White Paper

Cloud-Based Predictive Analytics Becoming a Critical Source of Vendor Differentiation in Enterprise Storage

Sponsored by: HPE And Intel

Eric Burgener
October 2018

IDC OPINION

With the maturing of the all-flash array (AFA) market, the established market leaders in this space are turning their attention to other ways to differentiate themselves from their competition besides just product functionality. Consciously designing and driving a better customer experience (CX) is a strategy being pursued by many of these vendors. CX is a more comprehensive concept than the older customer satisfaction — it takes a look at customer-vendor interactions starting with how a customer learns about a vendor and decides to evaluate product to buying, deploying, managing, scaling, supporting, and ultimately, refreshing it with a next-generation solution. IDC has identified 11 different areas where vendors are working to improve their CX — and cloud-based predictive analytics (CBPA) figures prominently in this list. What is important about cloud-based predictive analytics is not just that vendors support such a platform, but how they use that platform to drive meaningful value in the areas of performance, capacity utilization, availability, efficiency, and cost for their customers. Vendors also have to be effective at communicating the value they are driving for customers with these platforms, or their efforts will have less of an impact to the CX.

Cloud-based predictive analytics platforms are a relatively new phenomenon, and they go far beyond the remote monitoring systems of a prior generation. Three key features differentiate cloud-based predictive analytics — data sharing, scope of monitoring, and use of artificial intelligence/machine learning (AI/ML) to drive autonomous operations. To help familiarize the uninitiated with specifically what types of value these systems can drive, IDC discusses them at some length in this white paper. HPE has the most mature cloud-based predictive analytics platform in the industry today — InfoSight — and is establishing many best practices in its use to drive value for customers. Today, InfoSight covers both the 3PAR StoreServ and Nimble Storage portfolios — both are powered by Intel 3D NAND solid state drives (SSDs) — and HPE plans to expand coverage across more products going forward. IDC is suggesting to our enterprise storage clients that they become familiar with cloud-based predictive analytics, understand the types of value it can drive for them, and use the value that vendors can drive for them with these platforms as a differentiator when evaluating new enterprise storage purchases.

October 2018, IDC #US44365218

IN THIS WHITE PAPER

As enterprise storage vendors focus on driving an improved customer experience across their entire installed bases, the use of cloud-based predictive analytics platforms is replacing traditional remote monitoring for support purposes. This white paper defines cloud-based predictive analytics, discusses evolving storage requirements that are driving their rise, and takes a look at how these platforms are being used to drive incremental value for customers in the areas of performance, availability, management, recovery, and IT infrastructure planning. It then examines HPE's InfoSight cloud-based predictive analytics platform, highlighting how HPE is driving value for its customers with this AI/ML-driven platform.

SITUATION OVERVIEW

As the enterprise storage market has matured, the requirements for an enterprise storage platform targeted for general-purpose, mixed workload consolidation have become very clear cut. Many vendors offer enterprise storage platforms that meet these well-known functional requirements, and it is becoming more difficult for customers to differentiate between many vendors that offer similar products. Over the past several years, more forward-thinking enterprise storage vendors have been focusing on creating a more positive customer experience over the life of their systems and are using this to differentiate themselves. As part of the evolution of CX, cloud-based predictive analytics platforms have been introduced by these vendors, but IDC has noted significant variability between vendors not only in the capabilities of their cloud-based predictive analytics platforms but also in their ability to clearly explain how they benefit customers.

Vendors' efforts to improve CX are often communicated through a standardized metric to measure it called the Net Promoter Score. Enterprise storage managers who want to better understand this metric can review *Net Promoter Score Becoming an Important Metric for Enterprise Storage Managers to Understand* (IDC #US43896818, June 2018). Vendors' efforts to create a differentiating CX span 11 identified areas, and "cloud-based predictive analytics" is one of those areas. Cloud-based predictive analytics represents a significant departure from the way vendors interacted with their systems in the past, and for those vendors that are consciously leveraging the platform to drive value for both their customers and themselves, they have become a differentiating competitive weapon.

