

BigData Spark Hands-On 3 [part 1]

Spark on k8s
Podman Install

arnaud.nauwynck@gmail.com

Esilv 2024

Outline

Install podman (~Docker) on local Windows/Mac/Linux

Install Kubernetes Cluster "Kind" (=K8s in Docker)

Install Kubernetes Tools: k9s, helm, aliases

Install spark-operator for K8s using Helm + check/fixes

part 2/2

(Optional) Launch spark-shell on Docker, in mode "--master local[*]"

(Optional) Launch spark with "spark-submit --master k8s://localhost:port"

Launch spark with "kubectl apply -f file.yaml"

(Optional... and backup if local install fails) Using Azure Kubernetes

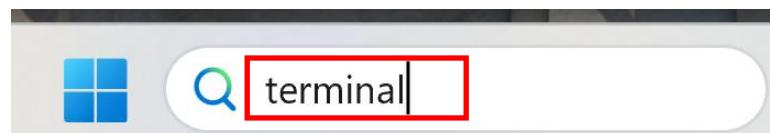
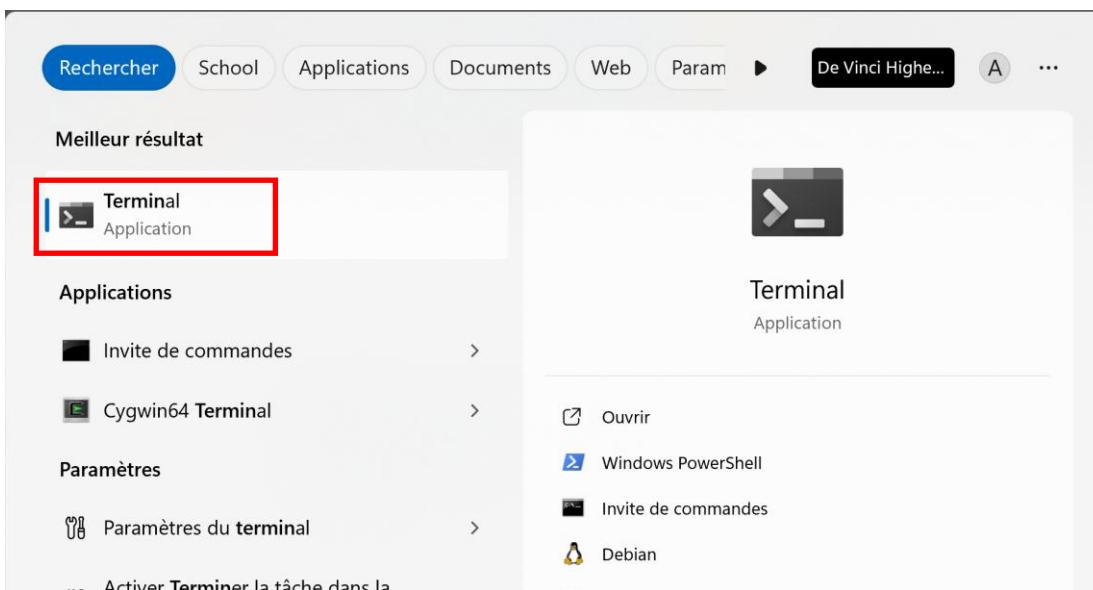
Pre-Requisites: install K8s
(even on Windows, with podman + Kind)

For Windows, Install a modern "Terminal"
(do not use "cmd", nor "Power Shell)

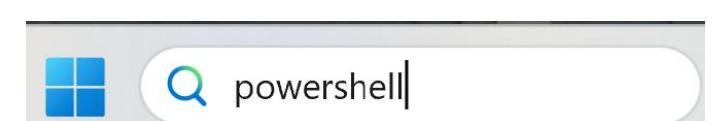
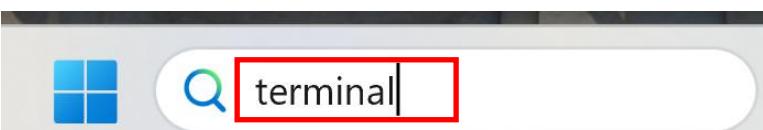
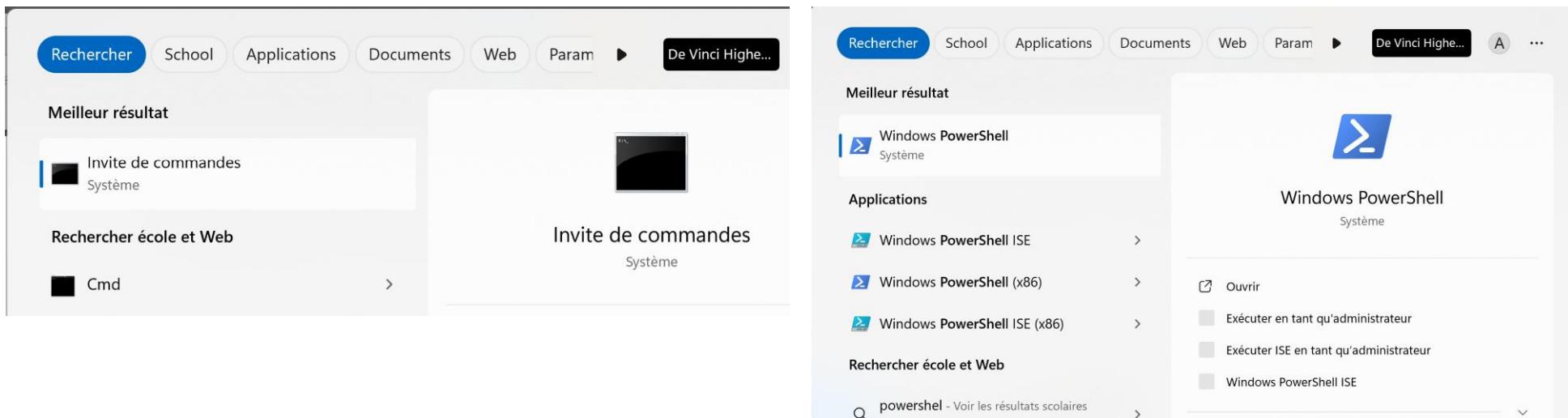
winget install Microsoft.WindowsTerminal

```
C:\apps\spark>winget install Microsoft.WindowsTerminal
Un package existant a déjà été installé. Tentative de mise à niveau du package installé...
Mise à niveau disponible introuvable.
Aucune version de package plus récente n'est disponible à partir des sources configurées.
```

Run Terminal

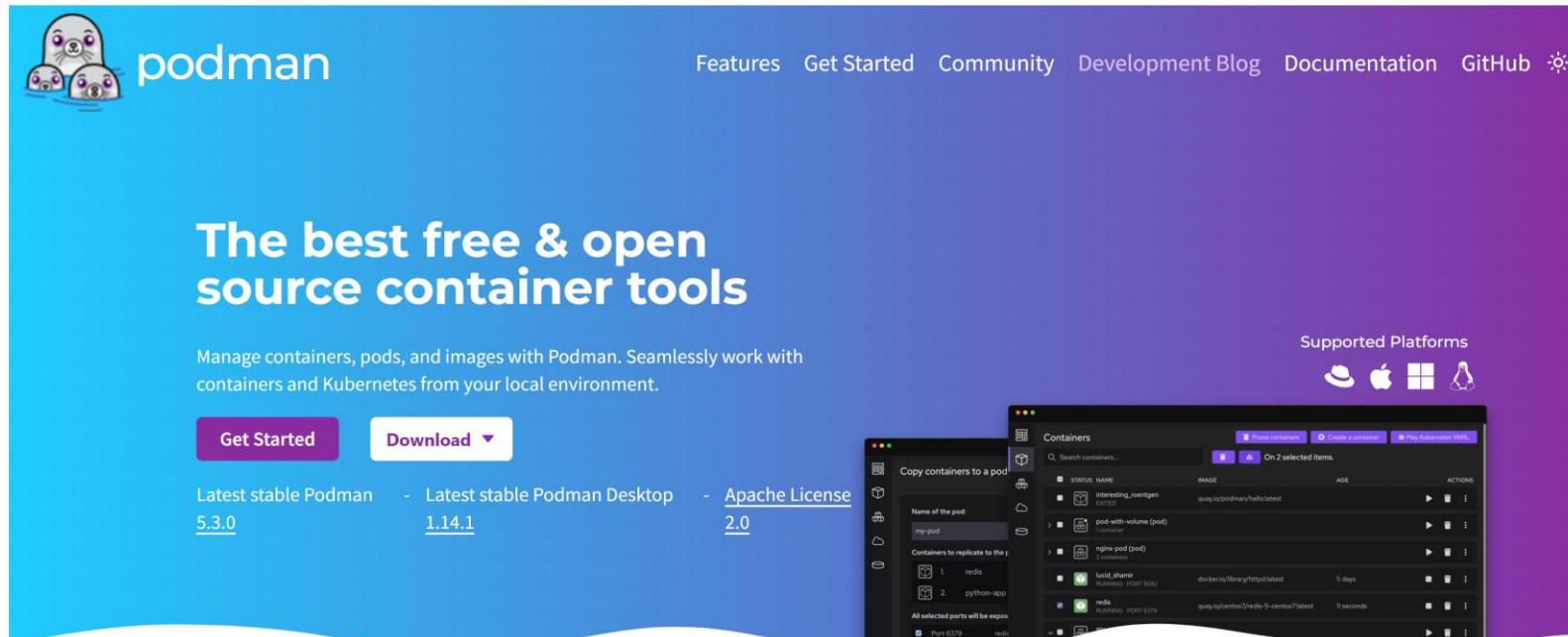


Terminal != Cmd Prompt (Invite de Command), != PowerShell



<https://podman.io/>

Podman is a replacement for "Docker" runtime, for Linux/Windows/Mac/, without root access
Podman also supports tools for easy Kubernetes interaction (minikube / kind / ..)



