Lecture HSE · May 23, 2026

Data Platform Basics

A basic introduction: why data platforms exist, how ETL/ELT loops work, and how maturity is measured

Alexander Polomodov

Speaker Context

Technical Director & Fellow, T-Technologies

- Architecture and engineering R&D.
- AI adoption in SDLC.
- Focus: data platform as product.



Part 1

01

Why a data platform exists

From raw events to reliable decisions, products and AI scenarios

Data becomes a production loop

Speed

Source → mart,
API, model, report

Trust

Ownership, quality,
lineage, access,
freshness

Scale

One loop serves
analytics, products,
ML and AI

Speed, trust and scale are properties of the whole system

Stores and Scripts Break

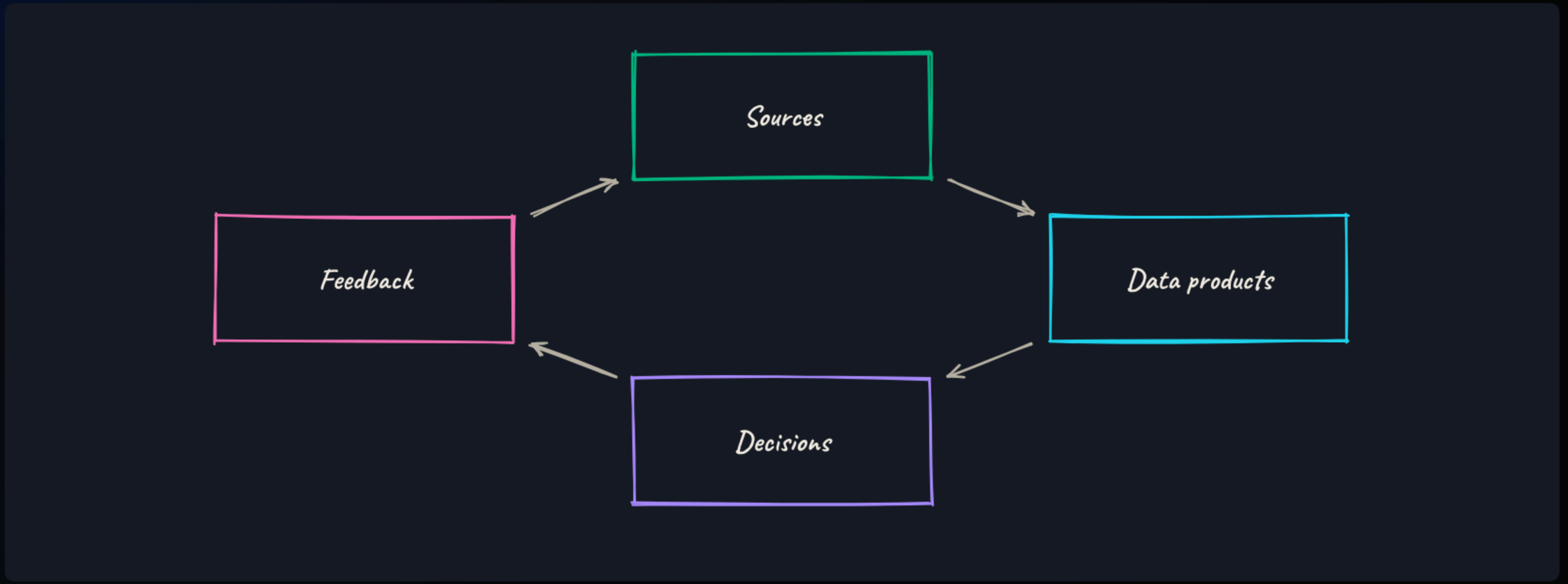
Problem | No platform

Accountability breaks

- Teams rediscover sources and access.
- Business logic hides in unowned SQL.
- Data incidents are noticed too late.
- Cost is not tied to value.

