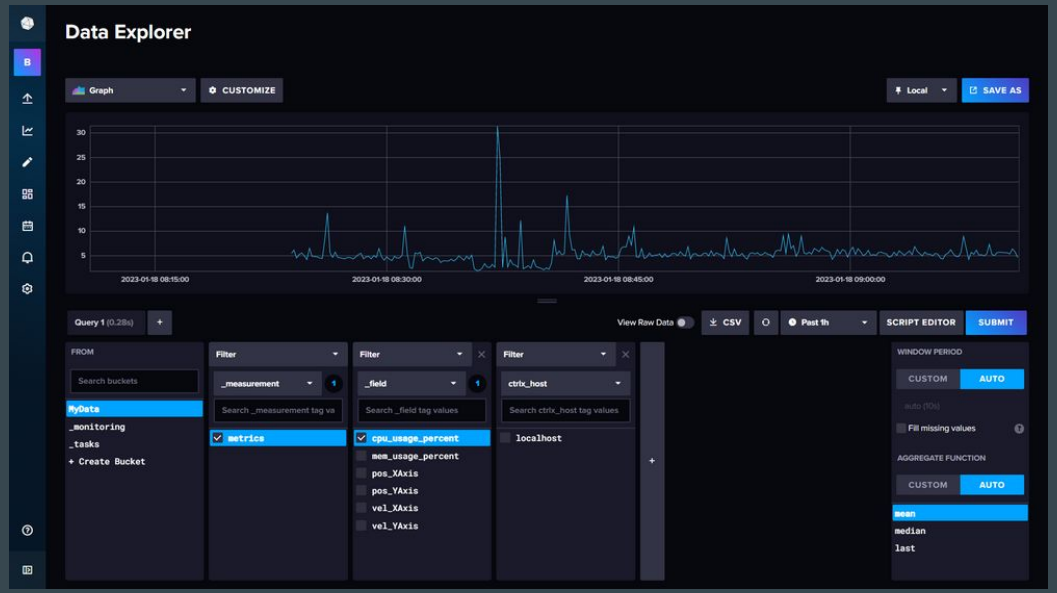
Diving into InfluxDB: Time Flow Analysis. Past, Present, and Future!

•••

Xi Chen, James D'Elia, Kaydence Lin, Kaito Minami, Yidi Wang
DS 4300: HW 6 mini-project

What is InfluxDB?

- Time Series Data: sequence of data points ordered chronologically
- Open Source Time-series Database
- Flux: time-series data query language
- Collect, store, process, and visualize time-series data
- Has cloud storage and services primarily for business customers
- Monitoring, analytics, and IoT applications



Use Cases

- **Predictive Analysis**: use historical data to predict future outcomes with algorithms and machine learning
- **Server or Machine Performance Metric**: keep track and predict breakdowns
- **Stock Market**: detect stock manipulation or predict stock prices
- **Health Care**: detect unusual patterns or predict changes in a person's health
- **Space Industry**: find optimal satellite communication, plan space exploration, detect changes in Earth's climate
- **Retail and E-Commerce**: determine which products should be marketed, shopping patterns during a certain time or season

History of InfluxDB

The InfluxDB was initiated by a company Errplane, in an attempt to create open source time-series database platform.

The Errplane was real-time metric and monitoring SaaS on Cassandra, which everybody else was also doing at the time.

2013 - InfluxDB announced. Given talks internationally (Kyoto, Sydney, Cologne, Dublin...)

2014 - InfluxDB 1.0 release

2015 - Company changed name to Influxdata Inc.

2016 - released a cloud offering

2018 - Flux, a query language, is introduced.

Who uses InfluxDB?

