Privacy-First Machine Learning

Bringing the Web on Par with Mobile

P J Łaszkowicz / W3C

Why is Privacy Important?

Inclusivity

- Web as an Egalitarian Ecosystem
- Content-first delivery
- Features included when available
- Devices, platforms, and browsers as distractions from inclusivity
- Progressive enhancement as a proven delivery model

Inclusivity

- Metrics can be harmful
- Humans become demographics
- Delivery becomes biased
- Web tilts towards the wealthier

> “Addressing accessibility, usability, and inclusion together can more effectively lead to a more accessible, usable, and inclusive web for everyone.”

Accessibility, Usability, and Inclusion | Web Accessibility Initiative (WAI) | W3C

State of Mobile

- Great accessibility features
- Backwards compatibility is limited
- Older APIs are costly to support
- Codebases can often diverge
- Dropping handsets is justifiable
- Solution: go web-first

Privacy and Inclusivity

Privacy as a Right

- Privacy as a feature is harmful
- Tracking by default is problematic
- Exploiting poorer users to upsell privacy or collect data is anti-inclusive
- Not providing a fair & inclusive choice creates a broken web

Privacy and UX

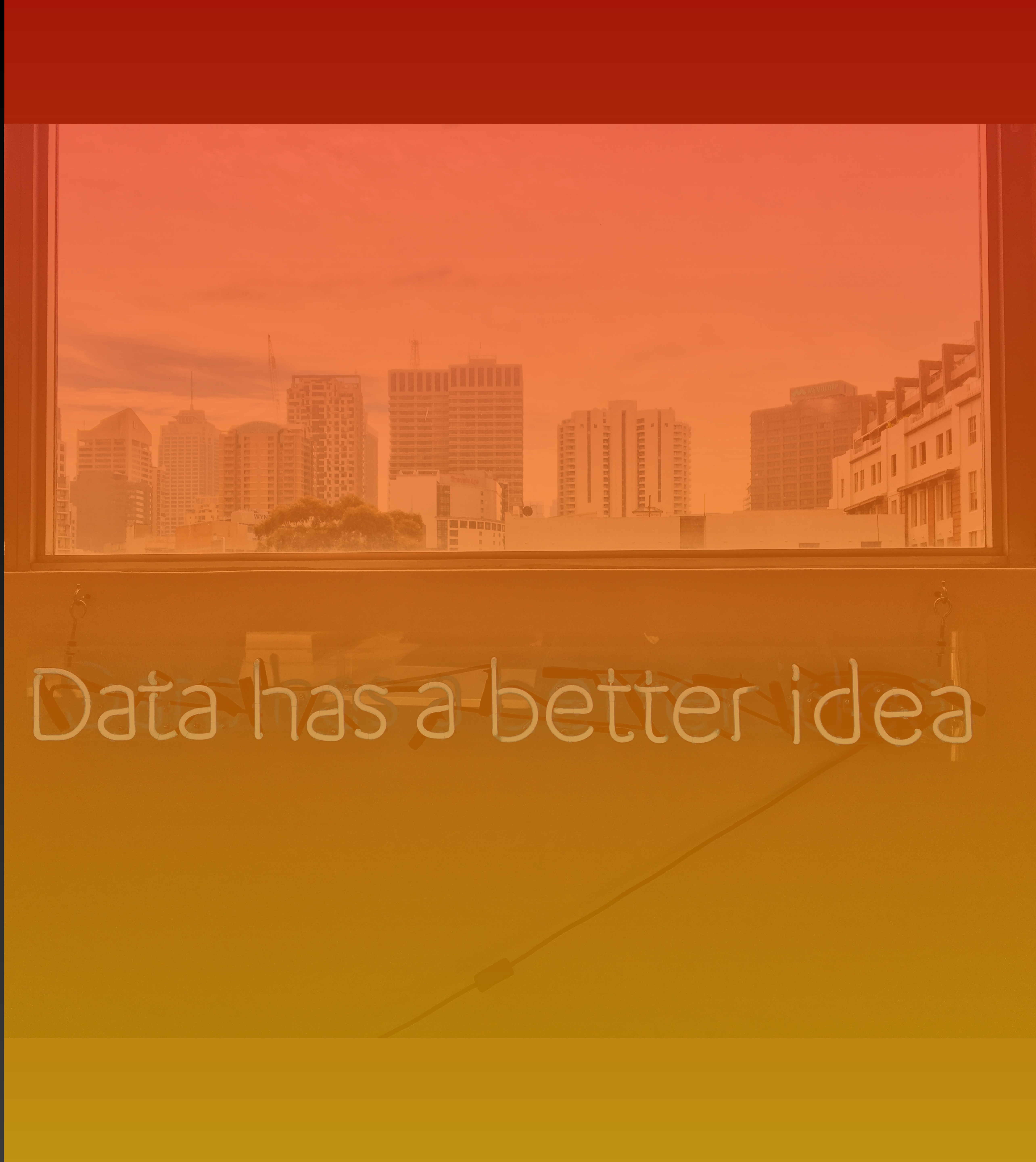
Privacy and UX

- Opt-ins for tracking are bad UX
- Opt-ins are created by defaulting to cookie & tracking on first-load
- To fix the UX, cookies and trackers can be moved in the UX flow
- Essentially reducing data collection to near-zero creates better experiences

