FlexPod para genomas

FlexPod

NetApp
March 25, 2024

Tabla de contenidos

FlexPod para genomas

TR-4911: Genómica de FlexPod

JayaKishore Esanakula, NetApp

Hay pocos campos de la medicina que son más importantes que la genómica para la salud y las ciencias biológicas, y la genómica se está convirtiendo rápidamente en una herramienta clínica clave para los médicos y las enfermeras. La genómica, cuando se combina con imágenes médicas y patología digital, nos ayuda a entender cómo los genes de un paciente pueden verse afectados por los protocolos de tratamiento. El éxito de la genómica en el sector sanitario depende cada vez más de la interoperabilidad de datos a escala. El objetivo final es comprender los enormes volúmenes de datos genéticos e identificar correlaciones y variantes clínicamente relevantes que mejoran el diagnóstico y hacen realidad la medicina de precisión. La genómica nos ayuda a comprender el origen de los brotes de enfermedades, cómo evolucionan las enfermedades y qué tratamientos y estrategias podrían ser eficaces. Claramente, la genómica tiene muchos beneficios que abarcan la prevención, el diagnóstico y el tratamiento. Las organizaciones de atención médica se enfrentan a varios desafíos, entre ellos los siguientes:

- Calidad de la atención mejorada
- Atención basada en el valor
- Explosión de datos
- Medicina de precisión
- Pandemias
- Prendas tecnológicas, supervisión remota y atención sanitaria
- Seguridad cibernética

Las vías clínicas estandarizadas y los protocolos clínicos son uno de los componentes críticos de la medicina moderna. Uno de los aspectos clave de la estandarización es la interoperabilidad entre los proveedores de atención médica, no solo para los registros médicos, sino también para los datos genómicos. La gran pregunta es ¿las organizaciones sanitarias renunciarán a la propiedad de los datos genómicos en lugar de que los pacientes sean propietarios de los datos personales de su genómica y los registros médicos relacionados?

La interoperabilidad de los datos de los pacientes es fundamental para posibilitar la medicina de precisión, una de las fuerzas que impulsan la reciente explosión del crecimiento de los datos. El objetivo de la medicina de precisión es hacer que el mantenimiento de la salud, la prevención de enfermedades, los diagnósticos y las soluciones de tratamiento sean más eficaces y precisos.

La tasa de crecimiento de los datos ha sido exponencial. A principios de febrero de 2021, los laboratorios estadounidenses secuenciaron aproximadamente 8,000 cepas del COVID-19 por semana. El número de genomas secuenciados había aumentado a 29,000 por semana en abril de 2021. Cada genoma humano secuenciado en su totalidad tiene un tamaño de alrededor de 125 GB. Por lo tanto, a una tasa de 29,000 genomas secuenciados por semana, el almacenamiento total del genoma en reposo sería de más de 180 petabytes al año. Varios países han comprometido recursos a la epidemiología genómica para mejorar la vigilancia genómica y prepararse para la próxima oleada de desafíos mundiales en materia de salud.

La reducción del costo de la investigación genómica está impulsando la investigación y las pruebas genéticas a un ritmo sin precedentes. Los tres P se encuentran en un punto de inflexión: Potencia informática, privacidad de datos y personalización de la medicina. En 2025, los investigadores estiman que se secuenciarán entre 100 y 2 millones de genomas humanos. Para que la genómica sea eficaz y sea una propuesta valiosa, las capacidades de genómica deben ser una parte integral de los flujos de trabajo de atención médica; debe ser fácil de acceder y ser aplicable durante la visita de un paciente. También es igualmente importante integrar los datos de historiales médicos electrónicos de los pacientes con los datos de genomas de los pacientes. Con la llegada de una infraestructura convergente de vanguardia como FlexPod, las organizaciones pueden incorporar sus funcionalidades de genómica en los flujos de trabajo diarios de médicos, enfermeras y gestores clínicos. Para obtener la información más reciente sobre la plataforma FlexPod, consulte este tema "Whitepaper de FlexPod Datacenter con Cisco UCS X-Series".

Para un médico, el verdadero valor de la genómica incluye la medicina de precisión y los planes de tratamiento personalizados basados en los datos genómicos de un paciente. Nunca ha habido tal sinergia entre médicos y científicos de datos en el pasado, y la genómica se está beneficiando de las innovaciones tecnológicas del pasado reciente, así como de asociaciones reales entre organizaciones sanitarias y líderes tecnológicos del sector.

Los centros médicos académicos y otras organizaciones de salud y ciencias de la vida están en camino de establecer el centro de excelencia (COE) en ciencia genómica. Según el Dr. Charlie Gersbach, Dr. Greg Crawford, y Dr. Tim E Reddy, de la Universidad de Duke, “sabemos que los genes no están activados o desactivados por un simple interruptor binario, sino que es el resultado de múltiples interruptores regulatorios genéticos que funcionan juntos. También han determinado que "ninguna de estas partes del genoma trabaja en aislamiento. El genoma es una red muy complicada que la evolución ha tejido” ("ref").

NetApp y Cisco llevan trabajando muy duro para implementar mejoras incrementales en la plataforma de FlexPod durante más de 10 años. Todos los comentarios de los clientes se escuchan, evalúan y están ligados a las fuentes de valor y los conjuntos de funciones de FlexPod. Este continuo bucle de comentarios, colaboración, mejora y celebración diferencia a FlexPod de la plataforma de infraestructura convergente de confianza de todo el mundo. Se ha simplificado y diseñado desde cero para ser la plataforma más fiable, robusta, versátil y ágil para organizaciones sanitarias.

Ámbito

La plataforma de infraestructura convergente FlexPod permite a una organización sanitaria alojar una o más cargas de trabajo genómica, junto con otras aplicaciones clínicas y no clínicas de la sanidad. En este informe técnico se utiliza una herramienta genómica de código abierto estándar en el sector denominada GATK durante la validación de la plataforma FlexPod. Sin embargo, una discusión más profunda sobre genómica o GATK está fuera del alcance de este documento.

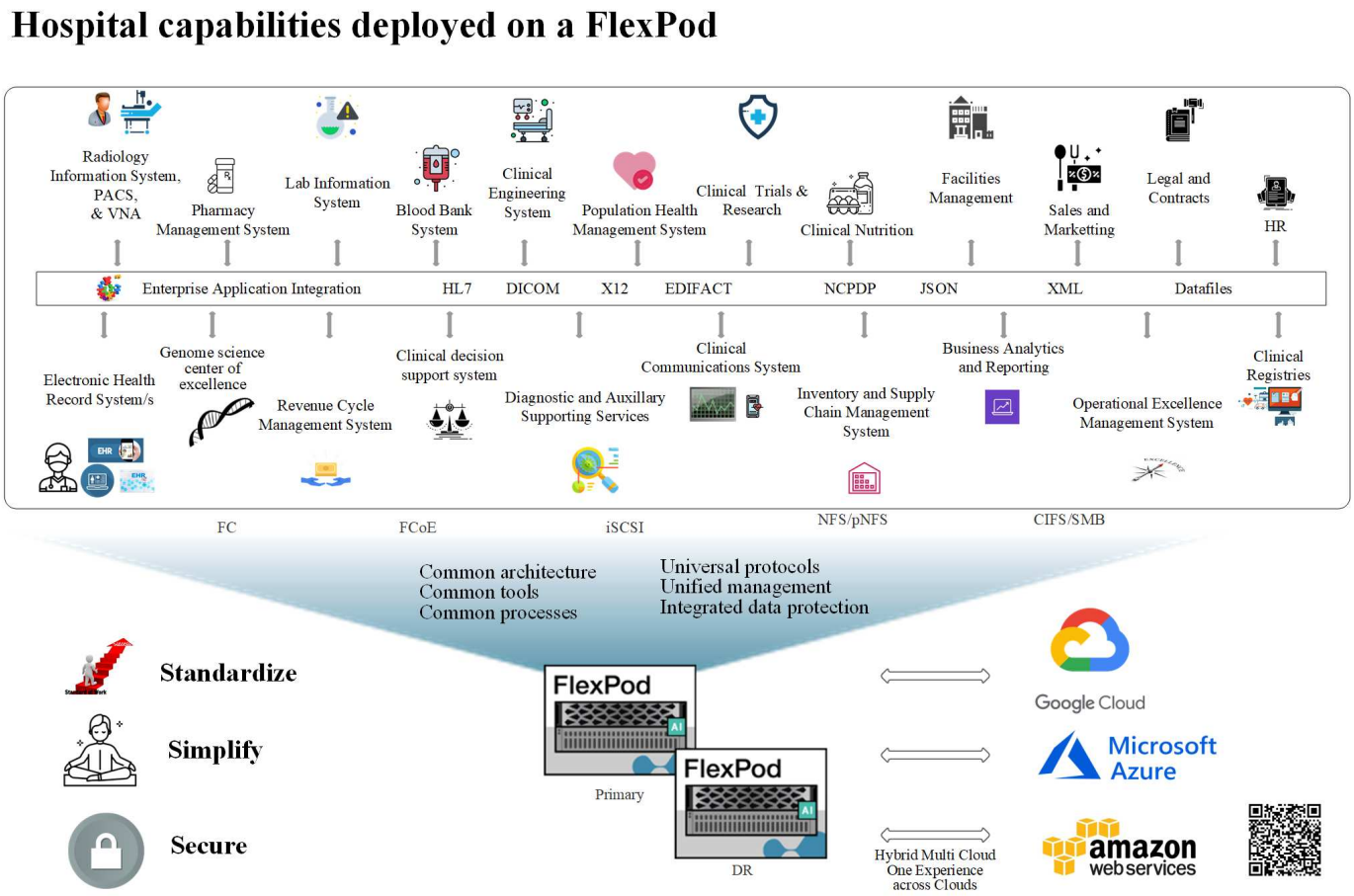
Destinatarios

Este documento está dirigido a líderes técnicos del sector sanitario y a ingenieros de soluciones de partners de Cisco y NetApp y personal de servicios profesionales. NetApp asume que el lector tiene un buen conocimiento de los conceptos de configuración de la computación y el almacenamiento, así como una familiaridad técnica con las amenazas para la salud, la seguridad sanitaria, los sistemas TECNOLÓGICOS para el sector sanitario, Cisco UCS y los sistemas de almacenamiento de NetApp.

Capacidades hospitalarias implementadas en FlexPod

Un hospital típico tiene un conjunto diversificado de sistemas DE TECNOLOGÍA. La mayoría de estos sistemas son adquiridos a un proveedor, mientras que muy pocos son construidos por el sistema hospitalario interno. Por tanto, el sistema hospitalario debe gestionar distintos entornos de infraestructura en sus centros de datos. Cuando los hospitales unifican sus sistemas en una plataforma de infraestructura convergente como

FlexPod, las organizaciones pueden estandarizar sus operaciones de centro de datos. Con FlexPod, las organizaciones sanitarias pueden implementar sistemas clínicos y no clínicos en la misma plataforma, unificando las operaciones del centro de datos.



"Siguiente: Ventajas de la puesta en marcha de cargas de trabajo genómicos en FlexPod"

Ventajas de la puesta en marcha de cargas de trabajo genómicos en FlexPod

"Anterior: Introducción."

En esta sección se proporciona una breve lista de ventajas para ejecutar una carga de trabajo genómica en una plataforma de infraestructura convergente FlexPod. Describiremos rápidamente las capacidades de un hospital. La siguiente vista de arquitectura empresarial muestra las capacidades de un hospital puestas en marcha en una plataforma de infraestructura convergente FlexPod preparada para cloud híbrido.

- * Evitar silos en la salud.* los silos en la salud son una preocupación muy real. A menudo, los departamentos se encuentran organizados en silos en su propio conjunto de hardware y software, no por elección, sino de forma orgánica, por evolución. Por ejemplo, radiología, cardiología, EHR, genómica, análisis, ciclo de ingresos y otros departamentos terminan con su conjunto individual de software y hardware dedicados. Las organizaciones sanitarias mantienen un conjunto limitado de profesionales DE TI para gestionar sus activos de hardware y software. El punto de inflexión se produce cuando se espera que este conjunto de personas gestionen un conjunto de hardware y software muy diversificado. La heterogeneidad se ve agravada por un conjunto incongruente de procesos llevados a la organización

sanitaria por los proveedores.

