El CRAG (Centro de Investigación en Agrigenómica) es un centro público dedicado a la investigación científica de frontera en el ámbito agrigenómico. Para satisfacer la demanda de velocidad y escalabilidad de su infraestructura de computación de alto rendimiento, que realiza tareas tan exigentes como la secuenciación del genoma de animales y plantas, el CRAG necesitaba remplazar su entorno de almacenamiento basado en disco. En 2019, la solución Pure Storage® FlashBlade™ les ha ayudado a multiplicar por 4 la velocidad de escritura, a reducir un 60 % el tiempo requerido para realizar copias de seguridad completas y a aumentar la estabilidad, eliminando las paradas imprevistas por sobrecargas del almacenamiento.

TRANSFORMACIÓN DE NEGOCIOS

El CRAG fue capaz de actualizar su entorno de computación de alto rendimiento (HPC, por sus siglas en inglés) para dotarlo de la velocidad, escalabilidad y estabilidad que requieren los proyectos de investigación de sus usuarios, gracias a la solución Pure FlashBlade.

REGION / PAIS

EMEA / España

INDUSTRIA

Agrigenómica

"Otras soluciones se han adaptado al uso de la tecnología flash, pero Pure Storage se ha construido por y para el almacenamiento flash, y esa es la gran diferencia".

Gonzalo Vera,
Jefe de Informática Científica

INVESTIGACIÓN AGRIGENÓMICA DE FRONTERA

El Centro de Investigación en Agrigenómica (CRAG) es un consorcio público formado por el Consejo Superior de Investigaciones Científicas (CSIC), el Instituto de Investigación y Tecnología Agroalimentarias (IRTA), la Universidad Autónoma de Barcelona (UAB) y la Universidad de Barcelona (UB). Desde 2011, ocupa unas modernas instalaciones en el campus de la Universidad Autónoma de Barcelona, donde más de 200 investigadores y personal de apoyo realizan una intensa actividad de investigación básica y aplicada en varios programas. Estos programas abarcan desde el metabolismo y la respuesta al estrés de las plantas al genoma de los animales de granja, y están respaldados por una plataforma tecnológica de vanguardia.

La investigación científica de frontera que se lleva a cabo en el CRAG usa grandes volúmenes de datos y requiere una infraestructura técnica avanzada. Para ello disponen de un centro de datos propio, con un clúster de computación de alto rendimiento (HPC) y varios tipos de nodos interconectados que acceden a una plataforma de almacenamiento compartido. A diario, un elevado número de usuarios concurrentes demandan grandes recursos de almacenamiento con altas velocidades, flexibilidad y escalabilidad. Sus necesidades pueden variar radicalmente con cada nuevo experimento o análisis que se realiza.

La plataforma de almacenamiento anterior, que venía de 2012, estaba totalmente basada en discos duros y ya no satisfacía las necesidades del centro. Tal y como explica Gonzalo Vera, jefe de Informática Científica del CRAG: "Estaba claro que teníamos que pasar de una solución basada 100% en disco a una solución híbrida que combinara la rapidez del almacenamiento flash para los datos en caliente o datos de análisis, con una solución basada en discos para el almacenamiento de los datos en frío o datos archivados. Al ser un centro público debemos conservar durante un periodo de diez años los datos de todos los proyectos".

UNA SOLUCIÓN QUE ENCAJA A LA PERFECCIÓN

Para conocer el estado del arte de la tecnología de almacenamiento flash, los responsables de informática científica del CRAG hicieron un estudio de mercado en el que destacó el Cuadrante Mágico de Gartner de Almacenamiento de Estado Sólido. Aunque valoraron otros proveedores, la posición de Pure Storage como "líder visionario" en el Cuadrante de Gartner llamó su atención. De modo que contactaron con los responsables de la compañía en España para solicitar más información: "Vimos que, desde el punto de vista de la arquitectura, la solución de Pure Storage

COMPAÑÍA:

CRAG
www.cragenomica.es

CASO DE USO:

- Computación de alto rendimiento (HPC) – secuenciación genómica

RETOS:

- Modernizar la plataforma de almacenamiento en disco.
- Mantener la topología actual (NFS y KVM).
- Reducir la ventana de copia de seguridad.
- Garantizar la máxima fiabilidad.

