

How Traton built a data lakehouse for autonomous driving

TRATON



Joachim Zetterman
Sr. Data & ML platform lead

TRATON

Sold vehicles in 2023

338,183

Employees as of 2023

103,621

Production and assembly sites world wide

33



Truck
Bus

AUTONOMOUS SOLUTIONS

Automation will have a revolutionary impact on transport, improving energy efficiency and safety as well as relieving congestion issues – profoundly changing the future of transportation.

MINING



HUB-TO-HUB



development of autonomous trucks for mining

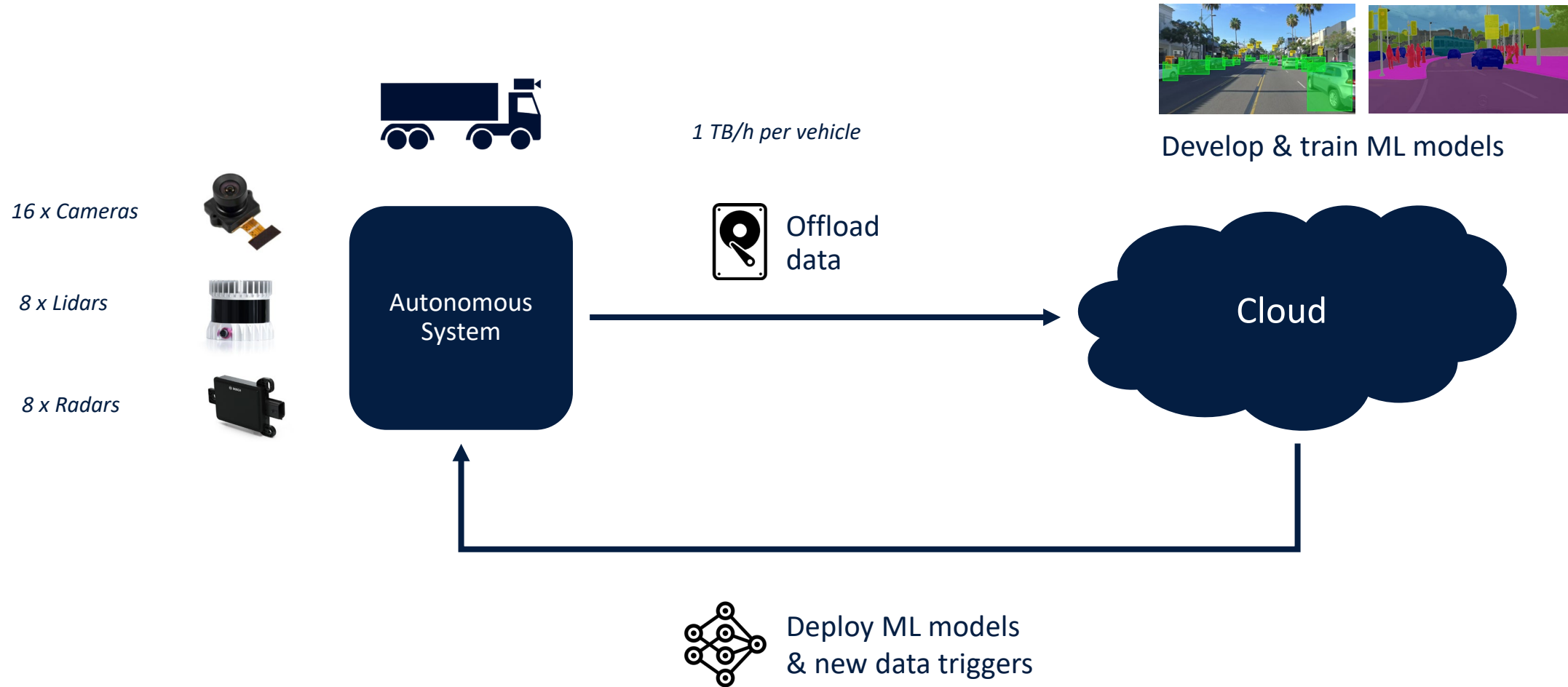


TRATON

self-driving trucks
ON PUBLIC ROADS

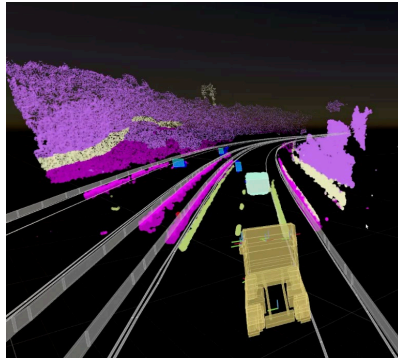


