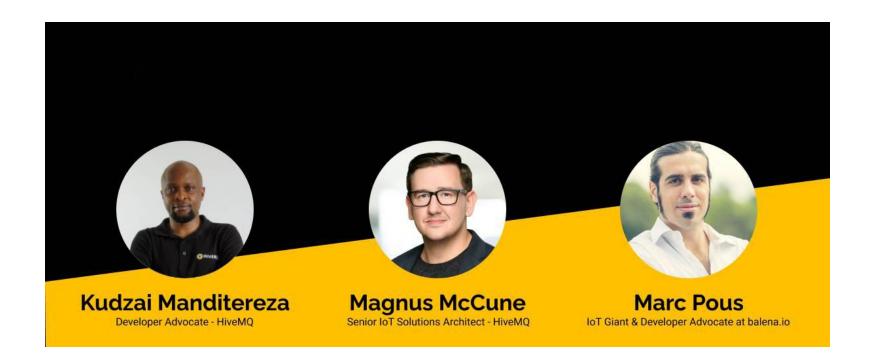




Speakers

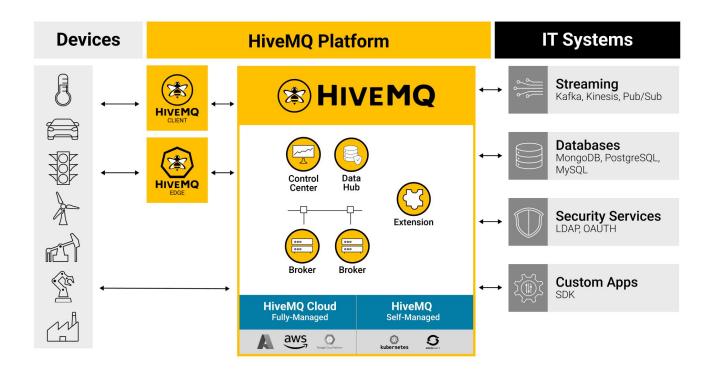


Agenda

- Introduction
- MQTT Standards for Integrating Edge AI Systems
- <u>Demo 1</u> Fully Integrated Pattern for Integrating Edge AI into MQTT
- <u>Demo 2</u> Unstructured Data
 Pattern for Integrating Edge Al into MQTT
- Q&A



The Enterprise MQTT Platform





Key Industries



Why Standards for MQTT in Edge Al are Critical?

Interoperability and Flexibility

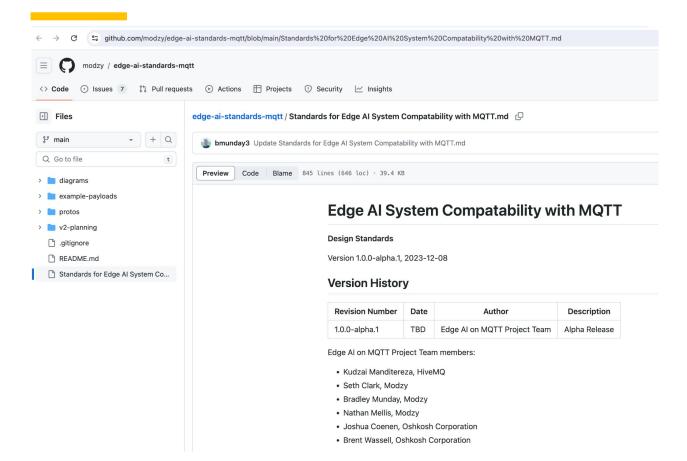
Cost Savings and Accelerated Deployment

Scalability and Consistency

Innovation and Community Engagement



Contributors and Resources

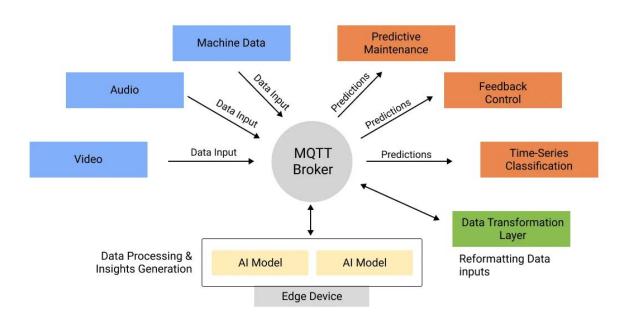


Common Patterns for Edge AI on MQTT

	MODEL INPUT: MQTT TOPIC(S) [SUBSCRIBE]	MODEL INPUT: OTHER PROTOCOLS, DIRECT SENSOR DATA, ETC.
Model Output: MQTT topic(s) [Publish]	The "Fully-integrated" pattern	The "Unstructured Data" pattern
Model Output: Other systems lacking MQTT support	The "Ambassador" pattern	Not covered by this standard