Download + exec "podman-{version} .setup.exe"

<https://github.com/containers/podman/releases>

The screenshot shows the GitHub releases page for the Podman repository. The URL in the address bar is `github.com/containers/podman/releases`. Below the address bar, there is a search bar and a download button. The main content area is titled "Assets" and shows 18 items. The assets listed are:

Asset	Size	Published
podman-5.3.0-setup.exe	30.5 MB	3 days ago
podman-installer-macos-amd64.pkg	55.8 MB	3 days ago
podman-installer-macos-arm64.pkg	54.3 MB	3 days ago
podman-installer-macos-universal.pkg	79.5 MB	3 days ago
podman-machine.aarch64.applehv.raw.zst	968 MB	2 days ago
podman-machine.aarch64.hyperv.vhdx.zst	968 MB	2 days ago
podman-machine.aarch64.qemu.qcow2.zst	967 MB	2 days ago
podman-machine.x86_64.applehv.raw.zst	1010 MB	2 days ago
podman-machine.x86_64.hyperv.vhdx.zst	1010 MB	2 days ago
podman-machine.x86_64.qemu.qcow2.zst	1010 MB	2 days ago
Source code (zip)		4 days ago
Source code (tar.gz)		4 days ago

Below the assets, there is a link "Show all 18 assets". At the bottom of the page, there are reaction icons for smiley, thumbs up, heart, and star, followed by the text "21 people reacted".

last week



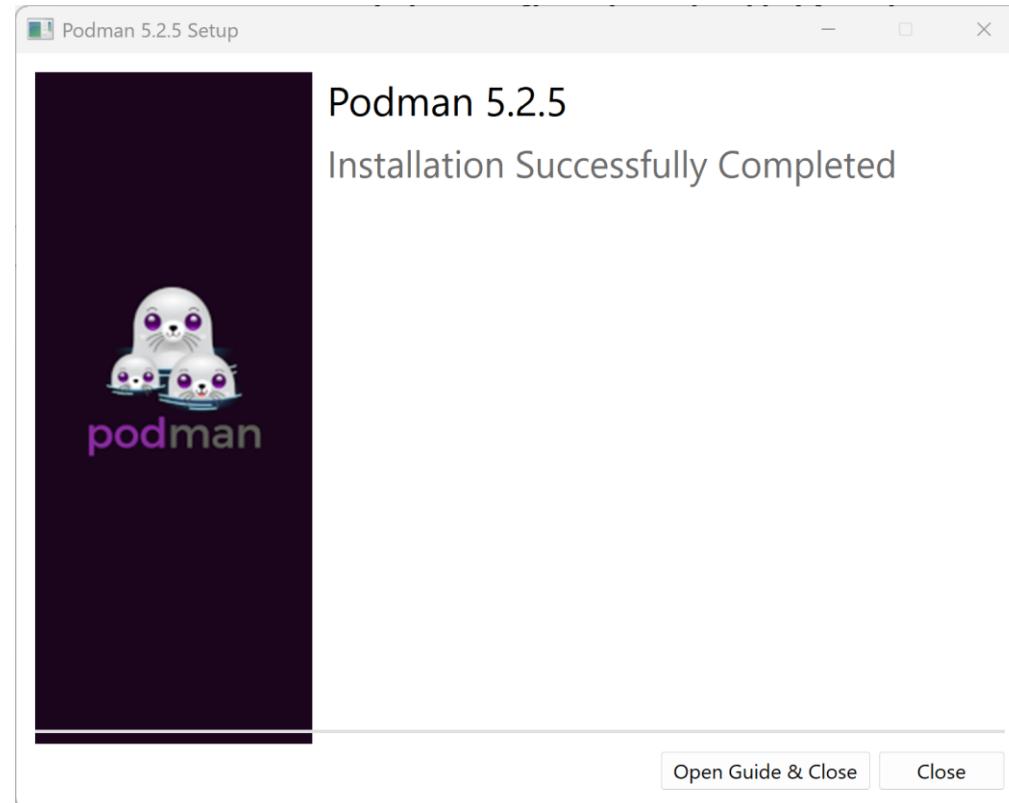
v5.3.0-rc3

v5.3.0-RC3

Pre-release

Features

Podman Installed



Launch & Explore interface



Podman Desktop

Dashboard

Learning Center

Working with Kubernetes

Understand how Podman's design philosophy can handle pods just as it handles containers, in consistent way with Kubernetes architectures.

Get started

podman

Podman Tutorial Zero to Hero

Full 1 Hour Course to learn everything you need to know about Podman and get the most of it.

Get started

Podman v5.2.5

RUNNING

podman-machine-default

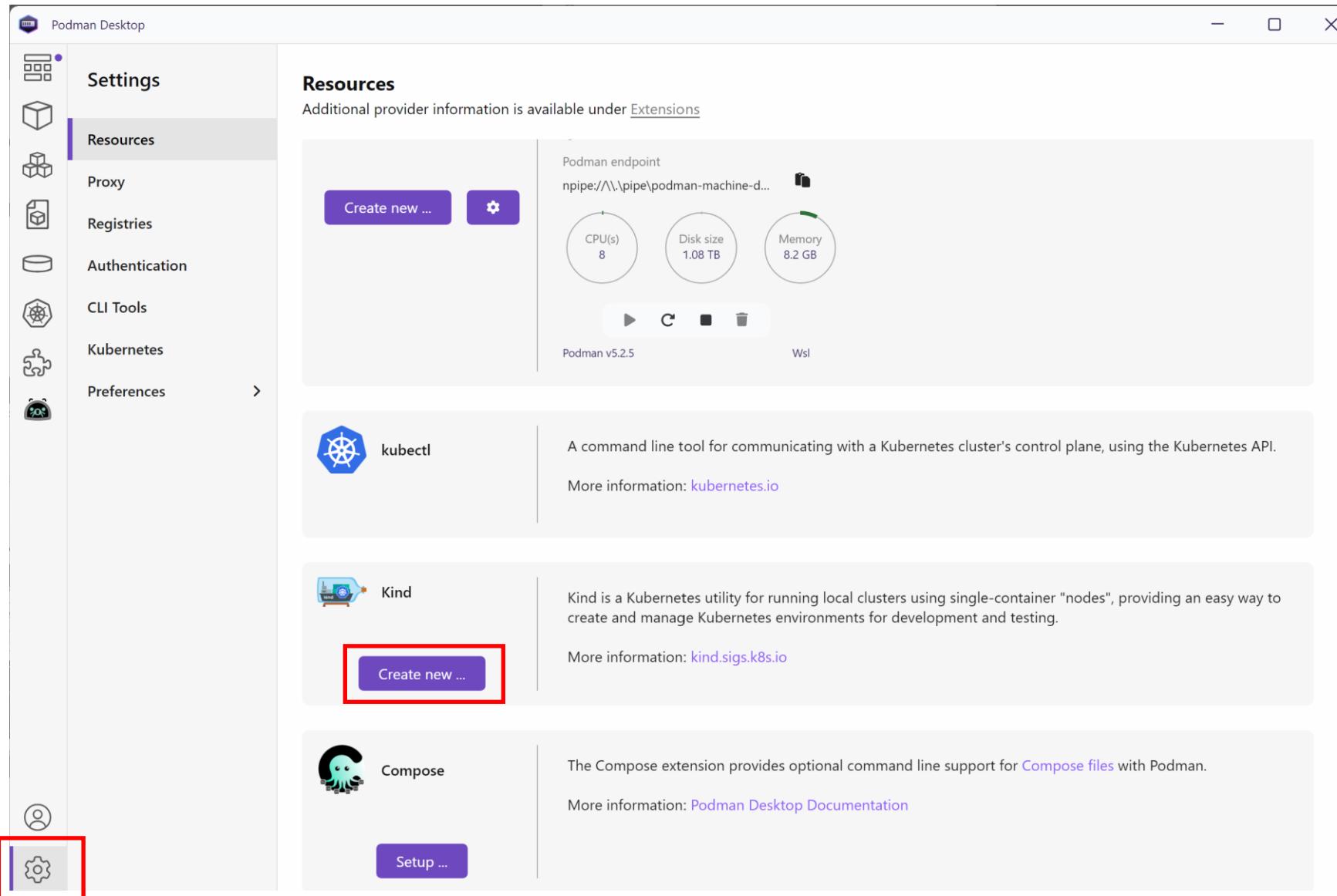
Website Installation guide Docker compatibility guide Join the community

kind-kind-cluster AI Lab API listening on port 10434

v1.14.1

A screenshot of the Podman Desktop application window. The title bar says "Podman Desktop". The left sidebar has icons for Home, Containers, Images, Snapshots, Networks, Orchestration, and Settings. The main area is titled "Dashboard" and shows the "Learning Center" section with two cards: "Working with Kubernetes" and "Podman Tutorial Zero to Hero". Below that is a summary of a running pod named "podman-machine-default" with version "v5.2.5" and status "RUNNING". At the bottom, there are links to "Website", "Installation guide", "Docker compatibility guide", and "Join the community". The footer shows "kind-kind-cluster" and "AI Lab API listening on port 10434", and the version "v1.14.1" along with some status icons.

