STORAGE FOR CONTAINER DEPLOYMENTS

CONTAINERS IN THE ENTERPRISE

Containers provide a nearly unparalleled ability to efficiently deploy and manage application workloads and give enterprises the tools required to quickly, safely and easily deploy workloads across multi-site, multi-cloud infrastructure. Built on a foundation of simplicity, isolation and efficient resource sharing, they have become an indispensable tool for IT administrators and DevOps practitioners.

Enterprises clearly understand the benefits that containers can bring. One recent Red Hat survey shows an expected increase of container adoption of nearly 89% by 2021.[1] Containers have become a fact of life in the enterprise. Enterprises are using container technology to deploy cloud-native applications, database workloads, stateless applications and Continuous Integration/Continuous Delivery (CI/CD) pipelines. Container technology is even becoming the preferred deployment model for deep learning workloads.

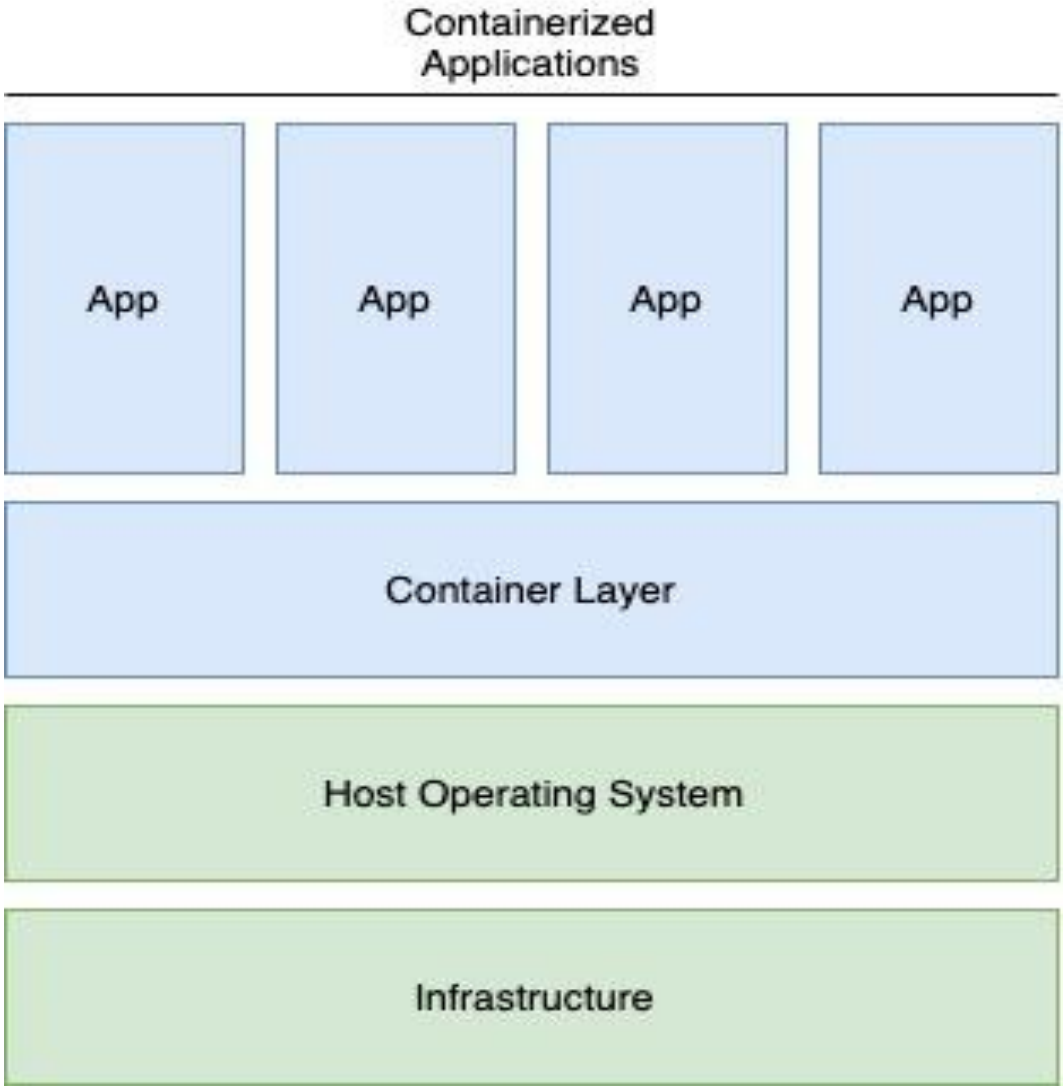
While containers bring simplicity to workload deployment, challenges remain in managing the resources outside the boundaries of the container. Traditional IT deployments, where applications and data are relatively static, are able to provision and tune storage based on the needs of the workload. Containers and hybrid-cloud environments, with flexible "run anywhere/migrate anywhere" models, change that dynamic. A shift in the nature of how data is stored, provisioned and served occurs.

