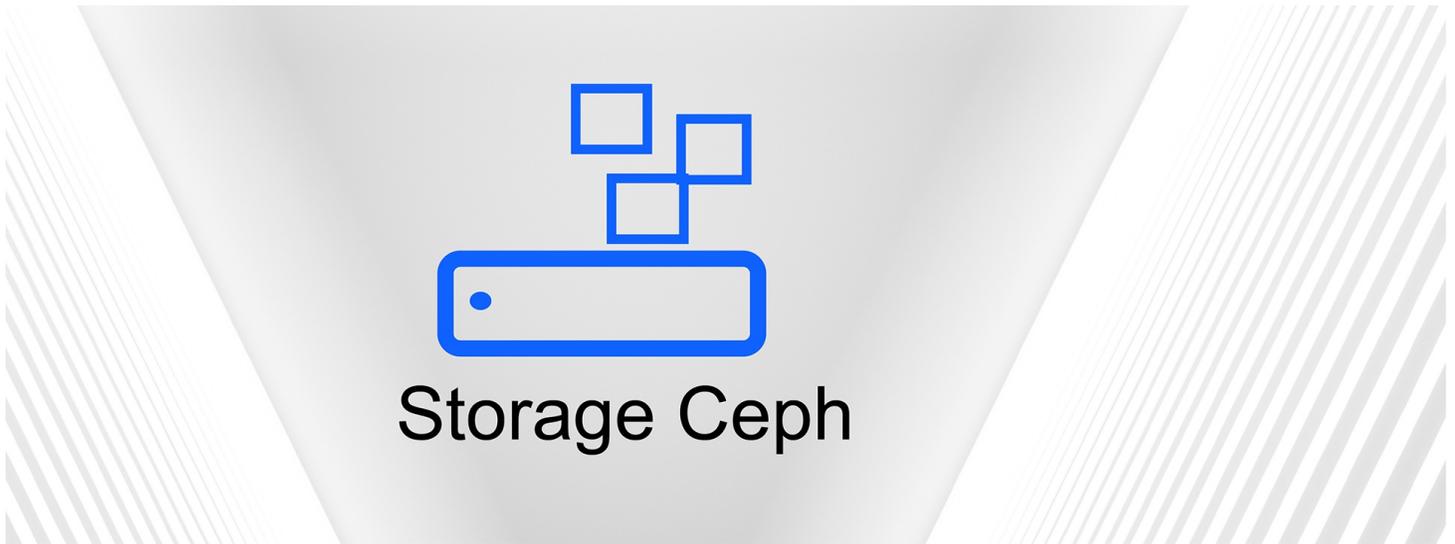


El crecimiento de IBM Storage Ceph: la base ideal para un data lakehouse moderno

Director técnico de gestión de productos, IBM Storage

Por [Gerald Sternagl](#) | Ceph

February 13, 2024



Ha pasado un año desde que IBM integró las hojas de ruta y a los equipos de productos de almacenamiento de Red Hat en IBM Storage. En ese tiempo, las organizaciones se han enfrentado a desafíos de datos sin precedentes para escalar la IA, debido al rápido crecimiento de los datos en más ubicaciones y formatos, pero con menos calidad. Ayudar a los clientes a combatir este problema ha significado modernizar su infraestructura con soluciones de vanguardia como parte de su transformación digital. En gran medida, esto implica ofrecer un almacenamiento de datos y aplicaciones confiables en entornos locales y de la nube. Además, fundamentalmente, esto incluye ayudar a los clientes a adoptar arquitecturas nativas de la nube para obtener las ventajas de la nube pública, como el costo, la velocidad y la elasticidad. Anteriormente Red Hat Ceph, ahora IBM Storage Ceph una plataforma de almacenamiento definida por software de código abierto de última generación, es una pieza clave en este esfuerzo.