Data & AI loop

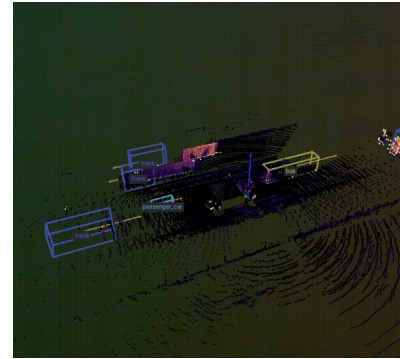


AI capabilities

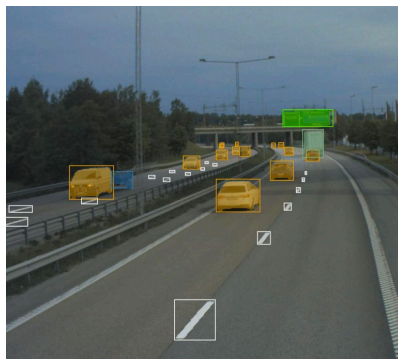
Perception



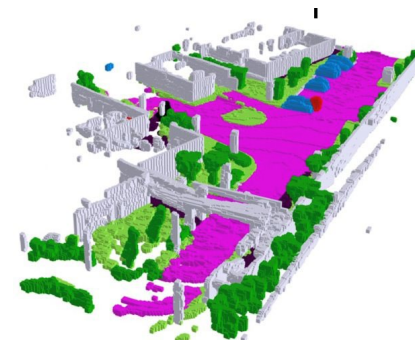
Semantic segmentation + stereo



Multimodal object detection

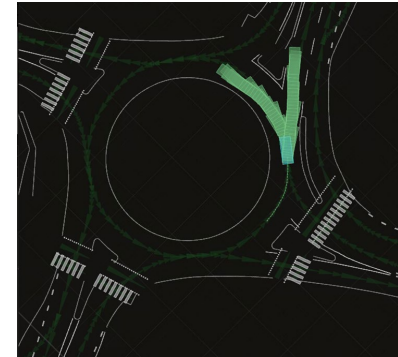


Road feature detection

