WHITE PAPER

# Building a Datacenter Infrastructure to Support Your Big Data Plans

Sponsored by: Cisco in collaboration with Intel

Richard L. Villars Dan Vesset

January 2014

## IDC OPINION

Companies rely on an expanding set of applications to compete in today's rapidly evolving business environment:

- They rely on a fast-growing array of applications and devices (email, collaboration tools, and smartphones/tablets) to communicate and conduct business with customers and business partners.
- They are creating, collecting, and repurposing large, unstructured data sets in life sciences, geophysics, media, and manufacturing.
- They are collecting, storing, and analyzing more social and sensor-generated data about environments, products, customers, and transactions.

The promise of better and faster data-driven decision making based on all this information is pushing big data and analytics (BDA) technology to the top of executive agendas. To succeed, CIOs must place a laserlike investment focus on datacenter solutions that allow them to deliver scalable, reliable, and flexible infrastructure for fast-growing BDA environments. They must:

- Improve business agility and reduce the cost of doing business through the aggressive use of technologies such as integrated systems in datacenters.
- Enable more effective use of all types of information by a wider range of users by making the information easily, but securely, accessible and minable through the use of big data and analytics solutions.

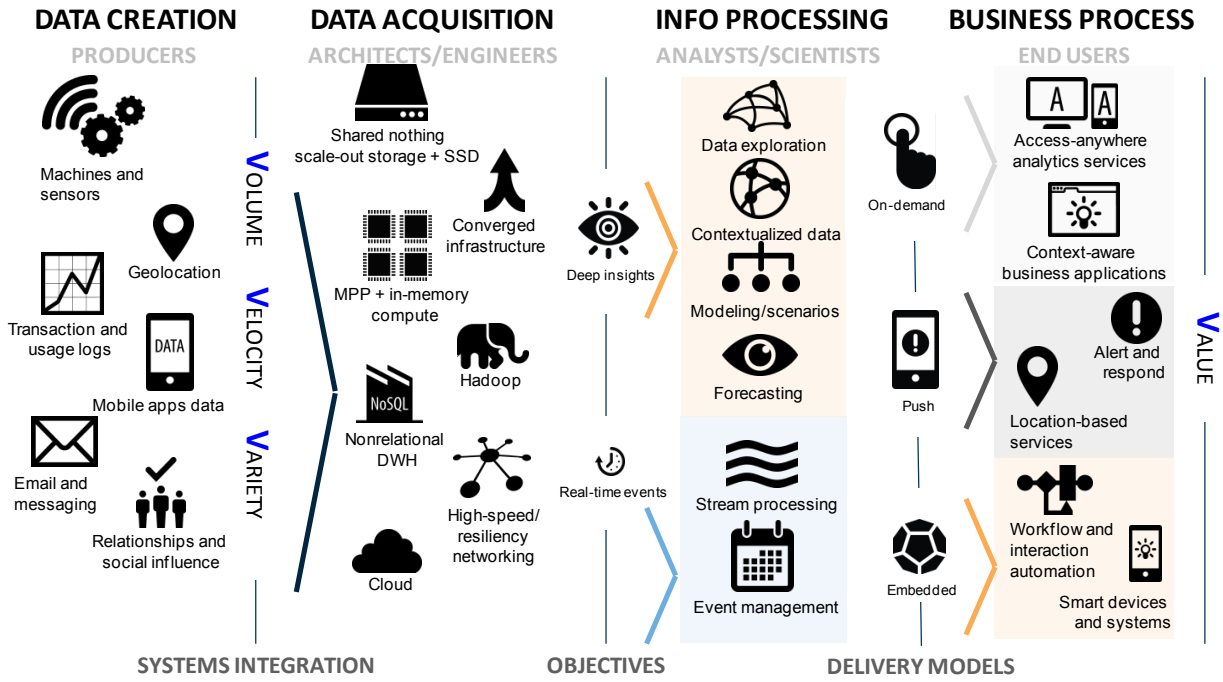
Solutions like Cisco's Common Platform Architecture based on UCS with Intel Xeon processors provide a predictable, flexible, and open infrastructure on which companies can build a broad portfolio of BDA solutions while minimizing capital and operating costs.

## THE GROWING IMPORTANCE OF BIG DATA

The growing focus on big data and analytics solutions as a basis for competitive advantage is both an opportunity and a challenge for most organizations. The promise of better and faster data-driven decision making is pushing BDA technology to the top of executive agendas. The BDA ecosystem, as shown in Figure 1, is a complex interplay of technology, data, processes, and people looking to derive value from the creation or use of data.

