

Microsoft



Azure

Tips & Tricks pour de la haute disponibilité avec Azure Kubernetes Services

@thomas_rannou

thomasrannou.azurewebsites.net

Introduction à Kubernetes

D'abord, les conteneurs :

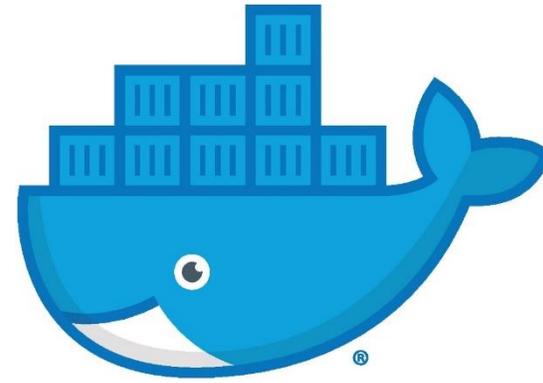
- Docker permet de créer et exécuter facilement nos conteneurs

Mais

- Comment gérer et planifier leur cycle de vie ?
- Comment monter en charge ?
- Comment manager une grande quantité de container ?

Que permet de faire Kubernetes ?

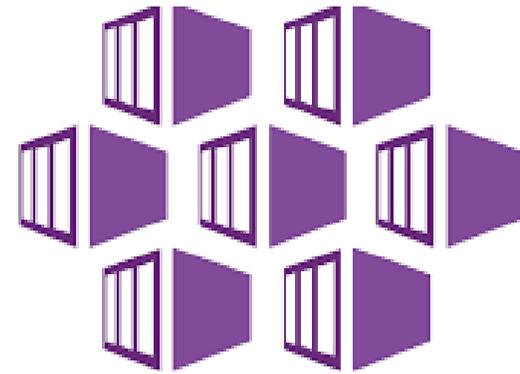
- Automatiser le déploiement et la réplication de conteneurs
- Gérer la montée en charge
- Détecter des problèmes sur un conteneur



