NAZIU
5FR.
2014
1000
Tausend Franken
Mille Franchi
Milli Francs
2 Fr.
1989
20
BANQUE NATIO
50
SCHWEIZ
BANCA NA
Le banconote sono
protette dal
diritto penale.



Diminishing Value of Data

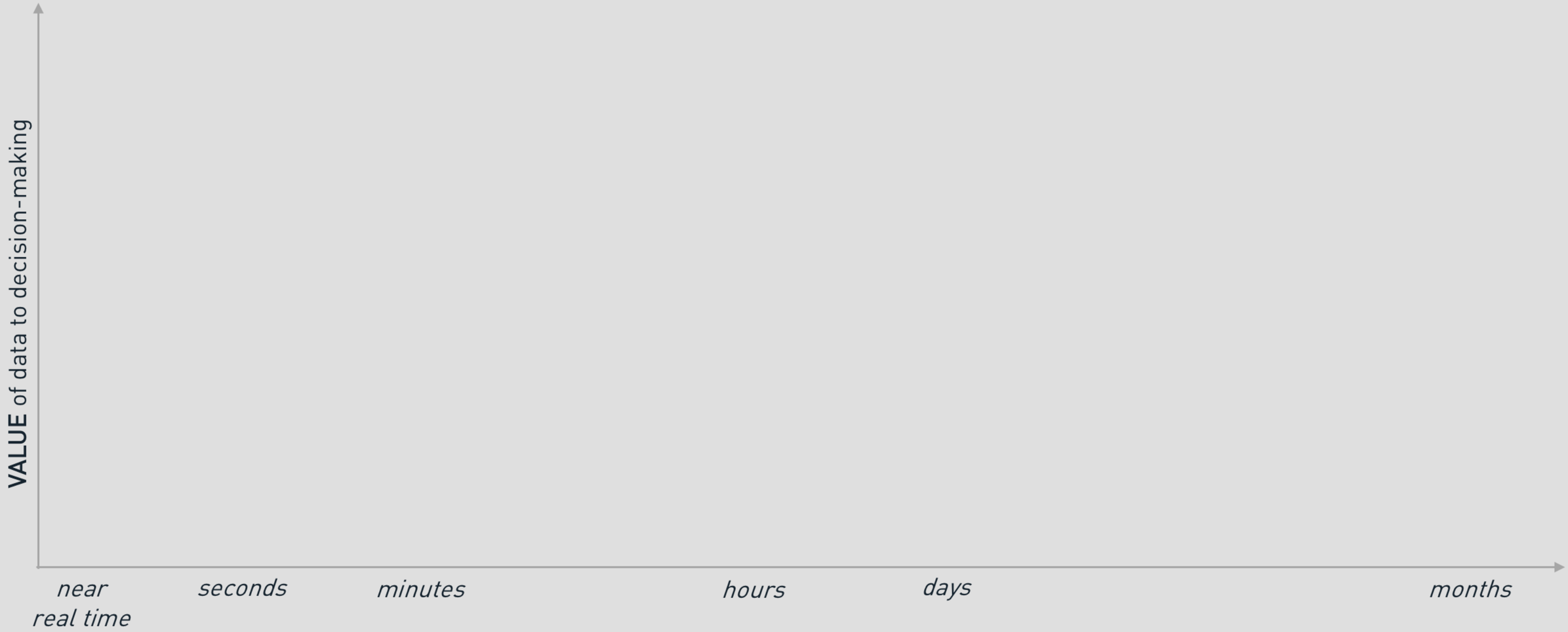
Source: Perishable Insights, Mike Gualtieri, Forrester

Diminishing Value of Data



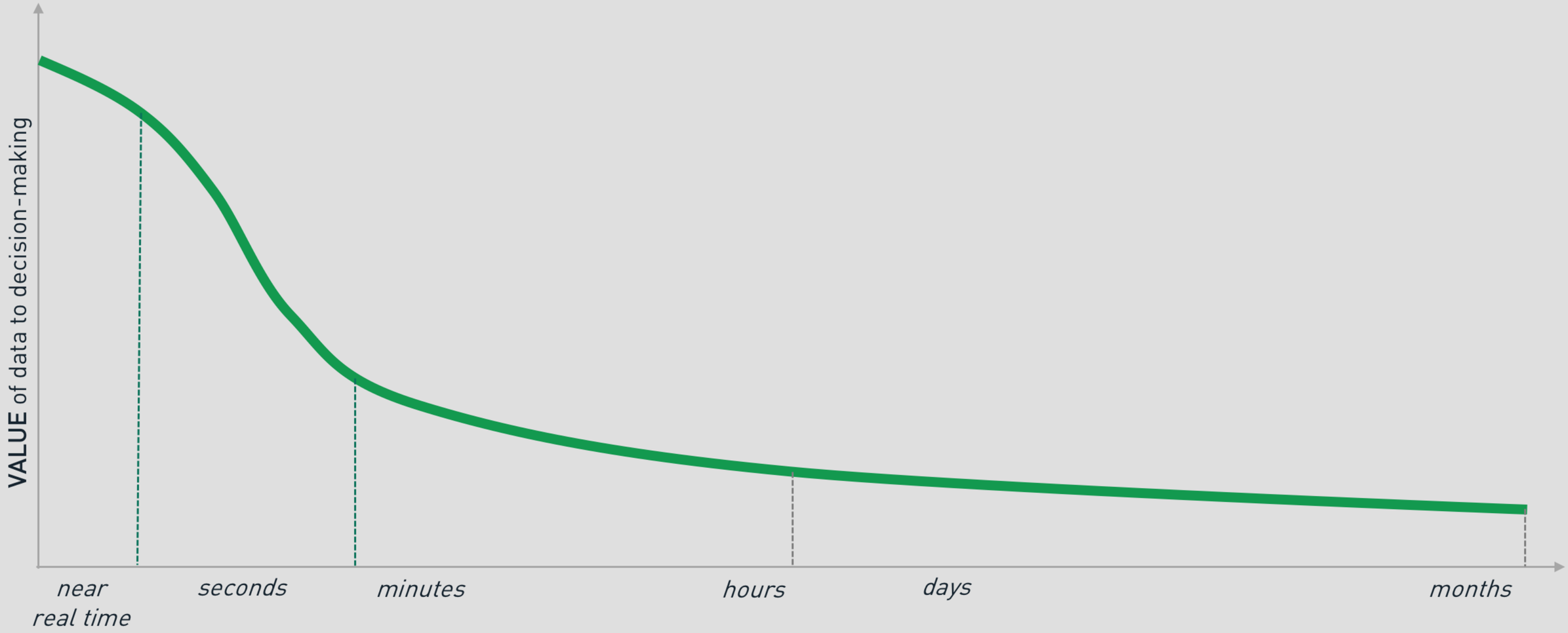
Source: Perishable Insights, Mike Gualtieri, Forrester

Diminishing Value of Data



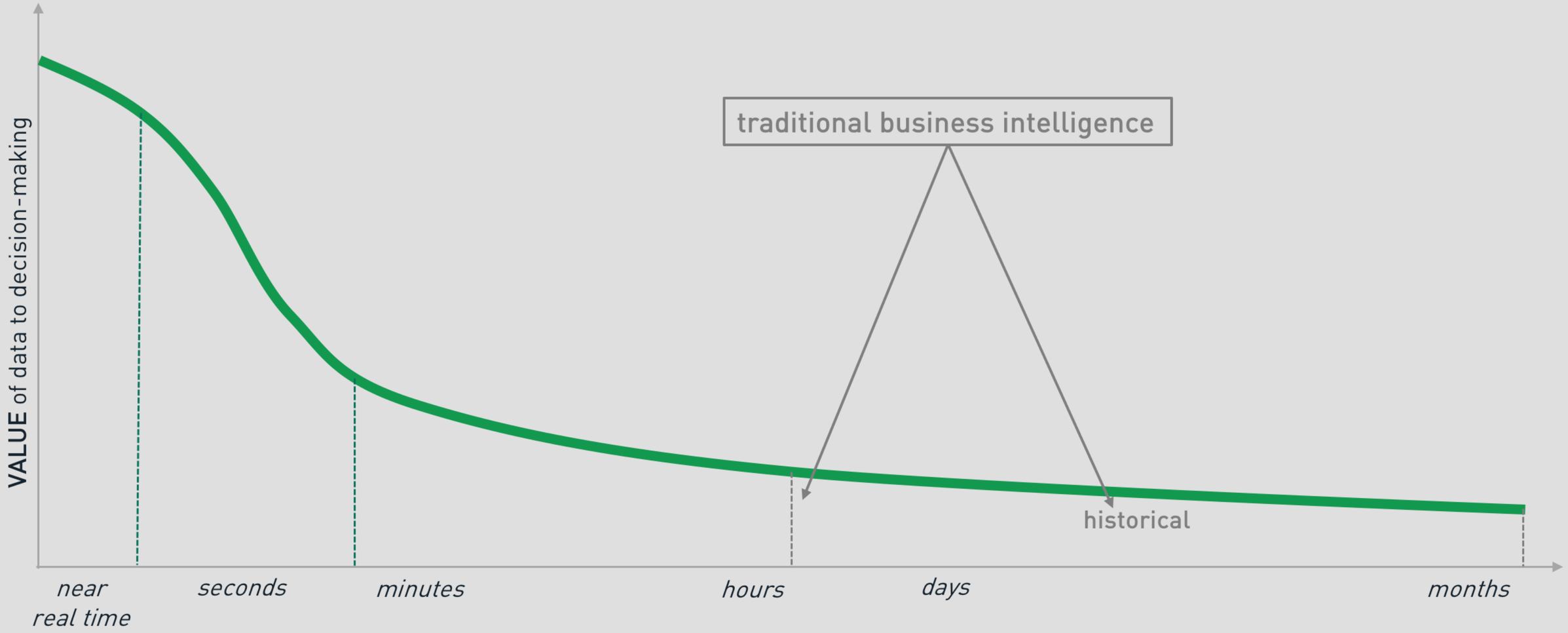
Source: Perishable Insights, Mike Gualtieri, Forrester

Diminishing Value of Data



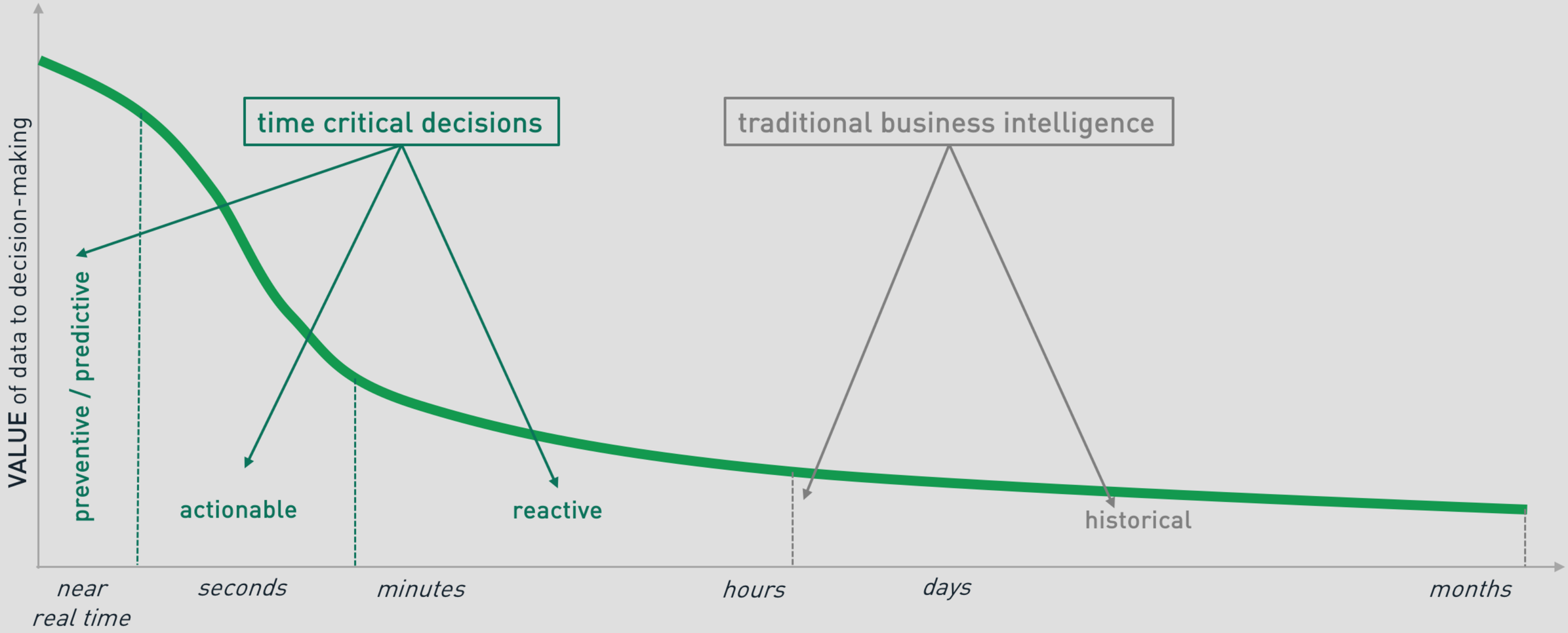
Source: Perishable Insights, Mike Gualtieri, Forrester

Diminishing Value of Data



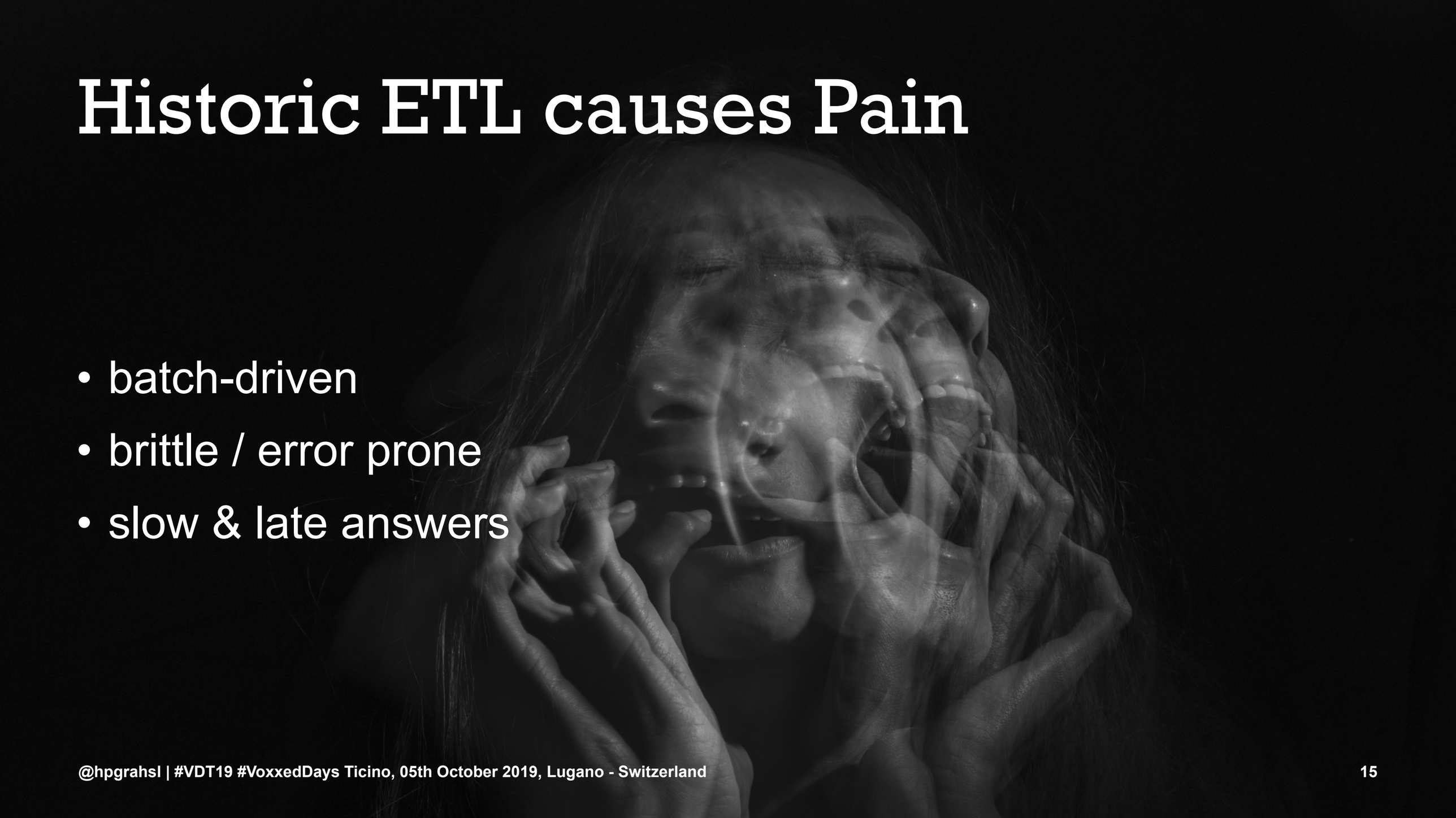
Source: Perishable Insights, Mike Gualtieri, Forrester

