L'orchestration de conteneurs avec Kubernetes

Virtualisation avancée - CM

Maxime Lambert

2023 / 2024

BUT Informatique - Semestre 5 - Ressource R5.09

Plan

- Introduction à Kubernetes
- Les principes fondamentaux
- Déploiement d'applications
- Gestion de la scalabilité
- Gestion du stockage
- Gestion de la sécurité et des autorisations
- Observabilité et traçabilité

Avant-propos

- Au semestre 4, au cours du module de virtualisation, vous avez étudié les conteneurs sous Linux en pratiquant avec la plateforme Docker.
- Vous avez déployé des conteneurs, réseaux et volumes en utilisant l'interface en ligne de commande de Docker. Vous avez également été initié à l'orchestration grâce à l'outil Docker Compose.
- Bien qu'acceptable dans un contexte de prototypage ou de développement, ces méthodes ne conviennent pas pour exploiter une application dans un environnement de production.
- Kubernetes est un outil largement adopté par l'industrie qui offre une approche normalisée pour la gestion des applications conteneurisées. C'est aussi un outil open source ce qui aide à éviter de se retrouver dans une situation de verrouillage du fournisseur (*vendor lock-in)*. C'est pour cette raison qu'il a été retenu dans la cadre de ce cours.
- Veuillez noter que nous nous placerons plutôt dans le rôle d'utilisateur de Kubernetes, l'administration d'un cluster de Kubernetes pour satisfaire aux exigences de la production est encore un autre sujet… Et il existe des alternatives que nous détaillerons plus tard.

Introduction à Kubernetes

Introduction à Kubernetes

Présentation de la plateforme

- Kubernetes est **une plateforme open-source** extensible et portable pour la gestion de charges de travail et de services conteneurisés.
- Le projet a été **rendu open-source par Google en 2014**, première version stable en 2015
- API standard pour la création d'applications ***cloud native***
  - Vous êtes invité à consulter la méthodologie *The Twelve-factor App*[1] qui dispense plusieurs bonnes pratiques qu'il est important de connaître.
- Permet de construire et déployer des **systèmes distribués fiables et évolutifs**

[1] https://12factor.net/fr

Introduction à Kubernetes

Intégration dans l'écosystème *Cloud*

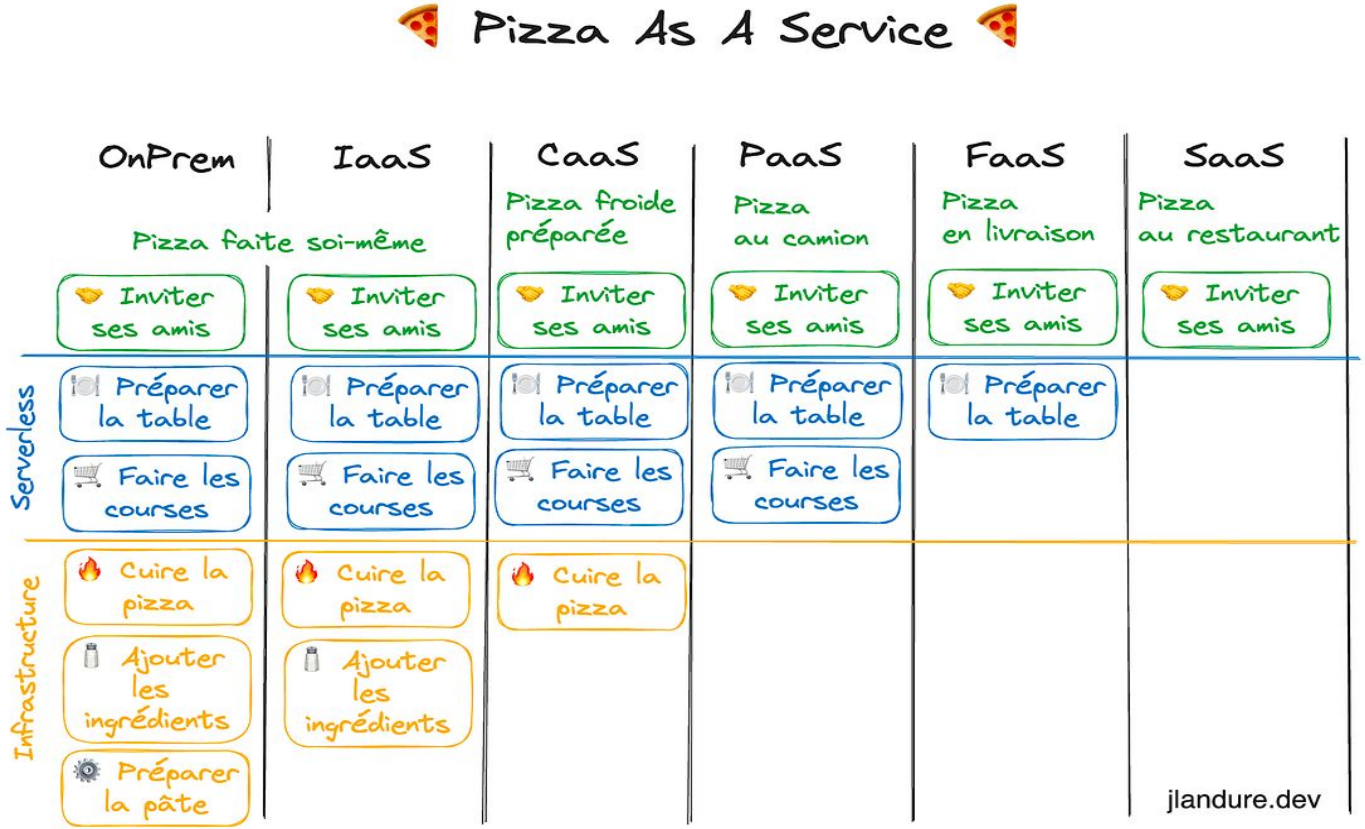
- Les services *Cloud* mettent à disposition à travers le réseau des ressources et fonctionnalités tout **en rendant abstrait les détails de l'infrastructure sous-jacente** pour les utilisateurs.

- Kubernetes fournit un **environnement de gestion focalisé sur le conteneur**. Il **orchestre** les ressources machines, la mise en réseau et l'infrastructure de stockage. On qualifie ce type de plateforme de **CaaS** (*Container as a Service*). Du fait de sa situation de monopole, vous rencontrerez également le terme KaaS (*Kubernetes as a Service*).

- Il est nécessaire de provisionner en amont une infrastructure pour y déployer Kubernetes. Ce qu'il est possible de faire manuellement, de manière automatisée par exemple grâce à Terraform que nous abordé au semestre 4, ou alors vous pouvez déléguer cette tâche à un tiers.