Create a k8s cluster: "Kind" (kind = "K"ubernetes "in" "D"ocker)



Create Kind

Podman Desktop

Settings Resources Proxy Registries Authentication CLI Tools Kubernetes Preferences >

Resources > Kind

 Create Kind cluster

Name: kind-cluster

Provider Type: podman

HTTP Port: 9090

HTTPS Port: 9443

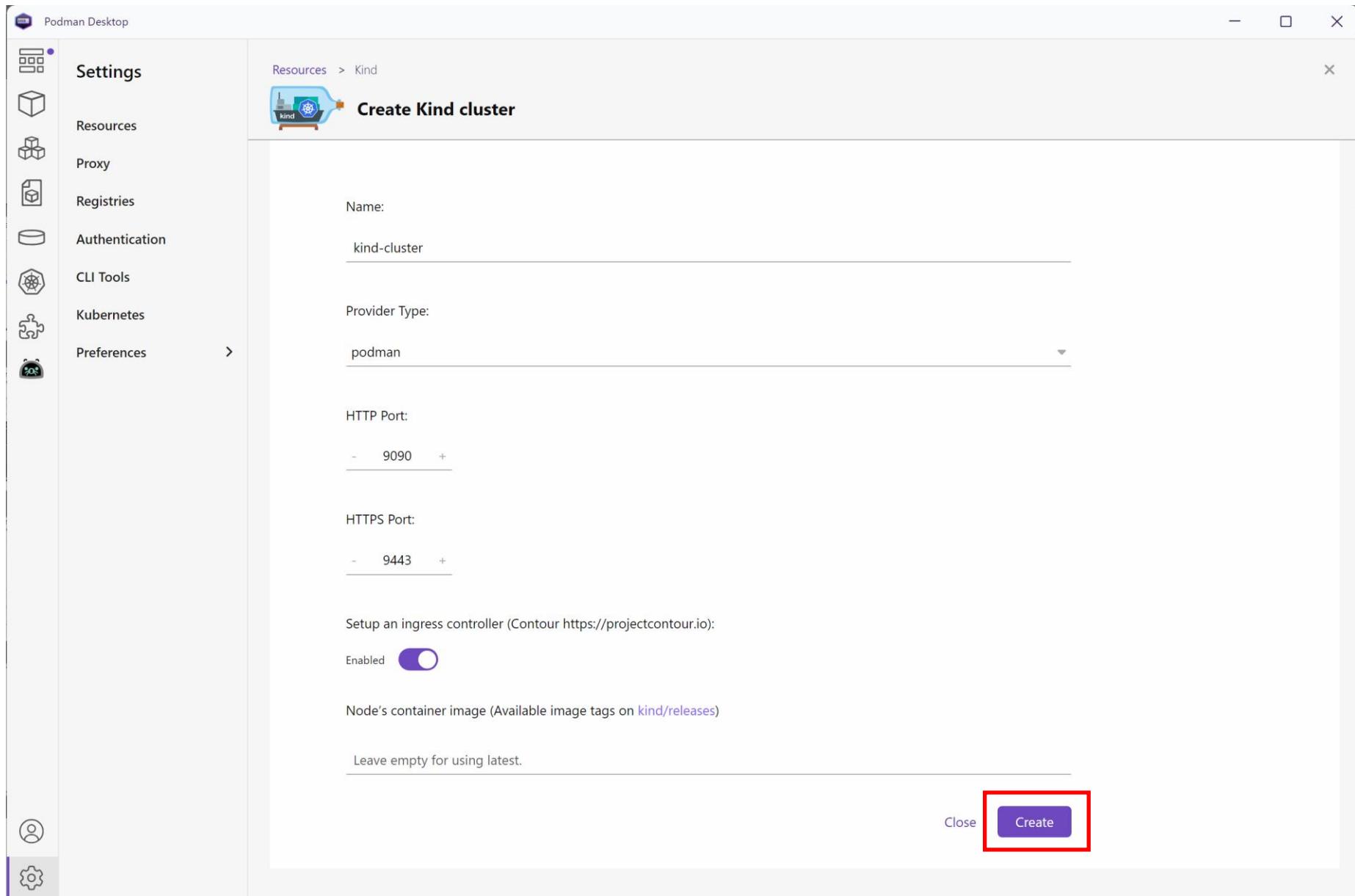
Setup an ingress controller (Contour <https://projectcontour.io>):

Enabled

Node's container image (Available image tags on [kind/releases](#))

Leave empty for using latest.

Close Create



Wait ...

Resources > Kind

 **Create Kind cluster**

Show Logs ^ Cancel

```
using podman due to KIND_EXPERIMENTAL_PROVIDER
enabling experimental podman provider
Creating cluster "kind-cluster" ...
  • Ensuring node image (kindest/node:v1.27.3) ...
  ✓ Ensuring node image (kindest/node:v1.27.3)
  • Preparing nodes ...
  ✓ Preparing nodes
  • Writing configuration ...
  ✓ Writing configuration
  • Starting control-plane ...

```

Resources > Kind

 **Create Kind cluster**



Creation

Successful operation

[Go back to resources](#)

Kind is RUNNING

Podman Desktop

— □ ×

Settings

Resources

Proxy

Registries

Authentication

CLI Tools

Kubernetes

Preferences >

Resources

Additional provider information is available under [Extensions](#)

Podman

Podman Machine RUNNING

Podman endpoint
npipe://\\.\pipe\podman-machine-d...

CPU(s) 8 | Disk size 1.08 TB | Memory 8.2 GB

Podman v5.2.5 | Wsl

kubectl

A command line tool for communicating with a Kubernetes cluster's control plane, using the Kubernetes API.

More information: [kubernetes.io](#)

Kind

kind-cluster RUNNING

Kubernetes endpoint
<https://localhost:53779>

The Kind provider is considered to be in an experimental state. It is not recommended for production use.

Kind - Nodes (only 1 : ControlPlane)

The screenshot shows the Podman Desktop application interface. On the left, there is a sidebar with the following navigation options:

- Podman Desktop
- Kubernetes
 - Dashboard (highlighted with a red box)
 - Nodes (highlighted with a red box)
 - Deployments
 - Services
 - Ingresses & Routes
 - Persistent Volume Claims
 - ConfigMaps & Secrets
 - Port Forwarding

The main area is titled "Nodes" and contains a table with the following data:

STATUS	NAME	ROLES	VERSION	OS	KERNEL	AGE
	kind-cluster-control-pl...	Control Plane	v1.27.3	Debian GNU/Linux 11 (bullseye)	5.15.167.4-microsoft-standard-WSL2	3 minutes

There is also a search bar labeled "Search nodes..." and a button labeled "Apply YAML". A status indicator shows "Connected".

