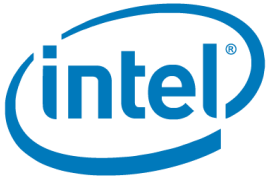
ANOMALY DETECTION IN GO

Marcin Spoczynski
Researcher @ Intel Labs
marcin.spoczynski@intel.com

GopherCon
15-07-2017



Legal Disclaimer

Co-funded by the Horizon 2020 Framework Programme
of the European Union

Who Am I

- Researcher on H2020 Mikelangelo project / Cloud Engineer
- Work/Research: telemetry, tools, cloud software management, scheduling
- now Researcher Intel Labs previously Python/C Developer

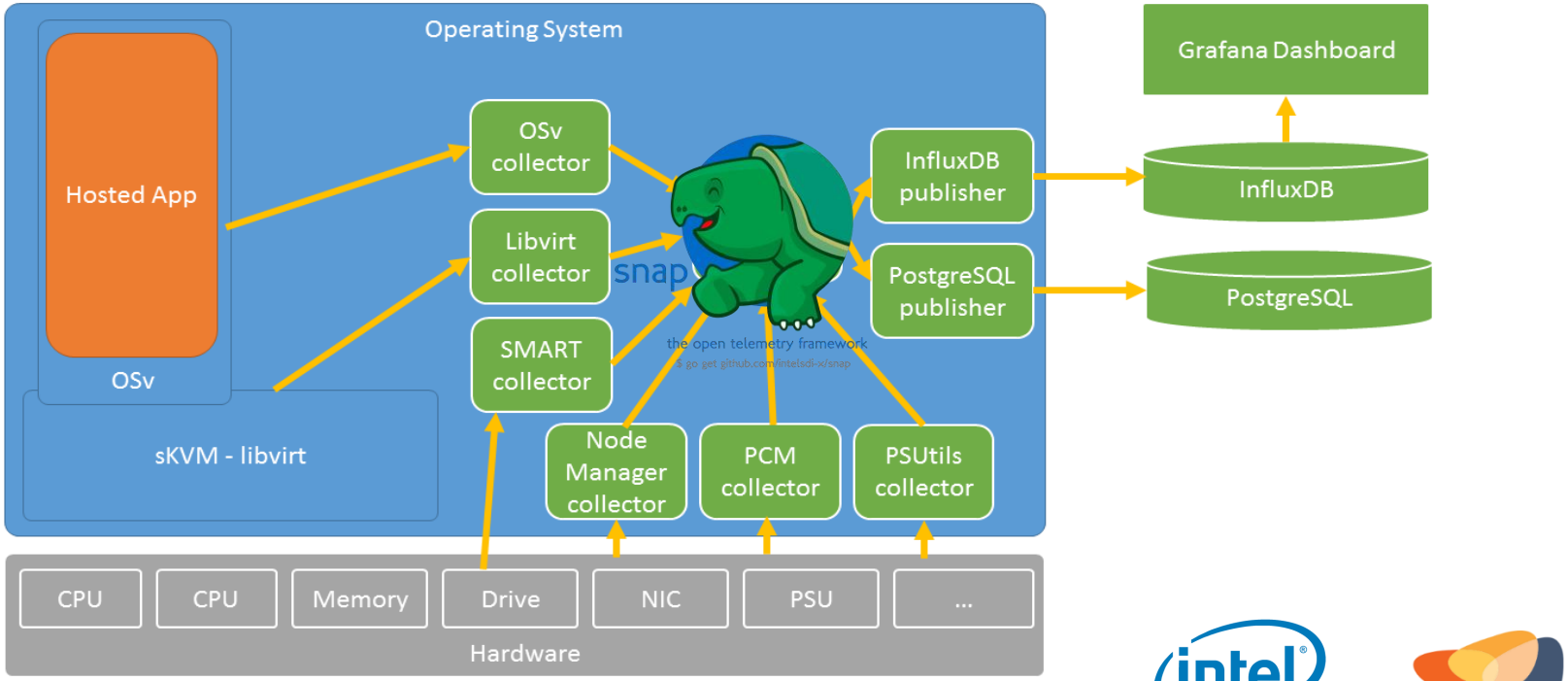
COLLECTING TELEMETRY FROM MIKELANGELO STACK

