

This Presentation



Machine learning & Kafka KSQL stream processing



[linkedin.com/in/simonaubury](https://www.linkedin.com/in/simonaubury)



@SimonAubury



Hello!

I am **Simon Aubury**

Principal Data Engineer @ ThoughtWorks

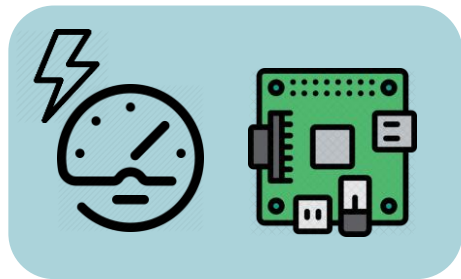
I am here because I love streaming & have spent too much time shopping on eBay



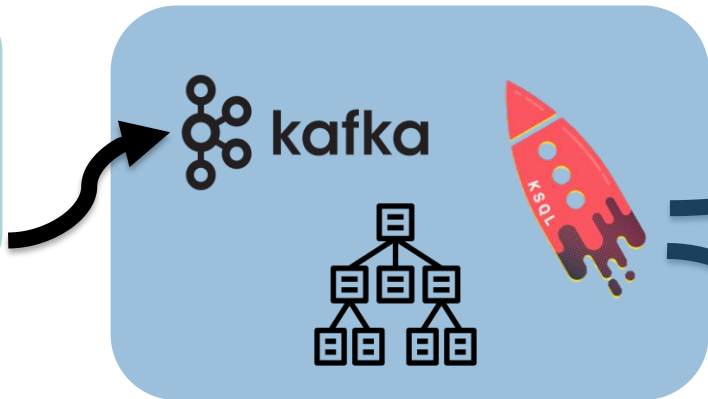
Goal of project

- Measure home power consumption
- Understand typical usage pattern
- Alert on unusual





