**CLABBY ANALYTICS**

# *Research Report*

## How IBM is Leveraging Storage, Hybrid Cloud and Software Defined Infrastructure to Drive Big Data, Analytics, and Cognitive Applications

### *Introduction*

In recent years profitability amongst the traditional leading manufacturers of storage arrays has dropped, for several reasons:

1. Numerous workloads moved to commodity arrays operated by cloud storage providers (one report indicates that for every dollar spent on public cloud storage, traditional vendors have lost three to four dollars in revenue opportunity);

2. Data-intensive Big Data/analytics/cognitive workloads have increased demand for faster storage array performance.  To meet this demand, tradition vendors have needed to revamp their storage product lines, introducing new offerings based on Flash-based solid state drive (SSD) technology including hybrid SSD/HDD (hard disk drive) arrays and all-Flash arrays.  Revenue opportunities were lost to niche vendors and cloud storage providers as traditional storage vendors retooled; and,

3. The increased use of commodity storage and software defined infrastructure has served to drive-down storage hardware revenues but has also contributed to storage-related software and service revenues/growth. (Note: software defined infrastructure includes software defined storage, software defined computing, and software defined networking, and implies that software is operating independently of the underlying hardware – increasing flexibility and efficiency.)

> *Traditional storage leaders have responded to changing storage market demands in two ways: 1) by expanding Flash array offerings (to dramatically improve performance); and, 2) by introducing software designed storage solutions (to drive down storage costs, improve cross-array management, and to give users more utilization flexibility).  And one storage vendor, IBM, has also focused on seamlessly blending public and private clouds using a "hybrid cloud" model.*

In this *Research Report*, *Clabby Analytics* takes a closer look at IBM's storage hardware line and its software defined infrastructure strategy.  What we find is that:

1. Over the past two years IBM has significantly revamped its storage array offerings by  introducing combined SSD/HDD arrays and expanding its all- Flash portfolio to include low-end and midrange Flash arrays to complement its high-end offerings;

2. IBM's revamped storage portfolio has become significantly more flexible for customers.  IBM customers can now purchase software-only solutions (to be run on commodity storage), storage appliances, or storage "as-a-service".  These offerings use the same code base regardless of delivery model. The IBM Spectrum Storage

Suite is packaged so that customers can use any or all of the products in the suite and pay based on capacity;

3.  IBM storage hardware combined with its software defined infrastructure portfolio (including IBM Spectrum Storage and IBM Spectrum Computing) has introduced new use-cases that are driving down costs, providing customers with increased flexibility, while also significantly improving performance, scalability, and agility. Accordingly, IBM has decided to invest a billion dollars to expand its software defined infrastructure offerings.  To IBM, software defined infrastructure represents a major growth market – and IBM is investing heavily in order to exploit market growth; and,

4.  *Hybrid cloud* has become the foundation of IBM's storage and software defined infrastructure strategy.

> *After evaluating IBM's storage offerings, our final takeaways are these:  1) IBM has corrected a deficiency in its Flash array product line by expanding its line-up to include mid-range and low-end offerings – and has introduced new SDD/HDD hybrid arrays as well as all-Flash arrays.  These moves have better positioned IBM storage to serve a range of application types from high input/output per second (IOPS) workloads (such as new generation analytics, big data, cognitive and high performance computing workloads) to entry-level applications;  2 ) IBM's storage is broad and flexible, enabling customers of all sizes in all industries to select a solution  and a delivery model that meets their respective price/performance requirements; 3) IBM's software revenues are growing as the company continues to introduce new, innovative software defined solutions that support new customer use cases; and, 4) IBM customers can now benefit from using a range of IBM solutions together in terms of ease of ordering and management, improved performance and cost savings*

### A Closer Look at IBM's Storage Array Offerings
IBM's storage array hardware offerings include hybrid arrays as well as all-Flash arrays.

### IBM Hybrid Storage
IBM complements its all-Flash portfolio with a complete range of hybrid IBM Storwize arrays built with IBM Spectrum Virtualize (part of the IBM Spectrum Storage offerings) software for modular, enterprise class storage. These hybrid arrays include software features such as Easy Tier storage tiering, Real-time Compression (RTC), in-line deduplication, snapshots and thin provisioning designed to maximize storage utilization and efficiency while still optimizing application performance. The IBM Spectrum Storage Suite (more on this next page), supported across the IBM storage family from hybrid to all-Flash, enables the easy evolution from traditional applications running on block-storage – based SANs to new generation workloads designed for object storage and hybrid and/or public cloud infrastructure.

