White Paper

Capgemini Builds on VCE to Deliver Advanced Services for Big Data and Internet of Things

By Nik Rouda, Senior Analyst

December 2015

This ESG White Paper was commissioned by VCE with the cooperation of Capgemini and is distributed under license from ESG.

Contents

A Case Study in Evaluating Big Data and IoT Solutions

Around the world today, many organizations are looking at how they can deploy big data and the Internet of Things (IoT) solutions to learn more about their operations, their customers, and new opportunities. Yet not every organization is confident that it understands the emerging technologies available for big data and IoT well enough to choose the right platform and deploy it successfully in a new initiative. ESG interviewed consulting services and systems integrator (SI) Capgemini to see how they approach this challenge, for themselves and as a foundation for a new offering for their large enterprise customers.

The goal for one specific offer in Europe was to understand which issues were of most importance in both the business proposition and the technology, and how they validated the capabilities of the options available to them. After extensive evaluation, Capgemini settled on VCE as the most appropriate solution for these customer business problems, because it provided a stable and flexible big data platform and business data lake-as-a-service, delivering a range of benefits for their end-users and for their own delivery teams.

Capgemini makes an interesting use case because a key element of its client approach is to work with clients collaboratively to build new solutions involving both business and technology, to find new value, reduce risk, increase capabilities, and transform organizations and processes. With 180,000 people in more than 40 countries, Capgemini appreciates the importance of selecting technology platforms that can support large enterprises at scale to meet rigorous requirements. As such, it's uniquely qualified to appreciate both the opportunities and problems around new initiatives for big data and IoT. The company expects that the "Industrial Internet" will be an area of rapid growth, and has recently invested in building its own expertise and focused teams.

Seeking Value from Big Data and IoT

Increasingly much of the world around us is instrumented to capture data about activities and behaviors, but utilizing this information is often a significant challenge. IoT magnifies the challenge by using sensor and machine-to-machine (M2M) information at a grand scale, and many of the use cases for IoT are still being defined today. Big data technologies provide the capabilities to store and analyze all of this IoT data, meaning that the two technology areas are often tightly intertwined. Inevitably to tackle the one requires parallel efforts in the other. While they are complementary disciplines, this dependency also increases the complexity of building a new solution, potentially increasing both the risks and the costs of the endeavor.

> Sixty percent say their spending on big data and analytics will increase in 2015 as compared to the previous year.

