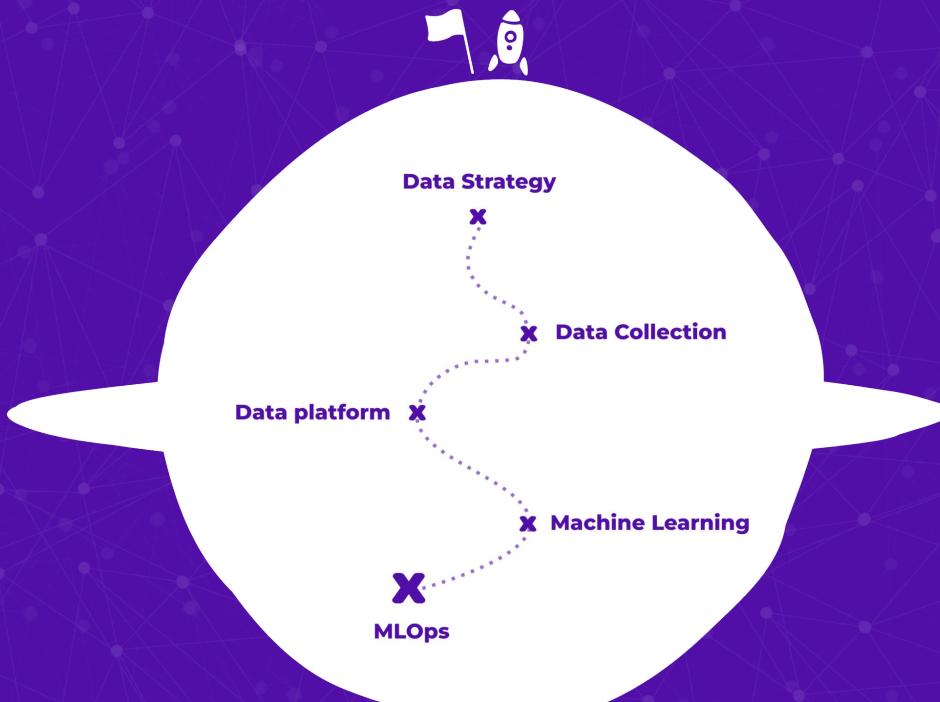


Cloud Native Data Platforms

The **why, what** and **how**

What we do



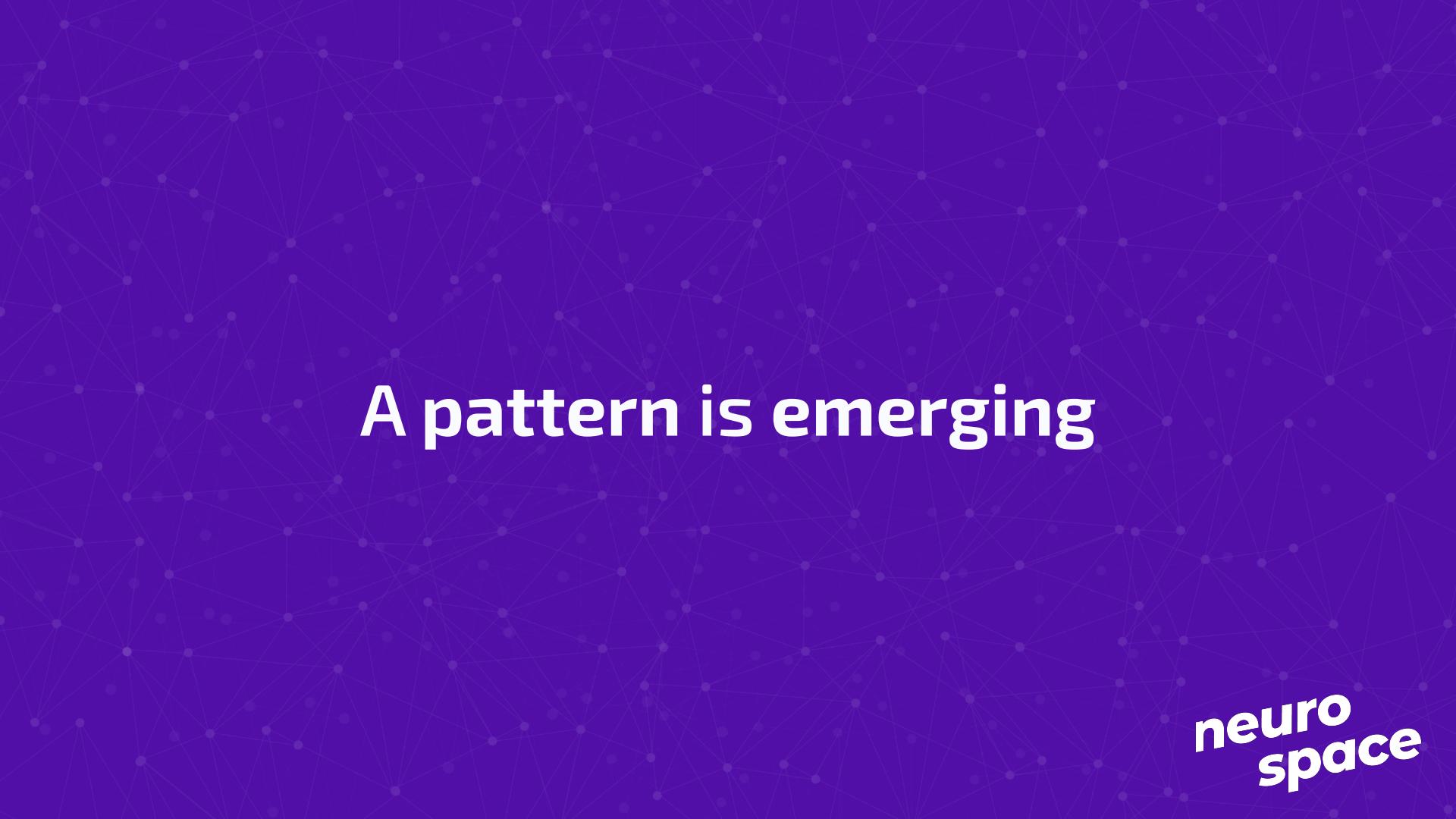
**neuro
space**

A dark purple background featuring a subtle, light-colored network graph composed of numerous small dots connected by thin lines.