Diminishing Value of Data

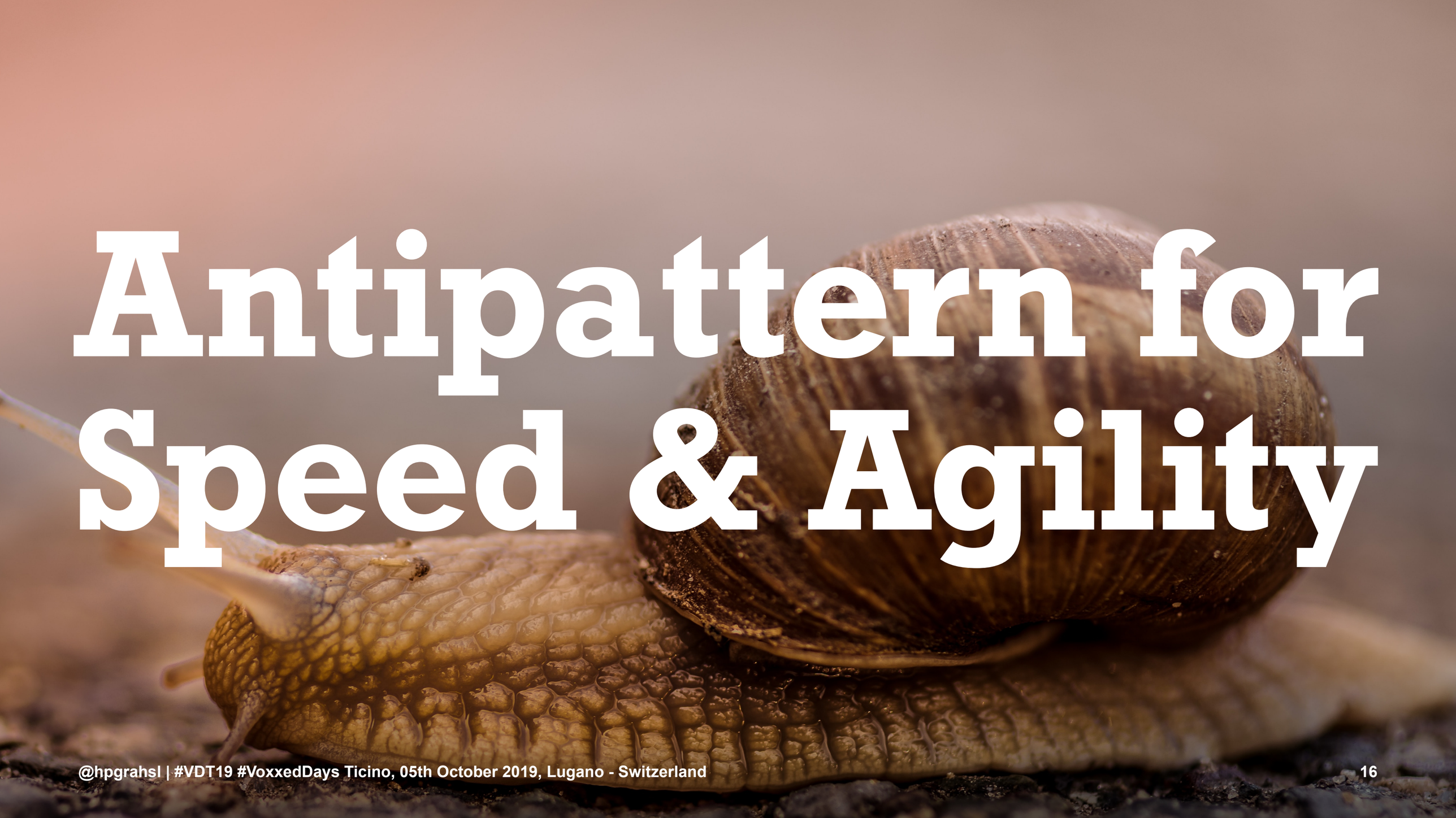


Source: Perishable Insights, Mike Gualtieri, Forrester

Historic ETL causes Pain



- batch-driven
- brittle / error prone
- slow & late answers



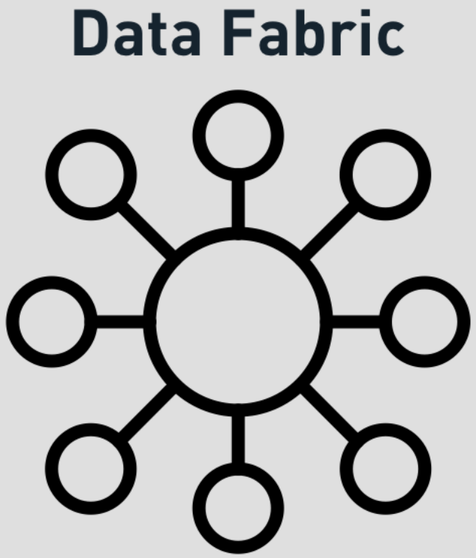
Antipattern for Speed & Agility

Streaming ETL alleviates Pain

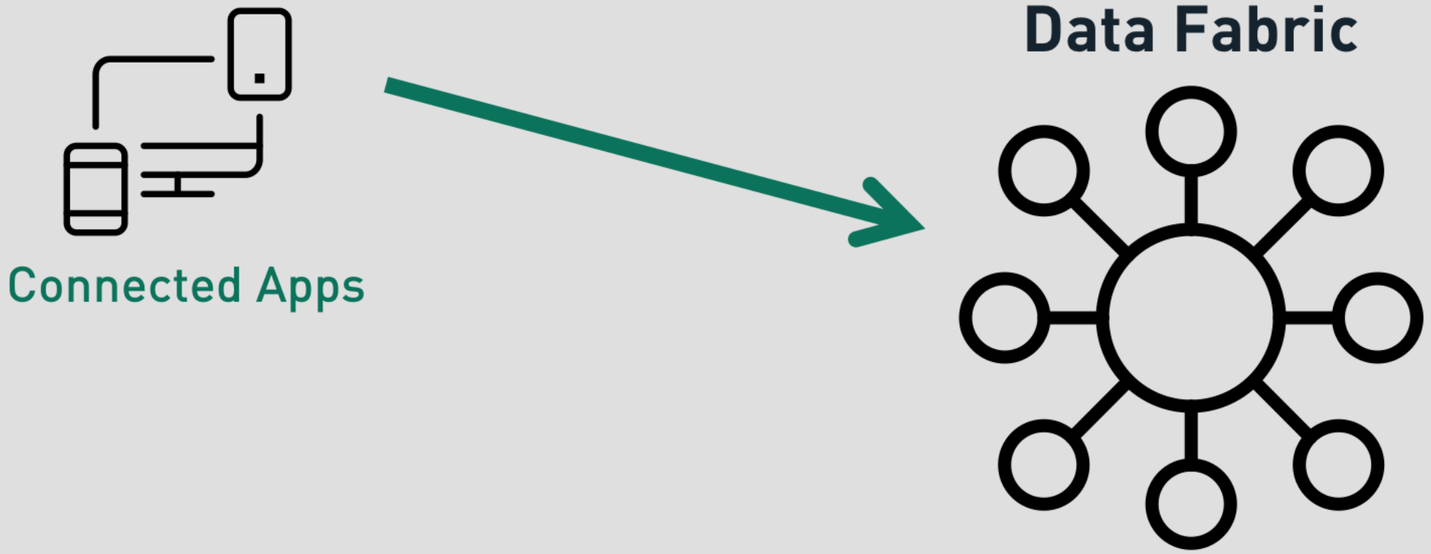
- event-centric
- stream-oriented
- fast & timely answers


Enabler for
Speed & Agility

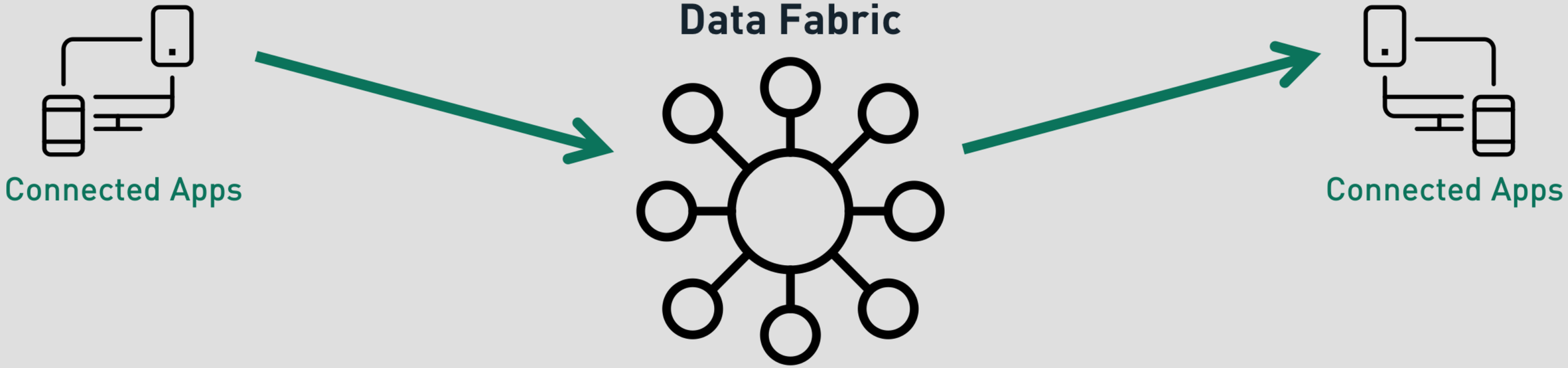
Modern Data Architecture?



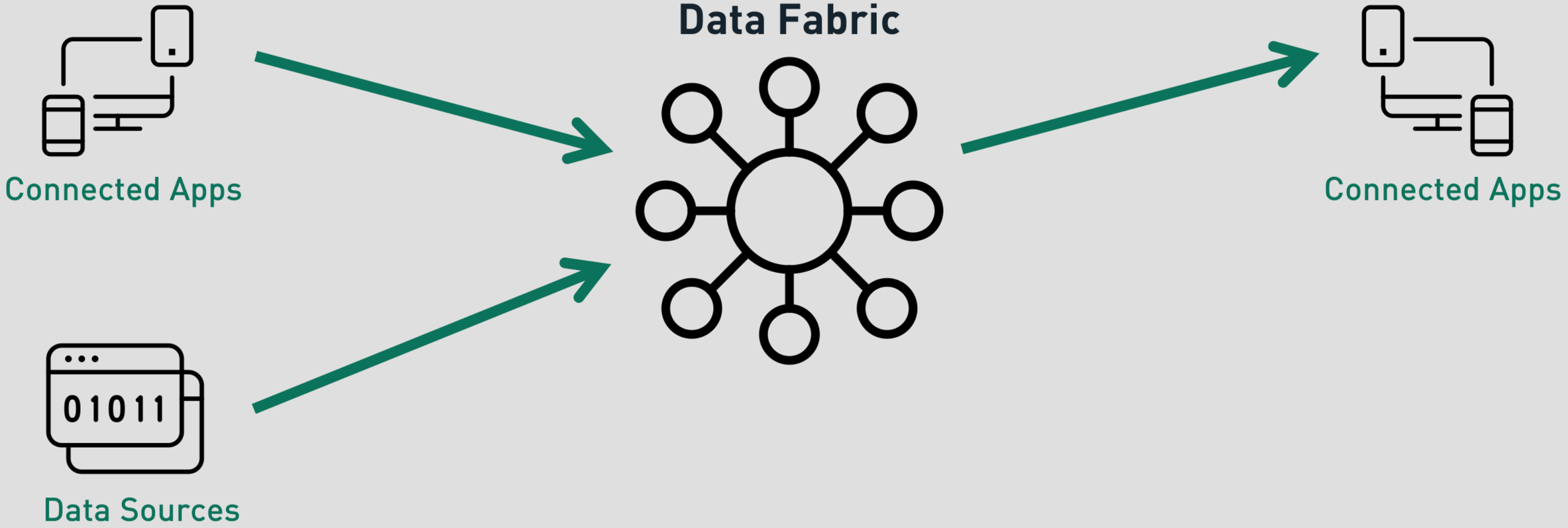
Modern Data Architecture?



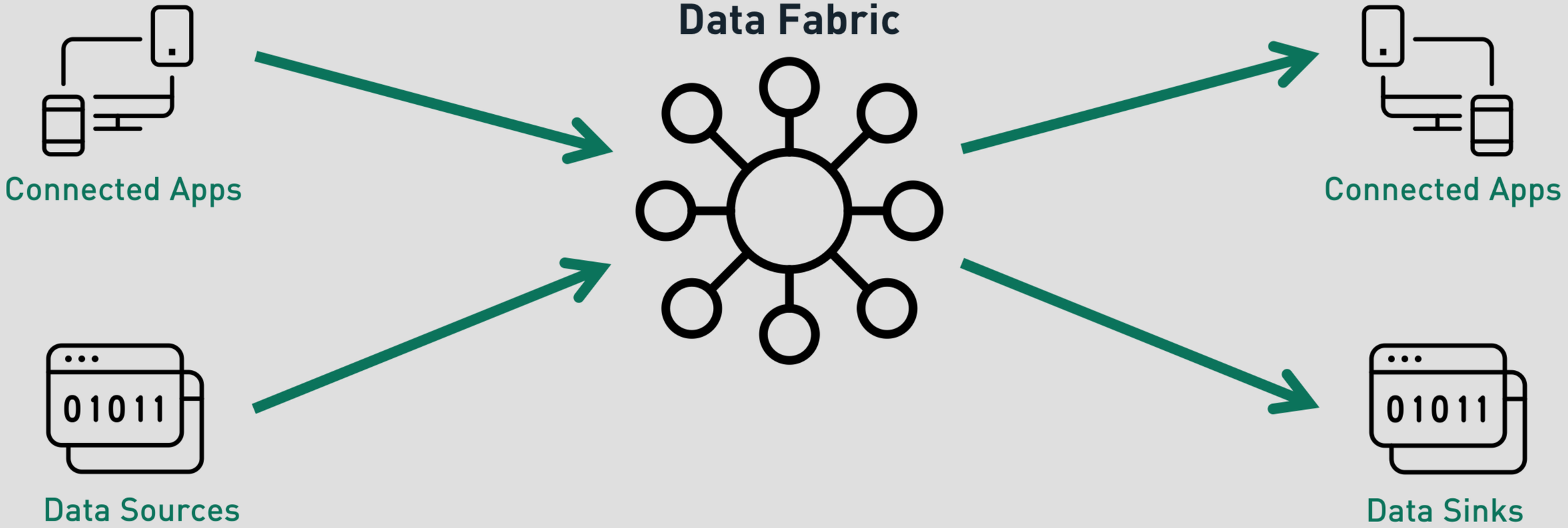
Modern Data Architecture?