- **Iniciar pequeño y crecer.** el kit de herramientas GATK está adaptado para la ejecución de CPU, que mejor paquete de plataformas como FlexPod. FlexPod permite una escalabilidad independiente de la red, el entorno informático y el almacenamiento. Empiece con poco y escale a medida que aumenten sus funcionalidades de genómica y el entorno. Las organizaciones sanitarias no tienen que invertir en plataformas especializadas para ejecutar cargas de trabajo genómicas. En su lugar, las organizaciones pueden aprovechar plataformas versátiles como FlexPod para ejecutar cargas de trabajo genómica y no genómicas en la misma plataforma. Por ejemplo, si el departamento de pediatría quiere implementar una capacidad genómica, los líderes TECNOLÓGICOS pueden aprovisionar computación, almacenamiento y redes en una instancia de FlexPod existente. A medida que crece la unidad de negocio genómica, la organización sanitaria puede escalar su plataforma FlexPod según sea necesario.
- **Un único panel de control y una flexibilidad sin igual.** Cisco Intersight simplifica significativamente las operaciones DE TI al unir aplicaciones con la infraestructura, proporcionando visibilidad y gestión desde servidores e hipervisores sin sistema operativo hasta aplicaciones sin servidor, reduciendo así los costes y mitigando el riesgo. Esta plataforma SaaS unificada utiliza un diseño de API abierta unificado que se integra de forma nativa con plataformas y herramientas de terceros. Además, permite la gestión desde su equipo de operaciones de centro de datos in situ o desde cualquier lugar mediante una aplicación móvil.

  Los usuarios liberan rápidamente un valor tangible en su entorno al aprovechar Intersight como su plataforma de gestión. Al permitir la automatización de muchas tareas manuales diarias, Intersight elimina los errores y simplifica las operaciones diarias. Además, las avanzadas capacidades de soporte que ofrece Intersight permiten a los usuarios mantenerse por delante de los problemas y acelerar la resolución de estos. En conjunto, las organizaciones dedican mucho menos tiempo y dinero a la infraestructura de aplicaciones y más tiempo al desarrollo empresarial principal.

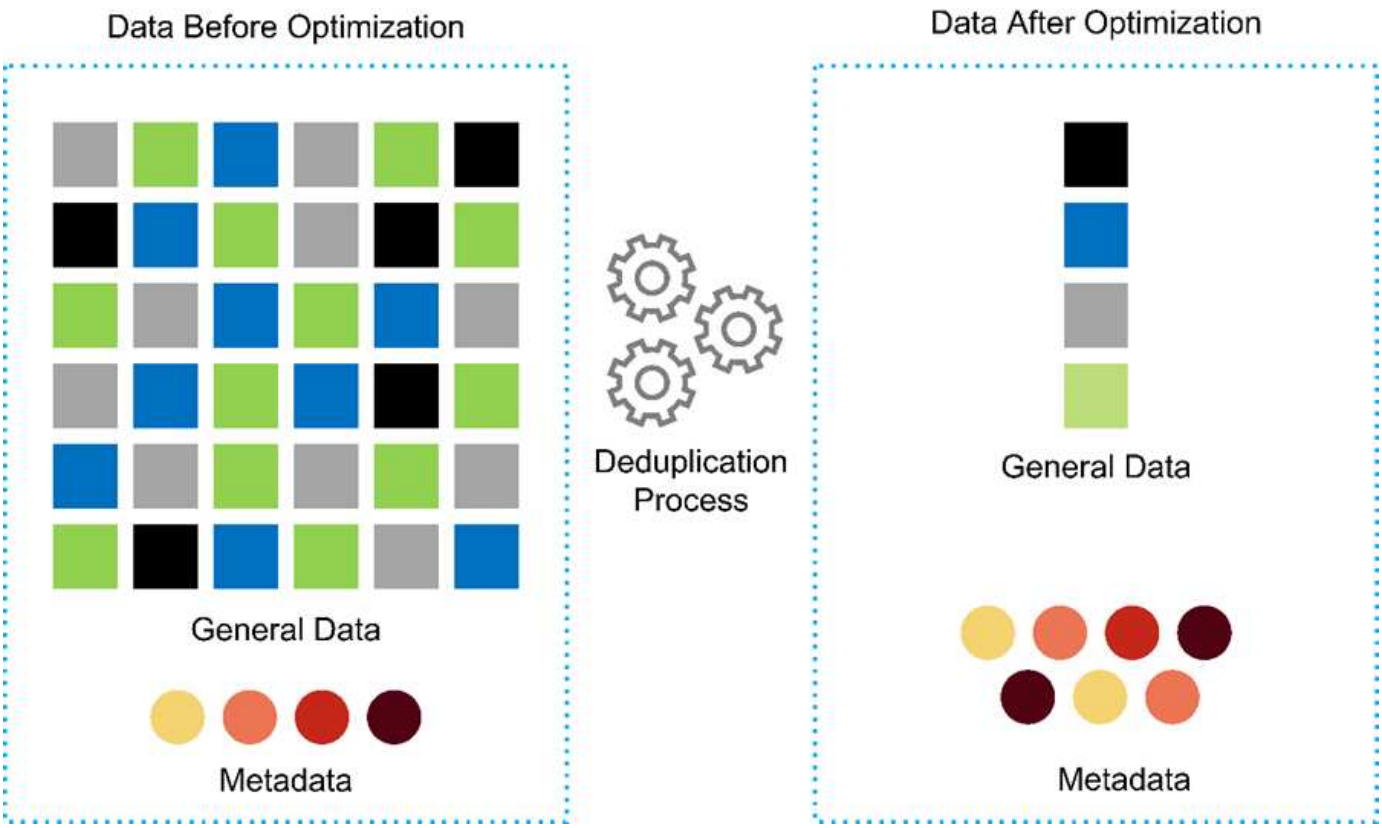
  Aprovechar la gestión de Intersight y la arquitectura de FlexPod, fácilmente escalable, permite a las organizaciones ejecutar varias cargas de trabajo de genoma en una única plataforma FlexPod, lo que aumenta el uso y reduce el coste total de propiedad (TCO). FlexPod permite un dimensionamiento flexible, con opciones que empiezan por nuestro pequeño FlexPod Express y se pueden escalar a implementaciones grandes de centros de datos FlexPod. Con las funcionalidades de control de acceso basadas en funciones integradas en Cisco Intersight, las organizaciones sanitarias pueden implementar mecanismos sólidos de control de acceso, lo que evita la necesidad de pilas de infraestructuras independientes. El uso de varias unidades de negocio en la organización sanitaria puede aprovechar la genómica como una competencia principal.

  En última instancia, FlexPod ayuda a simplificar las operaciones DE TI y a reducir los costos operativos y permite que los administradores de infraestructuras DE TI se centren en tareas que ayudan a los médicos a innovar en lugar de quedar relegados a mantener las operaciones básicas.

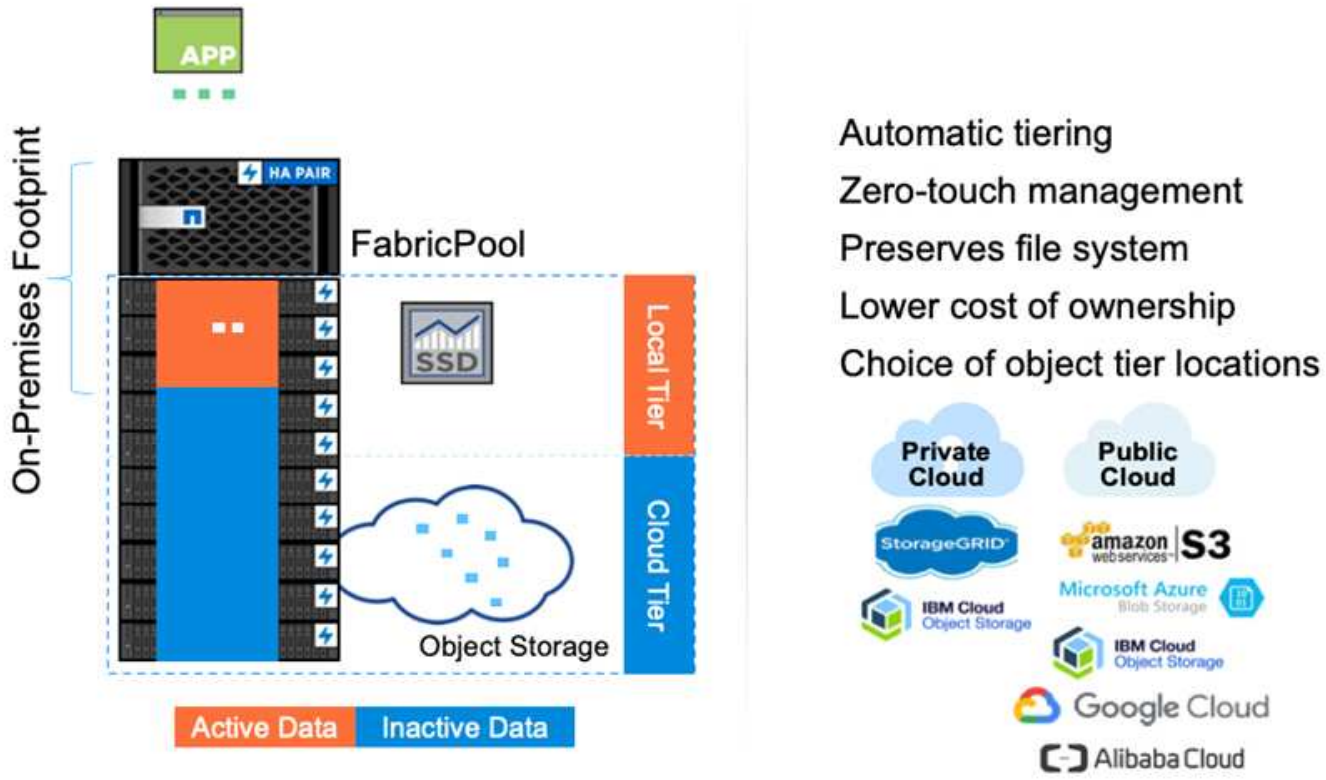
- **Diseño validado y resultados garantizados.** las guías de diseño e implementación de FlexPod se validan para ser repetibles, y cubren detalles completos de la configuración y las mejores prácticas del sector que se necesitan para implementar un FlexPod con confianza. Las guías de diseño validado, las guías de puesta en marcha y las arquitecturas de Cisco y NetApp ayudan a su organización de ciencias biológicas o sanitarias a eliminar las conjeturas de la implantación de una plataforma validada y de confianza desde el principio. Con FlexPod puede acelerar los tiempos de puesta en marcha y reducir costes, complejidad y riesgos. Los diseños validados y las guías de puesta en marcha de FlexPod establecen a FlexPod como la plataforma perfecta para diversas cargas de trabajo de genómica.
- * Innovación y agilidad.* se recomienda FlexPod como una plataforma ideal por EHR como Epic, Cerner, Meditech y sistemas de imagen como Agfa, GE, Philips. Para obtener más información acerca de "Rollo de honor épico" Y la arquitectura de la plataforma de destino; consulte la web de usuarios de Epic. Ejecutando la genómica "FlexPod" permite a las organizaciones sanitarias continuar su camino de innovación con agilidad. Con FlexPod, la implementación del cambio de organización viene de forma natural. Cuando las organizaciones se estandarizan en una plataforma de FlexPod, los expertos EN

TECNOLOGÍA del sector sanitario pueden disponer de tiempo, esfuerzo y recursos para innovar y, por lo tanto, ser tan ágiles como exige el ecosistema.

- **Datos liberados**. Con la plataforma de infraestructura convergente FlexPod y el sistema de almacenamiento ONTAP de NetApp, es posible acceder a los datos genómicos utilizando una amplia variedad de protocolos a escala desde una única plataforma. FlexPod con ONTAP de NetApp ofrece una plataforma de cloud híbrido sencilla, intuitiva y potente. Su Data Fabric, impulsado por ONTAP de NetApp, aúna los datos de distintos sitios, sin que los barreras físicas y de todas las aplicaciones. Su Data Fabric se ha creado pensando en empresas basadas en datos en un mundo donde todo gira alrededor de ellos. Los datos se crean y se usan en varias ubicaciones y, a menudo, hay que aprovecharlos y compartirlos con otras ubicaciones, aplicaciones e infraestructuras. Por tanto, se necesita una forma coherente e integrada de gestionarla. FlexPod controla a su equipo DE TECNOLOGÍA y simplifica la cada vez mayor complejidad TECNOLÓGICA.
- **Multitenancy seguro.** FlexPod utiliza módulos criptográficos compatibles con FIPS 140-2, lo que permite a las organizaciones implementar la seguridad como un elemento fundamental, no como una idea de última hora. FlexPod permite a las organizaciones implantar multi-tenancy seguro desde una única plataforma de infraestructura convergente, independientemente del tamaño de la plataforma. FlexPod con multi-tenancy seguro y calidad de servicio ayudan con la separación de cargas de trabajo y maximizan la utilización. Esto ayuda a evitar que el capital quede atrapado en plataformas especializadas que potencialmente están infrautilizadas y requiere un conjunto de habilidades especializadas para gestionarlas.
- **La eficiencia del almacenamiento.** la genómica requiere que el almacenamiento subyacente tenga las funcionalidades de eficiencia del almacenamiento líderes en el sector. Puede reducir los costes de almacenamiento con funciones de eficiencia del almacenamiento de NetApp como la deduplicación (inline y bajo demanda), la compresión de datos y la compactación de datos ("ref"). La deduplicación de NetApp proporciona deduplicación a nivel de bloque en un volumen FlexVol. En esencia, la deduplicación elimina los bloques duplicados y solo almacena los bloques únicos en el volumen FlexVol. La deduplicación funciona con un alto grado de granularidad y funciona en el sistema de archivos activo del volumen FlexVol. En la siguiente figura se muestra información general sobre el funcionamiento de la deduplicación de NetApp. La deduplicación es transparente para las aplicaciones. Por lo tanto, se puede usar para deduplicar datos originados por cualquier aplicación que utilice el sistema de NetApp. Puede ejecutar la deduplicación del volumen como un proceso en línea y como un proceso en segundo plano. Es posible configurarlo para que se ejecute automáticamente, para que se programe o manualmente a través de la CLI, ONTAP System Manager de NetApp o Active IQ Unified Manager de NetApp.



