Mise en situation professionnelle	**EasyLinux**
	Cours : Kubernetes
Sujet : Sécurisation du cluster	Numéro : 20 à 22 Version : 1.0

Objectifs :

Sécuriser le cluster

Prérequis :

aucun

Principales tâches à réaliser :

Sécurisation du cluster

20 Sauvegarder un master

Kubeadm est une boîte à outils de base qui vous aide à démarrer un cluster Kubernetes simple. Il est destiné à servir de base à des outils de déploiement de plus haut niveau, comme les playbooks Ansible. Un cluster Kubernetes typique configuré avec kubeadm se compose d'un seul maître Kubernetes, qui est la machine coordonnant le cluster, et de plusieurs nœuds Kubernetes, qui sont les machines exécutant la charge de travail réelle.

La gestion des défaillances de nœuds est simple : Lorsqu'un nœud tombe en panne, le maître détecte la panne et reprogramme la charge de travail aux autres nœuds. Pour revenir au nombre de nœuds souhaité, vous pouvez simplement créer un nouveau nœud et l'ajouter au cluster. Pour ajouter un nouveau nœud à un cluster existant, vous créez d'abord un jeton sur le maître avec kubeadm token create, puis vous utilisez ce jeton sur le nouveau nœud pour rejoindre le cluster avec kubeadm join.

Faire face à l'échec du master est plus compliqué. La bonne nouvelle, c'est que L'échec du maître n'est pas aussi grave qu'il n'y paraît. Le cluster et toutes les charges de travail continueront à fonctionner avec exactement la même configuration qu'avant l'échec. Les applications s'exécutant dans le cluster Kubernetes seront toujours utilisables. Cependant, il ne sera pas possible de créer de nouveaux déploiements ou de récupérer des pannes de nœuds sans le master.

21 Cluster HA vs Sauvegarde/Restauration

Une façon de faire face aux échecs du master est de mettre en place un cluster à haute disponibilité. L'idée est de mettre en place un cluster etcd répliqué, et d'exécuter une instance etcd avec chaque instance master. De cette façon, aucune donnée ne sera perdue si une seule instance master échoue.

L'approche HA n'est pas toujours meilleure qu'une configuration à master unique :

- La surveillance d'un seul master est beaucoup plus simple que la surveillance d'un cluster etcd répliqué. L'exécution d'un cluster etcd répliqué ne vous dispense pas de la nécessité de mettre en place une surveillance.
- Le temps de récupération n'est pas nécessairement plus rapide lorsqu'on compare une configuration HA avec une configuration mono maître. Dans une configuration à master unique, l'outil de surveillance détectera les défaillances du master et déclenchera un script de restauration automatique. Ce n'est pas nécessairement plus lent que le mécanisme de basculement de etcd.
- La configuration HA ne vous empêche pas d'implémenter la sauvegarde/restauration pour le master, car il est toujours possible que vous détruisiez accidentellement les données dans un cluster etcd répliqué.

Le principal inconvénient de l'approche de sauvegarde/restauration est qu'il n'y a pas de sauvegarde en temps réel de l'état du cluster. Comme indiqué ci-dessous, nous utiliserons un CronJob Kubernetes pour créer des sauvegardes du cluster etcd. Lorsque le master échoue, tous les changements dans la configuration du cluster après la dernière exécution de CronJob sont perdus. Le cluster sera restauré exactement dans le même état (déploiements, services, etc.) que lors de l'exécution du dernier CronJob. Bien que vous puissiez exécuter le CronJob toutes les minutes, vous n'obtiendrez pas la sauvegarde en temps réel qu'un cluster etcd répliqué fournit.

La pertinence d'un déploiement d'HA dépend de la façon dont vous utilisez le cluster. Si vous changez la configuration du cluster très fréquemment (déployez des applications si souvent que vous avez besoin d'une sauvegarde en temps réel de l'état du cluster), vous pourriez bénéficier de la duplication des données etcd dans une configuration HA. Si votre cluster change à un rythme plus lent, l'approche de sauvegarde/restauration pourrait être la meilleure option car elle simplifie les opérations.

22 Données à sauvegarder

Deux éléments de données à sauvegarder :

- Les fichiers de certificat racine /etc/kubernetes/pki/ca.crt et /etc/kubernetes/pki/ca.key.
- Les données etcd.

La sauvegarde du certificat racine est une opération unique que vous effectuerez manuellement après avoir créé le master avec kubeadm init. Le reste traite de la façon de sauvegarder les données etcd.