Open Terminal, check Kubectl

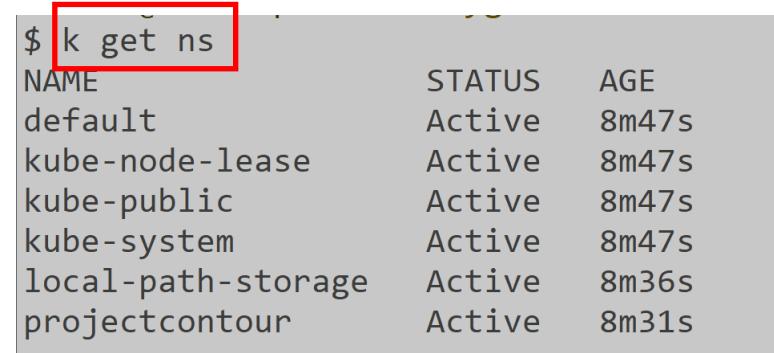
```
Invite de commandes x + ▾  
Microsoft Windows [version 10.0.26100.2314]  
(c) Microsoft Corporation. Tous droits réservés.  
  
C:\Users\arnaud>where kubectl  
C:\Users\arnaud\AppData\Local\Microsoft\WindowsApps\kubectl.exe  
  
C:\Users\arnaud>kubectl  
kubectl controls the Kubernetes cluster manager.  
  
Find more information at: https://kubernetes.io/docs/reference/kubectl/  
  
Basic Commands (Beginner):  
create Create a resource from a file or from stdin  
expose Take a replication controller, service, deployment or pod  
run Run a particular image on the cluster  
set Set specific features on objects  
  
Basic Commands (Intermediate):  
explain Get documentation for a resource  
get Display one or many resources  
edit Edit a resource on the server  
delete Delete resources by file names, stdin, namespaces and more
```

kubectl config get-contexts (cf standard aliases next: "kcgc")

```
arnaud@DesktopArnaud /cygdrive/c/arn/devPerso
$ kubectl config get-contexts
CURRENT      NAME           CLUSTER      AUTHINFO      NAMESPACE
*            kind-kind-cluster   kind-kind-cluster   kind-kind-cluster   default
                           minikube        minikube
```

kubectl get ns

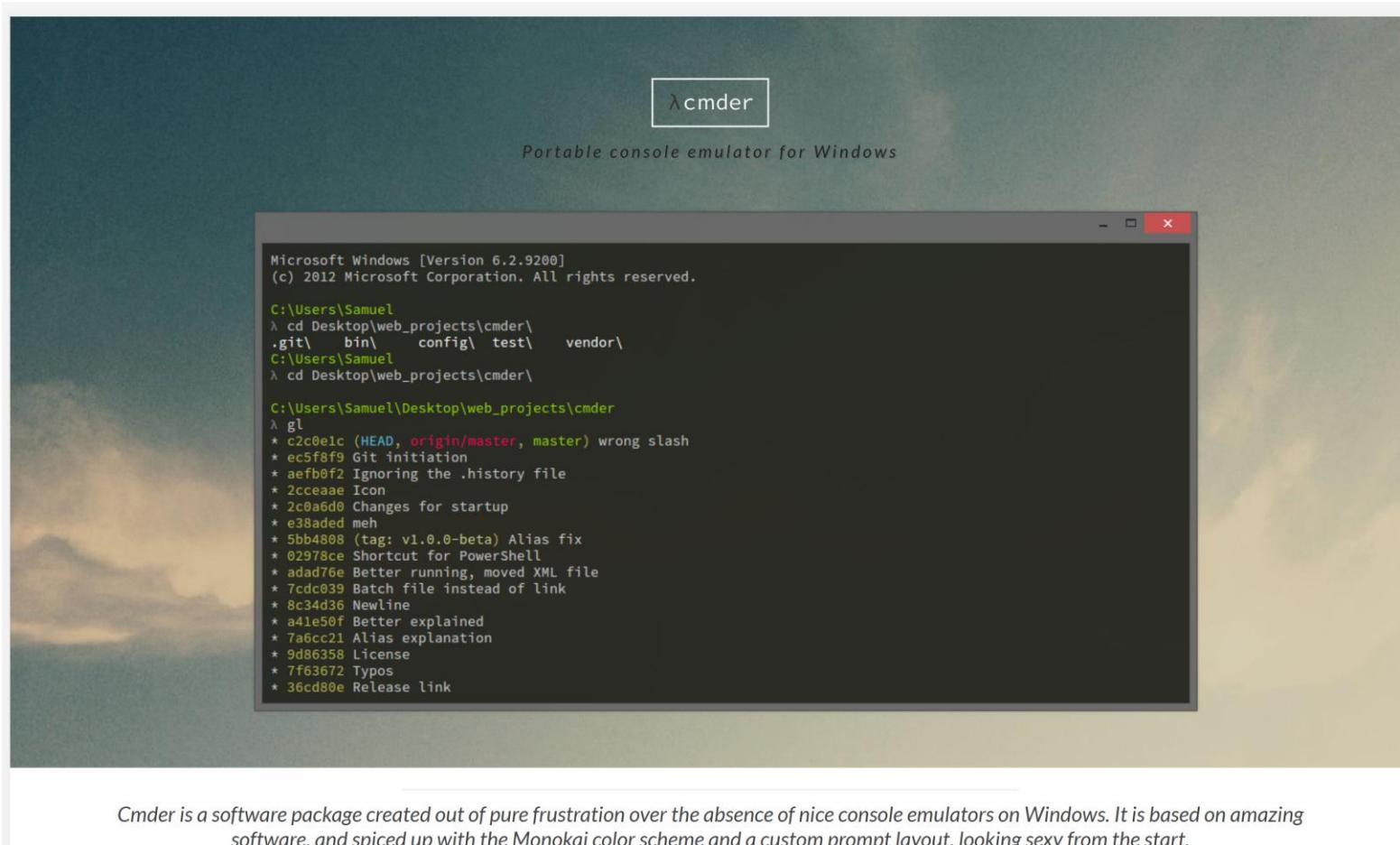
(or k get ns ... see next for aliases)



NAME	STATUS	AGE
default	Active	8m47s
kube-node-lease	Active	8m47s
kube-public	Active	8m47s
kube-system	Active	8m47s
local-path-storage	Active	8m36s
projectcontour	Active	8m31s

Optional: install K8s helper tools

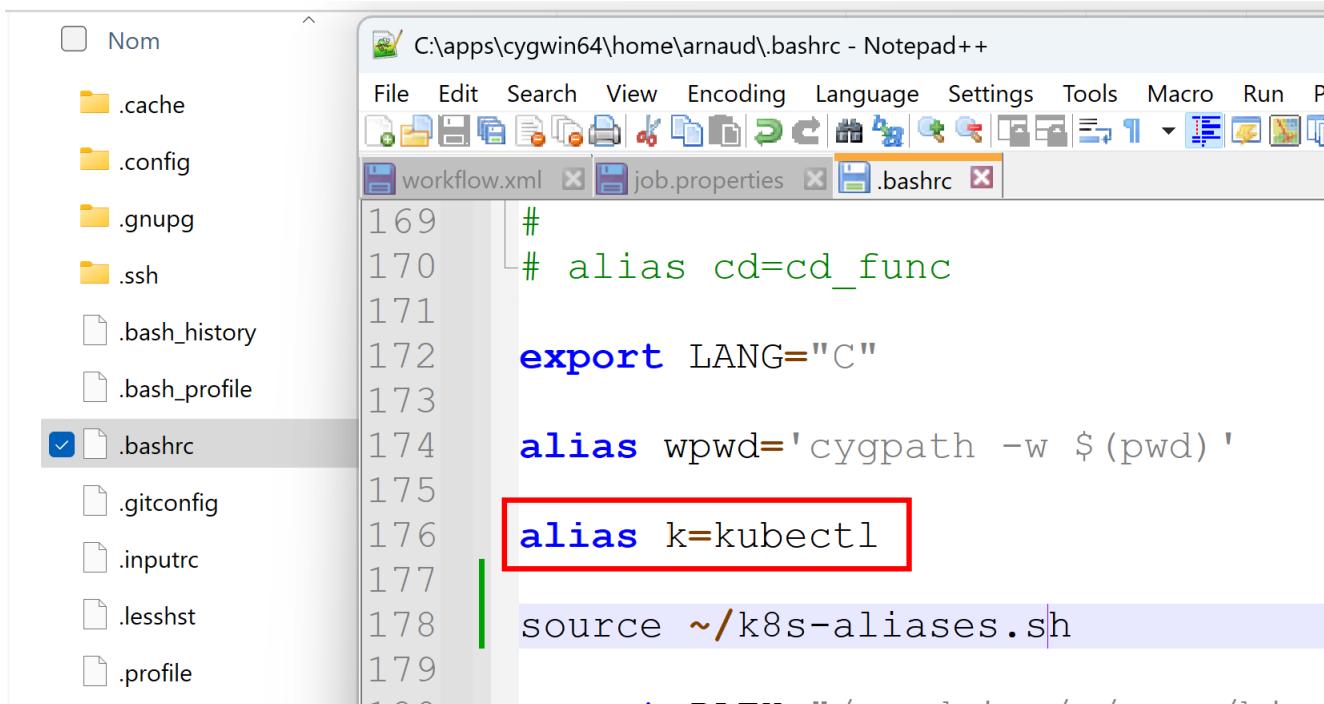
(Optional) Install Cmder <https://cmder.app/>
for "Bash" (without cygwin)
and do not use "cmd DOS" nor "PowerShell" !!



