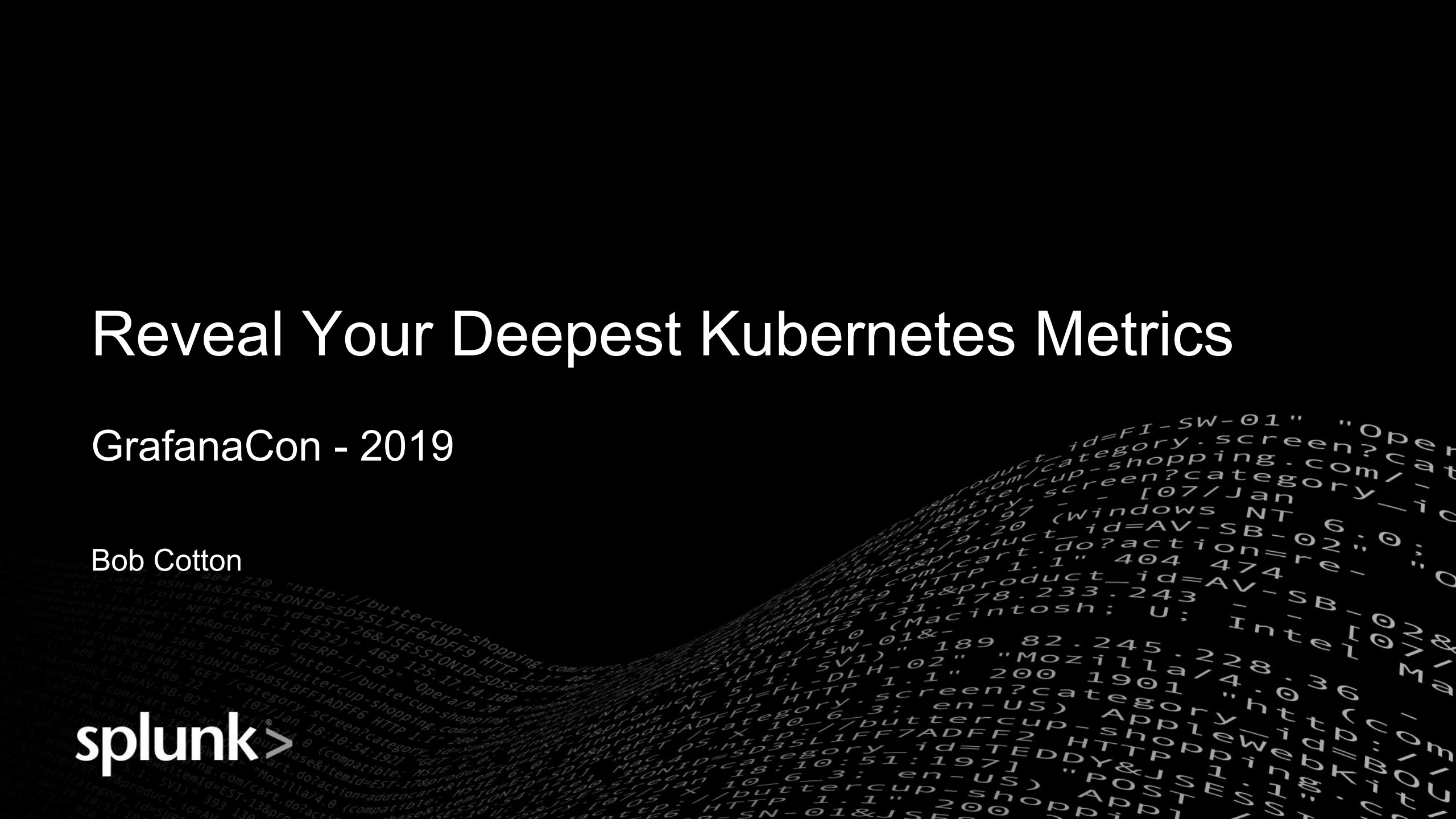
Reveal Your Deepest Kubernetes Metrics

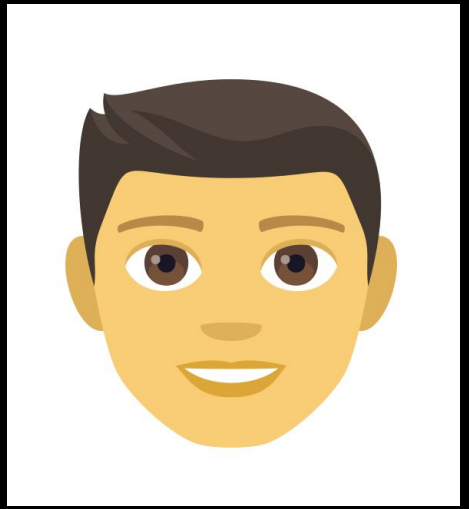
GrafanaCon - 2019

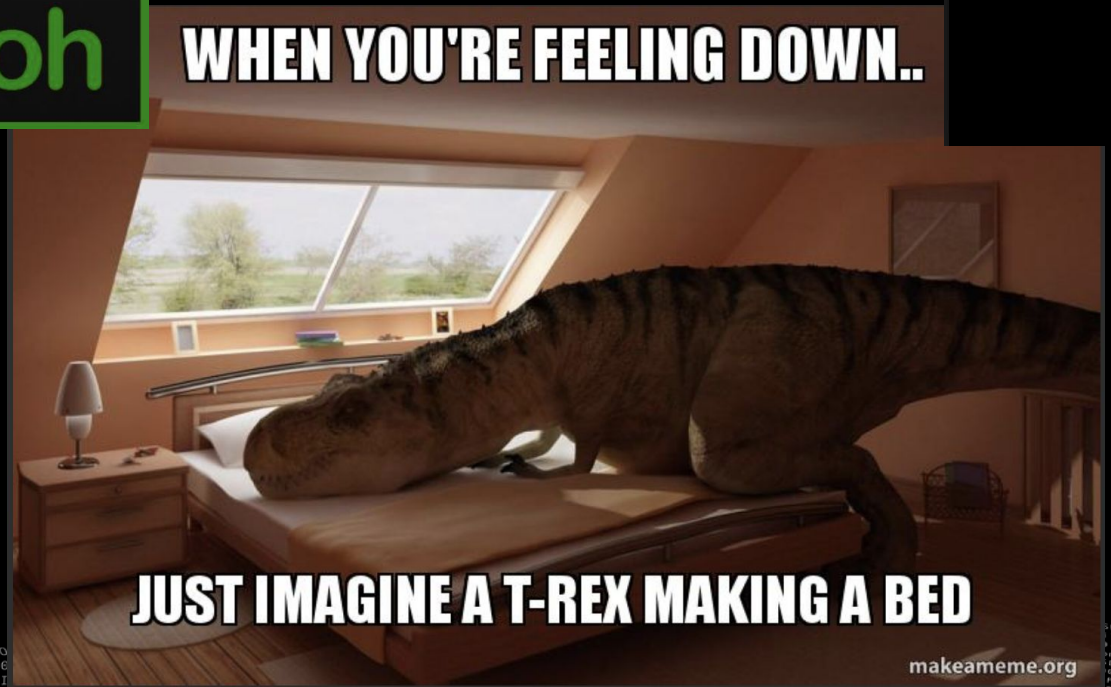
Bob Cotton



About Me

- Senior Principal Engineer - Splunk Inc.
- Working with data systems for 20+ years
  - FreshTracks.io, Rally Software
- @bob_cotton
- Father, Fly Fisher and Avid Homebrewer





> 275,000

Unique Series

10 Node Cluster

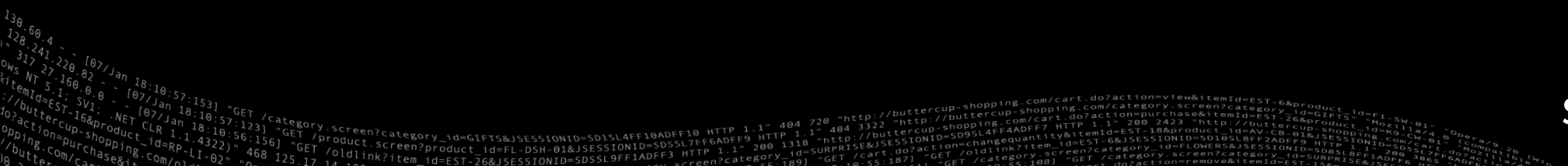
160 Containers

What are the Important Metrics?

Ways to approach all metrics

Four Golden Signals

- Latency
  - The time it takes to service a request.
- Errors
  - The rate of requests that fail, either explicitly, implicitly, or by policy
- Traffic
  - A measure of how much demand is being placed on your system
- Saturation
  - How "full" your service is.