Introduction à Kubernetes

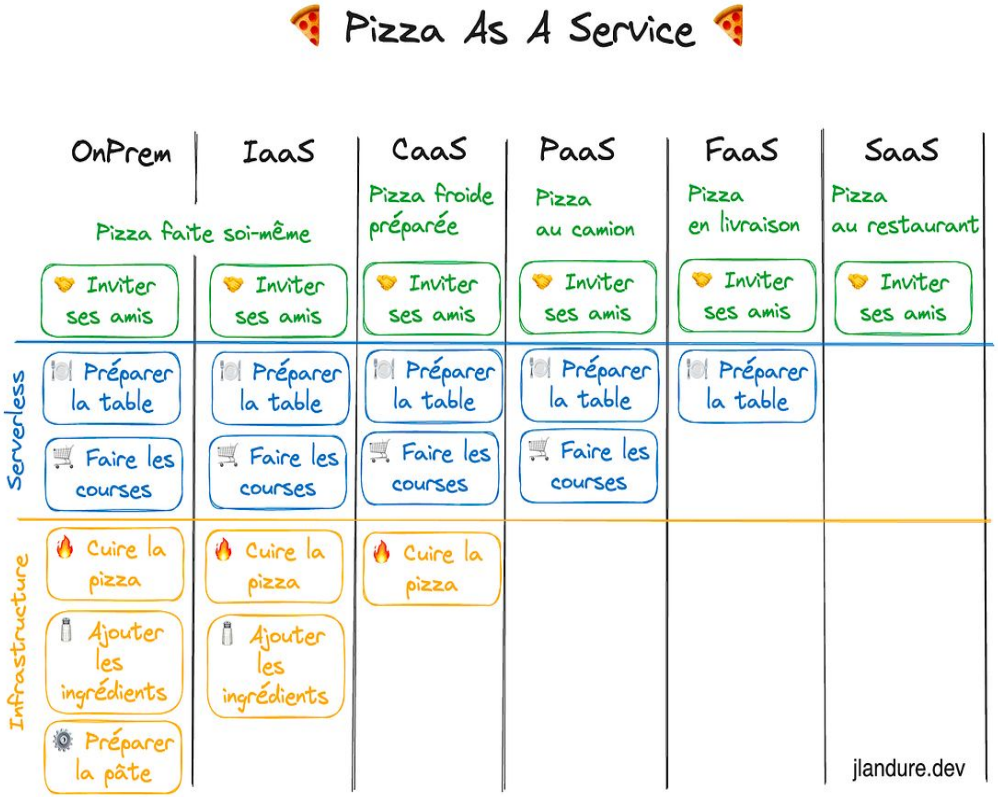
Comprendre le modèle “as a Service”



Introduction à Kubernetes

Comprendre le modèle “as a Service”

- ***On-Premise***: Contrôle totale sur l’infrastructure. Système informatique hébergés, gérés et maintenus directement par l’organisation.
- ***IaaS*** *(Infrastructure as a Service)*: Virtualisation des principaux composants d’infrastructure (calcul, stockage, réseau, sécurité et répartition de charge).
- ***CaaS*** *(Container as a Service)*: Mise à disposition par un fournisseur *cloud* d’un orchestrateur de conteneurs managé. Permet d’être utilisateur de Kubernetes et non administrateur de Kubernetes.



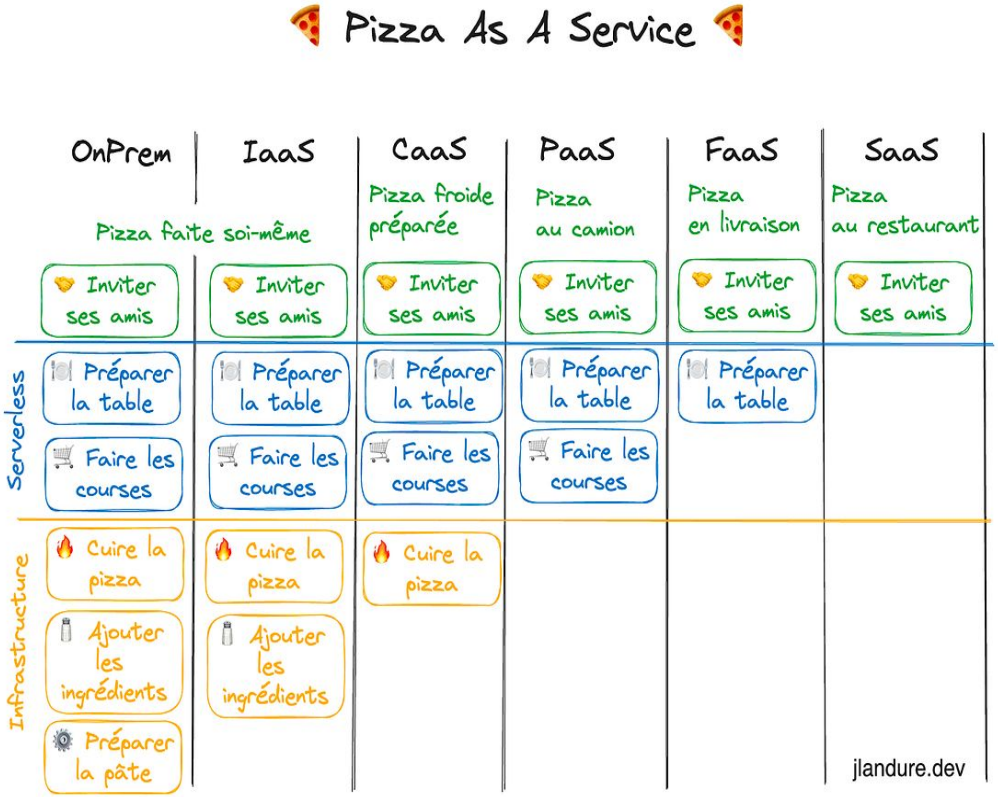
Introduction à Kubernetes

Comprendre le modèle “as a Service”

- ***PaaS*** *(Platform as a Service)*: Téléversement du code chez un fournisseur *cloud* qui disposera de l’outillage nécessaire pour construire et déployer l’application web, tout en proposant des fonctionnalités supplémentaires (bases de données, stockage objet…)

- ***FaaS*** *(Function as a Service)*: Déploiement d’un morceau de code qui réagit à un évènement (par exemple un appel HTTP)

- **SaaS**: Le logiciel n’est pas installé. Il est consommé au travers d’une interface (généralement web). Les données manipulées sont stockées sur les serveurs de l’éditeur.



Introduction à Kubernetes

Comment K8s aide à déployer rapidement et de manière fiable

- **Kubernetes encourage le concept d'immutabilité** en incitant les utilisateurs à créer de nouveaux déploiements au lieu de mettre à jour des conteneurs en cours d'exécution. Cette approche garantit que chaque modification est effectuée à travers une nouvelle version, préservant ainsi la stabilité de l'application. L'immutabilité facilite également le suivi des changements, la gestion des versions et la récupération en cas d'incident.

- **Kubernetes adopte un modèle de configuration déclarative**, où les utilisateurs décrivent l'état souhaité de leur application via des fichiers de configuration YAML. L'objet de configuration détermine la manière dont les ressources, telles que les pods, les services et les volumes, doivent être créées et gérées. Cette approche simplifie la gestion, la reproductibilité et le contrôle des ressources, tout en évitant la configuration manuelle.

