



Best Practices: How to Analyze IoT Sensor Data with InfluxDB

Anais Dotis Goergiou

November 2023



Agenda

- The basics of time series data and applications
- A platform overview — InfluxDB, Telegraf, and ecosystem compatibility
- How to start collecting data at the edge and use your preferred IoT protocol (i.e. MQTT)

Anais Dotis-Georgiou

Developer Advocate



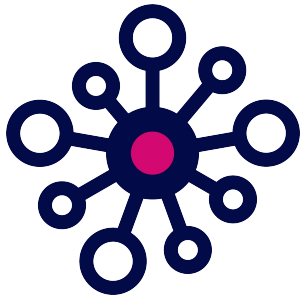
LinkedIn



Time Series Data

What is Time Series Data?

A sequence of data points, typically consisting of successive measurements made from the same source over a time interval.



**Sensors
in the physical world**



**Instrumentation
of the virtual world**

EXAMPLES

- Weather conditions
- Stock prices
- CPU use
- Healthcare Metrics
- Logs
- Traces

Metrics, Events, and Traces

Metrics

Usually derived through sampling, usually numeric, and typically regular in period.

Events

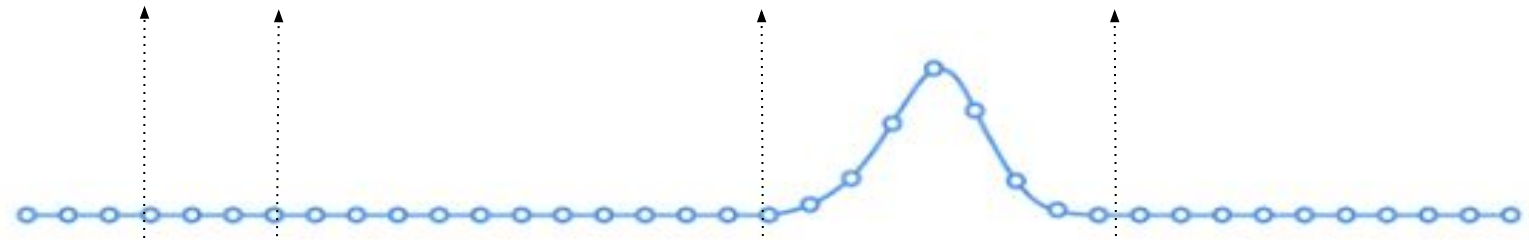
Usually emitted, on-event or on-exception. Can be either numeric or strings. Irregular period by nature.

Traces

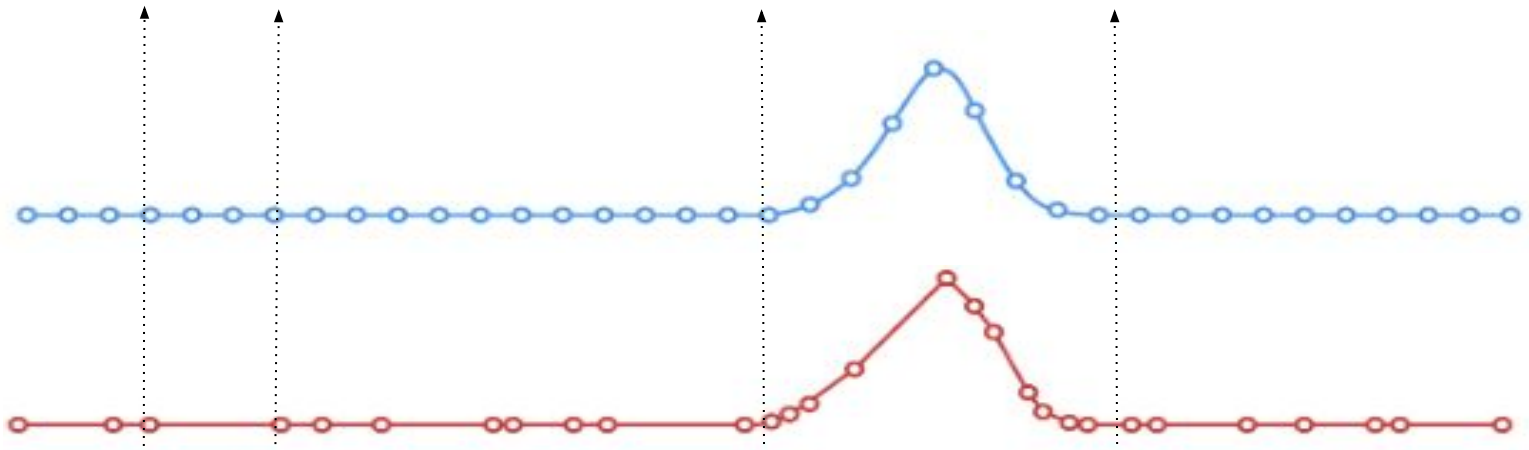
Bundled and uniquely labeled collections of related metrics and events related to a specific transaction or interaction. Irregular period and explicit duration.

