

White Paper

How to Capture More Value from an Enterprise Analytics Solution

By Nik Rouda, Senior Analyst

January 2016

This ESG White Paper is jointly commissioned by Hewlett Packard Enterprise and Intel, and is distributed under license from ESG.



Contents

Finding Value in Enterprise Data 3

 Using Analytics to Meet Business Goals 3

 Big Data Insights Are Not Easy to Gather 4

Changing the Rules for Data Warehouses 5

What to Look for in a Hadoop-ready Big Data Solution 6

HPE, Intel, and Microsoft Deliver a Converged Solution for Modern Data Warehouses..... 6

The Bigger Truth 8

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.

Finding Value in Enterprise Data

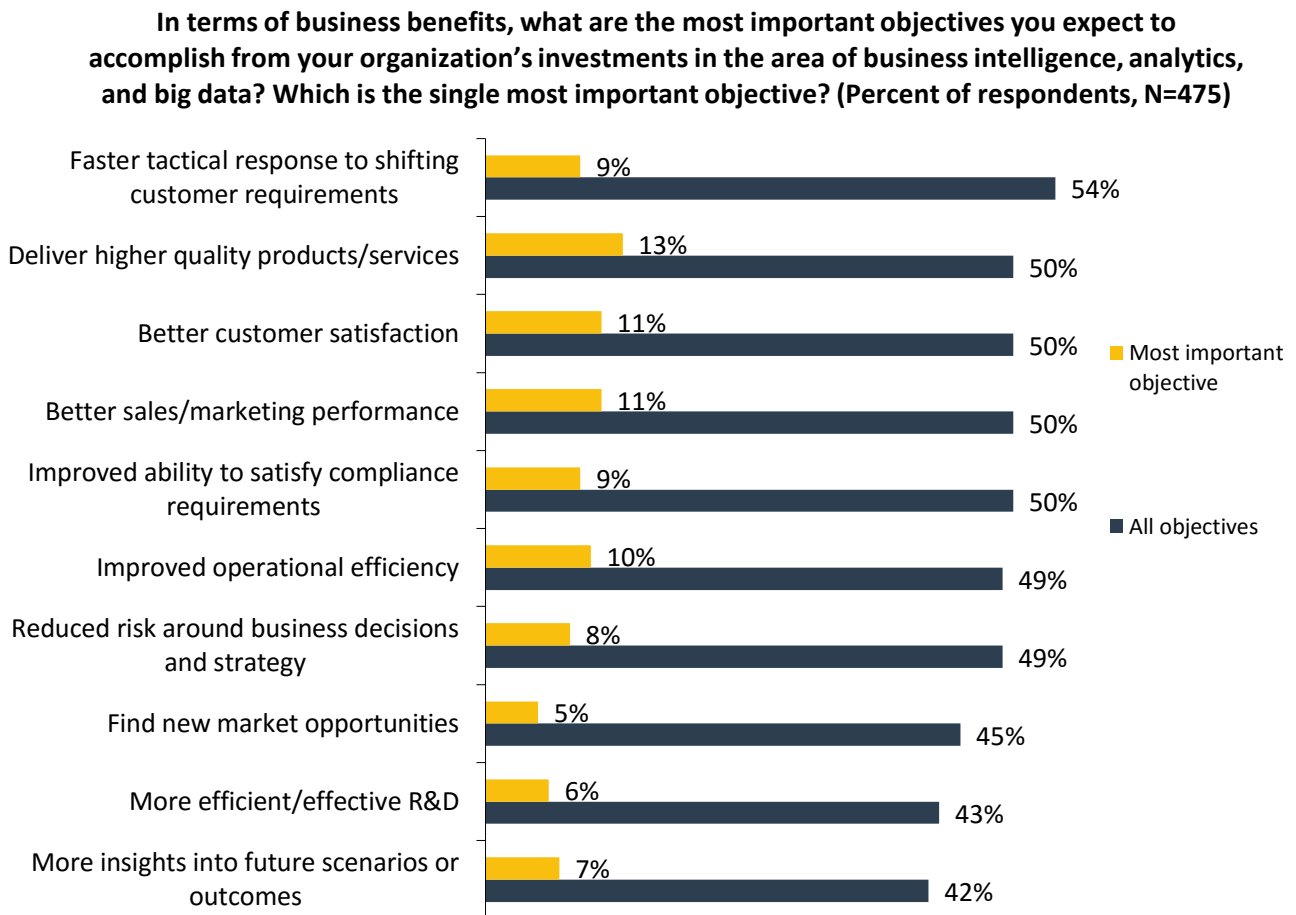
Modern enterprises are producing and gathering unprecedented quantities of data, which is a great opportunity—and a huge challenge. Within this data lie valuable business insights. There is plenty of potential to gain new customer insights, uncover operational efficiencies, and more closely monitor your business. However, old-school data warehouses often can't manage and analyze the ever-growing volume of data. New tools like Hadoop are exciting but unfamiliar to many in IT. Enterprise technology leaders Hewlett Packard Enterprise (HPE) and Microsoft have built a converged solution that promises to address this big data challenge.