Enterprise storage vendors have traditionally offered a remote support connection into their systems. For those customers that allowed it, a remote support connection would enable vendors to monitor a system in real time, primarily to rapidly identify failures, perform real-time troubleshooting, and institute other remedial responses. With the availability of this type of real-time remote monitoring, vendors were able to support faster resolution of problems when they occurred. But these systems were reactive and had almost no visibility into the ecosystem in which the storage platform was running. The fact that these systems looked only at the storage often made it difficult to comprehensively understand what was causing an identified problem. Still this approach represented an improvement over what had come before.

The first cloud-based predictive analytics platform was introduced by storage start-up Nimble Storage in 2010. Called InfoSight, this platform was bundled at no additional charge with the vendor's hybrid flash arrays (HFAs) and also later with all-flash arrays. The idea behind a cloud-based predictive analytics platform transcended the original goals of other vendors' remote monitoring capabilities across several areas. Whereas the older remote monitoring systems either retained log data on the

array itself or fed collected data into a private database behind the storage vendor's firewall, InfoSight used the cloud to store data. This was an important distinction, as the cloud made it easy to share data securely among multiple constituents and to expand the size of the platform as the vendor's installed base grew. InfoSight also significantly increased the amount of data collected from a particular storage platform, not only monitoring features like inline data reduction that had not been in wide usage among primary storage platforms before but also collecting data much more granularly. This distinction became important as the vendor applied AI/ML algorithms not only to the data it collected from a single storage array but to all the data it collected across its entire installed base. With the wealth of data and the use of AI/ML, InfoSight was able to perform feats of which traditional remote monitoring systems were not capable.

Within a few years, several other vendors came to understand the value of cloud-based predictive analytics and introduced their own versions. The three key features that distinguish cloud-based predictive analytics platforms from traditional offerings are:

- **Cloud based.** Collecting and retaining data in the cloud makes it very easy to share data among a number of different constituents, including not only technical support but also engineering and product management within the storage vendor and ultimately with external constituents like other customers. The ability to use anonymized data (discussed later in this white paper) opens up significant opportunities to drive additional value for customers with the data collected by cloud-based predictive analytics platforms.
- **Scope of monitoring.** The rise of cloud-based predictive analytics paralleled the rise of flash in general-purpose enterprise storage platforms. With flash, there were a number of new areas that could be monitored (inline data reduction, flash endurance, etc.), but the original design of cloud-based predictive analytics assumed the ability to sample on a much more granular basis, collecting more data in any given time period and, ultimately, to move beyond storage to collect data from other components and resources within the IT infrastructure besides just storage.
- **Use of AI/ML.** In 2010, big data concepts were just starting to break into use. The availability of vast amounts of data and the ability to analyze it using much more powerful CPUs and AI/ML techniques promised to uncover new correlations, create more accurate trend analysis that looked not only at longer periods but also at any given time period in much more detail, and improve the reach of predictive failure analytics (an ounce of prevention is worth a pound of cure). The key here is the extent of reliable, autonomous operations that can be driven with AI/ML that improve the overall CX.

It is interesting that, despite the vast promise of cloud-based predictive analytics, not all enterprise storage vendors seem to have grasped its importance. While some vendors are driving considerable new value for their customers with these platforms, other vendors seem satisfied to merely claim that they have a cloud-based predictive analytics platform too. Customers that have already seen the best of what these types of platforms offer are actively using them as differentiators when considering enterprise storage purchases. What is important about cloud-based predictive analytics is not whether a vendor has such a platform, but how the vendor uses that platform to drive additional value for its customers. IDC has noted a significant difference among vendors in not only how much value they drive for customers with their cloud-based predictive analytics platforms but also how effectively they communicate that value.