El almacenamiento definido por software (SDS, por sus siglas en inglés) ha surgido como una fuerza transformadora en lo que respecta a la gestión de datos, ofreciendo un gran número de ventajas sobre las matrices de almacenamiento heredadas tradicionales, incluidas una extrema flexibilidad y una escalabilidad adecuadas para manejar casos de uso modernos, como la IA generativa. Con IBM Storage Ceph, los recursos de almacenamiento se abstraen del hardware subyacente, lo que permite una asignación dinámica y una utilización eficiente del almacenamiento de datos. Esta flexibilidad no sólo simplifica la gestión, sino que también mejora la agilidad en la adaptación a las necesidades empresariales en evolución y el escalamiento de la capacidad y el cálculo a medida que se introducen nuevas cargas de trabajo. Esta plataforma de autorrecuperación y autogestión está diseñada para ofrecer servicios unificados de almacenamiento de archivos, bloques y objetos a escala en hardware estándar del sector.^[1]El almacenamiento unificado ayuda a brindar a los clientes un puente, de las aplicaciones heredadas que se ejecutan en archivos independientes o en almacenamiento en bloques a una plataforma común que incluye esas aplicaciones heredadas y el almacenamiento de objetos en un solo dispositivo.

Ceph está optimizado para grandes despliegues en uno y múltiples sitios y puede escalarse de forma eficiente para dar soporte a cientos de petabytes de datos y decenas de miles de millones de objetos,^[2] lo cual es clave para las cargas de trabajo tradicionales y las nuevas cargas de trabajo de la IA generativa. La escalabilidad, la

resiliencia y la seguridad de IBM Storage Ceph hacen que sea ideal para dar soporte a estructuras de código abierto de data lakehouse y de IA/ML, además de cargas de trabajo más tradicionales como MySQL y MongoDB en Red Hat OpenShift o RedHat OpenStack. Esta es una de las razones por las que la capacidad en bruto de 768 TiB de IBM Storage Ceph se incluye en [watsonx.data](#), la arquitectura de IBM de data lakehouse abierta, gobernada por y adecuada para fines específicos, optimizada para cargas de trabajo de datos, análisis e IA.

MÁS INFORMACIÓN SOBRE [IBM STORAGE CEPH](#)

La base adecuada para cargas de trabajo intensivas en cálculo y datos

El crecimiento explosivo de los datos no estructurados y la IA generativa comparten una relación simbiótica, en la que cada uno influye y beneficia al otro. En su informe Top Trends in Enterprise Data Storage 2023, Gartner afirma que "para 2028, las grandes empresas triplicarán su capacidad de datos no estructurados en sus instalaciones, en ubicaciones periféricas y en ubicaciones públicas de la nube, en comparación con mediados de 2023".^[3] La proliferación de datos no estructurados, como texto, imágenes y videos, proporciona una fuente amplia y diversa para entrenar modelos de IA generativa. A su vez, la IA generativa ayuda a dar sentido y extraer valiosos insights de la agrupación en constante expansión de datos no estructurados. Esta sinergia da lugar a un ciclo de retroalimentación en el que la IA generativa se nutre de la abundancia de datos no estructurados y la generación continua de datos realistas por parte de la IA enriquece y refina aún más para usted la comprensión de los conjuntos de datos no estructurados, fomentando la innovación y los avances.

Con el 70 % de los datos de archivos y objetos previstos para su despliegue en una plataforma de almacenamiento de datos no estructurados consolidada para 2028 (frente al 35 % en 2023) según el mismo informe de Gartner,^[4] las organizaciones necesitan una solución de gestión de almacenamiento capaz de acelerar la ingesta de datos, la limpieza y la clasificación de datos, la gestión y el aumento de metadatos, así como la gestión y el despliegue de capacidad a escala de la nube, como el almacenamiento definido por software. IBM Storage Ceph se escala sin problemas para satisfacer estas crecientes demandas de datos. Sus capacidades de autogestión garantizan que el sistema se adapte continuamente a las condiciones en constante cambio, lo que hace que la solución no tenga problemas, a la vez que mantiene fácilmente la integridad de los datos.

Para acelerar y escalar el impacto de los datos y la IA en toda la organización, y en última instancia mejorar sus resultados, las empresas deben ser híbridas por diseño. Esto incluye la capacidad de consumir servicios de almacenamiento en sus instalaciones con un modelo operativo nativo de la nube para abordar problemas como la necesidad de conjuntos de características empresariales no disponibles en la nube pública, consideraciones de soberanía de datos y costo. La arquitectura plug-and-play de IBM Storage Ceph simplifica la integración con las infraestructuras existentes, incluidas diversas plataformas, entornos de la nube, hipervisores, repositorios de datos de código abierto como Apache Iceberg o Apache Parquet y plataformas de solución completa, como watsonx.ai, watsonx.data y otros. Se pueden añadir nuevos nodos o dispositivos al clúster de forma transparente, sin interrupciones ni tiempo de inactividad del servicio. Y ofrece una forma fácil y eficiente para que los clientes creen un data lakehouse con watsonx.data y otras cargas de trabajo de IA de próxima generación.

