Utilisation de Kubernetes à la DIPSO

Retour d'expérience

Martin Souchal

24 Mai 2024

ID-ICS - DIPSO - INRAE

Utilisation de k8s à la DIPSO

clusters

- Cluster de tooling (monitoring, gestion des logs...)
- Onyxia (calcul sur GPU)
- Cahiers de laboratoires
- Forums Discourse

INRAE
la science pour la vie, l'humain, la terre

K8S

Pourquoi utiliser Kubernetes

- Adoption d'une démarche DevOps, usine logicelle
- Facilité de mettre en production des conteneurs
- Scalabilité
- Portabilité des déploiements
- Pas d'adhésion au hardware
- Facilité d'administration
- Partage de GPUs

INRAE
la science pour la vie, l'humain, la terre

Hébergement

- Plateforme OpenStack Orion
- Externe (Scaleway...)

INRAE
la science pour la vie, l'humain, la terre

Stack logicielle

- Référentiel sur Gitlab MIA (IaC + GitOps)
- Instanciation de clusters avec OpenTofu (Debian 12)
- Déploiement de K8S "Vanilla" avec Kubespray. Actuellement on utilise la version 1.28
- Runtime : containerd
- Réseau avec IPtable, calico
- ArgoCD pour GitOps
- Prometheus et Grafana pour monitoring
- Ingress Nginx
- Helm

INRAE
la science pour la vie, l'humain, la terre

Sécurité

- Gestion certificats avec CertManager
- PodSecurity
- Trivy pour audit sécu

INRAE
la science pour la vie, l'humain, la terre

Sécurité

Workload Assessment

Namespace	Resource	Vulnerabilities					Misconfigurations					Secrets				
		C	H	M	L	U	C	H	M	L	U	C	H	M	L	U
observability	Deployment/prometheus-kube-state-metrics							1		6						
observability	Deployment/prometheus-blackbox-exporter		6	17					1	7						
observability	StatefulSet/prometheus-alertmanager		8	28				1	2	7						
observability	ConfigMap/prometheus-server							1								
observability	Deployment/prometheus-server		20	52				2	4	15						
observability	Deployment/grafana	3	31	60	4			3	2	18						
observability	ConfigMap/prometheus-alertmanager							1	1							
observability	DaemonSet/prometheus-prometheus-node-exporter		4	15				3	2	9						
ingress-nginx	Job/ingress-nginx-admission-create									5						
ingress-nginx	Job/ingress-nginx-admission-patch									5						
ingress-nginx	Deployment/ingress-nginx-controller		6	28	3			2	1	4						
gitlab-runner	Deployment/gitlab-runner		8	49	2			1	1	7						
gitlab-runner	ConfigMap/gitlab-runner							1	1							
cinder-csi	Deployment/openstack-cinder-csi-controllerplugin							6	19	54						
cinder-csi	DaemonSet/openstack-cinder-csi-nodeplugin							7	11	27						
cert-manager	Deployment/cert-manager-cainjector									6						
cert-manager	Deployment/cert-manager									6						
cert-manager	Deployment/cert-manager-webhook									6						

Severities: C=CRITICAL H=HIGH M=MEDIUM L=LOW U=UNKNOWN

Infra Assessment

Namespace	Resource	Vulnerabilities					Misconfigurations					Secrets				
		C	H	M	L	U	C	H	M	L	U	C	H	M	L	U
kube-system	Service/coredns								1							
kube-system	Pod/nginx-proxy-node3		6	21	2			2	5	7						
kube-system	Service/kube-dns								1							
kube-system	Pod/kube-apiserver-node1							2	5	16						
kube-system	Pod/kube-scheduler-node1							2	4	9						
kube-system	DaemonSet/kube-proxy		5	8	17			3	5	9						
kube-system	DaemonSet/calico-node		24	20	38			9	14	32						
kube-system	ConfigMap/extension-apiserver-authentication								1							
kube-system	Deployment/coredns	2	16	18				1	3	3						
kube-system	Deployment/dns-autoscaler							1	3	6						
kube-system	Service/metrics-server								1							
kube-system	Deployment/calico-kube-controllers		2	1				1	4	5						
kube-system	Deployment/metrics-server								1	4						

