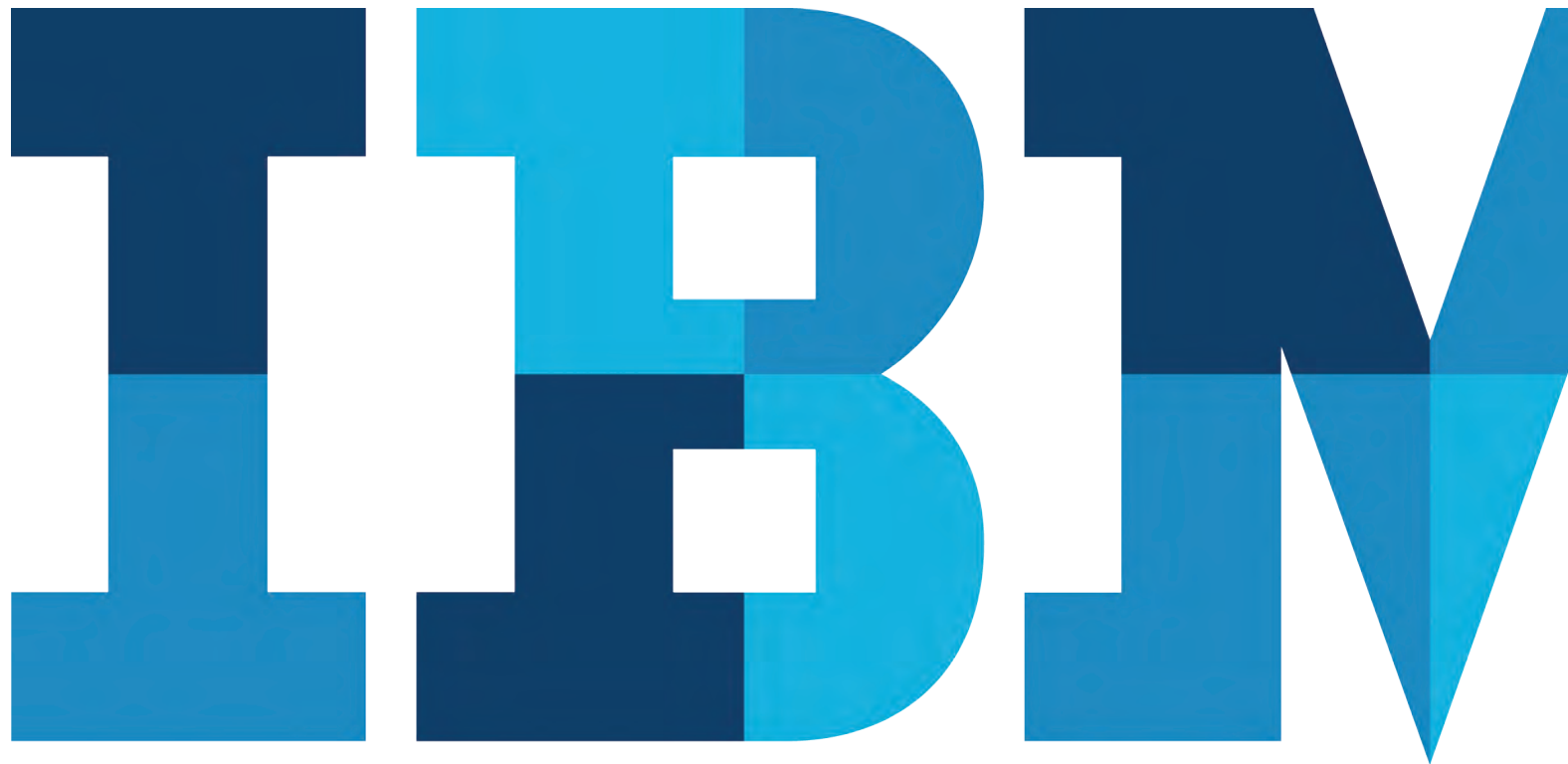# Building a strong storage foundation for big data and analytics

Enhance storage access, speed and availability

IBM

IBM

A confluence of technologies

Organizations are increasingly looking for ways to derive value from the tremendous volumes of data available today. In the past, they've focused solely on the structured data in databases and data warehouses.

But organizations now want to mine unstructured data as well, including data collected from new sources (such as social media and mobile applications) and traditional sources (such as email, supply chain and transactional systems). Fortunately, new technologies enable them to collect, manage and analyze both structured and unstructured data together.