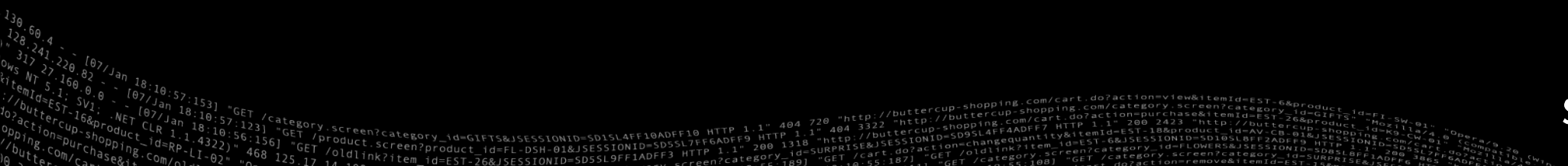


USE Method

- Introduced by Brendan Gregg for reasoning about system resources
  - Resources are all physical server functional components (CPUs, disks, busses…)
- Utilization
  - The average time that the resource was busy servicing work
- Saturation
  - The degree to which the resource has extra work which it can't service, often queued
- Errors
  - The count of error events

RED Method

- Introduced by Tom Wilkie
  - A subset of the Four Golden Signals for measuring Services
- Rate
  - The number of requests per second
- Errors
  - The number of errors per second
- Duration
  - The length of time required to service the request

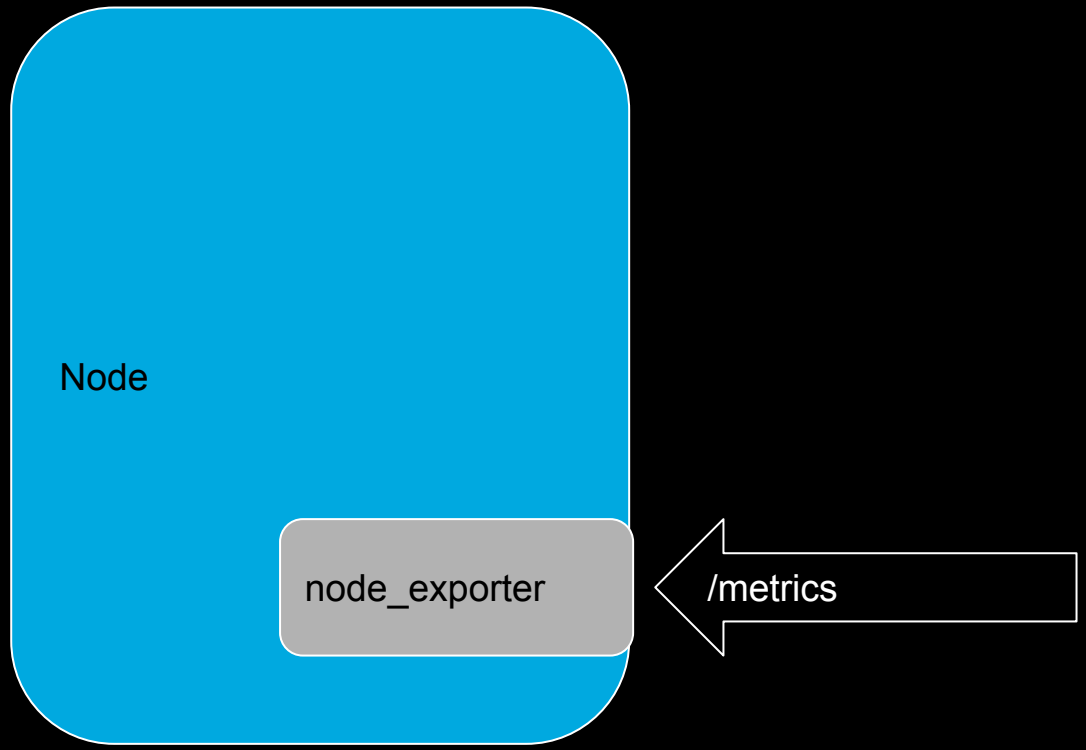


USE is for Resources
RED is for Services

Kubernetes Has Both!

