



Cutting Climate Costs with Kubernetes

Shiva Rezaie & Steve Francis, Sidero Labs







What does Sidero Labs do?

Talos Linux



What does Sidero Labs do?

Talos Linux



Sidero Metal



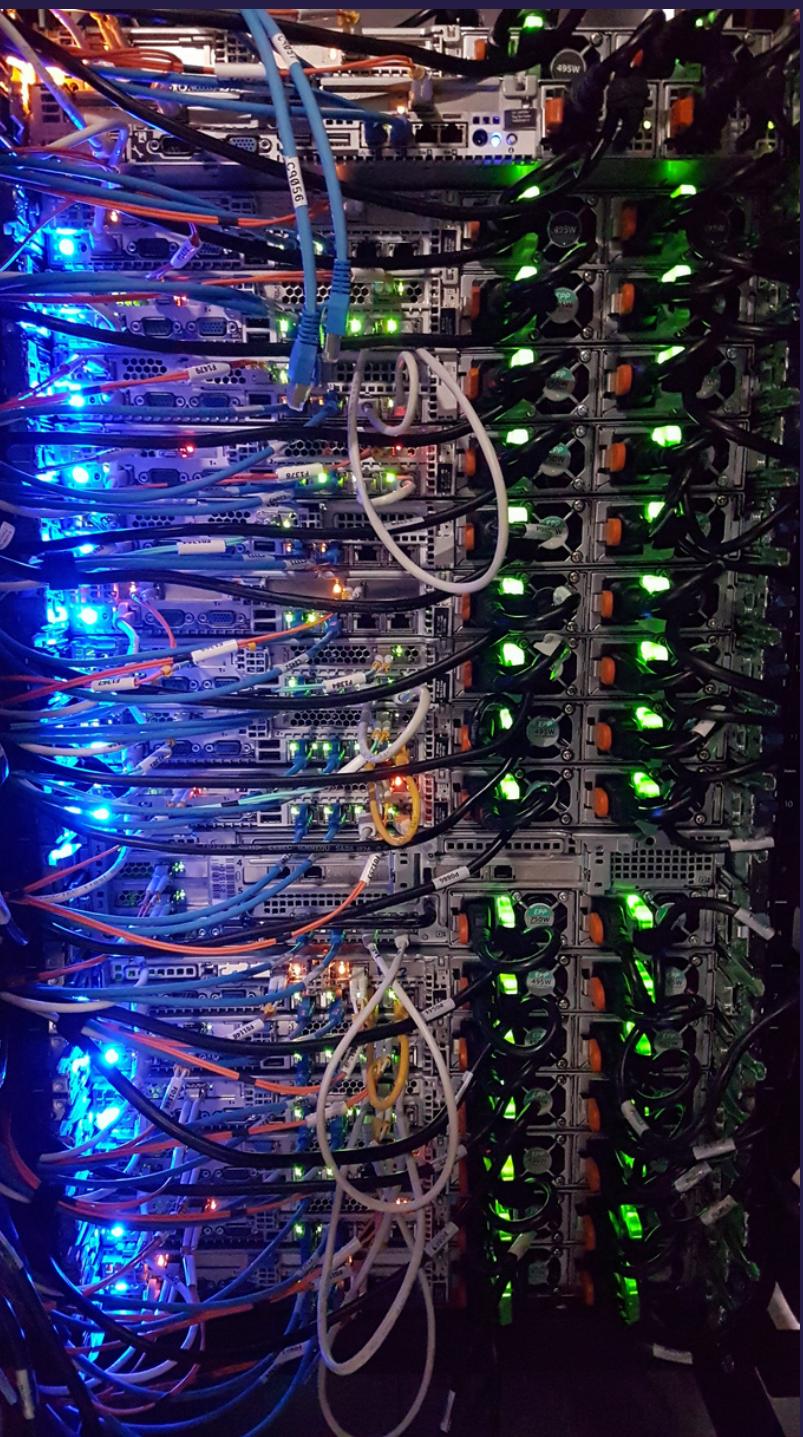
CAPI provider for bare metal

Sidero Omni



**SaaS for BYO compute K8s
For edge and bare metal**

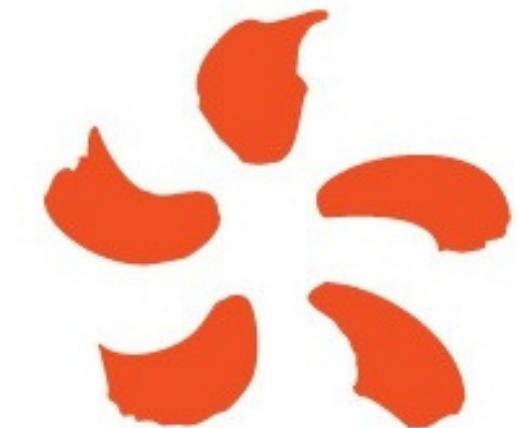
The Problem



NOKIA



EQUINIX



powerflex
EDF renewables

civo



adopt MX

Existing Scaling Options

Horizontal Pod AutoScaler

Vertical Pod AutoScaler

Cluster AutoScaler

Existing Scaling Options

Horizontal Pod AutoScaler

Vertical Pod AutoScaler

Cluster AutoScaler

Existing Scaling Options

Horizontal Pod AutoScaler
Vertical Pod AutoScaler
Cluster AutoScaler

An emissions aware Scheduler



An emissions aware Scheduler



Run all the workloads!

