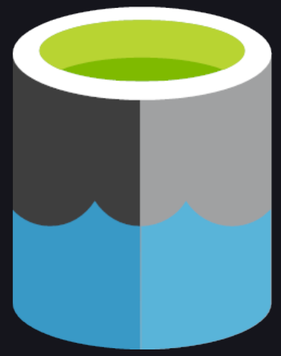
What

Taught me about



On

Lisa Hoving

Data Engineer at Rabobank

@Lizaard08

Premium sponsors

dataon
Power your data

twoday

exmon

Fellowmind

Standard sponsors

devoteam

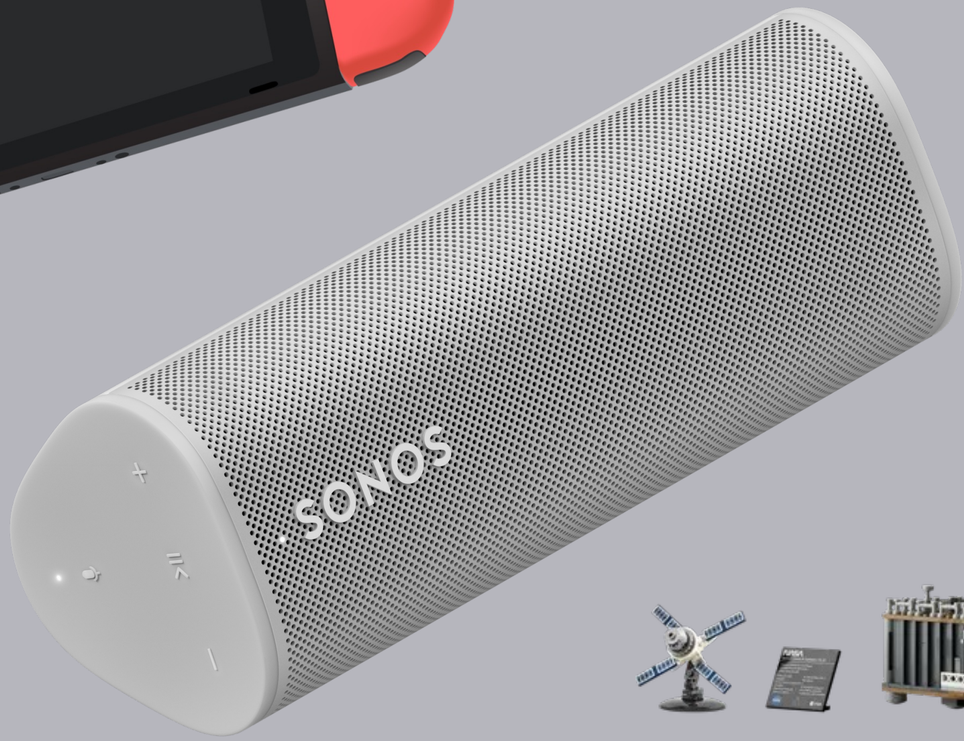
INVIXO

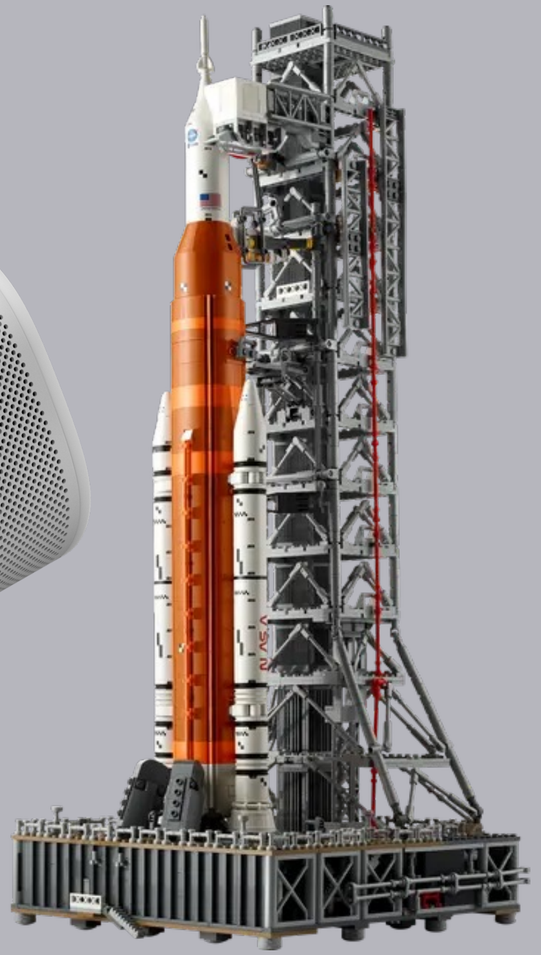
INSPARI
a valantic company

backstage

BIWISE
BASED ON INTELLIGENCE

Join the raffle!





drawing
body language
talking
writing
AAC devices
signs
All forms of communication are welcome
communication cards
sounds
and more...
@autism_happy_place