Telemetry

Telemetry is an automated communications process by which measurements and other data are collected at remote or inaccessible points and transmitted to receiving equipment for monitoring. The word is derived from Greek roots: *tele* = remote, and *metron* = measure.

source: wikipedia

What and why do we collect ?



intel

MIKELANGELO

Lots of data

After 1 day we collected around 1TB data from various sources:

- OS traces, syscall count
- Application layer traces
- Hardware layer traces

Collected data mostly static

1500135762 / cpu.utilization / 20

1500135462 / cpu.utilization / 20

1500135562 / cpu.utilization / 20

1500135962 / cpu.utilization / 20

1500137262 / cpu.utilization / 20

1500139362 / cpu.utilization / 29

1500141362 / cpu.utilization / 20

DO WE NEED TO STORE ALL THIS DATA ?

Anomaly Detection

In data mining, anomaly detection (also outlier detection) is the identification of items, events or observations which do not conform to an expected pattern or other items in a dataset.

source: http://www.wikipedia.com

Anomaly Detection

Typically the anomalous items will translate to some kind of problem such as software misconfiguration, attack, unexpected software behavior, hardware or software errors.

Anomaly Detection – Tukey Method

The intention of this method is to reduce the amount of data that needs to be transmitted without compromising the information that can be gained from potential usages of the data

https://github.com/intelsdi-x/snap-plugin-processor-anomalydetection

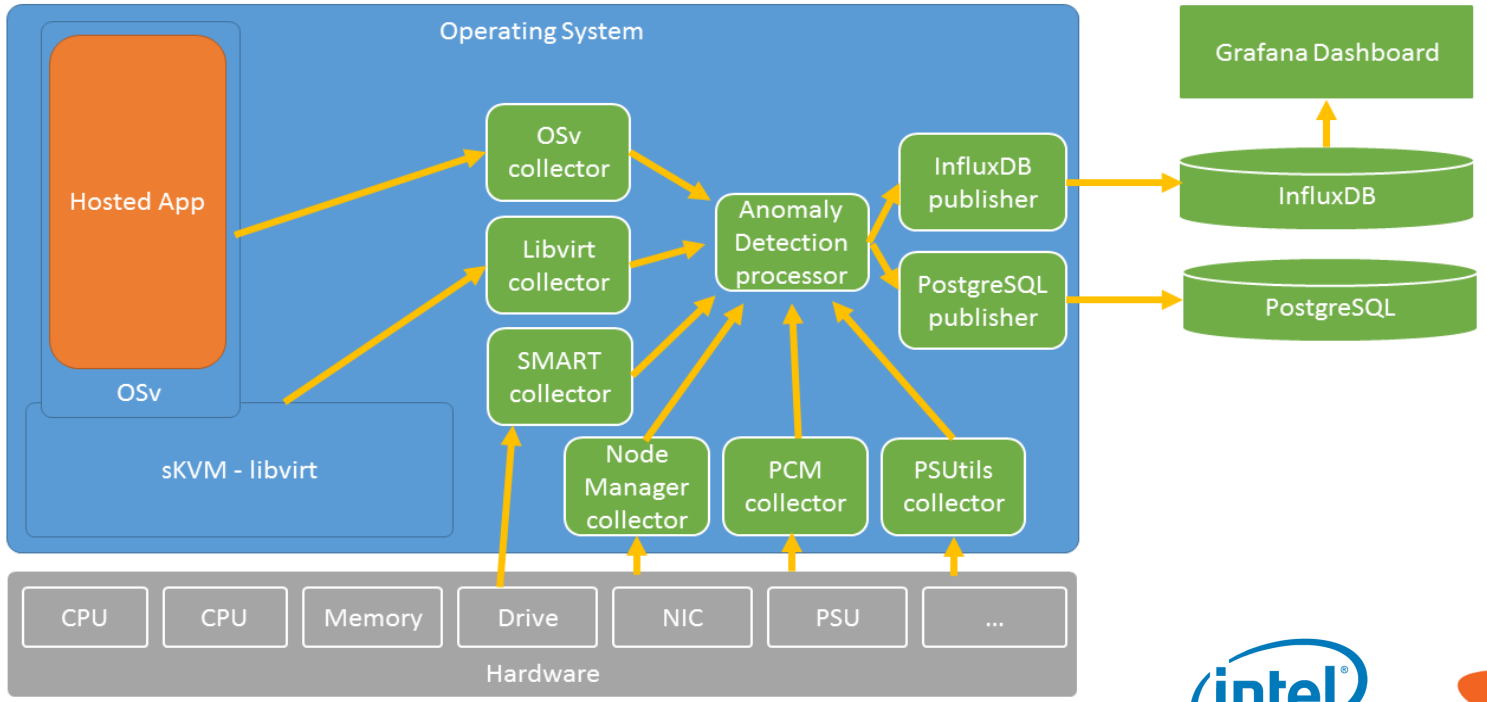
BETTER, CLOSER, WARMER.