"En Snap, nuestro requisito de almacenar más y más datos sigue expandiéndose y necesitamos una plataforma que pueda escalarse rápidamente, satisfacer nuestros KPI de rendimiento y ser rentable al mismo tiempo. IBM Storage Ceph es la plataforma preferida, con su sencilla arquitectura escalable, su interfaz fácil de gestionar y

su despliegue rentable definido por software. Contar con la experiencia y el soporte de primera clase de IBM es otra parte importante de nuestra decisión de utilizar IBM Storage Ceph para un componente tan crítico de nuestra empresa". -- Snap Inc.

Acceso rápido a datos con NVMe/TCP

En el último año, IBM ha introducido varias actualizaciones importantes en Ceph, incluida, más recientemente, IBM Storage Ceph 7.0. Esta plataforma Ceph de próxima generación prepara para las capacidades NVMe/TCP que están diseñadas para permitir una transferencia de datos más rápida entre dispositivos de almacenamiento, servidores y plataformas de la nube al conservar las características de baja latencia y gran ancho de banda alto de la NVMe tradicional. Esto lo hace adecuada para aplicaciones que exigen un acceso de almacenamiento ultrarrápido, como bases de datos, análisis y entrega de contenido, y simplifica la infraestructura, gracias a su compatibilidad con las inversiones en tecnologías tradicionales de red. Estas ventajas ayudarán a los clientes a adoptar un enfoque definido por software diseñado para ofrecer una experiencia similar a la nube en términos de velocidad, agilidad y rentabilidad.

NVMe/TCP puede ayudar a Ceph a salvar la brecha del almacenamiento en bloque tradicional con arquitecturas de escalamiento. Con NVMe/TCP, Ceph se diseñará para integrarse con plataformas como VMware para ayudar a las empresas a replicar arquitecturas de la nube en su propio centro de datos, alejándose de costosas y rígidas redes SAN y matrices de almacenamiento monolítico.

Nuevas características adicionales incluidas en Ceph 7.0:

- **Certificación de cumplimiento con SEC y FINRA para WORM con bloqueo de objetos**, que permite el cumplimiento de WORM para el almacenamiento de objetos
- Soporte de **NFS para el acceso al sistema de archivos CephFS** para clientes no nativos de Ceph
- Para obtener más detalles sobre las características, visite la comunidad de IBM Storage [aquí](#)

Economías de escala de la nube con IBM Storage Ceph

Puesto que IBM Storage Ceph almacena datos como objetos dentro de agrupaciones de almacenamiento lógico, un solo clúster puede tener varias agrupaciones, cada una ajustada a distintos requisitos de desempeño o capacidad. Esto permite que los clientes se beneficien con un acceso más fácil y rápido a los datos con clasificaciones de contenido y contexto, capacidad de almacenamiento limitada sólo por el tamaño de la infraestructura de la organización y reducciones de costos a escala, al eliminar las restricciones de hardware en comparación con las arquitecturas de matriz de almacenamiento tradicionales y heredadas.

Tiempo de generación de valor más rápido

IBM también ha hecho que el despliegue de Ceph sea más fácil que nunca. Con [IBM Storage Ready Nodes](#) for Ceph, la plataforma se puede desplegar como una solución completa de software y hardware y viene en diversas configuraciones de capacidad optimizadas para ejecutar cargas de trabajo de IBM Storage Ceph. Hemos sacado de la configuración todas las conjeturas, lo que hace que sea más fácil de digerir, configurar y administrar.

El crecimiento de IBM Storage Ceph es solo otro ejemplo de cómo el portafolio de hardware y software de almacenamiento de IBM ayuda a acelerar la generación de valor con capacidad y desempeño escalados para optimizar los costos para los clientes.

1, 2 <https://www.redhat.com/en/resources/data-solutions-overview>

[3], 4 Top Trends in Enterprise Data Storage 2023; Chandra Mukhyala, Julia Palmer, Jeff Vogel; 29 de junio de 2023



Article Categories

[Nube Híbrida](#)