Le GUIDE® GORILLA pour ...

Kubernetes multi-cloud avec Rancher

Comment choisir le bon outil pour gérer un environnement Kubernetes multi-cloud

DAN SULLIVAN

SUSE

OFFERT PAR ActualTech MEDIA

LE GUIDE GORILLA POUR...®

Gérer un environnement Kubernetes multi-cloud avec Rancher

Par Dan Sullivan

ACTUALTECH MEDIA

6650 Rivers Ave Ste 105 #22489
North Charleston, SC 29406-4829
www.actualtechmedia.com

REMERCIEMENTS DE L'ÉDITEUR

DIRECTEUR ÉDITORIAL
Keith Ward

DIRECTRICE DIFFUSION DE CONTENU
Wendy Hernandez

DIRECTRICE ARTISTIQUE
Olivia Thomson

DIRECTRICE PRINCIPALE CONTENU
Katie Mohr

PARTENAIRES ET VP CONTENU
James Green

CONTRIBUTIONS SPÉCIALES DE SUSE
Tom Callway – Senior Director, Product Marketing
Wendie Cheung – Senior Product Marketing Manager
Victor Estival – Head of Technical Marketing
Andrés Valero – Technical Marketing Manager

À PROPOS DE L'AUTEUR

Dan Sullivan est un ingénieur et architecte majeur, spécialisé dans l'architecture cloud, la science des données, l'apprentissage automatique et l'architecture de données. Il est l'auteur de six livres et de plusieurs cours en ligne, ainsi que des guides officiels de l'examen de certification de Google Cloud, édités par Sybex. Suivez Dan sur LinkedIn : linkedin.com/in/dansullivanpdx/.

SOMMAIRE

LÉGENDES UTILISÉES DANS CE GUIDE

L'ÉCOLE

Le gorille est naturellement pédagogue. Il aime aider les autres à apprendre. Ici, vous aborderez des thèmes parfois éloignés du sujet principal mais toutefois utiles.

BON À SAVOIR

Penchez-vous sur les sujets connexes présentés dans ce guide.

IDÉE BRILLANTE

Consultez d'excellentes idées.

EN PROFONDEUR

Approfondissez un sujet particulier.

BUREAU EXÉCUTIF

Découvrez des éléments d'intérêt stratégique pour les dirigeants d'entreprise.

ICÔNES UTILISÉES DANS CE LIVRE

DÉFINITION

Définit un mot, une expression ou un concept.

CONTRÔLE DES CONNAISSANCES

Teste les connaissances acquises.

ATTENTION

Attire votre attention sur un élément important.

GPS

Vous aide à acheminer vos connaissances au bon endroit.

AVERTISSEMENT !

À lire pour ne pas commettre d'erreur grave !

CONSEIL

Un conseil utile fondé sur ce que vous avez lu.

INTRODUCTION

Le cloud computing a radicalement changé la manière dont les entreprises fournissent leurs services digitaux, une évolution qui se poursuit avec l'adoption généralisée de Kubernetes. Offrant en tout lieu une plateforme informatique commune, Kubernetes permet aux développeurs de créer des services une seule fois et de les gérer. Pour beaucoup, l'association d'une infrastructure sur site et d'un ou plusieurs clouds publics constitue une solution optimale, mais utiliser ces ressources de la manière la plus efficace peut être un véritable défi.

Kubernetes, une plateforme d'orchestration de conteneurs, a considérablement modifié notre manière de gérer des services dans le cloud et sur site. Pour tirer le meilleur parti de Kubernetes, les entreprises se tournent vers les déploiements multi-cloud, ce qui peut occasionner d'autres défis en termes de gestion, comme la gestion de la configuration, la sécurité multi-cloud ou encore l'observabilité. Un grand nombre d'entre eux peuvent être relevés avec des systèmes comme Rancher, une plateforme de gestion multi-cluster permettant d'assurer la cohérence de l'ensemble des clusters, d'améliorer la gestion des charges de travail et de garantir une sécurité irréprochable, quelle que soit leur localisation.

Ce guide Gorilla® pour… gérer un environnement Kubernetes multi-cloud avec Rancher constitue votre feuille de route pour déployer efficacement Kubernetes au sein de votre entreprise et de son infrastructure variée. Nous commencerons par une présentation de Kubernetes, des conteneurs et du cloud computing, puis nous passerons en revue les avantages des clouds hybrides et des multi-clouds.

Les principaux fournisseurs de cloud prennent désormais en charge et ont adopté Kubernetes : nous examinerons les fonctionnalités

et avantages de leurs services Kubernetes gérés. La création d'une plateforme Kubernetes multi-cloud commence par une stratégie, c'est pourquoi nous aborderons les facteurs à prendre en compte pour élaborer une stratégie Kubernetes. Nous évoquerons ensuite comment choisir la solution Kubernetes adéquate et montrerons en quoi Rancher répond aux besoins des opérations Kubernetes multi-cloud. Nous conclurons par quelques conseils de ressources supplémentaires pour vous aider à commencer à travailler avec Kubernetes et Rancher.

En 2020, SUSE a fait l'acquisition de Rancher Labs, l'entreprise à l'origine de la plateforme de gestion Kubernetes open source la plus populaire, Rancher. Aujourd'hui, des services experts et d'assistance réparation jusqu'à 24 h/24 et 7 jours/7 sont proposés via un abonnement annuel appelé SUSE Rancher. Pour plus d'informations sur SUSE Rancher, rendez-vous sur https://www.suse.com/products/suse-rancher/.

CHAPITRE 1

Présentation de Kubernetes, des conteneurs et du cloud

DANS CE CHAPITRE :

- L'évolution de la virtualisation des ressources informatiques et de stockage
- Kubernetes pour l'orchestration du stockage
- Le modèle déclaratif

La virtualisation est devenue une technologie majeure pour améliorer l'efficacité de l'infrastructure IT, tout particulièrement des ressources informatiques. En remplaçant le matériel sous-jacent, les machines virtuelles ont constitué la première phase de la virtualisation à grande échelle.

Les conteneurs ont été développés en tant que moyen encore plus efficace d'utiliser les ressources informatiques, mais la généralisation des conteneurs s'est accompagnée d'une complexité supplémentaire. Kubernetes est une plateforme d'orchestration de conteneurs pouvant être exécutée sur site, sur des clouds ou en Edge. Kubernetes a profondément modifié notre manière de travailler avec des conteneurs.

Évolution de la virtualisation des ressources informatiques et de stockage

Les machines virtuelles sont des abstractions logiques de serveurs physiques. Elles sont exécutées sur des machines physiques qui

utilisent un hyperviseur afin de contrôler le serveur et effectuer des opérations en leur nom. Cela permet à un serveur physique unique d'exécuter plusieurs systèmes d'exploitation isolés les uns des autres.

Cette approche offre un avantage majeur : les applications ayant des exigences particulières en termes de système d'exploitation peuvent quand même utiliser le même serveur, car chaque application peut être exécutée sur sa propre machine virtuelle avec son propre système d'exploitation.

Avant les machines virtuelles, les administrateurs système exécutaient généralement une seule application sur un serveur afin d'éviter les éventuels conflits entre plusieurs applications exécutées dans le même système d'exploitation. Désormais, avec la virtualisation, chaque application est isolée et peut être gérée indépendamment des autres applications.

Cela permet une utilisation plus efficace des serveurs, puisque le matériel qui auparavant exécutait une seule application peut désormais en exécuter plusieurs. Lorsque des applications uniques étaient exécutées sur des serveurs, il n'était pas rare de constater une utilisation relativement faible du processeur sur les serveurs, tout particulièrement quand leur taille était adaptée à la capacité maximale.