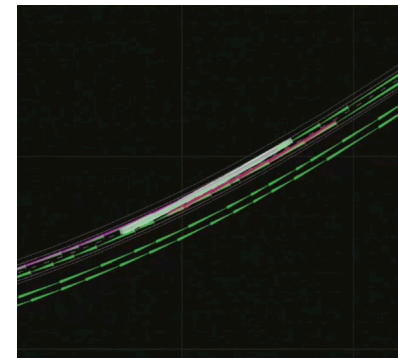


Occupancy

Prediction & planning

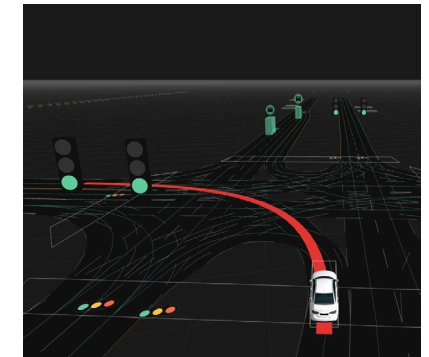


Agent prediction

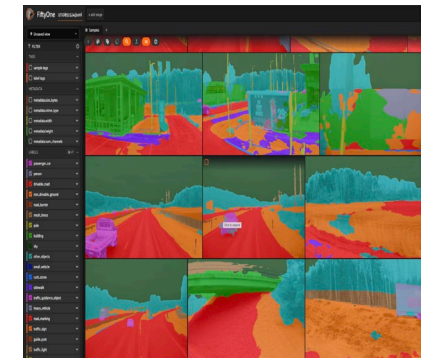


Motion planning

Tools

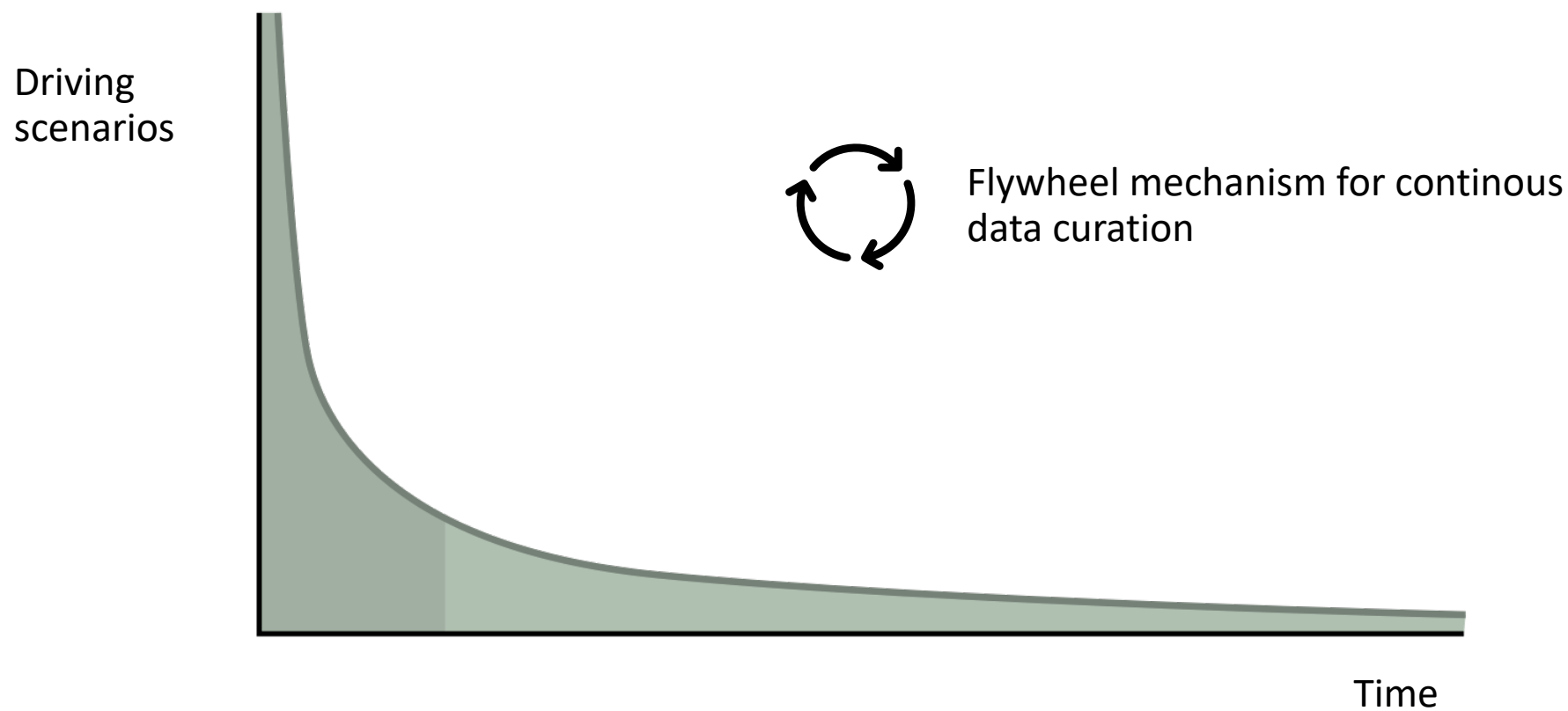


Verification / SBA



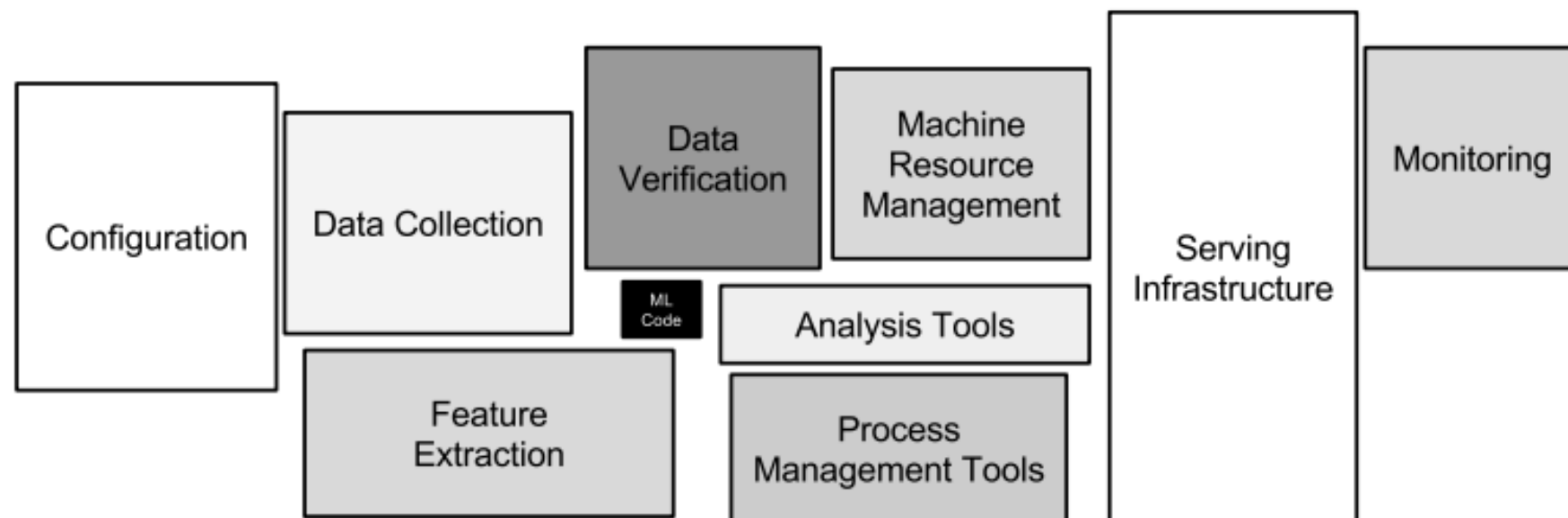