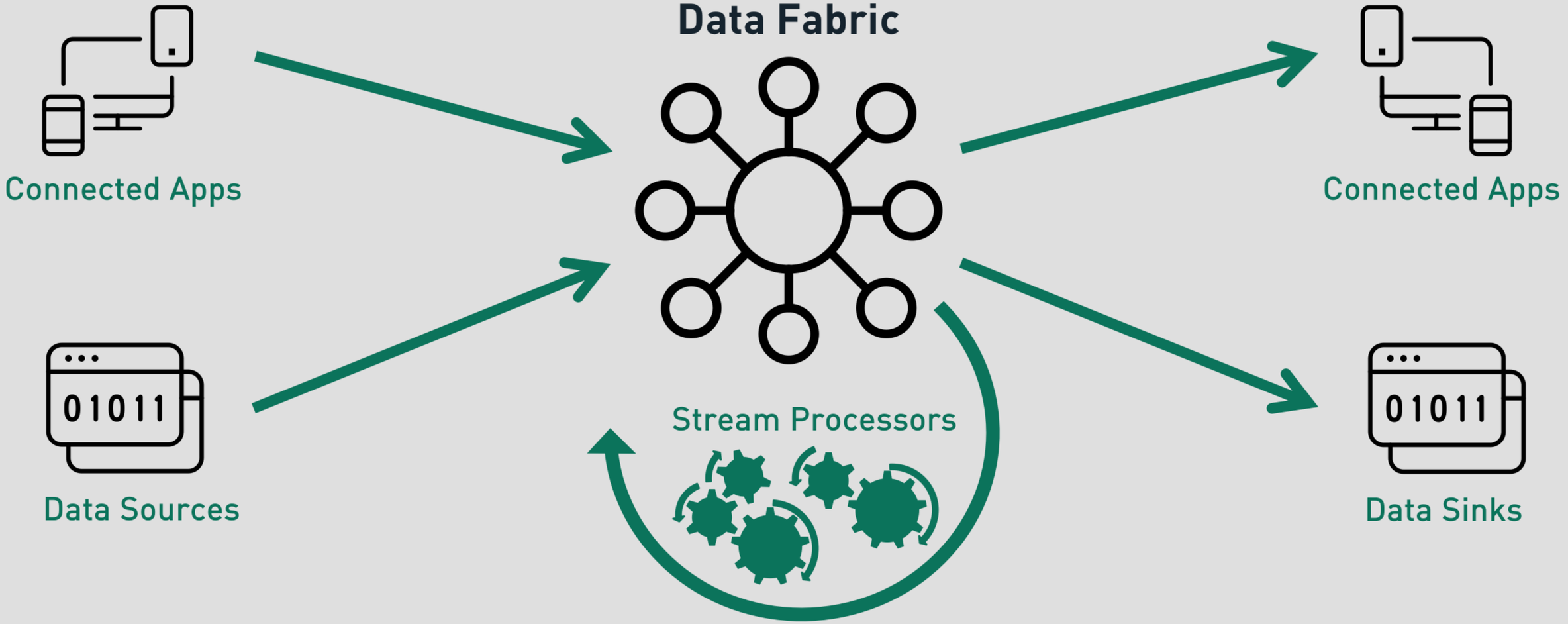
Modern Data Architecture?



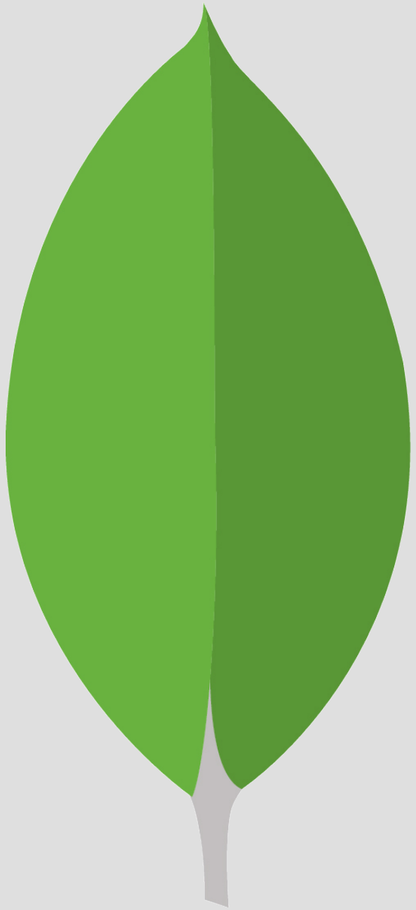
Modern Data Architecture?

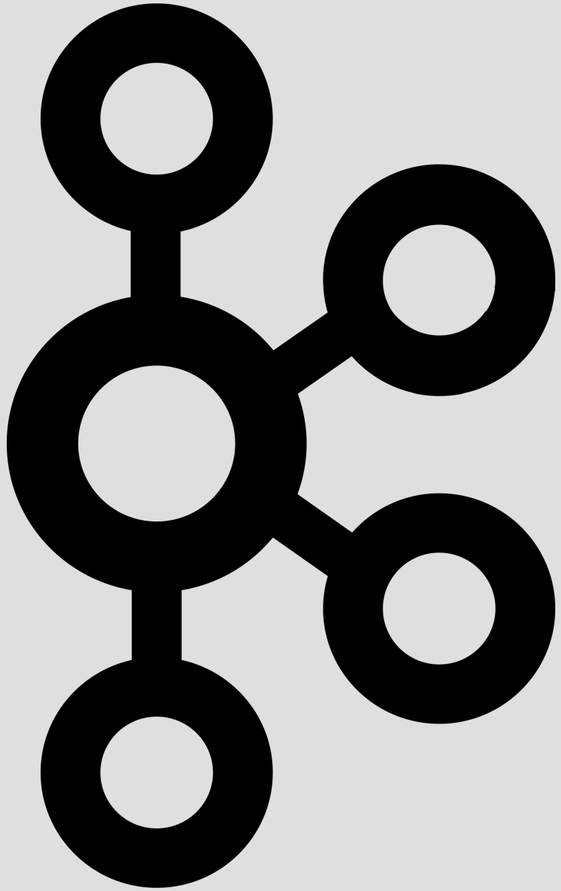


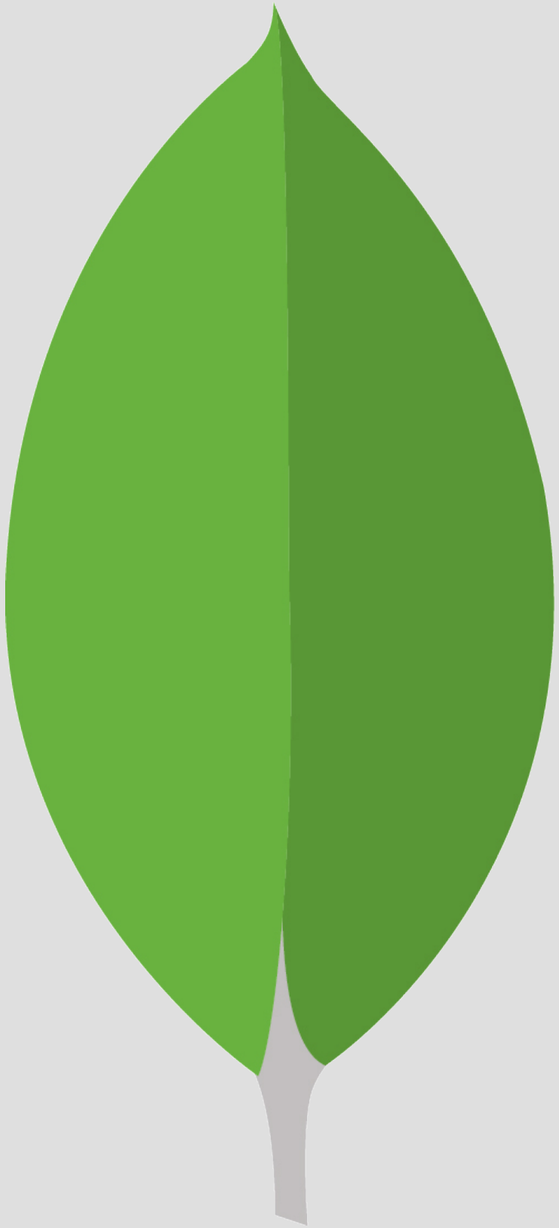
Modern Data Architecture?



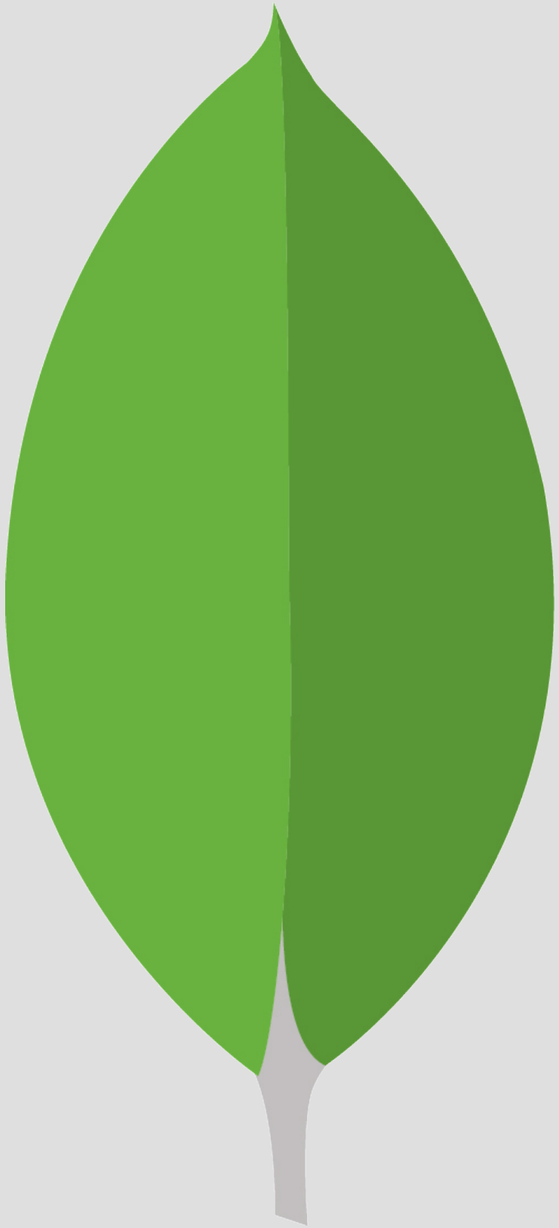
On the Shoulders of **G I A N T S**







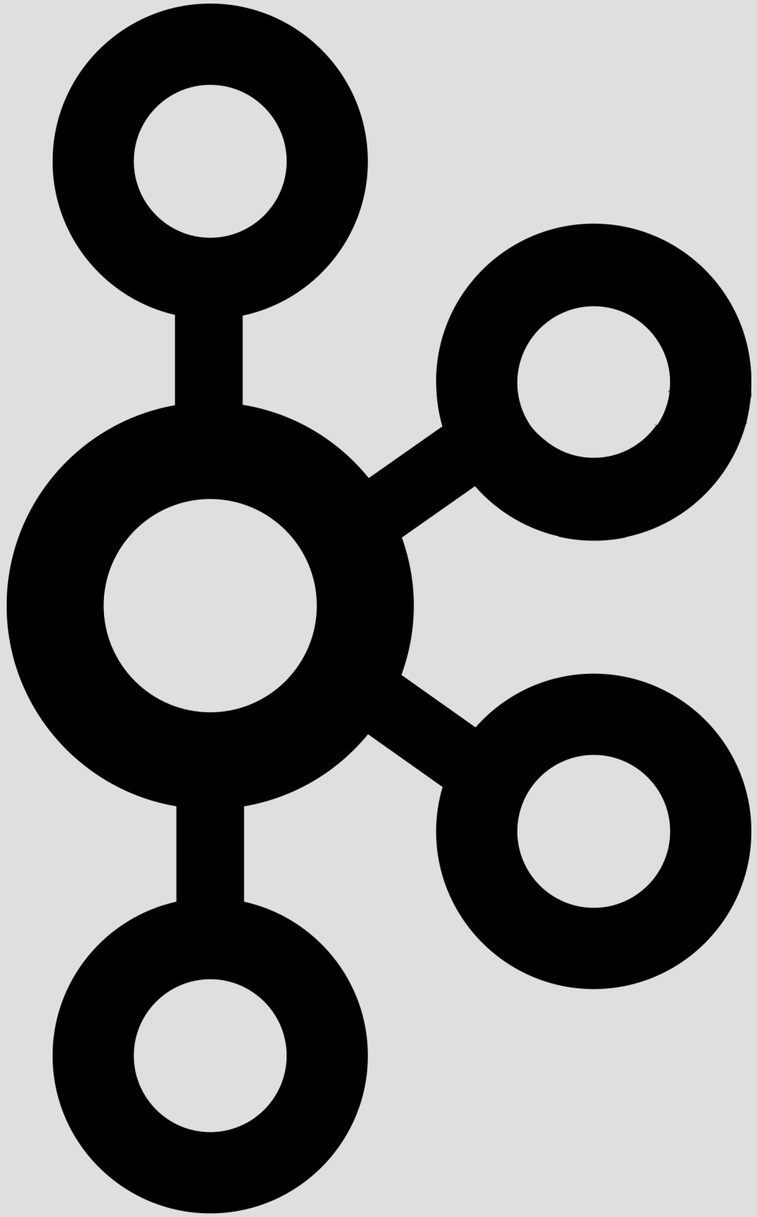
Operational Data Store



MongoDB

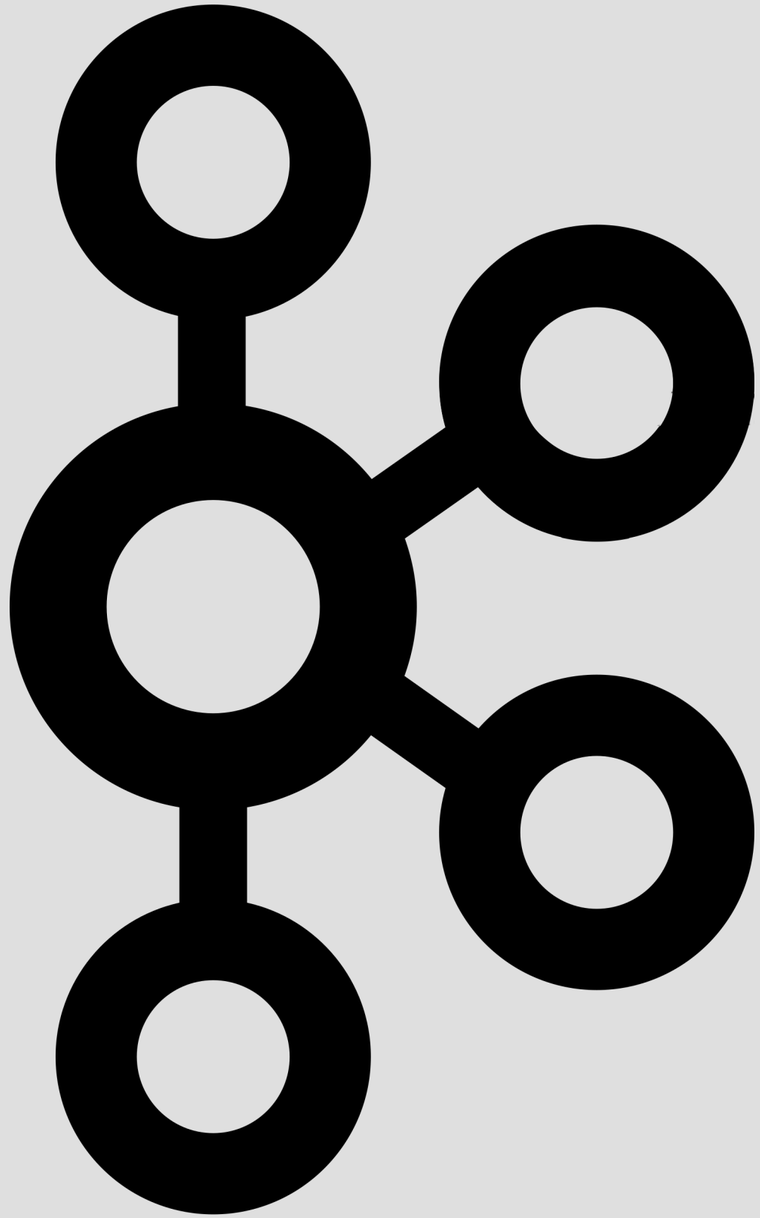
- rich **document** model
- powerful **queries & indexing**
- **ACID** transactions
- transparent **sharding & replication**