**We help bigger Danish companies with
Data and Machine Learning**

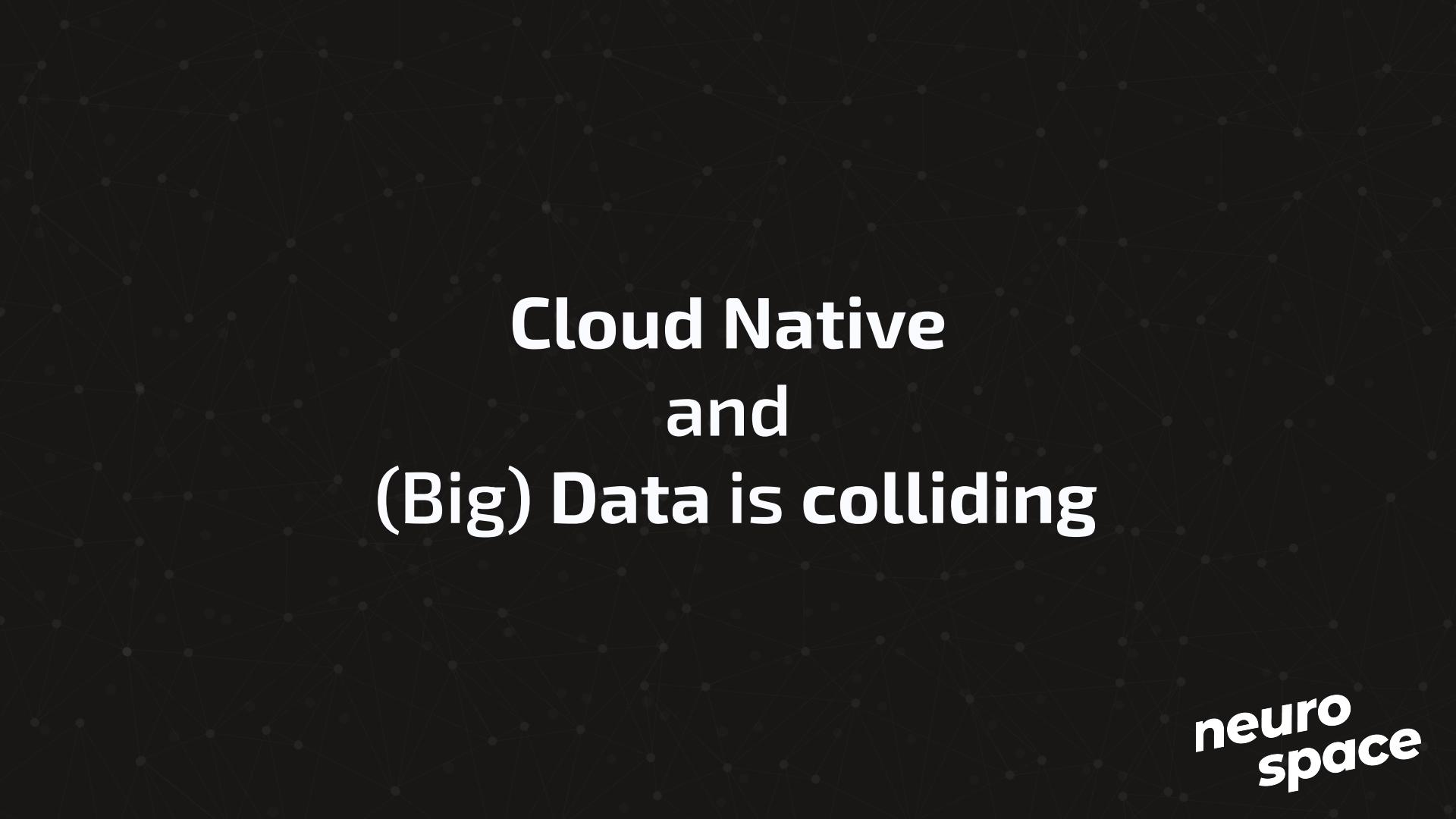
**neuro
space**





A pattern is emerging

**neuro
space**



Cloud Native and (Big) Data is colliding

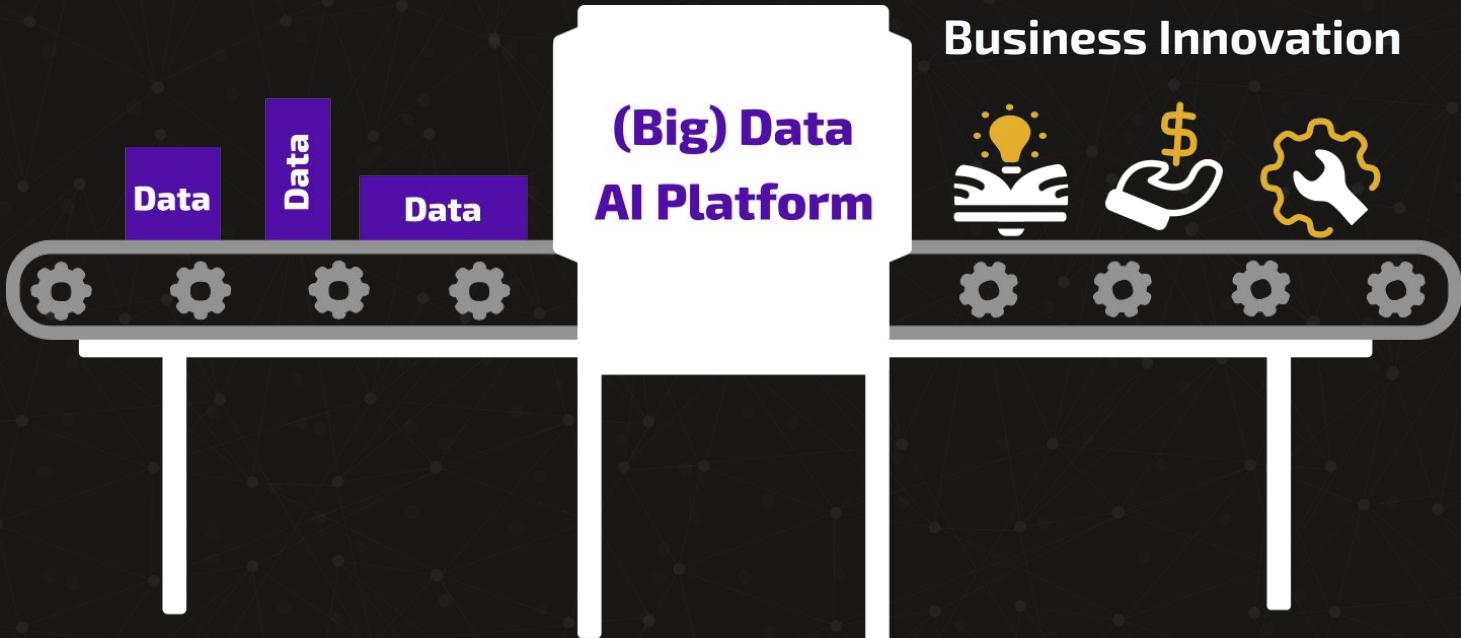
**neuro
space**

What the Business Wants

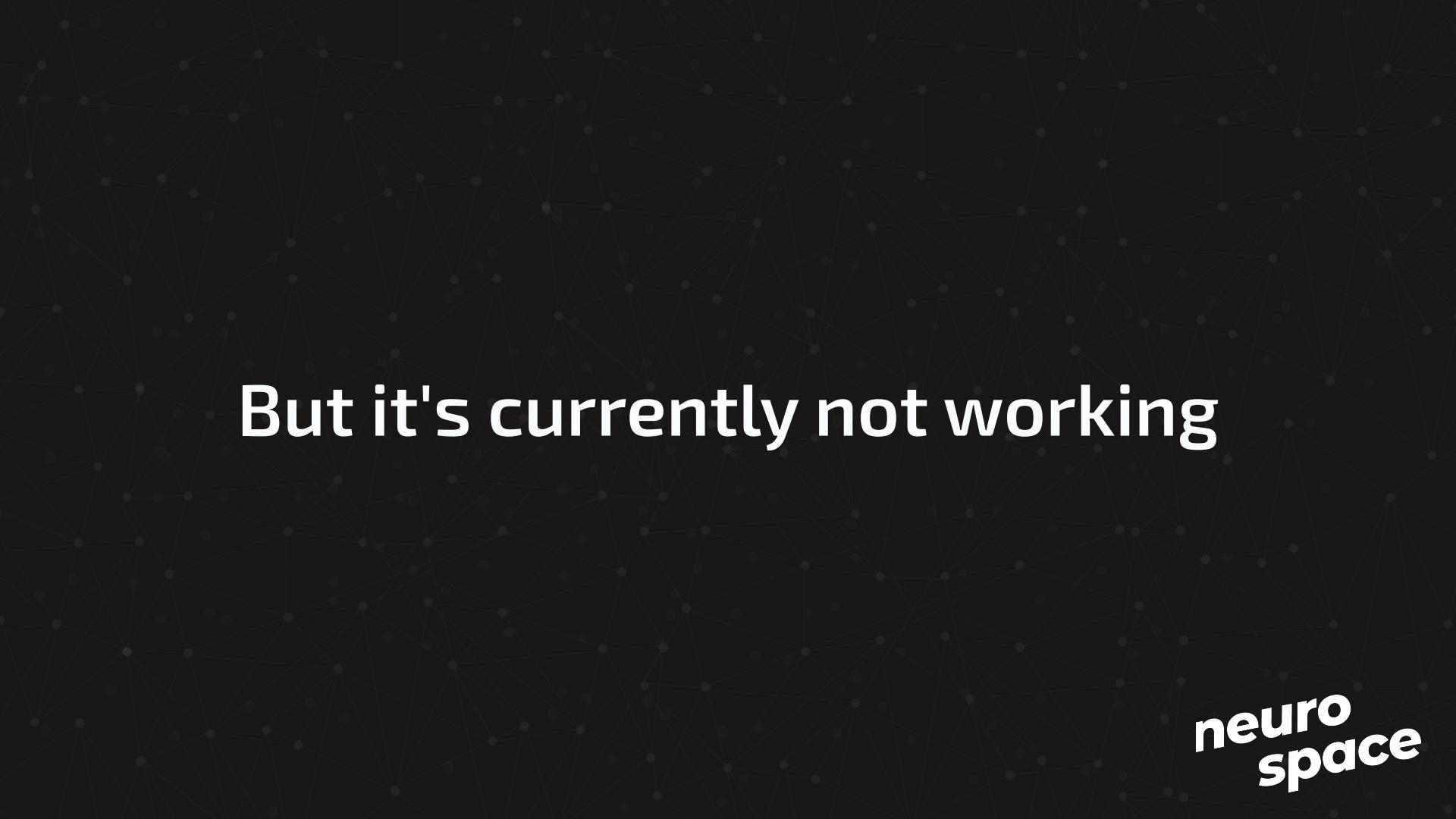
**neuro
space**

aka. the CFO and CEO

neuro
space



neuro
space



But it's currently not working

neuro
space



neuro
space



**Failure to materialize
Data-Driven Value**



**Failure to Scale Data
Sources**



**Failure to Scale Data
Consumers**

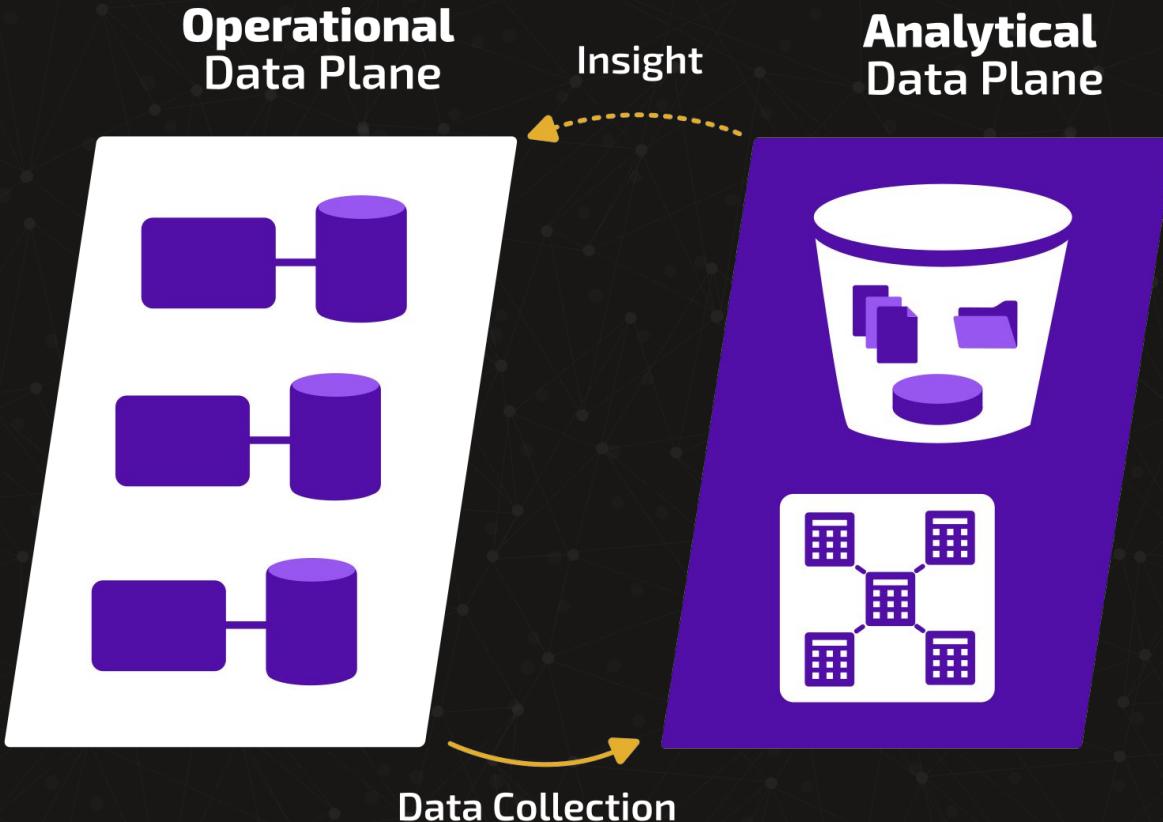


Failure to Bootstrap

Scaling a Data Platform is hard work!

neuro
space

We need to understand the general idea of a data platform and look at history



**neuro
space**

Analytical Data Platforms

Data Warehouse



1980

Data Mesh



2010

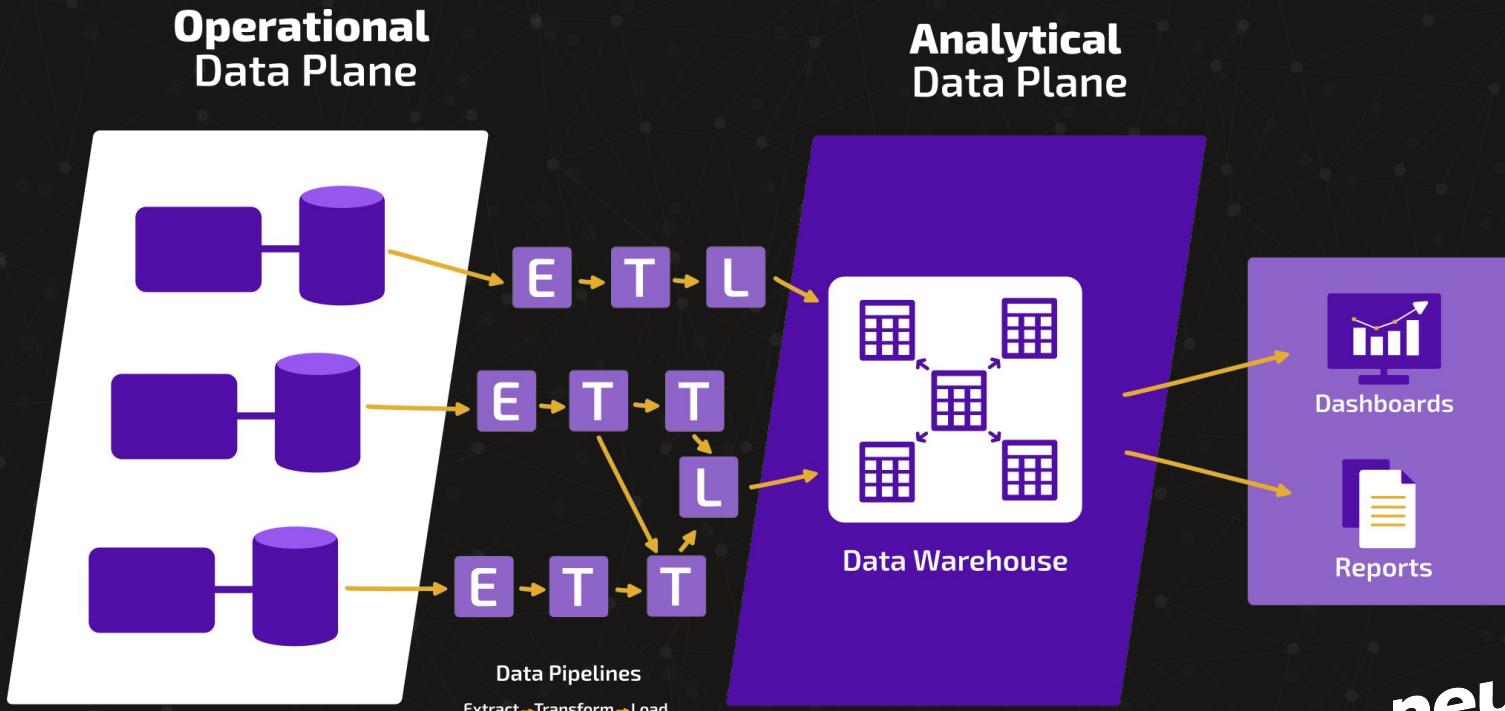
2020

Data Lake

A wide-angle photograph of a large industrial warehouse. The space is filled with tall, multi-tiered metal shelving units. The shelves are stocked with various products, including numerous white plastic barrels with red lids, cardboard boxes, and blue plastic tubs. The lighting is bright, coming from overhead fixtures, which creates a high-contrast scene with deep shadows in the recesses between the shelving units.

Data Warehouse

Data Warehouse

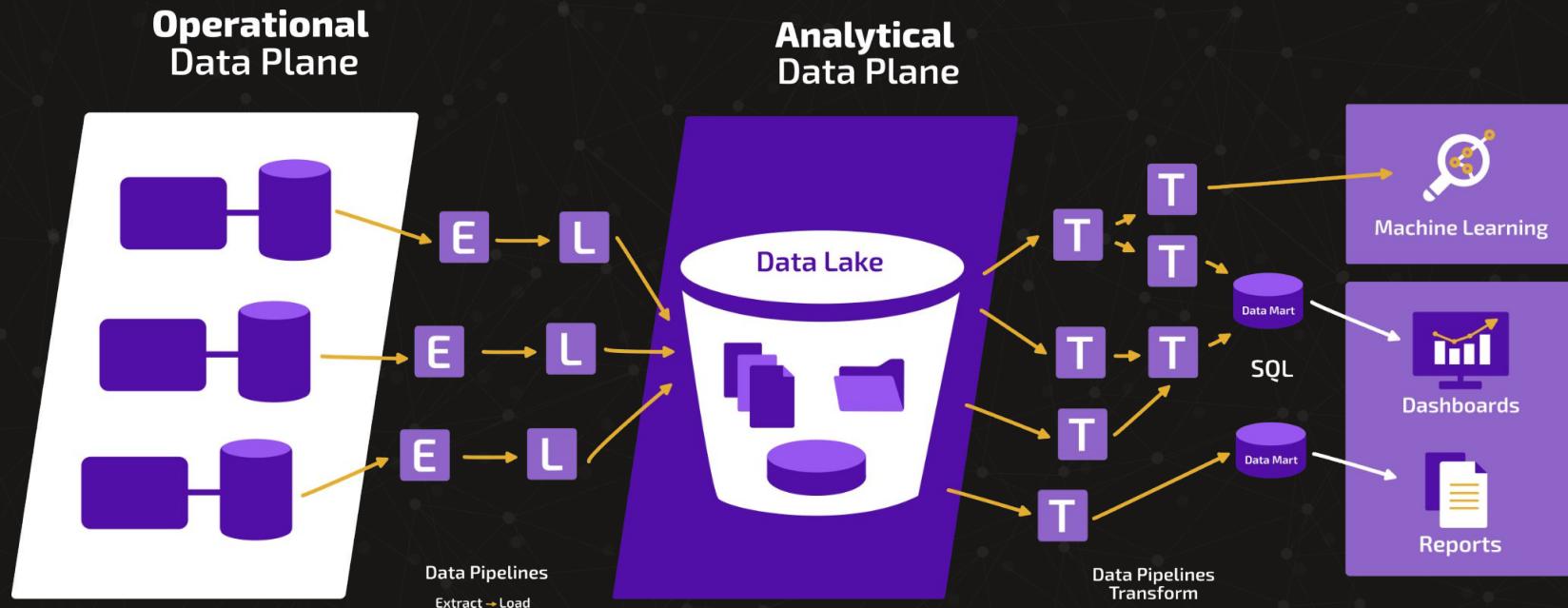


neuro
space

A wide-angle photograph of a serene landscape during sunset or sunrise. The sky is a gradient of deep blue at the top, transitioning into warm orange and pink hues near the horizon. Silhouettes of mountain ranges are visible against the bright sky. In the foreground, there's a calm body of water reflecting the colors of the sky. A small, dark island with some vegetation is positioned in the middle ground. On the right side of the frame, a traditional fishing structure made of bamboo poles is partially submerged in the water.

