



Democratising Data Storage

DIGITAL REPORT **2021**

IN ASSOCIATION WITH:





DEMOCRATISING *DATA STORAGE*

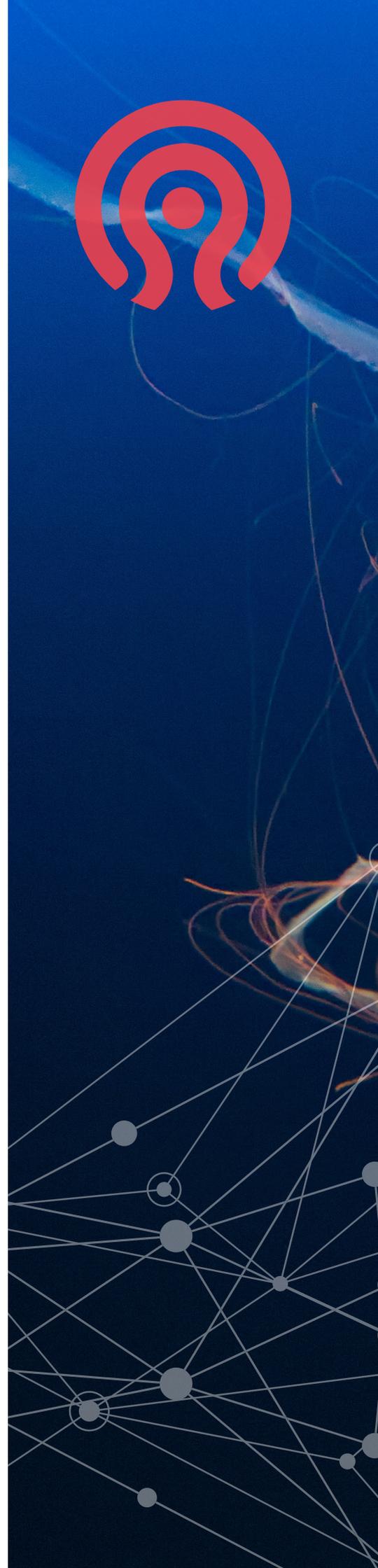


Meet Ceph: the open source data storage solution helping decision makers find the flexibility, reliability, and scale to navigate a changing IT landscape

It should come as absolutely no surprise to anyone who hasn't spent the past decade living under a rock that data has become the backbone, the body, the soul (the metaphor of your choice) of the modern world. Data storage, analysis, recovery, and management are mission critical capabilities for any enterprise - and the core value proposition for more than a few.

The data centre industry is experiencing an explosion of capacity throughout both mature and emerging markets, datasets are growing exponentially like some 1950s sci-fi special effect, and emerging tech trends like 5G, the IoT, artificial intelligence, machine learning, HPC, cold storage, and edge computing all conspire to pour gasoline on an already raging fire. Yet, the more critical that an effective data storage solution becomes to organisations and enterprises of all shapes and sizes, the more apparent it is that the solutions dominating the market today aren't necessarily the right tools for the job.

"In retrospect especially, but even at the time there was a glaring hole in the market. There was a clear need: everybody needed storage, it needed to be scalable, and there was no open source option; you had to buy expensive proprietary solutions," reflects Sage Weil, Principal Ceph architect at Red Hat, and the founder and chief architect of Ceph. "There needed to be an open source alternative that was good, and that's the niche we've tried to fill."







Red Hat Ceph Storage



Massively scalable storage for demanding applications

Red Hat® Ceph® Storage, an open, massively scalable, simplified storage solution for modern data pipelines. Engineered for data analytics, AI/ML, and emerging workloads, Red Hat Ceph Storage delivers software-defined storage on your choice of industry-standard hardware.

[Learn more](#)





The Ceph Foundation

Since the first prototype of Ceph was launched back in 2007, the community of enterprises, organisations, and users that use it has grown to touch a huge number of spaces, from small businesses to large scale enterprises; from the scientific community to regional telecom carriers.

