

2019

Poindexter Papers

How to Navigate the
Data Explosion

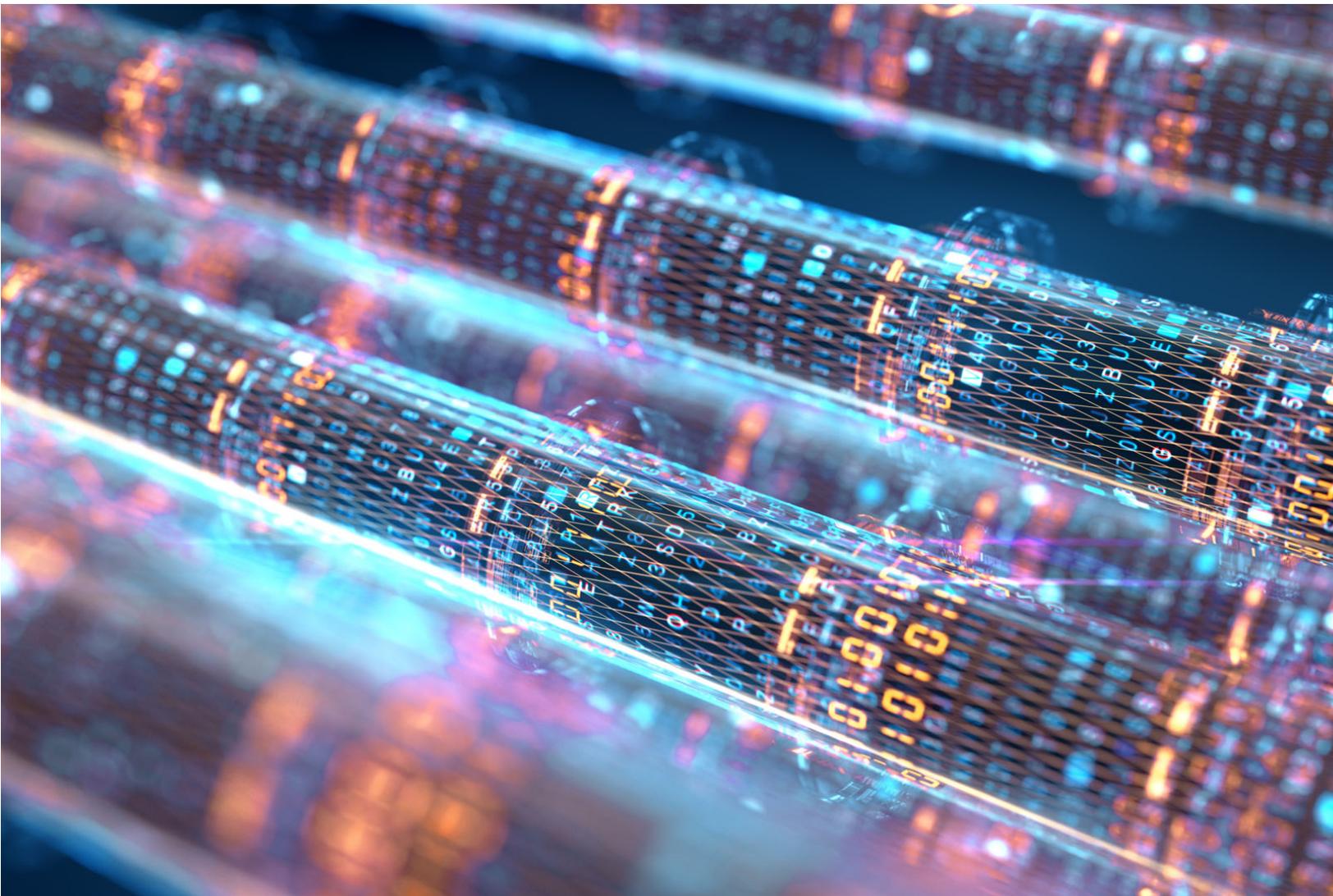


Table of Contents

INTRODUCTION	2
SOFTWARE-DEFINED STORAGE	3
OPPORTUNITIES FOR SDS	4
SUSE ENTERPRISE STORAGE	6
SERVICES YOU NEED AND WHY	8
CONCLUSION	9

Introduction

There are over 3.7 billion people using the internet to generate over 2.7 quintillion bytes of data every day. By 2025, IDC predicts that the total amount of data generated across the globe will increase to 175 zettabytes, almost 30% of all data will be generated real-time, and nearly half of the data will be stored in the public cloud¹. These are staggering predictions, which may leave IT managers wondering whether they have the strategy, solutions, or skillset to manage such massive amounts of data. In this paper, we'll look at why a very specific category of storage –software-defined storage (SDS), can be leveraged to provide value through simplicity, agility, flexibility, and improved total cost of ownership. We'll also discuss how engaging the right services partner can ensure you implement the best strategy, software-defined infrastructure, and reliable support structure to ensure your enterprise is ready to manage the data explosion.



