Detecting Workload Anomalies with Prometheus and Machine Learning

Anthony E. Nocentino

@nocentino

If you want to follow along: https://github.com/nocentino/MetricsML/

PURESTORAGE®

ScriptRunner®
The #1 for PowerShell Management

PDQ DataOn®

PATCH MY PC

MANNING

SNIA

Anthony E. Nocentino

Principal Field Solution Architect @ Pure Storage

- Specialize in system architecture, performance, SQL Server, Kubernetes, Containers, Microsoft Azure and VMware

- Masters Computer Science

email: anocentino@purestorage.com

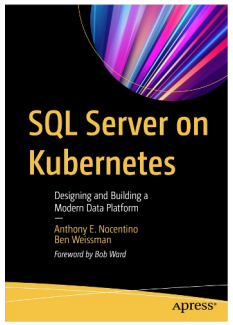
Blog: www.nocentino.com

Twitter: @nocentino

GitHub: https://github.com/nocentino/

Pluralsight Author: www.pluralsight.com

MVP Microsoft® Most Valuable Professional



PLURALSIGHT

EIGHTKB

Agenda

- Introduction to Telegraf, Prometheus and Grafana
- Building an Anomaly Detection Model
- Practical Use Cases for Anomaly Detection

If you want to follow along: https://github.com/nocentino/MetricsML

Why Did I Build This?

Customers often have no monitoring platforms ☹

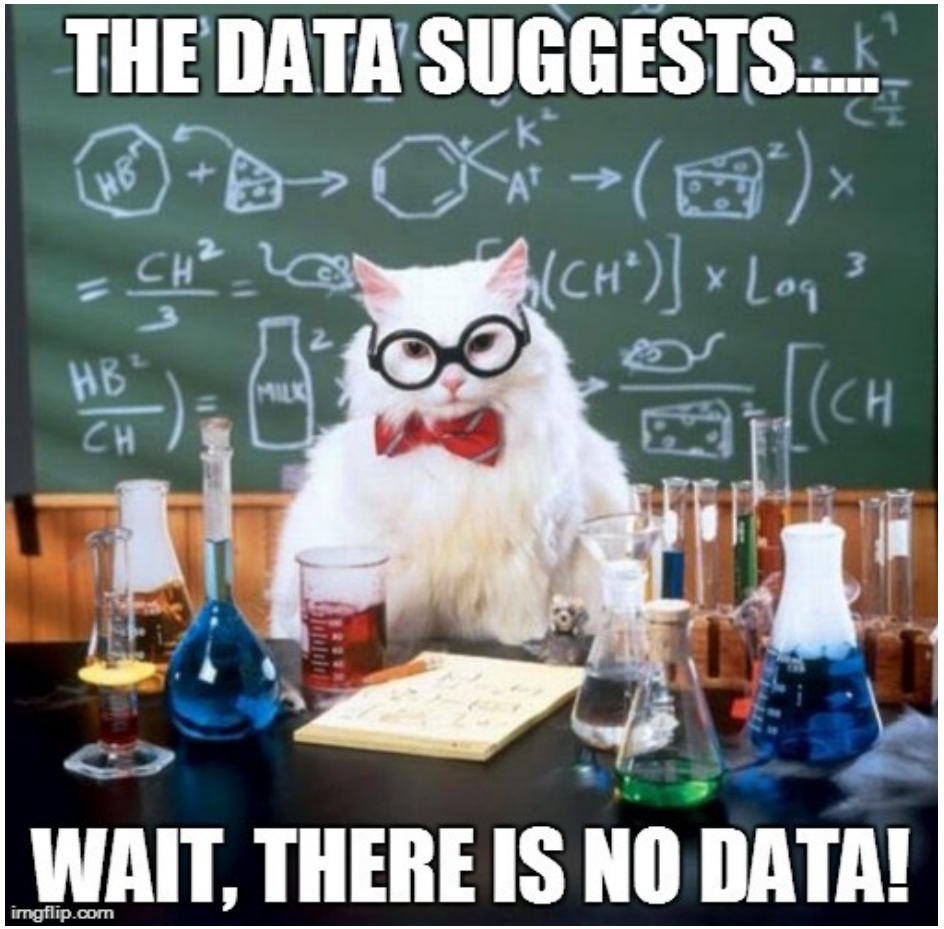
Defining what's normal and what's not normal

