Enhancing Data Quality in MQTT-Based IoT Data Pipelines

Webinar

Presented by HIVEMQ

Speakers

Stefan Frehse

Engineering Manager at HiveMQ

stefan.frehse@hivemq.com

https://www.linkedin.com/in/stefan-frehse-96b99421

Michal Piasecki

Product Manager at HiveMQ

michal.piasecki@hivemq.com

https://www.linkedin.com/in/mhpiasecki

Michael Parisi

Product Marketing Manager at HiveMQ

michael.parisi@hivemq.com

https://www.linkedin.com/in/michael-parisi

Agenda

- Importance of Data Quality
- MQTT and Data Quality
- Data Validation via Data Schemas and Data Policies
- Maximize the Business Value of your IoT Data
- Use Cases
- Demo
- Q&A

Importance of Data Quality

Data Quality

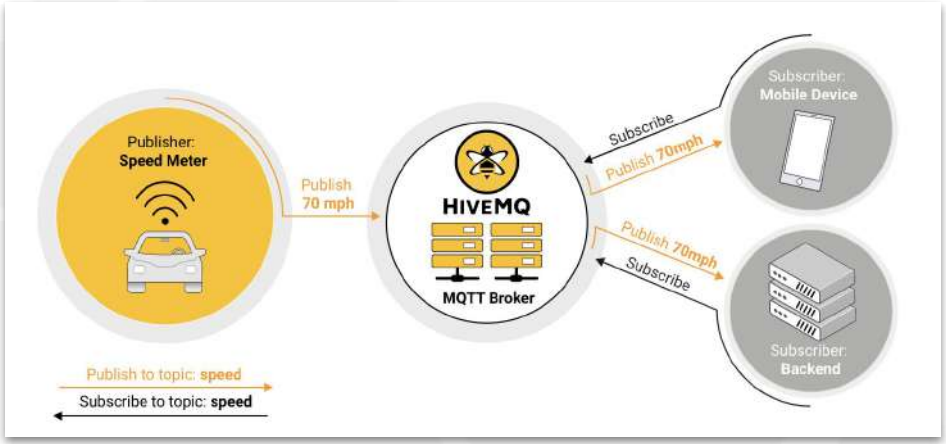
01 Defining Data Quality: Accuracy, Reliability, and Consistency

02 Significance of Data Quality in IoT

03 Role of Data Pipelines in Data Management

MQTT and Data Quality

MQTT & Data Quality



The MQTT protocol is the de facto standard for IoT

MQTT is the backbone of many IoT applications, but why is data quality so crucial in MQTT use cases?

IoT use cases often involve millions of producers and consumers, making client management challenging.

Devices are continuously updated, sometimes without coordination, leading to potential downtimes.

Enforcing behaviors is crucial to keep producers and consumers decoupled.