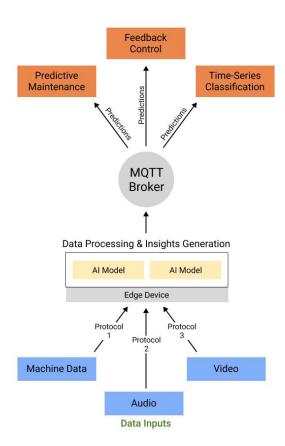


The Fully-Integrated Pattern

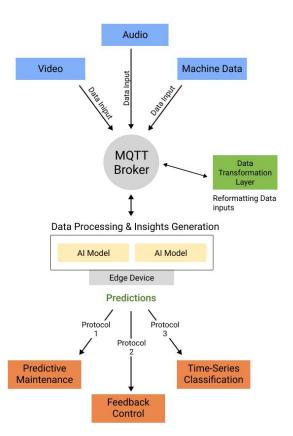




The Unstructured Data Pattern



The Ambassador Pattern





Guidelines for MQTT Topic Structure Design



Topic Namespaces for Edge Al

- Raw Data Namespace hold raw sensor data as it is captured. Al model input in the ambassador and fully-connected patterns.
- Inference Namespace hold metrics directly generated by a machine-learning model; do not necessarily provide business value on their own.
- Insight Namespace metrics with innate business value that are composed, calculated, or otherwise built on top of individual inference metrics.





Raw Data namespace Example

site/area/line/cell/milling_machine/raw

Messages to this topic might include the following metrics:

- Air temperature (K)
- Process temperature (K)
- Rotational speed (rpm)
- Torque (Nm)
- Tool wear (min)



Inference Namespace Example

Model name

site/area/line/cell/milling_machine/Machine Failure Prediction/0.0.1/inference

Messages to this topic might include the following metrics:

- Failure Likelihood (with confidence score measured from 0 to 1)
- Non-Failure Likelihood (with confidence score measured from 0 to 1)

Model version



Insight Namespace Example

Model name

site/area/line/cell/milling_machine/Machine Failure Prediction/0.0.1/insight

Messages to this topic might include the following metric:

 Maintenance Required: Published anytime the inference/failure score is larger than the inference/no_failure score, indicating that the Machine Failure Prediction model has found the milling machine to be likely to fail sometime soon.

Model version



Flat MQTT Topic Structure

[Customized MQTT topic structure]/Edge_DeviceID +/model_name/model_version/inference

- **[Customized MQTT topic structure]** Any existing top-level topic structure for a flat MQTT namespace
- **Edge_DeviceID** A unique identifier of some kind pointing to the device's location upon which an AI/ML model is running.
- model_name The name of a specific AI/ML model.
- model_version Version of AI/ML model
- **inference** The namespace element that will contain any messages generated by this version of this model.



MQTT Sparkplug Topic Structure

namespace/group_id/message_type/edge_node_id/[device_id]

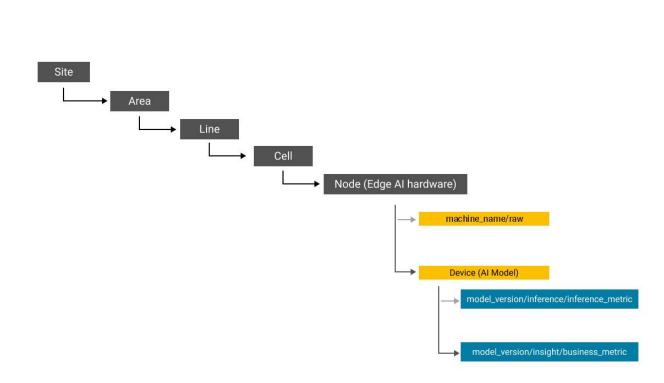
- namespace Sparkplug B root namespace
- group_id Recommends using a concatenation of the ISA-95 format of site:area:line:cell
- message_type Edge AI apps will primarily publish results using the DDATA message type, but other message types supported.
- edge_node_id Identity of the computing hardware that is running models or communicates to the MQTT broker.
- device_id Identity of the AI/ML model that is generating inferences.

ip_camera_1,
 edge_server_1,
 machinery_1

➤ Machine Failure Prediction



Unified namespace Snapshot





Guidelines for MQTT Payload Structure Design



Structured Payloads for Edge Al Outputs

- Predictions regarding potential future events.
- Structured insights (Classification or Detection) that are extracted from unstructured sources like audio and video.