CONTAINER ARCHITECTURE

Containers provide a lightweight, sand-boxed environment within which an application can execute. They leverage the host operating system to provide services to application while, at the same time, isolating the runtime environment for other applications or containers that might be running on the same host. This allows for easy configuration and flexible deployment for a container-based application. This is illustrated in the following figure:

[1] Red Hat Global Customer Tech Outlook 2019: https://www.redhat.com/en/blog/red-hat-global-customer-tech-outlook-2019-automation-cloud-security-lead-funding-priorities?source=bloglisting

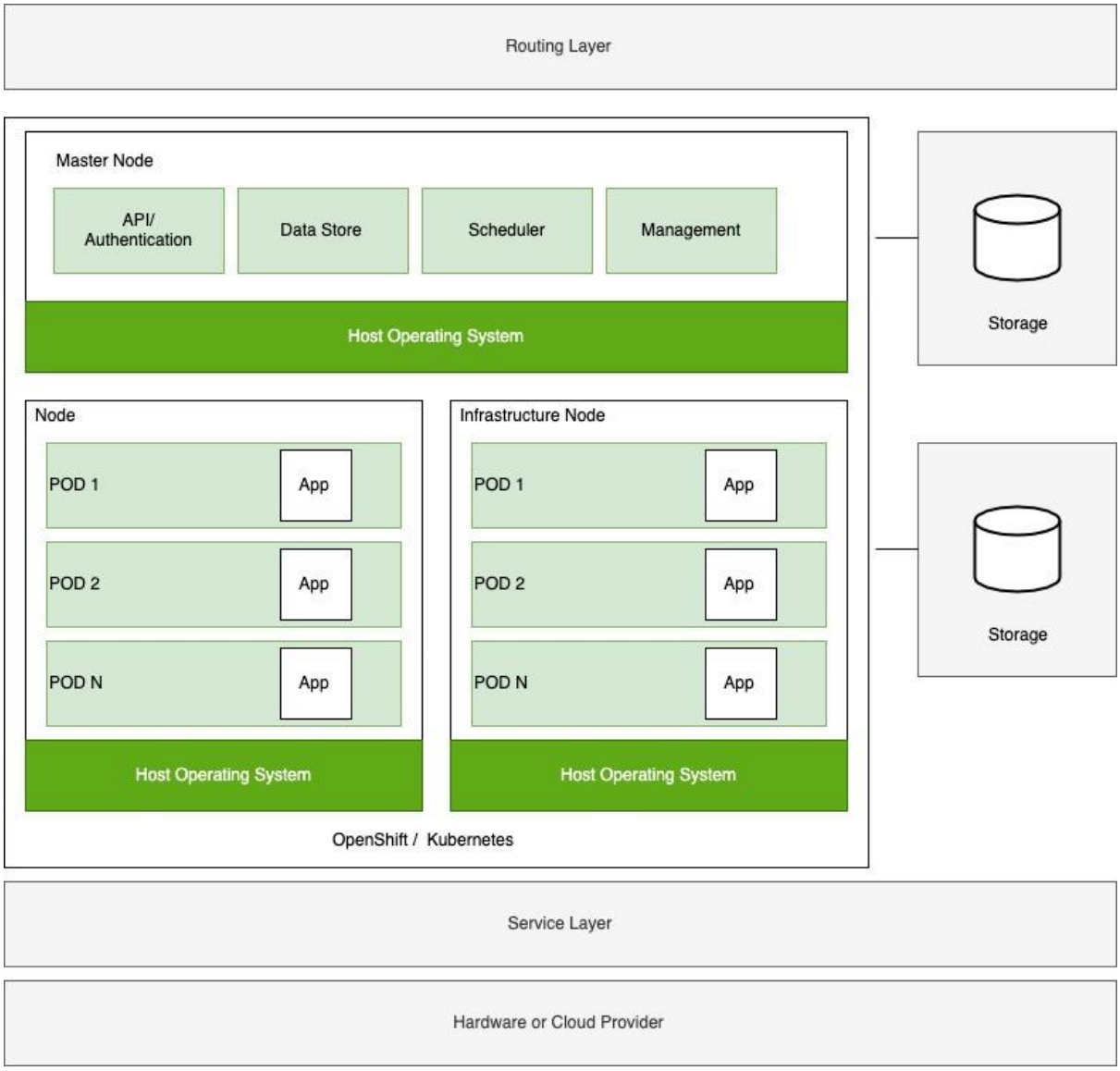
FIGURE 1: CONTAINER CONFIGURATION DIAGRAM