A couple disclaimers

- I'm not a data scientist
- Also, not a professional developer

But I had a problem I wanted to solve

https://github.com/nocentino/MetricsML

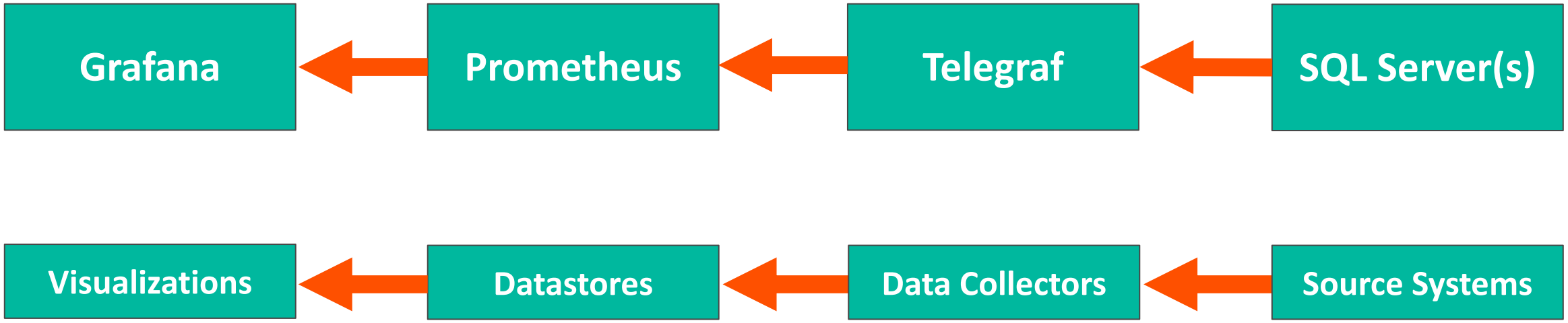


What’s Observability?

Understanding what’s happening and how it impacts the health of the system

Monitoring Architecture

Keeping an eye on complex systems



Container Based Deployment

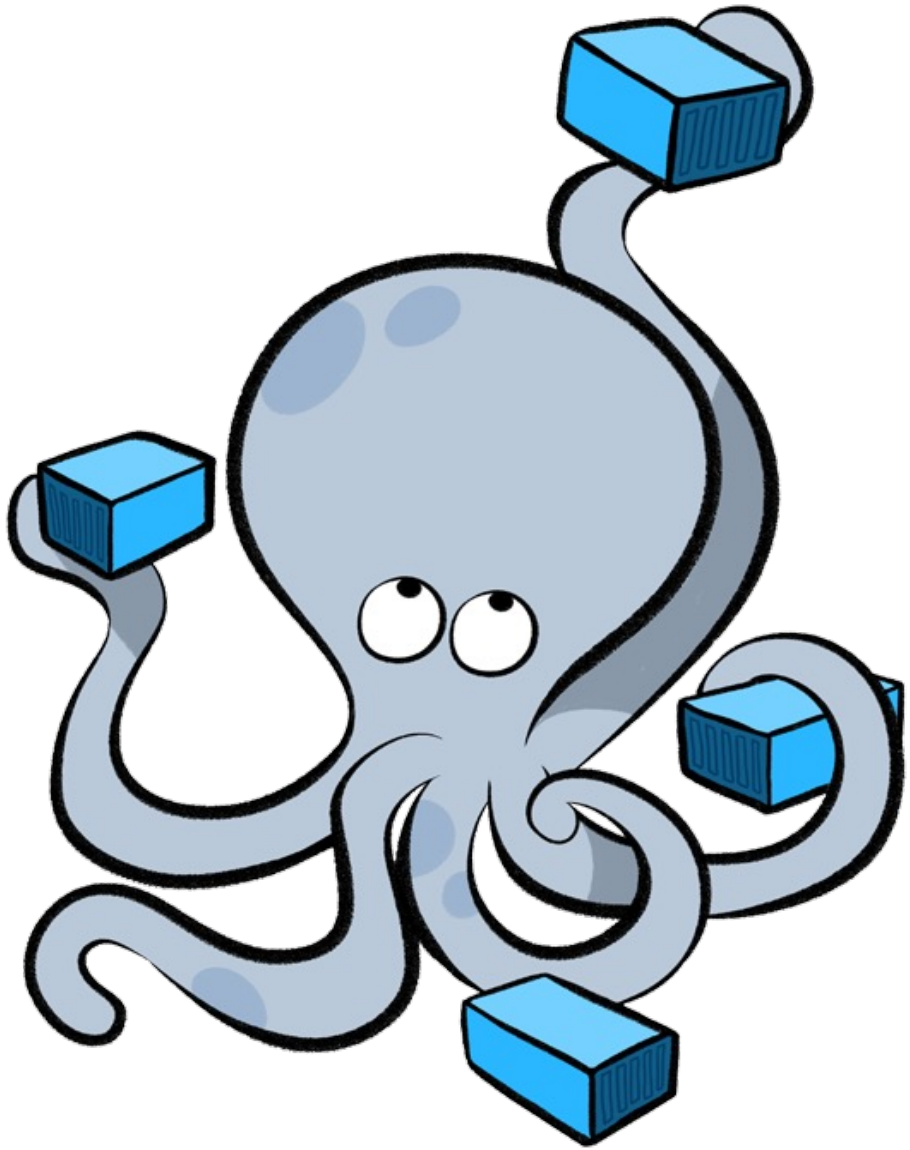
Container is a self-contained application

