



FlexPod pour la génomique

FlexPod

NetApp
October 30, 2025

Sommaire

FlexPod pour la génomique	1
Tr-4911 : génomique de l'FlexPod	1
Portée	2
Public	2
Les capacités de l'hôpital sont déployées sur FlexPod	2
Avantages liés au déploiement de workloads génomiques avec FlexPod	3
Composants matériels et logiciels de l'infrastructure de la solution	9
Génomique - configuration et exécution de la GATK	13
De l'identification de l'échantillon à la variante, de l'annotation et de la prédiction	13
Boîte à outils d'analyse génomique (GATK)	15
Définition FlexPod	15
Configuration et exécution de la GATK	19
Sortie pour l'exécution de GATK à l'aide du fichier jar	23
Sortie pour l'exécution de GATK à l'aide du script ./gatk	26
Sortie pour l'exécution de la GATK à l'aide du moteur Cromwell	28
Configuration des GPU	32
Conclusion	41
Où trouver des informations complémentaires	42
Historique des versions	43

FlexPod pour la génomique

Tr-4911 : génomique de l'FlexPod

JayaKishore Esanakula, NetApp

La génomique joue un rôle clé dans la recherche médicale et infirmière. Elle compte peu de domaines de médecine plus importants que la génomique au service de la santé et des sciences de la vie. La génomique, associée à l'imagerie médicale et à la pathologie digitale, nous permet de comprendre comment les gènes d'un patient peuvent être affectés par les protocoles de traitement. La réussite de la génomique dans le domaine de la santé dépend de plus en plus de l'interopérabilité des données à grande échelle. L'objectif final est de donner du sens aux énormes volumes de données génétiques et d'identifier des corrélations et variantes pertinentes sur le plan clinique qui améliorent le diagnostic et rendent la médecine de précision possible. La génomique nous aide à comprendre l'origine des épidémies, la façon dont les maladies évoluent et les traitements et stratégies susceptibles d'être efficaces. De toute évidence, la génomique offre de nombreux avantages qui couvrent la prévention, le diagnostic et le traitement. Les établissements de santé sont aux prises avec plusieurs défis, notamment :

- Amélioration de la qualité des soins
- Soins basés sur la valeur
- Explosion des données
- Médecine de précision
- Pandémies
- Wearables, télésurveillance et soins
- Cybersécurité

Les voies cliniques normalisées et les protocoles cliniques constituent l'une des composantes essentielles de la médecine moderne. L'interopérabilité entre les fournisseurs de soins, pas seulement pour les dossiers médicaux, mais aussi pour les données génomiques, est un des aspects clés de la standardisation. La question majeure est-ce que les établissements de santé vont abandonner la propriété des données génomiques au lieu de la propriété des patients pour leurs données génomiques personnelles et les dossiers médicaux connexes ?

L'interopérabilité des données des patients est essentielle pour permettre une médecine de précision, l'un des moteurs de la récente explosion des données. L'objectif de la médecine de précision est de rendre le maintien de la santé, la prévention des maladies, les diagnostics et les solutions de traitement plus efficaces et plus précis.

Le taux de croissance des données a été exponentiel. Début février 2021, les laboratoires américains ont séquencé environ 8,000 souches de COVID-19 par semaine. Le nombre de génomes séquencés était passé à 29,000 par semaine en avril 2021. Chaque génome humain complètement séquencé est d'environ 125 Go. Par conséquent, à un rythme de 29,000 génomes par semaine, le stockage total du génome au repos serait donc de plus de 180 pétaoctets par an. Divers pays ont consacré des ressources à l'épidémiologie génomique afin d'améliorer la surveillance génomique et de se préparer à la prochaine vague de défis mondiaux en matière de santé.

La réduction du coût de la recherche génomique favorise les tests et la recherche génétiques à un rythme sans précédent. Les trois PS se situent à un point d'inflexion : la puissance informatique, la confidentialité des données et la personnalisation de la médecine. D'ici 2025, les chercheurs estiment que 100 millions à près de 2 milliards de génomes humains seront séquencés. Pour que la génomique soit efficace et une proposition précieuse, les capacités génomiques doivent faire partie intégrante des flux de travail des soins ; il doit être facile d'accéder et d'être exploitable lors de la visite d'un patient. Il est également important d'intégrer les données médicales électroniques des patients dans les données génomiques des patients. Avec l'arrivée de l'infrastructure convergée de pointe comme FlexPod, les entreprises peuvent intégrer leurs fonctionnalités génomiques dans les workflows quotidiens des médecins, des infirmiers et des responsables des cliniques. Pour obtenir les toutes dernières informations sur la plate-forme FlexPod, reportez-vous à ce document ["Livre blanc FlexPod Datacenter avec Cisco UCS X-Series"](#).

Pour un médecin, toute la valeur de la génomique inclut une médecine de précision et des plans de traitement personnalisés basés sur les données génomiques d'un patient. Par le passé, les médecins et les scientifiques des données n'ont jamais connu une telle synergie, et la génomique bénéficie des innovations technologiques récentes, ainsi que de partenariats réels entre les organismes de santé et les leaders technologiques du secteur.

Les centres médicaux universitaires et d'autres organismes de soins de santé et de sciences de la vie sont en bonne voie pour établir un centre d'excellence (COE) en science du génome. Selon M. Charlie Gersbach, Dr Greg Crawford, et le docteur Tim E Reddy de l'Université Duke : « nous savons que les gènes ne sont pas activés ou désactivés par un simple commutateur binaire, mais plutôt le résultat de multiples commutateurs de régulation génétique qui fonctionnent ensemble. Ils ont également déterminé que « aucune de ces parties du génome ne fonctionne isolément. Le génome est un web très complexe que l'évolution a tissé » ("réf").

NetApp et Cisco se sont acharnés à mettre en œuvre des améliorations incrémentielles de la plateforme FlexPod depuis plus de 10 ans. Tous les commentaires des clients sont entendus, évalués et liés aux flux de valeur et aux fonctionnalités de FlexPod. C'est ce processus continu de retour d'informations, de collaboration, d'amélioration et de célébration qui fait de FlexPod une plateforme d'infrastructure convergée de confiance au monde entier. Elles ont été simplifiées et conçues dès le départ pour être la plateforme la plus fiable, robuste, polyvalente et agile des établissements de santé.

Portée

La plateforme d'infrastructure convergée FlexPod permet à un établissement de santé d'héberger un ou plusieurs workloads de génomique avec d'autres applications cliniques et non cliniques de santé. Dans ce rapport technique, nous fait appel à un outil génomique open source standard appelé GATK lors de la validation de la plateforme FlexPod. Toutefois, une discussion plus approfondie de la génomique ou de la GATK est en dehors de la portée de ce document.

Public

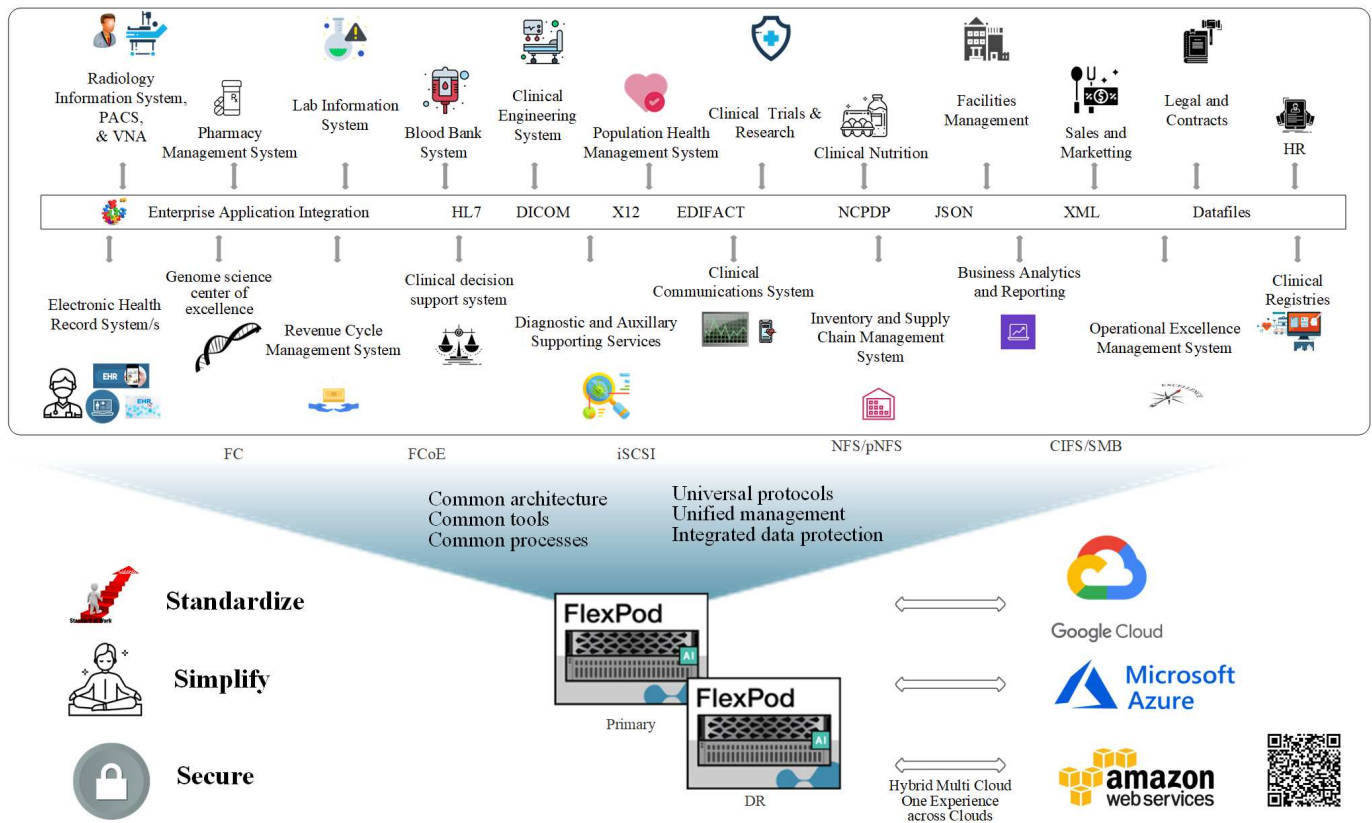
Ce document est destiné aux leaders techniques du secteur de la santé, aux ingénieurs solutions partenaires Cisco et NetApp et aux équipes des services professionnels. NetApp suppose que le lecteur connaît bien les concepts de dimensionnement du stockage et du calcul, ainsi que la connaissance technique des menaces médicales, de la sécurité sanitaire, des systèmes IT de santé, de Cisco UCS et des systèmes de stockage NetApp.

Les capacités de l'hôpital sont déployées sur FlexPod

Un hôpital typique possède un ensemble diversifié de systèmes INFORMATIQUES. La majorité de ces systèmes sont achetés auprès d'un fournisseur, alors que très peu sont construits par le système hospitalier en interne. C'est pourquoi le système hospitalier doit gérer un environnement d'infrastructure diversifié dans ses data centers. Lorsque les hôpitaux unifient leurs systèmes sur une plateforme d'infrastructure convergée

telle que FlexPod, les entreprises peuvent standardiser les opérations de leur data Center. Avec FlexPod, les établissements de santé peuvent mettre en place des systèmes cliniques et non cliniques sur une même plateforme, unifiant ainsi les opérations du data Center.

Hospital capabilities deployed on a FlexPod



"Ensuite, les avantages du déploiement de workloads génomiques avec FlexPod."

Avantages liés au déploiement de workloads génomiques avec FlexPod

"Précédent : introduction."

Dans cette section, vous trouverez un bref aperçu des avantages dont vous pouvez bénéficier pour exécuter un workload génomique sur une plateforme d'infrastructure convergée FlexPod. Décrire rapidement les capacités d'un hôpital. L'architecture métier suivante montre les capacités d'un hôpital déployées sur une plateforme d'infrastructure convergée FlexPod prête pour le cloud hybride.

- **Éviter les silos dans les soins de santé.** les silos dans les soins de santé sont une préoccupation très réelle. Elles sont souvent cloisonnées dans leur propre ensemble de matériel et de logiciels, et non pas selon leur choix, mais de façon organique par leur évolution. Par exemple, radiologie, cardiologie, EHR, génomique, les analyses, les cycles de revenus et les autres départements se terminent par leur jeu individuel de logiciels et de matériel dédiés. Les organismes de santé disposent d'un ensemble limité de professionnels IT pour gérer leurs ressources matérielles et logicielles. Le point d'inflexion s'est produit lorsque ce groupe de particuliers doit gérer un ensemble très diversifié de matériel et de logiciels. L'hétérogénéité est aggravée par un ensemble peu homogène de processus mis en place par les

fournisseurs dans l'organisation de soins de santé.

- **Démarrer petit et grandir.** le kit d'outils GATK est réglé pour l'exécution de la CPU, qui les meilleures suites plates-formes comme FlexPod. FlexPod offre une évolutivité indépendante du réseau, du calcul et du stockage. Commencez par une infrastructure de petite taille et faites-la évoluer à mesure que vos capacités en génomique et votre environnement se développent. Les organismes de santé n'ont pas à investir dans des plateformes spécialisées pour gérer des charges de travail génomiques. Les entreprises peuvent ainsi exploiter des plateformes polyvalentes telles qu'un système FlexPod pour exécuter des charges de travail génomiques et non génomiques sur une même plateforme. Par exemple, si le service de pédiatrie souhaite mettre en œuvre une fonctionnalité de génomique, le leadership INFORMATIQUE peut provisionner les ressources de calcul, de stockage et de réseau sur une instance FlexPod existante. À mesure que l'entité commerciale génomique se développe, le secteur de la santé peut faire évoluer sa plateforme FlexPod en fonction des besoins.
- **Panneau de contrôle unique et flexibilité inégalée.** Cisco Intersight simplifie considérablement les opérations INFORMATIQUES en rapprochant les applications de l'infrastructure, en fournissant la visibilité et la gestion des serveurs bare-Metal et des hyperviseurs aux applications sans serveur, réduisant ainsi les coûts et les risques. Cette plateforme SaaS unifiée s'intègre de façon native aux plateformes et outils tiers. Elle permet en outre de gérer l'ensemble des opérations de votre data Center sur site ou en tout lieu, à l'aide d'une application mobile.

Les utilisateurs valorisent rapidement leur environnement en utilisant Intersight comme plateforme de gestion. En permettant l'automatisation de nombreuses tâches manuelles quotidiennes, InterSight élimine les erreurs et simplifie vos opérations quotidiennes. En outre, les capacités de support avancées offertes par Intersight permettent aux utilisateurs de garder une longueur d'avance et d'accélérer la résolution des problèmes. Combinées, les entreprises réduisent considérablement leur infrastructure applicative et consacrent plus de temps au développement de leur activité principale.