kubernetes

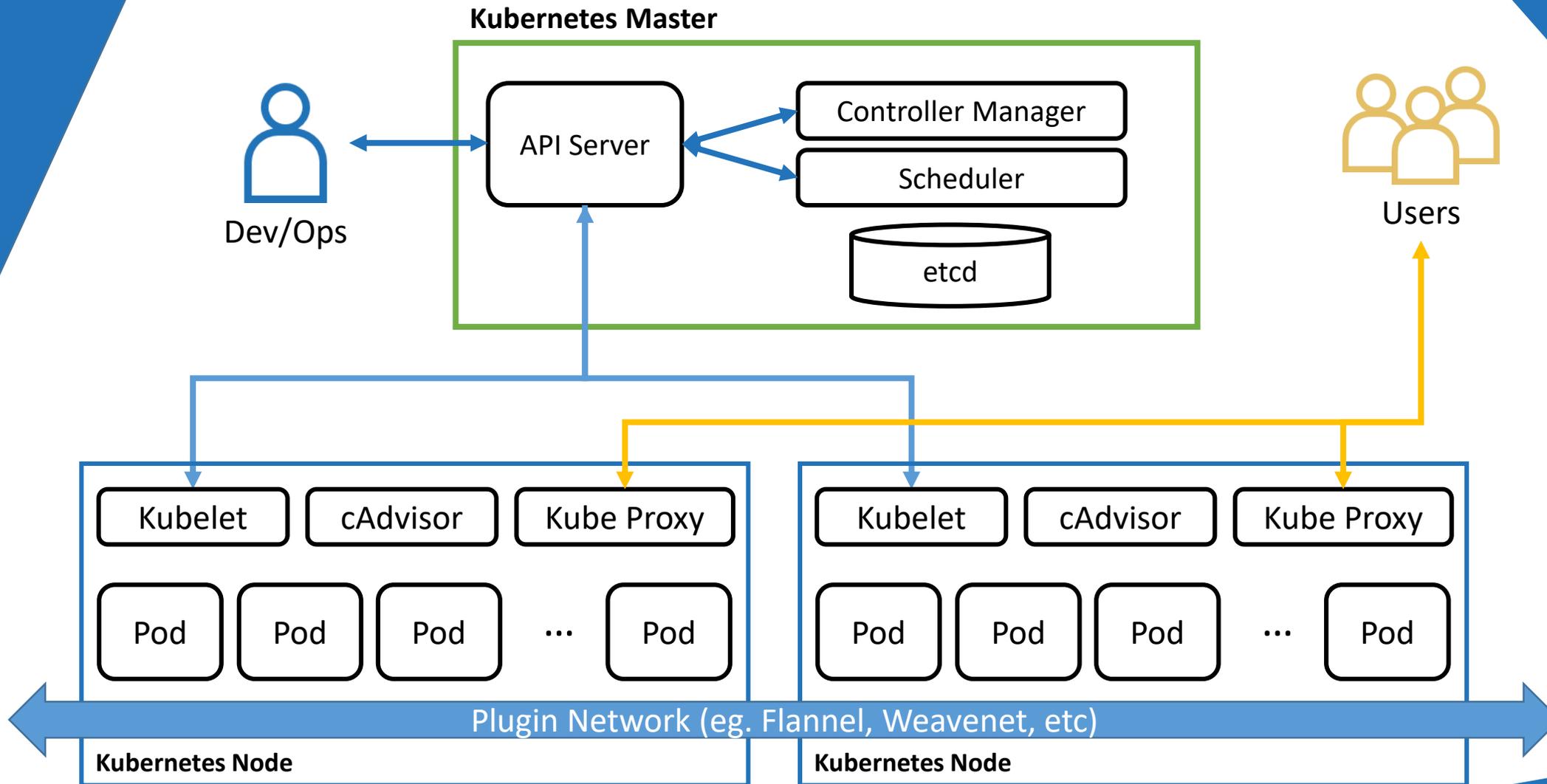
Azure Kubernetes Services

- **Service managé**
- Mise à disposition d'un cluster k8s sans la complexité de déploiement
- Gestion des mises à jour
- Scalabilité à grande échelle
- Intégration avec d'autres briques Azure :
 - Azure Monitor
 - Azure AD
