White Paper

RETHINK STORAGE

Transform the Data Center with EMC ViPR Software-Defined Storage

Abstract

The following paper opens with the evolution of the Software-Defined Data Center and the challenges of heterogeneous storage silos in making the SDDC a reality. It continues by introducing EMC ViPR Software-Defined Storage and details how it enables enterprise IT departments and service providers to transform physical storage arrays into pools of virtual shared storage resources enable the delivery of innovative data services across arrays.

May 2013

EMC²

Part Number H11749.1

Table of Contents

Executive Summary

The past decade and a half has seen virtualization technology transform applications, servers and networks into software abstractions that enable data center and IT managers to build adaptive and agile data centers. The rise of the Software-Defined Data Center (SDDC) promises to build on the progress of virtualization by completely abstracting every component of the data center from its underlying hardware so that IT can truly deliver IT resources as customizable, on-demand services. This is the transformative potential. However, the reality is that storage is still acts as a headwind to a truly virtual data center. Unlike applications, servers and networking, storage and its valuable data is still too often tied to proprietary hardware.

A major reason that storage has lagged behind server and network virtualization is its inherent heterogeneity. Storage hardware and operating systems vary much more than server, client or network platforms. Storage platforms are incredibly diverse; even different arrays from the same vendor will feature different operating systems, proprietary APIs and unique feature sets. Storage naturally evolved this way over decades as a response to new and different application workloads that require unique performance and protection characteristics. Every new IT endeavor required a new storage array - be it block, file or object-based - optimized for that purpose. Out of necessity, storage administrators have become storage managers who spend most of their time managing arrays rather than optimizing information storage for the business. If enterprises and service providers are going to break from this pattern and be part of the evolution to a SDDC they need to fundamentally rethink storage.

Disruption in storage is here. EMC ViPR brings the same virtualization benefits enjoyed by the compute and network elements of SDDC to storage. EMC ViPR is a revolutionary approach to storage automation and management that transforms existing heterogeneous physical storage into a simple, extensible and open virtual storage platform. The value proposition of the SDDC and cloud computing – easily consumed IT services, simple API access, and single management view – is now available for storage.

The Evolution of Software-Defined Data Center and the Challenge of Enterprise Storage

The Rise of the Empowered End User and Customer

The relationship between the consumers and providers of IT has been turned on its head. With the advent of public cloud “as-a-service” offerings, circumventing IT no longer requires an engineering degree, a high-end Windows or Linux server and expensive software. All it takes is an Internet connection and a credit card. The ease with which developers and end users can find ready technology alternatives means they can hold enterprise IT departments and traditional IT service providers to a much higher standard for service and delivery. Fair or not, IT departments and service

providers find it increasingly difficult to dictate technology choices to their internal and external customers; they have to compete for it.

The Storage Evolution: How We Got Here

Storage as it is deployed today often impedes the progress of data center virtualization and the SDDC, increases operational costs, and is not well suited to the next generation of Web, mobile and cloud applications. Neither enterprises nor service providers can easily offer differentiated data services or create the kind of open platform that attracts developers and facilitates innovation. Every data center workload is moving inexorably to a consumption-based, private, public or hybrid cloud service model. Storage must adapt and align with new data center operations and cloud stacks so that it too can be consumed as a service.

IT organizations have historically deployed different storage arrays optimized for specific application workloads, data types, access protocols, and performance and protection requirements. These applications take advantage of unique data services embedded in these storage arrays.

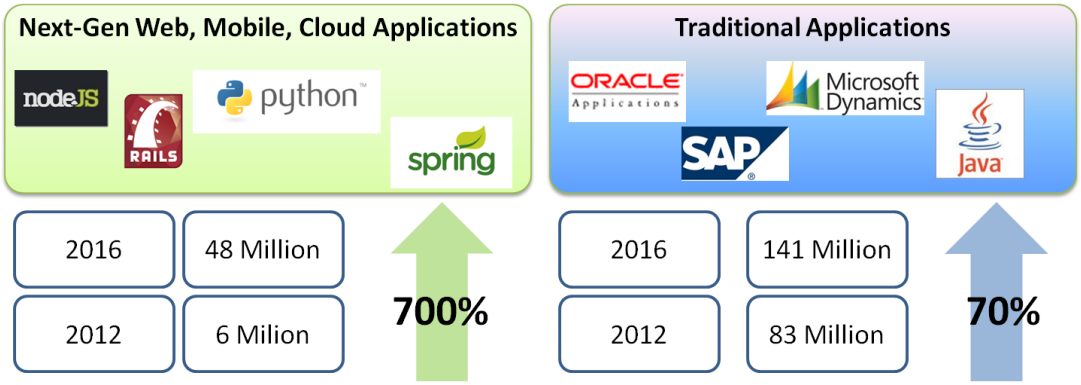
Consider, for example, a global financial trading application where:

- New trades must be transacted in real-time
- Historical trades must remain easily accessible for end of month, end of quarter, and end of year processing
- Yearly historical records must be archived for seven years for compliance but will never require real-time access
- All of the above must be protected against disaster enabling rapid application restart in an alternate site.

The sheer nature of these different data classifications forces different price/performance trade-offs on storage. The next generation of cloud applications will require storage with very different performance and reliability characteristics which will push IT to deploy yet more types of storage with the proper price/performance trade-offs.

Storage administrators are adept at intricately aligning the right storage array to the right data and the right application. But the resulting management complexity and operational rigidity fundamentally prevents IT from fully incorporating storage as a seamless component of the SDDC and cloud operations stacks.

Figure 1: Traditional and Web-scale application Growth Unabated



Source: Gartner, IDC, AWS workload estimates

Storage at a Tipping Point

If enterprises and service providers could centrally manage these varied workloads and offer the same user experience as public cloud providers, they would transform a perceived weakness into a sustainable competitive advantage. Getting to that advantage, however, has a host of immediate challenges including:

- **Storage environments must be designed for the Virtual Data Center and cloud.** The design premise of a traditional storage environments assumes the data path between applications, server, and storage is known, does not change frequently, exhibits a predictable traffic pattern, and is mainly confined within the boundaries of a data center. None of that holds true for a highly-virtualized data center or cloud. Traditional storage silos tightly tied to applications and servers are at odds with virtual data centers and dynamic applications and workloads.

- **Storage must operate at the proper level of abstraction.** Virtual machines (VMs) and virtual applications (vApps) and virtual storage volumes operate at the logical level while storage arrays still operate at the physical level (LUNs or file shares). Operating at the physical level requires intervention from storage administrators who must create LUNs and file shares and manage their lifecycle. Because of this, IT can't easily integrate storage into data center and cloud operations stacks nor can they give end users a simple, flexible self-service capability to consume storage on demand.

- **Data types and access protocols must be decoupled from the underlying hardware.** In traditional storage environments the data access protocol is tightly-coupled to the storage array. Consider a business application that wants to ingest data over the Internet as objects, process the data as files, and then re-publish to the internet as objects. They first need to put the data in the object store. Then, they need to copy it to a filer, and then copy it back to the object store. This massive duplication of data results in complex data processing workflows, longer execution time, and requires twice the amount of storage capacity.

- **There is too much existing data to just 'start over.'** Any solution to this problem, however innovative and transformative, must encompass existing data. If not, organizations will just end up with yet another storage silo. Storage must be viewed holistically.

- **Storage complexity increases costs and management burden**. The diversity of storage arrays necessary to satisfy business data requirements leads to multiple management control points. Each file, block and object-based storage array has its own management interface and its own provisioning and management processes. Multi-data center environments exacerbate the management headaches. Administrators have poor visibility into storage usage, available capacity, performance and system health. That lack of visibility and self-service capabilities means storage administrators spend an inordinate amount of time identifying available storage capacity and manually provisioning storage. This inefficient use of valuable staff time results in an IT department that can't deliver storage quickly or cost-effectively, inconvenienced users, and virtualization and automation investments that fail to reach their potential.

