



Reto empresarial

El Centre de Regulació Genòmica (CRG) tenía como objetivo ayudar a los centros de investigación de todo el mundo a compartir sus datos de secuencias genómicas. ¿Cómo se pueden transmitir estas grandes cantidades de datos de forma rápida y económica?

Transformación

Con una plataforma de transferencia de archivos de alta velocidad IBM® Aspera® el CRG puede compartir sus archivos a una velocidad máxima de 3 Gbps. De este modo, los investigadores disponen de un acceso rápido a los datos que necesitan para continuar con sus vanguardistas proyectos genéticos.

Beneficios:

500 TB

de datos de secuenciación en paquetes listos para compartir

3 Gbps

velocidad de transferencia garantizada

Soporte α

investigaciones genéticas de vanguardia

El Centre de Regulació Genòmica respalda una innovadora investigación con una transferencia de archivos a la velocidad máxima

El Centre de Regulació Genòmica (en castellano, Centro de Regulación Genómica) se fundó en el año 2000 y en la actualidad es un centro de investigación biomédica de referencia a nivel internacional. La organización, que cuenta con 600 empleados, tiene como objetivo el desarrollo de conocimientos para beneficio de la sociedad, la sanidad pública y la prosperidad económica.

“Nuestro trabajo con TmediaT para implementar IBM Aspera supuso un impulso crucial para el Archivo Europeo del Genoma-Fenoma”.

David Camargo

Director de TI

Centre de Regulació Genòmica

Compartir



Datos genéticos para todo el mundo

Actualmente, investigadores de todo el mundo dedican sus esfuerzos a identificar en qué medida afecta la variación de la composición genética de los individuos a la salud humana. A medida que las tecnologías de secuenciación genética se hacen más rápidas, asequibles y precisas, el volumen de datos disponible también aumenta a gran velocidad. ¿Cómo pueden los centros de investigación compartir estos datos de forma rápida, adecuada y económica?

David Camargo, Director de TI del Centre de Regulació Genòmica, continúa: “Aunque el coste de la secuenciación genética está decreciendo, la recopilación de datos para investigaciones para la detección, prevención y tratamiento de enfermedades sigue requiriendo una inversión considerable tanto de tiempo, como de recursos.

Para solventar este problema, el CRG se ha asociado con el Instituto Europeo de Bioinformática (EMBL-EBI) para mejorar los servicios del Archivo Europeo del Genoma-Fenoma (en lo sucesivo, EGA) y así ofrecer un acceso más sencillo a investigadores de todo el mundo. El propósito era mejorar el servicio del EGA de almacenamiento e intercambio de datos genómicos y fenotípicos provenientes de proyectos de investigación biomédicos. De este modo, los grupos de investigación pueden aprovechar datos de secuenciación existentes y relevantes para sus campos de estudio y, al mismo tiempo, se evita la duplicación de esfuerzos, se favorecen las investigaciones avanzadas y se reducen las barreras de participación de las instituciones más pequeñas”.

Ángel Carreño Torres, Gerente de Sistemas y Operaciones en el Centre de Regulació Genòmica y en el Archivo Europeo del Genoma-Fenoma, prosigue: “El EGA cuenta con un amplio abanico de formatos de datos, desde secuencias de nucleótidos de ADN sin procesar a formatos más depurados que contienen información sobre la relación entre las secuencias y el genoma de referencia humano.

Uno de los desafíos más recientes al que nos enfrentamos fueron el crecimiento acelerado y el gran tamaño del volumen y de la variedad del archivo del EGA, que en la actualidad contiene más de dos petabytes de datos. Para realizar su trabajo con eficacia, los centros de investigación requieren grandes cantidades de datos a gran velocidad, que pueden distribuirse en múltiples archivos pequeños, en un número reducido de archivos muy grandes o en una mezcla de los dos. De hecho, uno de nuestros estudios más solicitados está formado por 500 terabytes de datos genéticos. Desde el principio tuvimos claro que los protocolos tradicionales de transferencia de archivos no serían capaces de alcanzar el rendimiento necesario”.

Selección de una solución demostrada

Durante varios años, EMBL-EBI ha confiado en las soluciones de IBM Aspera High-Speed File Transfer para lograr satisfacer sus necesidades de transferencias internas de archivos. Cuando el CRG se unió al proyecto, la organización decidió realizar un ejercicio de prueba de concepto para decidir si la solución de IBM seguía siendo la plataforma correcta para conseguir los objetivos del proyecto.

“Nuestros colegas del EMBL-EBI ya usaban IBM Aspera de forma generalizada, así que nosotros queríamos averiguar si nosotros también podríamos ampliar los beneficios que el EMBL-EBI había visto en nuestro trabajo en el EGA” recuerda Ángel Carreño Torres.

El CRG contó con consultores expertos del Business Partner de IBM, TmediaT, para la prueba de concepto para llevar a cabo una prueba exhaustiva de la solución IBM Aspera. Durante esta fase, el CRG y TmediaT descubrieron, informaron y reprodujeron varios problemas técnicos, lo que permitió a la organización arreglarlos de antemano y reducir el riesgo de retraso de los proyectos.