- **Kubernetes est conçu pour être auto-guérisseur**, ce qui signifie qu'il surveille en permanence l'état des ressources et prend des mesures pour maintenir la santé de l'application. En cas de défaillance d'un pod ou d'une machine, Kubernetes redémarre automatiquement les pods ou planifie leur réparation sur d'autres nœuds sains. Cette capacité d'auto-guérison contribue à maintenir la disponibilité et la fiabilité des applications dans un environnement dynamique.

Introduction à Kubernetes

Gérer l'évolution d'une organisation avec K8s

- **Kubernetes favorise une architecture découplée en utilisant des API et des équilibreurs de charge** pour isoler chaque composant du système. Les API agissent comme des tampons entre les composants, permettant l'évolutivité des programmes sans ajuster les autres couches du service. Ce découplage simplifie également la gestion et la communication entre les équipes de développement, favorisant ainsi l'évolutivité des équipes.

- **L'évolutivité des applications est simple grâce à la nature immuable et déclarative de K8s.** Sous réserve de bénéficier des ressources disponibles dans le cluster; la modification d'un nombre dans un fichier de configuration suffit pour mettre à l'échelle un service.

- L'évolution des besoins humains peut compliquer l'organisation du travail en équipes. **Maintenir des équipes de petite taille est essentiel pour une communication efficace. L'architecture en micro-services offre une solution** en découpant le système en composants indépendants, permettant à chaque équipe de se concentrer sur un micro-service spécifique. Cela favorise l'évolutivité des équipes et la flexibilité pour gérer des besoins en constante évolution. Kubernetes offre des abstractions et des API qui facilitent ce type d'architecture..

- Kubernetes offre une cohérence accrue de l'infrastructure en séparant les problèmes et en découplant les conteneurs des machines. Cette séparation permet à une petite équipe de gérer de nombreuses machines de manière efficace. De plus, **K8s crée un contrat clair grâce à son API d'orchestration de conteneurs, délimitant clairement les rôles et responsabilités entre l'opérateur d'applications et l'opérateur d'orchestration de clusters.** Cela simplifie la gestion et l'évolutivité de l'infrastructure.

Introduction à Kubernetes

Une application cloud-native, qu'est-ce que c'est ?

“Une application cloud-native se compose de services plus petits, indépendants et faiblement couplés. Elle est conçue de façon à apporter une valeur métier incontestée, comme la capacité à prendre en compte rapidement l'avis des utilisateurs dans un effort d'amélioration continue. En d'autres termes, le développement d'applications cloud-native permet d'accélérer la création des nouvelles applications, d'optimiser les anciennes et de les connecter les unes aux autres. L'objectif est double : fournir aux utilisateurs les applications dont ils ont besoin tout en suivant le rythme imposé par leur activité.

Quel est le sens du mot « cloud » dans ce contexte ? Lorsque l'on dit d'une application qu'elle est « native pour le cloud », cela signifie qu'elle a été conçue spécialement pour offrir une expérience cohérente de développement et de gestion automatisée dans les clouds privés, publics et hybrides. Aujourd'hui, les entreprises adoptent le cloud computing pour améliorer l'évolutivité et la disponibilité de leurs applications, grâce à l'approvisionnement en libre-service et à la demande des ressources ainsi qu'à l'automatisation du cycle de vie des applications (de la phase de développement jusqu'à la production).

Toutefois, pour véritablement bénéficier de ces avantages, elles doivent mettre en place une nouvelle stratégie de développement des applications.

C'est tout l'enjeu du développement d'applications cloud-native : créer et mettre à jour rapidement des applications et, dans le même temps, améliorer la qualité et réduire les risques. Plus précisément, l'approche permet de développer et d'exécuter des applications réactives, évolutives et résistantes aux pannes dans toute architecture, que ce soit un cloud public, privé ou hybride.“

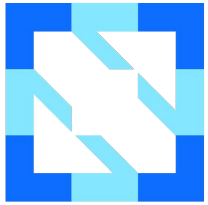
***source: redhat.com**/fr/topics/cloud-native-apps*

Introduction à Kubernetes

La Cloud Native Computing Foundation (CNCF)

- **La CNCF est une organisation à but non lucratif qui a pour mission de favoriser l'adoption de technologies cloud natives et de soutenir les projets open source liés à ces technologies.** Elle a été créée pour répondre aux besoins croissants des entreprises en matière de déploiement et de gestion d'applications dans des environnements cloud.

- La CNCF dispose de trois niveaux de maturité pour guider dans l'adoption de projets cloud native :
  - **Bac à sable:** La majorité des projets sont à ce statut. Il indique un projet encore dans un stade de développement précoce et il n'est pas recommandé de l'adopter.
  - **Incubation**: Projet ayant démontré sa viabilité, adopter par la communauté et conforme aux principes cloud natives.
  - **Graduation**: Seuls quelques projets ayant atteint un haut niveau de maturité, stabilité et d'adoption appartiennent à cette catégorie.



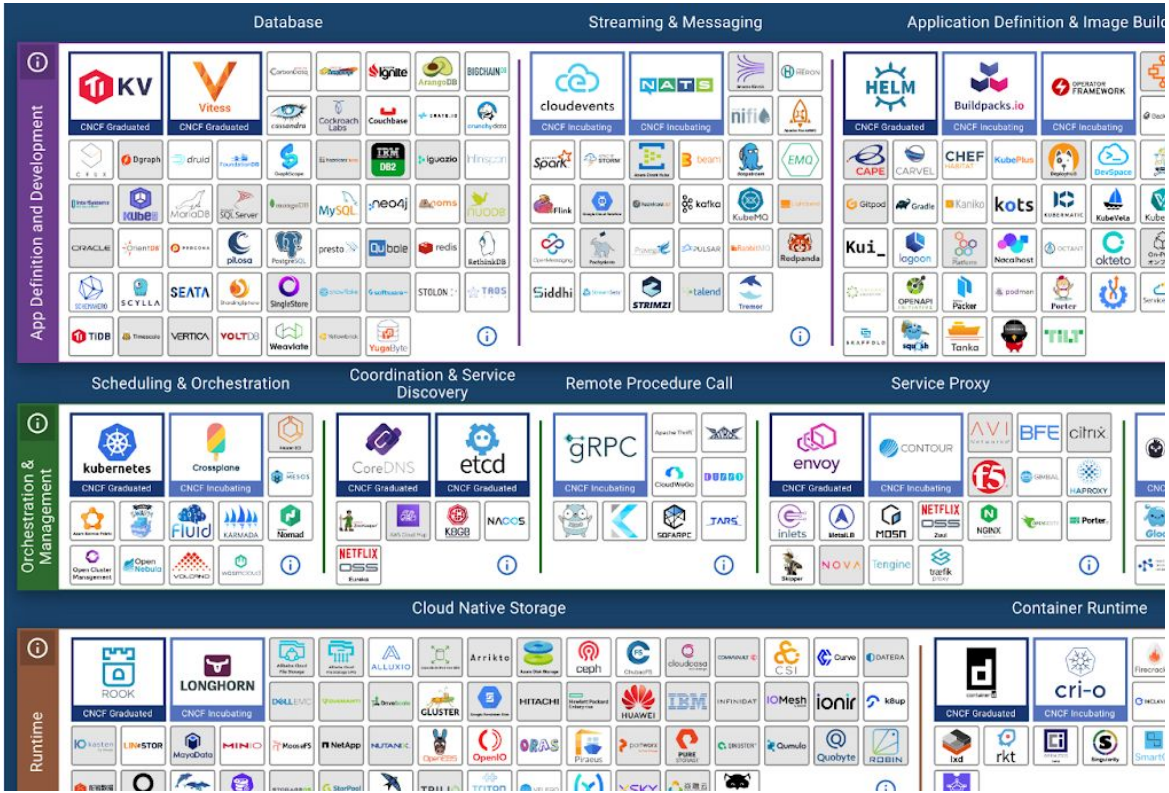
CLOUD NATIVE
COMPUTING FOUNDATION

Introduction à Kubernetes

La CNCF Landscape

- Après la sortie de Kubernetes, de nombreux outils ont été développés pour une variété de tâches, de l'apprentissage automatique aux modèles de programmation sans serveur.