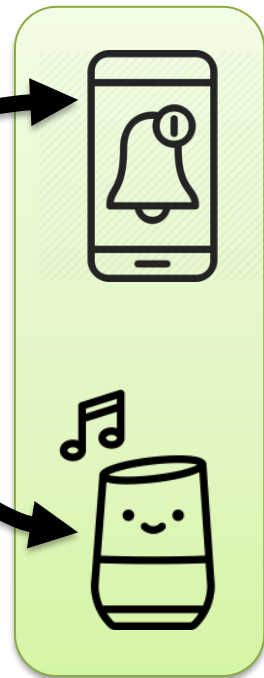
1. **Measure**



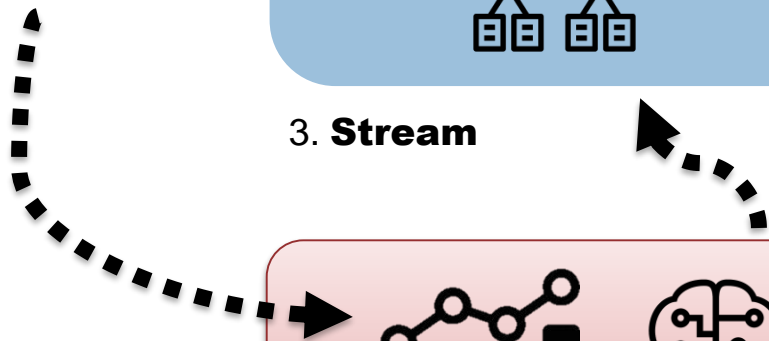
3. **Stream**

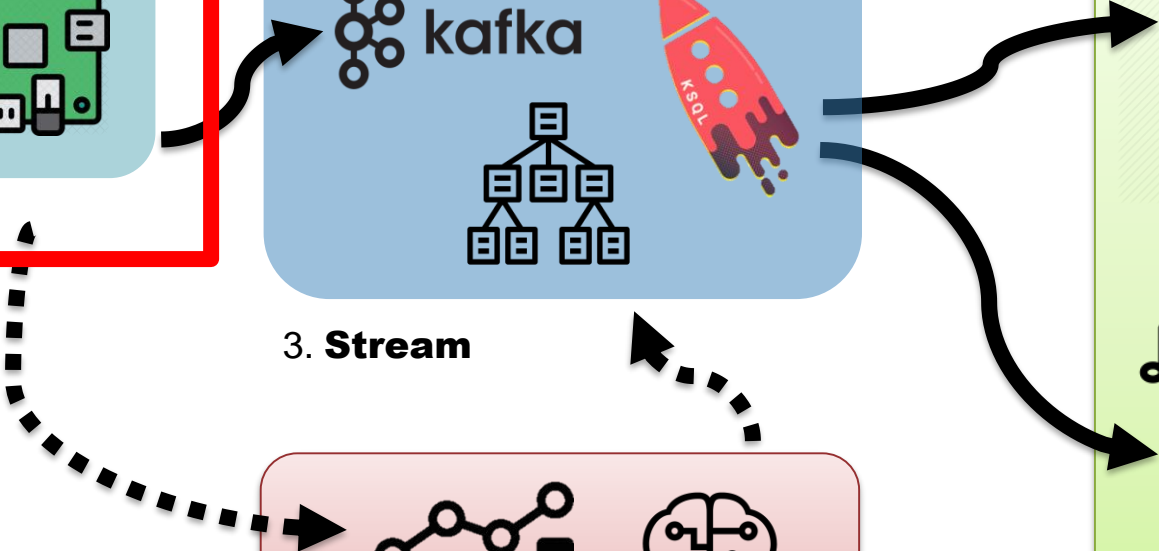
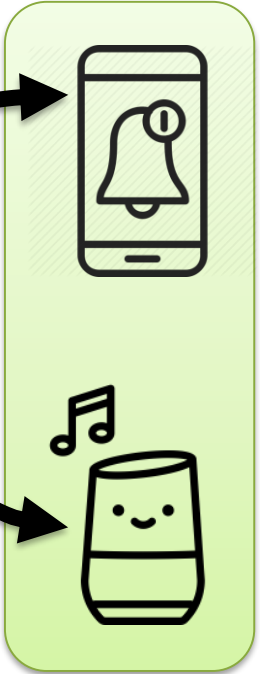
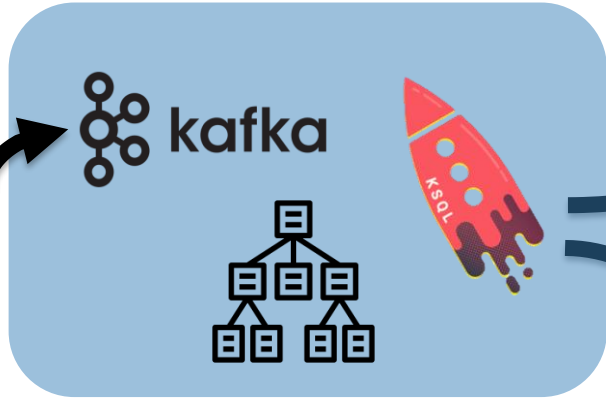
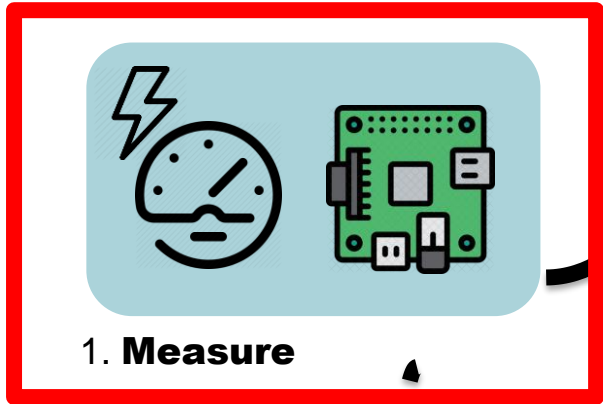