- **Permitir la interoperabilidad genómica.** ONTAP FlexCache es una capacidad de almacenamiento en caché remoto que simplifica la distribución de archivos, reduce la latencia WAN y reduce los costes de ancho de banda WAN ("ref"). Una de las actividades clave durante la identificación y anotación de variantes genómicas es la colaboración entre los médicos. La tecnología ONTAP FlexCache aumenta el rendimiento de los datos incluso cuando los médicos colaboradores se encuentran en diferentes ubicaciones geográficas. Dado el tamaño típico de un archivo *.BAM (de 1 GB a 100 s de GB), es fundamental que la plataforma subyacente pueda hacer que los archivos estén disponibles para los médicos en diferentes ubicaciones geográficas. FlexPod con ONTAP FlexCache hace que los datos genómicos y las aplicaciones estén realmente listos para los distintos sitios, lo que convierte la colaboración entre investigadores de todo el mundo en un proceso fluido con una baja latencia y un alto rendimiento. Las organizaciones sanitarias que ejecutan aplicaciones de genómica en un entorno multisitio pueden escalarse horizontalmente usando la estructura de datos para equilibrar la capacidad de gestión con el coste y la velocidad.
- **Uso inteligente de la plataforma de almacenamiento.** FlexPod con la clasificación automática por niveles de ONTAP y la tecnología Fabric Pool de NetApp simplifica la gestión de datos. FabricPool ayuda a reducir los costes de almacenamiento sin comprometer el rendimiento, la eficiencia, la seguridad o la protección. FabricPool es transparente para las aplicaciones empresariales y aprovecha las eficiencias del cloud al reducir el TCO del almacenamiento sin necesidad de rediseñar la arquitectura de la infraestructura de aplicaciones. FlexPod se puede beneficiar de las funcionalidades de almacenamiento por niveles de FabricPool para hacer un uso más eficiente del almacenamiento flash de ONTAP. Para obtener más información, consulte "FlexPod con FabricPool". El siguiente diagrama proporciona una descripción general de alto nivel de FabricPool y sus ventajas.



- **Análisis y anotación de variantes más rápidos.** la plataforma FlexPod es más rápida de implementar y operar. La plataforma FlexPod permite la colaboración entre los médicos, al poner los datos a escala con baja latencia y un mayor rendimiento. Una mayor interoperabilidad permite innovar. Las organizaciones sanitarias pueden ejecutar sus cargas de trabajo genómicos y no genómicas en paralelo, lo que significa que las organizaciones no necesitan plataformas especializadas para iniciar la transición a la genómica.

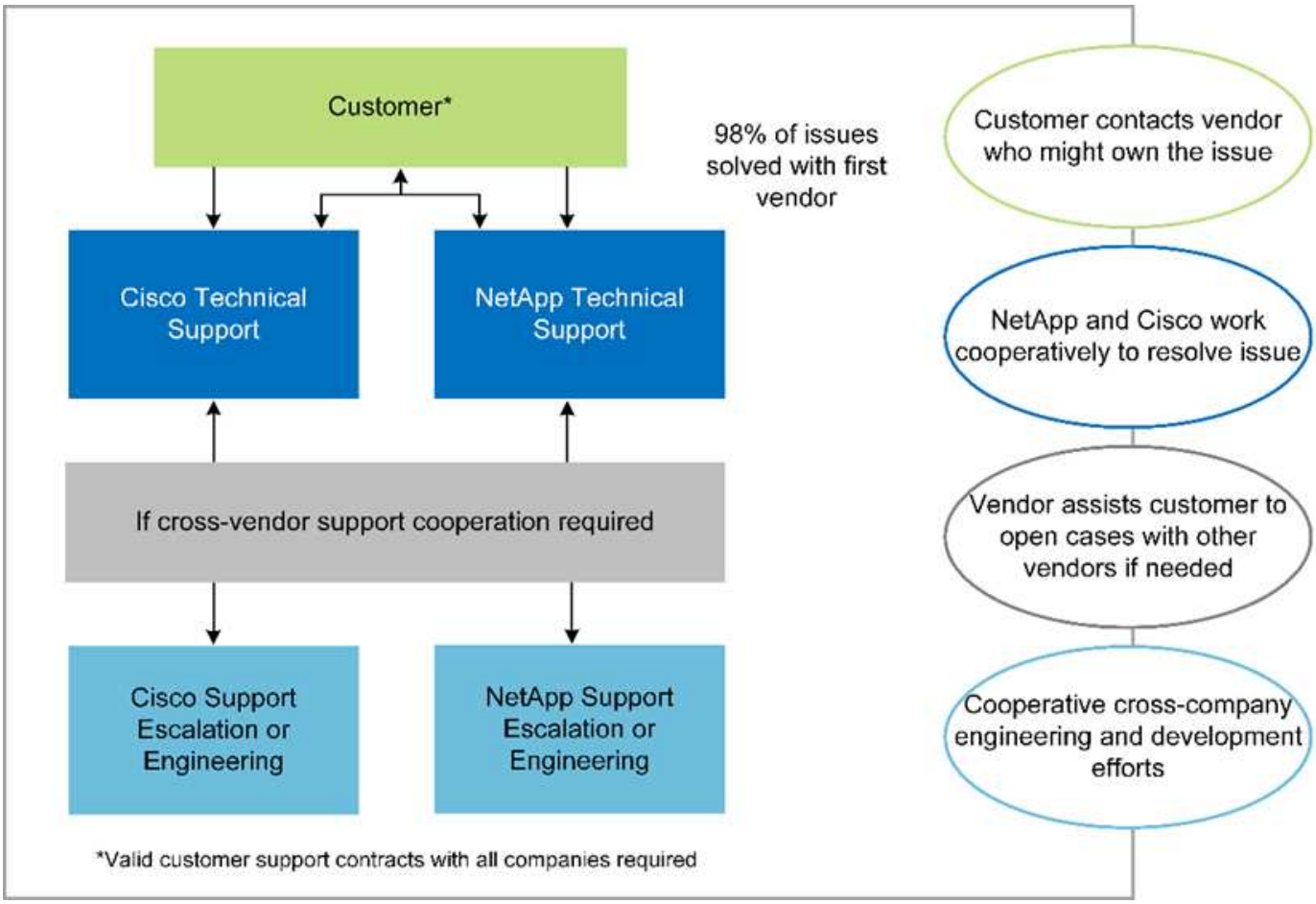
  FlexPod ONTAP añade habitualmente funciones innovadoras a la plataforma de almacenamiento. Centro de datos FlexPod es la base de infraestructura compartida óptima para poner en marcha FC- NVMe para permitir acceso al almacenamiento de alto rendimiento a aplicaciones que lo necesiten. A medida que FC-NVMe evoluciona para incluir alta disponibilidad, multivía y compatibilidad con sistemas operativos adicionales, FlexPod se adapta perfectamente a la plataforma preferida, lo que proporciona la escalabilidad y fiabilidad necesarias para respaldar estas funcionalidades. ONTAP con una I/o más rápida con NVMe integral permite que los análisis de genomas se completen con mayor rapidez ("ref").

  Los datos del genoma en secuencia RAW producen tamaños de archivo grandes, y es importante que estos archivos estén disponibles para los analizadores de variantes para reducir el tiempo total que toma de la colección de muestras a la anotación de variante. NVMe (memoria no volátil rápida) cuando se utiliza como protocolo de acceso al almacenamiento y transporte de datos proporciona niveles de rendimiento sin precedentes y los tiempos de respuesta más rápidos. FlexPod pone en marcha el protocolo NVMe a la vez que accede al almacenamiento flash a través del bus PCI Express (PCIe). PCIe posibilita la implementación de decenas de miles de colas de comandos, lo que aumenta la paralelización y el rendimiento. Un único protocolo del almacenamiento a la memoria acelera el acceso a los datos.

- **Agilidad para la investigación clínica desde el principio.** capacidad de almacenamiento flexible y ampliable y rendimiento permite a las organizaciones de investigación sanitaria optimizar el medio ambiente de una manera elástica o puntual (JIT). Al desacoplar el almacenamiento de la infraestructura informática y de red, la plataforma FlexPod se puede escalar horizontal y verticalmente sin interrupciones. Con Cisco Intersight, la plataforma FlexPod se puede gestionar con flujos de trabajo automatizados integrados y personalizados. Los flujos de trabajo de Cisco Intersight permiten a las organizaciones

sanitarias reducir los tiempos de gestión del ciclo de vida de las aplicaciones. Cuando un centro médico académico exige que los datos de los pacientes se anonimien y pongan a disposición de su centro de investigación informática y/o centro de calidad, su organización DE TI puede aprovechar los flujos de trabajo de Cisco Intersight FlexPod para realizar backups, clonarlos y restaurarlos con seguridad en cuestión de segundos y no de horas. Con Trident y Kubernetes de NetApp, las organizaciones TECNOLÓGICAS pueden suministrar nuevos científicos de datos y hacer que los datos clínicos estén disponibles para el desarrollo de modelos en cuestión de minutos, y en ocasiones incluso en segundos.

- **La protección de datos genómicos.** SnapLock de NetApp proporciona un volumen para una finalidad especial en el que los ficheros se pueden almacenar y realizar transacciones a un estado que no puede borrarse ni escribirse. Los datos de producción del usuario que se encuentran en un volumen FlexVol se pueden duplicar o realizar copias vault en un volumen SnapLock mediante la tecnología SnapMirror o SnapVault de NetApp. Los archivos del volumen de SnapLock, el volumen en sí y su agregado de alojamiento no se pueden eliminar hasta que finalice el período de retención. Mediante el software FPolicy de ONTAP, las organizaciones pueden evitar ataques de ransomware al dejar las operaciones en archivos con extensiones específicas. Es posible activar un evento de FPolicy para operaciones de archivos específicas. El evento está ligado a una política, que llama al motor que necesita utilizar. Puede configurar una política con un conjunto de extensiones de archivo que potencialmente puedan contener ransomware. Cuando un archivo con una extensión no permitida intenta realizar una operación no autorizada, FPolicy impide que esa operación se ejecute ("ref").
- **Soporte conjunto de FlexPod.** NetApp y Cisco han establecido el soporte conjunto de FlexPod, un modelo de soporte sólido, escalable y flexible para cumplir los requisitos de soporte exclusivos de la infraestructura convergente de FlexPod. Este modelo aprovecha la experiencia combinada, recursos y conocimientos de soporte técnico de NetApp y Cisco para ofrecer un proceso optimizado para identificar y solucionar problemas de soporte de FlexPod, independientemente de dónde se encuentre el problema. En la siguiente figura, se ofrece una descripción general del modelo de soporte conjunto de FlexPod. El cliente se pone en contacto con el proveedor que podría ser el propietario del problema y tanto Cisco como NetApp trabajan conjuntamente para resolverlo. Cisco y NetApp cuentan con equipos de desarrollo e ingeniería de múltiples empresas que trabajan conjuntamente para solucionar problemas. Este modelo de soporte reduce la pérdida de información durante la traducción, permite la confianza y reduce el tiempo de inactividad.

