IDC ANALYZE THE FUTURE

IDC TECHNOLOGY SPOTLIGHT

Sponsored by: Red Hat

Many strategic requirements and functional needs for enterprise storage today stem from the fact that the IT world is undergoing a digital transformation. Traditional proprietary, appliance-based storage offerings are no longer a viable long-term option for enterprises needing to modernize their datacenters.

Using Software-Defined Storage to Accelerate IT transformation

December 2018

Written by: Amita Potnis, Research Manager, Infrastructure & Software for Platforms & Technologies

Overview: DX, ITX, and the Road to Cloud

AT A GLANCE

KEY TAKEAWAY

Data growth and its specific requirements will drive end-users to adopt hybrid/multi-cloud strategy, data will reside across old and new infrastructure platforms making ease of deployment, data management and maintenance, and quality of service (QoS) top priority.

Digital transformation (DX) is reaching a macroeconomic scale. DX business objectives are balanced between tactical and strategic objectives and range from improvement in operational efficiencies and customer satisfaction, to increasing existing product revenue, to improving profit margins, to launching new digital revenue streams. IDC forecasts that by 2021, at least 50% of global GDP will be digitized, with growth in every industry driven by digitally enhanced offerings, operations, and relationships. Data is a new form of capital.

The accelerated pace of digitization and "datafication" of more value-creating business activities is leading to diverse, dynamic, and distributed data sets. There is proliferation of application deployment models and locations. To unlock the power of data capital and accelerate DX, businesses need modern data-centric IT services from edge to core to cloud. Organizations need to embrace and embark on an IT transformation (ITX) process and evolving journey. ITX initiatives are driven not just because of increasing datasets but also the variety of workloads and use cases this data growth has allowed that stems from the four pillars of the 3rd Platform: social, mobile, big data, and cloud. These four pillars and the supporting underlying technologies are transforming the way organizations strategize their infrastructure requirements for today and the future.

When planning ITX, it is imperative for IT organizations to identify gaps in business processes and profitability that stem from legacy infrastructure, and replace or augment it with efficient, reliable, modern, and future-proofed infrastructure. This also means that the way organizations traditionally procure and consume IT infrastructure will no longer fulfill the needs of today's digital world.

In recent years, the market has seen a steady shift from traditional proprietary storage infrastructure to a technology strategy that includes adoption of hybrid cloud/multi cloud storage environments. Data growth and its specific requirements will drive end-users to adopt a hybrid/multicloud strategy. Data will reside across old and new infrastructure platforms making ease of deployment, data management and maintenance, and quality of service (QoS) top priority. In addition, organizations are actively looking at augmenting or replacing their existing infrastructure with

software-defined storage that bears promise of increased flexibility, reduced costs, ease of operations and, more importantly, being a future-proof technology that is capable of easily catering to changing infrastructure demands.

As organizations strategize their IT transformation (ITX) plans, they should actively think about why they need new infrastructure and key criteria:

New infrastructure for:

- Supporting net new green field as well as current deployments by augmenting existing infrastructure
- Right-sizing applications by adopting private/public cloud, IaaS, and PaaS in addition to on/off-premises traditional storage
- Realizing long-term TCO savings and ROI (reduced management overheads, ease of scalability, etc.)
- Driving increased cost efficiency by reducing infrastructure silos and spend on proprietary solutions
- Establishing parity between on-premises storage infrastructure, private and public cloud for ease of management and usability

New Infrastructure should be:

- Scalable in terms of performance and capacity with the ability to load balance
- Flexible in terms of platform/solutions' ability to offer end-users the flexibility and agility to deploy traditional and next-gen workloads
- Leverage existing IT investment and strategically make new investments (including private & public cloud)
- Extensible with ease across on-premises and hybrid/multi-cloud (private/public cloud) environments along with new technologies like containers

IT Transformation with Hybrid and Multicloud

The notion that organizations are incorporating hybrid cloud/multi cloud infrastructure strategy in their IT roadmap is a reality indeed. Several IDC studies have indicated that "Cloud First" is emerging as a standard strategy for enterprise IT. For example, in 2017 IDC's Cloud Study – a survey of 1,000 respondents worldwide – indicate the following when it comes to main infrastructure strategies:

- 72% of the respondents are actively investing in IT to support digital transformation and growth
- 60% of the respondents use 3rd-party/outsourcing cloud firms to supplement in-house resources wherever possible
- 59% of respondents are proactive, broadly implementing a private or public cloud-first approach to new app deployments

The research findings are further validated in a recent survey called IDC's Buyer Behavior and Other Technology Adoption Trends, 2018. When asked why organizations are retiring their storage infrastructure, respondents indicate the following as top reasons:

- Introduction to private cloud
- Availability of next-generation (storage) systems
- Migration to public cloud

The cloud-first philosophy will drive the adoption and usage of public cloud IaaS, particularly for enterprises that do not have the scale of cloud usage to initially consider a private cloud option and for organizations that wish to evaluate new digital technology services like artificial intelligence. Many customers are comparing private cloud solutions and cost against public cloud options. Public clouds are often chosen by business managers who want to address a specific use case, or by IT departments to run applications and workloads that aren't critical to operations. But public clouds are gaining traction beyond just running non-critical workloads, especially with the addition of automation, unified management, and security that gives businesses the ability to scale as needed, manage complex deployments, and speed productivity.