2. **Train**



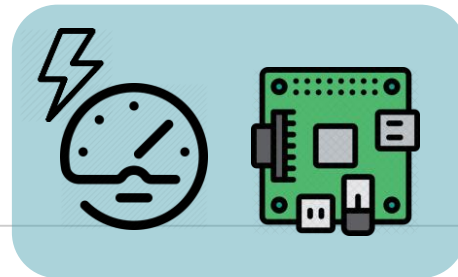
4. **Notify**



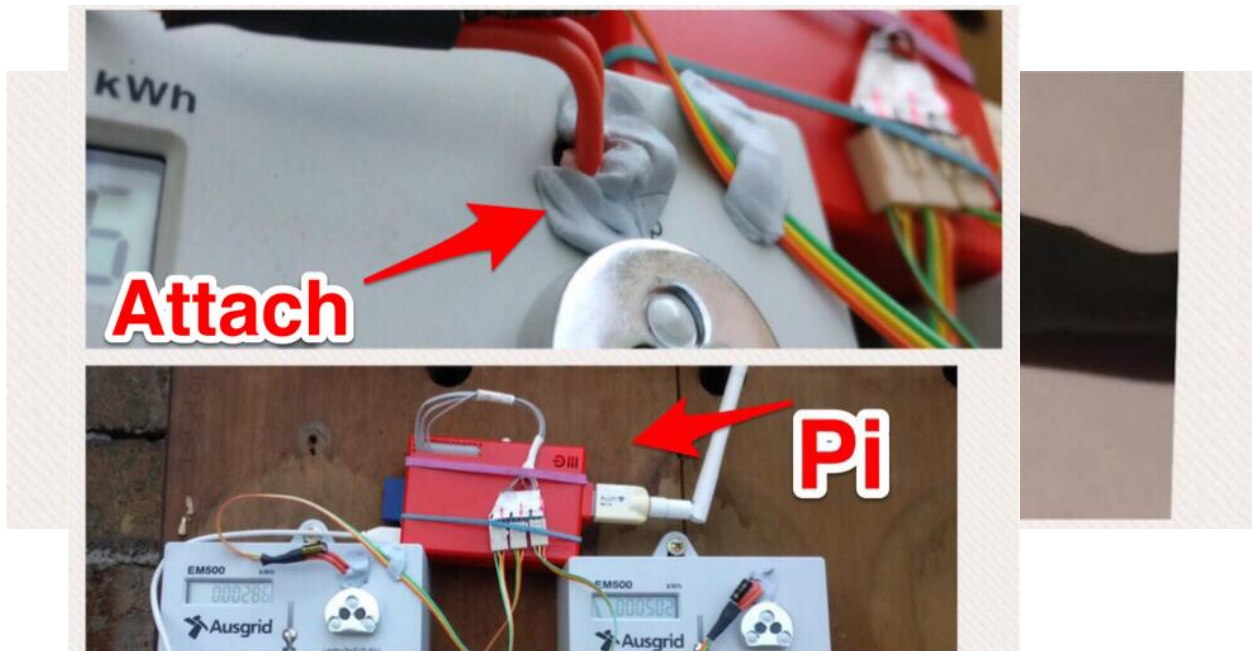




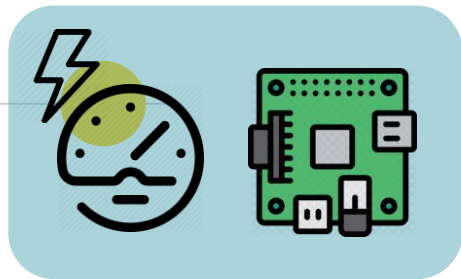
Home Power Monitoring using a **Raspberry Pi**.



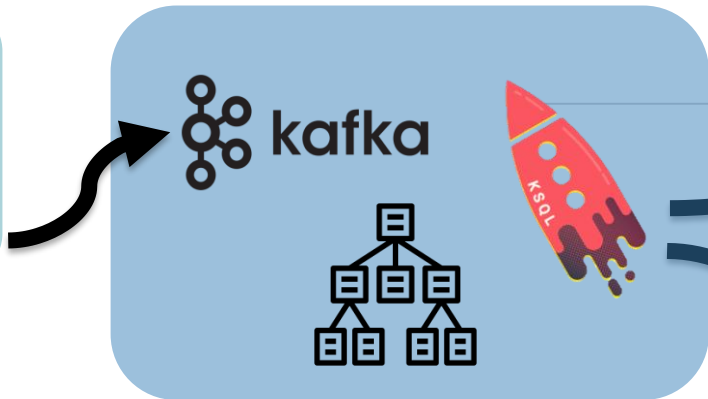
1. Measure



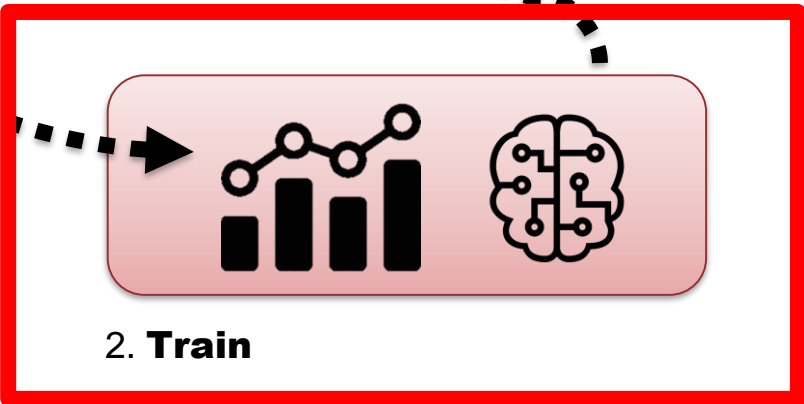
<https://medium.com/@simon.aubury/home-power-monitoring-65d0fded7769>



