

## Solution Showcase

# Primary Storage Data Reduction Trends

**Date:** June 2016 **Authors:** Mark Peters, Practice Director and Senior Analyst—Infrastructure; Adam DeMattia, Director—Custom Research; and Bill Lundell, Director—Syndicated Research

**Abstract:** Organizations are dealing with rapid data growth in their primary storage environments. This data growth is at the heart of many of the challenges organizations face with respect to storage. However, for many organizations, adoption of data reduction technologies offers IT a way to combat and ameliorate some of those challenges. Data reduction technologies can lower costs and make more efficient use of capacity (those are the obvious and expected impacts), *and* they can lead to better overall storage system performance (a less intuitive, but definitely valuable impact). Custom ESG research shows that IT organizations are gaining substantial capacity, cost, *and* performance benefits from data reduction—and that a majority are now identifying data reduction as a “must-have” feature of their primary storage systems. This dynamic of *smarter storage* rather than just *incremental capacity* aligns with IBM’s vision for modern storage infrastructure.

## Primary Storage Growth Is Significant and Brings Significant Challenges with It

We cannot continue to accept and manage data growth as a problem that simply “happens” to “inactive” storage infrastructures. The challenges of cost and management are simply not sustainable. Because the increase in data creation is evidently not going to stop, we must instead find ways to reduce the extent of the impact of that increase upon the infrastructure—and a key part of the answer lies in “active” storage elements that don’t just *accept* data, they *reduce* it to minimize capacity, cost, and management.

While the main elements within “data reduction” (that is, deduplication and compression) are heavily promoted, there has been little research into the absolute and comparative efficacy of data reduction tools (vendors’ claims and guarantees are often purely marketing efforts). Are they simply a “nice to have” item, or are they becoming a known and efficient necessity? ESG recently completed a market survey to get specific answers to these questions.<sup>1</sup>

Of course, data growth, and the pressure it places on IT organizations, is nothing new. IT vendors and even end-users often speak hyperbolically about “exponential” or “explosive” data growth. Although the hype around data growth is sometimes overstated in the headlines, ESG survey data bears out that the pace of storage capacity growth, which is obviously closely correlated to data growth, is significant. In a recent survey of 200 IT respondents with day-to-day data storage management responsibilities at their organizations, a plurality (45%) reported their deployed primary storage capacity to be growing at more than 40% year-over-year.

---

<sup>1</sup> The research was conducted in the first quarter of 2016, was sponsored by IBM, and covered 200 US-based IT professionals responsible for evaluating, purchasing, and managing data storage technologies for their organizations (respondents were 30% midmarket and 70% enterprise, representing a broad range of industry verticals); all respondent organizations were currently using or planning to use data deduplication or compression as a method of managing primary data.

## Fast-paced Data Growth Has Explicit and Implicit Links to Many Storage Environment Challenges

Organizations are facing numerous challenges with respect to their storage infrastructures (see Figure 1). However, many of those challenges can be tied back to the extreme rate at which data (and thus capacity) is growing, either *explicitly*, as is the case with challenges related to:

- **Hardware costs:** Storage systems are expensive and as capacity grows, so too does the associated hardware cost. Whether the organization elects to scale up existing storage systems or scale out the number of storage systems in place, it will incur a significant hardware cost.
- **Rapid data growth:** Data growth can be difficult to accommodate; many aspects of storage management are made more difficult in fast-growing environments: capacity planning, storage architecture and deployment, system tuning, etc. Thus, it is not surprising to see that rapid data growth, in and of itself, was cited frequently by respondents as a pain point.
- **Running out of physical capacity:** IT departments are bound by fixed budgets and refresh cycles. However, data growth is a constant pressure. This can lead to a mismatch in the physical capacity available and the physical capacity needed.