## FIGURE 1

### IDC's Big Data Ecosystem




Source: IDC, 2014

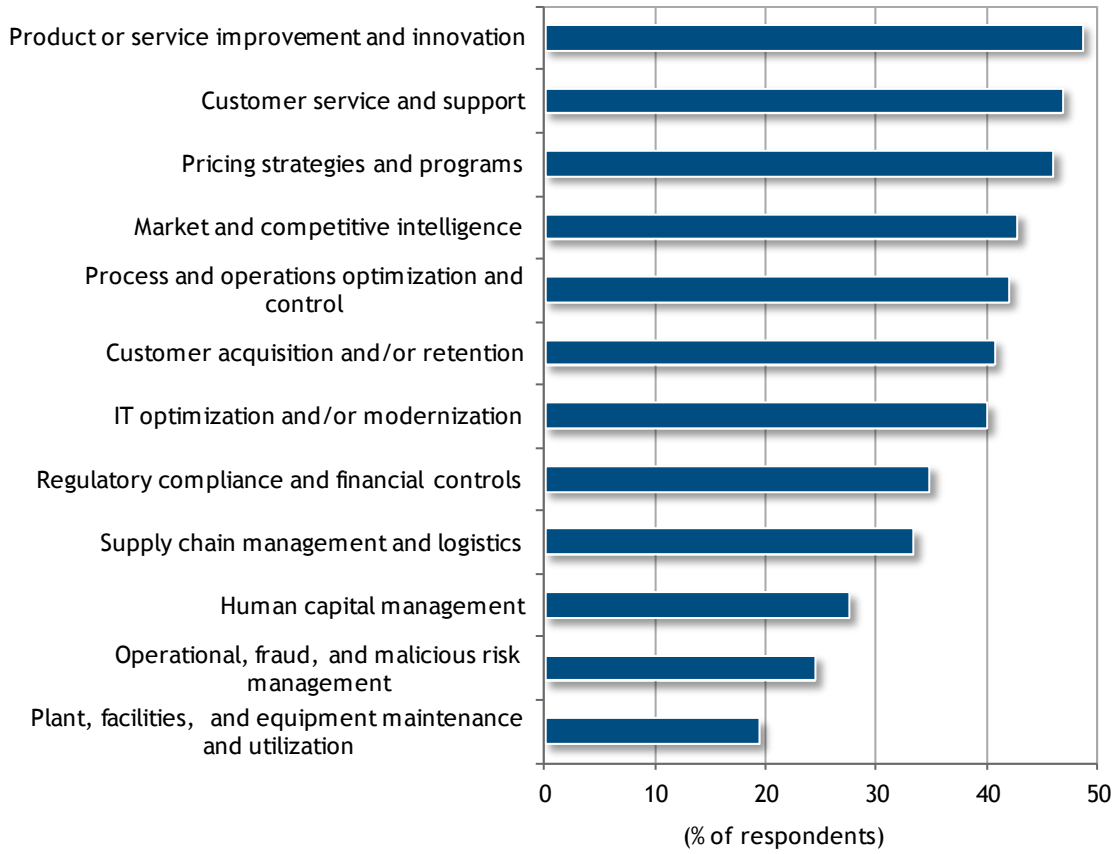
In this environment, it is not only access to information but also the ability to analyze and act upon the information in a timely manner that creates competitive advantage in retail and consumer marketplaces, enables sustainable management of communities and natural resources, and promotes appropriate delivery of social, healthcare, and educational services.

## Big Data Opportunities and Drivers

A broad range of business and technology needs drive organizations to invest in BDA projects and solutions. Recent IDC research identifies the most significant drivers of BDA initiatives across industries as those shown in Figure 2.

## FIGURE 2

### Most Significant Drivers of Big Data and Analytics Initiatives




n = 701

Source: IDC's *Big Data and Analytics Maturity Benchmarking Survey,* July 2013

These drivers differ in importance across industries, but their breadth is evidence that BDA solution benefits have the potential to affect both front- and back-office operations, introduction of new products or services, and development of new business models.