How They Relate

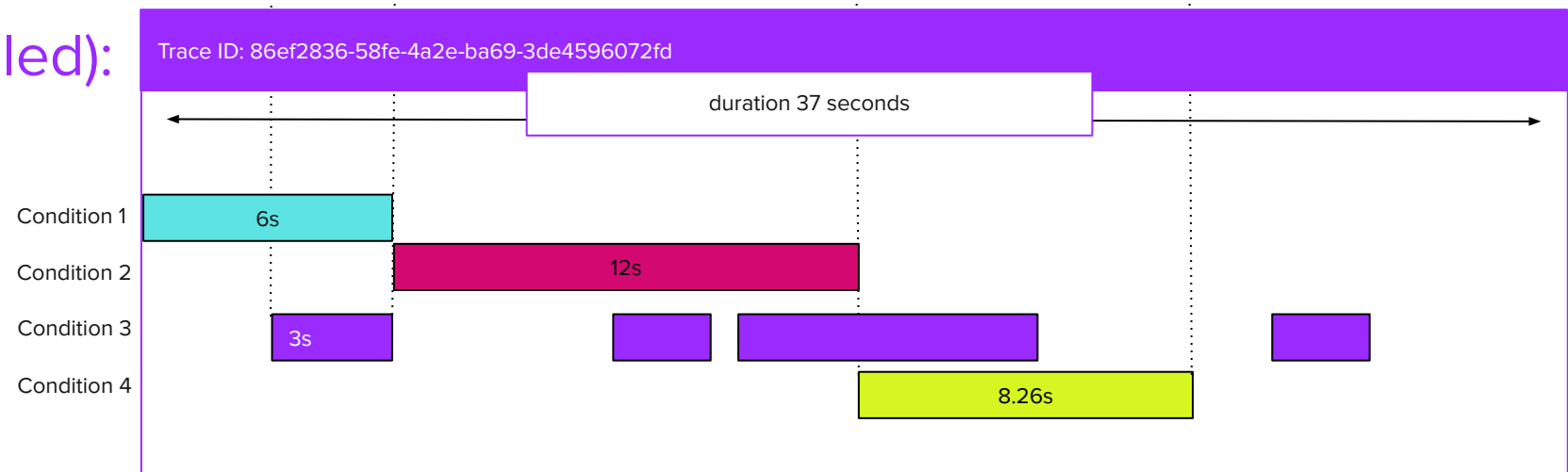
Metrics (one/sec):



Events (varied):



Traces (bundled):



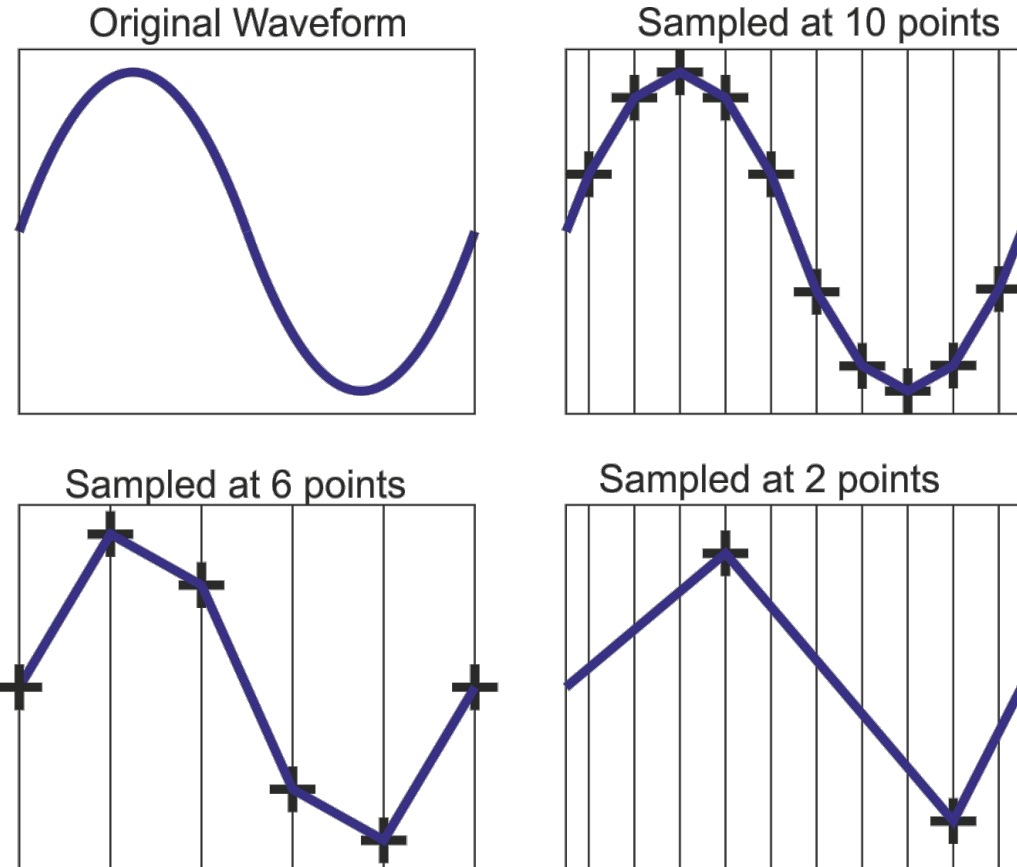
Timestamp Precision

- Relative time - :since epoch:
- Ideal to create early and maintain throughout pipeline
- Must consider implications of aggregation vs truncation
- Use cases: event ordering, correlation, time-bounded analytics

Precision	Description	Example
ns	Nanoseconds	1577836800000000000
us	Microseconds	1577836800000000
ms	Milliseconds	1577836800000
s	Seconds	1577836800

Data Granularity

- Loosely coupled to precision
- Also referred to as “sample rate”
- Refers to the number of discrete samples per series per window of time
- Reduce granularity through aggregation and roll-ups
- Ideally retain data shape through advanced post processing like SDA or Holt-Winters or raw-retention