As these organizations bolster their infrastructure to support big data solutions for better business decision-making, storage must be a key consideration. They need to make sure they can not only store large data volumes but also quickly and reliably access that data and make sense of it. Building a robust storage foundation is essential for maximizing their big data and analytics investment.

Why your customers need a big data and analytics infrastructure

Organizations have been collecting digital data for decades. To extract value from that data, they must explore and analyze it. As a result, 48 percent of organizations plan to allocate more IT spending for analytics.[1]
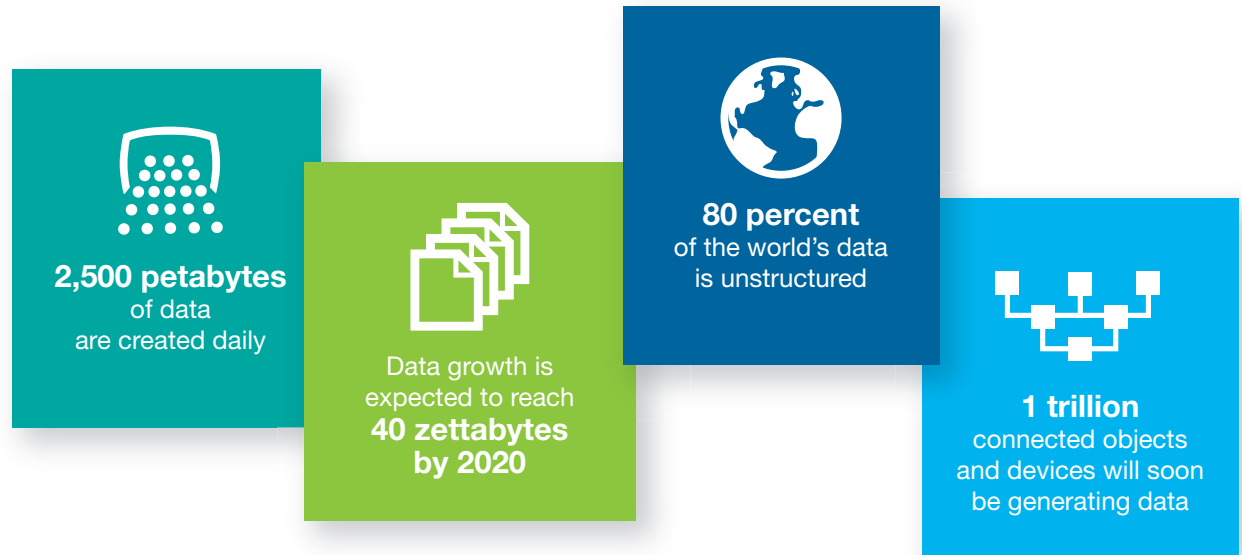
According to the IBM Institute for Business Value, 75 percent of business leaders believe analytics can boost innovation and generate revenue (see Figure 1). Yet according to the CMO Study of the IBM Institute for Business Value, 82 percent of CMOs feel underprepared to deal with the data explosion and only 20 percent have set up social networks for the purpose of engaging with customers.[2]