Only run high priority
workloads

Scheduler Logic



Schedule pods



Pod Manager Evicts
pods



Node Manager powers
on/off nodes

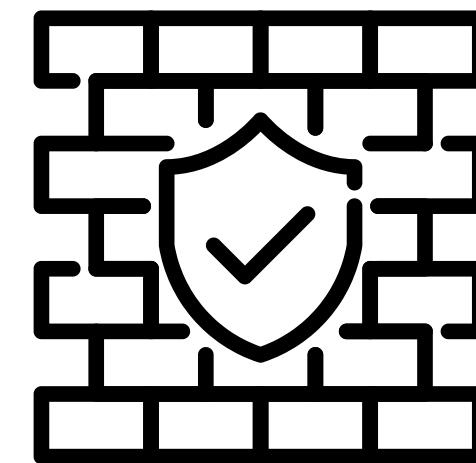
Scheduler Extension points

We only use the PreFilter

preFilter: used to pre-process or check information about a Pod or the cluster before filtering. They can mark a pod as unschedulable.



Single K8s Cluster



Bare metal Workers with IPMI



Control plane in Vultr

KubeSpan linking all nodes

Lets deploy!

From

<https://github.com/siderolabs/kube-scheduler>

- Create a WattTime account
- Create PriorityClass
- Annotate nodes with BMC info
- Deploy the scheduler
- Create pods with priorityClassName referencing the desired PriorityClass, and schedulerName set to kube-scheduler-siderolabs

Create WattTime account

WattTime.org

The screenshot shows the WattTime.org homepage. At the top, there's a navigation bar with links: WHAT WE DO, LEARN MORE, GET THE DATA, NEWS & INSIGHTS, and ABOUT. The main headline reads "The Power to Choose Clean Energy". Below it, a sub-headline states: "WattTime is a nonprofit that offers technology solutions that make it easy for anyone to achieve emissions reductions without compromising cost, comfort, and function." To the right, there's a data visualization for Chicago, IL, showing a gauge with a value of 68. The gauge has a color gradient from green to red, with 68 being a green value. The ZIP CODE input field contains "60616". The background features a stylized graphic of electrical transmission towers.

Create account at

<https://www.watttime.org/api-documentation/#register-new-user>

Create Priority Class

```
apiVersion: scheduling.k8s.io/v1
kind: PriorityClass
metadata:
  name: high-priority
value: 100
globalDefault: false
description: "This priority class is an example."
---
```

```
apiVersion: scheduling.k8s.io/v1
kind: PriorityClass
metadata:
  name: low-priority
value: 0
globalDefault: false
description: "This priority class is an example."
```

Apply Priority Class

```
[I] KubeCon 2023 > kubectl apply -f 02_priorities.yaml
priorityclass.scheduling.k8s.io/high-priority configured
priorityclass.scheduling.k8s.io/low-priority configured
[I] KubeCon 2023 > █
```

Annotate the nodes with BMC info

```
[root@kubecore ~]# KubeCon 2023 > kubectl edit node talos-c85-mm0
```

Edit Daemonset to add WattTime credentials

I

Apply the Scheduler DaemonSet

```
[1] KubeCon 2023 > kubectl apply -f 01_daemonset.yaml  
serviceaccount/kube-scheduler-siderolabs unchanged
```

What it does

Scheduler

Get the Emissions index (0-100)

Any pods to run with priority > Emissions? If so, schedule them

Pod Manager

Any pods running with priority less than the emissions? If so, evict them

Node Manager

And pods in the scheduling queue that have passed the PreFilter?

If so, turn on another node

Any nodes with no pods? If so, turn it off.