Docker Compose

- Starts up the containers
- Configures the applications
- Networking connecting the applications
- Expose Grafana to the network

Orchestrated solution defined in code

Can run anywhere you have Docker



Telegraf – What's a Metric?

Performance data

Plugins

Collecting metrics

Exposes a metrics HTTP/S endpoint

localhost:9273/metrics

```
# HELP sqlserver_cpu_other_process_cpu Telegraf collected metric
# TYPE sqlserver_cpu_other_process_cpu untyped
sqlserver_cpu_other_process_cpu{host="telegraf",measurement_db_type="SQLServer",sql_instance="sql1"} 100
sqlserver_cpu_other_process_cpu{host="telegraf",measurement_db_type="SQLServer",sql_instance="sql2"} 100
# HELP sqlserver_cpu_sqlserver_process_cpu Telegraf collected metric
# TYPE sqlserver_cpu_sqlserver_process_cpu untyped
sqlserver_cpu_sqlserver_process_cpu host="telegraf",measurement_db_type="SQLServer",sql_instance="sql1"} 0
sqlserver_cpu_sqlserver_process_cpu host="telegraf",measurement_db_type="SQLServer",sql_instance="sql2"} 0
# HELP sqlserver_cpu_system_idle_cpu Telegraf collected metric
# TYPE sqlserver_cpu_system_idle_cpu untyped
sqlserver_cpu_system_idle_cpu{host="telegraf",measurement_db_type="SQLServer",sql_instance="sql1"} 0
sqlserver_cpu_system_idle_cpu{host="telegraf",measurement_db_type="SQLServer",sql_instance="sql2"} 0
```

Other Telegraf Plugins

Windows Server 2022

MySQL

vmware

SQL

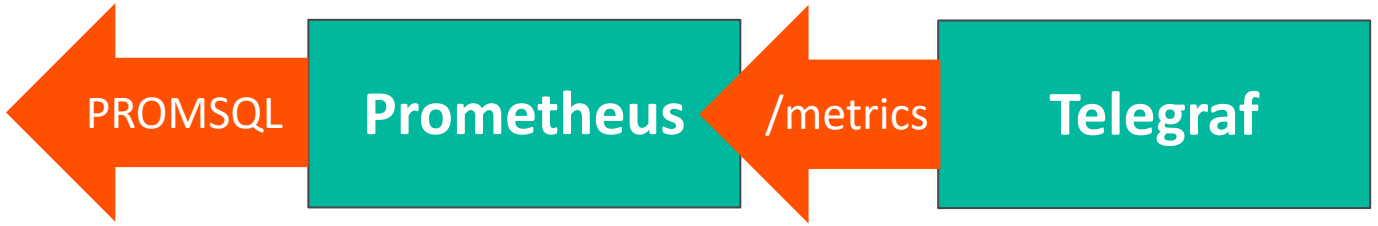
https://github.com/influxdata/telegraf

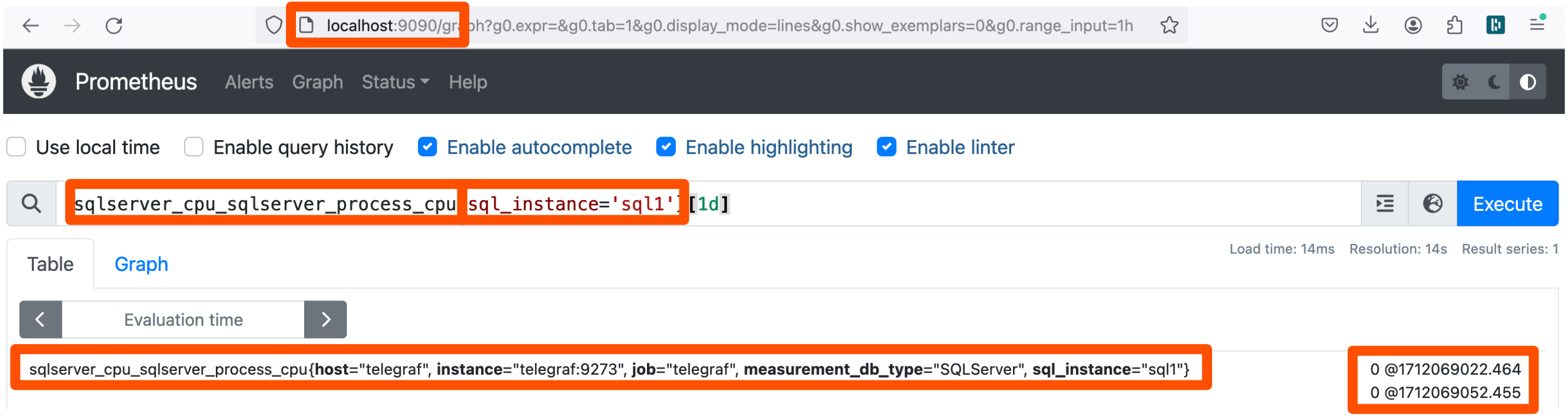
Prometheus – Time Series Database to Store Metrics