25%

Value outcomes

75%

■ Generates revenue/innovation
■ Controls costs/risks
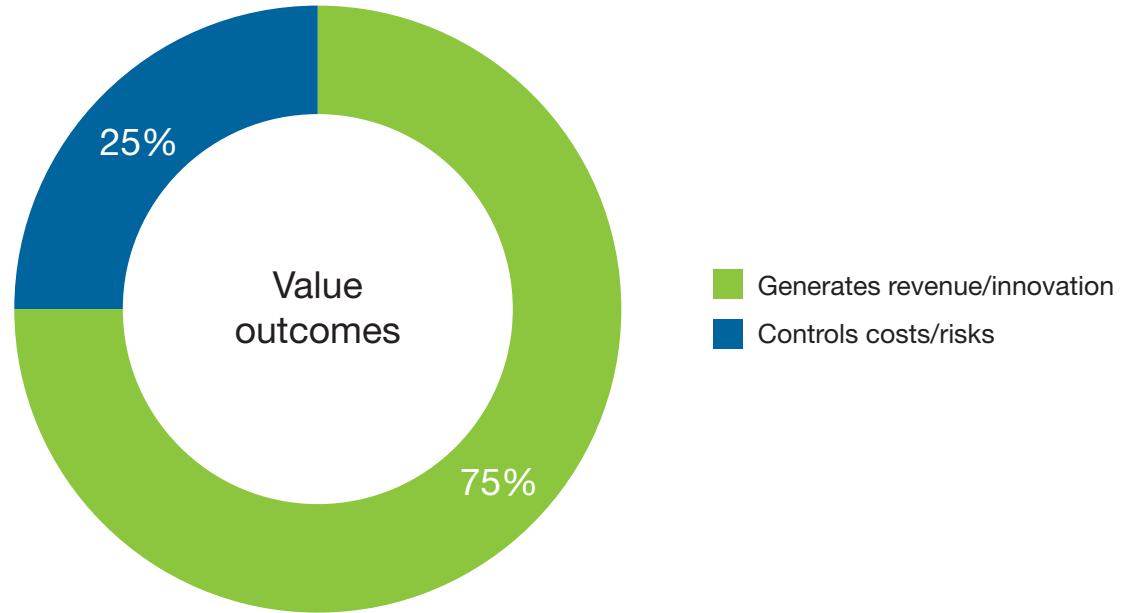
Note: Leaders were asked to describe the source of value derived from analytics. Consolidated response options shown; n = 156.
Source: IBM Institute for Business Value, "Analytics: A blueprint for value," 2013 Big Data and Analytics Study.

Figure 1. Value outcomes from big data and analytics.

IBM

There is no doubt that the biggest challenge organizations now face is determining how to harness the data explosion for greater competitive advantage. According to the IBM Institute for Business Value, the figures are staggering (see Figure 2). The last statistic speaks of even more challenges. As organizations begin to collect data from a larger variety of devices, such as the sensors that constitute part of the Internet of Things, data analysis will become more of a requirement.

2,500 petabytes
of data
are created daily

Data growth is
expected to reach
**40 zettabytes
by 2020**

80 percent
of the world's data
is unstructured

1 trillion
connected objects
and devices will soon
be generating data

Sources: Fidelity, "21st Century Investment Themes, Episode 5: Big Data," 2014 and IBM 2013 Annual Report, "What will we make of this moment?"

Figure 2. Trends contributing to the growth of big data.

Analysts from Gartner Research have argued, "Analytics will take center stage as the volume of data generated by embedded systems increases and vast pools of structured and unstructured data inside and outside the enterprise are analyzed. Organizations need to manage how best to filter the huge amounts of data coming from various sources and then deliver exactly the right information to the right person, at the right time. Analytics will become deeply, but invisibly embedded everywhere."[3]

Research firm IDC has also come to the same conclusion: In 2015, its analysts wrote, "Businesses [are] increasingly basing their decisions on data. This changing paradigm in which data creation and consumption govern business agility has given rise to a set of newer applications that in turn have forced the adoption of newer storage architectures and delivery models. Architectures and delivery models such as software-defined storage (SDS), hyper-converged systems, and intelligent data management platforms are inherently disruptive to traditional IT infrastructure."[4]

The takeaway is clear: analytics has evolved from business initiative to business imperative. Organizations must understand how big data and analytics can help them:

• Acquire, grow and retain customers
• Create new business models
• Transform financial and management processes
• Manage risk
• Optimize operations; counter fraud and threats
• Maximize trust, ensure insight and improve IT economics

Organizations must have the right storage infrastructure in place, one that will interact with their most valuable asset—their data. In a perfect world, data would move to the fastest storage for analysis, and then retreat to lower-cost storage when not in use. IBM uses policy engines and analytics-driven data management to keep data in the right place automatically, based on usage, and move data between storage systems without disrupting users or applications. **The result:** clients can run their big data and analytics projects and environments with much-faster performance at a much-lower total cost.

This infrastructure based on IBM storage will help employees make informed, useful decisions across the business, especially by gaining insights quickly to remain competitive and make in-the-moment actions based on this data. But besides storing data, IBM storage solutions will also manage, secure and protect organizational data no matter what kind of data it is and where it is stored.

There's some urgency to this effort: if the competition figures out the key to data analytics first, organizations will be at a disadvantage. According to a report from the IBM Center for Applied Insights, companies that effectively leverage data outperform their competitors by 2.6 times.[5]

Storage creates a crucial foundation for big data and analytics

What are the key components for building a storage foundation for big data and analytics? Access, speed and availability matter.