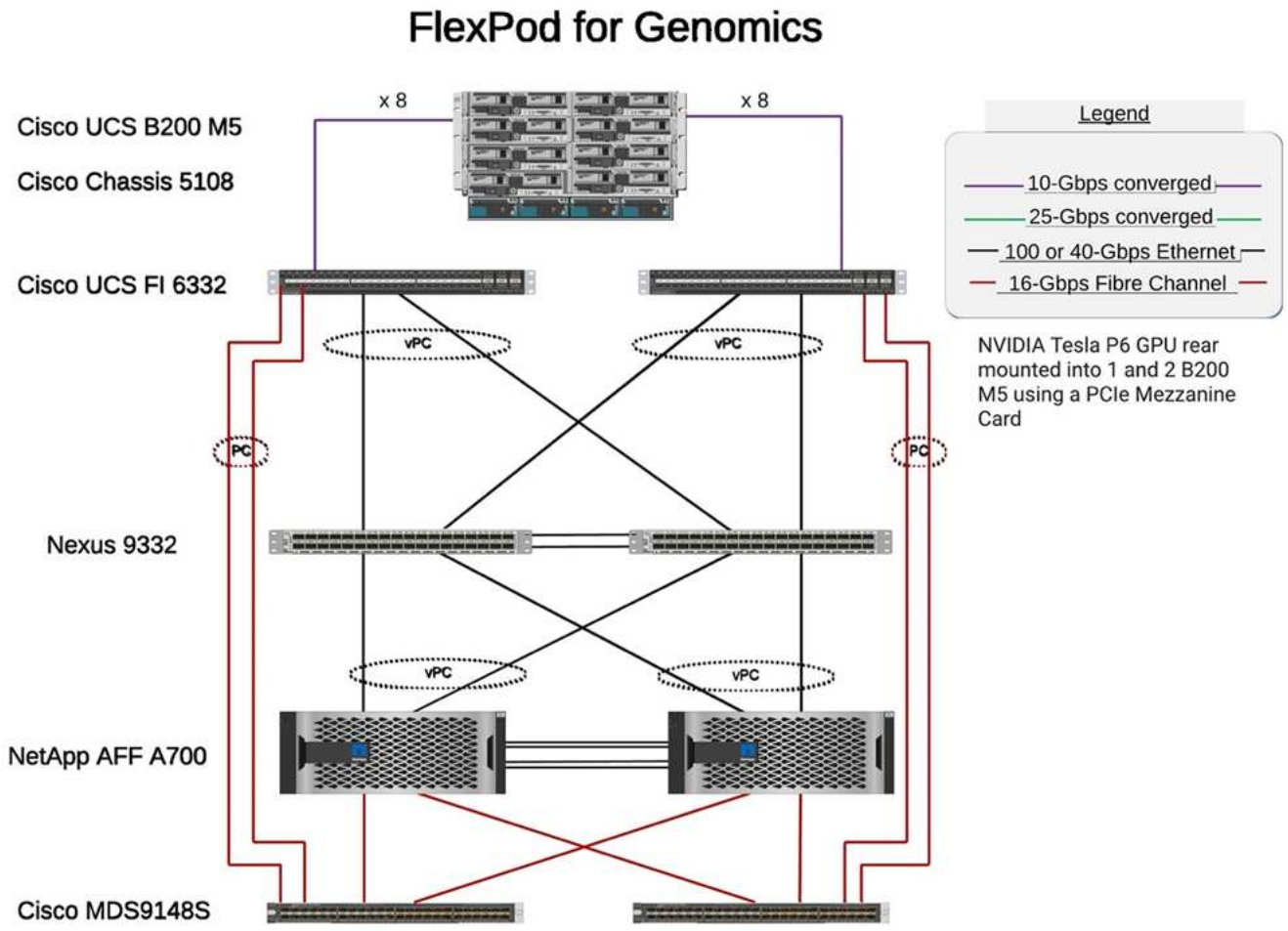


"Siguiente: Componentes de hardware y software de la infraestructura de la solución."

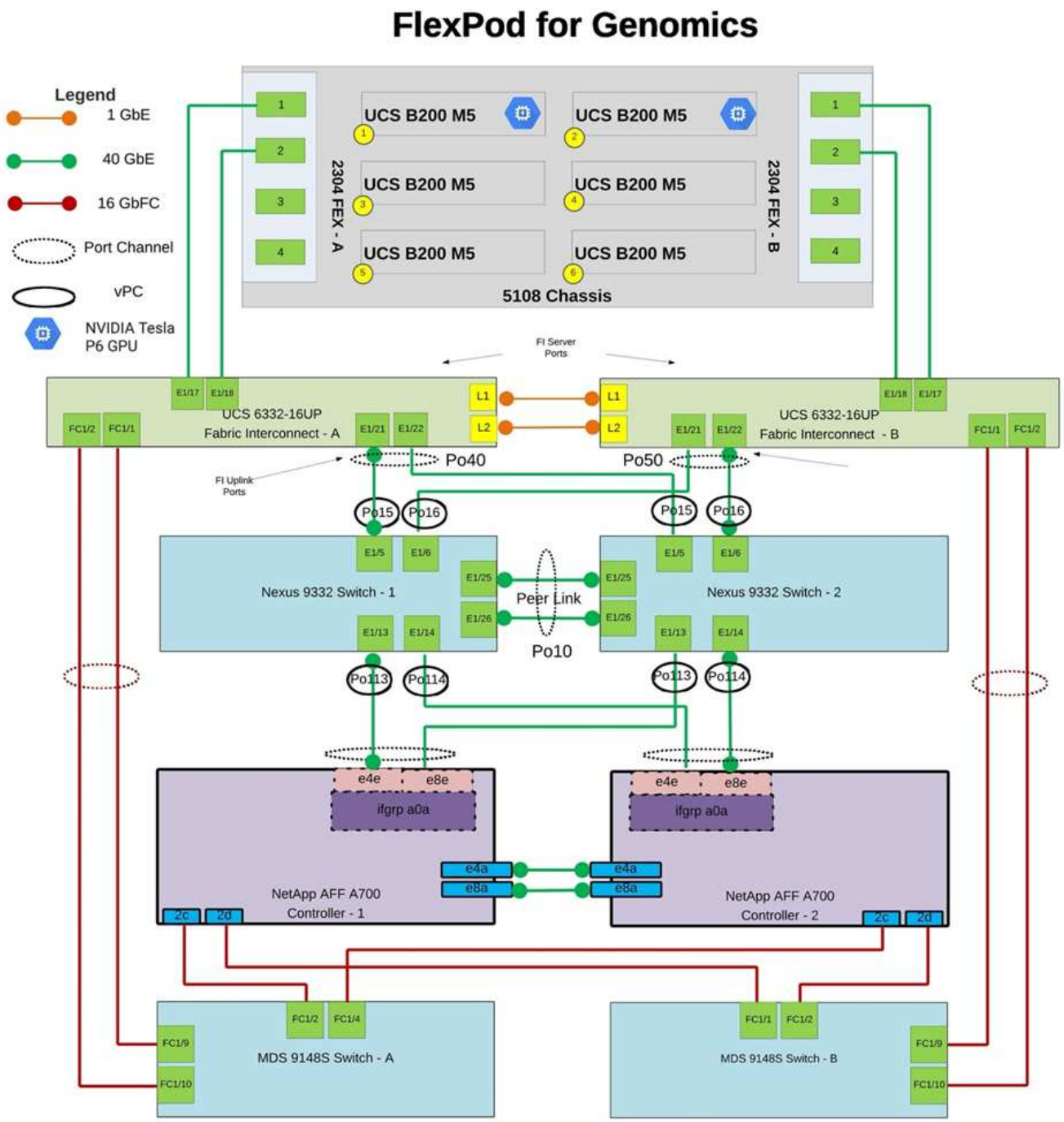
Infraestructura de la solución componentes de hardware y software

"Anterior: Ventajas de la puesta en marcha de cargas de trabajo genómicos en FlexPod"

La siguiente figura muestra el sistema FlexPod utilizado para la instalación y validación de GATK. Nosotros usamos "FlexPod Datacenter con VMware vSphere 7.0 y NetApp ONTAP 9.7 Cisco Validated Design (CVD)" durante el proceso de configuración.



En el siguiente diagrama se describen los detalles del cableado FlexPod.



La siguiente tabla enumera los componentes de hardware utilizados durante la prueba de GATK que se activa en un FlexPod. Aquí está la "Herramienta de matriz de interoperabilidad de NetApp" (IMT) y. "Lista de compatibilidad de hardware (HCL) de Cisco".

Capa	Familia de productos	Cantidad y modelo	Detalles
Informática	Chasis Cisco UCS 5108	1 o 2	
	Servidores blade Cisco UCS	6 B200 M5	Cada uno con 2 20 núcleos o más, 2,7 GHz y 128 GB de RAM

Capa	Familia de productos	Cantidad y modelo	Detalles
	Tarjeta de interfaz virtual (VIC) de Cisco UCS	Cisco UCS 1440	Consulte
	2 interconexiones de estructura Cisco UCS	6332	-
Red	Switches Cisco Nexus	2 switches Cisco Nexus 9332	-
Red de almacenamiento	Red IP para el acceso de almacenamiento mediante protocolos SMB/CIFS, NFS o iSCSI	Los mismos switches de red que se han descrito anteriormente	-
	Acceso a almacenamiento mediante FC	2 Cisco MDS 9148S	-
Reducida	Sistema de almacenamiento all-flash AFF A700 de NetApp	Clúster 1	Clúster con dos nodos
	Bandeja de discos	Una bandeja de discos DS224C o NS224	Totalmente lleno con 24 unidades
	SSD	24, 1,2 TB o más capacidad	-

Esta tabla enumera el software de infraestructura.

De NetApp	Familia de productos	Versión o versión	Detalles
Varios	Linux	RHEL 8.3	-
	Windows	Windows Server 2012 R2 (64 bits)	-
	ONTAP de NetApp	ONTAP 9.8 o posterior	-
	Interconexión de estructura Cisco UCS	Cisco UCS Manager 4.1 o posterior	-
	Switches de las series Cisco Ethernet 3000 o 9000	Para la serie 9000, 7.0(3)I7(7) o posterior para la serie 3000, 9.2(4) o posterior	-
	Cisco FC: Cisco MDS 9132T	8.4(1a) o posterior	-
	Hipervisor	VMware vSphere ESXi 7.0	-
Reducida	Sistema de gestión de hipervisores	VMware vCenter Server 7.0 (vcsa) o versiones posteriores	-
Red	Virtual Storage Console (VSC) de NetApp	VSC 9.7 o posterior	-

De NetApp	Familia de productos	Versión o versión	Detalles
	SnapCenter de NetApp	SnapCenter 4.3 o posterior	-
	Administrador de Cisco UCS	4.1(3c) o posterior	
Hipervisor	ESXi		
Gestión	Sistema de gestión del hipervisor VMware vCenter Server 7.0 (vcsa) o versiones posteriores		
	Virtual Storage Console (VSC) de NetApp	VSC 9.7 o posterior	
	SnapCenter de NetApp	SnapCenter 4.3 o posterior	
	Administrador de Cisco UCS	4.1(3c) o posterior	

"Siguiente: Genómica - instalación y ejecución de GATK."

Genómica - instalación y ejecución de GATK

"Anterior: Componentes de hardware y software de la infraestructura de la solución."

Según el Instituto Nacional de Investigación del Genoma humano ("NHGRI"), “la genómica es el estudio de todos los genes de una persona (el genoma), incluyendo las interacciones de estos genes entre sí y con el medio ambiente de una persona. ”

Según la "NHGRI", “El ácido Deoxirribonucleico (ADN) es el compuesto químico que contiene las instrucciones necesarias para desarrollar y dirigir las actividades de casi todos los organismos vivos. Las moléculas de ADN están hechas de dos filamentos de torsión, emparejados, a menudo llamados como una hélice doble". “El conjunto completo de ADN de un organismo se llama su genoma”.

La secuenciación es el proceso de determinar el orden exacto de las bases en una cadena de ADN. Uno de los tipos más comunes de secuenciación utilizados hoy se llama secuenciación por síntesis. Esta técnica usa la emisión de señales fluorescentes para ordenar las bases. Los investigadores pueden usar la secuenciación del ADN para buscar variaciones genéticas y cualquier mutación que pueda desempeñar un papel en el desarrollo o progresión de una enfermedad mientras una persona todavía está en la etapa embrionaria.