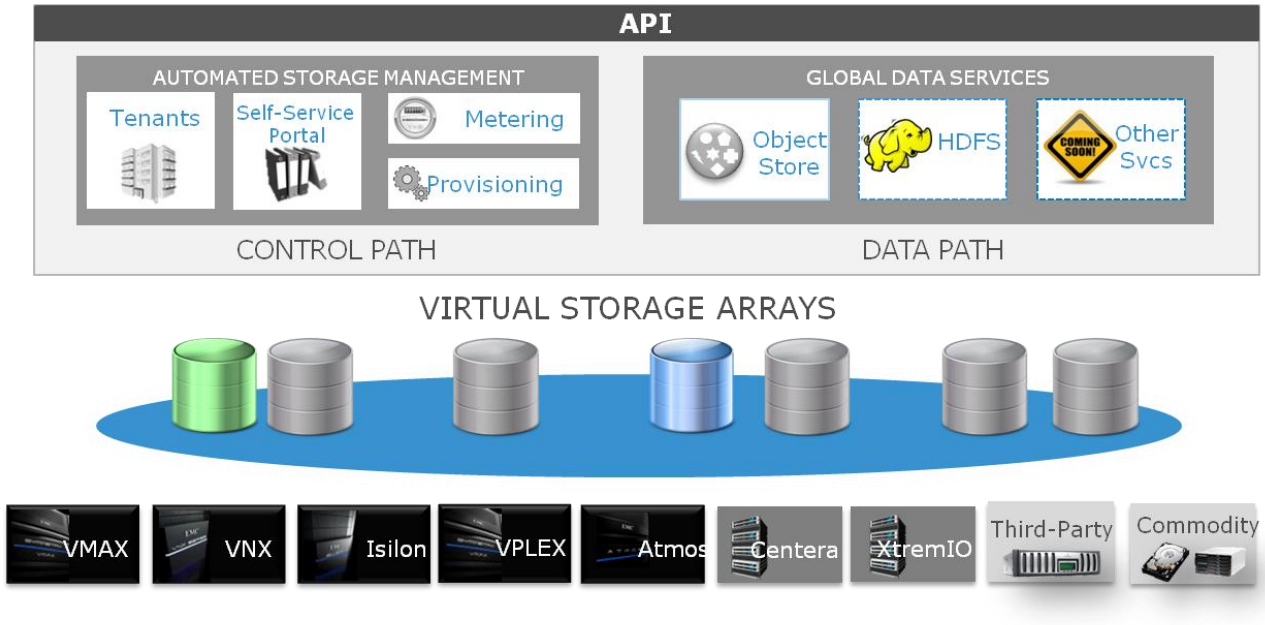
- **Proprietary storage APIs impede development and service delivery**. Consolidating data center infrastructure, outsourcing commodity IT tasks and shifting IT budget from CAPEX to OPEX remain important benefits of virtualization and cloud computing. But the real value is not just cost reduction; it is the ability to rapidly onboard new services and engage an ecosystem of application developers that create and launch new services. In the eyes of a developer, the beauty of cloud is the ability to simply write to an API. They don't need to concern themselves with the underlying infrastructure. Closed, proprietary storage platforms and the lack of open APIs force developers to write more code to access storage. Worse, they have to recode applications when infrastructure changes. It becomes much more difficult for enterprises and service providers to develop applications or to integrate third-party applications into their infrastructure. This ability to quickly develop and deploy applications may be the single most important factor in the success of the SDDC and giving enterprise IT and service providers the ability to compete with public cloud alternatives.

Rethink Storage: EMC ViPR Software-Defined Storage

The Software-Defined Storage Model

EMC ViPR is a storage virtualization software platform that abstracts storage from physical arrays – whether file, block or object-based - into a pool of virtual shared storage resources that enables a flexible storage consumption model across physical arrays and the delivery of applications and innovative data services. 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).

Figure 2: EMC ViPR Software-Defined Storage



Principally, a storage device features a control path and data path. In simplistic terms, the control path sets and manages the policies for the storage device and the data path performs the actual reads, writes and data services. ViPR is unlike previous attempts at storage virtualization in that it decouples the control path from the data path. By abstracting the control path, storage management operates at the virtual layer, which gives customers the ability to partition their storage pool into various virtual storage arrays and manage them uniquely by policy. This is analogous to partitioning a server into a number of virtual machines. However, ViPR does not sit in the data path for file and block storage. This decoupling of the control and data paths enables ViPR to centralize all data provisioning and data management tasks, but allows applications to access file and block data as they always have and continue to use the unique data services embedded in the storage arrays.