Access. Storage must be able to provide shared, secure access to all relevant information, regardless of data formats, sources and storage location. This access will help organizations achieve new levels of customer and operational visibility. For optimal access to big data and analytics workloads, the storage infrastructure must have scalability, fast and efficient data storage and access, data tiering and compression.

Speed. Advanced analytics isn't useful if it cannot deliver insights quickly. Storage must accelerate insights in real time to optimize decisions and capitalize on opportunities when they're most viable. With integrated, high-performance infrastructure, organizations can make decisions in real time (or near-real time) by embedding intelligence into operational processes and transactions. For maximum speed of big data and analytics workloads, a storage infrastructure must have not only low-latency capabilities, but the ability to scale up, in and out rapidly. The source of analytics must also be close to the systems of record.

Availability. Storage must consistently and securely deliver insights to the people and processes that require them. It must make data available to the right people, when and where they need it. With sufficient availability, organizations can empower employees with insights to improve collaboration, solve problems and capitalize on opportunities. For optimal availability of big data and analytics workloads, a storage infrastructure must have self-healing capabilities, enterprise-grade continuous infrastructure management, and single-site and multisite clustering solutions.

8

Storage for big data and analytics must also bring together systems of record and systems of engagement (see Figure 3). Systems of record benefit from simplified infrastructure. They require cost-efficiency through improved virtualization and automation, and drive controlled data growth. Systems of engagement, on the other hand, require massive scale and rapid pace so they can accelerate business insights. They also rely on data elasticity, including supporting diverse hardware.
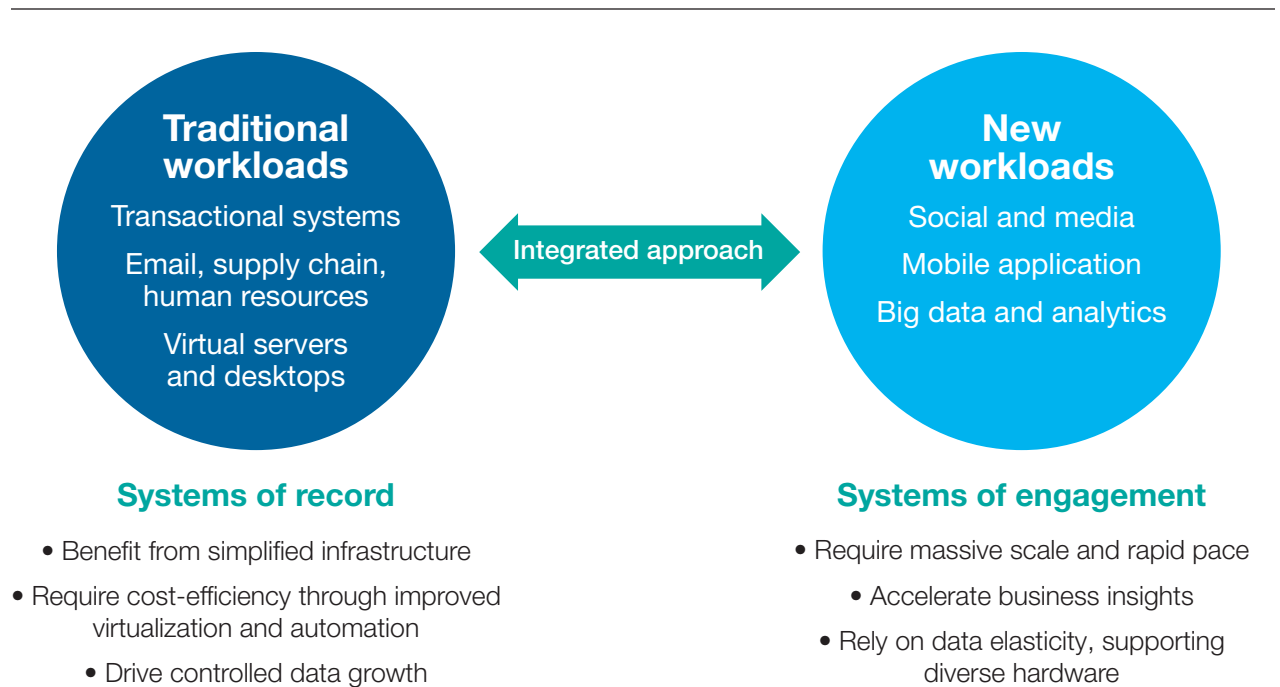
Traditional workloads

Transactional systems

Email, supply chain, human resources

Virtual servers and desktops

Integrated approach

New workloads

Social and media

Mobile application

Big data and analytics

Systems of record

- Benefit from simplified infrastructure
- Require cost-efficiency through improved virtualization and automation
- Drive controlled data growth

Systems of engagement

- Require massive scale and rapid pace
- Accelerate business insights
- Rely on data elasticity, supporting diverse hardware

Figure 3. Connecting systems of record and systems of engagement.