The platform connects data to decisions

Loop `Value loop`





Part 2

02

ETL, ELT and hybrid loop

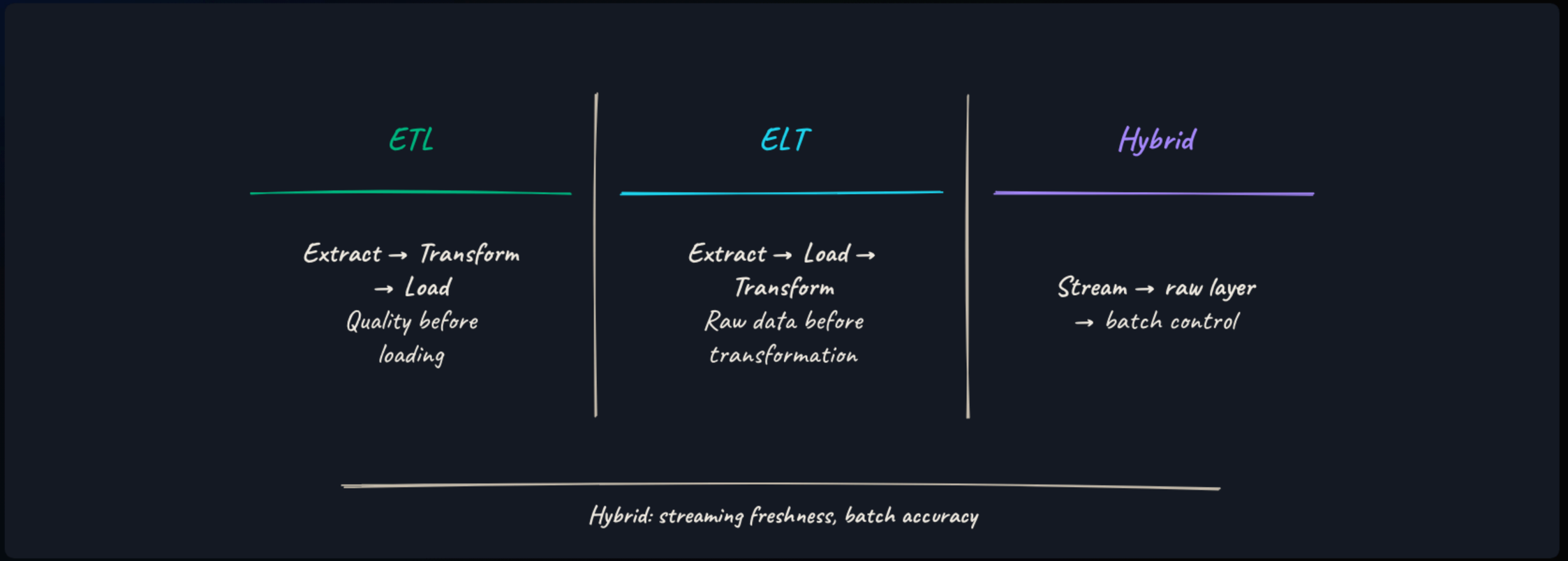
How data moves from source to consumer

ETL vs ELT


ETL
Extract
Transform
Load
ELT
Extract
Load
Transform
Transform before or after loading

