Streaming Anomaly Detection for Big Data / Internet of Things

Adam Drake
FOSSAsia 20160319

@aadrake
http://aadrake.com
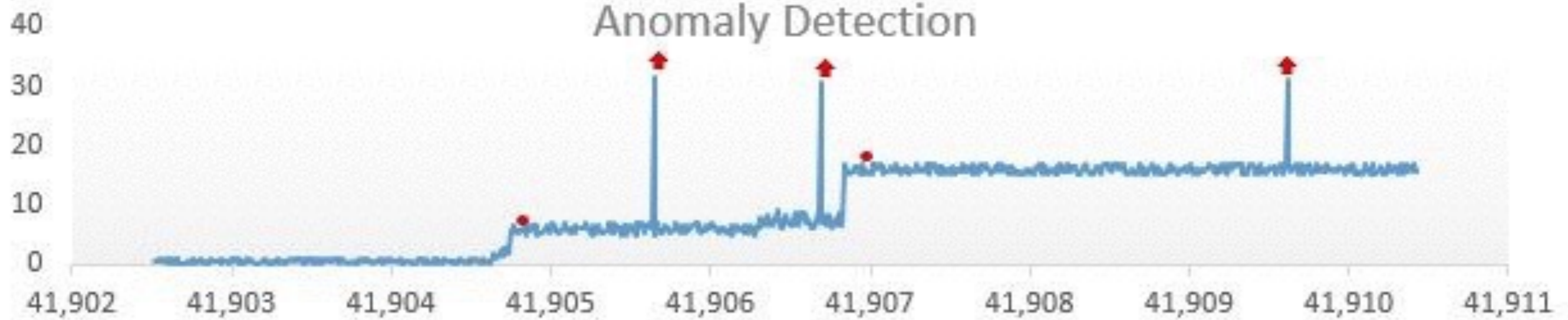
@aadrake #hacker