Using Analytics to Meet Business Goals

All businesses generate data, but many businesses don't use their data effectively. Today's businesses are generating more data from traditional structured data, including customer information, financial records, and production system data. But the rise in unstructured data, such as those from emails, IoT sensors, video, and other sources, has driven much of the interest around analytics.

To thrive and gain competitive advantage, every business should be looking for ways to get more from the data it already has. There are many reasons to do so, ranging from operational efficiency gains to better product development. Other reasons include reducing risk and entering new markets, as seen in recent ESG research on businesses' expected accomplishments stemming from investments in the area of business intelligence, analytics, and big data (see Figure 1).¹

Figure 1. *Top Ten Expected Business Benefits from an Analytics Solution*



Source: Enterprise Strategy Group, 2016.

¹ Source: ESG Research Report, *Enterprise Big Data, BI, and Analytics Trends: Redux*, to be published.

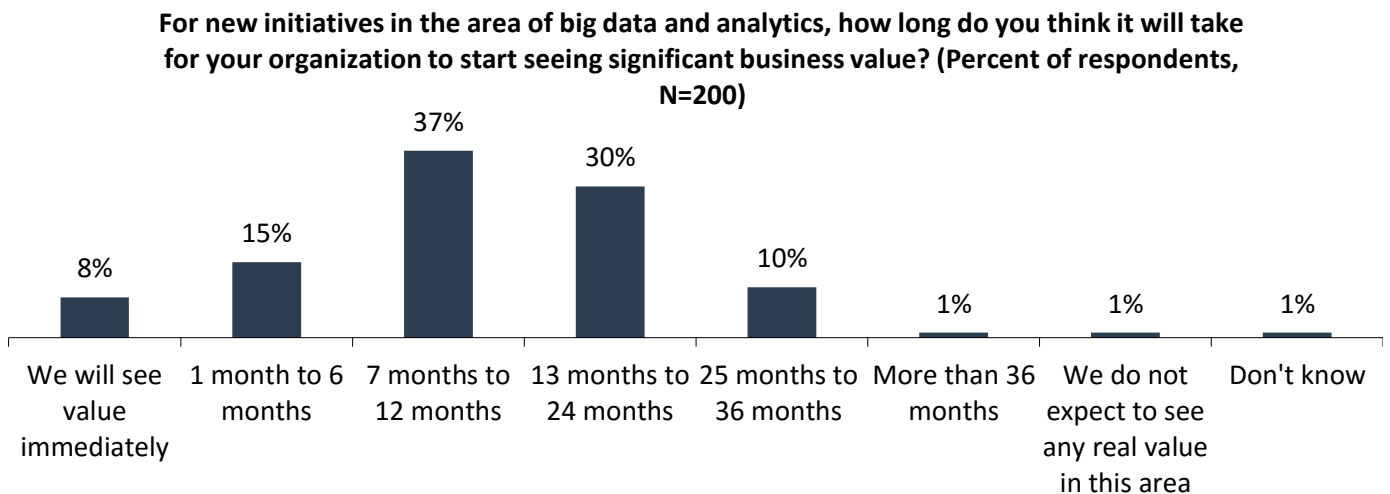
Big Data Insights Are Not Easy to Gather

These data insights can easily get lost in the shuffle without modern tools. Collected data is virtually worthless if the business doesn't analyze it. There are a few common challenges for today's enterprises around gathering data insights:

- No central system for analytics, just many fragmented sources of information and spreadsheets.
- Data warehouses are often expensive, unwieldy, and monolithic, without any of the necessary flexibility.
- Organizations have immature data lakes at this early stage of Hadoop, which can be hard to organize and leverage.