Source: Moor Insights & Strategy

Once coupled with orchestration software, the power of containers is fully realized. Orchestration software, such as Kubernetes, manages the configuration, deployment, scalability and availability of containers. Kubernetes is the dominant orchestration software for containers, with Red Hat's OpenShift providing an enterprise-grade implementation of Kubernetes. The illustration below models the OpenShift Kubernetes architecture:

FIGURE 2: OPENSHIFT KUBERNETES ARCHITECTURE



Source: Moor Insights & Strategy

At a high level, Kubernetes extends the container model by providing the mechanism to schedule and deploy container workloads across multiple server instances called *nodes*. Within each node, container-based applications are deployed in *pods*.

Availability and scalability are achieved by duplicating pods across nodes, with Kubernetes managing the state and access to each containerized workload. These

nodes may exist behind a load-balancer, with requests evenly distributed, or may exist as redundant copies of the applications acting as standby nodes for high availability.

Kubernetes also provides the connection of storage to the various elements within the system but does not intrinsically manage that storage for the user. Technical details of how all of this works can be found at Red Hat's OpenShift website.[2]

STORAGE & CONTAINERS

The idea that workloads are ephemeral, being easily replicated and migrated to provide high levels of flexibility and reliability, stands as a core tenet of container-based workloads. At the same time, the lightweight and throw-away nature of containers challenges an enterprise's responsibility to protect and preserve the data maintained and consumed by the container-based workloads. Most cloud native applications use object storage for persistent storage of unstructured data. Using Representational State Transfer (REST) application program interfaces (APIs) over HyperText Transfer Protocol Secure (https) provides application deployment flexibility but may not be suitable for all applications.

Deploying containers in the enterprise requires thinking about storage architecture as carefully as thinking about the workloads running within the containers. As containers enter the enterprises, it is crucial to consider container data as enterprise data, with the same considerations as all other enterprise data:

- Multi-Dimensional Performance.
- Data Protection, Compliance and Security.
- High Availability.
- Scalability.
- Performance and Service Level Guarantees.

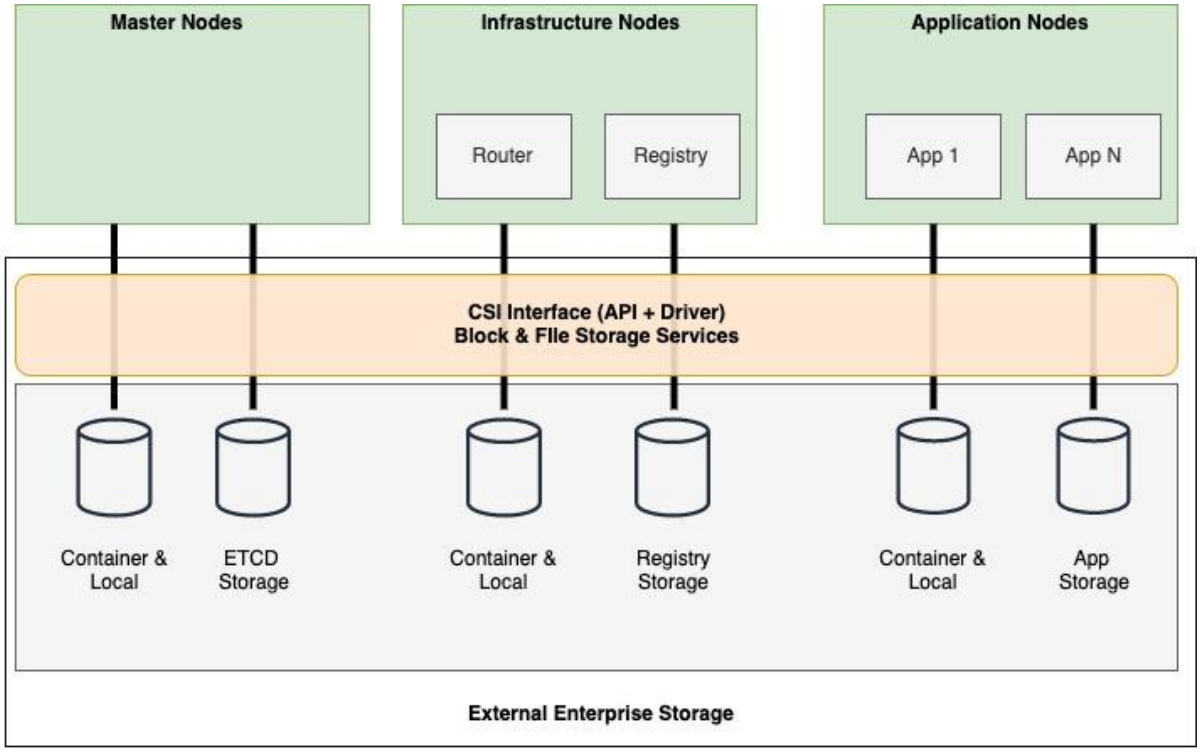
The use of shared storage systems provides the most effective way to deliver enterprise storage services to containerized applications.

HOW CONTAINERS USE PERSISTENCE

Kubernetes does not force a storage model on the user, but it does have a number of persistent storage needs. The architectural diagram, below, nicely illustrates this:

[2] https://www.openshift.com

FIGURE 3: KUBERNETES STORAGE DIAGRAM



Source: Moor Insights & Strategy

As you can see from the reference architecture, we have several buckets of storage. Data related to the operation of the cluster itself, image repositories for deployments into each node, data maintained by each node and, of course, data used by the applications running within each pod also exist in the architecture.

While Red Hat OpenShift and Kubernetes do not force a storage model on the user, they do lend themselves to deploying on enterprise storage systems. This allows for a wide range of data management services while also providing the utmost in flexibility to deliver data in a manner best required by the containerized workloads.

To allow flexible storage options for container users, the container development community has defined an open storage API called Container Storage Interface, or CSI. This replaces the original Persistent Data Volume drivers, which were proprietary and did not provide higher level data functions. CSI allows container orchestration systems, such as Kubernetes and Red Hat OpenShift, to seamlessly connect with block and file storage and will provide a standard interface for data management across vendors.

CSI is an evolving open source standard that defines a standard API[3] to allow storage providers to integrate block and file storage systems seamlessly into a Kubernetes deployment without modifying the core Kubernetes code. CSI gives Kubernetes users a myriad of options in deploying storage in a secure and flexible way, allowing users to:

- Automatically create and delete storage volumes as required.
- Provision and assign storage to containers as those containers are scheduled.
- Integrate with enterprise-level storage software, such as IBM's Spectrum Protect Suite, to manage snapshots and to provide data protection services.

Users of containers don't interact directly with CSI. The interfaces exist to allow storage providers such as IBM to seamlessly provide enterprise data services to container deployment. The important point is, as you look at architecting storage for containers, you need to consider the breadth and depth of CSI support in the storage systems you plan to deploy to support those containers.

Multi-Dimensional Performance

Containers are simply platforms in which applications execute. As such, they don't force performance characteristics on the underlying storage. It is the applications running within these containers that drive the requirements.

Containerized workloads range from the very lightweight, such as Lambda and other serverless functions, to more stateful applications such as machine learning, database, key-value caches and web applications.

Red Hat OpenShift provides an infrastructure on which to deploy these workloads, but each hosted application will have different service-level requirements for the underlying storage. These needs will range from the very high-throughput, low-latency needed by, for example, a machine learning container, to the low-latency, low-access characteristics of serverless functions.

One cannot predict how a container infrastructure will be used over time. The core value proposition of a container architecture is that it provides a flexible environment within which to dynamically deploy applications.

This provision of a flexible environment within which to dynamically deploy applications demands that the underlying storage be equipped to support the varying requirements

[3] https://github.com/container-storage-interface/spec/blob/master/spec.md

of the containerized workloads. Node-local storage, and even server-based shared storage, is often insufficient to meet the needs of an enterprise container infrastructure over time.

Data Protection

No one disputes that enterprise data must be protected. Whether from the threat of ransomware, in order to comply with local regulations or to safeguard from the still-too-common user error, rapid recovery of data has become a critical service of every IT organization. This extends equally to workloads running within a containerized infrastructure.

Data within a containerized cluster has varying requirements for data protection. The working-level persistent memory for each node, for example, may be short-lived and not subject to traditional data protection needs. Cluster configuration data, on the other hand, is long-lived and business critical. Thus, the full arsenal of enterprise data-protection strategies should protect cluster configuration data. Likewise, stateful workloads will require various levels of data protection.

Marrying a Red Hat OpenShift implementation with a storage system that supports the full range of data protection, such as that offered by the IBM Spectrum Protect Suite, provides the highest level of flexibility. IBM Spectrum Protect Suite can provide the right range of services for nearly any containerized workload, providing:

- Data lifecycle automation.
- Built-in cloud integration.
- Data replication.
- Comprehensive data protection.
- CSI snapshot support, which unifies container orchestration with storage.
- Pre-defined policies to help ensure Service Level Agreement (SLA) compliance.
- The ability to store container data in non-Kubernetes repositories to support cyber-resiliency.

Container Storage in a Hybrid-Cloud Environment

Enterprises embrace hybrid-cloud and multi-cloud architectures to ease deployment and deliver intelligent workload placement. These architectures include private clouds, public clouds and even infrastructure-as-a-service.

Coupling the intrinsic capabilities of containers to provide lightweight and portable environments for workloads, container storage in a hybrid-cloud environment becomes a natural element of any organization's hybrid-cloud strategy. Enterprises in all these environments deploy containerized applications.

Using tools such as Red Hat's OpenShift and Kubernetes allows for containers to be placed and managed anywhere within an enterprise infrastructure regardless of physical location but requires a storage strategy built with the above requirements in mind.

Shared storage solutions designed for hybrid-cloud deployments becomes a critical element in any enterprise container implementation.

Developing a Container Storage Strategy

You can't take storage for containers for granted. Deploying containers in the enterprise requires careful thought and planning. As you think about how to best deploy these technologies into production environments, the importance of partnering with technology providers who are expert in the intricacies of deploying both containers and storage in complex hybrid-cloud environments becomes apparent.

Developing a storage strategy for your organization's container infrastructure requires:

- Education in understanding the complexities of deploying an OpenShift implementation.
- Mapping the service-level requirements of the containerized workloads.
- Understanding how the container infrastructure intersects your enterprise's hybrid cloud strategy.
- Working with storage technologies that fully embrace CSI interfaces for storage.

IBM is an ideal example of a strong partner in architecting and deploying container solutions. IBM's storage products fully embrace the CSI interface to provide benefits and functionality that exceed the baseline goals of seamless integration.

IBM's integration with RedHat OpenShift and Kubernetes provide the following benefits:

- **Performance**: Rapid deployment and seamless integration with IBM FlashSystem that delivers 50% better IOPs in a fraction of the rack space and 2.5times higher throughput for DevOps and database workloads.
- **Security**: IBM's container security architecture reduces the cyber-security vulnerabilities by enabling selective storage volume mapping to containerized

applications. Beyond this, IBM's Storage Systems support FIPS 140-2 encryption and enterprise key management.

- **Productivity**: IBM's CSI drivers improve time-to-deployment by enabling and exposing dynamic storage provisioning to services classes, along with tools to enable automated, policy-based delegation.
- **Agility**: IBM's CSI drivers provide consistent interfaces to block and file storage, along with integration with data protection to allow workload and infrastructure flexibility.

IBM has also made it easy for IT organizations to quickly deploy solutions with the IBM Storage for Red Hat OpenShift Blueprint, a pre-tested and validated blueprint documenting the delivery and implementation of storage for Red Hat OpenShift into a hybrid cloud environment. It includes:

- IBM Spectrum Virtualize-based IBM FlashSystem non-volatile memory express (NVMe) arrays for primary storage.
- IBM Spectrum Scale and Elastic Storage System for Artificial Intelligence and big data workloads.
- IBM Cloud Object Storage for cloud-native applications.
- IBM Spectrum Protect Suite for data protection and re-use.

These elements combine to provide the utmost in security, service orchestration, infrastructure agility, performance, availability and data protection, all critical attributes of every storage architecture, including enterprise container architecture.

CONCLUDING THOUGHTS

Containers are a part of nearly every enterprise architecture. The data consumed by and supporting implementations of containers must be treated as enterprise data.

To successfully treat container data as enterprise data one must implement strategies for data protection, support multi-cloud and hybrid-cloud architectures and implement a storage system that can deliver scalable and multi-dimensional performance. Enterprise shared storage is the natural approach to providing data services to container deployments, while deep integration with CSI interfaces allows the storage to be treated as enterprise-class data.

Working with a vendor such as IBM, who has a strong portfolio and long legacy of delivering these capabilities into the enterprise, is a natural fit.

IMPORTANT INFORMATION ABOUT THIS PAPER

CONTRIBUTOR

Steve McDowell, Senior Analyst at Moor Insights & Strategy

PUBLISHER

Patrick Moorhead, Founder, President, & Principal Analyst at Moor Insights & Strategy

INQUIRIES

Contact us if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts but must be cited in-context, displaying author's name, author's title, and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

LICENSING

DISCLOSURES

This paper was commissioned by IBM Corporation. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

DISCLAIMER

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.