Data selection + augmentation

The long tail problem of autonomous driving



Challenges with data driven development

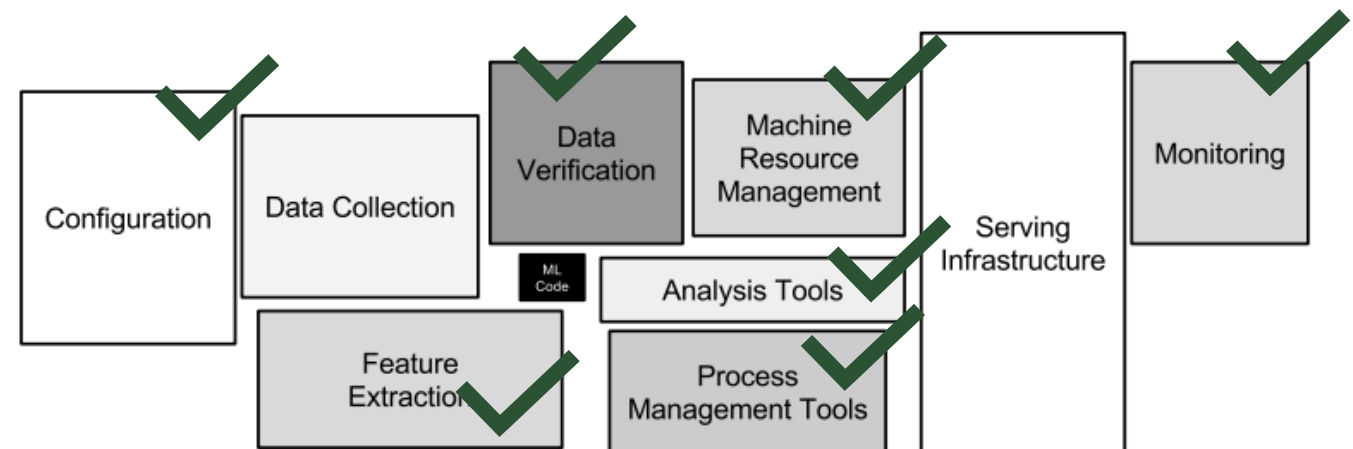
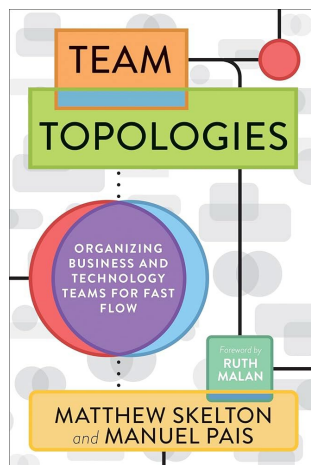
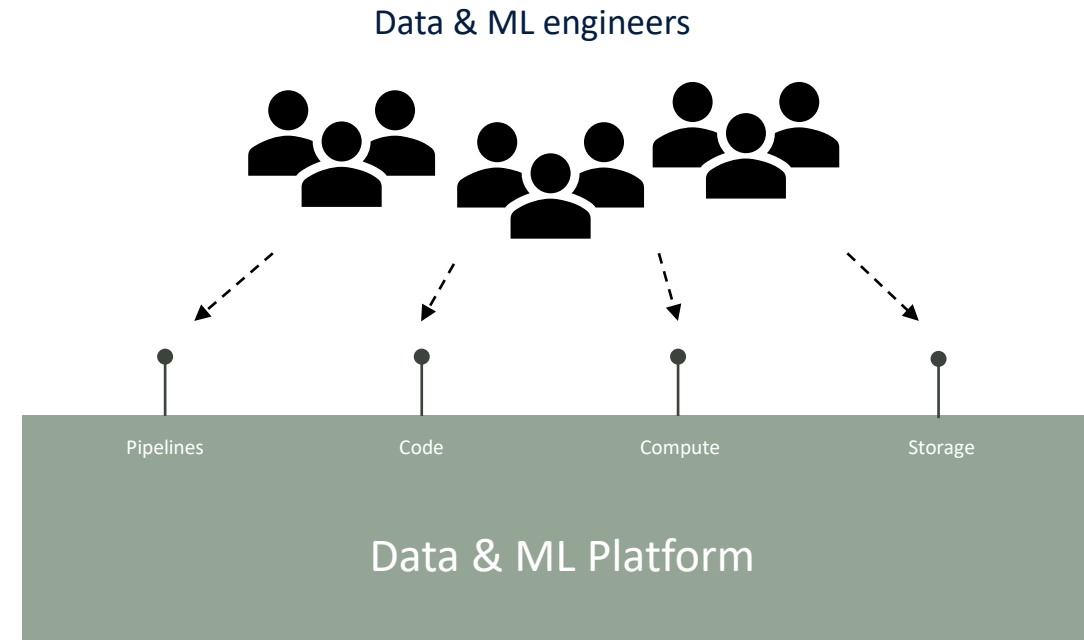
- Complex data and ML tooling
- Growing infrastructure
- Knowledge and competence