Recommended Formatting

- Protobuf –should be used whenever possible for its efficiency and future-proofing qualities
- **JSON** Should be used for compatibility with non-protobuf capable systems.





Flat MQTT Payload Template

```
"identifier": "inference-2HYZh8a4jtFi3xFc4e3TWRmclff",
"model":{
   "identifier": "brzrip6cxk",
   "version":"0.0.1",
   "name": "Machine Failure Prediction"
},
"tags":{
   "sourceTopic": "site:area:line:cell/node/device/raw/sensor_name",
   "sourceMessageID": "abcd1234",
   "inputSizeInBytes":32,
   "inputSha256Digest": "be01ef104fb88fd151132733e746fe29b997348bf34be875e25ba48c0d7436ca"
"resultType":"classPredictions",
"result":{
   "classPredictions":[
         "className":"no_failure",
         "score":0.974
         "className":"failure",
         "score":0.026
"explaination":{
```

23



MQTT Sparkplug DBIRTH Payload

```
"timestamp":1486144502122,
"metrics":[
      "name":"0.0.1/inference/identifier",
     "timestamp":1486144502122,
     "dataType":"string",
     "value": "inference-2HYZh8a4jtFi3xFc4e3TWRmclff"
     "name":"0.0.1/inference/model/identifier",
     "timestamp":1486144502122,
     "dataType":"string",
     "value":"brzrip6cxk"
      "name": "0.0.1/inference/model/version",
     "timestamp":1486144502122,
     "dataType":"string",
     "value":"0.0.1"
      "name": "0.0.1/inference/model/name",
     "timestamp":1486144502122,
     "dataType":"string",
     "value": "Machine Failure Prediction"
     "name":"0.0.1/inference/tags/sourceTopic",
     "timestamp":1486144502122,
     "dataType":"string",
     "value":"site:area:line:cell/node/device/raw/sensor_name"
```

MQTT Sparkplug Host Application

Metric		Value	Data Type
Site:Area:Line:Cell	/group_id		
^L Edge Node ID	/edge_node_id		
^L Edge Node Device ID	/device_id		
LInference			
^L identifier		inference-2HYZh8a4jtFi3xFc4e3TWRmclff	String
Lmodel			
Lidentifier		brzrip6cxk	String
Lversion		0.0.1	String
∟name		Machine Failure Prediction	String
^L tags			
∟sourceTopic		site:area:line:cell/node/device/raw/sensor_name	String
∟dataType		abcd1234	String
LinputSizeInBytes		32	Integer
LinputSha256Digest		be01ef104fb88fd1	String
∟resultType		classPredictions	String
∟result		{"classPredictions":[{"className":"no_ failure","score":0.974}, {"className":"failure","score":0.026}]}	String

MQTT Sparkplug DDATA Payload

```
"timestamp":1486144502122,
"metrics":[
      "name": "0.0.1/inference/identifier",
      "timestamp":1486144502122,
      "dataType": "string",
      "value": "inference-2HYZh8a4jtFi3xFc4e3TWRmclff"
      "name":"0.0.1/inference/tags/sourceTopic",
      "timestamp":1486144502122,
      "dataType":"string",
      "value": "site:area:line:cell/node/device/raw/sensor name"
      "name": "0.0.1/inference/tags/sourceMessageID",
      "timestamp":1486144502122,
      "dataType":"string",
      "value": "abcd1234"
      "name":"0.0.1/inference/tags/inputSizeInBytes",
      "timestamp":1486144502122,
      "dataType":"integer",
      "value":32
      "name":"0.0.1/inference/tags/inputSha256Digest",
      "timestamp":1486144502122,
      "dataType":"string",
      "value":"203nvi30fh4fb88fd120932733e746fe29b99732ifhi34be875e25ba48c0d7436ca'
      "name": "0.0.1/inference/result".
      "timestamp":1486144502122,
      "dataTvpe":"string",
      "value":'{"classPredictions":[{"className":"no failure", "score":0.087}, {"className":"failure", "score":0.913
"seq":1
```



Edge AI/ML Model Result Formats



Classification

Classification assigns a class to an individual piece of data. This might be useful for classifying an individual image, audio snippet, video frame, or a piece of machine data.