TRANSFORMACIÓN DE TI:

- Velocidad de escritura 4 veces superior (hasta 8 GiB/s vs 2,4 GiB/s anterior).
- Reducción del 60 % en la ventana de copia de seguridad.
- Nuevas funcionalidades, como snapshots.

> "El límite de saturación de la plataforma es tan alto que ni siquiera hemos llegado a vislumbrarlo. Tenemos la certeza de que nuestro almacenamiento no va a fallar".
>
> Gonzalo Vera,
> *Jefe de Informática Científica*

era idéntica a lo que estábamos intentando montar y por eso nos pusimos en contacto con ellos. Nos gustó especialmente que está diseñada totalmente con flash; no hay ni rastro de tecnologías de disco antiguas. Otras soluciones se han adaptado al uso de la tecnología flash, pero Pure Storage se ha construido por y para el almacenamiento flash y esa es la gran diferencia", explica Vera.

Finalmente, el CRAG optó por un centro de datos Pure FlashBlade, donde se almacenan todos los datos científicos del clúster del centro. Ocho blades de 17 TB están conectados por enlaces de 40GbE a una red de 100GbE. La plataforma mantiene el sistema de archivos NFS existente y las cargas de trabajo mixtas generadas en los nodos de cómputo del clúster científico y los nodos virtuales soportados por el hipervisor KVM.

Por política interna, los responsables de informática científica del centro quieren mantener un control total sobre todo el hardware y software de su infraestructura y depender lo mínimo posible de sus proveedores. Por esa razón, el equipo técnico del CRAG participó directamente en la instalación de FlashBlade, que fue rápida y no planteó dificultades especiales. "La configuramos por la mañana y por la tarde la conectamos a nuestro entorno. Tardamos un par de días en organizar las carpetas, realizamos una parada de 12 horas para sincronizar los archivos y desde entonces la solución está funcionando sin problemas", resume Vera.

AUMENTO DE VELOCIDAD Y ELIMINACIÓN DE CUELLOS DE BOTELLA

La instalación de Pure FlashBlade ha satisfecho las expectativas del departamento de TI del CRAG al proporcionar más velocidad, menos latencia, nuevas funcionalidades y, sobre todo, fiabilidad.

La velocidad de escritura de la solución anterior tenía un pico de 2,4 GiB/s, mientras que con la plataforma de Pure Storage ya se han alcanzado velocidades de 4 GiB/s y la configuración está dimensionada para alcanzar picos de hasta 8 GiB/s. Este rendimiento ha permitido eliminar los cuellos de botella que aparecían con las altas cargas del clúster de cálculo. También ha posibilitado agilizar otras tareas, como las copias de seguridad o el muestreo de datos de ocupación que se realiza cada dos horas para facturar los servicios de TI a los usuarios internos.

"La latencia de la cabina es extremadamente baja y eso que trabajamos con carpetas que pueden tener más de 200 millones de archivos. Antes tardábamos días en saber qué había dentro de un directorio; ahora lo sabemos en cuestión de minutos. Sin tocar nada más, hemos reducido el tiempo necesario para realizar una copia de seguridad completa de todo nuestro entorno, de las 250 horas que tardábamos antes a apenas 100 horas, y realizando algunos cambios esperamos hacerlo en menos de 24 horas", señala el jefe de Informática Científica del CRAG.

Otra ventaja es que, con la nueva plataforma de almacenamiento de Pure Storage, los responsables técnicos del CRAG ahora disponen de nuevas funcionalidades que antes no tenían. Por ejemplo, los snapshots, copias instantáneas de los volúmenes que permiten una recuperación completa de los datos en cualquier momento. Además, los especialistas de Pure Storage han seguido visitándoles tras la implementación para optimizar todavía más la solución.

"Lo más importante para nosotros es la tranquilidad que nos proporciona la solución, porque sabemos que, aunque aumentemos mucho la carga de la plataforma, no la llegaremos a saturar", concluye Gonzalo Vera, jefe de Informática Científica del CRAG. "El límite de saturación es tan alto que no hemos llegado a vislumbrarlo. Tenemos la certeza de que nuestro almacenamiento no va a fallar".

PURESTORAGE®
info@purestorage.com
www.purestorage.com/customers

© 2019 Pure Storage, Inc. All rights reserved. Pure Storage, FlashBlade, and the P Logo are trademarks of Pure Storage, Inc. All other trademarks are the property of their respective owners. ps_cs_CRAG_es_01