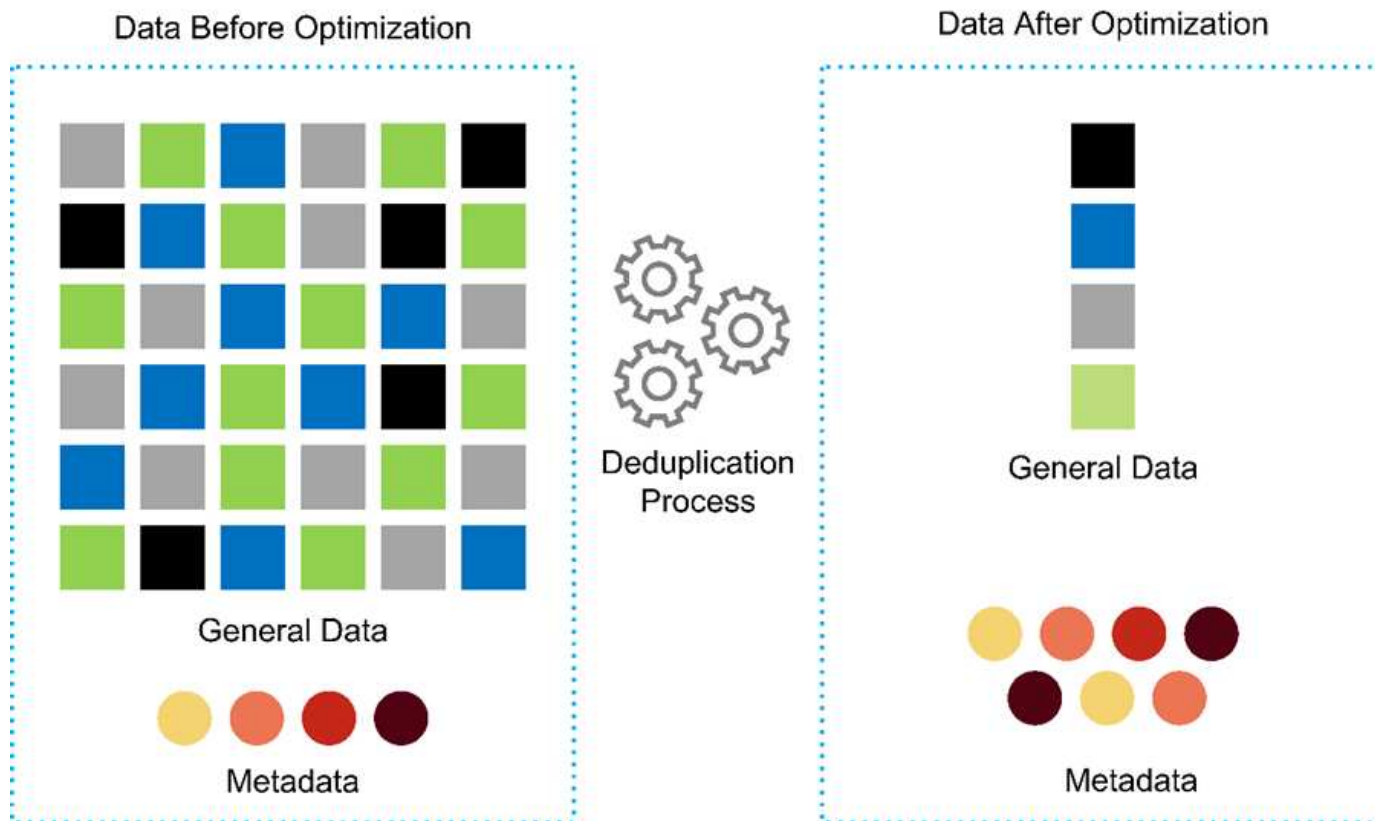
En tirant parti de la gestion Intersight et de l'architecture facilement évolutive de FlexPod, les entreprises peuvent exécuter plusieurs charges de travail du génome sur une plateforme FlexPod unique, ce qui améliore le taux d'utilisation et réduit le coût total de possession (TCO). FlexPod offre des possibilités de dimensionnement flexibles, avec des choix parmi notre petite infrastructure FlexPod Express et une évolutivité vers de grandes implémentations FlexPod Datacenter. Grâce aux fonctionnalités de contrôle d'accès basées sur des rôles intégrées dans Cisco InterSight, les établissements de santé peuvent mettre en place des mécanismes de contrôle d'accès robustes, ce qui évite d'avoir recours à des piles d'infrastructure distinctes. Plusieurs entités commerciales peuvent exploiter la génomique comme une compétence clé dans le domaine de la santé.

En définitive, FlexPod simplifie les opérations IT et réduit les coûts d'exploitation, et permet aux administrateurs D'infrastructure IT de se concentrer sur des tâches permettant aux médecins d'innover au lieu de rester relégués au second plan.

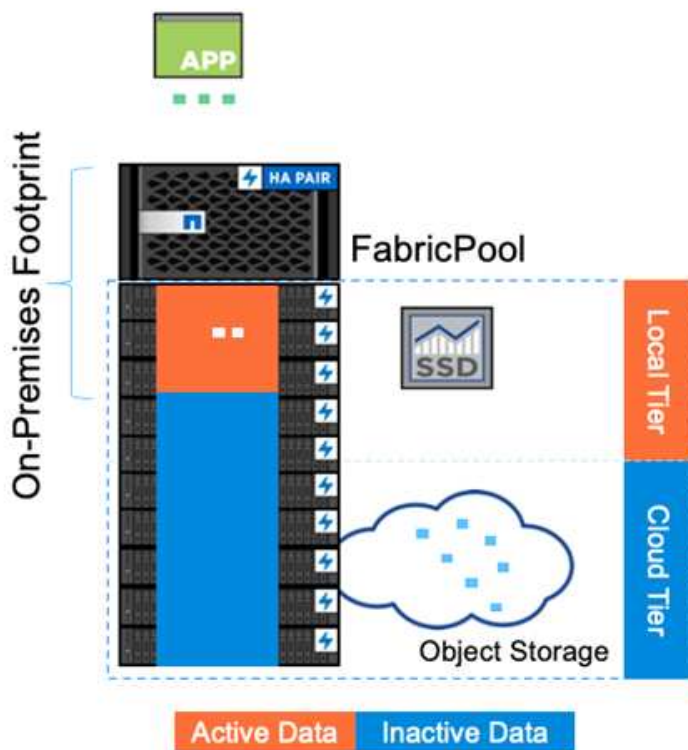
- **Conception validée et résultats garantis.** les guides de conception et de déploiement FlexPod sont validés de manière reproductible et couvrent les détails complets de la configuration et les meilleures pratiques du secteur nécessaires au déploiement en toute confiance d'un FlexPod. Les guides de conception validée par Cisco et NetApp, les guides de déploiement et les architectures aident votre département des soins de santé ou des sciences de la vie à éviter toute approximation lors de la mise en œuvre d'une plateforme validée et fiable dès le départ. Avec FlexPod, vous accélérez les délais de déploiement, tout en réduisant les coûts, la complexité et les risques. Les designs validés et les guides de déploiement de FlexPod établissent le FlexPod comme plateforme idéale pour de nombreuses charges de travail génomiques.
- **Innovation et agilité.** FlexPod est recommandé comme plate-forme idéale par les EHR comme Epic, Cerner, Meditech et les systèmes d'imagerie comme Agfa, GE, Philips. Pour plus d'informations sur "[Le programme « Epic Honor roll »](#)" Et l'architecture de la plateforme cible, consultez le site Web d'Epic userweb. Une bonne exécution de la génomique "[FlexPod](#)" permet aux établissements de santé de

poursuivre leur transition vers l'innovation avec agilité. Avec FlexPod, la mise en œuvre de changements organisationnels est naturelle. Lorsque les entreprises optent pour une plateforme FlexPod, les experts du secteur de la santé peuvent provisionner leur temps, leurs efforts et leurs ressources pour innover, et ainsi être aussi agiles que les besoins de l'écosystème.

- **Données libérées.** avec la plateforme d'infrastructure convergée FlexPod et un système de stockage NetApp ONTAP, les données génomiques peuvent être disponibles et accessibles à l'aide d'un large éventail de protocoles à grande échelle depuis une plateforme unique. FlexPod avec NetApp ONTAP propose une plateforme de cloud hybride simple, intuitive et puissante. Votre Data Fabric optimisé par NetApp ONTAP offre un maillage sur l'ensemble des sites, des emplacements physiques et des applications, Votre Data Fabric est conçue pour un monde centré sur la donnée. Les données étant créées et exploitées dans divers emplacements et, la plupart du temps, partagées avec d'autres sites, applications et infrastructures, C'est pourquoi il vous faut une méthode de gestion cohérente et intégrée. FlexPod permet à votre équipe IT de maîtriser et de simplifier une INFRASTRUCTURE IT toujours plus complexe.
- **Colocation sécurisée.** FlexPod utilise des modules cryptographiques conformes à la norme FIPS 140-2, ce qui permet aux entreprises de mettre en œuvre la sécurité comme élément fondamental, et non comme élément secondaire. FlexPod permet aux entreprises d'implémenter une colocation sécurisée à partir d'une plateforme d'infrastructure convergée unique, quelle que soit leur taille. FlexPod avec colocation sécurisée et qualité de service pour séparer les charges de travail et optimiser l'utilisation. Cela permet d'éviter que le capital soit dépendant de plateformes spécialisées sous-utilisées et que la gestion requiert une compétence spécialisée.
- **Efficacité du stockage.** la génomique exige que le stockage sous-jacent offre des fonctionnalités d'efficacité du stockage de pointe. Vous pouvez réduire les coûts du stockage grâce aux fonctionnalités d'efficacité du stockage NetApp, telles que la déduplication (à la volée et à la demande), la compression et la compaction des données ("réf"). La déduplication NetApp fournit une déduplication au niveau des blocs dans un volume FlexVol. Pour schématiser, la déduplication supprime les blocs dupliqués et ne stocke que les blocs uniques dans le volume FlexVol. La déduplication, qui fonctionne avec un niveau de granularité élevé, fonctionne sur le système de fichiers actif du volume FlexVol. La figure suivante présente le fonctionnement de la déduplication NetApp. La déduplication est transparente pour les applications. Par conséquent, elle peut être utilisée pour dédupliquer des données provenant de toute application qui utilise le système NetApp. Vous pouvez exécuter la déduplication volume comme processus à la volée ou en arrière-plan. Vous pouvez le configurer pour qu'il s'exécute automatiquement, de manière à être planifié ou manuellement via l'interface de ligne de commande, NetApp ONTAP System Manager ou NetApp Active IQ Unified Manager.



- Activer l'interopérabilité génomique.** ONTAP FlexCache est une fonctionnalité de mise en cache à distance qui simplifie la distribution de fichiers, réduit la latence WAN et réduit les coûts de bande passante WAN ("réf"). L'une des principales activités durant l'identification et l'annotation des variantes génomiques est la collaboration entre les cliniciens. La technologie ONTAP FlexCache augmente le débit de données même lorsque les médecins travaillent en collaboration à différents endroits. Étant donné la taille type d'un fichier *.BAM (1 Go à des centaines de Go), il est essentiel que la plate-forme sous-jacente puisse mettre des fichiers à la disposition des cliniciens dans différents emplacements géographiques. Avec FlexPod et ONTAP FlexCache, les données et les applications génomiques sont réellement prêtes pour plusieurs sites afin que la collaboration entre les chercheurs du monde entier s'effectue de manière fluide, avec une faible latence et un débit élevé. Les organismes de santé qui exécutent des applications génomiques dans un environnement multisite peuvent évoluer horizontalement avec la Data Fabric pour équilibrer la gestion avec la vitesse et les coûts.
- Utilisation intelligente de la plate-forme de stockage.** La gestion des données est simplifiée grâce à FlexPod avec le Tiering automatique ONTAP et la technologie NetApp FabricPool. FabricPool permet de réduire les coûts de stockage sans nuire aux performances, à l'efficacité, à la sécurité ou à la protection. FabricPool est transparent pour les applications d'entreprise et capitalise sur l'efficacité du cloud en réduisant le TCO du stockage sans devoir repenser l'architecture de l'infrastructure applicative. FlexPod bénéficie des fonctionnalités de hiérarchisation du stockage de FabricPool pour une utilisation plus efficace du stockage Flash ONTAP. Pour plus d'informations, voir "[FlexPod avec FabricPool](#)". Le diagramme suivant présente FabricPool et ses avantages.



Automatic tiering
 Zero-touch management
 Preserves file system
 Lower cost of ownership
 Choice of object tier locations



- **Analyse et annotation des variantes plus rapides.** la plate-forme FlexPod est plus rapide à déployer et opérationnaliser. Avec la plateforme FlexPod, les médecins peuvent collaborer avec les médecins en rendant les données à grande échelle, avec une faible latence et un débit plus élevé. L'interopérabilité accrue favorise l'innovation. Les établissements de santé peuvent traiter leurs workloads génomiques et non génomiques, ce qui signifie qu'elles n'ont pas besoin de plateformes spécialisées pour commencer leur transition vers la génomique.

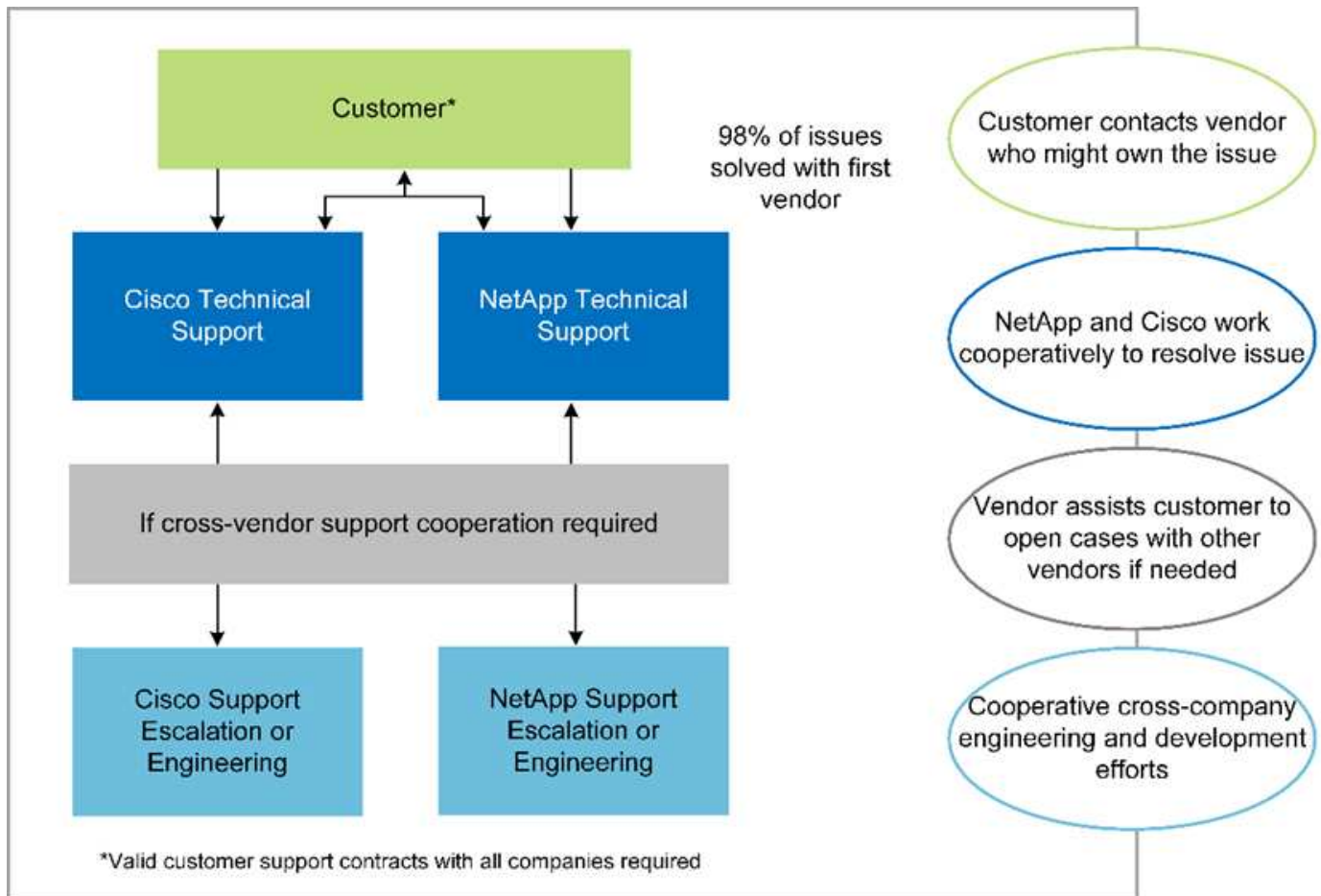
FlexPod ONTAP ajoute régulièrement des fonctionnalités de pointe à la plateforme de stockage. FlexPod Datacenter constitue une base d'infrastructure partagée idéale pour déployer la technologie FC- NVMe afin d'autoriser l'accès au stockage haute performance aux applications en besoin. Au fur et à mesure que la connectivité FC- NVMe évolue, incluant la haute disponibilité, les chemins d'accès multiples et une prise en charge supplémentaire du système d'exploitation, FlexPod convient aussi bien à la plateforme de choix, offrant l'évolutivité et la fiabilité nécessaires pour prendre en charge ces fonctionnalités. Grâce à la technologie ONTAP dotée d'E/S plus rapides et à la technologie NVMe de bout en bout, les analyses génomiques sont plus rapides ("réf").