In November 2018, a cluster of organisations actively involved in the development, support, and commercialisation of Ceph formed the Ceph Foundation, launching the new initiative under the umbrella of the Linux Foundation. The founding members included Amihan, Canonical, China Mobile, DigitalOcean, Intel, OVH, ProphetStor Data Services, Red Hat, SoftIron, SUSE, Western Digital, XSKY Data Technology, and ZTE.

“I was pretty naive back then. I thought you just built something, open sourced it, and people would just start appearing to develop it, fix bugs, etc. and that's not how it works,” laughs Weil. “We had spent several years trying to add all the features that we thought Ceph had to have before people would be willing to use it,” before the launch of the Foundation in 2018. “There are a lot of industry stakeholders that are selling or using Ceph as part of their business. So the Ceph Foundation became a way for those organisations to contribute funds that could be managed and spent to further Ceph's development and the community. Prior to the Foundation, it felt a lot more like begging - going around asking 'who wants to pay for X or Y',” he adds.



“Ceph is designed to provide a reliable storage service out of unreliable components”

SAGE WEIL
CEPH PRINCIPAL ARCHITECT,
RED HAT

“The Ceph Foundation is essential to the Ceph community and our customers because its members are all invested in the development and progression of Ceph,” says Aaron Joue, founder and CEO of Ambedded Technology - which combines Ceph technology with its own line of decentralised Arm servers.

The power of the Foundation, continues Kyle Bader, a Data Foundation Architect at Red Hat, lies in its ability to drive the industry to “deliver on the promise of democratising software defined storage through open source in a way that’s very similar in the way that Linux led to the democratisation of the operating system.”

That democratisation, adds Craig Chadwell, VP of Product at SoftIron, creates the necessary competition (centred around the foundation itself) to push the Ceph commercial ecosystem

to even greater heights. “The Ceph community is very large and robust. The Ceph Foundation helps to enliven and support that community, which in turn ensures that there will always be other options out there so that we can maintain that no vendor lock-in value proposition,” he explains. “It really forces us to continually challenge ourselves to deliver solutions that are uniquely solving customer problems, because the reality is, if a customer can move away and there's something providing more value out there, they will. It keeps us honest and on our toes.”

Philip Williams, Product Lead at Canonical, reflects that “a funny thing about the open source world is that essentially we’re all competitors, but we’re also all working together to make something that is available for free even better.”



Meet Ceph: Reliable, scalable, affordable. open source

Developed by Weil - in collaboration with data storage researchers at the University of California: Santa Cruz, as well as at researchers from the country's leading laboratories in Los Alamos and beyond - Ceph is a distributed, open source data storage solution that grew to fill that glaring hole in the market Weil and his colleagues saw back in the 2000s.

"Ceph is designed to provide a reliable storage service out of unreliable components. You take a bunch of individual hard drives that can fail, a bunch of networks that can fail, switches, servers that all individually are very fallible, you put them all together with Ceph and the net result is something that's highly reliable that tolerates any single point of failure - or in many cases many points of failure. It's highly available and highly scalable as well," Weil explains, adding that Ceph is

also capable of providing object, block, and file storage all in one system on the same hardware.

Ceph's distributed approach to data storage is hugely fault tolerant. Like a commercial airliner that can continue to fly with all but one engine out of commission, Ceph is robust enough to handle all but the most catastrophic of outages.

As a storage solution, Ceph's appeal also lies in its open source, software defined design that - in addition to delivering reliability and flexibility at scale - excels at adding up to far more than the sum of any somewhat meagre parts you might happen to have lying around. "Ceph is open source, software defined, and meant to be run on any commodity hardware you want to buy or already have," Weil says. "It doesn't matter which vendor you're buying your hardware from, whether you're using hard drives or SSDs, what kind of switches are in your network; it's fully software defined,"

Robust and Low Power Ceph Storage Appliance

A Ceph appliance integrates a purpose-built Arm micro-server platform, makes the ceph storage system furthermore robust, and enables enterprises to adopt a tiny cluster up to a large-scale one.