De muestra a identificación de variante, anotación y predicción

En líneas generales, la genómica se puede clasificar en los siguientes pasos. Esta no es una lista exhaustiva:

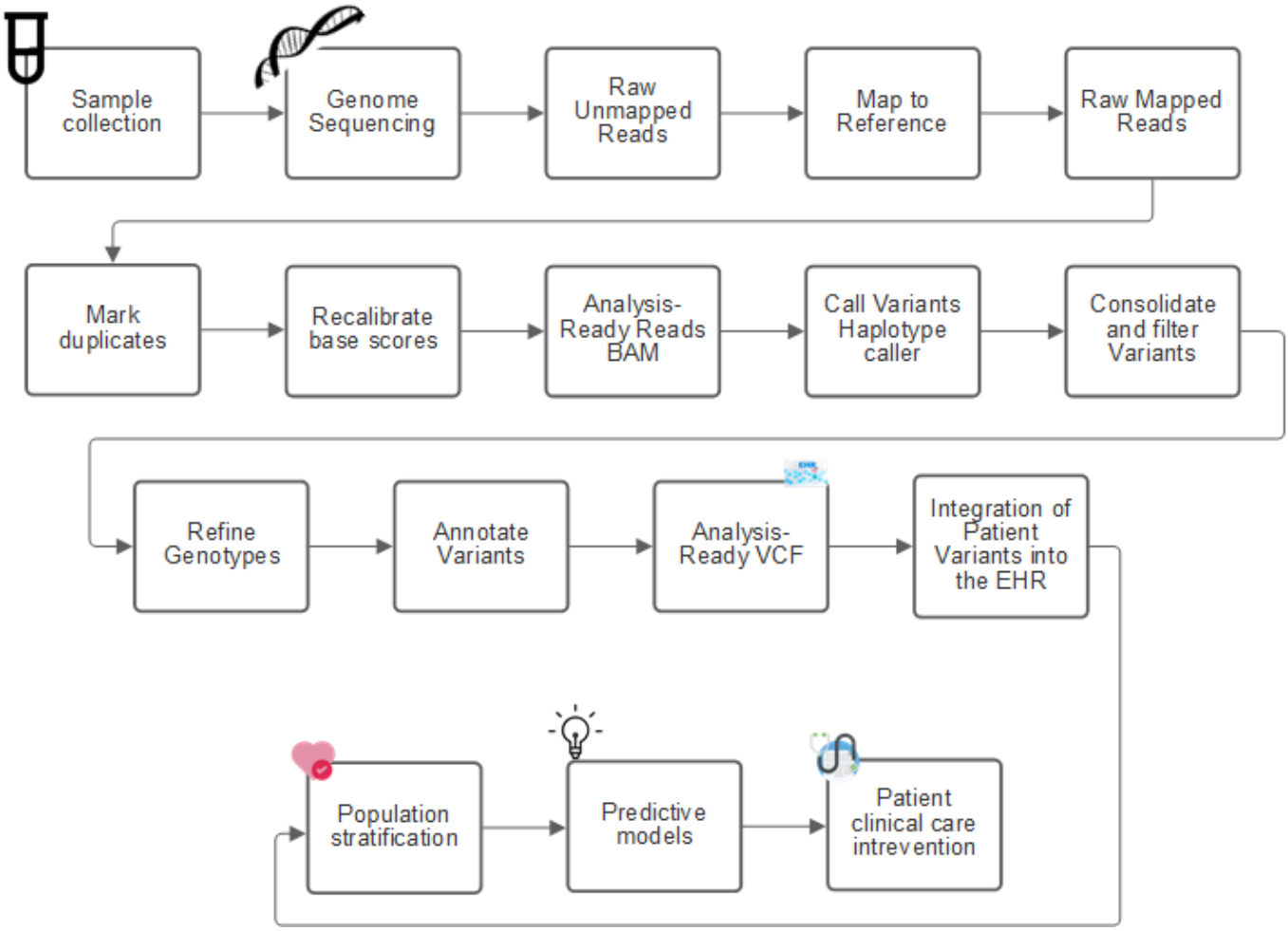
1. Recogida de muestras.
2. "Secuenciación genómica" uso de un secuenciador para generar los datos sin procesar.
3. Preprocesamiento. Por ejemplo: "deduplicación" uso "Picard".
4. Análisis genómico.

a. Asignación a un genoma de referencia.

b. "Variante" La identificación y la anotación se realizan normalmente con GATK y herramientas similares.

5. Integración en el sistema de historiales médicos electrónicos (EHR).
6. "Estratificación poblacional" e identificación de la variación genética a través de la ubicación geográfica y el origen étnico.
7. "Modelos predictivos" utilizando un polimorfismo significativo de un solo nucleótido.
8. "Validación".

En la siguiente figura se muestra el proceso desde el muestreo hasta la identificación de variantes, la anotación y la predicción.



El proyecto Genoma humano se completó en abril de 2003 y el proyecto hizo una simulación de muy alta calidad de la secuencia del genoma humano disponible en el dominio público. Este genoma de referencia inició una explosión en la investigación y el desarrollo de capacidades de genómica. Prácticamente cada dolencia humana tiene una firma en los genes de ese ser humano. Hasta hace poco, los médicos estaban aprovechando los genes para predecir y determinar defectos de nacimiento como la anemia drepanocítica, que es causada por un cierto patrón de herencia causado por un cambio en un solo gen. El tesoro de los datos que el proyecto del genoma humano ha puesto a su disposición llevó a la llegada del estado actual de las capacidades genómicas.

La genómica ofrece un amplio conjunto de ventajas. A continuación se presenta un pequeño conjunto de beneficios en los ámbitos de la salud y las ciencias de la vida:

- Mejor diagnóstico en el punto de atención
- Mejor pronóstico
- Medicina de precisión
- Planes de tratamiento personalizados
- Mejor control de enfermedades
- Reducción de acontecimientos adversos
- Mejora del acceso a las terapias
- Mejor control de enfermedades
- Participación efectiva en ensayos clínicos y mejor selección de pacientes para ensayos clínicos basados en genotipos.

La genómica es un "bestia con cuatro cabezas," debido a las demandas computacionales que se encuentran a lo largo del ciclo de vida de un conjunto de datos: adquisición, almacenamiento, distribución y análisis.

Kit de herramientas de análisis de genoma (GATK)

GATK fue desarrollado como una plataforma de ciencia de datos en el "Instituto amplio". GATK es un conjunto de herramientas de código abierto que permiten el análisis del genoma, específicamente el descubrimiento de variantes, la identificación, la anotación y el genotipado. Una de las ventajas de GATK es que el conjunto de herramientas y comandos se pueden encadenar para formar un flujo de trabajo completo. Los principales desafíos que enfrenta el instituto general son los siguientes:

- Entender las causas profundas y los mecanismos biológicos de las enfermedades.
- Identificar las intervenciones terapéuticas que actúen en la causa fundamental de una enfermedad.
- Entender la línea de visión de las variantes a la función en la fisiología humana.
- Crear estándares y políticas "marcos de trabajo" para representación de datos genómicos, almacenamiento, análisis, seguridad, etc.
- Estandarizar y socializar las bases de datos de agregación de genomas interoperables (gnomAD).
- Monitoreo, diagnóstico y tratamiento basados en el genoma de pacientes con mayor precisión.
- Ayudar a implementar herramientas que predicen enfermedades mucho antes de que aparezcan los síntomas.
- Crear y empoderar a una comunidad de colaboradores interdisciplinarios para ayudar a enfrentar los problemas más difíciles e importantes de la biomedicina.

Según GATK y el instituto general, la secuenciación del genoma debe ser tratada como un protocolo en un laboratorio de patología; cada tarea está bien documentada, optimizada, reproducible y consistente entre muestras y experimentos. El siguiente es un conjunto de pasos recomendados por el Instituto amplio, para más información, ver el "Sitio web de GATK".

Configuración de FlexPod

La validación de cargas de trabajo genómica incluye una configuración desde cero de una plataforma de infraestructura FlexPod. La plataforma FlexPod tiene una alta disponibilidad y se puede escalar de forma independiente; por ejemplo, red, almacenamiento y recursos informáticos se pueden escalar de forma independiente. Utilizamos la siguiente guía de diseño validado por Cisco como documento de arquitectura de referencia para configurar el entorno FlexPod: "FlexPod Datacenter con VMware vSphere 7.0 y ONTAP 9.7 de NetApp". Consulte los aspectos destacados de la siguiente configuración de la plataforma FlexPod:

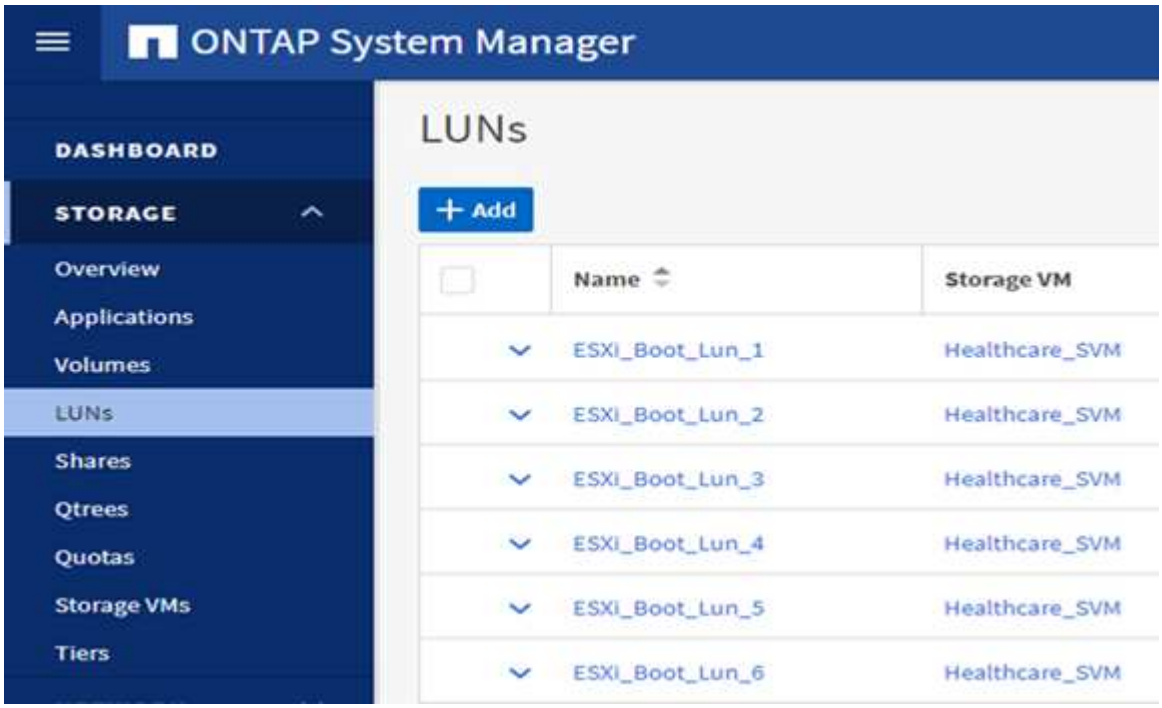
Para realizar la configuración de FlexPod Lab, realice los siguientes pasos:

1. La configuración y validación de FlexPod Lab utiliza las siguientes reservas IP4 y VLAN.

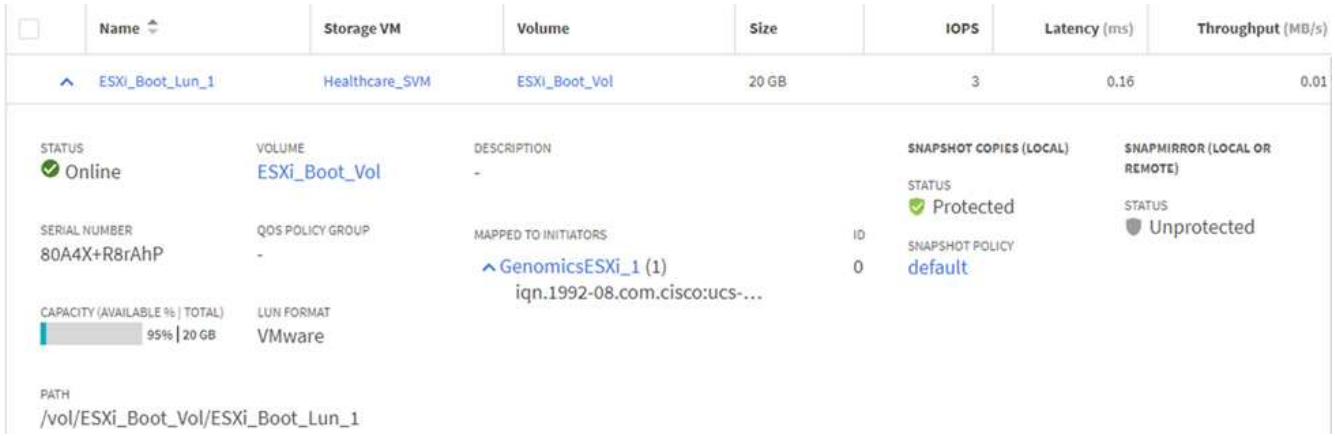
IP Reservations

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. Configure los LUN de arranque basados en iSCSI en la SVM ONTAP.