Les données de génome brut séquencé génèrent des fichiers de grandes tailles et il est important que ces fichiers soient mis à la disposition des analyseurs variantes pour réduire le temps total nécessaire de la collecte des échantillons à l'annotation des variantes. NVMe (Nonvolatile Memory Express), utilisé comme protocole d'accès au stockage et de transport de données, offre un débit sans précédent et des délais de réponse très courts. FlexPod déploie le protocole NVMe en accédant au stockage Flash via le bus PCI express (PCIe). Grâce à l'implémentation de dizaines de milliers de files d'attente de commandes, il est possible d'augmenter la parallélisation et le débit. Un protocole unique, du stockage à la mémoire, permet un accès rapide aux données.

- * L'agilité pour la recherche clinique depuis le départ.* la capacité de stockage flexible et extensible et la performance permet aux organismes de recherche en santé d'optimiser l'environnement de façon élastique ou juste-à-temps (JIT). En découplant le stockage de l'infrastructure de calcul et de réseau, la plateforme FlexPod peut évoluer verticalement et horizontalement sans perturbation. Grâce à Cisco

Intersight, la plateforme FlexPod peut être gérée à l'aide de flux de travail automatisés intégrés et personnalisés. Les workflows Cisco Intersight permettent aux établissements de santé de réduire la durée de gestion du cycle de vie des applications. Lorsqu'un centre médical universitaire exige que les données des patients soient anonymisées et mises à disposition par son centre d'informatique pour la recherche et/ou d'un centre de qualité, le service IT peut exploiter les workflows Cisco Intersight FlexPod pour effectuer des sauvegardes de données sécurisées, cloner et restaurer les données en quelques secondes, et non plus en quelques heures. Avec NetApp Trident et Kubernetes, les départements IT peuvent provisionner de nouveaux data Scientists et rendre les données cliniques disponibles pour le développement des modèles en quelques minutes, parfois même en quelques secondes.

- **Protection des données du génome.** NetApp SnapLock offre un volume spécial dans lequel les fichiers peuvent être stockés et archivés à un état non effaçable et non réinscriptible. Les données de production de l'utilisateur résidant dans un volume FlexVol peuvent être mises en miroir ou archivées sur un volume SnapLock grâce à la technologie NetApp SnapMirror ou SnapVault. Les fichiers du volume SnapLock, le volume lui-même et son agrégat d'hébergement ne peuvent pas être supprimés avant la fin de la période de conservation. Grâce au logiciel ONTAP FPolicy, les organisations peuvent empêcher les attaques par ransomware en désautorisant les opérations sur des fichiers avec des extensions spécifiques. Un événement FPolicy peut être déclenché pour des opérations de fichiers spécifiques. L'événement est lié à une politique, qui appelle le moteur qu'il doit utiliser. Vous pouvez configurer une règle avec un ensemble d'extensions de fichiers qui pourraient éventuellement contenir un ransomware. Lorsqu'un fichier doté d'une extension non autorisée tente d'effectuer une opération non autorisée, FPolicy empêche cette opération ("[réf](#)").
- **Support coopératif FlexPod.** NetApp et Cisco ont mis en place le modèle de support coopératif FlexPod, un modèle de support solide, évolutif et flexible, afin de répondre aux exigences de support uniques de l'infrastructure convergée FlexPod. Ce modèle tire parti de l'expérience, des ressources et de l'expertise de NetApp et de Cisco pour simplifier l'identification et la résolution des problèmes de support FlexPod, et ce, quelle que soit l'origine du problème. La figure suivante présente le modèle de support coopératif FlexPod. Le client contacte le fournisseur responsable du problème et travaille en collaboration avec Cisco et NetApp pour le résoudre. Cisco et NetApp ont des équipes d'ingénierie et de développement interentreprises qui travaillent main dans la main pour résoudre les problèmes. Ce modèle de support réduit la perte d'informations pendant la traduction, favorise la confiance et réduit les temps d'arrêt.



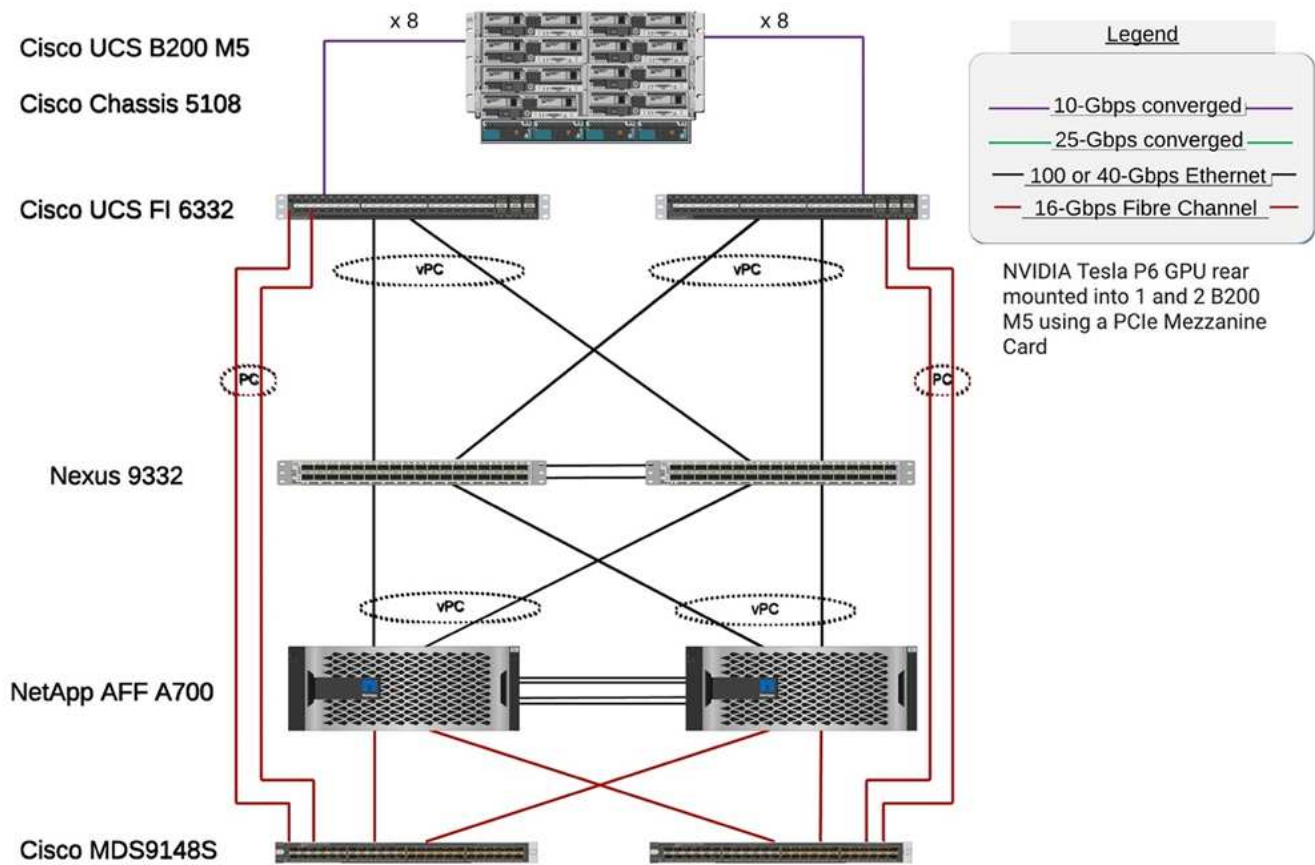
"Ensuite : composants matériels et logiciels de l'infrastructure de la solution."

Composants matériels et logiciels de l'infrastructure de la solution

"Précédent : avantages liés au déploiement des workloads génomiques avec FlexPod."

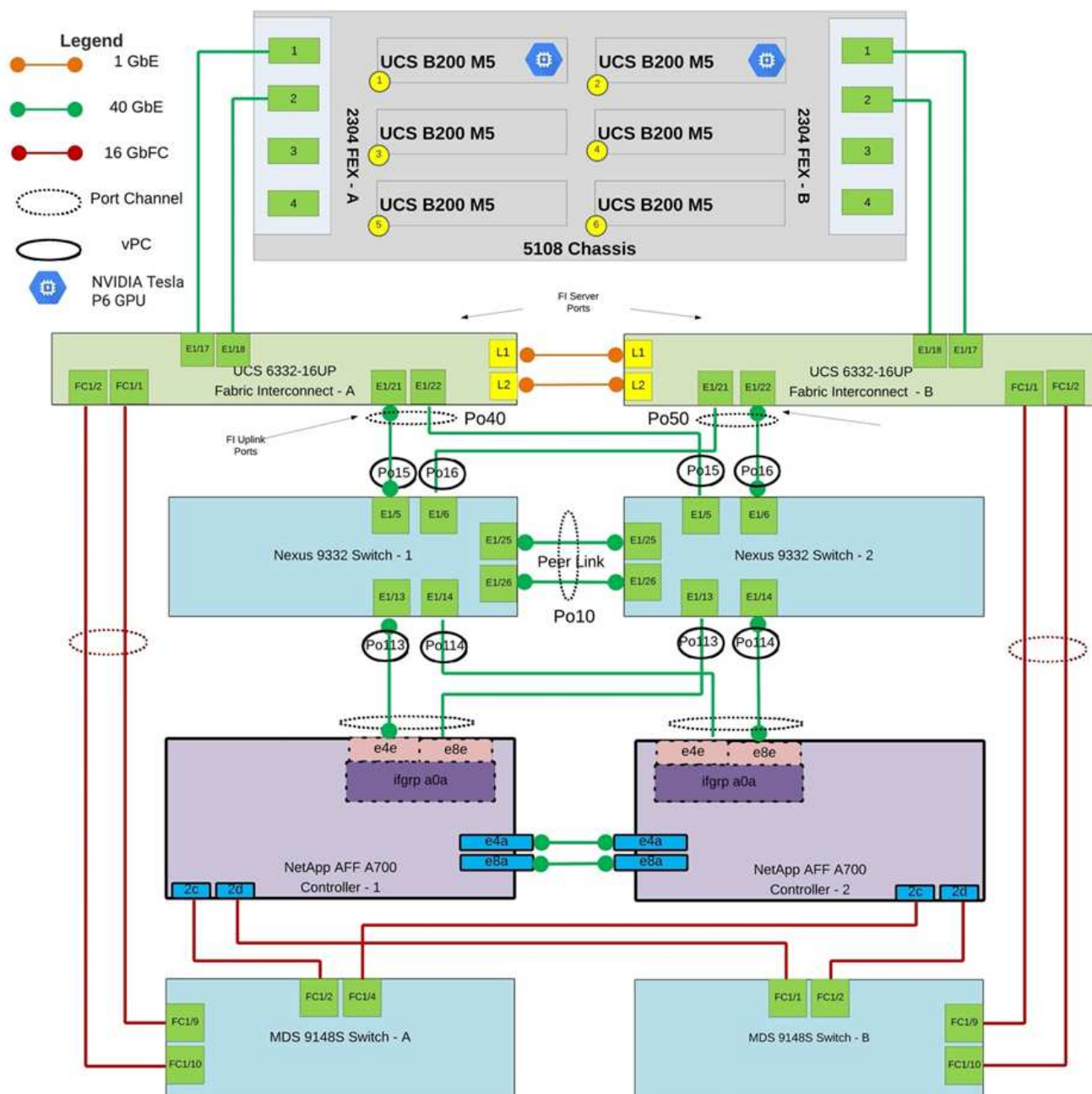
La figure suivante illustre le système FlexPod utilisé pour la configuration et la validation de la GATK. Nous avons utilisé "FlexPod Datacenter avec VMware vSphere 7.0 et NetApp ONTAP 9.7 conception validée par Cisco (CVD)" pendant le processus de configuration.

FlexPod for Genomics



Le diagramme suivant représente les détails du câblage FlexPod.

FlexPod for Genomics



Le tableau suivant répertorie les composants matériels utilisés lors du test GATK en activant sur un FlexPod. Voici le "Matrice d'interopérabilité NetApp" (IMT) et "Liste de compatibilité matérielle Cisco (HCL)".

Calque	Famille de produits	Quantité et modèle	Détails
Calcul	Châssis Cisco UCS 5108	1 ou 2	
	Les serveurs lames Cisco UCS	6 B200 M5	Chacun doté de 2 20 cœurs ou plus, de 2,7 GHz et de 128 Go de RAM

Calque	Famille de produits	Quantité et modèle	Détails
	Carte d'interface virtuelle Cisco UCS (VIC)	Cisco UCS 1440	Voir la
	2 interconnexions de fabric Cisco UCS	6332	-
Le réseau	Commutateurs Cisco Nexus	2 x Cisco Nexus 9332	-
Réseau de stockage	Réseau IP pour l'accès au stockage via les protocoles SMB/CIFS, NFS ou iSCSI	Mêmes commutateurs réseau que ci-dessus	-
	Accès au stockage via FC	2 x Cisco MDS 9148S	-
Stockage	Système de stockage 100 % Flash NetApp AFF A700	1 Cluster	Cluster à deux nœuds
	Tiroir disque	Un tiroir disque DS224C ou NS224	Plein avec 24 disques
	SSD	24, 1,2 To ou plus	-

Ce tableau répertorie le logiciel de l'infrastructure.

Logiciel	Famille de produits	Version ou version	Détails
Divers	Linux	RHEL 8.3	-
	Répertoires de base	Windows Server 2012 R2 (64 bits)	-
	NetApp ONTAP	ONTAP 9.8 ou version ultérieure	-
	Fabric Interconnect Cisco UCS	Cisco UCS Manager 4.1 ou version ultérieure	-
	Switchs Cisco Ethernet 3000 ou 9000	Pour la série 9000, 7.0(3)I7(7) ou ultérieure pour la série 3000, 9.2(4) ou ultérieure	-
	Cisco FC : Cisco MDS 9132T	8.4(1a) ou ultérieure	-
	Hyperviseur	VMware vSphere ESXi 7.0	-
Stockage	Système de gestion de l'hyperviseur	VMware vCenter Server 7.0 (vCSA) ou version ultérieure	-
Le réseau	NetApp Virtual Storage Console (VSC)	VSC 9.7 ou version ultérieure	-

Logiciel	Famille de produits	Version ou version	Détails
	NetApp SnapCenter	SnapCenter 4.3 ou version ultérieure	-
	Cisco UCS Manager	4.1(3c) ou ultérieure	
Hyperviseur	VMware ESXi		
Gestion	Système de gestion de l'hyperviseur VMware vCenter Server 7.0 (vCSA) ou version ultérieure		
	NetApp Virtual Storage Console (VSC)	VSC 9.7 ou version ultérieure	
	NetApp SnapCenter	SnapCenter 4.3 ou version ultérieure	
	Cisco UCS Manager	4.1(3c) ou ultérieure	

"Suivant: [Génomique - Configuration et exécution de la GATK.](#)"

Génomique - configuration et exécution de la GATK

"Précédent : [composants matériels et logiciels de l'infrastructure de la solution.](#)"

Selon l'Institut national de recherche sur le génome humain ("[NHGRI](#)"), « la génomique est l'étude de tous les gènes d'une personne (le génome), y compris les interactions entre ces gènes et avec l'environnement d'une personne. »

Selon le "[NHGRI](#)", "L'acide désoxyribonucléique (ADN) est le composé chimique qui contient les instructions nécessaires pour développer et diriger les activités de presque tous les organismes vivants. Les molécules d'ADN sont faites de deux torons couplés, souvent appelés « hélice double ». « L'ensemble complet d'ADN d'un organisme est appelé son génome. »

Le séquençage est le processus de détermination de l'ordre exact des bases dans un brin d'ADN. L'un des types de séquençage les plus courants utilisés aujourd'hui est appelé séquençage par synthèse. Cette technique utilise l'émission de signaux fluorescents pour commander les bases. Les chercheurs peuvent utiliser le séquençage de l'ADN pour rechercher des variations génétiques et des mutations susceptibles de jouer un rôle dans le développement ou la progression d'une maladie alors qu'une personne est encore au stade embryonnaire.