 - Azure Security Center
 - Azure Container Registry



Azure Kubernetes Service (AKS)

Architecture



Concepts

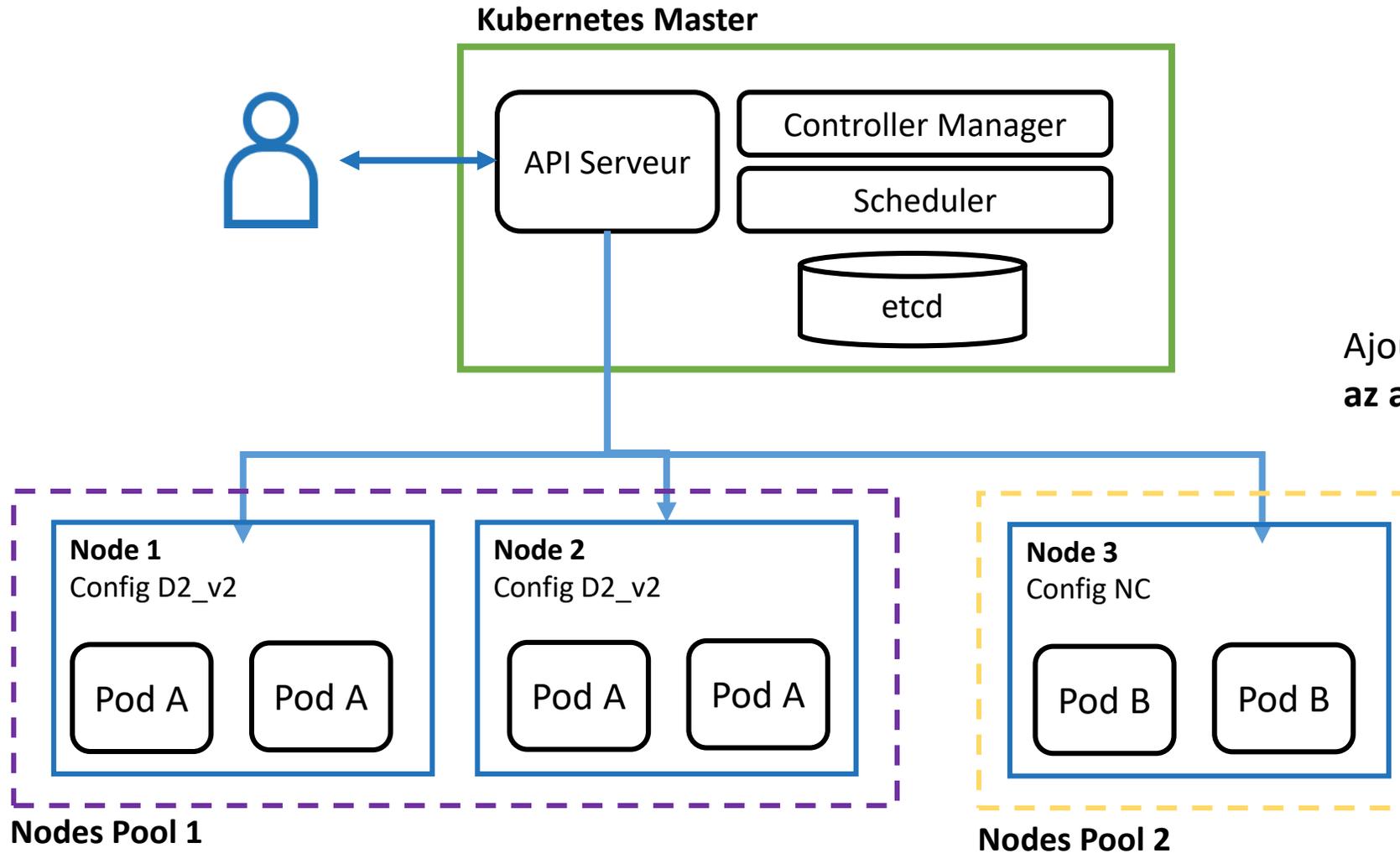
- **Qu'est ce que la haute disponibilité ?**

- Capacité d'un système à être toujours disponible.
- Capacité à s'adapter à des contraintes qui lui sont imposés.