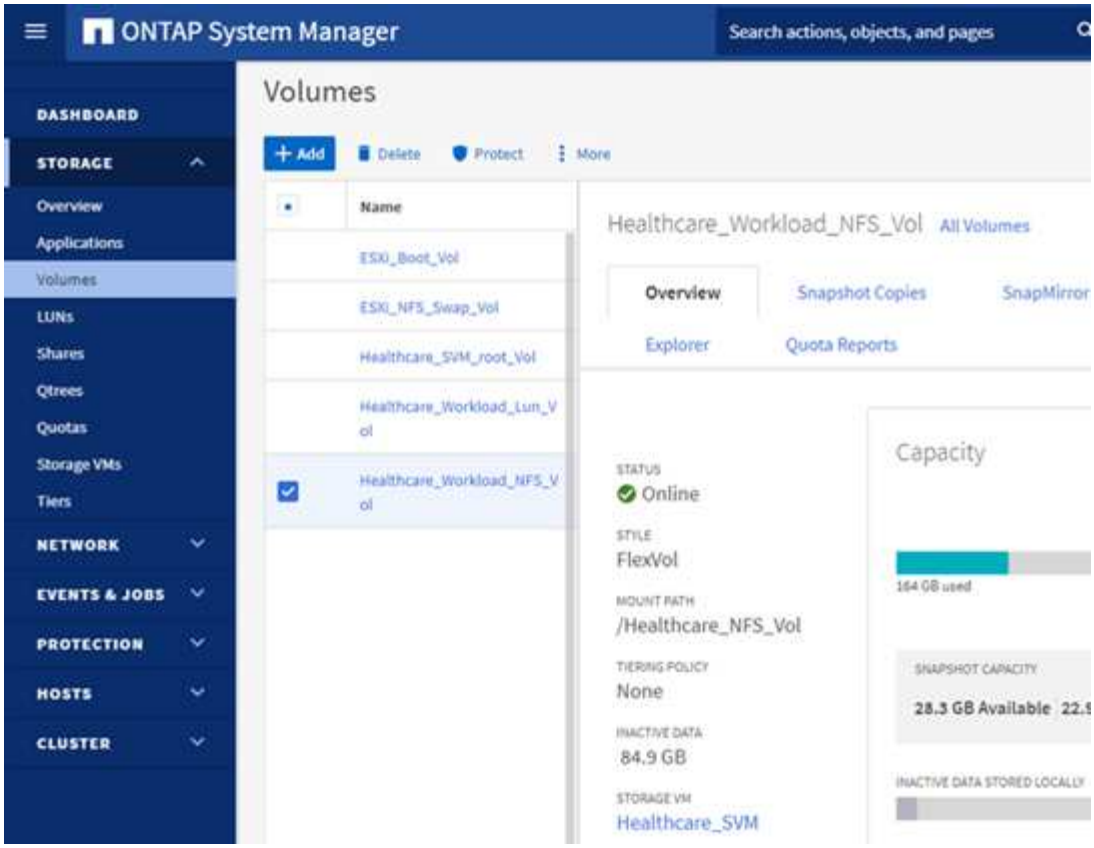


3. Asigne LUN a iGroups iSCSI.

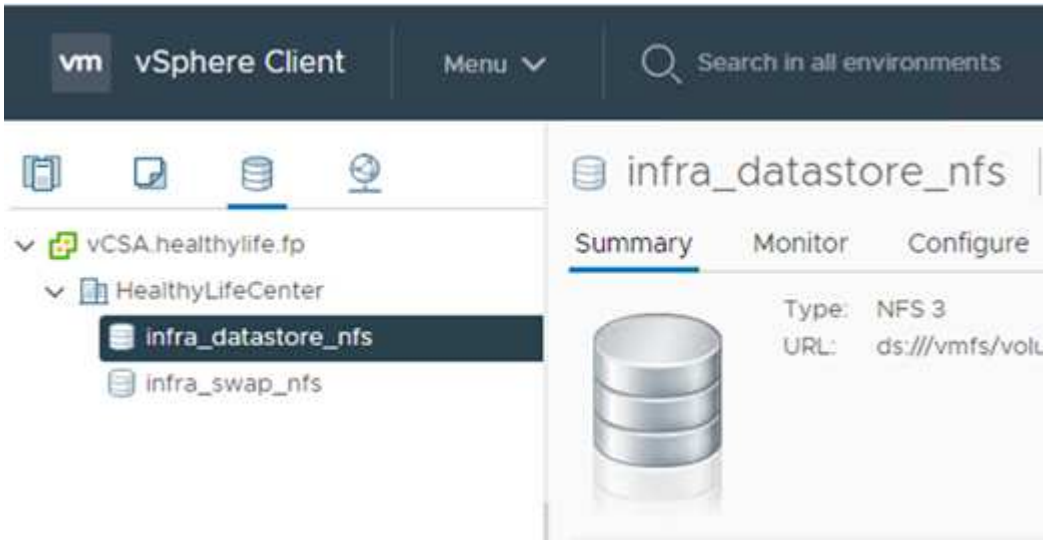


4. Instale vSphere 7.0 con arranque iSCSI.
5. Registre hosts ESXi en vCenter.

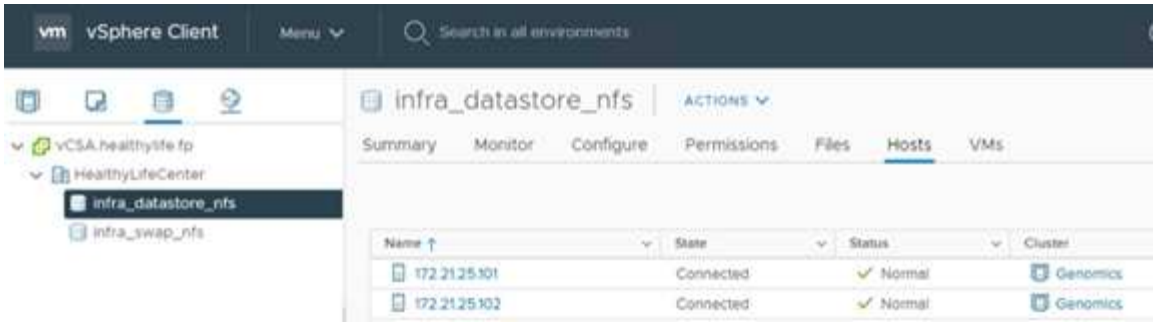
6. Aprovisione un almacén de datos NFS `infra_datastore_nfs` En el almacenamiento de ONTAP.



7. Añada el almacén de datos a vCenter.



8. Con vCenter, añada un almacén de datos NFS a los hosts ESXi.



9. Mediante el para vCenter, cree un equipo virtual Red Hat Enterprise Linux (RHEL) 8.3 para ejecutar GATK.
10. Un almacén de datos NFS se presenta a la máquina virtual y se monta en `/mnt/genomics`, Que se utiliza para almacenar ejecutables de GATK, secuencias de comandos, archivos de mapa de alineación binaria (BAM), archivos de referencia, archivos de índice, archivos de diccionario y archivos de salida para llamadas de variante.

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

Configuración y ejecución de GATK

Instale los siguientes requisitos previos en Red Hat Enterprise 8.3 Linux VM:

- Java 8 o SDK 1.8 o posterior
- Descargue GATK 4.2.0.0 de The Broad Institute "El sitio de GitHub". Los datos de la secuencia del genoma se almacenan generalmente en forma de una serie de columnas ASCII delimitadas por tabulaciones. Sin embargo, ASCII toma demasiado espacio para almacenar. Por lo tanto, un nuevo estándar evolucionó llamado un archivo BAM (*.bam). Un archivo BAM almacena los datos de secuencia en forma comprimida, indizada y binaria. Nosotros "descargado" Un conjunto de archivos BAM disponibles públicamente para la ejecución de GATK desde el "dominio público". También descargamos archivos de índice (*.bai), archivos de diccionario (*. dict) y archivos de datos de referencia (*. fasta) del mismo dominio público.

Después de la descarga, el kit de herramientas GATK tiene un archivo JAR y un conjunto de secuencias de comandos de soporte.

- `gatk-package-4.2.0.0-local.jar` ejecutable
- `gatk` archivo de script.

Descargamos los archivos BAM y los correspondientes archivos de índice, diccionario y genoma de referencia para una familia que consistía en archivos padre, madre e hijo *.bam.

Motor Cromwell

Cromwell es un motor de código abierto orientado a flujos de trabajo científicos que permite la gestión del flujo de trabajo. El motor Cromwell puede funcionar en dos "modos", Modo servidor o modo de ejecución de un único flujo de trabajo. El comportamiento del motor Cromwell se puede controlar mediante el "Archivo de configuración del motor Cromwell".

- **Modo servidor.** activado "Tranquilo" Ejecución de flujos de trabajo en motor Cromwell.

- **Modo de ejecución.** el modo de ejecución es más adecuado para ejecutar flujos de trabajo individuales en Cromwell, "ref" Para obtener un conjunto completo de opciones disponibles en el modo de ejecución.

Utilizamos el motor Cromwell para ejecutar los flujos de trabajo y las tuberías a escala. El motor Cromwell utiliza un motor fácil de usar "idioma de descripción del flujo de trabajo" Lenguaje de secuencias de comandos basado en (WDL). Cromwell también admite un segundo estándar de secuencias de comandos de flujo de trabajo denominado Common Workflow Language (CWL). A lo largo de este informe técnico, utilizamos el programa WDL. El WDL fue originalmente desarrollado por el Instituto amplio para oleoductos de análisis de genoma. Los flujos de trabajo de WDL se pueden implementar utilizando varias estrategias, incluidas las siguientes:

- **Encadenamiento lineal.** como su nombre indica, la salida de la tarea#1 se envía a la tarea #2 como entrada.
- **Multi-in/out.** esto es similar al encadenamiento lineal en que cada tarea puede tener múltiples salidas enviadas como entrada a tareas posteriores.
- **Scatter-gather.** esta es una de las estrategias de integración de aplicaciones empresariales (EAI) más potentes disponible, especialmente cuando se utiliza en arquitectura basada en eventos. Cada tarea se ejecuta de forma desacoplada y el resultado de cada tarea se consolida en el resultado final.

Hay tres pasos cuando se utiliza WDL para ejecutar GATK en modo independiente:

1. Validar la sintaxis con `womtool.jar`.

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

2. Generar entradas JSON.

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

3. Ejecute el flujo de trabajo con el motor Cromwell y. `Cromwell.jar`.

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

El GATK se puede ejecutar utilizando varios métodos; este documento explora tres de estos métodos.

Ejecución de GATK usando el archivo JAR

Veamos una sola versión de la ejecución de la canalización de llamadas usando la variante de haplotipo que llama.

```
[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
```

En este método de ejecución, utilizamos el archivo JAR de ejecución local GATK, utilizamos un solo comando java para invocar el archivo JAR y pasamos varios parámetros al comando.

1. Este parámetro indica que estamos invocando el `HaplotypeCaller` canalización de llamada variante.
2. `-- input` Especifica el archivo BAM de entrada.
3. `--output` especifica el archivo de salida de variante en formato de llamada variante (*.vcf) ("ref").
4. Con la `--reference` parámetro, estamos pasando un genoma de referencia.

Una vez ejecutada, los detalles de salida se pueden encontrar en la sección "Salida para la ejecución de GATK utilizando el archivo JAR."

Ejecución de GATK usando el script ./gatk

El kit de herramientas de GATK se puede ejecutar usando el `./gatk` guión. Examinemos el siguiente comando:

```
[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
```

Pasamos varios parámetros al comando.

- Este parámetro indica que estamos invocando el `HaplotypeCaller` canalización de llamada variante.
- `-I` Especifica el archivo BAM de entrada.
- `-O` especifica el archivo de salida de variante en formato de llamada variante (*.vcf) ("ref").
- Con la `-R` parámetro, estamos pasando un genoma de referencia.

Una vez ejecutada, los detalles de salida se pueden encontrar en la sección

Ejecución del GATK con el motor de Cromwell

Utilizamos el motor Cromwell para gestionar la ejecución de GATK. Examinemos la línea de comandos y sus parámetros.

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

Aquí, invocamos el comando Java pasando el `-jar` parámetro para indicar que queremos ejecutar un archivo jar, por ejemplo, `Cromwell-65.jar`. Se ha pasado el siguiente parámetro (`run`) Indica que el motor Cromwell está funcionando en modo de funcionamiento, la otra opción posible es modo de servidor. El siguiente parámetro es `*.wdl` Que el modo Run debe utilizar para ejecutar las tuberías. El siguiente parámetro es el conjunto de parámetros de entrada de los flujos de trabajo que se están ejecutando.