Key Drivers for Time Series Applications

ACCESS

Time Series Data
from Assets and
Applications

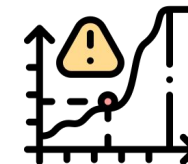


ANALYZE

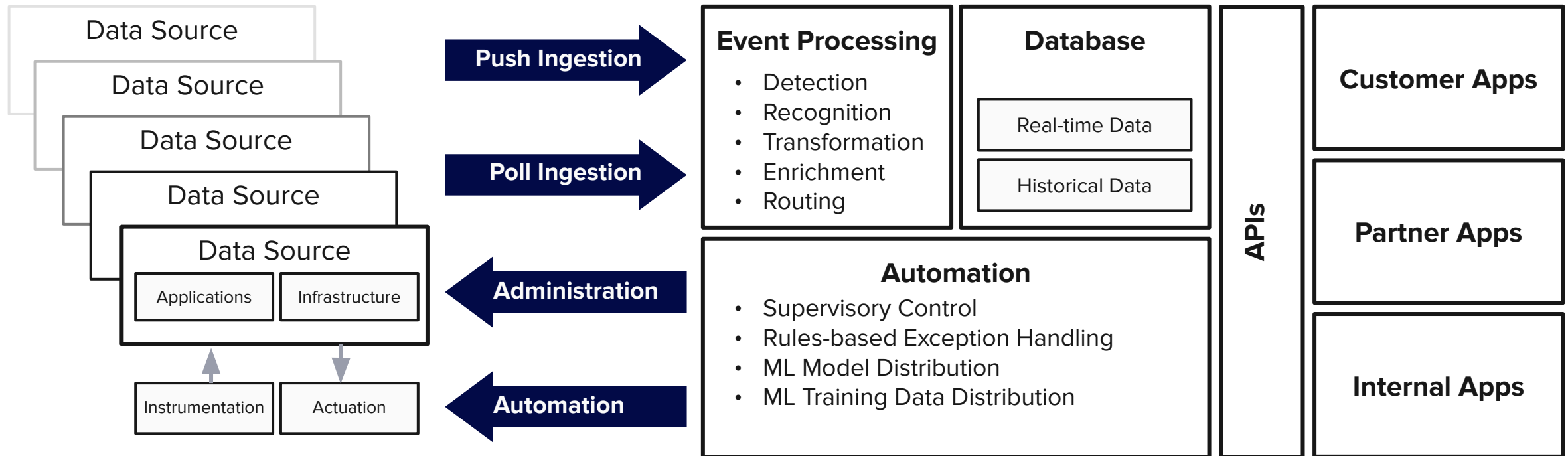
Performance,
Availability, and
Security

ACT

Improve and Expand
the Business



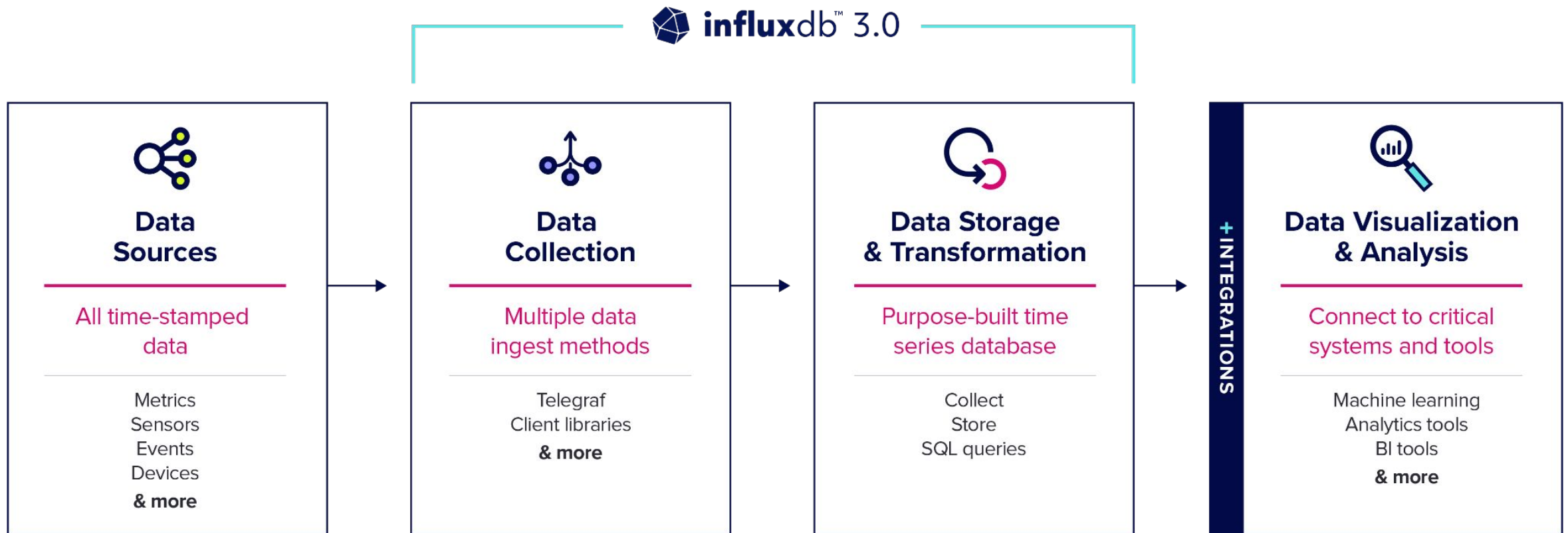
Key Components of Time Series Applications




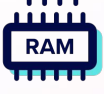



Time Series Data in InfluxDB

InfluxData Reference Architecture

InfluxDB Platform



InfluxDB's new storage engine is built on

-  Rust
-   Apache Arrow
-  Apache Parquet
- Arrow Flight
-  DataFusion

SQL and InfluxQL Support

The screenshot displays the InfluxDB SQL editor interface. At the top, the title "Basic SQL" is visible. Below the title, there are buttons for "+ New Script", "OPEN", "SAVE", and "EDIT". On the right side, there is a "Try New Script Editor" toggle and a "Provide Feedback" link.

