

Running Spark on Kubernetes

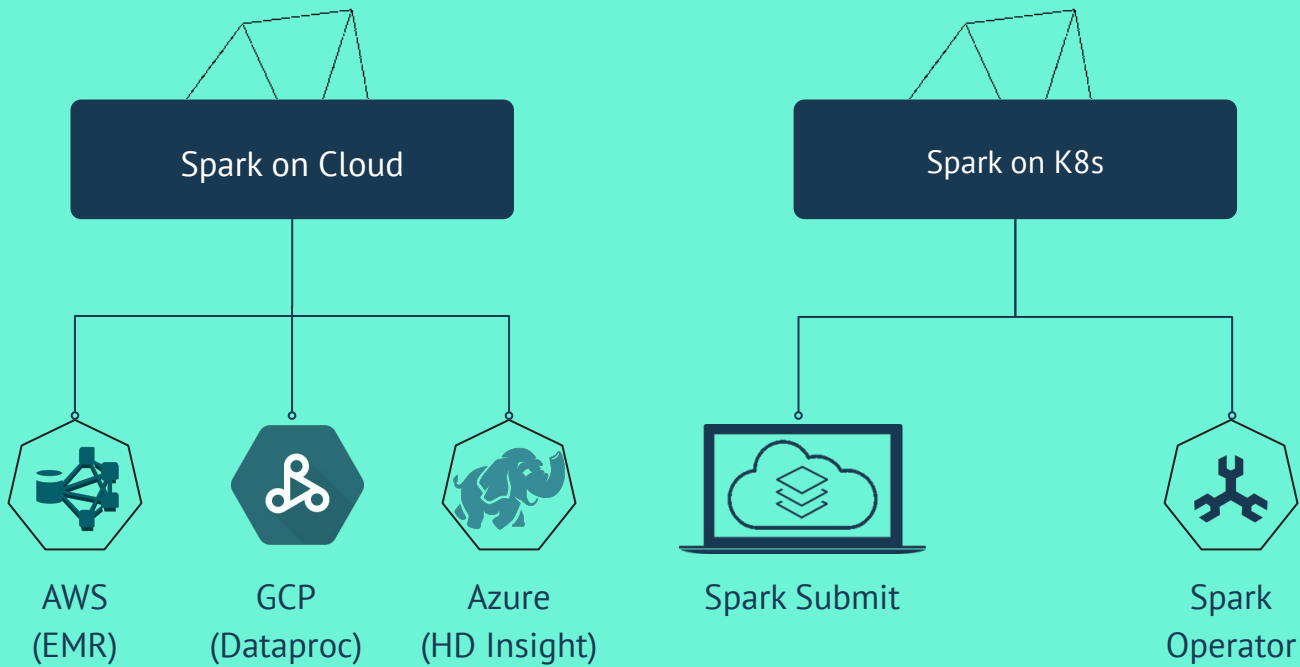
Ramiro Alvarez Fernandez



Running Spark on K8s

By Ramiro **Álvarez** Fernández

Opciones





Escalabilidad

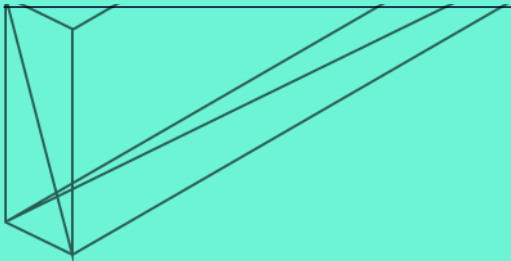
Menores costes

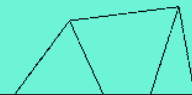
Mayor portabilidad

Monitorización

CI/CD

Agnosticismo



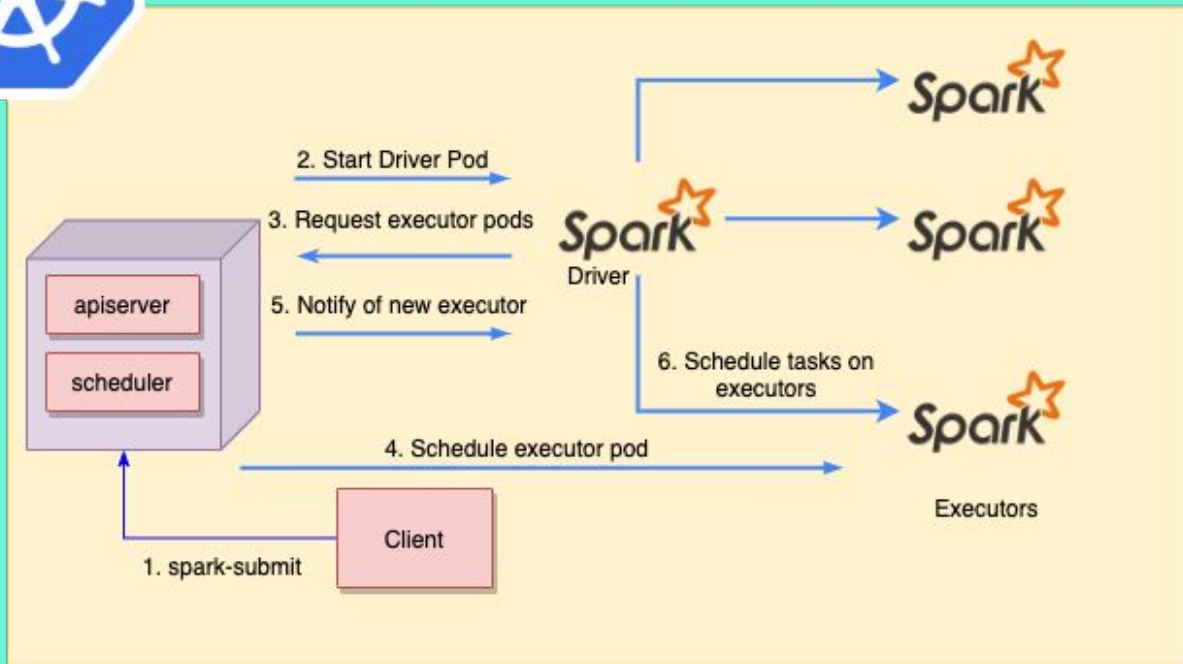


Running Spark on K8s

www.ramiroalvfer.dev

Spark on K8s

Funcionamiento



Spark Submit

1. Kubernetes version_

2. **Nativo Spark**

3. No es declarativo

4. Es necesario definir los recursos por nosotros mismos

5. No es necesario instalar un CRD

Spark Operator

1. Declarativo

2. Cron nativo

3. ConfigMaps

Volumes

Secrets

4. Reinicios
configurables

5. Inyección de
dependencias

6. Métricas

7. Gran soporte de la comunidad

Funcionamiento práctico

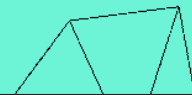


```
SparkSubmit x
1 # Run a Spark Application as a Job
2
3 $ /opt/spark/bin/spark-submit --master
  k8s://https://172.18.0.2:6443 --deploy-mode cluster
  --name spark-pi ...
4
5 # List Spark apps
6
7 $ kubectl get po -l job-name=spark-on-localhost-example
8
9 # Read logs
10
11 $ kubectl logs -f spark-on-localhost-example-lrvrb
12
13 # Describe app
14
15 $ kubectl describe po spark-on-localhost-example-lrvrb

SparkOperator x
1 # Run a Spark Application using Spark Operator
2
3 $ kubectl apply -f <spark-application>.yaml
4
5 # List Spark apps
6
7 $ kubectl get sparkApplication
8
9 # Read logs
10
11 kubectl logs -f <spark-application-name>-driver
12
13 # Describe app
14
15 kubectl describe sparkapplication <
  spark-application-name>
```