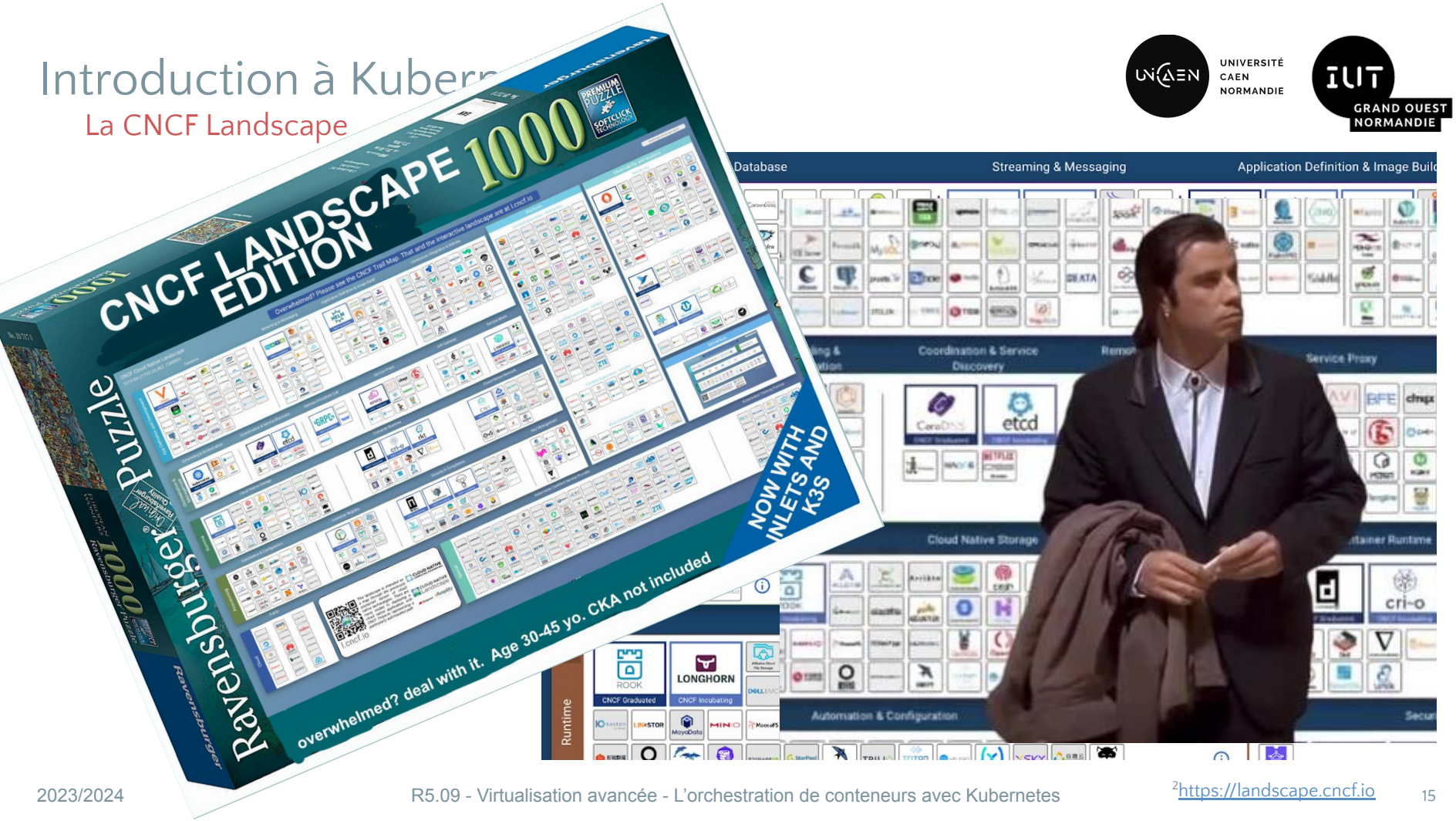
- Cependant, la difficulté réside souvent dans le choix de la meilleure solution parmi les nombreuses disponibles. Pour vous aider, vous avez à votre disposition la CNCF Landscape[2]...



[2] https://landscape.cncf.io

Introduction à Kuberr

La CNCF Landscape



[2] https://landscape.cncf.io

Introduction à Kubernetes

Les outils de base

- **kubectl** : interface en ligne de commandes officiel de Kubernetes. Avec cet outil open source vous pouvez déployer des applications, inspecter et gérer les ressources clusters ou encore afficher les journaux. Il est disponible sous Linux, macOS et Windows.

- **OpenLens** : interface graphique open source relativement conviviale qui facilite la visualisation et la gestion des ressources dans un cluster. Disponible sous Linux, macOS et Windows.

- **kind et minikube** : Ces deux outils permettent la création d'environnement de développement Kubernetes locaux mais présentent quelques différences. La principale est que le premier instancie un cluster à partir de conteneurs Docker alors que le second utilise une machine virtuelle.