It's clear that data warehouses can provide actionable insights and business value, but they must be properly designed and executed. The requirements for a data warehouse include the need for scalability, something many warehouses aren't designed to accommodate. In looking to expand an existing data warehouse, you should also evaluate new platforms, not just retrofit existing systems to support new data lakes. Though there's a lot of potential for data warehouses and analytics, there's also plenty of skeptics in the market who wonder how long it will take businesses to see any real value from analytics solutions, as illustrated by Figure 2.²

Figure 2. Length of Time Organizations Predict It Will Take to See Significant Business Value from Big Data



Source: Enterprise Strategy Group, 2016.

To address this doubt around successful analytics initiatives, it's critical to find an approach that can bring the benefits of data analytics solutions without the flaws. A successful approach has to include both IT and business stakeholders, enabling collaboration to reap the benefits of analytics tools. IT teams need to be comfortable with the technology that's used, while line of business teams have to be able to use that technology. There are some key guiding principles to follow in planning and building an analytics platform:

- End-users should have direct access to data, and easy-to-use business intelligence (BI) tools to interpret the data, without fragmentation of knowledge or limits on scale.
- IT should have a central data warehouse for accessing "one source of truth" without added cost and delay.
- Businesses should be able to incorporate lots more unstructured data in Hadoop, working toward eliminating any uncertainty or unfamiliarity.

² Source: ESG Research Report, [Enterprise Data Analytics Trends](#), May 2014.

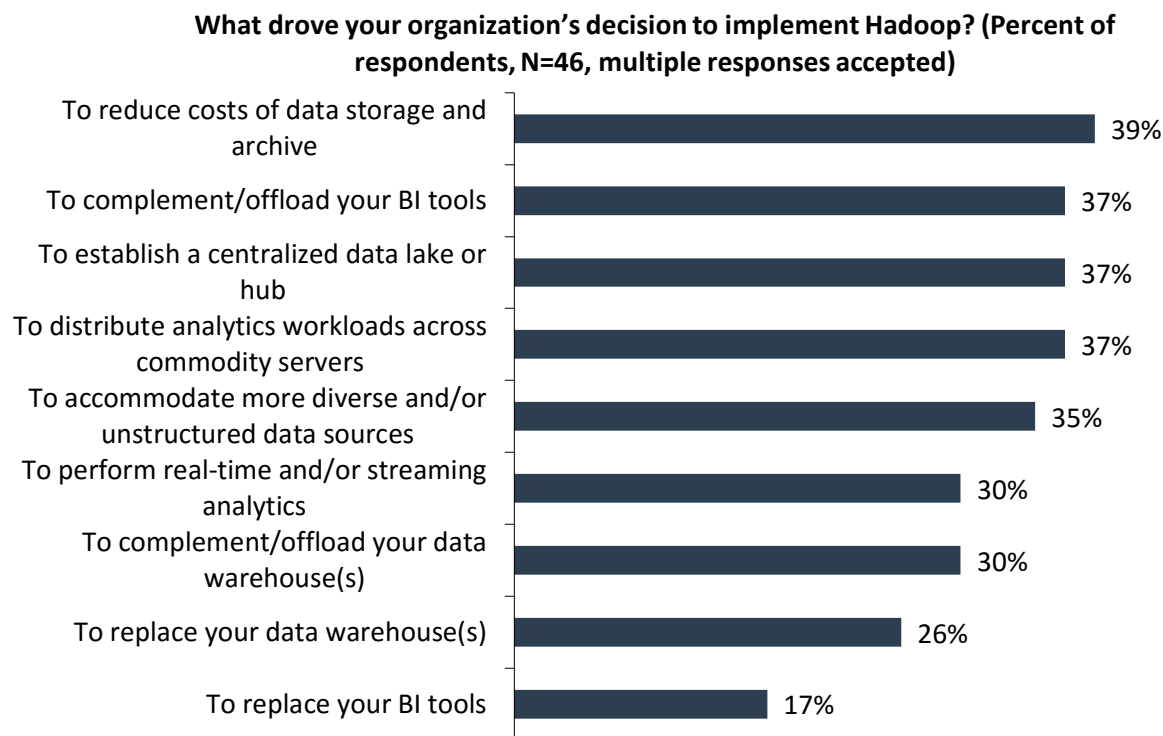
Changing the Rules for Data Warehouses

Succeeding at big data analytics often requires changes in mindset along with changes in technology. For an organization to be more successful, leaders may need to rethink some common assumptions and practices around centralized data analytics repositories. The effort involved in assembling and supporting a do-it-yourself big data approach can be a huge burden for IT, especially coupled with the cost of a traditional data warehouse.