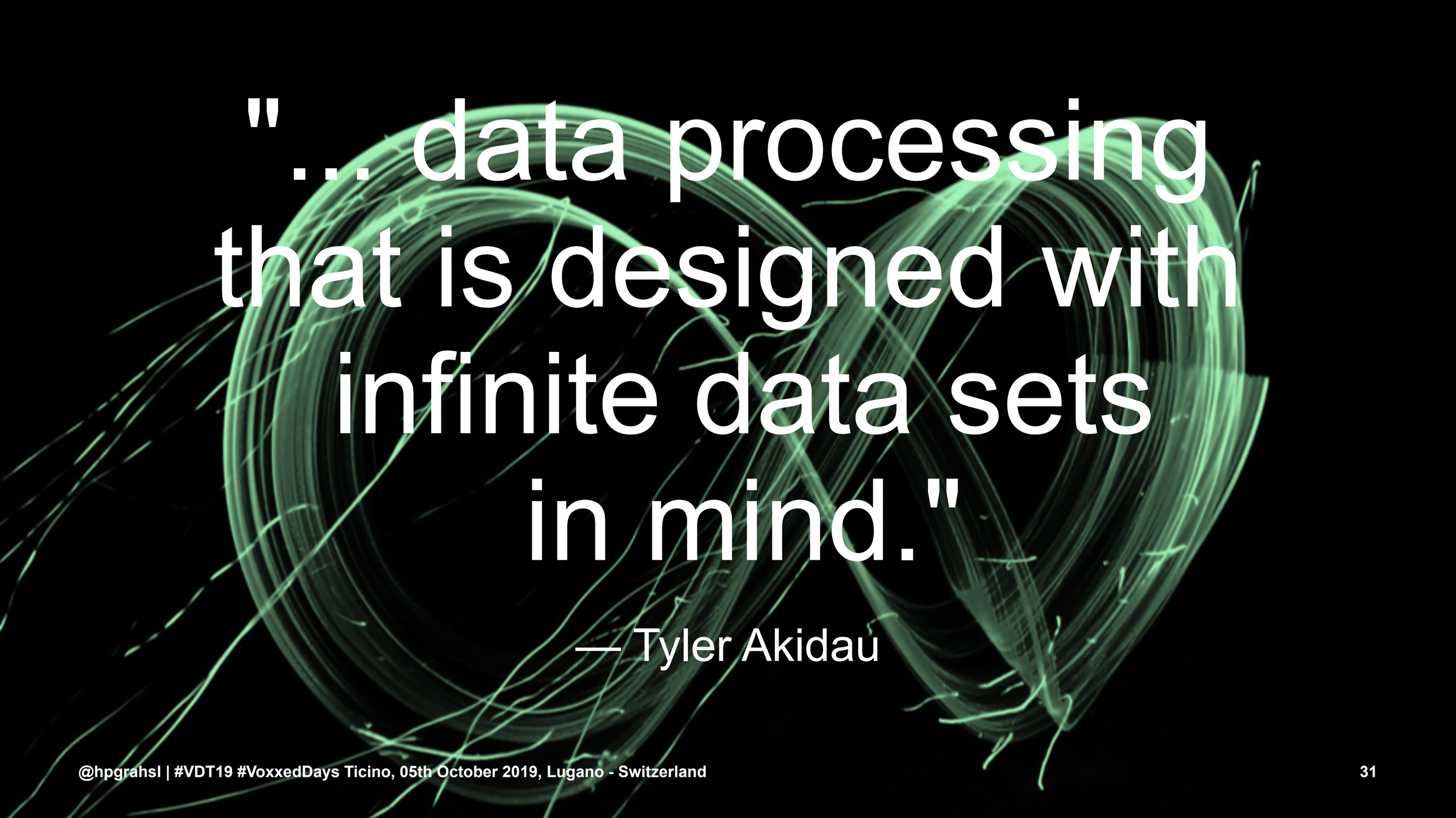
Streaming Platform



Apache Kafka

- **pub / sub** to event **streams**
- (permanently) **store** event **streams**
- event **streaming** in near **real-time**




"... data processing
that is designed with
infinite data sets
in mind."
— Tyler Akidau


EVENTS
SPACE RANGER
LIGHTYEAR
EVENTS EVERYWHERE!

Kafka APIs for "everything"

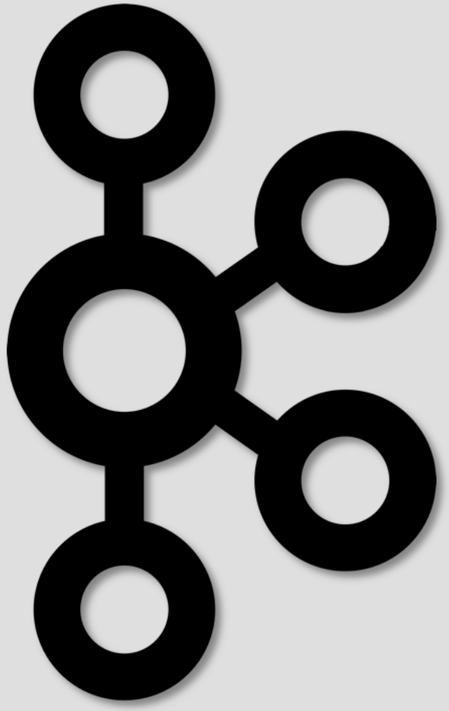
- simple pub / sub scenario ? **Producer & Consumer API**
- streaming data integration ? **Connect API**
- powerful stream processing ? **KStreams API + KSQL**


Kafka
Connect

Kafka Connect

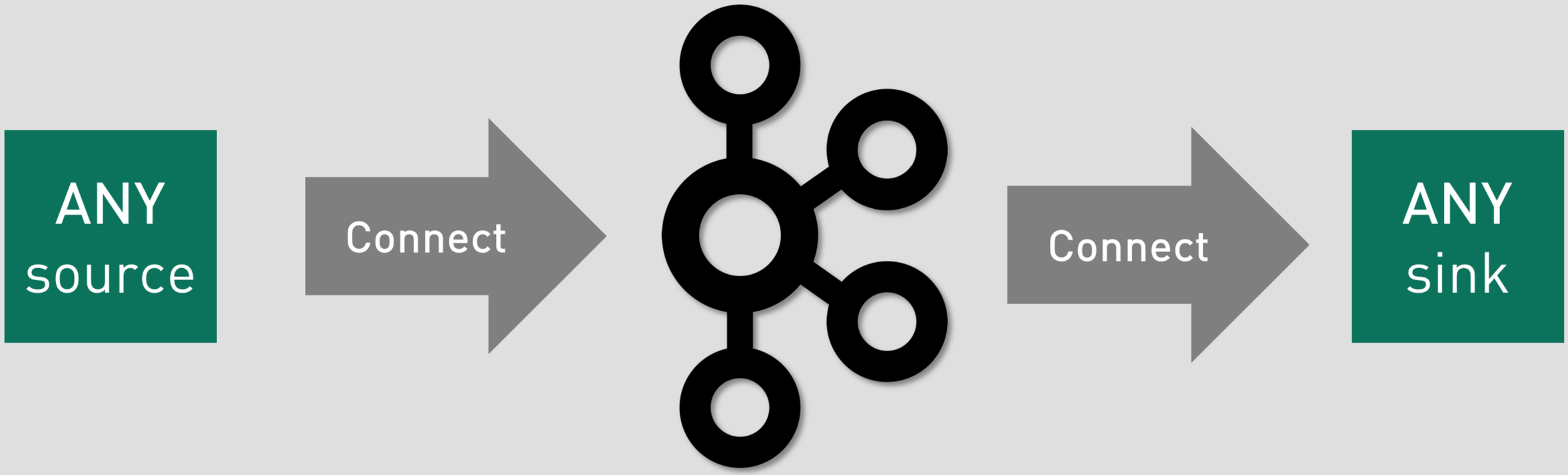
Kafka Connect

ANY source

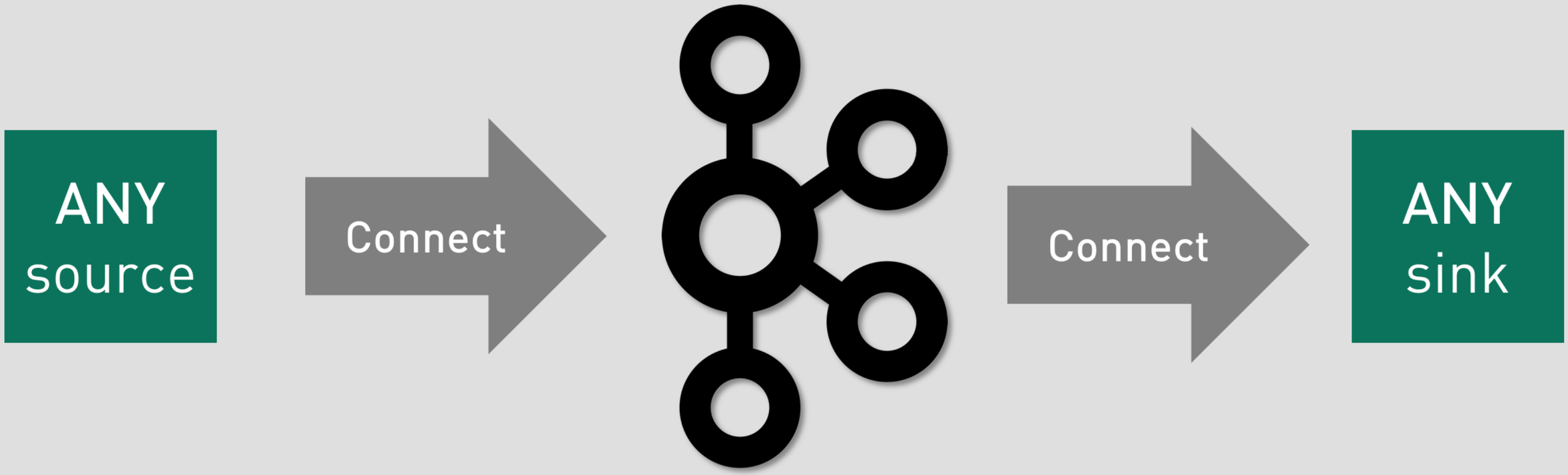


ANY sink

Kafka Connect



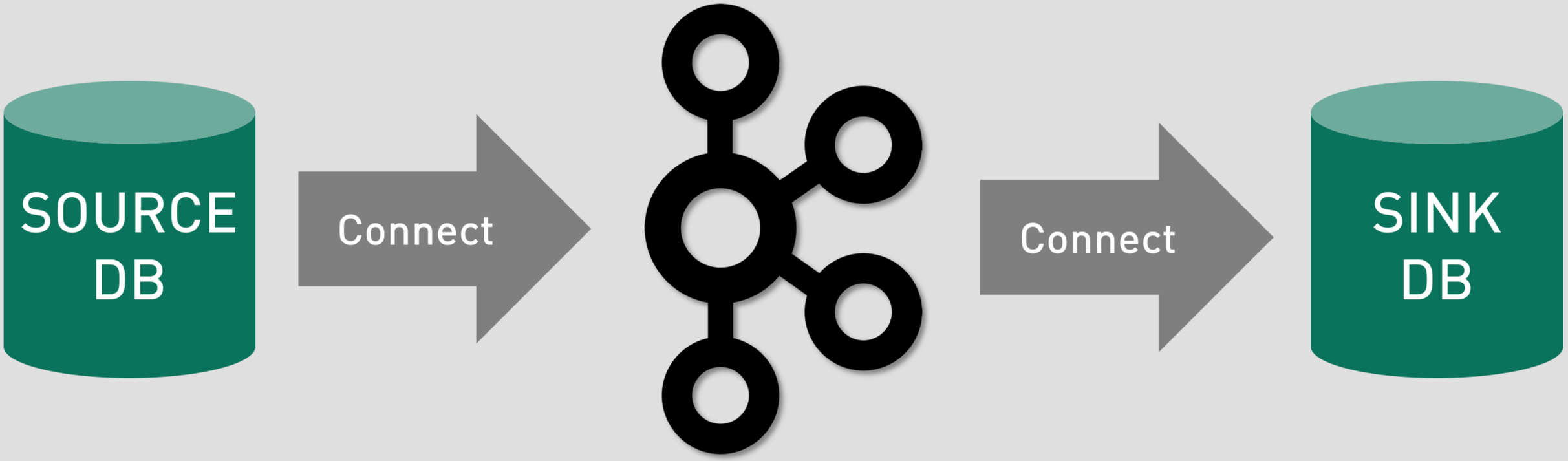
Kafka Connect



ANY → e.g. file systems, **data stores,** REST endpoints, ...