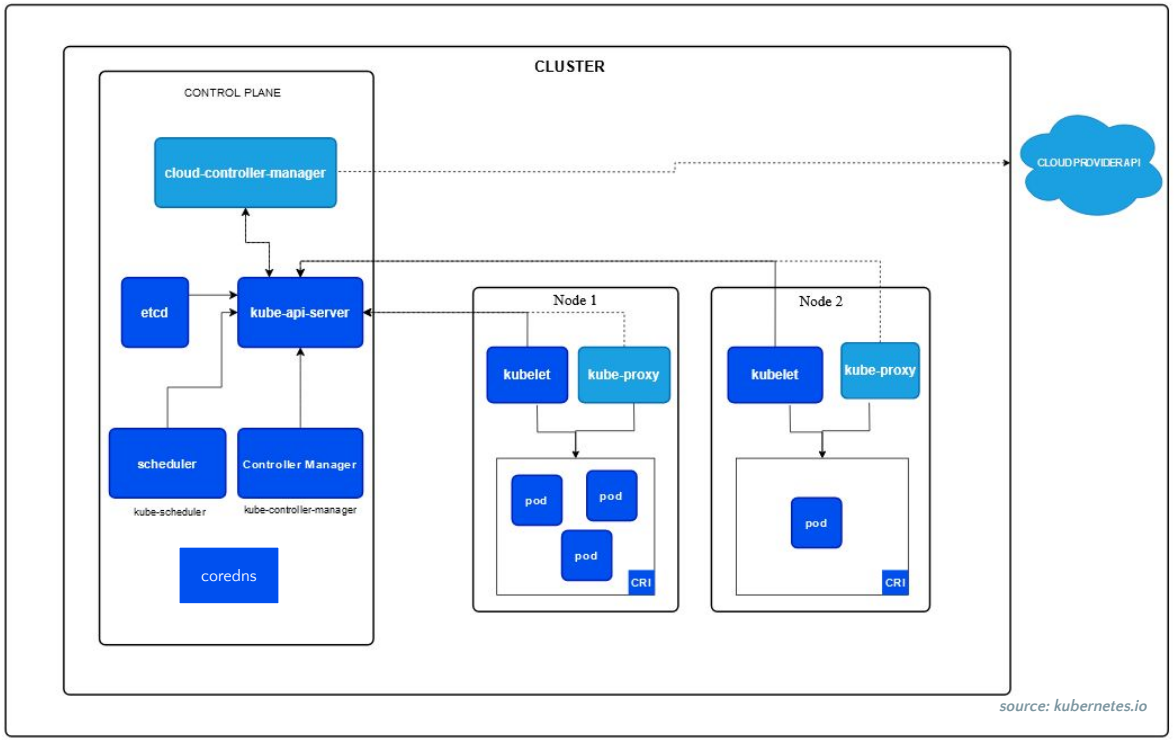
- **kubeadm** : outil destiné à l'initialisation de clusters kubernetes de production. Cet outil n'est donc pas nécessaire dans le cadre de ce module du fait de l'utilisation de kind en TP.

Les principes fondamentaux

Les principes fondamentaux

L'architecture d'un cluster

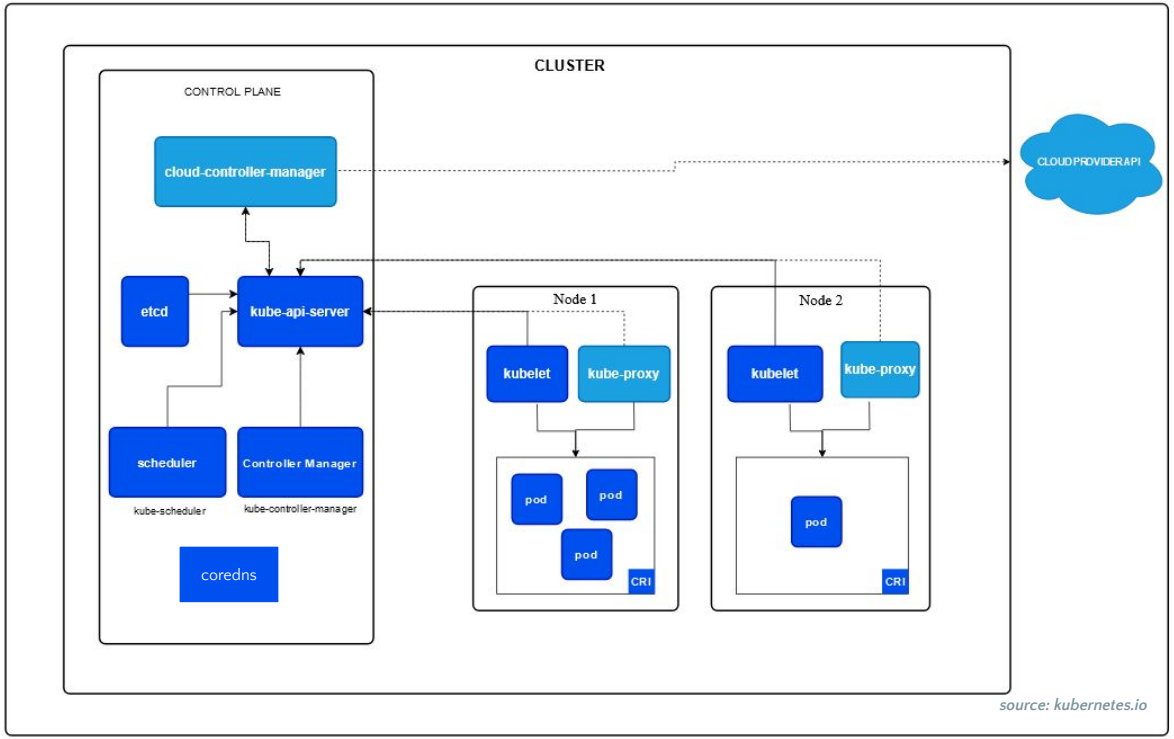
- ***Control plane***: Noeud portant la couche d'orchestration des conteneurs qui expose l'API et les interfaces pour définir, déployer et gérer le cycle de vie des conteneurs.
  - ***API server***: point d'entrée pour les commandes et opérations sur le cluster. Il expose l'API Kubernetes
  - ***Scheduler***: chargé de planifier les pods sur les noeuds disponibles en fonctions des exigences et des contraintes
  - **Etcd**: Magasin de données clé-valeur cohérent et distribué qui stocke les données du cluster
  - ***Controller Manager***: Surveille l'état du cluster et prend des mesures pour garantir l'état demandé



Les principes fondamentaux

L'architecture d'un cluster

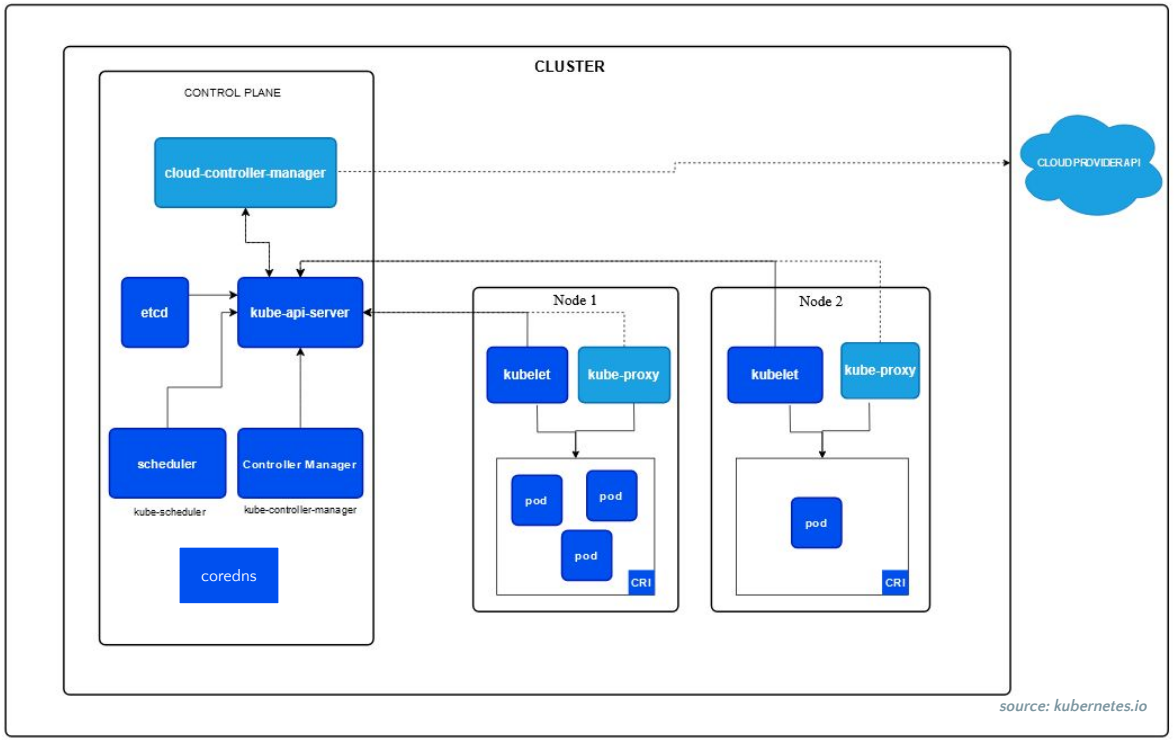
- ***Control plane***:
  - ***Cloud controller manager***: C'est un composant d'extension, distinct du controller manager. Son but est de gérer les ressources cloud spécifiques à un fournisseur de services cloud. Comme par exemple les volumes de stockage type S3 sur AWS
  - **Coredns**: serveur DNS de k8s. Il fournit un nommage et une fonctionnalité de découverte des services dans le cluster. Si vous vous connectez dans un conteneur, vous verrez que l'ip du serveur dns a été injectée dans `/etc/resolv.conf`