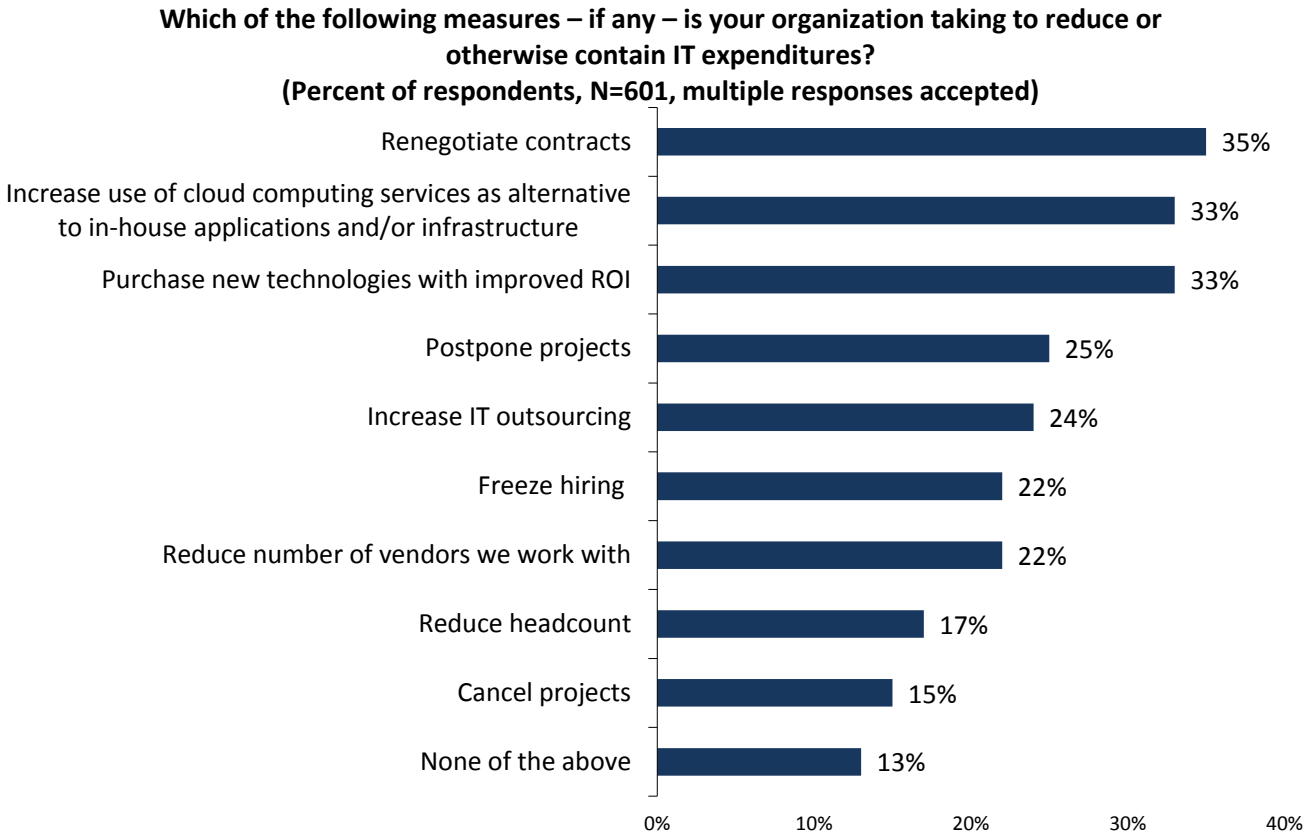
These concerns have not deterred spending, though, and 30% of ESG research respondents say that improved data analytics is one of the business initiatives that will drive the most technology spending in their organization in 2015.[1] Further, their investments are actually accelerating, with 60% saying that their spending on big data and analytics would increase in 2015 as compared to the previous year. Much is riding on the success of these new initiatives, and businesses are looking for ways to ensure they will see value and justify the large spending increases expected, based on the ESG interviews. One common way to reduce the risk of failure is to seek external expertise to assist. Indeed, 27% of those responsible for new projects indicated that systems integrators help provide the skills and manpower to implement and manage their initiatives around big data and analytics, while 26% said that management consultants provide the same service.[2] An organization like Capgemini will have learned many lessons as they built out their offerings, and can encode this wisdom into a repeatable process for customer success.

At the same time that spending is increasing for big data and analytics, many businesses are under pressure around IT cost containment, and this is reflected in ESG research on ways businesses are reducing their spending (see

[1] Source: ESG Research Report, *2015 IT Spending Intentions Survey*, February 2015.
[2] Source: ESG Research Report, *Enterprise Data Analytics Trends*, May 2014.

Figure 1).[3] While in the past, cost reduction often came from reducing staff, delaying projects, or other deferment techniques, this has largely shifted toward "smarter" spending now.

Figure 1. Measures to Reduce or Contain IT Expenditures



Source: Enterprise Strategy Group, 2015.

Capgemini's offerings in this area are aligned to these trends, including managed/hosted cloud offerings (with 33% interested) and increasing IT outsourcing (24%). Arguably, a well-executed initiative for big data and IoT will also deliver a significantly improved return on investment (ROI) versus continuing with only traditional approaches.

Business Drivers for Addressing the Industrial Internet

Like many businesses, Capgemini decided there was potential for new revenue streams in the areas of big data and IoT, yet revenue for its own sake is not enough; any new initiative must be a profitable practice to justify investment. While IoT can have both consumer and industrial applications, for Capgemini it was the industrial use cases that were identified as most appealing; the digitization of physical assets and the transformation of them into resources that are reportable, manageable, and predictable—all without direct human intervention— as a transformative event, ready for their clients to exploit and monetize.

All kinds of devices are now generating data, indicating how they're actually being used and behaving. Once this information is captured, it can be leveraged with analytics to generate actionable information and insights, and even provide automated, remote control of the devices. This is not necessarily a new idea, but a number of trends now are aligned to make it more applicable today, including:

- **The maturation of Hadoop** and related technologies that make it much more economical to store, manage, process, and analyze very large volumes of information.

[3] Source: ESG Research Report, *2015 IT Spending Intentions Survey*, February 2015.