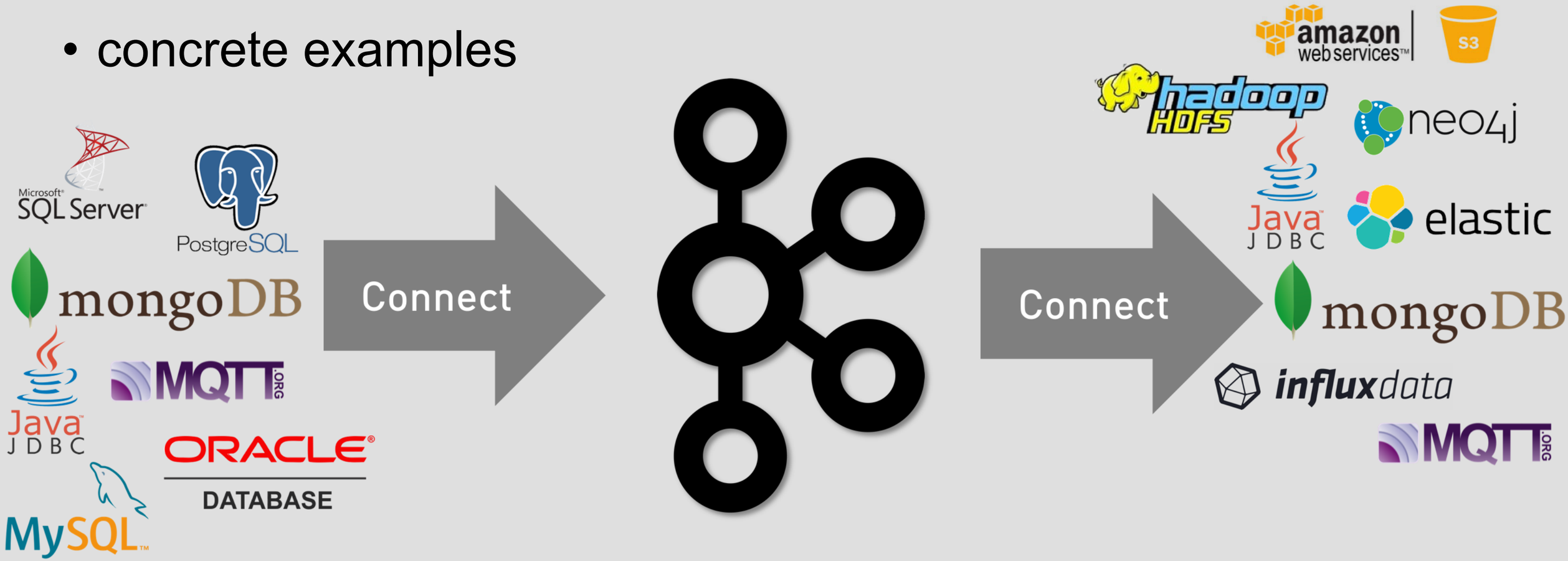
Kafka Connect

- often about data stores



Kafka Connect

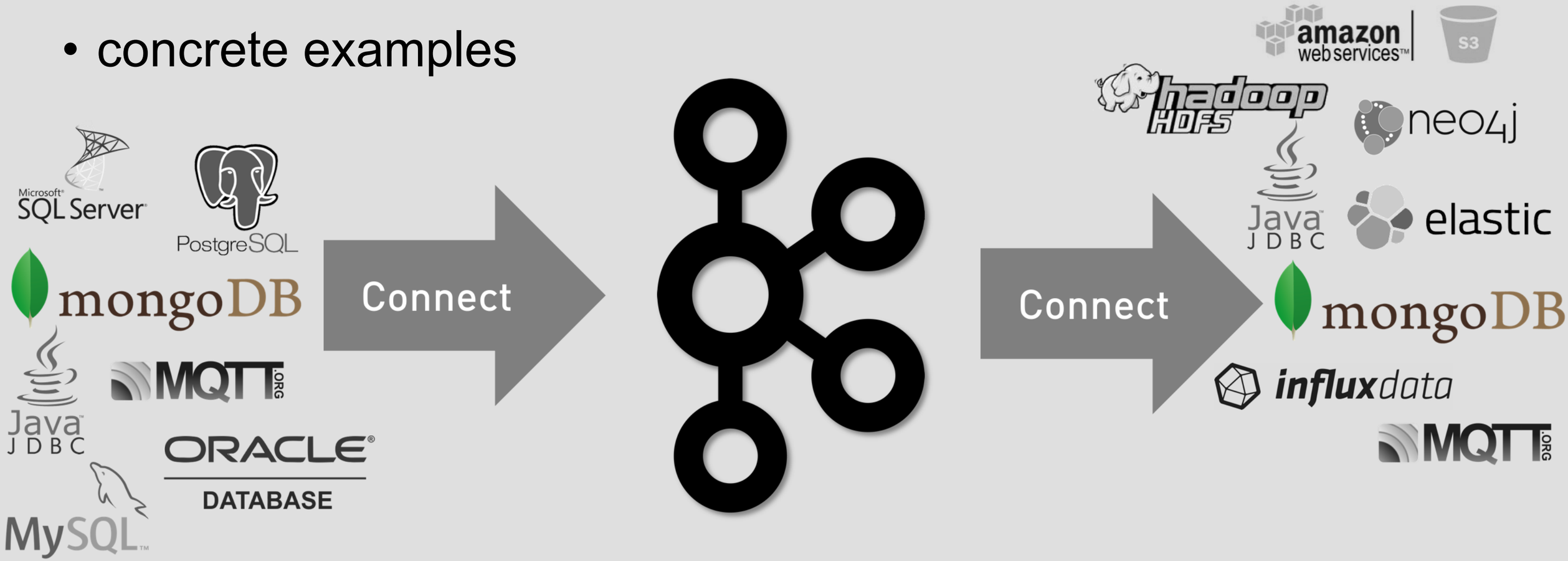
- concrete examples



https://hub.confluent.io → many many more

Kafka Connect

- concrete examples

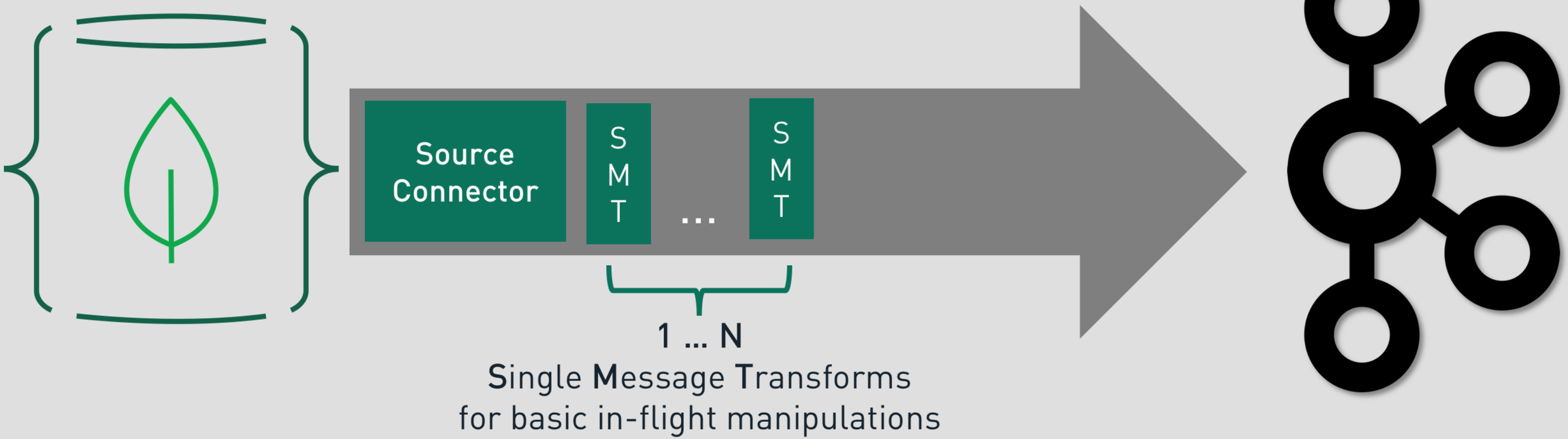


https://hub.confluent.io → many many more

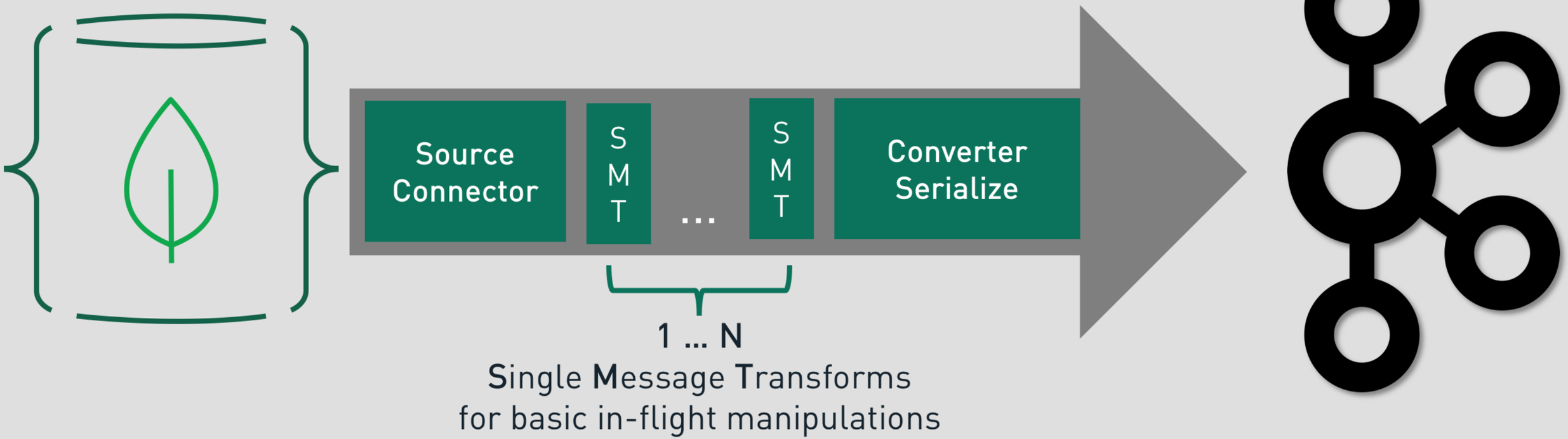
Source Connectors

Source Connectors

Source Connectors



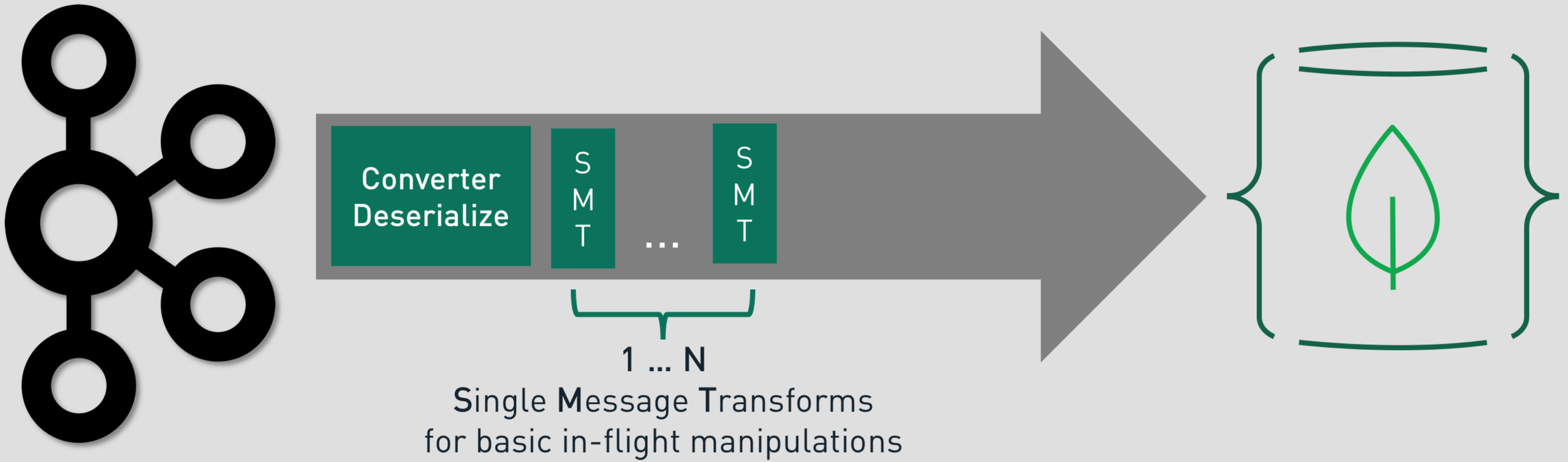
Source Connectors



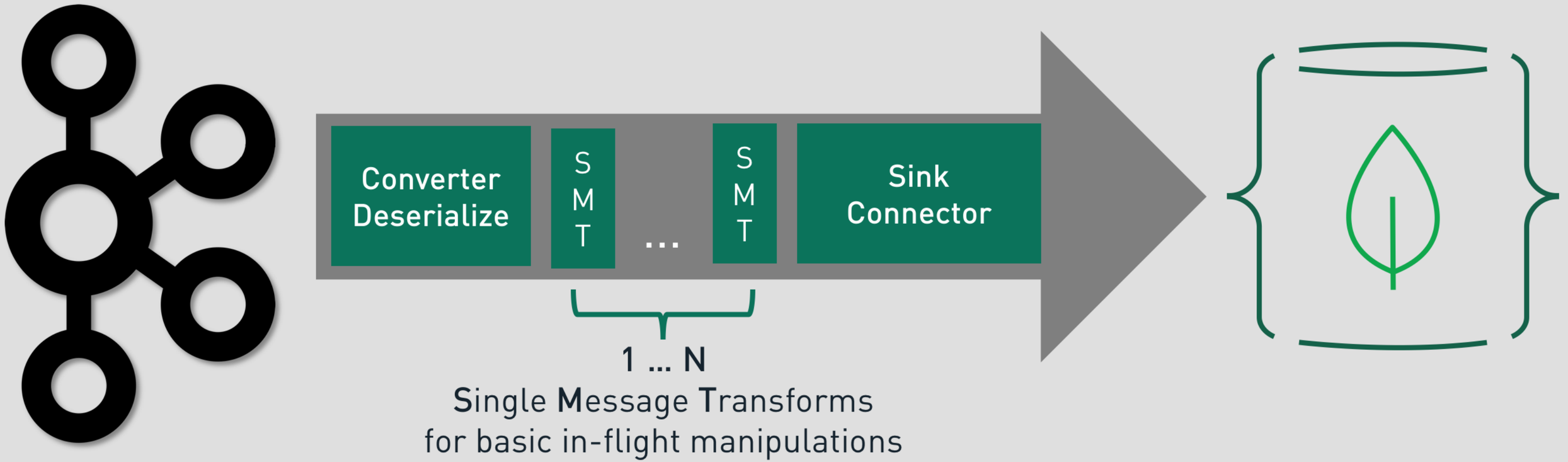
Sink Connectors

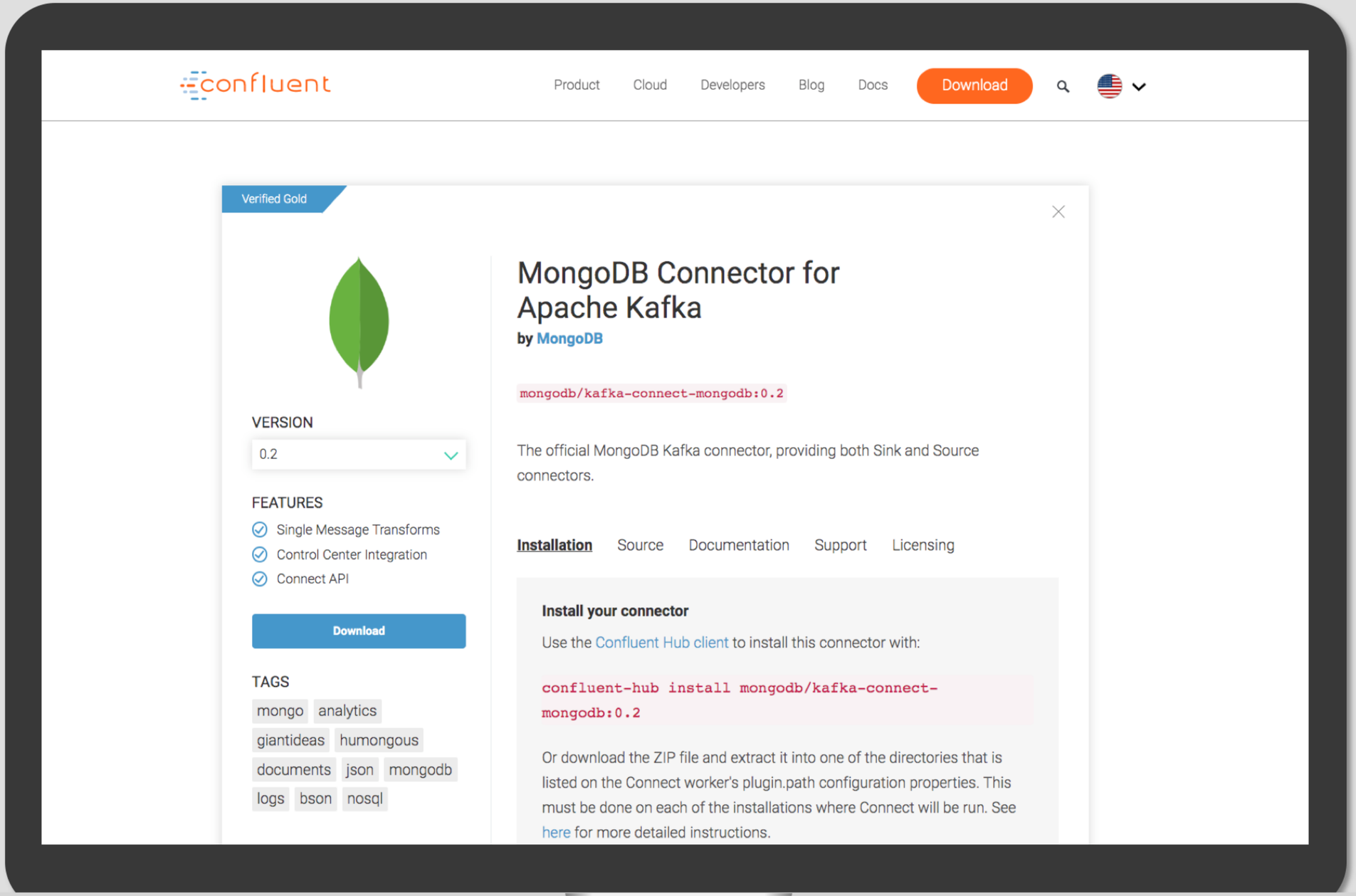
Sink Connectors

Sink Connectors



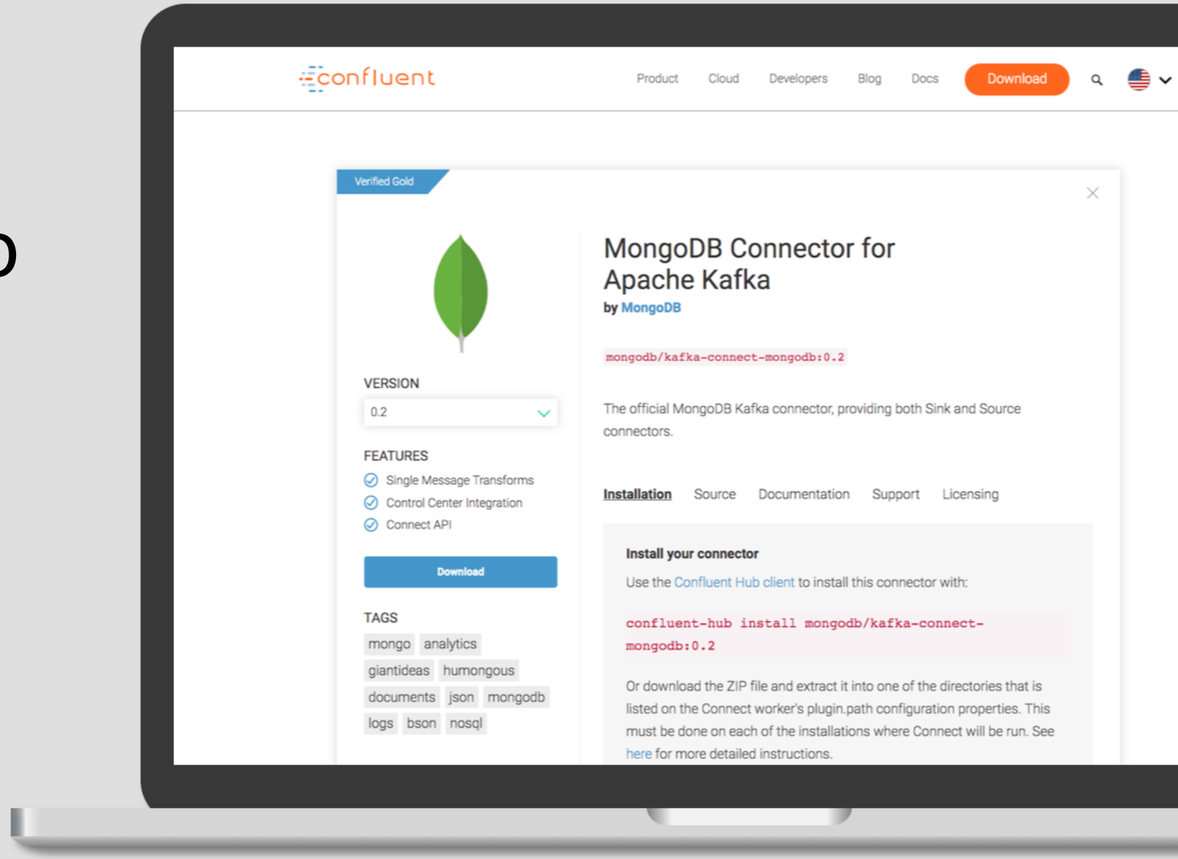
Sink Connectors




confluent
Product
Cloud
Developers
Blog
Docs
Download
Verified Gold
MongoDB Connector for Apache Kafka
by MongoDB
mongodb/kafka-connect-mongodb:0.2
VERSION
0.2
FEATURES
Single Message Transforms
Control Center Integration
Connect API
Download
TAGS
mongo
analytics
giantideas
humongous
documents
json
mongodb
logs
bson
nosql
The official MongoDB Kafka connector, providing both Sink and Source connectors.
Installation
Source
Documentation
Support
Licensing
Install your connector
Use the Confluent Hub client to install this connector with:
confluent-hub install mongodb/kafka-connect-mongodb:0.2
Or download the ZIP file and extract it into one of the directories that is listed on the Connect worker's plugin.path configuration properties. This must be done on each of the installations where Connect will be run. See here for more detailed instructions.

MongoDB Connector

- officially supported by MongoDB
- developed open-source on GitHub
- verified Gold by Confluent