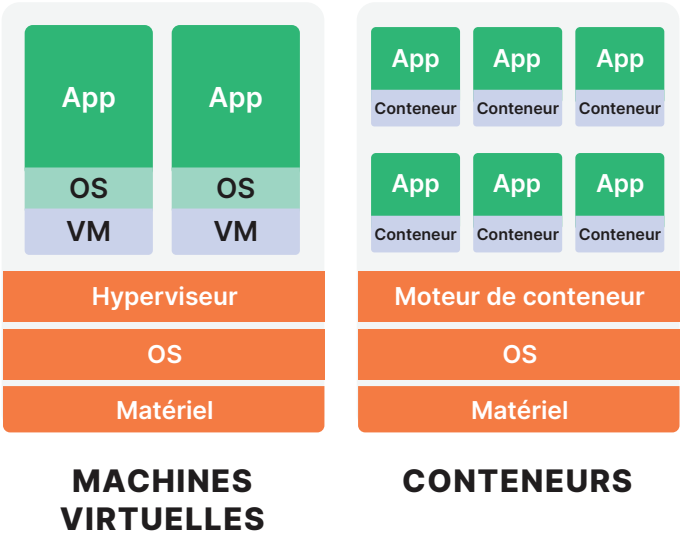


Figure 1 : Architecture de machines virtuelles et architecture de conteneurs

Certains problèmes peuvent apparaître lors de l'exécution d'applications sur des machines virtuelles. Si un système d'exploitation ou une application consomme trop de ressources, les autres systèmes d'exploitation et applications peuvent être impactés négativement et afficher des performances moindres que celles attendues. Ce phénomène est connu sous le nom de « voisin bruyant ». De plus, l'exécution de plusieurs systèmes d'exploitation consomme davantage de ressources que l'exécution d'un seul système d'exploitation. L'idéal serait de pouvoir exécuter des systèmes d'exploitation isolés tout en bénéficiant de l'utilisation des ressources d'un seul système d'exploitation. C'est là que les conteneurs entrent en jeu.

Les conteneurs permettent également d'exécuter plusieurs applications avec un haut niveau d'isolation sur le même serveur (voir la **Figure 1**). Il est possible d'exécuter plusieurs conteneurs sur un seul système d'exploitation tout en conservant des niveaux d'isolation satisfaisants.

Les conteneurs utilisent les fonctionnalités de Linux ou d'autres systèmes d'exploitation modernes pour isoler les ressources. Control Groups, couramment appelé *cgroups*, est une fonctionnalité du noyau Linux qui limite les ressources disponibles pour les processus. Cgroups peut limiter le volume de mémoire utilisée par un processus, donner la priorité aux processeurs et disques et offrir d'autres fonctionnalités afin de garantir l'isolation.

Parallèlement à ce contrôle de l'accès aux ressources, *seccomp* est une fonctionnalité de Linux qui limite les appels système que peuvent effectuer les processus. Restreindre les conteneurs à certains types d'appels permet d'améliorer la sécurité du système d'exploitation en limitant les éventuels dommages occasionnés par un conteneur malveillant ou défaillant.

Les conteneurs nécessitent un outil d'exécution. Auparavant, Docker constituait la référence pour les conteneurs, mais ce n'est plus le cas aujourd'hui. Sur Linux et Windows, *containerd* est un outil d'exécution de conteneurs de la Cloud Native Computing Foundation (CNCF) largement répandu. Cet outil gère le cycle de vie d'un conteneur exécuté sur le serveur hôte. CRI-O est un outil d'exécution de

conteneurs léger utilisé sur Minikube, une solution de mise en œuvre locale de Kubernetes souvent utilisée pour le développement de Kubernetes et la familiarisation avec la plateforme.

Outre ces outils d'exécution de conteneurs, il est possible d'utiliser un noyau applicatif, ou *sandbox*, tel que gVisor. gVisor utilise des fonctionnalités utilisateur pour émuler les appels système, qui sont ensuite exécutés par gVisor au nom de l'utilisateur. L'un des avantages de gVisor est que, puisque chaque conteneur possède son propre noyau applicatif, l'hôte est moins exposé aux éventuelles attaques d'un code malveillant exécuté dans un conteneur.

Les conteneurs permettent un bin packing et une utilisation efficaces des ressources mais, à mesure que leur nombre croît, ils sont plus difficiles à gérer. Ce qu'il faut, c'est un système capable d'orchestrer l'utilisation et la gestion de conteneurs. C'est ce besoin qui est à l'origine du développement de Kubernetes, la solution de référence du secteur.

Kubernetes pour l'orchestration de conteneurs

Kubernetes est un projet open source créé par Google pour gérer efficacement les conteneurs déployés sur un cluster de serveurs. Kubernetes est basé sur le système Borg, dont le développement a commencé au début des années 2000 et qui sert à exécuter un grand nombre de services de Google, notamment le moteur de recherche et Gmail. Sorti en 2014 en open source, Kubernetes est désormais un projet de la CNCF.

Kubernetes est souvent abrégé en « K8s », le chiffre 8 représentant les huit lettres présentes entre le « K » et le « s » de Kubernetes.

Kubernetes utilise un certain nombre d'abstractions dans sa conception. Il s'agit d'une de ses caractéristiques clés, puisque le système n'est

pas étroitement lié à une infrastructure spécifique. Par exemple, Kubernetes exécute les charges de travail informatiques sur des nœuds. Ces nœuds peuvent être mis en œuvre sur des machines bare metal ou en tant que machines virtuelles. Un cluster peut utiliser des ressources informatiques dans différentes zones ou régions, mais peut tout de même gérer les nœuds en tant que membres du même cluster.

Les nœuds possèdent différentes caractéristiques. Certains d'entre eux sont par exemple des instances ponctuelles sur un cloud public. Ces nœuds sont parfaitement adaptés pour les tâches capables de récupérer après l'arrêt d'un nœud. Les applications informatiques haute performance sont souvent conçues pour récupérer après ce type de défaillance.

Entraîner des modèles d'apprentissage automatique est un processus qui demande beaucoup de ressources informatiques et peut être facilité par l'utilisation de processeurs graphiques. Il est possible de déployer des nœuds dans un cluster avec des processeurs graphiques, mais il est préférable de s'assurer que les ingénieurs en apprentissage automatique disposent d'un accès privilégié à ces nœuds.

Il est possible de contrôler l'accès à différentes ressources, notamment grâce aux *espaces de noms*. Les espaces de noms permettent d'isoler les ressources dans un cluster, afin que les ressources puissent accéder uniquement aux ressources présentes dans le même espace de noms, sauf si une politique permet explicitement un tel accès.

Dans K8s, les conteneurs sont exécutés dans des pods. Un ou plusieurs conteneurs peuvent être exécutés dans un même pod. Le pod est la petite unité informatique de K8s et constitue une autre abstraction permettant certaines fonctionnalités importantes de K8s.

Par exemple, il est souvent nécessaire de déployer un conteneur avec un autre conteneur. Cela peut être le cas lorsque l'un des conteneurs exécute un code fournissant un proxy pour des services dans l'autre conteneur. Les deux conteneurs doivent toujours être déployés en même temps, arrêtés en même temps et, dans de nombreux cas, être traités en tant que paire étroitement liée. Le déploiement de conteneurs dans des pods permet à Kubernetes de mettre en œuvre

des fonctionnalités qui seraient, au mieux, difficiles à mettre en œuvre s'il n'était pas possible de regrouper des conteneurs étroitement liés.

Ces conteneurs étroitement liés sont appelés des *sidecars*. Les sidecars offrent des fonctionnalités couramment utilisées, comme des règles d'authentification, de suivi et de réseaux. L'un des avantages de l'utilisation des sidecars est qu'ils évitent aux développeurs de services d'avoir à mettre en œuvre des fonctions d'authentification, de surveillance et de collecte de journaux pour chaque nouveau service. Cela permet également de garantir plus facilement la cohérence de l'authentification, la surveillance et la collecte de journaux pour chaque service.

Le modèle déclaratif

Avec Kubernetes, il faut décrire la manière dont nous voulons qu'une application soit déployée, sans avoir à définir explicitement les premières étapes à suivre pour mettre en œuvre ce déploiement. On appelle cette approche le modèle déclaratif. Imaginons par exemple qu'une application a été déployée avec un déploiement et 10 répliques. Les répliques sont réparties sur différents nœuds du cluster. Le cluster rencontre ensuite un problème et l'un des nœuds sur lesquels les pods sont exécutés connaît une défaillance quelconque. Nous avions déclaré avoir 10 répliques, qui ont désormais disparu. Le planificateur de K8s va alors réévaluer les ressources dans le cluster et affecter de nouveaux pods sur les nœuds disponibles jusqu'à atteindre les 10 répliques.

Dans Kubernetes, le modèle déclaratif permet que ce type de réponse soit automatisable et c'est l'une des principales raisons pour lesquelles Kubernetes est facile à gérer en tant que code avec des outils comme Fleet. Kubernetes peut vérifier l'état des pods et détecter quand ils sont défaillants ou ne fonctionnent pas et les remplacer. Ces vérifications ainsi que la réparation automatique sont des fonctionnalités clés qui permettent une gestion plus automatisée des conteneurs à grande échelle.

Évolutivité des charges de travail

Lorsque l'on parle d'évolutivité dans Kubernetes, il convient de distinguer celle de l'infrastructure et celle des ressources des charges de travail. En ce qui concerne l'infrastructure, il est possible d'ajouter ou de supprimer des nœuds d'un cluster afin de fournir plus ou moins de ressources informatiques. Il s'agit d'une mise à l'échelle horizontale de l'infrastructure.