Data Lake

Data Lake

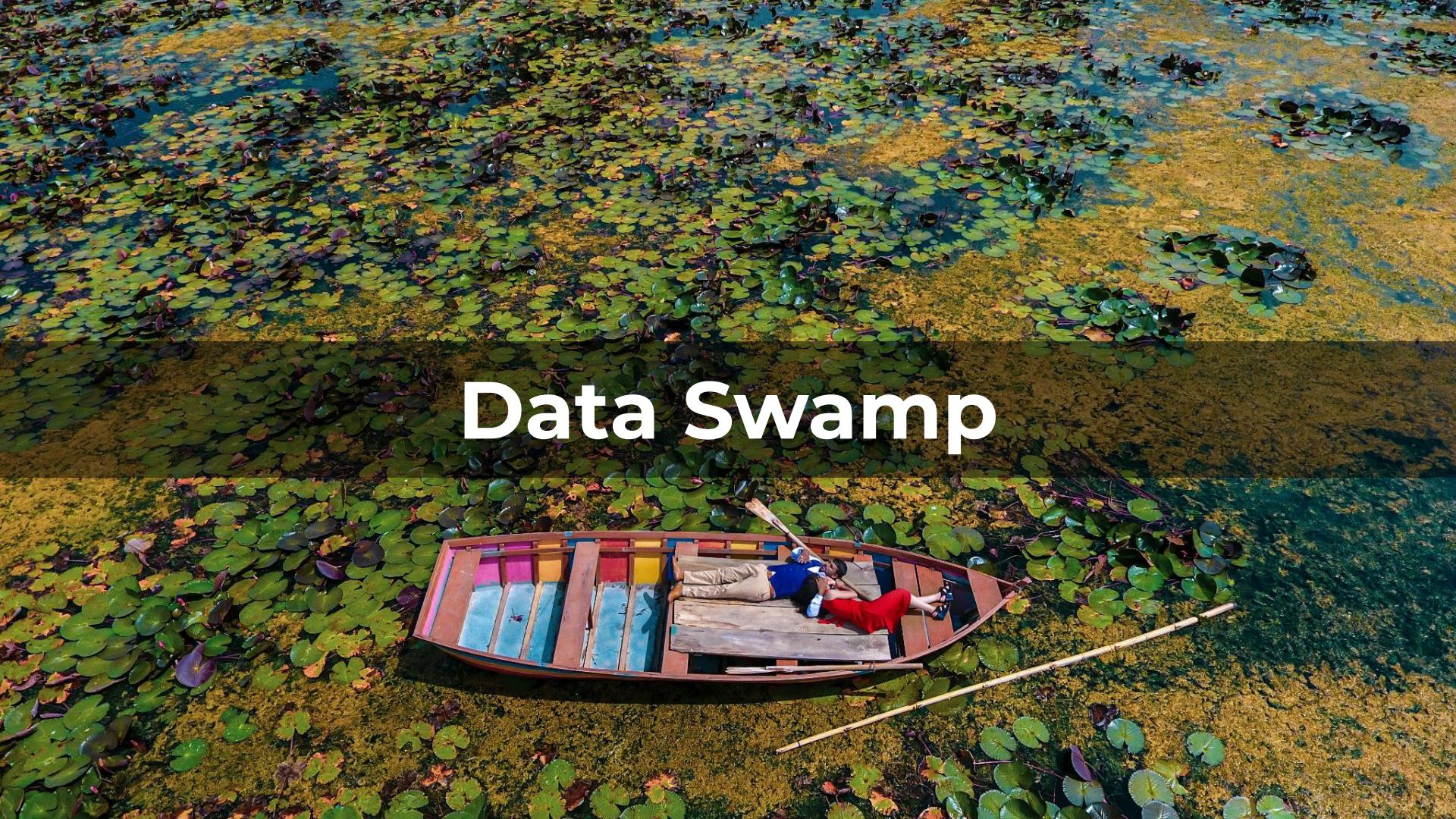


**neuro
space**



In reality though

neuro
space

A photograph of a man and a woman in a small wooden boat navigating through a dense field of green lily pads on a pond. The water is a mix of clear blue and greenish algae. The boat has colorful slats on its side. The man is at the stern, rowing, while the woman sits at the bow, looking towards the camera.

Data Swamp

An aerial photograph of a man and a woman in a small wooden boat on a pond. The boat is oriented diagonally, moving from the bottom left towards the top right. The man, wearing a blue shirt and khaki pants, is seated at the stern, holding a long wooden oar. The woman, wearing a red dress, is seated at the bow. They are both looking towards each other. The pond is covered with numerous green lily pads. A white arrow points from the text "Data Scientists comforting each other" to the man's hand holding the oar.