01 A python? Why?

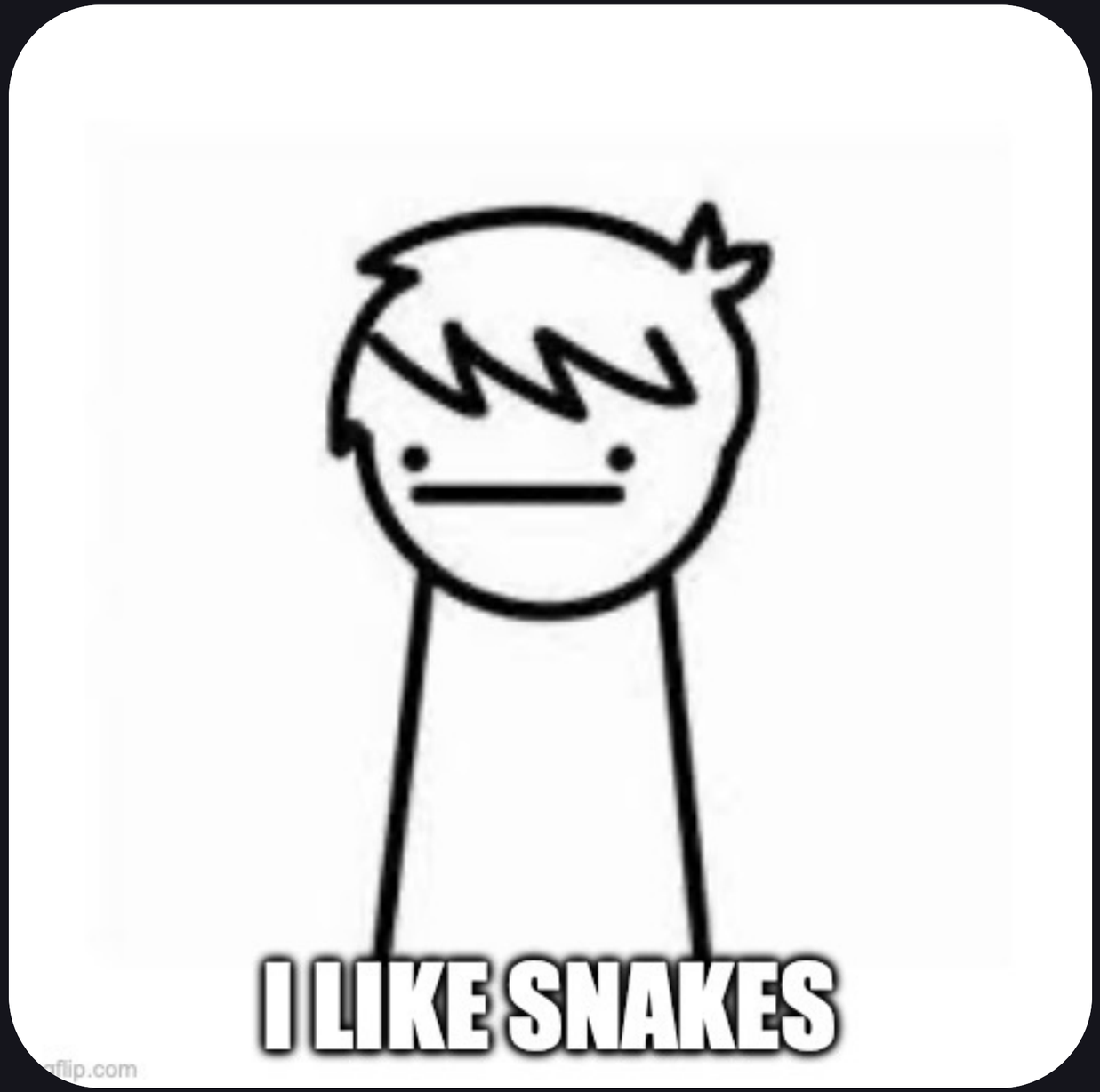
02 What is a Data Lake?

03 Lessons learned

04 Conclusion

01

A python, why?


I LIKE SNAKES

Good idea’s start on the internet

Source: Clint’s Reptiles

Snake DISCOVERY

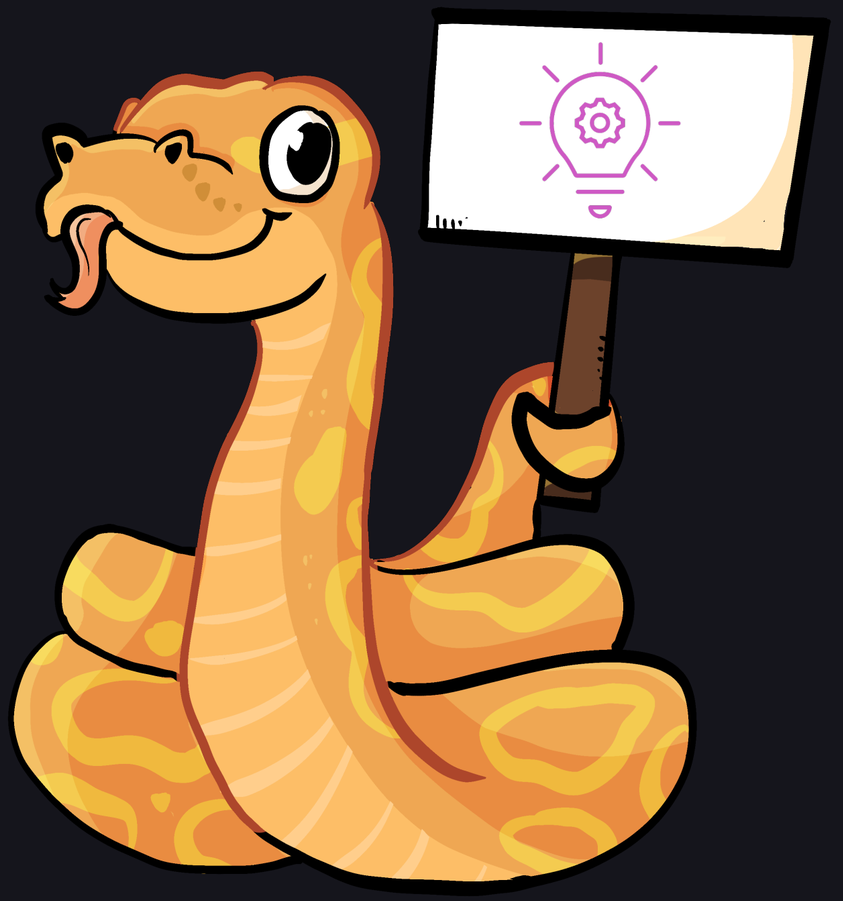
Good idea’s start on the internet

2019

- Bought a pet python
- Fluffy

Good care

- Technical knowledge required
- Not that easy!



Monitoring

What if technology fails?

- Alexa
- 2 wireless temperature sensors
- Home Assistant cloud

Call me when

- Temperature is off

Home Assistant Community Store

Search

Filtering by Downloaded CLEAR

Repository Name	Downloads	Stars	Updated	Category
HACS HACS gives you a powerful UI to handle downloads of all your custom needs.	211434	4427	last week	Integration
Alexa Media Player This is a custom component to allow control of Amazon Alexa devices in Home Assistant using the	23000	1267	last week	Integration
pyscript Pyscript adds rich Python scripting to HASS	16096	751	3 weeks ago	Integration

When

Trigger on the 1st second of every minute of every hour

+ ADD TRIGGER

And if (optional)

This list of conditions needs to be satisfied for the automation to run. A condition can be satisfied or not at any given time, for example: 'If lizaard08 is home'. You can use building blocks to create more complex conditions.