ETL, ELT and the hybrid lakehouse loop

Architecture Hybrid



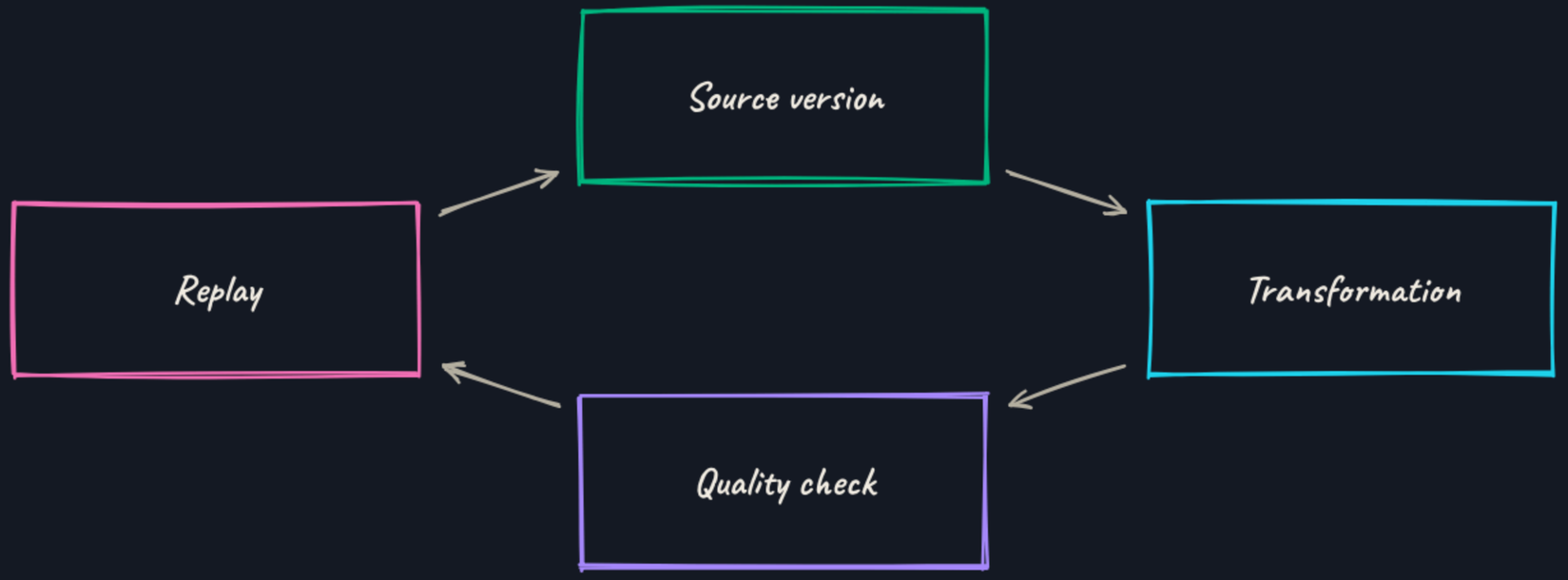
Reference data platform architecture

Layers Reference

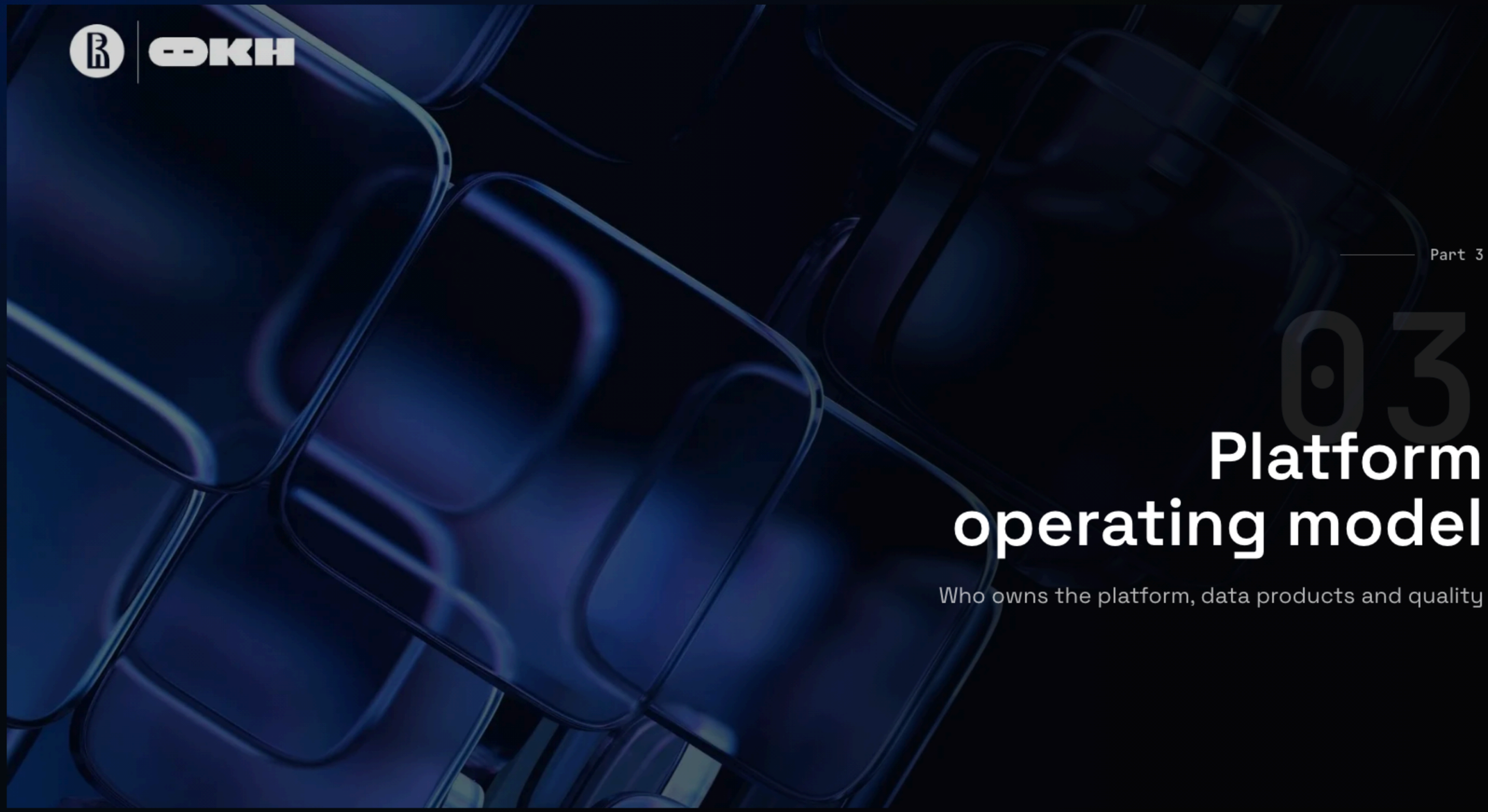


Pipelines: replayable and explainable

Reliability Pipeline



Exactly-once is rarely realistic; backfill needs a separate path



Part 3