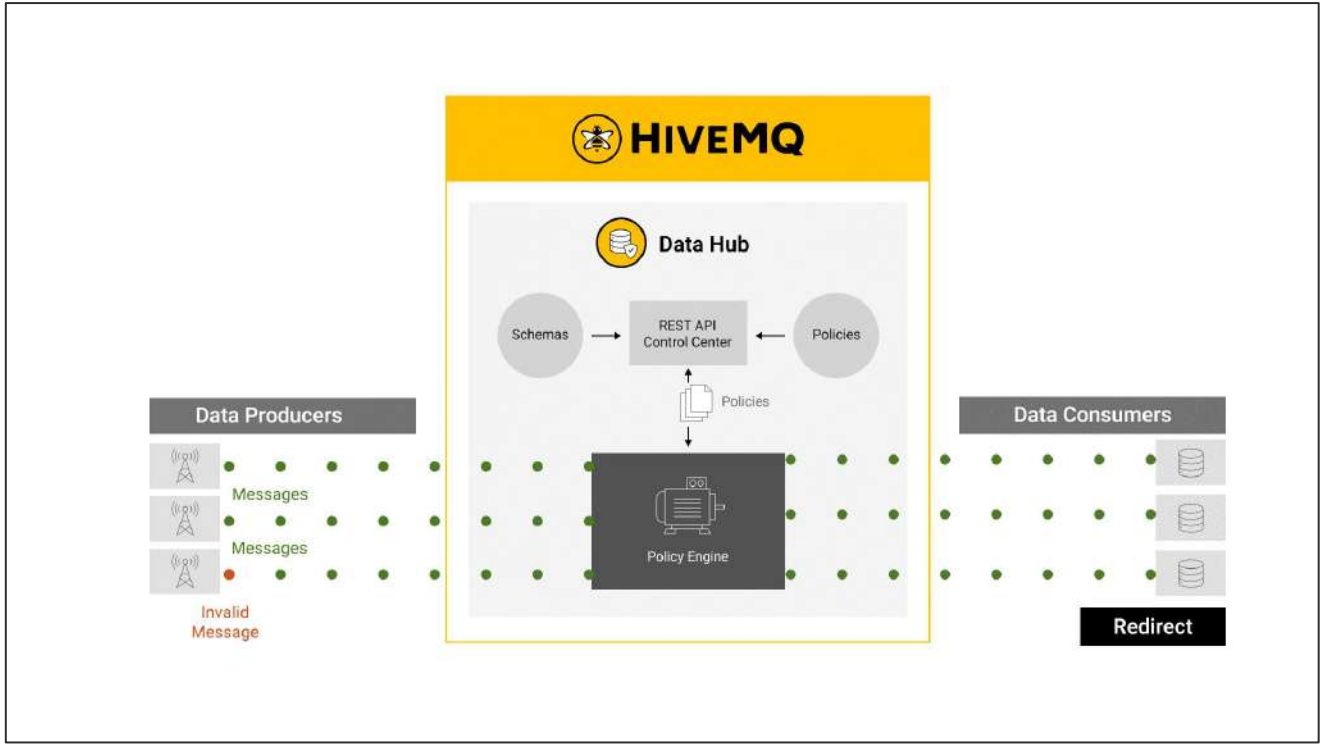
Data Validation via Data Schemas and Policies

Data Validation via Data Schemas and Policies

- MQTT's Data Agnosticism
- Importance of Data Validation
- HiveMQ Data Hub: Centralized MQTT Broker
- Role of Data Schemas and Policies
- Behavior Validation and Transformations

Maximize the Business Value of your Data

HiveMQ Data Hub- How it Works



HiveMQ Data Hub Benefits

- Faster Business Insights
- Increased Data Quality
- Operational Efficiency
- Cost Reduction

Use Cases

Stop bad data!

- What does ‘bad data’ mean?
- How can I identify ‘bad data’?
- What can I do when I identify ‘bad data’?
- How can I leverage this?

Stop bad clients!

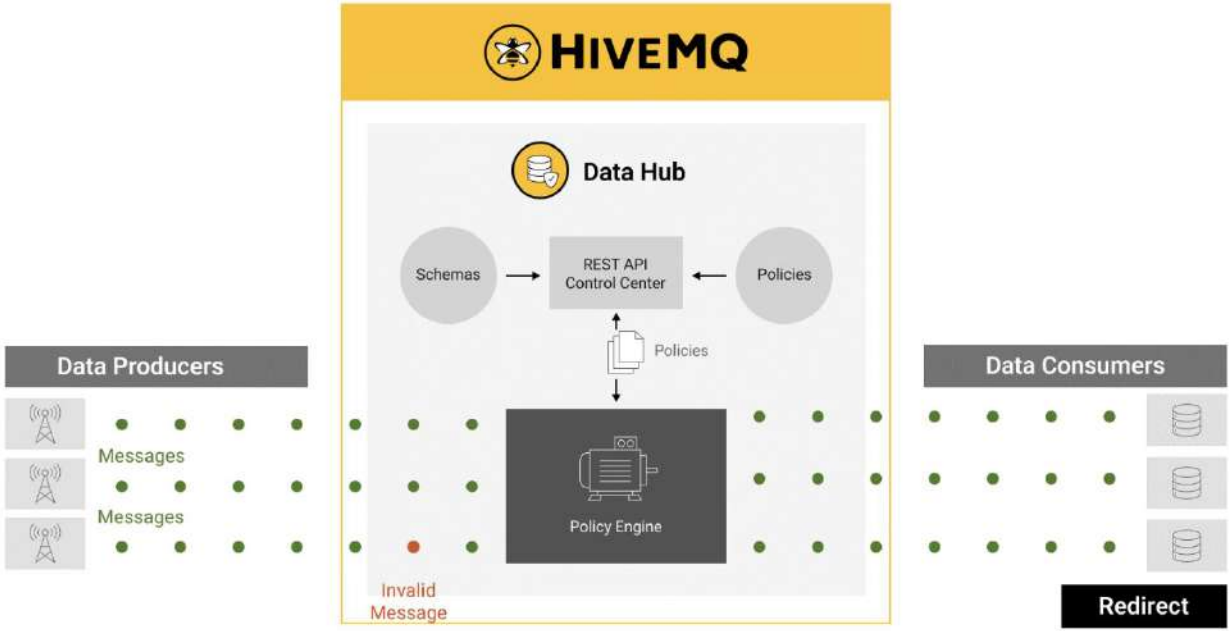
- What do ‘bad clients’ mean?
- How can I identify ‘bad clients'?
- What can I do when I identify ‘bad clients'?
- How can I leverage this? Examples