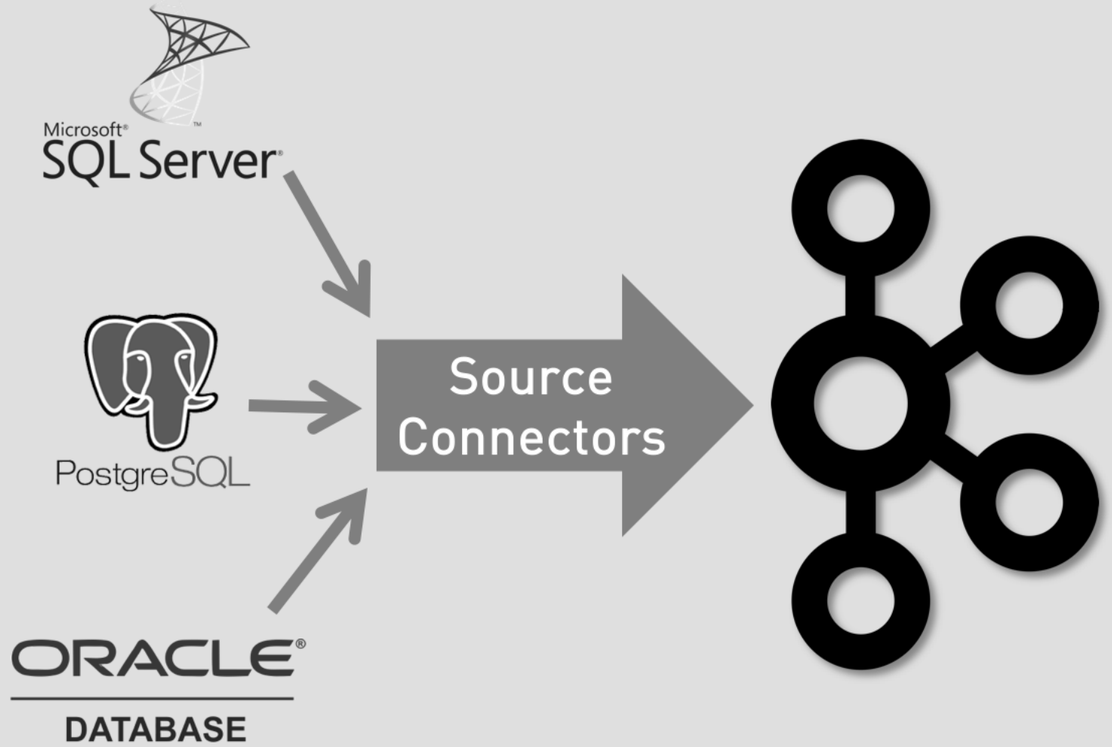


Exemplary Use Cases

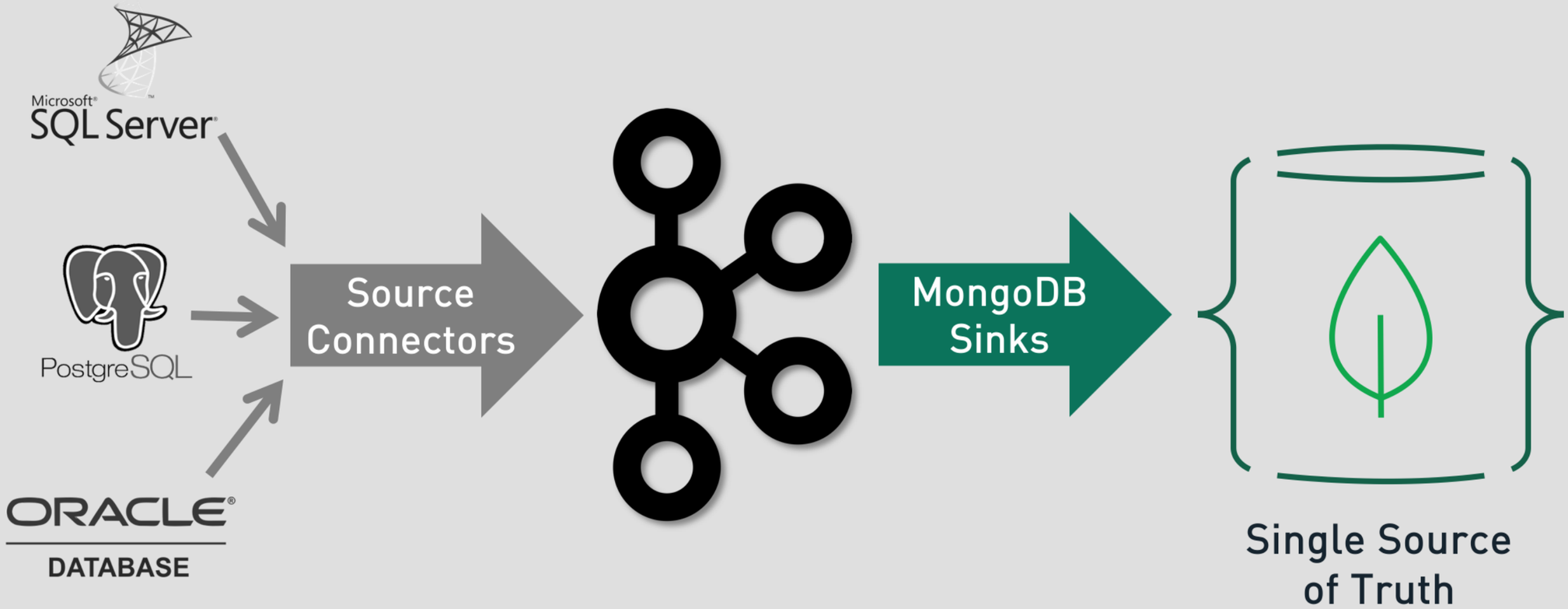
Single Customer View

Single Customer View

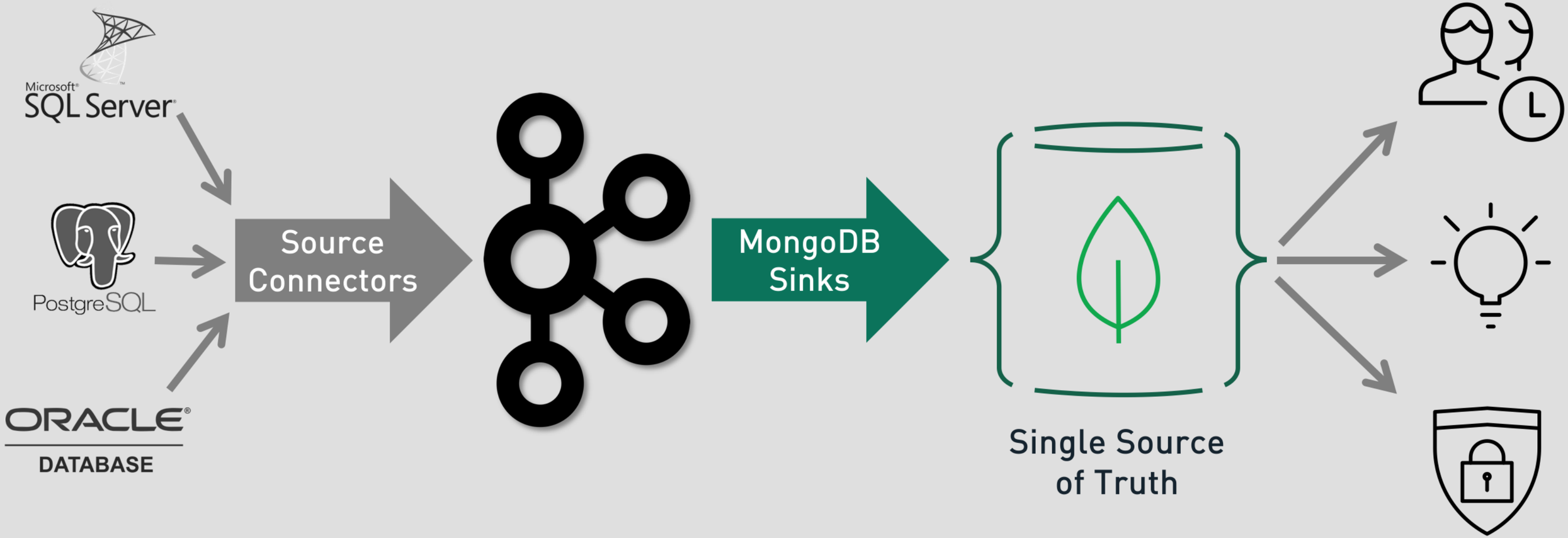
Single Customer View



Single Customer View

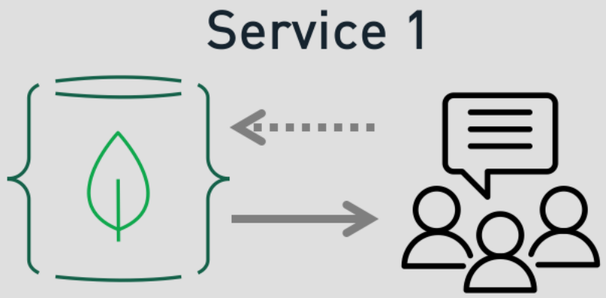


Single Customer View

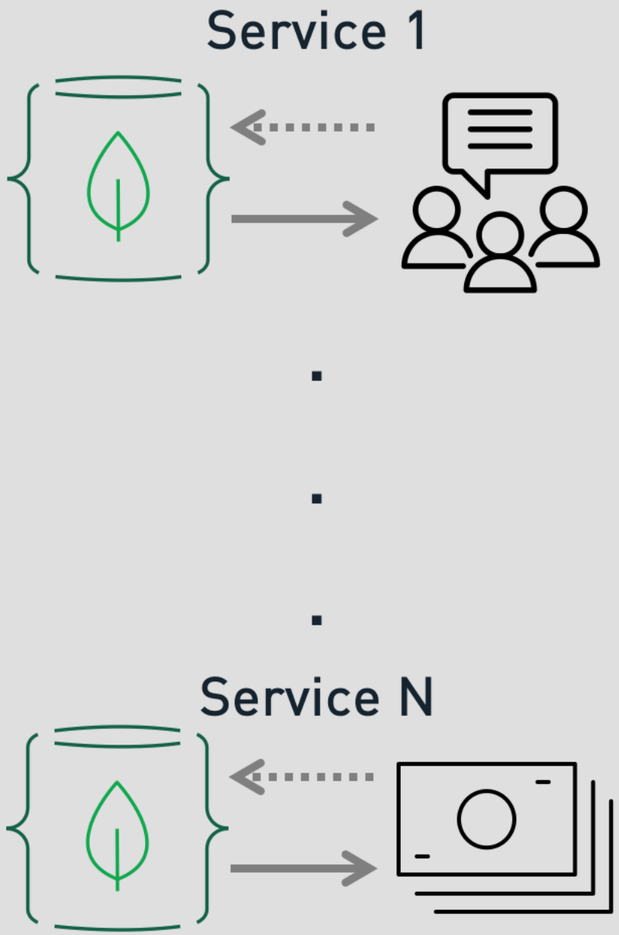


Synchronization across Services