03

Platform operating model

Who owns the platform, data products and quality

A data platform is also org design

Technology will not work without clear accountability

Models Ownership



Centralized

One team sets tools and rules.

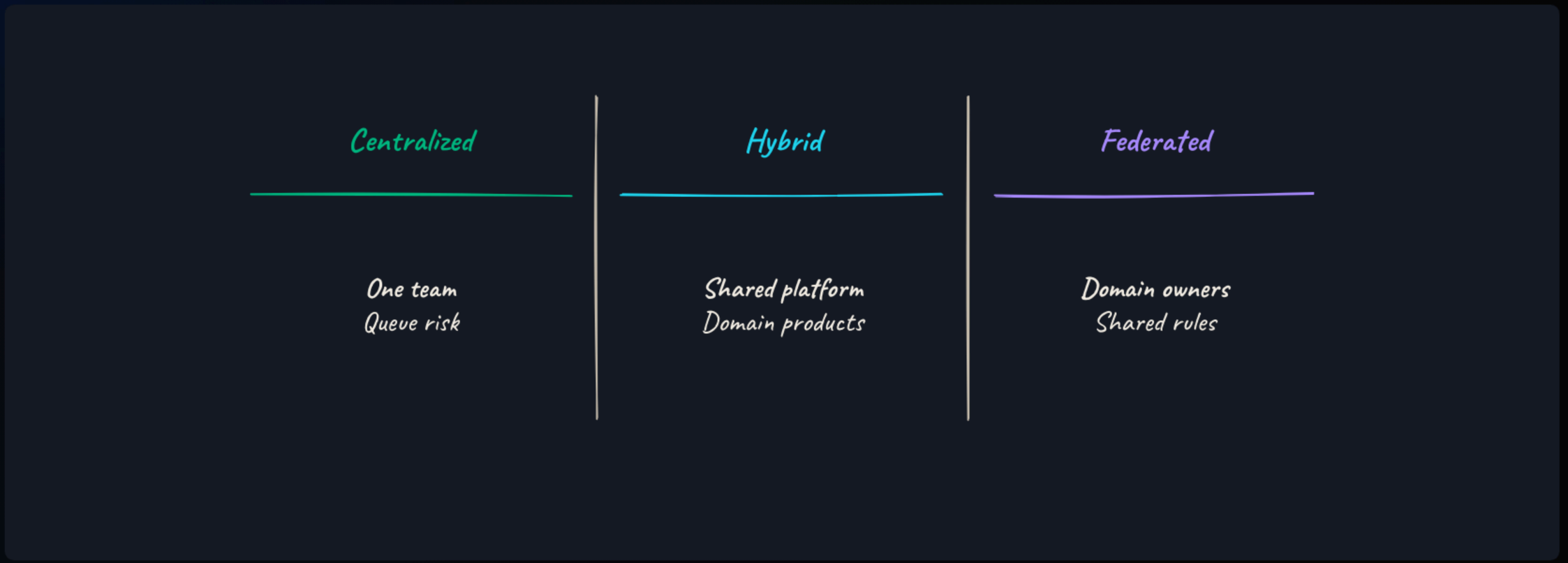
Hybrid

Platform owns capabilities; domains own products and SLAs.

