White Paper

Elastic Flash: Enabling Private Clouds and Analytics

By Mark Peters, Senior Analyst; and Monya Keane, Research Analyst

February 2015

This ESG White Paper was commissioned by IBM and is distributed under license from ESG.

© 2015 by The Enterprise Strategy Group, Inc. All Rights Reserved.

Contents

Introduction 3

IBM Wants to Empower Private Clouds and Analytics 3

IBM's Flash Storage Product Bundle—Pertinent Terminology 3

The Nature, Use, and Drivers Behind Analytics Applications 4

With Analytics, What You 'Run Your Cloud On' Matters 6

How Well Do Spectrum Scale and IBM FlashSystem Play Together? 7

The Bigger Truth 9

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of The Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.

© 2015 by The Enterprise Strategy Group, Inc. All Rights Reserved.

Introduction

Improved performance (in terms of reduced I/O latency and improved I/O throughput) is by far the most recognized and appreciated advantage that flash storage provides over mechanical disks. Within just two years after IT vendors started to offer contemporary, "enterprise-ready" flash-integrated arrays, more than a third of the organizations ESG surveyed were citing ultra-fast read/write performance as the biggest impetus driving their flash deployments. Nearly half of the potential adopters surveyed also named performance as the main reason their organizations were considering flash.[1]

The importance of performance has only grown since then, as more IT organizations realize that a spinning-disk architecture invented 60 years ago is insufficient to support a world in which:

- Data growth—and more importantly, our expectation to access components of that data quickly—hasn't abated and likely never will.
- Hardware and applications are virtualized and are changing frequently and rapidly.
- Applications need more speed and I/O to meet the demands of our mobile, distributed, always-on, information-hungry world.
- IT environments are distributed, often extending to private and public clouds. But regardless of geography, time zone, or physical environment, end-users still expect instant results.
- Big data analytics initiatives are in the limelight. Decision makers need affordable tools at their disposal that are able to provide timely support for their business-critical analytics.

The true challenge centers on getting performance *with* capacity while staying within IT budgets.

IBM Wants to Empower Private Clouds and Analytics

IBM has a comprehensive approach to high performance that it believes will give tremendous support to organizations that are pursuing demanding private cloud and/or analytics initiatives. An outcome of years of refinement to IBM's advanced, all-flash storage offerings and global data management capabilities, the approach combines IBM Spectrum Scale with IBM's FlashSystem offerings to meet cloud and analytics environments' demands for scale, economy, geographic dispersion, *and* performance.

IBM Spectrum Scale, based on the IBM GPFS filesystem software (see box) offers elasticity of scale, performance, placement, geographic location, and implementation style. It is the realization of something IT organizations have long desired—a better way to get the right data in the right place at the right time. Traditionally, that goal could be achieved only within numerous storage boxes inside the walls of individual data centers, if at all.

Spectrum Scale takes the "right data, right place, right time" concept and adds "right cost," "right performance," and a whole lot of other "rights." It supports data distribution to the cloud *and* to on-premises assets, including heterogeneous storage arrays with varying performance tiers. (Of course, all those "rights" may sound a bit suspect to a user community that has resigned itself to a traditional, reactive, largely manual approach requiring frequent tradeoffs in capacity, speed, affordability, and ease of shared access.)

IBM's Flash Storage Product Bundle—Pertinent Terminology

Here is a brief description of IBM Spectrum Scale and two associated elements of IBM's flash storage product bundle—IBM FlashSystem and IBM Elastic Flash:

A Highly Credible Foundation: GPFS

There is considerable value in the underlying, well-proven IBM GPFS, which is the foundation of the Spectrum Scale approach. GPFS makes IBM Elastic Flash highly credible because it has been around for years and has proven successful in supporting thousands of real-world, mission-critical deployments.

[1] Source: ESG Research Report, *Solid-state Storage Market Trends*, November 2011.

© 2015 by The Enterprise Strategy Group, Inc. All Rights Reserved.

