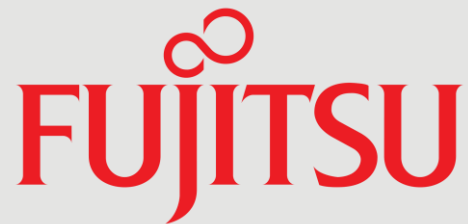




Inside Track Research Note

In association with



# Hyper-Scale Data Management

An open source-based approach  
to Software Defined Storage

January 2015

---

**About this Inside Track**

The insights presented in this document are derived from independent research conducted by Freeform Dynamics. Inputs include in-depth discussions on the latest technology developments with IT vendors and service providers, along with intelligence gathered from mainstream enterprises during broader market studies.

---

---

*Coming up with a single definition of SDS is hard. It's therefore better to focus on the shapes of solution that seem to be emerging.*

---

## In a nutshell

The term 'Software Defined Storage' is used to describe a wide range of ideas and offerings. This makes it hard to pinpoint specific opportunities. Fully supported enterprise class solutions based on open source technology, however, are driving practical and economic benefits in the context of today's storage challenges.

## Software defined everything

If you work in IT, the term 'software defined' will not have escaped you as one of the latest ideas to be promoted in the drive for greater efficiency and flexibility. It has surfaced across the industry in a number of different ways.

Server virtualisation led to private cloud architecture, which was in turn extended to the 'Software Defined Datacentre'. Some of the latest ideas in decoupling key elements of communications technology were then branded 'Software Defined Networking', and not to be left out, the data management community is now talking about 'Software Defined Storage', or 'SDS' for short.

The common thread that runs through all of this is the idea of moving control and management functionality from the hardware tier into an independent software layer. The way in which the infrastructure is configured and run is then no longer reliant on the many and varied proprietary embedded capabilities that exist in most data centres today. A more consistent set of software tools is used instead.

## From theory to practice

In theory, everything from policy definition, through resource provisioning and configuration, to ongoing optimisation, monitoring and administration, can be done centrally in a more joined up and flexible manner once your infrastructure is software defined. The challenge is that pretty much every vendor selling storage tooling or middleware has taken this as a licence to reposition their offerings as SDS. Put this together with the term also being used in the context of scale-out systems to handle huge data volumes, and the result is a lot of confusion over what SDS actually is.

Against this background, coming up with a single definition of SDS is hard. It's therefore better to focus on the shapes of solution that seem to be emerging:

**Mainstream SDS:** This is about abstracting functionality such as thin-provisioning, compression, deduplication, replication, snapshotting and backup/recovery into generic software that can operate in a heterogeneous hardware environment.

**Storage Virtualisation:** Some regard software that pools and virtualises capacity across devices as SDS, so we have included it here. However, this approach is best thought of as a complement to SDS. It is possible to implement Mainstream SDS with or without virtualising.

**Hyper-scale SDS:** The focus here is on high-performance distributed processing software that can be used to create a massively scalable, automated and resilient environment. The aim is to meet escalating storage needs in a flexible, unconstrained and cost-effective manner.

## Opportunity emerging from the noise

---

*Hyper-Scale SDS creates opportunities to tackle a whole new range of challenges stemming from the relentless growth in the volume and variety of data.*

---

Vendors and pundits look set to continue debating what constitutes SDS in a mainstream storage context, and whether storage virtualisation falls within the SDS category or is an aspect of the broader Software Defined Datacentre concept. As you explore such questions yourself, beware of those promoting closed, proprietary hardware and software combinations, as this is counter to the open spirit of SDS.

Also bear in mind that in most cases the Mainstream SDS proposition nets out to doing largely the same things as you were doing before, but in a more joined up, efficient, flexible and heterogeneous manner. The opportunity is therefore primarily to streamline and automate, and perhaps to reduce your reliance on specialist skill sets – particularly in relation to proprietary embedded capabilities.

In the meantime, Hyper-Scale SDS creates opportunities to tackle a whole new range of challenges stemming from the relentless growth in the volume and variety of data that many organisations are seeing. Now you have an alternative to buying traditional high capacity solutions based on single vendor stacks, which are typically expensive and continue to consume significant cash as requirements grow.

## A new breed of Hyper-Scale storage

The challenges of high-volume storage and rapid growth have been particularly felt within the internet and cloud service provider community. It is not uncommon for larger players in this sector to manage data from tens or even hundreds of millions of consumers and/or high numbers of business clients. The commercial and practical limits of proprietary storage solutions were identified in this context many years ago.

The result has been the emergence of a range of open source software (OSS) solutions designed for so-called ‘web-scale’ deployments. Many of the ideas, techniques and projects in this space originated from in-house developments by larger service providers themselves, before being picked up and further enhanced by the broader OSS community. As such, they were born out of the practical need for extreme levels of scalability, resilience, automation and economies of scale.

It is beyond the scope of our discussion to go through the many OSS SDS solutions that exist out there, but as you read around the topic, you’ll come across names such as Ceph, Gluster, Nutanix and Swift – the latter falling out of the much discussed OpenStack initiative. The range of offerings is pretty broad, so whether you need a distributed file system, scalable object storage, or something optimised for a particular need such as virtual server management, you can be sure of finding a relevant option. But working with fast moving OSS solutions isn’t easy.