Il est également possible d'adapter les ressources prévues pour les charges de travail, comme le nombre de pods exécutant un service. En cas de pic de demande pour un service, il peut être nécessaire d'ajouter des pods pour gérer la charge. Il s'agit d'une mise à l'échelle horizontale de la charge de travail.

L'un des autres avantages majeurs de Kubernetes, c'est qu'il peut s'exécuter partout : sur site, sur des clouds publics et en Edge. Kubernetes possède également un riche écosystème de projets connexes parrainés par la CNCF. Ces projets open source fournissent des outils et projets supplémentaires pour soutenir l'utilisation et le suivi de Kubernetes, ainsi que des processus de développement.

Kubernetes sur les clouds publics

Kubernetes est conçu pour exploiter l'infrastructure informatique et de stockage. Les clouds publics ont notamment pour but de fournir des services informatiques et de stockage à la demande. L'utilisation conjointe de ces deux éléments présente donc un avantage évident. Exécuter Kubernetes sur un cloud public permet d'accéder à une infrastructure flexible dont la taille peut s'adapter rapidement.

Le modèle de coût d'un cloud standard se base sur un paiement en fonction de l'utilisation. L'utilisation de clouds publics permet aux entreprises d'éviter ou de réduire les dépenses de capital qui seraient autrement nécessaires pour acquérir l'infrastructure nécessaire pour déployer Kubernetes.

Outre une infrastructure informatique, de stockage et de réseau, les clouds publics offrent des services gérés qui peuvent réduire les frais d'exploitation liés à la gestion de services IT. Par exemple, les fournisseurs de cloud possèdent des systèmes de gestion des identités et des accès qui peuvent être synchronisés avec votre fournisseur d'identité, ce qui permet de rationnaliser le processus de gestion des authentifications et autorisations.

D'autres services gérés portant sur la gestion du cycle de vie des données, la collecte de journaux et les scans de sécurité dans un pipeline CI/CD permettent de bénéficier de l'utilisation de Kubernetes sans surcharger vos équipes opérationnelles.

Les outils d'infrastructure as code, comme Terraform, permettent de préciser de manière déclarative l'infrastructure que l'on souhaite déployer sur un cloud. Préciser l'infrastructure de manière déclarative permet de réduire les erreurs de configuration de l'infrastructure et contribue au déploiement cohérent de l'infrastructure dans différents environnements.

Pour les entreprises qui craignent l'enfermement propriétaire et les clouds publics, Kubernetes a la solution. Kubernetes fonctionne partout et offre une plateforme commune qui s'exécute dans différents environnements. Une plateforme informatique commune permet aux organisations d'éviter l'enfermement propriétaire sur le cloud grâce à la portabilité élevée des charges de travail d'un environnement Kubernetes à l'autre. Cela permet aux utilisateurs de Kubernetes d'exécuter leurs charges de travail là où c'est le plus efficace. Cette flexibilité entraîne cependant une gestion plus complexe des charges de travail dans différents clusters sur le cloud, sur site et en Edge.

CHAPITRE 2

Valeur d'un environnement IT hybride multi-cloud

DANS CE CHAPITRE :

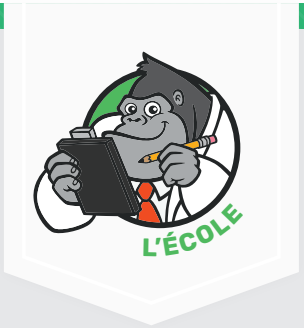
- Utilisation efficace des ressources cloud
- L'importance de l'automatisation
- Défis liés aux environnements IT hybrides et multi-cloud

Kubernetes offre l'opportunité de tirer parti des environnements IT hybrides et multi-cloud, sur site et sur les clouds publics. Des outils de gestion, tels que Rancher, permettent une expérience cohérente et homogène quelle que soit la localisation des clusters. Ces opportunités techniques se traduisent par des avantages clés pour l'entreprise, notamment la réduction des risques grâce à la généralisation de la technologie, la garantie d'une disponibilité optimale ainsi que la diversification des ressources informatiques et de stockage. Les avantages se divisent en trois catégories : portabilité des applications, cohérence des opérations et cohérence du développement et du déploiement. Bien entendu, certains défis sont également à prendre en compte.

Portabilité des applications

Puisque Kubernetes offre une plateforme commune pour exécuter des charges de travail sur des infrastructures mises à disposition par différents fournisseurs, il est logique de déplacer les charges de travail là où elles peuvent être exécutées de la manière la plus efficace et la plus rentable.

GitOps



La portabilité des applications est également assurée par des pratiques GitOps, notamment le fait de disposer d'un référentiel de code centralisé pouvant être déployé n'importe où avec Kubernetes. Cela permet de gérer et déployer des applications de manière flexible, peu importe le cloud.

L'un des facteurs clés pour déterminer si un déploiement est efficace est la localisation des données nécessaires à une charge de travail. Pour les charges de travail ne nécessitant pas une grande quantité de données, le choix de leur localisation dépendra d'autres facteurs, comme le coût et la latence.

Prenons par exemple les développements Web front-end. Généralement sans état, ils sont donc relativement faciles à adapter en termes de taille et certaines préoccupations sont inutiles, comme celle de veiller à ce que toutes les requêtes d'un client soient acheminées vers le même serveur back-end pendant la durée d'une session.

Les applications front-end collectent des données, mais il s'agit d'entrées de données à l'échelle humaine et non des ensembles de données générés par une machine. Si vous disposez d'une base d'utilisateurs mondiale, ou même régionale, la latence peut être un facteur. Selon les meilleures pratiques, il conviendrait alors de répartir le service à exécuter dans plusieurs régions en utilisant des équilibreurs de charge pour distribuer les requêtes vers la région la plus proche de l'utilisateur.

Utilisation efficace des ressources cloud

L'une des raisons de l'adoption généralisée des conteneurs est qu'ils sont plus efficaces que les machines virtuelles dans l'utilisation des ressources informatiques. Les machines virtuelles exigent un hyperviseur fournissant un niveau d'abstraction sur le matériel. L'hyperviseur permet l'exécution de plusieurs systèmes d'exploitation sur le même matériel, ce qui serait impossible si deux systèmes d'exploitation ou plus essayaient de contrôler le matériel.

Les conteneurs n'exigent pas d'hyperviseur, ils utilisent les fonctionnalités du système d'exploitation pour partager les ressources efficacement et en toute sécurité. Puisqu'il n'y a pas besoin d'hyperviseur et que l'on peut exécuter plusieurs conteneurs sur une instance unique de système d'exploitation, il est possible de déployer davantage d'applications sur des serveurs qu'avec des machines virtuelles.

Frais de sortie

Pour les opérations exigeant une grande quantité de données, comme l'analyse de données, l'entreposage de données ou l'apprentissage automatique, un autre facteur est à prendre en compte pour choisir où exécuter vos tâches. Vous devez en effet tenir compte des éventuels frais encourus pour la copie de données entre régions ou centres de données.

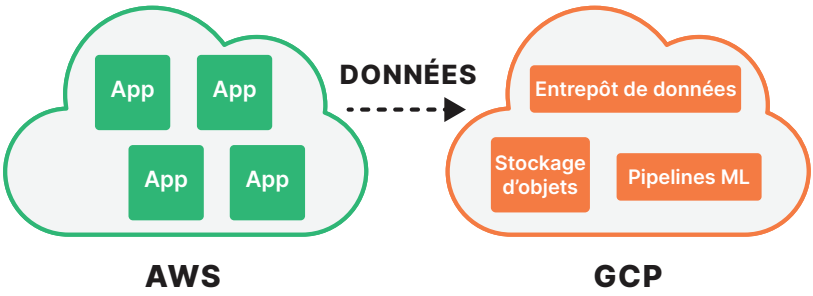


Figure 2 : Pensez à tenir compte des frais de sortie lorsque vous copiez de grands volumes de données dans un environnement multi-cloud ou multi-régional

Il est courant que les fournisseurs de cloud facturent les sorties de données à l'extérieur d'une région (voir la **Figure 2**). Ce n'est pas le cas si vous copiez des données au sein d'un cloud public unique, mais c'est autre chose lorsque vous les transférez vers différentes régions ou que vous les copiez d'un cloud à l'autre.

En raison de ces frais, le coût de transfert de données peut être supérieur aux économies réalisées par l'exécution de vos charges de travail dans un environnement informatique moins onéreux. Dans ce cas, il est plus judicieux de transférer les opérations informatiques et non les données. Heureusement, avec Kubernetes, ces déplacements des charges de travail sont une solution viable.

