CenturyLink® Business

EXECUTIVE BRIEF

Putting Big Data to Work in the Cloud 5 Key Success Factors

Challenge

Big data is moving rapidly from proof-of-concept to production — leading to "big workloads" for IT managers tasked with analyzing data from a range of diverse sources.

At Stake

Big data analysis is increasingly driving strategic business decision-making for large enterprises.

Solution

5 key success factors in this paper help guide development of big data apps optimized to support enterprise business opportunities — which often means turning to the cloud for scale and cost effectiveness.

Throwing together the terms "Big data" and "the cloud" poses more than a slight risk of hype overload. But hype notwithstanding, the truth is that many businesses are looking to big data as a source of customer insight and a way to bolster operational efficiencies. Consequently, the idea of using cloud infrastructure for data analytics is gaining traction with IT managers tasked with the challenges of analyzing large amounts of data from diverse sources.

A recent survey of IT managers found that 22% of their businesses were in the proof-of-concept stage for big data, 38% were at the budget or early rollout stage, and only 16% were actually in production. Still, production big data apps soon will be the norm in virtually every organization. It's coming.

Big data's many uses include marketing campaign analysis, fraud detection, customer experience analysis, and more. As Figure 1 (next page) shows, the ConsumerGoods.com survey found respondents' top business drivers for deploying big data were making operations more efficient, speeding revenue attainment, and connecting patterns and trends. These are non-trivial tasks; big data is synonymous with "big workloads."

As big data goes from proof-of-concept to production, it is being increasingly applied to enterprise-wide, mission-critical operations. The stakes have never been higher, and so it's time to work around the hype and get serious. Going from theory to practice, though, requires thinking about what it will take for big data to succeed in real-world corporate IT settings.

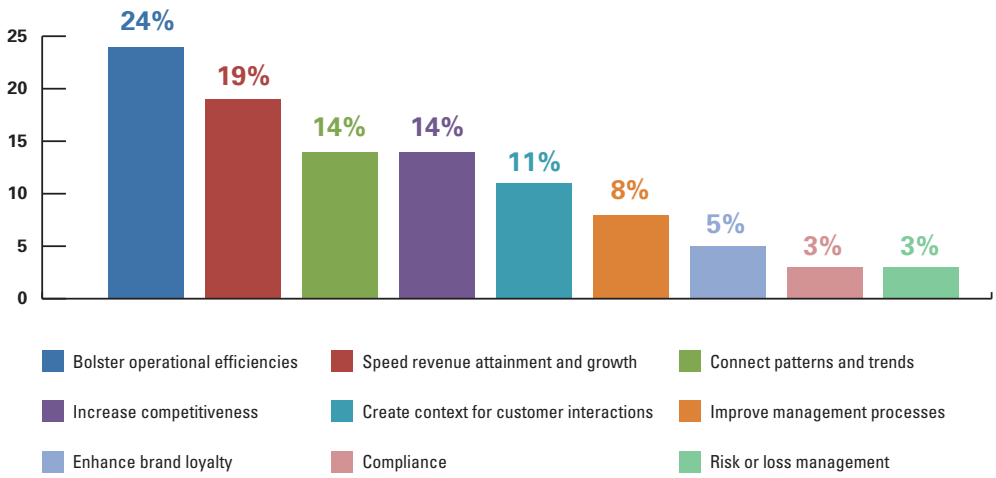

Figure 1: What is your company's main business driver for deploying a big data solution? (Source: Big Data Adoption research by ConsumerGoods.com, January 2014)

Five Key Success Factors

Making big data work involves building a complex stack, depicted in Figure 2. There are many challenges. For instance, as aggressive as big data activity may be, the pace of data collection is currently ahead of storage and utilization. If you're managing big data infrastructure, you will have to be out ahead of this issue (among many others).

Consideration must also be given to big data's massive infrastructure requirements. In this respect, the cloud offers many viable options that are difficult to match on-premise. That said, the "do it yourself" aspect of many current cloud infrastructure solutions has the potential to complicate big data efforts.

To help sort out these challenges, the following guidance is intended to give you a perspective on what's needed to accomplish your business objectives in big data — all while operating efficiently and reducing the risk of data loss, security problems, or poor performance due to infrastructure.