- **IBM Spectrum Scale:** Spectrum Scale is an approach to software-defined storage based on a highly credible foundation: IBM's well-known, field-proven General Parallel File System (GPFS) software.

  IBM envisioned the Spectrum Scale software architecture as a way to give users access to any type of data on any storage device, IBM or not, anywhere in the world. It also includes tools for task automation that provides dynamic tiering (for example) to place data on the best performance and cost tier at the right time. Spectrum Scale virtualizes storage so that multiple systems and applications can share common data pools as IT manages it all under a single global namespace. An organization could implement Spectrum Scale to boost its geo-distributed capabilities and manage its resources more dynamically across a range of physical media types and data protocols.

- **IBM FlashSystem:** As specific workloads are tackled, the FlashSystem family take center stage. These all-flash storage systems make applications and data centers faster and more efficient by offering extremely low data-access latency (which the vendor calls "IBM MicroLatency") to bolster OLTP environments, analytics efforts, virtual desktop infrastructures, and cloud deployments.

- **IBM Elastic Flash:** Together, IBM Spectrum Scale and IBM FlashSystem create what one might call an "elastic flash" storage environment, which IBM is branding as IBM Elastic Flash.

  Although Spectrum Scale and FlashSystem are both available by themselves and there's no requirement to combine them, doing so can result in the IT equivalent of "1+1=3." Each solution complements the other in supporting the needs of two of today's most common, fast-growing "application ecosystems"—private clouds and [big] data analytics. Both of those application ecosystems must deliver extreme capacity and extreme performance simultaneously *and* economically. If only one or two of the scale/speed/economy attributes is needed, then each of the IBM products can stand alone very well. But where all three are prerequisites, the full IBM solution shines.

> The addition of FlashSystem to a Spectrum Scale deployment can enable users to more effectively, efficiently, and economically achieve the business benefits of analytics.

It is worth remembering that *the physical hardware comprising a private cloud environment matters greatly.* Hardware is just as important as software in both analytics and cloud environments.