Multi-Classification

Similar to classification models, but used when model outputs are grouped into more than two distinct classes.

```
"classifications":[
      "classPredictions":[
            "class":"className",
            "score":1.0
```

Object Detection

Object detection is used to identify regions of interest within an image or video that are defined by a bounding box. Bounding boxes can have one or more classifications, and images can have one or more bounding boxes.

```
"detections":[
     "class":"className",
     "score":1.0,
     "boundingBox":{
         "x":100,
         "y":200,
         "width":300,
         "height":400
```

Named Entity Recognition

Named entity recognition is used to identify unique entities, such as names, organizations, and locations, within a larger corpus of text.

```
"entities":[
      "entityGroup":"B-LOC",
      "score":1.0,
      "textSpan":{
         "start":0,
         "end":5,
         "text": "Paris is a city."
```

Demos

Fully Integrated Pattern - Marc Pous, balena.io

Unstructured Data Pattern - Magnus McCune, HiveMQ



Demo 1: The **Fully-Integrated Pattern** with HiveMQ Edge Al balena

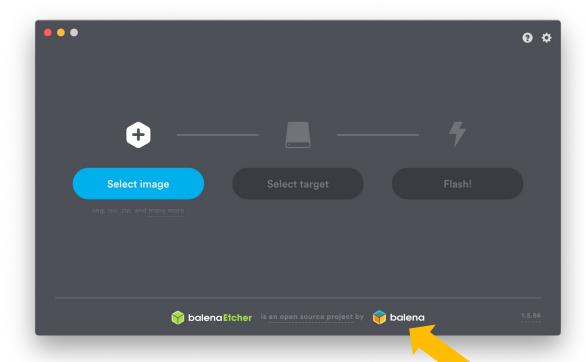
Marc Pous
IoT Giant & Developer Advocate

Anyone?





Anyone?



35



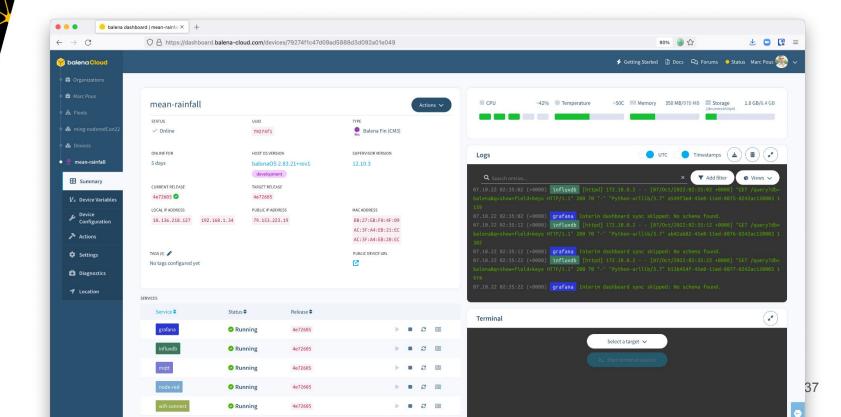
What is balena?

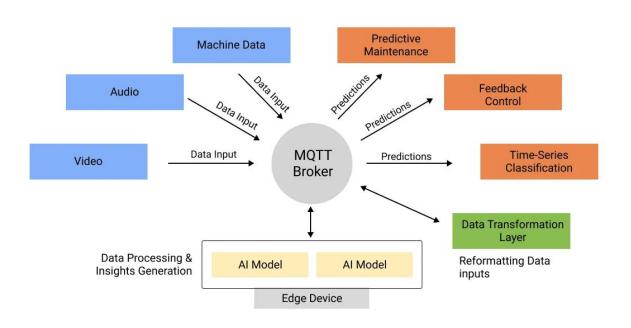


A secure container-based technology stack that enables you to **develop**, **deploy**, **manage** and **scale** large fleets of IoT Linux devices at any stage.