¹ Reinsel, Gantz & Rydling, *The Digitization of the World From Edge to Core. An IDC White Paper. 2018.*

Software-Defined Storage

Software-defined storage is a virtualization technique aimed at reducing the costs of managing growing data stores by decoupling storage software from its hardware to allow centralized management of cheaper commodity hardware. But, beyond this simplified definition of software-defined storage are nuances that create big differences in the solution you are ultimately getting. So, it's important to understand those subtleties.

As defined by Gartner, an SDS solution will use software to separate and abstract storage capabilities that are pulled from industry standard, commodity hardware, with the aim of delivering higher quality of service while reducing costs. IDC adds to the idea of hardware agnosticism by defining SDS as “any storage software stack that can be installed on commodity resources (x86 hardware, hypervisors, or cloud) and/or off-the-shelf computing hardware”. So, to qualify, a solution must run on generic, industry-standard hardware, without any proprietary hooks that ultimately lead to limitations. Gartner agrees with IDC's take by stating that SDS works regardless of class of storage.

But Gartner goes one step further by dividing software-defined storage into two distinct categories: management SDS, and infrastructure SDS. While both maintain strict hardware agnosticism, management SDS is more concerned with the coordination of data services for, and increasing the agility of existing storage systems, infrastructure SDS is more concerned with creating new data center services to replace or enhance more traditional storage arrays and deploying new storage platforms that ultimately lead to reduced CAPEX. The storage platforms created by infrastructure SDS can be accessed by block, file or object software, and can be

deployed by several means, including virtualized, containerized, or as software on bare-metal x86 hardware.

In this paper, whenever we use the term “software-defined storage”, or “SDS”, we are referring to infrastructure software-defined storage as defined by Gartner.

Benefits of SDS

As big data continues to grow, finding cost-effective ways to gain value from all that information will be a critical deciding factor for

companies that want to remain viable in the future. Because you can use commodity hardware

As big data continues to grow, finding cost-effective ways to gain value from all that information will be a critical deciding factor for companies that want to remain viable in the future.

to grow your storage infrastructure as your data grows, SDS provides a much more flexible and lower cost solution, no matter how massive your data storage needs become.

SDS also provides a no-single-point-of-failure design principle that simply asserts that no single part can stop the entire system from working. With traditional dedicated storage solutions, a storage array can't borrow capacity from another when demand for storage increases, which leads to data bottlenecks and a single point of failure. Software-defined storage uses commodity hardware devices and provides shared storage capabilities, such as mirroring and replication. It eliminates the need for dedicated storage arrays and storage area networks. And because it distributes the workload across multiple devices, if any single device or node ever fails, it doesn't bring down the entire system. This is how cloud computing providers, like AWS, are able to guarantee 99.9% uptime. AWS uses SDS for its Simple Storage Service (S3). S3 users don't know (or care) how data is stored in S3; all they care about is that their data is always available.

SDS utilizes commodity hardware to provide cost reduction, simplification, scalability and avoidance of a single point of failure through distributed workload. Now, let's look at some of the opportunities presented by software-defined storage.

Opportunities for SDS

In the past, SDS was often considered simply an alternative to the major brands of fully-integrated, expensive, on-prem enterprise storage solutions. But those days are long gone. Gartner defines SDS, as being deployable anywhere, including in the cloud: the public cloud. Precisely because it removes the constraints of data co-location it makes a great solution for hybrid cloud and multi-cloud approaches to data storage.

Because of all the increasingly complex and nuanced forms in which data occurs in business, unstructured data increasingly presents new challenges in identification and storage in a hybrid cloud, leading to object storage approaches that allow for an abundance of metadata. In these cases, metadata itself is used to organize data within an object store, which then allows object platforms superior scalability.