### IBM Flash Solutions
IBM offers the most extensive line of all-Flash arrays in the industry (see Figure 1 – next page).  This line includes the FlashSystem V9000 (designed to help virtualize the data center); the FlashSystem A9000 (designed for cloud service providers); The FlashSystem A9000R and the DeepFlash 150 (designed to serve Big Data applications).

---

**How IBM is Leveraging Storage, Hybrid Cloud and Software Defined to Drive Analytics and Cognitive Applications**

On August 23, 2016, IBM announced three new all-Flash systems that complement its high-end Flash announced and delivered over two years ago.  These new arrays include:

- IBM Storwize V7000F – cost optimized flash for heterogeneous environments with high performance CPU and more cache for up to 45% performance improvement;

- IBM V5030F – low cost for entry level and mid-size wds- Flash orkloaoptimized with Real-Time Compression, distributed RAID and multi-layer write caching; and,

- IBM FlashSystem V9000 – An ultra-low latency enterprise-class array designed for mission-critical workloads (with up to a 30% performance improvement over previous models).

The remainder of the Flash product line includes:

- The FlashSystem A9000 offers full time data reduction and has been designed to serve cloud, virtual desktop and VMware environments;

- The FlashSystem A9000R also offers full time data reduction, but has been designed and optimized to serve mixed workload and cloud environments; and,

- The DeepFlash 150 offers high capacity as well as high density.


*Figure 1 – IBM Flash Portfolio*
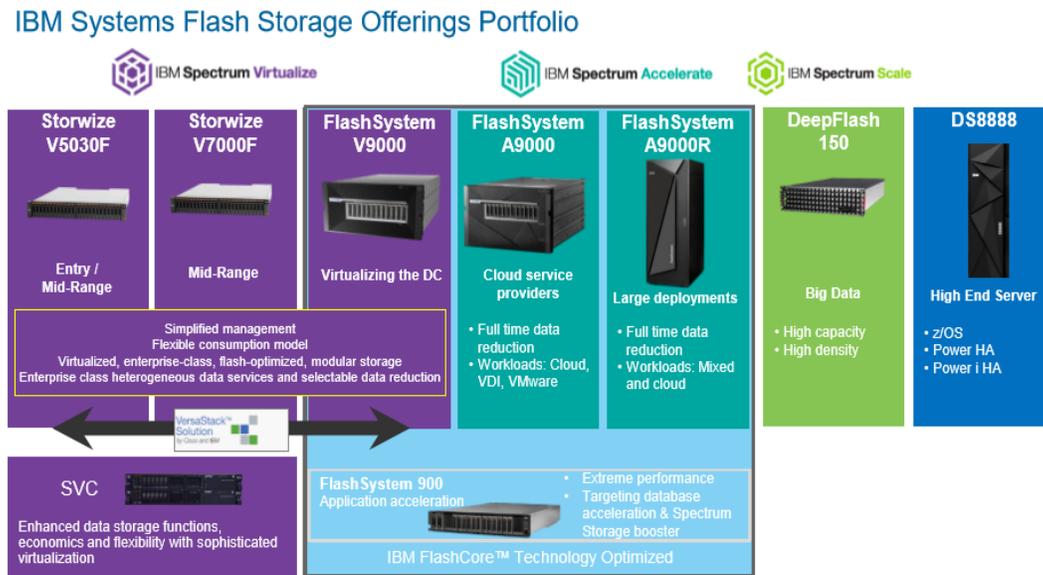


*Source: IBM - October, 2016*


*A Closer look at IBM Software Defined Storage – IBM Spectrum Storage*
The IBM Spectrum Storage Suite offers a family of software defined storage solutions that are available as software-only (for use on commodity storage or a wide range of storage available from other vendors including NetApp, HP, EMC and others), as an appliance, or as-a-service.  In addition, these solutions can leverage a range of underlying storage media

including Flash, NL-SAS or tape, enabling customers meet both high performance and/or high capacity needs with a single solution.

The flexible licensing model of the Spectrum Storage Suite is another plus – allowing customers to mix and match products within the suite as needs change, paying based on capacity. This can provide up to 40% savings over buying the products a la carte, and increases flexibility, enabling businesses to evolve to support new applications and use cases over time without incurring new software license fees. Not only can customers scale up as needs grow but they can also scale back as needs change. The approach is working– with IBM reporting 2900 new customers since launch.

