

FUJIFILM Object Archive

el software de almacenamiento de objetos en cinta devuelve a una de las universidades de la prestigiosa Liga Ivy de Estados Unidos el control de los datos almacenados



Hace quince años, una importante universidad de investigación del noreste de Estados Unidos utilizaba el almacenamiento en la nube para archivar los datos críticos de su investigación dejando de lado la tecnología de cinta para dicho propósito. Actualmente, la tecnología de cinta LTO Ultrium vuelve a ser una valiosa herramienta en la estrategia de almacenamiento de datos a largo plazo de la universidad, con el software FUJIFILM Object Archive que tiende un puente entre la arquitectura nativa S3 de la nube y la cinta para el archivo de datos.

La innovación genera un mayor crecimiento de datos

Esta universidad, miembro de la prestigiosa "Liga Ivy" con sede en Providence en los Estados Unidos, es una institución líder en investigación que se distingue por su aprendizaje centrado en el estudiante y su profundo sentido del bien común. Los estudiantes, el profesorado y el personal están motivados por la idea de que su trabajo tendrá un impacto en el mundo. La universidad cuenta con cerca de 9.000 estudiantes y 1.000 profesores (533 titulares y 147 con contrato indefinido).

En la universidad, los estudiantes trabajan junto al profesorado para generar nuevos descubrimientos y preparar el terreno para los avances científicos. Desde la física de alta energía hasta la ingeniería, pasando por la secuenciación del ADN, la investigación medioambiental, la economía y la astronomía, todos los departamentos importantes de la universidad generan y almacenan datos que hacen posible estos avances. Una cultura de colaboración anima a los estudiantes y al profesorado a compartir los datos entre los departamentos, lo que conduce a nuevos conocimientos e innovaciones.

Aunque gran parte de los datos generados por la universidad son dinámicos, muchos conjuntos de datos importantes no lo son. Uno de los muchos ejemplos es el de la cartografía estelar realizada con telescopios digitales, que proporciona puntos de datos que permanecerán inalterados durante millones de

años. Como los telescopios exploran los cielos, añadiendo constantemente nuevos detalles, aunque probablemente se acceda a estos datos en contadas ocasiones, será necesario almacenarlos a largo plazo para estudios posteriores.

A su vez, es posible que las patentes que posee la universidad no se consulten con frecuencia, pero deben conservarse durante décadas como prueba de la propiedad intelectual de la universidad en caso de que surjan problemas. Además, la mayoría de los investigadores trabajan con subvenciones que incluyen requisitos de almacenamiento de datos a largo plazo, y la política de la universidad es conservar todos los datos durante al menos diez años.

Hasta principios del siglo XXI, la universidad almacenaba todos estos datos en un sistema de almacenamiento conectado a la red (NAS), que incluía un uso limitado de archivo en cintas LTO de primera generación. En los últimos 22 años, el volumen de datos almacenados ha crecido considerablemente alcanzando unos 12 PB de capacidad. Una capacidad que hasta hace poco se almacenaba en varios dispositivos NAS en el campus y fuera de las instalaciones, en la nube pública, utilizando los servicios de almacenamiento en la nube de distintos proveedores. Como la mayoría de otras organizaciones, la universidad se ha enfrentado a un crecimiento exponencial de datos en los últimos años y ve que esa progresión sigue acelerándose año tras año.

Punto de inflexión: Los costes ocultos del almacenamiento en la nube

A medida que el almacenamiento de datos fuera de las instalaciones propias aumentaba rápidamente, el equipo responsable del archivo de datos de la universidad se enfrentó a un nuevo reto. Un desafío que no estaba relacionado con el coste del almacenamiento en la nube en sí mismo, ya que éste es relativamente asequible, sino a los costes excesivos e impredecibles de acceder y sacar grandes volúmenes de datos de la nube, como consecuencia tanto de las tarifas aplicadas por los proveedores de servicios en la nube como por las limitaciones de cambios de ancho de banda de la universidad. La mayoría de las organizaciones suelen pasar por alto estos costes suplementarios cuando evalúan los costes del almacenamiento en la nube.

Análisis de necesidades: La compatibilidad con el protocolo S3, un requisito indispensable

Nancy y su equipo se propusieron analizar las necesidades de gestión y almacenamiento de datos de la universidad y trabajaron directamente con el profesorado y los investigadores durante casi dos años en la búsqueda de soluciones. Los requisitos de almacenamiento cambiaban entre una base de usuarios «medios» y un subconjunto de investigadores «usuarios avanzados» que creaban conjuntos de datos mucho más grandes, de más de 500 TB con un crecimiento mensual de 100 TB.