Cohérence des opérations

L'utilisation de Kubernetes offre également l'avantage de garantir la cohérence des opérations sur différentes plateformes d'infrastructure. Cela est particulièrement vrai si l'on utilise des outils open source qui font partie de l'écosystème Kubernetes.

Outre l'orchestration des conteneurs, que Kubernetes prend en charge, vous aurez généralement besoin d'aide concernant l'observabilité de vos charges de travail et opérations. Dans ce domaine, il existe plusieurs outils open source extrêmement efficaces, matures et bien supportés.

Votre boîte à outils

Prometheus est une application open source de surveillance et d'alerte. Prometheus permet de collecter un grand nombre de métriques en temps réel et de les stocker dans une base de données de séries temporelles.

Prometheus s'exécute en tant qu'ensemble de services, avec notamment des exportateurs qui s'exécutent sur les hôtes à surveiller et exportent les métriques vers un magasin de données ; un gestionnaire d'alerte qui suit les données de séries temporelles et donne l'alerte lorsque certaines conditions sont remplies ; et un langage de requête, PromQL, pour interroger les données de séries temporelles. Grafana est un autre outil open source souvent utilisé avec Prometheus pour créer des tableaux de bord.

Fluentd est une application open source qui collecte les journaux d'événements, les journaux d'applications et les données relatives aux flux de clics sur différentes plateformes. Fluentd permet de créer un service unifié de collecte de journaux pour toutes les applications et infrastructures. Kubernetes utilise les agents de collecte de journaux Fluentd pour collecter et analyser les journaux.

Le traçage distribué est l'analyse d'une série d'appels de service qui, collectivement, mettent en œuvre un workflow. Par exemple, un appel à une fonction API afin de calculer la taxe de vente pour un site Web marchand peut entraîner l'appel à d'autres API qui authentifient l'appelant d'origine, interrogent la base de données des autorités fiscales, classent les articles afin de déterminer s'ils sont soumis à des taxes et calculent la taxe totale sur les articles.

Les détails de mise en œuvre sont souvent ignorés par le service appelant l'API, mais lorsque le service est lent et que les transactions prennent trop de temps, il est impératif de déterminer quel service de la série d'appels pose problème. C'est là qu'interviennent les outils de traçage distribué. OpenTracing fournit une API ouverte et un outil d'instrumentation pour le traçage distribué.

Il existe également des outils open source pour la gestion de l'infrastructure et de la configuration pouvant être utilisés avec Kubernetes. Mentionné plus haut, Terraform est un outil d'infrastructure as code (IaC) qui permet de gérer l'infrastructure à l'aide d'un logiciel. Terraform peut appeler les API des fournisseurs de cloud pour créer des infrastructures en votre nom. Les opérations vont de la plus simple à la plus complexe, de la mise en place d'une unique machine virtuelle à la création d'un cluster Kubernetes, de buckets

Terraform est un outil qui gagne en importance dans la mise en œuvre de l'infrastructure as a service. Terraform est utilisé sur des infrastructures cloud et sur site et peut servir à définir une large gamme de ressources, notamment les services sans serveur.

de stockage d'objets ou de bases de données NoSQL pour l'analyse de données à grande échelle.

Terraform utilise une configuration d'infrastructure et élabore un plan d'exécution pour créer cette infrastructure. Tout comme les configurations Kubernetes, les configurations Terraform sont déclaratives et peuvent être utilisées pour créer une infrastructure de A à Z ou mettre à jour une infrastructure existante en appliquant les changements nécessaires pour qu'elle soit conforme au nouvel état souhaité.

L'importance de l'automatisation

Une fois l'infrastructure créée, le logiciel est configuré et déployé dans cette infrastructure. Les outils de gestion de la configuration sont conçus pour automatiser ce processus. Ansible, Puppet, Chef et Salt sont quatre outils courants de gestion de la configuration pouvant être utilisés avec Kubernetes. Fleet est un autre outil de gestion de la configuration spécialement conçu pour Kubernetes afin de contribuer à automatiser la configuration des clusters à grande échelle.

Un tiers de la gestion opérationnelle, en plus de la gestion de l'infrastructure et de la configuration, est consacrée à l'intégration continue et au déploiement continu, ou CI/CD. Les architectes et ingénieurs logiciels sont nombreux à avoir abandonné les architectures applicatives monolithiques au profit d'architectures de microservices et CI/CD, qui permettent un développement plus agile et une maintenance plus aisée. Kubernetes est parfaitement adapté aux architectures de microservices, qui font partie des principes de conception originaux du cloud.

Outre les architectures de microservices, les équipes agiles utilisent des pratiques communes de développement et d'exploitation pour développer des applications cloud natives. Ces applications sont souvent créées par des équipes de développeurs qui travaillent en collaboration et utilisent des outils de gestion du code source tels que GitHub et Bitbucket.

Les applications peuvent contenir de nombreux éléments, il est donc important d'automatiser le processus de création afin d'éviter les

erreurs dues aux processus manuels et réduire le temps passé par les ingénieurs logiciels sur d'autres tâches que celle d'écrire du code.

Des outils tels que Jenkins et Spinnaker permettent à des pipelines automatisés de créer, tester et déployer le code de l'application jusqu'à la production. Certains des éléments utilisés dans des applications sont réutilisés par d'autres systèmes ou bibliothèques qui offrent des fonctionnalités supplémentaires. Ces éléments sont généralement stockés et gérés dans des référentiels d'artefacts comme Artifactory.

Les images des conteneurs sont souvent réutilisées et sont également stockés dans des référentiels centralisés. Les registres d'images peuvent être privés et réservés à un usage interne ou peuvent être publics, comme Docker Hub, qui propose des images accessibles au public. Comme indiqué précédemment, Prometheus et Grafana sont largement utilisés pour surveiller l'état des services applicatifs.

Défis liés aux environnements IT hybrides et multi-cloud

Lorsque des charges de travail sont déployées sur plusieurs clouds et environnements IT, des défaillances peuvent apparaître. Cela peut entraîner une perte et une sous-utilisation des ressources et, par conséquent, une augmentation des coûts. Dans le pire des cas, vous pourrez être confronté à une expansion incontrôlée de l'infrastructure et de la charge de travail, si vous n'utilisez pas d'outils pour gérer ce niveau de complexité supplémentaire. Vous pourriez donc être tenté d'éviter les approches hybrides et multi-cloud, mais ce n'est pas une bonne stratégie. Vous perdriez les avantages que nous venons d'aborder, tout particulièrement celui d'éviter l'enfermement propriétaire.

Certaines des difficultés liées aux environnements hybrides et multi-cloud peuvent être réduites par l'utilisation d'un service Kubernetes géré, mais cela vous lierait étroitement à un fournisseur unique. Vous pouvez aussi remplacer le plan de contrôle de Kubernetes par un gestionnaire fédéré tel que Rancher, qui unifie la gestion des clusters afin de fournir une gestion cohérente des opérations et charges de travail, tout en évitant l'expansion de l'infrastructure et la perte de ressources.

CHAPITRE 3

Kubernetes avec des services hébergés par des fournisseurs de cloud : Amazon EKS, Azure AKS, Google GKE

DANS CE CHAPITRE :

- AWS Elastic Kubernetes Service (EKS)
- Azure Kubernetes Service (AKS)
- Google Kubernetes Engine (GKE)

En ce qui concerne l'exécution de Kubernetes, certains utilisateurs choisissent de déployer et gérer leurs propres clusters Kubernetes, tandis que d'autres choisissent un service Kubernetes hébergé par un fournisseur de cloud. Un service géré peut permettre de réduire la quantité des tâches opérationnelles nécessaires pour gérer un environnement Kubernetes — mais comme nous allons le voir, Rancher fournit un grand nombre des avantages d'un service géré en termes de gestion des clusters tout au long de leur cycle de vie, sans les contraintes d'un service géré par un fournisseur spécifique.

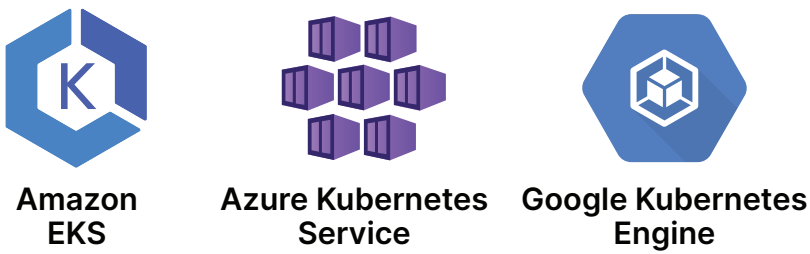


Figure 3 : Ensemble de services Kubernetes chez les principaux fournisseurs de cloud

Les trois principaux fournisseurs de cloud public — AWS, Azure et Google Cloud — proposent chacun des services Kubernetes prêts à l'emploi (voir la **Figure 3**).