Anomaly Detection – ESD

Seasonal Hybrid ESD (S-H-ESD) builds upon the Generalized ESD test for detecting anomalies. Note that S-H-ESD can be used to detect both global as well as local anomalies. This is achieved by employing time series decomposition and using robust statistical metrics, viz., median together with ESD.

https://github.com/intelsdi-x/snap-plugin-processor-anomalydetection

Where anomaly detection landed



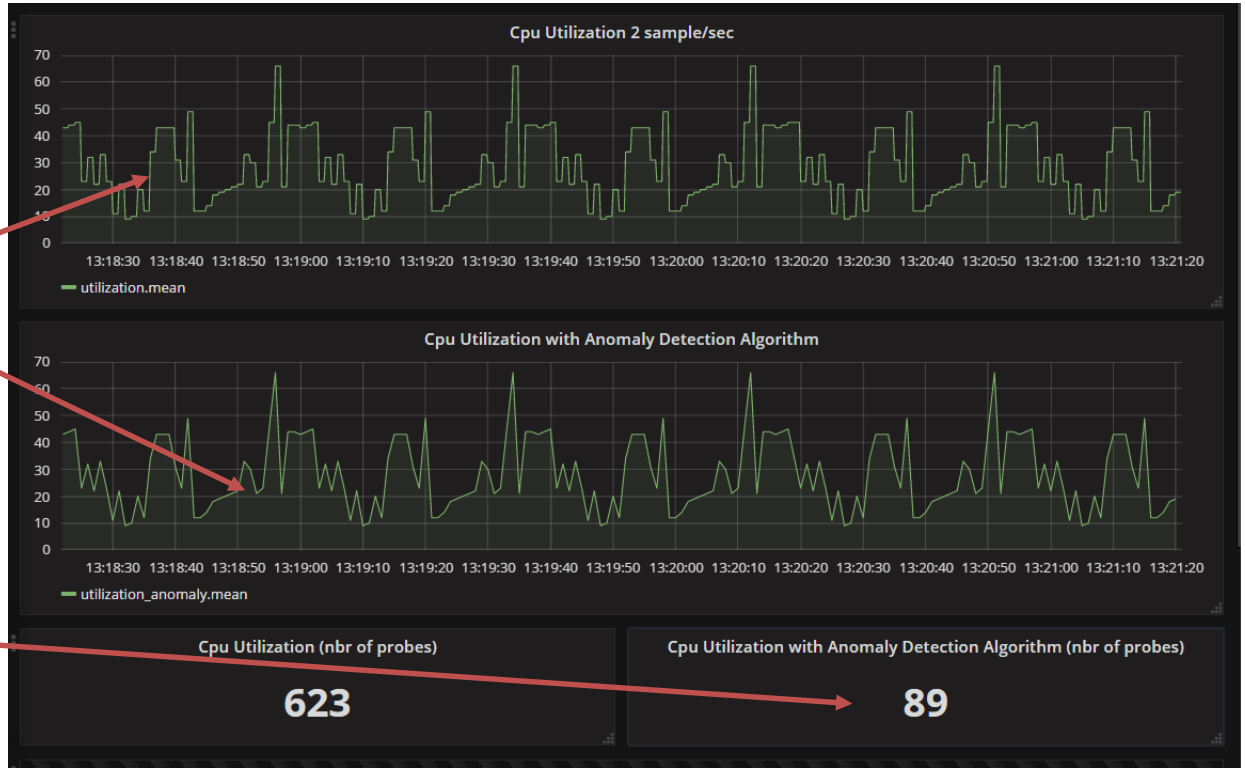
intel

MIKELANGELO

Anomaly Detection – Use Case

data shape preserved

8 x less samples



THANKS

marcin.spoczynski@intel.com
@sandlbn
github.com/sandlbn


MIKELANGELO