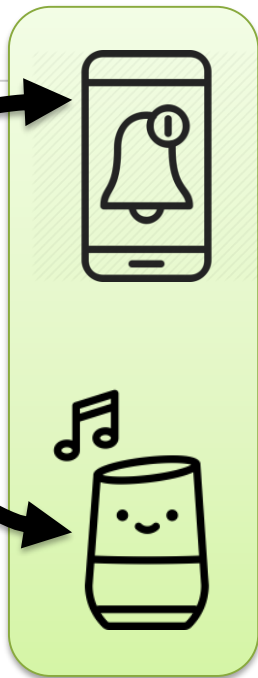
1. **Measure**



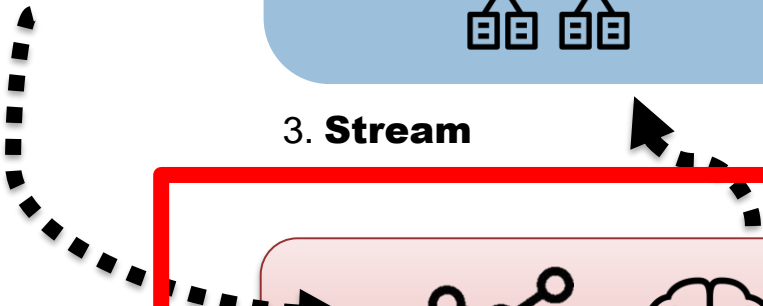
3. **Stream**



2. **Train**



4. **Notify**

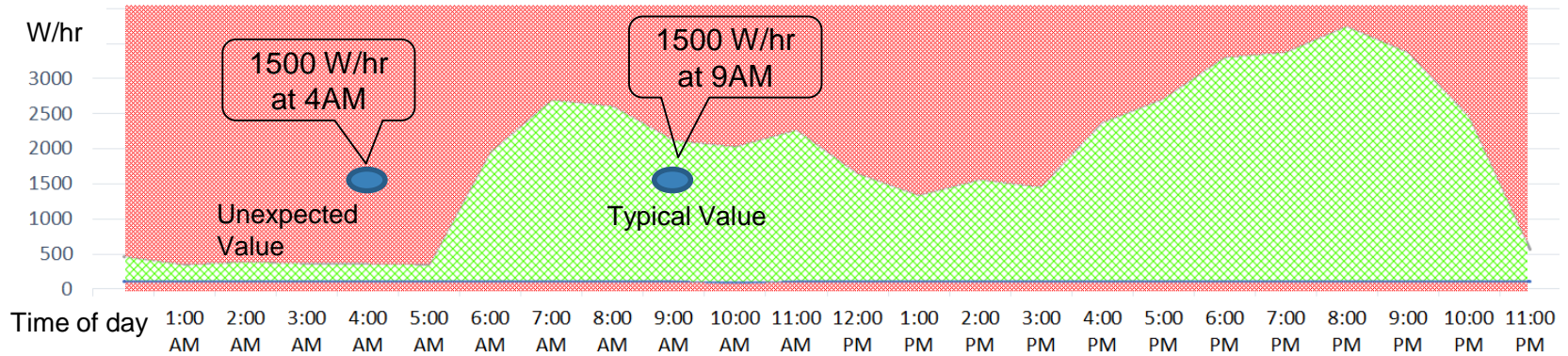




How to **train** your model



2. Train

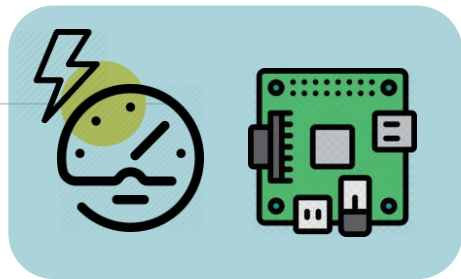




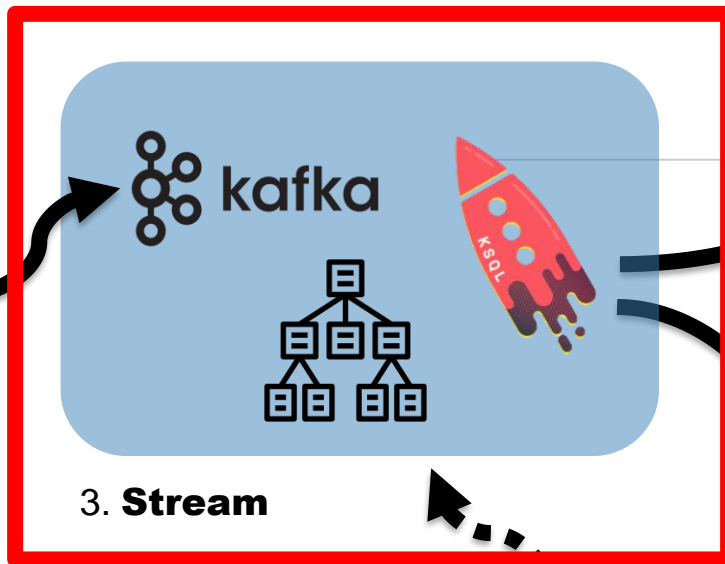
How to **train** your model

- 3 months of data
- Used H2O.ai
- score = AnomalyFunction (day, hour, power-usage)

	150 Wh	1500 Wh	3000 Wh
4 (4:00 am)	0.1647	5.4588	11.3412
9 (9:00 am)	0.0194	0.6914	1.4380
20 (8:00 pm)	0.0115	0.3827	0.7952



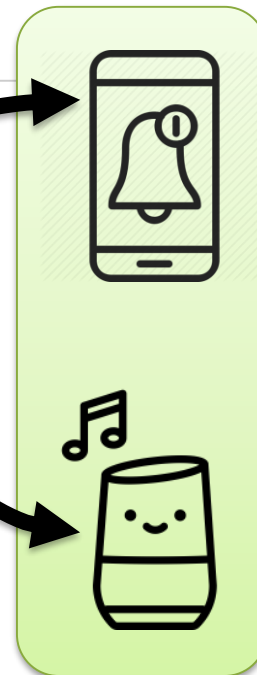
1. **Measure**



3. **Stream**



2. **Train**



4. **Notify**

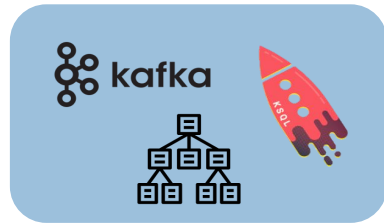




User Defined Function in **KSQL**

3. **Stream**

- KSQL is streaming SQL engine