Privacy and Machine Learning

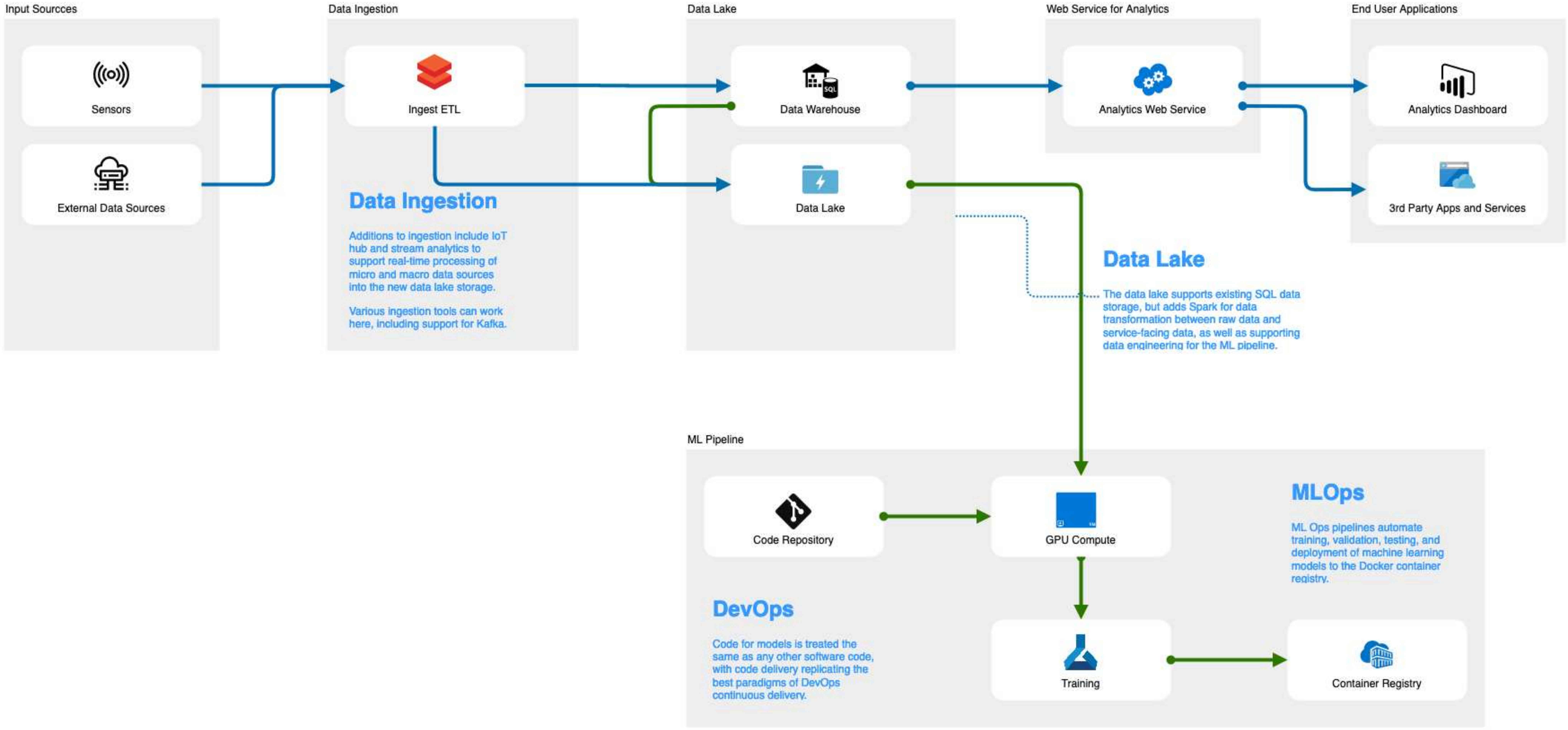
Privacy and Machine Learning

- Differential Privacy
- Federation
- Reduction of data collection
- All on Edge & IoT already
- Centralized learning is not necessary or privacy-enabled



Legacy ML Architecture

Simplified ML Service



Input Sourcces

Sensors

External Data Sources

Data Ingestion

Ingest ETL

Data Ingestion

Additions to ingestion include IoT hub and stream analytics to support real-time processing of micro and macro data sources into the new data lake storage.

Various ingestion tools can work here, including support for Kafka.