22.1 Sauvegarder Etcd

Un CronJob Kubernetes pour sauvegarder les données etcd

Comme indiqué dans la documentation etcd, vous créez une sauvegarde des données etcd avec

```
ETCDCTL_API=3
etcdctl --endpoints $ENDPOINT snapshot save snapshot.db
```

Nous allons créer un CronJob Kubernetes pour exécuter cette commande périodiquement. Il n'est pas nécessaire d'installer etcdctl sur le système hôte ou de configurer un travail cron sur le système hôte.

L'en-tête, nous voulons exécuter le CronJob dans l'espace de nommage kube-system :

```
apiVersion: batch/v1beta1
kind: CronJob
metadata:
  name: backup
  namespace: kube-system
```

Exemple un lancement toutes les trois minutes :

```
spec:
  schedule: "*/3 * * * *"
  jobTemplate:
    spec:
```

Un master Kubernetes utilise un module statique pour exécuter etcd. Nous réutilisons l'image du docker etcd de ce pod pour le CronJob. La définition du pod se trouve dans /etc/kubernetes/manifests/etcd.yaml.

```
      template:
        spec:
          containers:
          - name: backup
            # Same image as in /etc/kubernetes/manifests/etcd.yaml
            image: k8s.gcr.io/etcd-amd64:3.1.12
```

La commande ci-dessous créera des fichiers de sauvegarde du type :

/backup/etcd-snapshot-2018-05-24_21:54:03_UTC.db

Les paramètres additionnels à etcdctl spécifient les certificats et l'URL auxquels accéder etcd.

```
            env:
            - name: ETCDCTL_API
              value: "3"
            command: ["/bin/sh"]
            args: ["-c", "etcdctl --endpoints=https://127.0.0.1:2379
--cacert=/etc/kubernetes/pki/etcd/ca.crt --cert=/etc/kubernetes/pki/etcd/healthcheck-
client.crt --key=/etc/kubernetes/pki/etcd/healthcheck-client.key snapshot save
/backup/etcd-snapshot-$(date +%Y-%m-%d_%H:%M:%S_%Z).db"]
```

La commande etctctl ci-dessus suppose que les volumes suivants ont été mappés dans le conteneur Docker :

- /etc/kubernetes/pki/etcd : Le chemin sur le système hôte où les informations d'identification etcd sont stockées.
- /Backup : Le volume persistant où les sauvegardes sont créées.

Il faudra monter les volumes :

```
            volumeMounts:
            - mountPath: /etc/kubernetes/pki/etcd
              name: etcd-certs
              readOnly: true
            - mountPath: /backup
              name: backup
```

L'API etcd est disponible sur le port 2379 du système hôte. Nous devrons exécuter le CronJob dans l'espace de noms du réseau hôte pour qu'il puisse accéder à etcd en utilisant l'adresse IP loopback 127.0.0.1.

```
          hostNetwork: true
```

Normalement, Kubernetes empêche la planification des pod sur le master. Toutes les charges de travail sont exécutées sur les nœuds. Pour forcer le CronJob à s'exécuter sur le maître, nous spécifions un nodeSelector (voir affinité des nœuds), et nous spécifions que le CronJob doit "tolérer" l'effet NoSchedule qui empêche l'exécution des charges sur le maaster.

```
          nodeSelector:
            kubernetes.io/hostname: kube-master
          tolerations:
          - effect: NoSchedule
            operator: Exists
          restartPolicy: OnFailure
```

Enfin, nous devons définir les volumes utilisés dans le volumeMount. Le premier volume est le chemin sur le système hôte où les informations d'identification etcd sont stockées. Ces informations d'identification sont créées lorsque le cluster est configuré avec kubeadm init.

```
          volumes:
          - name: etcd-certs
            hostPath:
              path: /etc/kubernetes/pki/etcd
              type: DirectoryOrCreate
```

Le second volume est un volume persistant où la sauvegarde sera stockée. Ici, on va monter un partage CIFS donc j'utilise le plugin fstab/cifs. Cependant, il est possible d’utiliser un autre type de volume persistant, selon l'endroit où vous souhaitez stocker la sauvegarde.

```
          - name: backup
            flexVolume:
              driver: "fstab/cifs"
              fsType: "cifs"
              secretRef:
                name: "backup-volume-credentials"
              options:
                networkPath: "//my-server.com/backup"
                mountOptions: "dir_mode=0755,file_mode=0644,noperm"
```

Pour créer le CronJob, créer un fichier backup-cron-job.yml et exécuter

```
$ kubectl apply -f backup-cron-job.yml
```