The portfolio includes:

- IBM Spectrum Scale (based on IBM General Purpose File System (GPFS) – Scale-out file storage;
- IBM Spectrum Accelerate – Scale-out block storage;
- IBM Cloud Object Storage (formerly CleverSafe) – Scale-out object storage;
- IBM Spectrum Virtualize – Virtualized block storage;
- IBM Spectrum Protect – Data protection;
- IBM Spectrum Control  - Common management layer across block and object;
- IBM Spectrum Copy Data Management – a new solution that manages the proliferation of copies across the organization, not just those used for replication and data protection , but for other use cases as well.

"*Data Protection for Data Oceans*" is a new customer-driven solution being used currently by EU Water Company that combines the capabilities of IBM Spectrum Protect and IBM Spectrum Scale to improve data protection across "oceans" of data in a scale-out architecture.  EU Water Company has been able to back-up 3.75PB, eliminate redundant de-dupe appliances, and reduce licensing costs by 20%, by purchasing the IBM Spectrum Storage Suite. IBM will continue to identify these types of use-case-driven solutions, exploiting the synergies between IBM products to solve specific customer problems.

IBM plans to support IBM Spectrum Storage offerings, not only in the IBM SoftLayer Cloud, but also in other public cloud environments such as Google Cloud, Amazon AWS and Microsoft Azure.  For example, IBM Spectrum Protect for Amazon S3 Cloud Storage provides data protection leveraging a hybrid cloud model with Amazon S3 (Microsoft Azure and Google support will be added in 1H 2017).  This flexibility enables customers to move to hybrid cloud while supporting traditional and new generation workloads, and also enables IBM and its service providers to offer new services such as backup-as-a-service.

*A Closer Look at IBM's Hybrid Cloud Strategy*
IBM's software defined infrastructure solutions enable businesses to easily move to the hybrid cloud, which IDC reports will represent 80% of enterprise IT by 2017. Hybrid cloud environments optimize traditional applications by leveraging automation to free up resources for new applications, allowing organizations to modernize and transform.

**How IBM is Leveraging Storage, Hybrid Cloud and Software Defined to Drive Analytics and Cognitive Applications**

> *It is this ability to offer a flexible, hybrid cloud architecture that differentiates IBM from its competitors, including not only traditional competitors such as HP and Dell/EMC, but also Amazon and Google.*

*IBM Cloud Object Storage*
IBM Cloud Object Storage (formerly CleverSafe) is available:

- On-premise ;
- Dedicated (single tenant - compliant);
- Public cloud (multi-tenant).

A hybrid cloud model provides a consistent interface and management across all platforms, as well as the ability to manage data and workloads across the hybrid cloud. For example, on-premise storage can be replicated in a $3^{rd}$ data center in the public cloud and the public cloud can add capacity as-needed to an on-premise/dedicated solution.

Two new public (multi-tenant) cloud services have been recently announced by IBM:

- IBM Cloud Object Storage Standard Service -public cloud service for active workloads such as social media, mobile, collaboration,  and analytics; and,
- IBM Cloud Object Storage Vault Service –lower cost public cloud service for less-active workloads including archive, back-up, data retention and colder data.

These services will compete head-to-head with Amazon AWS storage offerings, but provide security and availability benefits not available on AWS. For example, AWS makes copies of data in availability zones within a single region. Data is encrypted but full copies can still be exposed.

With IBM Cloud Object Storage, data is written to a region, encrypted and using erasure coding *dispersed across data centers in other regions*, so data is secure and protected. As a result, a regional outage in AWS will impact applications and access to data. But with IBM Cloud Object Storage, applications and data access are unaffected by a regional outage.

To address cross-regional outages, AWS customers must purchase, set-up and maintain a second replicated copy of data in another region (no extra purchase or set-up needed in IBM Cloud Object Storage). IBM reports a 24% cost savings over Amazon S3 for 5PB of IBM Cloud Object Storage Vault Cross-Region when based on certain workload assumptions.

Future enhancements to IBM Cloud Object Storage will enable deployment over large user populations and capture user metrics. Other areas of focus will be performance optimization (latency/transactions per second) and providing data lifecycle management on object storage.