- Professional Ceph support services
- Smallest Failure Domain
- Extreme Power Saving

[Learn More Today >](#)



ARM Server



UVS Manager



CANONICAL - THE FULL SPECTRUM OF ENTERPRISE COMPUTE

» Open source is the core of everything that happens inside Canonical. “All of our work is out in the open; it’s free to use, which I think is super important for storage technology, because you can understand what is actually going on inside the system. That’s impossible when you’re using closed source or proprietary technologies,” says Philip Williams, Product Leader at Canonical. “One thing that you see with traditional storage systems from the big players is that their growth is usually around scale up. They’ll have arbitrary limits on the number of disks you can add to a system.”

With datasets inside every organisation growing larger every day, Williams notes that “for people with quite significant amounts of data, public cloud and those traditional proprietary storage options typically aren’t cost effective or feasible.” As a result, Canonical - along with the rest of the Ceph community - is embracing the infinite scalability of Ceph to “demonstrate to enterprise users that there is this viable alternative to the big players, and that their organisations don’t have to be developer centric to make use of Ceph.”

and therefore a legitimate and long-awaited answer to market demand for alternatives to restrictive, proprietary storage solutions.

“Storage is quite an interesting industry. It’s kind of hidden; people don’t really think about storage until it’s either too expensive or it’s not available and, worst case, all your data has been lost,” says Philip Williams, a Product Leader at Canonical. “So, it’s this funny little world that’s dominated by a number of very large players. The whole aim of the Ceph Foundation is not just to shepherd the upstream projects and this collaborative development work on Ceph itself, but also to demonstrate to enterprise users that there is this viable alternative to the big players, and that their organisations don’t have to be developer centric to make use of Ceph.”

Ceph’s open source, software defined nature means that organisations looking to deploy it can “choose any hardware you like, choose any vendor you like - or even no vendor at all - but if you build a Ceph system and you want to switch vendors or run things on your own, you can do that very easily.”

SOFTIRON - DELIVERING 21ST CENTURY INFRASTRUCTURE THE RIGHT WAY

» At its core, SoftIron asks the question of how to deliver 21st century infrastructure to meet the evolving demands centred around factors like resilience, performance, accessibility, and environmental impact. “We want to meet customers where they are. And we do that through building a set of task specific appliances that solve a variety of problems inside the data centre,” explains SoftIron’s VP of Product, Craig Chadwell. SoftIron’s solutions range from software defined infrastructure solutions that touch the computing and networking spaces to specific solutions for media encoding. “If you take a holistic view of solving a problem, you can deliver a solution that is resilient, but with performance that’s way beyond what a traditional commodity based solution could provide,” says Chadwell.

SoftIron’s suite of storage solutions are powered by Ceph, but the entire company’s ethos is firmly aligned with the open source community as well. “One of the core premises of SoftIron’s value proposition is this notion of no vendor lock-in. We provide solutions that enable our customers to effectively swap us out if they choose,” Chadwell continues. “Because of the way Ceph works and because organisations that comply with Ceph’s operating model can have products that work together seamlessly, it means that you can swap out a particular vendor’s hardware without ever having to swap out Ceph, which means that everything above the Ceph layer from a service delivery perspective is unaffected by the lower level technology changes.

“Everybody needed storage, it needed to be scalable, and there was no open source option; you had to buy expensive enterprise solutions”

SAGE WEIL
CEPH PRINCIPAL ARCHITECT,
RED HAT

In addition to offering the unparalleled freedom to start from scratch, move freely within its ecosystem, and avoid both the vendor lock in agreements and cumbersome, expensive upgrade cycles that define managed, proprietary storage solutions, Weil adds that the beauty of Ceph is that “Because it’s so flexible and built to scale, Ceph doesn’t require a lot of foreknowledge about where your organisation’s going to be in a couple of years time. You can just expand your hardware footprint in whatever direction you end up growing.”

Large storage systems - the kinds that are increasingly coming to define the cloud and data centre industries - are fundamentally dynamic. They grow and change in new and unexpected directions in response to the market and, with Ceph, organisations can grow and change with as little friction as possible. “You might start out with 10 servers from one vendor, and then five years later you’re storing 12 times

KEY PARTNERSHIPS

SAGE WEIL

TITLE: CEPH PRINCIPAL
ARCHITECT, RED HAT



Sage Weil is the lead architect and co-creator of the Ceph open source distributed storage system. Sage helped design Ceph as part of his PhD research at the University of California, Santa Cruz. Since finishing in 2007, he has continued to refine the system with the goal of providing a stable next generation distributed storage system for Linux. Sage co-founded Inktank in 2012 to bring enterprise-quality open source storage to the enterprise. After Red Hat acquired Inktank in 2014, Sage has worked in Red Hat's Office of the CTO to improve Ceph and help shape Red Hat's overall storage strategy.