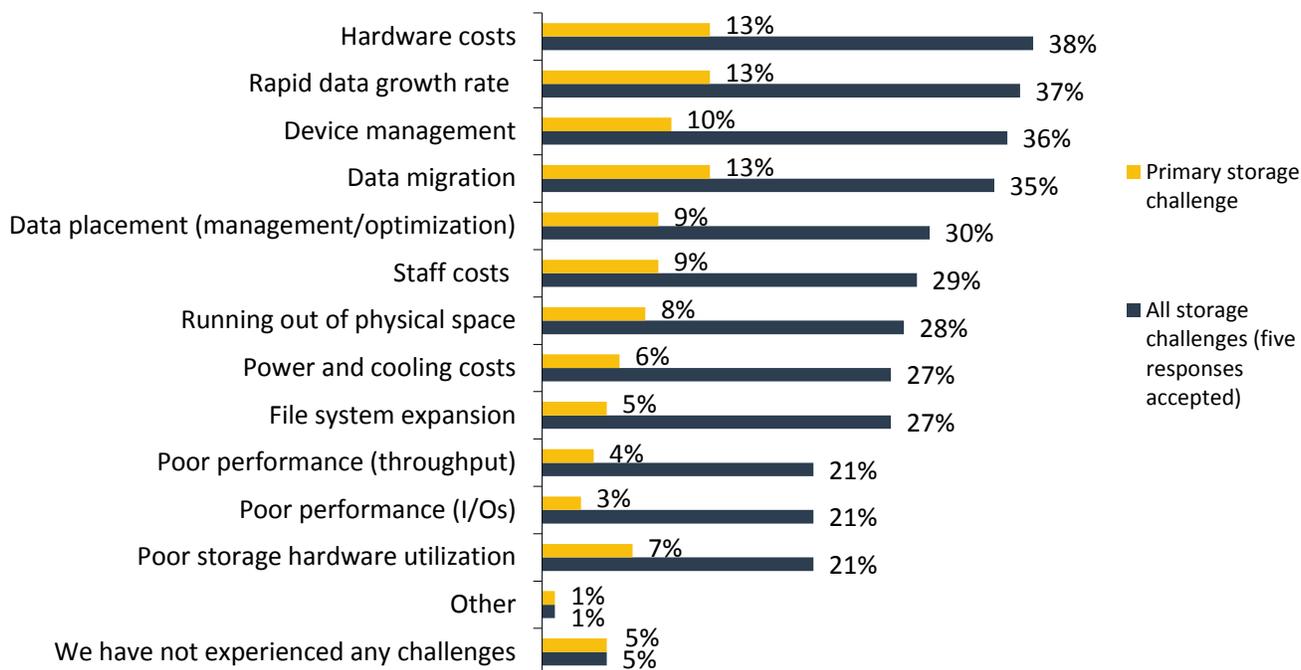
... or *implicitly*, as is the case for challenges related to:

- **Device management:** The number of storage devices is logically correlated to the amount of data to be housed. As data grows, managing the growing number of storage systems underlying that data becomes a greater burden on the IT organization.
- **Data migration:** As storage requirements outgrow the capabilities of the underlying storage system, painful data migrations can become commonplace.

If challenges are grouped broadly in terms of being either capacity- or performance-centric, the challenges aligned with capacity growth clearly are top of mind. Performance challenges (i.e., being IOPS- or throughput-bound) are less frequently experienced in customer environments. This phenomenon, too, can be seen in the ranking of storage environment challenges in Figure 1, regardless of whether one looks at the single main challenges respondents report (the yellow bars on the chart) or at the top five challenges to which they attest (the blue bars on the chart).

**Figure 1. Storage Environment Challenges**

**In general, what would you say are your organization’s biggest challenges in terms of its storage environment?  
(Percent of respondents, N=200)**