- **Qu'est ce que la résilience ?**

- Capacité d'un service à continuer de fonctionner malgré la défaillance d'un ou plusieurs éléments (base de données, serveurs Web ...)
- Capacité du service à revenir dans un mode nominal de façon automatisée.

Les Pools de Nœuds

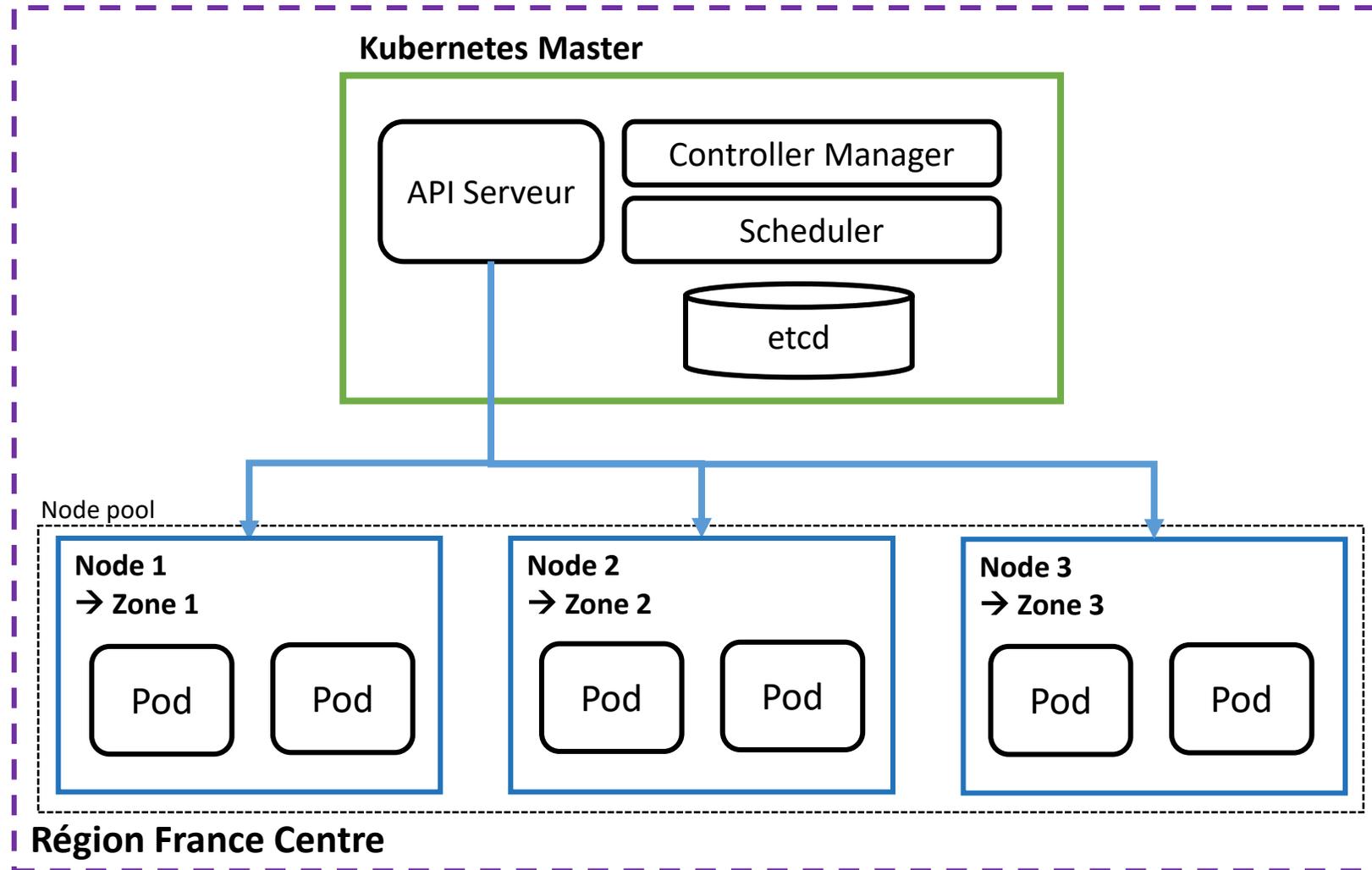


Ajouter un pool de nœud :
`az aks nodepool add ...`



Utiliser différents nodes pools pour mettre en place des nodes adaptés aux besoins applicatifs. Utiliser des taints et tolerations pour associer des pods aux nodes.