AWS Elastic Kubernetes Service (EKS)

AWS offre la possibilité d'exécuter des clusters Kubernetes gérés sur EC2, l'offre infrastructure as a service de machines virtuelles proposée par AWS, ou sur AWS Fargate, un service informatique sans serveur pour les conteneurs. La principale différence entre ces deux services est que pour un déploiement EC2, vous payez pour ce que vous mettez sur EC2, tandis qu'avec AWS Fargate, vous payez pour le temps que vous passez à exécuter des services sur Fargate.

Le service AWS EKS comprend des contrôleurs AWS pour Kubernetes qui intègrent EKS à d'autres services AWS, notamment une console de gestion intégrée. AWS EKS possède d'autres fonctionnalités, telles que Cloud Map pour la découverte de services, un maillage de services et un équilibreur de charge.

EKS s'intègre au vaste écosystème AWS, ce qui simplifie l'intégration aux applications et services créés par AWS, et offre un accord de niveau de service de 99,95 %. EKS est le service Kubernetes géré sur cloud public le plus utilisé. Cependant, par rapport à GKE et AKS, le service EKS exige davantage d'intervention manuelle. Par exemple, en tant qu'administrateur K8s vous devrez installer Calico CNI manuellement puisque ce plugin n'est pas préinstallé sur EKS.

Azure Kubernetes Service (AKS)

Azure propose également un service Kubernetes géré, Azure Kubernetes Service, ou AKS. Ce service permet la fourniture élastique de ressources. Tout comme AWS EKS, ce service Kubernetes s'intègre à d'autres services Azure, tels que les outils de développement Visual

Studio, Azure DevOps, Azure Monitor, Azure Active Directory ou encore Azure Policy.

La plateforme AKS fonctionne bien lorsque seuls des clusters Azure doivent être gérés, mais elle ne fournit pas d'outils pour gérer les clusters K8s exécutés sur d'autres clouds. Pour cela, vous pouvez opter pour Rancher, qui offre une interface commune pour gérer les clusters Kubernetes sur différents clouds, sur site ou quelle que soit la localisation de K8s.

L'un des avantages d'AKS est qu'il s'agit du service géré le plus rapide à fournir un accès aux nouvelles versions de Kubernetes. AKS ne facture pas les nœuds de contrôle. Microsoft est connu pour fournir un environnement de développement bien conçu et bien supporté pour AKS et Azure en général. Les stratégies réseau Azure et Calico peuvent être installées automatiquement lorsqu'un cluster est créé. La mise à niveau de Kubernetes sur AKS est semi-automatique. Pour atteindre les 99,95 % de l'accord de niveau de service d'EKS, il est obligatoire d'utiliser les zones de disponibilité, ce qui entraîne des coûts supplémentaires.

Google Kubernetes Engine (GKE)

Il n'est pas surprenant que Google, créateur de Kubernetes, offre également un service géré pour cette plateforme. Google Kubernetes Engine (GKE) est mis en œuvre à l'aide de machines virtuelles de Compute Engine. L'utilisateur peut choisir parmi différents types de machines.

GKE est disponible en deux modes opérationnels : standard et pilotage automatique. Le mode standard offre à l'utilisateur un contrôle total sur la gestion des nœuds. Comme pour AWS EKS exécuté sur EC2, vous payez en fonction de ce que vous y mettez.

L'autre mode, pilotage automatique, offre un support opérationnel plus important, notamment pour déterminer la configuration de cluster optimale en fonction des charges de travail. Ce mode est facturé comme AWS Fargate : vous payez pour les services que vous utilisez.

SUSE Rancher Hosted

Parallèlement à ces services Kubernetes, les fournisseurs de cloud et de services élargissent leur offre à des services de gestion hébergés. Découvrez par exemple SUSE Rancher Hosted.

GKE s'intègre à d'autres services Google Cloud, comme la gestion des identités et des accès, les services de sécurité et les services d'intelligence artificielle et d'apprentissage automatique. Les services de sécurité incluent la prévention de la perte de données, la protection contre les attaques DDoS et les Confidential VMs, qui assurent le chiffrement en mémoire. Les services d'intelligence artificielle et d'apprentissage automatique incluent l'intégration à Vertex AI et l'accès à Cloud TPU (Tensor Processus Unit).

GKE dispose de différentes versions de Kubernetes pour créer vos nouveaux clusters. Tout comme AKS, GKE répare automatiquement les nœuds. Il existe une intégration étroite avec la console et les outils de ligne de commande GCP, ce qui assure une expérience agréable pour les utilisateurs de GCP existants. GKE propose des mises à niveau entièrement automatisées de Kubernetes et peut fournir un accord de niveau de service de 99,95 % si vous utilisez des clusters régionaux. Cela entraîne des frais supplémentaires. GKE ne propose pas d'offre de cloud pour le gouvernement fédéral américain, contrairement à EKS et AKS.

Rancher peut également importer des clusters GKE, afin que vous puissiez gérer vos clusters GKE, AKS et EKS sur une seule console de gestion.

Avant de choisir un plan de mise en œuvre Kubernetes, il est important de disposer d'une stratégie d'entreprise pour Kubernetes et le multi-cloud.

CHAPITRE 4

Élaborer une stratégie Kubernetes multi-cloud adaptée à l'entreprise

DANS CE CHAPITRE :

- Évaluer l'utilisation actuelle de Kubernetes
- Définir les caractéristiques de la stratégie spécifiques à l'entreprise
- Déterminer où Kubernetes sera exécuté

L'élaboration d'une stratégie Kubernetes multi-cloud solide exige une évaluation complète des besoins organisationnels, ressources et besoins des utilisateurs. Il convient tout d'abord de poser les bonnes questions aux bons endroits :

- Qui utilise cela ?
- Où les déploiements Kubernetes seront-ils le plus souvent utilisés ?
- Quels sont les besoins d'exploitation et de gestion pour nos cas d'utilisation ?
- Quels types d'applications seront exécutées dans des clusters et de quels outils et ressources auront-elles besoin ?

Dans cette section, nous aborderons la manière d'envisager et d'évaluer les exigences relatives à un déploiement Kubernetes multi-cloud, et fournirons un cadre qui vous aidera à prendre les meilleures décisions pour votre entreprise (voir la **Figure 4**.)

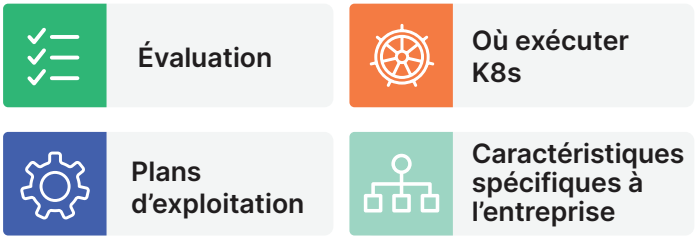


Figure 4 : Éléments d'une stratégie Kubernetes multi-cloud

Évaluation de l'utilisation actuelle de Kubernetes

La première étape de l'évaluation doit être l'analyse de l'utilisation actuelle de Kubernetes dans votre entreprise. Chaque entreprise a ses priorités concernant Kubernetes. Certaines utilisent Kubernetes avant tout pour le développement logiciel dans des environnements de développement ou pour l'analyse et l'entreposage de données, tandis que d'autres utilisent K8s dans des applications en contact direct avec les clients et créées sur une architecture de microservices.

Une stratégie de déploiement multi-cloud solide commence par définir les domaines où Kubernetes est déjà le plus sollicité. Cela permettra à la direction de prendre des décisions en tenant compte des besoins opérationnels existants.

Un autre aspect tout aussi important de l'utilisation actuelle de Kubernetes est la manière dont les déploiements sont utilisés. Une stratégie de déploiement complète doit déterminer si les charges de travail comprennent les applications Web, les pipelines de données par lot, les back-ends d'applications mobiles, etc. Différentes configurations multi-cloud pourront répondre à différentes exigences de charges de travail, il est donc important de comprendre parfaitement comment votre entreprise exécute ses charges de travail sur K8s avant de prendre des décisions définitives.

La dernière étape de l'analyse de l'utilisation actuelle de Kubernetes consiste à évaluer honnêtement la maturité de la gestion

opérationnelle au sein de votre entreprise. Pour le déterminer, une métrique simple et rapide peut aider : il s'agit de l'étendue de l'automatisation dans vos pipelines actuels.