Time Series Database

Stores collected metrics

Jobs scrape the metrics endpoints





https://prometheus.io/docs/prometheus/latest/querying/basics/

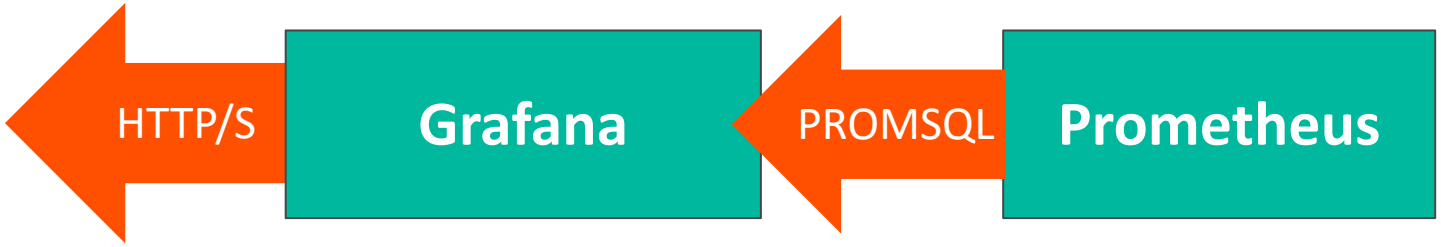
Grafana

Visualization tool

Define dashboards

Queries a datastore

See trends over time



Let’s hop into a demo

Start up the monitoring stack

Look at how metrics are exposed and collected

Check out how Grafana visualizes metrics

Machine Learning

What is Machine Learning (ML)?