## From DIY to enterprise solution

Most enterprises are understandably wary of adopting software straight from OSS projects. The core of the solution might be functionally stable and very well hardened, but it is not uncommon for features beyond this to vary in their level of mainstream readiness, or to still be on the ‘to do list’. In practical terms, you therefore need to work through what’s there and what’s not, what’s still ‘experimental’, and ultimately what’s wise to use or avoid for the moment.

---

*But working with fast-moving OSS solutions isn’t easy.*

---

---

*Enterprises are understandably wary of adopting software straight from OSS projects.*

---

---

*One approach is to identify a trusted supplier that takes care of all the complexity for you and protects you from the uncertainty.*

---

---

*The right vendor will package up the core capability, fill necessary gaps, and make the software available as a 'product'.*

---

---

*It is critical to define your needs precisely so you can assess the fit of available OSS SDS solutions and suppliers.*

---

As OSS projects are constantly evolving, there is then the challenge of keeping up with developments and managing change. Indeed to create a viable solution, you might even have to draw upon the output of multiple OSS projects, or sub-projects, which means establishing and keeping track of a whole set of integrations and dependencies too. And all of this, of course, assumes you have the specialist skills, bandwidth and inclination to work in such a 'Do It Yourself' (DIY) fashion.

If you are interested in taking advantage of OSS-based Hyper-Scale SDS solutions, which solve problems that are becoming increasingly common in an enterprise environment, one approach is to identify a trusted supplier that takes care of all the complexity for you and protects you from the uncertainty.

The right vendor will package up the core capability, fill necessary gaps, and make the software available as a 'product' with a support and maintenance arrangement delivering the safety and predictability required. Some may even pre-integrate the software with suitable hardware to produce a turnkey platform or appliance. With the right formula, you gain the benefits of an OSS-based scale-out storage environment, without the risks, overheads and the need for esoteric skill-sets.

## Taking advantage of the opportunity

But identifying a suitable supplier and solution is only part of the equation. You must make sure you have analysed your needs comprehensively and determined the type of solution that will best meet them. Considerations here include the mix of data you are intending to store, both in the shorter and longer term, your application and workload integration requirements, and, not least, the kind of access mechanisms and security model you need to support.

At the time of writing, while solutions might have a reasonably broad scope of functionality and support for standards and APIs, most have areas in which they are stronger or weaker depending on their origin. This is fine if current points of immaturity correspond to needs further out on your requirements roadmap, but it's important to make sure that efforts are being made to strengthen and enhance software capabilities where necessary.

There are then fundamental philosophical and practical differences to consider. Ceph, for example, is designed with strong consistency in mind – a write to disk is an immediate and permanent write. This makes it more suitable for hard-core enterprise application requirements than solutions favouring 'eventual' consistency for performance reasons. The opposite might be true if your primary aim is hosting virtual machines in a private cloud or VDI context, where performance takes precedent. With this in mind, it is critical to define your needs precisely so you can assess the fit of available OSS SDS solutions and suppliers.

## The bottom line

Despite the current confusion around software defined storage, Hyper-Scale SDS in particular can help to solve problems that have to date been difficult to manage cost-effectively. Significant opportunities therefore exist for mainstream enterprises. Bear in mind, however, that the DIY approach may require substantial specialist skills and resources in areas such as sizing, integration, tuning and maintenance. A packaged solution from the right supplier can reduce the burden and risks, and accelerate ROI.

## About Freeform Dynamics

Freeform Dynamics is an IT industry analyst firm. Through our research and insights, we aim to help busy IT and business professionals get up to speed on the latest technology developments, and make better informed investment decisions.

For more information, and access to our library of free research, please visit [www.freeformdynamics.com](http://www.freeformdynamics.com).

## About Fujitsu

Fujitsu is the leading Japanese information and communication technology (ICT) company offering a full range of technology products, solutions and services. Approximately 162,000 Fujitsu people support customers in more than 100 countries. We use our experience and the power of ICT to shape the future of society with our customers.

### Business-centric storage

Under the ETERNUS storage solution brand, Fujitsu provides disk systems and data protection appliances, which allow customers to flexibly manage their increasing data volumes at a lower cost of growth, and to benefit from a very reliable architecture and radically simplified operation.

### Hyper-scale storage

ETERNUS CD10000 provides unlimited, modular scalability of storage capacity and performance at zero downtime for instant and cost efficient online access to extensive data volumes. Integrating open-source Ceph software into a storage system delivered with end-to-end maintenance from Fujitsu, enables IT organisations to fully benefit from open standards without implementation and operational risks.

For more information, please see

<http://www.fujitsu.com/global/products/computing/storage/eternus-cd/>.

## Terms of Use

This document is Copyright 2015 Freeform Dynamics Ltd. It may be freely duplicated and distributed in its entirety on an individual one to one basis, either electronically or in hard copy form. It may not, however, be disassembled or modified in any way as part of the duplication process. Hosting of the entire report for download and/or mass distribution by any means is prohibited unless express permission is obtained from Freeform Dynamics Ltd or Fujitsu. The contents contained herein are provided for your general information and use only. Freeform Dynamics Ltd provides no warranty or guarantee as to the suitability of this document for any particular purpose.