Data Scientists comforting each other

O'REILLY®

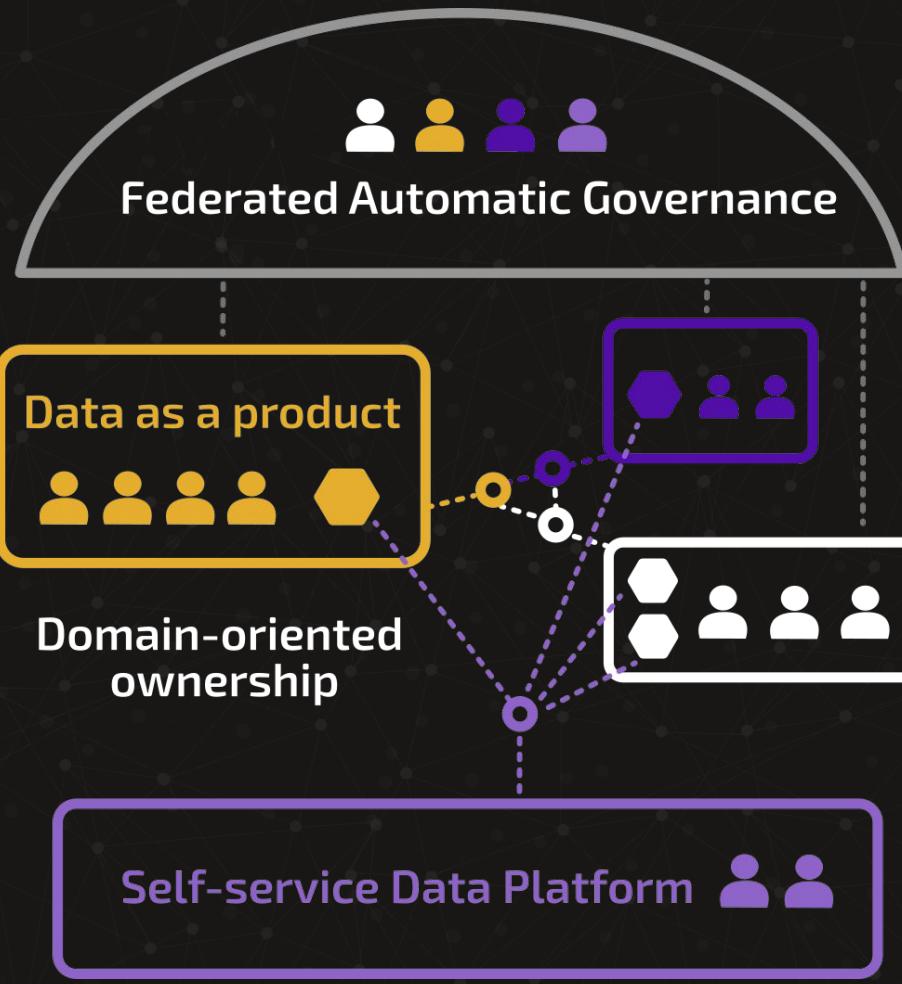
Data Mesh

Delivering Data-Driven Value at Scale



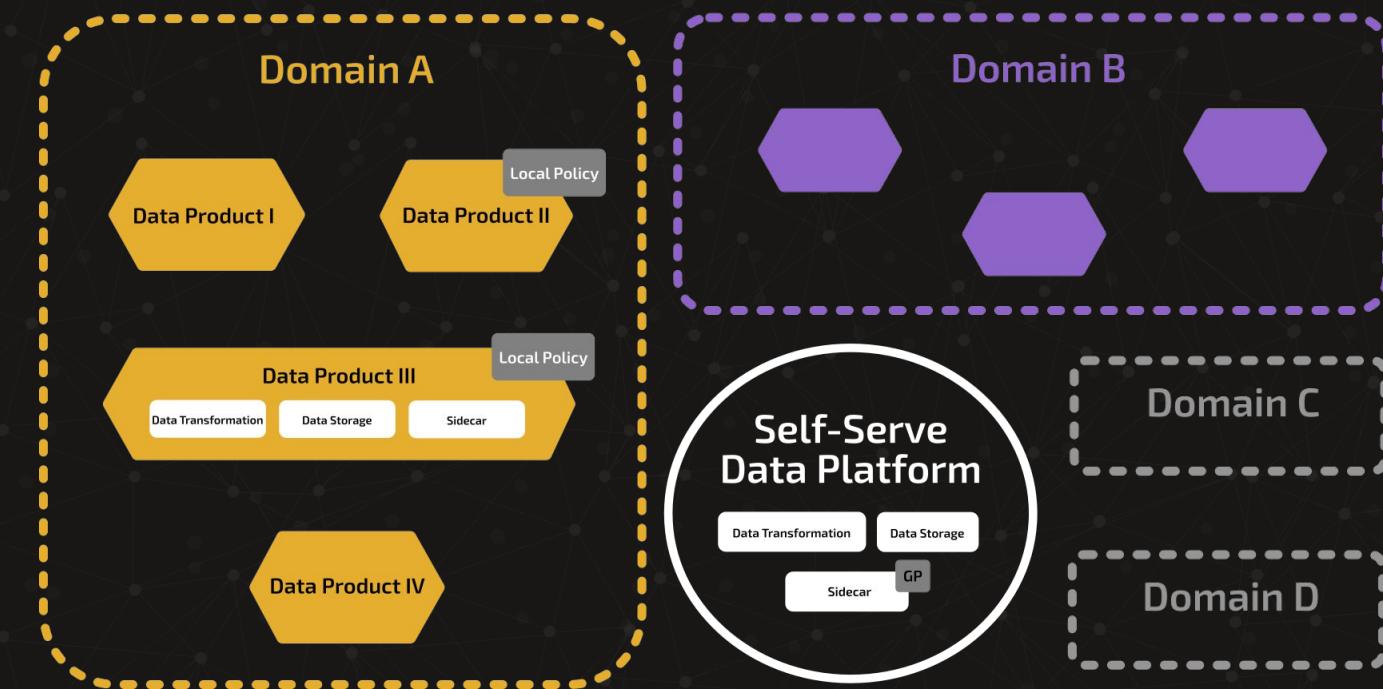
Zhamak Dehghani

neuro
space



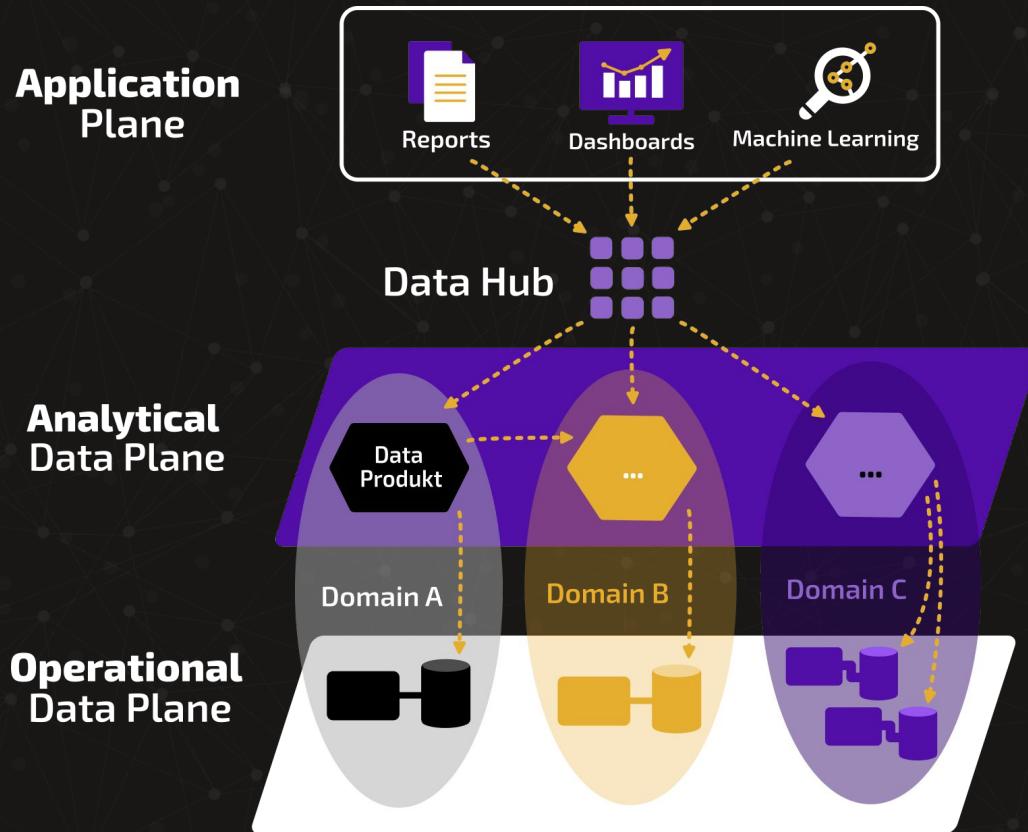
Federated Governance

Global Policy

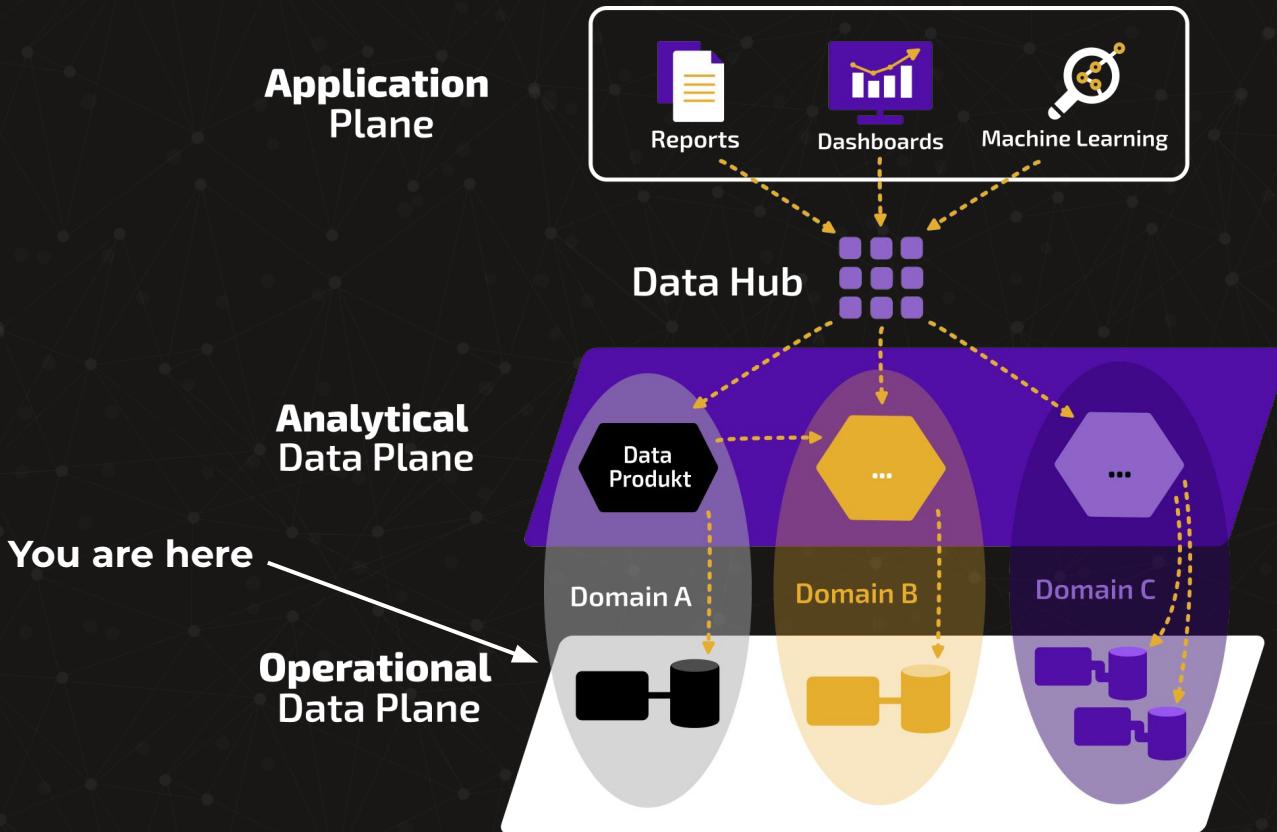


**neuro
space**

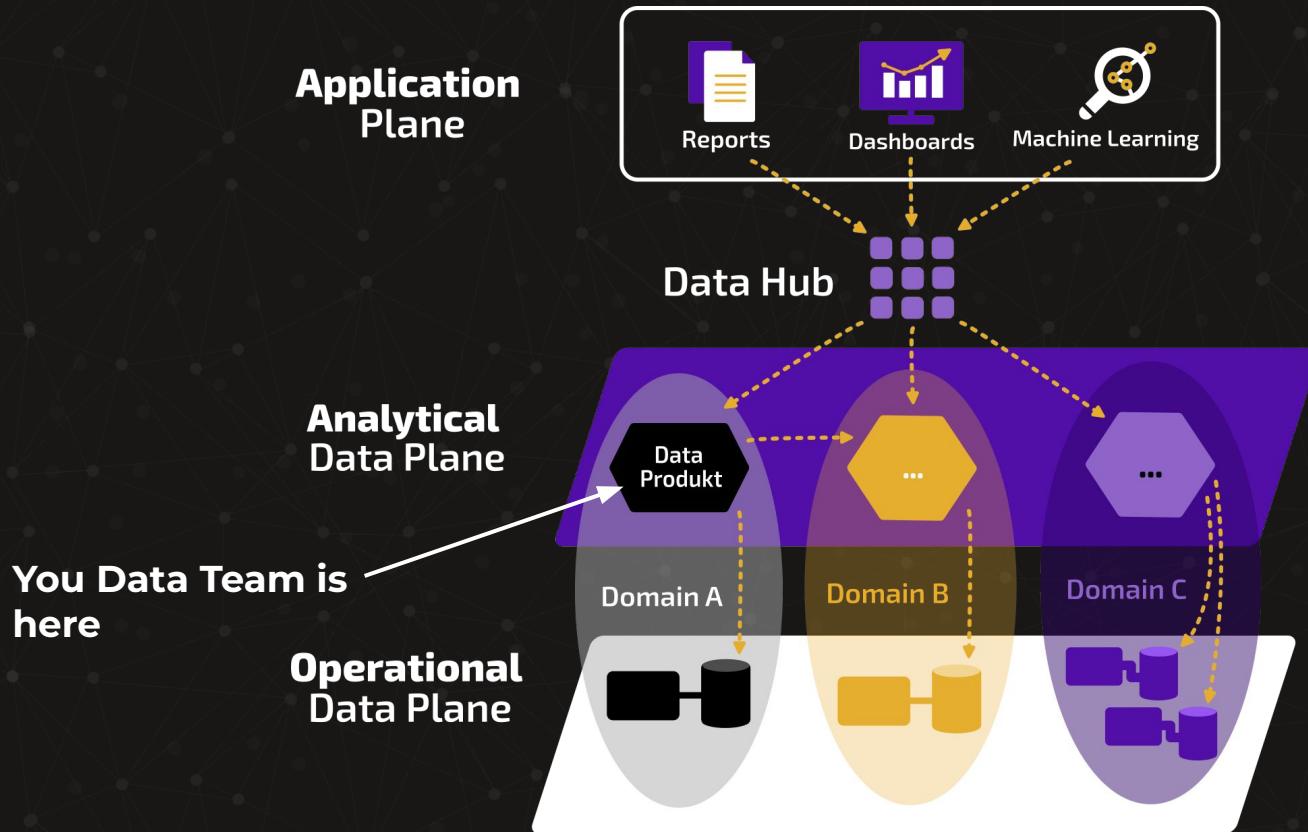
Data Mesh



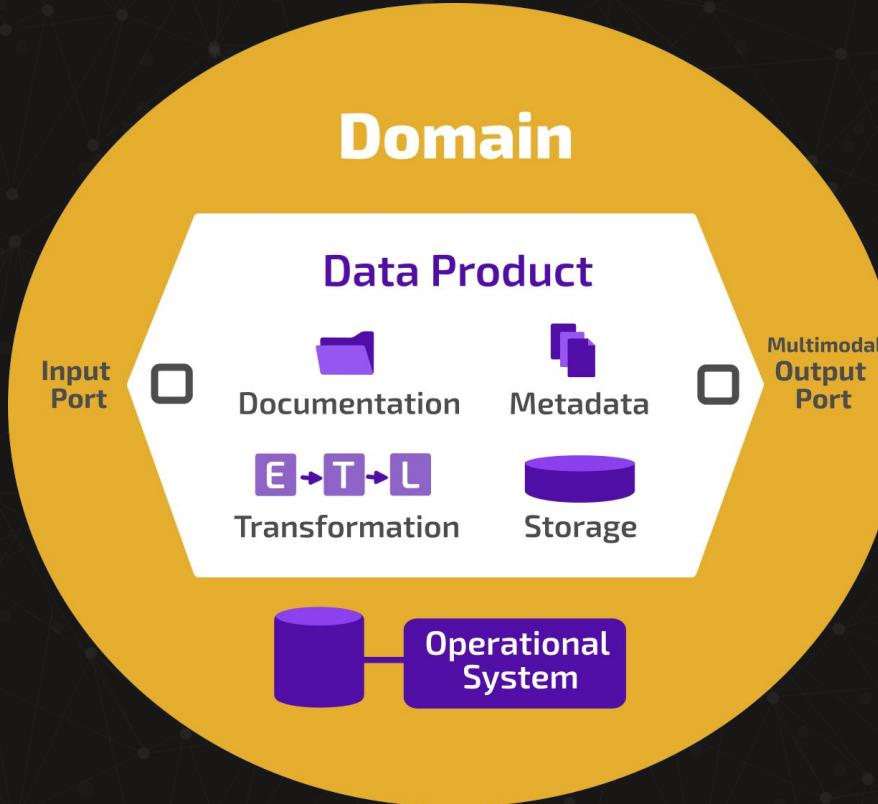
Data Mesh



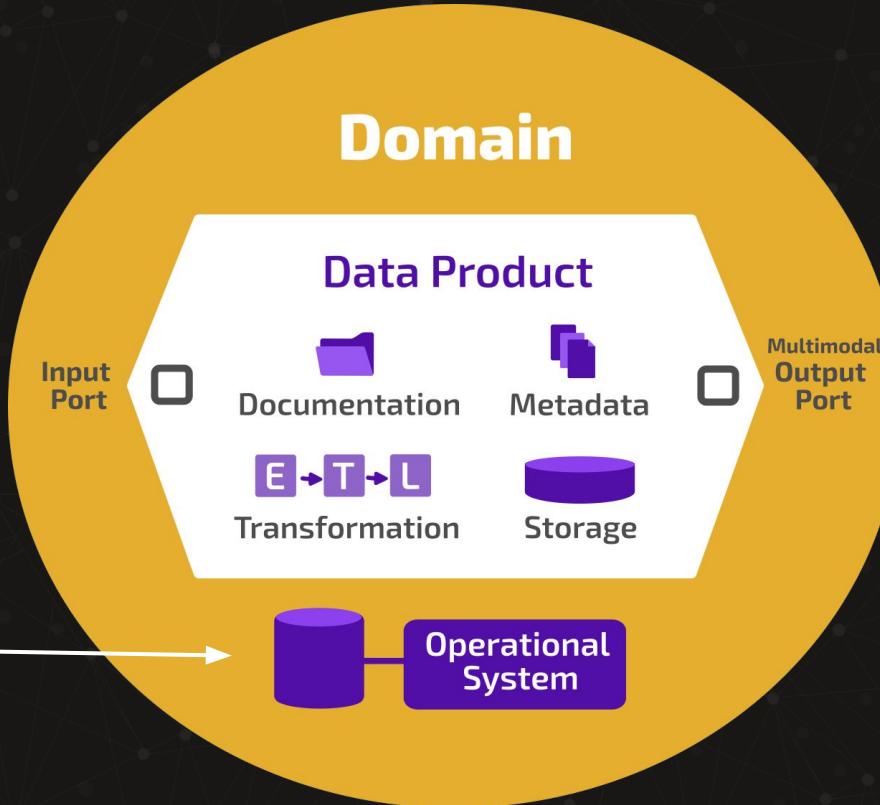
Data Mesh



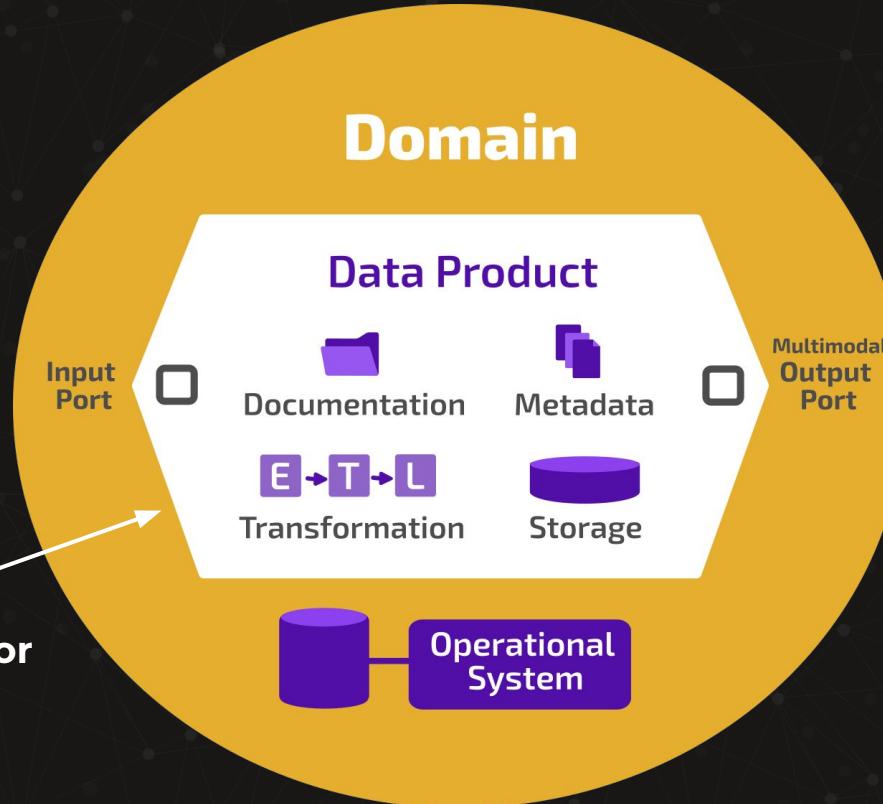
Data Product



Data Product



Data Product



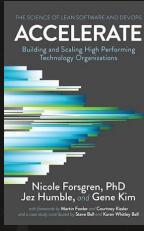
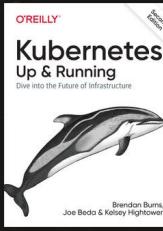
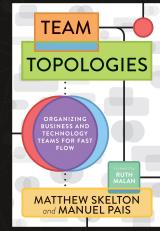
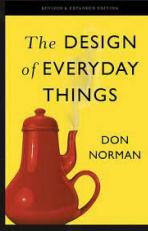
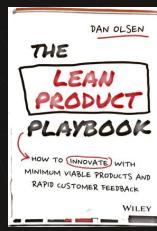
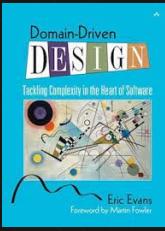
Four Core Data Mesh Principles