- User Defined Scalar Functions (UDFs)
- Anomaly score function can be exposed to the KSQL server — and executed against the Kafka stream



3. Stream

User Defined Function in **KSQL**

- TL;DR summary — compile some Java and place in the right directory

```
ksql> list functions;
```

```
Function Name          | Type
```

```
-----
```

```
ANOMOLY POWER
```

```
| SCALAR
```

```
<--- I need this one
```

```
ANOMOLY_WATER
```

```
| SCALAR
```



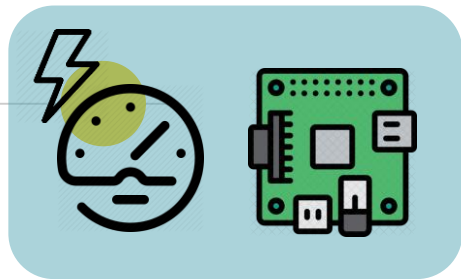
User Defined Function in KSQL



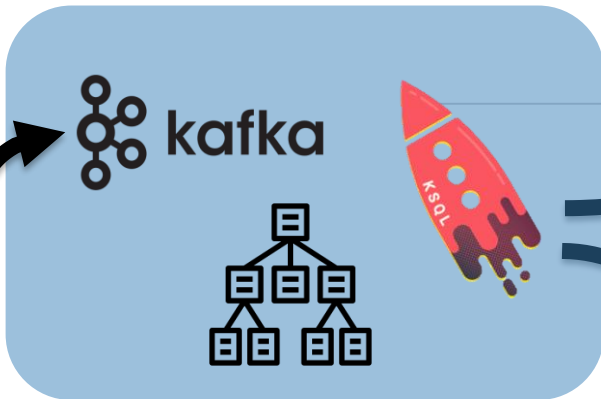
3. Stream

```
create stream raw_power_stream with (kafka_topic='raw_power',
value_format='avro');

create stream power_stream rekeyed as \
select rowtime, hour, kwh, anomaly_power(hour, kwh) as fn \
from raw_power_stream partition by rowtime;
```