Un haut niveau d'automatisation indique généralement un haut niveau de maturité opérationnelle. Cette information influencera la vitesse de déploiement de nouveaux services. En cas de faible automatisation, les ingénieurs DevOps auront certainement un grand nombre de tâches manuelles à réaliser avant de pouvoir déployer un nouveau service. Avec une automatisation élevée, en revanche, les développeurs pourront tirer parti d'un pipeline automatisé pour déployer rapidement un service, sans recourir à des ressources humaines plus utiles ailleurs.

La maturité de la gestion opérationnelle s'évalue également par le nombre de contrôles en place pour gérer les nouveaux déploiements. Les scans de vulnérabilité, la sécurité des conteneurs, en particulier en ce qui concerne la sécurité des outils d'exécution et la chaîne d'approvisionnement, les politiques d'autorisation de modification de services et les tests d'objectifs de niveau de services sont d'importants contrôles de déploiement pour garantir la stabilité, la sécurité et le respect des accords de niveau de service.

L'évaluation honnête de la maturité opérationnelle est un élément essentiel de la stratégie de déploiement multi-cloud, car elle contribue à déterminer les exigences opérationnelles et à identifier les éventuels points d'amélioration pour votre entreprise.

Définir les caractéristiques de la stratégie spécifiques à l'entreprise

L'étape suivante dans l'élaboration d'une stratégie de déploiement Kubernetes multi-cloud est de déterminer les caractéristiques de la stratégie spécifiques à l'entreprise. Il s'agit notamment des besoins en formation des collaborateurs, de la responsabilité du déploiement et des futurs objectifs relatifs à l'utilisation du déploiement K8s. Examinons tout d'abord comment évaluer la responsabilité du déploiement.

Bien que ce point soit similaire à ce qui a été mentionné plus haut sur l'utilisation de Kubernetes, il s'agit ici davantage d'une question de gestion et de responsabilité. La question qu'il convient de se poser est : « Qui est chargé du déploiement ? »

Si les développeurs et ingénieurs logiciels sont déjà les principaux utilisateurs de Kubernetes et que l'utilisation de K8s ne fait pas partie intégrante de la fourniture de services, donner la responsabilité du déploiement multi-cloud aux équipes de développement est une solution envisageable.

Il se peut que la direction préfère accorder cette responsabilité à une équipe d'infrastructure si votre stratégie opérationnelle est axée sur l'efficacité et la fiabilité de votre déploiement Kubernetes. Dans ce cas, votre stratégie a certainement catégorisé votre déploiement K8s en tant qu'outil de fourniture de services et non comme actif stratégique.

Cependant, si votre entreprise considère le déploiement comme un actif stratégique pour poursuivre ses objectifs, accorder la responsabilité de Kubernetes à un directeur de la technologie ou un directeur des systèmes d'information peut être la meilleure solution.

L'étape suivante consiste à définir les objectifs d'utilisation de Kubernetes pour votre entreprise. Cela dépend fortement de la stratégie de l'entreprise et des problématiques opérationnelles spécifiques à votre entreprise, mais certaines questions d'ordre général peuvent vous aider à définir vos besoins spécifiques :

- Quelles équipes utiliseront Kubernetes ?
- Quels services seront déployés ?
- Comment migrer les applications existantes qui seront exécutées sur K8s ?
- Quels sont les objectifs opérationnels de votre entreprises et ses exigences de niveau de service ?
- Comment les objectifs d'efficacité financière et opérationnelle s'inscrivent-ils dans vos déploiements K8s ?
- Et les objectifs de disponibilité et d'évolutivité ?

Il est difficile de trouver le juste équilibre entre limiter l'accès aux opérations Kubernetes pour les professionnels DevOps et donner aux développeurs la liberté d'effectuer des changements. Un trop grand nombre de restrictions augmente la charge de travail des équipes DevOps, mais un environnement trop permissif peut être plus difficile à contrôler.

Cela représente une quantité considérable d'informations à réunir et à prendre en compte, mais l'élaboration d'une stratégie complète axée sur les objectifs évitera à votre entreprise bien des difficultés futures.

Réfléchissez enfin à la formation des équipes existantes à l'utilisation du nouveau déploiement et à la migration des charges de travail vers les nouveaux environnements. Sans personne pour travailler dessus, vos nouveaux déploiements K8s ne mèneront nulle part. Vous devez donc prévoir une formation Kubernetes pour les développeurs, l'équipe de support opérationnel, les ingénieurs DevOps et toute autre personne ayant besoin d'une connaissance pratique de Kubernetes.

En ce qui concerne la migration, commencez par les services déjà exécutés dans des conteneurs. Ce seront les plus simples à déplacer et ils fourniront une base opérationnelle pour les futures migrations. Les services hautement prioritaires exécutés sur des machines virtuelles doivent également figurer en haut de la liste pour la migration. Bien que la migration de ces services soit plus délicate, leur exécution dans un déploiement Kubernetes leur sera extrêmement bénéfique, car ils ont généralement besoin d'un niveau de disponibilité et d'évolutivité élevé.

Déterminer où Kubernetes sera exécuté

Pour élaborer une stratégie de déploiement complète, il convient également de déterminer où exécuter vos déploiements. La première option est d'exécuter Kubernetes exclusivement sur site.

Bien que cela soit tentant de conserver le contrôle total de votre déploiement et de l'infrastructure sur lequel il est exécuté, cette option s'offre uniquement aux entreprises disposant d'un nombre de serveurs et d'une capacité de stockage suffisants pour gérer les déploiements à grande échelle. Héberger des déploiements Kubernetes exclusivement en interne requiert en outre une compréhension approfondie du fonctionnement de K8s.

Une autre option consiste à exécuter Kubernetes sur un cloud unique. C'est un bon début pour beaucoup d'entreprises, car cela permet de supprimer l'obstacle matériel de l'option sur site et de réduire les frais de gestion opérationnelle. Exécuter Kubernetes sur le cloud permet d'accéder à des ressources informatiques et de stockage hautement disponibles, mais l'option d'un cloud unique pose un problème d'enfermement propriétaire.

Pour éviter l'enfermement propriétaire, vous pouvez exécuter Kubernetes en tant que déploiement multi-cloud. Dans la plupart des cas, il s'agit de l'option la plus complexe. Les frais d'exploitation sont élevés et des ressources humaines importantes sont nécessaires pour une gestion adéquate, mais l'exécution d'un déploiement multi-cloud présente un avantage considérable.

L'exécution de Kubernetes sur deux clouds ou plus permet d'éviter l'enfermement propriétaire tout en bénéficiant de la disponibilité et de la flexibilité des services cloud. Et enfin, n'oubliez pas les déploiements en Edge. Si votre entreprise doit exécuter des déploiements Kubernetes en Edge, elle aura besoin d'un support spécialisé. L'exécution de K8s en Edge nécessite un plan d'intégration de la plateforme et une solution de gestion centralisée.

Définir les plans d'exploitation

Pour finir, intégrez vos plans d'exploitation à votre stratégie de déploiement Kubernetes. L'utilisation et la gestion de déploiements multi-cluster s'accompagnent de défis DevOps spécifiques. Vos pratiques de développement, de test et de déploiement doivent toutes s'intégrer à votre nouveau déploiement multi-cluster, qui peut demander beaucoup de travail de DevOps.

Veillez à automatiser autant que possible vos pipelines CI/CD afin de réduire les frais de DevOps et continuer à exploiter les avantages des déploiements multi-cluster. La gestion de l'infrastructure constitue un autre défi DevOps et soulève de nouvelles questions :

- Comment allez-vous gérer l'infrastructure sur site et sur le cloud ?
- Quels types de serveurs allez-vous utiliser ?
- Comment le stockage sera-t-il géré ?

Bien que ces questions soient courantes pour les ingénieurs DevOps, de nouvelles réponses sont souvent nécessaires lorsque l'on passe à un modèle multi-cluster. L'application des politiques, la surveillance et la collecte de journaux sont autant de défis DevOps supplémentaires.

La gestion des politiques est tout particulièrement importante lorsque l'on exécute différents services avec différentes exigences et priorités d'infrastructure, tandis que la surveillance et la collecte de journaux sont importantes pour comprendre comment Kubernetes utilise les ressources.

D'un point de vue opérationnel toujours, la gestion des identités et la sécurité doivent également être prises en compte. Lors d'un nouveau déploiement Kubernetes, vous devez identifier des rôles et autorisations afin de sécuriser les ressources et garantir la gestion adéquate du déploiement.