+ ADD CONDITION + ADD BUILDING BLOCK

Then do

Pyscript Python scripting 'pyscript.call_iothub'

Visual editor is not supported for this configuration

- Templates not supported in visual editor

You can still edit your config in YAML.

```yaml
service: pyscript.
metadata: {}
data:
  sensor1: "{{ states('sensor.temp_1_temperature') }}"
  sensor2: "{{ states('sensor.temp_2_temperature') }}"
  connection_string: >-

```

+ ADD ACTION + ADD BUILDING BLOCK

Exploring Data Lakes

2 data points every 2 minutes means:

- 60 per hour
- 1440 per day
- 525.600 per year

Do we want to waste all that data?

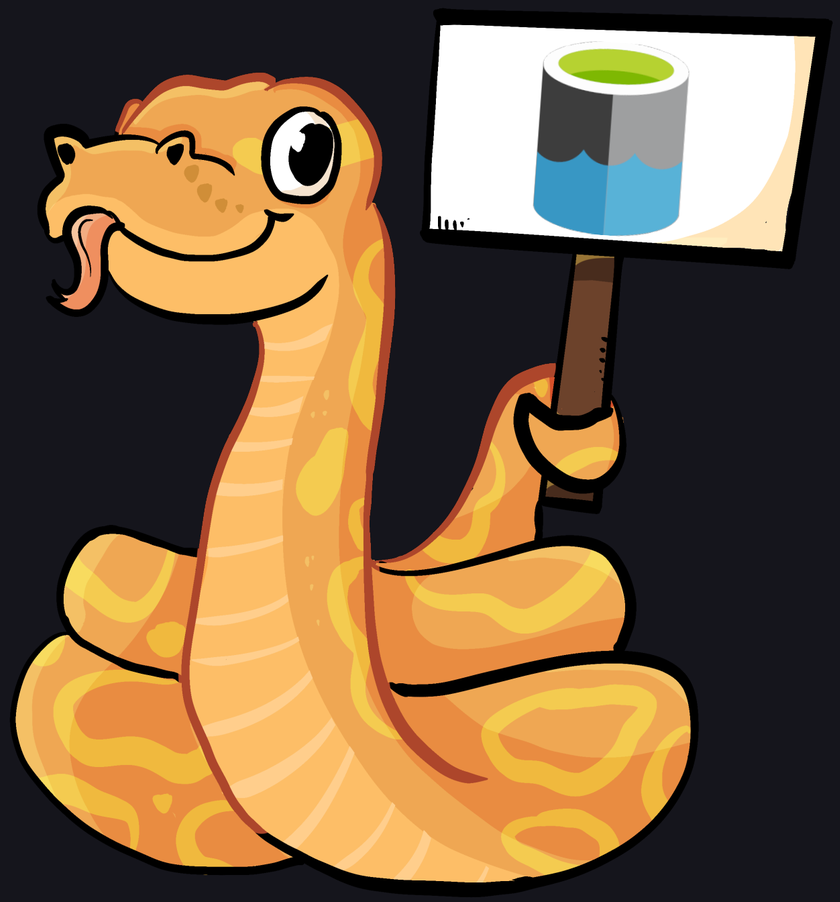
Exploring Data Lakes

What if we could

- Check trends

Platform requirements

- Flexible → more sensors
- Scalable → more terraria



02

What is a Data Lake?

Whatever,
Whenever,

Data & Lakes are meant to be together

https://en.wikipedia.org/wiki/Whenever,_Wherever

Whatever,
Whenever,

Data & Lakes are meant to be together

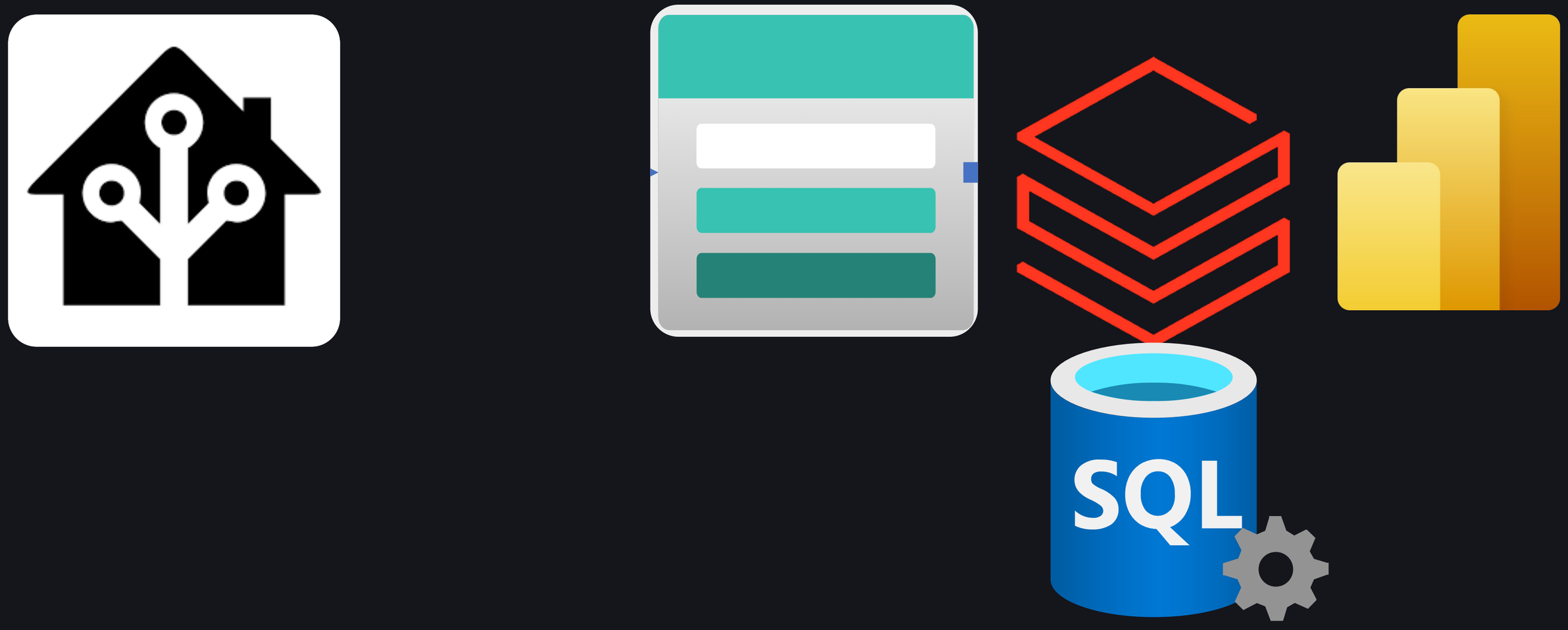
What is a Data Lake?