Understanding the Value Driven by Cloud-Based Predictive Analytics

The traditional management model is very familiar to those managing enterprise storage. A system is configured to meet a certain set of performance, capacity, and availability requirements, and processes are implemented to meet data protection and recovery requirements. Recurring processes are automated as much as possible to reduce the opportunity for errors. As long as the system is meeting defined service-level agreements (SLAs), it is generally left alone. New software and hardware releases may not be installed unless they fix a known problem or include a badly needed feature because of the risk associated with change, and strict change management procedures (which basically discourage any kind of change without a very clearly defined objective) are maintained and rigidly enforced. This approach assumes that, in general, risk reduction has a higher priority than the opportunity for any improvements to performance, capacity utilization, availability, and overall efficiency, which may follow a change.

Simply put, cloud-based predictive analytics can potentially change all that, allowing administrators to continuously optimize systems across a variety of metrics as workloads evolve without incurring undue risk. There are already a number of examples of how cloud-based predictive analytics can drive value for both enterprise and cloud customers:

- Big data analytics can extend the window of predictive fault identification. While it is pretty much table stakes to be able to identify a looming failure even with the older remote monitoring systems, big data analytics opens up the opportunity to identify impending failures earlier, which provides more time for administrators to address the issue before it becomes an acute problem. It also provides the opportunity to manage other metrics, which impact a system's performance, capacity utilization, availability, and efficiency of resource utilization that don't fit the classic definition of a "failure" but can improve its operation in a number of ways.
- A more complete data set that includes the entire history of a system and a broader set of data around the ecosystem in which the storage resides can help speed problem resolution when a customer must contact technical support. An industry best practice is to put a customer directly through to a level 3 technical support resource that is tasked with resolving the issue, bypassing level 1 and level 2 resources (that are no longer involved because the cloud-based predictive analytics platform already has everything that would be collected at those two levels). Any real-time remote diagnostics that must be performed benefit from the much more extensive data collection capabilities of cloud-based predictive analytics, which can extend beyond just the storage to help identify and resolve more complex issues. This leads to faster problem resolution for the customer and lower support costs for the vendor (which can also potentially lead to lower maintenance fees for the customer).
- Using AI/ML-driven pre-upgrade validation, any suggested firmware, software, or hardware update can be confirmed safe up front by analyzing all available information from the vendor's entire installed base as well as any relevant information released by other vendors available on the internet. Using this more informed approach, a vendor's track record of success with new releases can generate trust on the part of customers to move to newer releases more quickly. When this can be done without risk, the customer benefits by having a system that offers higher performance, better capacity utilization, increased availability, improved efficiencies, additional functionalities, and/or lower cost. The vendor benefits because later releases are more reliable and may offer improved instrumentation to make problem resolution faster and more reliable than with older systems. There is a clear correlation noted by IDC among enterprise storage vendors that have a higher percentage of their installed base on the latest releases and the level of performance, capacity utilization, availability, efficiency,

functionality, and cost those customers enjoy. For those vendors that exhibit high percentages on this metric, the effective use of cloud-based predictive analytics is a key factor.

- The more comprehensive scope of cloud-based predictive analytics can include other components and resources in an ecosystem beyond just storage. The best of these systems today monitors not just storage hardware and software but also compute resources, virtual machines, network infrastructure, and applications. This data, and the correlations that may be identified through its analysis, can help resolve more complex problems that go beyond any single IT resource.
- Using anonymized data from their entire installed base, vendors can suggest recommended configurations by workload type, help customers compare their metrics (such as their data reduction ratios) to other customers running similarly configured systems and workloads, and better disseminate best practices (which lead to improved performance, availability, and efficiency). Today, customers have access to provisioning templates that are specific to particular workload profiles based on actual configurations in use, making it faster and easier to deploy capacity for new workloads. This can improve the productivity of administrators while it helps them manage their systems to higher levels of performance, availability, and efficiency.
- Through monitoring and data collection over time, a cloud-based predictive analytics platform can use AI/ML to profile what is considered "normal" behavior for a given platform and send alerts out when behavior considered "abnormal" is seen. As the installed base grows and more data is collected, these platforms not only become better at predicting and recognizing issues but also evolve over time with developing trends. These systems can also make recommendations for configuration changes to improve performance, capacity utilization, availability, and efficiency without any increase in cost, as well as recommendations for remedial actions (which could be performed either by an administrator or automatically by the system itself) to address any identified issues.
- The ability to query the entire installed base set of anonymized data allows customers to perform "what if" analyses that can precede (and help confirm the impacts of) any contemplated change. This ability to "see into the future" can help avoid problems and inform better performance, capacity, and availability planning as IT infrastructure is scaled to meet business growth.