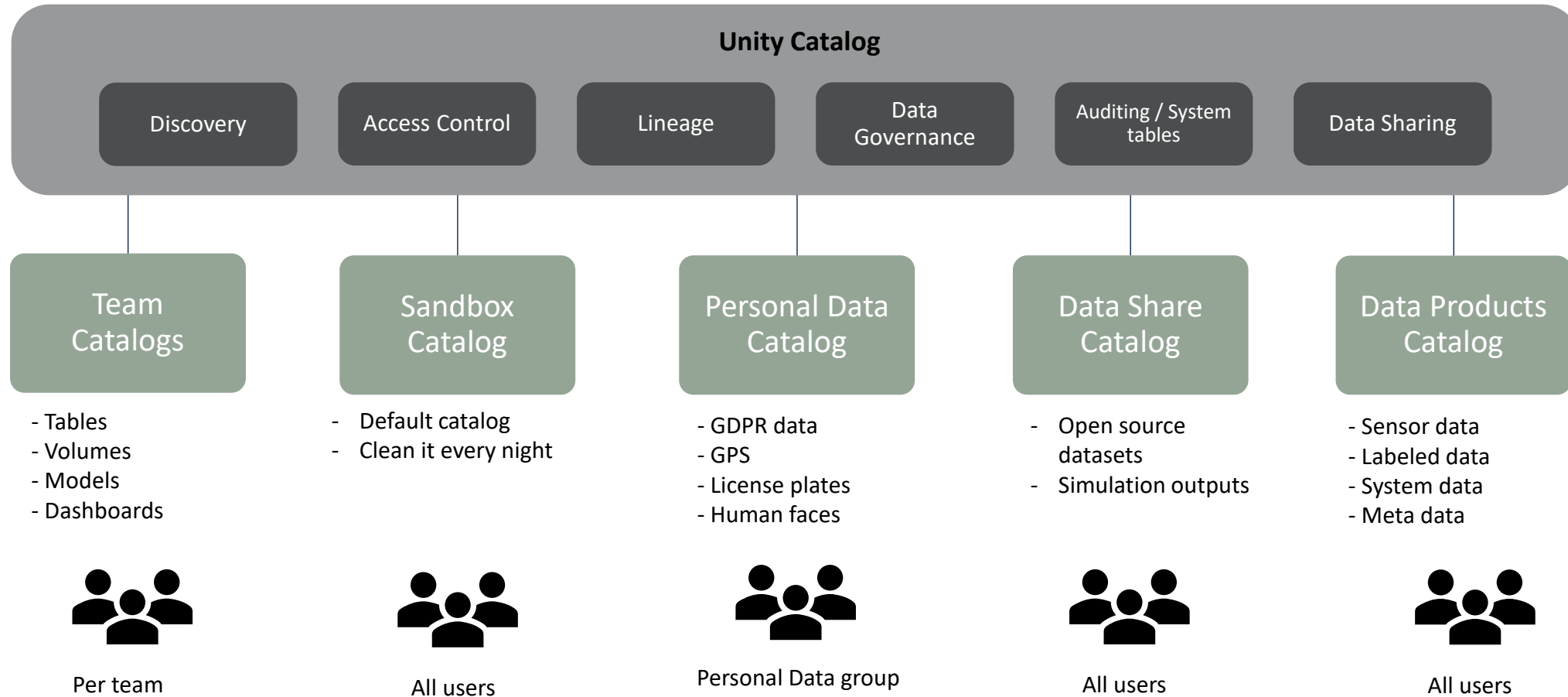
"Hidden technical debt in machine learning systems"