The Nature, Use, and Drivers Behind Analytics Applications

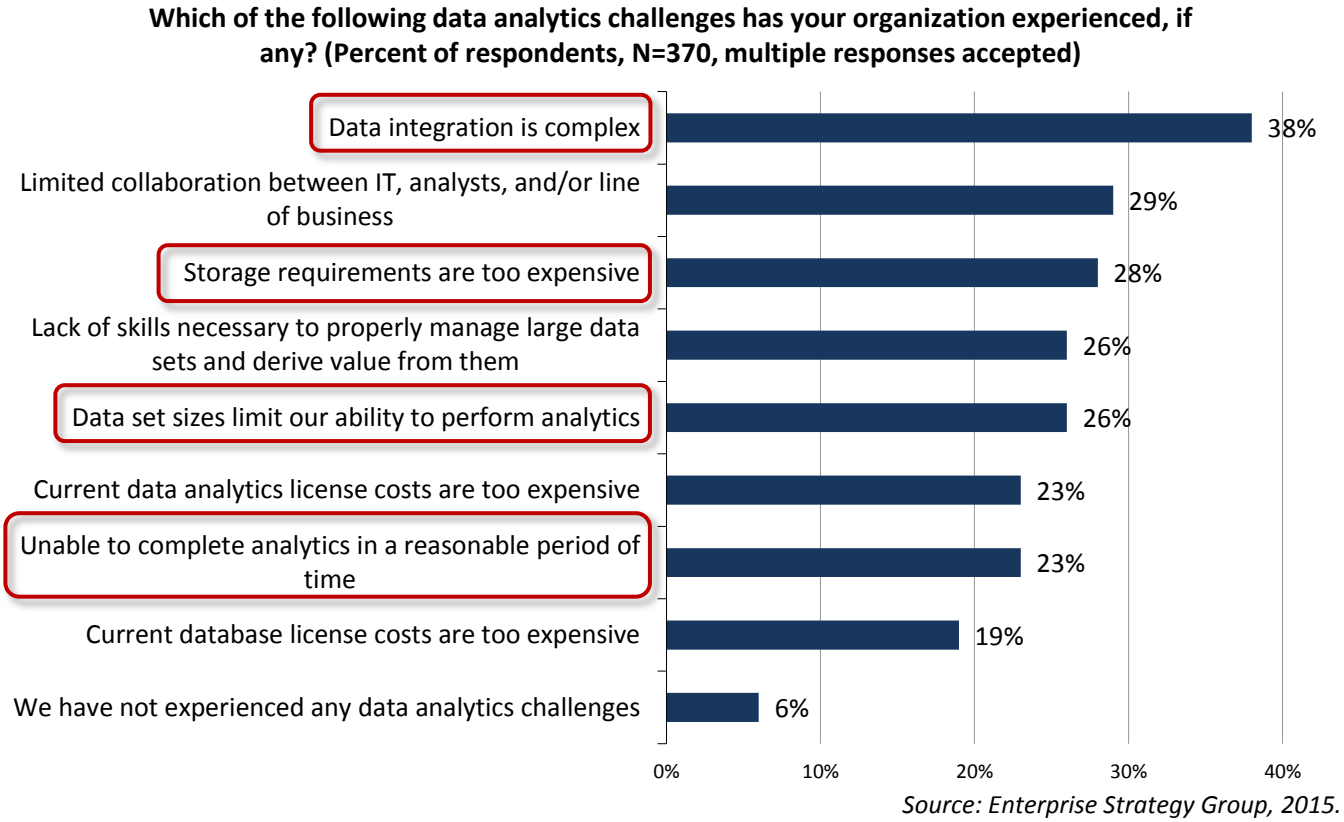
How many organizations "do" analytics? Basically, all of them do, even if the output just takes the form of an Excel spreadsheet. Companies analyze and manipulate their data for many reasons—to optimize their supply chains, understand their buyers' behavior, or otherwise strive for a competitive advantage. Unfortunately, storage (and storage performance) limitations can constrain their efforts.

Figure 1 shows the results of ESG research about the types of infrastructure-related concerns and roadblocks that hinder enterprises from moving forward with data analytics initiatives.[2] They are the kinds of obstacles that disappear when sufficient data is available economically and can be processed easily and rapidly. Of course, it sounds easy to make that claim on paper. Yet that quest has occupied IT professionals for decades and has sometimes cost their organizations hugely in terms of direct expenditures and opportunity costs.

[2] Source: ESG Research Report, *Enterprise Big Data, Business Intelligence, and Analytics Trends*, January 2015. All research results in this white paper are taken from that report unless otherwise noted.

© 2015 by The Enterprise Strategy Group, Inc. All Rights Reserved.

Figure 1. What Are the Biggest Data Analytics Challenges?



Source: Enterprise Strategy Group, 2015.

The same ESG research provides other insights on the need to deliver that elusive mix of data scale, speed, and economy. For instance, when ESG asked respondents, "What are the specific challenges your organization is facing with respect to its database size and/or growth?" (databases being a prime area for analytics, of course), the top responses were:

1. Performance degradation.
2. Increased storage capacity requirements.
3. Increased cost of infrastructure.

Those basic issues ranked ahead of such things as server capacity, software licenses, and governance. In simple terms, the storage world—whether in terms of performance, capacity, or cost—can truly be an anchor on analytics progress.