INRAE
la science pour la vie, l'humain, la terre

Sécurité

RBAC Assessment

Namespace	Resource	RBAC Assessment				
		C	H	M	L	U
kube-system	Role/system::leader-locking-kube-scheduler			1		
kube-system	Role/system:controller:cloud-provider			1		
kube-system	Role/system:controller:bootstrap-signer			1		
kube-system	Role/system:controller:token-cleaner			1		
kube-system	Role/system::leader-locking-kube-controller-manager			1		
kube-system	Role/kubeadm:kubeadm-certs			1		
kube-public	Role/system:controller:bootstrap-signer			1		
kube-public	RoleBinding/kubeadm:bootstrap-signer-clusterinfo	1				
ingress-nginx	Role/ingress-nginx-admission			1		
ingress-nginx	Role/ingress-nginx			1		
gitlab-runner	Role/gitlab-runner	1				
cert-manager	Role/cert-manager-webhook:dynamic-serving			2		
	ClusterRole/prometheus-kube-state-metrics	1				
	ClusterRole/system:aggregate-to-admin	1				
	ClusterRole/system:aggregate-to-edit	2	4	6		
	ClusterRole/system:controller:endpoint-controller		1			
	ClusterRole/system:dns-autoscaler			1		
	ClusterRole/system:controller:replicaset-controller			2		
	ClusterRole/cert-manager-controller-orders	1				
	ClusterRole/system:controller:namespace-controller	1				
	ClusterRole/cert-manager-controller-certificates	1				
	ClusterRole/system:controller:cronjob-controller			3		
	ClusterRole/system:controller:statefulset-controller			1		
	ClusterRole/cert-manager-controller-issuers	1				
	ClusterRole/system:controller:generic-garbage-collector	1				
	ClusterRole/prometheus-server		1			
	ClusterRole/system:controller:job-controller			2		
	ClusterRole/system:controller:replication-controller			2		
	ClusterRole/system:controller:persistent-volume-binder	1	2	1		
	ClusterRole/system:kube-controller-manager	5				
	ClusterRole/ingress-nginx-admission	1				

INRAE
la science pour la vie, l'humain, la terre

Gestion des GPUs

- Nvidia A100 sur Orion
- Attachés en passthrough sur les VMs
- Partagé dans K8S (Time-Slicing)

INRAE
la science pour la vie, l'humain, la terre

Déploiement

- CI/CD sur Gitlab
- Création des VMs avec OpenTofu
- Déploiement du cluster K8S avec Kubespray
- Déploiement d'un runner dans le cluster
- Push d'une release Helm depuis gitlab

INRAE
la science pour la vie, l'humain, la terre

Architecture

- Pas de LoadBalancer sur Orion -¿ Utilisation de HA Proxy hors Kube avec IP flottante
- Réseau privé sur CP et Worker
- Reverse Proxy tcp vers NodePort interne sur HA
- CSI Cinder pour les PVC
- Velero pour Backup

Orion - Problèmes rencontrés

- Bug de CSI Cinder sur les AZs (cf https://github.com/kubernetes/cloud-provider-openstack/issues/2185)
- IPv6 en interne non supporté
- Problème de DNS en interne - contournement avec Google DNS

INRAE
la science pour la vie, l'humain, la terre

Administration

- Monitoring avec kube-prometheus-stack
- Supervision avec k9s
- IDE : OpenLens

INRAE
la science pour la vie, l'humain, la terre

Améliorations

- Coté OpenStack : Magnum, LB...
- Travail sur la robustesse des pipelines et leur maintenabilité
- Consolider la base OS des machines (OS immutables)

INRAE
la science pour la vie, l'humain, la terre

Merci beaucoup!

Mail : martin.souchal@inrae.fr