## Big Data Defined

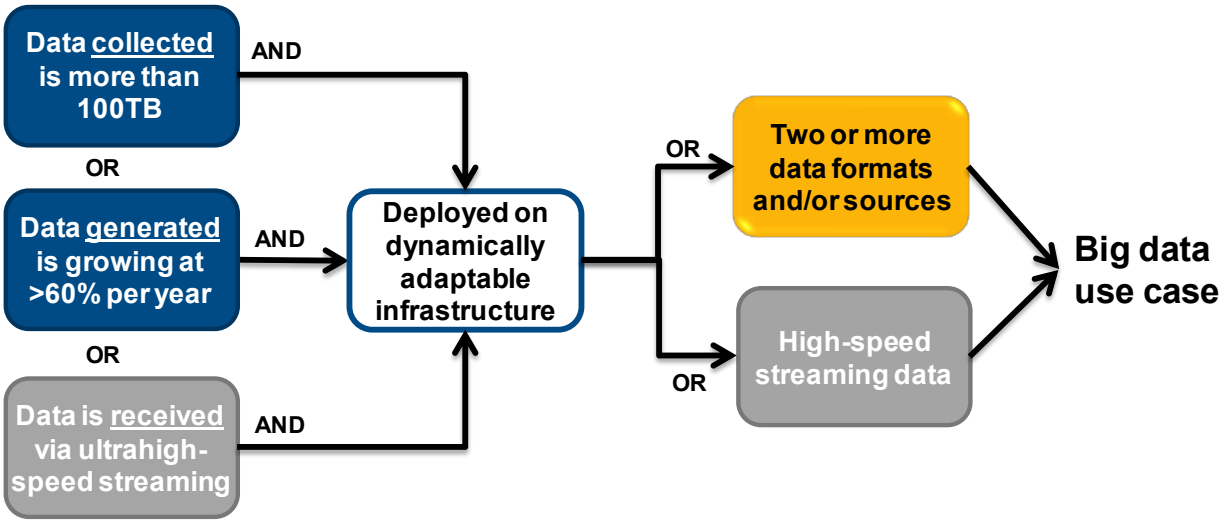
Today, commercial and social interactions among people, between people and machines, and among machines produce a constant stream of data to monitor and analyze. Social interactions, mobile devices, facilities, equipment, R&D, simulations, and physical infrastructure all contribute to the flow. In aggregate, this explosion in data sources is at the heart of big data.

IDC takes quite a restrictive approach to defining big data. We do not simply define big data as social media data or Web clickstream or machine-generated data, nor is our definition constrained to any one particular technology, such as, for example, Hadoop.

IDC's definition of big data technologies describes a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis. This technology could include a range of appropriate or fit-for-purpose software and hardware that are able to address the scenarios depicted in Figure 3.

## FIGURE 3

### IDC's Big Data Definition and Criteria




Source: IDC, 2014

The three scenarios include cases where data (of any structure) collected is more than 100TB; cases where there isn't 100TB of data but the rate at which the data is being generated is increasing at the rate of >60% per year; or cases where the data is being received via ultrahigh-speed streaming. The first two scenarios can be categorized as examples of data at rest, while the third scenario is an example of data in motion. In addition to these criteria, the data at rest should include or originate from two or more data formats and/or sources. In all cases, there is an expectation of a dynamically adaptable infrastructure to support the timely collection, analysis, and use of data in each scenario.

## Benefits and Value of Big Data Solutions

BDA solutions are already delivering benefits in both business and information management. 77% of organizations report that their current BDA projects are delivering quantified benefits that either meet or exceed expectations. These benefits come in the form of:

- Improved ability to predict equipment defects or failure in asset-intensive industries
- More personalized interactions with customers in the retail, financial services, and media and entertainment industries
- The ability to run through more iterations of analytic models during a fixed time frame
- Optimized pricing in the retail, hospitality, and travel industries
- More accurate fraud detection and subsequent fraud prevention processes in the financial services and government industries
- Improved delivery of personalized healthcare
- Faster research and development process cycles in the pharmaceutical and manufacturing industries

From the information management perspective, BDA solutions support and improve:

- Integration of multistructured data, such as structured data from transactional systems and unstructured customer interaction systems, sensor data from equipment, and so forth (48% of organizations surveyed by IDC indicate that they subscribe to external data sources to extract more value from internal data.)
- Storage of very large data sets that lend themselves to analytic experimentation and discovery by data scientists using nonrelational information management software
- Identification of nonobvious connections among customers through the application of graph analysis software
- Searching and entity extraction through vast repositories of textual information, including rich media metadata or audio-to-text transcription using text, audio, and video analytics
- Real-time monitoring of streaming data using event-based middleware software

These are just a few examples that highlight the breadth of use cases and technologies that come under the umbrella of big data.