“Para nosotros, trabajar con el equipo de TmediaT fue una experiencia muy positiva” comenta David Camargo. “Nos ayudaron a confirmar que IBM Aspera puede cumplir los niveles objetivos de rendimiento del EGA. Además, TmediaT organizó una reunión técnica para proporcionarnos los datos técnicos que necesitábamos sobre la API de Aspera y los procesos de desarrollo. Toda inversión en un componente de software principal supone una gran decisión y nuestra colaboración con TmediaT nos hizo estar seguros de que habíamos elegido correctamente”.

Transferencia de archivos a alta velocidad

Tras los resultados positivos de la fase de prueba, un equipo mixto del EGA y del CRG implementó IBM Aspera High-Speed File Transfer en un entorno de alto rendimiento y protección informática.

“La investigación de secuenciación de ADN puede producir grandes cantidades de datos extremadamente rápido. Hace 15 años, para secuenciar el genoma humano eran necesarios casi 3 mil millones de dólares de inversión y 13 años de investigación, ahora es posible hacer la misma tarea en cuestión de horas por menos de 1.000 dólares” añade Ángel Carreño Torres.

Hoy en día, el EGA y el CRG cuentan con la plataforma de alto rendimiento y segura que necesitan para conservar y compartir datos genéticos de gran valor, con otras instituciones de todo el mundo.

“Cuando un equipo de investigación recopila datos de secuenciación anónimos de individuos que autorizan su publicación para el uso científico, pueden subirse al EGA por nuestro portal en línea” explica Ángel Carreño Torres. “Otros centros de investigación pueden solicitar el uso de los datos. La organización que realizó el estudio en el que se recopilaban los datos es la encargada de revisar dichas solicitudes.

Una vez que la solicitud quede aprobada, volvemos a cifrar los datos con una clave privada que se concede previa solicitud y los cargamos en un contenedor. El solicitante podrá descargar posteriormente los contenidos del contenedor mediante el cliente de IBM Aspera, descifrarlo en un almacén seguro y realizar sus análisis”.

Un impulso a la investigación innovadora

Gracias a la solución IBM Aspera, el EGA y el CRG están cumpliendo su objetivo de hacer que los grupos de investigación de todo el mundo tengan acceso a datos de investigaciones a gran velocidad y escala.

“Nuestra velocidad de transferencia varía según el número de usuarios accediendo al sistema simultáneamente. Por lo general, contamos con una velocidad de 1 Gbps pero tenemos picos de 3 Gbps cuando la demanda del sistema es más elevada de lo habitual” comenta Ángel Carreño Torres. “Aunque hemos visto otras soluciones de transferencia de archivos que pueden ofrecer este excelente nivel de rendimiento, ninguna de ellas ofrecía un funcionamiento tan sencillo como el que obtenemos con IBM Aspera. Nuestros solicitantes solo tienen que instalar el cliente Aspera en sus sistemas y comenzar a descargar los datos a los que puedan acceder”.

Ángel Carreño Torres prosigue: “Es muy difícil pasar por alto los niveles de rendimiento que puede ofrecer IBM Aspera.

En una ocasión, observamos que el rendimiento de un solicitante descendió de repente y nos pusimos en contacto con ellos para averiguar si había algo que pudiéramos hacer para ayudarles. Finalmente, resultó que la transferencia de archivo era tan rápida y estaba ocupando tal ancho de banda que había saturado su red, ¡así que tuvimos que seguir con la transferencia a una velocidad inferior! Incluso hemos oído que algunos de nuestros solicitantes pudieron asegurar la inversión gracias a la mejora del hardware de su red como resultado de sus experiencias positivas con IBM Aspera”.

“En los próximos años, los centros que confíen en nuestros datos seguirán mejorando su infraestructura informática, red y almacenamiento. Creemos que nuestra solución les permitirá maximizar el valor de su inversión en estas tecnologías al aprovechar los beneficios de las velocidades de transferencia de archivos que ofrece el IBM Aspera”.

David Camargo concluye: “Trabajar con TmediaT para implementar IBM Aspera ha supuesto un impulso para el proyecto del EGA y nos ha permitido conseguir nuestro objetivo de respaldar investigaciones genéticas innovadoras”.

“Es muy difícil pasar por alto los niveles de rendimiento que puede ofrecer IBM Aspera”.

Ángel Carreño Torres, Gerente de Sistemas y Operaciones en el Centre de Regulació Genòmica y en el Archivo Europeo del Genoma-Fenoma

Componentes

- IBM® Aspera® High-Speed File Transfer

Dé el siguiente paso

Para conocer más información sobre las soluciones de TmediaT, visite: tmediat.es

Para conocer más información sobre el Archivo Europeo del Genoma-Fenoma, visite: <https://ega-archive.org>

Para conocer más información sobre las soluciones de IBM Aspera, póngase en contacto con su representante de IBM o su Business Partner de IBM, o visite el siguiente sitio web: ibm.com/software/aspera

Póngase en contacto



© Copyright IBM Corporation 2016, IBM Corporation, 1 New Orchard Road, Armonk, NY 10504 U.S.A. Produced in the United States of America, May 2016.

IBM, the IBM logo, ibm.com, and Aspera are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at ibm.com/legal/us/en/copytrade.shtml.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates. The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.



ZZC03364-ESES-02