- Cost effective
- Flexible
  - Any type of storage format
- Scalable
  - Any amount of data

03

Lessons learned

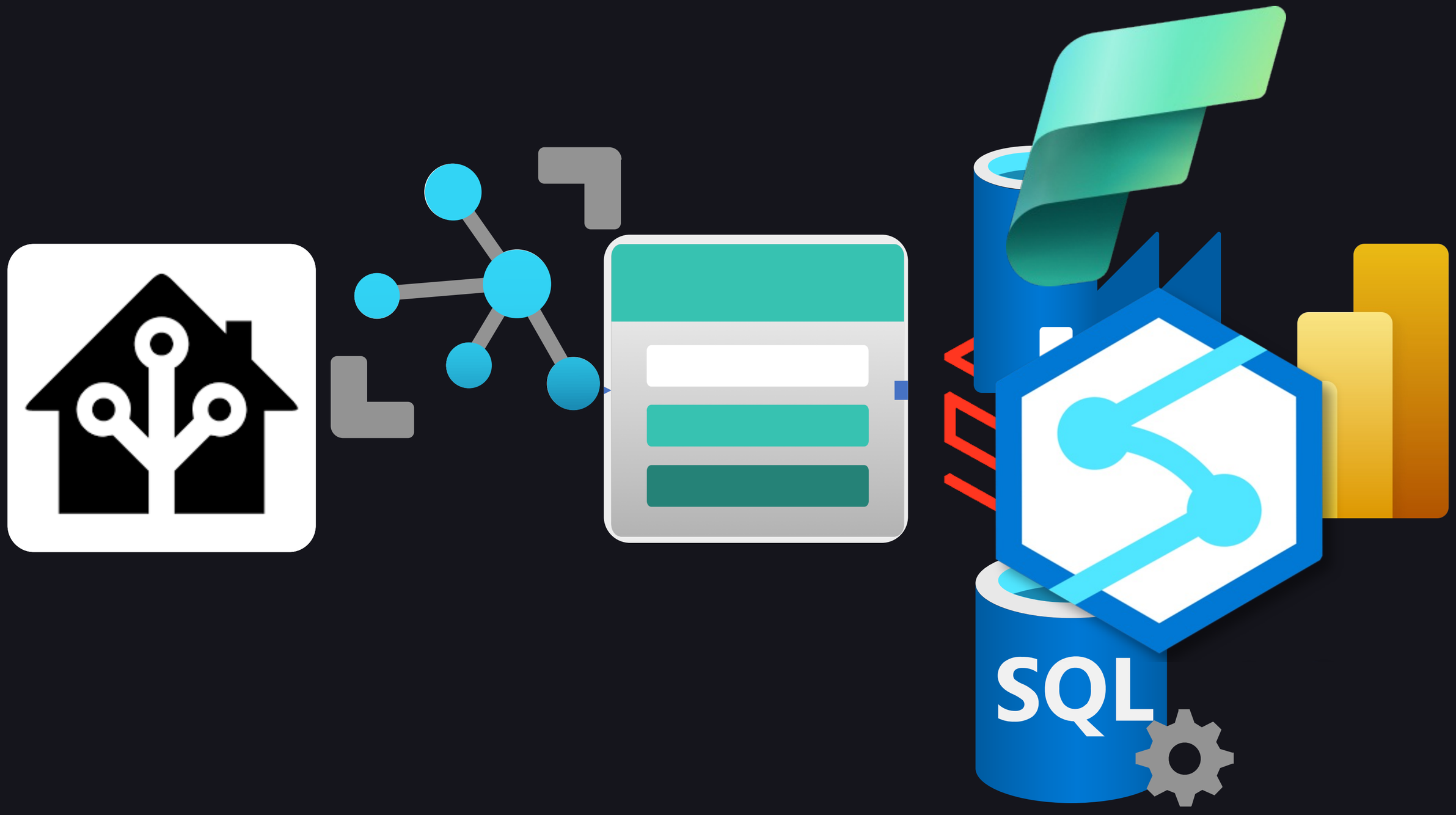
Choosing infrastructure

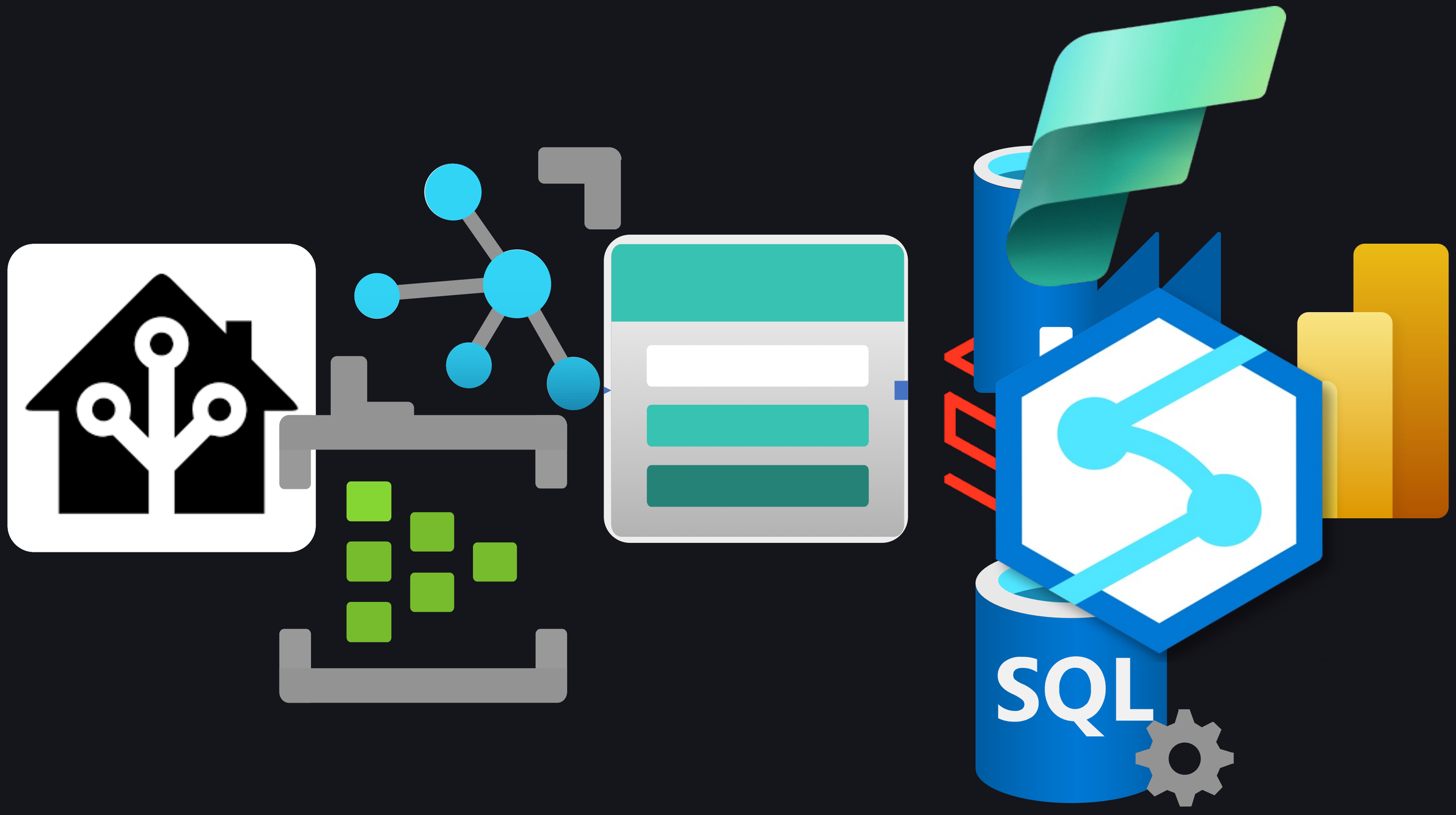

SQL

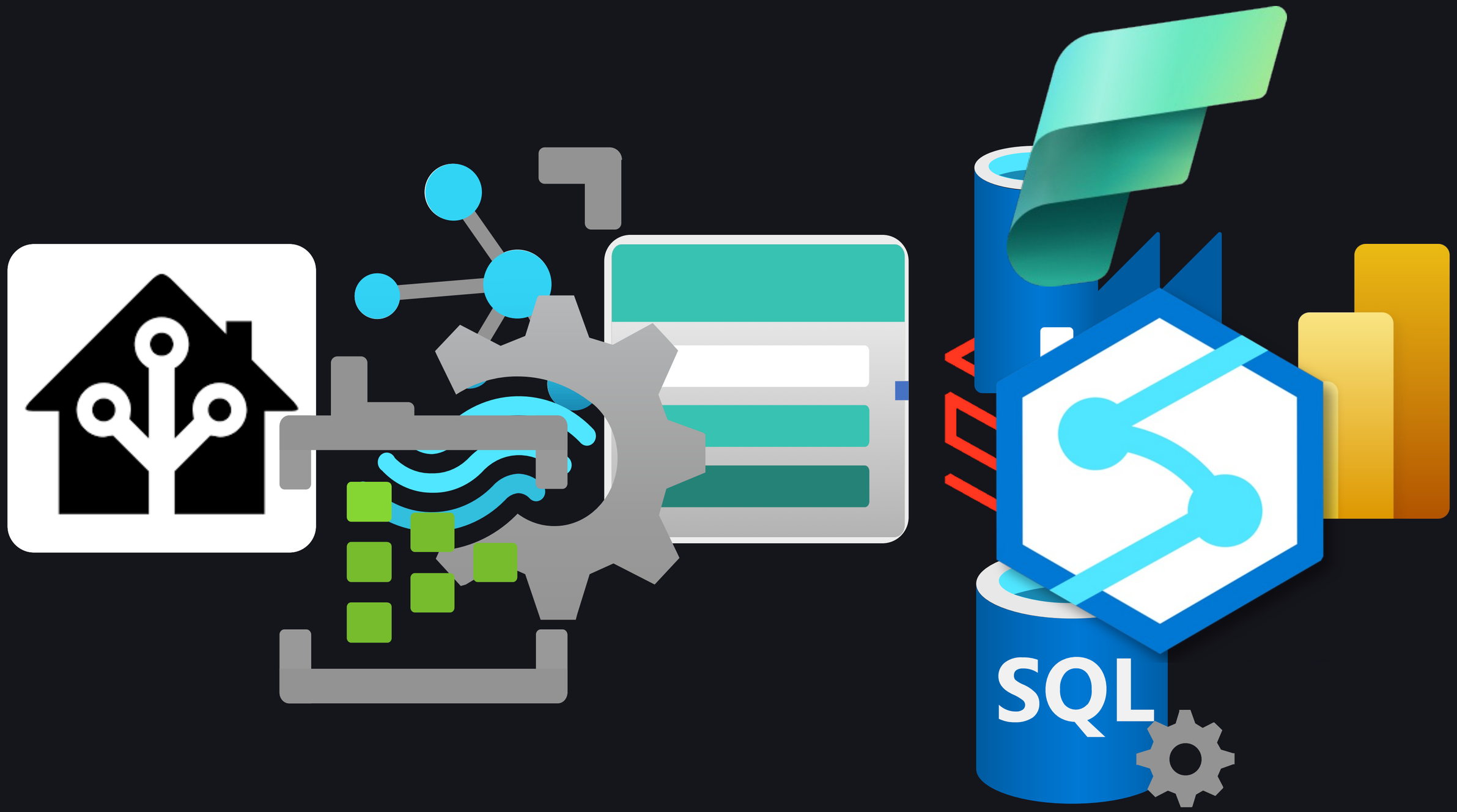
SQL

SQL

SQL

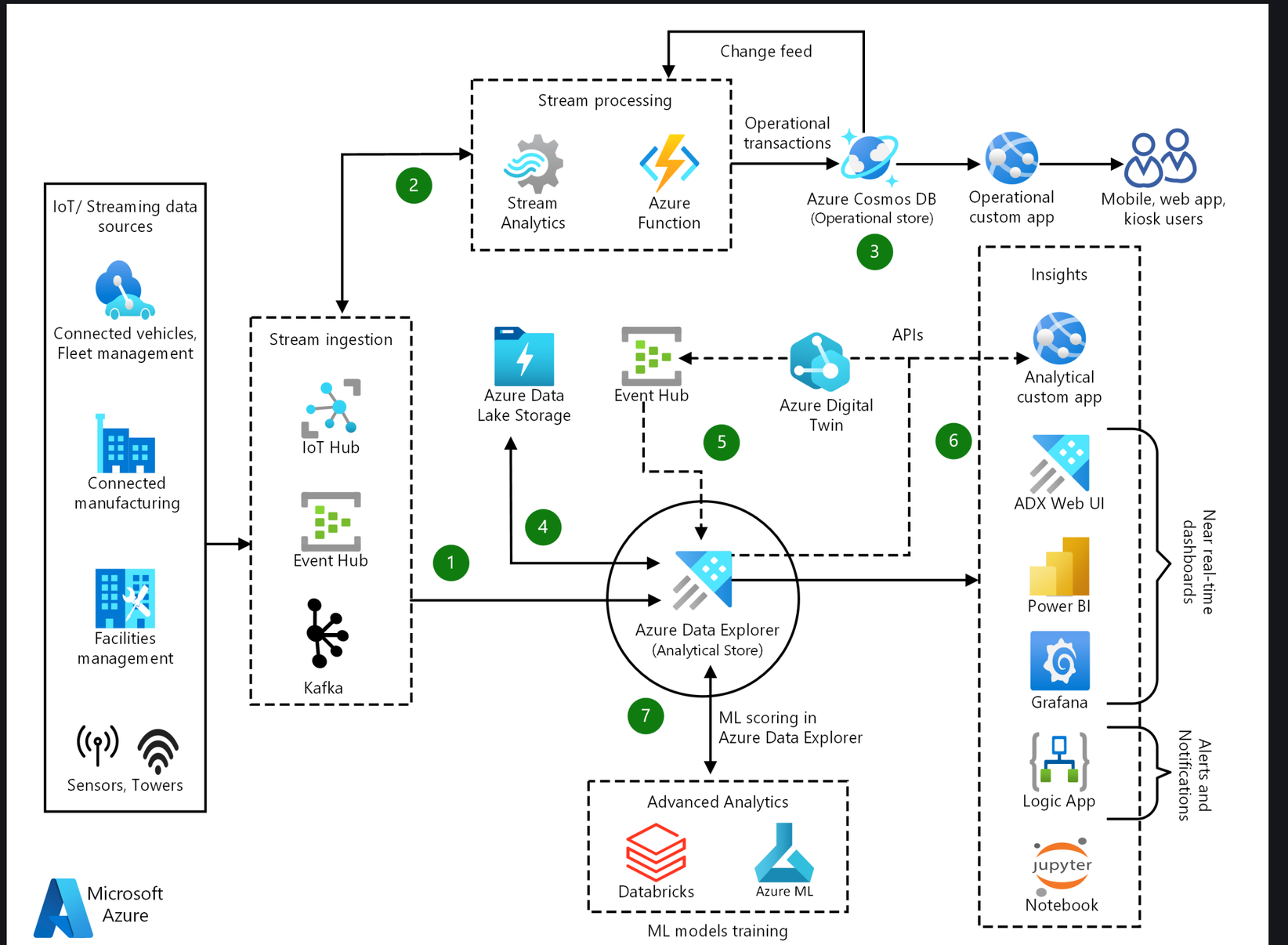

SQL


SQL


SQL

BUY ALL THE THINGS!

Choosing infrastructure



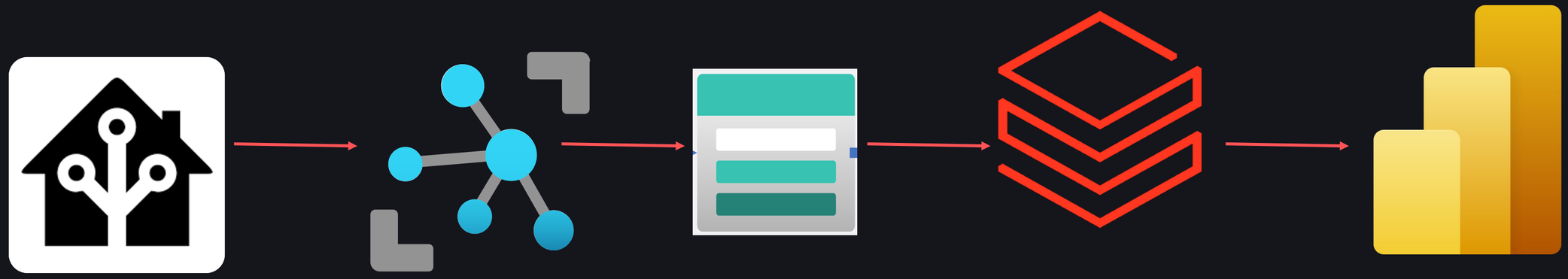
Source: learn.microsoft.com/en-us/azure/architecture

Choosing infrastructure

Keep it to a minimum!

- Stream vs batch processing
- Execution costs per use case
- Knowledge base

Choosing infrastructure



Home Assistant

IoT Hub

ADLS Gen 2

Databricks

Power BI

03

Lessons learned

Implementing infrastructure

Example JSON output


```
{
    "EnqueuedTimeUtc": "2024-05-29T04:51:01.5090000Z",
    "Properties": {},
    "SystemProperties": {
        "connectionDeviceId": "HomeAssistant",
        "connectionAuthMethod": "{\"scope\":\"device\",\"type\":\"sas\",\"issuer\":\"iothub\"}",
        "connectionDeviceGenerationId": "638398701538989601",
        "contentType": "application/json",
        "contentEncoding": "utf-8",
        "enqueuedTime": "2024-05-29T04:51:01.5090000Z"
    },
    "Body": {"sensor1": 21.12, "sensor2": 20.64
    }
}
```

03

Lessons learned

More data = more better?