Customers Favor AI/ML-Driven Predictive Analytics

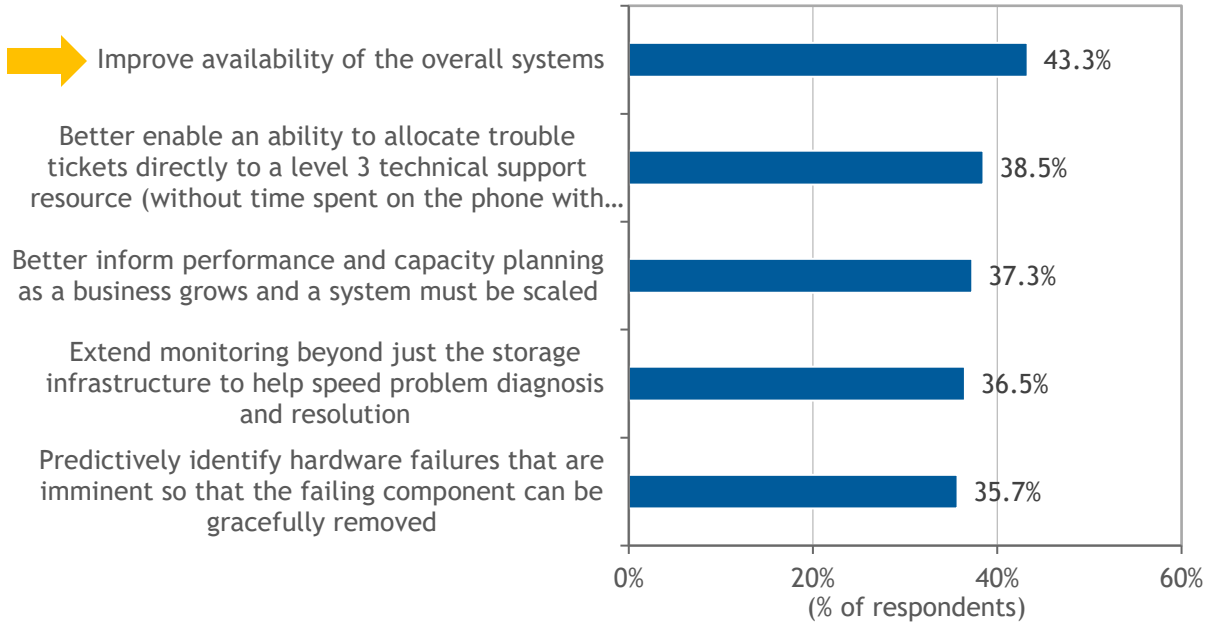
In June 2018, IDC completed a worldwide survey asking end-user customers about their experiences with and aspirations for cloud-based predictive analytics. Customers see the use of AI/ML algorithms to drive improvements to their infrastructure and processes in a very positive light. The use of AI/ML to drive increased customer value was desired by 85.7% of respondents. There was a high interest in increasingly autonomous operations (73.8%) to drive improved productivity, and 71.0% of respondents wanted to see more use of AI/ML to help drive autonomous operations. When asked about how they would like autonomous operations implemented, 20.2% of respondents wanted an AI/ML-driven recommendation engine to provide recommendations but leave the implementation to human administrators, 53.6% of respondents liked the idea of automated implementation of recommendations for certain types of operations but not others, and 26.2% of respondents were open to implementing more recommendations automatically as they gained trust with a vendor's ability to provide and reliably implement them.