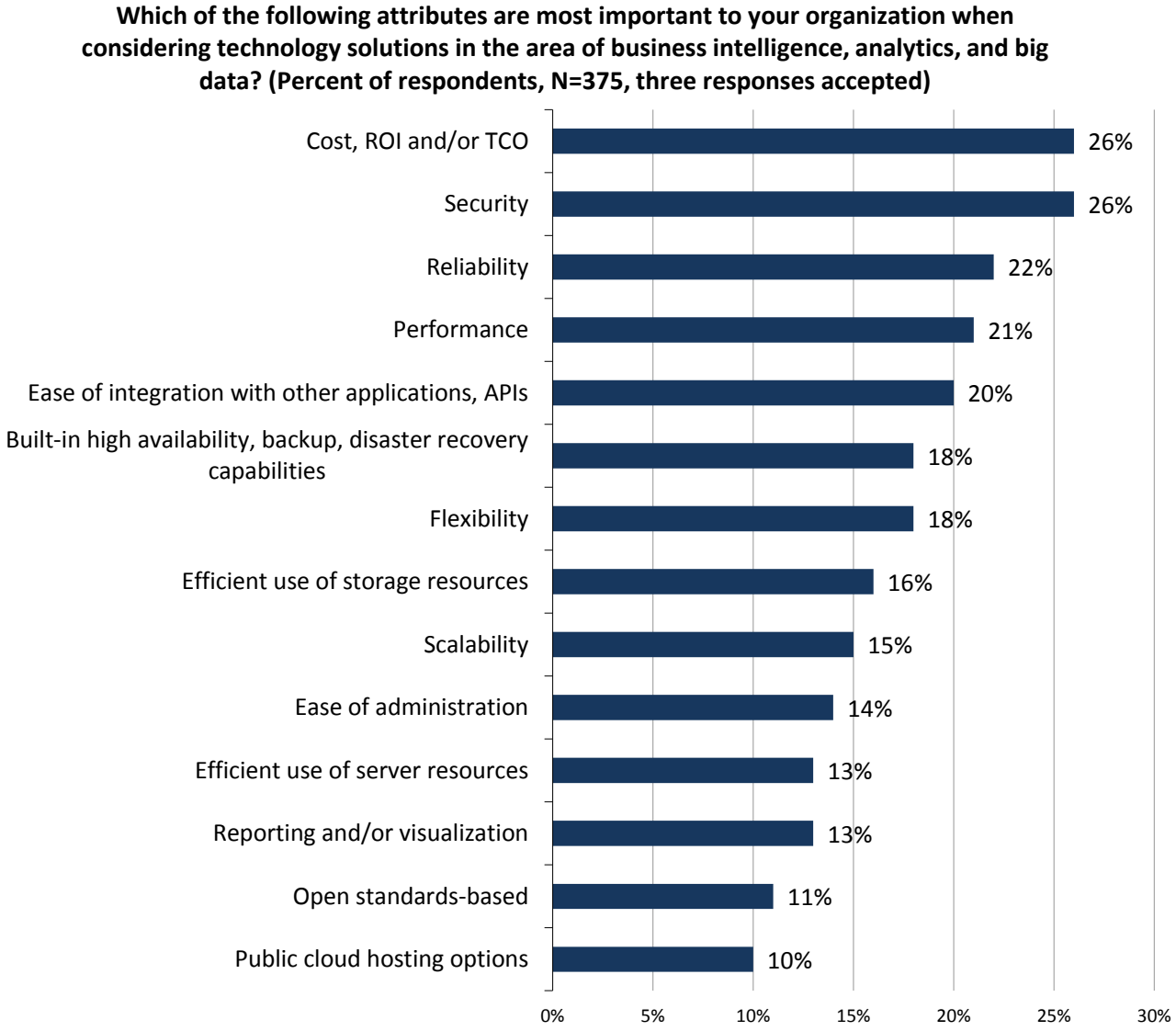
Furthermore, respondents were specifically asked about deployment models. (The question was, "In terms of net-new BI/analytics deployments, which of the following best describes the primary deployment strategy your organization will likely use going forward?") Their answers reflected a range of possibilities, from onsite (virtualized or not, or on purpose-built appliances) to public and hybrid clouds. Considering that data is the lifeblood of business, it makes sense to maintain organizational data in as secure, flexible, capable, and economic a manner as possible—i.e., a "corporate data cloud" with a software-defined storage solution such as IBM Spectrum Scale.