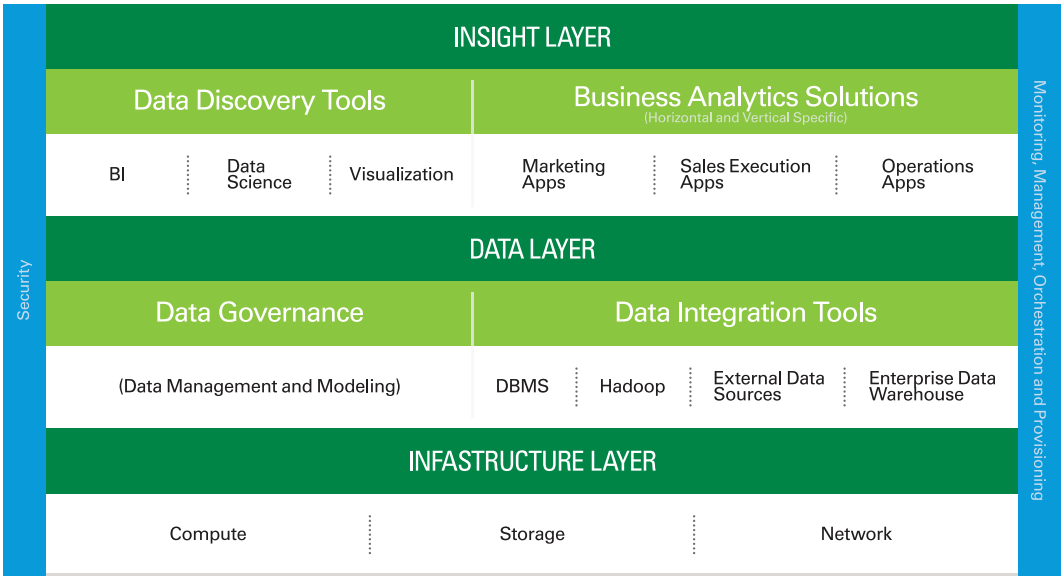

Figure 2: The big data model

1. Flexible, Elastic, and Scalable Infrastructure

Big data usually connotes large scale, though not always in the ways you might expect. True, big data is typically a matter of aggregating and analyzing mind-bending loads of information. However, in some cases, the scale is more eccentric and difficult to forecast. As you experience success with early initiatives, you will undoubtedly identify more data types and sources to funnel into the mix and more use cases for leveraging it. Your data will likely grow at a surprisingly fast rate.

Big data projects don't exist in a vacuum. Your big data infrastructure elements should have the flexibility to support multiple approaches to integration with one another, as well as with external elements. The data in your Hadoop application comes from many sources and is fed into a variety of applications in order to maximize their value for different constituencies. In addition, the constant shifting of big data workloads requires your big data infrastructure to scale elastically. Elastic infrastructure means being able to scale up and down at will. Cloud infrastructure is the best option for elasticity. The cloud typically gives you this ability, making it relatively simple to add or remove data analytics capacity on demand.

2. Infrastructure Availability and Reliability

Today, a lot of big data projects are peripheral to operations — big data is used to achieve insights that are acted on over time. But change is already occurring where operational, customer-facing systems analyze interactions in real time against a massive analytical capability that will drive real-time decisions. Real-time big data is the expected norm of the future.

This is already happening in online advertising. The ad that the web site user gets served is based on an instant, relatively deep analysis of numerous data points about the user's previous online behavior. There is a lot of room for improvement in this, however, and the process will also be extended to many other online and mobile customer interactions. As big data becomes an integral component of day-to-day customer relationships, it will need to be highly available and reliable. Otherwise, it will be revenue that takes a hit.

3. High-Performance Computing with Low Latency

Big data workloads function differently from traditional enterprise applications, especially when compared to applications that are typically cloud-based today. Big data projects generally have a higher level of expected uptime, low latency, and more robust integration across cloud and on-premise systems. Speed is essential in almost every case, though the optimal hardware configuration will vary. For example, some big data workloads require a large number of relatively small processing units while others demand huge amounts of concentrated compute power — the kind best handled by hyperscale servers with solid state drives.

Big data users must be able to access and analyze data from distant and disparate sources as though it were local. Distance and network connectivity also play a role in how your big data environment performs. The closer the data center is to your users, the better the performance. To effectively handle huge quantities of structured (e.g., relational databases), semi-structured (e.g., XML files) and unstructured (e.g., basic text files) data and enable ultra-fast decision-making, you need high-bandwidth, low-latency network connectivity. Cloud infrastructure is usually able to deliver this kind capability in the most economical terms.