More data = more better

Think about standards

- {iothub}/{partition}/{YYYY}/{MM}/{DD}/{HH}/{mm}

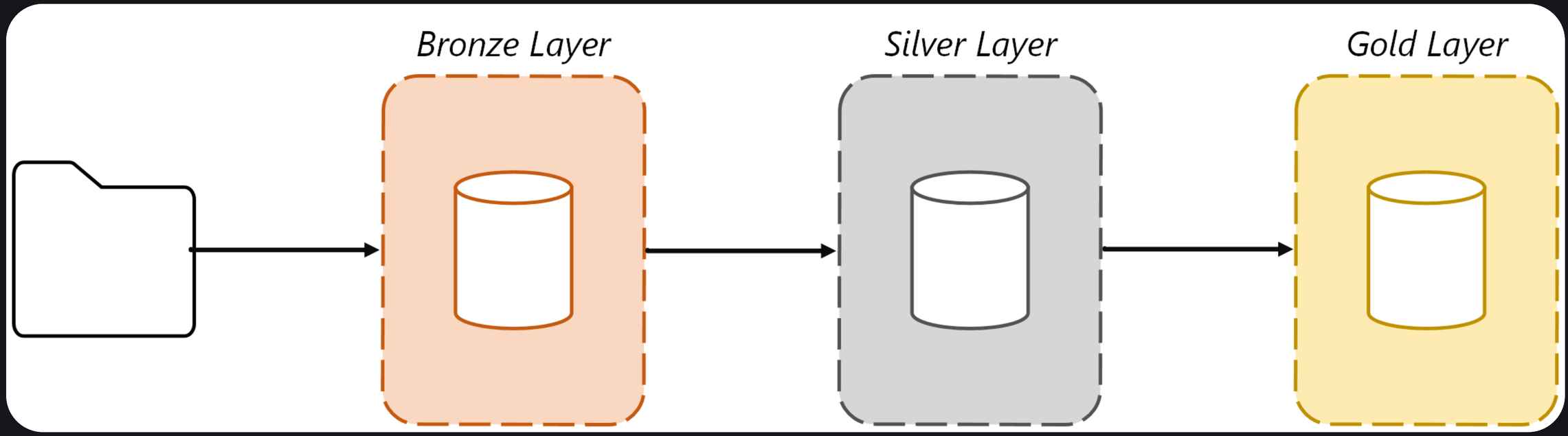
More data = more better

Think about standards

- {iothub}/{partition}/{YYYY}/{MM}/{DD}/{HH}/{mm}
- <source>/<processing_type>/<table>/<version>/loaddate=<loaddate>