Esto es lo que el contenido del `ghplo.wdl` aspecto del archivo:

```
[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"
  }
}
[root@genomics1 seq]#
```

Aquí está el archivo JSON correspondiente con las entradas del motor 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]#
```

Tenga en cuenta que Cromwell utiliza una base de datos en memoria para la ejecución. Una vez ejecutado, el registro de salida se puede ver en la sección "Salida para la ejecución del GATK utilizando el motor Cromwell."

Para ver un conjunto completo de pasos sobre cómo ejecutar GATK, consulte "Documentación de GATK".

"Siguiente: Salida para la ejecución de GATK usando el archivo JAR."

Salida para la ejecución de GATK utilizando el archivo JAR

"Anterior: Genómica - instalación y ejecución de GATK."

La ejecución de GATK usando el archivo JAR produjo la siguiente salida de muestra.

```
[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]#
```

Observe que el archivo de salida se encuentra en la ubicación especificada después de la ejecución.

Salida para la ejecución de GATK usando el script ./gatk

"Anterior: Salida para la ejecución de GATK usando el archivo JAR."

La ejecución de GATK con el útil `./gatk` la secuencia de comandos produjo la siguiente salida de ejemplo.

```
[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.healthylife.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]#
```

Observe que el archivo de salida se encuentra en la ubicación especificada después de la ejecución.

"Siguiente: Salida para la ejecución de GATK con el motor Cromwell."

Salida para la ejecución del GATK utilizando el motor Cromwell

La ejecución de GATK con el motor Cromwell produjo la siguiente salida de muestra.

```
[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]#
```

"Siguiente: Configuración de GPU."

Configuración de GPU

"Anterior: Salida para la ejecución de GATK con el motor Cromwell."

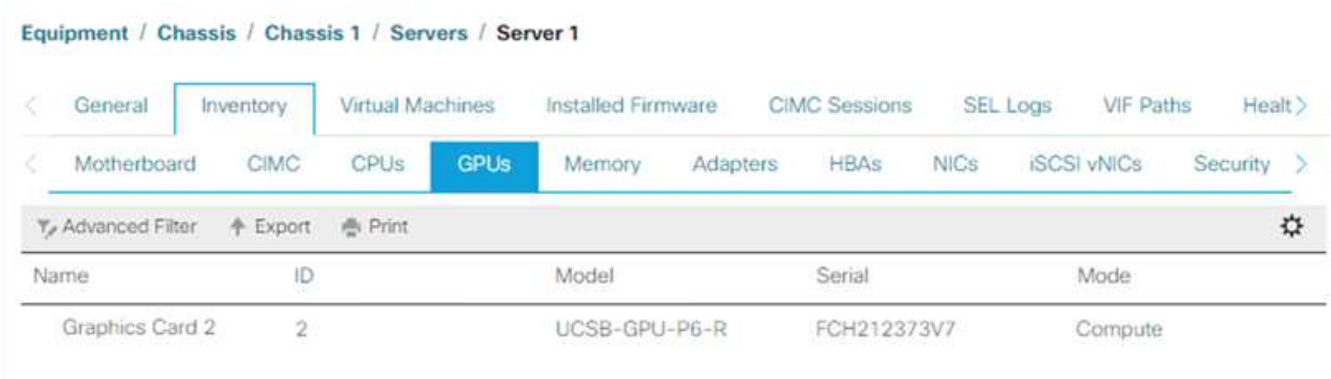
En el momento de su publicación, la herramienta GATK no tiene soporte nativo para la ejecución basada en GPU en las instalaciones. Se proporcionan las siguientes instrucciones y configuración para que los lectores puedan comprender lo sencillo que es utilizar FlexPod con una GPU NVIDIA Tesla P6 montada en la parte posterior utilizando una tarjeta intermedia PCIe para GATK.

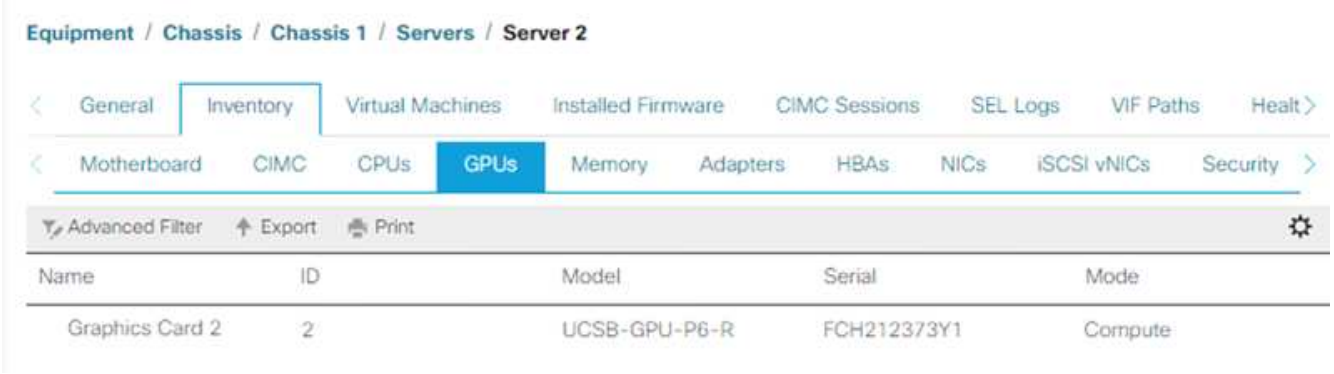
Utilizamos el siguiente diseño validado por Cisco (CVD) como arquitectura de referencia y guía de prácticas recomendadas para configurar el entorno FlexPod de modo que podamos ejecutar aplicaciones que utilizan GPU.

- "Centro de datos FlexPod para IA/ML con Cisco UCS 480 ML para aprendizaje profundo"

Este es un conjunto de puntos clave durante esta configuración:

1. Utilizamos una GPU PCIe NVIDIA Tesla P6 en una ranura intermedia en los servidores UCS B200 M5.





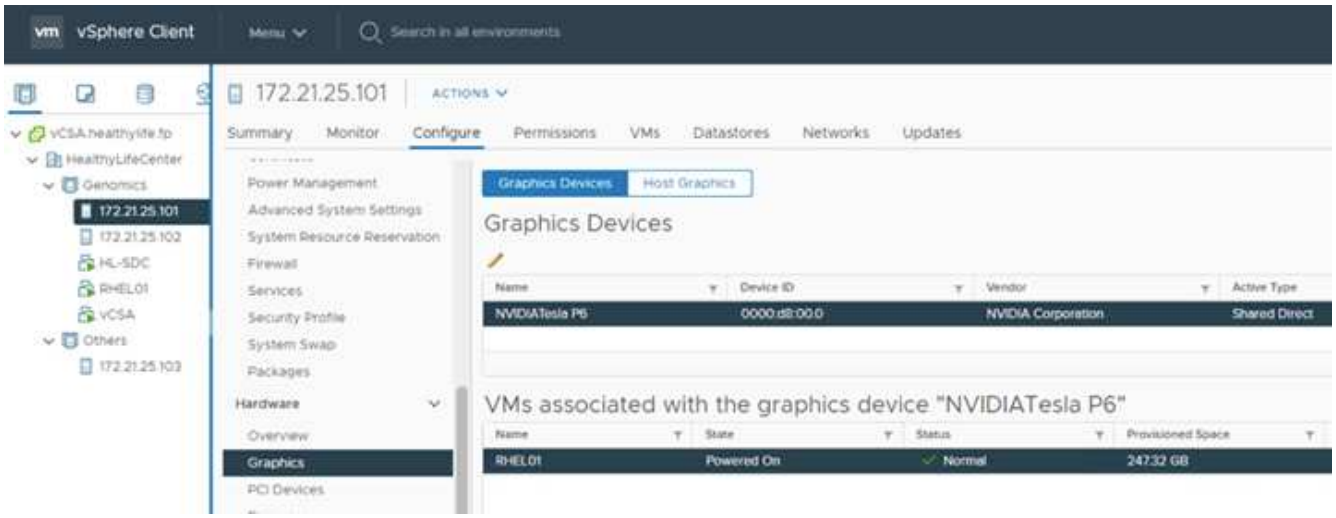
2. Para esta configuración, nos registramos en el portal de partners de NVIDIA y obtuvimos una licencia de evaluación (también conocida como autorización) para poder utilizar las GPU en el modo de computación.
3. Hemos descargado el software vGPU de NVIDIA requerido desde el sitio web para partners de NVIDIA.
4. Hemos descargado el derecho `*.bin` Archivo del sitio web del partner NVIDIA.
5. Instalamos un servidor de licencias NVIDIA vGPU y añadimos los derechos al servidor de licencias mediante `*.bin` Archivo descargado del sitio del partner de NVIDIA.
6. Asegúrese de elegir la versión correcta del software vGPU de NVIDIA para su puesta en marcha en el portal de partners de NVIDIA. Para esta configuración utilizamos la versión 460.73.02 del controlador.
7. Este comando instala el "Administrador vGPU de NVIDIA" En 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:
```