His experience spans product design, hardware, software, storage, computing, and Manufacturing.

Aaron founded Ambedded in 2012 to create an innovative platform for software-defined storage to improve energy efficiency, availability, and performance. Before this, he was ever an outstanding engineer, factory manager, and VP of product.

KYLE BADER

TITLE: DATA FOUNDATION
ARCHITECT, RED HAT



Kyle is the Data Foundation Architect covering both OpenShift Data Foundation and Red Hat Ceph Storage products at Red Hat. His focus is at the intersection of open source, distributed storage systems, data engineering, and machine learning.

PHILIP WILLIAMS

TITLE: PRODUCT LEADER,
CANONICAL



Philip is a Product Leader at Canonical responsible for Ceph and other storage solutions. He has over 15 years experience in the storage industry, having previously been responsible for storage infrastructure and products at a number of leading technology companies.

CRAIG CHADWELL

TITLE: VP OF
PRODUCT, SOFTIRON



Craig has spent over a decade engineering, marketing, and leading product management of cloud and software-defined data center solutions. Craig has held positions at Lenovo, NetApp, and High Point University where he gained first-hand buyer and administration experience across the lifecycle of data center operations. Craig has degrees in computer science, history, political science, and business administration.

AARON JOUE

TITLE: CEO, AMBEDDED



Aaron Joue is the founder of Ambedded Technology. He is responsible for the business strategy and ensures the product development and support satisfy customers' needs. He accumulated experience when he worked for the defense and information technology industry.



HyperDrive is purpose-built to optimize Ceph, so you don't have to.

Experience a true enterprise-class, turn-key Ceph solution with full interoperability and support, yet no vendor lock-in.

Test Drive a HyperDrive today.

[LEARN MORE](#)



softiron.com



as much data and you've been through three different hardware revisions all from different vendors, you've had to migrate data, change policy, and now you're storing a different type of data than you were before - it's all a total mess," Weil laughs. "Often, your net system is going to be a mixture of all sorts of different stuff, and open source lends itself to solving those problems really well because you have the neutrality to be flexible and adaptable. If you're buying a proprietary solution from a particular vendor, you're going to have to buy more of the X solution that they allow you to interoperate with. You're locked into a particular path." Ceph, he adds, not only frees organisations from those restrictive, vendor-defined upgrade paths, but opens up a huge, mature ecosystem of enterprises and community members to its user base.

Harnessing the Ceph community

When it comes to harnessing the true value of Ceph, its commercial ecosystem and user community are pivotal. From companies like Red Hat that deliver Ceph solutions to Fortune 500 companies, to SoftIron which simplifies the Ceph adoption process through curated, in-house designed

RED HAT - CEPH AT PETABYTE SCALE

» Kyle Bader, Red Hat's Data Foundation Architect, comes from an operations background. As such, he's no stranger to the headaches that can accompany endeavours to be flexible in a large IT environment. "If you have a handful of proprietary appliances, that's probably fine and quite manageable, but at Red Hat we have several customers that are managing north of a hundred petabytes of data," he explains. The growth in the amount of data being stored and managed, he continues, "is not stopping by any means. So, it becomes challenging once you get to larger scales."

Red Hat's business brings the power of open source ecosystems to large scale enterprises, leveraging everything from OpenShift to Kubernetes in order to take its customers' IT infrastructure to the next level. Engineered for data analytics, artificial intelligence and machine learning, Red Hat's Ceph Storage delivers software-defined storage on industry-standard hardware to organisations at any scale.



AMBEDDED - ROBUST AND LOW POWER CEPH APPLIANCE

» "We think reliability, scalability, and simplicity are keys to our offering to the enterprise storage market," explains Aaron Joue, founder and CEO of Ambedded Technology. "Ambedded integrates Ceph software, the Arm microservers, and its tailor-built ceph manager GUI (UVS manager). With the nature of arm microservers, enterprise storage customers could start deploying ceph from a tiny cluster to a petabyte-scale without an entry barrier.