Predictions based on past performance

Take a sample of data

Fit a model to data points

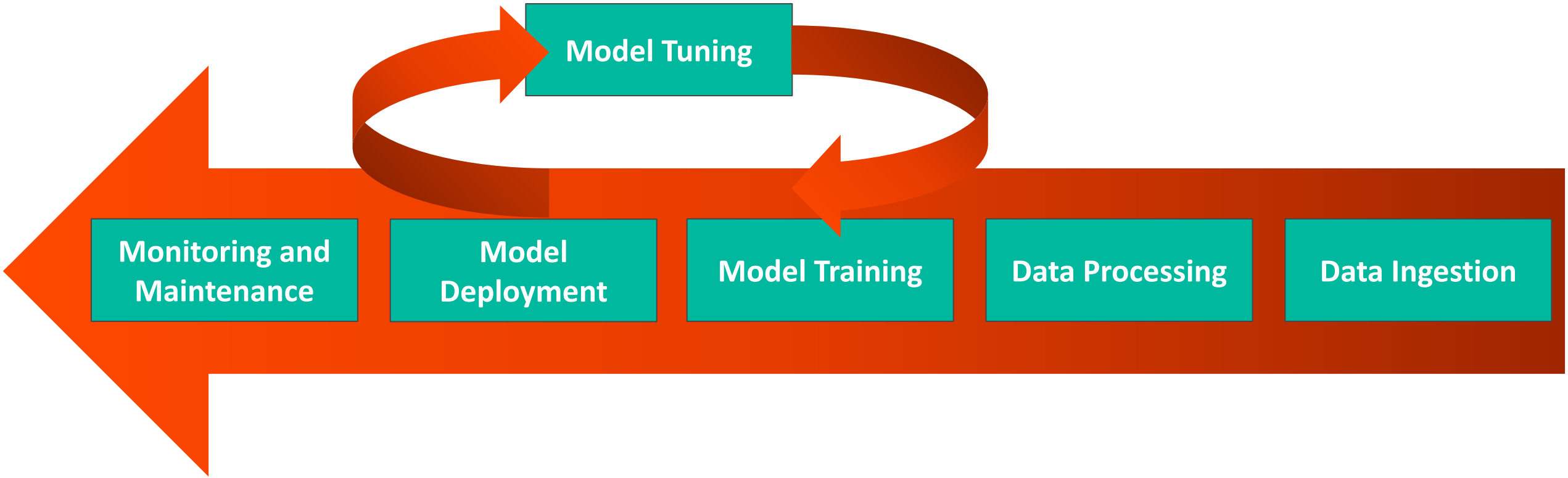
Parameters are used to tune the model

Python and R

I don’t know Python

So, I had to use Copilot for this

A Machine Learning Pipeline



Using Prophet

I not an expert at Machine Learning and I'm not a data scientist

Find something off the shelf

Time series machine learning model

Use for forecasting based on seasonality