Domain oriented
Decentralization

Data as a
Product

Self-serve
data infra as
a Platform

Federated
Governance



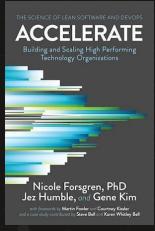
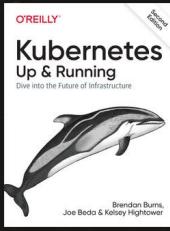
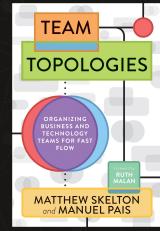
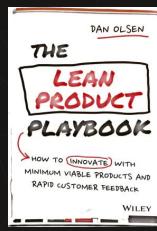
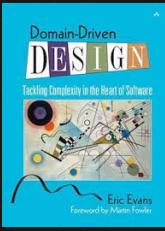
Domain oriented
Decentralization

Data as a
Product

Self-serve
data infra as
a Platform

Federated
Governance

neuro
space



Domain oriented
Decentralization

Data as a
Product

Self-serve
data infra as
a Platform

Federated
Governance

We already apply these in the Cloud Native Community

neuro
space

Software Engineering Principles

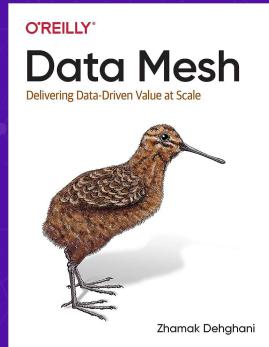
Decoupled
Micro
Services

Product
Thinking

Platform
Engineering

Community
driven
Standards
and Policies

Martin Fowler was the mentor for Zhamak Dehghani

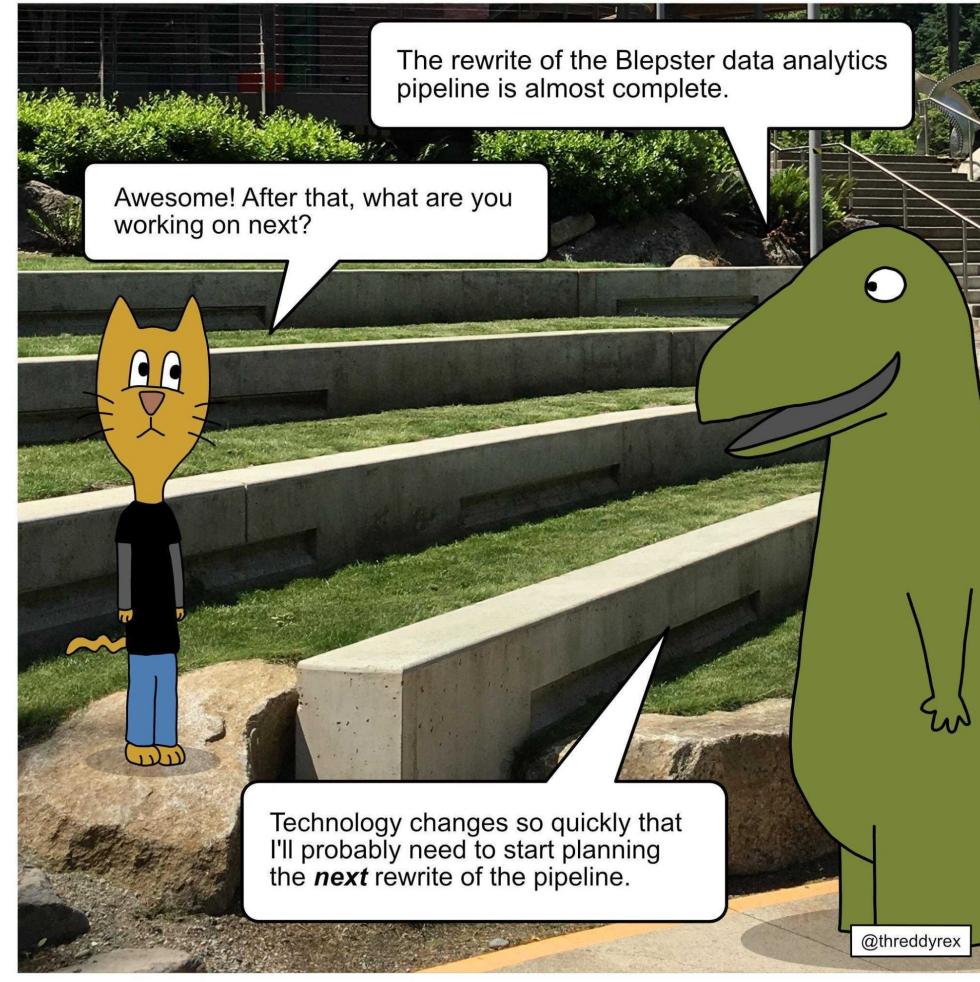


**neuro
space**

The Modern Data Stack

**neuro
space**

Distributed systems, stateless vs.
stateful, containerization, infra as
code, monitoring, APIs / Data
Contracts, etc.



neuro
space

Orchestration

neuro
space

A group of men are standing in a field under a cloudy sky. In the foreground, a man wearing a dark cap and a green jacket looks directly at the camera with a neutral expression. Behind him, several other men are visible, some looking towards the camera and others looking away. The background shows green trees and a fence.

DAGS. D'YA LIKE DAGS?



Apache
Airflow



DAGS. D'YA LIKE DAGS?



Apache
Airflow

=



Jenkins

DAGS. D'YA LIKE DAGS?

Imperative vs. Declarative

neuro
space



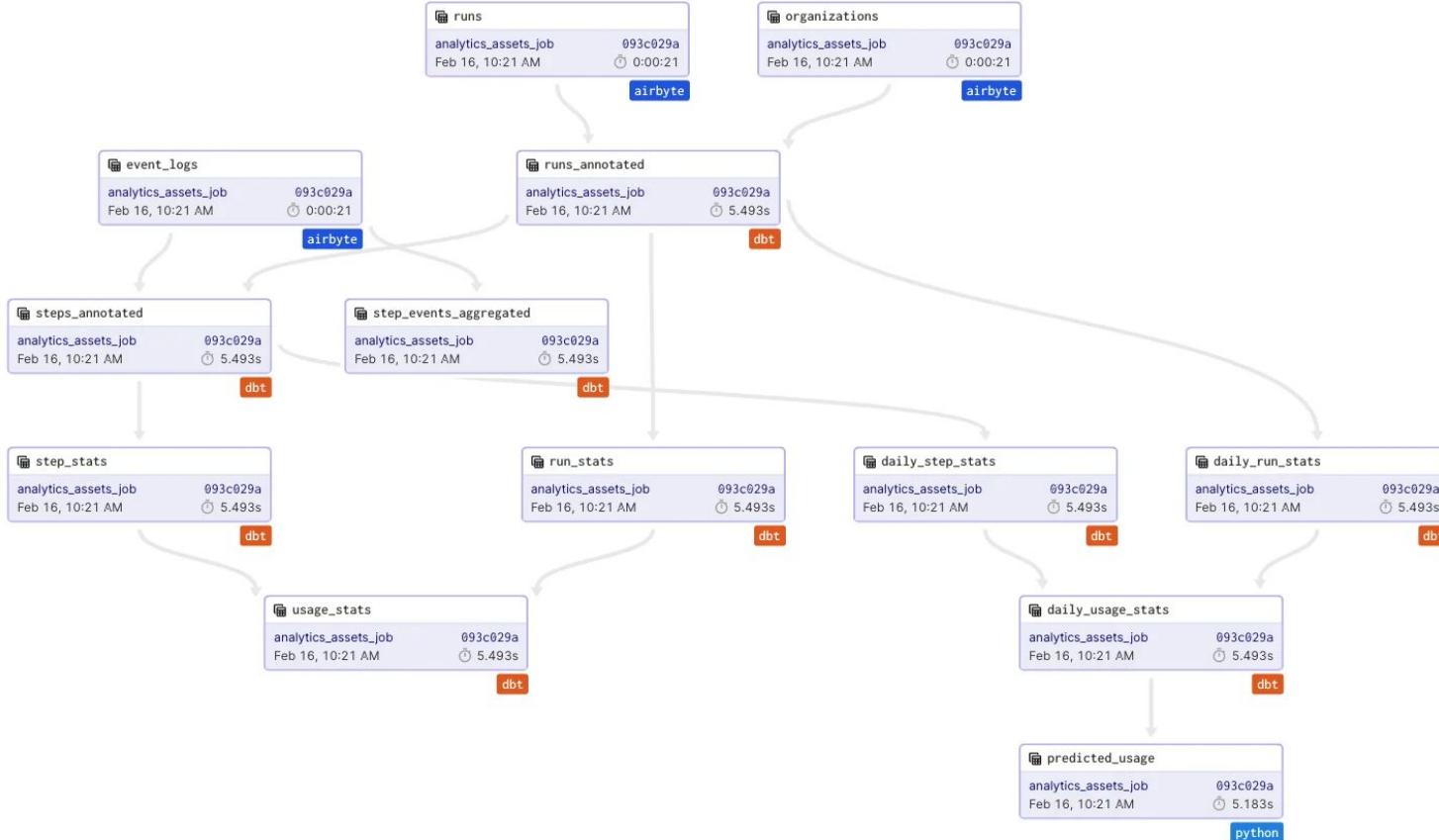
dagster

Disclaimer: We are contributors in Neurospace

**neuro
space**

Software Defined Data Assets

neuro
space



<https://dagster.io/blog/rebundling-the-data-platform>

neuro
space

○ ○ ○

```
DBT_PROJECT_DIR = 'path/to/dbt/project'

from dagster import AssetGroup, asset
from dagster_dbt import dbt_cli_resource, load_assets_from_dbt_project
from dagster_airbyte import airbyte_resource, build_airbyte_assets

airbyte_assets = build_airbyte_assets(
    connection_id="123456-abcd-ef34",
    destination_tables=["event_logs", "runs", "organizations"],
)

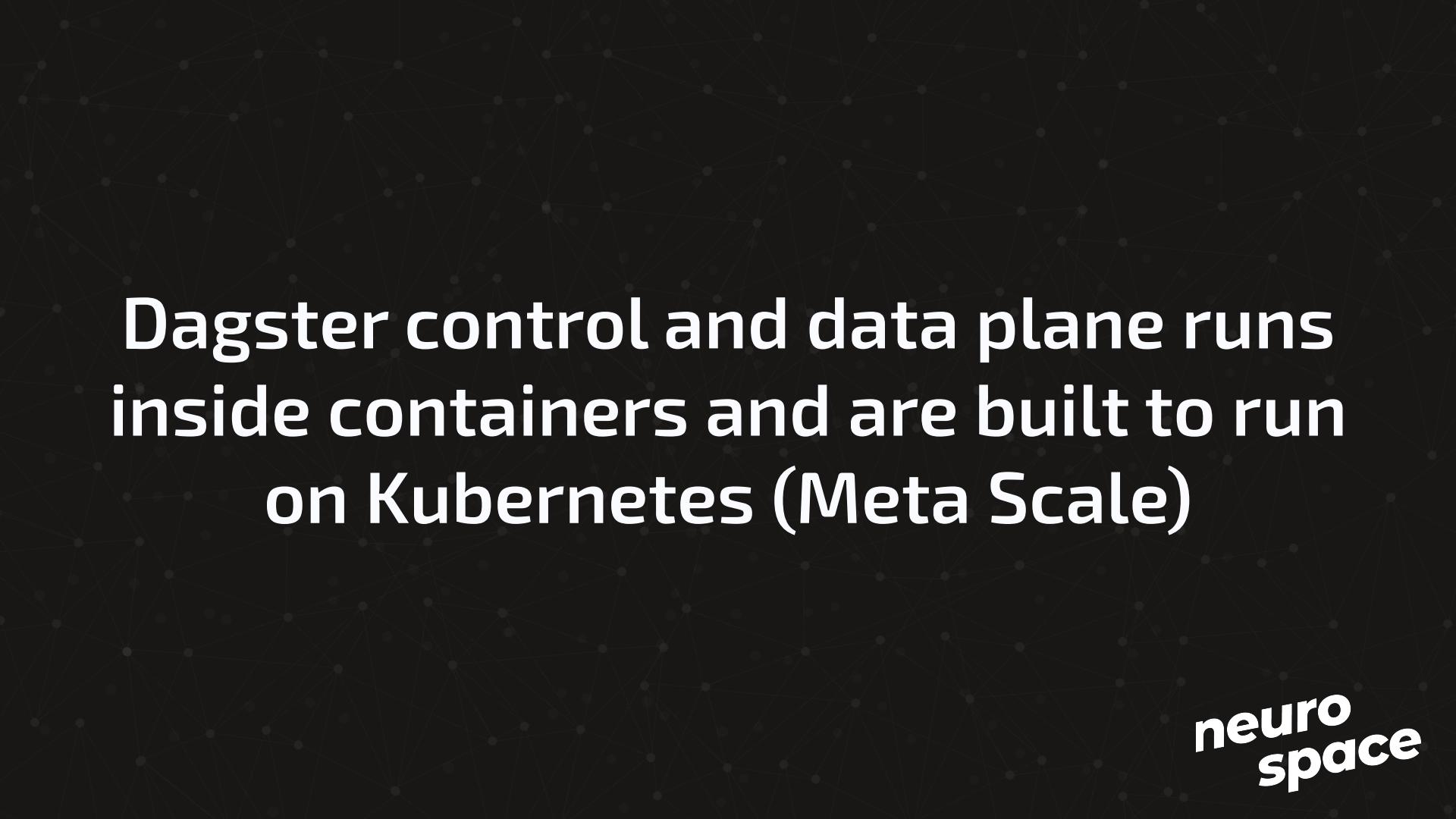
dbt_assets = load_assets_from_dbt_project(project_dir=DBT_PROJECT_DIR)

@asset(compute_kind="python")
def predicted_usage(daily_usage_stats):
    # Code to do stuff to daily_usage_stats goes here.

analytics_assets = AssetGroup(
    airbyte_assets + dbt_assets + [predicted_usage],
    resource_defs={
        "airbyte": airbyte_resource.configured({"host": "localhost", "port": "8000"}),
        "dbt": dbt_cli_resource.configured({"project_dir": DBT_PROJECT_DIR}),
    },
)
```