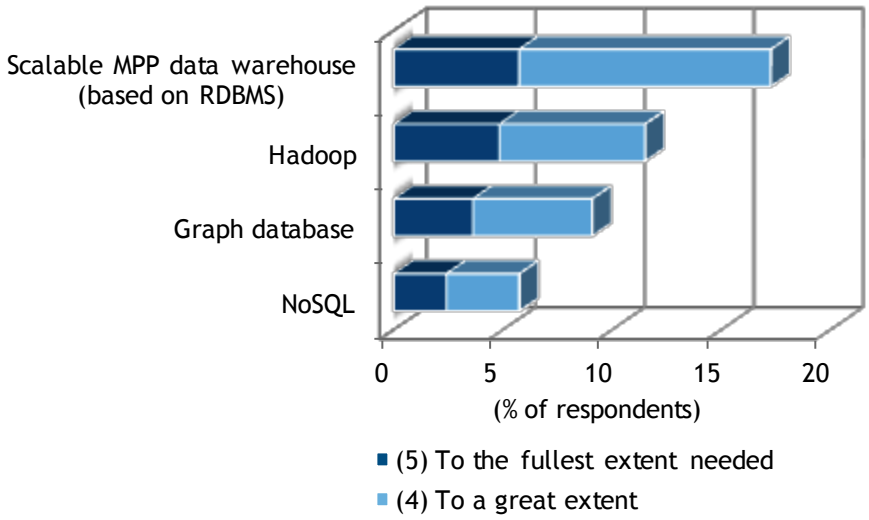
## MAKING BIG DATA AN INTEGRAL PART OF YOUR DATACENTER

Much of the focus of big data technology has been on the various software options such as Hadoop, NoSQL, and MPP database reference architectures and appliances, complex event processing (CEP) engines, graph databases, key value stores, and machine learning tools. Figure 4 shows the current level of adoption of some of these software tools.

## FIGURE 4

### Use of Specialized BDA Information Management Software

*Q. To what extent has your organization supplemented general-purpose data management technologies such as RDBMS with any of the following specialized BDA technologies?*




n = 330

Note: Use of each BDA technology was rated on a scale of 1 to 5, where 1 = to a very limited extent and 5 = to the fullest extent needed. Only results for options 5 and 4 are shown.

Source: IDC's and Computerworld's *Business Analytics and Big Data Survey,* August 2013

Today, organizations are moving beyond initial efforts that targeted specific data sets or business processes. To get the full benefit of their big data efforts, organizations need to integrate these technologies into the existing datacenter environment. They need to deploy solutions that optimize the performance of software tools across the range of BDA use cases while also building tighter links with existing datacenter infrastructure. Clearly, a one-size-fits-all approach doesn't work in terms of either software or hardware, but common strategies for connectivity and asset management will be critical.

## Problems and Challenges Big Data Poses for Today's Datacenters

Datacenter executives identify several major problems and challenges that expanded big data solutions (compute, network, storage) pose for their organizations. The major pain points fall into three major categories:

- **Enabling realistic capacity planning:** BDA solutions, whether delivering results in a batch process or real time, are highly elastic in nature. Loads on compute, storage, and network resources for any one solution can wax and wane rapidly in time scales ranging from days to seconds.
- **Improving availability:** BDA solutions are playing a greater role as a near-real-time component in customer-facing services, not just in supporting back-office decision making. Any delays in accessing data, reductions in available processing capacity, or delays in

delivering analysis can have a direct effect on business outcomes (revenue, profits, customer satisfaction).

- **Enabling more efficient and agile IT operations:** BDA solutions are proliferating as more parts of the organization seek to leverage data sets and analytic models. Without greater automation in IT asset administration and service provisioning, any business gains are soon overwhelmed by exploding IT operations costs.

## BIG DATA INFRASTRUCTURE STRATEGIES AND BEST PRACTICES

To avoid or overcome the challenges big data solutions can pose, datacenter planners and operations teams must start to incorporate big data requirements into architecture designs and IT operations best practices.

### Architect for Easy and Rapid Increases in Scale

- Standardize on validated systems that unify dense compute, memory, storage, and network resources for rapid deployment and expansion of modular compute/analytic pools.
- Enable a scalable/active data repository that is linked to existing structured and unstructured data storage assets within a common data security, replication, and management framework.

The combination of these two platforms will make it possible to address both historical and real-time data analysis workloads based on a common foundation that is readily scalable, flexible, and less expensive to operate.

### Automate for Utilization and Agility

- Take automation to the next level by leveraging automated service provisioning and orchestration to enable rapid (in minutes or seconds) deployment of compute assets across evolving and variable big data workloads.
- Implement a highly scalable, intelligent network solution that enables rapid network reconfiguration (intra- and inter-datacenter) for timely, secure movement of internal and third-party data sets to the appropriate locations.

