

Simple. Extensible. Open.

Unleash the Value of Data with EMC ViPR Global Data Services

Abstract

The following paper opens with the evolution of enterprise storage infrastructure in the era of Big Data and mobility and discusses the challenges of heterogeneous storage silos in supporting the new demands for cloud-scale applications. It continues by introducing EMC ViPR Software-Defined Storage and details how ViPR Global Data Services it enables enterprise IT departments and service providers to deliver innovative data services across arrays.

May 2013

Copyright © 2013 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is.” EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions. All other trademarks used herein are the property of their respective owners.

Part Number H11802 May 2013.

Table of Contents

Executive Summary	4
The Rise of Big Data, Cloud and Mobility	4
A Data Tsunami	4
The Challenges of a Mixed Workload Environment.....	5
Rethink Storage: EMC ViPR Software-Defined Storage.....	6
The Software-Defined Storage Model.....	6
ViPR Global Data Services	7
Building an Enterprise-Grade Cloud with ViPR Object Data Service	8
Big Data Analytics with ViPR HDFS Data Service.....	9
ViPR Global Data Services Benefits.....	9
Conclusion.....	10

Executive Summary

Driven by ubiquitous mobile devices, pervasive network connectivity and new Web, mobile and cloud applications, IT is facing a data tsunami. A typical enterprise data center experiences 60% annual data growth, primarily in the form of unstructured content. Data sources such as mobile phones, sensor devices, and other sources of telemetry continue to change the way we live. The content they produce continues to expand in frequency, fidelity and richness. Innovative companies have proven that all data, regardless of origin or age, can be converted into meaningful business insight and value.

The confluence of these factors is driving exponential growth in unstructured content. This content growth forces enterprises to think differently about their traditional applications, IT operations and infrastructure. Their existing platforms, operational processes, governance models, and technical expertise weren't developed with the current wave of content expansion and value creation opportunities in mind. However, building green field environments isn't practical, and even if it were, the technologies on which to do so are invariably immature and comparatively unproven when compared to their current infrastructure and experience. Consequently, the question for these enterprises is how to evolve existing investments and expertise to address these new realities.

EMC® ViPR Global Data Services, an integral part of EMC ViPR Software-Defined Storage, allows customers to deploy a rich set of application data services on top of existing infrastructure to meet the challenge of the data tsunami. At its core, ViPR aggregates existing infrastructure into a logical storage fabric and implements a massively scalable data services engine on top. EMC ViPR Global Data Services allow customers to extend the value of their current investments and deliver large scale applications and infrastructure management that meet the challenge and opportunity of the Big Data explosion.

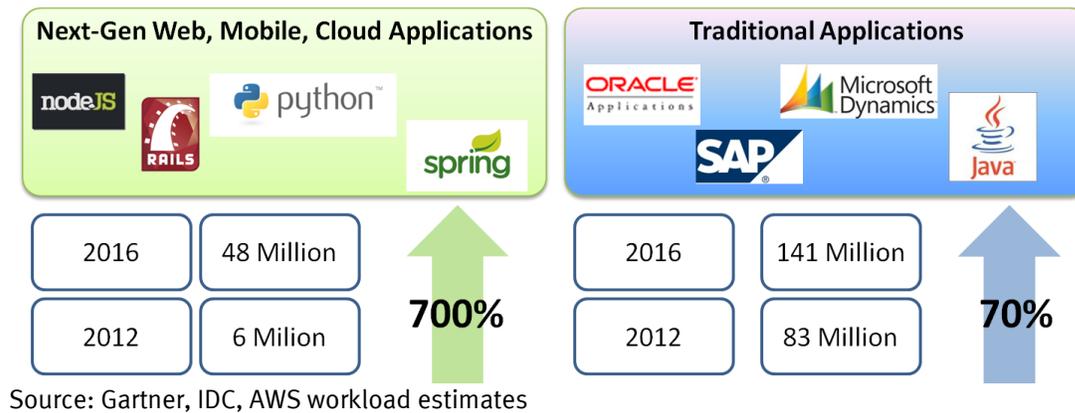
The Rise of Big Data, Cloud and Mobility

A Data Tsunami

Traditional IT has evolved in a world built around scale-up applications and thick-client software. However, this model has its limits and industry trends are forcing enterprises to re-evaluate their service catalogs.

Traditional applications such as customer relationship management (CRM) and enterprise resource planning (ERP) applications still make up the majority of enterprise workflows (Figure 1) and they're growing 70% per year. But next-gen Web, mobile and cloud applications are growing 700% per year. Even more importantly, these workloads generate tremendous amounts of unstructured data, with expected growth of more than 100x over the next decade.