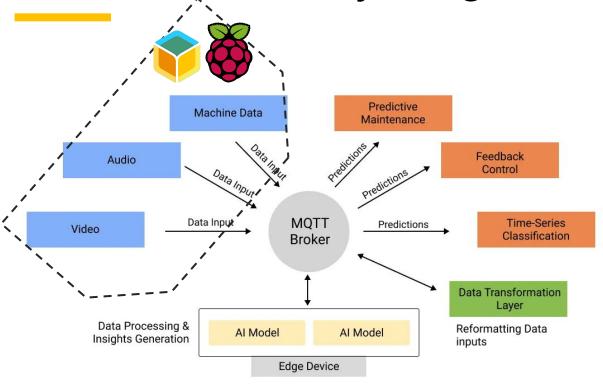


What is balena?

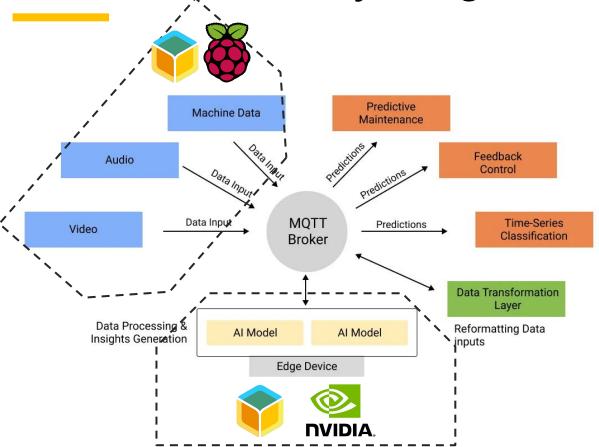


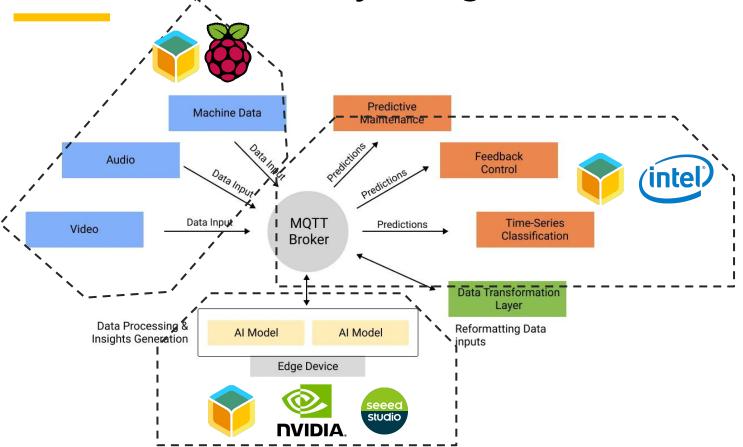


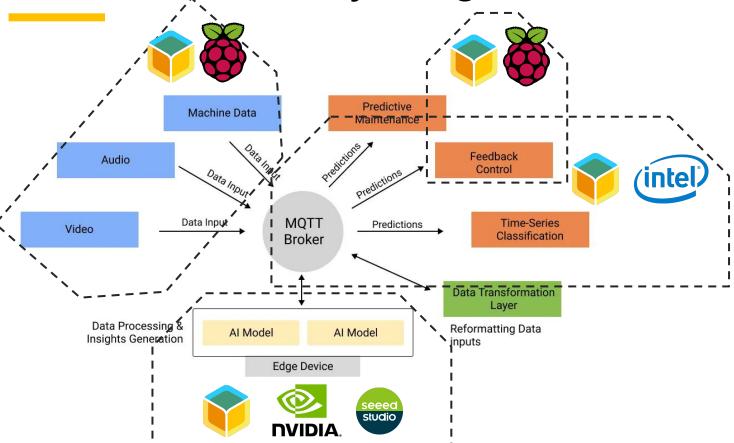






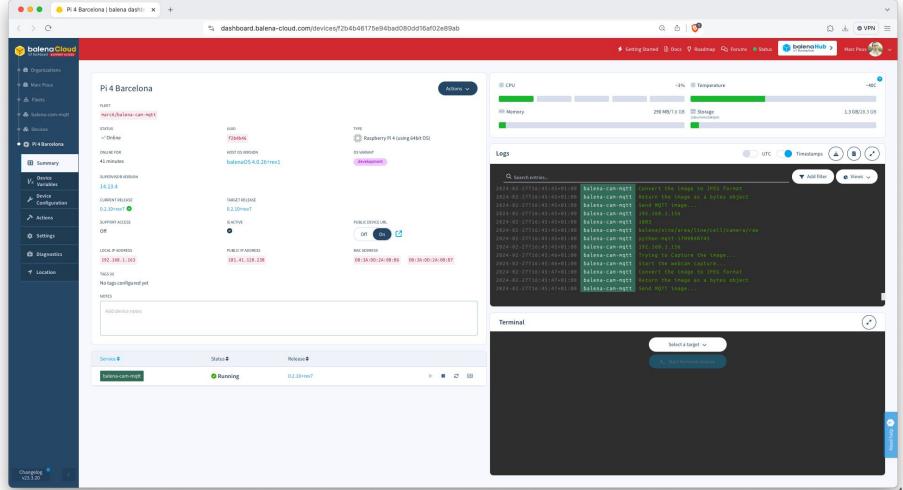


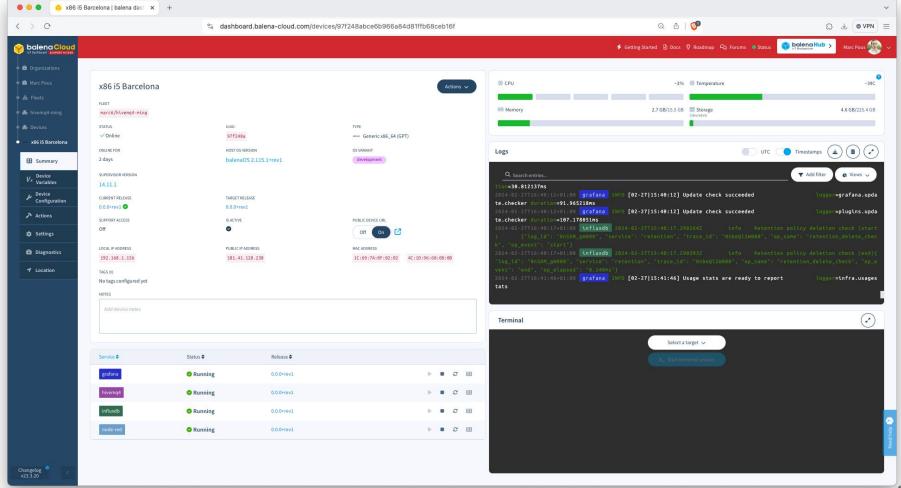


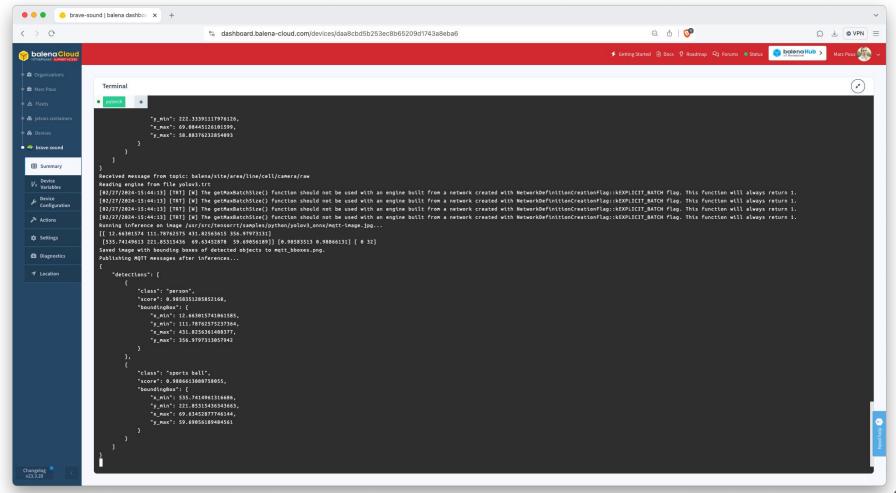


balena Demo

- balenaCam MQTT Raspberry Pi Camera publishing over MQTT.
 https://github.com/mpous/balena-cam-mqtt
- MING broker Intel NUC with the MING stack.
 https://github.com/mpous/ming
- Seeed Studio ReComputer J4012 / NVIDIA Jetson Orin 16GB Edge
 Al device with TensorRT Al model to recognize real-time images.
 https://github.com/mpous/J4012-pytorch-mqtt







```
• • •
    "detections": [
            "class": "person",
            "boundingBox": {
            "class": "sports ball",
            "boundingBox": {
```