ViPR features block and file control services that provide all the functionality of physical block and file storage arrays as virtual services. Block and file control services allows users to manage block volume, NFS file systems or CIFS shares, and advanced protection services such as snapshots, cloning, and replication. ViPR block and file control services offer full storage functionality as if the user were accessing a physical array. In contrast, block storage volumes typically provided for use by virtual compute instances in public clouds can forfeit many advanced array features in favor of using commodity disks for lower cost and operational simplicity. ViPR, however, does not require that sacrifice. The ViPR block and file control services deliver operational simplicity and maintain all the advanced features of the arrays such as mirrors, clones, snapshots, and multi-site high-availability, and replication. ViPR virtualizes storage without compromise.

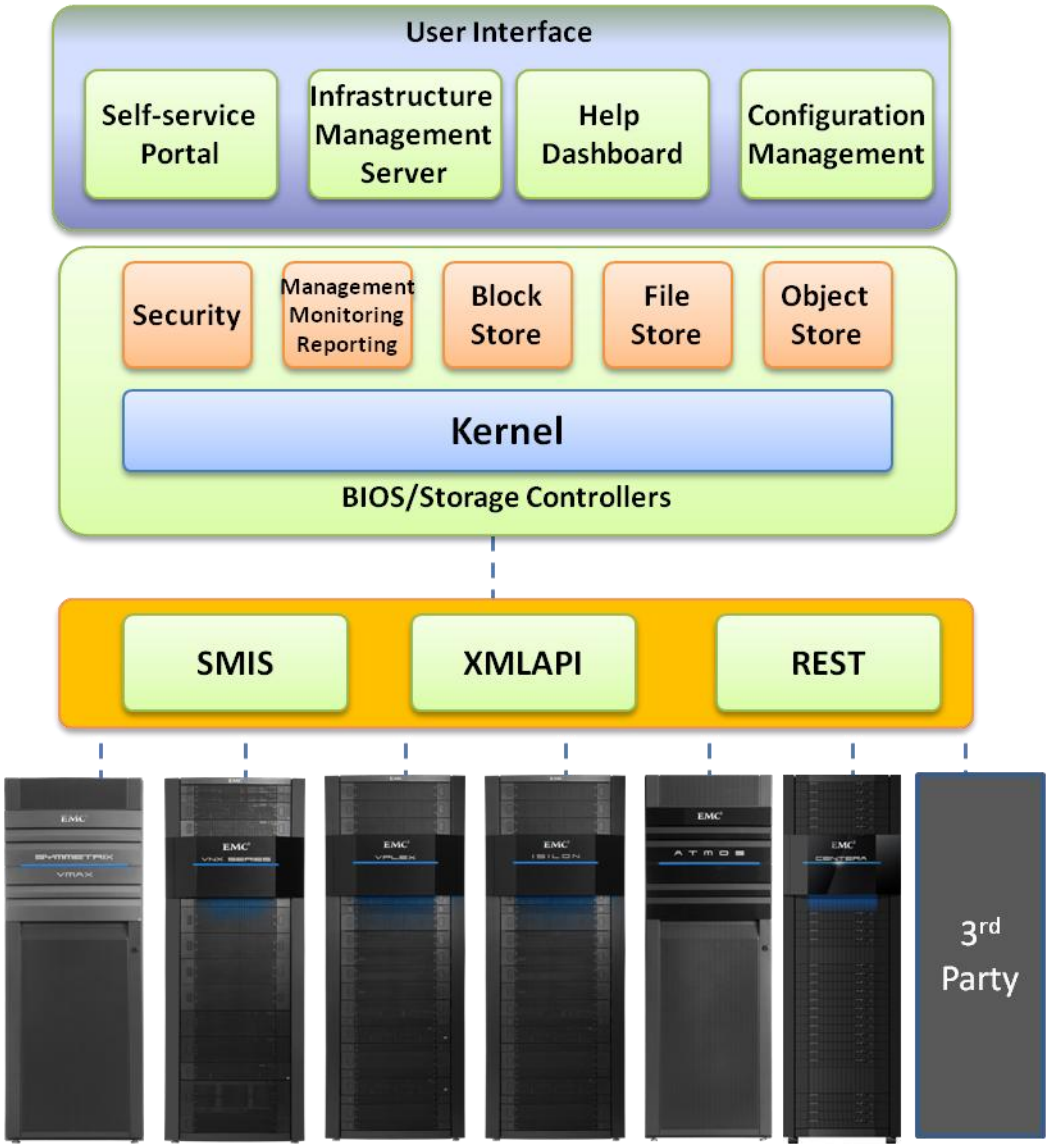
ViPR's unique approach not only maintains the capabilities of the underlying arrays but extends them. Administrators and developers can create new global services that