60.7% of respondents see room for improvement in the efficiency of customer-vendor interactions, and the comfort level with vendors using internet-connected, real-time cloud-based predictive analytics to help drive these improvements was high (78.6%). There was strong support for the use of anonymized data to streamline array management and operational planning (86.1%), and 83.9% of respondents were comfortable with anonymized data collected from their systems to be used to help improve support and better inform engineering and product road map planning at the vendors themselves. Anonymized data removes any identifiers that point to the identity of a customer or system while retaining relevant metrics for analysis. Interest was also high in using anonymized data to help address installed base queries (e.g., best practices for Oracle RAC deployment) and show aggregate installed base statistics (data reduction ratios achieved in VDI environments, etc.). 64.7% of survey respondents were either interested or highly interested in leveraging AI/ML specifically to improve their CX with vendors.

Among those customers that were already using some form of cloud-based predictive analytics, IDC asked about the value they were receiving from these platforms. The top preferred use was to improve the overall availability of the system, although speeding problem resolution and aiding in performance and capacity planning also ranked high. Figure 1 shows the top 5 responses, which also included monitoring additional IT infrastructure components beyond just storage and predictively identifying any impending failures. The percentages in Figure 1 indicate the number of customers that identified that response as a "top priority."

FIGURE 1

Top 5 Preferred Uses of CBPA Driving Customer Value



n = 252

Base = all respondents

Notes:

This survey is managed by IDC's Quantitative Research Group.

Data is not weighted.

Multiple responses were allowed.

Use caution when interpreting small sample sizes.

Source: IDC's *Cloud-Based Predictive Analytics Survey,* June 2018

On the topic of "top preferred uses for a cloud-based predictive analytics platform," customers also identified a number of other areas shown in Table 1. Percentages shown in Table 1 indicate the number of customers that identified that response as a "top priority." These responses do not necessarily indicate that the customer is already enjoying that benefit, only that they would find interest in it (and may or may not already be doing it, depending on which vendors' cloud-based predictive analytics platform they are using).

TABLE 1

Additional Preferred Uses of CBPA Driving Customer Value

	(%) of Respondents
More accurately size systems up front (prior to initial deployment)	35.3
Leverage AI/ML to suggest configuration changes or other optimizations that would lead to efficiency improvements in resource utilization	32.9
Pre-validate proposed system upgrades for reliability with my specific configuration	31.0
Leverage AI/ML to suggest configuration changes or other optimizations that would lead to performance improvements at no additional cost	28.6
Provide default templates for provisioning and application deployment based on installed base analysis	28.6
Enable the rapid and more complete dissemination of best practices within the installed base	27.8
Help better inform vendor product strategy and planning	24.6

Source: IDC's *Cloud-Based Predictive Analytics Survey,* 2018

Finally, there was significant interest in vendors extending their cloud-based predictive analytics platforms to cover more IT infrastructure components – 69.9% of respondents were interested or very interested in this enhancement, which was correlated with better overall performance, availability, and efficiency as well as faster problem resolution.