Node Metrics from node_exporter

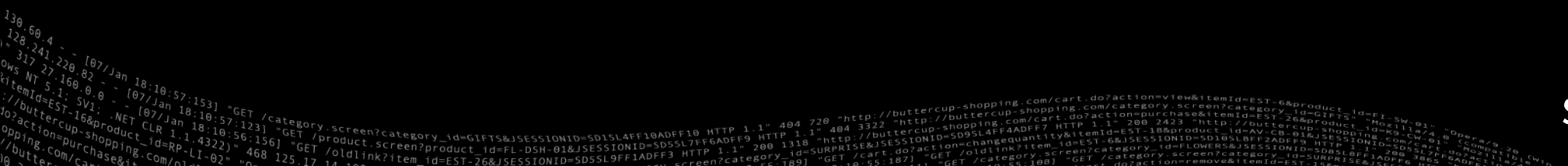
- node_exporter installs as a DaemonSet
  - One instance per node
- Standard Host Metrics
  - Load Average
  - CPU
  - Memory
  - Disk
  - Network
  - Almost anything in /proc
- ~1000 Unique series for a typical node



Nodes are a Resource - USE

Applied per-Node and per-Cluster

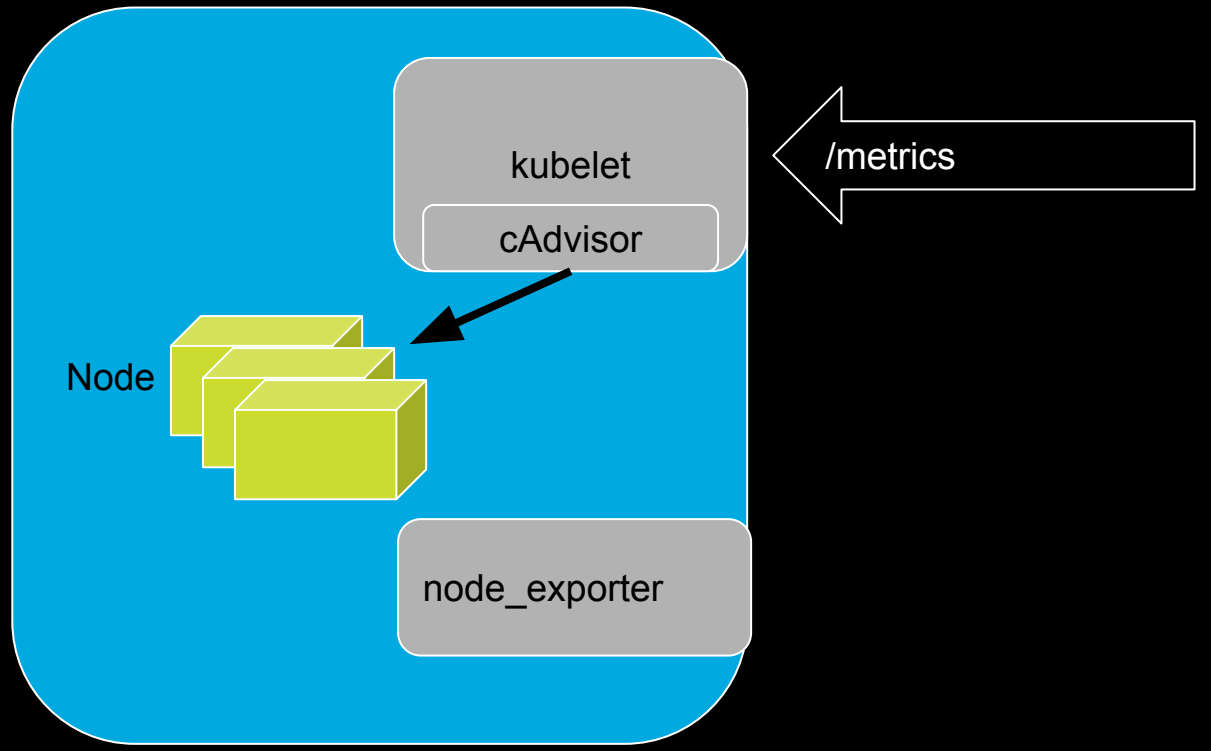
	Utilization Metrics	Saturation Metrics	Errors
CPU	node_cpu_seconds	node_load1 node_cpu_seconds_total	
Memory	node_memory_MemFree_bytes node_memory_MemCached_bytes node_memory_Buffers_bytes node_memory_MemTotal_butes	node_vmstat_pgpgin node_vmstat_pgpgout	
Disk IO	node_disk_io_time_seconds_total	node_disk_io_time_weighted_seconds_total	
Disk Usage	node_filesystem_size_bytes node_filesystem_avail_bytes		