(Optional) usefull on Bash shell : aliases

alias k=kubectl

alias kg=kubectl get; alias kgp=kubectl get pod;



The screenshot shows a Notepad++ window displaying the contents of the `.bashrc` file located at `C:\apps\cygwin64\home\arnaud\.bashrc`. The left pane shows a file tree with various configuration files like `.cache`, `.config`, `.gnupg`, `.ssh`, `.bash_history`, `.bash_profile`, and `.gitconfig`. The `.bashrc` file is selected and highlighted in the tree view. The right pane contains the code:

```
#  
# alias cd=cd_func  
  
export LANG="C"  
  
alias wppwd='cygpath -w $(pwd)'  
  
alias k=kubectl  
  
source ~/k8s-aliases.sh
```

The line `alias k=kubectl` is highlighted with a red rectangular box.

(Optional) usefull on Bash shell loading many aliases from "sourced" file

Google

kubectl aliases shell

```
176 alias k=kubectl
177
178 source ~/k8s-aliases.sh
179
```

github.com/dwertyent/alias-kubectl/blob/main/kubectl_aliases

Code Blame 120 lines (110 loc) · 3.92 KB Code 55% faster with GitHub Copilot

```
1 #!/bin/bash
2
3 # other
4 alias k='kubectl'
5 alias kl='kubectl logs'
6 alias kexec='kubectl exec -it'
7 alias kpf='kubectl port-forward'
8 alias kaci='kubectl auth can-i'
9 alias kat='kubectl attach'
10 alias kapir='kubectl api-resources'
11 alias kapiv='kubectl api-versions'
12
13 # get
14 alias kg='kubectl get'
15 alias kgns='kubectl get ns'
16 alias kgp='kubectl get pods'
17 alias kgs='kubectl get secrets'
18 alias kgd='kubectl get deploy'
19 alias kgrs='kubectl get rs'
20 alias kgss='kubectl get sts'
21 alias kgds='kubectl get ds'
22 alias kgcm='kubectl get configmap'
23 alias kgcj='kubectl get cronjob'
24 alias kgj='kubectl get job'
25 alias kgsvc='kubectl get svc -o wide'
26 alias kgn='kubectl get no -o wide'
27 alias kgr='kubectl get roles'
28 alias krbn='kubectl get rolebindings'
```

(Optional) Usefull Bash Completions



kubernetes.io/docs/reference/kubectl/generated/kubectl_completion/



kubernetes

Documentation

Kubernetes Blog

Training

Partners

kubectl_completion_bash > p(drew --prefix)/etc/bash_completion.d/kubectl_completion.bash



Search this site

- ▶ Documentation
- ▶ Getting started
- ▶ Concepts
- ▶ Tasks
- ▶ Tutorials
- ▼ Reference

Glossary

```
# Installing bash completion on Linux
## If bash-completion is not installed on Linux, install the
## via your distribution's package manager.
## Load the kubectl completion code for bash into the current
## shell session
source <(kubectl completion bash)

## Write bash completion code to a file and source it from
## kubectl completion bash > ~/.kube/completion.bash.inc
printf "
# kubectl shell completion
source '$HOME/.kube/completion.bash.inc'
" >> $HOME/.bash_profile
source $HOME/.bash_profile
```

Kubectl Bash Completions

source <(kubectl completion bash)

```
176 alias k=kubectl
177
178 source ~/k8s-aliases.sh
179
180 # source <(kubectl completion bash)
181 if [ ! -e ~/k8s-completion.sh ]; then
182     kubectl completion bash > ~/k8s-completion.sh;
183 fi
184 source ~/k8s-completion.sh
185
```

kubectl <cmd><tab> for autocomplete

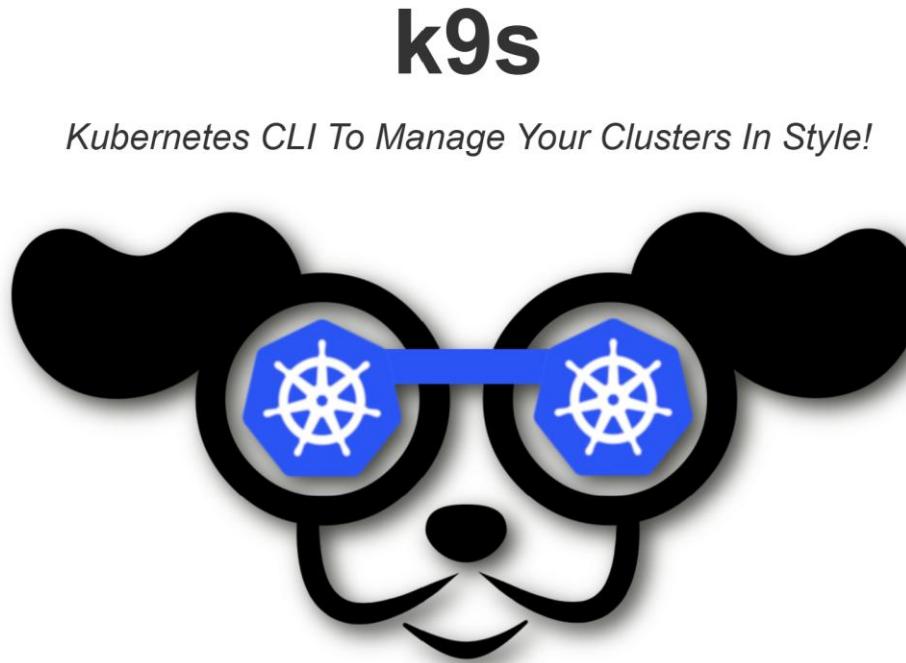
start typing "ap" then <tab><tab> => propose several auto completions

```
$ kubectl ap  
api-resources  (Print the supported API resources on the server)  
api-versions   (Print the supported API versions on the server, in the form of "group/version")  
apply          (Apply a configuration to a resource by file name or stdin)  
  
arnaud@DesktopArnaud /cygdrive/c/arn/devPerso  
$ kubectl ap
```

start typing "app" then <tab> => resolved unique completion "apply"

```
$ kubectl app  
  
$ kubectl apply 
```

Install Utility CLI tool : "k9s"



Who Let The Pods Out?

K9s is a terminal based UI to interact with your Kubernetes clusters. The aim of this project is to make it easier to navigate, observe and manage your deployed applications in the wild. K9s continually watches Kubernetes for changes and offers subsequent commands to interact with your observed resources.

Features

-
- Information At Your Finger Tips!

<https://k9scli.io/topics/install/>

winget install k9s

```
$ winget install k9s
Trouvé k9s [Derailed.k9s] Version 0.32.6
La licence d'utilisation de cette application vous est octroyée par son propriétaire.
Microsoft n'est pas responsable des paquets tiers et n'accorde pas de licences à ceux-ci.
Téléchargement en cours https://github.com/derailed/k9s/releases/download/v0.32.6/k9s\_Windows\_amd64.zip
[██████████] 30.6 MB / 30.6 MB
Le code de hachage de l'installation a été vérifié avec succès
Extraction des archives...
Extraction de l'archive réussie
Démarrage du package d'installation... Merci de patienter.
Variable d'environnement de chemin modifiée ; redémarrez votre interpréteur de commandes pour utiliser la nouvelle valeur.
Alias de ligne de commande ajouté : "k9s"
Installé correctement
```

using k9s

Cmder

```
Context: kind-kind-cluster
Cluster: kind-kind-cluster
User: kind-kind-cluster
K9s Rev: v0.32.6
K8s Rev: v1.27.3
CPU: n/a
MEM: n/a
```

<0> all <a> Attach <ctrl-k> Kill <o> Show Node
<1> default <ctrl-d> Delete <l> Logs <f> Show PortF
<d> Describe <p> Logs Previous <t> Transfer
<e> Edit <shift-f> Port-Forward <y> YAML
<?> Help <z> Sanitize
<shift-j> Jump Owner <s> Shell

NAME↑ PF READY STATUS RESTARTS IP NODE AGE

Pods(default)[0]

NAME↑	PF	READY	STATUS	RESTARTS	IP	NODE	AGE
<pod>							

bash.exe

Search

switch between pod, node, deployment: type "<ESCAPE> :node <ENTER>"

Context: kind-kind-cluster
Cluster: kind-kind-cluster
User: kind-kind-cluster
K9s Rev: v0.32.6
K8s Rev: v1.27.3
CPU: n/a
MEM: n/a

node.k8s.io/v1/runtimeclasses

Context: kind-kind-cluster Cluster: kind-kind-cluster User: kind-kind-cluster K9s Rev: v0.32.6 K8s Rev: v1.27.3 CPU: n/a MEM: n/a

<c> Cordon <u> Uncordon
<ctrl-d> Delete <y> YAML
<d> Describe
<r> Drain
<e> Edit
<?> Help

NAME↑	STATUS	ROLE	Taints	VERSION	PODS AGE
kind-cluster-control-plane	Ready	control-plane	0	v1.27.3	13 32m

Nodes(all)[1]

<node>

bash.exe

Search

Helm

<https://helm.sh>

The screenshot shows the official Helm website at <https://helm.sh>. The page has a white background with a faint, stylized blue topographic map pattern. At the top, there is a navigation bar with links for Home, Docs, Charts, Blog, and Community. To the right of the navigation bar are language selection ("English") and "Get Started" buttons. The main title "HELM" is displayed in large, bold, blue capital letters, with a small blue icon resembling a sun or crown above the letter "E". Below the title, the text "The package manager for Kubernetes" is written in a large, bold, blue sans-serif font. A callout box in the lower-left quadrant contains the text "Helm is the best way to find, share, and use software built for Kubernetes." In the bottom-left corner, there is a small illustration of a three-masted sailing ship on a blue base.