Source: Enterprise Strategy Group, 2016

### Data Reduction Technologies Deployment and Achieved Value

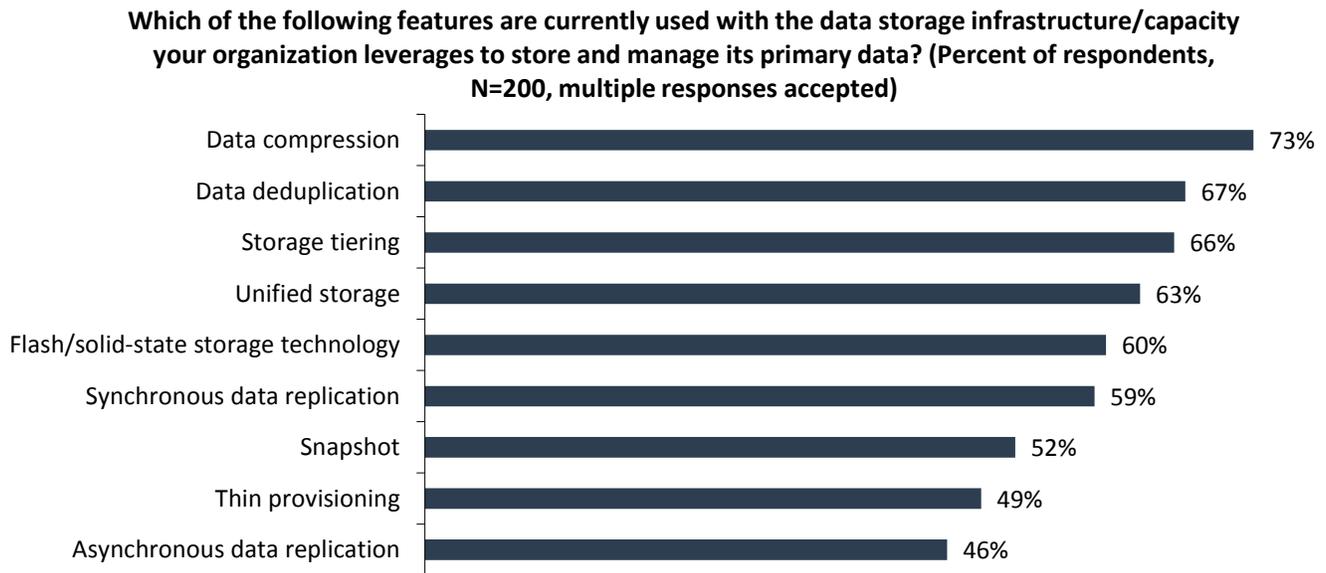
With data and capacity growth driving many of the challenges for IT with respect to storage, it is not surprising to find that data reduction technologies—compression and deduplication being the most often discussed and marketed today—are the most broadly utilized storage management features reportedly being leveraged by organizations. Data compression was reported to be in use by 73% of respondents, while two-thirds reported data deduplication is in use today (see Figure 2).

It is interesting to note that these data reduction technologies outstrip more obviously performance-centric technologies such as flash/SSD deployments or storage tiering in terms of adoption. This data point lends further credence to the hypothesis that data and capacity growth are the most central of the storage challenges that most organizations face today, ahead even of performance challenges. This should not be read as a negative comment on flash adoption for a couple of reasons: first is that data reduction technologies [can] apply across all storage systems and, hence, media types; secondly, as technologies advance and pricing improves, flash is increasingly becoming a genuine and affordable capacity play in addition to the turbo-boost performance fix that it started as.

Of course, not all data needs flash performance, so it may not need to be accelerated—but it *always* should be looked upon as a candidate for reduction.

Additionally, it is worth noting that data reduction technologies are not being deployed just for niche applications or only in corner cases, since 58% of the respondents reported that all of their organization’s primary data is or will have data reduction tools applied, irrespective of the workload.

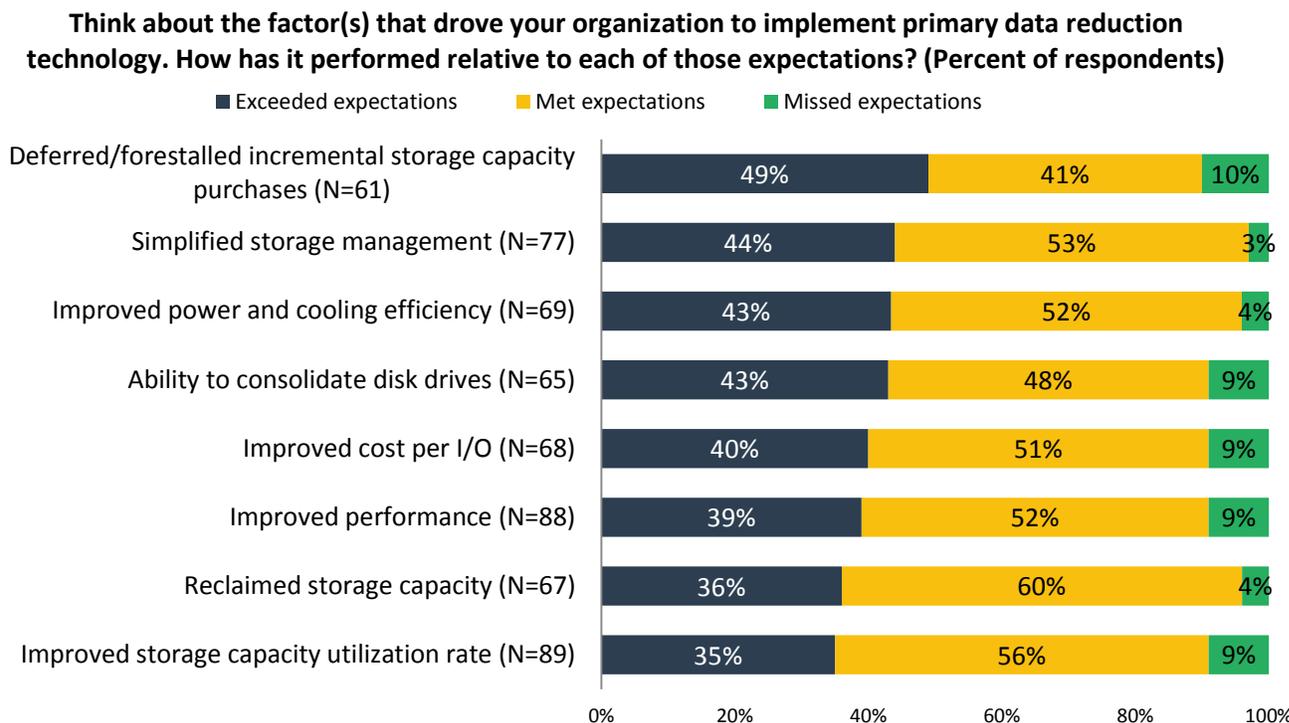
**Figure 2. Storage Features In Use Today**