splunk> listen to your data

Container Metrics from cAdvisor

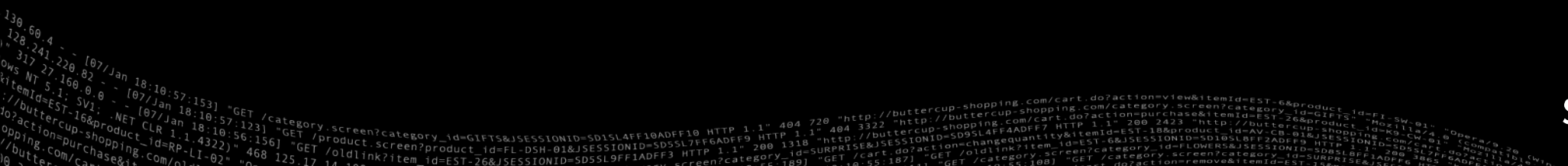
- cAdvisor is embedded in the kublet
- Each container reports:
  - CPU Usage and throttled
  - Filesystem read/writes/limits
  - Memory usage and limits
  - Network transmit/receive/dropped



Containers are a Resource - USE

Applied per-Node and per-Cluster

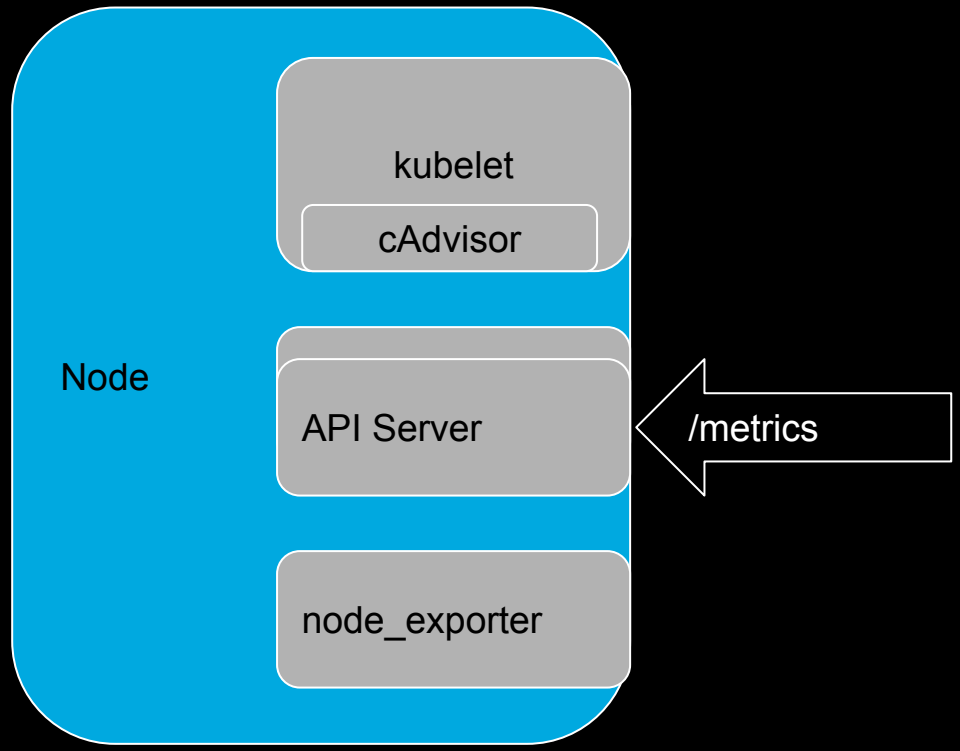
	Utilization Metrics	Saturation Metrics	Errors
CPU	container_cpu_usage_seconds_total	container_cpu_usage_seconds_total kube_pod_container_resource_requests_cpu_cores kube_pod_container_resource_limits_cpu_cores	
Memory	container_memory_usage_bytes**	container_memory_usage_bytes kube_pod_container_resource_requests_memory_bytes kube_pod_container_resource_limits_memory_bytes	container_memory_failcnt container_memory_failures_total



splunk> listen to your data

Kubernetes Metrics from the K8s API Server

- Metrics about the performance of the K8s API Server
  - Performance of controller work queues
  - Request Rates and Latencies
  - Etcd helper cache work queues and cache performance
  - General process status
    - (File Descriptors/Memory/CPU Seconds)
  - Golang status (GC/Memory/Threads)