9

Key stories to tell about storage

As a channel partner specializing in storage, you know the advancements in storage technology intimately, but you also know the connection between storage and data. You should ask your customers whether they're getting the insights they need from their data, or if they are just acquiring more data. That's why analytics is so important.

According to a recent study on analytics from the IBM Institute for Business Value, 63 percent of organizations see a positive return on analytics investments within a year. Another 69 percent of speed-driven analytics organizations created a significant, positive impact on their business through real-time IT infrastructure. And perhaps most important,

74 percent of respondents anticipate C-level expectations for new, data-driven insights will continue to increase.[6]

Many IBM customers are already capitalizing on storage systems for big data and analytics (see Figure 4).
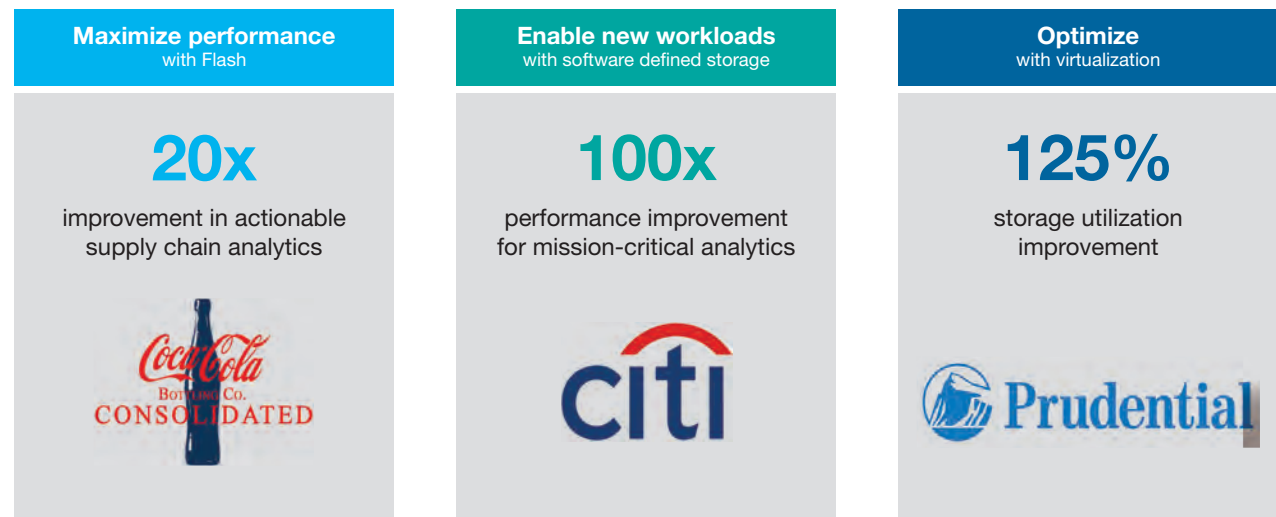
Maximize performance with Flash	Enable new workloads with software defined storage	Optimize with virtualization
20x improvement in actionable supply chain analytics	**100x** performance improvement for mission-critical analytics	**125%** storage utilization improvement

Figure 4. Maximizing and optimizing performance and redefining data economics.

Consider these examples:

- Coca-Cola reported an improvement of 20 times in analytics, batch processing by 4 times.
- Citibank experienced a hundredfold increase in mission-critical analytical performance.
- Prudential improved its storage utilization by 125 percent through virtualization.
- Vestas has reduced the time it takes to analyze petabytes of structured and unstructured data to identify wind turbine installation sites by 97 percent.
- Royal Caribbean saved USD77,000 in annual costs using real-time business analytics, enabling multiple marketing campaigns to attract and retain customers through improved quality of service.