HPE's Cloud-Based Predictive Analytics: InfoSight

Nimble Storage was the first vendor to introduce a cloud-based predictive analytics platform. It had more than four years of experience with the platform before the next external storage vendor (Pure Storage) entered the market with its version. Nimble Storage collected data from its installed base of HFA and AFA products, which were targeted primarily as general-purpose enterprise storage arrays for the midmarket. HPE finalized the acquisition of Nimble Storage in April 2017, and in IDC's opinion, the crown jewel of the acquisition was the InfoSight platform. HPE had not fielded a cloud-based predictive analytics platform for its enterprise storage prior to that, and with this acquisition, it had obtained what IDC views as industry-leading technology in that space. As part of the acquisition, HPE announced that it would be extending InfoSight coverage to other HPE storage product lines and, eventually, to other HPE IT infrastructure offerings (server, networking, etc.). In November 2017, HPE announced that InfoSight would be extended to cover the 3PAR product line and that it would be bundled at no additional charge with all 3PAR platforms (it was already included as part of a valid support contract with all Nimble platforms). In July 2018, HPE made good on that promise.

The InfoSight product was developed to handle big data analytics at scale. Today, the platform collects data from millions of sensors per second from tens of thousands of customers (the HPE Nimble Storage and 3PAR installed bases combined). AI/ML-driven algorithms correlate data within a single system and across systems to identify patterns, called predictive health signatures, that indicate potential trouble scenarios. InfoSight will then suggest possible remediations (or perform the fix itself if that has been allowed by the customer for that type of issue). The rest of the installed base can be scanned, based on identified predictive health signatures, to look for other systems at risk of having the same issue, getting ahead of any acute outages that may be a result. Patterns are also used to identify suboptimal configurations in terms of performance, capacity utilization, availability, and/or efficiency (relative to the other systems in the installed base running similar workloads), and AI/ML is used to recommend improvements, which can be implemented by either InfoSight or customers manually. InfoSight already looks beyond just the storage to improve operational metrics — 54% of identified problems to date have actually been outside storage. As autonomous operations become more widespread, customers recognize that the issues they do have left to resolve are becoming more complex, and they need additional instrumentation across more than just the storage infrastructure to troubleshoot and resolve them.

InfoSight includes a complete set of metrics that customers can view on a particular system, including latency, throughput, and bandwidth data that helps administrators understand how well the system is meeting defined SLAs; data reduction ratios by application, by group, and for the system as a whole; space savings associated with snapshot utilization, flash endurance, recovery point objectives achievable at the remote site in asynchronous replication configurations, and capacity consumption trends; and the identification of any failed hardware components (if and when that should occur). Other features include system-level health and wellness checks, per virtual machine–level statistics and heatmaps, and a utility called InfoSight Labs that allows users to experiment with their own data using reports created by InfoSight data scientists at HPE.

InfoSight already provides many of the features discussed previously when showcasing the promise of cloud-based predictive analytics to drive value for both customers and vendors. These include earlier and more comprehensive predictive analytics, a reduction in customer-involved support cases that get resolved on average more quickly than in the past, upgrade validations based on extensive installed base experience, access to certain installed base metrics that may be of interest to customers (e.g., the average data reduction ratio achieved on Oracle databases below 20TB in size), and the extension of data collection, analytics, and suggested remediation beyond just storage. Any new issue that arises is only likely to occur at most once in the entire installed base because as soon as it occurs, a predictive health signature is created, the installed base is analyzed to find other systems potentially at risk, and a resolution is suggested to potentially affected customers. With InfoSight, HPE is leveraging the power of big data analytics and an AI/ML-driven recommendation engine to relentlessly drive continuous improvements in a wide variety of visible metrics that drive an enhanced customer experience.

The proof is in the pudding for cloud-based predictive analytics platforms. With InfoSight, over 86% of issues are automatically predicted and resolved without requiring manual administrative intervention. With the assistance of InfoSight, covered storage arrays in production use have delivered in excess of "six-nines" availability over the past two years. The availability percentage has actually improved while the size of the covered installed base has increased over 900% over that same time period. There has also been a 19.3% decrease in customer-involved support cases over that same time period, despite the huge growth in the installed base. Trouble tickets get resolved faster in large part because when a customer logs a call to HPE, all the data that would normally be collected by level 1 and level 2

technical support personnel (system and ecosystem data, workloads, logs, telemetry, predictive health signature, and history) has already been collected by InfoSight, along with any AI/ML-suggested remediation, and is made available directly to the level 3 technical support engineer who will actually resolve the problem. Before its acquisition by HPE, Nimble Storage had in fact never employed level 1 or level 2 support engineers, and now this model is being extended to 3PAR as well through the use of InfoSight. This differentiates HPE from other vendors with its own cloud-based predictive analytics platforms that still employ level 1 and/or level 2 technical support tiers.