Demo

Setup

- Measure
- Identify bad data
- Enhance by enforcing

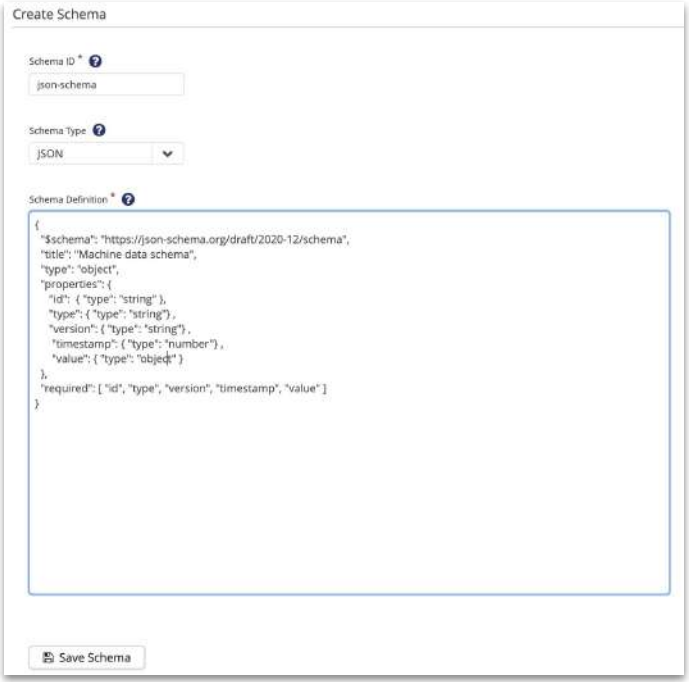
Measure



JSONschema Example

```
{
  "type": "object",
  "properties": {
    "id": {
      "type": "string"
    },
    "type": {
      "type": "string"
    },
    "version": {
    "type": "string"
    },
    "timestamp": {
      "type": "number"
    },
    "value": {
      "type": "object"
    }
  },
  "required": [ "id", "type", "timestamp", "value" ]
}
```