- Memorial Hermann Health System speeds access to patient information for 5,000 physicians at 12 hospitals and more than 100 outpatient locations to improve the level and quality of care each patient receives.
- The John F. Kennedy Center for the Performing Arts achieved a 90 percent reduction in storage costs per terabyte over three years using tape; reduced administration workload; and cut operational costs through automation.
- Allianz Australia eliminates the challenges of multiple backup environments with IBM Spectrum Protect™ for virtual environments and improves recovery time up to 90 percent.

IBM contributes to a new approach in storage solutions

IBM offers a broad product portfolio of software defined storage systems that can help organizations create a robust storage foundation for big data and analytics (see Figure 5). With IBM storage solutions, these organizations gain the agility and scalability to handle ever-growing data volumes. In the following five categories you'll see how the full range of IBM storage solutions address significant key big data and analytics requirements.

Data acceleration for agility and scalability. IBM® Spectrum Scale™ provides virtually unlimited scalable performance to manage big data—both structured and unstructured. Spectrum Scale uses an automatic policy-based storage tiering that moves data from flash through disk to tape

and cloud tiers to help accelerate analytics of new workloads, allocating key data to the highest-performing tiers and lowering costs

by moving "cold" data to lower-cost tiers. Spectrum Scale manages big data and analytics environments such as Hadoop

IBM Spectrum Storage Solutions Family	Based on technology from …
IBM Spectrum Control	Tivoli® Storage Productivity Center (TPC) and management layer of Virtual Storage Center (VSC)
IBM Spectrum Protect	Tivoli Storage Manager (TSM)
IBM Spectrum Archive	Linear Tape File System (LTFS)
IBM Spectrum Virtualize	SAN Volume Controller (SVC)
IBM Spectrum Accelerate	Software from XIV System
IBM Spectrum Scale	IBM Elastic Storage (based on IBM General Parallel File System [GPFS™] technology)

Figure 5. Family of software defined storage systems.

as a replacement for HDFS and is used for High Performance Computing for file and object or Technical Computing workloads.

IBM FlashSystem® delivers enhanced performance using solid-state disk technology for real-time analytical insights that enable in-the-moment actions. IBM FlashCore™ technology offers real-time analytical insight with up to 50 times better performance than enterprise disk systems.[7] FlashSystem offers scalable performance enabling scale up and out capabilities to address the widest range of big data workloads. FlashSystem enclosure with Spectrum Scale accelerates the performance of analytics clusters to make faster data-driven decisions.

Performance and availability for mainframe data environments. The integration between the DS8870 and IBM z Systems™ delivers accelerated analytics, 24x7 access to data and optimized data economics for mission-critical environments. Integrating IBM DB2® and specialty analytics "engines" leveraging DS8870 delivers three times better performance with 24x7 data reliability, security and agility.[8] Organizations can also optimize performance with all-flash and hybrid-flash configurations for fast transaction processing and real-time operational analytics.

Data workload diversity and flexibility. IBM XIV® delivers predictable performance that scales linearly without hot spots, delivering insights from analytics faster with tuning-free

data distribution. XIV helps optimize utilization, streamline management through automation and achieve high levels of data availability with enterprise resiliency capabilities.

At the same time, the IBM Storwize® family is designed to meet analytical demands, with the IBM Storwize V7000 Gen 2 improving performance up to two times for faster business decision making.[9] It optimizes traditional workloads with IBM Spectrum Virtualize™ software using IBM Real-time Compression™ to improve storage utilization and lower costs.

Data protection, data recovery and data retention. IBM Spectrum Protect offers advanced data protection for big data and analytics workloads. Spectrum Protect

13

provides a scalable infrastructure to protect and recover all types of data, including files, objects and databases hosted on-premises and in the cloud. Spectrum Protect gives data owners the choice of snapshots, incremental backups and policy-based space management, which helps control costs compared to one-size-fits-all solutions. Backup and recovery times can be reduced from hours to minutes for critical data.[10] Spectrum Protect offers copy management, database cloning and built-in efficiency for big data and analytics workloads. Organizations can reduce backup infrastructure costs by up to 38 percent by switching to Spectrum Protect.[11]

IBM Spectrum Archive™ Enterprise Edition can run any application designed for disk files on tape and can play a major role in reducing the cost of storage for data that does not need the access performance of primary disk. Spectrum Archive with tape can reduce TCO by up to 90 percent over disk for long-term retention of data at rest with a large, open-format tape repository.[12]