- **The emergence of ubiquitous, low cost sensors** to monitor all kinds of objects, in any location, and stream data over networks back to a central repository.
- **More frequent adoption of "DevOps" and "fail fast" cultures,** where new applications can be invented, tested, and either implemented or abandoned without penalty.
- **Excitement about "real-time" solutions** where an automated response is handled in the moment, as opposed to batch analytics and reporting at some later date.

Many businesses have long-existing pain points, but lately they've been able to identify where the new economics of the IT solutions, combined with the increased business opportunity, now make it compelling to move forward to implement Industrial Internet projects.

Some common examples of these applications are:

- Fraud detection for financial services.
- Yield optimization for energy production.
- Distribution and grid monitoring for utility companies.
- Connected cars in the automotive sector.
- Video analytics for safety and security.
- Improved marketing, sales, and operational efficiency across all industries.

While these may all be quite different in the ways that data is created and absorbed, the analytics models and infrastructure platforms are universally applicable.

Creating a Dedicated Team

For the Capgemini team, the main concerns were how to prioritize and develop these use cases and then standardize the offerings to make them more cost-effective and repeatable. A new team was formed to consolidate the efforts and expertise, not by line-of-business or geography, but around big data and IoT capabilities. This focuses on delivering prescriptive advice in a more agile fashion. While the actual applications may be sector-specific, having common platforms, approaches, and staff on the back-end makes it run all the more efficiently. The need for this kind of organizational transformation is often overlooked, but changing from independent silos of expertise to a common fabric is of great importance to overall success. Some financial engineering was also required to make the delivery model attractive to customers with both technical and commercial considerations. Many systems integrators are reluctant to disrupt their existing business, and typically wait for a significant market to emerge before investing to build a practice. Capgemini wanted to move faster. The company also realized that this would be an ecosystem play where developing meaningful alliances with leading technology vendors would be essential.

To measure success, some fundamental metrics and principles were defined. Of course bookings and margins remained among the key performance indicators, but to justify the significant new investments, cost of operations was also a focus. The faster customer solutions can be delivered, the better the net results will be. Even on the in-house IT side, staff efficiency is followed, with an ideal goal of having no one dedicated solely to tactically managing infrastructure, but instead having most staff spending their time more strategically tuning workloads and environments. All of these changes help align, to accelerate the market opportunity and capture a good share of new business.

Choosing the Right IT Platform for the Business

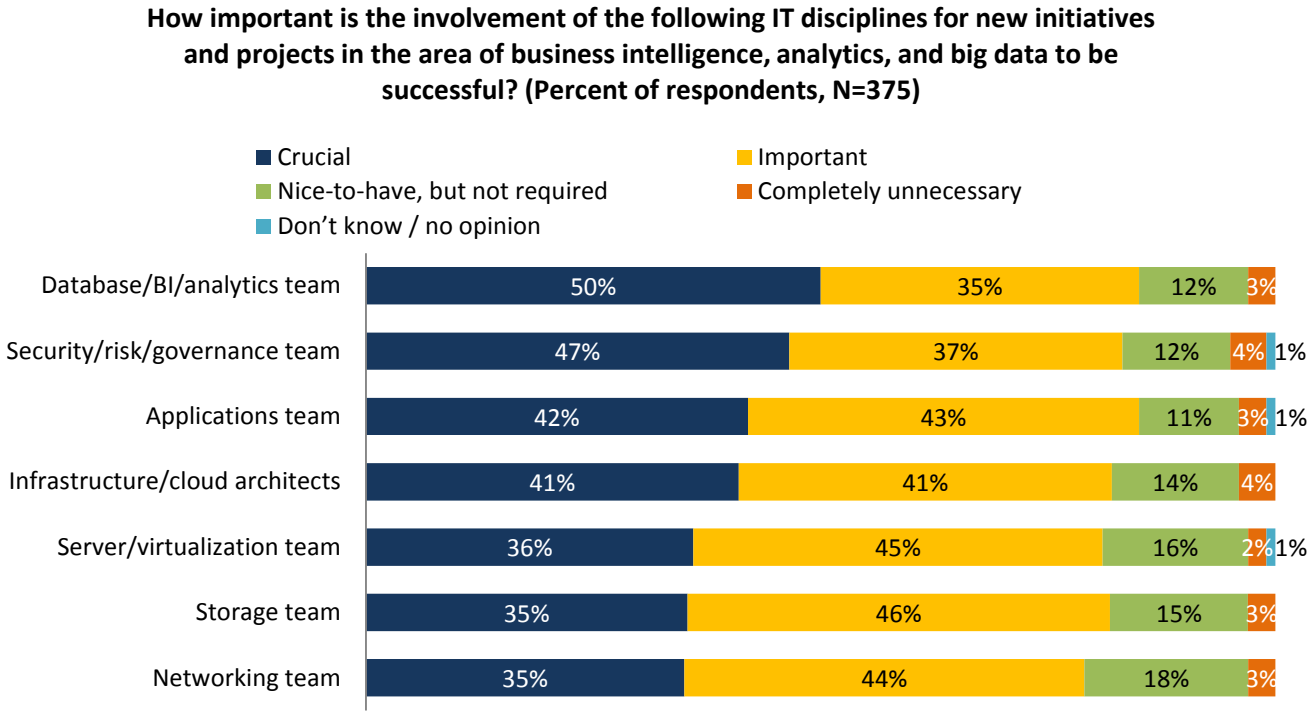
To a certain degree, the choice of infrastructure seems removed from the concerns of many of Capgemini's customers, who are more focused on the business outcomes. As one expert notes, "your business will grow 30%" is certainly a more compelling proposition than "you will have a Hadoop cluster." The conversation has shifted today, and it's no longer the CIO who controls all of the budget—increasingly money is held in the lines of business. Yet for

