Stop Monitoring Apps; Start Monitoring Behavior

- The Problem:
  A subtle memory leak ➔ slow DB queries ➔ HTTP 500 errors.
  i.e. 'unknown unknowns'!
- The Data:
  A mix of all raw server logs (syslog, auth, Nginx, DB)
  - Creates a high-dimensional 'data-fingerprint' of a healthy server.
  - No need for log parsing (=loss of information)
  - Model: A transformer-based model learns the "language" of a healthy server
- Federated vs Centralized (Why?):
  - Saves huge amounts of bandwidth (weights vs. logs).
  - The global model learns from a massive, diverse baseline of "normal" behavior.
  - Privacy: No raw, sensitive logs are ever shared.
- Stack: Flower, PyTorch, Exalsius.
- Data: 3 public datasets (~1GB each).
  - Datasets are mixed, and distributed between clients using flower simulation