Data management. IBM Spectrum Control™ uses proprietary analytics from IBM Research to optimize data performance and cost across flash and disk storage tiers. Spectrum Control can reduce users' cost of storage by up to 50 percent.[13] It features end-to-end visibility of file, block and

object storage, including Spectrum Scale and Spectrum Virtualize software defined storage environments. The on-premises version integrates with Spectrum Virtualize to migrate data between IBM and non-IBM storage systems, without disrupting users or applications. Spectrum Control helps improve capacity planning, storage utilization and data placement. New **IBM Spectrum Control Storage Insights** is a hybrid cloud storage and data management solution that can be up and running in as little as 30 minutes, provides nearly immediate insights, and delivers additional analysis over time as it learns about the organizational environment.

14

Conclusion

When you bring these needs and capabilities together, the result is a compelling storage solution for your customers. Consider these advantages IBM offers:

Simplified storage management.
IBM Storage solutions support all applications, data types and hundreds of heterogeneous storage systems, including cloud-based systems.

Scalability with data anywhere. IBM solutions enable elastic scalability with high performance for new analytics, big data, and social and mobile applications.

Improved data economics. IBM solutions leverage commodity hardware and intelligently move data to the right location at the right time, from flash for fast access, to tape and cloud for the lowest cost.

Openness. IBM supports industry standards, including OpenStack and Hadoop, to complement IBM innovations with ones from other providers and communities.

Speed. IBM storage solutions help you improve speed to data and gain new levels of meaningful visibility into data. They help accelerate business applications, reducing time to insight to help your customers make the data-driven decisions they need.

For more information
To learn more about IBM storage, big data and analytics solutions, visit **ibm.com**/ systems/storage/solutions/analytics

[1] Forrester Research, "Big Data Ups the Customer Analytics Game," February 2014, www.ibmbigdatahub.com/whitepaper/big-data-ups-customer-analytics-game

[2] IBM Institute for Business Value, "Stepping up to the challenge: CMO insights from the Global C-suite Study," March 2014, public.dhe.ibm.com/common/ssi/ecm/gb/en/gbe03593usen/GBE03593USEN.PDF

[3] Gartner, "Gartner Identifies the Top 10 Strategic Technology Trends for 2015," October 8, 2014, www.gartner.com/newsroom/id/2867917

[4] IDC, "IBM Spectrum Storage: Software-Defined Transformation for IBM Storage," February 2015, www.idc.com/getdoc.jsp?containerId=254453

[5] IBM Center for Applied Insights, "Outperforming in a data-rich, hyper-connected world," 2012, www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=XB&htmlfid=YTE03002USEN#loaded

[6] IBM Institute for Business Value, "Analytics: The speed advantage," 2014 Strategy and Analytics Study, ibm.com/services/us/gbs/thoughtleadership/2014analytics

[7] 50x performance improvement is based on customer experiences across various measures of performance. Individual customer experiences will vary based on data center architecture, workload and disk systems.

[8] Results of performance measurements conducted by the IBM Enterprise Storage performance team in Tucson, Arizona utilizing the IBM System Storage DS8870 with High-Performance Flash Enclosure. With the new High Performance Flash Enclosure the team was still able to measure up to 3.2x increase in IBM System z® database performance when compared to the same scenario with SSDs.

[9] IBM lab measurements, April 2014.

[10] Based on IBM internal testing. ibm.com/connections/blogs/tivolistorage/entry/backup_1000_virtual_machines_in_less_than_36_minutes7?lang=en_us and client feedback, including Franciscan Missionaries of Our Lady Health System, ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&appname=SWGE_SW_SW_USEN&htmlfid=SWC14089USEN&attachment=SWC14089USEN.PDF#loaded

[11] Average of individual customer Analysis Engine Reports from Butterfly Software, May 2013, n=450. The savings include cumulative 36-month hardware, hardware maintenance and electrical power savings. Excludes one-time TSM migration cost. ibm.com/software/products/sv/category/storage-software

[12] Three-year total cost of ownership comparison of IBM TS3500 Tape Library/IBM Spectrum Archive solution compared to IBM DS5100 disk storage solution using IBM Spectrum Control for data management.

[13] Based on IBM experience with its own implementations and on client feedback. See "IBM's Data: Managed and Protected," youtu.be/EBdr4yyoBhg