or

@aadrake #ThoughtLeader

Background
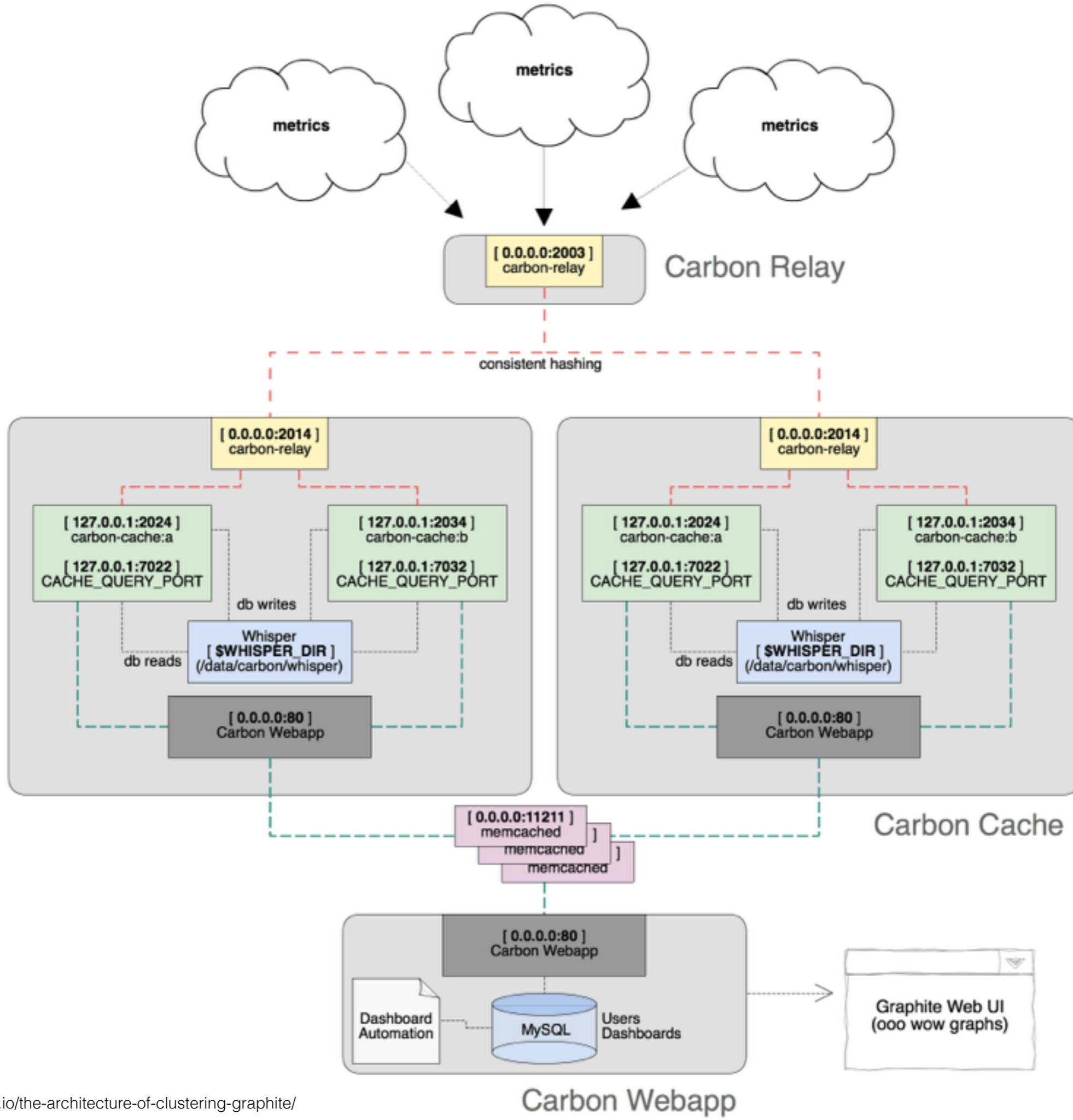
- Applied mathematics

- Startups

- Many industries

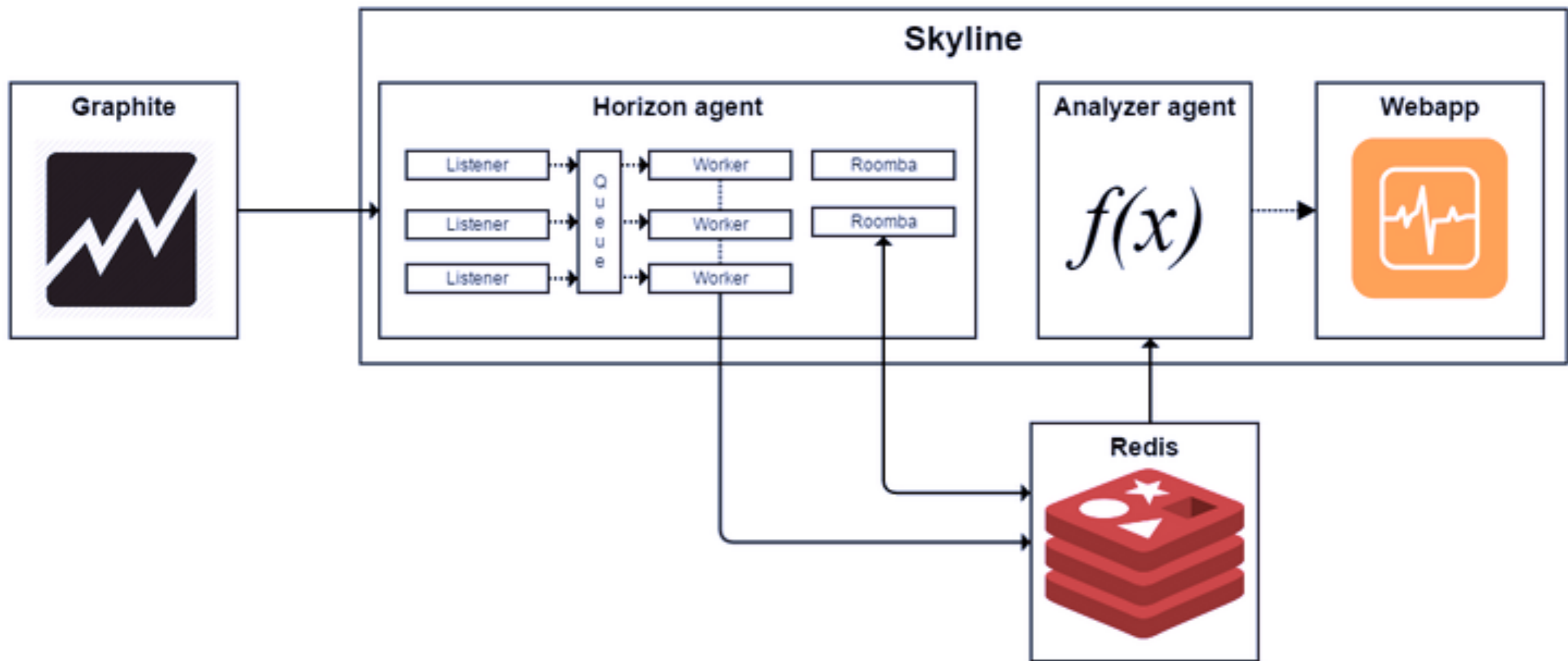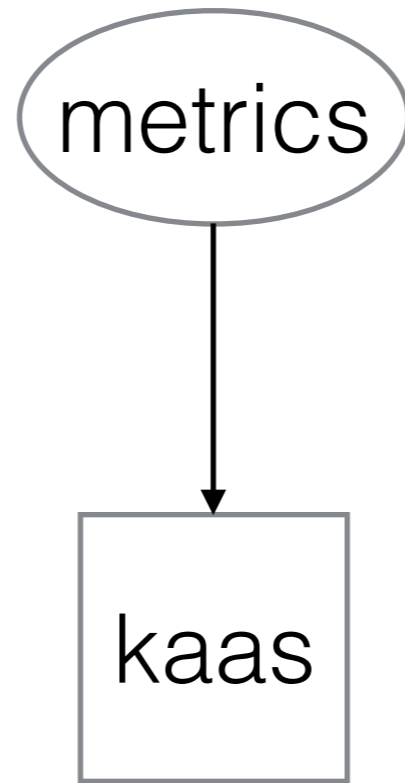# Time Series is the New Big Data