The main area is divided into three sections:

- Left Panel:** Contains a "Bucket" dropdown menu set to "anais-iox" and a "Measurement" dropdown menu set to "Select measurement...".
- Center Panel:** A code editor containing the following SQL query:

```
1 SELECT "temperature", "sensor_id", "time" FROM "airSensors"  
2 where time >= ('2022-12-01 19:05:41.000')::TIMESTAMP  
3 and time < now()::TIMESTAMP and sensor_id = 'TLM0100'
```
- Right Panel:** A configuration panel for the graph. It includes:
 - Data:** A section for configuring the data source.
 - X Column:** A dropdown menu set to "time".
 - Y Column:** A dropdown menu set to "temperature".
 - Adaptive Zoom:** A toggle switch that is currently turned on.
 - Options:** A section for additional graph options.
 - Time Format:** A dropdown menu set to "YYYY-MM-DD HH:mm:ss".
 - Interpolation:** A dropdown menu set to "Linear".
 - Line Colors:** A dropdown menu set to "Nineteen Eighty Four".
 - Hover Dimension:** A dropdown menu set to "auto".
 - Shade area below graph:** A toggle switch that is currently turned off.
 - X-Axis:** A section for X-axis configuration.
 - Generate X-Axis Tick Marks:** A button to generate X-axis tick marks.

Below the code editor, there is a status bar showing "Ready (143ms)" and buttons for "CSV" and "RUN".

The bottom section of the interface shows a graph visualization. The graph is titled "Graph" and has a "CUSTOMIZE" button. The graph displays a line chart with the following data points (approximate values):

Time	Temperature
2022-12-01 13:15:00	71.2
2022-12-01 13:30:00	71.8
2022-12-01 13:45:00	72.0
2022-12-01 14:00:00	72.2

Massive growth of instrumentation data

By 2025, data creation will grow to more than

180 zettabytes*

(An increase of 118.8 zettabytes since 2020)

1 Zettabyte = 1 Billion Terrabytes = 1 Trillion Gigabytes
= 1,000,000,000,000,000,000,000 bytes

Benchmarks for InfluxDB 3.0

InfluxDB 3.0 is up to 45x Faster for Recent Data Compared to InfluxDB Open Source

Dataset for Benchmark:

- Dataset duration: 24 hours
- Measurement interval: 10 seconds
- Cardinality: 160,000



Benchmarks for InfluxDB 3.0

InfluxDB 3.0 provides

45x

Better write
throughput

90%

Reduction in
storage costs

100x

Faster queries for
high cardinality data

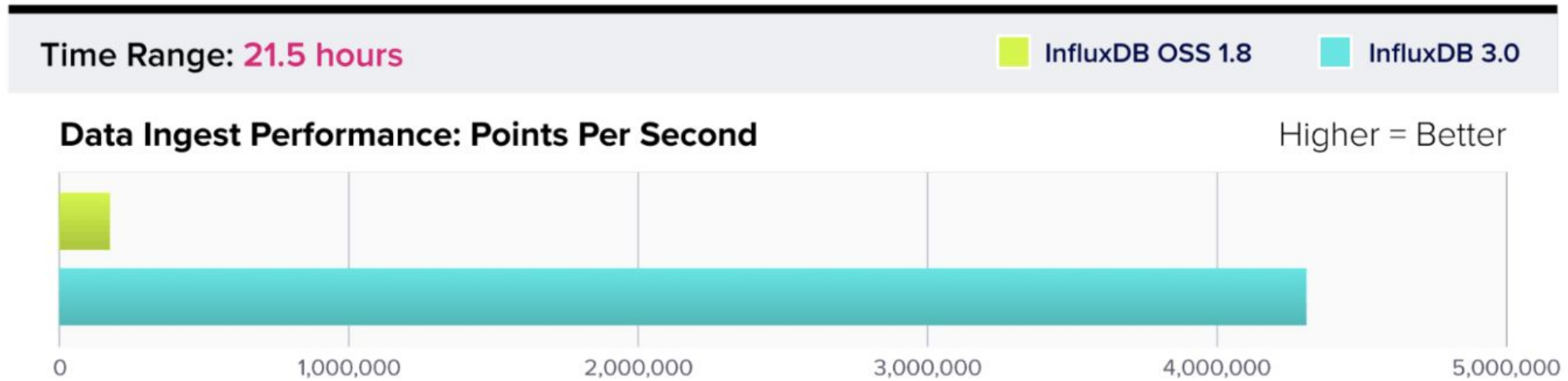
45x

Faster queries
for recent data

Compared to InfluxDB OSS

Data Ingest Performance

Results represent 21.5 hours of metrics reported from varying load of Telegraf instances.