Cela nécessite ensuite d'intégrer le nouveau déploiement Kubernetes à votre fournisseur d'identité. Dans le même ordre d'idée, citons les problématiques de conformité. Vous devrez peut-être créer des

politiques pour isoler l'exécution de certaines tâches ou le stockage de certaines données afin de répondre aux normes gouvernementales et sectorielles. Par ailleurs, des contrôles de sécurité, tels que les scans de vulnérabilité et la gestion des journaux, doivent être en place pour répondre aux exigences de conformité.

Le dernier aspect opérationnel à prendre en compte est la continuité et la reprise après sinistre. Bien que Kubernetes soit connu pour accroître la disponibilité, des interruptions de service sont toujours possibles. Avant le déploiement, définissez la perte de données maximale admissible (RPO) et la durée maximale d'interruption admissible (RTO). Ces métriques définissent une norme opérationnelle pour vos déploiements de service qui permettra à DevOps et aux autres équipes de mesurer l'efficacité et la résilience opérationnelles.

Pour accroître la continuité et favoriser la reprise après sinistre, il est possible d'utiliser des déploiements à la fois multi-régionaux et multi-cloud. Les déploiements multi-régionaux regroupent les ressources dans plusieurs régions, ce qui permet de continuer à répondre aux exigences informatiques et de stockage en cas d'interruption chez un fournisseur et les données peuvent être répliquées dans différentes régions, ce qui augmente la résilience et réduit la durée d'interruption.

Les déploiements multi-cloud sont similaires aux déploiements multi-régionaux, mais sont exécutés sur les clouds de différents fournisseurs, pas dans différentes régions géographiques. Bien que ces deux solutions soient très efficaces pour améliorer les RPO et RTO, attention aux frais liés aux sorties de données. Lorsque les données sont transférées entre les clouds et les régions, les frais de sortie peuvent rapidement s'accumuler.

CHAPITRE 5

Choisir les bonnes solutions pour votre stratégie Kubernetes multi-cloud

DANS CE CHAPITRE :

- Caractéristiques essentielles d'un outil de gestion multi-cluster
- Pourquoi Rancher est la bonne solution pour gérer un déploiement Kubernetes multi-cluster dans un environnement multi-cloud
- Autres avantages de Rancher

Les avantages de l'utilisation de Kubernetes sur plusieurs clouds et sur site sont bien établis, tout comme ses défis. Créer une stratégie multi-cloud d'entreprise complète et bien définie est la première étape pour transformer la manière de fournir vos services. Cela permet de guider les décisions organisationnelles et, sur bien des aspects, les décisions opérationnelles. Une autre étape importante pour optimiser les avantages de Kubernetes est d'avoir les bons outils multi-cloud en place.

Caractéristiques essentielles d'un outil de gestion multi-cluster

Au strict minimum, les outils multi-cloud doivent pouvoir garantir des opérations cohérentes, fournir des capacités de gestion des charges de travail robustes, prendre en charge les distributions Kubernetes multiples afin d'éviter l'enfermement propriétaire et soutenir la gestion de la sécurité. En ce qui concerne la cohérence des opérations, ces outils doivent vous aider pour le provisionnement, la gestion de versions, la

visibilité sur l'état des services et de l'infrastructure, les problèmes de services et de performances diagnostiques, le suivi des données d'événement et les capacités d'audit centralisées.

Les outils de gestion des charges de travail supportent le déploiement et la mise à jour des applications et services, y compris les services au long cours ainsi que les charges de travail par lot et les tâches ad hoc. Ces outils peuvent vous aider à gérer différents types de charges de travail, notamment les services sans état, les services avec état, les daemons et les tâches par lot.

Un outil de gestion Kubernetes multi-cloud doit aussi prendre en charge l'application des politiques de sécurité sur l'ensemble des clusters, quelle que soit leur localisation. Privilégiez les outils qui vous permettent de mettre en œuvre des contrôles d'accès cohérents pour les comptes utilisateurs et de service et d'appliquer les politiques applicables à l'ensemble des clusters. La gestion de la configuration inclut l'application des politiques et la sécurité, mais pas uniquement : une gestion de la configuration effectuée correctement, avec les bons outils, permet la standardisation et l'automatisation à grande échelle. Gérer des clusters à grande échelle représente un défi important, non seulement en termes de gestion de la configuration, mais aussi de gestion du cycle de vie des clusters. Des outils de gestion multi-cluster, comme Rancher, sont nécessaires pour répondre à ces défis.

Tous les outils de gestion des clusters ne prennent pas en charge l'intégralité du cycle de vie. Les solutions offrant une gestion du cycle de vie total sont capables de configurer, provisionner, sécuriser, auditer, mettre à niveau et donner l'accès à des applications tierces sur différents clusters.

Par exemple, OpenShift, un outil de gestion des clusters très répandu, propose des fonctionnalités d'observabilité limitées qui réduisent la capacité de l'opérateur à gérer entièrement le cycle de vie de ses clusters.

Bien sûr, n'importe quel outil de gestion Kubernetes devrait fonctionner correctement dans le vaste écosystème Kubernetes. Lorsque vous cherchez une solution Kubernetes multi-cloud, optez pour une solution qui fonctionne bien avec des outils open source cloud native. Cherchez tout particulièrement des solutions dotées de capacités de gestion du cycle de vie complet, comme celles de Rancher.

L'une des raisons pour lesquelles Kubernetes est devenu la référence non officielle de l'orchestration de conteneurs, c'est qu'un ensemble standard d'API sont prises en charge par les différentes implémentations. Selon la Cloud Native Computing Foundation, il existe plus de 90 offres Kubernetes certifiées et chacune d'entre elles prend en charge une API courante offrant un moyen standard d'interagir avec Rancher ou d'autres outils de gestion.

Pour être certifiée, l'offre doit être conforme aux spécifications établies par la CNCF. L'utilisation d'une distribution Kubernetes certifiée garantit la cohérence entre les implémentations certifiées, l'accès aux mises à jour de Kubernetes et la possibilité de confirmer que votre distribution est conforme grâce à l'exécution d'une application de conformité open source appelée Sonobuoy. Tous les outils de gestion Kubernetes multi-cloud doivent pouvoir prendre en charge n'importe quel cluster Kubernetes conforme.

L'écosystème Kubernetes évolue rapidement. Attendez-vous à voir apparaître de nouveaux outils pour vous aider à gérer et automatiser des environnements multi-cluster exécutés sur plusieurs clouds.

Pourquoi Rancher est la bonne solution pour gérer un déploiement Kubernetes multi-cluster dans un environnement multi-cloud

La liste des exigences pour un outil de gestion Kubernetes multi-cloud adéquat est longue. Les défis pour créer un tel outil sont importants, mais ils ont été relevés, notamment par Rancher.

Rancher est un outil de gestion Kubernetes multi-cloud proposé par SUSE. Conçu pour ne dépendre d'aucune plateforme, il est idéal pour la gestion multi-cluster et multi-cloud. Rancher permet aux opérateurs multi-cloud de maîtriser l'expansion de clusters qui peuvent croître, souvent de manière organique, sur plusieurs clouds. Rancher permet l'intégration native et la gestion du cycle de vie des clusters d'AWS EKS, Google Cloud GKE et Azure AKS. Il s'agit d'un avantage considérable pour les opérateurs qui ne disposent pas d'une connaissance approfondie des tâches spécifiques au cloud pour travailler avec Kubernetes.

Résilience

Un autre avantage du caractère multi-cloud de Rancher apparaît lorsqu'un problème survient. La sauvegarde et la récupération sont aussi importantes avec Kubernetes qu'avec n'importe quel autre système IT, mais en raison du caractère hautement distribué de Kubernetes, les méthodes de sauvegarde traditionnelles peuvent poser problème. Tout particulièrement sur plusieurs clouds.

Rancher fournit un mécanisme de sauvegardes automatiques régulières de la base de données etcd, qui enregistre l'état du cluster. En cas de sinistre, ces instantanés de sauvegarde peuvent être utilisés pour restaurer le cluster au dernier état correct connu. Bien entendu, vous devrez accéder à d'autres éléments, comme les fichiers de configuration du cluster, le stockage persistant et les images des conteneurs. Ceux-ci sont tous inclus dans les capacités de sauvegarde et de reprise après sinistre de Rancher.

Rancher offre non seulement son propre système de sauvegarde, mais également des outils de sauvegarde tiers fournis par les partenaires de SUSE et qui peuvent vous aider dans vos stratégies de sauvegarde et de récupération.

En outre, Rancher est déployé par défaut en haute disponibilité sur plus de trois nœuds, ce qui renforce encore sa résilience.

