ESG Brief

Data Storage Predictions for 2020

Date: December 2019 **Author:** Scott Sinclair, Senior Analyst

Abstract: As we approach the conclusion of 2019, it's time to reflect on the current state of data storage technology and try to forecast what's next for data storage in 2020.

What We Learned in 2019

Multiple storage technology storylines highlighted 2019. From the hardware side, we saw more announcements that included NVMe and NVMe over fabrics (NVMe-oF) technology. We heard announcements of systems that integrate storage-class memory (SCM), and several more identified as SCM-ready. We saw industry-wide consolidation with several smaller storage innovators getting acquired not only by their larger on-premises counterparts, but also public cloud providers, with Google acquiring Elastifile. What will the next 12 months bring?

Before we get to that, let's look back at last year's projections.

Did I Get Anything Right Last Year?

Last year at this time I made three, what I considered bold, predictions. I didn't want, and still don't, to be obvious in my predictions; I wanted to make it more difficult and try to forecast something outside of the norm. Last year, I predicted that the percentage of organizations that had pulled at least one workload back from the cloud would increase. I predicted that NVME over Fabrics adoption would increase with FC growing at a faster rate. And finally, I projected that many AI projects would fail due to lack of content awareness in tools and technologies.[1]

In a 2019 study of IT storage decision makers, 55% of organizations pulled a workload back from the cloud.[2] I am going to count that as something I got right. For anyone who has not read my earlier writing on workload repatriation: These moves ***are not*** an indictment of the cloud. The moves, though common, only happen with a handful of applications, and IT organizations that pull workloads back generally continue to use the cloud and think highly of public cloud services. These moves are often the result of overuse of a "lift and shift" approach to shifting workloads to the cloud, not conducting proper due diligence, and not refactoring the apps. These events will likely eventually fade overtime as cloud adoption and expertise grow, but until that time, these efforts increase the cost of the cloud adoption.

[1] See ESG Blog, *2019 Data Storage Predictions: More Cloud Missteps, FC Is Back, and Finding Data Holds Back AI*, January 2019.
[2] Source: ESG Master Survey Results, *2019 Data Storage Trends*, to be published November 2019.

On NVMe over fabrics, it is still early, but the results so far are different than I expected. Based on current research data, the preference for NVMe over fabrics adoption is trending toward Ethernet.[3] The prevalence of Fibre Channel combined with how easy it is to upgrade an existing FC environment to NVMe over Fibre Channel still suggests that NVMe over FC will have strong adoption moving forward, but the jury is still out on this.

On AI projects failing due to lack of content awareness, I might have been too bold on that claim. The value created by AI initiatives is so high that inefficient data management adds cost but is not causing the initiative to be considered a failure. I still expect, however, that understanding content will be a major storage focus area moving forward.

What's in Store for 2020?

First, let's get the obvious out of the way: Data will continue to grow. Public cloud adoption will increase. We will see increases in investment in application development and modern Dev/Ops practices. We will also see continued investment in highly valuable data-centric workloads such as analytics, machine learning, and deep learning. These trajectories started prior to 2019, and we will likely see them to continue into 2020 and beyond. Otherwise, here are my predictions for 2020:

One or More Public Cloud Providers Will Acquire an On-premises/Hybrid Data Storage Company

The long-held perception that eventually all IT will be public cloud-based is starting to fade. Hybrid cloud infrastructure is the de facto standard of modern IT. And for digital businesses, data is increasing on- and off-premises at similar rates. In other words, on-premises infrastructure will be a significant part of IT for the foreseeable future. In response, multiple public cloud providers have announced hybrid cloud solutions, such as AWS Outposts, Google Cloud's Anthos, and Microsoft Azure Stack. Then last year, Google acquired Elastifile to bolster its cloud file storage portfolio. I expect that this move is only the beginning. As the cloud space becomes more competitive and as hybrid offerings become more important, the need for advanced services and capabilities will fuel more technology acquisitions, with file storage being a likely target area for candidates.

IT Investment in Data Identification and Classification Will Accelerate

High-value workloads such as analytics, machine learning, and IoT workloads are driving data growth both on-premises and in the cloud. And IT organizations with extreme levels of data growth, such as 50% or more annually, are far more likely than slow data growth organizations to identify a lack of content awareness as a top data storage challenge.[4] I am stepping back a little from last year's prediction, as identifying and locating the right data is a major business problem *hindering* digital efforts, but not crippling them. This is still a major area of concern for digital businesses. I expect addressing data classification, especially with tools that expedite and maximize the use of metadata, across a hybrid cloud ecosystem to be a top three investment area for digital enterprises in 2020.

Consolidated Hybrid Cloud Operations Will Become the Standard IT Organization Model

As recently as two years ago, a majority of IT organizations managed public cloud services and on-premises infrastructure separately.[5] This year, I expect not only a significant majority of IT organizations to consolidate on- and off-premises responsibilities into one team, but also significant investment growth in management tools that span both on- and off-premises infrastructure environments. This prediction is based on an expectation that significant increases in Kubernetes and container-based workloads adoption will be the catalyst. Container-based workloads are far more likely to span on-

[3] ibid.
[4] ibid.
[5] Source: ESG Master Survey Results, *The Emergence of Multi-cloud Strategies*, April 2018.

and off-premises environments, demanding consistency across developers and IT. This need for consistency will fuel shifts in organizational structure and management.

Integrated Machine Learning Based on Telemetry Data Becomes Tablestakes For IT Infrastructure Buying

Neary every CIO wants an IT infrastructure solution tailored to his or her company's specific needs. Unfortunately, few, if any, organizations truly understand their own workload requirements. This lack of insight often becomes exposed when IT organizations shift applications to public cloud services, and the cost equation changes. Multiple infrastructure solutions have emerged which gather telemetry data based on the production application environment and then analyze that data with machine learning to provide recommendations as well as support automation. In 2020, these features will jump to the forefront of IT buying decisions.

The Bigger Truth

Taking a step back, I expect two high level themes to drive the storage industry over the next 12 months: Hybrid cloud will become the dominant IT design, management, and organizational model, and there will be a massive need for greater intelligence at the data level and at the application level. Both of these trends will change how businesses approach IT buying and architecture, which in turn will drive vendor behavior. Digital businesses are changing the rules for and the demands placed upon IT; over the next 12 months, we will see further evidence of that.