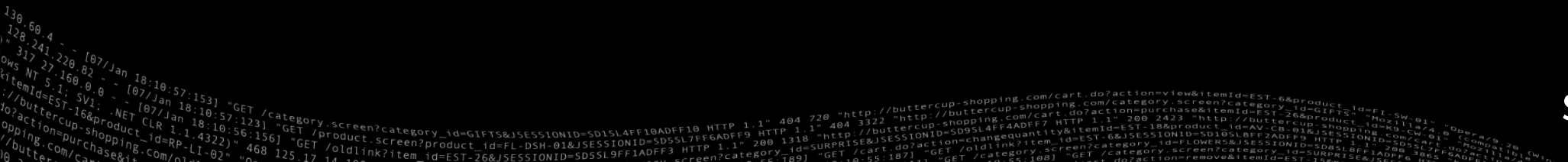
The API Server is a Service - RED

Applied per-Node and per-Cluster

Rate	Error	Duration
apiserver_request_count	apiserver_request_count{code=~"^(?:5..)$"}	apiserver_request_latencies_bucket

K8s Derived Metrics from kube-state-metrics

- Counts and metadata about many K8s types
  - Counts of many “nouns”
  - Resource Limits
  - Container states
  - ready/restarts/running/terminated/waiting
- *_labels series carries labels
  - Series has a constant value of 1
  - Join to other series for on-the-fly labeling using `left_join`



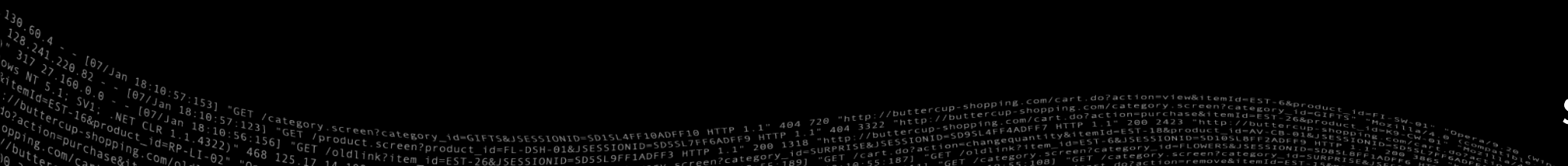
Etcd Metrics from etcd - RED

- Etcd is “master of all truth” within a K8s cluster
  - Leader existence and leader change rate
  - Proposals committed/applied/pending/failed
  - Disk write performance
  - Inbound gRPC stats

Rate	Error	Duration
etcd_http_received_total	etcd_http_failed_total	etcd_http_successful_duration_seconds_bucket

So Many Metrics

- Kubernetes Scheduler Metrics
- Kubernetes Proxy Metrics
- Admission Controller Metrics
- Istio Metrics

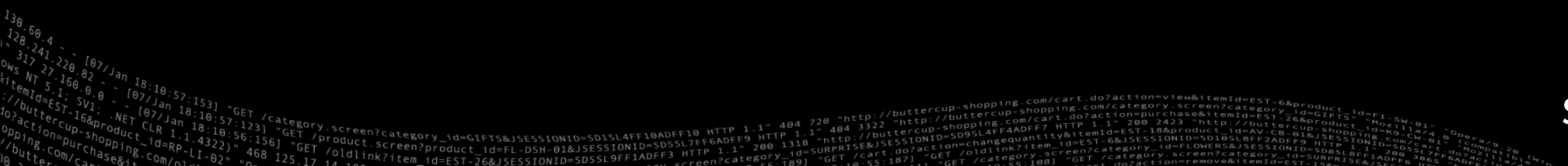


Tooling

Don’t do this by hand!