"Your business will grow 30%" is certainly a more compelling proposition than "you will have a Hadoop cluster."

any initiative to be successful, a high quality foundation is essential, and must be in place before any new big data or IoT application can be built upon it.

Selection of that ideal platform takes careful planning and evaluation, and interestingly spans many disciplines of IT, including the domains of storage, servers, networking, and security, not just analytics and applications. Figure 2 shows exactly how important it is to find a solution that meets the diverse needs of each group.[4] The procurement department may pay the bills, but it takes a village to develop the actual solution.

Figure 2. Importance of the Involvement of IT Groups for New Business Intelligence, Analytics, and Big Data



Source: Enterprise Strategy Group, 2015.

Taking this logic of interdisciplinary dependencies a step further, frequently businesses looking to deploy new big data solutions say they will choose an integrated system or appliance as the primary model for deployment, based on the ESG Research.[5] The theory is that the vendor's careful reengineering and design will accommodate all requirements in a more cohesive and coherent fashion than something assembled ad hoc in the data center. When assessing the project in Europe, Capgemini considered this approach in making their own decision, and came to the realization that, while they could build it themselves, it would require a significant investment in terms of cost and time. The goal was to have a reliable, utility-like service that is industrial-grade to support an Industrial Internet environment, not to re-invent the wheel.

Standardization of hardware platforms actually helps with reliability, supportability, and not least, the total costs of operation. A similar dynamic occurs not just at the hardware level of the technology stack, but also in the software. The relatively young Open Data Platform (ODP) alliance is now helping to build in the "consistency of experience" required, so that companies can more readily build their differentiation on top of a well-defined and well-tested Hadoop base. Participation by a wide range of leading vendors, including EMC, Pivotal, Hortonworks, IBM, Splunk, Teradata, and many others, reassures customers that they will avoid lock-in while maintaining quality. It's worth noting that Capgemini is also a participant in the ODP shared industry initiative.