DevOps is an emerging form of agile thinking and working applied in a way that merges the traditionally siloed functions of development and operations. Hence, "DevOps". It's more than a methodology for speeding up software development cycles; it's a way of thinking about deploying higher quality software faster and more frequently. It also helps improve operational efficiency by automating standardized deployments across your lower level environments. While not for everyone, done well, it can ultimately lead to a Continuous Integration, Continuous Development (CI/CD) approach. Any organization headed in that direction is also talking about some form of containerization, which, according to 451 Research, will be a \$2.7bn market by 2020.²

² 451 Research: Application containers will be a \$2.7bn market by 2020

The problem is that containers don't play well with traditional storage solutions because either the storage software tools can't handle containers, or they don't scale well-enough for containers. Precisely because SDS is so scalable and flexible, it ends up being a perfect fit for container technology.

With more and more data being generated, issues stemming from data fragmentation become governance issues. Suddenly, effective data management is a primary concern for all the unstructured data being generated. Not that it wasn't before, but with the data explosion, governance has taken on a whole new sense of urgency.

This can be alleviated by the simplified management offered by software-defined storage. When high performance isn't a critical requirement, and the simplified management of a lot of unstructured data is a top priority, open-source SDS, like **SUSE Enterprise Storage** is an ideal solution.

Another great use case for the application of software-defined storage is in the unification of an organization's many and seemingly unrelated data sources from different locations into a single, centralized view. With a myriad of data sources, an enterprise needs to be able to consolidate that data into a holistic and sensible picture from which to make critical decisions. SDS is an excellent way for unifying all these disparate data sources into a single storage platform.

There are many other use cases for SDS, such as high availability, data archival, IoT, VDI and others. There are too many, in fact, to cover here. So, let's move on to discuss a specific software-defined storage solution: SUSE Enterprise Storage.



SUSE Enterprise Storage

We've already noted the major benefit of open-source SDS as being able to provide storage services that are hardware agnostic, thereby reducing vendor dependency, as well as CAPEX and OPEX costs that are associated with proprietary storage software. Arguably, one of the major open source distributed storage solutions in the marketplace today is **SUSE Enterprise Storage**, a self-managing, self-healing, distributed software-based storage solution for enterprise customers, that was built to leverage the cost-effective and scaling capabilities of Ceph using commodity off-the-shelf servers and disk drives. It has all the features needed for handling object, block and file storage, including thin-provisioning, copy-on-write cloning, cache-tiering, and erasure-coding. In fact, according to a SUSE customer quoted on Gartner Peer Insights, "If you plan to mount a storage solution, SUSE Enterprise Storage offers you that storage solution in an agile and easy way to manage, in addition to reducing expenses using cheaper hardware."

SUSE Enterprise Storage is built on Ceph, the open-source, distributed object storage and filesystem that was purposefully designed as a single, integrated system that provides outstanding reliability, scalability, and performance for object, block and filesystem storage. Ceph is open-source, infinitely scalable and avoids single-point of failure risk posed by proprietary systems that don't allow sharing.

SUSE Enterprise Storage is a very scalable solution that also provides variable data sizes, as opposed to traditional, fixed-length block storage solutions. Data is distributed across the system and its location is tracked using the Controlled Replication Under Scalable Hashing (CRUSH) algorithm. In some conventional block storage systems, fixed-length blocks of data are tracked by a central management system. This can lead to bottlenecks as the centralized system breaks up data into evenly-sized blocks, and then looks for available slots into which it can insert the data. This can be considered inefficient when compared to the Ceph-based object approach, where the CRUSH algorithm takes care of managing the location of distributed data.

A client trying to retrieve data from a SUSE Enterprise Storage cluster is aware enough of the cluster topology thanks to the CRUSH map to ask the primary OSD for the object that underpins the data. This results in the ability for a cluster to scale horizontally without protocol gateways acting as choke points. The OSDs not only process client request, they also check on each other

SUSE Enterprise Storage is highly scalable, providing a clear business advantage as your data storage continues to shift and grow.

and send monitoring data back to the monitor node, which then updates the CRUSH map to reflect that status. This distributed approach eliminates bottlenecks that are present in centralized block-storage systems.

By using commodity servers and drives, and keeping them separate from the management system, SUSE Enterprise Storage allows an enterprise to build a highly-scalable storage system that avoids bottlenecks and single points of failure, and provides a consolidated, intuitive GUI for management.