Les principes fondamentaux

L'architecture d'un cluster

- ***Node***: Noeud travailleur pouvant être une VM ou une machine physique. Il dispose des démons locaux ou services nécessaires à l'exécution des Pods.
    - ***Kubelet***: Communique avec le *control plane* et gère les conteneurs sur le noeud pour s'assurer qu'ils sont en cours d'exécution
    - ***Kube proxy***: Gère la configuration réseau et la connectivité réseau entre les pods
    - ***Container runtime***: Logiciel responsable de l'exécution des conteneurs implémentant l'interface CRI.



Les principes fondamentaux

Container Runtime Interface (CRI)

- Comment est-il possible d'exécuter des conteneurs Docker (ou autre) sous Kubernetes ?
- Pour rappel, les conteneurs Docker sont conformes à la spécification "OCI Runtime Specification", qui garantie la standardisation du format d'exécution du conteneur et donc l'interchangeabilité des *runtimes*. Podman qui est un outil open source conforme à la spécification OCI peut être un *runtime* de substitution à celui de Docker pour exécuter des conteneurs à partir d'images construites avec Docker.
- La *Container Runtime Interface* est une interface standardisée de plugin qui permet au kubelet d'utiliser différents *runtimes* de conteneurs, sans avoir besoin de recompiler les composants du cluster. Il est nécessaire de disposer d'un exécuteur de conteneur par noeud du cluster afin que le kubelet puisse lancer les pods et leurs conteneurs.
- CRI-O, un *runtime* de conteneurs pour k8s, implémente CRI. Il est open source et conforme aux spécifications de l'OCI. Cela permet à Kubernetes de gérer des conteneurs conformes à OCI.

Les principes fondamentaux

Les espaces de noms (*namespaces*)

- **Les espaces de noms sont un concept qui permet d'organiser des objets dans le cluster. Cela permet de diviser un cluster Kubernetes en plusieurs espaces logiques et isolés.**
- Générale on isole les ressources de manière à les rendre inaccessible entre plusieurs espaces de noms.
- Les espaces de noms facilitent également la gestion des ressources au sein du cluster. Ils permettent de regrouper les ressources liées à une application ou à un service particulier dans un espace de noms spécifique. Cela simplifie la gestion, la surveillance, et la maintenance, car les administrateurs et les développeurs peuvent se concentrer sur un espace de noms à la fois, sans avoir à s'inquiéter des autres parties du cluster.
- Les espaces de noms sont particulièrement utiles dans les environnements multi-utilisateurs, où plusieurs équipes ou utilisateurs partagent un même cluster Kubernetes.
- Attention à ne pas confondre les espaces de noms Linux et Kubernetes.

Les principes fondamentaux

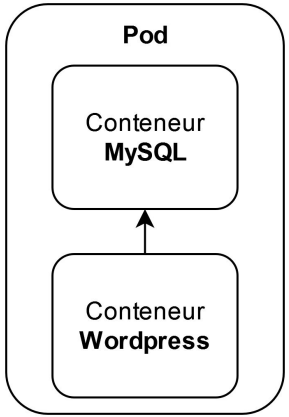
Les pods

- **Un pod est un ensemble atomique de conteneurs et volumes fonctionnant dans le même environnement d'exécution.**
- **Il s'agit de la plus petite unité déployable dans Kubernetes.** Les conteneurs d'un même pod partagent plusieurs espaces de noms Linux. Concrètement, ils partagent le même espace réseau, le même espace de stockage et les mêmes spécifications de ressources (RAM/CPU). Toutefois, ils s'exécutent dans leur propre cgroup.
- Il n'est pas possible de diviser un pod en parties plus petites et tous les conteneurs d'un pod sont situés sur le même noeud.

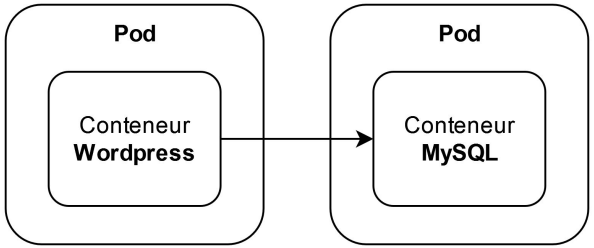
Les principes fondamentaux

Que mettre dans un pod ?

- Admettons que vous ayez deux conteneurs à déployer. Un pour site Wordpress et un second pour une base de données MySQL. Le conteneur Wordpress doit stocker les informations de son site dans le conteneur MySQL.
- Quelle solution choisiriez-vous entre les deux propositions suivantes ?



Solution A

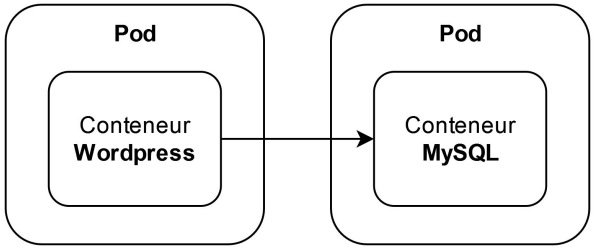


Solution B

Les principes fondamentaux

Que mettre dans un pod ?

- Lors de la conception des Pods, il est essentiel de se poser la question suivante : est-ce que les conteneurs peuvent opérer (de manière efficace et fiable) s'ils sont exécutés sur des machines distinctes ?
- Si la réponse est non, alors l'utilisation d'un seul Pod pour regrouper ces conteneurs est la solution appropriée.
- Cependant, si la réponse est oui, il serait plus judicieux d'envisager l'utilisation de plusieurs Pods.
- Dans notre exemple:
  - La base MySQL n'est pas fortement couplée au site Wordpress. En réalité, nous pourrions avoir plusieurs services consommant notre base de données.
  - Les deux conteneurs communiquent par le biais d'une interface réseau, ce n'est pas un obstacle qu'ils se trouvent sur deux machines différentes.
  - Notre site Wordpress est une application sans état, en cas de pics de charge, il est plus facile de créer plusieurs réplicats d'un Pod Wordpress que d'augmenter les ressources CPU/RAM d'un unique pod Wordpress/MySQL (le dimensionnement d'une base de données est plus délicat).