IT teams and business leaders should also consider how users will access and deploy big data tools and insights. Many users still like spreadsheets for their data. They are a very easy way to see and manipulate data. However, advanced functionality beyond Excel doesn't need to be scary. Tools like Microsoft PowerBI are an intuitive step up from Excel that result in users getting a lot more from data without requiring extra skills or knowledge.

Hadoop has taken the big data realm by storm with its possibilities. It offers flexibility and scale, and brings a broad ecosystem of innovative complements for a wide range of analytics for both structured and unstructured data. Of course, its newness is a double-edged sword. For now, reducing costs is the most-cited driver for Hadoop adoption, but there are many others of similar importance, according to recent ESG research (see Figure 3).³

Figure 3. Common Drivers for Hadoop Adoption



Source: Enterprise Strategy Group, 2016.

Hadoop is compelling, but enterprises certainly aren't abandoning their data warehouses—and they don't have to. They've been investing in legacy architectures for the past 15 years, and so they need ways to incorporate new technologies in smart, cost-effective ways. For example, legacy infrastructure can support data analytics using SQL and Hadoop.

IT teams are frequently lacking the skills to extract full value from Hadoop and may not know how to query and manage big data analytics activities. Many of the reasons cited for adopting Hadoop in Figure 3 are actually complementary. They can bring efficiency and optimization to data warehouses, which otherwise can squander some very expensive resources. For example, Hadoop complements and enhances the traditional data warehouse in cases like risk analysis modeling, transaction analysis, and predicting system failures.

³ Source: ESG Research Report, *Enterprise Big Data, BI, and Analytics Trends: Redux*, to be published.

What to Look for in a Hadoop-ready Big Data Solution

All of the above approaches can be merged, but it's important to consider the actual delivery mechanism. Many in IT are experimenting with large numbers of processors to perform massively parallel processing (MPP) analytics on commodity servers, especially with Hadoop and NoSQL databases. However, this can cause problems for IT infrastructure and operations teams. Twenty one percent of ESG research respondents say they would prefer an appliance for new initiatives, not wanting to attempt all the necessary design, component procurement, systems integration, and tuning themselves.⁴

Choosing a big data analytics platform may seem daunting, as there are options from major hardware and software IT vendors along with many new emerging startups and cloud services. It can be unclear which solution is the most effective and won't add infrastructure overhead or too many additional processes. Ultimately, the right analytics solution has to address the needs of the business without compromise.

Based on ESG research, we suggest evaluating offerings against the following factors. Each one is necessary in making an analytics decision.

- Carefully consider performance of the analytics tool, since the speed of queries and scans will have a big impact on effectiveness.
- Look at the potential for scalability of the platform. Future growth should happen seamlessly, without compromising performance or driving up costs.
- Take a look at the product's usability for both the business side and IT. It should enable insights with familiar queries, taking advantage of Polybase and other solutions.
- Examine the overall cost of the platform, including both up-front acquisition costs and ownership costs.
- Invest in a tool that can integrate tightly with Hadoop and SQL Server, including ETL for data movement and a tuned hardware and software platform.
- Take into account the solution's ease of administration for both data management and system management.
- Consider the expected time to value, including deployment and configuration time.
- Ask vendors for a proof of concept demonstration so that you can get a feel for what the tools provide and what they can do for your organization.

HPE, Intel, and Microsoft Deliver a Converged Solution for Modern Data Warehouses

While HPE offers customers a choice for data warehouses (SAP and its own Vertica and Autonomy Idol), the company has an impressive 30-year partnership with Microsoft. This partnership announced the first joint data warehouse appliance in November 2010, and has continued to invest and innovate to offer a converged analytics solution for customers who prefer Microsoft. Meeting the evaluation criteria for a Hadoop big data appliance requires a powerful product with both hardware and software optimized for data analytics usage. HPE's ConvergedSystem 300 for Microsoft Analytics Platform is a true collaboration between HPE and Microsoft, resulting in a complete big data analytics solution, purpose-built, factory integrated, and tuned to meet your most data-intensive big data, business intelligence, and analytics workloads. This turnkey data warehouse solution offers the power of massively parallel processing architecture and in-memory columnstore for dramatically faster query execution performance. Users can start with a single preconfigured rack and can linearly scale out to 6 petabytes of data. Integrated support with Microsoft and the unique HPE Support Pack designed to help customers with lifecycle management, which includes appliance validation, diagnostics, tested firmware, and driver packages, ensure that the solution runs at optimal performance and delivers expected return on investment (ROI).