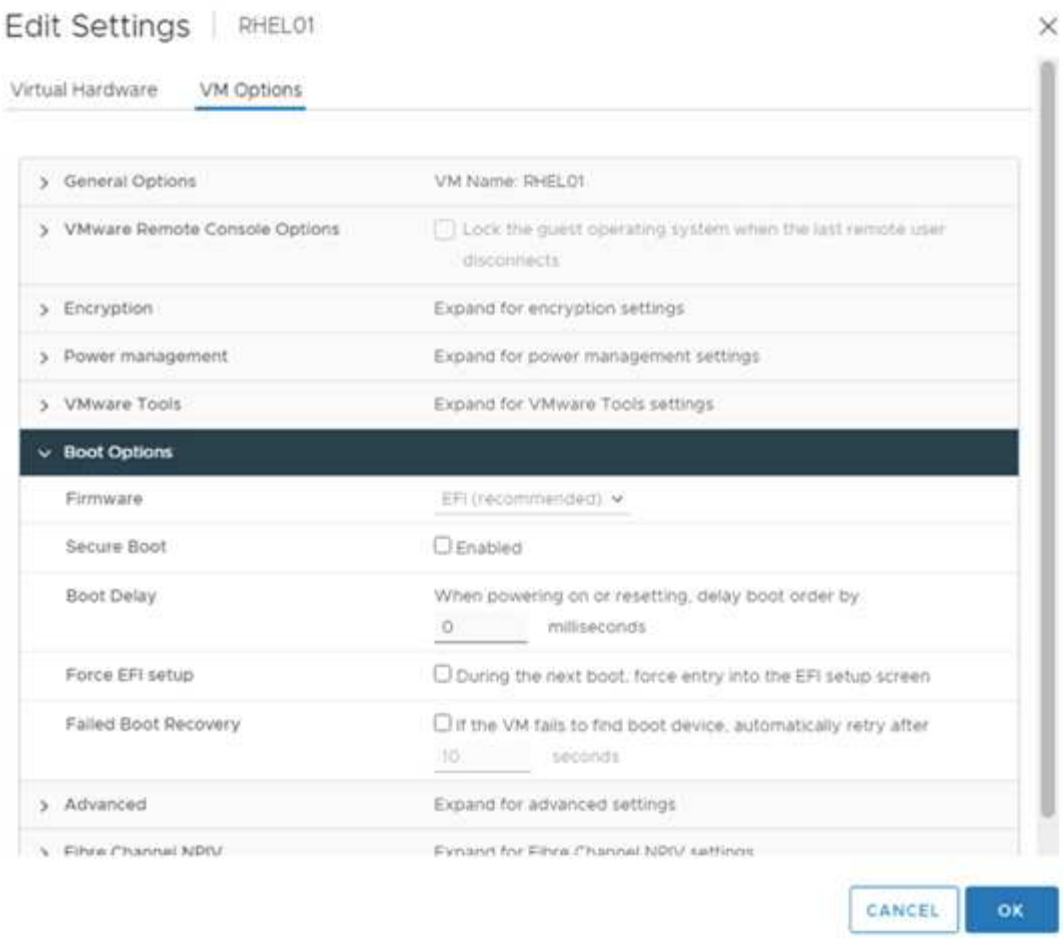
8. Después de reiniciar el servidor ESXi, ejecute el siguiente comando para validar la instalación y comprobar el estado de las GPU.

```
[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. Con vCenter, "configurar" La configuración del dispositivo gráfico a “Shared Direct”.



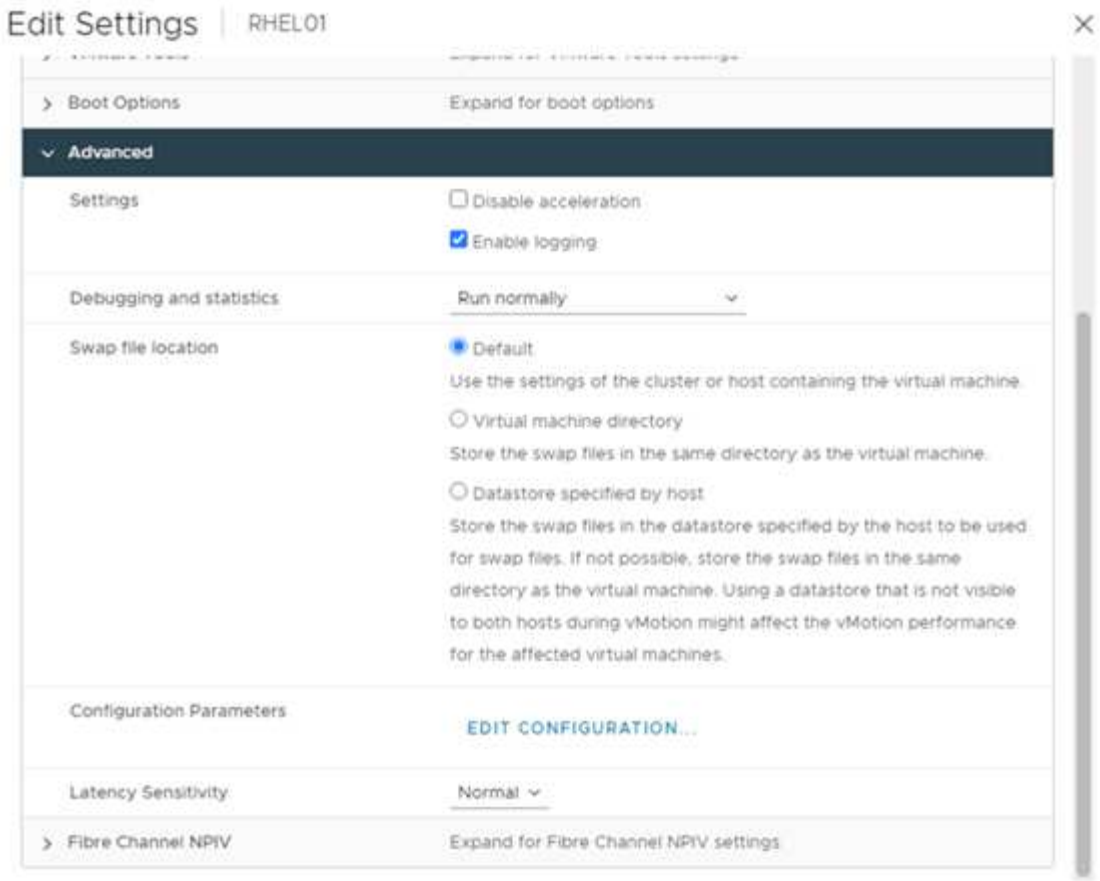
10. Asegúrese de que el arranque seguro está desactivado para el equipo virtual RedHat.
11. Asegúrese de que el firmware de VM Boot Options esté establecido en EFI ("ref").

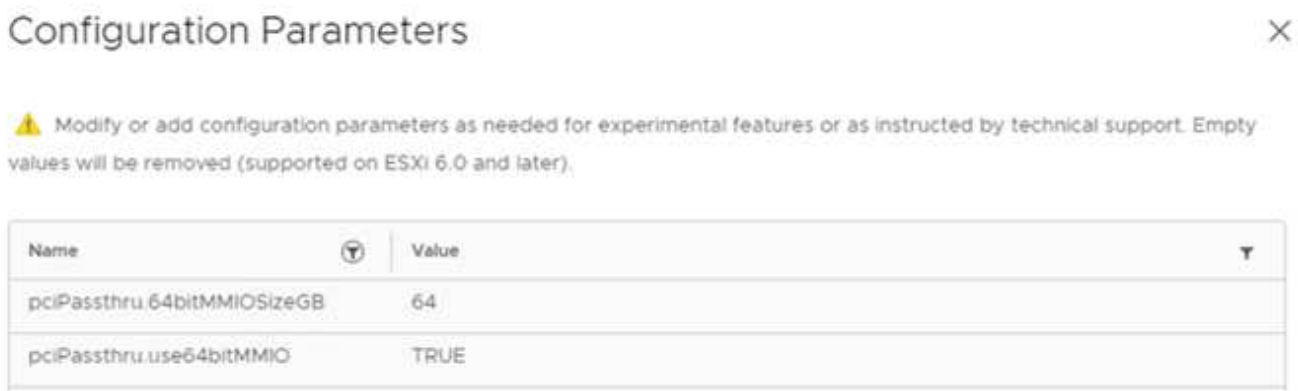


12. Asegúrese de que los siguientes PARÁMETROS se agreguen a la VM Options Advanced Edit Configuration. El valor de `pciPassthru.64bitMMIOSizeGB` El parámetro depende de la memoria de la GPU y del número de GPU asignadas a la máquina virtual. Por ejemplo:

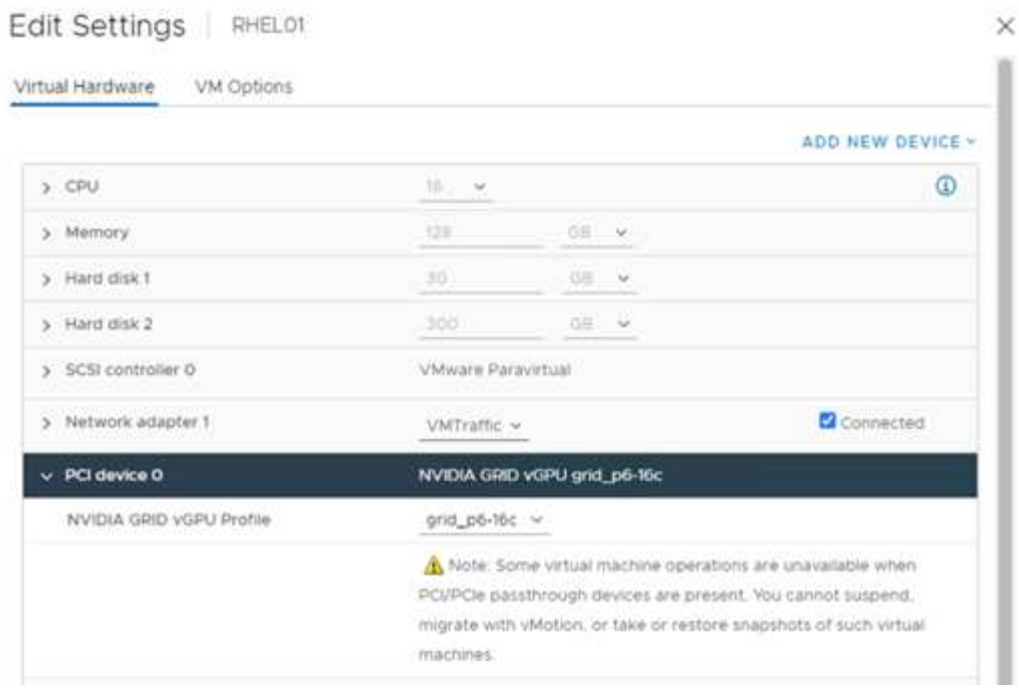
    a. Si se asigna una máquina virtual 4 GPU V100 de 32 GB, este valor debe ser 128.

    b. Si a una máquina virtual se le asignan 4 GPU P6 de 16 GB, este valor debe ser 64.





13. Cuando se añaden vGPU como un nuevo dispositivo PCI a la máquina virtual en vCenter, asegúrese de seleccionar NVIDIA GRID vGPU como tipo de dispositivo PCI.
14. Elija el perfil correcto de GPU que Suite la GPU que se va a usar, la memoria de la GPU y el objetivo de uso: Por ejemplo, los gráficos frente al procesamiento.



15. En el equipo virtual RedHat Linux, los controladores NVIDIA se pueden instalar ejecutando el siguiente comando:

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

16. Compruebe que se está reportando el perfil correcto de vGPU ejecutando el siguiente comando:

```
[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. Después de reiniciar, compruebe que se ha informado de la unidad vGPU de NVIDIA correcta junto con las versiones del controlador.

```
[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. Asegúrese de que la IP del servidor de licencias está configurada en el equipo virtual en el archivo de configuración de grid vGPU.
    a. Copie la plantilla.

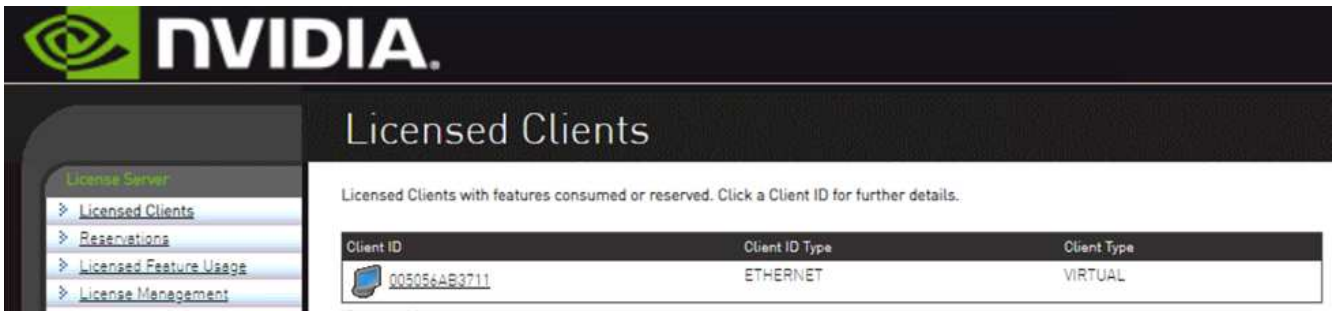
```
[root@genomics1 genomics]# cp /etc/nvidia/gridd.conf.template
/etc/nvidia/gridd.conf
```

b. Edite el archivo `/etc/nvidia/rid.conf`, Agregue la dirección IP del servidor de licencias y establezca el tipo de característica en 1.

```
ServerAddress=192.168.169.10
```

```
FeatureType=1
```

19. Después de reiniciar la máquina virtual, deberá ver una entrada en clientes con licencia en el servidor de licencias como se muestra a continuación.



20. Consulte la sección Configuración de soluciones para obtener más información sobre la descarga del software GATK y Cromwell.
21. Una vez que GATK puede utilizar las GPU en las instalaciones, el lenguaje de descripción del flujo de trabajo `*. wdl` tiene los atributos de tiempo de ejecución como se muestra a continuación.

```
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 addtional_disk_space_gb = 50
  }
  Int disk_size = ceil(size(input_bam, "GB")) + addtional_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}"
  }
}
```

"Siguiente: Conclusión."

Conclusión

"Anterior: Configuración de GPU."

Muchas organizaciones sanitarias de todo el mundo se han estandarizado en FlexPod como plataforma común. Con FlexPod, puede implementar capacidades en el sector sanitario con confianza. FlexPod con ONTAP de NetApp viene de serie con la capacidad de implementar un conjunto de protocolos líderes en el sector desde el primer momento. Con independencia del origen de la solicitud de ejecución de la genómica de un paciente dado, la interoperabilidad, la accesibilidad, la disponibilidad y la escalabilidad vienen de

serie con una plataforma FlexPod. Una solución estandarizada con la plataforma FlexPod es que la cultura de la innovación llega a ser contagiosa.

Dónde encontrar información adicional

Si quiere obtener más información sobre el contenido de este documento, consulte los siguientes documentos y sitios web:

- Centro de datos FlexPod para IA/ML con Cisco UCS 480 ML para aprendizaje profundo

  "https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/flexpod_480ml_aiml_deployment.pdf"

- FlexPod Datacenter con VMware vSphere 7.0 y ONTAP 9.7 de NetApp

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

- Centro de documentación de ONTAP 9

  "http://docs.netapp.com"

- Ágil y eficiente: Cómo impulsa FlexPod la modernización del centro de datos

  "https://www.flexpod.com/idc-white-paper/"

- IA en la sanidad

  "https://www.netapp.com/us/media/na-369.pdf"

- FlexPod para el sector sanitario facilita su transformación

  "https://flexpod.com/solutions/verticals/healthcare/"

- FlexPod de Cisco y NetApp

  "https://flexpod.com/"

- IA y análisis en el sector sanitario (NetApp)

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

- IA en el sector sanitario las opciones de infraestructura inteligente aumentan el éxito

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

- Centro de datos FlexPod con ONTAP 9.8, conector de almacenamiento ONTAP para Cisco Intersight y Cisco Intersight Managed Mode.

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

- Centro de datos FlexPod con plataforma OpenStack Red Hat Enterprise Linux

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

Historial de versiones

Versión	Fecha	Historial de versiones del documento
Versión 1.0	Noviembre de 2021	Versión inicial.