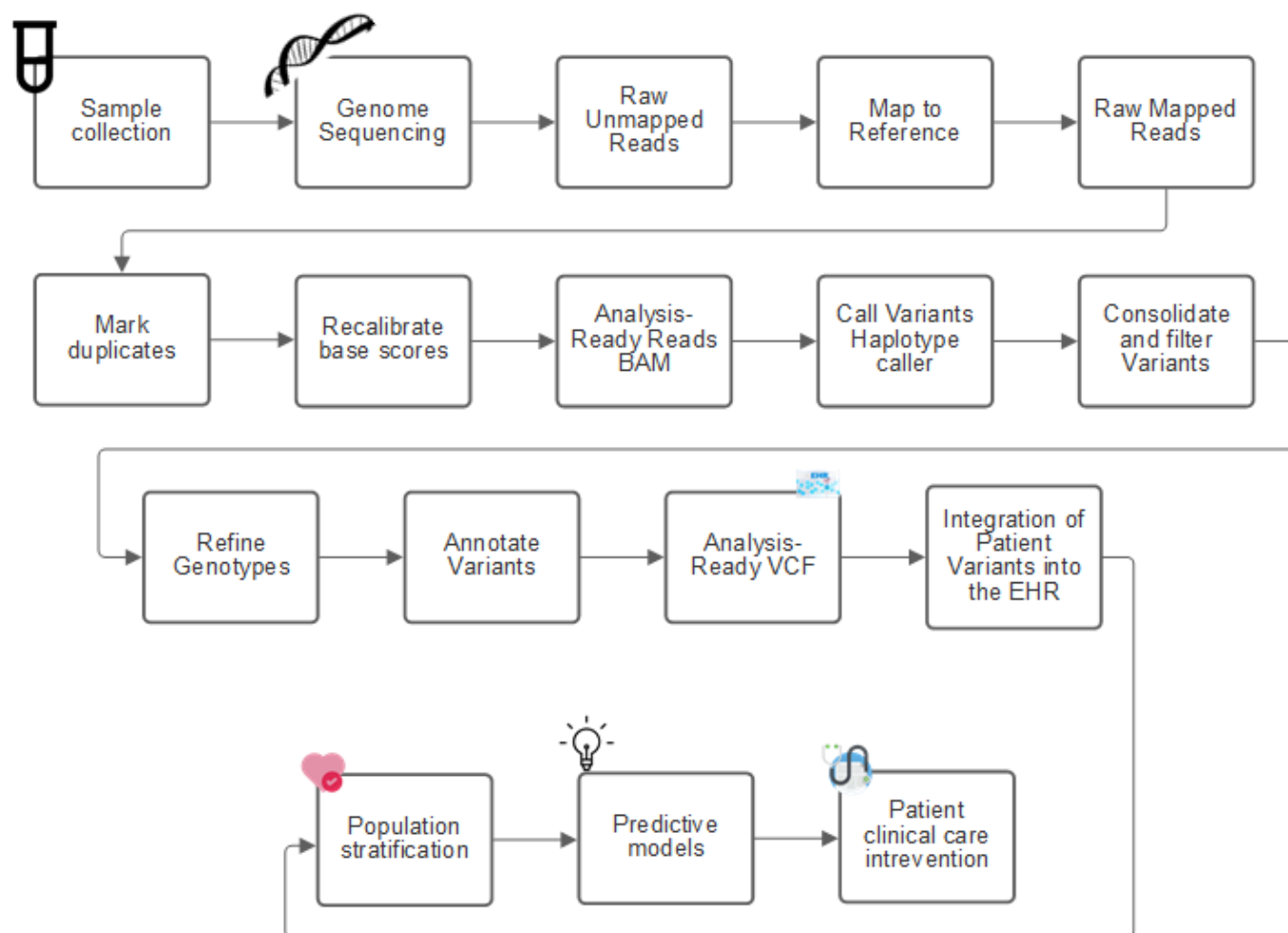
De l'identification de l'échantillon à la variante, de l'annotation et de la prédiction

À un niveau élevé, la génomique peut être classée comme suit. Cette liste n'est pas exhaustive :

1. Prélèvement d'échantillons.
2. "[Séquençage du génome](#)" utilisation d'un séquenceur pour générer les données brutes.
3. Prétraitement. Par exemple : "[déduplication](#)" à l'aide de "[Picard](#)".
4. Analyse génomique.

- a. Mappage sur un génome de référence.
- b. "**Variante**" Identification et annotation effectuées généralement à l'aide de la GATK et d'outils similaires.
5. Intégration au système de dossier médical électronique (EHR).
6. "**Stratification de la population**" et l'identification de la variation génétique entre la localisation géographique et le milieu ethnique.
7. "**Modèles prédictifs**" utilisant un polymorphisme significatif de nucléotide unique.
8. "**Validation**".

La figure suivante montre le processus de l'échantillonnage à l'identification de variante, à l'annotation et à la prédiction.



Le projet sur le génome humain a été achevé en avril 2003 et le projet a réalisé une simulation de très haute qualité de la séquence du génome humain disponible dans le domaine public. Ce génome de référence a initié une explosion de la recherche et du développement de capacités en génomique. Pratiquement chaque affection humaine a une signature dans les gènes de cet homme. Jusqu'à récemment, les médecins utilisaient les gènes pour prédire et déterminer les anomalies congénitales comme l'anémie falciforme, qui est causée par un certain modèle d'héritage causé par un changement dans un gène unique. Le Trésor des données mises à disposition par le projet du génome humain a conduit à l'avènement de l'état actuel des capacités en génomique.

La génomique offre de nombreux avantages. Voici quelques avantages dans le domaine de la santé et des sciences de la vie :

- Un meilleur diagnostic sur le lieu des soins
- Meilleur pronostic
- Médecine de précision
- Plans de traitement personnalisés
- Une meilleure surveillance des maladies
- Réduction des événements indésirables
- Amélioration de l'accès aux thérapies
- Amélioration de la surveillance des maladies
- Participation efficace aux essais cliniques et meilleure sélection des patients pour les essais cliniques basés sur les génotypes.

La génomique est une "[bête à quatre têtes](#)," compte tenu des besoins de calcul tout au long du cycle de vie d'un dataset : acquisition, stockage, distribution et analyse.

Boîte à outils d'analyse génomique (GATK)

La GATK a été développée comme plate-forme de science des données à l' "[Institut étendu](#)". GATK est un ensemble d'outils open source qui permettent l'analyse du génome, en particulier la découverte de variantes, l'identification, l'annotation et le génotypage. L'un des avantages de GATK est que l'ensemble d'outils et/ou de commandes peut être enchaîné pour former un workflow complet. Voici les principaux défis que le large institut doit relever :

- Comprendre les causes profondes et les mécanismes biologiques des maladies.
- Identifier les interventions thérapeutiques qui agissent à la cause fondamentale d'une maladie.
- Comprendre la ligne visuelle des variantes au fonctionnement dans la physiologie humaine.
- Création de normes et de règles "[frameworks](#)" pour la représentation, le stockage, l'analyse, la sécurité, etc. des données génomiques.
- Normaliser et socialiser les bases de données d'agrégation de génome interopérables (gnomAD).
- Surveillance, diagnostic et traitement basés sur des génomes pour le compte de patients avec plus de précision.
- Aider à mettre en œuvre des outils qui prédisent les maladies bien avant que les symptômes apparaissent.
- Créer et donner les moyens à une communauté de collaborateurs transdisciplinaires pour aider à s'attaquer aux problèmes les plus difficiles et les plus importants de la biomédecine.

Selon la GATK et le large institut, le séquençage du génome doit être considéré comme un protocole dans un laboratoire de pathologie. Chaque tâche est bien documentée, optimisée, reproductible et cohérente dans l'ensemble des échantillons et des expériences. Voici un ensemble de mesures recommandées par le large Institut pour plus d'informations, voir "[Site web de GATK](#)".

Définition FlexPod

La validation d'un workload génomique inclut une configuration complète d'une plateforme d'infrastructure FlexPod. La plateforme FlexPod est extrêmement disponible et peut évoluer indépendamment. Ainsi, vous pouvez faire évoluer indépendamment le réseau, le stockage et les ressources de calcul, par exemple. Nous avons utilisé le guide de conception validée Cisco suivant comme document d'architecture de référence pour configurer l'environnement FlexPod : "[FlexPod Datacenter avec VMware vSphere 7.0 et NetApp ONTAP 9.7](#)". Découvrez les points forts de la configuration de la plateforme FlexPod suivante :

Pour effectuer la configuration du laboratoire FlexPod, procédez comme suit :

1. La configuration et la validation du laboratoire FlexPod utilisent les réservations IP4 et les VLAN suivants.

IP Reservations

VLAN	IP Range	Subnet Mask	Purpose
3281	172.21.25 /24	255.255.255.0	IB-MGMT
3282	172.21.26 /24	255.255.255.0	vMotion
3283	172.21.27 /24	255.255.255.0	VM
3284	172.21.28 /24	255.255.255.0	NFS
3285	172.21.29 /24	255.255.255.0	iSCSI-A
3286	172.21.30 /24	255.255.255.0	iSCSI-B

2. Configuration des LUN de démarrage iSCSI sur le SVM ONTAP

The screenshot shows the ONTAP System Manager web interface. The left-hand navigation pane is expanded, showing the 'STORAGE' section with sub-items: Overview, Applications, Volumes, LUNs (highlighted), Shares, Qtrees, Quotas, Storage VMs, and Tiers. The main content area is titled 'LUNs' and features a '+ Add' button. Below this is a table with three columns: a checkbox, 'Name', and 'Storage VM'. The table lists six LUNs, all named 'ESXi_Boot_Lun_1' through 'ESXi_Boot_Lun_6', each associated with the 'Healthcare_SVM' Storage VM.

	Name	Storage VM
☐	ESXi_Boot_Lun_1	Healthcare_SVM
☐	ESXi_Boot_Lun_2	Healthcare_SVM
☐	ESXi_Boot_Lun_3	Healthcare_SVM
☐	ESXi_Boot_Lun_4	Healthcare_SVM
☐	ESXi_Boot_Lun_5	Healthcare_SVM
☐	ESXi_Boot_Lun_6	Healthcare_SVM

3. Mapper les LUN sur des groupes initiateurs iSCSI.

☐	Name	Storage VM	Volume	Size	IOPS	Latency (ms)	Throughput (MB/s)
☒	ESXi_Boot_Lun_1	Healthcare_SVM	ESXi_Boot_Vol	20 GB	3	0.16	0.01
STATUS  Online VOLUME ESXi_Boot_Vol DESCRIPTION - SNAPSHOT COPIES (LOCAL) STATUS  Protected SNAPMIRROR (LOCAL OR REMOTE) STATUS  Unprotected SERIAL NUMBER 80A4X+R8rAhP QOS POLICY GROUP - MAPPED TO INITIATORS ☒ GenomicsESXi_1 (1) iqn.1992-08.com.cisco:ucs-... ID 0 SNAPSHOT POLICY default CAPACITY (AVAILABLE % TOTAL) 95% 20 GB LUN FORMAT VMware PATH /vol/ESXi_Boot_Vol/ESXi_Boot_Lun_1							

CAPACITY (AVAILABLE % | TOTAL)

95% | 20 GB

LUN FORMAT

VMware

PATH

/vol/ESXi_Boot_Vol/ESXi_Boot_Lun_1

☐	Name	Storage VM	Volume	Size	IOPS	Latency (ms)	Throughput (MB/s)
	ESXi_Boot_Lun_1	Healthcare_SVM	ESXi_Boot_Vol	20 GB	1	0.25	0.01
	ESXi_Boot_Lun_2	Healthcare_SVM	ESXi_Boot_Vol	20 GB	4	0.18	0.02

STATUS		VOLUME	DESCRIPTION	SNAPSHOT COPIES (LOCAL)		SNAPMIRROR (LOCAL OR REMOTE)
Online		ESXi_Boot_Vol	-	STATUS		STATUS
				Protected		Unprotected
SERIAL NUMBER	QOS POLICY GROUP	MAPPED TO INITIATORS		ID	SNAPSHOT POLICY	
80A4X+R8rAhU	-	GenomicsESXi_2 (1) iqn.1992-08.com.cisco:ucs-...		0	default	
CAPACITY (AVAILABLE % TOTAL)		LUN FORMAT				
96% 20 GB		VMware				

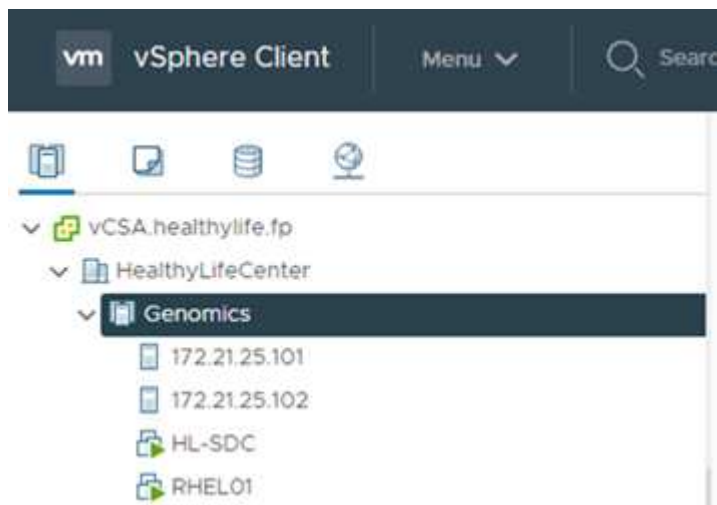
CAPACITY (AVAILABLE % | TOTAL)