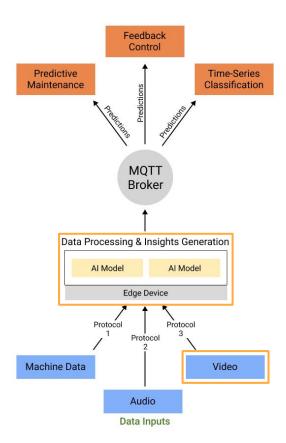


Demo 2:
The
Unstructured
Data Pattern
with HiveMQ

Magnus McCune
Senior IoT Solutions Architect

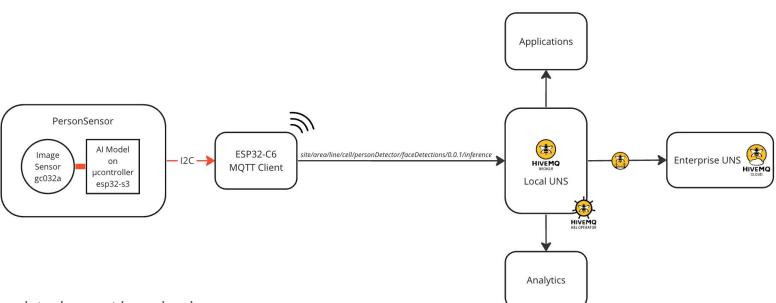


The Unstructured Data Pattern





Demo - Unstructured Data Pattern



*Image sensor data does not leave hardware

	ESP32-S3	ESP32-C6
Processor	Tensilica Xtensa 32 bit 240MHz dual-core	RISC V 32 bit 160MHz
SRAM	520KB	512KB
ROM	384KB	320KB
WiFi	WiFi4 - Absent	WiFi6



Output from Useful Sensor PersonSensor

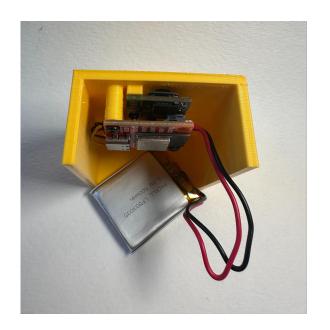
```
"facesDetected": 2,
"detections": [
    "class": "face0",
    "boundingBox": {
    "class": "face1",
    "facing": true,
    "boundingBox": {
```

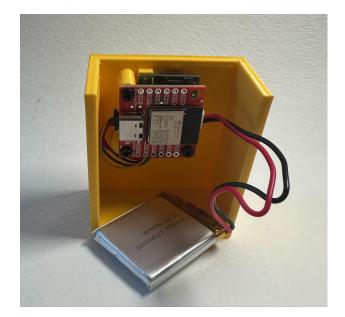
Inference - JSON payload following the format defined in spec with the full output from the Al Model

Insight - An example insight using this data might involve simple business logic calculation that compares the number of faces currently present to an expected number of faces.

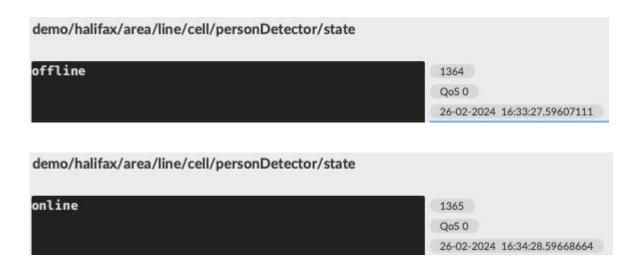












Not part of the EdgeAl spec, but it is always a good practice to include device state in our namespace.



```
demo/halifax/area/line/cell/personDetector/faceDetections/0.0.1/inference

{
    "detections": [],
    "facesDetected": 0
}
```

When no one is in frame, the `detections` array is empty and the `facesDetected` metric reads 0

Report By Exception written into the logic ensures that no repeat messages are published if the payloads would otherwise be identical



demo/halifax/area/line/cell/personDetector/faceDetections/0.0.1/inference

```
"detections": [
    "class": "faceθ",
    "score": 65,
    "facing": false,
    "boundingBox": {
      "x": 112,
      "v": 91,
      "width": 36,
      "height": 62
"facesDetected": 1
```

When a person is within frame the *facesDetected* metric indicates the number of identified faces and the *detections* array contains an object for each detected face.

The face0 object contains a confidence score, the *facing* bool and a nested object describing the bounding box.



Thank You!

Q&A

