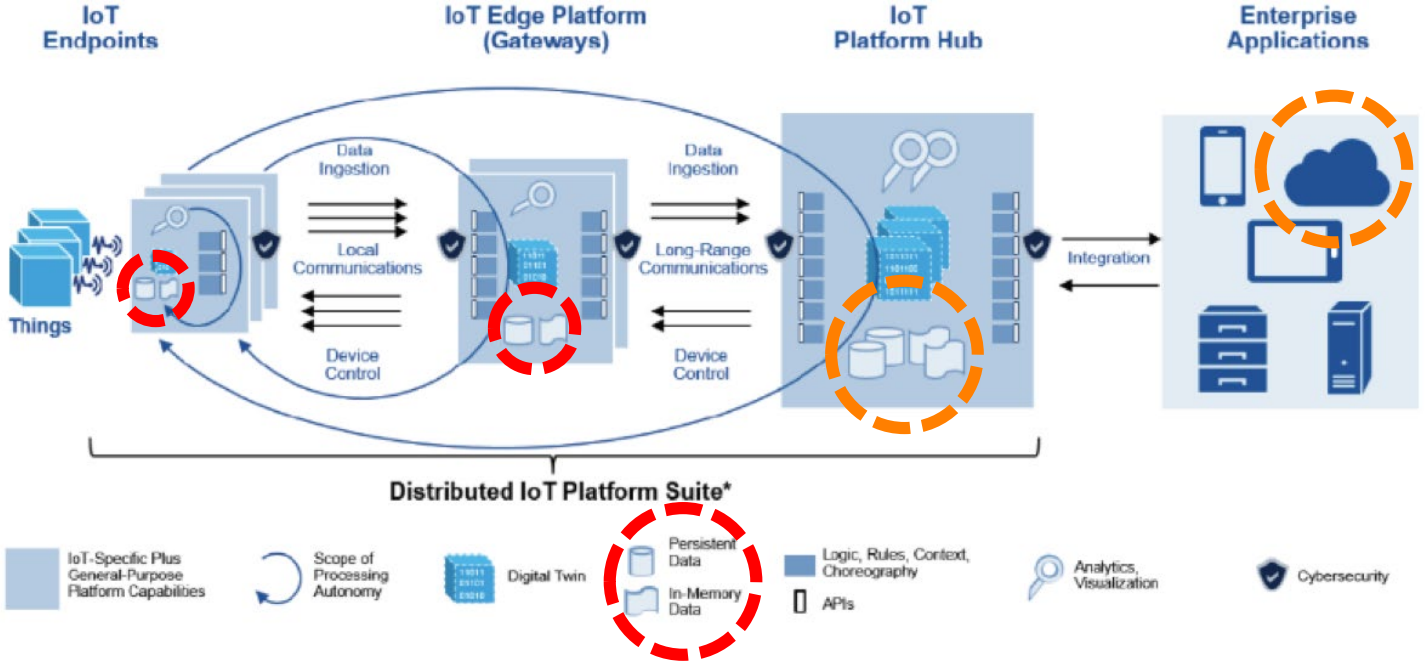
WEKA.IO
World's Fastest File System

High-Performance Flash Storage to Enable IoT and Accelerate AI

EMBD-302B-1: Flash Memory, IoT and AI — Bringing It All Together

Andy Watson
CTO, WekaIO
watson@weka.io, @the_andywatson

IoT — Where Does the Data Go?

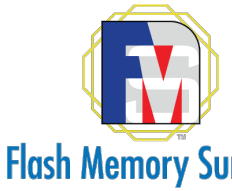


Gartner Reference Model for IoT Business Solutions: Edge, Datacenter(s) & Cloud

Flash for IoT at the Edge

Energy Utility	0.5 TB / day
Offshore Oil Field	0.75 TB / week
Oil & Gas Refinery	1 TB / day
Airliner in flight	20 TB / hour

- IoT data generated at the Edge
  - Arriving rapidly, too valuable to lose
    - Capture it all
    - Process locally to scrub & compress
    - Run simple analytics to flag major anomalies immediately
  - Use-cases have different requirements
    - Some will forward all scrubbed data
    - Others only send upstream exceptional or summary data
- Flash advantages: fast, small, low power, lightweight
  - Tolerates vibration & hi/low temp better than HDD



Flash for IoT in the Datacenter/Cloud

- IoT data sent upstream from the Edge
  - Even if distilled per source, in aggregate a continual flood
  - All must be captured and safely stored
    - Made available to multiple analytics purposes
    - At this point, category label is not so much “IoT” as “Data Lake”
    - Likely includes multi-site replication and tiering to an archival layer
- Flash advantages: low-latency and high-throughput
  - For both sequential or random access patterns
  - Any GPU-accelerated processing absolutely requires flash
    - Each individual GPU can ingest data at 100-gbit/s data rates

Flash Enabling IoT at the Edge

- IoT at the Edge — Medical Devices
  - Often in the form of patches worn by patients
    - Possibly locally configured/managed by smartphone software
    - Monitors vital signs, location & movement, blood levels
    - Notify emergency services (patient falls or in other serious events)
    - Homeostatic interaction (electronic pulses, insulin, etc.)
  - More ambitious future devices could take bolder action
    - Intervention when epileptic onset is detected
    - Enable improved motor skills for patients with spinal injuries, etc.
  - Data uploaded (usually at intervals) for clinical supervision

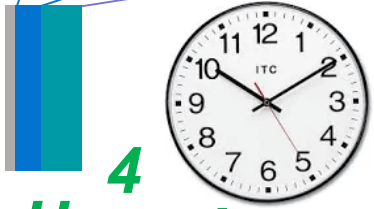
Flash Enabling IoT Upstream

- IoT in the Datacenter — Autonomous Vehicles
  - Vehicles outfitted with cameras & sensors collect data
    - Roaming streets and highways, about 10 TB/hour per car
      - This is not the same as data vicariously collected from owners' cars
      - This is data used as input to neural-net AI/ML training/validation to develop and refine the algorithms for self-driving capabilities
    - Multiple petabytes of new data per week added to the data lake
  - Many forms of analytics are applied to this IoT data
    - And also applied to owners' cars' IoT data being uploaded
  - GPU's are used to accelerate the AI/ML compute farm
    - NVMe flash storage can barely keep up with the GPU's IO demands

Leveraging NVMe Flash for GPU's

at an actual AI/ML customer site
(autonomous vehicle software development)

~80x Speedup



4 Hours!

Eliminating the
GPU IO bottleneck
with NVMe flash + WekaIO