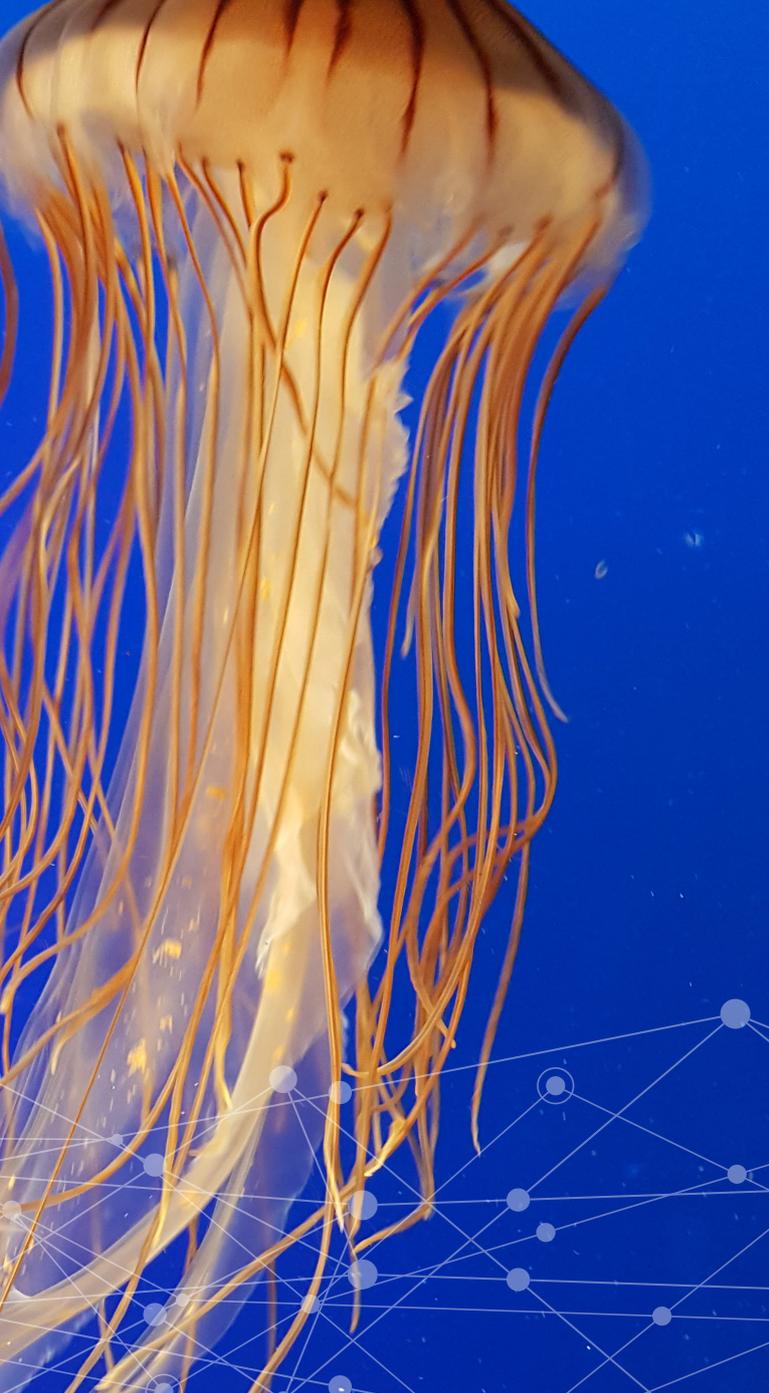
Also, this ARM-based appliance results in an easy to operate, ultra-robust, minimum failure domain and low-power offering that enables customers to operate their desired cluster in a cost efficient way without complications.