Check/Install "Helm" command line

```
[Cmder] $ helm
The Kubernetes package manager

Common actions for Helm:

- helm search:      search for charts
- helm pull:        download a chart to your local directory to view
- helm install:     upload the chart to Kubernetes
- helm list:        list releases of charts

Environment variables:

| Name           | Description
|-----|
| $HELM_CACHE_HOME | set an alternative location for storing cached files.
| $HELM_CONFIG_HOME | set an alternative location for storing Helm configuration.
| $HELM_DATA_HOME | set an alternative location for storing Helm data.
| $HELM_DEBUG | indicate whether or not Helm is running in Debug mode
```

helm list

```
$ helm list
```

NAME	NAMESPACE	REVISION	UPDATED	STATUS	CHART	APP VERSION
------	-----------	----------	---------	--------	-------	-------------

Nothing installed yet ... We need "spark operator for k8s"

Helm Install
Spark Operator for K8s

<https://github.com/kubeflow/spark-operator>

kubeflow / spark-operator

Type / to search | + - ⚡ 🔍 🚧 🚗

<> Code Issues 160 Pull requests 17 Actions Projects Security Insights

 spark-operator Public Watch 81 Fork 1.4k Star 2.8k

master ▾ 15 Branches 98 Tags Go to file Add file <> Code

 jacobsalway	Bump volcano.sh/apis to 1.10.0 (#2320)	22e4fb8 · 5 days ago	1,036 Commits
.github	Update issue and pull request templates (#2287)	3 weeks ago	
api	Support pod template for Spark 3.x applications (#2141)	3 weeks ago	
charts/spark-operator-chart	Allow the Controller and Webhook Containers to run with th...	last week	
cmd	implement an upper bound limit to the number of tracked e...	last month	
config	Support pod template for Spark 3.x applications (#2141)	3 weeks ago	
docs	Support pod template for Spark 3.x applications (#2141)	3 weeks ago	

About

Kubernetes operator for managing the lifecycle of Apache Spark applications on Kubernetes.

kubernetes spark apache-spark
kubernetes-operator kubernetes-controller
kubernetes-crd google-cloud-dataproc

Readme Apache-2.0 license Code of conduct Activity

Notice...

[https://github.com/GoogleCloudPlatform/
spark-on-k8s-operator](https://github.com/GoogleCloudPlatform/spark-on-k8s-operator)

(origin, now automatically redirected to kubeflow)



Google Cloud Community

<https://www.googlecloudcommunity.com> › spark-on-k... :

Solved: spark-on-k8s-operator

Spark-on-k8s-operator was donated to the **kubeflow** community, so it's being maintained by them now instead of Google. The github project you found is the correct ...

Notice: another operator... in apache spark official doc

The screenshot shows the Apache Spark homepage at spark.apache.org. The page features the Apache Spark logo and a large banner with the text "Unified engine for large-scale data analytics". A prominent orange "GET STARTED" button is visible. The navigation bar includes links for Download, Libraries, Documentation, Examples, Community, Developers, and GitHub. A dropdown menu for "GitHub" is open, showing options: spark, spark-connect-go, spark-docker, spark-kubernetes-operator, and spark-website. The "spark-kubernetes-operator" option is highlighted with a red box. The Apache Software Foundation logo is in the top right corner.

<https://github.com/apache/spark-kubernetes-operator>

The screenshot shows the GitHub repository page for 'spark-kubernetes-operator' under the 'apache' organization. The page includes a header with navigation links for Code, Pull requests, Actions, Security, and Insights. Below the header, there's a search bar and several status indicators: Watch (27), Fork (16), and Star (64). The main content area displays the repository's structure with a 'main' branch selected, showing 1 branch and 0 tags. A list of recent commits is shown, with the most recent one by 'dongjoon-hyun' upgrading Gradle to 8.11. The repository is described as the 'Apache Spark Kubernetes Operator' and includes links to its website and Java/Kubernetes/Spark dependencies. The footer contains a note about Apache-2.0 licenses.

apache / spark-kubernetes-operator

Type / to search

Code Pull requests Actions Security Insights

Watch 27 Fork 16 Star 64

main 1 Branch 0 Tags

Go to file Add file Code

dongjoon-hyun [SPARK-50305] Upgrade Gradle to 8.11 0df50bb · 3 days ago 150 Commits

.github [SPARK-49658] Refactor e2e tests pipelines 2 months ago

build-tools [SPARK-50305] Upgrade Gradle to 8.11 3 days ago

config [SPARK-491671] Enforce `UtilityClass` rule 3 months ago

About

Apache Spark Kubernetes Operator

[spark.apache.org/](#)

java kubernetes spark

Readme

Apache-2.0, Apache-2.0 licenses found

confusing ...

WARN ...

This TD is using the "kubeflow" operator
the kubeflow operator manage these types:

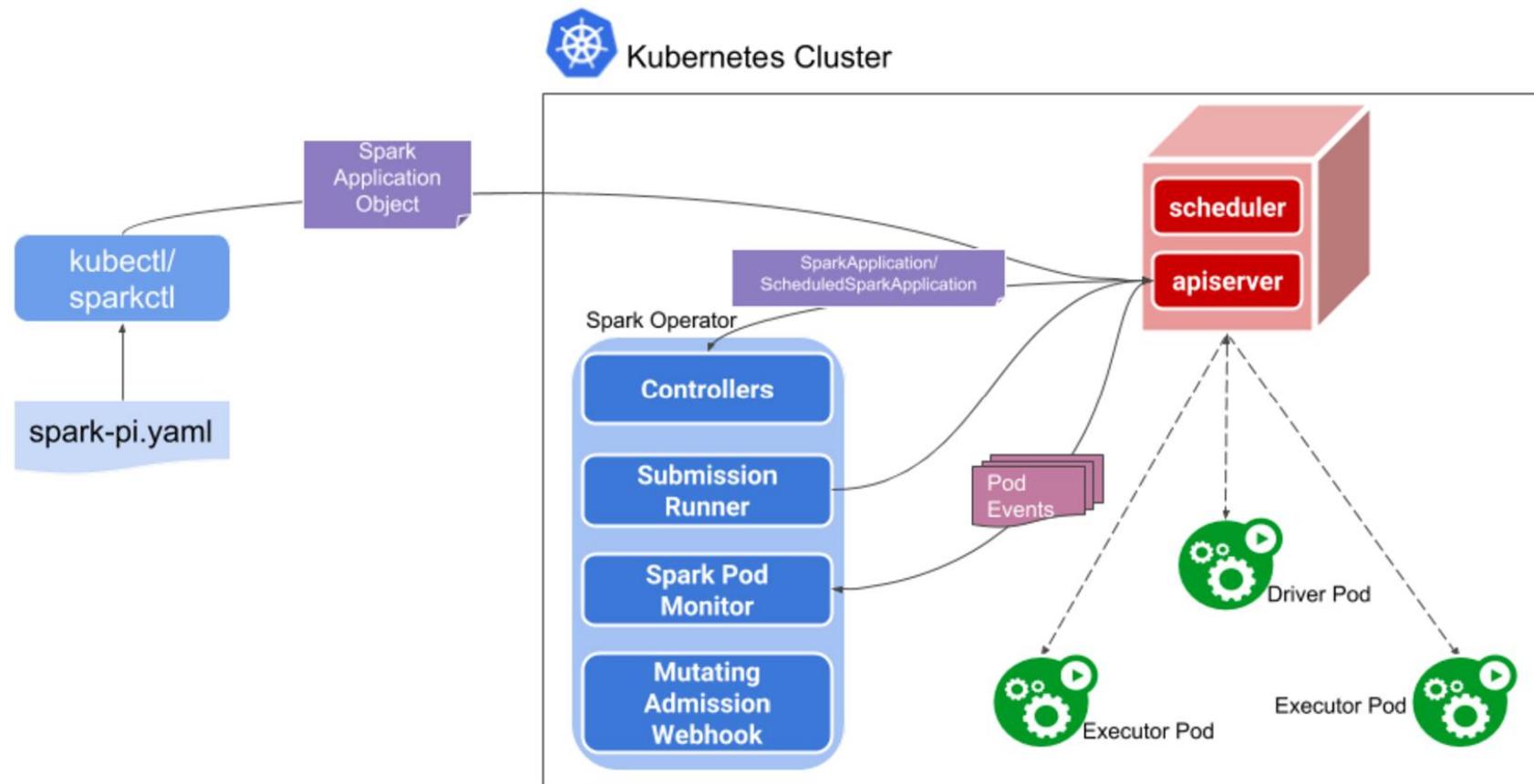
```
apiVersion: sparkoperator.k8s.io/v1beta2
kind: SparkApplication
kind: ScheduledSparkApplication
```

!= the "apache" one,
the apache operator manages theses types:

```
apiVersion: spark.apache.org/v1alpha1
kind: SparkCluster
kind: SparkApplication
```

doc

<https://www.kubeflow.org/docs/components/spark-operator>



<https://www.kubeflow.org/docs/components/spark-operator/getting-started/>

Documentation

- ▶ About
- ▶ Getting Started
- ▶ Concepts

▼ Components

- ▶ Central
Dashboard
 - ▶ Kubeflow
Notebooks
 - ▶ Kubeflow
Pipelines
 - ▶ Katib
 - ▶ Model Registry
- ### ▼ Spark Operator
- Overview

Installation

Add Helm Repo

```
helm repo add spark-operator https://kubeflow.github.io/spark-operator  
  
helm repo update
```

See [helm repo](#) for command documentation.