Source: Enterprise Strategy Group, 2016

Obviously the question that next begs to be asked is: How effective are data reduction technologies at blunting the pain associated with all this data growth? When ESG asked respondents to rate how successful they were at achieving the specific outcomes that led them to implement data reduction technologies, *the vast majority of respondents reported they have met or exceeded their expectations* in areas such as deferring storage purchases, simplifying storage management, and improving storage power and cooling needs, among others (see Figure 3).

**Figure 3. Primary Data Reduction Outcomes Relative to Expectations**



Source: Enterprise Strategy Group, 2016

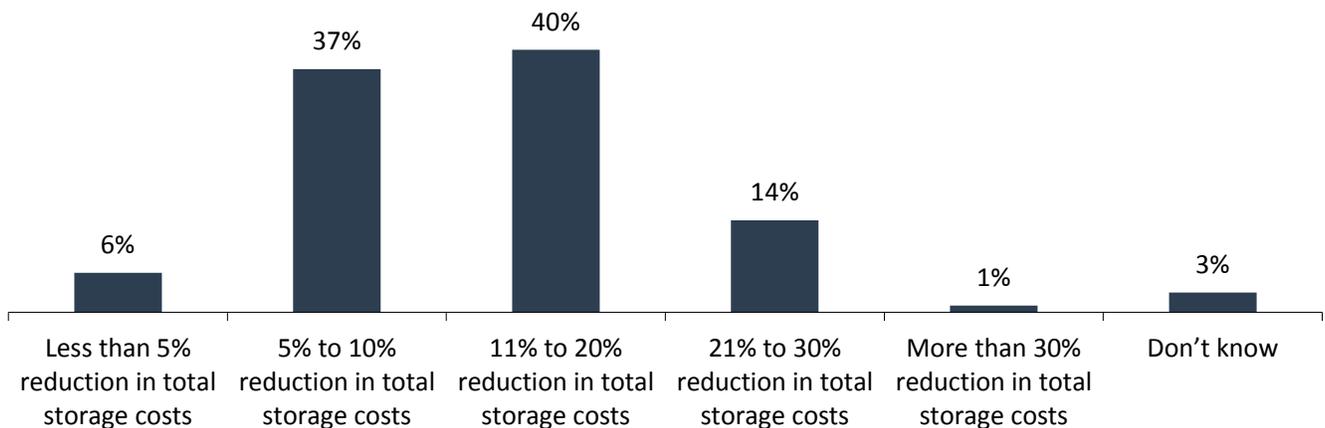
Most respondents reported extremely impressive capacity-reduction results from their data reduction deployment, and more than 70% reported having achieved capacity reductions exceeding 10X. Although that is an extremely impressive result, it may have been somewhat inflated by early “low-hanging fruit” and by the fact that many respondents (58%) report having tried, then disabled, data reduction on some or all of their workloads. That said, even allowing a little leeway in interpretation, it is clear that the claims of many vendors to achieve 3-5X data reduction are reasonable and are being met. All this said, though, there is no single, all-encompassing panacea in storage or indeed for data reduction (where, for example, data *deletion* is also a powerful and underutilized tool across much of IT). So a good measure of the true overall value and impact of data reduction tools can perhaps most fairly be represented by the financial impact. And the news is impressive.