Machine Data Schema



Control Center Schema Edit

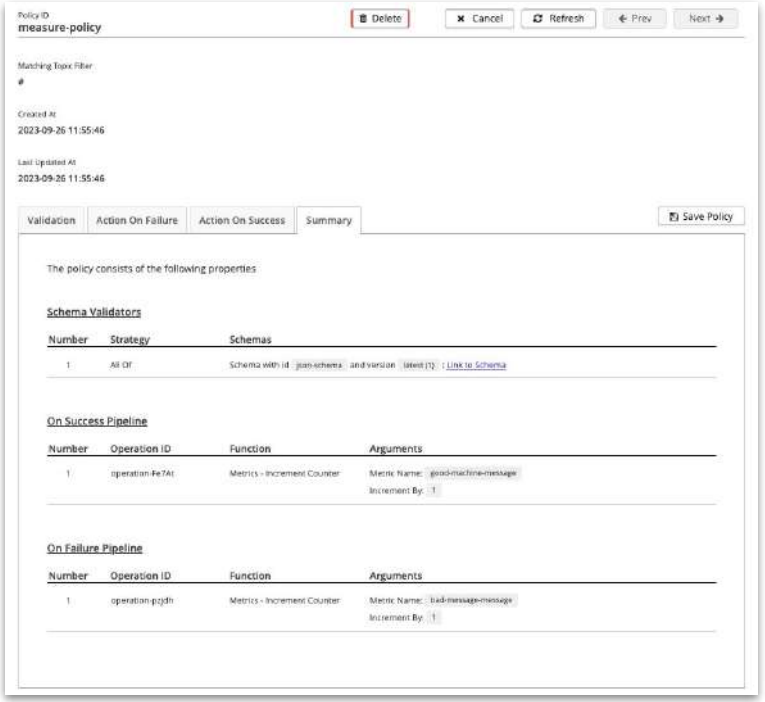
Data Policy Definition

- Define a policy for all incoming data by topicFilter ‘#’
- Use the created schema `machine-schema` to validate
- Increment a metric for valid and invalid data, respectively

Create a data-policy

```
{
  "id": "measure-policy",
  "matching": {...},
  "validation": {...},
  "onSuccess": {
    "pipeline": [
      {
        "id": "good-data-metric",
        "functionId": "Metrics.Counter.increment",
        "arguments": {
          "metricName": "good-machine-messages",
          "incrementBy": 1
        }
      }
    ]
  },
  "onFailure": {
    "pipeline": [
      {
        "id": "bad-data-metric",
        "functionId": "Metrics.Counter.increment",
        "arguments": {
          "metricName": "bad-machine-messages",
          "incrementBy": 1
        }
      }
    ]
  }
}
```

Data policy

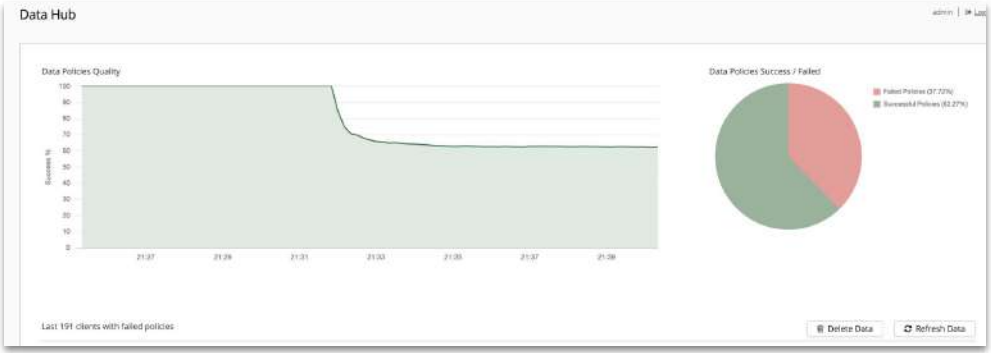


Control Center Data Policy Edit

Data Generation

⅔ good data

⅓ bad data



Stop Bad Data

- Policy Engine provides a function to drop incoming MQTT message directly at the broker: `Mqtt.drop`

Consumers get good data, only 🚀

Policies

Data Policy (demo)

Validates MQTT messages (payloads) and customize actions based on the outcome and drop the invalid data

Behavior Policy

Models the entire client lifecycle and checks against available behavior models

Demo: Generate duplicates

ANY QUESTIONS?

Resources

Blog | [HiveMQ Data Hub: Enhance the Value of IoT Data]

Blog | [Flagging Bad Actors At Scale in IoT and MQTT Deployments]

White Paper | [Measuring the Quality of Your Data Pipeline]

Series | [MQTT Essentials: Learn Basics of MQTT]

We make it easy for you to try HiveMQ

[Sign-up For Free]

[Evaluate for Free]

THANK YOU

Contact

Stefan Frehse

stefan.frehse@hivemq.com

https://www.linkedin.com/in/stefan-frehse-96b99421

Michal Piasecki

michal.piasecki@hivemq.com

https://www.linkedin.com/in/mhpiasecki

Michael Parisi

michael.parisi@hivemq.com

https://www.linkedin.com/in/michael-parisi