Solution B

Les principes fondamentaux

Les services

- **Un service dans Kubernetes est une ressource qui permet de fournir une abstraction réseau stable et constante pour les Pods, quel que soit l'endroit où ils sont déployés dans un cluster**.

- **Un équilibrage de charge interne est assuré par le service. Cela permet à un client de s'adresser à un groupe de Pods en utilisant un point d'entrée unique (l'ip du service) au lieu de pointer individuellement vers chaque Pod (l'ip du pod).** Cela facilite grandement la gestion des applications, car les Pods peuvent être créés et détruits dynamiquement sans perturber la communication des clients avec l'application.

- Les Services peuvent être de plusieurs types :
    - **ClusterIP**: service accessible uniquement à l'intérieur du cluster. Il est généralement utilisé pour exposer les services back-end ou des bases de données
    - **NodePort**: service exposant un port sur tous les nœuds du cluster
    - **LoadBalancer**: service utilisant un équilibreur de charge externe
    - **ExternalName**: service redirigeant vers un nom de domaine externe

Les principes fondamentaux

L'ingress

- **L'ingress est une ressource Kubernetes qui définit les règles d'acheminement du trafic HTTP et HTTPS entrant vers des services. Pour qu'il fonctionne, un contrôleur d'Ingress doit être déployé dans le cluster** comme par exemple Nginx Ingress Controller. Un contrôleur agit comme un gestionnaire de trafic qui interprète les règles Ingress et dirigent le trafic vers les services en conséquence.

- **L'ingress permet un routage basé sur des critères tels que les noms de domaine, les chemins d'URL, ou d'autres en-têtes HTTP.** Cela signifie que vous pouvez configurer des règles pour diriger le trafic vers des services spécifiques en fonction de l'URL demandée.

- Par exemple, vous pouvez rediriger le trafic vers un service de site Web, un service d'API ou d'autres applications en fonction de l'URL ou du nom de domaine. Cette flexibilité permet de gérer efficacement l'accès à plusieurs services à partir d'une seule passerelle d'entrée.

Les principes fondamentaux

Les étiquettes (*labels*)

- **Les étiquettes sont des métadonnées clés-valeurs que vous pouvez attacher à des objets Kubernetes tels que des pods, des services… Elles permettent d'associer des informations spécifiques à une ressource, facilitant ainsi la catégorisation, l'organisation et la recherche des ressources au sein d'un cluster.** C'est un élément omniprésent dans Kubernetes.

- Ces étiquettes permettent aux utilisateurs de faire correspondre leurs propres structures organisationnelles aux objets du système de manière souple, sans que les clients n'aient à stocker ces correspondances.

- **Dans un cluster k8s, il n'y a pas de notion de hiérarchie entre les composants.** Mais les objets doivent être reliés entre eux, et ces relations sont déterminés par les étiquettes et sélecteurs d'étiquettes.

- Les sélecteurs d'étiquettes permettent à un objet de faire référence à un ensemble d'autres objets Kubernetes.

- Plusieurs contraintes de nommage s'applique sur les étiquettes, vous les retrouverez dans la documentation:
  - https://kubernetes.io/docs/concepts/overview/working-with-objects/labels/

Les principes fondamentaux

Les annotations

- **Les annotations Kubernetes sont utilisées pour attacher des métadonnées arbitraires et non identifiantes aux objets.** Les clients tels que les outils et les bibliothèques peuvent récupérer ces métadonnées.
- **A la différence des étiquettes, il n'est pas possible de sélectionner des objets Kubernetes à l'aide d'une annotation.**
- Si vous avez un doute entre annotation ou étiquette, il est conseillé de créer une annotation. Si vous avez besoin de requêter votre objet à partir de cette méta-donnée, il suffira de modifier son type en étiquette.
- Quelques exemples d'utilisation:
    - Suivre l'état d'un déploiement (il s'agit d'une ressource Kubernetes que nous détaillerons plus tard)
    - Indiquer une information sur le build, la *release*, ou l'image
- Les mêmes contraintes de nommage que les étiquettes s'appliquent sur une annotation.

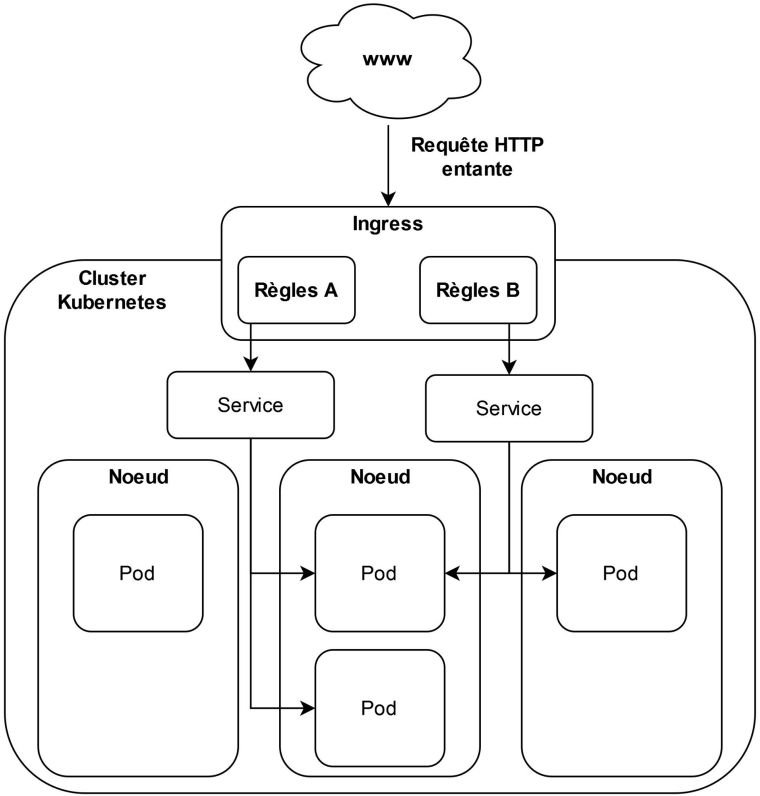
Les principes fondamentaux

Synthèse de la relation entre l'ingress, les services et les pods

- **Pods** : Les pods sont les unités de base de déploiement dans Kubernetes. Ils contiennent une ou plusieurs applications et sont les plus petits composants déployables. Les pods sont souvent exposés via des services pour permettre l'accès externe à leurs applications.

- **Services** : Les services Kubernetes fournissent une abstraction pour accéder aux pods. Ils définissent un point d'accès réseau stable pour un groupe de pods en utilisant des sélecteurs.