Aunque los datos variaban, desde mapas estelares hasta genomas, y el análisis se realizaba con una gran variedad de softwares, la ubicuidad de la plataforma Glacier con S3 estaba arraigada en los flujos de trabajo de los investigadores. La solución elegida tendría que coexistir e interactuar sin problemas con la arquitectura de un almacenamiento en la nube basado en objetos.

« Nos llevó mucho tiempo encontrar la solución adecuada... Gracias al sólido y apropiado asesoramiento de Fujifilm, en todo momento, ha sido una gran experiencia, en general »

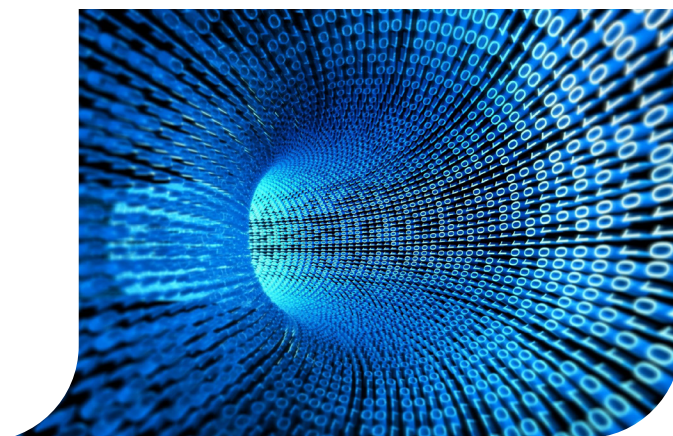
Amar, administrador principal de almacenamiento

Foco en el software FUJIFILM Object Archive combinado con la tecnología de cinta LTO Ultrium

Al principio, la tecnología de cinta para el almacenamiento de datos no estaba «en el radar» de Amar, el administrador principal de almacenamiento de la universidad. «Supongo que me consideraba parte de la generación «post-cinta», bromea ahora. Fue en un evento del sector TI a principios de 2020 cuando descubrió el software FUJIFILM Object Archive, su compatibilidad con S3 y S3 Glacier, y la capacidad de la solución para mover datos fácilmente entre la nube y la cinta. Hoy, Amar utiliza la cinta y es un claro defensor de esta tecnología para el archivo de datos.



El logotipo de «Hibernate», el sistema de archivo de la universidad, muestra al oso mascota de la escuela dormido, sano y salvo.



El entorno de archivo de la universidad

Actualmente, la infraestructura de almacenamiento in situ incluye una nueva librería de cintas con tecnología LTO-8 para una capacidad de 2 PB en dos ubicaciones del campus universitario. Los datos de archivo se leen desde su origen y se almacenan en ambos emplazamientos por motivos de redundancia y seguridad. Una vez almacenados en cinta, los datos inactivos pueden borrarse de su fuente original. A diferencia de las soluciones en cinta de hace 20 años, los sistemas de almacenamiento en cinta modernos ofrecen altos rendimientos y permiten operar en un entorno mucho más parecido al de los discos duros: «Simplemente no es un disco giratorio», comenta Amar. Los datos almacenados en cinta se conservan de forma permanente y nunca se eliminan ni se destruyen.

El software Starfish añade metadatos suplementarios que facilitan la localización de los archivos y proporcionan un acceso seguro a los datos en bruto cuando se necesitan. Mientras que el software Globus permite compartir los datos del sistema de almacenamiento de la universidad con colaboradores de otras instituciones de todo el mundo.

«Nos llevó mucho tiempo encontrar la solución adecuada, pero una vez identificada, fue un proceso fácil ponerla en marcha», señala Amar. «Gracias al sólido y apropiado asesoramiento de Fujifilm, en todo momento, ha sido una gran experiencia, en general».

Y ahora que está operativa, ¿Qué opinan los investigadores de la nueva solución?

«En relación al almacenamiento de datos, nuestros investigadores operan con un modelo de refacturación de gastos», señala Nancy. «Así que había un fuerte e inmediato incentivo por parte del usuario final para archivar los datos en soluciones más rentables. Sólo tuvimos que explicarles que si estaban pagando X dólares por almacenar los datos en la nube, pagarían 10X veces más por sacarlos de la nube». Un argumento de peso, sin duda, para apostar por una nube privada basada en cinta que permite reducir los costes de archivo y aumentar la seguridad de los datos almacenados, sin perder el control sobre ellos.