⁴ Source: ESG Research Report, [Enterprise Big Data, Business Intelligence, and Analytics Trends](#), January 2015.

HPE ConvergedSystem 300, an optimized converged solution, running Microsoft's Analytics Platform System (APS), brings together Microsoft's MPP data warehouse technology and the flexibility to deploy Hadoop any way you want—Linux or Windows; on-premises or in the cloud—with HPE's high-performing servers and storage in a converged package.

The HPE solution integrates in-memory computing performance, which is a modern technology that enables fast, efficient processing. Both relational and non-relational data are available in the solution.

This solution is well supported by HPE and Microsoft and was recently updated with HPE's ProLiant DL360 Gen9 servers for increased compute power with faster Intel Xeon processors. The MPP warehouse is composed of an orchestration server and a failover server, and can be built up in scale units of two servers and a D6000 storage unit (see Figure 4). In terms of storage, that's 70 D6000 drives of 1, 2, or 3 TB apiece. In total, the CS300 then can expand to a 6 PB maximum over seven racks, with a 5:1 compression ratio. It's also been updated to a software-defined storage model. Networking technology takes advantage of two InfiniBand switches from Mellanox, along with two of HPE's own Ethernet switches with redesigned top-of-rack HPE 5900 series switches.

The ConvergedSystem 300 for Microsoft Analytics Platform includes PolyBase, which enables seamless querying of data stored in Hadoop as well as the data in a SQL data warehouse using T-SQL. That eliminates the need for additional Hadoop skills or training.

The ConvergedSystem 300 for Microsoft Analytics Platform was built to scale in a linear way, and was tested with 10 TB of data containing 60 billion records. Users can follow best practices around database and table creation and set up column store indexes. Upgraded CPUs within the system mean performance has increased over the last version of the solution. Additionally, there's a central administrator reporting tool to view system and database activity, including queries and loads.

The optional HPE Data Integration Platform (DIP) provides an intermediate location for database backups, and also a staging location for extract/transformation/load (ETL) operations. Using the optional DIP server, data can be loaded, backed up, and restored to the HP ConvergedSystem 300 for Microsoft Analytics Platform at speeds up to 7 TB/hr. With up to 200 TB of raw storage, and high-speed InfiniBand and 10Gb Ethernet connections available, the DIP server provides a robust host for any and all ETL and ETL management architectures.

ConvergedSystem 300 for Microsoft Analytics Platform also includes the unique HPE Support Pack Utility to reduce complexity and improve performance. The support pack includes shell scripts for use in validation, diagnosis, reports, and systems management. Users can also view diagnostics and reports, and see data on PDW regions. Plus, they can set up shared volumes and cluster failover.

These features all contribute to a fast time to value for customers. Because this solution is converged and optimized, users can start gaining insights quickly and can expect high performance. The solution also extends existing SQL resources for a smooth integration process, which also allows the use of existing SQL licensing. The lack of proprietary hardware creates cost optimization.

Figure 4. Example Configuration of the CS300 with PDW Compute Nodes



The Bigger Truth

Microsoft and HPE have collaborated in some smart ways with this combined product. Each company has contributed from its areas of strength. HPE brings high-performing hardware, while Microsoft brings time-tested, user-friendly database and interface technology. This leverages most companies' in-house SQL expertise and even allows reuse of existing software licenses on the ConvergedSystem 300 for Microsoft Analytics Platform.

For IT teams, appliances and engineered systems can be attractive, since everything is collected into a manageable package. But those solutions can be extremely expensive. This offering strikes a balance between cost and effectiveness to please both IT and business leaders.

The ongoing partnership and joint engineering efforts of this Microsoft/HPE collaboration can only spell good news for any enterprise leader looking for more or better data insights. The world of Hadoop is evolving quickly, and the stalwart data warehouse continues to play an important role for enterprises. The concept of the data warehouse working seamlessly with Hadoop is efficient and practical, and may well become the best model for enterprise analytics in the modern market. Companies looking for a joint solution would be well advised to add the ConvergedSystem 300 for Microsoft Analytics Platform to their evaluation list for proofs of concept, and see how it fits with their data warehouse and analytics needs.



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