Anomaly Detection

Graphite
Prometheus (SoundCloud)
Nagios
BreakoutDetection (Twitter)
Surus (Netflix)
Atlas (Netflix)
Skyline (Etsy)
Bosun (Stack Exchange)

metrics

metrics

metrics

[0.0.0.0:2003]
carbon-relay

Carbon Relay

consistent hashing

[0.0.0.0:2014]
carbon-relay

[127.0.0.1:2024]
carbon-cache:a

[127.0.0.1:7022]
CACHE_QUERY_PORT

[127.0.0.1:2034]
carbon-cache:b

[127.0.0.1:7032]
CACHE_QUERY_PORT

db writes

Whisper
[$WHISPER_DIR]
(/data/carbon/whisper)

db reads

[0.0.0.0:80]
Carbon Webapp

[0.0.0.0:2014]
carbon-relay

[127.0.0.1:2024]
carbon-cache:a

[127.0.0.1:7022]
CACHE_QUERY_PORT

[127.0.0.1:2034]
carbon-cache:b

[127.0.0.1:7032]
CACHE_QUERY_PORT

db writes

Whisper
[$WHISPER_DIR]
(/data/carbon/whisper)

db reads

[0.0.0.0:80]
Carbon Webapp

Carbon Cache

[0.0.0.0:11211]
memcached
]
memcached
]
memcached

[0.0.0.0:80]
Carbon Webapp

Dashboard
Automation

MySQL

Users
Dashboards

Graphite Web UI
(ooo wow graphs)

Carbon Webapp

https://grey-boundary.io/the-architecture-of-clustering-graphite/

DON'T DO THIS

$$M_{2,n} = M_{2,n-1} + (x_n - \bar{x}_{n-1})(x_n - \bar{x}_n)$$

$$s_n^2 = \frac{M_{2,n}}{n-1}$$

Anomaly Detection