4. Secure, Compliant Big Data Infrastructure

Big data demands big security. Indeed, 35% of those surveyed by ConsumerGoods.com indicated that security was a "most important" feature to have in a big data solution. The implications of not having effective security are far-reaching, with repercussions touching on regulatory regimes, privacy, and customer loyalty. If security is breached and sensitive information falls into the wrong hands, the impact on brand standing and customer retention can be catastrophic.

According to IDC, 40% of all data in the world will require protection by 2020, up from about 30% today.[1] Of that, only about half actually has protection. The amount of data that requires protection is projected to grow by a factor of 26X in the next six years. To ensure your data is protected and can't be accessed by unauthorized users, there must be strict security at the physical data center, network and application levels. Given that many big data use cases involve data sources outside of your direct control, the choice of cloud-based big data infrastructure can ensure a consistent level of data security regardless of the data's point of origin.

5. A Manageable Big Data Ecosystem

The ConsumerGoods.com survey revealed an interesting tension. While 32% of respondents said they wanted to deploy big data on-premise and 24% prefer a hybrid of managed hosting and on-premise, 53% also stated that they don't have the people to manage such systems. That's not surprising. Big data management skills are new and still fairly rare. There aren't enough skilled people to go around. What can be done about this?

Cloud platform management tools can give infrastructure managers the ability to handle big data efficiently. As exemplified by CenturyLink Cloud's "blueprint" capability, the right cloud platform enables a user to establish identical big data environments with relatively little setup time. Advanced infrastructure management tools complement the blueprint, making it possible for managers to stay on top of shifting big data configurations to ensure performance and economical utilization of infrastructure.

Conclusion: Cloud is a Compelling Solution to Big Data Challenges

Big data is happening. It's happening in the real world, in your organization. The challenge is to develop a big data capability that will serve your business needs while remaining economical and flexible. Getting there involves finding the right mix of infrastructure, software, and people.

Then, while on task to create the required big data capability, IT managers need to focus on infrastructure that can scale elastically but not be overly complex to manage and secure. The cloud presents a compelling solution to this bundle of big data challenges.

CenturyLink Cloud and Big Data

CenturyLink Cloud specializes in running enterprise apps securely and reliably, 24/7. It offers Hyperscale, an instance type designed for distributed workloads that require maximum performance. With Hyperscale, CenturyLink Cloud delivers the top-tier CPU and RAM performance found in our standard servers while adding 100% flash storage to turbo-charge applications with at least 15,000 IOPS. Hyperscale is ideal for big data and other compute-intensive workloads.

CenturyLink Cloud can deploy and administer Hyperscale instances using our built-in cloud management, automation, and orchestration capabilities. This helps you keep your business running smoothly while significantly reducing your operational burden.

About CenturyLink Cloud

CenturyLink Cloud is the complete platform to easily manage your entire business application portfolio, from development to business-critical workloads. CenturyLink Cloud offers high-performance, scalable, self-service virtual machines across our global network of data centers, including Hyperscale instances for distributed workloads that require maximum performance. And CenturyLink Cloud provides built-in automation, orchestration, and management tools for an IT-ready and developer-friendly platform that is flexible, scalable, cost effective and highly manageable.

For more information, visit www.centurylinkcloud.com.

About CenturyLink Business

CenturyLink Business delivers innovative managed services for global businesses on virtual, dedicated and colocation platforms. It is a global leader in cloud infrastructure and hosted IT solutions for enterprise customers. Parent company CenturyLink, Inc. is the third largest telecommunications company in the United States, and empowers CenturyLink Business with its high-quality advanced fiber optic network. Headquartered in Monroe, LA, CenturyLink is an S&P 500 company and is included among the Fortune 500 list of America's largest corporations.

For more information visit www.centurylink.com/technology.

1 THE DIGITAL UNIVERSE IN 2020, IDC, December 2012. http://idcdocserv.com/1414

Global Headquarters
Monroe, LA
(800) 728-8471

EMEA Headquarters
United Kingdom
+44 (0)118 322 6000

Asia Pacific Headquarters
Singapore
+65 6591 8824

CenturyLink® Business