Storage—whether in terms of performance, capacity, or cost—can truly be an anchor on analytics progress.

Turning back to the technology needs for analytics, Figure 2 shows just how well the attributes of IBM Spectrum Scale with IBM FlashSystem match with the most important needs: Cost, security, reliability, and performance were the top four most-cited responses, and they all can be enhanced by the overall IBM data solution.

© 2015 by The Enterprise Strategy Group, Inc. All Rights Reserved.

Figure 2. What Matters Most in a Business Intelligence, Analytics, and Big Data Solution?



Source: Enterprise Strategy Group, 2015

With Analytics, What You 'Run Your Cloud On' Matters

Clouds—public or private—are not vapor; they are regular IT collections of hardware, software, and networking components that are simply consumed in a different fashion from the "direct-line, infrastructure-to-application" model that dominated IT for so long. And "software-defined storage" still requires physical storage infrastructure. Bottom line: The hardware that one chooses to deploy matters *a lot*. IBM has combined an approach tailored to cloud scale and flexibility (IBM Spectrum Scale) with technology perfect for supporting analytic speed and agility (the IBM FlashSystem family).

Although the approach and the storage family were designed for high-performance compute environments in general and not explicitly for analytics in particular, they are incredibly well-suited to supporting analytics use cases that help organizations *derive more value* from growing volumes of data. Spectrum Scale and FlashSystem work together to alleviate storage and data-access bottlenecks, providing the kind of performance a compute-intensive analytics application needs. After all, an analytics report that takes three weeks to produce isn't helpful. If you're going to do analytics right, you need speed.

A private cloud infrastructure is becoming a popular deployment model for geographically dispersed/system-dispersed analytics efforts, and Spectrum Scale supports geographically distributed compute/storage

© 2015 by The Enterprise Strategy Group, Inc. All Rights Reserved.

infrastructures. It offers granular storage management capabilities that *enable dynamic resource allocation*, something that's at the heart of the [private] cloud model.

Likewise, Spectrum Scale is perfectly designed for handling the file types that constitute the "raw material" used in many analytics efforts—unstructured file data, OLTP block data, and object data. Those dissimilar bits and bytes might be pulled in from all over the world, from a variety of array types, and from multiple groups/business units.

How Well Do Spectrum Scale and IBM FlashSystem Play Together?

Spectrum Scale provides the disparate data aggregation and management, and FlashSystem provides the speed. Together, they permit analytics to be conducted at a global level *and* at an immediate pace. You can have each without the other, but in a world with burgeoning data needing analysis, they work ideally in concert. That's true in part because Spectrum Scale knows a lot about the data it manages—for example:

- How "hot" is it? (i.e., how often has it been accessed?) Having such information means it can activate [IT controlled] policies that say, "This is hot data; move it to flash."
- Which groups or business units are accessing which data? Do certain groups have high priority and responsibilities in terms of keeping the business successful? If so, a policy can be invoked to move those groups' data to flash.

Basically, such actionable intelligence allows one to "tie the data" to a particular flash tier. And that, in turn, yet again makes IBM's approach ideal for supporting analytics; there may be particular analytics data that you really need to act on quickly, and IBM Elastic Flash enables—indeed protects—that effort.

As mentioned, certain requirements must be met for an organization to overcome the likely traditional barriers standing in the way of moving forward with private cloud deployments that can also effectively support analytics applications; and Spectrum Scale *with* FlashSystem can address those potential roadblocks:

- FlashSystem can accelerate Spectrum Scale: Any analytics application—whether it's big data analytics or something as simple as a batch process or data archiving analytics—will benefit by being speeded up.
- FlashSystem can make Spectrum Scale deployments more cost effective. FlashSystem operates as one tier in a tiered Spectrum Scale storage environment. Spectrum Scale deployments are excellent at providing dynamic/automated tiering of data, especially when leveraging FlashSystem to lower costs by making tiering more efficient.
- FlashSystem can store and dramatically accelerate Spectrum Scale metadata. As metadata are the "keys to the doors" of

Single Namespace—a Huge Differentiator

One especially attractive aspect of Spectrum Scale is that it offers a single namespace for everything. If all the data is in the same namespace, that means it's under the same management umbrella. And that advantage ties back to analytics.