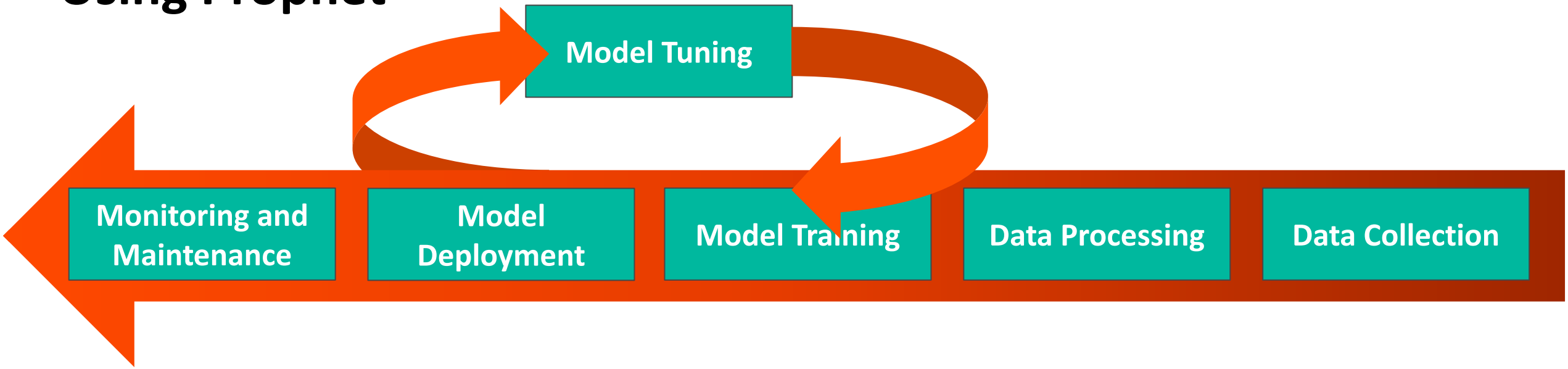
Tunable and Automated

Available in R or Python

Returns a **predicted value**, and an error range **high** and **low**

https://facebook.github.io/prophet/

Using Prophet



Let’s hop into a demo

Using Prophet at the CLI to generate predicted values

Building your own Metrics Exporter

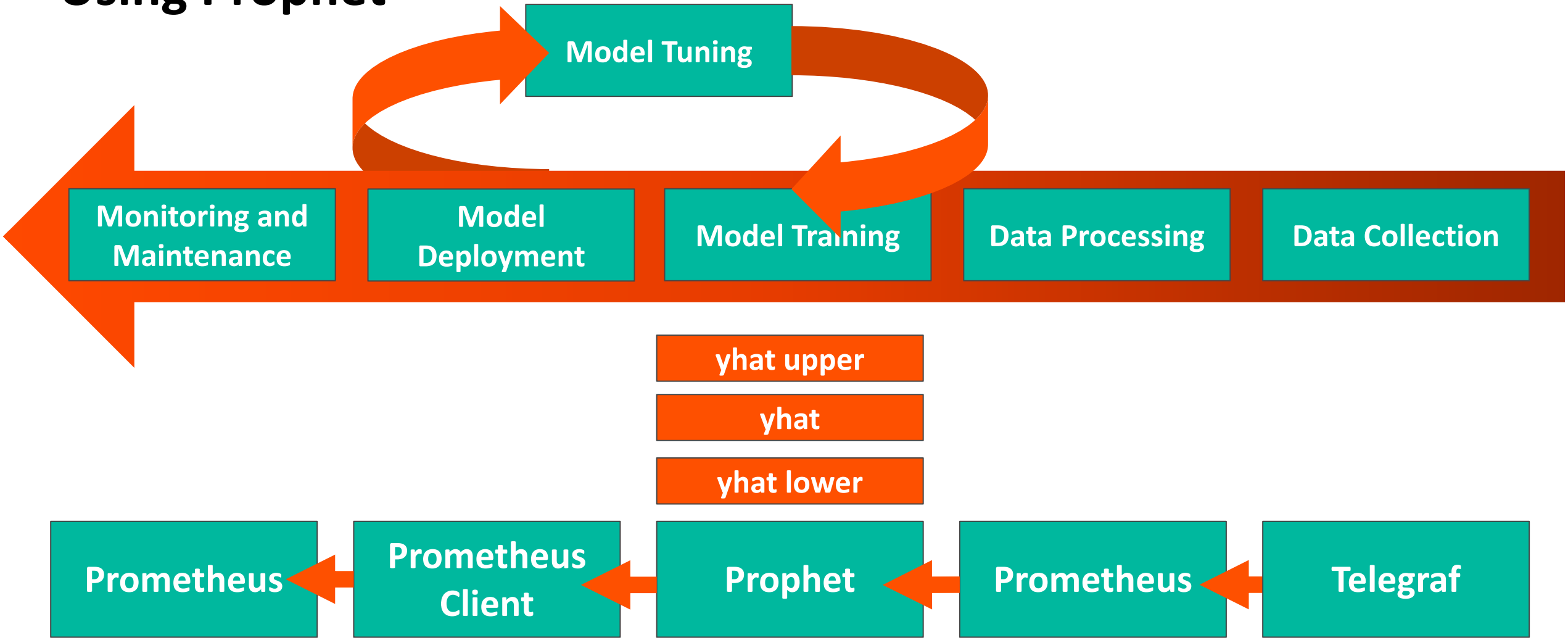
Prometheus Client

Defines an interface for the metric output

Gets metrics from the target via the network