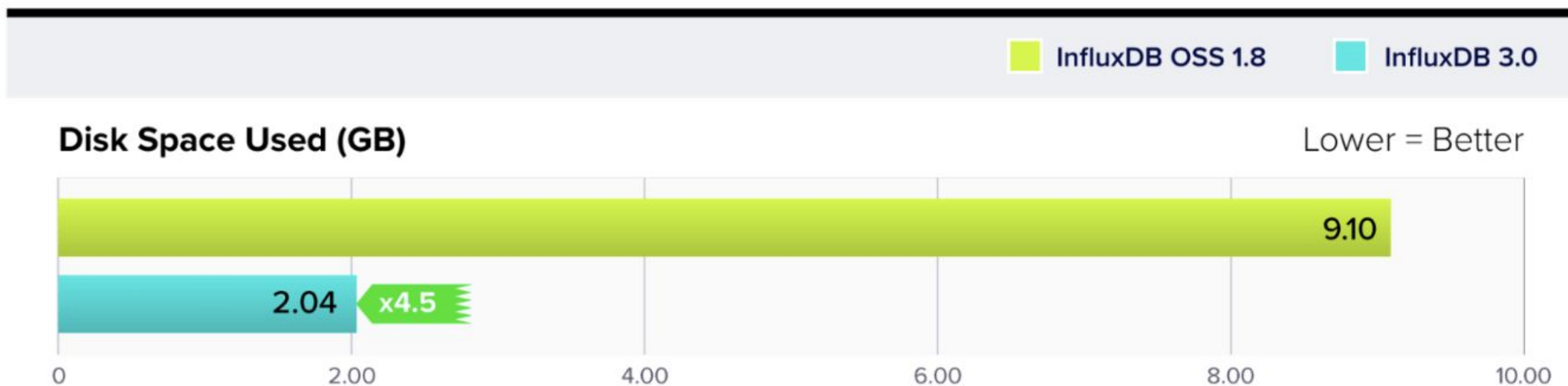


Ingest Vector	InfluxDB OSS 1.8	InfluxDB 3.0
Total Telegraf Clients	100	4,500 x45
Rows / Hour	8,859,600	329,206,806
Rows / Second	2,461	91,446
Values / Second	162,264	4,310,764

Storage performance

When it comes to data compression, InfluxDB 3.0 outperformed InfluxDB OSS by roughly **4.5x**. The cost saving is maximized because InfluxDB 3.0 uses object storage which is cheaper than the SSD (Solid State Devices) based storage used by InfluxDB OSS.