Federated

Data Mesh: domains + shared standard.

Central → hybrid → federated model



Data Mesh needs a platform

Data Mesh | No myth

Federation needs a shared layer

+ Domain ownership

- Meaning, rules, priorities.
- Data products and consumers.
- Freshness, completeness, compatibility.

+ Platform ownership

- Self-service ingestion, storage, compute, orchestration.
- Catalog, lineage, access, quality.
- Financial and engineering guardrails.



Part 4

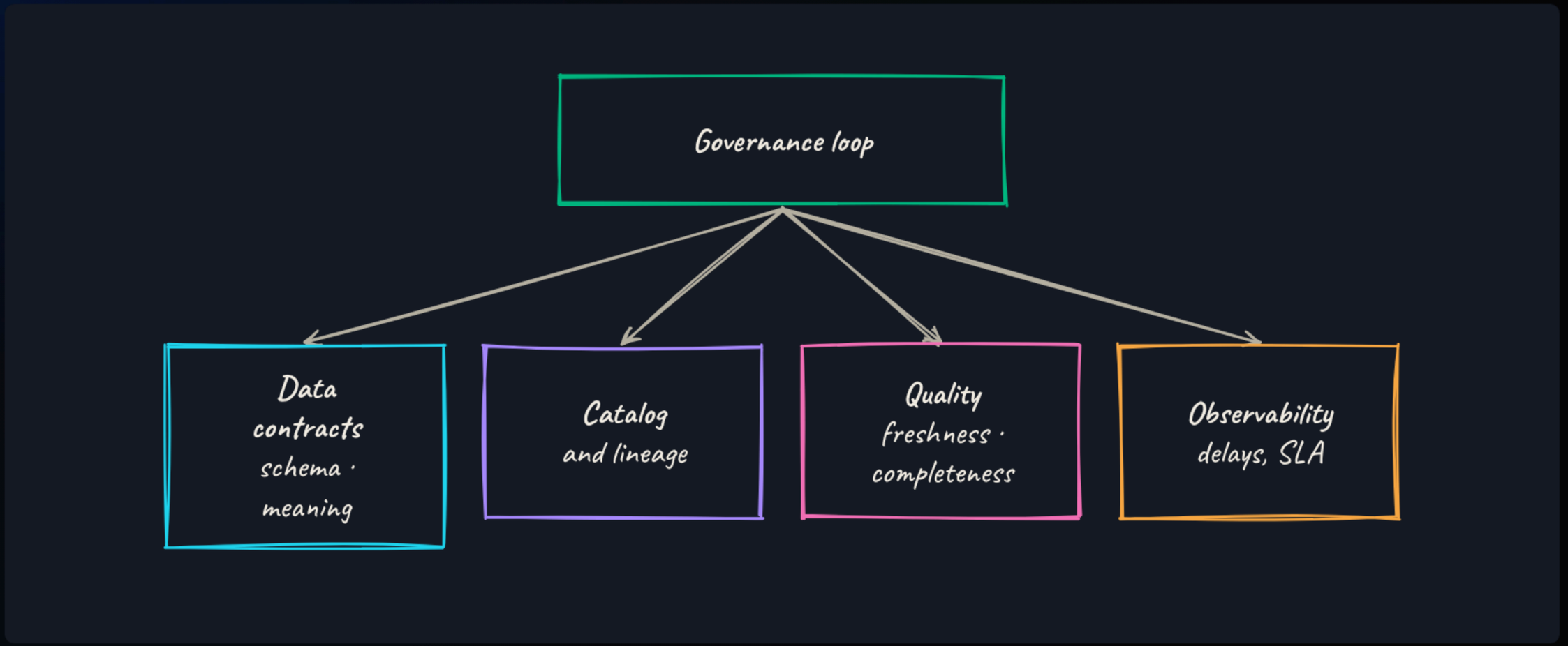
04

Governance and reliability

Contracts, lineage, quality, observability and cost

Basic Governance Loop

Governance Minimum



Where data platforms usually break

Anti-patterns Risks

Accountability breaks before tooling does

One giant DAG

Local change becomes systemic risk.