Private Cloud services are shared within a single enterprise or an extended enterprise, with restrictions on access and level of resource dedication, and defined/controlled by the enterprise beyond the control available in public cloud offerings. Enterprises will incorporate private cloud in their IT infrastructure strategy when they need to keep data on-premises but want cloud agility, ease of use, and cost advantages. The common use of private cloud today is modernizing or replacing existing datacenters and to reduce capital and operational expenses of existing workloads.

Given the diverse needs for infrastructure across various workloads within any organization, the shift is now leading towards a hybrid cloud (mix of private and public cloud with orchestration between platforms) and multi-cloud (two or more public or managed cloud computing services) infrastructure approach. It is important that the underpinning storage technology is flexible and scalable while lending itself to storage environments across all deployment locations (on-premises, private and public cloud). One such technology that offers the key benefits of scalability, affordability and flexibility is software-defined storage. Let us see how software-defined storage can support ITX and hybrid cloud strategies.

SDS: The Backbone of IT Transformation (ITX) and Hybrid Cloud Strategies

IDC defines Software-Defined Storage (SDS) as a system (hardware plus software) that delivers a full suite of persistent storage services via an autonomous software stack that can run on any industry-standard (rather than proprietary or custom) hardware platform (known as server-based storage). These industry-standard platforms must be commodity off-the-shelf (COTS) products, including (but not limited to) x86-based servers running Windows, Linux, or another off-the-shelf operating system. SDS solutions are available from start-ups, incumbents, and a vibrant open source community.

SDS solutions are especially beneficial to serve the needs of the 3rd Platform because these solutions are:

- **Extensible**. SDS solutions can be deployed on-premises, private and/or public cloud.
- **Multipurpose**. Can support dedicated block, file, or object platforms or some combination of block, file, and object. On the other hand, SDS solutions can serve as hyperconverged or dedicated storage based on the requirements of the user.
- **Flexible.** Can be procured as appliance or "software only"
- **Cost-effective**. Are built around scale-out architectures using commodity off-the-shelf hardware

SDS solutions that support block, object, and file interfaces that are full featured, scale seamlessly, and support high performance across deployment locations (traditional, private/public clouds) are integral to modern successful deployments. SDS's flexibility, automated services-rich framework, and unified data access help remove the physical boundaries to data repositories. These solutions truly help consolidate workloads and eliminate silos of storage. Such standardization improves cross-cloud manageability and improves their attractiveness as a platform for new service development.

Most cloud infrastructure uses SDS, thus accelerating the migration of workloads to the cloud by supporting the use of technologies such as containers, serverless computing and automation tools. Many SDS suppliers have a public cloud

option for customers adopting the "lift & shift" approach by supporting similar instances in the public cloud. For customers that want to retain their existing storage infrastructure but plan to integrate with the public cloud, SDS suppliers offer compatible APIs (e.g., S3) to seamlessly extend their environments across deployment location. In terms of private cloud, OpenStack is a widely adopted orchestration layer. SDS platforms that are full featured, scale seamlessly, and support high performance are integral to successful OpenStack private cloud deployments.

Organizations must ensure that the storage solutions they procure now will support their current and future objectives in a cost-efficient manner. IDC's research indicates that data analytics is a fast-developing use case that will be deployed on private or public cloud. IDC defines Big Data by three key attributes, volume, variety and velocity" and by the workload parameters of specific use cases rather than any individual technology. One of the key scenarios IDC's Big Data & analytics taxonomy, 2017 indicates that the critical qualifier for choice of infrastructure for Big Data is that it must be dynamically adaptable.

Today, most of the data resides in siloed, proprietary storage appliances which does not support the idea of Big Data that enable businesses to implement an organization-wide information fabric to solve many of urgent digital business problems through analytics. Hadoop and Big Data are synonymous with more than half of what Fortune 500 companies use for analytics. As Big Data deployments become critical to business operations, the top requirements from the data stores for Hadoop deployments are scale of performance to support both batch and real-time analytics workloads, resiliency at petabyte-scale deployments, easy and simple adherence to compliance and regulatory requirements, and lower cost of storing and managing huge volumes of data. Scale-out storage architectures dominate in the Big Data space based on the requirement for easy, non-disruptive scalability into the tens of petabytes (PBs) and beyond over time.