Key features

Database

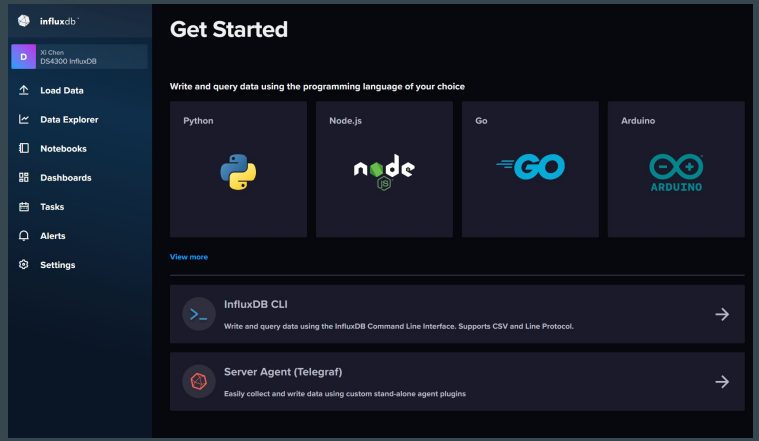
- Contains buckets
- Bucket
  - Contains measurements (tag, field, timestamp)
  - Tags: key value that does not change over time (ex. host, location. station)
  - Fields: key value that change over time (ex. temperature, pressure, stock prices)
  - Timestamp
    - Sorts data chronologically

Point: a data point with tag, field, timestamp

Series: a group of points with same tag key, tag values

Tutorial and Demo

1. Install/Run InfluxDB Instructions on https://docs.influxdata.com/influxdb/v2/install/?t=Windows
2. Open up http://localhost:8086/ in your local browser
3. Load Data & Visualize in InfluxDB UI



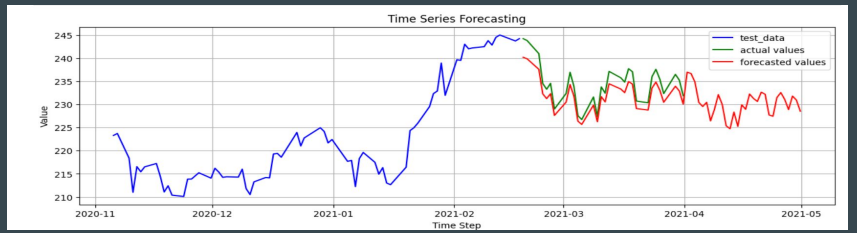
4. Use InfluxDB client library to retrieve data from InfluxDB

```
[3]:
# You can generate a Token from the "Tokens Tab" in the UI
# replace your token, org, bucket here
token = "gQObTKL-jd4gKC0ShdiaB5loHG3JWlqQX4elYomvWj77uzZk7RRGN-oOfa_r97kp-QTJ2YYgCOW467BppoB9Pw=="
org = "DS4300 InfluxDB"
bucket = "DS4300 HW6"

## connect to influxdb
client = InfluxDBClient(url="http://localhost:8086", token=token, org = org)

[4]:
# read the data
query_api = client.query_api()
tables = query_api.query("""from(bucket:"DS4300 HW6") |> range(start: -20y)
|> filter(fn: (r) => r["company"] == "Microsoft")
|> filter(fn: (r) => r["_measurement"] == "Stock_Price")
|> filter(fn: (r) => r["_field"] == "close")""")
```

5. Use Pytorch to build LSTM model to forecast close price for Microsoft's Stock



Strengths and Weaknesses

Strengths

- Horizontal/Elastic Scalability
- Displays a graph along with queries
- Simple UI with dropdown boxes
- Query automation (Tasks)
- Database as a Service (DBaaS)
- Data anonymization
- Granularity

Weaknesses

- No IDs for rows
- No duplicate data
- Restrictive update and delete permissions
- Querying continuously updating databases may not result in most up-to-date data

Conclusion and Takeaways

- InfluxDB excels in managing and analyzing time-series data, offering robust scalability and ease of use
- Suitable for projects where time-series data is crucial, with extensive support for data visualization and analysis.
- Provides a versatile platform for predictive analytics, monitoring, and real-time data analysis across various industries
- Active community and clear licensing terms support both open-source and commercial projects, ensuring a resourceful environment for developers.