Hidden logic

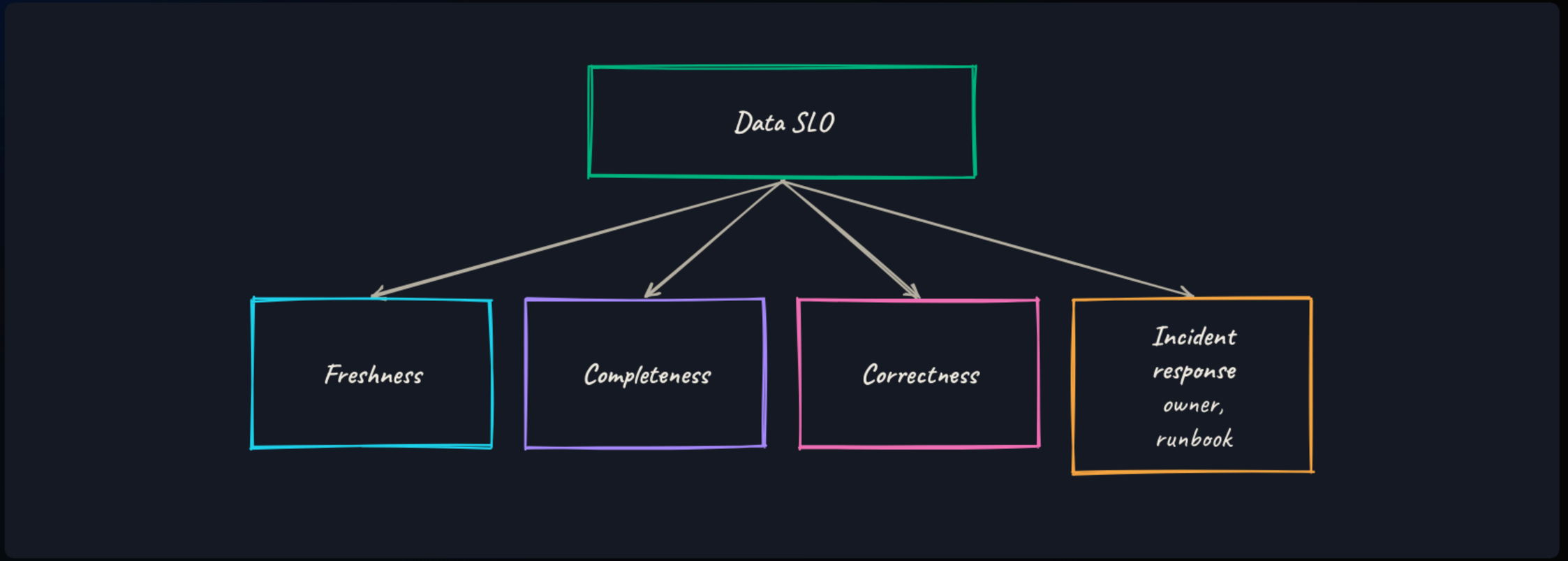
SQL rules lack tests, versions and review.

No economics

Mart and storage cost stays invisible.

Data Service Level Must Be Explicit

Practice SLO





Part 5

05

Platform maturity metrics

How to know the platform became a product, not a tool collection

Data platform maturity metrics

Maturity Scorecard

Speed	Reliability	Reuse	Economics
First dataset < 1 week	Pipeline ≥99.9% Freshness SLO ≥95% Schema issues ↓ QoQ	Reuse share > 60%	Cost per insight controlled

First 120 days Roadmap

First 120 Days of Maturity

Start with map, owners, change rules

- 0-30: sources, consumers, owners, incidents.
- 30-60: contracts, catalog, quality, freshness.
- 60-120: templates, runbooks, showback.
- Then: metrics → product value.

What to take away

Takeaways Next

A data platform is a managed path from source to decision

- ETL/ELT/hybrid answer different trade-offs.
- Operating model matters as much as stack.
- Maturity = speed, quality, freshness, economics.
- Next: DataOps/MLOps in production.

- Next: DataOps/MLOps

Links and materials

Sources References

Reference sources

- ETL/ELT pipeline architecture (Russian): ↗system-design.space
- Interview with Nikolay Golov, 2025 (Russian): ↗system-design.space
- OpenLineage, Apache Airflow and dbt documentation.
- Data Mesh and operating model materials.