<https://dagster.io/blog/rebundling-the-data-platform>

neuro
space



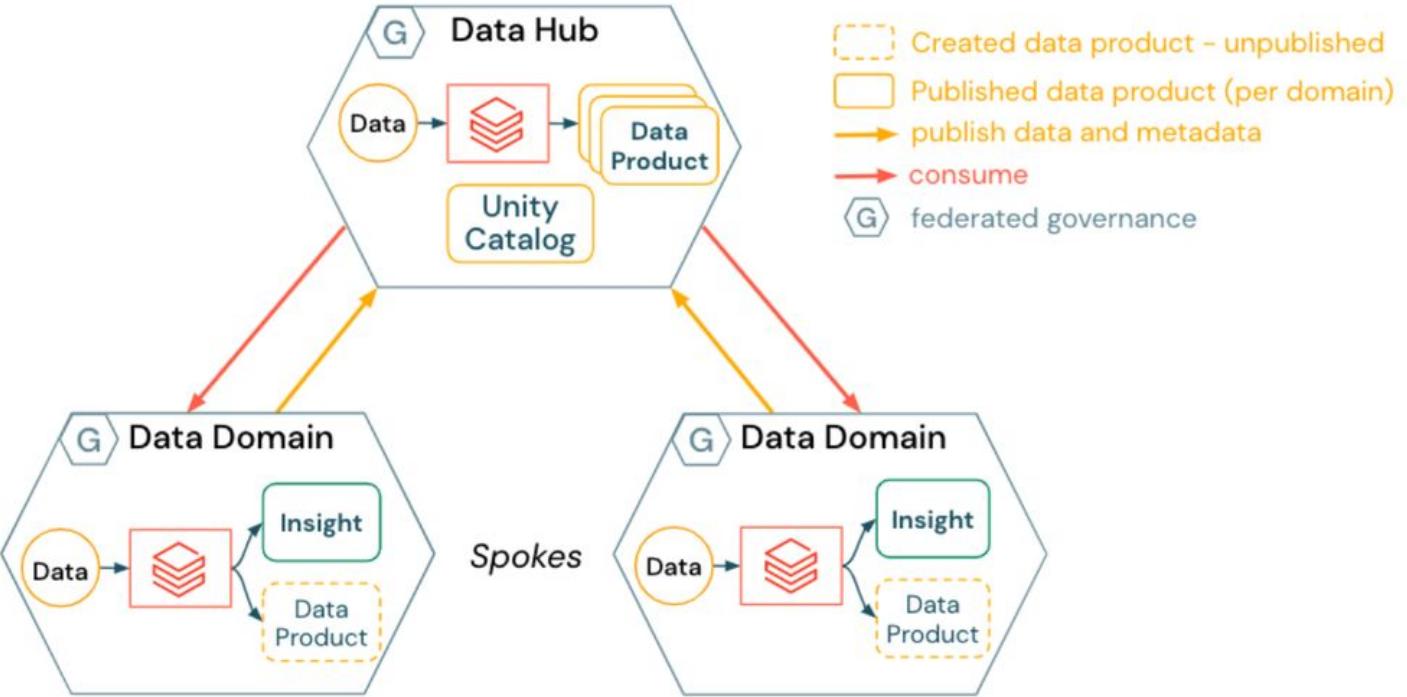
**Dagster control and data plane runs
inside containers and are built to run
on Kubernetes (Meta Scale)**

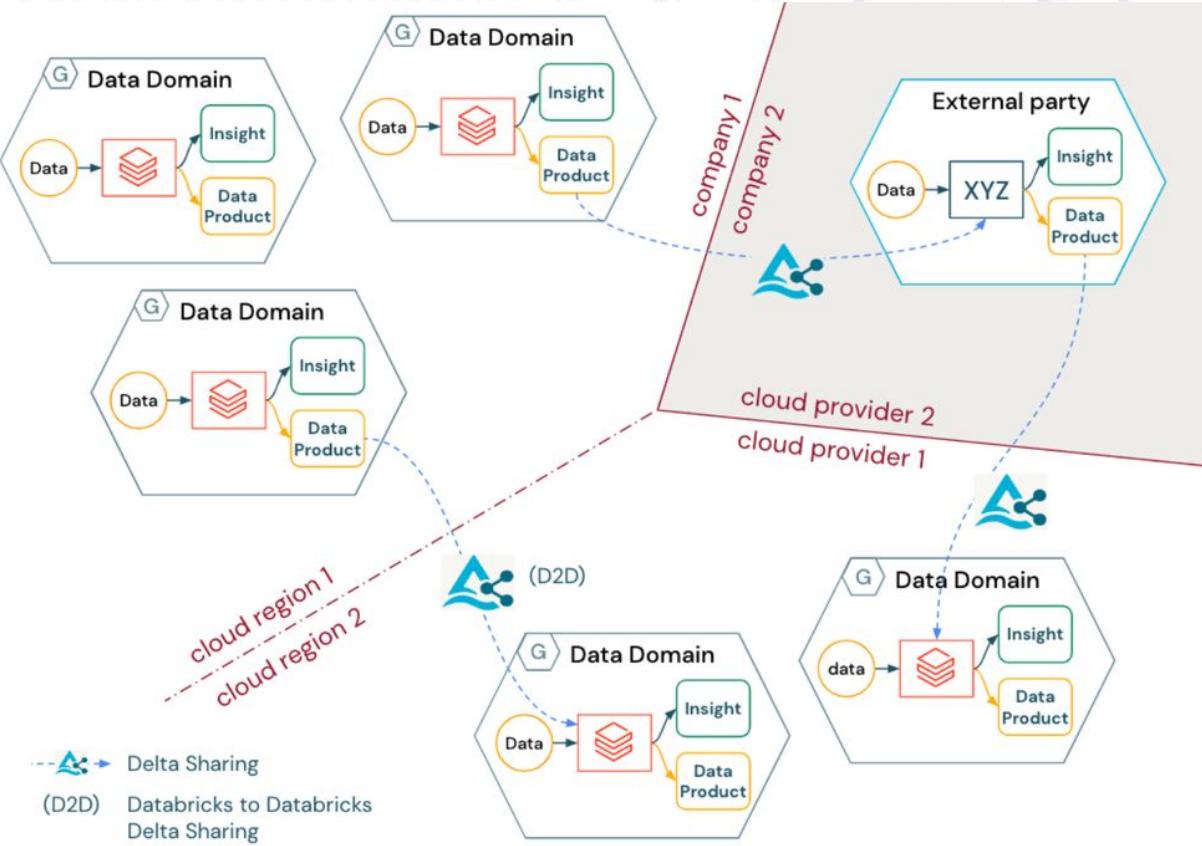


databricks

Disclaimer: We are a Databricks partner in Neurospace

**neuro
space**



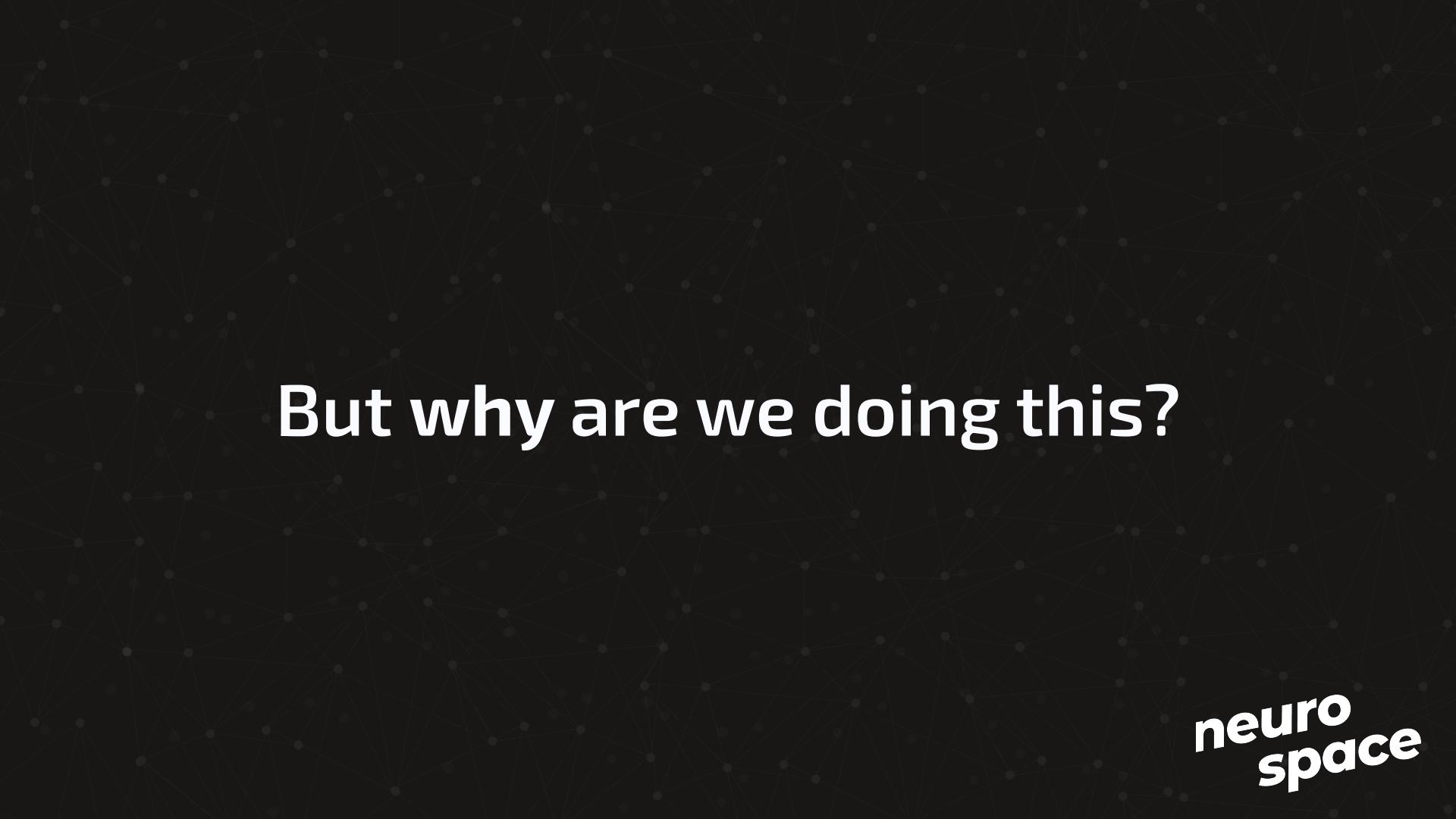


<https://www.databricks.com/blog/2022/10/19/building-data-mesh-based-databricks-lakehouse-part-2.html>



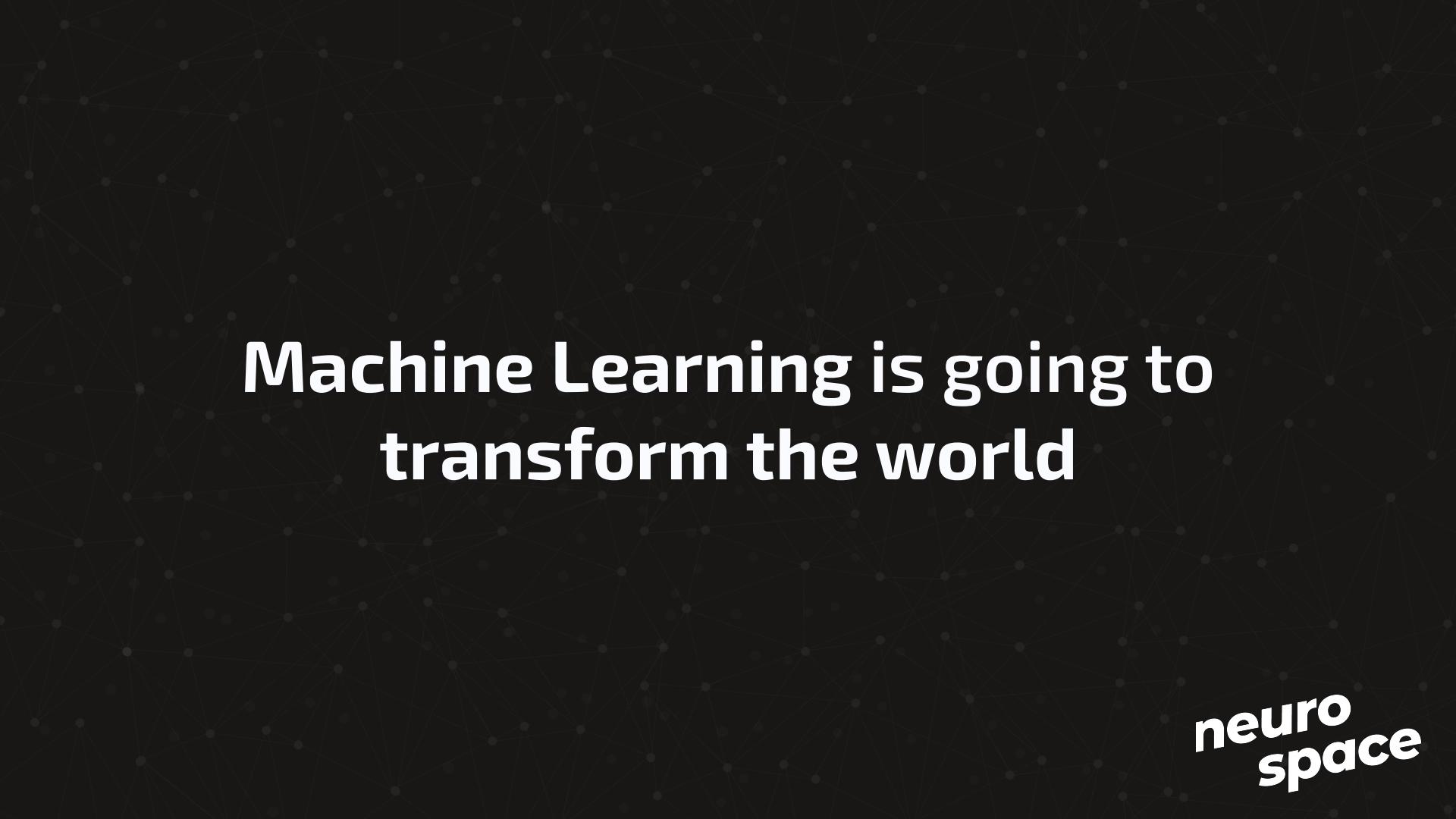
The (Big) Data world is going Cloud Native