"Efficient analytics" isn't just about speed; it's about being able to comb through lots of data looking for a nugget of insight. Disassociated systems in different parts of the world don't lend themselves to that effort.

The fact is, most enterprises nowadays have multiple physical locations, and they want to be able to move around their compute resources dynamically. Often, they want to do so in a private cloud rather than exposing data externally.

A bank with 50 locations could implement Spectrum Scale—and by doing so, energize analytics efforts both at headquarters and at the other 49 sites thanks to the single namespace and Spectrum Scale's data caching and synchronization capabilities.

Spectrum Scale plays strongly in private cloud scenarios, clouds with actual storage arrays that hold the data to make the scenario work. That's where FlashSystem comes in, enabling rather than limiting the enormous capabilities of IBM Spectrum Scale.

© 2015 by The Enterprise Strategy Group, Inc. All Rights Reserved.

Spectrum Scale's broad data repository, having that metadata stored on FlashSystem ensures that access to that data repository (both literally and figuratively!) is *not* subject to the "throttling" that could result if the metadata were on a less high-performing storage media.

- FlashSystem can be used as a very large storage cache for active data sets.

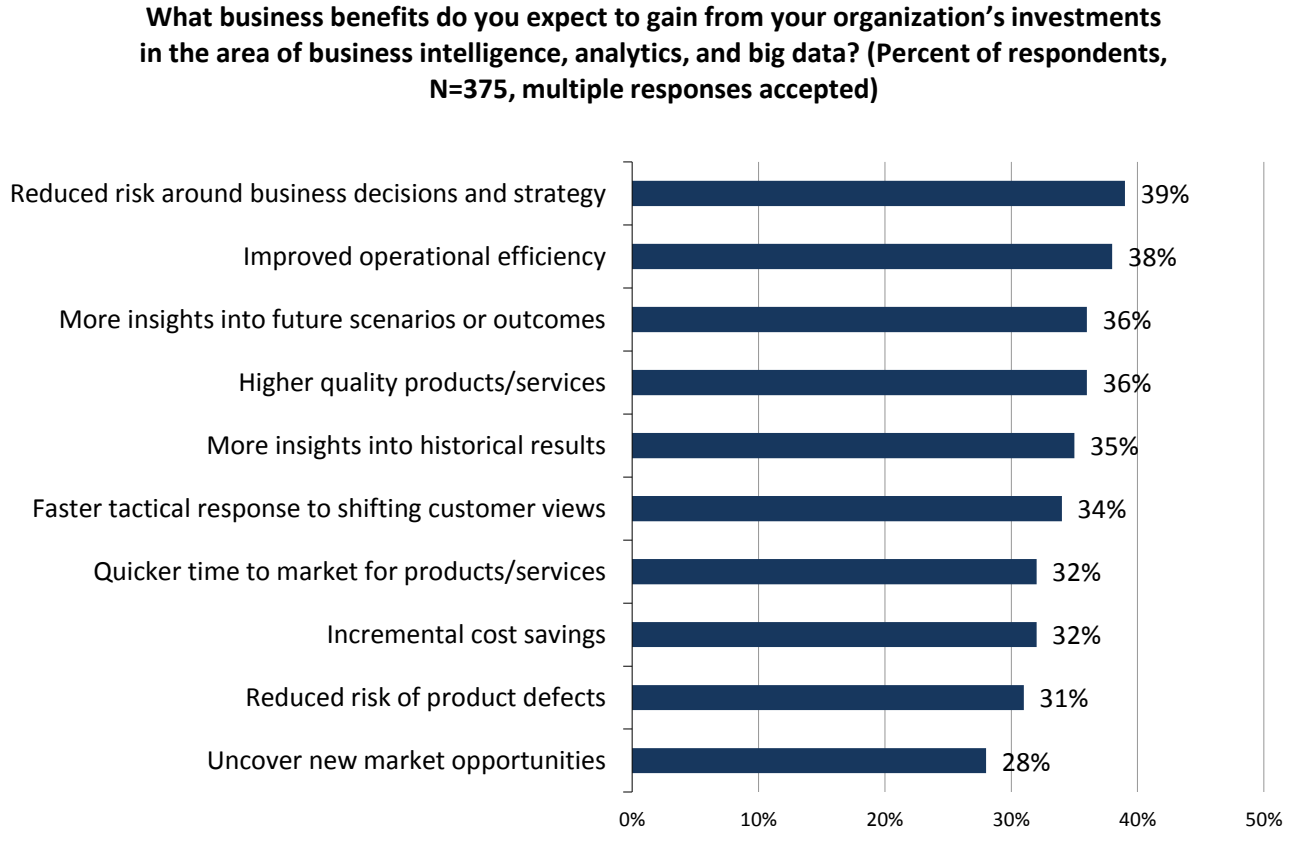
The result is that FlashSystem is adding the scalable power and performance (low-latency and high-bandwidth) that can turn analytics initiatives from failures into successes, without requiring the organization to overspend on other resources—i.e., buying disk, and then not using its capacity (short-stroking). You bring down your operational costs when you put your highly active data on flash and your less-active data elsewhere (i.e., disk or tape).