"We had a project in Hong Kong where a customer wanted a very high availability storage system for CCTV applications," Joue continues. "The storage system needed to keep working, even if up to four storage servers failed at the same time. It will be costly by using a traditional storage solution and hard to achieve to keep the cluster running with 4 servers failing at the same time." However, leveraging the ceph storage and the Ambedded microserver architecture, this HK customer easily achieved their goal even with a relatively small scale cluster.



hardware that's tailor made to support its deployment, Ambedded, Canonical, and beyond, the Ceph commercial ecosystem provides support and services that allow companies of any scale, maturity, or specialisation to deploy and benefit from distributed storage - all built on Ceph.

"When it comes to getting started with Ceph, it can be an issue knowing which servers to buy, which hard drives and how many," Weil acknowledges. "That's where companies in the commercial ecosystem really add a lot of value, not to mention the open source community at large."



Ceph for everyone

Since the dawn of the open-source approach to software design, open source solutions have often garnered “a reputation for being really complicated to use,” Weil admits - adding that he and the Ceph team have spent the past few years painfully aware of that fact. Now, however, “A lot of the stigma surrounding open source in general has gone away in recent years,” he explains, something that perfectly aligns with the latest evolution of Ceph’s General User Interface (GUI).

“These days, if you're a small business and you need 100 terabytes of storage,

you're going to want something with a nice GUI that just works,” Weil notes. “So, over the last three to four years, there's been a huge investment of time and resources in the Ceph community on the usability front. We've created a whole new, integrated GUI dashboard for Ceph for management. We've also developed an orchestrator layer for Ceph that can call out to whatever tools you use to deploy it, so that you can do just about anything you need to do from the new GUI. I think we've made huge progress.”

Challenges

Enterprise storage is full of challenges. Apart from the obvious spiralling quantity of data being generated, the applications that create and use that data are also increasingly diverse and changing almost daily. Storage, of course, is also not immune to the broader IT skills crisis that enterprises find themselves dealing with every day. Add to that the constant revolving door of mergers and acquisitions in the storage industry and it's hard not to find a storage manager that hasn't been burned by obsolete or sidelined proprietary solutions that have fallen out of favour. It's little wonder then that a platform like Ceph - able to flex and grow to meet ever changing demands across a huge variety of use cases - and do all that from within a vibrant open source community eliminating the lock-in problem, becomes deeply compelling.

The Ceph decade

Looking to the future, the intersection of market trends with Ceph’s constantly developing capabilities (Weil stresses that a sizable portion of the Foundation’s role is keeping up with cutting edge hardware developments to ensure Ceph continues to run smoothly, no matter what you plug it

FUTURE



Bloomberg

CEPH USER TESTIMONIALS

FUTURE PLC

NAME: **TOBY JACKSON,**
TITLE: **GLOBAL SRE LEAD**



“Future PLC has grown in recent years through acquisition of over a dozen media brands and technology companies, bringing a complex and diverse data ecosystem. Future operates services on both cloud and on-premises infrastructure, leveraging technologies that need to balance value and flexibility. Future's technology team has also grown and diversified to meet these demands, requiring their SRE team to provide ever more robust and flexible solutions without compromising on stability or performance. As a content publisher, data is integral to Future's operation, and Ceph was the clear choice to ensure its technology teams had the tools at their disposal to manage data at scale. Ceph allows Future to deploy storage across its datacenters, trusting that its data is distributed and available while enabling its development teams to self-manage storage requirements from Kubernetes with ease.

Future requires a storage solution that can not only grow in volume, but can handle diversity of locality, performance, distribution and access models; Ceph affords this flexibility under a common platform, reducing deployment complexity and simplifying its operational costs.”

DATA COMM CLOUD BUSINESS

NAME: **MR. LUK PHIN**
TIRTOKUNTJORO,
TITLE: **CTO**



“We have deployed, run and tested Ceph storage in our development environment for 2 years. Once confident with the solution, we decided

to build Openstack Cloud using Ceph storage to provide IaaS, PaaS and SaaS to SMEs in Indonesia. We have considered several storage solutions to work with our OpenStack cloud, including legacy storage and software-defined storage. As our cloud service targets SMEs, we need a solution that is cost-efficient, cloud-native ready, scalable, reliable, resilient, and with good performance. We found Ceph storage to be a perfect fit after 2+ year trial experience. Furthermore, as we have three data centers located in 3 different sites, the RBD mirroring and simplicity to backup data from one data center to another in the Ceph cluster enhances our belief in choosing Ceph storage.