As consolidation of Big Data and analytics continues, suppliers will look for newer logical abstraction mechanisms like software-defined storage, compute, and networking to make the Big Data infrastructure as operationally efficient as possible. Continued exponential growth of storage capacity owing to Big Data will likely drive migration of more data to the public cloud or on premise private cloud for analytics. Hybrid cloud needs true application and data mobility so that applications and data can be accessed where needed.

The storage infrastructure behind hybrid cloud must support unified data access, elasticity, and consistent performance. These infrastructures need to incorporate data workflow mobility not only within the datacenter, but also across datacenters, leveraging public or private clouds, and between on-premises and in-the-cloud storage targets. Traditional storage approaches were not intended for and thus do not fit in this new paradigm. SDS abstracts data from infrastructure and enables applications to run the same way, whether they are on-premises or in the cloud. It provides a flexible, automated, and services-rich framework to achieve agility and unified data access and helps remove the physical boundaries to data repositories. SDS is the stepping-stone to hybrid cloud.

Newer technologies such as containers allow for quick, mass deployment of isolated, modular applications across a hybrid and multi-cloud environment. Because of the lightweight nature of containers, a single server can generally host many more containers than virtual hosts. Today, early enterprise adopters are using containers for lift and shift of existing applications. Functionality continues to improve in this area, which widens the workloads that containers will be able to address in the future. Container-optimized storage offerings are software defined and container aware, run on industry-standard hardware, support elastic scalability, and have a rich set of data management capabilities, especially needed for cloud-native applications. Therefore, similar to private cloud, containers are integral to enterprise cloud strategies and go hand in hand with SDS deployments.

While there are several SDS suppliers in the market today, there are few that offer standalone products as well as a platform (portfolio of products) for specific infrastructure needs like hybrid cloud or multi-cloud deployment. Red Hat is one such vendor that offers SDS solutions as well as a surrounding portfolio to support varied infrastructure demands for several use cases including data lakes or Big Data analytics.

Considering Red Hat Ceph Storage: A Strategic SDS Offering for Hybrid/Multicloud

Red Hat is an experienced provider of open source software products and services to the enterprise community. The company's product portfolio includes Red Hat Enterprise Linux, Red Hat OpenShift Container Platform, Red Hat Storage (both Ceph and Gluster), Red Hat OpenShift Container Storage, Red Hat Hyperconverged Infrastructure (both for Virtualization and for Cloud), Red Hat Ansible Automation, and several other products that combined offer an infrastructure platform for its end users. Recently, Red Hat announced the Red Hat data analytics infrastructure solution that establishes a private infrastructure to complement big data analytics tools upon which data platforms can be built the same way as they are built in the public cloud. Users can target analytic workloads between public and private cloud infrastructure with a common S3 interface and reap more timely insight with less competition for resources.

Red Hat acquired Inktank in 2014 and later released Red Hat Ceph Storage (RHCS) as a part of its storage portfolio. Red Hat claims that its storage offerings, including RHCS, are based on foundation (community-driven innovation and strength of Red Hat platforms), focus (strong engineering centered around specific use cases), and flexibility (application portability, rolling upgrades, and no hardware vendor lock-in). RHCS is a software-defined storage platform that runs on industry standard x86 hardware and supports block-, file-, and object-based data organization capabilities. RHCS supports native file system capabilities through CephFS. RHCS' object-based layout lends itself to being a scalable unified storage platform where resources are provisioned either in a standalone fashion or via OpenStack. The solution supports current-generation applications as well as next-generation applications, deployed on bare metal or virtual machines or inside containers. RHCS enables multisite clusters for data distribution and disaster recovery.

RCHS supports a wide variety of use cases such as archival/backup, infrastructure as a service with OpenStack, and data lakes supporting data analytics.

The company’s greatest focus today for Ceph is on cloud infrastructure and emerging data-intensive workloads around analytics and artificial intelligence. As a result, Red Hat Ceph Storage serves as a primary solution underpinning many Red Hat offerings. For example, it’s a critical component of the Red Hat data analytics infrastructure solution that enables a cost-effective architecture for structured and unstructured data analytics that decouples compute from storage for establishing shared datasets in an object store. The solution offers on-demand provisioning of analytic tools and clusters activated through Red Hat OpenStack Platform and Red Hat OpenShift Container Platform. Red Hat Ceph Storage’s S3A compatibility and tightly integrated capability to provision Red Hat OpenStack Platform, including within Red Hat Hyperconverged infrastructure for Cloud, add to its popularity for multi-platform environments.