So in colloquial English, what impact can FlashSystem have in a solid, well-built private cloud (such as Spectrum Scale) to improve its analytics capabilities?

- More—and new—work can be done.
- Work can be done faster.
- Everything is made more efficient, lowering TCO and raising ROI.
- All workloads and users will get the data they need, when they need it, improving organizational insight.

In summary, the addition of FlashSystem to a Spectrum Scale deployment can enable users to more effectively, efficiently, and economically achieve the business benefits of analytics. Those benefits can be of significant organizational value, as Figure 3 demonstrates.[3]

Figure 3. The Benefits of Analytics



Source: Enterprise Strategy Group, 2015.

[3] Source: ESG Research Report, *Enterprise Data Analytics Trends*, May 2014.

© 2015 by The Enterprise Strategy Group, Inc. All Rights Reserved.

The Bigger Truth

If there's one thing we know about IBM, it's that the company has demonstrated time and again its ability to solve really, really big IT problems. Many other vendors make assertions that look great in a PowerPoint presentation, but can all those vendors genuinely support your operation in Tanzania, Tupelo, or Tyneside if it encounters a cloud connectivity glitch or otherwise needs help?

Spectrum Scale offers a proven, scalable private cloud data management system with the ability to use and integrate public clouds, too. IBM can provide "multi-dimensional elasticity," tying together any sort of data, anywhere in the environment or from places outside of the direct environment, and anywhere on the globe. When it is accelerated with IBM's FlashSystem (jointly, IBM Elastic Flash), greater value can be derived from analytics applications, spanning the latent knowledge and insights within any organization's "corporate data cloud."

Imagine accessing the world's biggest library using the world's best filesystem to find anything you want. You'd still need speed if you want your searching efforts to be worthwhile. That's where FlashSystem comes in. If you do *not* have a good card catalog (single namespace) and a fast way to extract information held in those ten billion books stacked in five million rooms worldwide, then you are left with a collection of ten billion paperweights.

IT thought leaders ruminate about moving from "data" to "information" to "insight" to "knowledge" to "wisdom." Without Spectrum Scale, that process will be restricted to one library or language. And without FlashSystem, any evolution from "data" to "wisdom" will take far too long to achieve. To put it another way, effective analytics are only possible when one has near-immediate access to all the information necessary to the analysis.

© 2015 by The Enterprise Strategy Group, Inc. All Rights Reserved.

Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel: 508.482.0188 Fax: 508.482.0218 | www.esg-global.com