BLOOMBERG

NAME: **MATTHEW LEONARD,**
TITLE: **ENGINEERING MANAGER,**
STORAGE

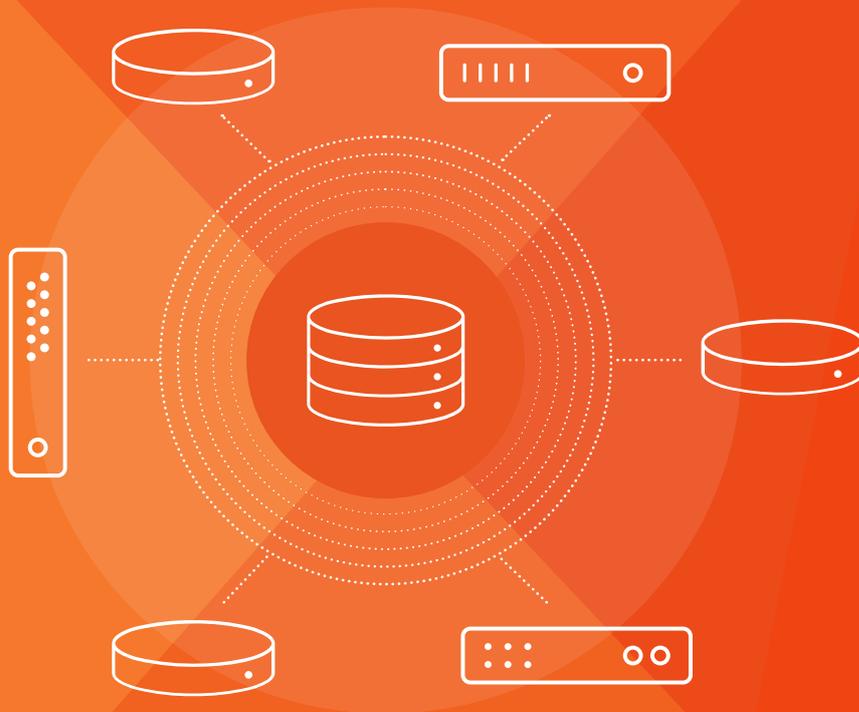


Data is at the heart of Bloomberg's technologies, which produce and distribute some of the most critical and valuable data in global business. Maintaining the systems which store and process this data requires a unique mix of commodity hardware, open source software-defined storage, and vendor-agnostic appliance-based storage solutions.

In our diverse storage ecosystem, Bloomberg utilizes Ceph's enterprise-level scalability and durability to support different applications and varied workloads across our organization. Ceph plays a significant role in Bloomberg's OpenStack-powered private cloud computing infrastructure, as well as underpins Bloomberg's private S3-based object stores. As Bloomberg works toward contributing back to the Ceph project, we feel it is the right distributed storage technology for us -- both now and into the foreseeable future.