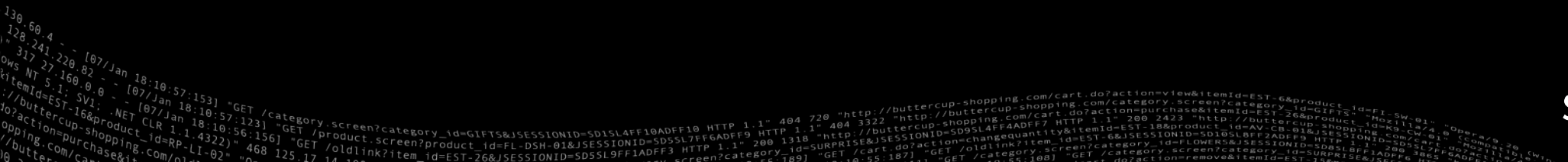
Prometheus Operator

- The Prometheus Operator from CoreOS
  - Prometheus
  - Alert Manager
  - Grafana
  - Custom Resource Definitions for Prometheus primitives



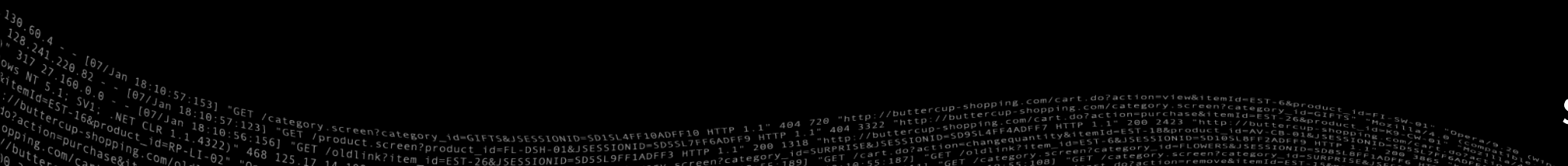
Monitoring Mixins

- Packaged monitoring configurations
  - Recording Rules (prometheus)
  - Dashboards (grafana)
  - Alerting Rules (prometheus)
- Written in jsonnet, adaptable to your environment
- Available for many projects:
  - Kubernetes
  - etcd
  - Consul
  - Vault
- Community maintained...

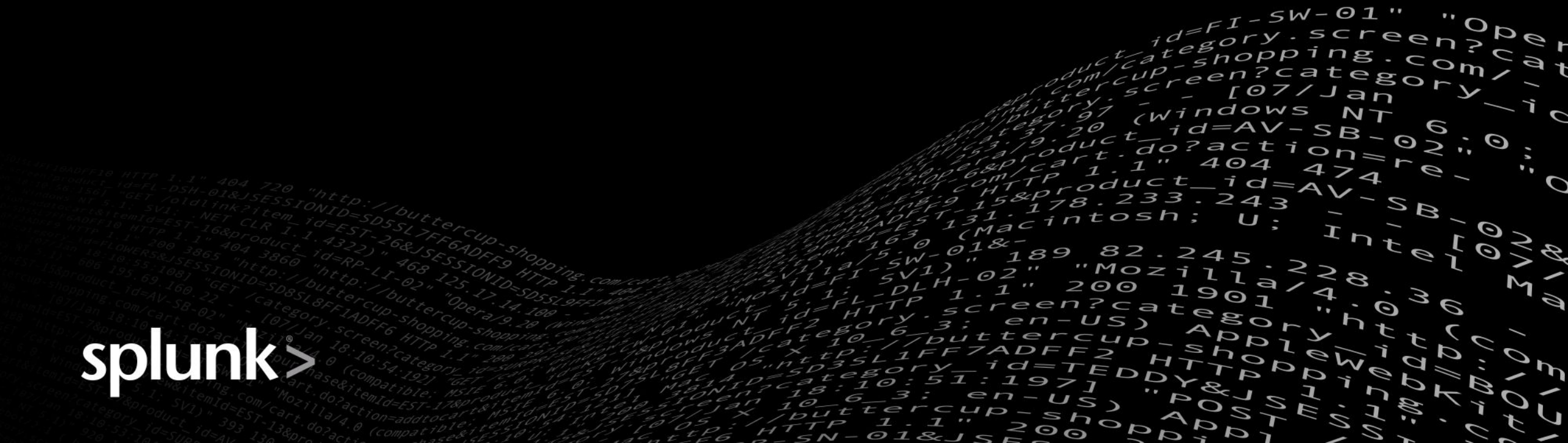


Kubernetes Metrics Overhaul

- Many metrics will be renamed
  - Consistency for naming and labelling
- Old metrics will be deprecated in 1.14
  - Removed in 1.15
- Kubernetes monitoring mixin will be updated
  - Another reason it use mixins!



Thank You!



Resources

- A Deep Dive into Kubernetes Metrics
- Everything you need to know about monitoring mixins
- Kubernetes Metrics Overhaul
-