**neuro
space**



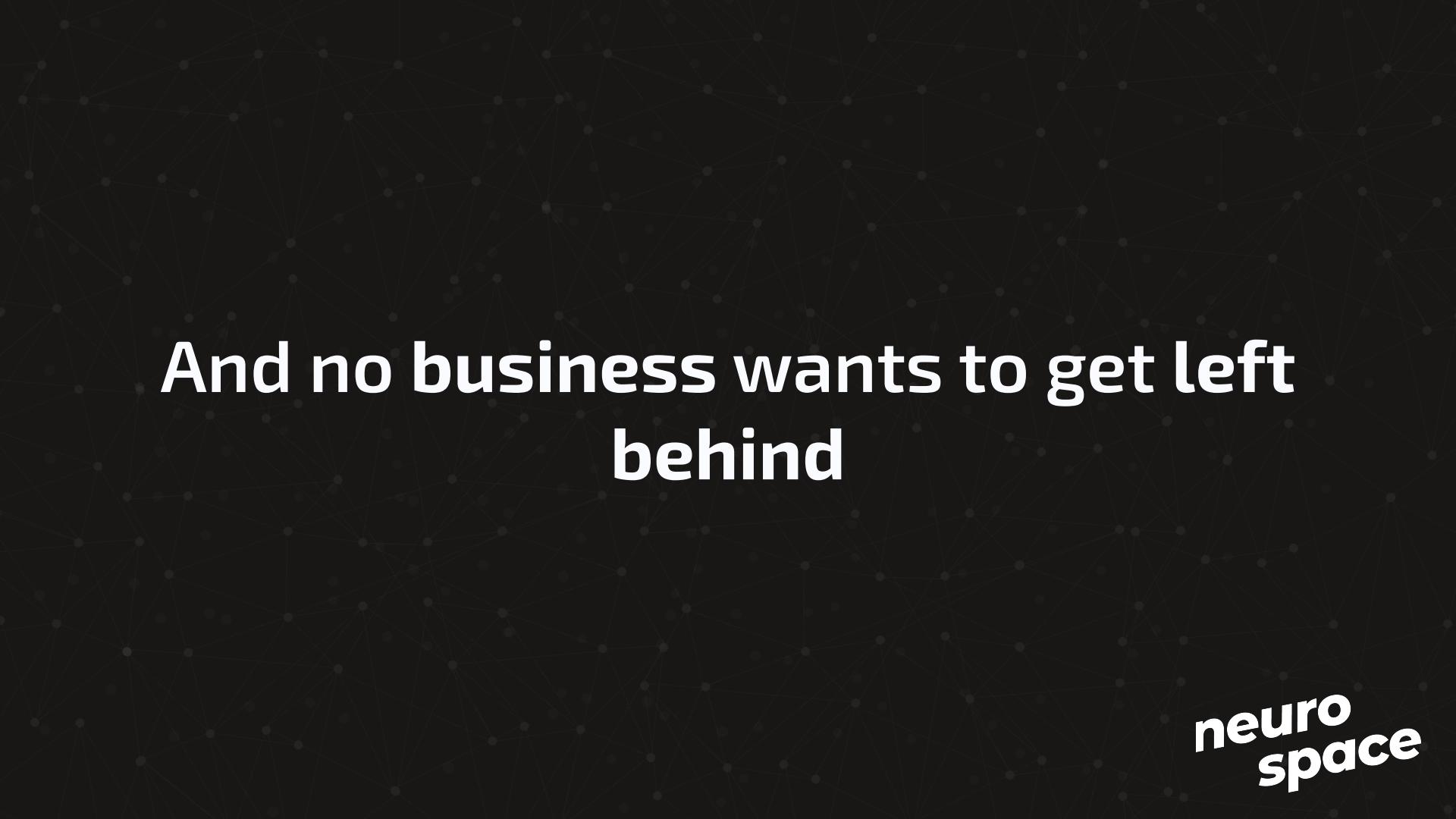
But why are we doing this?

**neuro
space**



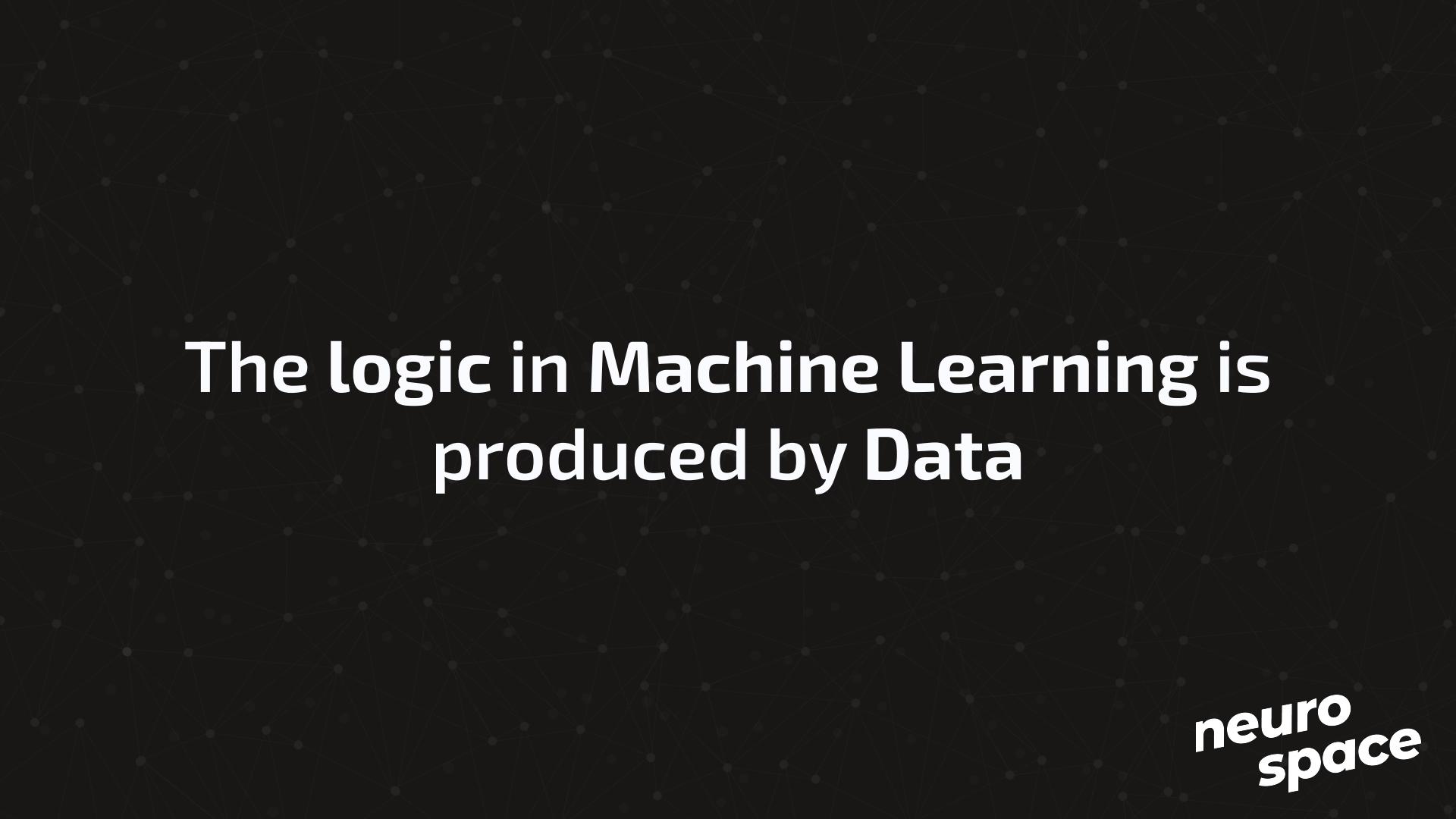
**Machine Learning is going to
transform the world**