96% | 20 GB

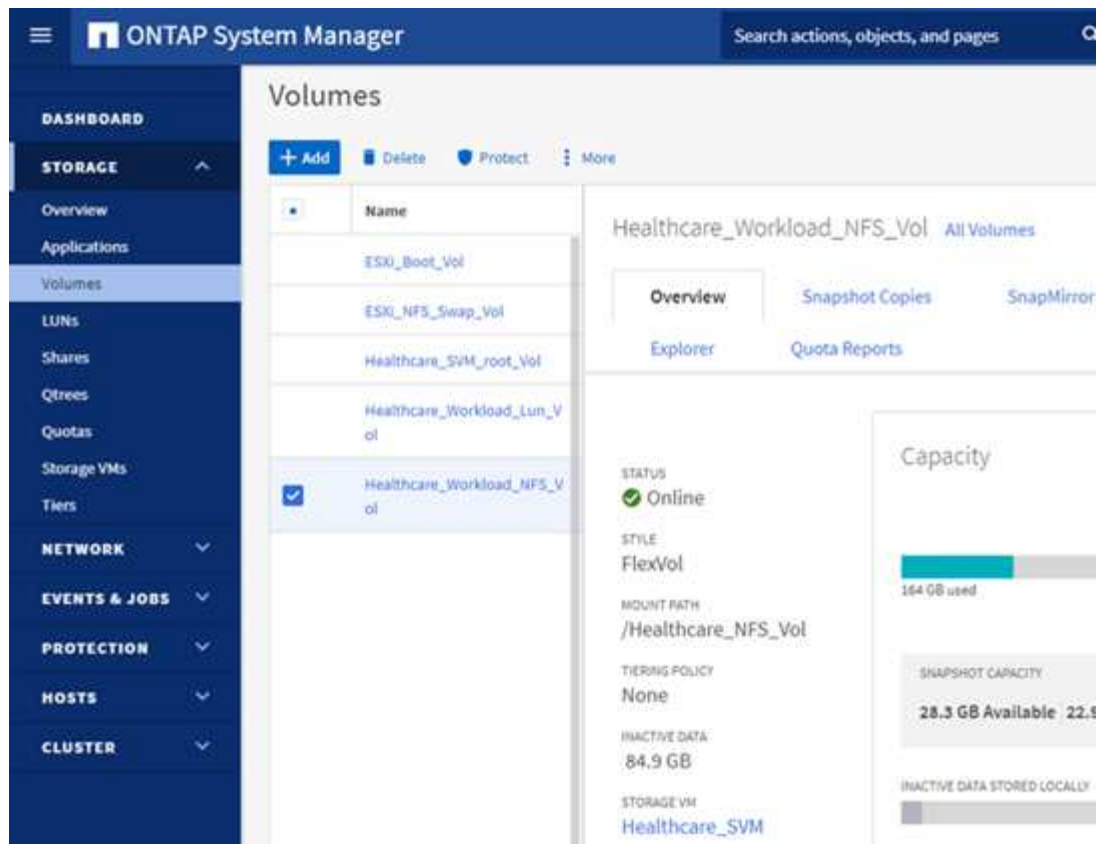
LUN FORMAT

VMware

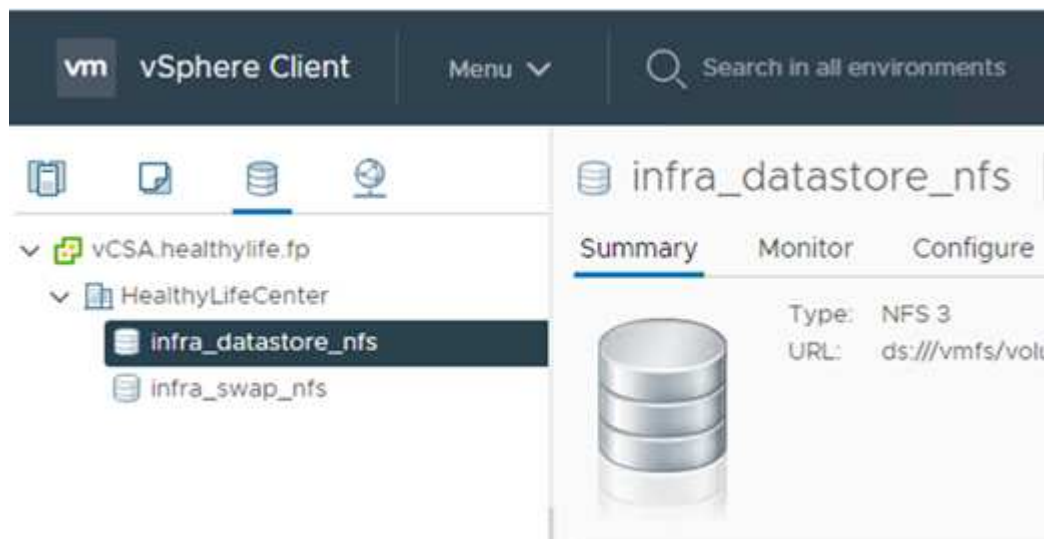
4. Installez vSphere 7.0 avec le démarrage iSCSI.
5. Enregistrez les hôtes ESXi avec vCenter.



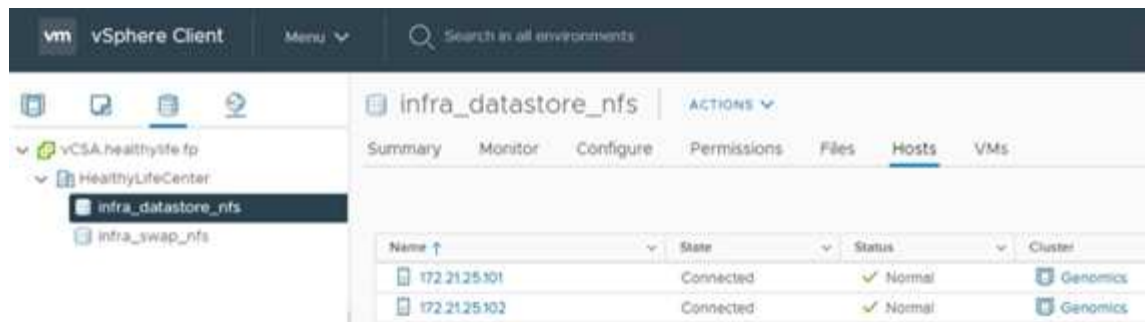
6. Provisionner un datastore NFS infra_datastore_nfs Sur le stockage ONTAP.



7. Ajoutez le datastore au vCenter.



8. À l'aide de vCenter, ajoutez un datastore NFS aux hôtes ESXi.



9. A l'aide de vCenter, créez une machine virtuelle Red Hat Enterprise Linux (RHEL) 8.3 pour exécuter GATK.
10. Un datastore NFS est présenté à la machine virtuelle et monté sur `/mnt/genomics`, Qui est utilisé pour stocker les exécutables GATK, les scripts, les fichiers de carte d'alignement binaire (BAM), les fichiers de référence, les fichiers d'index, les fichiers de dictionnaire et les fichiers de sortie pour les appels de variantes.

```
[root@genomics1 genomics]# df | grep genomics
/dev/sdb                308587328  5699492 287142812   2% /mnt/genomics
[root@genomics1 genomics]#
```

Configuration et exécution de la GATK

Installez les prérequis suivants sur la VM RedHat Enterprise 8.3 Linux :

- Java 8 ou SDK 1.8 ou version ultérieure
- Télécharger GATK 4.2.0.0 depuis le large Institute ["Site GitHub"](#). Les données de séquence du génome sont généralement stockées sous la forme d'une série de colonnes ASCII délimitées par des tabulations. Cependant, l'espace ASCII est trop important pour être stocké. Par conséquent, un nouveau standard évolué appelé fichier BAM (*.bam). Un fichier BAM stocke les données de séquence sous forme compressée, indexée et binaire. Nous ["téléchargé"](#) Un ensemble de fichiers BAM disponibles publiquement pour l'exécution GATK à partir de l' ["domaine public"](#). Nous avons également téléchargé des fichiers d'index (*.bai), des fichiers de dictionnaire (*.dict) et les fichiers de données de référence (*.fasta) du même domaine public.

Après le téléchargement, le kit d'outils GATK dispose d'un fichier jar et d'un ensemble de scripts de support.

- `gatk-package-4.2.0.0-local.jar` exécutable
- `gatk` fichier de script.

Nous avons téléchargé les fichiers BAM et les fichiers d'index, de dictionnaire et de génome de référence correspondants pour une famille composée de fichiers de père, mère et fils *.bam.

Moteur Cromwell

Cromwell est un moteur open-source axé sur les flux de travail scientifiques qui permet la gestion des flux de travail. Le moteur Cromwell peut fonctionner en deux ["modes"](#), Mode serveur ou mode d'exécution d'un seul flux de travail. Le comportement du moteur Cromwell peut être contrôlé à l'aide du ["Fichier de configuration du moteur Cromwell"](#).

- **Le mode serveur.** active ["RESTful"](#) Exécution de flux de travail dans le moteur Cromwell.

- **Mode d'exécution.** le mode d'exécution est le mieux adapté à l'exécution de flux de travail uniques dans Cromwell, "[réf](#)" Pour obtenir un ensemble complet d'options disponibles en mode Run.

Nous utilisons le moteur Cromwell pour exécuter les flux de travail et les pipelines à grande échelle. Le moteur Cromwell est convivial "[langage de description de workflow](#)" Langage de script basé sur WDL. Cromwell prend également en charge une deuxième norme de script de flux de travail appelée langage de flux de travail commun (CWL). Tout au long de ce rapport technique, nous avons utilisé WDL. À l'origine, le WDL a été développé par le large institut pour les pipelines d'analyse du génome. L'utilisation des workflows WDL peut être mise en œuvre à l'aide de plusieurs stratégies, notamment :

- **Chaînage linéaire.** comme le nom l'indique, la sortie de la tâche #1 est envoyée à la tâche #2 comme entrée.
- **Multi-in/out.** Ceci est similaire à la chaînage linéaire dans le fait que chaque tâche peut avoir plusieurs sorties envoyées en entrée aux tâches suivantes.
- **Scatter-rassembler.** il s'agit de l'une des stratégies d'intégration des applications d'entreprise les plus puissantes disponibles, surtout lorsqu'elle est utilisée dans une architecture basée sur des événements. Chaque tâche s'exécute de façon dissociée, et le résultat de chaque tâche est consolidé dans le résultat final.

Il existe trois étapes lorsque WDL est utilisé pour faire fonctionner GATK en mode autonome :

1. Valider la syntaxe à l'aide de `womtool.jar`.

```
[root@genomics1 ~]# java -jar womtool.jar validate ghplo.wdl
```

2. Générer des entrées JSON.

```
[root@genomics1 ~]# java -jar womtool.jar inputs ghplo.wdl > ghplo.json
```

3. Exécutez le flux de travail à l'aide du moteur Cromwell et `Cromwell.jar`.

```
[root@genomics1 ~]# java -jar cromwell.jar run ghplo.wdl --inputs ghplo.json
```

Le GATK peut être exécuté à l'aide de plusieurs méthodes; ce document explore trois de ces méthodes.

Exécution de GATK à l'aide du fichier jar

Examinons l'exécution d'un pipeline d'appel variante unique à l'aide de l'appelant variant en haplotype.

```
[root@genomics1 ~]# java -Dsamjdk.use_async_io_read_samtools=false \
-Dsamjdk.use_async_io_write_samtools=true \
-Dsamjdk.use_async_io_write_tribble=false \
-Dsamjdk.compression_level=2 \
-jar /mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-local.jar \
HaplotypeCaller \
--input /mnt/genomics/GATK/TEST\ DATA/bam/workshop_1906_2-
germline_bams_father.bam \
--output workshop_1906_2-germline_bams_father.validation.vcf \
--reference /mnt/genomics/GATK/TEST\ DATA/ref/workshop_1906_2-
germline_ref_ref.fasta
```

Dans cette méthode d'exécution, nous utilisons le fichier JAR d'exécution locale GATK, une seule commande Java pour appeler le fichier jar, et nous transmettons plusieurs paramètres à la commande.

1. Ce paramètre indique que nous invoons le HaplotypeCaller variante du pipeline appelant.
2. -- input Spécifie le fichier BAM en entrée.
3. --output spécifie le fichier de sortie de variante dans le format d'appel de variante (*.vcf) ("réf").
4. Avec le --reference paramètre, nous sommes en passe de passer un génome de référence.

Une fois exécuté, les détails de sortie se trouvent dans la section ["Sortie pour l'exécution de GATK à l'aide du fichier jar."](#)

Exécution de GATK à l'aide du script ./gatk

La trousse à outils GATK peut être exécutée à l'aide de l' ./gatk script. Examinons la commande suivante :

```
[root@genomics1 execution]# ./gatk \
--java-options "-Xmx4G" \
HaplotypeCaller \
-I /mnt/genomics/GATK/TEST\ DATA/bam/workshop_1906_2-
germline_bams_father.bam \
-R /mnt/genomics/GATK/TEST\ DATA/ref/workshop_1906_2-
germline_ref_ref.fasta \
-O /mnt/genomics/GATK/TEST\ DATA/variants.vcf
```

Nous transmettons plusieurs paramètres à la commande.

- Ce paramètre indique que nous invoons le HaplotypeCaller variante du pipeline appelant.
- -I Spécifie le fichier BAM en entrée.
- -O spécifie le fichier de sortie de variante dans le format d'appel de variante (*.vcf) ("réf").
- Avec le -R paramètre, nous sommes en passe de passer un génome de référence.

Une fois exécuté, les détails de sortie se trouvent dans la section

Exécution de la GATK à l'aide du moteur Cromwell

Nous utilisons le moteur Cromwell pour gérer l'exécution GATK. Examinons la ligne de commande et ses paramètres.

```
[root@genomics1 genomics]# java -jar cromwell-65.jar \  
run /mnt/genomics/GATK/seq/ghplo.wdl \  
--inputs /mnt/genomics/GATK/seq/ghplo.json
```

Ici, nous invoquons la commande Java en passant le `-jar` paramètre pour indiquer que nous avons l'intention d'exécuter un fichier jar, par exemple, `Cromwell-65.jar`. Le paramètre suivant a réussi (`run`) Indique que le moteur Cromwell fonctionne en mode RUN, l'autre option possible est le mode serveur. Le paramètre suivant est `*.wdl` Que le mode Exécuter doit utiliser pour exécuter les pipelines. Le paramètre suivant est l'ensemble des paramètres d'entrée des flux de travail exécutés.

Voici le contenu du `ghplo.wdl` type de fichier :

```
[root@genomics1 seq]# cat ghplo.wdl  
workflow helloHaplotypeCaller {  
  call haplotypeCaller  
}  
task haplotypeCaller {  
  File GATK  
  File RefFasta  
  File RefIndex  
  File RefDict  
  String sampleName  
  File inputBAM  
  File bamIndex  
  command {  
    java -jar ${GATK} \  
      HaplotypeCaller \  
      -R ${RefFasta} \  
      -I ${inputBAM} \  
      -O ${sampleName}.raw.indels.snps.vcf  
  }  
  output {  
    File rawVCF = "${sampleName}.raw.indels.snps.vcf"  
  }  
}
```

Voici le fichier JSON correspondant avec les entrées du moteur Cromwell.

```
[root@genomics1 seq]# cat ghplo.json
{
  "helloHaplotypeCaller.haplotypeCaller.GATK": "/mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-local.jar",
  "helloHaplotypeCaller.haplotypeCaller.RefFasta": "/mnt/genomics/GATK/TEST DATA/ref/workshop_1906_2-germline_ref_ref.fasta",
  "helloHaplotypeCaller.haplotypeCaller.RefIndex": "/mnt/genomics/GATK/TEST DATA/ref/workshop_1906_2-germline_ref_ref.fasta.fai",
  "helloHaplotypeCaller.haplotypeCaller.RefDict": "/mnt/genomics/GATK/TEST DATA/ref/workshop_1906_2-germline_ref_ref.dict",
  "helloHaplotypeCaller.haplotypeCaller.sampleName": "fatherbam",
  "helloHaplotypeCaller.haplotypeCaller.inputBAM": "/mnt/genomics/GATK/TEST DATA/bam/workshop_1906_2-germline_bams_father.bam",
  "helloHaplotypeCaller.haplotypeCaller.bamIndex": "/mnt/genomics/GATK/TEST DATA/bam/workshop_1906_2-germline_bams_father.bai"
}
[root@genomics1 seq]#
```

Veuillez noter que Cromwell utilise une base de données in-memory pour l'exécution. Une fois exécuté, le journal de sortie est visible dans la section ["Sortie pour l'exécution de la GATK à l'aide du moteur Cromwell."](#)

Pour un ensemble complet d'étapes sur la façon d'exécuter GATK, voir ["Documentation GATK"](#).

["Suivant : sortie pour l'exécution de GATK à l'aide du fichier jar."](#)

Sortie pour l'exécution de GATK à l'aide du fichier jar

["Précédent : génomique - Configuration et exécution de la GATK."](#)

L'exécution de GATK à l'aide du fichier jar a produit la sortie d'échantillon suivante.

```
[root@genomics1 execution]# java -Dsamjdk.use_async_io_read_samtools=false \
-Dsamjdk.use_async_io_write_samtools=true \
-Dsamjdk.use_async_io_write_tribble=false \
-Dsamjdk.compression_level=2 \
-jar /mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-local.jar \
HaplotypeCaller \
--input /mnt/genomics/GATK/TEST\ DATA/bam/workshop_1906_2-germline_bams_father.bam \
--output workshop_1906_2-germline_bams_father.validation.vcf \
--reference /mnt/genomics/GATK/TEST\ DATA/ref/workshop_1906_2-germline_ref_ref.fasta \
22:52:58.430 INFO NativeLibraryLoader - Loading libgkl_compression.so
from jar:file:/mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
```

Aug 17, 2021 10:52:58 PM

shaded.cloud_nio.com.google.auth.oauth2.ComputeEngineCredentials
runningOnComputeEngine

INFO: Failed to detect whether we are running on Google Compute Engine.