Install the chart

```
helm install [RELEASE_NAME] spark-operator/spark-operator
```

OLD repo already exist (but invalid) ... Failed to update

```
$ helm repo add spark-operator https://kubeflow.github.io/spark-operator
Error: repository name (spark-operator) already exists, please specify a different name

arnaud@DesktopArnaud /cygdrive/c/arn/devPerso
$ helm repo update
Hang tight while we grab the latest from your chart repositories...
...Unable to get an update from the "spark-operator" chart repository (https://googlecloudplatform.github.io/spark-on-k8s-operator):
    failed to fetch https://googlecloudplatform.github.io/spark-on-k8s-operator/index.yaml : 404 Not Found
Update Complete. Happy Helming!
```

remove + re-add

helm repo **remove** spark-operator

helm repo **add** spark-operator https://kubeflow.github.io/spark-operator

helm repo update

```
$ helm repo remove spark-operator
"spark-operator" has been removed from your repositories

arnaud@DesktopArnaud /cygdrive/c/arn/devPerso
$ helm repo add spark-operator https://kubeflow.github.io/spark-operator
"spark-operator" has been added to your repositories

arnaud@DesktopArnaud /cygdrive/c/arn/devPerso
$ helm repo update
Hang tight while we grab the latest from your chart repositories...
...Successfully got an update from the "spark-operator" chart repository
Update Complete. ⚡Happy Helming!⚡
```

helm install

```
helm install spark-operator spark-operator/spark-operator \
--namespace spark-operator \
--create-namespace
```

```
$ helm install spark-operator spark-operator/spark-operator --namespace spark-operator --create-namespace
NAME: spark-operator
LAST DEPLOYED: Sat Nov 16 14:13:45 2024
NAMESPACE: spark-operator
STATUS: deployed
REVISION: 1
TEST SUITE: None
```

Checking: helm list

```
$ helm list
NAME      NAMESPACE      REVISION      UPDATED      STATUS      CHART      APP VERSION
arnaud@DesktopArnaud /cygdrive/c/arn/devPerso
$ helm list -n spark-operator
NAME      NAMESPACE      REVISION      UPDATED      STATUS      CHART      APP VERSION
spark-operator  spark-operator  1          2024-11-16 13:22:32.0837027 +0100 CET  deployed    spark-operator-2.0.2  2.0.2
```

NOTICE: helm list => nothing in default namespace, need to specify "-n spark-operator"

installed spark-operator (2.0.2) Not Working ?

checking: kubectl get all -n spark-operator

```
$ kubectl get all -n spark-operator
NAME                                         READY   STATUS             RESTARTS   AGE
pod/spark-operator-controller-7c47b8c8d7-wnb19  0/1    CreateContainerError  0          103s
pod/spark-operator-webhook-7f77c86778-9prcj    0/1    CreateContainerError  0          103s

NAME                           TYPE      CLUSTER-IP      EXTERNAL-IP   PORT(S)      AGE
service/spark-operator-webhook-svc  ClusterIP  10.96.229.8   <none>        9443/TCP   103s

NAME                               READY   UP-TO-DATE   AVAILABLE   AGE
deployment.apps/spark-operator-controller  0/1     1           0           103s
deployment.apps/spark-operator-webhook       0/1     1           0           103s

NAME                         DESIRED   CURRENT   READY   AGE
replicaset.apps/spark-operator-controller-7c47b8c8d7  1         1         0       103s
replicaset.apps/spark-operator-webhook-7f77c86778     1         1         0       103s
```

2 errors ... CreateContainerError !

Investigating using k9s

```
Context: kind-kind-cluster
Cluster: kind-kind-cluster
User: kind-kind-cluster
K9s Rev: v0.32.6
K8s Rev: v1.27.3
CPU: n/a
MEM: n/a

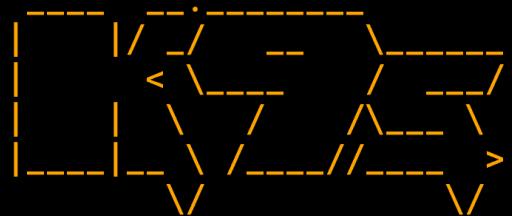
          <0> all           <a>     Attach      <ctrl-k> Kill       <o> Sh...
          <1> spark-operator <ctrl-d> Delete    <l> Logs       <f> Sho...
          <2> default        <d> Describe   <p> Logs Previous <t> Tra...
                                         <e> Edit       <shift-f> Port-Forward <y> YAM...
                                         <?> Help       <z> Sanitize   <shift-j> Jump Owner <s> Shell
                                         <shift-j> Jump Owner <s> Shell

Pods(spark-operator)[2]
NAME                PF  READY  STATUS           RESTARTS  IP          NODE          AGE
spark-operator-controller-7c47b8c8d7-wnbl9  ●  0/1   CreateContainerError  0  10.244.0.12  kind-cluster-control-plane  5m15s
spark-operator-webhook-7f77c86778-9prcj   ●  0/1   CreateContainerError  0  10.244.0.13  kind-cluster-control-plane  5m15s
```

k9s events

Context: kind-kind-cluster
Cluster: kind-kind-cluster
User: kind-kind-cluster
K9s Rev: v0.32.6
K8s Rev: v1.27.3
CPU: n/a
MEM: n/a

<0> all <d> Describe
<1> spark-operator <?> Help
<2> default <y> YAML



Events(spark-operator)[78]

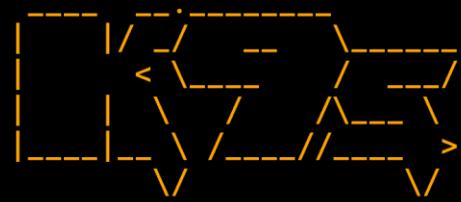
LAST SEEN↓	TYPE	REASON	OBJECT	COUNT
18m	Normal	SuccessfulCreate	replicaset/spark-operator-controller-7c47b8c8d7	1
18m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
18m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
18m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
18m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
18m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
18m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
18m	Normal	Scheduled	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
17m	Warning	Failed	pod/spark-operator-webhook-7f77c86778-9prcj	1
17m	Warning	Failed	pod/spark-operator-webhook-7f77c86778-9prcj	1
17m	Warning	Failed	pod/spark-operator-webhook-7f77c86778-9prcj	1
17m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
17m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
17m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	1
16m	Warning	Failed	pod/spark-operator-webhook-7f77c86778-9prcj	3
16m	Warning	Failed	pod/spark-operator-controller-7c47b8c8d7-wnbl9	3
3m51s	Normal	Pulled	pod/spark-operator-controller-7c47b8c8d7-wnbl9	71
3m49s	Normal	Pulled	pod/spark-operator-webhook-7f77c86778-9prcj	72

<event>

K9s - event - Failed

Context: kind-kind-cluster
Cluster: kind-kind-cluster
User: kind-kind-cluster
K9s Rev: v0.32.6
K8s Rev: v1.27.3
CPU: n/a
MEM: n/a

```
<c>          Copy
<e>          Edit
<n>          Next Match
<shift-n>    Prev Match
<r>          Toggle Auto-Refresh
<f>          Toggle FullScreen
```



```
Describe(spark-operator/spark-operator-controller-7c47b8c8d7-wnbl9.1808751db99de691)
```

```
Count: 1
Event Time: <nil>
First Timestamp: 2024-11-16T13:13:47Z
Involved Object:
  API Version: v1
  Field Path: spec.containers{spark-operator-controller}
  Kind: Pod
  Name: spark-operator-controller-7c47b8c8d7-wnbl9
  Namespace: spark-operator
  Resource Version: 10336
  UID: 9f5f438e-c662-488e-92e0-1ef10d539c0e
  Kind: Event
Last Timestamp: 2024-11-16T13:13:47Z
```