Gestion des identités

L'authentification et l'autorisation des utilisateurs représentent un autre défi plus difficile à relever dans des systèmes distribués que dans des applications monolithiques. Heureusement, Rancher assure une authentification centralisée.

Celle-ci vous permet de définir et gérer un ensemble d'utilisateurs pour tous vos clusters. Rancher utilise un proxy d'authentification pour authentifier les utilisateurs et envoyer des requêtes à vos clusters à l'aide d'un compte de service.

Ce proxy fonctionne avec une large gamme de services d'authentification, y compris Microsoft Active Directory, Microsoft Azure AD, OpenLDAP, GitHub, Okta, Google OAuth, et bien plus encore. Rancher utilise également le concept de groupes pour gérer les utilisateurs, ce qui permet de donner aux utilisateurs des rôles spécifiques.

Autres avantages de Rancher

Rancher rationalise un grand nombre de tâches de gestion des clusters. Notamment la gestion des membres des clusters, la modification et la mise à niveau des clusters, la gestion des nœuds, des volumes persistants et des classes de stockage, la gestion des projets et des espaces de noms, l'exécution de scans de sécurité et la rotation des certificats.

Rancher est 100 % open source et gratuit. Il prend en charge de manière native les distributions de Rancher Kubernetes Engine (RKE), K3s et RKE2— ainsi que toutes les distributions Kubernetes certifiées CNCF — ce qui permet aux opérateurs de choisir la meilleure distribution Kubernetes pour leur environnement. Par exemple, RKE s'exécute entièrement sur des conteneurs Docker et RKE2 est une distribution K8s dotée d'une sécurité renforcée conçue pour les environnements à haut risque qui donnent la priorité à la sécurité. K3s a été créée par Rancher et donnée à la CNCF en tant que distribution légère parfaite en Edge / pour les emplacements distants limités en ressources.

Figure 5 : Rancher et ses outils connexes offrent les bases d'une gestion automatisée tout au long du cycle de vie du cluster

De plus, toutes ces distributions sont prises en charge nativement dans les abonnements de support SUSE Rancher sans frais supplémentaires.

Rancher s'intègre également à d'autres composants de l'écosystème Kubernetes, comme Longhorn, Harvester ou Helm (voir la **Figure 5**).

Longhorn est un système de stockage de blocs distribué open source pour Kubernetes, qui rationalise l'ajout d'un stockage persistant à un cluster Kubernetes. Longhorn possède plusieurs fonctionnalités importantes de protection des données. Le système de stockage garantit des sauvegardes et instantanés incrémentiels, y compris des sauvegardes programmées. De plus, en cas de récupération après une sauvegarde, vous pouvez choisir de restaurer uniquement les données dont vous avez besoin, plutôt que de restaurer des volumes entiers. Les sauvegardes sont enregistrées sur des systèmes de stockage secondaires à l'extérieur du cluster, par exemple sur le volume NFS d'un système de stockage d'objets compatible AWS S3.

Harvester est une solution d'infrastructure hyperconvergée conçue pour être utilisée avec des serveurs bare metal. Elle permet de déployer les charges de travail de machines virtuelles et peut s'intégrer à Rancher afin de consolider les charges de travail de machines virtuelles et de conteneurs sur une plateforme unique.

Helm est un gestionnaire de paquets open source utilisé pour déployer et gérer les applications dans Kubernetes. Helm fonctionne avec une abstraction de gestion des paquets, les charts. Les charts comprennent

les informations nécessaires pour créer une instance d'une application Kubernetes. Les charts de Helm offrent une solution standard pour empaqueter et déployer les logiciels, tandis que les opérateurs constituent une solution non standard et légèrement compliquée d'empaqueter et distribuer les logiciels pour Kubernetes.

Il existe également une abstraction, une config, qui contient les informations de configuration pour l'exécution d'une application dans un environnement spécifique. Par exemple, une application qui effectue des transformations de données dans un ensemble de données d'apprentissage automatique aura un chart décrivant l'application.

Il peut y avoir plusieurs configs, par exemple, une pour chaque environnement de développement, de test, de production. Les charts et configs peuvent être combinés pour créer un objet pouvant être libéré. Une troisième abstraction, la release, exécute l'instance d'une application décrite dans un chart combiné à une config.

Ressources supplémentaires pour commencer

Les communautés Kubernetes et Rancher développent activement les plateformes ainsi que des ressources d'aide. En voici quelques-unes pour vous aider à commencer.

- La documentation Kubernetes officielle comprend des sections sur les bases, les concepts fondamentaux, les tâches courantes, des tutoriels, et bien plus encore.
- Le blog SUSE est une autre ressource incontournable pour Kubernetes, tout particulièrement si vous êtes à la recherche de conseils et meilleures pratiques. Par exemple, grâce à l'article Introduction to Kubernetes Workloads (en anglais), vous en saurez plus sur les éléments constitutifs des charges de travail dans Kubernetes. Il décrit également les liens entre services, déploiements, ensembles de réplicas, de deamons, d'applications avec état, endpoints, etc. et leur rôle dans les charges de travail.

- L'article Disaster Recovery Preparedness for Your Kubernetes Clusters (en anglais) explique en détail comment utiliser Rancher pour configurer automatiquement des sauvegardes automatiques de composants clés de Kubernetes, tels que la base de données etcd. Vous découvrirez également les éléments d'un scénario réussi de reprise après sinistre et sa mise en œuvre.

Ressources spécifiques à Rancher

- La communauté Rancher possède un canal Slack sur slack.rancher.io, où vous pouvez consulter des discussions, poser des questions et vous tenir au courant des dernières activités liées à Rancher.
- Le portail de la communauté SUSE & Rancher sur community.suse.com propose des liens vers des événements, vidéos et supports de formation. On y trouve aussi bien des introductions à Kubernetes que des master classes ou des cours approfondis de plusieurs semaines sur Rancher et K3s.
- Les documents Rancher officiels sont disponibles sur rancher.com/docs/. Il s'agit de présentations, guides de démarrage rapide, meilleures pratiques, conseils de gestion, etc.

C'est parti

Dans ce guide Gorilla, nous avons abordé la manière dont une plateforme de gestion Kubernetes telle que Rancher peut contribuer à optimiser la valeur de vos clusters sur des clouds hébergés et comment démarrer avec Kubernetes et Rancher.

Téléchargez Rancher ici pour commencer votre exploration. Si vous souhaitez en savoir plus sur la manière dont Rancher peut vous aider à optimiser Kubernetes pour votre entreprise, demandez une démo. Vous pouvez également rejoindre le canal Slack de Rancher ou la communauté SUSE, trouver un événement SUSE près de chez vous ou lire la documentation Rancher.

Merci de votre lecture !

À PROPOS DE SUSE

SUSE est un leader mondial en matière de solutions open source innovantes, fiables et adaptées aux besoins des entreprises. Plus de 60 % des entreprises du classement Fortune 500 font confiance à SUSE pour alimenter leurs environnements système essentiels. Nous sommes spécialisés dans les solutions Enterprise Linux, Kubernetes Management et Edge, et collaborons avec des partenaires et des communautés pour donner à nos clients les moyens d'innover partout - du datacenter, au cloud, en Edge et au-delà. SUSE redonne du sens au mot « open » dans « open source », en donnant aux clients l'agilité nécessaire pour relever les défis en matière d'innovation aujourd'hui et la liberté de faire évoluer leur stratégie et leurs solutions demain. La société emploie près de 2 000 personnes dans le monde et est cotée sur le marché réglementé (Prime Standard) de la Bourse de Francfort.

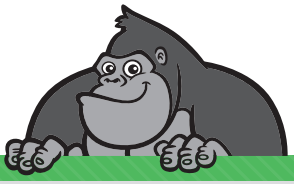
Pour plus d'informations, rendez-vous sur www.suse.com.

À PROPOS D'ACTUALTECH MEDIA

ActualTech Media est une entreprise de marketing technologique B2B qui met en relation les fournisseurs d'informatique d'entreprise et les acheteurs d'informatique par le biais de programmes de génération de leads innovants et de services de contenu personnalisé attractif.

L'équipe d'ActualTech Media sait parler au public de l'informatique d'entreprise car nous avons été le public de l'informatique d'entreprise.

Notre équipe de direction rassemble un grand nombre d'anciens DSI, IT managers, architectes, experts et professionnels du marketing grâce auxquels nos clients peuvent consacrer moins de temps à expliquer ce que fait leur technologie et plus à développer des stratégies efficaces.



Vous êtes responsable marketing informatique et vous souhaitez obtenir votre propre titre personnalisé Gorilla Guide® pour votre société, rendez-vous sur https://www.gorilla.guide/custom-solutions/