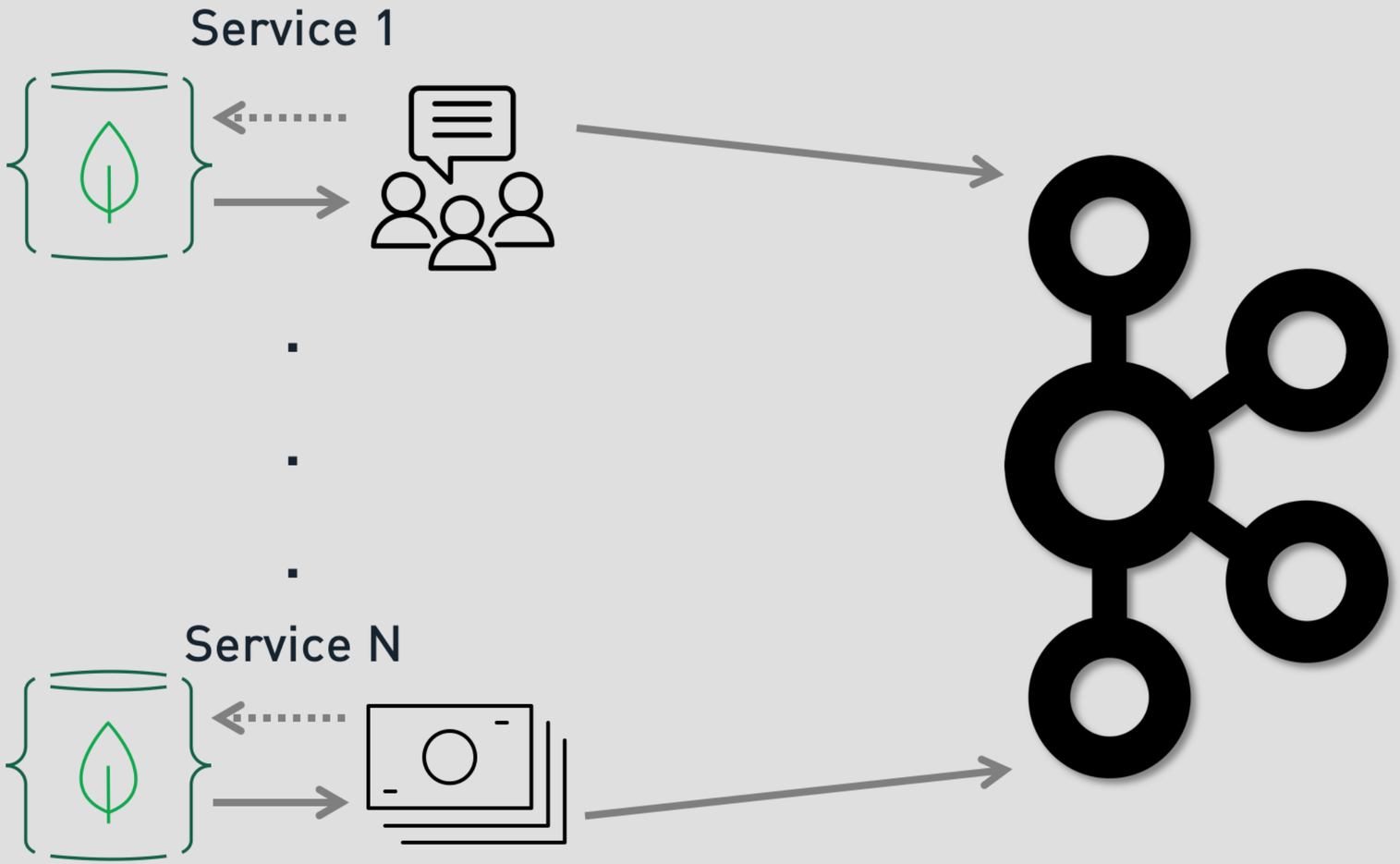
Synchronization across Services



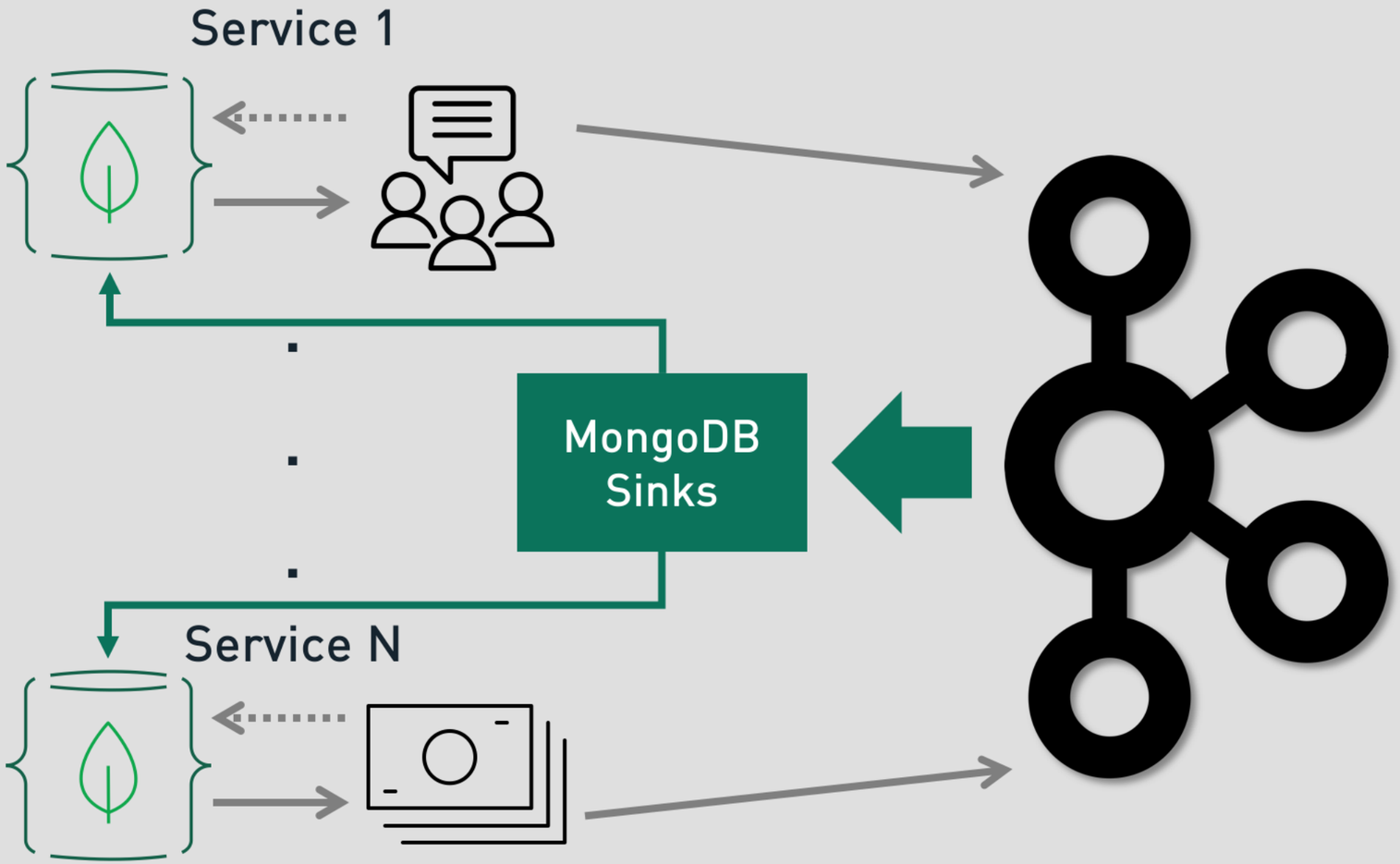
Synchronization across Services



Synchronization across Services

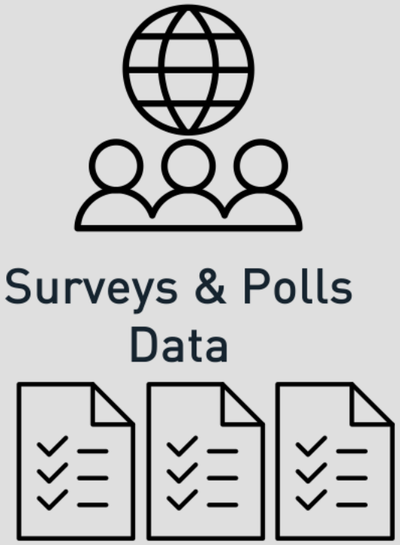


Synchronization across Services

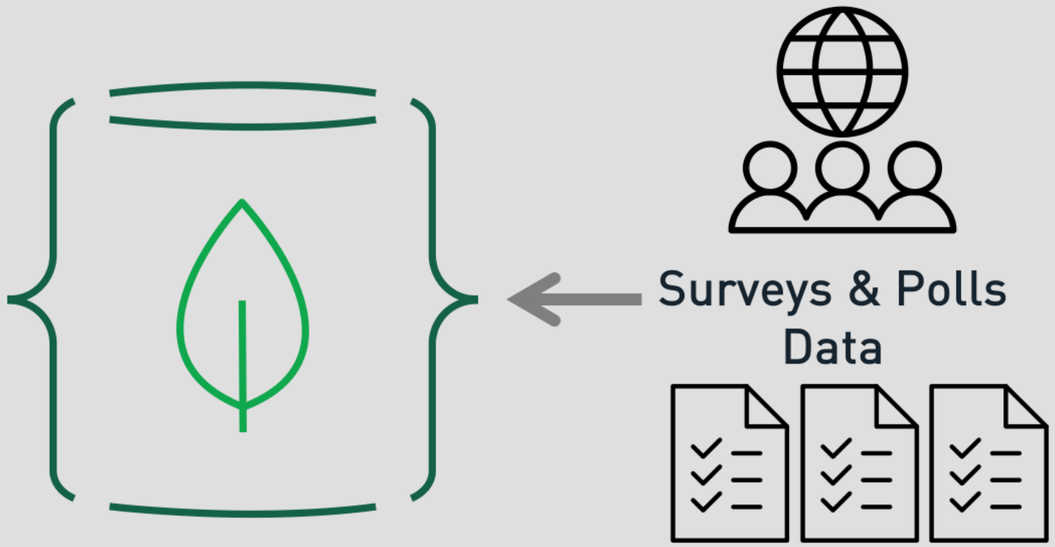


Real-Time Recommendations

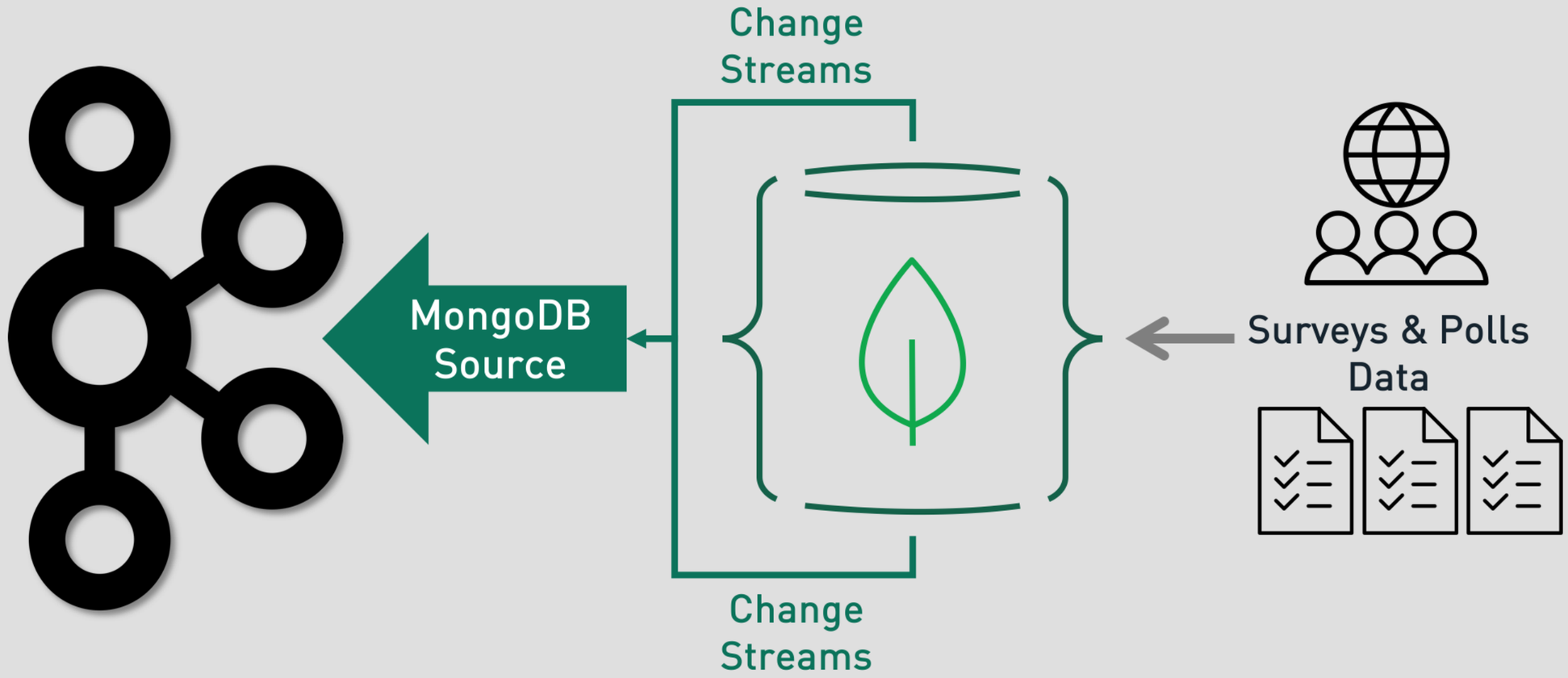
Real-Time Recommendations