Apply the scheduler and Priority to a deployment

[N] KubeCon 2023 > vim 03_high-priority-example.yaml

Deploy a High Priority Deployment

✗ fish /home/andrewrynhard/kube-scheduler/hack (ssh)

[I] KubeCon 2023 > █

✗ watch kubectl get nodes /home/andrewrynhard (ssh)

Every 2.0s: kubectl get nodes

NAME	STATUS	ROLES	AGE	VERSION
talos-2zu-zv8	Ready	omni-agent	4d20h	v1.28.3
talos-4u9-512	Ready	control-plane	11d	v1.28.3
talos-c85-mn0	NotReady,SchedulingDisabled	<none>	8d	v1.28.3

A low Priority Deployment

[I] KubeCon 2023 > |

Type Status
PodScheduled False

Volumes:

kube-api-access-jxm2d:

Type: Projected (a volume that contains injected data from multiple sources)
TokenExpirationSeconds: 3607
ConfigMapName: kube-root-ca.crt
ConfigMapOptional: <nil>
DownwardAPI: true
Class: BestEffort
<none>
Node-Selectors: node.kubernetes.io/not-ready:NoExecute op=Exists for 300s
node.kubernetes.io/unreachable:NoExecute op=Exists for 300s

Events:

Type	Reason	Age	From	Message
--	-----	----	----	-----
Warning	FailedScheduling	16s	kube-scheduler-siderolabs	0/3 nodes are available: pod priority (0) lower than index (44)

KubeCon 2023 > ■

Change high priority to low - eviction

× fish /home/andrewrynhard/kube-scheduler/hack (ssh)

[N] KubeCon 2023 > |

Change high priority to low - Node powered off

```
> watch kubectl get nodes /home/andrewrynhard/kube-scheduler (ssh)
```

```
Every 2.0s: kubectl get nodes
```

NAME	STATUS	ROLES	AGE	VERSION
talos-2zu-zv8	Ready	omni-agent	4d22h	v1.28.3
talos-4u9-512	Ready	control-plane	11d	v1.28.3
talos-c85-mm0	Ready	<none>	8d	v1.28.3

Cheap “Bin Packing”

High priority pods will (in the simple case) tend to run on the same node with the
MostAllocated strategy

This helps make nodes completely free when low priorities are shut down.

Things to Improve:

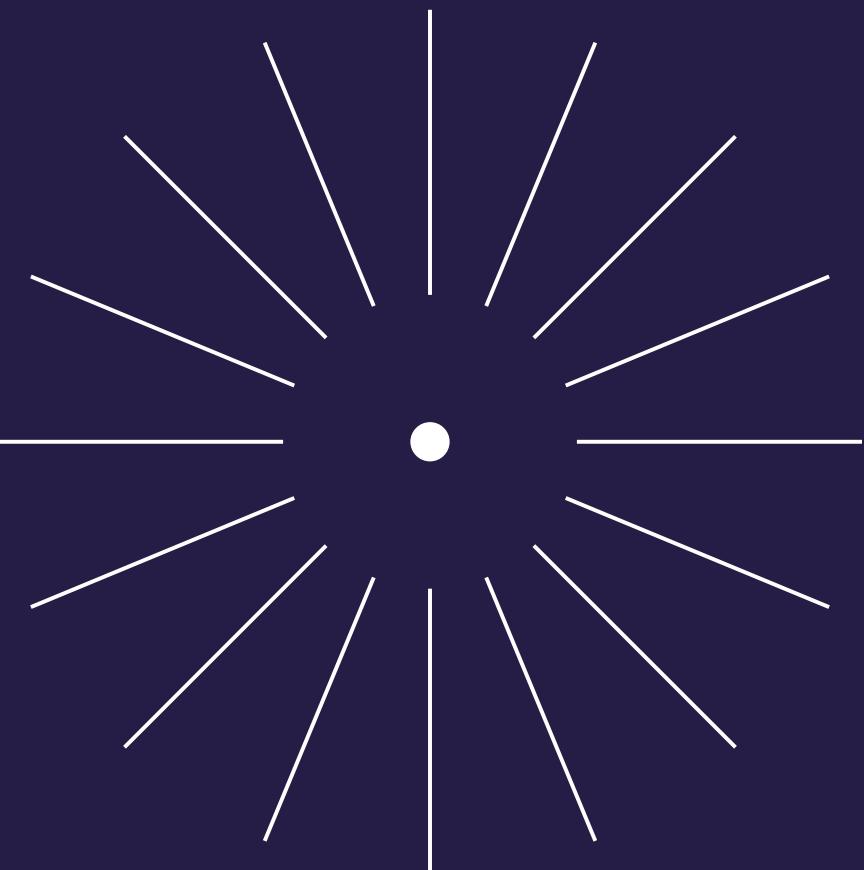
- Dynamic bin packing, to rearrange pods to free up more nodes
- Full Integration with Omni, Cluster Autoscaler, CAPI, and Sidero Metal
- Your ideas!

Use Cases

Bare metal workloads with periodic demand, or classes of Deployments that are not time critical

Questions

???



Project at <https://github.com/siderolabs/kube-scheduler>

Omni trial: <https://signup.siderolabs.io/>

Email info@siderolabs.com and say you want the KubeCon 4 week trial.

Leave Session Feedback



Come to our Customer Mixer now!

Simone's Bar

960 W 18th St, Chicago, IL 60608

6:30 pm – 8:30 pm.

**Come meet folks from Sidero Labs, Talos Linux users,
Chat, drink, food, and have a good time!**