Figure 1: Traditional and Web-scale application Growth Unabated



The Challenges of a Mixed Workload Environment

The rapid changes in infrastructure, competition and user behavior have created a host of new challenges for enterprises:

- **Unprecedented data growth increases costs.** A typical enterprise data center experiences 60% annual data growth. At that growth rate a 10 TB data center in 2001 would be a 120 PB data center by 2021. And much of that content is infrequently accessed unstructured content that need to be kept available and searchable for an increasingly long time.
- **Proliferation of data silos adds complexity.** The result of operating a mixed workload environment is a heterogeneous storage infrastructure to support it. Transaction-intensive applications rely on high-performance file and block-based storage arrays. Web and cloud applications like content sharing are more suited to low-cost commodity storage. Archiving requires yet another storage tier. The heterogeneity, while necessary, adds complexity and makes it difficult to share data across applications or automate provisioning and management tasks.
- **Public cloud providers are taking advantage.** Public IaaS providers make their low-cost, scale-out commodity infrastructure easily accessible via open APIs. Not only is the infrastructure suited to cloud-scale workloads, it's designed with the developer in mind. The success of public cloud has been driven by developer ecosystems that enhance the value of the public cloud service.
- **Lack of Big data analytics impedes business insights.** Businesses are increasingly trying to mine their data to gather more business intelligence. Many large enterprises are experimenting with Hadoop and Hadoop Distributed File System. But much of this is experimentation in a lab. Performing analytics across a heterogeneous, distributed storage infrastructure with an enterprise SLA is a far cry from a working lab implementation. Using a cloud provider poses another set of challenges like data ingest, data egress, SLAs, and security. Consequently, many enterprise analytics

implementations are stuck in the lab, or, if in limited production, are limited to specific data silos.

- **Complex workflows add inefficiency.** Enterprises have to create very inefficient workflows to get value from different data sets. For example, an organization may ingest a lot of content from a Web application that relies on an object storage platform, but have a need to manipulate that data with an application written to a file system. This requires that they move or copy the data from the object store to the file system that serves the file-based application. Or, they'd have to re-write the file-based app to run on the object store (and give up any unique attributes of the file array).

Until now, enterprises have had few options. They could build a cloud-scale infrastructure and analytics themselves, but that requires significant investment and technical expertise beyond the reach of most enterprises. They could tap into public cloud services, but the service level agreements (SLAs) for those services are largely inconsistent with traditional enterprise expectations. Lastly, many organizations will move or copy data from one silo to another to perform analytics or make it available to other applications and uses. But, again, this is inefficient and complicates workflows. Rather than perpetually moving data, re-writing applications and moving sensitive workloads to a public cloud, enterprises need a simple way of cloud-enabling their existing applications and data. Efficiency results from executing as many workloads as possible on the same infrastructure. If they can do that, they can continue to leverage the capabilities of their different arrays and open up new use cases for their data.

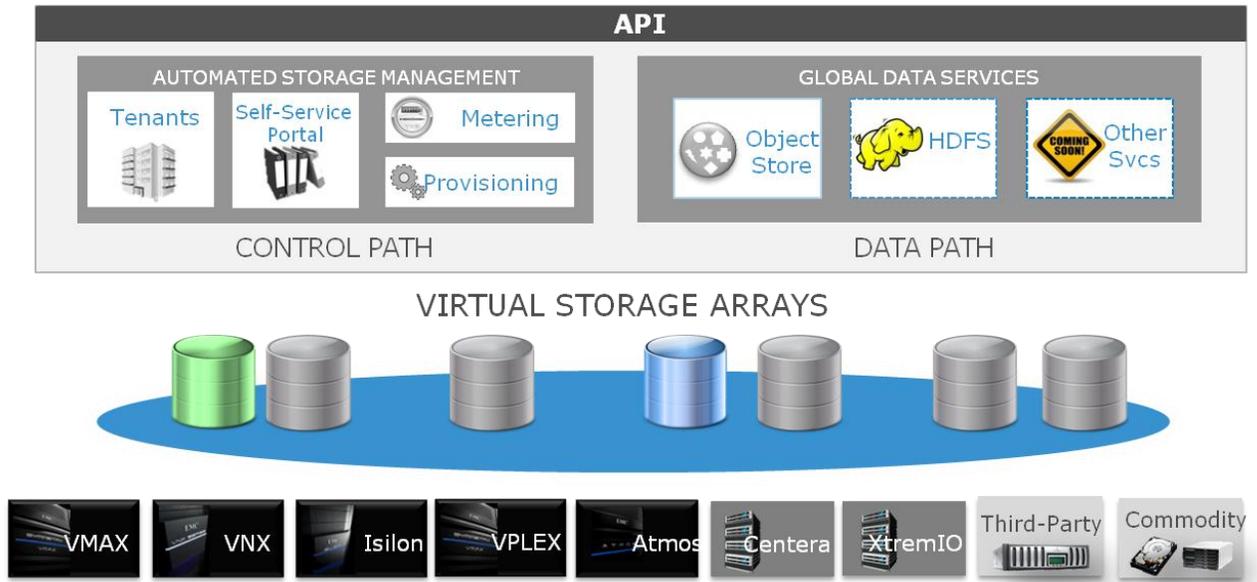
Rethink Storage: EMC ViPR Software-Defined Storage