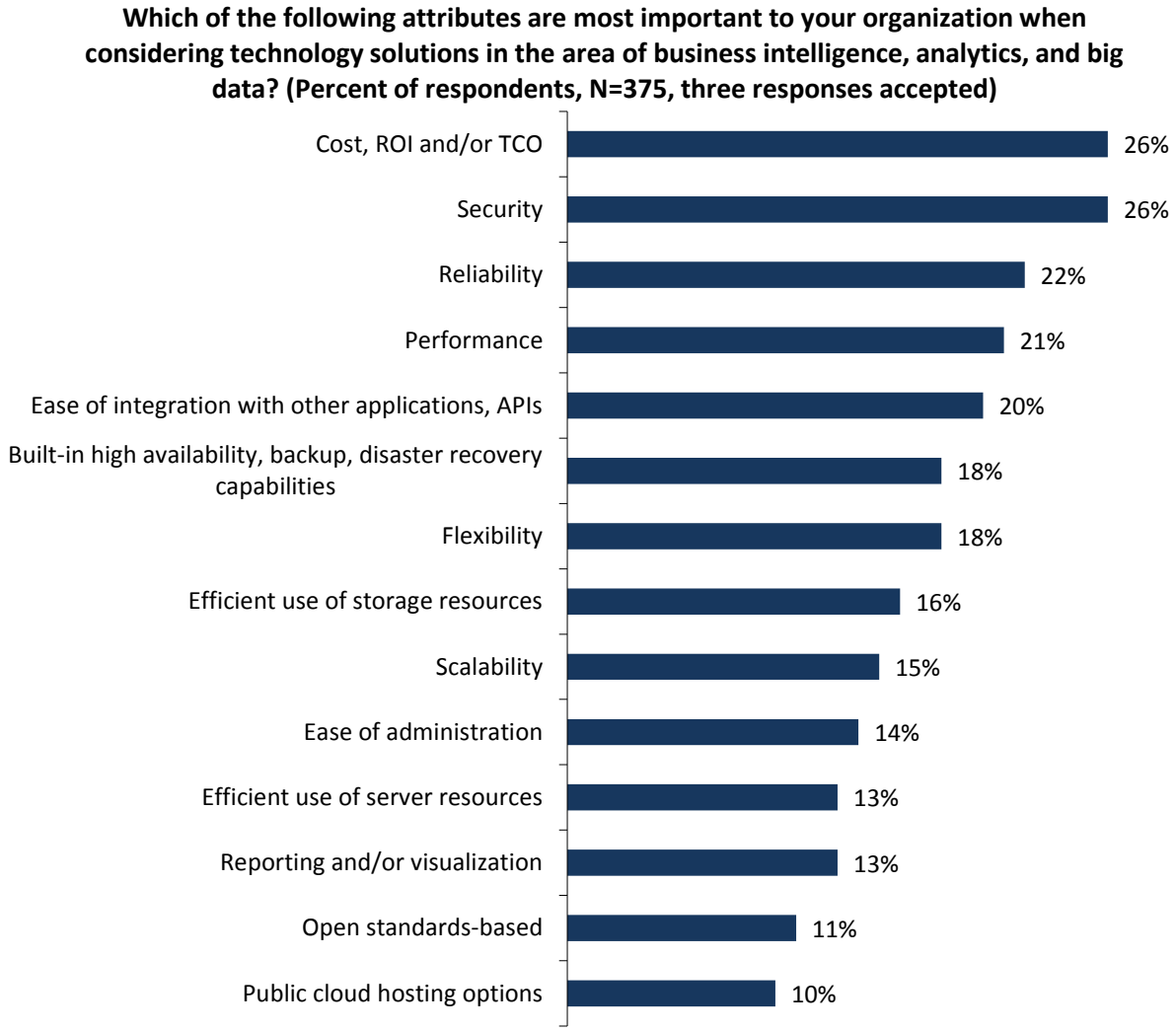
[4] Source: ESG Research Report, *Enterprise Data Analytics Trends*, May 2014.
[5] Source: ESG Research Report, *Enterprise Big Data, Business Intelligence, and Analytics Trends*, January 2015.

How VCE Was the Right Choice for Capgemini

All the above factors were steering Capgemini toward selecting a converged solution that would meet the specific needs of this use case. Yet there are still a number of choices on the market for engineered systems or appliances that have been designed for big data environments, so further research was needed to evaluate the technical underpinnings of each.

Despite vendors' tendencies to emphasize speed and capacity, these are not always the predominant concerns of most organizations. ESG research asked IT decision makers to identify their most important criteria for new solutions in the area of big data, and their responses are shown in Figure 3.[6]

Figure 3. Evaluation Criteria for Big Data and Analytics Solutions



Source: Enterprise Strategy Group, 2015.

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

VCE platform used for Capgemini Use Case:

Vblock System 340

- Cisco B-series blade servers
- VMware vSphere virtualization
- EMC VNX storage
- Cisco Nexus and MDS networking
- Cisco, EMC, and VMware management tools

VCE Technology Extension for EMC Isilon

Financial considerations were foremost on the list, followed closely by ease of deployment, pace of delivery, stability, solid integration up and down the stack, and ease of administration. The Capgemini projects had similar priorities and cited architectural efficiency as well.