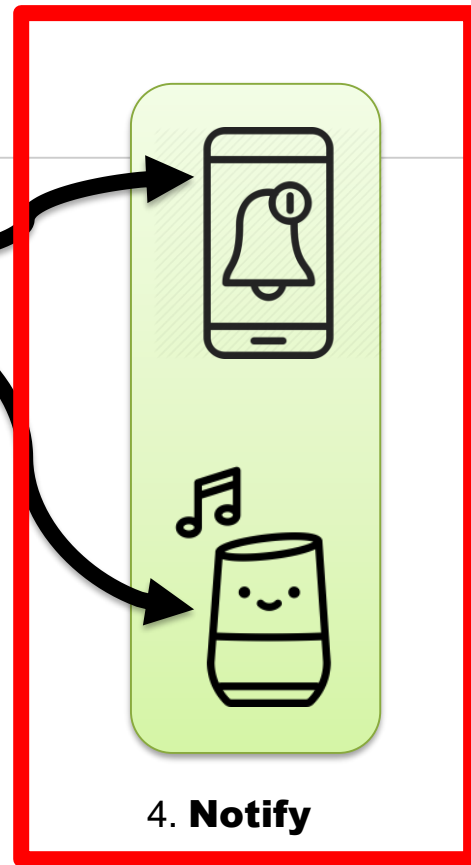
1. **Measure**



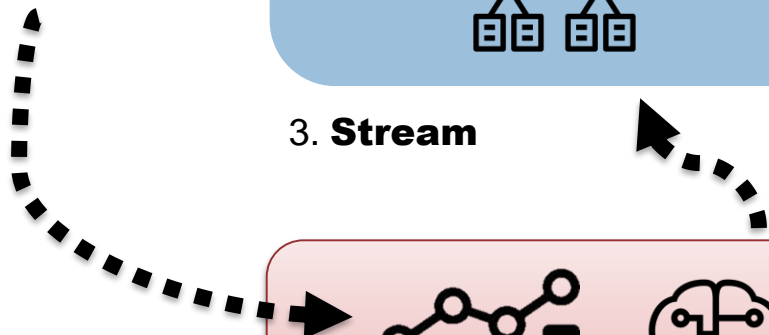
3. **Stream**



2. **Train**



4. **Notify**





Creating an anomaly topic



4. Notify

```
create stream anomaly_power with (value_format='JSON') as \  
select rowtime as event_ts, hour, kwh, fn \  
from power_stream_rekeyed where fn>1.0;
```



4. Notify



iOS **bug me**

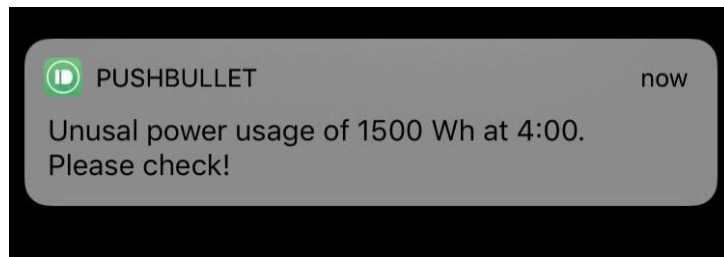
- Python consumes ANOMOLY_POWER topic
- Calls PushBullet API

```
c = Consumer(settings)
c.subscribe(['ANOMOLY_POWER'])

# Connect to pushbullet service
pb = Pushbullet(credentials.login['pushbullet_api_token'])

# Poll for messages; and extract JSON and call pushbullet for any messages
while True:
    msg = c.poll()
    app_json_msg = json.loads(msg.value().decode('utf-8'))

    # Send a push notification to phone via push-bullet
    push = pb.push_note('Unusual power usage of {:.0f} Wh at {:.0f}:00. Ple
```





Google yell at me



4. Notify

- Python consumes ANOMOLY_POWER topic
- Google Home Text-to-Speech (TTS) via Home Assistant

```
# Notifiy GoogleHome via Hass.io - Home Assistant
url = 'http://192.168.1.195:8123/api/services/tts/google_say?api_password={}'.format(c
data = '{"entity_id": "media_player.office_speaker", "message": "Warning. The power us
response = requests.post(url, data=data)
```





What did I learn?

The first legitimate alarm I received ...

This Presentation



Thanks!

Any **questions** ?



[linkedin.com/in/simonaubury](https://www.linkedin.com/in/simonaubury)



github.com/saubury/stream-smarts

Presentation template by SlidesCarnival