span arrays. For example, a developer can create an object data service that not only runs on commodity disks, but can also access and store objects on a filer and exploit the performance capabilities of that filer. In this example, ViPR also provides data path access to the object store, with data path access to the filer also available. As an example, organizations can store, access and manipulate objects on a high-performance NAS device without having to rewrite existing file-based applications. Developers can create new array capabilities in software once and then run on any array without having to re-write.

ViPR Architecture

ViPR makes a multi-vendor storage environment look like one, big virtual array. ViPR uses software adapters that connect to the underlying arrays, similarly to how device drivers enable universal device compatibility with a PC. ViPR exposes the APIs so any vendor, partner or customer can build new adapters to add new arrays. This creates an extensible "plug and play" storage environment that can automatically connect to, discover and map arrays, hosts and SAN fabrics (Figure 3).

Figure 3: ViPR Architecture



Once a storage administrator adds arrays, ViPR discovers the arrays and all their corresponding storage pools and ports. Once the Fibre Channel switches are added, ViPR automatically discovers and maps the Fibre Channel networks. And ViPR can accomplish this virtualization and mapping for EMC and non-EMC arrays, including EMC VMAX, EMC VNX, EMC Isilon, EMC VPLEX, EMC Atmos and NetApp. And ViPR will support additional EMC and commodity disks in addition to publishing the APIs.

ViPR hides the complexity of all the underlying storage arrays and exposes their core functionality as data services while retaining the unique attributes of the arrays. Storage administrators then create Virtual Storage Pools in ViPR that represent sets of capabilities required by unique application workloads. For example, a transactional workload would be best served by a Virtual Storage Pool that features the characteristics of high-performance block storage such as EMC VMAX. A cloud application such as online file and content sharing is not performance-sensitive and

would work just fine on commodity hardware that more economically provides the requisite level of data protection and availability. In either case, a user subscribes to a Virtual Storage Pool that meets their workload's demands. The user does not need to know or care about the underlying hardware and software that is providing the data service to their application.

Rather than provisioning space on a specific array, ViPR gives storage administrators the ability to expose a unique and customizable combination of hardware and software resources as consumable data services.

Global Data Services

EMC Software-Defined Storage is differentiated, in part, from storage virtualization and other software-defined storage solutions by its extensibility; administrators and developers can develop new global data services that can span arrays and support hybrid data types. Global data services are storage abstractions that reflect the combination of a data type (file, object, block of data or hybrid data type), access protocols (iSCSI, NFS, REST, etc.), and durability, availability, and security characteristics (snapshots, replication, etc.). Examples of data services include:

Object-on-File Data Service: The EMC ViPR Object-on-File Data Service provides the ability to store, access and manipulate unstructured data (e.g. images, video, audio, online documents) as objects on file-based storage such as EMC VNX and Isilon and NetApp storage systems without having to rewrite or rework existing file-based applications. The ViPR Object-on-File Data Service is a software layer that works transparently with different hardware platforms. Initially, the ViPR Object-on-File Data Service gives users the ability to manage object data using Amazon S3, OpenStack Swift and the EMC Atmos API and access data on file systems. The ViPR Object-on-file data service provides direct data path access to file arrays. Enterprises, in particular, can benefit from this capability since their existing applications that are written to file systems do not have to be recoded to take advantage of ViPR.

- **HDFS Data Service**: Hadoop Distributed File System (HDFS) support will apply location-awareness to data-intensive applications using object and file data services. Processing will be done on the worker node where the data resides without unnecessarily traversing the network thereby reducing backbone traffic.

- **Business Continuity/Mobility Data Service**: The EMC ViPR (virtual) block controller combined with VPLEX and RecoverPoint (physical) block data nodes deliver a global business continuity and mobility data service for VMAX and VNX block storage capable of supporting any workload with snapshots, replication, high-availability, and mobility in a metro area all managed from a single management control point.