When respondents were asked to report on their outcomes in terms of financial success as opposed to specific tactical successes, the results were unequivocally impressive. Just 6% of respondents reported less than a 5% reduction in their total storage infrastructure CapEx and OpEx tied to data reduction technologies (see Figure 4). This contrasts with the 37% of respondents who reported a 5-10% overall savings, and 40% who reported an 11-20% reduction. Indeed, there were also 15% who reported a greater than 20% reduction. Put simply, more than three quarters of respondents were able to save 5-20% of their overall storage expenditure by implementing data reduction technologies.

More than three quarters of the IT organizations surveyed by ESG reported that implementing data-reduction technologies saved them between 5% and 20% of their overall storage expenditures.

**Figure 4. Cost Savings Attributed to Primary Data Reduction Technology**

**Taking into account all of the costs (i.e., CapEx and OpEx) that go into the data storage infrastructure deployed to support your organization’s primary data, approximately what is the total cost savings that you believe primary data reduction technologies have delivered? (Percent of respondents, N=172)**



Source: Enterprise Strategy Group, 2016

### Opportunities—and Questions—for Data Reduction Technologies

The current state of data reduction technology usage, and the outcomes that usage can drive, is clear. Most organizations are utilizing data reduction technologies, oftentimes across all their primary data, to help ameliorate the many challenges brought on by rapid data growth.

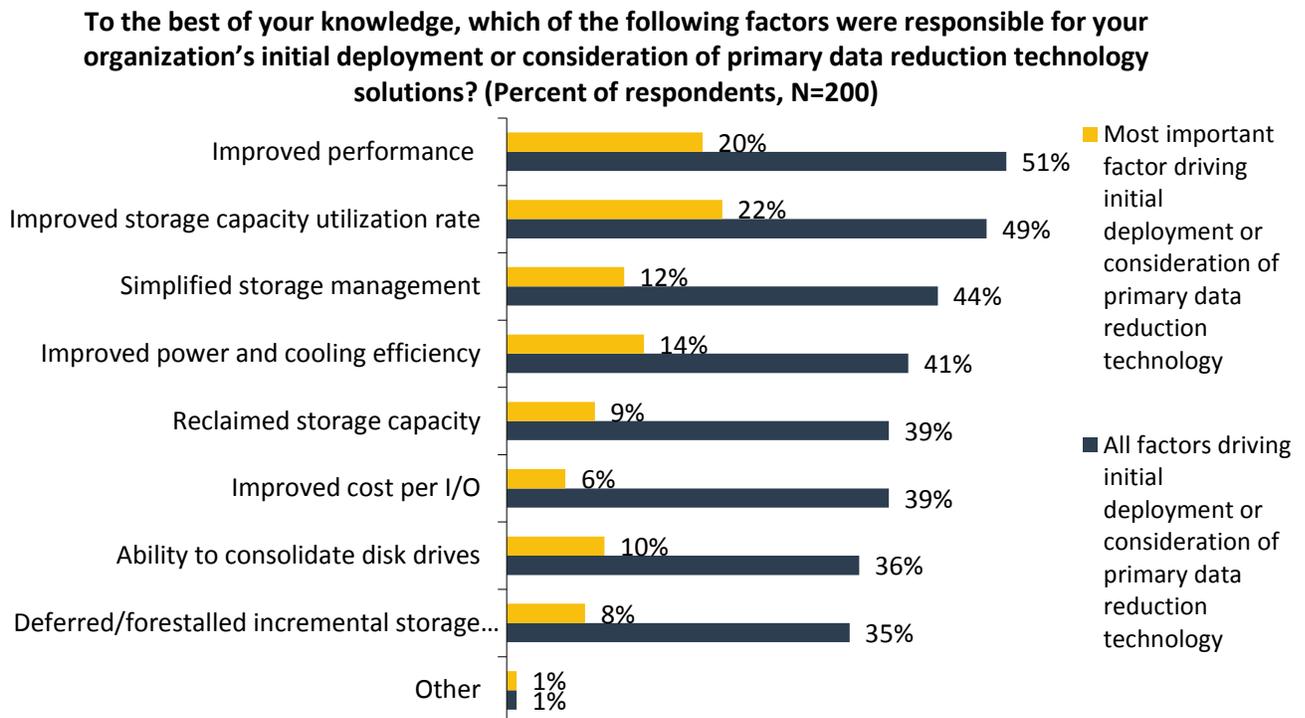
As with any technology, there are still questions and issues around data reduction:

- For example, a question about where data reduction technologies should sit in the technology stack garnered varying responses. Thirty six percent reported that data reduction should reside at the host OS or hypervisor layer (probably a

nod to the high degree of infrastructure virtualization present in most organizations today), while 27% suggested the storage system layer, 18% the application layer, and another 18% the software-defined storage layer. While the verdict is in with respect to the results data reduction technologies can deliver, the jury appears to still be out with respect to the “best” way to implement these technologies.

- Certainly the demand for data reduction technologies exists; when asked about a range of must-have storage system features, the largest percentage of respondents chose compression, making it the most popular response (tied with HA), and deduplication was the fourth most popular response.
- That relative ranking of the two features is interesting because a significant majority of the respondents reported the relative results of deduplication as superior to those achieved with compression; 83% said they had better or significantly better results with deduplication, whereas only 7% felt they had achieved more through compression. Given that compression works on a much larger percentage of workloads than deduplication, this result is intriguing—but it is probably affected by the lower number of vendors that offer compression as well as the massive marketing effort that has supported dedupe over the last few years.
- While reducing data capacity requirements and saving money are the obvious attractions of data reduction, there is also a very compelling user realization that reducing the sheer volume of stored data can also help “free up the system” and thereby improve performance. So while better storage utilization was the most mentioned *primary* factor driving data reduction deployment (the yellow bars), when respondents were allowed to provide all their reasons (the blue bars), improved performance rose to the top spot (see Figure 5).