Platform thinking

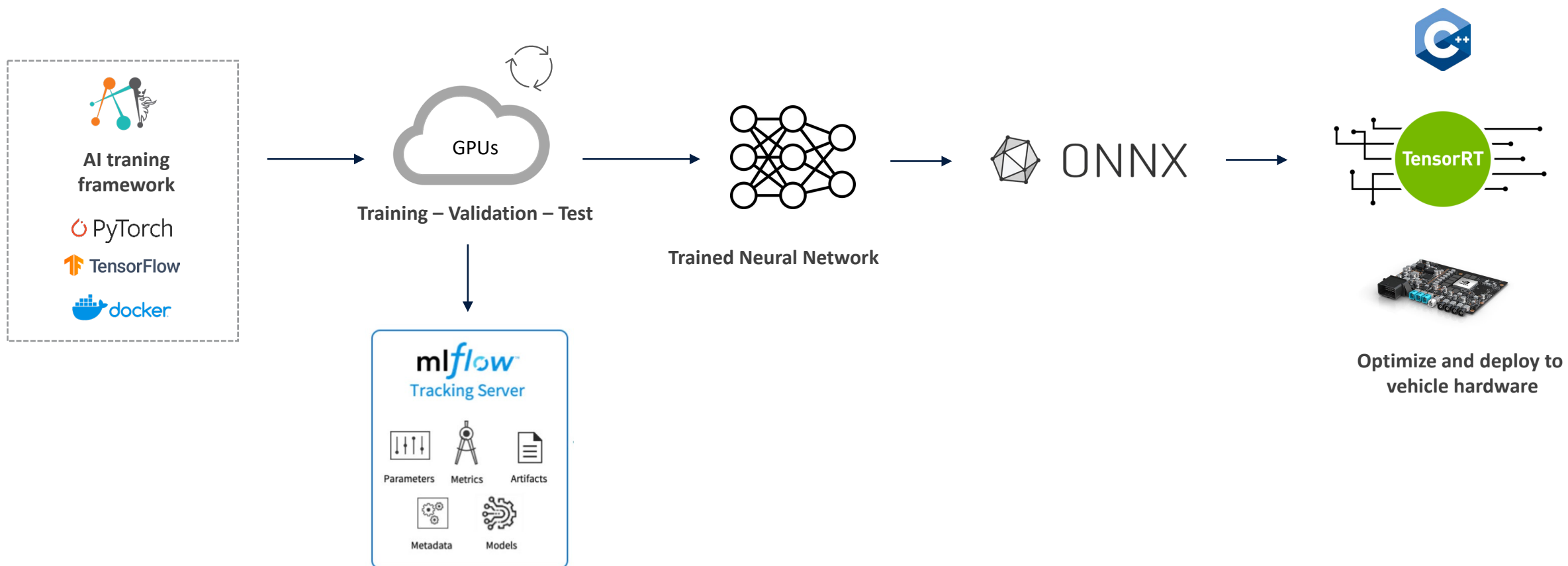
- Simplify complex technology and reduces cognitive load
- Self-service data and ML infrastructure
- Co-development with users



How we use Unity Catalog



ML development workflow

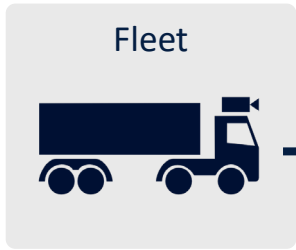




Autonomous Vehicle Lakehouse



TRATON



Upload to S3

Ingestion

Validated

Anonymization



Clustering

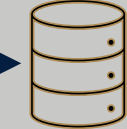
Inference (large DNNs)