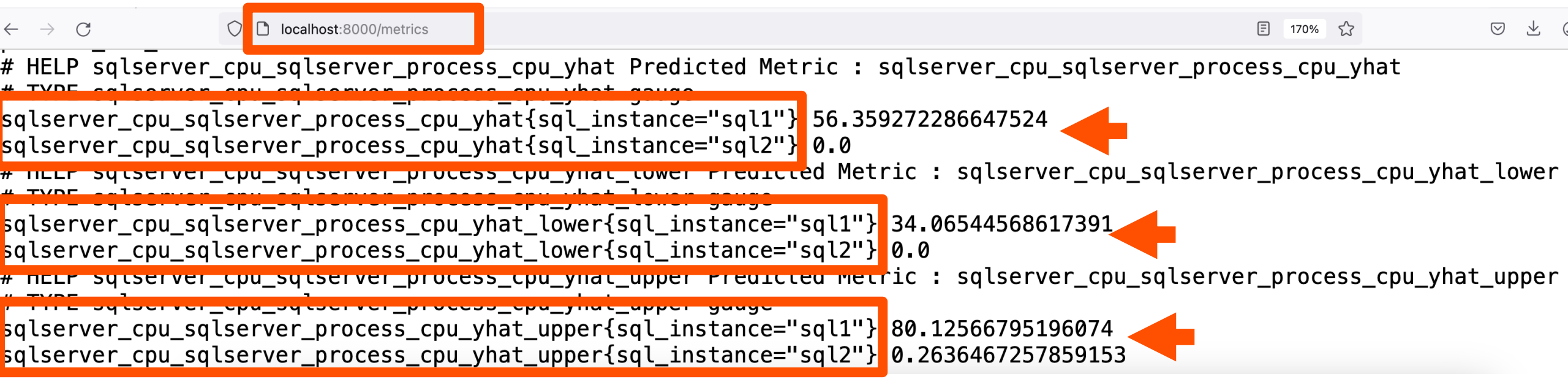
Generally, HTTP/HTTPS

Using Prophet



MetricsML – Bringing it all together

Use the Prometheus Client to expose our predicted values back



Let’s hop into a demo

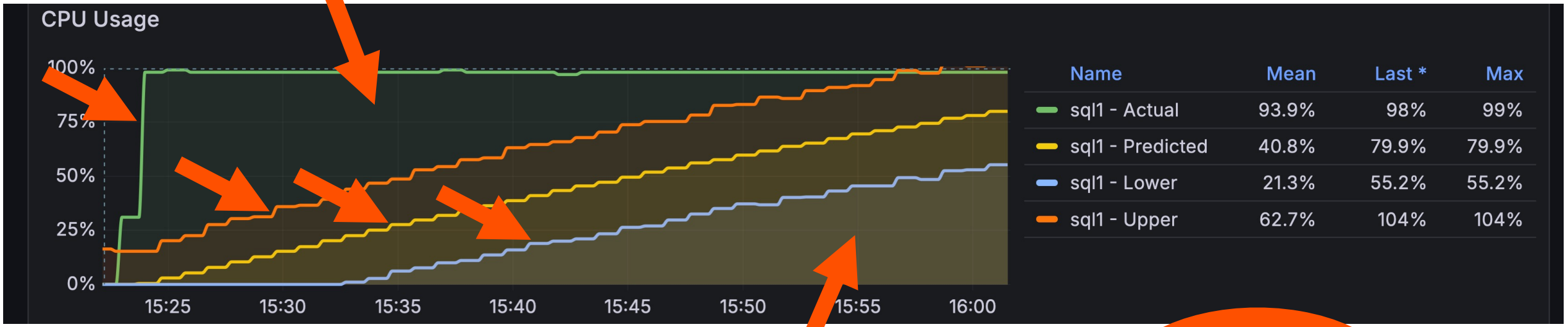
Getting data into Prophet from Prometheus using Python

Build some visualizations in Grafana

What's happening here?

Too high

When is too high an issue?

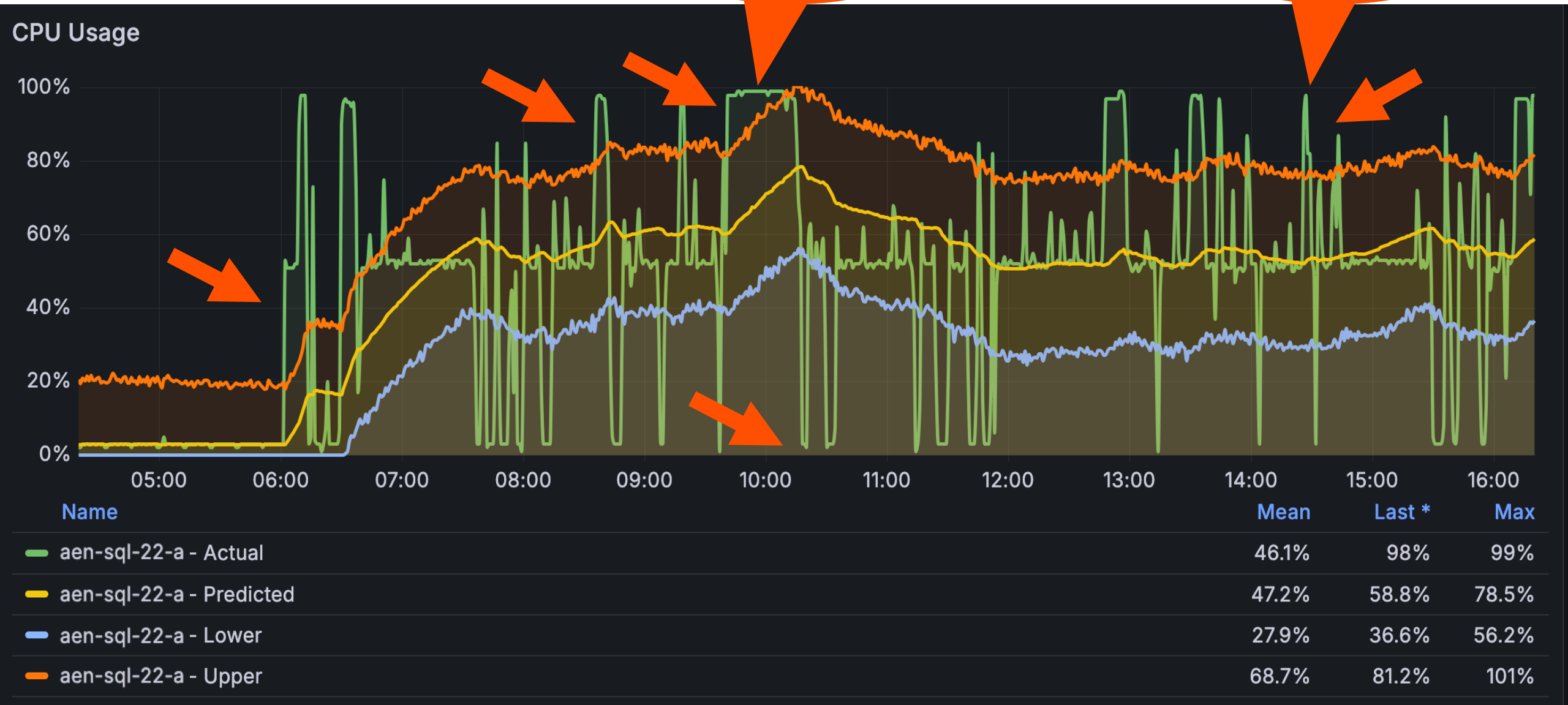


Too low

When is too low an issue?