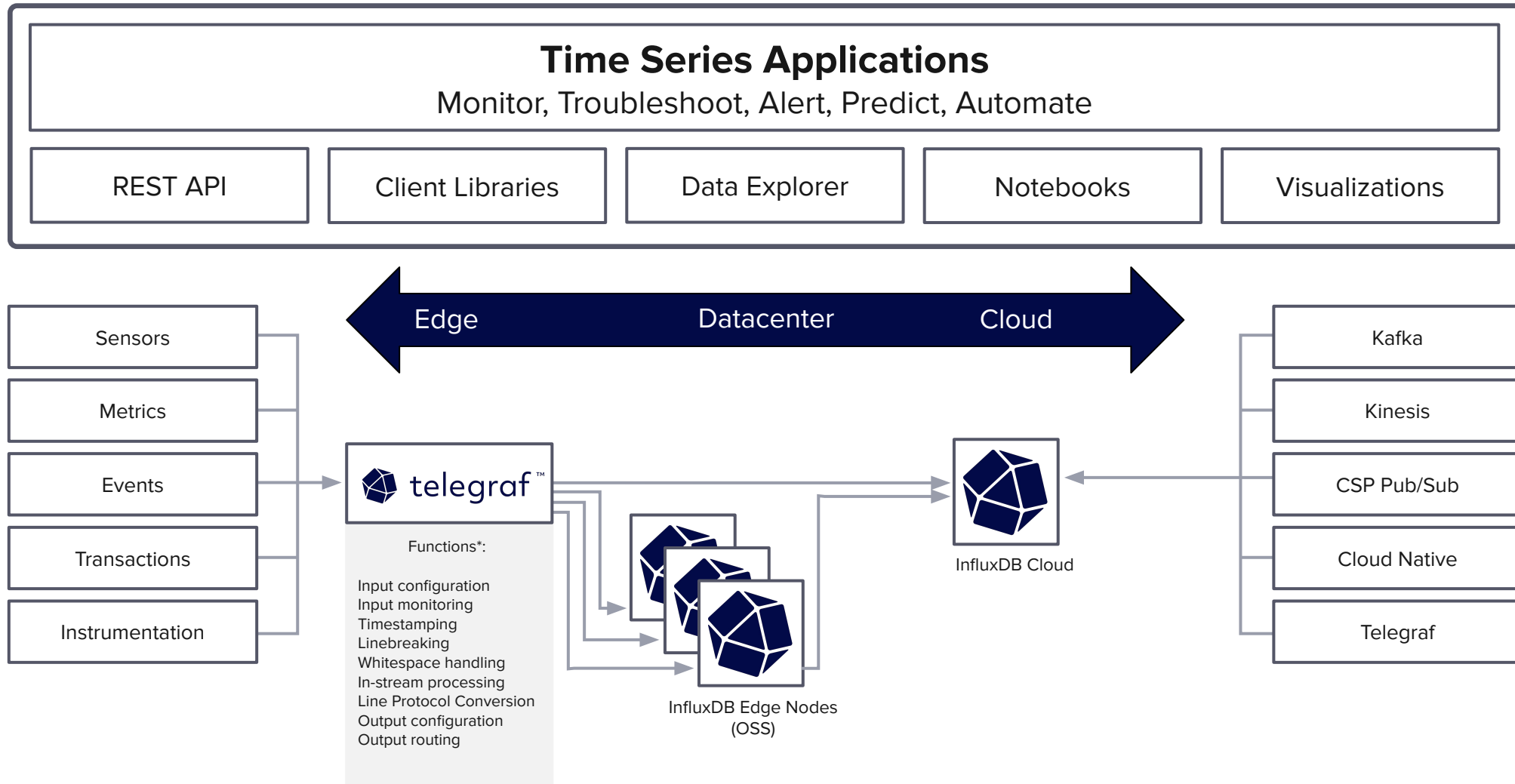
Storage Size on Disk



You have probably used
InfluxDB



Functional Architecture



Line Protocol

ALL time series data is written to InfluxDB using Line Protocol, and uses the following format:

```
<measurement> [ , <tag-key>=<tag-value> ] [ <field-key>=<field-value> ]  
[unix-nano-timestamp]
```

Measurement	Tag Set	Field Set	Timestamp
cpu_load,	hostname=server02, us_west=az	temp=24.5, volts=7	1234567890000000

Where data is formatted as line protocol is user and use case specific. Telegraf provides significant shortcuts for many popular data sources and formats, and new Cloud-Native Data Connectors provide JSONPath and Regex parsing solutions to work with more unstructured data. Customers often build telemetry pipelines specifically for LP delivery to InfluxDB

Fields and Tags

sample time: Friday July 15 2022 15:40:50 GMT

machine id: SN0039992 (TAG)

lab id: MEDF-MA-US01 (TAG)

patient id: 123456789A (TAG)

sample volume: 23.24 (FIELD)

sample units: ml (TAG)

cell count: 234 (FIELD)



- **Tags** are like labels, designed to further specify and disambiguate similar signals. They are indexed for fast seek, filtering, and grouping.
- **Fields** are the primary numerical and other values to be monitored. They are not indexed as stored values are highly variable.
- **Tags** are best for metadata, **Fields** are best for samples

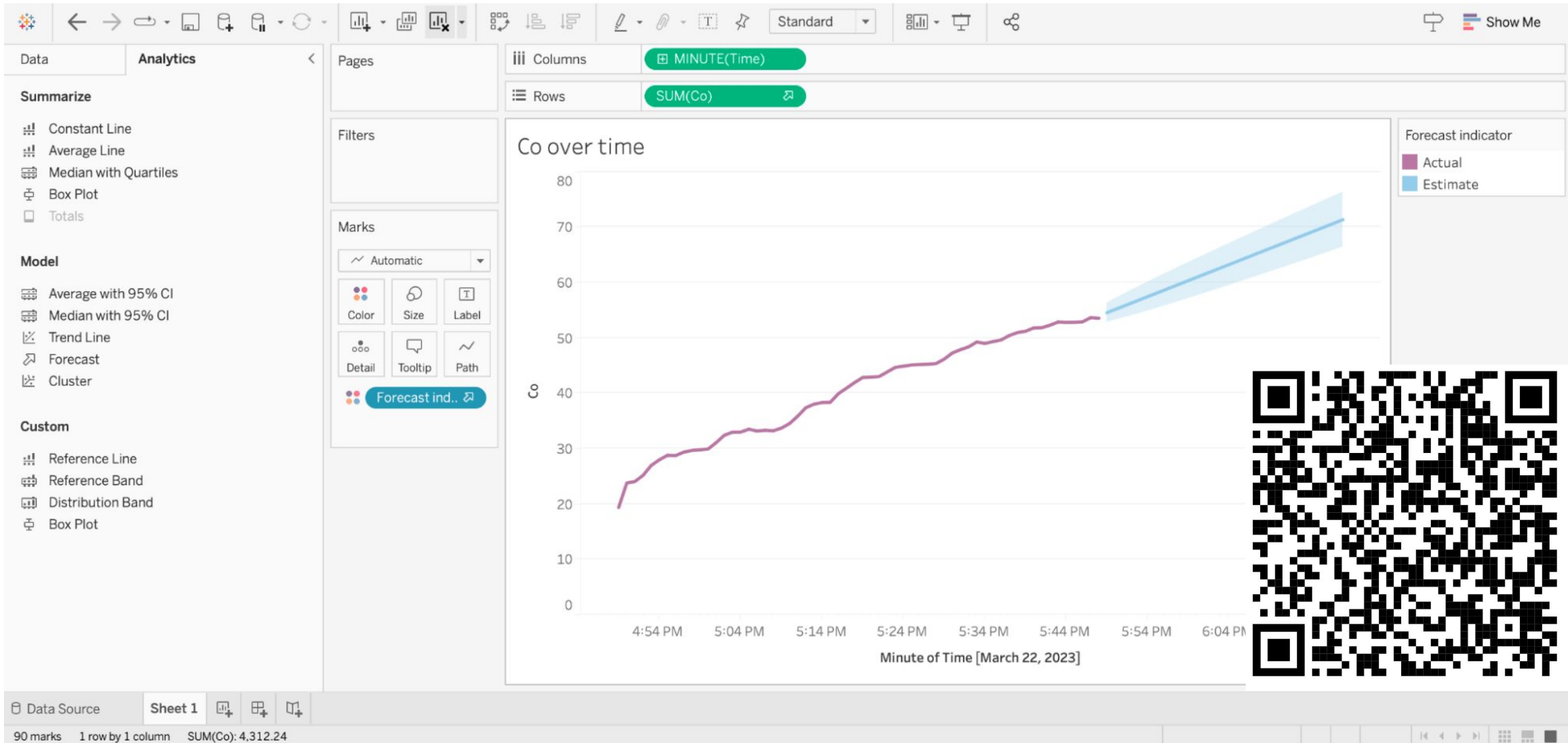
Interoperability with InfluxDB 3.0

Interoperability

- Client Libraries:
 - <https://docs.influxdata.com/influxdb/cloud-serverless/reference/client-libraries/v3/>