HPE uses Intel SSDs in both Nimble Storage and 3PAR StoreServ storage solutions. Intel SSDs enable faster performance and high efficiency at competitive cost, and HPE offers Intel SSDs at a number of different capacity points to provide configuration flexibility for HPE's customers. Intel also has a strong lead in emerging memory technologies, such as persistent and storage-class memory with its Optane technology and will be working closely with HPE to make those even higher-performance solid state storage products based on Optane available in well-integrated enterprise storage solutions. Intel components combine with InfoSight as well as other system components to drive the overall efficiency of HPE storage environments.

CHALLENGES/OPPORTUNITIES

The challenge with an offering like InfoSight is to generate more awareness among customers about the value that it can drive and to use that awareness to help create a meaningful differentiator for HPE. Customers need to understand that a cloud-based predictive analytics platform is not just a checkbox item on the list of purchase criteria, but there is a wide variability in the value that various vendors drive for their customers with these types of platforms, and customers should not let those vendors for whom it is little more than a glorified "remote monitoring" system determine their perception of what these systems can offer. Customers should look at not only what data is collected by cloud-based predictive analytics platforms but how it is used to drive higher performance, better capacity utilization (for increased storage density to drive lower energy and floor space consumption), increased availability (by avoiding problems before they occur and resolving them faster when they do), improved efficiencies, additional functionality, and lower costs. Those vendors that are the best at this should also be the ones that are most adept at communicating the value they deliver with these platforms — if they are not, they are missing a strong opportunity to differentiate themselves from their competition.

The opportunity with cloud-based predictive analytics is to help drive a meaningfully differentiating CX for customers. IDC has already identified several areas for near-term enhancement to these types of platforms that will enhance the value they provide to customers: profiling "normal" behavior for individual systems leading to the earlier detection of "abnormal" behavior and the ability to use anonymized data in queries by not only the vendor but also customers themselves (without vendor involvement). Vendors with good cloud-based predictive analytics platforms need to be creative in coming up with ways to drive value for their customers with the data they collect and the analyses they provide.

CONCLUSION

Cloud-based predictive analytics platforms are available from a number of enterprise storage providers, but how effectively vendors use those platforms to drive meaningful value for customers varies significantly. When considering a new enterprise storage purchase, prospective customers should challenge vendors to explain the value they are providing with these platforms, not just confirm that they have what they call a "cloud-based predictive analytics" platform. Customers should specifically be looking for what improvements these platforms drive for enterprise storage in performance, capacity utilization, availability, efficiency, and cost and what metrics the vendors can show that indicate the magnitude of improvements they provide.

IDC has noted a clear correlation between the length of time a vendor has offered a cloud-based predictive analytics platform and how much value they are driving for customers with it. Vendors with more experience drive more value for their customers — they have more data on which to draw, they are leveraging that data in more ways, and they have the right creative mind set to come up with new ways to drive value for customers with these platforms than vendors with less experience. Nimble Storage's InfoSight was the first platform of this type in the industry, and the company garnered years of experience using it before another external storage vendor even introduced one. When HPE acquired Nimble Storage in April 2017, this knowledge came with it, and with the resources HPE has at its disposal, it has a strong opportunity to further leverage InfoSight to drive an improved CX for its customers. With its stated strategy to extend InfoSight coverage to other HPE products, and the release of coverage for 3PAR in November 2017, HPE has made an excellent start.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com