Where are the anomalies?

Example Chart

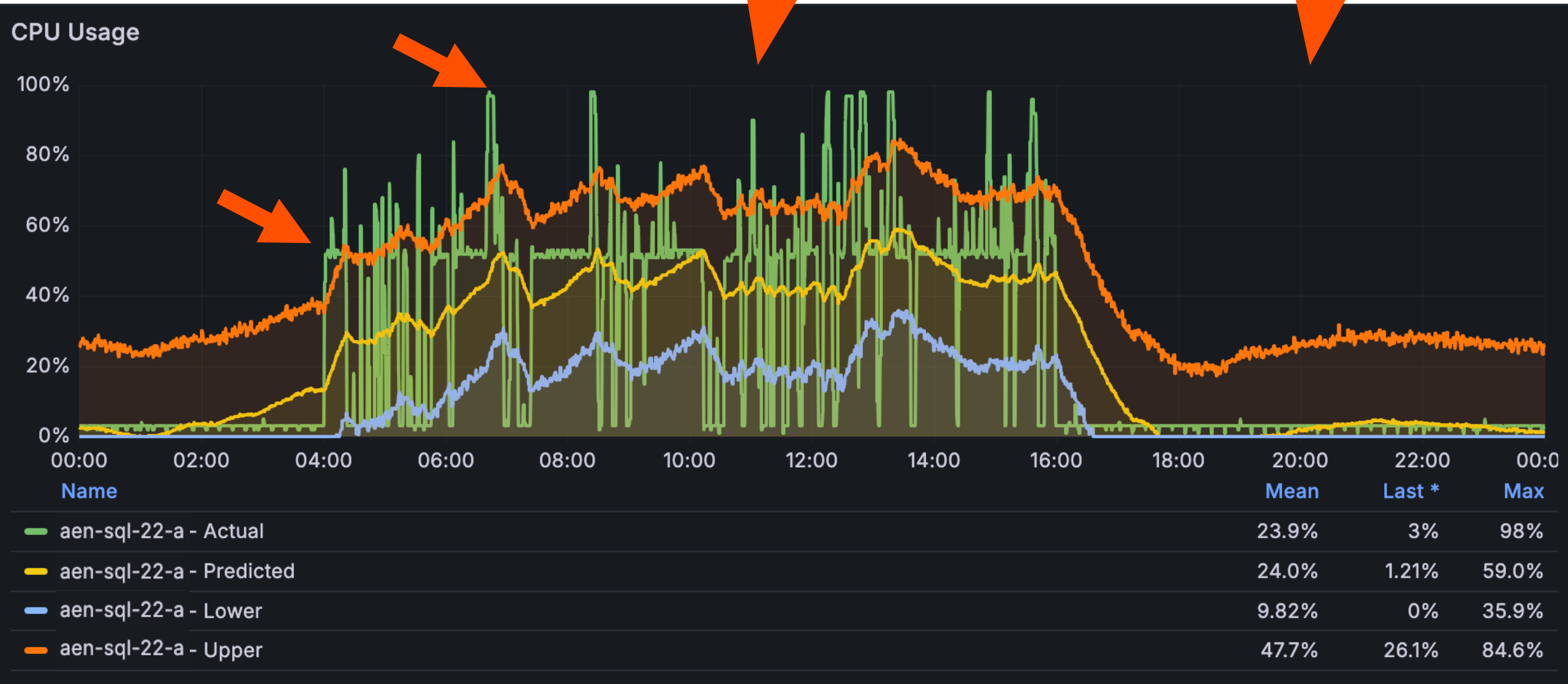


Name	Mean	Last *	Max
aen-sql-22-a - Actual	46.1%	98%	99%
aen-sql-22-a - Predicted	47.2%	58.8%	78.5%
aen-sql-22-a - Lower	27.9%	36.6%	56.2%
aen-sql-22-a - Upper	68.7%	81.2%	101%

Example Chart

A little more tuning is needed

Delayed alerting can help



Name	Mean	Last *	Max
aen-sql-22-a - Actual	23.9%	3%	98%
aen-sql-22-a - Predicted	24.0%	1.21%	59.0%
aen-sql-22-a - Lower	9.82%	0%	35.9%
aen-sql-22-a - Upper	47.7%	26.1%	84.6%

Next Steps…

- More tuning in Prophet
- Lots of tuning
- Generalize for all metrics
- Parallelize the model training/prediction
- Configure alerting

Review

- Introduction to Telegraf, Prometheus and Grafana
- Building an Anomaly Detection Model
- Practical Use Cases for Anomaly Detection

THANK YOU

Please review this session

Session Review

April 8-11, 2024