## BUILDING A BIG DATA FOUNDATION BASED ON CISCO UCS

Cisco Systems introduced the Unified Computing System in 2009 as a next-generation server and networking platform closely linked to partner storage platforms. The initial target was reducing the cost and complexity of deploying and managing fast-growing and increasingly critical pools of virtualized applications. The scale-out architecture of UCS was built with large-scale service provider and scale-out applications in mind.

The Cisco UCS Common Platform Architecture is emerging as an effective platform for enabling scalable BDA environments. Key characteristics that make UCS an attractive platform include:

- **Predictable deployment:** In rapid-scale BDA environments, ensuring that systems deployed today work the same as systems deployed last quarter or next year is critical. Organizations don't have the luxury of time to continually retune systems, and UCS solutions can be deployed in highly standardized, predefined pods to manage and monitor the systems from a single pane of glass using UCS' management suites including UCS Manager and UCS Central.
- **High performance:** Many BDA solutions are at the forefront of high compute, low network latency, and in-memory operations as they continually seek to boost performance. In addition, Cisco UCS servers include the Intel Xeon processor E5-2600 v2 product family, which delivers significant performance and efficiency gains. UCS also provides solid state solutions, and with Cisco's recent acquisition of WHIPTAIL's solid state product, Cisco will be able to extend capabilities much further.
- **Coexistence with enterprise applications:** Companies can host both enterprise applications and big data applications under the same UCS domains, allowing data movement into Hadoop, NoSQL, or MPP relational information management technology easily while eliminating the need for costly technology silos.
- **Massive scalability:** UCS Central federates (brings fabric interconnects together) UCS domains to accommodate up to 10,000 servers under a single management framework.
- **Advanced orchestration and automation:** The hallmark of BDA deployments is that any single application will wax and wane in terms of load, but the entire system needs to be constantly available. Cisco UCS provides the programmable framework for its partners to provide a set of advanced provisioning and automation functions that make it possible to redeploy compute and network assets in minutes, not days or hours.

## Challenges for Cisco

Most organizations are still in the early stages of incorporating BDA into their data management and business processes. The journey will take several years to complete. Today, rethinking datacenter infrastructure in terms of how it is designed, deployed, and managed is a top priority. Many efforts revolve around making greater use of converged infrastructure systems like Cisco's UCS to improve agility, availability, and operating efficiency.

Cisco, as a leading provider of core elements in current datacenters, must continue to expand its role in helping companies make the transition as painless and flexible as possible. Part of achieving this goal requires Cisco to stay focused on rolling out some key technical enhancements to existing product lines. These enhancements include:

- Extending support for higher-speed network links such as 10GbE, 40GbE, and 100GbE across all of its compute products
- Continuing to enhance the usefulness of automated data movement services offered by its storage technologies and partner ecosystem
- Widening the portfolio of BDA-specific converged systems by expanding partnerships with leading BDA software developers

Cisco must also continue to extend the scope of unified orchestration services across multiple converged systems and across multiple datacenters. Extending enterprisewide support further in other elements (e.g., server provisioning, performance monitoring, and resource management) will also be critical.

Extending BDA solutions across the enterprise presents technology challenges, but more important, it poses a significant number of challenges for IT organizations in terms of product evaluation, budgeting, and IT operations management. These organizations will ask Cisco what it is doing in terms of services, application development, and financing to help them navigate the transition in the areas of technology standards, facilities design, and IT staff retraining. Cisco and its business partners must educate customers about the broad set of professional services offerings available to help them navigate the change.

## CONCLUSION

Big data represents big opportunities for organizations but will also place more pressure on CIOs and their IT teams. Almost every CIO dreams about making IT a more valued asset to the organization. BDA projects are at the frontier of the business where many of the most significant business expansion or cost reduction opportunities lie. Taking a lead in BDA efforts provides the CIO with a chance to be a strategic partner with the business unit.

Because speed is strategically important, it is tempting for business unit teams to move forward on BDA efforts without IT support. CIOs need to both understand what their organization is planning and begin developing a BDA IT infrastructure strategy. IDC believes that building successful business cases around big data can be accomplished only through a tight alignment of critical thinking across both IT and the business. This will require out-of-the-box thinking as well as moving outside the traditional IT comfort zone.

BDA deployments also require new IT administration and application developer skill sets, and people with these skills are likely to be in short supply for quite a while. The biggest challenge is the cultural challenge. From a business and IT governance standpoint, initial BDA projects can quickly evolve to have companywide and industrywide business, organizational, and legal consequences. Business and IT leaders need to start preparing employees, partners, and customers now.

## About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1000 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For more than 48 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

## Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com

---