Edge Analytics & Streaming Flows

Problem

Collecting data from the field (e.g., IoT) is more complex than traditional analytics. It's intermittent, highly unstructured and typically real-time. Transferring all that data into cloud incurs latency, and bandwidth won't scale with the growth in volumes and data. Real-time analysis becomes less achievable.

Solution

Place IoT event processing and device management (firmware and updates) in one or more of the edge nodes. Since the node is located at the intersection point of consolidated access (cellular, broadband, internet, etc.)—including frequency networks (you install as more bands become available)—the edge node is the closest point to the field and the clouds. Placing device management and IoT analytical capabilities at the edge solves latency, bandwidth and device complexity constraints. It also provides multidestination control and choice in the types of network/providers you want to use, as well as which cloud analytic platforms are available. Participate in IoT ecosystems by connecting with partners and exchanging data. As data comes in it is validated, authenticated, inspected, pre-processed, stored in the global namespace, and then delivered to the next downstream processing step.

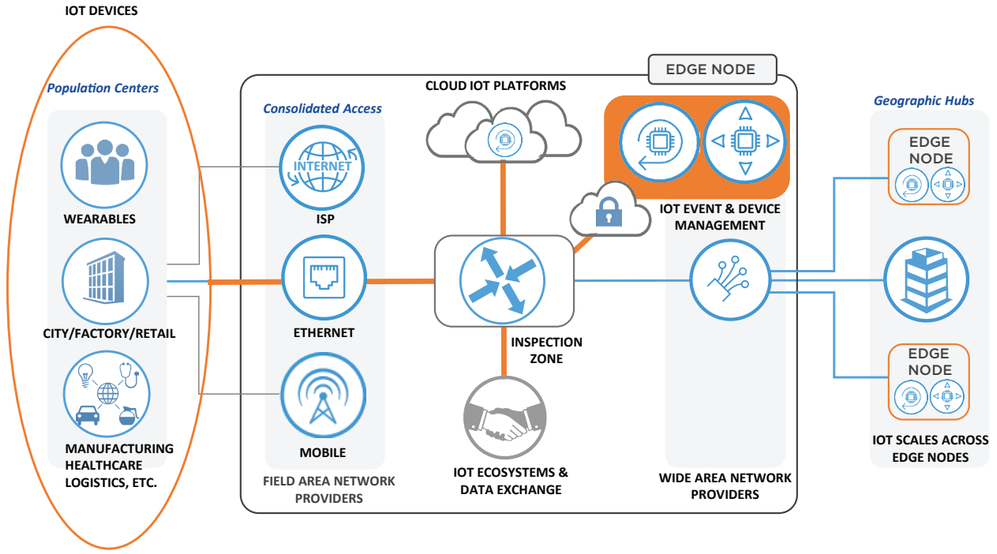
Constraints

1. Data from the field isn't the same as prepared data — it can be messy, intermittent, unstructured and somewhat dynamic (if the device is mobile).
2. The volumes of devices, variety of sources, frequency of samples and data are all growing, which will not scale using traditional approaches.
3. In many use cases latency matters, and delays between the event and the reaction need to be kept to near-real-time. Delays will become unacceptable as throughput is impacted by that growth.
4. Traditional approaches to centralizing all the data to run analytics (in the cloud) are not sustainable for real-time use cases. But in this architecture, where else can it be done?
5. Alternatives that put more machine learning intelligence in the device increase device complexity, draw more power and lead to the results in data being discarded. When we decide we need that data, we are back to where we started.

Steps

1. Establish segmentation flows from field area networks. Messages from IoT gateways (in the field) get published on the message bus; otherwise a local IoT gateway processes it first.
2. Boundary control validates and authenticates the source and message. (Security Step 1).
3. Valid messages go through an inspection zone and policy enforcement (Security Steps 2 and 3).
4. Messages are persisted to the data repository (Data Step 1), then delivered IoT event processing. Downstream messages are published to go onto cloud analytic platform(s) of choice, or your own cloud-agnostic repository.
5. IoT gateway "device requests" go through the same flow but are subscribed to by the device management function. Appropriate payload(s) (firmware, etc.) are loaded from the data repository and published back to the IoT gateway (or device directly).

Reference View



* Security Blueprint — IOAKB.com

Forces

- The growth of traffic and data at the edge is driving the need to reduce distance and bring analytics closer. Distributing computational processing is a known practice (data gravity).
- While bandwidth is cost prohibitive, latency causes the biggest problem, as there is a point where adding more bandwidth only marginally helps (physics).
- IoT and related field use cases are driving more simplicity into the field from the edge. It's easier to change something in 10 places than 10,000,000.
- Likewise, not all actions required on all data are equal, and investigation into staging processing allows a more balanced approach. Level 2 data does not need the heavy processing that Level 5 does.

Results

- IoT edge capabilities that can scale to billions of devices, at each metro location globally.
- Collected data can be validated, authenticated and inspected before being pre-processed. (Security Blueprint).
- All data can be stored in the global namespace (Step 1) — no need to discard.
- Most efficient use of bandwidth with lowest latency for real-time event reactions and the most efficient way to scale.
- As much processing as needed can be localized at the edge, with choice of cloud IoT platforms.
- Monetize data and sell access on a data exchange, gaining the insights and generating revenue.