**neuro
space**



**And no business wants to get left
behind**

**neuro
space**



The logic in Machine Learning is
produced by Data

neuro
space



ML Algorithm



Data

+

=

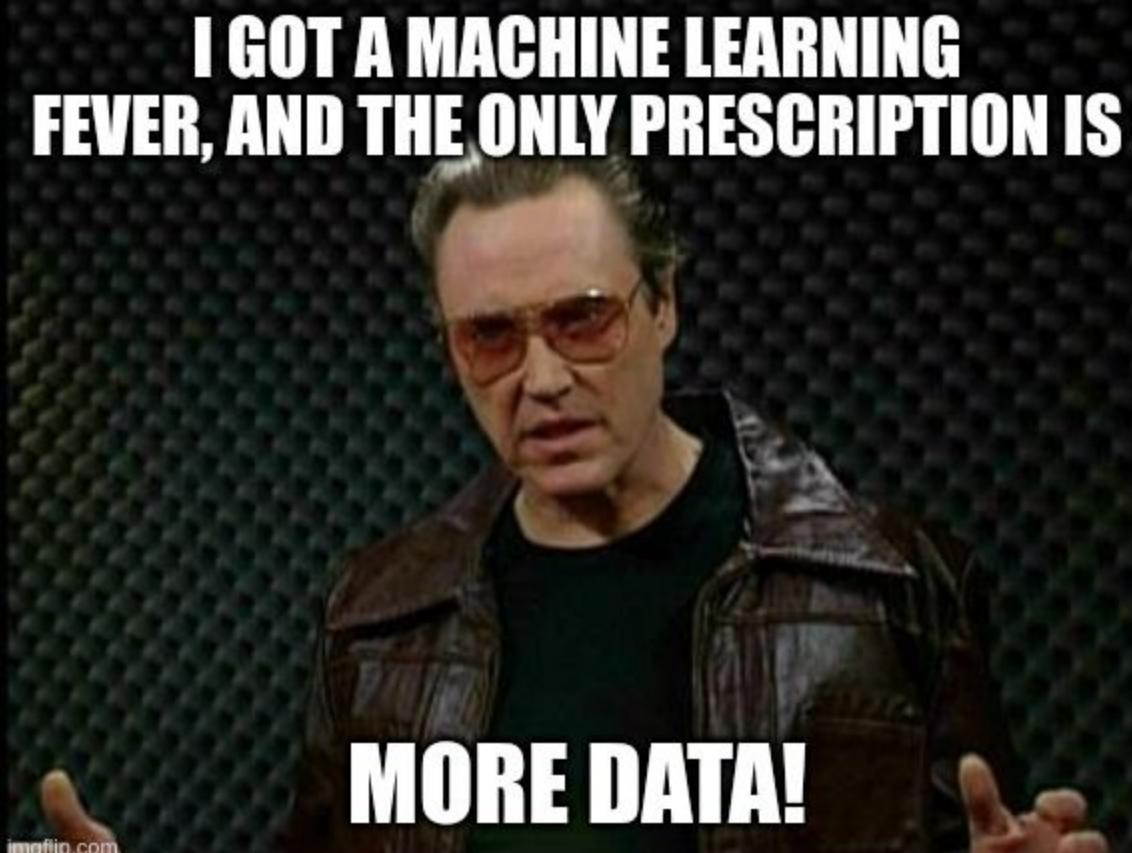


Model

neuro
space

**There have never been a higher
demand on good and timely data for
driving the businesses of tomorrow**

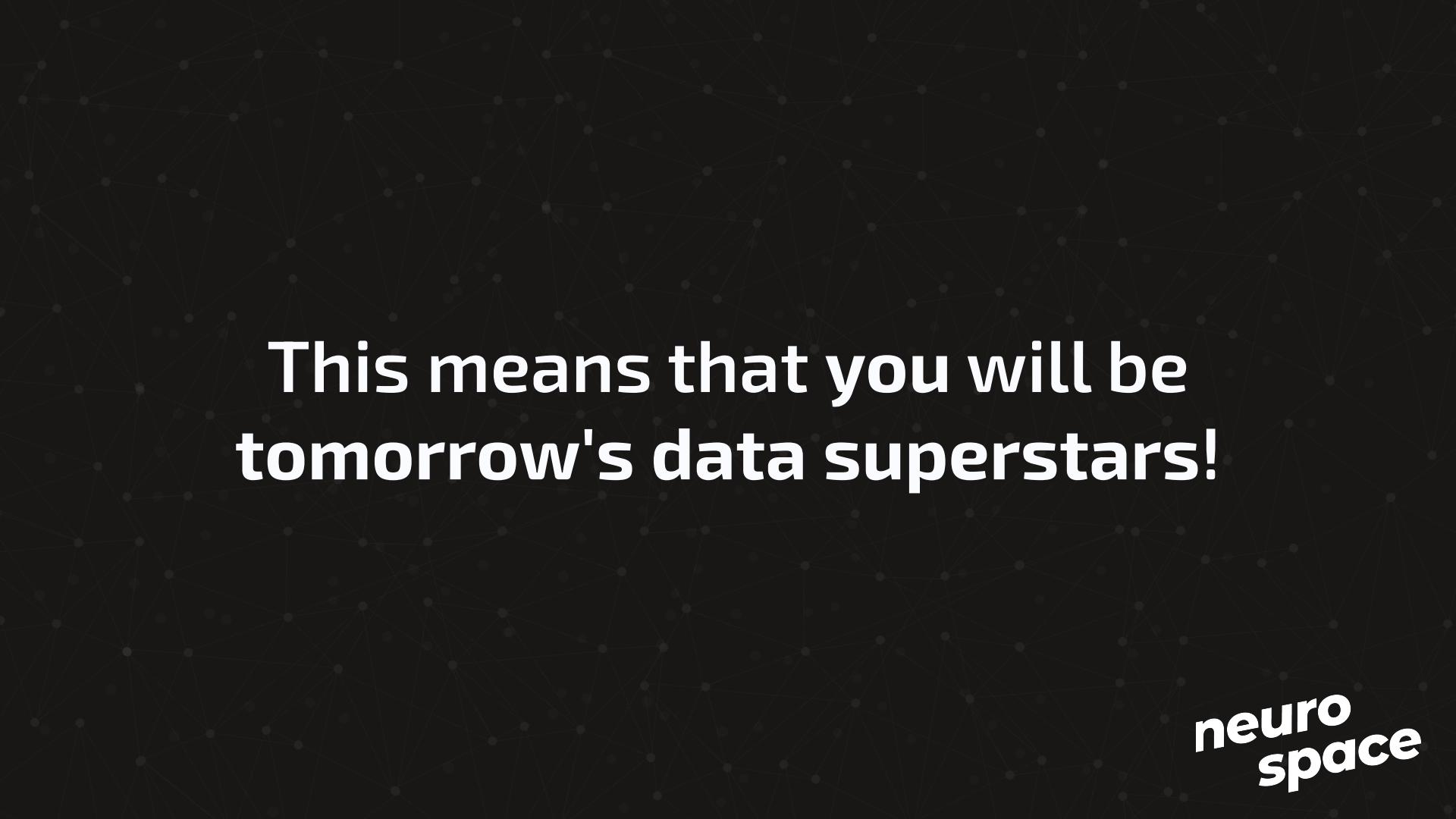
I GOT A MACHINE LEARNING
FEVER, AND THE ONLY PRESCRIPTION IS



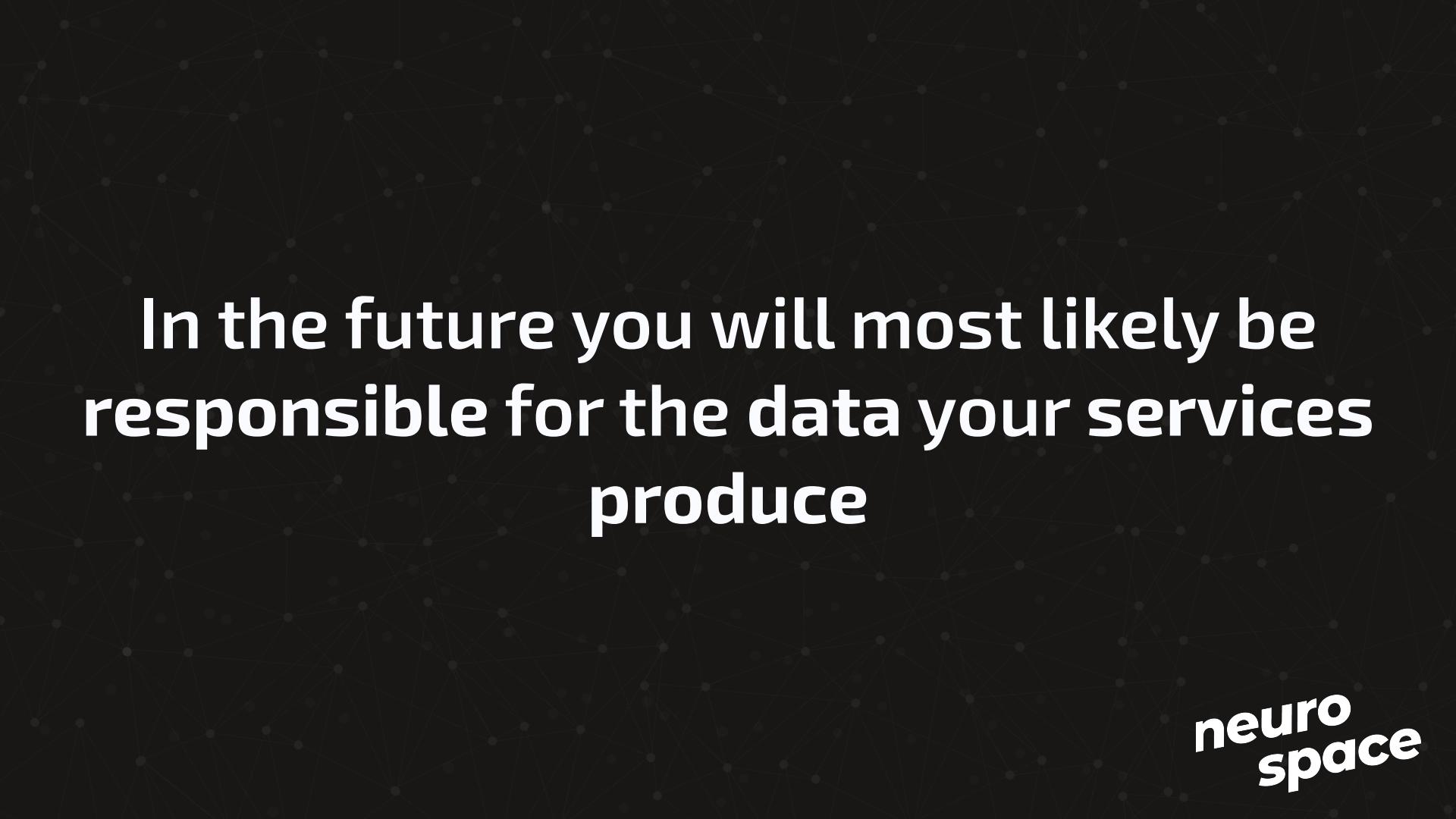
MORE DATA!

imgflip.com

neuro
space



**This means that you will be
tomorrow's data superstars!**

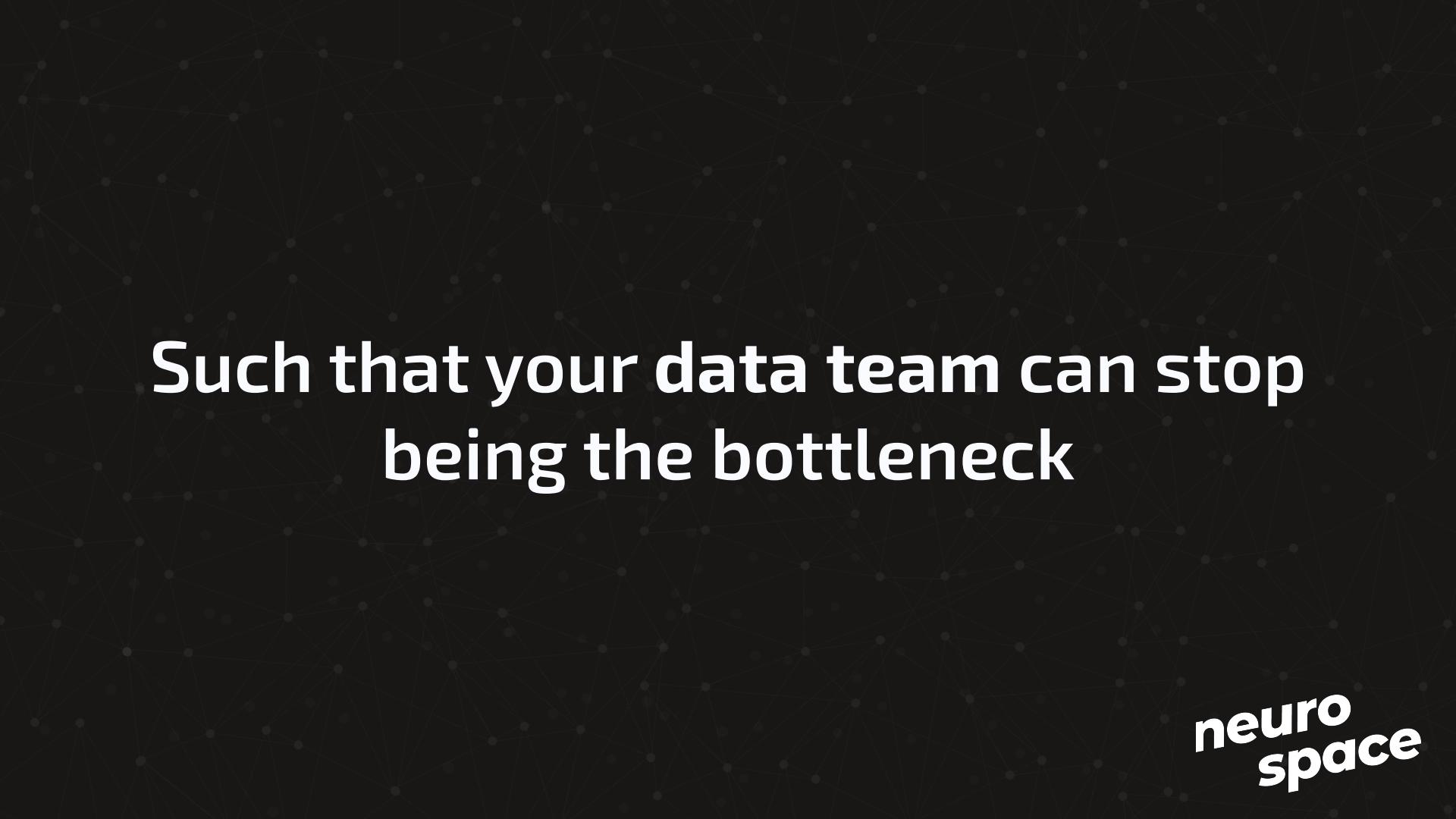


**In the future you will most likely be
responsible for the data your services
produce**

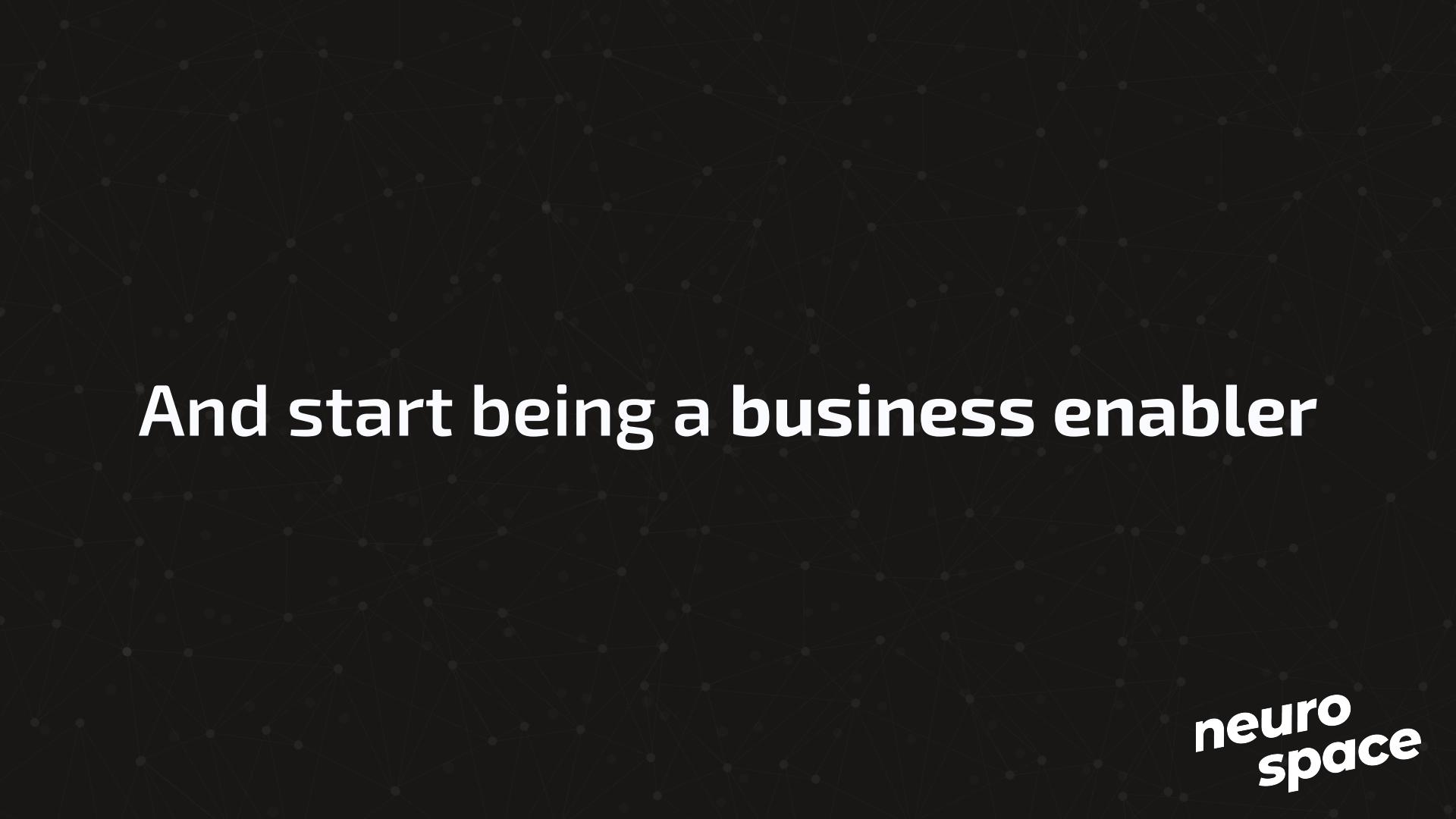


On a Data Product level

neuro
space



**Such that your data team can stop
being the bottleneck**



And start being a business enabler

**neuro
space**

Therefore, it is my hope that we can
talk more **Cloud Native Data** in this
community



Feedback

What are your questions?