Line 15, Column 55

Tab Size: 4 Plain Text

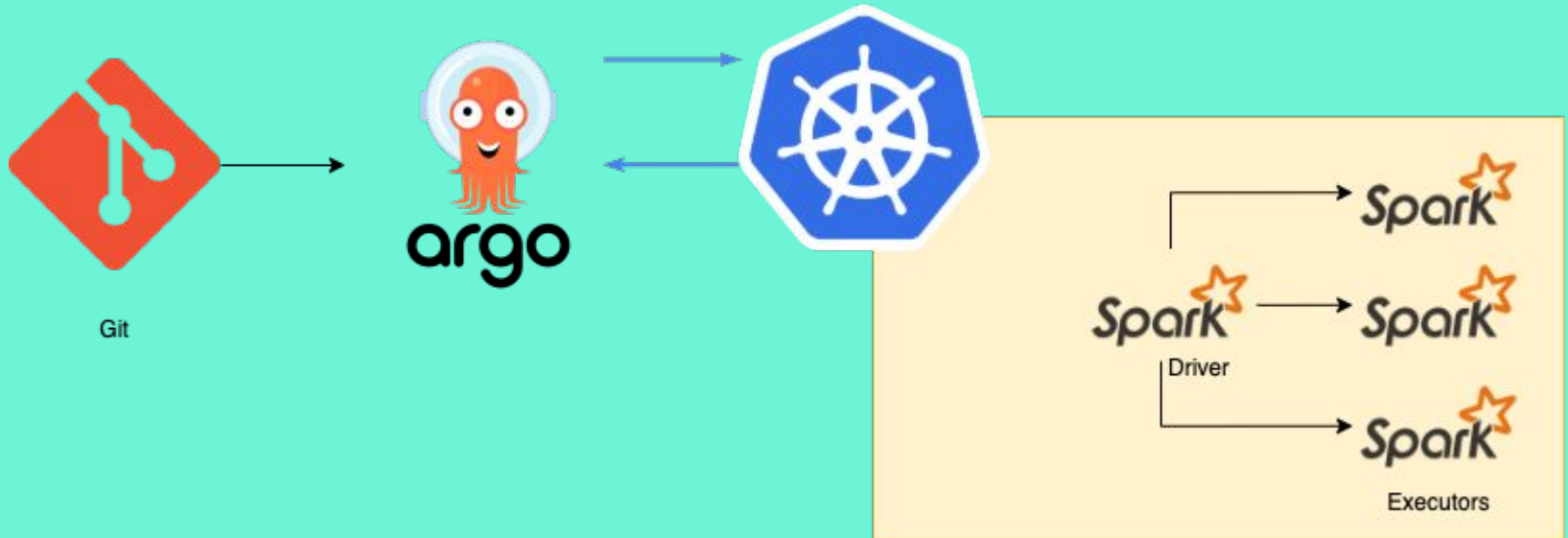


Running Spark on K8s

www.ramiroalvfer.dev

Orquestación

Orquestación



Argo Workflows

1. Orquestación de Jobs

2. Schedule workflows

3. Spark Applications

4. Workflow Templates

DRY

5. UI Visualizacion

Argo CD

1. GitOps

2. Declarativo

3.

Trazabilidad y
automatización

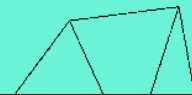
4.

Web UI

5..

Helm, Kustomize...

Escoge tu luchador



Running Spark on K8s

www.ramiroalvfer.dev

Demo

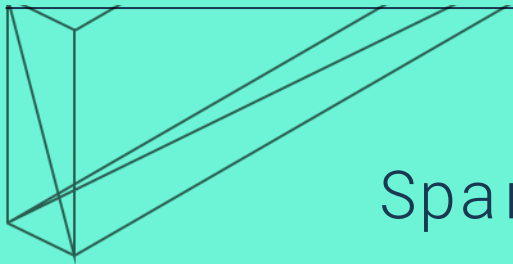


ArgoCD

SparkSubmit vs Spark Operator

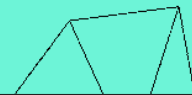
Argo Workflows

Monitorización



Spark History Server

Documentación



Running Spark on K8s

www.ramiroalvfer.dev

Demo

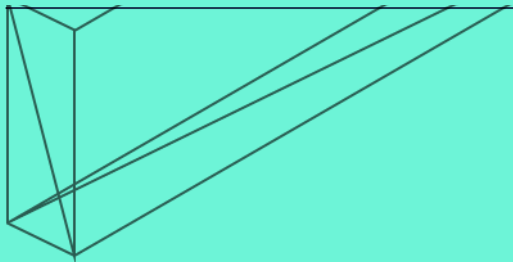


Escalabilidad

Menores costes

Mayor portabilidad

Monitorización



CI/CD

Autonomía



Ramiro
Álvarez
Fernández

Platform Engineering
Tech Lead
en **Empathy.co**



access





Gracias

empathy.co/company/careers/