Interoperability: Visualization Tools



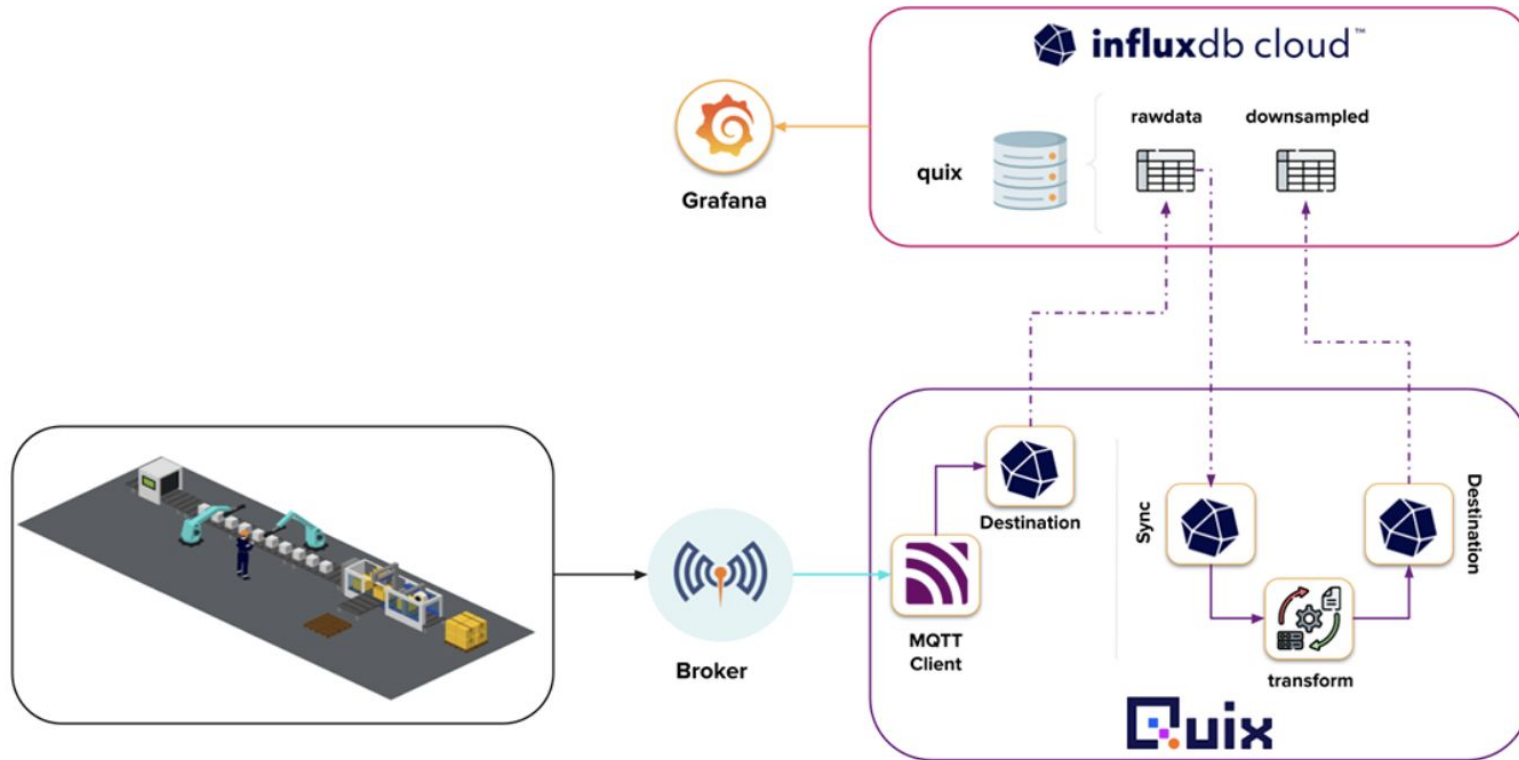
Projects

Mage and InfluxDB

```
#####  
#                               INPUT PLUGINS                               #  
#####  
  
[[inputs.mqtt_consumer]]  
  ## Broker URLs for the MQTT server or cluster. To connect to multiple  
  ## clusters or standalone servers, use a separate plugin instance.  
  ##   example: servers = ["tcp://localhost:1883"]  
  ##           servers = ["ssl://localhost:1883"]  
  ##           servers = ["ws://localhost:1883"]  
  #servers = ["tcp://telegraf_mqtt_mosquitto_1:1883"]  
  servers = ["tcp://mosquitto:1883"]  
  ## Topics that will be subscribed to.  
  topics = ["machine/#"]  
  qos = 2  
  ## Connection timeout for initial connection in seconds  
  connection_timeout = "30s"  
  data_format = "json_v2"  
  
[[inputs.mqtt_consumer.json_v2]]  
  measurement_name = "machine_data"  
  [[inputs.mqtt_consumer.json_v2.object]]  
    path = "@this"  
    disable_prepend_keys = true  
    tags = ["metadata_machineID", "metadata_provider"]  
    excluded_keys = ["metadata_barcode"]
```