Labeling

Data Preparation



Storage



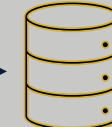
Extracted files



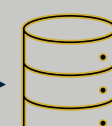
Sensor data



Selected data



Auto labels



Manual annotations



Unity Catalog

Governance

Lineage

Access control

Data Science

Dashboards

Data Exploration

2D/3D Visualization



ML Pipelines

Training

Evaluation

Model Registry



Users



Functional Developers



Data Engineers & Data Scientist



ML Engineers

Annotation & QA provider

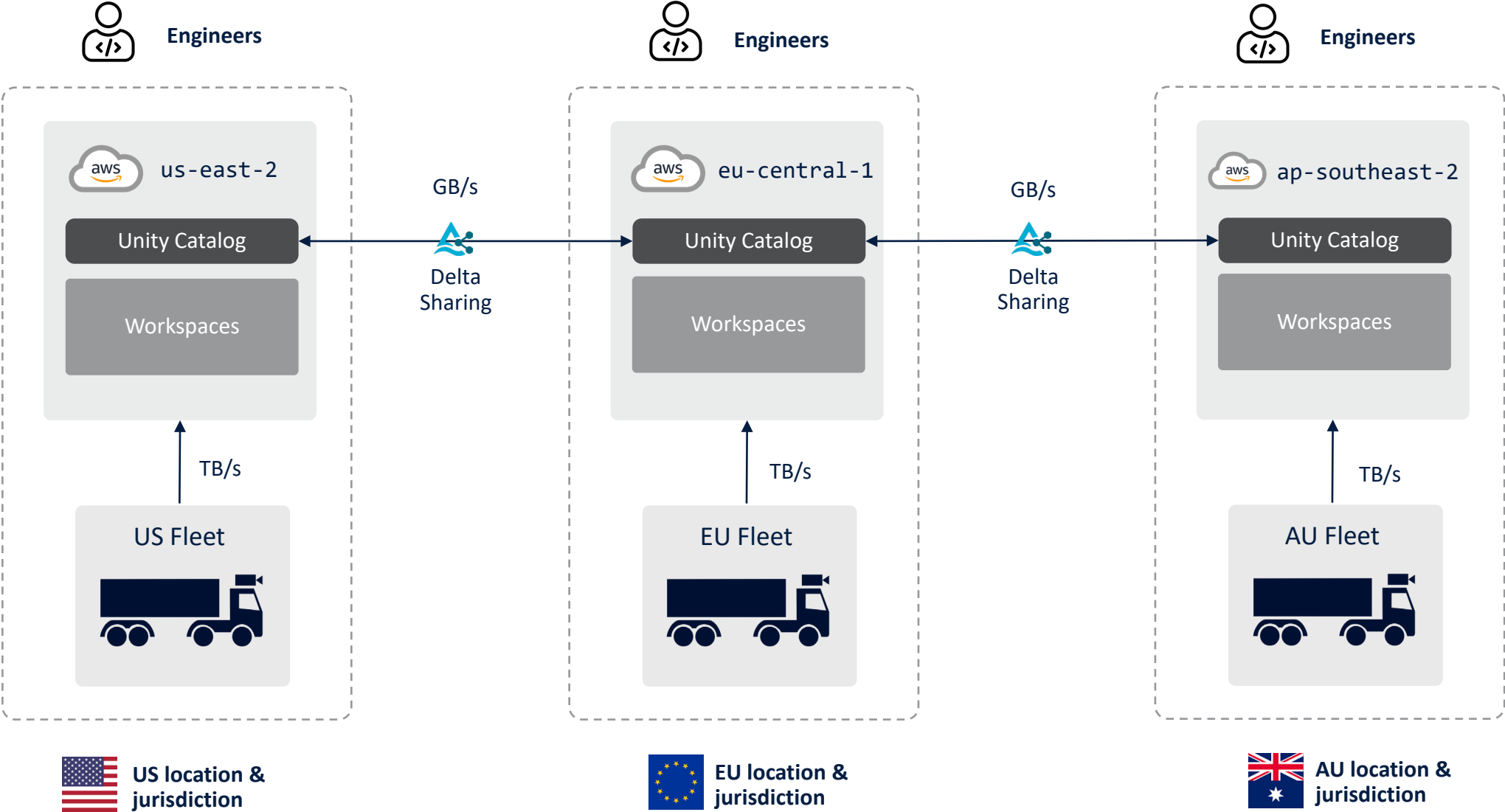


Fleet



CI/CD

Multitenant deployment



Final reflections

- Lakehouse architecture simplifies life as a platform engineer.
- Work with reducing cost.
- Terraform and CI/CD is your friend.
- Co-development with users to fosters collaboration and adoption.

Thank you for listening!

TRATON



Joachim Zetterman
Sr. Data & ML platform lead