Data Lake

Data Warehouse

Data Lake

Web Service for Analytics

Analytics Web Service

End User Applications

Analytics Dashboard

3rd Party Apps and Services

Data Lake

The data lake supports existing SQL data storage, but adds Spark for data transformation between raw data and service-facing data, as well as supporting data engineering for the ML pipeline.

ML Pipeline

Code Repository

GPU Compute

Training

Container Registry

MLOps

ML Ops pipelines automate training, validation, testing, and deployment of machine learning models to the Docker container registry.

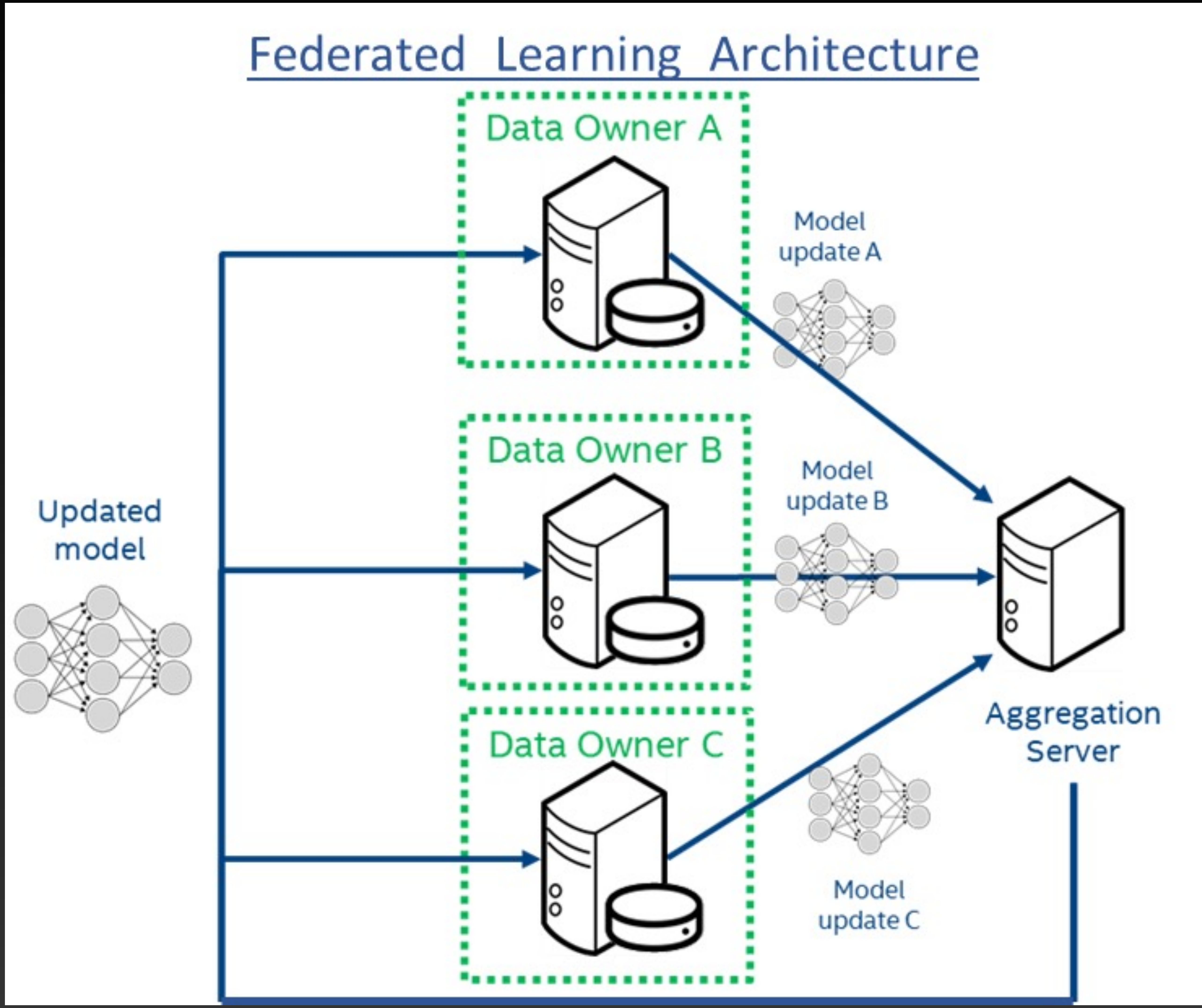
DevOps

Code for models is treated the same as any other software code, with code delivery replicating the best paradigms of DevOps continuous delivery.

Modern ML Architecture

ML Architecture

- Decentralize services
- Distribute training and inference
- Federate privacy-oriented models
- Inference on the edge (*when possible*)
- Add ML progressively (*ES Modules rule!*)



Web vs Native Privacy

Web vs Native Privacy

- Browsers *will* pick the best hardware available
- Wasm is a great companion
- Wasm is portable (re-use) & fast
- Browser models can be privacy-enabling and inclusive
- Stop defaulting to data collection

Thank you.