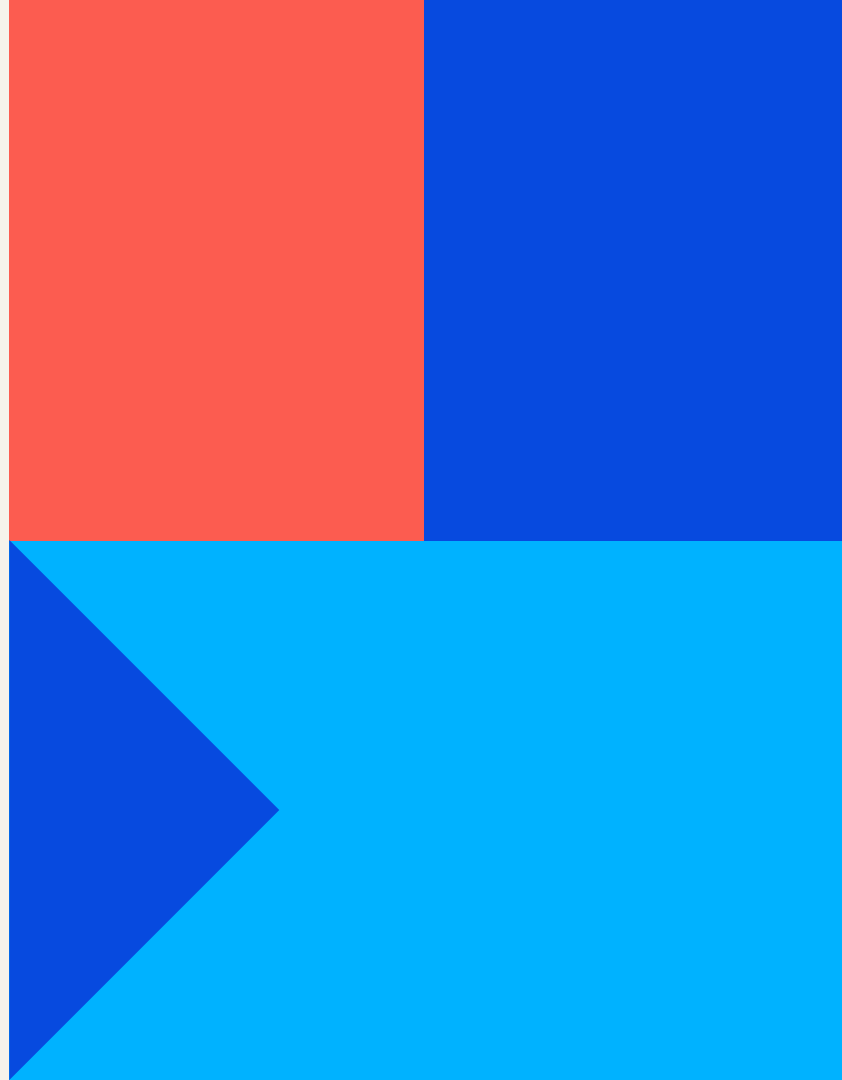


# Dump the Duck Tape

Unblocking data pipelines at scale in  
healthcare using DuckDB

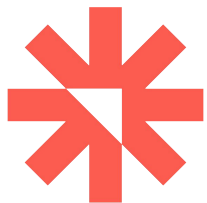
Tony Zeljkovic

Data Engineering Consultant





**Independent Data Engineering Consultant**  
Consulting biotech and medical companies



**curative**

**Leading healthcare services provider**  
delivering a first-of-its-kind employer-based  
health insurance plan through simplicity,  
engagement, and cost transparency

APIs



SFTP



Streams



Data bases



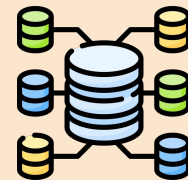
Sources



Loaders



Transformation



Presentation

# Advanced Data Protection

## From Compliance to Dynamic Masking

**PHI and PII:** Handling extensive Patient Health Information (PHI) and Personally Identifiable Information (PII).

**Regulatory Compliance:** Robust security and access control systems required to meet requirements like HIPAA/HITECH, SOC2, FedRAMP, etc...

**Dynamic Data Masking System:** Manage access through a dynamic masking system with column-level granularity for **±10,000 tables/views** and **100,000+ columns**.