- **Ingress** : L'ingress est un contrôleur de gestion du trafic entrant qui permet de configurer des règles d'acheminement pour le trafic HTTP et HTTPS. Il agit comme une passerelle d'entrée pour plusieurs services, en fonction de critères tels que les noms de domaine ou les chemins d'URL. L'ingress dirige le trafic vers les services appropriés, qui, à leur tour, redirigent le trafic vers les pods correspondants. Ainsi, l'ingress permet de gérer de manière centralisée l'accès externe à plusieurs services Kubernetes.



Les principes fondamentaux

Les configMaps

- **Le configMap est une ressource Kubernetes qui permet de stocker des données de configuration sous forme de paires clé-valeur.**

- **Il est utilisé pour stocker des données de configuration qui sont nécessaires aux applications déployées dans un cluster Kubernetes.** Cela permet de séparer les données de configuration de l'application elle-même, ce qui rend l'application plus portable et plus facile à gérer.

- Les configMaps sont souvent utilisés pour stocker des paramètres de configuration tels que les variables d'environnement, les fichiers de configuration, les URL, les adresses IP, etc.

- Une fois qu'un configMap est créé, il peut être référencé dans les spécifications des pods Kubernetes. Les pods peuvent utiliser ces références pour injecter les données de configuration stockées dans le configMap directement dans leurs conteneurs. Cela permet aux applications de récupérer dynamiquement les données de configuration dont elles ont besoin à partir du configMap.

Les principes fondamentaux

Les secrets

- **Le secret est une ressource Kubernetes qui permet de stocker des données sensibles sous forme de paires clé-valeur.** Il est relativement similaire au configMap qui concerne, lui, les données de configuration.
- **Si vous devez stocker une clé d'authentification, un mot de passe, un certificat ou une autre information confidentielle, vous devez passer par un secret et non par un configMap !**
- **Les données stockées dans un secret sont encodées et stockées de manière chiffrée, ce qui garantit leur confidentialité.** L'utilisation de secrets permet de s'assurer que les informations sensibles ne sont pas exposées de manière non sécurisée dans les fichiers de configuration ou dans les conteneurs eux-mêmes.
- Les secrets peuvent être référencés dans les spécifications des pods Kubernetes, ce qui permet aux applications d'accéder de manière sécurisée aux données sensibles.
- Par exemple, un pod peut référencer un secret pour obtenir un mot de passe de base de données.

Déploiement d'applications

Déploiement d'applications

Le manifeste de Pod

- **Un manifeste de Pod décrit la configuration d'un seul pod.**
- Le format du fichier peut-être le JSON ou YAML.
- Il est utilisé pour définir les caractéristiques spécifiques d'un pod, telles que :
  - les conteneurs
  - les volumes
  - les variables d'environnement…
- Les pods créés à partir de manifestes individuels n'offrent pas de mécanisme de mise à l'échelle automatique ni de gestion avancée des mises à jour.
- Vous utiliserez les commandes `kubectl` pour charger ce manifeste sur Kubernetes.

```
apiVersion: v1
kind: Pod
metadata:
  name: mon-pod
spec:
  containers:
    - name: conteneur-1
    image: mon-image:latest
    ports:
    - containerPort: 80
  restartPolicy: Always
```

Déploiement d'applications

Le déploiement (*Deployment*)

- Un déploiement permet de gérer la création, la mise à l'échelle et la mise à jour de Pods.
- Il utilise un modèle de Pod comme base, mais il encapsule la gestion des répliques, de la mise à jour et du contrôle de l'état.
- Ils sont conçus pour gérer des applications dont les Pods doivent être répliqués pour une haute disponibilité.
- Ils permettent également de réaliser des mises à jour et des rollbacks des applications de manière contrôlée.

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: mon-deployment
spec:
  replicas: 3
  selector:
    matchLabels:
      app: mon-app
  template:
    metadata:
      labels:
        app: mon-app
    spec:
      containers:
        - name: conteneur-1
          image: mon-image:latest
          ports:
            - containerPort: 80
```

Déploiement d'applications

Exposition dans le cluster d'un déploiement

- Vous trouverez ci-contre un manifeste de service nommé "mon-service", qui expose à l'intérieur du cluster les pods du déploiement précédent.
- Notez bien que nous référençons le déploiement par l' étiquette "app" ayant la valeur "mon-app".
- Dans cet exemple, le service écoute sur le port 80 et redirige le trafic vers le port 80 du pod.

```
apiVersion: v1
kind: Service
metadata:
  name: mon-service
spec:
  selector:
    app: mon-app
  ports:
    - protocol: TCP
      port: 80
      targetPort: 80
  type: ClusterIP
```

Déploiement d'applications

Exposition d'une application en dehors du cluster

- Vous trouverez ci-contre un manifeste d'ingress nommé "mon-ingress".
- Elle expose notre application en dehors du cluster au travers d'une règle pour le domaine "`mon-app.example.com`".
- Toutes les requêtes HTTP reçues sur ce domaine sont acheminées sur le port 80 de notre service, et donc transmis à un des pods exécutant notre application dans un conteneur.
- Notez qu'il est nécessaire d'avoir déployé en amont un ingress controller dans le cluster pour que cette règle soit applicable.

```yaml
apiVersion: networking.k8s.io/v1
kind: Ingress
metadata:
  name: mon-ingress
spec:
  rules:
    - host: mon-app.example.com
      http:
        paths:
          - path: /
            pathType: Prefix
            backend:
              service:
                name: mon-service
                port:
                  number: 80
```

Gestion de la scalabilité

Gestion de la scalabilité

La gestion des ressources CPU et mémoire

- TODO

Gestion de la scalabilité

Les principales sondes

- TODO

Gestion de la scalabilité

Les statefulsets

- TODO

Gestion de la scalabilité

Le concept de Horizontal Pod Autoscaling (HPA)

- TODO

Gestion de la scalabilité

Scalabilité verticale et évolutivité des noeuds

- TODO

Gestion du stockage

Gestion du stockage

TODO

- TODO

Gestion de la sécurité et des autorisations

Gestion de la sécurité et des autorisations

TODO

- TODO

Observabilité et traçabilité

Observabilité et traçabilité

TODO

- TODO

Bibliographie

BURNS Brendan, BEDA Joe, HIGHTOWER Kelsey, EVENSON Lachlan. Traduit de l'anglais par MANIEZ Dominique. Kubernetes Maîtriser l'orchestrateur des infrastructures du futur. 3e édition. DUNOD, 2023. 308p.

MERCIER Pierre-Olivier. *Conteneurs et technologies du DevOps.* Gentilly: Editions ALPO, 2022. 209 p.

CLOUX Pierre-Yves, GARLOT Thomas, KOHLER Johann. *Docker et conteneurs: architectures, développement, usages et outils.* 3e édition. Malakoff: Dunod, 2022. 317 p.

LANDURE Julien. Pizza As A Service : les différents modèles du Cloud [en ligne]. [Consulé le 02/11/2023]. Disponible à l'adresse: https://medium.zenika.com/pizza-as-a-service-les-differents-modeles-du-cloud-b18ffaa6906c

Merci de votre attention

Maxime Lambert

maxime.lambert@unicaen.fr