Quix



IoT Use Cases

- Industrial IoT 
- Enterprise IoT 
- Consumer IoT 

InfluxDB customers in IoT

Industrial



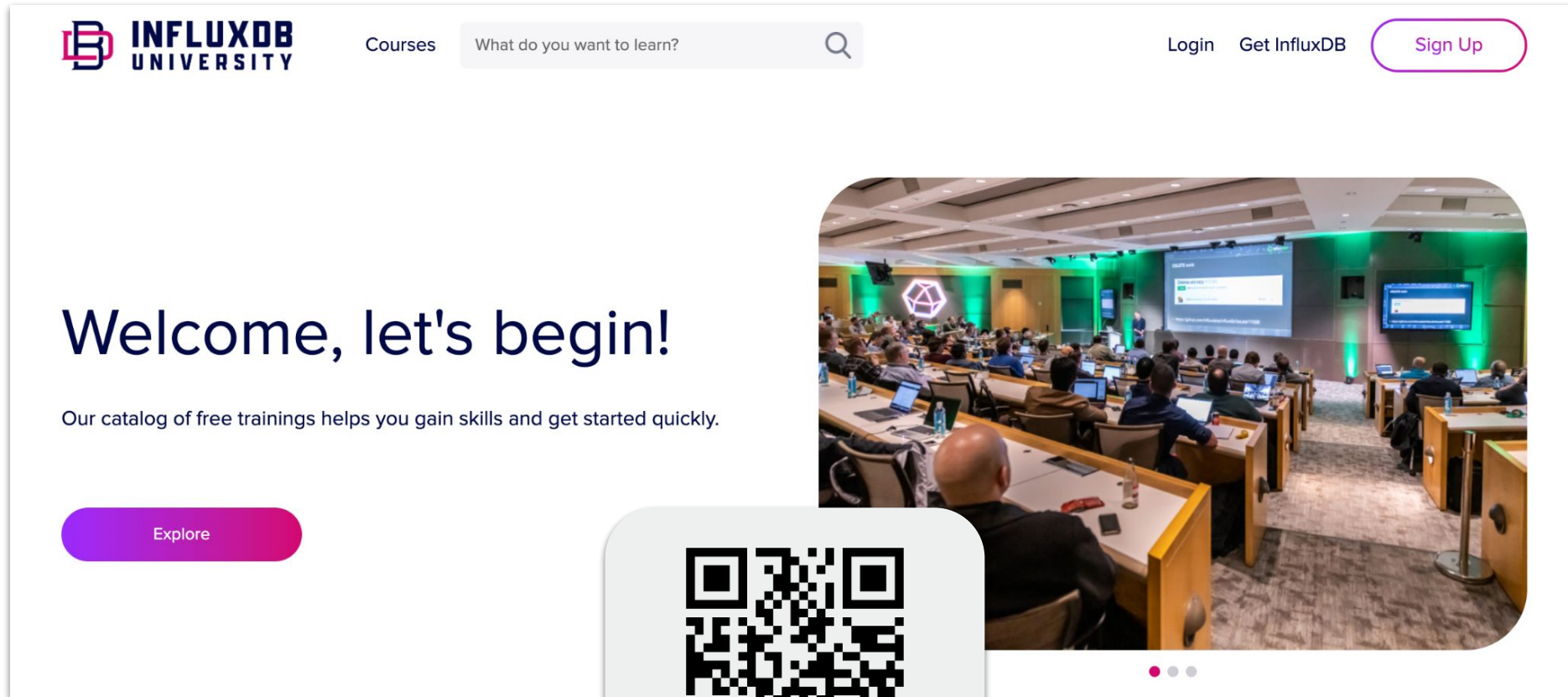
Consumer



Resources

- InfluxDB Community: <https://github.com/InfluxCommunity>
- Mage Demo:
 - https://github.com/InfluxCommunity/Mage_Demo
 - <https://www.influxdata.com/blog/mage-anomaly-detection-influxdb-half-space-trees/>
- Quix Demo:
 - <https://github.com/InfluxCommunity/quix-anomaly-detection-example>
 - <https://www.influxdata.com/resources/simplify-stream-processing-with-python-quix-and-influxdb/>
 - <https://www.influxdata.com/blog/quix-community-plugins-influxdb-build-streaming-task-engine/>
- Docs: <https://docs.influxdata.com/influxdb/cloud-serverless/>

Join InfluxDB University



The screenshot shows the InfluxDB University website. At the top left is the InfluxDB University logo. To its right is a 'Courses' link and a search bar containing the text 'What do you want to learn?'. Further right are links for 'Login', 'Get InfluxDB', and a 'Sign Up' button. The main content area features the heading 'Welcome, let's begin!' followed by the text 'Our catalog of free trainings helps you gain skills and get started quickly.' Below this is an 'Explore' button. On the right side of the main content is a large image of a conference room with people seated at desks, facing a stage with a presentation screen. A QR code is overlaid on the bottom left of the main content area.



influxdbu.com

InfluxDB Resources

Webinar: Gain Better Observability with OpenTelemetry and InfluxDB

Leverage OpenTelemetry and InfluxDB to collect and analyze metrics, logs, and traces, enabling better anomaly detection, root-cause analysis, and alerting.

Watch now

bit.ly/3qhemCw

Save 96% on Data Storage Costs:

Learn more

bit.ly/3NJEcGZ

Run a Proof of Concept:

Learn more

bit.ly/3puRsal





THANK YOU