The product’s ability to scale easily to hundreds of petabytes, support shared datasets (as opposed to duplicated), access predictive analytics tools (based on Red Hat Insights), and achieve high cost efficiencies make it an ideal candidate for not just on-premises infrastructure but also for hybrid cloud environments.

Advantages of Red Hat Ceph Storage:

- Core capabilities, including.
    - Unified storage capabilities to support block, object, and file storage under one roof

  - Seamless integration with OpenStack's modular architecture and key components
  - Massive scalability to manage petabytes of data on COTS hardware.
  - Flexible configuration to adjust storage needs as applications and deployment change
  - Rapid provisioning to deploy and integrate new virtual machines and nodes
  - No single point of failure, for maximum uptime
  - Self-management, the ability to self-heal, and fail in-place maintenance
  - Dynamic data placement and proportional data movement, for effective load balancing
- Advanced Functionality, including but not limited to striped erasure coding, Containerized Storage Daemons, and predictive analytics and insights
- Ecosystem. Within Red Hat and with external OEMs and ISVs as well as Ceph Foundation (31 inaugural members)
- Maturity. Ceph's history dates back more than 14 years. It began as a part of a research grant from the Department of Energy in cooperation with the Los Alamos and Lawrence Livermore National Laboratories in 2004. The open source code matured into a generally available production-ready solution in September 2012 following the launch of Inktank, a company focused on commercializing Ceph. Red Hat acquired Inktank in February 2014, and the product has been selling to global accounts under the name Red Hat Ceph Storage ever since.
- Integration and roadmap. RHCS is designed, tested, and incorporated as an integral part of Red Hat's larger infrastructure platform story in support of customers' pursuit of I.T. transformation/optimization and the adoption of hybrid/multi-cloud. Because siloed storage deployments are waning, this allows a broader set of initiatives to be addressed in tandem.

Strengths & Challenges

RHCS's strength comes from its deep integration with the broader Red Hat portfolio and supporting the company's vision to become a platforms company. In addition, Red Hat's strength as a company also lies in its engineering experience while commercializing several open source products and platforms such as OpenStack into which RHCS can be integrated. IDC believes that Red Hat's investment in RCHS will serve customers well as they look for an offering that can cater to traditional and next generation workloads in a multi/hybrid cloud environment.

RCHS's challenge stems from historical perception of performance and reliability issues related to Ceph. The Ceph community and RCHS have come a long way addressing these concerns and IDC believes that in time RCHS will be able to shed the old perception and position itself for increased adoption for many use cases and workloads. Customers should keep in mind that they must choose the recommended enterprise grade hardware for RCHS or any other SDS deployment to avoid unforeseen challenges and setbacks.

Conclusion and Recommendations

Enterprises continue to struggle with the task of managing data growth. Data is a critical corporate asset, and the value of data is tied to how it is ultimately used. Businesses in today's digital world are data centric (data becoming IP) and data driven (data used to make business decisions). The level of maturity that they have reached with their infrastructure strategy to accommodate growing amounts of data has to be reset when they shift to a hybrid cloud/multicloud paradigm. It's no longer just about managing data in a single location but also about managing it across multiple locations, applications, and stacks, and this presents an entirely new set of challenges.

In this era, infrastructure platforms cannot be designed to be standalone in nature. They must have a way to connect to the cloud. Nowhere is this more prominent than in the case of storage suppliers that are aggressively making a shift to deliver solutions that shift the conversation to end-to-end hybrid cloud data management, away from procuring yet another infrastructure product. Red Hat's vision for Red Hat Ceph Storage and its product capabilities are in line with the reality of the digitized world undergoing DX. Red Hat is enabling its customers to adopt a hybrid cloud strategy with ease, reliability, and cost effectiveness that is often desired but hard to deliver for vendors.

MESSAGE FROM THE SPONSOR

Red Hat Storage and Analytics Solutions

Visit https://www.redhat.com/en/technologies/storage/ceph for more information on Red Hat Ceph Storage. For more information on Red Hat Data Analytics Infrastructure, please visit https://www.redhat.com/en/resources/data-analytics-infrastructure-solution-detail.

About the analyst:

Amita Potnis, Research Manager

Amita Potnis is a Research Manager within IDC's Infrastructure & Software for Platforms & Technologies group. In this position, Ms. Potnis is responsible for conducting research in two key market segments: Infrastructure for Content & Cloud Native Applications and Infrastructure: Cognitive/Artificial Intelligence, Big Data & Analytics Workloads.

IDC Custom Solutions

IDC Corporate USA

5 Speen Street
Framingham, MA 01701, USA

T 508.872.8200

F 508.935.4015

Twitter @IDC

idc-insights-community.com

www.idc.com