A further consideration was the use of analytic tools. There is a certainty that analytics tools will continue to evolve or be replaced over time, and so it was important that those innovations shouldn't impact the platform either. Hadoop, real-time analytics, machine learning, and other new capabilities would need to tie into the existing data warehouses and enterprise applications in customers' environments.

Capgemini points out that the company's initial, exploratory investment in the VCE Vblock System 240 was relatively low, but the solution immediately delivered ample flexibility for dynamic, multi-tenant workloads. Resources could be allocated on-demand during client engagements, and scaled up as the possibilities were developed. The ability to later add Isilon scale-out storage or more rack-mounted servers helped in growing storage capacity or compute power independently as needed. This flexibility came as Hadoop was run in virtual servers, without any noticeable compromise of performance. The Pivotal Big Data Suite served as the full stack analytic capability in the proof-of-concept (POC) environment for prospective customers.

Building on the outcomes of the initial POC, the Capgemini team has switched to primarily using the VCE platform that consists of Vblock System 340 and VCE Technology Extension for EMC Isilon, a "variety-sized" platform supported by both the EMC VNX5600 plus seven nodes of Isilon storage with OneFS. Workloads did indeed evolve into different directions, with Pivotal's Gemfire in-memory database becoming popular, but this shift was readily accommodated by the 16 Cisco UCS B200 M3 blade servers with 512 GB of total memory. The Vblock System has also proven it can handle up to 6 TB of SAP HANA as required. Data platforms like Hadoop distributions from Pivotal, Cloudera, Hortonworks, and IBM, analytics packages like SAS data grids, ETL utilities like Informatica, and many more tools are also used to meet specific customer needs. The transfer rate of the integrated Cisco UCS Fabric Interconnect 6296, at the top of rack or straight through the fabric interconnects, enables more fluidity to move high volumes of data around between repositories and analytics engines. Capgemini also takes full advantage of the Pivotal solution, including Pivotal Cloud Foundry and Pivotal Big Data Suite, consisting of Pivotal HAWQ, Pivotal HD, Pivotal Gemfire, Pivotal Greenplum, and SpringXD. The VCE platform is modularized and adaptable to run the varying workload needs, and it exceeded the Capgemini team's expectations of business and technical considerations, including performance and scalability tests conducted on the platform.

Software level control of the whole stack enables allocation of specific resources in a very dynamic multi-tenant, multi-workload environment, while quality of service (QoS) and security eliminate concerns about doing this. For Capgemini's service delivery team, more consistency and more automation are essential to the way they provision environments from data lakes to IoT and still meet their service level agreements (SLAs) and client expectations. Together, the Vblock System and VCE Technology Extension for EMC Isilon gave Capgemini a blank slate to offer a high-quality infrastructure with consultancy, to determine which big data and IoT technologies are best for a client's own particular needs and preferences.

In order to address services integration, Capgemini added Service and Technical Orchestration layers to make the data lake fully automated. These layers transform the platform into a general (automated) data handling platform suited for a wide range of purposes.

Capgemini also added a data transport layer, mainly to handle IoT sources, as the practice of data qualification in IoT is not yet mature. When sensor data is used for different functional areas, each business function needs to

(dynamically) qualify these data and sources for those specific functions. It can also be that the same data can be used for operations, management reporting or legal tasks, but the mechanism to qualify these sources is still different. Defining the metadata is a complex task for the SI across the whole data supply chain. For Capgemini, this also implies the ability to organize feeder lines for traditional systems and to speed up all kinds of enterprise systems. In areas with hybrid clouds, the feeder lines will guard the data qualification function.

Capgemini further emphasized the confidence that comes from the engineered release certification matrix (RCM) by VCE as an important guarantee to knowing in advance that each component of the system will interoperate without issue. Updates can be managed by either VCE support or Capgemini's own delivery team, but either way, the technology stack will work seamlessly and consistently. This is a big advantage over "build-your-own" approaches on various hardware stacks, (be it for enterprise or commodity hardware), where the systems integration and maintenance effort can be extremely demanding.

Going Beyond Expectations with More Than a Platform

Capgemini has seen the value of innovation and embraced new data partnerships early on to develop clear thought leadership across IoT and wider big data opportunities. In industrializing for their enterprise clients, they are expanding their practice around big data, data science, analytics, and business intelligence.