The Software-Defined Storage Model

The EMC white paper, [“Rethink Storage: Transform the Data Center with EMC ViPR Software-Defined Storage”](#), introduces EMC ViPR and details the software-defined storage model. EMC ViPR is a storage virtualization software platform that abstracts and virtualizes storage while maintaining the unique capabilities of the underlying storage arrays. ViPR abstracts the storage control path from the underlying hardware arrays so that access and management of multi-vendor storage infrastructures can be centrally executed in software (Figure 2).

ViPR is much more than storage virtualization; it is true software-defined storage. Basic storage services such as file, block, and object storage and storage characteristics such as replication, compression, and high availability are defined in software and delivered as services. Storage administrators define various virtual storage pools that reflect different service levels. Examples could include a high-availability virtual storage pool, a transactional virtual storage pool or a geo-replicated object-based virtual storage pool. A user subscribes to the virtual storage pool that best meets the needs of their application workload.

Figure 2: EMC ViPR Software-Defined Storage

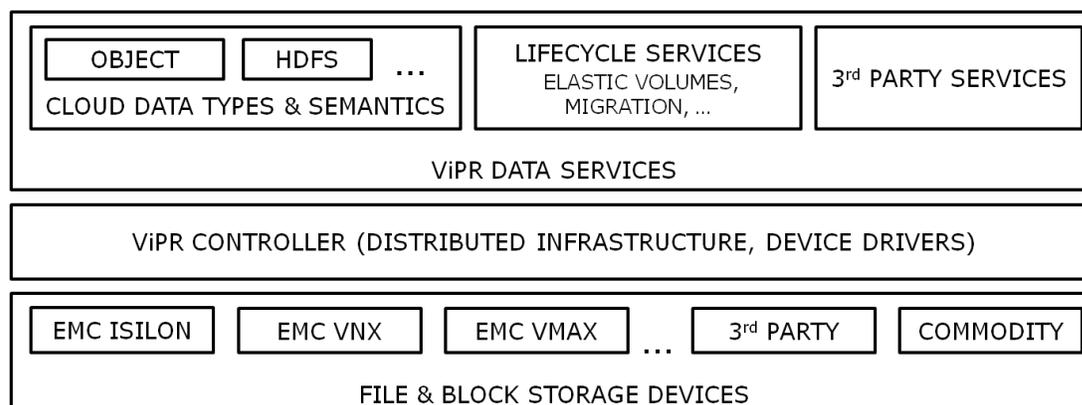


The immediate benefit of ViPR is its ability to automate storage management and provisioning and make storage available as a self-service, consumable resource within a software-defined data center (SDDC). But looking beyond the operational benefits, ViPR transforms how enterprises deliver data services. With storage arrays and storage services defined in software and managed by policy, ViPR enables organizations to deploy unique Global Data Services that cloud-enable existing infrastructure and extend the use cases for their data and the value of their storage investments.

ViPR Global Data Services

ViPR aggregates multi-vendor heterogeneous storage into a unified “storage fabric.” This storage fabric in turn can be leveraged as a logical scale-out layer which can serve as the underlying infrastructure for hosting a range of data services to support collecting, managing and utilizing unstructured content at massive scale (Figure 3).

Figure 3: ViPR Global data Services Architecture



ViPR Global Data Services are implemented in software and feature a simple, lightweight, low-touch, scale-out design. ViPR Global Data Services are critical enablers for Cloud and Big Data applications. Examples include:

- **Object-on-File Data Service:** The EMC ViPR Object-on-File Data Service provides the ability to store, access and manipulate unstructured data as objects on file-based storage such as EMC VNX® and Isilon® and NetApp® storage systems without having to rewrite existing file-based applications. Applications can access a set of objects as files, directly on the underlying file storage device, with native file system performance.
- **HDFS Data Service:** Hadoop Distributed File System (HDFS) support will allow enterprises to leverage their existing storage to build big data processing environments that are robust and effortlessly scalable.