GitHub  
Pull Request

Buildkite

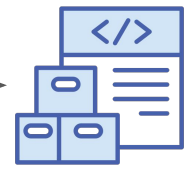


dbt

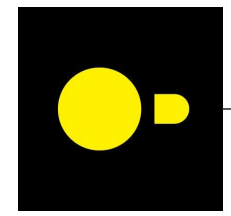
Transform data

snowflake

Produce artifacts



Execute SET TAG



Parse metadata

GitHub  
Pull Request

Buildkite

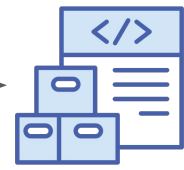


dbt

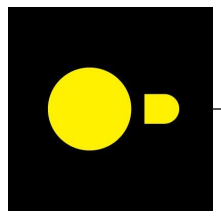
Transform data

snowflake

Produce artifacts



Execute SET TAG



Parse metadata

**Naive DBT solution:** 1-2 hours v.s **DuckDB solution:** 5 minutes

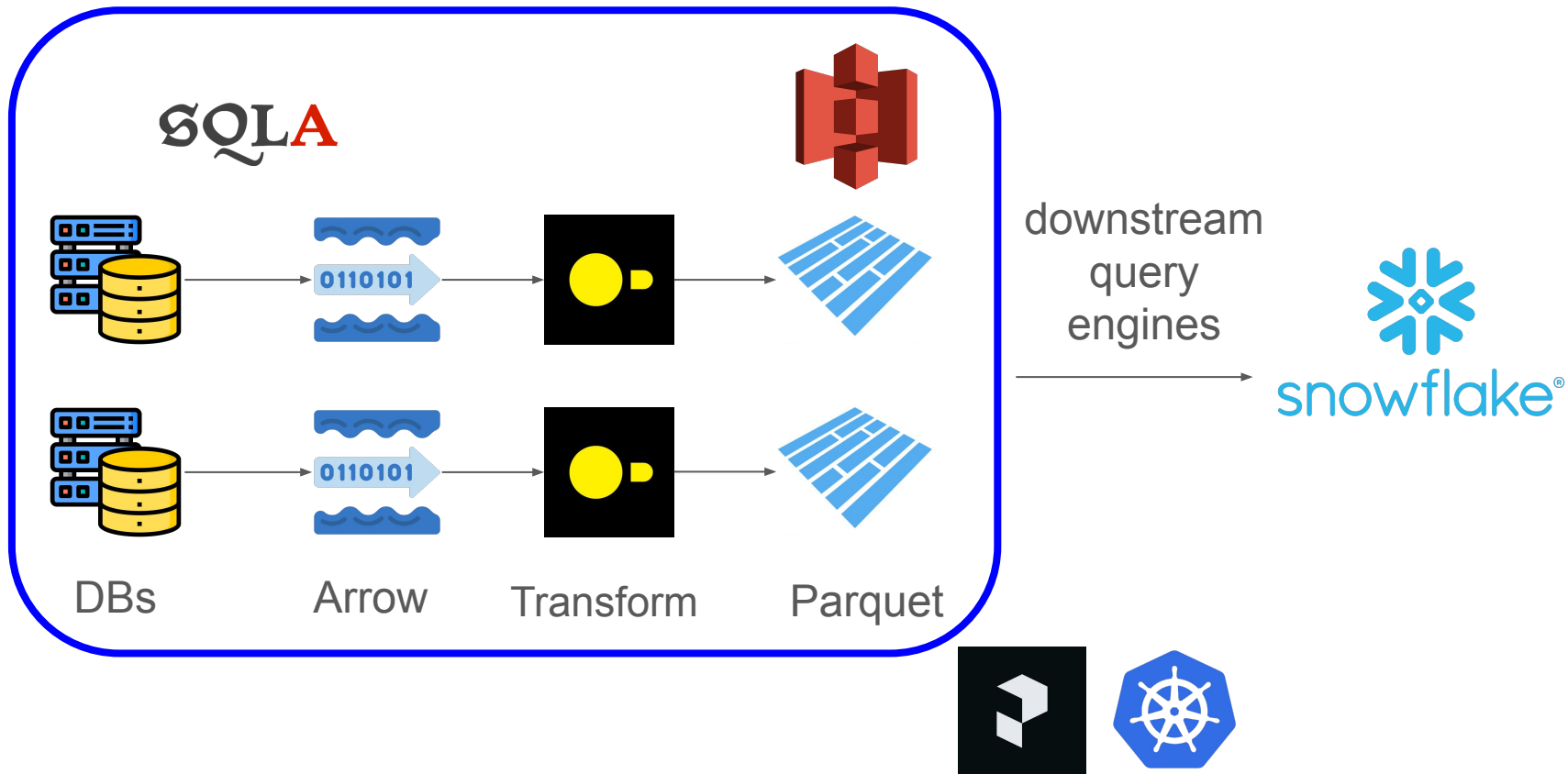
# Integrating Healthcare Data At lightspeed with DuckDB



**Operational Databases:** Integrating a diverse set of operational databases and data warehouses.

**Strict SLAs:** Performance, observability and robustness crucial for rapidly growing critical healthcare operations.

**In-House Connector Development:** Proprietary connectors powered by DuckDB meet stringent service-level agreements (SLAs) for our most vital operational systems.



SQLA



Application Performance Monitoring

---

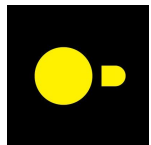
Queries, query latency, db load

---



CPU, memory, node load, cost

---



profiling?



S3 events, latency, datasets

---



snowflake®

Sf compute, query time, access

---



DATADOG

# Thank You!