With the storage resources virtualized and defined in software, storage administrators, no longer bound by physical storage constraints, can present their heterogeneous storage environment to users and applications as a set of services to which users can subscribe. ViPR's

open, extensible platform serves as a foundation for developing new and innovative data services that expose previously landlocked data to new use cases and applications and give enterprises and service providers the ability to attract an ecosystem of developers and ISVs that add new value-adding data services.

ViPR Global Data Services give customers choice. A workflow's data can reside partially in EMC VNX, partially in EMC Atmos, EMC Isilon, NetApp, Amazon S3 or any combination, completely transparent to the user. Different applications can access and interact with the same data in different ways, marking the end of data silos. Enterprise and service providers get more flexibility with less complexity.

The ViPR Value: Simple, Extensible, Open

EMC ViPR is the first, true storage virtualization platform that delivers on the promise of software-defined storage and the SDDC. ViPR **simplifies** and automates repetitive storage provisioning and management tasks by virtue of an abstracted, central control path. ViPR is an **extensible** platform that enables an organization to develop new data services and adapters to support additional arrays. And ViPR is an **open**, API-driven platform that facilities integration with cloud stacks and development of new and diverse data services and applications.

Simple

The abstracted, central control plane simplifies the operation of a diverse storage infrastructure. ViPR centralizes and automates the entire storage lifecycle process. Administrators can add, provision, manage, and share storage from a single software control point. Enterprises and service providers can operate multi-tenant environments and provide simple, self-service access to block, file, and object storage resources. ViPR monitors and reports on the health of the physical storage infrastructure as well as usage, available capacity, and performance. ViPR can meter storage usage, provide chargeback to tenants and integrate with existing billing systems.

ViPR not only simplifies management, it simplifies delivery and consumption. Rather than provisioning storage for users manually and via scripts, storage administrators can provide users/tenants with a self-service portal in which they allocate storage out of a Virtual Storage Pool. An authorized user or tenant simply selects the Virtual Storage Pool that meets the needs of their application workload. This helps storage administrators minimize user-IT interactions, automate the process of identifying available storage capacity, and better map an application workload's requirements to the right combination of software and hardware storage resources.

Extensible

To be truly software-defined storage, the platform must be programmatic. ViPR features a simple, powerful Representational State Transfer (REST) API that provides connectivity to multi-vendor storage resources as if they were one large storage pool. ViPR's extensibility enables organizations to create new adapters to support additional arrays and create new global data services that run on top of ViPR. Any

customer, partner or service provider can add support for more arrays or develop new data services.

With ViPR, the storage layer can now be another programmatic virtual resource in the SDDC. All data and resources managed by ViPR are accessible via the open API, which also integrates with VMware, Microsoft and OpenStack cloud environments. An organization can easily integrate ViPR into their existing data center operations. ViPR provides specific VMware integration with interfaces into the VMware vStorage API for Storage Awareness (VASA), vCenter Orchestrator and vCenter Operations. For example, a vCenter administrator has end-to-end visibility from the virtual machine to physical storage.

Open

ViPR supports multiple storage APIs including Amazon Simple Storage Service (S3), OpenStack Swift and EMC Atmos. Developers can write applications to multiple cloud APIs and execute those workloads on ViPR in an enterprise data center or a service provider's cloud. ViPR's object-on-file data services makes it possible to also store objects on existing file systems and provide access to these objects from legacy applications without having to modify or recode the applications.

EMC publishes the APIs to facilitate the addition of new array adapters and new data services. This lets developers focus on the on the functionality and value they are adding, rather than on the details of the underlying storage. The open API enables both enterprises and service providers to build developer communities that attract developers and ISVs that expand the universe of value-adding data services.

Conclusion

The rise of public cloud computing in the last few years has prompted executive leadership, line-of-business (LOB) managers, and developers to ask much more of their IT leadership, their service providers and, consequently, storage administrators. The continuing growth in traditional enterprise application workloads coupled with the explosive growth in Web, mobile and cloud applications demands a simple, automated way to align data center and storage resources to the myriad ways in which content is stored, protected and accessed.

The era of data silos is over. True software-defined storage is here. EMC ViPR software-defined storage abstracts storage from physical arrays into pools of virtual shared storage resources that that enable the delivery of innovative data services across arrays. Even complex, heterogeneous, multi-vendor environments with every conceivable application and workload can be made simple to manage, extensible, and open to innovation and value-adding data services.