*Integration with IBM Spectrum Storage*
Integration of IBM Cloud Object Storage with IBM Spectrum Protect, IBM Spectrum Control, and IBM Spectrum Scale enables seamless data back-up, a common management layer, and transparent cloud tiering.

IBM Spectrum Scale Transparent Cloud Tiering allows administrators to (without a gateway) add private or public cloud object storage as a target for IBM Spectrum Scale data.  Data can be tiered to tape, an object store or seamlessly to the cloud including IBM Cloud Object Store, Amazon S3, Google Cloud, Azure, (or any cloud with a RESTful API interface) in the same manner used to store data locally. A planned enhancement will provide Cloud Data Sharing–hybrid cloud persistent storage with native protocols–offering the ability to go from file to object and object to file, enabling new services, as well as a global index and universal access.

Other IBM Cloud Object Storage use cases include block snapshot target using IBM Spectrum Virtualize, log archives, compliance archives, file sync and share and cloud–native applications.

***A Closer Look at IBM Software Defined Infrastucture –IBM Spectrum Computing***
The other element of IBM's software defined infrastructure strategy is IBM Spectrum Computing workload management and resource scheduling (see Figure 2). IBM Spectrum Computing, originally designed to support the high-performance computing (HPC) market, is finding its way into a broader set of workloads and applications including big data and born-in-the-cloud workloads.

***Figure 2, The Big Picture -IBM Software Defined Infrastructure***
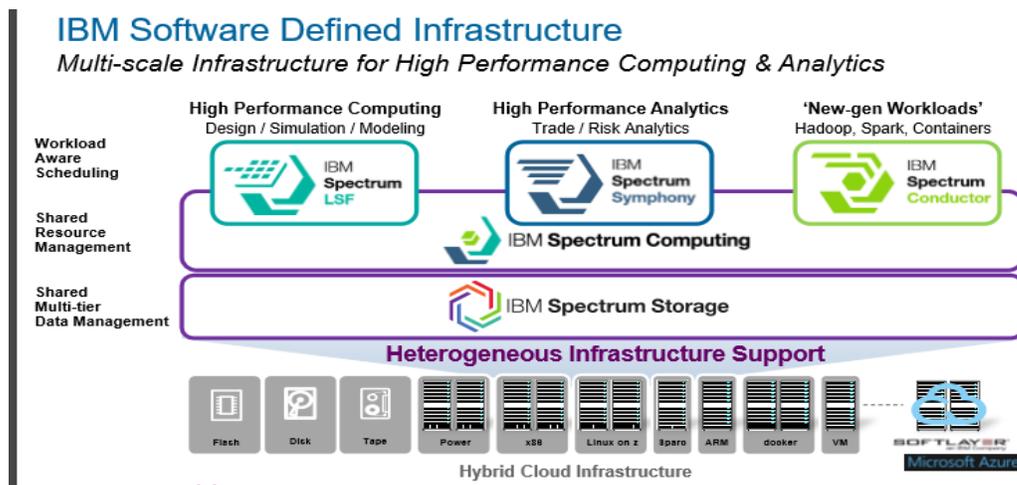


*Source: IBM – October 2016*

Providing intelligent workload and resource management, IBM Spectrum Computing, enables software defined computing – on-premise, on-the-cloud, or both – and is ideal for supporting traditional HPC and a new generation of Big Data, cognitive, and born-in-the cloud applications such as Hadoop, Spark, and Cassandra on a common infrastructure. IBM Spectrum Computing enables the consolidation of discrete IT silos into a multi-tenant

shared infrastructure: cluster sprawl (new workloads each with its own often underutilized cluster) is minimized; overall IT utilization, scalability, and performance are improved; IT agility is greatly improved while time-to-results is predictable and reliable. The portfolio includes:

- IBM Spectrum LSF -Intelligent workload management for HPC workloads, enabling up to 150 times performance gains
- IBM Spectrum Symphony- high performance, low latency grid management software for on-demand, high performance workloads including trade and financial risk analytics
- IBM Spectrum Conductor-Designed for new generation of big data, cognitive, and born-in-the-cloud workloads such as Hadoop, Cassandra, Mongo DB and others; and,
- IBM Spectrum Conductor for Spark – Enterprise-ready Apache Spark enabling better performance, scalability and scheduling fairness when compared to Apache Yarn and Apache Mesos. Src: STAC Research

*Software Defined Infrastructure– Customer Examples*
Examples of the *combined benefits* of IBM's software defined infrastructure solutions:

- *Red Bull Racing* is a designer and manufacturer of Formula One race cars that are used in 21 countries across 5 continents.  Cars must be both fast and safe and comply with industry regulation and auditing processes. To meet these demands, the company makes more than 30,000 engineering changes per year which must be tested in virtual simulations before going live.  The design team relies on petabytes of  data collected from more than 100 lightweight high-power sensors, on-board computing systems, and real-time telemetry as well as  external data (weather, for example) sources. The design team leverages IBM Spectrum Computing software to scale out the modeling, simulations and other analysis of this data to make complex data driven decisions in real-time. The company has standardized on IBM Spectrum LSF to run its HPC systems, increasing throughput for simulations from 30-50%. Other IBM solutions used include IBM Spectrum Scale, IBM Spectrum Symphony, IBM Elastic Storage Server on Power and IBM Spectrum Protect.

- A *Life Sciences* customer uses IBM Spectrum LSF (with support for AWS bursting and Docker containers), and a multi-tier storage system – IBM Spectrum Scale, IBM Cloud Object Storage and IBM FlashSystem – with over 60PB of storage to enable large-scale genomics and other bioinformatics research on over 10,000 compute cores. This solution is able to accommodate a range of users inside and outside the organization with a highly available, cost efficient infrastructure that supports upwards of 500,000 short- and long-running jobs daily, enabling queries to be run locally or remotely that access up to one (1) million files of all sizes.

- An *Oil & Gas* client has replaced an on-premise, high cost HPC environment with IBM Spectrum LSF and IBM High Performance Cloud to support 6 sites across 3 continents doing reservoir simulations for exploration and extraction, and simulation software dev/test and using from 400 to 1700 cores and close to 200TB of storage. Results are obtained more quickly and the solution scales easily with demand.

- A *Financial Services* customer uses IBM Conductor with Spark, IBM Spectrum Symphony, and IBM Spectrum Scale for Spark- and non-Spark based analytics for credit risk analysis, trade surveillance/analytics, market risk reporting and forecasting analytics. The IBM software-based system optimizes 20 (and growing) Spark applications and more

than 300 other analytical applications across a hybrid cloud environment with over 500TB of shared storage and a distributed computing grid of 5,000 servers with over 100,000 cores.

> *Businesses in a range of industries are using IBM Spectrum Computing intelligent resource and workload management for compute- and data-intensive applications such as analytics, simulations and cloud-native applications built on Apache Spark and Hadoop. Optimized for distributed scale-out shared computing environments on-premise and/or on-the-cloud, policy-based automation optimizes time-to-results, minimizes cost, and maximizes utilization of compute resources.*

### *Summary Observations*

A shift in workloads from traditional  workloads to include more compute- and data-intensive workloads (genomics, simulations, what/if analysis, cognitive), and next generation Big Data and born-in-the cloud applications such as NoSQL, Spark, and Docker are driving IT executives to evaluate new storage and software defined infrastructure options that provide better IT performance, scalability, and agility at a lower cost.  In addition, these executives are grappling with rationalizing which workloads belong on-premise – and which workloads can be shifted to low-cost public cloud storage.

To address traditional storage requirements while addressing the new generation of compute- and/or data-intensive applications, IBM has revamped its storage line to include a complete range of storage solutions (including storage software, services and a wide range of storage hardware options such as all-Flash arrays).

IBM has also honed its software defined infrastructure strategy, now focusing heavily on enabling software defined computing on the hybrid cloud – via IBM Spectrum Computing – in order to enable customers to easily support the range of workloads from traditional on-premise workloads to the new-generation of cloud-native applications on the IBM cloud or other popular public clouds. IBM Spectrum Storage including IBM Cloud Object Storage provides consistent data management across on-premise, dedicated and public cloud, giving customers choices based on security, cost and access requirements.

Last but not least, IBM's suite of software defined infrastructure offerings provides flexibility, enabling customers to use commodity storage or storage from other vendors.

> *At first glance, the IBM Storage portfolio can seem overwhelming. But Clabby Analytics sees it a different way. Businesses can address requirements from within the IBM portfolio, so there is no need to use different vendors, which actually reduces complexity associated with ordering, managing and support. And IBM solutions are easily integrated with existing infrastructure, including other vendor's storage and popular public clouds such as Amazon and Azure. Finally, customers benefit from using IBM products in conjunction with one-another, supporting a wide range of use cases to solve specific problems.*

---