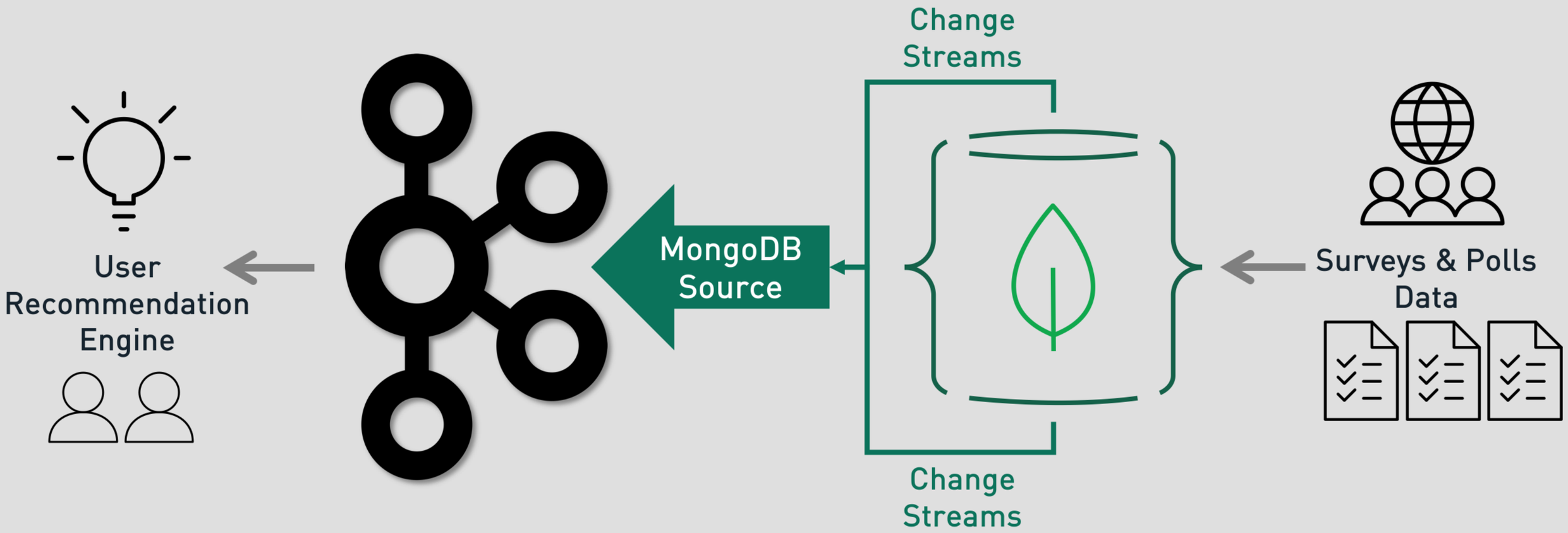
Real-Time Recommendations



Real-Time Recommendations

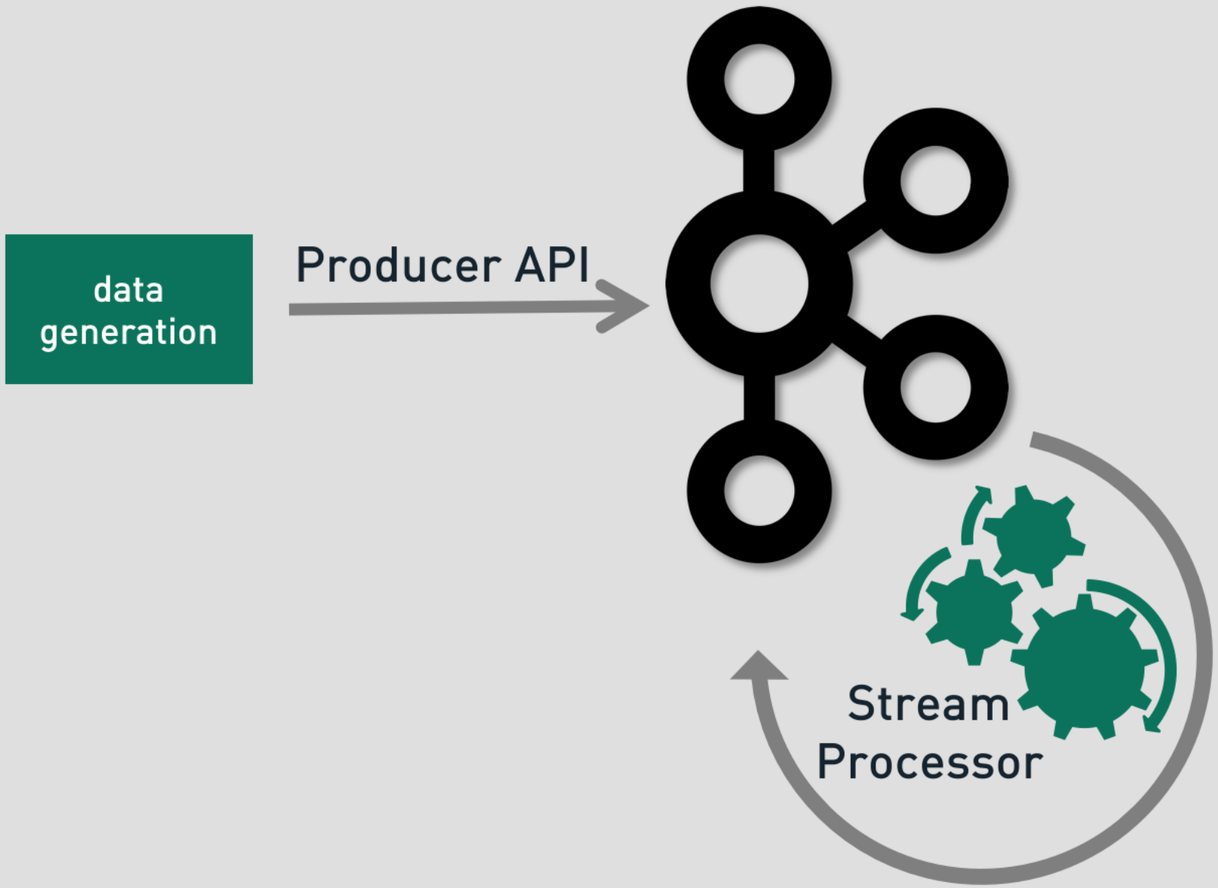


Real-Time Recommendations

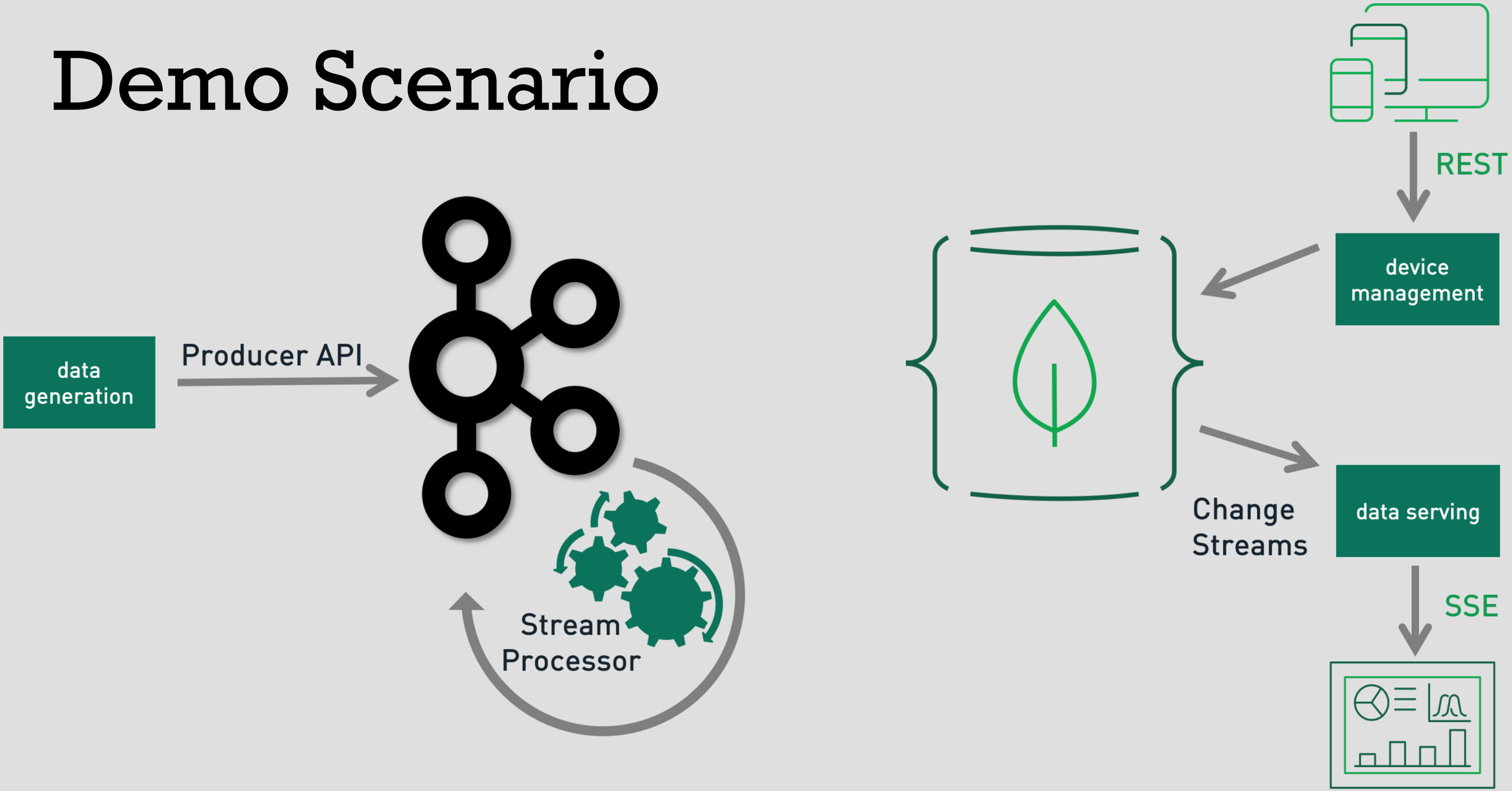


Demo Scenario

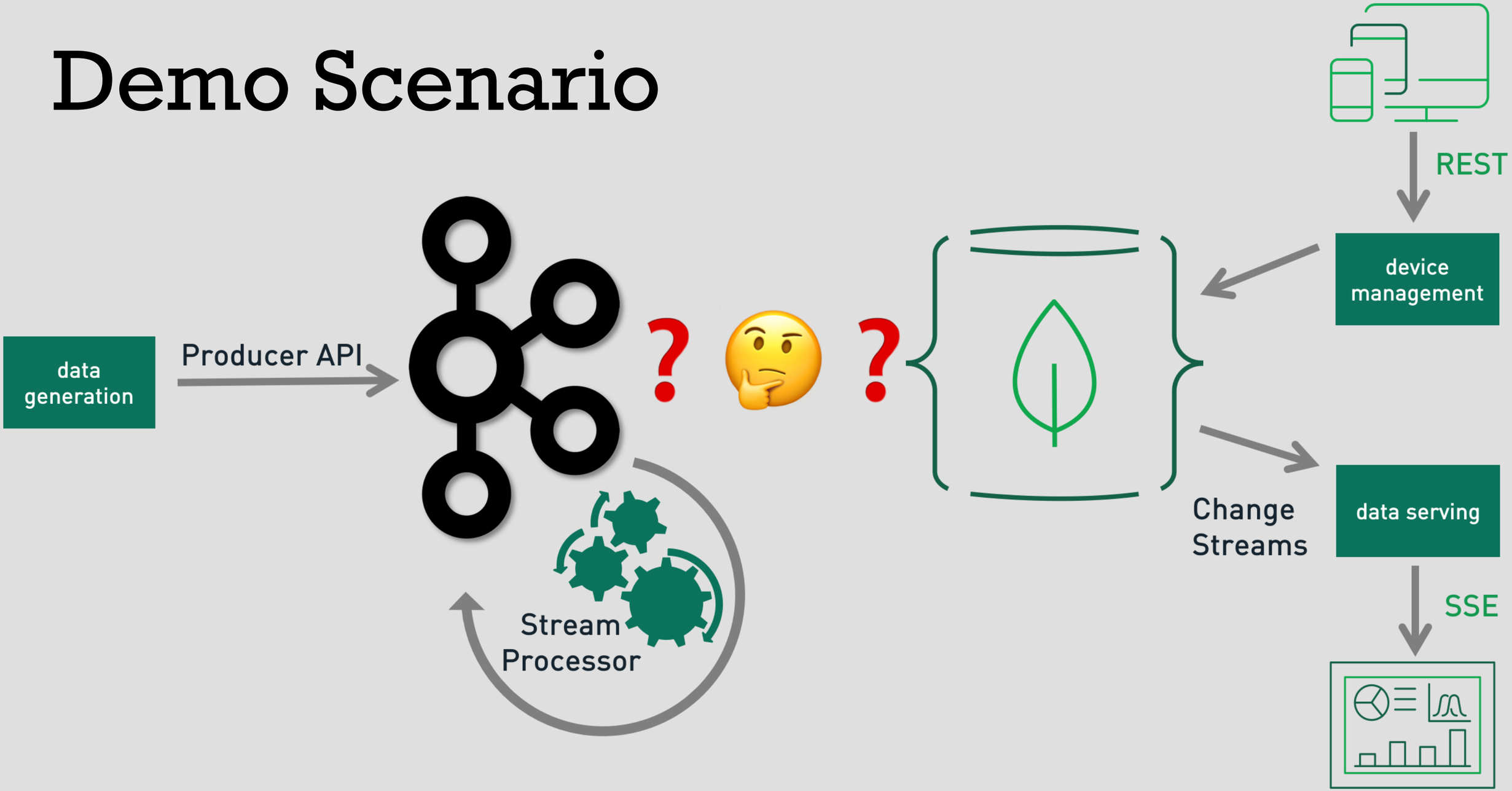
Demo Scenario



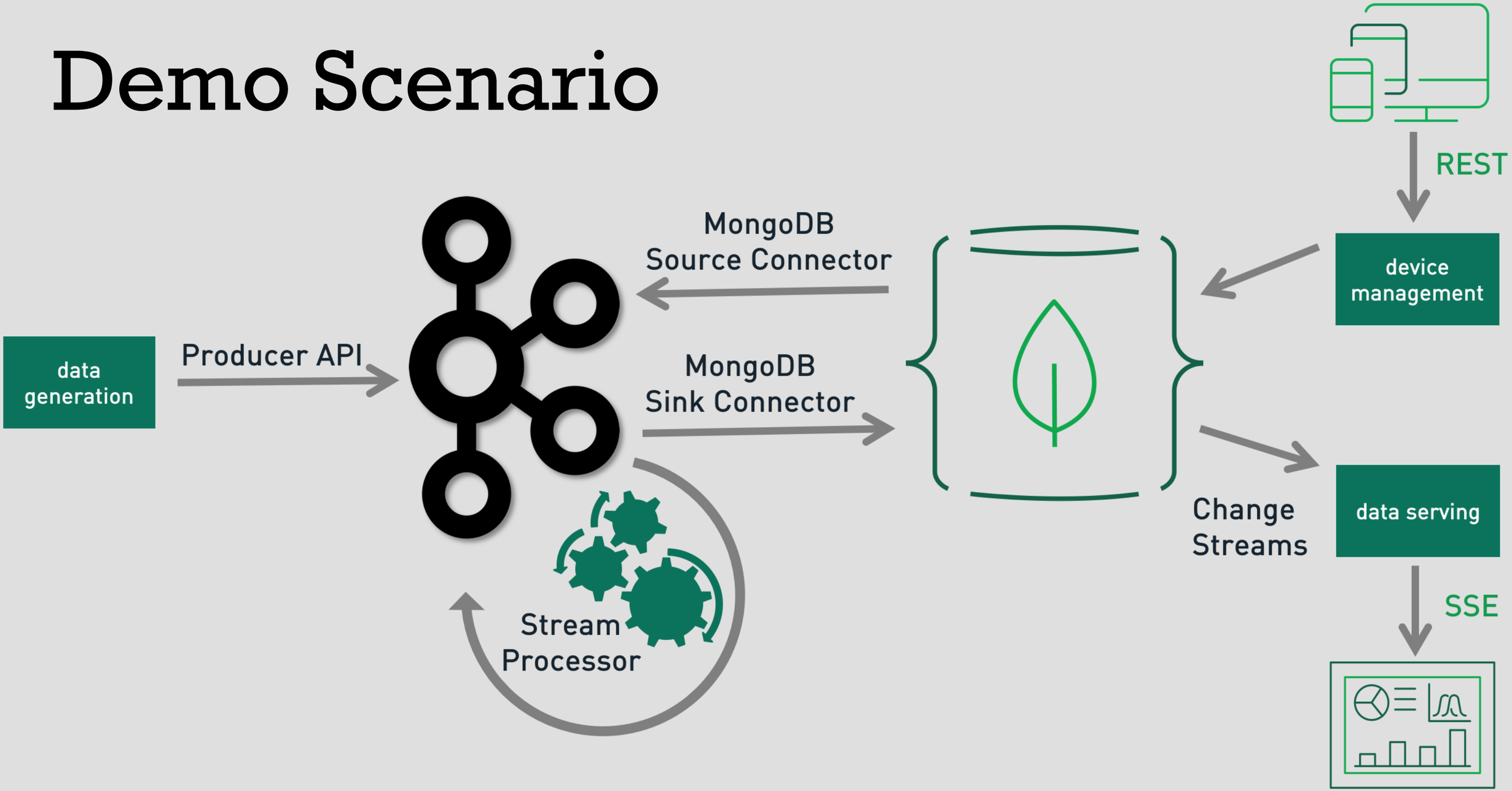
Demo Scenario



Demo Scenario



Demo Scenario



Thank you!