22:52:58.541 INFO HaplotypeCaller -

22:52:58.542 INFO HaplotypeCaller - The Genome Analysis Toolkit (GATK)
v4.2.0.0

22:52:58.542 INFO HaplotypeCaller - For support and documentation go to
<https://software.broadinstitute.org/gatk/>

22:52:58.542 INFO HaplotypeCaller - Executing as

root@genomics1.healthylife.fp on Linux v4.18.0-305.3.1.el8_4.x86_64 amd64

22:52:58.542 INFO HaplotypeCaller - Java runtime: OpenJDK 64-Bit Server
VM v1.8.0_302-b08

22:52:58.542 INFO HaplotypeCaller - Start Date/Time: August 17, 2021
10:52:58 PM EDT

22:52:58.542 INFO HaplotypeCaller -

22:52:58.542 INFO HaplotypeCaller -

22:52:58.542 INFO HaplotypeCaller - HTSJDK Version: 2.24.0

22:52:58.542 INFO HaplotypeCaller - Picard Version: 2.25.0

22:52:58.542 INFO HaplotypeCaller - Built for Spark Version: 2.4.5

22:52:58.542 INFO HaplotypeCaller - HTSJDK Defaults.COMPRESSION_LEVEL : 2

22:52:58.543 INFO HaplotypeCaller - HTSJDK

Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false

22:52:58.543 INFO HaplotypeCaller - HTSJDK

Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true

22:52:58.543 INFO HaplotypeCaller - HTSJDK

Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false

22:52:58.543 INFO HaplotypeCaller - Deflater: IntelDeflater

22:52:58.543 INFO HaplotypeCaller - Inflater: IntelInflater

22:52:58.543 INFO HaplotypeCaller - GCS max retries/reopens: 20

22:52:58.543 INFO HaplotypeCaller - Requester pays: disabled

22:52:58.543 INFO HaplotypeCaller - Initializing engine

22:52:58.804 INFO HaplotypeCaller - Done initializing engine

22:52:58.809 INFO HaplotypeCallerEngine - Disabling physical phasing,
which is supported only for reference-model confidence output

22:52:58.820 INFO NativeLibraryLoader - Loading libgkl_utils.so from
jar:file:/mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-
local.jar!/com/intel/gkl/native/libgkl_utils.so

22:52:58.821 INFO NativeLibraryLoader - Loading libgkl_pairhmm_omp.so
from jar:file:/mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-
local.jar!/com/intel/gkl/native/libgkl_pairhmm_omp.so

22:52:58.854 INFO IntelPairHmm - Using CPU-supported AVX-512 instructions

22:52:58.854 INFO IntelPairHmm - Flush-to-zero (FTZ) is enabled when

```

running PairHMM
22:52:58.854 INFO   IntelPairHmm - Available threads: 16
22:52:58.854 INFO   IntelPairHmm - Requested threads: 4
22:52:58.854 INFO   PairHMM - Using the OpenMP multi-threaded AVX-
accelerated native PairHMM implementation
22:52:58.872 INFO   ProgressMeter - Starting traversal
22:52:58.873 INFO   ProgressMeter -           Current Locus   Elapsed Minutes
Regions Processed   Regions/Minute
22:53:00.733 WARN   InbreedingCoeff - InbreedingCoeff will not be
calculated at position 20:9999900 and possibly subsequent; at least 10
samples must have called genotypes
22:53:08.873 INFO   ProgressMeter -           20:17538652           0.2
58900              353400.0
22:53:17.681 INFO   HaplotypeCaller - 405 read(s) filtered by:
MappingQualityReadFilter
0 read(s) filtered by: MappingQualityAvailableReadFilter
0 read(s) filtered by: MappedReadFilter
0 read(s) filtered by: NotSecondaryAlignmentReadFilter
6628 read(s) filtered by: NotDuplicateReadFilter
0 read(s) filtered by: PassesVendorQualityCheckReadFilter
0 read(s) filtered by: NonZeroReferenceLengthAlignmentReadFilter
0 read(s) filtered by: GoodCigarReadFilter
0 read(s) filtered by: WellformedReadFilter
7033 total reads filtered
22:53:17.681 INFO   ProgressMeter -           20:63024652           0.3
210522             671592.9
22:53:17.681 INFO   ProgressMeter - Traversal complete. Processed 210522
total regions in 0.3 minutes.
22:53:17.687 INFO   VectorLoglessPairHMM - Time spent in setup for JNI call
: 0.010347438
22:53:17.687 INFO   PairHMM - Total compute time in PairHMM
computeLogLikelihoods() : 0.259172573
22:53:17.687 INFO   SmithWatermanAligner - Total compute time in java
Smith-Waterman : 1.27 sec
22:53:17.687 INFO   HaplotypeCaller - Shutting down engine
[August 17, 2021 10:53:17 PM EDT]
org.broadinstitute.hellbender.tools.walkers.haplotypecaller.HaplotypeCalle
r done. Elapsed time: 0.32 minutes.
Runtime.totalMemory()=5561122816
[root@genomics1 execution]#

```

Notez que le fichier de sortie se trouve à l'emplacement spécifié après l'exécution.

Sortie pour l'exécution de GATK à l'aide du script ./gatk

"Précédent : sortie pour l'exécution de GATK à l'aide du fichier jar."

L'exécution de GATK à l'aide de l' ./gatk le script a produit l'exemple de sortie suivant.

```
[root@genomics1 gatk-4.2.0.0]# ./gatk --java-options "-Xmx4G" \
HaplotypeCaller \
-I /mnt/genomics/GATK/TEST\ DATA/bam/workshop_1906_2-
germline_bams_father.bam \
-R /mnt/genomics/GATK/TEST\ DATA/ref/workshop_1906_2-
germline_ref_ref.fasta \
-O /mnt/genomics/GATK/TEST\ DATA/variants.vcf
Using GATK jar /mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-
local.jar
Running:
    java -Dsamjdk.use_async_io_read_samtools=false
-Dsamjdk.use_async_io_write_samtools=true
-Dsamjdk.use_async_io_write_tribble=false -Dsamjdk.compression_level=2
-Xmx4G -jar /mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-local.jar
HaplotypeCaller -I /mnt/genomics/GATK/TEST DATA/bam/workshop_1906_2-
germline_bams_father.bam -R /mnt/genomics/GATK/TEST
DATA/ref/workshop_1906_2-germline_ref_ref.fasta -O /mnt/genomics/GATK/TEST
DATA/variants.vcf
23:29:45.553 INFO  NativeLibraryLoader - Loading libgkl_compression.so
from jar:file:/mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-
local.jar!/com/intel/gkl/native/libgkl_compression.so
Aug 17, 2021 11:29:45 PM
shaded.cloud_nio.com.google.auth.oauth2.ComputeEngineCredentials
runningOnComputeEngine
INFO: Failed to detect whether we are running on Google Compute Engine.
23:29:45.686 INFO  HaplotypeCaller -
-----
23:29:45.686 INFO  HaplotypeCaller - The Genome Analysis Toolkit (GATK)
v4.2.0.0
23:29:45.686 INFO  HaplotypeCaller - For support and documentation go to
https://software.broadinstitute.org/gatk/
23:29:45.687 INFO  HaplotypeCaller - Executing as
root@genomics1.healthyliife.fp on Linux v4.18.0-305.3.1.el8_4.x86_64 amd64
23:29:45.687 INFO  HaplotypeCaller - Java runtime: OpenJDK 64-Bit Server
VM v11.0.12+7-LTS
23:29:45.687 INFO  HaplotypeCaller - Start Date/Time: August 17, 2021 at
11:29:45 PM EDT
23:29:45.687 INFO  HaplotypeCaller -
-----
23:29:45.687 INFO  HaplotypeCaller -
```

```

-----
23:29:45.687 INFO HaplotypeCaller - HTSJDK Version: 2.24.0
23:29:45.687 INFO HaplotypeCaller - Picard Version: 2.25.0
23:29:45.687 INFO HaplotypeCaller - Built for Spark Version: 2.4.5
23:29:45.688 INFO HaplotypeCaller - HTSJDK Defaults.COMPRESSION_LEVEL : 2
23:29:45.688 INFO HaplotypeCaller - HTSJDK
Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
23:29:45.688 INFO HaplotypeCaller - HTSJDK
Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
23:29:45.688 INFO HaplotypeCaller - HTSJDK
Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
23:29:45.688 INFO HaplotypeCaller - Deflater: IntelDeflater
23:29:45.688 INFO HaplotypeCaller - Inflater: IntelInflater
23:29:45.688 INFO HaplotypeCaller - GCS max retries/reopens: 20
23:29:45.688 INFO HaplotypeCaller - Requester pays: disabled
23:29:45.688 INFO HaplotypeCaller - Initializing engine
23:29:45.804 INFO HaplotypeCaller - Done initializing engine
23:29:45.809 INFO HaplotypeCallerEngine - Disabling physical phasing,
which is supported only for reference-model confidence output
23:29:45.818 INFO NativeLibraryLoader - Loading libgkl_utils.so from
jar:file:/mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-
local.jar!/com/intel/gkl/native/libgkl_utils.so
23:29:45.819 INFO NativeLibraryLoader - Loading libgkl_pairhmm_omp.so
from jar:file:/mnt/genomics/GATK/gatk-4.2.0.0/gatk-package-4.2.0.0-
local.jar!/com/intel/gkl/native/libgkl_pairhmm_omp.so
23:29:45.852 INFO IntelPairHmm - Using CPU-supported AVX-512 instructions
23:29:45.852 INFO IntelPairHmm - Flush-to-zero (FTZ) is enabled when
running PairHMM
23:29:45.852 INFO IntelPairHmm - Available threads: 16
23:29:45.852 INFO IntelPairHmm - Requested threads: 4
23:29:45.852 INFO PairHMM - Using the OpenMP multi-threaded AVX-
accelerated native PairHMM implementation
23:29:45.868 INFO ProgressMeter - Starting traversal
23:29:45.868 INFO ProgressMeter -          Current Locus  Elapsed Minutes
Regions Processed  Regions/Minute
23:29:47.772 WARN InbreedingCoeff - InbreedingCoeff will not be
calculated at position 20:9999900 and possibly subsequent; at least 10
samples must have called genotypes
23:29:55.868 INFO ProgressMeter -          20:18885652          0.2
63390          380340.0
23:30:04.389 INFO HaplotypeCaller - 405 read(s) filtered by:
MappingQualityReadFilter
0 read(s) filtered by: MappingQualityAvailableReadFilter
0 read(s) filtered by: MappedReadFilter
0 read(s) filtered by: NotSecondaryAlignmentReadFilter
6628 read(s) filtered by: NotDuplicateReadFilter

```

```

0 read(s) filtered by: PassesVendorQualityCheckReadFilter
0 read(s) filtered by: NonZeroReferenceLengthAlignmentReadFilter
0 read(s) filtered by: GoodCigarReadFilter
0 read(s) filtered by: WellformedReadFilter
7033 total reads filtered
23:30:04.389 INFO ProgressMeter - 20:63024652 0.3
210522 681999.9
23:30:04.389 INFO ProgressMeter - Traversal complete. Processed 210522
total regions in 0.3 minutes.
23:30:04.395 INFO VectorLoglessPairHMM - Time spent in setup for JNI call
: 0.012129203000000002
23:30:04.395 INFO PairHMM - Total compute time in PairHMM
computeLogLikelihoods() : 0.267345217
23:30:04.395 INFO SmithWatermanAligner - Total compute time in java
Smith-Waterman : 1.23 sec
23:30:04.395 INFO HaplotypeCaller - Shutting down engine
[August 17, 2021 at 11:30:04 PM EDT]
org.broadinstitute.hellbender.tools.walkers.haplotypecaller.HaplotypeCalle
r done. Elapsed time: 0.31 minutes.
Runtime.totalMemory()=2111832064
[root@genomics1 gatk-4.2.0.0]#

```

Notez que le fichier de sortie se trouve à l'emplacement spécifié après l'exécution.

"Suivant : sortie pour l'exécution de la GATK à l'aide du moteur Cromwell."

Sortie pour l'exécution de la GATK à l'aide du moteur Cromwell

L'exécution de la GATK à l'aide du moteur Cromwell a produit la sortie d'échantillon suivante.

```

[root@genomics1 genomics]# java -jar cromwell-65.jar run
/mnt/genomics/GATK/seq/ghplo.wdl --inputs
/mnt/genomics/GATK/seq/ghplo.json
[2021-08-18 17:10:50,78] [info] Running with database db.url =
jdbc:hsqldb:mem:856a1f0d-9a0d-42e5-9199-
5e6c1d0f72dd;shutdown=false;hsqldb.tx=mvcc
[2021-08-18 17:10:57,74] [info] Running migration
RenameWorkflowOptionsInMetadata with a read batch size of 100000 and a
write batch size of 100000
[2021-08-18 17:10:57,75] [info] [RenameWorkflowOptionsInMetadata] 100%
[2021-08-18 17:10:57,83] [info] Running with database db.url =
jdbc:hsqldb:mem:6afe0252-2dc9-4e57-8674-
ce63c67aa142;shutdown=false;hsqldb.tx=mvcc

```



```

[2021-08-18 17:10:58,17] [info] Slf4jLogger started
[2021-08-18 17:10:58,33] [info] Workflow heartbeat configuration:
{
  "cromwellId" : "cromid-41b7e30",
  "heartbeatInterval" : "2 minutes",
  "ttl" : "10 minutes",
  "failureShutdownDuration" : "5 minutes",
  "writeBatchSize" : 10000,
  "writeThreshold" : 10000
}
[2021-08-18 17:10:58,38] [info] Metadata summary refreshing every 1
second.
[2021-08-18 17:10:58,38] [info] No metadata archiver defined in config
[2021-08-18 17:10:58,38] [info] No metadata deleter defined in config
[2021-08-18 17:10:58,40] [info] KvWriteActor configured to flush with
batch size 200 and process rate 5 seconds.
[2021-08-18 17:10:58,40] [info] WriteMetadataActor configured to flush
with batch size 200 and process rate 5 seconds.
[2021-08-18 17:10:58,44] [info] CallCacheWriteActor configured to flush
with batch size 100 and process rate 3 seconds.
[2021-08-18 17:10:58,44] [warn] 'docker.hash-lookup.gcr-api-queries-per-
100-seconds' is being deprecated, use 'docker.hash-lookup.gcr.throttle'
instead (see reference.conf)
[2021-08-18 17:10:58,54] [info] JobExecutionTokenDispenser - Distribution
rate: 50 per 1 seconds.
[2021-08-18 17:10:58,58] [info] SingleWorkflowRunnerActor: Version 65
[2021-08-18 17:10:58,58] [info] SingleWorkflowRunnerActor: Submitting
workflow
[2021-08-18 17:10:58,64] [info] Unspecified type (Unspecified version)
workflow 3e246147-b1a9-41dc-8679-319f81b7701e submitted
[2021-08-18 17:10:58,66] [info] SingleWorkflowRunnerActor: Workflow
submitted 3e246147-b1a9-41dc-8679-319f81b7701e
[2021-08-18 17:10:58,66] [info] 1 new workflows fetched by cromid-41b7e30:
3e246147-b1a9-41dc-8679-319f81b7701e
[2021-08-18 17:10:58,67] [info] WorkflowManagerActor: Starting workflow
3e246147-b1a9-41dc-8679-319f81b7701e
[2021-08-18 17:10:58,68] [info] WorkflowManagerActor: Successfully started
WorkflowActor-3e246147-b1a9-41dc-8679-319f81b7701e
[2021-08-18 17:10:58,68] [info] Retrieved 1 workflows from the
WorkflowStoreActor
[2021-08-18 17:10:58,70] [info] WorkflowStoreHeartbeatWriteActor
configured to flush with batch size 10000 and process rate 2 minutes.
[2021-08-18 17:10:58,76] [info] MaterializeWorkflowDescriptorActor
[3e246147]: Parsing workflow as WDL draft-2
[2021-08-18 17:10:59,34] [info] MaterializeWorkflowDescriptorActor
[3e246147]: Call-to-Backend assignments:

```

```

helloHaplotypeCaller.haplotypeCaller -> Local
[2021-08-18 17:11:00,54] [info] WorkflowExecutionActor-3e246147-b1a9-41dc-8679-319f81b7701e [3e246147]: Starting
helloHaplotypeCaller.haplotypeCaller
[2021-08-18 17:11:01,56] [info] Assigned new job execution tokens to the following groups: 3e246147: 1
[2021-08-18 17:11:01,70] [info] BackgroundConfigAsyncJobExecutionActor [3e246147helloHaplotypeCaller.haplotypeCaller:NA:1]: java -jar /mnt/genomics/cromwell-executions/helloHaplotypeCaller/3e246147-b1a9-41dc-8679-319f81b7701e/call-haplotypeCaller/inputs/-179397211/gatk-package-4.2.0.0-local.jar \
    HaplotypeCaller \
    -R /mnt/genomics/cromwell-executions/helloHaplotypeCaller/3e246147-b1a9-41dc-8679-319f81b7701e/call-haplotypeCaller/inputs/604632695/workshop_1906_2-germline_ref_ref.fasta \
    -I /mnt/genomics/cromwell-executions/helloHaplotypeCaller/3e246147-b1a9-41dc-8679-319f81b7701e/call-haplotypeCaller/inputs/604617202/workshop_1906_2-germline_bams_father.bam \
    -O fatherbam.raw.indels.snps.vcf
[2021-08-18 17:11:01,72] [info] BackgroundConfigAsyncJobExecutionActor [3e246147helloHaplotypeCaller.haplotypeCaller:NA:1]: executing: /bin/bash /mnt/genomics/cromwell-executions/helloHaplotypeCaller/3e246147-b1a9-41dc-8679-319f81b7701e/call-haplotypeCaller/execution/script
[2021-08-18 17:11:03,49] [info] BackgroundConfigAsyncJobExecutionActor [3e246147helloHaplotypeCaller.haplotypeCaller:NA:1]: job id: 26867
[2021-08-18 17:11:03,53] [info] BackgroundConfigAsyncJobExecutionActor [3e246147helloHaplotypeCaller.haplotypeCaller:NA:1]: Status change from - to WaitingForReturnCode
[2021-08-18 17:11:03,54] [info] Not triggering log of token queue status. Effective log interval = None
[2021-08-18 17:11:23,65] [info] BackgroundConfigAsyncJobExecutionActor [3e246147helloHaplotypeCaller.haplotypeCaller:NA:1]: Status change from WaitingForReturnCode to Done
[2021-08-18 17:11:25,04] [info] WorkflowExecutionActor-3e246147-b1a9-41dc-8679-319f81b7701e [3e246147]: Workflow helloHaplotypeCaller complete.
Final Outputs:
{
  "helloHaplotypeCaller.haplotypeCaller.rawVCF": "/mnt/genomics/cromwell-executions/helloHaplotypeCaller/3e246147-b1a9-41dc-8679-319f81b7701e/call-haplotypeCaller/execution/fatherbam.raw.indels.snps.vcf"
}
[2021-08-18 17:11:28,43] [info] WorkflowManagerActor: Workflow actor for 3e246147-b1a9-41dc-8679-319f81b7701e completed with status 'Succeeded'. The workflow will be removed from the workflow store.
[2021-08-18 17:11:32,24] [info] SingleWorkflowRunnerActor workflow

```

```

finished with status 'Succeeded'.
{
  "outputs": {
    "helloHaplotypeCaller.haplotypeCaller.rawVCF":
"/mnt/genomics/cromwell-executions/helloHaplotypeCaller/3e246147-b1a9-
41dc-8679-319f81b7701e/call-
haplotypeCaller/execution/fatherbam.raw.indels.snps.vcf"
  },
  "id": "3e246147-b1a9-41dc-8679-319f81b7701e"
}
[2021-08-18 17:11:33,45] [info] Workflow polling stopped
[2021-08-18 17:11:33,46] [info] 0 workflows released by cromid-41b7e30
[2021-08-18 17:11:33,46] [info] Shutting down WorkflowStoreActor - Timeout
= 5 seconds
[2021-08-18 17:11:33,46] [info] Shutting down WorkflowLogCopyRouter -
Timeout = 5 seconds
[2021-08-18 17:11:33,46] [info] Shutting down JobExecutionTokenDispenser -
Timeout = 5 seconds
[2021-08-18 17:11:33,46] [info] Aborting all running workflows.
[2021-08-18 17:11:33,46] [info] JobExecutionTokenDispenser stopped
[2021-08-18 17:11:33,46] [info] WorkflowStoreActor stopped
[2021-08-18 17:11:33,47] [info] WorkflowLogCopyRouter stopped
[2021-08-18 17:11:33,47] [info] Shutting down WorkflowManagerActor -
Timeout = 3600 seconds
[2021-08-18 17:11:33,47] [info] WorkflowManagerActor: All workflows
finished
[2021-08-18 17:11:33,47] [info] WorkflowManagerActor stopped
[2021-08-18 17:11:33,64] [info] Connection pools shut down
[2021-08-18 17:11:33,64] [info] Shutting down SubWorkflowStoreActor -
Timeout = 1800 seconds
[2021-08-18 17:11:33,64] [info] Shutting down JobStoreActor - Timeout =
1800 seconds
[2021-08-18 17:11:33,64] [info] Shutting down CallCacheWriteActor -
Timeout = 1800 seconds
[2021-08-18 17:11:33,64] [info] SubWorkflowStoreActor stopped
[2021-08-18 17:11:33,64] [info] Shutting down ServiceRegistryActor -
Timeout = 1800 seconds
[2021-08-18 17:11:33,64] [info] Shutting down DockerHashActor - Timeout =
1800 seconds
[2021-08-18 17:11:33,64] [info] Shutting down IoProxy - Timeout = 1800
seconds
[2021-08-18 17:11:33,64] [info] CallCacheWriteActor Shutting down: 0
queued messages to process
[2021-08-18 17:11:33,64] [info] JobStoreActor stopped
[2021-08-18 17:11:33,64] [info] CallCacheWriteActor stopped
[2021-08-18 17:11:33,64] [info] KvWriteActor Shutting down: 0 queued

```

```

messages to process
[2021-08-18 17:11:33,64] [info] IoProxy stopped
[2021-08-18 17:11:33,64] [info] WriteMetadataActor Shutting down: 0 queued
messages to process
[2021-08-18 17:11:33,65] [info] ServiceRegistryActor stopped
[2021-08-18 17:11:33,65] [info] DockerHashActor stopped
[2021-08-18 17:11:33,67] [info] Database closed
[2021-08-18 17:11:33,67] [info] Stream materializer shut down
[2021-08-18 17:11:33,67] [info] WDL HTTP import resolver closed
[root@genomics1 genomics]#

```

"Suivant : configuration du GPU."

Configuration des GPU

"Précédent : sortie pour l'exécution de la GATK à l'aide du moteur Cromwell."

Au moment de la publication, l'outil GATK n'a pas de prise en charge native de l'exécution basée sur GPU sur site. La configuration et les conseils suivants permettent aux lecteurs de comprendre à quel point il est simple d'utiliser FlexPod avec un GPU NVIDIA Tesla P6 monté à l'arrière au moyen d'une carte mezzanine PCIe pour le GATK.

Nous avons utilisé le CVD suivant pour l'architecture de référence et le guide des meilleures pratiques pour configurer l'environnement FlexPod afin que nous puissions exécuter des applications utilisant des GPU.

- ["FlexPod Datacenter pour l'IA/ML avec Cisco UCS 480 ML pour le deep learning"](#)

Voici un ensemble de messages clés à retenir lors de cette configuration :

1. Nous avons utilisé un processeur graphique PCIe NVIDIA Tesla P6 dans un slot mezzanine des serveurs UCS B200 M5.

Equipment / Chassis / Chassis 1 / Servers / Server 1

<

General

Inventory

Virtual Machines

Installed Firmware

CIMC Sessions

SEL Logs

VIF Paths

Health >

<

Motherboard

CIMC

CPUs

GPUs

Memory

Adapters

HBAs

NICs

iSCSI vNICs

Security >

Advanced Filter

Export

Print

Name

ID

Model

Serial

Mode

Graphics Card 2

2

UCSB-GPU-P6-R

FCH212373V7

Compute

Equipment / Chassis / Chassis 1 / Servers / Server 2

< General **Inventory** Virtual Machines Installed Firmware CIMC Sessions SEL Logs VIF Paths Health >

< Motherboard CIMC CPUs **GPUs** Memory Adapters HBAs NICs iSCSI vNICs Security >

Advanced Filter Export Print

Name	ID	Model	Serial	Mode
Graphics Card 2	2	UCSB-GPU-P6-R	FCH212373Y1	Compute

2. Pour cette configuration, nous nous sommes inscrits sur le portail partenaires NVIDIA et avons obtenu une licence d'évaluation (également appelée droit) pour pouvoir utiliser les GPU en mode de calcul.
3. Nous avons téléchargé le logiciel NVIDIA vGPU requis du site Web dédié aux partenaires NVIDIA.
4. Nous avons téléchargé le droit *.bin Fichier du site Web partenaire de NVIDIA.
5. Nous avons installé un serveur de licences NVIDIA vGPU et ajouté les droits au serveur de licences à l'aide de *.bin Fichier téléchargé du site partenaire NVIDIA.
6. Veuillez à choisir la version correcte du logiciel NVIDIA vGPU pour votre déploiement sur le portail partenaires NVIDIA. Pour cette configuration, nous avons utilisé la version 460.73.02 du pilote.
7. Cette commande installe le **"NVIDIA vGPU Manager"** Dans ESXi.

```
[root@localhost:~] esxcli software vib install -v
/vmfs/volumes/infra_datastore_nfs/nvidia/vib/NVIDIA_bootbank_NVIDIA-
VMware_ESXi_7.0_Host_Driver_460.73.02-1OEM.700.0.0.15525992.vib
Installation Result
Message: Operation finished successfully.
Reboot Required: false
VIBs Installed: NVIDIA_bootbank_NVIDIA-
VMware_ESXi_7.0_Host_Driver_460.73.02-1OEM.700.0.0.15525992
VIBs Removed:
VIBs Skipped:
```

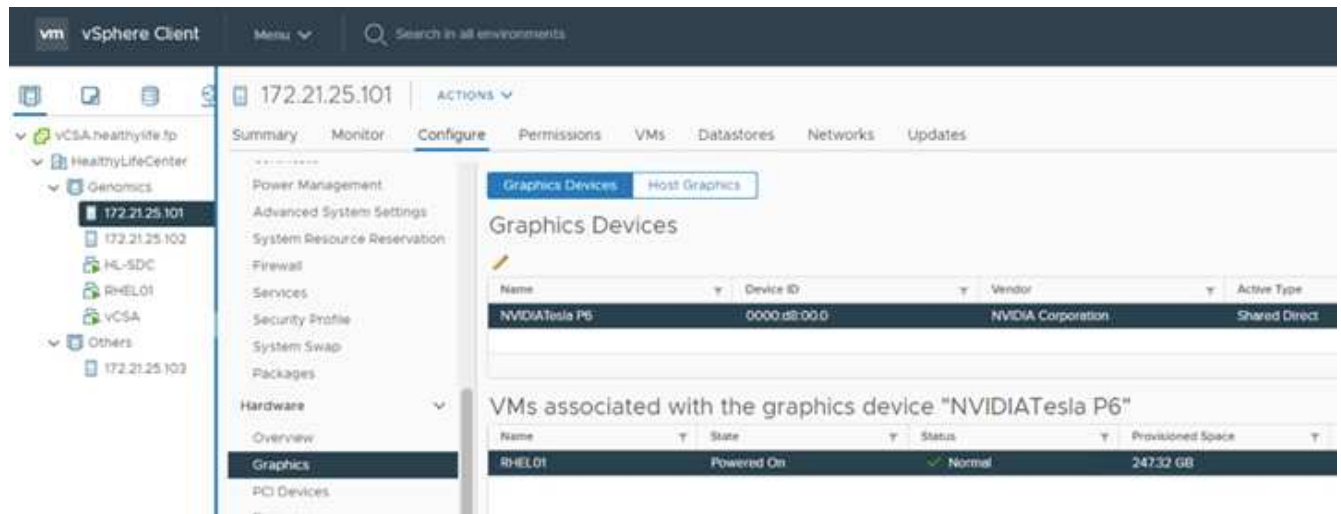
8. Après le redémarrage du serveur ESXi, exécutez la commande suivante pour valider l'installation et vérifier l'état des GPU.

```

[root@localhost:~] nvidia-smi
Wed Aug 18 21:37:19 2021
+-----+
+-----+
| NVIDIA-SMI 460.73.02      Driver Version: 460.73.02      CUDA Version: N/A
|
|-----+-----+
+-----+
| GPU   Name               Persistence-M| Bus-Id        Disp.A | Volatile
Uncorr. ECC |
| Fan   Temp   Perf   Pwr:Usage/Cap|      Memory-Usage | GPU-Util
Compute M. |
|
|-----+-----+
MIG M. |
|=====+=====+=====+
=====|
|    0   Tesla P6             On      | 00000000:D8:00.0 Off |
0 |
| N/A    35C    P8      9W /  90W |  15208MiB /  15359MiB |      0%
Default |
|
|-----+-----+
N/A |
+-----+-----+
+-----+
+-----+
+-----+
+-----+
| Processes:
|
| GPU   GI    CI          PID    Type    Process name          GPU
Memory |
|      ID    ID                 |          Usage
|
|=====+=====+=====+
=====|
|    0   N/A   N/A     2812553      C+G     RHEL01
15168MiB |
+-----+-----+
+-----+
[root@localhost:~]

```

9. À l'aide de vCenter, "[configurer](#)" Les paramètres du périphérique graphique sur « Shared Direct ».



10. Assurez-vous que le démarrage sécurisé est désactivé pour la machine virtuelle RedHat.
11. Assurez-vous que le micrologiciel des options de démarrage VM est défini sur EFI ("réf").

Edit Settings
RHEL01

Virtual Hardware
VM Options

> General Options	VM Name: RHEL01
> VMware Remote Console Options	☐ Lock the guest operating system when the last remote user disconnects
> Encryption	Expand for encryption settings
> Power management	Expand for power management settings
> VMware Tools	Expand for VMware Tools settings
> Boot Options	
Firmware	EFI (recommended) ▼
Secure Boot	☐ Enabled
Boot Delay	When powering on or resetting, delay boot order by 0 milliseconds
Force EFI setup	☐ During the next boot, force entry into the EFI setup screen
Failed Boot Recovery	☐ If the VM fails to find boot device, automatically retry after 10 seconds
> Advanced	Expand for advanced settings
> Fibre Channel NPIV	Expand for Fibre Channel NPIV settings

CANCEL
OK

12. Assurez-vous que les PARAMÈTRES suivants sont ajoutés à VM Options Advanced Edit Configuration. La valeur du `pciPassthru.64bitMMIOSizeGB` Le paramètre dépend de la mémoire du GPU et du nombre de GPU affectés à la machine virtuelle. Par exemple :

- Si une machine virtuelle est affectée à 4 GPU V100 de 32 Go, la valeur doit être 128.
- Si une machine virtuelle est affectée à 4 processeurs graphiques P6 de 16 Go, cette valeur doit être 64.