For example, the company was the first global systems integrator to announce formal relationships with Pivotal and Cloudera, and has since greatly expanded these kinds of alliances, both in number and in depth. No one vendor can offer everything a customer might want, so having a system integrator bring together the various elements greatly reduces complexity and risk.

Looking ahead, Capgemini sees much potential in the future of the company's relationship with the EMC federation including VCE, Pivotal, and VMware, with each company playing a valuable role in building comprehensive solutions. The recently announced Federation Business Data Lake is a great concept, being explored for its blueprint approach to address issues like time-to-value and industrialization of deployment for big data. The Cloud Foundry alliance will likely offer similarly valuable benefits: helping to speed application development, provisioning on-demand to multiple clouds, and scaling for both long-term growth and handling temporary spikes in workloads. All of these various initiatives were appreciated by Capgemini as evidence of how new opportunities are emerging through combinations of capabilities.

The Bigger Truth

You can certainly build your own big data and IoT solutions from open source software and commodity hardware, assuming you have the time, talent, and money. This approach may well be cheaper from a capital cost point of view, but it almost certainly will have significant hidden costs in staffing, support, and ongoing management of the environment. Big data and IoT as disciplines are too new, too complex, and too rapidly evolving to think that this will be easy either up front or over the years.

Instead, you can buy an engineered platform that's ready to go. Plug it in, and focus on inventing and developing new business capabilities. You'll have to evaluate whether your company's core competency is in IT systems integration or in creating new competitive differentiation in your industry sector. For most, the latter is where they want their investments concentrated. The IT platform is merely a means to an end.

Certainly, Capgemini enjoys a special relationship with EMC, VCE, and the federation of sister companies, with special access and influence on strategic direction, but the value of the EMC Federation, including VCE, to Capgemini is open to all comers. After careful comparison of the leading integrated systems against the Capgemini team's particular strict criteria, and after continued success using the systems, Capgemini determined that, for this use case, VCE delivers the right platform. "It just works."

Appendix

Element	Configuration

VCE Vblock System 340	VCE Management
	• Advanced Management Platform (AMP-2HA)
	Fabric Interconnect
	• Cisco UCS 6296UP Fabric Interconnect
	Servers
	• 16X Cisco UCS B200 M3
	Server Details
	• 2 X Xeon Intel E5-2660v2 (2.2 GHz)
	• 512GB memory (16 X 32 GB)
	• Host local storage: None (diskless; SAN Boot)
	• 1 X Cisco UCS VIC-1240
	• Cisco UCS 5108 Blade Server Chassis (with 2204XP fabric Extenders)
	Storage – EMC VNX 5600
	• Connectivity: Fibre Channel
	• Drive Count:
	  ○ 4 x 100 GB SSDs 2.5” disks fastcache + 2 Hotspares
	  ○ 10 x 100 GB SSDs 2.5”disks
	  ○ 20 x 600 GB 10K FC disks +2 Hotspares
	  ○ 24 x 2 TB 7.2K NL-SATA disks +1 Hotspare
	  ○ 45Tb useable
	• 8Gb Fibre Channel Networking—8 lanes from switch to VNX
	Networking
	• Switches: 48 ports Cisco
	• A pair of Nexus5548UP capable of 10 Gigabit Ethernet, Fibre Channel, and FCoE switch
	• Fabric Interconnect: 6296 switches
	• A pair of Cisco 6296 fabric switches
	• Nexus 1000v Virtual switch technology

EMC Isilon (offered as VCE Technology Extension for EMC Isilon storage)	• X410 – 102 TB HDD +1.6 TB SSD/64 Gb/ 2x10GE and 2x1GE • Nodes: 7 • OneFS 7.2.0.1 • Protection Policy: +3 • Dual 10 Gig Ethernet connectivity per Isilon node • Capacity Sizing: Not to exceed 85% usable • Storage Efficiency: 84.2%
Hypervisor	• vSphere 5.5 – EMC PowerPath/VE enabled • vCenter 5.5 • Storage: ○ VNX – VM datastores • Isilon – video storage

Virtual Machines	• VMware vSphere 5.5
	  ○ VM version 10
	  ○ VMware Tools installed
	• 2 VMs per server
	  ○ 8GB memory
	• 4 vCPUs

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