Message: Error: failed to create containerd container: mount callback failed on /var/lib/containerd/tmpmounts/containerd-mount1139155736: no users found

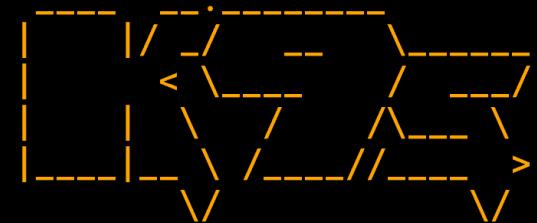
Metadata:
Creation Timestamp: 2024-11-16T13:13:47Z
Resource Version: 10355
UID: 19a66bd5-a449-46cf-8b99-5d63e3fc1d0e

<event> **<describe>**

k9s - pod - describe

```
Context: kind-kind-cluster  
Cluster: kind-kind-cluster  
User: kind-kind-cluster  
K9s Rev: v0.32.6  
K8s Rev: v1.27.3  
CPU: n/a  
MEM: n/a
```

```
<c> Copy  
<e> Edit  
<n> Next Match  
<shift-n> Prev Match  
<r> Toggle Auto-Refresh  
<f> Toggle FullScreen
```



```
Describe(spark-operator/spark-operator-webhook-7f77c86778-9prcj)
```

```
allback failed on /var/lib/containerd/tmpmounts/containerd-mount37773448: no users found  
  Warning Failed 16m kubelet Error: failed to create containerd container: mount c  
allback failed on /var/lib/containerd/tmpmounts/containerd-mount128027710: no users found  
  Warning Failed 16m kubelet Error: failed to create containerd container: mount c  
allback failed on /var/lib/containerd/tmpmounts/containerd-mount2920509095: no users found  
  Warning Failed 16m kubelet Error: failed to create containerd container: mount c  
allback failed on /var/lib/containerd/tmpmounts/containerd-mount3338808262: no users found  
  Warning Failed 16m kubelet Error: failed to create containerd container: mount c  
allback failed on /var/lib/containerd/tmpmounts/containerd-mount1761742618: no users found  
  Warning Failed 16m kubelet Error: failed to create containerd container: mount c  
allback failed on /var/lib/containerd/tmpmounts/containerd-mount2507996816: no users found  
  Warning Failed 15m kubelet Error: failed to create containerd container: mount c  
allback failed on /var/lib/containerd/tmpmounts/containerd-mount1222801735: no users found  
  Warning Failed 15m (x3 over 15m) kubelet (combined from similar events): Error: failed to crea  
te containerd container: mount callback failed on /var/lib/containerd/tmpmounts/containerd-mount3396310179: no users  
found  
  Normal Pulled 2m17s (x72 over 17m) kubelet Container image "docker.io/kubeflow/spark-operator:2.  
0.2" already present on machine
```

Find latest image

<https://hub.docker.com/r/kubeflow/spark-operator/tags>

The screenshot shows the Docker Hub interface for the `kubeflow/spark-operator` image. At the top, there's a search bar with the query `kubeflow/spark-operator`. Below the search bar, the repository name `kubeflow/spark-operator` is displayed, along with its owner `kubeflow` and the last update time, `23 days ago`. A red box highlights the `Tags` tab, which is currently selected. Another red box highlights the tag `2.1.0-rc.0`, which was pushed 23 days ago by `andreyvelichkevich`. To the right of this row is a command-line snippet: `docker pull kubeflow/spark-operator:2.1.0-rc.0` with a `Copy` button. Below the tag row, there are two more rows showing digests and OS/ARCH combinations: `b222d46627a0` for `linux/amd64` and `955b4ee91c00` for `linux/arm64`. The compressed size for each is listed as `537.23 MB` and `525.18 MB` respectively.

Digest	OS/ARCH	Compressed Size
b222d46627a0	linux/amd64	537.23 MB
955b4ee91c00	linux/arm64	525.18 MB

k9s deployment - edit

in K9, type

(if not already set, to select namespace :

<ESCAPE> :ns<ENTER> <ESCAPE> /spark-operator <ENTER>)

<ESCAPE> :deployment <ENTER>

scroll to "spark-operator-controller" (or <ESCAPE> / *controller)

type "e" to Edit

=> your Notepad editor opens, change "2.0.2" to "latest" (or "2.1.0-rc.0" as of 2024-11)

```
    . . .
    - --workqueue-ratelimiter-max-delay=6h
    # image: docker.io/kubeflow/spark-operator:2.0.2
    image: docker.io/kubeflow/spark-operator:2.1.0-rc.0
    imagePullPolicy: IfNotPresent
    livenessProbe:
```

then "SAVE" and close editor

... REPEAT same edit for other: "spark-operator-webhook"

deployment after edit image version

```
Context: kind-kind-cluster          <0> all           <ctrl-d> Delete
Cluster: kind-kind-cluster         <1> spark-operator <d>       Describe
User:   kind-kind-cluster          <2> default      <e>       Edit
K9s Rev: v0.32.6                   <?>       Help
K8s Rev: v1.27.3                   <shift-j> Jump Owner <l>       Logs
CPU:    n/a
MEM:   n/a
```

1/1 means OK... all expected "../1" are running "1/.."

Before, it was error: "0/1"

Check : operator pods "Running"

Context: kind-kind-cluster	<0> all	<a>	Attach
Cluster: kind-kind-cluster	<1> spark-operator	<ctrl-d>	Delete
User: kind-kind-cluster	<2> default	<d>	Describe
K9s Rev: v0.32.6		<e>	Edit
K8s Rev: v1.27.3		<?>	Help
CPU: n/a		<shift-j>	Jump Owner
MEM: n/a			

Pods(spark-operator)[2]							
NAME↑	PF	READY	STATUS	RESTARTS	IP	NODE	AGE
spark-operator-controller-7779b5c66d-zc4pr	●	1/1	Running	0	10.244.0.14	kind-cluster-control-plane	18m
spark-operator-webhook-9d465ccf5-qd5lg	●	1/1	Running	0	10.244.0.16	kind-cluster-control-plane	8m22

Pods are "Running", and Ready "1/1"

Notice also Restart=0 (they are not in a CrashLoop causing many Restarts)

Finalize Installation:
create ServiceAccount + role
for your spark app

Adding serviceAccount "spark"

```
kubectl create serviceaccount spark
```

```
kubectl create clusterrolebinding spark-role --clusterrole=edit --serviceaccount=default:spark  
--namespace=default
```

```
C:\apps\spark\spark-3.5.0> kubectl create serviceaccount spark  
serviceaccount/spark created
```

```
C:\apps\spark\spark-3.5.0>kubectl create clusterrolebinding spark-role --clusterrole=edit --serviceaccount=default:spark --namespace=default  
clusterrolebinding.rbac.authorization.k8s.io/spark-role created
```

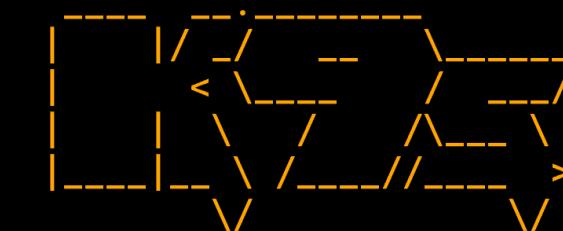
Checking ServiceAccount

Context: kind-kind-cluster	<0> all	<ctrl-d>	Delete
Cluster: kind-kind-cluster	<1> spark-operator	<d>	Describe
User: kind-kind-cluster	<2> default	<e>	Edit
K9s Rev: v0.32.6		<?>	Help
K8s Rev: v1.27.3		<shift-j>	Jump Owner
CPU: n/a		<enter>	Rules
MEM: n/a			
	Serviceaccounts (default)[3]		
NAME↑	SECRET	AGE	
default	0	5h56m	
spark	0	2m8s	
spark-operator-spark	0	156m	

Notice: other serviceAccounts created by "Helm" on namespace "spark-operator"

```
Context: kind-kind-cluster
Cluster: kind-kind-cluster
User: kind-kind-cluster
K9s Rev: v0.32.6
K8s Rev: v1.27.3
CPU: n/a
MEM: n/a
```

```
<0> all           <ctrl-d> Delete
<1> spark-operator <d>      Describe
<2> default        <e>      Edit
                           <?>      Help
                           <shift-j> Jump Owner
                           <enter> Rules
```



```
Serviceaccounts(spark-operator)[3]
```

NAME↑	SECRET	AGE
default	0	3h29m
spark-operator-controller	0	158m
spark-operator-webhook	0	158m

Next Steps: Part 2/2

open document
TD5-Spark-On-k8s-part2

part 2/2

- (Optional) Launch spark-shell on Docker, in mode "--master local[*]"
- (Optional) Launch spark with "spark-submit --master k8s://localhost:port"
- Launch spark with "kubectl apply -f file.yaml"
- (Optional... and backup if local install fails) Using Azure Kubernetes