Building an Enterprise-Grade Cloud with ViPR Object Data Service

A major challenge for enterprises seeking to deploy cloud infrastructure is that cloud storage platforms typically lack many features taken for granted in enterprise data centers such as snapshots, compliance, etc. ViPR Global Data Services support these critical functions to integrate seamlessly into traditional enterprise processes. As a result, ViPR Global Data Services transform existing infrastructure into cloud platforms.

As the nature of applications and the IT delivery model evolves, service providers are becoming an increasingly important part of every company's IT strategy. The ViPR approach enables service providers to pair EMC's technological expertise in data management with their own expertise in data centers, connectivity, and operations so they can offer a comprehensive and rich data services catalog to their customers. A broad and robust service catalog is critical for service providers to create profitable clouds. Layering more services on the same infrastructure reduces a service provider's total cost to serve and extends the range of applications they can

support. Service providers can improve their profitability by building a multi-tenant, high scale, low touch, cloud storage platform that supports both object and file workloads on one platform.

Big Data Analytics with ViPR HDFS Data Service

Hadoop has become a de-facto standard for companies that are investigating novel strategies for addressing their Big Data challenges. HDFS is the core distributed file system used by Hadoop. Many organizations have an HDFS project in their labs. However, many of these companies have found Hadoop to be difficult to deploy and manage at scale. The ViPR approach to HDFS takes advantage of proven storage hardware to overcome this challenge. Instead of building a discrete analytics silo with dedicated infrastructure, the ViPR HDFS data service leverages the existing ViPR virtualized storage environment and the backend storage platforms it utilizes.

The ViPR HDFS data service can provide a turnkey, highly scalable, highly available HDFS storage and namespace. The HDFS data service also enables unified access so organizations can run mixed workloads on the same set of data without moving data sets between storage devices. Administrators can use existing Hadoop® stacks or HDFS application ecosystem against any file storage device in their environment. Much like the ViPR Object-on-File data service can make existing data available to cloud use cases, the HDFS data service makes existing data available to Big Data analytics.

ViPR Global Data Services Benefits

Object-on-File and HDFS data services are just two examples that demonstrate how ViPR Global Data Services give customers choice and flexibility. ViPR Global Data Services address their key challenges when supporting a mix of cloud, Big Data and traditional enterprise workloads in the following ways:

- **More effective tiering to manage data growth.** The explosive rate of data growth makes it imperative that organizations have an automated tiering strategy. ViPR global data services will enable hybrid scenarios using multiple platforms to minimize TCO over the data lifecycle. ViPR enables a file or object to be stored on the storage platform that reflects the value of that data at that point in its lifecycle, and maintain its accessibility.
- **Break down data silos.** Again, the Object-on-File data service is an excellent example of breaking down data silos. A Web application can ingest content into an object store and the Object-on-File data service allows a file-based application to manipulate the objects in place without the need to re-write the application.
- **Offer a public cloud alternative.** Enterprises can offer the same user experience as a public cloud provider but maintain control of the infrastructure and leverage the value of the arrays to provide enterprise-grade cloud services.

- **Big Data analytics across the environment.** Enterprises can perform analytics with a turnkey HDFS implementation. Organizations can perform analytics across a heterogeneous, distributed storage infrastructure without having to move data among different arrays. They can offer analytics as a service to their internal constituents.
- **Simpler, efficient workflows.** Enterprises and service providers can increase the efficiency of workflows by allowing data to be manipulated in place through different access methods. Now the organization can ingest objects from a Web page, and then allow a file-based application to manipulate the object in place. They don't have to move the data or re-write their application.

The ViPR Object-on-File and HDFS data services are just the beginning. ViPR is an extensible platform that supports many additional data services. EMC will add additional data services over time and expose a rich set of platform primitives so that customers can develop new data services themselves or leverage 3rd party data services offered by partners.

Conclusion

ViPR Global Data Services provide storage services at cloud scale to transform their existing investments and expertise to address the demands of the unstructured content explosion and capitalize on the promise of Big Data analytics. ViPR provides a unified platform for data services that can be used as different semantic views of the same data. Enterprises and service providers can layer cloud data types over their mixed storage infrastructure, defined in software and delivered at scale. ViPR provides compatibility with existing applications by allowing native access to the underlying storage.

Enterprises and service providers now have more choice and flexibility when supporting mixed workloads. Data can reside on the storage infrastructure best suited to its value but be accessed and manipulated in place by cloud-scale and file-based applications seamlessly. Organizations can easily transform existing data into cloud data, perform analytics across multiple storage platforms and deliver enterprise-grade cloud services that can meet and exceed the most demanding end users and customers.