Les Availability Zone

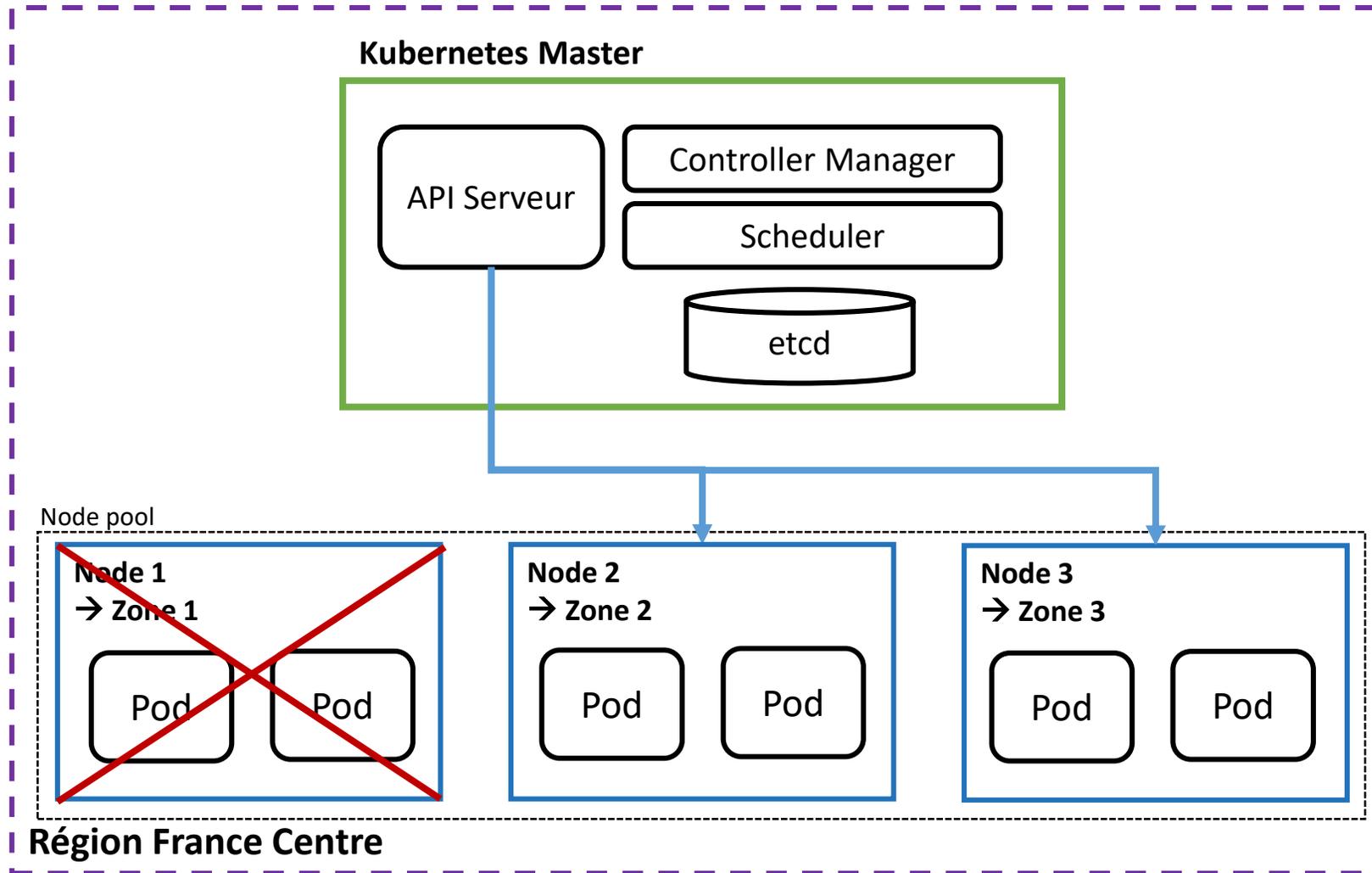


Utiliser les zones de disponibilité :
az aks create ...
--zones 1 2 3



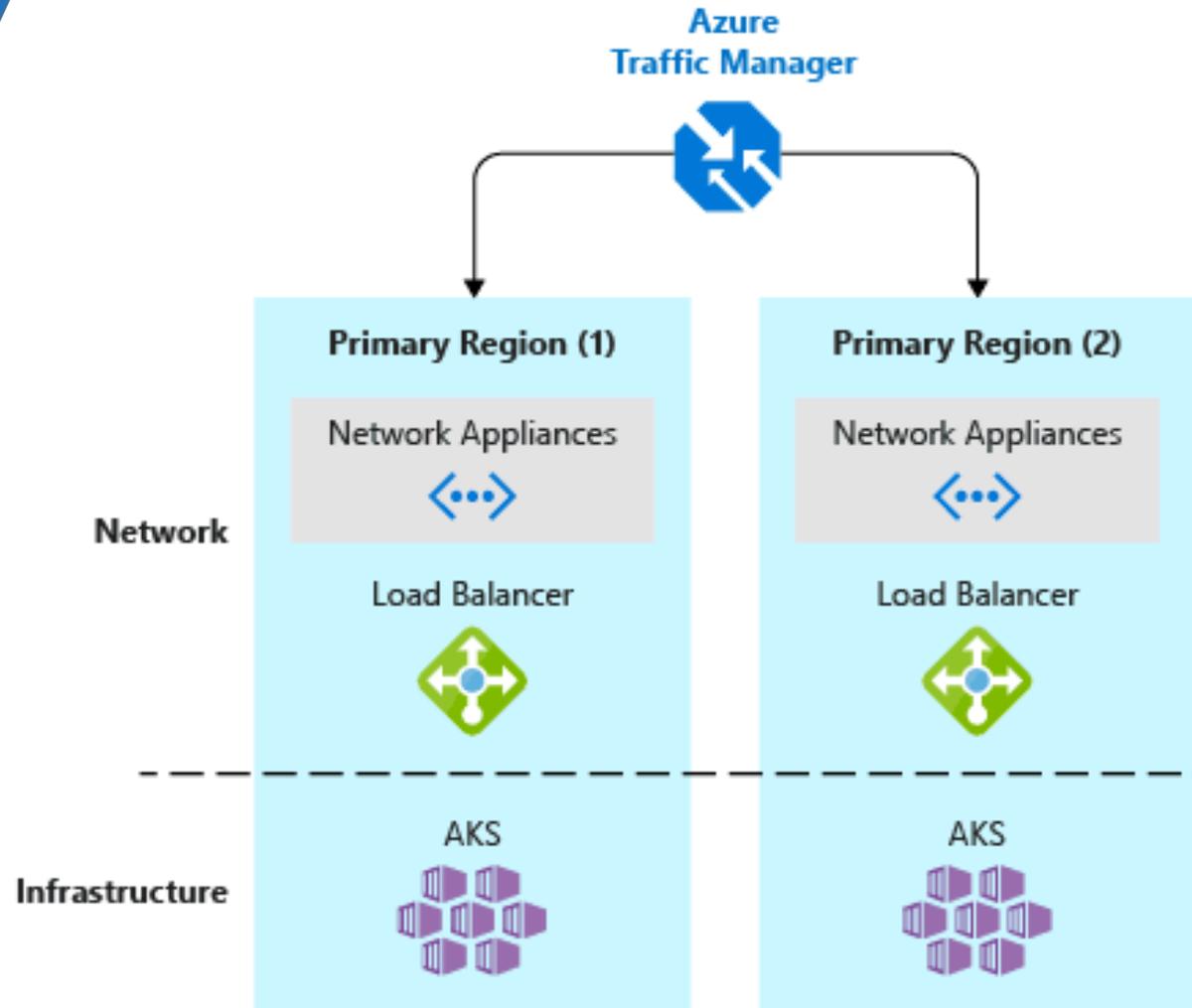
Utiliser des Availability Zone pour se prémunir de l'indisponibilité d'un datacenter d'une région.

Les Availability Zone : la zone 1 est indisponible.



Utiliser des topology spread constraints pour contrôler la propagation des pods sur les nœuds.

Déploiement multi régions

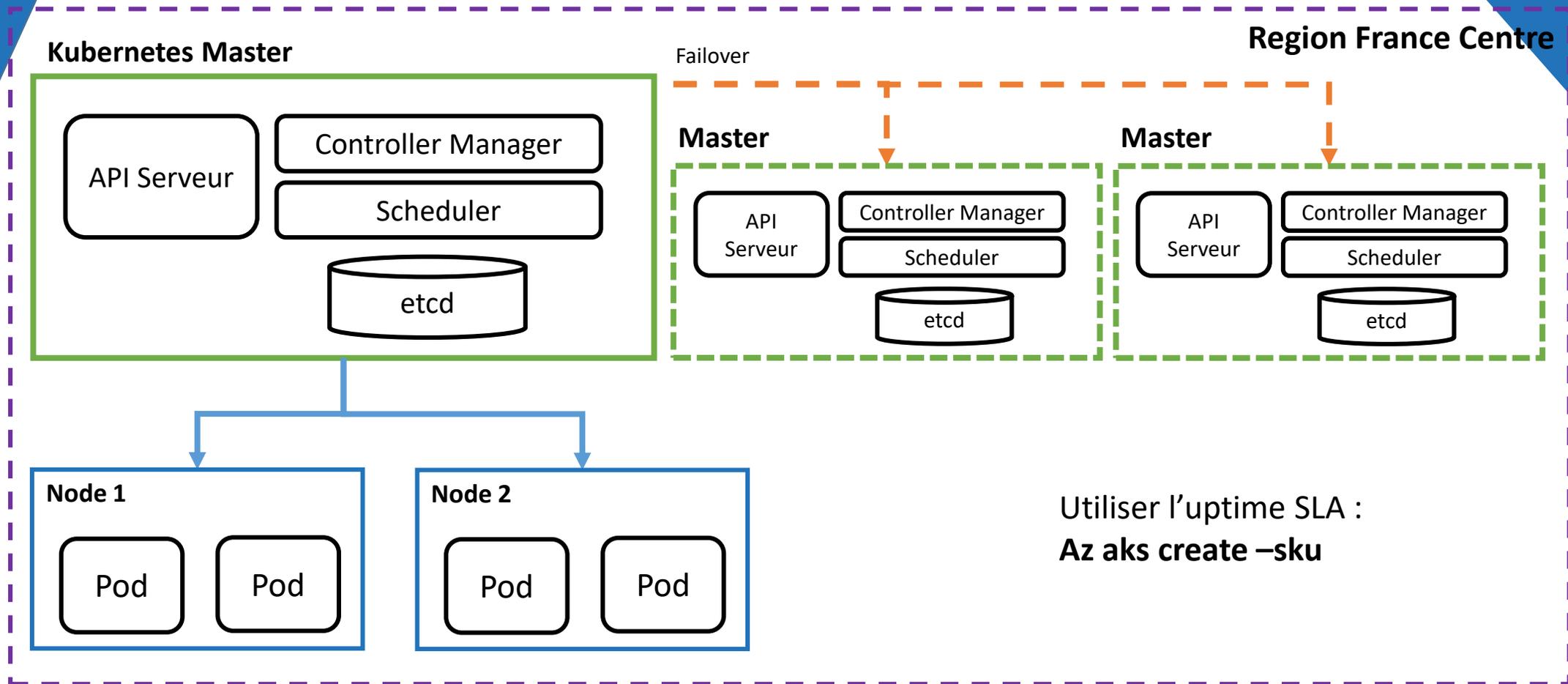


Attention : ici il faut bien gérer deux clusters !



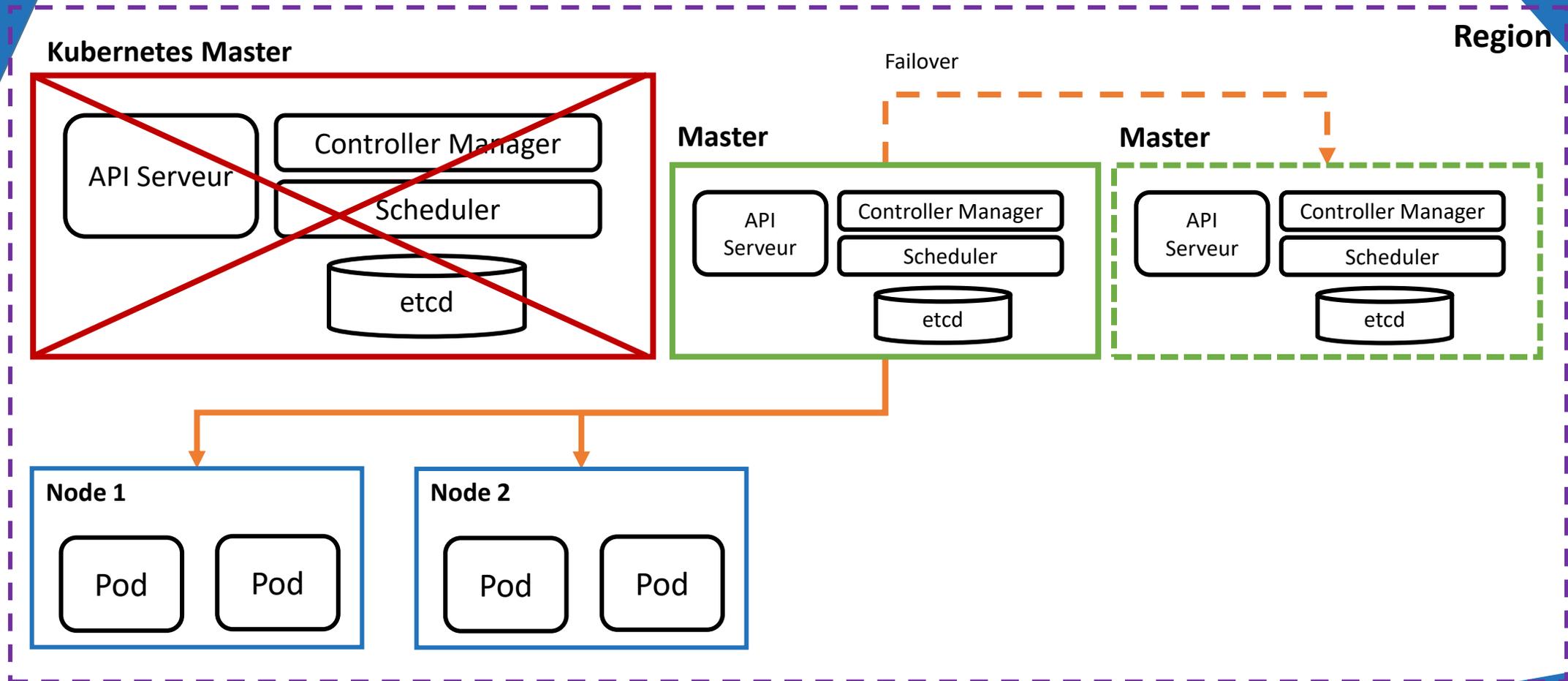
Utiliser un déploiement multi région pour se prémunir de l'indisponibilité d'une région entière.

Uptime SLA



Utiliser l'offre uptime SLA pour garantir la haute disponibilité de l'API Server.

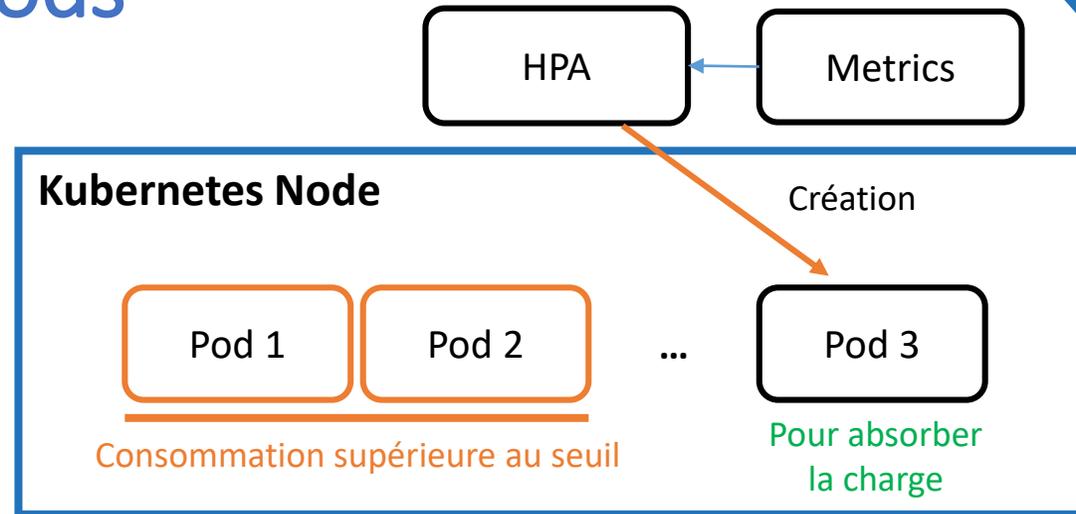
Uptime SLA



L'Autoscaling des Pods

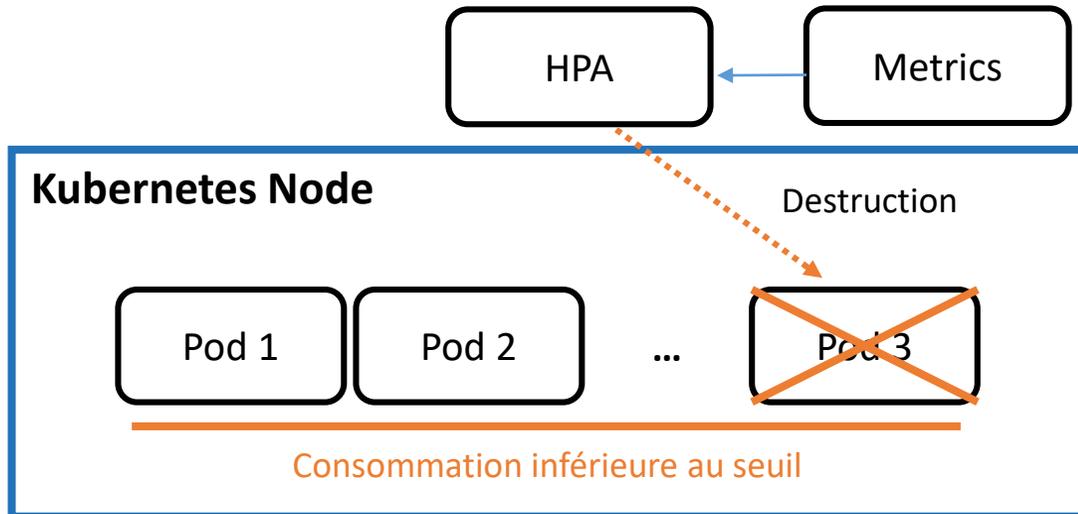
Scale Up

La consommation moyenne des pods d'un deployment > seuil autorisé (ex : 70%)



Scale Down

Tous les pods consomment moins que le seuil.



Créer un HPA :
`kubectl autoscale deployment monapp --max 100 --min 10 --cpu-percent 70`



Définir l'autoscale sur vos deployments pour gérer la montée/descente de charge sur vos applications

L'Autoscaling des Pods

```
apiVersion: v1
kind: Pod
metadata:
  labels:
    test: liveness
  name: liveness-http
spec:
  containers:
  - name: liveness
    image: k8s.gcr.io/liveness
    args:
    - /server
    livenessProbe:
      httpGet:
        path: /healthz
        port: 8080
      httpHeaders:
      - name: Custom-Header
        value: Awesome
      initialDelaySeconds: 3
      periodSeconds: 3
```

```
apiVersion: apps/v1
kind: Deployment
spec:
  replicas: 10
  template:
    spec:
      containers:
      - name: monapplication
        image: acrvacd.azurecr.io/applicationnetcore:v1
        ports:
        - containerPort: 80
        resources:
          requests:
            cpu: 250m
          limits:
            cpu: 500m
```

```
---
apiVersion: v1
kind: Service
spec:
  type: LoadBalancer
  ports:
  - port: 80
  selector:
    app: monapplication
```



Dans vos deployments définir des health probe et readiness probe pour gérer correctement le scaling.



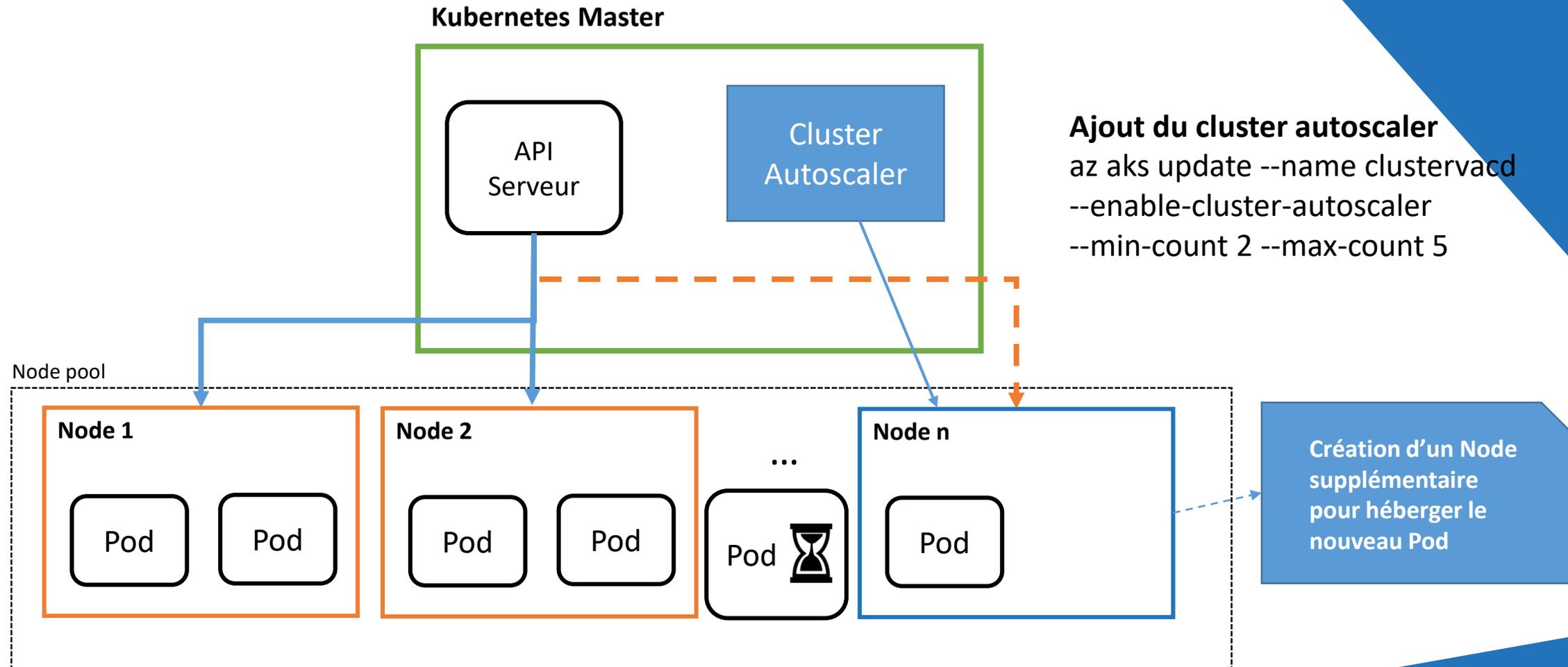
Microsoft



Azure

DEMO

Cluster autoscaler



Définir l'autoscale sur vos nodepools pour provisionner des Nodes en fonction de vos besoins applicatifs

Contexte client

Application de gestion des feuilles de matchs

Très peu utilisé en semaine

Par contre le week-end ...

Site de e-commerce

Journée « normale » et ...

Le Black Friday ou la sortie d'une console par exemple.

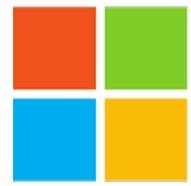
Récapitulatif

Coté infrastructure :

- Configuration des pools de nœuds
- Utilisation des Availability Zones
- Déploiement multi région
- Utilisation de l'offre Uptime SLA

Coté application :

- Autoscaling des pods
- Autoscaling des nodes



Microsoft



Azure

Merci

@thomas_rannou

thomasrannou.azurewebsites.net