Le plugin fstab/cifs nécessite un secret, c'est-à-dire le nom d'utilisateur et le mot de passe pour monter le volume CIFS. Vous n'en avez pas besoin si vous utilisez un autre type de volume persistant. Le secret est défini comme suit (nom d'utilisateur et mot de passe sont encodés en base64) :

```
apiVersion: v1
kind: Secret
metadata:
  name: backup-volume-credentials
  namespace: kube-system
type: fstab/cifs
data:
  username: 'ZXhhbXBsZQ=='
  password: 'bXktc2VjcmV0LXBhc3N3b3Jk'
```

Le fichier backup-volume-secret.yml est proposé ci-dessus, exécuter

```
$ kubectl apply -f backup-volume-secret.yml
```

NB : Le secret est spécifique au plugin volume fstab/cifs utilisé ici.

23 Restaurer un master

En cas de panne, créer un nouveau maître et l'initialiser avec les données de la sauvegarde. Avant d'exécuter kubeadm init sur le nouveau maître, il faut restaurer les données de la sauvegarde.

Commencer à restaurer les fichiers de certificat racine /etc/kubernetes/pki/ca.crt et /etc/kubernetes/pki/ca.key. Les permissions à donner sont 0644 pour ca.crt et 0600 pour ca.key.

Ensuite exécuter etcdctl pour restaurer la sauvegarde etcd. Il n'est pas indispensable d'installer etcdctl sur le système hôte, car on peut utiliser l'image Docker.

Supposons que la dernière sauvegarde soit stockée dans /mnt/etcd-snapshot-2018-05-24_21:54:54:03_UTC.db, Lancer :

```
mkdir -p /var/lib/etcd
docker run --rm \
    -v '/mnt:/backup' \
    -v '/var/lib/etcd:/var/lib/etcd' \
    --env ETCDCTL_API=3 \
    'k8s.gcr.io/etcd-amd64:3.1.12' \
    /bin/sh -c "etcdctl snapshot restore '/backup/etcd-snapshot-2018-05-
24_21:54:03_UTC.db' ; mv /default.etcd/member/ /var/lib/etcd/"
```

La commande ci-dessus crée un répertoire /var/lib/etcd/member/ avec les permissions 0700.

Enfin, lancer kubeadm init pour créer le nouveau master. Cependant, nous avons besoin d'un paramètre supplémentaire pour qu'il accepte les données etcd existantes :

```
kubeadm init –ignore-preflight-errors=DirAvailable--var-lib-etcd
```

En supposant que le nouveau master est accessible sous la même adresse IP ou le même nom d'hôte que l'ancien master, les nœuds se reconnecteront et le cluster sera à nouveau opérationnel.

24 Annexe

24.1 Fichier yaml complet

```
apiVersion: batch/v1beta1
kind: CronJob
metadata:
  name: backup
  namespace: kube-system
spec:
  # activeDeadlineSeconds: 100
  schedule: "*/3 * * * *"
  jobTemplate:
    spec:
      template:
        spec:
          containers:
          - name: backup
            # Same image as in /etc/kubernetes/manifests/etcd.yaml
            image: k8s.gcr.io/etcd-amd64:3.1.12
            env:
            - name: ETCDCTL_API
              value: "3"
            command: ["/bin/sh"]
            args: ["-c", "etcdctl --endpoints=https://127.0.0.1:2379
--cacert=/etc/kubernetes/pki/etcd/ca.crt --cert=/etc/kubernetes/pki/etcd/healthcheck-
client.crt --key=/etc/kubernetes/pki/etcd/healthcheck-client.key snapshot save
/backup/etcd-snapshot-$(date +%Y-%m-%d_%H:%M:%S_%Z).db"]
            volumeMounts:
            - mountPath: /etc/kubernetes/pki/etcd
              name: etcd-certs
              readOnly: true
            - mountPath: /backup
              name: backup
          restartPolicy: OnFailure
          nodeSelector:
            kubernetes.io/hostname: kube-master
          tolerations:
          - effect: NoSchedule
            operator: Exists
          hostNetwork: true
          volumes:
          - name: etcd-certs
            hostPath:
              path: /etc/kubernetes/pki/etcd
              type: DirectoryOrCreate
          - name: backup
            flexVolume:
              driver: "fstab/cifs"
              fsType: "cifs"
              secretRef:
                name: "backup-volume-credentials"
              options:
                networkPath: "//my-server.com/backup"
```

```
mountOptions: "dir_mode=0755,file_mode=0644,noperm"
```