More data = more better

Data Lake Layers

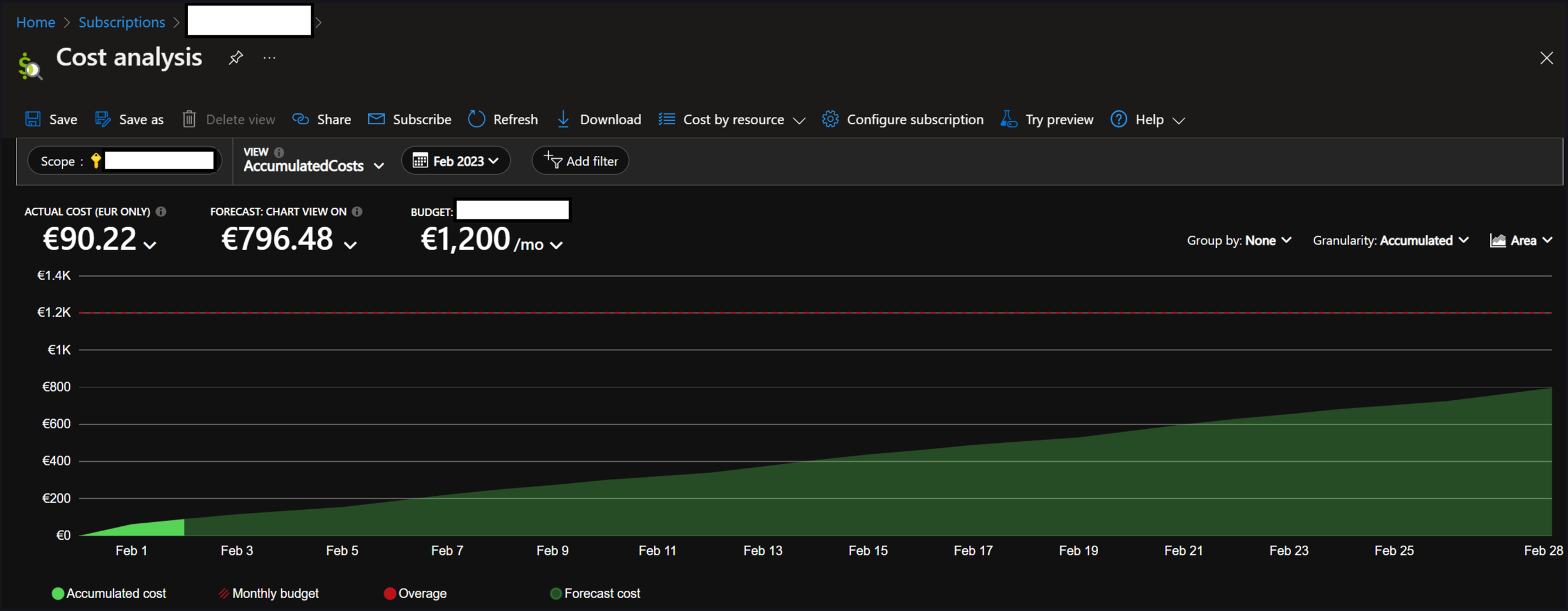


03

Lessons learned

Monitoring usage

Monitoring usage



Monitoring usage

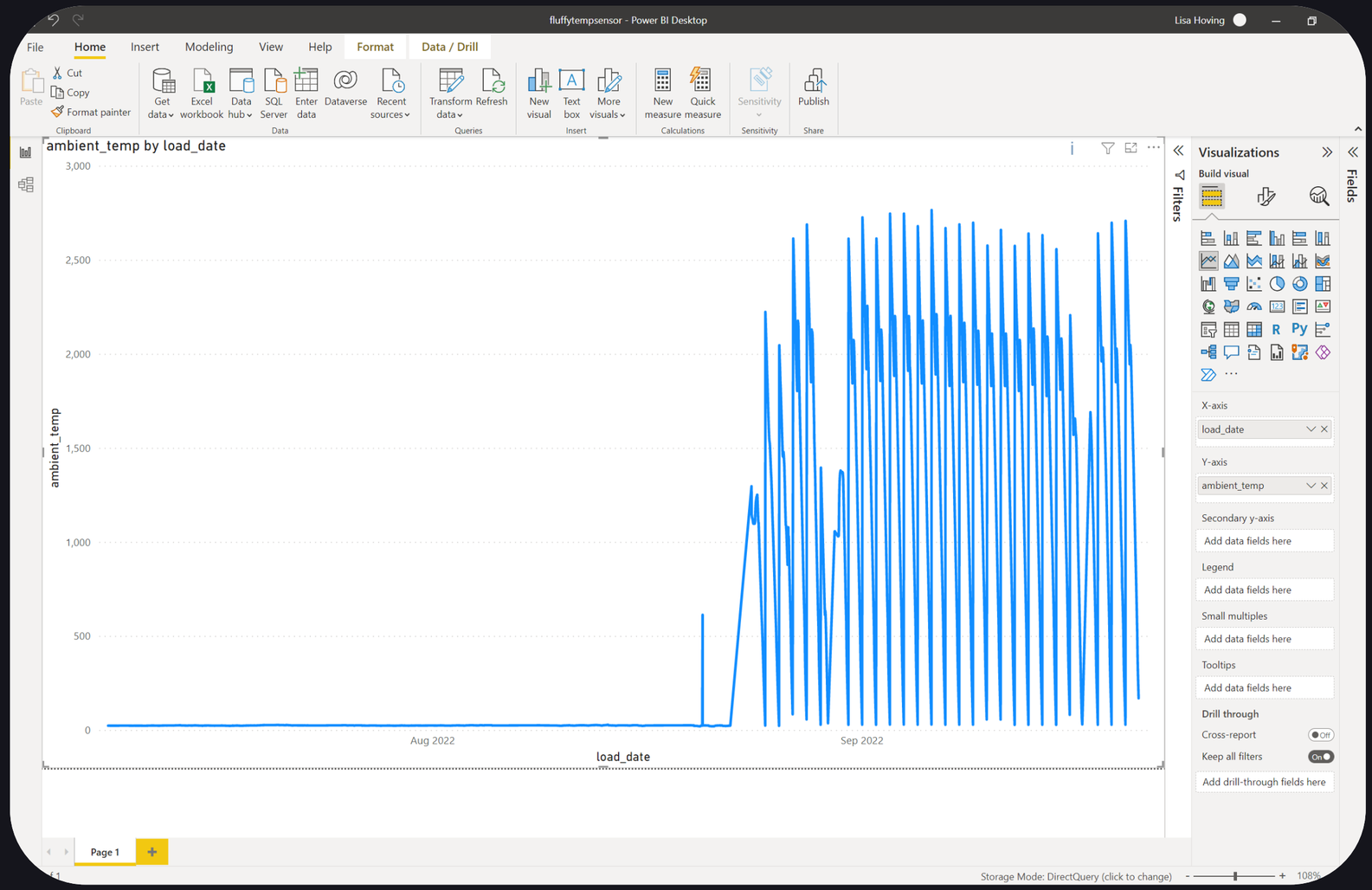
- Cost analysis
- Budgets
- Cost alerts

03

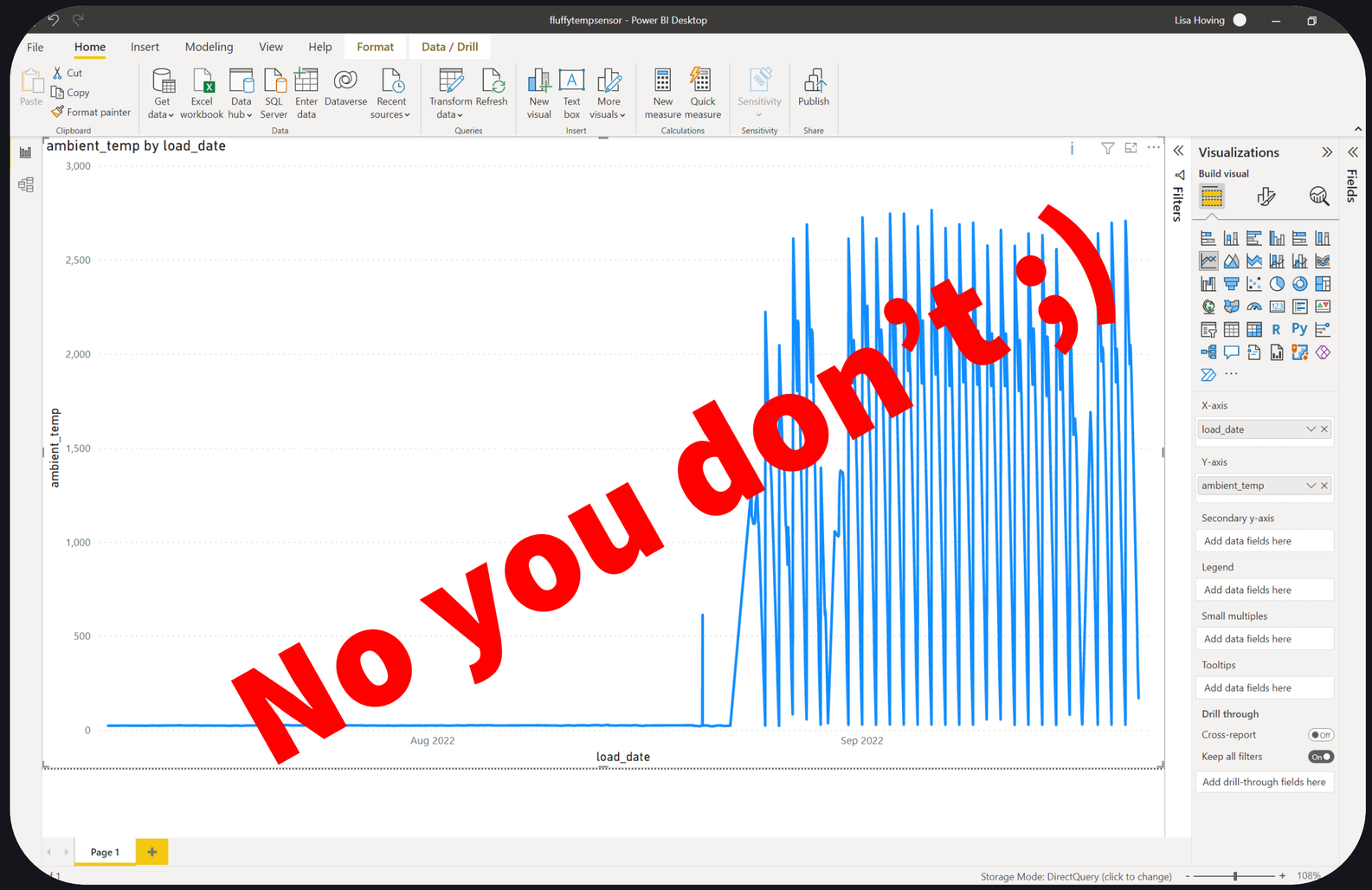
Lessons learned

Do you think you know your data?

Do you think you know your data?



Do you think you know your data?



04

Conclusion

Keep your head cool

Conclusion

- Don’t buy all the things
- Think about storage
  - *Standards*
  - *Paths*
  - *Compute*
- Monitor your usage!

MVP

Lisa Hoving

Big Data Engineer at Rabobank

@Lizaard08