Another feature offered by SUSE Enterprise Storage is scalability. The CRUSH algorithm helps balance storage across the entire storage cluster. And because it provides data access at the filesystem, object and block levels, SUSE Enterprise Storage is extremely flexible, allowing you to service a wide variety of use cases. For example, a virtual machine can be deployed on an RDB block device presented by the cluster. The same virtual machine can then use S3 object storage provided by the cluster to store backup or application data.

Once you've built and configured your SUSE Enterprise Storage cluster, it pretty much manages itself. Most of the time spent before putting SUSE Enterprise Storage into production should be utilized in planning the cluster topology. This includes power, network, etc. It also directly affects the CRUSH map, which when configured to properly reflect the topology enhances data protection and performance.

Tying all these features together is the management system of SUSE Enterprise Storage, which requires very little intervention because it constantly monitors data utilization and calculates the best possible data placement in accordance with the CRUSH map. This guarantees the best performance possible while minimizing the need for manual intervention. And while SUSE Enterprise Storage provides effective background monitoring, it also reports on the performance of your data storage.

SUSE Enterprise Storage provides a single, unified user interface (UI) to manage your storage. This in itself helps reduce OPEX because you don't have to worry about training users on multiple UIs. But there are other clear business benefits. SUSE Enterprise Storage is highly scalable, providing a clear business advantage as your data storage continues to shift and grow. Because you aren't committed to any single hardware vendor for your storage needs, and can make use of commodity hardware, your data storage solution is flexible and low-cost. SUSE Enterprise Storage also offers a flexible, node-based pricing model which means you don't have to pay more to add more storage capacity to your existing hardware.

Services You Need and Why

By now, all this discussion about the benefits of open-source SDS has probably got you thinking you should look into how to go about making the transition from your current, varied, and expensive storage solutions. This is where **SUSE Global Services** can add tremendous value. The truth is that your organization probably does not have the skillset to make the transition. According to research by the 451 Group, 84 percent of enterprises realize they don't have the in-house skills to enable the technology transformations they need to remain competitive.³

The amount of time and expense associated with trying to hire the skills you need makes any manager think twice, not to mention the risk of a bad hire, or having a great employee hired away just as progress was being made. SUSE Global Services can help you design, plan and deploy a solution by providing the right skills and experience to ensure your storage transformation is done right. In addition, engaging with an experienced and professional services provider will ensure your team receives the knowledge transfer and advocacy to make moving to a support model seamless, while ensuring your team has the knowledge base for future enhancements to your open-source SDS solution.

Not only does a services partner bring the right skills to bear on an open-source SDS solution, but augmenting your team with these professionals keeps your team free to focus on the highest priority: running the business. With SUSE Global Services your team won't feel the crunch of not having enough time to deliver on competing priorities but will feel the relief of having a highly competent, highly professional partner there to help them get where your enterprise needs to be.

The cost of doing it yourself can end up being an albatross that never delivers the promised savings or being prohibitive to even starting an SDS initiative. The learning curve of implementing a new technology for the very first time is just too huge for most organizations, often creating project delays, cost overruns, and technical debt once a solution is delivered. But having a knowledgeable and experienced partner can cut right through all of these risks to help ensure your open-source SDS solution delivers on its promises.

SUSE Global Services offers very flexible consulting, delivery and premium support services that are aimed specifically at creating value for their customers from their open-source SDS solutions. They start their process by helping customers create a data strategy that is clearly aligned to the needs of the business by assessing existing processes and infrastructure, and creating a structured roadmap and implementation for the storage solution. Once an approach that is clearly aligned to business strategy and direction has been defined, SUSE Global Services partners closely with their clients to select, implement and fully integrate an SDS that meets the needs of the business.

³ Dr. Katy Ring. "The Need for IT Services for Better Business Alignment". 451 Research. 2018.

CONCLUSION

The data explosion is here, happening in your data center right now. To stay competitive, you want to make the smartest use of your data you can. That requires a storage management solution that is scalable, reliable and cost-efficient. If this has you wondering whether you have the right data management strategy or in-house skillset, consider an open-source SDS solution, like SUSE Enterprise Storage. And be certain to engage with an experienced and proven partner, like SUSE Global Services to get the most out of your storage solution and your data.