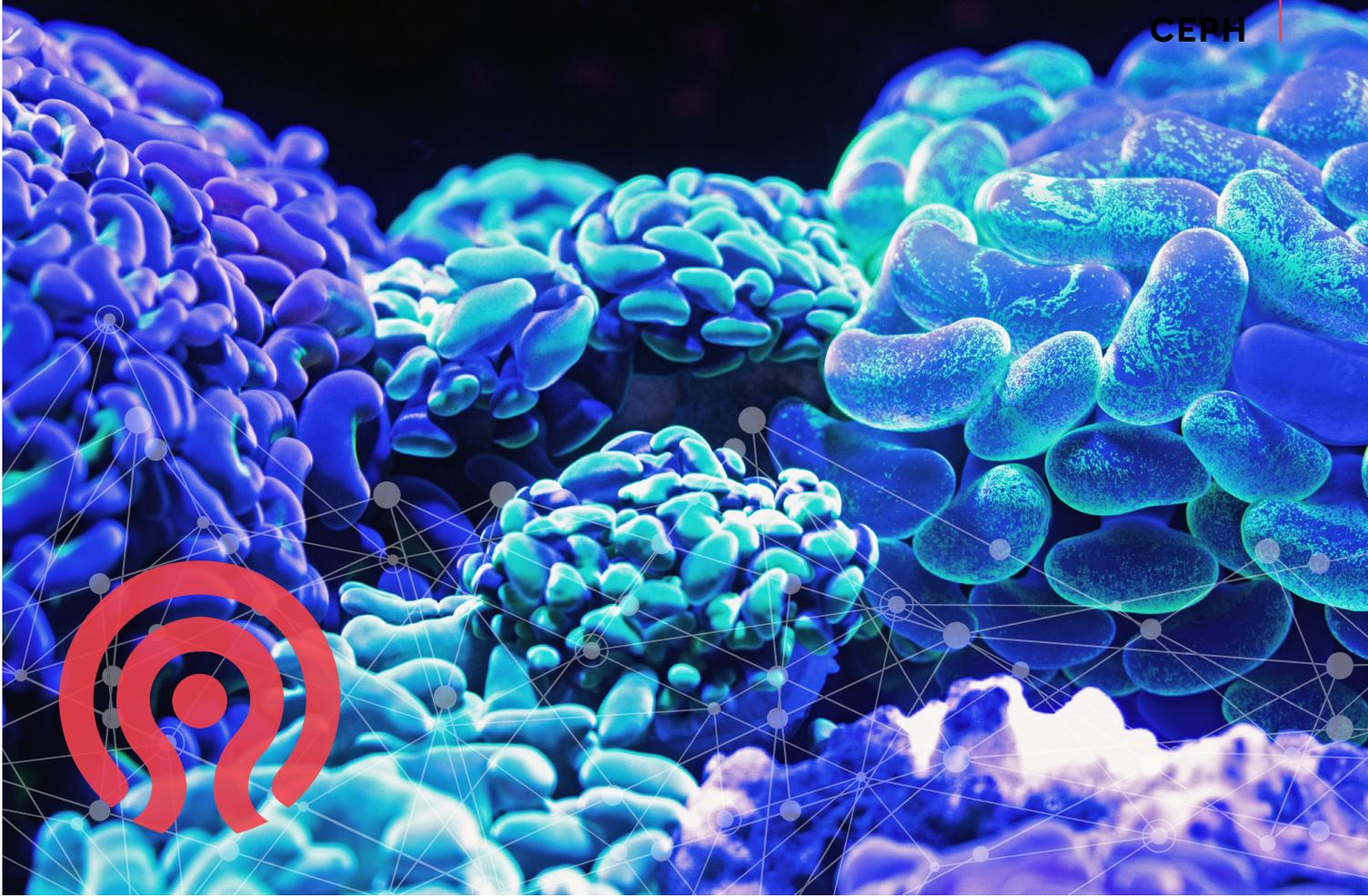
Enterprise-ready Ceph Software-defined Storage.

Open-source, reliable,
and scalable.



Plan, build, and grow your storage infrastructure using a production-hardened Ceph open-source solution. Model-driven automation drives down total cost of ownership, with simplified deployment and day 2 operations. 24x7 support with Ubuntu Advantage for Infrastructure.

[Click here to find out more](#)



“Because it's so flexible and built to scale, Ceph doesn't require a lot of foreknowledge about where your organisation's going to be in a couple of years time. You can just expand your hardware footprint in whatever direction you end up growing”

SAGE WEIL
CEPH PRINCIPAL ARCHITECT,
RED HAT

into), as well as an ever-expanding ecosystem of vendors, users, and developers positions it ideally for a decade of meteoric growth. “Ceph is a pretty mature piece of software at this point,” Weil reflects. “All of the important stuff is there and, in addition to building it out further, we’re starting to add a lot of polish.”

Craig Chadwell, VP of Product at SoftIron, reflects that “open-source infrastructure has rapidly evolved and matured over the last decade and is in all likelihood going to be the way that most organisations deploy their IT footprint going forward.”

“People like to call Ceph the Linux of storage, which I think is appropriate,” adds Weil. “Nobody thinks about which Unix they should buy because the open source one is the best, everyone's using it, and everyone is constantly improving it. Ceph is moving into that position in the storage space.” 





Learn more at ceph.io

POWERED BY:

