NetApp Distributed Content Repository Solution

Managing exploding data growth effectively in a collaborative world

Key Benefits

Manage Billions of Objects Across Multiple Sites

Store up to petabytes of data in a multisite global object namespace, and use metadata to intelligently manage storage.

Provide Global, Always-On Data Availability

Distribute content across multiple geographically dispersed sites for data protection, disaster recovery, and low-latency collaborative workflows.

Reduce Capex and Opex

Implement policy-driven storage management to align the cost of your storage with the value of your datasets. Modular, dense storage makes long-term retention affordable.

The Growth of Distributed Content

Enterprises across industries are finding that their content is growing at an unprecedented rate. Petabyte-sized content repositories are no longer a rarity as healthcare, media, financial, and other large institutions store digital objects that grow in size, quantity, and lifespan. More than ever, this data must be accessed by various constituencies in a variety of locations, and often in a fast-paced, collaborative manner. Enterprises depend on accessing a wide variety of data objects in order to run their mission-critical workflows. Managing this data intelligently across geographically dispersed locations cost effectively is a huge challenge.

The explosion of data comes from human as well as machine sources. While e-mail, home directories, and scanned documents are a huge area of data growth, there is also a vast quantity of machine-generated content from medical imaging systems, telecommunications and utility equipment, and financial transactions.

Even if your organization has less than a petabyte of unstructured content today, industry growth rates are so high that you could have over a petabyte in only five years. That means you have to think about implementing the right solution today in order to have a viable infrastructure for the long term.

Addressing the Challenges of Distributed Content Repositories

Distributed content repositories must provide the ability to store terabytes to petabytes of data for long periods of time, while enabling key management functionality required by IT departments as well as desired end-user functionality. Distributed content repository challenges include:

Intelligent management

To effectively manage very large distributed repositories, IT teams need a storage infrastructure that intrinsically manages content from multiple sources, distributed across different geographies. They must serve the needs of many different user groups with unique access and reliability requirements.

Global access and effective retrieval

Managing billions of objects created over periods of years or decades is complex enough, but finding the data you need when you need it becomes a daunting task without effective tools. Managing content from multiple sources across different geographies adds the challenges of content distribution, global access control, and end-user performance.

Reliable data protection

Providing global 24/7 data availability is a unique challenge when dealing with petabytes of data.

■ NetApp®

Cost-effective infrastructure
When dealing with multipetabyte repositories, the issues of cost are paramount. In some cases, data center space is a constraint as well.

Compliant retention
Corporate governance, e-discovery policies, and business processes all drive retention requirements, requiring an extremely robust and flexible infra- structure that works seamlessly with data protection, access control, and encryption.

The NetApp Distributed Content Repository Solution
The NetApp Distributed Content Repository solution addresses the needs of organizations requiring large, distributed, multisite content repositories. The solution consists of NetApp® StorageGRID® software running on a virtualized server infrastructure and NetApp E-Series storage systems for storing data. Additionally, NetApp offers comprehensive planning, implementation, and support services. The NetApp Distributed Content Repository solution addresses the challenges of managing petabytes of distributed objects as follows.

Intelligent management
The NetApp Distributed Content Repository solution provides you with the ability to store petabytes of data and billions of files or objects across hundreds of sites in a single, location-independent namespace. StorageGRID uses object- and file-level policies to manage where and how long data should reside while accommodating the need for a data infrastructure that enables secure, boundless retention policies—for years, decades, or "for- ever." In addition, it can identify and manage content by type to comply with corporate policy and industry regulations, including unauthorized access, alteration, or deletion. Finally, policy-driven storage management aligns the cost of your storage with the value of your datasets by migrating the data onto the appropriate type of storage— from high-performance drives to tape.

StorageGRID uses metadata to automate the management of data, lowering your management costs while allowing you to remain compliant with corporate, regulatory, or user require-ments. By providing a new metadata-based data management approach, StorageGRID transcends the limitations of traditional storage containers and hierarchical organization structures, allowing data to be organized, accessed, and managed using multiple user-defined criteria. NetApp StorageGRID eliminates the constraints of mapping data into predefined containers as blocks or volumes, allowing repositories to accommodate billions of files or objects and many petabytes of capacity in a single, unbounded repository that spans platforms, technologies, and sites.

In addition, StorageGRID enables intelligent data classification and access. Using metadata that describes the data stored in the repository, policies can be executed in a highly granular, yet efficient, way.

Global access and effective retrieval
Organizations that store petabytes of data across a number of sites typically exceed the capabilities of the traditional setup of a primary data center and a secondary (or disaster recovery) data center. Monolithic storage architectures are no longer adequate for these large, globally distributed content repositories. Existing storage constructs are typically too complex for large-scale content management. The administrative burdens are prohibitively expensive, and the requirement for geographically distributed content has increased management complexity to a point at which traditional storage products are no longer effective.

In addition, different constituencies across an enterprise will have different retrieval requirements, and different industries will have different storage and retrieval needs. For text-based data, keyword searches based on indexing the content can assist in finding data. But for nontext objects, traditional infrastructures with basic file-name conventions are inadequate for managing billions of objects.

The NetApp Distributed Content Repository solution helps resolve these issues by leveraging StorageGRID's rich metadata capabilities. Objects can be distributed based on metadata to appropriate physical locations in order to minimize latency. This enables a single, logical repository to be accessed from multiple geographically dispersed physical storage resources to support a collaborative workflow, or to simply improve end-user retrieval speed.

Similarly, objects of any type or from any source can be tagged with rich metadata, and then managed and retrieved simply by using that metadata, such as account numbers or the name of a customer, regardless of where the data is physically stored.

Reliable data protection
Meeting backup windows that are shrinking to nearly zero minutes and achieving recovery time and recovery point objectives that keep getting tighter are challenges that cause typical backup and replication technologies to collapse. Zero-downtime architecture that also includes nondisruptive upgrades are features that companies with content repositories are demanding in order to be productive, responsive, and competitive in today's unrelenting environments.

StorageGRID helps you to provide business continuity even in the event of disasters through multisite content federation, allowing you to meet your SLAs with confidence. Configurations can be designed for resilience against one or multiple simultaneous faults, even entire site losses and regional disasters. NetApp StorageGRID is suitable for single data centers or multi-data-center deployments with many sites across the globe.

In addition, you can leverage StorageGRID's ability to associate rich metadata at the object level to create highly granular storage policies to implement appropriate levels of recovery time and recovery point objectives based on data value.

NetApp Distributed Content Repository Solution
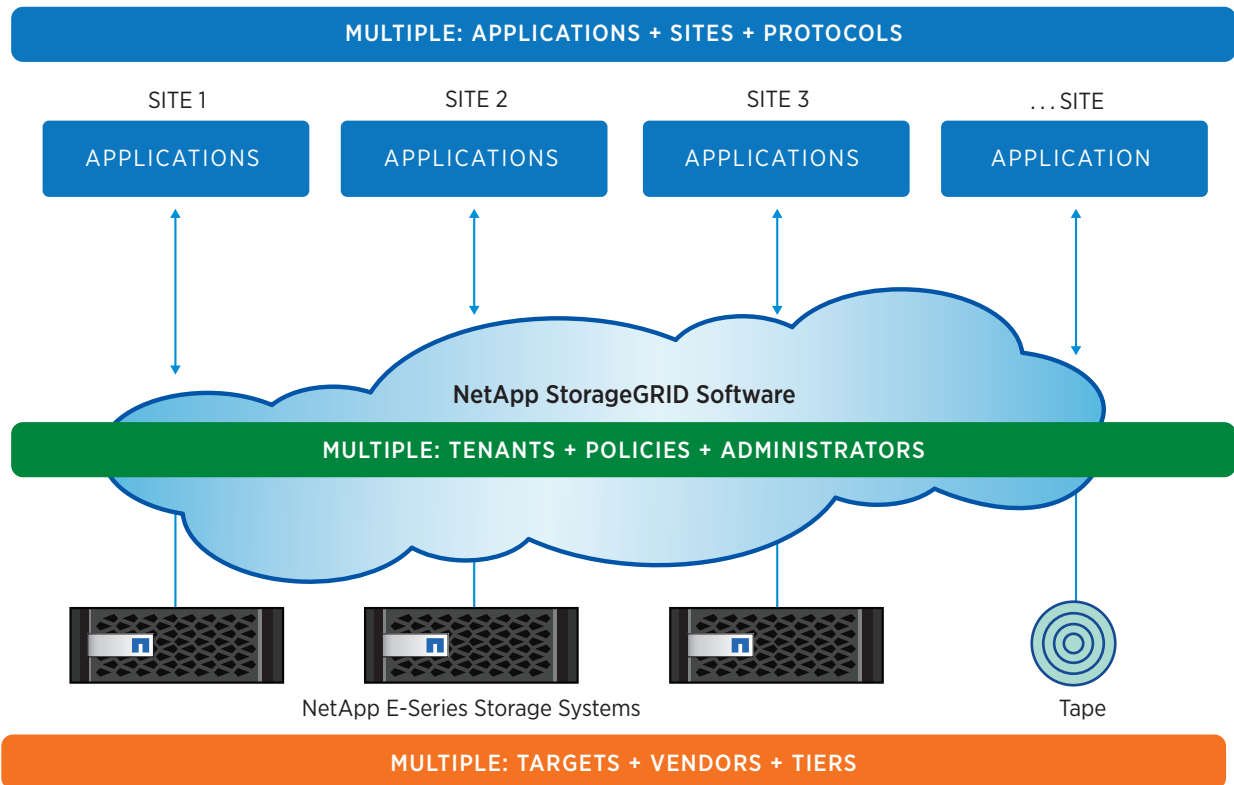
The E-Series storage system enables not only high-speed data access, but continuous access to the data as well. With over 20 years of storage development behind it, the E-Series is based on a field-proven architecture designed to provide high reliability and availability. Redundant components, automated path failover, and online administration keep organizations productive 24/7/365. Plus, extensive diagnostic capabilities and advanced protection features, including Dynamic Disk Pools and SANtricity® Snapshot technology, deliver high levels of data integrity.

Cost-effective infrastructure
In addition to the up-front acquisition costs for petabyte-sized repositories, there are long-term operational expense considerations, including power and cooling as well as data center real estate. Even if you aren't in an urban location with high square-footage costs, if you can't expand beyond your existing footprint at some point you will need to expand or move if no space is available. While keeping service-level agreements and increasing data demands in mind, organizations are looking for a better way to reduce their data center footprint and maintain high availability—all while keeping it simple and within their limited budget.

As noted earlier, StorageGRID's policy-driven storage management enables you to migrate data onto the appropriate storage tier based on metadata, allowing you to store high-value, low-latency data on high- performance drives and migrate to lower-cost disks and, eventually, to tape as business needs and policies dictate. Integrated object-level encryption means that sensitive data can be protected through the entire lifecycle, including storage on tape.

E-Series storage is purpose-built for capacity-intensive environments requiring optimal space utilization and reduced power/cooling requirements. Ultradense disk drive and shelf technology reduces space requirements by up to 60%, while intelligent power efficiency design can lower power and cooling costs by up to 10%.

Compliant retention
One key question customers ask is, "How long must I keep this data?" For some data, the answer is provided very clearly by government regulations. For other data, the answer is based on the nature of the data itself. As noted earlier, business, legal, and regulatory policies all drive a variety of retention requirements. Another issue is the fact that much of the data being generated will be retained for decades—far beyond the refresh cycle of the underlying infrastructure. For example, healthcare organizations must store very large patient content files, including patient records, CT and MRI scans and other imaging files, and diagnostic reports, for time frames that typically exceed the lifetime of the patient.

Metadata can be used to drive policy decisions around retention in ways that are not possible with traditional storage systems. The ability to add rich metadata to nontextual objects such as images and videos enables sophisticated content movement and access controls.

In addition, StorageGRID integrates with E-Series to provide object-level encryption to enhance overall security within a corporate environment as well as to support multi-tenant environments for service providers.

Getting Started

The NetApp Distributed Content Repository is a cost-effective, object-based storage solution that leverages multiple NetApp technologies and our years of content and storage expertise.

NetApp Professional Services can help you design and configure a tailored Distributed Content Repository solution for your enterprise, enabling you to achieve maximum productivity and value from your data. Our proven methodologies and best practices will enable a successful implementation with minimal disruption and risk. NetApp Professional Services speeds the delivery and

adoption of your object-based storage management solution through the implementation planning, installation, and configuration of your StorageGRID software and NetApp storage systems hardware and software.

To learn more about the NetApp Distributed Content Repository solution and how it can help you manage billions of files or objects and petabytes of content, contact your NetApp sales representative or NetApp channel partner today.

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of applications and data across cloud and on-premises environments to accelerate digital transformation. Together with our partners, we empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize their operations. For more information, visit www.netapp.com. #DataDriven