**Figure 5. Factors Responsible for Organizations’ Deployment or Consideration of Data Reduction**

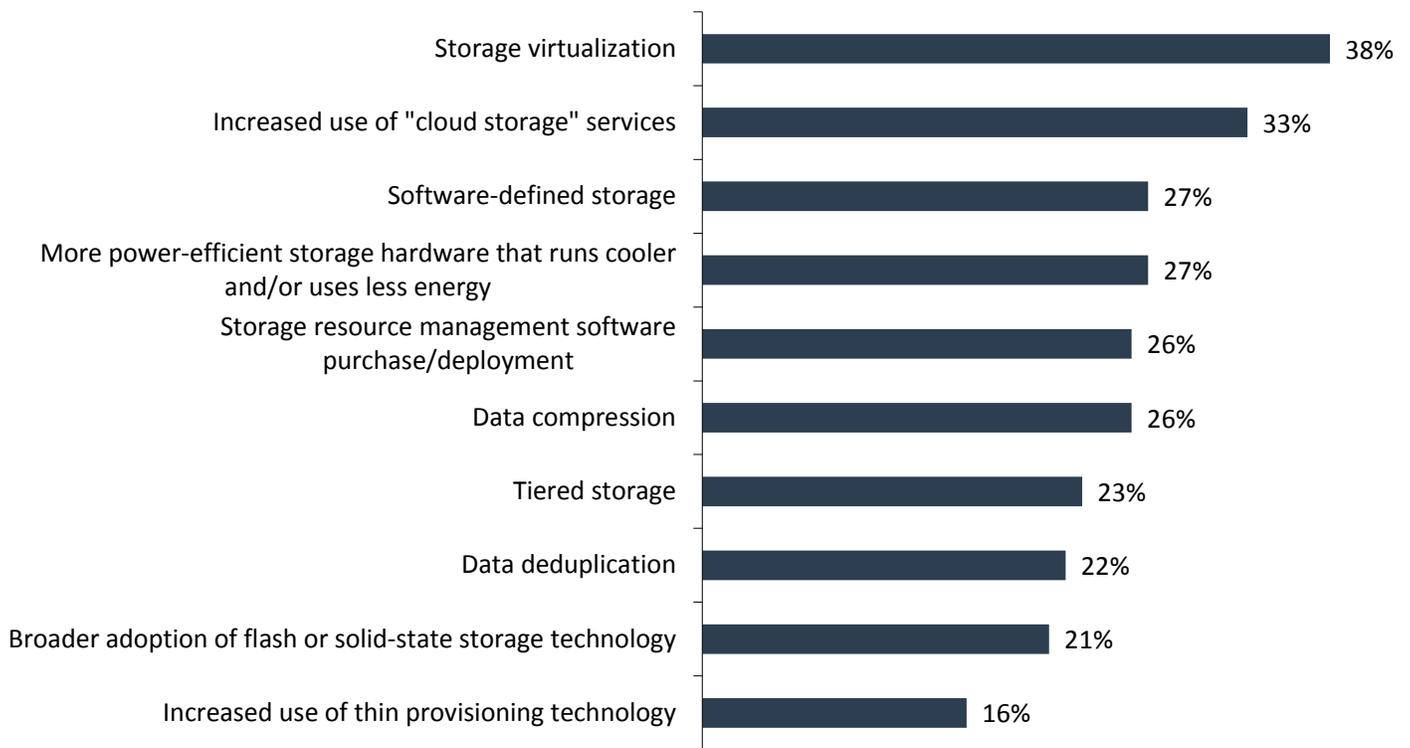


Source: Enterprise Strategy Group, 2016

- For all the impressive results data reduction technologies can help drive, they do not rate top in the hierarchy of storage technologies that are expected to drive storage efficiency for organizations over the next 24 months. Rather, concepts such as storage virtualization, cloud storage, and software-defined storage all outpaced both data compression and data deduplication (see Figure 6). Thus vendor offerings that can address most or all of these broader requirements (including data reduction) are likely to fare best. The sponsor of this *generic* market research, IBM, clearly has a *specific* play in this regard, with its latest A9000 and A9000R FlashSystem offerings.

**Figure 6. Storage Technologies Expected to Drive the Greatest Efficiency Impact 24 Months from Now**

**Please consider the following data storage technologies. Over the next 24 months, which three do you believe will have the greatest impact in terms of improving the efficiency (i.e., will result in reduced capital and/or operating costs) of the storage infrastructure your organization has deployed to store and manage its primary data? (Percent of respondents, N=200, three responses accepted)**



Source: Enterprise Strategy Group, 2016

- While data reduction technologies have made great strides of late, and achieved generally excellent results for those IT departments that use them (often exceeding their expectations), there is also clearly room for further market education. As such tools become—from many vendors—both no-added-cost and ever-easier-to-use, so too will their value to aid against the data tsunami become more compelling.

### A Note on IBM’s Flash Offerings

The latest IBM flash offerings (FlashSystem A9000 and FlashSystem A9000R) should resonate in a storage market that clearly not only values data reduction but also demands a great deal more from its chosen storage systems. The latest IBM FlashSystem platforms offer both comprehensive in-line data reductions (which include new compression algorithms, new flash-optimized dedupe, and also even pattern removal) and an extensive range of all the attributes from Figure 6 that are viewed as being needed to drive storage efficiency. As IBM continues to optimally mix abilities from within its Spectrum

family, FlashSystem A9000 and FlashSystem A900R are built on Spectrum Accelerate technology, which excels at internal data management, yet does so with low latency penalties.

## The Bigger Truth

With capacity efficiency a clear element of, and underlying, many top data storage challenges, the attraction of data reduction tools is not only obvious, but ESG's research also shows that most users who have deployed such tools are getting good results that often exceed their expectations.

The achieved outcomes from deduplication and compression to date have been impressive, and these technologies are moving from luxury to necessity. Data reduction technologies have been able to deliver substantial improvements in storage capacity/utilization and cost (as would be the obvious expectations), while helping with storage system performance (a less obvious outcome but one that is very motivational to users in adopting these technologies).

Of course, no part of the storage "engine" works in isolation to move the IT "vehicle" forward; while this research shows the tangible and desired benefits of data reduction, other storage technologies—such as virtualization and SDS—are perceived to drive equal or more value for overall storage efficiency over the next few years. The best advice for IT professionals is to look for systems that excel across the functional board—storage engines, in other words, that can fire on all cylinders.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of The Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.



**Enterprise Strategy Group** is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

© 2016 by The Enterprise Strategy Group, Inc. All Rights Reserved.