Edit Settings
RHEL01

> Boot Options
Expand for boot options

Advanced

Settings

☐ Disable acceleration
☒ Enable logging

Debugging and statistics

Run normally

Swap file location

☒ Default

Use the settings of the cluster or host containing the virtual machine.

☐ Virtual machine directory

Store the swap files in the same directory as the virtual machine.

☐ Datastore specified by host

Store the swap files in the datastore specified by the host to be used for swap files. If not possible, store the swap files in the same directory as the virtual machine. Using a datastore that is not visible to both hosts during vMotion might affect the vMotion performance for the affected virtual machines.

Configuration Parameters

EDIT CONFIGURATION...

Latency Sensitivity

Normal

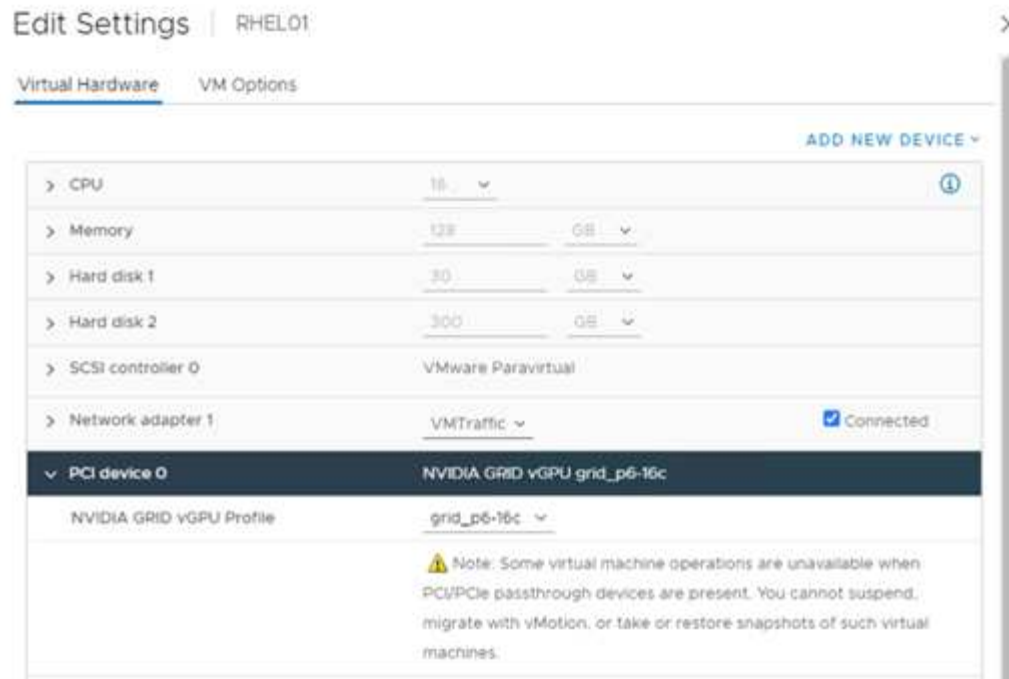
> Fibre Channel NPIV
Expand for Fibre Channel NPIV settings

Configuration Parameters

⚠ Modify or add configuration parameters as needed for experimental features or as instructed by technical support. Empty values will be removed (supported on ESXi 6.0 and later).

Name	Value
pciPassthru.64bitMMIOSizeGB	64
pciPassthru.use64bitMMIO	TRUE

- Lorsque vous ajoutez des vGPU en tant que nouveau périphérique PCI à la machine virtuelle dans vCenter, veillez à sélectionner NVIDIA GRID vGPU comme type de périphérique PCI.
- Choisissez le profil GPU correct qui suit le GPU utilisé, la mémoire GPU et son usage : par exemple, les graphiques plutôt que le calcul.



15. Sur la VM RedHat Linux, les pilotes NVIDIA peuvent être installés en exécutant la commande suivante :

```
[root@genomics1 genomics]# sh NVIDIA-Linux-x86_64-460.73.01-grid.run
```

16. Vérifiez que le profil vGPU correct est signalé en exécutant la commande suivante :

```
[root@genomics1 genomics]# nvidia-smi -query-gpu=gpu_name
-format=csv,noheader -id=0 | sed -e 's/ /-/g'
GRID-P6-16C
[root@genomics1 genomics]#
```

17. Après le redémarrage, vérifiez que le NVIDIA vGPU correct est signalé avec les versions du pilote.

```

[root@genomics1 genomics]# nvidia-smi
Wed Aug 18 20:30:56 2021
+-----+
+-----+
| NVIDIA-SMI 460.73.01      Driver Version: 460.73.01      CUDA Version:
11.2      |
|-----+-----+
+-----+
| GPU  Name           Persistence-M| Bus-Id        Disp.A | Volatile
Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|      Memory-Usage | GPU-Util
Compute M. |
|                               |                      |
MIG M. |
|=====+=====+=====
=====|
|   0  GRID P6-16C           On   | 00000000:02:02.0 Off |
N/A |
| N/A   N/A    P8     N/A /  N/A |   2205MiB / 16384MiB |      0%
Default |
|                               |                      |
N/A |
+-----+-----+
+-----+
+-----+
+-----+
+-----+
| Processes:
|
| GPU    GI    CI          PID    Type    Process name                  GPU
Memory |
|        ID    ID                                   Usage
|
|=====+=====+=====
=====|
|   0    N/A   N/A         8604      G   /usr/libexec/Xorg
13MiB |
+-----+-----+
+-----+
[root@genomics1 genomics]#

```

18. Assurez-vous que l'adresse IP du serveur de licences est configurée sur la machine virtuelle dans le fichier de configuration de la grille vGPU.

a. Copier le modèle.

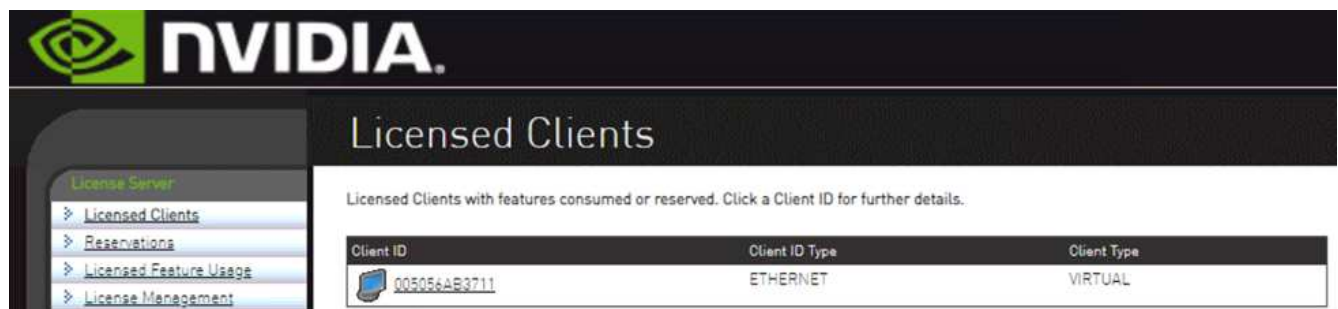
```
[root@genomics1 genomics]# cp /etc/nvidia/gridd.conf.template  
/etc/nvidia/gridd.conf
```

- b. Modifiez le fichier `/etc/nvidia/rid.conf`, Ajoutez l'adresse IP du serveur de licences et définissez le type de fonction sur 1.

```
ServerAddress=192.168.169.10
```

```
FeatureType=1
```

19. Après avoir redémarré la VM, vous devriez voir une entrée sous clients sous Licence dans le serveur de licences comme indiqué ci-dessous.



20. Se reporter à la section Configuration des solutions pour plus d'informations sur le téléchargement des logiciels GATK et Cromwell.
21. Une fois que la GATK peut utiliser des GPU sur site, le langage de description du workflow `*.wdl` possède les attributs d'exécution comme indiqué ci-dessous.

```

task ValidateBAM {
  input {
    # Command parameters
    File input_bam
    String output_basename
    String? validation_mode
    String gatk_path
    # Runtime parameters
    String docker
    Int machine_mem_gb = 4
    Int additional_disk_space_gb = 50
  }
  Int disk_size = ceil(size(input_bam, "GB")) + additional_disk_space_gb
  String output_name = "${output_basename}_${validation_mode}.txt"
  command {
    ${gatk_path} \
      ValidateSamFile \
      --INPUT ${input_bam} \
      --OUTPUT ${output_name} \
      --MODE ${default="SUMMARY" validation_mode}
  }
  runtime {
    gpuCount: 1
    gpuType: "nvidia-tesla-p6"
    docker: docker
    memory: machine_mem_gb + " GB"
    disks: "local-disk " + disk_size + " HDD"
  }
  output {
    File validation_report = "${output_name}"
  }
}

```

["Suivant: Conclusion."](#)

Conclusion

["Précédent : configuration du GPU."](#)

De nombreuses organisations du secteur de la santé à travers le monde ont choisi FlexPod comme plateforme commune. FlexPod vous permet de déployer des capacités médicales en toute confiance. FlexPod avec NetApp ONTAP est fourni de série avec la capacité d'implémenter un ensemble de protocoles de pointe, clé en main. Quelle que soit l'origine de la demande d'exécution de la génomique d'un patient donné, une plateforme FlexPod assure l'interopérabilité, l'accessibilité, la disponibilité et l'évolutivité.

Lorsqu'elle est standardisée sur une plate-forme FlexPod, la culture de l'innovation devient contagieuse.

Où trouver des informations complémentaires

Pour en savoir plus sur les informations données dans ce livre blanc, consultez ces documents et sites web :

- FlexPod Datacenter pour l'IA/ML avec Cisco UCS 480 ML pour le deep learning

["https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/flexpod_480ml_aiml_deployement.pdf"](https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/flexpod_480ml_aiml_deployement.pdf)

- FlexPod Datacenter avec VMware vSphere 7.0 et NetApp ONTAP 9.7

["https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/fp_vmware_vsphere_7_0_ontap_9_7.html"](https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/fp_vmware_vsphere_7_0_ontap_9_7.html)

- Centre de documentation ONTAP 9

["http://docs.netapp.com"](http://docs.netapp.com)

- Agile et efficace : FlexPod favorise la modernisation des data centers

["https://www.flexpod.com/idc-white-paper/"](https://www.flexpod.com/idc-white-paper/)

- L'IA dans le domaine de la santé

["https://www.netapp.com/us/media/na-369.pdf"](https://www.netapp.com/us/media/na-369.pdf)

- FlexPod pour le secteur de la santé facilite votre transformation

["https://flexpod.com/solutions/verticals/healthcare/"](https://flexpod.com/solutions/verticals/healthcare/)

- FlexPod de Cisco et NetApp

["https://flexpod.com/"](https://flexpod.com/)

- IA et analytique pour le secteur de la santé (NetApp)

["https://www.netapp.com/us/artificial-intelligence/healthcare-ai-analytics/index.aspx"](https://www.netapp.com/us/artificial-intelligence/healthcare-ai-analytics/index.aspx)

- Des choix d'infrastructure intelligents pour le secteur de la santé favorisent la réussite

<https://www.netapp.com/pdf.html?item=/media/7410-wp-7314.pdf>

- FlexPod Datacenter avec ONTAP 9.8, ONTAP Storage Connector for Cisco Intersight et Cisco Intersight Managed mode.

<https://www.netapp.com/pdf.html?item=/media/25001-tr-4883.pdf>

- FlexPod Datacenter avec Red Hat Enterprise Linux OpenStack Platform

["https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/flexpod_openstack_osp6.html"](https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/flexpod_openstack_osp6.html)

Historique des versions

Version	Date	Historique des versions du document
Version 1.0	Novembre 2021	Version initiale.

Informations sur le copyright

Copyright © 2025 NetApp, Inc. Tous droits réservés. Imprimé aux États-Unis. Aucune partie de ce document protégé par copyright ne peut être reproduite sous quelque forme que ce soit ou selon quelque méthode que ce soit (graphique, électronique ou mécanique, notamment par photocopie, enregistrement ou stockage dans un système de récupération électronique) sans l'autorisation écrite préalable du détenteur du droit de copyright.

Les logiciels dérivés des éléments NetApp protégés par copyright sont soumis à la licence et à l'avis de non-responsabilité suivants :

CE LOGICIEL EST FOURNI PAR NETAPP « EN L'ÉTAT » ET SANS GARANTIES EXPRESSES OU TACITES, Y COMPRIS LES GARANTIES TACITES DE QUALITÉ MARCHANDE ET D'ADÉQUATION À UN USAGE PARTICULIER, QUI SONT EXCLUES PAR LES PRÉSENTES. EN AUCUN CAS NETAPP NE SERA TENU POUR RESPONSABLE DE DOMMAGES DIRECTS, INDIRECTS, ACCESSOIRES, PARTICULIERS OU EXEMPLAIRES (Y COMPRIS L'ACHAT DE BIENS ET DE SERVICES DE SUBSTITUTION, LA PERTE DE JOUISSANCE, DE DONNÉES OU DE PROFITS, OU L'INTERRUPTION D'ACTIVITÉ), QUELLES QU'EN SOIENT LA CAUSE ET LA DOCTRINE DE RESPONSABILITÉ, QU'IL S'AGISSE DE RESPONSABILITÉ CONTRACTUELLE, STRICTE OU DÉLICTELLE (Y COMPRIS LA NÉGLIGENCE OU AUTRE) DÉCOULANT DE L'UTILISATION DE CE LOGICIEL, MÊME SI LA SOCIÉTÉ A ÉTÉ INFORMÉE DE LA POSSIBILITÉ DE TELS DOMMAGES.

NetApp se réserve le droit de modifier les produits décrits dans le présent document à tout moment et sans préavis. NetApp décline toute responsabilité découlant de l'utilisation des produits décrits dans le présent document, sauf accord explicite écrit de NetApp. L'utilisation ou l'achat de ce produit ne concède pas de licence dans le cadre de droits de brevet, de droits de marque commerciale ou de tout autre droit de propriété intellectuelle de NetApp.

Le produit décrit dans ce manuel peut être protégé par un ou plusieurs brevets américains, étrangers ou par une demande en attente.

LÉGENDE DE RESTRICTION DES DROITS : L'utilisation, la duplication ou la divulgation par le gouvernement sont sujettes aux restrictions énoncées dans le sous-paragraphe (b)(3) de la clause Rights in Technical Data-Noncommercial Items du DFARS 252.227-7013 (février 2014) et du FAR 52.227-19 (décembre 2007).

Les données contenues dans les présentes se rapportent à un produit et/ou service commercial (tel que défini par la clause FAR 2.101). Il s'agit de données propriétaires de NetApp, Inc. Toutes les données techniques et tous les logiciels fournis par NetApp en vertu du présent Accord sont à caractère commercial et ont été exclusivement développés à l'aide de fonds privés. Le gouvernement des États-Unis dispose d'une licence limitée irrévocable, non exclusive, non cessible, non transférable et mondiale. Cette licence lui permet d'utiliser uniquement les données relatives au contrat du gouvernement des États-Unis d'après lequel les données lui ont été fournies ou celles qui sont nécessaires à son exécution. Sauf dispositions contraires énoncées dans les présentes, l'utilisation, la divulgation, la reproduction, la modification, l'exécution, l'affichage des données sont interdits sans avoir obtenu le consentement écrit préalable de NetApp, Inc. Les droits de licences du Département de la Défense du gouvernement des États-Unis se limitent aux droits identifiés par la clause 252.227-7015(b) du DFARS (février 2014).

Informations sur les marques commerciales

NETAPP, le logo NETAPP et les marques citées sur le site <http://www.netapp.com/TM> sont des marques déposées ou des marques commerciales de NetApp, Inc. Les autres noms de marques et de produits sont des marques commerciales de leurs propriétaires respectifs.